Proceedings of the ASME 2022 41th International Conference on Ocean, Offshore and Arctic Engineering
OMAE 2022
June 5-10, 2022, Hamburg, Germany

OMAE2022-78608

AUTOMATIC CLUSTERING OF METOCEAN CONDITIONS IN THE BRAZILIAN COAST

Felipe M. Moreno
Researcher
Escola Politécnica
University of Sao Paulo
felipe.marino.moreno@usp.br

Eduardo A. Tannuri
Full Professor
Escola Politécnica
University of Sao Paulo

Fabio G. Cozman
Full Professor
Escola Politécnica
University of Sao Paulo

ABSTRACT

This paper introduces a pipeline that assembles a dataset of metocean conditions consisting of wind, wave and surface currents, and then clusters this data to find the characteristic environmental conditions of each region in the Brazilian coast and the associated Exclusive Economic Zone. Clustering uses the Partitioning Around Medoids algorithm with the silhouette coefficient. As examples, we first present an analysis for the whole Exclusive Economic Zone and then a focused analysis around the Santos port in Southeastern Brazil.

INTRODUCTION

The Brazilian coastal waters carry significant international trade and valuable natural resource exploration, corresponding to about 95% of Brazilian oil, 80% of gas and 45% of fishery. Due to its size and importance, that region of the ocean is also called the Blue Amazon. It is thus important to characterize the typical meteorological and oceanic (metocean) conditions of this region, as such a characterization can yield insights and help planning activities in the sea, for instance, the construction of oceanic and coastal structures, vessel operations, fishing, extraction of minerals, ecological and environmental forecasting. Metocean conditions of particular interest are the currents at the sea surface, the gravity waves and winds close to the sea level.

Several previous efforts have applied cluster analysis to characterize metocean parameters, usually to obtain the input for another process (such as fatigue simulations for structures at the sea) or to gain insight into the environmental elements of a region. Most studies in the literature focus on clustering a single environmental phenomenon, such as surface waves [1–3], and ocean currents [4–6], while a few studies cluster more than one environmental phenomena such as currents and waves [7–9] or current, waves and wind [10]. Such existing proposals apply different clustering algorithms: k-means, Ward's algorithm, and self-organizing maps. Results are either subjectively evaluated by an expert, or objectively evaluated through a metric, for instance the quantization error or the decrease in accuracy of a fatigue simulation by using the clustering result as an input instead of the entire dataset. When more than one variable is used, the variables are either normalized or converted to a comparable quantity (e.g. ocean waves can be seen as an oscillatory water current for some applications).

Most applications in the literature only focus on small geographical regions or in a single environmental element. Hence, our contribution here is both to build a large dataset of diverse simultaneous metocean environmental elements encompassing the totality of the Blue Amazon, and to generate descriptions that capture the large-scale characteristic conditions of the entire region.

To achieve those objectives, we merged metocean data from the NOAA's and HYCOM's databases of reanalysis models for the entire Brazilian Coast with respect to a period of 11 years. We characterized the most common environmental conditions on the Brazilian coast in the year of 2009 by the application of Partitioning Around Medoids (PAM, also known as k-medoids). The best number of clusters was obtained by the silhouette method. Results are presented in this paper. In future work, this pipeline will be integrated into the BLue Amazonia Brain project (BLAB) now in progress, so as to provide a basic tool for extracting the most common characteristic conditions of regions in the Blue Amazon [11].

The next section discusses the datasets we have merged. The third section describes the clustering method we have employed, and the fourth section describes the application of the clustering method to the data. The last section presents the conclusion and next steps for this work.

Data and Datasets

The dataset used in this paper was assembled from two distinct datasets available on online repositories of agencies responsible for meteorology at a large scale.

The first dataset is provided by HYCOM [12] and consists of measurements of water elevation, current horizontal velocities, temperatures and salinity for 40 different depth levels and for the ocean floor. That reanalysis data was obtained by assimilating satellite and in situ data in the numerical models of the global ocean flow using the Navy Coupled Ocean Data Assimilation (NCODA) based on multi-variate optimum interpolation. That data encompasses the entire ocean between latitudes 90° North and 80° South in a rectangular grid with a resolution of about 1/12°. The dataset for this model encompasses the period between 1^{st} January 1994 and 31^{st} December 2015. More information about the hydrodynamic model can be found in the work of Chassignet and co-authors [13]. More information about the data assimilation is available in the works of Cummings [14], and Cummings & Smedstad [15]

The second dataset is provided by the National Oceanic and Atmospheric Agency [16] and consists of mean wind speed and mean wind direction at 10m height and significant wave height, mean period and direction. These values were obtained from the Global Multi-Grid Wave Model product that uses the WaveWatch III at its core, where the wave spectral equations are solved by taking into account the interaction between the sea and the winds. The Global Multi-Grid Wave Model is run with wind boundary conditions from the Global Forecasting System (GFS) from NOAA. The data for this model is available between 1^{st} February 2005 and 11^{th} March 2019. This data is available for the entire globe with latitudes between 77.5° South and 77.5° North, with a resolution of 0.5°. More information is available in the WaveWatch III user manual [17].

