Remote Sensing of Environment 283 (2022) 113295

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A Bayesian approach for remote sensing of chlorophyll-*a* and associated retrieval uncertainty in oligotrophic and mesotrophic lakes

Mortimer Werther [a,b,*], Daniel Odermatt [b], Stefan G.H. Simis [c], Daniela Gurlin [d], Moritz K. Lehmann [e], Tiit Kutser [f], Remika Gupana [b], Adam Varley [a], Peter D. Hunter [a], Andrew N. Tyler [a], Evangelos Spyrakos [a]

[a] Earth and Planetary Observation Sciences (EPOS), Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling, United Kingdom
[b] Swiss Federal Institute of Aquatic Science and Technology, Department of Surface Waters - Research and Management, Dübendorf, Switzerland
[c] Plymouth Marine Laboratory, Plymouth, United Kingdom
[d] Wisconsin Department of Natural Resources, Madison, WI, United States of America
[e] Xerra Earth Observation Institute and the University of Waikato, Hamilton, New Zealand
[f] Estonian Marine Institute, University of Tartu, Tallinn, Estonia

ARTICLE INFO

Edited by Dr. Menghua Wang

Keywords:
Chlorophyll-*a*
Lakes
Uncertainty
Bayesian machine learning
Remote sensing

ABSTRACT

Satellite remote sensing of chlorophyll-*a* concentration (chla) in oligotrophic and mesotrophic lakes faces uncertainties from sources such as atmospheric correction, complex inherent optical property compositions, and imperfect algorithmic retrieval. To improve chla estimation in oligo- and mesotrophic lakes, we developed Bayesian probabilistic neural networks (BNNs) for the Sentinel-3 Ocean and Land Cover Instrument (OLCI) and Sentinel-2 MultiSpectral Imager (MSI). The BNNs were built using an *in situ* dataset of oligo- and mesotrophic water bodies (1755 observations from 178 systems; median chla: 5.11 mg m^{-3}, standard deviation: 10.76 mg m^{-3}) and provide a per-pixel uncertainty percentage associated with retrieved chla. Shifts of oligo- and mesotrophic systems into the eutrophic regime, characterised by higher biomass levels, are widespread. To account for phytoplankton biomass fluctuation, a set of eutrophic lakes (167 observations from 31 systems) were included in this study (maximum chla 68 mg m^{-3}). The BNNs were evaluated through five assessments including single day and time series match-ups with OLCI and MSI. OLCI BNN accuracy gains of >25% and MSI BNN accuracy gains of >15% were achieved in the assessments when compared to chla reference algorithms for oligotrophic waters (chla $\leq$ 8 mg m^{-3}). In comparison to the reference algorithms, the accuracy gains of the BNNs decreased as chla and trophic levels increased. To measure the quality of the provided BNN uncertainty estimate, we calculated the prediction interval coverage probability (PICP), Sharpness and mean absolute calibration difference (MACD) metrics. The associated BNN chla uncertainty estimate included the reference *in situ* chla values for most observations (PICP $\geq$ 75%) across the different performance assessments. Further analysis showed that the BNN chla uncertainty estimate was not constantly well-calibrated across different evaluation strategies (Sharpness 1.7–6, MACD 0.04–0.25). BNN uncertainties were used to test two chla improvement strategies: 1) identifying and filtering uncertain chla estimates using scene-specific thresholds, and 2) selecting the most accurate prior atmospheric correction algorithm per individual satellite observation to retain chla with the lowest BNN uncertainty. Both strategies increased the quality of the chla result and demonstrated the significance of uncertainty estimation. This study serves as research on Bayesian machine learning for the estimation and visualisation of chla and associated retrieval uncertainty to develop harmonised products across OLCI and MSI for small and large oligo- and mesotrophic lakes.

* Corresponding author at: Swiss Federal Institute of Aquatic Science and Technology, Department of Surface Waters - Research and Management, Dübendorf, Switzerland.
E-mail address: mortimer.werther@eawag.ch (M. Werther).

https://doi.org/10.1016/j.rse.2022.113295
Received 16 May 2022; Received in revised form 21 September 2022; Accepted 25 September 2022

1. Introduction

The phytoplankton pigment chlorophyll-*a* concentration (chla) is widely used as a proxy of phytoplankton biomass and net primary production (Carlson, 1977; Huot et al., 2007; Poikane et al., 2010; Vörös and Padisák, 1991). Chla is thus an important indicator for the ecological integrity of aquatic ecosystems (Boyer et al., 2009). Optical sensors such as the MultiSpectral Imager (MSI) and Ocean and Land Cover Instrument (OLCI) aboard the Sentinel-2 and 3 satellites, respectively, are used to estimate chla and other optically active constituents (OACs) in lakes (Odermatt et al., 2018; Pahlevan et al., 2020; Toming et al., 2016). Oligotrophic and mesotrophic lakes have low to moderate levels of phytoplankton biomass (chla $\leq$ 25 mg m^{-3}) and provide ecosystem services critical to human and animal welfare, such as drinking water, biodiversity, sediment retention and processing, fishery, and climate change mitigation (Grizzetti et al., 2016; Sterner et al., 2020).

Spectral remote sensing reflectance ($R_{rs}(\lambda)$) contains information about water column properties and components such as phytoplankton (Mobley, 1999; O'Reilly et al., 1998). Many chla estimation algorithms relate optical features in the red and near-infrared (NIR) areas of the reflectance spectrum to chla (Odermatt et al., 2012). For example, the contribution from phytoplankton backscattering to the reflectance spectrum forms distinct reflectance peaks near 560 and 700 nm (Suits, 1975). Further, a phytoplankton pigment feature at 675 nm can be offset against the NIR scattering peak around 700 nm (Gitelson, 1992; Mittenzwey et al., 1992). A variety of two, three and four-band combination algorithms were developed over the last decades based on the 700 and 665 nm band ratio to capture the red absorption signature of chla (Dall'Olmo et al., 2003; Gilerson et al., 2010; Gons, 1999; Gurlin et al., 2011; Mishra and Mishra, 2012; Moses et al., 2009; Pepe et al., 2001). Chla absorption and fluorescence peaks in the 665–685 nm region are also at the basis of line height algorithms such as the fluorescence line height (FLH) (Gower, 1980; Gower et al., 2004; Gupana et al., 2021), maximum peak height (MPH) (Matthews et al., 2012; Matthews and Odermatt, 2015) and maximum chlorophyll index (MCI) (Binding et al., 2013; Gower et al., 2005).

In recent decades, the majority of inland water quality studies have concentrated on eutrophic water bodies (Clark et al., 2017; Coffer et al., 2021; Dekker and Peters, 1993; Simis et al., 2005; Urquhart et al., 2017). In comparison, oligo- and mesotrophic water bodies are under-sampled globally (Filazzola et al., 2020), limiting the development and validation of designated chla estimation methods. Furthermore, recent research indicates that chla estimation in oligo- and mesotrophic systems is associated with higher uncertainties than in eutrophic conditions (Liu et al., 2021; Mouw et al., 2013; Neil et al., 2019; Werther et al., 2022). The causes of estimation uncertainty are manifold. Absorption by coloured dissolved organic matter ($a_{CDOM}(\lambda)$) and non-pigmented particulate absorption ($a_{NAP}(\lambda)$) may be larger than by phytoplankton pigments ($a_{\phi}(\lambda)$), particularly at short wavelengths where band arithmetic algorithm specificity then breaks down (Härmä et al., 2001; Kutser et al., 2016; Mouw et al., 2013; Neil et al., 2019; Schalles, 2006). For analytical algorithms that invert a reflectance spectrum into inherent optical properties (IOPs), $a_{\phi}(\lambda)$ must be precisely estimated to relate it to chla (Werdell et al., 2018). Effects such as pigment packaging impact the linearity of the relationship between chla and $a_{\phi}(\lambda)$(Grzymski et al., 1997; Kirk, 1994; McKee et al., 2014). The scaling of $a_{\phi}(\lambda)$ to chla further depends on phytoplankton type and particle size (Johnsen et al., 1994; Lutz, 2001; Simis et al., 2007). Therefore, for analytical algorithm application to estimate chla, prior information about $a_{\phi}(\lambda)$ variability is required. Without system-specific calibration, the underlying bio-optical models use default pigment mass-specific phytoplankton absorption ($a_{\phi}^{*}(\lambda)$) values. $a_{\phi}^{*}(\lambda)$ natural variability spans an order of 4 magnitudes in the ocean (Bricaud et al., 1995, 1998, 2004), and its variability is unknown for global lakes, yet subject to measurement uncertainty (McKee et al., 2014). For algorithms that focus on the red/NIR spectral area, the phytoplankton absorption peak at 675 nm in lakes may be undetectable in conditions where CDOM and NAP absorption is high (Kutser et al., 2016). For oligotrophic lakes, the red-NIR area of the spectrum may have low signal to noise ratios, causing chla-related peaks to be below the detection limit.

In recent years, machine learning (ML) methods have been developed to estimate chla in inland water bodies. Neural networks (NNs) developed for lakes overcome some of the issues encountered with band ratio methods and analytical approaches when assessed over large chla ranges (Pahlevan et al., 2020, 2022; Schaeffer et al., 2022; Smith et al., 2021). For a ML approach to accurately handle unseen observations, large training datasets of *in situ* and/or simulated measurements are necessary. Because the spatial and temporal scope of the training set is frequently constrained, ML techniques are prone to produce estimation errors when a so-called dataset shift happens. A dataset shift occurs when the IOPs and OACs of a particular, unknown input deviate considerably from the measurements utilised during ML model training (Moreno-Torres et al., 2012; Ovadia et al., 2019). In lake remote sensing, chla estimation uncertainty generated by a dataset shift or other factors such as preceding AC and surrounding land effects on the radiance distribution are typical. Current algorithms designed for lakes, however, do not account for the uncertainty associated with estimated chla. Because chla estimation uncertainty is larger in oligo- and mesotrophic lakes than in eutrophic lakes (Neil et al., 2019), a provision of confidence is a desirable feature to enable the detection and handling of questionable chla estimates.

Bayesian probability theory offers a mathematical tool to reason about uncertainty (Ghahramani, 2015). Bayesian probabilistic reasoning applied to NNs usually comes at a prohibitive computational cost. A NN with distributions placed over the weights has long been studied as a Bayesian Neural Network (BNN) (MacKay, 1992; Neal, 1996) commonly through variational inference (VI) (Jordan et al., 1999; Saul et al., 1996), but with limited success (Gal and Ghahramani, 2016; Graves, 2011). An approach based on VI brings higher model complexity, which for the field of aquatic remote sensing is prohibitive: to represent uncertainty, the number of parameters in BNNs based on VI is doubled when compared to the same non-Bayesian NN size. To train a network with millions of parameters large input datasets are required. However, for inland aquatic remote sensing large datasets are sparse and only just mature through community-wide efforts to collate datasets ready for use with recent ML approaches (Cao et al., 2020; Pahlevan et al., 2022; Werther et al., 2021).

Here, we investigate a potential solution in BNNs developed through Monte Carlo dropout (Gal and Ghahramani, 2016) to obtain chla and a well-calibrated uncertainty estimate from a limited set of input training samples. Monte Carlo (MC) dropout refers to the process of randomly turning off all outgoing connections from a neuron in a NN through a previously determined probability p (Srivastava et al., 2014). Dropout can be activated during the application to unknown observations, which results in an ensemble of NNs to estimate a final chla value: for a single input observation, MC dropout generates a previously defined number (S) of NN variants, each with a different set of activated neurons during prediction, producing S unique chla estimates. The S estimates form a predictive distribution which is then averaged to obtain a final chla value.

Based on the largest available *in situ* dataset covering typical oligo- and mesotrophic lake optical properties, we aim to: 1) improve the chla estimation accuracy *via* OLCI and MSI sensors over both small and large oligo- and mesotrophic lakes through Bayesian probabilistic machine learning, 2) provide a well-calibrated, per-pixel uncertainty percentage associated with estimated chla, and 3) demonstrate how the provided uncertainty can be used to deal with dataset shifts and AC errors commonly affecting satellite remote sensing of chla.

2. Datasets

2.1. Development dataset

A dataset of 1755 *in situ* observations from 178 lakes and reservoirs was compiled from six sources to develop and evaluate the BNNs: (i) LIMNADES (Lake Bio-optical Measurements and Matchup Data for Remote Sensing (https://limnades.stir.ac.uk/)), (ii) Wisconsin DNR (U.S.), (iii)

Earth and Planetary Observation Sciences (EPOS) (Scotland, UK), (iv) the University of Tartu (Estonia), (v) the University of Waikato (New Zealand (NZ)) and the (vi) LéXPLORE Platform (Lake Geneva, Switzerland). Only samples containing above-water $R_{rs}(\lambda)$ and chla measurements were considered for this study. Where available, total suspended matter (TSM) dry weight and the absorption by CDOM at 443 nm ($a_{CDOM}(443)$) were included in the dataset. For algorithm development and validation we separated the measurements from these datasets into 10 regions (Fig. 1) to evaluate the BNNs on independent geographical regions, as previously demonstrated for ML algorithm development (O'Shea et al., 2021; Pahlevan et al., 2022). Details about the regions and included water bodies are provided in Appendix 1. Datasets in LIMNADES were described in Spyrakos et al. (2018) and their original publications (Binding et al., 2008, 2010; Bradt, 2012; Bresciani et al., 2011; Dall'Olmo et al., 2005; Giardino et al., 2005, 2013, 2014, 2015; Gitelson et al., 2007, 2008; Gons et al., 2008; Guanter et al., 2010; Li et al., 2013, 2015; Manzo et al., 2015; Matsushita et al., 2015; Matthews and Bernard, 2013; Moore et al., 2014; Ruiz-Verdú et al., 2005, 2008; Schalles and Hladik, 2012; Wang et al., 2018; Yacobi et al., 2011). Measurements from Wisconsin DNR were taken across the U.S. State of Wisconsin between spring 2014 and autumn 2016. The Wisconsin DNR dataset was used in recent algorithm development and inter-comparison studies (Pahlevan et al., 2020, 2021b). Collection methods and details are provided in Werther et al. (2022). Measurements provided by the University of Tartu were previously detailed in Kutser et al. (2013); Soomets et al. (2020). The measurements from lakes in NZ refer collectively to two datasets. The first one was collected by the University of Waikato between 2015 and 2019 and has been previously published (Pahlevan et al., 2020, 2022). The second dataset refers to data collected jointly by the University of Waikato and University of Stirling in 2020 (referred to as NZ 2020). The NZ 2020 dataset is unpublished and described in Appendix 2. Surface chla obtained through calibration of *in vivo* fluorescence measurements and $R_{rs}(\lambda)$ from the LéXPLORE platform (https://lexplore.info/) in Lake Geneva were measured between 2018/10 and 2021/09. Measurement details are provided in Appendix 3.

Optical Water Types (OWTs) of Spyrakos et al. (2018) were used to define the development and application range of the BNNs (Fig. 2). OWT membership was derived through estimation of the spectral angle (Kruse et al., 1993; Liu et al., 2021). For all $R_{rs}(\lambda)$ an OWT membership score was calculated and the OWT with the highest similarity score assigned. Observations assigned to OWT 5 were small ($n = 40$) in this dataset. The corresponding systems were high in CDOM absorption, TSM and low in chla, known to be challenging for retrieval algorithms (Kallio et al., 2015; Kutser et al., 2016; Toming et al., 2016; Werther et al., 2022).

Phytoplankton productivity in lakes is rarely steady and changes with season and external forcings such as wind (Anneville et al., 2004; Rusak et al., 2018). Lakes that are oligotrophic might suddenly become eutrophic and *vice versa* (Gilarranz et al., 2022; Minaudo et al., 2021). To account for changes in phytoplankton biomass, the collected data was focussed on oligo- and mesotrophic lakes, with an expansion to eutrophic systems (chla median 5.11 mg m^{-3}, mean 9.45 mg m^{-3}, standard deviation 10.76 mg m^{-3}) (Fig. 3). 167 observations (9.5%) from 31 systems in the dataset exhibit chla >25 mg m^{-3} with a maximum of 68 mg m^{-3}. In this study, we adopt the OECD (1982) trophic status scheme, which serves as an international standard for assessing the ecological status of lakes (Carvalho et al., 2013). OECD (1982) classified water trophic status as oligotrophic for chla $\leq$ 8 mg m^{-3}, mesotrophic until 25 mg m^{-3}, and eutrophic thereafter. We also refer to oligo- and mesotrophic systems as low - moderate biomass lakes until 25 mg m^{-3}, because eutrophic lakes can quickly reach phytoplankton chla >100 mg m^{-3} (Matthews, 2014; Matthews et al., 2012; Neil et al., 2019; Pahlevan et al., 2020).

2.2. Satellite data processing

Match-ups were generated for the period 2018/10 to 2021/09 between coinciding *in situ* measurements taken on the LéXPLORE platform (Lake Geneva) and satellite observations from Sentinel-3A/B (S3 A/B) OLCI 300 m resolution and Sentinel-2A/B (S2 A/B) MSI 20 m resolution. Surface *in situ* chla from a Thetis profiler was obtained between 9 and 12.30 a.m. (UTC) (see Appendix 3 for details). OLCI 3A and B overpassed the platform between 10 and 11.30 a.m. (UTC). To generate the match-up, the nearest OLCI and MSI sensor overpass within ± 3 hours of the *in situ* Thetis chla measurement was chosen. The S2 A/B overpass

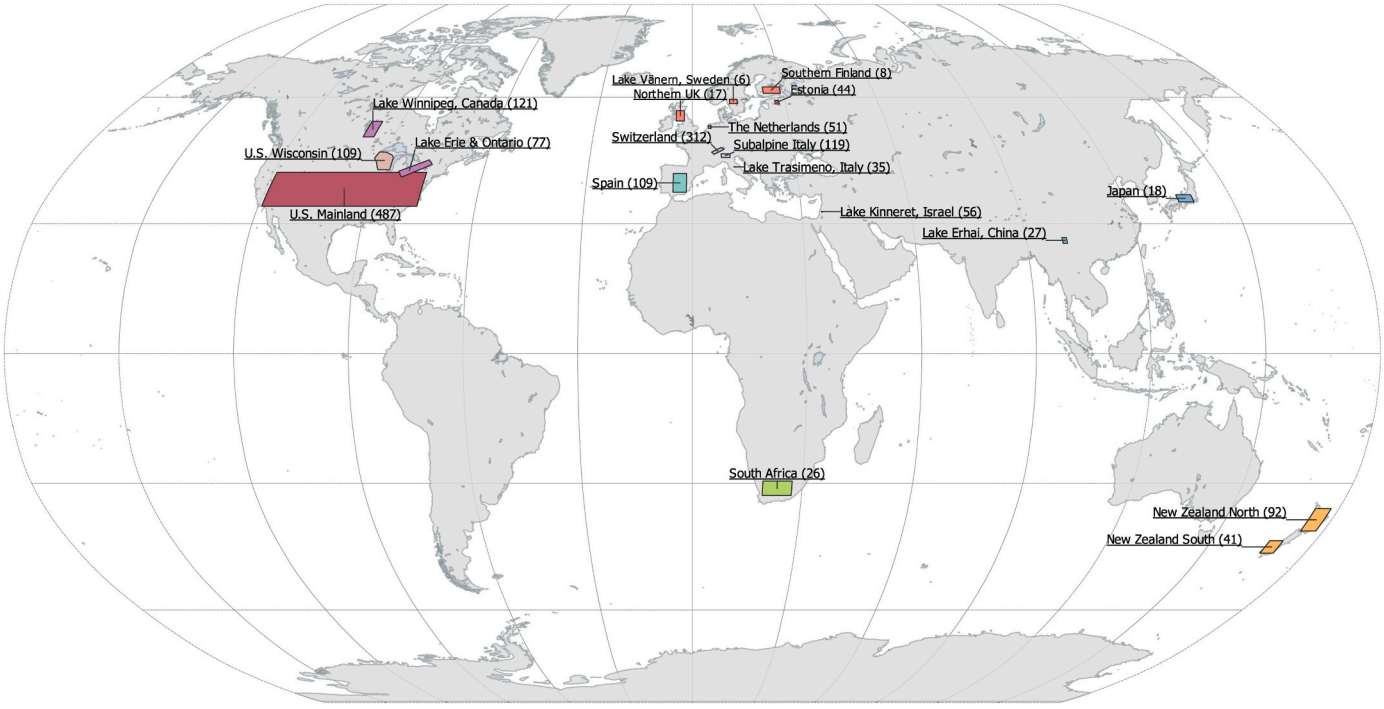

Fig. 1. Spatial distribution of the 10 *in situ* measurement regions constituting the dataset of this study. Grouped regions share the same colour. Number of samples taken in the indicated polygon is shown in brackets. See Appendix 1 for a detailed description of the dataset regions.

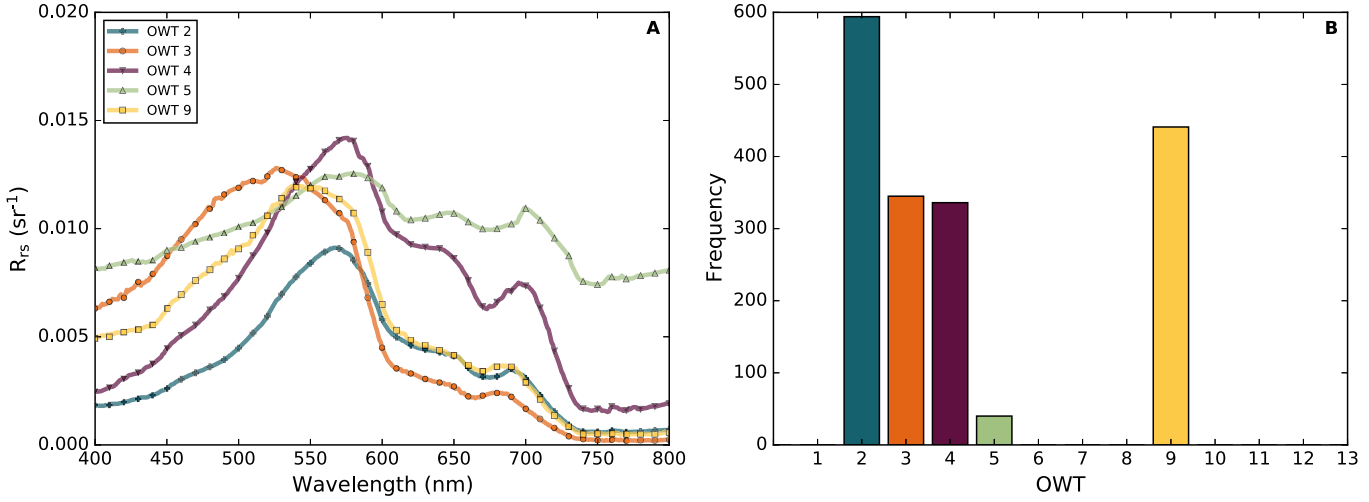

Fig. 2. OWTs 2, 3, 4, 5 and 9 used for the development and application of the BNNs. (A) Spectral medians. (B) Frequency per OWT.

