

Coastal regionalization with self-organizing maps-Water quality variables applied to cluster formation

Oliveira, Bruno Meirelles^{1*}; Carneiro, Cleyton de Carvalho²; Harari, Joseph³; Belosevich, Pablo de Sosa⁴

¹Institute of Energy and Environment of the University of São Paulo, Brazil.

²Polytechnic School of the University of São Paulo, Brazil.

³Oceanographic Institute of the University of São Paulo, Brazil.

⁴Institute of Energy and Environment of the University of São Paulo, Brazil.

*Corresponding author

Abstract— Human sewage disposal can interfere with water quality and thus diminish the ecosystem services provision, including Phytoplankton lifespan. Understanding the role played by sewage disposal in water quality can be useful not only for tourism planning but also for characterizing beaches based on water quality using secondary data and avoiding the costs of sampling and monitoring. The objectives of this paper were to understand the water quality behavior at several small bays in a coastal city of Brazil and to test the use of self-organizing maps in forming clusters similar to those derived from geomorphology and to understand how representative these maps were of the water quality of the whole city. According to our results, self-organizing maps showed similar behavior to geomorphological processes, confirm the hypothesis of cluster formation due to quality and also presented a new pattern of data variation related to seasonality that was not noticed before in the sampling.

Keywords— Bathability, self-organizing maps, coastal management, water quality.

I. INTRODUCTION

Marine ecosystem services (ES) such as supplying fisheries, carbon sequestration, food provision, and recreation—all of which make an undeniable contribution to human well-being—are being affected by changes in the climate system (Costanza, 1997; Pauly, 2005; Beaumont et al., 2007; de Groot, 2012). Ocean ES contribute more than 60% of the total economic value of the biosphere (equivalent to almost US\$21 trillion per year [1994 US\$]; Costanza et al., 1997). De Groot (2012) shows an average income from coastal zones of \$2,384.00/ha/year from food provision plus \$256/ha/year from recreation. These values reinforce the importance and irreplaceability of marine ecosystem services, putting their management firmly on the decision-making agenda. However, despite various initiatives in this direction, including the development of an ecosystem approach to fisheries management (Pauly, 2005) and an assessment of the state of health of the global ocean (Halpern et al., 2012), ocean management is still neglected by governments, even at the highest international level.

One of the issues most relevant to ocean ecosystem services is that related to phytoplankton

(photosynthetic microalgae), which are responsible for 50% of global annual marine net primary production (NPP). Phytoplankton, which link the atmospheric and ocean carbon cycles via the biological carbon pump, have crucial importance in trophic chains and ecological balance (Rither, 1969; Field et al., 1998; Falkowski and Oliver, 2007; Falkowski and Raven, 2007; Behrenfeld, 2014). The study of phytoplankton within the marine realm is of vital importance, given the threat of climate change and its knock-on effects on local oceanographic regimes (Armbrecht et al., 2014).

Human sewage disposal can interfere with phytoplankton communities and affect the ecosystem services they provide (KIMOR, 1992), including the recreational use of beaches. Sewage, because of its organic contents, impacts the marine ecosystem when discharged into the ocean, providing high nutrient loads to the coastal zone, especially of nitrogen (N) and phosphorous (P). Furthermore, a high seasonal flow of tourists, together with their related economic attributes, contributes to a significant increase in sewage rates, and this directly interferes with the nutrient rates available for phytoplankton.

This problem is not new, but the perspective of a coastal city losing income due to human sewage in the water is still alive. That is why, since the 1980s, the environmental protection agency of the state of São Paulo in Brazil has had a program dedicated to monitoring seawater quality and to informing the population of the batheability of coastal waters. We used their data from 2004 to 2015.

Nevertheless the amount of data produced by this monitoring program is overwhelming and then the use of some sort of artificial intelligence is necessary. Although sewage discharge is a global problem, our case study focuses on Ubatuba, a small coastal city in southeast São Paulo state, Brazil, with a 200 km long coastline. The city has been designated a priority zone by Brazil's National Council of Tourism, through a Federal Decree: the diversity of its natural resources makes it a place of high ecological importance, with tourism as its main economic activity (IBGE, 2015).

The city has several beaches with low human interference, as well as beaches with a moderate to high human presence which presents some issues regarding scale and representativeness of each of those beaches in the overall picture of the city.

In this paper we discuss the formation of clusters of beaches along the coastline, created by several natural

bays, using water quality data. Then, the individual participation of the clusters in the overall picture of the city is presented and discussed.

Finally the goal of this paper was to discuss the application of self-organizing maps to batheability data to understand variations in coastal attributes. More specific questions relate to: i) the representatives of geomorphology in the overall settings; ii) the possibility of artificial clusters being correlated and representing a coherent group of data; iii) the use of SOMs to create an overall picture of Ubatuba; iv) how individual collaborations fit into the overall picture; and finally v) the emergence of unnoticed patterns in the data. This paper does not represent a novelty in artificial neural network research but may be useful for local management and sustainability.

