Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 6, No. 1, 2025

Analysis and Prediction of Factors Affecting Marine Chlorophyll Based on Gray Correlation and Random Forest Modeling

Xiaoyi Liu*

College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin 300457, China
* **Corresponding author**: Xiaoyi Liu (Email: 15028101055@163.com)

Abstract: In this paper, the relationship between marine chlorophyll content and environmental factors is deeply investigated by utilizing ocean observation big data. Through data preprocessing, gray correlation analysis, Spearman correlation analysis and random forest modeling, this study identified the key factors affecting the chlorophyll content. The results showed that carbon dioxide levels, seawater temperature, salinity, pH and dissolved oxygen were the key environmental drivers, with carbon dioxide and temperature being significantly negatively correlated with chlorophyll content. These findings emphasize the potential role of marine chlorophyll and associated aquaculture in carbon sequestration. This work provides valuable insights into marine resource management and ecological conservation, providing a scientific basis for further research and policy development.

Keywords: Marine Chlorophyll, Grey Correlation Analysis, Spearman Correlation Coefficient, Random Forest Model.

1. Introduction

The marine environment is a vital component of the Earth's ecosystem, and marine chlorophyll content serves as a crucial indicator of marine primary productivity and ecosystem health. Understanding the factors that influence marine chlorophyll content is essential for marine resource management and ecological conservation. With the advancement of technology, the integration of satellite remote sensing and machine learning has become a significant direction in marine monitoring, enhancing our ability to predict and analyze marine environmental information.

Previous studies have made substantial contributions to this field. Zhang et al. [1] investigated the impact of small-scale eddies on chlorophyll distribution from a marine dynamics perspective, revealing how atmospheric circulation and oceanic heat energy influence marine ecological dynamics. AlHossainy et al. [2] utilized satellite imagery and machine learning algorithms to infer bathymetry, providing a new approach for marine monitoring in the absence of ground-measurement data. Wen et al. [3] constructed a time-series prediction network combining long-and short-term memory networks with convolutional neural networks, improving the prediction of long-term trends in marine environmental information. Li et al. [4] proposed a multi-agent deep reinforcement learning model that significantly enhanced the prediction performance of marine environmental monitoring. Ke et al. [5] demonstrated the application of convolutional neural networks in oceanic carbon-sink monitoring, enabling near-real-time monitoring of oceanic carbon sinks. Jin et al. [6] achieved a breakthrough in multi-step prediction of chlorophyll concentration using a multi-attention collaborative network, enhancing prediction accuracy. Yuan et al. [7] focused on the impact of marine heatwaves on chlorophyll concentration, analyzing their vertical structures and revealing the relationship between marine heatwaves and chlorophyll concentration changes [8].

While these studies have provided valuable insights, there remains a need for a more comprehensive and integrated analysis framework that can effectively identify and predict the key factors affecting marine chlorophyll content. This paper aims to fill this gap by employing a combination of gray correlation analysis, Spearman correlation analysis, and random forest modeling to investigate the relationship between marine chlorophyll content and environmental factors. The study utilizes ocean observation big data and focuses on the Bohai Sea in China, providing a detailed analysis of the environmental drivers of marine chlorophyll content. Our work not only identifies the key factors influencing chlorophyll content but also highlights the potential role of marine chlorophyll and associated aquaculture in carbon sequestration, offering a scientific basis for further research and policy development in marine resource management and ecological conservation.

2. Fundamentals of marine chlorophyll impact factors

2.1. Data processing and feature engineering

2.1.1. Source of Data

In order to study the changes in the content of various factors in the ocean big data and its influence and correlation with the chlorophyll and CO_2 content that need to be studied in this paper, and to carry out the study of chlorophyll and CO_2 content in the next three years, this study obtained the data from the Ocean Science Data Center of the Chinese Academy of Sciences (http://msdc.qdio.ac.cn), in which the data used the CO_2 content, chlorophyll content, pH, salinity, dissolved oxygen, nitrate, phosphate, silicate, alkalinity, and content of the Pacific Ocean from 1973 to 2021 as the main objects of the study, and used the latitude and longitude to locate the specific position, and limited the experimental data to the Bohai Sea in China, as shown in Table 1.

