2-16-2023

GLORIA - A globally representative hyperspectral *in situ* dataset for optical sensing of water quality

Moritz K. Lehmann
et al.

scientific data

OPEN

DATA DESCRIPTOR

GLORIA - A globally representative hyperspectral *in situ* dataset for optical sensing of water quality

Moritz K. Lehmann *et al.*[#]

The development of algorithms for remote sensing of water quality (RSWQ) requires a large amount of *in situ* data to account for the bio-geo-optical diversity of inland and coastal waters. The GLObal Reflectance community dataset for Imaging and optical sensing of Aquatic environments (GLORIA) includes 7,572 curated hyperspectral remote sensing reflectance measurements at 1 nm intervals within the 350 to 900 nm wavelength range. In addition, at least one co-located water quality measurement of chlorophyll *a*, total suspended solids, absorption by dissolved substances, and Secchi depth, is provided. The data were contributed by researchers affiliated with 59 institutions worldwide and come from 450 different water bodies, making GLORIA the *de-facto* state of knowledge of *in situ* coastal and inland aquatic optical diversity. Each measurement is documented with comprehensive methodological details, allowing users to evaluate fitness-for-purpose, and providing a reference for practitioners planning similar measurements. We provide open and free access to this dataset with the goal of enabling scientific and technological advancement towards operational regional and global RSWQ monitoring.

Background & Summary

Light from the sun reflected back across the water-air interface carries characteristic spectral signatures of several key water quality constituents due to their wavelength-specific absorption and scattering properties[1,2]. Chlorophyll *a*, total suspended solids, and colored dissolved organic matter are the dominant optically active constituents in inland and coastal waters[3,4], and common measures of water quality used for the management of ecosystem and public health[5–8]. Accurate measurements of spectral reflectance (i.e., the upwelling radiance normalized by the downwelling solar irradiance) are the foundation for synoptic and cost-effective environmental monitoring applications using satellite sensors, automated sensors installed near the water surface and portable instruments for manual field surveys[9].

Space-borne instruments have been providing accurate estimates of chlorophyll *a* and particle backscattering in the open ocean since the late 1990s with data from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) followed by many others, including the MEdium Resolution Imaging Spectrometer (MERIS) and Moderate Resolution Imaging Spectroradiometer (MODIS) in the 2000s, and the Ocean and Land Colour Instrument (OLCI) and Visible Infrared Imaging Radiometer Suite (VIIRS) over the last decade[10–17]. However, in coastal and inland waters, uncertainties in these estimates are typically much higher due to factors that include diverse atmospheric contributions, stray light from adjacent land areas, potentially uncorrelated variability of optically active constituents, and, in optically shallow water, bottom reflection[9,18–20]. Further, coarse-resolution imagers with a nominal resolution near 1 km are limited in nearshore and narrow systems where modern high-resolution missions like Landsat-8 and Sentinel-2 offer valid observations[21]. Overall, the retrieval of water quality in lakes, rivers, estuaries, lagoons and nearshore coastal waters remains an active area of research where improvements are needed so that satellite observations can fulfill their potential and become part of routine monitoring programs for ecosystem states, trends, and public-health alerting systems[22–26].

Large and globally representative *in situ* datasets are essential for the development and validation of bio-optical algorithms to support large-scale monitoring using satellite Earth observation technologies. Such

[#]A full list of authors and their affiliations appears at the end of the paper.

datasets are particularly scarce and geographically fragmented from inland and coastal waters as radiometric measurements are not part of most routine sampling programs and many lakes are remote and difficult to access.

We address these shortcomings with our GLObal Reflectance community dataset for Imaging and optical sensing of Aquatic environments (GLORIA). GLORIA includes over 7000 curated hyperspectral remote sensing reflectance (R_{rs}, sr^{-1}) and co-located chlorophyll *a* (Chl*a*, mg m^{-3}), total suspended solids (TSS, g m^{-3}), absorption by colored dissolved organic matter (CDOM) at 440 nm wavelength ($a_{CDOM}(440)$, m^{-1}) and Secchi depth (m) measurements. The data were contributed by researchers affiliated with 59 institutions in 20 countries who made the measurements for a range of objectives under diverse funding sources and resource levels, but shared attention to strict sampling protocols, tenacity to reach remote and inaccessible sites, commitment to establish long-term trend monitoring sites, and the recognition of the value of open-access datasets for public benefit. With its almost global coverage, geomorphic range of water bodies, and 30-year time span (Fig. 2), GLORIA represents the *de-facto* state of knowledge of *in situ* coastal and inland water bio-geo-optical diversity. Subsets of the data have already produced significant contributions to global algorithm development for the satellite-based estimation of Chl*a*, TSS, and $a_{CDOM}(440)$ using data-intensive machine-learning methods[27–31] or global semi-analytical approaches[32]. Where they were available, we also provide uncertainty estimates of R_{rs} and water quality measurements as standard deviations and means from replicate measurements. Nevertheless, some methodological detail which is currently considered relevant may not have been recorded at the time of observation, which limits our ability to retrospectively assess sources of uncertainty to subsets of the global dataset.

GLORIA builds upon the existing data repositories aimed at remote sensing studies of aquatic environments. We address poorly represented optically complex coastal and inland waters in existing open-data platforms such as the SeaWiFS Bio-optical Archive and Storage System (SeaBASS, https://seabass.gsfc.nasa.gov)[33,34]. In contrast to other relevant data repositories, such as the Lake Bio-optical Measurements and Matchup Data for Remote Sensing (LIMNADES, https://limnades.stir.ac.uk) database, GLORIA is open-access. By carrying out consistent quality control across the entire dataset, and providing comprehensive methodological details associated with each measurement, we have produced an analysis-ready, standalone data package for the community.

The commitment of space agencies towards maintaining and enhancing optical Earth observing systems and the burgeoning fleet of commercial platforms indicate that our coupled reflectance-water quality attribute dataset fills a strong need to facilitate algorithm and application development. We anticipate that our collection of field setups and methodologies will encourage targeted data collection for the calibration and validation of upcoming satellite sensors[35–37], as well as the growth of *in situ* observatories[38–40].

Methods

The GLORIA dataset was collated from the aquatic optics community of researchers or research groups working towards a range of goals including the routine monitoring of high-priority sites, one-off bio-optical characterization of a range of water bodies, data gathering to support algorithm development, or designated sampling for validating equivalent satellite-derived products. Efforts to gather this data started in 2018 with the second atmospheric correction intercomparison exercise (ACIX-II Aqua), an international collaboration to test processors that generate aquatic reflectance products from radiance measurements made at the top of the atmosphere[19]. Requests for contributions were made at pertinent conference sessions and via the research networks of individuals. These requests were for quality assured remote sensing reflectance spectra at 1 nm intervals within the 350 to 900 nm wavelength range and at least one co-located water quality attribute (Chl*a*, TSS, $a_{CDOM}(440)$, or Secchi depth), and associated uncertainties. The sections below provide more details of the data and processing.

Radiometric data collection and processing. The central radiometric quantity reported in our dataset is remote sensing reflectance, R_{rs} (sr^{-1}). It is defined as the ratio of the water-leaving radiance just above the water surface ($L_w(0+)$, W m^{-2} sr^{-1} nm^{-1}) over above-water downwelling irradiance (E_s, W m^{-2} nm^{-1})(Eq. 1, Fig. 1). We use the symbology of Ruddick *et al.*[41] with slight modifications:

$$R_{rs}(\lambda, \Theta, \phi) = \frac{L_w(\lambda, \Theta, \phi)}{E_s(\lambda)} \quad (1)$$

R_{rs} and L_w are dependent on the viewing nadir angle Θ (measured from the downward vertical axis) and azimuth viewing angle ϕ (measured clockwise from the sun); λ identifies the wavelength dependence. For aquatic remote sensing applications, it is conventional to define R_{rs} as derived from a sensor looking straight down $L_w(\lambda, \Theta = 0)$ where ϕ is not defined[42]. Therefore, we omit λ, Θ, and ϕ for notational brevity. Several methods and instruments were used for the measurement of the downwelling and upwelling radiometric quantities reported in our dataset. Here we provide brief descriptions of the broad types of methodologies used for their measurement, and a list at the end of this section gives a formal summary.

L_w can be measured directly using a radiometer just above the water surface, looking vertically down and shielded from light reflected off the water surface[43]. Other common techniques include measurement of the upwelling radiance at nadir below the water surface ($L_u(0\text{-})$)[44], or from above the water surface where the sensor is directed at a non-zero nadir angle (L_t)[45]. Both of these radiance measurements require conversions to L_w, which are referenced in the list at the end of this section. In brief, $L_u(0\text{-})$ can be derived by extrapolating upwelling radiance from measurements at practical depths below the water surface to just below the water surface[46]. Propagation through the water-air interface by accounting for the reduction of radiance by internal reflection off the water surface yields L_w. The estimation of L_w from L_t is more involved, as L_t contains a considerable amount of sky radiance reflected off the water surface into the sensor field of view (reflected sky radiance) in addition to $L_w(\Theta, \phi)$, where we note the angular dependence to emphasize the need for conversion to $\Theta = 0$.

