European Water 57: 57-62, 2017.

Cluster analysis and classification of storm events at Rethymno

N. Martzikos[1*], A. Lykou[2], C. Makropoulos[2] and V. Tsoukala[1]
[1] *Laboratory of Harbor Work, National Technical University of Athens, Athens, Greece*
[2] *Laboratory of Hydrology & Water Resources Utilization, National Technical University of Athens, Athens, Greece*
[*] *e-mail: nmartzikos@central.ntua.gr*

Abstract: Coastal urban areas, shorelines, ports and coastal structures are being increasingly threatened by violent hydro - meteorological phenomena such as storms, especially nowadays due to climate change and the climate variability. An extended analysis of storm events in Rethymno city of Crete Island, is attempted, under the frames of PEARL (Preparing for Extreme And Rare events in coastaL regions) an undergoing EU funded project. The aim of this paper is to present an integrated analysis of storm events and their thresholds by enhancing the storm definition for this location and finally to classify them by means of clustering methods. The storm is usually defined as the event exceeding a minimum significant wave height with a certain minimum duration, but the thresholds for the wave height, the duration and the calm period between two successive storms vary depending on the area of interest. Investigating extreme historical storm events and their future projections, the thresholds are defined as 90^{th} percentile of the data set rather to describe the most rare events, checking simultaneously the independence between two successive events. Since the thresholds are selected, a storm's definition is created and the storm events are classified into five classes, depending on their severity. The classification is accomplished with cluster analysis based on the storm energy and the storm period. The different clustering algorithms and methods (hierarchical, k-means, partitioning around medoids and fuzzy clustering) are validated via appropriate indices and the results are presented, checking also their impacts on the coastal area.

Key words: extreme events, storms, classification, clustering evaluation

1. INTRODUCTION

Storm events are one of the most destructive natural hazards which affect low-lying coastal areas and they are responsible, among others, for urban coastal flooding (Costas et al., 2015). The destructiveness of these extreme events depends on their energy and their duration, while due to climate change they are of great importance for many scientists over the past decades. PEARL is an undergoing EU funded project "Preparing for Extreme And Rare events in coastaL regions" which examines the case study of Rethymno city of Crete Island. Extreme storm events at Rethymno cause wave overtopping affecting the stability of breakwaters as well as the safety of human population and the coastal economic activity, as evidenced the last years.

In this study an analysis of storm events and their thresholds is attempted with the main purpose to present an integrated storm classification through clustering methods and their evaluation, extending in this way the previous study of Tsoukala et al. (2016).

2. DATA

The available data in this study were projections of the wave climate to the future which were derived by SWAN wave model (Booij et al. 1999; Ris et al. 1999), a model which was set up and used in the context of the research project: "Estimating the effects of Climate Change on SEA level and WAve climate of the Greek seas, coastal Vulnerability and Safety of coastal and marine structures - CCSEAWAVS" (Prinos 2014) and discussed thoroughly by Athanassoulis et al. (2015) and Tsoukala et al. (2016). The wind data used for the hindcast and climate simulations were proceeded by using a dynamical down-scaling approach of the atmospheric simulations conducted with the regional model RegCNET (Pal et al. 2007), driven by hindcast carbon dioxide emissions during 1960-2000 while integrating AR4-A1B emission scenarios for the period 2000-2100 (Vagenas et al. 2014; Velikou et al. 2014). Those simulations consisted of 3-hourly timeseries of various wave parameters in deep waters, i.e. the significant wave, the spectral peak wave period, the wave direction (Figure 1).

