Revision of the manuscript: "Wave extremes characterization using Self-Organizing Maps"

Barbariol, Francesco; Falcieri, Francesco Marcello; Scotton, Carlotta; Benetazzo, Alvise; Carniel, Sandro; and Sclavo, Mauro

MS No.: os-2015-63

January, 12th 2016

Dear Sirs,

enclosed you can find the changes to our paper following the comments of the Referees. Please, consider that the title has been changed from "Self-Organizing Maps approaches to analyze extremes of multivariate wave climate" to "Wave extremes characterization using Self-Organizing Maps".

The Referees' comments are written in black, our answers and comments are written in blue color. We have reported major changes to the manuscript within the replies (in *italics*), while for minor changes you are invited to refer to the marked-up manuscript version attached.

We would like to thank the Referees because of their comments and suggestions, which have been useful to improve the original manuscript.

Sincerely yours,

Francesco Barbariol
Institute of Marine Sciences - National Research Council
Venice, ITALY

Response to Referee #1

This work presents a very interesting approach to improve the existing tool to characterize and visualize multivariate wave climate (Camus et al 2011), introducing an extra step to put the eye (zooming) in specific areas of the trivariate space (Hs,Tm,Dir). In particular, the authors focus in extremes of Hs. I am very familiar with these techniques, and one of the most important issues is the correct definition of extremes and this work contributes to a better knowledge of extremes. However, it would be a pity not showing in this work the potentiality to describe extremes of a variable which is a function of the initial variables (Hs,Tm,Dir). For instance, the authors could show the ability of the proposed method to describe the alongshore component of energy flux (Fy) in shallow water (the authors could use some of the concepts explained in Reguero et al (2013) to propagate wave climate to local areas to shallow water), and use as objetive function local Fy. In my opinion, this added work would improve the quality of the work enormously.

We agree with the Referee that showing a potential application of the proposed technique could be interesting for the readers. Therefore, following the Referee's advice we have included an additional section in the revised manuscript providing an application of TSOM to the estimate of the alongshore component of the wave energy flux. The manuscript has been modified accordingly in the "Abstract", "Introduction" and "Conclusions" sections (see attached marked-up manuscript version). Section 6, called "Application of TSOM" and including a new figure (Figure 14), is the following:

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):

$$P = Ec_g \sin\alpha \cos\alpha \tag{1}$$

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 $\overline{P}_-$.*

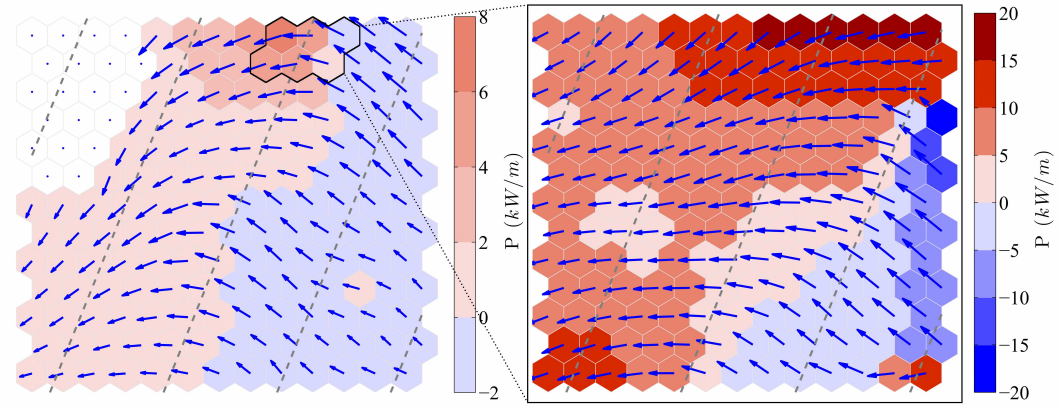

Figure 1: *Application of TSOM: assessment of the longshore flux of wave energy in shallow-water P. Mean wave directions at AcquaAlta tower (blue arrows) indicate contributions of different meteorological conditions: positive mainly due to Bora ($180 \leq \theta_m \leq 270°N$), negative to Sirocco ($270 < \theta_m \leq 360°N$). Land wind events (white cells) have been excluded, and the direction of the shoreline ($270°N$) is shown as gray dashed lines.*

Response to Referee #2

The paper 'Self-Organizing Maps approaches to analyze extremes of multivariate wave climate' by Barbariol, Falcieri, Scotton, Benetazzo, Carniel and Sclavo, deals with a SOM mapping technique, to assess and describe the multivariate sea climate. The analysis is performed on a long wave time series of data, collected off shore Venice at the Acqua Alta oceanographic tower. The multivariate sea state presentation is a key topic in the OS, the technique here proposed (SOM) is well supported by the literature and is well presented. The analysis is performed on the entire data set and on several sub set of data, and the differences are discussed. The paper is suitable for publication on Ocean Sciences after several small discussion and correction on the actual version of the paper.

- Some other methodologies and techniques are used in the description and estimation of joint probability dependence among variables. Of particular interest is the use of copulas. Several application on this application on sea state variables can be found in literature, i.e. De Michele et al., 2007, (De Michele, G. Salvadori, G. Passoni, R. Vezzoli. A multivariate model of sea storms using copulas. Coastal Engineering, 54, (10), 2007, 734-751) and more recently in Masina et al., 2015 (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, 2015, 37-52). A comment is welcome.
  We have appreciated the Referee's suggestion, because it has given us the opportunity of enriching the description of the theoretical framework and of the alternatives to estimate joint probability distributions. We have indeed included a comment in the "Introduction" of the revised manuscript about the possibility of estimating joint probability distributions using copulas and the reference to the papers indicated by the Referee:

  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.

- More in detail, the authors select and discuss the results on several storms, for example considering the first data set and SOM application (page 1981-1982), the authors discuss the storm with Hs= 4.46 m, 6.7 s, 275N; and results are presented in Fig.5 (and later in Fig. 8 and Fig. 11). Can the authors give more information on the date of the events? Why this event is here discussed and how was selected? What #118 BMU refers to?
  The series of events shown in Figure 5, Figure 8 and Figure 11 occurred in the period from November 26th 1983 15.00UTC to December 14th 1983 03.00UTC. In the revised manuscript the time axis now shows the days of the year 1983, to allow an easier identification of the events. The modified Figure 5 now appears as (please refer to the marked-up manuscript attached for Figures 8 and 11):

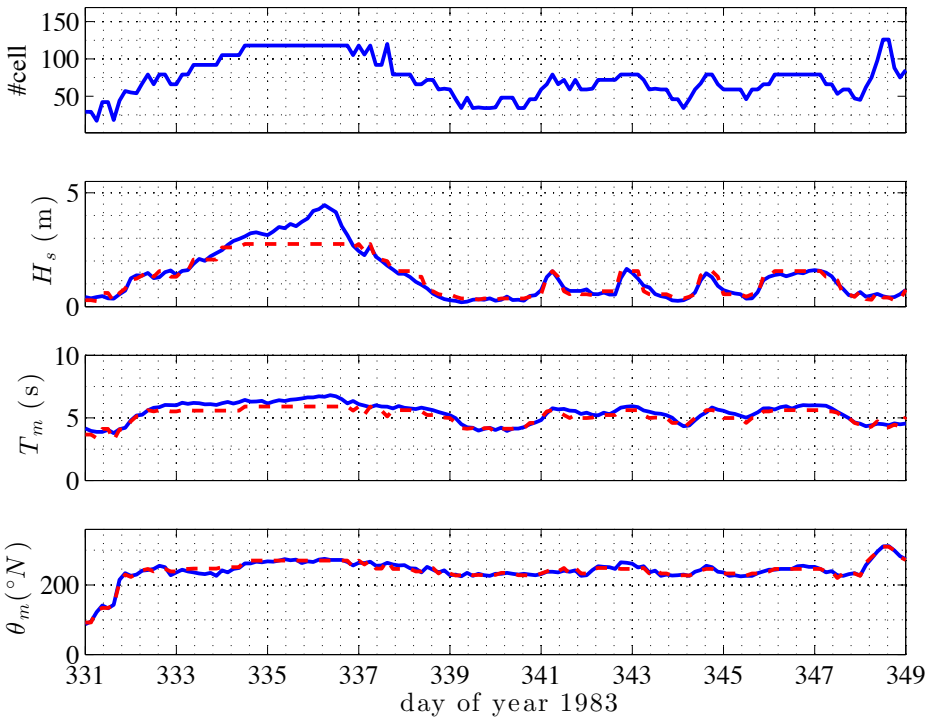

Figure 2: *Single-step SOM: BMU cells (top panel) and comparison between original (blue solid lines) and reconstructed (red dashed lines) time series of H_s (central-top panel), T_m (central-bottom panel) and θ_m (bottom panel), for a chosen sequence of events.*

The series has been chosen because it was representative of sea states well reproduced by the traditional SOM technique (i.e. frequent and low/moderate H_s) and, at the same time, of sea states that are not properly reproduced due to the limitations pointed out in the SOM technique (i.e. extreme H_s). The

#118 BMU refers to one of the 169 (=13x13) BMUs composing the output map and representing the analyzed sea states. In particular, #118 refers to the position occupied by the BMU having the highest H_s, which indeed is placed at the first row and tenth column (the BMUs' 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). We thank the Referee, because actually this was not properly explained in the first submission of the manuscript. Hence, we have provided a better explanation in the revised version:

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}.

