UNIL | Université de Lausanne

University of Lausanne
School of Geosciences and Environment
LAKES group

PREDICTING THE EPILIMETIC PHOSPHORUS DYNAMICS OF LAKE GENEVA USING A HYBRID MODEL APPROACH

MSc Thesis
Environmental Sciences

Gabriel Juri Alvarez

Supervisor:
Prof. Dr. Marie-Elodie Perga

June 14th, 2023

UNIL | Université de Lausanne
Faculté des géosciences et de l'environnement
MSc in Environmental Science
bâtiment Géopolis bureau 4613
1015 Lausanne

Formulaire dépôt du mémoire – Master en sciences de l'environnement MSc

Ce formulaire est à déposer (en format papier ou numérique) au secrétariat du master deux semaines avant la soutenance orale

Données personnelles de l'étudiant·e

Titre : ○ Madame ◉ Monsieur

Nom : Juri Alvarez Prénom : Gabriel

Adresse postale complète (en Suisse) : Les rives de la morges 3 1110

N° d'immatriculation : 20439980 N° de téléphone / Portable : +77418158489

Dans quelle orientation du MSc êtes-vous ? Aquatic-science

Données du mémoire

Titre du mémoire en français :
PRÉDIRE LA DYNAMIQUE ÉPILIMÉTIQUE DU PHOSPHORE DU LAC LÉMAN PAR UNE APPROCHE DE MODÈLE HYBRIDE

Titre du mémoire en anglais :
PREDICTING THE EPILIMETIC PHOSPHORUS DYNAMIC OF LAKE GENEVA USING AN HYBRID MODEL APPROACH

Directeur·trice(s) du mémoire : Marie-Elodie Perga

Expert·e (s): Tomy Doda

Si l'expert est externe de l'Université de Lausanne, merci de spécifier son adresse professionnelle :
tomy.doda@eawag.ch

Données défense orale

Date de la défense orale : 28/6/2023 Heure : 11:00

Date : 6/14/2023 **Signature :**

Les modalités et délais du dépôt de mémoire se trouvent dans les directives de mémoire du master MSc.

Ne pas remplir par l'étudiant·e

Date de dépôt du mémoire, signature et tampon :

Acknowledgments

I would like to express my gratitude to my supervising professor, Dr. Marie- Elodie Perga, for contributing to this study with her considerable expertise and giving me the guidance and encouragement to carry out this thesis and giving me the opportunity to work in the LAKES Group at the University of Lausanne.

I would like to extend my thanks to Dr. Gaël Many who provided me with the data to conduct this work.

I would also like to acknowledge Paul Hanson who not only took the time to guide my research process, but he kindly facilitated his code that helped me to build my own.

Furthermore, I would like to express my appreciation to my friends Trini, Matta, Rosi, Ami, Rafa, Chad, Gi, and Peter for their unwavering support, love, and belief in my abilities have been a constant source of motivation.

Finally, I would like to dedicate this achievement to my parents. Thank you for being my guiding light, for supporting me along the road, and for influencing my life for the better. Specially, I owe my deepest gratitude to my partner Camila, that has been here for me every single time that I needed support, care, and companionship. I could not have made it this far if it was not for you.

Gabriel Juri Álvarez

Lausanne, June 14th, 2023

Abstract

Lake Geneva represents one of many successful cases of phosphorus manage that by controlling P input, the concentration substantially diminished, and a mesotrophic state could be recovered, however, P concentrations are still above the CIPEL objectives. Since the relevance of phosphorus in the Lakes water quality, mechanistic models have been developed to predict and understand the phosphorus concentration dynamics, and hence develop better management strategies. Mechanistic models are hardly generalizable and prompt to bias, and they are derived from shallow or moderately deep lakes, showing a poor performance or a not integrated intra/inter annual P variations for Lake Geneve.

In this context Recurrent Neural Networks (RNN), Machine Learning approach have been shown to improve accuracy and predictability, however since internal structure is in general not accessible ("black-box") it may generate not-real results or impractical solutions. As an alternative, Process-Guided Recurrent Neural Network or hybrid model (PGRNN) combines the strengths of both RNN and theory-driven approaches, performing highly accurate predictions with little biased, without losing complexity and remaining acceptable for the physical/natural laws. In this way, The present work aims to enhance the current knowledge state of the epilimnetic phosphorus dynamics in Lake Geneva by building a mechanistic approach (PLGM), RNN and PGRNN.

Despite the concerns from overestimations and uncertainties found in the predictions, this work showed that through a simple mechanistic model it is still possible to simulate the epilimnetic phosphorus in Lake Geneva with high accuracy and proportion of variance. Moreover, by integrating PLGM knowledge in a hybrid model the P concentration dynamics were greatly improved, overpassing the obtained with RNN. When analyzing the relevance of the driven variables, The Rhône River and Entrainment appears as main external and internal phosphorus load, reaffirming its described relevance for Lake Geneva. In addition, the vertical diffusion flux of phosphorus (Hvd) appears as a new important internal P load in Lake Geneva when entrainment is suppressed by a strong thermocline, but which can be perturbed by the wind force.

Keywords: Phosphorus dynamic, Lake Geneva, Mechanistic model, Process-guided Machine Learning, Mixing, Stability

Contents

List of Figures iv

List of Tables vi

1. Introduction 1

    1.1 Aims & Objectives 6

2. Methods 8

    2.1 Study site 8

    2.2 Mechanistic models and Machine learning approach 8

        2.2.1 Mechanistic model of Lake Geneva 8

        2.2.2 Recurrent Neural Network (RNN). 17

        2.2.3 Process-Guided Recurrent Neural Network 19

        2.2.4 Models training and optimization. 21

        2.2.5 Metrics performance 25

        2.2.6 Sensitivity analysis 26

    2.3 Data 27

3. Results and Discussion 28

    3.1 Mechanistic model results 28

        3.1.1 Overall Assessment 28

        3.1.2 Seasonal Assessment 32

    3.2 Benchmarking the performance for PGLM, RNN, and PGRNN 37

        3.2.1 Performance comparison 37

    3.3 Sensitivity analysis 40

4. Conclusion 45

5. References 47

List of Figures

Figure 1. a) Observed vs. modeled phosphorus (μgL−1) for Lake Geneva, using three dynamic models: LakeMab, the JPJS, and PCLake. The three thin-dashed lines describe empirical annual average data ± one standard deviation (Figure modified from Bryhn & Håkanson., 2007). b) Observed vs. modeled surface phosphorus (μgL−1) for Lake Geneva (from January of 1964 to 1984) using: LakeMab (modeled lake), Vollenweider model, and the OECD model (Figure extracted from Håkanson& Bryhn., 2008). 4

Figure 2. A simplified structural description of a P dynamic model of three compartments (Figure extracted from Bryhn & Håkanson., 2007). TP refers to total phosphorus, ET-sediments refer to TP in erosion and transport sediments, A-sediments to TP 5

Figure 3. Iterative relationship between model-building steps (Modified from Jakeman et al., 2006). 9

Figure 4. Visual representation of the internal process that governs the phosphorus dynamic in Lake Geneva when it is stratified. $Sedn_E$ and $Sedn_H$ denote the sedimentation process from the epilimnion to the hypolimnion, and hypolimnion to the sediments. EntrP and Hvd represent the two mechanisms of phosphorus reincorporation from the hypolimnion into the epilimnion, while release does the same but from the sediments into the hypolimnion. The black line represents the thermocline in Lake Geneva. 10

Figure 5. Visual representation of the phosphorus dynamic components when Lake Geneva is not stratified, thus fully mixed. *Sedn* denotes the sedimentation process from the entire water column to the sediments. Release refers to the phosphorus reincorporation from the sediments to the water and burial corresponds to the no longer available P in the sediment. 12

Figure 6. Unfold neural network of RNN with one hidden layer (Ren et al., 2020). 18

Figure 7. Structure of the RNN and GRU (Modified from Ren et al., 2020). 19

Figure 8. Left: The flow of the Phosphorus PGRNN model (Figure modified from Jia et al., 2019). Right: PGRNN architecture (Hanson et al., 2020). 19

Figure 9. Boxplot of the data distribution for the predicted phosphorus concentrations by PLGM in the epilimnion, hypolimnion, and whole water column (red lines) compared to the observed data (black lines)..... 29

Figure 10. Phosphorus concentrations evolution between 1981 and 2016 for the predicted phosphorus in the whole water column (top panel) the epilimnion (middle panel) and, the hypolimnion (bottom panel). The time series in colors represent the PLGM predicted values, whereas the observed data is presented in grey. 31

Figure 11. Boxplot of seasonal data distribution for the predicted phosphorus concentrations by PGLM in the epilimnion, hypolimnion, and whole water column (red lines), compared to the observed data (black lines)..... 33

Figure 12. Boxplot of epilimnetic phosphorus concentrations distribution for the predictions done by PLGM, RNN, and PGRNN. Here LG refers to the observed data in Lake Geneva (LG)..... 37

Figure 13. Epilimnetic phosphorus dynamic comparison between PLGM, RNN and PGRN models. 40

Figure 14. PLGM, RNN and PGRNN sensitivity analysis..... 41

List of Tables

Table 1. Reference table on phosphorus-oriented models with the respective reference, model type (n° of dimension), name, and the time scope (Long or short term). 3

Table 2. Free parameters used for the mechanistic model, based on the work of (Hanson et al., 2020) 23

Table 3. Explored hyperparameters and their values range. '-' denotes the absence of the parameter in the model..... 25

Table 4. Variables used to model the mechanistic model, recurrent neural network and the process guided recurrent neural network. On the Table it is indicated the respective abbreviation, description, units and the model in which is implemented (Indicated by 'x')..... 27

Table 5. Interquartile ranges (IQR), coefficients of variation (CV), Mean Error Percentage (MEP), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Coefficient of Determination (R^2) obtained for PLGM phosphorus in the epilimnion (EpiP), hypolimnion (HypP) and whole water column (WcP). 30

Table 6. Interquartile ranges (IQR), coefficients of variation (CV), Mean Error Percentage (MEP), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Coefficient of Determination (R^2) obtained by seasons for PLGM phosphorus in the epilimnion (EpiP), hypolimnion (HypP) and whole water column (WcP)..... 34

Table 7. Interquartile ranges (IQR), coefficients of variation (CV), Mean Error Percentage (MEP), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Coefficient of Determination (R^2) obtained for PGLM, RNN, and PGRNN..... 38

Table 8. Average seasonal phosphorus dynamic simulated by PLGM, RNN and PGRNN models 39

Table 9. Overall seasonality of phosphorus flux showed by the entrainment (EntrP), vertical diffusion (Hvd), and external phosphorus load from Rhône River (LoadP) 43

1. Introduction

Human activities have historically caused the increment of the phosphorus concentration, eutrophication, and the detriment of the water quality of several lakes, detonating the concerns and multiple actions from governments to address this problem (Siuda et al., 2020). In this context, numerical and mechanistic models have been developed to predict and understand the lake phosphorus (P) concentration (Table 1) to develop better strategies to manage these aquatic systems. Nonetheless, this is complex phenomena since many variables influence differently the lake P dynmics depending on lakes characteristics such as climate (E.g., ambient temperature), biogeochemical processes, wind disturbance, inflow-outflow relationship, lake morphometry, trophic state, mixing degree, light, and water temperature (Krishna et al., 2021; Rocha & Neto., 2022). In this way, it is not rare the lack of consensus among specialists regarding the most relevant driven variables and the needed complexity to accurately describe the lake P dynamics, even if the aim is to simulate it at the local spatial scale (Jakeman et al., 2006; Khaither & Erechtchoukova et al., 2007; Taranu et al., 2017; Liu et al., 2017).

Most phosphorus models are numerical, empirical, and mechanistic approaches divisible into two classes: static and dynamic models (Bryhn & Håkanson., 2007; Fornarelli et al., 2013). The static models consist of inferential relationships between the driven variables and the target variable (phosphorus concentration), assuming an utterly mixed water column with a constant long-history phosphorus input. In contrast, dynamic models incorporate an ensemble of ordinary or partial differential equations of driven and target variables over time, simulating the dynamic's physical and biogeochemical processes (Bryhn & Håkanson., 2007; Fornarelli et al., 2013; Hanson et al., 2020; Dresti et al., 2021). Dynamic models have also been developed to model deeper aquatic systems by incorporating two or more compartments. The latter considers the external (phosphorus input from rivers, pipes, etc.) and internal phosphorus sources (E.g., hypolimnetic and sediment's P into the epilimnion), hence being more complex than static approaches since they incorporate non-linear relationships and variables. These types of models are applied when the lake's hydrodynamics modulate the interaction between the water body and the sediment, which is usually

the case in moderately/deep lakes (Janse et al., 2012; Hirayama., 2003; Jensen et al., 2006; Hanson et al., 2020).

Static and dynamic models can have zero up to three dimensions, and one up to multiple compartments. In contrast to dimensional models (one to three dimensions), the zero-dimensional model (0D) lacks physical dynamics. One-dimensional models (1D) assume horizontal homogeneity of the water column since the vertical hydrodynamics are much slower, as in deep and stratified lakes (Dresti et al., 2021). Two-dimensional models (2D) consider a horizontal or vertical homogeneity, restricted to shallows and elongated reservoirs in narrow valleys, respectively. Finally, the three-dimensional models (3D) simulate the physical and biogeochemical processes in a spatial and temporal continuity, requiring extensive field measurements, computational resources, and both spatial calibration and validation (Dresti et al., 2021).

Most developed P models are mechanistic, one-dimensional, and built based on empirical data of shallow lakes (mean depth Zm < 15m) or, to a lesser extent, based on narrow and moderately deep lakes (Zm ≤ 70m) (Ahlgren et al., 1988; Bryhn & Håkanson., 2007; Fornarelli et al., 2013; Tasnim., 2020; Hanson et al., 2020; Khorasani & Zhu., 2021). To the best of this work knowledge, there are at least five phosphorus models oriented to understand and simulate the lake's P dynamics (Table 1), from which some incorporate two or multiple compartments that separate the P sediment from the water column. These models are the PCLake model (Aldenberg et el., 1995), LEEDS model - also called LakeMab- (Malmaeus & Håkanson., 2004), JPJS model (Jensen et al., 2006), and a nameless model by Hanson et al. (2020), referred here as the authors' last names, Hanson model (Hanson et al., 2020). Ecosystem models, as previously mentioned, have not only allowed us to understand these aquatic systems, but also to elaborate effective strategies to recover the water quality of many lakes.

Table 1. Reference table on phosphorus-oriented models with the respective reference, model type (n° of dimension), name, and the time scope (Long or short term).

Article	Model type	Model	Time
Aldenberg et el. (1995)	0D	PCLake	Long-term
Malmaeus & Håkanson. (2004)	1D	LakeMab	Long-term
Jensen et al., (2006)	1D	JPJS model	Long-term
Hanson et al. (2020)	1D	Hanson model	Long-term
Krishna et al. (2021)	1D	GOTM-FABM-ERGOM	Short-term

Lake Geneva represents one of many successful cases of the lake's nutrient water management since the nutrient recovery after the intense nutrient input, algal bloom, and water detriment where the governmental efforts reduced a phosphorus (P) concentration of 90 µgL-1 and revert the lake eutrophic state in the 1970s to 16 µgL-1 and mesotrophic condition nowadays (Tadonleke., 2009; CIPEL., 2019). Despite the latter, the P concentration in Lake Geneva is still considered above the objectives fixed by CIPEL for the recovery of water quality, which is between 10 to 15 µgP/L. This may be attributed to the phosphorus introduced by the Rhone River which represents about 70 to 75% of its water inflow (Schwefel., et al 2016) and the main external P source (CIPEL), which originated in 90% of from non-point sources. In addition, complete mixing events that sporadically occur every decade in sum to deep mixing events fertilize the epilimnion and are essential in the phosphorus concentration measured in the epilimnion (Schwefel et al., 2016., Krishna). For example, Krishna et al. (2021) modeled the P content in Lake Geneve to analyze the three times maximum phosphorus in 2012 compared to 2013, from which infer that greater P turnovers linked to the meteorological conditions is probably explaining this difference. Internal P load in combination with changes in food web structure and climate patterns could heavily delay the remedial efforts focused on P point sources (Jarvie et al., 2017; Bol et al., 2018; Moisset et al., 2019).

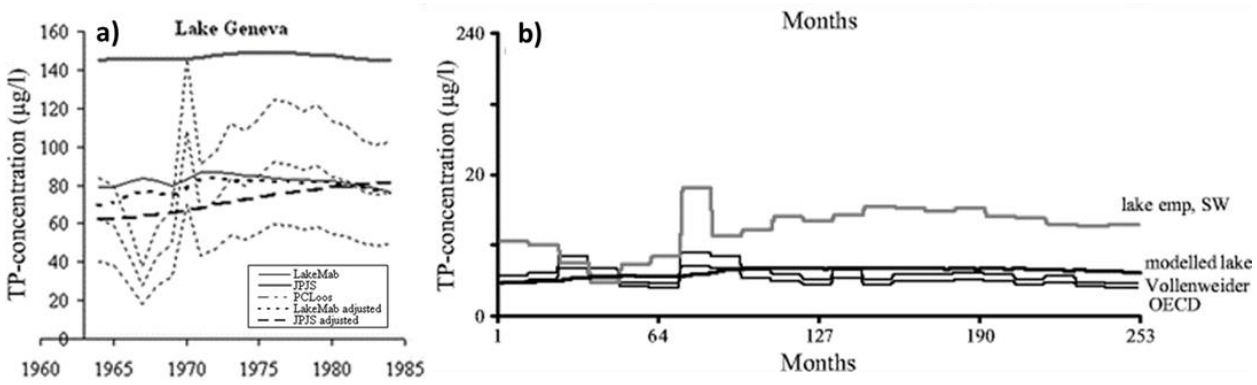

Figure 1. a) Observed vs. modeled phosphorus (µgL−1) for Lake Geneva, using three dynamic models: LakeMab, the JPJS, and PCLake. The three thin-dashed lines describe empirical annual average data ± one standard deviation (Figure modified from Bryhn & Håkanson., 2007). b) Observed vs. modeled surface phosphorus (µgL−1) for Lake Geneva (from January of 1964 to 1984) using: LakeMab (modeled lake), Vollenweider model, and the OECD model (Figure extracted from Håkanson& Bryhn., 2008).

