Resolving phytoplankton pigments from spectral images using convolutional neural networks

Pauliina Salmi [1]*, **Ilkka Pölönen,**[1] **Daniel Atton Beckmann,**[2] **Marco L. Calderini,**[3] **Linda May,**[4] **Justyna Olszewska,**[4] **Laura Perozzi,**[2] **Salli Pääkkönen,**[1] **Sami Taipale,**[3] **Peter Hunter**[2]

[1]Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland
[2]Faculty of Natural Sciences, University of Stirling, Stirling, Scotland
[3]Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland
[4]Centre for Ecology & Hydrology, Edinburgh, Scotland

Abstract

Motivated by the need for rapid and robust monitoring of phytoplankton in inland waters, this article introduces a protocol based on a mobile spectral imager for assessing phytoplankton pigments from water samples. The protocol includes (1) sample concentrating; (2) spectral imaging; and (3) convolutional neural networks (CNNs) to resolve concentrations of chlorophyll *a* (Chl *a*), carotenoids, and phycocyanin. The protocol was demonstrated with samples from 20 lakes across Scotland, with special emphasis on Loch Leven where blooms of cyanobacteria are frequent. In parallel, samples were prepared for reference observations of Chl *a* and carotenoids by high-performance liquid chromatography and of phycocyanin by spectrophotometry. Robustness of the CNNs were investigated by excluding each lake from model trainings one at a time and using the excluded data as independent test data. For Loch Leven, median absolute percentage difference (MAPD) was 15% for Chl *a* and 36% for carotenoids. MAPD in estimated phycocyanin concentration was high (102%); however, the system was able to indicate the possibility of a cyanobacteria bloom. In the leave-one-out tests with the other lakes, MAPD was 26% for Chl *a*, 27% for carotenoids, and 75% for phycocyanin. The higher error for phycocyanin was likely due to variation in the data distribution and reference observations. It was concluded that this protocol could support phytoplankton monitoring by using Chl *a* and carotenoids as proxies for biomass. Greater focus on the distribution and volume of the training data would improve the phycocyanin estimates.

Seasonal development of phytoplankton is of central interest because phytoplankton are principal primary producers in water bodies and have a major role in cycling of nutrients and energy in aquatic food webs. Phytoplankton exploit photosynthetically active radiation (PAR), and in variable light environments different light-harvesting strategies are favored (Kirk 2011; Reynolds 2006). Chlorophylls,

especially Chl *a*, are the most abundant light-harvesting pigments in eukaryotic phytoplankton. In contrast, carotenoid pigments of phytoplankton have a significant effect on their optical properties (Hoepffner and Sathyendranath 1991; Nair et al. 2008; Brito et al. 2015). In addition to chlorophylls and carotenoids, cyanobacteria and cryptophytes have light-harvesting complexes that consist of phycobiliproteins. Phycocyanin is an essential phycobiliprotein in almost all cyanobacteria, and is, therefore, referred as c-phycocyanin (Watanabe and Ikeuchi 2013; Stadnichuk et al. 2015). Broadly, the pigment combinations of phytoplankton confirm the taxonomic composition, although the taxonomic resolution of pigment analysis remains often at the level of phylum or in some cases non-phyletic group (Irigoien et al. 2004; Reynolds 2006; Wright and Jeffrey 2006). However, concentrations and ratios of total chlorophylls or Chl *a*, total or major carotenoids and c-phycocyanin could be informative enough to indicate the rough structure of phytoplankton and occurrence of a harmful algae bloom (Millie et al. 1992; Sathyendranath et al. 2005). This strategy is a

*Correspondence: pauliina.u.m.salmi@jyu.fi

Additional Supporting Information may be found in the online version of this article.

potential tool for monitoring that could help to detect harmful algae blooms in their initial stages enabling rapid management actions.

Microscopy is a routinely used and standardized method for monitoring phytoplankton composition and biomass (EN 15204 2006) and spectrophotometry to assess Chl *a* biomass (ISO 10260 1992), but spectroscopic methods, targeted to resolve the pigment-related signals, have also been developed as important on-site tools for aquatic monitoring (Brient et al. 2008; Rode et al. 2016; Möller et al. 2019). Recent technological developments have enabled sensors to be developed for higher spectral resolution; in addition to single-channel fluorometers, multiexciters, and spectroradiometers yield robust estimates of the occurrences of phytoplankton pigment groups (Möller et al. 2019). In addition to on-site methods, satellites carrying hyperspectral imagers are currently being launched (Giardino et al. 2019; Chabrillat et al. 2020). A hyperspectral imager produces a dataset, also called a data cube, in which each point on the imaged surface contains information about its reflectance or transmittance across a variety of wavelengths. The functional principle and practical size of spectral imagers offer a possibility for laboratory use, too (Legleiter et al. 2022). Spectral imagers yield information of the target's spatial distribution, and the measurement geometry is adjustable to different sample sizes from individual samples to the Earth's orbit (Legleiter et al. 2022). This, together with a capability of estimating both phytoplankton composition and biomass with low effort compared to the standardized methods, makes spectral imagers attractive as new tools for providing environmental monitoring solutions.

Laboratory-based biomolecule or microscopy techniques are laborious and require special expertise, leading to low spatial and temporal coverage. Especially microscopy-based assessments could be biased based on the person conducting the analysis (Vuorio et al. 2007) or have wide confidence intervals even when conducted by one person (Salonen et al. 2021). On-site sensors require frequent maintenance, calibration, and protection against vandalism (Rode et al. 2016). Remote sensing of inland waters has limitations due spatial, temporal, and spectral resolutions; and although these are currently being developed (Mouw et al. 2015; Giardino et al. 2019), weather conditions may still limit observations. Therefore, there is a niche for new, rapid, and robust monitoring methods that are applicable year-round in a variety of water bodies.

All spectroscopic methods, from laboratory to the field and remote sensing, involve the challenge of resolving phytoplankton-related signals from the measurements when each pixel includes information from several different sources. Light propagates non-linearly through a phytoplankton community, as it scatters from particle to particle, and is refracted and absorbed by different substances, including dissolved compounds (Kirk 2011). Therefore, physical models to resolve, for example, cell or pigment concentrations, have been described as being simplifications of complex phytoplankton

communities (Bricaud et al. 2007; Pyo et al. 2019; Werther et al. 2022). Machine learning algorithms, in contrast, are computational models that make predictions or generate new data based on the data that they have been trained with. Among machine learning algorithms, convolutional neural networks (CNNs) have potential application for resolving information from spectral images because the convolution filters separate the irrelevant information when trained successfully (Goodfellow et al. 2016). The filtered data are then fed into a neural network that consist of nodes arranged in layers, with the output layer being the final predictions. During model training, optimization algorithms try to minimize the error between the model estimates and the observed values. CNNs are widely used to classify phytoplankton from images (Henrichs et al. 2021; Kraft et al. 2022) but have also been used to establish calibrations between reflectance spectra and Chl *a* (Aptoula and Ariman 2021) or phycocyanin (Pyo et al. 2019). The combination of a spectral imager and machine learning are becoming the direction of development in monitoring cultivated microalgae (Solovchenko 2023); however, this approach could be beneficial for environmental motoring, too.

This paper introduces and evaluates the use of a commercial hyperspectral imager in the laboratory to image water samples that had been concentrated by centrifugation. Three separate one-dimensional (1D) CNNs were trained and tested to resolve concentrations of Chl *a*, carotenoids and phycocyanin from the spectral data. The protocol developed was demonstrated and tested by sampling 20 Scottish lakes across a variety of water colors and trophic states, with special emphasis on monitoring the development of early summer phytoplankton in Loch Leven, Scotland. Loch Leven is an important loch in many respects, including the natural habitats and recreational opportunities that it provides (Spears et al. 2022). Therefore, protecting the lake and mitigating reoccurring cyanobacteria blooms is a priority around Loch Leven. The protocol described here could support the current microscopy and on-site monitoring, and remote sensing practices by offering a robust, fast, and taxonomically informative method with minimal, and non-destructive, sample processing or need for instrument maintenance. High-performance liquid chromatography (HPLC) that yields information on composition and concentrations of carotenoids and Chl *a* in the most detailed level of the current analysis methods (Wright and Jeffrey 2006) was used as the reference method for chlorophyll and carotenoids. Phycocyanin is typically assessed with spectroscopic methods from extracts (Horváth et al. 2013; Sobiechowska-Sasim et al. 2014), and it was chosen as the reference method for c-phycocyanin.

