SMARTLAGOON

DELIVERABLE 3.3

Report on results of machine learning approaches for flood, algal bloom and hypoxia predictions

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

SMARTLAGOON

Innovative modelling approaches for predicting Socio-environMentAl evolution in highly anthRopized coasTal LAGOONs

Deliverable 3.3

Deliverable No.:	3.3
Project Acronym:	SMARTLAGOON
Full Title:	Innovative modelling approaches for predicting Socio-environMental evolution in highly anthRopized coastal LAGOONs
Grant Agreement No.:	101017861
Workpackage/Measure No.:	WP3
Workpackage/ Measure Title:	Innovative modelling of environmental processes
Responsible Author(s):	Javier Senent-Aparicio (UCAM)
Responsible Co-Author(s):	Patricia Jimeno Sáez (UCAM)
Date:	30/06/2023
Status:	
Dissemination level:	Public

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

HISTORY OF CHANGES		
Date	Content	Author(s)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Abstract

SMARTLAGOON Project benefits from the support of the Horizon 2020 programme of the European Union within the framework of an Innovation Action. To ensure the development of their activities to achieve the expected results the eight organizations members of the consortium will release some deliverables during the timeline of the project (48 months in total). These deliverables are numbered and depend on each Work Package.

The present deliverable D3.3 is the third deliverable of the Work-Package 3 (WP3). The focus of this Work Package is to develop innovative modelling of the Mar Menor environmental processes in a holistic way. The first (task 3.1) and second tasks (task 3.2) of the WP3 were described in deliverables 3.1 and 3.2, respectively. The present deliverable reports the results of the third task (task 3.3) of WP3. The task 3.3 addresses the objective of developing new modelling approaches, combining different data collections, physics-based modelling and artificial intelligence for highly anthropised coastal environments. The focus of this task is to develop innovative approaches to creating environmental models will be developed based on the capacity of machine learning (ML) techniques to estimate nonlinear relationships between environmental variables. ML techniques can be used as an alternative to simulate environmental processes without the need to know the cause-effect relationship with the catchment or lake characteristics. These techniques will be combined with physically based models to build more efficient hybrid predictive models.

Specifically, these tasks consist of the development of machine learning approaches for enhancing predictions of flood, algal bloom and hypoxia. These tasks seek to develop new ML algorithms for predicting stream discharge, the likelihood of flood events, algal blooms and hypoxia. A combination of sensing data from Work Package 2 (WP2), and model output from WP3 (task 3.1 and 3.2), are utilized in the search for new predictive algorithm. Initially, for some of the tasks, the large dataset already available from Erken Lake was used, which helped to narrow down the field of different ML approaches that might be useful, before applying a selection of these approaches to the Mar Menor. We tested and applied a number of techniques available through comprehensive ML libraries implemented in Python (e.g. tensorflow, xgboost.ai and scikit-learn). Given the recent installation of the streamflow and quality measurement instruments in the framework of the SMARTLAGOON project, this document presents all the modelling carried out to date and will of course be continuously updated as new model capabilities are tested and new observational data are obtained.

In this deliverable we explain how each of the elaborated tasks that are part of this task 3.3 has been developed. First, we describe the motivation, objectives, interest of the results and methodology of each task. The different study areas (Erken Lake, Mar Menor Lagoon and other experimental areas) are described, and the required input data are analyzed. It is explained in detail how the development of the different ML models has been carried out and how they are combined with the outputs of the physics-based models of the other WP tasks. Finally, the results and conclusions reached are detailed.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

TABLE OF CONTENT

SMARTLAGOON

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

1. Introduction

1.1 Motivation

This document is the third deliverable of the Work-Package 3 (WP3) of the SMARTLAGOON project, led by the UCAM. The focus of this Work-Package is to develop innovative modelling of the Mar Menor environmental processes in a holistic way. This implies considering the water balance of the watershed, the ecological, physical and chemical dynamics of the water body, and its connections with the Mediterranean Sea. This will be achieved by combining physical and ecological models. The first (task 3.1) and second tasks (task 3.2) of the WP3 were described in deliverables 3.1 and 3.2, respectively, and are focused on the development of physically based catchment, lake and lagoon models. The present deliverable reports the results of the third task (task 3.3) of WP3.

The task 3.3 addressed the objective of developing new modelling approaches, which combine sensing data, physically based models and artificial intelligence techniques for highly anthropized coastal environments. The focus of this task is to develop innovative approaches to creating environmental models will be developed based on the capacity of machine learning (ML) techniques to estimate nonlinear relationships between environmental variables. As an alternative to physically based models, estimation of hydrological and ecological variables by ML techniques has also gained much attention among researchers. Different ML techniques, which have been introduced and widely applied to problems affecting dynamic environmental systems, were found to be powerful tools for the modelling of these processes. The advantage of the ML is that it can be trained to learn complex relationships without requiring a priori knowledge of the physical characteristics of the process. This feature makes ML an effective tool for modelling complex environmental processes. ML techniques can be used as an alternative to simulate environmental processes without the need to know the cause-effect relationship with the catchment or lake characteristics.

Specifically, the task 3.3 consists of the development of machine learning approaches for enhancing predictions of flood, algal bloom and hypoxia. The fist task aims to develop machine learning algorithms to predict the hydrographs in the ephemeral streams which drain the Mar Menor lagoon because of the more and more frequent episodes of extreme storms. The prediction of the chlorophyll variable directly related to algae blooms and of the dissolved oxygen variable for the prediction of hypoxia episodes is also analysed. A combination of sensing data from Work Package 2 (WP2), and model output from WP3 (task 3.1 and 3.2), will be utilized in the search for new predictive algorithms. ML techniques will be combined with physically based models to build more efficient hybrid predictive models. The data initially available for the development of all these tasks in the Mar Menor lagoon were not sufficient, so the large dataset already available from the Erken Lake and other lakes was used for some of them. This helped to refine the range of different ML approaches that might be useful, before applying a selection of these approaches to the Mar Menor. Several techniques available through comprehensive ML libraries implemented in Python were tested and applied. In addition to the tasks already described, other tasks related to the prediction

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

of water quality variables with ML were carried out. The prediction of suspended sediments in the water by means of physically based models and their improvement by means of ML was analyzed.

This report presents all the modelling carried out to date and will of course be continuously updated as new model capabilities are tested and new observational data are obtained given the recent installation of the flow measurement (video-monitoring) and water quality (monitoring buoy) instruments in the framework of the SMARTLAGOON project.

1.2 Goals

The aim of this deliverable is to report on the progress made in the development of the machine learning (ML) models to estimate hydrological and water quality variables of the Mar Menor. These techniques were combined with physically based models to build more efficient hybrid predictive models. As there was initially insufficient data available in the Mar Menor, the Erken Lake was chosen to test the different approaches that could later be applied in the coastal lagoon. For some of the tasks, we did not have enough data in Erken Lake either, so the techniques were tested in other experimental areas.

The general objectives of the task 3.3 are:

- Develop new ML methodologies to enhance flood, algal bloom and hypoxia prediction.
- Combine outputs from physically based models such as SWAT+ and GOTM-WET models (task 3.1 and 3.2) with ML models to improve predictions of target variables.
- Use for prediction with ML models the data collected by the technologies installed in the framework of the SMARTLAGOON project (data obtained from the video-based monitoring and data monitored by the buoy).
- Develop methodologies using machine learning models for improved prediction of additional water quality variables such as suspended sediments.

1.3 Interest in results

Coastal lagoons are ecosystems of great socioeconomic and environmental value. However, they are exposed to great anthropogenic and environmental pressures, mainly due to climate change, which threaten their sustainability.

One of the main effects of climate change is flooding. During the last few years, floods have caused a lot of personal and economic damage in the Campo de Cartagena (drainage basin to the Mar Menor), affecting the deterioration of the Mar Menor. Every time there is heavy rainfall in the area around the Mar Menor there are two negative effects. On the one hand, the risks associated with the civilian population and on the other hand, the massive entry of runoff with high nutrient loads into the Mar Menor. For this reason, managing the risk of flooding is an issue of interest in the recovery of the lagoon. Accurate prediction of hourly runoff during storms or heavy rainfall is an important task for better flood risk management. For this purpose, different hydrological models have been developed to provide a better understanding of the characteristics, processes, and responses of the catchment during storm events. In the SMARTLAGOON project, one of the world's

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

most popular physically based semi-distributed hydrological models, the Soil and Water Assessment Tool (SWAT) model, has been tested for this task 3.3. Nowadays, advances in modelling based on machine learning methodologies have led to good results in the field of hydrology. Therefore, the results obtained from the physical model have been compared with those obtained with different machine learning algorithms.

Another consequence of climate change as well as increased nutrient loading from anthropogenic sources is the eutrophication in lakes and lagoons. Eutrophication influences the level of algal blooms which are considered very harmful and can also cause hypoxic conditions in lakes. Harmful algal blooms are a serious threat to natural water systems and have been increasing worldwide (Burford et al., 2020; Watson et al., 2016). The chlorophyll (Chl) concentration is a useful indicator for measuring the abundance and variety of phytoplankton and/or algal biomass (Boyer et al., 2009). Because all photosynthetic algae include Chl, algal bloom can be easily predicted by investigation of Chl concentration in waterbodies. In the specific case of the Mar Menor, since 2016 there have been several episodes of unprecedented eutrophication caused by an abrupt increase in the average concentration of nutrients and Chl. Chl is an indirect measure of the health of the lagoon as several episodes of fish mortality have been attributed to lack of oxygen in the water due to massive algal blooms. Understanding and modeling the level of a eutrophication indicator, such as Chl, benefits the management of this complex system. To better manage and mitigate the effects of algal blooms, methods are needed to predict their timing and magnitude. Machine learning techniques have been used in this project for the prediction of algal blooms.

The third task of this deliverable focuses on hypoxia prediction. Hypoxia is a condition that occurs when the dissolved oxygen (DO) in water falls below 2-4 mg/l, which can have adverse consequences for ecosystem functioning. DO is a critical parameter for the health and functioning of lagoons and lakes. It refers to the amount of oxygen gas dissolved in water, which is essential for the survival of aquatic organisms and the overall ecosystem balance. Hypoxic conditions can arise due to various factors, including excessive nutrient input, algal blooms, and water pollution. Hypoxia severely impacts the survival and health of fish, invertebrates, and other aquatic organisms. This can disrupt the natural balance of the ecosystem and lead to changes in species composition. In hypoxic conditions, the release of nutrients from sediments increases, exacerbating eutrophication. These nutrients can fuel the growth of harmful algal blooms, further depleting oxygen as they decay. Eutrophication episodes in the Mar Menor are characterized by an increase in turbidity and a decrease in DO. Identifying the environmental conditions that induce hypoxia is an issue of great ecological importance in the case of the Mar Menor. In this study, machine learning algorithms were used to predict the OD concentration and thus hypoxia in the Mar Menor.

In a first phase, insufficient data were available for the Mar Menor. Therefore, different machine learning techniques were tested in the Erken lagoon and in other areas, which have an automated long-term monitoring programme that provides hourly data. These trials allowed us to perform comprehensive tests and establish the methodology to be used for the prediction of water quality variables.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

The runoff into the lagoon brings a large amount of sediment, which also contributes to habitat degradation and deterioration of the biota. In this SMARTLAGOON project, different techniques, including artificial intelligence, were tested for the prediction of sediments from stream discharge.

1.4 Theoretical framework

1.4.1 Flood prediction

Determining the relationship between precipitation and runoff in river basins is considered one of the most important problems facing hydrologists and engineers (Nor et al., 2007). Overall, this is essential for effective water management, flood protection, and climate change adaptation. It enables informed decision-making, supports sustainable development, and helps mitigate the risks associated with water-related hazards.

Understanding the rainfall-runoff relationship is crucial for designing effective flood protection measures in urban areas and agricultural lands. By accurately estimating runoff help to design infrastructure such as dams, levees, and drainage systems that can handle the expected flow of water during heavy rainfall events, thereby reducing the risk of flooding and associated damages. On the other hand, climate change is altering rainfall patterns, leading to more frequent and intense precipitation events in certain regions. Understanding the rainfall-runoff relationship is crucial for assessing the potential impacts of climate change on water resources, including changes in flood risk, drought conditions, and overall water availability. This information is vital for developing adaptation strategies and designing resilient water management systems.

Flood forecasting refers to the anticipated estimation of streamflow in a river or stream during a flood event. It consists of predicting the magnitude and time of occurrence of the flood in order to take preventive and mitigation measures. In the context of flood forecasting, hydrographs are a fundamental tool for understanding and visualising how the flow of a stream responds to intense precipitation and how the flood develops over time. Flood prediction is therefore based on the analysis of historical hydrographs and the application of models.

Intermittent rivers and ephemeral streams (IRES) are common in semi-arid Mediterranean climates and flash floods caused by extreme storm events (Jaeger et al., 2017) are becoming more severe due to climate change. Accurate simulation of sub-hourly flows in these streams is crucial for risk management and flood control.

Focusing on our case study. Numerous streams, almost all of them ephemeral, flow into the Mar Menor Lagoon. These streams play the role of draining rainwater into the lagoon. Flooding in this area is the result of several factors, among them the intense rainfall regime that occurs in the region, especially during "cold drops". During episodes of torrential rain, the Mar Menor streams can carry a very high flow of water, dragging sediments, vegetation and other materials. This can generate overflows in areas close to the channels and cause flooding in urban and agricultural areas. These events have raised concern and awareness of the need to take measures to prevent and mitigate the effects of flooding. Hence the importance of finding the relationship between precipitation and discharge for flood prediction.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

There are several approaches and approaches used to predict floods. These rainfall-runoff modelling frameworks can be divided into two main groups: physically based and/or data based approaches (Young et al., 2017). Physically based models are used to simulate the behaviour of water flow in a river basin. These models consider precipitation, topography, soil and other parameters to predict runoff and potential flooding. Examples of widely used hydrological models include HEC-HMS, SWAT, and MIKE SHE. Data-driven models use ML and statistical techniques to analyse historical flood data and generate predictions. Physics-based modelling aims to reproduce hydrological processes in a physically realistic way and is based on physical laws, whereas data-driven modelling aims to directly model the input-output relationship based on statistical correlations rather than physical principles (Nguyen et al., 2014).

There are several studies comparing these two types of models. In Nor et al. (2007), the radial basis function (RBF) method and HEC-HMS Model are applied to model the rainfall-flow relationship of two river basins in Malaysia. Both models are used to predict the flow hydrograph as a function of storm events. Young et al. (2017) developed a hybrid approach based on physics and machine learning for rainfall-runoff modelling during extreme typhoons. Chang et al. (2017) concluded that ANFIS is a reliable alternative to the physically based model in cases where less data is available.

1.4.2 Algal bloom prediction

Harmful algal blooms, which are a serious threat to natural water systems, have been increasing throughout the world (Burford et al., 2020; Watson et al., 2016), primarily because of both climate change and increased nutrient loading from anthropogenic activities (Brookes and Carey, 2011; Paerl and Huisman, 2008). Moreover, as indicated by Carey et al. (2012) and Huisman et al. (2018), more intense and longer periods of thermal stratification could potentially specifically favour blooms of toxic Cyanobacteria. To improve management and mitigation of the effects of algal blooms, methods are needed to predict their timing and magnitude. However, the factors regulating algal blooms are complex, variable and site-specific, and often involve high-order interactions between environmental factors and biogeochemical processes (Reichwaldt and Ghadouani, 2012; Richardson et al., 2018).

Process-based models encode our understanding of biogeochemical processes in a series of numerical formulations, but these are unavoidable simplifications that lead to an incomplete description of complex biogeochemical interactions and a low level of confidence in the models (Elliott, 2012). Based on novel data mining and statistical techniques, data-driven machine learning (ML) models have been implemented to identify patterns in observed data (Mellios et al., 2020), and with the increasing number of lake monitoring data (Marcé et al., 2016), ML models have been applied as an alternative to process-based models for the prediction of blooms (Rousso et al., 2020).

Considering the difficulty of modelling with a process-based model, multiple previous studies applying ML models, including random forest models (Nelson et al., 2018), support vector machine (Jimeno-Sáez et al., 2020) and artificial neural network models (Xiao et al., 2017) among others, have been found to improve chlorophyll predictions. However, a drawback of ML models is that they lack the interpretability and generalisability of process-based models. In recent years, process-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

driven deep learning (PGDL) modelling has emerged and has been applied to water temperature (Read et al., 2019) and water quality (Hanson et al., 2020), which explicitly combine well-defined physical theories in training ML models, improving their interpretability. Although this approach has demonstrated promising results, it is difficult to apply to phytoplankton dynamics due to the many non-linear interactions within biogeochemical cycles and the difficulty of defining measurable processes or mass balances that can be used as a physical constraint on knowledge-driven decisions. In addition, the scarcity of lake water quality observations (e.g. nutrients and chlorophyll concentration) may limit the application of ML models in algal bloom modelling (Rousso et al., 2020).

In order to overcome these data limitations, remote sensing using satellite sources (SRS), based on sensors on board satellites that detect and record reflected or emitted energy, is emerging as an increasingly widespread system for monitoring different natural environments (Huang et al., 2018; Lechner et al., 2020). These systems provide high-resolution images with high-quality geometric and radiometric information. In addition, biogeophysical parameters can be obtained from the processing of these images (Vos et al., 2019; Yang et al., 2013), developing different products for observation of the earth, oceans, etc. (Justice et al., 2002). ESA's Sentinels (Malenovský et al., 2012) and NPP VIIRS (Justice et al., 2013) products offer ever better spatial and temporal resolutions, as well as the development of new services.

SRS-based products that generate biogeochemical data are developed on a global scale and often need to be adapted to more local or regional contexts. In fact, there can be significant discrepancies between satellite-derived products and the actual state of surface parameters (Wu et al., 2019). This implies that data generated by satellite products are not useful due to large deviations when compared to in situ measurements (i.e. ground-truth). This reduces not only their usefulness in applications, but also the accuracy of their derived products. Adapting these SRS systems in local/regional contexts is therefore essential to improve their reliability. SRS data is becoming popular for near real-time (NRT) monitoring systems to address natural and societal challenges such as shedding light on global and regional challenges, thereby supporting a number of initiatives (e.g. Sendai Framework, Paris Agreement, Sustainable Development Goals) (Grainger and Kim, 2020; Noe et al., 2022; Walz et al., 2020).

Several studies have been proposed to develop SRS-based monitoring systems for Chl in the Mar Menor lagoon (Caballero et al., 2022; Erena et al., 2019; Gómez et al., 2021; González-Enrique et al., 2023). However, these works assume the validity of the data obtained by satellite systems, focusing mainly on one or two satellite sources, mainly Sentinel-derived products, without going into the validation process and/or coverage of all existing products for this gap. This modeling process is critical in the current unique ecosystem. The reasons are the low quantity and quality of data available in situ, the spatial mismatch between these measurements and remotely sensed pixels, the disregarding intrinsic heterogeneity of land and ocean surfaces that release nutrients into the lagoon, the shortcomings and deficiencies of theories on the scale problem in the validation.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

1.4.3 Hypoxia prediction

Dissolved oxygen (DO) is a parameter often used to assess water quality, responding to changes in phytoplankton photosynthesis, ecosystem respiration and lake mixing. The increases of duration and frequency of hypolimnetic hypoxia is often used as an indicator of aquatic ecosystem health. The decline of DO can have significant impact on water quality by increasing internal nutrient loading and further modifying lake trophic level (Orihel et al., 2017), promoting harmful algal bloom (Paerl and Paul, 2012), and also reducing fish habitat and even leading to fish kills (Rao et al., 2014).

Nowadays, hypoxia is becoming more common in the hypolimnion of lakes due to intensifying thermal stratification and loss of water clarity (Jane et al., 2021; North et al., 2014). The variability of DO is affected by the combined effect of physical, chemical and biological conditions acting on the lake systems. Factors that regulate DO concentration include atmospheric drivers, heat convection, algal biomass, nutrient loading, sediment resuspension, and sediment oxygen demand (Ladwig et al., 2021; Müller et al., 2012).

A variety of modelling tools, including analytical models (Bouffard et al., 2013; Cortés et al., 2021), and numerical models, i.e., coupled hydrodynamic-water-quality models (Ladwig et al., 2021; Leon et al., 2011) have been developed to simulate lake DO dynamics via parameterizing air-water interaction, photosynthesis, respiration, biochemical oxygen demand, and sediment oxygen demand. However, water quality simulations tend to be restricted to individual lake due to the unique biogeochemical circle within each lake.

State-of-art data driven ML models have been applied in the wide range of water resource research, simulating lake water temperature (Read et al., 2019; Yousefi and Toffolon, 2022) and water quality parameters, e.g., phosphorus (Hanson et al., 2020) and algal bloom (Lin et al., 2023). To solve the OD prediction task in different lakes, we developed three ML model approaches based on two ML models, Gradient Boosting Re-gressor (GBR) and Long-short-term-memory (LSTM) network to simulate multi-year seasonal-scale surface and background DO concentrations in 5 lakes with various sizes and trophic levels. This hybrid approach has achieved promising results in predicting algal blooms in a mesotrophic lake (Lin et al., 2023), and can also improve predictions of lake DO and hypoxia, which is highly dependent on lake hydrodynamics.

1.4.4 Other water quality variables prediction

- **Sediment Load prediction:**

Sediments transported by streams are of additional interest in environmental engineering, especially if the sediments transport pollutants. Suspended sediment inputs are among the main factors contributing to water quality degradation (Zeiger and Hubbart, 2016). Sediments from the watercourses into the lagoons can contain a significant organic pollutants and/or absorbed heavy metals, becoming a vehicle for the diffusion of harmful substances that can compromise water quality and aquatic ecosystems in the receiving lagoons.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

To assess the extent of the soil erosion problem and identify problem areas within a basin, the amount of sediment transported by streams or rivers must first be reliably quantified. However, finding accurate tools to estimate suspended sediment load (SSL) is challenging since fluvial sediment transport is a complex, non-linear process influenced by hydrographic, hydraulic, climatic, and anthropogenic factors in the river basin (Zounemat-Kermani et al., 2020). To simulate sediment transport processes in river basins and water bodies, several empirical, physically based, and conceptual methods are available (Borrelli et al., 2021; Merritt et al., 2003). Such models are useful for estimating sediment concentrations and/or loads generated under different climatic conditions, land use or management strategies, or in ungauged river basins (Fu et al., 2019).

The Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998) is an efficient tool for many types of water resource and land management applications (Gassman et al., 2014) and the most popular physically based model applied at the river-basin scale (Fu et al., 2019). Numerous studies conducted worldwide have concluded that the SWAT model provides a satisfactory estimate of sediment load (Duru et al., 2018; Dutta and Sen, 2018; López-Ballesteros et al., 2019; Pulighe et al., 2019). Physically based models simulate sediment produced by rainfall events in a river basin based on the laws of conservation of mass and energy. They provide an understanding of river basin processes, which is beneficial for assessing the impacts of soil and water conservation measures (Singh et al., 2014). However, a disadvantage of these models is that they require extensive information for development, calibration, and validation. In addition, calibrating the numerous parameters in such models is highly complex, due to the non-linearity of sediment transport processes, and requires expert knowledge and high computational time compared to data-driven models (Hamaamin et al., 2016; Khosravi et al., 2020). These drawbacks demand the exploration of more efficient methods of sediment computation.

Due to their capacity to model complex non-linear systems, data-driven methods such as machine learning (ML) have emerged as a powerful alternative to physically based models. In water resource management, such models can be used to estimate variables such as streamflow (Jimeno-Sáez et al., 2018; Minns and Hall, 1996; Srivastava et al., 2006) and sediment transport (Chen and Chau, 2016; Gupta et al., 2021; Kumar et al., 2016; Olyaie et al., 2015; Zounemat-Kermani et al., 2020). In contrast to physical models, ML models uses mathematical functions to connect inputs to outputs, ignoring the physical, logical relationship between variables (Ji et al., 2021). Kisi (2005) used a neural network approach and a neuro-fuzzy technique to estimate current suspended sediment values based on previous streamflow and sediment data, finding that the best results were obtained using neuro-fuzzy techniques. Al-Mukhtar (2019) predicted suspended sediments in the Tigris-Baghdad river using random forest (RF), support vector machine, and neural network techniques. The results showed that RF performed best. Ghasempour et al. (2021) used ML techniques to predict sediment, presenting a kernel-based approach based on the Gaussian process and the extreme learning kernel. The former was used for linear processes, while the latter was applied to non-linear processes. Sihag et al. (2021) developed a study to evaluate the best model using M5P and RF regression techniques to estimate sediment. The M5P-based model performed best in the study.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Few studies have compared SWAT and ML approaches. Singh et al. (2014, 2012) compared SWAT with a multilayer perceptron artificial neural network model and a radial basis neural network by simulating monthly sediment yields and obtained better results using the artificial neural network. Kim et al. (2012) observed that using artificial neural networks to predict total suspended solids was a useful alternative to SWAT. Sirabahenda et al. (2020) evaluated the SWAT model and the adapted neuro-fuzzy inference system (ANFIS) for SS prediction concluding that ANFIS obtained higher accuracy than SWAT.

2. Materials

2.1 Study areas

2.1.1 Lake Erken

Lake Erken is a mesotrophic lake located in east-central Sweden that has a surface area of 24 km^2, a maximum depth of 21 m, and an average retention time of 7 years. Its surface is usually ice-covered during winter, whereas during the summer, the water is stratified. It is a shallow lake, the mean depth is around 9 m. The lake is dimictic, with seasonal stratification commonly beginning in May–June and ending in August–September. The onset of ice cover usually begins in December–February, and the loss of ice occurs in March–April (Persson and Jones, 2008). Located near the Baltic coast, Lake Erken is windexposed and susceptible to periodic wind-induced turbulent mixing. Changes in algal Chl in Lake Erken have a typical seasonal pattern, with spring and summer peaks in concentration (Pettersson et al., 2003). Spring blooms are dominated by dinoflagellates and diatoms (Pettersson, 1985) and initiated by overwinter species from the last autumn (Yang et al., 2016). Cyanobacteria dominate summer peaks in Chl, given that they can optimize their vertical position with regard to nutrients and light (Paerl, 1988; Pierson et al., 1992).

2.1.2 Mar Menor lagoon

The Mar Menor is a 135 km^2 coastal lagoon located in Región de Murcia (southeastern Spain). This lagoon is one of the largest coastal lagoons in Europe and the largest in the Iberian Peninsula. This lagoon has an average depth of 3.6m and a maximum depth of 7m. It is isolated from the Mediterranean Sea by a 22 km long sandy coastal barrier called La Manga. La Manga is crossed by several gullies, which determines the semi-confined nature of its waters and gives them unique salinity and temperature characteristics. The Mar Menor, in addition to its environmental importance, is a key place for the economy of the Region of Murcia. Its unique climatic conditions and abundant natural resources have attracted tourism, recreational uses, and fishing, as well as the importance of agriculture to the local economy. The basin that drains the Mar Menor is called Campo de Cartagena and is an extensive plain of more than 1,200 km^2 with a network of ephemeral streams that collect the scarce but intense rainfall (Senent-Aparicio et al., 2015). The Mar Menor has been characterized by its oligotrophic waters and by its strong resistance to eutrophication. The transparent waters of the lagoon have been its main characteristic but during the last decade its waters have become eutrophic (Conesa and Jiménez-Cárceles, 2007). In recent decades, the environment of the Mar

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Menor has undergone significant changes due to the intensification of tourism and the expansion of intensive agriculture (Álvarez-Rogel et al., 2020). These changes have led to an increase in non-point source pollutant (NPS) loads in the lagoon, which has had a negative impact on its ecological status, such as increased levels of algae, more frequent episodes of hypoxia and consequent fish kills (López-Ballesteros et al., 2023). Agriculture has been recognised as the main source of NPS among the main sources of nutrients, such as nitrogen and phosphorus (Liu et al., 2013). Several studies (Domingo-Pinillos et al., 2018; Senent-Aparicio et al., 2021) show that the flow transfer from the Campo de Cartagena aquifer to the lagoon has a significant impact due to the entry of nitrates and other agrochemical elements from agricultural fertilizers. In addition, in the Mar Menor watershed, surface runoff can be very high due to common torrential rainfall events (Senent-Aparicio et al., 2021) generating massive inflows of water and pollution loads into the coastal lagoon and further aggravating its vulnerability. This surface runoff comes through the ephemeral streams that are the main drainage system of the Campo de Cartagena, which flow only during intense rainfall events (Alcolea et al., 2019; Senent-Aparicio et al., 2015). Among all ephemeral streams, the eight most relevant in the Campo de Cartagena are: Mirador, Peraleja, La Maraña, Albujón, Miranda, Miedo, Beal and La Carrasquilla. Large quantities of water and NPS pollutant loads move through these ephemeral streams into the coastal lagoon during these extreme rainfall events (Velasco et al., 2006). The return from crop irrigation recharges the aquifer by increasing the piezometric level and consequently, the Albujón stream, which is the most relevant of all the ephemeral streams, maintains a base flow in the final reach of the outlet, fed by the discharge from the aquifer (García-Pintado et al., 2009). These inputs caused a drastic increase in pollution in the lagoon (García-Ayllón, 2019) and resulted in a process of eutrophication, causing a decrease in water quality (Jimeno-Sáez et al., 2020).

In 2016, an extreme eutrophication event resulted in sharp bloom of phytoplankton (Mercado et al., 2021) that produced a significant change in the quality of the waters, with a substantial increase in turbidity and a loss of transparency. The algal blooms reduce the bottom light, enhancing benthic photosynthesis inhibition (Cloern et al., 2016). The death of benthic flora and the decomposition of their organic material increases the oxygen demand and can, eventually, lead to hypoxia-anoxia events. These generated important concerns among the environmental community and the tourism industry, with significant socio-economic repercussions (Garcia-Ayllon, 2018). Subsequently, after the "Santa Maria" flood of September 2019, attributed to the meteorological phenomenon known as "cold drop", the ecological degradation of the Mar Menor was aggravated by the massive input of nitrogen, phosphorus and organic matter, exceeding even the peak of Chl observed in 2016 (Jimeno-Sáez et al., 2020). Later, in August 2021, there was a massive die-off of fish and crustaceans (Caballero et al., 2022). Under this scenario, early intervention is required to reduce nutrient and other pollutant entry into the lagoon.

For section 3.1. of the tasks contained in this deliverable, the study area of study is the Mosa basin. The study area covers about 7.15 km^2. The basin has an altitude ranging from 560 m to 230 m at its lowest point (Fig. 1) and is characterized by steep slopes and rapid response to storms This basin is a headwater basin draining the left bank of the Albujón stream. The stream in the basin is ephemeral

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

and only carries water during specific periods in response to rainfall. This stream is dry for prolonged periods.

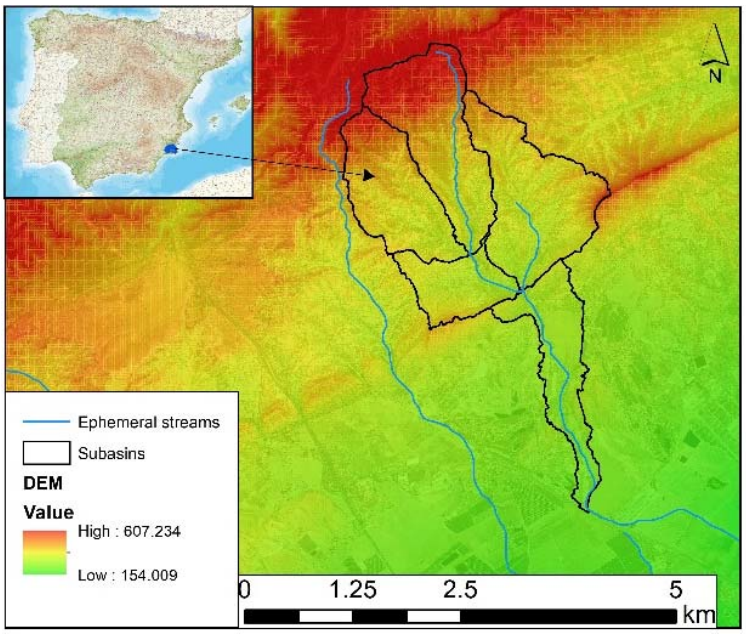


Fig. 1. Mosa basin at the headwaters of the Mar Menor basin.

2.1.3 Other study areas

For the task of predicting DO and hypoxia episodes, experiments were carried out on 4 other lakes (Lake Müggelsee (Germany), Lake Furesø (Denmark), Lake Mendota (U.S.), and Lake Ekoln (Sweden)) in addition to Lake Erken (Sweden). The characteristics of each lake are described in Table 1.

Table 1. Physical characteristics of the lakes.

Characteristics	Erken[1]	Müggelsee[2]	Furesø[3]	Mendota[4]	Ekoln[5]
Lake area (km^2)	23.7	7.4	9.4	39.6	29.8
Mean/Max depth (m)	9/ 21	4.9/8	7.4/37.7	12.7/ 25	15.4/ 50
Residence time (years)	7	0.12-0.15	10	4.3	<1
Lake type	dimictic	Polymictic[6]	dimictic	dimictic[4]	dimictic
Trophic state	Mesotrophic	Eutrophic	Mesotrophic	Eutrophic	Eutrophic

[1] Pierson et al. (1992); [2] Kakouei et al. (2022); [3] Gurkan et al. (2006); [4] Farrell et al. (2020); [5] Goedkoop et al. (2011); [6] Shatwell and Köhler (2019).

In Lake Furesø, a major restoration project started in 2003 to control the internal loading of phosphorus from the sediment during stratification period with the actions including the combination of hypolimnetic aeration and biomanipulation (Gurkan et al., 2006). Since then, the hypoxia has been reduced in the bottom waters of the lake.

The invasive aquatic plant species Nuttall's waterweed (*Elodea nuttallii*) was first detected in 2011 in Lake Müggelsee, spreading rapidly, and was the most abundant macrophyte species by 2017. The abundance of Another invasive species, the dreissenid mussel, increased with the increasing invasive waterweed, following the invasion meltdown hypothesis (Wegner et al., 2019). *E. nuttallii* can

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

largely increase the oxygen production via photosynthesis but also can result in extremely low DO in the bottom water of lakes, developing night-time anoxic conditions (Vilas et al., 2017).

In Lake Mendota, the grazing activity of freshwater zooplankton, i.e., Daphnia, leads to a reduction in algal biomass in late spring-early summer, which is so-called clear water phase (Carpenter and Kitchell, 1988). However, the lake also experienced the invasion from the spiny water flea, *Bythotrephes longimanus* (hereafter *Bythotrephes*) since the fall of 2009, which led to the decline in water clarity after 2009. Although one of the major *Daphnia* species (i.e*., Daphnia pulicaria*) is a preferred prey of *Bythotrephes*, another smaller-bodied *Daphnia Mendotae*, which competes differently with *Bythotrephes* increased in spring after the invasion (see Fig. 2). The combined result of these changes was that grazing on phytoplankton and probably accelerating organic matter mineralization and hypolimnetic oxygen depletion before summer stratification (Ladwig et al., 2021; Matsuzaki et al., 2021; Walsh et al., 2017).

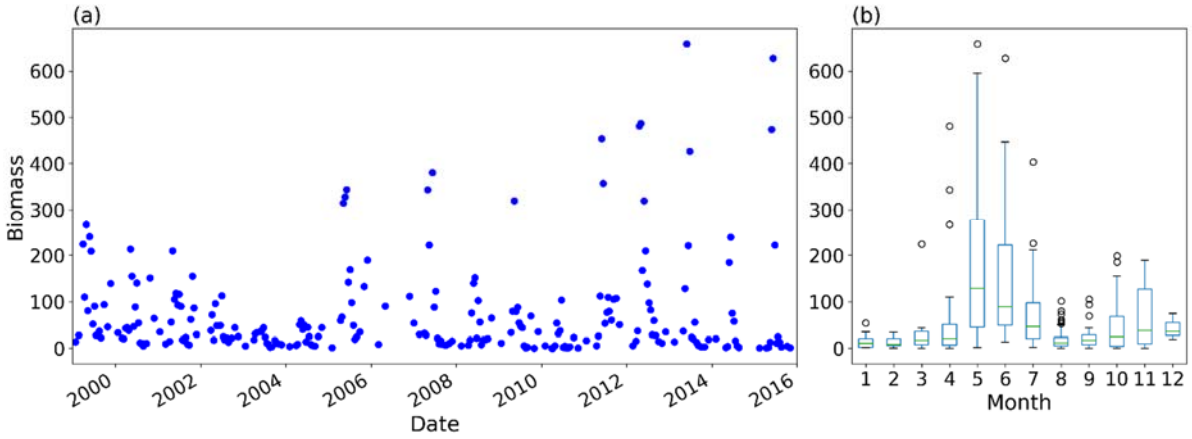


Fig. 2. (a) Timeseries of Daphnia Mendota biomass in Lake Mendota, and (b) shows the boxplots of biomass over the month.

For the study of sediment as the target variable of the prediction task, the study area of the Mar Menor and the Erken lagoon did not have sufficient data to carry out a preliminary experiment. After a hard task of finding areas where to test the methodology, an experimental area was chosen, the Oskotz basin (Navarra, northern Spain), where sufficient quality data was available. In Navarra, Spain, soil erosion is a major problem on agricultural land (Casalí et al., 1999; De Santisteban et al., 2006). For this reason, the Government of Navarra created a network of experimental basins to obtain data on water quality and soil erosion and evaluate the impact of agriculture in different areas of the region. In terms of morphology, soils, climate, land use, and management, these experimental river basins are representative of large areas of Navarra and Spain (Casalí et al., 2008). Oskotz is one of the four pilot basins in this network. This task uses data recorded at the Oskotz river basin to study SSL. Oskotz river basin is a small experimental basin covering 16.74 km^2 More details of the basin can be found at (Casalí et al., 2010; Jimeno-Sáez et al., 2022).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

2.2 Data

2.2.1 Lake Erken

Lake Erken has a long-running automated monitoring program that provides hourly meteorological data, water temperature profiles, and the flow from the inflow and outflow (Fig. 3).

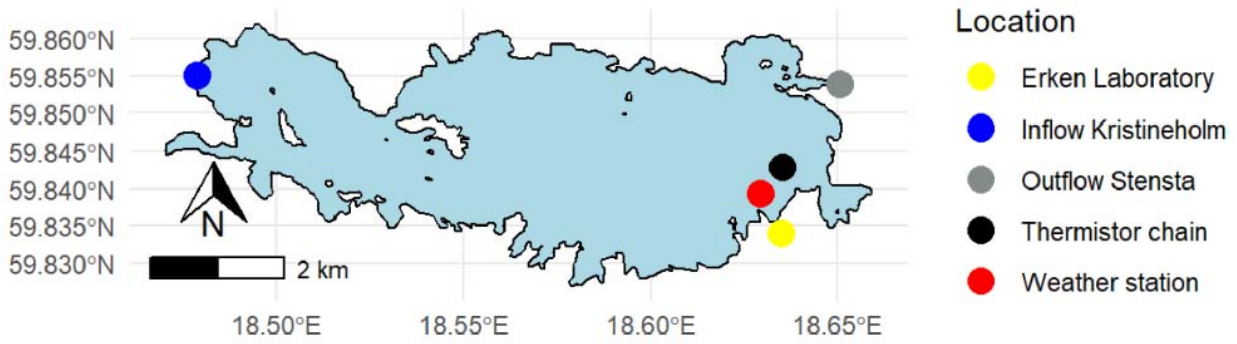


Fig. 3. Map of Lake Erken. The locations of the monitoring systems are shown.

A meteorological station on an island offshore from Uppsala University's Erken Laboratory provides measurements of wind speed, solar radiation, and air temperature. An automated water temperature monitoring system records water temperature profiles at a depth of 15 m with sensors placed at 0.5 m intervals. Water discharge is measured entering the lake from the largest input at Kristineholm, and the outflow at Stensta (Fig. 3). These data have been further quality controlled and combined with data from other nearby meteorological stations to provide a long-term dataset that is suitable as input for model simulations (Moras et al., 2019).

Since 1991, a consistent (1-2 week) monitoring program has collected integrated water samples from the epilimnion and hypolimnion during stratified conditions or from the entire water column during isothermal conditions. Stream samples are collected from the main inflow at Kristineholm and the outflow of the lake. All samples are analyzed by the Erken Laboratory for all major nutrient concentrations (e.g., NOX, NH4, PO4, Total P, Si, etc.), dissolved oxygen (O2), and Chl concentration.

Water and nutrient loads input to the GOTM/SELMAPROTBAS model were calculated from the discharge and nutrient concentrations measured at Kristineholm (Fig. 3) which accounted for 50.7 % of the lake watershed. Inputs from the remaining watershed were estimated from the measured Kristineholm inputs that were scaled by area to account for the remaining 49.3 % of the watershed area.

2.2.2 Mar Menor lagoon

The data used in the tasks included in this deliverable come from various sources. The data available in the Mar Menor are much more limited than those available for the Erken lake test area.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

For the flood prediction task, precipitation and streamflow data of ephemeral streams that drain into the Mar Menor were used as input data. The precipitation data were obtained from the SAIH-Segura pluviometry network service, the Automatic Hydrological Information System of the Segura Hydrographic Basin, which collects and stores the hydrological, hydraulic, and complementary information of the entire basin, at any time and under any circumstances, even in real time and with centralised control (https://www.chsegura.es/es/cuenca/redes-de-control/saih/).

Collecting observed flow data in ephemeral stream presents a significant challenge with classical hydrometric techniques due to the high water velocities and the presence of floating objects, which not only pose a safety risk to operators, but also have the potential to damage any instrument in contact with the water. To address these challenges and within the framework of the SMARTLAGOON project, an in-situ monitoring system based on computer vision using Particle Image Velocimetry to monitor flows at a fixed location was installed in the several studied streams. With this monitoring system it is possible to obtain flow data on a sub-hourly scale (5 minutes). In this way, the hydrographs of the storms can be recorded. These observed flow data are used together with the precipitation data at the same time scale as inputs to the ML models to simulate the hydrograph produced by the increasingly severe storms in the study area. Another more economical monitoring system was installed at the same point in the stream to provide support for the fixed camera. This system called Freestation. More details on these systems can be found in section 3.1.