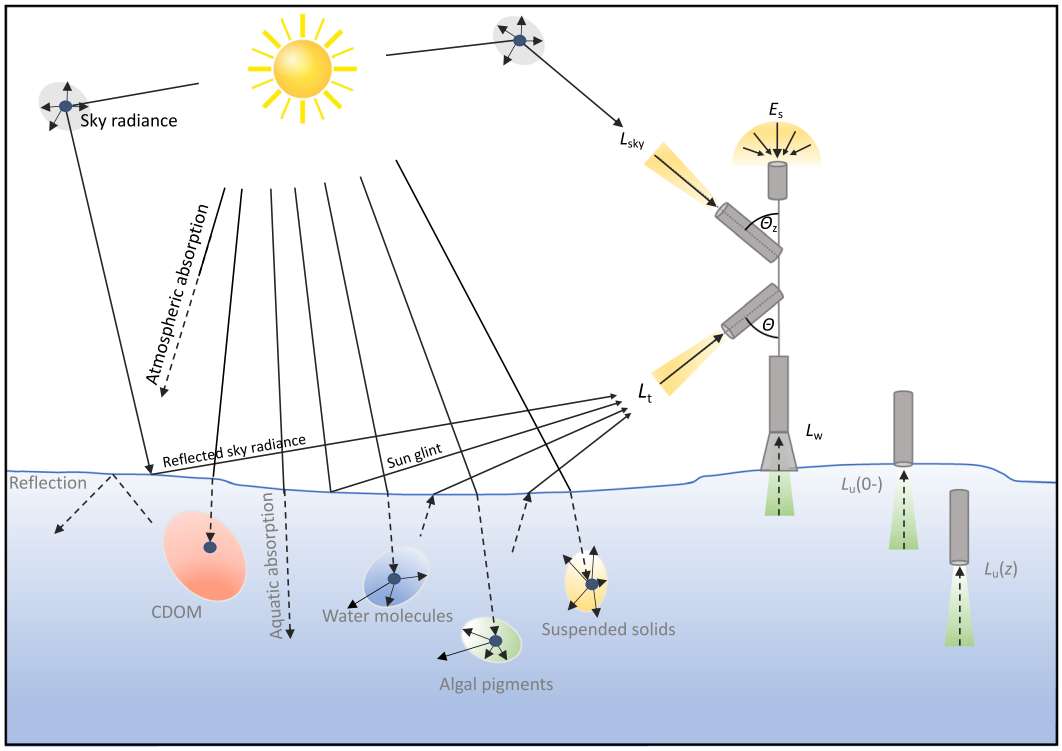

Fig. 1 Optical processes of absorption and scattering in the atmosphere and the water determine the amount and spectral nature of light received by a sensor. Remote sensing reflectance, the central radiometric quantity of the GLORIA dataset, is the ratio of the water-leaving radiance just above the water surface (L_w) over above-water downwelling irradiance (E_s).

Sky radiance (L_{sky}) is therefore usually measured simultaneously with L_t at the same azimuth angles and at zenith angles Θ_z (from the upward vertical axis) near 40°[42].

Three different approaches were used to measure E_s in the present dataset and a detailed review is provided by Ruddick *at al.*[47]. Most commonly E_s was measured directly using a plane irradiance sensor above the water surface directed straight upwards. The second most used method employed a downward pointing radiance sensor measuring the reflectance of a horizontally held Lambertian plaque with known reflective properties. This method has the advantage that a single sensor can be used for all measurements needed for the calculation of R_{rs}, potentially reducing cost, equipment load and uncertainties from the intercalibration of several sensors. In some cases, E_s was estimated from irradiance measurements below the water surface (just below the surface: $E_d(0-)$, or at depth z: $E_d(z)$). These measurements are typical of autonomous installations on vertical sensor chains or a single sensor package on a vertically profiling platform[44].

The instruments used for the radiometric measurements for each entry of the GLORIA dataset are part of the metadata (file *GLORIA_meta_and_lab.csv*) and are provided in the list at the end of this section. These include those customarily used for validation of satellite-derived aquatic reflectance, such as RAMSES (TriOS, Germany), HyperOCR (manufactured by Sea-Bird Scientific, USA; previously manufactured by Satlantic Inc., Canada) and C-OPS (Biospherical Instruments Inc., USA). The RAMSES and HyperOCR have 256 channel silicon photodiode array detectors with a 10 nm spectral resolution and a spectral sampling of 3.3 nm per pixel. The typical setup for RAMSES instruments for our dataset is an above-surface installation with a vertical E_s sensor and L_{sky} and L_t sensors at 40–42° zenith and nadir angles, respectively (Fig. 1). HyperOCR instruments are typically installed on a floating frame to measure E_s, and L_u or L_w at zero nadir angle while the HyperPRO (and HyperPro II) are free-falling setups of the HyperOCR designed to measure vertical profiles in the water column. The C-OPS configuration is similar to the HyperOCR, but the instrument only has 19 spectral bands of 10 nm width. The HyperSAS is a three-sensor setup of the HyperOCR for above-surface installation on structures overlooking the water or ships, similar to the RAMSES setup. The Water Insight WISP-3 is a self-contained handheld unit with optical inputs for E_s, L_{sky} and L_t leading to separate spectrometers[48].

A number of instruments used accommodate a single optical input into handheld units or portable instruments and need to be pointed to provide the different radiometric measurements (ASD FieldSpec range, Satlantic HyperGun, Spectra Vista, Spectral Evolution, Spectron Engineering and Photo Research SpectraScan devices).

Some investigators integrated compact spectrometers (manufactured by Ocean Insight, Inc., formerly known as Ocean Optics, Inc., USA) with data loggers and optical fibers on frames or poles that can be pointed away from observation platforms. Measurements would either be accomplished through several instruments and optical fibers oriented for the respective radiometric quantities, or a single sequentially reoriented fiber.

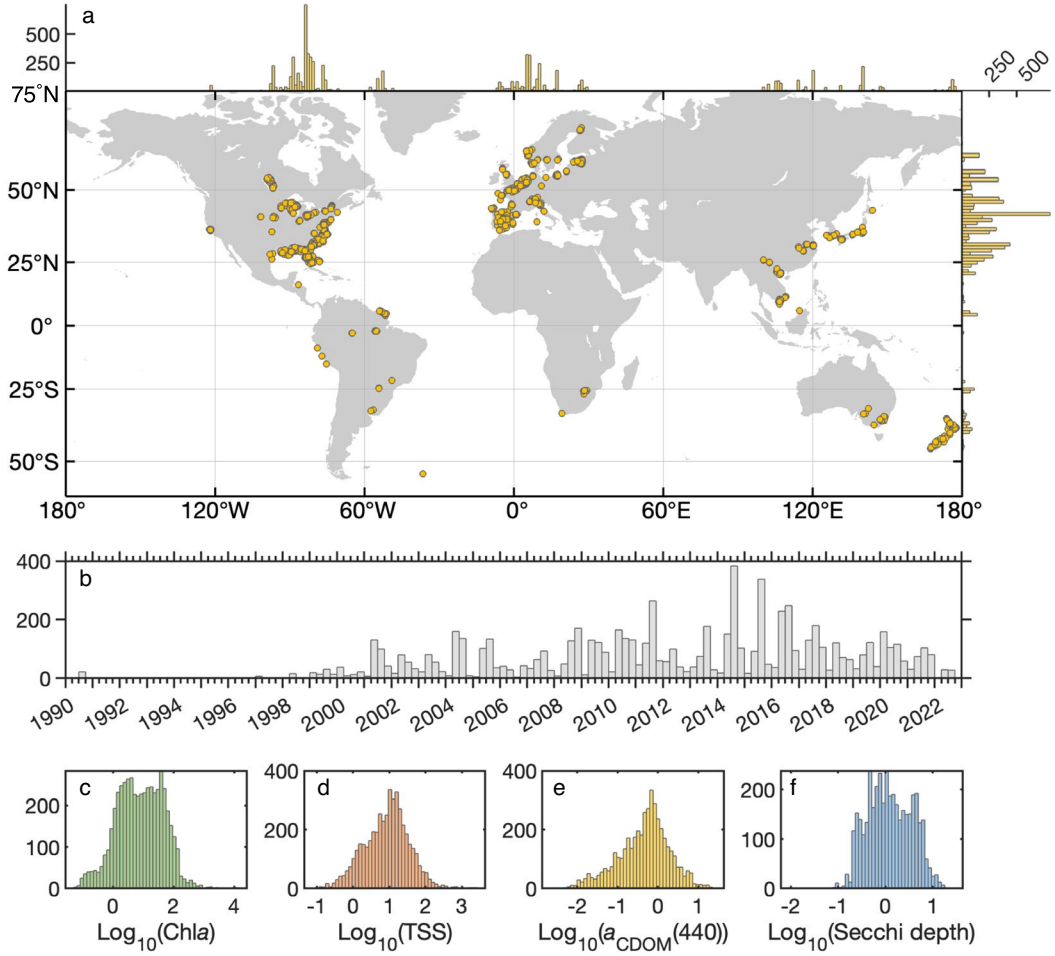

Fig. 2 Summary of the geographical, temporal and water quality distributions of the GLORIA samples. (**a**) Dots mark the location of each sample and the histograms on the edges of the map show the longitudinal and latitudinal distributions of the dataset. (**b**) The earliest samples were collected in 1990 and the sampling effort has been steady since 2001. (**c**–**f**) Histograms of log-transformed water quality attributes illustrate the extreme range of values and their typical log-normal distributions.

Data contributors provided radiometric measurements interpolated to 1 nm intervals over the 350 to 900 nm wavelength range. The instrument-specific bandwidths of the original measurements are provided in the data table (file *GLORIA_meta_and_lab.csv*, column 'Spectral_resolution_nm'). Due to instrument and processing constraints, some spectra span the range from 400 to 750 nm, or nearby bounds. The radiometric data for each GLORIA entry may be from a single measurement, or the mean or median of several measurements over a time interval. When available, the data contributors provided the spectral R_{rs} means, standard deviations, and numbers of measurements for sampling events. Quality control was conducted on all received spectra (see section *Technical validation*).

The measurement setups and instruments used for radiometric measurements are listed below. The number of the method corresponds to the column 'Measurement_method' in *GLORIA_meta_and_lab.csv*. References to published descriptions of the approach and applications are provided where available.

