Coastal Education & Research Foundation, Inc.

Segmentation of Marine Forecast Zone in Korea by Using Wave Data

Author(s): Myung-Seok Kim, Doyoun Kim, Hyunmin Eom, Sung Hyup You, Jae-Soon Jeong and Seung-Buhm Woo

Source: *Journal of Coastal Research*, FALL 2021, Special Issue No. 114. Crisis and Integrated Management for Coastal and Marine Safety (FALL 2021), pp. 251-255

Published by: Coastal Education & Research Foundation, Inc.

Stable URL: https://www.jstor.org/stable/10.2307/48638757

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/10.2307/48638757?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms



Coastal Education & Research Foundation, Inc. is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Coastal Research*

This content downloaded from
165.246.122.112 on Mon, 11 Mar 2024 07:20:34 +00:00
All use subject to https://about.jstor.org/terms

Journal of Coastal Research	SI	114	251–255	Coconut Creek, Florida	2021

Segmentation of Marine Forecast Zone in Korea by Using Wave Data: Based on K-Means Clustering Algorithm

Myung-Seok Kim[†], Doyoun Kim[‡], Hyunmin Eom[§], Sung Hyup You[§], Jae-Soon Jeong[†], and Seung-Buhm Woo[†*]

[†]Department of Oceanography
Inha University
Incheon, Republic of Korea

[‡]Oceanography Division
Ara Consulting & Technology
Incheon, Republic of Korea

[§]Marine Meteorology Division
Korea Meteorological Administration
Seoul, Republic of Korea

www.cerf-jcr.org

www.JCRonline.org

ABSTRACT

Kim, M.-S.; Kim, D.; Eom, H.; You, S.H.; Jeong, J.-S., and Woo, S.-B., 2021. Segmentation of marine forecast zone in Korea by using wave data: Based on k-means clustering algorithm. *In:* Lee, J.L.; Suh, K.-S.; Lee, B.; Shin, S., and Lee, J. (eds.), *Crisis and Integrated Management for Coastal and Marine Safety*. *Journal of Coastal Research*, Special Issue No. 114, pp. 251–255. Coconut Creek (Florida), ISSN 0749-0208.

The Korea Meteorological Administration (KMA) classified the forecast zone based on the administrative districts, weather characteristics, civil complaints, and location of observation points. A cluster analysis was performed using the wave observation and model data to check the validity of the current forecast zone for the wave forecast. The cluster analysis is a method used to group areas with common characteristics. In this study, a set of n objects was decomposed into k clusters (groups) and a distance-based method, k-means clustering algorithm, was used. As the data structure of the cluster analysis, the mean and standard deviation of the significant wave height and wave period during the period of 2014 to 2016 were selected. When performing the cluster analysis with wave heights and wave periods observed at ocean buoys, 11–12 cluster conditions have relatively significant internal evaluation indices suggesting the distance and variance ratio between points and points within each cluster or between distinct clusters. However, since there were few points to perform the segmentation of the marine forecast zone using only the observation data, there was a problem that clusters concentrated in a specific area were distributed. Therefore, it is necessary to perform the cluster analysis by using the KMA wave model for the segmentation of the marine forecast zone. As a result, 10 cluster condition showed the best internal evaluation indices. Although the results of the cluster analysis did not differ significantly from the current forecast zone, the segmentation was made more distinct and detailed.

ADDITIONAL INDEX WORDS: *Marine forecast zone, wave observation and model data, cluster analysis, internal evaluation indices*.

INTRODUCTION

Korea Meteorological Administration (KMA) has been operated the marine zone for wave forecasting along the West, South, and East Sea of Korea. The separation of the marine forecast zone was determined based on the administrative districts and the distance from the coastline. It has been updated based on long-term forecasting experience. In this study, cluster analysis was performed using wave observations and wave model data to confirm the validity of the segmented marine forecast zone. It was possible to find a limited previous study on the segmentation of the marine forecast zone in Korea based on cluster analysis. Son, Byon, and Lee (2015) discussed the validity of the marine forecast zone through cluster analysis of wind speed and wave height which are key variables in the marine forecast. They suggested that when the group number was 6, the segmentation of the zone was most similar to the current marine forecast zone and it showed the most distinct cluster analysis.