- In Fig. 10 and Fig. 13 the right panels present the map for the highest values, so for the extreme events. I suggest to create new color palettes for the right panels, both for Hs and %, in order to better present the results for the extremes.
  We thank the Referee for the suggestion, but we have preferred to keep the current color palettes (for H_s and %) in both the left and right panels of Fig. 10 and 13. Indeed, although SOM output maps can potentially provide a lot of information (e.g. H_s, T_m, θ_m and F), we want to keep the readability of the maps as easy as possible. We have verified that using a dedicated color palette for frequency in the right panels of Fig. 10 and 13 worsens the readability of the map, mainly because of the non-continuous values of frequency associated to the extreme sea states. Indeed, the SOM technique is applied to the wave parameters, hence the topological ordering is principally achieved on these, rather than on frequency which is a derived quantity. At the same time, he have preferred to keep the same palette for H_s too, because we want to stress the difference between the maps, and highlight that the right one is the extension of the left one.

- Line 9 page 1981: 0.36 s, seems to be a mistake, and the correct value is 3.6 s, please review this value.
  Yes, it is definitely a mistake. Thanks for having noticed it. We have corrected it in the revised manuscript.

Response to Referee #3

The study applies Self-Organizing Maps (SOM) method to long term wave measurements in the northern Adriatic, describing multivariate sea wave climate through three different approaches which enable better representation of extreme states. Namely, as SOM technique is not efficient where the density of events is low, the authors introduce an extra step for extreme wave states and discuss differences and benefits of varying strategies in representing the extreme wave climate. General comments: I find the paper appropriate for this special issue of Ocean Science journal, with a valuable dataset presented and an interesting method and strategies proposed and discussed to better represent multivariate wave climate in the north-

ern Adriatic. The analysis is based on the SOM method, which is described in appropriate manner, particularly focusing on the SOM publications dealing with the multivariate wave climate.

- However, the Introduction (page 1973) should include broader view of the SOM in climatology, oceanography, climatology,...and present/cite at least some important SOM oceanographic publications outside of the Adriatic.
  We agree with the Referee, therefore in the revised manuscript the "Introduction" has been enriched with reference to SOM applications over different oceanic contexts (e.g. Morioka et al. 2010).

- Page 1975, line 15: please use kilometers instead of miles, particularly since it is not clear if these are nautical miles or not.
  In the submitted manuscript the distance of the Acqua Alta tower from the coastline was expressed in nautical miles. However, we have accepted the Referee's suggestion and in the revised manuscript we have written that "Acqua Alta" tower is located approximately 15 km off the Venice coast.

- Page 1975, line 25-: The direction of waves in the Table 1, and throughout the text, tables and figures is given in oacenographic convention. This is not obvious for wave direction in general, and the authors should emphasize this in the text, that the angles given correspond to the direction (from North) the waves are propagating towards.
  The mean wave direction θ_m is defined at page 1975, line 20 as mean direction of wave propagation. In the revised manuscript, we have added the convention we used throughout the text, i.e. the convention indicating the direction of wave propagation from the geographical North.

- The authors should also give more information related to the extreme Hm (5.23 m), by detailing when did the event occur, giving corresponding wave period and direction and related meteorological/wind characteristics that caused this extreme event.
  In the revised manuscript, we have added more information about the most extreme observed sea state (i.e. H_s = 5.23 m) that occurred on December 9th 1992 at 00.00UTC. In this respect, the complete triplet of the sea states is $\{H_s,T_m,\theta_m\}$=$\{5.23$ m, 5.36 s, 242°N$\}$, suggesting that it verified during a strong Bora storm. Unfortunately, we could not add more details about meteorological conditions related to the event since they are not available. Details about the most extreme events have been included in the revised manuscript ("Data" section) as:

  the most intense event (H_s = 5.23 m, T_m = 5.36 s, θ_m = 242°N) occurred on December 9th, 1992 during a storm forced by winds coming from north-east.

- Page 1976, line 28: three extreme states are mentioned, however, Hm=5.23 m is missing both in figure and in the text?
  The three sea states mentioned in the text are obtained from the bivariate (H_s-T_m) diagram. These sea states, therein represented by $\{H_s,T_m\}$ pairs, are not sea states that have been necessarily observed. Indeed, to build the diagram, the ranges of observed H_s and T_m have been first divided in classes, and then each sea state was attributed to a proper class. Hence, the $\{H_s,T_m\}$ we mentioned are centroids of the classes that are representative of the most

extremes sea states. According to the bivariate experimental statistics, the most extreme sea state belongs to the class represented by $\{H_s,T_m\}$=$\{5.20$ m, 5.38 s$\}$, which is not colored in Figure due to its very low frequency of occurrence (the sea state in discussion is the only one in this class). The white color has been added to the color palette in order to indicate the lowest frequency pairs (please, refer to the Figure reported in the next reply), and details on the event have been added in the "Data" section (as shown in the previous reply).

- Does the difference in period between these extreme events correspond to different winds at the time (Bora – shorter periods, Sirocco – longer periods)? Which directions correspond to these pairs?
  Yes, this is correct, i.e. shorter periods are generally associated to Bora events, longer periods to Sirocco events. Actually, we have split the (H_s-T_m) diagram in Figure 3 into 2 diagrams (one related to Bora and one related to Sirocco) in order to classify $\{H_s,T_m\}$ pairs that pertain to similar directions of propagation within the western quadrants. Wave periods corresponding to such directions are compared in Figure 3 (bottom-right panel) and it can be seen that Bora is indeed associated to shorter periods than Sirocco. The modified Figure 3 is:

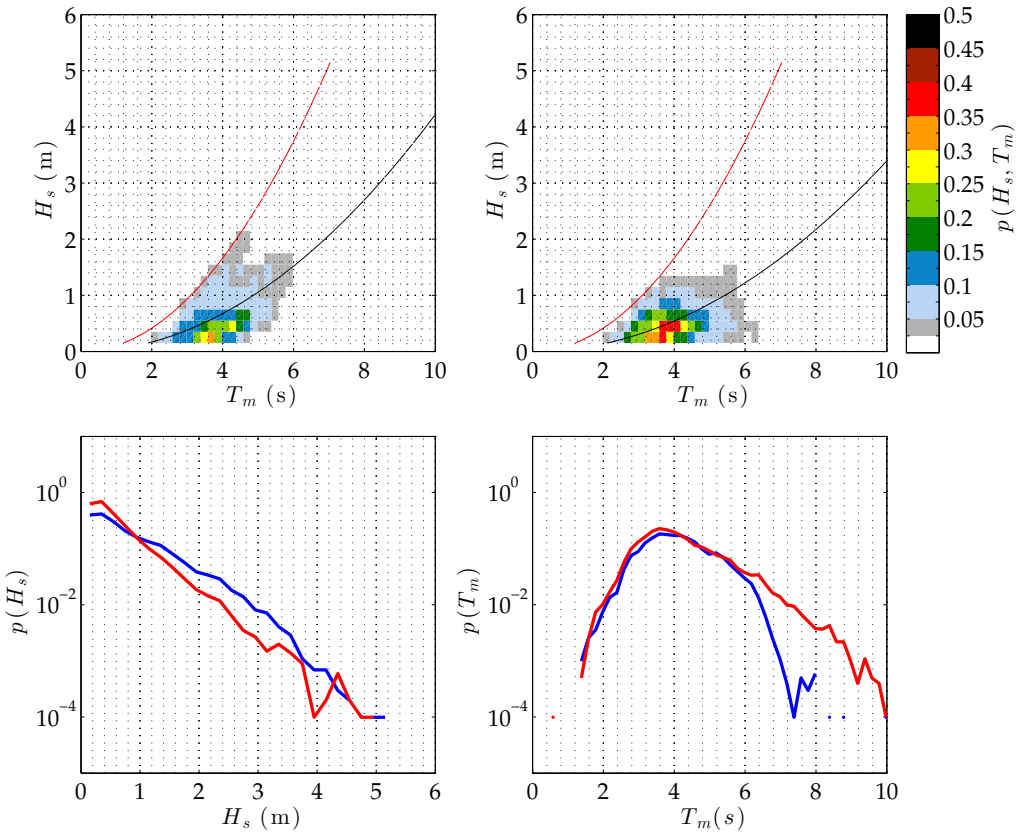

Figure 3: *Observed bivariate wave climate at AcquaAlta: joint pdf of H_s and T_m for Bora (top-left panel) and Sirocco (top-right panel) sea states, and corresponding marginal pdfs of H_s (bottom-left panel; blue for Bora, red for Sirocco) and T_m (bottom-right panels; blue for Bora, red for Sirocco). 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.*

- Page 1980, line 21: Fig. 4 and similar figures (Figs. 7, 10, 13, where wave period is presented by the vector) should include a unit vector (in seconds), so that for each BMU the period could also be relatively easily deduced from the figure.
  We thank the Referee for this comment. We have added a reference vector (10 seconds) for mean wave periods to make the map reading easier (please, refer to the marked-up manuscript attached).

