Check for updates

Application of carbon emission prediction based on a combined neural algorithm in the control of coastal environmental pollution in China

Bin Wang*

School of Naval Architecture and Ocean Engineering, Jiangsu University of Science and Technology, Zhenjiang, China

The marine ecosystem provides the environment, resources, and services necessary for the development of every human society. In recent years, China's coastal zone has been polluted to varying degrees, which has seriously affected its development. The characteristics of marine environmental data include the variety of data types, the complexity of factors affecting the marine environment, and the unpredictability of marine pollution. Currently, there are few studies applying the clustering analysis algorithm to marine environmental monitoring. Then, carbon emissions (CEs) from coastal areas are predicted using marine environmental data. Therefore, this paper mainly studies the spatial and temporal accumulation characteristics of marine environmental data and uses the fuzzy c-means (FCM) algorithm to mine the data monitored by the marine environment. Meanwhile, it has been focused on the prediction of coastal CEs, and the grey model-back propagation (GM-BP) algorithm has been developed to predict CEs from coastal areas, which solves the problem that the traditional back propagation neural network (BPNN) cannot fully learn data features, which leads to a decline in accuracy. The experimental results showed that the FCM algorithm can divide the marine sample data into corresponding categories to distinguish polluted and unpolluted samples. The improved neural network model has a higher degree of non-linear fit and lower prediction error than a back propagation (BP) neural network. The main contribution of this paper is to first study the spatial and temporal accumulation characteristics of marine environmental data. The academic contribution of this study is to substitute the predictions of the three gray models (GMs) with the neural network structure simulation to finally obtain more accurate predictions. From a practical point of view, this study is helpful to a certain extent in alleviating the pressure of climate change due to increased CEs in global coastal zones. This study can also provide a new method of measuring environmental governance for marine environmental regulatory authorities.

KEYWORDS

coastal zone, environmental governance, CE, FCM, GM-BP

1. Introduction

The coastal zone is a transitional zone between land and sea. It is a special zone extending from the coastline to both sides of land and sea, with a strong interaction between land and sea. It has the characteristics of two different environments: land and sea. With rich biodiversity and a comfortable climate, the coast is suitable for the survival and development of human beings and is the focus of the national economy and population. More than half of the world's coastal areas are related to the development of the human population, which is the key to the development of every human society. However, with the intensification of human activities, a large number of pollutants are discharged into the coastal zone in a variety of ways. The environmental pollution caused by high-intensity human activities has led to the degradation of the coastal zone, the earth's key zone. Therefore, exploring the current situation of coastal pollution in China and putting forward reasonable and effective pollution control measures will play an important role in ensuring the ecological security of coastal zones, ensuring public health, and maintaining the sustainable development of coastal zones. The global carbon cycle is constantly changing under the influence of natural changes and human beings, and the carbon cycle in coastal waters is also constantly changing. The diversity of carbon sources and carbon sinks and the interaction between them increase the complexity of research (Li, 2018; Liu et al., 2018). Experiments showed that increased temperature and decreased precipitation will slow down the decomposition of soil organic matter, thus slowing down the positive feedback effect of human-induced changes on the climate. Human impacts on the coastal zone and the ocean are becoming more and more intense, and some human activities have weakened the carbon sink in the coastal zone (Vondolia et al., 2020; Fu et al., 2021).

Data mining technology can be used for statistics and the analyses of data so as to guide the decision-making of practical problems in real life, discover the relationships between things, and predict the future work. The marine environment monitoring involves a wide range of sea areas, and a large amount of monitoring data will be obtained. It is also possible to monitor only a certain station or sea area, and the data acquisition may be a long-term series or a specific time period (Sun et al., 2018). Data mining technology is an application-oriented research field. When dealing with the marine environment, we need to know who the specific users are, and the knowledge obtained according to the needs of users is useful and valuable. Although data mining technology has achieved certain results in the processing of marine environmental monitoring data, the development of a data mining system is not perfect because of the particularity of marine environmental data, the diversity of data types, the complexity of factors affecting the marine environment, and the unpredictability of marine pollution. Some data mining methods still have some problems in marine data processing, such as unclear mining tasks, sensitivity to noise, and low efficiency (Zhang, 2018).

In this study, the spatial and temporal accumulation of marine environmental data is studied; at the same time, it focuses on the prediction of coastal carbon emissions (CEs) and uses the GM-BP algorithm to predict the CEs in coastal areas, which provides a new method for the marine environmental supervision department.

2. Literature review

2.1. Coastal environment management based on data mining

Some developed countries have achieved some theoretical research results in the fields of marine data management and marine information processing (Zhuang et al., 2016; Tanhua et al., 2019; Lou et al., 2021). The United States is one of the first countries to develop digital mapping products for deep sea areas. As early as 1990, the United States established the digital bathymetric data center of the international island survey organization. It is expected to establish a global digital bathymetric data warehouse to store, manage, and mine the data of areas deeper than 100 m. China has also gradually established the information system for sea area use and marine environmental protection management information system so as to make reasonable use of and store the information resources of the marine information center. In terms of data mining technology, more and more scholars use the characteristics of marine information data to further explore the data mining technology applied in the marine environment. Chen (2014) improved the clustering algorithm using the concepts of specified distance and reachable distance of sub-trajectories in a trajectory division, which solved the problem of the sensitivity of the algorithm to input parameters, and applied the improved algorithm to the clustering of hurricane tracks to obtain an internal clustering structure of hurricane tracks. Zhou puts forward a fuzzy clustering algorithm based on weight. By clustering the physical and chemical factors of red tide, the occurrence and development of red tide were divided into four stages (Zhou and Yang, 2016). Liu also analyzes the physical and chemical factors of red tide, designs an open-source, cross-platform marine environment data mining system, improves the fuzzy algorithm based on similarity relationships, and processes the marine data to obtain the main factors leading to the occurrence of red tide, which can give an early warning on whether red tide occurs (Liu L. et al., 2016). Geman et al. (2016) analyze the ecological environment remote sensing data of Bohai Bay using the k-means algorithm in cluster analysis and the Apriori algorithm in correlation analysis, respectively, to analyze and predict the environmental conditions of Bohai Bay and

put forward reasonable suggestions (Geman et al., 2016). In addition to the application of data mining to the analysis of the marine environment at the algorithmic level, there is also the storage of massive marine environmental monitoring data at the technical level. Liu gridded the marine environmental monitoring data using spatial interpolation and a spatiotemporal difference algorithm, making it convenient for users to analyze data at multiple levels and dimensions (Liu L. et al., 2016).