To develop the SWAT hydrological model in the Mosa basin described in section 2.2.2 different data from different sources were collected. With these data the sub-basins, the drainage network and the Hydrologic Response Units (HRUs) were determined. The tool designed for QGIS Mapswat (López-Ballesteros et al., 2022) was used for the preparation of input data. These data were obtained from the following sources: (i) DEM: Digital terrain model made from the LIDAR point clouds of the first Cover with a mesh pitch of 25 meters, from Spanish National Geographic Institute (IGN); (ii) SOILS: Harmonized World Soil Database v 1.2 (Nachtergaele et al., 2010a). 30 arc-second raster database; and (iii) LANDUSE: Corine Land Cover 2018 of European Environment Agency (EEA) under the framework of the Copernicus programme. A graphical summary of the process used for SWAT data preparation is shown in Figure 4.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

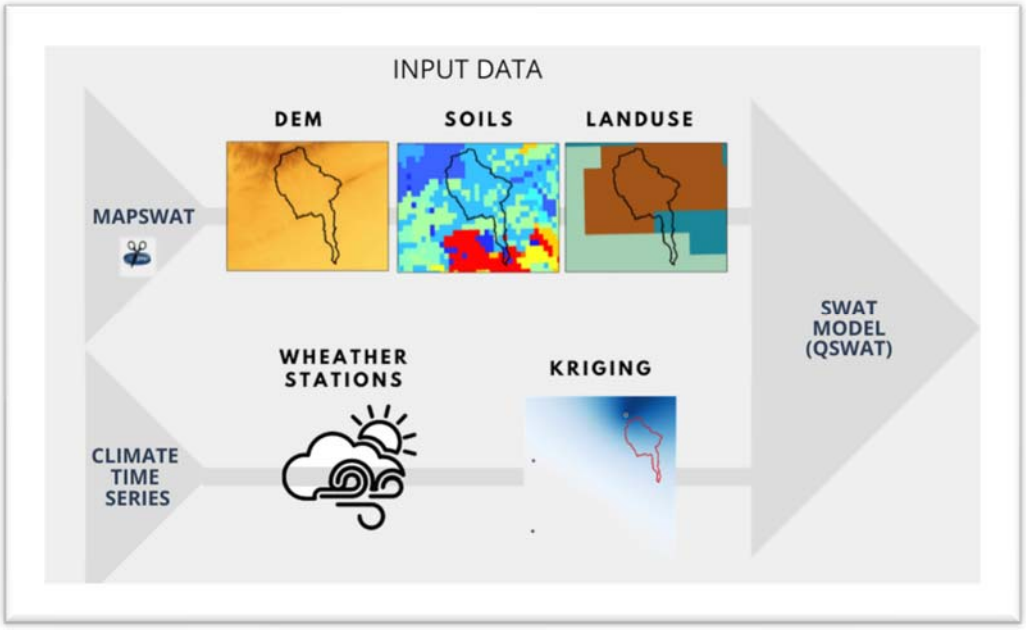


Fig. 4. Graphical overview of SWAT data preparation

Regarding the data on the lagoon itself, thanks to the buoy installed in the SMARTLAGOON project, data collection has started with the aim of using them in the developed models, but there is not yet sufficient data available to use them in the calibration of the models. Therefore, the in-situ data on water quality parameters of the Mar Menor were obtained from the Regional Government of Murcia (CARM). The CARM has promoted periodic measurements of the lagoon since 2016 when the degradation of the lagoon became evident. This dataset provides almost weekly measurements at different depths for twelve points along the Mar Menor (Fig. 5). These points are called in situ measurements points (ISMPs) and they represent the heterogeneity of the lagoon, having different characteristics such as depth, and distances to different interesting points such as sea sore and wadis. Among the data provided by CARM, it is included Chl, turbidity, Oxygen, salinity, and pH. CARM provides filtered data so no cleaner nor outlier search is needed.

Other in-site data used are those collected by management the Dirección General del Mar Menor, in compliance with the urgent measures approved by CARM, carrying out weekly environmental monitoring by means of gauging and analyses in various sections of the Rambla del Albujón, the main stream that provides runoff water to the lagoon. The records of discharge and the concentration and load of nutrients (nitrates and phosphates) of the Albujón stream is available on the Mar Menor information service website (https://canalmarmenor.carm.es/monitorizacion/monitorizacion-de-parametros/aforos/).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Fig. 5. In situ monitoring points (ISMP) carried out by Regional Government of Murcia (CARM) at Mar Menor lagoon.

In the Chl prediction task, the publicly available SRS data provided by ESA satellite products have been analysed. The European Copernicus Marine Service (CMS) that provides free, regular, and systematic authoritative information on the state of the Blue (physical), White (sea ice), and Green (biogeochemical) ocean, on a global and regional scale. Among several products that Copernicus offers through their different platforms, data from Sentinel-3 A and B were used in this study. Table 2 describes the S3 missions. Both satellites have on board the imaging spectrometer Ocean and Land Colour Instrument (OLCI) that measures solar radiation reflected by the Earth in 21 spectral bands at a spatial resolution of 300 m. It is important to note that this work focuses on these Sentinel-3-related products as they offer daily data from the same world region around the same hour (see Table 2) and having a high temporal resolution is essential for the development of a monitoring system.

Table 2. Main satellite instruments features

Satellite/Instrument	Revisit time (day)	Scanning time (UTM)*	Resolution at nadir (m)	Chl Algorithm	Start Mission (year)	End Mission (year)
Sentinel-3A OLCI	1	10:10 +- 1h	~300	Chl_nn Chl_nomc	2016	2023
Sentinel-3B OLCI	1	10:10 +- 1h	~300	Chl_nn	2018	2025

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Sentinel-based data are processed and split into products depending on the stage of the process. Thus, OLCI provides three different processing levels. This work focuses on Level-2 (L2) data that provides the water Chl-a concentration in mg/m^3. OLCI L2 product also provides additional data such as reflectances, Aerosol Optical Thickness (AOT), Diffuse Attenuation coefficient (Kd940), just to mention a few. Despite these parameters are not considered in this work, they are used by ESA algorithm to calculate the Chl-a data provided. Another important feature for OLCI products is the resolution. In this work, we use Water Full Resolution (WFR) product that provides the data related with water masses at 300 meters resolution. Regarding time granularity, the CMS download tool selects files by timeless, which indicates the period from the image was acquired until it becomes public. Particularly, we focus on NTC (Non-Time Critical) files that refer to products delivered within the next 24/48 hours after data acquisition by satellite, in contrast to NRT (Near Real Time) which refers to products delivered less than 3 hours after acquisition. NTC products contain data adjusted and collated using information from other platforms thus seems more precise to use this product in the present analysis. It is also important to note that we consider the best-case scenario for data coverage provided by SRS sources; i.e., NTC data instead of NRT. Anyway, a monitoring system could provide the data in NRT and then rewrite it with its NTC data when it becomes available.

To obtain the processed Chl-a historical data from Sentinel-3 A and B, we use CREODIAS (Malinowski et al., 2020); a platform that contains the data generated by Sentinel, Envisat, Lansat, and other EODATA satellites. Its design allows third-party users to prototype and builds their own value-added services and products. This service allows programmers to request the previous Sentinel products easily from Sentinel-hub. In this case, the API returns a set of netCDF4 files that contain parameters related to Chl-a, GPS coordinates, light reflectances, atmosphere light absorption, etc.

2.2.3 Other study areas

For the task of predicting DO and hypoxia episodes, 4 lakes in addition to Erken were selected as study areas. Since 2007, a multi-parameter YSI profiling sensor was installed in Lake Müggelsee observation stations providing hourly DO concentration measurements. And since 2015, Lake Erken improved its automated monitoring program to include a YSI profiling system that collects hourly profiles of DO concentrations. For these two lakes, the hourly surface DO concentrations have been averaged to daily values, and the minimal bottom DO concentrations in a day were used to represent the daily values.

For the sediment prediction task, data from the experimental basin of Oskotz were used. This basin is equipped with an automatic meteorological station and a hydrological station that measures streamflow and water quality parameters and takes water samples. The data from these stations are available in the experimental basins portal from the local Government of Navarra (http://cuencasagrarias.navarra.es/) and include precipitation (mm), maximum and minimum temperatures (°C), and observed streamflow and SSL data. Hydrometeorological data are available from 2002 to the present while sediment data are available from 2004. These data, obtained from observation stations in the basin, were used to construct the ML methods and the hydrological model. The SWAT hydrological model also required spatial inputs such as digital elevation model (DEM)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

data, land use, and soil data. The 25 x 25 m DEM was obtained from the National Geographic Institute (IGN) in Spain. Land-use map was extracted from Corine Land Cover (2012) with a scale of 1:100,000, and a soil map with a resolution of 1 km was implemented from the Harmonized World Soil Database (HWSD) (Nachtergaele et al., 2010b).

3. Methods

3.1 Machine learning approaches for predictions of flood

In this task, ML methods such as LSTM, GRU and CNN, and a physically based model such as SWAT have been used to model the sub-hourly flows during storm events in a headwater basin of the catchment draining the Mar Menor lagoon. There are two stages, the first is the calibration of model parameters using data from specific storm events and a second stage is the validation of these calibrated models using other storm events. The observed streamflow data needed to calibrate and validate the model were obtained using Particle Image Velocimetry methods. The scale used is sub-hourly (5 minutes) because storms typical of semi-arid climates tend to be brief but intense.

- **Generation of hourly hydrographs:**

To develop an accurate hydrological model, it is essential to measure flows during hydrological extremes as these data provide the basis for calibrating and validating the model. Collecting observed flow data under such circumstances presents a significant challenge with classical hydrometric techniques due to the high water velocities and the presence of floating objects, which not only pose a safety risk to operators, but also have the potential to damage any instrument in contact with the water. A computer vision-based in-situ monitoring system using Particle Image Velocimetry (PIV) to monitor flows at a fixed location has been employed in this task (Fig. 6). This monitoring system was installed at the outlet of the basin to determine the discharge during rainfall events. Due to the unavailability of stream gauges, the purpose of this streamflow data is to obtain observed data to calibrate sub-hourly models using SWAT and ML modelling. The system has proven to be a valuable tool in water resource management and environmental monitoring and provide continuous and accurate measurements enables decision-makers to have a comprehensive understanding of stream behavior and flow dynamics (Peña-Haro et al., 2021). In conclusion, represents a significant advancement in optical measurement systems for monitoring stream flow. Its robust and reliable operation, combined with its flexibility in different stream types and flow conditions, makes it a valuable tool for water resource management.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

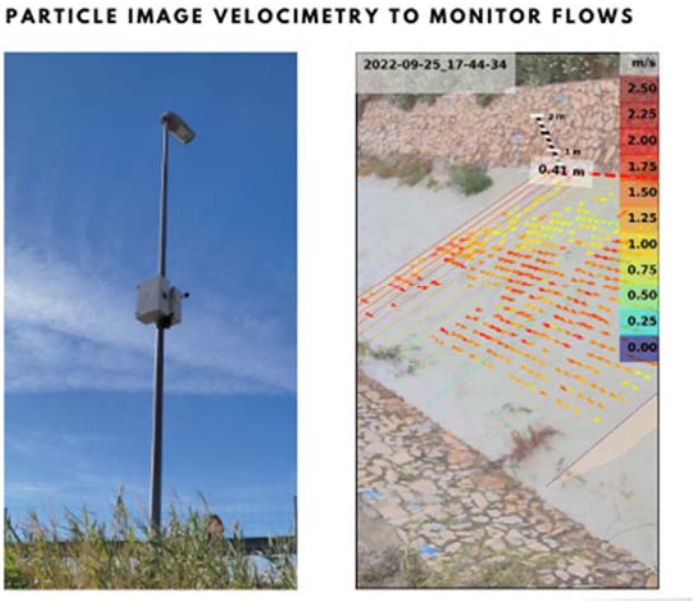


Fig. 6. PIV to monitor flows in the study area (Mosa).

Another more economical monitoring system was installed at the same point in the stream to provide support for the fixed camera. This system called FreeStation is an economical DIY environmental monitoring system (Mulligan et al., 2021). Figure 7 shows the various Freestations that have been installed in the Mar Menor basin, such as the one installed at the Spanish Institute of Oceanography (IEO), Mosa and Torre Pacheco.

Fig. 7. Freestations installed in SMARTLAGOON project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

The FreeStation and FreeSensor initiatives employ open source 3-D printing technology, open source software, and open source hardware to create and install dependable environmental data loggers and sensors at the lowest possible cost. These are intended to increase access to trustworthy, detailed local environmental data in locations that may lack the resources and know-how to do so. The designs are open source and free to use, as their names suggest, so anyone can build them.(Chan et al., 2021).

These monitoring systems (Fig. 7) measure precipitation, wind speed and direction, temperature, humidity, atmospheric pressure, and solar radiation. In addition, freestations have been installed that incorporate sensors that can measure distances positioned at various points in the Mar Menor watershed in order to determine the distance between the sensor and the water surface present in the channel in order to determine the discharge from a flow curve in the channel. Figure 8 shows the output interface of the freestations installed in our study area.

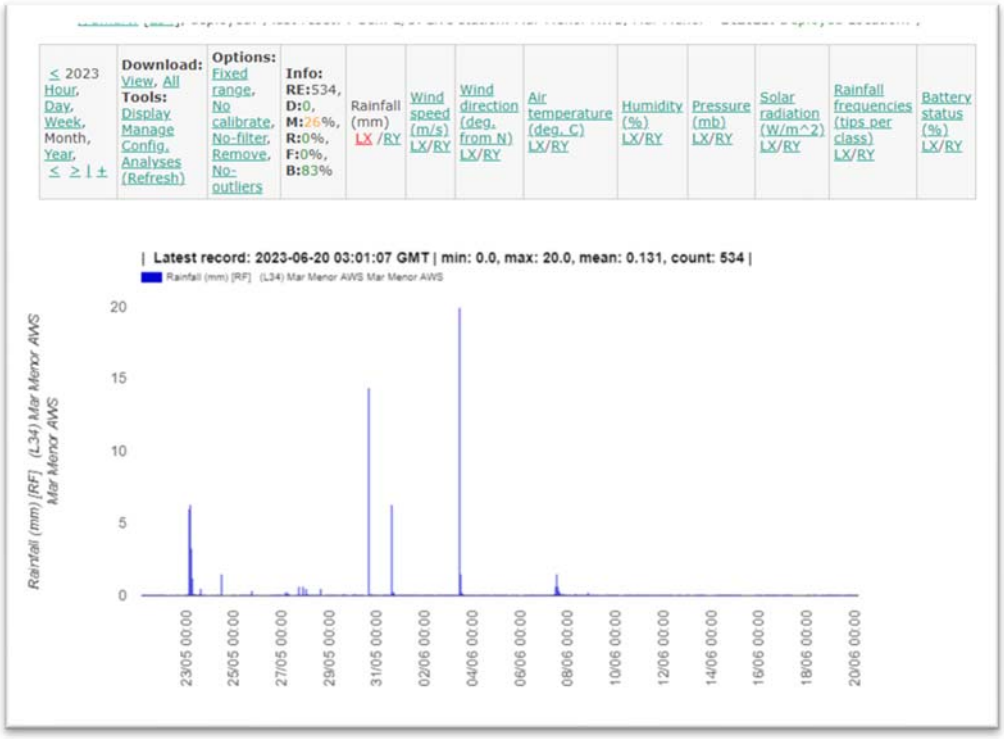


Fig. 8. Graphical output of the freestation interface for recorded precipitation in the MOSA basin.

- **SWAT model:**

SWAT is a physically based, hydrological model developed to simulate water, sediment, and agricultural chemical production in a river basin in a semi-distributed form. The SWAT model divides the basin into sub-basins, each of which is further divided into several Hydrological Response Units (HRUs), areas of land that are homogeneous and have similar responses to meteorological inputs. Each HRU is a combination of a specific soil type, land use, and slope. The hydrological part of the model simulates a catchment's hydrological cycle, based on the water balance equation, and calculates the runoff from each HRU. The curve number and Muskingum methods are employed for runoff computation and channel routine respectively.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

In this research, SWAT model has been used to simulate the flow of an ephemeral stream in a Mosa basin at the headwaters of the Mar Menor watershed. The QSWAT plugin for QGIS has been used for this purpose.

In this study, SWAT was employed to model the hydrographs of the ephemeral stream generated during storm events. Using streamflow data obtained using PIV techniques, this model has been calibrated. The validation is performed with another observed hydrograph from a different storm event than the calibration event. Until now, it has only been possible to calibrate it, so the validation is in progress due to the availability of data.

From the DEM, the stream and the different sub-basins were obtained. DEM, soil and landuse map were used to determinate the HRUs. HRUs are spatially-defined areas within a watershed that exhibit similar hydrological characteristics, such as land use, soil type, and slope. HRUs serve as the fundamental modeling unit in SWAT and play a crucial role in simulating various hydrological processes. Finally, the meteorological data required as input data for SWAT model were introduced into the model. In this way the SWAT model is ready to run its simulation.

- **Selection of inputs for ML models:**

On the one hand, the ML models were analysed univariate and on the other hand multivariate. Univariate models use a single input variable to make predictions. These models focus on obtaining the relationship between a single characteristic and the target variable. Multivariate models use multiple input variables to make predictions considering the interaction between different characteristics. The hydrographs to be modelled have a sub-hourly scale, because they are those generated by flash storms, many extreme, in an ephemeral stream. For our flood flow prediction task three variables were considered, the discharge itself, the precipitation at different instants (t, t-1, etc) and the accumulated precipitation (from 5 minutes up to two hours).

The selection of variables is an important step in the development of ML models and is an iterative process that varies depending on the problem and the specific data. Linear correlation is a useful measure to preliminarily assess the relationship between multiple variables. The linear correlation between variables has been analysed in this study using Spearman's correlation and Kendall's correlation. However, this correlation does not capture non-linear relationships so it is also important testing other variables even if their correlation is weak.

- **ML model:**

– Long Short-Term Memory (LSTM): The LSTM network is part of a class of deep learning architectures, called recurrent neural networks (RNNs), built for sequential and time series modelling (Hochreiter and Schmidhuber, 1997). The core concepts of LSTM are the cell and hidden states and its three gates (input gate, forget gate, and output gate; see Fig. 9). Essentially, the LSTM model defines a transition relationship for a hidden representation through a LSTM cell which combines the input features at each time step with the inherited information from previous time steps. This architecture is suitable for extracting information from sequential data (Rahmani et

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

al., 2021; Read et al., 2019). At its core, LSTM is designed to overcome the vanishing gradient problem faced by traditional RNNs. The vanishing gradient problem occurs when gradients diminish exponentially as they propagate back through time, making it difficult for RNNs to capture long-term dependencies. LSTM addresses this issue by incorporating a memory cell and several specialized gates.

The memory cell in an LSTM unit is responsible for storing and updating information over time. It has a self-loop, allowing it to retain information from previous time steps. The cell's state is selectively updated using various gates, which control the flow of information into, out of, and within the LSTM unit. The primary gates in an LSTM are: (i) Forget Gate: Determines which information from the previous cell state should be discarded; (ii) Input Gate: Controls the update of the cell state with new information; and (iii) Output Gate: Modulates the output of the LSTM unit based on the current cell state. The gates of an LSTM unit are trained to adaptively learn how to retain important information and discard irrelevant or noisy information.

- Gated Recurrent Unit (GRU): GRU is another type of recurrent neural network (RNN) architecture that is widely used. GRU is like LSTM network but has a simpler structure with fewer gates, making it computationally efficient and easier to train. The key components of a GRU unit are: (i) Update Gate: Controls the flow of information from the previous hidden state to the current state. It determines how much of the previous state should be retained and how much of the new information should be integrated into the current state; and (ii) Reset Gate: Determines how much of the past information should be forgotten when computing the current state. It helps the model decide how much historical information is relevant to the current time step.
- Convolutional Neural Network (CNN): a CNN is a deep learning architecture. The key characteristic of CNNs is their ability to automatically learn and extract hierarchical features from raw input data. This is achieved using convolutional layers, pooling layers, and fully connected layers. The convolutional layer applies a set of learnable filters (also known as kernels) to the input image, performing element-wise multiplications and summations to produce feature maps. These filters capture local patterns and spatial dependencies in the data. Multiple filters are used to learn different features at different locations in the image. The pooling layer downsamples the feature maps, reducing their spatial dimensions while preserving the most salient information. The most common pooling operation is max pooling, which selects the maximum value within a defined pooling window. Pooling helps in reducing the dimensionality of the data, making the network more efficient and invariant to small spatial transformations. Towards the end of the CNN architecture, one or more fully connected layers are used to combine the extracted features and make predictions. These layers have connections between all neurons, similar to those in a traditional neural network.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

3.2 Machine learning approaches for predictions of algal bloom

In the SMARTLAGOON project, the objectives related to the algal bloom prediction task are (1) to apply ML models to predict algal blooms in a well monitored lake such as Lake Erken and in a less monitored lagoon such as the Mar Menor, (2) to evaluate the performance and uncertainties of the models, and (3) to explore approaches to improve model performance and broaden model applications.

3.2.1 Modelling methods for predicting algal chlorophyll concentrations in Erken Lake

We first tested the ability of ML models in predicting algal chlorophyll (Chl) concentrations via available environmental factors, including observed lake nutrient data, and then proposed a two-step ML approach for predicting algal dynamics that first estimates lake nutrient concentrations which often have limited observations and secondly predicts variations in algal Chl using these pregenerated nutrient concentrations combined with other observed environmental factors that are collected at higher frequency. We also tested a simple hybrid model architecture that, by adding hydrodynamic features derived from the PB model into the training features of the two-step ML approach, allowed us to include additional information describing physical lake processes expected to affect variations in algal growth and succession in the machine learning prediction. We applied the above workflows to predict changing Chl concentration, as a proxy for the occurrence of algal blooms, via the gradient boost regressor (GBR) and long short-term memory network (LSTM). Two shuffling year tests were conducted. One assessed the uncertainty of ML models in predicting Chl during the same 2-year period, and the other evaluated the sensitivity of ML accuracy to various training–testing year combinations and lake nutrient sampling intervals.

- **Modelling methods:**

A Process-based (PB) hydrodynamic lake model, GOTM (General Ocean Turbulence Model) (Burchard et al., 1999), was used to generate water temperature profiles and other hydrodynamic metrics. GOTM also served as the foundation of water quality simulations made with the SELMAP-ROTBAS model (Mesman et al., 2022) that is coupled to GOTM through the Framework for Aquatic Biogeochemical Models (FABM) (Bruggeman and Bolding, 2014).

Tree models have been widely applied in modelling phytoplankton dynamics in freshwater systems (Fornarelli et al., 2013; Harris and Graham, 2017; Rousso et al., 2020). The gradient boosting regressor (GBR) is one of these tree models, iteratively generating an ensemble of estimator trees with each tree improving upon the performance of the previous. Details about the GBR model can be found in Friedman(2001). The hyperparameters in GBR are optimized via the *RandomizedSearchCV* function within the Scikit-Learn library. The loss function of model is chosen as “huber”, which is a combination of the squared error and absolute error of regression. Since the target variable in our research Chl concentration has peak values during algal blooms, which could be regarded as outliers, the huber loss function is more robust and gives greater weight to peak values than the mean squared error function.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101017861.

More details on LSTM networks can be found in section 3.1. Different combination of numbers of layers and neurons (1-3 layers, 20-200 neurons) have been tried d, but larger numbers of layers and neurons did not obviously improve the results but increased the computational time a lot, and worse results were achieved when the number of layers and neurons were decreased. Eventually, we used 3 hidden LSTM layers with 100 neurons in each layer, and each of them is followed by a dropout layer with 0.01-0.03 dropout rate for regularizing the network. The numbers of batch and epoch are set as 10 and 100, respectively. Thus, the training samples are divided into 10 batches, and the internal model parameters will update after working through one batch. And the deep learning algorithm will work through the entire training dataset 100 (epochs) times. The 'MinMaxScaler' was used to pre-process the data for generalization purposes, and 'Mean Absolute Error' was used as loss function.

Compared to the GBR model, LSTM has more complex model architectures, carrying the "memory" from the previous time steps. In this study, the GBR and LSTM were applied, respectively, to assess the performance of ML models with and without memory.

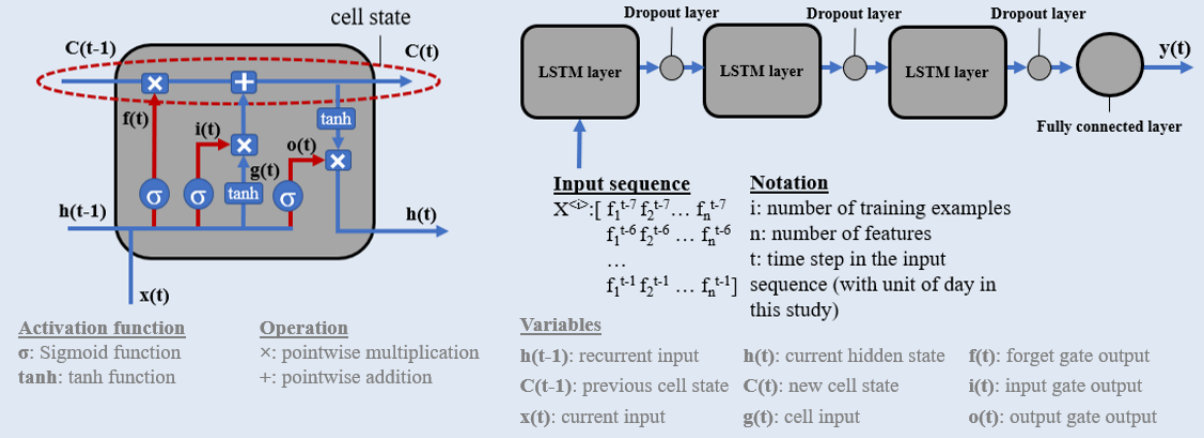


Fig. 9. Left: Detail of a LSTM cell. Right: The LSTM model architecture (based on Hochreiter and Schmidhuber, (1997)).

- **Design of predictive workflows and shuffling year data sparsity tests:**

Three workflows have been tested using a dataset split for training (years 2004–2016) and testing (years 2017–2020). In all three workflows, a 5-fold cross-validation using the training dataset was used to optimize the hyperparameters in the ML models. The training features and target variables used in each workflow are defined in Table 1.

Workflow 1 directly predicts Chl concentration based on available environmental observations. The training and testing datasets were limited by the frequency of lake nutrient observations, which resulted in 5–7 d gaps between data points. The time step of LSTM was set to 1; that is, the environmental factors on the target date and previous observation date, which may be 5–7 d ago, were used to train the model and make predictions. In workflows 2 and 3, a two-step approach was implemented. Daily measurements of physical factors were employed to pre-generate daily variations in lake nutrients using separate ML models. The ML models were trained on a daily time step using

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

the measured environmental factors and pre-generated nutrient concentrations. The LSTM time step was set to 7 days. In workflow 3, three hydrodynamic features, i.e., mixing layer depth (z_e), Wedderburn number (Wn), and seasonal thermocline depth (thermD), derived from the GOTM model, were considered as daily training features in the two-step ML approach.

Following the two-step approach and using workflow 3, we set up two tests. (1) To assess the uncertainty induced by variations in the data used to train the ML models, we shuffled the training years, randomly taking 13 years out of the 2004–2018 dataset 30 times, and tested the model predictions of Chl during 2019–2020. And, (2) to test if the workflow could be used for other water systems which may have less frequent lake nutrient monitoring data, we conducted a data sparsity test that evaluated the sensitivity of models to the lake nutrient and Chl sampling interval. For this test the lake nutrient and Chl concentration observations in the training dataset were downsampled to a 7, 14, 21, 28, and 35 d sampling interval. Then for each sampling interval using the 2004–2020 dataset, Chl was predicted for different consecutive 4-year periods when the ML models were trained by the remaining 13 years of data. Data shuffling was conducted 13 times so that every 4-year period in our dataset was tested.

Table 3. *List of training features and target variables in each workflow. Stars (*) indicate training features, circles (o) indicate target variables, and squares (□) indicate the variables are the target variables in step 1 used to daily produce a training feature for use in step 2. The order of nutrient model sequence is from the top to bottom based on its position in the table (NO_x to Si).*

Variables	Sample interval	Workflow 1	Workflow 2		Workflow 3	
			Step 1	Step 2	Step 1	Step 2
Inflow	Daily	*	*	*	*	*
Meteorological data (air temperature, wind speed, short-wave radiation, precipitation, humidity, and cloud cover)	Daily	*	*	*	*	*
ΔT	Daily	*	*	*	*	*
Ice duration	Daily	*	*	*	*	*
Days from ice-off date	Daily	*	*	*	*	*
z_e	Daily				*	*
Wn	Daily				*	*
thermD	Daily				*	*
NO_x	1–2 weeks	*	□	*	□	*
O_2	1–2 weeks	*	□	*	□	*
PO_4	1–2 weeks	*	□	*	□	*
Total P	1–2 weeks	*	□	*	□	*
NH_4	1–2 weeks	*	□	*	□	*
Si	1–2 weeks	*	□	*	□	*
Chl	1–2 weeks	o		o		o

SMARTLAGOON

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Model performance was evaluated by comparing the simulated and measured Chl concentrations and by calculating the mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (R^2). To evaluate the accuracy of the model in detecting the onset of an algal bloom, we calculated a confusion matrix in workflows 2 and 3, where the observations were linearly interpolated to daily values, and predicted daily Chl concentration were smoothed with a 7 d rolling mean. Using these data, the onset of a bloom was categorized as occurring when the daily change of Chl (ΔChl) exceeded a threshold, 0.35 mg m^{-3} d^{-1}. This works well in Lake Erken where Chl concentrations are frequently monitored (near weekly), and the linear interpolation can be expected to be reasonably representative of the Chl concentrations between measured samples. Considering the randomization in the ML models, we also add a 3 d window on the bloom onset prediction; that is, we considered the prediction of a bloom valid if the measured data suggested a bloom the day before or after the simulated onset. We used the true positive rate (TPR), false positive rate (FPR), and modified accuracy (kappa).

- **Feature selection:**

The feature selection process is based on some a priori knowledge of the underlying phenomena related to algal blooms. All workflows made use of the daily automated monitoring data. In addition, the temperature difference (ΔT) between surface water (averaged over the upper 3 m) and bottom water (15 m) was also used to represent the thermal structure of the lake, and the duration of ice cover in the previous winter and the number of days from ice-off date were used.

In workflow 2 and 3 nutrients are predicted sequentially, with each pre-generated nutrient prediction included in the training data of the next nutrient prediction (Table 3). The calculation of the hydrodynamic features for workflow 3 is described below.

The mixing layer depth (z_e) was computed using the GOTM simulated vertical eddy diffusivity (K_z) profiles, and was defined as the first depth, from the lake surface, where K_z fell below the predefined threshold value (Wilson et al., 2020), and can be describe as:

$$z_e = z_i + (K_Z^{threshold} - K_{zi})(\frac{z_{i+1}-z_i}{K_{zi+1}-K_{zi}}),$$

where z_i and K_{zi} are the depth from the lake surface, and the eddy diffusivity, respectively, in the i^{th} layer within the model. The threshold value $K_z^{threshold}$ was set to 5×10^{-5} m^2s^{-1}, based on the value described in Wüest and Lorke (2009) and Lin et al. (2021).

Unlike the dynamically varying mixing layer depth derived from the modelled K_z profiles, the calculation of the seasonal thermocline depth was estimated using Lake Analyzer (Read et al., 2011) based on the modelled temperature profile. A movement of thermocline can allow nutrient released from the sediment to enter the upper water column, leading to nutrient enrichment. It also can lead to resuspension of cells or dormant forms of cyanobacteria into the water column, encouraging bloom development (Reichwaldt and Ghadouani, 2012).

The Wedderburn number W_n is used to estimate the chance of upwelling occurring in the lake. It is described as:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

$$W_n = \frac{g' z_e^2}{u_*^2 L_s},$$

where $g' = g\frac{\Delta\rho}{\rho_h}$ is the reduced gravity due to the change in water density $\Delta\rho$ between the hypolimnion (ρ_h) and epilimnion (ρ_e). L_s is the lake fetch length (2700 m for Lake Erken) and u_* is the wind stress induced water friction velocity, defined as:

$$u_* = \sqrt{\frac{\tau_w}{\rho_e}},$$

where τ_w is the wind shear (N m^{-2}) on the water surface, computed by $\tau_w = C_D \rho_{air} U^2$. U is wind speed (m s^{-1}) measured at 10 m above the water surface. C_D is drag coefficient, given as 10^{-3} for $U < 5$ m s^{-1}, and 1.5×10^{-3} for $U >= 5$ m s^{-1}.

3.2.2 Modelling methods for predicting algal chlorophyll concentrations in Mar Menor lagoon

Until a few months ago the Mar Menor Lagoon was less monitored and therefore the availability of historical data is much more limited than in the Erken Lake. At present, with the data available, it is not possible to replicate the methodology developed in Erken Lake, so another methodology is being evaluated in the Mar Menor. As data are limited, high-resolution spatial and temporal monitoring systems are mandatory to (1) identify the threats to the lagoon, (2) understand the main problems affecting this ecosystem and (3) predict how the lagoon will behave in the future. In this task, a monitoring system based on the European remote sensing service Copernicus that allows daily monitoring of Chl for the Mar Menor lagoon is analysed. In addition, several ML and deep learning (DL) models are analysed to adapt the Chl satellite data generated by Sentinel-3 to the particular context of the shallow and highly saline Mar Menor.

The main contributions of this task are as follows: (1) Evaluating ML and DL models accuracy focused on Chl concentration in Mar Menor by using remote sensing. (2) Finding a global model for Mar Menor to allow us to estimate Chl concentration that will provide a continuous health lagoon monitoring and better dataset to Chl forecasting. The model is able to detect Chl concentrations peaks. (3) Finding a cluster based on depths.

This task focused on the Chl-a of the Mar Menor in its twelve ISMPs. Table 4 shows the main dataset descriptors. Moreover, in this work the experiments manage the data by depths, which means the models will be fitted using different depths, but never mixing them. Thus, the measures taken between 0 and 1-meter depth appear in this work as depth 0, those taken between 1 and 2 meters are depth 1, and so on.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Table 4. *Chl-a data description provided by CARM*

	Depth(m)	**Chl-a** (mg/l)
count	6146	6146
mean	2.225	2.651
std	1.66	3.865
min	0	0.022
25%	1	0.726
50%	2	1.367
75%	4	2.964
max	6	28.112

Sentinel-based products use two different algorithms to obtain Chlorophyll concentration from this raw information. We focus on the Inverse Radiative Transfer Model-Neural Network (IRTM-NN) to estimate from normalized waterleaving reflectance at OLCI bands b440 to b674, b709, θs and Δφ (namely *Chlor nn*). Unlike other chl a estimations algorithms, the S3 NN method provides data even when some reflectances provided at level-2 are negatives. Therefore, IRTM-NN algorithm can deal with these values, something that the classical algorithm does not allow and which forces negative values to be discarded. This consideration is taken to allow us to match as many days as possible between the S3 data and the in situ data in order to obtain a larger data set to train and test the models. Similarly, we consider that the twin S3 A-B reflectances provide similar reflectance information as they use the same OLCI instrument and the same processes to obtain L2 products, so we merge the data from both sources into one dataset.

It is worth mentioning IRTM-NN relies on neural networks to address the problem as a non-linear multiple regression method that drastically reduces computational time, once the network is properly trained to fit the coefficients. Fortunately, CMS already provides the IRTM-NN model that outputs water-inherent optical properties such as total backscattering coefficient (BBP443), total absorption coefficient (ATOT443), phytoplankton absorption coefficient (APH443), colored Detrital and Dissolved Material absorption coefficient (ADG443), Chlorophyll (Chl2) concentration, Total Suspended Matter (TSM) concentration. These S3 products also provide a flag value for every pixel which indicates if valid the kind of terrain: water, land or snow. More information can be found at Pahlevan et al. (2020). Anyhow, in this study all data used and associated to each ISMP are classified by S3 as water. Those invalid has been discarded. An initial analysis shows that S3 Chl values has a middle correlation with CARM in-situ data. The correlation depends on the ISMP, that means S3 does not provide reliable results for random point. Also, in the whole sentinel dataset there is not Chl-a peaks that means its model is not able to detect algal blooms that has happened several times during the study period.

- **ML models for Chl-a forecasting:**

The prediction of Chl-a is considered in this task as a regression problem using satellite data as input and in-situ observations as output. To find a specific model to fit the SRS data obtained through the Copernicus system to the Chl-a measurements obtained for the Mar Menor, the use of different ML

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

models is proposed to analyse which of them provides the best results. Five ML algorithms were selected and are described below:

- Random Forest Regression (RFR): RFR is a supervised learning technique, using a set of decision trees. To avoid overfitting, RF uses the bagging technique whereby each tree is trained with different data samples for the same problem. No single tree sees all the training data. In this way, by combining their results, some errors are compensated by others and we get a prediction that generalises better (Breiman, 2001).
- Regression decision tree (DT): DT are a sub-type of prediction trees that are applied when the response variable is continuous. In training a regression decision tree, instances are distributed through the nodes that generate the tree structure until a terminal node is reached, the condition for generating or splitting a node is usually the Residual Sum of Squares. When a new instance is to be predicted, the nodes of the tree are traversed according to the attributes of the new observation until a terminal node is reached. When the terminal node is reached, the output is the average of the output attribute of that node (Breiman, 2017).
- K-Nearest Neighbors Regressor (KNN): KNN is an instance-based method that tries to classify or predict a new instance *x* based on the k-nearest to that value. To do so, it uses different distance measures, depending on the types of attributes that make up the instance. In the case of regression prediction, the value returned is usually the average of the output attribute values of the k-nearest neighbors to the given instance. It is worth mentioning that KNN method does not require training, since the model is created at the same time as a new instance is inferred (Hackeling, 2014).
- Multi-layer Perceptron Regressor (MLP): MLP is a supervised learning algorithm that learns a non-linear function approximator. It consists of an input layer and an output layer, with one or multiple hidden layers in between. Each layer consists of neurons that are trained with the back-propagation learning algorithm. MLPs are designed to approximate any continuous function and can solve problems that are not linearly separable (Taud and Mas, 2018).
- Convolutional Neural Network (CNN): CNNs are a type of neural network that consists of a deep learning network architecture that learns directly from data, without the need for manual feature extraction. More details on CNN in section 3.1. A 1D convolutional layer has been used. The network finds the relationship between adjacent reflectances (Harbola and Coors, 2019).

- **Experimental set up:**

The dataset used in this work contains the pairing data S3-AB and ISMPs by day and coordinate. Thus, we generate several datasets by ISMP and depths containing the S3-AB reflectances and CARM Chl-a measurements. The output unit is the mg/m^3 of Chl-a, considering as inputs the remote sensing reflectances. Table 5 resumes the number of instances available for each dataset.

Table 5. *Number of instances by ISMP and depth.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

	Depth						
ISMP	0	1	2	3	4	5	6
1	87	87	87				
2	105	105	105	105	105	98	
3	108	108	108	108	108	108	42
4	112	112	112	112	112	30	4
5	102	102	102	102	102	87	
6	103	103	103	103	103	103	46
7	108	109	109	108	106		
8	103	103	103	103	102		
9	96						
10	98	99	99	99	99		
11	104	104	105	103	54		
12	101	101	101	101	101	98	

A total of four experiments have been carried out aiming to get an appropriate model. ML techniques are initially used but also CNN model can improve the results. For the experiments, algorithms are fitted using the best parameter optimization to obtain the best fitting models. The following, we summarize the steps follows to carry us to find the model: (1) The first experiment performs a 5-fold cross-validation randomly on the ML algorithm proposed. It evaluates a local model for each ISMP and depth. (2) The second experiment performs a 5-fold cross-validation randomly on ML by using all ISMPs indiscriminately to find a global model. Always, the depths are handling separately. (3) The previous experiments suggest a cluster existence. Thus we repeat the previous ones excluding ISMPs to evaluate a the model by cluster. (4) Finally we repeat all the experiments by using a CNN Conv1D.

3.3 Machine learning approaches for predictions of hypoxia

Due to the limited availability of data in the Mar Menor lagoon, the first ML experiments for hypoxia prediction have been carried out in five lakes where enough data were available for testing. Once data on DO in hypoxic episodes are available, we will be able to carry out specific tests and experiments for the Mar Menor lagoon.

In this task, we developed three ML model approaches based on two ML models, Gradient Boosting Regressor (GBR) and Long-short-term-memory (LSTM) network to simulate multi-year seasonal-scale surface and bottom DO concentrations in 5 lakes with various size and trophic levels. Except for applying GBR and LSTM directly, we designed a 2-step mixed model workflow by inputting the results from GBR model into LSTM. These ML approaches were trained with available meteorological forcing. We also used a one-dimensional (1-D) hydrodynamic model forced with the same meteorological and hydrological data to provide addition information on lake thermal structure, stratification, and ice cover that could also be included as ML model training features. Except for evaluating these ML approaches in simulating the variability of DO and detecting hypolimnetic hypoxia in the lakes, this study also aims to explore the significant factors regulating DO concentrations in each individual lake. In the following sections, the comparison against process-based models, limitations, and future applications of the ML approaches in the water management are discussed.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

- **Process-based hydrodynamic models:**

Simple one-dimensional process-based hydrodynamic models were used to estimate variables that describe lake thermal structure and mixing could also serve as training inputs to the ML models. The 1-D hydrodynamic model, GOTM (General Ocean Turbulence Model; (Burchard et al., 1999); was applied in Lake Erken (Mesman et al., 2022; Moras et al., 2019), Müggelsee, Furesø, and Ekoln) and GLM (General Lake Model (Hipsey et al., 2019) was applied in Lake Mendota (Ladwig et al., 2021)., 2021)). The meteorological and inflow data used to train the ML models were also the inputs of the hydrodynamic models (Fig. 10). We used the daily vertical profiles of simulated water temperature and eddy diffusion (Kz) obtained from the process-based models to derive daily features to train the ML models. Mixed layer depth (MLD) was defined as the first depth, from the lake surface, where Kz fell below the 5×10-5 m2s-1 (Wilson et al., 2020), and Wedderburn number (Wn) was computed based on MLD, and indicates the magnitude of upwelling. We used Lake Analyzer (Read et al., 2011) to estimate thermocline depth (thermD) and Schmidt stability (St). These parameters based on the daily temperature profiles indicate the extent of mixing, hypolimnetic thickness and the intensity of stratification, which further relating to the variability of bottom DO concentrations (Cortés et al., 2021; Foley et al., 2012; North et al., 2014).