Table 1 Main data of the dataset

Data Set Name	Data Significance	Data Set Name	Data Significance
G2chla	Chlorophyll concentration	G2nitrate	Nitrate
G2phts25p0	Acid-base at 25°C 0d	G2phosphate	Phosphate
G2latitude	Latitude	G2silicate	Silicate
G2longitude	Longitude	G2tco2	Total Carbon Dioxide
G2bottomdepth	Bottom Depth	G2talk	Alkalinity
G2maxsampdepth	Maximum sampling depth	G2fco2	Carbon dioxide content

2.1.2. Pre-processing of marine big data

The initial data were merged from CSV files and CN files. In this paper, latitude and longitude are extracted as indicators, and the data of carbon dioxide content and chlorophyll content, pH value, salinity, and oxygen content of the sea areas that are in the same latitude and longitude and in the same time period are merged in one dataset. In the processing of temporal data, this paper combined the three time columns of year, month and day of the original data and created DATE-type temporal variables. At the same time, this paper found that there are a large number of -9999 in the data, this data has not been detected except the real data, this paper will be used to replace the missing data with zero.

2.1.3. Normalization of data

This method scales the data so that it falls into a small specific interval. The formula is.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where: X is the value of the original data point. X_{min} is the minimum value in the data set. X_{max} is the maximum value in the data set, X_{norm} is the normalized value.

The above treatment removes the discrepancies between the quantities, making the conclusions more scientifically valid.

2.1.4. Quartile outlier treatment

(1) IQR Calculation

IQR is the difference between the third quartile and the first quartile, which indicates the distribution range in the middle of the data.

$$IQR = Q3 - Q1 \quad (2)$$

(2) Definition of outliers

Outliers are usually defined as data points that lie outside the following ranges.

$$\text{Lower Bound} = Q1 - k \times IQR \quad (3)$$

$$\text{Upper Bound} = Q3 + k \times IQR \quad (4)$$

where Lower Bound means the lower bound of the outlier. Upper Bound is the upper bound of the outlier. Upper Bound is Upper Bound of the outliers. where is a constant, a multiplier based on the actual data, used to adjust the "strictness" of the outliers. This is used to identify extreme outliers.

2.1.5. Gray correlation feature dimensionality reduction

(1) Gray correlation modeling

First of all, data preprocessing is carried out. The dataset D of the original ocean contains mfeatures, and each feature i corresponds to a sequence of features $X_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$, which needs to be converted to the normalized dataset D_{norm} in this paper. Second, determine the reference series. The leaf green content of the ocean is selected as the reference sequence to set the reference series $R = \{r_1, r_2, \ldots, r_n\}$. Then, calculate the correlation degree. Construct the absolute difference sequence: for each feature sequence $X_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$, calculate its correlation with the reference series R. In order to quantify the similarity between the feature sequence and the reference sequence, first calculate the absolute difference between them.

$$\Delta_i(k) = |x_{ik} - r_k|,\ i = 1,2,\ldots,m;\ k = 1,2,\ldots,n \quad (5)$$

Then determine the coefficient of resolution to be sought: the coefficient of resolution reflects the relative trend and strength of change between the series. For each data point, the resolution coefficient is calculated by the following formula.

$$\xi_i(k) = \frac{\Delta_{min}^{min} + \rho\Delta_{max}^{max}}{\Delta_i(k) + \rho\Delta_{max}^{max}} \quad (6)$$

where $\Delta_{min}^{min} = \min_i \min_k \Delta_i(k)$ is the smallest of all the differences. $\Delta_{max}^{max} = \max_i \max_k \Delta_i(k)$ is the maximum value of all the differences. ρ is the differentiation factor (usually 0.5), which regulates the relative importance.

Calculate the correlation: For each feature, calculate its correlation with the reference series.

$$\Gamma_i = \frac{1}{n}\sum_{k=1}^{n} \xi_i(k) \quad (7)$$

The closer the correlation Γ_i is to 1, the more closely the feature is associated with the reference series.

(2) Feature extraction based on correlation

The results obtained in this paper are shown in the results of Table 2 below. For this reason, this paper's get as above bolded color data as this paper's grey correlation after the initial feature extraction data. This paper also initially found that there are many factors with high degree of correlation with chlorophyll content, of which CO_2 has a very high degree of correlation, is one of the highest degree of correlation, the table name in the gray correlation model of CO_2 content and chlorophyll content shows a strong correlation between the chlorophyll and its corresponding marine algae kelp, etc. may have a strong correlation with the role of CO_2 carbon sequestration.

