Research on Sea Surface temperature Reconstruction from long-term MODIS data

Yan Zhou *, Yingying Gai and Junxiao Li

Institute of Oceanographic Instrumentation, Qilu University of Technology (Shandong Academy of Sciences), Shandong Provincial Key Laboratory of Marine monitoring instrument equipment technology, National Engineering and Technological Research Center of Marine Monitoring Equipment, Qingdao, China

*Corresponding author e-mail: zhouyan_oers@qlu.edu.cn

Abstract. Sea surface temperature (SST) is one of the important dynamic parameters to characterize the physical characteristics of seawater, and it is also an important basic auxiliary parameter for the quantitative observation of other ocean parameters. In view of the lack of remote sensing data mainly caused by cloud cover over the ocean, based on the long-term MODIS Aqua L3 8-day mean SST products from January 2003 to March 2015, the missing data in the East China Sea are reconstructed by Data Interpolating Empirical Orthogonal Function (DINEOF) method, and the reconstruction results are evaluated by error analysis before and after cross-validation data reconstruction. It shows that the image reconstructed by DINEOF method can reflect the temporal and spatial variation characteristics of sea surface temperature in the study area, and the root mean square error of reconstruction is 0.4272. DINEOF method can reconstruct large area missing data without any prior information and has high accuracy.

1. Introduction

Sea surface temperature (SST) is one of the important dynamic parameters to characterize the physical characteristics of seawater, and it is also an important basic auxiliary parameter for the quantitative observation of other ocean parameters. The sea surface temperature data products are retrieved by thermal infrared band, but there are some objective factors, such as cloud cover, thin cloud detection algorithm is not ideal, sensor noise and so on, which makes the missing rate of water temperature remote sensing products is relatively high, which seriously affects the practical application of ocean research and operational work [1].

In recent years, domestic and foreign scholars have done a lot of research on the reconstruction of missing remote sensing data. In 2003, Beckers and Rixen jointly proposed a parameter-free adaptive EOF decomposition-based method to reconstruct the missing data in the time series data set, and respectively analyze the simulated time series and the actual Adriatic Sea. The missing data was reconstructed in the AVHRR dataset of ocean remote sensing [2]. Alvera-Azcárate et al. (2005) taken a data interpolation method DINEOF (Data Interpolating Empirical Orthogonal Function) based on EOF decomposition [3], which carried out missing data on 135 SST images of ocean remote sensing AVHRR data in the Adriatic Sea The reconstruction of, and the introduction of the cross correction set, in the process of establishing EOF, the optimal cutoff and the estimated error of the default value are obtained.
from the cross correction set [4]. Alvera-Azcárate et al. (2007) applied the DINEOF method on multivariate elements to reconstruct the sea surface temperature, chlorophyll concentration and wind field on the west coast of Florida, USA, and achieved good results [5]. Sheng Zheng et al. (2009) used the DIEOF method to reconstruct the missing data of sea surface temperature data in the sea area of the Yangtze River Estuary [6]. Ding (2009) also specially carried a reconstruction method combining empirical mode decomposition and empirical orthogonal functions, and reconstructed the missing data in the remote sensing products of sea surface temperature and surface suspended sediment concentration in the Yangtze River Estuary in 2003 [7]. In the process of reconstructing the missing data of the sea surface temperature data set, Wang Jianle et al. (2012) proposed a new method of missing data reconstruction combining self-organizing mapping algorithm and empirical orthogonal function [8]. Ji Chenxu et al. (2018) used the DINEOF method to reconstruct the East China Sea sea surface chlorophyll a concentration products from 2003 to 2016, and used correlation analysis and regression analysis to evaluate the relationship between the chlorophyll a concentration and sea surface temperature [9].

In this paper, MODIS Aqua L3 SST (4km, 8day) is used to reconstruct remote sensing data in the East China Sea, and the DINEOF method is used to reconstruct large areas of missing data to provide reliable data reconstruction services for the later application of remote sensing data.

2. Method

The DINEOF method is self-adaptive for data reconstruction and does not require any prior information. Through EOF decomposition, the spatial feature mode and temporal feature mode reflecting the data change are extracted from a large amount of data, and the best mode is retained. Data reconstruction can effectively impute missing data. The basic principle is as follows [10]:

Assume that $X^0$ is a two-dimensional data matrix with size of $m \times n$ for multiple time-series data image data, where $m$ is the number of pixel points of sea remote sensing data, and $n$ is the number of time-series images. $X^0$ includes some missing points without data, such as cloud coverage, unreliable data points, etc., which are represented by NaN.

(1) Convert the raw matrix $X^0$ into an anomaly matrix, and obtain the mean matrix $X_{\text{mean}}$. Randomly select part of the effective data point set as the cross-correction set for judging the best reconstruction mode number, and assign NaN to the position in the cross-correction set.

(2) Replace the points with NaN value in the anomaly matrix with 0, make the initial value of the missing point the unbiased estimate value of the data set, set the EOF mode retention number $P=1$, and the reconstruction iteration number $k=1$.