Materials and procedures

Sampling sites

Water samples were collected from 20 lochs in Scotland by grab sampling from the surface water layers between May and June 2022 (Supplementary Table 1, sampling times and

locations). Sampling sites were located along the lake shores, except at Loch Leven, where boats were used to collect samples from the open water. From one to three 2-liter containers were submerged until full, being careful not to disturb the lake bottom. Samples were stored at $+4°C$ and processed in the laboratory within 48 h of collection. Samples from separate containers were combined and mixed carefully before being sub-sampled for imaging or reference pigment assessments. Each lake was sampled once, apart from Airthrey Loch and Gartmorn Dam Reservoir, which were sampled twice, and Loch Leven which was sampled at one to three different sampling sites (Harbour basin, Harbour Jetty, Pelagial, Kirkgate Park, Reed Bower, Sluices) on five different sampling occasions.

Sample concentrating and spectral imaging

Samples for hyperspectral imaging were concentrated by centrifuging 30–480 mL of lake water (3500 g, 10 min, room temperature) in 15-mL centrifuge tubes using a swing out rotor to ensure formation of a pellet at the bottom of the tube. The pellet was suspended in 2 mL of the supernatant and up to three replicates of each lake water sample was pipetted onto a 24-well plate (Sarstedt) for the hyperspectral imaging. One replicate of supernatant from each sampling occasion was placed similarly in a sample well as a background to represent the substances that did not settle during the centrifugation.

Spectral images were taken of the 24-well plates under transmission light with Specim IQ imager (Specim, Finland). Specim IQ is a mobile spectral camera for VNIR 400–1000 nm range with

204 spectral bands and spectral resolution of 7 nm full width at half maximum (FWHM). The imaging arrangement (Fig. 1) included a broadband halogen light source (Fiber-Lite, DC-950, Dolan-Jenner) with a diffusor plate (Dolan-Jenner). Light setup was 50% of the halogen's maximum and imager's exposure time 12 ms. The distance between the imager and the diffusor plate was 15 cm, producing a spatial pixel size of approximately 0.2 mm × 0.2 mm.

Raw images were normalized between the imager's internal dark reference and a spectral image of the illuminated diffusor plate, alone. The lower and upper ends of the spectra contained higher levels of radiometric noise; therefore, the spectral data were truncated to between 420 and 800 nm resulting to 150 spectral bands. Absorbance (A) was calculated from transmittance (T) in accordance with Eq. 1:

$$A = -\log T \tag{1}$$

A region of interest (ROI) of 50×50 pixels was cropped from each sample, taking care not to include the edges of the wells (Fig. 2A).

Observed pigment concentrations

Chlorophylls and carotenoids were assessed with HPLC. Three replicate samples of each lake (or six from Westfield Power Plant Reservoir) were collected on glass fiber filters (GF/F Whatman or Fisherbrand MF300) so that the filters were not clogged but were notably colored. The filters were folded to enclose the samples and wrapped in aluminum foil before flash-freezing them by submerging them in liquid nitrogen and then storing them at $-80°C$. Samples were shipped on dry-ice to the analysis laboratory (DHI, Denmark), where they were stored at $-80°C$ and analyzed withing 4–6 months of collection. In the analysis laboratory, the filters were transferred to vials containing 6 mL of 95% acetone with an internal standard (Vitamin E) added. The samples were mixed on a vortex mixer, sonicated on ice, extracted at $4°C$ for 20 h, and then mixed again. The samples were then filtered through 0.2-μm Teflon syringe filter into HPLC vials and placed in the cooling rack of the HPLC. Buffer and extract were injected into the HPLC (Shimadzu LC-10A HPLC system with LC Solution software) in the ratio 5 : 2 using a pretreatment program and mixing before injection. The HPLC method used was according to Van Heukelem and Thomas (2005). The analysis included the following pigments: chlorophyll $c2$, chlorophyll $c1$, chlorophyllide a, pheophorbide a, peridinin, ocillaxanthin, fucoxanthin, neoxanthin, aphanizophyll, violaxanthin, astaxanthin, diadinoxanthin, dinoxanthin, myxoxanthophyll, antheraxanthin, alloxanthin, diatoxanthin, zeaxanthin, lutein, canthaxanthin, chlorophyll b (Chl b), Chl a, pheophytin a, α-carotene, and β-carotene. After these initial analyses, Chl a was assessed separately because it is used widely as a proxy for phytoplankton biomass. Other carotenoids analyzed were pooled together and referred as carotenoids; these exclude other chlorophylls,
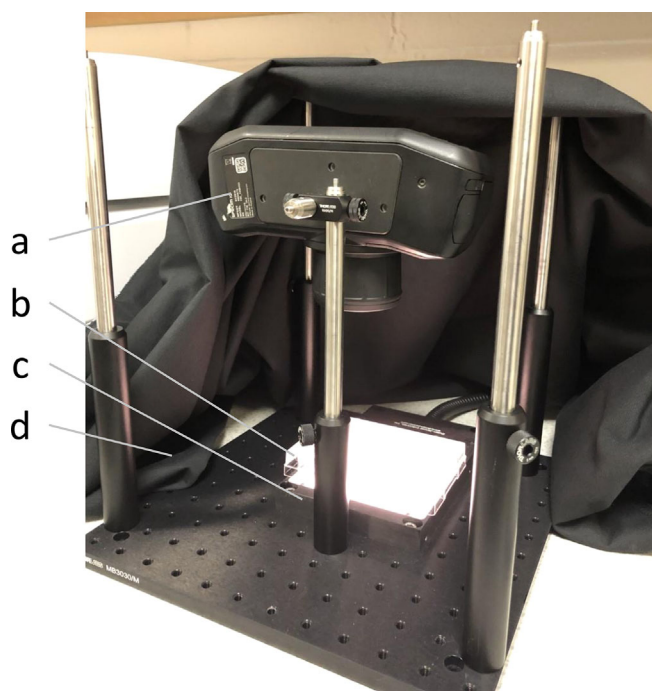

Fig. 1. The imaging arrangement: a—spectral imager, b—sample plate, c—diffusor plate, d—light impermeable hood.
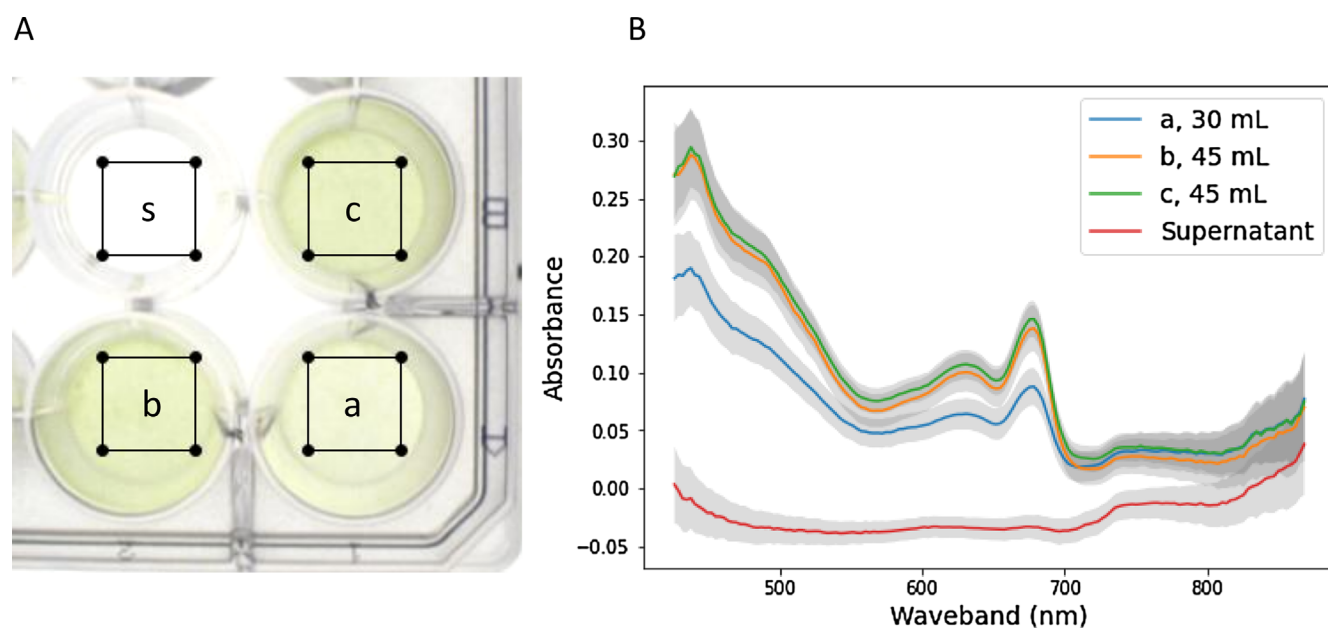
Fig. 2. An example of a sample plate visualized by combining a red, blue, and green wavebands to form an RGB image (**A**) and mean spectra calculated across the region of interest (**B**). The black rectangles in panel A illustrate the region of interest. The example shows three replicate centrifuged samples (a–c) and the supernatant (s) from a Loch Leven sample. The volumes in panel B are the initial sample volume that was concentrated into the 2 mL sub-samples in the sample wells.