Table 2 Gray correlation of each feature of the dataset

Dataset name	Relevance value	Dataset name	Relevance value
G2donf	0.023	G2temperature	0.654
G2tdn	0.011	G2salinity	0.478
G2pressure	0.594	G2oxygen	0.380
G2chla	1.000	G2nitrate	0.721
G2phts25p0	0.672	G2phosphate	0.465
G2latitude	0.678	G2silicate	0.489
G2longitude	0.578	G2tco2	0.811
G2bottomdepth	0.398	G2talk	0.015
G2maxsampdepth	0.097	G2fco2	0.322

Note: Other unknown and unfocused data have been omitted. Which has been omitted part of the data of low relevance of the data, this data features higher dimensions and difficult to expand.

Upon completion of the initial data preprocessing, this study transitions to the next phase, which involves conducting in-depth data analysis and constructing statistical models for predictive purposes. The dataset employed in this phase consists of the preprocessed data, as detailed in Table 3 below.

Table 3 Data after initial feature extraction for the dataset

Data set name	Data Significance	Data set name	Data Significance
G2temperature	temperature	G2salinity	Salinity
G2pressure	Pressure	G2oxygen	Dissolved Oxygen
G2chla	Chlorophyll concentration	G2nitrate	Nitrate
G2phts25p0	Acid-base at 25°C 0 d	G2phosphate	Phosphate
G2latitude	Latitude	G2silicate	Silicate
G2longitude	Longitude	G2tco2	Total Carbon Dioxide
G2bottomdepth	Depth	G2talk	Alkalinity

2.2. A study of the factors influencing marine chlorophyll

2.2.1. Continuous type and normality test of data

(1) Continuous type differential test

For a given data sequence s, the difference between its consecutive data points is defined as

$$d_i = s_i - s_{i-1} \tag{8}$$

where d_i denotes the difference between the ith observation in sequence s and the previous observation s_{i-1}. This step is designed to capture the instantaneous change between any two points in the sequence.

After calculating the resulting difference series d, the next critical step is to check the uniqueness of these difference values. Theoretically, if all d_i values are equal, it indicates that the sequence varies consistently from one observation interval to the next, thus proving the continuity of the data. This assumption can be verified by counting the number of distinct values in the difference sequence if then it indicates that the data is continuous, and if $0.5 < \text{nunique}(d) < 3$ does not fall into this range, then the data is discontinuous.

(3) Normality JB test

In order to assess whether the marine parameter data follow a normal distribution, the Jarque-Bera (JB) test was used in this study. The JB test is based on the degree of matching between the skewness and kurtosis of the samples to determine whether the data conform to a normal distribution or not. The Jarque-Bera (JB) test is particularly well-suited for analyzing large datasets. The statistics of JB test is calculated by the following formula:

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right) \tag{9}$$

where n represents the number of samples, S is the sample skewness, and k is the sample kurtosis. According to the central limit theorem, when the sample size is sufficiently large, the JB statistic approximately obeys a chi-square distribution (with a degree of freedom of 2).

$$S = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^3}{\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^{3/2}} \tag{10}$$

where: n is the number of samples and S is the sample skewness indicating the symmetry of the distribution.

$$K = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^4}{\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^{2}} \tag{11}$$

where: k is the sample kurtosis, which indicates how sharp or flat the distribution is, and $\bar{X}$ is the sample mean.

The larger the JB statistic, the stronger the evidence for rejecting the hypothesis that the data are from a normal distribution. In the JB test, the null hypothesis is assumed to be that the data are from a normal distribution and the alternative hypothesis is that the data are not from a normal distribution. This paper rejects the null hypothesis if the JB statistic corresponds to a p-value that is less than a predetermined level of significance.

Continuum and normality analysis

The test for continuity and normality of the data is shown in Table 4:

Table 4 Data set test results

Data name	nunique(d)	Continuous	Statistic	p-value
G2latitude	572.57	No	0.92151	0
G2longitude	452.34	No	0.75523	0
G2chla	45.237	No	0.08389	0
G2tco2	76.732	No	0.91951	0
G2temperature	78.691	No	0.81528	0
G2salinity	478.79	No	0.64747	0
G2oxygen	38.377	No	0.98612	0
G2pressure	876.73	No	0.80922	0
G2depth	576.25	No	0.81054	0
G2phts25p0	145.24	No	0.93666	0
G2nitrate	75.421	No	0.86858	0
G2nitrite	6445.2	No	0.21566	0
G2phosphate	437.87	No	0.89063	0
G2silicate	150.75	No	0.90477	0

Through the above table, this paper found that the data has discontinuous nature and does not conform to the normal distribution, for this this paper used in order to study its correlation can be used to conform to the discontinuous non-normal distribution of Spearman's correlation analysis model.