(3) The singular value decomposition (SVD) method is selected for the anomaly matrix to perform EOF decomposition, and the most important $P$ EOF modes are obtained.

$$X = USV^T$$

Where, $U$ is the spatial eigenmode after SVD decomposition, $S$ is the singular value matrix, $V$ is the time eigenmode, and $T$ represents the matrix transposition. The reconstruction value of the missing point is calculated as follows:

$$X_{i,j} = \sum_{i=1}^{P} \rho_i (u_i)(v_i^T)$$

Where, $i$ and $j$ are the space and time subscripts of matrix $X$, $u_i$ is the column of space mode $U$, $v_i$ is the column of time mode $V$, $\rho_i$ and is the corresponding singular value.

(4) Calculate the root mean square error RMSE between the reconstructed value of the cross-correction set and the raw image, take $P=2,\ldots,P_{\text{max}}$ and repeat step (3) to calculate the corresponding RMSE, and compare the EOF mode retention number $P$ with the minimum RMSE, where $P_{\text{max}}$ is determined by the time dimension of the observation data.

(5) After confirming $P=P$, let $k=1,\ldots,N_{\text{ITEMAX}}$, repeat steps (1)-(3) for data reconstruction, where $N_{\text{ITEMAX}}$ is the maximum number of iterations preset by the program, calculate the root mean
3rd International Conference on Air Pollution and Environmental Engineering IOP Publishing
IOP Conf. Series: Earth and Environmental Science 631 (2021) 012032 doi:10.1088/1755-1315/631/1/012032

square error RMSE between the reconstruction value of the cross correction set and the original value, and compare The number of reconstruction iterations $\bar{k}$ with minimum RMSE is obtained.

(6) Let $P = \tilde{P}$ and $k = \tilde{k}$, and repeat steps (1)~(3) to reconstruct the data, calculate the reconstruction value of all points, set $X_{re} = X_0 + X_{mean}$, the valid point data of the original data remains unchanged, and the missing point data is replaced with the reconstruction value to obtain the reconstruction dataset $X_{re}$, and $X_{re}$ would be finally transformed into a time series image.

Reconstruction error of the original data set is calculated as:

$$\bar{R} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (X_{re} - X^0_t)^2}$$

Where, N is the total number of valid points, $X^0$ and $X_{re}$ are the original data set and reconstructed data set, respectively.

3. Data
Select the MODIS Aqua L3 SST (4km, 8day) product provided by NASA OceanColor website, use the DINEOF method to reconstruct the data in the East China Sea area, and select part of the valid data in the original data as the cross-validation data set to verify and evaluate the reconstructed data accuracy. The image size of the study area is 408×336, and the time period is from January 2003 to March 2015, a total of 563 time periods of 147 months. The annual average SST image of the study area is shown in Figure 1(a), and the missing data coverage statistics As shown in Figure 1(b).

Figure 1. The annual average SST data image in the study area (a) and the time missing rate statistics of the raw dataset (b).

Due to SST data products mainly retrieved from thermal infrared channel data, this band is easily affected by clouds, fog and haze. Affected by the weather over the sea at all times, there will be large areas of irregular data missing and abnormal. It can be seen from Figure 3 that the 8-day average SST data missing is smaller in the 27th to 33rd 8 days of the year, that is, the cloud cover in autumn from August to September is small, and the inversion is due to the cloud cover at other times The availability of water temperature and water color products has greatly reduced, which seriously affects the application of remote sensing data products in marine environment monitoring. Therefore, it is necessary to solve the problem of data missing caused by factors such as satellite remote sensing data cloud coverage, and reconstruct to obtain spatio-temporal continuous data.

4. Results and discussions
The DINEOF method interpolates the missing data of the target element based on the transformation law of the first few main modal time coefficients after the EOF decomposition of one or more reference elements. When reconstructing the anomaly matrix, it is necessary to determine the number of EOF modal retention and the number of iterations, and the optimal value of the parameter is selected through
cross-validation. Through cross-validation, the optimal retention number of the EOF mode can be obtained as 3 and the error is 1.5198. The optimal value of the number of reconstruction iterations is 17, and the RMSE is 0.8972. The final reconstruction error of the original observation data is calculated from the error between the original data and the reconstruction data of all valid points.

The first three main spatial modes and temporal modes used for reconstruction are shown in Figure 2 and Figure 3. The first eigenvector of the spatial mode is distributed in a belt in the latitude direction, and the temperature is shown as a state of increasing temperature from northwest to southeast. At the same time, the Yellow Sea masses and the Taiwan warm current and Kuroshio tributaries intersect in the horizontal direction at 30 north latitude. In the performance of the temperature front zone formed on the East China Sea shelf, the spatial mode index is all negative, and there is an obvious antiphase relationship. The time coefficient of the first mode shows obvious seasonal changes, indicating that the season is the strongest factor affecting the sea surface temperature. There are obvious regional differences in the influence corresponding to the second spatial mode. The spatial distribution is bounded by its zero line, reflecting the influence of land on ocean temperature, and coastal water temperature is higher than that of adjacent seas. The second mode shows that the factors that affect this mode have different effects on different latitudes. This factor may be related to some effects brought by the Coriolis parameters, or it may be the influence of the sun. The distribution of the third spatial mode has obvious regional differences, but there is no obvious periodicity in the time mode, and its influencing factors may be more complicated.

