Article

Exploring Spatial Aggregations and Temporal Windows for Water Quality Match-Up Analysis Using Sentinel-2 MSI and Sentinel-3 OLCI Data

Tanja Schröder [1], Susanne I. Schmidt [1,*], Rebecca D. Kutzner [2,3], Hendrik Bernert [4], Kerstin Stelzer [5], Kurt Friese [1] and Karsten Rinke [1]

1   Helmholtz-Centre for Environmental Research—UFZ, 39114 Magdeburg, Germany
2   Institut für Seenforschung, 88085 Langenargen, Germany
3   Forschungsinstitut für Bergbaufolgelandschaften e.V., 03238 Finsterwalde, Germany
4   EOMAP GmbH & Co. KG, 82229 Seefeld, Germany
5   Brockmann Consult GmbH, 21029 Hamburg, Germany
*   Correspondence: susanne.i.schmidt@ufz.de

Abstract: Effective monitoring and management of inland waterbodies depend on reliable assessments of water quality through remote sensing technologies. Match-up analysis plays a significant role in investigating the comparability between in situ and remote sensing data of physical and biogeochemical variables. By exploring different spatial aggregations and temporal windows, we aimed to identify which configurations are most effective and which are less effective for the assessment of remotely sensed water quality data within the context of governmental monitoring programs. Therefore, in this study, remote sensing data products, including the variables of Secchi depth, chlorophyll-a, and turbidity, derived from the Copernicus satellites Sentinel-2 and Sentinel-3, were compared with in situ laboratory data from >100 waterbodies (lakes and reservoirs) in Germany, covering a period of 5 years (2016–2020). Processing was carried out using two different processing schemes, CyanoAlert from Brockmann Consult GmbH and eoapp AQUA from EOMAP GmbH & Co. KG, in order to analyze the influence of different processors on the results. To investigate appropriate spatial aggregations and time windows for validation (the match-up approach), we performed a statistical comparison of different spatial aggregations (1 pixel; 3 × 3, 5 × 5, and 15 × 15 macropixels; and averaging over the whole waterbody) and time windows (same day, ±1 day, and ±5 days). The results show that waterbody-wide values achieved similar accuracies and biases compared with the macropixel variants, despite the large differences in spatial aggregation and spatial variability. An expansion of the temporal window to up to ±5 days did not impair the agreement between the in situ and remote sensing data for most target variables and sensor–processor combinations, while resulting in a marked rise in the number of matches.

Keywords: validation; match-up; water quality; inland waters; satellite data

1. Introduction

Inland waterbodies provide essential services for various human uses, particularly water supply and recreation, as well as habitat and ecosystem-regulating services, including nutrient and carbon cycling, or effects on the local climate [1–5]. However, inland waterbodies are increasingly threatened by anthropogenic exploitation and multiple environmental pressures such as organic and inorganic pollution, eutrophication, climate change effects, and toxic cyanobacteria blooms [2,6]. Therefore, the monitoring of inland waterbodies, with a special focus on water quality and their ecosystem status, is of global concern and a major prerequisite to better understand the effects of environmental changes on inland waters and to identify drivers of future change [7].

Despite the increasing need for more frequent and comprehensive monitoring arising, for example, from legislation such as the European Union's Water Framework Directive [8],

only a small fraction of inland waterbodies are part of in situ monitoring networks [7,9]. Although the Water Framework Directive was a major breakthrough in the monitoring of surface water, its implementation is limited by logistic and economic factors [10,11]. For example, samplings are not realized every year and usually occur only monthly during the growing season. The data may allow an assessment of the status of the ecosystem but do not provide deeper insights into the underlying dynamics and stressor–response relationships. It is therefore difficult to identify appropriate sustainable management strategies to improve ecological and chemical status [12,13].

In support of classical monitoring of inland water quality, remote sensing can be a valuable tool to provide water quality variables at a relatively low cost, at spatial scales from local to global, and at an improved temporal resolution through relatively frequent temporal revisits [10,14]. In this way, remote sensing can assist in identifying long-term trends and effects of climate change, point- and non-point source contaminants [11], or the emergence of extreme events such as algal blooms [15,16]. For the latter, near real-time and frequent information can be provided on algal dynamics, enabling early detection of phytoplankton blooms, and early warnings or tailored management reactions [11,16]. This could be relevant and helpful for urban waterbodies or bathing sites where algal blooms may impose health risks, and may strengthen the conversion of monitoring algal blooms according to regulations such as the European Union's Bathing Water Directive [17] and its locally derived decrees. Satellite-based remote sensing can also provide data on the ecological classification and assessment of trophic status, e.g., in the form of the product types applied, such as eutrophication indices [18–22].

However, the application of satellite-based remote sensing is limited in the number of measurable variables, the frequency of satellite overpasses, and the spatial resolution of the sensors on board, and mostly reflect the conditions just at the water's surface (i.e., the visible water column, depending on the pixel size and penetration depth of light) [7,10,23,24]. Technically, the optical complexity of inland waters, referring to the intricate and fluctuating interactions of light with the varied and dynamic composition of optically active constituents in the water, still presents a challenge [7,16,24]. In addition, atmospheric correction, adjacency effects, bottom reflectance effects, and the sensor's design are challenging factors [7,9,16,24,25]. Besides these, weather conditions such as clouds, rainfall, ice coverage, or waves during storm events may interfere with satellite-based remote sensing for certain applications [11,14,26].

On the basis of the information presented, it is evident that remote sensing could enhance water quality information through increased spatial and temporal coverage, cost-effectiveness, and relatively quick availability of data [10,14,26], thereby complementing in situ data. The synergies between in situ and remote sensing have not been fully realized due to challenges such as limited temporal capacities and a lack of support from the organizational management within water administrations, concerns about products' accuracy and data continuity, and the absence of legal frameworks explicitly incorporating or permitting remote sensing-derived observations [27,28].

The integration of processed remote sensing products by authorities and water management authorities requires a rigorous comparison of satellite-based and in situ observations, and an analysis and interpretation of the quality, accuracy, and uncertainties (validation [29]). Quantifying the reliability of remote sensing products and accounting for all components of uncertainty, however, is a challenging task due to the challenging process associated with every satellite-derived variable from sensor-level signals to mass concentrations [14,29]. Additional uncertainties are introduced through the in situ dataset as well as the spatiotemporal sampling mismatch between the satellite data and the in situ data [29,30]. Even though in situ measurements are often referred to as "ground truth" measurements, they also come with measurement uncertainties themselves, also caused by the different methods of sampling and analysis applied [31].

While the aim of validation is quite clear, the implementation often involves various steps that are subject to assumptions and potentially require the user's decisions, which affect the validation of the results [29]. One aspect of this is the choice of the optimal spatial and temporal

scales for in situ and remote sensing data. This requires a decision on which mismatch in scales can be accepted [23,25,29,32]. As a general rule, match-ups between in situ and remote sensing data should be as close to each other as possible in time and space (horizontal, vertical, temporal) so they represent the same or at least comparable conditions [29]. However, to allow for robust statistical analysis, a sufficient number of match-ups is also needed. Therefore, a choice has to strike a balance between minimizing the spatiotemporal mismatching and producing a large number of match-ups in order to have a representative sampling size [25,29]. Moreover, in situ monitoring is rarely aligned with satellites' overpasses and a full match down to minutes of both monitoring activities in space and time is the exception. Even worse, in the case that both take place on the same day and thereby may appear to be temporally aligned, this can be misleading in cases when the observed water parcels may differ substantially or one sample takes place in the morning and the other in the evening [33].

In order to address these complications in the validation of remote sensing products for water quality, we designed this study. We collected in situ observations of 112 inland waterbodies from different monitoring agencies in Germany from 2016 to 2020. The waterbodies varied in their morphometry and trophic state, resulting in 37.930 observations for all three variables prior to preprocessing. We generated the corresponding satellite-based products from the Copernicus satellites Sentinel-2 MSI and Sentinel-3 OLCI, and used two different processing schemes (CyanoAlert® from Brockmann Consult Ltd. in Hamburg, Germany and eoapp® AQUA from EOMAP GmbH & Co. KG in Seefeld, Germany). Subsequently, we compared the remote sensing data for different levels of spatial aggregation (pixel windows or "macropixels" of 1×1, 3×3, 5×5, and 15×15 as well as at the scale of the whole waterbody) (see Figure 1) and for temporal windows from the same day to up to ± 5 days for different variables of water quality. We intentionally focused on processing workflows that are commercially available because applications in governmental or societal contexts are usually carried out in a consultancy setting. We quantified the deviations between satellite and in situ observations by using three different error measures. Thus, the research questions of this study were to find (i) what spatial aggregations of remote sensing data are appropriate for validation, (ii) which time window is adequate when comparing remote sensing and in situ data, and (iii) whether there are systematic differences in these characteristics depending on the variable of water quality, the satellite, or the processor.
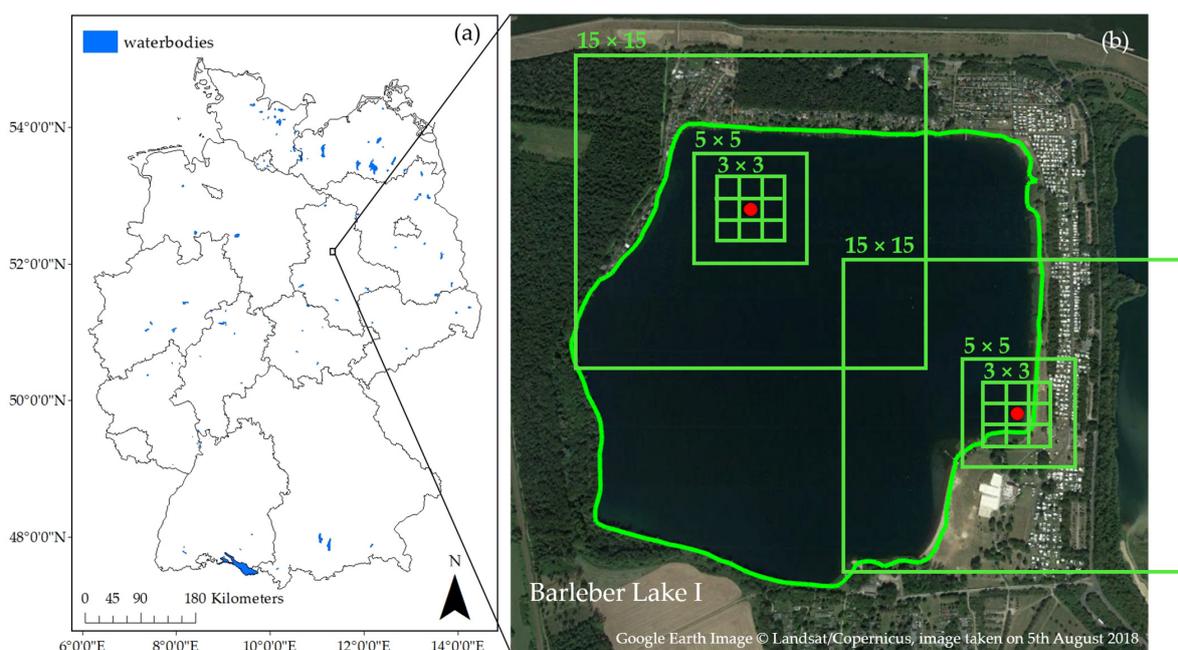
Figure 1. (**a**) Inland waterbodies investigated. (**b**) Exemplary illustration of the spatial aggregations extracted for MSI data.