1. **Sequential L_t, L_{sky}, and E_s via a plaque on MP (moving platform)**
   Instruments: ASD FieldSpec, Photo Research PR-650 SpectraScan Colorimeter, Sea-Bird Scientific/Satlantic HyperGun, Spectra Vista GER1500, Spectral Evolution SR-3500/PSR-1100f, Spectron Engineering SE-590, TriOS RAMSES
   Approach: Mobley[45]
   Applications: Bresciani *et al.*[49]; Kudela *et al.*[50]; Zolfaghari *et al.*[51]
2. **L_t, L_{sky}, and E_s on MP**
   Instruments: Water Insight WISP-3
   Approach: Mobley[45]
   Applications: Hommersom *et al.*[48]
3. **L_u(0-) and E_s on pole connected to a spectrometer via fiber optics from MP or water edge**
   Instruments: Ocean Insight/Ocean Optics USB2000/USB2000 + /USB4000

Approach: Chipman *et al.*[52]
Applications: Gurlin *et al.*[53]; Schalles and Hladik[54]; Li *et al.*[55]; Mishra *et al.*[56]; Brezonik *et al.*[57]; Werther *et al.*[58]

4. **$L_w(0+)$ skylight blocked and E_s afloat away from MP**
Instruments: Sea-Bird Scientific/Satlantic HyperOCR
Approach: Lee *et al.*[55]
Applications: Wei *et al.*[59]
5. **$L_u(0-)$ afloat away from MP, E_s on MP**
Instruments: Sea-Bird Scientific/Satlantic HyperOCR, TriOS RAMSES
6. **L_t, L_{sky}, and E_s on MP**
Instruments: Sea-Bird Scientific/Satlantic HyperSAS, TriOS RAMSES
Approach: Mobley[45]; Simis and Olsson[60]
Applications: Qin *et al.*[61]; Warren *et al.*[62]
7. **L_t, L_{sky}, and E_s on a frame deployed on MP**
Instruments: TriOS RAMSES
Approach: Mobley[45]; Mobley[63]
Applications: Maciel *et al.*[64]; Cairo *et al.*[65]; da Silva *et al.*[66]
8. **$L_u(0-)$ and $E_d(0-)$ in-water profiling from MP, E_s on MP**
Instruments: Biospherical C-OPS, Sea-Bird Scientific/Satlantic HyperOCR, TriOS RAMSES
Approach: Mueller *et al.*[44]; Lubac and Loisel[67]
Applications: Binding *et al.*[68]
9. **$L_u(0-)$ and $E_d(z)$ units on a depth adjustable bar (measurements at −0.21 and −0.67 m) on a frame afloat away from MP, E_s unit lifted above water surface for E_s**
Instruments: TriOS RAMSES
Approach: Fritz *et al.*[69]
10. **$L_u(0-)$ and $E_d(0-)$ from winch on MP, E_s on MP**
Instruments: TriOS RAMSES
Approach: Zibordi and Talone[70]
11. **L_t and E_s on pole from water edge**
Instruments: TriOS RAMSES
Approach: Kutser *et al.*[71]
12. **$L_u(0-)$ and $E_d(0-)$ autonomous in-water profiling from a fixed platform**
Instruments: Sea-Bird Scientific/Satlantic HyperOCR
Approach: Mueller *et al.*[44]
Applications: Minaudo *et al.*[72]
13. **Sequential L_t and E_s via a plaque, mounted on gimbal stabilized pole from MP**
Instruments: Ocean Insight/Ocean Optics STS-VIS
14. **$L_u(0-)$ (and $E_d(0-)$ only for depth information) from in-water profiling from MP, E_s recorded simultaneously from same MP very close to profiler deployment**
Instruments: TriOS RAMSES
Approach: Mueller *et al.*[44]; Stramski *et al.*[73]
Applications: Bracher *et al.*[74]; Tilstone *et al.*[75]
15. **L_t, L_{sky}, E_s, combined with one L_u unit (aperture at −0.05 to −0.10 m) placed on a pole**
Instruments: TriOS RAMSES
16. **Sequential $L_u(0-)$ and E_s via a plaque, both measurements using an optical fiber to a black masked perspex tube**
Instruments: Spectron Engineering SE-590
Approach: Dekker[76]
17. **$L_u(0-)$ and $E_d(z)$ units on a floating frame (measurements at −0.4 m (L_u) and −0.1 m (E_d)) drifting 10 m away from vessel**
Instruments: TriOS RAMSES
Approach: Fritz *et al.*[69]

SeaBASS data. GLORIA includes approximately 1100 entries from SeaBASS[33]. We searched SeaBASS for reflectance spectra with concomitant water quality measurements and ensured that these are from inland and coastal waters only by mapping sampling locations of all records from water depths less than 200 m. Where water depth was not part of the SeaBASS record, we assigned it based on the General Bathymetric Chart of the Ocean (GEBCO_2021 Grid sub-ice topo/bathy)[77]. Several metadata fields were unavailable for this data, but SeaBASS dataset identifiers are provided to allow further research if needed. All SeaBASS data were included in our quality control process. While SeaBASS allows the upload of uncertainty data for radiometry and water quality, the entries we located for inland and coastal waters did not contain this information.

Water sample analysis. Water quality attributes Chl*a*, TSS and $a_{CDOM}(440)$ were determined using well established high-accuracy laboratory methods. The method for each analysis is identified in the columns 'Chl_method', 'TSS_method', and 'aCDOM_method' in the file *GLORIA_meta_and_lab.csv* and method details are provided in *GLORIA_variables_and_methods.xlsx*. Where available, data means and standard deviations from replicate analyses of Chl*a*, TSS and $a_{CDOM}(440)$ are provided in separate files.

The most frequently used methods for Chl*a* were via solvent-based pigment extraction from filter pads followed by fluorometric (U.S. EPA 445.0) or spectrophotometric (U.S. EPA 446.0) analysis. In the majority of samples, pigments were extracted in 90% acetone with the aid of mechanical tissue grinding. Modifications of

these methods included the use of 90% acetone buffered with $MgCO_3$ and different approaches to support the mechanical breakdown of the algal cells. Other methods for Chl*a* followed national and international standards (DIN 38412-16:1985-12, NEN 6520, HJ 897–2017, SL88-2012 and ISO 10260:1992). Methods which included a correction for phaeophytin, a degradation product of Chl*a*[78], are indicated by a flag ('1') in the data table (column 'Phaeophytin_correction') and the corresponding Chl*a* value is found in column 'Chla'; where phaeophytin was not corrected for the flag is '0' and Chl*a* is provided in column 'Chl_plus_phaeo' unless the correction for phaeophytin was not applicable as for certain fluorometric instrument setups[79]. Many investigators also used high-pressure liquid chromatography (HPLC) for Chl*a* determination and the Chl*a* value is found in column 'Chla'. The only exception to lab-determined Chl*a* are measurements from the Thetis profiler in Lake Geneva (Switzerland) where Chl*a* associated with R_{rs} measurements was estimated from absorption line height at 676 nm[80] and the linear relationship between the night-time fluorometric Chl*a* (measured by a WetLabs ECO Triplet BBFL2W) with absorption line height (average coefficient of determination: $R^2 = 0.92$)[72].

TSS concentration was measured gravimetrically by weighing the dried residue of a water sample filtered on a pre-combusted and pre-weighed filter pad. $a_{CDOM}(440)$ was generally quantified following Mitchell *et al.*[81]. Therefore, the optical density of water samples, typically filtered through 0.2 μm pore size polycarbonate membranes to remove particulates, was measured in a spectrophotometer and converted to absorption. Secchi depth was determined as the depth at which a disk, typically black and white of 20 or 30 cm in diameter, is no longer visible by an observer when it is lowered into the water[82,83].

Ancillary and metadata. Each data entry is associated with fields identifying the data contributor, cross-references to other databases, and details describing the sampling site and environmental conditions. Several categorical variables allow cursory stratification of the dataset according to water body type (lake, estuary, coastal ocean, river or other), data collection purpose (e.g., routine surface water monitoring or event-driven sampling), dominant biogeochemical water type (e.g., sediment-dominated or algal-dominated), and optical stability (e.g., low for shallow lakes, rivers and estuaries or high for deep lakes and some coastal ocean environments).

Specific characteristics of the sampling event such as geocoordinates, date and time stamps, environmental conditions (e.g., cloud cover, wind speed and wave height), and environmental settings (e.g., elevation above sea level, dominant land cover and slope) are provided where known. Several metadata fields provide cross references to details of instrumentation, measurement and processing methods for all radiometric and water quality data.

Data Records

The GLORIA dataset is hosted at the PANGAEA Data Publisher for Earth & Environmental Science[84]. The data is contained in several comma-separated value (csv) files and a Microsoft Excel file provides keys to column names and method details (Table 1). Individual data points are identified across all files using the GLORIA_ID.

The 7,572 GLORIA R_{rs} spectra originate from 31 countries over an almost global geographical range from 67°N to 54°S and from 122°W to 178°E (Fig. 2) with the majority of samples from lakes (60%), followed by coastal waters (32%), estuaries (4%), and the remainder from rivers and other water body types. The wide range of radiometric and water quality measurements in GLORIA (Fig. 2) is consistent with the global diversity of R_{rs} spectral shapes with respect to optical water types[85,86] and visual color ranges[87,88] (Fig. 3). The range of water quality attributes is comprehensive and their frequency distributions are shown in (Fig. 2).

Technical Validation

All data submitted for inclusion into this compilation had undergone quality control by the providers. Our curation process included detailed information recovery with them to ensure sampling, sample processing, and laboratory analysis methods are fit for purpose. Further checks on the gathered data were carried out as described below.

Reflectance spectra. Reflectance spectra were checked for outliers and unrealistic spectral shapes using a series of quality control indicators (Table 2). By flagging, but keeping, spectra with moderate or suspected quality issues, we were able to retain a larger dataset and we advise the user to inspect the flags to evaluate the dataset for their purposes. The quality control methods are described below. Data entries with quality issues are identified by setting the corresponding quality flag to one (1) in the file *GLORIA_qc_flags.csv*.

The first round of quality control was a procedural detection of high-frequency variability (suspected noise), baseline shifts (e.g., from suboptimal glint removal), the presence of an oxygen absorption feature near 762 nm (e.g., from sensor intercalibration issues), and negative slopes in the ultraviolet to blue part of the spectrum (e.g., from suboptimal diffuse sky radiance correction). These are the first five flags in Table 2.

Additionally, we calculated the Quality Water Index Polynomial (QWIP) score[89]. This approach was developed to identify hyperspectral aquatic reflectance data that fall outside general trends observed in a large dataset from optically deep waters. Briefly, the QWIP is a 4th order polynomial which describes a well-formed central tendency for a spectrally integrated metric (Apparent Visible Wavelength[90], AVW) to predict a Normalized Difference Index (NDI; $\lambda = 492, 665$ nm) across a continuum of water types. For a given spectrum, the difference between the calculated NDI and that predicted by the AVW is known as the QWIP score. If a given QWIP score exceeded a prescribed deviation from the polynomial relationship, in this case |0.2|, the data was identified by the flag 'QWIP_fail' in the file *GLORIA_qc_flags.csv* (Table 2). AVW and the QWIP score are provided in the file *GLORIA_qc_ancillary.csv* (Table 3).