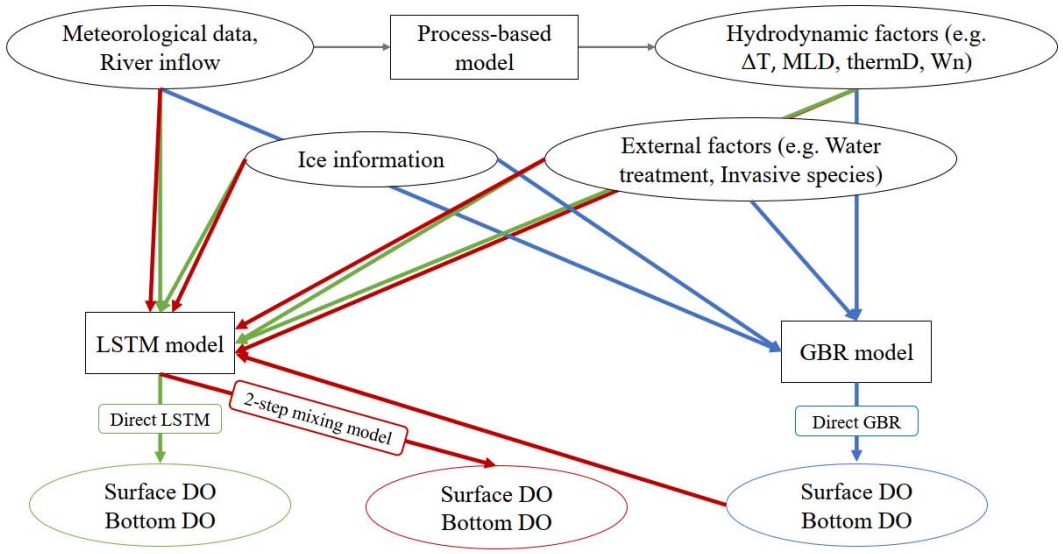


Fig. 10. Workflow of three machine learning models. Green arrows represent direct LSTM model, blue arrows represent direct GBR model, and red arrows represent 2-step mixed model.

- **Direct LSTM and GBR models:**

This study applied two ML models, LSTM and GBR, built by the Scikit-Learn (https://scikit-learn.org/stable/, last access: May, 2023) and TensorFlow (https://www.tensorflow.org/, last access: May, 2023) libraries in Python.

GBR is a type of tree models, a class of ML models that are most applied in the water resource studies, including DO prediction (Heddam and Kisi, 2018; Kisi et al., 2020). The GBR model can rank

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

the feature importance for each predictive target, illustrating the key factors which regulate the DO concentrations. More details in section 3.2.1 of this document.

LSTM belongs to recurrent neural network, built for sequential and timeseries modelling. This model architecture has achieved promising results in the harmful algal bloom simulation in Lake Erken (Lin et al., 2023). More details in section 3.2.1 of this document. There are 3 hidden LSTM layers with 50 neurons in each layer, and each of them is followed by a dropout layer with 0.2 dropout rate for regularizing the network. The numbers of batchs and epochs are set as 20 and 100, respectively. The 'MinMaxScaler' was used to pre-process the data for generalization purposes, and 'Mean Absolute Error' was used as loss function. The time step of LSTM was then set to 7 days, which means the memory of all the training features within the previous 7 days was used to train the model and predict the targets.

The direct applications of these two models involved using the training features described below (Table 6) and corresponding targets (Surface and bottom DO concentrations) along the timeseries in the training periods to train the models, and testing (validating) the model performance by inputting the features along the timeseries in the testing periods and comparing the predictive targets with the measurements (Fig. 10).

- **2-step mixed ML model workflow:**

By using both the GBR and LSTM models, the 2-step mixed ML model workflow was developed. First the GBR was used to simulate both surface and bottom DO concentrations, and then ranking the importance of features affecting surface and bottom DO, respectively. Secondly, only the features ranked as important in the GBR simulations were retained in the training of LSTM models so that the more computationally demanding LSTM training was accelerated and only considered the significant features. Also, the predictive daily values of DO concentration from GBR were added into the training dataset of LSTM (Fig. 10). Based on the results from direct GBR model, the seasonal variability of DO concentrations could be represented, so the idea is to take GBR results as the reference and the initial values into the training processes of LSTM models.

- **Training features**

The general training features used in every tested lake are daily meteorological data, river inflow data, ice information, and derived hydrodynamic factors. However, given that the physical and biogeochemical characteristics and data availability varied in each lake, each lake has its unique training features (Table 6).

The duration of ice cover period could affect the frequency, intensity and occurrence of winter mixing events, and further impact the renewal of deep water. Ice related features, i.e., ice duration and days from ice-off date, were included as training features for the lakes that were routinely ice covered (Lake Erken, Müggelsee, Mendota, Ekoln; Table 6) to indicate the extent and timing of spring mixing and deep-water renewal. The ice duration feature was converted into 4-level categorical feature according to the length of ice duration in the previous winter (Table 6). Note that we try to

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

avoid the direct usage of time-related features, since the seasonality related to time strongly affects DO and tends to overwhelm the impact from other features. Thus, the days from ice-off date feature was also converted into 7-level categorical feature according to the days before or after ice-off date (Table 6).

In Lake Mendota, zooplankton, i.e., Daphnia, density which were recorded biweekly were interpolated linearly into daily values. Original data shows the biomass of Daphnia per m3, but since the data were sparse, we also categorize Daphnia into 4 levels (Table 6).

To consider other external factors in the specific lakes, we added external training factors, to account for invasive species in Lake Müggelsee, and Mendota, and hypolimnetic aeration in Lake Furesø. These factors were set as binary number, with 1 representing period after the invasion in Lake Müggelsee and Mendota or water treatment operation on going in Lake Furesø. In addition, DO in the surface and bottom waters were predicted sequentially by the ML models, with predictive surface DO being included in the training features of the bottom DO.

Table 6. *Training features in each lake.*

Features (Data types)	Erken	Müggelsee	Furesø	Mendota	Ekoln
River discharge (metrics)	■	■	■	■	■
Air temperature (metrics)	■	■	■	■	■
Air pressure (metrics)		■	■		■
Precipitation (metrics)	■	■		■	■
Wind speed (metrics)	■	■	■	■	■
Humidity (metrics, 0-100)	■	■	■		■
Shortwave radiation (metrics)	■	■		■	■
Cloud cover (metrics, 0-1)	■	■	■		■
delT (metrics)	■	■	■	■	■
Accumulated bottom water temperature (metrics)	■				
Ice duration (category)[a]	■	■		■	■
Days from ice-off date (category)[b]	■	■		■	■
MLD (metrics)	■	■	■	■	■
Wn (metrics)	■	■	■	■	■
Schmidt Stability (metrics)	■				
ThermD (metrics)	■	■	■		■
Water treatment (binary)			■		
Invasive species (binary)		■		■	
Daphnia (category)[c]				■	
Accumulated Phosphate from river loading (metrics)				■	
Dissolved organic nutrients from river loading (metrics)				■	
Inflow temperature (metrics)				■	

a. 4 levels: ice duration over 60 days, 30-60 days, less than 30 days, and no ice duration
b. 7 levels: over 30 days before ice-off date, 30-20 days before ice-off date, 10-20 days before ice-off date, 10 days before or 10 days after ice-off date, 10-20 days after ice-off date, 20-30 days after ice-off date, over 30 days after ice-off date.
c. 4 levels: Daphnia biomass > 400/m^3, 200-400/m^3, 50-200/m^3, <50/m^3.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

- **Model evaluation**

To assess the uncertainty induced by variations in the training data, we randomly removed 2 years data out of the whole training periods 30 times and tested the model performances in the fixed testing periods (Table 7). The results of these 30 times model runs were aggregated to assess the model performance in each lake.

To further evaluate model performance in detecting hypoxia, we define the hypolimnetic hypoxia when bottom DO concentrations decreased below the specific thresholds. Given that the restoration actions have been taken in Lake Furesø since 2003, the hypoxia phenomena in hypolimnion have been reduced. Also, the sample interval of hypolimnetic DO in Lake Ekoln is over a month, not sufficient to interpolate the exact timing of anoxia. Thus, we only used Lake Erken, Müggelsee, and Mendota, to evaluate model performance in detecting hypoxia events. DO < 2 mg/L was used as the criterion for hypoxia in Lake Mendota which experience serious eutrophication and anoxia in the seasonal stratified period and DO < 3 mg/L was used as the criterion in Lake Erken and Müggelsee (Nürnberg et al., 2013; Scavia et al., 2014).

Table 7. Data available for the OD prediction task

Data	Erken	Müggelsee	Furesø	Mendota	Ekoln
Averaged DO sampling interval during ice-free period (days)	8 (2004-2014) 1 (2015-2020)	1	18	16	31
Training period	2004-2016	2004-2016	1990-2009	1999-2009	1987-2008
Testing period	2017-2020	2017-2020	2010-2017	2010-2015	2009-2019

3.4 Machine learning approaches for predictions for other water quality variables

3.4.1 Predictions of suspended sediment load

Modeling the specific processes that occur in a basin is highly complex, and no model works perfectly in all basins. Identifying models that accurately simulate the complexity of basin processes using available data is a challenge for decision-makers in basin management (Nguyen et al., 2019). This study individually tests two ML methods, M5P and RF, as alternatives to the SWAT model to estimate SSL in the Oskotz river basin. The main objective of this task is to determine efficient models for SSL estimation by comparing the results obtained using ML models with those obtained using the physically based SWAT model.

- ***Hydrological model (SWAT)***

SWAT calculates the sediment yield for each HRU using the Modified Universal Soil Loss Equation (MUSLE) (Williams, 1975), which predicts erosion as a function of a runoff factor representing the energy used in the detachment and transport of sediment (Neitsch et al., 2009). SWAT computes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

water and sediment yield for each HRU individually and aggregates them at the sub-basin level. Details of this model in section 3.1.

The Oskotz river basin was subdivided into four sub-basins and homogenous sections, resulting in 53 HRUs. The Hargreaves-Samani approach was chosen to estimate potential evapotranspiration. The SUFI-2 algorithm was used to automatically calibrate the SWAT model parameters in SWAT-CUP (Abbaspour et al., 2007). Since sediment transport is dependent on runoff, a sequential calibration approach was applied (recommended by Arnold et al., (2015)), in which streamflow generation parameters were calibrated first, followed by sediment parameters. One thousand simulations were performed twice, and the parameters were readjusted after the second iteration. The selected parameters were calibrated using daily time steps and adjusted using the Nash-Sutcliffe efficiency (NSE) as an objective function to ensure that the simulation results were as close as possible to the streamflow and SSL observations. The periods 2002–2012 (11 years) and 2013–2020 (8 years) were used to calibrate and validate the streamflow, respectively. In the case of SSL, 2004–2012 (9 years) was the calibration period and 2013–2020 (8 years) was the validation period. In both calibrations, three years were used as a warm-up period.

- **Machine learning algorithms**

– M5P: The M5P technique (Wang and Witten, 1997) is a remodeling of Quinlan's M5 (Quinlan, 1992) for induction trees in regression models. This technique combines a traditional decision tree with the possibility of performing linear regression functions at the nodes. First, an induction decision tree is constructed, applying the splitting criterion, which minimizes the variance of a subset of class values at each branch, at each node. This process is stopped if the values of each branch vary slightly or there are a minimum number of instances at the node. Secondly, a pruning process is performed, in which a regression function converts the internal nodes into a leaf node. Finally, to avoid discontinuities, a smoothing process is applied that combines the leaf model prediction with each node encountered on the way to the root node.

– Random forest (RF): RF (Breiman, 2001) is defined as an ensemble based on decision trees. Decision trees have the advantage of good interpretability in both the constructed model and inference. However, they have the disadvantage of bias and variance problems. These complications are resolved using the ensemble to merge and combine information from the decision trees. On the one hand, the data variability and amount of stored information are increased. On the other hand, the interpretability of the constructed model is maintained, although in a more complex manner. In general, the RF ensemble has the following characteristics:

  - Given a dataset of |N| samples to construct each tree a, |N| cases are randomly selected using replacement as a training dataset. The process of sampling with replacement is called bootstrapping. One third of the data is excluded from training and used for testing (Schonlau and Zou, 2020). These data are known as out-of-bag (OOB) samples. Each tree has an OOBa set with which it is tested. The testing result provides a weighting for each tree used in the combination of the information.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

- To choose the splitting decision for a node, the M attributes to be studied are chosen at random from the M input variables at each tree node.
- Each tree grows to its maximum possible extent and can be configured so that no pruning is required. New instances are predicted by aggregating the predictions of the A trees (i.e., a majority vote for classification, an average for regression).

- **ML model inputs:**

Concerning ML model inputs, streamflow and precipitation were selected to estimate SSL according to previous studies (Cobaner et al., 2009; Kumar et al., 2016; Singh et al., 2014). The precipitation and streamflow variables are correlated but complementary since precipitation transports sediment in the drainage basin and streamflow regulates concentrations and downstream transport (Sirabahenda et al., 2020). Correlation coefficients (CCs) were calculated to analyze the dependence between the observed sediment data and the flow and precipitation data on the same and previous days (Table 8).

Table 8. *The correlation coefficients (CC) between the input variables and SSLt.*

Variable	Q_t	Q_{t-1}	Q_{t-2}	Q_{t-3}	Q_{t-4}	Q_{t-5}	P_t	P_{t-1}	P_{t-2}	P_{t-3}	P_{t-4}	P_{t-5}
CC	0.64	0.34	0.18	0.16	0.15	0.11	0.40	0.36	0.17	0.12	0.12	0.11

The SSL for day t (SSL_t) was strongly correlated with the streamflow on that day (Q_t), with a CC of 0.64. The CC for SSL_t and the precipitation for day t (P_t) had a value of 0.40. From time t-1, the correlations between the Q and P variables and SSL (t) decreased considerably, with values below 0.2. Therefore, to estimate daily SSL, several input combinations were constructed and tested (Table 9), including the daily streamflow and precipitation of the current day t and previous days.

Table 9*. Input scenarios for the ML models.*

Scenario	Model inputs
I	Q_t
II	Q_t, P_t
III	Q_t, P_t, P_{t-1}
IV	Q_t, Q_{t-1}
V	Q_t, P_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, P_{t-1}, P_{t-2}
VI	Q_t, P_t, Q_{t-1}
VII	Q_t, P_t, Q_{t-1}, Q_{t-2}
VIII	Q_t, P_t, Q_{t-1}, Q_{t-2}, P_{t-1}, P_{t-2}
IX	Q_t, P_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, P_{t-1}, P_{t-2}, P_{t-3}

The first scenario considered the streamflow for day t (i.e. the variable most strongly correlated with SSL_t) as the only predictor variable. Scenario II considered the two variables most correlated with sediment (Q_t and P_t). Scenario III was similar to scenario two but with the addition of the third most correlated variable (P_{t-1}). Scenario IV considered only the hydrological variables on the same day and the previous day (Q_t and Q_{t-1}). In addition, scenario V included all variables with a CC greater than 0.15. The remaining scenarios were other combinations of these variables. The advantage of

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

employing only streamflow and precipitation as inputs is that the collection and availability of these data are often easier than other data in many basins. Other studies (Kumar et al., 2016; Sihag et al., 2021; Zounemat-Kermani et al., 2020) have also used sediment from previous days as inputs with the disadvantage that this type of data is measured with less frequency and is, therefore, more difficult to collect. Therefore, nine scenarios were tested for each ML model. The training period was from September 2004 to December 2012, and the test period was from January 2013 to December 2020.

3.5 Model performance metrics

The performance of the models was evaluated using the statistics defined in table 10, indicators calculated from measured and simulated data.

Table 10. *Performance metrics.*

Performance Metric	Equation	Range
Coefficient of determination (R^2)	$\frac{[\sum_{i=1}^{n}(O_i-\bar{O})\cdot(E_i-\bar{E})]^2}{[[\sum_{i=1}^{n}(O_i-\bar{O})^2]^{0.5}\cdot[\sum_{i=1}^{n}(E_i-\bar{E})^2]^{0.5}]^2}$	[0, 1]
Mean squared error (MSE)	$\frac{\sum_{i=1}^{n}(O_i-E_i)^2}{n}$	[0, ∞]
Root mean squared error (RMSE)	$\sqrt{\frac{\sum_{i=1}^{n}(O_i-E_i)^2}{n}}$	[0, ∞]
Mean absolute error (MAE)	$\frac{\sum_{i=1}^{n}\lvert O_i-E_i\rvert}{n}$	[0, ∞]
Percent bias (PBIAS)	$\frac{\sum_{i=1}^{n}(O_i-E_i)\cdot 100}{\sum_{i=1}^{n}(O_i)}$	[−∞, ∞]
Root mean square error observations standard deviation ratio (RSR)	$\sqrt{\frac{\sum_{i=1}^{n}(O_i-E_i)^2}{\sum_{i=1}^{n}(O_i-\bar{E})^2}}$	[0, ∞]
Nash-Sutcliffe efficiency (NSE)	$1-\frac{\sum_{i=1}^{n}(O_i-E_i)^2}{\sum_{i=1}^{n}(O_i-\bar{O})^2}$	[−∞, 1]

O_i is the *ith* observed data, $\bar{O}$ is the mean of the observed data, E_i is the *ith* estimated data, $\bar{E}$ is the mean of the estimated data and n is the total number of observations.

R^2 is the proportion of the total variance of the variable explained by the regression. This measure reflects the well-fitting of a model to the variable it intends to explain, the closer to the value 1, the better the fit of the model. The MSE is a measure derived from the square of the Euclidean distance, it is always a positive value that decreases as the error approaches zero. RMSE is a measure that indicates the average magnitude of errors between the observed values and the predicted values in a regression model and is expressed in the same units as the target variable. The lower the RMSE value, the better the fit of the model to the observed data.

The RSR uses the standard deviation of the observations to standardize the mean square error. The variance between the observed and simulated data is quantified using NSE. The PBIAS indicates the average tendency of the simulated data to be higher (negative values) or lower (positive values) than the observed data.

The MAE is used to quantify the accuracy of a prediction technique by comparing the predicted versus observed values. Since the models were tested in the lakes with various trophic level, the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

MAE were normalized ($NMAE$) by the range of observed concentrations to be used to conduct inter-lake comparison:

$$NMAE = MAE/(\max(y) - \min(y))$$

In the Chl prediction task in Lake Erken, the confusion matrix and kappa statistic were used to identify the potential of the ML models to correctly capture the algal bloom onset (Table 11). The kappa statistic uses the information from the confusion matrix to calculate a coefficient that measures the agreement between observed and model-predicted classifications, considering the expected agreement by chance. A model with 100% TPR, 0% FPR and 100% kappa would be a perfect fit.

Table 11. *Confusion matrix and metrics based on it for bloom onset.*

	Modeled onset	**Modeled no onset**
Observed onset	True Positive (TP): Model predicted the bloom onset when there was an onset	True Negative (TN): Model predicted no bloom onset when there was no onset.
Observed on onset	False Positive (FP): Model predicted the bloom onset when there was no onset	False Negative (FN): Model did not predict bloom onset when in fact there was an onset
True positive rate (TPR) = TP / (TP+FN); What proportion of all events were correctly detected **False positive rate (FPR)** = FP / (TN+FP); What proportion of no events were incorrectly defined as bloom onset **Kappa** = $(P_o-P_e)/(1-P_e)$; The modified accuracy that considers the possibility of the agreement occurring by chance. P_o = (TP + TN)/(TP+TN+FP+FN); Actual accuracy P_e = ((TP+FP)/(TP+TN+FP+FN) * (FN+TN)/(TP+TN+FP+FN)) + ((TP+FN)/(TP+TN+FP+FN) * (FP+TN)/ (TP+TN+FP+FN)); Chance agreement		

TPR and FPR were also used to identify the potential of ML models to accurately capture hypoxia (see Table 12).

Table 12. *Confusion matrix and metrics based on it for hypoxia event .*

	Modeled onset	Modeled no onset
Observed onset	True Positive (TP): Model predicted hypoxia when hypoxia happened	True Negative (TN): Model predicted no hypoxia when hypoxia happened
Observed on onset	False Positive (FP): Model predicted hypoxia when hypoxia happened when hypoxia did not happen	False Negative (FN): Model did not predict hypoxia when in fact hypoxia happened
True positive rate (TPR) = TP / (TP+FN); What proportion of all events were correctly detected		
False positive rate (FPR) = FP / (TN+FP); What proportion of no events were incorrectly defined as hypoxia event		

SMARTLAGOON

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

4. Results and Discussion

4.1 Flood predictions

SWAT and ML models were used to find the relationship between precipitation and runoff in the Mosa basin (Headwaters of the Mar Menor basin). These models simulated the streamflow hydrograph based on storm events. The evaluation of the ML models is determined from the errors between the predicted value and the actual value and compared with the results obtained with the SWAT model.

At the time of writing this document only one flow hydrograph event was available and was used for calibration. But we are currently collecting more information and this task is still in progress. When we have a set of hydrographs, they will be randomly selected for calibration and the rest for validation and this document will be updated. The hydrograph available to date was that generated by the storm event of 25 September 2022. In that event a total of 70 mm was recorded in 3 hours of which 66 mm fell in one hour (Fig. 11).

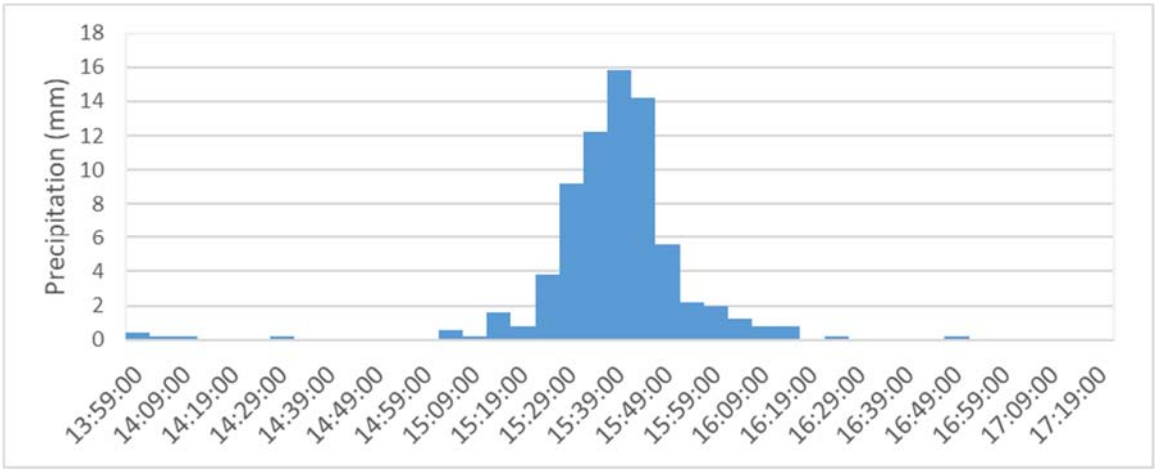


Fig. 11. The hietogram of the 25 September 2022 storm event. Data collected at the "El Relojero" rain gauge.

- **Hydrological SWAT model:**

Once the SWAT model was built with all the input data, the model was run and preliminary uncalibrated results were obtained. The flow hydrograph obtained with SWAT for the recorded storm event (Fig. 11) is shown in Figure 12. Here it can be seen that the model simulates peak flow close to 18 m^3/s at 15:39 PM, from that time onwards the hydrograph begins to descend progressively until it ends at approximately 20:00, with a duration of approximately 4-5 hours. This Figure 12 also shows the actual or observed streamflow, that is, the results obtained by the fixed camera using the PIV techniques, represented by the orange line. Compared with the hydrograph simulated by SWAT, it is worth noting that both the peak and the duration of the hydrograph are similar in time and volume, although the peak of the observed hydrograph is more accentuated,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

reaching 24 m^3/s. The overall initial performance is good although the highest volume differences are observed between the flows collected between 16:30 and 18:00 hours. These two small peaks, increases in discharge, are not due to an increase in precipitation, so the SWAT model is not able to simulate them.

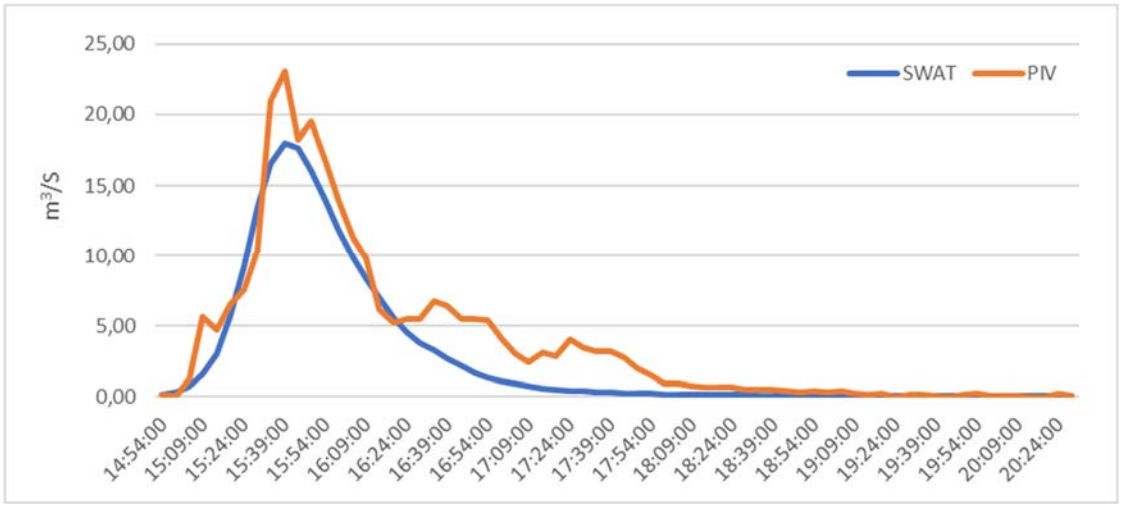


Fig. 12. Observed versus uncalibrated simulated flow hydrograph with SWAT.

After simulating the model, a manual calibration process was carried out in order to adjust the simulated peak flow under (Abbaspour et al., 2015). For this purpose, the AWC parameter has been raised by 9% of its initial estimated value, the CN2 parameter has been lowered by 2% and the ESCO parameter has been decreased from its initial value of 0.95 to 0.75. In this way we obtain the following hydrograph (Fig. 13) in which it can be seen that the peak of the hydrograph practically reaches the peak of the observed flow, achieving a good adjustment of the simulated hydrograph compared to the observed one (PIV). It has not been possible to adjust more precisely for disturbances caused by the flow observed between 16.30 and 18.00 hours.

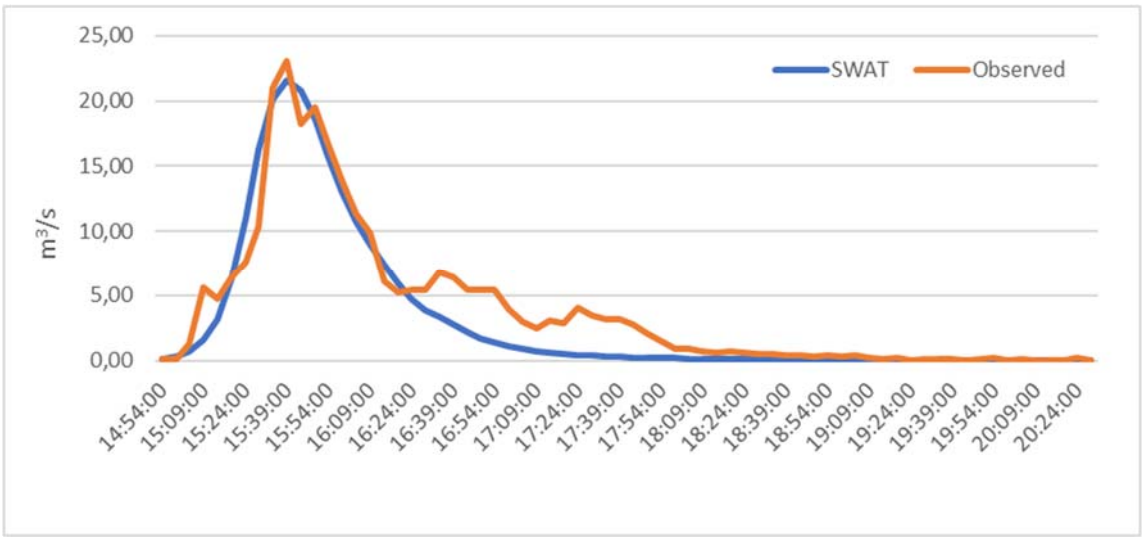


Fig. 13. Observed versus uncalibrated simulated flow hydrograph with SWAT.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Table 13 shows the statistics of the SWAT model after calibration. The performance is very good despite the challenge of simulating a storm event on such a small time scale (5 minutes). Therefore, the SWAT model could be considered valid for short and intense rainfall events, as is the case for those in the Mar Menor catchment area.

Table 13. *Performance of the SWAT model after calibration.*

Statistical	Value
R^2	0.92
NSE	0.89
PBIAS (%)	21.09
KGE	0.78

- **Univariate ML models:**

The previous step to the construction of the ML models was to analyse the input variables and thus select the inputs. In the univariate analysis, the target variable, streamflow, was analysed. The discharge was tested by plotting the autocorrelation factors of the historical observations (Fig. 14). These graph allowed us to identify a suitable lag period for the prediction of discharge in the study basin (Hussain and Khan, 2020).

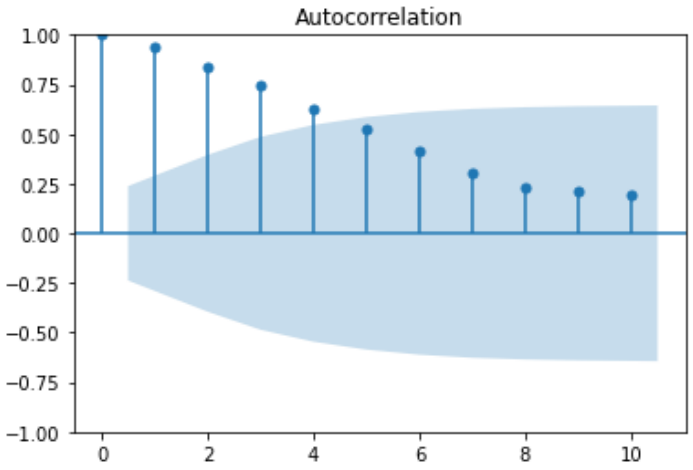


Fig. 14. Autocorrelation function (ACF) for streamflow time series of the Mosa river basin.

The coefficients up to a lag of 5 are greater than 0.5 and have a 95% probability of having a certain impact on the target variable. Therefore, to predict the streamflow at time t, the five previous flow values were selected, that is, a time lag of 1 to 5 days.

Next, the different deep learning models were constructed and trained using *TensorFlow* and *Keras* library (Python): LSTM, CNN and GRU. Different network configurations with different parameters were tested. The configuration selected was the one with the lowest MSE value. The ML models were constructed so that the input layer of each model expected sequences of length 5, representing historical discharge data. The final layer of each model (output layer) was a dense layer with 1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

neuron and linear activation function. This layer produces the output prediction for the flow rate of intant *t*. The input and output layer were common in all three models. What changes for each model were the intermediate or hidden layers. The models were compiled with a MSE loss function, Adam optimizer with a learning rate of 0.001, and RMSE as the evaluation metric. Adam is an adaptive learning rate optimization algorithm that combines the benefits of both AdaGrad and RMSProp. It dynamically adjusts the learning rate during training to optimize the model's performance. Once the model is compiled, the model is trained using the provided training data (inputs and outputs). The training process involves iteratively adjusting the model's parameters over a specified number of epochs (300 in these cases). During training, the model updates its internal weights and biases based on the optimization algorithm (Adam) and the loss function specified during compilation.

The best performing LSTM model consisted of an LSTM layer after the input layer with 12 neurons. This layer serves as the first hidden layer of the model and is responsible for capturing temporal dependencies in the input data. Next, a dense layer with 10 neurons and a rectified linear unit (ReLU) activation function was added to the model. This layer acts as the second hidden layer, introducing non-linearity and higher-level representations.

The optimal CNN network consisted of a convolutional layer after the input layer. A Conv1D layer is added to the model with 10 filters and a kernel size of 2. This layer applies a one-dimensional convolution operation on the input data, extracting local patterns or features. The number of filters represents the number of distinct features the model can learn from the input data.

The following, a flatten layer is inserted to convert the output of the convolutional layer into a one-dimensional tensor, which can be fed into the subsequent fully connected layers. Two Dense layers are added to the model. The first Dense layer consists of 10 neurons and applies the ReLU activation function. The second Dense layer is the output layer with a single neuron.

The best configuration of the GRU model was with a GRU layer after input layer. The GRU layer was added to the model with 20 units. The GRU is a type of RNN that can capture long-term dependencies in sequential data. It uses gating mechanisms to control the flow of information within the network, allowing it to selectively remember or forget information from previous time steps. Following the GRU layer, there were two Dense layers. The first Dense layer consisted of 10 neurons and applies the ReLU activation function. The second Dense layer is the output layer.

In terms of computational training time, there are no significant differences between the three models due to the small amount of data used for training. Table 14 shows the performance statistics of the three models. All three models obtained very similar performance results. The results show that the models fitted very acceptably. However, these data could not be validated because more data were not yet available. Therefore, the work is still in progress. The performance of the three univariate ML models was better than that of SWAT for all statistics. The most considerable improvement was in the PBIAS statistic. This large difference is since the ML models were able to simulate the two small peaks between 16:30 and 18:00 (see Fig. 15) which SWAT was not able to do. Graphically, the three simulated hydrographs show a small lag, of a few minutes, with the observed hydrograph, reaching the peaks later. LSTM and GRU had a similar behaviour, both staying below

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

the peak of the observed hydrograph. CNN obtained a peak closer to the observed value, but with the delay mentioned above.

Table 14. *Performance of univariate ML models*

Statistical	LSTM	CNN	GRU
R^2	0.96	0.92	0.97
NSE	0.95	0.92	0.96
PBIAS (%)	1.92	0.78	1.36
MSE	1.46	2.37	1.09
RMSE	1.21	1.54	1.04
MAE	0.63	0.78	0.60

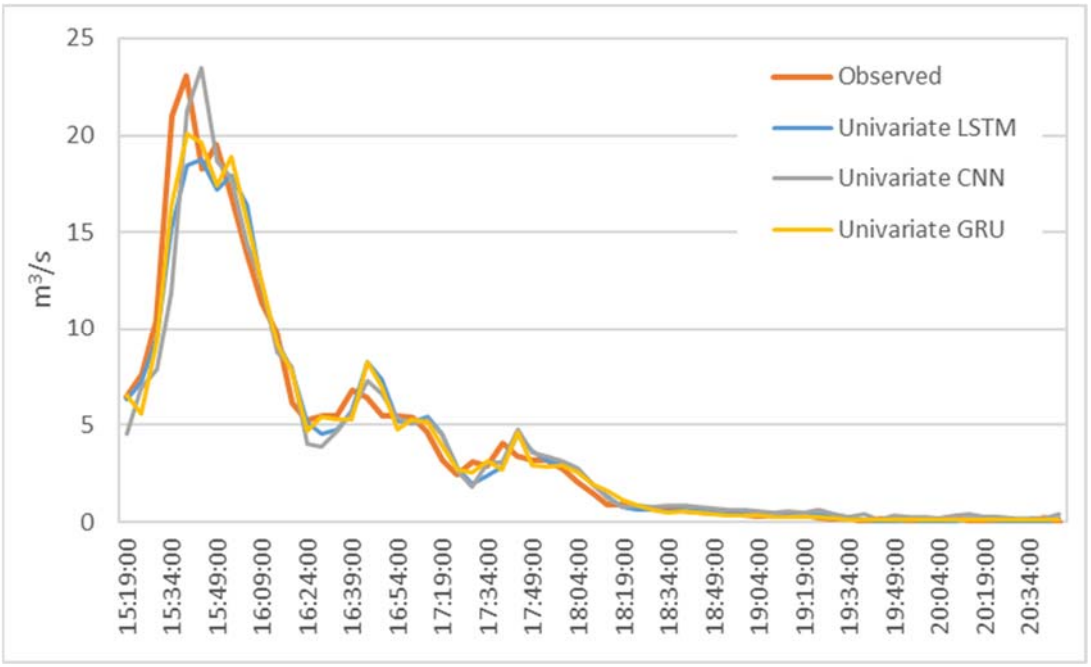


Fig. 15. Observed versus uncalibrated simulated flow hydrograph with univariate ML models.

- **Multivariate ML models:**

Regarding multivariate modelling, the relationship between flow and rainfall at different instants and cumulative rainfall at different intervals was analysed. First, the Shapiro-Wilk test and D'Agostino's K-squared test were utilized, yielding p-values below the predefined significance level of 0.05. These findings strongly support the rejection of the null hypothesis of normality for the analysed variables. Consequently, to gain a comprehensive understanding of their relationship, the correlation between the variables was assessed using Spearman's rho coefficient. The Spearman's rho results showed that the precipitation at times t-13 (65 minutes before), t-14 (70 minutes before) and t-15 (75 minutes before) have correlation values greater than 0.70 with the discharge at instant t. In the case of accumulated rainfall, the accumulated value at 2 hours obtained a higher correlation with the discharge with a value of 0.85. All these precipitation variables together with the discharge

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

in the five previous instants (used in the univariate case) were finally used as inputs to predict discharge. The data were standardized to eliminate the influence of units and the three ML models were trained. The structures of the best performing models are shown in Figure 16. The input layer of all models provides the input shape for the subsequent layers with a window size of 1 and with 9 variables. For the first model, an LSTM layer, the first hidden layer, was added with 10 neurons and followed by a dense layer with 10 neurons which ReLu activation function prior to the output layer. The second model was a CNN with a 1D convolutional layer with 12 neurons, a flatten layer and a dense pre-output layer with 5 neurons. The third model consists of a GRU layer downstream of the input layer with 10 neurons followed by a dense layer with 5 neurons. The performance of the three models is shown in table 15. The difference between the models appears to be minimal. Figure 16 also shows similar results graphically.

The results were generally satisfactory when introducing the exogenous variable of precipitation. The improvement with respect to the univariate case can be seen in the statistics but is particularly apparent graphically. Figure 16 shows how all models were able to reach the peak of the hydrograph when considering the precipitation variable as an input. As in the univariate case, LSTM and GRU reproduced the flow in a very similar way.

Table 15. *Performance of univariate ML models*

Statistical	LSTM	CNN	GRU
R^2	0.98	0.98	0.98
NSE	0.98	0.98	0.98
PBIAS (%)	0.32	0.29	0.47
MSE	0.40	0.39	0.49
RMSE	0.63	0.63	0.70
MAE	0.41	0.39	0.43

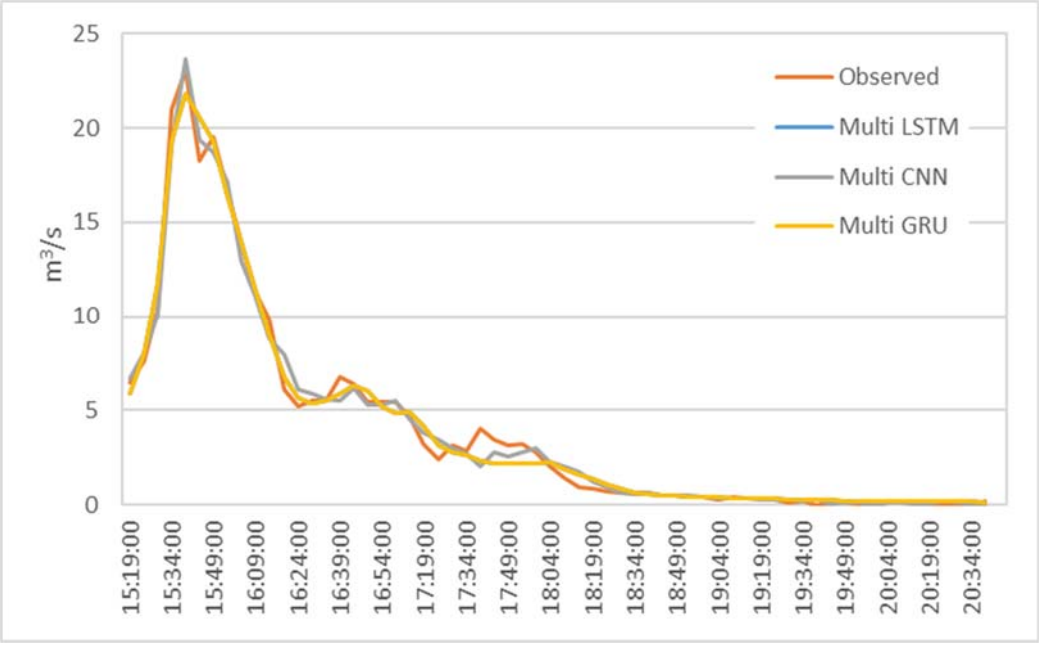


Fig. 16. Observed versus uncalibrated simulated flow hydrograph with multivariate ML models.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

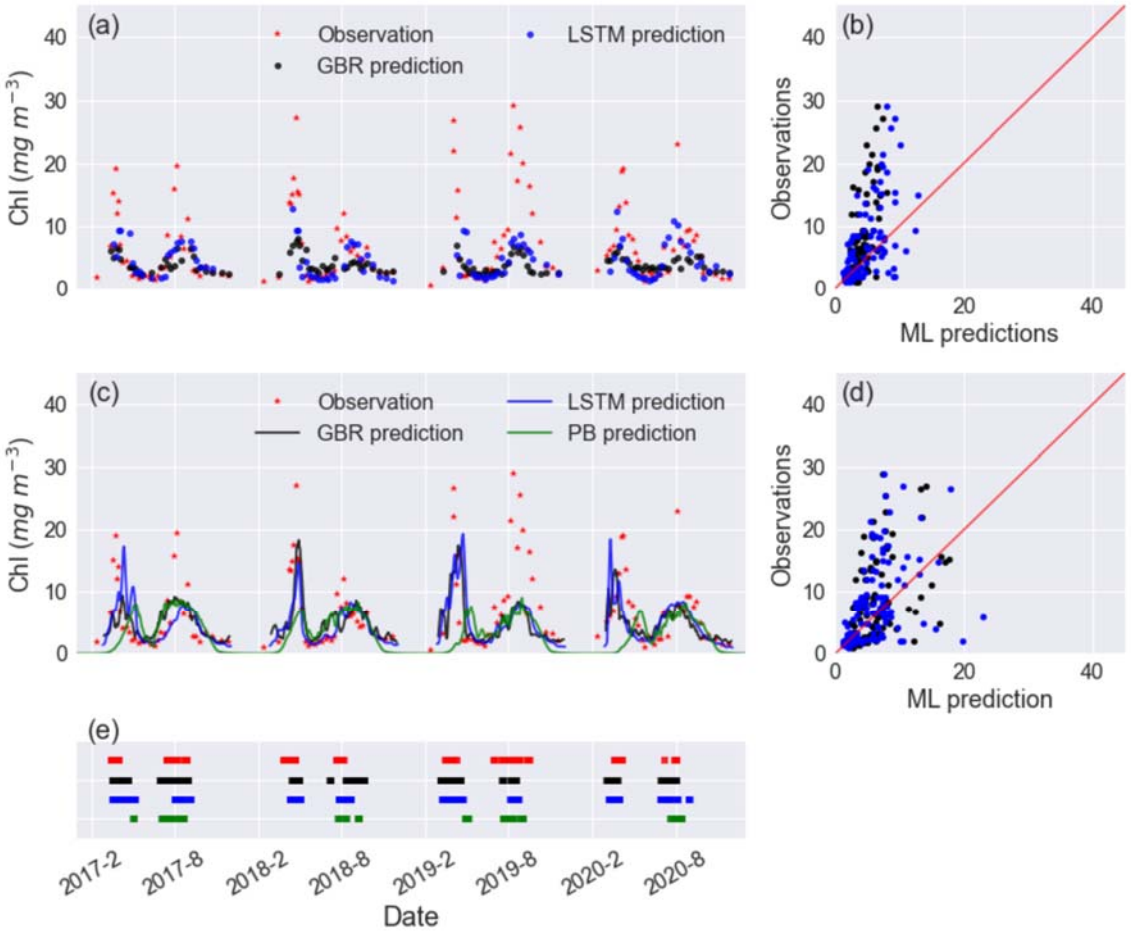
4.2 Algal bloom predictions

4.2.1 Predictions of algal bloom in Lake Erken

The results of each workflow, the results of Effects of shuffling training years on 2019-2020 predictions and the results of Shuffling year data sparsity test are presented following.

