Remote Sensing of Environment 334 (2026) 115195

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Highest quality remote sensing reflectance database compiled from 20+ years of MODIS-aqua measurements

Longteng Zhao, Zhongping Lee *, Xiaolong Yu, Tianhao Wang, Daosheng Wang, Shaoling Shang

State Key Laboratory of Marine Environmental Science College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China

ARTICLE INFO

Editor: Menghua Wang

Keywords:
Highest quality database
Remote sensing reflectance
MODIS-aqua

ABSTRACT

Remote sensing reflectance (R_{rs}) is a fundamental property in satellite ocean color remote sensing, which is critical for retrieving optical-biogeochemical properties and data-driven atmospheric correction algorithms. In this study, with three criteria applicable to ~91% of the global ocean, we compiled a database of the highest quality R_{rs} (HQ_{MODISA}-R_{rs}) of oceanic waters based on 20+ years of ocean color measurements by the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Aqua satellite. While removing a large number of daily "standard" data products, our evaluation showed that the criteria for the highest-quality R_{rs} (CHQR) improved MODIS R_{rs} data consistency with benchmark *in situ* R_{rs} datasets, such as those from MOBY and AERONET-OC. After applying CHQR, analysis of imagery products in the South Pacific Ocean revealed that the coefficient of variation (CV) of R_{rs} among pixels reduced from 0.042 (standard quality control) to 0.030, along with enhanced temporal consistency, which indicates that this approach effectively filters abnormal data products. While such a dataset played a key role in the development of the cross-satellite atmospheric correction algorithm (Lee et al., 2024), we here further demonstrate that applications of HQ_{MODISA}-R_{rs} have ~21.0% of oceanic areas between 50°S and 50°N showing reversed long-term trends of R_{rs} compared to the trend based on the standard R_{rs} product. We anticipate that this highest-quality R_{rs} database would not only improve our evaluation and understanding of long-term changes in various R_{rs}-derivative bio-optical properties of the global ocean, but also help to obtain consistent products among various satellite ocean color missions.

1. Introduction

Data play a pivotal role in daily life as well as in scientific research. It is the base for us to discover new findings, validate theories, and make sound decisions (Karpatne et al., 2017; Sidirourgos et al., 2013). Especially in the AI era, a large and representative volume of data is the key for the development of AI-based algorithms (Whang et al., 2023). In this context, the ability to effectively collect, store, and manage precise data, or the establishment of a database, becomes crucial for advancing both AI-based research and broader scientific inquiries (Carbotte et al., 2022; Marchuk et al., 2013; Schwing, 2023; Tanhua et al., 2019). Realizing the importance and necessity of oceanic databases, various data portals, such as the World Ocean Database (WOD) (Boyer et al., 2013), ESA's Copernicus Marine Service (Le Traon et al., 2019), and NASA's SeaBASS (Werdell et al., 2002), started decades ago to systematically accumulate and organize a wide range of oceanic data for ocean sciences. These have become indispensable tools and resources for oceanography researchers and managers worldwide, and have helped to discover numerous breakthrough findings regarding our oceans and the Earth system (Behrenfeld et al., 2006; Schmidtko et al., 2017; Zhang et al., 2023).

One of the data products included in many data portals, including ESA's Copernicus Marine Service (Le Traon et al., 2019) and NASA's SeaBASS (Werdell et al., 2002), is the remote sensing reflectance (R_{rs}, sr^{-1}), which is a measure of water's color spectrum. R_{rs} can be derived from satellite ocean color measurements (Gordon and Wang, 1994; Morel and Prieur, 1977), which serve as a key bridge to obtain optical and biogeochemical properties of the ocean *via* remote sensing (Lee et al., 2002; O'Reilly and Werdell, 2019; Wei et al., 2021; Werdell et al., 2013). In addition, R_{rs} is an important input of many big-data models (Hu et al., 2021; Wang and Li, 2024). It is thus not surprising to see R_{rs} as a core product for satellite ocean color missions and that of multi-mission merged datasets, such as that of OC-CCI (Ocean Colour

* Corresponding author.
E-mail address: ZhongPing.Lee@umb.edu (Z. Lee).

https://doi.org/10.1016/j.rse.2025.115195
Received 24 June 2025; Received in revised form 15 November 2025; Accepted 8 December 2025
Available online 13 December 2025

Climate Change Initiative) (Sathyendranath et al., 2019).

The available R_{rs} data products can be broadly categorized into three types: those derived from *in situ* measurements, those obtained through ocean color satellites, and those from models rooted in radiative transfer. *In situ* measurements have the advantage of avoiding atmospheric correction errors, but they are limited in spatial and temporal coverages. Additionally, many elements of the measurement protocol differ between observation programmes, such as instrument, measurement geometry, and data processing (Lee et al., 2010; Ruddick et al., 2019). These inherent challenges make it difficult to obtain *in situ* R_{rs} from different groups with the same quality. While efforts like Valente et al. (2019) and Lehmann et al. (2023) have successfully compiled *in situ* data spanning 1990s-present, covering a wide range of locations in the global ocean, along with rigorous quality control and standardization, such datasets are still discrete in space and time. On the other hand, data from modeling *via* radiative transfer (Fan et al., 2021; IOCGG, 2006) will always face the challenge of representativeness of the bio-optical variations in the real world.

To overcome these constraints, satellite-derived products offer extensive spatial and temporal coverage of the global ocean. This need has driven the establishment of projects like GlobColour and OC-CCI, which focus on developing long-term bio-optical property records for oceanic waters. However, as indicated in Lee et al. (2024), there are inconsistencies between the R_{rs} products distributed by GlobColour and by OC-CCI, likely originated from different atmospheric correction algorithms used, as well as differences in merging methods (van Oostende et al., 2022). These inconsistencies introduce doubts or controversies on the response of the oceanic ecosystem under a changing climate (Pauthenet et al., 2024). It calls for the establishment of a robust, highest-quality, satellite R_{rs} database that can be considered as the standard. Such a database will not only be important for a reliable evaluation of the oceanic ecosystem, but also serve as a benchmark for other ocean-color missions or agencies to assess or calibrate their R_{rs} products.

In the late 1990s and early 2000s, the international community put a series of ocean-color satellites into space, with the SeaWiFS, MODIS-Aqua, and MERIS the most well-known, along with data products widely used (Esaias et al., 1998; McClain, 2009; Rast et al., 1999). Among the three, in addition to the longest period in service, MODIS-Aqua (MODIS-A hereafter) has a much higher signal-to-noise (SNR) ratio compared to SeaWiFS and MERIS (Hu et al., 2012), which is critical for high-quality R_{rs} from satellite measurements (Franz et al., 2012). After years of great effort and investment, the R_{rs} product from MODIS-A has shown superior quality and consistency (Zhang et al., 2022). In view of these characteristics and strengths regarding MODIS-A, it is thus natural to develop the "highest-quality" R_{rs} global database from the 20+ years of MODIS-A measurements. Note that the use of the word "highest", although subjective to any criteria, is in order to distinguish the dataset from that filtered with the Level-2 flags (l2-flags hereafter) adopted by NASA's SeaDAS (see below), where the resulted data are generally considered as "high quality". Such a database will not only serve as a benchmark for other ocean-color missions, cross-satellite calibration or atmospheric correction (Lee et al., 2024), but also be of great value for long-term climate research, ecosystem monitoring and marine resource management.

2. Criteria to determine the highest-quality R_{rs} from satellite measurements

2.1. Criteria for the highest-quality R_{rs}

Due to that a wide range of environmental factors influence the generation of R_{rs} from satellite measurements, and at the same time, R_{rs} just makes ~20% or less of the top-of-atmosphere signal, the R_{rs} product from an ocean-color satellite is prone to noise or errors (IOCCG, 2010). As a result, only a fraction of the daily global R_{rs} data can be considered valid for scientific applications. Traditionally, the l2-flags of NASA's Ocean Color Level-2 products have been used to identify and exclude questionable data, with the absence of any flags related to data quality control indicating the high-quality data (Hooker et al., 2003). However, this traditional method, which relies on predefined flags to exclude conditions likely to introduce various uncertainties, often fails to ensure the quality of retained R_{rs} data, leading to the inclusion of noisy R_{rs} products (Chen et al., 2016; Hu et al., 2013).

To address these shortcomings, two separate approaches were proposed in the past to assess the quality of an R_{rs} spectrum: the constraint based on chlorophyll concentration (CBCC) (Hu et al., 2013), and the quality assurance (QA) system (Wei et al., 2016). CBCC evaluates R_{rs} quality by comparing the chlorophyll-a value (Chl) derived from band ratio and band difference algorithms (Hu et al., 2013). If the Chl values from the two independent algorithms agree within approximately 5%, the R_{rs} is classified as "error-free". This method has been widely used in validating satellite-derived R_{rs} and improving data correction (Chen et al., 2016; Chen et al., 2023; Chen et al., 2021). However, further applications found that at times an R_{rs} spectrum with poor spectral quality (*e.g.*, a completely negative R_{rs} spectrum) may still have the band-ratio- and band-difference-calculated Chl satisfying the 5% criterion, thus a stricter constraint for the highest-quality R_{rs} is required to ensure a robust database.

