Water Research X 28 (2025) 100386

Contents lists available at ScienceDirect

Water Research X

journal homepage: www.sciencedirect.com/journal/water-research-x

Improving algal bloom modelling in eutrophic lakes by calibrating the General Lake Model with satellite remote sensing products

Maud A.C. Siebers [a,*], Mortimer Werther [b], Daniel Odermatt [b,c], Eleanor Mackay [d], Linda May [e], Thomas Shatwell [f,g], Ian Jones [a], Matthew Blake [a], Peter D. Hunter [a,h]

[a] Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling FK9 4LA, Stirling, UK
[b] Swiss Federal Institute of Aquatic Science and Technology, Department of Surface Waters - Research and Management, Duebendorf, Switzerland
[c] Department of Geography, University of Zurich 8057 Zurich, Switzerland
[d] Centre for Ecology & Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster LA1 4AP, Lancashire, UK
[e] Centre for Ecology & Hydrology, Bush Estate, Penicuik EH26 0QB, UK
[f] Helmholtz Centre for Environmental Research - UFZ, Department of Lake Research 39114 Magdeburg, Germany
[g] Ostwestfalen-Lippe University of Applied Sciences and Arts, Department of Environmental Engineering and Applied Computer Science 37671 Höxter, Germany
[h] Scotland's International Environment Centre, Faculty of Natural Sciences, University of Stirling, Stirling, UK

ARTICLE INFO

Keywords:
Algal bloom forecasting
Lake modelling calibration
Earth observation
Conformal prediction

ABSTRACT

Accurate forecasting of algal blooms in lakes can support effective freshwater management. However, observational datasets for calibrating and validating algal bloom forecasting models such as the General Lake Model - Aquatic Eco Dynamics (GLM-AED) are often scarce, which impedes robust model calibration and forecasting ability. Satellite remote sensing can help fill these gaps by offering high-frequency, large-scale measurements of phytoplankton chlorophyll-*a* concentration (mg m^{-3}), but satellite chl-*a* products often carry high uncertainty. Here we introduce a novel approach to quantify uncertainty in satellite chl-*a* based on conformal prediction, with the aim of integrating robust chlorophyll-*a* products into GLM-AED. Using Sentinel-2 imagery from two eutrophic lakes in the UK, Esthwaite Water and Loch Leven, we obtain remotely sensed chlorophyll-*a* with low systematic signed percentage bias (-1.22 % and 0.38) and moderate median symmetric accuracy (15.87 and 43.02 %) using Polymer atmospheric correction. We effectively flag potentially uncertain chlorophyll-*a* estimates (coverage factor: 75.6 - 81 %). Integrating the screened remotely sensed chlorophyll-*a* estimates improved GLM-AED algal bloom forecasts by 50 % in Loch Leven and 13 % in Esthwaite Water, with the greater improvement in Loch Leven attributed to its higher initial model errors. In contrast, incorporating unscreened chlorophyll-*a* estimates into GLM-AED increases validation errors on average by 32 %.

Our findings show that process-based model predictions can substantially benefit from incorporating additional satellite-derived chlorophyll-a estimates. At the same time, they highlight a crucial need for robust uncertainty quantification to support downstream applications such as algorithm validation, biological monitoring in data-scarce regions, and water management decision-making.

Moreover, because conformal prediction is model-*a*gnostic and satellite-derived chlorophyll-*a* products are globally accessible, our study paves the way for large-scale, well-calibrated bloom forecasting through process-based models.

Abbreviations: ACOLITE, Atmospheric Correction for OLI 'lite'; Chl-a, Chlorophyll-*a*; CP, Conformal Prediction; CV, Coefficient of Variation; C2RCC, Case 2 Regional CoastColour (atmospheric correction); EO, Earth Observation; GLM-AED, General Lake Mode - AquaticEcodynamics; MAE, Mean Absolute Error; MdSA, Median Symmetric Accuracy; MSI, MultiSpectral Instrument; NC, Non-conformity; OWT, Optical Water Type; Polymer, POLYnomial-based approach of MERIS data (atmospheric correction); SSPB, Symmetric Signed Percentage Bias; S2, Sentinel-2.

* Corresponding author.
E-mail address: maud.siebers@stir.ac.uk (M.A.C. Siebers).

https://doi.org/10.1016/j.wroa.2025.100386
Received 3 April 2025; Received in revised form 3 July 2025; Accepted 23 July 2025
Available online 23 July 2025
2589-9147/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Algal blooms, defined as high accumulation of phytoplankton or cyanobacteria in aquatic systems, have become a pressing issue in lakes and reservoirs worldwide (Beaulieu et al., 2013; Feng et al., 2024; Paerl & Huisman, 2008). Blooms can appear swiftly under favourable conditions driven by warm temperatures (Jöhnk et al., 2008; Paerl & Huisman, 2008), abundant nutrients (Beaulieu et al., 2013) and stable water column conditions, such as thermal stratification (Bormans et al., 2005). These complex and interconnected drivers highlight the dynamic and often unpredictable nature of algal blooms.

Process-based models, like the General Lake Model (GLM) coupled with the Aquatic Eco Dynamics model (AED) (Hipsey, 2022; Hipsey et al., 2019), have the capabilities to resolve algal bloom dynamics and their temporal heterogeneity. GLM-AED can simulate temperature (Bruce et al., 2018; Bueche et al., 2017; Mesman et al., 2020), dissolved oxygen, total phosphorus, nitrate, ammonium and a proxy for algal biomass, phytoplankton chlorophyll-*a* concentration (chl-*a*) (Fadel et al., 2019; Fenocchi et al., 2019; Krinos et al., 2019; Soares & Calijuri, 2021; Ward et al., 2020). While GLM-AED demonstrates the potential to simulate algal blooms, its effectiveness remains contingent on the availability of high-quality observational datasets.

Meteorological and hydrological data, such as air temperature, wind speed, solar radiation, inflow rates, and nutrient concentrations, are essential for driving the GLM-AED model and can often be obtained from models or in situ sensors at relatively low cost. In contrast, chl-*a* measurements, which are used for calibrating or validating phytoplankton dynamics within the model, are expensive and labour-intensive to collect, as they rely on laboratory analysis. This leads to limited spatial and temporal coverage of consistent chl-a observations across lakes.

This scarcity of in situ chl-*a* measurements limits our ability to model algal dynamics, including forecasting of blooms through process-based models and widens the gap between advances in hydrodynamic and ecological lake modelling. To close the observation gap between physical and biological variables, satellite remote sensing of lakes could provide direct estimates of water constituent concentrations such as chl-*a* (Bukata et al., 1974; Morel, 1980). In particular, the Sentinel-2 (S2) MultiSpectral Instrument (MSI) constellation provides high spatial resolution imagery (10 – 60 m) with a short revisit cycle of 2-3 days, with the A,B and C satellites in orbit (10-day nominal global coverage).

However, the estimation of chl-*a* through S2 MSI is subject to various sources of uncertainty including the necessary correction for atmospheric disturbances (Warren et al., 2019), adjacency effects of the surrounding environment (Santer & Schmechtig, 2000), complex water constituent compositions (Werther, et al., 2022a) and errors in observational processing schemes (Burggraaff, 2020). Algorithmic approaches for chl-*a* retrieval are subject to model-specific uncertainties, and only a handful of methods explicitly quantify them (Werther et al., 2022b). As multiple sources of uncertainty are inherent within aquatic remote sensing, accurately quantifying uncertainty in satellite-derived products for downstream uses, such as an input to GLM-AED, is pivotal to reliable modelling (Werther & Burggraaff, 2023).

To enable the use of satellite-derived chl-*a* products for improved GLM-AED calibration and validation, we introduce an approach rooted in conformal prediction (CP) (Vovk et al., 2005). Using a separate baseline dataset, CP evaluates how closely new chl-*a* estimates conform to established patterns. CP not only produces well-calibrated, spatially explicit uncertainty estimates but also offers a data-driven threshold for filtering unreliable observations Although CP is well established in applied machine learning and terrestrial remote-sensing tasks, such as estimating canopy height and identifying tree species (Fontana et al., 2023; Papadopoulos, 2008; Singh et al., 2024), its application in aquatic remote sensing is only indirect (Waczak et al., 2024). Here, we apply CP to a range of S2 MSI chl-*a* retrieval approaches and illustrate how its model-agnostic design allows seamless integration with these approaches.

We hypothesise that incorporating MSI-derived chl-*a* estimates into the GLM-AED model will enhance both its calibration and estimation accuracy relative to using only in situ chl-*a* measurements. This assumption is driven by evidence from prior research suggesting that satellite-based observations can refine ecosystem models (Deng et al., 2024; Hedger et al., 2002). We also hypothesise that filtering MSI chl-*a* estimates through CP-derived uncertainty thresholds will result in greater levels of improvement than relying on unfiltered estimates, given that discarding highly uncertain observations may prevent error propagation.

Accordingly, this study addresses the following questions:

(1) To what extent can chlorophyll-a be accurately estimated from MSI data over two eutrophic lakes for integration into the GLM-AED model?
(2) How effective is the conformal prediction (CP) method in filtering MSI-derived chl-*a* estimates?
(3) Does incorporating these filtered chl-*a* estimates into the calibration of GLM-AED improve the accuracy of algal bloom forecasting and detection?

2. Methods

2.1. In situ data

2.1.1. Study lakes

Two freshwater lakes in the United Kingdom (UK) were used as case study sites (Fig. 1). Esthwaite Water is a small (surface area of 0.96 km^2), shallow (mean depth 6.9 m), eutrophic lake located in the English Lake District (Mackay et al., 2012, 2014; Mortimer, 1942). Loch Leven is a comparably larger (surface area13.3 km^2) and shallower (mean depth 3.9 m) eutrophic lake in lowland Scotland (Carvalho et al., 2012). Both lakes have a long history of algal blooms consisting of cyanobacterial and diatom species (Maberly et al., 2011; May & Spears, 2012) with mean annual chl-*a* values of 18.5 mg m^{-3} in Esthwaite Water and 25.1 mg m^{-3} in Loch Leven (Carvalho et al., 2012; Maberly et al., 2011; Madgwick et al., 2006; Spears et al., 2012). In addition, both lakes have been studied using Earth observation (Hedger et al., 2002; Hunter et al., 2010; Liu et al., 2021).

2.1.2. Chlorophyll-a concentration

Fortnightly in situ chl-*a* for Esthwaite Water (n = 156) (Feuchtmayr et al., 2021, 2024, 2025) and Loch Leven (n = 145) (Dudley et al., 2013; Taylor et al., 2021, 2022a, 2022b, 2023) were extracted for the period January 2016 to December 2022 from an ongoing long-term monitoring programme by the UK Centre for Ecology & Hydrology (UKCEH) (see sampling points indicated in Fig. 1). Samples were unavailable or could not be collected in June and July 2019 and March to July 2020 due to the Covid-19 pandemic. For Loch Leven, chl-*a* data were only available until November 2021.

