Automatic Modeling of Dynamical Interactions Within Marine Ecosystems

Omar Iken[1], Maxime Folschette[1], Tony Ribeiro[2]

[1] Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
[2] Université de Nantes, Centrale Nantes, CNRS, LS2N, F-44000 Nantes, France

1. Introduction

Aim

Build a causal model from the observation of dynamical transitions of a marine ecosystem

Phytoplankton

Microscopic organisms that live in aquatic environments

- Basis of aquatic food
- Essential for the carbon cycle
- Produces a large part of the planet's oxygen

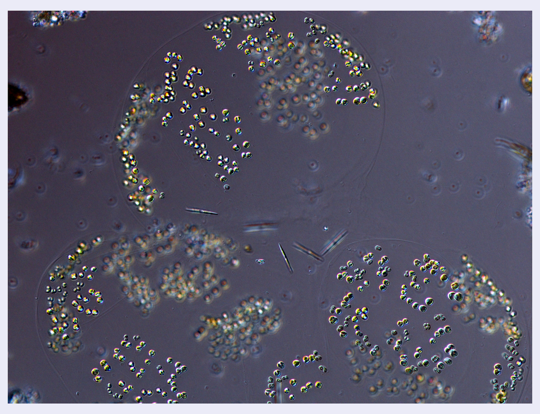
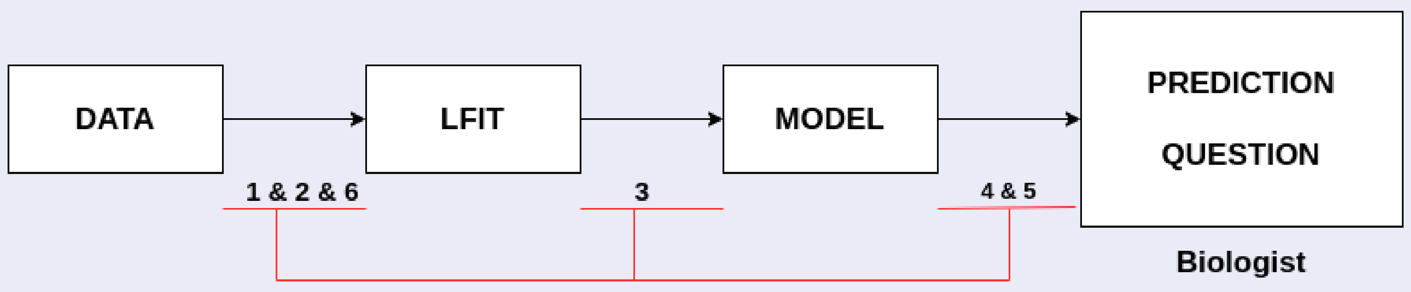


Figure: Phaeocystis (phytoplankton)

Figure: Effects of phytoplankton bloom

2. Methodology



- **1** Data cleaning
- **2** Discretize features
- **3** Use *PyLFIT* API
- **4** Improve rules using Pareto front
- **5** Build influence graph
- **6** Data augmentation

