Front. Earth Sci. 2024, 18(3): 463–487
https://doi.org/10.1007/s11707-022-1022-1

RESEARCH ARTICLE

Assessing the quality of chlorophyll-a concentration products under multiple spatial and temporal scales

Zheng WANG[1,2], Qun ZENG[3], Shike QIU[1], Chao WANG[1], Tingting SUN[1], Jun DU (✉)[1]

1 Institute of Geographical Science, Henan Academy of Sciences, Zhengzhou 450052, China
2 State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography (Ministry of Natural Resources), Hangzhou 310012, China
3 Editorial Department of Journal of Central China Normal University, Wuhan 430079, China

Abstract The chlorophyll-a concentration data obtained through remote sensing are important for a wide range of scientific concerns. However, cloud cover and limitations of inversion algorithms of chlorophyll-a concentration lead to data loss, which critically limits studying the mechanism of spatial-temporal patterns of chlorophyll-a concentration in response to marine environment changes. If the commonly used operational chlorophyll-a concentration products can offer the best data coverage frequency, highest accuracy, best applicability, and greatest robustness at different scales remains debatable to date. Therefore, in the present study, four commonly used operational multi-sensor multi-algorithm fusion products were compared and subjected to validation based on statistical analysis using the available data measured at multiple spatial and temporal scales. The experimental results revealed that in terms of spatial distribution, the chlorophyll-a concentration products generated by averaging method (Chl1-AV/AVW) and GSM model (Chl1-GSM) presented a relatively high data coverage frequency in Case I water regions and extremely low or no data coverage frequency in the estuarine coastal zone regions and inland water regions, the chlorophyll-a concentration products generated by the Neural Network algorithm (Chl2) presented high data coverage frequency in the estuarine coastal zone Case 2 water regions. The chlorophyll-a concentration products generated by the OC5 algorithm (ChlOC5) presented high data coverage frequency in Case I water regions and the turbid Case II water regions. In terms of absolute precision, the Chl1-AV/AVW and Chl1-GSM chlorophyll-a concentration products performed better in Class I water regions, and the Chl2 product performed well only in Case II estuarine coastal zones, while presenting large errors in absolute precision in the Case I water regions. The ChlOC5 product presented a higher precision in Case I and Case II water regions, with a better and more stable performance in both regions compared to the other products.

Keywords remote sensing, chlorophyll-a concentration, data coverage frequency, accuracy, validation, multiple spatial and temporal scales

Received January 22, 2022; accepted September 13, 2022

E-mail: dujun@igs-has.cn

1 Introduction

Chlorophyll-a is the main pigment responsible for photosynthesis of phytoplankton (Xi et al., 2020). Therefore, the chlorophyll-a concentration at the sea surface represents an important variable when characterizing phytoplankton biomass (Zhang et al., 2018). Moreover, the pattern of changes occurring in marine phytoplankton, the factors influencing these changes, and the response mechanism of phytoplankton toward the changes in the marine environmental factors could be comprehensively understood through a quantitative evaluation of chlorophyll-a concentration in the ocean. In this context, rapidly developing remote sensing technology has provided a novel and irreplaceable approach to monitor and study the chlorophyll-a concentration in the ocean. Ocean color remote sensing offers advantages that are lacking in traditional monitoring approaches. A series of ocean color satellites with high sensitivity and a high signal-to-noise ratio (s/n) have been launched since the late 1990s, including SeaWiFS, MODIS, MERIS, VIIRS, OLCI, etc. These satellites have generated massive amounts of ocean color remote sensing data that can be used to monitor and study marine environmental parameters such as ocean chlorophyll-a concentrations, thereby establishing a solid database foundation for research on marine phytoplankton and the issues that go along with it (Yamada et al., 2004; Liu and Chen, 2014; O'Reilly and Werdell, 2019).

Decades of progress in ocean color remote sensing technology have generated several operational chlorophyll-a concentration data sets/products that are applicable in different scenarios, including the Chl1, Chl2, ChlOC5, and CI products, among others (O'Reilly et al., 2000; Gohin et al., 2002; Doerffer and Schiller, 2007; Hu et al., 2012). It is noteworthy that extensive research has already been conducted on determining the chlorophyll-a concentration in phytoplankton using the remote sensing technology in the following major areas of study: 1) the relationship between increasing chlorophyll-a concentration determined using remote sensing and upwelling (Tang et al., 1999; Wang et al., 2010); 2) the relationship between increasing chlorophyll-a concentration determined using remote sensing and typhoon (Shang et al., 2008; Zhao et al., 2008; Li et al., 2009); 3) the relationship between the chlorophyll-a concentration determined using remote sensing and mesoscale eddies (Chen et al., 2007; Ning et al., 2008; Lin et al., 2010; Lin et al., 2014; Shang et al., 2015); 4) the relationship between increasing chlorophyll-a concentration determined using remote sensing and ocean nutrients (Shan and Hui, 2008; Lin et al., 2009; Grosse et al., 2010; Liu et al., 2010; Shen et al., 2010; Xu et al., 2010; Wang et al., 2012; Kim et al., 2014); 5) the relationship between increasing chlorophyll-a concentration determined using remote sensing and climate change (Isoguchi et al., 2005; Tang et al., 2011); 6) the relationship between increasing chlorophyll-a concentration determined using remote sensing and Asian monsoons (Chen et al., 2006; Yang et al., 2012); 7) the relationship between the variation in the chlorophyll-a concentration determined using remote sensing and variation in the monsoon speed (Zhao et al., 2008; Gai et al., 2012); 8) the relationship between variation in the chlorophyll-a concentration determined using remote sensing and temperature of the sea surface (Wang and Tang, 2010); 9) the relationship between variation in the chlorophyll-a concentration determined using remote sensing and light intensity (Eilers and Peeters, 1988; Lewandowska and Sommer, 2010; Yuan et al., 2011; Winder et al., 2012). The majority of these studies have concentrated on short-term mechanisms underlying the spatial and temporal distribution patterns of chlorophyll-a concentration, as well as the investigation and analysis of typical ocean phenomena using incomplete data.

However, it is reported that at any particular moment, over 66% of the Earth is covered by clouds (Wang et al., 2020). Cloud cover significantly interferes with remote sensing measurements. Moreover, most of the chlorophyll-a concentration inversion algorithms developed thus far are based on blue–green bands. According to the Mie scattering theory, the shorter wavelength bands are influenced to a greater degree by the objects in the atmosphere compared to the longer wavelength bands, which implies that the sensors based on the blue–green bands have a relatively higher sensitivity to all kinds of clouds, fog, haze, etc., compared to the sensors based on other bands (Wang et al., 2020). The factors that influence the application potential of remote sensing measurements include, in addition to cloud coverage, the limitations of the inversion algorithms on which the sensors are based and the limited resolution of the output, which would lead to huge losses in the chlorophyll-a concentration inversion data generated (Wang and Liu, 2014; Huynh et al., 2016; Shropshire et al., 2016).

On the one hand, the chlorophyll-a concentration data determined through remote sensing might include missing details in spatial distribution. On the other hand, irregularities in the missing details in the time series affect the acquisition and application of the time-series data. Fortunately, in the past 20 years, a series of ocean color satellites have been launched under the Earth Observation Program (EOS). These satellites employ operational ocean color remote sensing products that present good consistency and complementarity (Brewin et al., 2014). Moreover, the limitations of the multi-source data collected by these satellites could be overcome through a complementary application of specific analytical methods and fusion algorithms (such as AV, AVW, and GSM) (Maritorena and Siegel, 2005; Maritorena et al., 2010). After applying the fusion algorithms, the data sets would effectively improve data coverage frequency in terms of spatial distribution and present a certain degree of improvement in absolute accuracy and temporal resolution. So far, several sets of time-series operational chlorophyll-a concentration data sets have been generated, such as the Chl1-AV/AVW, Chl1-GSM, Chl2, and ChlOC5 products, which have a wide range of applications (Gohin et al., 2002). However, several questions remain unanswered. What is the data coverage frequency of the commonly used operationalized multi-sensor fusion products on different time scales and different spatial scales? What is the useful data coverage frequency of these fusion products on the daily, 8-day, and monthly scales? Which available operational ocean color chlorophyll-a concentration products presents a higher absolute accuracy? Which of the operational ocean color remote sensing chlorophyll-a concentration products are best suited for Case I and Case II waters? Which of the available operational chlorophyll-a concentration product presents the highest overall effective data coverage frequency, absolute accuracy, and most efficient performance in different scenarios and regions?

The main purpose of this study is to determine which of the four commonly used operational chlorophyll-a concentration products presented the highest available data coverage frequency, accuracy, and optimal robustness on

regional and global scales to support global change and biogeochemical cycle studies. The remainder of this work is organized as follows: Section 2 presents the data sets employed and the methods used in the present study; Section 3 evaluates different operationalized chlorophyll-a concentration products at daily, 8-day, and monthly scales with the average available data coverage frequency over 20 years at the global scale and in the East Asian monsoon region. In addition, evaluations of the average data coverage frequency of different operationalized chlorophyll-a concentration products on the daily, 8-day, and monthly scales in the South China Sea region and its surrounding areas from 1998 to 2018 (time-series scale) are also demonstrated; Section 4 discusses the quantitative evaluation of the absolute accuracy of the different chlorophyll-a concentration products on different scales based on the NOMAD data, Bio Argo data, and other measured chlorophyll-a concentration data; Sections 5 and 6 presents the results of all evaluations and a final discussion on the results.

2 Data and methods

2.1 Ocean color remote sensing chlorophyll-a concentration products

The synthetic chlorophyll-a concentration data measured via remote sensing on daily, 8-day, and monthly scales and subjected to L3-level fusions in the current study were freely obtained from the Global Color website. This website offers products with resolutions of 4 km, 9 km, and 25 km, including MERIS, MODIS, SeaWiFS, VIIRS, and OLCI, and different multi-sensor synthetic products. The spatial coverage of the data available on this website is at a global-scale, and the duration of data collection was from September 1, 1997, to the present. The website also provides chlorophyll-a concentration data obtained using different inversion algorithms; the main ones include Chl1, Chl2, and ChlOC5 products (Table 1). Comparing these algorithms is not possible as each is based on a different sensor, making them applicable to different water types and regions. Therefore, different algorithms should be considered for different water body components and regions.

2.2 Measured chlorophyll-a concentration data used in present study

2.2.1 Bio-Argo data sets

The Bio-Argo buoy's new automated ocean observation platform brings a revolutionary technological leap toward *in situ* observations of ocean optics and biogeochemistry. It is capable of long-term automated observations in different global sea areas and has a high vertical resolution (1 m). The simultaneous measurement of physical and bio-optical parameters on the Bio-Argo buoy makes it particularly suitable for analyzing bio-optical and biogeochemical seasonal processes in local marine areas and studying of coupled eco-physical processes. The Bio-Argo buoy can drift freely in the ocean for 10 days and surface once to transmit signals to the satellite. Bio-Argo buoys measure chlorophyll-a concentration, seawater temperature, salinity, and water depth between the sea surface and a predetermined depth (typically 2000 m) on a 10-day cycle. Bio-Argo data are accurate to 0.1 m for water depth and 0.001°C for temperature. The Bio-Argo chlorophyll-a concentration data used in the present study are the actual measured surface chlorophyll-a concentrations within 5 m of the multiple floats in a time sequence (Fig. 1). A total of 9502 survey Bio-Argo data points matched the remote sensing products from October 2012 to January 2016 (Fig. 1). A few of the chlorophyll-a concentration data from Bio-Argo buoys used in this study were obtained from the Global Marine Argo Scattershot Data set, a Bio-Argo data set comprised of data quality-controlled by the China Argo Real-Time Information Center.

2.2.2 NASA bio-Optical Marine Algorithm Data set (NOMAD)

The NOMAD is a publicly available global high-quality bio-optical data set provided by NASA's Ocean Biology Processing Group (OBPG) for the validation and development of ocean color remote sensing algorithms (Werdell and Bailey, 2005). NOMAD is a high-quality data set with global coverage of measurement data points, including Apparent Optical Properties (AOPs) and Inherent Optical Properties (IOPs) data points of water bodies measured in different sea regions across the world.

Table 1 Parameters of different satellite-inversed chlorophyll-a concentration products

Product	Inversion algorithm	Time coverage	Fusion algorithm	Sensor
Chl1	OC4V5, OC4Me, OC3v5	From September 1997 to date	AV/AVW	MER, MOD, SWF, VIR
Chl1	OC4V5, OC4Me, OC3v5	From September 1997 to date	GSM	MER, MOD, SWF, VIR
ChlOC5	OC5	From September 1997 to date	AVW	MER, MOD, SWF, VIR
Chl2	C2R-Neural Network	From April 2002 to April 2012	AV	MER

Notes: MER, MERIS; MOD, MODIS; SWF, SeaWiFS; VIR, VIIRS.

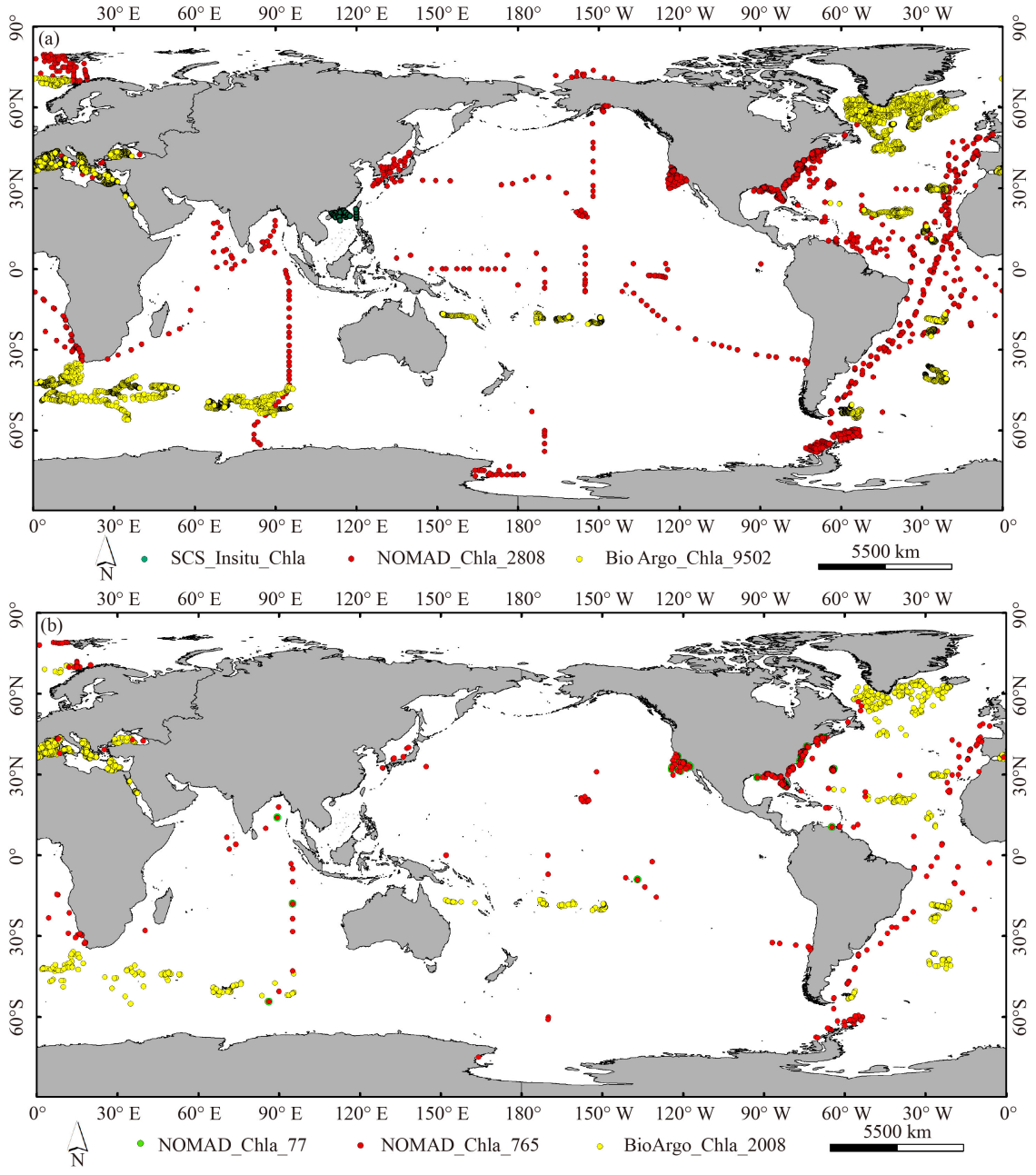

Fig. 1 Distribution of in situ measured chlorophyll-a concentration data for accuracy evaluation in this paper. (a) Distribution of all NOMAD (2808), Bio-Argo (9502), and *in situ* measured (northern South China Sea) data points matched to the remote sensing inversion data (Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5 products). (b) Spatial distribution of measured NOMAD (765) and Bio-Argo (2008) data points under the same criteria after the selection.