- **Performance of ML and hybrid PB-ML models:**

In workflow 1, which performs direct prediction based on observations, both the GBR and LSTM clearly reproduced spring and summer blooms (Fig. 17a) but underestimated the intensity of blooms (Fig. 17a, b). Neither ML model captured the extraordinarily high Chl (~15–30 mg m^{-3}) in the summer of 2019. Although the abnormal summer bloom in 2019 could contribute to the higher RMSE and MAE in the testing dataset than the mean values in the training dataset, the cross-validation on the training dataset (see Table 16) shows what appears possibly to be an overfitting issue in both models. The achieved accuracy of models is attributed to the daily availability of physical inputs and the fact that in Lake Erken water samples are collected frequently at 5–7 d intervals. Workflow 1 may be most valuable in reconstructing previous variations in algal Chl, filling the gaps between measured Chl observations and feature importance ranking (see Fig. 18). But when using this workflow, future forecasts will be limited by the absence of future nutrient data.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Fig. 17. Time series of observed and predicted Chl from GBR and LSTM models in (a) workflow 1 and (c) workflow 3, and the corresponding scatter plots of observations vs ML predictions of Chl in workflow 1 and workflow 3 are shown in panels (b) and (d), with the black and blue dots and lines representing the predictions from GBR and LSTM, respectively. Panel (e) shows the observed and predicted algal bloom onsets in 2017–2020 using the same colour coding as the previous panels. Results from the PB model simulation in Mesman et al. (2022) are also shown in (c) and (e).

Table 16. *Comparisons of ML models' performance based on RMSE, MAE, and R2 in training dataset (via 5-fold cross validation) and testing dataset.*

Workflow	GBR			LSTM		
	MAE	RMSE	R2	MAE	RMSE	R2
1 (training)	2.86	4.30	0.18	2.66	4.38	0.31
1 (testing)	3.55	5.77	0.13	3.58	5.64	0.20
2 (training)	2.78	4.07	0.33	2.71	4.73	0.31
2 (testing)	4.22	6.27	0.05	3.87	6.00	0.13
3 (training)	2.79	4.10	0.32	2.64	4.51	0.40
3 (testing)	3.99	5.94	0.14	3.71	5.81	0.18

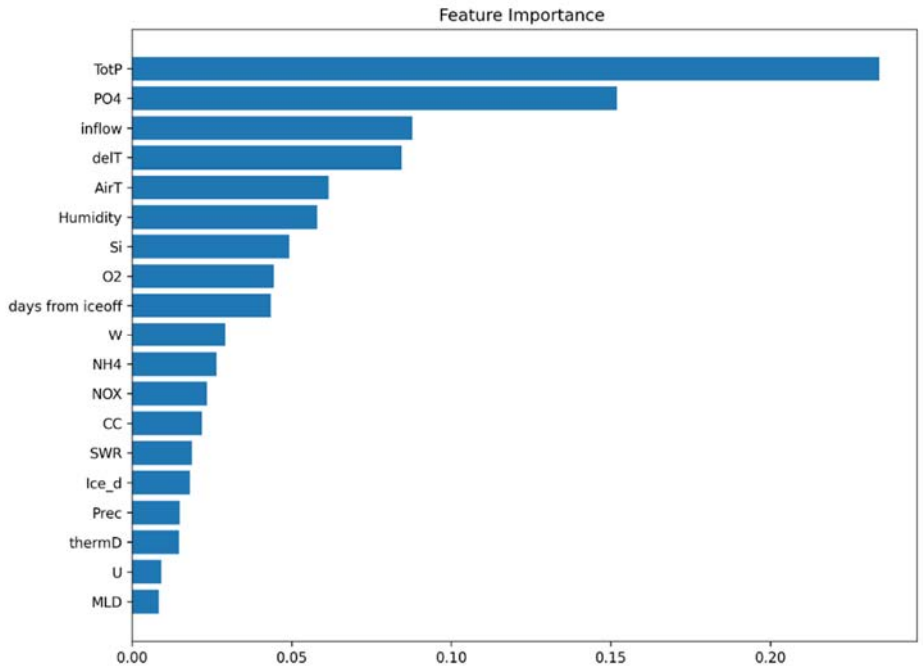


Fig. 18. *Feature importance based on 'feature_importances' function from GBR model in scenario 1.*

Workflow 2 consisted of two-step ML models based on previously generated daily nutrients and observed physical factors. As in workflow 1, both ML models in workflow 2 had poor fit in the summer of 2019 and suffered from overfitting leading to higher MAE and RMSE and lower R2 in testing datasets than training datasets (see Table 16).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Overall, both the GBR and LSTM showed slightly higher MAE (4.22 mg m^{-3} vs. 3.87 mg m^{-3}) and RMSE (6.27 mg m^{-3} vs. 6.00 mg m^{-3}) when compared to workflow 1 (Table 17). But they also showed improved performance in terms of capturing the peak values of Chl during spring blooms (Figs. 17 and 19). Both workflows outperformed the SELMAPROTBAS PB model in simulating concentrations of lake nutrients (see Fig. 20). The ML models were more accurate in predicting the low values of NO_x and peak values of PO_4 and total P. However, both ML models and the PB model failed in predicting the extremely high values of measured lake nutrients, such as the autumn peak of NH_4 in 2017 (Fig. 20e) and the spring peak of O_2 in 2018 (Fig. 20c). Thus, higher workflow 2 MAE and RMSE (Table 17) are presumably due to the inaccuracies in the pre-generated nutrient training data, but the improved daily predictions that better capture the bloom events overshadow these flaws.

Table 17. Comparisons of model performance during the testing period based on RMSE, MAE, and R2. The unit of Chl is milligrams per cubic metre (mg m-3). In bold are the best fits of each statistical metric. For comparison of training and testing periods, see Table 16.

Model	PB	ML-workflow 1		ML-workflow 2		ML-workflow 3	
		GBR	LSTM	GBR	LSTM	GBR	LSTM
RMSE	7.18	5.77	**5.64**	6.27	6.00	5.94	5.81
MAE	4.77	**3.55**	3.58	4.22	3.87	3.99	3.71
R^2	−0.25	0.13	**0.20**	0.05	0.13	0.14	0.18

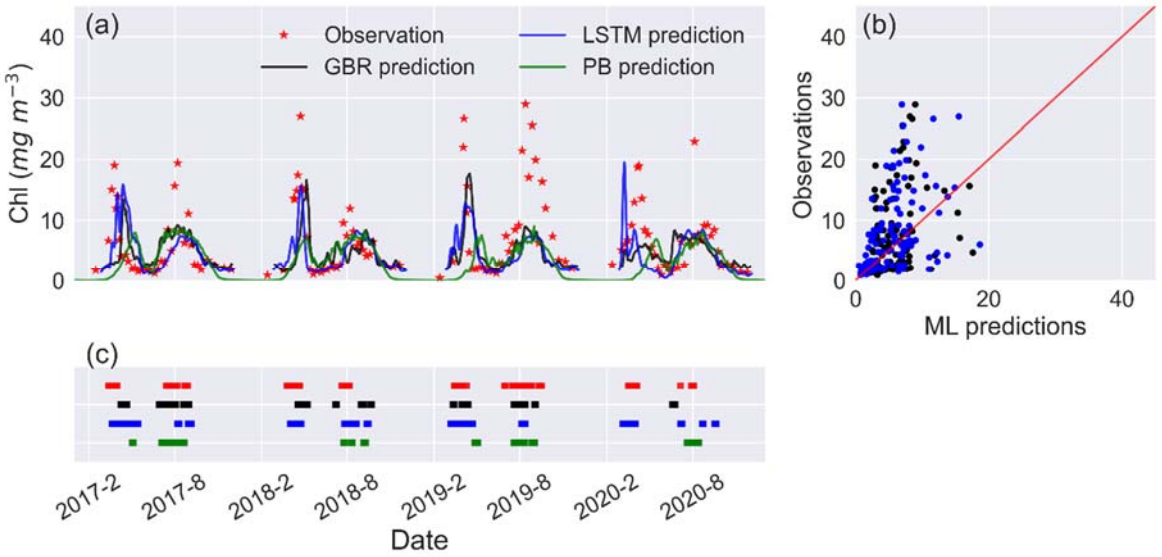


Fig. 19. *(a) Timeseries of observed and predicted Chl from GBR and LSTM models in workflow 2, (b) scatter plots of observations vs GBR and LSTM models. Penal (c) shows the observed and predicted algal bloom onsets in 2017-2020 using the same color coding as the previous panels. Results from the PB model simulation in Mesman et al. (2022) are also shown in (a) and (c).*

SMARTLAGOON

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

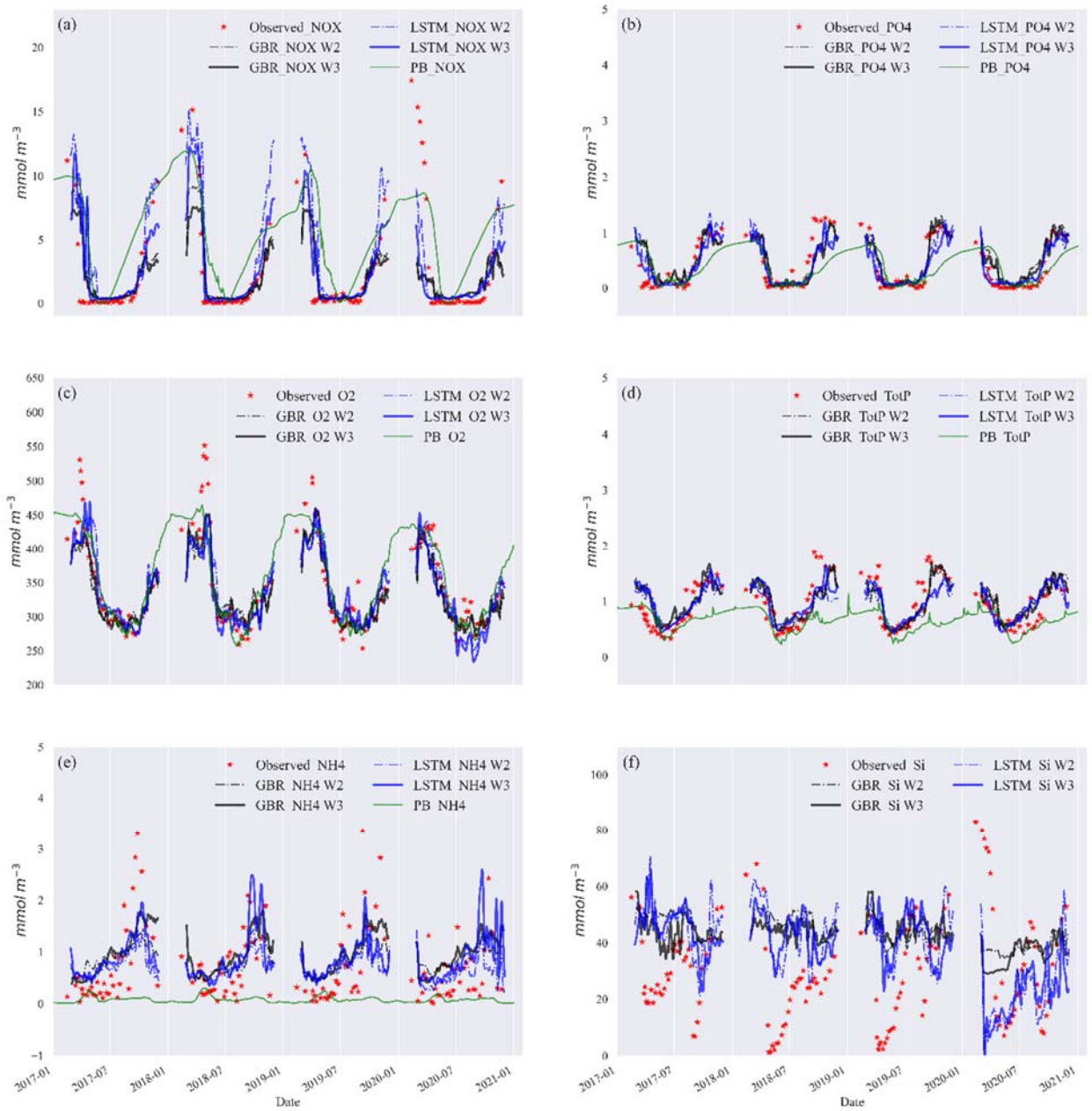


Fig. 20. Timeseries of six observed and predicted nutrients (a) NOX, (b) PO4, (c) O2, (d) Total P, (e) NH4, (f) Si, at surface (-3 m) from GBR, LSTM in workflow 2 (W2) and 3 (W3), and PB models. The Si simulations in the PB model had not been optimized, so these are not shown in the figure.

The workflow 3 is based on workflow 2 and including hydrodynamic training features derived from the GOTM model. Including hydrodynamic training information in workflow 3 did not significantly improve lake nutrient predictions compared to workflow 2 (see Fig. 20), and when using workflow 3 both ML models showed comparable performance in Chl predictions compared to workflow 1. However, the predictions of the spring bloom in all years improved compared to workflows 1 and 2, in terms of the magnitude and timing of the spring bloom (Fig. 17e). This was the case in 2019–2020 (Fig. 17a), which was an abnormally warm winter with only 5 d ice cover and had an unusually early spring algal bloom. Both the GBR and LSTM in workflows 2 and 3 did not capture the extremely intensive bloom (with peak values close to 30 mg m^{-3}) in summer of 2019, and neither did the PB

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

model. Furthermore, adding hydrodynamic features derived from the PB model improved predictions of the onset of algal blooms (Figs. 17e and 21), with the overall TPR increasing by 15% and 5% and FPR increasing around 5% and 3% in the GBR and LSTM models, respectively. Compared with the PB model, which showed lower TPR (15 %) and FPR (6 %), ML models are more likely to predict algal bloom at the correct time. The optimal TPR was from LSTM in workflow 3, which could detect the onset of algal blooms with TPR closed to 50 %. However, the concomitant higher FPRs indicating an incorrect warning of algal bloom is also more likely to occur in the ML models, since the PB model is more like to miss the bloom entirely. The kappa values of both ML models and the PB model are close to 80 %, showing that all models simulated the entire period (blooms and the periods between blooms) to a moderate–strong level (McHugh, 2012).

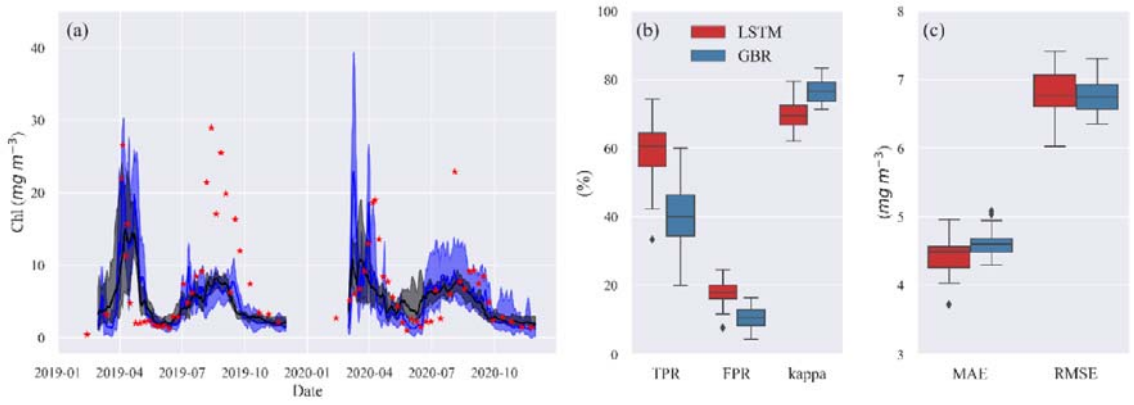


Fig. 21. (a) Time series of observed (red stars) and predicted Chl from GBR (black) and LSTM (blue) models in the shuffling training year test. The shades represent the range between minimum and maximum prediction, and the solid lines represent the median prediction. Panel (b) shows the box plot of TPR, FPR, and kappa, and panel (c) shows the box plot of MAE and RMSE of both models in the shuffling training year test.

The one clear failure of both the ML and PB-based model predictions was that during July–August 2019, Chl concentrations in integrated samples collected between the surface and 6–12m exceeded 20 mg m^{-3} over a 5-week period. Neither the PB model nor ML models captured this unusually persistent bloom (Figs. 17 and 22). At this time the phytoplankton were dominated by the Cyanobacteria Gloeotrichia and Anabaena, that form a resting akinete life stage at the end of their yearly bloom, which can initiate the following year's bloom as they are transformed to vegetative cells that migrate from the sediment to the upper water column. We hypothesize that the large summer bloom in 2019 was the result of unusually large recruitment of akinetes in this year (Karlsson-Elfgren et al., 2005, 2004). The life cycle of Cyanobacteria is not a process included in the PB model (but see Hense and Beckmann (2006) and Jöhnk et al., (2011)), so increased recruitment of akinetes could explain the underestimation of the 2019 summer bloom. Even the LSTM algorithms could not account for previous conditions so far back in time as to affect the formation and deposition of Cyanobacteria akinetes (this may require the memory of the last ice-free season). The consequent poor fit of summer bloom in 2019 partially led to the higher MAE and RMSE in the testing dataset compared to the training dataset in all three workflows, in both GBR and LSTM models.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

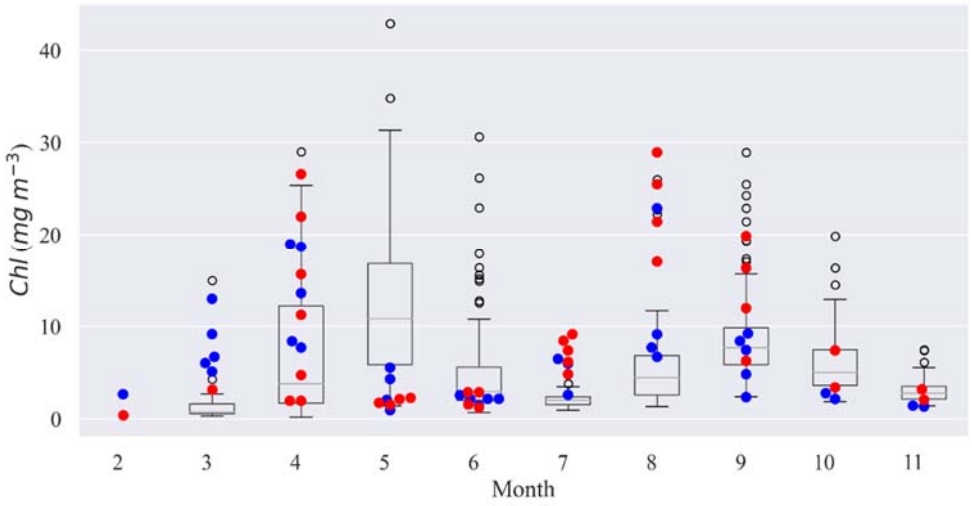


Fig. 22. Comparison of Chl concentrations in every month over 2004-2018, the red and blue dots represent the data from 2019 and 2020, respectively.

Warm winters can initiate a chain of events, i.e. shortening the ice cover duration, extending spring circulation, affecting nutrient availability, and causing an earlier spring bloom (Adrian et al., 2006; Yang et al., 2016). According to the ice record in Lake Erken (see Fig. 23), in 2020, the lake was covered by very thin ice for only 5 d, which is the shortest duration since observations were first recorded in 1954. The spring bloom in 2020 did occur earlier than other years (see Fig. 22), and both ML models which considered the timing of lake ice show fairly good performance in predicting the timing and magnitude of this abnormally early spring bloom (Figs. 17 and Fig. 24).

The hybrid model structure tested here provides a richer set of input data, leading to more accurate ML predictions of algal Chl at little additional computational cost or data requirements. Using data from the hydrothermal PB model allowed for the seasonal deepening of the thermocline and variations in the surface mixing layer depth and upwelling events, represented by *Wn*, to be encoded into the ML algorithms. These factors can affect the underwater light climate, the internal loading of phosphorus, and the transport of resting Cyanobacteria colonies from the hypolimnion into the epilimnion, favouring summer blooms of Cyanobacteria (Pettersson, 1998; Pierson et al., 1992). The inclusion of these factors did increase the accuracy of the ML models, especially in the case of unusual environmental conditions (e.g. spring of 2020, Figs. 17 and 24) that did not frequently occur in the remaining meteorological, hydrological, and biogeochemical training data.

The best model performance in predicting the timing of algal blooms was obtained after adding hydrodynamic features derived from a PB model in workflow 3, with TPR above 45% in detecting the onset of algal bloom during 2017–2020 and a modified accuracy (kappa) around 80 %, indicating a moderate–strong level of prediction.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

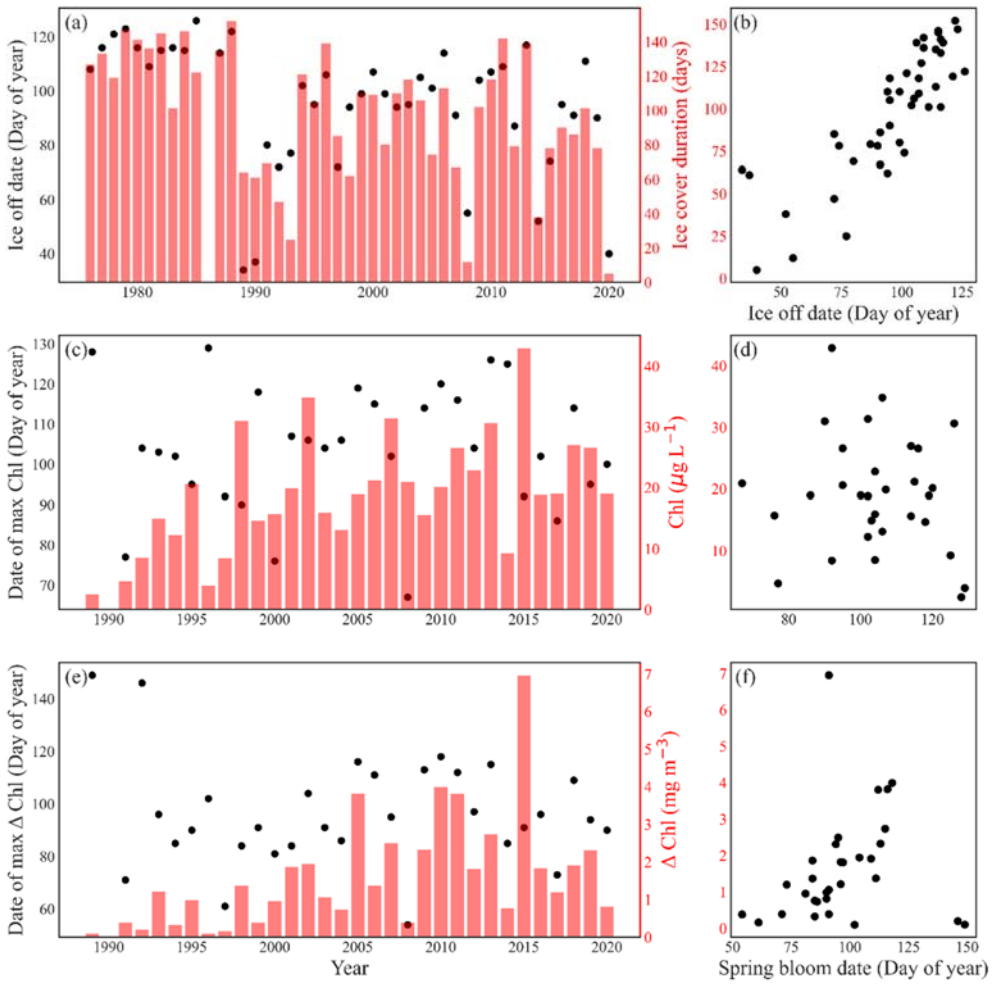


Fig. 23. (a, b) Ice break-up dates and ice cover durations since 1975 (Part of data from Weyhenmeyer et al. 1999). The timing of spring bloom in Lake Erken defined by (c, d) maximum Chl peak, and (e, f) steepest daily change of Chl.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

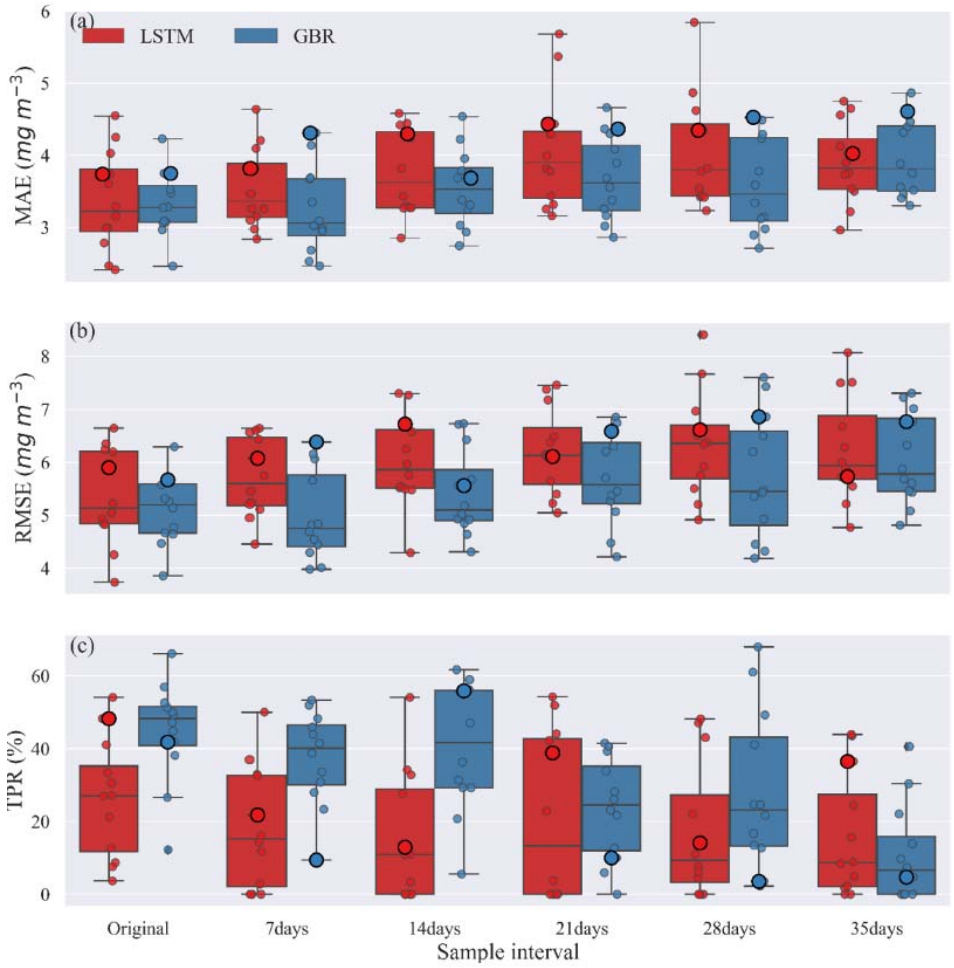


Fig. 24. Comparisons of (a) MAE, (b) RMSE, and (c) TPR between GBR and LSTM during the testing period created under various sample intervals. Circles along the box show the result from the testing period of all shuffled training–testing year combinations, and the bigger circles represent 2004–2016 training and 2017–2020 testing year combination, as used in Fig. 17.

- **Effects of shuffling training years on 2019–2020 predictions:**

The results presented so far are based on a typical strategy of training ML models for a historical period, in this case 2004–2016, and then accessing model performance in a second period between 2017–2020. The accuracies of the model predictions were to some extent related to the range and variability in the training data. To evaluate the importance of this, we randomly removed 2 years from a 2004–2018 training dataset and made 30 different predictions of Chl during 2019–2020 when the models had difficulties predicting spring and summer blooms (Fig. 24). When trained with the various shuffled combinations, both ML models were capable of reproducing the seasonal variations in algal Chl with a 4.5% and 5.8% coefficient of variation (CV) in MAE and a 24.0% and 16.4% CV in TPR of GBR and LSTM, respectively (see Table 18). This provides an indication of the uncertainty that may arise as a consequence of differences in the training datasets used for in our workflows. And, it also shows that even a relatively long training period of 13 years can not totally capture the system behaviour in such a way as to lead to nearly similar bloom predictions. Although none of the model

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

runs captured the intensive summer bloom in 2019, the spring bloom in both years was well represented, especially by LSTM, in terms of timing and magnitude.

Despite comparable RMSE and MAE in LSTM and the GBR (Fig. 21c), both higher TPRs (with median of 60 %) and FPRs (with median of 18 %) in LSTM indicate that the LSTM model was more aggressive in making algal bloom predictions. The GBR model's apparent advantage in FPRs (with median 10 %) is largely the result of it making a lower number of bloom predictions since the low concentrations between spring and summer blooms in 2020 were not well represented (Fig. 21b).

Table 18. Coefficient of variation of evaluating metrics in shuffling training years to test 2019-2020.

Model	*MAE* (%)	*RMSE* (%)	TPR (%)	FPR (%)	Kappa (%)
GBR	4.49	4.00	23.98	31.77	4.53
LSTM	5.80	5.21	16.36	21.41	6.30

- **Shuffling year data sparsity test:**

To examine the possible use of workflow 3 when data are less frequently available, lake nutrient and Chl data were downsampled so that the effects of sampling frequency on model predictions could be evaluated. Each downsampled dataset was also rearranged into 13 different 13-year training periods and 4-year testing periods. The variability in predictions provided a measure of model performance and uncertainty. Figure 24 shows the uncertainty in model predictions as a consequence of the chosen sampling intervals. The MAEs and RMSEs of both GBR and LSTM models tended to increase with the longer sample intervals. The median MAE was always slightly higher for the LSTM model, except when trained with the original dataset (Fig. 24a). While our initial evaluation of TPR using 2017–2020 as the testing period and 2004–2016 as the training period suggested the LSTM model was more accurate in turns of detection of algal bloom onsets (Fig. 25), Fig. 24c showed the median TPR of GBR model calculated by the shuffling year test was over 50 %, higher than that found when using the original testing and training periods. This can be explained by the fact that the 2017–2020 testing period as in Fig. 25 and shown as large points in Fig. 24 was unusually difficult for the GBR to simulate. Consequently, even though the GBR model usually performs better in the shuffled data test in Fig. 24, Fig. 25, which show the results of the 2017–2020 testing period, presented the opposite result. This illustrates the importance of the sequence of training and testing years for evaluating model performance.

For the first three sampling intervals the GBR model clearly had better TPR values than the LSTM model. The median TPRs of GBR model started to drop below 30% once the sample interval reached 21 d. For LSTM, medium TPRs remained lower than 30 %, for all sampling intervals, but also showed a much wider range of variability (Table 19) dependent on the training and tested datasets used. In some training and testing year combinations, TPRs are close to 0% (Fig. 24), and CVs of the TPRs are highly variable, even at the original sample interval, being over 30% for GBR and over 60% for LSTM, indicating that the correct detection of algal blooms in both models is highly dependent on the years used to train the models.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

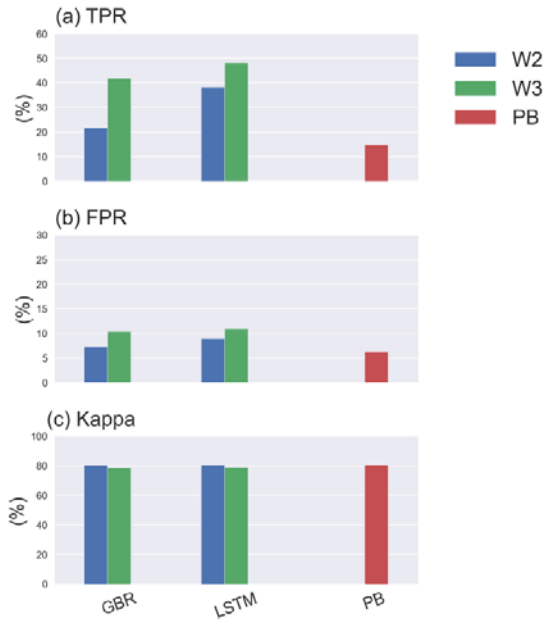


Fig. 25. TPR, FPR, and kappa of GBR and LSTM models in workflow 2, 3, and the PB model.

Table 19. Coefficient of variation of MAEs, RMSEs, and TPRs in shuffling year data sparsity test.

Model	Sample interval	MAE (%)	RMSE (%)	TPR (%)
GBR	Original	13.82	12.88	31.62
	7 days	18.60	17.08	34.63
	14 days	15.17	15.12	43.94
	21 days	15.73	15.22	59.51
	28 days	18.30	20.65	77.09
	35 days	13.63	14.11	118.61
LSTM	Original	20.52	16.98	62.12
	7 days	15.71	13.05	91.63
	14 days	15.97	14.32	113.53
	21 days	19.83	13.08	107.39
	28 days	19.15	15.81	110.40
	35 days	14.44	16.12	106.99

The resulting variability provided a more accurate estimate of the model performance at each downsampled data interval and showed that increasing sample interval led to reduced performance for both ML models, in terms of MAE, RMSE, and the CV of TPR. These tests also highlighted that the performance of both ML models, especially LSTM, varied with the sampled history of events in the training period for evaluating a specific pattern of change in the testing period. We suggest that testing strategies similar to the shuffle methods used in this study are needed to accurately evaluate the expected accuracy of ML models when applied to any given site. The estimated uncertainty in

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

shuffling training year tests (Fig. 21) and shuffling training–testing year tests (Fig. 24) can be used to better represent the uncertainty of ML derived forecasts.

