



Technical note: Euclidean Distance Score (EDS) for algorithm performance assessment in aquatic remote sensing

Amanda de Liz Arcari¹, Juliana Tavora², Daphne van der Wal^{1,3}, and Mhd. Suhyb Salama¹

¹Department of Water Resources, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, OV, 7522 NH, Netherlands

²IFREMER, DYNECO, Hydrosedimentary Dynamics Laboratory (DHYSED), 29280 Plouzané, France

³Department of Estuarine and Delta Systems, NIOZ Royal Netherlands Institute for Sea Research, Yerseke, ZE, 4401 NT, Netherlands

Correspondence: Amanda de Liz Arcari (a.delizarcari@utwente.nl)

Abstract. In the absence of community consensus, there remains a gap in standardized, consistent performance assessment of remote-sensing algorithms for water-quality retrieval. Although the use of multiple metrics is common, whether reported individually or combined into scoring systems, approaches are often constrained by statistical limitations, redundancy, and dataset- and context-dependent normalizations, leading to subjective or inconsistent interpretations. To address this, we propose the Euclidean Distance Score (EDS), which integrates five statistically appropriate and complementary metrics into a composite score. Capturing three core aspects of performance (regression fit, retrieval error, and robustness), EDS is computed as the Euclidean distance from an idealized point of perfect performance, providing a standardized and interpretable measure. We demonstrate the applicability of EDS in three scenarios: assessing a single algorithm for different retrieved variables, comparing two algorithms on shared retrievals, and evaluating performance across contrasting trophic conditions. By offering an objective framework, EDS supports consistent validation of aquatic remote sensing algorithms and transparent comparisons in varied contexts.

1 Introduction

Assessing the quality of remote sensing products is vital to monitoring and understanding aquatic ecosystems. The growing availability of algorithms that estimate water quality variables from optical signals, along with their varying transferability across regions and conditions, underscores the need for rigorous performance metrics to evaluate them (Werdell et al., 2018). In the absence of consensus in the community, studies often report multiple metrics (e.g., Pearson's r , r^2 , slope, mean absolute error, mean percentage error, bias, root-mean-squared error) to quantify agreement between retrievals and observations. While reporting several metrics is important to capture complementary aspects of performance (Stow et al., 2009; IOCCG, 2019), some are partially redundant in what they represent, and not all are well-suited for the statistical properties of bio-optical data (Werdell et al., 2018). This can introduce subjectivity and risk biased or inconsistent assessments (Simão et al., 2024; Seegers et al., 2018).



Bio-optical data (i.e., optical proxies of biogeochemical state such as absorption, scattering, attenuation coefficients, and constituent concentrations) typically follow log-normal distributions and span wide dynamic ranges (Campbell, 1995). Consequently, several common error metrics are less appropriate. For example, root-mean-squared error is most informative when residuals are approximately Gaussian and homoscedastic. Because it squares residuals, it disproportionately penalizes large deviations and becomes sensitive to skew and outliers (Morley et al., 2018). Therefore, for this type of data, simple deviation metrics are preferred over sums of squares (Seegers et al., 2018), and calculations on log-transformed values are recommended (IOCCG, 2019; EUMETSAT, 2021; Mélin and Franz, 2014). Additional good practices include using medians to reduce outlier influence, ensuring symmetric treatment of over- and under-prediction, and favoring relative errors to maintain meaning across large ranges (O'Shea et al., 2023; Pahlevan et al., 2022).

To mitigate subjectivity from multiple standalone metrics, scoring-based and comparative approaches have been explored. Brewin et al. (2015) introduced a composite score aggregating multiple metrics, later adapted by Neil et al. (2019). However, potential redundancy among metrics may bias comparisons (Seegers et al., 2018). Seegers et al. (2018) also proposed a pairwise comparison metric (% wins), identifying the algorithm with the lowest residual error per observation. The method also accounts for algorithm robustness, as failed retrievals penalize the score. While valuable for ranking multiple algorithms, these comparative methods are less suited for evaluating a single algorithm across variables, environmental conditions, or datasets.

Graphical multidimensional tools are also used to summarize performance across metrics. Taylor diagrams have supported cross-dataset comparisons in aquatic remote sensing (Arabi et al., 2020), but are limited to a fixed set of metrics and lack scoring capability. Radar/star plots offer more flexibility and are increasingly used to derive composite scores: Tran et al. (2023) defined a summary indicator based on the area enclosed by polygons linking normalized metrics; Simão et al. (2024) incorporated the distance from the polygon centroid to the center to reward both high and balanced performance; Subirade et al. (2024) summed normalized scores to rank algorithms. These enhance within-study interpretation, but reliance on normalization constrains broader applicability: scores become tied to dataset-specific ranges. Shifts in value distributions or context (e.g., different optical water types/trophic states, oceanic vs. coastal) change the normalization baseline, and heterogeneity in variables and units further limits comparability. As a result, such indices remain inherently relative and are not comparable across broader contexts.

Considering these challenges, we propose the Euclidean Distance Score (EDS), designed to provide a standardized and broadly applicable framework for evaluating aquatic remote sensing algorithms. EDS quantifies performance as the Euclidean distance from an idealized result, integrating complementary metrics selected for their suitability to bio-optical data. By avoiding redundancy and dataset-specific normalization, EDS supports consistent evaluation of water quality retrievals and strengthens the reliability of remote sensing products that support aquatic biogeochemistry and ecosystem monitoring. The following sections review Euclidean distance in composite scoring (2), detail metric selection (3), introduce the EDS formulation (2), and demonstrate its application (5).



