

Hybrid Deep Learning and Physics Models for Lake Water Quality Forecasting

David Métivier¹ Céline Casenave¹ Brigitte Vinçon-Leite²

Interns: Ali Elhishi & Mathieu Le Séac'h

¹UMR MISTEA, INRAE

²LEESU, École nationale des ponts et chaussées, Université Paris-Est Créteil



Objective

- “Simple” hybrid IA-Physics model for lake temperature and variables related to water quality (chlorophyll, oxygen, etc.)

Objective

- “Simple” hybrid IA-Physics model for lake temperature and variables related to water quality (chlorophyll, oxygen, etc.)
- + predicting chlorophyll concentration
- ++ anticipating algal blooms
- Oxygen depletion in the lake
- Toxicity for certain species, including humans



Algal bloom in Lake Créteil.

Data: Créteil lake

Variable	Meaning	Depth [m]	Units
$T^{(\text{air})}$	Air temperature	-	°C
T_w	Water temperature	0.5, 1.5, 2.5, 3.5, 4.5	°C
Chl	Chlorophyll concentration	1.50	$\mu\text{g L}^{-1}$
Cyano	Cyanobacteria concentration	1.50	$\mu\text{g L}^{-1}$
O2	Oxygen saturation	1.50	%

Summary of available variables.

Roughly 18 months of data,

Sampling frequency of 1 measurement / 15min.

Aggregation to hourly data to get rid of the noise and get a more manageable dataset for training.

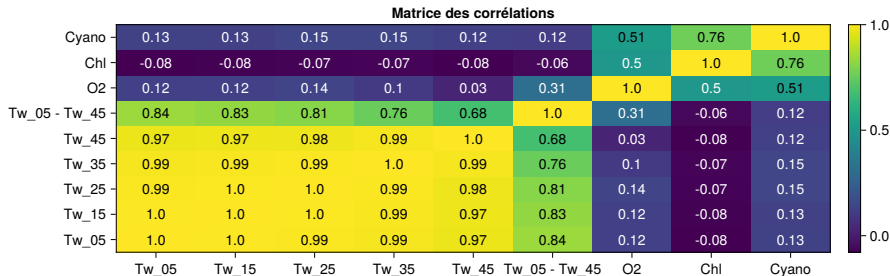
Factors promoting algal blooms

- Temperature increase

Factors promoting algal blooms

- Temperature increase
- Lake stratification → need to model surface and bottom temperature

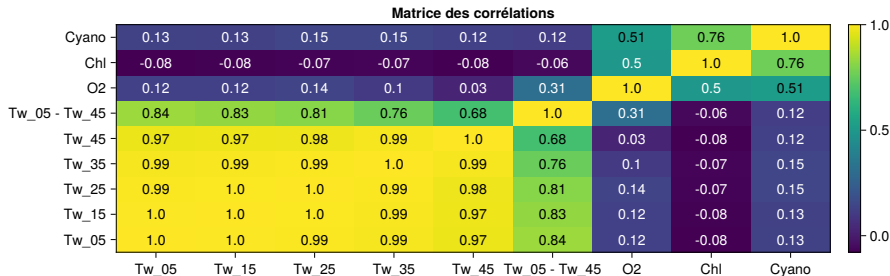
$$\Delta T_w = (T_w^{(0.5)} - T_w^{(4.5)})$$



Correlation matrix between the variables of Lake Créteil.

Factors promoting algal blooms

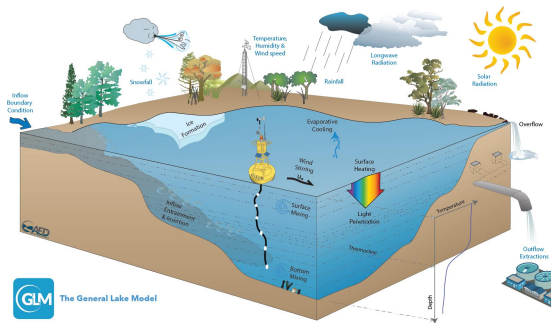
- Temperature increase
- Lake stratification → need to model surface and bottom temperature
$$\Delta T_w = (T_w^{(0.5)} - T_w^{(4.5)})$$
- Increase in nutrients (nitrogen and phosphorus) in the lake → not measured and not explicitly modeled for now



Correlation matrix between the variables of Lake Créteil.