- Page 1981, line 9: 0.36 seconds? Do you mean 3.6 s?
  Yes, definitely. Thanks for having noticed it. We have corrected the typo in the revised manuscript.

- Page 1981, line 24: please provide more information about the event (storm) depicted in Fig. 5. Temporal axes should include exact time of the event (also in Figs. 8 and 11) and more details about meteorological conditions related to it should be given in the manuscript.
  The series of events shown in Figure 5, Figure 8 and Figure 11 occurred in the

period from November 26th 1983 15.00UTC to December 14th 1983 03.00UTC. In the revised manuscript the time axis now shows the days of the year 1983, to allow an easier identification of the events. The modified Figure 5 now appears as (please refer to the marked-up manuscript attached for Figures 8 and 11):

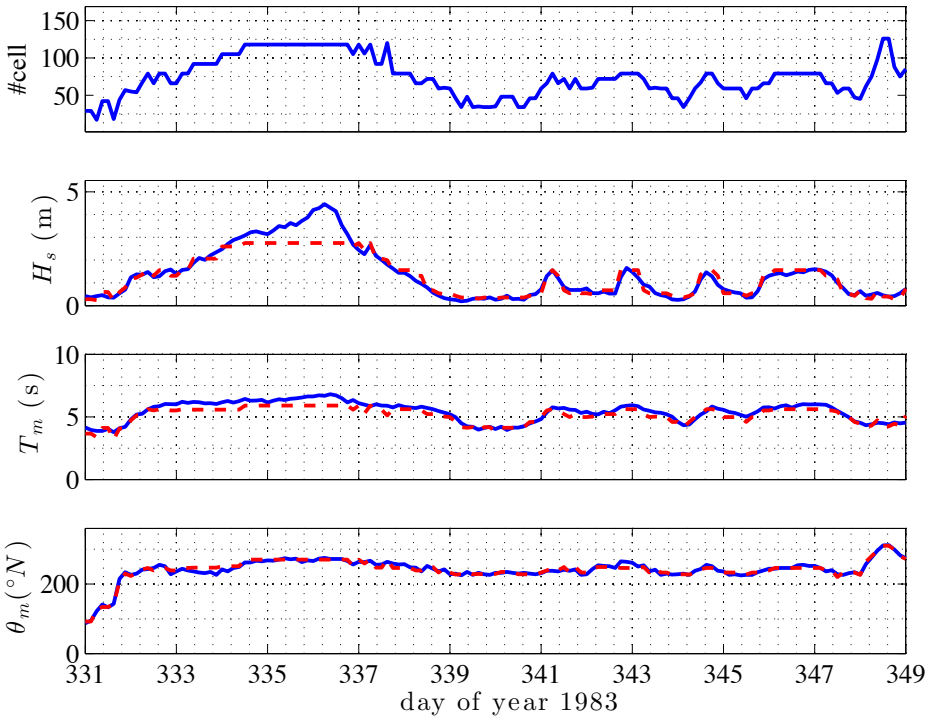

Figure 4: *Single-step SOM: BMU cells (top panel) and comparison between original (blue solid lines) and reconstructed (red dashed lines) time series of H_s (central-top panel), T_m (central-bottom panel) and θ_m (bottom panel), for a chosen sequence of events.*

The series has been chosen because it was representative of sea states well reproduced by the traditional SOM technique (i.e. frequent and low/moderate H_s) and, at the same time, of sea states that are not properly reproduced due to the limitations pointed out in the SOM technique (i.e. extreme H_s). We could not add details about the meteorological conditions related to the events since they are not available.

- Page 1983, line 11: is it "40% error on 99th percentile Hs" or "4% error on 99th percentile Hs", Table 2 says it is 4%?
  Yes, the value in Table 2 (i.e. 4%) is correct. We mistook in reporting the correct value in the text. Thanks for having noticed it.

- Page 1984, lines 18-28: the left side map in Fig. 10 is the same as the map in Fig. 4 (and the same as the left side map in Fig. 13) , i.e. it gives the SOM representation of the entire dataset, not only climate below Hs, since some

of the BMUs have Hs larger than the threshold prescribed? The line 20 on this page should be reformulated, as it says: "...the first map on the left side describes the climate below Hs...", and is not consistent with the rest of the paragraph. The figure caption is not clear either, as it says: "Wave climate below the threshold (left panel)...". Since the left panel also includes 6 BMUs with Hs above the threshold, than the left panel partly also describes climate above the threshold. Please check throughout the manuscript.
Yes, the Referee is right, and we thank her/him for giving us the opportunity to reformulate the sentences and the Figure captions, which are not clear as they appear in the submitted manuscript. Indeed, the left map describes the entire wave climate at Acqua Alta, not only the climate below the threshold (herein defined according to the 97th percentile of Hs), even if extremes are poorly described for the reasons in discussion. Hence, it provides a good representation of low/moderate sea states (excluding the six cells exceeding the threshold). The right map, instead, provides a more accurate description of the extreme wave climate, which is represented in the left map by the SOM cells encompassed by the black line.

- Page 1985, line 12: "Sirocco" events as discussed here are in bottom-right part of the right map in Fig. 10, not bottom-left?
  Yes, the Referee is right. The Sirocco events are in the bottom-right part of the right map of Figure 10, not bottom-left. We apologize for this mistake and thank the Referee for having noticed it. The revised manuscript has been corrected.

- Page 1985, lines 13-17: please discuss the most severe sea states along the diagonal, especially those with the highest Hs values. Apparently the largest number of extreme BMUs has direction towards 260-270. Still, some of them (to the lower-right in top-right corner) obviously correspond to Sirocco. This could be then related to extreme Hs-Tm pairs mentioned at the bottom of page 1976 (including the one that is missing, 5.23 m, see previous comment). Which directions correspond to those extreme pairs? What was the meteorological situation during those extremes? Please provide some details.
  It is not straightforward to assign directions of propagation to the $\{H_s,T_m\}$ pairs of Figure 3. In addition, we could rely on wave propagation direction only to discuss the meteorological conditions (e.g. Bora or Sirocco) corresponding to those extremes. To facilitate the discussion and isolate Bora and Sirocco (the prevailing conditions, especially during extremes), we have split (H_s-T_m) diagram in Figure 3 into 2 diagrams (one related to Bora and one related to Sirocco).

- Does the majority of most extreme wave states at the Acqua Alta station correspond to Bora winds, as 260-270 seems to be more related to Bora than Sirocco? Putting a limit between the Bora and Sirocco exactly at 270 could be somewhat confusing?
  Yes, the majority of most extreme wave states at Acqua Alta propagate towards the third and fourth quadrant, i.e. the propagation directions that pertain to Bora and Sirocco waves. More precisely, Bora is the northeastern (coming from) wind, and Sirocco is the southeastern wind blowing only along the main basin axis (i.e. approximately 315°N). The 270 limit, despite arbitrary, is conventionally adopted to discriminate between this two meteoro-

logical conditions. Therefore for the purposes of a quantitative classification, we have preferred to leave the 270° limit to discriminate between Sirocco (from 270° to 360°) and Bora (from 180° to 269°).

- Page 1985, Fig. 10: colorbar for the frequency of individual BMUs in the right panel is not really useful, as all the frequencies appear almost the same. A separate colorbar (different color pallets) should be related to the right-side map BMU frequencies, perhaps for Hs too. The same thing should be done in Fig. 13, although the color visibility is not as bad as in Fig. 10.
  We thank the Referee for the suggestion, but we have preferred to keep the current color palettes (for H_s and %) in both the left and right panels of Fig. 10 and 13. Indeed, although SOM output maps can potentially provide a lot of information (e.g. H_s, T_m, θ_m and F), we want to keep the readability of the maps as easy as possible. We have verified that using a dedicated color palette for frequency in the right panels of Fig. 10 and 13 worsens the readability of the map, mainly because of the non-continuous values of frequency associated to the extreme sea states. Indeed, the SOM technique is applied to the wave parameters, hence the topological ordering is principally achieved on these, rather than on frequency which is a derived quantity. At the same time, he have preferred to keep the same palette for H_s too, because we want to stress the difference between the maps, and highlight that the right one is the extension of the left one.

- Page 1987, line 13: is it completely correct/precise to tell that the left map/panel comprises only low/moderate sea states (also on page 1975, lines 1-10 where the authors propose a double-sided map). Namely, the most extreme BMUs correspond to the highest Hs values (when looking on temporal evolution of BMUs, e.g. Fig. 5 between 30th and 50th hour), but corresponding BMU reconstructions are relatively far from measurements due to the SOM characteristics/problems you described and emphasized in Introduction (rare events and distant from the others in multidimensional input data space)? Therefore the left side map represents the entire dataset, but it does not properly describe extreme sea states? That is what the authors essentially say at the top of page 1985: "Without such BMUs, the map on the left represents the low/moderate wave climate...". Please make sure to be consistent with this throughout the manuscript (e.g., Page 1988, line 23 again says that the left map represents low/moderate wave climate).
  The left map of Figure 13 (i.e. the result of a single-step SOM, also in Figure 4) comprises all the sea states that occurred at Acqua Alta in the period we analyzed. However, due to SOM tendency to better describe the more frequent and less intense sea states, the map in discussion cannot properly represent the most extreme sea states. As recalled by the Referee, this is shown for instance in Figure 5, where extreme sea states are not properly described by the cells composing the map of Figure 4. Hence, we have better explained these concepts in section 4 ("SOM strategies to characterize wave extremes") and in "Conclusions" of the revised manuscript (please, refer to marked-up manuscript attached).