2 Euclidean Distance in Composite Scores

55 The Euclidean distance quantifies the separation between two points in n -dimensional space. For points $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$, it is defined as:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (1)$$

This provides a straightforward and interpretable measure of dissimilarity and forms the basis for composite performance scores across scientific domains. When multiple performance indicators are treated as spatial dimensions, a single score can be derived from their distance to an idealized reference point.

In hydrology, the widely used Kling-Gupta Efficiency (KGE) (Gupta et al., 2009) combines correlation, bias, and variability into one score, computed as the Euclidean distance from a perfect model. Similarly, Hu et al. (2019) introduced the DISO metric, which integrates correlation, absolute error, and RMSE to evaluate climate model performance (Hu et al., 2019). These examples demonstrate the usefulness of Euclidean distance in synthesizing multiple indicators into a robust, interpretable score.

65 Developing effective composite scores requires careful selection of individual components. Chosen metrics should be relevant to the application domain and reflect distinct performance aspects to avoid redundancy. While equal weighting supports balance, formulations could also be adapted to emphasize specific aspects based on application needs. Additionally, limiting the number of metrics helps preserve robustness, as high dimensionality can reduce the reliability of distance-based scores (Samaratay et al., 2024). The next section outlines the rationale for selecting the individual performance metrics to be used in our composite approach.

3 Selection of Individual Metrics

For the construction of our composite score, we define three core dimensions of retrieval performance: (A) regression fit, which reflects how well the estimated values follow the pattern of the observed data; (B) retrieval error, which quantifies the magnitude and direction of deviations between estimates and observations; and (C) retrieval robustness, which describes the algorithm's ability to produce valid outputs consistently. For each dimension, we selected metrics suited to characterizing performance in the context of bio-optical data.

3.1 Regression Fit

To characterize the regression fit between estimated and observed values, we employ a Type II regression (reduced major axis) to account for uncertainty in both in-situ and retrieved values. Calculations are performed in logarithmic space to accommodate the log-normal distribution of bio-optical data (Campbell, 1995). The following metrics are selected:

- **Pearson correlation coefficient (r):** Quantifies the strength and consistency of the linear association between log-transformed estimated and observed values.



- **Slope (m):** Describes the proportionality of estimates relative to observations, indicating whether the dynamic range is well represented.

85 3.2 Retrieval Error

To characterize retrieval error, we selected metrics that quantify the typical magnitude of deviation and the systematic bias between estimated and observed values. We adopt two metrics proposed by Morley et al. (2018) to address limitations of traditional measures such as root-mean-squared error and mean average percentage error (Salama et al., 2022), which are advised for variables that span over orders of magnitude. These have been increasingly adopted in recent aquatic remote sensing studies (O'Reilly and Werdell, 2019; O'Shea et al., 2021, 2023; Pahlevan et al., 2020). Advantages of these metrics include their relative (percentage) format for interpretability and comparability, symmetric treatment of over- and under-estimations, and robustness to outliers and skewed distributions (Morley et al., 2018).

Both metrics are based on the accuracy ratio (Q_i), defined as:

$$Q_i = \frac{E_i}{O_i} \quad (2)$$

95 where E_i and O_i are the estimated and observed values, respectively.

Based on the accuracy ratio, the retrieval error metrics are calculated as follows:

- **Median Symmetric Accuracy (ϵ):** Measures the typical relative error between estimates and observations. It is defined as:

$$\epsilon = 10^{\text{median}_{i=1,\dots,n}(|\log_{10}(Q_i)|)} - 1 \quad (3)$$

- 100 – **Signed Symmetric Bias (β):** Quantifies the systematic tendency of the algorithm to overestimate or underestimate observations. The median of the logarithmic accuracy ratio is first defined as:

$$M = \text{median}_{i=1,\dots,n}(\log_{10}(Q_i)) \quad (4)$$

Using this definition, β is computed as:

$$\beta = \text{sign}(M) \times (10^{|M|} - 1) \quad (5)$$

105 It should be noted that although Morley et al. (2018) formulated these metrics using the natural logarithm, the choice of logarithm base is arbitrary. Furthermore, since $\log(Q) = \log(E) - \log(O)$, these metrics are conceptually analogous to the mean absolute error and bias calculations as recommended by Seegers et al. (2018). The differences are the use of median instead of mean aggregation (as the distribution of $\log(Q)$ may not be symmetric (Morley et al., 2018)), and that 1 is directly subtracted to rescale the value into a percentage deviation term.



110 3.3 Retrieval Robustness Metrics

Retrieval robustness characterizes the algorithm's ability to consistently produce valid outputs across the dataset. In real-world remote sensing applications, retrieval algorithms may fail to converge or generate valid estimates under challenging observational conditions. To quantify this aspect of performance, we use:

– Valid Retrieval Ratio (n):

$$115 \quad n = \frac{n_{\text{valid}}}{n_{\text{total}}} \quad (6)$$

where n_{valid} is the number of retrievals yielding valid outputs and n_{total} is the total number of retrieval attempts.