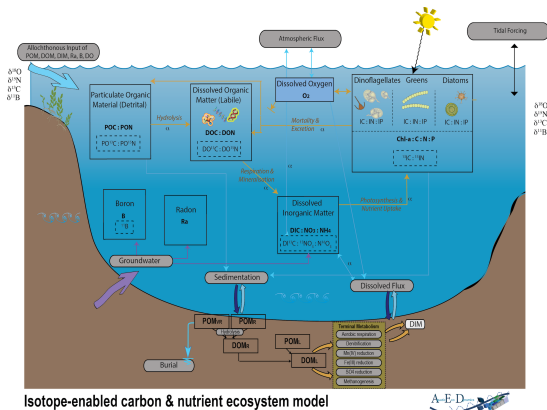
State of the art: Process-based

General Lake Model (GLM): 1D hydrodynamics model EDP based



State of the art: Process-based

General Lake Model (GLM): 1D hydrodynamics model EDP based
Aquatic Eco-Dynamics (AED): also EDP based



State of the art: Process-based

General Lake Model (GLM): 1D hydrodynamics model EDP based
Aquatic Eco-Dynamics (AED): also EDP based

Pros & Cons

- + Interpretable and physics based
- Hard to calibrate and require a lot of information e.g. bathymetry, nutriment etc
- Hard to couple

State of the art: Process-based

General Lake Model (GLM): 1D hydrodynamics model EDP based
Aquatic Eco-Dynamics (AED): also EDP based

Pros & Cons

- + Interpretable and physics based
- Hard to calibrate and require a lot of information e.g. bathymetry, nutriment etc
- Hard to couple

Reduced models e.g. Air2Temp for surface temperature

- + Easy to train on data
- ± Qualitative but not quantitative

State of the art: Data-driven

Different types of models (Random forest, MLP, RNN, etc.)

- + Learn with different source of training data
- + Can learn very complex dynamic
- + Beneficiate from AI boom
 - Times series with seasonal patterns, rare events, changing environment
 - Do not know physics
 - Sensitive to quantity and quality of data
 - Interpretability

State of the art: Data-driven

Different types of models (Random forest, MLP, RNN, etc.)

- + Learn with different source of training data
- + Can learn very complex dynamic
- + Beneficiate from AI boom
- Times series with seasonal patterns, rare events, changing environment
- Do not know physics
- Sensitive to quantity and quality of data
- Interpretability
- Not state of the art for lake temperature forecast



Journal of Hydrology

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

Research papers

Forecasting surface water temperature in lakes: A comparison of approaches

Senlin Zhu^{a,b,*}, Mariusz Ptak^c, Zaher Mundher Yaseen^d, Jiangyu Dai^{a,*}, Bellie Sivakumar^e

^a State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China

^b Department of Civil and Environmental Engineering, Cullen College of Engineering, University of Houston, Houston 77204, USA

State of the art: Hybrid models

Neural Networks + physics models

- + Best of both worlds
- + More robust and need less data
- + Satisfy better physics
- ? Can improve reduced models to make them quantitative
- ? Times series with seasonal patterns, rare events, changing environment
- ± Interpretability
 - **How?** → Under active research
 - Harder to train

Universal Differential Equations (Neural ODEs)

For a variable of interest $u(t)$ and an external forcing $g(t)$ we can write the following ODE system,

$$\frac{du}{dt}(t) = \underbrace{f_{\phi}^{(\text{PHY})}(u(t), g(t))}_{\text{known physics}} + \underbrace{f_{\theta}^{(\text{NN})}(u(t), g(t))}_{\text{Neural Network}} .$$

- + Already good with simple NN
- + Physics structure is there to guide the learning
- + Adapted for dynamical systems and forecasting
- Training
- No Universal approximation theorem

e.g. $\dot{u} = f_{\theta}^{(\text{NN})}(u)$ cannot produce a periodic $u(t)$

Universal Differential Equations (Neural ODEs)

For a variable of interest $u(t)$ and an external forcing $g(t)$ we can write the following ODE system,

$$\frac{du}{dt}(t) = \underbrace{f_{\phi}^{(\text{PHY})}(u(t), g(t))}_{\text{known physics}} + \underbrace{f_{\theta}^{(\text{NN})}(u(t), g(t))}_{\text{Neural Network}} .$$

- + Already good with simple NN
- + Physics structure is there to guide the learning
- + Adapted for dynamical systems and forecasting
 - Training
 - No Universal approximation theorem

PINNs vs UDEs

- PINNs learn the solution $u_{\theta}^{\text{NN}}(t)$ with a physical loss term
- + Dynamics ($\sim$ derivative) is often simpler than the solution

Universal Differential Equations (Neural ODEs)

For a variable of interest $u(t)$ and an external forcing $g(t)$ we can write the following ODE system,

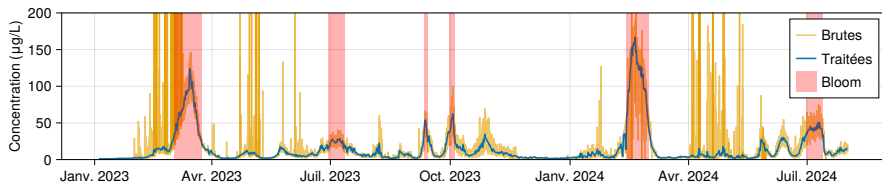
$$\frac{du}{dt}(t) = \underbrace{f_{\phi}^{(\text{PHY})}(u(t), g(t))}_{\text{known physics}} + \underbrace{f_{\theta}^{(\text{NN})}(u(t), g(t))}_{\text{Neural Network}}.$$

