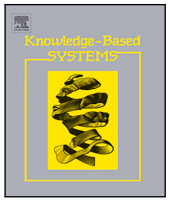
Knowledge-Based Systems 301 (2024) 112279

Contents lists available at ScienceDirect

Knowledge-Based Systems

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



Deep learning methods for multi-horizon long-term forecasting of Harmful Algal Blooms

Silvia Martín-Suazo [a], Jesús Morón-López [b], Stanislav Vakaruk [a], Amit Karamchandani [a], Juan Antonio Pascual Aguilar [c], Alberto Mozo [a,*], Sandra Gómez-Canaval [a], Meritxell Vinyals [d], Juan Manuel Ortiz [c]

[a] E.T.S. de Ingeniería de Sistemas informáticos, Universidad Politécnica de Madrid, Calle Alan Turing s/n, Madrid, 28031, Spain
[b] European Regional Centre for Ecohydrology of the Polish Academy of Sciences, 3 Tylna, Łódź, 90-364, Poland
[c] IMDEA Water Institute, Avenida Punto Com, 2, Parque Científico Tecnológico de la Universidad de Alcalá, Alcalá de Henares, 28805, Spain
[d] Université Fédérale de Toulouse, INRAE, MIAT, 24 Chem. de Borde Rouge, Castanet-Tolosan, 31326, France

ARTICLE INFO

Dataset link: https://github.com/stanislavvakaruk/Chlorophyll_soft-sensor_machine_learning_models

Keywords:
Deep learning
Chlorophyll
Harmful algae bloom
Monitoring
Time-series
Forecasting

ABSTRACT

The increasing occurrence of Harmful Algal Blooms (HABs) in water systems poses significant challenges to ecological health, public safety, and economic stability globally. Deep Learning (DL) models, notably Convolutional Neural Networks (CNN) and Long-Short Term Memory (LSTM), have been widely employed for HAB prediction. However, the emergence of state-of-the-art multi-horizon forecasting DL architectures such as Neural Basis Expansion Analysis for Interpretable Time Series Forecasting (N-BEATS) provides a novel solution for long-term HAB prediction. This study compares the performance of N-BEATS with LSTM and CNN models using high temporal granularity water quality data from As Conchas reservoir (NW Spain) to forecast chlorophyll-a (Chl-a) concentrations, a key indicator of HABs. The evaluation encompasses one-day and one-week prediction horizons, aligning with World Health Organization (WHO) HAB alert criteria. Results indicate that N-BEATS outperforms LSTM and CNN models for one-week predictions and when forecasting multiple consecutive days within a week. Furthermore, augmenting input data with additional variables does not significantly enhance predictive accuracy, challenging the assumption that complexity always improves model performance. The study also explores the transferability of trained models across different monitoring buoys within the same water body, emphasizing the adaptability and broad applicability of predictive models in diverse aquatic environments. This research underscores the potential of N-BEATS as a valuable tool for HAB prediction, particularly for longer-term forecasting.

1. Introduction

In the Anthropocene era, in which human activity has had a significant impact on the planet's climate and ecosystems, the relationship between societal needs and water resources has become very complex [1]. This is evidenced by the widespread and record-breaking increase of Harmful Algal Blooms (HABs) in inland water bodies and coastal areas around the world [2,3]. HABs are intensifying as a result of rising temperatures and high nutrient loads mainly from agriculture and livestock runoff [4]. They, in turn, endanger the environment, animals and humans due to toxin production and dissolved oxygen depletion following bloom collapse [5,6]. Estimates point to an annual cost of billions of dollars in public health and loss of ecosystem services (e.g. fisheries, tourism and recreation) [7,8]. Therefore, if the United Nations Sustainable Development Goal on Water is to be achieved (Goal 6: Ensure access to water and sanitation for all) [9], it is imperative to intensify the monitoring of HABs and effectively coordinate climate and water-related actions.

In this perspective, Artificial Intelligence (AI) is increasingly proposed to modulate integrated water management [10]. Through a subset of algorithmic models based on techniques such as Machine Learning (ML) and Deep Learning (DL), AI is able to learn and predict outcomes through the observation of environmental changes. The use of AI for HABs forecasting has been explored mainly with hyperspectral data collected by remote sensing [11]. Instruments aboard satellites and aircrafts are capable of measuring water reflectance signals resulting from light absorption and scattering by algal cells. Numerous studies have applied ML and DL models to interpret these water-leaving reflectance signals and obtain information on the presence, abundance

* Corresponding author.
E-mail addresses: j.moronlopez@erce.unesco.lodz.pl (J. Morón-López), a.mozo@upm.es (A. Mozo).

https://doi.org/10.1016/j.knosys.2024.112279
Received 22 March 2024; Received in revised form 7 July 2024; Accepted 24 July 2024
Available online 27 July 2024

and spatial distribution of HABs in surface waters [12–15]. However, although remote sensing offers wide spatial coverage, it only provides daily or weekly data due to overflight limitations or interference caused by cloud cover, complex water optics or weather conditions [16]. In addition, geographical resolution poses challenges for small water bodies, as pixels may overlap with those of the Earth, making spatial estimation difficult. This may lead to inaccurate predictions because the spatial and temporal distribution of HABs may show short-term variations in surface waters. It is therefore necessary to combine the strengths of remote sensing with complementary technologies operating at near real-time scales to mitigate temporal uncertainty [16,17].

High temporal granularity of water quality data is possible by deploying sensors directly in the water. There are sensors to measure the Chlorophyll-a (Chl-*a*), a pigment considered a proxy indicator of algal concentration in surface waters, as well as other physicochemical parameters related to HAB dynamics, such as pH, temperature, dissolved oxygen, conductivity, and turbidity [18]. These sensors may be integrated into buoys as Automatic High-Frequency Monitoring (AHFM) systems to collect and transmit massive data in real time [19–21]. The vast amount of data that water quality sensors are capable of collecting at scales of minutes or hours successfully meets AI requirements for complex water quality assessment [22]. Several studies have trained ML and DL models with long-term data to infer Chl-*a* concentration or algae cell density [23–26]. These models may be particularly suitable to triggering various levels of alert before algae cells exceeds certain preset thresholds, thus serving as critical indicators of potential risks and guiding appropriate responses to HABs [24,27]. Typical alert levels for short-term responses proposed by the World Health Organization (WHO) are based on factors such as Chl-*a* concentration, algal type and abundance, toxin levels, and environmental conditions [28]. At vigilance level, routine monitoring continues with heightened awareness. The alert level 1 involves increased monitoring at least once a week and advisories as algal cell counts rise. Increasing the frequency of monitoring is important considering that certain harmful algal species (especially cyanobacteria) can two- or threefold in a few days in natural water bodies [28]. A warning arises with high cell counts, visible blooms, and concerning toxin levels, leading to public advisories and potential restrictions. The alert level 2, triggered by severe conditions, demands immediate action, including water body closures and evacuation directives. In the recently established alert levels framework for HABs in recreational water bodies, the vigilance level is associated with Chl-*a* values ranging from 3–12 μg/L. Alert 1 and Alert 2 are activated when Chl-*a* levels fall within the range of 12–24 μg/L, or if they surpass these values with a prevalence of harmful algal species. For drinking water, the threshold values are significantly lower. The vigilance level is triggered at concentration of Chl-*a* of 1 μg/L, and alert level 1 and 2 at concentrations of Chl-*a* between 1–12 μg/L or higher. Although there are regional variations in specific alert levels and corresponding actions, the imperative for continuous vigilance, timely communication and collaborative efforts remains constant. Therefore, the use of predictive models to aid in sampling and triggering alerts is a pivotal strategy for implementing accurate and cost-effective early warning systems to guide coordinated decision making in water bodies vulnerable to HABs.

Conventional ML algorithms (i.e. non-deep), such as Linear Regression (LR), Classification And Regression Tree (CART) [25], Random Forest (RF) [25,29–31], and Support Vector Machine (SVM) [32,33], have proven to be adequate for predictions at very short timescales (minutes) or to build low-cost soft-sensors. However, these models may fall short in dealing with complex nonlinear relationships and short- to medium-term temporal dependencies (i.e. days to weeks) present in water quality data. In this context, DL algorithms promise to overcome the limitations of conventional ML models. For instance, a basic DL architecture as Multi-Layer Perceptron (MLP) has been applied in studies related to algae concentration prediction [34–37]. More advanced DL models, including Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models, have also been investigated within this context for their potential in managing temporal dependencies in the data [23,38,39]. Notably, Barzegar et al. [23] proposed a hybrid LSTM and Convolutional Neural Network (CNN) model, which demonstrated superior performance in predicting Chl-*a*. Traditional architectures, such as LSTM and CNN, have long been the stalwarts in tasks requiring predictions over a fixed horizon. However, the advent of modern DL architectures specifically engineered for multi-horizon forecasting, represents a significant stride forward. These architectures can be trained in a sequence-to-sequence manner to produce results across multiple prediction horizons in a single iteration, eliminating the necessity to train separate models for each individual horizon. Not only do they offer enhanced computational efficiency, but these models also bestow a greater degree of flexibility compared to their fixed-horizon counterparts. This allows end-users to select the most suitable prediction horizon based on their unique requirements and the tolerable bounds of prediction error. Recently, the domain of interpretable DL algorithms has witnessed the inception of "Neural Basis Expansion Analysis for Interpretable Time Series Forecasting" (N-BEATS), a recently multi-horizon prediction framework proposed by Oreshkin et al. [40]. This architecture consists of a succession of densely stacked blocks, each incorporating numerous fully connected layers interconnected through residual links. This model design offers unique advantages: (i) it evidences a broad adaptability across a diverse range of target domains without the need for specific modifications; (ii) it is rapid to train; and (iii), importantly, it has the capacity to produce interpretable results. The N-BEATS architecture has been widely employed to address various traditional forecasting challenges, and its applications extend to various sectors such as energy [41], healthcare [42,43] and telecommunications [44]. In light of these improvements and advances in other fields, it is relevant to evaluate the potential of the N-BEATS algorithm for HAB forecasting and alert generation.

The aim of this research is to improve HABs prediction by examining the effectiveness of the N-BEATS algorithm in relation to the most commonly used LSTM and CNN models. Specifically, this research delves into the application of DL techniques to a multi-horizon forecasting problem, thus generating predictions with an hourly granularity and extending up to a one-week horizon. The choice of these time scales is in line with WHO recommendations, which advocate an intensification of sampling frequency as higher alert levels are reached [28]. A high degree of accuracy and breadth of predictions could transform water body management by reducing near-future uncertainty. In particular, the inherent versatility of the multi-horizon forecasting model allows experts to select from a variety of future projections, a feature that may be highly beneficial in light of the typically observed deterioration of prediction accuracy over extended horizons. A highlight of our research is the exploration of the ability to detect the initial alarm level of HABs within a multi-horizon forecasting framework. The development of the HABs alert level based on a forecasting of Chl-*a* concentration may provide experts with valuable early warnings, thus enabling more responsive approaches to water body monitoring. This advance in future alert prediction methodologies could greatly improve coordination between mitigation strategies under a holistic approach.

This study is drawn from a unique dataset obtained from the As Conchas reservoir (Galicia, NW Spain). This dataset, collected assiduously over a period of three years at 15-min intervals from two buoys equipped with sensors, serves as an excellent context for examining the effectiveness of different DL techniques. Several DL techniques under two distinct scenarios were evaluated: the first employs solely historical Chl-*a* measurements as input, while the second incorporates additional exogenous variables from the buoys, such as water temperature, pH, and conductivity. The objective of comparing these two scenarios was to assess whether augmenting the input data with these supplementary variables significantly enhances the predictive accuracy of the models. The trend and periodicity of these additional variables (which correlate

with Chl-*a* as observed in our previous work [25]) are expected to provide useful information about the future behavior of Chl-*a* and enhance the predictive capability of the data. Additionally, we delved into the concept of model transferability, assessing whether a model trained on data from one buoy may sustain its accuracy when applied to another. The findings from this component of the study could provide insight into the generalizability of the model across water bodies and potentially pave the way for more efficient methods of training the model, obviating the need for individual buoy-based training.

In summary, this study represents a significant advancement in the state-of-the-art for predicting algae concentration, thanks to the following noteworthy contributions:

- A comprehensive comparison is conducted between the recent DL technique N-BEATS and other well-established DL methods, namely LSTM and CNN networks. This evaluation focuses on the prediction of HABs, which serves to enrich existing literature while offering valuable insights into the relative performance of these sophisticated methods.
- A method is proposed to compare different forecasting methods in the context of a multi-horizon forecasting problem. This research challenges traditional methodologies in predictive modeling of HABs by combining a granularity of one hour and a forecasting horizon extending up to one week.
- A distinctive aspect of this research lies in the evaluation of forecasting techniques with respect to their ability to predict the HABs alert level in a multi-horizon forecasting context (e.g., rather than relying on more standard metrics such as value prediction error). The introduction of these advanced techniques offers a novel approach to effectively plan water body sampling, with the potential to significantly increase the effectiveness of HABs management and mitigation efforts.
- This research represents a pioneering application of these advanced techniques to a meticulously collected dataset from As Conchas reservoir. This unique dataset allows for a comparison not only based on models trained on historical Chl-*a* values, but also on a set of additional exogenous variables acquired from sensors on-board buoys. This comprehensive examination allows for a better understanding of the benefits and limitations associated with integrating different sensors into AHFM systems.
- This study also provides an analysis of the transferability of trained models between different buoys located in the same water body. The results of this analysis have substantial implications for the acceleration and efficiency of model training, leading to new strategies for the deployment of these models in various environments.

This article is structured as follows. Firstly, a theoretical background is provided in Section 2 in order to facilitate the knowledge to comprehend the solution proposed in this article. Secondly, the proposed solution is presented in Section 3, offering a detailed description of the procedures undertaken to prepare the experiments performed. In Section 4, the results of the proposed methods are exhaustively evaluated and discussed. Finally, the study's main findings and conclusions are detailed in Section 5.

2. Theoretical framework

In this section, we will delve into the concept of time-series forecasting and explore the various methods employed in this domain. Additionally, we provide an in-depth explanation of the theoretical aspects related to several ML/DL architectures widely used in the forecasting field, and we examine the fundamental principles and mechanisms underlying these architectures and highlight their respective strengths and applications in the context of forecasting tasks.

2.1. Time-series forecasting description

Time series represents a collection of data points sequentially gathered over uniform time intervals, characterized by their chronological order. This temporal ordering is essential, as each point is interdependent and contextually linked to its time of occurrence. Central to diverse scientific and practical fields, time-series analysis is a statistical methodology employed to examine and predict data points gathered over a specific time span. Within this analysis, a fundamental distinction is made between endogenous and exogenous variables. Endogenous variables are contingent upon the system under investigation, while exogenous variables are influenced by external factors that can affect the endogenous variables. In the context of time-series forecasting, endogenous variables (Y_i) refer to the data points that are predicted, whereas exogenous variables (X_i) are not forecasted but can be supplied as input to the model to enhance the accuracy of the forecast.

Within the realm of forecasting methodologies, several forecasting methods are typically employed. One-step-ahead forecasting consists of predicting one time step ahead from the current moment using observed data and prior predictions to generate the forecast, which can be utilized iteratively to produce indefinite future predictions (see the top result One-step-ahead in Fig. 1). Nonetheless, this method may not be optimal for long-term forecasting problems due to the potential accumulation of errors from each intermediary prediction. Formally, we can define one-step-ahead forecasting as follows. Let I represent the distinct set of samples in a specific time-series dataset. Each sample i within the set I corresponds to a scalar input $X_{i,t} \in \mathbb{R}$ and a target $Y_{i,t} \in \mathbb{R}$ at every time step $t \in [0, T_i]$, where T_i signifies the size of each complete sample i (i.e., the length of the time series). One-step-ahead forecasting aims to predict the value of the target variable $Y_{i,t+1}$ in the subsequent time step, given the prior predicted steps and historical data up to time t.

$$\hat{Y}_{i,t+1} = f(Y_{t-K:t}, X_{i,t-K:t}) \tag{1}$$

In this equation, K denotes the size of the previous temporal window, and f represents a function that maps the inputs and history to the predicted output at time $t+1$. The objective is to identify an appropriate function f that minimizes the error between the predicted output $\hat{Y}_{i,t+1}$ and the ground truth value $Y_{i,t+1}$ on a validation set. The size of the past temporal window K relies on the problem and must be ascertained empirically.