4 The Proposed Euclidean Distance Score (EDS)

The Euclidean Distance Score (EDS) aggregates the selected individual metrics into a single composite score that quantifies the overall retrieval performance by calculating the Euclidean distance from an idealized result. The score is computed as:

$$120 \quad \text{EDS} = 1 - \sqrt{(r-1)^2 + (m-1)^2 + \epsilon^2 + \beta^2 + (n-1)^2} \quad (7)$$

In this formulation, an ideal algorithm would achieve $r = 1$, $m = 1$, $\epsilon = 0$, $\beta = 0$, and $n = 1$, yielding an EDS of 1. As deviations from these values increase, the EDS decreases proportionally, reflecting poorer algorithm performance.

To help interpret the EDS, benchmarks can be set based on typical thresholds for acceptable retrieval performance. For example, with ϵ and β at 30% (commonly accepted error in aquatic remote sensing (IOCCG, 2019)), $r = 0.5$, slope = 1.2 (or
125 0.8), and $n = 0.9$, the EDS is approximately 0.31. This suggests that results exceeding this benchmark could be considered generically indicative of good retrieval performance in this context.

EDS can be extended to evaluate overall performance when algorithms retrieve multiple variables by averaging individual EDS values into an aggregated score. This approach enables flexible comparisons depending on evaluation priorities: if one variable is of primary interest, the highest individual EDS can guide selection; if balanced performance is desired, the aggregated EDS identifies the most consistent algorithm across variables. Additionally, weights can be applied to reflect variable
130 importance, allowing tailored assessments to specific application needs.

4.1 Relative influence of performance components

The design of the EDS balances three aspects of retrieval performance: regression fit, retrieval error, and retrieval robustness. When errors (ϵ and β) are below 1 (i.e., within 100%), the slope (m) is within ± 1 of the ideal (i.e., $0 \leq m \leq 2$), and the
135 correlation is positive ($r \geq 0$), the maximum contribution to the distance calculation is 2 for regression fit (if $r = 0$ and $m = 0$ or $m = 2$), 2 for retrieval error (if $\epsilon = 1$ and $\beta = 1$), and 1 for retrieval robustness (if $n = 0$). This means that regression fit and retrieval error typically weigh equally and collectively more than retrieval robustness, prioritizing output quality while



still accounting for algorithmic stability. However, if errors or slope deviations are excessive, or correlation is negative, these factors can contribute disproportionately to the distance calculation, strongly penalizing the final EDS.

140 4.2 Metrics dependency

The EDS is designed to capture complementary aspects of retrieval performance while minimizing metric redundancy. Although the selected metrics are conceptually independent, correlations may still arise in practice due to shared responses to specific data patterns or algorithm behavior. For example, ϵ and β may co-vary when there is consistent over- or underestimation, and in some datasets, r and m may show strong empirical correlation. However, in reduced major axis regression, m is
145 not directly dependent on r , as it is governed by the ratio of standard deviations and only adopts the sign of r (Smith, 2009). Such empirical correlations do not indicate conceptual redundancy, as each metric captures a distinct and relevant aspect of retrieval performance.

5 Application Examples

We demonstrate the EDS by evaluating bio-optical retrievals against in situ reference data, focusing on inherent optical properties (IOPs) commonly used to characterize water quality.
150

The analysis illustrates three main use cases of the EDS framework. First, we evaluate the retrieval of different variables using a single algorithm, including three IOPs—phytoplankton absorption ($a_\phi(443)$), detrital and gelbstoff absorption ($a_{dg}(443)$), and particulate backscattering ($b_{bp}(443)$)—and the diffuse attenuation coefficient ($K_d(489)$) (a quasi-IOP (Yu et al., 2016)). Second, we compare the performance of two algorithms retrieving the same variables. Finally, we examine how performance
155 for $K_d(489)$ varies across water types, defined by chlorophyll- a thresholds (0.1 and 1 $\text{mg}\cdot\text{m}^{-3}$) for oligotrophic, mesotrophic, and eutrophic conditions.

5.1 Dataset and Processing

The dataset used is the NASA bio-Optical Marine Algorithm Dataset (NOMAD) (Werdell and Bailey, 2005), a publicly available, globally distributed collection of 4,459 in situ bio-optical observations across diverse aquatic environments. Each record
160 includes spectral upwelling radiance (L_w) and downwelling irradiance (E_s) at 21 nominal wavelengths (411–683 nm), alongside coincident optical property measurements and metadata.

Remote sensing reflectance (R_{rs}) was calculated as L_w/E_s . The IOPs of interest— $a_\phi(443)$, $a_{dg}(443)$, and $b_{bp}(443)$ —were derived using standard relationships. Solar geometry was computed from coordinates and time, with a fixed sensor viewing geometry (40° zenith, 135° azimuth relative to the sun (Mobley, 1999)).

165 R_{rs} spectra and geometry serve as input to 2SeaColor algorithm (Salama and Verhoef, 2015), to retrieve the three IOPs and $K_d(489)$, and to the Quasi-Analytical Algorithm (QAA) (Lee et al., 2002), to retrieve only the IOPs.

Post-processing involved filtering retrievals with variable-specific thresholds to exclude unrealistic values and define invalid retrievals:



- $a_{\phi}(443)$: 0.0001–11 m⁻¹
- $a_{dg}(443)$: 0.0001–9 m⁻¹
- $b_{bp}(443)$: 0.000215–12 m⁻¹
- $k_d(489)$: 0.001–10 m⁻¹

170

175 limits are based on extreme values reported in the literature (Jorge et al., 2021).