Both agencies provide a server where a URL query can be sent by the users, and if it is valid, the server returns a netCDF file. The query can define a bounding box for the data, the time spawn for the data, the desired variables to be included, among other parameters. An overview of the entire pipeline, from the data acquisition to the cluster analysis is shown in Figure 1.

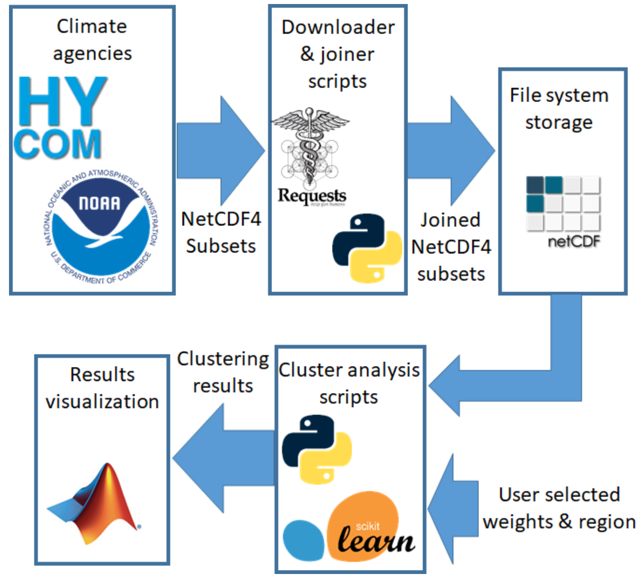


FIGURE 1: Overview of the pipeline used to assemble the dataset and to find the environmental conditions of the regions.

By joining both sources of environmental data we obtained a combined dataset containing values of surface ocean currents velocities, wind velocities and surface waves' significant height, mean period and mean direction encompassing the period between February 2005 and December 2015, having values for ever three hours starting at midnight of each day. As both sources provide data for the same hours of the day (00:00, 03:00, 06:00, ... , 21:00), no interpolation was made with respect to time. For spatial coordinates we used the NOAA's coarser grid (resolution of 0.5° in latitude and longitude) in our dataset. The HYCOM grid coincides with NOAA's grid for integer latitudes and longitudes, so no interpolation was needed; for the half angles between the integers we selected the HYCOM currents in the closest point available. The only data cleansing made was to remove incomplete time-steps (the ones with missing currents), as the original data is provided already clean by the agencies.

This data is available for the entire bounding box shown in Figure 2 with a spatial resolution of 0.5° in latitude and longitude. The dataset's geographic bounding box was selected to fully encompass the Brazilian Exclusive Economic Zone. For the entire period each one of the 2307 geographical points has 31523 complete time-steps, and 490 time-steps with missing currents (that were discarded in the assembled dataset). An overview of the assembled dataset structure is depicted in Table 1.

Clustering the Blue Amazon

Due to the size of the dataset, we must handle, we must use an automatic clustering procedure where parameter fine-tuning

FIGURE 2: Geographic region of interest where ocean data was collected. The bounding box has latitudes between 8°N and 36°S and longitudes between 26°W and 55°W, thus encompasses the Brazilian Exclusive Economic Zone [18]

is reduced to a minimum. The most crucial parameter that we desire to tune automatically is the number of groups in the data, henceforth defined as k, as the number of groups together with its members characterize the region.

Some authors have proposed methods to automatically identify the number of groups in a dataset. Examples are the density-based algorithm OPTICS — Ordering Points to Identify the Clustering Structure [19], and Gaussian Mixture [20]. Another possible approach is to use a single metric, such as the silhouette coefficient or Davies-Bouldin index, to evaluate the best number of clusters in the data. We used the silhouette coefficient as the criterion for selecting the best number of clusters. The silhouette coefficient was selected due to its good performance under different scenarios [21].

The clustering algorithm selected was the Partitioning Around Medoids, also known as k-medoids. This clustering algorithm assumes that each group in the data has a central element called medoid, and medoids are selected so as to minimize the sum of dissimilarities between the elements of the data and their respective medoids [22].

The k-medoids algorithm was selected instead of the k-means due to its ability to cluster data using any arbitrary dissimilarity matrix with guaranteed convergence. This property will allow the direct use of modular distance for circular variables such as incidence angles and time of the year.

TABLE 1: Structure of the dataset for one pair of latitudes/longitudes used for cluster analysis. The directions (denoted by $\nearrow$) follow the oceanic convention: North-East-Down frame, angles are 0^o at North and increase clockwise. Waves and wind directions are the directions from where the wind and waves come from. Wave directions are the peak directions.