Also, there has been active research on segmentation of area using cluster analysis in other research areas. Based on cluster analysis, Ahmed and Akter (2017) estimated changes in landcover in floods affected coastal areas by flood events in the southwest coastal area of Bangladesh. Similarly, Carvalho *et al.* (2016) identified regions with consistent climate changes in precipitation in Europe. Rehman, Burton, and Weatherill (2014) also used cluster analysis to define seismicity partitions for Pakistan. Burton and Hall (2014) explained the segmentation of the Sumatran fault by finding the optimal number of the fault derived from the cluster analysis.

Based on the previous study (Son, Byon, and Lee, 2015), we try to find more optimal group number for the segmentation of the marine forecast zone in Korea by using the clustering algorithm. Again, this study was conducted not to criticize the current marine forecast zone as incorrect, but to verify the feasibility of the segmented region. In particular, the forecast zone is over-subdivided according to administrative districts. In some cases, different risk levels can be set by grouping the forecast zone based on wave characteristics of each region. The results of this study can be utilized as a kind of forecast guidance.

METHODS

For cluster analysis, a set of n objects was decomposed into k clusters and a distance-based method, k-means clustering algorithm, was used (MacQueen, 1967). The clustering

DOI: 10.2112/JCR-SI114-051.1 received 20 November 2020; accepted in revision 18 January 2021.
*Corresponding author: sbwoo@inha.ac.kr
©Coastal Education and Research Foundation, Inc. 2021

This content downloaded from
165.246.122.112 on Mon, 11 Mar 2024 07:20:34 +00:00
All use subject to https://about.jstor.org/terms

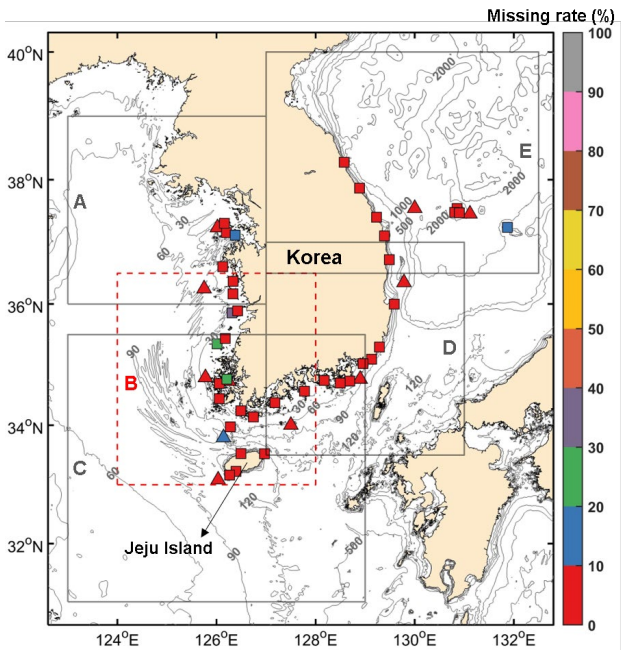


Figure 1. Study area, depth contours (m), locations of 49 ocean buoys (11 triangles: ocean data buoy, 38 squares: coastal wave buoy), and 5 wave model domains (A-E square bounding boxes). The observations and model were operated by Korea Meteorological Administration (KMA). Color of the triangles and squares denotes the missing rate of wave observations at each buoy during the period of 2014 to 2016.

algorithm for the cluster C_k and the k centroids μ_k can be expressed as follows (Carvalho *et al.*, 2016):

$$\text{Minimize } \sum_{k=1}^{K}\sum_{x_n \in C_k}\left\|x_n - \mu_k\right\|^2 \text{ with respect to } C_k, \mu_k \quad (1)$$

In addition, to evaluate the goodness of clustering results, internal evaluation for each cluster condition was performed. Internal evaluation gives a high score to a cluster condition with high intra-cluster similarity and low inter-cluster similarity between clusters. As the internal evaluation indices, the higher the average silhouette value and the lower the Davies–Bouldin Index, the better the clustering results (Arroyo *et al.*, 2017).

