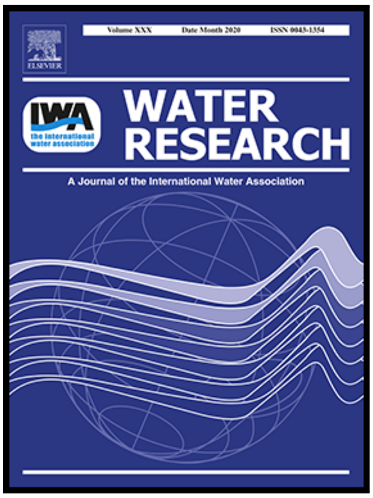
Journal Pre-proof



Time-series modelling of harmful cyanobacteria blooms by convolutional neural networks and wavelet generated time-frequency images of environmental driving variables

Hyo Gyeom Kim , Kyung Hwa Cho , Friedrich Recknagel

PII: S0043-1354(23)01102-8
DOI: https://doi.org/10.1016/j.watres.2023.120662
Reference: WR 120662

To appear in: *Water Research*

Received date: 5 July 2023
Revised date: 4 September 2023
Accepted date: 21 September 2023

Please cite this article as: Hyo Gyeom Kim , Kyung Hwa Cho , Friedrich Recknagel , Time-series modelling of harmful cyanobacteria blooms by convolutional neural networks and wavelet generated time-frequency images of environmental driving variables, *Water Research* (2023), doi: https://doi.org/10.1016/j.watres.2023.120662

Highlights

- Time windows for cyanobacteria blooms in rivers were identified.
- Time-frequency images of environmental drivers were utilized as predictors for blooms.
- Image-driven CNN models identified bloom intensities qualitatively.
- Image-driven CNN models predicted *Microcystis* densities quantitatively.
- CNN-models prove to be feasible for one-month-ahead forecasts of cyanobacterial blooms.

Time-series modelling of harmful cyanobacteria blooms by convolutional neural networks and wavelet generated time-frequency images of environmental driving variables

Hyo Gyeom Kim [a,b*], Kyung Hwa Cho [b], Friedrich Recknagel [a*]

[a] School of Biological Sciences, The University of Adelaide, Adelaide, 5005 Australia

[b] School of Civil, Environmental, and Architectural Engineering, Korea University, Seoul, 02841, The Republic of Korea

*Corresponding authors:

E-mail addresses:

1004kyeom@gmail.com (H.G. Kim); friedrich.recknagel@adelaide.edu.au (F. Recknagel)

Abstract

Early warning systems for harmful cyanobacterial blooms (HCBs) that enable precautional control measures within water bodies and in water works are largely based on inferential time-series modelling. Among deep learning techniques, convolutional neural networks (CNNs) are widely applied for recognition of pictorial, acoustic and thermal images. Time-frequency images of environmental drivers generated by wavelets may provide crucial signals for modelling of HCBs to be recognized by CNNs. This study applies CNNs for time-series modelling of HCBs of *Microcystis* sp. in four South Korean rivers between 2016 and 2022 by means of time-frequency images of environmental drivers within the lead time of HCBs. After estimating the cardinal dates of beginning, peak, and ending of HCBs, wavelet analysis identified key drivers by phase analysis and generated time-frequency images of the drivers within the cardinal dates for 3, 4 and 5 years. Performances of CNNs were compared in terms of four determinants of input images: methods of estimating critical timings, the number of segments, time-series continuity, and image size. The resulting CNNs predicted high or low intensities of HCBs with a mean accuracy of 97.79 ± 0.06% and F1-score 97.49 ± 0.06% for training dataset, and a mean accuracy of 95.01 ± 0.06% and F1-score 93.30 ± 0.07% for testing dataset. Predictions of *Microcystis* abundances by CNNs achieved a mean MSE of 2.58 ± 2.46 and a mean R^2 of 0.78 ± 0.20 for training, and a mean MSE of 2.76 ± 2.42 and a mean R^2 of 0.55 ± 0.20 for testing dataset. Precipitation and discharge appeared to be the best performing drivers for qualitative and quantitative predictions of HCBs pointing at the nonstationary nature of river habitats. This study highlights the opportunities of time-series modelling by CNNs driven by wavelet generated time-frequency images of key environmental variables for forecasting of HCBs.