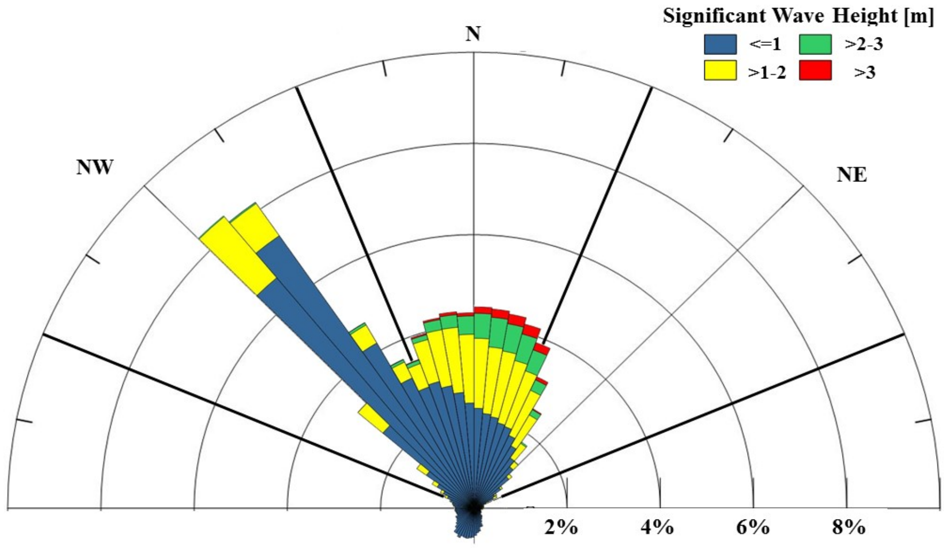

Figure 1. Direction wave occurrence with associated significant wave height at Rethymno, for the period 1960-2000.

3. METHODOLOGY

3.1 Storm thresholds

Lin-Ye et al. (2016) explicitly state that a storm determination should include the storm intensity, the thresholds of H_S, the minimum duration, as well as the calm period between two consecutive storm events. Considering the storm as an extreme event which rarely occurs, the significant wave height's threshold is defined as the 90th percentile of the data set, as it has already been proposed by previous studies (Rangel-Buitrago and Anfuso 2011; Eastoe et al. 2013; Bernardara et al. 2014). Thus, taking into account the wave predictions from hindcast data, it is confirmed that the 10% of the total wave data exceed the threshold of 2 m for the significant wave height (H_S).

The duration of a storm event is defined as the time period in which the H_S remains over the threshold (Boccotti 2000). Contrary to previous studies, where the minimum duration of a storm event is determined without a clarification (see also Table 1.), in this study the minimum duration is estimated to be 9 hours, following again the rule of 90th percentile, having already discarded all the single occurrences with duration of 3 hours.

Regarding the calm phase or the inter-arrival time (De Michele et al. 2007; Corbella and Stretch 2012) between two consecutive storms, it is of great importance to investigate the independence of two events. For this purpose all the events with inter-arrival time not exceeding 48 hours are checked out about their correlation, estimating Kendall's tau and Spearman's rho statistic test. The p-value should be lower than 0.05 in order to reject the null hypothesis of having dependent events. The results show that the correlation has no evident trend concerning the duration between two events, but the first low p-values are indicated for the calm phase of 18 hours. According to this threshold the events which are developed at time interval less than 18 hours, are considered as the same storm event.

Based on the above analysis and for the specific area of Rethymno, a storm event is defined as an extreme wave event which has a significant wave height over the threshold of 2 m, for a minimum duration of 9 hours and a minimum inter-arrival time of 18 hours from the next event. In this way the definition presented by Tsoukala et al. (2016) for the same location is extended.

Next, the data are sorted by wave direction, creating three categories for North, Northeast and Northwest incident waves at the Rethymno coast. As shown in Figure 1, the waves of these directions have the highest frequencies of occurrence and additionally as vertically incident waves they are considered of high-risk for the coastal zone (Allsop et al. 2008). It should be noted that the direction of a storm event is identified with the mean direction of each event, while the standard deviation and the coefficient of variation show that there are no big fluctuations around the mean.