For both lakes, chl-*a* was determined by filtering a known volume of water from a 0-5 m integrated sample onto a Whatman GF/C filter on the day of sampling with the filter then frozen for later analysis. Filters were defrosted and extracted in a faster hot methanol method for Esthwaite Water and a slower cold methanol for Loch Leven, which reduces thermal degradation. These were measured spectrophotometrically with absorbance values converted to chl-*a* following the protocols of Talling (1974) for Esthwaite Water by and Rice et al. (2017) for Loch Leven.

2.1.3. GLM-AED driving data

Latitude, longitude, depth, area and elevation for each lake were provided by UKCEH (Mackay et al., 2014; Spears et al., 2012). Hourly shortwave radiation, longwave radiation, air temperature, relative humidity, windspeed, rain and snow were obtained from ERA 5 simulations for 2016 to 2022 for the respective pixel (latitude-longitude 0.25 °x 0.25 ° grid) in which the lake coordinate falls (Hersbach et al., 2018).

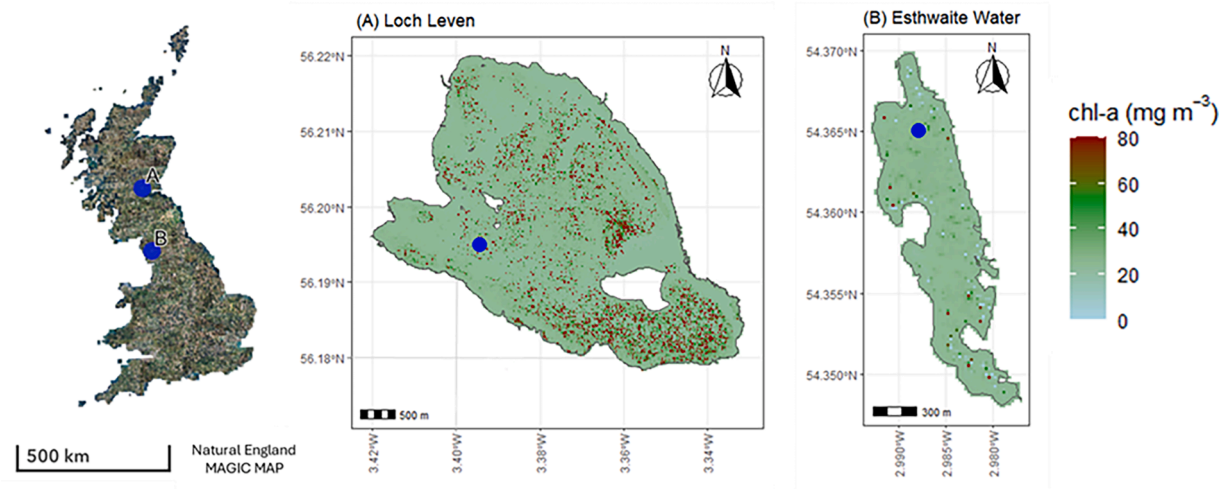

Fig. 1. MSI chl-*a* maps of Loch Leven (A) and Esthwaite Water (B) and their respective locations in the UK, using Polymer atmospheric correction and lake-optimised Gurlin et al. (2011) on 2016-06-09 (A) and 2016-07-19 (B). The blue points indicate in situ sampling sites.

For the same period, daily lake inflow or outflow data were obtained from the UK National River Flow Archive for the main inflowing rivers (Esthwaite Water: Black Beck; Loch Leven: North Queich, South Queich and Greens Burn) or the outflow. In both lakes, outflow was assumed to be equal to the inflow. Fortnightly water temperature and chemistry in situ measurements (pH, phosphorus, nitrate, ammonia and silica) were provided by UKCEH and linearly interpolated to provide daily resolution.

2.2. Satellite data processing

S2 MSI images were atmospherically corrected to derive surface water remote-sensing reflectance (R_{rs} (λ)), followed by a valid pixel identification to exclude non-water pixels using IdePix (Identification of Pixel features) (Skakun et al., 2022). We then assessed the spatial homogeneity of the water surface area around the fortnightly sampling locations in Esthwaite Water and Loch Leven. We used thirteen inland water Optical Water Types (OWTs) defined in Spyrakos et al. (2017), representing clusters of water bodies with similar optical characteristics, to inform chl-*a* retrieval algorithm selection for each lake (Atwood et al., 2024; Eleveld et al., 2017; Liu et al., 2021; Neil et al., 2019). Finally, we quantified the uncertainties associated with MSI chl-*a* estimates through CP, resulting in filtered MSI chl-*a* to calibrate the GLM-AED model.

2.2.1. Atmospheric correction and valid pixel identification

S2 MSI tiles (30UVF for Esthwaite Water and 30VVH for Loch Leven) were processed using three different atmospheric correction algorithms: Atmospheric Correction for OLI 'lite' (ACOLITE) (Vanhellemont & Ruddick, 2016), POLYnomial-based approach of MERIS data (Polymer) (Ramon et al., 2011) and Case 2 Regional CoastColour atmospheric correction (C2RCC) (Brockmann et al., 2016) for the period of 2016 and 2024. A 3 × 3 MSI pixel grid around the respective in situ coordinate was extracted for match-up generation with the fortnightly in situ sampling following the match-up protocol of Concha et al. (2021). We excluded pixels flagged as invalid due to cloud (including ambiguous, sure, and a buffer) or cloud shadow, snow/ice, bright, coastline land, white and glint risk.

2.2.2. Spatial homogeneity test and S2 matchups

A spatial homogeneity test was conducted to identify MSI scenes affected by high spatial R_{rs} variability within 3 × 3 kernels of valid pixels (Zibordi et al., 2018). Spatial coefficient of variation (CV), defined as the ratio of the standard deviation to the mean, was calculated for the 560 nm band (Zibordi et al., 2018). Pixel kernels with a CV greater than 0.25 (25 %) were deemed to have high spatial variability and removed from the dataset.

2.2.3. Remote sensing algorithm calibration and validation

The resulting MSI scenes were matched with in situ chl-*a* within a ± 3-day range of the in situ sampling date following previous studies (Ansper & Alikas, 2018; Riddick et al., 2019). A 3-day window resulted in significantly more match ups (±50 matchups) than a 1 (±22) or 2 (±34) day window. For all valid R_{rs} (λ) an OWT membership score was calculated by estimating of the spectral angle (Kruse & Lefkoff, 1994; Liu et al., 2021) and the OWT with the highest similarity score was assigned (Spyrakos et al., 2017), resulting in one OWT (Table 1) per pixel. The OWTs were assigned to one of three clusters based on common cluster characteristics (Table 1; Neil et al., 2019). Cluster A represents the most transparent water (in terms of biomass and turbidity), B represents

Table 1
Optical Water Types (OWT), characteristics, trophic status, and clustering. The table summarises the characteristics and trophic status of different OWTs based on their reflectance features and associated water quality properties. Each OWT is categorised into a trophic status (hypereutrophic, meso-eutrophic, or oligotrophic) and assigned to a cluster (A, B, or C) based on that trophic status. A is Oligotrophic, B is Meso-eutrophic and C is Hypereutrophic.

OWT	Dominant characteristics	Trophic status	Cluster
1	Hypereutrophic waters with scum of cyanobacterial bloom and vegetation like R_{rs}.	Hypereutrophic	B
2	Diverse reflectance shape, marginal dominance of pigments and CDOM over inorganic suspended particles	Meso-eutrophic	B
3	Clear waters	Oligotrophic	A
4	Turbid waters with high organic content	Meso-eutrophic	B
5	Sediment-laden waters	Meso-eutrophic	B
6	Balanced effects of optically active constituents at shorter wavelength	Hypereutrophic	C
7	Highly productive waters with high cyanobacteria abundance and elevated reflectance at the red/near-infrared spectra	Hypereutrophic	C
8	Productive waters with cyanobacteria presence and with R_{rs} peak close to 700 nm	Hypereutrophic	C
9	Optically neighbouring to OWT2 but higher R_{rs} at shorter wavelength	Oligotrophic	A
10	CDOM rich waters	Hypereutrophic	C
11	High in CDOM with cyanobacteria presence and high absorption efficiency by NAP	Meso-eutrophic	B
12	Turbid, moderately productive waters with cyanobacteria presence	Meso-eutrophic	B
13	Clear blue water	Oligotrophic	A

moderately transparent water conditions associated with mesotrophic status and C represents turbid and/or hypereutrophic conditions.

2.2.4. Chlorophyll-a algorithms

Nine widely used chl-*a* algorithms (Table 2) were evaluated for the case study lakes. Algorithms 1 to 7 are empirical band-ratio and (semi-) analytical algorithms, 8 and 9 are probabilistic neural networks. All algorithms use multispectral R_{rs} (λ) in selected bands as input (Table 2).

Algorithm parameters were optimised for the case study conditions within each OWT cluster (Table 1) using a nonlinear fit for algorithms 3, 4 and 6 and a linear model fit for algorithm 5, because global algorithm coefficients often fail to generalise effectively to individual lakes (Werther et al., 2022b). Algorithms 1, 2 and 7 have fixed parameters, and the two probabilistic algorithms (8 and 9) are calibrated based on large-scale application and therefore cannot be optimised.

MSI estimates in oligo-mesotrophic lakes with low to moderate chl-*a* concentrations are particularly susceptible to inaccuracies in satellite-derived retrievals (Neil et al., 2019; Werther et al., 2022a). To address this, we introduce a separate classification, "BNN < 10," which uses the Bayesian Neural Network (BNN) approach by Werther et al. (2022b). This approach targets lakes with chl-*a* below 10 mg m^{-3} in OWTs 2, 3, 4, 5, and 9, providing estimates that carry a BNN-specific uncertainty below 60 %. Whenever these BNN-based low uncertainty estimates are available, they take precedence over other algorithm values.

The algorithm performance for each OWT cluster was assessed using three widely used metrics: the Pearson correlation coefficient (R), to measure the strength of the linear relationship between in situ and MSI chl-*a* estimates; the mean absolute error (MAE); and the Median Symmetric Accuracy (MdSA), which quantifies the median error between in situ chl-*a* and estimated chl-*a* (Morley et al., 2018), and is commonly used in remote sensing literature (Pahlevan et al., 2020; Werther et al., 2022b):

Table 2
Chl-*a* estimation algorithms used in this study. Ranges are adjusted based on findings of Odermatt et al. (2012).