It is important to select an appropriate item condition to construct the data structure used for cluster analysis. We determined the significant wave height and wave period, which are representative variables of wave characteristics, as the data structure of the cluster analysis. As shown in Figure 1, wave observations from the ocean data buoys and the coastal wave buoys at 1-hour intervals provided by KMA were used. The missing rate of most wave observations was less than 10%. In addition, cluster analysis based on the Coastal Wave Watch III (CWW3), an officially verified KMA wave model, was performed for the segmentation of the marine forecast zone in high resolution. This wave model has a spatial resolution of about 1 km and an output interval of 3 hours. Currently, KMA

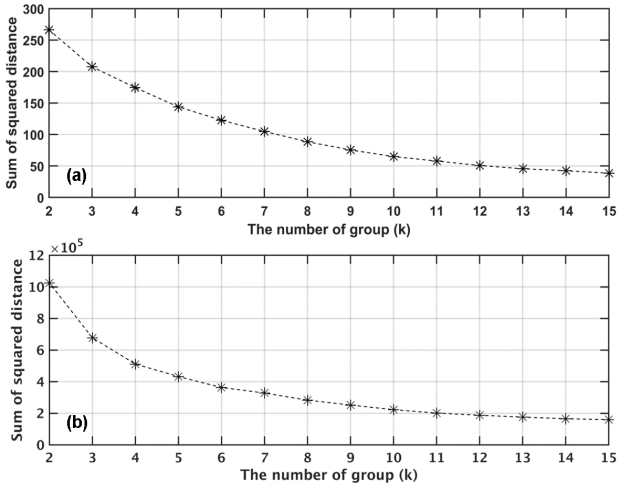


Figure 2. Sum of squared distance according to each cluster condition (group: k) based on k-means clustering algorithm of (a) wave observations (ocean data buoy and coastal wave buoy) and (b) wave model (Coastal Wave Watch III, CWW3).

operates CWW3 divided into five model domains (A-E), which covers the West, South, and East Sea of Korea. Of the five domains, the B domain overlapped with the A, C, and D domains in a large area (Figure 1), so it was excluded from cluster analysis using the wave model. Accordingly, a blank area occurred near 36°N latitude band between A and C (Figure 6–7).

Since it is not possible to use all the time series of wave data for three years from 2014 to 2016, it was necessary to perform cluster analysis with several statistical figures representing each location. Of the various statistical figures (*e.g.*, mean, standard deviation, range, median, minimum, and maximum), the mean and standard deviation, which are relatively less sensitive to outliers, were used for cluster analysis. Therefore, the cluster analysis data structure was constructed with the mean and standard deviation of significant wave height and wave period of each point including the longitude and latitude. In order to reduce the bias due to outliers in a specific variable, normalization by z-score was performed. The simulated wave data was assumed that there will be no location-related bias because they exist at each grid point, unlike observation. Accordingly, the longitude and latitude were not normalized. Afterward, when finding the optimal group number from the clustered results using the wave data, the cluster analysis results were monitored by increasing the number of groups (k) sequentially starting from k=8. This is because the sum of the distance of each data from the center of the cluster (Figure 2) stabilized from the condition above k=8.

RESULTS

When clustering with the observations, overall, results of more detailed and complex segmentation were shown in the West Sea than in the East Sea (Figure 3). Because the number of buoys relative to the zone was limited when performing the segmentation of the marine forecast zone, the cluster analysis results were highly dependent on the buoy location. Most of the

This content downloaded from
165.246.122.112 on Mon, 11 Mar 2024 07:20:34 +00:00
All use subject to https://about.jstor.org/terms