The types of marine data samples are diverse. The marine monitoring data include the most basic environmental data, such as temperature, seawater salinity, and dissolved oxygen concentration, as well as remote sensing data, marine culture, and other statistical data related to the marine economy (Zhang, 2018). These categories can also be subdivided into many small categories, which leads to the fact that the data types of marine data samples are very complex and huge. At the same time, there is little research on the application of clustering analysis algorithms to marine environmental monitoring. This study considered the complex factors affecting the marine environment, with characteristics of diversity, suddenness, disorder, and randomness, which cause fuzziness in the marine data, and used the fuzzy C-means (FCM) algorithm to mine some basic monitoring data of the marine environment.

2.2. Prediction of CE

Carbon dioxide emitted by human beings mainly comes from urban areas. Approximately 50% of the world's population lives in cities, especially in economically developed coastal areas. At least 70% of carbon dioxide released by fossil fuels and a large amount of methane generated by human activities come from this source (Krause-Jensen and Duarte, 2016). In the construction of the smart ocean, with the continuous development of the shipping industry, there are an increasing number of ships on the sea, and CEs from ships has increased significantly, which is the main source of CE in the coastal zone. CEs from ships is based on the ship energy efficiency operation index, which represents carbon dioxide emissions per unit voyage and is directly related to the ship's fuel consumption. The index can be used as an objective calculation method for monitoring the energy efficiency of operating ships. Zheng et al. established the relationship between CE, fuel consumption, and ship Automatic Identification System (AIS) information. Through the obtained AIS data information, the variables needed to calculate the EEIO formula were converted into the data contained in AIS to estimate the value of CEs (Zheng et al., 2014). Yao et al. calculated CEs from ships according to the ship type, linked the length of the ship with the ship power, and directly calculated the CE value using the static AIS data (Yao et al., 2017).

Summarizing the current research on China's CE peak prediction and total amount control, it can be found that

scholars mainly analyze the influencing factors of total CEs through the following methods (Chai et al., 2017; Dong et al., 2019), namely, the STIRPAT model, the logarithmic mean Divisia index (LMDI) factor decomposition model, the environmental Kuznets curve (EKC) curve, and other means, that were established to decompose CEs of energy consumption so as to put forward corresponding countermeasures and suggestions for China to eliminate backward production capacity and accelerate the upgrading of the industrial structure. For the prediction of total CEs, most of the existing studies used the partial least squares regression method, the STIRPAT model, and the scenario analysis method to predict the peak of China's CE, and some scholars combined various methods to predict China's CEs (Yang et al., 2018; Li et al., 2021). Jin et al. used the radial basis function (RBF) to predict the urban CE content in China from 2027 to 2032 (Jin, 2021). The experimental results showed that the number of samples will affect the prediction ability and network structure complexity of RBF, and the network is not suitable for large sample data prediction. Kashi et al. used a multilayer perceptron, RBF, and adaptive neuro-fuzzy inference system model to predict the dissolved oxygen, biochemical oxygen demand, and chemical oxygen demand levels of Iran's Karun river (Emamgholizadeh et al., 2014). Zhang et al. (2018) proved that the back propagation (BP) network can be used to predict the evolution of sea ice (Zhang et al., 2018). The experimental results showed that the BP algorithm is more satisfactory than the least squares method in predicting the ice melting process. Although research on the shallow neural network (SNN) in predicting marine data is quite effective, this kind of model has a single structure and can only capture simple data features, which is suitable for processing the mapping relationships of small sample data. In the face of marine environmental data with mass and diversity characteristics, SNN cannot fully learn the data features, which leads to a decline in accuracy (Jiang et al., 2018; Erichson et al., 2020).

Through the abovementioned analysis, it can be seen that, at present, neural networks have been widely used in predicting CEs, but most of them are for urban CEs. At the same time, the accuracy of SNN is reduced because it cannot fully learn the data features. Therefore, this study attempts to use the gray neural network prediction model to predict CEs in the coastal zone.

3. Clustering model of coastal environmental monitoring data based on FCM

3.1. Mining process of monitoring data

The types of coastal data samples are diverse. The marine monitoring data include the most basic environmental data, such as temperature, salinity, dissolved oxygen concentration,

etc., as well as remote sensing data, marine culture, and other statistical data related to the marine economy. The whole data mining process can be divided into three parts: the data preparation stage, the data mining stage, and the pattern evaluation and knowledge representation stage. In the first stage, the data preparation stage mainly includes data cleaning, data integration, and data transformation. The specific process is shown in Figure 1.

3.2. FCM clustering algorithm

Fuzzy C-means algorithm can deal with the problem of difficult batch processing under the condition of ocean big data, and it can realize the optimization of clustering centers under the condition of ocean big data. FCM achieves data clustering by minimizing the objective function and determines the number of genetic centers under the condition of marine big data via piecewise processing of a specific analysis through the genetic population of individuals under the condition of marine big data, combined with the characteristics of marine big data.

Let the n-dimensional sample data set be $X = \{x_1, x_2, x_j, x_n\}$, and each sample data have a S-dimensional property $x_k = \{x_{k1}, x_{k2}, x_{ks}\}$. If the FCM algorithm is used to cluster the sample data into C classes, V is the cluster center matrix, and U is the fuzzy matrix, the corresponding objective function of the algorithm is as follows:

$$J_m(U, V) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m D_{ij}^2 \quad (1)$$

where u_{ij} represents the membership degree of some sample data j and the ith category, and $0 \leq j \leq n, 0 \leq i \leq c$.