The QA system developed by Wei et al. (2016), on the other hand, compares an R_{rs} spectrum to that in an R_{rs} database developed from *in situ* measurements, which quantifies the spectral similarity of the targeted R_{rs} against the R_{rs} in the database. The system assigns a score close to 1 for the highest similarity, subsequently the quality of the target R_{rs} is quantified based on this score (Wei et al., 2016). This quality assurance scheme has shown excellent applications in many studies (Liu et al., 2021; Men et al., 2023). However, this QA system mainly evaluates the spectral shape of an R_{rs}, thus an R_{rs} spectrum with small bias may still pass this filter.

In view of the incompleteness of the above schemes, we propose the following criteria for the highest-quality R_{rs} (termed CHQR in the following), shown in Fig. 1, which includes the following three aspects: 1) filtering based on observation condition or failure in data processing, 2) high assurance in spectral shape, and 3) minimum to no residual errors. Specifically, these aspects are:

1) Filtering based on observation condition or failure in data processing: Any R_{rs} product having the following level-2 quality-control

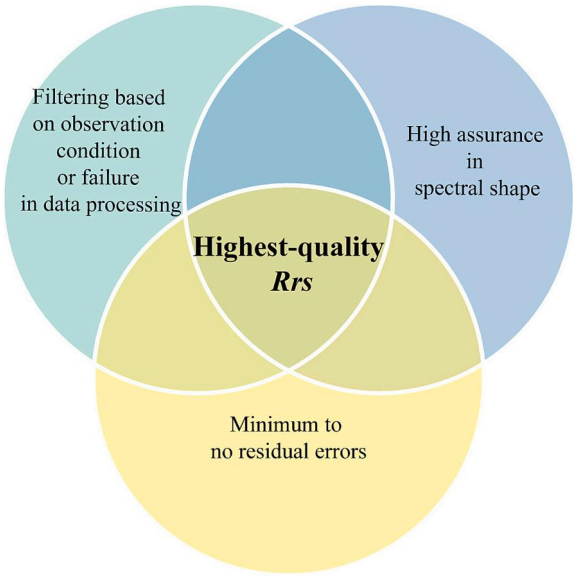

Fig. 1. Criteria for the highest-quality R_{rs} (CHQR).

flags (l2-flags) are excluded: ATMFAIL (1), LAND (2), HIGLINT (4), HILT (5), HISATZEN (6), STRAYLIGHT (9), CLDICE (10), COCCOLITH (11), HISOLZEN (13), LOWLW (15), CHLFAIL (16), NAVWARN (17), MAXAERITER (20), CHLWARN (22), ATMWARN (23), and NAVFAIL (26). In the past, data products after being screened by these flags were considered high quality and further used for the generation of Level-3 products.

2) Spectral shape consistency: The QA system will be applied to a target R_{rs} spectrum, and only QA = 1 will be kept. This helps ensure the data product represents realistic waters.
3) Minimum to no residual errors: Numerical constraints similar to CBCC will be employed to filter out data with residual errors. The commonly used CBCC is limited to oceans where Chl is less than ~0.2 mg/m^3, covering roughly ~70% of the global ocean. Recently, a new algorithm based on band differences of R_{rs} for estimating the absorption coefficient at the 443 nm band (a(443)) was developed, and it was found that this algorithm could be applied to ~91% of the global ocean (Lee et al., 2023). Since a(443) from the band-difference algorithm (Lee et al., 2023) is insensitive to spectrally covarying errors in an R_{rs} spectrum, while a(443) from the band-ratio algorithm (Lee et al., 1998) is sensitive to such a disturbance in R_{rs}, we can implement a constraint (termed as CBAC) based on the difference of a(443) obtained from the two algorithms to screen "error-free" R_{rs}. Details of CBAC are presented in Section 2.2.

2.2. Development of the constraint based on absorption coefficient (CBAC)

For the estimation of the total absorption coefficient at 443 nm (a(443)), Lee et al. (1998) developed a band-ratio algorithm, which is modified to meet the MODIS spectral bands and updated with more available data. Specifically, the band-ratio algorithm for a(443) ($a(443)_{BR}$) is expressed as:

$$a(443)_{BR} = 10^{\alpha_0+\alpha_1\ \rho_{25}+\alpha_2\ \rho_{25}^2+\alpha_3\ \rho_{35}+\alpha_4\rho_{35}^2} \tag{1a}$$

$$\rho_{25} = log10\left[\frac{R_{rs}(443)}{R_{rs}(547)}\right], \rho_{35} = log10\left[\frac{R_{rs}(488)}{R_{rs}(547)}\right] \tag{1b}$$

where α_{0-4} are the fitting coefficients (Lee et al., 1998).

On the other hand, the band-difference algorithm for a(443) ($a(443)_{BD}$) is:

$$MBD_{Rrs(443)} = R_{rs}(547) - \left[R_{rs}(443) + \frac{547-443}{667-443}(R_{rs}(667) - R_{rs}(443))\right] \tag{2a}$$

$$a(443)_{BD} = 10^{\beta_0+\beta_1\ \mathrm{Exp}(\beta_2\ MBD_{Rrs(443)})} \tag{2b}$$

with β_{0-2} the fitting coefficients (Lee et al., 2023).

To obtain the coefficients of α_{0-4} and β_{0-2}, the R_{rs} data from the BIOSOPE (Claustre et al., 2008) and NOMAD (Werdell and Bailey, 2005) datasets were combined and utilized. As in Lee et al. (2023), the a(443) values obtained by the quasi-analytical algorithm (Lee et al., 2002) were considered as the ground "truth" for the derivation of the fitting coefficients. As these data were generally collected in oceanic waters where Raman scattering can make strong contributions in the longer wavelengths (Hu and Voss, 1997; Morel et al., 2002), a correction of the Raman contribution was carried out first (Lee et al., 2013) before implementing the quasi-analytical algorithm. The resulted fittings are presented in Fig. 2 a and b, with the coefficients for the band-difference algorithm as $\beta_0 = -2.12$, $\beta_1 = 0.91$, and $\beta_2 = 247.38$; whereas the coefficients for the band-ratio algorithm are $\alpha_0 = -0.883$, $\alpha_1 = -0.955$, $\alpha_2 = -0.010$, $\alpha_3 = -0.450$, and $\alpha_4 = 0.345$. Fig. 2c shows that there is a high consistency between $a(443)_{BD}$ and $a(443)_{BR}$ for these data, where the upper limit of the absorption coefficient is 0.1 m^{-1}, effectively covering ~91% of the global ocean.

Following the CBCC scheme (Hu et al., 2013), the relative difference between $a(443)_{BD}$ and $a(443)_{BR}$ is calculated as:

$$\delta a(443) = \frac{|a(443)_{BD} - a(443)_{BR}|}{a(443)_{BD}} \tag{3}$$

To screen the highest-quality R_{rs}, it is necessary to determine a threshold for δa(443). A larger value of δa(443) for the threshold might compromise data quality, while a lower value might be overly stringent, filtering out a large amount of reasonable or good-quality data. An appropriate threshold helps ensure the credibility of the data while maintaining its richness.

To determine a reasonable threshold for δa(443), we constructed a high-confidence R_{rs} dataset from MODIS-Aqua. This dataset (approximately 1.37 million data points), which also represents "highest quality", was generated from measurements in February 1, 2007 and was selected from very clear open-ocean waters (Chl < 0.2 mg/m^3) using a combination of three criteria: l2-flags, QA score = 1.0, and CBCC. We calculated δa(443) of this dataset, with its distribution shown in Fig. 3. It is found that approximately 99% of the samples have δa(443) less than 0.15. Based on this result, we set the δa(443) threshold for CBAC as 0.15. This value is aimed at achieving an optimal balance between maintaining high data quality and ensuring data inclusivity, although it could be updated after gaining more insights from the database.

3. Evaluation of the criteria for the highest-quality R_{rs}

3.1. Evaluation using high-quality field observations

For robust criteria, the screened highest-quality R_{rs} should be consistent with data or knowledge that is already known or determined. Therefore, we first used the data measured at the MOBY (Marine Optical Buoy) site as one of the evaluation procedures of the above-proposed criteria for the highest-quality R_{rs}, due to the recognized reliability and wide applications of MOBY R_{rs} in ocean color studies (Clark et al., 2002; Clark et al., 2003). In this research, we exclusively selected those marked "high-quality" R_{rs} from MOBY, spanning from 2003 to the

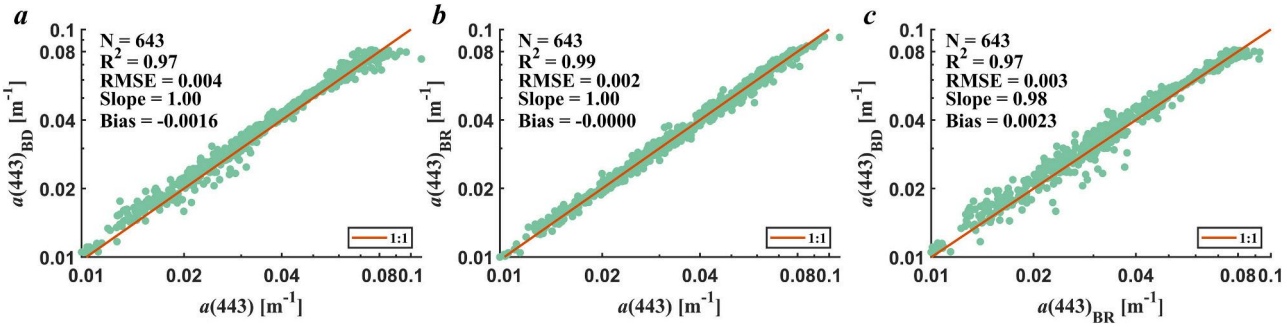

Fig. 2. Fitting results for the band ratio algorithm and the band difference algorithm for a(443). The units for RMSE and Bias are m^{-1}.

