Proceedings of the 5th International Conference on Materials Chemistry and Environmental Engineering
DOI: 10.54254/2755-2721/2025.29848

Selection and Analysis of Prediction Elements for Arctic Sea Ice Extent Prediction Model

Haoxu Yin[1†], Qiyang Chen[2*†], Yanfei Yang[3†], Zujian Liu[4†]

[1]Guangdong Experimental School, Guangzhou, China
[2]College of Hydrology and Water Resource, Hohai University, Nanjing, China
[3]College of Remote Sensing and Surveying Engineering, Nanjing University of Information Science and Technology, Nanjing, China
[4]School of Earth Sciences and Engineering, Hohai University, Nanjing, China
**Corresponding Author. Email: 1851780484@qq.com*
[†]These authors contributed equally to this work and should be considered as co-first author.

Abstract. The phenomenon of global warming has resulted in significant alterations to the conditions of the Arctic Ocean, giving rise to a heightened degree of unpredictability in the behaviour of sea ice cover. The present study employed a comparative and analytical approach to examine the monthly and annual trends of Arctic sea ice extent, as well as a range of other environmental variables, including carbon dioxide(CO2),global average temperature (GAT), precipitation(PP),Sea ice extent (SIE),two-metre mean temperature (T2M). Furthermore, the forecast albedo was analysed. The CO2 concentration was employed as the variable for analysis, and linear models, moving averages, Granger causality tests, and collinearity tests were used to compare and analyse the relationship between various variables and sea ice extent, as well as the multicollinearity among various variables. The results demonstrated tha GAT,PP,and CO2 can be regarded as the primary drivers of SIE alterations. Meanwhile, SIE can be identified as the instigator of changes in T2M, GAT, and PP, yet not as a causal factor in CO2 fluctuations. The issue of collinearity between PP, GAT, CO2, and T2M is relatively minor and can be employed as a predictive element in the construction of Arctic sea ice prediction models.

Keywords: Arctic sea ice extent, Influencing factors, Granger causality test, Collineearity test.

1. Introduction

According to AMAP 2021, the extent of Arctic sea ice has exhibited a notable decline [1], with the study's findings indicating the lowest end-of-winter sea ice volume during the past decade [2]. Though sea ice in Arctic covers only approximately 4.5% of the ocean area, Meier, W. N. [3], it is recognized as an important component of the global climate and temperature system. This is due to its high albedo, which allows the same area of sea ice to reflect more solar radiation than seawater. Furthermore, sea ice has an important impact on the atmospheric heat balance.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.29848

All the results above caused a phenomenon that the changes in sea ice extent in the Arctic have attracted considerable interest from researchers around the globe. The study of changes in sea ice extent and the prediction of future changes using models represents a prominent area of research. Previous studies have discussed a number of potential causes affecting the extent of Arctic sea ice, including sea surface temperature, long wave radiation, Arctic temperature, surface albedo, sea level pressure, and the drift of the sea ice itself. Deser et al. [4] found that sea ice in some Arctic waters is closely related to the change of surface temperature. Tian et al. [5] studied the Arctic Effects of cyclones on sea ice. Zhang et al. [6] mainly analyzed the characteristic changes of sea ice type and thickness.

In previous studies, researchers have constructed various models for sea ice extent based on predictions of sea surface temperature, two-metre mean temperature and carbon dioxide(CO_2) concentration. For example, based on the characteristics of high albedo of Arctic sea ice, Ren Shuai et al. constructed a statistical model of sea ice albedo-atmospheric circulation. From the vantage point of deterministic and probabilistic prediction, Chang Le assessed the forecasting of Arctic temperature and sea ice on a seasonal-interannual scale by means of the high-resolution prediction system DePreSys3. These models can be used to predict future Arctic sea ice extent. For example, Kim et al. [7] selected Kara Sea and Barents Sea as research areas, compared the prediction effect of deep neural network and multiple regression algorithm on sea ice density, and pointed out that the former is obviously superior potential.