$D_{ij} = \| x_j - v_i \|$ in Equation (1) is expressed as the Euclidean distance between the jth data sample and the ith cluster center. The biggest difference between fuzzy clustering and hard clustering lies on this m, which represents the fuzzy weighting index of membership degree. If the value of m is 1, then it becomes a hard cluster, and as the value of m becomes larger, it means that it is fuzzier. When the FCM algorithm achieves the best clustering effect, the objective function obtains the minimum value. Therefore, using the Lagrangian multiplier method on $m \in [1, \infty)$, when the objective function satisfies the constraint $\sum_{i=1}^{c} u_{ij} = 1$ to obtain the minimum value and the optimal clustering effect is achieved, the corresponding fuzzy matrix and clustering center can be obtained.

$$J_m^*(U, V) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m D_{ij}^2 + \lambda \left(\sum_{i=1}^{c} u_{ij} - 1 \right) \quad (2)$$

The meaning of u_{ij} and D_{ij} is consistent with what is described in Equation (1). The partial derivatives in Equation (2) are calculated, and the fuzzy matrix and cluster center corresponding to the minimum value of the objective function are obtained via simplification, as shown in Equations (3) and (4).

$$U_{ij} = \left[\sum_{k=1}^{c} \left[\frac{D(x_j, v_i)}{D(x_j, v_k)} \right]^{\frac{1}{m-1}} \right]^{-1} \quad (3)$$

$$V_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m} \quad (4)$$

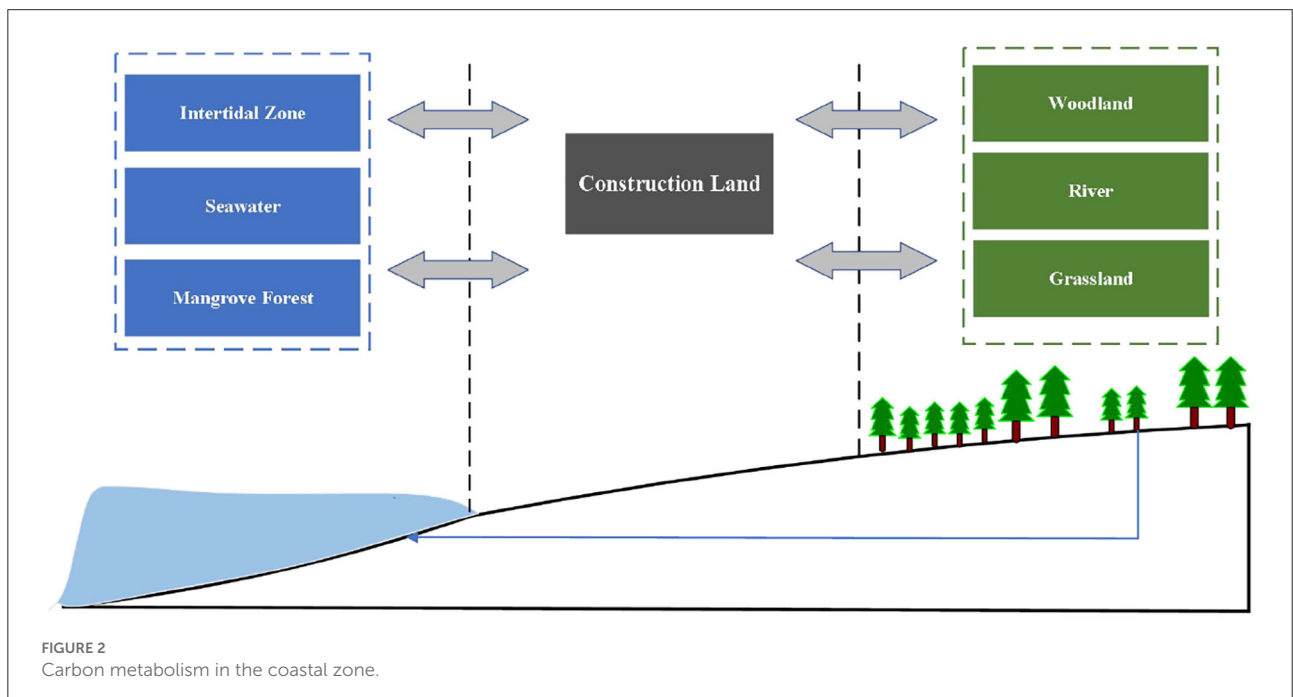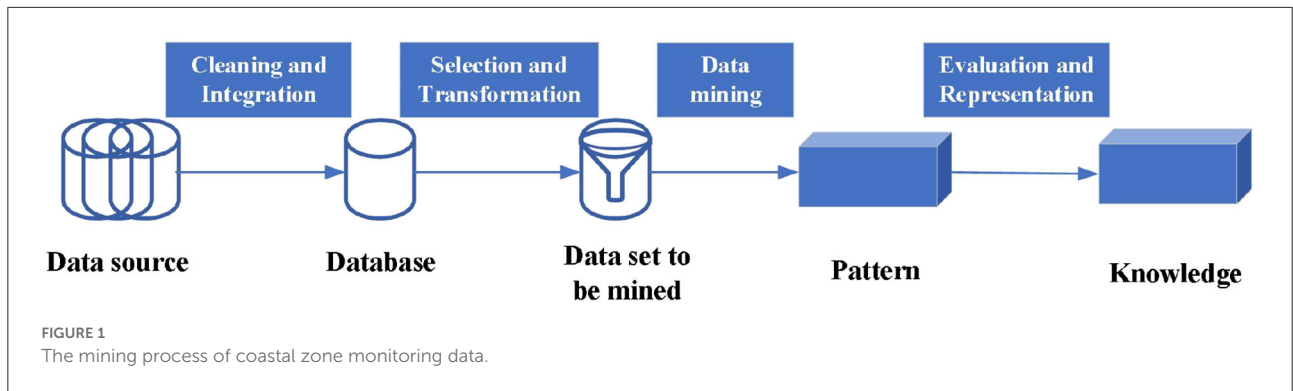## 3.3. Algorithm implementation process

A specific clustering process of the FCM clustering algorithm is as follows: the input is the sample object data n, the number of clustering categories c, and the output is the cluster center matrix V and the fuzzy matrix U. The specific steps are as follows:

(1) Set the specific number of categories of the sample data to be clustered, and initialize a series of parameters: the number of iterations, the threshold of the objective function ε, and the fuzzy weighting exponent m. Given any fuzzy matrix of initial classification and C initial cluster centers, it should be noted that the initial fuzzy matrix only needs to satisfy the constraint $\sum_{i=1}^{c} u_{ij} = 1$.