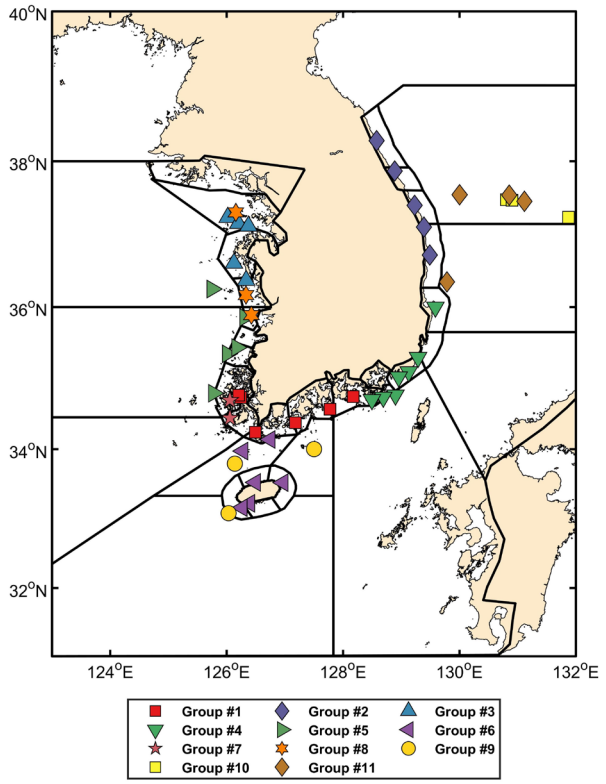


Figure 3. An example of cluster analysis results (k=11) based on k-means clustering algorithm using significant wave height and wave period observed from ocean buoys. Black solid line indicates current marine forecast zone operated by KMA.

buoys are located near the coastline, and cluster analysis results also focused on the coast. In addition, as the group number was smaller, the bias in a certain buoy, which is a kind of outlier, had a significant effect on the cluster analysis result. Therefore, in cluster analysis with only the wave observation data, it was not possible to determine which cluster condition was the optimal cluster from the perspective of segmentation of marine forecast zone. However, in the internal evaluation indices for each cluster condition, the average silhouette value and the Davies-Bouldin index suggested that the most optimal conditions in common were k=11–12 (Figure 4).

As a result of performing cluster analysis including the B domain (B bounding box in Figure 1), there was a problem that the overlapped regions were clustered separately, as shown as the red square bounding box in Figure 5. In particular, the result of the separated cluster near the 36°N latitude band (red contour in Figure 5) continued to occur regardless of the number of groups (k). Therefore, cluster analysis was performed excluding the B domain to minimize bias due to overlap of model domains. On the other hand, each overlap of model domains between C and D, and between E and D (Figure 1) seems to have little effect on segmentation for each cluster condition, based on the fact that cluster analysis was performed regardless of each model domain.

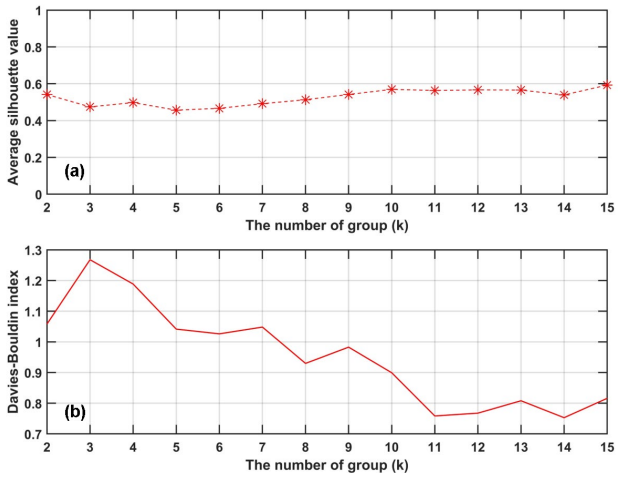


Figure 4. Internal evaluation indices for cluster analysis results of the wave observations. (a) Average silhouette value, (b) Davies-Bouldin index.