Additional attempts have been made to simulate in a long term the P dynamics in Lake Geneva, for instance the LakeMab model is intended to work for deep and shallow lakes, however it has shown the worst performance for Lake Geneva among the P-oriented models in Figure 1a. LakeMab showed to overestimated the phosphorus concentration, whereas the rest of the model predictions fell in the ±1 standard deviation (SD) of the observed data but without inter-annual P cycles observed in Lake Geneva (Robarts et al., 2005; Xu et al., 2017; Roubeix et al., 2020). The modeling of Lake Geneva phosphorus dynamics gets even impossible by applying static models, such as the OCDE and Vollenweider models (Figure 1b) (Håkanson& Bryhn., 2008). Considering that mechanistic models are hardly generalizable, one may expect the results observed in Figure 1a and 2b, especially considering that they were not only built based on specific lakes, but on shallow or moderately deep lake (compared to the large dimensions and deep depth of Lake Geneva).

Consequently, by implementing knowledge base modeling approaches (PROCESS or mechanistic models) with specific knowledge on Lake Geneva's phosphorus dynamics, one may be able to accurately predict, better explain and understand this phenomenon. In this sense, the meromictic nature and the long-history dynamics of the input-phosphorus of Lake Geneva one should consider a dynamic model of at least three compartments (surface water, bottom water, and sediments, Figure 2) to predict the lake phosphorus concentration (Bryhn & Håkanson., 2007; Hanson et al., 2020). Indeed, Bryhn & Håkanson (2007) found that dynamic and three-compartments models (Epilimnion, hypolimnion, and sediment) predict phosphorus concentrations with much higher certainty than static models or dynamic models with one- or two-compartment(s) in deeper lakes,

including the deep Lake Geneva. To avoid the risk of overfitting and bias one should limit the model complexity to the data paucity and interpretability through reducing the number of parameters, model dimensionality, dropping too detailed/not-relevant processes, and by assuming regular structural behavior that average generalizable compartments of the system (Jakeman et al., 2006; Khaither, & Erechtchoukova., 2007).

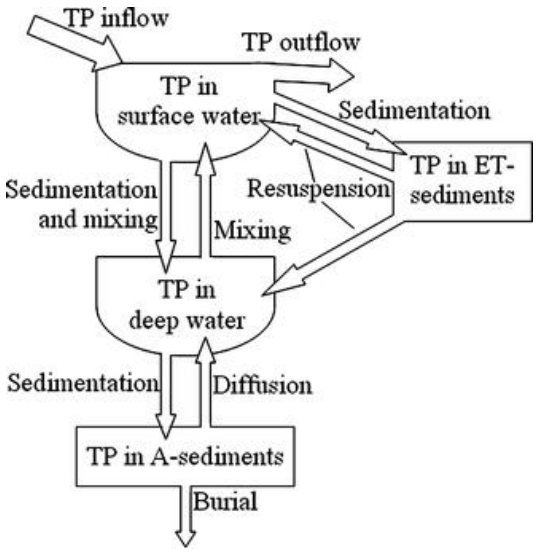

Figure 2. A simplified structural description of a P dynamic model of three compartments (Figure extracted from Bryhn & Håkanson., 2007). TP refers to total phosphorus, ET-sediments refer to TP in erosion and transport sediments, A-sediments to TP

In this line, Hanson et al. (2020) applied a relatively simple and one-dimensional (1D) dynamic model to predict, in the long-term, the phosphorus (P) concentration of Lake Mendota by only considering a few parameters and three compartments, for three generalizable lake sections (Epilimnion, hypolimnion and sediment). Surprisingly, the P prediction accounted for most of the observed P variance, where Hanson et al. (2020) explained it based on the inclusion of the most critical hydrodynamic process (external load, sedimentation, recycling, and export) and P fluxes of the studied aquatic system. Despite that Hanson model was built for a shallow lake and hence it is expected a similar result as the models in Figure 1, it still considers volumetric, hydro and thermodynamics to model the phosphorus content of three generalizable lake segments (Epilimnion, hypolimnion and the sediment) that may result as a good starting point for Lake Geneva.

Even though a mechanistic model may accurately predict Lake Geneva phosphorus dynamics, base knowledge approaches are prompt to bias and high levels of uncertainties. With the advances in

mathematical models and computing capacity, new perspectives, such as Machine Learning algorithms (ML) have shown to improve accuracy, efficiency, and predictability in ecosystem modeling problems (Wu & Marceau., 2002; Yang et al., 2008; Van der Molen et al., 1994). In this sense, given the temporal nature of the phosphorus dynamic, one is dealing with a time series forecasting problem, in which Gated Recurrent Neural Networks (RNN), an improved RNN version, have shown outstanding success in sequential modeling, achieving high prediction accuracy in the water quality prediction of lakes, including phosphorus prediction (Yan et al.,2021; Yu et al., 2022).

Despite the growing excitement about machine learning algorithms, they do not count in cause-effect mechanisms, ignore prior knowledge, and often their internal structure is not accessible ("black-box"), which may generate not-real results or impractical solutions for environmental management (Schuwirth et al., 2019). As an alternative, one could combine the strengths of mechanistic model and RNN (ML) into a Process-Guided Recurrent Neural Network or hybrid model (PGRNN), to perform highly accurate predictions with little biased, without losing complexity and remaining acceptable for the physical/natural laws. In the context of environmental sciences, hybrid models have been shown to improve the prediction accuracy in Lake modeling compared with mechanistic and machine learning models (Hanson et al., 2020), achieving good generalizability without violations of physical laws (Read et al., 2019). Beyond the predictability potential of a hybrid approach, together with a base knowledge model one may be able to unravel determinant variables and complex relationships (Hanson et al., 2020) that may give new insights to develop better manage strategies of Lake Geneva phosphorus.

1.1 Aims & Objectives

Consequently, the present work aims to improve the current knowledge state of the epilimnetic phosphorus dynamics in Lake Geneva, and to accomplish this, the following objectives are addressed:

a) Build a simplified phosphorus mechanistic model based on the main processes and driven variables behind the phosphorus dynamics of Lake Geneva.

b) Implement a Gated Recurrent neural network (RNN) and Hybrid model (PGRNN) with an optimized architecture for Lake Geneva, to then carry out a performance comparison of the mechanistic model, RNN, and PGRNN.

c) Analyze the relative relevance of the driven variables used in the three implemented models to gain new insights on Lake Geneva phosphorus dynamic, and their possible implications on the current state of Lake Geneva water quality.

The PROCESS Lake Geneva model (PLGM) will be conducted under the hypothesis that it is possible to simulate and explain, in a long-term, the epilimnetic phosphorus concentrations of the deep Lake Geneve by building a simple model. This approach considers the essential driven variables, hydrodynamic processes, and P fluxes.

Moreover, the PGRNN is performed under the hypothesis that by combining PLGM with a recurrent neural network algorithm (RNN), one can obtain a better phosphorus model performance and understanding, compared with an isolated implementation of a mechanist or machine learning approach.

2. Methods

2.1 Study site

Lake Geneva is a peri-Alpine lake located between the Alps and the Jura Mountains, where it is hydrologically supplied mainly by the Rhône River (70-75% of all inflows, Schwefel et al., 2016). Lake Geneva is known to remain stratified most of the time without being frozen in winter, and hence it is cataloged as a warm oligomictic lake (Perroud et al., 2009; Bengtsson., 2012; Gaillard et al., 2022). Complete winter mixing occasionally reaches the bottom of the lake (once every seven years on average), reaching 179 ± 20 m depth on average (Schwefel et al., 2016) . In addition, Lake Geneva has a curved and elongated shape, and it is considered as a large and deep lake based on the morphometric characteristics of this aquatic ecosystem such as its maximum and mean depth (310 m and 157 m, respectively) length of 70 km, a maximum width of 14 km, a volume of 90 km^3, and a surface area of 582 km^2 (Soulignac et al., 2021). These notorious morphometric characteristics change as one approaches the littoral zone of Geneva City, where the lake reaches its shallower maximum depth of 70 km, conveying to a narrower width of 4 km and a short length of 13.8km (Lemmin et al., 2005). High mountains characterize the basin of Lake Geneva, however it still experiences the influence of powerful winds across the entire expanse of the lake.

Lake Geneva has been a measured since 1958 producing the accumulation of a large data set with great temporal scale and quality (Rimet et al., 2020) presenting a great opportunity to study the dynamics of deep lakes.

2.2 Mechanistic models and Machine learning approach

2.2.1 Mechanistic model of Lake Geneva

In order to build/modify a phosphorus model for a deep lake as Lake Geneva, it was necessary to follow good model-development practices, considering the previous related works and the specific context/knowledge of these aquatic systems (morphometric characteristics, inflow sources, and outflows loss of phosphorus, hydrodynamics, etc). In the present work, the first steps of Jakeman's scheme indicated in Figure 3 were used to guide the construction of Lake Geneva P model (Jakeman et al., 2006).

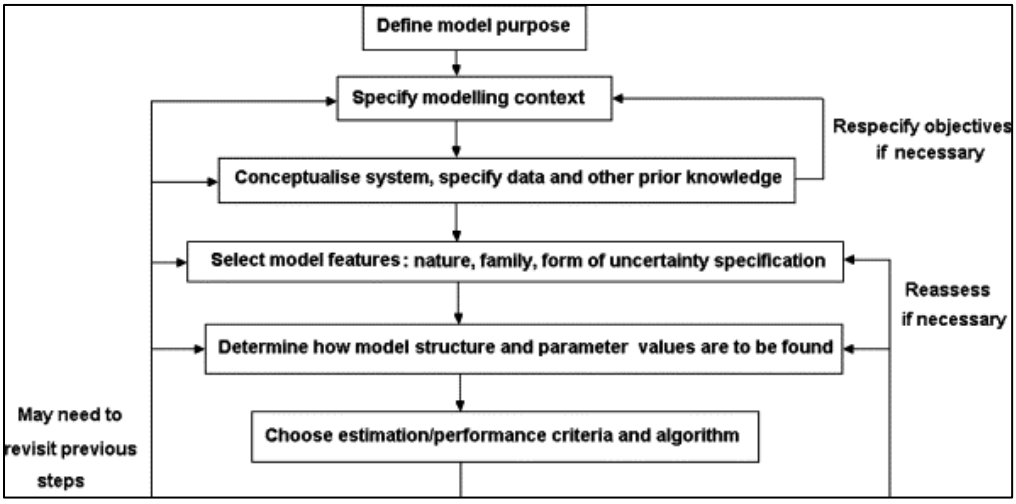

Figure 3. Iterative relationship between model-building steps (Modified from Jakeman et al., 2006).

The development of Lake Geneva phosphorus model was intended to follow a simple approach to perform a long-term phosphorus (P) simulation on a daily time step. The P model was done considering a simplified approach that includes the overall most essential driven variables, hydrodynamic processes, and P fluxes.

In this sense, the simplest appropriate model structure for Lake Geneva corresponded to a one-dimensional model (1D) composed of three-component that follows a mass-balance approach, where each compartment generalizes the phosphorus content of three generalizable lake segments: epilimnion, hypolimnion and the sediment in which the corresponding phosphorus concentration is given by the P difference between the input and output of each compartment.

Consequently, PROCESS Lake Geneva model (PLGM) was based on the same basic architecture as the Hanson or LakeMab model. Beyond the already proven poor P prediction LakeMab performance for Lake Geneva, The Hanson model was chosen as the starting point since it is a much simpler long-term model that not only considers volumetric dynamics but also incorporates thermal dynamics of the water through conditional statements that determine the seasonal behavior of P input and output of each compartment (Malmaeus & Håkanson., 2004; Perroud et al., 2009; Hanson et al., 2020; Soulignac et al., 2021; Krishna et al., 2021).

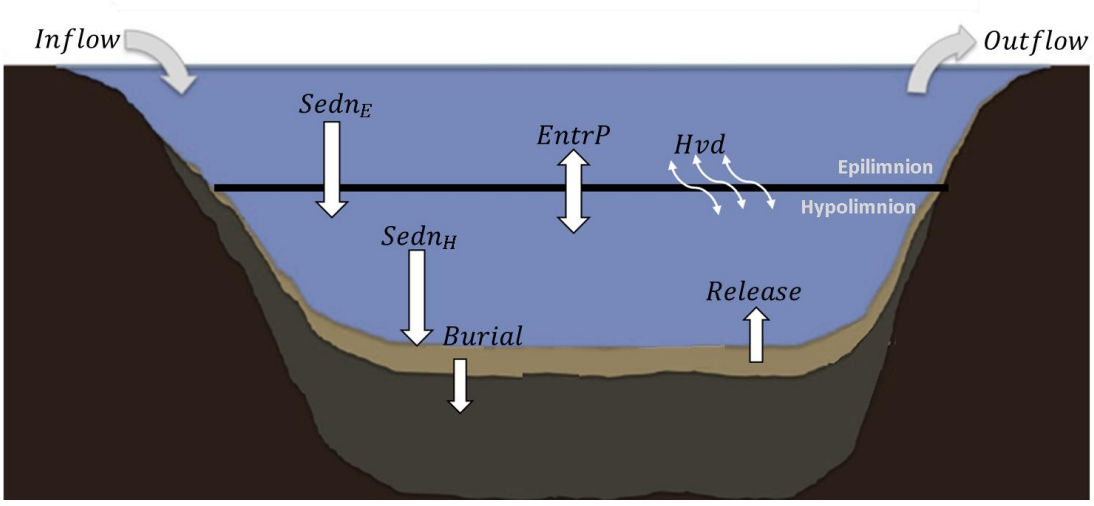

Figure 4. Visual representation of the internal process that governs the phosphorus dynamic in Lake Geneva when it is stratified. $Sedn_E$ and $Sedn_H$ denote the sedimentation process from the epilimnion to the hypolimnion, and hypolimnion to the sediments. EntrP and Hvd represent the two mechanisms of phosphorus reincorporation from the hypolimnion into the epilimnion, while release does the same but from the sediments into the hypolimnion. The black line represents the thermocline in Lake Geneva.

In the previous figure, the phosphorus inflow refers to the total external phosphorus load ($Load_P$) from Rhône River, the injection depth for this input corresponded to the one where the water density of the input was equal to the water density of the lake where both densities are computed based on Martin & McCutcheon approach (1998) assuming zero salinity. If during stratification this depth is within the epilimnion then the epilimnetic compartment in PLGM (Fig 4) receives the total external phosphorus load ($Load_E$), otherwise it is assigned to the hypolimnion ($Load_H$). The internal input is a crucial process in Lake Geneva (Krishna et al., 2021) which is also considered by Hanson's model by computing the phosphorus entrainment caused by the physical deepening of the mixed layer (Hanson et al., 2020). Nevertheless, the vertical transport into the epilimnion is incomplete without the P vertical diffusion (*Hvd*) since it can incorporate phosphorus content from the hypolimnion into the epilimnion upon the wind speed, delta P concentration, and water stability. The *Hvd* phenomenon has been described to be as important as the entrainment in other deep monomictic lakes (Merino-Ibarra et al., 2021), and a potential key factor in Lake Geneva (Michalski, J., & Lemmin., 1995; Lemmin., 2020).

The phosphorus losses in the epilimnion depend on the phosphorus *sedimentation* ($Sedn_E$), the phosphorus outflow, phosphorus export (E_p), the *Entrainment* (when the thermocline decreases in

depth), and the negative vertical diffusion from the epilimnion to the hypolimnion. In this way, the epilimnion compartment for the phosphorus is mathematically formulated as follows:

$$\frac{d\boldsymbol{P}_{epi}}{dt} = Load_E + Entr_P + Hvd - Sedn_H - E_P \qquad \text{/Eq. 1}$$

The hypolimnetic phosphorus content (P_{hyp}) receives as input $Load_H$, $Sedn_E$, *Hvd*, $Entr_P$ from the epilimnion when the thermocline decreases in-depth, and the P release from the sediment (*R*). Regarding P_{hyp} loss, it depends on the $Entr_P$, of phosphorus into the epilimnion, positive *Hvd*, and the sedimentation of phosphorus from the hypolimnion to the sediment ($Sedn_{Hyp}$) (Nürnberg., 2007; Hanson et al., 2020; Silvonen et al., 2021).

$$\frac{d\boldsymbol{P}_{hyp}}{dt} = Load_H + Sedn_E + R - Hvd - Entr_p - Sedn_H \qquad \text{/Eq. 2}$$

Regarding the sediment phosphorus (P_{sed}), it depends on the total phosphorus that comes from the hypolimnion. The P_{sed} loss depends on the phosphorus release into the hypolimnion (*R*) and the burial of phosphorus (Hanson et al., 2020).

$$\frac{d\boldsymbol{P}_{sed}}{dt} = Sedn_H - R - B \qquad \text{/Eq. 3}$$

Finally, under mixing conditions (Fig. 5) the water column of Lake Geneva is fully mixed, therefore the epilimnetic and hypolimnetic phosphorus concentration is assumed to be homogenous along the whole water body. In such case, $Entr_P$ and *Hvd* were set to zero since the absence of compartments differentiation, and the phosphorus concentration homogeneity. Regarding the sedimentation, instead of having two separate processes $Sedn_E$ and $Sedn_H$, it is considered as a unique sedimentation *Sedn*.