Finally, the EDS is computed from valid retrievals, integrating the individual metrics described in Section II.

5.2 Use Case 1: Evaluating Variable-Specific Retrieval Performance

The first use case demonstrates how EDS captures performance differences across multiple retrieval targets using 2SeaColor (Fig. 1). Among the evaluated variables, $K_d(489)$ achieved the highest EDS (0.80), reflecting excellent overall performance with perfect convergence (3346/3346), low errors, and strong correlation. In contrast, $b_{bp}(443)$ shows the lowest EDS (0.06), despite low bias ($\beta = -4\%$) and acceptable median symmetric accuracy ($\epsilon = 24\%$). The poor score is mostly driven by an exaggerated slope (1.86). Intermediate scores for $a_{\phi}(443)$ (0.45) and $a_{dg}(443)$ (0.25) reflect strong correlation but higher errors.

180

5.3 Use Case 2: Comparing Algorithms

This second use case illustrates how EDS supports direct comparison of algorithm performance for the same variables. Fig. 2 shows QAA retrievals for $a_{\phi}(443)$, $a_{dg}(443)$, and $b_{bp}(443)$, while 2SeaColor results are in Fig. 1. For $a_{\phi}(443)$, both algorithms achieve the same EDS (0.45), despite differences in individual metrics. QAA performs better for $a_{dg}(443)$ (EDS = 0.41 vs. 0.25), with stronger correlation and better alignment with the 1:1 line. For $b_{bp}(443)$, both perform poorly, but QAA yields worse results due to a steeper slope and stronger negative bias ($\beta = -37\%$ vs. -4%).

190

Aggregated EDS scores are nearly identical (0.25 for 2SeaColor and 0.24 for QAA) indicating comparable overall performance. However, if one variable is particularly important for a given application, the algorithm with the highest EDS for that variable should be preferred.

5.4 Use Case 3: Evaluating Retrieval Performance Across Water Types

The third use case illustrates how EDS enables performance evaluation across water types with varying dynamic ranges. Fig. 3 shows $K_d(489)$ results grouped by trophic state: oligotrophic, mesotrophic, and eutrophic. In oligotrophic waters, EDS is 0.50 despite the lowest correlation ($r = 0.55$), driven by low errors. Mesotrophic waters show the highest overall performance (EDS = 0.76), with a balanced metric profile. Eutrophic waters also yield high performance (EDS = 0.73), with the highest r due to greater variability, though increased scatter is reflected in a higher ϵ .