2. Materials and Methods

2.1. In Situ Data Observations of More than 100 Lakes and Reservoirs

Laboratory in situ measurements of three target variables of water quality, namely Secchi depth, chlorophyll-a, and turbidity, for 112 waterbodies (218 measurement stations) were collected from various water authorities of 13 federal states and research institutes in Germany (see Figure 1). The measurements cover 5 years from 2016 to 2020. Since the monitoring of water quality and the effective implementation of water protection policies in Germany is the responsibility of the federal states, the number of in situ samples and the frequency of sampling differs among waterbodies, federal states, and target variables. All available data were collected for the purpose of assessing water quality to meet the objectives of existing policies at the EU or federal level, and ultimately to ensure the good ecological status of the waterbodies. It should be noted that the in situ data were collected across federal states according to different methods and protocols. In the following section, we describe the steps of homogenizing the data we undertook to mitigate the differences in the protocols.

Several steps were taken to optimally prepare the in situ data for comparison with satellite data. The maximum depth of the in situ data was restricted, measurement sites located in shallow water zones as well as probe data were excluded, and extreme values were removed from the dataset. These steps are described in detail in the following section.

Remote sensing data capture events only within the visible water column. Therefore, to improve vertical comparability between in situ and remote sensing data, discrete in situ measurements with a sampling depth of more than 2 m from the water's surface were removed from the dataset for chlorophyll-a and turbidity. In case of integral measurements (taken between 0 m and 25 m max.), all values were retained and arithmetically averaged per time point. Only 10% had a maximum sampling depth of more than 10 m so that the potential contribution from deep chlorophyll or maxima of turbidity, which can hardly be detected by satellites, remained small.

In addition, optically shallow waters, where the penetration depth of light exceeds the physical depth of the water column, are prone to detection errors arising from the contributions of bottom reflectance. Therefore, in this study, measurement points located in shallow water zones were removed from the dataset by visually inspecting high-resolution satellite images together with the measurement points.

Furthermore, probe data with nearly daily measurements were not integrated into the analysis to balance the number of matches across the waterbodies. Finally, extreme values in both the in situ and satellite-based water quality data were removed if they were outside the data ranges as given in Table A3. The resulting in situ dataset consisted of 7286 data points in total and encompassed both discrete ($n = 5437$) and integral ($n = 1849$) measurements with varying numbers of observations per site and parameter (Tables 1 and S1).

Table 1. Minimum, maximum, and mean number of in situ measurements per waterbody and variables, and the total number of measurements (2016–2020).

In Situ Data	Chlorophyll-A	Turbidity	Secchi Depth
Unit	µg/L	FNU	m
Minimum	2	2	2
Maximum	173	241	231
Average	30	89	33
Total n	2956	1151	3179

2.2. Satellite-Based Detection of Water Quality with Sentinel-2 MSI and -3 OLCI

The remote sensing data originated from optical sensors on board the Copernicus satellites Sentinel-2 and Sentinel-3 over the timespan from 2016 to 2020. Sentinel-2 satellites (A and B), launched in 2015 and 2017, with multispectral instruments (MSI) on board, offer high-resolution imagery at spatial resolutions of 10 m, 20 m, and 60 m, depending on the spectral band [34,35]. The MSI measure 13 spectral bands and provide data every 2–5 days depending on the latitude [34,35]. Likewise, the Sentinel-3 mission consists of two satellites (A and B), that were launched in 2016 and 2017, respectively [36]. Sentinel-3 satellites carry ocean and land color instruments (OLCI) on board [36,37]. The OLCI measure 21 spectral bands at a spatial resolution of 300 m and provide temporal coverage every 2 days at the equator and up to twice a day in midlatitudes [36,37].

For the purpose of quantifying the selected variables of water quality (chlorophyll, turbidity, Secchi depth), the radiance leaving the water, i.e., the spectrally resolved light, needed to be determined on the basis of the signal detected by the sensor [14,38]. For this, the influence of the atmosphere and the reflection of light at the air–water interface had to be determined and subtracted from the top-of-atmosphere radiance to derive estimates of the radiance leaving the water [38–41]. The spectrally resolved remotely sensed signal was then used to retrieve the variables of water quality [40,42].

In this study, two commercially available but scientifically documented operational processing schemes, CyanoAlert® (Brockmann Consult GmbH) and eoapp® AQUA (EOMAP GmbH & Co. KG), were applied. These are both based on analytical expressions, incorporating the radiative transfer equation to retrieve the concentrations of optically active constituents of water from the radiance leaving the water, which was derived from top-of-atmosphere measurements of various optical satellite-based data sources. Both processing schemes are globally applicable for almost any kind of water and without a priori knowledge of the particular waterbody. The core element of the processing chain applied within CyanoAlert® is C2RCC (Case 2 Regional CoastColour), which is composed of a set of neural nets (NNs) to derive the atmospheric and in-water properties. The NNs are trained by a set of approximately 10 million simulated reflectance spectra representing a wide range of in-water and atmospheric conditions [43]. Within the processing chain, C2RCC is complemented by cloud detection (Idepix), MPH (maximum peak height) algorithms for the detection of chlorophyll-a (only OLCI data), and the Nechad algorithm for determining the turbidity. Further information on this processing scheme was presented by Brockmann et al. [43] (C2RCC), Matthews et al. [18,44] (MPH), Wevers et al. [45,46] (Idepix), and Nechad et al. [47]. The various processing elements of this processing scheme are referred to as CyanoAlert. In comparison, the processor underlying eoapp® AQUA is called MIP (modular inversion and processing system). MIP is purely physics-based and consists of a sensor-independent suite of algorithms and databases to derive atmospheric and in-water properties [28,48]. The model encompasses all the relevant processing steps and necessary corrections such as detection of the surface type (land, water, or cloud), correction of adjacency and sun glint, and atmospheric correction. Retrieval of the variables of water quality was performed by modeling the influence of their respective optically active components on the measured radiance [28,48]. Details of this processing scheme were documented by Heege et al. [48] (MIP). The processor is referred to as EOMAP-MIP. Note that in accordance with company recommendations, the Sentinel-2 MSI data processed by CyanoAlert have a resolution of 60 m, whereas the Sentinel-2 MSI data processed by EOMAP-MIP have a resolution of 10 m.

Due to the lower spatial resolution of Sentinel-3 OLCI (300 m pixels versus 10 or 60 m pixels in Sentinel-2 MSI), only a selection of 38 waterbodies were suited for OLCI images, as prescribed by area and shape of the waterbody. It was assumed that a spatial window of 5×5 water surface pixels, i.e., about 1.5×1.5 km in dimension, should be present inside a waterbody to be suitable for evaluation based on OLCI. For selection of a waterbody, the point furthest away from the entire shoreline was calculated and buffered with a radius of 750 m, based on which, a bounding box with an edge length of 1500 m was calculated.

The bounding box, corresponding to a macropixel of 5×5 grid cells, was intersected with the respective shape of the waterbody. All waterbodies whose shoreline shapes intersected with the bounding box were considered to be unsuitable.

Each processing scheme incorporates algorithms to mask out clouds, cloud shadows, and haze, and to handle adjacency effects or sun glint. The data processed by CyanoAlert came with a quality indicator (quality band) to differentiate between valid (quality indicator == 1) and invalid (quality indicator == 0) pixels. The quality indicator is composed of a combination of quality flags generated by the C2RCC processor raised for invalid processing conditions and by the Idepix pixel classification scheme identifying disturbed water pixels. Data processed by EOMAP-MIP contained a quality score ranging from 0 to 100 (low to high quality), in which case, pixels with quality score values smaller than 50 were removed from the data. This score was calculated from the influences of atmospheric and surface effects, the angles of the sun and the sensor view, and detectable concentration limits defined within the processor's definitions. In addition, pixel outliers within the macropixel were removed when they were outside the range of mean ± 1.5-times the standard deviation, as suggested by Bailey and Werdell [31]. Only if more than 30% of pixels within the macropixel or waterbody were valid, the respective scene was evaluated; otherwise, it was excluded. The remaining valid pixels were spatially aggregated by calculating the median, 25th percentile, and 75th percentile, as well as the coefficient of variation (CV; standard deviation/mean). The CV was calculated for characterizing the level of spatial variability within the different spatial aggregations. As an example, Table 2 shows an overview of the amount of data extracted for chlorophyll-a and the associated spatial aggregations. Data on turbidity and Secchi depth were in the same order of magnitude and are given in the Appendix A (Tables A1 and A2).

Table 2. Overview of the valid remote sensing data for chlorophyll-a, given as the number of scenes per sensor, processor, and spatial resolution, for all waterbodies over the period considered (2016–2020).

Chlorophyll-A	S2-MSI				S3-OLCI		
	3×3	5×5	15×15	Waterbody Scale	1×1	3×3	Waterbody Scale
CyanoAlert	14,639	15,053	13,534	17,403	10,802	11,867	14,508
EOMAP-MIP	15,895	16,004	16,083	19,630	16,990	18,542	20,730

Finally, extreme values in both the in situ and satellite-based water quality data were removed if they were outside the data ranges as given in Table A3.

2.3. Statistical Analyses and Comparison of In Situ and Satellite-Based Observations

In this study, temporal windows of in situ and remote sensing data from the same day up to ± 5 days were considered. In this process, the optimal temporal match was determined for each in situ data point, ensuring that each satellite-based data point appeared only once in the dataset. Note that the time window was expanded for this research, meaning that the ± 5-day window included all matches from the same day to ± 5 days. The number of matches generated per waterbody varied greatly due to difference in both the available in situ and usable remote sensing data, ranging from 1 to 108 matches per waterbody. In addition, since the literature on remote sensing of water quality reports a wide range of temporal matches to be considered near-coincident, ranging from ± 3 h to ± 10 days [23], we also formed all possible matches for each in situ data point up to ± 10 days. For each time lag, we calculated the residuals along with the associated statistics (mean, median, and interquartile range). Note that the in situ data did not have temporal information specified to the hour.