Keywords: *Microcystis*; *rivers*; *wavelet analysis*; *time frequency images*; *convolutional neural networks*; *blooming time window*

1. Introduction

The synergy of anthropogenically induced eutrophication and global warming enables massive growth of toxic cyanobacterial species that pose a major threat to drinking and irrigation water supplies, fishing, and recreational use of inland waters worldwide (Paerl and Huisman, 2009; Glibert, 2020). 'Despite advances in scientific understanding of cyanobacteria and associated compounds, many unanswered questions remain about occurrence, environmental triggers for toxicity, and the ability to predict the timing, duration, and toxicity of harmful cyanobacteria blooms (HCB). Scientific data and mechanistic understanding of environmental factors — as well as the related adverse effects of cyanotoxin exposure — are necessary to develop reliable early warning systems and predictive tools that guide management decisions. Advanced warning at time scales relevant to HCB management (hours to days), allow proactive, rather than reactive, responses to these events' (Graham et al., 2016).

Great efforts are undertaken to develop early warning tools suitable for near- and short-term forecasting. Most promising results are being achieved by applying cyanotoxin-encoding genes for the prediction of cyanotoxin production in lakes (Duan et al., 2022), processing remotely sensed hyperspectral images by machine learning (e.g., Hill at al., 2020; Pyo et al., 2021), and time series modelling by machine learning (e.g., Teles et al., 2006; Recknagel et al., 2017; Pyo et al., 2020; Henrichs et al., 2021). Most of these tools perform well in terms of forecasting the timing and magnitude of peak concentrations of HCBs 5 to 20 days ahead (e.g., Recknagel et al., 2017). However, knowing the cardinal dates for beginning, peak, and ending of HCBs will allow to link early warning closer to mitigation of them (Adrian et al., 2012; Beltran-Perez and Waniek, 2021). Since the phenology of species-specific HCBs varies temporally and spatially driven by local and seasonal water quality, hydrology, and climate conditions (Beal et al., 2021), it is important to take local and temporal environmental dynamics into account that largely determine the formation and growth of HCBs (Díaz et al., 2016; Beltran-Perez and Waniek, 2021).

HCBs are consequences from intrinsic interactions in phytoplankton communities as well as responses to complex water quality, meteorological, and hydrological fluctuations, which complicate predictive modelling within the effective lead time. Time-frequency images of time-series can indicate both short-term abnormal events, such as concentrated rainfall, pulsed flow, and nutrient load from sediment resuspension (Wang et al., 2012; Stumpf et al., 2016; Adeyeri et al., 2020), and anomalies in long-term trends like seasonal and annual cycles in climatology (Grinsted et al., 2004). Thus, time-frequency images of driving variables can inform models about short-term and long-term fluctuations that may improve HCBs forecasts. Wavelet analysis has proven to be an effective tool for signal and image processing, providing information within one signal on both stochastic and periodic events simultaneously due to its time-frequency localisation characteristic (Sundararajan, 2016; Nourani and Partoviyan, 2018). Several studies linked the wavelet analysis to artificial neural networks for prediction of precipitation (Nourani et al., 2009), water temperature, dissolved oxygen, and conductivity (Saber et al., 2020), and cyanobacterial cell density (Xiao et al., 2017; Heddam et al., 2022; Jiang et al., 2021). Since wavelet transformed images from time series of environmental variables can indicate and quantify sources of variation and time lags, predictions based on feature extraction considering the whole time and frequency dimension are expected to outperform those considering a manually selected or specific spectrum. Furthermore, as a signal processing method, segmentation or the segment combination with short time-series of interest would allow to interpret a signal more accurately and improve signal classifier accuracy (Grandy et al., 2016; Ho et al., 2017; Sabour and Benezeth, 2022).

Deep learning techniques are recent advances in machine learning by neural networks in terms of the depth or number of hidden layers, which indicates that they can automatically and adaptively learn from data representations by extracting and selecting features from unstructured or unlabeled data (LeCun et al., 2015). Traditional machine learning algorithms have limitations that the quality of the extracted features determines the model performances, and the feature extraction depends on the experience of researcher. Deep learning by convolutional neural networks (CNNs) does not require separate feature extraction and modelling (classification or estimation) tools. By utilizing a small grid of parameters called the kernel matrices, an optimizable feature extractor, these algorithms construct high-level features with the convolution operation (Hinton et al., 2006). The motivation to use CNNs is to utilize CNNs in combination with wavelet analysis is to utilize the ability of extracting salient features of a signal in both time and frequency components. These have proven to be successful when integrated with wavelet transformation in solving a binary classification problem of electroencephalograph signals in medicine (Morabito et al., 2019). In ecology, CNNs are commonly applied to recognition of pictorial, acoustic and thermal images, whereby time-series properties exposed by wavelet transformation open new opportunities for time-series modelling by CNNs (Recknagel, 2023).