On visual inspection, some spectra that passed the above criteria still appeared to have subtle problems. Further issues may be caused by instrument drift, instrument shading, stray light contamination, or errors during sky glint correction, and are often exacerbated by environmental conditions[59]. Such suspicious spectra can be recognized by experienced practitioners familiar with how inherent optical properties of surface waters vary

Filename	Description
GLORIA_variables_and_methods.xlsx	Excel file with several sheets:
	Data headers: Key to columns and units in the ancillary and metadata table (*GLORIA_meta_and_lab.csv*).
	Radiometry methods: Details of the instruments and their setups for the radiometric measurements.
	Chla methods, TSS methods, aCDOM Methods: Method details for the respective water quality measurements. Cross-referencing to data entries requires the dataset ID and methodology name from the table *GLORIA_meta_and_lab.csv*.
	References: List of references cited in this file.
GLORIA_meta_and_lab.csv	Ancillary information, metadata and water quality measurements associated with each R_{rs} spectrum. The data fields in this file are defined in *GLORIA_variables_and_methods.xlsx* sheet 'Data headers'.
GLORIA_Rrs.csv	Remote sensing reflectance (R_{rs}, sr^{-1}) spectra (Rrs_350, Rrs_351, …, Rrs_900). The first column is the GLORIA sample ID.
GLORIA_Es.csv	Above-water downwelling irradiance (E_s, W m^{-2} nm^{-1}) spectra (Es_350, Es_351, …, Es_900). The first column is the GLORIA sample ID.
GLORIA_Lw.csv	Water-leaving radiance just above the water surface (L_w, W m^{-2} sr^{-1} nm^{-1}) spectra (Lw_350, Lw_351, …, Lw_900). The first column is the GLORIA sample ID.
GLORIA_Lt.csv	Above-water upwelling radiance (L_t, W m^{-2} sr^{-1} nm^{-1}) spectra (Lt_350, Lt_351, …, Lt_900). The first column is the GLORIA sample ID.
GLORIA_Lu.csv	Upwelling radiance just below the water surface (L_u, W m^{-2} sr^{-1} nm^{-1}) spectra. The first column is the GLORIA sample ID.
GLORIA_Lsky.csv	Sky radiance (L_{sky}, W m^{-2} sr^{-1} nm^{-1}) spectra (Lsky_350, Lsky_351, …, Lsky _900). The first column is the GLORIA sample ID.
GLORIA_qc_flags.csv	Quality control (QC) flags for each QC procedure described in Table 2. A value of 1 indicates that the issue has been detected.
GLORIA_qc_ancillary	Ancillary information for quality control flags listed in Table 3.
GLORIA_Rrs_mean	Mean of R_{rs} measurements and the number of replicates of entries where the standard deviation is available.
GLORIA_Rrs_std	Standard deviation of R_{rs} measurements. The number of replicates for the calculation of the standard deviation is provided in *GLORIA_Rrs_mean*. Standard deviation is only available for a subset of the dataset.
GLORIA_waterqual_uncert	Mean, standard deviation and number of replicates for water quality measurements. The replicate type is specified as 'field' (separate water samples taken in the field) or 'lab' (replicate analyses of the same water sample). This information is available for a subset of the dataset.

Table 1. Files of the GLORIA dataset and their content.

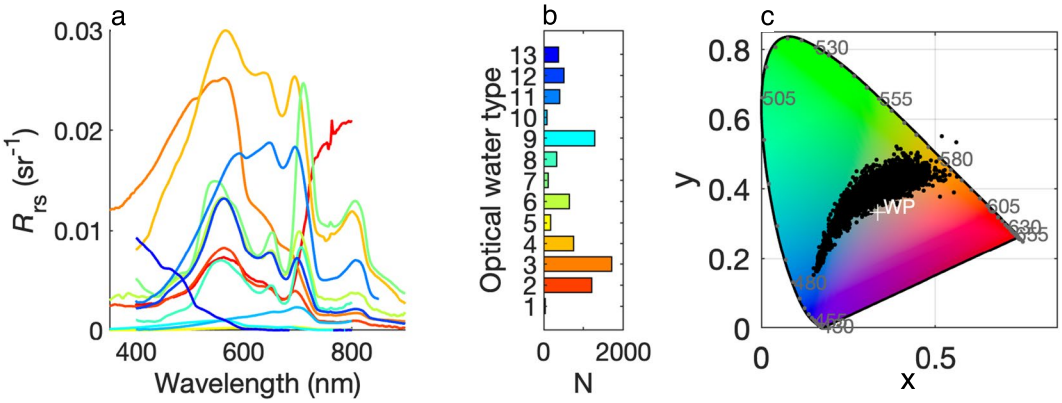

Fig. 3 Summary of the diversity of GLORIA's R_{rs} spectra. (**a**) Thirteen R_{rs} spectra chosen at random, one from each optical water type displayed in b. (**b**) Bar chart of the number of GLORIA spectra assigned to each optical water type from Spyrakos *et al.*[85]. (**c**) Chromaticity diagram[98] showing the visual color derived from each GLORIA R_{rs} spectrum using the tristimulus weighting functions according to the Commission Internationale de l'Éclairage (CIE)[99]; WP: white point.

naturally and determine reflectance through radiative transfer processes (Fig. 1)[91]. Utilizing this knowledge within the co-author community, we conducted systematic expert elicitation by randomly dividing the R_{rs} spectra into batches of 400 to 700 and assigning each batch to an expert for identifying suspicious looking data. The spectra that were flagged 'Suspect' were then evaluated by three more experts for the purpose of improving

Flag name Number of cases	Description and method
Noisy_red 40	High-frequency variability, potentially instrument noise, near the red end: spectra were standardized to zero mean and unit standard deviation. A 4th order polynomial was fitted over the interval 750–900 nm. Spectra with a root-mean square error (RMSE) >0.2 were flagged. This threshold was determined using visual inspection of the distribution of RMSEs with respect to spectral shapes.
Noisy_blue 15	High-frequency variability, potentially instrument noise, near the blue end: spectra were standardized to zero mean and unit standard deviation. A 4th order polynomial was fitted over the interval 350–400 nm. Cases where root-mean square error >0.15 were flagged (threshold determined using visual inspection of the distribution of RMSEs with respect to spectral shapes).
Baseline_shift 164	Spectra shifted up are those where the minimum R_{rs} is 60% of its median. This percentage corresponds approximately to 1.5 times the interquartile range above the upper quartile of the baseline-percent distribution of the entire GLORIA dataset. Spectra shifted down are those with at least 20 negative values and either: • a negative linear slope in the interval 765–900 nm $<-8.75\times10^{-7}$ sr^{-1} nm^{-1} (the slope threshold was determined as the bound of the lower quartile) and >50% negative R_{rs} values in this spectral region; or • >70% negative R_{rs} in the interval 765 nm-900 nm; or • at least 20 negative R_{rs} at in the interval 350–450 nm.
Oxygen_signal 1311	Spectra where Oxygen_peak_height >0.1 (Table 3). This threshold was determined using visual inspection of the distribution of peak heights with respect to spectral shapes.
Negative_uv_slope 139	Negative slopes in the ultraviolet to blue end: The spectra were standardized to zero mean and unit standard deviation. A straight line was fitted over the interval 350–420 nm and spectra with slopes <−0.005 were flagged. This threshold was determined using visual inspection of the distribution of slope values.
QWIP_fail 278	Spectra failing a statistical quality control metric based on Apparent_visible_wavelength (Table 3). The QWIP score exceeded a value of |0.2|.
Suspect 226	Spectra identified during expert elicitation as potentially fraught with measurement problems.
Flagged 1779	A one in this column indicates the presence of at least one flag from the tests described in this table.

Table 2. Quality control tests and associated flag names in table *GLORIA_qc_flags.csv*.

Column name	Description and method
Oxygen_peak_height	Local maximum or minimum in R_{rs} near 762 nm due to absorption of light by oxygen: The spectra were standardized to zero mean and unit standard deviation. A straight line was fitted to the interval between the median values of (745–755 nm) and (775–785 nm). The maximum absolute residual of the standardized spectrum near 762 nm was recorded and is provided as a value in this column.
Apparent_visible_wavelength	Defined as the weighted harmonic mean of the visible (400–700 nm) reflectance wavelengths. This metric is used to assess the directionality and magnitude of shifts in the spectral shape of remote sensing reflectance.
QWIP_score	The QWIP score represents the difference between a calculated Normalized Difference Index (NDI; $\lambda = 492, 665$ nm) and that estimated empirically from the spectrum's Apparent Visible Wavelength (AVW). Data producing absolute QWIP scores exceeding a value of ±0.2 have been found to exhibit spectral shapes that deviate from central tendencies typically observed in aquatic reflectance data.

Table 3. Ancillary information for quality control flags in table *GLORIA_qc_ancillary.csv*.

consistency across the batches from different individuals. The resulting set of suspect spectra are identified by the flag 'Suspect' in the file *GLORIA_qc_flags.csv* (Table 2).

Uncertainty in R_{rs} from above-surface measurements by means of reconstruction with a coupled water-atmospheric radiance model. Determining the uncertainty inherent in R_{rs} observations is challenging because of the variable nature of illumination and water surface conditions during repeat observations. This is especially true for measurements of upwelling light made above the water surface where sun glint and reflected sky radiance contribute to L_t, which applies to about 42% of the samples in GLORIA. To a large extent, spurious observations resulting from such random effects were already removed at source, such that the remaining variability is the result of various quality screening procedures and expert interpretation. However, it is possible to use models of atmospheric irradiance and bio-optical properties to model the most likely contribution of sun glint and reflected sky radiance on the R_{rs} observation, and thereby test the reported R_{rs} for physical consistency. To this end, we used the 3C algorithm[92] to reconstruct R_{rs} from records where L_t, L_{sky} and E_s were available.