II. METHODS

Ubatuba case study

The Ubatuba municipality in São Paulo, Brazil, is located on the northern coast of São Paulo state (Figure 1) and has an approximate area of 723,883 square kilometers: 87.04% of the area is covered by native vegetation and 68% lies within a protected area (IBGE, 2015). Ubatuba's economy is seasonal, its predominant development factor being tourism (SMA / CPLEA, 2005).

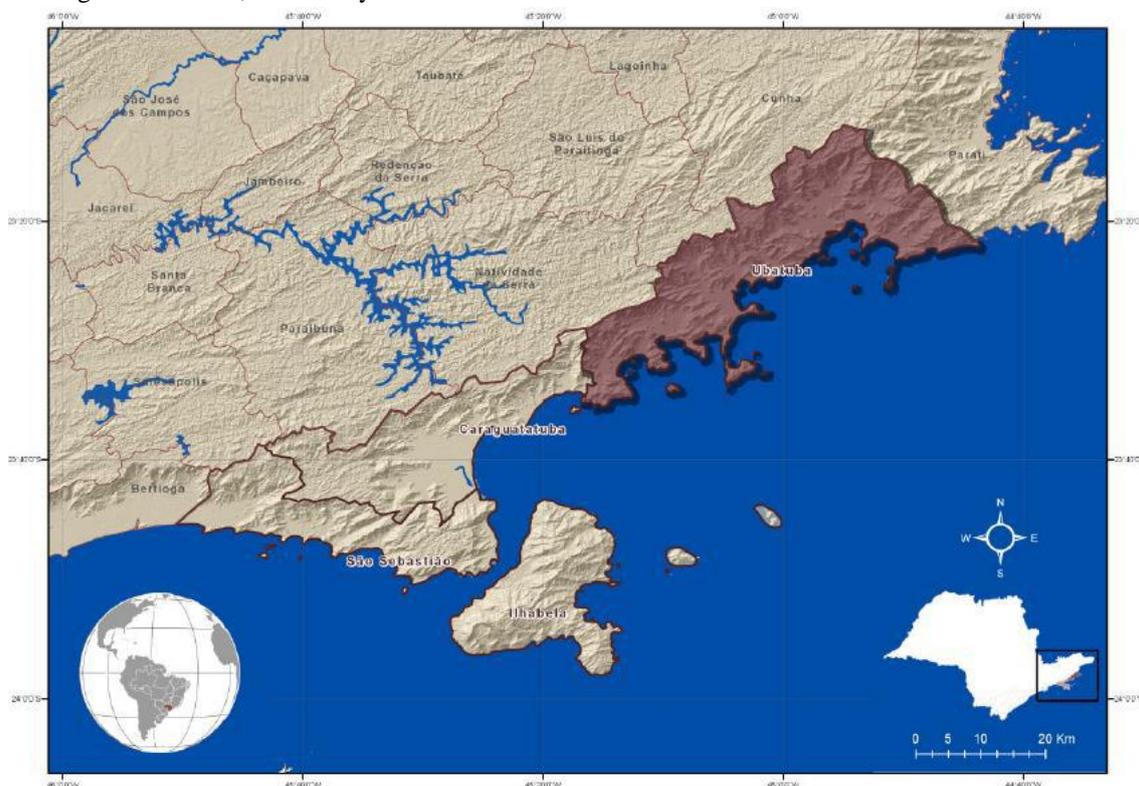


Fig.1: Location map of the north coast of São Paulo. Ubatuba is the dark-shaded area

Source: Karlla Arruda (2017)

It has been estimated that, over the last ten years, the city has welcomed more tourists each year than its actual number of inhabitants (CETESB, 2013; SEADE, 2015). In recent decades, the coastal region of São Paulo has been undergoing significant environmental changes because of intense land use transformation, demographic expansion, and investment inflows into large projects.

Among the major impacts suffered by the region, are tourism and fishing-related impacts on marine ecosystems, as well as impacts caused by high load effluents released to water bodies. Note that the sewage system of the city has only 27.65% coverage (IBGE, 2010) which has increased to 50% currently (CETESB, 2016).

To understand the complex and dynamic behavior of water quality, we directed our focus to sewage disposal as a hypothetically influential factor with respect to the marine ecosystem.

The data set we used to analyze the sewage discharge to the ocean was the annual report on batheability published by the environmental agency of the state of São Paulo. The annual report presents weekly data on the amount of thermo tolerant coliforms¹ collected at 26 sampling points on beaches along the entire city coastline. One was discarded because the sampling point—although very close to the beach—was on a river, which was considered to be a different environment.