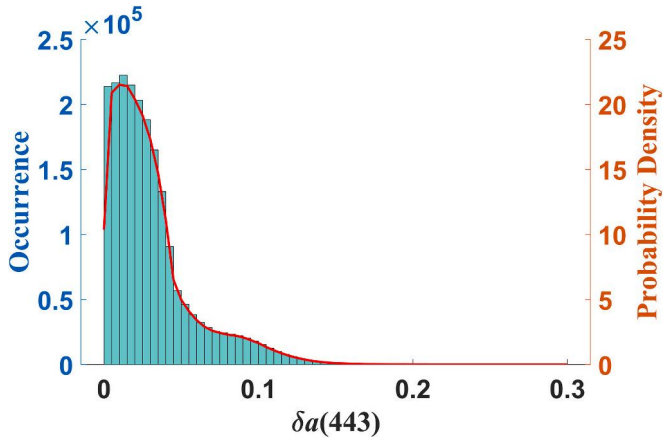

Fig. 3. Distribution of δ*a*(443) of a MODIS-Aqua dataset satisfying no l2-flags, QA score = 1.0, and CBCC, measured in January 2007. The red line indicates the probability density function (PDF) of the δ*a*(443) values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

present, using data downloaded from NOAA's CoastWatch (https://coastwatch.noaa.gov/cwn/products/ocean-color-radiances-moby-field-observations.html), which resulted in 2304 "high-quality" R_{rs} spectra.

As shown in Fig. 4*a*, approximately 81% of the high-quality MOBY data points satisfied CHQR, suggesting a general mutual consistency of the two independent schemes in determining the quality of an R_{rs} spectrum. For the excluded 19% by CHQR, it was found that a portion (~5%) of these spectra were rejected due to QA scores slightly below 1.0. Another portion (~14%) was excluded because the δ*a*(443) values exceeded the 0.15 threshold, even though the actual differences between the two *a*(443) estimates were small. This is primarily because MOBY samples very clear oceanic waters, where *a*(443) values are inherently low.

Further, it was found that most of the excluded cases were concentrated in 2020 and 2021 (see Fig. 4*a*), a period during which the R_{rs} values were substantially lower than those in other years. For example, Fig. 4b presents the CHQR-rejected R_{rs} spectrum (the red line, at the MODIS bands) measured by MOBY on April 29, 2021, which was classified as "good data" by MOBY's procedure. However, this R_{rs} spectrum is considerably lower across all bands compared to the climatological mean (2003–2020) R_{rs} spectrum in April (the blue line). In addition, this rejected R_{rs} spectrum is also much lower than the MODIS-A R_{rs} of the same day (the green line), which is an average over an area of 3 × 3 pixels centered on the MOBY site. Note that the green line matches well with the blue line, meaning that the MODIS-A R_{rs} spectrum aligns very well with the climatological monthly mean R_{rs} obtained by MOBY. These observations indicate that some of the MOBY R_{rs} spectra for the period around 2021, although marked as "good data", are indeed showing some abnormality, which were picked up by the CHQR system.

Furthermore, we assessed the effectiveness of our criteria using matchup datasets between MODIS-A R_{rs} and high-quality *in situ* observations. While the MOBY dataset primarily represents a single site in very clear oceanic waters, we complemented this analysis with matchup data from the Aerosol Robotic Network – Ocean Color (AERONET-OC) program. AERONET-OC is known for its rigorous calibration and validation protocols and offers broader spatial coverage, encompassing a wider range of water conditions (Zibordi et al., 2021). These matchup data were obtained through NASA's official Ocean Color Matchup Search Tool (https://matchup.oceancolor.gsfc.nasa.gov/), ensuring consistency in the selection procedure and quality control. The tool applies a series of default filtering criteria to ensure the reliability of satellite-to-*in situ* comparisons, including:

1) minimum valid satellite pixels of 50%,
2) maximum solar zenith angle of 75°,
3) maximum satellite zenith angle of 60°,
4) maximum time difference between satellite overpass and *in situ* measurement of 3 h,
5) maximum coefficient of variation for satellite pixels of 0.15,
6) maximum irradiance difference between measured and modeled values of 20%,
7) maximum wind speed of 35 m/s.

To ensure compatibility with our oceanic screening framework, we further constrained the dataset by requiring $MBD_{Rrs(443)} < 0.0005\ sr^{-1}$, a necessary condition for applying the band-difference algorithm to estimate *a*(443) (Lee et al., 2023). Notably, for AERONET-OC, only observations from stations located in optically clear waters were used, specifically the USC_SEAPRISM and Venice sites, as they satisfy both the required spectral coverage (across all MODIS ocean color bands) and water-type criteria.

To evaluate the R_{rs} consistency between MODIS-Aqua and *in situ* observations, the Type-I regression (ordinary least squares) was used to calculate the coefficient of determination (R^2), where errors were primarily assumed in MODIS-A measurements. Further, root mean square difference (RMSD), mean absolute percentage difference (MAPD), and

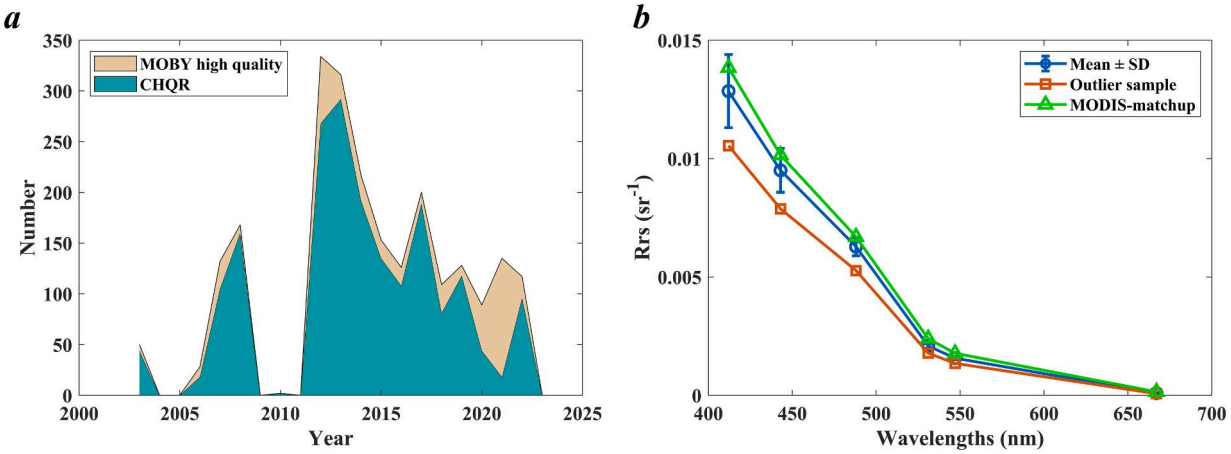

Fig. 4. (*a*) Yearly distribution of data points of MOBY high-quality R_{rs} data and those after applying CHQR. (*b*) An example (measured on April 29, 2021) illustrates a CHQR-rejected MOBY R_{rs} spectrum (red) compared to the multi-year average (+ standard deviation, SD) obtained in April (blue), along with the MODIS-A product (green) of the same day. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

bias were calculated as follows:

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}\# \quad (4)$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)\# \quad (5)$$

$$MAPD = \frac{1}{n}\left|\frac{y_i - x_i}{x_i}\right| \times 100\%\# \quad (6)$$

where x_i and y_i represent *in situ* and MODIS-A measurements, respectively.

Using $R_{rs}(443)$ as an example, Fig. 5 shows the consistency measure between MODIS-A and *in situ* measurements (from MOBY and AERONET-OC). Before applying the CHQR criteria, the matchup dataset contained 1318 samples. The agreement between MODIS-A and *in situ* $R_{rs}(443)$ was reasonable, with $R^2 = 0.91$, RMSD = 0.0010 sr^{-1}, MAPD = 11.9%, and bias = 0.0001 sr^{-1} (Fig. 5*a*), but it is more than two times the 5% desired uncertainty for these blue waters at the blue bands (Brown et al., 2007; Cazzaniga and Zibordi, 2023). After applying the CHQR criteria to both satellite and *in situ* data, 492 high-quality matchups remained. While this dual-filtering process substantially reduced the sample size, it resulted in a notable improvement in the agreement of $R_{rs}(443)$ between satellite and field measurements (see Fig. 5*b*), especially the MAPD dropped to 7.4%, a value that is very close to the desired 5% uncertainty for blue waters. In contrast, the subset data excluded by CHQR showed clearly much wider disagreement (MAPD = 14.6%; Fig. 5*c*), confirming that lower-quality or abnormal observations were effectively and objectively removed.