	Name	Abbreviation	Used bands	Chl-*a* application range (mg m^{-3})	Reference
1	Two Band MERIS	Gil2band	665, 708	0-80	(Gilerson et al., 2010)
2	Three Band Meris	Gil3band	665, 708 and 753	0-80	(Gilerson et al., 2010)
3	Ocean Colour 2-band	OC2	490,560	0.012-15	(O'Reilly, 2000)
4	Ocean Colour 3-band	OC3	443,490,560	0.012-15	(O'Reilly, 2000)
5	Moses 2-band	Moses	665,708	10-100	(Moses et al., 2009)
6	Gurlin 2-Band	Gurlin	665,708	2.3-200.8	(Gurlin et al., 2011)
7	Gons05	Gons	665,708, 783	0-100	(Gons et al., 2002)
8	Bayesian Neural Network based on Monte Carlo Dropout	BNN	All bands (443 – 783)	0-68	(Werther et al., 2022b)
9	Mixture Density Network	MDN	All bands (443 – 783)	0-1000	(Pahlevan et al., 2020)

$$MdSA = \left(e^{median\left(\left[\log\left(\frac{y_{est}}{y_{true}}\right)\right]\right)} - 1 \right) \times 100 \quad (1)$$

and the Symmetric Signed Percentage Bias (SSPB), which evaluates whether an algorithm systematically over- or under-estimates chl-*a* (Morley et al., 2018):

$$SSPB = sign\left(median\left(\log\left(\frac{y_{est}}{y_{true}}\right)\right)\right)\left(e^{\left[median\left(\log\left(\frac{y_{est}}{y_{true}}\right)\right)\right]} - 1\right) \times 100 \quad (2)$$

In which Y_{est} represent the estimated chl-*a* concentrations from the model, algorithm or MSI and Y_{true} represents the in situ chl-*a* concentrations. The best-performing algorithms were identified for each OWT cluster and subsequently combined to calculate MAE, MdSA, and SSPB values over the whole chl-*a* range. These metrics facilitated the selection of the optimal OWT-clustered chl-*a* algorithm and atmospheric correction combination for each lake.

2.3. Conformal prediction uncertainty quantification

Conformal prediction (CP) (Angelopoulos & Bates, 2023, Vovk et al., 2005) was adapted here to quantify uncertainty in MSI-derived chl-*a* estimates (Fig. 2). We used a *non-conformity percentage* as a proxy for uncertainty, based on the idea that the uncertainty associated with a new MSI chl-*a* retrieval relates to how the chl-*a* retrieval error compares with those of the reference observations. In this context, error is the non-conformity score (α_i) defined as the absolute error between an MSI chl-*a* estimate and its corresponding match up in situ measurement.

The absolute error for MSI chl-*a* estimates with corresponding in situ observations (matchups) was calculated and the dataset was split into two sets: conformity (reference dataset) and test (new observations). In the CP framework, the terms conformity and test are used instead of calibration and validation to avoid confusion with their traditional use in chl-a algorithms and GLM-AED model tuning. While these dataset splits may appear similar, their functions differ: calibration and validation involve adjusting and evaluating model parameters, whereas conformity and test assess the consistency of new data without modifying the model. The terminology reflects these distinct roles.

One year was assigned as the test set, while the remaining years (six for Esthwaite Water and five for Loch Leven) formed the conformity set (Fig. 2). As noted in section 2.2.3, not all MSI chl-*a* estimates had corresponding in situ chl-*a* measurements (i.e., no match-up); these estimates were assigned to the test set. Their absolute errors for these observations were estimated using the following procedure:

1. Identifying the three in situ chl-*a* measurements closest in time (day of the year).
2. Averaging these measurements by weighting same-year observations twice to account for inter-annual variability.
3. Calculating the absolute difference between the MSI estimate and this weighted average.

We then sorted the non-conformity scores from the conformity set ($\alpha_{conformity}$) in ascending order. For each test observation, we determined the position of its non-conformity score ($\alpha_{test,i}$) within the sorted conformity set and calculated its percentile rank, termed the non-conformity percentage.

Prediction intervals were derived using the 5th and 95th percentiles of $\alpha_{conformity}$ and scaled by the ratio of $\alpha_{test,i}$ to the median of $\alpha_{conformity}$ to accommodate varying uncertainty levels. Calibration quality was evaluated using a coverage factor, i.e. the proportion of intervals that successfully encompassed the interpolated in situ chl-*a* values, including

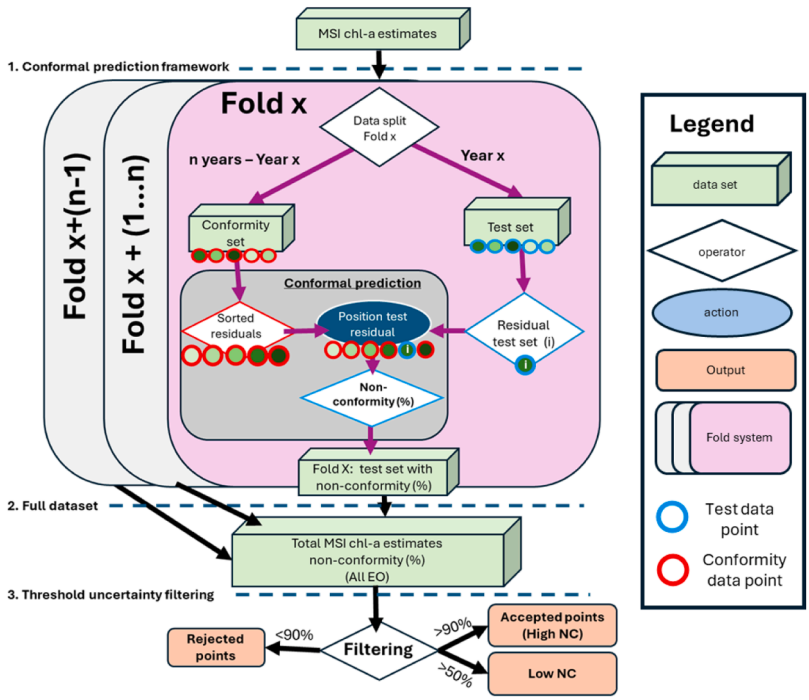

Fig. 2. Overview of the conformal prediction approach developed based on Angelopoulos & Bates (2023) and Vovk et al. (2005).

those without a direct match-up (Saranathan et al., 2024; Werther, et al., 2022a; Yao et al., 2019).

An automatic rejection threshold ($\hat{q}$) was set at the 90th percentile of $\alpha_{conformity}$: if a test sample's non-conformity percentage exceeded this threshold, it was classified as an outlier. Additionally, to illustrate the impact of uncertainty filtering, a stricter manual threshold of 50 % was applied, resulting in three filtering scenarios for MSI-derived chl-*a* estimates: All Earth Observations (EO) (no filtering), High NC (filter out values above 90 %) and Low NC (filter out values above 50 %).

The conformal prediction process was executed in folds that resampled the data, ensuring that each year served as the test set once so that every MSI chl-*a* estimate received a non-conformity percentage. For estimates appearing in multiple folds, the final non-conformity percentage was computed as the average across folds.

Our CP approach includes two adaptations to quantify uncertainty in satellite-derived chl-a estimates. First, rather than splitting data into three sets (training, calibration, and test) as required by typical CP applications in machine learning settings (Angelopoulos & Bates, 2023; Singh et al., 2024; Waczak et al., 2024), we leverage the fact that most chl-a algorithms used in this study are semi-analytical or empirical. These methods do not require training; hence, we partition the available dataset directly into a calibration set and an evaluation (test) set.

Second, exchangeability, a key assumption for the validity of CP (Angelopoulos & Bates, 2023; Singh et al., 2024; Vovk et al., 2005), requires that the order of observations does not affect their joint distribution. To approximate this condition in our dataset, we split the dataset by year rather than using random sampling. This approach naturally accounts for seasonal variations by ensuring that observations within each year are drawn from the same underlying distribution, which is essential for the CP guarantees to hold. Our CP implementation is publicly available on GitHub (see the Code Availability section).

2.4. GLM-AED recalibration

The GLM-AED model was applied to Esthwaite Water and Loch Leven to simulate chl-*a* concentrations. Each lake was manually calibrated by adjusting key parameters within realistic ranges: the light extinction coefficient (*Kw*, 0.1–1.5 m^{-1}), which affects light penetration and thermal stratification, and the wind factor ($f_{wind,}$ 0 - 2), which scales wind speed and influences mixing and surface cooling.

Sensitivity analyses identified the maximum growth rates of phytoplankton groups (R_{growth}) as key parameters for optimisation. These parameters were found to offer the greatest potential for improving model performance but are difficult to determine accurately from existing literature. For this study, two functional phytoplankton groups were defined: one representing diatom and the other cyanobacteria, as these were identified as the dominant bloom-forming taxa in the databases of the two lakes.A systematic parameter calibration was conducted to optimise the maximum growth rates for cyanobacteria (R_{cyano}) and diatoms (R_{diatom}). A total of four hundred combinations of the two parameters were evaluated using a 20 × 20 grid, with simulations run for the period 2016 to 2022. This approach allowed for the identification of parameter combinations that produced the best agreement with in situ chl-*a* dynamics. For Esthwaite Water (R_{diatom}: 0.06 – 2.8, R_{cyano}: 0.26 – 1.16) and Loch Leven (R_{diatom}: 0.1 – 4, R_{cyano}: 0.1 – 4) the growth rate ranges were thus lake-specific and ensured that the parameter space was representative of the ecological conditions in each system.

GLM-AED model calibration was conducted on even years (2016, 2018, 2020, and 2022), while validation was performed on odd years (2017, 2019, and 2021) following the method used by Essou et al., (2016) and as suggested by Xu, (2021) for hydrological models due to its spatial and temporal varying data quality and to consider climatic trends. For each of the 400 runs, the MdSA and the Mean Absolute Error (MAE):

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}, \quad (3)$$

were calculated using the calibration years. The top three runs for each scenario, based on these metrics, were selected for further validation. GLM-AED model validation was conducted by calculating the MdSA and MAE based on the model-derived surface chl-*a* and in situ chl-measurements.

2.4.1. Evaluation scenarios

GLM-AED model calibration was undertaken using seven different

scenarios. These scenarios were based on the different uncertainty filtering scenarios of the MSI-derived chl-*a* estimates (Section 2.3) per lake (Table 3).

The seven scenarios allow us to evaluate three key aspects. First, they assess whether using MSI chl-a estimates enhances the calibration of GLM-AED (scenario a). Second, they examine whether filtering MSI chl-a estimates further improves GLM-AED accuracy (comparison of scenarios b, c, and d). Third, they evaluate whether a calibration approach that incorporates in situ and MSI chl-a estimates is more effective than one that relies solely on MSI chl-a estimates (comparison of scenarios b, c, and d versus e, f, and g).