The remaining 25 samples showed the presence of coliform concentrations. There were two distinct issues regarding the use of these data in further dynamic analysis. First, if we were to use statistical analysis (average values), all variation would disappear (Figure 8), and the variations are where the batheability problems can best be seen. Second, not all the data can be considered in the same analysis because the quantity of information is colossal (Figure 2).

Figure 2 shows the distribution of batheability data from one sample point. Although the volume of data, just for one point, is huge, no pattern can be perceived. We then converted weekly data into monthly data, transforming 572 samples into a more manageable 143 samples. We obtained the linear tendency, shown as the dotted line in Figure 2.

When analyzing the database for the entire coast, we found an issue related to scale in the sense that we could not use whole city scale. Merging all the data meant

losing peaks of sewage disposal and lack of batheability, making the city seem like an ecological paradise. Moreover, using every monitored beach as an individual study meant losing the overall picture. Thus, to analyze the batheability of the entire coast, we had to cluster sampling points to make the analysis feasible.

Artificial neural networks and simulations

The information revolution during the last decades has altered the traditional water quality management, planning and decision making (Chau, 2006). Same author claims that four types of models have been used to help researches in coastal water quality management: knowledge-based systems (where the decision making can be simulated in an automatic algorithm); Genetic algorithm (simulating natural evolutionary processes and applying them in solving problems); Fuzzy inference systems (when objectives and constrains are vague and the systems are imprecise); and Artificial Neural Networks - ANN (using an information-processing paradigm to simulate relationships that are not fully understood). This paper uses one type of ANN analysis because the objectives are to understand patterns presented in data and not well understood by the researchers and considering it has been used before by other researchers (Maier and Dandy, 1996, Muttill and Chau, 2006; Singh et al., 2009; Najah et al., 2013)

Self-organizing maps

Self-organizing maps (SOMs) are a computer algorithm dedicated to analyzing and interpreting large data sets. The technique is also known as Kohonen maps in honor of the developer of the method.

The main goals of SOMs are to understand and analyze big data and propose results in a “meaningful fashion” (Fraser and Dickson, 2007). Since their discovery, SOMs have been used in finance, industrial control, speech analysis, astronomy, to analyze seismic activity, and in the geochemical and petroleum industry (Fraser and Dickson, 2007). A broad review applied to ecology showed SOMs being used at several hierarchical scales within biology, such as molecules and genes, organisms and ecosystems, and in different ways, ranging from molecular response to poisons to patterning macro invertebrates in coastal ecosystems (Choon, 2011).

Aguilera et al. (2001) also used SOMs to forecast water quality variations due to disposal of human sewage from tourist cities off the Spanish coast. In a broad comparative study using SOMs and other algorithms focusing on ecological data, it was concluded that SOMs are a powerful machine that is perfectly suited to ecological studies, Giraudel and Lek (2001) also recommended SOMs

¹ The thermotolerant coliforms are used as indicators of recent fecal pollution because they present high densities of feces, which are collected by the sewage network (CETESB, 2013). Available on: <<http://www.cetesb.sp.gov.br/agua/praias/25-publicacoes/-relatorios>>

“to be used in an exploratory approach in which unexpected structures might be found.”

One of the main advantages of SOMs is the simultaneous clustering of objects and variables (sampling

locations; Olkowska et al., 2014). The method can also be used to predictor estimate data, pattern recognition, noise reduction, classification, and clustering (Fraser and Dickson, 2007).

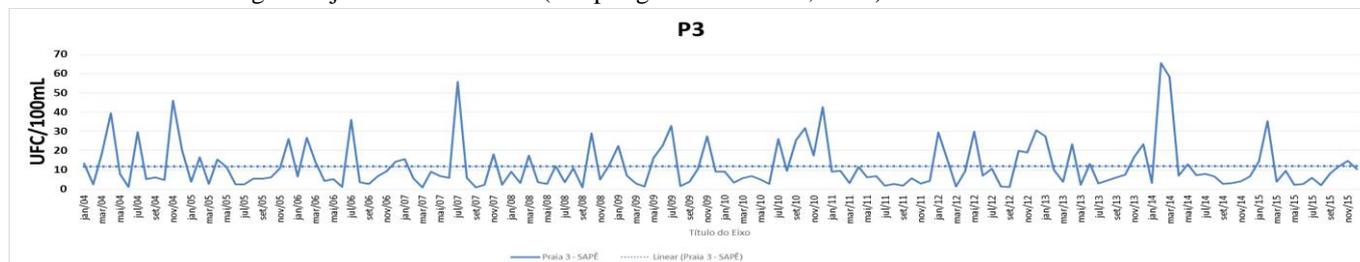


Fig.2: Distribution of sample point P1

Source. The authors

Kohonen maps, unlike normal maps, are formed by a regular grid of (commonly hexagonal) cells. These cells are called neurons and the number of neurons is proportional to the size of the samples ($5\sqrt{\text{number of samples}}$).

