

Development of Monthly-Seasonal-Annual Dataset of Sea Surface Chlorophyll-a Concentration for 21 Years (1998–2018)

Li, L. W.¹ Fu, Y. X.^{1,2} Xue, C. J.^{2,3*} Cui, J. Y.¹ Zhang, Y. Y.^{1,2} Xu, Y. F.^{1,2}

1. College of Oceanography and Space Informatics, China University of Petroleum, Qingdao 266580, China;
2. Key Laboratory of Digital Earth Science, Chinese Academy of Sciences, Beijing 100094, China;
3. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

Abstract: Based on the Chlorophyll-a concentration products retrieved from SeaWIFS, Terra-MODIS, Aqua-MODIS, MERIS and VIIRS from January 1998 to December 2018. This paper combines the wavelet transform and Kalman filtering to develop the multi-source remote sensing data fusion technology, and integrates the look-up table and the maximum value composite to design the data fusion method. The algorithm completes the global sea surface chlorophyll-a concentration fusion dataset (1998–2018) at three temporal scales: monthly, seasonal and annual. The spatial resolution of the dataset is 4 km × 4 km, the data format is TIFF. The dataset includes 357 data files and the data size is 50.1 GB, and the compressed dataset is 23 data files and the compressed data size is 19.4 GB. Comparisons with the measured data and GSM (Graver-Siegel-Maritorena) products from European Space Agency show that the correlation between the fusion product and the measured data is 79% in 2008, while the correlation between the GSM product and the measured data is only 35%. The results prove the advantages of this fused products.

Keywords: sea surface chlorophyll-a; data fusion; global dataset; 1998–2018

DOI: <https://doi.org/10.3974/geodp.2021.02.15>

CSTR: <https://cstr.escience.org.cn/CSTR:20146.14.2021.02.15>

Dataset Availability Statement:

The dataset supporting this paper was published and is accessible through the *Digital Journal of Global Change Data Repository* at: <https://doi.org/10.3974/geodb.2021.05.05.V1> or <https://cstr.escience.org.cn/CSTR:20146.11.2021.05.05.V1>.

1 Introduction

Chlorophyll-a is a key element of water constituents, its value can reflect the primary productivity of water body and it is an index to evaluate the degree of eutrophication^[1].

Received: 06-04-2021; **Accepted:** 30-05-2021; **Published:** 25-06-2021

Foundation: Chinese Academy of Sciences (XDA19060103)

***Corresponding Author:** Xue, C. J. 0000-0003-3605-6578, Aerospace Information Research Institute, Chinese Academy of Sciences, xuecj@aircas.ac.cn

Data Citation: [1] Li, L. W., Fu, Y. X., Xue, C. J., *et al.* Development of monthly-seasonal-annual dataset of sea surface chlorophyll-a concentration for 21 years (1998–2018) [J]. *Journal of Global Change Data & Discovery*, 2021, 5(2): 219–225. <https://doi.org/10.3974/geodp.2021.02.15>. <https://cstr.escience.org.cn/CSTR:20146.14.2021.02.15>.

[2] Li, L. W., Fu, Y. X., Xue, C. J., *et al.* Global ocean surface chlorophyll-a concentration fusion 4-km grid dataset (1998–2018) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2021. <https://doi.org/10.3974/geodb.2021.05.05.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2021.05.05.V1>

Therefore, it is of great significance to know the status and change of chlorophyll-a for the maintenance of ecological quality. Various countries in the world have launched many marine satellites, which can conduct large-scale and high-precision observation of the marine environment. It can provide massive data for researchers in various countries to study the marine environment, and also provide a solid data foundation for the detection of marine chlorophyll-a concentration^[2-4].

Due to the variations in spatial and temporal resolution of different sensors, single satellite image has various shortcomings in coverage, resolution and utilization. In order to overcome the existing defects and to make full use of the advantages of multiple satellite images, researchers have proposed multi-source data fusion technology^[5]. Multi-source data fusion can increase the data quantity, expand the spatial-temporal resolution, and improve the spatial-temporal continuity, consistency and reliability^[4].