2.3. Spearman rank correlation coefficient

The Spearman rank correlation coefficient, usually denoted as ρ (rho), is a nonparametric correlation coefficient that measures the monotonic correlation between the ranks of two variables and is applicable to both ordinal and continuous scale data.

The calculation of Spearman's correlation coefficient is based on the ranking of the data. Given two variables X and Y with corresponding observations x_i and y_i, this paper first converts each observation to its rank r_{x_i} and r_{y_i} in the respective variables. The Spearman's correlation coefficient is computed as:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{12}$$

where $d_i = r_{x_i} - r_{y_i}$ is the difference between the two rankings and n is the total number of data points.

In cases where there are duplicate rankings (i. e. , "rank ties"), the calculation needs to be adjusted to account for the effect of ties. Specifically, for each group of tied rankings, their average rankings need to be used instead.

Computerized calculations yielded the following results of the correlation experiment, as shown in Figure 1 below:

The correlations between chlorophyll concentration (G2chla) and marine environmental parameters were as follows: both total carbon dioxide content (G2tco2) and temperature (G2temperature) showed a strong negative correlation with chlorophyll (about -0.51), which may be related to the effects of ocean acidification and temperature on productivity. Dissolved oxygen content (G2oxygen) was moderately positively correlated with chlorophyll (~0.37), reflecting oxygen release from photosynthesis. Salinity (G2salinity) was weakly correlated with latitude (G2latitude), and longitude (G2longitude) was weakly positively correlated (~0.28). Nitrate (G2nitrate), phosphate (G2phosphate) and silicate (G2silicate) contents were weakly correlated with chlorophyll and may be influenced by a variety of factors. Chlorophyll is a key indicator of marine photosynthesis, and its concentration is affected by a variety of factors, and has a strong negative correlation with carbon dioxide content, suggesting that it may be important for carbon dioxide sequestration, as shown in Figure 1.

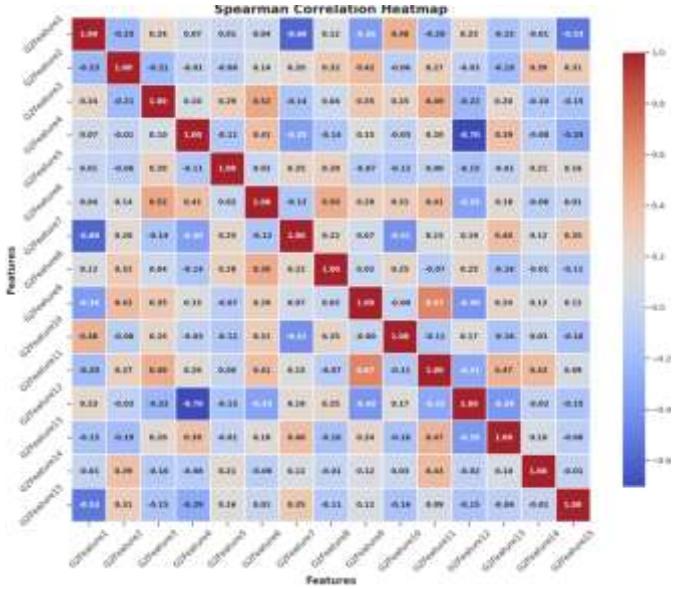

Figure 1 Spearman's Correlation Heat Map Matrix

3. Results

3.1. A study of the importance of random forest features

In this study, this paper delves into the mathematical principles of the random forest regression model and its application to predicting the chlorophyll content of the target variable. The random forest model consists of multiple decision trees, and the predictive stability and accuracy of a single model is improved by integrated learning methods.

3.1.1. Composition of Random Forest

A random forest consists of N decision trees, each independently modeling the same prediction problem. The predictive output of a random forest is the average of the predictions of these N trees. For regression problems, this averaging method can be mathematically expressed as

$$y(x) = \frac{1}{N}\sum_{i=1}^{N} T_i(x;\Theta_i) \tag{13}$$

where $T_i(\mathbf{x};\Theta_i)$ is the output of the i-th tree based on the parameter set Θ_i and x is the feature vector.