To highlight the improvements obtained with CHQR, Fig. 6 (also see Table 1) presents the statistical comparison across all spectral bands using matchup data between MODIS-A R_{rs} and high-quality *in situ* observations (including both MOBY and AERONET-OC), constrained by $MBD_{Rrs(443)} < 0.0005$ sr^{-1}. The comparison includes R^2, RMSD, MAPD, and bias values before and after applying the CHQR criteria. The results show that R^2, RMSD, and MAPD consistently improved across all bands after applying CHQR, indicating enhanced agreement between satellite-derived and *in situ* R_{rs} measurements. This is in particular evidenced by the reduction of MAPD at the 412 nm band, changed from 15.3% to ~8.6%, indeed showing the high quality of MODIS-A R_{rs} products after strict quality control, as the 8.6% (or the 7.4% at 443 nm) difference including contributions of the inherent imperfect "matchups" in time and space between satellite and field measurements.

It is necessary to point out that the initial matchup dataset was already filtered using various quality control measures—such as limits on viewing geometry, pixel variability, and solar conditions—ensuring that only relatively reliable matchups were included. The fact that CHQR still significantly improved the agreement, particularly in reducing MAPD, underscores the value and necessity of implementing more stringent criteria for identifying the "highest-quality" R_{rs} products.

3.2. Evaluation with satellite images

We further evaluated CHQR using MODIS-A R_{rs} data products in the open ocean, as water properties in such regions have low spatial variation due to limited to no impacts from human activities and land runoffs, therefore, the highest-quality R_{rs} product of such a region should show minimal spatial variances. To examine the spatial variability of R_{rs} before and after the application of CHQR, we selected data from an area in the South Pacific Gyre (120° W to 110° W and 25° S to 35° S). This region is known for having some of the clearest natural waters on Earth, with generally uniform spatial distribution and stable seawater conditions over short periods (Morel et al., 2007).

As an example that was randomly selected, Fig. 7 shows the spatial distribution of the coefficient of variation (CV) of $R_{rs}(547)$ of April 6, 2005, which was calculated for all boxes having 3 × 3 pixels. Four different screening methods were compared: (*a*) l2-flags alone, (*b*) QA = 1.0 alone, (*c*) the CBAC scheme with $\delta a(443) < 0.15$ alone, and (*d*) the CHQR system. Fig. 7*e* presents the probability density function (PDF) of the CV values for each method. As illustrated in Fig. 7*e*, and further confirmed by our analysis of the pixel-wise CV distribution, more than 10% of the pixels screened by QA = 1.0 alone exhibit higher spatial variability (CV > 0.1), compared to ~6% by l2-flags alone, and ~ 5% by CBAC alone. In contrast, the CHQR method results in fewer than 1% of pixels with CV > 0.1, indicating a marked improvement in spatial coherence.

Meanwhile, Table 2 provides complementary statistics on the mean and standard deviation of CV values, showing that the CHQR system yields the lowest average CV (0.030) and the lowest variability (standard deviation of 0.016), again reflecting a sizeable improvement in spatial homogeneity. These results confirm that CHQR offers the highest-quality and most spatially consistent R_{rs} products among the methods evaluated, which are also found for the other ocean color bands (see Table S1 in the Supplementary Information).

In addition to the evaluation of the relationship between spatial coherence and screening methods, we further analyzed the relationship between temporal consistency of R_{rs} and screening methods for waters in the South Pacific Gyre, as it is well known that water properties in these ocean deserts vary very slowly, or remain nearly a "constant", from day to day (Claustre et al., 2008; Morel et al., 2007). To evaluate short-term stability, we extracted daily $R_{rs}(547)$ values from a small region (120°W to 118°W and 32°S to 30°S) during a one-month period from April 1 to May 1, 2005. The daily values were also generated using four different

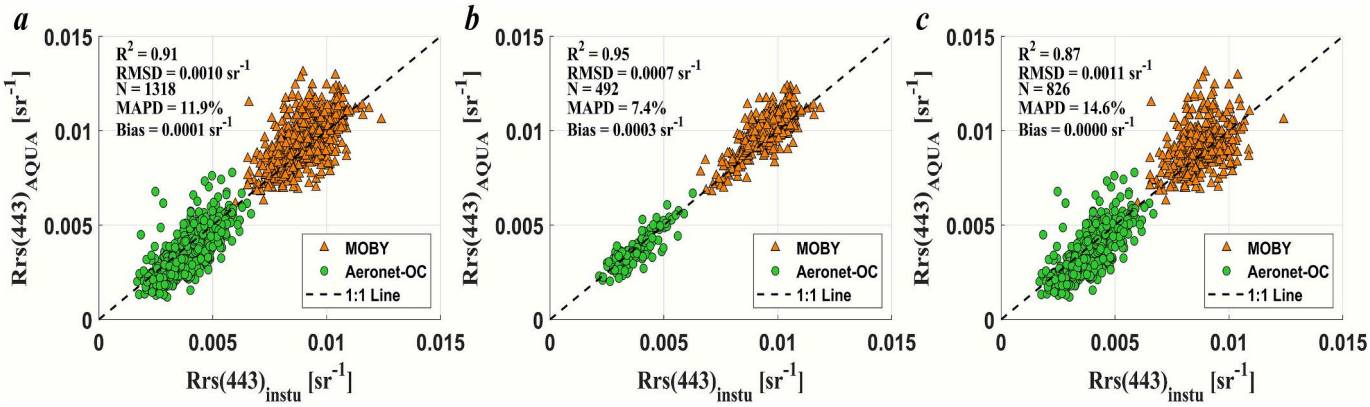

Fig. 5. Comparison of $R_{rs}(547)$ between MODIS-A and *in situ* measurements. (*a*) Before the application of CHQR; (*b*) After the application of CHQR; (*c*) Results for the data excluded by CHQR.

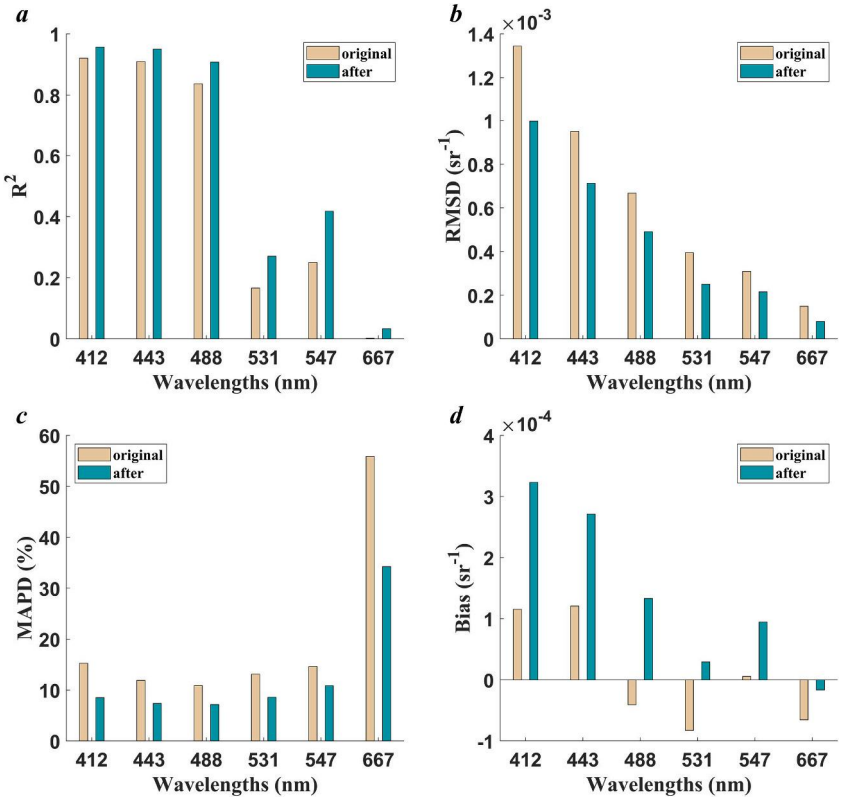

Fig. 6. Statistical measures assessing the consistency between MODIS-A R_{rs} and *in stu* data across various bands before and after applying the CHQR criteria. The parameters include R^2, RMSD, MAPD, and bias (*a-d*).

Table 1
R^2, RMSD (sr^{-1}) MAPD (%) and Bias (10^{-5} sr^{-1}) of MODIS-A R_{rs} compared with *in situ* observations from MOBY and AERONET-OC, before and after CHQR filtering.