Neurons are special cells that represent an amount of data (input vectors or seed vectors) inserted (seeded) into the machine. The algorithm will then classify the data—a process called training. In this phase, all input vectors are translated into neurons. This process occurs in two steps, the first being competitive and the second cooperative.

The algorithm sees all the input vectors (in our case water quality values) displaced as a layer within a two-dimensional form (with this being repeated many times, as the variables demand). The normal hexagonal grid is applied over this distribution in such a way that every input vector underlies one or more hexagons on the neuron layer. The closer a hexagon is to the input vector, the higher the probability of this hexagon becoming the so-called best-matching unit (BMU). This occurs in a competitive way between the hexagons (neurons), meaning that the closer the neuron is from the input vector, the higher its probability of winning the representation of that vector. At the end of this competitive phase, every input vector is replaced by its best-matching neuron.

The cooperative phase moves all the best-matching neurons within a given radius in the direction of the data they represent, inside the data space, changing a small percentage of their attributes so that they better represent the data they are replacing. In other words, the data topology is preserved from the competitive phase, but cooperation means that every neuron will move toward the data it represents, with few changes in its attributes, and this movement will influence all the other neurons to move along a little. The movement in the cooperative phase is performed individually for each neuron, but as each neuron

pushes all its adjacent neurons, the movement subsequently affects all neurons.

The starting point is important for the final result. The final overview will be different for each neuron depending on its program starting point. The topology remains invariable, independent of the stochastic characteristics of the process. After hundreds or thousands of iterations have been run, the final result is a trained (self-organized) map.

This self-organized map is a “2D representation of a complex multi parameter data set” (Fraser and Dickson, 2007), and some visual exploration of the data can be made (U-matrix and component plots). Unified Distance Matrix (U-matrix) indicates how close adjacent nodes are on the map, typically using Euclidean distance. Component plots are another visualization of the neurons where it is possible to see each contribution for a particular variable (beaches in our study) and to display the values using a color-temperature scale so that low values are blue and high values are red.

The errors in the process are measured in two forms, the topographic error (TE) and the quantization error (QE). TE is a measure of the topological preservation errors of input vectors; QE is a measure of the average distance between each input vector and its BMU. Topologies and distances are very important, as they assume that “close placed planes are indication for similar behavior or correlation between respective variables” (Olkowska et al., 2014).

One of the best features of SOMs and the main reason for their use in this type of work is that SOMs are an unsupervised method of cluster formation. This means that there is no need to observe the algorithm working, or to eventually help it with some parameterization and decision (supervision). SOM works alone.

III. RESULTS

Clustering process—Batheability time series.

The bathing data cover the period from 2004 to 2015, using the best available data from the State of São

Paulo environment protection agency (CETESB) - number of colony-forming units (CFU/100 mL) for thermo tolerant coliforms. The distribution of the sampling points can be seen on the map in Figure 3.



Fig.3: Location of sample points

Source: CETESB, 2016



Fig.4: Location of bays

Source: Google maps, modified by the authors

The geomorphological criterion we adopted was based on the hypothesis that the bays and coves in the region tend to have similar characteristics in terms of a lower water circulation rate than the more open regions on the coast.

However, the question arises as to whether this criterion, based on geographic observation and the characteristics of the bays, would be the best one for analyzing all the region’s beaches with respect to the load of pollutants presented by each. To address these issues, we developed two approaches: first, we performed a statistical analysis and second, we compared the results of this with self-organizing maps.

For the statistical analysis, the correlation between the beaches comprising each bay was verified. All the data were tested for their normality with Minitab® statistical software, and their on-parametric distribution was noted. Because of this, the Spearman correlation, which is appropriate for this type of data set, was applied. The level of significance was set at least 5% (p value <0.05), rejecting the hypothesis that there is no statistically significant correlation for cases where the p value is less than 0.05. The results are shown in Table 1, where the present value of all analyses is less than 0.05: this supported the existence of a statistically significant correlation and that the geomorphological criterion adopted made sense from the statistical point of view.