3. SRN Dataset

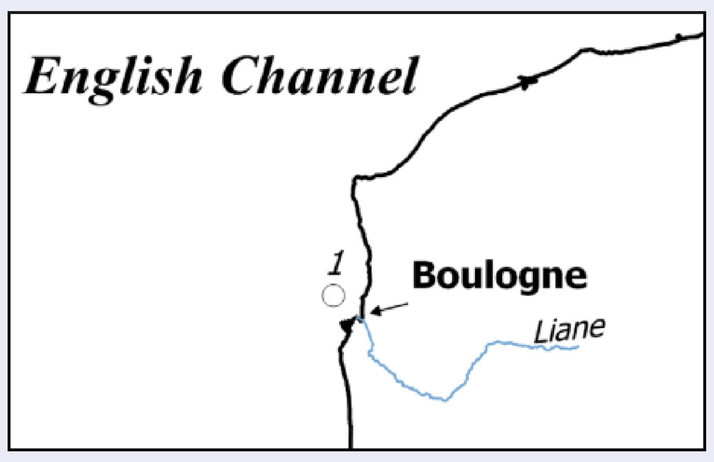


Figure: Chosen observation station

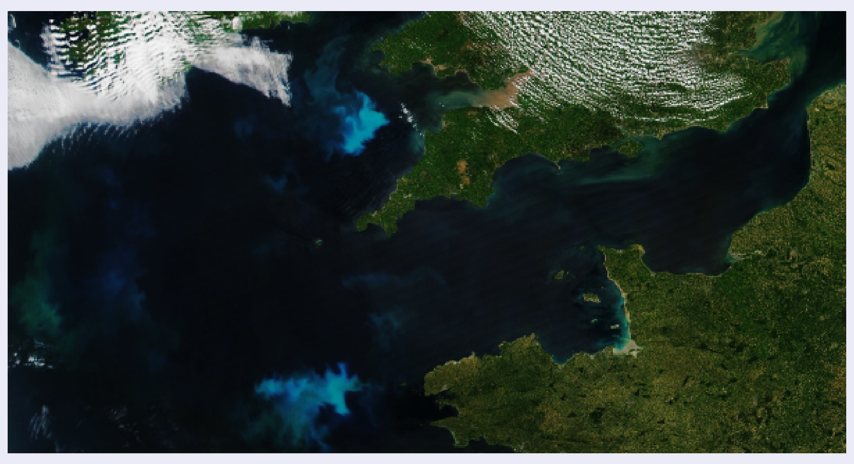


Figure: Blooms of phytoplankton

Sampling location	Sampling date	Taxon	Value	Type	Sampling depth
001-P-015	1992-05-18	CHLOROA	6.0	Hydro	Surface (0-1m)
006-P-001	2019-12-02	Chaetoceros	1000.0	Plankton	Surface (0-1m)
002-P-007	1994-05-25	Pleurosigma	100.0	Plankton	Surface (0-1m)

Table: Example of samples from the *SRN* dataset

4. Data Preprocessing

Discretization

Hydrological factors

- Level 0 if $X(t) \leq \text{avg}(X)$
- Level 1 if $X(t) > \text{avg}(X)$

Phytoplankton species

- Level 0 if $X(t) = 0$
- Level 1 if $X(t) \leq \text{avg}(X)$
- Level 2 if $X(t) > \text{avg}(X)$

Machine Learning Settings

- 253 training transitions
- 53 testing transitions
- features: Hydro & Plankton
- targets: Plankton

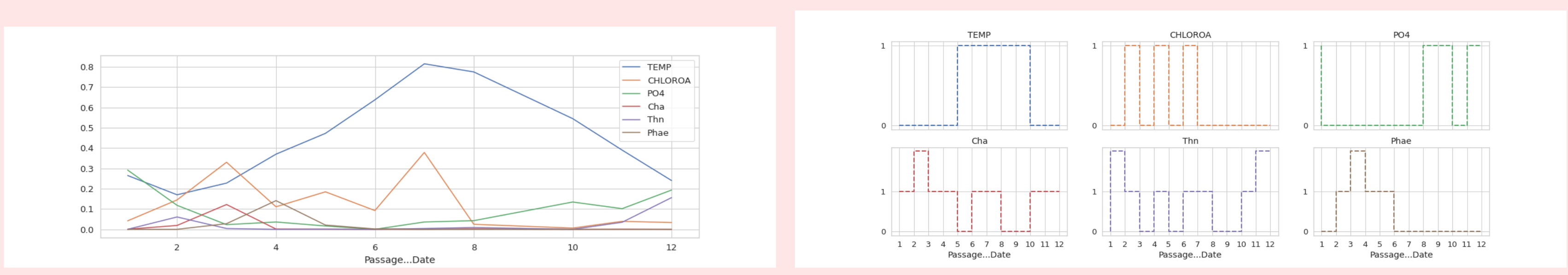


Figure: Variation of some features over 2012 and their discretization

5. Learning From Interpretation Transition

LFIT infers a *logic program* that models the dynamics of a system as a set of logic rules that explain the observed transitions.

Transition: **(feature state)** $\rightarrow$ **(target state)**

Rule: $\underbrace{\mathrm{v}_0^{\mathrm{val}_0}}_{\text{head}} \leftarrow \underbrace{\mathrm{v}_1^{\mathrm{val}_1} \wedge \mathrm{v}_2^{\mathrm{val}_2} \wedge \mathrm{v}_3^{\mathrm{val}_3} \wedge \cdots \wedge \mathrm{v}_m^{\mathrm{val}_m}}_{\text{body}}$

- variables : $\mathrm{v}_0, \ldots, \mathrm{v}_\mathrm{m}$
- values : $\mathrm{val}_0, \ldots, \mathrm{val}_\mathrm{m}$
- head: target variables (step t)
- body: feature variables (step $t-1$)