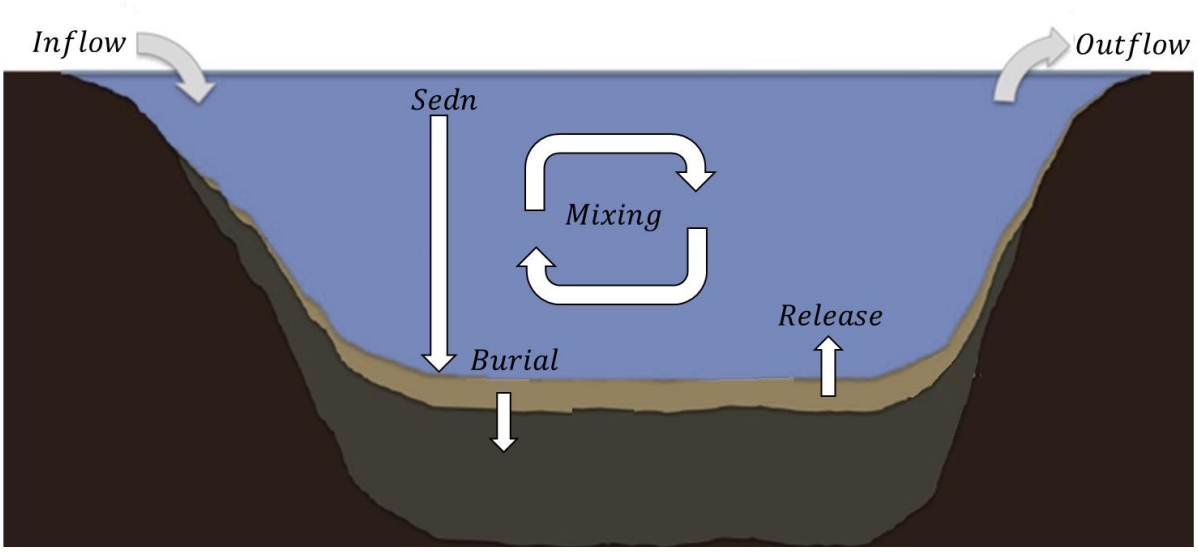

Figure 5. Visual representation of the phosphorus dynamic components when Lake Geneva is not stratified, thus fully mixed. *Sedn* denotes the sedimentation process from the entire water column to the sediments. Release refers to the phosphorus reincorporation from the sediments to the water and burial corresponds to the no longer available P in the sediment.

- Total epilimnetic and sediment phosphorus input

The external phosphorus inputs ($Load_P$) are the result of multiple sources such as streams, waterfowls (for P load), precipitation, wastewater treatment plants (WWTPs), and groundwater infiltration. In this sense, the major P source in Lake Geneva is originated from tributaries during surface-runoff events, accounting for 84% of the total input (Robertson., 2002). From Geneva tributaries, Rhône River represents the most import river in terms of water discharge, delivering approximately more than 85% of the river nutrient input (Klein., 2015). On the other hand, waterfowls and precipitation were the following most important P sources, contributing about 5 to 12 percent each. Regarding WWTPs, they contributed between 2 and 8 percent of the P between 1975 to 1998, but since the government regulation nowadays the percentual contribution should be much less relevant (Rapin & Gerdeaux., 2013).

Since the river is the most relevant phosphorus source, here it is assumed a safe neglection of the mentioned contributions. $Load_P$ is assigned to either the epilimnion or hypolimnion based on the depth at which the Rhône River water density is equal or lower the water column density. If that depth is inside the epilimnion then Rhône River phosphorus input is assigned to this compartment ($Load_E$), or in the contrary to the hypolimnion ($Load_H$).

- Entrainment

As proposed by Hanson et al. (2020), the *Entrainment* can take a positive or negative sign, depending on an increasing/growing epilimnion volume or decreasing/shrinking epilimnion volume as a function of the lake depth.

Under a growing epilimnion, the phosphorus *Entrainment* is the daily change in the epilimnetic volume d*EpiV* divided by the hypolimnetic volume *HypV* (m^3) and then multiplied by the respective daily hypolimnetic phosphorus concentration (P_{Hyp}), as expressed in Eq. 4.

$$Entr = \frac{dEpiV}{HypV} \times \boldsymbol{P}_{Hyp} \qquad \text{/Eq. 4}$$

In contrast, under a shrinking epilimnion, the *Entrainment* for the phosphorus is calculated by dividing *dEpiV* by the epilimnetic volume *EpiV* and then multiplied by the respective epilimnetic phosphorus concentration as expressed in Eq.5.

$$Entr = \frac{dEpiV}{EpiV} \times \boldsymbol{P}_{Epi} \qquad \text{/Eq. 5}$$

- Vertical diffusion

The vertical diffusive phosphorus movement (*Hvd*) between the center of the epilimnion and hypolimnion is calculated as:

$$Hvd = K_z \times \frac{\Delta C}{\Delta Z} \qquad \text{/Eq. 6}$$

Where $\Delta C/\Delta Z$ is the phosphorus concentration difference between the center of the epilimnion and hypolimnion, and *Kz* is the vertical eddy diffusivity ($m^2\ d^{-1}$) at the center of the metalimnion. There are many ways to estimate *Kz*, however, the one proposed by Yeates & Imberger (2003) (Eq. 7) was considered, since it is derived from a mechanistic scaling argument that captures the specific physical processes governing the stratification and mixing, being more suitable for a mechanistic model.

$$K_z = K_m * \frac{C \times N_z^2}{LN \times N_{max}^2} \quad \text{/Eq. 7}$$

Where K_m is the molecular diffusivity of heat (1.4x10 ms^{-2}), C is the unitless vertical mixing coefficient, 700 for Lake Geneva (Perroud et al., 2009), N^2 is Buoyancy frequency, and LN is the lake number. N_z^2 is N^2 computed at the bottom of the metalimnion, while N_{max}^2 is the maximum value that N^2 can take inside the metalimnion. Both N^2 and LN were calculated based on the same procedure used by Read et al (2011).

- Sedimentation

To compute the epilimnetic and hypolimnetic sedimented phosphorus ($Sedn_E$ and $Sedn_H$), it was assumed the same formulation done by Hanson et al (2020) for both $Sedn_E$ and $Sedn_H$, where the sedimented epilimnetic and hypolimnetic phosphorus are equal to their phosphorus concentration (P_{epi} and P_{hyp}), multiplied by the optimizable Arrhenius coefficient for temperature scaling of sedimentation (θ_{ESed} or θ_{HSed}), between 1 to 1.4, and the daily epilimnetic or hypolimnetic phosphorus sedimentation rate (C_{Psed} *or* C_{Hsed}, d^{-1}) as expressed below:

$$Sedn_E = \boldsymbol{P}_{Epi} \times C_{\boldsymbol{E}Sed} \times \theta_{\boldsymbol{E}Sed}^{\Delta T_{\boldsymbol{EpiSed}}} \quad \text{/Eq. 8}$$

$$Sedn_H = \boldsymbol{P}_{Hyp} \times C_{\boldsymbol{H}Sed} \times \theta_{HSed}^{\Delta T_{\boldsymbol{HypSed}}} \quad \text{/Eq. 9}$$

θ_{ESed} and θ_{HSed} were obtained from the model optimization constrained between 1.0-1.4. The temperature scaling coefficients (ΔT_{EpiSed} *and* ΔT_{HypSed}) were calculated as the daily observed temperature (°C d^{-1}) minus the base temperature for sedimentation ($T_{\boldsymbol{Eb}}$ and $T_{\boldsymbol{Hb}}$), as expressed below:

$$\Delta T_{EpiSed} = T_{\boldsymbol{Epi}} - T_{\boldsymbol{Eb}} \quad \text{/Eq. 10}$$

$$\Delta T_{HypSed} = T_{\boldsymbol{Hyp}} - T_{\boldsymbol{Hb}} \quad \text{/Eq. 11}$$

Since the sedimentation can change upon the characteristic of a lake, it was calculated based on the mathematical expression reported by Müller et al. (2014). Consequently, the phosphorus sedimentation rate (*Csed*, m yr^{-1}) is determined by the net P sedimentation *NS*, lake volume *V*, and lake phosphorus P, as expressed in Eq. 12.

$$C_{sed} = \frac{NS}{V \times P} \qquad \text{/Eq. 12}$$

Regarding *NS*, it was calculated as:

$$NS = \frac{C_{SMax} \times P}{\hat{Z}} \qquad \text{/Eq. 13}$$

Where C_{SMax} is the maximum sedimentation rate (14.5 m yr^{-1}). Consequently, the P sedimentation rate (C_{sed}) can be estimated as it is shown in the following equation:

$$C_{sed} = \frac{14.5}{V \times \hat{Z} \times 365} \qquad \text{/Eq. 14}$$

Where *V* is the lake volume and *Z* corresponds to the mean lake depth. In this manner, under unstratified conditions the water column of Lake Geneva is fully mixed, and the sedimentation process is considered as a unique process (Eq. 14). On the other hand, when Lake Geneva is stratified, the sedimentation is split into the epilimnetic and hypolimnetic sedimentation (C_{Psed} and C_{Hsed}) and calculated as:

$$C_{Esed} = \frac{14.5}{V_{Epi} \times \hat{Z}_{Epi} \times 365} \qquad \text{/Eq. 15}$$

$$C_{Hsed} = \frac{14.5}{V_{Hyp} \times \hat{Z}_{Hyp} \times 365} \qquad \text{/Eq. 16}$$

From equations, V_{Epi} and V_{Hyp} are epilimnetic and hypolimnetic volume, respectively, whereas Z_{epi} and Z_{hyp} represent the mean epilimnetic and hypolimnetic depth, respectively.

- Export

Hanson et al. (2020) calculated the daily phosphorus export E_P (in g d^{-1}), by multiplying the daily observed outflow Q_{out} (m^3d^{-1}) with the daily epilimnetic P content, P_{Epi} (in g d^{-1}), and dividing it by the daily epilimnetic volume V_{Epi} (m^3d^{-1}), as shown in equation Eq.17:

$$E_P = \frac{Q_{out}}{V_{Epi}} \times P_{Epi} \qquad \text{/Eq. 17}$$

Where V_{Epi} is equal to the Lake Volume (LV) multiplied by the volume vector of proportion (V^p) as expressed in Eq. 18, while V_{Hyp} was calculated as the LV minus V_{Epi} (Eq. 19).

$$V_{Epi} = LV \times V^p \qquad \text{/Eq. 18}$$

Where V^p is based on the accumulative volume (V_{acc}) at each depth from the surface (0m) until the bottom (309m) which is then divided by the total volume of Lake Geneva (89km^3). V_{acc} was computed at each depth as a linear interpolation between 0m and 309m.

$$V_{Hyp} = \text{LV} - V_{Epi} \qquad \text{/Eq. 19}$$

- Release

To integrate the phosphorus release into the mechanistic model it is considered the approach from the Modelling Aquatic Eco-Dynamics (Hipsey et al., 2020) which describes this process equal to the phosphorus in the sediment multiplied (P_{sed}), by the optimizable release rate (C_R), between 0.1 and $1e^{-5}$, by the optimizable Arrhenius coefficient for temperature scaling of sedimentation (θ^{T_R}), between 1 and 1.4, and by φ_{o_2} as the oxygen limiting factor:

$$R = P_{sed} \times C_R \times \theta^{\Delta T_R} \times \varphi_{o_2} \qquad \text{/Eq. 20}$$

In the temperature scaling the ΔT_R was determined by the daily temperature (°C d^{-1}) observed in the hypolimnion T_{opt} and the scaling base temperature (°C) as a free parameter.

$$\Delta T_R = T_{opt} - T_R \qquad \text{/Eq. 21}$$

Finally, the oxygen limiting factor is equal to the optimizable half-saturation (K_{sed}) of oxygen, between 20 – 100 mmol O_2 m^{-3}, divided by K_{sed} added to the oxygen concentration in the sediment of Lake Geneva (O_2):

$$\varphi_{O_2} = \frac{K_{sed}}{K_{sed} + O_2} \qquad \text{/Eq. 22}$$

- Burial

The burial (*B*) represents the phosphorus fraction in sediment that is permanently retained and no longer available for the water column dynamic. *B* was calculated as indicated by Hanson et al (2020):

$$B = P_{Sed} \times C_{Burial} \times SedDepth^{-1} \times 365 \qquad \text{/Eq. 23}$$

Where *Psed* is the phosphorus content in the sediment, C_{Burial} is the optimizable phosphorus burial rate (between 0.1 to 0.5), and *SedDepth* is the sediment depth available for the release of phosphorus (0.1m; Hanson et al., 2020).

2.2.2 Recurrent Neural Network (RNN).

The Recurrent neural network (RNN) was derived from the Artificial Neural Network (ANN) theory, and it is based on the memory of the experience. The RNN takes a previous computation or information as the input that will influence the current output, giving a network memory function. From the structure of RNN (Figure 6) one can observe that the nonlinear transformations of the inputs (x) done by the hidden layer (*st*) at time *t* will produce the corresponding wights (*W*) by combining it with the previously hidden layer output ($s_{t-1,}$). The latter will have the output weights (*V*) and outcome of the current hidden layer at time *t* (s_t) (Ren et al., 2020; Hanson et al., 2020).

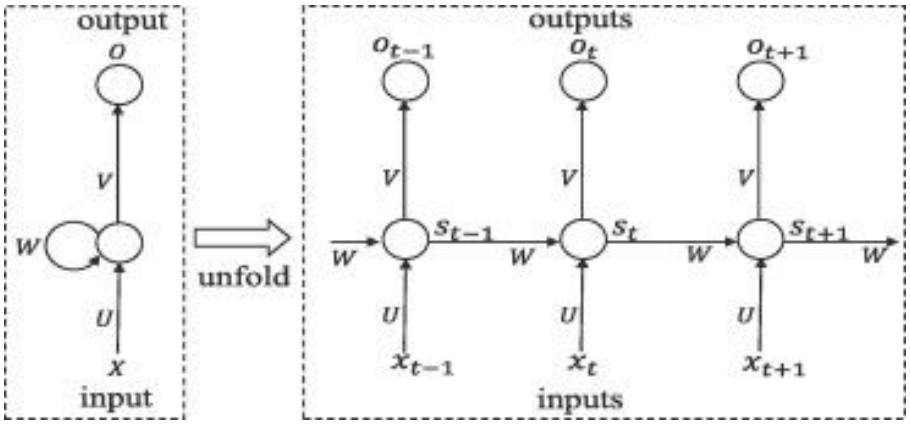

Figure 6. Unfold neural network of RNN with one hidden layer (Ren et al., 2020).

The hidden representation can be computed using the transition function as follows:

$$S_t = \sigma(Ux_t + Ws_{t-1} + b) \qquad \text{/Eq. 24}$$

Where σ is the sigmoidal function, Ux_t is the weight between the current hidden layer and input, W_{st-1} is the weight parameter between the current and previous hidden layers, and b is a model parameter to be tuned using true observations.

With hidden representation calculation, the output at the current time is represented as follows:

$$O_t = \sigma(V_t S_t + c_t) \qquad \text{/Eq. 25}$$

Combining the previous two equations makes it possible to observe the memory influence of many previous input values on the current output computation in the RNN. However, Since RNN has only one activation function *tanh*, the RNN long-term memory will be affected after using the Back Propagation Through Time algorithm for multiple time-steps training. Gated Recurrent Unit (GRU) can effectively define transition functions to memorize short-term and long-term computations (Ren et al., 2020; Hanson et al., 2020).

The GRU is characterized by having two gate components, update Z_t, and Reset r_t^j gates, that determine the amount of information to be updated and the amount of last hidden information to be deleted, respectively. Then, the update and reset gate information are scaled and added to the activation function *tanh* (Huang et al., 2019; Ren et al., 2020) (Figure 7).

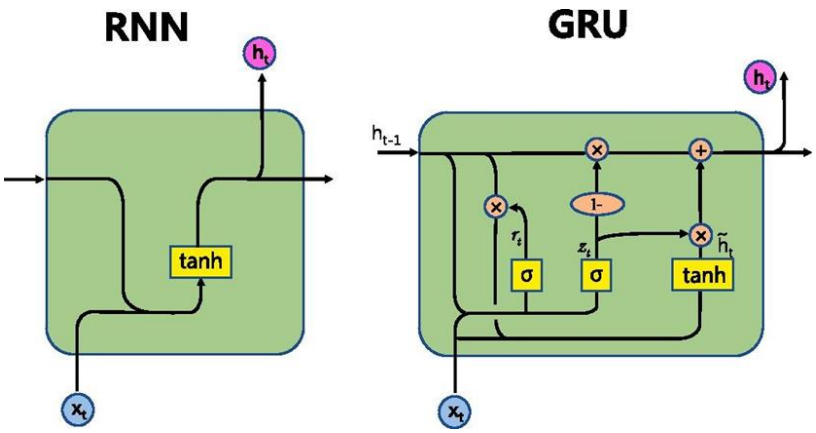

Figure 7. Structure of the RNN and GRU (Modified from Ren et al., 2020).

2.2.3 Process-Guided Recurrent Neural Network

Process-Guided Machine Learning is a hybrid model (HM) between the Recurrent Neural Network (RNN) and the PROCESS/mechanistic model, referred to as Process-Guided Recurrent Neural Networks model PGRNN (Hanson et al., 2020).

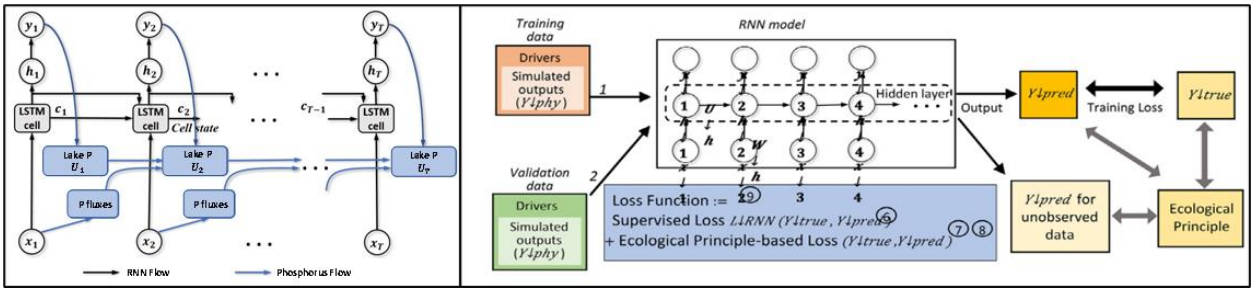

Figure 8. Left: The flow of the Phosphorus PGRNN model (Figure modified from Jia et al., 2019). Right: PGRNN architecture (Hanson et al., 2020).