2.4.2. Algal bloom agreement

In situ and model-derived chl-*a* were labelled as 'Bloom' if they exceeded 25 mg m^{-3}. The threshold is based on Alert level 2 (>24 mg m^{-3}) for cyanobacteria of the World Health Organization (World Health Organization, 2021).To evaluate the model's alignment with in situ chl-*a* bloom occurrences (*Bloom agreement*) for each scenario, instances in which in situ chl-*a* values and GLM-AED chl-*a* outputs simultaneously indicated a bloom were counted ($N_{correct}$) and divided by the total number of bloom instances in the in situ chl-*a* ($N_{observed}$):

$$Bloom\ Agreement\ (\%) = \left(\frac{N_{correct}}{N_{observed}}\right) * 100. \quad (4)$$

The bloom agreement percentage per scenario, per lake provides insight into the impact of filtered MSI chl-*a* estimates on bloom forecasting.

3. Results

3.1. MSI spatial homogeneity

Polymer derived reflectance produced the lowest spatial variability (n=105, n=172 scenes with CV<0.25) in Esthwaite Water and Loch Leven, when compared to the atmospheric correction algorithms ACOLITE (n=100, n=97) and C2RCC (n=87, n =153).

The difference in spatial variability across the atmospheric correction methods is an inherent indicator of their limitations. The most spatially variable scenes (CV>0.25) in Esthwaite Water and Loch Leven were removed. Match-ups were available for approximately 25 % of scenes in Loch Leven (Polymer = 47, C2RCC = 39, ACOLITE = 24) and 36 % of scenes in Esthwaite Water (Polymer = 37, C2RCC = 29, ACOLITE = 40).

3.2. Optical Water Type classification

All OWTs, except 10, were observed at least once across all atmospheric correction approaches and lakes, but their relative frequencies varied strongly across the different atmospheric correction methods (Fig. 3). Polymer and C2RCC aligned in that they excluded certain OWT classes (1, 5, 7 and 8 for Esthwaite Water, and 1 and 7 for Loch Leven), whilst ACOLITE observes classes 1 in Esthwaite Water and 5 in both lakes (Fig. 3). C2RCC and Polymer show highest representation of clusters A and B with cluster C being absent in C2RCC. ACOLITE shows equal representation of clusters A, B and C in Esthwaite Water. In Loch Leven, ACOLITE assigns most scenes to cluster B, showing ACOLITE to be the outlier.

Table 3
The seven different evaluation scenarios used for model recalibration. Each scenario has a letter (a-g), a Name and a column explaining which chl-*a* data is part of that scenario. Non-Conformity is later abbreviated as NC and Earth Observation as EO (Scenario e, f and g).

Scenario	Name	Description
a	In situ	All available in situ chl-a values
b	All Earth Observation	All MSI chl-a estimates
c	High Non-Conformity	MSI chl-a estimates with a non-conformity scare ≤90 %
d	Low Non-Conformity	MSI chl-a estimates with a non-conformity score ≤ 50 %
e	In situ + All EO	a + b
f	In situ _ High NC	a + c
g	In situ + Low NC	a + d

3.3. MSI chl-a estimation

The MSI chl-*a* retrieval resulted in higher R values in Esthwaite Water compared to Loch Leven and a better visual fit (Fig. 4). Moderate chl-*a* range (10- 40 mg m^{-3}) of Loch Leven was inadequately retrieved by the chl-*a* algorithms. The OWT cluster-based optimisation improved the retrieval for both lakes compared to the non-optimised MSI chl-*a* estimates (grey triangles, Fig. 4), which showed up to 50 % lower R values, 3.8 times higher MdSA's and 16 times higher SSPB values.

The diversity between atmospheric correction methods was visible in the error metrics of the chl-*a* retrieval results (Fig. 4). The C2RCC SSPB was 3 to 10 times higher compared to Polymer and ACOLITE, with the MDSA (higher) and R (lower) metrics deterioration as well. The ACOLITE SSPB for Loch Leven and MdSA for Esthwaite Water were observed to be the lowest. However, Polymer outperforms across the remaining error metrics and therefore the better choice.

3.4. Uncertainty quantification

The uncertainty quantification using CP resulted in an expected range of non-conformity scores for MSI chl-*a* estimate from 0 to 100 %. A rejection rate (non-conformity > 90 %) of ± 29 % was observed with 29 MSI chl-*a* estimates in Esthwaite Water and 26 in Loch Leven (Fig. 5).

Removal of rejected values resulted in a significant reduction of the MdSA from 53.83 % to 36.55 % in Esthwaite Water (MAE: 9.2 to 4.95) and in Loch Leven from 39.17 % to 24.93 % (MAE: 12.6 to 7.65). Adjusting the rejection rate further to 50 % (Low NC) changed the MdSA of Esthwaite Water minimally from 36.55 % to 37.33 % (MAE: 4.95 to 3.81). However, reduced it significantly for Loch Leven from 39.17 % to 16.21 % (MAE: 7.65 to 5.38). The coverage factor is lower for Esthwaite Water (75.58 %) than for Loch Leven (80.95 %), indicating that the uncertainty percentages in Loch Leven were slightly better calibrated than those for Esthwaite Water.

3.5. GLM application

Esthwaite Water resulted in lower errors between in situ and GLM-AED compared to Loch Leven. An overestimation of the chl-*a* peaks by 25 mg m^{-3} in Esthwaite Water and 50 to 75 mg m^{-3} in Loch Leven was observed in all scenarios.

Calibration scenarios relying solely on MSI chl-*a* estimates (b,d,f) consistently reduced the MdSA in Loch Leven compared to scenarios using in situ data. In Esthwaite Water, the scenario d resulted in a reduction in MdSA, making it the sole scenario to demonstrate a decrease across both lakes (Fig. 6).

Scenarios combining MSI chl-*a* estimates and in situ chl-*a* (c,e,g) resulted in decreased or equal MdSA values in 5 out of 6 cases, with g showing no change for either lake. In Esthwaite Water, High NC + In situ (e) was the only scenario involving in situ data that reduced the MdSA.

Calibration using in situ (a) captured fifty percent of blooms in Esthwaite Water and 36 % in Loch Leven (Fig. 6). In Loch Leven, the MSI chl-*a* estimate scenarios increased bloom agreement up to 80 % (b and d) and 64 % (f). Esthwaite Water bloom agreement increased in two scenarios (d and e) up to 63.6 %, making the High NC scenario the best performing across the lakes. Despite improving the MdSA in both lakes, the All EO scenario (f) resulted a significant decrease in bloom agreement to 31 % in Esthwaite Water and reduced improvement in Loch

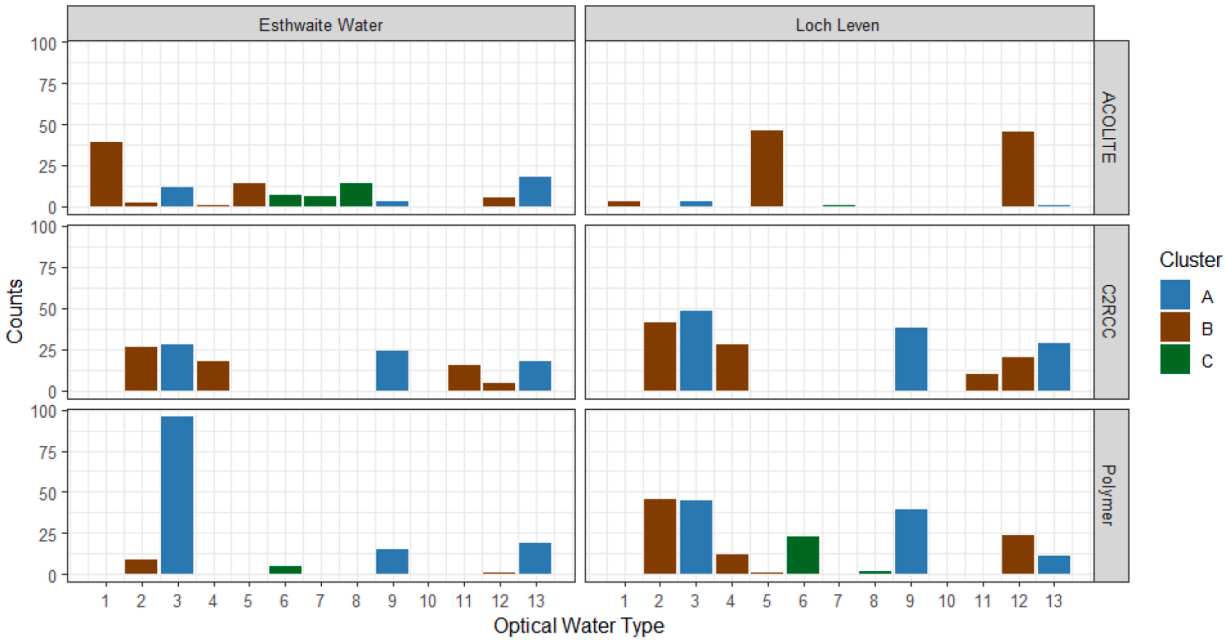

Fig. 3. Total number of images per Optical Water Type, per lake and, per AC method. The colours represent the different OWT clusters: A (clear), B (intermediate), and C (turbid), providing a visual distinction of the distribution of individual OWT across the different clusters.