The first version of NOMAD data sets was released in September 2005, and the current second version, NOMAD-2a, was released in June 2008. The detailed description and validation of the NOMAD were introduced by Werdell and Bailey (2005). This new data set, NOMAD, includes over 3400 stations of spectral water-leaving radiances, surface irradiances, and diffuse downwelling attenuation coefficients, encompassing chlorophyll- a concentration ranging from 0.012 to 72.12 mg/m^3. This study utilized 2808 matching data points from October 1997 to May 2007 of NOMAD-2a. The spatial distribution of these matched data points is demonstrated in Fig. 1. However, the actual matching data points would be even fewer when the impact of clouds and the algorithms are considered. Figure. 1(b) depict the spatial distribution of the used data points, which were in the same date and same location, matched to the three (without Chl2, *NOMAD_Chl_765*), four (with Chl2, *NOMAD_Chl_77*) chlorophyll-a concentration remote sensing data.

2.2.3 Other data sets of measured chlorophyll-a concentration data

The measured chlorophyll-a concentration data of the South China Sea region were obtained from a previously published report (Cai et al., 2015). These data were collected in the region of the upper water depths at 5 m, 25 m, 50 m, 75 m, 100 m, 125 m, and 150 m. In the present study, the chlorophyll-a concentration data of the 5 m water depth layer were selected as the representative data in the surface layer of the South China Sea. These data were collected during four different seasons from 30 April to 24 May in 2011 (spring), 18 July to 16 August in 2009 (summer), 26 October to 24 November in 2010 (autumn), and 6–30 January in 2010 (winter). These four seasonal voyages collected data at 123 different data collection stations, covering 33 stations during the spring voyage, 41 stations during the summer voyage, 19 stations during the autumn voyage, and 30 stations during the winter voyage. Only the chlorophyll-a concentration data measured at a depth of 5 m were chosen and compared with the data from satellite inversion products to determine the absolute accuracy profile of each product when applied in the South China Sea region. Figure 2 depicts the distribution of these stations visited during each season.

2.3 Data merging

The Glob Color website was employed to generate the chlorophyll-a concentration data using various fusion algorithms and inversion algorithms. The main fusion algorithms used for the L3-level product were simple averaging (AV), weighted averaging (AVW), and the Garver-Siegel-Maritorena (GSM) model. The fusion algorithms used for the Chl1 chlorophyll-a concentration product were AV/AVW and GSM. The Chl1 chlorophyll-a concentration product was a fusion of four main sensors – MERIS, MODIS, SeaWiFS, and VIIRS. The Chl2 product combined the MERIS sensor, with AV as the fusion algorithm. The fusion algorithm used for the ChlOC5 chlorophyll-a concentration product was AVW. The ChlOC5 chlorophyll-a concentration product was a fusion of four main sensors – MERIS, MODIS/Aqua,

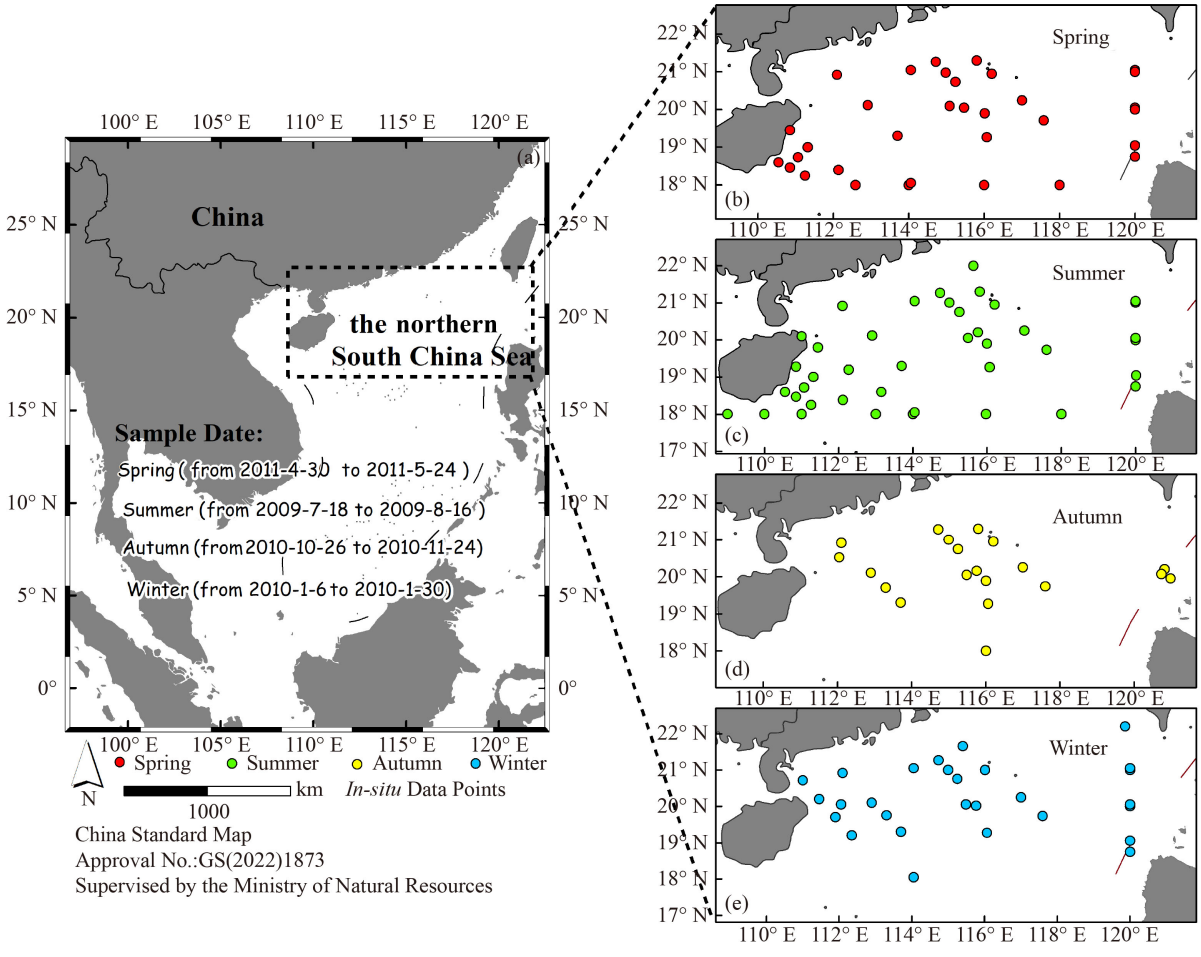

Fig. 2 The measured chlorophyll-a concentration data from the spring, summer, autumn, and winter voyages in the South China Sea. (b–e) *In situ* chlorophyll-a concentration data collected during different seasons in the South China Sea region.

SeaWiFS, and VIIRS. The fusion steps of L3-level products utilized in this paper were performed for the multi-sensors as demonstrated in Fig. 3 (ACRI-ST GlobColour Team, 2017).

2.4 Method of data accuracy validation

Based on the *in situ* measurements of the chlorophyll-a concentration data from Bio-Argo, NOMAD, and the South China Sea voyages, the correlation coefficient (R), the root mean square error (RMSE), and the mean absolute deviation (MAD) were used to compare and evaluate the performance of the four chlorophyll-a concentration estimation products (Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5) in different regions, waters, and periods. The aim was to identify which of the four remote sensing chlorophyll-a concentration products had higher accuracy, better stability, and stronger applicability. The calculation formulas for RMSE and MAD are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i^{\mathrm{estimated}} - x_i^{\mathrm{measured}}\right)^2}{n}}, \tag{1}$$

$$\mathrm{MAD} = \frac{\sum_{i=1}^{n}\left|x_i^{\mathrm{estimated}} - x_i^{\mathrm{measured}}\right|}{n}, \tag{2}$$

where, $x_i^{\mathrm{estimated}}$ denotes the chlorophyll-a concentration from the remote sensing inversion, x_i^{measured} denotes the *in situ* measurement of the chlorophyll-a concentration, and n denotes the number of matching data points.

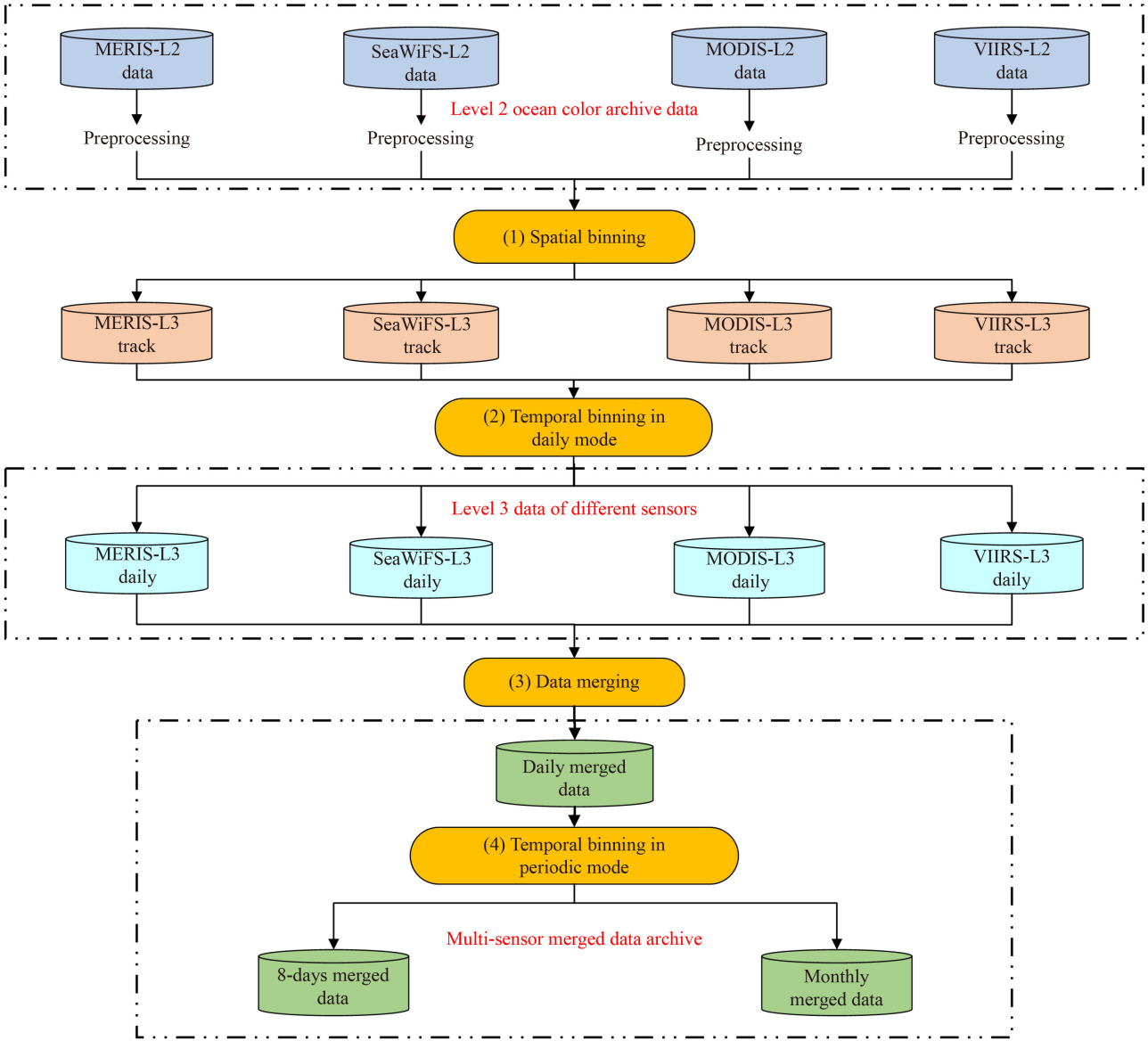

Fig. 3 Flow chart of the multi-sensor fusion of global chlorophyll-a concentration data.

3 Evaluation of average coverage of the remote sensing products for chlorophyll-a concentration at spatial scales

3.1 The coverage of remote sensing products for chlorophyll-a concentration at the global scale

In the present study, the data coverage frequency of four different remote sensing inversed chlorophyll-a concentration products was validated at the global scale. Chl1-AVW, Chl1-GSM, and Chl-OC5 products include data from 1998 to 2018, while Chl2 include data between 2002 and 2012. Four products were assessed in daily, 8-day, and monthly average.