This study applies CNNs for time-series modelling of HCBs based on time-frequency maps of environmental drivers transformed by wavelets in four South Korean rivers where HCBs events were recorded between 2016 and 2022. In particular, to evaluate the suitability of the modelling framework, performances of CNNs were compared in terms of four determinants of input images: methods of estimating critical timings, the number of segments, time-series continuity, and image size. The predictability of time-series signals of environmental drivers on HCBs was evaluated by the following sequences: First, estimation of cardinal dates corresponding to the beginning and peak of HCBs based on historical data. Second, determining key environmental drivers by phase analysis. Third, calculation of the lead time of key environmental drivers related to *Microcystis* cell densities during HCBs. Third, training of lead time related time-frequency maps from time series of key drivers for HCBs. Fourth, determining HCBs occurrence and intensities from time-frequency maps of key drivers by CNNs.

2. Materials and Methods

2.1. Study site and data collection

The study sites include the four large rivers of the Republic of Korea, situated between 33º N and 43º N latitudes and between 124º E and 132º E longitudes. Located in the temperate zone, this region has four distinct seasons and the rainy season of the East Asian monsoon in summer. As about 60 percent of precipitation falls in summer, series of weirs were constructed in the main channel of the rivers to secure water resources and control flood risk by maintaining a water level and regulating river discharge (see Fig. 1). The Nakdong River is the longest with a length of 510 km, and the annual outflow of the river is highest in the Han River (174 billion m^3). Those rivers are considered to be eutrophic with dissolved inorganic phosphate concentrations ranging between 17.1 μg L^{-1} (Han River) and 64.4 μg L^{-1} (Yeongsan River) resulting in frequent HCBs by *Microcystis aeruginosa*. Table S1 summarizes limnological properties of the rivers.

Cyanobacteria cell density was monitored weekly along ten stations of the rivers from 2016 to 2022 (Fig. 1). Following official test standards for environmental pollution (details at http://law.go.kr), water samples were collected from the monitoring site using van Dorn sampler, and samples for cell enumeration were fixed with Lugol's iodine solution. The quantitative assessment of phytoplankton was performed using a Sedgewick Rafter counting chamber under multiple magnifications (200 × and 400 ×) of an upright microscope (Axioskop, Carl Zeiss, Oberkochen, Germany). In particular, colonies of *Microcystis* cells were completely separated into individual spherical ones to measure their exact densities.

