**PAPER • OPEN ACCESS**

# Spatio-Temporal Wave Pattern using Multi-dimensional Clustering Method for Exploring Ocean Energy Potential

To cite this article: N A Rohana and N Yusof 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1051** 012013

IOP Conf. Series: Earth and Environmental Science 1051 (2022) 012013 doi:10.1088/1755-1315/1051/1/012013

# Spatio-Temporal Wave Pattern using Multi-dimensional Clustering Method for Exploring Ocean Energy Potential

**N A Rohana[1], N Yusof[1]**

[1] Universiti Teknologi Malaysia, Skudai, Johor, Malaysia.

tikahrohana88@gmail.com , norhakim@utm.my

**Abstract**. Wave is formed from the movement of air caused by pressure variations that make airflow move from high pressure toward places of low pressure**.** Understanding the wave patterns is challenging since it is highly changeable in space as they travel in variety of directions and heights. Wave are also changing over time especially during the monsoon seasons. Hence, to extract significant information from this highly changeable behaviour of wave this study has utilized a multi-dimensional clustering technique called co-clustering. This technique is able to cluster spatio-temporal data with similar behavior into spatial and temporal components simultaneously. To reveal the spatial and temporal patterns, an algorithm called Bregman Block Average co-clustering with I-divergence (BBAC_I) has been implemented for extracting wave patterns. In order to discover the wave behaviour, the extracted wave patterns were visualized in the form of heatmap that contain information of co-clusters; spatial clusters and temporal clusters dimensions. Then, both spatial and temporal clusters from the heatmap were transformed into geographical maps to depict the variation of wave patterns based on their individual dimension. From these maps, we could observe the distribution of 8 different group of clusters that representing the spatial wave patterns. Furthermore, 5 individual maps have been produced to depict the temporal wave patterns across the study area. Finally, the obtained maps were interpreted in the form of wave height which were found to be within 0.4 to 1.4 m heights. The wave height information can be used for identifying their potential for ocean energy harvesting along the coastal area. In generally, the generated spatio-temporal wave patterns from this study could aid Malaysian marine agencies to provide strategic planning for proposing future ocean energy in Malaysian coastal area.

## 1. Introduction

Wave is a geo-referenced time series dataset that is highly changeable across space and time. Due to this dynamic nature of wave, it is become very challenging to identify significant behaviour for exploring the possibility to harvesting energy. In fact, Malaysia has the 29th longest coastline in the world, with a total length of 4675 kilometres (kilometres) and making it a great ocean energy resource [1]. Malaysian seas are characterised by low significant wave height from (0.5–1 m) and long wave periods from (4–5.5 s). But, in monsoon periods, Malaysian wave can ranging from 8 to 20 kW/m [2]. There are various types wave patterns guided from the World Meteorological Organization sea state. A better understanding of waves movement and their potential for power harvesting are necessary due to the various characteristics of waves that are vary over time and space. This dynamic behaviour can lead to an exploration of wave patterns. Thus, an appropriate data mining approach is required to allow data exploration in both space and time components simultaneously.

Spatial-temporal data exploration are becoming increasingly important in a wide variety of applications in providing better understanding about the nature of the data [3]. Data mining is advantageous for spatio-temporal data exploration because it discovers hidden patterns in huge datasets. Clustering is a key element of spatio-temporal data mining. It organizes relevant data components together and also provides a higher-level perspective of the data [4]. Previously, most clustering studies have concentrated only on a single dimension of geographic information for data exploration that are either spatial or temporal dimension [5],[6]. This clustering technique is performed over a single dimension for grouping things based on their spatial and temporal characteristics, respectively. For instance, spatial clustering is used to identify clusters in attribute values across the spatial component. The resulting clusters are groups of locations that exhibit similar behavior. As a result, the temporal element was absent within these patterns. Similarly, to temporal clustering, the data is grouped according to similar characteristics that occurred over time without taking into account spatial elements within the data. Hence, clustering from both spatial and temporal dimensions could provide a new insight in performing clustering over geospatial dataset. This type of clustering method is called co-clustering. Hagenauer and Helbich [4] defined the co-clustering as the simultaneous grouping of columns and rows in a matrix dataset. In general, the co-clustering performs significantly better than one-dimensional clustering [4],[5]. Spatio-temporal clustering is important since it enables the extraction of a series of groups based on both spatial and temporal data simultaneously [7]. The idea of co-clustering has grown in relevance since it was first proposed by Hartigan [8], and now it has extensively been applied in many applications include bioinformatics, crime behaviour prediction, traffic dynamic analysis, climate and metrological data [9]. For instance, Kluger [6] has studied using spectral bi-clustering technique that simultaneously groups genes into common expression patterns across different type of genes and gene conditions.