As seen in Fig. 4, except for an average data coverage frequency of approximately 20%–30% in the subtropical waters of the southern and northern hemispheres, the daily average data coverage frequency of the Chl1-GSM (a, c, e) and Chl1-AVW (b, d, f) products was generally in the range of 10%–20% from 1998 to 2018 (a, b). The 8-d time scale average spatial coverage of the Chl1-GSM and Chl1-AVW chlorophyll-a concentration products was approximately 40% in the equatorial waters and approxi-

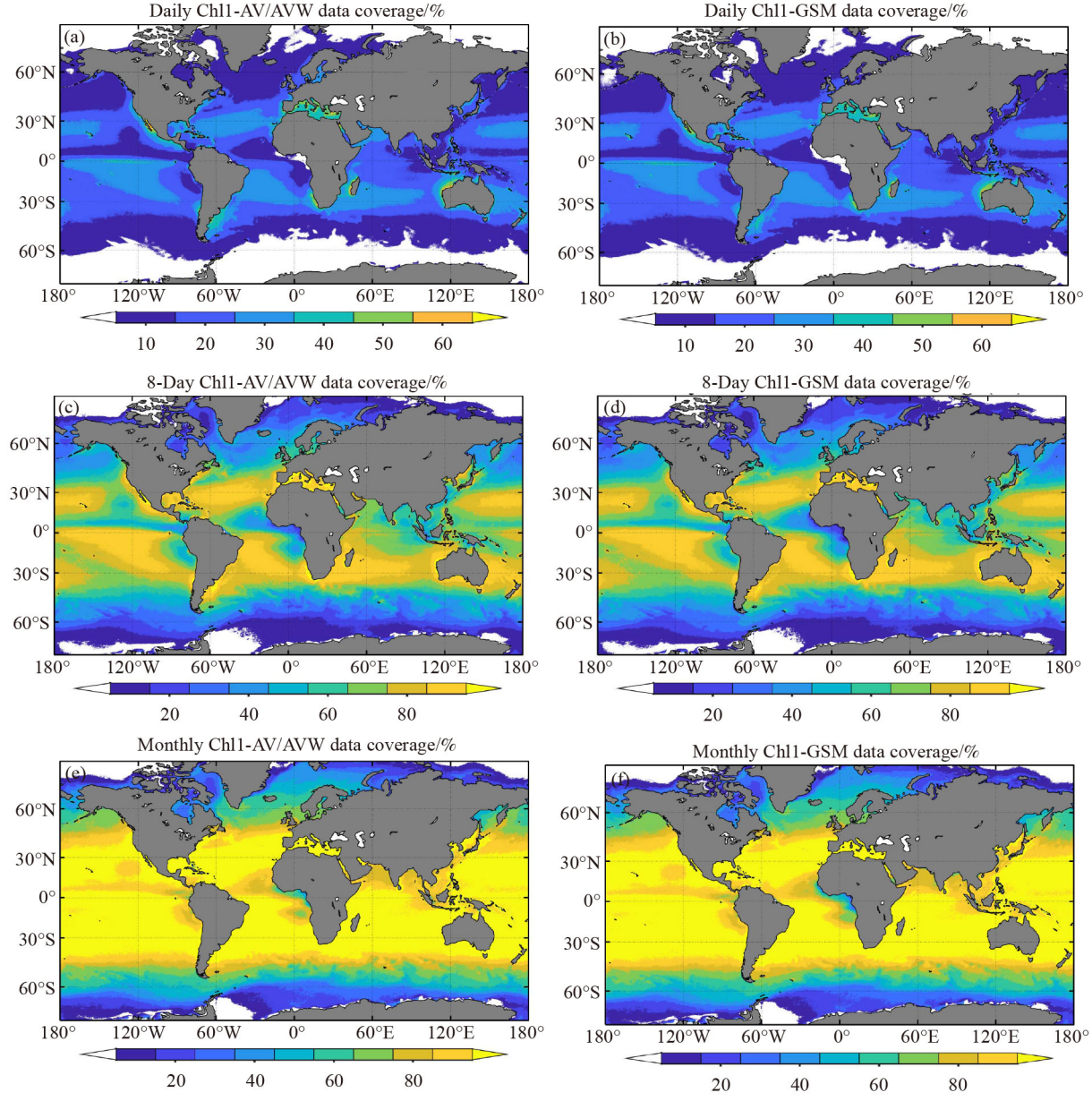

Fig. 4 Average global data coverage frequency of Chl1 product at different time scales. (a)–(b) Daily average Chl1 chlorophyll-a concentration data coverage frequency between the years 1998 and 2018; (c)–(d) 8-d time scale average Chl1 chlorophyll-a concentration data coverage frequency between the years 1998 and 2018; (e)–(f) Monthly average Chl1 chlorophyll-a concentration data coverage frequency between the years 1998 and 2018.

mately 70%–80% above in the subtropical regions. With an average daily coverage of approximately 20%–30%, the data coverage frequency near the latitudes of 60 degrees north and south of the equator was relatively low. In terms of the monthly data, the data coverage frequency at around 60 degrees north and south of the equator was approximately 50%–70%. Because of the comparatively high cloud coverage, the coverage of the chlorophyll-a concentration data in the regions close to the equator, which have a rainy tropical climate, was approximately 60%–80% between 1998 and 2018, slightly lower than that of the surrounding regions. In the subtropical seas, this rate was in the range of 80%–90%. Therefore, at different time scales from 1998 to 2018, the Chl1-GSM and Chl1-AVW chlorophyll-a concentration estimation products were consistent in their spatial coverage.

The Chl2 product depicted in Fig. 5 covers the data from 2002 to 2012, i.e., the in-orbit operation period of the MEIRS sensor. The ChlOC5 product covers the data from September 1997 to date. The present study, selected spatial averages at different time scales for the 21 years between 1998 and 2018. The daily chlorophyll-a concentration data from the Chl2 product had approximately 5%–15% coverage in most of the regions worldwide, with a maximum data coverage frequency of approximately 25% (Fig. 5(a)). The 8-day average data coverage frequency was approximately 50%–60% (Fig. 5(c)),

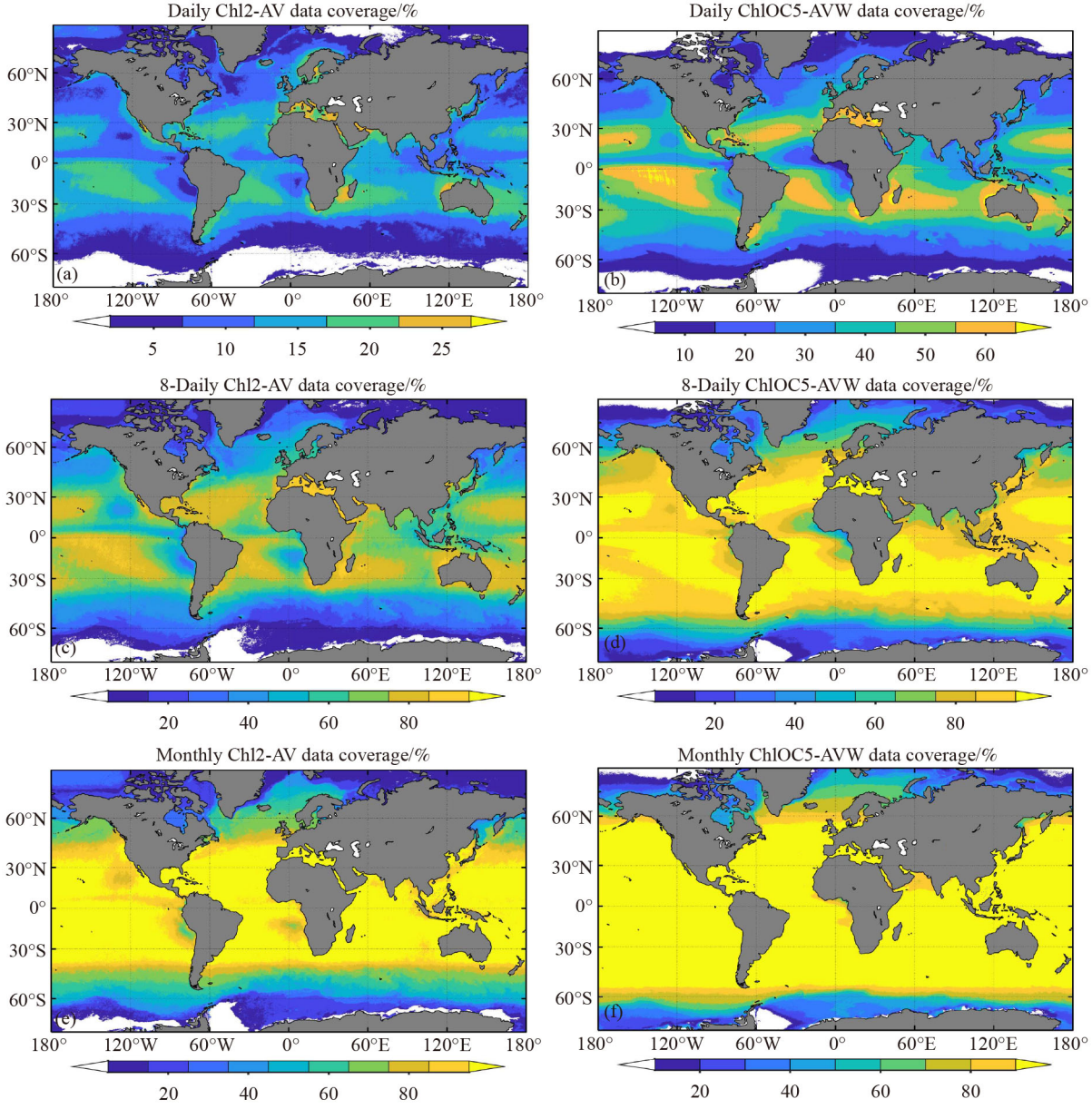

Fig. 5 The average global spatial coverage of Chl2 and ChlOC5 products at different time scales. Chlorophyll-a concentration data coverage frequency of the Chl2 (a, c, e) and ChlOC5 (b, d, f) products at daily, 8-d, and monthly time scales, respectively.

while the monthly data coverage frequency was approximately 70%–80%. Unlike the Chl1 and Chl2 products, the ChlOC5 product presented a significantly better daily data coverage frequency of approximately 30% in the tropical rain forest regions near the equator, 55% in the subtropical waters, and 20%–30% at 60 degrees latitude north and south of the equator (Fig. 5(b)). The 8-d average spatial coverage of the ChlOC5 product was approximately 70% or even beyond 80% (excluding the higher latitudes). The ChlOC5 product's monthly average data coverage frequency in space was approximately 90%, which remained at approximately 80% even near the tropical rainforest regions close to the equator (Fig. 5(f)).

Comparing the global daily, 8-d, and monthly data of the products from 1998 to 2018 shows that the effective data coverage is diverse. The Chl2 product from a single sensor delivered the lowest coverage in any given region, while the fused Chl1-GSM and Chl1-AVW products from multiple sensors presented a slightly higher spatial coverage. The ChlOC5 product presented a comparatively high data coverage frequency in most regions.

3.2 Evaluation of the coverage of chlorophyll-a concentration estimation products in the South China Sea and its surrounding waters

Statistical studies and data processing were performed for multiple products to validate the data coverage frequency provided by the fused products at different temporal scales using different algorithms in the South China Sea and its surrounding waters.

As depicted in Fig. 6, the data coverage frequency of the Chl1-AVW (b, d, f) and Chl1-GSM (a, c, e) products was consistent throughout, although the coverage of the former was slightly higher. In the South China Sea, the daily data coverage frequency was approximately 10%,

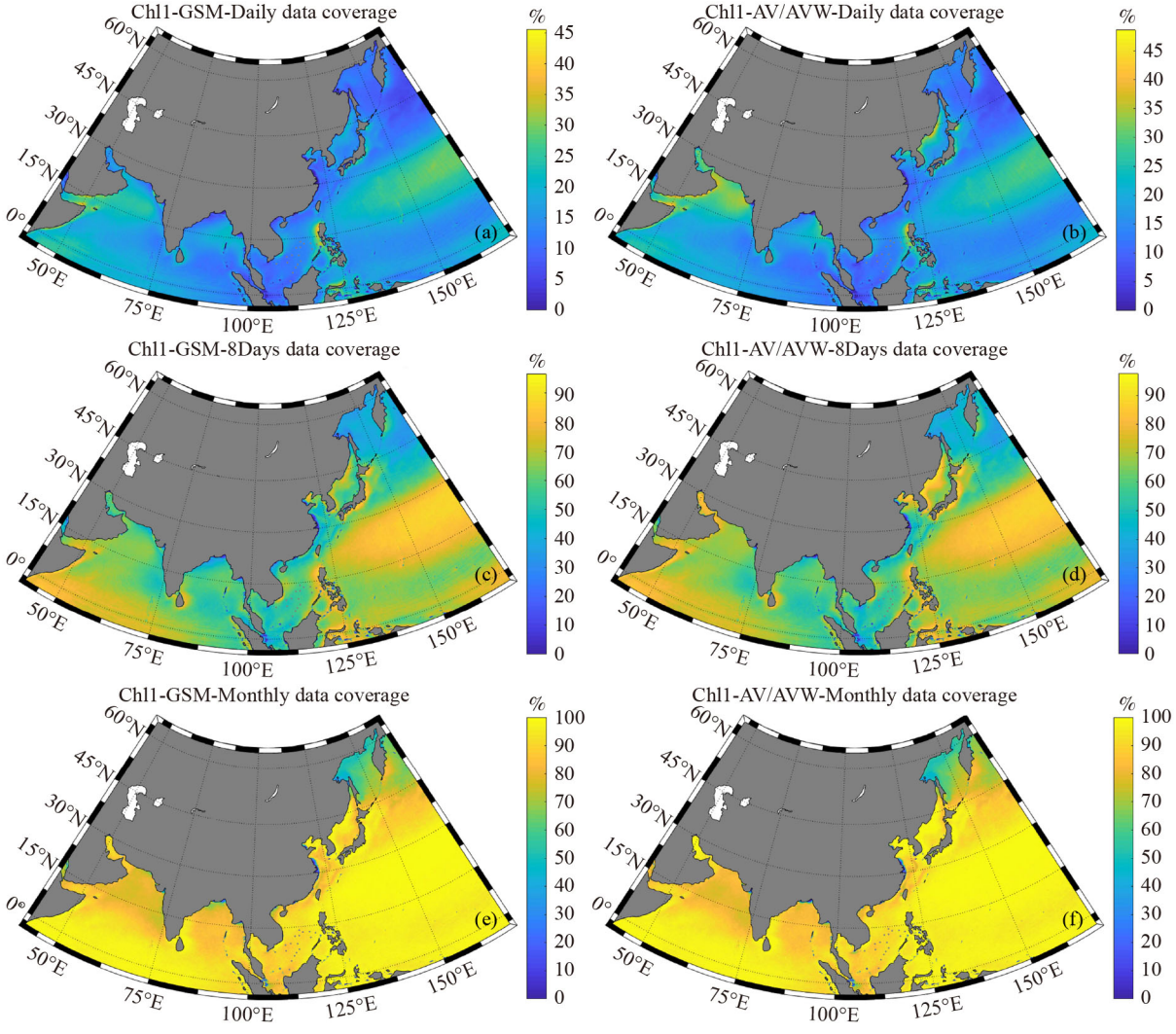

Fig. 6 The average data coverage frequency of Chl1 product in the South China Sea and its surrounding waters at different time scales. Panels (a, c e) and panels (b, d, f) present the chlorophyll-a concentration data coverage frequency of the Chl1-GSM and Chl1-AVW products at daily, 8-d, and monthly time scales, respectively.

similar to that in the East China Sea, while lower than that in the Bay of Bengal, the Arabian Sea, and the north-west Pacific Ocean (Figs. 6(a) and 6(b)). The 8-day data coverage frequency in the South China Sea was slightly higher to the west of Luzon, while most other areas in the South China Sea had coverage of approximately 40%–50%, which was similar to that in the Bay of Bengal and the East China Sea, slightly lower than that in the Arabian Sea, and significantly lower than that in the subtropical regions of the north-west Pacific Ocean. The monthly data coverage frequency in the South China Sea was mostly in the range of 70%–80%, which was at the same level as that in the East China Sea, the Bay of Bengal, and the Arabian Sea, but lower than that in the subtropical regions of the north-west Pacific Ocean.