The multivariate LSTM model constructed by Yang et al. demonstrably enhances the precision of Arctic summer sea ice extent predictions, offering a valuable reference point for future Arctic sea ice forecasting. However, due to the limited number of selected variables, this paper suggests that it would be beneficial to screen additional prediction elements that are closely related to sea ice extent change and conduct a more in-depth analysis of the change law in subsequent research. This would help to further reduce the uncertainty of the model prediction results and improve the prediction accuracy of sea ice extent. In this paper, 11 research variables, including PP and GAT, were selected for analysis. The model was constructed from a mathematical statistical perspective, and the linear regression model was employed to identify variables with a strong correlation and to elucidate the causal relationship between these variables and the Arctic sea ice extent. The VAR model and causality were employed to test the aforementioned variables, and the influencing factors of the Arctic sea ice extent change from 1980 to 2024 were subjected to further analysis. The feasibility of the selected variables for a sea ice prediction model was preliminarily verified by a collinearity test [8,9].

2. Data and methods

2.1. Data

The data on sea ice extent utilised in this article is derived from the Sea Ice Index, Version 3, as provided by the National Snow and Ice Data Center (NSIDC) in the United States. The temporal scope of the study encompasses the period between 26 October 1978 and the present. The data is available at a daily temporal resolution and covers an area between 30.98° N and 90° N, with a spatial resolution of 25km × 25km. The spatial reference frame employed for this data set is the NSIDC sea ice polar stereo map, North EPSG: 3411. The data were processed on a monthly basis in order to obtain the relevant sea ice extent data.

The data on carbon dioxide concentration is derived from the Global Monthly Mean Mole Fractions provided by the World Data Centre for Greenhouse Gases (WDCGG), which includes

DOI: 10.54254/2755-2721/2025.29848

CO_2, CH_4, and N_2O. The temporal resolution of this data set is monthly, with the selected time range spanning from 1984.1 to 2022.12.

The elevation data for the Arctic region is sourced from the A Big Earth Data Platform for Three Poles. Additionally, observational data from the ERA5 reanalysis dataset of the European Centre for Medium Range Weather Forecasts (ECMWF) were employed in this study. These data included snowmelt, surface net solar radiation, mean surface net longwave radiation flu, sea surface temperature, total cloud cover, 10m u-component of wind, 10m v-component of wind, 2m temperature, sea surface temperature, surface pressure, total precipitation, and forecast albedo. The data set spans the period from January 1980 to the present, with a monthly time resolution. The data are presented at a spatial resolution of 0.25°× 0.25°. The data encompass a spatial range of 90° north, -180° west, 60° south, and 180° east.

2.2. Methods

2.2.1. Linear regression model

In order to gain a more intuitive understanding of the trend of each independent variable over time, our study constructed linear regression models of each independent variable with time. The basic form of the constructed linear regression model is as follows:

$$y_i = \hat{a}x_i + \hat{b} \tag{1}$$

Among them, $\hat{a}$, $\hat{b}$ are the undetermined coefficient, which can be determined by using the least squares method:

$$\hat{a} = \frac{\sum_{i=1}^{n} y_i x_i - \frac{1}{n}\left(\sum_{=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{\sum_{i=1}^{n} {x_i}^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} \tag{2}$$

$$\hat{b} = \overline{y} + \hat{a}\overline{x} \tag{3}$$

Among them:

$$y = \frac{1}{n}\sum_{i=1}^{n,} y_i, \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4}$$

2.2.2. Moving average

In order to mitigate the effects of seasonality, our study employed a straightforward averaging technique on the aforementioned data, utilising a moving window of 12, which represents the shortest possible interval for variables and parameters.

The formula for a simple moving average is as follows:

DOI: 10.54254/2755-2721/2025.29848

$$F_t = (A_{t-1} + A_{t-2} \ldots + A_{t-n})/n \tag{5}$$

In the formula: F_t - The predicted value for the forthcoming period; n - The number of periods comprising the moving average; A_{t-1} - The actual value at the outset; A_{t-2} & A_{t-n} The actual values for the initial two periods, the first three periods, and up to the first n periods

2.2.3. Granger causality test

In order to ascertain the causal relationship between the selected variables and sea ice extent with greater precision, our study employed Granger causality tests on each variable and sea ice.

2.2.3.1. VAR model construction

The VAR model is a statistical tool that can be used to test for the existence of a causal relationship between variables. In other words, it is employed to ascertain whether changes in a given variable are influenced by its own past behaviour and the past behaviour of other variables. The model primarily measures the dynamic relationship between variables through the use of actual data, as opposed to relying on theoretical experience. Typically, regression models are constructed by treating all variables in the system as functions of variable lag terms.