(2) According to Equation (3) and the value of the initial cluster center, the distance between the sample data and the cluster center is calculated to obtain the fuzzy matrix.

(3) According to Equation (4), the cluster center is updated by the fuzzy matrix.

(4) Loop process judgment if the objective function difference is less than the set objective function threshold and then exit the loop. Otherwise, return to Step (2).

4. CE prediction model based on the gray neural network

4.1. Carbon balance analysis of coastal cities

Urban carbon balance in coastal zones studies the relationship between CEs and carbon absorption. When urban CEs are equal to its carbon absorption, it is in equilibrium. Carbon input and output include vertical and horizontal carbon flow. In addition, carbon will stay in the carbon pool for a certain time. Carbon inputs include plant photosynthesis, fossil fuels, food, and building material transport; the carbon pool includes plants, soils, buildings, furniture, books, and living

FIGURE 1
The mining process of coastal zone monitoring data.

FIGURE 2
Carbon metabolism in the coastal zone.

organisms; and carbon output includes the respiration of plants, soils, and other living things, the burning of fossil fuels, and the decomposition of waste.

Carbon input and output can be divided into horizontal flow and vertical flow. The vertical flow of carbon includes vegetation and algae, among others, which is produced by the processes of carbon fixation by photosynthesis, fossil energy combustion, and biological respiration; the horizontal flow of carbon is the movement of carbon containing materials in the horizontal direction, such as the transfer of food and fiber from cultivated land, oceans, and forests to urban areas and the transfer of domestic waste from urban areas to land and sea landfill sites. Meanwhile, the horizontal carbon flow starts from or finally participates in the carbon cycle through the vertical flow. For example, organisms are transported to the market in a horizontal direction after being captured and then reemitted into the atmosphere through vertical biological respiration after

being consumed. There is no carbon absorption or CEs in the horizontal direction. CEs studied in this study are vertical. The simplified carbon metabolism in coastal cities is shown in Figure 2.

4.2. Gray neural network algorithm

The gray prediction method can eliminate the influence of random interference *via* the accumulation of a small number of sample data, and the accumulated series show a monotonically increasing trend, which can better predict the overall trend (Liu C. et al., 2016; Zhuang et al., 2021). In the modern emerging technology of information processing, the combination of gray model (GM) prediction and the BP neural network prediction can effectively predict CEs. Based on GM (1, 1), unbiased GM (1, 1) (WPGM), unequal GM (1, 1), the weighted GM (1, 1)

FIGURE 3
The topology of gray neural network prediction model.
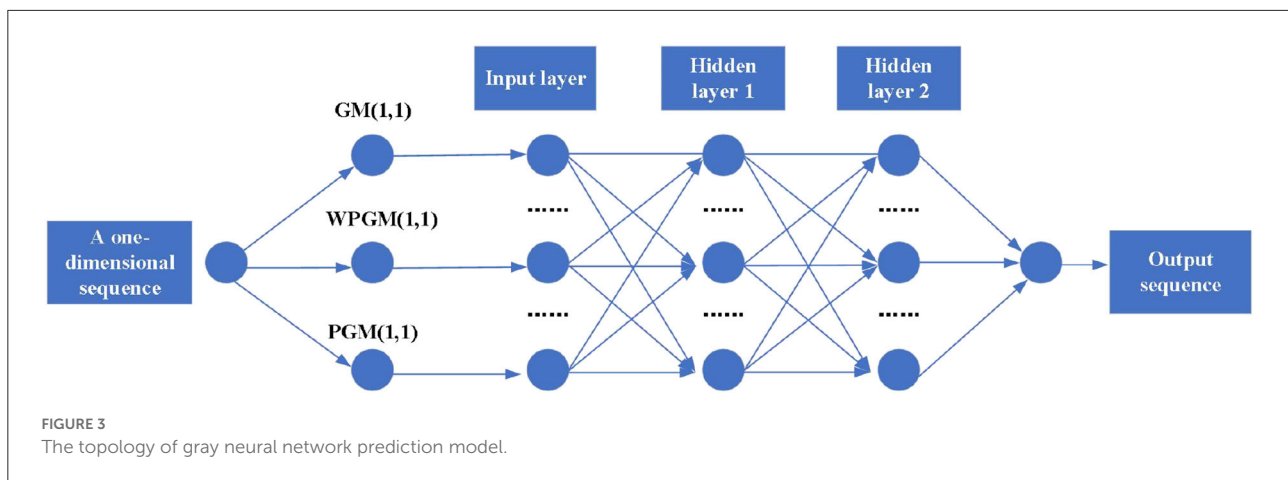
(PGM), and BP neural networks, the GM-BP model is proposed in this study. The optimal neural network structure is obtained by training the three sets of simulated values obtained from the one-dimensional sequence through three GMs as the input mode and from the original sequence as the output mode. The predictions of the three GMs are substituted into the neural network structure simulation, and the predictions are finally obtained. In regression problems, the mean square error (MSE) loss function is used to measure the distance between sample points and the regression curve, and the sample points can fit the regression curve better by minimizing the square loss. The MSE loss function is adopted in the model in this study. The structure of the GM-BP prediction model is shown in Figure 3.