chlorophyllide *a*, pheophorbide *a* and pheophytin *a* because their absorbance spectra are similar to Chl *a* in the visual wavebands (Clementson and Wojtasiewicz 2019).

Phycocyanin concentration was analyzed spectrophotometrically from extracts following protocol "E" of Horváth et al. (2013). Samples for phycocyanin were collected on glass fiber filters (GF/F Whatman or Fisherbrand MF300) and stored at $-20°$C before shipping on dry ice to the analysis laboratory; here they were again stored at $-20°$C and analyzed within 3–6 months of collection. The stock solutions for the extraction buffers were made by weighing 15.7 g of monobasic $NaH_2PO_4 \bullet 2H_2O$ and filling to 500 mL with ultrapure H_2O, and by adding 26.825 g of dibasic $Na_2HPO_4 \bullet 7H_2O$ to a separate flask filled to 500 mL with ultrapure H_2O. The extraction buffer was made by combining 306 mL of the monobasic stock and 294 mL of the dibasic stock resulting in a pH of 6.7–6.8. Fresh stock solutions and extraction buffer were prepared for each day when phycocyanin was extracted from the samples. Each frozen sample filter was chopped in a mortar, then immersed in the buffer and ground to a pulp with a pestle. After that grinding was continued for 1 min. The mortar was rinsed with the buffer so that the total extraction volume became 9–18 mL. Samples were frozen at $-20°$C and thawed. Immediately after thawing, the samples were sonicated in a water bath (VWR Utrasonic Cleaner) for 10 min. Crushed ice was added to the sonicator to keep the samples cool. Sonicated samples were centrifuged for 10 min in a centrifuge with a 3500*g* with swing out rotor. The supernatant containing the extracted pigments was decanted into a 10-mL syringe

with a syringe filter of 0.2 μm pore size. The first nine samples were filtered without centrifugation; however, it was found that centrifugation was a practical of preventing the syringe filters from becoming clogged. The samples were injected directly in a quartz cuvette with 1 or 5 cm diameter. Absorbance (*A*) was measured at 615, 652 and 750 nm against the extraction buffer using a Shimadzu UV-1800 spectrophotometer. Phycocyanin concentration (C_e) in the sample cuvette was calculated according to the equation of Bennett and Bogorad (1973):

$$C_e = [(A_{615} - A_{750}) - 0.474 \times (A_{652} - A_{750})]/5.32 \qquad (2)$$

Data augmentation for modeling

The samplings yielded 1–3 imaged replicate sample wells. Image data were augmented by subsampling and simulating more data by using a spectral mixture model, similar to that used by Salmi et al. (2022), to train the 1D CNN efficiently. The 50×50-pixel ROIs were constrained to smaller, 10×10 pixel, ROIs so that each 50×50-pixel ROI resulted to 25 smaller ROIs. Samples of the supernatants were included in the data, with the corresponding pigment concentrations being zero. This resulted in 2225 samples. Mean absorbance spectra were calculated across the smaller ROIs. A spectral mixture model was created to randomly sum two mean spectra ($x_i + x_j$, where $i, j \in [0, 2224]$) and the corresponding ground truths ($m_i + m_j$, where $i, j \in [0, 2224]$). The mixture model resulted in groups of spectra (X_{sim}) and ground truths (M_{sim}). The sizes of X_{sim} and M_{sim} were 10,000 samples. Generating data this way is possible

because absorbance is additive; the created data contain the same sources of variation as the original data, although the new combinations might not always be ecologically likely.

Pigment estimation using 1D CNN

Modeling was done with Python version 3.8.8, Jupyter notebooks, Keras 2.4.0 library, and Tensorflow 2.4.1 backend. Nvidia Tesla V100-SXM2 16 GB GPU units were used for computing. The modeling demonstrations of this study consisted of (1) model hyperparameter validation and (2) tests using the L-O-O method. All of the spectral data were min-max normalized using the minimum and maximum of the X_{sim} to facilitate learning.

1. *Hyperparameter validation*: Hyperparameter validations were done using the X_{sim} and M_{sim} that were divided so that 80% were used for training and 20% for validation. Different model architectures were trained by adding one convolution or dense layer at a time until validation loss (mean squared error [MSE]) stopped decreasing (Supplementary Table 2). The values set for different hyperparameters had to be limited, because the number of possible combinations would have been computationally impractical. Filter and node counts were powers of two (32, 64, 128, 256, 512) to make the iteration efficient. A maxpooling layer of size 2 was added after each convolution layer to simplify the data. After the expedient numbers of convolution and dense layers were obtained (Supplementary Table 2), further hyperparameter tuning was done using Keras Tuner Random Search. The hyperparameters tuned with the Random Search were convolution filter count, node count in the dense layers, kernel size and learning rate. Filter and node counts between 32 and 512 were validated with step 32. Kernel sizes three and five and learning rates 0.01 and 0.001 were validated. A 100 different models per pigment were trained for a 100 epochs with a patch size 32. MSE was used as the loss function and the hyperparameters were tuned using minimum validation loss as the objective. The most valid model architectures (Tables 1–3, see Supplementary Fig. 1 for training and loss curves) selected this way was used for the leave-one-out tests.

2. *Leave-one-out tests*: The capability of the most adequate model architectures (Tables 1–3) to adapt to different lakes was tested using the L-O-O method, where the lake that was left out was used as independent test data and excluded from the X_{sim} and M_{sim}. Therefore, the trained model had not used data from the test lake, even as a part of a modeled mixture. The data used for testing were non-augmented mean spectra from the 50×50-pixel ROIs and their corresponding observed pigment concentrations. The models were tested on the 50×50-pixel ROIs because these represent how the method would be used in practise, in contrast to the augmented training data. The mean spectra were min-max normalized using the minimum and maximum of X_{sim}. The L-O-O tests resulted to 20 different

trainings of the architectures. Loch Leven was sampled five times between May and June to study the potential of the protocol to support regular phytoplankton monitoring; therefore, the L-O-O test results for Loch Leven are given separately from the results for the other lakes.

Error metrics

1. *Measurement error*: Coefficient of variation (CV, %) was determined for mean absorbance spectra and HPLC and spectrophotometry assessments where three replicates were prepared. Coefficient of variation was calculated as follows:

$$CV, \% = \left(\frac{SD}{\mu} \right) \times 100 \qquad (3)$$

where μ is the mean of the replicate assessments and SD the standard deviation. Replicates were from separate filtrations and separate tubes during centrifugations; this way, the coefficient of variation included variation due to sample processing and measurement. Average distance from the mean was calculated for samplings that had two replicates instead of three. The coefficient of variation or distance from the mean for the imaging was calculated as the average absorbance spectrum across the 50×50 pixel area of the imaged sample well (Fig. 2B) and with the metrics for the mean spectra being calculated from replicate samples wells.