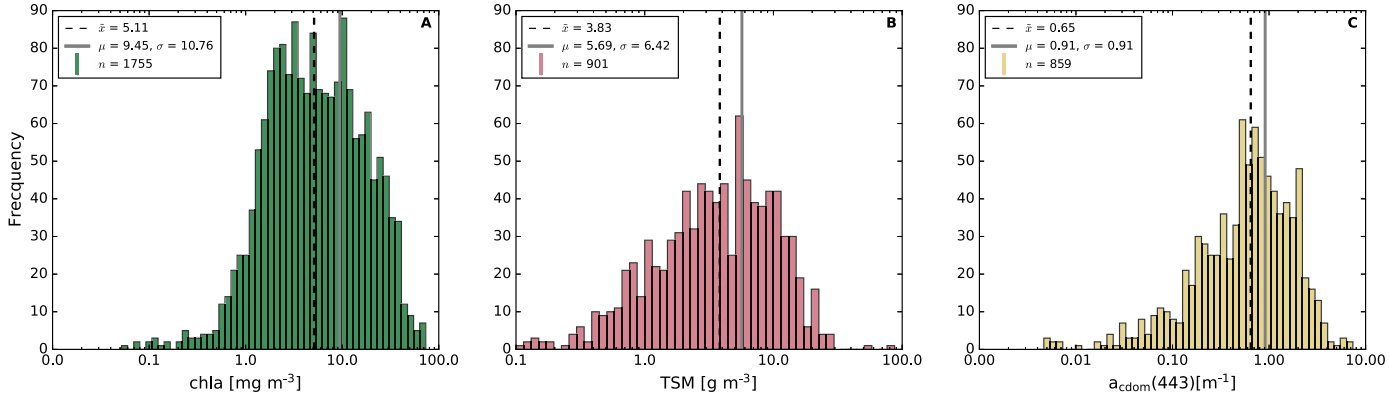

Fig. 3. Log-distributions of (A) chla, (B) TSM and (C) $a_{CDOM}(443)$ for the entire dataset. Denoted are median ($\widehat{x}$), mean (μ) and standard deviation (σ) of the respective parameter.

frequency over LéXPLORE is one-fifth of the S3 A/B OLCI frequency but occurs in the same time frame. For OLCI and MSI (in their respective spatial resolution) valid pixels south of the platform were identified through the IdePix (Identification of Pixel features) algorithm (Skakun et al., 2022; Wevers et al., 2021). The centre pixel of 3 × 3 valid pixels was used as the match-up location. Data flagged by IdePix as invalid, cloud (including ambiguous, sure, and a 2-pixel buffer) or cloud shadow, snow/ice, bright, coastline land, white and glint risk were excluded.

For AC of OLCI and MSI data, the POLYMER (Steinmetz et al., 2011) and Case 2 Region CoastColour (C2RCC) (Brockmann et al., 2016) algorithms were selected. POLYMER and C2RCC AC performances were previously compared for OLCI and MSI and found to perform well for the OWTs included in this study (Liu et al., 2021; Pahlevan et al., 2021a; Warren et al., 2019, 2021). Both ACs are widely used in combination with algorithms for the retrieval of chla in lakes (Kravitz et al., 2020; Liu et al., 2021; O'Shea et al., 2021; Pahlevan et al., 2020, 2022; Pereira-Sandoval et al., 2019; Smith et al., 2021). A comparison of $R_{rs}(\lambda)$ obtained from the LéXPLORE platform *versus* OLCI/MSI POLYMER/C2RCC is provided in Appendix 4.

3. Methodology

3.1. Bayesian Neural Networks

At the heart of all BNNs is the Bayesian theorem:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}, \tag{1}$$

where $p(\theta|D)$ is the posterior, *i.e.*, the probability of a certain value of a neural network parameter θ given the data D, $p(D|\theta)$ the likelihood, $p(\theta)$ the prior and $p(D)$ the quantity (also known as the marginal likelihood) (Bolstad and Curran, 2016). In Bayesian statistics, θ is not defined by one value but has an uncertainty described by a probability distribution $p(\theta)$. This distribution, $p(\theta)$, defines the probability value of each parameter value θ. For a Bayesian model to obtain a predictive chla distribution, the weight distribution of the neural network is averaged:

$$p(y|x_{test}, D) = \sum_i p(y|x_{test}, \theta_i) \bullet p(\theta_i|D), \tag{2}$$

where x_{test} is an input test $R_{rs}(\lambda)$ observation and θ_i denote the weights w_i of the NN. Replacement of NN weights with distributions is computationally costly and requires large amounts of training data, which is scarce in lake remote sensing. In this study, we therefore used Monte Carlo (MC) dropout applied to NN layers (Gal and Ghahramani, 2016). MC dropout replaces each fixed θ_i of a deterministic NN with a binary distribution, whereby either zero or the value θ_c for a NN connection are obtained. Using a dropout chance of 25%, the NNs estimated S-times chla for a single input reflectance spectrum. Each of the S estimates originated from a different NN variant that corresponded to a sampled network constellation. Combining the S dropout estimates resulted in a chla probability distribution:

$$p(y|x_{test}, D) = \frac{1}{S}\sum_{i=1}^{S} p(y|x_{test}, \theta_i), \quad (3)$$

which is an empirical approximation to Eq. 2 that captured both the NN model (epistemic) and the data intrinsic aleatoric uncertainty (Abdar et al., 2021). S is a user-supplied value. The higher S, the more NN dropout variants and chla estimates are generated to form the probability distribution. We chose 50 in this study as a trade-off between required computation time and performance. We then sampled from all determined Gaussian distributions: $y \sim N(\mu_{x,\theta}, \sigma_{x,\theta})$, where μ_θ is the chla mean and σ_θ the standard deviation (Fig. 4). Because σ was a distribution, for each chla estimate y we calculated the 95% confidence interval (CI) and then estimated the width of the CI (CI_w) to get an absolute number. CI_w was divided by two to obtain uncertainty $\pm$ of y and the entire term was multiplied by 100 to obtain a BNN percentage uncertainty for estimated chla:

$$\varepsilon(\%) = \left(\frac{\frac{CI_w}{2}}{y}\right) \bullet 100. \quad (4)$$

Determination of the BNN hyper-parameters, such as the probability that a neuron was dropped, is explained in section 3.3.1.

3.2. BNN processing

In situ $R_{rs}(\lambda)$ were convolved to the relative spectral response (RSR) of the OLCI instruments on Sentinel-3A and 3B and the MSI instruments on Sentinel-2A and B, and were used as inputs to the BNNs. Distinct BNNs were developed for OLCI S3A/B and MSI S2A/B. For the OLCI BNNs we selected the 12 bands from 413 through 778 nm (omitting the oxygen bands at 761, 764 and 767 nm). The waveband centred at 400 nm was excluded because the majority of *in situ* $R_{rs}(\lambda)$, did not extend to <400 nm and therefore could not be convolved to match the OLCI band. For MSI, all 7 bands from 443 to 783 nm were selected.

For oligotrophic waters with chla <5 mg m^{-3}, negative $R_{rs}(\lambda)$ in one or more red-NIR wavebands may occur because of low signal to noise ratios. Negative values in these wavelengths can cause unpredictable behavior of the BNNs that may lead to high chla estimation uncertainty. To enable the BNNs to produce chla estimates even when negative values in red-NIR wavelengths occur, all negative $R_{rs}(\lambda)$ values $\geq$ 665 nm in the *in situ* dataset were set to zero for BNN training and during pre-processing of test observations (including AC-derived $R_{rs}(\lambda)$). $R_{rs}(\lambda)$ features were then normalized to fall between the 0–1 range. Each feature was treated individually such that it was in the given range of the dataset between 0 and 1. Normalisation reduced the required training time and smoothed the process of minimising the loss function of the network, the negative log-likelihood (NLL) in this study. NLL is a standard loss function to measure the quality of a probabilistic model outcome (Hastie et al., 2001).

Since the BNNs are based on MC dropout, a stochastic process is involved. To obtain a representative BNN for the different evaluation strategies (see section 3.3 below), 10 BNNs were trained for each sensor and applied to the respective test observations once. The median of the 10 different networks was used as the final estimate we report, following recent practise of NNs with stochastic elements (O'Shea et al., 2021; Pahlevan et al., 2020, 2022). For the evaluation on *in situ* data, the S3A and S2A BNNs versions were used. For match-ups with MSI 2A/B and OLCI 3A/B the BNN version corresponding to the respective sensor was selected.

3.3. BNN performance assessment

Five different assessments were conducted to evaluate the quality of the chla and uncertainty BNN products under varying conditions: (i) a randomised 50/50 training/test data split, (ii) a leave-one-out exercise (LOO) using *in situ* observations, (iii) BNN OLCI and MSI LéXPLORE match-ups, (iv) a LéXPLORE 3 ½ year time-series (including the match-ups) and lastly (v) single-day product visualisations over both large and small lakes in Europe, New Zealand, Africa and Canada. The BNNs were compared to state-of-the-art chla algorithms. Details of these assessments are provided in the following sections.

3.3.1. 50/50 training/test data split

For an assessment of overall BNN performance on *in situ* data, the development dataset ($n = 1755$) was split into 50% training ($n = 878$) and 50% test ($n = 877$) sets, following recent ML assessment practises (O'Shea et al., 2021; Pahlevan et al., 2020). The observations in the two sets were

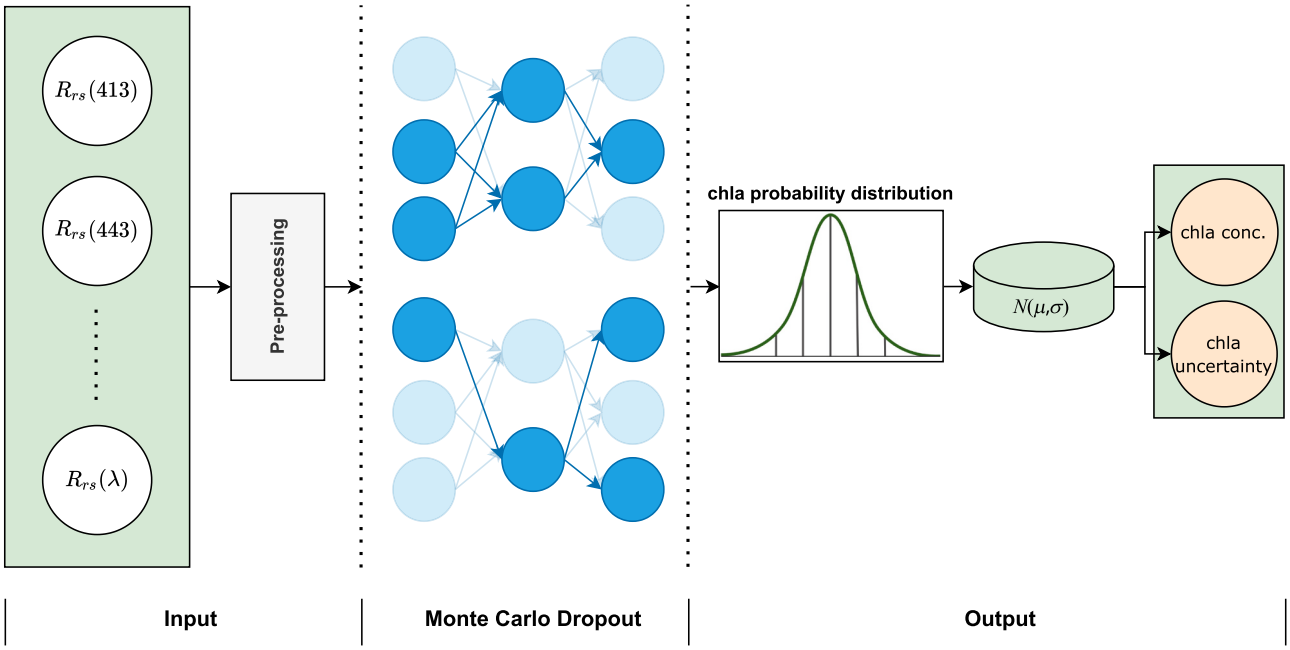

Fig. 4. Processing scheme of the BNNs based on Monte Carlo dropout. For MSI the first band is 443 nm. Pre-processing includes the spectral convolution of the training data, normalisation of both training and unknown (test) observations and treatment of negative $R_{rs}(\lambda)$ values.

randomly drawn from all regions of the entire dataset. The 50% training data were further split into a training (60%; 526 observations) and validation set (40%; 352 observations). An initial OLCI and MSI BNN was constructed using this training subset ($n = 526$), and the performance of these BNN versions was evaluated on the validation set ($n = 352$) to find an optimal architecture and to tune the hyper-parameters of the BNNs through Bayesian optimisation implemented in the Python package Weights & Biases (Biewald, 2020; Werther et al., 2021). BNNs were continually improved by tracking the loss on the validation data. Once the validation loss did not decrease further, the evaluation on the validation set was concluded. The BNNs were then re-trained with the entire training set (878 / 1755 observations) and subsequently applied to the test set using the previously determined architecture and hyper-parameters. The hyper-parameters identified in this 50/50 assessment were used across the other performance assessments to allow comparisons of the same model architecture with different training and test sets. Furthermore, the OLCI and MSI BNNs distributed in conjunction with this article conform to the single hyper-parameter configuration utilised throughout the assessments. The OLCI and MSI BNN performances on the test set across OWTs and TS classes were tabulated.

A randomised 50/50 split of training and test data has two major limitations. First, training and test datasets share the same distribution of OACs (assured through prior randomisation of the overall dataset), which may not occur in a situation outside of model development. Performance under dataset shifts can therefore not be analysed. Second, the performance of the BNN OLCI and MSI versions in this assessment does not represent their final models, as these versions were only trained with 50% of the training data. The final BNN versions made publicly available with this article were re-trained with the entire dataset ($n = 1755$). The final hyper-parameter configuration of the OLCI and MSI BNNs is listed in Appendix 5.

3.3.2. Leave-one-out

We simulated dataset shifts with all available *in situ* measurements through a leave-one-out (LOO) strategy (O'Shea et al., 2021; Pahlevan et al., 2022). In LOO, a BNN was trained with measurements from all regions except for one region. The BNN was then applied to the left-out region and the entire process was repeated until each region was left-out. Through the LOO strategy we assess the generalisation performance of the BNN for independent systems (represented by a region) that were not part of the training process, and which may thus not share the same OAC distribution. We note that the performance of the BNNs for different regions may underestimate overall model performance. Whilst different *in situ* measurement techniques between the datasets were used, the measurement consistency between regions is unknown. Individual regions may thus carry varying degrees of measurement uncertainty.

3.3.3. Lake Geneva match-ups and time series assessments

To evaluate the BNNs with independent satellite match-ups, a training set excluding observations from Lake Geneva was created. OLCI and MSI BNNs were trained with the exact same set of bands and architecture as in the 50/50 training/test split and LOO assessments. The match-up evaluation on Lake Geneva facilitated to assess how well the BNNs trained with a semi-global *in situ* dataset of lake properties transfer to an individual system measured through satellite sensors, thereby representing a common use-case of the newly developed algorithm.

3.3.4. Comparison with other chla algorithms

Using the 50/50 split, LOO and match-up performance assessments we evaluated the accuracy of the OLCI and MSI BNNs *versus* five state-of-the-art algorithms designed for large chla concentration ranges (hereafter reference algorithms). The five algorithms are the OLCI and MSI Mixture Density Networks (MDNs)(Pahlevan et al., 2020), the red/NIR semi-analytical Gons05 (Gons et al., 2005), the red/NIR band ratio G11 (Gurlin et al., 2011), the blue/green band ratio OC3 (O'Reilly et al., 1998; O'Reilly and Werdell, 2019), and the Blend algorithm (Smith et al., 2018) which switches between OCI (Hu et al., 2012) and the red/NIR semi-analytical approach by Gilerson et al. (2010), that blends their estimates using empirically derived thresholds.

To reduce bias resulting from calibration of the reference algorithms on different datasets, the G11, OC3 and Blend model coefficients were optimised (denoted as -opt) through non-linear least squares fitting against the same training datasets used in the evaluation strategies (identical to the observations for training the BNNs). We used the original OLCI and MSI MDN versions that were designed for general applicability to inland waters (including low - moderate biomass lake conditions). For the OLCI and MSI MDNs specifically, in the LOO assessment regions 5 and 9 were excluded from the evaluation because these were already used in full as part of the training set of said algorithms. Their inclusion would thus not represent an independent comparison per region. Gons05 optimisation was not undertaken as it requires measurements of phytoplankton mass-specific absorption at 665 nm ($a_\phi^*(665)$) and backscatter, which the collated dataset did not sufficiently include.

Algorithms using OLCI wavebands were included in their originally published configuration (denoted as -org) and with optimised coefficients (denoted as -opt). For Blend in the OLCI configuration only the OC4 algorithm was optimised, and for Blend in the MSI configuration the OC3 algorithm was used instead of OC4 because the 510 nm band in OC4 is not available on MSI. MSI band positions of Gons05, G11 and OC3 (also used in Blend) were slightly shifted, namely from 490, 708 and 778 nm to 492, 704 and 783 nm, which was successfully demonstrated previously in Warren et al. (2021). Optimised model coefficients are tabulated in Appendix 6. For the Lake Geneva match-ups, all measurements were provided for optimisation, except for the observations measured in Lake Geneva.

3.4. Accuracy and uncertainty calibration metrics

Common performance metrics were calculated to assess the chla retrieval accuracy between algorithm estimated (e) and observed (o) *in situ* chla: the median symmetric accuracy (MdSA, in %), the symmetric signed percentage bias (Bias, in %), the mean absolute difference (MAD) and median absolute percentage difference (MAPD, in %) (Morley et al., 2018; Seegers et al., 2018, 2021; Smith et al., 2021):

$$MdSA : \left(10^{median\left(\left|log_{10}\left(\frac{e}{o}\right)\right|\right)} - 1\right) \times 100. \tag{5}$$

$$Bias : \left(10^{|MR|} - 1\right) \times sign(MR) \times 100, \quad \text{where } median\ ratio\ (MR) = median\left(log_{10}\left(\frac{e}{o}\right)\right). \tag{6}$$

$$MAD : 10^{Z} - 1,\ Z = \left[\frac{\sum_{i=1}^{n}\left|log_{10}(e_i) - log_{10}(o_i)\right|}{N}\right]. \tag{7}$$

$$MAPD : median\left(\left|\frac{e_i - o_i}{o_i}\right|\right) \times 100 \text{ for } i = 1, \ldots, N. \tag{8}$$

BNN uncertainty estimates may not capture the true data distribution (Lakshminarayanan et al., 2017). For example, a 90% confidence interval (CI) may not contain the reference *in situ* chla concentration in 9/10 scenarios. The BNN uncertainty estimate is miscalibrated when the CI does not include the reference *in situ* chla. To assess the quality of the BNN chla uncertainty calibration we calculated three metrics between the BNN estimated (e) and observed (o) *in situ* chla:

1. Prediction interval coverage probability (PICP; denoted ρ). The percentage of observations for which *in situ* chla lies within the 95% confidence interval of the BNN chla estimate:

$$\rho = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{o_n \le e_n^{high}} \bullet \mathbb{1}_{o_n \ge e_n^{low}}, \quad (9)$$

where e_n^{high} is the 97.5% percentile and e_n^{low} is the 2.5% percentile of the BNN estimated chla for an input x_n (Yao et al., 2019). The higher the ρ [%] value, the more *in situ* chla values were covered by the BNN uncertainty estimate.

2. Sharpness (denoted σ). The standard deviation (var) of a chla estimate whose cumulative distribution function (F_N) should be small:

$$\sigma = \sqrt{\frac{1}{N}\sum_{n=1}^{N} var(F_n)^2}. \quad (10)$$

Sharpness represents the average of the BNN estimated chla standard deviations (Tran et al., 2020).

3. Mean absolute calibration difference (MACD; denoted τ). MACD is a statistic to measure calibration relative to an ideal reliability diagram. Reliability diagrams show expected observation accuracy as a function of confidence (Degroot and Fienberg, 1983; Niculescu-Mizil and Caruana, 2005). To compute MACD, chla estimates were sorted and divided into B equally-spaced bins ($B = 100$ in this study) with an approximately equal number of BNN chla estimates in each bin:

$$\tau = \sum_{b=1}^{B} \frac{n_b}{N} \bullet |acc(b) - conf(b)|, \quad (11)$$

where n_b is the number of chla estimates in bin b, N is the total number of observations, and $acc(b)$ and $conf(b)$ are the accuracy and confidence in bin b (Guo et al., 2017; Nixon et al., 2019). The smaller the values of Sharpness and MACD, the more accurate the uncertainty calibration.