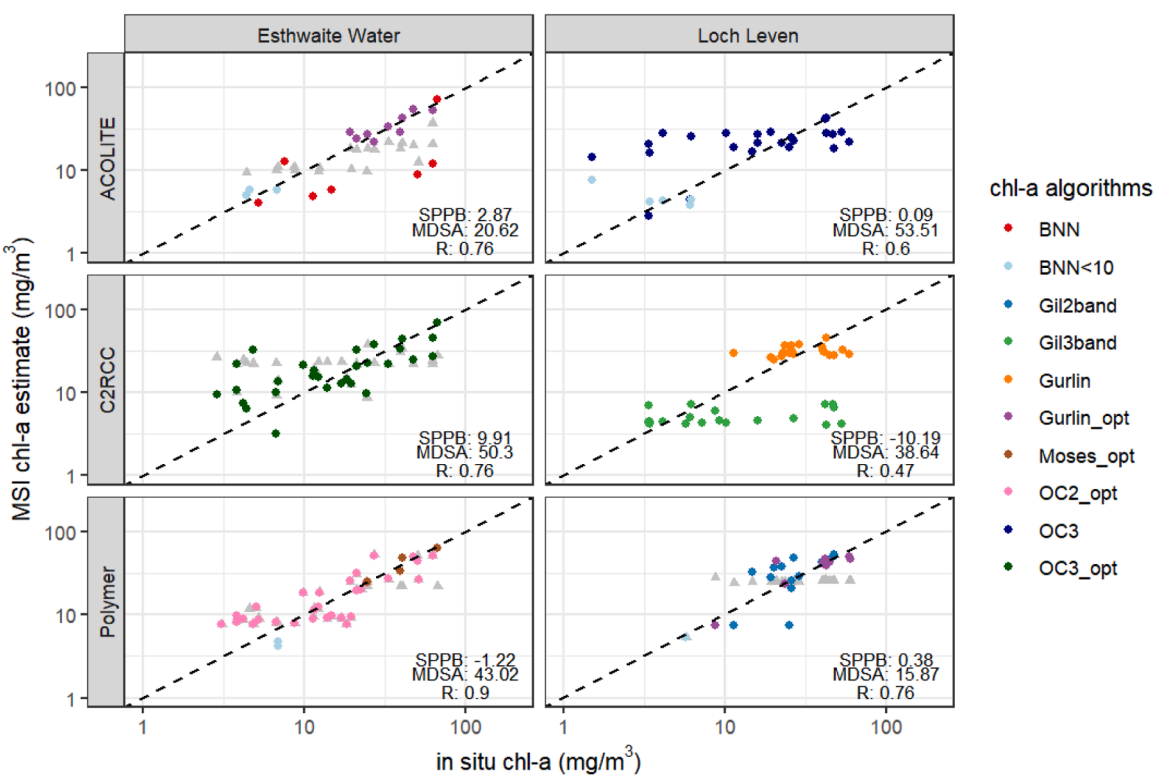

Fig. 4. Performance of the best chl-*a* algorithms for each lake and atmospheric correction method. Grey triangles represent non-optimised algorithms, highlighting the comparison between optimised (coloured points) and non-optimised algorithm performance across the different conditions. Performance is evaluated.

Leven.

4. Discussion

The GLM-AED model chl-*a* output for Loch Leven and Esthwaite Water is line with the current literature. For comparison with (Fenocchi et al., (2019) who report errors in terms of nMAE, the MAE of 19.73 mg m^{-3} for Loch Leven and 12.61 mg m^{-3} for Esthwaite Water correspond to a normalised MAE values of 88.3 % and 56.5 %, respectively. These values are in line with the nMAE of 56 % ± 24 % found in Fenocchi et al (2019) for a range of different lakes. Loch Leven exhibits a larger error compared to Esthwaite Water, with higher MAE, nMAE, and MdSA values, reflecting greater challenges in accurately representing this lake as well as the differences between lakes characterised by the same

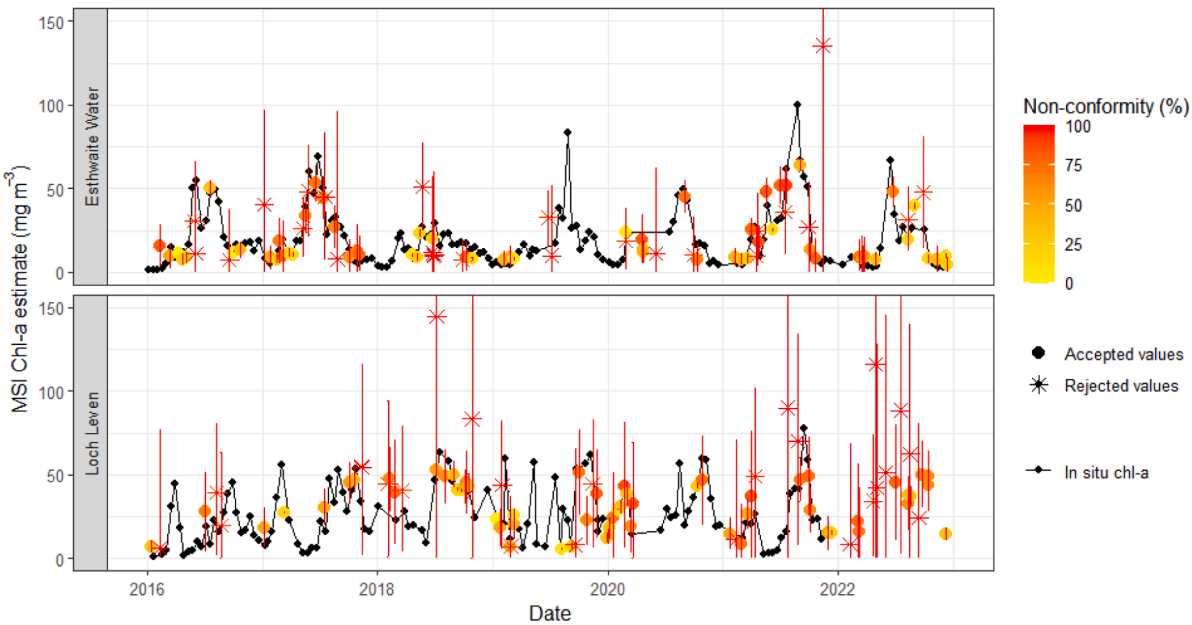

Fig. 5. Non-conformity quantification of the MSI chl-*a* estimates, along with their corresponding prediction intervals. In situ chl-*a* values are shown in black, with MSI chl-*a* estimates coloured based on their conformity. Rejected estimates, where non-conformity exceeds 90 %, are marked displayed as an asterisk and accepted estimates are represented as points. The data is separated by lake to illustrate the variability in conformity.

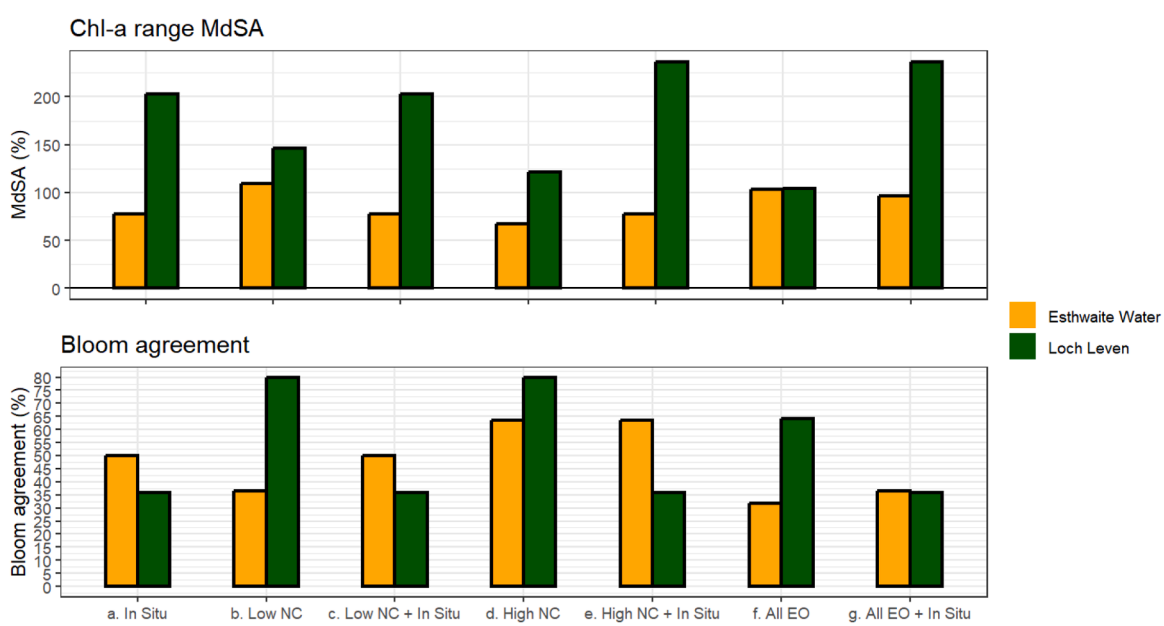

Fig. 6. Median Absolute Scaled Error (MdSA, %) of the full chl-*a* range for each scenario and lake (a-g) shown on the left. On the right, the percentage of bloom agreement between in situ chl-*a* concentrations and the GLM-AED scenario outputs (a-g) is displayed. Both Fig.s highlight the accuracy and agreement of chl-*a* estimations across different scenarios and lakes.

model.

4.1. MSI derived chl-a

The MSI derived chl-*a* also reported higher errors in Loch Leven, compared to Esthwaite Water across atmospheric correction and chl-*a* algorithms. This reoccurring finding across different data sources can be explained by the shallower nature of Loch Leven resulting in more dynamic clear and turbid states (Scheffer & Van Nes, 2007).

Overall, MSI-derived chlorophyll-a estimates achieved reasonable accuracy across the two case study lakes using the Polymer atmospheric correction (R: 0.76–0.90; MdSA: 15.67 %–43.02 %). Optimal performance required the combined use of OWT clustering, appropriate atmospheric correction, and careful selection of chl-a retrieval algorithms.

4.2. Uncertainty quantification through conformal prediction

The high coverage factor (75 – 80 %) and improved error metrics after CP filtering showed that the approach is an effective tool for quality control of MSI derived chl-*a* concentrations. While some inaccuracies were observed, in particular chl-*a* estimates showed high nonconformity scores despite being close to in situ measurements, these variations

provided valuable insights into the dataset. The anomalous climatic conditions of 2017, the fifth warmest year since 1910 (Royal Meteorological Society, 2018), likely contributed to these discrepancies, alongside the limited number of in situ matchups available for that period. Understanding these factors helps refine the CP approach and improve uncertainty estimates. Importantly, since these errors occurred at chl-*a* values below 30 mg m^{-3}, their impact on bloom detection outcomes remained minimal.

The uncertainty estimates from our CP approach align with those reported in previous studies. For instance, Saranathan et al., (2023) reported uncertainties of 10 to 45 % for chl-*a* using Mixture Density Networks, while Werther et al. (2022b) found uncertainties ranging from approximately 30 % to 80 % for chl-*a* over oligotrophic and mesotrophic lakes, depending on prior atmospheric correction and lake characteristics. Similarly, global chl-*a* estimates by Liu et al. (2021) yielded comparable uncertainty ranges, suggesting that our findings reflect common capabilities of chl-*a* estimation via Sentinel-2 MSI.

A wide range of uncertainty quantification methods for chl-*a* retrieval exist (IOCCG, 2019), but in practice are often built on machine learning–based approaches (e.g., Saranathan et al., 2024) techniques tailored to specific models or optical water types (e.g., Liuet al., 2021), and uncertainty propagation calculations (e.g., Xi et al., 2021). These methods are usually not model agnostic, therefore not easily transferrable. In contrast, CP fills a critical gap by providing uncertainty estimates for any chl-*a* algorithm, provided a sufficient dataset is available, even for algorithms originally published without uncertainty metrics.