To create a forecasting sequence using one-step-ahead forecasting, we start with an initial input $X_{i,0}$ and apply the model f to predict the output $Y_{i,1}$. Subsequently, we use the predicted output as the input for the next time step, that is, $X_{i,1} = Y_{i,1}$, and iterate the process to generate a sequence of predicted outputs $Y_{i,2}, Y_{i,3}, \ldots, Y_{i,T_i}$. As evident, errors in each intermediary prediction can propagate to later predictions, causing a compounding effect on the overall forecast accuracy.

In opposition, fixed-horizon forecasting is tailored to predict a specific future time step with optimized accuracy, thereby avoiding error accumulation (see the central result Fixed-horizon in Fig. 1). However, this technique is restricted to predicting only the particular step and cannot be employed for earlier or later predictions. Fixed-horizon forecasting focuses on predicting the target variable $Y_{i,T}$ at a specific future time T, given the input variables $X_{i,T}$ and the history of the time series up to time t, i.e., $(X_{i,0}, Y_{i,0}), \ldots, (X_{i,t}, Y_{i,t})$. This can be formalized as follows:

$$\hat{Y}_{i,T} = f(Y_{t-K:t}, X_{i,t-K:t}) \tag{2}$$

In this formula, T represents the time at which the prediction occurs, $Y_{t-K:t}$ denotes the past values of the time series up to time t, and $X_{i,t-K:t}$ corresponds to the input variables associated with the time series over an identical time window. The function f maps the

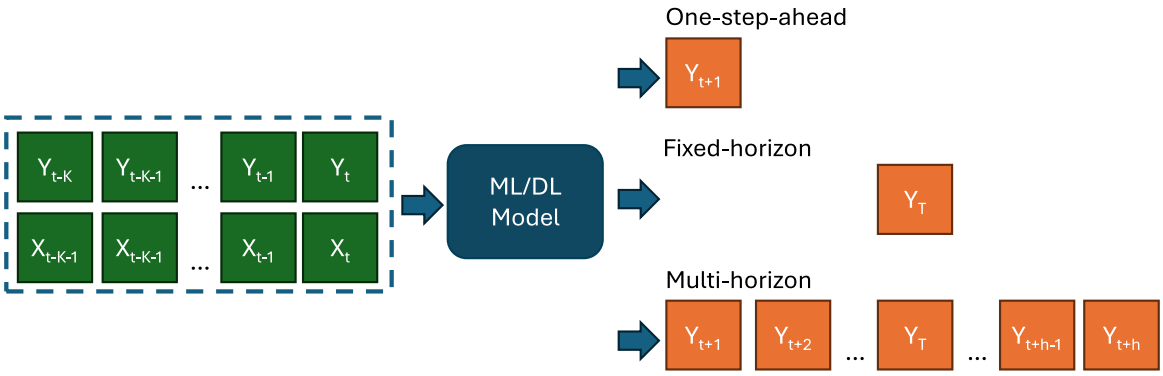


Fig. 1. Graphical representation of the input–output structure of a prediction model according to the type of horizon: One-step-ahead, Fixed-horizon, and Multi-horizon.

historical values and inputs to the predicted value $\hat{Y}_{i,T}$ at time T. The goal is to find an appropriate function f that minimizes the prediction error between the predicted value $\hat{Y}_{i,T}$ and the ground truth value Yi,T on a validation set.

Multi-horizon forecasting offers an alternative approach that encompasses predicting a series of future steps (see the bottom result Multi-horizon in Fig. 1). This method strives to provide flexibility in the forecasting process and minimize cumulative errors, although it may not achieve the same level of accuracy as the fixed-horizon method for predicting a singular point in time [45]. The multi-horizon forecasting technique involves predicting the target variable for a series of future time steps, i.e., $Y_{i,t+1}, Y_{i,t+2}, \ldots, Y_{i,T}$. This can be formalized as follows:

$$\hat{Y}_{i,t+1:t+h} = f_h(Y_{t-K:t}, X_{i,t-K:t}) \tag{3}$$

In the expression above, h denotes the forecast horizon, $\hat{Y}_{i,t:t+h}$ represents the predicted sequence of the h-step-ahead forecast with respect to the time step t, K indicates the size of the past temporal window, and f represents a function that maps the inputs and history up to time t to the predicted sequence $\hat{Y}_{i,t+1:t+h}$. The objective is to find a suitable function f_h for each horizon h that minimizes the error between the predicted output $\hat{Y}_{i,t:t+h}$ and the ground truth value $Y_{i,t:t+h}$ on a validation set, typically using a metric that assesses the overall accuracy of the predictions, such as the mean squared error or the mean absolute error. However, in the case of using DL, a single model can be trained to predict multiple horizons concurrently, an approach that is referred as Multi-output Multi-input (MIMO) forecasting [46]. In this approach, the model is trained to predict the target variable for multiple future time steps, i.e., $Y_{i,t+1}, Y_{i,t+2}, \ldots, Y_{i,T}$, simultaneously. This technique allows for increased flexibility and accuracy in long-term forecasting, as it can capture complex nonlinear relationships between the input and output variables, and errors do not accumulate over time.

2.2. Selection of methods for time-series forecasting

This study aims to assess the performance of the N-BEATS architecture as representative of state-of-the-art sequence-to-sequence time-series forecasting method, in comparison to traditional ML/DL architectures for the given problem. The selection of N-BEATS as the DL architecture was motivated by its exceptional performance in well-established forecasting competitions [40]. As commented above, N-BEATS is specifically designed for time-series forecasting and has demonstrated remarkable results with univariate time-series data. The N-BEATS architecture offers several advantageous characteristics. It can be directly applied to various problem types without requiring extensive feature engineering. Furthermore, it exhibits faster training and greater scalability compared to other DL architectures like LSTM. Additionally, it possesses model interpretability capabilities that are lacking in other DL architectures [40]. Moreover, studies have shown that the N-BEATS architecture demonstrates better generalization abilities than other DL models when trained on a specific source time-series dataset and applied to a different target dataset [40,42–44,47].

In order to ensure a fair and meaningful comparison between N-BEATS and other well-established machine learning and deep learning models, we conducted a thorough review of recent literature on time-series forecasting. The primary objective of this review was to identify the most relevant models for comparison, taking into consideration their performance and applicability in the context of time-series forecasting. Among the wide range of architectures available, the Long Short-Term Memory (LSTM) neural network architecture was selected due to its exceptional performance in capturing temporal dependencies within univariate and multivariate time-series forecasting, as highlighted in several studies [48–50]. LSTMs are particularly suitable for time-series forecasting tasks that rely on long-term dependencies. They excel at modeling sequential data by effectively capturing trends, seasonality, and irregular patterns. Unlike traditional recurrent neural networks, LSTMs are capable of handling input sequences of varying lengths and are resistant to the issues of vanishing and exploding gradients. Their ability to retain information over extended periods makes them well-suited for forecasting tasks that require the consideration of extensive historical context.

Convolutional Neural Networks (CNNs), originally designed for image processing, have also gained attention in the domain of time-series forecasting [47,48]. By leveraging the concept of convolutional operations, CNNs can effectively extract local features and patterns from sequential data. This characteristic makes CNNs suitable for capturing temporal dependencies within time-series data, such as trends and recurring patterns. Several studies have highlighted the effectiveness of CNNs in time-series forecasting tasks [47,48]. By employing convolutions across different dimensions of the input data, CNNs can effectively capture single variable and cross-variable temporal dependencies. This ability is particularly advantageous in time-series analysis, as it enables the network to learn complex relationships and interactions between different variables, thereby potentially enhancing overall predictive performance. Additionally, CNNs are computationally efficient due to shared weights and parallel processing, enabling them to handle large-scale time-series datasets efficiently and making them well-suited for real-time applications.

This study also combined CNNs and LSTM networks to build a hybrid model. The integration of CNNs and LSTMs aims to leverage the complementary strengths of both architectures and enhance the predictive capabilities of the overall model in time-series forecasting tasks. The hybrid model combining CNNs and LSTMs offers the potential to capture both local and global temporal dependencies within time-series data. By employing CNNs for initial feature extraction and leveraging the ability of LSTMs to capture long-term dependencies, the hybrid model can effectively capture both short-term patterns and long-term trends in the data. Several studies have demonstrated the effectiveness of this hybrid approach in improving the predictive performance of time-series forecasting models [51–53].

2.3. Overview of selected methods for time-series forecasting

In this subsection, we present a comprehensive analysis of the theoretical aspects of several Machine Learning (ML) and Deep Learning (DL) architectures employed for forecasting tasks.

2.3.1. Fully connected neural network

Fully Connected Neural Networks (FCNNs), also referred as multilayer perceptrons, are a fundamental form of feed-forward neural networks extensively utilized in DL. These FCNNs consist of interconnected neurons organized into layers, with each neuron in a given layer connected to all neurons in the subsequent layer, resulting in a dense network structure [47]. The layers within an FCNN can be categorized into three types: the input layer, hidden layers, and output layer.

The choice of the number of neurons in the hidden layers is crucial and involves striking a balance between model complexity and generalization capability. A higher number of neurons allows the FCNN to learn complex patterns by capturing a larger number of features, but too many can cause overfitting, leading to poor performance on unseen data. On the other hand, a lower number of neurons may limit the expressive power of the FCNN, hindering its ability to capture complex relationships in the data. Proper initialization and regularization techniques (e.g., Dropout, L2, etc.) are often employed to prevent issues such as vanishing or exploding gradients, ensuring stable and effective learning within the hidden layers. The output layer of an FCNN receives the results of the consecutive non-linear data transformations achieved through the hidden layers and learns how to map these non-linear feature representation, or embedding, to the desired target [47]. These output neurons generate the final output of the FCNN, which could be a prediction, classification, or any other relevant outcome, based on the specific task the model is designed to perform.

One noteworthy limitation of FCNNs is their inability to effectively leverage the inherent data structures commonly present in forecasting applications [47]. Moreover, their fixed number of inputs and outputs makes them unsuitable for problems that involve varying input and output sizes, as typically encountered in forecasting tasks [47]. In order to overcome these limitations, specialized architectures have been developed [48] that incorporate FCNNs as fundamental building components in the form of Fully Connected (FC) layers. In the following discussion, we explore these advanced architectures that effectively address the challenges mentioned above.

2.3.2. Long-short term memory

Long Short-Term Memory (LSTM) [54] networks were introduced to address the vanishing gradient problem present in RNNs. This issue arises from the exponential decay of gradients over time, hindering learning over lengthy sequences [47]. The LSTM architecture enables the model to capture and preserve long-term dependencies, making them particularly effective for tasks that involve complex patterns across long periods of time. To that end, as depicted in Fig. 2, the internal structure of the LSTM introduces specialized memory cells and gating mechanisms that allow it to selectively remember or forget information at different time steps. These gating mechanisms consist of three main gates: the input gate, the forget gate, and the output gate.

The input gate controls the incorporation of new information into the memory cell. It applies a sigmoid activation function to the linear combination of the current input, denoted as X_t, and the previous hidden state, denoted as H_{t-1}. Mathematically, the input gate activation I_t is defined as:

$$I_t = \sigma(W_{xi} \cdot X_t + W_{hi} \cdot H_{t-1} + b_i) \tag{4}$$

where W_{xi}, W_{hi}, and b_i are the weight matrix and bias term associated with the input gate.

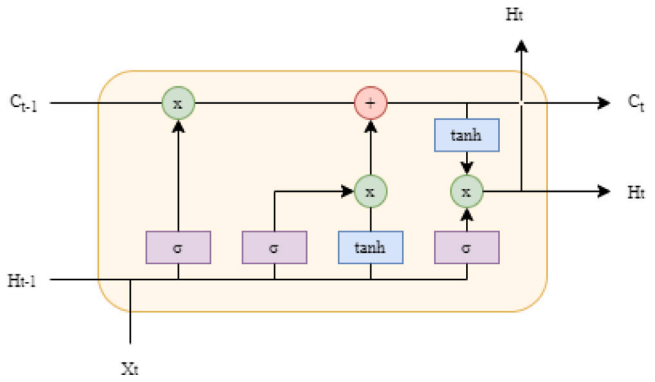


Fig. 2. Architecture of a long short-term memory cell, showing the interactions between the input X_t, the previous hidden state H_{t-1}, and the previous cell state C_{t-1}. σ denotes the sigmoid function, while *tanh* corresponds to the hyperbolic tangent function. The symbols $\times$ and $+$ represent element-wise multiplication and addition, respectively. H_t denotes the current hidden state, and C_t denotes the current cell state, both of which are updated based on the inputs and the outputs of the forget, input, and output gates. The updated values H_t and C_t are then passed to the next LSTM cell in the sequence.

The forget gate determines the extent to which the previous memory cell state, denoted as C_{t-1}, should be forgotten or preserved. The forget gate activation F_t is computed as:

$$F_t = \sigma(W_{xf} \cdot X_t + W_{hf} \cdot H_{t-1} + b_f) \tag{5}$$

where W_{xf}, W_{hf}, and b_f are the weight matrix and bias term associated with the forget gate.

The memory cell C_t stores the information over time and is updated using the input and forget gates. The updated memory cell state is computed as:

$$C_t = F_t \cdot C_{t-1} + I_t \cdot \tanh(W_{xc} \cdot X_t + W_{hc} \cdot H_{t-1} + b_c) \tag{6}$$

where W_{xc}, W_{hc}, and b_c are the weight matrix and bias term associated with the memory cell.

The output gate determines the extent to which the current memory cell state C_t should be exposed to the next hidden state H_t. The output gate activation O_t is computed as:

$$O_t = \sigma(W_{xo} \cdot X_t + W_{ho} \cdot H_{t-1} + b_o) \tag{7}$$

where W_{xo}, W_{ho}, and b_o are the weight matrix and bias term associated with the output gate.

The hidden state H_t is the output of the LSTM at time step t. Mathematically, the hidden state H_t is computed as:

$$H_t = O_t \cdot \tanh(C_t) \tag{8}$$

LSTMs effectively address the problem of gradient vanishing, preventing gradients from dissipating as they traverse the network [48,55]. This is achieved through the aforementioned gate-based mechanism, which allows the cell state to be preserved and updated as necessary. LSTMs are commonly trained using a variation of the classical backpropagation algorithm known as Backpropagation Through Time (BPTT). While similar to traditional backpropagation, BPTT accounts for the LSTM's ability to predict sequences comprising multiple time steps rather than a single value. To accomplish this, BPTT unfolds the LSTM network over time and computes the gradients of the cell state and gates at each time step. Consequently, the loss gradient with respect to the cell state at time t is backpropagated through time to the cell state at time t_{-1}. Notably, the network parameters are shared across all time steps. Once the gradients are computed, they propagate through the unrolled network [48].

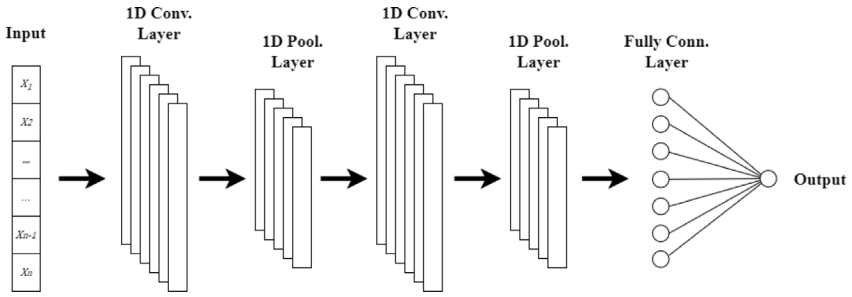


Fig. 3. CNN network architecture and layers.