The formula for the variable-angle regression (VAR) model with two variables is expressed as follows:

$$Y_t = \sum_{i=1}^{m} \alpha_i y_{t-i+} \sum_{i=1}^{m} \beta_i x_{t-i} + c_{1t} + \varepsilon_{1t}$$

$$X_t = \sum_{i=1}^{m} \lambda_i y_{t-i+} \sum_{i=1}^{m} \delta_i x_{t-i} + c_{2t} + \varepsilon_{2t} \tag{6}$$

2.2.3.2. Causal relationship test

The Granger causality method is employed in the vector autoregression (VAR) model to ascertain whether a causal relationship exists between variables. This is achieved by evaluating the significance of coefficients, which entails determining whether all lagged variables exert an influence on the present values of other variables. The null hypothesis of the Granger causality test is that the tested variable is not a causal factor in relation to the dependent variable. If the probability value of the test is less than the set confidence level, the null hypothesis is rejected and the tested variable is deemed to constitute a causal relationship. Conversely, the tested variable is not considered to constitute a causal relationship with the dependent variable.

The following four scenarios may be considered:

The first situation is characterised by a unidirectional causal relationship from x to y, where α is non-zero overall and λ is zero overall.

The second situation is defined by a unidirectional causal relationship from y to x, where α is zero overall and λ is non-zero overall.

DOI: 10.54254/2755-2721/2025.29848

A bidirectional causal relationship from x to y is observed, concurrently with a unidirectional causal relationship from y to x. Both α and λ are non-zero as a whole.

No causal relationship is identified between x and y, with both α and λ being zero as a whole.

2.2.4. Collinearity diagnostics

In order to minimise the impact of correlation between the selected independent variables on model establishment, a collinearity test was conducted on the independent variable data using the variance inflation factor.

The principal testing methods employed are those of tolerance and variance inflation factor (VIF). The most frequently employed method is that of calculating the variance inflation factor (VIF), which is achieved through the utilisation of the following formula:

$$VIF = \frac{1}{1 - R_i^2} \tag{7}$$

In regression analysis, R_i represents the negative correlation coefficient between the independent variable X_i and the other independent variables.

The value of VIF is greater than 1. As VIF approaches 1, the degree of multicollinearity increases, and vice versa.

In the majority of cases, there are four main scenarios to consider.

When VIF is within the range of (0,5], it can be concluded that there is no evidence of multicollinearity between the independent variables.

Similarly, when VIF falls within the interval of (5,10], it can be deemed that there is a minimal level of multicollinearity between the independent variables.

In the event that VIF is within the range of (10, 100], it can be posited that there is moderate multicollinearity between the independent variables.

Conversely, when VIF is situated within the range of (100, ∞), it can be inferred that there is severe multicollinearity between the independent variables.

3. Results

3.1. Temporal trend of Arctic sea ice extent

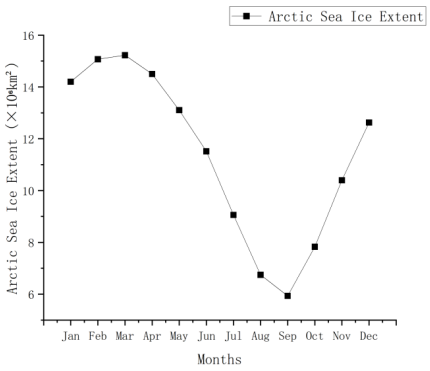


Figure 1. The line chart of mean Arctic sea ice extent over month ($10^6 km^2$)

DOI: 10.54254/2755-2721/2025.29848

As illustrated in the figure, there is a distinct seasonal variation in sea ice extent, with the maximum value occurring between February and March and the minimum value occurring between August and October. It is evident that the extent of sea ice will decrease from March to September and subsequently increase from September to March in the following year. These are the fundamental characteristics of the monthly variation in Arctic sea ice extent.