In this study, three error metrics were calculated for evaluation of the performance, namely the mean absolute error of the log-transformed data (MAE) and bias, in line with [49], as well as the root mean square error (RMSE). The bias quantified systematic differences between the two datasets, namely systematic over- or underestimation, and was defined as the difference of the mean values for the in situ and satellite-based values, and hence was not sensitive to random errors [29,49,50]. The RMSE and the MAE are both metrics describing the accuracy or the pairwise agreement between matched in situ and satellite-based observations [49]. The RMSE is frequently used in validation analyses of remote sensing data but can become strongly influenced by larger deviations [51], in contrast to MAE, which is more robust against outliers [49,51]. In addition, our data showed a logarithmic distribution of error (see Figure A1), and we therefore followed the recommendation of Seegers et al. [49] and calculated the MAE and bias using log-transformed data, followed up by back-transformation to linear space to facilitate interpretation. RMSE was calculated using untransformed data, i.e., in the linear space, to facilitate comparability with other studies and for easier interpretation. For details on the performance metrics for the assessment of satellite-based data products, see Seegers et al. [49]. The error metrics were computed using Equations (1)–(3).

$$MAE = 10\hat{}\left(\frac{\sum_{i=1}^{n} \left| \log_{10}(i_{satellite}) - \log_{10}(i_{insitu}) \right|}{n} \right) \tag{1}$$

$$Bias = 10\hat{}\left(\frac{\sum_{i=1}^{n} \log_{10}(i_{satellite}) - \log_{10}(i_{insitu})}{n} \right) \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (i_{satellite} - i_{insitu})^2}{n}} \tag{3}$$

We defined different variants for spatial and temporal matching (Table 3) in order to identify the performance of satellite-borne data and to derive recommendations for practical matching. The selection of smaller spatial aggregations was made (1 pixel, 3×3 macropixels, and 5×5 macropixels) following the recommendations of EUMETSAT [52], which suggested spatial aggregations of 1 pixel, 3×3 macropixels, and 5×5 macropixels, depending on the local conditions. A larger spatial aggregation (15×15 macropixels) was added for the S2-MSI data to approximate comparability with the S3-OLCI data. Finally, a waterbody-scale variant was added to analyze to what extent the in situ data represented the entirety of the waterbody. With regard to the selection of temporal windows, the usage of same-day matching constituted the minimum possible temporal window. This was extended to time windows of ± 1 day or ± 5 days as alternatives in temporal matching, following previous studies that have applied temporal intervals of in situ and remote sensing data ranging from ± 3 h to ± 10 days [23]. For the comparison of spatial aggregations, the temporal windows were kept to same day matches in order to restrict the variability only to the effects of spatial aggregation. Accordingly, in a comparison of the temporal windows, the spatial aggregation was kept constant at 3×3 macropixels. The abovementioned error metrics were used to compare and evaluate these different variants. Moreover, the respective sample sizes, i.e., the number of matches between in situ and satellite-based observations, were included in our evaluation. Note that the applied spatial aggregations differed between S3-OLCI and S2-MSI in order to account for the different spatial resolutions of these sensors.

All analyses and visualizations were performed using R Statistical Software (v4.3.0; R Core Team 2021). Geospatial analysis was performed using ArcGIS Desktop (Esri 2019).

Table 3. Different variants for spatial and temporal matching. Note the different spatial resolutions of the two sensors as outlined in the text. For example, a 3×3 macropixel from S3-OLCI covers 900×900 m, which is the same as a 15×15 macropixel from S2-MSI (when S2 is processed at a 60 m resolution). Note also that the comparison of spatial matches only used same-day temporal matches.

		Spatial Matching				
	Temporal Matching	1 Pixel	3×3 Pixels	5×5 Pixels	15×15 Pixels	Waterbody Scale
S2-MSI	Same day (0 d)		X	X	X	X
	± 1 day		X			
	± 5 days		X			
S3-OLCI	Same day (0 d)	X	X			X
	± 1 day		X			
	± 5 days		X			

3. Results

3.1. Spatial Aggregations

In the comparison of in situ and satellite-based water quality variables for different spatial aggregations, the emerging differences remained relatively small and random overall. This was true for all three target variables (chlorophyll-a, turbidity, and Secchi depth), both sensors (OLCI and MSI), and both processors (EOMAP-MIP and CyanoAlert) (Figures 4, A2 and A3). In most cases, it did not make a difference if either finely resolving the macropixels at the scale of a few hundred meters or the waterbody-scale variant were used in determinations of the error metrics. A further complicating factor was that the different error measures applied to the same variable, sensor, and processor favored different spatial aggregations (Figure 4). As an effect of scale, data on the Secchi depth produced the least pronounced differences overall in the error metrics among spatial aggregations simply because their range was more restricted than that of, say, chlorophyll. Errors for chlorophyll-a appeared to be generally higher, but again, no systematic differences among the applied spatial aggregations emerged. For turbidity, the error estimates were slightly better for waterbody-wide values. This evaluation was, however, based on a smaller number of waterbodies and matches compared with the other target variables. In most cases, the waterbody-scale variant yielded a slightly larger number of matches for both sensors. But this effect remained small: on average, the waterbody scale led to an increase in matches by 16.8% compared with the finest spatial aggregation applied.

A closer inspection of the different spatial aggregations of all variables of water quality for MSI-based data products (Figure 2) and OLCI-based products (Figure 3) generally confirmed the highly diverse and partly random performance of the different spatial aggregations. Irrespective of the spatial aggregations applied, however, the absolute estimates differed between the different processors and sensors. For example, chlorophyll concentrations yielded lower concentration values on average when processed by EOMAP-MIP compared with being processed by CyanoAlert. Scatterplots, however, sometimes showed severe cases of overestimation (above the 1:1 line) and underestimation (below the 1:1 line). This appeared to be independent of the spatial aggregation applied (Figures 2 and 3) and occurred for all processor–sensor combinations.

In one case, the consistent superior performance of the waterbody scale appeared in our analysis. The RMSE when the data were processed by EOMAP-MIP was always the lowest for the waterbody scale (see Figure 4). No such consistent response was detectable for any macropixel-based aggregation.
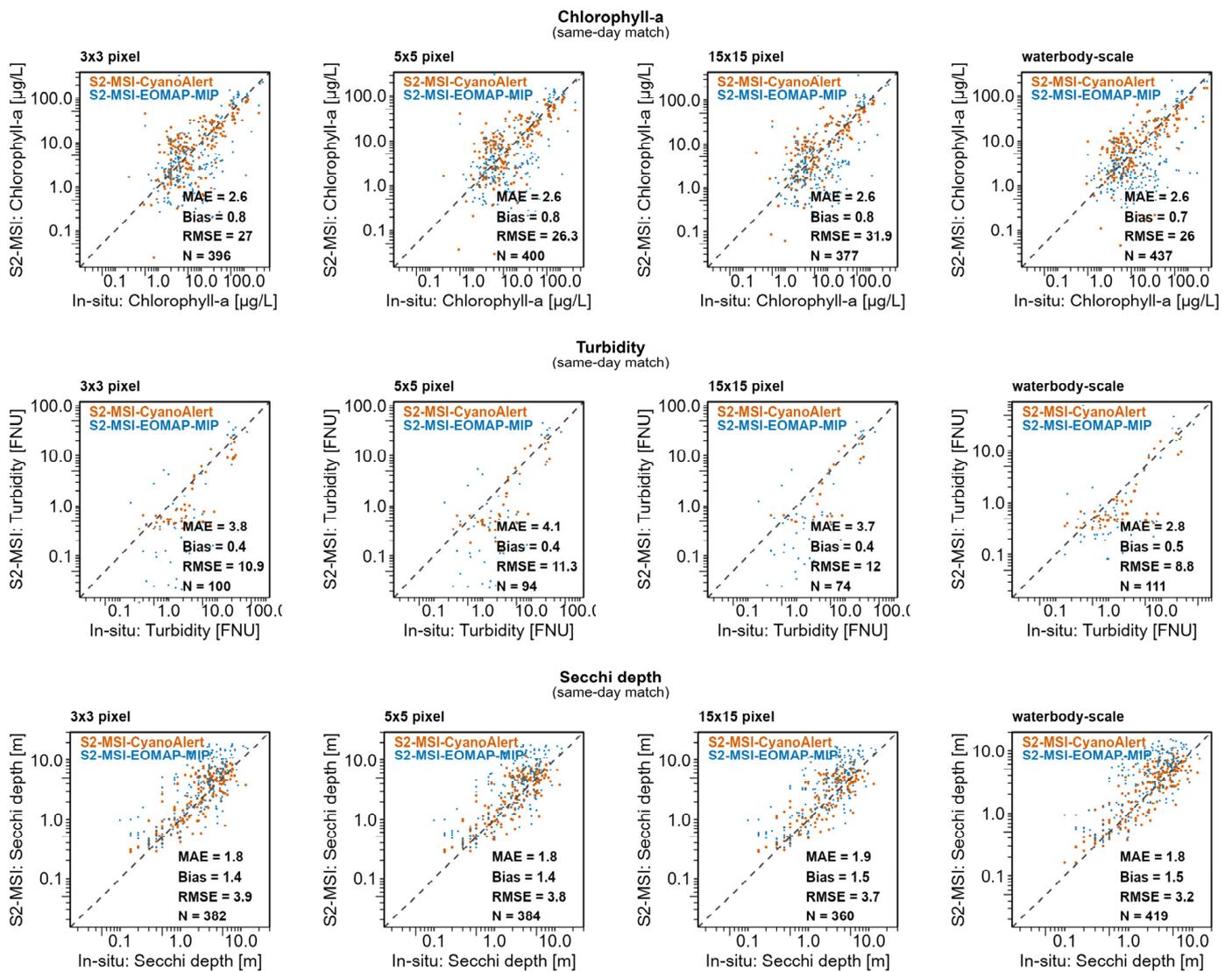
Figure 2. Overall performance of different spatial aggregations using the match-up dataset (same day) with MSI data processed by CyanoAlert (orange) and EOMAP-MIP (blue). Scatterplots are shown in log-log scale with selected error metrics and the number of observations (both processors combined) for the three target variables of chlorophyll-a (**top**), turbidity (**middle**), and Secchi depth (**bottom**). The grey dashed line refers to the 1:1 line.