2. *Performance of the CNNs*: The error between the observed (*obs*) and estimated (*est*) pigment concentrations obtained using the 1D CNN were calculated using typical error metrics (Morley et al. 2018), that is, median absolute percentage difference (MAPD, %), mean absolute percentage error (MAPE, %) and root MSE (RMSE). Also, median symmetric accuracy (MdSA) was calculated to provide a robust error estimate for biased data (Morley et al. 2018). The metrics were defined as follows, where *N* is the number of samples:

$$MAPD, \% = median \left(\frac{|est_i - obs_i|}{obs_i} \right) \times 100, \ i = 1, \ldots, N \qquad (4)$$

$$MAPE, \% = mean \left(\frac{|est_i - obs_i|}{obs_i} \right) \times 100, \ i = 1, \ldots, N \qquad (5)$$

$$RMSE, \mu g\,L^{-1} = \sqrt{\frac{1}{n} \left(\sum_i^n (est - obs)^2 \right)}, i = 1, \ldots, N \qquad (6)$$

$$MdSA, \% = \left(10^{median \left(\left| \log_{10} \left(\frac{est}{obs} \right) \right| \right)} - 1 \right) \times 100 \qquad (7)$$

Assessment

Measurement error

The mean coefficient of variation in the absorbance spectra was 20.9% (SD = 21.2) and the median was 15.1% (Table 4). Mean coefficients of variation in the HPLC assessments of Chl

Table 1. Model architecture selected to estimate chlorophyll *a* concentration.

Layer	Kernel or pool size	Activation	Output shape	Parameters	Learning rate (Adam)
Conv1D	3	Relu	(None, 148, 384)	1536	0.001
MaxPooling1D	2		(None, 74, 384)	0	
Conv1D	5	Relu	(None, 70, 32)	61,472	
MaxPooling1D	2		(None, 35, 32)	0	
Conv1D	3	Relu	(None, 33, 32)	3104	
MaxPooling1D	2		(None, 16, 32)	0	
Conv1D	5	Relu	(None, 12, 384)	61,824	
MaxPooling1D	2		(None, 6, 384)	0	
Flatten			(None, 2304)	0	
Dense		Relu	(None, 480)	1,106,400	
Dense		Relu	(None, 320)	153,920	
Dense		Relu	(None, 320)	102,720	
Dropout (0.2)			(None, 320)	0	
Dense		Linear	(None, 1)	321	

Table 2. Model architecture selected to estimate the carotenoid concentration.

Layer	Kernel or pool size	Activation	Output shape	Parameters	Learning rate (Adam)
Conv1D	3	Relu	(None, 148, 96)	384	
MaxPooling1D	2		(None, 74, 96)	0	0.001
Conv1D	3	Relu	(None, 72, 384)	110,976	
MaxPooling1D	2		(None, 36, 384)	0	
Conv1D	5	Relu	(None, 32, 32)	61,472	
MaxPooling1D	2		(None, 16, 32)	0	
Conv1D	3	Relu	(None, 14, 32)	3104	
MaxPooling1D	2		(None, 7, 32)	0	
Flatten			(None, 224)	0	
Dense		Relu	(None, 384)	86,400	
Dense		Relu	(None, 352)	135,520	
Dense		Relu	(None, 480)	169,440	
Dropout (0.2)			(None, 480)	0	
Dense		Linear	(None, 1)	481	

a and carotenoids were 6.7% (SD = 6.0) and 5.7% (SD = 4.0), respectively. For phycocyanin assessments, the mean coefficient of variation was 16.6% (SD = 17.5, median 9.3). The errors estimated as the distance between the means of two replicates of mean absorbance spectra were of the same order of magnitude as the corresponding coefficient of variation (Table 5). For phycocyanin assessments, distances from the mean were slightly lower than, or the same order of magnitude as, the error calculated as the coefficient of variation (Table 5). These levels of measurement error determine the accuracy that the 1D CNN could be expected to provide in estimating pigment concentrations.

Observed pigment concentrations

The observed total pigment concentrations based on the reference HPLC and spectrophotometry assessments varied across two orders of magnitude; 3.2–335.7 μg L^{-1} in the non-condensed lake water samples (Fig. 3). Median concentration of Chl *a* in the dataset was 6.6 μg L^{-1}, carotenoids 4.3 μg L^{-1}, and phycocyanin 4.6 μg L^{-1} indicating that most of the sites sampled were relatively oligotrophic. The respective mean concentrations were 17.9, 8.2^{-}, and 22.5 μg L^{-1}. Total pigment concentrations were highest in Airthrey Loch, Loch Leven, and Monikin Reservoir (Fig. 3). The observed Chl *a* and phycocyanin concentrations were 58.7–58.9 and 162.3–244.8 μg L^{-1} in Airthrey Loch, respectively, 10.4–58.9 and 4.6–45.9 μg L^{-1} in Loch Leven and 21.5 and 32.4 μg L^{-1} in Monikin Reservoir. Pigment concentrations decreased in Loch Leven in mid-June (Fig. 3), but otherwise Chl *a* and phycocyanin concentrations stayed relatively high in these lakes. The blooming taxa were identified as coiled Nostocales species based on qualitative scrutiny using an inverted light microscope (see description

Table 3. Model architecture selected to estimate the phycocyanin concentration.

Layer	Kernel or pool size	Activation	Output shape	Parameters	Learning rate (Adam)
Conv1D	5	Relu	(None, 146, 160)	960	
MaxPooling1D	2		(None, 73, 160)	0	0.001
Conv1D	3	Relu	(None, 71, 384)	184,704	
MaxPooling1D	2		(None, 35, 384)	0	
Conv1D	5	Relu	(None, 31, 32)	61,472	
MaxPooling1D	2		(None, 15, 32)	0	
Conv1D	3	Relu	(None, 13, 352)	34,144	
MaxPooling1D	2		(None, 6, 352)	0	
Conv1D	5		(None, 2, 192)	338,112	
Flatten			(None, 384)	0	
Dense		Relu	(None, 192)	73,920	
Dense		Relu	(None, 448)	86,464	
Dense		Relu	(None, 320)	143,680	
Dense		Relu	(None, 352)	112,992	
Dense		Relu	(None, 384)	135,552	
Dense		Relu	(None, 288)	110,880	
Dropout (0.2)			(None, 288)	0	
Dense		Linear	(None, 1)	289	

Table 4. Coefficient of variation (CV, %) in different assessments calculated from three replicate samplings. Imaging$_{absorbance(\lambda)}$ denotes absorbance spectra calculated over the 50 × 50 pixel ROI, Carotenoids denotes total concentration of carotenoids, Spectrophotometry$_{PC}$ denotes spectrophotometric assessments of phycocyanin.

CV (%)	Imaging$_{absorbance(\lambda)}$	HPLC$_{Chl}$	HPLC$_{Carotenoids}$	Spectrophotometry$_{PC}$
Mean	21	6.7	5.7	17
SD	21	6.0	4.0	18
Median	15	5.0	5.0	9
Samplings	14	31	31	3
Replicates	3	3	3	3

of the microscopy-based observations in Supplementary Table 1). The algal blooms in Loch Leven exceeded the first Chl *a* alert level of 10 μg L^{-1} set by World Health Organization (WHO 2003) and once also the moderate health alert of 50 μg L^{-1} (WHO 2003) in late May.

Estimated pigment concentrations

When Loch Leven was excluded in the L-O-O tests, the CNNs trained with the data from the 19 other lakes yielded adequate estimates of Chl *a* and carotenoid concentrations for Loch Leven (Fig. 4). Median absolute percentage difference was 15% for Chl *a* (MAPE 22%, SD 23%) and 36% for carotenoids (MAPE 35%, SD 29%). However, the model failed to predict the mid- and late-June carotenoid concentrations in Loch Leven resulting to 64%–80% error (Fig. 4). Phycocyanin estimates were typically higher than the observed phycocyanin levels in Loch Leven (Fig. 4), with the median absolute percentage difference being 102% (MAPE 174%, SD 199%)

Table 5. Distance from the mean (%) in different assessments calculated from two replicate samplings. Imaging$_{absorbance(\lambda)}$ means absorbance spectra calculated over the 50 × 50 pixel ROI, Carotenoids denotes total concentration of the major carotenoids, Spectrophotometry$_{PC}$ denotes spectrophotometric assessments of phycocyanin.