As depicted in Fig. 7, the single-sensor Chl2 product (Figs. 7(a), 7(c), 7(e)) had a daily spatial coverage of approximately 5%–10% in the South China Sea, which was smaller than that in the Bay of Bengal, the Arabian Sea, the East China Sea, and the north-west Pacific Ocean. The 8-day data coverage frequency in the South China Sea was also evidently smaller (approximately 40%) than in the surrounding waters. On the contrary, the monthly Chl2 products data coverage frequency in the South China Sea exhibited little difference in the adjacent regions. The ChlOC5 product (Figs. 7(b), 7(d), 7(f)) fused from multiple sensors presented a daily chlorophyll-a concentration data coverage frequency of 20%–30% in the South China Sea. This is a value higher than that in the East China Sea, at the same level as that in the Bay of Bengal, and slightly lower than that in the Arabian Sea (Fig. 7(b)). This product's 8-day average coverage data in the South China Sea was approximately 70%–80% approximately higher than in the East China Sea, the Bay of Bengal, and the Arabian Sea (Fig. 7(d)). The monthly data coverage frequency of the ChlOC5 product in the

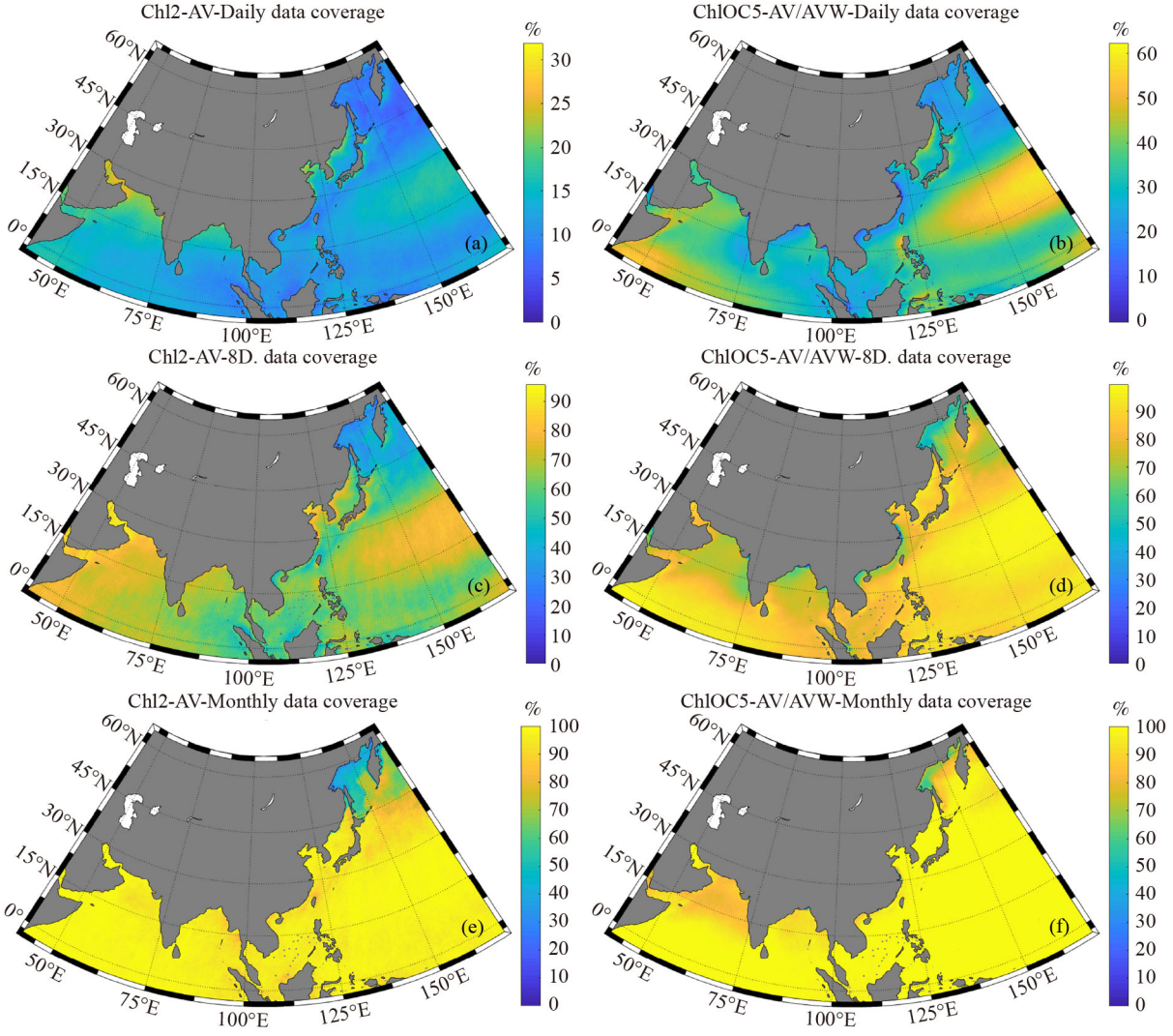

Fig. 7 Data coverage frequency of the Chl2 and ChlOC5 products in the South China Sea and its surrounding regions at different time scales. Panels (a, c, e) and panels (b, d, f) present the average data coverage frequency at daily, 8-d, and monthly scales for the Chl2 and ChlOC5 products.

South China Sea was approximately 90%, which was similar to that in the East China Sea but higher than that in the Bay of Bengal and the Arabian Sea.

The data coverage frequency of different regions, time scales, and algorithms presented in Figs. 6 and 7 reveal that the data coverage frequency of the single-sensor Chl2 product was relatively low on daily, 8-d, and monthly time scales. The results for the Chl1-GSM, Chl1-AVW, and ChlOC5 products demonstrated that the fusion of multiple sensors could effectively enhance the coverage of the available data. In addition, the ChlOC5 product could deliver significantly improved data coverage frequency compared to Chl1-AVW and Chl1-GSM, confirming that the data coverage frequency can be effectively increased using an improved algorithm. Overall, the ChlOC5 product presented the highest data coverage frequency, followed by Chl1-AVW and Chl1-GSM; the Chl2 product presented the lowest data coverage frequency. In particular, ChlOC5 demonstrated the most evident superiority in the 8-day average coverage.

3.3 Evaluation and comparison of the chlorophyll-a concentration products in the South China Sea at different temporal scales

To further investigate the performance of different products in terms of data coverage frequency in the South China Sea at different time scales, a statistical study of data coverage frequency was performed for multiple cases, including the 1998–2018 data from the Chl1-GSM, Ch1-AVW, and ChlOC5 products, the 2003–2011 data from the Chl2 product at the 0°–23.5°N, 99°E–122°E in the South China Sea region, and the daily, 8-day, and monthly data coverage frequency over the 21 years for different algorithm products. Figures 8–10 present (a) the coverages of the Chl1-AVW product at different time scales, (b) the results for the Chl1-GSM product at different time scales, (c) the results for the Chl2 product at different time scales, and (d) the results for the ChlOC5 product at different time scales.

As visible in Figs. 8(a) and 8(b), the temporal patterns of the daily data coverage frequency were basically the same in the South China Sea. A relatively low data coverage frequency of approximately 10% was demonstrated consistently before May 2002, after which the data coverage frequency was significantly enhanced and reached a value of approximately 20%. The reason for this was that, previously, the operational chlorophyll-a concentration data were solely provided by the SeaWiFS until the MERIS satellite data were added in May 2002, thereby improving the data coverage frequency after this time. The 2003 to 2010 period experienced a comparatively high coverage of effective data in the South China Sea region. In addition to the sensor data from SeaWiFS and MERIS, MODIS also provided operational data, and this fusion from multiple sensors significantly enhanced data coverage frequency. Another inflection point occurred in December 2010 when the retiring of the SeaWiFS sensor led to a decrease in the data coverage frequency. In 2011, the data coverage frequency in the South China Sea was slightly lower, which was further reduced in 2012 due to the loss of MERIS. The operational running of VIIRS ameliorated this reduction in the available data in 2012, which ensured data continuity. As seen in Fig. 8(c), the Chl2 product provided a generally stable data coverage frequency of approximately 10% from 2003 to 2011. This was because Chl2 is a single-sensor product based on MERIS. It is noteworthy that, as depicted in Fig. 8(d), for the ChlOC5 product, the single-sensor data coverage frequency was 18% before May 2002, which increased to approximately 33% until December 2010. In 2011, the data coverage frequency dropped to 22%, although after 2012, this rate increased to an average of 34%. Overall, as depicted in Fig. 8, the single-sensor Chl1-GSM, Chl1-AVW, and Chl2 products presented a data coverage frequency of approximately 10% in the South China Sea, while ChlOC5 had a data coverage frequency of 18%. The multi-sensor fused Chl1-GSM and Chl1-AVW products, on the other hand, demonstrated a data coverage frequency of 20%, while the daily multi-sensor ChlOC5 product had a data coverage frequency of approximately 33%. As a result, one may conclude that the ChlOC5 algorithm played a significant role in increasing the frequency of data coverage at daily time scales.

Figure 9(a) presents the temporal and spatial data coverage similar to Fig. 9(b). Specifically, before May 2002, the data coverage frequency was approximately 40% from the single-sensor product, which increased to 67% in 2010 due to the inclusion of multi-sensor fusion in May 2002. The average data coverage frequency for the whole period was approximately 60%, as depicted in the figure. The 9-year average data coverage frequency of the single-sensor Chl2 product was close to 55% (Fig. 9(c)). Before May 2002, the 8-day data coverage frequency of the single-sensor ChlOC5 product was approximately 70%, which, by 2010, improved to approximately 85%. During the year 1998–2018, the average data coverage frequency of ChlOC5 was 80% (Fig. 9(d)). Overall, at the 8-d time scale, the data coverage frequency in the South China Sea region over the 21 years was best with the ChlOC5 product, significantly better than that of the Chl1-GSM, Chl1-AVW, and Chl2 products.

As depicted in Fig. 10, the monthly data coverage frequency of the Chl1-GSM and Chl1-AVW products was approximately 75% before May 2002, noting that it was from a single sensor. This increased to approximately 95% by 2010 upon the inclusion of multi-sensor fusion in May 2002. The monthly data coverage frequency from 1998 to 2018 was approximately 90% (Figs. 10(a) and 10(b)). The monthly data coverage frequency of the Chl2

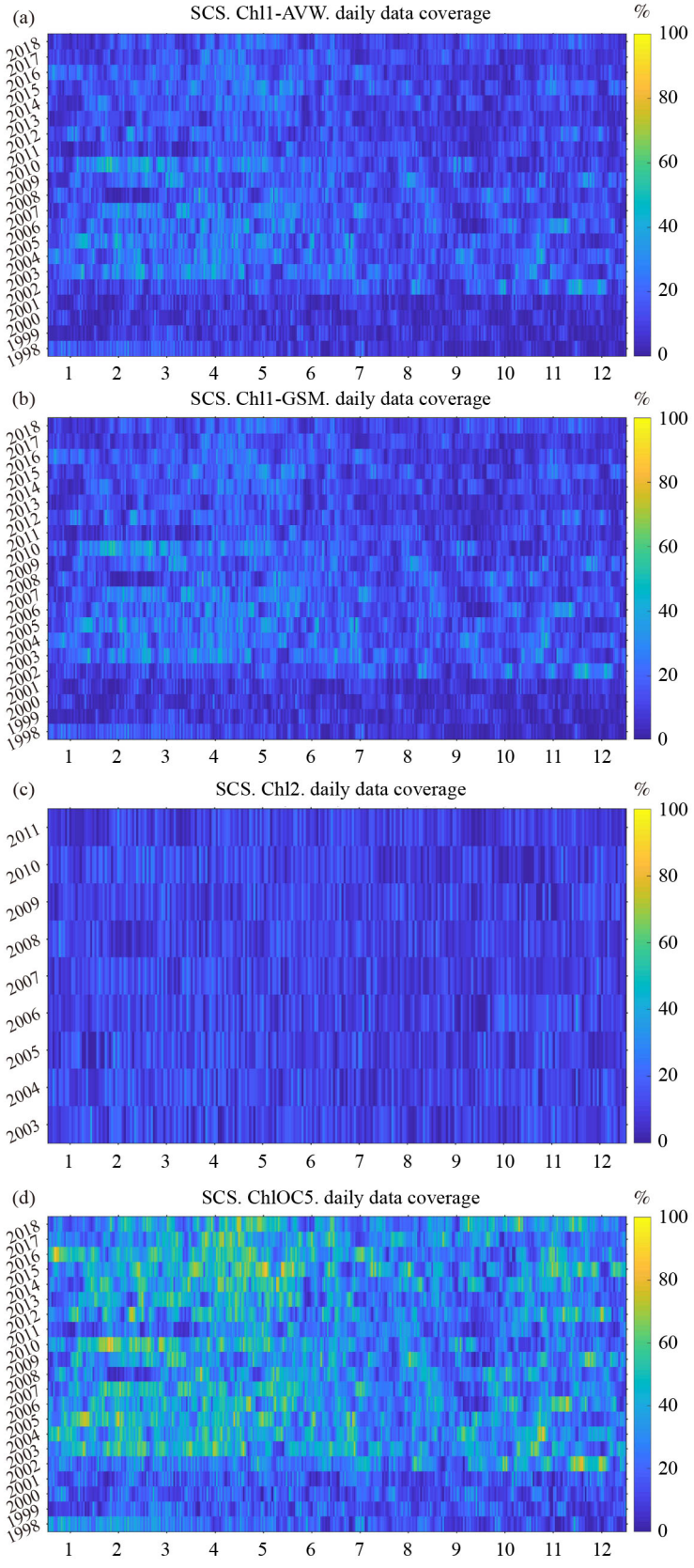

Fig. 8 Daily chlorophyll-a concentration data coverage frequency for different chlorophyll-a concentration inverse algorithm products over 21 years. Panels (a–d) present the regional average daily available chlorophyll-a concentration data obtained through inverting the Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5 products.

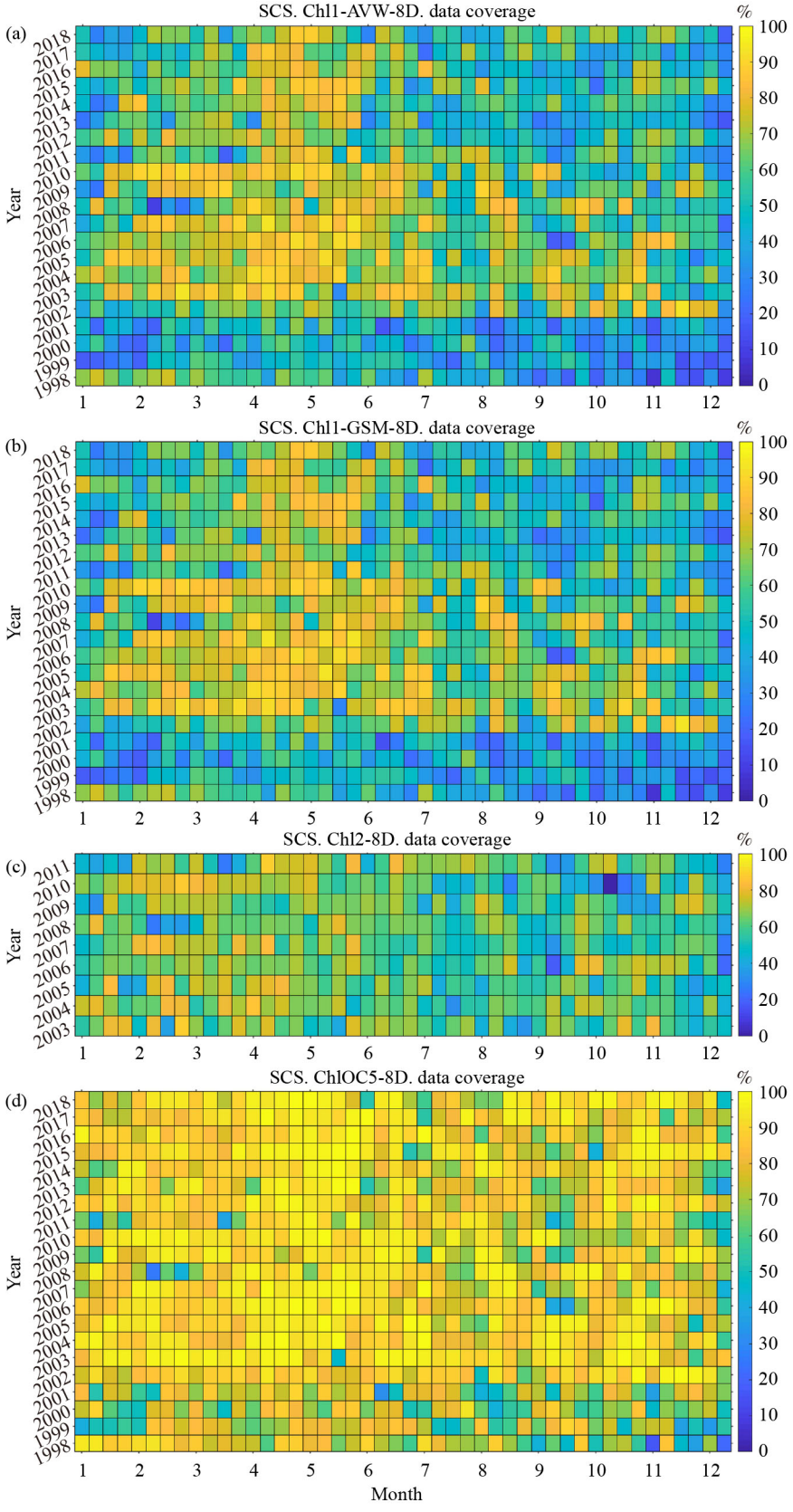

Fig. 9 Regionally averaged 8-d scale data coverage frequency over 21 years for different algorithms. Panels (a–d) present the average regional 8-d available chlorophyll-a concentration data through the inversion of the Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5 products, respectively.