Figure 2 below illustrates the structure of the random forest model:

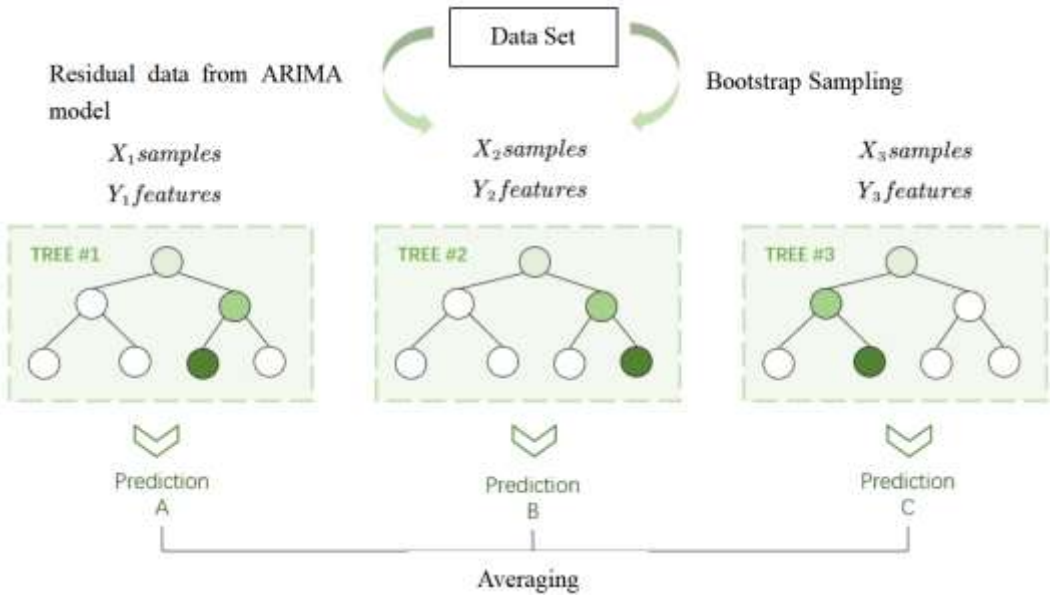

Figure 2 Model Structure of Random Forests

3.1.2. Decision Tree Partitioning Principles

Each tree is constructed by recursively splitting the training data to form a tree structure. At each segmentation point, the best segmentation features and segmentation thresholds are chosen to maximize the objective function, which is usually a reduction in node impurity. For regression problems, the objective function typically used is a reduction in the mean squared error of.

$$\Delta \mathrm{MSE} = \mathrm{MSE}_{\mathrm{parent}} - \left(\frac{n_{\mathrm{left}}}{n_{\mathrm{parent}}}\mathrm{MSE}_{\mathrm{left}} + \frac{n_{\mathrm{right}}}{n_{\mathrm{parent}}}\mathrm{MSE}_{\mathrm{right}}\right) \quad (14)$$

where, $\mathrm{MSE}_{\mathrm{parent}}$ is the MSE of the parent node, $\mathrm{MSE}_{\mathrm{left}}$ and $\mathrm{MSE}_{\mathrm{right}}$ are the MSE of the two child nodes. n_{left} and n_{right} are the number of samples of the left and right child nodes and n_{parent} is the number of samples of the parent node.

3.1.3. Random characteristic sampling

At each attempt to segment a node, the algorithm randomly draws a feature subset F' of size m from F. The set size is m , and the random subset F' can be represented by the random sampling function.

$$F' = \mathrm{RS}(F, m) \quad (15)$$

The potential segmentation points for each feature are then evaluated in this randomly selected subset of features F', RS denotes random also sampling, to select the features and segmentation thresholds that maximize the segmentation effect. For regression problems, the usual goal is to minimize the total mean squared error of the left and right child nodes.

3.1.4. Significance of features

For regression tasks, the most commonly used impurity measure is the mean squared error (MSE). At each node split, this paper can compute the amount of impurity reduction due to that feature with the following formula:

$$\Delta\,\mathrm{Ip}(t, f) = \mathrm{Ip}(t) - \left(\frac{n_{\mathrm{left}}}{n_t}\mathrm{Ip}\,(t_{\mathrm{left}}) + \frac{n_{\mathrm{right}}}{n_t}\mathrm{Ip}\,(t_{\mathrm{right}})\right) \quad (16)$$

where: t is the current node. f is the feature used to segment t. Ip(t) is the impurity of node t. t_{left} and t_{right} are the left and right child nodes after segmentation. n_{left} and n_{right} are the number of samples of the left and right child nodes. n_t is the number of samples of node t.