Distance from the mean (%)	Imaging$_{absorbance(\lambda)}$	Spectrophotometry$_{PC}$
Mean	18	6
SD	17	2
Median	14	7
Samplings	7	4
Replicates	2	2

for phycocyanin estimates at this site. The estimates of Chl *a* exceeded the first alert level of WHO at each sampling time, and the second, level of moderate health alert in late May and
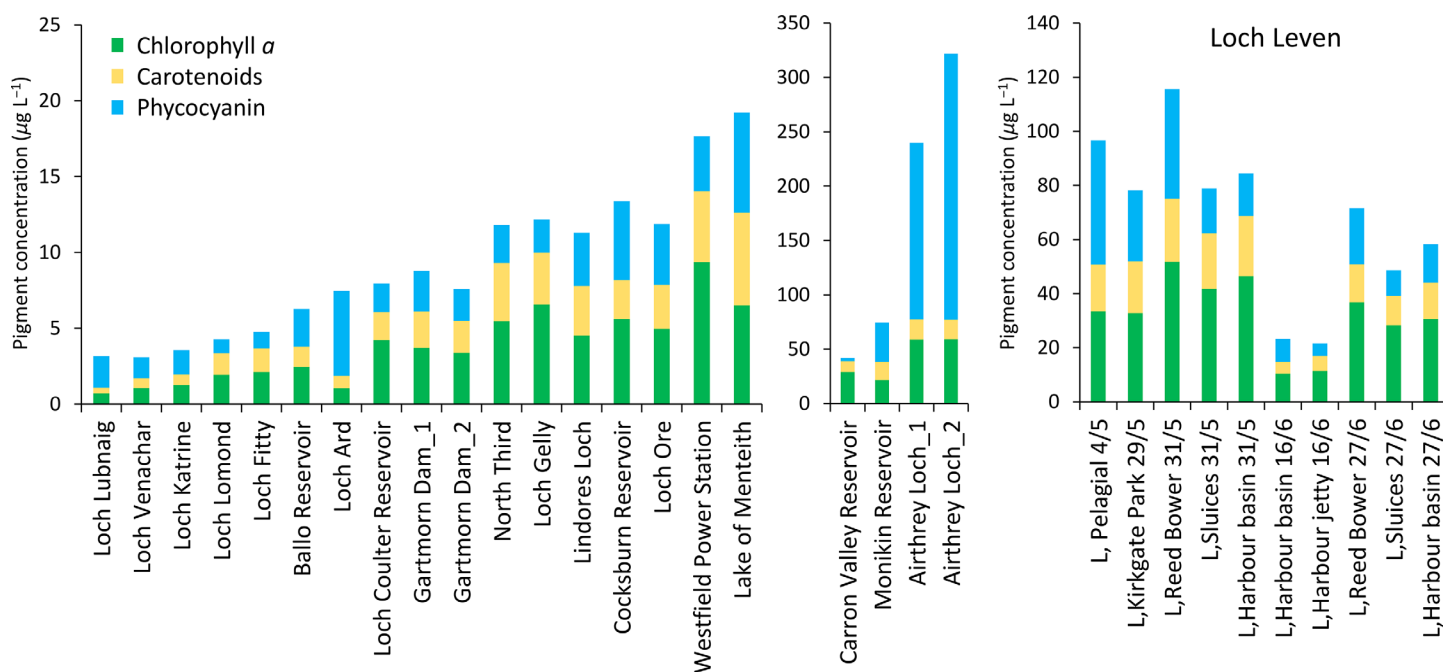
Fig. 3. Observed pigment composition in the 20 sampled lakes. Samples with high concentrations and samples from Loch Leven are shown in separate panels for clarity. Chl *a* and carotenoid concentrations are averages of three replicate filtrations (six replicates from Westfield Power Station Reservoir), and phycocyanin 1–3 filtrations.

late June (Fig. 5). At the times when Chl *a* moderate health alert levels were exceeded, also phycocyanin estimates were high (Fig. 4). The RMSE for Chl *a* estimates at Loch Leven was 8.4 μg L^{-1}, for carotenoids 4.5 μg L^{-1}, and for phycocyanin 31.7 μg L^{-1}. MdSA was 15%, 39%, and 107%, for Chl *a*, carotenoids and phycocyanin, respectively, in Loch Leven.

When the 19 lakes other than Loch Leven were scrutinized, the congruence between the observations and the estimations given by the 1D CNN was adequate for Chl *a* and total carotenoids (Fig. 5), median absolute percentage difference being 26% (MAPE 30% SD 23%) and 27% (MAPE 47%, SD 93%), respectively. The highest prediction errors were for samples from Westfield Power Station Reservoir (Fig. 5), which, according to qualitative microscopic inspection, was blooming with the green alga *Monoraphidium* sp. This deviating observation is corroborated also by the notably high proportion of Chl *b* (Supplementary Fig. 2) and the separation of the water body in the primary component analysis of the pigment profiles (Supplementary Fig. 2). Median absolute percentage difference for the phycocyanin concentrations was high (75%, MAPE 193%, SD 274%). RMSE was 4.6 μg L^{-1} for Chl *a* estimates, 5.6 μg L^{-1} for carotenoid estimates and 50.0 for μg L^{-1} for phycocyanin estimates (Fig. 5). MdSA was 26%, 35%, and 215%, for Chl *a*, carotenoids and phycocyanin, respectively, in the lakes, excluding Loch Leven. The highest concentrations in the plots (Fig. 5) were extrapolations by the trained model, because samples containing spectra from the test lake were removed from the model training data (X_{sim}) for the L-O-O tests.

Discussion

A proposed protocol to resolve Chl *a*, carotenoid and phycocyanin concentrations, based on a mobile spectral imager and 1D CNNs, was developed and evaluated against HPLC and spectrophotometry-based assessments. The imaging setup was simple to operate, and the samples required minimal processing in laboratory. However, there are some important sources of variation, especially in the sample processing and in the volume and distribution of training data, that need to be considered when applying this protocol.

Measurement error and model performance

One of the most prominent sources of variation in the imaging protocol is likely to have been the cell distribution resulting from centrifugation. The centrifugation force was selected so that it did not destroy phytoplankton cells and pigments could be expected to stay inside intact cells. However, the centrifugation force used here was probably not strong enough to settle very buoyant or small cells, such as picophytoplankton. In this study, *Volvox* colonies in samples from Lake of Menteith recovered quickly from the centrifugation and started to swim up into the centrifuge tube. Therefore, it is very important that samples are pipetted quickly to minimize the loss of flagellated cells.