4. Results

4.1. 50/50 dataset split

OLCI and MSI BNNs outperformed the reference chla algorithms (MdSA difference > 18%) (Figs. 5, 6). The OLCI BNN was slightly more accurate than its MSI version (1.44% MdSA difference). In the employment of the reference chla algorithms, lower accuracy in the estimation of chla from MSI compared to OLCI was also observed (Fig. 6). For OLCI, optimisation of model coefficients from G11, OC3 and Blend improved the estimates by 35–75% (MdSA) for oligotrophic waters and 15–35% (MdSA) for mesotrophic waters (Table 1). For eutrophic lakes the original algorithm configurations were more accurate, because the dataset of this study used to optimise the algorithms only included a small set of high biomass observations. The coefficients of the original formulations were obtained from larger eutrophic datasets and are thus more applicable. The optimisation process of the coefficients is considered transferable to MSI. Large over- and underestimates of chla were observed for concentrations between 0 and 1 mg m^{-3} across all the algorithms. Only 65 data points (7.4% of 877 test observations) were available in this concentration range. Chla estimates from the algorithms used for comparison became markedly more accurate for concentrations >10 mg m^{-3}. Varying retrieval accuracies became more apparent when the assessment was tabulated per TS class (Table 1) as the OLCI BNN was >45% and the MSI BNN > 38% more accurate for oligotrophic waters (chla 0–8 mg m^{-3}) than the reference algorithms. Red/NIR algorithms started to become highly accurate in mesotrophic waters (chla 8–25 mg m^{-3}) (Table 1). For eutrophic waters (chla > 25 mg m^{-3}) the red/NIR OLCI Gons05 algorithm was most accurate (MdSA 13.98%), but its performance did not transfer to MSI. Across OWTs, the BNNs consistently outperformed the reference algorithms (Table 2). For OWT 5 it is to note that only 13 observations were assigned to it, thus the evaluation is not representative. The low number in this OWT originated from the randomisation of the entire dataset prior to training and evaluation of the algorithms. Uncertainty

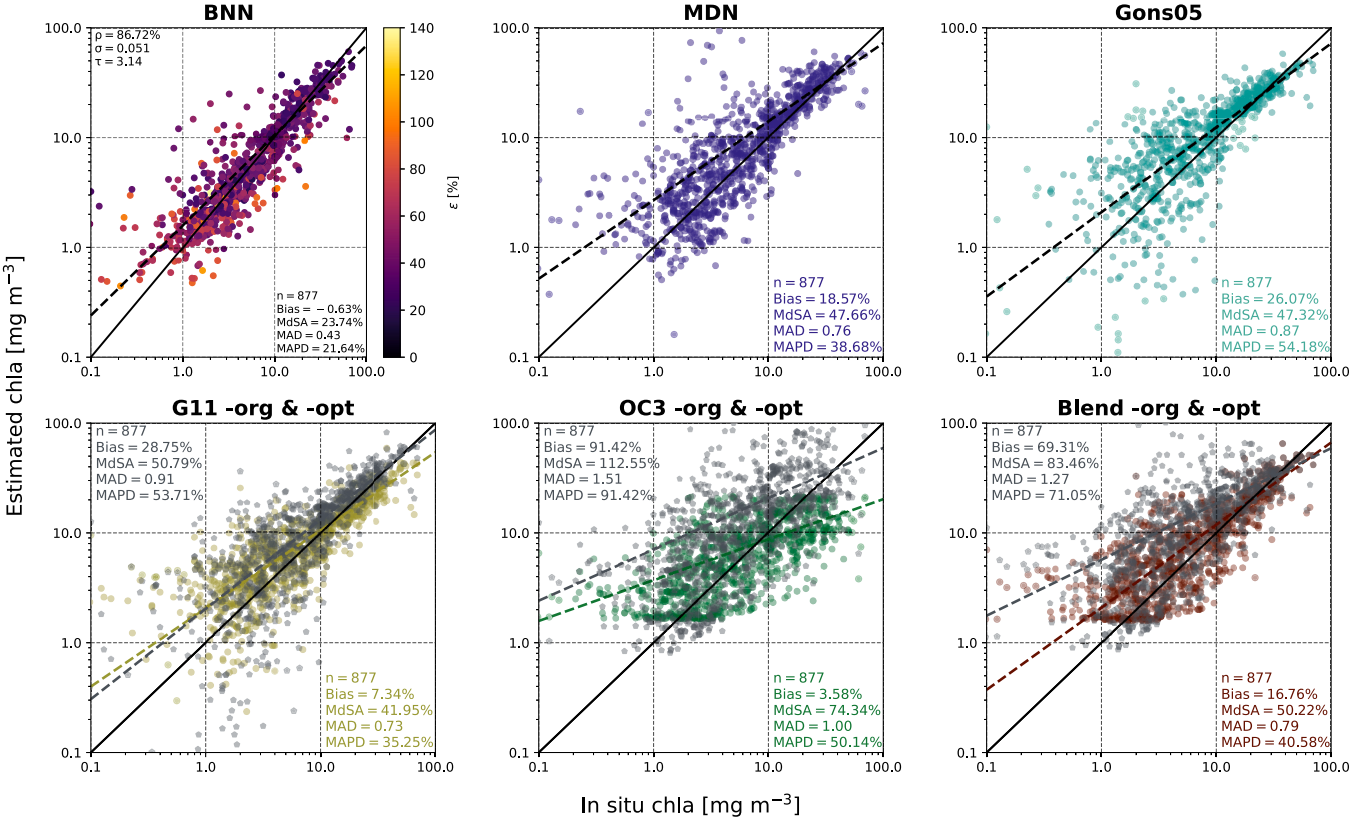

Fig. 5. Chla retrieval results by the OLCI BNN and reference algorithms in the 50/50 training/test split assessment. For G11, OC3 and Blend chla retrievals through original coefficients (−org) are shown in grey, and optimised coefficients (−opt) in colour.

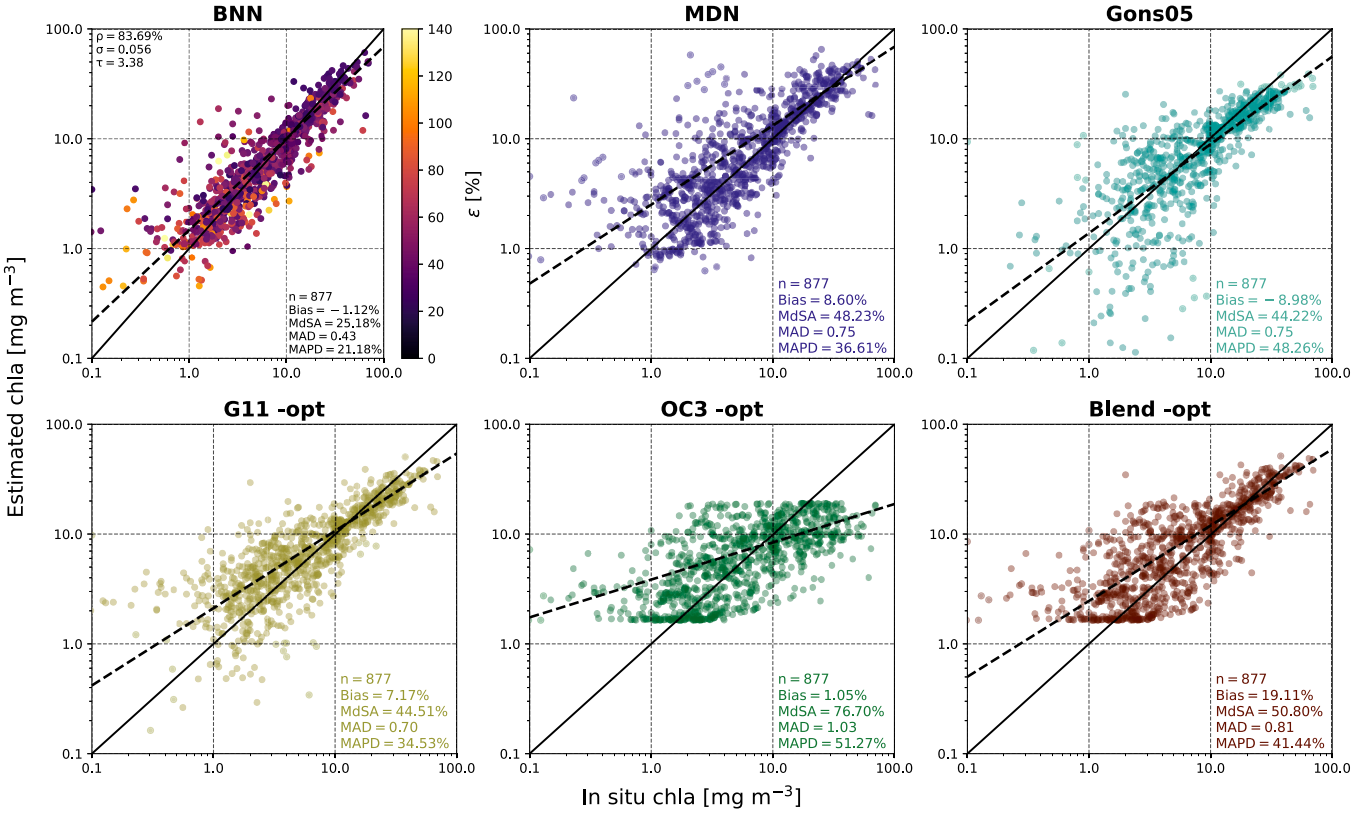

Fig. 6. Chla retrieval results by the MSI BNN and reference algorithms in the 50/50 training/test split assessment.

Table 1
Retrieval accuracies of the chla algorithms per TS class for the test set, compared through the MdSA (in %) metric. Lowest MdSA achieved by an algorithm for each TS in bold.

Algorithm	**Oligotrophic** (chla ≤ 8 [mg m^{-3}]; n = 548)	**Mesotrophic** (chla >8 and ≤ 25 [mg m^{-3}]; n = 253)	**Eutrophic** (chla >25 [mg m^{-3}]; n = 76)
		OLCI	
BNN	**28.78**	**18.95**	17.28
MDN	76.43	25.33	19.37
Gons05	109.33	30.00	**13.98**
G11 -org	110.35	35.08	17.40
G11 -opt	75.09	19.26	34.85
OC3 -org	154.43	80.95	57.31
OC3 -opt	78.95	44.97	191.88
Blend -org	147.98	49.73	16.88
Blend -opt	83.21	24.74	19.84
		MSI	
BNN	**32.42**	19.14	**19.77**
MDN	71.13	26.43	20.94
Gons05	81.57	20.28	46.41
G11 -opt	72.58	**17.52**	35.98
OC3 -opt	79.54	47.20	207.57
Blend -opt	81.97	27.15	21.27

calibration by the OLCI and MSI BNNs was accurate (MACD 0.051 and 0.056, respectively) and PICP included >83% of the *in situ* chla measurements. Sharpness of 3.14 (OLCI) and 3.38 (MSI) across the large concentration range was moderate - low. Highly uncertain chla estimates were identified by both BNN configurations.

4.2. Leave-one-out: BNN generalisation ability

In the LOO assessment, the OLCI BNN MdSA was expectedly higher for some regions (MdSA between 19 and 77%) than for the 50/50 training/test split assessment (Fig. 7). In comparison the OLCI BNN outperformed the reference algorithms in 7/10 regions and remained competitive for the other regions. Highest gains in retrieval accuracy by the OLCI BNN were made for region 4 consisting of the UK, Sweden, Finland and Estonia (MdSA 19% lower), the South African reservoirs representing region 7 (MdSA 18% lower) and Lake Erie, Ontario, and Winnipeg of region 10 (MdSA 40% lower). For most regions, MSI MdSA was only slightly higher than for OLCI. However, the largest difference in retrieval accuracy between the MSI and OLCI BNN configurations was linked to regions 4 and 10, whereby the MSI configuration was about 20% less accurate. The results of the independent region evaluation emphasise that measurements included in this dataset are representative for other regions of the globe. Consequently the dataset enabled a robust

Table 2
Retrieval accuracies of the chla algorithms for OLCI (top) and MSI (bottom) for the test set per OWT class, compared through the MdSA (in %) metric. Lowest MdSA achieved by an algorithm for each OWT in bold.

Chla algorithm	**OWT 2 (n = 261)**	**OWT 3 (n = 257)**	**OWT 4 (n = 149)**	**OWT 5 (n = 13)**	**OWT 9 (n = 197)**
			OLCI		
BNN	**23.21**	**27.73**	**20.82**	42.51	**23.67**
MDN	36.87	64.72	47.53	109.67	44.73
Gons05	37.48	136.90	31.90	60.51	49.07
G11 -org	37.98	148.50	42.58	67.88	48.68
G11 -opt	32.71	82.80	34.47	47.48	39.66
OC3 -org	146.81	74.03	134.55	262.57	135.89
OC3 -opt	65.25	66.49	134.44	443.42	57.21
Blend -org	118.24	64.83	66.59	44.39	96.49
Blend -opt	41.11	69.44	41.31	**24.82**	55.09
			MSI		
BNN	**24.27**	**23.14**	**18.22**	**19.67**	**28.49**
MDN	30.16	67.32	51.76	83.30	41.10
Gons05	32.02	107.92	40.09	63.58	66.14
G11 -opt	29.72	63.43	39.23	36.52	46.04
OC3 -opt	72.57	59.26	135.18	433.24	57.65
Blend -opt	44.06	102.98	36.62	37.20	51.26

generalisation of the BNNs to systems that shared the same OWTs, but that were from regions not included in the respective training dataset.

Through LOO, the effectiveness of the uncertainty calibration can be analysed (Fig. 8). MACD and Sharpness did not co-vary linearly with chla accuracy. For example, OLCI MdSA in regions 2, 5 and 9 was the smallest, but the corresponding uncertainty metrics were not consistently small, too. In region 9, chla accuracy with MdSA <20% was achieved, but MACD (> 0.20) and Sharpness (> 5) were high in comparison to other regions. Nevertheless, region 9 PICP was 100%, indicating that all BNN chla estimates fell within the estimated confidence intervals. Overall, OLCI PICP was on average 7% higher than MSI (85.24% and 78.36%, respectively). Although the MSI BNN was only slightly less accurate in the retrieval of chla, the algorithm was less certain about its estimates resulting in a lower PICP percentage and higher MACD.

4.3. Lake Geneva match-ups

Satellite-derived $R_{rs}(\lambda)$ for *in situ* match-ups over Lake Geneva were generated through POLYMER and C2RCC applied to OLCI and MSI products. POLYMER calculated negative values in several red bands of the spectrum for 21 OLCI and 12 MSI match-ups. To assure a better match-up comparison of the chla algorithm performances, the 21 OLCI and 12 MSI observations obtained through POLYMER were excluded. Further, Gons05 and G11 were excluded because these red/NIR algorithms showed the largest retrieval uncertainties in the 50/50 split and LOO assessments for chla <10 mg m^{-3}. Chla retrievals from OLCI POLYMER and C2RCC were more accurate than for MSI from the evaluated algorithms (see Table 3). OLCI POLYMER OC3 -org and Blend -org configurations were most accurate (MdSA 48.41% and 36.61%, respectively), closely followed by the BNN (MdSA 49.75%). Optimising the OC3 and Blend algorithms coefficients did not improve the chla estimation for the match-ups, unlike in the 50/50 split and LOO assessments. The optimisation of the OC3 and Blend algorithm coefficients systematically increased the underestimation of chla (Bias < −70%). The reason for this difference lays in the large value range used for the algorithm coefficient optimisation, which was naturally higher than the variation in Lake Geneva. Algorithms using OLCI imagery corrected through POLYMER underestimated chla (Bias between −21% for BNN and –175% for MDN). For OLCI C2RCC match-ups, the BNN was four times more accurate than the OC3 and Blend configurations (Fig. 9). The reference algorithms performed worse on OLCI observations compared to the *in situ* 50/50 training/test set split and LOO assessments, which indicates that the reference algorithms were more affected by AC uncertainties than the BNN. Contrary to POLYMER, C2RCC OLCI and MSI BNNs were more balanced (Bias −0.48% and 22.20%, respectively), resulting in less variability of estimated chla (Fig. 10). PICP was high (> 76%) across all four sensor and AC combinations, and MACD values were similar to the previous LOO analysis on independent region data. Sharpness values were lower than in the 50/50 split and LOO assessments, indicating small average standard deviations and thus a better uncertainty calibration. OLCI and MSI MDNs systematically underestimated chla across all four sensor and AC configurations (MdSA >175%, Bias < −175%).

4.4. Lake Geneva time series

Phytoplankton vertical variability is seasonal in Lake Geneva (Minaudo et al., 2021) and a major factor driving optical properties (Nouchi et al., 2018). The impacts of phytoplankton vertical variability and its effect on remote sensing retrievals is still poorly understood. Seasonal variations over the low to moderate chla range (0–8 mg m^{-3}) were accurately estimated by the OLCI BNNs during the four-year period, independent of the prior AC algorithm. The confidence intervals in most cases (PICP >80%) captured the corresponding *in situ* chla value (see Figs. 11, 12). For a spring bloom in 2020, moderate chla (8–12 mg m^{-3}) was not estimated with similar precision when compared to low chla across the four years. The confidence intervals were too small and did not contain most of the high peaks during the bloom. Conversely, *in situ* estimates in August – September 2021 with chla >5 mg m^{-3} were closely matched by the OLCI and MSI BNNs.

The MSI BNNs showed a larger spread in the low to moderate chla range compared to the OLCI BNNs, with wider confidence intervals. Larger and less precise intervals (higher Sharpness) *via* MSI resemble the results of the 50/50 split and the match-up assessments. The uncertainty estimates by the OLCI BNNs were throughout the exercises more accurately calibrated. Across the time series, $R_{rs}(\lambda)$ derived through POLYMER were associated with higher BNN chla uncertainty than through C2RCC. Between April and May 2021, POLYMER estimated negative $R_{rs}(\lambda)$ values in red bands (not shown) which led to higher uncertainty in chla estimates compared to more stable and lower chla estimates following from C2RCC for the same period.

4.5. OLCI and MSI BNN chla and uncertainty products

The BNNs were applied to several OLCI and MSI images over lakes with varying optical properties to assess horizontal consistency and visualise chla estimates alongside retrieval uncertainties. Single-day products over Lake Geneva, the Southern Island in New Zealand (SNZ), Lake Turkana in Kenya and boreal areas of Sweden and Canada were processed. Fig. 13 illustrates OLCI BNN chla through prior C2RCC AC and associated uncertainty products over Lake Geneva from March 2020, November 2020, and August 2021. The maps using the BNN match the time series chla patterns and ranges for Lake Geneva from the oligotrophic states in March 2020 to mesotrophic levels in August 2021 (Soulignac et al., 2018). The retrieval uncertainties ranged between 20 and 60% in Lake Geneva similar to the results of the previous match-up and time series assessments. The results over the LéXPLORE location can thus be considered transferable to other regions of the lake. During March 2020, areas with high uncertainty are observed near the north

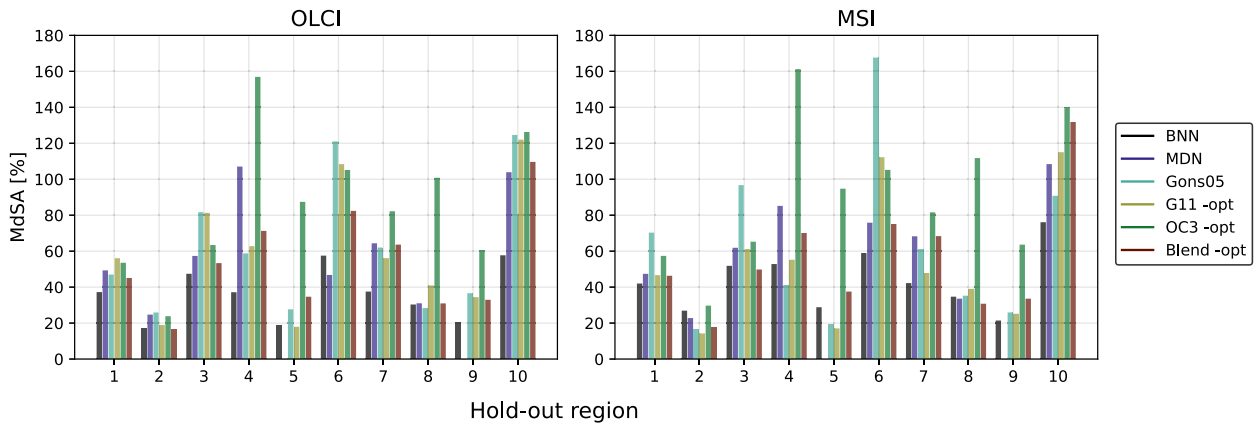

Fig. 7. Change in MdSA of the chla algorithms for each region as part of the LOO assessment.

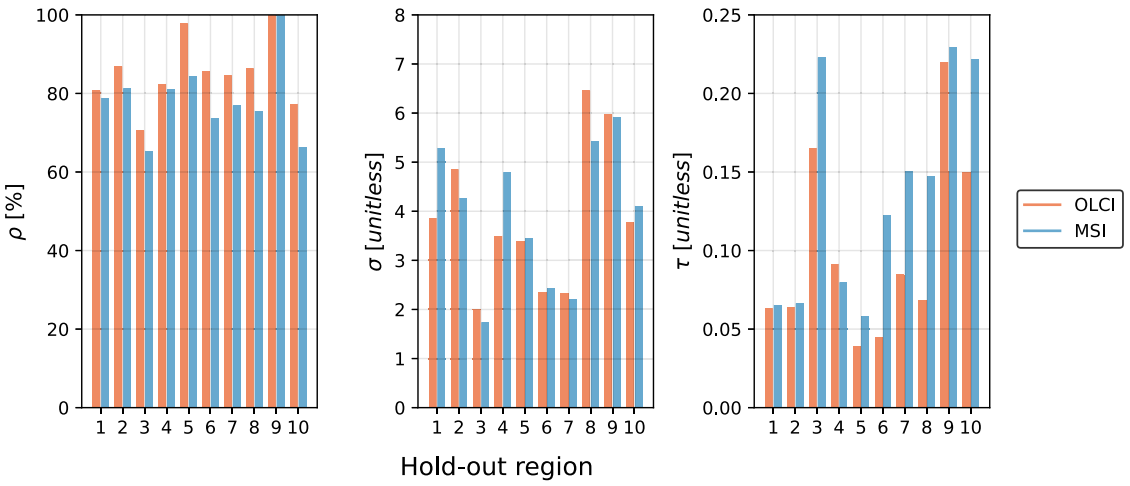

Fig. 8. BNN uncertainty calibration metrics PICP (ρ), Sharpness (σ) and MACD (τ) per region as part of the LOO assessment.

Table 3
Performance metrics of the chla algorithms for the OLCI and MSI match-ups over LéXPLORE. Results for POLYMER are coloured in orange, C2RCC in blue. Lowest MdSA per combination in bold.

OLCI POLYMER / C2RCC

Algorithm	Bias [%]		MdSA [%]		MAD []		MAPD [%]	
BNN	-21.16	-0.48	49.79	**59.07**	0.59	0.71	36.51	44.53
MDN	-175.18	-318.68	175.18	323.23	1.76	3.39	63.66	76.51
OC3 -org	-44.24	-253.43	48.41	253.43	0.53	2.43	34.33	67.70
OC3 -opt	-76.27	-193.06	76.27	193.06	0.83	2.70	43.27	65.88
Blend -org	-23.75	-202.80	**36.61**	209.60	0.54	2.29	28.23	67.70
Blend -opt	-73.40	-226.33	78.22	235.89	0.90	2.95	44.08	70.23

MSI POLYMER / C2RCC

Algorithm	Bias [%]		MdSA [%]		MAD []		MAPD [%]	
BNN	-84.26	22.20	**84.26**	**60.72**	0.75	0.84	45.73	51.48
MDN	-283.02	-634.84	283.02	634.84	2.50	7.60	73.89	87.30
OC3 -org	-87.66	-535.99	93.65	541.88	1.03	8.35	49.61	84.42
OC3 -opt	-137.40	-370.75	137.40	370.75	1.30	5.14	57.88	78.87
Blend -org	-213.79	-1471.02	213.79	1471.02	2.65	27.38	71.87	94.20
Blend -opt	-133.36	-373.17	133.36	381.39	1.24	8.54	57.88	79.23

shore of Lake Geneva, likely associated with contamination of water-leaving radiance by bottom reflectance, adjacent land, or both.