4.3. Incorporation of MSI derived chl-a

Calibration using MSI chl-*a* estimates filtered using the High NC scenario produced significant reductions in MdSA, demonstrating two key points. First, MSI chl-*a* estimates can improve the calibration of the chl-*a* model; second, filtering these estimates prior to downstream use is not only beneficial but necessary to prevent calibration deterioration.

Although we expected that combining MSI chl-*a* estimates with in situ data would yield even better results, our findings indicate that the combined use provided no additional model improvement or bloom agreement. We speculate that the in situ observations are dominated by low chl-*a* values (<25 mg m^{-3}), whilst the MSI estimates show an even distribution after matchups. In addition, satellite estimates are constrained to surface chl-*a* values only, whilst in situ tends to be a mixed sample from 0 to 5 m which could result in a more diluted sample. As no negative effects were observed when combining the two sources, structured integration of chl-*a* values creates potential for further enhancement.

The High NC scenario improved bloom agreement in both lakes, while scenarios that incorporated unfiltered MSI chl-*a* estimates (All EO, scenarios B and E) or in situ data (scenarios F and G) showed diminished or no improvements. This indicates that filtering MSI chl-*a* estimates is essential for optimising their predictive utility, and we propose filtering based exclusively on the High NC scenario metric of rejected values.

4.4. Towards large-scale algal bloom forecasting

Despite the merits of each method, neither GLM-AED nor EO alone would be able to support large-scale algal bloom forecasting. GLM-AED's capacity to estimate bloom events is often constrained by limited calibration datasets, especially for regions lacking consistent in situ observations. Meanwhile, EO platforms can only provide near-real-time or retrospective measurements, heavily dependent on clear-sky conditions. Consequently, relying on GLM-AED alone risks calibration inaccuracies, and solely using EO forfeits the ability to forecast.

The combined use offers a novel pathway for scaling algal bloom forecasting to a global level. Its core strength lies in leveraging complementary capabilities: GLM-AED grounds bloom estimation in physical principles constrained by hydrodynamic forcing parameters, while satellite observations, such as those derived from Sentinel-2 MSI, provide spatially extensive estimates of chl-*a* based on bio-optical relationships.

A critical first step for broad application of this combined approach is establishing a robust hydrodynamic foundation in GLM, ensuring that water balance and temperature/oxygen profiles are accurately modelled prior to refining chl-*a* calibration. In cases where hydro-meteorological observations are lacking, consistent input variables (e.g. inflow, outflow, meteorological drivers) can be derived from large-scale reanalysis products, such as the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) (Gelaro et al., 2017) or ERA-5 for atmospheric drivers and Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998) for flow and nutrients. These often have broad spatial coverage and high temporal resolution, thereby overcoming observational scarcity and provide reliable forcing conditions for the model.

Whilst Bruce et al., (2018) demonstrated multi-lake applications of GLM, no parallel large-scale study using GLM-AED with EO datasets has been completed. A logical next step would be to apply the approach developed in this study across a broader set of lakes globally, moving toward large-scale algal bloom forecasting. However, scaling up to the global level presents substantial challenges. These include the selection and adaptation of atmospheric correction models, the lack of globally validated in situ datasets, differences in satellite mission coverage, and the absence of a globally consistent Sentinel-2 water-leaving reflectance product for inland waters. Atmospheric correction performance is highly sensitive to the OWT encountered (Pahlevan et al., 2021), which complicates algorithm transferability. The European Space Agency Lakes Climate Change Initiative (CCI) project is a key initiative in this space, aiming to generate consistent global-scale water quality products from satellite data, and has been successfully used for variables including chl-*a* (Caroni et al., 2025). Our results further show that even among two eutrophic lakes, optimal chl-*a* algorithm performance and GLM model behaviour differ substantially. This highlights a core limitation of global process-based modelling: either lakes must be parameterised individually, requiring considerable effort, or generalised models must be accepted with reduced confidence in their outputs. Addressing this trade-off will be crucial for the development of robust, global-scale bloom forecasting systems. As reanalysis datasets improve and satellite product accuracy increases, near-real-time and broad-scale monitoring and forecasting will become a more feasible option for lake management.

5. Conclusions

This study demonstrates that integrating MSI chl-*a* estimates, when quality-filtered using CP, significantly enhances GLM-AED model calibration and algal bloom detection. The ability to quantify uncertainties in remote sensing-derived chl-*a*, alongside reductions in validation model errors and significant improvements in bloom agreement, underscores the value of this approach and the broader role of remote sensing in ecological modelling.

The negative impact of unfiltered MSI chl-*a* estimates on model performance highlights the necessity of quality filtering methods such as CP to minimise downstream errors. These findings emphasise the potential of CP-filtered MSI chl-*a* estimates for improving bloom forecasting and supporting more informed lake management decisions. This approach provides a robust framework for integrating EO-based water quality data into predictive models, with applications extending beyond algal blooms. Future research could explore its effectiveness across different water systems, EO datasets, and modelling techniques, further validating its role in ecological forecasting.

Code, model and data availability

The code for the conformal prediction framework is available on: https://github.com/maudsiebers/Conformal_prediction_Framework

GLM-AED is freely available. In this work Version 3.1.9 of the GLM-AED was downloaded and executed in RStudio. For running and visualisation purposes the packages 'glmtools'(https://github.com/robertladwig/glmtools) and 'GLM3r' (https://github.com/robertladwig/GLM3r) were installed from GitHub.

Data for lakes is freely available from the UK Lakes Portal (https://uklakes.ceh.ac.uk), managed by the UK Centre for Ecology & Hydrology (UKCEH). For additional data requests, enquiries can be made directly to UKCEH. These are the published datasets of relevance for Esthwaite Water: (Feuchtmayr et al., 2021, 2024, 2025) and Loch Leven (Dudley et al., 2013; Taylor et al., 2021, 2022a, 2022b, 2023)

Sentinel-2 data is publicly available through multiple platforms, with the primary source being the Copernicus Open Access Hub (https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-2). Additional access points include ESA's Sentinel Hub, Google Earth Engine, NASA Earthdata Search, and AWS Open Data, each offering various tools for visualisation and analysis.

Financial support

This work is a part of the Marie Sklodowska-Curie Actions ITN inventWater (Grand no: 956623 H2020-MSCA-ITN).

AI Use

During the preparation of this work the author(s) used Chat-GPT in order to improve language clarity and sentence structure. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Maud A.C. Siebers: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mortimer Werther:** Writing – review & editing, Supervision, Methodology. **Daniel Odermatt:** Writing – review & editing, Supervision. **Eleanor Mackay:** Writing – review & editing, Data curation. **Linda May:** Writing – review & editing, Data curation. **Thomas Shatwell:** Writing – review & editing, Supervision. **Ian Jones:** Supervision. **Matthew Blake:** Data curation. **Peter D. Hunter:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Acknowledgements

The results contain modified Copernicus Climate Change Service information 2020. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. The sampling and processing were supported by the Natural Environment Research Council award number NE/R016429/1 as part of the UK-SCAPE programme delivering National Capability.

We thank the two anonymous reviewers for their supportive and constructive comments on the manuscript.

Data availability

Links to data are described in manuscript.

References

Angelopoulos, A.N., Bates, S., 2023. Conformal prediction: a gentle introduction. Found. Trends® Mach. Learn. 16 (4), 494–591. https://doi.org/10.1561/2200000101.

Ansper, A., & Alikas, K. (2018). Retrieval of chlorophyll a from Sentinel-2 MSI data for the European Union water framework directive reporting purposes. Volume 11, Issue 1, 11(1). https://doi.org/10.3390/rs11010064.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: Model Development1. JAWRA J. Am. Water Resour. Assoc. 34 (1), 73–89. https://doi.org/10.1111/J.1752-1688.1998.TB05961.X.

Atwood, E.C., Jackson, T., Laurenson, A., Jönsson, B.F., Spyrakos, E., Jiang, D., Sent, G., Selmes, N., Simis, S., Danne, O., Tyler, A., Groom, S., 2024. Framework for regional to global extension of optical water types for remote sensing of optically complex transitional water bodies. Remote Sens. 16 (17), 3267. https://doi.org/10.3390/RS16173267, 2024, Vol. 16, Page 3267.

Beaulieu, M., Pick, F., Gregory-Eaves, I., 2013. Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. Limnol. Ocean. 58 (5), 1736–1746. https://doi.org/10.4319/LO.2013.58.5.1736.

Bormans, M., Ford, P.W., Fabbro, L., 2005. Spatial and temporal variability in cyanobacterial populations controlled by physical processes. J. Plankton. Res. 27 (1), 61–70. https://doi.org/10.1093/PLANKT/FBH150.

Brockmann, C., Roland, Peters, M., Kerstin, Sabine, & Ruescas, A. (2016). 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.

Bruce, L.C., Frassl, M.A., Arhonditsis, G.B., Gal, G., Hamilton, D.P., Hanson, P.C., Hetherington, A.L., Melack, J.M., Read, J.S., Rinke, K., Rigosi, A., Trolle, D., Winslow, L., Adrian, R., Ayala, A.I., Bocaniov, S.A., Boehrer, B., Boon, C., Brookes, J. D., Hipsey, M.R., 2018. A multi-lake comparative analysis of the General Lake Model (GLM): stress-testing across a global observatory network. Environ. Model. Softw. 102, 274–291. https://doi.org/10.1016/J.ENVSOFT.2017.11.016.

Bueche, T., Hamilton, D.P., Vetter, M., 2017. Using the general lake model (GLM) to simulate water temperatures and ice cover of a medium-sized lake: a case study of Lake Ammersee, Germany. Env. Earth Sci. 76 (13), 1–14. https://doi.org/10.1007/S12665-017-6790-7/TABLES/5.

Bukata, R.P., Harris, G.P., Bruton, J.E., 1974. The detection of suspended solids and chlorophyll "A" utilizing digital multispectral ERTS-1 data. In: Proc. Second Canadian Symposium on Remote Sensing, 2, pp. 551–564.

Burggraaff, O., 2020. Biases from incorrect reflectance convolution. Opt. Express. 28 (9), 13801. https://doi.org/10.1364/OE.391470.

Caroni, R., Greife, A.J., Bresciani, M., Giardino, C., Tellina, G., Carrea, L., Liu, X., Simis, S., Albergel, C., Pinardi, M., 2025. Examining global trends of satellite-derived water quality variables in shallow lakes. Remote Sens. Appl. 38, 101565. https://doi.org/10.1016/J.RSASE.2025.101565.

Carvalho, L., Miller, C., Spears, B.M., Gunn, I.D.M., Bennion, H., Kirika, A., May, L., 2012. Water quality of Loch Leven: responses to enrichment, restoration and climate change. Hydrobiologia 681 (1), 35–47. https://doi.org/10.1007/s10750-011-0923-x.