While the multi-source data fusion technology has been widely used^[6-8], most of them still have problems in terms of accuracy, coverage rate or time span. This paper used the chlorophyll-a concentration products retrieved from SeaWiFS, Terra, Aqua, MERIS and VIIRS sensors, and designed a fusion algorithm based on wavelet transform and Kalman filtering technology. The fused data are at monthly, seasonal and annual scales and have a span from Jan. 1998 to Dec. 2018.

2 Metadata of the Dataset

The dataset name, authors, geographical region, years of dataset, temporal resolution, spatial resolution, data files, data publishing and sharing service platform, data sharing policy and other information of the Global ocean surface chlorophyll-a concentration fusion 4-km grid dataset (1998–2018)^[9] are shown in table 1.

3 Methods

3.1 Data Collection

Global oceanic chlorophyll-a concentration dataset comes from the Ocean Color official network of NASA^[11]. According to the sensor life cycle, the global sea surface chlorophyll-a concentration data from January 1998 to December 2018 were downloaded, a total of 2,921 scenes are shown in table 2 and the life cycle of each sensor is shown in Figure 1.

The measured data of chlorophyll-a concentration on the global sea surface come from SeaBass^[12]. A total of 5,711 observation files were recorded in each observation file, including site, number, date, time, latitude, longitude, water depth, chlorophyll concentration and other information. The statistical table of measured data is shown in Table 3.

3.2 Data Processing

Based on the adaptive weighted fusion algorithm, the fusion dataset of global sea surface chlorophyll-a concentration is generated. The core idea of the algorithm is: (1) Taking the regional optimum as the evaluation criterion of weight selection, the data collected by each sensor can adaptively find the corresponding optimal weighting factor; (2) Dynamic weighted fusion based on wavelet transformation. The design process of adaptive weighted fusion algorithm is shown in Figure 2.

The specific steps of the fusion algorithm include:

(1) firstly, the threshold iterative method is used to segment the image according to whether there have measured values or not;

(2) the regions with no measured values are divided into the similar (homogeneous)

regions and the variable (heterogeneous) ones according to their variances;

Table 1 Metadata summary of the Global ocean surface chlorophyll-a concentration fusion 4-km grid dataset (1998–2018)

Items	Description
Dataset full name	Global ocean surface chlorophyll-a concentration fusion 4-km grid dataset (1998–2018)
Dataset short name	Global_Chla_1998-2018
Authors	Li, L. W., College of Oceanography and Space Informatics, China University of Petroleum, lilianwei78@163.com Fu, Y. X., College of Oceanography and Space Informatics, China University of Petroleum, 624002974@qq.com Xue, C. J. 0000-0003-3605-6578, Aerospace Information Research Institute, Chinese Academy of Sciences, Key Laboratory of Digital Earth Science, Chinese Academy of Sciences, xuecj@aircas.ac.cn Cui, J. Y., College of Oceanography and Space Informatics, China University of Petroleum, xjuzhxcjy@163.com Zhang, Y. Y., College of Oceanography and Space Informatics, China University of Petroleum, 1529142841@qq.com Xu, Y. F., College of Oceanography and Space Informatics, China University of Petroleum, xuyf187627@163.com
Geographical region	Global sea surface
Year	From January 1998 to December 2018
Temporal resolution	Month / season / year
Spatial resolution	4 km × 4 km
Data format	Geo-TIFF
Data size	50.1 GB (19.4 GB after compression)
Data files	Chlorophyll-a concentration fusion monthly scale dataset of global ocean surface Chlorophyll-a concentration fusion seasonal scale dataset of global ocean surface Chlorophyll-a concentration fusion annual scale dataset of global ocean surface
Foundations	Chinese Academy of Sciences (XDA19060103)
Data publisher	Global Change Research Data Publishing & Repository, http://www.geodoi.ac.cn
Address	No. 11A, Datun Road, Chaoyang District, Beijing 100101, China
Data sharing policy	Data from the Global Change Research Data Publishing & Repository includes metadata, datasets (in the <i>Digital Journal of Global Change Data Repository</i>), and publications (in the <i>Journal of Global Change Data & Discovery</i>). Data sharing policy includes: (1) Data are openly available and can be free downloaded via the Internet; (2) End users are encouraged to use Data subject to citation; (3) Users, who are by definition also value-added service providers, are welcome to redistribute Data subject to written permission from the GCdataPR Editorial Office and the issuance of a Data redistribution license; and (4) If Data are used to compile new datasets, the ‘ten per cent principal’ should be followed such that Data records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset ^[10]
Communication and searchable system	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS

Table 2 Information of input remote sensing datasets of global sea surface chlorophyll-a concentration

Sensor	Algorithm	Start time	End time	Quantity	Temporal resolution	Spatial resolution
SeaWiFS	OCI	Sep. 18, 1997	Nov. 17, 2010	644	8 days	9 km
Terra	OCI	Jul. 4, 2002	Apr. 13, 2018	736	8 days	4 km
Aqua	OCI	Jul. 4, 2002	Apr. 13, 2018	736	8 days	4 km
MERIS	OCI	Apr. 29, 2002	Mar. 15, 2012	506	8 days	4 km
VIIRS	OCI	Jan. 2, 2012	Sep. 21, 2018	299	8 days	4 km

(3) The data in similar region is fused by fixed weight;

(4) In variable regions, the weighted coefficient optimization method is used to fuse the low-frequency data, while the edge-preserving smoothing method is used to fuse the high-frequency data;

(5) Finally, the global sea surface chlorophyll-a concentration is fused by the wavelet

inverse transform.

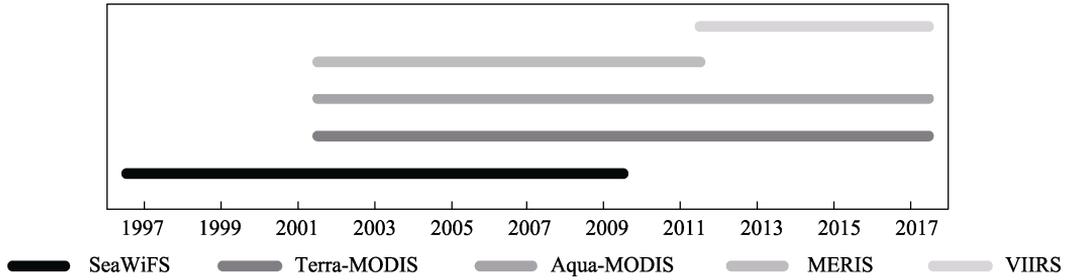


Figure 1 Time range of multi-sensor remote sensing dataset for chlorophyll-a concentration

The maximum value composite method and look-up table method are combined to develop the fusion dataset of global sea surface chlorophyll-a concentration. The specific process includes:

(1) firstly, for each sensor, the dataset of every eight days is fused, and 46 images per year are obtained, thus, five global sea surface chlorophyll-a concentration datasets are generated, i.e., Terra, Aqua, MERIS, SeaWiFS, VIIRS;

(2) the maximum value of each month, season and year in each pixel is selected to replace the final fusion dataset;

(3) for those dataset with only one data source, e.g., 1998–2002 with only SeaWiFS sensor, the dataset is calculated using the lookup table method to make the data consistent with the fused dataset.

Table 3 Statistics for measured data of global sea surface chlorophyll-a concentration

Sensor	Measured points
TERRA	1,576
AQUA	925
SeaWiFS	2,280
MERIS	781
VIIRS	149
Total	5,711

