



Automatic Modeling of Dynamical Interactions Within Marine Ecosystems

Omar IKEN¹, Maxime FOLSCHETTE¹, Tony RIBEIRO²

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

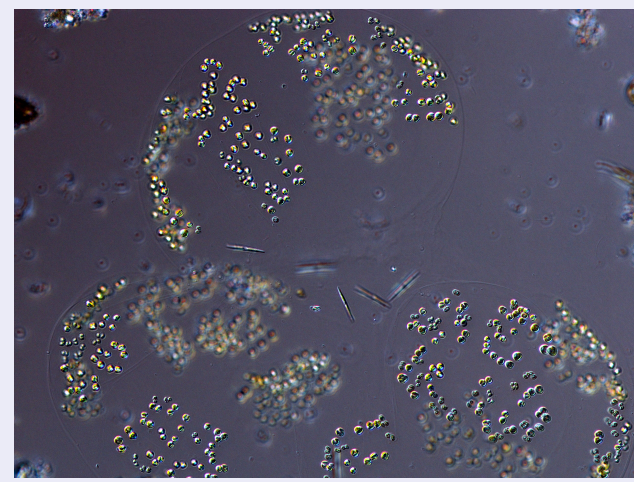


Figure: Phaeocystis (phytoplankton)

Phytoplankton

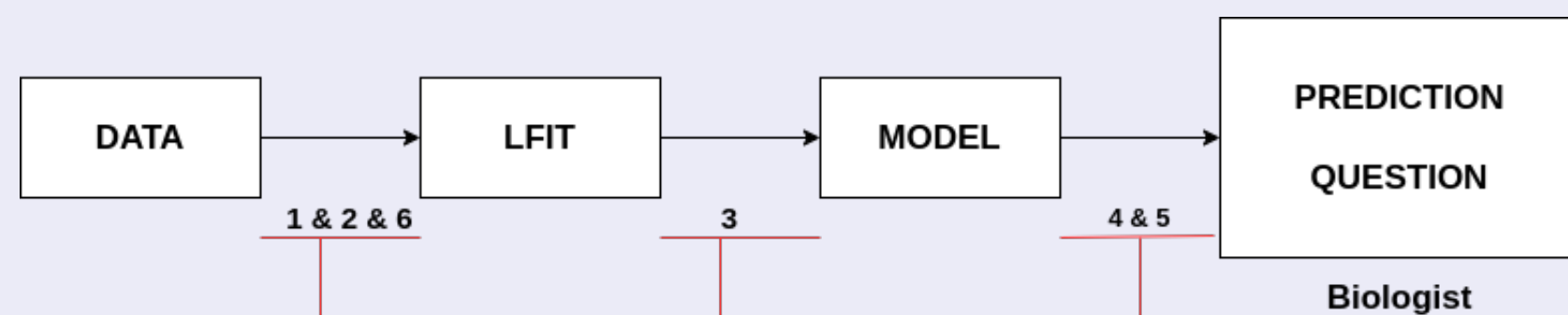
Microscopic organisms that live in aquatic environments

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



Figure: 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

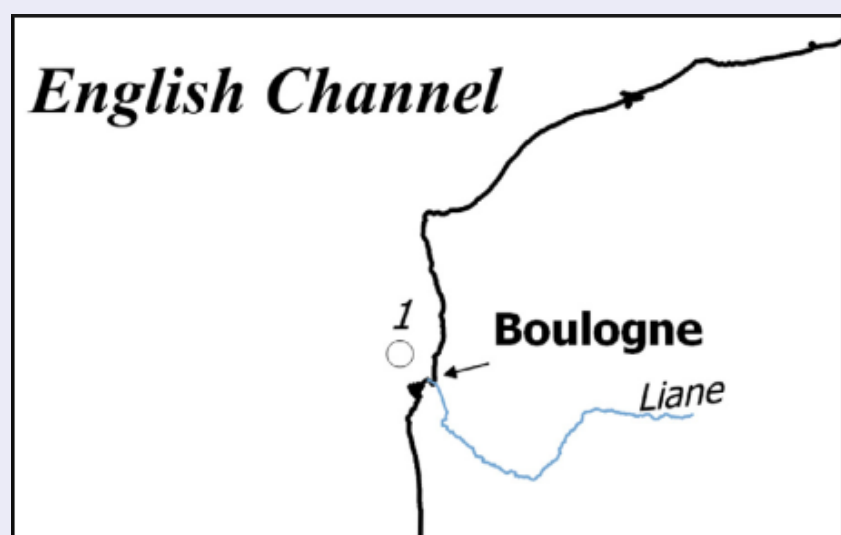


Figure: Chosen observation station



Figure: Blooms of phytoplankton

Sampling location	Sampling date	Taxon	Value	Type	Sampling depth
001-P-015	1992-05-18	CHLOROAO	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