In PGRNN, the observed values can be simulated by the PROCESS model ($Y_{Process}$), which will be an incomplete presentation given the limitation and biases of the mechanistic models discussed before. Here the principal benefit of PGRNN is leveraging the strength of RNN and PROCESS models, improving the overall prediction accuracy without physical laws violations (Jia et al., 2019; Hanson 2020). PGRNN architecture (Figure 8) integrates the mass-balance and conservation principles of the phosphorus through the PROCESS model ($f_{Process}$) to ensure that predicted values

follow those principles, meanwhile, PGRNN can also map the relationships from a series of input drivers to the phosphorus (target variable) across the time using the RNN model (f_{RNN}) (Jia et al., 2019; Hanson 2020). As a result, we obtain the PRGNN model (f_{PGRNN}):

$$f_{PGRNN} = [X, Y_{Process}] \rightarrow Y \qquad \text{/Eq. 26}$$

To close the gap between the $Y_{Process}$ and the true observations, an additional function is fitted in the residual modeling:

$$f_{Residual}(x) = f_{Process}(x) + g(x) \qquad \text{/Eq. 27}$$

The ecological constraint for PGRNN model is the power scaling law, which eliminates high and low phosphorus values that do not exist in the real world. Furthermore, to regularize daily predictions, an additional term in the loss function will be added to penalize predicted values that are larger than the 80 percentiles of the changes of the first differences of observed values, as follows (Hanson et al., 2020):

$$L_{deltas} = \sum_t ReLU(|\Upsilon_{t+1} - \gamma_t| - p_{0.8}(\Delta y))/T \qquad \text{/Eq. 28}$$

In addition, Hanson et al (2020) incorporated an ecological constraint term through the penalization of large power discrepancy between the frequency spectrum over time of the predicted and observed output (which was obtained by a Short Time Fourier Transform *STFT* with a window length 256) as follow (Hanson et al., 2020):

$$L_{stft} = \frac{\sum_t \sum_f |S_{t,f} - s_{t,f}|}{TF} \qquad \text{/Eq. 29}$$

Where S and s represent the obtained spectrogram for $\Upsilon_t$and γ_t, *TF* is the total number of frequency levels, $S_{t,f}$ and $s_{t,f}$ represent the power of the spectrum at time *t* and frequency *f*.

2.2.4 Models training and optimization.

PLGM aims to accurately predict and dynamically describe the epilimnetic total phosphorus and to achieve this, it needs a loss function ($PLGM_{Loss}$) as a metric to fit the model. Despite the main aim of this work towards the epilimnetic phosphorus of Lake Geneva, $PLGM_{Loss}$ sum of the epilimnetic and hypolimnetic losses (L_{Epi}^{x,ℓ_e} and $M_{Hyp}^{y,\ell_\hbar}$, respectively) as shown in Eq 30. This was done to get a balanced optimization between the accuracy and proportion of variance explained by the model.

$$L_{PLGM} = L_{Epi}^{x,\ell_e} + L_{Hyp}^{y,\ell_\hbar} \qquad \text{/Eq. 30}$$

L_{Epi}^{x,ℓ_e} is equal to RMSE added to R^2, where both metrics are multiplied by their corresponding regularization term (ℓ_{1e} and ℓ_{2e}, respectively) as formulated below:

$$L_{Epi}^{x,\ell_e} = RMSE_{EPI} \times \ell_{1e} + R^2{}_{EPI} \times \ell_{2e} \qquad \text{/Eq. 31}$$

Where RMSE is the sum of the squared differences between the predicted (y_i) and observed ($\hat{y}_i$i) values, which is divided by the number of observations (n). To this operation the squared root is applied as expressed in Eq. 32.

$$L_{RNN} = RMSE = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad \text{/Eq. 32}$$

R^2 instead is calculated as one minus the squared differences between y_i minus $\hat{y}_i$ (Sum of squares residual, RSS), and divided by the squared differences between y_i minus the average of all $\hat{y}_i$ values (Sum of squares total, SST).

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \hat{y}_i^{Mean})^2} \qquad \text{/Eq. 33}$$

From L_{Epi}^{x,ℓ_e}, the ℓ_{1e} and ℓ_{2e} are computed as is described in Eq 34 and 35, respectively.

$$\ell_{1e} = \alpha \times \ell_e \qquad \text{/Eq. 34}$$

$$\ell_{2e} = \ell_{e} - \ell_{1e} \quad \text{/Eq. 35}$$

As the epilimnetic loss, $L_{Hyp}^{y,\ell_{\hbar}}$ it is the hypolimnetic RMSE multiplied by its regularization $\ell_{1\hbar}$, plus the hypolimnetic R^2 times $\ell_{2\hbar}$.

$$L_{Hyp}^{y,\ell_{\hbar}} = RMSE_{HYP} * \ell_{1\hbar} + R^2{}_{HYP} * \ell_{2\hbar} \quad \text{/Eq. 36}$$

Where $\ell_{1\hbar}$ and $\ell_{1\hbar}$ are the regularization terms for RMSE and R^2 for the hypolimnetic phosphorus.

$$\ell_{1\hbar} = \ell_{h} \times \beta \quad \text{/Eq. 37}$$

$\ell_{\hbar}$ is just one minus ℓ_{e} minus, and $\ell_{2\hbar}$ is computed as:

$$\ell_{2\hbar} = \ell_{h} - \ell_{1\hbar} \quad \text{/Eq. 38}$$

Here ℓ_{e} is a free parameter, between 0.1 and 0.5, that balance the optimization of $L_{Epi}^{x,\ell_{e}}$ and $L_{Hyp}^{y,\ell_{\hbar}}$, and α and β are the weights for the respective hypolimnetic and epilimnetic RMSE and R^2, that balance the optimization between accuracy and proportion of the observed variance explained by the model. The PLGM free parameters (Table 2) are minimized using a least squared function. The latter should result in a more reliable representation of Lake Geneva phosphorus dynamics, and if L_{PGRNN} results in a worse performance, it would indicate missing relevant driven variables and/or relationships to be further investigated, which may improve the model performance and its generalization.

Table 2. Free parameters used for the mechanistic model, based on the work of (Hanson et al., 2020)

FreeParameter	Values range	Units
θ_{Esed}	1 to 1.4	unitless
θ_{Hsed}	1 to 1.4	unitless
T_{EpiSed}	5 to 15	°C
T_{HypSed}	5 to 15	°C
T_R	5 to 15	°C
θ_R	1 to 1.2	unitless
K_{sedP}	20 to 100	gO_2
C_{psed}	0 to 0.9	unitless
PSED	1 to 100	gP
C_{burial}	1e-1 ro 1e-10	mm y^{-2}
ℓ_e	0 to 0.5	unitless
β	0 to 0.5	unitless
α	0 to 0.5	unitless

As done by Hanson et al (2020), the loss function for the recurrent neural network (L_{RNN}) is just RMSE, while the PGRNN loss function (L_{PGRNN}) is expressed as the sum of L_{RNN} with the weighted Short Time Fourier Transform (L_{deltas}), and the weighted penalization of the 80 percentiles between predicted and observed values (L_{stft}).

$$L_{PGRNN} = L_{RNN} + L_{deltas} * \jmath_1 + \mu L_{stft} * \jmath_2 \qquad \text{/Eq. 39}$$

Where $\jmath_1$ and $\jmath_2$ are the regularization terms for L_{deltas} and L_{stft} , receptively, that controls their relative error that contributes to L_{PGRNN}.

Both machine learning models are trained and compiled with their respective loss functions and use the features scaled and reshaped using a MinMax (in a range from 0 to 1) and 100-time steps, respectively. The loss function for both machine learning processes depends upon a wide range of hyperparameters to explore and tweak (known as hyperparameter tuning) such as: the number of hidden layers and neurons, the activation and loss function, the optimizer, learning rate, regularization(s), epochs, and batch size, among many others (Domhan et al., 2015; Zhu et al., 2019). Hence, an optimum internal model parameter (Neuron weights) capable of maximizing a certain task performance, requires the pre-setting of a personalized model structure with a specific

hyperparameter set, which cannot be learned by the ML itself (Yang & Shami., 2020; Alsharef et al., 2022).

Here the manual hyperparameter tuning is a common approach (Müller et al., 2021), which usually requires expertise and an exhaustive search of multiple not linearly related combinations, making it a complex and time-consuming process (Snoek et al., 2012; Yang & Shami., 2020; Alsharef et al., 2022). The hyperparameter tuning is a crucial and complex step that turns ineffective and vulnerable to human bias when it is done manually, even by experts (Guo et al., 2019; Yang & Shami., 2020). In this sense, the automatization of hyperparameter tuning has become popular to automate the hyperparameter optimization to make ML a more feasible approach (Elshawi et al., 2019). Among the hyperparameter optimization approaches (HPO), the Bayesian optimization (BO) is an automatic algorithm that has been shown to outperform classic optimization techniques in the configuration of neural networks, resulting not only in a good performance but also requiring fewer iterations (Snoek et al., 2012; Cho et al., 2020; Yang & Shami., 2020).

In this way, the training optimization is implemented by using a Bayesian optimization approach with a 6-fold cross-validation, where the features are divided into training and validation data, by using 85% and 15% of the data, respectively, without shuffling. At each fold, the fitting is done with 83.3% of the training data and then tested with the remaining 16.7% of the training data. Afterward, the fitted model is evaluated with unseen validation data. The BO with a 6-fold cross-validation is then applied for both RNN and PGRNN models to optimize their hyperparameters and architecture and to finally perform the training process.

Table 3. Explored hyperparameters and their values range. '-' denotes the absence of the parameter in the model.

Hyperparameters	RNN	PGRNN
Batch size	60 to 240	
Learning rate	1e-2 to 1e-7	
Optimizer	SGD, ADAM, RMSprop	
Epochs	20 to 60	
Number of hidden layers	1 to 3	
Number of hidden neurons	15 to 100	
Activation function	Tanh, Sigmoid, Relu	
Dropoute	0 to 0.5	
$\jmath_1$	-	0 to 0.5
$\jmath_2$	-	

2.2.5 Metrics performance

Since the total phosphorus values are large in magnitude, a min-max normalization was applied to scale the data and obtain more interpretable performance values that will reflect the error of the model in standard deviations. Regarding the performance metrics are calculated overall and by season through the analysis of the mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R^2).

In addition, the mean relative error proportion (MEP) is calculated to estimate the phosphorus concentration error prioritizing the long-term patterns of the data. To accomplish this, high frequency variations (on the daily basis) are smoothed out by grouping and averaging the data by month and year (overall assessment), and for a seasonal analysis it is first filtered by season.

That said, MSE is calculated by just taking the square of RMSE, and hence $MSE = RMSE^2$. Regarding MAE, it is calculated by computing n^{-1} multiplied by the sum of the absolute difference between y_i and $\hat{y}_i$:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad \text{/Eq. 40}$$

Finally, the mean error percentage (MEP) is the inverse of n, multiplied by the sum of the absolute subtraction between each y_i and $\hat{y}_i$, divided by the absolute of $\hat{y}_i$:

$$MEP = \frac{1}{n}\sum_{i=1}^{n-1}\frac{y_i - \hat{y}_i}{\hat{y}_i} \qquad \text{/Eq. 41}$$

Beyond the accuracy determination of the model, additional statistical tools were used to provide a more complete perspective on how well the predictions are representing the observation. As a first approach, the difference between the observed and predicted distribution is overall acquired through the mean comparison of the predictions by a *t*-test (for normally distributed data) or Mann-Whitney test for one-to-one comparison, or their alternatives for multiple mean comparisons (Anova or Kruskal-Wallis test, respectively). Then the interquartile ranges (IQR) and coefficient of determination (CV) are computed to compare the variation at the central tendency and along the width range. IQR shows how the predictions within the 50% of the distribution are dispersed compared to the observations, and hence the uncertainty around the mean predicted values. CV can serve as an indicator of the uncertainty and robustness of the model output. Finally, the Wasserstein distance is used to evaluate how similar is the overall distribution of two variables, being the same when its result is zero and totally different when it is one (Lehmann et al., 2021; Talukdar et al., 2023).

2.2.6 Sensitivity analysis

To perform the sensitivity analysis, the change in magnitude of the performance is quantified by iterating one variable scaled from 0 to 10 with a 1-step (11 iterations in total), while the remaining variables are not changed. In this way, the variable that shows the highest deviation in the performance metrics will have a higher importance in the modeling process. This process is repeated for all the considered features and performed for the mechanistic, RNN, and PGRNN models.

2.3 Data

In order to simulate the phosphorus dynamic of Lake Geneva through a mechanistic, machine learning, and hybrid model, different combinations of variables were considered as it is shown in Table 4. The output of the 1D hydrodynamic SIMSTRAT model, previously obtained by the LAKE group at University of Lausanne was used to acquire the vertical water temperature profile and seasonal thermocline depth, from which it can be computed the Schmidt Stability, water density, epilimnetic and hypolimnetic volume. The phosphorus as the oxygen information is obtained from the French Long Term Ecological Observatory of Lakes. Finally, the variables range from 1981 to 2016.

Table 4. Variables are used to model the mechanistic model, recurrent neural network and the process guided recurrent neural network. On the Table it is indicated the respective abbreviation, description, units and the model in which is implemented (Indicated by 'x')

Variable			Models		
Abreviation	Description	Units	PLGM	RNN	PGRNN
Qout	lake outflow	m^3d^{-1}	x	x	x
St	Schimidt Stability	kJ m^{-2}	x	x	x
Pload	phosphorus load	g	x	x	x
SWR	short wave radiation	W m^{-2}	x	x	x
U	wind speed	m s^{-1}	x	x	x
HypV	Hypolimnetic Volume	m^3	x	x	x
EpiV	Epilimnetic Volume	m^3	x	x	x
SedV	Sediment's Volume	m^3	x		
EpiT	Epilimnetic temperature	°C	x	x	x
HypT	Hypolimnetic temperature	°C	x	x	x
SedT	Sediment's Temperatre	°C	x		
Pepi	Epilimnetic Phosphorus	g	x	x	x
Phyp	Hypolimnetic phosphorus	g	x		
Psed	Sediment's phosphorus	g	x		
EpiO	Hypolimnetic Oxygen	g	x		
HypO	Epilimnetic Oxygen	g	x		
SedO	Sediment's Oxygen	g	x		
Hvd-PLGM	Vertical diffusion phosphorus	g			
EntrP-PLGM	Entrainment phosphorus	g			
EpiP-PLGM	Epilimnetic Phosphorus	g			x
HypP-PLGM	Hypolimnetic phosphorus	g			x

3. Results and Discussion

3.1 Mechanistic model results

3.1.1 Overall Assessment

The epilimnetic phosphorus distribution predicted by the mechanistic model (PLGM) ranged from 2 µgP L^{-1} to 106 µgP L^{-1} with a median concentration of 29 µgP L^{-1}(Fig. 9). In contrast, the PLGM hypolimnetic P concentration (HypP) ranged from 14 µgP L-1 to 104 µgP L-1, with a median P of 36 µgP L^{-1}. The PLGM phosphorus for the whole water column (WcP) ranged from 14 µgP L^{-1} to 104 µgP $L^{-1}$with a median concentration of 36 µgP L^{-1}.

Moreover, comparing the median values of phosphorus distribution in Lake Geneva (LG), all the median P concentrations of the observed distribution are significantly different from the predicted median values, based on the Mann Whitney-u test results. In this sense, the epilimnetic phosphorus (EpiP) predicted by PLGM was 38% above the mean P observed for LG (18 µgP L^{-1}), the predicted HypP was 55% below the observed concentrations in LG (80 µgP L^{-1}), and WcP was 50% lower respect to the observed value (72 µgP L^{-1}).

Despite the latter, the predicted right-skewed distribution for EpiP, HypP, and WcP successfully matched their respective observed distribution in Lake Geneva. This was further confirmed by the Wasserstein distance test (wdt) for EpiP, HypP, and WcP predictions, resulting in wdt values (0.11, 0.12, and 0.1) that reflect high similarity to their respective observed distribution in Lake Geneva (LG).

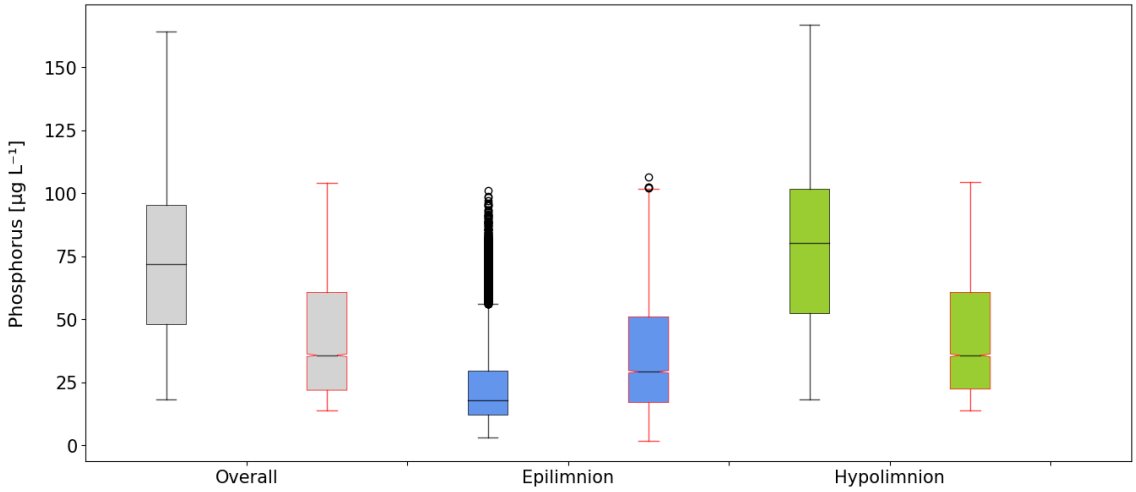

Figure 9. Boxplot of the data distribution for the predicted phosphorus concentrations by PLGM in the epilimnion, hypolimnion, and whole water column (red lines) compared to the observed data (black lines)

A further inspection of the descriptive statistic observed in phosphorus concentrations (Table 5), reveals that the coefficient of variation (CV) for observed and the predicted EpiP differ by just 9% (72 and 65 µgP L^{-1}, respectively), however, the predicted interquartile range (IQR) is higher by almost 100% with respect to the observed value (34.0 and 17.6 µgP L^{-1}, respectively).