195

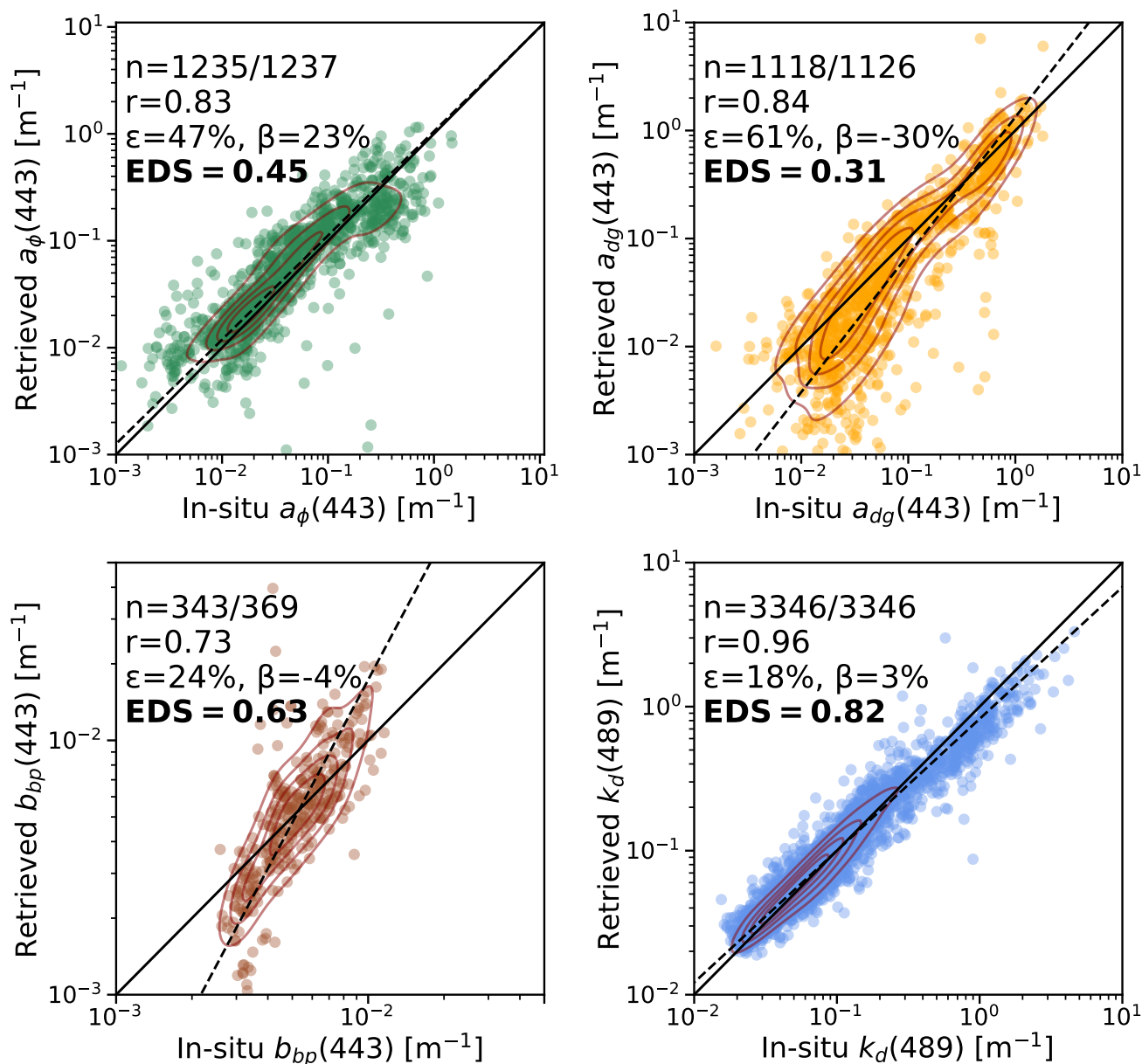


Figure 1. Scatterplots of in situ vs. retrieved values using 2SeaColor, with annotations for the correlation coefficient (r), regression slope (m), median symmetric accuracy (ϵ), signed symmetric bias (β), valid retrieval ratio (n , valid retrievals/all input points), and the Euclidean Distance Score (EDS). Red contours indicate data density.

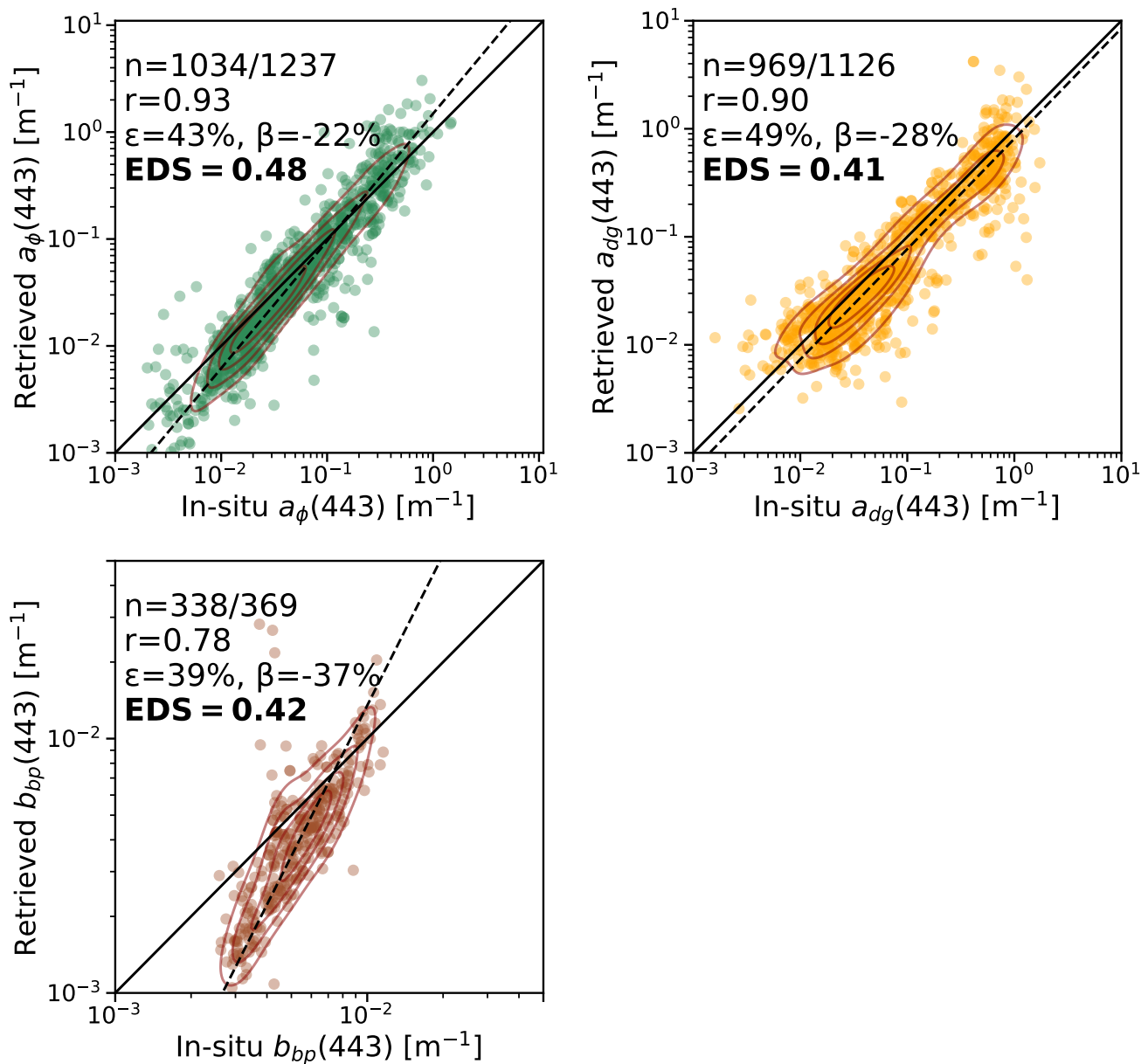


Figure 2. Scatterplots of in situ vs. retrieved values using QAA. Panel annotations follow the same convention as in Figure 1.