2.3.3. Convolutional neural networks

Convolutional neural networks (CNNs) have emerged as a powerful and increasingly popular DL architecture for forecasting tasks. The inception of CNNs can be traced back to the year 1989 when [56] initially proposed this concept, drawing inspiration from Fukushima's Neocognitron model introduced in 1980 [57]. Since then, CNNs have been widely adopted for image classification, object detection, and other computer vision tasks. Although initially developed for image analysis, CNNs have proven their effectiveness in a variety of fields where data possess a known ordinal structure [47], such as time-series prediction.

CNNs consist of multiple layers that operate hierarchically to capture temporal patterns and make predictions [47]. As depicted in Fig. 3, these layers include convolutional layers, pooling layers, and FC layers, each serving a specific role in the network [47]. Convolutional layers are responsible for extracting relevant features from the input time-series data by applying convolution operations over sliding windows of the input sequence. This allows the network to automatically learn meaningful representations at different temporal scales. Pooling layers play a crucial role in downsampling the extracted features while retaining their essential characteristics and enhancing the network's ability to generalize and capture important temporal patterns. The output of the convolutional and pooling layers is then fed into FC layers, which integrate the learned representations and make predictions based on the extracted features.

One of the key advantages of CNNs for forecasting is their ability to automatically learn hierarchical representations from sequential data. By exploiting the hierarchical nature of convolutional layers, CNNs can initially capture local temporal correlations, progressively learning more abstract and high-level features in deeper layers. This hierarchical learning allows CNNs to extract relevant information across multiple levels of abstraction in the temporal dimension, thereby enhancing their predictive capabilities. Furthermore, the shared weights of CNNs enable the network to learn spatially invariant features, facilitating the application of the same learned feature across different regions of the input sequence. Additionally, the local connectivity inherent to CNNs ensures that each neuron is responsible for a specific receptive field, effectively capturing local dependencies within the input data. This characteristic makes CNNs well-suited for analyzing sequential information, as neighboring time steps are more likely to exhibit stronger relationships compared to those that are further apart.

The application of CNNs in time-series forecasting requires certain adaptations to accommodate the temporal nature of the data. Unlike 2D CNNs used on two-dimensional grids, 1D CNNs process one-dimensional sequences [48]. Convolutional and pooling operations in 1D CNNs are applied along the temporal axis, rather than across spatial dimensions as in 2D CNNs. Moreover, in contrast to conventional ML/DL, CNNs can efficiently handle multivariate data by treating each variable as a separate channel, with convolutions applied independently to each channel [58]. FC layers at the end of the CNN then extract dependencies between these channels, enabling comprehensive multivariate time-series analysis.

2.3.4. N-BEATS

The Neural Basis Expansion Analysis for Interpretable Time Series Forecasting (N-BEATS) architecture, introduced by Oreshkin et al. [40], presents a novel approach to time-series forecasting that has shown exceptional performance in renowned forecasting competitions and benchmarks. N-BEATS is a DL model explicitly designed for time-series forecasting, exhibiting impressive results particularly with univariate time-series data. One aspect that sets N-BEATS apart from other DL architectures is its ability to provide interpretability of the predictions. This is achieved by explicitly breaking down the forecasting task into interpretable components, including trend and seasonality. This explicit breakdown allows analysts and practitioners to gain insights and understanding into the underlying factors driving the forecasts. By offering this level of transparency and interpretability, N-BEATS surpasses traditional statistical techniques, making it a preferred choice in practical scenarios where interpretable forecasting is crucial [40].

The basic block of the N-BEATS architecture consists of two components: FCNNs and basis layers. These blocks are arranged vertically in an arbitrary number of stacks to form the complete architecture. In the first block, represented as x_l, the input X_l corresponds to a historical window of observations with a specific length l, ending with the most recent measurement. Originally, l is chosen as a multiple of the forecast horizon denoted as H. Each block takes an input, denoted as x_l, and produces two outputs: a forward estimate $\hat{y}_l$ (forecast) and a backward estimate $\hat{x}_l$ (backcast) of x_l. $\hat{y}_l$ represents the block's predicted values for the future time steps, while the $\hat{x}_l$ represents the best approximation of the input x_l based on the constraints imposed on the block's functional space. In subsequent blocks, the input x_l is obtained as the residual output of the previous block, specifically the backcast generated by the preceding block.

The aggregation of forecasts in the N-BEATS architecture follows a hierarchical doubly residual stacking principle based on the topology originally proposed by other well-known architectures such as DenseNet [59]. This principle involves two residual branches: one for the backcast estimate and another for the forecast branch within each block. The backcast branch allows for sequential analysis of the input signal, progressively removing approximated components at each block to assist downstream blocks in their forecasting tasks. This structural design facilitates gradient backpropagation and ensures smooth information flow throughout the network. The final forecast is obtained by summing all the partial forecasts generated by the blocks.

Each block is composed of multiple FC layers, with each layer being followed by a Rectified Linear Unit (ReLU) activation function. These stacked FC layers are responsible for predicting the forward expansion coefficients (θ_l^f) and backward expansion coefficients (θ_l^b). Subsequently, the basis layers, denoted as g_l^b and g_l^f, perform a linear transformation based on an element-wise product between the expansion coefficients of each branch and the corresponding weights of the basis layers to generate the backcast and forecast outputs of the block. A complete diagram of the architecture and its different components is presented in Fig. 4.

The flexibility of the N-BEATS architecture allows for the choice of learnable basis layers or predefined functional forms, depending on the

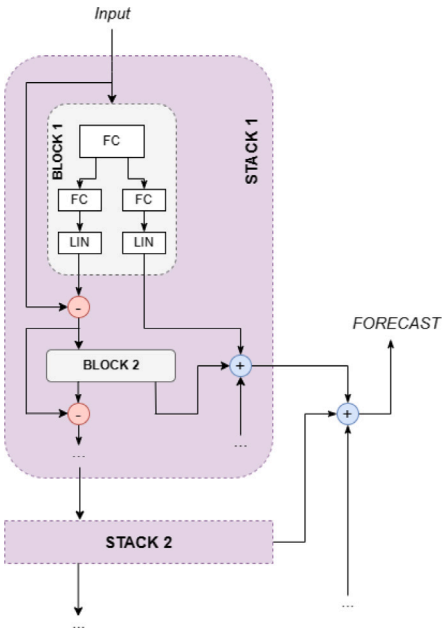


Fig. 4. Diagram of the N-BEATS architecture showing the internal structure of a stack, including the residual connections between blocks and the layers within a block.

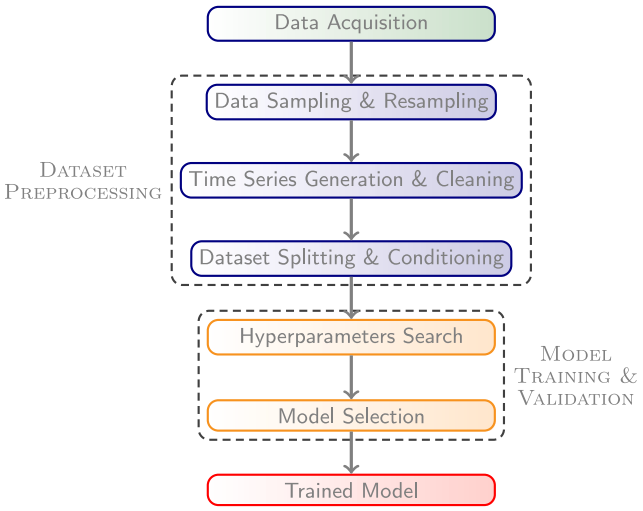


Fig. 5. Summary of the data acquisition, preprocessing and deep learning models training and validation workflow.

specific problem and desired inductive biases for the task. By imposing constraints on the weights of the basis layers, the architecture can be adjusted to reflect problem-specific constraints in the generated outputs and enable subsequent analysis of the predictions by decomposing them into trend and seasonality components, as well as other relevant factors, and consequently allowing for interpretation of the decisions made by the model. However, a generic implementation of N-BEATS commonly employs learnable basis layers, as they offer greater flexibility and adaptability to various forecasting tasks.

The N-BEATS architecture possesses numerous advantageous characteristics. Firstly, it can be directly applied to diverse problem types without requiring extensive feature engineering, thus saving considerable time and effort. Moreover, N-BEATS outperforms other DL architectures such as LSTM in terms of training efficiency and scalability, which is especially advantageous when handling extensive datasets. Additionally, the interpretability aspect of N-BEATS is of significant value, allowing users to gain insights into the model's decision-making process. Furthermore, studies have demonstrated the superior generalization capabilities of the N-BEATS architecture when trained on a specific source time-series dataset and applied to a distinct target dataset [40,42–44,47]. These findings underscore the robustness and adaptability of the N-BEATS model, further reinforcing its suitability for time-series forecasting tasks.

While originally conceived as a univariate model, a novel redesign of N-BEATS that empowers it to handle multivariate series was previously proposed by [60]. This redesign allows for the consideration of variables based on their relationship with independent variables, whether they are endogenous or exogenous. To that end, a concatenation layer that merges multiple inputs into a one-dimensional input layer is introduced. By adopting this approach, the model is optimized solely to capture the dynamics of endogenous variables, with exogenous variables serving as support for model learning rather than being the forecasting target. This design choice ensures that the model's capabilities are entirely dedicated to modeling the endogenous dynamics, without being encumbered by the entire multivariate context.

3. Multi-horizon prediction framework setup

The objective of this study to develop the first multi-horizon forecasting model for Chl-*a* levels in a freshwater body, with a granularity of one hour and a forecast horizon of one week into the future. For the first time, the N-BEATS architecture is introduced for Chl-*a* forecasting and its performance is evaluated against traditional DL architectures such as CNN and LSTM. Previous research has demonstrated the superior performance of N-BEATS in multi-horizon forecasting across various domains [60]. Furthermore, this study explores an innovative application that utilizes a multi-horizon Chl-*a* forecasting model to identify algae proliferation alert levels, thereby transforming the problem into a multi-horizon classification task.

In this section, we will outline the steps taken from data acquisition to the training and evaluation of multi-horizon forecasting models for Chl-*a* values in a freshwater body. Section 3.1 will provide details on the environment and sensors used to collect data from the freshwater body over a period of three years. Section 3.2 will describe the steps followed to process the captured sensor data. Section 3.3 will explain how the models were trained, selected, and evaluated. Finally, Section 3.4 will specify the technology and dataset used for conducting the experiments performed in this study. A summary of the workflow of this process can be appreciated in Fig. 5.

3.1. Data acquisition

Located in the Miño-Sil River Basin District of Galicia, NW Spain, the As Conchas reservoir forms an integral part of the "Baixa Limia-Serra do Xurés" Natural Park. This eutrophic freshwater body can store up to 80 Hm^3 of water, displaying pronounced depth variations, from shallow edges to a central depth of 32 m at full capacity. According to the Miño-Sil Hydrographic Confederation, the depth of the photic zone of the reservoir measured with the Secchi disk (indicative of water clarity) ranges between 2 and 6 m from May to October over a three-year observation period. Two centrally anchored EM1250 buoys from Xylem Analytics, designated as the beach buoy and the dam buoy, facilitate reservoir monitoring. Equipped with YSI multiparametric EXO3 probes, these buoys housed sensors that measured Chl-*a* using a total algae fluorescence sensor, pH, water temperature, and EC, as detailed by Morón-López et al. [19]. These probes and sensors were subjected to regular maintenance and calibration according to manufacturer's instructions, ensuring reliable data collection. Solar panels provided the necessary power for this autonomous system, with battery levels serving as an indicator for daylight hours. The probes, positioned at a standard depth of approximately 1 m, were set to log data every 15 min.

Table 1
Statistical overview of the dataset characteristics for the beach buoy, which comprises approximately 28,000 samples.

Statistic	Temperature (°C)	pH	Conductivity (μS/cm)	Chlorophyll (μg/L)
Average	15.10	7.74	66.85	8.86
Standard deviation	6.07	0.87	15.66	7.67
Minimal value	5.38	6.56	1.03	0.00
Percentile 25%	9.38	7.11	55.34	3.98
Percentile 50%	14.36	7.44	70.74	6.56
Percentile 75%	21.03	8.11	77.22	11.22
Maximal value	27.75	10.26	99.20	78.65

Table 2
Statistical overview of the dataset characteristics for the dam buoy, which comprises approximately 27,000 samples.

Statistic	Temperature (°C)	pH	Conductivity (μS/cm)	Chlorophyll (μg/L)
Average	15.11	7.43	67.69	9.09
Standard deviation	5.94	0.94	9.04	8.30
Minimal value	6.15	4.93	4.35	0.00
Percentile 25%	9.41	6.88	62.05	3.59
Percentile 50%	14.15	7.16	67.82	7.07
Percentile 75%	21.10	7.63	73.06	11.95
Maximal value	28.48	10.30	97.57	71.66

3.2. Data preprocessing

DL models require appropriate data formatting to ensure effective training for the specific problem at hand. For the multi-horizon, it was crucial to adapt the data to a time series format and split it into separate training, testing, and validation datasets. Additionally, DL models can benefit from outlier data cleaning and conditioning procedures. In this study, the dataset used was collected remotely from two sensor-equipped buoys. Due to occasional environmental effects, there were instances of data loss in the sensor-to-server communication, and occasional erroneous measurements were recorded by the sensors. As a result, prior to training the DL models for Chl-*a* level forecasting, it was necessary to perform data cleaning and correction processes.

The data preprocessing process employed in this study consists of the following steps: (i) feature selection for the training process; (ii) resampling the data from 15 min to 1 h intervals; (iii) constructing the time series; (iv) cleaning and removing invalid data; (v) splitting the data into training, testing, and validation sets; (vi) normalizing the data and removing seasonality. In what follows, we describe in detail each of preprocessing steps.

Feature selection. The Chl-*a* variable acts as the target forecast variable, and following standard forecasting systems practice, its historical observed values are utilized as inputs for the DL models. In this context, Chl-*a* is regarded as the endogenous variable. Furthermore, exogenous variables were incorporated to enhance the quality of the trained models. More specifically, temperature, pH, and conductivity were selected as exogenous variables due to their high correlation with Chl-*a*, as reported in the study by Mozo et al. [25] using the same dataset. These four features were employed in conjunction with the Chl-*a* endogenous variable to improve the models obtained solely using the endogenous variable. However, when training solely with exogenous variables, there was no observed improvement in the results compared to the other variable combinations. Hence, in our experiments, we considered Chl-*a* alone or Chl-*a* with exogenous variables (often referenced as Chl-*a*+Exo). For a more detailed analysis of the statistical nature of the endogenous and exogenous variables, please refer to Tables 1 and 2, corresponding to the beach and dam datasets, respectively.

Data resampling. The buoys measured and transmitted data to a server every 15 min. These data exhibited low variation between sequential samples due to minimal changes in the water masses within a 15-min period. Therefore, it was decided to resample the dataset at a coarser granularity, to hourly intervals, using the statistical function of median. Resampling the data helps expediting the training of deep learning models without sacrificing important information. Additionally, the median function serves to mitigate the number of outliers that may arise due to measurement errors over an hour-long period.

Table 3
Total count of data rows associated with each dataset segment (training, weight validation, hyperparameter validation, and testing) following the data partitioning process for both beach and dam buoys. The final column specifies the cumulative total of rows for each respective buoy.