4.3. Modeling steps

Let m be the original sequence length and n be the predicted sequence length. The structure of the two hidden layers of a four-layer neural network is (p, q), where p is the number of nodes in the first hidden layer and q is the number of nodes in the second hidden layer.

(1) For a sequence, $GM(1, 1)$, $WPGM(1, 1)$, $\ and\ PGM(1, 1)$, were used, respectively, to obtain m simulation values and n forecast;

(2) The preset structure of a neural network (p, q) is $(1, 1)$.

(3) The simulated values of the three groups of GMs were used as the input modes. The lumped error E of the corresponding neural network structure samples was recorded (Wang, 2020).

(4) Increase p and q by 1, respectively. If $p = q = 11$, continue to Step (5); otherwise, go back to Step (3).

(5) Select the (p, q) corresponding to the minimum value of the abovementioned E, denoted as (p_0, q_0).

(6) Respectively according to $(p_0, 1)$, $(p_0, 2)$, , $\cdots$, $(p_0, 10)$, $(1, q_0)$, $(2, q_0)$, $\cdots$, $(10, q_0)$,

select the (p, q) corresponding to the minimum value of the abovementioned 20 E, denoted as (p_1, q_1).

(7) The three groups of n predictions in Step (1) are used as input modes, and the network simulation is carried out according to the structure of (p_1, q_1).

4.4. Prediction of CE

Suppose that it is necessary to predict the marine pollution at time t_A at a point $A(x_A, y_A, z_A)$ in the target sea area. Using the marine CE monitoring data of the target marine area at time t_1, t_2, $\cdots$, t_n, polyhedral models of the marine pollution situation at time t_1, t_2, $\cdots$, t_n, respectively, were constructed. The polyhedron model is used to interpolate the ocean CE data f_1, f_2, $\cdots$, f_n of point A. The above model is trained, the relevant parameters in the network are determined, $f_1, f_2, \cdots, f_n$ is taken as the input data, and the organic gray neural network model is the input, that is, the predicted value of the marine pollution at point A can be obtained.

Suppose that it is necessary to predict the Marine pollution status of the target sea area at time t_A. Taking a proper number of sampling points in the target area and ensuring the rationality of the spatial distribution of sampling points, the predictions of marine pollution at time t_A of each sampling point are solved one by one (Di Luccio et al., 2020). Using the sampling points to construct a three-dimensional (3D) polyhedral model, the status of CEs in the whole target sea area at time t_A can be described.

5. Experiment and analysis

5.1. Data acquisition

The data used in this study are from the National Marine Science Data Center of China (http://mds.nmdis.org.cn/), among which the indicators of marine monitoring data
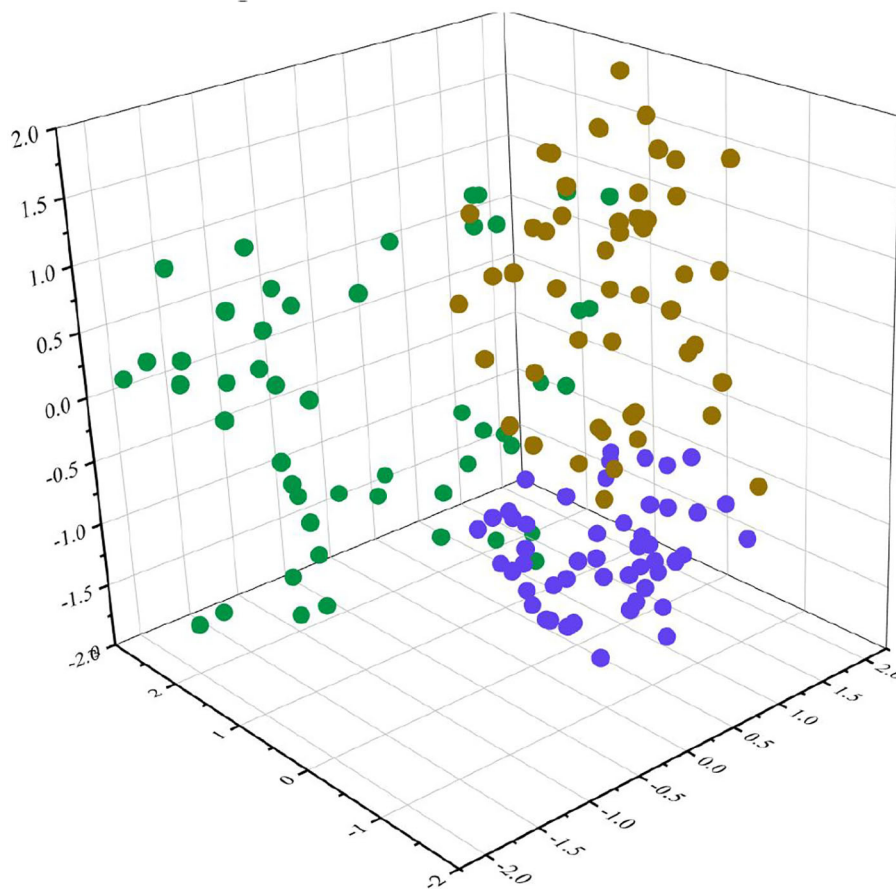
FIGURE 4
Fuzzy clustering results.

mainly include: seawater temperature, salinity, pH value, and dissolved oxygen concentration. The FCM algorithm was used to mine the marine environmental data by time and region.

5.2. Results and discussion

5.2.1. FCM clustering result analysis

Based on the fuzzy cluster analysis of 150 groups of marine environmental monitoring data from January to May, the influence of salinity on the clustering results can be ignored because there is no clear change in the salinity from January to May. The corresponding coefficient of the algorithm is set: clustering category $C = 3$, $m = 2$, and the maximum iteration number l is 20. The fuzzy clustering result is presented in Figure 4.