Wavelength (nm)	R^2 before CHQR	R^2 after CHQR	RMSD before CHQR (sr^{-1})	RMSD after CHQR (sr^{-1})	MAPD before CHQR (%)	MAPD after CHQR (%)	Bias before CHQR (10^{-5} sr^{-1})	Bias after CHQR (10^{-5} sr^{-1})
412	0.92	0.96	0.0013	0.0010	15.3	8.5	12	32
443	0.91	0.95	0.0010	0.0007	11.9	7.4	12	27
488	0.84	0.91	0.0007	0.0005	10.9	7.2	−4	13
531	0.17	0.27	0.0004	0.0002	13.1	8.6	−8	3
547	0.25	0.42	0.0003	0.0002	14.6	10.9	0.6	9
667	0.00	0.03	0.0001	0.0001	55.9	34.2	−7	−2

screening methods: l2-flags alone, QA = 1.0 alone, CBAC alone, and CHQR.

As shown in Fig. 8 and summarized in Table 3, all methods yielded similar monthly mean values, but their day-to-day variability differed significantly. Among them, CHQR yielded the lowest standard deviation (0.00011 sr^{-1}), indicating the highest temporal consistency. The CBAC method alone also showed low variability (0.00012 sr^{-1}), slightly higher than CHQR, while l2-flags and QA = 1.0 methods alone, resulted in noticeably higher variability (0.00024 sr^{-1} and 0.00025 sr^{-1}, respectively). The much higher standard deviations are results of the abnormal highs on April 3, April 15, *etc.* (See Fig. 8), which were objectively filtered out by CHQR.

For comparison, the R_{rs} product at 560 nm (R_{rs}(560)) produced by the Ocean Colour Climate Change Initiative (OC-CCI, (Sathyendranath et al., 2019) is also included, which is a merged dataset combining R_{rs} observations from multiple satellite missions (*e.g.*, SeaWiFS, MODIS, MERIS, VIIRS). Despite the multi-source merging and gap-filling procedures (Sathyendranath et al., 2019), it is found that this product has a higher standard deviation (0.00017 sr^{-1}) than that of MODIS-A R_{rs}(547) product after applying the CBAC or CHQR schemes, indicating a propagation of abnormal R_{rs} product from individual missions to the merged R_{rs} products (see the sharp R_{rs}(560) peak on April 28, 2005). These results underscore the effectiveness of CHQR in delivering temporally consistent and high-quality R_{rs} products. The abnormal values, if not filtered out, will not only be propagated to the merged product, but also significantly impact the evaluations of the long-term trend in the ocean's bio-optical properties.

3.3. Impact on data volume by CHQR

For satellite remote sensing, it is not just the quality of a data product that matters, but also the volume of high-quality data products, as the volume affects the effectiveness of such a high-cost platform. CHQR, as it takes multiple filtering criteria (see Fig. 1), inevitably will remove more satellite observations than the individual schemes. To get a measure of the impact of CHQR on data volume, we compared the retained products for data showing in Fig. 7. It is found that, among the four schemes, QA = 1 alone retained the most (~584 thousand), with CHQR retained ~415 thousand, which is just about 29% less than that of QA = 1 (or, ~25% less than that of l2-flags). To further assess CHQR's impact on a global scale, we collected 158 level-2 MODIS-Aqua images acquired on February 1, 2007, totaling approximately 46 million R_{rs} spectra. After applying CHQR, ~ 14.6 million data points were kept as the "highest" quality, which is ~60% of the data points when the l2-flags scheme was

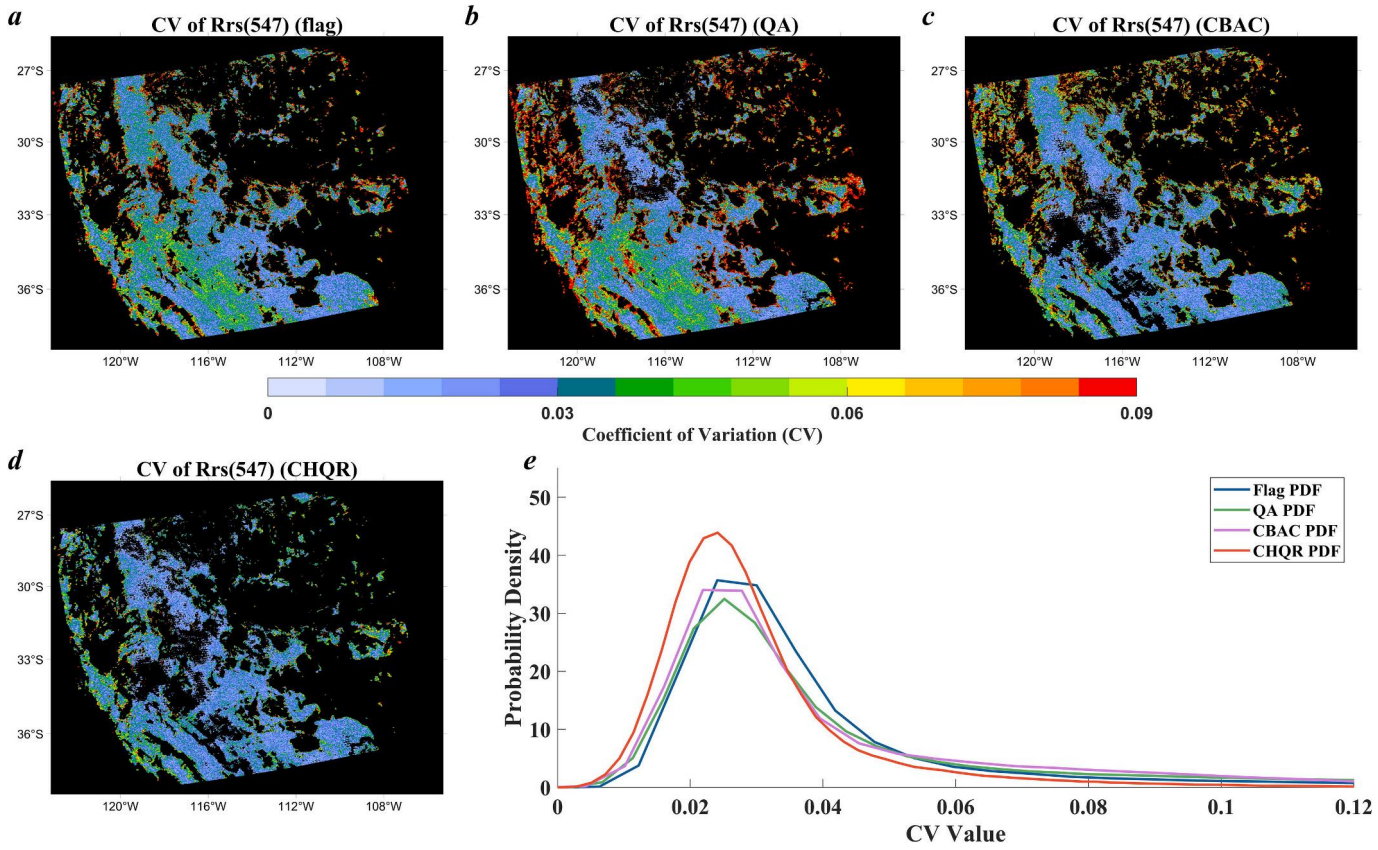

Fig. 7. Spatial distribution of the coefficient of variation (CV) of MODIS level-2 R_{rs}(547) on April 6, 2005, after applying four different screening schemes. (*a*) by l2-flags alone, (*b*) by QA = 1.0 alone, (*c*) by CBAC alone, and (*d*) by CHQR. (*e*) The probability density function (PDF) of CV values resulting from the four screening methods.

Table 2
Spatial variability of R_{rs}(547) in the South Pacific Gyre (120° W to 110° W and 25° S to 35° S) from April 6, 2005, after applying different screening methods.

Screening methods	Mean Coefficient of Variation (CV)	Standard Deviation of CV	Number of Data Points retained
l2-flags	0.042	0.039	559,855
QA = 1.0	0.050	0.048	584,198
CBAC	0.040	0.027	546,280
CHQR	0.030	0.016	415,029

applied (which retained ~24.6 million data points). These results indicate that CHQR achieves a reasonable balance between data quality and data volume, with a sufficient number of the "highest" quality satellite observations retained for further applications. On the other hand, as would be expected, more data (~70% of the filtered-out data) were excluded in high-latitude regions (>60° N/S) due to challenges in atmospheric correction under high solar zenith angles, fog, and other factors (He et al., 2018; Khanal and Wang, 2018), indicating a necessity to further improve the data products for such regions by the community.

Table 3
Statistics of R_{rs}(547) for data in Fig. 8.