Coordinates (Latitude,Longitude) = (24.5S,36.0W)						
Wind		Currents		Waves		
Vel. (m/s)	$\nearrow$	Vel. (m/s)	$\nearrow$	Hs (m)	Tp (s)	$\nearrow$
7.4	218º	0.95	155º	1.05	8.9	130º
6.8	193º	0.80	144º	1.12	8.6	122º
5.3	160º	0.79	162º	1.11	8.2	125º

The wind velocity, current velocity, wave significant height, and wave mean period attributes of the dataset were normalized between 0 and 1 considering the maximum and minimum values in each attribute, enforcing a maximum distance between elements of 1 per attribute. Circular attributes such as angles and time of the year wrap at some value (360°=0° for angles and 24:00 -31st-December = 00:00 -1st-January for time), so they are normalized between 0 and 2 to keep the maximum distance equal to 1, as indicated through Equation (1). By keeping the maximum distance equal to one for all attributes, we guarantee that all attributes have the same priority in the clustering phase. Despite the possibility of correlation between the attributes (eg. wind has influence in waves and currents), we took into account all attributes in the cluster analysis.

With the data ready, the clustering process was run through a k-medoids algorithm available in the sklearn library for Python. The k-medoids allows the use of a dissimilarity matrix, composed by pairwise dissimilarities between elements of the dataset. In this case the dissimilarity between two datapoints o_h and o_j is defined as a weighted taxicab distance (Equation 1), where the distance d_i is calculated differently if the attribute i is linear (wind and current velocity, wave significant height and mean period) or circular (direction, time of the year), according to Equation (2), where o_{hi} is the value of the i-th attribute of point o_h, and o_{ji} is the value of the i-th attribute of point o_j. The weights $w_i \geq 0$ for attributes are user selected, and they depend on the relative importance of each attribute for the analysis being made. Hence we have:

$$dissimilarity(o_h, o_j) = \sum_{i=1}^{8} w_i d_i, \tag{1}$$

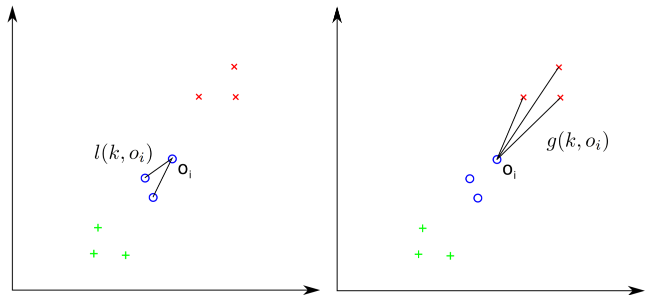


FIGURE 3: Example of the distances between elements of the same group ($l(k,o_i)$) and distances to the closest neighbouring group ($g(k,o_i)$) used to calculate the silhouette coefficient. The closest neighbouring group is the one with the smallest mean distances between its elements and o_i. [23]

where

$$\begin{cases} d_i = |o_{hi} - o_{ji}| \text{ if variable is linear,} \\ d_i = min(|o_{hi} - o_{ji}|, (2 - |o_{hi} - o_{ji}|)) \text{ if variable is circular.} \end{cases} \tag{2}$$

The best number of clusters is obtained from the silhouette coefficient metric implemented in the sklearn library, for a number of clusters ($k \in \{2, \ldots, 10\}$). The silhouette coefficient, shown in Equation (3), is the average ratio between the mean distances of the elements of the same group $l(k,o_i)$ and the elements of the closest group that the elements are not part of $g(k,o_i)$, where o_i is the i-th element of a total of N elements in the data:

$$S(k) = \frac{1}{N} \sum_{i=1}^{N} \frac{g(k,o_i) - l(k,o_i)}{max(g(k,o_i), l(k,o_i))}. \tag{3}$$

Results

In this section we apply our clustering pipeline to different regions of the Blue Amazon.

Largest clusters in the Brazilian coast

The process described was applied individually to geographical regions in a rectangular grid with spatial resolution of 1^o, totaling 845 locations. In order to reduce computational time a time window of only one year was used. The randomly selected year was 2009, so the data period of this analysis starts on 1^{st} January 2009, and goes until 31^{st} December of the same year. For this specific year there are 26 dates with missing currents from a total of 2921 dates; these dates were not used for the cluster analysis. In the present paper, the cluster analysis considered only the direction of the environmental agents, by giving these dimensions unitary weight and the others zero weight.