3C provides a reconstruction of R_{rs} using nonlinear optimization of atmospheric and water optical models, allowing for a range of optical properties to solve the relationship between the upwelling radiance and downwelling irradiance provided as input. Due to the flexibility of the surface corrections, 3C is proposed to enable robust R_{rs} to be obtained across a wide range of measurement conditions. The resultant 3C-R_{rs} is expected to have reduced propagation of error from the variable spectral shape of sky reflectance and glint. This provides an advantage over methods which consider these corrections either constant, or a function of wind speed[60], which is the case for the majority of R_{rs} from above-surface measurements reported in the GLORIA database (column 'Skyglint_removal' in *GLORIA_meta_and_lab.csv*). The difference between 3C-R_{rs} and the originally reported R_{rs} is, therefore, an approximate measure of algorithmic uncertainty. A close match between the 3C reconstruction and the originally reported R_{rs} provides confidence that the reported observation was physically consistent.

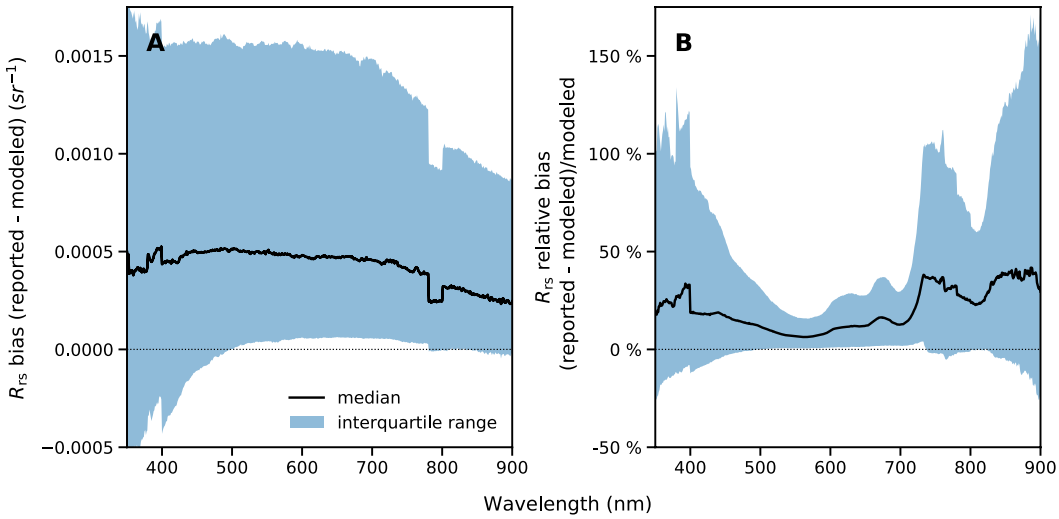

Fig. 4 Spectral bias of reported R_{rs} compared with 3C-modeled R_{rs} from 1589 spectra for which L_t, L_{sky} and E_s were available. (**A**) Median and interquartile (reported - modeled). (**B**) Relative bias in R_{rs} (reported - modeled)/modeled. Discontinuities in the bias spectrum are caused by sensors having different wavelength ranges within parts of the dataset.

Larger discrepancies are assumed to be associated with challenging observation conditions, resulting in suspect L_{sky}, L_t or E_s, but can also be caused by water or atmospheric properties which the model cannot reconstruct.

For this analysis, we used the 1589 spectra which included L_t, L_{sky}, E_s, observation time, and geographic location, and for which the method to calculate R_{rs} was not already based on 3C. This analysis is also independent from the quality flagging in the previous section, so that all observations were included and the results present a worst-case scenario which best represents the algorithmic uncertainty inherent to calculating R_{rs}, albeit without knowledge of quality control criteria applied prior to the data being reported. The 3C water optical model was configured with wide bounds for the concentration of Chl*a* (initial condition 5 mg m^{-3}, range 0.01–1000 mg m^{-3}) and TSS (initial condition 10 g m^{-3}, range 0–1000 g m^{-3}) whilst otherwise configured as detailed in Groetsch *et al.*[92] and Jordan *et al.*[93].

The median bias between reported and 3C-R_{rs} was in the order of 0.0005 sr^{-1}, with 3C yielding lower R_{rs}, as should be expected because incomplete correction relying on a static correction factor for surface reflections leads to higher R_{rs} (Fig. 4A). Bias gradually decreased with wavelength, which suggests the reported data have been suboptimally corrected for diffuse sky radiance. There is considerable spread in the model-observation bias, in the order of 0.00004 to 0.0016 sr^{-1} for R_{rs}(560) in the interquartile range.

In relative terms (Fig. 4B), median bias in R_{rs} between observed and 3C-R_{rs} is smallest in the green spectral range (order of 6.4%), where peak R_{rs} amplitude is typically observed in this dataset, and largest in the UV and NIR regions of the spectrum where R_{rs} is typically lower. The spread (interquartile range) in the relative bias in R_{rs}(560) is 5–16%, but much wider in the UV and NIR range, exceeding −30% and 170%.

The largest differences in R_{rs} bias between reported and 3C spectra were found between contributed datasets, rather than between observation methods. The majority of datasets showed absolute relative differences in R_{rs}(400–800) in the 0–10% range, but there are also cases where the difference exceeds 100%.

This analysis points to an overall high degree of uncertainty in the methods using above-water L_t measurements and the need for rigorous quality control by observers. For future work, we suggest adding R_{rs} model reconstruction as part of the data collection effort, which allows inspection of glint terms to objectively flag observations as suspect, before other quality controls are implemented. Furthermore, to support future algorithmic improvements (e.g., to elaborate bidirectional reflectance distribution functions), all component spectra and observation geometries should be included in datasets and these should be reported at the native resolution of each sensor involved to avoid convolution error when calculating R_{rs}[94].

Water quality. The water quality measurements were investigated using frequency distributions to identify outliers. Separate frequency distributions were created by 'Water_type', a subjective classification assigned by the data contributors according to the dominant optical constituent for each water body (TSS-dominated, Chl*a*-dominated, CDOM-dominated, Chl*a* + CDOM-dominated, moderately turbid coastal, clear). Any measurements above three standard deviations from the water-type specific mean were reevaluated to ensure they were of high confidence.

Usage Notes

References to method details. The methods used for radiometric measurements and laboratory analyses are identified in the columns 'Measurement_method', 'Chl_method', 'TSS_method', and 'aCDOM_method' in the file *GLORIA_meta_and_lab.csv*. Associated details with references are provided in separate sheets in the file *GLORIA_variables_and_methods.xlsx*. Looking up the method for a particular measurement requires the 'Dataset_ID' and the method name.

Quality flags. Each R_{rs} measurement is associated with quality flags (file *GLORIA_qc_flags.csv*). The quality flags are binary and indicate the presence ('1') or absence ('0') of the quality issue described in Table 2. Missing values indicate that the flag could not be determined because the spectrum did not include the required wavelength range. Some numerical values generated during the quality control are provided in the file *GLORIA_qc_ancillary.csv* (Table 3).

Cross-references to other datasets. Some of the data in GLORIA is part of other data publications, or is also included in the community repositories SeaBASS[33] and/or LIMNADES. The columns 'SeaBASS_ID', 'LIMNADES_ID', and 'LIMNADES_UID' in the data table (*GLORIA_meta_and_lab.csv*) provide identifiers used in the respective datasets to facilitate cross referencing entries, for example for the avoidance of duplicates. Other references to prior publication of the data are provided in the 'Comments' column in *GLORIA_meta_and_lab.csv* in the form of a digital object identifier (DOI).

Code availability

The code to conduct the quality control flagging described in the section *Technical validation* is written in R[95] and available on Zenodo[96]. The 3C code is available at https://gitlab.com/pgroetsch/Rrs_model_3C. The code for QWIP is on Zenodo[97].

Received: 12 July 2022; Accepted: 17 January 2023;
Published online: 16 February 2023