In contrast, the predicted CV and IQR for P concentrations in the hypolimnion (58.2 and 38.4 µgP L^{-1}, respectively) were respectively 33% higher and 22% lower than the observed (CV: 39 µgP L^{-1} and IQR: 49 µgP L^{-1}). In the same line, the CV and IQR for the WcP (58 and 38.8 µgP L^{-1}) were 30% higher and 18% lower than the observed CV and IQR (41 and 47.4µg µgP L^{-1}). The moderate differences between the observed and predicted hypolimnetic CV and IQR reflect a moderate discrepancy between the observations and moderate levels of uncertainty along the whole predicted data distribution. In contrast, the higher difference in IQR for the predicted epilimnetic P gives a low level of confidence for the P predictions at the central distribution, but good confidence outside the 75th and 25th percentiles of the data (Heggy et al., 2022; Talukdar et al., 2023).

Table 5. Interquartile ranges (IQR), coefficients of variation (CV), Mean Error Percentage (MEP), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Coefficient of Determination (R^2) obtained for PLGM phosphorus in the epilimnion (EpiP), hypolimnion (HypP) and whole water column (WcP). Performance metrics values are normalized and hence they are dimensionless.

	Observed		Predicted						
	CV [µgP L^{-1}]	IQR [µgP L^{-1}]	CV [µgP L^{-1}]	IQR [µgP L^{-1}]	MEP	RMSE	MSE	MAE	R^2
WcP	41.0	47.4	58.3	38.8	-46.0%	0.206	0.043	0.168	49%
EpiP	71.7	17.6	64.9	34.0	43.6%	0.171	0.029	0.127	95%
HypP	38.9	49.4	58.2	38.4	-50.9%	0.211	0.045	0.177	55%

Regarding the mechanistic EpiP phosphorus concentration dynamics showed its variability well simulated within and among the years as it is displayed in Fig. 10. Despite this, EpiP maximum concentrations are clearly overestimated from 1982 to 1985, while the P minimum concentration are overestimated from 1986 to 1999. On the other hand, WcP and HypP dynamic concentrations were not totally incorporated by PLGM, and therefore, no intra-annual variations are seeing when comparing to the observed data.

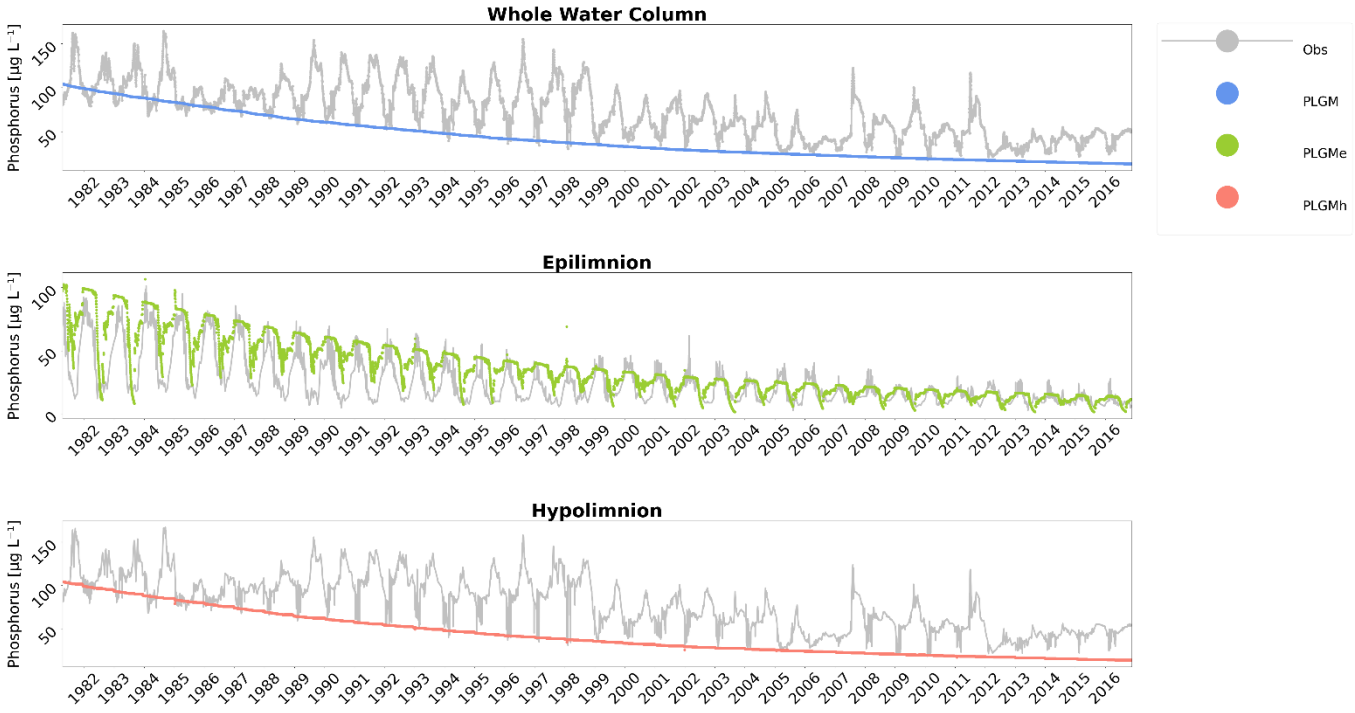

Figure 10. Phosphorus concentrations evolution between 1981 and 2016 for the predicted phosphorus in the whole water column (top panel) the epilimnion (middle panel) and, the hypolimnion (bottom panel). The time series in colors represent the PLGM predicted values, whereas the observed data is presented in grey.

The phosphorus concentration dynamics predicted by PLGM show a mean error percentage (MEP) that overestimated the average mean observed epilimnetic phosphorus concentration by 43%, and also considerably underestimated WcP and HyP by around 46 and 51%, respectively. The underestimation of WcP and HypP concentrations could be related to SIMSTRAT difficulties in accurately predicting the thermocline depth during low water column stability and a deeper stratification event, which may cause an incomplete representation of the water movements and its associated phosphorus turnover (Schwefel et al., 2016; Gaudard et al., 2017). However, the estimation discrepancies are homogeneously presented along the entire simulation of both, WcP and HypP, which may be indicating missing additional phosphorus sources either from the lack of information or overlooked processes. Despite the latter, PLGM still manages to predict the phosphorus in similar scales and relatively close to the EpiP, HypP, and WcP observations, and hence being successful at simulating overall phosphorus patterns in the epilimnion of Lake Geneva.

In terms of model accuracy (Table 5), PLGM demonstrated good general performance for EpiP, HypP, and WcP, as the Root Mean Square Error (RMSE), Mean Square Error (MSE), and Mean Absolute Error (MAE) values fell in a standard deviation range lower than ± 0.25. The goodness of fit showed a high proportion of the variance is explained for the epilimnion (R^2 95%), while moderate values were obtained for the hypolimnion (HypP) and whole water column (WcP) (R^2 55% and 49%, respectively). In this sense, there is a high agreement between the driven variables and processes with respect to the observed phosphorus for the epilimnion, but moderate for the hypolimnion, and for the whole water column (Heggy et al., 2022).

Wrapping up the overall assessment, the modeled IQR and CV for the HypP and WcP indicate moderate deviations in its predictions which give lower but still good confidence in the midrange predicted values, while its performance metrics (RMSE, MSE, and MAE) reflect a high accuracy for both, HypP and WcP. In addition, PLGM showed moderate explanatory power and hence incorporating a moderate variability proportion of the observation. In this way, PLGM gives a reasonable understanding of Lake Geneva P concentrations from the considered variables and underlying processes for HypP and WcP (see Section 3.3).

3.1.2 Seasonal Assessment

The seasonality of the epilimnetic phosphorus distribution predicted by the mechanistic model (PLGM) is illustrated in Figure 11. The results showed the lowermost EpiP median in summer (20 µgP L^{-1}), while the highest was present in spring and winter (34 and 36 µgP L^{-1}), and autumn had an intermediate concentration of 28 µgP L^{-1}. The PLGM mean phosphorus prediction in the hypolimnion (HypP) and the whole water column (WcP) was around 36 µgP L^{-1}through all seasons. The median phosphorus predicted in HypP and WcP contrasts with the real observations, since in summer and autumn Lake Geneva presented the maximum median phosphorus (91 and 87 µgP L^{-1}), while in winter the minimum (63 and 51 µgP L^{-1}) and slightly above the median P during spring (65 and 73 µgP L^{-1}). In this way, HypP and WcP median are 43 and 30% lower compared to observed concentrations in winter, around 60% lower for both, HypP and WcP, in summer and autumn, meanwhile in spring concentrations were 50 and 45% higher for HypP and WcP, respectively.

Despite that predicted EpiP showed a better match with respect to the observations, its median P during spring and autumn was 40% and 100% times higher than the observed values (21 and 14 µgP L^{-1}). Additionally, in summer the median P of the predicted distribution was 35% higher than the observed value (13 µgP L^{-1}), while winter was just 16% above respect to the real value (30 µgP L^{-1}).

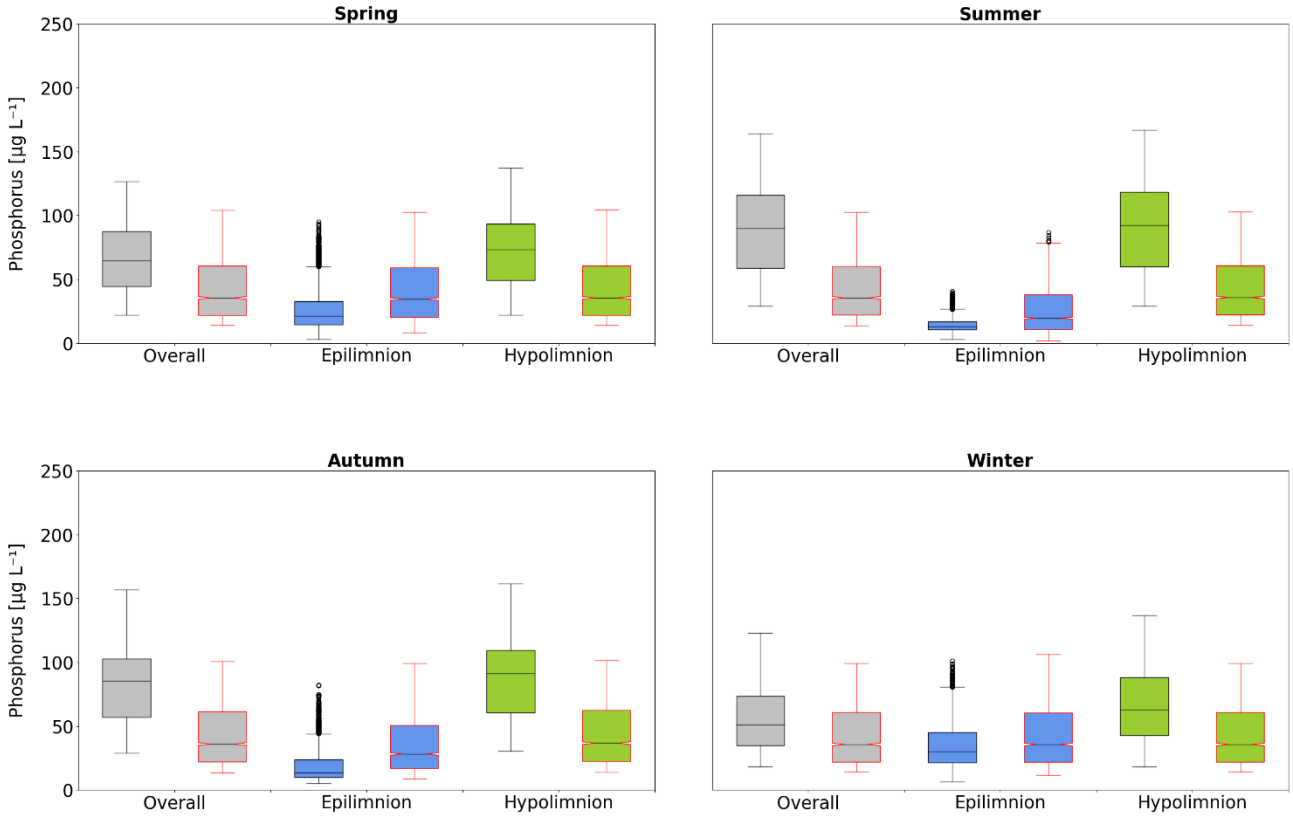

Figure 11. Boxplot of seasonal data distribution for the predicted phosphorus concentrations by PGLM in the epilimnion, hypolimnion, and whole water column (red lines), compared to the observed data (black lines).

From the discrepancies obtained in these results, it is expected that Wasserstein distance values (wdv) for HypP and WcP in winter reflect that their P distributions are quite similar to the respective real phosphorous distributions (wdv ~ 0.09), with slightly poorer results for the remaining seasons (wdv ~ 0.13). In contrast, PLGM epilimnetic phosphorus during summer and winter maintained its respective distributions almost identical to the observed in LG (wdv 0.07 and 0.05). summer, followed by autumn and spring (wdv 0.1).

Continuing with the distribution of PLGM phosphorus data, the seasonal interquartile ranges (IQR) for the epilimnion during spring and autumn were more than twice the values observed in LG, three

times in summer, and 40% above in winter. The predicted EpiP coefficient of variations (CV) during spring and autumn are ~ 10% lower than observed values, and 46% above during summer.

Table 6. Interquartile ranges (IQR), coefficients of variation (CV), Mean Error Percentage (MEP), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Coefficient of Determination (R^2) obtained by seasons for PLGM phosphorus in the epilimnion (EpiP), hypolimnion (HypP) and whole water column (WcP). Performance metrics values are normalized and hence they are dimensionless.

		Observed		Predicted						
		CV [µgP L^{-1}]	IQR [µgP L^{-1}]	CV [µgP L^{-1}]	IQR [µgP L^{-1}]	MEP	RMSE	MSE	MAE	R^2
WcP	spring	36.6	42.9	58.6	38.7	-41.4%	0.206	0.043	0.173	59.8%
	summer	36.9	57.0	58.8	38.2	-57.3%	0.232	0.054	0.197	51.3%
	autumn	33.6	46.1	58.2	39.3	-54.5%	0.191	0.036	0.161	64.4%
	winter	39.4	38.4	57.3	38.9	-27.9%	0.140	0.019	0.109	80.0%
EpiP	spring	66.4	18.1	59.8	38.9	47.9%	0.178	0.032	0.131	95.3%
	summer	38.9	6.4	71.5	27.4	60.3%	0.181	0.033	0.148	75.6%
	autumn	67.9	13.6	60.2	33.8	69.2%	0.163	0.027	0.115	95.9%
	winter	54.7	23.7	57.9	38.6	19.3%	0.097	0.009	0.078	90.0%
HypP	spring	36.2	44.3	58.6	38.8	-48.1%	0.229	0.052	0.197	58.6%
	summer	37.1	58.7	58.8	38.4	-57.6%	0.234	0.055	0.199	50.7%
	autumn	33.3	48.9	58.2	40.0	-56.8%	0.197	0.039	0.167	63.6%
	winter	39.2	45.3	57.2	38.9	-40.1%	0.183	0.033	0.151	62.9%

Consecutively, it is possible to detect a high uncertainty at the central tendency of EpiP prediction through all seasons, which is highly diminished as the predicted values fall outside the interquartile range with the exception of summer, according to the CV values obtained in this compartment. A more detailed analysis of the magnitude discrepancies in PLGM reveals that the mean absolute error percentages or MEP (Table 6) are more than 60% during summer and autumn, 48% in spring, and just 19% in winter. Despite the latter, RMSE, MSE, and MAE values for the predicted EpiP, HypP, and WcP are lower than ± 0.3 SD through the seasons, while their determination coefficients (R^2) are highly variable with values above 50%. In this sense, EpiP phosphorus predictions resulted in a high explained variance proportion during spring and autumn (R^2 ~ 95%), which was slightly

lower during winter (90%), and moderate in summer (R^2 ~ 75%). In contrast, the predicted HypP and WcP had moderate R^2 during autumn (R^2 ~ 64%) and moderate to good in winter (R^2 80 and 63%, respectively), while slightly lower values in spring (R^2 ~ 60%), and summer (R^2 ~ 50%).

Despite the discrepancies in the seasonal P average values (MEP) in the epilimnion, the time series concentrations produced by PLGM showed to simulate the seasonal logic behind the Lake Geneva P dynamic. For instance, during winter the relatively highest phosphorus concentration within the epilimnion, and lowest concentration within the hypolimnion can be explained through the positive relationship between the temperature and Schmidt Stability (*St*) where low water column temperatures are coupled with less stability, and hence a lake water column prompt to perturbations (Katsev el al., 2010; Read et a., 2011; Cossu, R & Wells., 2013; Rose et al., 2014). In winter *St* can reach such low values that the complete mixing of the water column is thermodynamically spontaneous, while during less intense cold periods as autumn, relevant upwelling can still happen when under low to medium *St* combined with important wind stress events trigger the P turnover that overcomes the thermocline. Complete mixing roughly happens every 10 years in Lake Geneva, and instead, the epilimnion is fertilized by the rise of water masses that transport phosphorus when the thermocline is moving to deeper depth (Schwefel et al., 2016; CIPEL., 2019). Because autumn is also characterized by cold but higher average temperate in winter, the physical forces would be weaker and less prompt to reincorporation equivalent (but still relevant) phosphorus levels into the epilimnion, as observed in the present results and previous studies in Lake Geneva (Schwefel et al. 2016). In contrast to winter, during warm periods the water stability and higher temperature set a strong and stable thermocline at shallower depths causing an impenetrable barrier for the hypolimnetic phosphorus, and hence explaining the PLGM lowest epilimnetic phosphorus in summer and moderate concentrations during spring compared to the other simulated seasons (Schwefel et al. 2016).

In general, high coefficients of determination and low errors often lead to accurate predictions that explain a high proportion of the observed variance. However, despite these values were observed for the predicted EpiP, HypP and WcP, there are still three relevant concerns to address regarding the confidence of these results: The overestimation of EpiP, and underestimation for both HypP and WcP through the seasons, the considerable discrepancies in EpiP, HypP, and WcP distribution

reflected by their CV, and the moderate explanatory power during warm seasons for WcP and HypP predictions.