Response to Referee #4

The paper "Self-Organizing Maps approaches to analyze extremes of multivariate wave climate (Barbariol et al.)" presents several interesting strategies to deal with extremes using SOM due to its visualization properties. The best solutions seem to be the TSOM and POT-SOM. However, the POT-SOM presents the disadvantage of the extreme data discontinuity in order to reproduce the time series and probability density. Both methods require working with two SOMs.

- Have been analyzed if a SOM of a higher size (e.g. 25 x 25, more clusters than two SOMs of 13x13) might detect extremes with a similar range of variation of Hs? A preselection using MDA to avoid many clusters in areas with high data density could have a more significant effect in a SOM of a higher size.
  To answer to the Referee's comment, the following text has been added to the "Discussion" section of the revised manuscript:

  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.

- One suggestion is not used "BMU" to refer to centroid, prototype. BMU is "the neuron whose weight vector is the closest to the input vector" during the training process and in the final classification. BMU is a term related to each input data.
  We thank the Referee for the suggestion that we have accepted. Hence, in the revised manuscript we have changed "BMUs" into "cells" when referring to the centroids (i.e. the elements) that constitute the SOM map.

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 Mapsapproaches to analyze extremes of multivariate wave climate

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. In this paper the The Self-Organizing Map technique (SOM) technique is considered and extended to assess the extremes of a multivariate sea wave climate at a siteis analyzed and discussed with the aim of . Main purpose is to obtain a more complete representation which includes the most severe sea of the sea states, including the most severe 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 if the frequency of events is high (e.g. for low/moderate and frequent sea states), while SOM fails where the density if the frequency is low (e.g. for severe and rare 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 the standard SOM have been tested in order to widen the range of applicability to capture the most extreme events. The first strategy contemplates a pre-processing of the input dataset with the Maximum Dissimilarity Algorithmaimed at reducing redundancies; the secondand the third strategies focus , based on the post-processing of SOM outputs, resulting consists 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 applied on the events exceeding a threshold (either taking all the events over the threshold or only the peaks of storms)given threshold. A complete graphical representation of the outcomes of two-steps SOM is proposed. Results suggest that the post-processing strategies are strategy is 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 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 outcomes oftwo-steps SOM as double-sided maps is proposed. 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 ~~the~~ human activities in the marine environment. ~~To this end, in~~ In the last decades, the observational network ~~, that mostly relies~~ (mostly relying on buoys, satellites and other probes~~from fixed platforms,)~~ has been integrated with numerical models outputs ~~, which allow to compute~~ 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 ~~hardly obtained~~not straightforward. For individual waves, for instance, the bivariate joint pdf ~~for~~ of wave height and period was derived by Longuet-Higgins (1983) and the bivariate joint pdf ~~for~~ of wave height and direction was obtained by Isobe (1988). A trivariate joint pdf ~~for~~ 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 ~~for~~ 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) ~~technique~~ has been successfully applied to represent the multivariate wave climate around the Iberian peninsula (Camus et al., 2011a, b) and the ~~South America~~ 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 ~~a~~ map. The SOM technique was originally developed in the 1980s, and ~~over three decades~~ 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 ~~properties~~ 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. ~~Hence,~~ Indeed, SOM is principally forced by 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~~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 ~~typically~~ 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, ~~individuation~~ identification of typical sea states for wave energy resource assessment and wave farm optimization. In addition the ~~experimental~~ 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~~ "competitive" stage of the SOM learning) and distant from the others in the multidimensional space of input data (poorly influenced during the ~~cooperative~~ "cooperative" stage).

Moving from this ~~fact we have asked~~ evidence, the scientific question being asked is: how can we employ SOM with its visualization ~~properties~~ capabilities to improve representation of the ~~edges~~ extremes 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 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, ~~the improved extremes representation has been searched by running SOM on a dataset of events exceeding a prescribed threshold. Hence,~~ we have applied a two-steps SOM approach (herein called TSOM), by firstly running ~~a standard~~ SOM to get a reliable representation of the low/moderate (i.e. the most frequent) wave climate, and then ~~we have run~~ by running a second SOM on a reduced input sample. This new sample has been obtained ~~in one case by taking the events that exceed~~ by taking from first-step SOM results the events exceeding a prescribed threshold (e.g. 97^{th} 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,)~~. To present results of two-steps SOMs, we ~~propose a double-side map, which shows~~ have proposed a double-sided map, showing on the left the SOM map with the reliable representation of the low/moderate sea states, ~~while~~ 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 ~~consist~~ 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)~~,~~. *Acqua Alta* is located in the northern Adriatic Sea (Italy, northern Mediterranean Sea), ~~8 miles~~ approximately 15 km off the Venice coast at 17 m depth (Figure 1) ~~. The tower~~ 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 ~~, that were measured by~~ (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 ~~, given in Table 1,~~ (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~~($max(H_s) = 5.23$ m)~~: 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 ~~quite rare~~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~~°N, indeed most of the waves propagate towards the western quadrants.

This is represented more in detail in the ~~experimental~~ pdf of θ_m (Figure 2, bottom panel), ~~where~~ which shows that the most frequent directions of propagation are indeed in the range 180< θ_m <360~~° N~~ °N (western quadrants), with peaks at 247.5 and 315~~° N~~°N. Directions associated with the most intense sea states ($H_s > 4.5$ m) can be ~~attained~~ obtained from the bivariate diagram ($H_s - \theta_m$) representing the ~~experimental~~ joint pdf of H_s and θ_m (Figure 2, ~~upper~~ top panel): 247.5, 270 and 315~~° N~~°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 ~~$H_s,\theta_m\{H_s,\theta_m\}$=\{0.25 m, 315°~~ N~~°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 "~~ Bora and Sirocco winds (Signell et al., 2005; Benetazzo et al., 2012). ~~"Bora "~~ 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 "~~ 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 "~~ Bora which come from the southern part of the basin. Denoting ~~as "Bora " conventionally as~~ Bora the events with $180 \leq \theta_m \leq 270$°N, and as ~~"Sirocco "~~ Sirocco the events with $270 < \theta_m \leq 360$°N, it ~~results that "Bora "~~ follows that Bora storms have an occurrence of 12%, while ~~"Sirocco "~~ Sirocco storms of 8%.

~~The bivariate wave diagram~~ The most frequent $\{H_s, T_m\}$ occurred in the Bora and Sirocco quadrants are shown in the bivariate $(H_s - T_m)$ ~~illustrates the experimental joint pdf of H_s and T_m~~ diagrams (Figure 3), ~~where it emerges that the most common sea states are represented by H_s, T_m=0.2~~ and are {0.15 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 marginal pdfs of~~ {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 ~~and T_m are also shown in Figure 3, confirming that the most probable~~ shows that Sirocco events with H_s ~~at *Acqua Alta* is 0.2 m, and the most probable T_m is 3.6 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 ~~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~~ 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 ~~the learning~~ functions and parameters that control learning is crucial to obtain ~~stable and~~ 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 ~~the~~ a 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 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~~as done by the original incremental SOM algorithm, where input data are presented~~ , i.e. presenting the input data one at a time to the map, ~~or batchwise~~ as done by the ~~more recent batch algorithm , where the whole~~ original incremental SOM algorithm. A more recent algorithm performs a batchwise learning, presenting the 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 2D 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, in order to achieve reliable maps, 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 have ~~adjacent first (the BMU) and second units~~ first and second best matching units adjacent 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~~ *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~~ 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, ~~SOM have produced outputs as~~ we have chosen bi-dimensional squared ~~lattices~~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. 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.~~ 1 during a single phase learning process.

Input data have been normalized ~~in oder~~ 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:

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

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:

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

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

4 ~~Results~~SOM strategies to characterize wave extremes

In this Section, results of the standard SOM approach ~~and~~ (applied one time, hence called single-step SOM) and results of the different strategies proposed to improve extremes representation are presented. The performances of ~~standard~~ single-step SOM, MDA-SOM and TSOM are assessed by comparing the wave parameters time series and their ~~experimental~~ empirical marginal pdfs to the time series reconstructed from the results of the different strategies and relative pdfs, respectively. ~~POT-TSOM~~ 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 ~~Standard~~ Single-step SOM

~~Standard~~ 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~~{H_s, T_m, ~~θ_m~~θ_m} triplet. Hence, one can have an immediate sight on the wave climate features and on the ~~experimental~~ empirical joint pdf thanks to ~~the powerful visual properties~~ visual capabilities of SOM's output. Gradual and continuous ~~changing~~ change in wave parameters over the ~~BMUs~~ 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~~ °N}, which substantially 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,~~ (Figure 3), though these two diagrams are not formally related one to the other. ~~The most of the BMUs~~ Most cells show wave propagation directions ~~that point~~ pointing towards the western quadrants, as also displayed in the joint and marginal pdfs of θ_m ~~. The BMUs~~ (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 ~~very~~ low frequencies of occurrence (individual and cumulated). The frequency of occurrence of calms is 80%, while that of ~~"Bora "~~ Bora storms is 12% and that of ~~"Sirocco "~~ Sirocco storms is 8% (using definition of calms, ~~"Bora " and "Sirocco "~~ 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 ~~3.0~~ 2.75 m. Indeed, the triplet with the highest H_s produced by