Data after different screening methods	Mean R_{rs}(547) (sr^{-1})	Standard deviation of R_{rs}(547) (sr^{-1})
l2-flags	0.0018	0.00024
QA = 1.0	0.0019	0.00025
CBAC	0.0017	0.00012
CHQR	0.0017	0.00011
OC-CCI (560 nm)	0.0013	0.00017

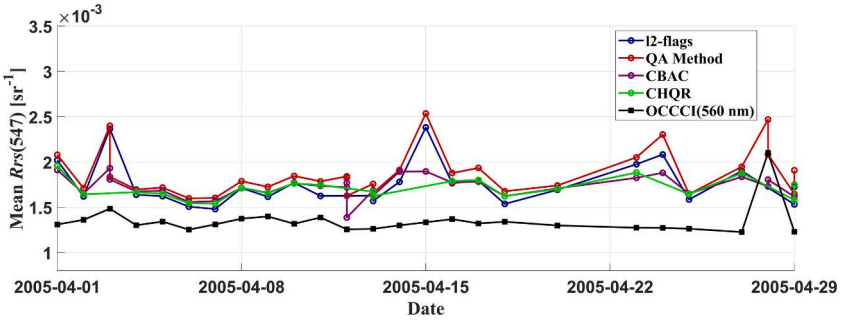

Fig. 8. Time series of R_{rs}(547) in the South Pacific Gyre after applying different screening methods.

4. Examples of using the products

There could be many valuable applications of this highest-quality R_{rs} database, such as its use in developing the cross-satellite atmospheric correction algorithm (Lee et al., 2024). Here, another example is presented. Utilizing the monthly composite R_{rs} data product from MODIS-A, covering the period from July 2002 to July 2024, we derive time series products from the original R_{rs} data product distributed by NASA and after applying CHQR, respectively. As case studies, we randomly selected two small areas in the Pacific Ocean, specifically 40–42°N, 140–138°W and 32–30°S, 124–122°W. To be consistent with the influential analysis presented in Cael et al. (2023), we also limited our study to the period of 2002 to 2022.

Fig. 9 shows the time series of monthly averaged $R_{rs}(547)$ before and after CHQR screening for the two selected 2° × 2° regions. For each case, only months in which more than 30% of the 4 km pixels contained non-missing R_{rs} values were retained for the analysis. Interestingly, as illustrated in Fig. 9, it is found that the long-term trends of $R_{rs}(547)$ obtained from the original data are completely opposite to those derived from the highest-quality data. Specifically, in Fig. 9*a*, the trend of $R_{rs}(547)$ from the original data is -3.0×10^{-6} sr^{-1} yr^{-1} ($P > 0.05$), while the trend from the highest-quality data is 2.1×10^{-6} sr^{-1} yr^{-1}, although both show no statistical significance ($P > 0.05$). Similarly, in Fig. 9*b*, the original data trend is -0.6×10^{-7} sr^{-1} yr^{-1} ($P > 0.05$), contrasting with a trend of 9.0×10^{-7} sr^{-1} yr^{-1} ($P > 0.05$) for the highest-quality data. Although these differences are not statistically significant ($P > 0.05$), the trend reversal highlights the potential for bias in long-term analyses due to the inclusion of low-quality or abnormal data. Note that the decreasing trends of both locations found with the original data are also presented in Cael et al. (2023). In calculating these trends, we employed the Cochrane-Orcutt method (Cael et al., 2023) to address autocorrelation issues, thereby enhancing the reliability of these results. The time series of $R_{rs}(547)$ showing in Fig. 9 *a* and *b* suggest that the contrasting trends are a result of low-quality or abnormal data (especially the extremely high peak around 2006 for the North Pacific) embedded in the originally distributed R_{rs} data products, which then result in biased signals and impact the reliability of long-term trend assessments.

Importantly, this pattern of opposing trends is not limited to a few specific areas. A broader analysis within the 50°S–50°N latitude band was conducted, where the highest-quality data are sufficient for long-term trend analysis. Based on long-term trend calculations of $R_{rs}(547)$ at a 2° × 2° grid resolution (Cael et al., 2023), 21.0% of the grids exhibited opposing trend directions before and after applying CHQR. This percentage was calculated relative to the total number of grids that contained more than 10 years of valid highest-quality data for trend analysis. Among trends with the same direction and exhibiting statistical significance ($P < 0.05$), the average relative difference in trend magnitude reached 20.3%, and 2.2% of these had the difference exceeding 100%. Such substantial variations—especially when averaged across large regions and over many days—can collectively distort large-scale interpretations of long-term oceanic changes. These findings emphasize that the abnormal values included in the original data products could introduce serious biases in long-term trend analyses.

5. Summary

In this study, we combined three filtering criteria for the determination and compilation of the highest-quality R_{rs} products, collectively termed as CHQR: 1) filtering based on observation condition or failure in data processing, 2) spectral shape consistency, and 3) minimum to no residual errors. These criteria, when applied jointly, ensure the selection of the highest-quality R_{rs} product. We thus applied CHQR to 20+ years of MODIS-Aqua measurements and compiled a global database of the highest-quality R_{rs}, termed as HQ_{MODISA}-R_{rs} (three formats: daily, 1 km resolution; 8-day, 4 km resolution; and monthly, 4 km resolution). Compared to databases created from *in situ* R_{rs}, in addition to avoiding the consistency issue in data quality among the R_{rs} measurements by different methods, HQ_{MODISA}-R_{rs} has enormously larger spatial and temporal coverages. The availability of such an R_{rs} database will be of great value for many studies in oceanography and ocean color remote sensing, such as the response of phytoplankton to climate change and cross-satellite calibration or atmospheric correction. With HQ_{MODISA}-R_{rs}, trending results in the literature obtained from data products without the application of CHQR could be reversed. These highlight the critical role of data quality in long-term trend analyses and underscore the importance of employing more rigorous quality control methods.

On the other hand, it is necessary to keep in mind that CHQR is designed for clear-ocean waters ($a(443) < 0.1$ m^{-1}; ~91% of the ocean) and removes more (~25–40%) data products compared to the present NASA operational l2-flags. For the other ~9% of global water, mostly coastal and highly productive waters, criteria to determine the highest-quality R_{rs} from MODIS-A measurements are lacking. In addition, the HQ_{MODISA}-R_{rs} database excludes data collected under unfavorable observational conditions, such as high solar zenith angles. This limitation particularly affects high-latitude regions, where high-quality data are relatively scarce. A truly global coverage HQ_{MODISA}-R_{rs} database demands great efforts for the development of robust criteria for

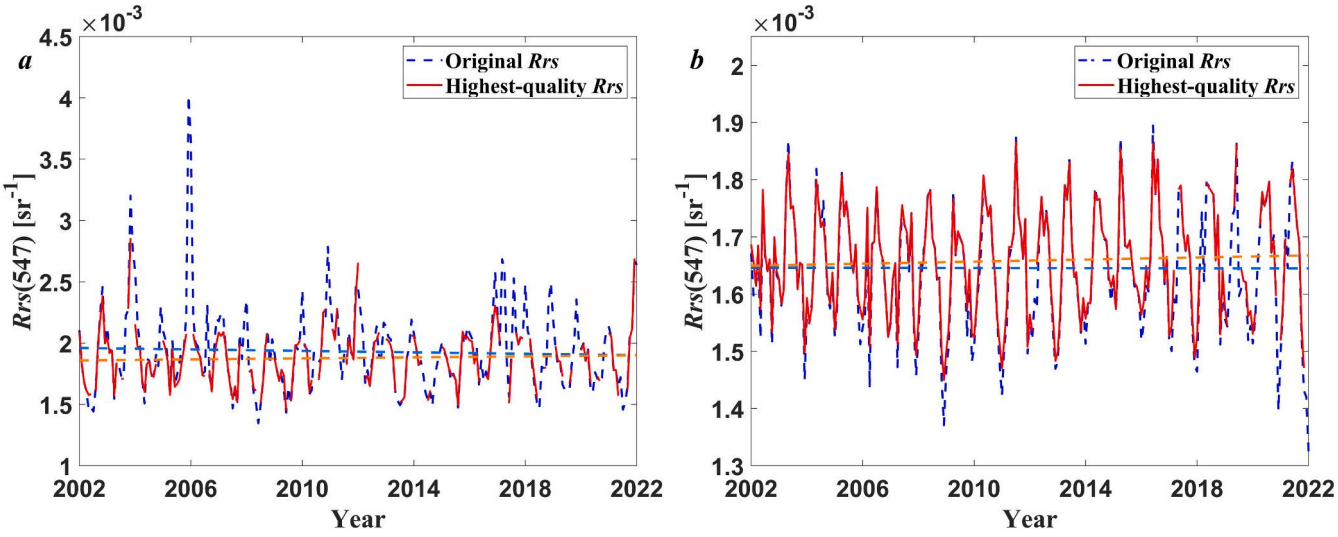

Fig. 9. Time series of $R_{rs}(547)$ for two localized areas in the Pacific Ocean: (*a*) 40–42°N, 140–138°W; (*b*) 32–30°S, 124–122°W.

objective determination of the highest-quality R_{rs} in coastal and high-latitude waters from MODIS-Aqua. Further, in the era of hyperspectral satellite ocean color remote sensing (*e.g.*, PACE, HY-1E), it is necessary and useful to expand the multi-band HQ_{MODISA}-R_{rs} data to a hyperspectral R_{rs} product. This would enable the extension of present-day hyperspectral observations back to earlier decades, offering long-term and continued hyperspectral data for ocean color studies.