References

1. Seyhan, E. & Dekker, A. Application of remote sensing techniques for water quality monitoring. *Hydrobiological Bulletin* **20**, 41–50, https://doi.org/10.1007/BF02291149 (1986).
2. Kirk, J. T. O. *Light & Photosynthesis in Aquatic Ecosystems*. 2nd edn, (Cambridge University Press, 1994).
3. Muller-Karger, F. E. Remote sensing of marine pollution: A challenge for the 1990s. *Mar. Pollut. Bull.* **25**, 54–60, https://doi.org/10.1016/0025-326X(92)90186-A (1992).
4. Bukata, P. R., Jerome, J. H., Kondratyev, K. Y. & Pozdnyakov, D. *Optical Properties and Remote Sensing of Inland and Coastal Waters*. **362** (CRC Press, 1995).
5. Hassan, R., Scholes, R. & Ash, N. in *Millennium Ecosystem Assessment* (Island Press, Washington, DC., 2005).
6. Health Canada. Guidelines for Canadian Recreational Water Quality, Third Edition. (Water, Air and Climate Change Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, 2012).
7. UNEP. A Framework for Freshwater Ecosystem Management. Volume 4: Scientific Background. (UN Environment, 2018).
8. WHO. Guidelines on recreational water quality. Volume 1: coastal and fresh waters. (World Health Organization, Geneva, 2021).
9. IOCCG. Earth Observations in Support of Global Water Quality Monitoring. (International Ocean Colour Coordinating Group, Dartmouth, Canada, 2018).
10. McClain, C. R. *et al*. Science quality SeaWiFS data for global biosphere research. *Sea Technol.* **39**, 10–16 (1998).
11. Gordon, H. R., Clark, D. K., Mueller, J. L. & Hovis, W. A. Phytoplankton Pigments from the Nimbus-7 Coastal Zone Color Scanner: Comparisons with Surface Measurements. *Science* **210**, 63–66, https://doi.org/10.1126/science.210.4465.63 (1980).
12. Stramski, D., Reynolds, R. A., Kahru, M. & Mitchell, B. G. Estimation of particulate organic carbon in the ocean from satellite remote sensing. *Science* **285**, 239–242 (1999).
13. O'Reilly, J. E. *et al*. Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research: Oceans* **103**, 24937–24953, https://doi.org/10.1029/98JC02160 (1998).
14. IOCCG. Why Ocean Colour? The Societal Benefits of Ocean- Colour Technology. (IOCCG, Dartmouth, Canada, 2008).
15. Werdell, P. J. *et al*. An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing. *Prog. Oceanogr.* **160**, 186–212, https://doi.org/10.1016/j.pocean.2018.01.001 (2018).
16. Esaias, W. E. *et al*. An overview of MODIS capabilities for ocean science observations. *IEEE Transactions on Geoscience and Remote Sensing* **36**, 1250–1265, https://doi.org/10.1109/36.701076 (1998).
17. Hlaing, S. *et al*. Evaluation of the VIIRS ocean color monitoring performance in coastal regions. *Remote Sens. Environ.* **139**, 398–414, https://doi.org/10.1016/j.rse.2013.08.013 (2013).
18. IOCCG. Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex, Waters. (IOCCG, Dartmouth, Canada, 2000).
19. Pahlevan, N. *et al*. ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. *Remote Sens. Environ.* **258**, 112366, https://doi.org/10.1016/j.rse.2021.112366 (2021).
20. Moses, W. J., Sterckx, S., Montes, M. J., De Keukelaere, L. & Knaeps, E. Chapter 3 - Atmospheric Correction for Inland Waters in *Bio-optical Modeling and Remote Sensing of Inland Waters* (eds D. R., Mishra, I., Ogashawara, & A. A., Gitelson) 69–100 (Elsevier, 2017).
21. Pahlevan, N., Chittimalli, S. K., Balasubramanian, S. V. & Vellucci, V. Sentinel-2/Landsat-8 product consistency and implications for monitoring aquatic systems. *Remote Sens. Environ.* **220**, 19–29, https://doi.org/10.1016/j.rse.2018.10.027 (2019).
22. Schaeffer, B. A. *et al*. Barriers to adopting satellite remote sensing for water quality management. *Int. J. Remote Sens.* **34**, 7534–7544, https://doi.org/10.1080/01431161.2013.823524 (2013).
23. Dekker, A. G. & Pinnel, N. (Committee on Earth Observation Satellites (CEOS), Australia, 2018).
24. El Serafy, G. Y. H. *et al*. Integrating Inland and Coastal Water Quality Data for Actionable Knowledge. *Remote Sensing* **13**, 24, https://doi.org/10.3390/rs13152899 (2021).
25. Schaeffer, B. A. *et al*. Mobile device application for monitoring cyanobacteria harmful algal blooms using Sentinel-3 satellite Ocean and Land Colour Instruments. *Environ. Model. Software* **109**, 93–103, https://doi.org/10.1016/j.envsoft.2018.08.015 (2018).
26. Binding, C. E., Pizzolato, L. & Zeng, C. EOLakeWatch; delivering a comprehensive suite of remote sensing algal bloom indices for enhanced monitoring of Canadian eutrophic lakes. *Ecol. Indicators* **121**, 106999, https://doi.org/10.1016/j.ecolind.2020.106999 (2021).
27. Pahlevan, N. *et al*. Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. *Remote Sens. Environ.* **270**, 112860, https://doi.org/10.1016/j.rse.2021.112860 (2022).
28. Smith, B. *et al*. A Chlorophyll-a Algorithm for Landsat-8 Based on Mixture Density Networks. *Frontiers in Remote Sensing* **1**, https://doi.org/10.3389/frsen.2020.623678 (2021).
29. Balasubramanian, S. V. *et al*. Robust algorithm for estimating total suspended solids (TSS) in inland and nearshore coastal waters. *Remote Sens. Environ.* 111768, https://doi.org/10.1016/j.rse.2020.111768 (2020).
30. Pahlevan, N. *et al*. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. *Remote Sens. Environ.* 111604, https://doi.org/10.1016/j.rse.2019.111604 (2020).
31. Smith, B. *et al*. in *International Ocean Colour Science Meeting* (Busan, South Korea, 2019).

32. Jiang, D. *et al*. Remotely estimating total suspended solids concentration in clear to extremely turbid waters using a novel semi-analytical method. *Remote Sens. Environ.* **258**, 112386, https://doi.org/10.1016/j.rse.2021.112386 (2021).
33. Werdell, P. J. *et al*. Unique data repository facilitates ocean color satellite validation. *Eos, Transactions American Geophysical Union* **84**, 377–387 (2003).
34. Mouw, C. B. *et al*. Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. *Remote Sens. Environ.* **160**, 15–30, https://doi.org/10.1016/j.rse.2015.02.001 (2015).
35. Cawse-Nicholson, K. *et al*. NASA's surface biology and geology designated observable: A perspective on surface imaging algorithms. *Remote Sens. Environ.* **257**, 112349, https://doi.org/10.1016/j.rse.2021.112349 (2021).
36. Werdell, P. J. *et al*. The Plankton, Aerosol, Cloud, Ocean Ecosystem Mission: Status, Science, Advances. *Bulletin of the American Meteorological Society* **100**, 1775–1794, https://doi.org/10.1175/bams-d-18-0056.1 (2019).
37. Guanter, L. *et al*. The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. *Remote Sensing* **7**, 8830–8857 (2015).
38. Bresciani, M. *et al*. The use of multisource optical sensors to study phytoplankton spatio-temporal variation in a Shallow Turbid Lake. *Water* **12** https://doi.org/10.3390/w12010284 (2020).
39. Bailey, S. W. & Werdell, P. J. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sens. Environ.* **102**, 12–23, https://doi.org/10.1016/j.rse.2006.01.015 (2006).
40. Goyens, C., Vis, P. D. & Hunt, S. E. in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. 7920–7923.
41. Ruddick, K. G. *et al*. A Review of Protocols for Fiducial Reference Measurements of Water-Leaving Radiance for Validation of Satellite Remote-Sensing Data over Water. *Remote Sensing* **11**, 2198 (2019).
42. Zibordi, G., Voss, K. J., Johnson, B. C. & Mueller, J. Protocols for Satellite Ocean Colour Data Validation: *In Situ* Optical Radiometry. (IOCCG, Dartmouth, NS, Canada, 2019).
43. Lee, Z., Pahlevan, N., Ahn, Y. H., Greb, S. & O'Donnell, D. Robust approach to directly measuring water-leaving radiance in the field. *Appl. Opt.* **52**, 1693–1701, https://doi.org/10.1364/Ao.52.001693 (2013).
44. Mueller, J. L. In-water radiometric profile measurements and data analysis protocols in *Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 4, Volume III: Radiometric Measurements and Data Analysis Protocols* Ch. 2, 7–20 (NASA/TM, 2003).
45. Mobley, C. D. Estimation of the remote-sensing reflectance from above-surface measurements. *Appl. Opt.* **38**, 7442–7455, https://doi.org/10.1364/AO.38.007442 (1999).
46. Voss, K. J. *et al*. A Method to Extrapolate the Diffuse Upwelling Radiance Attenuation Coefficient to the Surface as Applied to the Marine Optical Buoy (MOBY. *Journal of Atmospheric and Oceanic Technology* **34**, 1423–1432, https://doi.org/10.1175/JTECH-D-16-0235.1 (2017).
47. Ruddick, K. G. *et al*. A Review of Protocols for Fiducial Reference Measurements of Downwelling Irradiance for the Validation of Satellite Remote Sensing Data over Water. *Remote Sensing* **11**, 1742 (2019).
48. Hommersom, A. *et al*. Intercomparison in the field between the new WISP-3 and other radiometers (TriOS Ramses, ASD FieldSpec, and TACCS). *Journal of Applied Remote Sensing* **6**, 063615, https://doi.org/10.1117/1.JRS.6.063615 (2012).
49. Bresciani, M. *et al*. Analysis of within- and between-day chlorophyll-a dynamics in Mantua Superior Lake, with a continuous spectroradiometric measurement. *Marine and Freshwater Research* **64**, 303–316, https://doi.org/10.1071/MF12229 (2013).
50. Kudela, R. M. *et al*. Application of hyperspectral remote sensing to cyanobacterial blooms in inland waters. *Remote Sens. Environ.* **167**, 196–205, https://doi.org/10.1016/j.rse.2015.01.025 (2015).
51. Zolfaghari, K. *et al*. Impact of Spectral Resolution on Quantifying Cyanobacteria in Lakes and Reservoirs: A Machine-Learning Assessment. *IEEE Transactions on Geoscience and Remote Sensing* **60**, 1–20, https://doi.org/10.1109/TGRS.2021.3114635 (2022).
52. Chipman, J. W., Olmanson, L. G. & Gitelson, A. A. *Remote sensing methods for lake management: a guide for resource managers and decision-makers*. (North American Lake Management Society, 2009).
53. Gurlin, D., Gitelson, A. A. & Moses, W. J. Remote estimation of chl-a concentration in turbid productive waters — Return to a simple two-band NIR-red model. *Remote Sens. Environ.* **115**, 3479–3490, https://doi.org/10.1016/j.rse.2011.08.011 (2011).
54. Schalles, J. F. & Hladik, C. M. Mapping phytoplankton chlorophyll in turbid, Case 2 estuarine and coastal waters. *Isr. J. Plant Sci.* **60**, 169–191, https://doi.org/10.1560/IJPS.60.1-2.169 (2012).
55. Li, L. *et al*. An inversion model for deriving inherent optical properties of inland waters: Establishment, validation and application. *Remote Sens. Environ.* **135**, 150–166, https://doi.org/10.1016/j.rse.2013.03.031 (2013).
56. Mishra, S., Mishra, D. R., Lee, Z. & Tucker, C. S. Quantifying cyanobacterial phycocyanin concentration in turbid productive waters: A quasi-analytical approach. *Remote Sens. Environ.* **133**, 141–151, https://doi.org/10.1016/j.rse.2013.02.004 (2013).
57. Brezonik, P. L., Olmanson, L. G., Finlay, J. C. & Bauer, M. E. Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. *Remote Sens. Environ.* **157**, 199–215, https://doi.org/10.1016/j.rse.2014.04.033 (2015).
58. Werther, M. *et al*. Characterising retrieval uncertainty of chlorophyll-a algorithms in oligotrophic and mesotrophic lakes and reservoirs. *ISPRS Journal of Photogrammetry and Remote Sensing* **190**, 279–300, https://doi.org/10.1016/j.isprsjprs.2022.06.015 (2022).
59. Wei, J., Lee, Z. & Shang, S. A system to measure the data quality of spectral remote sensing reflectance of aquatic environments. *Journal of Geophysical Research: Oceans* **121**, 8189–8207, https://doi.org/10.1002/2016JC012126 (2016).
60. Simis, S. G. H. & Olsson, J. Unattended processing of shipborne hyperspectral reflectance measurements. *Remote Sens. Environ.* **135**, 202–212, https://doi.org/10.1016/j.rse.2013.04.001 (2013).
61. Qin, P., Simis, S. G. H. & Tilstone, G. H. Radiometric validation of atmospheric correction for MERIS in the Baltic Sea based on continuous observations from ships and AERONET-OC. *Remote Sens. Environ.* **200**, 263–280, https://doi.org/10.1016/j.rse.2017.08.024 (2017).
62. Warren, M. A. *et al*. Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. *Remote Sens. Environ.* **225**, 267–289, https://doi.org/10.1016/j.rse.2019.03.018 (2019).
63. Mobley, C. D. Polarized reflectance and transmittance properties of windblown sea surfaces. *Appl. Opt.* **54**, 4828–4849, https://doi.org/10.1364/AO.54.004828 (2015).
64. Maciel, D. *et al*. Retrieving Total and Inorganic Suspended Sediments in Amazon Floodplain Lakes: A Multisensor Approach. *Remote Sensing* **11**, 1744 (2019).
65. Cairo, C. *et al*. Hybrid Chlorophyll-a Algorithm for Assessing Trophic States of a Tropical Brazilian Reservoir Based on MSI/Sentinel-2 Data. *Remote Sensing* **12**, 40, https://doi.org/10.3390/rs12010040 (2020).
66. da Silva, M. P., Sander de Carvalho, L. A., Novo, E., Jorge, D. S. F. & Barbosa, C. C. F. Use of optical absorption indices to assess seasonal variability of dissolved organic matter in Amazon floodplain lakes. *Biogeosciences* **17**, 5355–5364, https://doi.org/10.5194/bg-17-5355-2020 (2020).
67. Lubac, B. & Loisel, H. Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea. *Remote Sens. Environ.* **110**, 45–58, https://doi.org/10.1016/j.rse.2007.02.012 (2007).
68. Binding, C. E., Zastepa, A. & Zeng, C. The impact of phytoplankton community composition on optical properties and satellite observations of the 2017 western Lake Erie algal bloom. *J. Great Lakes Res.* **45**, 573–586, https://doi.org/10.1016/j.jglr.2018.11.015 (2019).
69. Fritz, C., Dörnhöfer, K., Schneider, T., Geist, J. & Oppelt, N. Mapping Submerged Aquatic Vegetation Using RapidEye Satellite Data: The Example of Lake Kummerow (Germany). *Water* **9**, 510 (2017).