SOM is {2.75 m, 5.9 s, 270° ~~N~~°N}. However, Table and diagrams in Section 2 have shown that ~~at~~ ~~*Acqua Alta*~~ H_s can exceed 5.0 m ~~. Therefore~~ at *Acqua Alta*. Therefore, sea states with $H_s > 2.75$ m are represented by ~~BMUs~~ 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 ~~representative~~ corresponding BMU. In Figure 5 sea states with $H_s > 2.75$ m are represented by ~~BMU #118 (the one~~ 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~~°N}. Quantitatively, for this particular event, ~~standard~~ 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 ~~experimental~~ empirical 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 m to 0.17 — 2.75~~ [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~~ [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*~~ [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 ~~standard~~ 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 $\langle H_s \rangle$, ~~40%~~ 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 ~~BMUs~~sea states. However, the most frequent ~~BMU has H_s~~sea state has $\{H_s$, T_m, ~~θ_m~~$\theta_m\}$=$\{$0.28 m, 2.8 s, 328~~° N, which is more distant to~~ °N$\}$, which still resembles what have emerged from diagrams of Section 2, even if T_m is less in agreement with respect to ~~standard SOM, especially for T_m~~single-step SOM. Also, the ~~BMU~~ sea state with highest H_s has the triplet equal to $\{$2.8 m, 6.0 s, 275~~° N~~°N$\}$, 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, ~~for the sequence of events already shown, the~~ the comparison has been extended to the results of 60% MDA-SOM (smaller error on 99^{th} percentile H_s, see Table ~~7~~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 ~~($~~$\{$2.93 m, 5.8 s, 258~~° N)~~°N$\}$. 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~~ 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 ~~standard~~ 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 ~~standard~~ 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 "~~ Bora storms is 12% and that of ~~"Sirocco "~~ Sirocco storms is 7%. Hence, except for a minor change in the frequency of calms and ~~"Sirocco "~~ Sirocco events, the overall statistics resembles that one ~~observed at *Acqua Alta*~~ 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^* ~~was~~ has been fixed, and the ~~BMUs~~ 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 ~~BMUs~~ cells exceeding the threshold. Hence, a second SOM has been run on the new dataset (second step). ~~The SOM~~ Using the same 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~~ (Figure 10~~, for instance~~): the first map ~~on the left side describes the climate below H_s^*,~~ (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 ~~on the right side focuses~~ (right panel), focusing on the climate over H_s^*. Three thresholds ~~were~~ 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 ~~already~~ shown in Figure 4, representing the whole wave climate at *Acqua Alta*. On that map, the ~~BMUs~~ six cells with $H_s > 2.12$ m have been encompassed by a black line. Without such ~~BMUs~~cells, the map on the left represents the low/moderate ~~wave climate~~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 ~~BMUs~~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 ~~$249 < \theta_m < 299°N$~~$249 < \theta_m < 299°$N. Hence, according to SOM, the most severe sea states pertain to a rather ~~tight~~ narrow directional sector (50°) ~~making hard~~ hardly allowing to discriminate between ~~"Bora " and "Sirocco " events~~Bora and Sirocco conditions. A more detailed representation of such ~~extreme events~~ extremes is provided by the second map in Figure 10, on the right, where extreme ~~"Bora " and "Sirocco "~~ Bora and Sirocco events are more widely described by ~~BMUs~~cells. Indeed, a sort of diagonal (from the top-right corner to the bottom-left corner of the map) divides the ~~BMUs. "Bora "~~ cells. Bora events are clustered on the left of this diagonal (top-left part of the map), while ~~"Sirocco " events~~ Sirocco ones on the right of that (~~bottom-left~~ bottom-right part of the map). On the diagonal, ~~BMUs~~ 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 ~~$224 < \theta_m < 316°N$~~$224 < \theta_m < 316°$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 ~~the~~ 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 ~~yield~~ 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 ~~found also~~ also found in the marginal ~~experimental~~ 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 ~~tightest~~ 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 "~~ Bora storms is 11% and that of ~~"Sirocco "~~ Sirocco storms is 8%. Hence, except for a minor change in the frequency of calms and ~~"Bora "~~ Bora events, the overall statistics resembles that one observed at *Acqua Alta*.

4.4 ~~Comparison among strategies~~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 ~~we tested until here~~ is given in Table 3. There, ~~standard~~ 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 ~~BMUs.~~ 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 ~~one~~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). ~~Though~~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.

6 ~~Peak-Over-Threshold Two-steps SOM (POT-TSOM)~~

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.

~~As an additional strategy to provide a more complete representation of the wave climate through SOM~~

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):

$$P = Ec_g \sin\alpha \cos\alpha \tag{3}$$

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), ~~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 the applied to the peaks of the storms defined by means of Peak-Over-Threshold technique. Storms were identified according to the definition of : a storm is the sequence of H_s that remains at least 12 hours over a given threshold H_s^* that corresponds to 1.5 times the mean~~ 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. ~~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.POTH-SOM 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~~ The longshore component of the wave energy flux P at 5-meter depth is shown in Figure ~~13. As expected stormy events are "Bora " and "Sirocco " events : the former are~~ 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 ~~upper and middle part of 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 ~~, the latter in the lower part of it. The most severe storms, concentrated on the right of the map , are both "Bora " and "Sirocco "~~ of Figure 10, the sea states exceeding the 97^{th} percentile threshold on H_s are Bora and Sirocco events. The ~~triplet~~ 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 ~~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~~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. ~~4.28 m~~ 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 $\overline{P}_-$. ~~Hence, 99% of the stormy events are included within the represented range, resembling what observed for the original dataset and standard SOM.~~

7 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~~ improving the characterization of multivariate wave climate 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~~ single-step SOM approach misses the most severe sea states, which are hidden in ~~BMUs~~ SOM cells with H_s even considerably smaller than the extreme ones. ~~For instance, standard SOM classified as 2.75 m, 5.9 s, 270° N~~

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 ~~a few BMUs which hardly allow to discriminate between "Bora " and "Sirocco " events.~~ 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 ~~standard SOM results~~ 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, ~~a wider range of the wave parameters within BMUs was obtained, in particular for H_s. This allowed a~~ TSOM allowed a more accurate and complete representation of the sea states with ~~TSOM more accurate and complete with~~ respect to the one furnished by MDA-SOM, ~~both for~~ because it provided a wider range of the wave parameters (particularly H_s), and more reliable a posteriori reconstructions of 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~~capabilities of SOM. In this context, we proposed a double-sided map representation~~for two-steps SOM strategies, which summarizes~~ , which provides on the left ~~the~~ a description of the whole wave climate that is particularly reliable for the low/moderate ~~wave climate and~~ events, completed on the right ~~the most severe one~~by the description of the extreme wave climate. This novel representation was ~~employed also to show the results of the POT-TSOM, which classified the storms peaks at *Acqua Alta* tower~~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. ~~Statistics of the 1979-2008 wave parameters dataset~~ Wave climate at *Acqua Alta* ~~tower~~ in the period 1979-2008. Mean ($\langle - \rangle$~~: mean~~), ~~*std:*~~ standard deviation (std), ~~*min:* minmum~~minimum (min), ~~perc:~~ p^{th} percentile (p^{th} perc), ~~*max:*~~ and maximum (max) of wave parameters.

	$\langle - \rangle$	~~*std*~~ std	~~*min*~~ min	50^{th} perc	95^{th} perc	97^{th} perc	99^{th} perc	~~*max*~~ 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~~°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}: ~~time series~~ ratio of time series averages, r_{std}: ~~time-series~~ 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}	$RMSE_{pdf}$
~~Standard~~ Single-step SOM	0.98	0.91	0.95	0.18	[0.17~~-2.75~~,2.75]	1.00	0.04
MDA-SOM (80%)	1.00	0.90	0.95	0.19	[0.21~~-2.82~~,2.82]	0.99	0.04
TSOM (97^{th} perc)	0.99	0.95	0.96	0.16	[0.17~~-4.26~~,4.26]	1.00	0.04
T_m	r_{av}	r_{std}	CC	$RMSE$ (s)	range (s)	CC_{pdf}	$RMSE_{pdf}$
~~Standard~~ Single-step SOM	1.00	0.89	0.95	0.34	[2.4~~-7.4~~,7.4]	0.99	0.02
MDA-SOM (80%)	1.00	0.90	0.95	0.37	[2.4~~-7.4~~,7.4]	0.95	0.05
TSOM (97^{th} perc)	1.00	0.90	0.95	0.32	[2.4~~-8.3~~,8.3]	0.99	0.02
θ_m	r_{av}	r_{std}	CC	$RMSE$ (°N)	range (°N)	CC_{pdf}	$RMSE_{pdf}$
~~Standard~~ Single-step SOM	1.00	0.92	0.95	23	~~41-323~~ [41,323]	0.97	0.00
MDA-SOM (80%)	0.99	0.95	0.96	20	~~30-328~~ [30,328]	0.98	0.00
TSOM (97^{th} perc)	1.00	0.92	0.95	23	~~41-323~~ [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

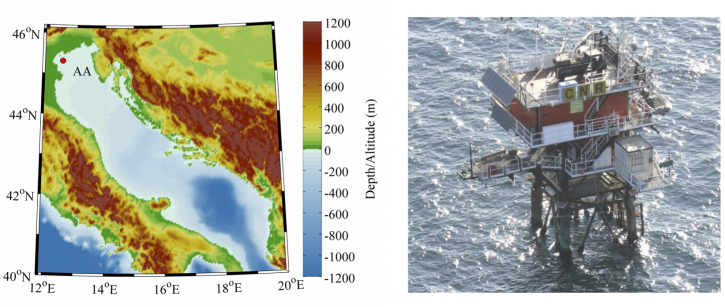

Figure 1. *Acqua Alta* (AA) oceanographic tower location in the northern Adriatic Sea, Italy (left panel). The tower is depicted in the right panel.