The model performance evaluated with the median percentage metrics was on a similar level to the coefficients of variation in the assessments. This indicated that the expected performance had been achieved. In this study, despite sampling the
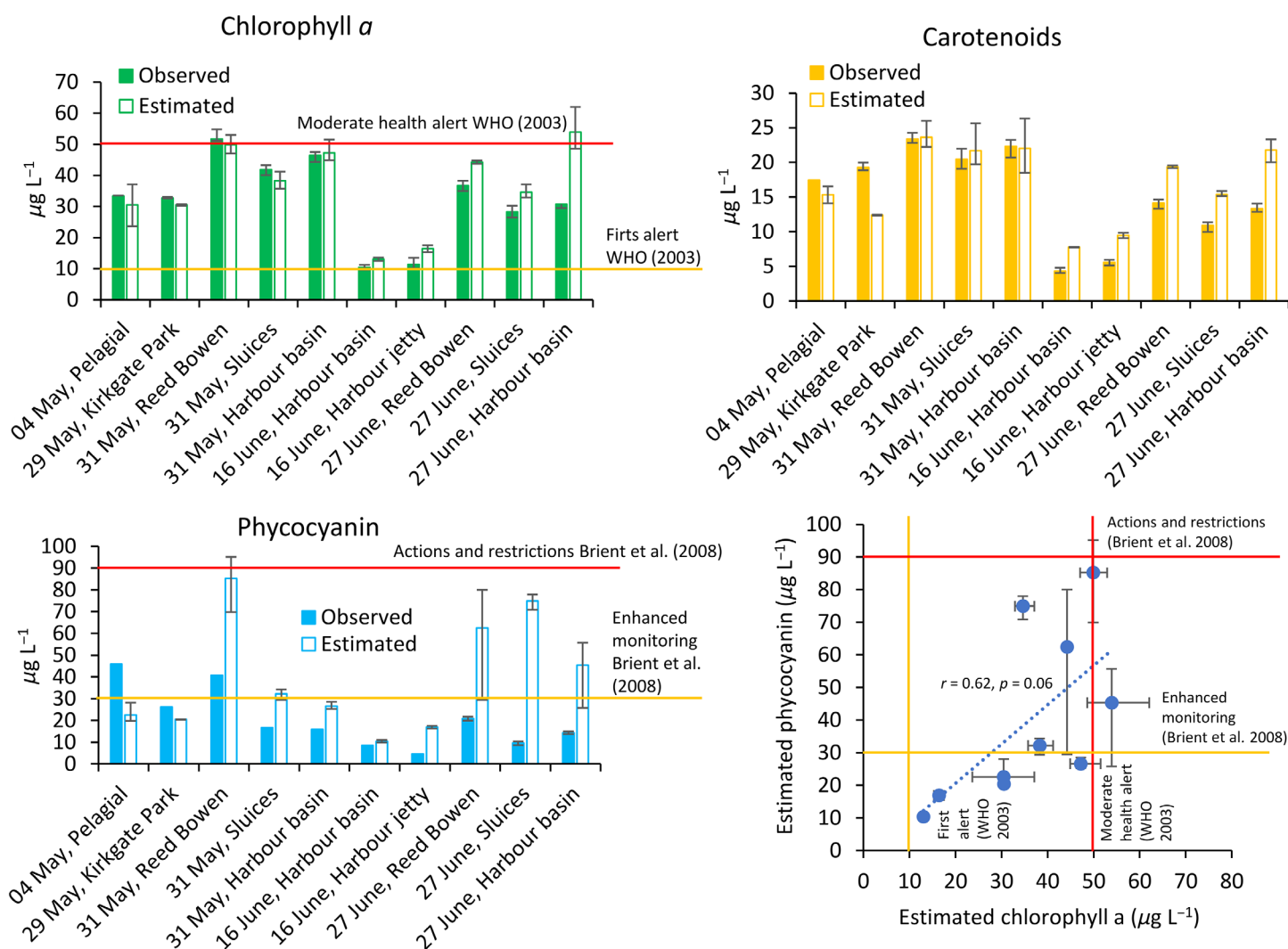
Fig. 4. Monitoring of pigments in Loch Leven in May–June 2022. "Observed" refers to HPLC assessments of chlorophyll *a* and carotenoids, or spectro-photometric assessment of phycocyanin. "Estimated" refers to the concentrations predicted by the 1D CNN. The 1D CNNs were trained with mixtures (X_{sim}) of samples from 19 lakes, excluding Loch Leven. Error bars show the minimum and maximum of the observed or estimated values. Chl *a* alert concentrations by WHO (2003) and corresponding phycocyanin concentrations according to Brient et al. (2008) are marked on the panels with yellow (lower alert levels) and red (higher alert levels) lines.

lakes as a gradient from western to eastern Scotland, most of the lakes were oligo-mesotrophic and rather similar in their pigment profiles apart from the phycocyanin-rich Airthrey Loch, Loch Leven and Monikin Reservoir, and the Chl *a*-rich Westfield Power Station Reservoir (*see* Supplementary Fig. 2 for a PCA). Werther et al. (2022) noted that that algorithms to estimate Chl *a* from satellite images have been developed, typically, for eutrophic lakes. In their study, these authors used a Bayesian neural network to estimate Chl *a* from multispectral satellite data, including oligotrophic and mesotrophic lakes, and reported MdSA values of 17.28–32.42%. The MdSAs for Chl *a* and carotenoids in this study were on the same level (26–35%) in the L-O-O tests.

Recently, Aptoula and Ariman (2021) demonstrated the use of a CNN to resolve Chl *a* concentration from

multispectral satellite images of Lake Balik, Turkey. They reported RMSEs of 4.07–13.64 $\mu g\ L^{-1}$ for estimations across different temporal and spatial scales. Pyo et al. (2019) constructed a CNN to resolve Chl *a* and phycocyanin concentrations from hyperspectral satellite images of Baekje weir in Geum River, Korea. Their validation results showed good consistency between the observed concentrations and CNN-based estimates (R^2: 0.73–0.86). The RMSE was 8.34 $\mu g\ L^{-1}$ for Chl *a* estimates and 9.39 $\mu g\ L^{-1}$ for phycocyanin estimates (Pyo et al. 2019). In this study, RMSE values of Chl *a* and carotenoids from the L-O-O tests (4.6 and 5.6 $\mu g\ L^{-1}$, respectively) were lower than, or at a similar level to, the results presented by Aptoula and Ariman (2021) and Pyo et al. (2019). In this study, condensation of the oligotrophic lake samples was no doubt instrumental to detecting the
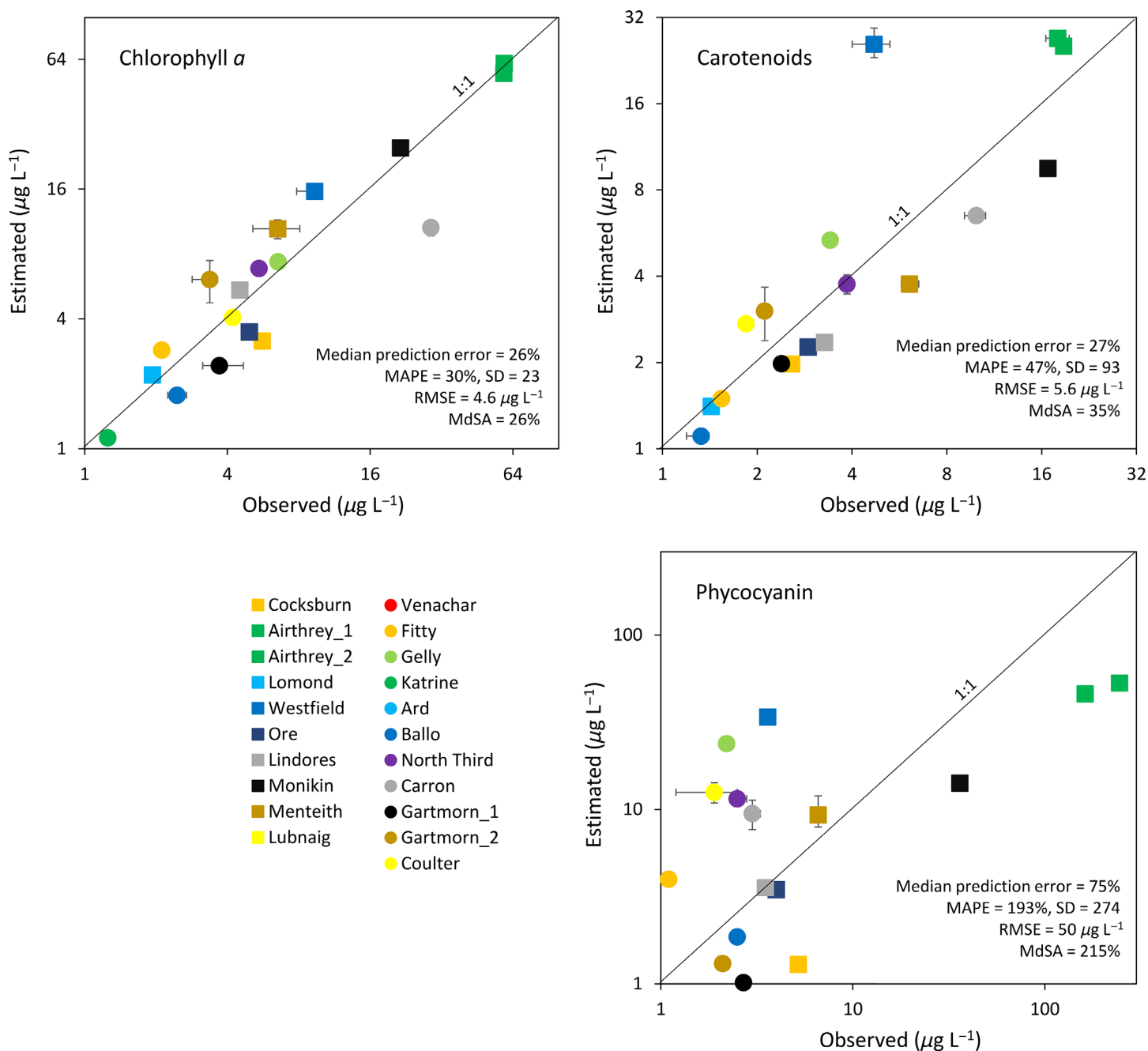
Fig. 5. Ratios of observed concentrations and estimations made by the leave-one-out CNNs. "Observed" refers to the HPLC assessments of chlorophyll *a* and carotenoids, or the spectrophotometric assessment of phycocyanin. "Estimated" refers to the concentrations predicted by the 1D CNN. Values are means of 1–3 replicates (6 HPLC replicates for Westfield Power Plant). Error bars show the minimum and maximum of the observed or estimated values. Values below 1 μg L^{-1} are omitted for the logarithmic scales. "Loch" prefixes or "Reservoir" suffixes are omitted from the legend to improve readability. The models were trained with X_{sim} where data from one lake at a time (the test lake for the estimations) had been removed.

pigment-related signal from the spectral images. For Loch Leven, RMSEs were at the similar level or lower for Chl *a* and carotenoids (8.4 and 4.5 μg L^{-1}) than the results by Aptoula and Ariman (2021) and Pyo et al. (2019). Error in phycocyanin estimates from this study were notably higher than those reported by Pyo et al. (2019).