OLCI products processed with C2RCC and POLYMER for SNZ are shown in Fig. 14. BNN chla products varied spatially between the lakes and AC algorithms. Moreover, uncertainties of C2RCC derived-chla were significantly higher (between 50 and 100%) than for POLYMER(Fig. 14). Unlike for Lake Geneva, POLYMER did not produce negative $R_{rs}(\lambda)$ in any wavebands over this area. A surface measurement of *in situ* chla in Lake Hawea (stars on Fig. 14) was 0.89 mg m^{-3} on the 9th of February during the Sentinel-3 overpass, which closely matched the BNN chla of this scene. Overall, highest uncertainties were generated for chla <1 mg m^{-3}, a range for which training data were scarce. The scarcity became apparent in the top left corner of the imagery in the clear coastal waters. The uncertainties for chla <1 mg m^{-3} were similar to the results of the initial 50/50 split analysis. Highest chla retrieval uncertainties from the BNN and the reference algorithms are observed in this concentration range.

The MSI BNN generated high-quality chla products for small water bodies (SWBs). Fig. 15 contains MSI BNN products over SWBs of Canada and Sweden. Satellite-derived $R_{rs}(\lambda)$ from SWBs are prone to adjacency effects which degrade the quality of the retrievals (Paulino et al., 2022; Sterckx et al., 2011). In Fig. 15 D-F, small lakes contained water pixels (white squares) for which the MSI BNN produced high uncertainties. Associated reflectance spectra of small lakes in this scene used to retrieve chla were influenced by the proximity to land, resulting in high uncertainty. The MSI BNN products demonstrate sensitivity to changes in chla between the lakes captured on this scene. Consistent with OLCI BNN retrievals, lake areas depicted in MSI BNN products exhibited higher uncertainties where chla was lowest (Fig. 15 A, B, C). In November 2021 (Fig. 15 G-I) the RGB of "Clearwater Lake" appears extremely bright (see yellow square), which the BNN associated with high chla uncertainty. OLCI BNN products from December 2021 and January 2022 (Fig. 16) cover the largest desert lake in the world, Lake

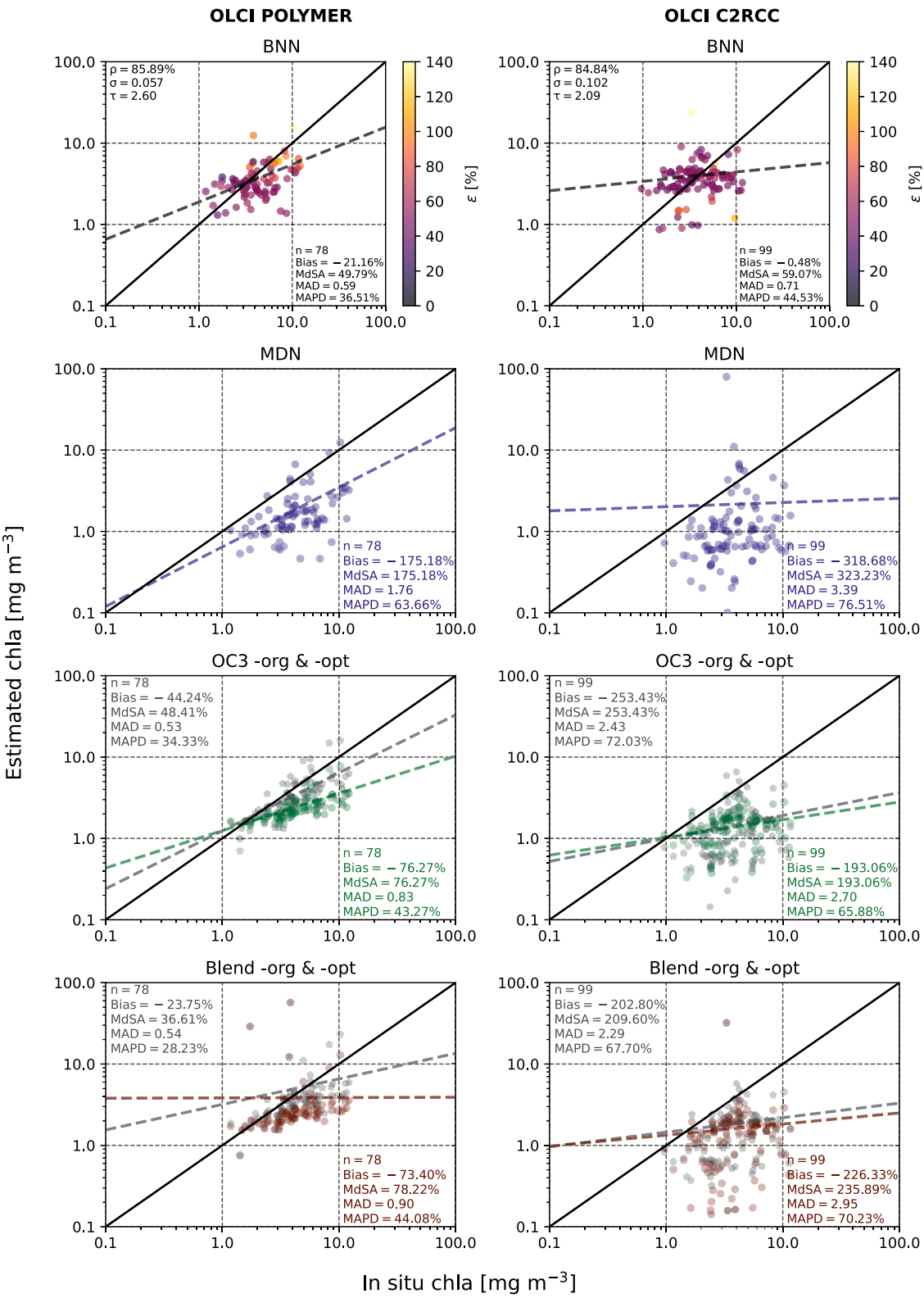

Fig. 9. LéXPLORE platform (Lake Geneva) OLCI match-up chla estimates from BNN, MDN, OC3 -org, OC3 -opt, Blend -org and Blend -opt algorithms. For OC3 and Blend, grey markers indicate the original and coloured markers the optimised algorithm coefficients.

Turkana (Kenya), known for its spatially variable optical properties (Liu et al., 2021; Tebbs et al., 2020). OLCI BNN products over Lake Turkana documented a gradient in chla of the more turbid waters in the north of the lake that decreases towards the lake's southern end. BNN chla following C2RCC (Fig. 16 B, I) was higher (2–6 mg m^{-3}) than through POLYMER (2–3.5 mg m^{-3}; Fig. 16 E, L). Chla uncertainties from C2RCC were approximately 50% lower than by POLYMER for both days.

The product uncertainties were used to filter retrievals and exclude areas of water bodies for which the BNN assigned high uncertainty. $R_{rs}(\lambda)$ from POLYMER caused the BNN to produce unrealistic patches of

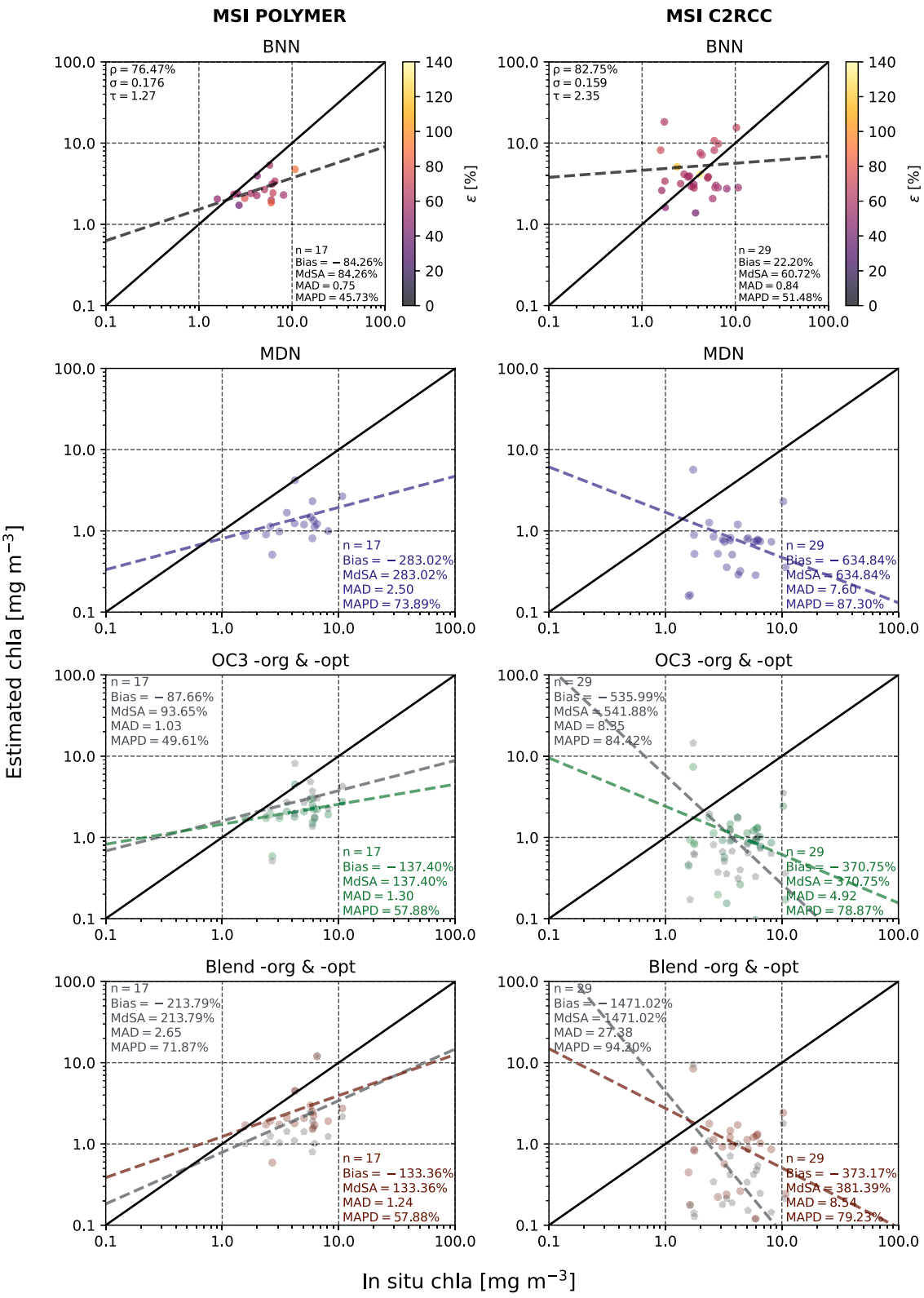

Fig. 10. LéXPLORE platform (Lake Geneva) MSI match-up chla estimates from the BNN, MDN, OC3 -org & -opt and Blend -org and -opt algorithms. For OC3 and Blend, grey markers indicate the original and coloured markers the optimised algorithm coefficients.

chla around 1 mg m^{-3} (Fig. 16 F, M). These patches did not resemble the optical gradient apparent in the area and were not produced through prior C2RCC AC. In contrast, for the Jan. 2022 product C2RCC reflectance spectra caused the BNN to generate unrealistically low chla in the middle of the lake and nearshore when compared to neighbouring pixels (Fig. 16 I). The unusual BNN chla patches and areas caused by the POLYMER and C2RCC ACs were consistently associated with 40–100% higher uncertainties than surrounding pixels (Fig. 16 F, J, M). These highly uncertain areas over Lake Turkana were treated and resulted in new products. First, the corresponding pixels were removed by imposing

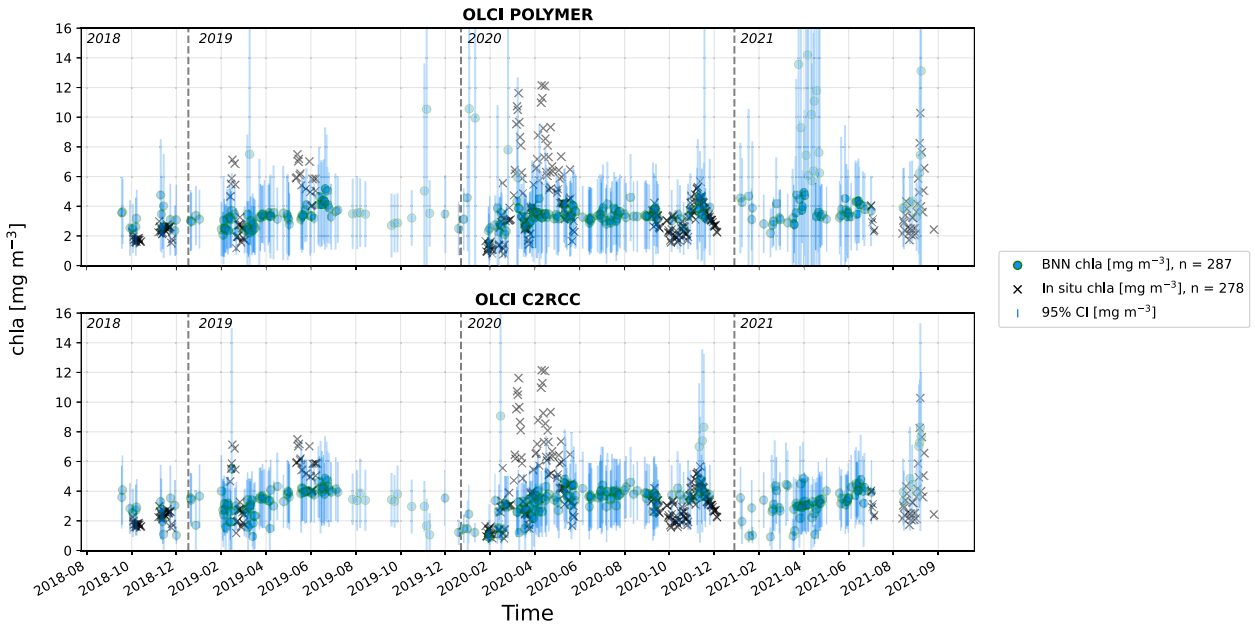

Fig. 11. BNN OLCI time series of LéXPLORE (Lake Geneva), including all match-ups.

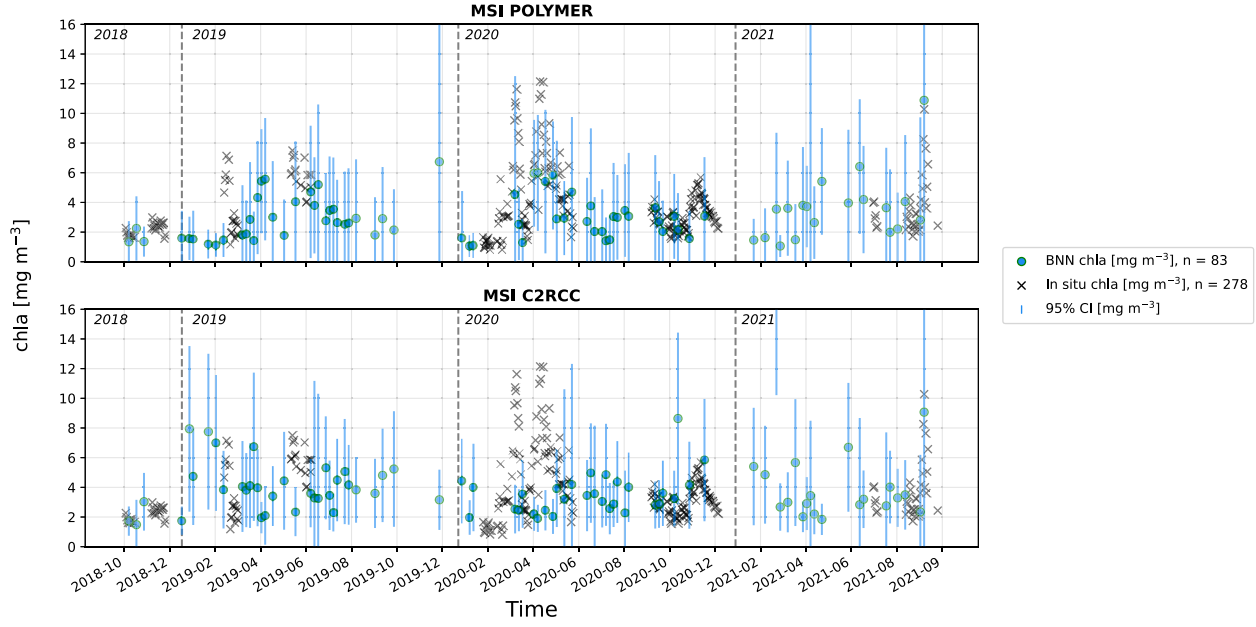

Fig. 12. BNN MSI time series of LéXPLORE (Lake Geneva), including all match-ups.

BNN uncertainty thresholds (higher than 85% for Dec. 2021 and 65% for Jan. 2022) on the entire scene. The procedure eliminated almost all observations with high uncertainty and resulted in more homogenous products (Fig. 16 D, G, K, N). Second, the BNN chla uncertainties associated with each AC algorithm were used to generate merged chla and uncertainty products (Fig. 17), by selecting the chla result per observation with the lowest BNN uncertainty from either AC. The resulting, merged product replaced the highly uncertain areas of the AC processors (see Fig. 16 for the products used in the merge) and led to an improved product quality both in terms of chla and uncertainty.

5. Discussion

Chla estimation uncertainties have repeatedly been shown to be highest in oligotrophic and mesotrophic lakes (Liu et al., 2021; Neil et al., 2019; Werther et al., 2021). The retrieval results over the chla range considered here confirmed these findings. Uncertainty is common in lake remote sensing and unlike the reference chla algorithms, the BNNs indicated when retrieved chla was uncertain. Here we discuss the BNN chla and uncertainty estimation, the performance of the reference algorithms, the uncertainty calibration and the handling of uncertain estimates as exemplified for Lake Turkana.

5.1. OLCI and MSI BNN chla estimation

The uncertainty associated with BNN chla varied between the five performance assessments. Lowest BNN chla retrieval uncertainty was found for the 50/50 split assessment over established chla algorithms. The 50/50 split performance of the BNN resembles results reported in the MDN publications (O'Shea et al., 2021; Pahlevan et al., 2020). In this assessment strategy, the systems in both training and test sets are not strictly independent. Observations from the same system can occur in

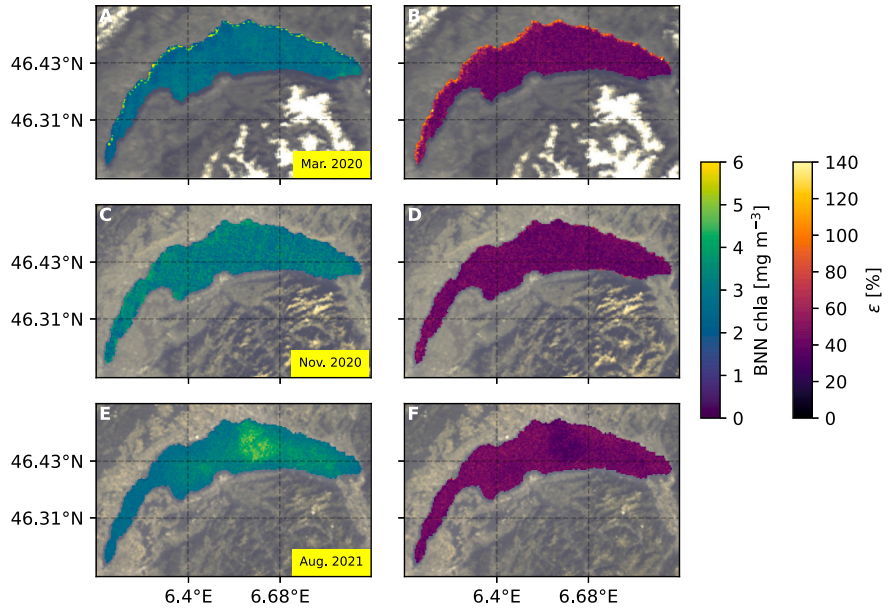

Fig. 13. OLCI BNN chla and uncertainty products obtained through C2RCC AC over Lake Geneva on 13th of March 2020 (A, B), 11th of November 2020 (C, D) and 21st of August 2021 (E, F).

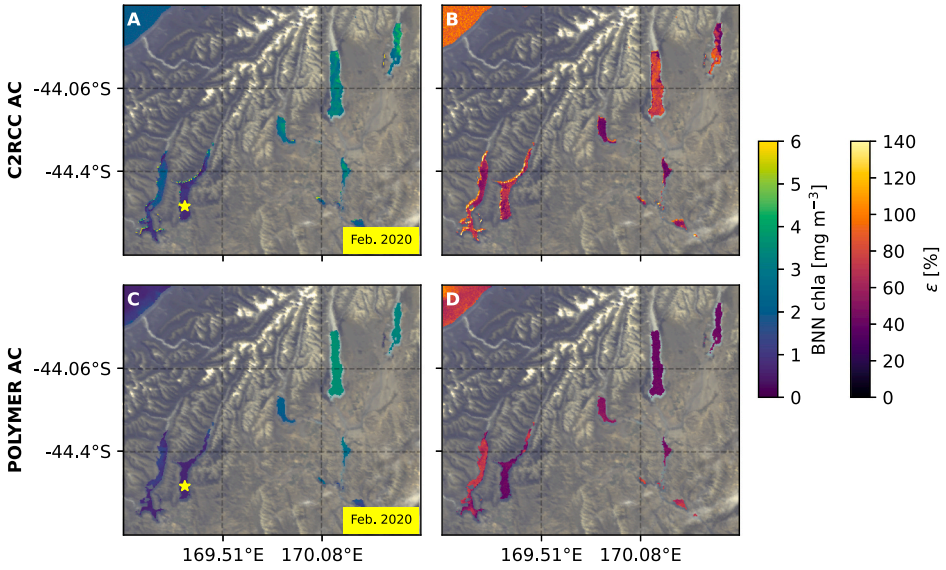

Fig. 14. OLCI BNN chla and uncertainty products over southern New Zealand, 9th of February 2020. Chla was measured *in situ* at the location of Lake Hawea (star symbol) as 0.89 mg m^{-3} during the overpass.

both sets due to the prior randomisation of the entire dataset, thereby introducing knowledge about similar optical properties in the algorithm training stage. Controlled randomisation prevents dataset shifts and the 50/50 dataset split results therefore represent an idealistic evaluation scenario. However, dataset shifts are common during satellite application of a developed ML algorithm. Therefore, the region-wise LOO assessment was conducted, and the results provided further insight into the BNN capabilities. The LOO assessment corroborated the 50/50 split results, whereby for 7 out of 10 regions the BNN outperformed reference algorithms. Expectedly, overall accuracy for LOO was worse than in the 50/50 split assessment. Regional chla accuracy varied from 19% to nearly 77% MdSA for the BNNs but was consistently higher for the reference chla algorithms. Moreover, the reference chla algorithm performances varied strongly between regions. LOO results highlight that the training dataset of this study is representative for measurements sharing the same OWTs, even if from different regions. The LOO results also suggest that varying degrees of uncertainty between the observations of the dataset exist that propagated into the BNN chla estimation. This source of observational uncertainty is irreducible for the BNNs and sets a realistic baseline for expected performance. Compared to the *in situ* assessments, overall BNN accuracy was expectedly lower for the match-ups over Lake Geneva. Although there has been a decline in performance, this was not surprising given that the ACs induced uncertainty ranging from 18 to 60% in the blue to red bands (see Appendix 4). The MdSA of the OLCI BNN match-ups over Geneva were 49.7% (POLYMER) and 59.07% (C2RCC) which is considered within an acceptable performance range. The decrease in accuracy was greater for MSI because the BNN model was overall less precise, as was also seen in the *in situ* data analyses. Additionally, this is consistent with a decline in the accuracy of MSI AC-derived R_{rs} when compared to OLCI.