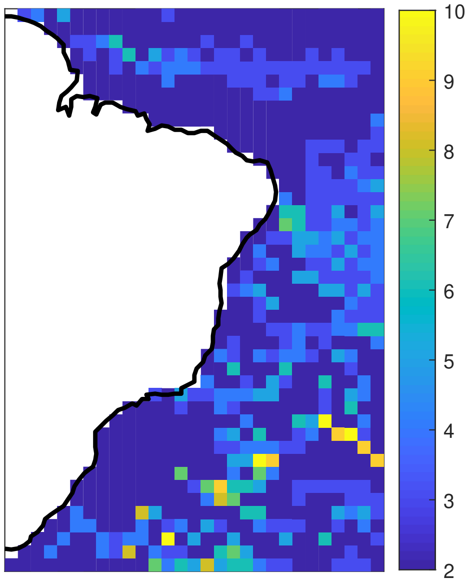
The first analysis was made by plotting the best number of clusters in each region shown in Figure 4a. It is possible to note the predominance of regions with two and three groups and only few geographical points have more than 5 clusters. The silhouette coefficients, provided in Figure 4b, indicate how distinct are the groups found in each region; lower values mean that there is less difference between different clusters. In the figure it is possible to note that regions closer to the coast have slightly higher silhouette coefficients than the open sea, and notably the northeast region has the most well-defined clusters accordingly to the silhouette metric. The increase in the Silhouette coefficient near the shore can be attributed to the fact that current directions tend to become parallel to the coast, assuming two opposite directions along the coast, this mean that the clusters associated with those currents are on average better separated and thus higher Silhouette coefficients.

The seasonality is also plotted as an indicator of how the dominant cluster's environmental conditions are distributed throughout the year. Figure 4c shows a measure of how unequal is the occurrence of the cluster throughout the year; this measure was obtained by Equation (4), by counting the number of occurrences of the cluster in each month, n_i, minus the expected number of occurrences if the cluster occurrence was equally distributed across all months $n/12$, where n is the total number of elements in the cluster. As there are 8 data points for each day, the value was then multiplied by 0.125 in order to obtain this metric in days. In this specific formula the months' duration is considered equal to exactly $1/12$ of the year.

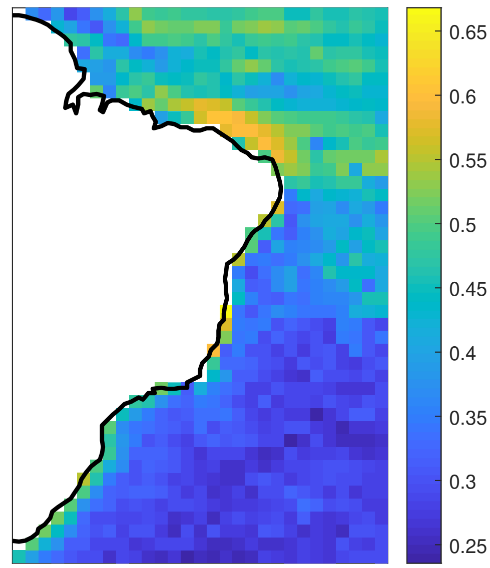
Given this, we have:

$$Sc(lat,lon) = 0.125 \sum_{i=1}^{12} \left| n_i - \frac{n}{12} \right|. \tag{4}$$

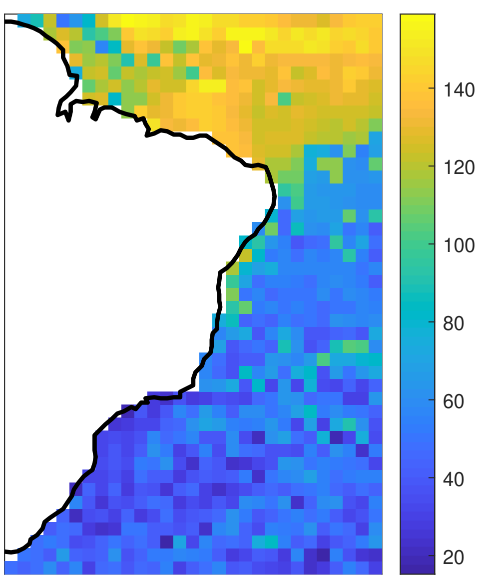
A further analysis was made by plotting the mean direction of each environmental agent in the largest clusters, as well as the mean velocity for wind and currents, mean wave significant height and mean period. These plots are shown in Figures 5a, 5b, 5c, 5d. In the North and Northeast regions, the winds in the main clusters tend to come from the East to the West, in accordance with the Trade Winds. In the South and Southeast coast, winds go parallel to the coast towards the South. An analysis of the currents in the largest clusters shows that surface current velocities at the open sea are significantly lower than closer to the coast, and there is notably a strong stream that goes through the Northern sea. The waves in the main clusters also go towards the coast for the most part, with higher heights in open sea regions, most noticeably in the extreme South of the considered region; also lower wave periods are found closer to the coast. It is also possible to notice in Figure 5a that the wind clusters have a larger variability on the southeast region. This can be attributed to the cold fronts and pressure centers below the tropical zone.