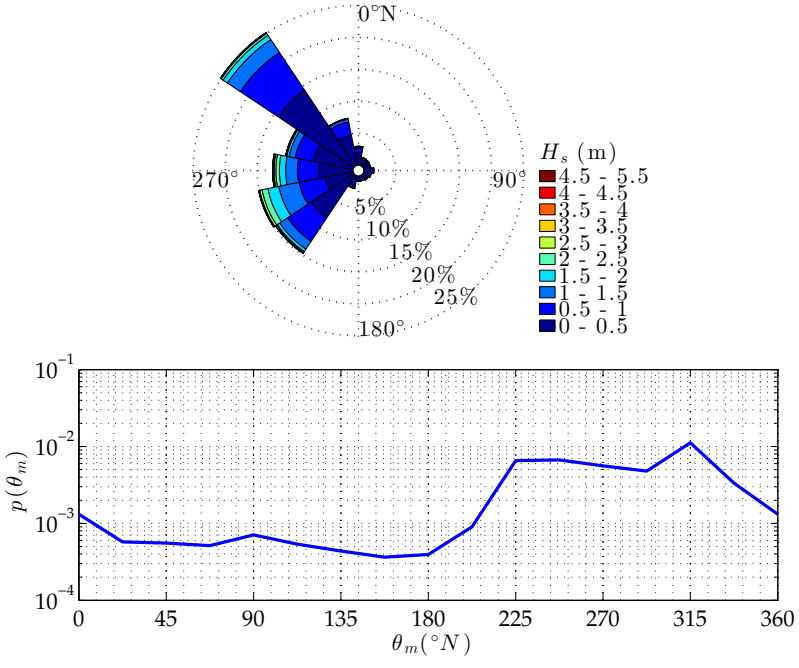

Figure 2. ~~Bivariate~~ Observed bivariate wave climate at *Acqua Alta*~~tower representing the~~: joint ~~(~~pdf of H_s ~~and~~ θ_m ~~) experimental pdf~~ with radially distributed frequencies (top panel)~~; resolutions are $\Delta H_s = 0.5$ m~~, and ~~$\Delta\theta_m = 22.5^\circ$~~. Experimental~~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

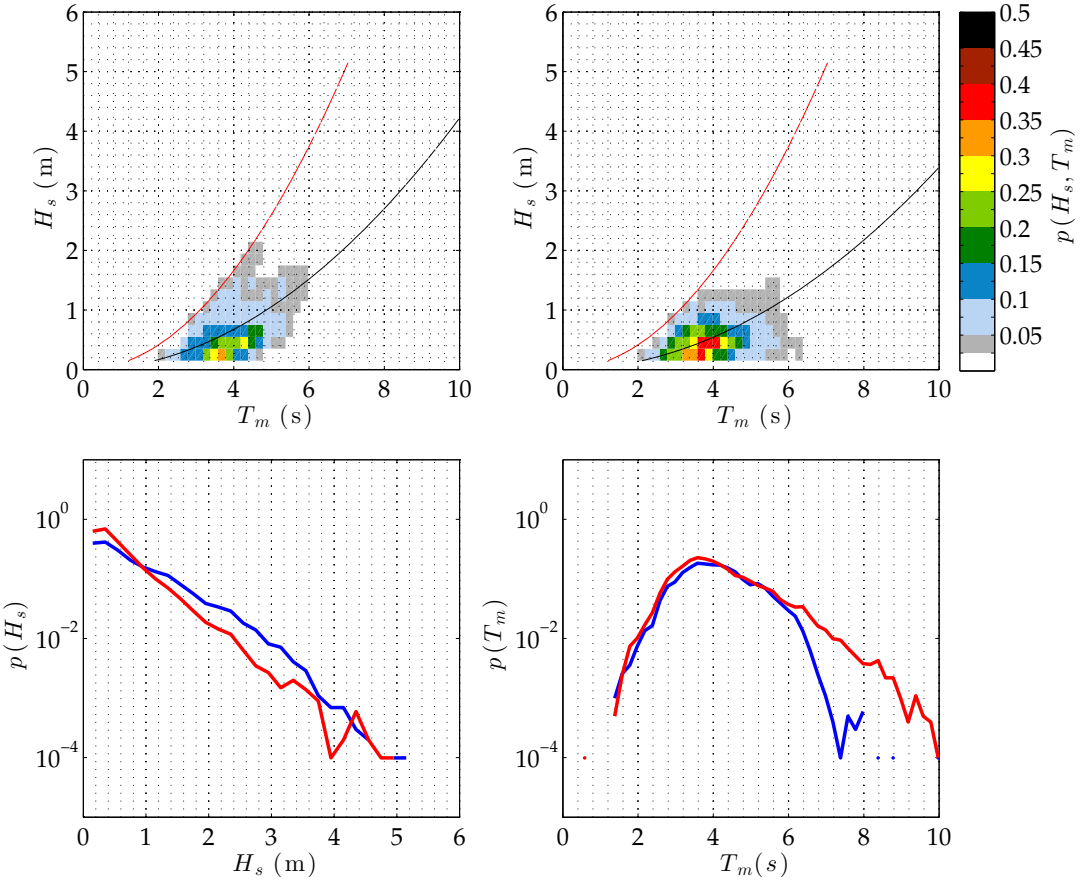

Figure 3. ~~Bivariate~~ Observed bivariate wave climate at *Acqua Alta*~~tower representing the~~ : joint ~~(~~pdf of H_s ~~-~~and T_m for Bora (top-left panel) ~~experimental pdf~~ and Sirocco (top-right panel) ~~; resolutions are $\Delta H_s = 0.1$ m~~ sea states, and ~~$\Delta T_m = 0.2$ s. Experimental~~ corresponding marginal pdfs of H_s (bottom-left panel; blue for Bora, red for Sirocco) and T_m ~~pdfs at *Acqua Alta* tower (top-left and~~ bottom-right panels; blue for Bora, ~~respectively~~red for Sirocco). ~~Solid~~ Black solid lines in the ~~top-right panel~~ top panels denote average wave steepness $2\pi H_s/g/T_m^2$ ~~equal to 2% (average observed~~3% for Bora, ~~black) and 72~~% ~~(breaking limit~~for Sirocco, ~~red),~~ 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.

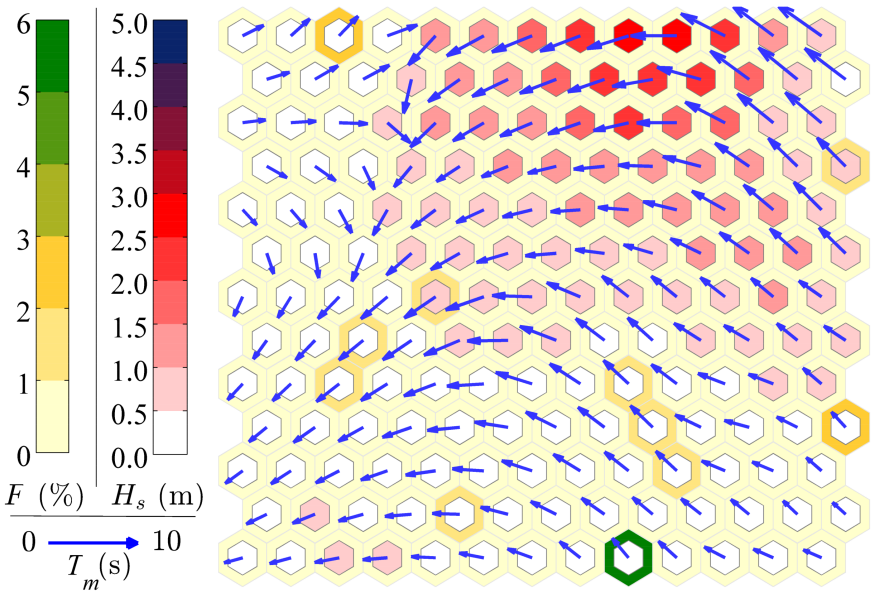

Figure 4. ~~Strandard~~ Single-step SOM output map. H_s: inner hexagons color, T_m: vectors' length, θ_m: vectors' direction, F: outer hexagons color. Mean quantization error: 0.06; topographic error: 22%.