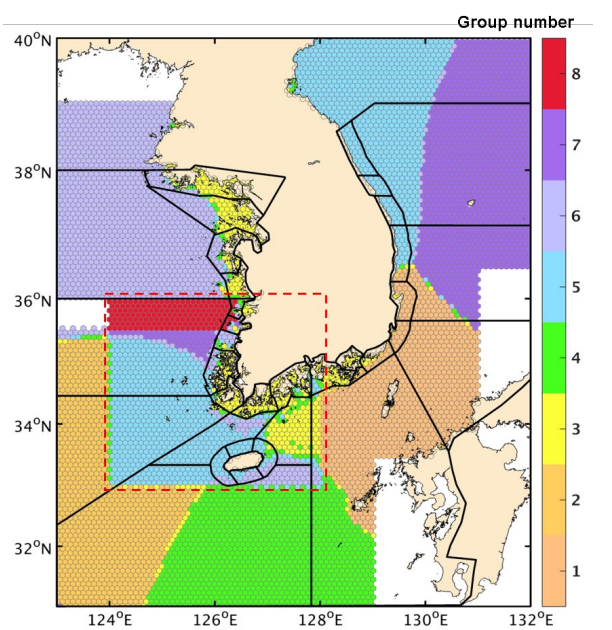


Figure 5. When the B domain was added, an example of cluster analysis results (k=8) using significant wave height and wave period of wave model (CWW3). Red square bounding box indicates biased region due to the B domain.

As suggested by Son, Byon, and Lee (2015), as the number of groups (k) increased, the segmentation was performed mainly in the West and South Sea of Korea (Figure 6–7). Conversely, the East Sea of Korea maintained three cluster conditions, only the location of segmentation slightly changed even when the number of groups (k) increased. And based on the Davies-Bouldin index, it was

This content downloaded from
165.246.122.112 on Mon, 11 Mar 2024 07:20:34 +00:00
All use subject to https://about.jstor.org/terms

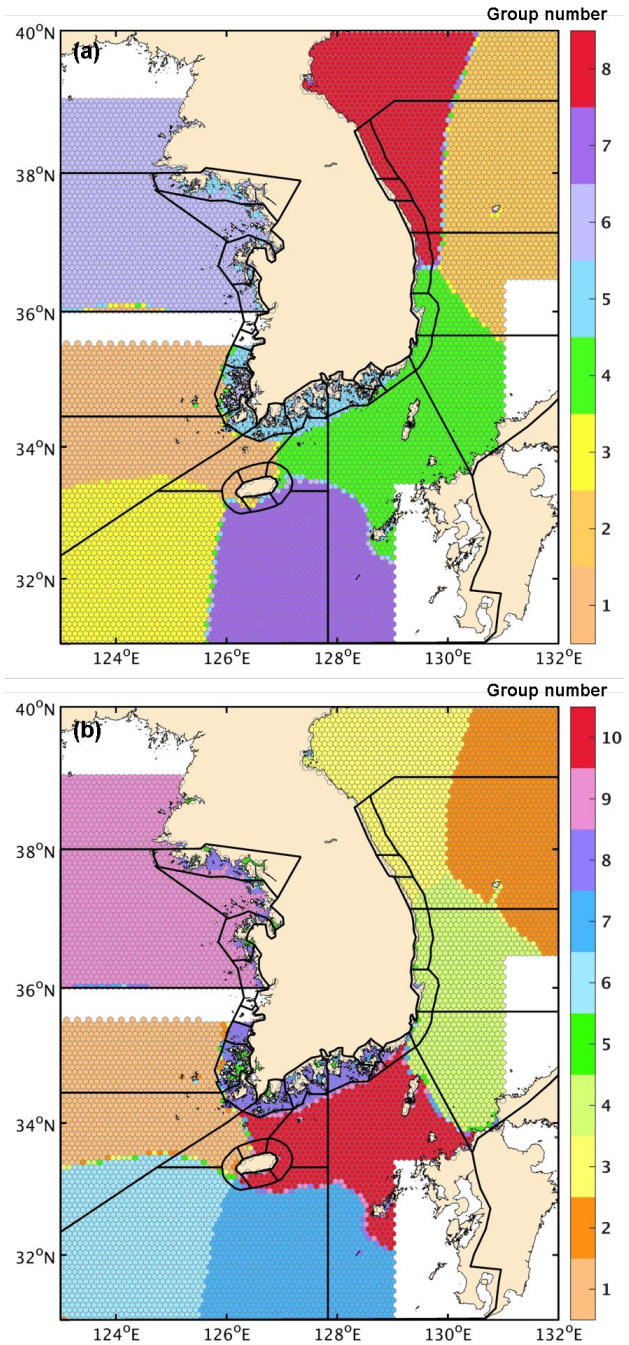


Figure 6. With the B domain excluded, cluster analysis results using significant wave height and wave period of wave model (CWW3) according to the number of groups (k). (a) k=8, (b) k=10. Black solid line indicates current marine forecast zone operated by KMA.