For detail wave data exploration, the co-clustering method can suitably applies for extracting spatio-temporal patterns that will exhibit the significant behaviour of wave and their potential for converting into energy [10],[11]. Furthermore, an appropriate geo-visualization approaches are required to thoroughly explore the extracted spatio-temporal patterns to gain better understanding of wave behaviour.

Therefore, this study has introduced the implementation of co-clustering algorithm for extracting spatio-temporal wave patterns across Malaysian seas. Specifically, the Bregman block average co-clustering algorithm was utilized in this study. Even though, Malaysian seas are highly influenced by the monsoon seasons (southwest monsoon and northeast monsoon) [12], the obtained co-clustering results have detected the invariably changes in wave dataset. These finding were interpreted into ocean energy potential. Hopefully, these spatio-temporal wave patterns discovery could aid Malaysian marine agencies in their strategic planning for proposing future ocean energy.

## 2. Concept of Co-Clustering

The concept of co-clustering is illustrated in Figure 1, where partitions in rows (R) $I_1$,..., $I_n$ and columns (C) $J_1$,..., $J_n$ are clustered separately. Crossing the row and column partitions results in co-clustering clusters. Co-clustering clusters, also called biclusters, will extract a set of submatrices from the original data matrix. When rows and columns are partitioned into partitions, there are two types of co-cluster structures: diagonal block and checkerboard structures. The diagonal block is defined as a bicluster that is exclusive of rows and columns (rectangular diagonal blocks after row and column reordering) [13]. These blocks are subsets of rows and columns whose expression values are similar. A perfect matrix rearrangement would produce an image with K diagonal rectangular blocks (Figure 2). The checkerboard structure (Figure 3) identifies each row using all column clusters and each column using all row clusters. The rows and columns attributes contain information about the membership of the row and column clusters. Each of these checkerboard partitions has its own cluster, and the distinction between them is based on their similarity within the groups.

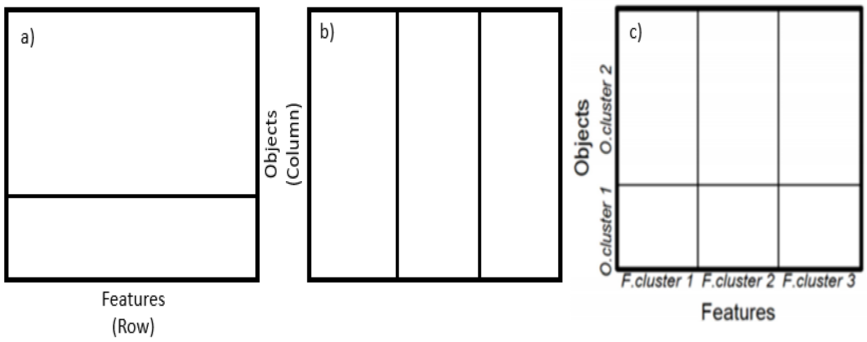


**Figure 1.** Concept co-clustering. a) Features for row input; b) objects for columns and c)the co-clustering when combine with row and columns [14].

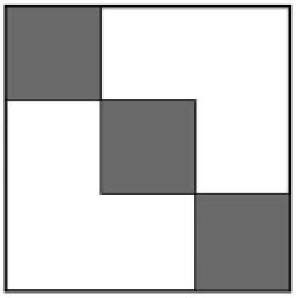

**Figure 2.** Diagonal Biclusters in rows and columns [13].

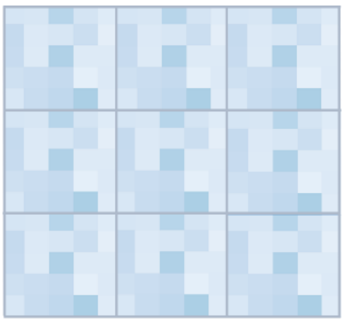

**Figure 3.** Checkerboard Structure for Co-clustering [6].

The Bregman block average co-clustering algorithm with I-divergence (BBAC_I) were applied in a study [10] to implement the simultaneous analysis of the spatial and temporal configuration in geo-referenced time series. The I-divergences were chosen cause from the study [15] it analytically proved its superiority. The variations among the wave values within each co-cluster along with both spatial (location) and temporal (years) dimensions. This co-clustering applies both equally to the objects and attributes [16]. For this study, the co-clustering techniques reduce distortion by assigning each spatial and years to the nearest spatial- and years-clusters. This process ends when convergence occurs, resulting in optimal co-clustering. The co-clustering results then generate a re-ordered data matrix. That is, all spatial or years in the same co-cluster are combined by exchanging rows and columns. A co-cluster is the intersection of two locations/years clusters. In this way, co-clustering allows for the analysis of spatial patterns with the time-varying of years (temporal patterns).