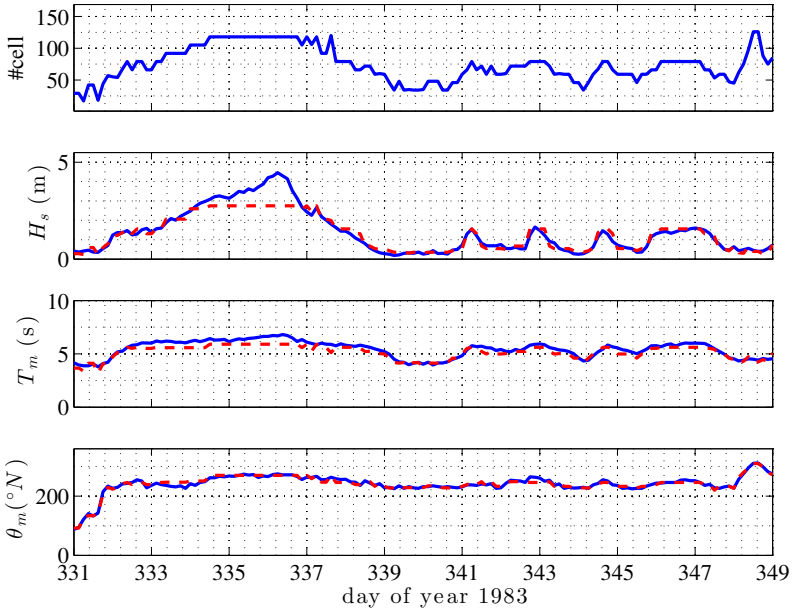

Figure 5. ~~Strandard~~ Single-step SOM: BMU cells (top panel) and comparison between original (blue solid lines) and reconstructed (red dashed lines) time series of H_s (central-top panel), T_m (central-bottom panel) and θ_m (bottom panel), for a chosen sequence of events.

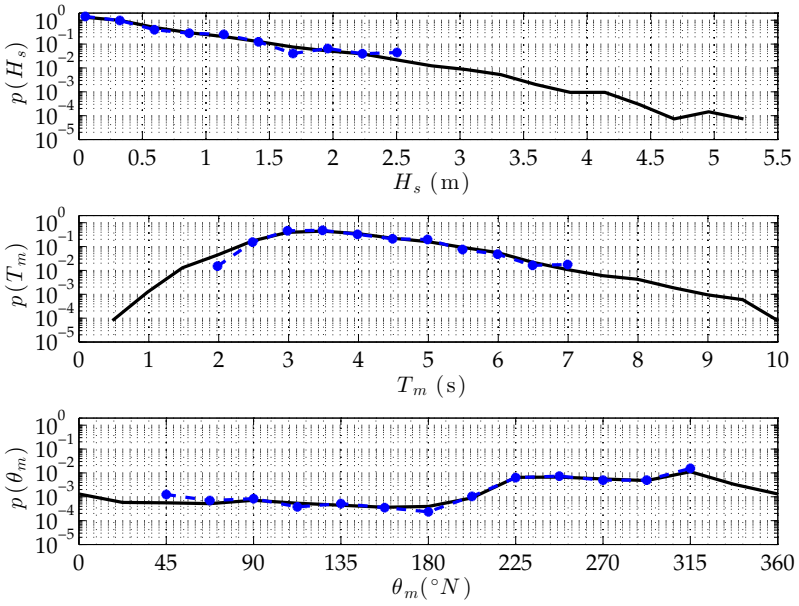

Figure 6. ~~Strandard~~ Single-step SOM: comparison of original (black solid line) and resulting (blue dashed-circles) pdfs of H_s (top panel), T_m (central panel) and θ_m (bottom panel), for the whole dataset.

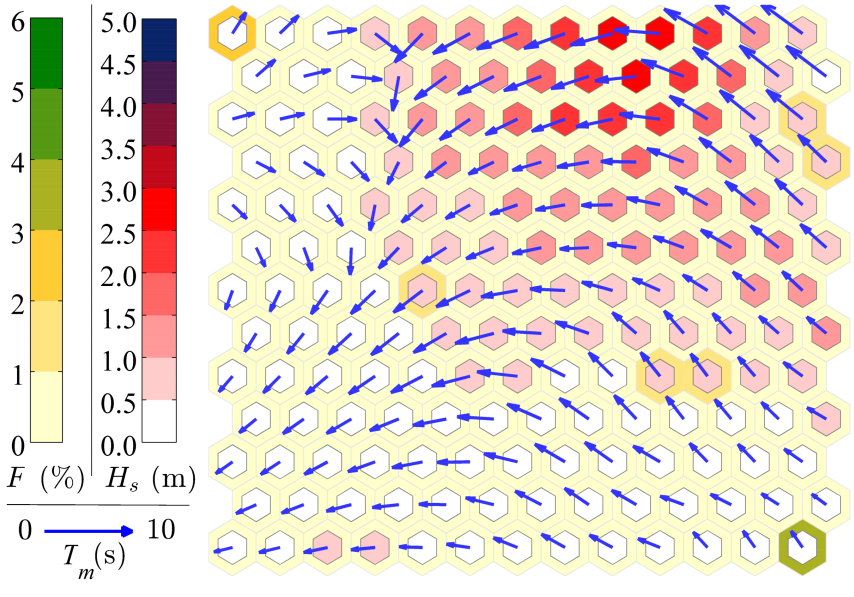

Figure 7. MDA-SOM output map, 80% reduction of the original dataset. H_s: inner hexagons color, T_m: vectors' length, θ_m: vectors' direction, F: outer hexagons color. Mean quantization error: 0.06; topographic error: 15%.

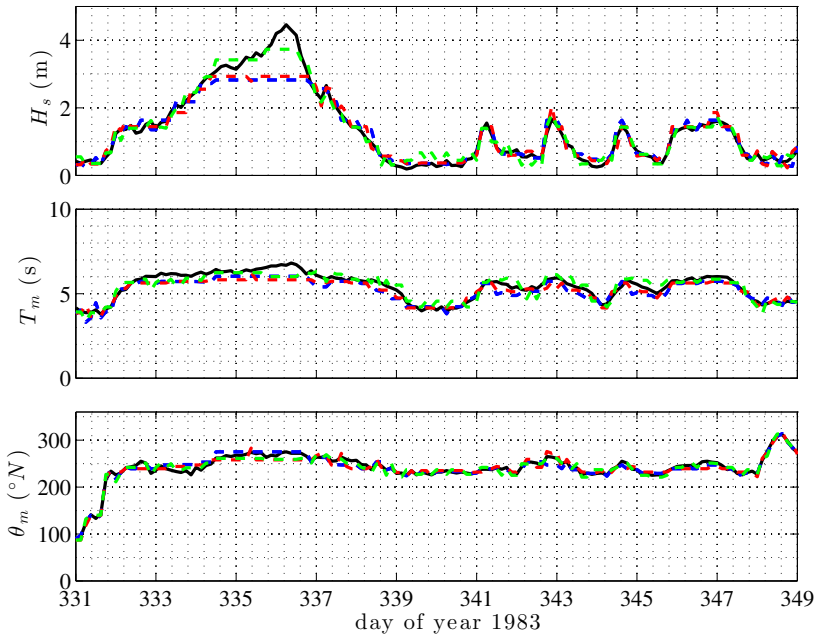

Figure 8. MDA-SOM: comparison between original (black solid lines) and reconstructed time series of H_s (top panel), T_m (central panel) and θ_m (bottom panel), for a chosen sequence of events. Dataset reduction: 80% (blue dashed line), 60% (red dashed line) and 10% (green dashed line).

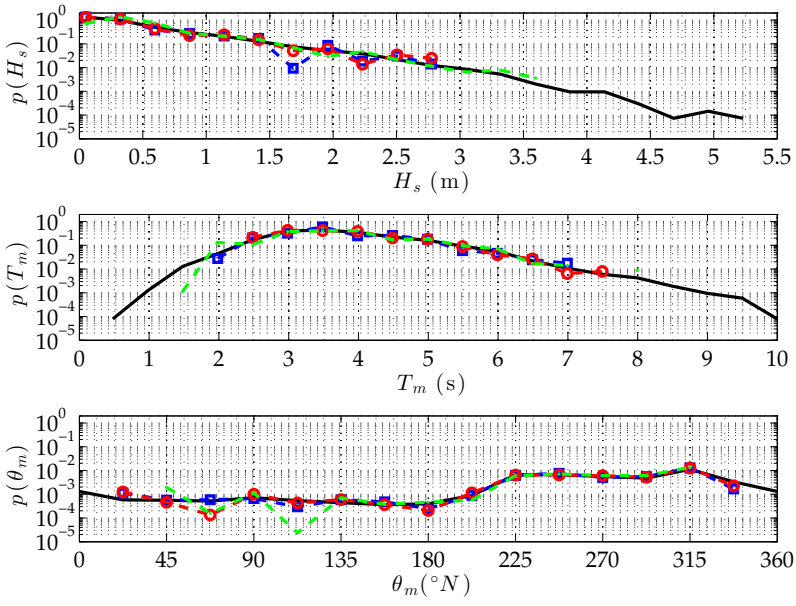

Figure 9. MDA-SOM: comparison between original (black solid lines) and resulting pdfs of H_s (top panel), T_m (central panel) and θ_m (bottom panel), for the whole period of observations. Dataset reduction: 80% (blue dashed-squares line), 60% (red dashed-circles line) and 10% (green dashed-dots line).