Dataset	Training	Weights validation	Hyperparameter validation	Testing	Total
Beach buoy	15 640	1912	1646	8398	27 596
Dam buoy	14 989	1912	1646	8397	26 944

Generation of time series. Diverse window sizes have been selected according to the number of time steps used to attain the forecasts (input window size) and the future time steps that the models will forecast (forecast horizon size). Input window sizes were studied in the set {12, 24, 72, 168, 336} whereas forecast horizon sizes were studied in the set {24, 168}. The alert level 1 involves increased monitoring at least once a week, then the one-day or one-week ahead forecasting would aid the planning of the water-body monitoring. It is important to note that the maximum input window size was set to be at least twice the maximum forecasting horizon size, and the minimum window size was set to be at least half of the minimum forecasting horizon size. Additionally, intermediate window sizes were defined to assess the impact of window size on the deep learning models.

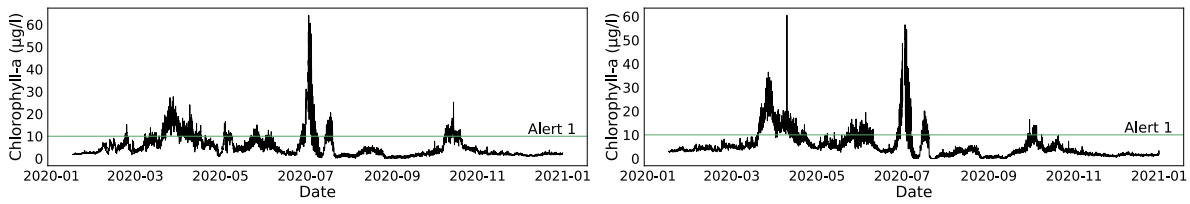
Time series cleaning. To ensure the models are not affected by noise and can effectively learn, all rows in the time window and forecasting horizon that contained any *invalid* data points were excluded. These *invalid* data points can arise from occasional sensor errors or missing data due to communication interference. In this particular case, approximately 4% of the data was identified as *invalid* and subsequently discarded.

Dataset splitting. To mitigate overfitting during model training, the original dataset was split into four distinct subsets, while maintaining the sequential order of the data for each buoy. These subsets were designated as follows: training, weights validation, hyperparameter validation, and testing. To ensure models' generalizability, the data from the last year was reserved exclusively for the testing dataset. Prior to that, 10% of the data before the last year was allocated for weights validation, and another 10% was set aside for hyperparameter validation. The remaining data comprised the training dataset. Fig. 6 presents a detailed depiction of the temporal evolution of Chl-*a* concentrations within the testing partition, encompassing data from both the beach and dam buoy locations.

The model was initially trained using the training dataset, and the loss function optimization was monitored using the validation dataset. Subsequently, the model was evaluated using the hyperparameter validation dataset to determine the optimal set of hyperparameters. Table 3 provides a breakdown of the number of data rows in each of the resulting datasets obtained from the splitting process.

Data conditioning. The data was preprocessed using standard scaling and seasonality removal techniques. Standard scaling is a method that normalizes the data's mean to zero and standard deviation to one. Seasonality removal involves eliminating patterns that periodically repeat to mitigate their influence. Preconditioning the dataset in this manner facilitates the training of DL models and prioritizes the learning of more relevant patterns.

In this study, a total of eight preprocessed datasets were generated from two distinct buoy sources: the beach buoy and the dam buoy. These datasets were meticulously prepared to facilitate the experiments



(a) Dam dataset (b) Beach dataset

Fig. 6. Temporal progression of Chl-a concentrations within the testing dataset, incorporating data from both the beach and dam buoy sites. The horizontal axis denotes time in hours, while the vertical axis represents Chl-a concentration in µg/l.

conducted as part of this research. For each buoy, four datasets were processed, comprising training, validation, hyperparameter validation, and testing datasets. From these eight datasets, a careful selection was made, incorporating all four datasets from the beach buoy and exclusively the testing dataset from the dam buoy.

The primary objective of this study was to showcase the predictive capacity of modern DL models in forecasting chlorophyll levels up to one week ahead in a body of water, while also evaluating their generalization capabilities. Consequently, the models were trained and rigorously evaluated using data exclusively from the beach buoy, followed by thorough testing on the dam buoy. This approach was adopted to demonstrate the models' remarkable ability to generalize effectively across different locations within the body of water.

3.3. Deep learning models training and evaluation

In supervised problems, DL models are typically trained iteratively by optimizing their weights using observed labels from a given training dataset, while simultaneously monitoring convergence through a separate validation dataset. Prior to training, it is crucial to define a set of hyperparameters. However, since there is no universally optimal set of hyperparameters that applies to all problems, it is necessary to explore different combinations and identify the most suitable ones for each specific model and scenario. To address this challenge, a random search heuristic is employed to generate diverse sets of hyperparameter values randomly sampled from predefined hyperparameter grids for each model or scenario. Each of these models is then trained on the training dataset and evaluated on a separated validation dataset. Ultimately, the best-performing model is selected based on its performance on the validation dataset.

The random search phase explored various hyperparameters, in selected ranges, and the models they were applied to. Table 4 presents the detailed hyperparameter configurations. The models investigated in this study include N-BEATS, 1D-CNN (abbreviated as CNN), LSTM, and CNN-LSTM. The 1D-CNN or CNN model is a variant of convolutional neural networks specifically designed to operate on one-dimensional data, in contrast to the standard 2D CNNs. The CNN-LSTM model is a hybrid architecture that commences with a multi-layer 1D-CNN, and its output is subsequently fed into a multi-layer LSTM model. An additional hyperparameter, Skipping Connections, is considered in the CNN-LSTM model, which determines whether the output of the initial 1D-CNN network is concatenated with the input of the subsequent 1D-CNN network. The other hyperparameters listed in Table 4 are specific to each architecture, while the hyperparameters under the "Common" row are shared across all architectures. To ensure comprehensive coverage of the parameter space, the sets of hyperparameters were randomly sampled from a uniform distribution.

The hyperparameters explored in this random search phase, along with their ranges, can be observed in Table 4. The architectures explored include N-BEATS, 1D-CNN (abbreviated as CNN), LSTM, and CNN-LSTM (see Section 2 for further details about the architectures). The CNN-LSTM model is a hybrid model that starts with a multi-layer 1D-CNN, and its output serves as the input for a multi-layer LSTM model. In this CNN-LSTM hybrid model, an additional hyperparameter called *skipping connections* is considered. It controls whether the output of the first 1D-CNN network is concatenated with the input of the subsequent 1D-CNN network. The other hyperparameters in Table 4 are standard for each architecture, while the hyperparameters in the "Common" row are shared among all architectures. All sets of hyperparameters were sampled following a uniform distribution to cover the breadth of the ranges more effectively.

In these experiments, various combinations of the architectures listed above, input datasets, input window sizes, and forecast horizon sizes were evaluated. Two types of input datasets were used: Chl-*a* only and Chl-*a* with exogenous variables. Five different input window sizes were explored: 12, 24, 72, 168, and 336 h. Two forecast horizon sizes were examined: 24 and 168 h. For each model and scenario, up to 30 sets of random hyperparameters were tested, resulting in a total of approximately 2400 trained models. Table 5 provides a summary of the hyperparameter configurations explored in this study. Subsequently, the best-performing set of hyperparameters, based on the performance on the hyperparameter validation dataset, was selected for each model, resulting in 80 scenarios for comparison on the test dataset.

Multi-horizon forecasting models are commonly assessed using metrics for measuring the accuracy of regression models, such as Mean Absolute Error (MAE) or Mean Squared Error (MSE), which provide a measure of the average deviation between the model's predictions and the expected values. While MAE considers the absolute difference, MSE additionally penalizes larger errors more severely, making it particularly suitable for problems where significant deviations have higher importance. In this study, both metrics were employed to evaluate the accuracy of the regression models. Additionally, the MSE was utilized as the loss function during the training phase to optimize the model weights, thereby placing a greater penalty on larger errors.

$$\begin{aligned} \text{MAE} &= \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \\ \text{MSE} &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \end{aligned} \tag{9}$$

Furthermore, this study examines the models' capability to detect the HABs alert levels in several future horizons, which can be regarded as a multi-horizon classification problem. WHO considers different HABs vigilance and alert levels based on the intended use of water. Likewise, the implementation of these levels may vary according to national and local regulations. To simplify the various alert scenarios and facilitate comparisons between the different models, this study opted for a Chl-*a* alert level of 10 µg/L. This particular threshold falls within the current range of the vigilance level for recreational waters and alert level 1 for drinking water [28]. Additionally, it aligns with the guidance level 1 outlined in the previous WHO 1999 edition [61]. It is important to highlight that the tool developed for this study can be fine-tuned based on specific thresholds as needed.

Evaluation of this classification problem commonly employs the F1-score metric, which represents the harmonic mean of precision and recall. The F1-score metric assesses whether the model's predictions

Table 4
Hyperparameters for each DL architecture used in the study, along with their respective data types and the range of values considered. The hybrid CNN-LSTM architecture incorporates hyperparameters from both the CNN and LSTM architectures. Hyperparameters represented with square brackets indicate that they were sampled from a uniform distribution for Integer types and a logarithmic uniform distribution for Real types. Values from ranges without specified intervals were uniformly selected.

Architecture	Parameter	Type	Range
Common	Dropout	Real	[0, 0.7]
	Batch normalization	Integer	[0, 1]
	L2 regularization	Real	[0, 1]
LSTM	Number cells	Integer	[1, 200]
	Number layers	Integer	[0, 5]
CNN	Number filters	Integer	8, 16, 31, 64, 128
	Kernel size	Integer	[2, 12]
	Max pooling	Integer	[1, 2]
	Number layers	Integer	[0, 8]
CNN-LSTM	Skipping connections	Boolean	True, False
N-BEATS	Number stacks	Integer	[1, 10]
	Number of blocks per stack	Integer	[1, 10]
	Theta's dimension	Integer	[1, 10]
	Non-linear FC layers units	Integer	[32, 1024]
	Non-linear FC layers per block	Integer	[1, 10]
	Shared weights in stack	Boolean	True, False

correctly identify the HAB alert level as compared to the original sensor. In this study, the F1-score metric was employed to select the best models for both the Random Search with hyperparameter validation and the test dataset. This choice stems from the practical significance of predicting future HAB alert in the environmental domain. However, metrics such as MAE and MSE were also utilized to enable experts to compare the performance of models on the test dataset.

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}} \\ \text{Recall} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \\ F1\text{-score} &= \frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \end{aligned} \tag{10}$$

$\text{TP} = \text{True Positives}, \text{FP} = \text{False Positives}, \text{FN} = \text{False Negatives}$

3.4. Experimental environment

Deep learning models were trained and tested on an off-the-shelf computer with the following configurations: an Intel(R) Xeon(R) E5-2630 v2 @ 2.60 GHz CPU, 64 GB of RAM, and a GeForce GTX 1080 Ti GPU. The software libraries used along with their respective versions were as follows: Python 3.9.6, TensorFlow 2.4.1, Keras 2.4.0, Optuna 2.10.0, Pandas 1.3.1, and NumPy 1.18.5.

Please note that all the software utilized in this study is freely available to the public. The data used to conduct the experiments in this work can be accessed from the following repository: https://github.com/stanislavvakaruk/Chlorophyll_soft-sensor_machine_learning_models

4. Results and discussion

In this study, various DL models have been trained to predict Chl-*a* levels over multiple time horizons and identify the preliminary alarm threshold consistent with WHO guidelines for HAB control in freshwater bodies [28]. Specifically, these models were designed to predict at both daily and weekly intervals, delivering forecasts for 24 h in the daily scenario and 168 h for the weekly scenario (hourly data granularity). Throughout the training, various past window lengths were employed for each model type, namely N-BEATS, LSTM, CNN, and CNN+LSTM. These lengths spanned 12, 24, 72, 168, and 336 h. The endogenous variable, i.e. Chl-*a*, was used both in isolation and alongside exogenous variables of temperature, pH, and conductivity. For every model category, we explored up to 30 unique hyperparameter sets, resulting in a total of over 2000 trained models. An in-depth overview of the training methodology and model configurations can be referenced in the preceding Section 3.

This section synthesizes the results from the evaluation of the models once trained and validated, and examines the practical implications of the developed prediction system. The models were trained using data collected from the beach buoy over a period of two years, and subsequently tested against data from the following year of both the beach and dam buoys. This strategy was formulated to ascertain the generalizability of the models over different sampling points within the same water body. Notably, the N-BEATS model outperformed all the other models evaluated in this study, showing superior accuracy in forecasting Chl-*a* levels and accurately identifying the critical WHO alert threshold for Chl-*a* concentrations.

The following subsections offer an in-depth analysis of model results, categorized based on their predictive time intervals. Focusing initially on the short-term models trained for 24-h predictions, a majority of models consistently produced highly accurate predictions. In the following section, we compare the initial 24-h predictions of the long-range models, which provide forecasts for 168 h (or a week), with those of their daily counterparts. This comparison highlights a slight decrease in the predictive accuracy of the long-term models. In the category of long-range forecasts, the N-BEATS model stands out as particularly noteworthy, achieving the lowest Mean Absolute Error (MAE) and approaching F1-scores of 70%, even on the third day of prediction. Notably, when these models were fine-tuned using beach buoy data and then applied to dam buoy data, they retained their effectiveness with only a modest 10% decrease in their ability to identify the initial alert threshold at this distinct location. In the final section, we reflect on the practical benefits and extensive versatility of the predictive approach introduced in this study.

4.1. One-day ahead forecast

The selected DL models in this subsection have been trained with the primary aim of predicting Chl-*a* values a day ahead at an hourly granularity. These models have been primarily assessed based on their ability to predict the Chl-*a* value (evaluated using the MAE metric) as well as their capability to identify the first level of HABs according to WHO standards (evaluated using the F1-score metric) in the last hour of the prediction (t + 24 h into the future). It should be noted that the models generated up to 24 values, each corresponding to an hour from time t + 1 up to t + 24 in the future. Initially, the models were filtered based on the validation dataset to identify first level alert 24 h into the future. For each model type, window size, and input features set, the optimal hyperparameters were selected. The results of these optimal hyperparameters, tested on both the beach and dam buoy testing datasets, are presented in this section.

The testing phase results for the best models, which were selected based on the highest mean F1-score obtained on the validation dataset, are shown in Fig. 7. The results presented in this figure indicate the mean F1-score and mean MAE of the 24-h forecast (mean result of 24 values) for each buoy test dataset. As observed, three of four models yielded high results, with a negligible difference in F1-score values (less than 1%). Specifically, the N-BEATS (83.5% mean F1-score), CNN (83.7% mean F1-score), and CNN-LSTM (84% mean F1-score) models achieved the best results for the beach buoy. Therefore, these models were capable of forecasting and identifying Chl-*a* values at the alert threshold in a more effective manner than the LSTM model (81% mean F1-score). The decrease of approximately 5% in mean F1-score from beach buoy to dam buoy suggests a strong generalization capability of all the models. These three models were close in their predictions even in the intermediate values, as it can be seen in Fig. 8. However, N-BEATS obtained the lowest values, and therefore the best results according to the mean MAE metric (1.069 in beach buoy and 1.100 in

Table 5
Summary of the DL architecture, input, and output variables configurations explored in this study. For each of these combinations, up to 20 random hyperparameter configurations were evaluated. Ultimately, up to 2000 deep learning models were trained.