Variation in phycocyanin estimates

An explicit level of risk for phycocyanin concentration in lakes cannot be estimated because several factors add variation to the detection of phycocyanin. Cyanobacterial species and proportion, as well as variability in their intracellular phycocyanin concentration, affect the observation of the pigment

(Stumpf et al. 2016). The WHO has set a Chl a concentration of 10 μg L^{-1} as the first alert level for irritative and allergenic reactions due to possible cyanobacteria blooms and 50 μg L^{-1} as the second, moderate health alert level (WHO 2003). Also, based on the sampling of 35 water bodies in western France, Brient et al. (2008) estimated that these levels of alert could correspond to approximately 30 μg L^{-1} ($\pm$ 2 μg L^{-1}) and 90 μg L^{-1} ($\pm$ 2 μg L^{-1}) of phycocyanin. In Loch Leven, the phycocyanin estimates increased with Chl a estimates (Pearson $r = 0.62$, $p = 0.06$, Spearman's $\rho = 0.77$, $p = 0.009$). Each time, when Chl a estimate exceeded the level of moderate health warning, phycocyanin estimates exceeded at least the first level of alert by Brient et al. (2008, Fig. 4). However, the results from Loch Leven indicated that the first phycocyanin alert level of 30 μg L^{-1} might be too high for Loch Leven, and the alerts should be given earlier. Therefore, although the protocol demonstrated here failed to estimate phycocyanin concentration adequately, it could be a promising application for early detection of cyanobacteria blooms.

The reasons for the high variation in phycocyanin estimate in this study are likely to have been the variation in the spectrophotometry-based phycocyanin assessments and, more, prominently the imbalanced data distribution. Phycocyanin concentrations were generally low in the data from the 20 lakes, with few containing phycocyanin concentrations above the alert level. Therefore, there were unlikely to have been enough data for the L-O-O tests to make adequate predictions of concentrations. In the reference assessments by spectrophotometry, the variation could have been caused by incomplete extraction of the pigment from cells as cyanobacteria are known to be resistant to the mechanical degradation on which the extraction protocol of the water-soluble pigments is based (Horváth et al. 2013). Carotenoid aphanizophyll was observed in the phycocyanin-rich Airthrey Loch, Loch Leven, and Monikin Reservoir (Supplementary Fig. 2). In addition to phycocyanin, aphanizophyll is a biomarker of cyanobacteria (Peltomaa et al. 2023). In the dataset of all the lakes and samplings, phycocyanin concentration correlated with aphanizophyll concentration ($r = 0.6$, $p < 0.001$) and with Chl a concentration ($r = 0.7$, $p < 0.001$), suggesting that in addition to phycocyanin analytics other sources contributed to the variation. Tests with a Loch Leven sample and Grad CAM algorithm (https://www.kaggle.com/discussions/general/286454, accessed 06 October 2023) showed that the phycocyanin model was focusing on the region of expected phycocyanin absorbance (Supplementary Fig. 4). Based on these observations, both the relatively limited training data and the variation in the phycocyanin analytics probably caused the high error in phycocyanin estimates.

Outlook

The possibility to resolve carotenoid concentrations, in addition to the traditionally monitored Chl a concentrations, in a rapid and robust way could open up new possibilities for monitoring. On-site spectrofluorometers are applicable to monitor Chl a and phycocyanin based on the pigments' autofluorescence. In contrast, carotenoids are a diverse group of auxiliary pigments that do not have autofluorescence and whose absorbance spectra overlap with each other (Clementson and Wojtasiewicz 2019) and with humic substances, making their on-site or remote sensing assessment less straightforward. Total carotenoids were estimated successfully in this study, but expedient training data distribution and model testing are crucial when applying this approach in the future. In addition, quantitative microscopy-based assessments in parallel with spectral imaging could be a potential way to develop the protocol towards spectral imaging-based taxonomic assessment.

Recently, Legleiter et al. (2022) combined spectral imaging from the laboratory with satellite data to resolve bloom-forming cyanobacteria. They obtained reflectance spectra of different cyanobacteria species in the laboratory with a spectral imager attached to a microscope and used those as spectral endmembers to resolve to taxa level using data from satellite images. Instead of a neural network, they used a linear model-based spectral unmixing. Their results were promising because their algorithm indicated the presence of possible toxin-producing taxa. Their results also demonstrated that laboratory-based spectral imaging is scalable and has the potential to support phytoplankton monitoring from satellites (Legleiter et al. 2022). The models trained in this study most likely will not generalize as such. By using the full survey protocol, similar results can be obtained at other times of the year and elsewhere. Openly available software to analyze phytoplankton spectra could be an interesting future direction. The protocol described here, could also be tested for rapid assessment of cyanobacteria blooms in inland waters with high water color, because one of the benefits of the preprocessing was to enable the capability of introducing water color as a background to the 1D CNN. Another benefit is that spectral imaging in the laboratory enables further processing of the samples, downstream, such as biochemical or microscopy-based analysis. This is because the sample processing and imaging were non-destructive.

Comments and recommendations

A new protocol based on a mobile hyperspectral imager was introduced to support phytoplankton monitoring. In addition to sampling, this protocol contains three steps: (1) sample condensing, (2) spectral imaging, and (3) modeling with CNNs. Based on assessments using samples from 20 Scottish lakes, the protocol was demonstrated to have the potential to resolve Chl a and carotenoids from lake water samples and indicate cyanobacteria blooms. Further studies are need to improve the predictability of phycocyanin assessments and to test the protocol more widely across different lakes and setups.

Data availability statement

Pigment data and corresponding descriptive metadata are available via The Environmental Information Data Centre

(EIDC) with doi: https://doi.org/10.5285/7794cbfa-1d74-4885-b303-8c20f4608728. Spectral images and Python code to image processing are available via Fairdata IDA with doi: https://doi.org/10.23729/20a62a08-e776-47f4-96e5-e8e3475d8c4a. CNNs to resolve algae pigments are available via Fairdata IDA with doi: https://doi.org/10.23729/0d4b355b-8ed9-466a-a6db-86843e62262a.

References

Aptoula, E., and S. Ariman. 2021. Chlorophyll-a retrieval from sentinel-2 images using convolutional neural network regression. IEEE Geosci. Remote Sens. Lett. **19**: 1–5. doi:10.1109/LGRS.2021.3070437

Bennett, A., and L. Bogorad. 1973. Complementary chromatic adaptation in a filamentous blue-green alga. J. Cell Biol. **58**: 419–435. doi:10.1083/jcb.58.2.419

Bricaud, A., C. Mejia, D. Blondeau-Patissier, H. Claustre, M. Crepon, and S. Thiria. 2007. Retrieval of pigment concentrations and size structure of algal populations from their absorption spectra using multilayered perceptrons. Appl. Optics **46**: 1251–1260. doi:10.1364/AO.46.001251

Brient, L., and others. 2008. A phycocyanin probe as a tool for monitoring cyanobacteria in freshwater bodies. J. Environ. Monit. **10**: 248–255. doi:10.1039/B714238B

Brito, A. C., C. Sá, V. Brotas, R. J. Brewin, T. Silva, J. Vitorino, T. Platt, and S. Sathyendranath. 2015. Effect of phytoplankton size classes on bio-optical properties of phytoplankton in the Western Iberian coast: Application of models. Remote Sens. Environ. **156**: 537–550. doi:10.1016/j.rse.2014.10.020

Chabrillat, S., L. Guanter, K. Segl, and others. 2020. The Enmap German spaceborne imaging spectroscopy mission: Update and highlights of recent preparatory activities. IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium. doi:10.1109/IGARSS39084.2020.9324006

Clementson, L. A., and B. Wojtasiewicz. 2019. Dataset on the absorption characteristics of extracted phytoplankton pigments. Data Brief **24**: 103875. doi:10.1016/j.dib.2019.103875

EN 15204. 2006. European standard. In water quality—Guidance standard for the routine analysis of phytoplankton abundance and composition using inverted microscopy (Utermöhl technique). European Committee for Standardization.