BNN chla uncertainties were highest for chla <1 mg m^{-3} across larger and smaller water bodies captured on OLCI and MSI products over

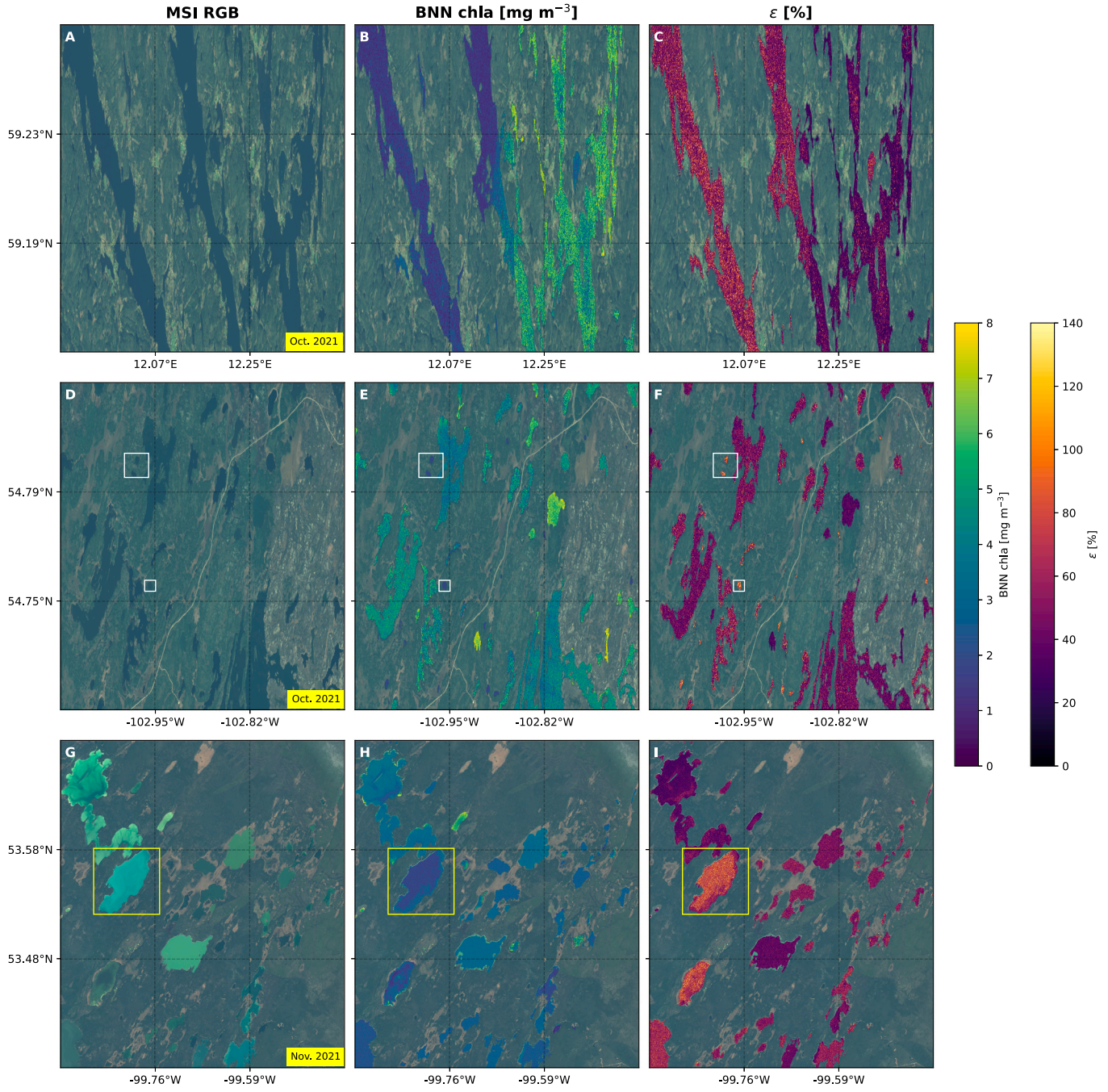

Fig. 15. BNN-C2RCC chla and uncertainty products for MSI (20 m) over Sweden (A, B, C) on 11th of October 2021, and Canada on 16th of October (D, E, F) and 4th of November 2021 (G, H, I). A, D, G represent MSI L1 products with true water colours obtained from top-of-atmosphere sensor radiance, while B, E, H depict produced chla. C, F, I represent associated chla uncertainty. White and yellow boxes surround small water bodies with high BNN uncertainty. See explanations in this section for the boxes.

SNZ, Canada and Lake Turkana. Because measurement availability is globally scarce in the clearest lakes, measurement coverage was low in the training dataset. The low amount of training data and resulting high chla uncertainty over this range clearly indicate a measurement gap. Radiative transfer simulations or coastal data sets may be able to fill this measurement gap to further reduce chla estimation uncertainties. In addition to measurement scarcity, systematic measurement error is prone to lowest chla and phytoplankton absorption coefficients (McKee et al., 2014). With regards to higher biomass conditions, only 9.5% of the dataset consisted of chla >25 mg m^{-3} (maximum 68 mg m^{-3}), therefore chla estimation uncertainty will rise with increasing chla. Oligo- and mesotrophic waters most likely do not exceed this threshold even when experiencing seasonal, short-term phytoplankton blooms such as those observed in Lake Geneva. Observations characterised by high CDOM (> 2 m^{-1}) were limited in the collated dataset. High CDOM is common in boreal and tundra lakes with low - moderate chla but these systems are globally under-sampled and as such only constitute a small part of the training data available for the BNNs.

5.2. Performances of the chla algorithms

We compared the constructed BNNs to five reference chla algorithms that differed in their approach to estimating chla. The OLCI and MSI MDNs were created using a bigger community-dataset (n = 2943) than

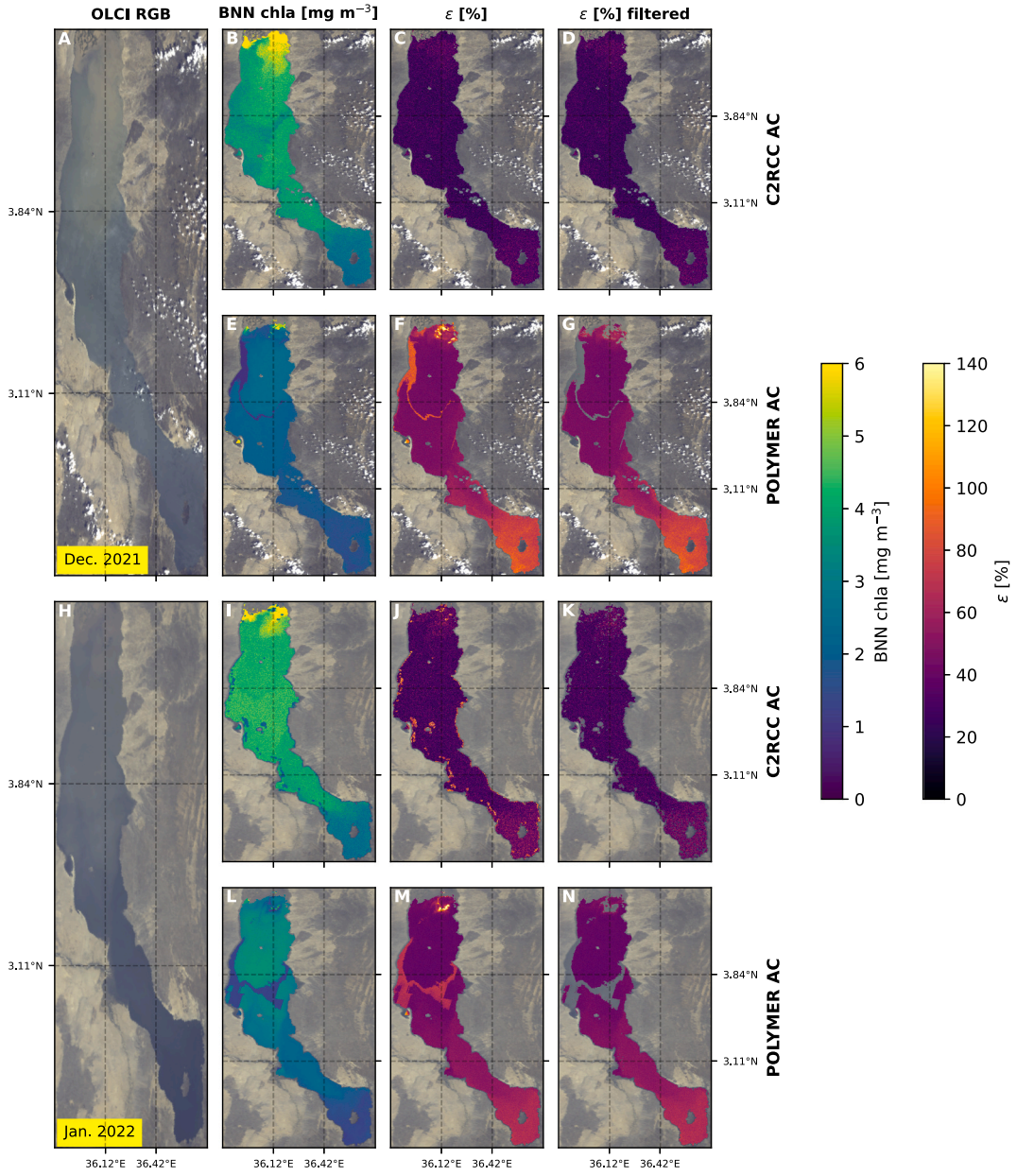

Fig. 16. OLCI BNN chla and uncertainty products for Lake Turkana (Kenya) on 13th of December 2021 and 14th of January 2022 using C2RCC and POLYMER ACs. A, H: OLCI L1 RGB. B, E, I, L: BNN chla. C, F, J, M: Associated uncertainty. D, G, K, N: Filtered pixels based on uncertainty higher than 85% for Dec. 2021 and 65% for Jan. 2022.

the BNNs ($n = 1755$). In contrast to the dataset used in this work, the MDN training dataset was centred on eutrophic waters (chla dataset mean 21.7 mg m^{-3}, median 8.9 mg m^{-3} and standard deviation of 47.5 mg m^{-3}). Accordingly, the performance of the MDNs became steadily more accurate with increasing chla for the observations of this study. A recent chla retrieval uncertainty characterisation of the MDNs by Werther et al. (2022) revealed that MDN overestimation in low biomass waters (chla 0–10 mg m^{-3}) is primarily caused by a lack of training observations from waters with moderate - high NAP and CDOM absorption.

To achieve higher generalisation across global lakes, it appears that training sets for ML algorithms must be further expanded, however gathering *in situ* measurements will remain expensive and time-consuming. As a result, novel techniques to overcome the lack of *in situ* data are required. Recent developments grounded in semi-supervised (Zhu and Goldberg, 2009) or self-supervised learning (Jing and Tian, 2021) extend the training set size through the transformation of unlabelled observations into available training data. Both learning approaches show promise in adjacent remote sensing fields (Wang et al., 2022), but their potential remains untapped for aquatic remote sensing of biogeochemical proxies.

The red/NIR algorithms using 665 and 708 nm such as G11 -opt, Gons05 and Blend became highly accurate (MdSA <20%) at chla > 10 mg m^{-3}. The results match the findings of earlier studies across larger chla ranges reporting inaccuracies of red/NIR algorithms in oligotrophic concentration ranges (Neil et al., 2019; Pahlevan et al., 2020; Werther et al., 2022). The OC3 -org and -opt versions showed moderate - high estimation uncertainties (MdSA >44%) across all trophic states in the 50/50, LOO and match-up assessments. For oligo- and mesotrophic lakes with highly varying OAC concentrations, phytoplankton estimation based only on a blue/green band ratio appears severely constrained.

The developed BNNs are an ensemble of individual NNs, with the size of the ensemble specified by the parameter *S*. Through MC dropout, *S* network variants (50 in this study) with differing numbers of neurons were produced, resulting in a varied chla estimate for an identical input observation. Ensembles consistently outperform individual algorithms, either

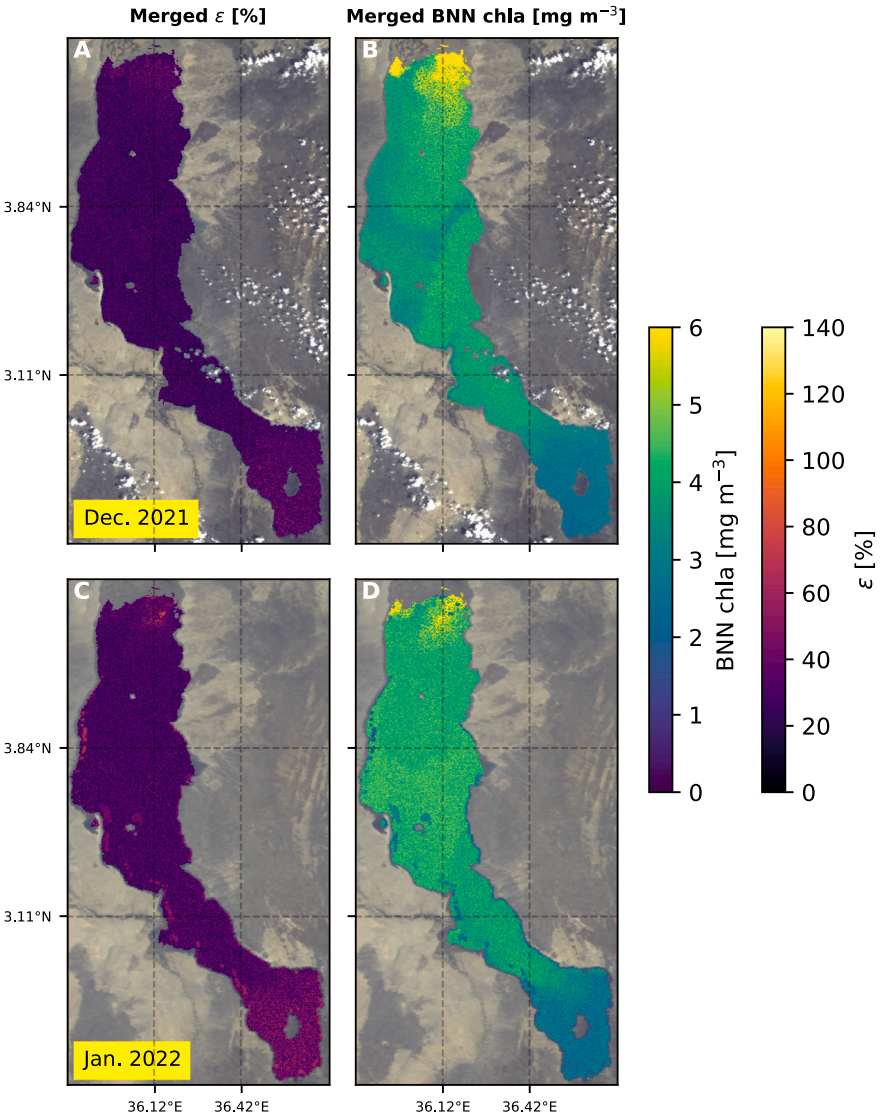

Fig. 17. OLCI BNN merged uncertainty and chla products from 13th of December 2021 and 14th of January 2022 over Lake Turkana. See Fig. 16 for the products used for the merge.

when combined through OWTs (Neil et al., 2019; Liu et al., 2021), statistical merging or blending (Smith et al., 2018; Schaeffer et al., 2022) or stacking of individual algorithms (Werther et al., 2021). However, using multiple algorithms increases model complexity at the expense of feature interpretability. The 25% dropout probability determined S distinct NN variants in this study. The 50 distinct NNs each had a different number of neurons (because different neurons were dropped in each NN variant). Due to their uniqueness, the importance of individual features (the multispectral bands) varied among the 50 NNs. The final chla value was determined by the chla distribution created by all of the different NNs. As a result, this chla distribution is a blend of distinct feature importances, making it a complex task to separate the individual NN contributions in order to identify which bands contributed meaningfully to the model.

5.3. Calibration of uncertainty

The quality of the uncertainty calibration was measured through the PICP, Sharpness and MACD metrics. Across the different assessments, the confidence intervals covered the *in situ* reference chla value in 3/4 cases (PICP >75%). However, the provided uncertainty estimate was not consistently well-calibrated across the assessments. The Sharpness metric specifically measured the range of the uncertainty estimate per observation and small Sharpness values were not achieved in all performance assessments. Moreover, even though OLCI performance did not differ significantly from MSI, the OLCI uncertainties were consistently better calibrated. More research is needed to understand under which conditions the BNN uncertainty calibration was systematically inaccurate and how it correlated with optical properties and training data availability.

The source of uncertainty can be the underlying BNN statistical model (the epistemic uncertainty) or an external source (the aleatoric uncertainty), such as prior AC, an adjacency effect or other unknown error as part of the measurements contained in the dataset (IOCCG, 2019). Whereas epistemic uncertainty can be improved through better data coverage and incremental improvements to the underlying statistical model, aleatoric uncertainty sources are often more difficult to identify, quantify and to correct for as they are methodologically de-coupled from the BNN. AC, for example, is often performed using a distinct algorithm. Uncertainty in the AC method for calculating $R_{rs}(\lambda)$ may thus propagate into the BNN chla estimation. It is difficult to identify and correct for this source of aleatoric uncertainty impacting BNN chla estimation accuracy without particular knowledge about the uncertainty introduced by the AC. To provide precise pathways for BNN improvements and to simplify the identification of uncertainty drivers, it is preferable to separate the aleatoric and epistemic uncertainty contributions entirely. Kendall and

Gal (2017) conducted preliminary work in this direction using MC dropout in computer vision tasks, but it is still far from established research practise in the field of remote sensing.

Besides the investigation into uncertainty origins, computational methods exist to improve the uncertainty calibration as a post-processing step. The approaches usually require a separate dataset, which in practise can be a subset of the overall data that was set aside, and which is consequently excluded from training or evaluation. Example techniques include histogram binning (Zadrozny and Elkan, 2001), isotonic regression (Kuleshov et al., 2018; Zadrozny and Elkan, 2002) and Platt scaling (Platt, 1999).

5.4. AC algorithm selection through BNN uncertainty

Atmospheric correction is the greatest source of uncertainty in satellite remote sensing of OACs in lakes (Pahlevan et al., 2021a; Pereira-Sandoval et al., 2019; Wang et al., 2019; Warren et al., 2019). Observations associated with high uncertainty differed between the BNN estimates for the same products over Lake Geneva, SNZ and Lake Turkana. To deal with varying degrees of uncertainty introduced by AC algorithms, the BNN chla uncertainties were used as a selection mechanism by imposing thresholds (as undertaken in Fig. 16) or by retaining the observation associated with the lowest BNN uncertainty (Fig. 17). The former approach omits observations and is desirable when AC quality is unknown, during unsupervised processing of imagery or when a comparison to another AC algorithm is not possible. When two or more ACs for the same product are available, the latter approach to obtain a merged product is the favoured option. Moreover, the merging of products exemplified in Fig. 17 provides a pathway to decide between different ACs. Since AC performance largely varies with faced atmospheric properties and water column IOPs, it is *a priori* impossible to know which AC performs most accurate. A more thorough analysis with multiple products, areas and atmospheric conditions should be undertaken to determine whether AC selection through BNN uncertainty consistently improves the chla product quality.

6. Conclusion

The BNNs developed in this study represent a novel approach to derive chla and associated uncertainty estimates from $R_{rs}(\lambda)$ from Sentinel-2 (MSI) and Sentinel-3 (OLCI) over oligo- and mesotrophic lakes with the ability to account for phytoplankton abundance shifts into the eutrophic regime. Using the largest *in situ* dataset over oligo- and mesotrophic lakes and reservoirs available to date, the BNNs were trained, validated, and compared against state-of-the-art and community-established algorithms. In this study we demonstrated that the developed BNN algorithms work similarly for both MSI and OLCI observations (Bias $<1.5\%$, MdSA $<26\%$, MAPD $<22\%$), suggesting their potential to produce high-quality chla products across multiple missions. We found that chla uncertainties for MSI were on average 10–25% higher than for OLCI reaffirming that additional bands on OLCI not only benefit the chla retrieval accuracy, but effectively reduce the associated uncertainty. Lowering uncertainty by increasing the number of available bands may be important for the growing number of hyperspectral Earth observation missions, such as the Italian PRISMA (PRecursore IperSpettrale della Missione Applicativa) and U.S. PACE (Plankton, Aerosol, Cloud and ocean Ecosystem).

As evidenced through long-term match-ups over Lake Geneva, the BNNs were able to improve upon the chla retrieval of established methods. Across all assessment strategies, the largest (MdSA $> 25\%$) performance improvements by the BNNs were made over oligotrophic lakes (chla ≤ 8 mg m^{-3}), whereas for chla >25 mg m^{-3} established algorithm approaches performed similarly. As for any other chla algorithm relying on satellite-derived $R_{rs}(\lambda)$, the chla accuracy of the OLCI and MSI BNNs varied with the prior AC algorithm. The BNN uncertainty estimate enabled to handle high uncertainty introduced by AC algorithms and other sources prior to chla retrieval. We exemplified common issues that are caused by ACs over oligo- and mesotrophic lakes: negative $R_{rs}(\lambda)$ in red bands, uncertainty induced by the adjacency effect and implausible $R_{rs}(\lambda)$ values stemming from extreme variation in bio-optical lake properties. To deal with inconsistent results obtained by the different AC algorithms, we suggest further research into the selection of AC products through BNN uncertainties to retain accurate chla retrievals. Moreover, research into strategies to improve the uncertainty calibration is recommended. Further downstream products depending on accurate chla, such as trophic status and net primary production, may impose uncertainty quality thresholds to retain high-quality chla retrievals. The BNN architecture presented in this study can be extended to other multi- and hyperspectral missions such as the Landsat missions, PRISMA and PACE and it supports the retrieval of other OACs and IOPs such as TSM and CDOM.