Table 1. Correlation analysis for each cluster

Baía 1												
	Pulso			Maranduba			Sapê			Lagoinha Rua Engenho		
	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?
Maranduba	0,098	0,014	SIM									
Sapê	0,114	0,004	SIM	0,581	0,0000	SIM						
Lagoinha rua engenho	0,085	0,034	SIM	0,489	0,0000	SIM	0,454	0,0000	SIM			
Lagoinha camping	0,105	0,009	SIM	0,409	0,0000	SIM	0,434	0,0000	SIM	0,452	0,0000	SIM
Baía 2												
	Dura			Domingas Dias			Lázaro					
	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?			
Domingas Dias	0,427	0,0000	SIM									
Lázaro	0,543	0,0000	SIM	0,397	0,0000	SIM						
Sununga	0,353	0,0000	SIM	0,418	0,0000	SIM	0,369	0,0000	SIM			
Baía 3												
	Perequê - Mirim			Santa Rita								
	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?						
Santa Rita	0,475	0,0000	SIM									
Enseada	0,41	0,0000	SIM	0,427	0,0000	SIM						
Baía 4												
	Toninhas			Praia Grande			Tenório					
	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?			
Praia Grande	0,471	0,0000	SIM									
Tenório	0,398	0,0000	SIM	0,503	0,0000	SIM						
Praia Vermelha	0,297	0,0000	SIM	0,324	0,0000	SIM	0,27	0,0000	SIM			
Baía 5												
	Itaguá 1			Itaguá 2			Iperoig			Perequê açu		
	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?
Itaguá 2	0,594	0,0000	SIM									
Iperoig	0,436	0,0000	SIM	0,508	0,0000	SIM						
Perequê açu	0,403	0,0000	SIM	0,475	0,0000	SIM	0,48	0,0000	SIM			
Vermelha do Norte	0,338	0,0000	SIM	0,336	0,0000	SIM	0,359	0,0000	SIM	0,412	0,0000	SIM
Baía 6												
	Rio Itamambuca			Itamambuca			Félix					
	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?	Correlação	p Valor	Rejeita H0 ?			
Itamambuca	0,343	0,0000	SIM									
Félix	0,316	0,0000	SIM	0,436	0,0000	SIM						
Prumirim	0,094	0,0180	SIM	0,134	0,0010	SIM	0,169	0,0000	SIM			

Source. The authors

Results of Self-organizing maps

Ubatuba unified matrix presents the distribution of the data after the treatment with SOM algorithm. It is presented in three visual forms (Figure 5): i) node representation; ii) smoothed; and iii) 3D. This U-matrix is

a spatially explicit representation of the neurons trained by the SOM algorithm and ultimately represents the data set inserted into the program. Red cells represent great dissimilarity between data and blue cells represent great similarity.

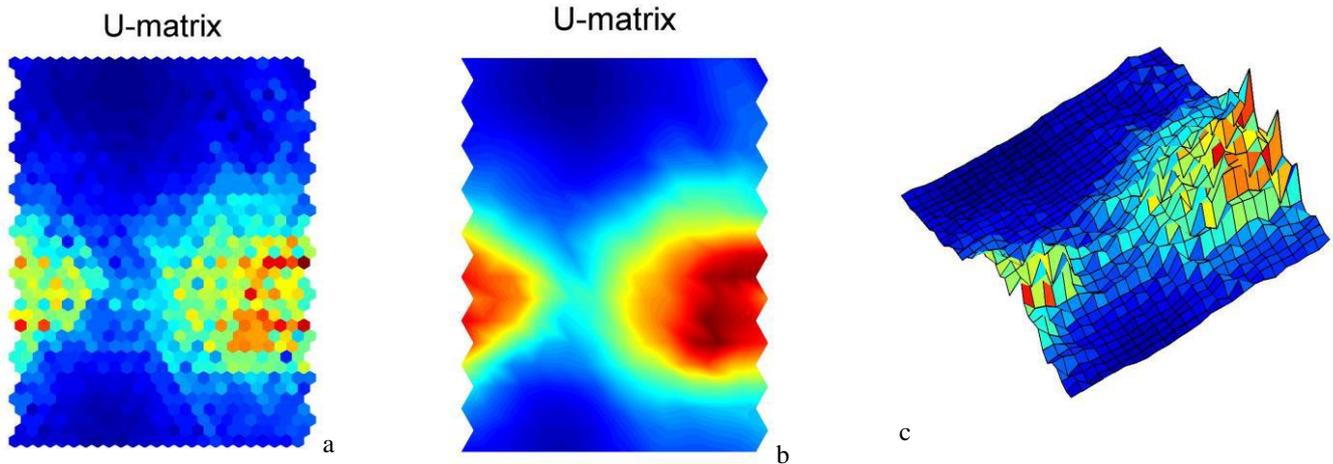


Fig.5: Ubatuba U-matrix for all data sets: i) U-matrix showing neuron patterning; ii) smoothed version; iii) 3D plot of the same pattern

Source. The authors

The U-matrix allows a comparative study of the groups of data (sample points) included in the analysis. The universe of several U-matrices is made up of

component plots which provide visual information on the particular contribution of every sample point of the whole formation of the Ubatuba U-matrix (Figure 6).