## 3. Methodology

### *3.1. Data and Study Area*

This study used time series wave data starting from 2011 until 2021(10 years) and these data were obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF). The downloaded data was in the format Network Common Data Form (NetCDF). The ECMWF consist of multidimensional variables such as wind data, wave data, temperature, and others. The study area for this analysis was covered within the South China Sea (Figure 4), where this ocean area is frequently battered by strong winds from both the Indian Ocean and the South China Sea. In addition, the South China Sea's ocean energy resources have great potential for exploitation [17].


**Figure 4.** South China Sea (Study Area).

In the context of co-clustering using spatio-temporal data, the majority of research has used the same algorithm which is Bregman Block Average Co-Clustering with I-divergence (BBAC_I). This BBAC_I was used to capture the interactions of wave patterns between the location in the South China Sea and the years. In the beginning, the wave data were read using Python in NetCDF format which has three dimensions (time, latitude, longitude). The variable in wave data was extracted which is significant wave height using Python. The dimensions for variable wave are (132,41,65) and the latitude and the longitude in this analysis is assigned as space(location). It will read as a 2D which is space and the time is assigned as time.

### 3.2 *The Bregman Block Average Co-Clustering with I-Divergence (BBAC_I).*

In this section, the co-clustering analysis was run using a BBAC_I algorithm. The BBAC_I algorithm begins by choosing a random co-clustering scheme for row clustering (p) and column clustering (c) (y). For this analysis, the number of column and the number of rows were set to 8 and 5. Each iteration of the algorithm updates one of the parameters, p or y, in order to minimize the loss function. The loss function is one of the BBAC_I parameters that must be set during the clustering process [18]. Additionally, the loss function is referred to as the cost function or error function. Second step, the loss information between the original and co-clustered matrix is calculated by the BBAC_I (I-divergence). In the iteration step, the algorithm will update a space and time into the space-clusters and time-clusters. The iteration for this analysis is set as 10 where it will update the space and time 10 times. The last part is to recalculate the co-clustering loss of information. The result is in the checkerboard where the spatial-temporal of a matrix was grouped into the same space cluster and the time cluster.

The convergence was set to the threshold of the changes in the loss of information which is 0.1.The maximum number of iterations was set to 10 times to guarantee the convergence stability for the co-clustering results. Table 1 shows the four steps BBAC_I algorithm for an extraction wave pattern.

**Table 1.** Steps in Co-Clustering Method.

| |
|---|
| *Step 1: Choosing a random co-clustering scheme for row clustering (p) and column clustering (c) (y) where row as space and column as a time. It will randomly state the location of the space clusters and time to a time-clusters.* |
| *Step 2: From that, it will calculate the loss in mutual information between the original matrix and co-clustered matrix which is calculated by the I-divergence.* |
| *Step 3: The iteration will update a space and time into the space-clusters and time-clusters.* |
| *Step 4: The generated loss of mutual information will recalculate before and after the co-clustering.* |
| *If the change of the value loss in mutual information is smaller than the threshold, need the co-clustering was obtained. However, if not, need to start a new loop and repeat Step 2.* |

As a result, the implemented of BBAC_I allow for the identification of significant patterns of wave that go through radical changes in composition over time and space as they evolve simultaneously. Furthermore, the patterns that have been extracted were further analyzed in the form of wave heights in order to identify their suitability for energy harvesting with regard to their locations along the coastal area.

## 4. Result and Discussion

The extracted spatio-temporal wave patterns using the BBAC_I algorithm were converted into geographical maps in order to represent the patterns in both space and the year's cluster. The main co-clusters output is presented in the checkerboard matrix with the size of 8 columns x 5 rows. Figure 5 shows the checkerboard for space and years co-clusters. The blue indicates the value of the wave is 0.2m. the red color shows the highest value which is 1.4m. the cluster is divided into 5 colors which are red, peach, light grey, light blue, and blue. Each of the colors represents a different cluster with different value of wave height.