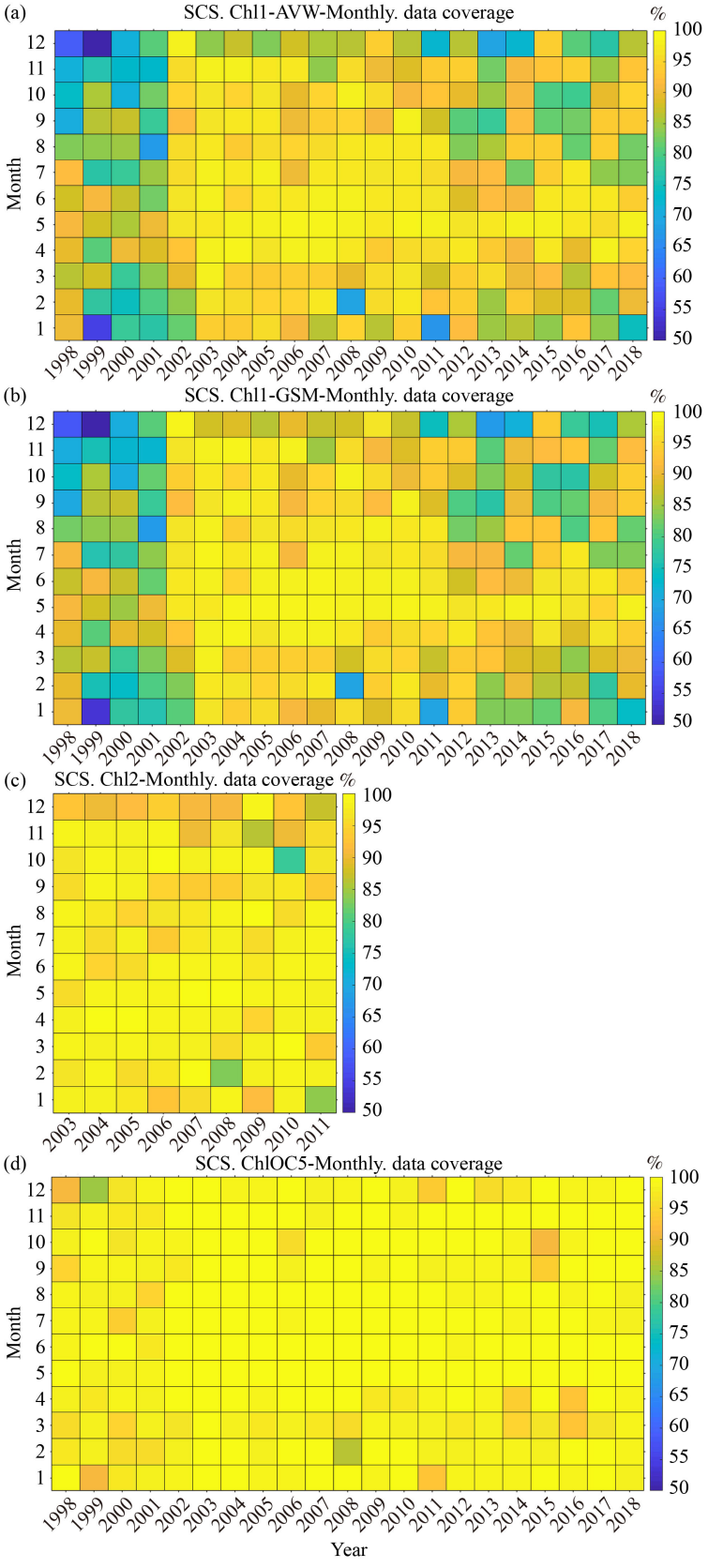

Fig. 10 The long-time-scale monthly data coverage frequency for the different algorithms. Panels (a–d) present the regionally averaged monthly available chlorophyll-a concentration data through the inversion of the Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5 products, respectively.

product was approximately 90%. The monthly data coverage frequency of the ChlOC5 product remained higher than 90% before and after May 2002, with an average of approximately 95%. Therefore, in terms of the long-time-scale monthly data coverage frequency in the South China Sea, ChlOC5 outperformed the Chl1-GSM, Chl1-AVW, and Chl2 products.

In summary, Figs. 8–10 revealed that the ChlOC5 product could deliver a higher data coverage frequency than the other products at daily, 8-d, and monthly time scales, both for single-sensor and multi sensors fused products. Moreover, the ChlOC5 product presented the greatest advantage compared to the Chl1-GSM, Chl1-AVW, and Chl2 products in terms of the 8-day average, exhibiting an approximate improvement of 20%–25%.

4 Accuracy evaluation of chlorophyll-a concentration products on different scales

To assess and compare the absolute accuracy of the different algorithm products, the measured data from several sources were used in the present study, including the surface chlorophyll-a concentration data retrieved globally from Bio-Argo, the chlorophyll-a concentration data from the NASA bio-Optical Marine Algorithm Data set (NOMAD) (Fig. 1), and the surface chlorophyll-a concentration data from the four South China Sea voyages conducted in spring, summer, autumn, and winter, respectively (Fig. 2).

4.1 Comparison and validation of the chlorophyll-a concentration values at a global scale

4.1.1 Comparison of all matching points

Bio-Argo commenced its data retrieval on October 26, 2012, while the Chl2 product ceased data acquisition in April 2012 due to a fault in the MERIS sensor, leading to matching points between the Bio-Argo *in situ* data and the Chl2 product. Therefore, Figure 11 presents only the absolute accuracies of the survey points that could be matched between the Chl1-AVW, Chl1-GSM, and ChlOC5 products and the Bio-Argo data. The three products yielded a total of 9502 survey points matching in time with the Bio-Argo data. Influenced by the algorithm and cloud coverage, the products present varied matching data points in practice. A larger number of matching points reflects a higher data coverage frequency of the algorithm, a closer relationship with the *in situ* Bio-Argo data, a higher inversion accuracy, and a more accurate algorithm. The Chl1-AVW product depicted in Fig. 11(a) presented 2167 matching points in total, the Chl1-GSM product depicted in Fig. 11(b) presented 2011 matching points, and the ChlOC5 product depicted in Fig. 11(c) presented the highest number of matching survey points of 3931. When all matching points were compared to actual measured data, the absolute accuracy for Chl1-AVW, Chl1-GSM, and ChlOC5 was 0.80, 0.84, and 0.79, indicating insignificant differences. Overall, ChlOC5 presented the maximum number of matching

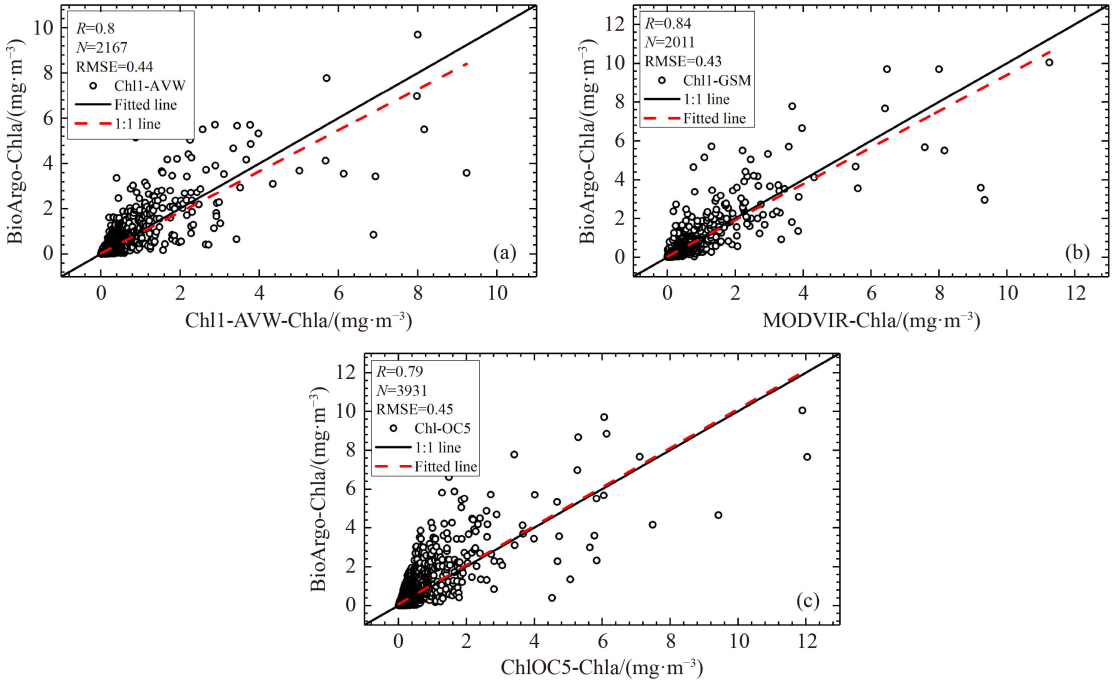

Fig. 11 Comparative evaluation of the chlorophyll-a concentration data retrieved from Bio-Argo and satellite remote sensing. Panels (a–c) compare the Bio-Argo *in situ* chlorophyll-a concentration data within 5 m of depth, and the satellite observed chlorophyll-a concentration data of Chl1-AVW, Chl1-GSM, and ChlOC5, respectively.

points, while the Chl1-GSM product delivered the highest accuracy.

The Bio-Argo data were mostly accessible after 2010. Therefore, for evaluating the inversed chlorophyll-a data between 1997 and 2010, the NOMAD data set was used in the present study. Between the years 1997 and 2008, there were 2808 measuring points in the NOMAD data, matching with Chl1-AVW (Fig. 12(a)), Chl1-GSM (Fig. 12(b)), Chl2 (Fig. 12(c)), and ChlOC5 (Fig. 12(d)) products. A higher number of matching points corresponds to a higher frequency of data coverage. Furthermore, a higher correlation coefficient between the inversed and actual data indicates a better outcome for the satellite remote sensing inversion algorithm.

As visible in Fig. 12, the Chl1-AVW and Chl1-GSM products presented 874 and 763 matching data points, respectively, with the NOMAD data. As for the Chl2 product, which was limited by the time range, only 122 *in situ* matching data points were obtained. ChlOC5, on the other hand, presented 1293 matching points. The correlation coefficients between products of Chl1-AVW, Chl1-GSM, Chl2, ChlOC5, and NOMAD data set were 0.71, 0.73, 0.7, and 0.78, respectively. As depicted in Fig. 12, the ChlOC5 product yielded the highest number of matching data points and the largest correlation coefficient. The scatter plots of all the survey points revealed that ChlOC5 had most of its points close to the 1:1 line, and the performance of Chl1-AVW was inferior to that of ChlOC5 but better than that of Chl1-GSM and Chl2 products. In general, whether comparing *in situ* Bio-Argo data or actual NOMAD measurements, ChlOC5 produced more matching data points than Chl1-AVW, Chl1-GSM, and Chl2 products. Furthermore, the absolute accuracy of ChlOC5 products was generally satisfactory.

4.1.2 Comparison and validation of all matching points under the same criteria

Figures 11 and 12 present the validation results of the number of matching data points of different products and an evaluation of the absolute accuracy based on these matching points. However, the number of data points was not consistent, lowering their comparability and the significance of this comparison. Therefore, all the data points during the same interval were obtained for the different products, as presented in Figs. 13 to 15. In other words, the data to be evaluated were compared with a unified standard, thereby validating the performance of the different products.

Figure 1(b) presents all the matching points of the satellite remote sensing inversion products and Bio-Argo data with the same time mark, longitude, and latitude. The 2008 survey points fulfilling these criteria were then used to evaluate the data from different products. It was revealed that the absolute accuracy of Chl1-AVW, Chl1-GSM, and ChlOC5 changed from 0.8, 0.84, and 0.79 (Fig. 11) before the standardization to 0.73, 0.85, and 0.84 (Fig. 13) after standardization, respectively. The variation in the absolute accuracy of the Chl1-GSM product was not obvious, while the Chl1-AVW and

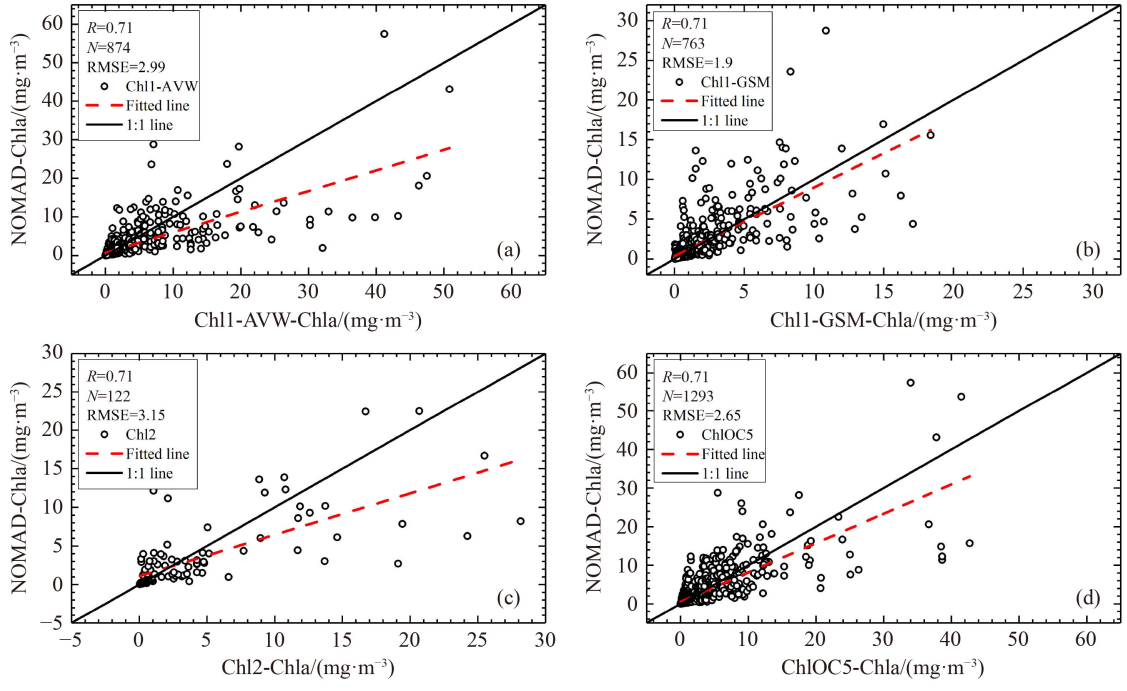

Fig. 12 NOMAD's comparative evaluation of the different satellite inversion products. Panels (a–d) compare the NOMAD *in situ* chlorophyll-a concentration data and the satellite observed chlorophyll-a concentration data of Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5 products, respectively.