![Figure 2. The spatial characteristic of the first three modes using reconstruction](image)

![Figure 3. The temporal characteristic of the first three modes using reconstruction](image)
Using the results of the optimal parameter selection, the missing data matrix is reconstructed according to the DINEOF reconstruction algorithm, it retains the large and medium-scale information in the original data, smooths part of the small-scale information, and obtains a remote sensing product with full spatial distribution, as shown in Figure 4.

![Figure 4. The 8-day average SST raw images of 2004017 (a), 2004081 (b), and 2004361(c), and corresponding reconstructed images (d-f).](image)

In Figure 5, the data missing rate of each original image is quite high, 62.83%, 32.27% and 73.27% respectively. Especially when there is a large area missing in the image (as shown in Figure c), it is difficult for the traditional data interpolation method to achieve such an interpolation effect. The DINEOF algorithm not only can effectively reconstruct the missing observation data with higher reconstruction accuracy, but also the temporal and spatial distribution characteristics of the remote sensing data set can be maintained to facilitate the analysis of the data set.

After reconstruction, error analysis is performed on the reserved data randomly selected from the non-missing data part of the missing data matrix, and the accuracy parameter values of the measured and predicted data of the detected pixel points before and after reconstruction are calculated. The accuracy parameters have selected RMSE (root mean square error) and MAE (average absolute error), and the accuracy detection result after the abnormal value removal is shown in Figure 6. The original value and the reconstructed predicted value are concentrated in the vicinity of the y=x line. At the same time, the three accuracy index values are also ideal, the RMSE is 0.4272, The MAE is 0.5953, indicating the effectiveness of the reconstruction results on the whole.
5. Conclusion

This study selects MODIS 8-day average sea surface temperature L3 product to reconstruct missing remote sensing data in the East China Sea using the DINEOF method, and analyzes and evaluates the reconstruction results from temporal and spatial changes and precision index calculations. The DINEOF method can effectively reconstruct the missing data in sea surface temperature remote sensing images, especially for images with large area missing or a high percentage of missing data, and can have a better reconstruction effect without prior error statistical information. The reconstructed sea surface temperature image can better retain the temporal and spatial characteristics of the time series data, and cross-validate the reserved data. The final reconstructed root mean square error is 0.4272. Due to the lack of SST data, we can consider using SST data as auxiliary data in future research work, and use the DINEOF algorithm to reconstruct some other ocean remote sensing water color physical quantities, such as chlorophyll concentration and suspended solids concentration. The result images obtained by other missing data interpolation methods adopt a certain fusion principle for image fusion to obtain the final missing data reconstruction result.

Acknowledgments

This work was financially supported by, the National Key Research and Development Program of China, grant number 2017YFC1404802, the Key Research and Development Program of Shandong, grant number 2019GHY112017, State Key Laboratory of Tropical Oceanography, South China Institute of Oceanology, Chinese Academy of Sciences, grant number LTO2017, and the Natural Science Foundation of Shandong, grant number ZR2019PD021.

References

[1] Sirjacobs, D.; Alvera-Azcarate, A.; Barth, A.; Lacroix, G.; Park, Y.; Nechad, B.; Ruddick, K.; Beckers, J.M. Cloud filling of ocean colour and sea surface temperature remote sensing products over the Southern North Sea by the Data Interpolating Empirical Orthogonal Functions methodology. J Sea Res 2011, 65, 114-130.

[2] Beckers, J.M.; Rixen, M. EOF calculations and data filling from incomplete oceanographic datasets. Journal of Atmospheric & Oceanic Technology 2003, 20, 1839-1856.

[3] Alvera-Azcarate, A.; Barth, A.; Beckers, J.M.; Weisberg, R.H. Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields. J Geophys Res-Oceans 2007, 112.

[4] AlveraAzcarate, A.; Barth, A.; Rixen, M.; Beckers, J. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature. Ocean Model 2005, 9, 325-346.

[5] AlveraAzcarate, A.; Barth, A.; Beckers, J.; Weisberg, R.H. Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields. Journal of Geophysical Research 2007, 112.
[6] Sheng, Z.; Shi, H.Q.; Ding, Y.Z. Missing satellite-based sea surface temperature data reconstructed by DINEOF method. Advances in Marine Science 2009, 27, 243-249.

[7] Ding, Y.Z. Data reconstruction and assimilation experiment of satellite sea surface temperature and suspended sediment concentration. Doctor, Nanjing University of Science & Technology, 2009.

[8] Wang, J.L. A reconstruction method study of ocean color and SST based on remote sensing data. Master, Tianjin University, 2012.

[9] Hilborn, A.; Costa, M. Applications of DINEOF to Satellite-Derived Chlorophyll-a from a Productive Coastal Region. Remote Sens-Basel 2018, 10.

[10] Beckers, J.M.; Barth, A.; Alvera-Azcarate, A. DINEOF reconstruction of clouded images including error maps - application to the Sea-Surface Temperature around Corsican Island. Ocean Sci 2006, 2, 183-199.