Figure 6 shows the space cluster in South China Sea area. In addition, the total number cluster is 8 and each the clusters represent a different color such as red, orange, bright orange, peach, light grey, light blue, cornflower-blue, royal blue, and blue. All these colors are from matplotlib in Python. The West Malaysia, Sabah, and Sarawak were located from 2.5 – 7.5 (latitude) and from 109 – 119 (longitude). This result was generated from the BBAC_I algorithm which indicates the space with a different cluster. Moreover, the result shows the location with different clusters and colors that are assigned in each location. For example, cluster 1 with blue color is assigned from latitude -5.0 to -7.5 and from longitude 95 to 113, at 7.5 to 10 latitude and ranging from 90 to 97 longitude. The red color from cluster 8 was located from -2.5 to -5.0 latitude and for longitude is ranging from 107 to 125 longitude. Next, the South China Sea area has a variety of color that indicates cluster 3,4,5, and 6. The white color represents West Malaysia, Sabah, Sarawak, and West Sumatera which do not have any cluster because there is no significant wave height value (null) at the mainland.

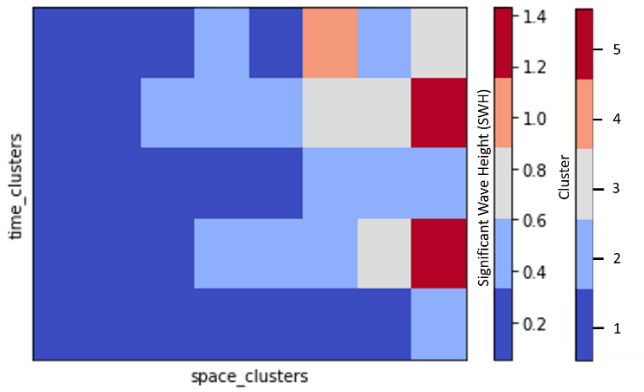


**Figure 5.** Checkerboard co-cluster for space and time of wave pattern.

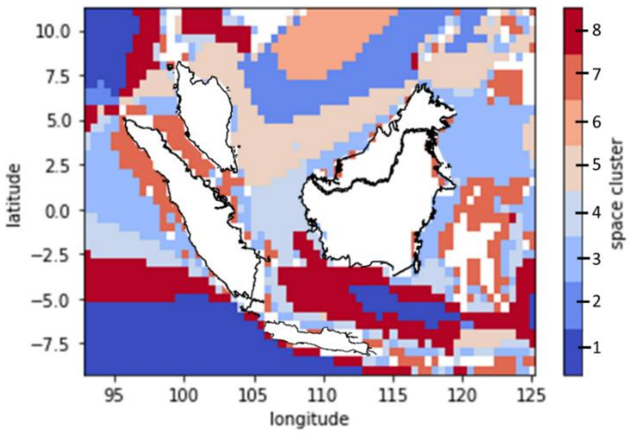


**Figure 6.** Space cluster for the South China Sea.

Next, figure 7 depicts the time clusters for each cluster. The total clusters that are available are 5 and each of the time clusters indicate the significant wave height. All the discussion for this result in figure 7 will focus on South China Sea area. For example, the blue color from cluster 1 depicts that the significant wave height is 0.2 – 0.4 m. For time cluster 2, in the South China Sea area, the significant wave height is from 0.2, 0.4 and 0.8 which indicates blue, cornflower-blue, and light grey. Next, for time cluster 3, the significant wave height is range from 0.2 – 0.4 m. In addition, the color that is involved in cluster 3 is blue and cornflower-blue. Next, in time cluster 4, the range of significant wave height is from 0.2 – 1.2 m where the color involved is blue, cornflower-blue, peach and light grey. However, there is one location that depicts the highest wave height which is 1.4 m in time cluster 2 and time cluster 5. Lastly, for time cluster 5, the wave height is between range 0.2, 0.4, 0.8 and 1.4 m.