- + Already good with simple NN
- + Physics structure is there to guide the learning
- + Adapted for dynamical systems and forecasting
 - Training
 - No Universal approximation theorem

PINNs vs UDEs

- PINNs learn the solution $u_{\theta}^{\text{NN}}(t)$ with a physical loss term
- + Dynamics ($\sim$ derivative) is often simpler than the solution
- PINNs learn the solution only on a fixed interval $t \in [t_0, t_f] \rightarrow$ not good for forecasting **VS** UDEs learn the dynamics

Training



Cleaned chlorophyll.

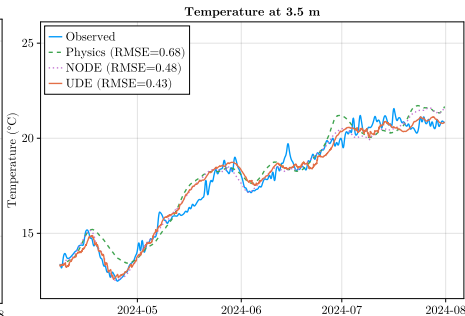
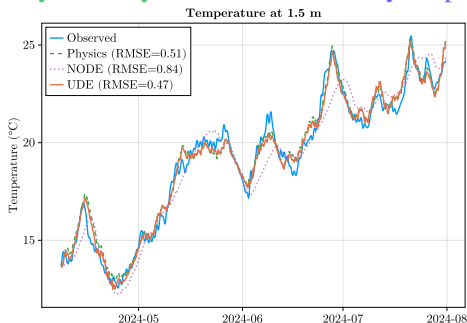
Data from January 2023 to August 2024, split into 80% for validation (≈ 15 months) and 20% for validation (≈ 4 months).

Temperature at multiple depth

Physics: **Diffusion model**

$$\frac{\partial T}{\partial t} = \underbrace{A_b \frac{\partial^2 T}{\partial z^2}}_{\text{diffusion}} - \underbrace{w \frac{\partial T}{\partial z}}_{\text{advection}} + \underbrace{\frac{1}{\rho c} \frac{\partial I}{\partial z}}_{\text{source radiative}}$$

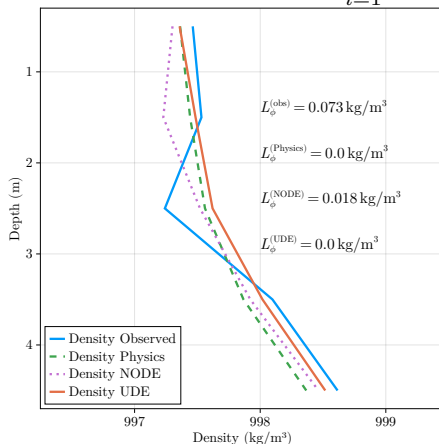
Physics only, **NODE: IA only no physics,** **UDE: physics + IA**



Physics consistency of the models

Density of water $\rho(T)$ should **not decrease with depth**.

$$\text{Physical consistency loss } \mathcal{L}_{\text{phys}}(T) = \sum_{i=1}^{N_{\text{Layers}}-1} \max(-\Delta\rho_i, 0)$$



Physics consistency of the models

Density of water $\rho(T)$ should **not decrease with depth**.

$$\text{Physical consistency loss } \mathcal{L}_{\text{phys}}(T) = \sum_{i=1}^{N_{\text{Layers}}-1} \max(-\Delta\rho_i, 0)$$

Model	Validation <i>RMSE [°C]</i>	Train <i>RMSE [°C]</i>	$L_{\phi}^{(\text{val})}$ [kg/m ³]	$L_{\phi}^{(\text{train})}$ [kg/m ³]
UDE	0.5248	1.0736	3.3e-04	9.2e-04
Physics	0.6238	1.1857	2.3e-03	3.1e-03
NODE	0.6636	1.1767	7.7e-03	1.5e-02

- **UDE** → better physics consistency (density increases with depth).
- More than good RMSE: physics-credible model.

Conclusions

- UDEs embed physics in the structure of the model and learn the missing physics with a Neural Network.
 - UDEs can learn lake temperature dynamics and are more physics-consistent than physics-only or data-only models.
 - Hard to train
- **(Ongoing)** Encouraging results for chlorophyll or oxygen concentration but more work needed to get both at the same time.

Conclusions

- UDEs embed physics in the structure of the model and learn the missing physics with a Neural Network.
- UDEs can learn lake temperature dynamics and are more physics-consistent than physics-only or data-only models.
- Hard to train
- **(Ongoing)** Encouraging results for chlorophyll or oxygen concentration but more work needed to get both at the same time.

Thank You!