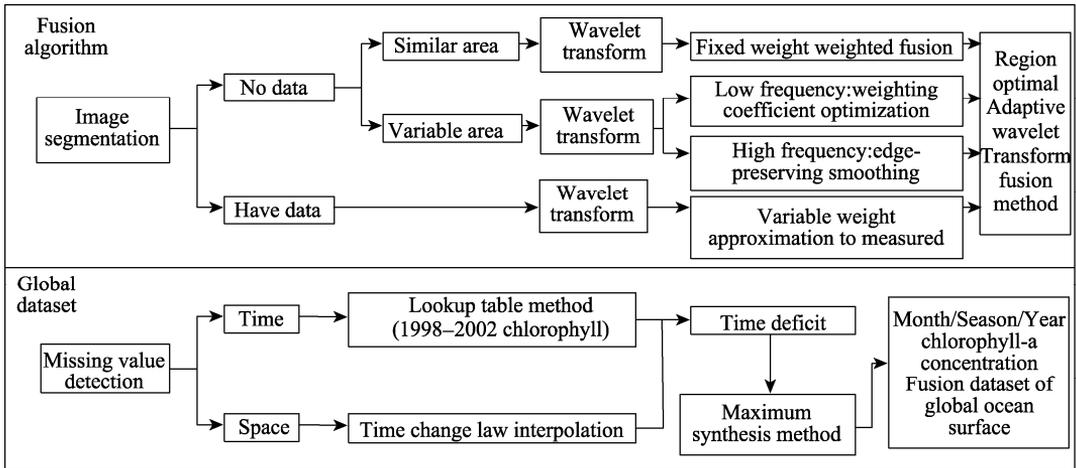


Figure 2 Workflow of the data fusing algorithm

4 Data Results and Validation

4.1 Data Composition

Global fusion dataset of sea surface chlorophyll-a concentration (1998–2018) includes monthly scale dataset, seasonal scale dataset and annual scale dataset. The number and amount of products contained in the dataset are shown in Table 4.

4.2 Data Pre-processing

Data Pre-processing is to establish a unified spatio-temporal dataset, which provides the basis for designing the algorithm data fusion and the algorithm of generating products. The preprocessing includes the data format transformation, the spatiotemporal resampling and the coordinate system projection.

Table 4 Composition of global fusion dataset of sea surface chlorophyll-a concentration

Classification	Number of Products	Data Size
Monthly Global fusion dataset of sea surface Chlorophyll-a Concentration	252	35 GB
Seasonal Global fusion dataset of sea surface Chlorophyll-a Concentration	84	11.8 GB
Annual Global fusion dataset of sea surface Chlorophyll-a Concentration	21	2.9 GB

(1) Data format transformation

The original dataset of Terra-MODIS, Aqua-MODIS, SeaWiFS and VIIRS are all files in NetCDF (suffix name. nc) format, and the dataset of MERIS are files in HDF4. The data format conversion program is written based on ENVI-IDL language and Python language, and the format of remote sensing data of chlorophyll-a concentration on the global sea surface from multiple sensors is converted to TIFF format.

(2) Spatiotemporal resampling

In order to ensure the accuracy and utilization of the dataset, the bilinear interpolation method is used to interpolate the low-resolution data. After interpolation, the spatial resolution of remote sensing data of chlorophyll-a concentration on the global sea surface is resampled to 4 km.

(3) Coordinate system projection

The global sea surface chlorophyll-a concentration remote sensing dataset from multiple sensors were projected into WGS-84 geographic coordinate system.

4.3 Data Results

The datasets includes three temporal scales, there are monthly, seasonal and annual global datasets of sea surface chlorophyll-a concentration. The annual dataset is shown in Figure 3.

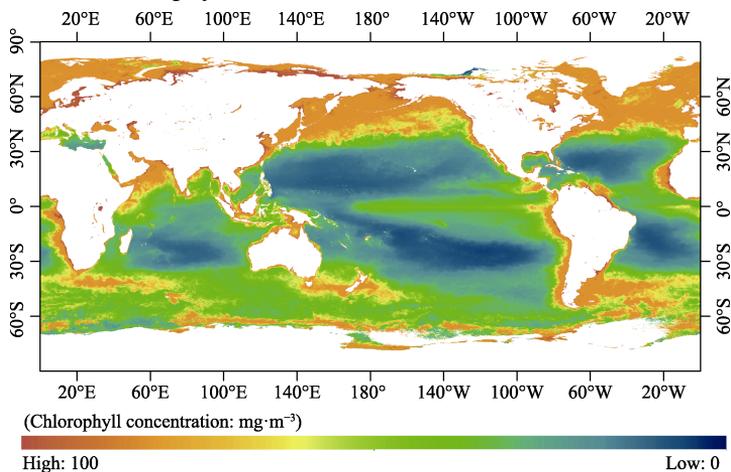


Figure 3 Annual fusion datasets of sea surface chlorophyll-a concentration (2008)

4.4 Data Validation

(1) Comparison between the fusion dataset and the measured data

The comparison between the fusion dataset and the measured data was analyzed. Firstly,

the maximum value and the corresponding coordinate value of each eight days were selected from the measured dataset found from the fusion dataset image. Then calculate the logarithm value with 10 as the bottom, and the chlorophyll-a concentration less than 1 becomes negative. Finally, the correlation between the two data points is calculated. The matching degree of the two data points is high, and the point fitting degree reaches nearly 87%.