The coefficient of variation (CV) is given in Table 4. It is important to note that the CV is not necessarily a measure of quality but is rather an indicator of spatial heterogeneity, something that is normal in natural water and that can rarely be covered with in situ measurements. Moreover, the different spatial scales at which the MSI data were processed had a direct impact on that measure. The spatial variability increased with increasing spatial scale and was at a maximum at the waterbody scale. This increase in spatial CV was stronger for EOMAP-MIP-based processing when applied to the products of MSI but were higher for CyanoAlert-based processing when applied to the products of OLCI. Moreover, the increase in spatial CV was more pronounced for chlorophyll-a than for the other two water quality products.
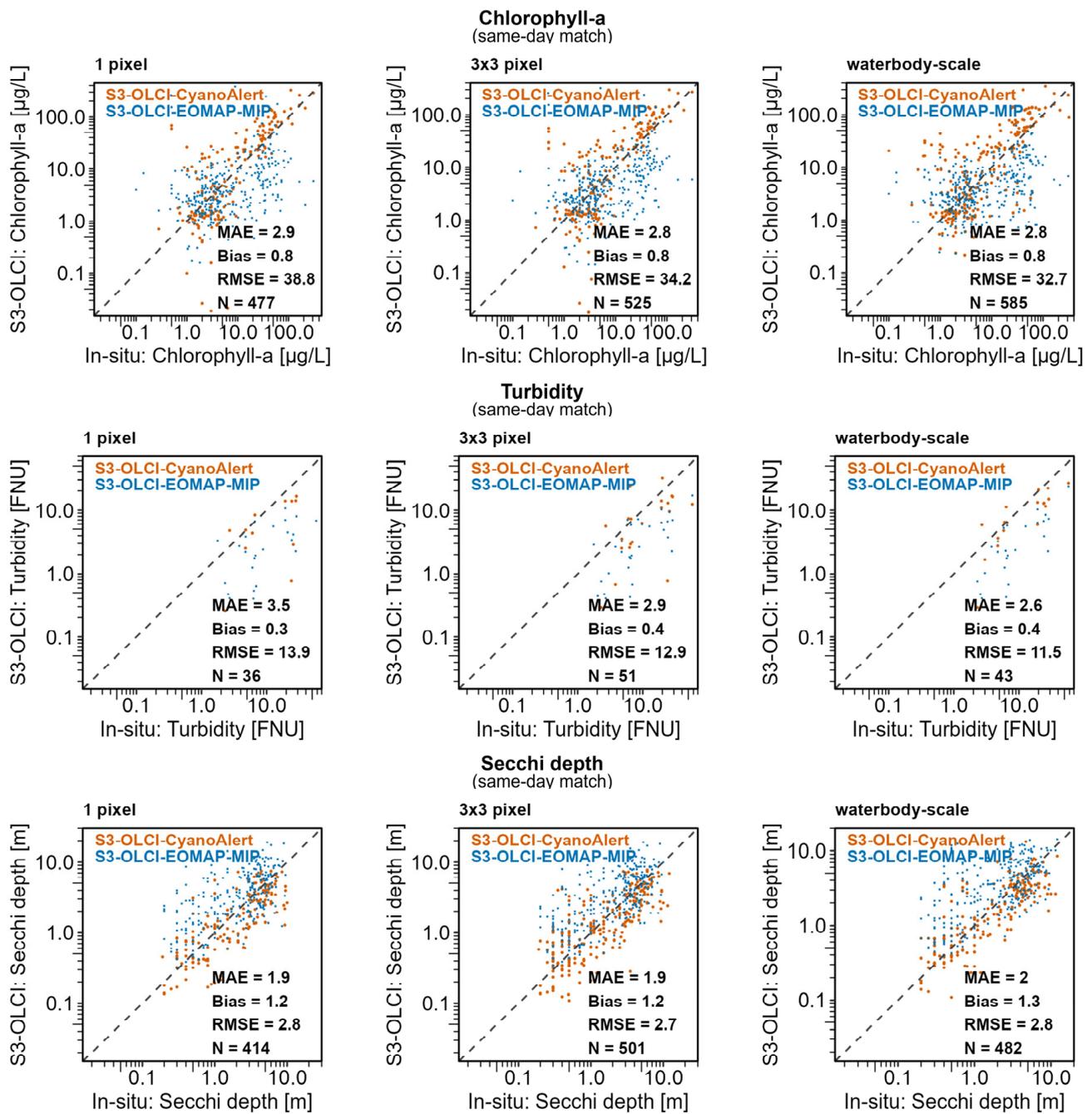
Figure 3. Overall performance of different spatial aggregations (left column, 1 pixel; middle column, 3 × 3 macropixels; right column, waterbody scale) using the match-up dataset (same day) with OLCI data processed by CyanoAlert (orange) and EOMAP-MIP (blue). Scatterplots are shown in log-log scale with selected error metrics and the number of observations (both processors combined) for the three target variables of chlorophyll-a (**top** row), turbidity (**middle** row), and Secchi depth (**bottom** row). The grey dashed line refers to the 1:1 line.

Error metrics - Spatial aggregations

Model	Sensor	Target-variable	MAE 1	3x3	5x5	15x15	all	Bias 1	3x3	5x5	15x15	all	RMSE 1	3x3	5x5	15x15	all	N 1	3x3	5x5	15x15	all	N Lakes 1	3x3	5x5	15x15	all
CyanoAlert	MSI	Chlorophyll-a	-	2.3	2.3	2.1	2.3	-	1.0	1.1	1.2	1.1	-	18.5	19.2	13.9	21.8	-	194	198	168	206	-	64	64	52	62
		Turbidity	-	2.4	2.8	2.3	2.2	-	0.5	0.4	0.5	0.5	-	4.7	4.3	5.1	3.5	-	48	42	23	54	-	11	9	6	10
		Secchi depth	-	1.6	1.6	1.6	1.6	-	1.2	1.2	1.2	1.2	-	2.0	2.0	2.0	2.0	-	184	185	156	197	-	64	64	51	64
	OLCI	Chlorophyll-a	2.5	2.5	-	-	2.5	1.0	1.1	-	-	1.3	29.6	27.3	-	-	36.8	175	196	-	-	239	31	32	-	-	29
		Turbidity	2.7	2.5	-	-	1.8	0.4	0.5	-	-	0.7	11.4	13.6	-	-	10.3	12	21	-	-	18	2	2	-	-	2
		Secchi depth	1.7	1.8	-	-	1.7	0.8	0.9	-	-	0.9	1.9	1.9	-	-	2.0	160	228	-	-	190	29	30	-	-	28

Model	Sensor	Target-variable	MAE 1	3x3	5x5	15x15	all	Bias 1	3x3	5x5	15x15	all	RMSE 1	3x3	5x5	15x15	all	N 1	3x3	5x5	15x15	all	N Lakes 1	3x3	5x5	15x15	all
EOMAP-MIP	MSI	Chlorophyll-a	-	3.0	2.9	3.0	2.9	-	0.6	0.6	0.6	0.5	-	33.2	31.7	40.9	29.2	-	202	202	209	231	-	65	65	67	68
		Turbidity	-	5.7	5.6	4.6	3.5	-	0.3	0.3	0.4	0.5	-	14.5	14.7	14.1	11.8	-	52	52	51	57	-	10	10	10	11
		Secchi depth	-	2.1	2.1	2.1	2.1	-	1.7	1.7	1.7	1.8	-	5.0	5.0	4.7	3.9	-	198	199	204	222	-	65	66	67	69
	OLCI	Chlorophyll-a	3.1	3.0	-	-	3.1	0.7	0.7	-	-	0.6	43.2	37.7	-	-	29.5	302	329	-	-	346	33	33	-	-	32
		Turbidity	4.0	3.2	-	-	3.4	0.3	0.3	-	-	0.3	15.0	12.4	-	-	12.3	24	30	-	-	25	2	2	-	-	2
		Secchi depth	2.1	2.0	-	-	2.2	1.6	1.6	-	-	1.7	3.2	3.2	-	-	3.2	254	273	-	-	292	31	32	-	-	31

Figure 4. Overall performance of different spatial aggregations using the match-up dataset (same day) with MSI and OLCI data processed by CyanoAlert (top) and EOMAP-MIP (bottom). The table shows selected error metrics, the number of observations (matches), and the number of waterbodies (N Lakes) for the three target variables of chlorophyll-a, turbidity, and Secchi depth. Dark grey shades indicate the poorer performance of the variant, while the lightest shade represents the best performance within each processing scheme.

Table 4. Average coefficient of variation depending on the spatial aggregations, target variables, sensors, and processors. Note that OLCI was not evaluated for 5×5 and 15×15, and that for 1×1, no CV could be calculated.

Sensor	Target Variable	CV (CyanoAlert) 3×3	5×5	15×15	All	CV (EOMAP-MIP) 3×3	5×5	15×15	All
S2-MSI	Chlorophyll-a	36.7	44.0	78.9	87.8	13.6	18.0	33.3	204.8
	Turbidity	18.3	56.7	26.3	32.8	18.2	24.4	51.3	188.2
	Secchi depth	17.8	20.8	28.2	30.5	8.8	12.1	20.3	52.1
S3-OLCI	Chlorophyll-a	13.6	-	-	115.4	23.5	-	-	109.2
	Turbidity	18.2	-	-	63.2	15.2	-	-	24.3
	Secchi depth	8.8	-	-	55.4	14.9	-	-	36.0