The coordinate axes in Figure 4 represent the values of seawater temperature, dissolved oxygen concentration, and P_H, respectively. As shown in Figure 4, the clustering effect of this algorithm for the marine monitoring data of the sea area is

TABLE 1 Comparison of clustering results.

	k-means	FCM
Accuracy/%	80.34	89.30
Execution time/s	0.190	0.034
Iterations	28	30

good, and it shows strong time characteristics. The same marine environmental data samples were selected, the FCM and k-means algorithms were used to cluster the marine monitoring sample data, and the execution efficiency and accuracy of the two algorithms were compared. The results are presented in Table 1.

The k-means and FCM algorithms were used to select and cluster the historical monitoring data of the marine environment into the corresponding sea areas. To a certain extent, there was a deviation. The reason may be related to a sudden change in the sea environment and the geographical location of the three sea areas. Some attributes of the nearby sea area
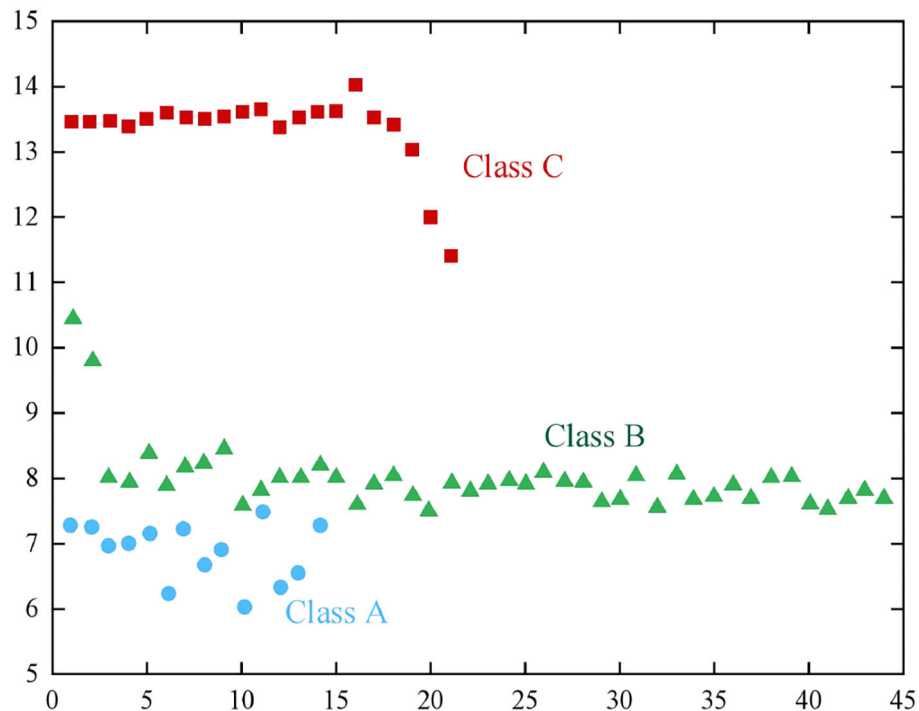
FIGURE 5
The clustering results of pH values of seawater.

were similar. However, when processing a large number of sea area monitoring data, the two algorithms can basically cluster the data from different sea areas in terms of accuracy. Compared with the *k*-means algorithm, the accuracy rate of the FCM algorithm for the data processing of different sea areas was 89.30%. Moreover, when a large number of marine environmental data are classified, the FCM algorithm can still achieve the desired effect, which shows that the FCM algorithm has more advantages in clustering marine environmental monitoring data.

In this study, the pH value of seawater in some historical monitoring data of a certain sea area was taken as the data sample, and some new historical monitoring data of pH value were added. The clustering results obtained by simulation are shown in Figure 5.

The coordinate axes in Figure 5 represent pH and temperature. From the simulation diagram, it is clear that the FCM algorithm divides the sample data into three categories according to the pH value, which corresponds to normal seawater (B), acidic polluted seawater (A), and alkaline polluted seawater (C). The pH of class A seawater is acidic, indicating that the content of carbon dioxide in seawater is high and that there is a great possibility of a jellyfish outbreak crisis in the sea area. Class C represents that there may be a large number of algal plants breeding. The growth of algae will consume a lot of carbon dioxide, which will lead
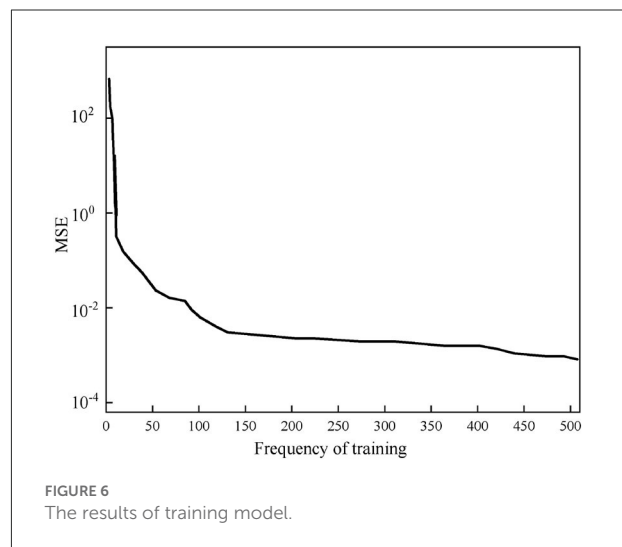

FIGURE 6
The results of training model.

to an increase in the pH value of seawater and make the seawater alkaline.