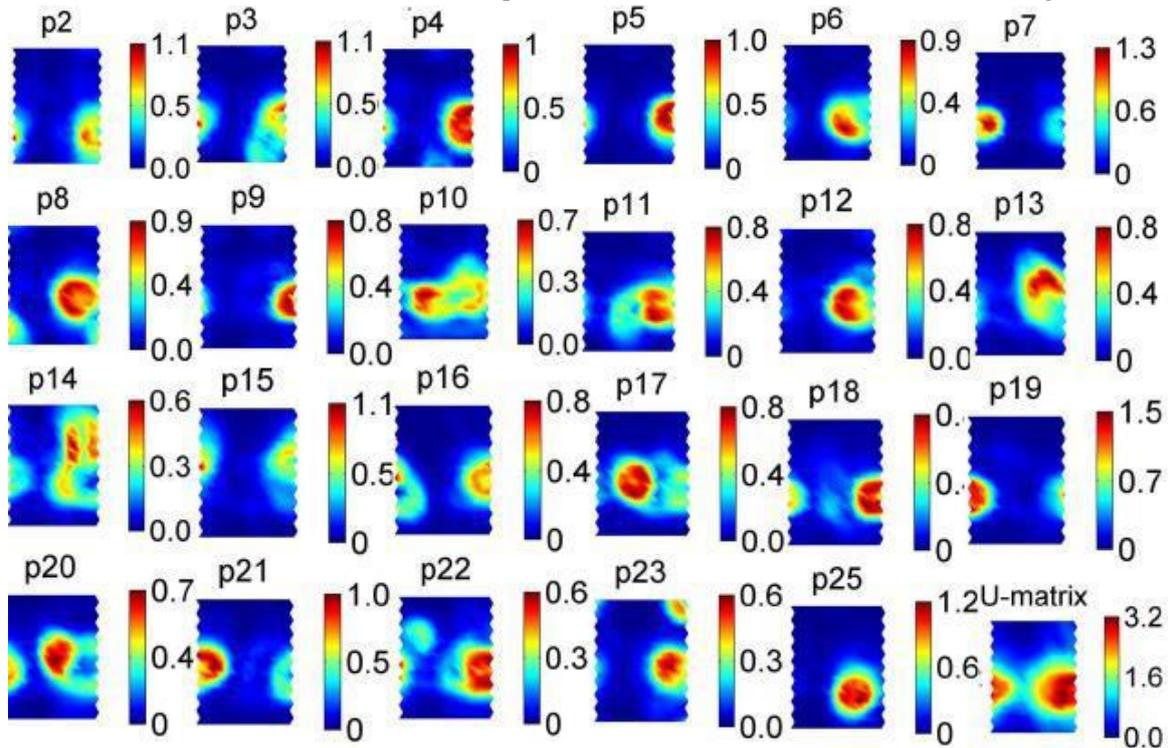


Fig.6: Ubatuba component plots showing contributions of every sample point to the total profile (U-matrix)

Source: The authors

The visual analysis of those beach pro files allowed us to grasp that some of the sample points are highly representative of Ubatuba's general profile, namely, points p2, p4, p5, p6, p7, p8, p9, p12, p15, p16, p18, p19, p22, p23, and p25. These are understood as being the cleanest beaches or even the least frequently polluted.

Other points clearly present other distribution patterns, namely, p3, p10, p11, p13, p14, p17, p20, and p21. These are taken to be the most polluted points or those with a more variable pollutant-dispersion pattern throughout the year.

Table 2. Names and codes for each sample point

Name and reference number of Ubatuba Sample Points									
Name	Nº	Name	Nº	Name	Nº	Name	Nº	Name	Nº
Pulso	1	Dura	6	Santa Rita	11	Praia Vermelha	16	Vermelha do Norte	21
Maranduba	2	Domingas Dias	7	Enseada	12	Itaguá1	17	Itamambuca	22
Sapê	3	Lázaro	8	Toninhas	13	Itaguá2	18	Félix	23
Lagoinha	4	Sununga	9	Praia Grande	14	Iperoig	19	Prumirim	24
Lagoinha (Camping)	5	Perequê Mirim	10	Tenório	15	Perequê-Açu	20	Picinguaba	25

Source. The authors

The map in Figure 5 (a and b) is a 2D representation of a toroid, that is, an nD figure. To visualize this, join the upper border to the lower border to form a horizontal tube. Then link the beginning and the end of the tube to form a never-ending tube or toroid.

The figure thus formed raises the suspicion that is possible to have different clusters of data within the samples. However, the assumption that the U-matrix is produced stochastically cannot be confirmed without a further specific test—the k-means clustering test—an algorithm created for the analysis of clustering processes.