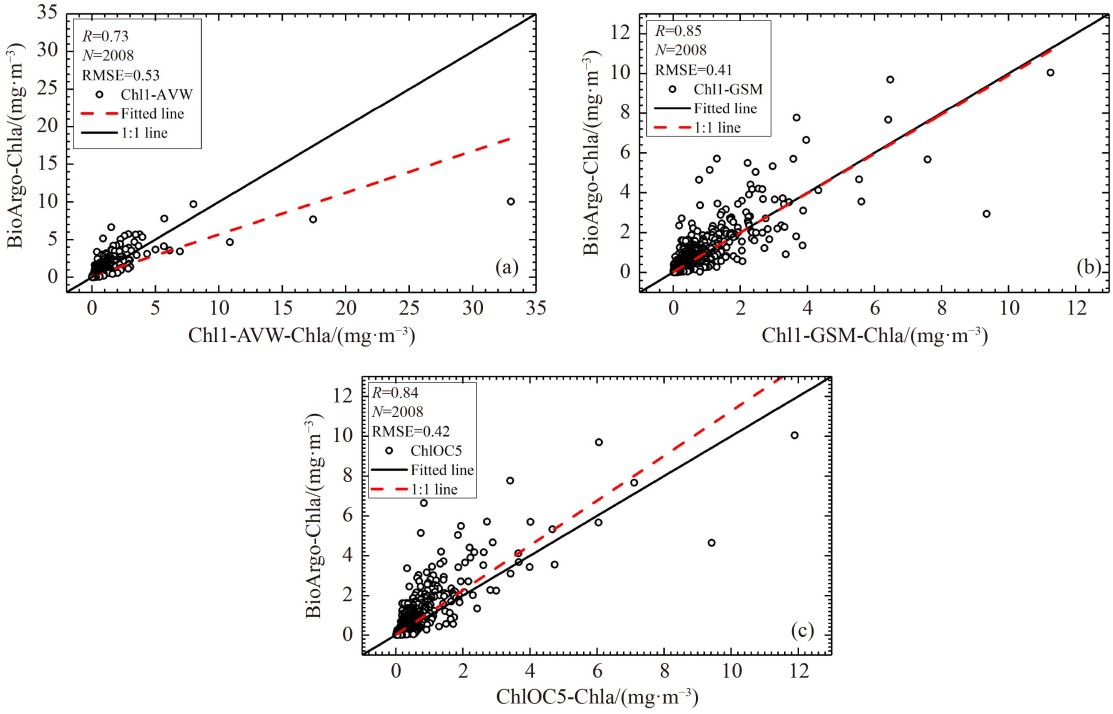

Fig. 13 The Bio-Argo data set compares the different satellite inversion products after standardization. Panels (a–c) compare the Bio-Argo *in situ* chlorophyll-a concentration data within 5 m of depth and the satellite-observed chlorophyll-a concentration data of Chl1-AVW, Chl1-GSM, and ChlOC5, respectively, with the same sample point number and location.

ChlOC5 products demonstrated significant changes; the former exhibited reduced accuracy, while the latter exhibited improved absolute accuracy.

The points with available data from the same day and location were identified for the four chlorophyll-a concentration products. Their absolute accuracy was compared to assess their performance in various locations. As depicted in Figs. 14 and 1(b), the four products and the NOMAD data presented 77 matching points distributed mostly in the coastal waters, and the corresponding absolute accuracies for Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5 products at these 77 points were 0.71, 0.62, 0.73 and 0.85, respectively. It is noteworthy that the Chl2 products achieved higher absolute accuracy compared to Chl1-GSM or Chl1-AVW as it was specifically aimed at Case II waters near the coast. However, the accuracy of Chl2 products (0.73) was much smaller than that of ChlOC5 (0.85) products. Therefore, the ChlOC5 products inversion product was more applicable to the coastal Case II waters. Figure 14 presents the performance of different products in coastal regions, although their performance in the regions containing both Case I and Case II waters remains to be determined. As validated, the algorithm used by the Chl2 product was unsuitable for Case I waters.

Figures 15 and 1(b) depict the matching points with the same time, longitude, and latitude between NOMAD and the Chl1-AVW, Chl1-GSM, and ChlOC5 products (excluding the Chl2 data). There are 765 points in total (Fig. 1(b)), including Case II waters in the coastal areas, Case I waters in the open oceans, oligotrophic waters in the ocean, and eutrophic waters at high latitudes. These representative data were diverse, with abundant samples. The correlation coefficients of the absolute accuracy between the NOMAD and the Chl1-AVW, Chl1-GSM, and ChlOC5 products were 0.72, 0.67, and 0.76, respectively. It is obvious that the ChlOC5 product performed well in these different regions (Fig. 15).

In summary, the ChlOC5 product delivered the maximum number of matching data points with both the Bio-Argo data set (the matching data were collected mostly between 2012 to 2016) and the NOMAD data set (the matching data points were distributed mainly between 1997 to 2007), which were 3931 and 1293, respectively. These numbers were much higher than those obtained for Chl1-AVW, Chl1-GSM, or Chl2 products. Consequently, ChlOC5 delivered the best result for both Bio-Argo and NOMAD data sets in terms of absolute accuracy. Regarding the absolute accuracy performance in different regions, ChlOC5 excelled in Case II, Case I, and mixed waters. These results suggest that the ChlOC5 product has relatively higher accuracy and strong adaptability.

4.2 Comparison and validation of the chlorophyll-a concentration values in the South China Sea

Section 4.1 discusses the performance of the different chlorophyll-a concentration inversion products at the

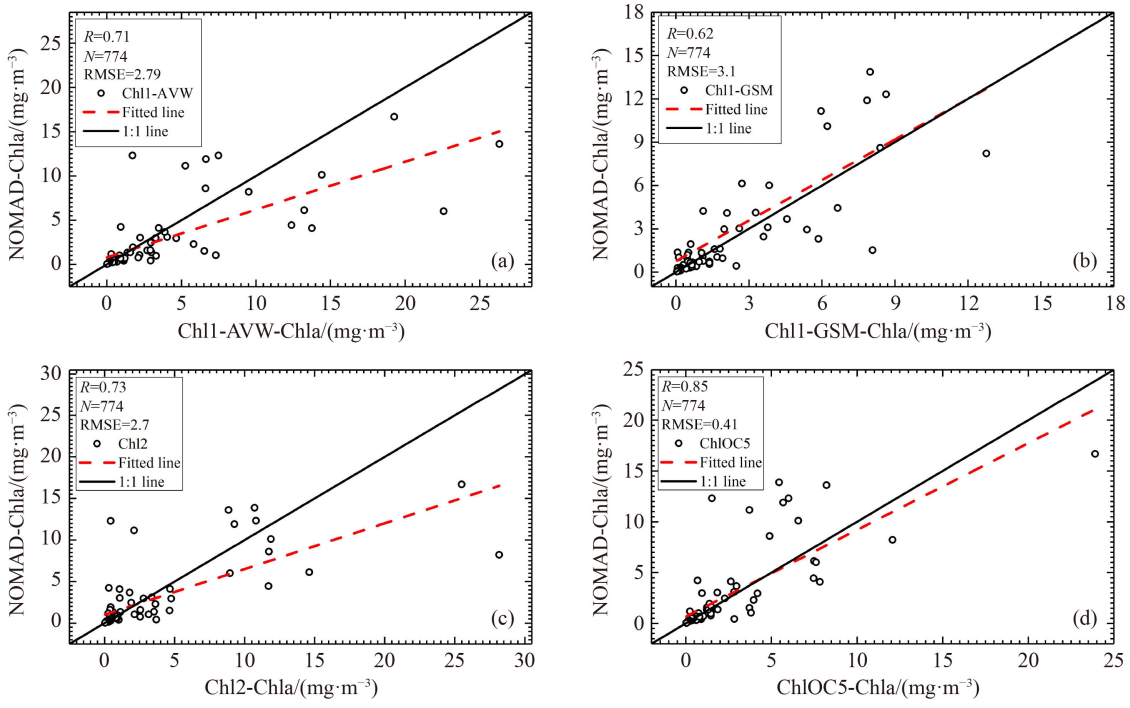

Fig. 14 Comparison of the NOMAD data set with the different satellite inversion products (including the Chl2 product) after standardization.

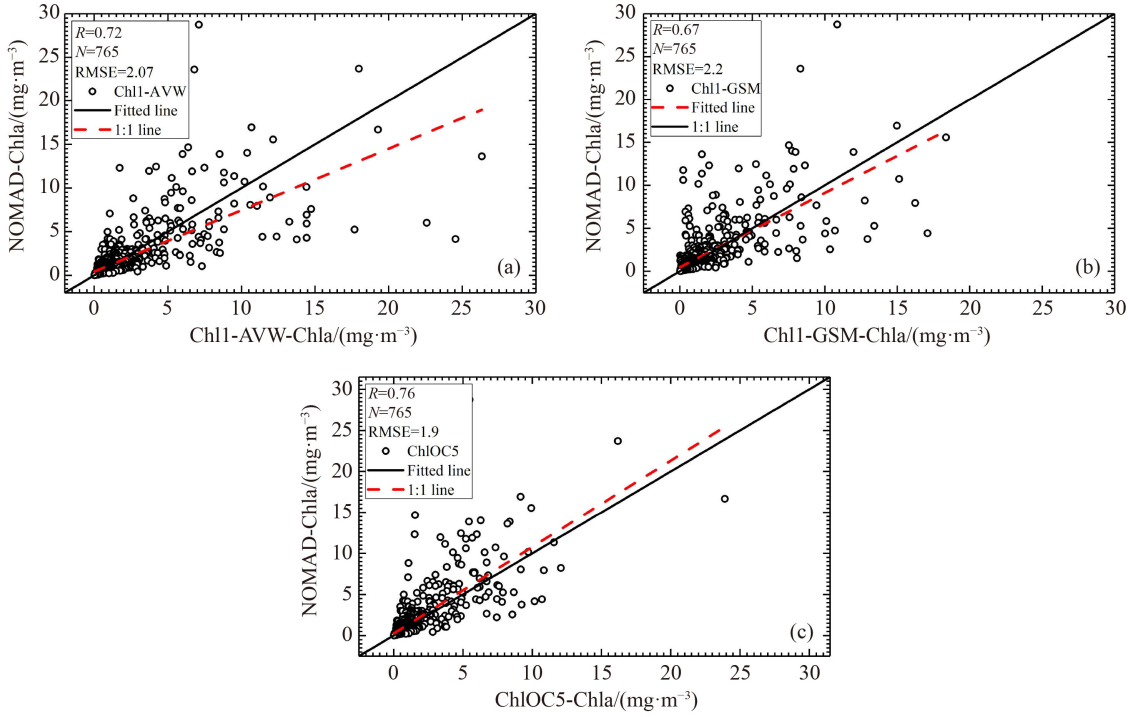

Fig. 15 Comparison of the NOMAD data set with the different satellite inversion products (excluding the Chl2 product) after standardization.

global scale. However, owing to regional differences, the optical characteristics of the waters differ for different regions. Therefore, it is necessary to specifically investigate the performance of the chlorophyll-a concentration estimation products in different seasons in the study area of the South China Sea to validate the accuracy of the products. In the present study, the actual measurements of chlorophyll-a concentrations in the South China Sea are

not specified with an exact date but rather with an approximate period. Therefore, the 8-day time scale data that matched the period were selected for sampling and quantitative evaluation. Specifically, the spring, summer, autumn, and winter voyages delivered 32, 41, 13, and 26 matching survey points, respectively, and their spatial distributions are provided in Fig. 2.

As shown in Fig. 16, the algorithms performed differently in terms of absolute accuracy among the 36 matching points in the springtime. The actual measurement correlation coefficients were 0.72, 0.74, 0.58, and 0.78 for the Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5 products, respectively. As the majority of the area of the South China Sea belongs to the oligotrophic Case I waters, the Chl2 product, more suitable for Case II waters, demonstrated a subpar performance. Otherwise, Chl1-AVW, Chl1-GSM, and ChlOC5 products demonstrated similar performances in terms of absolute accuracy, with ChlOC5 product being the best (Fig. 16). Overall, except for Chl2, all products could achieve errors smaller than 30% and, therefore, met the requirements.

Figure 17 presents the summer data, containing 41 matching points between the four products and the measured data (Fig. 2). The correlation coefficients between Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5 products and the measurements were 0.94, 0.40, 0.59, and 0.74, respectively. As depicted in Fig. 17, the Chl2 products, suitable for Case II waters, presented a correlation coefficient of 0.59, which is relatively stable, considering the spring scenario of 0.58. The ChlOC5 product was also quite stable, with a correlation coefficient of 0.74 in summer and 0.78 in spring. More dramatic changes occurred for the Chl1-AVW and Chl1-GSM products.

Figure 18 presents the correlation coefficients between the four products and the measured chlorophyll-a concentration in autumn. The largest error occurred in the Chl2 product, in which the inverse values were significantly larger than the actual measurements, as depicted in Fig. 18(c). A large error occurred in the Chl1-AVW and Chl1-GSM products relative to the measured values in the autumn scenario. ChlOC5 was the only product that presented a relatively high correlation coefficient. One reason for this could be the limited number of samples. Another reason could be that the data were mostly collected in November, when the surface chlorophyll-a concentration in the South China Sea was dramatically changing as the season changed from autumn to winter. As a result, the error could have been introduced when the 8-d average satellite products were compared to the measured values. Nonetheless, ChlOC5 had the highest correlation coefficient and the greatest seasonal stability.

Figure 19 presents the correlation coefficients between the four satellite inversion products and the measured chlorophyll-a concentration for winter, which were 0.78, 0.80, 0.50, and 0.74 for Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5 products, respectively. The actual measurements were taken in January when the situation was stable, and the products performed satisfactorily, except

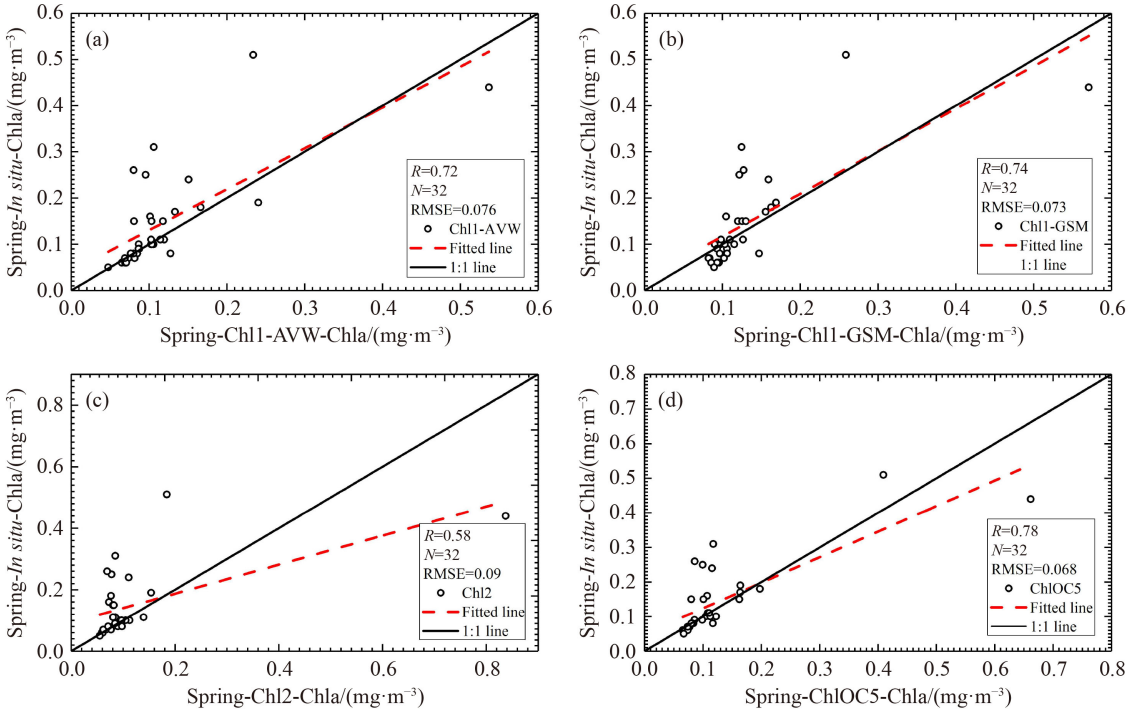

Fig. 16 Comparative analysis between the measured chlorophyll-a concentration and the data from four satellite remote sensing inversion products (i.e., Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5) during spring voyage in the South China Sea.