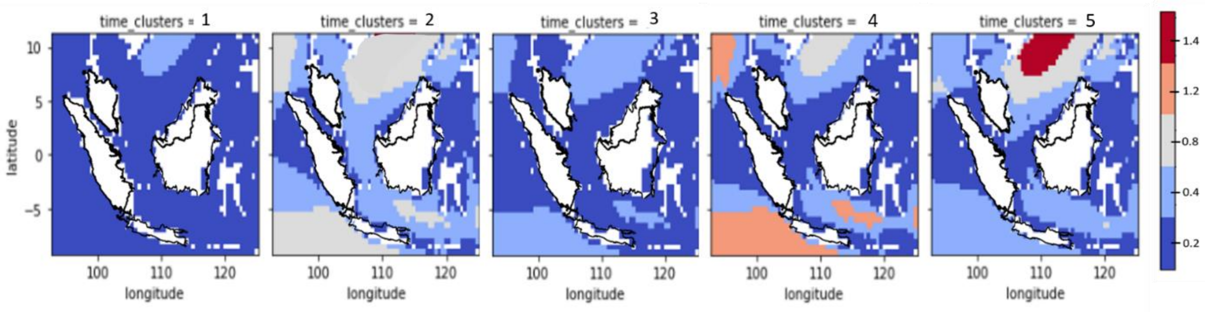


**Figure 7.** Time cluster for the South China Sea.

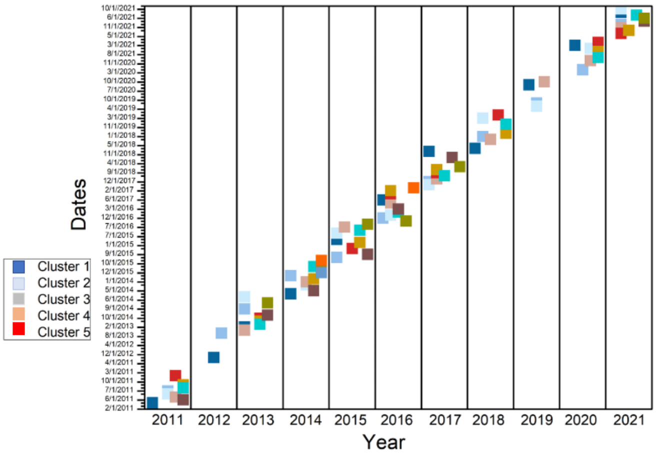


**Figure 8.** Cluster Plot for the extracted Cluster and Dates for time cluster.

Figure 8 depicts the cluster plot result for each cluster and corresponds to specific dates (temporal). This figure shows the exact cluster were define by year and related to the exact date. For example, the intersection for cluster 1 between the space and temporal clusters is highlighted in cornflower-blue (Figure 7), and the exact dates assigned for cluster 1 are 2/1/2011, 6/1/2011, 12/1/2012, 2/1/2013, 5/1/2014, 6/1/2014, 7/1/2015, 3/1/2016, 6/1/2016, 4/1/2017, 11/1/2017, 11/1/2018, 7/1/2019, 10/1/2019, 3/1/2020, 5/1/2020, and 6/1/2021(Figure 8). Another example is that there is only one peach color depicted in figure 5, and the same for figure 7 in cluster 4. This is due to the spatial distribution. For cluster 5, the red color is separated in Figure 5, this is because of the spatial distribution and the time cluster which is assigned in time cluster 5. Similarly, the information obtained from each result is that each cluster is assigned to a specific cluster, the area in South China Sea, and exact dates based on significant wave height, and lastly, they are generally related to each other.

## 5. Conclusion

This study has implemented a multi-dimensional clustering method to extract the spatial and temporal of wave patterns using the Bregman Block Average Co-Clustering with I-Divergence algorithm. The main co-cluster result depicts the location and years of wave patterns in the form of heatmap. The finding showed that, clusters that were distributed in South China Sea area are belong to cluster 2, 3, 4,5 and 6. Furthermore, the significant wave heights for these clusters were range from 0.4 to 1.2 m. For the time clusters, the highest wave is 1.4 m height and it just appears only in one location in time cluster 5. Since this study has only provided the preliminary finding of co-clustering wave patterns, hence, estimation of ocean energy that can be converted from the wave patterns is not further discuss in this work. The energy estimation from the spatio-temporal wave patterns can be potentially explored in future work. Furthermore, further study of wave patterns can be elaborated in detail by considering the monsoon seasons period; the Southwest Monsoon and Northeast Monsoon and better visualization to present the cluster. This new discovery could help in identifying which period of monsoon has the highest potential for harvesting ocean energy along the Malaysian coastal area.

## References

[1] E.Chuah, "DESIGN AND MODELLING OF WAVE ENERGY CONV ERTER AN D POWER TA KE -O FF SYSTEM E. Chuah 1 , K. Magaswaran 2 and M.Z. Hassan 3 1," 2017.