3.2 Storm clustering

Following the definition of the storm energy (Eq. 1) as it was proposed by Dolan and Davis (1992, 1994) and was also used by Mendoza et al. (2011) and Kokkinos et al. (2014) for the Mediterranean Sea, a classification of storm events can be derived with a cluster analysis.

$$E = \int_{t_1}^{t_2} H_s^{\,2} dt \tag{1}$$

In previous studies, hierarchical agglomerative cluster analysis was carried out with Ward's method (Kokkinos et al. 2014) or average linkage method (Dolan and Davis 1992, 1994; Mendoza et al. 2011), using the euclidean distance between two objects and the energy content as a classification variable.

In this study the cluster analysis is based upon the energy and the average peak period, with aim to associate a storm event with the most important variables, i.e. H_S, duration and T_P and finally to classify a total of 563 and 1463 storm events respectively for two periods: 1960 - 2000 (past climate) and 2000 - 2100 (future climate). The classification is accomplished into five classes, trying to have an agreement with previous studies but especially with Dolan-Davis Scale and the Saffir-Simpson Hurricane Wind Scale (Dolan and Davis 1992) which are used extensively by many scientists and organizations (Kelman 2013).

However many studies in Computer Science (de Morsier et al. 2015; Arbelaitz et al. 2013; Jain 2010; Halkidi et al. 2001) have proposed a lot of indices in order to evaluate the clustering results and find the optimum method. An integrated analysis consequently is conducted for the purpose of clarifying the clustering procedure.

First, before trying to find patterns in our data it is necessary to investigate if it is highly clusterable. The Hopkins Statistic value (Banerjee and Dave 2004) indicates, especially if it is lower than 0.05, that there are clusters in our data and so it is worth to proceed. Next for the validation of clustering methods, the connectivity, Dunn and silhouette indices are estimated. The connectivity (Handl et al. 2005) indicates the degree of clusters' association and should be closer to zero. Dunn index (Dunn 1974) is intended to specify the density of clusters and how well-seperated they are; Dunn value should be maximized. The Silhouette Width (Kaufman and Rousseeuw 2005) is the mean of each Silhouette value which estimates the degree of confidence in a specific cluster, when the values are close to one indicate well clustered data without outliers. Correspondingly, many measures evaluate the stability of a clustering results removing one column at each time and comparing the results. Thus the average proportion of non-overlap (APN), the average distance (AD), the average distance between means (ADM), and the figure of merit (FOM) are also estimated as they were proposed and described by Datta and Datta (2003), while small values for all of them are preferred.

All the above estimations are carried out by R (R Core Team 2016) and mainly based on specific packages of Brock et al. 2008, Maechler 2016, Kassambara and Mundt 2016. Here the typical cluster algorithms which are examined are Hierarchical Agglomerative clustering, K-means, Fuzzy and Partitioning around Medoids, for all the clustering methods (i.e. Ward, Single – Complete and Average linkage) and for two types of distance Euclidean and Manhattan. At this stage authors would like to highlight that a detailed description of clustering procedure and their methods is absent in this study, prompting the reader for an extensive study to basic books (i.e. Theodoridis and Koutroumbas 2009) and references therein.

The results show that the optimum clustering algorithm is the Hierarchical Agglomerative, having always the best combination of all examined indices. In addition, the Euclidean distance indicates in any case better results than Manhattan, contrary to the clustering method which differs according to the direction of the storm events.

More specifically, for the period of 1960-2000, the storm events clustered by:

- Average linkage for N direction
- Single linkage for NE direction
- Ward method for NW direction.

And for the period of 2000-2100 the storm events clustered by:

- Single linkage for N direction
- Average linkage for NE direction
- Ward method for NW direction.

4. STORM CLASSIFICATION

In conclusion, the use of foregoing clustering methods contributes to a well-developed storm classification for past and future wave climate and for each direction N, NE, NW. The storm events are classified into five classes depending on their severity, (I - Weak, II - Moderate, III - Significant, IV - Severe, V - Extreme). All the results for the period 1960-2000 are described below in Tables 1-3 in association with the wave height, the wave peak period and the duration for each storm class.