5.2.2. Analysis of CE prediction results

The training process of the enhanced neural network model using the GM-BP model is shown in Figure 6. It can be seen from
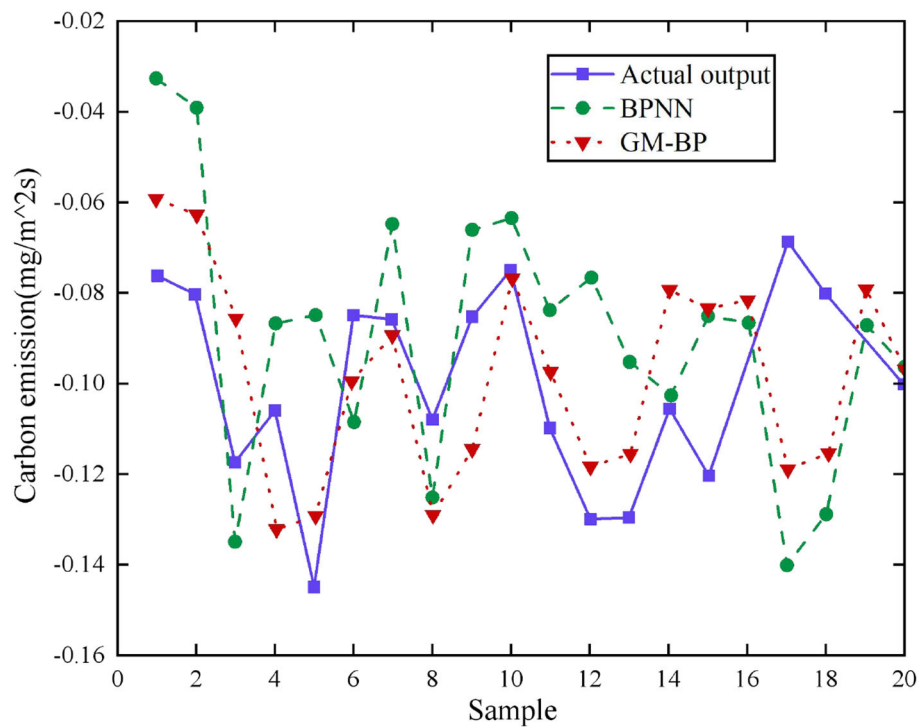
FIGURE 7
Comparison of the predicted and actual values of coastal carbon emission (CE). The prediction error of different algorithms is presented in Table 2.

TABLE 2 Comparison of carbon emission (CE) prediction errors.

	MSE	MAPE
BP	0.0264	19.532
GM-BP	0.0134	9.88

the figure that, after 510 times of training, the network stops training, and the network error reaches the goal of 0.001.

It can be seen from Figure 7 that the predicted value of CEs based on the GM-improved BP neural network algorithm is closer to the real value than that of the ordinary BP neural network algorithm.

The MSE and the mean absolute percentage error (MAPE) were used to analyze the CE errors of different models. From the test data, this algorithm can produce relatively accurate coastal CEs, while the BP neural network can only roughly predict the general trend. At the same time, from the perspective of prediction performance, the improved neural network algorithm based on GM has the best prediction effect. The maximum relative error is 0.4386, the root MSE of the prediction data is 0.0264, and the average relative error is 19.532%. It can be seen that the prediction error of the improved neural network is better than that of

the ordinary BP neural network, which indicates that the model has a higher degree of non-linear fit and is scientific and feasible.

6. Conclusion

Coastal cities are the most densely populated and economically populated areas, as well as the most concentrated areas of CEs. The growth of CEs in the global urban areas will continue to exert pressure on climate change. The FCM algorithm is used to mine the monitoring of data of the marine environment, and the GM-BP algorithm is used to predict the CEs in the coastal area. By analyzing the sea area data in the first half of a year, the clustering results showed that the parallel FCM algorithm has a good clustering effect in processing a certain scale of historical marine monitoring data. The GM-enhanced neural network has a higher degree of non-linear fit, and its prediction error is lower than that of the BP neural network. This study focused on the effective prediction of CEs in the coastal zone, which can provide a new method of measuring environmental governance for offshore environmental regulatory authorities.

The status of marine pollution changes gradually and continuously. The change in the results obtained by the

interpolation method occurs only at the boundary; that is, the results produced are abrupt at the boundary and homogeneous inside the boundary, i.e., unchanged. In the absence of sufficient sampling points, this assumption is not appropriate. In addition, future work will be devoted to studying the changes in carbon metabolism in coastal cities and mapping them onto the classified fine spatial data so as to analyze the spatial pattern characteristics of carbon metabolism in coastal cities more fully. New research (Yao et al., 2022) suggests that the use of social media data to clean up carbon pollution in the marine environment is also a good idea.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

This study was reviewed and approved by the School of Shipbuilding and Ocean Engineering, Jiangsu University of Science and Technology. All participants provided their written informed consent to participate in the study.

Author contributions

BW contributed to the writing of this paper, data collection, and data pre-processing.

Acknowledgments

The author thanks the reviewers whose comments and suggestions helped to improve the manuscript.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Chai, Q., Fu, S., Zheng, X., Zhao, X., and Xu, H. (2017). Research on total carbon emission control target and policy of key sectors and industries in China. *China Populat. Resour. Environ.* 27, 1–7. doi: 10.12062/cpre.20170608

Chen, J. (2014). *Research on Clustering Algorithm in Marine environment [D]*. Nanjing: Nanjing University of Aeronautics and Astronautics.

Di Luccio, D., Riccio, A., Galletti, A., Laccetti, G., Lapegna, M., Marcellino, L., et al. (2020). Coastal marine data crowdsourcing using the internet of floating things: improving the results of a water quality model. *IEEE Access.* 8, 101209–23. doi: 10.1109/ACCESS.2020.2996778

Dong, B. B., Li, L., Tang, H. S., Su, Y., and Tang, H. (2019). Environmental regulation, FDI and peak carbon emission prediction of energy consumption: a case study of five provinces in Northwest China. *Arid Land Geogr.* 42, 689–697. doi: 10.12118/j.issn.1000-6060.2019.03.26

Emamgholizadeh, S., Kashi, H., Marofpoor, I., and Zalaghi, E. (2014). Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *Int. J. Environ. Sci. Technol.* 11, 645–56. doi: 10.1007/s13762-013-0378-x

Erichson, N. B., Mathelin, L., Yao, Z., Brunton, S. L., Mahoney, M. W., Kutz, J. N., et al. (2020). Shallow neural networks for fluid flow reconstruction with limited sensors. *Proc. Royal Soc. A.* 476, 20200097. doi: 10.1098/rspa.2020.0097