3.2. Temporal Windows

Similar to the analysis of different spatial resolutions, the expansion of the time window from 0 to 5 days did not yield a consistent pattern if all error measures were considered from a synoptic view (Figures 5 and 6). Nevertheless, as in many tested combinations of sensors, processors, and variables of water quality, a 5 d delay resulted in elevated RMSE errors (Figures 5, 6 and A4). This was, for instance, noticeable in the measurements of chlorophyll-a from MSI or the turbidity assessments of OLCI. In some

cases, the results for different temporal windows hardly differed, such as in the Secchi depth and the OLCI chlorophyll-a data products, indicating that a temporal extension to ±5 d did not necessarily come with higher errors (Figure 6). It is important to note that the expansion of the time window from the same day to ±5 days resulted in a significant increase in the number of matches (right-hand side of Figure 6), which were, on average, about 2.5 times (±1 day) or 5 times (±5 days) greater compared with the same-day matches.
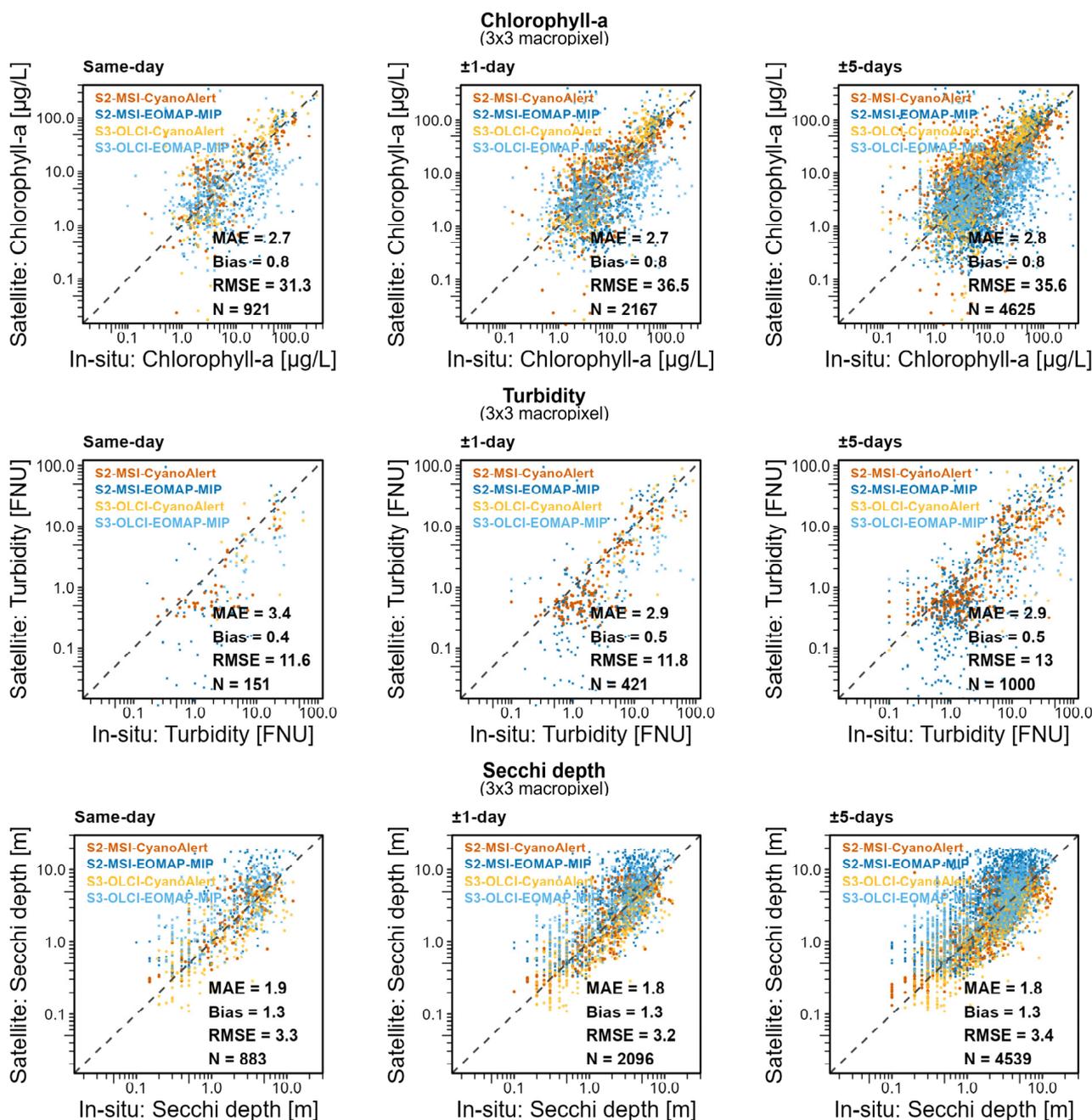
Figure 5. Overall performance of different temporal windows (left column, same day; middle column, ±1 day; right column, ±5 days) using the match-up dataset (3 × 3 macropixel) with MSI data processed by CyanoAlert (orange) and EOMAP-MIP (dark blue), and OLCI data processed by CyanoAlert (yellow) and EOMAP-MIP (light blue). Scatterplots are shown in log-log scale with selected error metrics and the number of observations for the three target variables of chlorophyll-a (**top** row), turbidity (**middle** row), and Secchi depth (**bottom** row). The grey dashed line refers to the 1:1 line.

Error metrics - Temporal windows

			MAE			Bias			RMSE			N			N lakes		
Model	Sensor	Target-variable	±0d	±1d	±5d	±0d	±1d	±5d	±0d	±1d	±5d	±0d	±1d	±5d	±0d	±1d	±5d
CyanoAlert	MSI	Chlorophyll-a	2.3	2.3	2.3	1.0	1.1	1.2	18.5	20.7	22.0	194	557	1267	64	83	91
		Turbidity	2.4	2.1	2.1	0.5	0.6	0.6	4.7	7.8	9.0	48	158	370	11	11	12
		Secchi depth	1.6	1.5	1.6	1.2	1.1	1.1	2.0	2.0	2.1	184	554	1298	64	81	91
	OLCI	Chlorophyll-a	2.5	2.2	2.4	1.1	1.1	1.0	27.3	30.6	25.7	196	414	849	32	33	34
		Turbidity	2.5	2.1	2.3	0.5	0.7	0.6	13.6	14.8	16.5	21	42	80	2	2	2
		Secchi depth	1.8	1.7	1.7	0.9	0.9	0.9	1.9	1.7	1.7	228	454	874	30	31	33

			MAE			Bias			RMSE			N			N lakes		
Model	Sensor	Target-variable	±0d	±1d	±5d	±0d	±1d	±5d	±0d	±1d	±5d	±0d	±1d	±5d	±0d	±1d	±5d
EOMAP-MIP	MSI	Chlorophyll-a	3.0	3.2	3.4	0.6	0.6	0.6	33.2	49.4	47.0	202	588	1486	65	84	92
		Turbidity	5.7	3.7	3.4	0.3	0.4	0.5	14.5	10.7	12.3	52	172	458	10	10	12
		Secchi depth	2.1	2.0	2.1	1.7	1.6	1.6	5.0	4.7	4.8	198	590	1506	65	84	91
	OLCI	Chlorophyll-a	3.0	2.9	3.0	0.7	0.7	0.6	37.7	37.0	37.0	329	608	1023	33	35	35
		Turbidity	3.2	4.0	4.6	0.3	0.3	0.3	12.4	20.5	22.8	30	49	92	2	2	2
		Secchi depth	2.0	2.0	2.0	1.6	1.6	1.6	3.2	3.1	3.2	273	498	861	32	34	34

Figure 6. Overall performance of different temporal windows using the match-up dataset with MSI and OLCI data processed by CyanoAlert (top) and EOMAP-MIP (bottom). The table shows selected error metrics, the number of observations (N, "matches"), and the number of waterbodies (N Lakes) for the three target variables of chlorophyll-a, turbidity, and Secchi depth. Dark grey shades indicate poorer performance of the variant, while the lightest shades represent the best performance within each processing scheme.

Besides the highly diverse behavior of the error indices among the variables, sensors, and processors, a few consistent patterns emerged. The RMSE tended to increase with large time windows (±5 d), while the log-based MAE and bias were more random (see Section 4.5). Among the variables, the Secchi depth reacted rather insensitively to expansion of the time windows, and in many cases, a time window of ±5 d did not come with an elevated error. An increase in the sample size, i.e., the number of valid matches, necessarily increased with an increasing time window. This increase in sample size was higher for MSI products (roughly a factor of 6 to 7) than for OLCI products (roughly a factor of 3–4).

The highly diverse behavior of the error metrics for different temporal matchings was slightly reduced when both sensors (MSI, OLCI) and processors (EOMAP-MIP, CyanoAlert) were combined for each product (Figure A4). In this aggregated view, the RMSE suggested that larger temporal windows (i.e., ±5 d) increased the errors. In addition, there were only minor differences in CV among the different temporal windows (Table A4).

4. Discussion

Restating the established objectives, the study aimed to identify (i) appropriate spatial aggregations of remote sensing data for a good representation of observations obtained through in situ measurements, (ii) adequate temporal windows when comparing remote

sensing and in situ data, and (iii) systematic differences in these characteristics depending on the variable of water quality, satellite, or processor.

4.1. Spatial Aggregations

Regarding Objective (i), the study has shown that no clear pattern emerged regarding the spatial aggregation. This finding was invariant to the processing scheme used and was confirmed when all results for both processors (EOMAP-MIP, CyanoAlert) were merged into one analysis (only possible for 3 × 3 macropixel and the waterbody-scale variants) (Figures A2 and A3). In detail, the pairwise agreement (MAE, RMSE) between the different spatial aggregations (different sizes of macropixels versus the waterbody scale) was surprisingly similar overall and showed no clear patterns, despite large differences in their spatial extent and variability, whereby the waterbody-scale variant performed similarly well overall compared with the macropixel variants (see Figure 4). As an exception, the performance of the waterbody-scale evaluation was significantly better for OLCI data processed with EOMAP-MIP when evaluated with the RMSE. Similarly, systematic differences (bias) between the in situ and remote sensing data showed only minor differences among the different spatial aggregations. This is astonishing, because the CV in the satellite data clearly showed that the variability in the target variables (e.g., chlorophyll-a) increased with an increase in the spatial scale and was at maximum at the waterbody scale. This spatial variability at the scale of the whole waterbody actually suggests that macropixels should be superior in terms of validation, which could not be confirmed statistically. Note that in this context, the comparison of spatial aggregations only used same-day matches (Figure 4) and the temporal dynamics of the ecosystem had only a limited impact (as discussed in Section 4.2).

In addition, the results showed only minor differences among the macropixel variants. However, to account for processing errors and to avoid the risk of operating with a faulty pixel, very small spatial windows (e.g., only one pixel) are not recommended [31,53]. This was also partly reflected in this study, with the 1-pixel variant of aggregated OLCI EOMAP-MIP data performing slightly worse than the 3 × 3 macropixel variant for all target variables. Different macropixel variants of MSI data showed noticeable differences in the number of matches compared with the 15 × 15 macropixel variant. On average, 8% fewer matches were produced when the 15 × 15 macropixel results were compared with the smaller macropixel variants, probably due to higher number of invalid pixels when the measurement points were closer to land.

In the context of governmental monitoring programs, where the assessment of the status of the entire waterbody is the goal, aggregating at the waterbody scale can be advantageous for comparing in situ and remote sensing data. The spatial variability is averaged away, and thus the spatial heterogeneity is considered. Another advantage of waterbody-wide aggregation is the mitigation of effects occurring primarily close to the shore such as adjacency effects, shade from hills, or bottom reflection. This is particularly valid for waterbodies where the proportion of number of pixels close to the shore is high compared with the total number of pixels. Furthermore, waterbody-scale extractions usually yield slightly higher numbers of matches (Figure 4), because it is not so important where exactly the invalid pixels are located for the water-body-scale extraction still to be valid. For statistical purposes, it may thus also be advantageous to aim for waterbody-scale extractions.