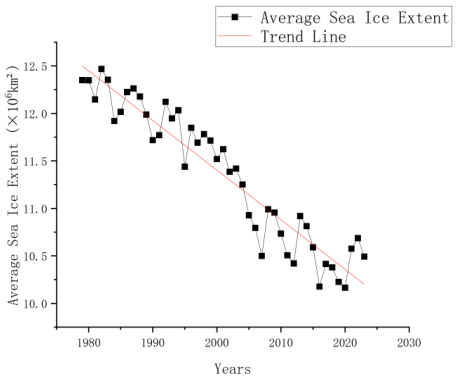


Figure 2. The line chart and Trend line of mean Arctic sea ice extent over time (10^6km^2)

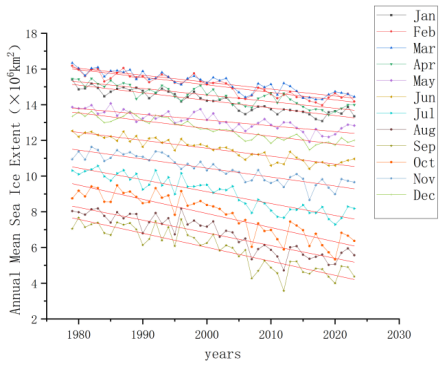


Figure 3. The line chart and trend lines of mean sea ice extent by month (10^6km^2)

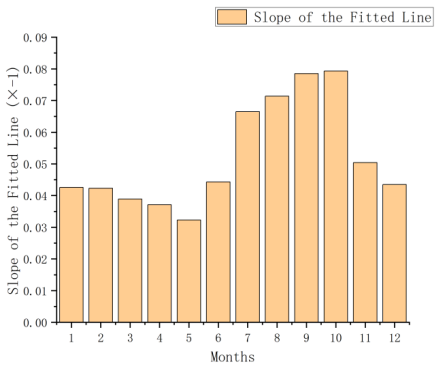


Figure 4. The comparison between the slope of the fitted line by month (-1)

As illustrated by Figure 2, the annual average sea ice extent in the Arctic demonstrates a fluctuating downward trend with increasing years. A linear fit was performed on the curve, with an intercept of 115.93471 ± 5.30 ($10^6/km^2$/year) and a slope of -0.05227 ± 0.002 ($10^6/km^2$/year). The

DOI: 10.54254/2755-2721/2025.29848

results of the fitting indicate a year-on-year decrease in the annual average sea ice extent in the Arctic region.

Furthermore, as illustrated in Figure 3, the Arctic sea ice extent has exhibited a consistent decline on an annual basis across various months. Furthermore, by employing a linear regression analysis to fit the twelve months, our study obtained the slope of the fitting line for the annual variation of the Arctic sea ice extent in each month, as illustrated in Figure 4.

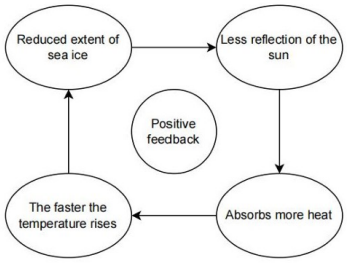


Figure 5. The possible mechanism of the abnormally high value

It is notable that the slope of the fitting line in September and October is significantly higher than that observed in other months. This phenomenon may be attributed to the fact that a reduction in sea ice extent during the summer months results in a decrease in solar radiation, which subsequently leads to an increase in heat absorption. This results in a greater absorption of heat, which in turn causes a more rapid increase in temperature. This, in turn, contributes to a greater reduction in sea ice extent. It is this positive feedback loop that is responsible for the accelerated reduction in Arctic sea ice extent observed in September and October. The mechanism of which is illustrated in Figure 5.