Giardino, C., V. E. Brando, P. Gege, and others. 2019. Imaging spectrometry of inland and coastal waters: State of the art, achievements and perspectives. Surv. Geophys. **40**: 401–429. doi:10.1007/s10712-018-9476-0

Goodfellow, I., Y. Bengio, and A. Courville. 2016. Deep learning. MIT press.

Henrichs, D. W., S. Anglès, C. C. Gaonkar, and L. Campbell. 2021. Application of a convolutional neural network to improve automated early warning of harmful algal blooms. Environ. Sci. Pollut. Res. **28**: 28544–28555. doi:10.1007/s11356-021-12471-2

Hoepffner, N., and S. Sathyendranath. 1991. Effect of pigment composition on absorption properties of phytoplankton. Mar. Ecol. Prog. Ser. **73**: l–23. doi:10.3354/meps073011

Horváth, H., A. W. Kovács, C. Riddick, and M. Présing. 2013. Extraction methods for phycocyanin determination in freshwater filamentous cyanobacteria and their application in a shallow lake. Eur. J. Phycol. **48**: 278–286. doi:10.1080/09670262.2013.821525

Irigoien, X., B. Meyer, R. Harris, and D. Harbour. 2004. Using HPLC pigment analysis to investigate phytoplankton taxonomy: The importance of knowing your species. Helgoland Mar. Res. **58**: 77–82. doi:10.1007/s10152-004-0171-9

ISO 10260 1992 Water quality—Measurement of biochemical parameters—Spectrometric determination of the chlorophyll-a concentration. International Organization for Standardization.

Kirk, J. T. 2011. Light and photosynthesis in aquatic ecosystems. Cambridge Univ. Press.

Kraft, K., and others. 2022. Towards operational phytoplankton recognition with automated high-throughput imaging, near-real-time data processing, and convolutional neural networks. Front. Mar. Sci. **9**: 867695. doi:10.3389/fmars.2022.867695

Legleiter, C. J., and others. 2022. Spectral mixture analysis for surveillance of harmful algal blooms (SMASH): A field-, laboratory-, and satellite-based approach to identifying cyanobacteria genera from remotely sensed data. Remote Sens. Environ. **279**: 113089. doi:10.1016/j.rse.2022.113089

Millie, D. F., M. C. Baker, C. S. Tucker, B. T. Vinyard, and C. P. Dionigi. 1992. High-resolution airborne remote sensing of bloom-forming phytoplankton. J. Phycol. **28**: 281–290. doi:10.1111/j.0022-3646.1992.00281.x

Möller, K., W. Petersen, and R. Nair. 2019. Report on best practice in the utilization of sensors used for measuring nutrients, biology related optical properties, variables of the marine carbonate system, and for coastal profiling. JERICO-NEXT Deliverable D2.5. Version 1.0. doi:10.25607/OBP-898

Morley, S. K., T. V. Brito, and D. T. Welling. 2018. Measures of model performance based on the log accuracy ratio. Space Weather **16**: 69–88. doi:10.1002/2017SW001669

Mouw, C. B., and others. 2015. Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. Remote Sens. Environ. **160**: 15–30. doi:10.1016/j.rse.2015.02.001

Nair, A., S. Sathyendranath, T. Platt, J. Morales, V. Stuart, M. H. Forget, E. Devred, and H. Bouman. 2008. Remote sensing of phytoplankton functional types. Remote Sens. Environ. **112**: 3366–3375. doi:10.1016/j.rse.2008.01.021

Peltomaa, E., H. Asikainen, J. Blomster, H. Pakkanen, C. Rigaud, P. Salmi, and S. Taipale. 2023. Phytoplankton

group identification with chemotaxonomic biomarkers: In combination they do better. Phytochemistry **209**: 113624. doi:10.1016/j.phytochem.2023.113624

Pyo, J., H. Duan, S. Baek, M. S. Kim, T. Jeon, Y. S. Kwon, H. Lee, and K. H. Cho. 2019. A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery. Remote Sens. Environ. **233**: 111350. doi:10.1016/j.rse.2019.111350

Reynolds, C. S. 2006. The ecology of phytoplankton. Cambridge Univ. Press.

Rode, M., and others. 2016. Sensors in the stream: The high-frequency wave of the present. Environ. Sci. Technol. **50**: 10297–10307. doi:10.1021/acs.est.6b02155

Salmi, P., M. Calderini, S. Pääkkönen, S. Taipale, and I. Pölönen. 2022. Assessment of microalgae species, biomass, and distribution from spectral images using a convolution neural network. J. Appl. Phycol. **34**: 1565–1575. doi:10.1007/s10811-022-02735-w

Salonen, K., P. Salmi, and J. Keskitalo. 2021. How to reach optimal estimates of confidence intervals in microscopic counting of phytoplankton? J. Plankton Res. **43**: 846–852. doi:10.1093/plankt/fbab062

Sathyendranath, S., V. Stuart, T. Platt, H. Bouman, O. Ulloa, and H. Maass. 2005. Remote sensing of ocean colour: Towards algorithms for retrieval of pigment composition. Indian J. Mar. Sci. **34**: 333–340. http://nopr.niscpr.res.in/handle/123456789/1568

Sobiechowska-Sasim, M., J. Stoń-Egiert, and A. Kosakowska. 2014. Quantitative analysis of extracted phycobilin pigments in cyanobacteria—An assessment of spectrophotometric and spectrofluorometric methods. J. Appl. Phycol. **26**: 2065–2074. doi:10.1007/s10811-014-0244-3

Solovchenko, A. 2023. Seeing good and bad: Optical sensing of microalgal culture condition. Algal Res. **71**: 103071. doi:10.1016/j.algal.2023.103071

Spears, B. M., D. P. Hamilton, Y. Pan, C. Zhaosheng, and L. May. 2022. Lake management: Is prevention better than cure? Inland Waters **12**: 173–186. doi:10.1080/20442041.2021.1895646

Stadnichuk, I. N., P. M. Krasilnikov, and D. V. Zlenko. 2015. Cyanobacterial phycobilisomes and phycobiliproteins. Microbiology **84**: 101–111. doi:10.1134/S0026261715020150

Stumpf, R. P., T. W. Davis, T. T. Wynne, J. L. Graham, K. A. Loftin, T. H. Johengen, D. Gossiaux, D. Palladino, and A.

Burtner. 2016. Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria. Harmful Algae **54**: 160–173. doi:10.1016/j.hal.2016.01.005

Van Heukelem, L., and C. S. Thomas. 2005. The HPL Method, p 86-92. In The second SeaWiFS HPLC analysis round-Robin experiment (SeaHARRE-2). NASA Tech. Memo. 2005: 212785. NASA Goddard Space Flight Center.

Vuorio, K., L. Lepistö, and A. L. Holopainen. 2007. Intercalibrations of freshwater phytoplankton analyses. Boreal Environ. Res. **12**: 561–569. doi:ber12-561

Watanabe, M., and M. Ikeuchi. 2013. Phycobilisome: Architecture of a light-harvesting supercomplex. Photosynth. Res. **116**: 265–276. doi:10.1007/s11120-013-9905-3

Werther, M., and others. 2022. A Bayesian approach for remote sensing of chlorophyll-a and associated retrieval uncertainty in oligotrophic and mesotrophic lakes. Remote Sens. Environ. **283**: 113295. doi:10.1016/j.rse.2022.113295

WHO. 2003. Guidelines for safe recreational water environments. Volume 1: Coastal and fresh waters. World Health Organization. doi:10.1002/tox.10114

Wright, S. W., and S. W. Jeffrey. 2006. Pigment markers for phytoplankton production, p. 71–104. *In* Marine organic matter: Biomarkers, isotopes and DNA. The handbook of environmental chemistry. Springer. doi:10.1007/698_2_003