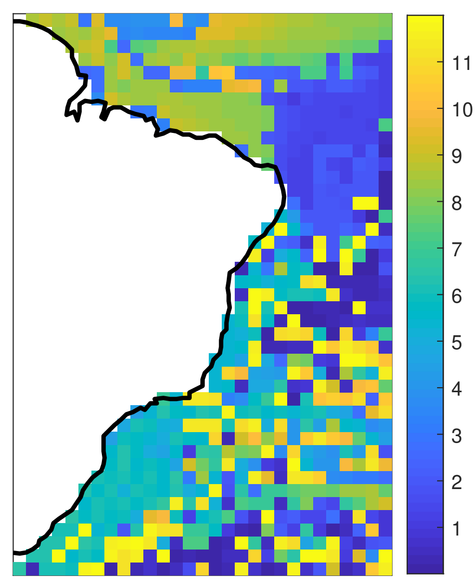
(a) Best number of groups for each geographical region; it is noticeable that most areas closer to the coast have two or three characteristic environmental conditions.



(b) The silhouette coefficient shows low values for most regions, with clear exceptions being the coastline region and the northeast sea.

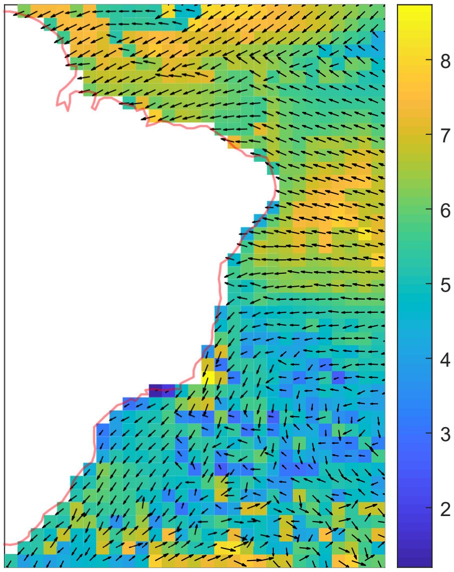


(c) A measurement of the seasonality strength; the scale is obtained by Equation (4), the larger the number, stronger is the seasonality.

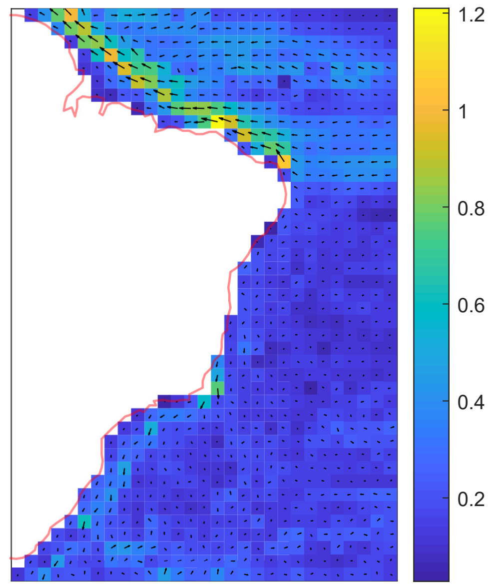


(d) The mean month when the largest cluster occurs; the scale value represents the months, with 0 being 1^{st} January and 12 being 31^{st} December.

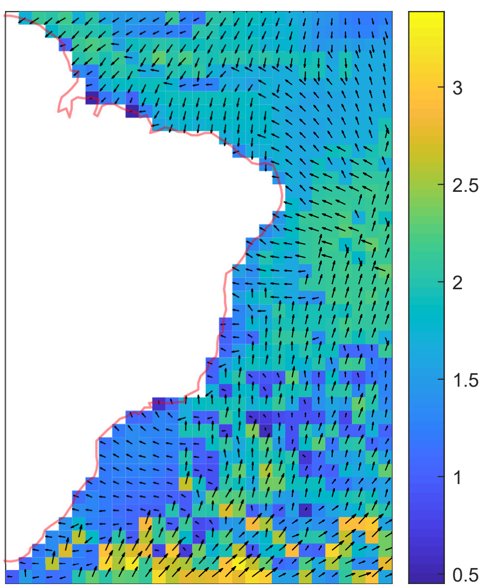
FIGURE 4: Plots of some of the outputs of the cluster analysis.



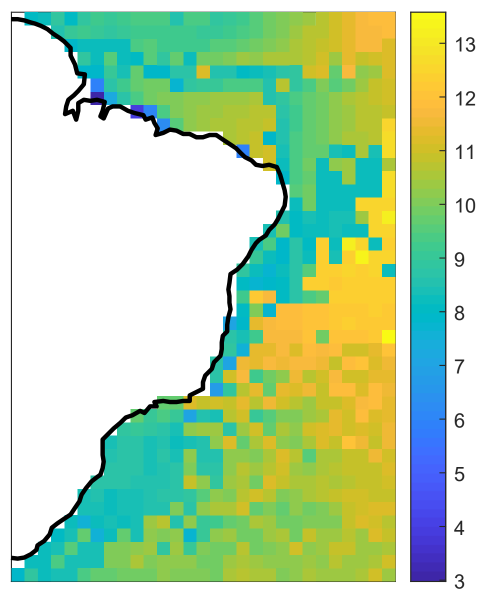
(a) Mean wind directions and mean velocity in the largest clusters, units in the scale are given in ms^{-1}.