Code availability

The OLCI and MSI BNNs are available under https://github.com/mowerther/BNN_2022

CRediT author statement

Conceptualisation: MW, ES, DO, SGHS
Data curation: MW, ES, DO, ML, DG, TK, PH, AT, RG, AV
Methodology: MW
Formal analysis: MW
Investigation: MW, ES, DO, SGHS, DG
Software: MW
Visualisation: MW
Writing – Original Draft: MW
Writing – Review & Editing: SGHS, DO, ES, DG, TK, ML, PH, AT, RG, AV

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.

Data availability

Data will be made available on request.

Acknowledgements

This study has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 776480 (MONOCLE). The fieldwork in New Zealand (2020) was funded by the EUMETSAT Copernicus Collaborative Exchange Award, the University of Stirling (UK) and University of Waikato (NZ). We thank three anonymous reviewers for their constructive and thorough comments on the manuscript. We are grateful to the LIMNADES data providers for their measurements that we used in this study: Stewart Bernard, Caren E. Binding, Mariano Bresciani, Claudia Giardino, Anatoly A. Gitelson, Luis Guanter, Kari Y.O. Kallio, Tiit Kutser, Ciro Manzo, Mark W. Matthews, John F. Schalles and Antonio Ruiz-Verdú. We thank Camille Minaudo for assistance with processing chla and $R_{rs}(\lambda)$ measurements obtained on LéXPLORE.

Appendix 1. Dataset

Regions and inland water bodies constituting the dataset of this study. Colour shading matches Fig. 1.

Region	Countries	Dataset measure-ments (n = 1755)	Inland water bodies (n = 178)	References
1	Spain	109	Aguilar reservoir, Alarcon reservoir, Lake Albufera, Alcantara reservoir, Alcorlo reservoir, Almendra reservoir, Bornos reservoir, Burguillo reservoir, Canelles reservoir, Cernadilla reservoir, Cijara reservoir, Contreras reservoir, Cortes de Pallas reservoir, Cuerda del Pozo reservoir, El Atazar reservoir, Giribaile reservoir, Guadalcacin reservoir, Guadalen reservoir, Guadalteba reservoirs, Iznajar reservoir, La-Serena reservoir, Negratin reservoir, Pinilla reservoir, Rialb reservoir, Riano reservoir, Ricobayo reservoir, San Juan reservoir, Lake Sanabria, Santa Teresa reservoir, Terradets reservoir, Pantano de Tremp reservoir, Uribarri Ganboako urtegia reservoir, Valdecanas reservoir, Valmayor reservoir, Valparaiso reservoir, Valuengo reservoir	(Ruiz-Verdú et al., 2005, 2008; Simis et al., 2007)
2	The Nether-lands, Is-rael	107	IJsselmeer, Lake Kinneret	(Simis et al., 2005; 2007; Yacobi et al., 2011)
3	Switzer-land, Italy	466	Lake Biel, Lake Garda, Lake Idro, Lake Como, Lake Iseo, Lake Geneva, Lake Maggiore, Lake Trasimeno	(Bresciani et al., 2011; Giardino et al., 2005, 2013, 2014, 2015; Guanter et al., 2010; Manzo et al., 2015)
4	United Kingdom, Sweden, Finland, Estonia	75	Lake Antu, Lake Bassenthwaite, Lake Coniston, Loch Katrine, Loch Lomond, Lake Maleren, Loch Ness, Lake Peijänne, Lake Peipsi, Lake Pyhäjärvi, Lake Ullswater, Lake Vänern, Lake Vesijärvi	(Kallio et al., 2015; Kutser et al., 2013)
5	China, Japan	45	Lake Biwa, Lake Erhai, Lake Suwa	(Matsushita et al., 2015)
6	New Zealand	133	Lake Brunner, Lake Hauroko, Lake Hawea, Lake Ellesmere, Lake Humuhumu, Lake Kahuparere, Lake Kai-iwi, Lake Kanono, Lake Kegonsa, Lake Mokeno, Lake Okareka, Lake Okaro, Lake Okataina, Lake Ototoa, Lake Pearson, Lake Pupuke, Lake Rotoehu, Lake Rotoiti, Lake Rotokakahi, Lake Rotokare, Lake Rotokawau, Lake Rotoma, Lake Rotonuiaha, Lake Rotoroa, Lake Rotorua, Lake Rototuna, Lake Sheppard, Lake Tarawera, Lake Taupo, Lake Tutuira, Lake Waikaremoana, Lake Waikere, Lake Mapourika, Lake Moeraki, Lake Monowai, Lake Paringa, Lake Te Anau, Lake Wahapo, Lake Wakatipu	(Pahlevan et al., 2020, 2022; Wang et al., 2018)
7	South Africa	26	Loskop reservoir, Theewaterskloof reservoir	(Matthews, 2014; Matthews and Bernard, 2013)
8	U.S. Mainland	487	Lake Amos, Lake BranchedOak, Lake Collins, Lake CorpusCristi, Lake Crescent, Lake DiamondPond, Lake EagleCreek, Lake EastTwin, Lake Ewell, Lake Fivemile, Flatriver reservoir, Lake Forest, Lake Fremond, Lake Fresmond, Lake Geist, Lake GingerCove, Lake GoosePond, Lake Granite, Great Salt Lake, Green Lake, Groton Lake, Hickleys Lake, Kettlebrook reservoir, Lake Attitash, Lake Champlain, Lake Hayward, Lake Manawa, Lake Mattawa, Lake Sunapee, Lake Wentworth, Lake LittleBigWood, Long Lake, Long Pond, Lowell Lake, Lake Lowerbranch, Lake Mantova, Lake Morse, Norton reservoir, Lake Okoboji, Perch Pond, Pine Lake, Pleasant Lake, Quacumquasit Pond, Quinebaug Lake, Sebago Lake, Sebasticook Lake, Silver Lake, Squam Lake, Tilden Pond, Tyler Lake, Lake Whalum, Wallum Lake	(Bradt, 2012; Dall'Olmo et al., 2003, 2005; Gitelson et al., 2007; Gitelson et al., 2008; Li et al., 2013, 2015; Moore et al., 2014; Schalles, 2006; Schalles and Hladik, 2012)
9	U.S. Wisconsin	109	Big Saint Germain Lake, Big Sand Lake, Butternut Lake, Fence Lake, Fox Lake, Geneva Lake, Lac Courte Oreilles, Lac Vieux Desert, Lake Chippewa, Lake Mendota, Lake Monona, Lake Waubesa, Lake Winnebago, Metonga Lake, Pelican Lake, Rock Lake, Round Lake, Shawano Lake, Trout Lake	(Pahlevan et al., 2022, 2021, 2020)
10	U.S., Canada	198	Lake Erie, Lake Ontario, Shadow Lake, Lake Winnipeg	(Binding et al., 2008, 2010; 2011; 2013)

Appendix 2. *In situ* measurements NZ 2020

Chla water samples were filtered (using a 0.45 μm GF/F filter) and then extracted using 95% methanol (spiked with Tocopheryl acetate as an internal standard) before being sonicated and frozen to complete extraction of all chla (Hooker et al., 2009). Finally, the extract was filtered through a 0.45 μm Teflon filter. Resultant extracts were then run on an Agillent Infinity 1290 uHPLC using the method by Van Heukelem and Thomas (2001). TSM was measured gravimetrically from the dried and combusted residue on pre-combusted and pre-weighed filter pads (APHA 2540D and APHA 2540E). Laboratory triplicates of $a_{CDOM}(\lambda)$ were transferred to a 0.1 m cuvette and optical densities of the filtrates were measured for a wavelength range from 350 to 800 nm using the method described in Stedmon et al. (2000).

Radiometric measurements were made with a ship-mounted set of three hyperspectral sensors (TriOS RAMSES) installed just above the water surface. The first sensor pointed at the water surface and collected upwelling radiance $L_u(\lambda)$, which comprises both water-leaving radiance $L_w(\lambda)$ and the reflected sky irradiance $\rho_s L_s(\lambda)$. The sky irradiance $L_s(\lambda)$ was measured with a second sensor, while a third sensor measured downwelling irradiance $E_d(\lambda)$. $R_{rs}(\lambda)$ was then calculated through Simis and Olsson (2013):

$$R_{rs}(\lambda) = L_w(\lambda, 0^+)/E_d(\lambda), \tag{A2}$$

$$L_w(\lambda, 0^+) = L_t(\lambda) - \rho_s L_s(\lambda). \tag{A2.1}$$

Measurements were kept in this dataset if they were taken under cloud free conditions and on calm waters to ensure a correct alignment of the ship at a viewing azimuth angle (φ_v) > 90° (ideally at 135°) (Hooker and Morel, 2003). All $R_{rs}(\lambda)$ were corrected for residual reflected skylight through the method described in Jiang et al. (2020).

Appendix 3. LéXPLORE platform chla measurements

The autonomous Thetis profiler (Seabird Scientific Inc.) is deployed at the LéXPLORE platform in Lake Geneva. The profiler is comprised of hyperspectral radiometers, absorption, backscattering and attenuation sensors and a fluorescence probe for chla measurements. The top 50 m of the water column are measured by the profiler at 3-h intervals and a vertical resolution of 10 cm. A total of 278 profiles taken between October 2018 and September 2021 were used in this study. *In situ* measurements from the Thetis profiler are available in the Datalakes portal (https://www.datalakes-eawag.ch/).

Fluorescence probe-derived chla (CHL_F) from the Thetis profiler were measured using the WetLabs ECO Triplets BBFL2W, which has excitation and emission wavelengths at 470 and 695 nm, respectively. CHL_F estimates are impacted by non-photochemical quenching (NPQ) of fluorescence leading to an underestimation of chla in high irradiance conditions. To compensate for the underestimation, we applied the chla correction procedure developed by Roesler and Barnard (2013). We obtained a linear relationship between CHL_F and absorption line height (aLH) based chla estimates (CHL_A) during low light conditions which ensured minimal impact of NPQ to CHL_F. This linear relationship was used for daytime aLH measurements to correct NPQ-impacted CHL_F estimates. aLH was calculated using the following equations:

$$\mathrm{a_{BL}}\left(\lambda_{\mathrm{ref}}\right) = \frac{a(715) - a(650)}{715 - 650} * \left(\lambda_{ref} - 650\right) + a(650), \mathrm{a_{LH}}(676) = \mathrm{a}(676) - \mathrm{a_{BL}}(676). \tag{A3.1}$$

where a_{BL} is the absorption measured from the AC-S at the reference wavelength λ_{ref} of 676 nm. CHL_A was calculated for the upper 50 m of the water column and the top 5 m were averaged to represent *in situ* surface chla.

Appendix 4. OLCI/MSI POLYMER/C2RCC remote sensing reflectance validation

The underwater upwelling radiance (L_u) and downwelling irradiance (E_d) measurements from the Thetis profiler on LéXPLORE were measured at nadir using the Satlantic HOCR RO8W and ICSW, respectively. Both radiometers span 180 spectral channels from 300 to 1200 nm. The underwater reflectance (R_m) was calculated using the following equation:

$$R_m\left(\lambda, z\right) = \pi\left(\frac{L_u(\lambda, z)}{E_d(\lambda, z)}\right)\left[\mathrm{dl}\right], \tag{A4.1}$$

where λ is the wavelength region from 400 to 800 nm and z is depth in meters. The R_m closest to the water surface was converted to above-water remote sensing reflectance using a transmission factor following Doxaran et al. (2004):

$$R_{rs}(\lambda, z) = 0.544 * R_m(\lambda, z) \tag{A4.2}$$

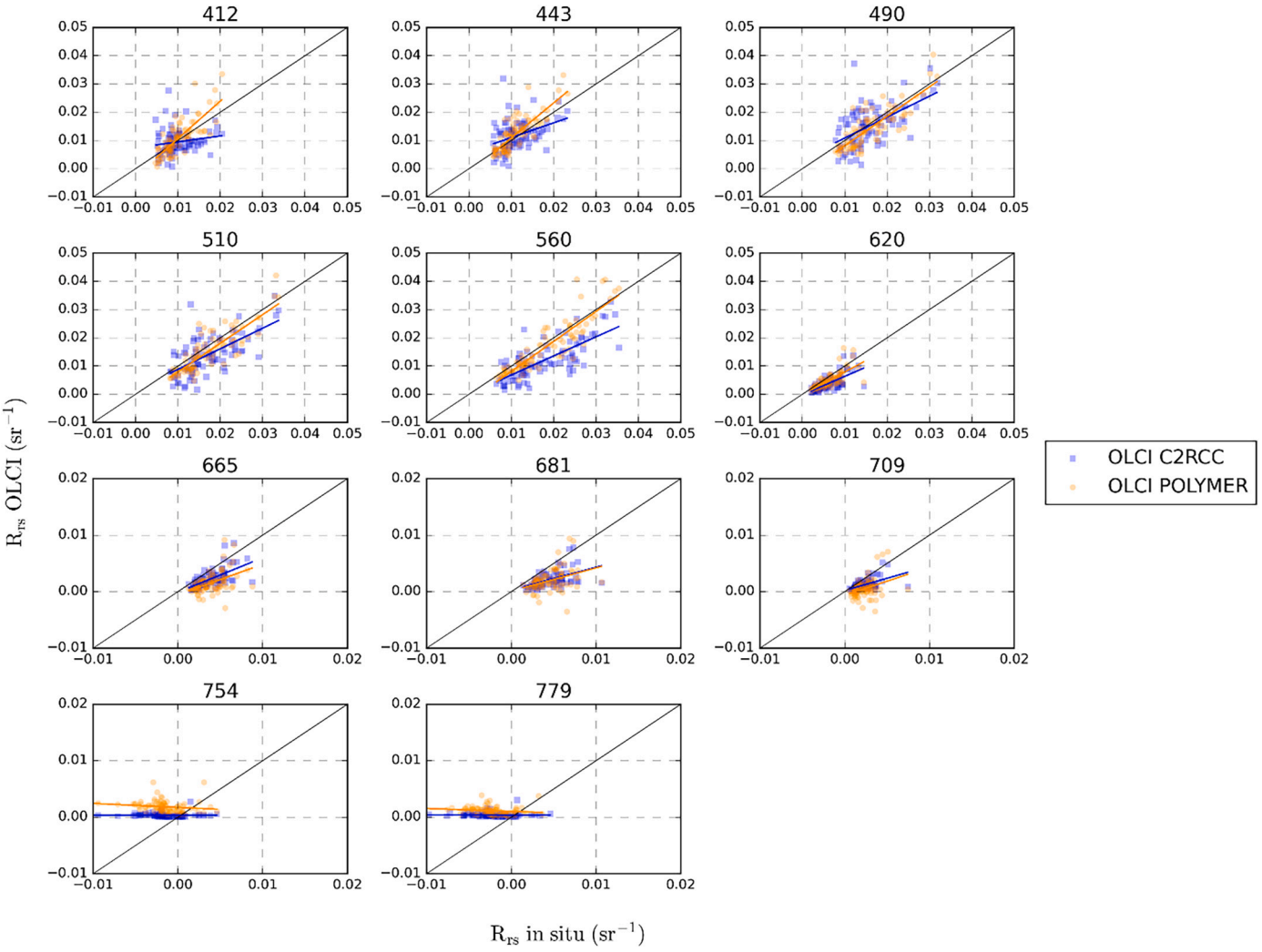

Fig. A4.1. LéXPLORE *in situ* $R_{rs}(\lambda)$ *versus* OLCI C2RCC and POLYMER $R_{rs}(\lambda)$ for all available match-ups ($n = 90$). The black line represents the theoretical 1:1 regression, coloured lines linear fits through the observations. See Fig. A4.3 for the performance metrics per band. The y-axis scale changes $\geq$ 665 nm.

In this section, we compare the LéXPLORE *in situ* $R_{rs}(\lambda)$ *versus* Sentinel-3A/B OLCI and Sentinel-2A/B MSI. The POLYMER (Steinmetz et al., 2011) and C2RCC (Brockmann et al., 2016) algorithms were used to atmospherically correct all OLCI and MSI scenes. In total, 99 OLCI match-ups with LéXPLORE chla were available (see section 2.2. for a description of the match-up details). However, $R_{rs}(\lambda)$ was unavailable for nine of these OLCI chla match-ups due to radiometer maintenance, resulting in 90 observations available for this validation (Fig. A4.1). POLYMER estimated negative $R_{rs}(\lambda) \geq$ 665 nm for 21 chla OLCI match-ups that were omitted from the chla algorithm evaluation (section 4.3) to allow for a focus on the chla algorithm mechanics. These negative OLCI POLYMER reflectance spectra, however, were used in this validation to investigate the AC performance. Similarly, 29 MSI chla match-ups were generated. 12 negative MSI POLYMER $R_{rs}(\lambda)$ were removed from the chla algorithm comparison but included in this $R_{rs}(\lambda)$ validation. *In situ* $R_{rs}(\lambda)$ was unavailable for three of the chla match-ups, yielding a total of 26 MSI match-ups for validation of AC-derived $R_{rs}(\lambda)$ (Fig. A4.2). Until 665 nm, both ACs scored the highest accuracies (Fig. A4.3). Over the blue-green bands, OLCI POLYMER was more accurate (MdSA $\leq$ 25%) than OLCI C2RCC (MdSA 10–30% higher) (Fig. A4.3).

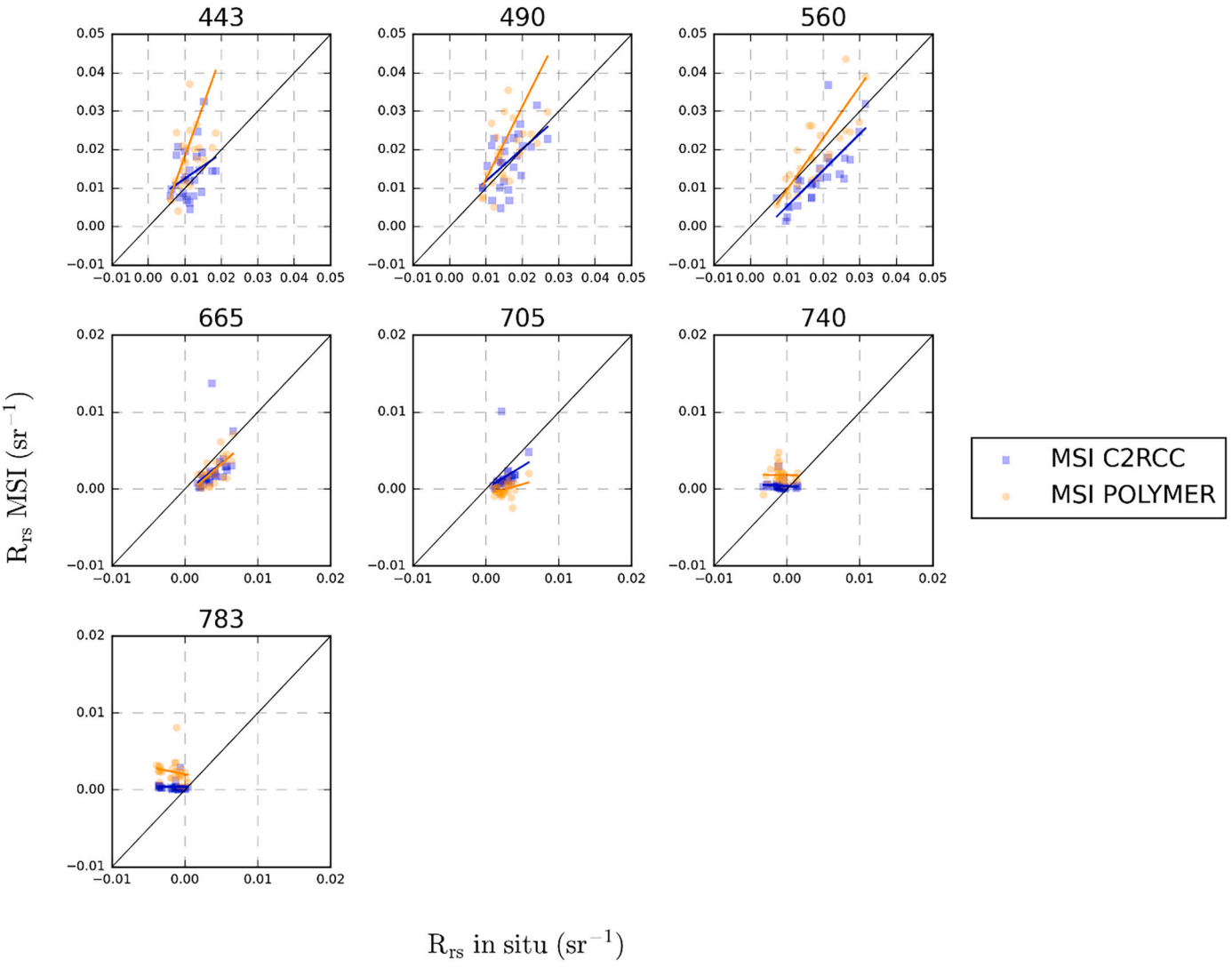

Fig. A4.2. LéXPLORE *in situ* $R_{rs}(\lambda)$ *versus* MSI C2RCC and POLYMER $R_{rs}(\lambda)$ for all available match-ups ($n = 26$). See Fig. A4.1 for a description of the lines and A4.3 for performance metrics. The y-axis scale changes $\geq$ 665 nm.

The accuracy of the ACs in 443, 490, and 560 nm for MSI remained similar to OLCI, with minor differences in the metrics due to the smaller number of match-ups. Across the match-ups, $R_{rs}(\lambda)$ decreased sharply towards red wavelengths and reached near-zero $\geq$ 665 nm. Because of low signal-to-noise ratios in the red/NIR bands, *in situ* $R_{rs}(\lambda)$ became partially negative $\geq$ 740 nm (MSI) and 754 nm (OLCI). Similarly, POLYMER calculated a small number of negative observations $\geq$ 665 nm. C2RCC did not produce negative $R_{rs}(\lambda)$ because the C2RCC neural networks were originally trained using reflectance spectra with values greater than zero for wavelengths $\geq$ 665 nm to constrain the output (Brockmann et al., 2016). C2RCC provided more accurate $R_{rs}(\lambda)$ in these red/NIR wavelengths than POLYMER (see Fig. A4.3). The chla match-ups over LéXPLORE did not include *in situ* $R_{rs}(\lambda)$ observations. Because of the negative *in situ* $R_{rs}(\lambda)$, the AC-derived $R_{rs}(\lambda)$ cannot be properly compared at 754 and 779 nm (OLCI) and 740 and 783 nm (MSI). Nonetheless, C2RCC and POLYMER values near zero appear to be feasible in these wavelengths. Because negative values obtained using POLYMER or another AC may arise for oligotrophic lakes with low signal-to-noise ratios in red/NIR, observations with negative values were set to zero in the BNNs pre-processing, preventing unpredictable behavior that would result in inflated chla estimates and excessive uncertainty (see section 3.2 for pre-processing details).