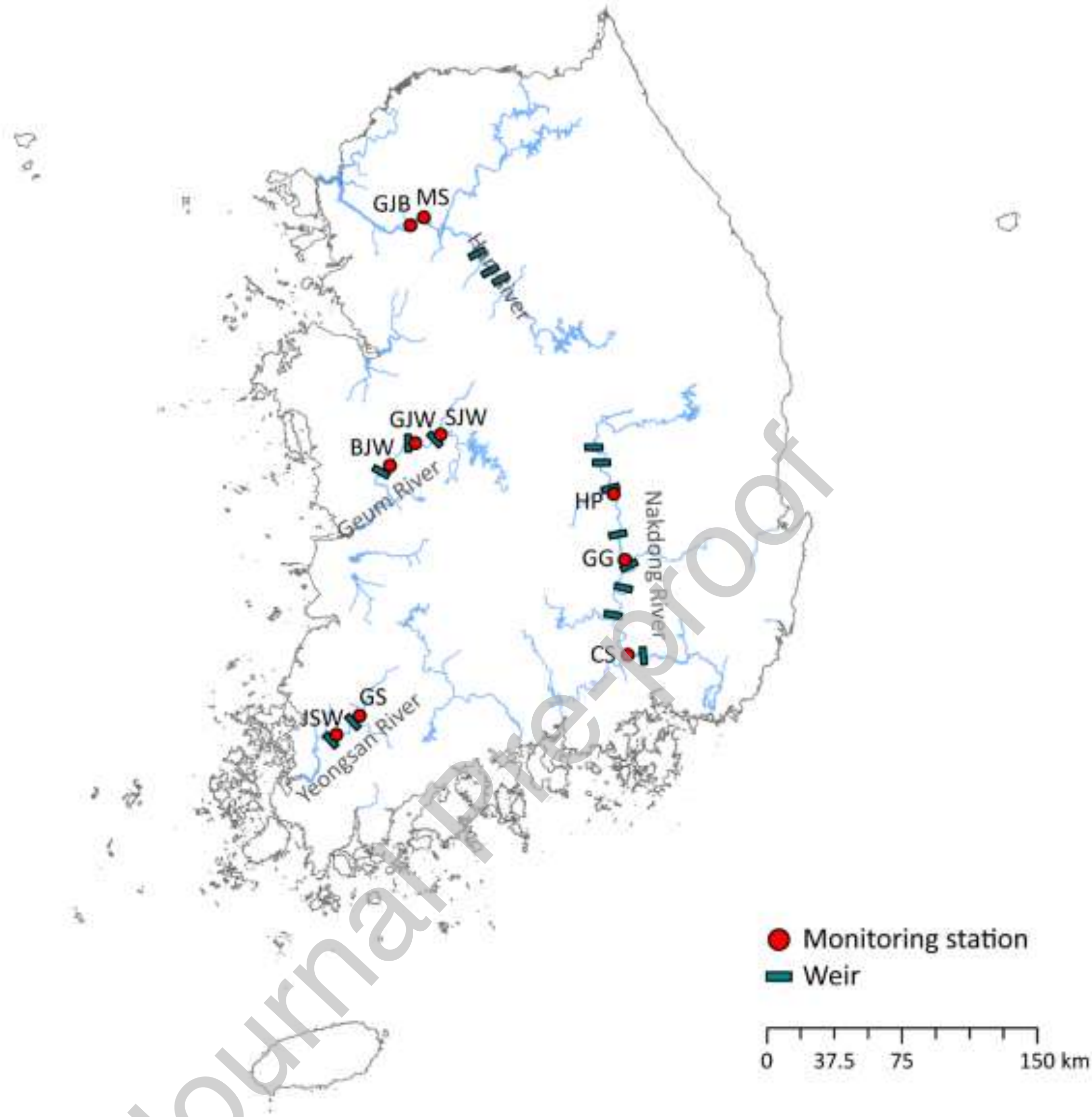


Figure 1. Map of study sites in the four rivers of the Republic of Korea.

As summarized in Table S1, the data base of the four rivers included thirteen environmental variables (water temperature (WT), dissolved oxygen (DO), suspended solids (SS), pH, conductivity (EC), total nitrogen (TN), total phosphorus (TP), nitrate (NO_3), ammonia (NH_4), and dissolved inorganic phosphorus (DIP) sampled daily from spring to autumn, and monthly in winter. Daily discharge (DIS) data were collected at the nearest station of each study site, measured by the National Water Resource Management Information System. Precipitation (PRE) and average wind velocity (Wind) of the day were gathered at the nearest station to each site, recorded by Korea Meteorological Administration. Since the intervals for measuring the historical data ranged between daily, weekly, and biweekly, and sampling dates differed for each variable, the data were interpolated to suit daily time steps using linear interpolation. These variables will be investigated as potential drivers in this study since they are either directly or indirectly linked to HCBs.

2.2. Workflow for time-series modelling of HCBs by convolution neural networks

The proposed modelling procedure for of HCBs by CNNs included three steps: (1) characterization of HCBs with determination of cardinal dates, (2) wavelet analyses of environmental variables, and (3) design and applications of CNNs on classification and regression tasks. The Fig. 2 illustrates the workflow along the three steps and four determinants of input images. The following sections from 2.2.1 to 2.2.3 provide details of these procedural steps.

4. Discussion

This study investigated the feasibility of time-series modelling of HCBs by means of wavelet transformed time-frequency images and CNNs. To derive the timing of beginning and peak of HCBs, the three methods POI, WF and LS were applied that are typically used for analysing phenology of spring blooms, cyanobacterial blooms, and phytoplankton and zooplankton dynamics (Rolinski et al., 2007; Adrian et al., 2012; Beltran-Perez and Waniek, 2021). The methods POI and WF proved to deviate within a reasonable range for the time of start and peak of HCBs, which is in agreement with Rolinski et al. (2007). Generic high-density blooms reached their maximum in mid-August, while the beginning and end dates were slightly different between the estimation methods.

The phase analysis by wavelets revealed the lead time of 32 days for WT, 29 days for SS, DIS, and PRE, and 36 days for TP, suggesting that the periods from mid-May to mid-July and from late-April to mid-July were of interest. Since the lead time of HCBs can vary according to their environmental conditions (Beltran-Perez and Waniek, 2021), environmental variables play a central role in regulating the phenology of HCBs. Time lags of growth of cyanobacteria and dinophyta in response to water quality and meteorological variables have also been identified using wavelets by Recknagel et al. (2013) and Zhang et al. (2014). Several modelling studies suggested a variety of drivers of HCBs, including nutrient loads, hydrodynamics, and meteorological conditions at the different time scales (Stumpf et al., 2016; Beal et al., 2021). While short-term and near-term forecasts should enable operational control of HCBs (Wynne et al., 2013), longer-term forecasts based on nutrient loads or hydroclimatic variables as drivers allow preventative management not feasible at short timescales (Beal et al., 2023). Thus, trained by drivers within the time window that favors the development of HCBs, CNN models can serve as tools for predicting lead time signals of environmental drivers for HCBs.