Table 1. Storm events and their characteristics for the period 1960-2000 North wind direction

Storm Class	Significant Wave Height [m]		Peak Period [s]			Energy [m^2h]			Duration [h]	Events
	max	mean	min	max	mean	min	max	mean	mean	
I	3.12	2.18	4.99	9.67	7.26	36.27	135.38	81.16	16.31	183
II	4.09	2.40	5.25	9.16	7.50	137.14	269.91	198.83	32.68	122
III	4.47	2.57	5.71	9.34	7.71	274.26	472.79	345.54	49.25	84
IV	5.26	2.89	4.84	10.36	7.95	490.10	981.40	652.01	72.10	60
V	6.31	3.45	5.28	10.82	8.38	1173.10	1862.63	1517.76	114.86	14

Table 2. Storm events and their characteristics for the period 1960-2000 Northeast wind direction

Storm Class	Significant Wave Height[m]		Peak Period [s]			Energy [m^2h]			Duration [h]	Events
	Max	mean	min	max	mean	min	max	mean	mean	
I	2.82	2.13	5.31	7.93	7.07	37.27	105.83	67.65	14.36	28
II	2.88	2.29	5.90	9.45	7.36	122.66	213.85	167.83	30.94	16
III	3.70	2.49	5.83	8.81	7.45	239.86	310.29	270.45	41.45	11
IV	4.15	2.93	5.85	9.02	7.88	361.34	442.05	411.76	44.25	4
V	4.29	2.99	6.27	9.09	7.92	504.73	698.02	582.86	60.75	4

Table 3. Storm events and their characteristics for the period 1960-2000 Northwest wind direction

Storm Class	Significant Wave Height [m]		Peak Period [s]			Energy [m^2h]			Duration [h]	Events
	Max	mean	min	max	mean	min	max	mean	mean	
I	3.02	2.18	6.56	8.97	7.71	39.02	93.41	61.53	12.16	19
II	3.55	2.39	5.45	9.15	7.59	113.66	154.41	134.19	21.90	10
III	3.80	2.58	4.74	9.48	7.89	173.55	204.20	191.25	26.25	4
IV	4.38	2.88	4.62	10.34	8.35	273.85	287.91	280.88	30.00	2
V	4.24	3.19	6.19	9.64	6.19	404.12	404.12	404.12	36.00	1

5. STORM IMPACTS

Finally in order to connect the storm classes with their impacts (Coco et al. 2014; Del Rio et al. 2012; Almeida et al. 2012) the wave data of each storm event is used by MIKE21 PMS so as to simulate the wave propagation into shallower water regions and estimate the corresponding nearshore wave characteristics, giving also their spatial evolution in the coastal zone of Rethymno. With purpose to study unfavourable sea states, one storm event of class V for each direction is selected for simulation and then the mean wave overtopping is estimated, in one known structure section.

6. CONCLUSION

An integrated analysis of storm events at Rethymno was investigated in this study and a storm classification by means of clustering methods was also examined. The significant wave height and the inter-arrival period thresholds were estimated according to the 90th percentile rule, in order to describe rare and extreme events, such as storms. Consequently a new definition of storm event for the coastal zone of Rethymno was derived. Well-known clustering algorithms and their methods were validated using many indices which are generally used in computer and data science community for the cluster results evaluation. All the above analysis and procedure should be used by anyone who wants to cluster wave data in any specific area. Finally numerical simulations could be used as a tool in order to associate the storm impacts with each class.

ACKNOWLEDGMENTS

This study is part of Mr. Martzikos' PhD thesis which is co-funded through the Action: «Enhancing the Human Research Potential through the Implementation of Doctoral Research» of the Operational Programm «Human Resources Development, Education and Lifelong Learning» of NSRF 2014-2020, co-financed by the European Social Fund and the Greek Government. The research leading to these results has also received partial funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant agreement n° 603663 for the research project PEARL (Preparing for Extreme And Rare events in coastaL regions). The research and its conclusions reflect only the views of the authors and the European Union is not liable for any use that may be made of the information contained herein.