These concerns may be the result of missing or poor address variables and processes that control the special phosphorus interaction (between the three Lake compartments) and the temporal phosphorus succession (from cold to warm periods). In addition, the difficulties of SIMSTRAT to accurately predict the thermocline depth during low water column stability and a deeper stratification event could be translated into an incomplete representation of the water movements and its associated phosphorus turnover (Schwefel et al., 2016; Gaudard et al., 2017). The estimation discrepancies of P turnover events may mislead the minimization algorithm in its task to find an optimum solution. In this sense, the lowest cost function may be inclined toward the optimization of PLGM epilimnetic phosphorus concentration, with its optimized free parameters according to the phosphorus movement from the hypolimnion to the epilimnion.

Another relevant factor to consider before the PLGM implementation is its simplicity level and its implication for the well-represented phosphorus dynamics of Lake Geneva. Even though recent studies had considered that Lake Geneva phytoplankton community is not limited by the P concentration but rather by the light intensity (Anneville et al., 2019), Phy may be an additional explanatory variable for P variations at high frequencies through Phy consumption, respiration, sinking, etc.

In the same line, another more probably determinant aspect is the dimensionality of PLGM, since it only considered the vertical component of the P dynamics and hence a 2D description. It is well known the relevance of the vertical mixing in the internal P supply for the epilimnion, however, studies have shown that 3D transport processes are also playing an important role in Lake Geneva phosphorus concentration (Beres et al., 2003; Reiss et al., 2022). This is especially important since horizontal dispersion can happen at an interbasin level and be exacerbated under prolonged winter stratification and strong wind perturbation (Umlauf & Lemmi., 2005). Indeed, Reiss et al (2022) argued that based on the interaction between the “Petite” and “Grand lac” 1-D models cannot correctly simulate Lake Geneva, and hence PLGM could not be describing well enough the phosphorus dynamics of Lake Geneva.

3.2 Benchmarking the performance for PGLM, RNN, and PGRNN

3.2.1 Performance comparison

The trained recurrent neural network (RNN) and process-guided recurrent neural network (PGRNN) of the epilimnetic phosphorus distribution are illustrated in Figure 12. The models had median P concentration at 24.6 and 15.9 µgP L^{-1}, being 28 and 11% above the P median observed for LG, and hence much closer than PLGM phosphorus median (63% higher than observed data in LG). However, a Mann-Whitney U test still showed a significant difference ($p<0.05$) between the machine learning models and the true observed phosphorus concentrations. In terms of distribution, a Wasserstein distance test reveals that RNN (wdt 0.17) is similar to the observed epilimnetic P distribution but lower than the PLGM value (wdt 0.11). In the case of PGRNN, the Wasserstein distance test showed its distribution was the most similar to LG phosphorus concentrations. Despite the latter, these three model approaches successfully reconstructed the overall underlying tendency, spread, and skewness of the observed data.

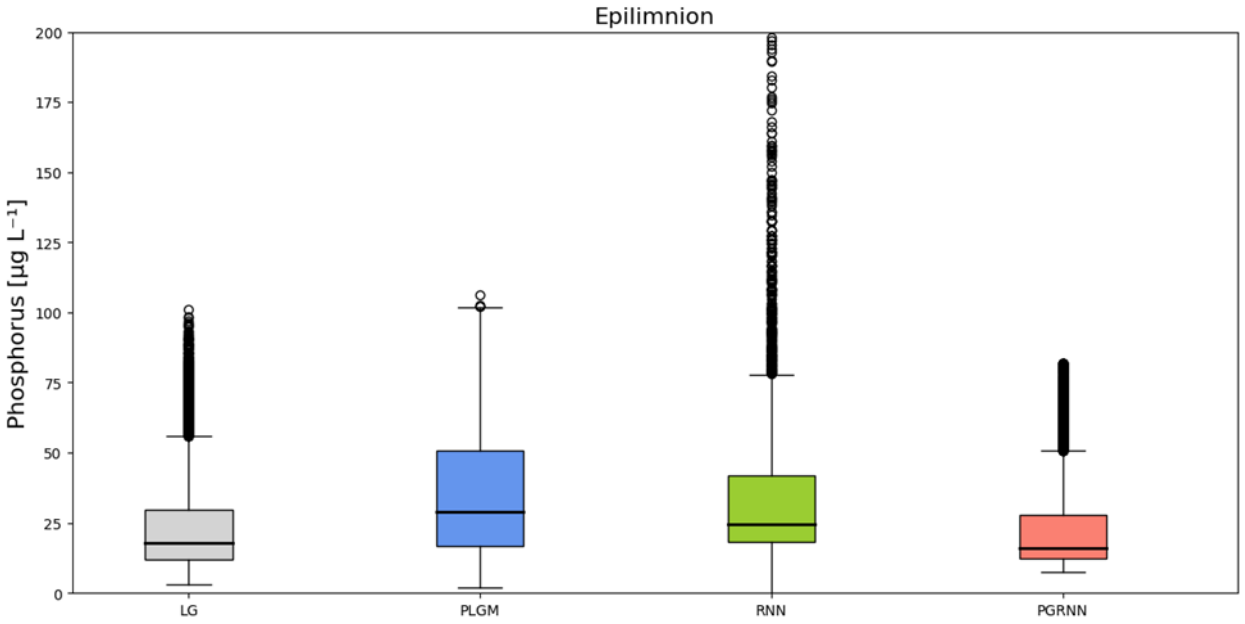

Figure 12. Boxplot of epilimnetic phosphorus concentrations distribution for the predictions done by PLGM, RNN, and PGRNN. Here LG refers to the observed data in Lake Geneva (LG)

Regarding the interquartile ranges (IQR), the RNN and PLGM values (24 and 34 µgP L^{-1}) were moderately and highly deviated from the observed P (17.6 µgP L^{-1}). In contrast, PGRNN had the

closest IQR (15.3 µgP L^{-1}), with less than 13% difference from the IQR for the observed data. Moreover, the coefficient of variations (CV) for both machine learning models slightly differed by 10% with respect to the observed data. Furthermore, the performance metrics being ≤ 0.01 SD and their coefficients of determination (R^2) relatively high (≥ 90%), the yearly average P concentration error (MEP) for PGRNN was marginally different from the observations (<1%) compared to the slight and moderate overestimation in the respective RNN and PLGM predictions (44 and 16%).

In this sense, the accurate predictions of the models are coupled with a highly explained variance proportion of the P observations. As stated in the previous section, the IQR values for RNN and PGRNN had lower uncertainty and probably more outlier resistance at mid-range predictions. Nevertheless, RNN showed a moderate CV deviation from LG data compared to the almost identical CV value presented by PGRNN. In this way, the hybrid model had the highest and more consistent confidence in the predicted values along its P distribution. In addition, PGRNN is obtained the lowest MEP values and hence the least monthly average P underestimation along the simulation.

Table 7. Interquartile ranges (IQR), coefficients of variation (CV), Mean Error Percentage (MEP), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Coefficient of Determination (R^2) obtained for PGLM, RNN, and PGRNN. Performance metrics values are normalized and hence they are dimensionless.

	Observed		Predicted						
	CV [µgP L^{-1}]	IQR [µgP L^{-1}]	CV [µgP L^{-1}]	IQR [µgP L^{-1}]	MEP	RMSE	MSE	MAE	R^2
PLGM	71.7	17.6	64.9	34.0	43.6%	0.171	0.029	0.127	95.2%
RNN	71.7	17.6	78.4	23.9	16.3%	0.239	0.057	0.171	89.5%
PGRNN	71.7	17.6	68.0	15.3	0.8%	0.083	0.007	0.062	96.5%

From the latter, one may expect comparable or even better PGRNN and RNN phosphorus dynamics representation than the obtained from PLGM. In this sense, both machine learning models presented accurate predictions and high explained variance proportion of P observation, which combined with their better overall IQR and CV values, more outlier resistance is expected at mid-range predictions. Nevertheless, RNN showed a moderate CV deviation from LG data compared to the almost identical CV value presented by PGRNN. In this sense, the hybrid model had the

highest and more consistent confidence in their predicted values and along its P distributions. In addition, PGRNN obtained the lowest overall MEP and the most identical predicted P distribution, reflecting lower average P underestimation that may be also resistant to the effects of outliers.

Table 8. Average seasonal phosphorus dynamic simulated by PLGM, RNN and PGRNN models. Performance metrics values are normalized and hence they are dimensionless.

		Observed		Predicted						
		CV [µgP L^{-1}]	IQR [µgP L^{-1}]	CV [µgP L^{-1}]	IQR [µgP L^{-1}]	MAE	MEP	MSE	RMSE	R^2
PLGM	spring	66.4	18.1	59.8	38.9	0.13	48%	0.03	0.18	95%
	summer	38.9	6.4	71.5	27.4	0.15	60%	0.03	0.18	76%
	autumn	67.9	13.6	60.2	33.8	0.11	69%	0.03	0.16	96%
	winter	54.7	23.7	57.9	38.6	0.08	19%	0.01	0.10	90%
RNN	spring	66.4	18.1	96.6	30.6	0.22	8%	0.08	0.29	89%
	summer	38.9	6.4	42.8	28.6	0.25	169%	0.09	0.30	0%
	autumn	67.9	13.6	45.8	8.9	0.10	16%	0.02	0.15	74%
	winter	54.7	23.7	75.8	15.5	0.27	-18%	0.10	0.32	77%
PGRNN	spring	66.4	18.1	57.0	16.1	0.07	3%	0.01	0.09	96%
	summer	38.9	6.4	15.6	2.5	0.18	-4%	0.05	0.23	54%
	autumn	67.9	13.6	55.6	8.0	0.05	-5%	0.01	0.07	97%
	winter	54.7	23.7	47.5	26.8	0.07	5%	0.01	0.09	92%

Nonetheless, this presumption is ultimately extended for RNN to its well-matched phosphorus variation in time, since its predictions also considered negative values and much more outliers than expected. In addition, during summer RNN had zero explanatory power and an IQR value that was almost four times the one in the observations. While in summer, the CV values for RNN were moderate but more deviated from LG compared to PLGM. In contrast, PGRNN had the best IQR and CV values than the other models for all seasons but not in summer, where the hybrid model coefficient of variation was 150% lower than the CV in LG observations.

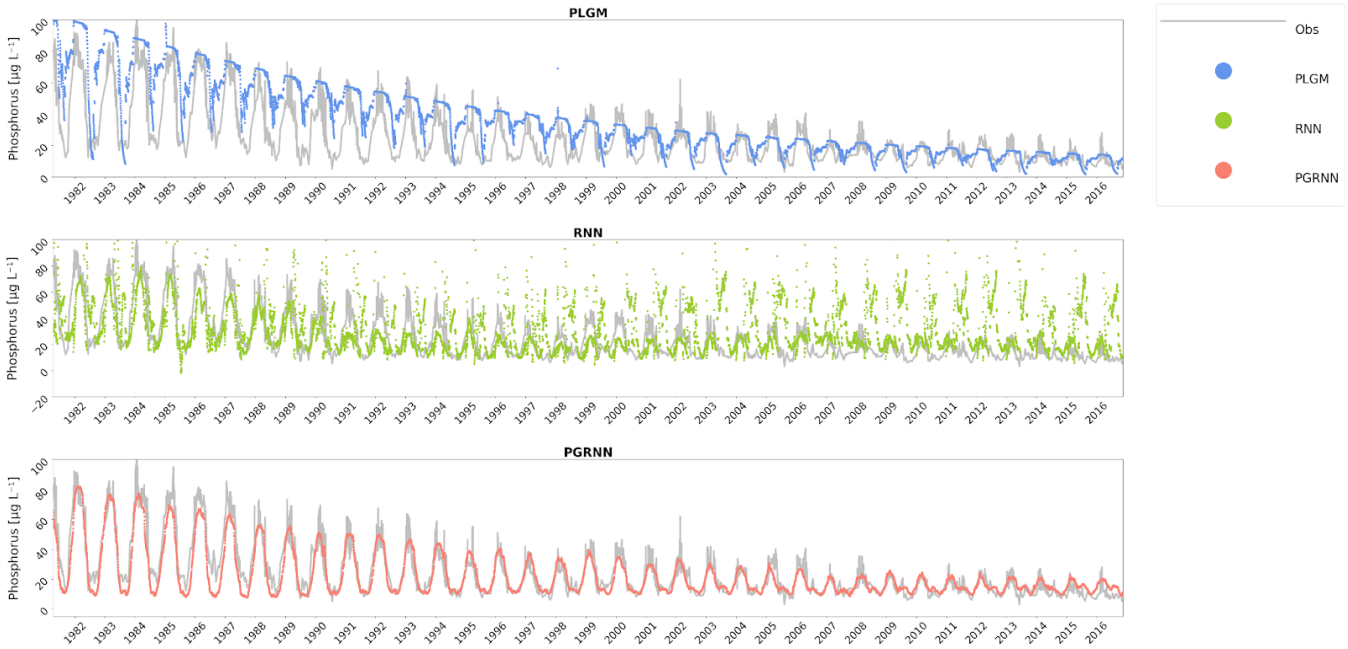

Figure 13. Epilimnetic phosphorus dynamic comparison between PLGM, RNN and PGRN models.

These results showed better performance of hybrid models versus purely machine learning or mechanistic model approaches since they leverage the information of both model approaches by complementing their strengths and diminishing their weaknesses (Zhu et al., 2019). On one hand, PGRNN could capture the underlying principles and physical laws governing the phosphorus dynamic, which is captured through the predicted epilimnetic and hypolimnetic phosphorus from PLGM (Yang & Shami., 2020; Alsharef et al., 2022). Since PGRNN phosphorus is dynamic and predicted concentration in comparison to RNN and PLGM, it shows its better ability to capture the complexities in such systems. In addition, the artificial intelligence component from RNN shows its advantage in accurately predicting P values with less uncertainty and lower underestimation, which is probably the result of the overcoming of PLGM bias.

3.3 Sensitivity analysis

From PLGM, RNN, and PGRNN model architecture and predictions, a sensitivity analysis was performed to extract the most important variables, in which the total loaded phosphorus (LoadP) from the Rhône River had a high sensitivity index value (S_{ix}) for PLGM and RNN, and hence appears as one of the most important variables in Figure 14. Even though non-point P sources and

other fluvial P contributions (Dranse, Aubonne, Venoge, or other small rivers) may be still relevant, Rhône River is the main water inflow (80%) and probably the main primary phosphorus source for Lake Geneva (Perga et al., 2016; Soomets et al., 2019). Despite that the contrary is found in PGRNN, the incorporated epilimnetic phosphorus from PLGM ($EpiP_{PLGM}$) is mainly derived from the differential sum of LoadP among other variables, and hence even with LoadP set to zero, its P contribution is a portion of $EpiP_{PLGM}$.

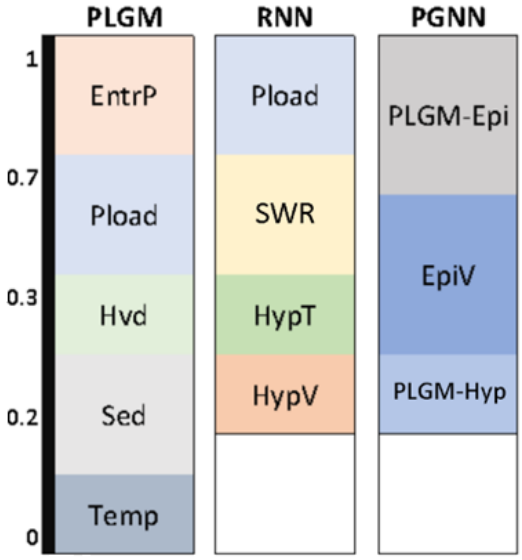

Figure 14. PLGM, RNN and PGRNN sensitivity analysis

Similarly to Rhône River P input, epilimnetic volume (EpiV) is another major component that undoubtedly has a relevant S_{ix} for PGRNN (0.7 and 0.5, respectively). The latter can be explained by the use of EpiV to proportionally compute the epilimnetic phosphorus entrainment (EntrP) in PLGM, in which it has the highest importance (0.6). In this way, the volume relevance reflects the phosphorus fluxes into the epilimnion, carried by the moving water volume which is produced by the deepening of the thermocline (Gibbs., 1977; Krishna et al., 2021). Since $EpiP_{PLGM}$ differential sum also considers EntrP, one may infer that when EpiV (as a surrogate of the phosphorus entrainment) is set to zero its associated P contribution should still be accounted by $EpiP_{PLGM}$. Nevertheless, the variations in EpiV determine the magnitude of EntrP which produces the main proportion of the internal phosphorus flux in PLGM, and hence be as important as $EpiP_{PLGM}$. The high EpiV and $EpiP_{PLGM}$ relevance may also reflect that their values are highly weighted within

hidden neurons and among the hidden layers in the PGRNN architecture in order to predict the epiliminetic phosphorus. Therefore, even with $EpiP_{PLGM}$ set to zero, PGRNN would predict EpiP in great proportion from EpiV values.

As mentioned before, EntrP resulted to be the principal internal phosphorus flux in Lake Geneva, which is associated to happen principally during the deepened thermocline events in winter, and to a less extent in spring and fall (Table 9). On the contrary, during summer a remarkably steady and shallow thermocline prevents the EntrP, which can be linked to the short-wave radiation (SWR) and HypT as relevant variables for RNN. In this line, SWR intensity is one factor that can determine the temperature of the air and as a consequence the temperature of the water column (Abed-Zaid., et al., 2021), as the air temperature increment the water temperature and density gradient between the hypolimnion and epilimnion, and with the strength and duration of the stratification while its depth is decreased suppressing the nutrient supply from deeper layers (Goldman et al 1989; Behrenfeld et al 2006; Woolway & Merchant., 2019).