70. Zibordi, G. & Talone, M. On the equivalence of near-surface methods to determine the water-leaving radiance. *Optics Express* **28**, 3200–3214, https://doi.org/10.1364/OE.28.003200 (2020).
71. Kutser, T., Vahtmae, E., Paavel, B. & Kauer, T. Removing glint effects from field radiometry data measured in optically complex coastal and inland waters. *Remote Sens. Environ.* **133**, 85–89, https://doi.org/10.1016/j.rse.2013.02.011 (2013).
72. Minaudo, C. *et al.* The Imprint of Primary Production on High-Frequency Profiles of Lake Optical Properties. *Environ. Sci. Technol.* **55**, 14234–14244, https://doi.org/10.1021/acs.est.1c02585 (2021).
73. Stramski, D. *et al.* Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans. *Biogeosciences* **5**, 171–201, https://doi.org/10.5194/bg-5-171-2008 (2008).
74. Bracher, A. *et al.* Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. *Ocean Sci.* **11**, 139–158, https://doi.org/10.5194/os-11-139-2015 (2015).
75. Tilstone, G. *et al.* Field Intercomparison of Radiometer Measurements for Ocean Colour Validation. *Remote Sensing* **12**, 1587, https://doi.org/10.3390/rs12101587 (2020).
76. Dekker, A. G. *Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing* PhD thesis, Vrije Universiteit Amsterdam, (1993).
77. GEBCO Bathymetric Compilation Group 2021. GEBCO 2021 Grid. *NASA National Snow and Ice Data Center DAAC* https://doi.org/10.5285/c6612cbe-50b3-0cff-e053-6c86abc09f8f (2021).
78. Yentsch, C. & Menzel, D. A method for the determination of phytoplankton chlorophyll and phaeophytin by fluoroescence. *Deep Sea Research* **10**, 221–231 (1963).
79. Welschmeyer, N. A. Fluorometric analysis of chlorophyll-a in the presence of chlorophyll-b and pheopigments. *Limnol. Oceanogr.* **39**, 1985–1992 (1994).
80. Roesler, C. S. & Barnard, A. H. Optical proxy for phytoplankton biomass in the absence of photophysiology: Rethinking the absorption line height. *Methods in Oceanography* **7**, 79–94, https://doi.org/10.1016/j.mio.2013.12.003 (2013).
81. Mitchell, B. G., Kahru, M., Wieland, J. & Stramska, M. Determination of spectral absorption coefficients of particles, dissolved material and phytoplankton for discrete water samples - Chapter 4. 39–64 (NASA Goddard Space Flight Center, Greenbelt, MD, 2003).
82. Tyler, J. E. The Secchi disc. *Limnol. Oceanogr.* **13**, 1–6, https://doi.org/10.4319/lo.1968.13.1.0001 (1968).
83. Wernand, M. R. On the history of the Secchi disc. *Journal of the European Optical Society-Rapid Publications* **5**, https://doi.org/10.2971/jeos.2010.10013s (2010).
84. Lehmann, MK. *et al.* GLORIA - A global dataset of remote sensing reflectance and water quality from inland and coastal waters, *PANGAEA*, https://doi.org/10.1594/PANGAEA.948492 (2022).
85. Spyrakos, E. *et al.* Optical types of inland and coastal waters. *Limnol. Oceanogr.*, https://doi.org/10.1002/lno.10674 (2018).
86. Eleveld, A. M. *et al.* An Optical Classification Tool for Global Lake Waters. *Remote Sensing* **9**, https://doi.org/10.3390/rs9050420 (2017).
87. Lehmann, M. K., Nguyen, U., Allan, M. & van der Woerd, H. Colour Classification of 1486 Lakes across a Wide Range of Optical Water Types. *Remote Sensing* **10**, https://doi.org/10.3390/rs10081273 (2018).
88. Hou, X. *et al.* Global mapping reveals increase in lacustrine algal blooms over the past decade. *Nature Geoscience* **15**, 130–134, https://doi.org/10.1038/s41561-021-00887-x (2022).
89. Dierssen, H. M. *et al.* QWIP: A Quantitative Metric for Quality Control of Aquatic Reflectance Spectral Shape Using the Apparent Visible Wavelength. *Frontiers in Remote Sensing* **3**, https://doi.org/10.3389/frsen.2022.869611 (2022).
90. Vandermeulen, R. A., Mannino, A., Craig, S. E. & Werdell, P. J. 150 shades of green: Using the full spectrum of remote sensing reflectance to elucidate color shifts in the ocean. *Remote Sens. Environ.* **247**, 111900, https://doi.org/10.1016/j.rse.2020.111900 (2020).
91. Mobley, C. D. *Light and Water: Radiative Transfer in Natural Waters*. 608 (Academic Press, 1994).
92. Groetsch, P. M. M., Gege, P., Simis, S. G. H., Eleveld, M. A. & Peters, S. W. M. Validation of a spectral correction procedure for sun and sky reflections in above-water reflectance measurements. *Optics Express* **25**, A742–A761, https://doi.org/10.1364/OE.25.00A742 (2017).
93. Jordan, T. M., Simis, S. G. H., Grötsch, P. M. M. & Wood, J. Incorporating a Hyperspectral Direct-Diffuse Pyranometer in an Above-Water Reflectance Algorithm. *Remote Sensing* **14** (2022).
94. Burggraaff, O. Biases from incorrect reflectance convolution. *Optics Express* **28**, 13801–13816, https://doi.org/10.1364/OE.391470 (2020).
95. R Core Team. *R: A language and environment for computing*. (R Foundation for Statistical Computing, 2022).
96. Maciel, D. R code for GLORIA quality control flags, *Zenodo*, https://doi.org/10.5281/zenodo.7372445 (2022).
97. Vandermeulen, RA. QWIP: v1.1 (v1.1), *Zenodo*, https://doi.org/10.5281/zenodo.7373840 (2022).
98. Wyszecki, G. & Stiles, W. S. *Color science: concepts and methods, quantitative data, and formulae*. (John Wiley & Sons, 2000).
99. CIE. *Commission Internationale de l'Éclairage proceedings, 1931*. (Cambridge University Press, 1932).

Acknowledgements

Making measurements and taking samples in inland and coastal waters has considerable challenges. These include overland travel to and from water bodies, launching boats across natural shorelines, and processing samples at the end of long days. Therefore, this data is the result of the hard work by many more people than can practically be acknowledged individually here, and we sincerely thank field personnel, lab technicians, students, captains and skippers, and all other supporting personnel without whom this dataset would not have been possible. In some cases, specific acknowledgements of people and/or funding can be found in the 'Comments' column in the file *GLORIA_meta_and_lab.csv*.

Funding sources include: Estonian Ministry of Education and Research; European Commission FP7, H2020, FP7-ENV-2007-1-226224; Estonian Research Council; Helmholtz Infrastructure Initiative FRAM; BMBF 03G0218A; New Zealand Ministry for Business, Innovation & Employment grants UOWX1503, UOWX1802, KENTR1601, NASA ROSES grants 80HQTR19C0015, 80NSSC 21K0499, 80NSSC22K1389, and USGS Landsat Science Team Award 140G0118C0011, Vietnam National Foundation for Science and Technology Development (NAFOSTED), grant number 105.08-2019.329, Federal Ministry for Economic Affairs and Energy, Germany, Award: LAKESAT 50EE1340, EnMAP CalVal 50EE1923, TypSynSat 50EE1915.