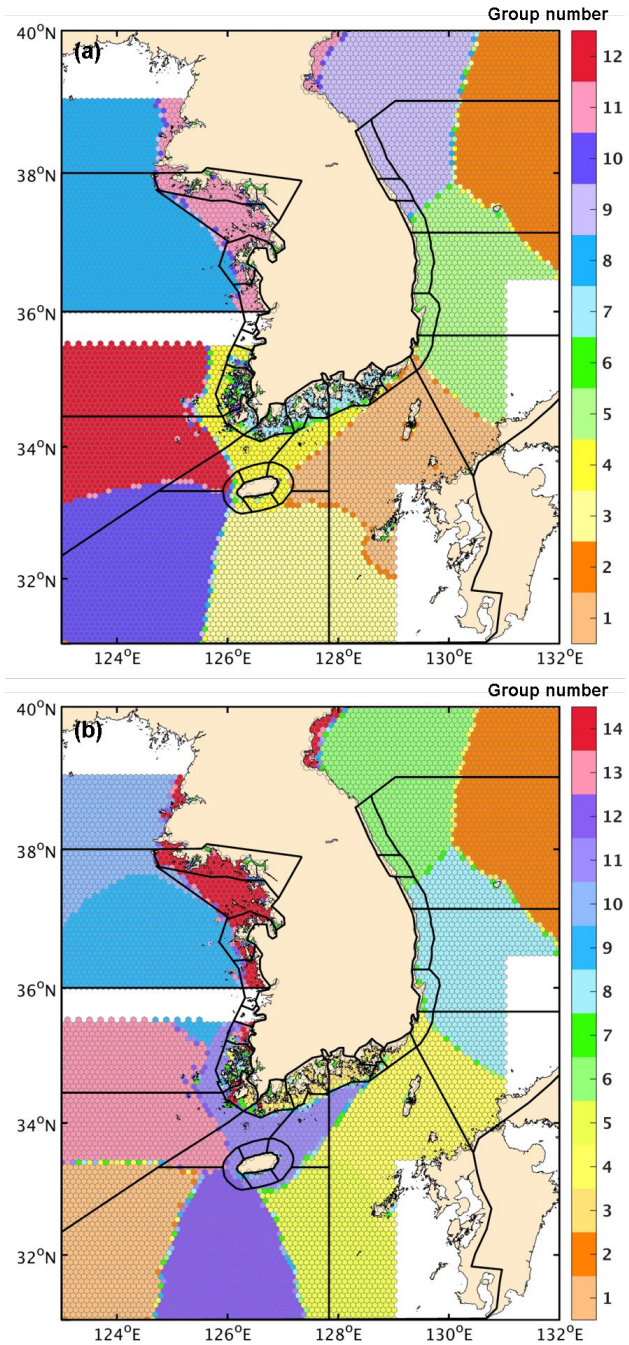


Figure 7. As in Figure 6, but for (a) k=12, (b) k=14.

confirmed that the clustering condition was the most prominent when k=10 (Figure 8b). The average silhouette value remained constant between 0.5 and 0.6 regardless of cluster conditions (Figure 8a). Overall, when k=10 (Figure 6b), the characteristics of the segmentation were as follows: (i) Boundaries of each segmentation in the open sea were most similar to boundaries of the current marine forecast zone. And (ii) the segmentation near the coastline was divided into one cluster in the West and South Sea, and two clusters in the East Sea.

This content downloaded from
165.246.122.112 on Mon, 11 Mar 2024 07:20:34 +00:00
All use subject to https://about.jstor.org/terms

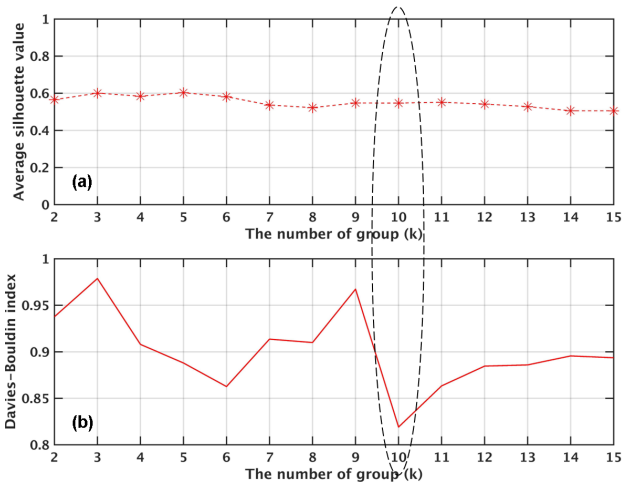


Figure 8. Internal evaluation indices for cluster analysis results of the wave model. (a) Average silhouette value, (b) Davies-Bouldin index.