In general, both models preformed best at the original and 7 d sampling interval but then showed slightly worse performance that was consistent up to a sample interval of 21 d. In terms of the errors evaluated over the entire 4-year testing period (Fig. 24a,b) the GBR model had lower errors and, therefore, better predicted the seasonal variations of Chl concentration. The time series comparison of observed and predicted Chl from this shuffling year data sparsity test can be found in the Figs. 26, 27 y 28. The data sparsity test (Fig. 24) showed that, at least for Lake Erken, the ML models can still detect the seasonal algal dynamics even for sample intervals approaching 1 month (Figs. 26–28).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

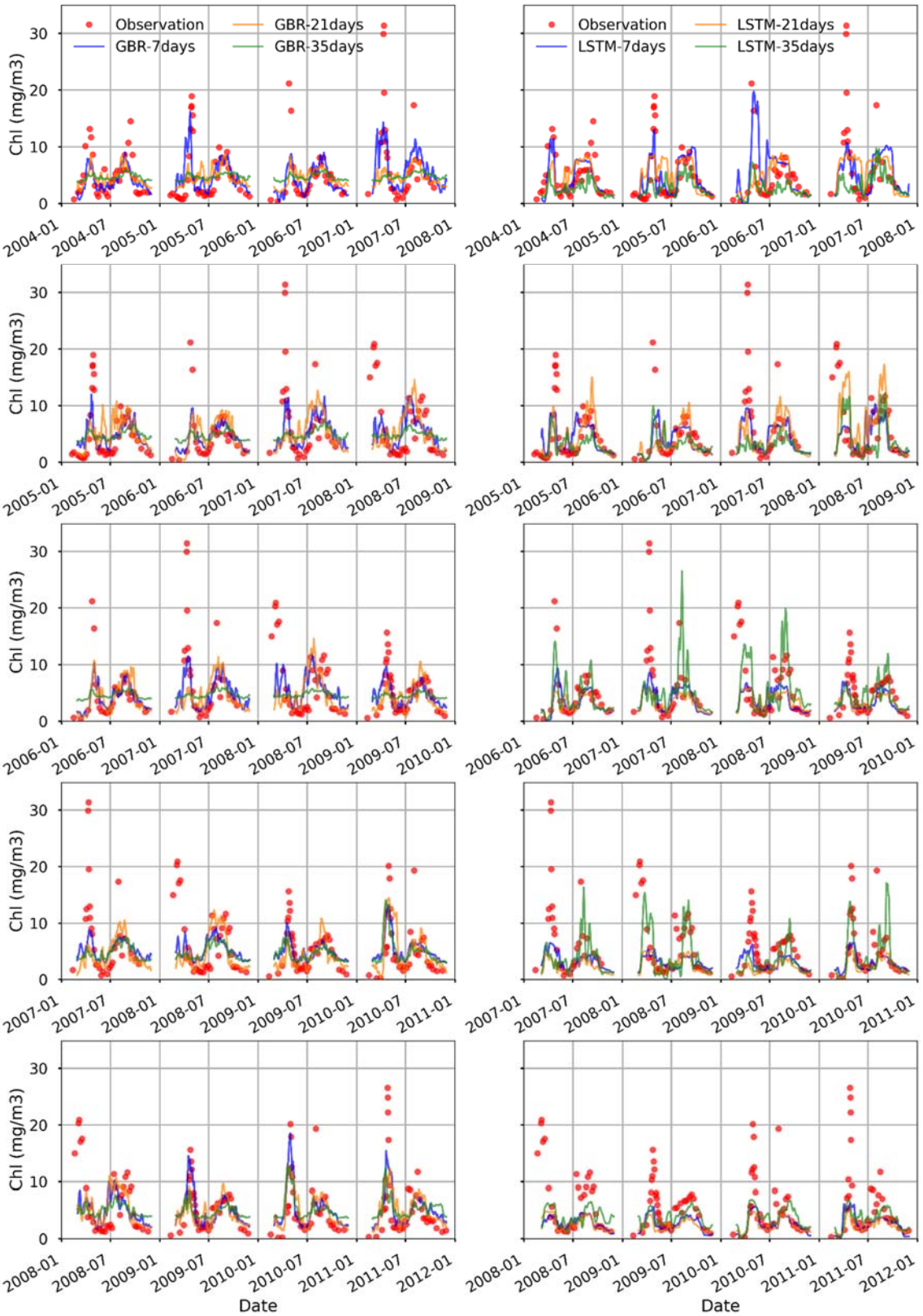


Fig. 26. *Timeseries of observed and predicted Chl from GBR (panels on the left) and LSTM (panels on the right) models based on 7-day, 21-day, and 35-day sample intervals, via leave-four-year-out shuffling year test. Each row is a different 4-year period.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

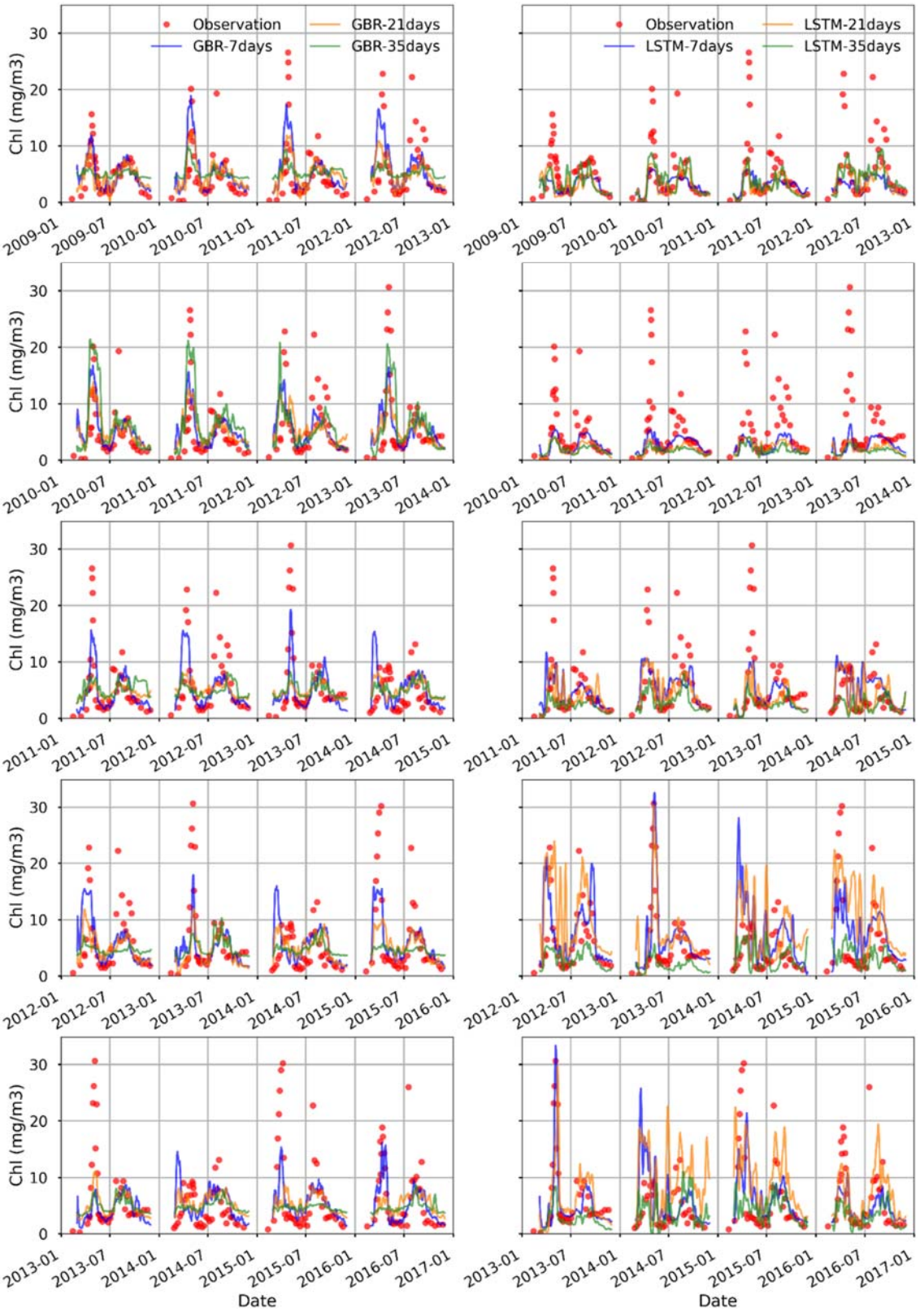


Fig. 27. *Timeseries of observed and predicted Chl from GBR (panels on the left) and LSTM (panels on the right) models based on 7-day, 21-day, and 35-day sample intervals, via leave-four-year-out shuffling year test (Same as Figure 20, but with different x-axis).*

SMARTLAGOON

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

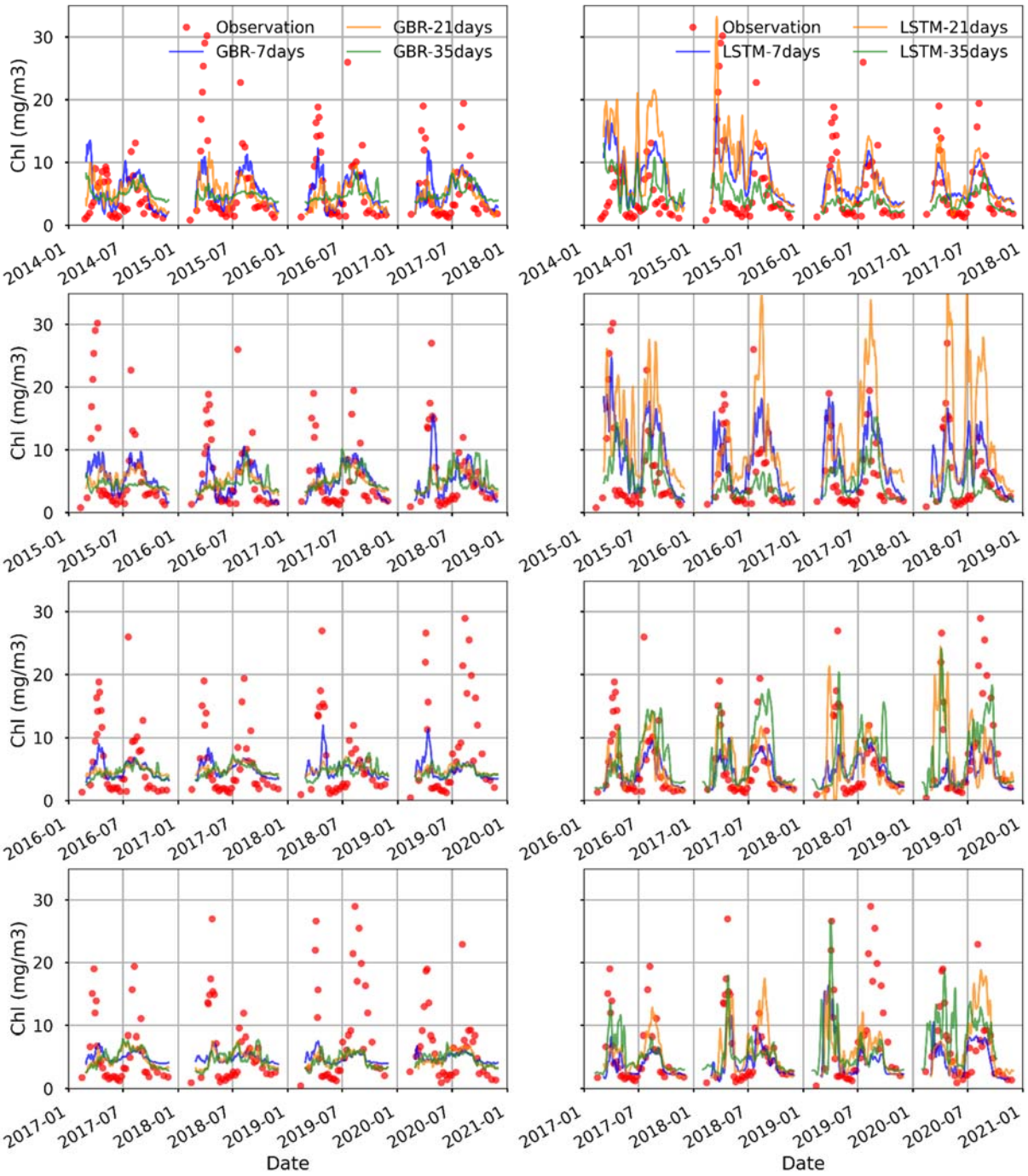


Fig. 28. *Timeseries of observed and predicted Chl from GBR (panels on the left) and LSTM (panels on the right) models based on 7-day, 21-day, and 35-day sample intervals, via leave-four-year-out shuffling year test (Same as Figure 20, but with different x-axis).*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

4.1.2 Predictions of algal bloom in Mar Menor lagoon

The first experiment performed evaluated a local model for each ISMP and depth. The second experiment used all ISMPs indistinctly to find a global model. Then a global model was evaluated by clustering. Finally, the experiments were performed using Deep Learning.

- **Creating ML local model for ISMP:**

This first experiment aims to train ML models for each station individually. Due to the heterogeneity of the Mar Menor, each monitoring point (i.e., ISMP) may have different characteristics that may benefit from modelling using ML methods. Thus, the main hypothesis at this section is that each ISMP requires a different ML model to be trained for this particular point.

To obtain the best results we set up ML models using different parameterisations. Table 20 presents the best results for each ISMP and the depth where it is most accurate, indicating the best ML technique. In the selection of the best results, we have taken the highest R^2 as a reference. In this first experiment, the ML models for ISMP 1 and 9 do not perform well, have R2 values of 0 and large error values for both MAE and MSE. In fact, these ISMPs are the shallowest ISMPs in the Mar Menor and, therefore, the remote sensing reflectances introduce many errors, resulting in a high disparity with respect to in situ monitoring.

***Table 20**. The best performance model for each station and its depth wherein R2 is max to predict Chl-a. MAE and MSE are measured in mg/m^3.*

Technique	ISMP	Depth(m)	R^2	MAE	MSE
DT	1	0	0.0	1.526	12.919
MLP	2	0	0.863	1.141	3.473
MLP	3	0	0.762	1.132	4.171
MLP	4	1	0.808	1.1	2.75
MLP	5	4	0.808	1.1	2.750
RFR	6	1	0.864	0.932	2.88
MLP	7	0	0.79	0.943	3.966
RFR	8	0	0.877	0.84	1.589
KNN	9	0	0.0	4.682	53.934
RFR	10	0	0.738	0.919	4.481
RFR	11	1	0.861	1.047	3.1
RFR	12	1	0.887	0.965	2.823

Another relevant aspect to note is that for all ISMPs except ISMP 5, the best results are obtained at shallow depths of 0 to 1 meter or 1 to 2 meters. For ISMP 5, the best results are obtained at depths of 4 to 5 meters. This makes sense since ISMP 5 is located on the other side of the lagoon (see Fig. 5). Finally, there is no single model that is better than the others for all ISMPs. The best performing techniques are MLP and RFR for 5 different ISMPs and DT and KNN for one different ISMP. Although the results of most ISMPs, excluding 1 and 9, are satisfactory, the lack of a single prediction model creates difficulties and complexity for the prediction system. Hence the need to find a global model of Chl-a in the Mar Menor. Therefore, different ML techniques for all ISMPs and all depths were then tested to find a global model.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

- **Creating global ML model:**

In this second experiment, ISMPs are no longer considered in isolation, so a global prediction model is created for each ML technique. For these global models, a new dataset is created with data from all ISMPs and depths together. Table 21 shows, for the first 4 depth levels, the results of the global models for each ML technique. In this experiment, depths greater than 4m are discarded due to the low accuracy produced also in the previous experiment.

***Table 21**. Results for ML at 4 depths using all stations to predict Chl-a. MAE and MSE are measured in mg/m^3.*

Technique	Depth(m)	R^2	MAE	MSE
DT	0	0.553	1.249	7.481
	1	0.671	1.217	6.237
	2	0.264	1.582	10.101
	3	0.266	1.57	8.386
KNN	0	0.662	1.159	6.528
	1	0.692	1.161	4.921
	2	0.48	1.326	5.847
	3	0.409	1.519	8.036
MLP	0	0.587	1.344	6.091
	1	0.596	1.351	7.561
	2	0.578	1.397	7.224
	3	0.526	1.391	5.894
RFR	0	0.644	1.251	6.456
	1	0.631	1.148	5.562
	2	0.61	1.212	5.093
	3	0.555	1.309	4.956

Table 21 shows that the best result obtained is with the KNN technique at depths between 1 and 2 meters for Chl-a prediction. The R^2 value is 0.6919 and the MAE and MSE are 1.1614 and 4.9211 mg/m^3 respectively. The next best performing technique is DT at the same depth with an R^2 of 0.6707. The RFR and MLP techniques are the worst performers, contrary to the local models. This is because the DT and KNN techniques are less affected by the noise introduced by ISMP 1, 5 and 9. To verify this theory of noisy data, the same experiment is performed using a 5-fold cross-validation, but without including the data from these ISMPs. As mentioned above, ISMP 1 and 9 have very shallow data and a small amount of data. ISMP 5 is also discarded because the best results were obtained in the local models at great depths and also, being outside the lagoon, satellite data can provide possible noise in the results.

- **Creating global model by clustering:**

Taking into account the above reasons, we can assume the existence of a cluster related to the maximum depth of the ISMP. Thus, we can consider two clusters initially: (1) Cluster-0 less than 2 meters deep (ISMP 1 and 9), (2) Cluster-1 more than 2 meters deep. Thus, the accuracy of the ML models for Cluster-1 improves in general, as shown in the table 22. For the DT and KNN techniques, the improvement is only about 3 or 4 hundredths of a fit in the reference to the R^2 value, and the MAE and MSE errors decrease slightly or remain the same. However, the MLP and RFR techniques obtain much better results, reaching R^2 fit values of about 0.8 and MAE and MSE errors of about 1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

mg/m^3 at shallow depths of 0 to 1 meter. Figure 29 shows the dispersion graph for RFR model where this model is able to detect trend changes in Chl-a data. Actually, RFR introduces the best model able to detect these peaks. From these ML experiments several conclusions can be drawn. The first is a possible cluster based on the depth of ISMP. Also that it is possible to discard ISMP 5 as it is located outside the lagoon. Finally, the global ML models fit slightly worse than the local ones. However, local models are not homogeneous and complicate the system. It is justified to try other techniques such as Deep Learning to find a model.

Table 22. *Models' precision at different depths to predict Chl-a in Cluster-1. MAE and MSE are measured in mg/m^3.*

Technique	Depth(m)	R^2	MAE	MSE
DT	0	0.639	1.159	5.733
	1	0.692	1.239	5.776
	2	0.629	1.436	6.458
	3	0.488	1.45	6.417
KNN	0	0.765	1.103	5.041
	1	0.682	1.303	6.215
	2	0.558	1.402	6.882
	3	0.451	1.381	6.077
MLP	0	0.772	1.041	4.012
	1	0.75	1.267	5.543
	2	0.677	1.334	5.829
	3	0.556	1.347	5.790
RFR	0	0.833	0.964	3.647
	1	0.81	1.0783	3.969
	2	0.724	1.296	5.163
	3	0.548	1.314	5.769

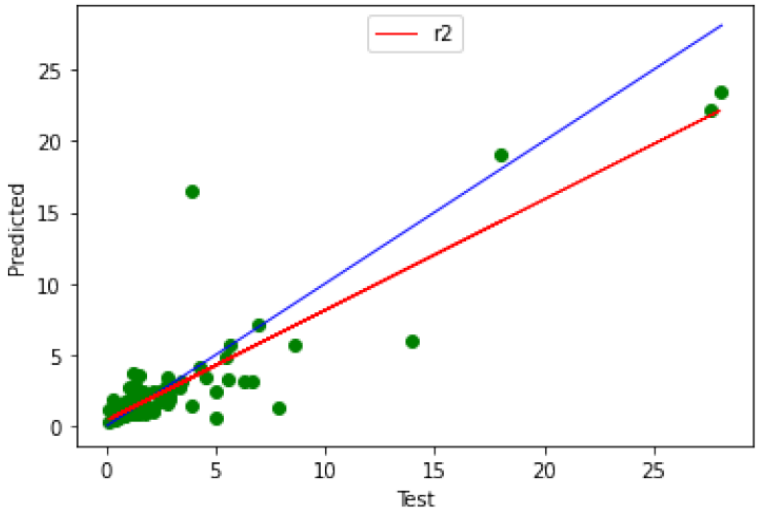


Fig. 29. *RFR dispersion plot for depth 0 at ISMP 6*

- **Using Deep Learning:**

Although the accuracy of the ML models is not bad, a DL model could perform better. Moreover, it could work without a depth grouping. Therefore, we repeat the above steps but using CNN Conv1D. To find the relation among adjacent reflectances we set a kernel windows to 4 to build the layer. Initially CNN performs well, but for single stations the results are not very significant compared to

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

the ML models. Similar to ML models, the best accuracy is achieved using CARM depth 0. However, it shows better accuracy than ML when using all ISMPs, i.e. when CNN is trained to predict the values of a complete ISMP using the rest of the ISMPs. The table 23 show how the behaviour of ISMPs 1, 5 and 9 are unrelated to each other. Therefore, clustering by depth seems necessary and also justifies the exclusion of ISMP 5.

Table 23. *Results for CNN technique at 0 depth using a global model to predict Chl-a of whole station. MAE and MSE are measured in mg/m3.*

ISMP	R^2	MSE	MAE
1	0.00	15.43	2.08
2	0.90	2.58	0.94
3	0.85	2.60	0.92
4	0.88	1.72	0.74
5	0.00	0.83	0.59
6	0.91	1.80	0.75
7	0.89	2.00	0.77
8	0.90	1.27	0.69
9	0.00	48.82	3.86
10	0.76	4.18	0.88
11	0.87	2.14	0.90
12	0.87	2.81	0.80

Taking in mind these results, we look for a global model trying to predict randomly whatever ISMP value into Cluster-1. Thus, training the model with 80% of data and using a cross-validation of 5 provided by the model the results are R^2=0.89, MSE=1.95 mg/m^3 and MAE = 0.78 mg/m^3, which shows a good performance that outperforms the previous ML models. Also, the Figure 30 shows the CNN behaviour in Chl-a peaks in depth 0, when algal blooms phenomenon occur, that indicates this model can represent the Mar Menor Chl-a behaviour.

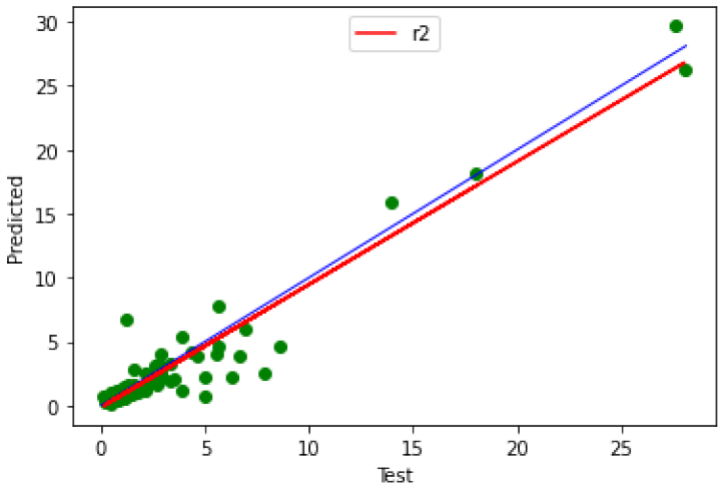


Fig. 30. *CNN dispersion plot for depth 0 at ISMP 6*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

4.3 Hypoxia predictions

- **Model performance:**

Despite diversity in the physical size, lake type and trophic level of the lakes tested in the study, three model approaches well simulated the seasonal variation of both surface and bottom DO. In Fig. 31 and 32, R^2 values of the 30 separate model runs were plotted for each lake in order to illustrate the effects of variations in the training data set on the models predictive power.

For surface DO simulation in the testing periods, three approaches performed best in Lake Erken with direct LSTM and 2-step mixed model presenting averaged $R^2 > 0.8$ and averaged *NMAE* < 0.05 mg/l, and for bottom DO simulation, direct LSTM and GBR shows best testing results in Lake Mendota with averaged $R^2 > 0.8$ and averaged *NMAE* < 0.06 mg/l. The accuracies of these ML models improved those obtained in previous studies of Erken Lake (Mesman et al., 2022) and Lake Mendota (Ladwig et al., 2021). However, the 2-step mixed model achieved even better testing results in Lake Müggelsee with averaged R^2 similar as that in Lake Mendota > 0.8 and averaged *NMAE* < 0.03 mg/l.

Direct GBR approach showed more stable model performance with less variation in both training and testing period (Fig. 31, 32, also see coefficient of variation in Table 24) than direct LSTM and 2-step mixed model. Combined the GBR and LSTM, 2-step mixed model improved the accuracy of both surface and bottom DO simulation in Lake Müggelsee by largely increasing R^2 and decreasing NMAE, and it also decreased the variation in the surface DO simulations in Lake Ekoln. However, in some cases, the 2-step mixed model reduced model accuracy. For example, both surface and bottom DO simulations in Lake Furesø and bottom DO simulations in Lake Ekoln.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

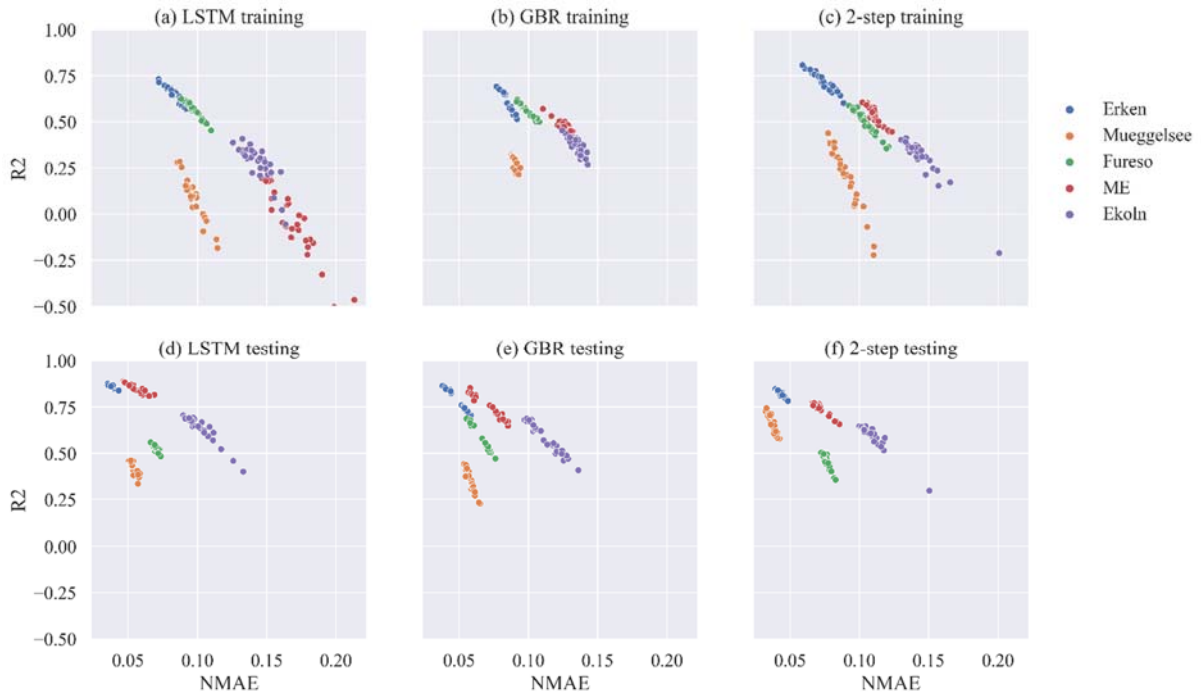


Fig. 31. *Evaluating metrics of three ML model approaches for DO in the surface water in (a-c) training dataset, and (d-f) testing dataset. Each point represents one model run, and the points located in the left upper corner of the figure means better model performance.*

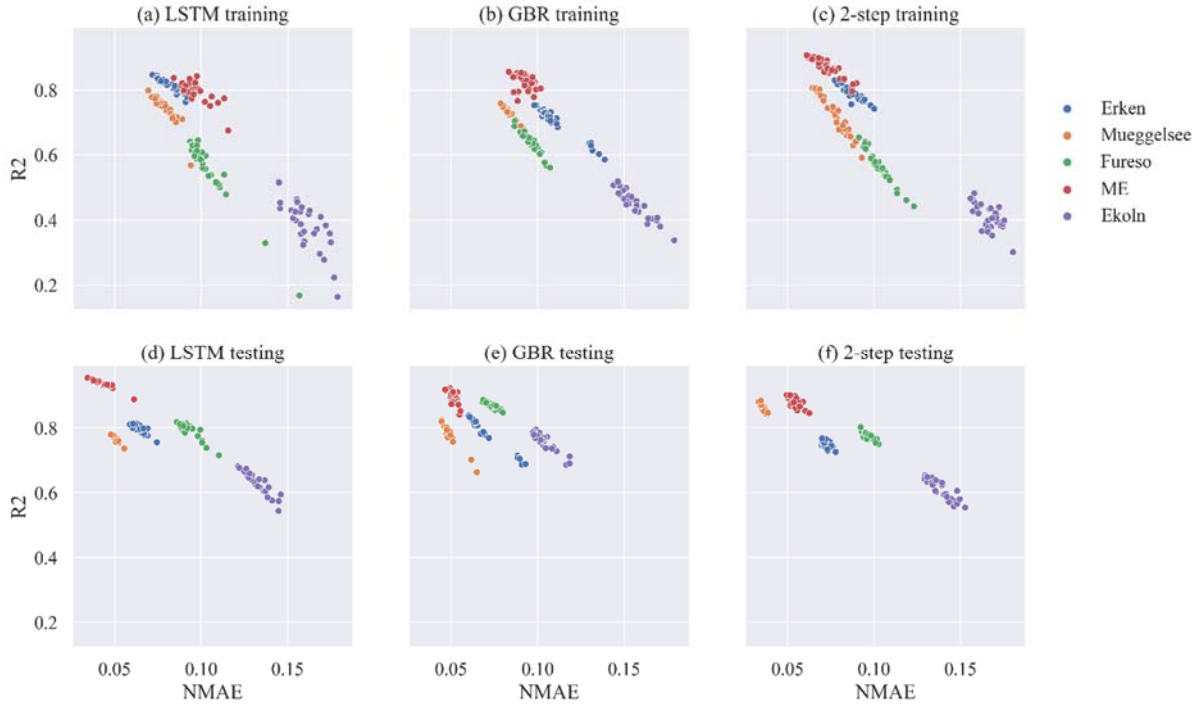


Fig. 32*. Similar as Figure 31, but for DO in the bottom water in (a-c) training dataset, and (d-f) testing dataset.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

***Table 24**. coefficient of variation (%) in three model approaches.*

Lake	Variable/ model approach		(%) NMAE	(%) R2
Erken	Surface DO	Direct LSTM	8.23	7.96
		Direct GBR	5.09	9.28
		2-step mixing	9.67	6.59
	Bottom DO	Direct LSTM	7.57	2.7
		Direct GBR	10.65	6.64
		2-step mixing	6.49	2.75
Müggelsee	Surface DO	Direct LSTM	7	113.47
		Direct GBR	1.65	10.74
		2-step mixing	9.98	85.38
	Bottom DO	Direct LSTM	7.31	5.62
		Direct GBR	3.73	2.52
		2-step mixing	10.75	8.86
Furesø	Surface DO	Direct LSTM	4.66	6.18
		Direct GBR	5.05	6.49
		2-step mixing	6.45	11.59
	Bottom DO	Direct LSTM	12.12	17.73
		Direct GBR	5.03	5.67
		2-step mixing	6.97	10.04
Mendota	Surface DO	Direct LSTM	9.38	333.35
		Direct GBR	3.67	6.35
		2-step mixing	4.42	7.79
	Bottom DO	Direct LSTM	7.71	3.58
		Direct GBR	5.41	2.77
		2-step mixing	8.58	2.64
Ekoln	Surface DO	Direct LSTM	6.38	34.45
		Direct GBR	3.97	15.14
		2-step mixing	8.31	29.15
	Bottom DO	Direct LSTM	5.78	18.89
		Direct GBR	5.4	9.44
		2-step mixing	3.82	9.51

The polymictic characteristics of Lake Müeggelsee makes the prediction of hypoxia more variable and therefore more challenging, while, the 2-step mixed model workflow stands out among the three ML approaches, showing its advantages in this lake by successfully capturing the fluctuations in bottom DO concentrations that were recorded by the high frequency YSI sensor.

Among the five lakes, the simulations of DO concentrations in Lake Ekoln were the worst with surface DO R^2 < 0.4 in the training data and around 0.6 in the testing data, bottom DO R^2 < 0.5 in the training data and around 0.7 in the testing data. Also, the results have larger variance than that in other lakes (Table 24). This presumably was due to the relatively large sampling interval, which suggests that the performance of ML models to some extent relies on the temporal resolution of training data. More results are shown in the following figures (Figs. 33-42).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

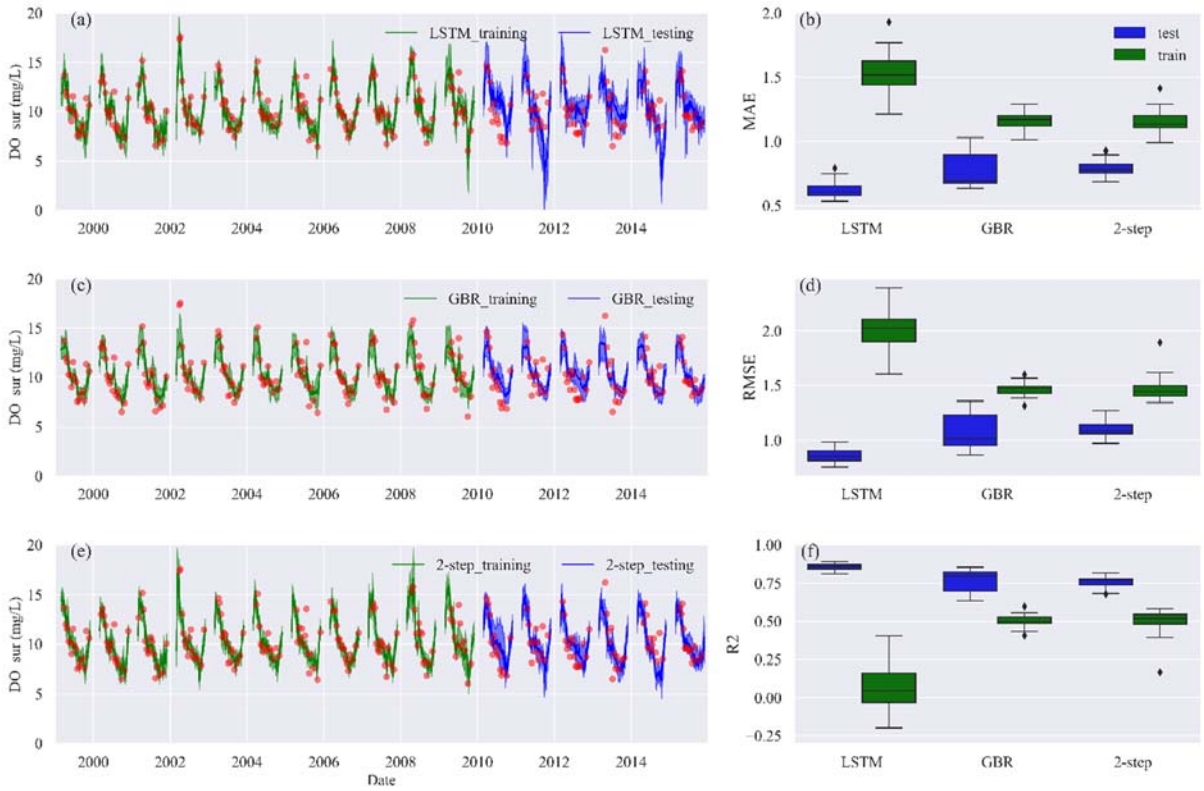


Fig. 33. *(a, c, e) Timeseries of observed and modeled surface DO via three ML approaches in Lake Mendota. Green lines are calibration period and blue lines are validation period. Min and max predictive values among 30 times run contour the shaded areas, representing the uncertainty level of model prediction. (b, d, f) show the evaluating metrics of the models.*

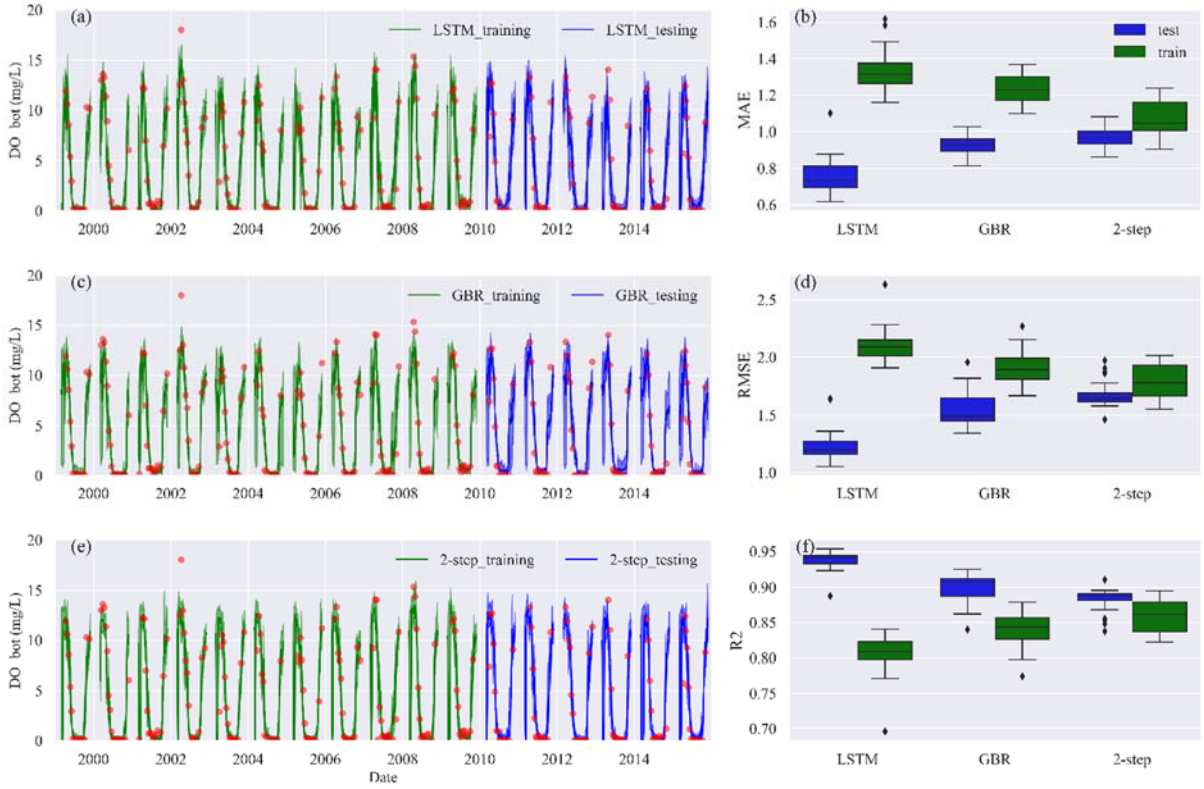


Figure 34. *Same as Fig. 33, but for bottom DO in Lake Mendota.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

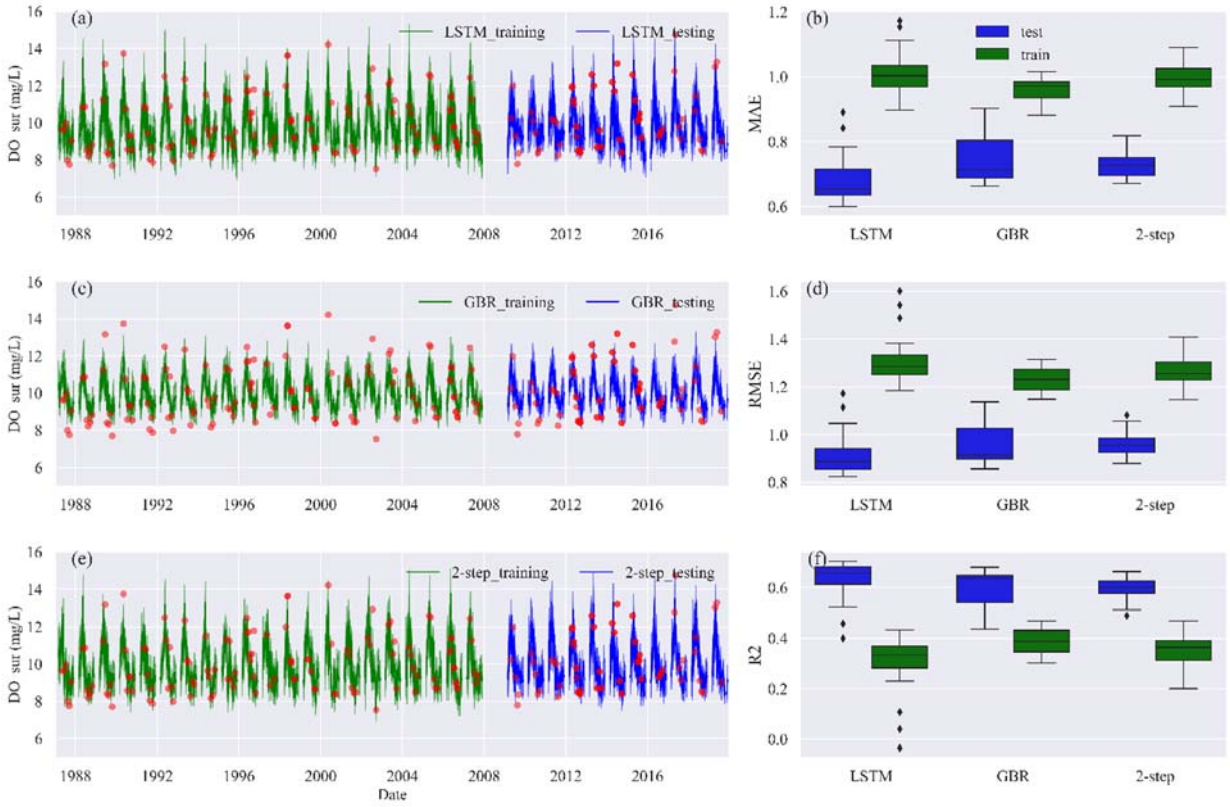


Fig. 35. *Same as Fig. 33, but for surface DO in Lake Ekoln.*

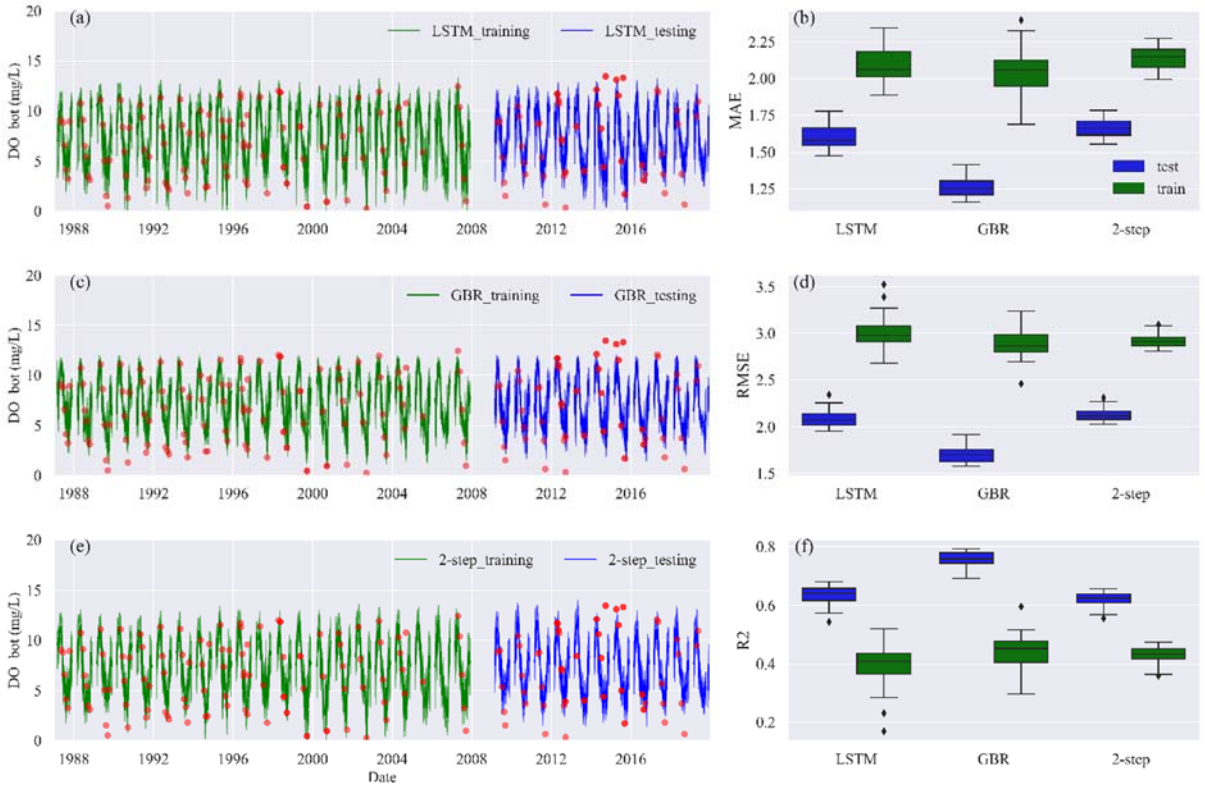


Fig. 36. *Same as Fig.33, but for bottom DO in Lake Ekoln.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