Declaration of AI in the writing process

During the initial preparation of this work, the lead author used ChatGPT, an AI language tool, to improve the readability and language of the manuscript. After the initial draft, the authors carefully reviewed and significantly edited the content as needed and take full responsibility for the contents of the published article.

CRediT authorship contribution statement

Longteng Zhao: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation. **Zhongping Lee:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Xiaolong Yu:** Writing – review & editing. **Tianhao Wang:** Writing – review & editing. **Daosheng Wang:** Writing – review & editing, Data curation. **Shaoling Shang:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful for the financial support from the National Natural Science Foundation of China (#42430107, #42376173), the National Key Research and Development Program of China (#2023YFC3107603, #2024YFF0808901, #2024YFF0808902), the Fundamental Research Funds for the Central Universities (#20720240102), Fujian Satellite Data Development, Co., Ltd., Fujian Haisi Digital Technology Co., Ltd., and Fujian Ocean Innovation Laboratory. We thank Ms. Ying Su for her assistance during the execution of the above projects. We also acknowledge NASA for providing MODIS-Aqua ocean color products. We greatly appreciate the constructive comments and suggestions from three anonymous reviewers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2025.115195.

Data availability

The highest-quality R_{rs} data of monthly and 8-day, 4-km resolution can be found at https://doi.org/10.5281/zenodo.17877392.

References

Behrenfeld, M.J., O'Malley, R.T., Siegel, D.A., McClain, C.R., Sarmiento, J.L., Feldman, G.C., Milligan, A.J., Falkowski, P.G., Letelier, R.M., Boss, E.S., 2006. Climate-driven trends in contemporary ocean productivity. Nature 444, 752–755.

Boyer, T.P., Antonov, J.I., Baranova, O.K., Garcia, H.E., Johnson, D.R., Mishonov, A.V., O'Brien, T.D., Seidov, D., Smolyar, I., Zweng, M.M., Paver, C.R., Locarnini, R.A., Reagan, J.R., Forgy, C., Grodsky, A., & Levitus, S., 2013. World Ocean database 2013. In N.A.N. 72 (ed.). NOAA National Centers for environmental information.

Brown, S.W., Flora, S.J., Feinholz, M.E., Yarbrough, M.A., Houlihan, T., Peters, D., Kim, Y.S., Mueller, J.L., Johnson, B.C., Dennis, K., 2007. The marine optical BuoY (MOBY) radiometric calibration and uncertainty budget for ocean color satellite sensor vicarious calibration. In: Conference on Sensors, Systems, and Next-Generation Satellites XI. Florence, ITALY.

Cael, B., Bisson, K., Boss, E., Dutkiewicz, S., Henson, S., 2023. Global climate-change trends detected in indicators of ocean ecology. Nature 619, 551–554.

Carbotte, S.M., O'Hara, S., Stocks, K., Clark, P.D., Stolp, L., Smith, S.R., Briggs, K., Hudak, R., Miller, E., Olson, C.J., Shane, N., Uribe, R., Arko, R., Chandler, C.L., Ferrini, V., Miller, S.P., Doyle, A., Holik, J., 2022. Rolling deck to repository: supporting the marine science community with data management services from academic research expeditions. Frontiers in marine. Science 9.

Cazzaniga, I., Zibordi, G., 2023. AERONET-OC LWN uncertainties: revisited. J. Atmos. Ocean. Technol. 40, 411–425.

Chen, J., Lee, Z., Hu, C., Wei, J., 2016. Improving satellite data products for open oceans with a scheme to correct the residual errors in remote sensing reflectance. J. Geophys. Res. Oceans 121, 3866–3886.

Chen, J., Quan, W.T., Duan, H.T., Xing, Q.G., Xu, N., 2021. An improved inherent optical properties data processing system for residual error correction in turbid natural waters. Ieee J. Selected Top. App. Earth Observ. Remote Sens. 14, 6596–6607.

Chen, J., Li, J.F., He, X.Q., Tang, J.W., Pan, D.L., 2023. Neural network spectral relationship to improve an inherent optical properties data processing system for residual error correction. Opt. Express 31, 39583–39605.

Clark, D.K., Feinholz, M., Yarbrough, M., Johnson, B.C., Brown, S.W., Kim, Y.S., Barnes, R.A., 2002. Overview of the radiometric calibration of MOBY. In: In, Earth Observing Systems VI. SPIE, pp. 64–76.

Clark, D.K., Yarbrough, M.A., Feinholz, M., Flora, S., Broenkow, W., Kim, Y.S., Johnson, B.C., Brown, S.W., Yuen, M., Mueller, J.L., 2003. MOBY, a radiometric buoy for performance monitoring and vicarious calibration of satellite ocean color sensors: measurement and data analysis protocols. In: Ocean Optics Protocols for Satellite Ocean Color Sensor Validation. Special Topics in Ocean Optics Protocols and Appendices, vol. 6.

Claustre, H., Sciandra, A., Vaulot, D., 2008. Introduction to the special section bio-optical and biogeochemical conditions in the south East Pacific in late 2004: the BIOSOPE program. Biogeosciences 5, 679–691.

Esaias, W.E., Abbott, M.R., Barton, I., Brown, O.B., Campbell, J.W., Carder, K.L., Clark, D.K., Evans, R.H., Hoge, F.E., Gordon, H.R., Balch, W.M., Letelier, R., Minnett, P.J., 1998. An overview of MODIS capabilities for ocean science observations. IEEE Trans. Geosci. Remote Sens. 36, 1250–1265.

Fan, Y., Li, W., Chen, N., Ahn, J.-H., Park, Y.-J., Kratzer, S., Schroeder, T., Ishizaka, J., Chang, R., Stamnes, K., 2021. OC-SMART: a machine learning based data analysis platform for satellite ocean color sensors. Remote Sens. Environ. 253.

Franz, B.A., Bailey, S.W., Meister, G., Werdell, P.J., 2012. Quality and Consistency of the NASA Ocean Color Data Record. Proc, Ocean Optics XXI.

Gordon, H.R., Wang, M., 1994. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. Appl. Opt. 33, 443–452.

He, X., Stamnes, K., Bai, Y., Li, W., Wang, D., 2018. Effects of earth curvature on atmospheric correction for ocean color remote sensing. Remote Sens. Environ. 209, 118–133.

Hooker, S., Firestone, E.R., Patt, F.S., Barnes, R.A., Eplee Jr., R.E., Franz, B.A., Robinson, W.D., Feldman, G.C., Bailey, S.W., 2003. SeaBASS biogeochemical and bio-optical measurements data system. In: In. NASA Goddard Space Flight Center, Greenbelt, MD.

Hu, C.M., Voss, K.J., 1997. In situ measurements of Raman scattering in clear ocean water. Appl. Opt. 36, 6962–6967.

Hu, C.M., Feng, L., Lee, Z., Davis, C.O., Mannino, A., McClain, C.R., Franz, B.A., 2012. Dynamic range and sensitivity requirements of satellite ocean color sensors: learning from the past. Appl. Opt. 51, 6045–6062.

Hu, C., Feng, L., Lee, Z., 2013. Uncertainties of SeaWiFS and MODIS remote sensing reflectance: implications from clear water measurements. Remote Sens. Environ. 133, 168–182.

Hu, C.M., Feng, L., Guan, Q., 2021. A machine learning approach to estimate surface chlorophyll a concentrations in global oceans from satellite measurements. IEEE Trans. Geosci. Remote Sens. 59, 4590–4607.

IOCCG, 2010. Atmospheric correction for remotely-Sensed Ocean-colour products. In: MengHua, W. (Ed.), Reports of the International Ocean-Colour Coordinating Group. IOCCG, Dartmouth, Canada.

IOCGG, 2006. Remote sensing of inherent optical properties: Fundamentals, tests of algorithms, and applications. In: ZhongPing, L. (Ed.), Reports of the International Ocean-Colour Coordinating Group. IOCCG, Dartmouth, Canada.

Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2017. Theory-guided data science: a new paradigm for scientific discovery from data. IEEE Trans. Knowl. Data Eng. 29, 2318–2331.

Khanal, S., Wang, Z., 2018. Uncertainties in MODIS-based cloud liquid water path retrievals at high latitudes due to mixed-phase clouds and cloud top height inhomogeneity. J. Geophys. Res. Atmos. 123.

Le Traon, P.Y., Reppucci, A., Alvarez Fanjul, E., Aouf, L., Behrens, A., Belmonte, M., Bentamy, A., Bertino, L., Brando, V.E., Kreiner, M.B., 2019. From observation to information and users: the Copernicus marine service perspective. Front. Mar. Sci. 6, 234.

Lee, Z.P., Carder, K.L., Steward, R.G., Peacock, T.G., Davis, C.O., Patch, J.S., 1998. An empirical algorithm for light absorption by ocean water based on color. J. Geophys. Res. Oceans 103, 27967–27978.

Lee, Z.P., Carder, K.L., Arnone, R.A., 2002. Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. Appl. Opt. 41, 5755–5772.