DISCUSSION AND CONCLUSIONS

In this study, cluster analysis using wave data (significant wave height and wave period) was performed to find the optimal group number for segmentation of the marine forecast zone in Korea. The segmentation was conducted by using the wave model (CWW3) due to the limitation of the distribution of wave observations too concentrated on the coast. Conclusively, we determined that the most reasonable segmentation was performed when k=10 among several group numbers (k=2–15) based on the following criteria:

(1) As shown in an ellipse in Figure 8b, when k=10, a significant internal evaluation index (Davies-Bouldin index) was derived. The Davies-Bouldin index is defined as a function of the ratio of the within-cluster scatter to the between cluster separation, a lower value indicates that the clustering is better (Arroyo *et al.*, 2017).

(2) Unlike segmentation of the West and South Sea of Korea, segmentation of the East Sea showed no distinction between the coastal and the open sea (Figure 6b) due to different depth profile (depth contours in Figure 1). In addition, the southwestern coast of Korea, which is characterized as shallow water depth of less than 30 m and likely to have similar wave characteristics, was divided into the same cluster.

(3) Similar to previous study (Son, Byon, and Lee, 2015), the segmentation was performed, which was divided into each different zone (Figure 6b), especially around Jeju Island (Figure 1).

The optimal group number in this paper was different from the optimal group number (k=6) suggested in the previous study, but interestingly, the segmented results were similar to the current marine forecast zone. Based on these results, further discussion and consideration of the subdivision of the marine forecast zone are required when updating the forecast zones or predicting waves. Also, it is expected that regions of interest or relatively less divided regions (*e.g.*, region of the A domain in Figure 6b) can be further subdivided according to needs. In the future, it is necessary to supplement the current segmentation results by performing additional cluster analysis for a relatively long period.

ACKNOWLEDGMENTS

This research was a part of the project titled 'Gyeonggi Sea Grant Program', funded by the Ministry of Oceans and Fisheries, South Korea. Additionally, this research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2019R1A6A3A13097301).

LITERATURE CITED

Ahmed, K.R. and Akter, S., 2017. Analysis of landcover change in Southwest Bengal delta due to floods by NDVI, NDWI and k-means cluster with landsat multi-spectral surface reflectance satellite data. *Remote Sensing Applications: Society and Environment*, 8, 168–181.

Arroyo, Á.; Herrero, Á.; Tricio, V., and Corchado, E., 2017. Analysis of meteorological conditions in Spain by means of clustering techniques. *Journal of Applied Logic*, 24, 76–89.

Burton, P.W. and Hall, T.R., 2014. Segmentation of the Sumatran fault. *Geophysical Research Letters*, 41(12), 4149–4158.

Carvalho, M.J.; Melo-Gonçalves, P.; Teixeira, J.C., and Rocha, A., 2016. Regionalization of Europe based on a k-means cluster analysis of the climate change of temperatures and precipitation. *Physics and Chemistry of the Earth*, 94, 22–28.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. *In*: Cam, L.L. and Neyman, J. (eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.

Rehman, K.; Burton, P.W., and Weatherill, G.A., 2014. K-means cluster analysis and seismicity partitioning for Pakistan. *Journal of Seismology*, 18, 401–419.

Son, G.; Byon, J.-Y., and Lee, J.-H., 2015. The classification of marine forecast zone and their characteristics around the Korean Peninsula using cluster analysis. *Journal of the Korean Data Analysis Society*, 17(4), 2129–2138.

This content downloaded from
165.246.122.112 on Mon, 11 Mar 2024 07:20:34 +00:00
All use subject to https://about.jstor.org/terms