Parameter	Range
Deep learning architecture	N-BEATS, LSTM, CNN, CNN-LSTM
Input variables set	Chl-*a*, Chl-*a*+Exo (Chlorophyll-a + Exogenous variables)
Input window size	12, 24, 72, 168, 336
Forecast window size	24, 168

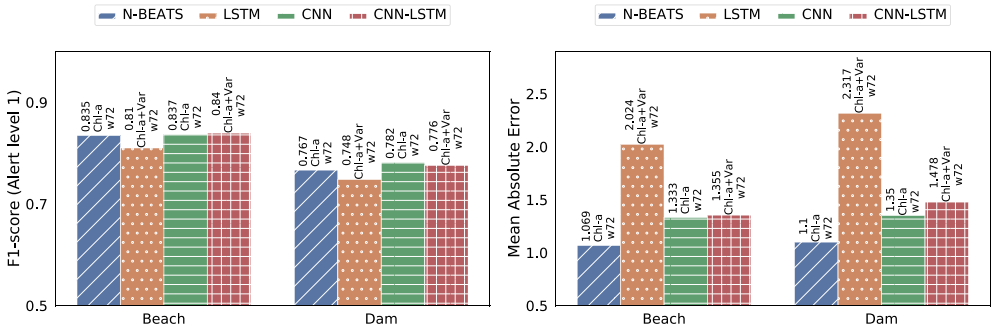
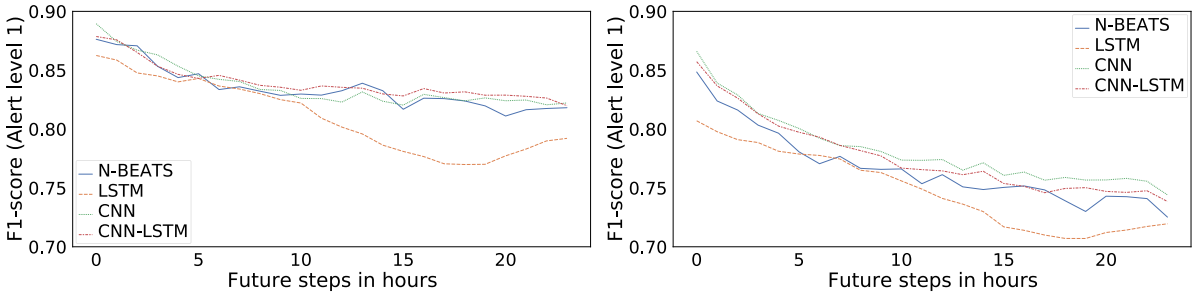


Fig. 7. Mean F1 and MAE results for the best 1-day forecast model configurations. Both buoys used as sources of data are illustrated on the *x*-axis, whereas the mean F1-score and mean MAE over 24 time steps are shown in the *y*-axis on the left and right figure, respectively. Each of the bars represents one of the models selected for comparison. In addition, the exact value represented on the *y*-axis, the window and the variables used for the model are indicated on top of each bar. For the F1-score metric, higher values are better. On the contrary, lower MAE values signify higher accuracy. The top-performing models for the mean F1-score and mean MAE over 24 time steps, averaged between the two buoys, are CNN and N-BEATS, respectively.



(a) Beach dataset (b) Dam dataset

Fig. 8. F1 score for the best 1-day forecast model configurations on each time step. Steps into the future and F1 are shown in *y*-axis and *x*-axis accordingly. The best configuration per model is selected and assigned a different color line. Higher value is directly correlated to better accuracy. Beach buoy results are on the left and dam buoy on the right. The best model on the last hour is CNN.

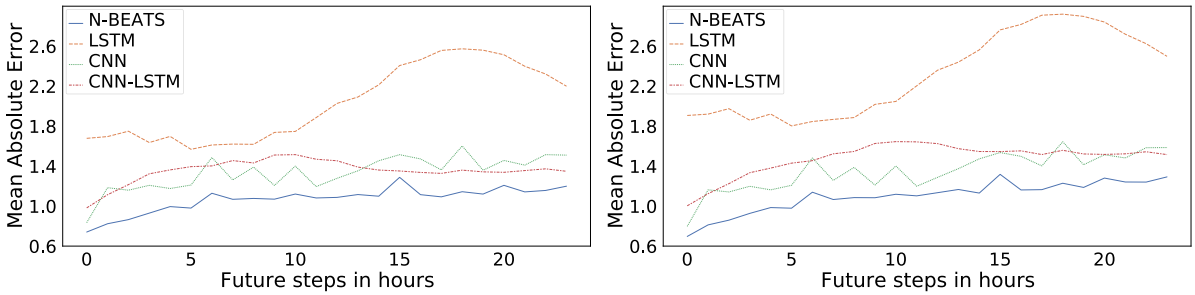
dam buoy), which implies its better capability to infer the exact value of Chl-*a*. The mean MAE of N-BEATS was 24% better than that obtained by the second best CNN model (1.333 MAE). This result became clear in Fig. 9, where N-BEATS' line is clearly below the other lines even in the intermediate values.

In terms of window size, all the best models used a window size of 72 h, which points to better performance at this value. This results could be explained as the need for 3 days of past data to correctly predict and identify abnormal Chl-*a* levels. With larger time windows, the results start to deteriorate, which implies that the models need to be tuned taking into account the amount of historical information. Interestingly, the N-BEATS and CNN models obtained their best results using only the endogenous variable (i.e. Chl-*a*) as input variable , unlike the CNN-LSTM and LSTM models, which performed better considering Chl-*a* and exogenous variables. The LSTM architecture, tailored for sequential data, excels in discerning and assimilating the trend and periodicity inherent in the Chl-*a* level time series. As a result, the nuances of trends and periodicity become increasingly salient, overshadowing the contributions of exogenous variables in extended forecasting scenarios. Conversely, in short-term forecasting, the input from exogenous variables gains prominence, providing crucial information for achieving high accuracy in LSTM-based models predictions with respect to their endogenous-only versions. In any case, these exogenous variables did not seem to significantly differentiate these models from the others.

These results highlighted the superior performance of N-BEATS over conventional DL models in predicting day-ahead Chl-*a* values. N-BEATS not only achieved the minimal mean MAE but also showcased an F1-score with a difference of less than 1% from the top-tier model, CNN-LSTM. Thus, this study suggests that the N-BEATS model could be the ideal model for the prediction of short-term Chl-*a* and identification of alert levels based on WHO guidance.

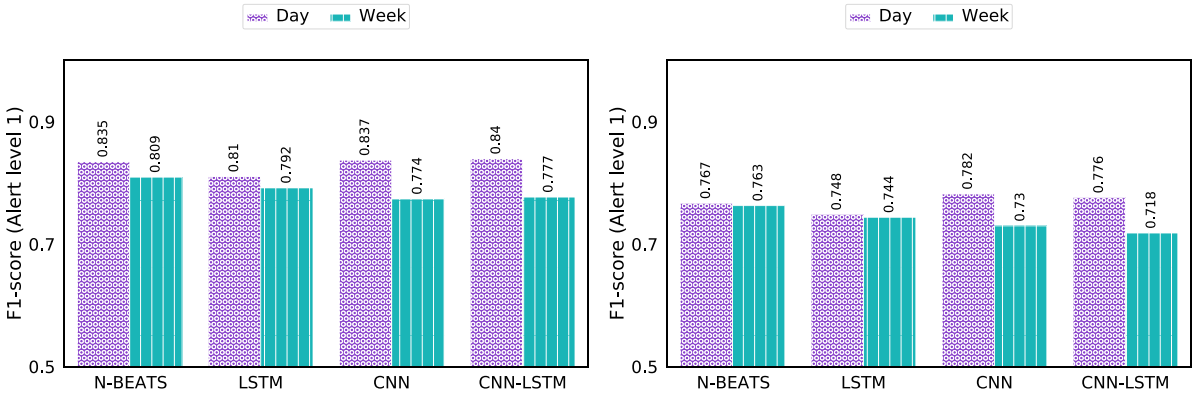
4.2. One-week ahead forecast

One-week ahead forecasting is focused on predicting 168 timesteps into the future. As previously mentioned, each of these time steps corresponds to one hour. Hence, the difficulty of this task is greater than for one day forecasting and higher error is expected the further into the future it is predicted. Longer-term forecasting allows us to have a better idea of how Chl-*a* levels will evolve, evaluate possible next steps and take earlier action. The experiments conducted for this subsection were analogous to the day-ahead prediction experiments. The best model of each DL architecture was selected based on the



(a) Beach dataset (b) Dam dataset

Fig. 9. MAE for the best 1-day forecast model configurations on each time step. The steps into the future are represented in the x-axis, and MAE is shown in the y-axis. Each of the best configurations per model are given a color. Lower value suggests higher accuracy. Graphs correspond to beach and dam from left to right. N-BEATS obtains the lowest MAE on the 24th step.



(a) Beach dataset (b) Dam dataset

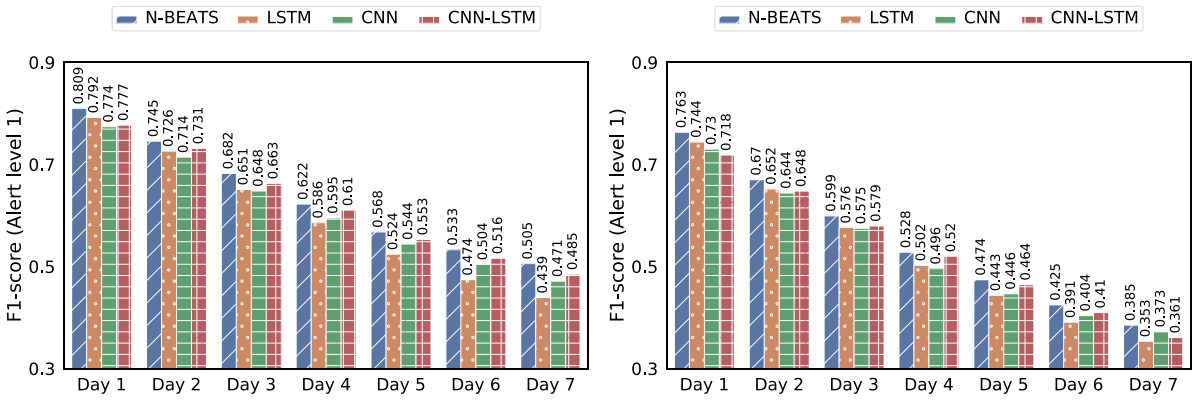
Fig. 10. Comparison of the 1st day forecast for one-day and one-week ahead forecasting. The x-axis indicates the model to which each of the results refers and the mean F1 over the first 24 time steps is shown at the top of each bar. The purple bars illustrate one-day ahead forecasting mean, while turquoise represents one-week ahead forecasting results. 10(a) shows the results for beach buoy while 10(b) shows the results for dam buoys. A higher F1 means a higher model exactness. CNN demonstrates superior performance for both one-day and one-week ahead forecasting across both beach and dam buoy. However, N-BEATS exhibits the best performance for one-week ahead forecasting in both scenarios.

F1-score results obtained from the validation dataset and were subsequently evaluated using the testing datasets from both the beach and dam buoys (last year). Then, the best model configuration was chosen for each DL model architecture, window size and input features set. Additionally, the one-week ahead forecasting models were compared with the one-day ahead prediction models discussed in the previous subsection.

The first analysis in this section shows the flexibility of employing a long-term multi-horizon forecasting model instead of a short-term forecasting model. In a multi-horizon model, we can select the forecast horizon, allowing us to seamlessly use a forecasting model designed for one-week ahead predictions for one-day ahead forecasting. In this scenario, we only select the 24th prediction step of the long-term forecasting model from its 168 prediction steps. In order to compare the long-term to the short-term forecasting model, we assess the 24th step prediction of the long-term model alongside the last prediction step (24th) of the short-term forecasting model, showcased in the previous Section 4.1. In Fig. 10, the results for the best hyperparameter configuration of each architecture on the 24th step in one-week forecasting are shown in turquoise color, while the results for one-day ahead forecasting are shown in purple color. In general terms, N-BEATS obtains the best results for the first day in one-week ahead forecasting on both buoys with a F1-score of 80.9% on beach buoys and 76.3% on dam buoy, respectively. Although the short-term version of N-BEATS differs less than 1% from the CNN and CNN+LSTM models, in the long-term version it is more than 3% higher than them in both beach and dam buoys. The sequence based architectures such as N-BEATS and LSTM outperform the other models as they exploit the temporal data structure in multi-horizon forecasting problems in a more effective way. In general, the use of long-term multi-horizon forecasting models instead of short-term forecasting models results in a 3% decrease, which can be considered a negligible cost for the flexibility of training only one long-term forecasting model instead of potentially hundreds or thousands of short-term forecasting models.

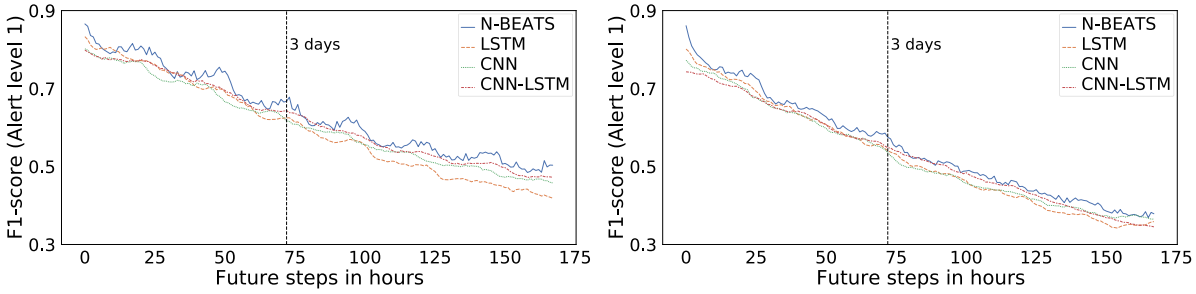
To evaluate the forecasting reliability over a one-week ahead forecast horizon, we analyzed the performance metrics for each day's predictions. As depicted in Fig. 11, the mean F1-score for the best configurations across different models is presented as the mean value of each day. Notably, N-BEATS stands our for its consistently superior results, with an F1-score approaching 70% by the close of day 3. This dominance is further emphasized in Fig. 12, where N-BEATS predominantly surpasses the observed performance of the other models. This linear plot, however, highlights intermittent decreases in the model's performance. In particular, the Chl-a level showed less consistency during the day, resulting in less accurate predictions. There are fluctuations in the level of Chl-a between the day and night time due to the vertical migration of algae along the water column in search of sunlight. Even so, the N-BEATS model continued to show its robustness by exhibiting lower MAE values compared to the other models (Fig. 13). This advantage is further evidenced in Fig. 14, where the trajectory of N-BEATS (indicated by a blue line) persistently ranks below the competing models throughout the observed duration.

As depicted in the figure presented above, N-BEATS consistently outperforms the other models. By the end of day 2, particularly on the dam buoy, N-BEATS approaches an F1-score of approximately 70%. As elaborated in Section 4.1, a typical decline in generalization



(a) Beach dataset (b) Dam dataset

Fig. 11. Daily mean of the F1-score for the best one-week forecasting model configurations. Days in the forecasted week are shows in the *x*-axis, while the mean F1 per day is shown in *y*-axis. Each color bar is assigned to the best configuration for each model. Beach buoy is represented in (a) and dam buoy is represented in (b). Higher values indicate better alarm 1 classification. N-BEATS obtains the best overall results.