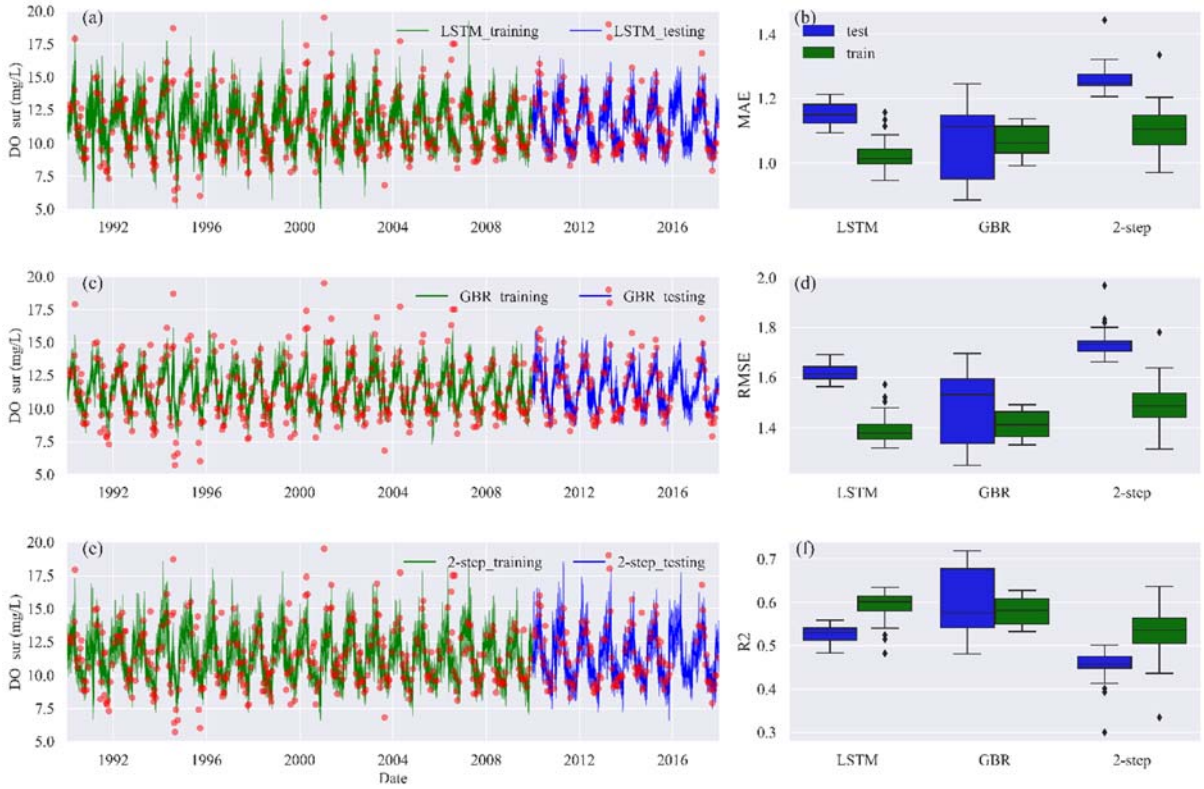


***Fig. 37**. Same as Fig. 33, but for surface DO in Lake Furesø.*

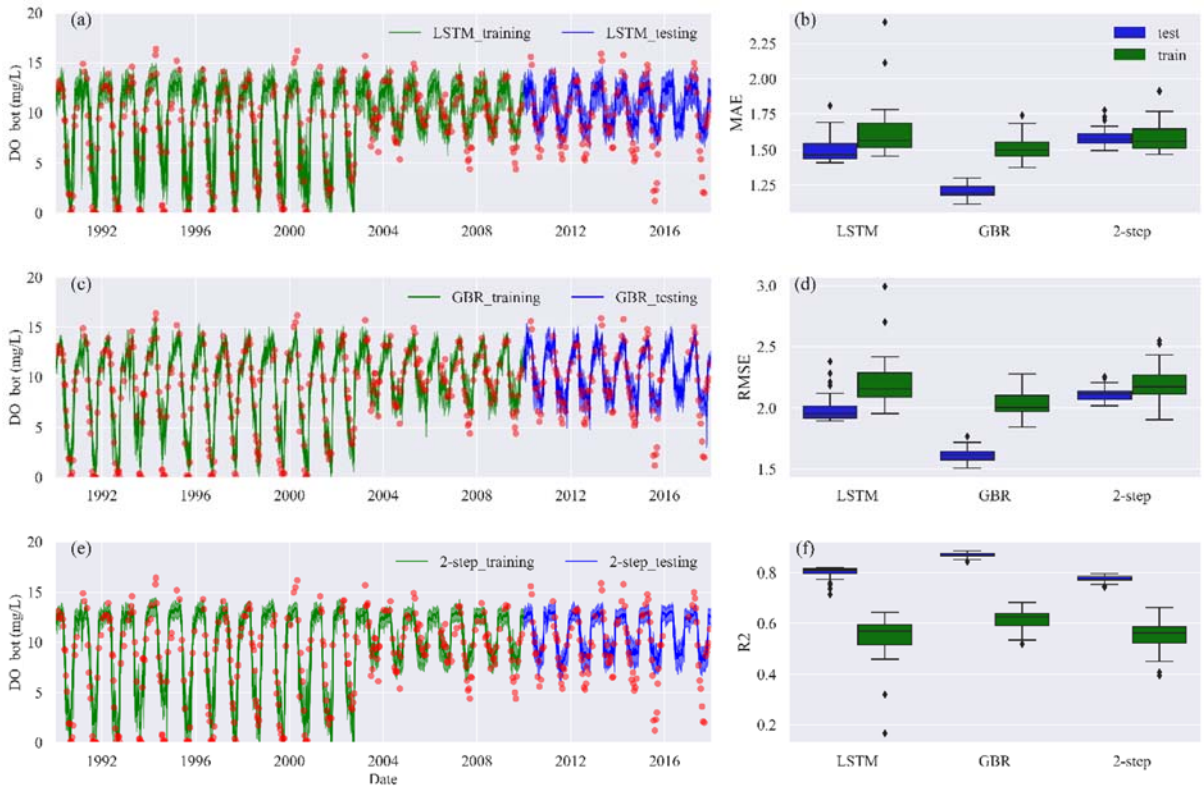


***Fig. 38.** Same as Fig. 33, but for bottom DO in Lake Furesø.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

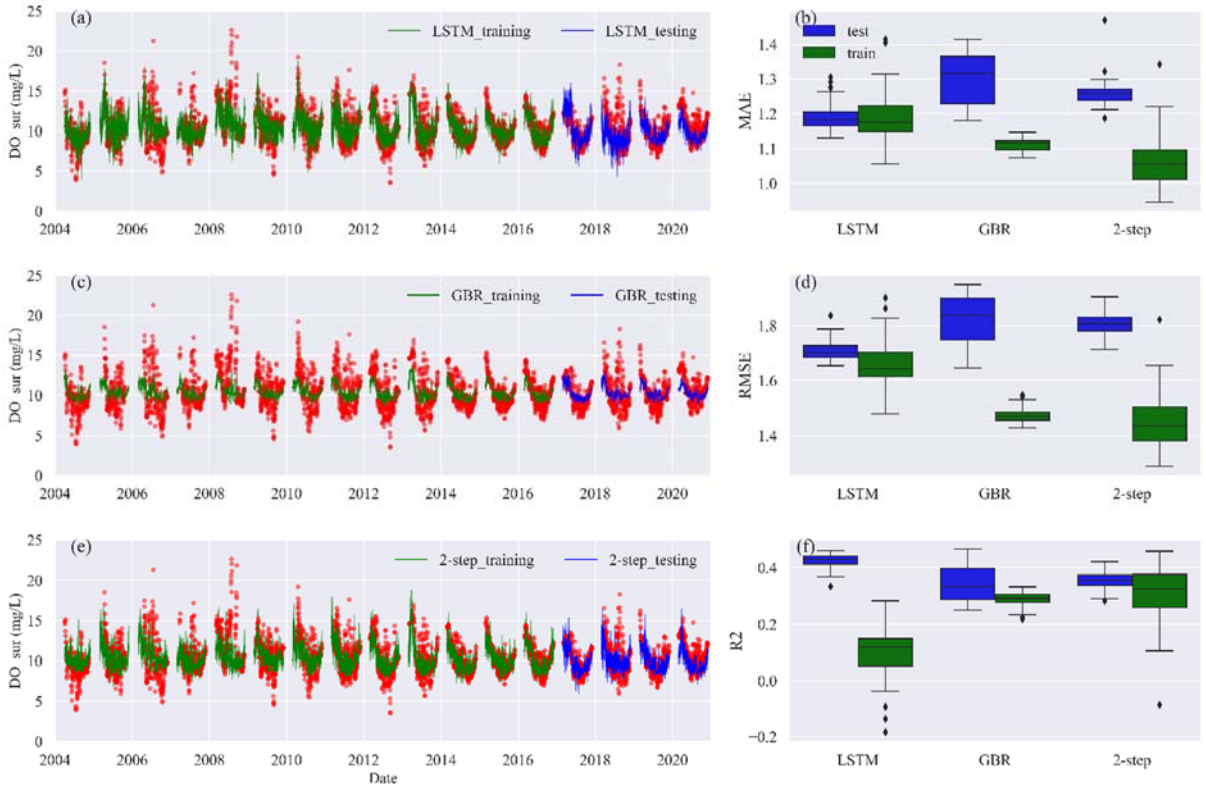


Fig. 39. *Same as Fig. 33, but for surface DO in Lake Müggelsee.*

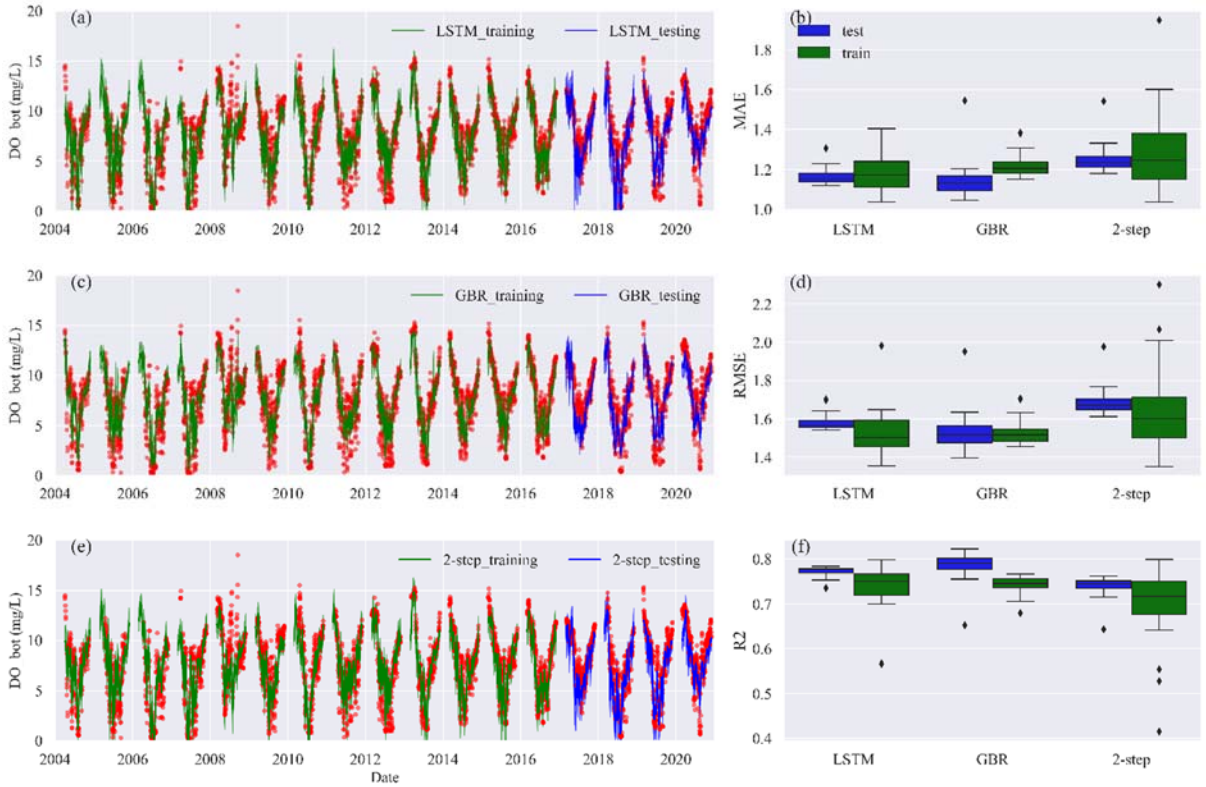


Fig. 40. *Same as Fig. 33, but for bottom DO in Lake Müggelsee.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

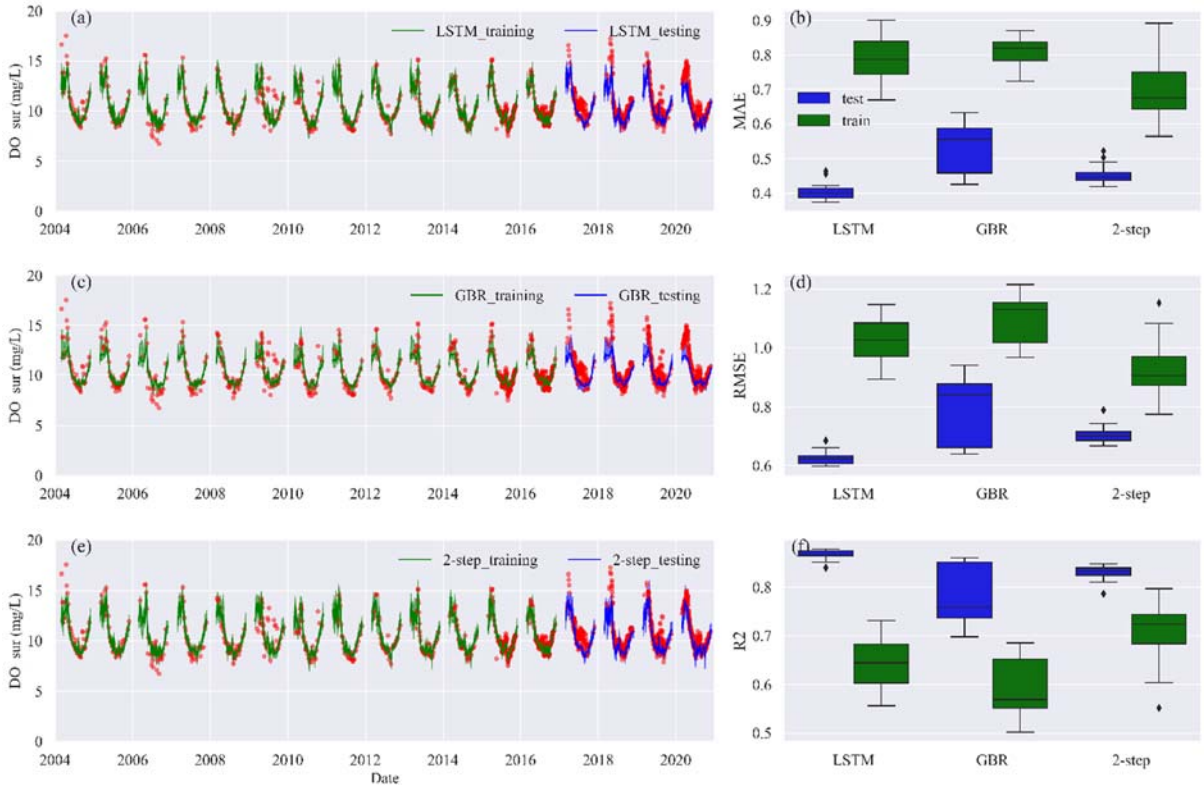


Fig. 41. *Same as Fig. 33, but for surface DO in Lake Erken.*

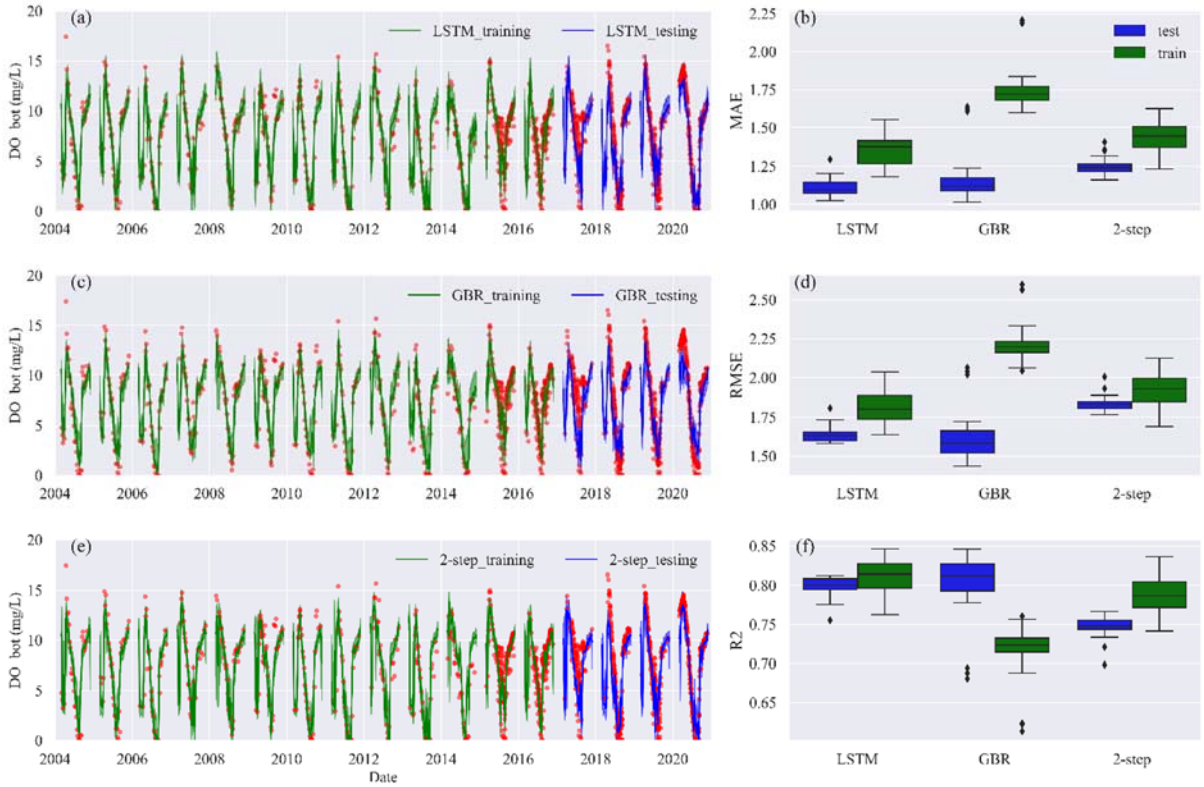


Fig. 42. *Same as Fig. 33, but for bottom DO in Lake Erken.*

In Lake Furesø, although all three ML models did not capture the abnormalities that oxidation treatment failed to improve the hypoxia largely in 2015 and 2017 (Fig. 38), most testing results showed that ML models learned the effect of oxidation treatment from training data and achieved generally

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

promising R2 and NMAE values. The oversimplification of the hypolimnetic oxidation to a single binary factor could be one of the reasons of inaccuracy. We are going to discuss more in the following content. Overall, the evaluating metrics in both training and testing datasets are comparable, and in most of lakes the testing data even show better accuracy. Thus, overfitting issue derived from training/testing data split can be ignored in this study. In particular, the 2-step mixed model further narrow down the difference in evaluating metrics between training and testing data compared to direct LSTM models (Figs. 31-42)

- **Hypoxia detection:**

In general, all three model approaches captured the low bottom DO values during the stratified season (Fig. 43-45). In Lake Erken and Müggelsee, models were able to reproduce the trends in declining bottom DO concentration, but the magnitudes of simulated bottom DO decline were not enough to be counted as hypoxia events (e.g., the hypoxia in the July of 2019 in Lake Erken and in the June of 2020 in Lake Müggelsee). In Lake Müggelsee, the hypolimnetic hypoxia in the stratified season of 2020 was only captured by 2-step mixed model, which also showed better representation of bottom DO variation in the whole testing period than direct LSTM and GBR approaches (Fig. 44), and the 2-step mixed model outperformed the other two approaches in Lake Müggelsee with averaged TPR over 70% and highest TPR closed to 90 % (Fig. 46b). In Lake Erken, direct LSTM presented the highest averaged (~ 70%) and optimal (~ 80%) TPRs of hypoxia detection among the three approaches (Fig. 46a). In these two lakes, the FPRs of three approaches are below 10 %, indicating that the possibilities of sending the wrong hypoxia warning are low (Fig. 46d, e).

Compared to Lake Erken and Müggelsee, hypoxia at the bottom of Lake Mendota lasted almost the entire summer stratified period every year in the testing period. For this lake three ML approaches do not show any significant difference (Fig. 45). The TPRs are relatively higher in Lake Mendota than the other two lakes with mean over 80% for all three ML approaches, presumably due to the longer and more stable hypoxia condition during stratified season in the bottom of Lake Mendota (Fig. 45, 46c). However, the concomitant higher FPRs (over 30%) due to slight errors in the timing of the decline and rising in DO during spring and fall indicating an incorrect warning of hypolimnetic anoxia is also more likely to be sent out (Fig. 46f). Overall, the majority of the hypoxic period is correctly predicted by all three ML approaches.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

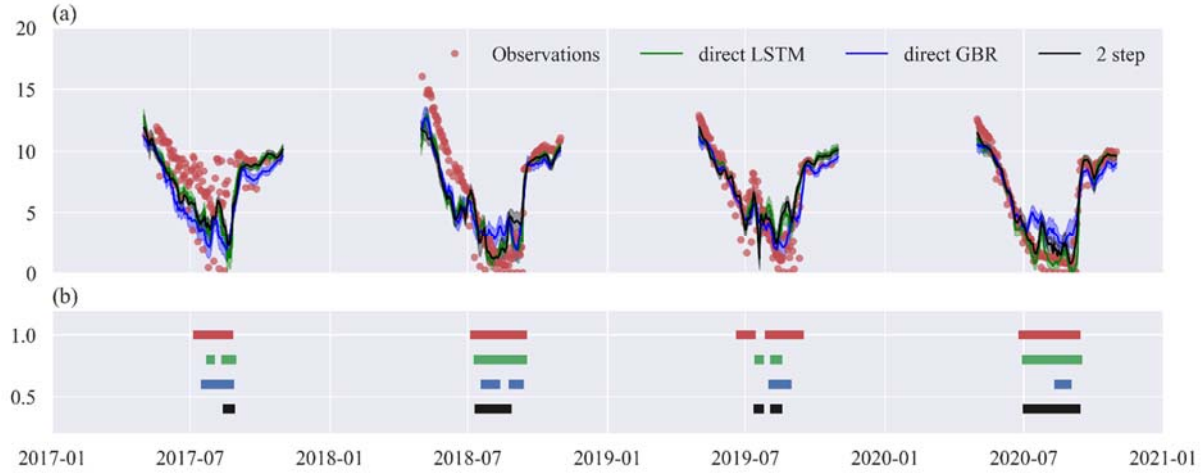


Fig. 43. *(a)Timeseries of YSI observed and modeled (averaged over 30-time model runs) bottom DO concentration in Lake Erken during May to September. The shaded areas represent the 95 % confidential level of each model approach. (b) Observed and modeled hypoxia events in the testing period.*

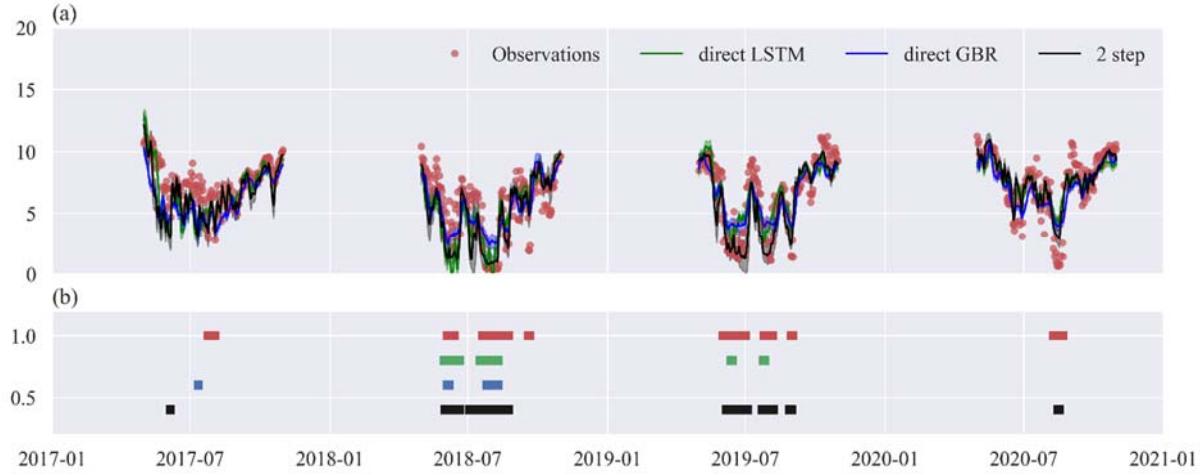


Fig. 44. *(a)Timeseries of YSI observed and modeled bottom DO concentration (May-September) in Lake Müggelsee. The shaded areas represent the 95 % confidential level of each model approach. (b) Observed and modeled hypoxia events in the testing period.*

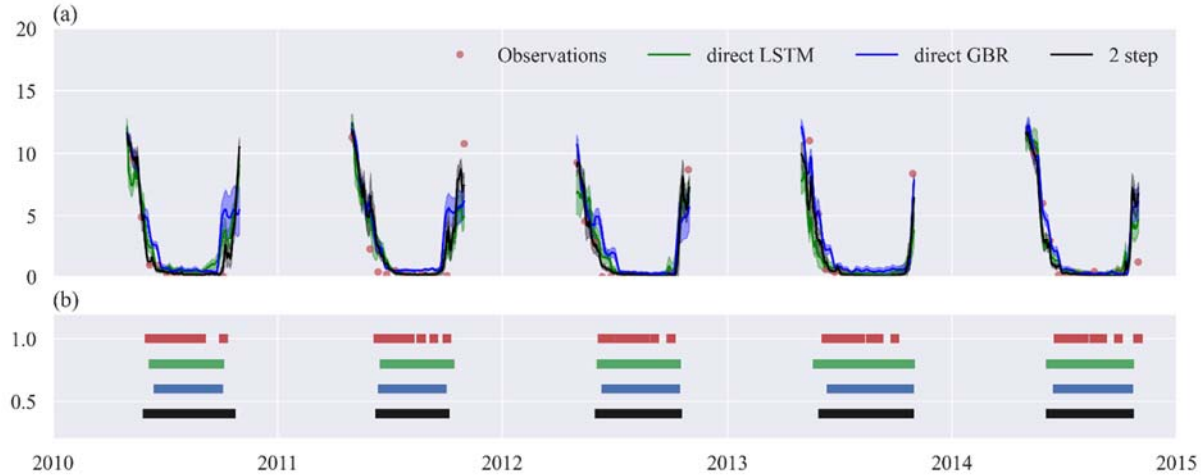


Fig. 45*. (a)Timeseries of YSI observed and modeled bottom DO concentration (May-September) in Lake Mendota. The shaded areas represent the 95 % confidential level of each model approach. (b) Observed and modeled hypoxia events in the testing period.*

SMARTLAGOON

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

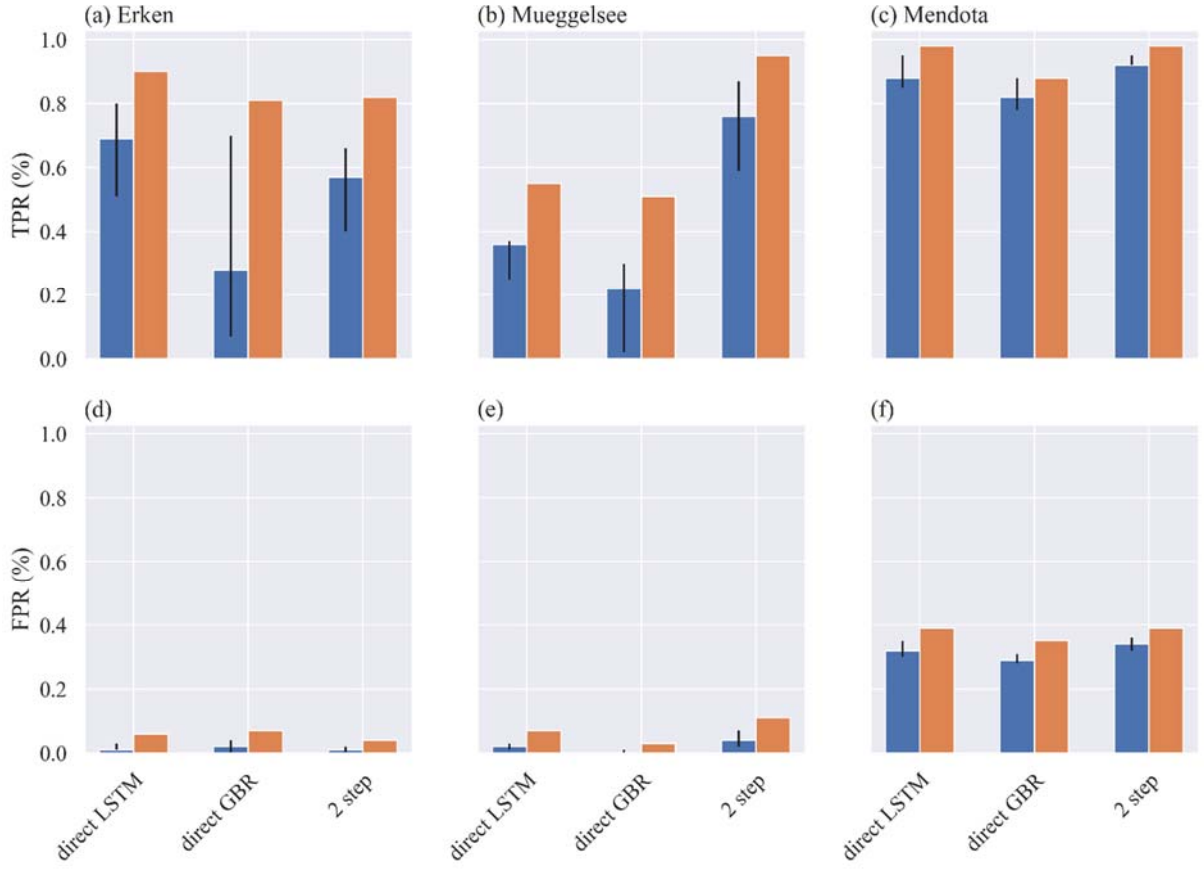


Fig. 46. *Evaluation of hypoxia detection in 3 lakes. The blue bars represent the TPRs and FPRs from the averaged hypolimnetic DO concentration predictions with error bars representing the 95% confidential level, and the orange bars represent the TPRs and FPRs from minimal hypolimnetic DO concentration predictions over the 30-time model runs, indicating the optimal hypoxia detection rates.*

In principle carrying the memory of inputs in the previous week should allow the LSTM model to better represent hypolimnetic DO depletion (Foley et al., 2012). However, due to the more complicated model architecture, it takes longer time to train LSTM model than GBR model, and the model training time is highly dependent on the length of training data and the number of features. We found that the design of 2-step mixed model workflow that first prescreening the training features for the final LSTM model using the feature ranking from the GBR model saves the computational cost. Furthermore, the prescreening seems to lead to an overall improvement in the LSTM performance as can be seen by the difference in the R2 values between direct LSTM and the 2-step workflow ML results in Figs. 31 and 32. The 2-step mixed model workflow which also uses the LSTM model showed better performance in detecting hypoxia events in all three tested lakes where high frequency measurements were available particularly for the polymictic Lake Mueggelsee (Fig. 46). Accuracy of the 2-step approach is promising, with optimal TPRs in Lake Erken, Müegelsee, and Mendota. reaching over 80%.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

- **Feature ranking:**

The GBR model also ranked the dominant features controlling the DO concentration variations and hypoxia events. Table 25 summarizes the top 3 important features for surface and bottom DO concentration simulations from 30-time GBR model runs. It turns out that surface DO concentration is the essential factor for simulating bottom DO in all the tested lakes, demonstrating the necessity of adding the predictive surface DO into training features of bottom DO. In Lake Müggelsee, the only tested polymictic lake, the top 3 important features in simulating surface and bottom DO are consistent over the 30 test runs. The value of the hybrid modeling approach that makes use of information from the process-based models (Fig. 10) is demonstrated by the importance of the derived hydrothermal variables for the prediction of DO. Wn, described as the ratio between the wind friction and the gradient of pressure established by the stratification (Patterson et al., 1984), is the dominant feature in bottom DO simulation except Lake Furesø which has the anthropogenic disturbance in the bottom water environment, also, ΔT, which indicates the intensity of stratification play the major role in predicting the bottom DO in three out of five lakes.

***Table 25**. Top 3 important features in simulating surface and bottom DO in five lakes. The numbers within the parentheses are times of the feature was ranked as the top 3 features in the 30 test runs.*

	Erken	**Müggelsee**	**Furesø**	**Mendota**	**Ekoln**
Surface DO	Accum BotT[1] (30) Outflow (16) Inflow (13) *ThermD* (10) St[2] (9) W_n (6) ΔT (6)	SWR (30) Air temp (30) ΔT (30)	ΔT (30) *ThermD* (30) W_n (15) Inflow (15)	Accum P[3] (30) ice-off date 3[4] (30) Inflow temp (24) Air temp (3) *MLD* (2) OGM_dop[5] (1)	SWR (30) W_n (29) Humidity (22) *ThermD* (9)
Bottom DO	W_n (30) Surface DO (29) Air temp (19) Inflow (8) Accum BotT (4)	Air Temp (30) Surface DO (30) W_n (30)	ΔT (30) Surface DO (28) Oxidization treatment (21) Inflow (11)	ΔT (30) Accum P[3] (29) W_n (21) Surface DO (6) Inflow temp (4)	Surface DO (30) W_n (27) ΔT (21) Air Temp (11) Humidity (1)

1. Accumulated bottom water temperature over 10 days
2. Schmidt stability
3. Accumulated phosphate from river loading
4. Over 30 days from ice-off date
5. Concentration of dissolved organic phosphorus

In Lake Mendota, ΔT and Wn play the major roles in controlling bottom DO, indicating that hypoxia was largely regulated by stratification dynamics. In addition, accumulated phosphate from river loading (Accum P) is also one of the top features for both surface and bottom DO in Lake Mendota, demonstrating that ML models could account for relationships between DO concentration and external loading of nutrients without explicitly specifying the detailed biogeochemical relationships that would be needed in a process-based water quality model.

The results from process-based modeling (Ladwig et al., 2021) revealed that external nutrient loading has a minor effect on the onset and duration of anoxia in the hypolimnion of eutrophic lake like Lake Mendota. However, the feature ranking from our study shows that accumulated external loading of Phosphate (i.e. Accum P) regulates both surface and bottom DO concentration here (Table

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

25). However, these results do not completely contradict each other, since external nutrient loadings may not affect the overall duration of anoxia in the lake but may still play a role in controlling the variations in DO throughout the ice-free period.

Further, by accounting the external factors, like water treatment operation in Lake Furesø as binary training features, three ML approaches can well reproduce the relief of hypoxia in the bottom of lake adapting to the treatment action (Fig. 38). The training data in Lake Furesø spans 20 years (1990-2009) with 7 years (2003-2009) conducting oxidation treatment. Such a long-term historical data recording the conditions with and without oxidation treatment provided the models with sufficient training materials, and the binary classification is easy for ML model, especially GBR, to learn the pattern. However, further quantifying the external factor features may be required for reproducing the abnormalities in 2015 and 2017 observations of bottom DO (Fig. 38).

The observed DO feedbacks to invasive species in Lake Mendota and Müeggelsee are not as clear as that to oxidation treatment in Lake Furesø. The high frequency observations in Müeggelsee did not show a clear changing trend of surface DO (Fig. 39). We observed the slight increase of lowest bottom DO concentration in Lake Müeggelsee during 2011-2017 when Nuttall's waterweed dominated the macrophytes species. All three ML approaches did not clearly capture this minor trend, but they did show the slight relief of hypoxia in 2014-2017 (Fig. 40). The effect of invasive species definitely involves more complex ecological processes which caused the difficulties for ML models to simulate the its impact.

- **Model limitations:**

The major features we used to train the ML models are physical factors (e.g., Wind speed, ΔT, thermD, etc.) which have been proved to largely explain the DO variations (Bouffard et al., 2013; Cortés et al., 2021), but very few of them indicate the external nutrients loading which could have huge impact on oxygen depletion and water quality especially for shallow lakes. The representation of surface DO could be affected by not accounting for oxygen depletion due to the oxidation DOM or oxygen production due to photosynthesis. In Lake Müeggelsee and Mendota, the two most eutrophic lakes, the accuracy of surface DO model was lower than bottom DO model. Presumably, this was related to the lack of training data that would be more directly related to the processes affecting surface photosynthesis and respiration. Since biological factors, like chlorophyll concentrations, are not explicitly considered, the ML models may still represent the effects of ecosystem production implicitly based on relationships with the physical training features, but this is unexplainable and induces uncertainties. Given that the lake nutrients, nutrients loadings, and biological observations are relatively sparse, it is inaccurate to interpolate their daily values so that they can be included into training features and paired to the target DO concentrations. The DO dynamics usually are nonlinear, strongly associated with the seasonal and diurnal variability of lake metabolism which involves complex reactions among physical, chemical, and biological factors. Although the ML approaches are able to reproduce the DO dynamics based on the present training features, lacking biogeochemical factors could induce inaccuracies and uncertainties.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

In addition, this study assumes a fixed daily DO concentration, which in reality can vary even over a short sub-daily time scales. To account for these variations high frequency observations are required for improving the temporal resolution of the model, which could be challenging in most lakes due to lack of high-frequency monitoring. On another hand, the accuracy of process-based models which generate the temperature and eddy diffusivity profiles need to be quantified (Farrell et al., 2020) so that the uncertainties and inaccuracies of the inputs of ML models are understood.

The lakes used to test the model are small-medium lakes with surface area < 50 km2, maximum depth < 50 m and the major hydrodynamic processes resolvable by a 1-D approach. One of our tested lakes, Lake Ekoln, is a basin of Lake Mälaren, the third largest lake in Sweden, and the application of the model in this lake is an example of applying the ML approach to large and complex water system. In this case, training variables representing the circulation pattern or water mass exchange between the target region and other parts of the lake may be required if the interactions among the regions are significant to the hypoxia events. Also, in this case not only the local meteorological conditions, but the meteorological conditions at the other parts of the large lake that could lead to bottom water mass exchange and further trigger the hypoxia issue in the interested region should be considered (Jabbari et al., 2019; Rao et al., 2008).

Since the ML model approaches applied here take the lake as a horizontally uniform system, they only resolve the temporal variations of surface and bottom DO concentrations. However, this assumption may not hold in large water systems with complex transport processes, and the training features (e.g., meteorological inputs, MLD, thermD, etc.) at a single point may not be sufficient to simulate the hypoxia in the whole lake, especially when the hypoxia has the spatial variation and related to the horizontal water mass transportation (Valipour et al., 2021). Simply coupled the ML approaches with the three-dimensional model may not address the issue since more complex ML model architecture design, training features selection, and model workflow are required to resolve the spatial interactions among water columns and spatial variation of DO concentrations.

4.4 Water quality variables predictions

4.4.1 Predictions of suspended sediment load

- ***SWAT calibration:***

Fourteen commonly used streamflow calibration parameters and their ranges were selected based on available studies close to the study area (Epelde et al., 2015; Meaurio et al., 2015) and our previous experience. Following hydrological calibration, sediment calibration was performed using six specific parameters. The calibration process substantially reduced the disparity between the observed and simulated streamflow and SSL. The specific parameters used to calibrate streamflow and SSL are listed in Table 26, which describes each parameter, its range, and its final calibrated value.

Table 26*. The SWAT model parameters used to calibrate streamflow and sediment.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Parameter	Description	Initial range used in calibration	Calibrated value
Parameters used to calibrate streamflow			
r_CN2.mgt	SCS runoff curve number	–0.2 to 0.2	-0.16
v_ESCO.bsn	Soil evaporation compensation factor	0.1 to 1	0.89
v_EPCO.bsn	Plant uptake compensation factor	0.1 to 1	0.56
v_SURLAG.bsn	Surface runoff lag time (days)	0.05 to 24	15.40
v_LAT_TTIME.hru	Lateral flow travel time (days)	0 to 30	2.20
v_ALPHA_BF.gw	Baseflow alpha factor (days-1)	0 to 1	0.53
v_GW_DELAY.gw	Groundwater delay (days)	0 to 100	13
v_GW_REVAP.gw	Groundwater revap coefficient	0.02 to 0.20	0.09
v_GWQMN.gw	Threshold depth of water in the shallow aquifer for return flow to occur (mm)	0 to 5000	1863.12
v_REVAPMN.gw	Threshold depth of water in the shallow aquifer for revap to occur (m)	0 to 1000	252.13
v_RCHRG_DP.gw	Deep aquifer percolation fraction	0 to 1	0.07
r_SOL_AWC.sol	Available water capacity of the soil layer (mm H_2O/mm soil)	–0.2 to 0.2	0.12
r_SOL_BD.sol	Moist bulk density	–0.2 to 0.2	0.11
r_SOL_K.sol	Saturated hydraulic conductivity (mm h-1)	–0.2 to 0.2	0.12
Parameters used to calibrate SSL			
v_CH_COV1.rte	Channel erodibility factor	0 to 1	0.35
v_CH_COV2.rte	Channel cover factor	0 to 1	0.42
v_SPCON.bsn	Linear parameter for calculating the maximum amount of sediment that can be re-entrained during channel sediment routing	0.0001 to 0.01	0.0001
v_SPEXP.bsn	Exponent parameter for calculating sediment re-entrained in channel sediment routing	1 to 1.5	1.45
v_USLE_P.mgt	USLE equation support practice factor	0 to 1	0.85
r_USLE_K.sol	USLE equation soil erodibility (K) factor	-0.2 to 0.2	0.04

The calibrated parameter analysis presented in Table 26 shows that CN2 decreased by 16% compared to the default value, thus increasing infiltration and decreasing runoff. The adjusted ESCO value was 0.89, a high value typical of climates in which evapotranspiration is not highly relevant (Jimeno-Sáez et al., 2018). The LAT_TTIME value was low, indicating that water pathways through the soil profile are short. The RCHRG_DP value was very low at 0.07 and characteristic of land with no relevant groundwater storage (Senent-Aparicio et al., 2019). The fitted values for ESCO, LAT_TTIME, RCHRG_DP, SPCON, and SPEXP in the Oskotz river basin were similar to those used in nearby basins (Epelde et al., 2015; Meaurio et al., 2015).

- **ML model parameters and computational time:**

The following M5P parameters were used in the experiments: a minimum number of four examples for the leaf nodes and no pruning for the trees. Regarding the parameters used in the RF experiments, there were 100 trees (A), each having a maximum depth expansion, and for each node, the number of attributes selected to divide de node was $\log_2(M)+1$, where M was the maximum number of input attributes. The computational times were very low for both the M5P and RF techniques. In the scenario with the most variables, the model building times for M5P and RF were 0.33 and 0.58 seconds respectively. The inference was 0.08 and 0.03.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

- **SWAT streamflow estimation:**

The calibrated SWAT model simulated daily streamflow using R2 values of 0.71 and 0.79 for the calibration and validation periods, respectively. The RSR, NSE, and PBIAS were respectively 0.54, 0.70, and -20.95% for the calibration period and 0.48, 0.76, and -24.38% for the validation period. At the monthly scale, the performance statistics were more satisfactory, displaying R2 values of 0.84 for calibration and 0.94 for validation. The monthly RSR, NSE, and PBIAS values were 0.46, 0.81, and -20.55% for the calibration period and 0.36, 0.86, and -24.85% for the validation period. Therefore, the model was evaluated as very good regarding monthly streamflow simulation during both calibration and validation, according to the RSR and NSE criteria described by Moriasi et al. (2007). The good agreement between simulated and observed monthly streamflow is presented graphically in Fig. 47, which displays both the calibration and validation periods. However, the model was classified as satisfactory according to PBIAS (Moriasi et al., 2007) since calibration and validation both obtained negative values, indicating that the model overestimated the streamflow (see Fig. 47). However, the simulated streamflow fitted the observed flows very satisfactorily, matching the low flows and most of the peaks well.