Fu, C., Li, Y., Zeng, L., Zhang, H., Tu, C., Zhou, Q., et al. (2021). Stocks and losses of soil organic carbon from Chinese vegetated coastal habitats. *Glob. Chang. Biol.* 27, 202–14. doi: 10.1111/gcb.15348

Geman, O., Chiuchisan, I., and Covasa, M. (2016). "Data mining and knowledge discovery tools for human microbiome big data[C]," in *International Conference on Computers Communications and Control* (Piscataway, NJ: IEEE). p. 91–96. doi: 10.1109/ICCCC.2016.7496744

Jiang, G. Q., Xu, J., and Wei, J. (2018). A deep learning algorithm of neural network for the parameterization of typhoon—ocean feedback in typhoon forecast models. *Geophys. Res. Lett.* 45, 3706–16. doi: 10.1002/2018GL077004

Jin, H. (2021). Prediction of direct carbon emissions of Chinese provinces using artificial neural networks. *PLoS ONE.* 16, e0236685. doi: 10.1371/journal.pone.0236685

Krause-Jensen, D., and Duarte, C. M. (2016). Substantial role of macroalgae in marine carbon sequestration. *Nat Geosci.* 9, 737–42. doi: 10.1038/ngeo2790

Li, Y., Li, T., and Lu, S. (2021). Forecast of urban traffic carbon emission and analysis of influencing factors. *Energy Efficiency.* 14, 1–15. doi: 10.1007/s12053-021-10001-0

Li, Z. (2018). *Spatial and Temporal Patterns of Carbon Metabolism in Coastal Zone During Urbanization*. Xiamen: Xiamen University.

Liu, C., Shu, T., Chen, S., Wang, S., Lai, K. K., and Gan, L. (2016). An improved grey neural network model for predicting transportation disruptions. *Expert Syst. Appl.* 45, 331–40. doi: 10.1016/j.eswa.2015.09.052

Liu, L., Sun, S. Z., Yu, H., Yue, X., and Zhang, D. (2016). A modified Fuzzy C-Means (FCM) Clustering algorithm and its application on carbonate fluid identification. *J. Appl. Geophys.* 129, 28–35. doi: 10.1016/j.jappgeo.2016.03.027

Liu, Q., Guo, X., Yin, Z., Zhou, K., Roberts, E. G., and Dai, M.-H. (2018). Current status and prospect of carbon fluxes in marginal seas of China science. *China Earth Sci.* 48, 1422–1443. doi: 10.1007/s11430-017-9267-4

Lou, R., Lv, Z., Dang, S., Su, T., and Li, X. (2021). Application of machine learning in ocean data. *Multimedia Syst.* 1–10. doi: 10.1007/s00530-020-00733-x

Sun, M., Fu, Y., Lu, L., and Jiang, X. (2018). Application of deep learning in ocean big data mining. *Sci. Technol. Rev.* 36, 83–90. doi: 10.3981/j.issn.1000-7857.2018.17.010

Tanhua, T., Pouliquen, S., Hausman, J., O'brien, K., Bricher, P., De Bruin, T., et al. (2019). Ocean FAIR data services. *Front. Mar. Sci.* 6, 440. doi: 10.3389/fmars.2019.00440

Vondolia, G. K., Chen, W., Armstrong, C. W., and Norling, M. D. (2020). Bioeconomic modelling of coastal cod and kelp forest interactions: co-benefits of habitat services, fisheries and carbon sinks. *Environ. Resource Econ.* 75, 25–48. doi: 10.1007/s10640-019-00387-y

Wang, X. (2020). A neural network algorithm based assessment for marine ecological environment. *J. Coastal Res.* 107, 145–148. doi: 10.2112/JCR-SI107-037.1

Yang, L., Xia, H., Zhang, X., and Yuan, S. (2018). What matters for carbon emissions in regional sectors? A China study of extended STIRPAT model . *J. Cleaner Prod.* 180, 595–602. doi: 10.1016/j.jclepro.2018.01.116

Yao, Q., Li, R. Y. M., and Song, L. (2022). Carbon neutrality vs. neutralité carbone: a comparative study on French and English users' perceptions and social capital on Twitter. *Front. Environ. Sci.* 1632. doi: 10.3389/fenvs.2022.969039

Yao, X., Mou, J., Chen, P., and Zhang, X. (2017). Study on ship emission inventory in Yangtze estuary based on AIS data. *J. Safety Environ.*17, 1510–1514. doi: 10.13637/j.issn.1009-6094.2017.04.058

Zhang, N., Ma, Y., and Zhang, Q. (2018). Prediction of sea ice evolution in Liaodong Bay based on a back-propagation neural network model.

Cold Regions Sci. Technol. 145, 65–75. doi: 10.1016/j.coldregions.2017.10.002

Zhang, S. (2018). *Research on Data Mining Algorithm based on Marine Environment [D]*. Qingdao: Qingdao University of Science and Technology.

Zheng, H., Wang, S., and Liu, H. (2014). Analysis of ship carbon Emission based on AIS data [J]. *China Water Transport (Second Half)*. 14, 157–158.

Zhou, K., and Yang, S. (2016). *Exploring the Uniform Effect of FCM Clustering[M]*. Amsterdam: Elsevier Science Publishers B V. doi: 10.1016/j.knosys.2016.01.001

Zhuang, X., Yu, X. L., Zhao, Z. M., Zhou, D., Zhang, W. J., Li, L., et al. (2021). *Application of Laser Ranging and Grey Neural Network (GNN) for the Structure Optimization of Radio Frequency Identification (RFID) Tags*. Philadelphia, PA: Lasers in Engineering (Old City Publishing). p. 48.

Zhuang, Y., Wang, Y., Shao, J., Chen, L., Lu, W., Sun, J., et al. (2016). D-Ocean: an unstructured data management system for data ocean environment. *Front. Comput. Sci.* 10, 353–69. doi: 10.1007/s11704-015-5045-6