REFERENCES

Allsop, W., T. Bruce, T.A. Pullen, J.W. Van der Meer., 2008. Direct hazards from wave overtopping-the forgotten aspect of coastal flood risk assessment? In: 43rd Defra Flood and Coastal Management Conference, 1-3 July 2008, Manchester University. ID code: 215.

Almeida, L.P., Vousdoukas, M.V., Ferreira, O., Rodrigues, B.A., Matias, A., 2012. Thresholds for storm impacts on an exposed sandy coastal area in southern Portugal, Geomorphology, 143–144: 3–12

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J., Perona, I., 2013. An extensive comparative study of cluster validity indices. Pattern Recognition, 46(1): 243-256

Athanassoulis, G.A., Belibassakis, K.A., Gerostathis, Th.P., Kapelonis, Z.G., 2014. Application of SWAN wave model for climatic simulation of sea condition at coastal areas of the Mediterranean. In: 6th Panhellenic Conf. Coastal Zones Manage. Improvement, 24-27 November 2014, Athens, Greece: 345-364.

Banerjee, A., Dave, R., 2004. Validating clusters using the Hopkins statistic. IEEE International Conference On Fuzzy Systems (IEEE Cat. No.04CH37542), doi: 10.1109/fuzzy.2004.1375706

Bernardara, P., Mazas, F., Kergadallan, X., L., Hamm, 2014. A two-step framework for over-threshold modelling of environmental extremes. Nat. Hazards Earth Syst. Sci. 14: 635-647.

Boccotti, P., 2000. Wave mechanics for ocean engineering. Elsevier Oceanography Series. Elsevier Science, Amsterdam:183-184

Booij, N., Ris, R.C., Holthuijsen, L.H., 1999. A third-generation wave model for coastal regions: 1. Model description and validation. Journal of Geophysical Research, 104(C4): 7649.

Brock, G., Pihur, V., Datta, S., Datta, S., 2008. clValid: An R Package for Cluster Validation. Journal of Statistical Software, 25(4): 1-22

Coco, G., Senechal, N., Rejas, A., Bryan, K., Capo, S., Parisot, J. et al., 2014. Beach response to a sequence of extreme storms. Geomorphology, 204: 493-501

Corbella, S., Stretch, D.D., 2012. Predicting coastal erosion trends using non-stationary statistics and process-based models. Coastal Engineering, 70: 40-49.

Costas, S., Ferreira, O., Martinez, G., 2015. Why do we decide to live with risk at the coast?. Ocean & Coastal Management, 118: 1-11

Datta, S., Datta, S., 2003. Comparisons and validation of statistical clustering techniques for microarray gene expression data. Bioinformatics 19(4): 459-466.

De Michele, C., Salvadori, G., Passoni, G., Vezzoli, R., 2007. A multivariate model of sea storms using copulas. Coastal Engineering 54: 734-751.

De Morsier, F., Tuia, D., Borgeaud, M., Gass, V., Thiran, J., 2015. Cluster validity measure and merging system for hierarchical clustering considering outliers. Pattern Recognition, 48(4): 1478-1489

Del Rio, L., Plomaritis, T.A., Benavente, J., Valladares, M., Ribera, P., 2012. Establishing storm thresholds for the Spanish Gulf of Cádiz coast. Geomorphology, 143–144: 13–23

Dolan, R., Davis, R.E., 1992. An intensity scale for Atlantic coast northeast storms. J. Coast. Res., 8(4): 40-853

Dolan, R., Davis, R.E., 1994. Coastal storm hazards. J. Coast. Res., (SI 12): 103-114

Dunn, J.C., 1974. Well separated clusters and fuzzy partitions. Journal on Cybernetics, 4: 95-104.

Eastoe, E., Koukoulas, S., Jonathan, P., 2013. Statistical measures of extremal dependence illustrated using measured sea surface elevations from a neighbourhood of coastal locations. Ocean Eng., 62: 68-77.