(a) Beach dataset (b) Dam dataset

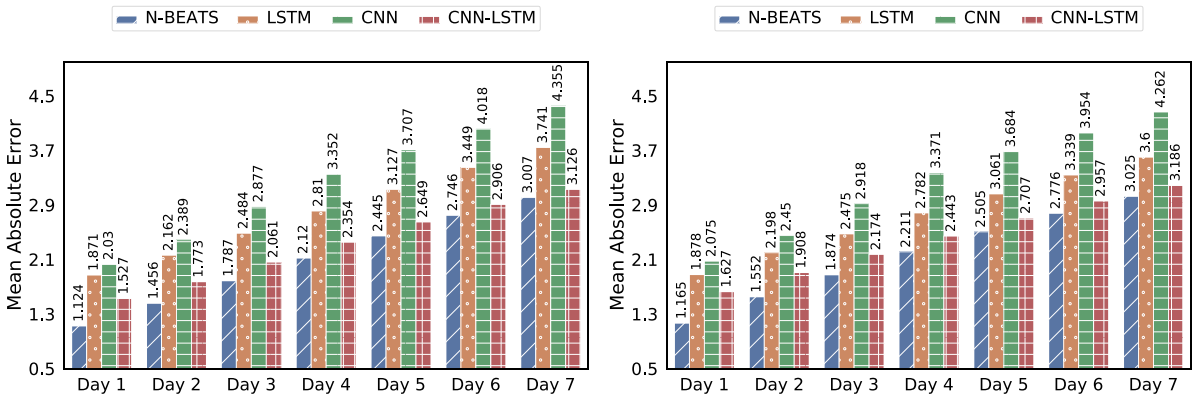
Fig. 12. F1 for the best one-week ahead forecasting model configurations. Time steps are indicated in the *x*-axis and F1 is shown in *y*-axis. The best model configurations are assigned a unique color line. Higher F1 score equates to better model exactness. Beach buoy results are shown in (a) and dam buoy results on the (b). The best model in the majority of horizons including the last 50 h is N-BEATS.

capability, roughly around 10%, is observed on the second buoy, even over extended periods. However, the results in Fig. 12 reinforces the supremacy of N-BEATS, consistently placing it above its counterparts for the dam buoy. A similar trend is observed when evaluating the MAE as shown in Figs. 13 and 14. Notably, Fig. 14 reveals a mere 5% difference in MAE between the beach and dam buoy for N-BEATS, signifying an even more formidable generalization capability than what is inferred from the F1-score. Broadening the lens to a long-term view, N-BEATS' prowess in generalization becomes undeniably apparent. With reductions of approximately 10% and 5% in F1-score and MAE, respectively, N-BEATS demonstrates its ability to estimate the level of Chl-*a* at various locations in the water body.

4.3. Significance of deep learning models for chlorophyll-a forecasting

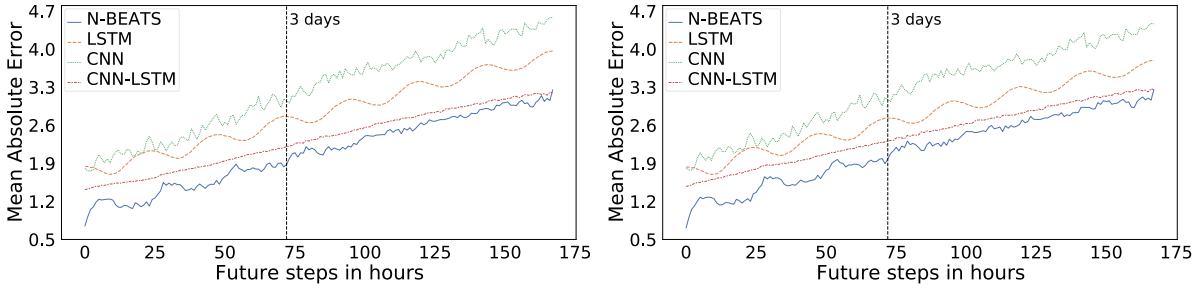
The one-day and one-week ahead forecasting results demonstrate that our proposed methodology produced DL model configurations capable of predicting Chl-*a* values over time, maintaining high accuracy in identifying the alert threshold at 10 μg/L up to the third day. Accurate predictions threshold based on WHO guidance are crucial for enabling informed decision-making related to sampling, protection, and risk management in water bodies affected by algal blooms. In this subsection, we analyze the results by focusing on the time windows and prediction speeds. False negatives and positives are then examined and modifications to the prediction threshold are evaluated to optimize detection of WHO-defined alert.

Table 6 compiles the findings presented in the prior sections, providing a comparative analysis with the results obtained using the Random Walk method as a benchmark. This naïve approach utilizes the last observed value as the forecasting for all future timesteps. The table shows optimal configurations for each buoy, prediction horizon, and associated DL architecture, contingent on the feature set and the model's input window size. It should be noted that, the configurations for the dam buoy align with the beach buoy, since predictions for the dam buoy leveraged the best-performing models initially trained on beach buoy data. Each optimal configuration features presents average MAE, MSE, and F1-score values, computed both daily and weekly (for more details about the metrics see the Section 3.3). While models with daily forecast horizons cannot directly produce weekly forecasts, those designed for weekly predictions can offer insights for their first predicted day. The table also shows inference speeds, with all models registering under one second per prediction. As previously emphasized, N-BEATS consistently outperforms other models in terms of MAE and MSE, indicating its superiority in closely approximating future Chl-*a* values. In numerous instances, N-BEATS demonstrates a substantially reduced MSE, suggesting fewer pronounced errors relative to other models. In relation to the identification of the 10 μg/L Chl-*a*, N-BEATS stands out, particularly for long-term, one-week ahead forecasts, and only marginally diverges from top models for daily forecasts by less than 1%. These long-term forecasting models hold pronounced value due to their inherent flexibility of forecast selection. N-BEATS, when trained on a one-week ahead forecast horizon, is notably competitive at one-day ahead forecasts. As depicted in Fig. 12, N-BEATS obtained a F1-score approaching 70% on the third day of prediction for the beach buoy, i.e., the one used to train the model. Additionally, its robust generalization capabilities shine through, as evidenced by nearly matching performance on the second day for the dam buoy situated



(a) Beach dataset (b) Dam dataset

Fig. 13. Daily average of the MAE for the best one-week ahead forecasting model configurations. Days are shown on the *x*-axis. The *y*-axis illustrates the mean MAE of the corresponding day. A unique color bar is assigned to the best configuration for each distinct architecture. (a) shows the results for beach buoy and (b) the results for dam buoy. Lower values suggest better forecasting accuracy. N-BEATS obtains the best overall results.



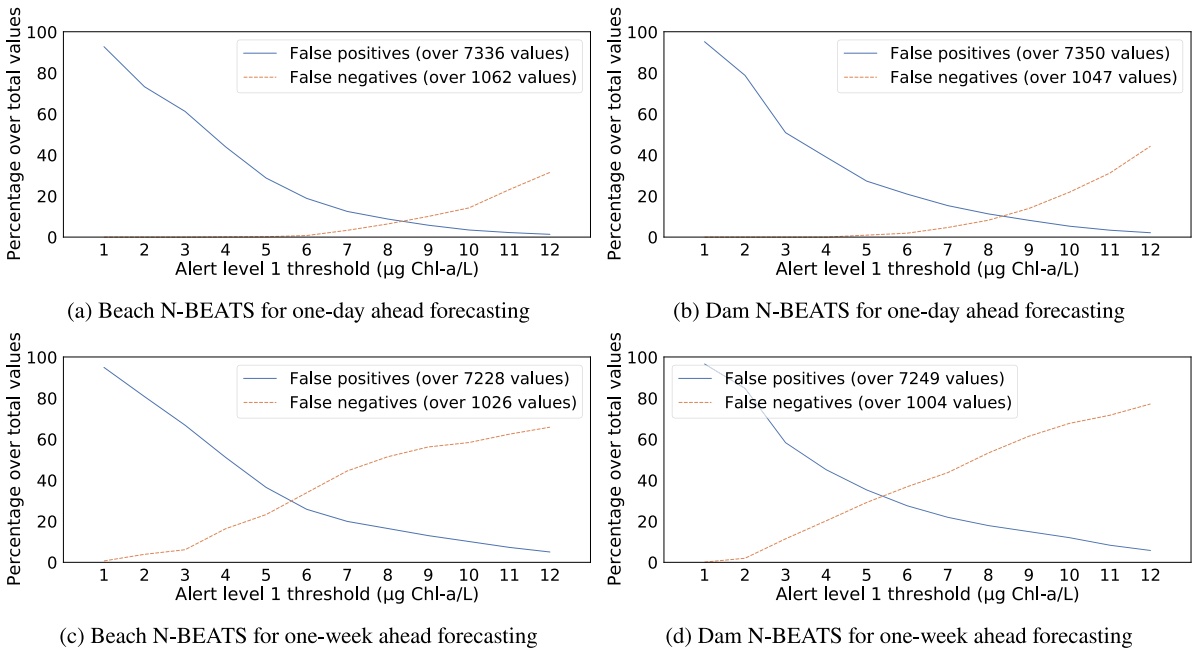
(a) Beach dataset (b) Dam dataset

Fig. 14. MAE on each forecasted time step for the best one-week ahead forecasting model configurations. Time steps are indicated in the *x*-axis and F1 is shown in *y*-axis. Unique colors are assigned to the best model configurations. Lower MAE equates to better forecasting accuracy. (a) shows the results for the beach buoy while (b) shows the results of dam buoy. The best model on the on all predicted hours is N-BEATS.

Table 6
Summary of results categorized by buoy testing dataset, forecast horizon, and architecture. The table represents optimal trained models based on the best feature set and input window size. Configurations for the dam buoy are consistent with those of the beach buoy. Displayed metrics include average MAE, MSE, and F1 values for both daily and weekly forecast horizons. It is worth noting that while models with a daily horizon are not designed for weekly predictions, those with a weekly horizon offer results for the first day. Additionally, inference time in seconds are provided.

Buoy	Horizon	Architecture	Features	Window	MAE (day)	MSE (day)	F1 (day)	MAE (week)	MSE (week)	F1 (week)	Inference time (s)
Beach	Day	N-BEATS	Chl-*a*	w72	**1.069**	**4.394**	0.835	–	–	–	0.0025
		LSTM	Chl-*a*+Exo	w72	2.024	6.866	0.810	–	–	–	0.0052
		CNN	Chl-*a*	w72	1.333	4.439	0.837	–	–	–	0.0026
		CNN-LSTM	Chl-*a*+Exo	w72	1.355	4.665	**0.840**	–	–	–	0.008
		Random Walk	Chl-*a*	w12	1.319	8.530	0.774	–	–	–	–
	Week	N-BEATS	Chl-*a*	w12	**1.124**	**4.410**	**0.809**	2.098	**13.354**	**0.638**	0.0023
		LSTM	Chl-*a*	w72	1.871	6.650	0.792	2.806	16.805	0.599	0.0074
		CNN	Chl-*a*	w24	2.030	6.758	0.774	3.247	18.213	0.607	0.0045
		CNN-LSTM	Chl-*a*	w72	1.527	6.191	0.777	2.342	15.745	0.619	0.014
		Random Walk	Chl-*a*	w12	1.257	7.688	0.769	**2.027**	17.568	0.636	–
Dam	Day	N-BEATS	Chl-*a*	w72	**1.100**	**4.303**	0.767	–	–	–	0.0026
		LSTM	Chl-*a*+Exo	w72	2.317	8.293	0.748	–	–	–	0.0052
		CNN	Chl-*a*	w72	1.350	4.405	**0.782**	–	–	–	0.0025
		CNN-LSTM	Chl-*a*+Exo	w72	1.478	4.877	0.776	–	–	–	0.0077
		Random Walk	Chl-*a*	w12	1.363	8.326	0.699	–	–	–	–
	Week	N-BEATS	Chl-*a*	w12	**1.165**	**4.697**	**0.763**	2.158	**12.279**	**0.549**	0.0023
		LSTM	Chl-*a*	w72	1.878	7.705	0.744	2.762	14.885	0.523	0.0075
		CNN	Chl-*a*	w24	2.075	7.369	0.730	3.245	16.832	0.524	0.0044
		CNN-LSTM	Chl-*a*	w72	1.627	7.781	0.718	2.429	14.400	0.528	0.0139
		Random Walk	Chl-*a*	w12	1.322	7.994	0.702	**2.135**	17.821	0.531	–

on the opposing lake end. The conclusions were validated applying a Model Confidence Set (MCS) [62] to the non-conditional forecasts obtained from the endogenous models. This procedure narrowed down the initial set of models to a subset with equivalent predictive ability for each dataset and forecasting horizon. As previously concluded, N-BEATS consistently demonstrates superior performance within every model confidence set. Detailed results of this validation can be found in Appendix.



(a) Beach N-BEATS for one-day ahead forecasting

(b) Dam N-BEATS for one-day ahead forecasting

(c) Beach N-BEATS for one-week ahead forecasting

(d) Dam N-BEATS for one-week ahead forecasting

Fig. 15. Assessment of alert detection based on Chl-*a* threshold variations. This figure juxtaposes detection thresholds from N-BEATS model predictions with the selected Chl-*a* threshold (10 µg/L). The *x*-axis enumerates the alternative thresholds in µg/L. On the *y*-axis, the false positive rates, represented by the blue line and based on 7000 possible values from the testing dataset, and the false negative rates, indicated by the orange line and based on 1000 possible samples, are visualized in the context of 10 µg/L Chl-*a* alert detection. Fig. 15(a) depicts the performance of the model trained with a one-day forecast horizon for the beach buoy. Similarly, Fig. 15(b) portrays results for the dam buoy under the same forecast horizon. For models trained with a one-week ahead forecast horizon, Fig. 15(c) provides insights for the beach buoy, and Fig. 15(d) highlights findings for the dam buoy.

In our study, as shown in Table 6, a distinct preference emerged for adopting a 72-h input time window when making one-day ahead forecasts. However, for week-long forecasts, no such clear trend is observed. For projections spanning a week, DL models appeared to be more effective at discerning trend and periodicity patterns intrinsic to the time series. These patterns, more prominent over a week than a single day, increased the model's performance metrics during training on a week-long forecast horizon. Models such as N-BEATS excel at identifying trend and periodicity patterns without necessitating exhaustive input window size. On the other hand, for daily forecasts, the relevance of trend or periodicity patterns of the time series was reduced, leading to a direct prediction challenge. Here, models took advantage of the larger data sets, often resorting to the largest possible window size and occasionally integrating exogenous variables. Interestingly, the fact that the models used exclusively the previous Chl-*a* values to predict suggests that, at least in certain water bodies such as the As Conchas reservoir, the use of Chl-*a* sensors for model training and alerting would be sufficient. This simplifies and reduces the cost of AHFM system design and operation, since the fewer sensors used, the less investment and maintenance of the nodes. Furthermore, the rapid prediction capabilities presented by these models pave the way for further reductions in associated hardware expenses.

All the models, notably those based on the N-BEATS architecture, are capable of producing predictions in as little as every 2 ms, a speed that vastly outpaces the conventional 15-min data collection interval. Leveraging the robust generalization ability of these models across different buoys, N-BEATS can achieve predictions for Chl-*a* levels and the WHO alert thresholds for an impressive array of up to 300,000 sampling points within each 15-min span, given the hardware discussed in this study (refer to Section 3.4). In the future context of environmental digital twins, where virtual replicas of physical systems or processes are created, the fast predictive capabilities of N-BEATS are perfectly aligned. This suggests that there is a substantial realm of opportunities to reduce computational costs for predictive models, potentially leading to an embedded solution within the buoy alongside the sensors.

On the other hand, predicting algal blooms presents an inherent challenge: balancing false negatives (i.e. HABs that do not trigger the alert) against false positives (i.e. erroneously anticipated HABs). The first pose public health risks, while the latter could have economic consequences, especially if they cause unnecessary restrictions of lakes used for tourism, drinking water supply or for industrial activities that depend on clean water (e.g., hydro-power production) [28]. Fig. 15 contrasts these errors, with blue denoting false positive percentages (based on 7000 possible instances of the testing dataset) and orange representing false negatives (based on 1000 possible instances). It was evident that daily forecasts manifested fewer discrepancies (both false positive and negatives) than their weekly counterparts. This is likely because long-term forecasts inherently face greater complexities.

Typically, there is a tendency that, as the alert threshold increases, false negatives increase and false positives decrease, and vice versa. Generally, at the 10 µg/L threshold, false positives were surpassed by false negatives (Fig. 15). In other words, as previously explained, the potential economic impact (related to the false positives) was outweighed by health risk (related to the false negatives). However, the model's adaptability empowers users to recalibrate this balance by opting for an alternate threshold in line with their specific needs. For instance, in water bodies for drinking water purposes, considering lower thresholds could refine detection fidelity. Pairing such threshold adjustments with the optimal prediction horizon can fine-tune both the prediction quality and the balance of false indicators. Yet, irrespective of threshold nuances, it is of paramount importance to corroborate any generated alerts with rigorous laboratory assessments for harmful species and toxin concentration, safeguarding both precision and public health.