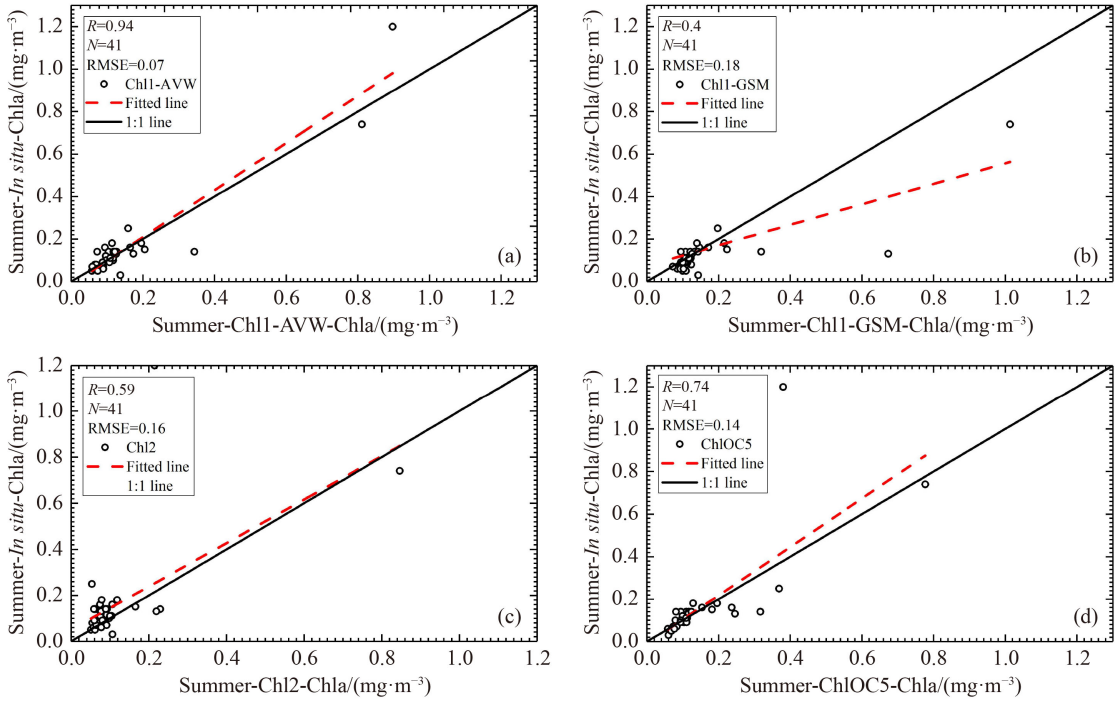

Fig. 17 Comparative analysis between the measured chlorophyll-a concentration and the data from the four remote sensing inversion products (i.e., Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5) during a summer voyage in the South China Sea.

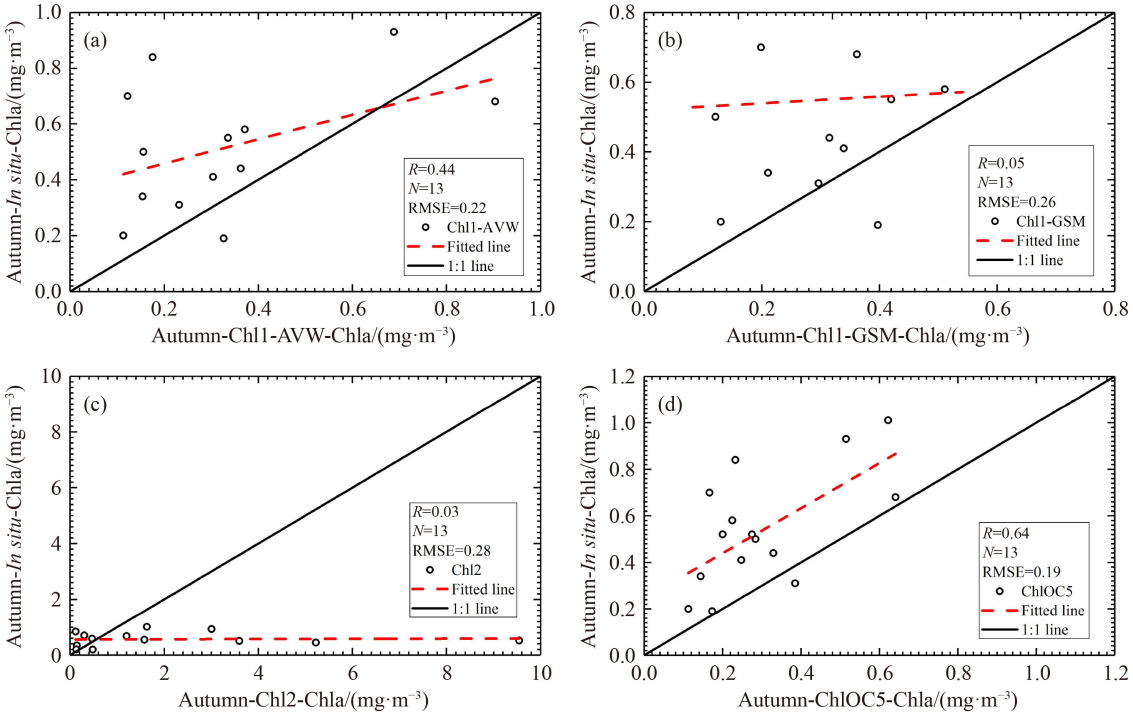

Fig. 18 Comparative analysis between the measured chlorophyll-a concentration and the data from the four satellites remote sensing inversion products (i.e., Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5) during the autumn voyage in the South China Sea.

for Chl2.

In summary, we infer from the seasonal comparison between the actual measurements in the South China Sea and data from the four satellite inversion products that a partial error was introduced due to the matching of the 8-d data with the measured data as there was a lack of accurate daily matching data points. However, overall, in contrast to the dramatic changes exhibited by the other algorithms, the ChlOC5 product demonstrated a stable performance over the four seasons.

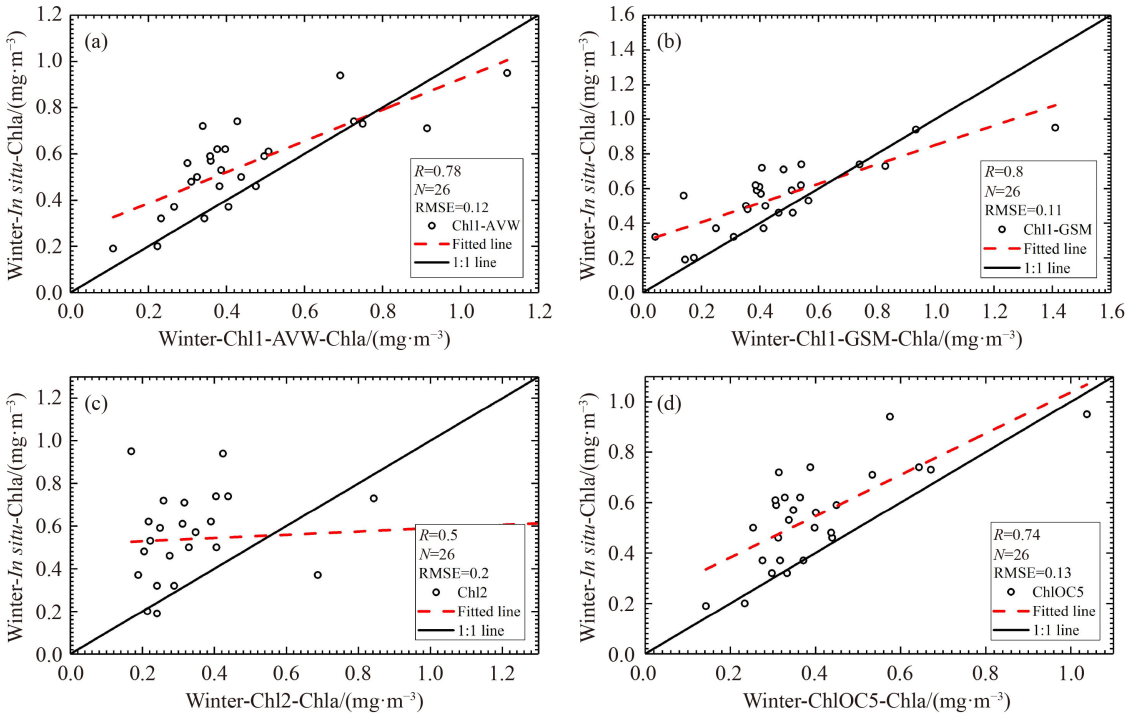

Fig. 19 Comparative analysis between the measured chlorophyll-a concentration and the data from the four remote sensing inversion products (i.e., Chl1-AVW, Chl1-GSM, Chl2, and ChlOC5) during a winter voyage in the South China Sea.

4.3 Comparison of the data from remote sensing products and actual measurements of chlorophyll-a concentration

As the time series of the satellite remote-sensing data used in the present study was relatively long, the collectible data matching with the *in situ* chlorophyll-a concentrations were limited. The Bio-Argo and NOMAD data sets were strictly matched according to the same location and time (same day). Regarding the measured data in the South China Sea, a relatively accurate spatial matching was ensured, although, in the time domain, the 8-day scale data were adopted to match with the actual measurements taken during the voyages, which inevitably introduced some error. These data may present some deviation in terms of the absolute value, although they are still useful for representing the trend and the relative size of the errors.

Table 2 presents the comparative statistical results of the four products when the same standard was applied (the same date and the same longitude and latitude; three or four kinds of products included). The 2008 data points of Bio-Argo and the 765 data points of NOMAD covered various regions, water types, and periods. The results reveal that ChlOC5 had relatively higher accuracy and small errors. The South China Sea voyages in the four seasons spanned three years. During that time, the ChlOC5 product generally provided the most accurate, stable, and adaptive results, regardless of the season or the performance of the other products. Overall, ChlOC5 proved to be the best product for this region in terms of temporal and spatial matching between measured data and data from remote sensing inversion products.

5 Discussion

Based on the differences in their water components, water bodies can be broadly classified as Case I waters or Case II waters. In general, phytoplankton and related organisms dominate the optical properties of Case I waters. As a result, oceanic waters are the typical Case I waters. On the other hand, Case II waters are characterized by a combination of phytoplankton, yellow suspended particulate matter, etc., and present more complex optical properties than Case I waters. Case II waters are commonly located in inland, estuarine, and offshore regions. The most commonly used chlorophyll-a concentration inversion algorithms that could be extended operationally to a larger scale include empirical, semi-analytical, and neural network algorithms. Owing to the differences in the band settings of these commonly used ocean color sensors, the inversion algorithms of their chlorophyll-a concentration products also differ. The most widely used algorithms are the OC3/OC4-based Chl1-AVW, Chl1-GSM, and ChlOC5 products and the neural network-based Chl2 products. However, the application potential of the Chl2 products is limited. This is because the chlorophyll-a concentration inversion algorithm of the Chl2 product has been designed and validated mainly for highly turbid coastal waters and,

Table 2 Comparative statistical evaluation of the remote sensing inversion chlorophyll-a concentration (Chl_a) with the actual measured chlorophyll-a concentration

Measured Chl-a concentration	Inversion products of Chl-a	Number of matching points	*R*	RMSE	MAD
Bio-Argo	Chl1-AVW	2008	0.73	0.53	0.1996
	Chl1-GSM	2008	0.85	0.41	0.1662
	ChlOC5	2008	0.84	0.42	0.1835
NOMAD	Chl1-AVW	77	0.71	2.79	1.8674
	Chl1-GSM	77	0.62	3.1	1.4998
	Chl2	77	0.73	2.7	1.6745
	ChlOC5	77	0.85	0.41	1.3459
	Chl1-AVW	765	0.72	2.07	0.9122
	Chl1-GSM	765	0.67	2.2	0.9158
	ChlOC5	765	0.76	1.9	0.7687
SCS-Spring	Chl1-AVW	32	0.72	0.076	0.0445
	Chl1-GSM	32	0.74	0.073	0.0463
	Chl2	32	0.58	0.09	0.1706
	ChlOC5	32	0.78	0.068	0.0426
SCS-Summer	Chl1-AVW	41	0.94	0.07	0.0355
	Chl1-GSM	41	0.4	0.18	0.0742
	Chl2	41	0.59	0.16	0.0668
	ChlOC5	41	0.74	0.14	0.0463
SCS-Autumn	Chl1-AVW	13	0.44	0.22	0.2684
	Chl1-GSM	13	0.05	0.26	0.2582
	Chl2	13	0.03	0.28	1.9822
	ChlOC5	13	0.64	0.64	0.2593
SCS-Winter	Chl1-AVW	26	0.78	0.12	0.1426
	Chl1-GSM	26	0.8	0.11	0.1408
	Chl2	26	0.5	0.2	0.2991
	ChlOC5	26	0.74	0.13	0.1586

therefore, provides better data coverage frequency and highly accurate of chlorophyll-a concentration values particularly in the coastal and offshore regions, while producing large errors in both spatial distribution and absolute values when applied to the clear Case I waters. Consequently, despite certain advantages of the Chl2 product in terms of the data coverage frequency and accuracy in the offshore region, the algorithm has limitations in data accuracy, data coverage frequency, and time scales of coverage in clear water regions. Moreover, the Chl2 product was initially designed for the MERIS sensor, which was relieved of its service in 2012, and there are no associated data available now. These issues reflect why the chlorophyll-a concentration inversion method of Chl2 products demonstrates the worst overall performance in terms of data coverage frequency and absolute accuracy among the various commonly used inversion algorithms.