Lee, Z.P., Ahn, Y.H., Mobley, C., Arnone, R., 2010. Removal of surface-reflected light for the measurement of remote-sensing reflectance from an above-surface platform. Opt. Express 18, 26313–26324.
Lee, Z., Hu, C., Shang, S., Du, K., Lewis, M., Arnone, R., Brewin, R., 2013. Penetration of UV-visible solar radiation in the global oceans: insights from ocean color remote sensing. J. Geophys. Res. Oceans 118, 4241–4255.
Lee, Z., Zhao, L., Hu, C., Wang, D., Lin, J., Shang, S., 2023. Absorption coefficient and chlorophyll concentration of oceanic waters estimated from band difference of satellite-measured remote sensing reflectance. J. Remote Sens. 3, 0063.
Lee, Z., Wang, T., Zhao, L., Wang, D., Ye, X., Shang, S., Yu, X., 2024. Cross-satellite atmospheric correction for consistent remote sensing reflectance from Multiple Ocean color satellites: concept and demonstrations. J. Remote Sens. 4, 0302.
Lehmann, M.K., Gurlin, D., Pahlevan, N., Alikas, K., Conroy, T., Anstee, J., Balasubramanian, S.V., Barbosa, C.C., Binding, C., Bracher, A., 2023. GLORIA-A globally representative hyperspectral in situ dataset for optical sensing of water quality. Sci. Data 10, 100.
Liu, X.Y., Yang, Q., Liu, Q.J., 2021. Adaptability analysis of various versions of GDPS based on QA score for GOCI data processing in the Yellow Sea. Spectrosc. Spectr. Anal. 41, 2233–2239.
Marchuk, G.I., Paton, B.E., Korotaev, G.K., Zalesny, V.B., 2013. Data-computing technologies: a new stage in the development of operational oceanography. Izvestiya Atmos. Ocean. Phys. 49, 579–591.
McClain, C.R., 2009. A decade of Satellite Ocean color observations. Annu. Rev. Mar. Sci. 1, 19–42.
Men, J., Feng, L., Chen, X., Tian, L.Q., 2023. Atmospheric correction under cloud edge effects for Geostationary Ocean color imager through deep learning. ISPRS J. Photogramm. Remote Sens. 201, 38–53.
Morel, A., Prieur, L., 1977. Analysis of variations in ocean color 1. Limnol. Oceanogr. 22, 709–722.
Morel, A., Antoine, D., Gentili, B., 2002. Bidirectional reflectance of oceanic waters: accounting for Raman emission and varying particle scattering phase function. Appl. Opt. 41, 6289–6306.
Morel, A., Gentili, B., Claustre, H., Babin, M., Bricaud, A., Ras, J., Tièche, F., 2007. Optical properties of the “clearest” natural waters. Limnol. Oceanogr. 52, 217–229.
O'Reilly, J.E., Werdell, P.J., 2019. Chlorophyll algorithms for ocean color sensors-OC4, OC5 & OC6. Remote Sens. Environ. 229, 32–47.
Pauthenet, E., Martinez, E., Gorgues, T., Roussillon, J., Drumetz, L., Fablet, R., Roux, M., 2024. Contrasted trends in chlorophyll a satellite products. Geophys. Res. Lett. 51.
Rast, M., Bézy, J.L., Bruzzi, S., 1999. The ESA medium resolution imaging spectrometer MERIS -: a review of the instrument and its mission. Int. J. Remote Sens. 20, 1681–1702.
Ruddick, K.G., Voss, K., Boss, E., Castagna, A., Frouin, R., Gilerson, A., Hieronymi, M., Johnson, B.C., Kuusk, J., Lee, Z., Ondrusek, M., Vabson, V., Vendt, R., 2019. A review of protocols for fiducial reference measurements of water-leaving radiance for validation of satellite remote-sensing data over water. Remote Sens. 11, 2198.
Sathyendranath, S., Brewin, R.J., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., Cipollini, P., Couto, A.B., Dingle, J., Doerffer, R., 2019. An ocean-colour time series for use in climate studies: the experience of the ocean-colour climate change initiative (OC-CCI). Sensors 19, 4285.
Schmidtko, S., Stramma, L., Visbeck, M., 2017. Decline in global oceanic oxygen content during the past five decades. Nature 542, 335–339.
Schwing, F.B., 2023. Modern technologies and integrated observing systems are “instrumental” to fisheries oceanography: a brief history of ocean data collection. Fish. Oceanogr. 32, 28–69.
Sidirourgos, L., Kersten, M., Boncz, P., 2013. Scientific discovery through weighted sampling. In: IEEE International Conference on Big Data (Big Data), pp. 300–306. Santa Clara, CA, USA.
Tanhua, T., Pouliquen, S., Hausman, J., O'Brien, K., Bricher, P., de Bruin, T., Buck, J.J. H., Burger, E.F., Carval, T., Casey, K.S., Diggs, S., Giorgetti, A., Glaves, H., Harscoat, V., Kinkade, D., Muelbert, J.H., Novellino, A., Pfeil, B., Pulsifer, P.L., Van de Putte, A., Robinson, E., Schaap, D., Smirnov, A., Smith, N., Snowden, D., Spears, T., Stall, S., Tacoma, M., Thijsse, P., Tronstad, S., Vandenberghe, T., Wengren, M., Wyborn, L., Zhao, Z.M., 2019. Ocean FAIR data services. Front. Mar. Sci. 6.
Valente, A., Sathyendranath, S., Brotas, V., Groom, S., Grant, M., Taberner, M., Antoine, D., Arnone, R., Balch, W.M., Barker, K., Barlow, R., Bélanger, S., Berthon, J. F., Besiktepe, S., Borsheim, Y., Bracher, A., Brando, V., Canuti, E., Chavez, F., Cianca, A., Claustre, H., Clementson, L., Crout, R., Frouin, R., García-Soto, C., Gibb, S.W., Gould, R., Hooker, S.B., Kahru, M., Kampel, M., Klein, H., Kratzer, S., Kudela, R., Ledesma, J., Loisel, H., Matrai, P., McKee, D., Mitchell, B.G., Moisan, T., Muller-Karger, F., O'Dowd, L., Ondrusek, M., Platt, T., Poulton, A.J., Repecaud, M., Schroeder, T., Smythe, T., Smythe-Wright, D., Sosik, H.M., Twardowski, M., Vellucci, V., Voss, K., Werdell, J., Wernand, M., Wright, S., Zibordi, G., 2019. A compilation of global bio-optical in situ data for ocean-colour satellite applications - version two. Earth Syst. Sci. Data 11, 1037–1068.
van Oostende, M., Hieronymi, M., Krasemann, H., Baschek, B., Röttgers, R., 2022. Correction of inter-mission inconsistencies in merged ocean colour satellite data. Front. Remote Sens. 3, 882418.
Wang, H.Y., Li, X.F., 2024. DeepBlue: advanced convolutional neural network applications for ocean remote sensing. Ieee Geosci. Remote Sens. Magaz. 12, 138–161.
Wei, J.W., Lee, Z.P., Shang, S.L., 2016. A system to measure the data quality of spectral remote-sensing reflectance of aquatic environments. J. Geophys. Res. Oceans 121, 8189–8207.
Wei, G.M., Lee, Z.P., Wu, X.L., Yu, X.L., Shang, S.L., Letelier, R., 2021. Impact of temperature on absorption coefficient of pure seawater in the blue wavelengths inferred from satellite and in situ measurements. J. Remote Sens. 2021.
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, 122–140.
Werdell, P.J., Fargion, G.S., McClain, C.R., Bailey, S.W., 2002. The SeaWiFS bio-optical archive and storage system (SeaBASS): Current architecture and implementation. In: In. NASA Goddard Space Flight Center, Greenbelt, MD.
Werdell, P.J., Franz, B.A., Bailey, S.W., Feldman, G.C., Boss, E., Brando, V.E., Dowell, M., Hirata, T., Lavender, S.J., Lee, Z.P., Loisel, H., Maritorena, S., Mélin, F., Moore, T.S., Smyth, T.J., Antoine, D., Devred, E., d'Andon, O.H.F., Mangin, A., 2013. Generalized Ocean color inversion model for retrieving marine inherent optical properties. Appl. Opt. 52, 2019–2037.
Whang, S.E., Roh, Y., Song, H., Lee, J.-G., 2023. Data collection and quality challenges in deep learning: a data-centric AI perspective. VLDB J. 32, 791–813.
Zhang, M.W., Ibrahim, A., Franz, B.A., Ahmad, Z., Sayer, A.M., 2022. Estimating pixel-level uncertainty in ocean color retrievals from MODIS. Opt. Express 30, 31415–31438.
Zhang, Y., Shen, F., Sun, X., Tan, K., 2023. Marine big data-driven ensemble learning for estimating global phytoplankton group composition over two decades (1997–2020). Remote Sens. Environ. 294, 113596.
Zibordi, G., Holben, B.N., Talone, M., D'Alimonte, D., Slutsker, I., Giles, D.M., Sorokin, M.G., 2021. Advances in the ocean color component of the aerosol robotic network (AERONET-OC). J. Atmos. Ocean. Technol. 38, 725–746.