Table 9. Overall seasonality of phosphorus flux showed by the entrainment (EntrP), vertical diffusion (Hvd), and external phosphorus load from Rhône River (LoadP)

		Phosphorus [µg d^{-1}]		
		EntrP	Hvd	LoadP
Season	Month			
autumn	November	0.4	1.8e-04	1.6e-05
	October	0.7	4.5e-05	2.2e-05
	September	0.8	2.0e-04	4.1e-05
spring	April	-20.4	9.9e-03	2.1e-05
	March	-6.8	1.1e-03	6.9e-06
	May	-14.6	1.4e-03	4.2e-05
summer	August	0.5	1.0e-04	8.0e-05
	July	-0.6	4.2e-03	9.4e-05
	June	-6.5	2.4e-04	7.4e-05
winter	December	-0.8	7.7e-03	9.5e-06
	February	-0.6	4.8e-03	4.1e-06
	January	-6.0	8.0e-02	6.2e-06

Vertical loading of phosphorus into the epilimnion is often considered just as EntrP since it has been historically considered as the most relevant -if not the unique- factor controlling the internal phosphorus loading (MacKay., 2011). Despite the latter, the vertical diffusion flux (Hvd) of phosphorus is placed in third place of importance for PLGM epilimnetic phosphorus prediction. In contrast to the entrainment, Hvd mainly depends on the concentration gradient between the hypolimnion and epilimnion, water stability, and water friction produced by wind. When intermediate to low temperatures are combined with moderate to high wind stress, vertical diffusion is expected to provide the energy to diffuse phosphorus through the thermocline, and hence bring phosphorus to shallows water (MacKay., 2011; Lohmann et al., 2013).

Despite Hvd is not comparable to EntrP in terms of phosphorus flux, it may be still relevant as an internal vertical P loading in Lake Geneva, especially during spring where PLGM shows that Hvd drives on average more phosphorus into the epilimnion than EntrP. A potential key role of Hvd when EntrP weakens is totally plausible when strong wind events occur, such as the Vent (wind from the southwest) and the Bise (wind from the northeast), which can occur during spring

(Soulignac et al., 2018). Vent and Bise indeed are strong perturbances that act over the entire surface of Lake Geneva, being a principal agent in water circulation, circular inertial motions, and energy transfer from the atmosphere (Cimatoribus et al., 2019; Lemmin el al., 2020; Soulignac et al., 2021). Indeed, the vertical eddy diffusivity constant for Lake Geneva (C = 700) is comparable to values found for deep ocean layers that are associated with relevant turbulent mixing mechanisms (Lemmin., 2020). Moreover, in lakes with large surface areas, Hvd is expected to be stronger and relatively more important when the stratification is strong enough to block vertical water movement while it can be weak enough to potentially perturb the Lake Geneva water column through strong wind stress (MacKay et al.,2011; Merino-Ibarra et al., 2021). In this way, PLGM shows Hvd of phosphorus, as an internal load that may occur in Lake Geneva, and to the best of our knowledge, this work is the first to show vertical diffusion as a potentially relevant internal phosphorus flux for Lake Geneva during the suppression of EntrP.

4. Conclusion

By considering a simple mechanistic approach that incorporates the principle variables and processes controlling the Lake Geneva phosphorus, it is possible to accurately predict and simulate its P concentration. Nevertheless, the confidence in PLGM predictions is highly dependent on the compartment that is being analyzed (whole water column, epilimnion or hypolimnion), the seasonality of the estimation, and the relative distribution of the values within the predicted phosphorus.

PLGM showed an overall high performance and large incorporation of the observed variability for the epilimnetic phosphorus, which, however, presented not depreciable overall P overestimations and indications of outlier sensitivity. In addition, relevant levels of uncertainty and even further estimation discrepancies are seasonally expressed, presenting the lowest prediction confidence at mid-predictions and highest overestimation, mainly during summer and autumn, while the best P representations are principally in winter, followed by spring which however is likely prompt to uncertainties at mid-range predicted values. If the predictions fall beyond the interquartile range the confidence considerably improves but this may also indicate the effect of outliers and possibly the overfitting of noisy data.

Despite high accuracy and moderate explanatory power for P in the hypolimnion and whole water column, PLGM predictions are as uncertain and deviating from the observations, as the epilimnetic PLGM phosphorus. Nevertheless, during summer and autumn, PLGM shows more confidence in its middle predictions compared to the predicted epilimnetic phosphorus. The predictions that fall outside IQR however are homogeneously more uncertain through all seasons, which is also coupled with underestimation for both HypP and WcP

The mixed results for PLGM arise some concerns regarding the overall PLGM reliability, which may be the result of not considering variables (Phytoplankton and horizontal dynamic) and not fully addressing the process (Thermocline depth estimation). These challenges are not new in mechanistic models, and hence they have been extensively addressed by alternatives and complementary approaches, such as the pure machine learning models (RNN) or hybrid approaches (PGRNN). In this work, both alternatives followed the already described better performance of

hybrid approaches compared to the individual application of the PLGM or RNN. However, further analysis has to be done in order to have better confidence in the predictions made in this model.

Through sensitivity analysis, it was possible to detect the main variables for PLGM, PGRN, and RNN. From this Rhône River is the main phosphorus input for the epilimnion, hence reaffirming its importance for Lake Geneva. In addition, the entrainment (also reflected by the change in volumes) is also identified as the main variable in P prediction since it represents the main intel phosphorus mechanism as expected for Lake Geneva. Finally, this analysis showed the vertical diffusion flux of phosphorus (Hvd) as another relevant variable for PLGM, indicating that it plays an important role in Lake Geneva which may be new insights to consider in Lake Geneva P dynamic, especially when EntrP is suppressed by a strong thermocline which may be weak enough to be perturbed by wind and lead to Hvd phosphorus flux into the epilimnion.

In conclusion, despite the concerns and mixed results obtained from PLGM, the present work shows that through a simple mechanistic model it is still possible to address a complex task such as the modeling of the epilimnetic phosphorus in Lake Geneva. Furthermore, by comparing and combining this mechanistic approach with artificial intelligence, the performance can be greatly improved and most importantly give more understanding by analyzing the sensitivity of the variables. In this context, PLGM can successfully match with relative precision the general seasonal and dynamic of the phosphorus in Lake Geneva, which is also reflected and confirmed from the sensitivity analysis.

5. References

Abed-Zaid, D. S., & Al-Zubaidi, H. A. (2021, November). Surface Heat Budget estimation for Laurance Lake, US. In *IOP Conference Series: Earth and Environmental Science* (Vol. 877, No. 1, p. 012005). IOP Publishing.

Ahlgren, I., Frisk, T., & Kamp-Nielsen, L. (1988). Empirical and theoretical models of phosphorus loading, retention, and concentration vs. lake trophic state. In Phosphorus in Freshwater Ecosystems (pp. 285-303). Springer, Dordrecht.

Ahmad, H. R., Loeschcke, H. H., & Woidtke, H. H. (1978). Three compartments model for the bicarbonate exchange of the brain extracellular fluid with blood and cells. In The regulation of respiration during sleep and anesthesia (pp. 195-209). Springer, Boston, MA.

Aldenberg, T., Janse, J. H., & Kramer, P. R. G. (1995). Fitting the dynamic model PCLake to a multi-lake survey through Bayesian Statistics. Ecological Modelling, 78(1-2), 83-99.

Ahlgren, J., Reitzel, K., De Brabandere, H., Gogoll, A., & Rydin, E. (2011). Release of organic P forms from lake sediments. Water research, 45(2), 565-572.

Alsharef, A., Aggarwal, K., Kumar, M., & Mishra, A. (2022). Review of ML and AutoML Solutions to Forecast Time-Series Data. Archives of Computational Methods in Engineering, 1-15.

Anneville O., Chang C.-W., Dur G., Souissi S., Rimet F., Hsieh C.-h.(2019).The paradox of re-oligotrophication: the role of bottom–up versus top–down controls on the phytoplankton community

Anneville, O., & Leboulanger, C. (2001). long-term changes ln the vertical distribution of phytoplankton blomass and primary production ln lake genev a: a response to the oligotrophica tion.

Beale, E. M. L. (1962). Some uses of computers in operational research. Industrielle Organisation, 31, 51-52.

Bengtsson, L. (2012). Classification of lakes from origin processes. In Encyclopedia of Lakes and Reservoirs (pp. 166-168). Springer.

Beres, M., Scheidhauer, M., Marillier, F., Girardclos, S., Baster, I., Wildi, W., ... & Rachoud-Schneider, A. M. (2003). Bottom-current and wind-pattern changes as indicated by Late Glacial and Holocene sediments from western Lake Geneva (Switzerland). *Lake Systems from the Ice Age to Industrial Time*, 39-48.

Bol, R., Gruau, G., Mellander, P. E., Dupas, R., Bechmann, M., Skarbøvik, E., ... & Gascuel-Odoux, C. (2018). Challenges of reducing phosphorus based water eutrophication in the agricultural landscapes of Northwest Europe. *Frontiers in Marine Science*, *5*, 276.

Bresciani, M., Stroppiana, D., Odermatt, D., Morabito, G., & Giardino, C. (2011). Assessing remotely sensed chlorophyll-a for the implementation of the Water Framework Directive in European perialpine lakes. Science of the Total Environment, 409(17), 3083-3091.

Bryhn, A. C., & Håkanson, L. (2007). A comparison of predictive phosphorus load-concentration models for lakes. Ecosystems, 10(7), 1084-1099.

Cerco, C. F. (1989). Measured and modelled effects of temperature, dissolved oxygen and nutrient concentration on sediment-water nutrient exchange. Hydrobiologia, 174(3), 185-194.

Cimatoribus, A. A., Lemmin, U., & Barry, D. A. (2019). Tracking Lagrangian transport in Lake Geneva: A 3D numerical modeling investigation. Limnology and Oceanography, 64(3), 1252-1269.

Cho, H., Kim, Y., Lee, E., Choi, D., Lee, Y., & Rhee, W. (2020). Basic enhancement strategies when using Bayesian optimization for hyperparameter tuning of deep neural networks. IEEE Access, 8, 52588-52608

Chou, L. (1995). Physics and chemistry of lakes. A. Lerman, D. M. Imboden, & J. R. Gat (Eds.). Berlin: Springer-Verlag.

Chorus, I., Köhler, A., Beulker, C., Fastner, J., van de Weyer, K., Hegewald, T., & Hupfer, M. (2020). Decades needed for ecosystem components to respond to a sharp and drastic phosphorus load reduction. Hydrobiologia, 847(21), 4621-4651.

CIPEL. (2019). Rapports sur les études et recherches entreprises dans le bassin lémanique, Campagne 2018. Commission internationale pour la protection des eaux du Léman (CIPEL). Last accessed 30 May 2022

Cooke, G. D., Welch, E. B., Peterson, S., & Nichols, S. A. (2016). *Restoration and management of lakes and reservoirs*. CRC press.

Corcoran, M., Sherif, M. I., Smalley, C., Li, A., Rockne, K. J., Giesy, J. P., & Sturchio, N. C. (2018). Accumulation rates, focusing factors, and chronologies from depth profiles of 210Pb and 137Cs in sediments of the Laurentian Great Lakes. Journal of Great Lakes Research, 44(4), 693-704.

Correll, D. L. (1999). Phosphorus: a rate limiting nutrient in surface waters. Poultry science, 78(5), 674-682.

Commission Internationale pour la protection des eaux du Léman, CIPEL. (2022). Le brassage complet des eaux du Léman n'a pas eu lieu pour la 10ème année consécutive : quelles conséquences sur l'état de santé du lac ? Communiqué de presse, March.

Cossu, R., & Wells, M. G. (2013). The interaction of large amplitude internal seiches with a shallow sloping lakebed: observations of benthic turbulence in Lake Simcoe, Ontario, Canada. PloS one, 8(3), e57444

Dearing, J. A., & Jones, R. T. (2003). Coupling temporal and spatial dimensions of global sediment flux through lake and marine sediment records. Global and Planetary Change, 39(1-2), 147-168.

Dimberg, P. H., & Bryhn, A. C. (2015). Predicting total nitrogen, total phosphorus, total organic carbon, dissolved oxygen and iron in deep waters of Swedish lakes. Environmental Modeling & Assessment, 20(5), 411-423.

Domhan, T., Springenberg, J. T., & Hutter, F. (2015). Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves. In Twenty-fourth international joint conference on artificial intelligence.

Dresti, C., Fenocchi, A., & Copetti, D. (2021). Modelling physical and ecological processes in medium-to-large deep European perialpine lakes: a review. JOURNAL OF LIMNOLOGY, 80(3).

Elshawi, R., Maher, M., & Sakr, S. (2019). Automated machine learning: State-of-the-art and open challenges. arXiv preprint arXiv:1906.02287.

Evans, R. D. (1994). Empirical evidence of the importance of sediment resuspension in lakes. Hydrobiologia, 284(1), 5-12.

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 interbasin water transfers. Water Resources Research, 49(6), 3626-3641.

Gaillard, R., Perroud, M., Goyette, S., & Kasparian, J. (2022). Multi-column modelling of Lake Geneva for climate applications. Scientific reports, 12(1), 1-12.

Gaudard, A., Schwefel, R., Vinnå, L. R., Schmid, M., Wüest, A., & Bouffard, D. (2017). Optimizing the parameterization of deep mixing and internal seiches in one-dimensional hydrodynamic models: a case study with Simstrat v1. 3. *Geoscientific Model Development*, *10*(9), 3411-3423.

Guinaldo, T., Munier, S., Le Moigne, P., Boone, A., Decharme, B., Choulga, M., & Leroux, D. J. (2021). Parametrization of a lake water dynamics model MLake in the ISBA-CTRIP land surface system (SURFEX v8. 1). Geoscientific Model Development, 14(3), 1309-1344

Gibbs, R. J. (1977). Transport processes in lakes and oceans. In Transport Processes in Lakes and Oceans (pp. 1-7). Springer, Boston, MA

Guo, H., Pasunuru, R., & Bansal, M. (2019). Autosem: Automatic task selection and mixing in multi-task learning. arXiv preprint arXiv:1904.04153.

Graham, B. M., & Adler, A. (2006). Objective selection of hyperparameter for EIT. Physiological measurement, 27(5), S65..

Håkanson, L. (2005). The importance of lake morphometry and catchment characteristics in limnology–ranking based on statistical analyses. Hydrobiologia, 541(1), 117-137.

Håkanson, L., & Bryhn, A. C. (2008). A dynamic mass-balance model for phosphorus in lakes with a focus on criteria for applicability and boundary conditions. Water, air, and soil pollution, 187(1), 119-147.

Hamilton, D. P., Collier, K. J., Quinn, J. M., & Howard-Williams, C. (2018). Lake Restoration Handbook. Springer International Publishing, Cham, 10, 978-3.

Hanson, P. C., Stillman, A. B., Jia, X., Karpatne, A., Dugan, H. A., Carey, C. C., & Kumar, V. (2020). Predicting lake surface water phosphorus dynamics using process-guided machine learning. Ecological Modelling, 430, 109136.

Heggy, E., Bermudez, V., & Vermeersch, M. (Eds.). (2022). Sustainable Energy-Water-Environment Nexus in Deserts: Proceeding of the First International Conference on Sustainable Energy-Water-Environment Nexus in Desert Climates. Springer Nature.

Hirayama, A. 2003. Development of decision-making tools for eutrophic lakes. UNESCO–EOLSS. UNESCO.

Hipsey, M. R., Boon, C., Paraska, D., Bruce, L. C., & Huang, P. (2020). Modelling aquatic eco-dynamics: Overview of the AED modular simulation platform.

Huang, Z., Yang, F., Xu, F., Song, X., & Tsui, K. L. (2019). Convolutional gated recurrent unit–recurrent neural network for state-of-charge estimation of lithium-ion batteries. *Ieee Access*, *7*, 93139-93149.

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software, 21(5), 602-614.-

Janse, J. H., Aldenberg, T., & Kramer, P. R. G. (2012). a mathematical model of the phosphorus cycle in lake loosdrecht and. in restoration and recovery of shallow eutrophic lake ecosystems in the netherlands: proceedings of a conference held in amsterdam, the netherlands, 18–19 april 1991 (vol. 74, pp. 119-136). springer science & business media.

Jarvie, H. P., Johnson, L. T., Sharpley, A. N., Smith, D. R., Baker, D. B., Bruulsema, T. W., & Confesor, R. (2017). Increased soluble phosphorus loads to Lake Erie: Unintended consequences of conservation practices?. *Journal of Environmental Quality*, *46*(1), 123-132.

Jensen, J. P., Pedersen, A. R., Jeppesen, E., & Søndergaard, M. (2006). An empirical model describing the seasonal dynamics of phosphorus in 16 shallow eutrophic lakes after external loading reduction. Limnology and Oceanography, 51(1part2), 791-800.

Jia, X., Willard, J., Karpatne, A., Read, J., Zwart, J., Steinbach, M., & Kumar, V. (2019). Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles. In *Proceedings of the 2019 SIAM International Conference on Data Mining* (pp. 558-566). Society for Industrial and Applied Mathematics.

Kane, D. D., Conroy, J. D., Richards, R. P., Baker, D. B., & Culver, D. A. (2014). Re-eutrophication of Lake Erie: Correlations between tributary nutrient loads and phytoplankton biomass. *Journal of Great Lakes Research*, *40*(3), 496-501.

Katsev, S., Crowe, S. A., Mucci, A., Sundby, B., Nomosatryo, S., Douglas Haffner, G., & Fowle, D. A. (2010). Mixing and its effects on biogeochemistry in the persistently stratified, deep, tropical Lake Matano, Indonesia. Limnology and Oceanography, 55(2), 763-776.cs

Kamarainen, A. M., Yuan, H., Wu, C. H., & Carpenter, S. R. (2009). Estimates of phosphorus entrainment in Lake Mendota: a comparison of one-dimensional and three-dimensional approaches. Limnology and Oceanography: Methods, 7(7), 553-567.

Khaither, P. A., & Erechtchoukova, M. G. (2007). From complex to simple in environmental simulation modelling. In MODSIM 2007–International Congress on Modelling and Simulation (pp. 2069-2075).

Khorasani, H., & Zhu, Z. (2021). Phosphorus retention in lakes: A critical reassessment of hypotheses and static models. Journal of Hydrology, 603, 126886.

Klein, A. (2015). Les apports par les affluents au Leman et au Rhone a l'aval du Geneve et leur qualite. Rapp. Comm. int. prot. eaux Leman contre pollut., Campagne, 2016, 108-114.