Cumulative importance of features: The total importance of a feature f is the cumulative reduction in impurity caused by that feature at all split points in all trees. The impurity reduction of each feature is accumulated in each tree of the random forest and eventually averaged over all trees. The average importance of a feature f is given by the following equation:

$$\mathrm{I}(f) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t\in T_i}\Delta\mathrm{Ip}\,(t, f) \quad (17)$$

where: N is the total number of trees. T_i is the set of all nodes in the ith tree that are partitioned using feature f. I denotes the importance.

Normalized Feature Importance: In order to make the comparison of feature importance more interpretable, the importance scores of all features are usually normalized to ensure that the sum of all the importance scores is one. The normalized importance is calculated as

$$\mathrm{NI}(f) = \frac{\mathrm{I}\,(f)}{\sum_{f'\in F}\mathrm{I}\,(f')} \quad (18)$$

where F is the set of all features. NI(f) denotes the normalized feature importance.

The following result is obtained as a visualization of feature importance as Figure 3 shown below:

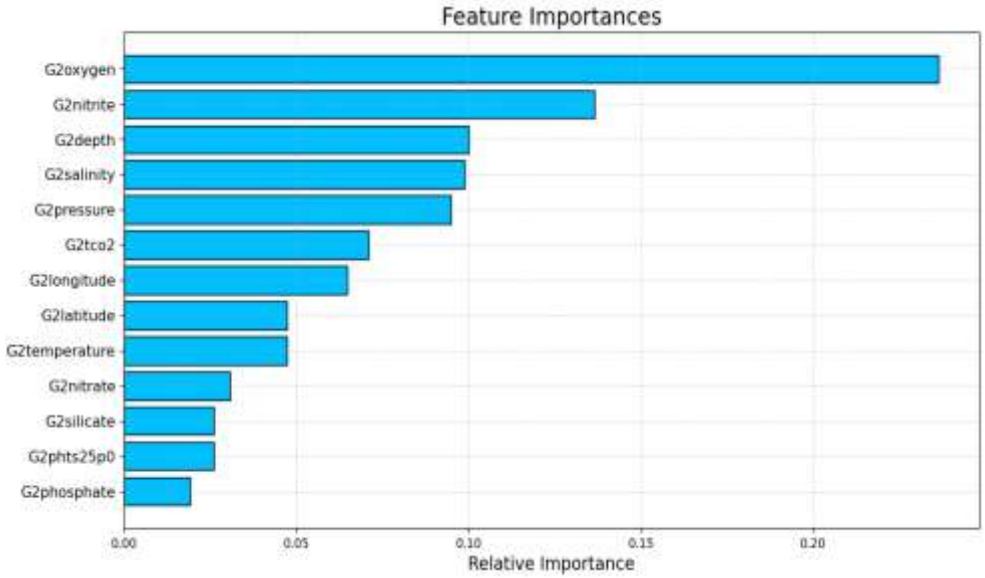

Figure 3 Importance of Random Forest Data Features

The decision variable for this paper was selected as chlorophyll content (G2chla, and other factors as features), and this paper's analysis of the above resulted in a chart that demonstrates the relative importance of different features, which is usually used to illustrate the magnitude of each feature's contribution in a predictive model. From the graphs this paper can be seen:

In the random forest model, dissolved oxygen content (G2oxygen) is the most important factor affecting chlorophyll content, nitrate content (G2nitrite) and depth (G2depth) also have high importance. Salinity (G2salinity), pressure (G2pressure) and total carbon dioxide content (G2tco2) were of moderate importance, while longitude (G2longitude), latitude (G2latitude), temperature (G2temperature), silicate content (G2silicate), acid-base at 25°C (G2phts25p0) and phosphate content (G2phosphate) are less important.

Based on this graph, this paper can infer that dissolved oxygen content, nitrate content, and depth are the most important factors affecting the prediction results in random forest models that make predictions of ocean chemistry parameters. This information is particularly valuable for understanding how specific environmental variables affect marine ecosystems and may also provide important insights for further research and marine resource management strategy development. In particular, chlorophyll content was analyzed in relation to the various chemicals in the oceans.