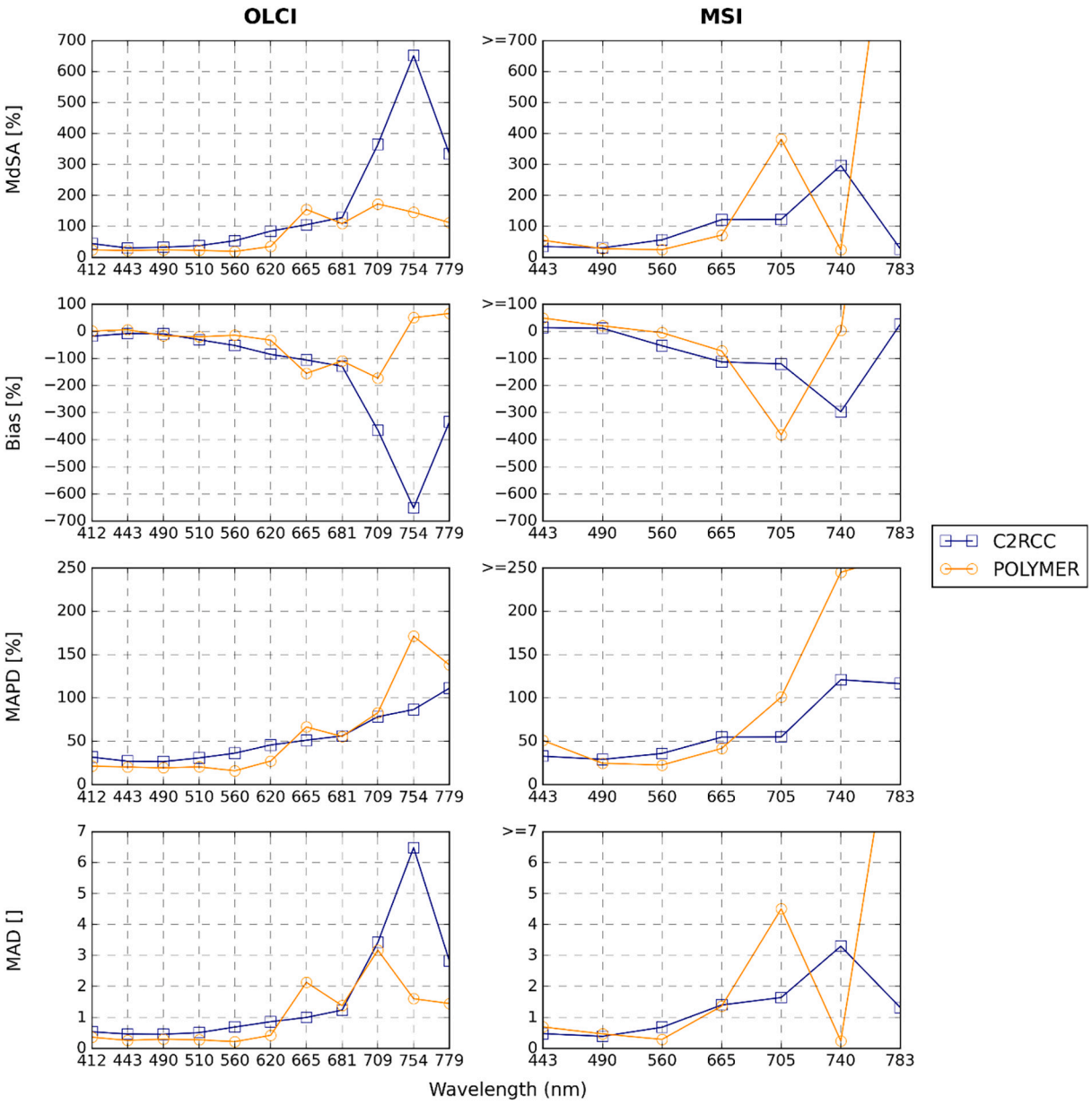

Fig. A4.3. Performance statistics per wavelength for the available OLCI and MSI match-ups generated with POLYMER and C2RCC over LéXPLORE.

Appendix 5. BNN hyper-parameters

Hyper-parameter	Value
Training epochs	9000
Hidden layers	6
Neurons in hidden layers 1–6	200, 500, 500, 500, 200, 2
Batch-size	512
Activation function	Rectified Linear Unit (ReLU)
Dropout rate	25%
Learning rate	1.5×10^{-4}
Optimiser	Adam
Loss function	Negative-log likelihood (NLL)

Appendix 6. Optimised coefficients of reference chla algorithms

Algorithm	Coefficient 1	Coefficient 2	Coefficient 3	Coefficient 4	Coefficient 5
			OLCI		
OC3 -opt	0.07406589	−2.02001282	0.16124464	−17.55852465	−33.62782537
OC4 -opt	0.12359851	−2.80056003	3.94756167	−23.76880782	−92.02561384
G11 -opt	−7.47220546	47.51051798	−22.15274869		
			MSI		
OC3 -opt	0.07090613	−2.15824786	1.41248107	−25.01903886	−60.82627189
G11 -opt	2.22937993	38.80134518	−22.82347274		

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., Makarenkov, V., Nahavandi, S., 2021. A review of uncertainty quantification in deep learning: techniques, applications and challenges. Information Fusion 76, 243–297. https://doi.org/10.1016/j.inffus.2021.05.008.

Anneville, O., Souissi, S., Gammeter, S., Straile, D., 2004. Seasonal and inter-annual scales of variability in phytoplankton assemblages: comparison of phytoplankton dynamics in three peri-alpine lakes over a period of 28 years. Freshw. Biol. 49, 98–115. https://doi.org/10.1046/j.1365-2426.2003.01167.x.

Biewald, L., 2020. Experiment Tracking with Weights and Biases [WWW Document]. https://www.wandb.com/.

Binding, C., Jerome, J., Bukata, R., Booty, W., 2008. Spectral absorption properties of dissolved and particulate matter in Lake Erie. Remote Sens. Environ. 112, 1702–1711. https://doi.org/10.1016/j.rse.2007.08.017.

Binding, C.E., Greenberg, T.A., Bukata, R.P., 2013. The MERIS Maximum Chlorophyll Index; its merits and limitations for inland water algal bloom monitoring. J. Great Lakes Res. 39, 100–107. https://doi.org/10.1016/j.jglr.2013.04.005.

Binding, C.E., Jerome, J.H., Bukata, R.P., Booty, W.G., 2010. Suspended particulate matter in Lake Erie derived from MODIS aquatic colour imagery. Int. J. Remote Sens. 31, 5239–5255. https://doi.org/10.1080/01431160903302973.

Bolstad, W.M., Curran, J.M., 2016. Introduction to Bayesian Statistics, Third edition. John Wiley & Sons, Inc, Hoboken, NJ, USA.

Boyer, J.N., Kelble, C.R., Ortner, P.B., Rudnick, D.T., 2009. Phytoplankton bloom status: chlorophyll a biomass as an indicator of water quality condition in the southern estuaries of Florida, USA. Ecol. Indic. 9, S56–S67. https://doi.org/10.1016/j.ecolind.2008.11.013.

Bradt, S.R., 2012. Development of bio-optical algorithms to estimate chlorophyll in the Great Salt Lake and New England lakes using in situ hyperspectral measurements. The Uni. of New Hampshire.

Bresciani, M., Stroppiana, D., Odermatt, D., Morabito, G., Giardino, C., 2011. Assessing remotely sensed chlorophyll-a for the implementation of the Water Framework Directive in European perialpine lakes. Sci. Total Environ. 409, 3083–3091. https://doi.org/10.1016/j.scitotenv.2011.05.001.

Bricaud, A., Babin, M., Morel, A., Claustre, H., 1995. Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: analysis and parameterization. J. Geophys. Res. Oceans 100, 13321–13332. https://doi.org/10.1029/95JC00463.

Bricaud, A., Claustre, H., Ras, J., Oubelkheir, K., 2004. Natural variability of phytoplanktonic absorption in oceanic waters: influence of the size structure of algal populations. J. Geophys. Res. Oceans 109. https://doi.org/10.1029/2004JC002419.

Bricaud, A., Morel, A., Babin, M., Allali, K., Claustre, H., 1998. Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters : analysis and implications for bio-optical models. J. Geophys. Res. 103, 31033–31044.

Brockmann, C., Doerffer, R., Peters, M., Kerstin, S., Embacher, S., 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. In: Living Planet Symposium, ESA Special Publication, p. 54.

Cao, Z., Ma, R., Duan, H., Pahlevan, N., Melack, J., Shen, M., Xue, K., 2020. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. Remote Sens. Environ. 248, 111974 https://doi.org/10.1016/j.rse.2020.111974.

Carlson, R.E., 1977. A trophic state index for lakes. Limnol. Oceanogr. 22, 361–369. https://doi.org/10.4319/lo.1977.22.2.0361.

Carvalho, L., Poikane, S., Lyche Solheim, A., Phillips, G., Borics, G., Catalan, J., De Hoyos, C., Drakare, S., Dudley, B.J., Järvinen, M., Laplace-Treyture, C., Maileht, K., McDonald, C., Mischke, U., Moe, J., Morabito, G., Nõges, P., Nõges, T., Ott, I., Pasztaleniec, A., Skjelbred, B., Thackeray, S.J., 2013. Strength and uncertainty of phytoplankton metrics for assessing eutrophication impacts in lakes. Hydrobiologia 704, 127–140. https://doi.org/10.1007/s10750-012-1344-1.

Clark, J.M., Schaeffer, B.A., Darling, J.A., Urquhart, E.A., Johnston, J.M., Ignatius, A.R., Myer, M.H., Loftin, K.A., Werdell, P.J., Stumpf, R.P., 2017. Satellite monitoring of cyanobacterial harmful algal bloom frequency in recreational waters and drinking water sources. Ecol. Indic. 80, 84–95. https://doi.org/10.1016/j.ecolind.2017.04.046.

Coffer, M.M., Schaeffer, B.A., Salls, W.B., Urquhart, E., Loftin, K.A., Stumpf, R.P., Werdell, P.J., Darling, J.A., 2021. Satellite remote sensing to assess cyanobacterial bloom frequency across the United States at multiple spatial scales. Ecol. Indic. 128, 107822 https://doi.org/10.1016/j.ecolind.2021.107822.

Dall'Olmo, G., Gitelson, A.A., Rundquist, D.C., 2003. Towards a unified approach for remote estimation of chlorophyll-a in both terrestrial vegetation and turbid productive waters. Geophys. Res. Lett. 30 https://doi.org/10.1029/2003GL018065.

Dall'Olmo, G., Gitelson, A.A., Rundquist, D.C., Leavitt, B., Barrow, T., Holz, J.C., 2005. Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands. Remote Sens. Environ. 96, 176–187. https://doi.org/10.1016/j.rse.2005.02.007.

Degroot, M.D., Fienberg, S.E., 1983. The comparison and evaluation of forecasters. In: Proceedings of the 1982 .I.O.S. Annual Conference on Practical Bayesian Statistics, pp. 12–22. https://doi.org/10.2307/2987588. Carnegie-Mellon University, Pittsburgh, PA, Department of Statistics; TR-244.

Dekker, A.G., Peters, S.W.M., 1993. The use of the thematic mapper for the analysis of eutrophic lakes: A case study in the netherlands. Int. J. Remote Sens. 14, 799–821. https://doi.org/10.1080/01431169308904379.

Doxaran, D., Nagur Cherukuru, R.C., Lavender, S.J., 2004. Estimation of surface reflection effects on upwelling radiance field measurements in turbid waters. J. Opt. A: Pure Appl. Opt. 6, 690–697. https://doi.org/10.1088/1464-4258/6/7/006.

Filazzola, A., Mahdiyan, O., Shuvo, A., Ewins, C., Moslenko, L., Sadid, T., Blagrave, K., Imrit, M.A., Gray, D.K., Quinlan, R., O'Reilly, C.M., Sharma, S., 2020. A database of chlorophyll and water chemistry in freshwater lakes. Scientific Data 7, 1–10. https://doi.org/10.1038/s41597-020-00648-2.

Gal, Y., Ghahramani, Z., 2016. Dropout as a bayesian approximation: representing model uncertainty in deep learning. Proceedings of the 33rd International conference on machine learning 48, 1050–1059.

Ghahramani, Z., 2015. Probabilistic machine learning and artificial intelligence. Nature 521, 152–459. https://doi.org/10.1038/nature14541.

Giardino, C., Bresciani, M., Stroppiana, D., Oggioni, A., Morabito, G., 2013. Optical remote sensing of lakes: an overview on Lake Maggiore. J. Limnol. 73, 201–214. https://doi.org/10.4081/jlimnol.2014.817.

Giardino, C., Bresciani, M., Stroppiana, D., Oggioni, A., Morabito, G., 2014. Optical remote sensing of lakes: an overview on Lake Maggiore. J. Limnol. 73, 201–214. https://doi.org/10.4081/jlimnol.2014.817.

Giardino, C., Bresciani, M., Valentini, E., Gasperini, L., Bolpagni, R., Brando, V.E., 2015. Airborne hyperspectral data to assess suspended particulate matter and aquatic vegetation in a shallow and turbid lake. Remote Sens. Environ. 157, 48–57. https://doi.org/10.1016/j.rse.2014.04.034.

Giardino, C., Candiani, G., Zilioli, E., 2005. Detecting chlorophyll-a in Lake Garda using TOA MERIS radiances. Photogramm. Eng. Remote Sens. 71, 1045–1051. https://doi.org/10.14358/PERS.71.9.1045.

Gilarranz, L.J., Narwani, A., Odermatt, D., Siber, R., Dakos, V., 2022. Regime shifts, trends, and variability of lake productivity at a global scale. Proc. Natl. Acad. Sci. 119 https://doi.org/10.1073/pnas.2116413119.

Gilerson, Alexander A., Gitelson, A.A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., Ahmed, S.A., 2010. Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. Opt. Express 18, 24109. https://doi.org/10.1364/oe.18.024109.

Gitelson, A.A., 1992. The peak near 700 nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll. Int. J. Remote Sens. 13, 3367–3373. https://doi.org/10.1080/01431169208904125.

Gitelson, A.A., Dall'Olmo, G., Moses, W., Rundquist, D.C., Barrow, T., Fisher, T.R., Gurlin, D., Holz, J., 2008. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. Remote Sens. Environ. 112, 3582–3593. https://doi.org/10.1016/j.rse.2008.04.015.

Gitelson, A.A., Schalles, J.F., Hladik, C.M., 2007. Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake Bay case study. Remote Sens. Environ. 109, 464–472. https://doi.org/10.1016/j.rse.2007.01.016.

Gons, H., Auer, M., Effler, S., 2008. MERIS satellite chlorophyll mapping of oligotrophic and eutrophic waters in the Laurentian Great Lakes. Remote Sens. Environ. 112, 4098–4106. https://doi.org/10.1016/j.rse.2007.06.029.

Gons, H., Rijkeboer, M., Ruddick, K., 2005. Effect of a waveband shift on chlorophyll retrieval from MERIS imagery of inland and coastal waters. J. Plankton Res. 27.

Gons, H.J., 1999. Optical teledetection of chlorophyll a in turbid inland waters. Environ. Sci. Technol. 33, 1127–1132. https://doi.org/10.1021/es9809657.

Gower, J., King, S., Borstad, G., Brown, L., 2005. Detection of intense plankton blooms using the 709 nm band of the MERIS imaging spectrometer. Int. J. Remote Sens. 26, 2005–2012. https://doi.org/10.1080/01431160500075857.
Gower, J.F.R., 1980. Observations of in situ fluorescence of chlorophyll-a in Saanich Inlet. Bound.-Layer Meteorol. 18, 235–245. https://doi.org/10.1007/BF00122022.
Gower, J.F.R., Brown, L., Borstad, G.A., 2004. Observation of chlorophyll fluorescence in west coast waters of Canada using the MODIS satellite sensor. Can. J. Remote. Sens. 30, 17–25. https://doi.org/10.5589/m03-048.
Graves, A., 2011. Practical variational inference for neural networks. Adv. Neural Inf. Proces. Syst. 24, 2348–2356.
Grizzetti, B., Lanzanova, D., Liquete, C., Reynaud, A., Cardoso, A.C., 2016. Assessing water ecosystem services for water resource management. Environ. Sci. Pol. 61, 194–203. https://doi.org/10.1016/j.envsci.2016.04.008.
Grzymski, J., Johnsen, G., Sakshaug, E., 1997. The significance of intracellular self-sharding on the biooptical properties of bronw, red and green macroalgae. J. Phycol. 33, 408–414. https://doi.org/10.1111/j.0022-3646.1997.00408.x.
Guanter, L., Ruiz-Verdú, A., Odermatt, D., Giardino, C., Simis, S., Estellés, V., Heege, T., Domínguez-Gómez, J.A., Moreno, J., 2010. Atmospheric correction of ENVISAT/MERIS data over inland waters: validation for European lakes. Remote Sens. Environ. 114, 467–480. https://doi.org/10.1016/j.rse.2009.10.004.
Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q., 2017. On Calibration of Modern Neural Networks. arXiv. doi.org/10.48550/arXiv.1706.04599.
Gupana, R.S., Odermatt, D., Cesana, I., Giardino, C., Nedbal, L., Damm, A., 2021. Remote sensing of sun-induced chlorophyll-a fluorescence in inland and coastal waters: current state and future prospects. Remote Sens. Environ. 262, 112482 https://doi.org/10.1016/j.rse.2021.112482.
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 Sens. Environ. 115, 3479–3490. https://doi.org/10.1016/j.rse.2011.08.011.
Härmä, P., Vepsäläinen, J., Hannonen, T., Pyhälahti, T., Kämäri, J., Kallio, K., Eloheimo, K., Koponen, S., 2001. Detection of water quality using simulated satellite data and semi-empirical algorithms in Finland. Sci. Total Environ. 268, 107–121. https://doi.org/10.1016/S0048-9697(00)00688-4.
Hastie, T., Friedman, J., Tibshirani, R., 2001. The Elements of Statistical Learning, Springer Series in Statistics. Springer, New York, New York, NY. https://doi.org/10.1007/978-0-387-21606-5.
Hooker, S.B., Morel, A., 2003. Platform and environmental effects on above-water determinations of water-leaving radiances. J. Atmos. Ocean. Technol. 20, 187–205. https://doi.org/10.1175/1520-0426(2003)020<0187:PAEEOA>2.0.CO;2.
Hooker, S.B., Van Heukelem, L., Thomas, C.S., Claustre, H., Ras, J., Schlüter, L., Clementson, L., Van Der Linde, D., Eker-Develi, E., Berthon, J.-F., Barlow, R., Sessions, H., Ismail, H., Perl, J., 2009. The Third SeaWiFS HPLC Analysis Round-Robin Experiment (SeaHARRE-3).
Hu, C., Lee, Z., Franz, B., 2012. Chlorophyll *a* algorithms for oligotrophic oceans: a novel approach based on three-band reflectance difference. J. Geophys. Res. Oceans 117. https://doi.org/10.1029/2011JC007395.
Huot, Y., Babin, M., Bruyant, F., Grob, C., Twardowski, M.S., Claustre, H., 2007. Does chlorophyll a provide the best index of phytoplankton biomass for primary productivity studies? Biogeosci. Discuss. 4, 707–745.
IOCCG, 2019. Uncertainties in Ocean Colour Remote Sensing. International Ocean-Colour Coordinating Group (IOCCG). Dartmouth, Canada.
Jiang, D., Matsushita, B., Yang, W., 2020. A simple and effective method for removing residual reflected skylight in above-water remote sensing reflectance measurements. ISPRS J. Photogramm. Remote Sens. 165, 16–27. https://doi.org/10.1016/j.isprsjprs.2020.05.003.
Jing, L., Tian, Y., 2021. Self-supervised visual feature learning with deep neural networks: a survey. IEEE Trans. Pattern Anal. Mach. Intell. 43, 4037–4058. https://doi.org/10.1109/TPAMI.2020.2992393.
Johnsen, G., Nelson, N.B., Jovine, R.V.M., Prézelin, B.B., 1994. Chromoprotein- and pigment-dependent modeling of spectral light absorption in two dinoflagellates, Prorocentrum minimum and Heterocapsa pygmaea. Mar. Ecol. Prog. Ser. 114 (3), 245–258.
Jordan, M.I., Ghahramani, Z., Jaakkola, T.S., Saul, L.K., 1999. An introduction to variational methods for graphical models. Mach. Learn. 37, 183–233. https://doi.org/10.1023/A:1007665907178.
Kallio, K., Koponen, S., Ylöstalo, P., Kervinen, M., Pyhälahti, T., Attila, J., 2015. Validation of MERIS spectral inversion processors using reflectance, IOP and water quality measurements in boreal lakes. Remote Sens. Environ. 157, 147–157. https://doi.org/10.1016/j.rse.2014.06.016.
Kendall, A., Gal, Y., 2017. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? arXiv: 1703.04977v2.
Kirk, J.T.O., 1994. Light and photosynthesis in aquatic ecosystems, third edition, Light and Photosynthesis in Aquatic Ecosystems, third edition. Cambridge University Press. https://doi.org/10.1017/CBO9781139168212.
Kravitz, J., Matthews, M., Bernard, S., Griffith, D., 2020. Application of Sentinel 3 OLCI for chl-a retrieval over small inland water targets: successes and challenges. Remote Sens. Environ. 237, 111562 https://doi.org/10.1016/j.rse.2019.111562.
Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P. J., Goetz, A.F.H., 1993. The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. In: AIP Conference Proceedings. AIP Publishing, pp. 192–201. https://doi.org/10.1063/1.44433.
Kuleshov, V., Fenner, N., Ermon, S., 2018. Accurate uncertainties for deep learning using calibrated regression. In: 35th International Conference on Machine Learning, ICML 2018 6, pp. 4369–4377.
Kutser, T., Paavel, B., Verpoorter, C., Ligi, M., Soomets, T., Toming, K., Casal, G., 2016. Remote sensing of black lakes and using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters. Remote Sens. 8 https://doi.org/10.3390/rs8060497.
Kutser, T., Vahtmäe, E., Paavel, B., Kauer, T., 2013. 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.
Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 6405–6416.
Li, Linhai, Li, Lin, Song, K., 2015. Remote sensing of freshwater cyanobacteria: an extended IOP Inversion Model of Inland Waters (IIMIW) for partitioning absorption coefficient and estimating phycocyanin. Remote Sens. Environ. 157, 9–23. https://doi.org/10.1016/j.rse.2014.06.009.
Li, Linhai, Li, Lin, Song, K., Li, Y., Tedesco, L.P., Shi, K., Li, Z., 2013. 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.
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. Environ. 267, 112710 https://doi.org/10.1016/j.rse.2021.112710.
Lutz, V.A., 2001. Changes in the in vivo absorption and fluorescence excitation spectra with growth irradiance in three species of phytoplankton. J. Plankton Res. 23, 555–569. https://doi.org/10.1093/plankt/23.6.555.
MacKay, D.J.C., 1992. Bayesian Interpolation. Neural Comput. 4, 415–447. https://doi.org/10.1162/neco.1992.4.3.415.
Manzo, C., Bresciani, M., Giardino, C., Braga, F., Bassani, C., 2015. Sensitivity analysis of a bio-optical model for Italian lakes focused on Landsat-8, Sentinel-2 and Sentinel-3. Eur. J. Remote Sens. 48, 17–32. https://doi.org/10.5721/EuJRS20154802.
Matsushita, B., Yang, W., Yu, G., Oyama, Y., Yoshimura, K., Fukushima, T., 2015. A hybrid algorithm for estimating the chlorophyll-a concentration across different trophic states in Asian inland waters. ISPRS J. Photogramm. Remote Sens. 102, 28–37. https://doi.org/10.1016/j.isprsjprs.2014.12.022.
Matthews, M.W., 2014. Eutrophication and cyanobacterial blooms in South African inland waters: 10years of MERIS observations. Remote Sens. Environ. 155, 161–177. https://doi.org/10.1016/j.rse.2014.08.010.
Matthews, M.W., Bernard, S., 2013. Characterizing the absorption properties for remote sensing of three small optically-diverse south african reservoirs. Remote Sens. 5, 4370–4404. https://doi.org/10.3390/rs5094370.
Matthews, M.W., Bernard, S., Robertson, L., 2012. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. Remote Sens. Environ. 124, 637–652. https://doi.org/10.1016/j.rse.2012.05.032.
Matthews, M.W., Odermatt, D., 2015. Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters. Remote Sens. Environ. 156, 374–382. https://doi.org/10.1016/j.rse.2014.10.010.
McKee, D., Röttgers, R., Neukermans, G., Calzado, V.S., Trees, C., Ampolo-Rella, M., Neil, C., Cunningham, A., 2014. Impact of measurement uncertainties on determination of chlorophyll-specific absorption coefficient for marine phytoplankton. J. Geophys. Res. Oceans 119, 9013–9025. https://doi.org/10.1002/2014JC009909.
Minaudo, C., Odermatt, D., Bouffard, D., Rahaghi, A.I., Lavanchy, S., Wüest, A., 2021. 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.
Mishra, S., Mishra, D.R., 2012. Normalized difference chlorophyll index: a novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. Remote Sens. Environ. 117, 394–406. https://doi.org/10.1016/j.rse.2011.10.016.
Mittenzwey, K.-H., Ullrich, S., Gitelson, A.A., Kondratiev, K.Y., 1992. Determination of chlorophyll a of inland waters on the basis of spectral reflectance. Limnol. Oceanogr. 37, 147–149. https://doi.org/10.4319/lo.1992.37.1.0147.
Mobley, C.D., 1999. Estimation of the remote-sensing reflectance from above-surface measurements. Appl. Opt. 38, 7442. https://doi.org/10.1364/ao.38.007442.
Moore, T.S., Dowell, M.D., Bradt, S., Verdú, A.R., 2014. An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. Remote Sens. Environ. 143, 97–111. https://doi.org/10.1016/j.rse.2013.11.021.
Moreno-Torres, J.G., Raeder, T., Alaiz-Rodríguez, R., Chawla, N.V., Herrera, F., 2012. A unifying view on dataset shift in classification. Pattern Recogn. 45, 521–530. https://doi.org/10.1016/j.patcog.2011.06.019.
Morley, S.K., Brito, T.V., Welling, D.T., 2018. Measures of model performance based on the log accuracy ratio. Space Weather 16, 69–88. https://doi.org/10.1002/2017SW001669.
Moses, W.J., Gitelson, A.A., Berdnikov, S., Povazhnyy, V., 2009. Estimation of chlorophyll-a concentration in case II waters using MODIS and MERIS data - successes and challenges. Environ. Res. Lett. 4, 45005. https://doi.org/10.1088/1748-9326/4/4/045005.
Mouw, C.B., Chen, H., McKinley, G.A., Effler, S., O'Donnell, D., Perkins, M.G., Strait, C., 2013. Evaluation and optimization of bio-optical inversion algorithms for remote sensing of Lake Superior's optical properties. J. Geophys. Res. Oceans 118, 1696–1714. https://doi.org/10.1002/jgrc.20139.
Neal, R.M., 1996. Bayesian learning for neural networks. Lecture Notes in Statistics, 118. University of Toronto, pp. 1–185.
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. Environ. 229, 159–178. https://doi.org/10.1016/j.rse.2019.04.027.