(b) Mean current directions and mean velocity in the largest clusters, units in the scale are given in ms^{-1}.



(c) Mean wave significant height and directions in the largest clusters, units in the scale are given in meters.



(d) Mean wave period in the largest cluster in each region, units in the scale are given in seconds.

FIGURE 5: Plots of directions and intensities of the environmental agents in the largest cluster in each region.

FIGURE 6: The geographic point considered in the regional analysis [18]

By analyzing Figure 4c, it is notable that there is a larger seasonality in the Northeast and North seas, while in the other areas the largest cluster of environmental conditions is more equally distributed throughout the year. Figure 4d shows the average month of the cluster occurrences.

Regional Analysis: Santos Port Region

A further example using different types of plots is provided for one specific region of economic importance. The selected region is the coastal region close to the Santos port in São Paulo. The analysis is run for the point 24.5°S 46°W, which is the closest grid point to the port, the selected point is shown in Figure 6. The point is around 66km distant from the shore, and only the directions of environmental elements directions in the year of 2009 are considered for the cluster analysis, as in the previous subsection. Since this analysis focuses only on one geographic point it is feasible to use directional histograms and time series to get a deeper insight into the data patterns.

The cluster analysis took 2 as the best number of groups. The first set of plots we use are directional histograms for each group obtained, to study the distribution of the environmental conditions in each cluster. These plots are shown in Figures 7a and 7b, where the correlation between wind and current directions is apparent. The current goes most of the time parallel to the coast following the direction of the wind. The waves come mainly from the South in both groups, but in the first group one can note a larger percentage of waves coming from the East, that might also be related to the wind.

The cluster's mean conditions and their number of elements are shown in Figures 7c and 7d. The cluster 1 is slightly more common than cluster 2, and the main difference between the clusters is the wind and current directions. The seasonality can be seen in Figures 7e and 7f, where it is noticeable that both clusters occur throughout the entire year, but the cluster 2 occurs more frequently during the fall season. The cluster 2 is normally associated to cold fronts coming from the South, which is more frequent (in the Brazilian Southeast coast) during the fall season.

Conclusion

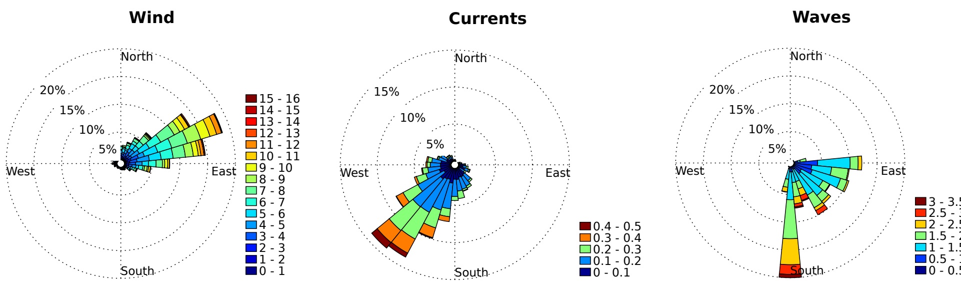
We have developed a pipeline to download and assemble metocean data from climate agencies and to cluster this data in order to characterize the most common environmental conditions in the Brazilian coast. This is an important applied task that is connected with a region of utmost value to Brazil's economy and population.

The k-medoids algorithm was used to cluster the data and the silhouette coefficient was employed as the only metric to find the best number of clusters due to their simplicity and lack of need to fine-tune parameters to cluster large quantities of geographical points. The method was applied to find typical metocean characteristics in a grid in the Blue Amazon, as we have discussed in this paper. An example of a more regional analysis was also presented. It is also noteworthy to consider that the data used here comes from numerical models. While those models are calibrated to represent the real world phenomena, it is hard to assess how well the data represents the ground truth, and real world measurements might be needed for some applications.

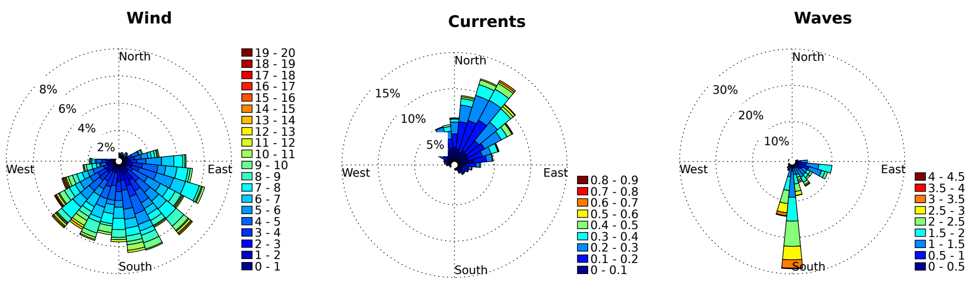
The cluster analysis was separately applied to each geographical point, but is also possible to apply the cluster analysis simultaneously to all the dataset points to obtain clusters that encompass multiple geographical points. In this paper this was not done due to the large volume of data (around 2.2 million datapoints per year for a grid with points separated 1° in latitude and longitude). Depending on the application of interest, it is also possible to use other clustering methods that may yield alternative results, such as hierarchical methods or Gaussian mixture models. And while that methodology's focus was to find the typical metocean conditions of the largest clusters, for some applications the detection of smaller clusters is crucial. In those cases it is possible to expanding the user defined weights to give more weight to the conditions deemed critical to the specific application.