On the other hand, macropixel-based products may be, on one hand, able to account for local conditions but, on the other hand, can be severely influenced by the neighboring pixels in a locally heterogeneous environment. The latter becomes even more influential when part of the spatial heterogeneity is attributable to random errors. Note that in this respect, in situ samples are always point samples attributed to precise locations with a very small spatial extent. The fact that waterbody-wide averaging does not deteriorate the statistical performance of satellite-based products in most cases (Figure 4) may indicate

that spatial dynamics at small scales are, in many cases, not strong enough that large-scale averaging yields similar patterns.

It needs to be noted that we conducted a broad-scale validation study using administrative data. Therefore, the validation here aimed to examine whether in situ and remote sensing data can be used comparably well for official monitoring. In a different context, for example, when it comes to optimizing algorithms, the existing protocols should be relied upon, which typically rely on macropixel variants and well-positioned sampling sites not too close to shore, where the spatial variability of the variables of water quality in question is relatively stable over the whole macropixel. In addition, a very tight match of satellite overpasses and in situ sampling is crucial (at temporal scales far below daily; see the discussion below) for these purposes; otherwise, the sampled water parcels do not exhibit the same characteristics of water quality due to possible temporal fluctuations in the concentrations at small time scales.

4.2. Temporal Windows

Regarding Objective (ii), different time windows performed similarly well with only minor differences and without clear patterns when evaluated with the MAE and bias (see Figure 6). However, RMSE showed an increase for various target variables (except Secchi depth) and sensor–processor combinations. The differences between MAE (log scale) and RMSE (linear) indicated an increase in the skewness of data's distribution and the presence of outliers with increasing time windows (see Section 4.5). This indicated that expanding the time window to up to 5 days can be useful in validation studies for all target variables because of a noticeable increase in the number of matches, as long as special attention is given to the appropriate treatment of outliers.

In addition, an expansion of the time lags for up to 10 days showed that the distribution of residuals between in situ and remote sensing data and the associated summary statistics (mean, median, and interquartile range) were relatively stable even for time lags of more than 5 days (Figures S1–S3).

4.3. Interplay between Temporal and Spatial Scales

A waterbody surface can be seen as a heterogenous system in motion, so that the spatial patterns depend on the time window. Two hours may lead to different spatial patterns at the surface due to the high advective transport in waterbodies, particularly under windy conditions [33] and during ice-melt or phytoplankton blooms. Such comparably rapid events impair the agreement between in situ and remote sensing data more strongly for macropixel variants than for evaluations based on the whole waterbody [9]. It would therefore seem advisable, particularly for the macropixel approaches, to match only in situ and satellite data that are within 2 h of each other [7]. In the present study, this was not possible due to the time information in the in situ data not being specified to the hour. The smallest time window available for a comparison of different spatial aggregations was the same day, which translated to a temporal mismatch between in situ and remote sensing data of up to 8 h. This is one explanation for some errors being relatively high, which complicated the evaluation of the spatial patterns.

4.4. Systematic Differences among Variables of Water Quality, Satellites, or Processors

The agreement between in situ and remote sensing data depends on multiple factors, e.g., the environmental and in-water conditions, the placement of the sensors, and the sensitivity of the algorithms, among others. For this study, two operational processing schemes were applied with different approaches to address atmospheric influences, sun and sky glint, and other interfering factors from the radiance leaving the water to derive the concentrations of the variables of water quality. These operations have to work over a wide range of geographical regions, may be affected by unpredictable uncertainties, and are further developed continuously. The abovementioned aspects led to the differences between sensor–processor combinations and the occurrence of occasionally strong outliers

irrespective of the spatial aggregation or time window applied. However, integrating all the alternative satellite products (S2-MSI or S3-OLCI, EOMAP-MIP or CyanoAlert) into one waterbody-specific average yielded a robust value when compared with the in situ-based values (unpublished data).

4.5. Error Metrics

MAE, bias, and RMSE are often applied in validation studies in parallel, although they are only partially complementary [29]. MAE and RMSE differ in their sensitivity to outliers and the distribution of the data [49]. RMSE, in contrast to MAE, penalizes high absolute deviations and an uneven distribution of the error due to it being calculated in linear space and the squaring of errors. However, they both addressed the pairwise agreement (accuracy) between remote sensing and in situ data. Since the residuals in regressions of satellite versus in situ observations increase with the mean—an argument for using log scale—the RMSE for the concentration of chlorophyll-a in eutrophic waterbodies should be much higher than for oligotrophic waterbodies. The opposite holds true for the log-based MAE, where smaller deviations had a big impact if the chlorophyll value was low. To give an example, the log-based MAE resulted in the same magnitude of error when the detected/measured chlorophyll was 0.1 or 1 µg/L compared with 10 and 100 µg/L because both have the same ratio. But the RMSE would weigh the pair at 10 and 100 µg/L to be far more erroneous because it is based on squared deviations at the linear scale. From the limnological point of view, however, a difference between 0.1 and 1 µg/L is irrelevant and hardly measurable in the laboratory, while a difference between 10 and 100 µg/L makes a huge difference in evaluations of the status. In summary, RMSE stresses errors at larger values (e.g., high chlorophyll) while MAE emphasizes deviations at low true values (e.g., low chlorophyll).

Therefore, the small and random differences between the different spatial aggregations across all error measures, with MAE and RMSE being not very different from each other, indicated that the presence of outliers or the distribution of the errors did not differ significantly among different spatial aggregations. However, we noted that RMSE was increasingly higher at large time windows for various target variables and sensor–processor combinations. This suggests an increase in the influential outliers and the skewness of the distribution of the errors with an increasing time window, which should be addressed when larger time windows are applied. This is also partially reflected in Figure S1, with the median and interquartile range being relatively stable up to a time lag of ±5 days, whereas the mean increased more.

Lastly, it needs to be considered that both datasets, remote sensing and in situ, contained errors and uncertainties. Therefore, the assumption that the error metrics fully evaluated the dataset is disputable. Given these results and occurrences of error, we agree with Papathanasopoulous et al. [54] and IOCCG [55] that satellite-based monitoring will complement rather than replace in situ sampling to increase the understanding of the spatiotemporal development of waterbodies, as well as to provide information where in situ data were (in the past) or are still sparse or non-existent.

5. Conclusions

In general, our results provide clear evidence that satellite-based products reflect the average condition of a given waterbody and are therefore suited for assessments of the status, such as are required in national and international legislation for the protection of water. The results showed that it does not necessarily pay to focus spatially on the exact sampling point. In many cases, waterbody-scale values achieved similar or slightly better accuracies and biases despite the large differences in the spatial aggregation and spatial variability. They also provided slightly larger sample sizes. Data on Secchi depth produced the least pronounced differences among all spatial aggregations, independent of the sensor and processor. Overall, the results did not show huge differences among different spatial aggregations, and no clear preferences for the type of spatial aggregation emerged.

The study has also shown that an expansion of the time window of up to ± 5 days can be practiced under certain conditions. Data products for Secchi depth showed only small and random differences among different temporal windows. In contrast, data products for turbidity and chlorophyll-a showed an increase in outliers with increasing time window for various sensor–processor combinations. Therefore, if applicable, the increasing occurrence of outliers must be considered in case of extending the time window. In summary, the results indicated that an expansion of the time window of up to ± 1 day or ± 5 days can be useful in validation studies because of the marked rise in number of matches, which were, on average, about 2.5 times (± 1 day) or 5 times (± 5 days) greater than same-day matches.

Besides these details on procedures of validation for satellite-based monitoring of water quality, our data on more than 100 waterbodies showed that averaging all the available values for a given waterbody can very likely reflect the status of the waterbody. Satellite-based information can therefore supplement the information collected via in situ monitoring, which can only be conducted at limited temporal and spatial scales, and thus can improve assessments of the ecosystem status of a given waterbody.

Appendix A

Table A1. Overview of the valid remote sensing scenes for turbidity, given as the number of scenes per sensor, processor, and spatial resolution, for all waterbodies over the period considered (2016–2020).

Turbidity	S2-MSI				S3-OLCI		
	3×3	5×5	15×15	Waterbody Scale	1×1	3×3	Waterbody Scale
CyanoAlert	14,639	16,250	13,534	17,403	10,803	14,301	13,114
EOMAP-MIP	15,895	16,004	16,083	19,630	16,990	18,542	20,730

Table A2. Overview of the valid remote sensing scenes for Secchi depth, given as the number of scenes per sensor, processor, and spatial resolution, for all waterbodies over the period considered (2016–2020).

Secchi Depth	S2-MSI				S3-OLCI		
	3×3	5×5	15×15	Waterbody Scale	1×1	3×3	Waterbody Scale
CyanoAlert	14,639	15,053	13,534	17,403	11,808	16,250	13,114
EOMAP-MIP	15,890	15,997	16,077	19,623	16,890	18,440	20,618

Table A3. Thresholds for the identification of outliers in the three target variables applied to both in situ and remote sensing data. Values were identified by expert judgement and visual inspection of the distributions of probability. Identification of the outliers resulted in removal of <<1% of the observations.

Target Variable	Unit	Minimum	Maximum
Chlorophyll-a	μg/L	0.01	400
Turbidity	FNU	0.01	100
Secchi depth	m	0.05	20
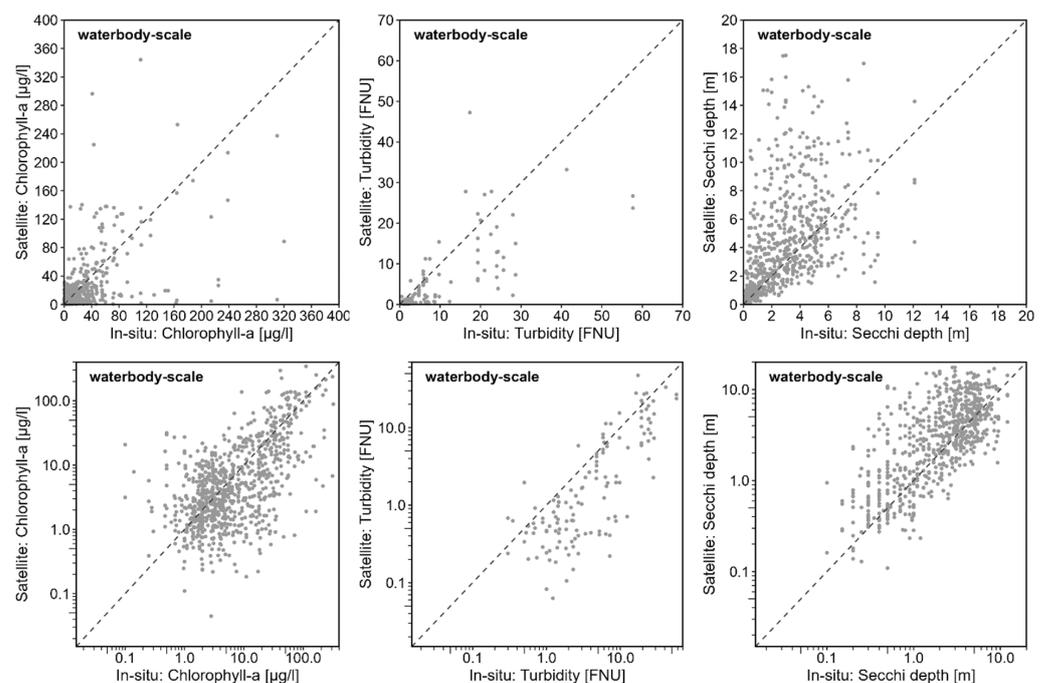
Figure A1. Scatterplots of chlorophyll-a (**left**), turbidity (**middle**), and Secchi depth (**right**) match-up dataset, log10 transformed (**bottom**) and untransformed (**top**).