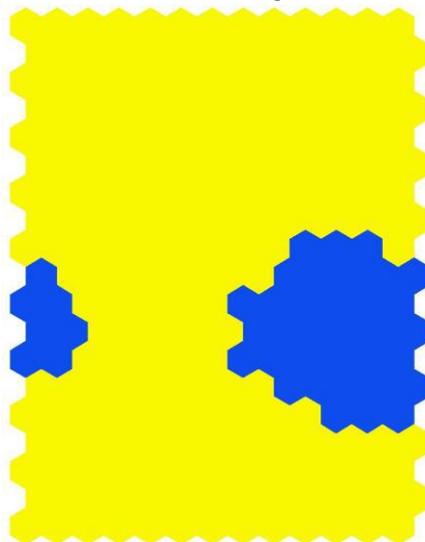


Fig.7: K-means representing clustering of Ubatuba bathability data.

Source: The authors

The k-means is represented in Figure 7. The cluster formation is determined by the David-Boulding Index (DBI), a subroutine on the k-means algorithm. The DBI represents the number of clusters found in the analysis, in this case, 2. Assuming that the DBI is stochastic and that the results depend on which data in the data set the algorithm begins the calculations with, the procedure to obtain the DBI was repeated 70 times and the most frequent number was selected (2).

The maps show an island of dissimilarity within an ocean of similarity. Considering local reality, this can mean two different possibilities: first, that the data vary as a function of geomorphology, meaning that the most populated beaches have a different pattern of sewage disposal compared with the most isolated ones; or, second, that

there is a temporal pattern of waste disposal occurring only within a time interval determined by the data, in other words, there is seasonal variation

IV. DISCUSSION

The results obtained using statistical analysis were clear and corroborate the geomorphological hypothesis of clustering. This result could be useful for grasping the behavioral characteristics of each individual bay and what locally adapted policies need to be developed to enhance water quality and displace sewage pollution.

The SOM clustering does not show whether a bay is polluted or not, as expected. However, results did give us several insights into the dynamics of the complex sewage dispersal system on the coast.

SOMs organized the information for all beaches and showed that there is a strong pattern of sample division into two main realms (Figure 7). At first glance, we could not understand if this was due to seasonality or to the north-south position of the bay. However, when we compared individual collaborations to overall behavior using component plots (Figure 6), the latitudinal variation of samples does not make sense—the two groups formed have interpolated samples, which discards that possibility. The results show a similar group (formed by p2, p4, p5, p6, p7, p8, p9, p12, p15, p16, p18, p19, p22, p23, and p25) and also a dissimilar group (formed by p3, p10, p11, p13, p14, p17, p20, and p21)

Understanding that the two groups are relative to seasonal variations makes much more sense and also allows us to focus on the problem group in order to prevent pollution and expand sewage treatment.

Another positive application was that all the variations in each point were organized into a suitable view that not only allows the overall picture to be understood (Figure 5) but also the particular collaborations involved (Figure 6). To understand the city, it does not make sense to analyze each sample point individually. Analyzing every bay is possible (Table 1), but there are still some variations that can perturb the analysis.

Figure 8 exhibits the k-means and David Boulding index for every cluster formed using statistics. These clusters are tested using SOMs, but they were artificially formed using the geomorphological hypothesis and the statistical analysis presented in Table 1. Nevertheless, they present many more variations internally when compared to the whole Ubatuba scenario obtained in Figures 5 and 7.

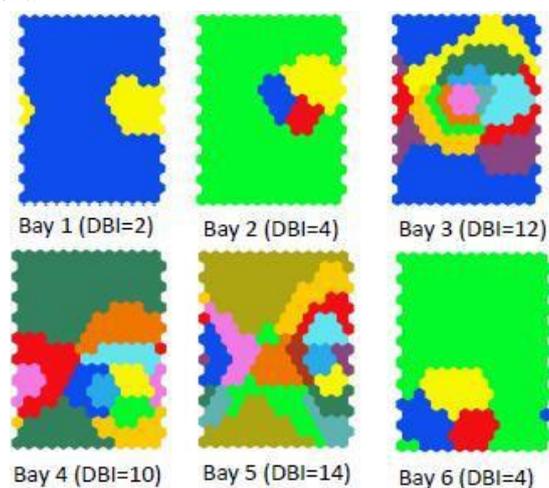


Fig.8: K-means and David-Boulding index for six clusters.

Source. The authors

This outcome represents the possibility of exploring this tool to create representative maps of more regional- and country-scale features, albeit ignoring some local variations. This could help direct policy development.

V. CONCLUSIONS

In this paper we carried out clustering and pattern analysis of a complex dynamic coastal system in Brazil, using batheability time series (2004–2015).

To reach our goals, we used self-organizing maps—a technology deployed to mine big data—to form and analyze clusters by means of a competitive/collaborative algorithm, and to compare outcomes with traditional statistics (Spearman correlation).

The results show that geomorphology can be used as a bias for understanding similarities within batheability data and cluster formation. SOM was shown to be a powerful tool for cluster formation when it was applied to

coastal batheability, and it resulted in unexpected cluster formations. The program was able to separate the whole coast into two groups (pristine and seasonally influenced areas) and was also used to test the remaining variations on that six divisions pattern suggested by geomorphology.