ACKNOWLEDGMENT

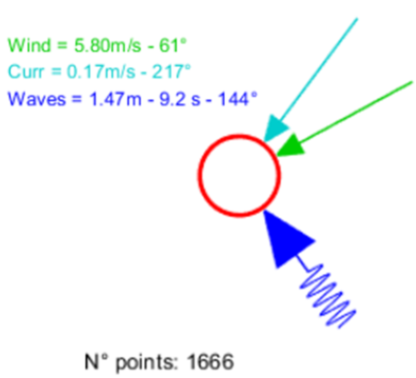
This work was carried out at the Center for Artificial Intelligence (C4AI-USP), with support by the São Paulo Research Foundation (FAPESP grant #2019/07665-4) and by the IBM Corporation. The first author acknowledges FAPESP for research grant 2020/16746-5. The work of Eduardo A. Tannuri was supported by Brazilian National Council for Scientific and Technological Development (CNPq) under Grant 310127/2020-3. The



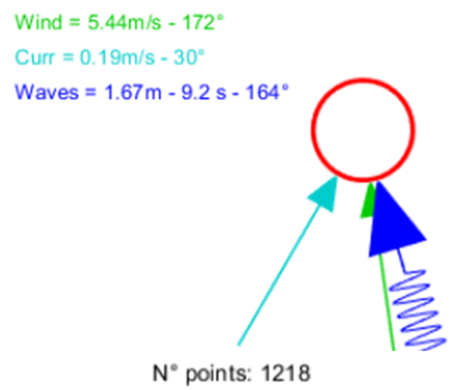
(a) Directional histograms for the first cluster.



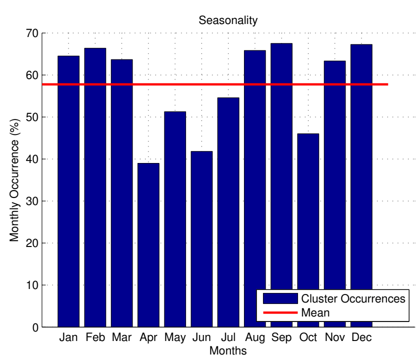
(b) Directional histograms for the second cluster.



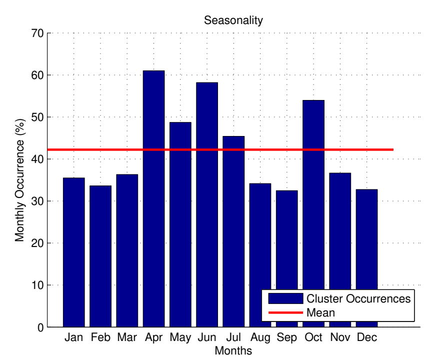
(c) Mean environmental conditions for cluster 1.



(d) Mean environmental conditions for cluster 2.



(e) Monthly occurrences of cluster 1 throughout the year.



(f) Monthly occurrences of cluster 2 throughout the year.

FIGURE 7: Results for the regional analysis, containing directional histograms, mean environmental conditions and the monthly cluster occurrence. Directional histograms were made with the windRose script for MatLab. Wind and waves direction is the one from where these agents come from. Current directions is the one to where the current goes. [24]

work of Fabio G. Cozman was supported in part by CNPq under grant #312180/2018-7.

REFERENCES

[1] Vogel, M., Hanson, J., Fan, S., Forristall, G.Z., Li, Y., Fratantonio, R. and Jonathan, P.., 2016. " Efficient Environmental and Structural Response Analysis by Clustering of Directional Wave Spectra". Vol. Day 3 Wed, May 04, 2016 of *OTC Offshore Technology Conference, Houston, Texas, USA*.

[2] Camus, P., Mendez, F. J., Medina, R. and Cofiño, A. S., 2011. "Analysis of clustering and selection algorithms for the study of multivariate wave climate". *Coastal Engineering, **58***(6), pp. 453–462.

[3] Romeu, M.A.R. and Fontoura, J.A.S. and Melo, E., 2015. "Typical scenarios of wave regimes off rio grande do sul, southern brazil". *Journal of Coastal Research, **31***(1), 1, pp. 61–68.