LFIT Pipeline

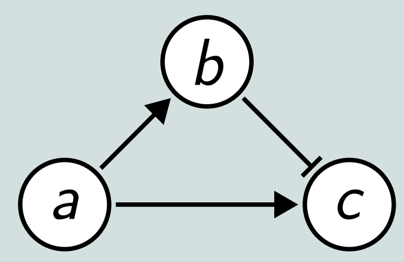
Input: Transitions

a_{t-1}	b_{t-1}	c_{t-1}	b_t	c_t
1	0	0	1	1
1	1	0	1	1
1	1	0	1	0
...	...	...	...	...
1	1	1	1	1

Output: Logic program

$$\{\, b^0 \leftarrow a^0,\quad b^1 \leftarrow a^1,\quad c_t^0 \leftarrow a^0,\quad b^1 \leftarrow b^1,\quad c^0 \leftarrow b^1,\quad c^1 \leftarrow a^1,\quad c^1 \leftarrow b^0 \wedge c^1 \,\}$$

Discrete model:



Hydro: a
Plankton: b, c

LFIT learns both likeliness & unlikeliness rules. Weighting rules by the number of verified observations produces a predictive model.

6. Improving rules

1,683 likeliness rules and 1,981 unlikeliness rules, accuracy: 0.67

Pareto frontier on subsets of the body of rules:

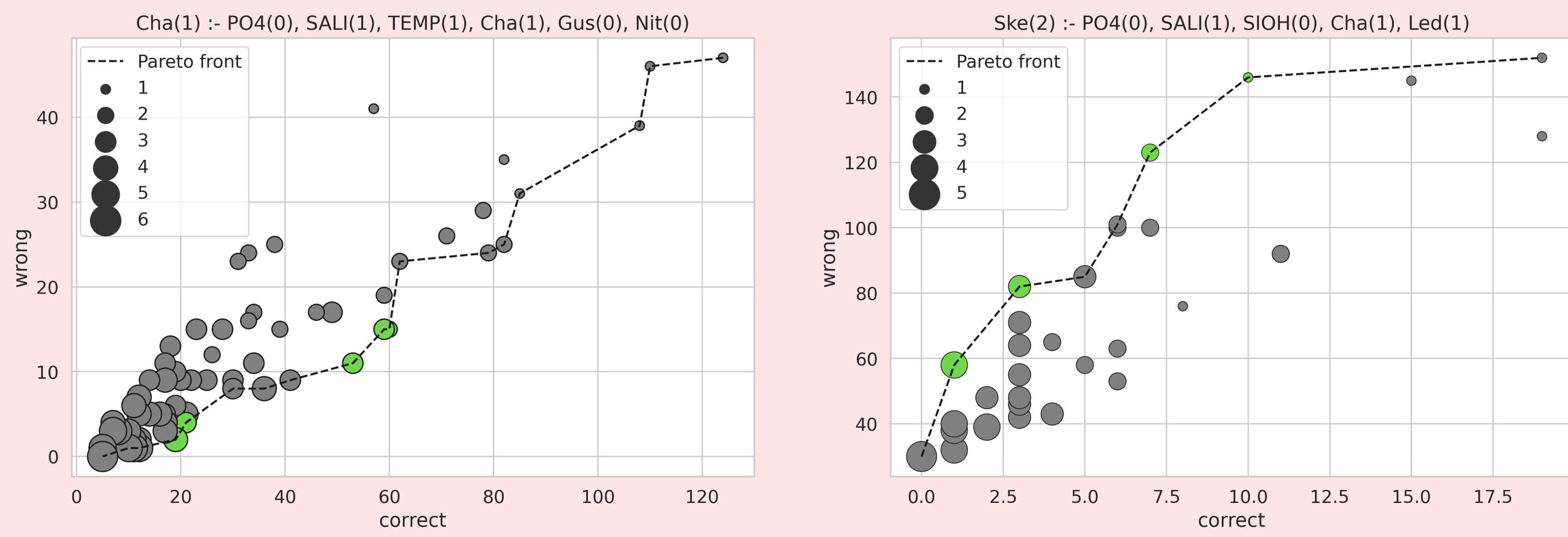


Figure: Pareto frontier of weighted combinations for two examples rules

1,609 likeliness rules and 1,405 unlikeliness rules, accuracy: 0.72

7. Influence Graph

Possible influences: activation ($a \xrightarrow{+} b$) or inhibition ($a \xrightarrow{-} b$)