5. Conclusions and future work

The study delves into predicting Harmful Algal Blooms (HABs), focusing on Chlorophyll-a levels and their threats to environmental, animal, and human health in the Anthropocene era. With the application of novel Deep Learning (DL) architectures such as Neural

Basis Expansion Analysis for Interpretable Time Series Forecasting (N-BEATS), it improves Chlorophyll-a level forecasting with high temporal resolution and extends the horizon to one week. Utilizing a publicly available dataset from the As Conchas reservoir, the study trains and evaluates an extensive array of models, totaling up to 2400 distinct configurations, with N-BEATS standing out for its ability to capture temporal dependencies efficiently. This DL architecture demonstrates superior long-term forecasting while also proving effective for short-term predictions, suggesting its potential as a substitute for short-term models with minimal loss in precision.

The main conclusions of this study are the following:

- Innovative Deep Learning Application: By employing advanced DL architectures, specifically N-BEATS, for predicting Chlorophyll-a levels, a high temporal resolution and an extension of the forecasting horizon to one-week are achieved. This approach shows promise in replacing short-term forecasting models with minimal loss of accuracy.
- Practical Advancements in HAB Management: A novel methodology for predicting initial alarm levels of HABs within a multi-horizon Chlorophyll-a level forecasting framework is proposed. This methodology is aligned with World Health Organization (WHO) guidelines and offers practical advantages for refined HAB management and mitigation initiatives.
- Effectiveness of Supplementary Variables: The integration of supplementary variables such as water temperature, pH, and conductivity has been found to not substantially improve predictive accuracy, affirming the validity of simpler models and challenging assumptions about increased complexity leading to enhanced performance.
- Model Portability: Models trained on data from one buoy exhibited minimal degradation in precision when applied to another buoy within the same water body, indicating potential for streamlined model training methodologies and broadened utility across various water body locations.
- Refined Alarm Level Inference: Interesting opportunities have been studied to refine predictive model alarm level inference by adjusting prediction thresholds, ensuring more timely and precise alarm identification, and ultimately preserving human health.

As future work, building upon the insights obtained from this study, there are multiple directions for future research to enhance the forecasting and understanding of HABs. A primary area of focus should be to increase data collection efforts, including a variety of aquatic environments and capturing a wider spectrum of variables that have known associations with HAB occurrences. This augmented dataset could potentially enable more complex models to accurately determine higher-risk alarm levels, which denotes a more severe threat. The extension of the forecast horizon to span one month may provide valuable insights into long-term HAB trends, although this would necessitate access to extensive water quality and climate data, especially in relation to nutrient inputs such as nitrogen and phosphorus. Employing broader temporal granularities, such as several hours or a full day, would also afford a more comprehensive view of short-term HAB patterns, contributing to the development of response strategies and ultimately improving our ability to manage and mitigate the impacts of HABs. The implementation of transfer learning methodologies presents a promising direction for future work. This approach, which uses pre-trained models on similar tasks or datasets, could simplify the training process, facilitating a quicker and more efficient path to accurate HAB forecasting models by capitalizing on previously learned patterns and knowledge. Additionally, there is potential for developing specialized classification models explicitly designed for HAB alarm detection. Such models could bypass the forecasting stage altogether, enabling direct alarm-level classification. While this might yield faster and potentially more precise assessments, it would also limit the flexibility of the approach compared to the methodology outlined in this study, as it would constrain the utilization of predicted Chlorophyll-a levels for other analytical purposes, such as examining the evolution and dynamics of HABs.

Code availability

The code and data used to conduct the experiments presented in this work can be found in the following publicly accessible repository: https://github.com/stanislavvakaruk/Chlorophyll_soft-sensor_machine_learning_models.

CRediT authorship contribution statement

Silvia Martín-Suazo: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jesús Morón-López:** Writing – review & editing, Writing – original draft, Validation, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization,. **Stanislav Vakaruk:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Amit Karamchandani:** Writing – review & editing, Writing – original draft, Validation, Investigation, Formal analysis, Conceptualization. **Juan Antonio Pascual Aguilar:** Writing – review & editing, Writing – original draft, Validation, Resources, Formal analysis, Data curation, Conceptualization. **Alberto Mozo:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Project administration, Investigation, Funding acquisition, Conceptualization,. **Sandra Gómez-Canaval:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Meritxell Vinyals:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Juan Manuel Ortiz:** Writing – review & editing, Writing – original draft, Validation, Resources, Funding acquisition, Conceptualization.

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

The dataset used in this study is publicly available and can be accessed at https://github.com/stanislavvakaruk/Chlorophyll_soft-sensor_machine_learning_models. The dataset includes measurements of Chl-*a*, pH, water temperature, electrical conductivity, and other relevant parameters collected from the As Conchas reservoir in the Miño-Sil River Basin District of Galicia, NW Spain. Detailed information regarding buoy location and data acquisition can be found in Section 3.1.

Acknowledgments

This work has been supported in part by Fundación Biodiversidad, Spain, the Spanish Ministry for Ecological Transition and the Demographic Challenge with grant CA-CC-2018 (CianoMOD), and partially by the Spanish Ministry of Economy, Industry and Competitiveness (MINECO), Spain, co-financed by the European Union (FEDER), under the grant RTC-2016-5087-2 (CIANOALERT). The co-authors thank the Miño-Sil Hydrographic Confederation for their support in the development of the project and for facilitating the access to the water quality data. This work was partially supported by Universidad Politécnica de Madrid, Spain under the project MLSEC with reference RP2161220029. The authors would like to thank M.Sc. student Miriam Guindel Gómez for her help with the design and coding of the preliminary experiments.

Table A.1
Results of the MCS relative to the mean MAE of all timesteps on the unconditional forecasts. Each forecasting horizon's models were evaluated using data from both buoys. The column denoted as S represents the critical value threshold, while the critical value indicates the predictive equivalence of the actual confidence set. The p-value reflects the statistical significance of these results. Additionally, the excluded model and resulting confidence set for each step are detailed.

Dataset	Horizon	S	Critical value	p-value	Removed	Model confidence set
Beach	Day	182.910	9.210	0.000	CNN	[N-BEATS, Random Walk]
		16.770	6.635	0.000	Random Walk	[N-BEATS]
	Week	336.190	13.277	0.000	CNN	[N-BEATS, LSTM, CNN-LSTM, Random Walk]
		142.448	11.345	0.000	LSTM	[N-BEATS, CNN-LSTM, Random Walk]
		30.469	9.210	0.000	CNN-LSTM	[N-BEATS, Random Walk]
		1.211	6.635	0.271	–	[N-BEATS, Random Walk]
Dam	Day	93.408	9.210	0.000	Random Walk	[N-BEATS, CNN]
		52.270	6.635	0.000	CNN	[N-BEATS]
	Week	284.010	13.277	0.000	CNN	[N-BEATS, LSTM, CNN-LSTM, Random Walk]
		137.200	11.345	0.000	LSTM	[N-BEATS, CNN-LSTM, Random Walk]
		31.968	9.210	0.000	CNN-LSTM	[N-BEATS, Random Walk]
		0.106	6.635	0.745	–	[N-BEATS, Random Walk]

Table A.2
Results of the MCS relative to the mean MSE of all timesteps on the unconditional forecasts. Each forecasting horizon's models were evaluated using data from both buoys. The column denoted as S represents the critical value threshold, while the critical value indicates the predictive equivalence of the actual confidence set. The p-value reflects the statistical significance of these results. Additionally, the excluded model and resulting confidence set for each step are detailed.

Dataset	Horizon	S	Critical value	p-value	Removed	Model confidence set
Beach	Day	9.510	9.210	0.009	Random Walk	[N-BEATS, CNN]
		0.048	6.635	0.827	–	[N-BEATS, CNN]
	Week	86.763	13.277	0.000	CNN	[N-BEATS, LSTM, CNN-LSTM, Random Walk]
		15.714	11.345	0.001	Random Walk	[N-BEATS, LSTM, CNN-LSTM]
		13.932	9.210	0.001	LSTM	[N-BEATS, CNN-LSTM]
		2.552	6.635	0.110	–	[N-BEATS, CNN-LSTM]
Dam	Day	3.337	9.210	0.189	–	[N-BEATS, CNN, Random Walk]
	Week	73.515	13.277	0.000	Random Walk	[N-BEATS, LSTM, CNN, CNN-LSTM]
		108.245	11.345	0.000	CNN	[N-BEATS, LSTM, CNN-LSTM]
		17.186	9.210	0.000	LSTM	[N-BEATS, CNN-LSTM]
		8.763	6.635	0.003	CNN-LSTM	[N-BEATS]

Appendix

A.1. Model confidence set results

The results presented in Section 4 were subsequently validated through a Model Confidence Set (MCS). This procedure aims to obtain a set of models with equal predictive ability by conducting several rounds of exclusions. Tables A.1 and A.2 provide an overview of each exclusion step until the formation of the final confidence set. This procedure was performed for each forecasting horizon and buoy. Furthermore, the analysis included both MAE and MSE metrics. Only models utilizing Chl-a as an input feature were considered, as conditional and non-conditional models cannot be compared. On the first step of the procedure, the critical value and S of the initial set are calculated. If the critical value exceeds the threshold S, it signifies a substantial difference between the models, thereby necessitating the exclusion step. The p-value quantifies the probability of obtaining the difference measured by S or an even larger difference, under the assumption of no true distinction between the models (i.e., when the null hypothesis holds). A p-value below a threshold, set at $\alpha = 0.01$, signifies the rejection of the null hypothesis, which indicates no substantial difference between the models under comparison. If a substantial difference is observed, pairwise comparisons of models using the multivariate Diebold–Mariano test are conducted and the model with highest anticipated future loss is excluded. This exclusion process continues until a subset of models with equal predictive capabilities is achieved, or until only one model remains in the subset.

References

[1] Maciej Zalewski, Edyta Kiedrzyńska, Iwona Wagner, Katarzyna Izydorczyk, Joanna Mankiewicz Boczek, Tomasz Jurczak, Kinga Krauze, Piotr Frankiewicz, Małgorzata Godlewska, Adrianna Wojtal-Frankiewicz, Małgorzata Łapińska, Magdalena Urbaniak, Agnieszka Bednarek, Zbigniew Kaczkowski, Ilona Gągała, Liliana Serwecińska, Sebastian Szklarek, Renata Włodarczyk-Marciniak, Arnoldo Font-Nájera, Elżbieta Mierzejewska, Małgorzata Połatyńska-Rudnicka, Kamila Belka, Paweł Jarosiewicz, Ecohydrology and adaptation to global change, Ecohydrol. Hydrobiol. 21 (3) (2021) 393–410, http://dx.doi.org/10.1016/j.ecohyd.2021.08.001.

[2] Jeff C. Ho, Anna M. Michalak, Nima Pahlevan, Widespread global increase in intense lake phytoplankton blooms since the 1980s, Nature 574 (7780) (2019) 667–670, http://dx.doi.org/10.1038/s41586-019-1648-7, ISBN: 4158601916 Publisher: Springer US.

[3] Zorica Svirčev, Dijana Lalić, Gorenka Bojadžija Savić, Nada Tokodi, Damjana Drobac Backović, Liang Chen, Jussi Meriluoto, Geoffrey A. Codd, Global geographical and historical overview of cyanotoxin distribution and cyanobacterial poisonings, in: Archives of Toxicology, Vol. 93, Springer Berlin Heidelberg, 2019, pp. 2429–2481, http://dx.doi.org/10.1007/s00204-019-02524-4, ISSN: 0340-5761 Number: 9.

[4] Anna M. Michalak, Study role of climate change in extreme threats to water quality, 2016, http://dx.doi.org/10.1038/535349a, URL http://go.nature.com/29ye4ul. ISSN: 14764687 Number: 7612 Pages: 349–350 Publication title: Nature Volume: 535.

[5] Cihelio Alves Amorim, Ariadne do Nascimento Moura, Ecological impacts of freshwater algal blooms on water quality, plankton biodiversity, structure, and ecosystem functioning, Sci. Total Environ. 758 (2021) 143605, http://dx.doi.org/10.1016/j.scitotenv.2020.143605, Publisher: Elsevier B.V..

[6] Mulalo Mutoti, Jabulani Gumbo, Afam Israel Obiefuna Jideani, Occurrence of cyanobacteria in water used for food production: A review, Phys. Chem. Earth A/B/C 125 (2022) 103101, http://dx.doi.org/10.1016/J.PCE.2021.103101, Publisher: Pergamon.

[7] Elisa Berdalet, Lora E. Fleming, Richard Gowen, Keith Davidson, Philipp Hess, Lorraine C. Backer, Stephanie K. Moore, Porter Hoagland, Henrik Enevoldsen, Marine harmful algal blooms, human health and wellbeing: Challenges and opportunities in the 21st century, J. Mar. Biol. Assoc. UK 96 (1) (2016) 61–91, http://dx.doi.org/10.1017/S0025315415001733.

[8] Isabella Sanseverino, Diana Conduto, Algal bloom and its economic impact, ISBN: 978-92-79-58101-4, 2016, http://dx.doi.org/10.2788/660478.

[9] The water crisis does not stop in New York, Nat. Water 1 (4) (2023) 301, http://dx.doi.org/10.1038/s44221-023-00078-5.

[10] Catherine E. Richards, Asaf Tzachor, Shahar Avin, Richard Fenner, Rewards, risks and responsible deployment of artificial intelligence in water systems, Nat. water (2023) 1–11, http://dx.doi.org/10.1038/s44221-023-00069-6, URL https://www.nature.com/articles/s44221-023-00069-6. Publisher: Nature Publishing Group.

[11] Silvia Beatriz Alves Rolim, Bijeesh Kozhikkodan Veettil, Antonio Pedro Vieiro, Anita Baldissera Kessler, Clóvis Gonzatti, Remote sensing for mapping algal blooms in freshwater lakes: a review, Environ. Sci. Pollut. Res. 30 (8) (2023) 19602–19616, http://dx.doi.org/10.1007/S11356-023-25230-2, URL https://link.springer.com/article/10.1007/s11356-023-25230-2. Publisher: Springer.

[12] Haobin Cen, Jiahan Jiang, Guoqing Han, Xiayan Lin, Yu Liu, Xiaoyan Jia, Qiyan Ji, Bo Li, Applying deep learning in the prediction of chlorophyll-a in the east china sea, Remote Sens. 14 (21) (2022) http://dx.doi.org/10.3390/rs14215461.

[13] Moein Izadi, Mohamed Sultan, Racha El Kadiri, Amin Ghannadi, Karem Abdelmohsen, A remote sensing and machine learning-based approach to forecast the onset of harmful algal bloom, Remote Sens. 13 (19) (2021) http://dx.doi.org/10.3390/rs13193863.

[14] Hamed Karimian, Jinhuang Huang, Youliang Chen, Zhaoru Wang, Jinsong Huang, A novel framework to predict chlorophyll-a concentrations in water bodies through multi-source big data and machine learning algorithms, Environ. Sci. Pollut. Res. (0123456789) (2023) http://dx.doi.org/10.1007/s11356-023-27886-2, ISBN: 1135602327886 Publisher: Springer Berlin Heidelberg.

[15] Ting Zhou, Yan Li, Bo Jiang, Juha M. Alatalo, Chen Li, Cheng Ni, Tracking spatio-temporal dynamics of harmful algal blooms using long-term MODIS observations of Chaohu Lake in China from 2000 to 2021, Ecol. Indic. 146 (2023) 109842, http://dx.doi.org/10.1016/J.ECOLIND.2022.109842, Publisher: Elsevier.