Error metrics - Spatial aggregations

Sensor	Target-variable	MAE					Bias					RMSE					N					N Lakes				
		1	3x3	5x5	15x15	all	1	3x3	5x5	15x15	all	1	3x3	5x5	15x15	all	1	3x3	5x5	15x15	all	1	3x3	5x5	15x15	all
MSI	Chlorophyll-a	-	2.6	2.6	2.6	2.6	-	0.8	0.8	0.8	0.7	-	27.0	26.3	31.9	26.0	-	396	400	377	437	-	70	70	71	69
MSI	Turbidity	-	3.8	4.1	3.7	2.8	-	0.4	0.4	0.4	0.5	-	10.9	11.3	12.0	8.8	-	100	94	74	111	-	11	11	11	11
MSI	Secchi depth	-	1.8	1.8	1.9	1.8	-	1.4	1.4	1.5	1.5	-	3.9	3.8	3.7	3.2	-	382	384	360	419	-	69	70	71	71
OLCI	Chlorophyll-a	2.9	2.8	-	-	2.8	0.8	0.8	-	-	0.8	38.8	34.2	-	-	32.7	477	525	-	-	585	34	34	-	-	32
OLCI	Turbidity	3.5	2.9	-	-	2.6	0.3	0.4	-	-	0.4	13.9	12.9	-	-	11.5	36	51	-	-	43	2	2	-	-	2
OLCI	Secchi depth	1.9	1.9	-	-	2.0	1.2	1.2	-	-	1.3	2.8	2.7	-	-	2.8	414	501	-	-	482	32	33	-	-	31

Figure A2. Overall performance of different spatial aggregations using the match-up dataset (same-day) with MSI and OLCI data separated, as in Figure 4, but the two processors EOMAP-MIP and CynaoAlert combined. Table shows selected error metrics, number of observations (matches) and number of waterbodies (N Lakes) for the three target variables chlorophyll-a, turbidity and Secchi depth. Dark-grey shades indicate poorer performance of variant, while lightest shades represent the best performance.

Error metrics - Spatial aggregations

	MAE		Bias		RMSE		N		N Lakes	
Target-variable	3x3	all	3x3	all	3x3	all	3x3	all	3x3	all
Chlorophyll-a	2.7	2.7	0.8	0.8	31.3	30.0	921	1022	76	75
Turbidity	3.4	2.7	0.4	0.5	11.6	9.6	151	154	11	11
Secchi depth	1.9	1.9	1.3	1.4	3.3	3.0	883	901	75	76

Figure A3. Overall performance of two different spatial aggregations (3 × 3 macropixels and waterbody-scale variants) using the match-up dataset (same day) with data from both sensors (S2-MSI, S3-OLCI) and processors (EOMAP-MIP, CyanoAlert) combined. The table shows selected error metrics, the number of observations (matches; N), and the number of waterbodies (N Lakes) for the three target variables of chlorophyll-a, turbidity, and Secchi depth.

Error metrics - Temporal windows

Target-variable	MAE			Bias			RMSE			N			N lakes		
	±0d	±1d	±5d	±0d	±1d	±5d	±0d	±1d	±5d	±0d	±1d	±5d	±0d	±1d	±5d
Chlorophyll-a	2.70	2.66	2.79	0.80	0.81	0.79	31.32	36.54	35.61	921	2167	4625	76	93	97
Turbidity	3.44	2.86	2.86	0.39	0.48	0.50	11.63	11.84	13.00	151	421	1000	11	11	12
Secchi depth	1.89	1.81	1.82	1.31	1.27	1.31	3.25	3.17	3.38	883	2096	4539	75	91	96

Figure A4. Overall performance of different temporal windows using the match-up dataset (3 × 3 macropixels) with data of both sensors (MSI, OLCI) and processors (EOMAP-MIP, CyanoAlert)

combined. The table shows selected error metrics, the number of observations (matches), and the number of waterbodies (N Lakes) for the three target variables of chlorophyll-a, turbidity, and Secchi depth. Dark grey shades indicate poorer performance of the variant, while the lightest shades represent the best performance.

Table A4. Coefficient of variations (average) dependent on the temporal window, the target variables, the sensors, and the processors.

Sensor	Target Variable	CV (CyanoAlert)			CV (EOMAP-MIP)		
		Same Day	1 Day	5 Days	Same Day	1 Day	5 Days
S2-MSI	Chlorophyll-a	36.7	38.5	39.5	13.6	13.5	14.1
	Turbidity	18.3	19.4	20.7	18.2	15.2	14.4
	Secchi depth	17.8	19.0	19.7	8.8	8.4	8.5
S3-OLCI	Chlorophyll-a	25.1	22.8	27.9	23.5	23.5	23.2
	Turbidity	53.0	60.5	64.2	15.2	16.1	17.4
	Secchi depth	43.6	39.2	40.9	14.9	14.9	15.1

References

1. Postel, S.L. Entering an Era of Water Scarcity: The Challenges Ahead. *Ecol. Appl.* **2000**, *10*, 941. [CrossRef]
2. Brönmark, C.; Hansson, L.-A. Environmental issues in lakes and ponds: Current state and perspectives. *Environ. Conserv.* **2002**, *29*, 290–307. [CrossRef]
3. Stendera, S.; Adrian, R.; Bonada, N.; Cañedo-Argüelles, M.; Hugueny, B.; Januschke, K.; Pletterbauer, F.; Hering, D. Drivers and stressors of freshwater biodiversity patterns across different ecosystems and scales: A review. *Hydrobiologia* **2012**, *696*, 1–28. [CrossRef]
4. Tranvik, L.J.; Downing, J.A.; Cotner, J.B.; Loiselle, S.A.; Striegl, R.G.; Ballatore, T.J.; Dillon, P.; Finlay, K.; Fortino, K.; Knoll, L.B.; et al. Lakes and reservoirs as regulators of carbon cycling and climate. *Limnol. Oceanogr.* **2009**, *54*, 2298–2314. [CrossRef]
5. Williamson, C.E.; Saros, J.E.; Vincent, W.F.; Smol, J.P. Lakes and reservoirs as sentinels, integrators, and regulators of climate change. *Limnol. Oceanogr.* **2009**, *54*, 2273–2282. [CrossRef]
6. Dudgeon, D.; Arthington, A.H.; Gessner, M.O.; Kawabata, Z.-I.; Knowler, D.J.; Lévêque, C.; Naiman, R.J.; Prieur-Richard, A.-H.; Soto, D.; Stiassny, M.L.J.; et al. Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biol. Rev. Camb. Philos. Soc.* **2006**, *81*, 163–182. [CrossRef] [PubMed]
7. Palmer, S.C.; Kutser, T.; Hunter, P.D. Remote sensing of inland waters: Challenges, progress and future directions. *Remote Sens. Environ.* **2015**, *157*, 1–8. [CrossRef]
8. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 Establishing a Framework for Community Action in the Field of Water Policy. *Off. J. Eur. Union* **2000**, *L327*, 1–72. Available online: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2006:0397:FIN:EN:PDF (accessed on 27 April 2024).
9. Toming, K.; Kutser, T.; Laas, A.; Sepp, M.; Paavel, B.; Nõges, T. First Experiences in Mapping Lake Water Quality Parameters with Sentinel-2 MSI Imagery. *Remote Sens.* **2016**, *8*, 640. [CrossRef]
10. Odermatt, D.; Danne, O.; Philipson, P.; Brockmann, C. Diversity II water quality parameters from ENVISAT (2002–2012): A new global information source for lakes. *Earth Syst. Sci. Data* **2018**, *10*, 1527–1549. [CrossRef]
11. Sòria-Perpinyà, X.; Vicente, E.; Urrego, P.; Pereira-Sandoval, M.; Tenjo, C.; Ruíz-Verdú, A.; Delegido, J.; Soria, J.M.; Peña, R.; Moreno, J. Validation of Water Quality Monitoring Algorithms for Sentinel-2 and Sentinel-3 in Mediterranean Inland Waters with In Situ Reflectance Data. *Water* **2021**, *13*, 686. [CrossRef]
12. Reyjol, Y.; Argillier, C.; Bonne, W.; Borja, A.; Buijse, A.D.; Cardoso, A.C.; Daufresne, M.; Kernan, M.; Ferreira, M.T.; Poikane, S.; et al. Assessing the ecological status in the context of the European Water Framework Directive: Where do we go now? *Sci. Total Environ.* **2014**, *497–498*, 332–344. [CrossRef] [PubMed]
13. Hestir, E.L.; Brando, V.E.; Bresciani, M.; Giardino, C.; Matta, E.; Villa, P.; Dekker, A.G. Measuring freshwater aquatic ecosystems: The need for a hyperspectral global mapping satellite mission. *Remote Sens. Environ.* **2015**, *167*, 181–195. [CrossRef]
14. Zheng, G.; DiGiacomo, P.M. Uncertainties and applications of satellite-derived coastal water quality products. *Prog. Oceanogr.* **2017**, *159*, 45–72. [CrossRef]
15. Ho, J.C.; Michalak, A.M.; Pahlevan, N. Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature* **2019**, *574*, 667–670. [CrossRef] [PubMed]
16. Dörnhöfer, K.; Oppelt, N. Remote sensing for lake research and monitoring–Recent advances. *Ecol. Indic.* **2016**, *64*, 105–122. [CrossRef]

17. Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 Concerning the Management of Bathing Water Quality and Repealing Directive 76/160/EEC 2006 Feb 15. Available online: https://eur-lex.europa.eu/eli/dir/2006/7/oj (accessed on 27 April 2024).