Niculescu-Mizil, A., Caruana, R., 2005. Predicting good probabilities with supervised learning. In: Proceedings of the 22nd International Conference on Machine Learning, pp. 625–632. https://doi.org/10.1145/1102351.1102430.
Nixon, J., Dusenberry, M., Jerfel, G., Nguyen, T., Liu, J., Zhang, L., Tran, D., 2019. Measuring Calibration in Deep Learning.
Nouchi, V., Odermatt, D., Wüest, A.J., Bouffard, D., 2018. Effects of non-uniform vertical constituent profiles on remote sensing reflectance of oligo- to mesotrophic lakes. Eur. J. Remote Sens. 51, 808–821. https://doi.org/10.1080/22797254.2018.1493360.
O'Reilly, J.E., Werdell, P.J., 2019. Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. Remote Sens. Environ. 229, 32–47. https://doi.org/10.1016/j.rse.2019.04.021.
O'Reilly, Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., McClain, C., 1998. Ocean color chlorophyll algorithms for SeaWiFS. J. Geophys. Res. Oceans 103, 24937–24953. https://doi.org/10.1029/98JC02160.
O'Shea, R.E., Pahlevan, N., Smith, B., Bresciani, M., Egerton, T., Giardino, C., Li, L., Moore, T., Ruiz-Verdu, A., Ruberg, S., Simis, S.G.H., Stumpf, R., Vaičiūtė, D., 2021. Advancing cyanobacteria biomass estimation from hyperspectral observations: demonstrations with HICO and PRISMA imagery. Remote Sens. Environ. 266, 112693 https://doi.org/10.1016/j.rse.2021.112693.
Odermatt, D., Danne, O., Philipson, P., Brockmann, C., 2018. Diversity II water quality parameters from ENVISAT (2002-2012): a new global information source for lakes. Earth Syst. Sci. Data 10, 1527–1549. https://doi.org/10.5194/essd-10-1527-2018.
Odermatt, D., Gitelson, A., Brando, V.E., Schaepman, M., 2012. Review of constituent retrieval in optically deep and complex waters from satellite imagery. Remote Sens. Environ. 118, 116–126. https://doi.org/10.1016/j.rse.2011.11.013.
OECD, 1982. Eutrophication of waters. Monitoring, assessment and control. Final report of the OECD cooperative programme on monitoring of inland waters (eutrophication control). Paris.
Ovadia, Y., Research, G, Fertig, E., Ren, J., Nado Google Research, Z, Nowozin Google Research, S, Dillon Google Research, J. V, Lakshminarayanan, B., Snoek, J., 2019. Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift, in. In: NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp. 14003–14014.
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., Ma, R., Martin-Lauzer, F.R., Olmanson, L., Oppelt, N., Pan, Y., Peters, S., Reynaud, N., Sander de Carvalho, L.A., Simis, S., Spyrakos, E., Steinmetz, F., Stelzer, K., Sterckx, S., Tormos, T., Tyler, A., Vanhellemont, Q., Warren, M., 2021a. 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.
Pahlevan, N., Smith, B., Alikas, K., Anstee, J., Barbosa, C., Binding, C., Bresciani, M., Cremella, B., Giardino, C., Gurlin, D., Fernandez, V., Jamet, C., Kangro, K., Lehmann, M.K., Loisel, H., Matsushita, B., Hà, N., Olmanson, L., Potvin, G., Simis, S. G.H., VanderWoude, A., Vantrepotte, V., Ruiz-Verdù, A., 2022. 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.
Pahlevan, N., Smith, B., Binding, C., Gurlin, D., Li, L., Bresciani, M., Giardino, C., 2021b. Hyperspectral retrievals of phytoplankton absorption and chlorophyll-a in inland and nearshore coastal waters. Remote Sens. Environ. 253, 112200 https://doi.org/10.1016/j.rse.2020.112200.
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. Environ. 240, 111604 https://doi.org/10.1016/j.rse.2019.111604.
Paulino, R.S., Martins, V.S., Novo, E.M.L.M., Barbosa, C.C.F., de Carvalho, L.A.S., Begliomini, F.N., 2022. Assessment of adjacency correction over inland waters using Sentinel-2 MSI images. Remote Sens. 14, 1829. https://doi.org/10.3390/rs14081829.
Pepe, M., Giardino, C., Borsani, G., Cardoso, A.C., Chiaudani, G., Premazzi, G., Rodari, E., Zilioli, E., 2001. Relationship between apparent optical properties and photosynthetic pigments in the sub-alpine Lake Iseo. Sci. Total Environ. 268, 31–45. https://doi.org/10.1016/S0048-9697(00)00691-4.
Pereira-Sandoval, M., Ruescas, A., Urrego, P., Ruiz-Verdú, A., Delegido, J., Tenjo, C., Soria-Perpinyà, X., Vicente, E., Soria, J., Moreno, J., 2019. Evaluation of atmospheric correction algorithms over Spanish inland waters for Sentinel-2 multi spectral imagery data. Remote Sens. 11, 1469. https://doi.org/10.3390/rs11121469.
Platt, J.C., 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in Large Margin Classifiers. MIT Press, pp. 61–74.
Poikane, S., Alves, M.H., Argillier, C., Van Den Berg, M., Buzzi, F., Hoehn, E., De Hoyos, C., Karottki, I., Laplace-Treyture, C., Solheim, A.L., Ortiz-Casas, J., Ott, I., Phillips, G., Pilke, A., Pádua, J., Remec-Rekar, S., Riedmüller, U., Schaumburg, J., Serrano, M.L., Soszka, H., Tierney, D., Urbanic, G., Wolfram, G., 2010. Defining chlorophyll-a reference conditions in European Lakes. Environ. Manag. 45, 1286–1298. https://doi.org/10.1007/s00267-010-9484-4.
Roesler, C.S., Barnard, A.H., 2013. Optical proxy for phytoplankton biomass in the absence of photophysiology: rethinking the absorption line height. Methods Oceanogr. 7, 79–94. https://doi.org/10.1016/j.mio.2013.12.003.
Ruiz-Verdú, A., Dominguez-Gomez, J.-A., Pena-Martinez, R., 2005. Use of CHRIS for monitoring water quality in rosarito reservoir. In: ESA Special Publication, ESA Special Publication, p. 26.
Ruiz-Verdú, A., Simis, S.G.H., de Hoyos, C., Gons, H.J., Peña-Martínez, R., 2008. An evaluation of algorithms for the remote sensing of cyanobacterial biomass. Remote Sens. Environ. 112, 3996–4008. https://doi.org/10.1016/j.rse.2007.11.019.
Rusak, J.A., Tanentzap, A.J., Klug, J.L., Rose, K.C., Hendricks, S.P., Jennings, E., Laas, A., Pierson, D., Ryder, E., Smyth, R.L., White, D.S., Winslow, L.A., Adrian, R., Arvola, L., de Eyto, E., Feuchtmayr, H., Honti, M., Istvánovics, V., Jones, I.D., McBride, C.G., Schmidt, S.R., Seekell, D., Staehr, P.A., Zhu, G., 2018. Wind and trophic status explain within and among-lake variability of algal biomass. Limnol. Oceanogr. Lett. 3, 409–418. https://doi.org/10.1002/lol2.10093.
Saul, L.K., Jaakkola, T., Jordan, M.I., 1996. Mean field theory for sigmoid belief networks. J. Artif. Intell. Res. 4, 61–76. https://doi.org/10.1613/jair.251.
Schaeffer, B., Salls, W., Coffer, M., Lebreton, C., Werther, M., Stelzer, K., Urquhart, E., Gurlin, D., 2022. Merging of the Case 2 Regional Coast Colour and Maximum-Peak Height chlorophyll-a algorithms: validation and demonstration of satellite-derived retrievals across US lakes. Environ. Monit. Assess. 194, 1–25. https://doi.org/10.1007/s10661-021-09684-w.
Schalles, J.F., 2006. Optical remote sensing techniques to estimate phytoplankton chlorophyll a concentrations in coastal waters with varying suspended matter and cdom concentrations. In: Remote Sensing and Digital Image Processing, Remote Sensing and Digital Image Processing. Springer International Publishing, pp. 27–79.
Schalles, J.F., Hladik, C.M., 2012. Mapping phytoplankton chlorophyll in turbid, Case 2 estuarine and coastal waters. Israel J. Plant Sci. 60, 169–191. https://doi.org/10.1560/IJPS.60.1-2.169.
Seegers, B.N., Stumpf, R.P., Schaeffer, B.A., Loftin, K.A., Werdell, P.J., 2018. Performance metrics for the assessment of satellite data products: an ocean color case study. Opt. Express 26, 7404. https://doi.org/10.1364/oe.26.007404.
Seegers, B.N., Werdell, P.J., Vandermeulen, R.A., Salls, W., Stumpf, R.P., Schaeffer, B.A., Owens, T.J., Bailey, S.W., Scott, J.P., Loftin, K.A., 2021. Satellites for long-term monitoring of inland U.S. lakes: the MERIS time series and application for chlorophyll-a. Remote Sens. Environ. 266, 112685 https://doi.org/10.1016/j.rse.2021.112685.
Simis, S.G.H., Olsson, J., 2013. Unattended processing of shipborne hyperspectral reflectance measurements. Remote Sens. Environ. 135, 202–212. https://doi.org/10.1016/j.rse.2013.04.001.
Simis, S.G.H., Peters, S.W.M., Gons, H.J., 2005. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. Limnol. Oceanogr. 50, 237–245. https://doi.org/10.4319/lo.2005.50.1.0237.
Simis, S.G.H., Ruiz-Verdú, A., Domínguez-Gómez, J.A., Peña-Martinez, R., Peters, S.W. M., Gons, H.J., 2007. Influence of phytoplankton pigment composition on remote sensing of cyanobacterial biomass. Remote Sens. Environ. 106, 414–427. https://doi.org/10.1016/j.rse.2006.09.008.
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., Roger, J.-C., Scaramuzza, P., Vermote, E., Vesel, N., Zupanc, A., Žust, L., 2022. Cloud Mask Intercomparison eXercise (CMIX): an evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. Remote Sens. Environ. 274, 112990 https://doi.org/10.1016/j.rse.2022.112990.
Smith, B., Pahlevan, N., Schalles, J., Ruberg, S., Errera, R., Ma, R., Giardino, C., Bresciani, M., Barbosa, C., Moore, T., Fernandez, V., Alikas, K., Kangro, K., 2021. A chlorophyll-a algorithm for landsat-8 based on mixture density networks. Front. Remote Sens. 1, 5. https://doi.org/10.3389/frsen.2020.623678.
Smith, M.E., Robertson Lain, L., Bernard, S., 2018. An optimized chlorophyll a switching algorithm for MERIS and OLCI in phytoplankton-dominated waters. Remote Sens. Environ. 215, 217–227. https://doi.org/10.1016/j.rse.2018.06.002.
Soomets, T., Uudeberg, K., Jakovels, D., Brauns, A., Zagars, M., Kutser, T., 2020. Validation and comparison of water quality products in baltic lakes using Sentinel-2 MSI and Sentinel-3 OLCI data. Sensors 20, 742. https://doi.org/10.3390/s20030742.
Soulignac, F., Danis, P.-A., Bouffard, D., Chanudet, V., Dambrine, E., Guénand, Y., Harmel, T., Ibelings, B.W., Trevisan, D., Uittenbogaard, R., Anneville, O., 2018. Using 3D modeling and remote sensing capabilities for a better understanding of spatio-temporal heterogeneities of phytoplankton abundance in large lakes. J. Great Lakes Res. 44, 756–764. https://doi.org/10.1016/j.jglr.2018.05.008.
Spyrakos, E., O'Donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis, G.H.S., 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. M., Ogashawara, I., Ruiz-Verdú, A., Schalles, J.F., Tebbs, E., Zhang, Y., Tyler, A.N., 2018. Optical types of inland and coastal waters. Limnol. Oceanogr. 63 (2), 846–870. https://doi.org/10.1002/lno.10674.
Srivastava, N., Hinton, G., Krizhevsky, A., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (56), 1929–1958.
Stedmon, C.A., Markager, S., Kaas, H., 2000. Optical properties and signatures of chromophoric dissolved organic matter (CDOM) in Danish coastal waters. Estuar. Coast. Shelf Sci. 51, 267–278. https://doi.org/10.1006/ecss.2000.0645.
Steinmetz, F., Deschamps, P.-Y., Ramon, D., 2011. Atmospheric correction in presence of sun glint: application to MERIS. Opt. Express 19, 9783. https://doi.org/10.1364/oe.19.009783.
Sterckx, S., Knaeps, E., Ruddick, K., 2011. Detection and correction of adjacency effects in hyperspectral airborne data of coastal and inland waters: The use of the near infrared similarity spectrum. Int. J. Remote Sens. 32, 6479–6505. https://doi.org/10.1080/01431161.2010.512930.

Sterner, R.W., Keeler, B., Polasky, S., Poudel, R., Rhude, K., Rogers, M., 2020. Ecosystem services of Earth's largest freshwater lakes. Ecosystem Services 41, 101046. https://doi.org/10.1016/j.ecoser.2019.101046.

Suits, G.H., 1975. The nature of electromagnetic radiation. In: Reeves, R.G. (Ed.), Manual of Remote Sensing. American Society of Photogrammetry, Falls Church, Va, pp. 51–73.

Tebbs, E.J., Avery, S.T., Chadwick, M.A., 2020. Satellite remote sensing reveals impacts from dam-associated hydrological changes on chlorophyll- *a* in the world's largest desert lake. River Res. Appl. 36, 211–222. https://doi.org/10.1002/rra.3574.

Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., Nõges, T., 2016. First experiences in mapping lakewater quality parameters with sentinel-2 MSI imagery. Remote Sens. 8, 1–14. https://doi.org/10.3390/rs8080640.

Tran, K., Neiswanger, W., Yoon, J., Zhang, Q., Xing, E., Ulissi, Z.W., 2020. Methods for comparing uncertainty quantifications for material property predictions.

Urquhart, E.A., Schaeffer, B.A., Stumpf, R.P., Loftin, K.A., Werdell, P.J., 2017. A method for examining temporal changes in cyanobacterial harmful algal bloom spatial extent using satellite remote sensing. Harmful Algae 67, 144–152. https://doi.org/10.1016/j.hal.2017.06.001.

Van Heukelem, L., Thomas, C.S., 2001. Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. J. Chromatogr. A 910, 31–49. https://doi.org/10.1016/S0378-4347(00)00603-4.

Vörös, L., Padisák, J., 1991. Phytoplankton biomass and chlorophyll-a in some shallow lakes in central Europe. Hydrobiologia 215, 111–119. https://doi.org/10.1007/BF00014715.

Wang, D., Ma, R., Xue, K., Loiselle, S.A., 2019. The Assessment of Landsat-8 OLI Atmospheric Correction Algorithms for Inland Waters. Remote Sens. 11, 169. https://doi.org/10.3390/rs11020169.

Wang, S., Li, J., Zhang, B., Spyrakos, E., Tyler, A.N., Shen, Q., Zhang, F., Kuster, T., Lehmann, M.K., Wu, Y., Peng, D., 2018. Trophic state assessment of global inland waters using a MODIS-derived Forel-Ule index. Remote Sens. Environ. 217, 444–460.

Wang, Y., Albrecht M., C., Braham A., A., N., Mou, L., Zhu, X., 2022. Self-Supervised Learning in Remote Sensing: A Review. IEEE Geoscience and Remote Sensing Magazine 2–36. https://doi.org/10.1109/MGRS.2022.3198244.

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. Environ. 225, 267–289. https://doi.org/10.1016/j.rse.2019.03.018.

Warren, M.A., Simis, S.G.H., Selmes, N., 2021. Complementary water quality observations from high and medium resolution Sentinel sensors by aligning chlorophyll-a and turbidity algorithms. Remote Sens. Environ. 265, 112651 https://doi.org/10.1016/j.rse.2021.112651.

Werdell, P.J., McKinna, L.I.W., Boss, E., Ackleson, S.G., Craig, S.E., Gregg, W.W., Lee, Z., Maritorena, S., Roesler, C.S., Rousseaux, C.S., Stramski, D., Sullivan, J.M., Twardowski, M.S., Tzortziou, M., Zhang, X., 2018. An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing. Prog. Oceanogr. https://doi.org/10.1016/j.pocean.2018.01.001.

Werther, M., Odermatt, D., Simis, S.G.H., Gurlin, D., Jorge, D.S.F., Loisel, H., Hunter, P. D., Tyler, A.N., Spyrakos, E., 2022. Characterising retrieval uncertainty of chlorophyll-a algorithms in oligotrophic and mesotrophic lakes and reservoirs. ISPRS J. Photogramm. Remote Sens. 190, 279–300. https://doi.org/10.1016/j.isprsjprs.2022.06.015.

Werther, M., Spyrakos, E., Simis, S.G.H., Odermatt, D., Stelzer, K., Krawczyk, H., Berlage, O., Hunter, P., Tyler, A., 2021. Meta-classification of remote sensing reflectance to estimate trophic status of inland and nearshore waters. ISPRS J. Photogramm. Remote Sens. 176, 109–126. https://doi.org/10.1016/j.isprsjprs.2021.04.003.

Wevers, J., Müller, D., Scholze, J., Kirches, G., Quast, R., Brockmann, C., 2021. IdePix for Sentinel-2 MSI Algorithm Theoretical Basis Document. https://doi.org/10.5281/ZENODO.5788067.

Yacobi, Y.Z., Moses, W.J., Kaganovsky, S., Sulimani, B., Leavitt, B.C., Gitelson, A.A., 2011. NIR-red reflectance-based algorithms for chlorophyll-a estimation in mesotrophic inland and coastal waters: Lake Kinneret case study. Water Res. 45, 2428–2436. https://doi.org/10.1016/j.watres.2011.02.002.

Yao, J., Pan, W., Ghosh, S., Doshi-Velez, F., 2019. Quality of Uncertainty Quantification for Bayesian Neural Network Inference. arXiv.

Zadrozny, B., Elkan, C., 2002. Transforming classifier scores into accurate multiclass probability estimates. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery (ACM), New York, New York, USA, pp. 694–699. https://doi.org/10.1145/775047.775151.

Zadrozny, Bianca, Elkan, C., 2001. Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. In: Proceedings of the Eighteenth International Conference on Machine Learning, pp. 609–616.

Zhu, X., Goldberg, A.B., 2009. Introduction to Semi-Supervised Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning 3, 1–130. https://doi.org/10.2200/S00196ED1V01Y200906AIM006.