6 Conclusions

200 The Euclidean Distance Score (EDS) provides an objective framework for synthesizing algorithm performance in aquatic remote sensing. By integrating regression fit, retrieval error, and robustness into a single, interpretable score, we show how EDS

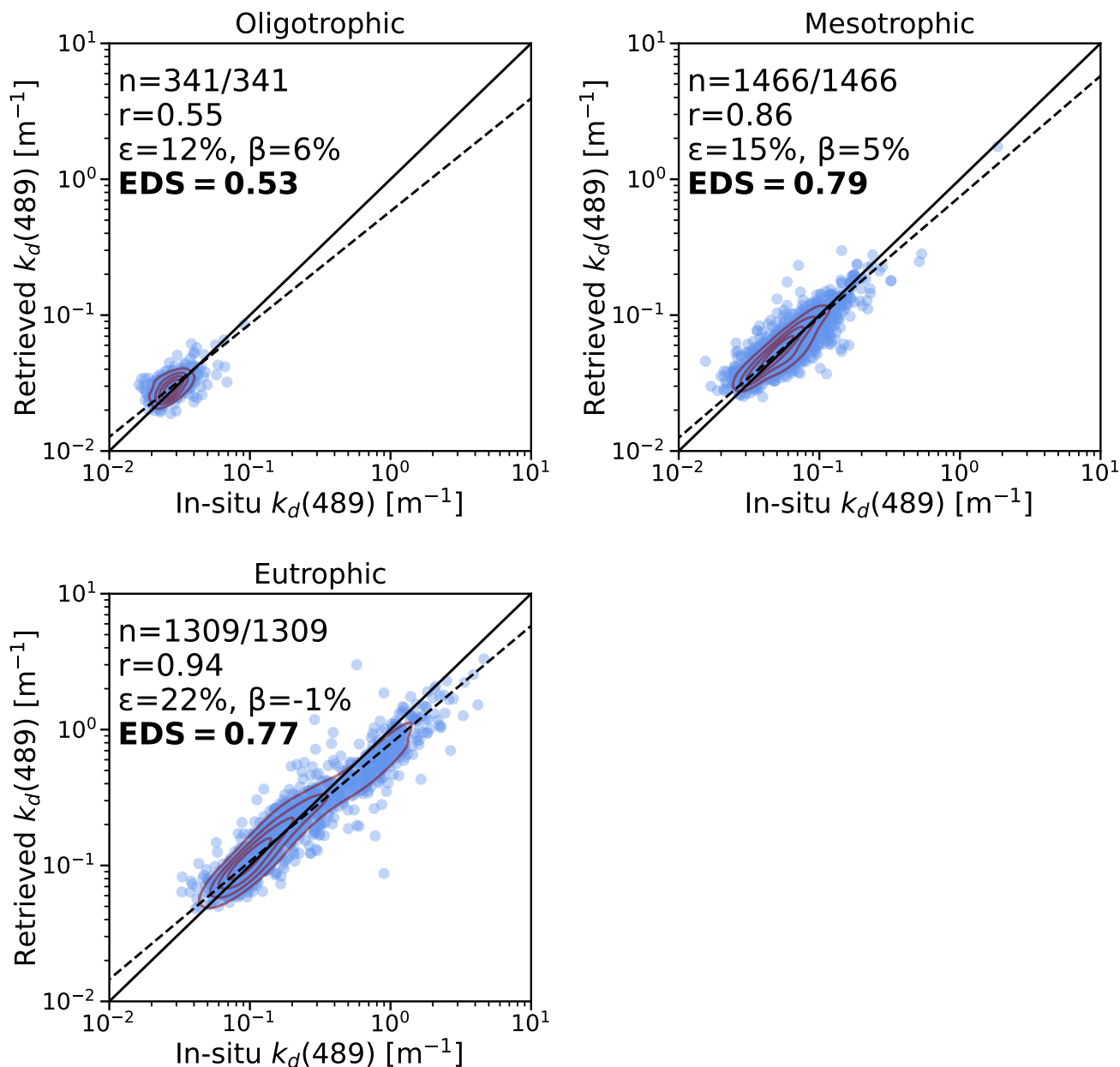


Figure 3. Scatterplots of in situ vs. retrieved $K_d(489)$ for three water types: oligotrophic, mesotrophic, and eutrophic. Panel annotations follow the same convention as in Figure 1.

enables consistent comparisons across variables, algorithms and environmental contexts. Its transparent formulation based on established metrics supports reproducibility, while its simplicity facilitates broader adoption in both research and operational



205 contexts. In addition, the framework can be adapted to emphasize specific aspects of performance when required, ensuring flexibility across diverse applications. Although EDS offers a unified view of performance and supports decision-making, reporting individual metrics remains essential for diagnosing specific retrieval limitations and guiding targeted improvements of algorithms.

Data availability. The bio-Optical Marine Algorithm Dataset (NOMAD) used in this study is publicly available from NASA's Ocean Biology Processing Group at: <https://seabass.gsfc.nasa.gov/wiki/NOMAD>

210 *Author contributions.* Amanda de Liz Arcari: conceptualization, methodology, software, formal analysis, data curation, visualization, and writing – original draft. Juliana Tavora: validation and writing – review and editing. Daphne van der Wal: supervision and writing – review and editing. Mhd. Suhyb Salama: conceptualization, methodology, supervision, funding acquisition, and writing – review and editing.