The wavelet images of time-frequency domain mainly reflected in the frequency band of 2-58 days for discontinuous segments by converting the raw environmental signal into a re-organized 2D feature matrix. The key of wavelet analysis is partitioning the variation of signals into two domains, frequency and time location, which allows us to zoom in or out some detailed variations occurring at a specific temporal scale and time location (Sundararajan, 2016; Nourani and Partoviyan, 2018). By extracting features (e.g., short-term abnormal events or anomalies in long-term trends) from the raw signals, wavelet analysis has been used to explain inter-annual variability and detect short periodicities as related to hydroclimatic factors such as rainfall, discharge, and temperature (Wang et al., 2012; Stumpf et al., 2016; Adeyeri et al., 2020). Because cyanobacterial development is determined by seasonal and annual periodicities in environmental drivers, wavelet transformed image feature provides an opportunity to analyze relationships by decomposing a time series into a time-frequency space to explain both, the dominant modes of variability and how these vary in time. Furthermore, HCBs classification and estimation across discontinuous segments were comparably accurate and precise, despite segment length. Several studies suggest that short time-series of interest might be not necessarily acquired continuously, because it is more accurate to eliminate a noisy signal or avoid interference from the other time window (Ho et al., 2017; Chen et al., 2020; Sabour and Benezeth, 2022). Our results confirm that wavelet transformed time series signals of the time window between the dates of beginning and peak of HCBs can serve as sufficient data for early warning modelling of HCBs.

CNNs classification and regression of wavelet transformed time-frequency maps has the potential to forecast HCBs with high accuracy. Although wavelet transformation provides information on the strength and persistence of patterns in environmental signals, it can be difficult or subjective to capture variation in multiple scales in terms of each frequency and time unit. CNNs, which were employed for feature extraction, characterized time-series wavelet coefficients, such as mean values, standard deviation, and skewness (Morabito et al., 2019). By automatically extracting, selecting, and fusing features, the CNN models reduced the classification and prediction errors of HCBs from the time-frequency images. CNNs have shown the superior performance on extracting and processing information from image data with a bottom-up approach, where small and less complex features are successively combined to larger and complex features (Richter et al., 2021). This indicates that the size of the input image resolution should be scaled to achieve high efficiency (Tan and Le, 2019). The dimension of wavelet power spectrum can be different according to the length of time-series data, which impose complexity and redundancy in computation and calculation. Comparison results between input image sizes suggested that by convention resized images to a fixed square can provide enough information for HCBs classification and prediction. Thus, reconstructing wavelet images automatically in the convolution layer of the CNNs to enable learning non-periodic abrupt changes as well as periodic gradual changes of time series, which allows CNNs to model HCBs both qualitatively and quantitatively.

The CNNs allowed to test combined effects of environmental drivers on HCBs. Results have shown that CNN models considering both DIS and PRE as drivers for 3 year-segments performed better than models with sole drivers or other combinations of drivers. These findings correspond with other studies suggesting that model performances can be degraded by less-correlated drivers (Hill et al., 2020; Lee et al., 2022). Also, the identification of DIS and PRE as key drivers reflects the impact of hydrodynamic processes on river habitats on cyanobacteria growth distinctly caused by flow regulation and the monsoon season typical for the studied rivers (Kim et al., 2021). The CNNs trained by wavelet transformed time-frequency images of single drivers discriminated successfully high bloom and low bloom events, achieving a classification accuracy and F1 coefficient of more than 90% and 0.90 respectively. Proving successful modelling by means of a simplified but relevant dataset not only makes monitoring and prediction more efficient, but also simplifies model applications (Xiao et al., 2017). Thus, the CNN models can provide accurate and reliable forecasts of HCBs events as well as a cost-effective modelling framework.

The proposed approach to model HCBs based on wavelet signals from drivers has the potential to predict likely intensities and densities of HCBs in advance. In view of the facts that it is still difficult utilizing long-term data with pertaining time-frequency components in one ecosystem (Lovett et al., 2007), and that HCBs are highly variable in terms of magnitude and timing when considering multiple stations (Lee et al., 2022), this study synthesized historical data from multiple monitoring stations of four rivers with similar climate, hydrologic, and trophic conditions. When the approach was applied to sequentially structured monitoring data of ten sites of four rivers, the model performances were significantly low. These experiments indicated that CNNs for long-term forecasts of HCBs require continuously measured data with high-frequency and -resolution based on at least daily or hourly time steps to represent and analyse temporal patterns of the aquatic environment in new ways (Chen et al., 2015; Xiao et al., 2017; Jiang et al., 2021). This study demonstrated the feasibility of developing early warning systems for HCBs by means of CNNs by processing multivariate time-series features.