3.2. Pearson correlation coefficient heatmap

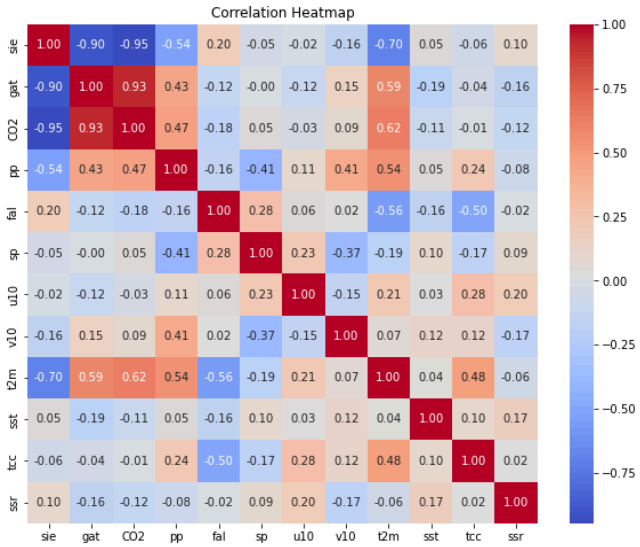


Figure 6. Pearson correlation coefficient heatmap based on various variables

DOI: 10.54254/2755-2721/2025.29848

Table 1. The full names, ranges, and units of each variable in the heatmap

Abbreviation	Full name in English	range	unit
Sie	Sea Ice Extent	arctic	$\times 10^6 km^2$
Gat	Average Global Temperature	Global	K
CO_2	Carbon Dioxide	Global	ppm
Pp	Precipitation	Latitude 60°-90°N	mm/day
Fal	Forecast albedo	Latitude 60°-90°N	degrees_east
Sp	Surface pressure	Latitude 60°-90°N	Pa
u10	10 metre U wind component	Latitude 60°-90°N	m/s
v10	10 metre V wind component	Latitude 60°-90°N	m/s
t2m	2 metre temperature	Latitude 60°-90°N	K
sst	Sea surface temperature	Latitude 60°-90°N	K
tcc	Total cloud cover	Latitude 60°-90°N	(0 - 1)
ssr	Surface net short-wave (solar) radiation	Latitude 60°-90°N	degrees_east

The present study conducted a basic analysis of the factors affecting Arctic sea ice extent as mentioned in previous research. Furthermore, the Pearson correlation coefficients between various independent variables were calculated and a Pearson correlation coefficient heatmap was plotted.

The heatmap demonstrated that the independent variables exhibiting the strongest correlation with sea ice extent were CO_2 concentration, GAT, and 2-metre temperature. The independent variable that demonstrated a moderate correlation with sea ice extent was precipitation. The independent variables exhibiting a weak correlation with sea ice extent were total cloud cover, sea surface temperature, 10-metre meridional wind component, 10-metre radial wind component, surface pressure, and predicted albedo.

3.3. The results of Granger causality test between variables and sea ice extent

In light of the findings above, our study proceeded to undertake further Granger causality testing on the selected variables: carbon dioxide concentration, global average temperature, 2-meter temperature, and precipitation.

Table 2. The Granger causality test results of VAR model

X	Y	Chi-sq	df	Prob	H_0
SIE	T2M	10.6676	11	0.4715	Accept
	CO_2	14.5040	3	0.0023	Reject
	GAT	29.7029	6	0.0000	Reject
	PP	28.7559	12	0.0043	Reject
T2M	SIE	38.1043	11	0.0001	Reject
CO_2		39.4394	3	0.2676	Accept
GAT		17.7495	6	0.0069	Reject
PP		36.7608	12	0.0002	Reject

H0: Assuming Y is the cause of X's occurrence

DOI: 10.54254/2755-2721/2025.29848

Table 2 presents the findings of the VAR model construction and Granger causality test for SIE with T2M, CO_2, GAT, and PP, respectively. The results demonstrate that the probability values of the causality tests for the effects of GAT, PP, and CO_2 on SIE are 0.0000, 0.0043, and 0.0023, respectively. All of these values are less than 0.05, indicating a Granger causality on SIE at the 5% level. Consequently, GAT, PP, and CO_2 can be regarded as the direct causes of the observed changes in SIE. This indicates that an increase in global temperature and the greenhouse effect result in alterations to the sea ice extent. In the physical dimension, changes in precipitation directly influence the sea ice. The probability of the causality test of the effect of T2M on SIE is greater than 0.05, and it does not constitute a Granger causality on SIE at the 5% level. Therefore, T2M is not considered a direct cause of SIE. However, considering the correlation results between SIE and T2M, as T2M increases, SIE exhibits a notable downward trend, and T2M is significantly correlated with changes in SIE. Therefore, it is believed that the statistical results of T2M on SIE do not align with reality.