[2] N. H. Idris, "Wave energy resource assessment with improved satellite altimetry data over the Malaysian coastal sea," *Arab. J. Geosci.*, vol. 12, no. 15, 2019, doi: 10.1007/s12517-019-4670-z.

[3] C. T. Kuo, J. Bailey, and I. Davidson, "A framework for simplifying trip data into networks via coupled matrix factorization," *SIAM Int. Conf. Data Min. 2015, SDM 2015*, pp. 739–747, 2015, doi: 10.1137/1.9781611974010.83.

[4] J. Hagenauer and M. Helbich, "Hierarchical self-organizing maps for clustering spatiotemporal data," *Int. J. Geogr. Inf. Sci.*, vol. 27, no. 10, pp. 2026–2042, 2013, doi: 10.1080/13658816.2013.788249.
[5] G. Govaert, "Simultaneous clustering of rows and columns," *Control Cybern.*, vol. 24, no. 4, pp. 437–457, 1995.
[6] Y. Kluger, R. Basri, J. T. Chang, and M. Gerstein, "Spectral biclustering of microarray data: Coclustering genes and conditions," *Genome Res.*, vol. 13, no. 4, pp. 703–716, 2014, doi: 10.1101/gr.648603.
[7] Y. Mellek and A. Kaya, "Spatio-temporalclustering:Neighbourhoodsbasedonmedianseasonalentropy." 2020.
[8] J. A. Hartigan, "Direct clustering of a data matrix," *Journal of the American Statistical Association*, vol. 67, no. 337. pp. 123–129, 1972, doi: 10.1080/01621459.1972.10481214.
[9] M. Das and S. K. Ghosh, "Data-Driven Approaches for Spatio-Temporal Analysis: A Survey of the State-of-the-Arts," *J. Comput. Sci. Technol.*, vol. 35, no. 3, pp. 665–696, 2020, doi: 10.1007/s11390-020-9349-0.
[10] X. Wu, R. Zurita-Milla, and M. J. Kraak, "Co-clustering geo-referenced time series: exploring spatio-temporal patterns in Dutch temperature data," *Int. J. Geogr. Inf. Sci.*, vol. 29, no. 4, pp. 624–642, 2015, doi: 10.1080/13658816.2014.994520.
[11] S. Kisilevieh, F. Mansmann, M. Nanni, and S. Rinzivillo, "Spatio-temporal c1ustering," pp. 855–874, 2010.
[12] F. M. Saimi, F. M. Hamzah, M. E. Toriman, O. Jaafar, and H. Tajudin, "Trend and linearity analysis of meteorological parameters in peninsular Malaysia," *Sustain.*, vol. 12, no. 22, pp. 1–19, 2020, doi: 10.3390/su12229533.
[13] S. C. Madeira and A. L. Oliveira, "Biclustering algorithms for biological data analysis: A survey," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 1, no. 1, pp. 24–45, 2004, doi: 10.1109/TCBB.2004.2.
[14] T. Tokuda *et al.*, "Multiple co-clustering based on nonparametric mixture models with heterogeneous marginal distributions," *PLoS One*, vol. 12, no. 10, pp. 1–29, 2017, doi: 10.1371/journal.pone.0186566.
[15] A. Banerjee, "A Generalized Maximum Entropy Approach to Bregman Co-clustering and Matrix Approximation," pp. 1–76, 2007.
[16] J. P. Jiawei Han, Micheline Kamber, *Data mining: Data mining concepts and techniques*. 2012.
[17] G. Lin *et al.*, "Assessment of wave energy in the South China sea based on GIS technology," *Adv. Meteorol.*, vol. 2017, 2017, doi: 10.1155/2017/1372578.
[18] X. Wu, C. Cheng, R. Zurita-Milla, and C. Song, "An overview of clustering methods for geo-referenced time series: from one-way clustering to co- and tri-clustering," *Int. J. Geogr. Inf. Sci.*, vol. 34, no. 9, pp. 1822–1848, 2020, doi: 10.1080/13658816.2020.1726922.
[19] I. Ismail, W. S. W. Abdullah, A. @ M. S. M. Muslim, and R. Zakaria, "Physical Impact of Sea Level Rise To the Coastal Zone Along the East Coast of Peninsular Malaysia," *Malaysian J. Geosci.*, vol. 2, no. 2, pp. 33–38, 2018, doi: 10.26480/mjg.02.2018.33.38.

**Acknowledgment**
This research is supported by TIER 2 research grant (PY/2019/00505: Q.J130000.2652.17J11) from Universiti Teknologi Malaysia.