5. Conclusions

The suggested modelling approach by means of CNNs proved to be applicable to qualitative predictions of the degree of HCBs and to quantitative predictions of HCB abundances. The CNNs were trained by time-frequency images of the drivers between the beginning and peak dates of HCBs for three-, four-, and five-years that were transformed from their time-series by wavelets. The major findings of this research are as follows:

- Wavelet transformation of time-series of drivers within the time windows of HCBs with the length of 29 to 36 days proved to be suitable for training CNNs to perform feature extraction, image classification and regression.
- The CNN models performed well in identifying intensities of HCBs qualitatively in training images with mean of accuracy 97.79 ± 0.06% and F1-score 97.49 ± 0.06%, and testing images with mean of accuracy 95.01 ± 0.06% and F1-score 93.30 ± 0.07%.
- Results of the CNNs for quantitative predictions of *Microcystis* cell densities of the last year of modelled data were also satisfying with a mean MSE of 2.58 ± 2.46 and a mean R^2 of 0.78 ± 0.20 for training and the mean MSE of 2.76 ± 2.42 and the mean R^2 of 0.55 ± 0.20 for testing dataset.
- Precipitation and discharge appeared to be the best performing drivers for qualitative and quantitative prediction of HCBs, being in accordance with the fact that cyanobacteria in rivers are largely controlled by hydrodynamic turbulences and nutrient supplies (see also Kim et al., 2021). CNNs trained by combinations of drivers revealed the best results in terms of accuracy and MSE values when discharge and precipitation were combined.
- Combinations of discontinuous data segments indicating critical timings can classify and estimate HCBs with smaller and less complex re-sized input time-frequency images.
- One-month-ahead forecasts of intensities and densities of HCBs enable to apply

operational control measures of HCBs within water bodies and in water works.

This study highlighted the feasibility of CNNs on the classification and quantification of wavelet transformed images of ecological time series with the potential to diagnose and estimate HCBs with high accuracy. It can be expected that outcomes of the proposed modelling approach for forecasting HCBs will significantly be enhanced when applied to long-term time-series monitored at high-frequency and spatial resolution across lakes or rivers. The here presented results open new opportunities for developing early warning systems for HCBs by CNNs based on driver images within lead times extracted from time series by wavelets.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant number: NRF-2021R1A6A3A03042582). The authors are grateful to Qiuwen Chen and Cheng Chen for their valuable comments on the first draft of the manuscript. The authors thank two anonymous reviewers for their very instructive comments that improved significantly the clarity of the manuscript.

References

Adeyeri, O.E., Laux, P., Lawin, A.E., Arnault, J., 2020. Assessing the impact of human activities and rainfall variability on the river discharge of Komadugu-Yobe Basin, Lake Chad Area. Environmental Earth Sciences 79, 1–12.

Adrian, R., Gerten, D., Huber, V., Wagner, C., Schmidt, S.R., 2012. Windows of change: temporal scale of analysis is decisive to detect ecosystem responses to climate change. Marine Biolology 159, 2533–2542.

Allaire, J.J., Chollet, F., Tang, Y., Falbel, D., Van Der Bijl, W., Studer, M., Allaire, M.J., 2022. Package 'keras.' R Interface to 'Keras'.

Beal, M.R.W., O'Reilly, B., Hietpas, K.R., Block, P., 2021. Development of a sub-seasonal cyanobacteria prediction model by leveraging local and global scale predictors. Harmful Algae 108, 102100.

Beal, M.R.W., Wilkinson, G.M., Block, P.J., 2023. Large scale seasonal forecasting of peak season algae metrics in the Midwest and Northeast U.S. Water Research 229, 119402.

Beltran-Perez, O.D., Waniek, J.J., 2021. Environmental window of cyanobacteria bloom occurrence. Journal of Marine Systems 224, 103618.

Chen, Q., Guan, T., Yun, L., Li, R., Recknagel, F., 2015. Online forecasting chlorophyll a concentrations by an auto-regressive integrated moving average model: Feasibilities and potentials. Harmful Algae 43, 58–65.

Díaz, P.A., Ruiz-Villarreal, M., Pazos, Y., Moita, T., Reguera, B., 2016. Climate variability and *Dinophysis acuta* blooms in an upwelling system. Harmful Algae 53, 145–159.

Duan, X., Zhang, C., Struewing, I., Li, X., Allen, J., Lu, J., 2022. Cyanotoxin-encoding genes as powerful predictors of cyanotoxin production during harmful cyanobacterial blooms in an inland freshwater lake: Evaluating a novel early-warning system. Science of The

Total Environment 830, 154568.

Grandy, T.H., Garrett, D.D., Schmiedek, F., Werkle-Bergner, M., 2016. On the estimation of brain signal entropy from sparse neuroimaging data. Scientific Reports 6, 23073.