(2) Comparison between the fusion dataset and the original dataset

In order to compare and analyze the data values of fusion dataset and the data values of original dataset, two tests are carried out. One test selected from the fusion datasets, Aqua, Terra, MERIS and SeaWIFS from 2005 to 2010, and the other selected from fusion datasets, Aqua, Terra and VIIRS from 2012 to 2016. Firstly, the remote sensing values corresponding to the measured maximum values are respectively calculated, and then the data values of the fusion dataset and multi-source sensor data of the same day are selected, and the correlation analysis is carried out. The results are shown in Figure 4 and Figure 5.

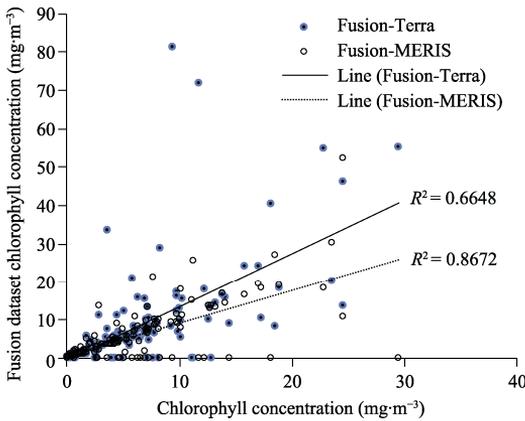


Figure 4 Comparison between the fusion dataset and the real dataset

The first test shows that the MERIS data have the best correlation with the dataset, and the fitting degree reaches nearly 87%. Terra data have the worst fitting degree, and the fitting degree is only 66%. And the second test shows that the fitting degree of Aqua, Terra, VIIRS and other three kinds of data with the fusion dataset is better than that of the first group. The minimum fitting degree is nearly 83%, and the maximum fitting degree is nearly 93%, indicating that Aqua, Terra, VIIRS original datasets have good fitting with the fusion dataset.

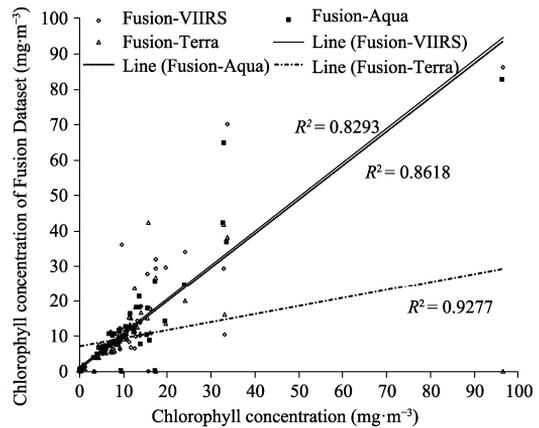


Figure 5 Comparison between the fusion dataset and the original dataset

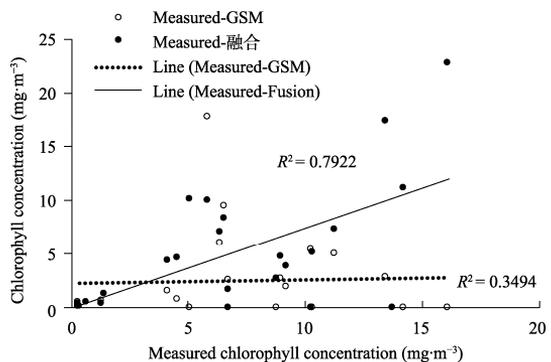


Figure 6 Comparisons between the fusion datasets, GSM and the real dataset

(3) Comparison between the fusion dataset and prevailing datasets

The fusion dataset, the GSM dataset and the real(original) dataset in 2008 are used for comparison. There are a total of 24 matched pairs of points, and the comparison is shown in Figure 6, which shows that the fitting degree of the fusion dataset and the real dataset is 79%, while the fitting degree of GSM and the real dataset is only 35%. The fitting degree of the

fusion dataset and the real dataset is much higher than that of the GSM product, indicating that the fusion dataset has a high quality and a good matching with the real dataset.