Competing interests. The authors declare that they have no conflict of interest.

215 *Acknowledgements.* The authors thank the NASA Ocean Biology Processing Group and the individual data contributors for compiling, maintaining, and providing public access to the SeaWiFS Bio-optical Archive and Storage System (SeaBASS), and in particular the NASA bio-Optical Marine Algorithm Dataset (NOMAD). Cruise-level details and contributor information are available through the SeaBASS database (<https://seabass.gsfc.nasa.gov/wiki/NOMAD>).



References

- Arabi, B., Salama, M. S., Pitarch, J., and Verhoef, W.: Integration of in-situ and multi-sensor satellite observations for long-term water quality
 220 monitoring in coastal areas, *Remote Sensing of Environment*, 239, 111 632, <https://doi.org/10.1016/j.rse.2020.111632>, 2020.
- Brewin, R. J., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P. Y., Devred, E., Doerffer, R., Fomferra, N., Franz, B., Grant, M., Groom, S., Horseman, A., Hu, C., Krasemann, H., Lee, Z. P., Maritorena, S., Mélin, F., Peters, M., Platt, T., Regner, P., Smyth, T., Steinmetz, F., Swinton, J., Werdell, J., and White, G. N.: The Ocean Colour Climate Change Initiative: III. A round-robin comparison on in-water bio-optical algorithms, *Remote Sensing of Environment*, 162, 271–294, <https://doi.org/10.1016/J.RSE.2013.09.016>, 2015.
- 225 Campbell, J. W.: The lognormal distribution as a model for bio-optical variability in the sea, *Journal of Geophysical Research*, 100, 13 237–13 254, <https://doi.org/10.1029/95JC00458>;WGROU:STRING:PUBLICATION, 1995.
- EUMETSAT: Recommendations for Sentinel-3 OLCI Ocean Colour product validations in comparison with in situ measurements-Matchup Protocols, Tech. rep., EUMETSAT, Darmstadt, ISBN 4961518077, 2021.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria:
 230 Implications for improving hydrological modelling, *Journal of Hydrology*, 377, 80–91, <https://doi.org/10.1016/J.JHYDROL.2009.08.003>, 2009.
- Hu, Z., Chen, X., Zhou, Q., Chen, D., and Li, J.: DISO: A rethink of Taylor diagram, *International Journal of Climatology*, 39, 2825–2832, <https://doi.org/10.1002/JOC.5972>, 2019.
- IOCCG: Uncertainties in Ocean Colour Remote Sensing, in: Reports of the International Ocean-Colour Coordinating Group, edited
 235 by Mélin, F., vol. No. 18 of *Reports of the International Ocean Colour Coordinating Group*, IOCCG, Dartmouth, Canada, <https://doi.org/10.25607/OBP-696>, 2019.
- Jorge, D. S., Loisel, H., Jamet, C., Dessailly, D., Demaria, J., Bricaud, A., Maritorena, S., Zhang, X., Antoine, D., Kutser, T., Bélanger, S., Brando, V. O., Werdell, J., Kwiatkowska, E., Mangin, A., and d’Andon, O. F.: A three-step semi analytical algorithm (3SAA) for estimating inherent optical properties over oceanic, coastal, and inland waters from remote sensing reflectance, *Remote Sensing of Environment*, 263, 112 537, <https://doi.org/10.1016/J.RSE.2021.112537>, 2021.
- 240 Lee, Z., Carder, K. L., and Arnone, R. A.: Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters, *Applied Optics*, Vol. 41, Issue 27, pp. 5755–5772, 41, 5755–5772, <https://doi.org/10.1364/AO.41.005755>, 2002.
- Mélin, F. and Franz, B. A.: Assessment of Satellite Ocean Colour Radiometry and Derived Geophysical Products, *Experimental Methods in the Physical Sciences*, 47, 609–638, <https://doi.org/10.1016/B978-0-12-417011-7.00020-9>, 2014.
- 245 Mobley, C. D.: Estimation of the remote-sensing reflectance from above-surface measurements, *Applied Optics*, Vol. 38, Issue 36, pp. 7442–7455, 38, 7442–7455, <https://doi.org/10.1364/AO.38.007442>, 1999.
- Morley, S. K., Brito, T. V., and Welling, D. T.: Measures of Model Performance Based On the Log Accuracy Ratio, *Space Weather*, 16, 69–88, <https://doi.org/10.1002/2017SW001669>, 2018.
- Neil, C., Spyarakos, E., Hunter, P. D., and Tyler, A. N.: A global approach for chlorophyll-a retrieval across optically complex inland waters
 250 based on optical water types, *Remote Sensing of Environment*, 229, 159–178, <https://doi.org/10.1016/J.RSE.2019.04.027>, 2019.
- O’Reilly, J. E. and Werdell, P. J.: Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6, *Remote Sensing of Environment*, 229, 32–47, <https://doi.org/10.1016/J.RSE.2019.04.021>, 2019.