Author contributions

Moritz K. Lehmann, Daniela Gurlin and Nima Pahlevan contributed equally to this work; M.K. Lehmann wrote the manuscript, contributed to data collation, managed the data files, and conducted quality control of the radiometric and laboratory data; D. Gurlin edited the manuscript, contributed to data collation, checked and validated all measurement methodologies, and supported quality control of the radiometric and laboratory

data; N. Pahlevan conceived the effort, invited each data contributor, led preliminary studies based on subsets of GLORIA, and supported manuscript writing, data collation, and communication. The following authors contributed to quality control of the entire dataset (alphabetical by first name): Andrea Vander Woude, Astrid Bracher, Caren Binding, Claudia Giardino, Dalin Jiang, Daniel A. Maciel, Hendrik J. van der Woerd, Jeremy A. Kravitz, Lin Li, Mortimer Werther, Nathan Drayson, Ryan A. Vandermeulen, Sachidananda Mishra, Salem I. Salem, Stefan G.H. Simis, Thomas Jordan and Zhigang Cao. Brandon Smith and Sundarabalan V. Balasubramanian extracted and readied the SeaBASS data. Authors not listed by name in this section made significant contributions to data collection and sample processing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.K.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Moritz K. Lehmann[1,2]✉, Daniela Gurlin[3], Nima Pahlevan[4,5], Krista Alikas[6], Janet Anstee[7], Sundarabalan V. Balasubramanian[8], Cláudio C. F. Barbosa[9], Caren Binding[10], Astrid Bracher[11,12], Mariano Bresciani[13], Ashley Burtner[14], Zhigang Cao[15], Arnold G. Dekker[16], Courtney Di Vittorio[17], Nathan Drayson[7], Reagan M. Errera[18], Virginia Fernandez[19], Dariusz Ficek[20], Cédric G. Fichot[21], Peter Gege[22], Claudia Giardino[13], Anatoly A. Gitelson[23], Steven R. Greb[24], Hayden Henderson[25], Hiroto Higa[26], Abolfazl Irani Rahaghi[27], Cédric Jamet[28], Dalin Jiang[29], Thomas Jordan[30], Kersti Kangro[6], Jeremy A. Kravitz[31], Arne S. Kristoffersen[32], Raphael Kudela[33], Lin Li[34], Martin Ligi[6], Hubert Loisel[28], Steven Lohrenz[35], Ronghua Ma[15], Daniel A. Maciel[9], Tim J. Malthus[36], Bunkei Matsushita[37], Mark Matthews[38], Camille Minaudo[39], Deepak R. Mishra[40], Sachidananda Mishra[41], Tim Moore[42], Wesley J. Moses[43], Hà Nguyễn[44], Evlyn M. L. M. Novo[9], Stéfani Novoa[45], Daniel Odermatt[27], David M. O'Donnell[46], Leif G. Olmanson[47], Michael Ondrusek[48], Natascha Oppelt[49], Sylvain Ouillon[50,51], Waterloo Pereira Filho[52], Stefan Plattner[22], Antonio Ruiz Verdú[53], Salem I. Salem[54], John F. Schalles[55], Stefan G. H. Simis[30], Eko Siswanto[56], Brandon Smith[4,5], Ian Somlai-Schweiger[22], Mariana A. Soppa[11], Evangelos Spyrakos[29], Elinor Tessin[32], Hendrik J. van der Woerd[57], Andrea Vander Woude[18], Ryan A. Vandermeulen[4,58], Vincent Vantrepotte[28], Marcel R. Wernand[45,60], Mortimer Werther[27,29], Kyana Young[17] & Linwei Yue[59]

[1]Xerra Earth Observation Institute, PO Box 400, Alexandra, 9340, New Zealand. [2]School of Science, University of Waikato, Private Bag 3105, Hamilton, 3240, New Zealand. [3]Wisconsin Department of Natural Resources, Bureau of Water Quality, 101 S Webster Street, Madison, WI, 53707, USA. [4]Science Systems and Applications, Inc. (SSAI), Lanham, MD, USA. [5]NASA Goddard Space Flight Center, Greenbelt, MD, USA. [6]Tartu Observatory of the University of Tartu, Tartumaa, 61602, Estonia. [7]Coasts and Oceans Systems Program (COS), CSIRO Environment Business Unit, Acton, ACT, 2601, Australia. [8]GeoSensing and Imaging Consultancy, Trivandrum, Kerala, India. [9]Instrumentation Lab for Aquatic Systems (LabISA), National Institute for Space Research (INPE), São José dos Campos, Brazil. [10]Environment and Climate Change Canada, Burlington, ON, Canada. [11]Phytooptics Group, Physical Oceanography of Polar Seas, Climate Sciences, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany. [12]Department of Physics and Electrical Engineering, Institute of Environmental Physics, University of Bremen, Bremen, Germany. [13]National Research Council of Italy, Institute for Electromagnetic Sensing of the Environment, CNR-IREA, Milano, Italy. [14]Cooperative Institute for Great Lakes Research, University of Michigan, 4840 South State Road, Ann Arbor, MI, 48108, USA. [15]Nanjing Institute of Geography and Limnology,

Chinese Academy of Sciences, Nanjing, 210008, China. [16]SatDek Pty Ltd, 99 Read Rd, Sutton, NSW, 2620, Australia. [17]Wake Forest University, Engineering, 455 Vine Street, Winston-Salem, NC, 27101, USA. [18]NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, MI, USA. [19]Department of Geography, Universidad de la República, Montevideo, Uruguay. [20]Institute of Biology and Earth Sciences, Pomeranian University, Arciszewskiego 22, 76-200, Slupsk, Poland. [21]Department of Earth and Environment, Boston University, Boston, MA, USA. [22]German Aerospace Center (DLR), Remote Sensing Technology Institute, Wessling, Germany. [23]University of Nebraska-Lincoln, School of Natural Resources, 3310 Holdrege Street, Lincoln, NE, 68503, USA. [24]University of Wisconsin-Madison, Aquatic Sciences Center, 1975 Willow Drive, Madison, WI, 53706, USA. [25]Michigan Technological University, Great Lakes Research Center, 100 Phoenix Drive, Houghton, MI, 49931, USA. [26]Faculty of Urban Innovation, Yokohama National University, Tokiwadai 79-5, Hodogaya, Yokohama, Kanagawa, Japan. [27]Swiss Federal Institute of Aquatic Science and Technology, Department of Surface Waters – Research and Management, Dübendorf, Switzerland. [28]Université du Littoral Côte d'Opale, CNRS, Univ. Lille, IRD, UMR 8187 - LOG - Laboratoire d'Océanologie et de Géosciences, F-62930, Wimereux, France. [29]Earth and Planetary Observation Sciences (EPOS), Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling, UK. [30]Plymouth Marine Laboratory, Plymouth, PL1 3DH, UK. [31]NASA Ames Research Center, Moffett Field, CA, USA. [32]Department of Physics and Technology, University of Bergen, Bergen, Norway. [33]University of California-Santa Cruz, Ocean Sciences Department, Institute of Marine Sciences, 1156 High Street, Santa Cruz, CA, 95064, USA. [34]Department of Earth Sciences, Indiana University-Purdue University, Indianapolis, IN, USA. [35]University of Massachusetts-Dartmouth, School for Marine Science and Technology West, 706 South Rodney French Blvd., New Bedford, MA, 02744, USA. [36]Coasts and Oceans Systems Program (COS), CSIRO Environment Business Unit, Ecosciences Precinct, 41 Boggo Road, Dutton Park, QLD, 4102, Australia. [37]Faculty of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan. [38]CyanoLakes (Pty) Ltd, Sydney, 2126, Australia. [39]Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Spain. [40]Department of Geography, University of Georgia, Athens, GA, 30602, USA. [41]National Centers for Coastal Ocean Science, National Oceanic and Atmospheric Administration, 1305 East-West Hwy, Silver Spring, MD, 20910, USA. [42]Harbor Branch Oceanographic Institute, Florida Atlantic University, Fort Pierce, FL, USA. [43]U.S. Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC, 20375, USA. [44]Faculty of Geology, VNU University of Science, Ha Noi, Vietnam. [45]Royal Netherlands Institute for Sea Research, Physical Oceanography, Marine Optics & Remote Sensing, Den Burg, Texel, Netherlands. [46]Upstate Freshwater Institute, Syracuse, NY, USA. [47]Department of Forest Resources, University of Minnesota, St. Paul, MN, USA. [48]NOAA Center for Satellite Applications and Research, College Park, MD, USA. [49]Earth Observation and Modelling, Kiel University, Department of Geography, 24118, Kiel, Germany. [50]UMR LEGOS, University of Toulouse, IRD, CNES, CNRS, UPS, 14 Avenue Edouard Belin, 31400, Toulouse, France. [51]Department Water-Environment-Oceanography, University of Science and Technology of Hanoi (USTH), Vietnamese Academy of Science and Technology (VAST), 18 Hoang Quoc Viet, Hanoi, 100000, Vietnam. [52]Department of Geosciences, Federal University of Santa Maria, Av. Roraima, 1000, 97105-900, Santa Maria, Rio Grande do Sul, Brazil. [53]Laboratory for Earth Observation, University of Valencia, Catedrático Agustín Escardino 9, Paterna (Valencia), 46980, Spain. [54]Faculty of Engineering, Kyoto University of Advanced Science (KUAS), 18 Yamanouchi Gotanda, Ukyo, Kyoto, Japan. [55]Creighton University, Department of Biology, Omaha, NE, 68178, USA. [56]Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Showa-machi 3173-25, Yokohama, Kanagawa, 2360001, Japan. [57]Department of Water & Climate Risk, Institute for Environmental Studies (IVM), Vrije Universiteit, Amsterdam, Netherlands. [58]Ocean Ecology Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA. [59]China University of Geosciences, School of Geography and Information Engineering, Wuhan, China. [60]Deceased: Marcel R. Wernand. ✉e-mail: moritz.lehmann@gmail.com