5 Discussion and Conclusion

Based on the chlorophyll-a data of five sensors, i.e., SeaWiFS, Terra, Aqua, MERIS and VIIRS, a data fusion algorithm is designed and a fusion dataset is generated. The fusion dataset of global sea surface chlorophyll-a concentration covers from January 1998 to December 2018, with a spatial resolution of 4 km × 4 km, the temporal resolution of monthly, seasonal and yearly. The amount of dataset size is 50.1 GB (19.5 GB after compressed). Based on the real (original) dataset and the GSM dataset of ESA, three validations of the fusion dataset are carried out. The results show that the fitting degree of the fusion dataset is higher than that of GSM dataset. As a large number of observation data is needed to optimize the fusion algorithm and there is a lack of observation data in the opening ocean, the fusion dataset may have much room to improve.

Author Contributions

Li, L. W. is responsible for data validation. Xue, C. J. is responsible for the overall planning and design of datasets. Cui, J. Y. is responsible for dataset fusion algorithm design and algorithm implementation. Fu, Y. X., and Zhang, Y. Y. participated in data download and preprocessing. Xu, Y. F. is responsible for dataset fusion. Li, L. W. and Fu, Y. X. are responsible for data paper writing.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Wang, J. L., Zhang, Y. J., Yang, F., *et al.* The seasonal chlorophyll-a dataset of Poyang Lake, China (2009–2012) [J]. *Journal of Global Change Data & Discovery*, 2017, 1(2): 208–215.
- [2] Li, X. X., Zhang, T. L., Tian, L., *et al.* Merging chlorophyll-a data from multiple ocean color sensors in South China Sea [J]. *Journal of Remote Sensing*, 2015, 19(4): 680–689.
- [3] Shi, Y. N., Zhang, T. L., Shi, L. J., *et al.* Objective analysis for merging multisensory chlorophyll-a data [J]. *Haiyang Xuebao*, 2016(3): 82–87.
- [4] Cui, J. Y., Liu, X. D., Yue, Z. Y., *et al.* Multi-source ocean remote sensing chlorophyll data fusion [J]. *Remote Sensing Information*, 2020, 35(3): 31–36.
- [5] Chen, Y. Z., Wang, X. Q., Wu, B., *et al.* Ocean color data merging based on adaptive weighted averaging [J]. *Remote Sensing Technology and Application*, 2012, 27(3): 333–338.
- [6] Gregg, W. W., Conkright, M. E. Global seasonal climatologies of ocean chlorophyll: blending in situ and satellite data for the coastal zone color scanner era [J]. *Journal of Geophysical Research*, 2001, 106(C2): 2499–2516.
- [7] Qu, L. Q., Guan, L., He, M. X. The global availabilities of SeaWiFS, MODIS and merged chlorophyll-a data [J]. *Periodical of Ocean University of China*, 2006, 36(2): 321–326.
- [8] Chen, Z. Y., Zheng, G. Q., Wang, X. Q., *et al.* Retrieval of chlorophyll a concentration with multi-sensor data by GSM01 merging algorithm [J]. *Journal of Geo-information Science*, 2013, 15(6): 911–917.
- [9] Li, L. W., Fu, Y. X., Xue, C. J., *et al.* Global ocean surface chlorophyll-a concentration fusion 4-km grid dataset (1998–2008) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2021. <https://doi.org/10.3974/geodb.2021.05.05.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2021.05.05.V1>.
- [10] GCdataPR Editorial Office. GCdataPR data sharing policy [OL]. <https://doi.org/10.3974/dp.policy.2014.05> (Updated 2017).
- [11] MERIS Reprocessing Information. <https://oceancolor.gsfc.nasa.gov/cgi/browse.pl>.
- [12] SeaBASS. <https://seabass.gsfc.nasa.gov/>.