SOMs of individual beaches, visually compared with whole-city data results, showed that one group (more pristine beaches) was more significant in the overall picture. One final conclusion is that SOMs are more than a substitute for statistics; they can be an additional tool for working with coastal data.

ACKNOWLEDGEMENTS

The authors would like to thank the University of São Paulo for the infrastructure available for the modeling process, the São Paulo State Environment Secretary for the continuous process of measuring and communicating the environmental data used in this work.

REFERENCES

- [1] AIDAR E., GAETA S., GIANESELLA-GALVÃO S., KUTNER M., TEIXEIRA C. (1993). Ecossistema costeiro subtropical: nutrientes dissolvidos, fitoplâncton e clorofila-a e suas relações com as condições oceanográficas na região de Ubatuba, SP. Publicação especial Instituto oceanográfico, São Paulo, (10): 9-43[in Portuguese].
- [2] AGUILERA, P. A. et al. (2001). Application of the Kohonen neural network in coastal water management: Methodological development for the assessment and prediction of water quality. *Water Research*, 35(17), 4053-4062.
- [3] ARMBRECHT L. H. ROUGHAN M., R V., SCHAEFFER A., DAVIES P., WAITE A., ARMAND L. (2014). Phytoplankton composition under contrasting oceanographic conditions: Upwelling and down welling (Eastern Australia). *Continental Shelf Research* 75, pp.54–67.
- [4] BEHRENFELD M. (2014). Climate-mediated dance of the plankton. *Nature Climate Change*, 4(10):880-887
- [5] CHAU, Kwok-wing. A review on integration of artificial intelligence into water quality modelling. **Marine pollution bulletin**, v. 52, n. 7, p. 726-733, 2006.
- [6] CETESB (São Paulo) (2013). Relatório de qualidade das praias litorâneas no estado de São Paulo. Available at: <http://www.cetesb.sp.gov.br/agua/praias/25-publicacoes/-relatorios>[in Portuguese].
- [7] CHON, TAE-SOO.(2001). Self-organizing maps applied to ecological sciences. *Ecological Informatics*, 6(1), 50-61.
- [8] FALKOWSKI, P.G. and OLIVER, M.J. (2007). Mix and match: how climate selects phytoplankton. *Nat. Rev. Microbiol* 5(10), 813–819.
- [9] FALKOWSKI, P.G. and RAVEN, J.A. (2007). *Aquatic Photosynthesis*, 2nd edition. Princeton University Press, Princeton.
- [10] FIELD, C.B. BEHRENFELD, M.J. RANDERSON, J.T., and FALKOWSKI, P. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*(281), 237–240.
- [11] FRASER, S. J. DICKSON, B. L. (2007). A new method for data integration and integrated data interpretation: self-organizing maps. In: *Proceedings of Exploration*, pp.907-910.
- [12] GIRAUDEL, J. L. and LEK, S. A. (2001). Comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. *Ecological Modelling*, 146(1) 329-339.
- [13] HALPERN, B.S. (2012). An index to assess the health and benefits of the global ocean, *Nature* 488(7413):615-20.
- [14] KIMOR, B. The impact of eutrophication on phytoplankton composition in coastal marine ecosystems. In: **Marine Coastal Eutrophication**. Elsevier, 1992. p. 871-878.
- [15] MILLS M., RIDAME C., DAVEY M., La ROCHE, and GEIDER R. (2004). Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature*, 429.
- [16] MAIER, Holger R.; DANDY, Graeme C. The use of artificial neural networks for the prediction of water quality parameters. **Water resources research**, v. 32, n. 4, p. 1013-1022, 1996.
- [17] MUTTIL, Nitin; CHAU, Kwok-Wing. Neural network and genetic programming for modelling coastal algal blooms. **International Journal of Environment and Pollution**, 2006
- [18] NAJAH, A. et al. Application of artificial neural networks for water quality prediction. **Neural Computing and Applications**, v. 22, n. 1, p. 187-201, 2013.
- [19] OLKOWSKA, E., et al. (2014) Assessment of the water quality of Kłodnica River catchment using self-organizing maps. *Science of the Total Environment*, 476, 477-484.
- [20] RYTHER J. (1969). Photosynthesis and fish production in the sea: The production of organic matter and its conversion to higher forms of life vary throughout the world ocean. *Science*, 166.
- [21] SINGH, K P., et al. "Artificial neural network modeling of the river water quality—a case study." *Ecological Modelling* 220.6 (2009): 888-895.
- [22] SVERDRUP, H. U. (1955). The place of physical oceanography in oceanographic research. *Journal Marine Research*, 14, 287–294
- [23] SMETACEK V. and NICOL S. (2005). Polar ocean ecosystems in a changing world. *Nature*, 437.