[16] Andrea Vander Woude, Steve Ruberg, Thomas Johengen, Russ Miller, Dack Stuart, Spatial and temporal scales of variability of cyanobacteria harmful algal blooms from NOAA GLERL airborne hyperspectral imagery, J. Gt. Lakes Res. 45 (3) (2019) 536–546, http://dx.doi.org/10.1016/j.jglr.2019.02.006, Publisher: International Association of Great Lakes Research.

[17] Katja Dörnhöfer, Philip Klinger, Thomas Heege, Natascha Oppelt, Multi-sensor satellite and in situ monitoring of phytoplankton development in a eutrophic-mesotrophic lake, Sci. Total Environ. 612 (2018) 1200–1214, http://dx.doi.org/10.1016/j.scitotenv.2017.08.219.

[18] Rafael Marce, Glen George, Paola Buscarinu, Melania Deidda, Julita Dunalska, Elvira de Eyto, Giovanna Flaim, Hans-Peter Grossart, Vera Istvanovics, Mirjana Lenhardt, Enrique Moreno-Ostos, Biel Obrador, Ilia Ostrovsky, Donald C Pierson, Jan Potuz, Sandra Poikane, Karsten Rinke, Sara Rodríguez-Mozaz, Peter A Staehr, Guido Waajen, Gesa A Weyhenmeyer, Kathleen C Weathers, Mark Zion, Bas W Ibelings, Eleanor Jennings, Automatic high frequency monitoring for improved lake and reservoir management, 50 (2016) 10780–10794, http://dx.doi.org/10.1021/acs.est.6b01604.

[19] Jesus Moron-Lopez, Cristina Rodriguez-Sanchez, Francisco Carreno, Joaquin Vaquero, Angel G. Pompa-Pernia, Myriam Mateos-Fernandez, Juan Antonio Pascual Aguilar, Implementation of smart buoys and satellitebased systems for the remote monitoring of harmful algae bloom in inland waters, IEEE Sens. J. (2020) 1–8, http://dx.doi.org/10.1109/JSEN.2020.3040139.

[20] Abhiram S.P. Pamula, Hamed Gholizadeh, Mark J. Krzmarzick, William E. Mausbach, David J. Lampert, A remote sensing tool for near real-time monitoring of harmful algal blooms and turbidity in reservoirs, JAWRA J. Am. Water Resour. Assoc. (2023) http://dx.doi.org/10.1111/1752-1688.13121, Publisher: John Wiley & Sons, Ltd.

[21] Jingwei Yang, Andreas Holbach, Andre Wilhelms, Yanwen Qin, Binghui Zheng, Hua Zou, Boqiang Qin, Guangwei Zhu, Stefan Norra, Highly time-resolved analysis of seasonal water dynamics and algal kinetics based on in-situ multi-sensor-system monitoring data in Lake Taihu, China, Sci. Total Environ. 660 (2019) 329–339, http://dx.doi.org/10.1016/j.scitotenv.2019.01.044, Publisher: Elsevier B.V..

[22] Mengyuan Zhu, Jiawei Wang, Xiao Yang, Yu Zhang, Linyu Zhang, Hongqiang Ren, Bing Wu, Lin Ye, A review of the application of machine learning in water quality evaluation, Eco-Environ. Health (2022) http://dx.doi.org/10.1016/j.eehl.2022.06.001.

[23] Rahim Barzegar, Mohammad Taghi Aalami, Jan Adamowski, Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model, Stoch. Environ. Res. Risk Assess. 34 (2) (2020) 415–433, http://dx.doi.org/10.1007/S00477-020-01776-2, URL https://link.springer.com/article/10.1007/s00477-020-01776-2. Publisher: Springer.

[24] Hu Li, Chengxin Qin, Weiqi He, Fu Sun, Pengfei Du, Improved predictive performance of cyanobacterial blooms using a hybrid statistical and deep-learning method, Environ. Res. Lett. 16 (12) (2021) http://dx.doi.org/10.1088/1748-9326/ac302d.

[25] Alberto Mozo, Jesús Morón-López, Stanislav Vakaruk, Ángel G. Pompa-Pernía, Ángel González-Prieto, Juan Antonio Pascual-Aguilar, Sandra Gómez-Canaval, Juan Manuel Ortiz, Chlorophyll soft - sensor based on machine learning models for algal bloom predictions, Sci. Rep. (2022) 1–23, http://dx.doi.org/10.1038/s41598-022-17299-5, ISBN: 0123456789 Publisher: Nature Publishing Group UK.

[26] Jiabao Wen, Jiachen Yang, Yang Li, Liqing Gao, Harmful algal bloom warning based on machine learning in maritime site monitoring, Knowl.-Based Syst. 245 (2022) 108569, http://dx.doi.org/10.1016/J.KNOSYS.2022.108569, Publisher: Elsevier.

[27] Jin Hwi Kim, Jae Ki Shin, Hankyu Lee, Dong Hoon Lee, Joo Hyon Kang, Kyung Hwa Cho, Yong Gu Lee, Kangmin Chon, Sang Soo Baek, Yongeun Park, Improving the performance of machine learning models for early warning of harmful algal blooms using an adaptive synthetic sampling method, Water Res. 207 (2021) 117821, http://dx.doi.org/10.1016/J.WATRES.2021.117821, Publisher: Pergamon.

[28] Ingrid Chorus, Martin Welker (Eds.), Toxic Cyanobacteria in Water. A Guide to Their Public Health Consequences, Monitoring and Management, second ed., CRC Press, Geneva, 2021.

[29] María Castrillo, Álvaro López García, Estimation of high frequency nutrient concentrations from water quality surrogates using machine learning methods, Water Res. 172 (2020) 115490, http://dx.doi.org/10.1016/j.watres.2020.115490, URL https://www.sciencedirect.com/science/article/pii/S0043135420300269.

[30] Hiroshi Yajima, Jonathan Derot, Application of the Random Forest model for chlorophyll- a forecasts in fresh and brackish water bodies in Japan, using multivariate long-term databases, J. Hydroinform. 20 (2017) jh2017010, http://dx.doi.org/10.2166/hydro.2017.010.

[31] Bing Li, Guishan Yang, Rongrong Wan, Georg Hörmann, Jiacong Huang, Nicola Fohrer, Lu Zhang, Combining multivariate statistical techniques and random forests model to assess and diagnose the trophic status of Poyang Lake in China, Ecol. Indic. 83 (2017) 74–83, http://dx.doi.org/10.1016/j.ecolind.2017.07.033, URL https://www.sciencedirect.com/science/article/pii/S1470160X17304466.

[32] Xue Li, Jian Sha, Zhong-Liang Wang, Application of feature selection and regression models for chlorophyll-a prediction in a shallow lake, Environ. Sci. Pollut. Res. 25 (2018) http://dx.doi.org/10.1007/s11356-018-2147-3.

[33] Jingjing Xia, Jin Zeng, Environmental factor assisted chlorophyll-a prediction and water quality eutrophication grade classification: a comparative analysis of multiple hybrid models based on a SVM, Environ. Sci.: Water Res. Technol. 7 (6) (2021) 1040–1049, http://dx.doi.org/10.1039/D0EW01110J, Publisher: The Royal Society of Chemistry.

[34] Li Wang, Xiaoyi Wang, Xuebo Jin, Jiping Xu, Huiyan Zhang, Jiabin Yu, Qian Sun, Chong Gao, Lingbin Wang, Analysis of algae growth mechanism and water bloom prediction under the effect of multi-affecting factor, Saudi J. Biol. Sci. 24 (3) (2017) 556–562, http://dx.doi.org/10.1016/j.sjbs.2017.01.026, URL https://www.sciencedirect.com/science/article/pii/S1319562X17300359.

[35] Fang Lu, Zhi Chen, Wenquan Liu, Hongbo Shao, Modeling chlorophyll-a concentrations using an artificial neural network for precisely eco-restoring lake basin, Ecol. Eng. 95 (2016) 422–429, http://dx.doi.org/10.1016/j.ecoleng.2016.06.072, URL https://www.sciencedirect.com/science/article/pii/S0925857416303974.

[36] Gooyong Lee, Faridah Othman, Shaliza Ibrahim, Min Jang, Determination of the forecasting-model parameters by statistical analysis for development of algae warning system, Desalin. Water Treat. 57 (55) (2016) 26773–26782, http://dx.doi.org/10.1080/19443994.2016.1190106, Publisher: Taylor & Francis tex.eprint: https://doi.org/10.1080/19443994.2016.1190106.

[37] P.J. García Nieto, E. García-Gonzalo, J.R. Alonso Fernández, C. Díaz Muñiz, Water eutrophication assessment relied on various machine learning techniques: A case study in the Englishmen Lake (Northern Spain), Ecol. Model. 404 (2019) 91–102, http://dx.doi.org/10.1016/j.ecolmodel.2019.03.009, URL https://www.sciencedirect.com/science/article/pii/S0304380019301061.

[38] Sangmok Lee, Donghyun Lee, Improved prediction of harmful algal blooms in four major south koreaś rivers using deep learning models, Int. J. Environ. Res. Public Health 15 (7) (2018) http://dx.doi.org/10.3390/ijerph15071322, URL https://www.mdpi.com/1660-4601/15/7/1322. Number: 1322 tex.pubmedid: 29937531.

[39] Zhenyu Yu, Kun Yang, Yi Luo, Chunxue Shang, Spatial-temporal process simulation and prediction of chlorophyll-a concentration in Dianchi Lake based on wavelet analysis and long-short term memory network, J. Hydrol. 582 (2020) 124488, http://dx.doi.org/10.1016/j.jhydrol.2019.124488, URL https://www.sciencedirect.com/science/article/pii/S0022169419312235.

[40] Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, Yoshua Bengio, N-BEATS: Neural basis expansion analysis for interpretable time series forecasting, in: Eighth International Conference on Learning Representations, 2020.

[41] Boris N. Oreshkin, Grzegorz Dudek, Paweł Pełka, Ekaterina Turkina, N-BEATS neural network for mid-term electricity load forecasting, Appl. Energy 293 (2021) 116918, http://dx.doi.org/10.1016/j.apenergy.2021.116918.

[42] B. Puszkarski, K. Hryniów, G. Sarwas, Comparison of neural basis expansion analysis for interpretable time series (N-BEATS) and recurrent neural networks for heart dysfunction classification, Physiol. Meas. 43 (6) (2022) 064006, http://dx.doi.org/10.1088/1361-6579/ac6e55.

[43] Thierry Rock Jossou, Zakaria Tahori, Godwin Houdji, Daton Medenou, Abdelali Lasfar, Fréjus Sanya, Mêtowanou Héribert Ahouandjinou, Silvio M. Pagliara, Muhammad Salman Haleem, Aziz Et-Tahir, N-beats as an EHG signal forecasting method for labour prediction in full term pregnancy, Electronicsweek 11 (22) (2022) 3739, http://dx.doi.org/10.3390/electronics11223739, URL https://www.mdpi.com/2079-9292/11/22/3739.

[44] Lijie Deng, Ke Ruan, Xun Chen, Xiaoying Huang, Yongqing Zhu, Weihao Yu, An IP network traffic prediction method based on ARIMA and N-BEATS, in: 2022 IEEE 4th International Conference on Power, Intelligent Computing and Systems, ICPICS, 2022, pp. 336–341, http://dx.doi.org/10.1109/ICPICS55264.2022.9873564, tex.eventtitle: 2022 IEEE 4th international conference on power, intelligent computing and systems (ICPICS).

[45] Stanislav Vakaruk, Amit Karamchandani, Jesús Enrique Sierra-García, Alberto Mozo, Sandra Gómez-Canaval, Antonio Pastor, Transformers for multi-horizon forecasting in an industry 4.0 use case, Sensors 23 (7) (2023) 3516, http://dx.doi.org/10.3390/s23073516, Publisher: Multidisciplinary Digital Publishing Institute.

[46] Souhaib Ben Taieb, Antti Sorjamaa, Gianluca Bontempi, Multiple-output modeling for multi-step-ahead time series forecasting, Neurocomputing 73 (10–12) (2010) 1950–1957, http://dx.doi.org/10.1016/j.neucom.2009.11.030.

[47] Konstantinos Benidis, Syama Sundar Rangapuram, Valentin Flunkert, Yuyang Wang, Danielle Maddix, Caner Turkmen, Jan Gasthaus, Michael Bohlke-Schneider, David Salinas, Lorenzo Stella, François-Xavier Aubet, Laurent Callot, Tim Januschowski, Deep learning for time series forecasting: Tutorial and literature survey, ACM Comput. Surv. 55 (6) (2022) 121:1–121:36, http://dx.doi.org/10.1145/3533382.

[48] Fatoumata Dama, Christine Sinoquet, Time series analysis and modeling to forecast: A survey, 2021, arXiv:2104.00164 [cs]..

[49] Sima Siami-Namini, Neda Tavakoli, Akbar Siami Namin, A comparison of ARIMA and LSTM in forecasting time series, in: 2018 17th IEEE International Conference on Machine Learning and Applications, ICMLA, 2018, pp. 1394–1401, http://dx.doi.org/10.1109/ICMLA.2018.00227.

[50] Hansika Hewamalage, Christoph Bergmeir, Kasun Bandara, Recurrent neural networks for time series forecasting: Current status and future directions, Int. J. Forecast. 37 (1) (2021) 388–427, http://dx.doi.org/10.1016/j.ijforecast.2020.06.008.

[51] Taoying Li, Miao Hua, Xu Wu, A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM2.5), IEEE Access : Pract. Innov. Open Solut. 8 (2020) 26933–26940, http://dx.doi.org/10.1109/ACCESS.2020.2971348.

[52] Xifeng Guo, Qiannan Zhao, Di Zheng, Yi Ning, Ye Gao, A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price, Energy Rep. 6 (2020) 1046–1053, http://dx.doi.org/10.1016/j.egyr.2020.11.078.

[53] Musaed Alhussein, Khursheed Aurangzeb, Syed Irtaza Haider, Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting, IEEE Access : Pract. Innov. Open Solut. 8 (2020) 180544–180557, http://dx.doi.org/10.1109/ACCESS.2020.3028281.

[54] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.

[55] Bryan Lim, Stefan Zohren, Time-series forecasting with deep learning: A survey, Phil. Trans. R. Soc. A 379 (2194) (2021) 20200209, http://dx.doi.org/10.1098/rsta.2020.0209, Publisher: Royal Society.

[56] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural Comput. 4 (1) (1989) 541–551.

[57] K. Fukushima, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biol. Cybernet. 36 (4) (1980) 193–202.

[58] Alberto Mozo, Bruno Ordozgoiti, Sandra Gomez-Canaval, Forecasting short-term data center network traffic load with convolutional neural networks, PLoS One 13 (2) (2018) e0191939.

[59] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, Kilian Q. Weinberger, Densely Connected Convolutional Networks, IEEE Computer Society, 2017, pp. 2261–2269, http://dx.doi.org/10.1109/CVPR.2017.243, URL https://www.computer.org/csdl/proceedings-article/cvpr/2017/0457c261/12OmNBDQbld. tex.eventtitle: 2017 IEEE conference on computer vision and pattern recognition (CVPR).

[60] Amit Karamchandani, Alberto Mozo, Stanislav Vakaruk, Sandra Gómez-Canaval, J Enrique Sierra-García, Antonio Pastor, Using N-BEATS ensembles to predict automated guided vehicle deviation, Appl. Intell. (2023) 1–66, Publisher: Springer.

[61] Ingrid Chorus, Jaime Bartram, Toxic cyanobacteria in water. a guide to their public health consequences, monitoring, and management, in: Limnology and Oceanography, Vol. 45, 1999, p. 1212, http://dx.doi.org/10.4319/lo.2000.45.5.1212, ISSN: 00243590 Number: 5.

[62] Peter Hansen, James Nason, Asger Lunde, The model confidence set, Econometrica 79 (2010) 453–497, http://dx.doi.org/10.2139/ssrn.522382.