The probability values of the SIE to T2M, GAT, and PP influence causality test are 0.0001, 0.0069, and 0.0002, respectively. These values are less than 0.05 and constitute a Granger causality to T2M, GAT, and PP, respectively, at the 5% level. This suggests that SIE is the cause of the changes in T2M, GAT, and PP. The probability of CO_2 influencing SIE is 0.2676, which is greater than 0.05 and does not constitute a Granger causality on SIE at the 5% level. This suggests that SIE is not a direct cause of changes in CO_2; however, it is possible that CO_2 may indirectly influence SIE. This indicates that a reduction in SIE will result in additional alterations to GAT, PP, and T2M, which will have a more pronounced impact on SIE itself.

3.4. The results of Collinearity diagnostics between variables

Table 3. The results of multicollinearity test

Variable	R^2	VIF
PP	0.3203	1.1143
GAT	0.8728	4.1978
CO_2	0.8825	4.5209
T2M	0.4629	1.2727

The calculated variance inflation factors (VIFs) for the four covariates PP, GAT, CO_2 and T2M are 1.1143, 4.1978, 4.5209 and 1.2727, respectively. This indicates that the problem of covariance among the four covariates is relatively minor and can be used as a predictive element.

4. Conclusions

Seasonal variation in sea ice extent, with the maximum value occurring between February and March and the minimum value occurring between August and October. The annual average sea ice extent in the Arctic exhibits a fluctuating downward trend with increasing years, and the period with the fastest change in sea ice extent is from August to October. By undertaking a detailed analysis of each variable, our study has finally identified the predictive elements that can be involved in the construction of a model for predicting the Arctic sea ice extent. These are: The variables identified as predictive elements in the construction of a model for predicting the Arctic sea ice extent are CO_2, GAT, PP and T2M.

DOI: 10.54254/2755-2721/2025.29848

It is hoped that further relevant variables can be included in the discussion in order to obtain more accurate predictors of the Arctic sea ice extent. It is also hoped that future researchers will construct mechanistic learning models based on the four elements of CO_2, GAT, PP and T2M, with the aim of developing models with higher accuracy.

Authors' contributions

Haoxu Yin: Conceptualization, Writing - review & editing; Qiyang Chen: Conceptualization, Investigation, Writing - original draft, Writing - review & editing; Yanfei Yang: Formal analysis, Validation, Software, Visualization; Writing - original draft; Zujian Liu: Conceptualization, Data curation, Formal analysis, Validation, Visualization

Acknowledgement

Haoxu Yin, Qiyang Chen, Yanfei Yang and Zujian Liu contributed equally to this work and should be considered co-first authors.

References

[1] AMAP, 2021. Arctic Climate Change Update 2021: Key Trends and Impacts. Summary for Policy-makers. Arctic Monitoring and Assessment Programme (AMAP), Tromsø, Norway. 16 pp
[2] Arctic report card: update for 2021: rapid and pronounced warming continues to drive the evolution of the Arctic environment [R]. 2021. https: //arctic.noaa.gov/Report-Card/Report-Card-2021.
[3] Meier, W. N. . (2016). Losing Arctic sea ice: observations of the recent decline and the long-term context. John Wiley & Sons, Ltd.
[4] DESER C, WALSH J E, TIMLIN M S. Arctic sea ice variability in the context of recent atmospheric circulation trends [J]. Journal of Climate, 2000, 13(3): 617-633.
[5] TIAN Z, LIANG X, ZHANG J, et al. Thermodynamical and dynamical impacts of an intense cyclone on Arctic sea ice [J]. Journal of Geophysical Research: Oceans, 2022, 127(12): e2022JC018436.
[6] ZHANG T, ZHANG J, ZHANG X. Analysis of Arctic Sea ice thickness from 2014 to 2018 based on CryoSat-2Data [J].Advances in Marine Science, 2020, 3(38): 425-434.
[7] Kim J, Kim K, Cho J, et al. Satellite-based prediction of Arctic sea ice concentration using a deep neural network with multimodel ensemble [J]. Remote Sensing, 2018, 11(1) ; 19.
[8] Yang Yanrui, Zhang Yining, Lu Zhiwei, et al. Research on Arctic Sea Ice Range Prediction Based on Atmospheric and Oceanic Multivariables and Sea Ice Time Series Characteristics [J]. Ocean Bulletin, 2024, 43 (04): 521-533
[9] Yue Handong, Dou Tingfeng, Li Runkui, et al. Monthly scale Arctic sea ice prediction based on multiple machine learning and stacked ensemble methods [J]. Glacier and Frozen Soil, 2023, 45 (03): 893-901