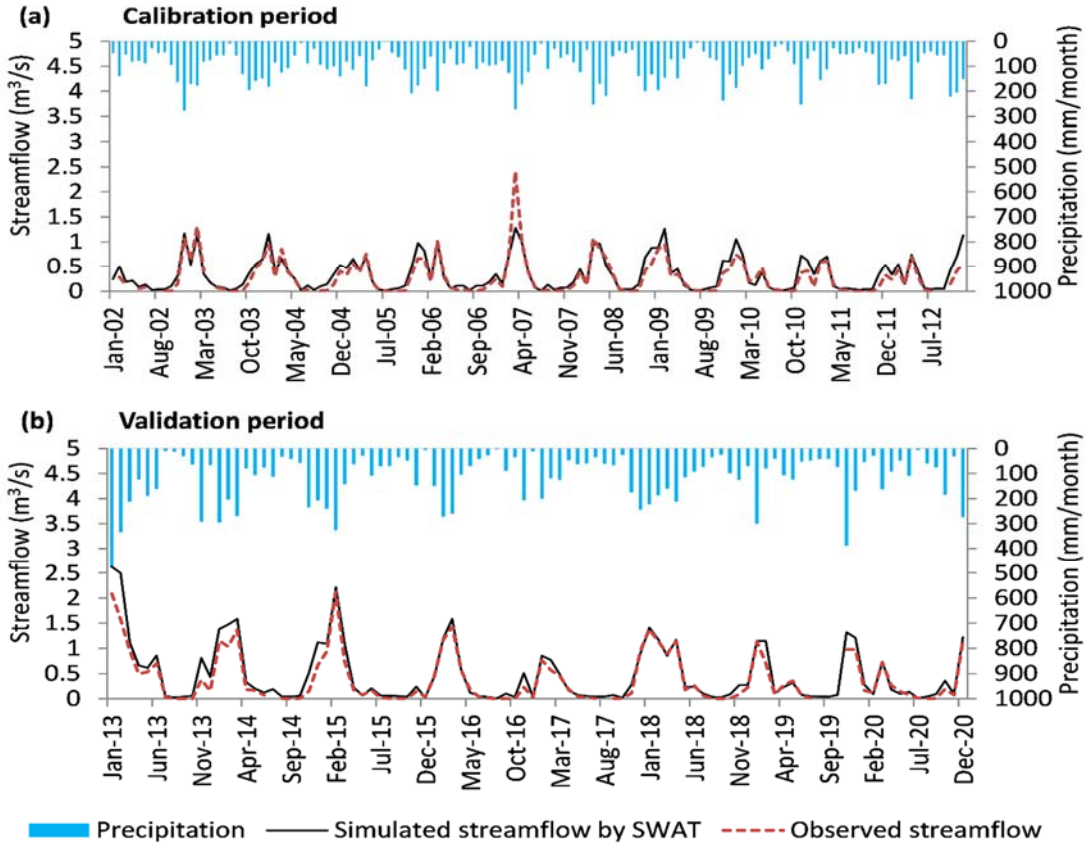


Fig. 47. *Monthly precipitation, observed streamflow, and streamflow simulated by SWAT at a hydrological station for (a) the calibration period and (b) the validation period.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

- **SWAT and ML sediment estimation:**

Nine combinations of inputs were used in the ML models. The daily calibration and validation results for each model are presented in Table 27. In all scenarios, the M5P and RF models obtained better daily performance metrics than the SWAT model, during both calibration and validation. In the training (calibration) phase, RF performed particularly well compared to the other models. However, RF was less reliable in the test phase (validation) due to model overfitting during training. This problem can occur if the model is not tested using cross-validation (Cai et al., 2020). The M5P models were more stable, demonstrating similar results during both phases. Compared to the RF models, the M5P models improved the test-phase statistics in many scenarios, indicating that the model did not overfit during the training phase. Regarding the different scenarios, at the daily level, scenario III, which included the three variables most correlated with sediment, presented the best results during M5P and RF model validation. Scenario V obtained one of the best results for both models, albeit during calibration only. Scenario IX was the most successful of the nine scenarios for both ML models in both phases (training and validation). The SWAT model did not reach the NSE value of 0.3. However, the ML models obtained an NSE value greater than 0.3 in all cases, reaching the satisfactory classification according to the daily criteria established by Kalin et al. (2010). The two ML models and SWAT achieved PBIAS values of less than 25%, the criterion for very good daily models set by Kalin et al. (2010).

***Table 27**. The daily performance of the models in SSL estimation.*

Input scenario	Model	Performance metrics Calibration (Validation)				
		R^2	MAE (ton/day)	RSR	NSE	PBIAS (%)
-	SWAT	0.27 (0.28)	5.64 (11.17)	0.86 (0.86)	0.28 (0.26)	-7.60 (-11.98)
I	M5P	0.41 (0.49)	4.63 (8.83)	0.77 (0.75)	0.41(0.44)	-5.25 (-5.79)
	RF	0.88 (0.31)	2.15 (10.14)	0.37 (0.83)	0.86 (0.32)	1.58 (-7.55)
II	M5P	0.41 (0.48)	4.62 (8.86)	0.77 (0.75)	0.41 (0.44)	-6.12 (-6.84)
	RF	0.90 (0.41)	1.97 (8.95)	0.36 (0.78)	0.87 (0.39)	0.12 (2.11)
III	M5P	0.50 (0.51)	4.12 (8.02)	0.71 (0.71)	0.50 (0.50)	-0.60 (0.50)
	RF	0.91 (0.45)	1.94 (8.58)	0.35 (0.77)	0.88 (0.41)	-0.47 (4.79)
IV	M5P	0.42 (0.50)	4.55 (8.73)	0.76 (0.75)	0.42 (0.44)	-3.11 (-1.23)
	RF	0.89 (0.37)	1.99 (9.52)	0.37 (0.80)	0.86 (0.35)	0.93 (-2.78)
V	M5P	0.54 (0.44)	4.03 (8.93)	0.68 (0.75)	0.54 (0.44)	-0.21 (-17.87)
	RF	0.92 (0.41)	1.91 (8.94)	0.34 (0.79)	0.88 (0.38)	-0.73 (0.40)
VI	M5P	0.42 (0.50)	4.55 (8.73)	0.76 (0.75)	0.42 (0.44)	-3.11 (-1.23)
	RF	0.92 (0.34)	1.92 (9.23)	0.35 (0.82)	0.88 (0.33)	1.2 (0.99)
VII	M5P	0.42 (0.50)	4.55 (8.74)	0.76 (0.75)	0.42 (0.44)	-3.06 (-1.25)
	RF	0.91 (0.41)	1.91 (8.85)	0.35 (0.79)	0.88 (0.38)	-0.4 (2.06)
VIII	M5P	0.43 (0.49)	4.70 (8.87)	0.75 (0.75)	0.43 (0.44)	-7.13 (-3.71)
	RF	0.90 (0.43)	1.95 (8.80)	0.34 (0.78)	0.88 (0.40)	-1.91 (1.60)
IX	M5P	0.68 (0.51)	3.55 (8.10)	0.58 (0.71)	0.66 (0.50)	0.68 (-6.80)
	RF	0.92 (0.43)	1.96 (8.77)	0.34 (0.78)	0.88 (0.39)	-1.4 (-2.52)

The statistics in Table 28 indicate that the SWAT and ML approaches performed well in SSL estimation at the monthly scale. According to criteria set by Moriasi et al. (2007), SWAT simulated monthly SSL well during calibration and satisfactorily during validation. Conversely, PBIAS had very good values in both cases. The monthly SWAT results were significantly better than the daily results. Similar findings have been observed in other studies (Choukri et al., 2020; Nunes et al., 2018), in which

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

problems in SWAT's sediment transport module caused inaccuracies in daily sediment estimates, which were averaged out and smoothed on a monthly scale. The M5P model was superior to SWAT during calibration in scenarios III, V, and IX but better in all scenarios during validation. In all cases, RF outperformed SWAT. As expected, the best results for both models were obtained in scenarios III, V, and IX, which is similar to the daily scale results. Again, both ML models performed best in scenario IX, with performance statistics classified as very good during calibration and good during validation, according to the monthly criteria established by Moriasi et al. (2007).

Table 28. *The monthly performance of the models in SSL estimation.*

Input scenario	Model	Performance metrics Calibration (Validation)				
		R^2	MAE (ton/day)	RSR	NSE	PBIAS (%)
-	SWAT	0.75 (0.57)	116.20 (181.24)	0.52 (0.65)	0.72 (0.57)	1.25 (-1.27)
I	M5P	0.62 (0.62)	118.25 (153.12)	0.62 (0.62)	0.62 (0.62)	8.55 (4.27)
	RF	0.90 (0.58)	59.97 (161.97)	0.40 (0.60)	0.84 (0.58)	12.13 (1.08)
II	M5P	0.64 (0.64)	116.72 (153.09)	0.60 (0.61)	0.64 (0.63)	7.75 (3.30)
	RF	0.92 (0.59)	59.81 (162.63)	0.37 (0.65)	0.86 (0.57)	9.79 (12.25)
III	M5P	0.75 (0.72)	101.08 (136.08)	0.50 (0.55)	0.74 (0.70)	8.83 (6.04)
	RF	0.94 (0.65)	58.33 (157.04)	0.35 (0.63)	0.88 (0.60)	8.62 (14.24)
IV	M5P	0.62 (0.62)	118.70 (157.55)	0.62 (0.63)	0.61 (0.60)	11.21 (7.76)
	RF	0.92 (0.61)	58.17 (158.25)	0.38 (0.63)	0.86 (0.60)	11.19 (4.39)
V	M5P	0.78 (0.69)	92.90 (142.02)	0.46 (0.56)	0.78 (0.69)	4.86 (-4.48)
	RF	0.93 (0.69)	60.56 (146.24)	0.36 (0.58)	0.87 (0.66)	9.36 (8.99)
VI	M5P	0.62 (0.62)	118.70 (157.55)	0.62 (0.63)	0.61 (0.60)	11.21 (7.76)
	RF	0.93 (0.64)	59.58 (157.31)	0.39 (0.61)	0.85 (0.62)	11.66 (10.41)
VII	M5P	0.62 (0.62)	118.68 (157.60)	0.62 (0.63)	0.61 (0.60)	11.25 (7.74)
	RF	0.92 (0.67)	59.35 (151.35)	0.38 (0.56)	0.85 (0.63)	10.31 (10.63)
VIII	M5P	0.66 (0.65)	119.16 (159.31)	0.60 (0.57)	0.64 (0.62)	8.14 (5.06)
	RF	0.92 (0.72)	60.45 (145.51)	0.38 (0.53)	0.86 (0.67)	8.83 (10.10)
IX	M5P	0.83 (0.73)	89.13 (135.04)	0.43 (0.54)	0.81 (0.71)	8.90 (5.19)
	RF	0.94 (0.72)	57.95 (143.39)	0.35 (0.57)	0.88 (0.67)	8.56 (11.60)

The monthly results of the SWAT and ML models for the best-case scenario (scenario IX) are presented graphically in Fig. 48 and Fig. 49. The temporal variations in monthly SSL at the Oskotz river basin are given for the calibration period (Fig. 48) and the validation period (Fig. 49). The effect of precipitation variability on SSL production can be observed in Fig. 50, which presents the mean monthly values for precipitation, streamflow, and SSL in the basin for the calibration and validation periods.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

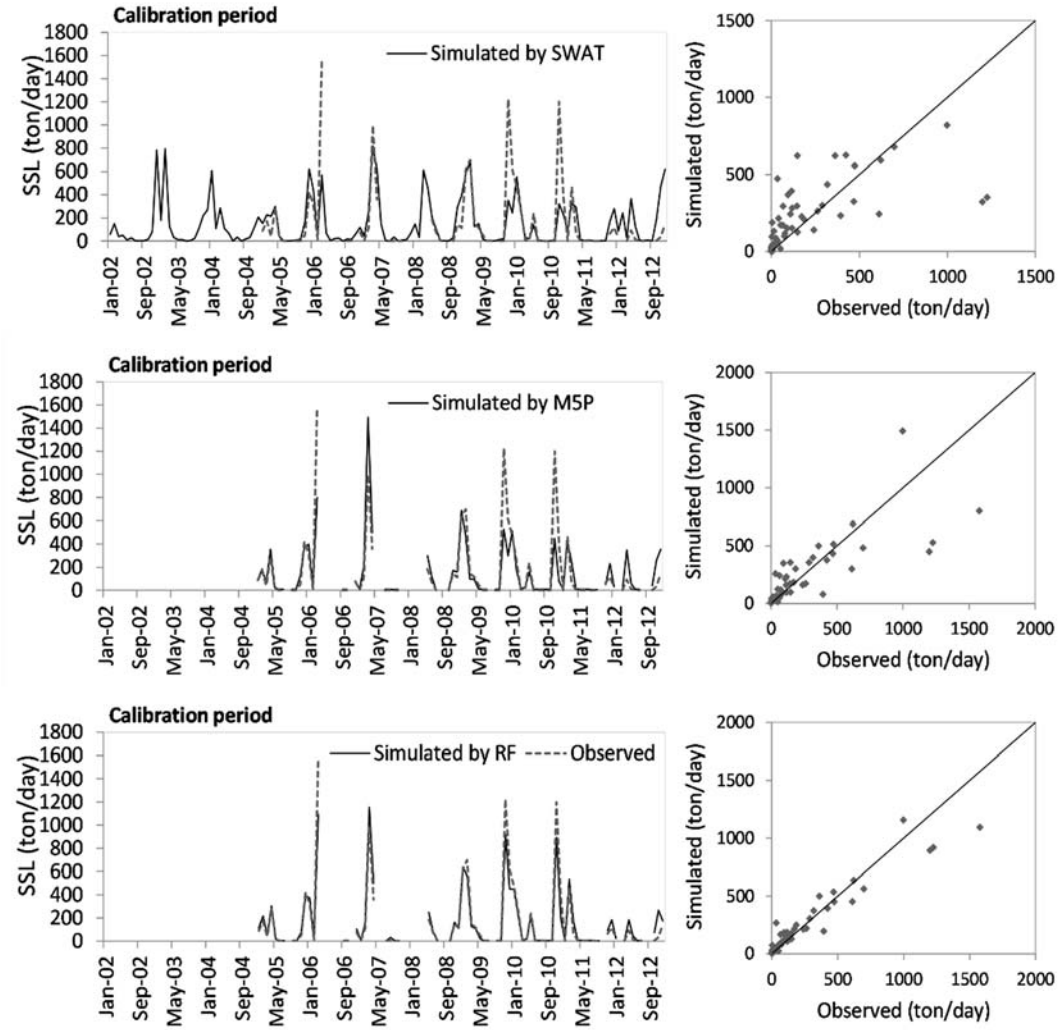


***Fig. 48**. A comparison of observed SSL and monthly SSL simulated using SWAT, M5P, and RF during the calibration period.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

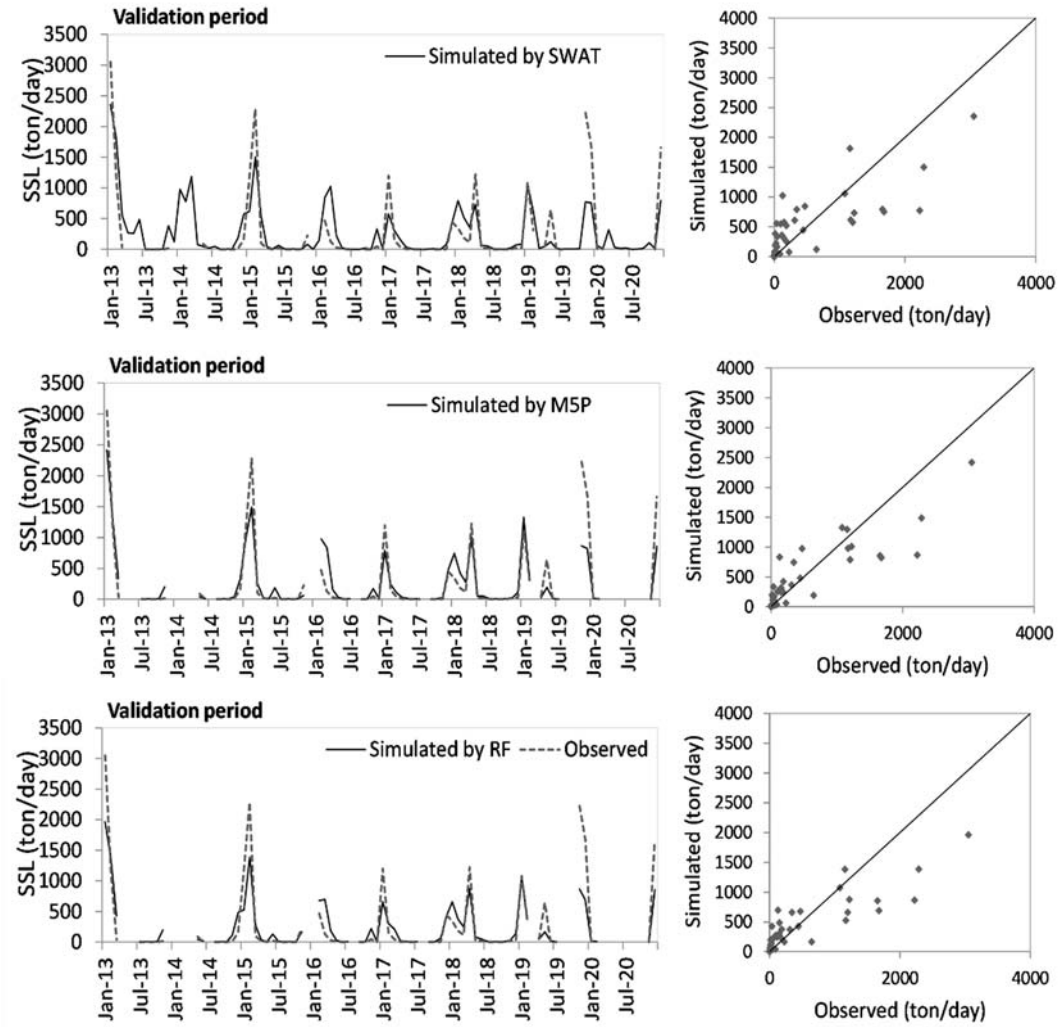


Fig. 49. *A comparison of observed SSL and monthly SSL simulated using SWAT, M5P, and RF during the validation period.*

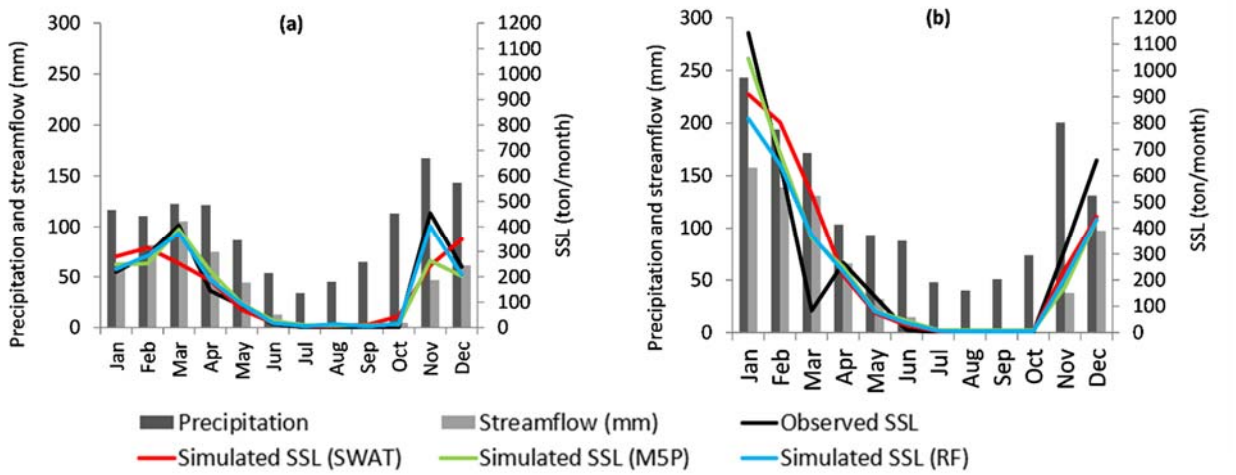


Fig. 50*. The seasonal distribution of observed precipitation, streamflow, and SSL and SSL simulated using SWAT and ML models for (a) calibration and (b) validation.*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

5. Conclusions

5.1 Flood predictions

The Campo de Cartagena region (draining basin to the Mar Menor) is characterized by an arid climate with limited rainfall. However, when torrential rains, typical of this climate, occur in the area, they can have severe consequences, including flooding causing, among other damages, the contamination of the Mar Menor. This influx of pollutants can disrupt the delicate ecological balance of the Mar Menor, negatively impacting its water quality, marine life, and overall ecosystem health.

The use of models to help us predict and assess the effects of storms and heavy rainfall in the region is of vital importance. Through modelling the hydrographs, it is possible to estimate the magnitude and duration of peak flows, as well as the velocity and quantity of runoff flowing into the Mar Menor. This provides valuable information for flood management and decision making related to the protection and conservation of this fragile ecosystem as well as avoiding damage to the population of the area. In addition, modelling of discharge after heavy rainfall events can contribute to the identification of risk areas and the planning of flood mitigation and control measures. This includes the design of appropriate infrastructures, such as diversion channels or dams, to regulate water flow and reduce negative impacts on the Mar Menor and surrounding areas.

In this WP3 (Task 3.3) the discharge of an ephemeral stream draining into the Mar Menor was modelled using physically based models such as the SWAT hydrological model and machine learning models. Modelling streamflow using hydrological models and machine learning techniques offers distinct advantages and disadvantages. Hydrological models, grounded in physical principles, provide a solid theoretical foundation, and enable the interpretation and explanation of underlying hydrological processes. They can be generalized to different conditions and used for simulating extreme events, such as floods. However, they require substantial data for calibration and validation and can be complex to implement. Additionally, hydrological models may face challenges in capturing patterns and trends in non-stationary conditions. On the other hand, machine learning techniques excel at modeling complex relationships and can handle large datasets effectively. They are not dependent on specific assumptions about physical processes, which is advantageous in uncertain or poorly understood hydrological contexts. However, machine learning models often lack interpretability and explanation, and the quality and completeness of input data greatly influence their performance. Overfitting is another concern, as these models may excessively fit training data and struggle to generalize to unseen conditions. In conclusion, a combination of hydrological models and machine learning techniques can leverage the strengths of each approach, offering a comprehensive framework for streamflow modeling and analysis.

Specifically, the results obtained in this task showed better statistics for ML than for SWAT. The three models tested, LSTM, CNN and GRU seem to be very promising but still work in progress due to lack of data and these results need to be validated.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

5.2 Algal bloom predictions

The GBR and LSTM models successfully reproduced the Erken Lake seasonal Chl patterns with all three workflows tested, capturing spring and summer bloom events with better averaged statistics than those obtained with a PB model. And in all three workflows, the LSTM model always showed slightly better performance in predicting Chl concentrations than the GBR model and better detection of the onset of algal bloom events.

Workflow 1, which predicted Chl based on all available environmental factors including lake nutrient observations, showed that both ML models can reproduce the seasonal dynamics of algal Chl with promising accuracy via the direct input of available environmental observations. These ML models can be applied to reconstruct past patterns of algal Chl, fill the gaps between measured Chl observations, and interpret the mechanisms that drive phytoplankton dynamics. Workflows 2 and 3 adopted a two-step approach, first using separate ML models to estimating daily changes in lake nutrient concentration and in Workflow 3 also including PB model derived physical factors as training features of the algal ML model. These two workflows allowed for daily predictions of changes in algal Chl concentration using both observations and pre-generated lake nutrient concentrations at a consistent daily time step, and at only a minor decrease in performance compared to workflow 1, workflow 2 and 3 demonstrated a wider potential range of applications (e.g. interpolation, reconstruct historical data, and algal bloom forecast) via making daily forecasts with less-than-daily measured nutrient observations.

For the purposes of water management, it may be most important to first predict the potential occurrence of a bloom and then once underway improve predictions of its magnitude. For predicting the onset of algal bloom, the best result was obtained after adding hydrodynamic features of the PB model in workflow 3, with a moderate-strong level of prediction. The correct detection of algal blooms is highly dependent on the years used to train the models. Thus, while the ML models can be better than the PB models at predicting the onset of algal blooms, they still may not be good enough for operational forecasting.

To reach the goal of incorporating ML models into operational forecasts either for short-term management support or longer-term evaluation and planning, two steps must occur. First the ML model must be developed, trained, and evaluated on the water body of interest due to the unique physical characteristics and water quality dynamics in different systems. Secondly, future forcing data for the model must be obtained and integrated into a workflow that makes the future predications. In regards to the second point, a lack of frequent water monitoring (Stanley et al., 2019) is a major deterrence to applying ML models to many lakes. The data sparsity test showed that, at least for Lake Erken, the ML models can still detect the seasonal algal dynamics even for sample intervals approaching 1 month. If this result holds for other lakes, the use of the two-step ML workflow could offer a method of forecasting seasonal variations in algal Chl, even in lakes with relatively infrequent nutrient monitoring but higher frequency meteorological and hydrological data, as is the case of the Mar Menor lagoon.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

The hybrid PB/ML models have the potential to provide reasonably accurate and timely short-term algal bloom forecasts, working as part of an early-warning system for the water resource management (Baracchini et al., 2020), and clearly have the ability to predict border seasonal variations in algal Chl concentration. However, since a large number of water temperature and water quality samples are required for ML training, and since our results only apply to one well studied lake, obtaining more datasets to test and evaluate the workflows developed here is necessary. Monitoring networks (e.g., Global Lake Ecological Observatory Network, GLEON; https://gleon.org/, last access: 10 June 2022), could provide the data to allow more extensive testing and application of hybrid PB/ML models, and we are presently working in the GLEON network to test the methods developed in this paper on many other lakes.

Due to the scarcity of data currently available for the Mar Menor lagoon, remore sensing data were analysed. Remote sensing, thanks to its ubiquity, offers a powerful tool for monitoring large natural areas. Improvements in the performance of these systems offer systems with lower and lower spatial and temporal resolution, providing the opportunity to develop warning systems for the conservation of natural areas. However, satellite systems are global and their signals need to be reprocessed to increase their accuracy in more local/regional contexts. Different machine learning and deep learning models have been analysed to improve the accuracy of Chl-a measurements from the reflectances of the European Sentinel-3 observation system in a particular highly anthropised context such as the Mar Menor lagoon. Our results show that the Sentinel-3 reflectance data correlate reasonably well with the available in situ observations. This good correlation suggested that using remote sensing it was possible to obtain a good estimate of Chl-a concentration in the Mar Menor. The presented CNN model represents a global model that can handle well the remote sensing information in this environment. Therefore, it can provide a unique model capable of describing the whole lagoon. However, there is still a lack of accuracy at near-shore or shallow points, so a better method for estimating values in these cases will have to be sought. This fact seems to clash with the findings that best model results occur using surface measurements, but it is due to the reflectance light phenomenon in water and is beyond the scope of the present research. Moreover, we acknowledge that only Chl-a has been targeted and other parameters such as turbidity can be included in the model to provide a multivariate model of the lagoon. These other parameters are being studied for future work as well as test the CNN model approach on other similar environments suggest an interesting address.

5.3 Hypoxia predictions

The mechanisms that lead to variations in lake DO concentration and hypolimnetic hypoxia are complex and vary from lake to lake. This study tested the performance of ML approaches in simulating surface and bottom DO concentration in various lake systems. In addition to meteorological and hydrological inputs, hydrodynamic and ice-related variables, as well as other external disturbance factors (e.g., water treatment action, invasive species) have been considered in the model training.

As the key parameter to indicate water quality, DO concentration particularly changes in bottom water DO is fundamental for understanding the biogeochemical processes in the lake systems.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Based on two machine learning models, Gradient Boost Regressor (GBR) and Long-short-term-memory (LSTM) network, this study developed three machine learning (ML) model approaches (i.e., direct GBR, direct LSTM, and 2-step mixed ML model workflow) to simulate the multi-year surface and bottom DO concentrations in 5 lake systems.

Successful ML predictions depend on the availability of long-term high-frequency data sets that can serve as training features for the ML algorithms. This can limit training data to measurements of meteorology and stream discharge which are routinely available at a daily measurement frequency, and have been collected over long historical periods. Here we demonstrate a hybrid modeling approach that uses these same model inputs to first force a simple 1-D hydrodynamic model to provide additional information describing the thermal structure and mixing dynamics of lakes. This hybrid model workflow therefore preprocesses available environmental data to provide additional information that is known to influence the ecology and biogeochemistry of lakes, which has been demonstrated to be successful in algal bloom prediction in Lake Erken. In this task, hydrodynamic training features (e.g. Wn and ΔT) were found to be of importance for the prediction of DO in all lakes studied. This was particularly true for the bottom DO that would be of greatest interest for water management. The feature ranking provided by GBR model in the mixed model workflow can also help to understand the interactions between DO dynamics and physical or biogeochemical processes, to calibrate the process-based numerical models, and to better design process-based models in the specific water systems.

The three ML approaches tested in the study are shown to be powerful tools for reproducing and predicting DO concentrations, opening the possibility for similar algorithms to be incorporated into forecasting workflows that would predict lake DO and the onset loss and duration of hypoxia events. Such a forecasting system could be the cost-effective choice for early warning and short-term forecast of anoxia events, supporting the decision making in the drinking water plants, and providing the reference of potential fish habitat loss and internal loading (Nürnberg et al., 2013; Orihel et al., 2017). In the long term, the warmer lake surface temperature and stronger stratification are the undergoing changes as a result of climate warming for numerous lakes across the world, resulting in the potential increases of the occurrence, duration, and extent of hypoxia in the hypolimnion (Ladwig et al., 2021; North et al., 2014), the ML approaches could also play a part in projections of DO under the future climate (Jane et al., 2021). The results also suggest that surface DO concentration is an important feature for predicting bottom DO and hypoxia in our workflow. Addition of this single easily automated measurement into a water quality modeling program could therefore, potentially improve the accuracy of future forecast systems.

5.4 Other water quality variables predictions

- **Suspended sediment load prediction**

Due to the non-linear behavior of sediment transport, ML algorithms demonstrate considerable potential for accurately estimating sediment loads. This study has explored the performance of two

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

ML models (M5P and RF) and compared them to the SWAT hydrological model for SSL simulation, using data collected in an experimental basin.

The sediment loads estimated by the three models generally provide satisfactory approximations. Regarding the ML models, nine input scenarios were tested exclusively using daily precipitation and streamflow as input data and considering scenarios with and without past data. The best scenario included streamflow and precipitation data from day t to day t-3 as inputs. Moreover, both ML models were equally valid with similar statistics. At the daily scale, the ML models outperformed the SWAT hydrological model in all scenarios. At the monthly scale, both ML models achieved better validation results than SWAT. The ML models significantly reduced the time and computational effort required compared to SWAT, making sediment estimation easier. Since ML models are data-driven, they can be trained to describe complex processes without spatial data. In contrast, SWAT requires multiple spatial data and a complete description of the physical processes that govern the hydrological behaviour of a basin.

Overall, the results indicate that ML techniques provide a more accurate prediction of SSL than the SWAT model. Therefore, for water basin managers and stakeholders, ML can be a useful method for simulating sediment production, analyzing soil degradation, and designing appropriate measures for soil and water conservation. Such models could be particularly useful in basins where limited spatial data are available or knowledge of the processes taking place in the basin is limited or unknown. The novelty of this study is that it is the first to compare the performance of SWAT with M5P and RF models to estimate SSL. Moreover, no previous studies using these models have been carried out in the study area. Furthermore, SSL estimation can be extended to other ML models or the estimation of other water quality variables. In addition, future research could improve the models by exploring the influence of other inputs related to uptake and sedimentation processes.

APPENDIX I: Bibliography

Abbaspour, K.C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., Kløve, B., 2015. A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. Journal of Hydrology 524, 733–752. https://doi.org/10.1016/j.jhydrol.2015.03.027

Abbaspour, K.C., Vejdani, M., Haghighat, S., 2007. SWAT-CUP calibration and uncertainty programs for SWAT. In Proceedings of the Modsim 2007: International Congress on Modelling and Simulation, Christchurch, New Zealand, 3–8 December 2007; pp. 1603–1609.

Adrian, R., Wilhelm, S., Gerten, D., 2006. Life-history traits of lake plankton species may govern their phenological response to climate warming. Global Change Biology 12, 652–661. https://doi.org/10.1111/j.1365-2486.2006.01125.x

Alcolea, A., Contreras, S., Hunink, J.E., García-Aróstegui, J.L., Jiménez-Martínez, J., 2019. Hydrogeological modelling for the watershed management of the Mar Menor coastal lagoon (Spain). Science of The Total Environment 663, 901–914. https://doi.org/10.1016/j.scitotenv.2019.01.375

Al-Mukhtar, M., 2019. Random forest, support vector machine, and neural networks to modelling suspended sediment in Tigris River-Baghdad. Environ. Monit. Assess. 12.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Álvarez-Rogel, J., Barberá, G.G., Maxwell, B., Guerrero-Brotons, M., Díaz-García, C., Martínez-Sánchez, J.J., Sallent, A., Martínez-Ródenas, J., González-Alcaraz, M.N., Jiménez-Cárceles, F.J., Tercero, C., Gómez, R., 2020. The case of Mar Menor eutrophication: State of the art and description of tested Nature-Based Solutions. Ecological Engineering 158, 106086. https://doi.org/10.1016/j.ecoleng.2020.106086

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development. J. Am. Water Resour. Assoc. 34, 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Arnold, J.G., Youssef, M.A., Yen, H., White, M.J., Sheshukov, A.Y, Sadeghi, A.M, Moriasi, D.N., Steiner, J.L, Amatya, D.M., Skaggs, R.W, Haney, E.B., Jeong, J., Arabi, M., Gowda, P.H, 2015. Hydrological Processes and Model Representation: Impact of Soft Data on Calibration. Trans. ASABE 58, 1637–1660. https://doi.org/10.13031/trans.58.10726

Baracchini, T., Wüest, A., Bouffard, D., 2020. Meteolakes: An operational online three-dimensional forecasting platform for lake hydrodynamics. Water Research 172, 115529. https://doi.org/10.1016/j.watres.2020.115529

Borrelli, P., Alewell, C., Alvarez, P., Anache, J.A.A., Baartman, J., Ballabio, C., Bezak, N., Biddoccu, M., Cerdà, A., Chalise, D., Chen, S., Chen, W., De Girolamo, A.M., Gessesse, G.D., Deumlich, D., Diodato, N., Efthimiou, N., Erpul, G., Fiener, P., Freppaz, M., Gentile, F., Gericke, A., Haregeweyn, N., Hu, B., Jeanneau, A., Kaffas, K., Kiani-Harchegani, M., Villuendas, I.L., Li, C., Lombardo, L., López-Vicente, M., Lucas-Borja, M.E., Märker, M., Matthews, F., Miao, C., Mikoš, M., Modugno, S., Möller, M., Naipal, V., Nearing, M., Owusu, S., Panday, D., Patault, E., Patriche, C.V., Poggio, L., Portes, R., Quijano, L., Rahdari, M.R., Renima, M., Ricci, G.F., Rodrigo-Comino, J., Saia, S., Samani, A.N., Schillaci, C., Syrris, V., Kim, H.S., Spinola, D.N., Oliveira, P.T., Teng, H., Thapa, R., Vantas, K., Vieira, D., Yang, J.E., Yin, S., Zema, D.A., Zhao, G., Panagos, P., 2021. Soil erosion modelling: A global review and statistical analysis. Sci. Total Environ. 780, 146494. https://doi.org/10.1016/j.scitotenv.2021.146494

Bouffard, D., Ackerman, J.D., Boegman, L., 2013. Factors affecting the development and dynamics of hypoxia in a large shallow stratified lake: Hourly to seasonal patterns. Water Resources Research 49, 2380–2394. https://doi.org/10.1002/wrcr.20241

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

Breiman, L., 2017. Classification and Regression Trees. Routledge, New York. https://doi.org/10.1201/9781315139470

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brookes, J.D., Carey, C.C., 2011. Resilience to Blooms. Science 334, 46–47. https://doi.org/10.1126/science.1207349

Bruggeman, J., Bolding, K., 2014. A general framework for aquatic biogeochemical models. Environmental Modelling & Software 61, 249–265. https://doi.org/10.1016/j.envsoft.2014.04.002

Burchard, H., Bolding, K., Villarreal, M.R., 1999. GOTM, a General Ocean Turbulence Model: Theory, Implementation and Test Cases. Space Applications Institute.

Burford, M.A., Carey, C.C., Hamilton, D.P., Huisman, J., Paerl, H.W., Wood, S.A., Wulff, A., 2020. Perspective: Advancing the research agenda for improving understanding of cyanobacteria in a future of global change. Harmful Algae, Climate change and harmful algal blooms 91, 101601. https://doi.org/10.1016/j.hal.2019.04.004

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Caballero, I., Roca, M., Santos-Echeandía, J., Bernárdez, P., Navarro, G., 2022. Use of the Sentinel-2 and Landsat-8 Satellites for Water Quality Monitoring: An Early Warning Tool in the Mar Menor Coastal Lagoon. Remote Sensing 14, 2744. https://doi.org/10.3390/rs14122744

Cai, J., Xu, K., Zhu, Y., Hu, F., Li, L., 2020. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. Appl. Energy 262, 114566. https://doi.org/10.1016/j.apenergy.2020.114566

Carey, C.C., Ibelings, B.W., Hoffmann, E.P., Hamilton, D.P., Brookes, J.D., 2012. Eco-physiological adaptations that favour freshwater cyanobacteria in a changing climate. Water Research, Cyanobacteria: Impacts of climate change on occurrence, toxicity and water quality management 46, 1394–1407. https://doi.org/10.1016/j.watres.2011.12.016

Carpenter, S.R., Kitchell, J.F., 1988. Consumer Control of Lake Productivity: Large-scale experimental manipulations reveal complex interactions among lake organisms. BioScience 38, 764–769. https://doi.org/10.2307/1310785

Casalí, J., Gastesi, R., Álvarez-Mozos, J., De Santisteban, L.M., Lersundi, J.D.V. de, Giménez, R., Larrañaga, A., Goñi, M., Agirre, U., Campo, M.A., López, J.J., Donézar, M., 2008. Runoff, erosion, and water quality of agricultural watersheds in central Navarre (Spain). Agric. Water Manage. 95, 1111–1128. https://doi.org/10.1016/j.agwat.2008.06.013

Casalí, J., Giménez, R., Díez, J., Álvarez-Mozos, J., Del Valle de Lersundi, J., Goñi, M., Campo, M.A., Chahor, Y., Gastesi, R., López, J., 2010. Sediment production and water quality of watersheds with contrasting land use in Navarre (Spain). Agric. Water Manage. 97, 1683–1694. https://doi.org/10.1016/j.agwat.2010.05.024

Casalí, J., Lopez, J.J., Giraldez, J.V., 1999. Ephemeral gully erosion in southern Navarra (Spain). CATENA 36, 20. https://doi.org/10.1016/S0341-8162(99)00013-2

Chan, K., Schillereff, D.N., Baas, A.C., Chadwick, M.A., Main, B., Mulligan, M., O'Shea, F.T., Pearce, R., Smith, T.E., van Soesbergen, A., Tebbs, E., Thompson, J., 2021. Low-cost electronic sensors for environmental research: Pitfalls and opportunities. Progress in Physical Geography: Earth and Environment 45, 305–338. https://doi.org/10.1177/0309133320956567

Chang, T.K., Talei, A., Alaghmand, S., Ooi, M.P.-L., 2017. Choice of rainfall inputs for event-based rainfall-runoff modeling in a catchment with multiple rainfall stations using data-driven techniques. Journal of Hydrology 545, 100–108. https://doi.org/10.1016/j.jhydrol.2016.12.024

Chen, X.Y., Chau, K.W., 2016. A Hybrid Double Feedforward Neural Network for Suspended Sediment Load Estimation. Water Resour. Manag. 30, 2179–2194. https://doi.org/10.1007/s11269-016-1281-2

Choukri, F., Raclot, D., Naimi, M., Chikhaoui, M., Nunes, J.P., Huard, F., Hérivaux, C., Sabir, M., Pépin, Y., 2020. Distinct and combined impacts of climate and land use scenarios on water availability and sediment loads for a water supply reservoir in northern Morocco. Int. Soil Water Conserv. Res. 8, 141–153. https://doi.org/10.1016/j.iswcr.2020.03.003

Cloern, J.E., Abreu, P.C., Carstensen, J., Chauvaud, L., Elmgren, R., Grall, J., Greening, H., Johansson, J.O.R., Kahru, M., Sherwood, E.T., Xu, J., Yin, K., 2016. Human activities and climate variability drive fast-paced change across the world's estuarine–coastal ecosystems. Global Change Biology 22, 513–529. https://doi.org/10.1111/gcb.13059

Cobaner, M., Unal, B., Kisi, O., 2009. Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydro-meteorological data. J. Hydrol. 367, 52–61. https://doi.org/10.1016/j.jhydrol.2008.12.024

Conesa, H.M., Jiménez-Cárceles, F.J., 2007. The Mar Menor lagoon (SE Spain): A singular natural ecosystem threatened by human activities. Marine Pollution Bulletin 54, 839–849. https://doi.org/10.1016/j.marpolbul.2007.05.007

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Cortés, A., Forrest, A.L., Sadro, S., Stang, A.J., Swann, M., Framsted, N.T., Thirkill, R., Sharp, S.L., Schladow, S.G., 2021. Prediction of Hypoxia in Eutrophic Polymictic Lakes. Water Resources Research 57, e2020WR028693. https://doi.org/10.1029/2020WR028693

De Santisteban, L.M., Casalí, J., López, J.J., 2006. Assessing soil erosion rates in cultivated areas of Navarre (Spain). Earth Surf. Process. Landforms 31, 487–506. https://doi.org/10.1002/esp.1281

Domingo-Pinillos, J.C., Senent-Aparicio, J., García-Aróstegui, J.L., Baudron, P., 2018. Long Term Hydrodynamic Effects in a Semi-Arid Mediterranean Multilayer Aquifer: Campo de Cartagena in South-Eastern Spain. Water 10, 1320. https://doi.org/10.3390/w10101320

Duru, U., Arabi, M., Wohl, E.E., 2018. Modeling stream flow and sediment yield using the SWAT model: a case study of Ankara River basin, Turkey. Phys. Geogr. 39, 264–289. https://doi.org/10.1080/02723646.2017.1342199

Dutta, S., Sen, D., 2018. Application of SWAT model for predicting soil erosion and sediment yield. Sustain. Water Resour. Manag. 4, 447–468. https://doi.org/10.1007/s40899-017-0127-2

Elliott, J.A., 2012. Is the future blue-green? A review of the current model predictions of how climate change could affect pelagic freshwater cyanobacteria. Water Research, Cyanobacteria: Impacts of climate change on occurrence, toxicity and water quality management 46, 1364–1371. https://doi.org/10.1016/j.watres.2011.12.018

Epelde, A.M., Cerro, I., Sánchez-Pérez, J.M., Sauvage, S., Srinivasan, R., Antigüedad, I., 2015. Application of the SWAT model to assess the impact of changes in agricultural management practices on water quality. Hydrol. Sci. J. 1–19. https://doi.org/10.1080/02626667.2014.967692

Erena, M., Domínguez, J.A., Aguado-Giménez, F., Soria, J., García-Galiano, S., 2019. Monitoring Coastal Lagoon Water Quality through Remote Sensing: The Mar Menor as a Case Study. Water 11, 1468. https://doi.org/10.3390/w11071468

Farrell, K.J., Ward, N.K., Krinos, A.I., Hanson, P.C., Daneshmand, V., Figueiredo, R.J., Carey, C.C., 2020. Ecosystem-scale nutrient cycling responses to increasing air temperatures vary with lake trophic state. Ecological Modelling 430, 109134. https://doi.org/10.1016/j.ecolmodel.2020.109134

Foley, B., Jones, I.D., Maberly, S.C., Rippey, B., 2012. Long-term changes in oxygen depletion in a small temperate lake: effects of climate change and eutrophication. Freshwater Biology 57, 278–289. https://doi.org/10.1111/j.1365-2427.2011.02662.x

Fornarelli, R., Galelli, S., Castelletti, A., Antenucci, J.P., Marti, C.L., 2013. An empirical modeling approach to predict and understand phytoplankton dynamics in a reservoir affected by inter-basin water transfers. Water Resources Research 49, 3626–3641. https://doi.org/10.1002/wrcr.20268

Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics 29, 1189–1232.