Concha, J.A., Bracaglia, M., Brando, V.E., 2021. Assessing the influence of different validation protocols on Ocean Colour match-up analyses. Remote Sens Env. 259. https://doi.org/10.1016/j.rse.2021.112415.

Deng, Y., Zhang, Y., Pan, D., Yang, S.X., Gharabaghi, B., 2024. Review of recent advances in remote sensing and machine learning methods for Lake water quality management. Remote Sens. 16 (22), 4196. https://doi.org/10.3390/RS16224196, 2024, Vol. 16, Page 4196.

Dudley, B.J., May, L., Spears, B.M., Kirika, A., 2013. Loch Leven Long-Term Monitoring Data: Phosphorus, Silica And Chlorophyll Concentrations, Water Clarity And Temperature, 1985-2007. NERC Environmental Information Data Centre. https://doi.org/10.5285/2969776d-0b59-4435-a746-da50b8fd62a3.

Eleveld, M.A., Ruescas, A.B., Hommersom, A., Moore, T.S., Peters, S.W.M., Brockmann, C., 2017. An optical classification tool for Global Lake Waters. Remote Sens. 9 (5), 420. https://doi.org/10.3390/RS9050420, 2017, Vol. 9, Page 420.

Essou, G.R.C., Arsenault, R., Brissette, F.P., 2016. Comparison of climate datasets for lumped hydrological modeling over the continental United States. J. Hydrol. 537, 334–345. https://doi.org/10.1016/J.JHYDROL.2016.03.063.

Fadel, A., Sharaf, N., Siblini, M., Slim, K., Kobaissi, A., 2019. A simple modelling approach to simulate the effect of different climate scenarios on toxic cyanobacterial bloom in a eutrophic reservoir. Ecohydrol. Hydrobiol. 19 (3), 359–369. https://doi.org/10.1016/J.ECOHYD.2019.02.005.

Feng, L., Wang, Y., Hou, X., Qin, B., Kuster, T., Qu, F., Chen, N., Paerl, H.W., Zheng, C., 2024. Harmful algal blooms in inland waters. Nat. Rev. Earth. Environ. 5 (9), 631–644. https://doi.org/10.1038/s43017-024-00578-2, 2024 5:9.

Fenocchi, A., Rogora, M., Morabito, G., Marchetto, A., Sibilla, S., Dresti, C., 2019. Applicability of a one-dimensional coupled ecological-hydrodynamic numerical model to future projections in a very deep large lake (Lake Maggiore, Northern Italy/Southern Switzerland). Ecol. Modell. 392, 38–51. https://doi.org/10.1016/J.ECOLMODEL.2018.11.005.

Feuchtmayr, H., Clarke, M.A., De Ville, M.M., Dodd, B.A., Fletcher, J., James, J.B., Mackay, E.B., Pereira, M.G., Rhodes, G., Thackeray, S.J., Maberly, S.C., 2021. Surface Temperature, Surface Oxygen, Water Clarity, Water Chemistry And Phytoplankton Chlorophyll A Data From Esthwaite Water, 2014 to 2018. NERC Environmental Information Data Centre. https://doi.org/10.5285/89ff964d-4aed-4b66-8669-71d42c5add52.

Feuchtmayr, H., Clarke, M.A., De Ville, M.M., Dodd, B.A., Guyatt, H., Hunt, A.G., James, J.B., Rhodes, G., Thackeray, S.J., Maberly, S.C., 2024. Surface Temperature, Surface Oxygen, Water Clarity, Water Chemistry And Phytoplankton Chlorophyll A

Data From Esthwaite Water, UK, 2019-2020. NERC EDS Environmental Information Data Centre. https://doi.org/10.5285/26ea7dfb-59f2-4dc5-b1ee-e9270167a6a2.
Feuchtmayr, H., Clarke, M.A., Dodd, B.A., Guyatt, H., Hunt, A.G., James, J.B., Kelsall, M., Mackay, E.B., McShane, G., Rankin, G., Rhodes, G., Thackeray, S.J., Maberly, S.C., 2025. Surface Temperature, Surface Oxygen, Water Clarity, Water Chemistry And Phytoplankton Chlorophyll A Data From Esthwaite Water, UK, 2021-2022. NERC EDS Environmental Information Data Centre. https://doi.org/10.5285/97256972-e6a5-471c-9236-3a2162075640.
Fontana, M., Zeni, G., Vantini, S., 2023. Conformal prediction: a unified review of theory and new challenges. Bernoulli 29 (1), 1–23.
Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Zhao, B., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J. Clim. 30 (14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1.
Gilerson, A.A., Gitelson, A.A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., Ahmed, S.A., Bailey, W., Franz, B.A., Harding, L.W., Feldman, G.C., McClain, C.R., Stumpf, R.P., Tyler, M.A., Le, C., Li, Y., Zha, Y., Sun, D., Huang, C., Lu, H., 2010. Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. Optics Express 18 (23), 24109–24125. https://doi.org/10.1364/OE.18.024109.
Gons, H.J., Rijkeboer, M., Ruddick, K.G., 2002. A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. J. Plankton Res. 24 (9). https://academic.oup.com/plankt/article/24/9/947/1519859.
Gurlin, D., Gitelson, A.A., Moses, W.J., 2011. Remote estimation of chl-a concentration in turbid productive waters - Return to a simple two-band NIR-red model? Remote Sensing Environ. 115 (12), 3479–3490. https://doi.org/10.1016/j.rse.2011.08.011.
Hedger, R.D., Olsen, N.R.B., Malthus, T.J., Atkinson, P.M., 2002. Coupling remote sensing with computational fluid dynamics modelling to estimate lake chlorophyll-a concentration. Remote Sens Env. 79 (1), 116–122. https://doi.org/10.1016/S0034-4257(01)00244-9.
Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J.N., 2018. ERA5 Hourly Data On Single Levels From 1940 to Present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://cds.climate.copernicus.eu/cdsapp#!/dataset/10.24381/cds.adbb2d47?tab=overview.
Hipsey, M. R. (2022). Modelling aquatic eco-dynamics: overview of the AED modular simulation platform. Zenodo repository. https://aquaticecodynamics.github.io/aed-science/index.html.
Hipsey, M.R., Bruce, L.C., Boon, C., Busch, B., Carey, C.C., Hamilton, D.P., Hanson, P.C., Read, J.S., De Sousa, E., Weber, M., Winslow, L.A., 2019. A General Lake model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake ecological observatory network (GLEON). Geosci. Model. Dev. 12 (1), 473–523. https://doi.org/10.5194/GMD-12-473-2019.
Hunter, P.D., Tyler, A.N., Carvalho, L., Codd, G.A., Maberly, S.C., 2010. Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. Remote Sens Env. 114 (11), 2705–2718. https://doi.org/10.1016/J.RSE.2010.06.006.
IOCCG, 2019. Uncertainties in Ocean Colour Remote Sensing: Vol. No.18 (F. Mélin, Ed.; Issue IOCCG Report Series). International Ocean Colour Coordinatin Group. https://doi.org/10.25607/OBP.
Jöhnk, K.D., Huisman, J., Sharples, J., Sommeijer, B., Visser, P.M., Stroom, J.M., 2008. Summer heatwaves promote blooms of harmful cyanobacteria. Glob. Chang. Biol. 14 (3), 495–512. https://doi.org/10.1111/J.1365-2486.2007.01510.X.
Krinos, A. I., Farrell, K. J., Daneshmand, V., Subratie, K. C., Figueiredo, R. J., & Carey, C. C. (2019). Including variability in air temperature warming scenarios in a lake simulation model highlights uncertainty in predictions of cyanobacteria. BioRxiv, 734285. https://doi.org/10.1101/734285.
Kruse, F.A., Lefkoff, A.B., 1994. Knowledge-based geologic mapping with imaging spectrometers. Remote Sens. Rev. 8 (1–3), 3–28. https://doi.org/10.1080/02757259309532188.
Liu, X., Steele, C., Simis, S., Warren, M., Tyler, A., Spyrakos, E., Selmes, N., Hunter, P., 2021. Retrieval of Chlorophyll-a concentration and associated product uncertainty in optically diverse lakes and reservoirs. Remote Sens Env. 267, 112710. https://doi.org/10.1016/J.RSE.2021.112710.
Maberly, S. C., De Ville, M. M., Kelly, J., & Thackeray, S. J. (2011). The state of Esthwaite Water in 2010.
Mackay, E.B., Folkard, A.M., Jones, I.D., 2014. Interannual variations in atmospheric forcing determine trajectories of hypolimnetic soluble reactive phosphorus supply in a eutrophic lake. Freshw. Biol. 59 (8), 1646–1658. https://doi.org/10.1111/FWB.12371.
Mackay, E.B., Jones, I.D., Folkard, A.M., Barker, P., 2012. Contribution of sediment focussing to heterogeneity of organic carbon and phosphorus burial in small lakes. Freshw. Biol. 57 (2), 290–304. https://doi.org/10.1111/J.1365-2427.2011.02616.X.
Madgwick, G., Jones, I.D., Thackeray, S.J., Elliott, J.A., Miller, H.J., 2006. Phytoplankton communities and antecedent conditions: high resolution sampling in Esthwaite Water. Freshw. Biol. 51 (10), 1798–1810. https://doi.org/10.1111/J.1365-2427.2006.01607.X.
May, L., & Spears, B. M. (2012). Loch leven: 40 years of scientific research. www.springer.com/series/5842.
Mesman, J.P., Ayala, A.I., Adrian, R., De Eyto, E., Frassl, M.A., Goyette, S., Kasparian, J., Perroud, M., Stelzer, J.A.A., Pierson, D.C., Ibelings, B.W., 2020. Performance of one-dimensional hydrodynamic lake models during short-term extreme weather events. Environ. Model. Softw. 133, 104852. https://doi.org/10.1016/J.ENVSOFT.2020.104852.
Morel, A., 1980. In-water and remote measurements of ocean color. Bound. Layer Meteorol. 18 (2), 177–201.
Morley, S.K., Brito, T.V., Welling, D.T., 2018. Measures of model performance based on the log accuracy ratio. Space Weather. 16 (1), 69–88. https://doi.org/10.1002/2017SW001669.
Mortimer, C.H., 1942. The exchange of dissolved substances between mud and water in lakes. J. Ecol. 30 (1), 147–201. https://doi.org/10.2307/2256691.
Moses, W.J., Gitelson, A.A., Berdnikov, S., Povazhnyy, V., 2009. Satellite estimation of chlorophyll-a concentration using the red and NIR bands of MERISThe azov sea case study. IEEE Geosci. Remote Sensing Lett. 6 (4), 845–849. https://doi.org/10.1109/LGRS.2009.2026657.
Neil, C., Spyrakos, E., Hunter, P.D., Tyler, A.N., 2019. A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. Remote Sens. Env. 229, 159–178. https://doi.org/10.1016/J.RSE.2019.04.027.
O'Reilly, J.E., 2000. SeaWiFS postlaunch calibration and validation analyses, part 3. NASA Center for AeroSpace Information.
Paerl, H.W., Huisman, J., 2008. Climate: blooms like it hot. Science 320 (5872), 57–58. https://doi.org/10.1126/SCIENCE.1155398.
Pahlevan, N., Mangin, A., Balasubramanian, S.V., Smith, B., Alikas, K., Arai, K., Barbosa, C., Bélanger, S., Binding, C., Bresciani, M., Giardino, C., Gurlin, D., Fan, Y., Harmel, T., Hunter, P., Ishikaza, J., Kratzer, S., Lehmann, M.K., Ligi, M., Warren, M., 2021. ACIX-Aqua: a Global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. Remote Sens Env. 258, 112366. https://doi.org/10.1016/J.RSE.2021.112366.
Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M.K., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: a machine-learning approach. Remote Sens. Env. 240. https://doi.org/10.1016/j.rse.2019.111604.
Papadopoulos, H., 2008. Inductive conformal prediction: theory and application to neural networks. Tools In Artificial Intelligence. Citeseer.
Ramon, D., Steinmetz, F., Deschamps, P.Y., 2011. Atmospheric correction in presence of sun glint: application to MERIS. Opt. Express. 19 (10), 9783–9800. https://doi.org/10.1364/OE.19.009783. Vol. 19, Issue 10, Pp. 9783-9800.
Rice, E. W., Baird, R. B., & Eaton, A. D. (2017). Standard Methods for the examination of water and Wastewater, 23rd edition.
Riddick, C.A.L., Hunter, P.D., Domínguez Gómez, J.A., Martinez-Vicente, V., Présing, M., Horváth, H., Kovács, A.W., Vörös, L., Zsigmond, E., Tyler, A.N., 2019. Optimal cyanobacterial pigment retrieval from ocean colour sensors in a highly turbid, optically complex lake. Remote Sens. 11 (13), 1613. https://doi.org/10.3390/RS11131613, 2019, Vol. 11, Page 1613.
Royal Meteorological Society. (2018, January 29). Review of UK weather for 2017. https://www.rmets.org/metmatters/review-uk-weather-2017.
Santer, R., Schmechtig, C., 2000. Adjacency effects on water surfaces: primary scattering approximation and sensitivity study. Appl. Opt. 39 (3), 361–375. https://doi.org/10.1364/AO.39.000361. Vol. 39, Issue 3, Pp. 361-375.
Saranathan, A.M., Smith, B., Pahlevan, N., 2023. Per-pixel uncertainty quantification and reporting for satellite-derived chlorophyll-a estimates via mixture density networks. IEEe Trans. Geosci. Remote Sens. 61. https://doi.org/10.1109/TGRS.2023.3234465.
Saranathan, A.M., Werther, M., Balasubramanian, S.V., Odermatt, D., Pahlevan, N., 2024. Assessment of advanced neural networks for the dual estimation of water quality indicators and their uncertainties. Front. Remote Sens. 5, 1383147. https://doi.org/10.3389/FRSEN.2024.1383147/BIBTEX.
Scheffer, M., Van Nes, E.H., 2007. Shallow lakes theory revisited: various alternative regimes driven by climate, nutrients, depth and lake size. Hydrobiologia 584 (1), 455–466. https://doi.org/10.1007/S10750-007-0616-7/FIG.S/4.
Singh, G., Moncrieff, G., Venter, Z., Cawse-Nicholson, K., Slingsby, J., Robinson, T. B., & Author, C. (2024). Uncertainty quantification for probabilistic machine learning in earth observation using conformal prediction.
Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., López-Puigdollers, D., Louis, J., Lubej, M., Mateo-García, G., Osman, J., Peressutti, D., Pflug, B., Puc, J., Richter, R., Žust, L., 2022. Cloud mask intercomparison eXercise (CMIX): an evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. Remote Sens Env. 274, 112990. https://doi.org/10.1016/J.RSE.2022.112990.
Soares, L.M.V., Calijuri, M.C., 2021. Sensitivity and identifiability analyses of parameters for water quality modeling of subtropical reservoirs. Ecol. Modell. 458, 109720. https://doi.org/10.1016/J.ECOLMODEL.2021.109720.
Spears, B.M., Carvalho, L., Perkins, R., Kirika, A., Paterson, D.M., 2012. Long-term variation and regulation of internal phosphorus loading in Loch Leven. Hydrobiologia 681 (1), 23–33. https://doi.org/10.1007/s10750-011-0921-z.
Spyrakos, E., O'donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis, S.G.H., Neil, C., Barbosa, C.C.F., Binding, C.E., Bradt, S., Bresciani, M., Dall'olmo, G., Giardino, C., Gitelson, A.A., Kutser, T., Li, L., Matsushita, B., Martinez-Vicente, V., Matthews, M. W., Tyler, A.N., 2017. Optical types of inland and coastal waters. Limnol. Ocean. https://doi.org/10.1002/lno.10674.
Talling, J.F., 1974. Photosynthetic pigments: general outline of spectrophotometric methods; specific procedures. In: Vollenweider, R.A. (Ed.), A Manual on Methods of Measuring Primary Production in Aquatic Environments. Blackwell, pp. 22–26.
Taylor, P.J., Dobel, A., Grant, A., Gunn, I.D.M., Olszewska, J.P., May, L., 2022a. Loch Leven Monitoring Data (Water Quality, Chemistry, Crustacean Zooplankton), 2020. NERC EDS Environmental Information Data Centre. https://doi.org/10.5285/2ce166c7-bff2-419b-a17f-0a5d53896416.