The OC*x* algorithm utilizes the statistical relationship between the measured chlorophyll concentration data and the ratio of the blue–green band (440–670 nm) during the remote sensing process. Among the commonly used operationalized, global-scale, and empirical inversion algorithms, the OC4 algorithm is the chlorophyll concentration inversion algorithm commonly used by SeaWiFS, the OC4E algorithm is used by MERIS, the OC3M algorithm is utilized by MODIS, and the OC3v algorithm is employed by VIIRS. The parameters and the inversion formulas of the OC3/OC4 algorithm are presented on the website and Eq. (3), respectively,

$$\log_{10}(\text{chl\_}a) = a_0 + \sum_{i=1}^{4} a_i \left(\log_{10}\left(\frac{R_{rs}\lambda_{\text{blue}}}{R_{rs}\lambda_{\text{green}}} \right) \right)^i, \qquad (3)$$

where λ_{blue} and λ_{green} represent the bands closest to 443 nm and 555 nm, respectively, and $R_{rs}\lambda_{\text{green}}$ represents the band with the best-measured correlation coefficient value among the several blue bands selected for performing the chlorophyll-a concentration inversion using the algorithm. However, in the coastal regions, the

inversion of chlorophyll-a concentration values based on the OC3/OC4 algorithm produces inversion result values that are higher compared to the actual values, with severe overestimations observed during the period between late summer to the early spring season when the yellow matter and suspended matter dominate the optical properties of the water column being studied. As a result, during the validation of the Chl1-AVW and Chl1-GSM products using previously measured data, the inverse chlorophyll-a concentration values produced by these algorithms were found to be more accurate in the Case I water regions and less accurate in the Case II water regions. In addition, the OC3/OC4 algorithm leads to atmospheric overcorrection, which is thought to be largely responsible for the low data coverage frequency in coastal water regions of high turbidity. To resolve the overestimation issue encountered when measuring chlorophyll-a concentration in the nearshore region using the OC3/OC4 algorithm, an empirical algorithm named the ChlOC5 algorithm was developed, which was based on the parameterized relationship between the ratio of the results of the OC4 algorithm, the 412 nm band, and the 555 nm band for different chlorophyll concentrations (Gohin et al., 2002). The ChlOC5 algorithm correlates the OC4 algorithm results, the 412 nm band, and the 555 nm band with the measured chlorophyll-a concentration values through the construction of a look-up table based on the data sets of measured ocean optical data of various coastal areas, such as the English Channel and Bay of Biscay. The limitations of the OC4 algorithm were overcome by another algorithm, named the OC5 algorithm, which combines the 412 nm and 555 nm channel bands. In this algorithm, the 555 nm channel band is used mainly to reveal and correct the effect of suspended matter on the band ratio results in the OC4 algorithm. The 412 nm channel band, on the other hand, is primarily used to correct and adjust the effects of yellow matter and atmospheric overcorrection issues encountered in the OC4 algorithm. Eq. (4) presents the expression of the OC5 algorithm:

$$\mathrm{OC5}_a = \mathrm{OC4}(C_1) - \mathrm{A1}(\mathrm{OC4}(C_1) - 0.55)^{A2}, \qquad (4)$$

where the values of parameters A1 and A2 are 0.18 and 2.0, respectively. The parameters of the OC4 algorithm are presented on the website and given in Eq. (3). The atmospheric overcorrection issue encountered during measurements in the nearshore highly turbid water regions could lead to zero or negative values of chlorophyll-a concentration obtained for the region using algorithm inversion, consequently affecting data coverage frequency. On the other hand, overestimation of the chlorophyll-a concentration values could occur in the coastal water bodies considering the domination of yellow matter and suspended matter. The OC5 algorithm combines the advantages offered by the OC3/OC4 algorithm in Case I water regions with the improvements brought by the OC3/OC4 algorithm by resolving the atmospheric overcorrection issue and the overestimated values in the coastal Case II water regions. Consequently, ChlOC5 achieves better data coverage frequency and improved absolute accuracy than the other algorithms.

Therefore, the ChlOC5 algorithm offers the highest data coverage frequency at the daily, 8-day, and monthly time scales among the available operational chlorophyll-a concentration inversion algorithms. Moreover, the chlorophyll-a concentration results generated using the ChlOC5 algorithm have the highest accuracy in different water types and regions.

6 Conclusions

The most common remote sensing inversed chlorophyll-a concentrations data are Chl1-AVW, Chl1-GSM, Chl2, and Chl-OC5 products. The current study compared these four products in terms of their global coverage and coverage across the South China Sea region and its surrounding areas. In the analysis results, the Chl1 product presented relatively high data coverage frequency and high accuracy only in Case I water regions. In contrast, the Chl2 product only presented high data coverage frequency and accuracy in the estuarine coastal Case 2 water regions. The current study found that the ChlOC5 product presented much higher data coverage frequency, the highest accuracy, and the strongest stability globally and locally in the Case I water regions and the turbid Case II water regions. Specifically, the ChlOC5 product presented better data coverage frequency, particularly evident in the 8-day data, in which this product exhibited approximately 20% higher data coverage than the other three products. It was also observed that the ChlOC5 product exhibited the highest accuracy and the strongest stability, both globally and locally, in coastal Case II waters and open oceans, as well as in oligotrophic oceans and eutrophic waters at high latitudes.

The study of climate change demands data with a relatively high time resolution to collect relevant information related to climate change on finer time scales. Because of its high temporal resolution, an 8-day time-scale product would be best suited for this purpose. The 8-day average data, on the other hand, may present massive data loss issues due to cloud interference and algorithm limitations, which would severely impede climate change research. In the present study as well, the analysis of the data from the South China Sea region and its surrounding areas, which were collected between the years 1998 and 2018, revealed that the chlorophyll-a concentration data collected on the daily, 8-day, and monthly time-scale presented a data loss rate of 90%–80%, 60%–30%, and approximately 30%–10%,

respectively, which is considerably high. Therefore, data reconstruction becomes necessary due to the serious impact of data loss (Liu and Wang, 2019; Martinez et al., 2020; Ma et al., 2021).

Acknowledgments This research was funded by the Project for Fostering Outstanding Young talents of Henan Academy of Sciences (No. 210401001), Special Project for Team Building of Henan Academy of Sciences (No. 200501007), Science and Technology Research Project of Henan Province (Nos. 212102310424, 222102320467, and 212102310024); Major Scientific Research Focus Project of Henan Academy of Sciences (No. 210101007). The authors would like to thank the reviewers and the editor for their constructive comments.

References

ACRI-ST GlobColour Team (2017). GlobColour Product User Guide, GC-UM-ACR-PUG-01, Version 4.1. (Sophia-Antipolis)

Brewin R J W, Mélin F, Sathyendranath S, Steinmetz F, Chuprin A, Grant M (2014). On the temporal consistency of chlorophyll products derived from three ocean-colour sensors. ISPRS J Photogramm Remote Sens, 97: 171–184

Cai P, Zhao D, Wang L, Huang B, Dai M (2015). Role of particle stock and phytoplankton community structure in regulating particulate organic carbon export in a large marginal sea. J Geophys Res Oceans, 120(3): 2063–2095

Chen C, Shiah F, Chung S, Liu K (2006). Winter phytoplankton blooms in the shallow mixed layer of the South China Sea enhanced by upwelling. J Mar Syst, 59(1–2): 97–110

Chen Y L, Chen H, Lin I, Lee M, Chang J (2007). Effects of cold eddy on Phytoplankton production and assemblages in Luzon Strait bordering the South China Sea. J Oceanogr, 63(4): 671–683

Doerffer R, Schiller H (2007). The MERIS Case 2 water algorithm. Int J Remote Sens, 28(3–4): 517–535

Eilers P H C, Peeters J C H (1988). A model for the relationship between light intensity and the rate of photosynthesis in phytoplankton. Ecol Modell, 42(3–4): 199–215

Gai S, Wang H, Liu G, Huang L, Song X (2012). Chlorophyll a increase induced by surface winds in the northern South China Sea. Acta Oceanol Sin, 31(4): 76–88

Gohin F, Druon J N, Lampert L (2002). A five channel chlorophyll concentration algorithm applied to SeaWiFS data processed by SeaDAS in coastal waters. Int J Remote Sens, 23(8): 1639–1661

Grosse J, Bombar D, Doan H N, Nguyen L N, Voss M (2010). The Mekong River plume fuels nitrogen fixation and determines phytoplankton species distribution in the South China Sea during low- and high-discharge season. Limnol Oceanogr, 55(4): 1668–1680

Hu C, Lee Z, Franz B (2012). Chlorophyll-a algorithm for oligotrophic oceans: a novel approach based on three-band reflectance difference. J Geophys Res Oceans, 117(C1): C01011

Huynh H T, Alvera-Azcárate A, Barth A, Beckers J (2016). Reconstruction and analysis of long-term satellite-derived sea surface temperature for the South China Sea. J Oceanogr, 72(5): 707–726

Isoguchi O, Kawamura H, Ku-Kassim K (2005). El Niño–related offshore phytoplankton bloom events around the Spratley Islands in the South China Sea. Geophys Res Lett, 32(21): L21603

Kim T, Lee K, Duce R, Liss P (2014). Impact of atmospheric nitrogen deposition on phytoplankton productivity in the South China Sea. Geophys Res Lett, 41(9): 3156–3162

Lewandowska A, Sommer U (2010). Climate change and the spring bloom: a mesocosm study on the influence of light and temperature on phytoplankton and mesozooplankton. Mar Ecol Prog Ser, 405: 101–111

Li G, Wu Y, Gao K (2009). Effects of Typhoon Kaemi on coastal phytoplankton assemblages in the South China Sea, with special reference to the effects of solar UV radiation. J Geophys Res Oceans, 114(G4): G04029

Lin I I, Wong G T F, Lien C, Chien C, Huang C, Chen J (2009). Aerosol impact on the South China Sea biogeochemistry: an early assessment from remote sensing. Geophys Res Lett, 36(17): L17605

Lin I, Lien C, Wu C, Wong G T F, Huang C, Chiang T (2010). Enhanced primary production in the oligotrophic South China Sea by eddy injection in spring. Geophys Res Lett, 37(16): L16602

Lin J, Cao W, Wang G, Hu S (2014). Satellite-observed variability of phytoplankton size classes associated with a cold eddy in the South China Sea. Mar Pollut Bull, 83(1): 190–197

Liu F, Chen C (2014). Seasonal variation of chlorophyll a in the South China Sea from 1997–2010. Aquat Ecosyst Health Manage, 17(3 3SI): 212–220

Liu K K, Tseng C M, Yeh T Y, Wang L W (2010). Elevated phytoplankton biomass in marginal seas in the low latitude ocean: a case study of the South China Sea. Adv Geosci, 18: 1–17

Liu X, Wang M (2019). Filling the Gaps of Missing Data in the Merged VIIRS SNPP/NOAA-20 Ocean Color Product Using the DINEOF Method. Remote Sens (Basel), 11(2): 178

Ma C, Zhao J, Ai B, Sun S (2021). Two-decade variability of sea surface temperature and chlorophyll-a in the northern South China Sea as revealed by reconstructed cloud-free satellite data. IEEE Trans Geosci Remote Sens, 59(11): 9033–9046

Maritorena S, d'Andon O H F, Mangin A, Siegel D A (2010). Merged satellite ocean color data products using a bio-optical model: characteristics, benefits and issues. Remote Sens Environ, 114(8): 1791–1804

Maritorena S, Siegel D A (2005). Consistent merging of satellite ocean color data sets using a bio-optical model. Remote Sens Environ, 94(4): 429–440

Martinez E, Gorgues T, Lengaigne M, Fontana C, Sauzede R, Menkes C, Uitz J, Lorenzo E, Fablet R (2020). Reconstructing global chlorophyll-a variations using a non-linear statistical approach. Front Mar Sci, 7: 464

Ning X, Peng X, Le F, Hao Q, Sun J, Liu C, Cai Y (2008). Nutrient limitation of phytoplankton in anticyclonic eddies of the northern South China Sea. Biogeosci Discuss, 5(6): 4591–4619

O'Reilly J E, Werdell P J (2019). Chlorophyll algorithms for ocean color sensors – OC4, OC5 & OC6. Remote Sens Environ, 229: 32–47

O'Reilly J E, Maritorena S, Siegel D A, O'Brien M C, Toole D, Mitchell B G, Kahru M, Chavez F P, Strutton P, Cota S G F,

Hooker S B, McClain C R, Carder K L, Müller-Karger F, Harding L, Magnuson A, Phinney D, Moore G F, Aiken J, Arrigo K R, Letelier R, Culver M (2000). Ocean color chlorophyll a algorithms for SeaWiFS, OC2, and OC4: Version 4. In: Hooker S B, Firestone E R, eds. SeaWIFS Postlaunch Calibration and Validation Analyses, Part 3, NASA Technical Memorandum, 2000-206892, vol. 11, NASA Goddard Space Center, Greenbelt, MD(2000), 9–27

Shan G, Hui W (2008). Seasonal and spatial distributions of phytoplankton biomass associated with monsoon and oceanic environments in the South China Sea. Acta Oceanol Sin, 27(6): 17–32

Shang S, Li L, Sun F, Wu J, Hu C, Chen D, Ning X, Qiu Y, Zhang C, Shang S (2008). Changes of temperature and bio-optical properties in the South China Sea in response to Typhoon Lingling, 2001. Geophys Res Lett, 35(10): L10602

Shang X, Zhu H, Chen G, Xu C, Yang Q (2015). Research on cold core eddy change and phytoplankton bloom induced by typhoons: case studies in the South China Sea. Adv Meteorol, 2015: 1–19

Shen P P, Tan Y H, Huang L M, Zhang J L, Yin J Q (2010). Occurrence of brackish water phytoplankton species at a closed coral reef in Nansha Islands, South China Sea. Mar Pollut Bull, 60(10): 1718–1725

Shropshire T, Li Y, He R (2016). Storm impact on sea surface temperature and chlorophylla in the Gulf of Mexico and Sargasso Sea based on daily cloud-free satellite data reconstructions. Geophys Res Lett, 43(23): 12199–12207

Tang D L, Ni I H, Kester D R, Müller-Karger F E (1999). Remote sensing observations of winter phytoplankton blooms southwest of the Luzon Strait in the South China Sea. Mar Ecol Prog Ser, 191(3): 43–51

Tang S, Dong Q, Liu F (2011). Climate-driven chlorophyll-a concentration interannual variability in the South China Sea. Theor Appl Climatol, 103(1–2): 229–237

Wang J, Tang D, Sui Y (2010). Winter phytoplankton bloom induced by subsurface upwelling and mixed layer entrainment southwest of Luzon Strait. J Mar Syst, 83(3–4): 141–149

Wang S, Hsu N C, Tsay S, Lin N, Sayer A M, Huang S, Lau W K (2012). Can Asian dust trigger phytoplankton blooms in the oligotrophic northern South China Sea? Geophys Res Lett, 39(5): L05811

Wang S, Tang D (2010). Remote sensing of day/night sea surface temperature difference related to phytoplankton blooms. Intern J Remote Sens, 31(17–18): 4569–4578

Wang Y, Liu D (2014). Reconstruction of satellite chlorophyll-a data using a modified DINEOF method: a case study in the Bohai and Yellow seas, China. Int J Remote Sens, 35(1): 204–217

Wang Z, Du J, Xia J, Chen C, Zeng Q, Tian L, Wang L, Mao Z (2020). An effective method for detecting clouds in GaoFen-4 images of coastal zones. Remote Sens (Basel), 12(18): 3003

Werdell P J, Bailey S W (2005). An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. Remote Sens Environ, 98(1): 122–140

Winder M, Berger S A, Lewandowska A, Aberle N, Lengfellner K, Sommer U, Diehl S (2012). Spring phenological responses of marine and freshwater plankton to changing temperature and light conditions. Mar Biol, 159(11): 2491–2501

Xi H, Losa S N, Mangin A, Soppa M A, Garnesson P, Demaria J, Liu Y, d'Andon O H F, Bracher A (2020). Global retrieval of phytoplankton functional types based on empirical orthogonal functions using CMEMS GlobColour merged products and further extension to OLCI data. Remote Sens Environ, 240: 111704

Xu H, Paerl H W, Qin B, Zhu G, Gao G (2010). Nitrogen and phosphorus inputs control phytoplankton growth in eutrophic Lake Taihu, China. Limnol Oceanogr, 55(1): 420–432

Yamada K, Ishizaka J, Yoo S, Kim H, Chiba S (2004). Seasonal and interannual variability of sea surface chlorophyll a concentration in the Japan/East Sea (JES). Prog Oceanogr, 61(2–4): 193–211

Yang Y, Xian T, Sun L, Fu Y (2012). Summer monsoon impacts on chlorophyll-a concentration in the middle of the South China Sea: climatological mean and annual variability. Atmos Ocean Sci Lett, 5(1): 15–19

Yuan X, Yin K, Harrison P J, Zhang J (2011). Phytoplankton are more tolerant to UV than bacteria and viruses in the northern South China Sea. Aquat Microb Ecol, 65(2): 117–128

Zhang M, Zhang Y, Shu Q, Zhao C, Wang G, Wu Z, Qiao F (2018). Spatiotemporal evolution of the chlorophyll a trend in the North Atlantic Ocean. Sci Total Environ, 612: 1141–1148

Zhao H, Tang D, Wang Y (2008). Comparison of phytoplankton blooms triggered by two typhoons with different intensities and translation speeds in the South China Sea. Mar Ecol Prog Ser, 365: 57–65