Kremer, K., Hilbe, M., Simpson, G., Decrouy, L., Wildi, W., & Girardclos, S. (2015). Reconstructing 4000 years of mass movement and tsunami history in a deep peri-Alpine lake (Lake Geneva, France-Switzerland). Sedimentology., 62, 1305–1327.

Krishna, S., Ulloa, H. N., Kerimoglu, O., Minaudo, C., Anneville, O., & Wüest, A. (2021). Model-based data analysis of the effect of winter mixing on primary production in a lake under reoligotrophication. Ecological Modelling, 440, 109401.

Lang, C. (2016). Phosphorus decreases in Lake Geneva but climate warming hampers the recovery of pristine oligochaete communities whereas chironomids are less affected. Journal of Limnology, 75(2).

Lemmin, U. (2020). Insights into the dynamics of the deep hypolimnion of Lake Geneva as revealed by long-term temperature, oxygen, and current measurements. Limnology and Oceanography, 65(9), 2092-2107.

Lemmin, U., Jiang, R., & Thorpe, S. A. (1998). Finescale dynamics of stratified waters near a sloping boundary of a lake. Coastal and Estuarine Studies, 461-474.

Lemmin, U., Mortimer, C. H., & Bäuerle, E. (2005). Internal seiche dynamics in Lake Geneva. Limnology and Oceanography, 50(1), 207-216.

Lehmann, M. K., Schütt, E. M., Hieronymi, M., Dare, J., & Krasemann, H. (2021). Analysis of recurring patchiness in satellite-derived chlorophyll a to aid the selection of representative sites for lake water quality monitoring. International Journal of Applied Earth Observation and Geoinformation, 104, 102547.

Li, Y., Liu, Y., Zhao, L., Hastings, A., & Guo, H. (2015). Exploring change of internal nutrients cycling in a shallow lake: a dynamic nutrient driven phytoplankton model. Ecological Modelling, 313, 137-148.

Liu, Y., Bi, J., Lv, J., Ma, Z., & Wang, C. (2017). Spatial multi-scale relationships of ecosystem services: A case study using a geostatistical methodology. Scientific reports, 7(1), 1-12.

Liu, W., Zhang, Q., & Liu, G. (2011). Effects of watershed land use and lake morphometry on the trophic state of Chinese lakes: implications for eutrophication control. CLEAN–Soil, Air, Water, 39(1), 35-42.

Lods-crozet, B., & Reymond, O. (2004). réponses des communautés benthiques du léman à l'amélioration de l'état trophique du léman entre 1983 et 2003 responses of the benthic communities in lake geneva to the improved trophic state of the lake. sur les études et recherches entreprises dans le bassin lémanique.

Loizeau, J. L., Dominik, J., Luzzi, T., & Vernet, J. P. (1997). Sediment core correlation and mapping of sediment accumulation rates in Lake Geneva (Switzerland, France) using volume magnetic susceptibility. Journal of Great Lakes Research, 23(4), 391-402.

MacKay, E. B. (2011). Heterogeneity in Esthwaite Water, a Small, Temperate Lake: Consequences for Phosphorus Budgets. Lancaster University (United Kingdom).

Malmaeus, J. M., & Håkanson, L. (2004). Development of a lake eutrophication model. Ecological Modelling, 171(1-2), 35-63.

Martin, J. L., & McCutcheon, S. C. (1998). Hydrodynamics and transport for water quality modeling. CRC press.

Meraj, G., Singh, S. K., Kanga, S., & Islam, M. N. (2021). Modeling on comparison of ecosystem services concepts, tools, methods and their ecological-economic implications: A review. Modeling Earth Systems and Environment, 1-20.

Merino-Ibarra, M., Ramírez-Zierold, J. A., Valdespino-Castillo, P. M., Castillo-Sandoval, F. S., Guzmán-Arias, A. P., Barjau-Aguilar, M., ... & Quintanilla-Terminel, J. G. (2021). Vertical Boundary Mixing Events during Stratification Govern Heat and Nutrient Dynamics in Windy Tropical Lakes with High Water-Level Fluctuations: A Long-Term (2001–2018)

Moisset, S. A., Cabanes, D. J., Blanco-Ameijeiras, S., & Hassler, C. S. (2019). Response of phytoplankton from the metalimnetic chlorophyll maximum to macro-and micro-nutrients amendments in Lake Geneva. Journal of Great Lakes Research, 45(2), 290-299

Müller, J., Park, J., Sahu, R., Varadharajan, C., Arora, B., Faybishenko, B., & Agarwal, D. (2021). Surrogate optimization of deep neural networks for groundwater predictions. Journal of Global Optimization, 81(1), 203-231.

Müller, B., Gächter, R., & Wüest, A. (2014). Accelerated water quality improvement during oligotrophication in peri-alpine lakes. Environmental science & technology, 48(12), 6671-6677.

Neumann, T., Fennel, W., & Kremp, C. (2002). Experimental simulations with an ecosystem model of the Baltic Sea: a nutrient load reduction experiment. Global biogeochemical cycles, 16(3), 7-1.

Nürnberg, G. K. (2007). Lake responses to long-term hypolimnetic withdrawal treatments. *Lake and Reservoir Management*, *23*(4), 388-409.

Och, L. M., Müller, B., Voegelin, A., Ulrich, A., Göttlicher, J., Steiniger, R., & Sturm, M. (2012). New insights into the formation and burial of Fe/Mn accumulations in Lake Baikal sediments. Chemical Geology, 330, 244-259

Perga, M. E., Maberly, S. C., Jenny, J. P., Alric, B., Pignol, C., & Naffrechoux, E. (2016). A century of human-driven changes in the carbon dioxide concentration of lakes. *Global Biogeochemical Cycles*, *30*(2), 93-104.

Perolo, P., Fernández Castro, B., Escoffier, N., Lambert, T., Bouffard, D., & Perga, M. E. (2021). Accounting for surface waves improves gas flux estimation at high wind speed in a large lake. Earth System Dynamics, 12(4), 1169-1189.

Perroud, M., Goyette, S., Martynov, A., Beniston, M., and Anneville, O. (2009). Simulation of multiannual thermal profiles in deep Lake Geneva: A comparison of onedimensional lake models, Limnol. Oceanogr., 54, 1574–1594.

Qin, B., Zhou, J., Elser, J. J., Gardner, W. S., Deng, J., & Brookes, J. D. (2020). Water depth underpins the relative roles and fates of nitrogen and phosphorus in lakes. Environmental science & technology, 54(6).

Rapin, F., & Gerdeaux, D. (2013). La protection du Léman. Archives des sciences, 66, 103-116.

Rapin, F., & Klein, A. (2013). Les apports par les affluents au léman et au rhône à l'aval de genève assessment of the input from the tributaries into the lake geneva and into the rhône downstream of geneva.

Read, J. S., Hamilton, D. P., Jones, I. D., Muraoka, K., Winslow, L. A., Kroiss, R., ... & Gaiser, E. (2011). Derivation of lake mixing and stratification indices from high-resolution lake buoy data. Environmental Modelling & Software, 26(11), 1325-1336.

Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., & Kumar, V. (2019). Process-guided deep learning predictions of lake water temperature. Water Resources Research, 55(11), 9173-9190.

Reardon, K. E., Moreno-Casas, P. A., Bombardelli, F. A., & Schladow, S. G. (2016). Seasonal nearshore sediment resuspension and water clarity at Lake Tahoe. Lake and Reservoir Management, 32(2), 132-145.

Reiss, R. S., Lemmin, U., & Barry, D. A. (2022). Wind-Induced Hypolimnetic Upwelling Between the Multi-Depth Basins of Lake Geneva During Winter: An Overlooked Deepwater Renewal Mechanism?. Journal of Geophysical Research: Oceans, 127(6), e2021JC018023.

Ren, T., Liu, X., Niu, J., Lei, X., & Zhang, Z. (2020). Real-time water level prediction of cascaded channels based on multilayer perception and recurrent neural network. *Journal of Hydrology*, *585*, 124783.

Rimet, F., Anneville, O., Barbet, D., Chardon, C., Crépin, L., Domaizon, I., ... & Monet, G. (2020). The Observatory on LAkes (OLA) database: Sixty years of environmental data accessible to the public. J. Limnol, 79(2), 164-178.

Robarts, R. D., Waiser, M. J., Arts, M. T., & Evans, M. S. (2005). Seasonal and diel changes of dissolved oxygen in a hypertrophic prairie lake. Lakes & Reservoirs: Research & Management, 10(3), 167-177.

Robertson, D. M. (2002). Hydrology and water quality of Geneva Lake, Walworth County, Wisconsin (No. 2). US Department of the Interior, US Geological Survey.

Rocha, M. D. J. D., & Neto, I. E. L. (2022). Internal phosphorus loading and its driving factors in the dry period of Brazilian semiarid reservoirs. *Journal of Environmental Management*, *312*, 114983.

Rosen, M. R. (2015). The influence of hydrology on lacustrine sediment contaminant records. In Environmental Contaminants (pp. 5-33). Springer, Dordrecht.

Rose, N. L., Morley, D., Appleby, P. G., Battarbee, R. W., Alliksaar, T., Guilizzoni, P., ... & Punning, J. M. (2011). Sediment accumulation rates in European lakes since AD 1850: trends, reference conditions and exceedence. Journal of Paleolimnology, 45(4), 447-468.

Rose, K. C., Winslow, L. A., Read, J. S., Read, E. K., Solomon, C. T., Adrian, R., & Hanson, P. C. (2014). Improving the precision of lake ecosystem metabolism estimates by identifying predictors of model uncertainty. Limnology and Oceanography: Methods, 12(5), 303-312

Roubeix, V., Minaudo, C., Prats, J., Reynaud, N., Zhang, Q., Moatar, F., & Danis, P. A. (2020). Adapting the dynamic LakeMab model to simulate seasonal variations of phosphorus concentration in reservoirs: a case study of Lake Bultière (France). Limnology, 21(2), 233-244.

Schiefer, E. R. I. K., Reid, K. E. V. I. N., BURT, A., & LUCE, J. (2001). Assessing Natural Sedimentation Patterns and Impacts of Land Use on Sediment Yield: A Lake-sediment–based Approach. In Watershed assessment in the Southern Interior of British Columbia: Workshop proceedings. BC Ministry of Forests, Research Branch, Victoria, BC Working Paper (No. 57).

Schuwirth, N., Borgwardt, F., Domisch, S., Friedrichs, M., Kattwinkel, M., Kneis, D., ... & Vermeiren, P. (2019). How to make ecological models useful for environmental management. Ecological Modelling, 411, 108784.

Schwefel, R., Gaudard, A., Wüest, A., & Bouffard, D. (2016). Effects of climate change on deepwater oxygen and winter mixing in a deep lake (Lake Geneva): Comparing observational findings and modeling. Water Resources Research, 52(11), 8811-8826.

Schwefel, R., Müller, B., Boisgontier, H., & Wüest, A. (2019). Global warming affects nutrient upwelling in deep lakes. Aquatic Sciences, 81(3), 50.

Silvonen, S., Niemistö, J., Csibrán, A., Jilbert, T., Torma, P., Krámer, T., ... & Horppila, J. (2021). A biogeochemical approach to evaluate the optimization and effectiveness of hypolimnetic withdrawal. Science of the Total Environment, 755, 143202.

Simmons, J. A., Harley, M. D., Marshall, L. A., Turner, I. L., Splinter, K. D., & Cox, R. J. (2017). Calibrating and assessing uncertainty in coastal numerical models. Coastal Engineering, 125, 28-41.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. Advances in neural information processing systems, 25.

Soomets, T., Kutser, T., Wüest, A., & Bouffard, D. (2019). Spatial and temporal changes of primary production in a deep peri-alpine lake. *Inland Waters*, *9*(1), 49-60.

Soulignac, F., Danis, P. A., Bouffard, D., Chanudet, V., Dambrine, E., Guénand, Y., ... & Anneville, O. (2018). Using 3D modeling and remote sensing capabilities for a better

understanding of spatio-temporal heterogeneities of phytoplankton abundance in large lakes. *Journal of Great Lakes Research*, *44*(4), 756-764.

Siuda, W., Grabowska, K., Kaliński, T., Kiersztyn, B., & Chróst, R. J. (2020). Trophic state, eutrophication, and the threats for water quality of the great Mazurian Lake System. Polish River Basins and Lakes–Part I: Hydrology and Hydrochemistry, 231-260.

Soulignac, F., Lemmin, U., Ziabari, S. M. H., Wynn, H. K., Graf, B., & Barry, D. A. (2021). Rapid changes in river plume dynamics caused by advected wind-driven coastal upwelling as observed in Lake Geneva. Limnology and Oceanography, 66(8), 3116-3133

Steinsberger, T., Schmid, M., Wüest, A., Schwefel, R., Wehrli, B., & Müller, B. (2017). Organic carbon mass accumulation rate regulates the flux of reduced substances from the sediments of deep lakes. Biogeosciences, 14(13), 3275-3285.

Steinsberger, T., Schwefel, R., Wüest, A., & Müller, B. (2020). Hypolimnetic oxygen depletion rates in deep lakes: Effects of trophic state and organic matter accumulation. Limnology and Oceanography, 65(12), 3128-3138.

Stigebrandt, A., & Wulff, F. (1987). A model for the dynamics of nutrients and oxygen in the Baltic proper. Journal of Marine Research, 45(3), 729-759.

Steele, J. H. (1962). Environmental control of photosynthesis in the sea. Limnology and oceanography, 7(2), 137-150.

Southeastern Wisconsin Regional Planning Commission (SEWRPC). (2008). A Lake Management Plan for Geneva Lake, Walworth County, Wisconsin.

Tadonleke, R. (2004). Production et biomasse phytoplanctoniques dans le Léman. Rapp. Comm. int. prot. eaux Léman contre pollut., Campagne, 91-99.

Tadonleke, R. D., Lazzarotto, J., Anneville, O., & Druart, J. C. (2009). Phytoplankton productivity increased in Lake Geneva despite phosphorus loading reduction. Journal of Plankton Research, 31(10), 1179-1194.

Taranu, Z. E., Gregory-Eaves, I., Steele, R. J., Beaulieu, M., & Legendre, P. (2017). Predicting microcystin concentrations in lakes and reservoirs at a continental scale: A new framework for modelling an important health risk factor. Global Ecology and Biogeography, 26(6), 625-637.

Tasnim, B. (2020). Enhancement and Redevelopment of the Regional Lake Water Quality Model with Applications.

Taube, C. M. (2000). Three methods for computing the volume of a lake. Manual of Fisheries Survey Methods II: With Periodic Updates; Schneider, JC, Ed, 175-179.

Tu, L., Jarosch, K. A., Schneider, T., & Grosjean, M. (2019). Phosphorus fractions in sediments and their relevance for historical lake eutrophication in the Ponte Tresa basin (Lake Lugano, Switzerland) since 1959. *Science of the total environment*, *685*, 806-817.

Talukdar, S., Ahmed, S., Naikoo, M. W., Rahman, A., Mallik, S., Ningthoujam, S., ... & Ramana, G. V. (2023). Predicting lake water quality index with sensitivity-uncertainty analysis using deep learning algorithms. Journal of Cleaner Production, 406, 136885

Umlauf, L., & Lemmin, U. (2005). Interbasin exchange and mixing in the hypolimnion of a large lake: The role of long internal waves. *Limnology and Oceanography*, *50*(5), 1601-1611.

Van der Molen, D. T., Los, F. J., Van Ballegooijen, L., & Van der Vat, M. P. (1994). Mathematical modelling as a tool for management in eutrophication control of shallow lakes. In Nutrient Dynamics and Biological Structure in Shallow Freshwater and Brackish Lakes (pp. 479-492). Springer, Dordrecht.

Vernet, J. P., Dominik, J., & Favarger, P. Y. (1983). Texture and sedimentation rates in Lake Geneva. Environmental Geology, 5(3), 143-149.

Vinçon-Leite, B., Tassin, B., & Jaquet, J. M. (1995). Contribution of mathematical modeling to lake ecosystem understanding: Lake Bourget (Savoy, France). Hydrobiologia, 300, 433-442.

Woolway, R. I., & Merchant, C. J. (2019), Worldwide alteration of lake mixing regimes in response to climate change. Nat. Geosci. 12, 271–276

Wu, J., & Marceau, D. (2002). Modeling complex ecological systems: an introduction. Ecological Modelling, 153(1-2), 1-6.

Xu, H., Paerl, H. W., Zhu, G., Qin, B., Hall, N. S., & Zhu, M. (2017). Long-term nutrient trends and harmful cyanobacterial bloom potential in hypertrophic Lake Taihu, China. Hydrobiologia, 787(1), 229-242.

Yan, J., Liu, J., Yu, Y., & Xu, H. (2021). Water quality prediction in the luan river based on 1-drcnn and bigru hybrid neural network model. Water, 13(9), 1273.

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing, 415, 295-316.

Yang, X. E., Wu, X., Hao, H. L., & He, Z. L. (2008). Mechanisms and assessment of water eutrophication. Journal of zhejiang university Science B, 9(3), 197-209.

Yankova, Y., Neuenschwander, S., Koster, O., & Posch, T. (2017). Abrupt stop of deep water turnover with lake warming: Drastic consequences for algal primary producers. Scientific Reports, 7(1), 13770.

Yeates, P. S., & Imberger, J. (2003). Pseudo two-dimensional simulations of internal and boundary fluxes in stratified lakes and reservoirs. *International Journal of River Basin Management*, *1*(4), 297-319.

Yu, J. W., Kim, J. S., Li, X., Jong, Y. C., Kim, K. H., & Ryang, G. I. (2022). Water quality forecasting based on data decomposition, fuzzy clustering and deep learning neural network. Environmental Pollution, 303, 119136.

Zia, A., Schroth, A. W., Hecht, J. S., Isles, P., Clemins, P. J., Turnbull, S., & Pinder, G. (2022). Climate change-legacy phosphorus synergy hinders lake response to aggressive water policy targets. Earth's Future, e2021EF002234.

Zhu, W., Yeh, W., Chen, J., Chen, D., Li, A., & Lin, Y. (2019). Evolutionary convolutional neural networks using abc. In Proceedings of the 2019 11th International Conference on Machine Learning and Computing (pp. 156-162)