Halkidi, M., Batistakis, Y., Vazirgiannis, M., 2001. On Clustering Validation Techniques. Journal of Intelligent Information Systems 17: 107

Handl, J., Knowles, K., Kell, D., 2005. Computational cluster validation in post-genomic data analysis. Bioinformatics, 21(15): 3201-3212.

Jain, A. K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 8: 651-666

Kassambara, A., Mundt, F., 2016. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.3., https://CRAN.R-project.org/package=factoextra

Kaufman, L., Rousseeuw, P., 2005. Finding Groups in Data: An Introduction to Cluster Analysis (1st ed.). Hoboken: John Wiley & Sons, Inc., Chapter 2: 68.

Kelman, I., 2013. Saffir-Simpson Hurricane Intensity Scale. In: Bobrowsky, P. T., 2013. Encyclopedia of Natural Hazards. Springer Netherlands: 882-883

Kokkinos, D., Prinos, P., Galiatsatou, P., 2014. Assessment of coastal vulnerability for present and future climate conditions in coastal areas of the Aegean Sea. In: Paper Presented at the 11th International Conference on Hydroscience & Engineering: Hydro-Engineering for Environmental Challenges, http://vzb.baw.de/e-medien/iche-2014/PDF/15%20Mini-Symposium%20Impacts%20of%20Climate%20Change/15_04.pdf.

Lin-Ye, J., Garcia-Leon, M., Gracia, V., Sanchez-Arcilla, A., 2017. A multivariate statistical model of extreme events: An application to the Catalan coast. Coastal Engineering, 117: 138-156

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2016. Cluster Analysis Basics and Extensions. R package version 2.0.4.

Mendoza, E.T., Jiménez, J.A., Mateo, J., 2011. A coastal storms intensity scale for the Catalan sea (NW Mediterranean). Nat. Hazards Earth Syst. Sci., 11(9): 2453-2462

Pal, J., Giorgi, F., Bi, X., Elguindi, N., Solmon, F., Rauscher, S. et al., 2007. Regional Climate Modeling for the Developing World: The ICTP RegCM3 and RegCNET. Bulletin of the American Meteorological Society, 88(9): 1395-1409

Prinos, P., 2014. Climate change effects on the Greek seas and coastal areas - the research project THALIS-CCSEAWAVS. In: 6th Panhellenic Conference on Coastal Zones Management and Improvement, 24-27 November 2014, Athens: 315-324.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

Rangel-Buitrago, N., Anfuso, G., 2011. An application of Dolan and Davis (1992) classification to coastal storms in SW Spanish littoral. J. Coast. Res., SI 64: 1891-1895.

Ris, R.C., Holthuijsen, L.H., Booij, N., 1999. A third-generation wave model for coastal regions: 2. Verification. J. Geophys. Res., 104 (C4): 7667-7681

Theodoridis, S., Koutroumbas, K., 2009. Pattern recognition (4th ed.). London: Academic Press.

Tsoukala, V., Chondros, M., Kapelonis, Z.G., Martzikos, N., Lykou, A., Belibassakis, K., Makropoulos, Ch., 2016. An integrated wave modelling framework for extreme and rare events for climate change in coastal areas - the case of Rethymno-Crete, Oceanologia, 58 (2): 71-89

Vagenas, C., Anagnostopoulou, C., Tolika, K., 2014. Climatic study of the surface wind field and extreme winds over the Greek seas. In COMECAP 2014. Abstracts and e-contributions. 12th International Conference on Meteorology, Climatology and Physics of the Atmosphere, Heraklion 28–31 May 2014. Hereklion, Crete: Crete University Press: 283–288.

Velikou, K., Tolika, K., Anagnostopoulou, C., Tegoulias, I., Vagenas, C., 2014. High resolution climate over Greece: assessment and future projections. In: 12th International Conference on Meteorology, Climatology and Atmospheric Physics (COMECAP 2014), Heraklion.