- O'Shea, R. E., Pahlevan, N., Smith, B., Bresciani, M., Egerton, T., Giardino, C., Li, L., Moore, T., Ruiz-Verdu, A., Ruberg, S., Simis, S. G., Stumpf, R., and Vaičiūtė, D.: Advancing cyanobacteria biomass estimation from hyperspectral observations: Demonstrations with HICO and PRISMA imagery, *Remote Sensing of Environment*, 266, 112 693, <https://doi.org/10.1016/J.RSE.2021.112693>, 2021.
- O'Shea, R. E., Pahlevan, N., Smith, B., Boss, E., Gurlin, D., Alikas, K., Kangro, K., Kudela, R. M., and Vaičiūtė, D.: A hyperspectral inversion framework for estimating absorbing inherent optical properties and biogeochemical parameters in inland and coastal waters, *Remote Sensing of Environment*, 295, 113 706, <https://doi.org/10.1016/J.RSE.2023.113706>, 2023.
- Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M. K., Ondrusek, M., Oppelt, N., and Stumpf, R.: Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach, *Remote Sensing of Environment*, 240, 111 604, <https://doi.org/10.1016/J.RSE.2019.111604>, 2020.
- Pahlevan, N., Smith, B., Alikas, K., Anstee, J., Barbosa, C., Binding, C., Bresciani, M., Cremella, B., Giardino, C., Gurlin, D., Fernandez, V., Jamet, C., Kangro, K., Lehmann, M. K., Loisel, H., Matsushita, B., Hà, N., Olmanson, L., Potvin, G., Simis, S. G., VanderWoude, A., Vantrepotte, V., and Ruiz-Verdù, A.: Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3, *Remote Sensing of Environment*, 270, 112 860, <https://doi.org/10.1016/J.RSE.2021.112860>, 2022.
- Salama, M. S. and Verhoef, W.: Two-stream remote sensing model for water quality mapping: 2SeaColor, *Remote Sensing of Environment*, 157, 111–122, <https://doi.org/10.1016/j.rse.2014.07.022>, 2015.
- Salama, M. S., Spaia, L., Poser, K., Peters, S., and Laanen, M.: Validation of Sentinel-2 (MSI) and Sentinel-3 (OLCI) Water Quality Products in Turbid Estuaries Using Fixed Monitoring Stations, *Frontiers in Remote Sensing*, 2, <https://doi.org/10.3389/frsen.2021.808287>, 2022.
- Samantaray, A. K., Mooney, P. A., and Vivacqua, C. A.: Bergen metrics: Composite error metrics for assessing performance of climate models using EURO-CORDEX simulations, *Geoscientific Model Development*, 17, 3321–3339, <https://doi.org/10.5194/GMD-17-3321-2024>, 2024.
- Seegers, B. N., Stumpf, R. P., Schaeffer, B. A., Loftin, K. A., and Werdell, P. J.: Performance metrics for the assessment of satellite data products: an ocean color case study, *Optics Express*, 26, 7404, <https://doi.org/10.1364/oe.26.007404>, 2018.
- Simão, R., Távora, J., Salama, M. S., and Fernandes, E.: Optimizing Optical Coastal Remote-Sensing Products: Recommendations for Regional Algorithm Calibration, *Remote Sensing* 2024, Vol. 16, Page 1497, 16, 1497, <https://doi.org/10.3390/RS16091497>, 2024.
- Smith, R. J.: Use and misuse of the reduced major axis for line-fitting, *American Journal of Physical Anthropology*, 140, 476–486, <https://doi.org/10.1002/AJPA.21090>, 2009.
- Stow, C. A., Jolliff, J., McGillicuddy, D. J., Doney, S. C., Allen, J. I., Friedrichs, M. A., Rose, K. A., and Wallhead, P.: Skill assessment for coupled biological/physical models of marine systems, *Journal of Marine Systems*, 76, 4–15, <https://doi.org/10.1016/J.JMARSYS.2008.03.011>, 2009.
- Subirade, C., Jamet, C., Tran, M. D., Vantrepotte, V., and Han, B.: Evaluation of twelve algorithms to estimate suspended particulate matter from OLCI over contrasted coastal waters, *Optics Express*, Vol. 32, Issue 26, pp. 45719–45744, 32, 45 719–45 744, <https://doi.org/10.1364/OE.529712>, 2024.
- Tran, M. D., Vantrepotte, V., Loisel, H., Oliveira, E. N., Tran, K. T., Jorge, D., Mériaux, X., and Paranhos, R.: Band Ratios Combination for Estimating Chlorophyll-a from Sentinel-2 and Sentinel-3 in Coastal Waters, *Remote Sensing* 2023, Vol. 15, Page 1653, 15, 1653, <https://doi.org/10.3390/RS15061653>, 2023.
- Werdell, P. J. and Bailey, S. W.: An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation, *Remote Sensing of Environment*, 98, 122–140, <https://doi.org/10.1016/J.RSE.2005.07.001>, 2005.



- Werdell, P. J., McKinna, L. I., Boss, E., Ackleson, S. G., Craig, S. E., Gregg, W. W., Lee, Z., Maritorena, S., Roesler, C. S., Rousseaux, C. S., Stramski, D., Sullivan, J. M., Twardowski, M. S., Tzortziou, M., and Zhang, X.: An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing, *Progress in Oceanography*, 160, 186–212, <https://doi.org/10.1016/J.POCEAN.2018.01.001>, 2018.
- 295 Yu, X., Salama, M. S., Shen, F., and Verhoef, W.: Retrieval of the diffuse attenuation coefficient from GOCI images using the 2SeaColor model: A case study in the Yangtze Estuary, *Remote Sensing of Environment*, 175, 109–119, <https://doi.org/10.1016/J.RSE.2015.12.053>, 2016.