Fu, B., Merritt, W.S., Croke, B.F.W., Weber, T.R., Jakeman, A.J., 2019. A review of catchment-scale water quality and erosion models and a synthesis of future prospects. Environ. Modell. Softw. 114, 75–97. https://doi.org/10.1016/j.envsoft.2018.12.008

García-Ayllón, S., 2019. New Strategies to Improve Co-Management in Enclosed Coastal Seas and Wetlands Subjected to Complex Environments: Socio-Economic Analysis Applied to an International Recovery Success Case Study after an Environmental Crisis. Sustainability 11, 1039. https://doi.org/10.3390/su11041039

Garcia-Ayllon, S., 2018. The Integrated Territorial Investment (ITI) of the Mar Menor as a model for the future in the comprehensive management of enclosed coastal seas. Ocean & Coastal Management, Maritime Spatial Planning, Ecosystem Approach and Supporting Information Systems (MapSIS 2017) 166, 82–97. https://doi.org/10.1016/j.ocecoaman.2018.05.004

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

García-Pintado, J., Barberá, G.G., Erena, M., Castillo, V.M., 2009. Calibration of structure in a distributed forecasting model for a semiarid flash flood: Dynamic surface storage and channel roughness. Journal of Hydrology 377, 165–184. https://doi.org/10.1016/j.jhydrol.2009.08.024

Gassman, P.W., Sadeghi, A.M., Srinivasan, R., 2014. Applications of the SWAT Model Special Section: Overview and Insights. J. Environ. Qual. 43, 1–8. https://doi.org/10.2134/jeq2013.11.0466

Ghasempour, R., Roushangar, K., Sihag, P., 2021. Suspended sediment load prediction in consecutive stations of river based on ensemble pre-post-processing kernel based approaches. Water Supply 17.

Goedkoop, W., Naddafi, R., Grandin, U., 2011. Retention of N and P by zebra mussels (Dreissena polymorpha Pallas) and its quantitative role in the nutrient budget of eutrophic Lake Ekoln, Sweden. Biol Invasions 13, 1077–1086. https://doi.org/10.1007/s10530-011-9950-9

Gómez, D., Salvador, P., Sanz, J., Casanova, J.L., 2021. A new approach to monitor water quality in the Menor sea (Spain) using satellite data and machine learning methods. Environmental Pollution 286, 117489. https://doi.org/10.1016/j.envpol.2021.117489

González-Enrique, J., Ruiz-Aguilar, J.J., Madrid Navarro, E., Martínez Álvarez-Castellanos, R., Felis Enguix, I., Jerez, J.M., Turias, I.J., 2023. Deep Learning Approach for the Prediction of the Concentration of Chlorophyll α in Seawater. A Case Study in El Mar Menor (Spain), in: García Bringas, P., Pérez García, H., Martinez-de-Pison, F.J., Villar Flecha, J.R., Troncoso Lora, A., de la Cal, E.A., Herrero, Á., Martínez Álvarez, F., Psaila, G., Quintián, H., Corchado Rodriguez, E.S. (Eds.), 17th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2022), Lecture Notes in Networks and Systems. Springer Nature Switzerland, Cham, pp. 72–85. https://doi.org/10.1007/978-3-031-18050-7_8

Grainger, A., Kim, J., 2020. Reducing Global Environmental Uncertainties in Reports of Tropical Forest Carbon Fluxes to REDD+ and the Paris Agreement Global Stocktake. Remote Sensing 12, 2369. https://doi.org/10.3390/rs12152369

Gupta, D., Hazarika, B.B., Berlin, M., Sharma, U.M., Mishra, K., 2021. Artificial intelligence for suspended sediment load prediction: a review. Environ. Earth Sci. 80, 346. https://doi.org/10.1007/s12665-021-09625-3

Gurkan, Z., Zhang, J., Jørgensen, S.E., 2006. Development of a structurally dynamic model for forecasting the effects of restoration of Lake Fure, Denmark. Ecological Modelling 197, 89–102. https://doi.org/10.1016/j.ecolmodel.2006.03.006

Hackeling, G., 2014. Mastering machine learning with scikit-learn: apply effective learning algorithms to real-world problems using scikit-learn, Packt open source. Packt Publ, Birmingham.

Hamaamin, Y., Nejadhashemi, A., Zhang, Z., Giri, S., Woznicki, S., 2016. Bayesian Regression and Neuro-Fuzzy Methods Reliability Assessment for Estimating Streamflow. Water 8, 287. https://doi.org/10.3390/w8070287

Hanson, P.C., Stillman, A.B., Jia, X., Karpatne, A., Dugan, H.A., Carey, C.C., Stachelek, J., Ward, N.K., Zhang, Y., Read, J.S., Kumar, V., 2020. Predicting lake surface water phosphorus dynamics using process-guided machine learning. Ecological Modelling 430, 109136. https://doi.org/10.1016/j.ecolmodel.2020.109136

Harbola, S., Coors, V., 2019. One dimensional convolutional neural network architectures for wind prediction. Energy Conversion and Management 195, 70–75. https://doi.org/10.1016/j.enconman.2019.05.007

Harris, T.D., Graham, J.L., 2017. Predicting cyanobacterial abundance, microcystin, and geosmin in a eutrophic drinking-water reservoir using a 14-year dataset. Lake and Reservoir Management 33, 32–48. https://doi.org/10.1080/10402381.2016.1263694

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Heddam, S., Kisi, O., 2018. Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. Journal of Hydrology 559, 499–509. https://doi.org/10.1016/j.jhydrol.2018.02.061

Hense, I., Beckmann, A., 2006. Towards a model of cyanobacteria life cycle—effects of growing and resting stages on bloom formation of N2-fixing species. Ecological Modelling 195, 205–218. https://doi.org/10.1016/j.ecolmodel.2005.11.018

Hipsey, M.R., Bruce, L.C., Boon, C., Busch, B., Carey, C.C., Hamilton, D.P., Hanson, P.C., Read, J.S., de Sousa, E., Weber, M., Winslow, L.A., 2019. A General Lake Model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake Ecological Observatory Network (GLEON). Geoscientific Model Development 12, 473–523. https://doi.org/10.5194/gmd-12-473-2019

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Computation 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, Y., Chen, Z., Yu, T., Huang, X., Gu, X., 2018. Agricultural remote sensing big data: Management and applications. Journal of Integrative Agriculture 17, 1915–1931. https://doi.org/10.1016/S2095-3119(17)61859-8

Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B.W., Verspagen, J.M.H., Visser, P.M., 2018. Cyanobacterial blooms. Nat Rev Microbiol 16, 471–483. https://doi.org/10.1038/s41579-018-0040-1

Hussain, D., Khan, A.A., 2020. Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. Earth Sci Inform 13, 939–949. https://doi.org/10.1007/s12145-020-00450-z

Jabbari, A., Ackerman, J.D., Boegman, L., Zhao, Y., 2019. Episodic hypoxia in the western basin of Lake Erie. Limnology and Oceanography 64, 2220–2236. https://doi.org/10.1002/lno.11180

Jaeger, K.L., Sutfin, N.A., Tooth, S., Michaelides, K., Singer, M., 2017. Chapter 2.1 - Geomorphology and Sediment Regimes of Intermittent Rivers and Ephemeral Streams, in: Datry, T., Bonada, N., Boulton, A. (Eds.), Intermittent Rivers and Ephemeral Streams. Academic Press, pp. 21–49. https://doi.org/10.1016/B978-0-12-803835-2.00002-4

Jane, S.F., Hansen, G.J.A., Kraemer, B.M., Leavitt, P.R., Mincer, J.L., North, R.L., Pilla, R.M., Stetler, J.T., Williamson, C.E., Woolway, R.I., Arvola, L., Chandra, S., DeGasperi, C.L., Diemer, L., Dunalska, J., Erina, O., Flaim, G., Grossart, H.-P., Hambright, K.D., Hein, C., Hejzlar, J., Janus, L.L., Jenny, J.-P., Jones, J.R., Knoll, L.B., Leoni, B., Mackay, E., Matsuzaki, S.-I.S., McBride, C., Müller-Navarra, D.C., Paterson, A.M., Pierson, D., Rogora, M., Rusak, J.A., Sadro, S., Saulnier-Talbot, E., Schmid, M., Sommaruga, R., Thiery, W., Verburg, P., Weathers, K.C., Weyhenmeyer, G.A., Yokota, K., Rose, K.C., 2021. Widespread deoxygenation of temperate lakes. Nature 594, 66–70. https://doi.org/10.1038/s41586-021-03550-y

Ji, H., Chen, Y., Fang, G., Li, Z., Duan, W., Zhang, Q., 2021. Adaptability of machine learning methods and hydrological models to discharge simulations in data-sparse glaciated watersheds. J. Arid Land. https://doi.org/10.1007/s40333-021-0066-5

Jimeno-Sáez, P., Martínez-España, R., Casalí, J., Pérez-Sánchez, J., Senent-Aparicio, J., 2022. A comparison of performance of SWAT and machine learning models for predicting sediment load in a forested Basin, Northern Spain. CATENA 212, 105953. https://doi.org/10.1016/j.catena.2021.105953

Jimeno-Sáez, P., Senent-Aparicio, J., Cecilia, J.M., Pérez-Sánchez, J., 2020. Using Machine-Learning Algorithms for Eutrophication Modeling: Case Study of Mar Menor Lagoon (Spain). IJERPH 17, 1189. https://doi.org/10.3390/ijerph17041189

Jimeno-Sáez, P., Senent-Aparicio, J., Pérez-Sánchez, J., Pulido-Velazquez, D., 2018. A Comparison of SWAT and ANN Models for Daily Runoff Simulation in Different Climatic Zones of Peninsular Spain. Water 10, 192. https://doi.org/10.3390/w10020192

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Jöhnk, K.D., Brüggemann, R., Rücker, J., Luther, B., Simon, U., Nixdorf, B., Wiedner, C., 2011. Modelling life cycle and population dynamics of Nostocales (cyanobacteria). Environmental Modelling & Software 26, 669–677. https://doi.org/10.1016/j.envsoft.2010.11.001

Justice, C., Román, M., Csiszar, I., Vermote, E., Wolfe, R., Hook, S., Friedl, M., Wang, Z., Schaaf, C., Miura, T., Tschudi, M., Riggs, G., Hall, D., Lyapustin, A., Devadiga, S., Davidson, C., Ed, M., 2013. Land and cryosphere products from Suomi NPP VIIRS: Overview and status. Journal of Geophysical Research: Atmospheres 118. https://doi.org/10.1002/jgrd.50771

Justice, C.O., Townshend, J.R.G., Vermote, E.F., Masuoka, E., Wolfe, R.E., Saleous, N., Roy, D.P., Morisette, J.T., 2002. An overview of MODIS Land data processing and product status. Remote Sensing of Environment, The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring 83, 3–15. https://doi.org/10.1016/S0034-4257(02)00084-6

Kakouei, K., Kraemer, B.M., Adrian, R., 2022. Variation in the predictability of lake plankton metric types. Limnology and Oceanography 67, 608–620. https://doi.org/10.1002/lno.12021

Kalin, L., Isik, S., Schoonover, J.E., Lockaby, B.G., 2010. Predicting Water Quality in Unmonitored Watersheds Using Artificial Neural Networks. J. Environ. Qual. 39, 12.

Karlsson-Elfgren, I., Hyenstrand, P., Riydin, E., 2005. Pelagic growth and colony division of Gloeotrichia echinulata in Lake Erken. Journal of Plankton Research 27, 145–151. https://doi.org/10.1093/plankt/fbh165

Karlsson-Elfgren, I., Rengefors, K., Gustafsson, S., 2004. Factors regulating recruitment from the sediment to the water column in the bloom-forming cyanobacterium Gloeotrichia echinulata. Freshwater Biology 49, 265–273. https://doi.org/10.1111/j.1365-2427.2004.01182.x

Khosravi, K., Cooper, J.R., Daggupati, P., Thai Pham, B., Tien Bui, D., 2020. Bedload transport rate prediction: Application of novel hybrid data mining techniques. J. Hydrol. 585, 124774. https://doi.org/10.1016/j.jhydrol.2020.124774

Kim, R.J., Loucks, D.P., Stedinger, J.R., 2012. Artificial Neural Network Models of Watershed Nutrient Loading. Water Resour. Manage. 26, 2781–2797. https://doi.org/10.1007/s11269-012-0045-x

Kisi, O., 2005. Suspended sediment estimation using neuro-fuzzy and neural network approaches/Estimation des matière. Hydrol. Sci. J. 50, 683–696. https://doi.org/10.1623/hysj.2005.50.4.683

Kisi, O., Alizamir, M., Docheshmeh Gorgij, A., 2020. Dissolved oxygen prediction using a new ensemble method. Environ Sci Pollut Res 27, 9589–9603. https://doi.org/10.1007/s11356-019-07574-w

Kumar, D., Pandey, A., Sharma, N., Flügel, W.-A., 2016. Daily suspended sediment simulation using machine learning approach. CATENA 138, 77–90. https://doi.org/10.1016/j.catena.2015.11.013

Ladwig, R., Hanson, P.C., Dugan, H.A., Carey, C.C., Zhang, Y., Shu, L., Duffy, C.J., Cobourn, K.M., 2021. Lake thermal structure drives interannual variability in summer anoxia dynamics in a eutrophic lake over 37 years. Hydrology and Earth System Sciences 25, 1009–1032. https://doi.org/10.5194/hess-25-1009-2021

Lechner, A.M., Foody, G.M., Boyd, D.S., 2020. Applications in Remote Sensing to Forest Ecology and Management. One Earth 2, 405–412. https://doi.org/10.1016/j.oneear.2020.05.001

Leon, L.F., Smith, R.E.H., Hipsey, M.R., Bocaniov, S.A., Higgins, S.N., Hecky, R.E., Antenucci, J.P., Imberger, J.A., Guildford, S.J., 2011. Application of a 3D hydrodynamic–biological model for seasonal and spatial dynamics of water quality and phytoplankton in Lake Erie. Journal of Great Lakes Research 37, 41–53. https://doi.org/10.1016/j.jglr.2010.12.007

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Lin, S., Boegman, L., Rao, Y.R., 2021. Characterizing spatial and temporal distributions of turbulent mixing and dissipation in Lake Erie. Journal of Great Lakes Research 47, 168–179. https://doi.org/10.1016/j.jglr.2020.11.014

Lin, S., Pierson, D.C., Mesman, J.P., 2023. Prediction of algal blooms via data-driven machine learning models: an evaluation using data from a well-monitored mesotrophic lake. Geoscientific Model Development 16, 35–46. https://doi.org/10.5194/gmd-16-35-2023

Liu, R., Zhang, P., Wang, X., Chen, Y., Shen, Z., 2013. Assessment of effects of best management practices on agricultural non-point source pollution in Xiangxi River watershed. Agricultural Water Management 117, 9–18. https://doi.org/10.1016/j.agwat.2012.10.018

López-Ballesteros, A., Senent-Aparicio, J., Srinivasan, R., Pérez-Sánchez, J., 2019. Assessing the Impact of Best Management Practices in a Highly Anthropogenic and Ungauged Watershed Using the SWAT Model: A Case Study in the El Beal Watershed (Southeast Spain). Agronomy 9, 576. https://doi.org/10.3390/agronomy9100576

López-Ballesteros, A., Trolle, D., Srinivasan, R., Senent-Aparicio, J., 2023. Assessing the effectiveness of potential best management practices for science-informed decision support at the watershed scale: The case of the Mar Menor coastal lagoon, Spain. Science of The Total Environment 859, 160144. https://doi.org/10.1016/j.scitotenv.2022.160144

Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M.E., García-Santos, G., Fernandes, R., Berger, M., 2012. Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. Remote Sensing of Environment, The Sentinel Missions - New Opportunities for Science 120, 91–101. https://doi.org/10.1016/j.rse.2011.09.026

Malinowski, R., Lewiński, S., Rybicki, M., Gromny, E., Jenerowicz, M., Krupiński, Michał, Nowakowski, A., Wojtkowski, C., Krupiński, Marcin, Krätzschmar, E., Schauer, P., 2020. Automated Production of a Land Cover/Use Map of Europe Based on Sentinel-2 Imagery. Remote Sensing 12, 3523. https://doi.org/10.3390/rs12213523

Marcé, R., George, G., Buscarinu, P., Deidda, M., Dunalska, J., de Eyto, E., Flaim, G., Grossart, H.-P., Istvanovics, V., Lenhardt, M., Moreno-Ostos, E., Obrador, B., Ostrovsky, I., Pierson, D.C., Potužák, J., Poikane, S., Rinke, K., Rodríguez-Mozaz, S., Staehr, P.A., Šumberová, K., Waajen, G., Weyhenmeyer, G.A., Weathers, K.C., Zion, M., Ibelings, B.W., Jennings, E., 2016. Automatic High Frequency Monitoring for Improved Lake and Reservoir Management. Environ. Sci. Technol. 50, 10780–10794. https://doi.org/10.1021/acs.est.6b01604

Matsuzaki, S.-I.S., Lathrop, R.C., Carpenter, S.R., Walsh, J.R., Vander Zanden, M.J., Gahler, M.R., Stanley, E.H., 2021. Climate and food web effects on the spring clear-water phase in two north-temperate eutrophic lakes. Limnology and Oceanography 66, 30–46. https://doi.org/10.1002/lno.11584

McHugh, M.L., 2012. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 22, 276–282.

Meaurio, M., Zabaleta, A., Uriarte, J.A., Srinivasan, R., Antigüedad, I., 2015. Evaluation of SWAT models performance to simulate streamflow spatial origin. The case of a small forested watershed. J. Hydrol. 525, 326–334. https://doi.org/10.1016/j.jhydrol.2015.03.050

Mellios, N., Moe, S., Laspidou, C., 2020. Machine Learning Approaches for Predicting Health Risk of Cyanobacterial Blooms in Northern European Lakes. Water 12, 1191. https://doi.org/10.3390/w12041191

Mercado, J.M., Cortés, D., Gómez-Jakobsen, F., García-Gómez, C., Ouaissa, S., Yebra, L., Ferrera, I., Valcárcel-Pérez, N., López, M., García-Muñoz, R., Ramos, A., Bernardeau, J., Belando, M.D., Fraile-Nuez, E., Ruíz, J.M., 2021. Role of small-sized phytoplankton in triggering an ecosystem disruptive algal bloom in a Mediterranean hypersaline coastal lagoon. Marine Pollution Bulletin 164, 111989. https://doi.org/10.1016/j.marpolbul.2021.111989

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Merritt, W.S., Letcher, R.A., Jakeman, A.J., 2003. A review of erosion and sediment transport models. Environ. Modell. Softw. 18, 761–799. https://doi.org/10.1016/S1364-8152(03)00078-1

Mesman, J.P., Ayala, A.I., Goyette, S., Kasparian, J., Marcé, R., Markensten, H., Stelzer, J.A.A., Thayne, M.W., Thomas, M.K., Pierson, D.C., Ibelings, B.W., 2022. Drivers of phytoplankton responses to summer wind events in a stratified lake: A modeling study. Limnology and Oceanography 67, 856–873. https://doi.org/10.1002/lno.12040

Minns, A.W., Hall, M.J., 1996. Artificial neural networks as rainfall-runoff models. Hydrol. Sci. J. 41, 399–417. https://doi.org/10.1080/02626669609491511

Moras, S., Ayala, A.I., Pierson, D.C., 2019. Historical modelling of changes in Lake Erken thermal conditions. Hydrology and Earth System Sciences 23, 5001–5016. https://doi.org/10.5194/hess-23-5001-2019

Moriasi, D.N., Arnold, J.G., Liew, M.W.V., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Trans. ASABE 50, 885–900. https://doi.org/10.13031/2013.23153

Müller, B., Bryant, L.D., Matzinger, A., Wüest, A., 2012. Hypolimnetic Oxygen Depletion in Eutrophic Lakes. Environ. Sci. Technol. 46, 9964–9971. https://doi.org/10.1021/es301422r

Mulligan, M., Douglas, C., van Soesbergen, A., Shi, M., Burke, S., van Delden, H., Giordano, R., Lopez-Gunn, E., Scrieciu, A., 2021. Environmental Intelligence for more Sustainable Infrastructure Investment, in: Proceedings of the Conference on Information Technology for Social Good, GoodIT '21. Association for Computing Machinery, New York, NY, USA, pp. 225–229. https://doi.org/10.1145/3462203.3475916

Nachtergaele, F., van Velthuizen, H., Batjes, N., Dijkshoorn, K., van, V., Fischer, G., Jones, A., Montanarella, L., Petri, M., Prieler, S., Teixeira, E., Wiberg, D., 2010a. The harmonized world soil database.

Nachtergaele, F., van Velthuizen, H., Batjes, N., Dijkshoorn, K., van, V., Fischer, G., Jones, A., Montanarella, L., Petri, M., Prieler, S., Teixeira, E., Wiberg, D., 2010b. The harmonized world soil database 4.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2009. Soil and Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resources Institute Technical Report No. 406.

Nelson, N.G., Muñoz-Carpena, R., Phlips, E.J., Kaplan, D., Sucsy, P., Hendrickson, J., 2018. Revealing Biotic and Abiotic Controls of Harmful Algal Blooms in a Shallow Subtropical Lake through Statistical Machine Learning. Environ. Sci. Technol. 52, 3527–3535. https://doi.org/10.1021/acs.est.7b05884

Nguyen, P., Shearer, E.J., Tran, H., Ombadi, M., Hayatbini, N., Palacios, T., Huynh, P., Braithwaite, D., Updegraff, G., Hsu, K., Kuligowski, B., Logan, W.S., Sorooshian, S., 2019. The CHRS Data Portal, an easily accessible public repository for PERSIANN global satellite precipitation data. Sci Data 6, 180296. https://doi.org/10.1038/sdata.2018.296

Nguyen, P.K.-T., Chua, L.H.-C., Son, L.H., 2014. Flood forecasting in large rivers with data-driven models. Nat Hazards 71, 767–784. https://doi.org/10.1007/s11069-013-0920-7

Noe, S.M., Tabakova, K., Mahura, A., Lappalainen, H.K., Kosmale, M., Heilimo, J., Salzano, R., Santoro, M., Salvatori, R., Spolaor, A., Cairns, W., Barbante, C., Pankratov, F., Humbert, A., Sonke, J.E., Law, K.S., Onishi, T., Paris, J.-D., Skov, H., Massling, A., Dommergue, A., Arshinov, M., Davydov, D., Belan, B., Petäjä, T., 2022. Arctic observations and sustainable development goals – Contributions and examples from ERA-PLANET iCUPE data. Environmental Science & Policy 132, 323–336. https://doi.org/10.1016/j.envsci.2022.02.034

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Nor, N.I., Harun, S., Kassim, A.H., 2007. Radial Basis Function Modeling of Hourly Streamflow Hydrograph. Journal of Hydrologic Engineering 12, 113–123. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:1(113)

North, R.P., North, R.L., Livingstone, D.M., Köster, O., Kipfer, R., 2014. Long-term changes in hypoxia and soluble reactive phosphorus in the hypolimnion of a large temperate lake: consequences of a climate regime shift. Global Change Biology 20, 811–823. https://doi.org/10.1111/gcb.12371

Nunes, J.P., Naranjo Quintanilla, P., Santos, J.M., Serpa, D., Carvalho-Santos, C., Rocha, J., Keizer, J.J., Keesstra, S.D., 2018. Afforestation, Subsequent Forest Fires and Provision of Hydrological Services: A Model-Based Analysis for a Mediterranean Mountainous Catchment: Mediterranean Afforestation, Forest Fires and Hydrological Services. Land Degrad. Develop. 29, 776–788. https://doi.org/10.1002/ldr.2776

Nürnberg, G.K., LaZerte, B.D., Loh, P.S., Molot, L.A., 2013. Quantification of internal phosphorus load in large, partially polymictic and mesotrophic Lake Simcoe, Ontario. Journal of Great Lakes Research 39, 271–279. https://doi.org/10.1016/j.jglr.2013.03.017

Olyaie, E., Banejad, H., Chau, K.-W., Melesse, A.M., 2015. A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: a case study in United States. Environ. Monit. Assess. 187, 189. https://doi.org/10.1007/s10661-015-4381-1

Orihel, D.M., Baulch, H.M., Casson, N.J., North, R.L., Parsons, C.T., Seckar, D.C.M., Venkiteswaran, J.J., 2017. Internal phosphorus loading in Canadian fresh waters: a critical review and data analysis. Can. J. Fish. Aquat. Sci. 74, 2005–2029. https://doi.org/10.1139/cjfas-2016-0500

Paerl, H.W., 1988. Nuisance phytoplankton blooms in coastal, estuarine, and inland waters1. Limnology and Oceanography 33, 823–843. https://doi.org/10.4319/lo.1988.33.4part2.0823

Paerl, H.W., Huisman, J., 2008. Blooms Like It Hot. Science 320, 57–58. https://doi.org/10.1126/science.1155398

Paerl, H.W., Paul, V.J., 2012. Climate change: Links to global expansion of harmful cyanobacteria. Water Research, Cyanobacteria: Impacts of climate change on occurrence, toxicity and water quality management 46, 1349–1363. https://doi.org/10.1016/j.watres.2011.08.002

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M.K., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. Remote Sensing of Environment 240, 111604. https://doi.org/10.1016/j.rse.2019.111604

Patterson, J.C., Hamblin, P.F., Imberger, J., 1984. Classification and dynamic simulation of the vertical density structure of lakes1. Limnology and Oceanography 29, 845–861. https://doi.org/10.4319/lo.1984.29.4.0845

Peña-Haro, S., Carrel, M., Lüthi, B., Hansen, I., Lukes, R., 2021. Robust Image-Based Streamflow Measurements for Real-Time Continuous Monitoring. Frontiers in Water 3.

Persson, I., Jones, I.D., 2008. The effect of water colour on lake hydrodynamics: a modelling study. Freshwater Biology 53, 2345–2355. https://doi.org/10.1111/j.1365-2427.2008.02049.x

Pettersson, K., 1998. Mechanisms for internal loading of phosphorus in lakes. Hydrobiologia 373, 21–25. https://doi.org/10.1023/A:1017011420035

Pettersson, K., 1985. The Availability of Phosphorus and the Species Composition of the Spring Phytoplankton in Lake Erken. Internationale Revue der gesamten Hydrobiologie und Hydrographie 70, 527–546. https://doi.org/10.1002/iroh.19850700407

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Pettersson, K., Grust, K., Weyhenmeyer, G., Blenckner, T., 2003. Seasonality of chlorophyll and nutrients in Lake Erken – effects of weather conditions. Hydrobiologia 506, 75–81. https://doi.org/10.1023/B:HYDR.0000008582.61851.76

Pierson, D.C., Pettersson, K., Istvanovics, V., 1992. Temporal changes in biomass specific photosynthesis during the summer: regulation by environmental factors and the importance of phytoplankton succession. Hydrobiologia 243, 119–135. https://doi.org/10.1007/BF00007027

Pulighe, G., Bonati, G., Colangeli, M., Traverso, L., Lupia, F., Altobelli, F., Dalla Marta, A., Napoli, M., 2019. Predicting Streamflow and Nutrient Loadings in a Semi-Arid Mediterranean Watershed with Ephemeral Streams Using the SWAT Model. Agronomy 10, 2. https://doi.org/10.3390/agronomy10010002

Quinlan, J.R., 1992. Learning with Continuous Classes. Presented at the Proceedings of Australian Joint Conference on Artificial Intelligence, Hobart, pp. 343–348.

Rahmani, F., Lawson, K., Ouyang, W., Appling, A., Oliver, S., Shen, C., 2021. Exploring the exceptional performance of a deep learning stream temperature model and the value of streamflow data. Environ. Res. Lett. 16, 024025. https://doi.org/10.1088/1748-9326/abd501

Rao, Y.R., Hawley, N., Charlton, M.N., Schertzer, W.M., 2008. Physical processes and hypoxia in the central basin of Lake Erie. Limnology and Oceanography 53, 2007–2020. https://doi.org/10.4319/lo.2008.53.5.2007

Rao, Y.R., Howell, T., Watson, S.B., Abernethy, S., 2014. On hypoxia and fish kills along the north shore of Lake Erie. Journal of Great Lakes Research 40, 187–191. https://doi.org/10.1016/j.jglr.2013.11.007

Read, J.S., Hamilton, D.P., Jones, I.D., Muraoka, K., Winslow, L.A., Kroiss, R., Wu, C.H., Gaiser, E., 2011. Derivation of lake mixing and stratification indices from high-resolution lake buoy data. Environmental Modelling & Software 26, 1325–1336. https://doi.org/10.1016/j.envsoft.2011.05.006

Read, J.S., Jia, X., Willard, J., Appling, A.P., Zwart, J.A., Oliver, S.K., Karpatne, A., Hansen, G.J.A., Hanson, P.C., Watkins, W., Steinbach, M., Kumar, V., 2019. Process-Guided Deep Learning Predictions of Lake Water Temperature. Water Resources Research 55, 9173–9190. https://doi.org/10.1029/2019WR024922

Reichwaldt, E.S., Ghadouani, A., 2012. Effects of rainfall patterns on toxic cyanobacterial blooms in a changing climate: Between simplistic scenarios and complex dynamics. Water Research, Cyanobacteria: Impacts of climate change on occurrence, toxicity and water quality management 46, 1372–1393. https://doi.org/10.1016/j.watres.2011.11.052

Richardson, J., Miller, C., Maberly, S.C., Taylor, P., Globevnik, L., Hunter, P., Jeppesen, E., Mischke, U., Moe, S.J., Pasztaleniec, A., Søndergaard, M., Carvalho, L., 2018. Effects of multiple stressors on cyanobacteria abundance vary with lake type. Global Change Biology 24, 5044–5055. https://doi.org/10.1111/gcb.14396

Rousso, B.Z., Bertone, E., Stewart, R., Hamilton, D.P., 2020. A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes. Water Research 182, 115959. https://doi.org/10.1016/j.watres.2020.115959

Scavia, D., David Allan, J., Arend, K.K., Bartell, S., Beletsky, D., Bosch, N.S., Brandt, S.B., Briland, R.D., Daloğlu, I., DePinto, J.V., Dolan, D.M., Evans, M.A., Farmer, T.M., Goto, D., Han, H., Höök, T.O., Knight, R., Ludsin, S.A., Mason, D., Michalak, A.M., Peter Richards, R., Roberts, J.J., Rucinski, D.K., Rutherford, E., Schwab, D.J., Sesterhenn, T.M., Zhang, H., Zhou, Y., 2014. Assessing and addressing the re-eutrophication of Lake Erie: Central basin hypoxia. Journal of Great Lakes Research 40, 226–246. https://doi.org/10.1016/j.jglr.2014.02.004

Schonlau, M., Zou, R.Y., 2020. The random forest algorithm for statistical learning. Stata J. 20, 3–29. https://doi.org/10.1177/1536867X20909688

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Senent-Aparicio, J., Jimeno-Sáez, P., Bueno-Crespo, A., Pérez-Sánchez, J., Pulido-Velázquez, D., 2019. Coupling machine-learning techniques with SWAT model for instantaneous peak flow prediction. Biosyst. Eng. 177, 67–77. https://doi.org/10.1016/j.biosystemseng.2018.04.022

Senent-Aparicio, J., López-Ballesteros, A., Nielsen, A., Trolle, D., 2021. A holistic approach for determining the hydrology of the mar menor coastal lagoon by combining hydrological & hydrodynamic models. Journal of Hydrology 603, 127150. https://doi.org/10.1016/j.jhydrol.2021.127150

Senent-Aparicio, J., Pérez-Sánchez, J., García-Aróstegui, J.L., Bielsa-Artero, A., Domingo-Pinillos, J.C., 2015. Evaluating Groundwater Management Sustainability under Limited Data Availability in Semiarid Zones. Water 7, 4305–4322. https://doi.org/10.3390/w7084305

Shatwell, T., Köhler, J., 2019. Decreased nitrogen loading controls summer cyanobacterial blooms without promoting nitrogen-fixing taxa: Long-term response of a shallow lake. Limnology and Oceanography 64, S166–S178. https://doi.org/10.1002/lno.11002

Sihag, P., Sadikhani, M.R., Vambol, V., Vambol, S., Prabhakar, A.K., Sharma, N., 2021. Comparative study for deriving stage- discharge–sediment concentration relationships using soft computing techniques. Journal of Achievements in Materials and Manufacturing Engineering 104, 20.

Singh, A., Imtiyaz, M., Isaac, R.K., Denis, D.M., 2012. Comparison of soil and water assessment tool (SWAT) and multilayer perceptron (MLP) artificial neural network for predicting sediment yield in the Nagwa agricultural watershed in Jharkhand, India. Agric. Water Manage. 104, 113–120. https://doi.org/10.1016/j.agwat.2011.12.005

Singh, A., Imtiyaz, Mohd., Isaac, R.K., Denis, D.M., 2014. Assessing the performance and uncertainty analysis of the SWAT and RBNN models for simulation of sediment yield in the Nagwa watershed, India. Hydrol. Sci. J. 59, 351–364. https://doi.org/10.1080/02626667.2013.872787

Sirabahenda, Z., St-Hilaire, A., Courtenay, S.C., van den Heuvel, M.R., 2020. Assessment of the effective width of riparian buffer strips to reduce suspended sediment in an agricultural landscape using ANFIS and SWAT models. CATENA 195, 104762. https://doi.org/10.1016/j.catena.2020.104762

Srivastava, P., McNair, J.N., Johnson, T.E., 2006. Comparison of process-based and artificial neural network approaches for streamflow modeling in an agricultural watershed. J. Am. Water Resour. Assoc. 42, 545–563. https://doi.org/10.1111/j.1752-1688.2006.tb04475.x

Stanley, F.K.T., Irvine, J.L., Jacques, W.R., Salgia, S.R., Innes, D.G., Winquist, B.D., Torr, D., Brenner, D.R., Goodarzi, A.A., 2019. Radon exposure is rising steadily within the modern North American residential environment, and is increasingly uniform across seasons. Sci Rep 9, 18472. https://doi.org/10.1038/s41598-019-54891-8

Taud, H., Mas, J.F., 2018. Multilayer Perceptron (MLP), in: Camacho Olmedo, M.T., Paegelow, M., Mas, J.-F., Escobar, F. (Eds.), Geomatic Approaches for Modeling Land Change Scenarios, Lecture Notes in Geoinformation and Cartography. Springer International Publishing, Cham, pp. 451–455. https://doi.org/10.1007/978-3-319-60801-3_27

Valipour, R., León, L.F., Howell, T., Dove, A., Rao, Y.R., 2021. Episodic nearshore-offshore exchanges of hypoxic waters along the north shore of Lake Erie. Journal of Great Lakes Research 47, 419–436. https://doi.org/10.1016/j.jglr.2021.01.014

Velasco, J., Lloret, J., Millan, A., Marin, A., Barahona, J., Abellan, P., Sanchez-Fernandez, D., 2006. Nutrient And Particulate Inputs Into The Mar Menor Lagoon (Se Spain) From An Intensive Agricultural Watershed. Water Air Soil Pollut 176, 37–56. https://doi.org/10.1007/s11270-006-2859-8

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Vilas, M.P., Marti, C.L., Adams, M.P., Oldham, C.E., Hipsey, M.R., 2017. Invasive Macrophytes Control the Spatial and Temporal Patterns of Temperature and Dissolved Oxygen in a Shallow Lake: A Proposed Feedback Mechanism of Macrophyte Loss. Frontiers in Plant Science 8.

Vos, K., Splinter, K.D., Harley, M.D., Simmons, J.A., Turner, I.L., 2019. CoastSat: A Google Earth Engine-enabled Python toolkit to extract shorelines from publicly available satellite imagery. Environmental Modelling & Software 122, 104528. https://doi.org/10.1016/j.envsoft.2019.104528

Walsh, J.R., Lathrop, R.C., Vander Zanden, M.J., 2017. Invasive invertebrate predator, Bythotrephes longimanus, reverses trophic cascade in a north-temperate lake. Limnology and Oceanography 62, 2498–2509. https://doi.org/10.1002/lno.10582

Walz, Y., Min, A., Dall, K., Duguru, M., Villagran de Leon, J.-C., Graw, V., Dubovyk, O., Sebesvari, Z., Jordaan, A., Post, J., 2020. Monitoring progress of the Sendai Framework using a geospatial model: The example of people affected by agricultural droughts in Eastern Cape, South Africa. Progress in Disaster Science 5, 100062. https://doi.org/10.1016/j.pdisas.2019.100062

Wang, Y., Witten, I.H., 1997. Induction of Model Trees for Predicting Continuous Classes. Presented at the 9th Eur Conf on Machine Learning, Prague (Czech Republic).

Watson, S.B., Miller, C., Arhonditsis, G., Boyer, G.L., Carmichael, W., Charlton, M.N., Confesor, R., Depew, D.C., Höök, T.O., Ludsin, S.A., Matisoff, G., McElmurry, S.P., Murray, M.W., Peter Richards, R., Rao, Y.R., Steffen, M.M., Wilhelm, S.W., 2016. The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia. Harmful Algae 56, 44–66. https://doi.org/10.1016/j.hal.2016.04.010

Wegner, B., Kronsbein, A.L., Gillefalk, M., van de Weyer, K., Köhler, J., Funke, E., Monaghan, M.T., Hilt, S., 2019. Mutual Facilitation Among Invading Nuttall's Waterweed and Quagga Mussels. Frontiers in Plant Science 10.

Williams, J.R., 1975. Sediment routing for agricultural watersheds. Water Resour. Bull. 11, 965–975.

Wilson, H.L., Ayala, A.I., Jones, I.D., Rolston, A., Pierson, D., de Eyto, E., Grossart, H.-P., Perga, M.-E., Woolway, R.I., Jennings, E., 2020. Variability in epilimnion depth estimations in lakes. Hydrology and Earth System Sciences 24, 5559–5577. https://doi.org/10.5194/hess-24-5559-2020

Wu, X., Xiao, Q., Wen, J., You, D., Hueni, A., 2019. Advances in quantitative remote sensing product validation: Overview and current status. Earth-Science Reviews 196, 102875. https://doi.org/10.1016/j.earscirev.2019.102875

Wüest, A., Lorke, A., 2009. Small-Scale Turbulence and Mixing: Energy Fluxes in Stratified Lakes, in: Likens, G.E. (Ed.), Encyclopedia of Inland Waters. Academic Press, Oxford, pp. 628–635. https://doi.org/10.1016/B978-012370626-3.00084-3

Xiao, X., He, J., Huang, H., Miller, T.R., Christakos, G., Reichwaldt, E.S., Ghadouani, A., Lin, S., Xu, X., Shi, J., 2017. A novel single-parameter approach for forecasting algal blooms. Water Research 108, 222–231. https://doi.org/10.1016/j.watres.2016.10.076

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., Dickinson, R., 2013. The role of satellite remote sensing in climate change studies. Nature Clim Change 3, 875–883. https://doi.org/10.1038/nclimate1908

Yang, Y., Stenger-Kovács, C., Padisák, J., Pettersson, K., 2016. Effects of winter severity on spring phytoplankton development in a temperate lake (Lake Erken, Sweden). Hydrobiologia 780, 47–57. https://doi.org/10.1007/s10750-016-2777-8

Young, C.-C., Liu, W.-C., Wu, M.-C., 2017. A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. Applied Soft Computing 53, 205–216. https://doi.org/10.1016/j.asoc.2016.12.052

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

Yousefi, A., Toffolon, M., 2022. Critical factors for the use of machine learning to predict lake surface water temperature. Journal of Hydrology 606, 127418. https://doi.org/10.1016/j.jhydrol.2021.127418

Zeiger, S.J., Hubbart, J.A., 2016. A SWAT model validation of nested-scale contemporaneous stream flow, suspended sediment and nutrients from a multiple-land-use watershed of the central USA. Sci. Total Environ. 572, 232–243. https://doi.org/10.1016/j.scitotenv.2016.07.178

Zounemat-Kermani, M., Mahdavi-Meymand, A., Alizamir, M., Adarsh, S., Yaseen, Z.M., 2020. On the complexities of sediment load modeling using integrative machine learning: Application of the great river of Loíza in Puerto Rico. J. Hydrol. 585, 124759. https://doi.org/10.1016/j.jhydrol.2020.124759

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.

SMARTLAGOON

End of Deliverable 3.3

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017861.