Taylor, P.J., Dobel, A., Grant, A., Gunn, I.D.M., Olszewska, J.P., May, L., 2022b. Loch Leven monitoring data (water quality, chemistry, crustacean zooplankton), 2021. NERC EDS Environmental Information Data Centre. https://doi.org/10.5285/f5dc095c-e3f6-4d9e-be80-40e35c298142.

Taylor, P.J., Dobel, A., Grant, A., Gunn, I.D., Olszewska, J., May, L., 2023. Loch Leven Monitoring Data (Water Quality, Chemistry, Crustacean Zooplankton), 2022. NERC EDS Environmental Information Data Centre. https://doi.org/10.5285/6dd7ac13-c3e3-4156-a0d8-929227dbf9e8.

Taylor, P.J., Dobel, A., Gunn, I.D.M., Olszewska, J.P., May, L., 2021. Loch Leven Monitoring Data (Water Quality, Chemistry, Crustacean Zooplankton), 2019. NERC Environmental Information Data Centre. https://doi.org/10.5285/e404f64c-ddbc-4e3e-8dca-9bea3d68959a.

Vanhellemont, Q., & Ruddick, K. (2016). Acolite for Sentinel-2: aquatic applications of MSI imagery. In ESA special publication SP. https://odnature.naturalsciences.be/remsem/acolite-.

Vovk, V., Gammerman, A., Shafer, G., 2005. Algorithmic learning in a random world, 29. Springer.

Waczak, J., Aker, A., Wijeratne, L.O.H., Talebi, S., Fernando, A., Dewage, P.M.H., Iqbal, M., Lary, M., Schaefer, D., Lary, D.J., 2024. Characterizing water composition with an autonomous robotic team employing comprehensive In situ sensing, hyperspectral imaging, machine learning, and conformal prediction. Remote Sens. 16 (6), 996. https://doi.org/10.3390/RS16060996/S1.

Ward, N.K., Steele, B.G., Weathers, K.C., Cottingham, K.L., Ewing, H.A., Hanson, P.C., Carey, C.C., 2020. Differential responses of maximum versus median chlorophyll-a to air temperature and nutrient loads in an oligotrophic lake over 31 years. Water. Resour. Res. 56 (7), e2020WR027296. https://doi.org/10.1029/2020WR027296.

Warren, M.A., Simis, S.G.H., Martinez-Vicente, V., Poser, K., Bresciani, M., Alikas, K., Spyrakos, E., Giardino, C., Ansper, A., 2019. Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. Remote Sens Env. 225, 267–289. https://doi.org/10.1016/J.RSE.2019.03.018.

Werther, M., Burggraaff, O., 2023. Dive into the unknown: embracing uncertainty to advance aquatic remote sensing. J. Remote Sens. 3. https://doi.org/10.34133/REMOTESENSING.0070/ASSET/287DF95B-D1C2-43BA-A104-E675BB933CE4/ASSETS/GRAPHIC/REMOTESENSING.0070.FIG.001.JPG.

Werther, M., Odermatt, D., Simis, S.G.H., Gurlin, D., Jorge, D.S.F., Loisel, H., Hunter, P. D., Tyler, A.N., Spyrakos, E., 2022a. Characterising retrieval uncertainty of chlorophyll-a algorithms in oligotrophic and mesotrophic lakes and reservoirs. ISPRS. J. Photogramm. Remote Sens. 190, 279–300.

Werther, M., Odermatt, D., Simis, S.G.H., Gurlin, D., Lehmann, M.K., Kutser, T., Gupana, R., Varley, A., Hunter, P.D., Tyler, A.N., 2022b. A Bayesian approach for remote sensing of chlorophyll-a and associated retrieval uncertainty in oligotrophic and mesotrophic lakes. Remote Sens Env. 283, 113295.

World Health Organization. (2021). Guidelines on recreational water quality. Volume 1: coastal and fresh waters. http://apps.who.int/bookorders.

Xi, H., Losa, S.N., Mangin, A., Garnesson, P., Bretagnon, M., Demaria, J., Soppa, M.A., Hembise Fanton d'Andon, O., Bracher, A., 2021. Global Chlorophyll a concentrations of phytoplankton functional types with detailed uncertainty assessment using multisensor ocean color and sea surface temperature satellite products. J. Geophys. Res. 126 (5), e2020JC017127. https://doi.org/10.1029/2020JC017127.

Xu, C.Y. (2021). Issues influencing accuracy of hydrological modeling in a changing environment. https://doi.org/10.1016/j.wse.2021.06.005.

Yao, J., Pan, W., Ghosh, S., & Doshi-Velez, F. (2019). Quality of uncertainty quantification for Bayesian Neural network inference. https://arxiv.org/abs/1906.09686v1.

Zibordi, G., Melin, F., Berthon, J.F., 2018. A regional assessment of OLCI data products. IEEe Geosci. Remote Sens. Lett. 15 (10), 1490–1494. https://doi.org/10.1109/LGRS.2018.2849329.