18. Matthews, M.W.; Bernard, S.; Robertson, L. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sens. Environ.* **2012**, *124*, 637–652. [CrossRef]

19. Mozafari, Z.; Noori, R.; Siadatmousavi, S.M.; Afzalimehr, H.; Azizpour, J. Satellite-Based Monitoring of Eutrophication in the Earth's Largest Transboundary Lake. *Geohealth* **2023**, *7*, e2022GH000770. [CrossRef] [PubMed]

20. Meng, D.; Mao, J.; Li, W.; Zhu, S.; Gao, H. An Integrated Framework for Remote Sensing Assessment of the Trophic State of Large Lakes. *Remote Sens.* **2023**, *15*, 4238. [CrossRef]

21. Hu, M.; Ma, R.; Cao, Z.; Xiong, J.; Xue, K. Remote Estimation of Trophic State Index for Inland Waters Using Landsat-8 OLI Imagery. *Remote Sens.* **2021**, *13*, 1988. [CrossRef]

22. Seegers, B.N.; Werdell, P.J.; Vandermeulen, R.A.; Salls, W.; Stumpf, R.P.; Schaeffer, B.A.; Owens, T.J.; Bailey, S.W.; Scott, J.P.; Loftin, K.A. Satellites for long-term monitoring of inland U.S. lakes: The MERIS time series and application for chlorophyll-a. *Remote Sens. Environ.* **2021**, *266*, 112685. [CrossRef] [PubMed]

23. Hansen, C.; Burian, S.; Dennison, P.; Williams, G. Spatiotemporal Variability of Lake Water Quality in the Context of Remote Sensing Models. *Remote Sens.* **2017**, *9*, 409. [CrossRef]

24. Topp, S.N.; Pavelsky, T.M.; Jensen, D.; Simard, M.; Ross, M.R.V. Research Trends in the Use of Remote Sensing for Inland Water Quality Science: Moving Towards Multidisciplinary Applications. *Water* **2020**, *12*, 169. [CrossRef]

25. Zibordi, G.; Berthon, J.-F.; Mélin, F.; D'Alimonte, D.; Kaitala, S. Validation of satellite ocean color primary products at optically complex coastal sites: Northern Adriatic Sea, Northern Baltic Proper and Gulf of Finland. *Remote Sens. Environ.* **2009**, *113*, 2574–2591. [CrossRef]

26. Morales, J.; Stuart, V.; Platt, T.; Sathyendranath, S. (Eds.) *Handbook of Satellite Remote Sensing Image Interpretation: Applications for Marine Living Resources Conservation and Management*; IOCCG: Dartmouth, NS, Canada, 2011.

27. Schaeffer, B.A.; Schaeffer, K.G.; Keith, D.; Lunetta, R.S.; Conmy, R.; Gould, R.W. Barriers to adopting satellite remote sensing for water quality management. *Int. J. Remote Sens.* **2013**, *34*, 7534–7544. [CrossRef]

28. Bresciani, M.; Giardino, C.; Stroppiana, D.; Dessena, M.A.; Buscarinu, P.; Cabras, L.; Schenk, K.; Heege, T.; Bernet, H.; Bazdanis, G.; et al. Monitoring water quality in two dammed reservoirs from multispectral satellite data. *Eur. J. Remote Sens.* **2019**, *52* (Suppl. 4), 113–122. [CrossRef]

29. Loew, A.; Bell, W.; Brocca, L.; Bulgin, C.E.; Burdanowitz, J.; Calbet, X.; Donner, R.V.; Ghent, D.; Gruber, A.; Kaminski, T.; et al. Validation practices for satellite-based Earth observation data across communities. *Rev. Geophys.* **2017**, *55*, 779–817. [CrossRef]

30. Thouvenin-Masson, C.; Boutin, J.; Vergely, J.-L.; Reverdin, G.; Martin, A.C.H.; Guimbard, S.; Reul, N.; Sabia, R.; Catany, R.; Fanton-D'andon, O.H. Satellite and In Situ Sampling Mismatches: Consequences for the Estimation of Satellite Sea Surface Salinity Uncertainties. *Remote Sens.* **2022**, *14*, 1878. [CrossRef]

31. Bailey, S.W.; Werdell, P.J. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sens. Environ.* **2006**, *102*, 12–23. [CrossRef]

32. Kallio, K.; Koponen, S.; Ylöstalo, P.; Kervinen, M.; Pyhälahti, T.; Attila, J. Validation of MERIS spectral inversion processors using reflectance, IOP and water quality measurements in boreal lakes. *Remote Sens. Environ.* **2015**, *157*, 147–157. [CrossRef]

33. Rinke, K.; Huber, A.M.R.; Kempke, S.; Eder, M.; Wolf, T.; Probst, W.N.; Rothhaupt, K.-O. Lake-wide distributions of temperature, phytoplankton, zooplankton, and fish in the pelagic zone of a large lake. *Limnol. Oceanogr.* **2009**, *54*, 1306–1322. [CrossRef]

34. European Space Agency (ESA). Sentinel-2 User Handbook. 2015. Available online: https://sentinels.copernicus.eu/web/sentinel/user-guides/document-library/-/asset_publisher/xlslt4309D5h/content/sentinel-2-user-handbook (accessed on 21 March 2024).

35. European Space Agency (ESA). Sentinel-2 Factsheet. 2017. Available online: https://esamultimedia.esa.int/docs/EarthObservation/SENTINEL-2_sheet_170125.pdf (accessed on 21 March 2014).

36. European Space Agency (ESA). Sentinel-3 Factsheet. 2017. Available online: https://www.d-copernicus.de/fileadmin/Content/pdf/Sentinel-3_factsheet_jan2017.pdf (accessed on 21 March 2024).

37. EUMETSAT Sentinel-3 OLCI Marine User Handbook. 2021. Available online: https://www-cdn.eumetsat.int/files/2021-03/Sentinel-3%20OLCI%20Marine%20User%20Handbook.pdf (accessed on 21 March 2024).

38. Chen, J.; Chen, S.; Fu, R.; Li, D.; Jiang, H.; Wang, C.; Peng, Y.; Jia, K.; Hicks, B.J. Remote Sensing Big Data for Water Environment Monitoring: Current Status, Challenges, and Future Prospects. *Earth's Future* **2022**, *10*, e2021EF002289. [CrossRef]

39. IOCCG. Atmospheric correction for remotely-sensed ocean-colour products. In *Reports of the International Ocean-Colour Coordinating Group*; Wang, M., Ed.; No. 10; IOCCG: Dartmouth, NS, Canada, 2010; Available online: https://www.ioccg.org/reports/report10.pdf (accessed on 15 February 2024).

40. Werdell, P.J.; Bailey, S.W.; Franz, B.A.; Harding, L.W., Jr.; Feldman, G.C.; McClain, C.R. Regional and seasonal variability of chlorophyll-a in Chesapeake Bay as observed by SeaWiFS and MODIS-Aqua. *Remote Sens. Environ.* **2009**, *113*, 1319–1330. [CrossRef]

41. 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.; et al. Generalized ocean color inversion model for retrieving marine inherent optical properties. *Appl. Opt.* **2013**, *52*, 2019–2037. [CrossRef] [PubMed]

42. Mouw, C.B.; Greb, S.; Aurin, D.; DiGiacomo, P.M.; Lee, Z.; Twardowski, M.; Binding, C.; Hu, C.; Ma, R.; Moore, T.; et al. Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. *Remote Sens. Environ.* **2015**, *160*, 15–30. [CrossRef]

43. Brockmann, C.; Doerffer, R.; Peters, M.; Stelzer, K.; Embacher, S.; Ruescas, A. Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. *ESA Living Planet. Symp.* **2016**, *2016*, 740.

44. Pitarch, J.; Ruiz-Verdú, A.; Sendra, M.D.; Santoleri, R. Evaluation and reformulation of the maximum peak height algorithm (MPH) and application in a hypertrophic lagoon. *J. Geophys. Res. Ocean.* **2017**, *122*, 1206–1221. [CrossRef]

45. Wevers, J.; Müller, D.; Kirches, G.; Quast, R.; Brockmann, C. IdePix for Sentinel-3 OLCI Algorithm Theoretical Basis Document. Zenodo. 2022. Available online: https://zenodo.org/records/6517333#.YnkDfIzP33k (accessed on 22 February 2024).

46. Wevers, J.; Müller, D.; Scholze, J.; Kirches, G.; Quast, R.; Brockmann, C. IdePix for Sentinel-2 MSI Algorithm Theoretical Basis Document. Zenodo. 2021. Available online: https://zenodo.org/records/5788067 (accessed on 22 February 2024).

47. Nechad, B.; Ruddick, K.G.; Park, Y. Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sens. Environ.* **2010**, *114*, 854–866. [CrossRef]

48. Heege, T.; Kiselev, V.; Wettle, M.; Hung, N.N. Operational multi-sensor monitoring of turbidity for the entire Mekong Delta. *Int. J. Remote Sens.* **2014**, *35*, 2910–2926. [CrossRef]

49. Seegers, B.N.; Stumpf, R.P.; Schaeffer, B.A.; Loftin, K.A.; Werdell, P.J. Performance metrics for the assessment of satellite data products: An ocean color case study. *Opt. Express* **2018**, *26*, 7404–7422. [CrossRef] [PubMed]

50. Walther, B.A.; Moore, J.L. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* **2005**, *28*, 815–829. [CrossRef]

51. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]

52. EUMETSAT. *Recommendations for Sentinel-3 OLCI Ocean Colour Product Validations in Comparison with In Situ Measurements-Matchip Protocols*; EUMETSAT: Darmstadt, Germany, 2021.

53. Concha, J.A.; Bracaglia, M.; Brando, V.E. Assessing the influence of different validation protocols on Ocean Colour match-up analyses. *Remote Sens. Environ.* **2021**, *259*, 112415. [CrossRef]

54. Papathanasopoulou, E.; Simis, S.G.H.; Alikas, K.; Ansper, A.; Anttila, J.; Barillé, A.; Barillé, L.; Brando, V.; Bresciani, M.; Bučas, M.; et al. Satellite-Assisted Monitoring of Water Quality to Support the Implementation of the Water Framework Directive. Zenodo. 2019. Available online: https://zenodo.org/records/3463051 (accessed on 22 February 2024).

55. IOCCG. Earth observations in support of global water quality monitoring. In *Reports of the International Ocean-Colour Coordinating Group*; Greb, S., Dekker, A., Binding, C., Eds.; No. 17; IOCCG: Dartmouth, NS, Canada, 2018; 125p. [CrossRef]