[4] Prevosto, M., Forristall, G.Z., Jeans, G., Herry, C., Harte, G., Harrington-Missin, L. and Dooley, P., 2012. "Worldwide approximations of current profiles for steel riser design: The wacup project". Vol. 3 - Pipeline and Riser Technology of *International Conference on Offshore Mechanics and Arctic Engineering, Rio de Janeiro, Brazil*, pp. 261–268.

[5] Liu, Y. and Weisberg, R.H., 2005. "Patterns of ocean current variability on the west florida shelf using the self-organizing map". *Journal of Geophysical Research: Oceans, **110***(C6).

[6] Jeans, G., Gibson, R. and Jones, O., 2015. "A new quantitative assessment of current profile clustering methods for riser engineering". Vol. 1 - Offshore Technology; Offshore Geotechnics of *International Conference on Offshore Mechanics and Arctic Engineering, St. John's, Newfoundland, Canada*.

[7] Kanner, S., Aubault, A., Peiffer, A. and Yu, B., 2017. "Efficient Algorithm for Discretization of Metocean Data Into Clusters of Arbitrary Size and Dimension". Vol. 10 - Ocean Renewable Energy of *International Conference on Offshore Mechanics and Arctic Engineering. Trondheim, Norway*.

[8] Kanner, S., Aubault, A., Peiffer, A, and Yu, B., 2018. "Maximum Dissimilarity-Based Algorithm for Discretization of Metocean Data Into Clusters of Arbitrary Size and Dimension". Vol. 10 - Ocean Renewable Energy of *International Conference on Offshore Mechanics and Arctic Engineering. Madrid, Spain*.

[9] Gibson, R. and Christou, M., 2019. "Metocean Criteria for the Fatigue Analysis of Subsea Pipelines". Vol. 7B: Ocean Engineering of *International Conference on Offshore Mechanics and Arctic Engineering. Glasgow, Scotland, UK*.

[10] Moreno, F.M. and Tannuri, E.A., 2021. "Clustering Applied to Large Sets of Environmental Conditions for Selecting Typical Scenarios for Ship Maneuvering Real-Time Simulations". Vol. Volume 6: Ocean Engineering of *International Conference on Offshore Mechanics and Arctic Engineering. Online Conference*.

[11] Center for Artificial Intelligence., 2020. KELM - Knowledge-Enhanced Machine Learning for Reasoning about Ocean Data. https://c4ai.inova.usp.br/keml-en/. Last acessed: 15-Jun-2021.

[12] HYbrid Coordinate Ocean Model., 2019. HYCOM THREDDS Server. https://tds.hycom.org/thredds/catalogs/GLBv0.08/expt_53.X.html. Last acessed: 20-Jun-2021.

[13] Chassignet, E.P. and Hurlburt, H.E. and Smedstad, O.M. and Halliwell, G.R. and Hogan, P.J. and Wallcraft, A.J. and Baraille, R. and Bleck, R., 2007. "The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system". *Journal of Marine Systems, **65***(1-4 SPEC. ISS.), 3, pp. 60–83.

[14] Cummings, J.A., 2005. "Operational multivariate ocean data assimilation. quarterly journal of the royal meteorological society". Vol. 131, pp. 3583–3604.

[15] Cummings, J.A. and Smedstad, O.M., 2013. *Variational Data Assimilation for the Global Ocean*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 303–343.

[16] National Centers for Environmental Information., 2017. NCEI - Coasts & Oceans THREDDS Server. https://www.ncei.noaa.gov/thredds-ocean/catalog/ncep/nww3/catalog.html. Last acessed:20-Jun-2021.

[17] Tolman, H.L., 2009. User manual and system documentation of WAVEWATCH III TM version 3.14 †. Tech. rep.

[18] Racicot, A., 2019. BboxFinder. http://bboxfinder.com/. Last acessed: 18-Jun-2021.

[19] Ankerst, M., Breunig, M.M., Kriegel, H.P. and Sander, J., 1999. "OPTICS: Ordering Points to Identify the Clustering Structure". In Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, SIGMOD '99, Association for Computing Machinery, p. 49–60.

[20] Attias, M., 1999. "A variational bayesian framework for graphical models". In Proceedings of the 12th International Conference on Neural Information Processing Systems. Denver, USA, NIPS'99, MIT Press, p. 209–215.

[21] Vendramin, L. and Campello, R.J.G.B. and Hruschka, E.R., 2010. "Relative clustering validity criteria: A comparative overview". *Statistical Analysis and Data Mining, **3***(4), 8, pp. 209–235.

[22] Kaufman, L. and Rousseeuw, P.J., 1987. "Clustering by means of medoids". In Proc. Stat. Data Anal. Based L1 Norm Conf, pp. 405–416.

[23] Rousseeuw, P.J., 1987. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. journal of

computational and applied mathematics". Vol. 20, pp. 53–65.
[24] MA, M., 2020. Wind rose. https://www.mathworks.com/matlabcentral/fileexchange/17748-windrose. last acessed: 20-jun-2021.