Grinsted, A., Moore, J.C., Jevrejeva, S., 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Processes in Geophysics 11, 561–566.

Glibert, P.M., 2020. Harmful algae at the complex nexus of eutrophication and climate change. Harmful Algae 91, 101583.

Graham, J.L., Dubrovsky, N.M., Ebert, S.M., 2016. Cyanobacterial harmful algal blooms and U.S. Geological Survey science capabilities. US Department of the Interior, US Geological Survey.

Heddam, S., Yaseen, Z.M., Falah, M.W., Goliatt, L., Tan, M.L., Sa'adi, Z., Ahmadianfar, I., Saggi, M., Bhatia, A., Samui, P., 2022. Cyanobacteria blue-green algae prediction enhancement using hybrid machine learning–based gamma test variable selection and empirical wavelet transform. Environmental Science and Pollution Research 29, 77157–77187.

Henrichs, D.W., Anglès, S., Gaonkar, C.C., Campbell, L., 2021. Application of a convolutional neural network to improve automated early warning of harmful algal blooms. Environmental Science and Pollution Research 28, 28544–28555.

Hill, P.R., Kumar, A., Temimi, M., Bull, D.R., 2020. HABNet: Machine learning, remote sensing-based detection of harmful algal blooms. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 3229–3239.

Hinton, G.E., Osindero, S., Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. Neural Computation 18, 1527–1554.

Ho, T.-W., Lin, F.-C., Lin, C.-M., Lai, F., 2017. Smart computing mechanism for noise detection and elimination in ECG signal, in: 2017 IEEE International Conference on Big Data and Smart Computing (BigComp). Presented at the 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 28–33.

Isles, P.D., Pomati, F., 2021. An operational framework for defining and forecasting phytoplankton blooms. Frontiers in Ecology and the Environment 19, 443–450.

Jiang, P., Huang, Y., Liu, X., Zhang, J., Gin, K.Y.-H., 2021. A feature reconstruction-based multi-task regression model for cyanobacterial distribution forecasting along the water column. Journal of Cleaner Production 292, 126025.

Kim, H.G., Recknagel, F., Kim, H.-W., Joo, G.-J., 2021. Implications of flow regulation for habitat conditions and phytoplankton populations of the Nakdong River, South Korea. Water Research 207, 117807.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86, 2278–2324.

Lee, D., Kim, M., Lee, B., Chae, S., Kwon, S., Kang, S., 2022. Integrated explainable deep learning prediction of harmful algal blooms. Technological Forecasting and Social Change 185, 122046.

Lovett, G.M., Burns, D.A., Driscoll, C.T., Jenkins, J.C., Mitchell, M.J., Rustad, L., Shanley, J.B., Likens, G.E., Haeuber, R., 2007. Who needs environmental monitoring? Frontiers in Ecology and the Environment 5, 253–260.

Morabito, F.C., Campolo, M., Ieracitano, C., Mammone, N., 2019. Deep learning approaches to electrophysiological multivariate time-series analysis, in: Artificial Intelligence in the Age of Neural Networks and Brain Computing. Elsevier, pp. 219–243.

Morlet, J., Arens, G., Fourgeau, E., Glard, D., 1982a. Wave propagation and sampling theory—Part I: Complex signal and scattering in multilayered media. Geophysics 47, 203–221.

Morlet, J., Arens, G., Fourgeau, E., Giard, D., 1982b. Wave propagation and sampling theory; Part II, Sampling theory and complex waves. Geophysics 47, 222–236.

Nourani, V., Alami, M.T., Aminfar, M.H., 2009. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Engineering Applications of Artificial Intelligence 22, 466–472.

Nourani, V., Partoviyan, A., 2018. Hybrid denoising-jittering data pre-processing approach to enhance multi-step-ahead rainfall–runoff modeling. Stochastic Environmental Research and Risk Assessment 32, 545–562.

Paerl, H.W., Huisman, J., 2009. Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. Environmental Microbiology Reports 1, 27–37.

Pyo, J.C., Duan, H., Baek, S., Kim, M.S., Jeon, T., Kwon, Y.S., Lee, H., Cho, K.H., 2019. A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery. Remote Sensing of Environment 233, 111350.

Pyo, J.C., Park, L.J., Loo, J., Pachepskyc, Y., Baek, S.O., Kim, K., Cho, K.H., 2020. Using convolutional neural network for predicting cyanobacteria concentrations in river water. Water Research 186, 116349.

Pyo, J.C., Cho, K.H., Kim, K., Baek, S.O., Nam, G., S. Park, 2021. Cyanobacteria cell prediction using interpretable deep learning model with observed, numerical, and sensing data assemblage. Water Research 203, 117483.