Process: compare values of targets with each feature atom

Influence Extraction

$$a^0 \leftarrow a^0 \wedge b^1 \wedge c^2 \implies \begin{cases} a \xrightarrow{+} a \\ b \xrightarrow{?} a \\ c \xrightarrow{-} a \end{cases}$$

$$a^0 \leftarrow a^1 \implies a \xrightarrow{-} a$$

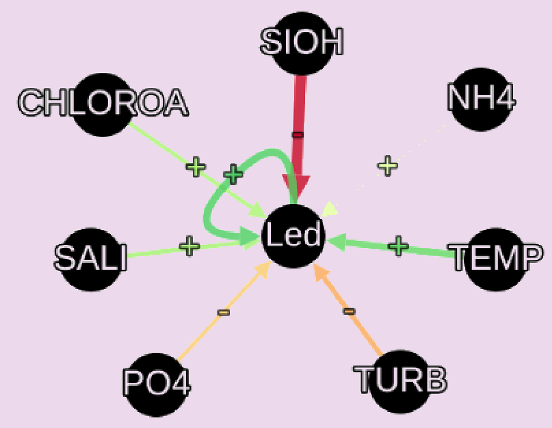


Figure: Influence extraction and influence graph of the species Led.

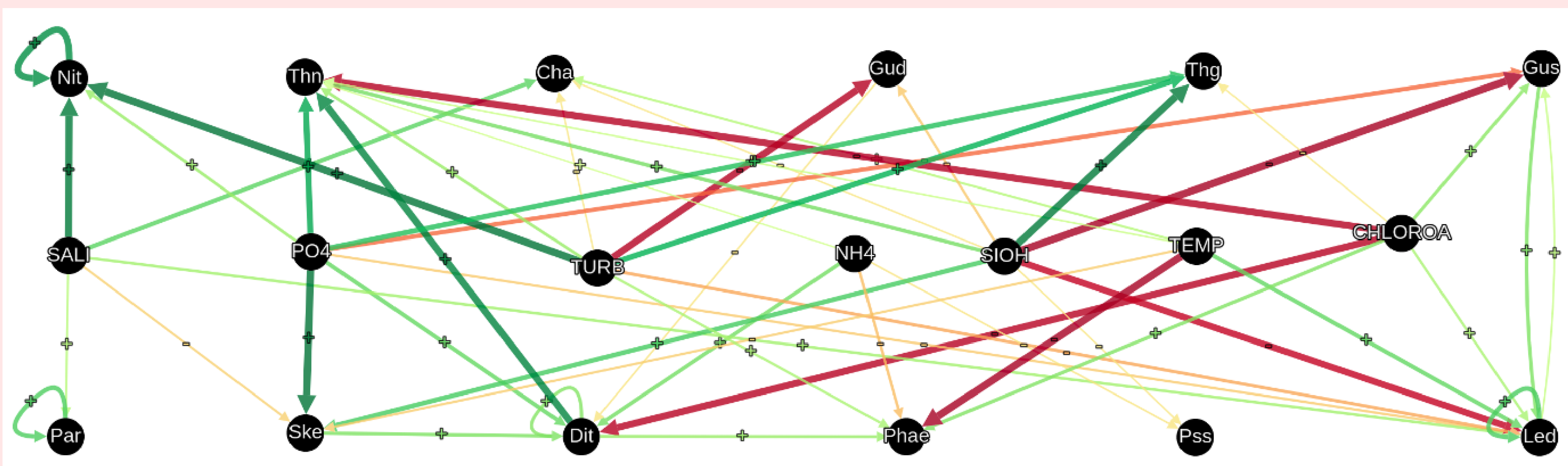


Figure: Influence graph between hydrological factors and phytoplankton species.

8. Perspectives

- Develop heuristics for discretization/Pareto frontier extraction.
- Adapt data augmentation methods to improve model accuracy.
- Provide theoretical guarantees over extracted influences.

Karasiewicz, S., Breton, E., Lefebvre, A., Hernandez Farinas, T., Lefebvre, S.: Realized niche analysis of phytoplankton communities involving HAB: Phaeocystis spp. as a case study. *Harmful Algae* 72, 1–13 (2018)

Ribeiro, T., Folschette, M., Magnin, M., Roux, O., Inoue, K.: Learning dynamics with synchronous, asynchronous and general semantics. In: International Conference on Inductive Logic Programming. pp. 118–140. Springer (2018)

SRN – Regional Observation and Monitoring program for Phytoplankton and Hydrology in the eastern English Channel: SRN dataset. 1992-2016. (2017)