3.2. Performance evaluation of regression tasks

In order to validate the reasonableness of the results of the feature importance described above in this paper, this paper performs a performance evaluation of the model for the regression task. Performance evaluation is a key step in the development of a machine learning model that allows this paper to quantify the predictive power of the model for new data. In random forest regression models, performance evaluation typically involves a variety of statistical metrics to provide a comprehensive understanding of the model's accuracy, consistency, and prediction error.

3.2.1. Results for Random Forests

In this paper, we get the results of random forest regression evaluation as shown in Table 5:

Table 5 Randomized forest regression evaluation results

Evaluation indicator	MSE	MAE	R^2
Value	0.0041	0.0421	0.6834

Mean Squared Error (MSE - Mean Squared Error): the value of MSE is 0.0041, the value of MSE is relatively low which indicates that the model gives good accuracy in predicting the G2chla values. Mean Absolute Error (MAE - Mean Absolute Error): the MAE value is 0.0421, the MAE value indicates that the model gives relatively accurate predictions. Coefficient of Determination (R^2 - R-squared): The value of R^2 is 0.6834, the value of R^2 shows that the model has a good fit to the data, but there is still room for improvement. Through this test result, this paper preliminarily judged that the results of regression fitting of different characteristics on the effect of chlorophyll content at the time of regression fitting in this paper have a high degree of accuracy and scientific validity.

4. Conclusions

This study has provided significant insights into the factors influencing marine chlorophyll content through the application of gray correlation analysis, Spearman's rank correlation coefficient, and random forest modeling. The research has confirmed that carbon dioxide levels, seawater temperature, salinity, pH, and dissolved oxygen are key environmental drivers affecting marine chlorophyll content, with carbon dioxide and temperature showing substantial negative correlations with chlorophyll levels. These findings highlight the crucial role of marine chlorophyll and associated aquaculture in carbon sequestration, offering valuable insights for marine resource management and ecological conservation.

The integrated analysis framework developed here can

serve as a template for future studies in other marine regions. However, there is still room for improvement. For future research, it would be beneficial to expand the dataset to include more diverse oceanic regions and a wider range of environmental variables. Additionally, further refinement of the random forest model could enhance its predictive capabilities. This could involve incorporating more complex interactions between variables and leveraging advanced machine learning techniques to better capture the nuanced relationships within marine ecosystems. Such advancements would further our understanding of marine chlorophyll regulation and support more effective marine environmental management strategies.

References

[1] Zhang Z, Qiu B, Klein P, et al. The influence of geostrophic strain on oceanic ageostrophic motion and surface chlorophyll[J].Nature Communications, 2019, 10(1):1-11.

[2] AlHossainy H R, Saber A, Ghany E A R, et al. Inferring Bathymetry from Sentinel-2 Satellite Images Using Machine Learning Algorithms Based on Chlorophyll Concentration Data in the Absence of Ground Measurement[J]. Arabian Journal for Science and Engineering, 2025, (prepublish): 1-18.

[3] Wen J, Yang J, Jiang B, et al. Big data driven marine environment information forecasting: a time series prediction network[J]. IEEE Transactions on Fuzzy Systems, 2020, 29(1): 4-18.

[4] Li Z, Wang J, Cao D, et al. Investigating Neural Activation Effects on Deep Belief Echo-State Networks for Prediction Toward Smart Ocean Environment Monitoring[J]. Arabian Journal for Science and Engineering, 2021, 46(4): 3913-3923.

[5] Ke P, Gui X, Cao W, et al. Near-real-time monitoring of global ocean carbon sink based on CNN[R]. Copernicus Meetings, 2024.

[6] Jin Y, Zhang F, Wang X, et al. Multi-Step Forecasting of Chlorophyll Concentration with Multi-Attention Collaborative Network[J].Journal of Marine Science and Engineering, 2025, 13(1): 151-151.

[7] Yuan T, Zhang J, YangS, et al. Vertical Structures of Marine Heatwaves in the South China Sea: Characteristics, Drivers and Impacts on Chlorophyll Concentration[J]. Journal of Geophysical Research: Oceans, 2024, 129(12): e2024JC021091-e2024JC021091.

[8] Francisco G, Isabel F, Lidia Y, et al. Two decades of satellite surface chlorophyll a concentration (1998–2019) in the Spanish Mediterranean marine waters (Western Mediterranean Sea): Trends, phenology and eutrophication assessment[J]. Remote Sensing Applications: Society and Environment, 2022, 28.