Recknagel, F., Ostrovsky, I., Cao, H., Zohary, T., Zhang, X., 2013. Ecological relationships, thresholds and time-lags determining phytoplankton community dynamics of Lake

Kinneret, Israel elucidated by evolutionary computation and wavelets. Ecological Modelling 255, 70–86.

Recknagel, F., Orr, P., Bartkow, M., Swanepoel, A., Cao, H., 2017. Early warning of limit-exceeding concentrations of cyanobacteria and cyanotoxins in drinking water reservoirs by inferential modelling. Harmful Algae 69, 18–27.

Recknagel, F., 2023. Cyberinfrastructure for sourcing and processing ecological data. Ecological Informatics 75, 102039.

Richter, M.L., Byttner, W., Krumnack, U., Wiedenroth, A., Schallner, L., Shenk, J., 2021. (Input) Size matters for CNN classifiers, in: Farkaš, I., Masulli, P., Otte, S., Wermter, S. (Eds.), Artificial Neural Networks and Machine Learning – ICANN 2021, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 133–144.

Roesch, A., Schmidbauer, H., Roesch, M.A., 2014. Package 'WaveletComp.' The Comprehensive R Archive Network2014.

Rolinski, S., Horn, H., Petzoldt, T., Paul, L., 2007. Identifying cardinal dates in phytoplankton time series to enable the analysis of long-term trends. Oecologia 153, 997–1008.

Saber, A., James, D.E., Hayes, D.F., 2020. Long-term forecast of water temperature and dissolved oxygen profiles in deep lakes using artificial neural networks conjugated with wavelet transform. Limnology and Oceanography 65, 1297–1317.

Sabour, R.M., Benezeth, Y., 2022. Gated recurrent unit-based RNN for remote photoplethysmography signal segmentation, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Presented at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, New Orleans, LA, USA, pp. 2201–2209.

Si, B.C., Zeleke, T.B., 2005. Wavelet coherency analysis to relate saturated hydraulic properties to soil physical properties. Water Resources Research 41.

Song, C., Yao, L., Hua, C., Ni, Q., 2021. A novel hybrid model for water quality prediction based on synchrosqueezed wavelet transform technique and improved long short-term memory. Journal of Hydrology 603, 126879.

Stumpf, R.P., Johnson, L.T., Wynne, T.T., Baker, D.B., 2016. Forecasting annual cyanobacterial bloom biomass to inform management decisions in Lake Erie. Journal of Great Lakes Research 42, 1174–1183.

Sundararajan, D., 2016. Discrete wavelet transform: a signal processing approach. John Wiley & Sons.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning. PMLR, pp. 6105–6114.

Teles, L., Vasconcelos, V., Teles, L., Pereira, E., Saker, M., Vasconcelos, V., 2006. Time series forecasting of cyanobacteria blooms in the Crestuma Reservoir (Douro River, Portugal) using artificial neural networks. Environmental Management 38, 227–237.

Wang, F., Wang, X., Zhao, Y., Yang, Z., 2012. Nutrient response to periodic hydrological fluctuations in a recharging lake: a case study of Lake Baiyangdian. Fresenius Environmental Bulletin 21, 1254–1262.

WHO (World Health Organization). 2003. Guidelines for safe recreational water environments. Coastal and Fresh Waters 1, 1–219.

Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Fahnenstiel, G.L., Dyble, J., Schwab, D.J., Joshi, S.J., 2013. Evolution of a cyanobacterial bloom forecast system in western Lake Erie: development and initial evaluation. Journal of Great Lakes Research 39, 90–99.

Xiao, X., He, J., Huang, H., Miller, T.R., Christakos, G., Reichwaldt, E.S., Ghadouani, A., Lin, S., Xu, X., Shi, J., 2017. A novel single-parameter approach for forecasting algal blooms. Water Research 108, 222–231.

Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K., 2018. Convolutional neural networks: an overview and application in radiology. Insights into Imaging 9, 611–629.

Yu, Z., Yang, K., Luo, Y., Shang, C., 2020. Spatial-temporal process simulation and prediction of chlorophyll-a concentration in Dianchi Lake based on wavelet analysis and long-short term memory network. Journal of Hydrology 582, 124488.

Zar, J.H., 1999. Biostatistical analysis. Pearson Education India.

Zhang, X., Chen, Q., Recknagel, F., Li, R., 2014. Wavelet analysis of time-lags in the response of cyanobacteria growth to water quality conditions in Lake Taihu, China. Ecological Informatics 22, 52–57.

Declaration of interests

☐ 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.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Hyo Gyeom Kim reports financial support was provided by National Research Foundation of Korea.