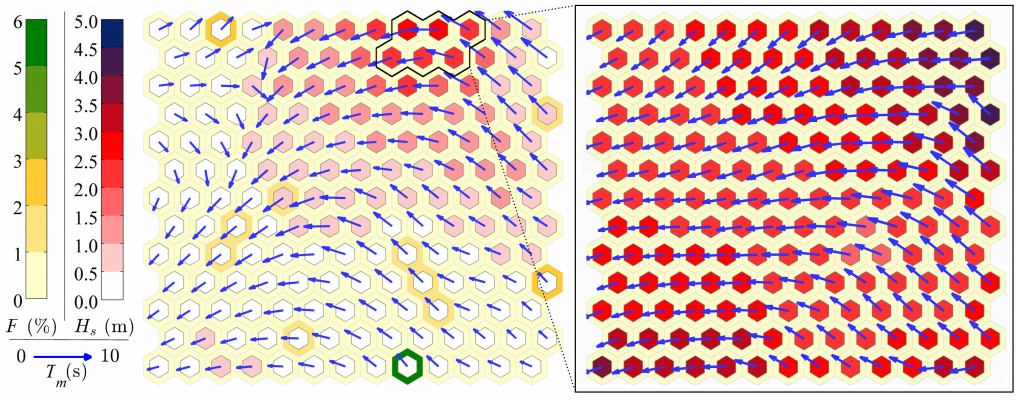

Figure 10. TSOM output map with threshold $H_s^* = 2.12$ m (97^{th} percentile of H_s). H_s: inner hexagons color, T_m: vectors' length, θ_m: vectors' direction, F: outer hexagons color. Wave climate ~~below the threshold~~ after single-step SOM (left panel) and TSOM extreme wave climate (i.e. over the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

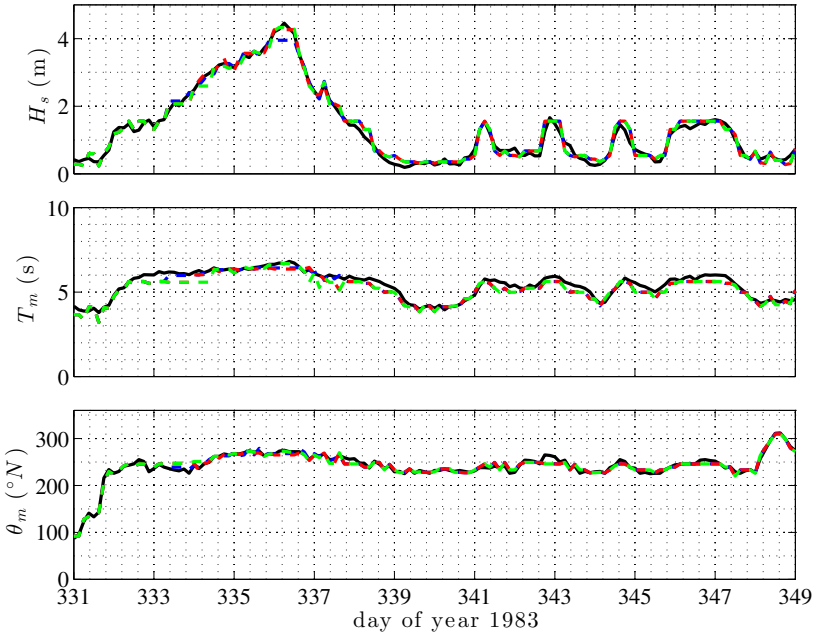

Figure 11. TSOM: comparison between original (black solid lines) and reconstructed time series of H_s (top panel), T_m (central panel) and θ_m (bottom panel), for a chosen sequence of events. Thresholds: 95^{th} ~~percentile~~ (blue dashed line), 97^{th} ~~percentile~~ (red dashed line) and 99^{th} ~~percentile~~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

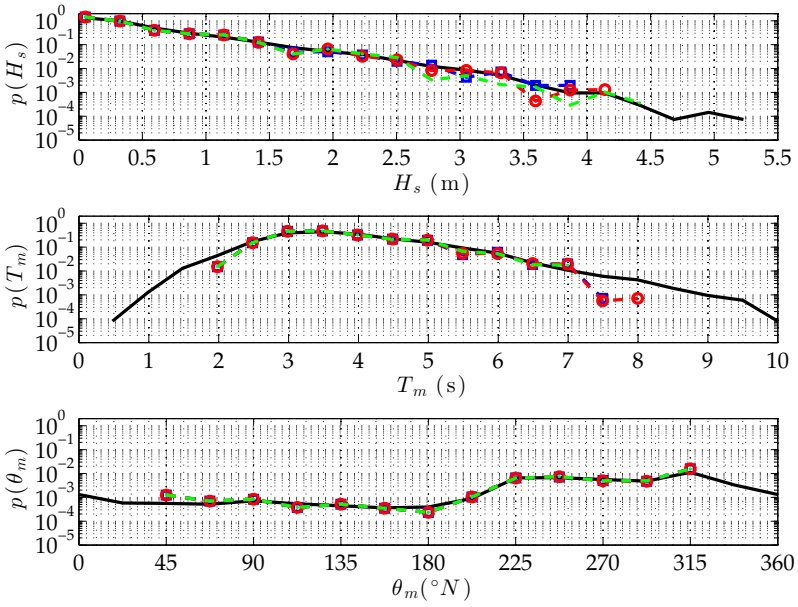

Figure 12. TSOM: comparison of original (black solid line) and resulting pdfs of H_s (top panel), T_m (central panel) and θ_m (bottom panel), for the whole dataset. Thresholds: 95^{th} ~~percentile~~ (blue dashed-squares line), 97^{th} ~~percentile~~ (red dashed-circles line) and 99^{th} ~~percentile~~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

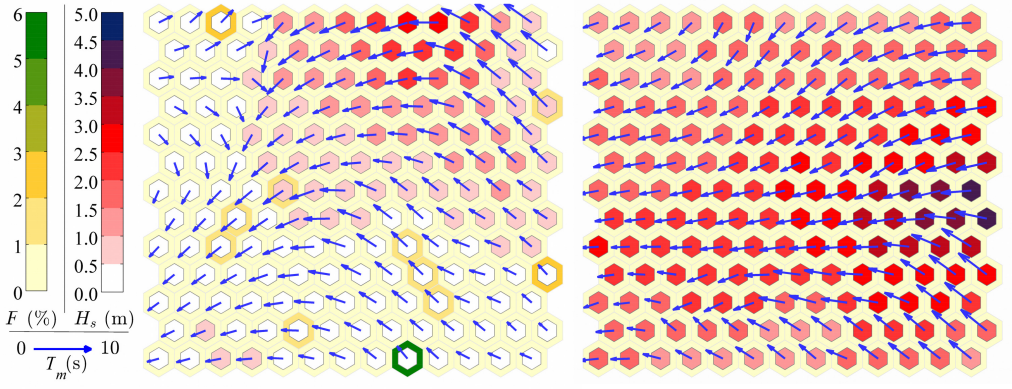

Figure 13. ~~POT-TSOM~~ POT-SOM output map. H_s: inner hexagons color, T_m: vectors' length, θ_m: vectors' direction, F: outer hexagons color. ~~Low/moderate wave~~ Wave climate after single-step SOM (left panel) and stormy wave climate (right panel). For the right panel map, mean quantization error: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

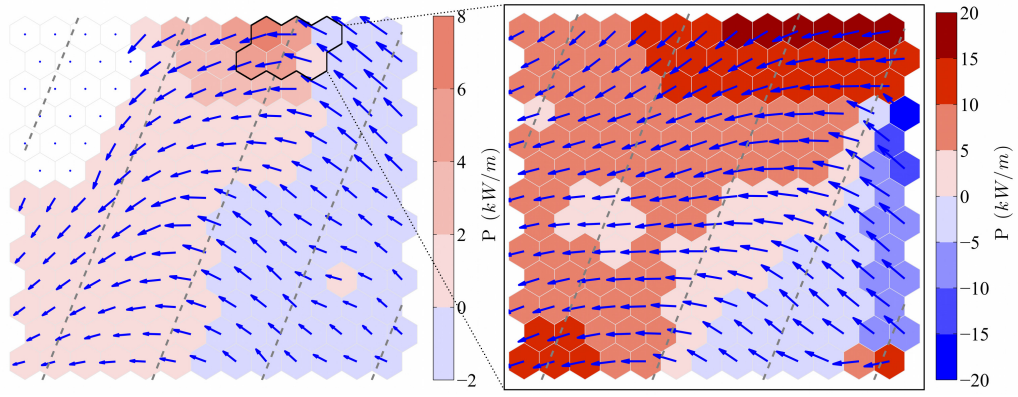

Figure 14. Application of TSOM: assessment of the longshore flux of wave energy P in shallow-water, after single-step SOM (left panel) and resulting from the TSOM extreme wave climate (right panel and cells within black solid line in the left panel). Mean wave directions at *Acqua Alta* tower (blue arrows) indicate contributions of different meteorological conditions: positive mainly due to Bora ($180 < \theta_m \leq 270$°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.