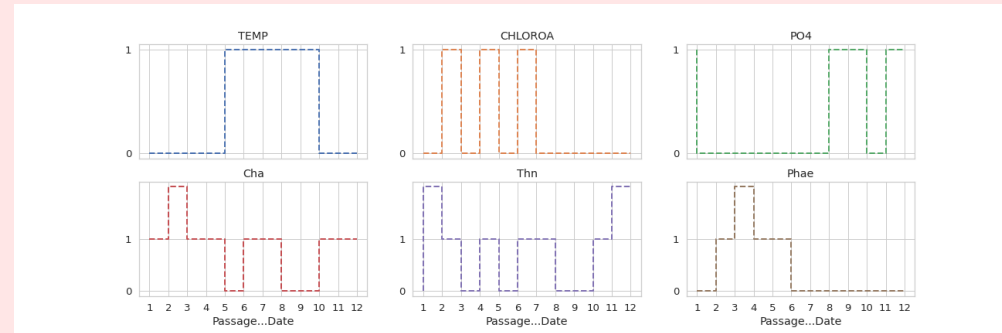
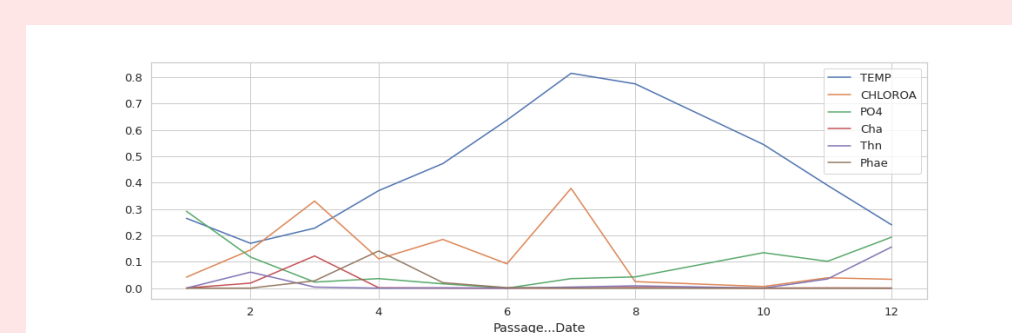


Figure: Variation of some features over 2012 and their discretization

8. Perspectives

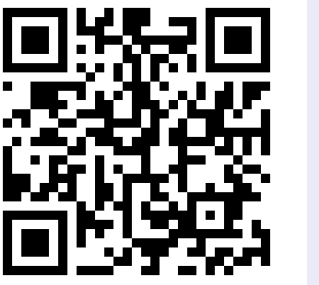
- Develop heuristics for discretization/Pareto frontier extraction.
- Adapt data augmentation methods to improve model accuracy.
- Provide theoretical guarantees over extracted influences.

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)

$$\text{Rule: } \underbrace{v_0^{\text{val}_0}}_{\text{head}} \leftarrow \underbrace{v_1^{\text{val}_1} \wedge v_2^{\text{val}_2} \wedge v_3^{\text{val}_3} \wedge \dots \wedge v_m^{\text{val}_m}}_{\text{body}}$$



- variables : v_0, \dots, v_m
- values : $\text{val}_0, \dots, \text{val}_m$
- head: target variables (step t)
- body: feature variables (step $t - 1$)

LFIT Pipeline

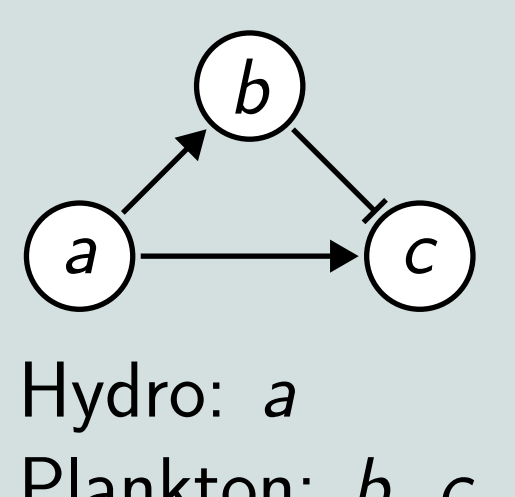
Input: Transitions

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

$$\left\{ \begin{array}{l} b^0 \leftarrow a^0, \quad b^1 \leftarrow a^1, \\ c_t^0 \leftarrow a^0, \quad b^1 \leftarrow b^1, \\ c^0 \leftarrow b^1, \quad c^1 \leftarrow a^1, \\ c^1 \leftarrow b^0 \wedge c^1 \end{array} \right\}$$

Discrete model:



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:

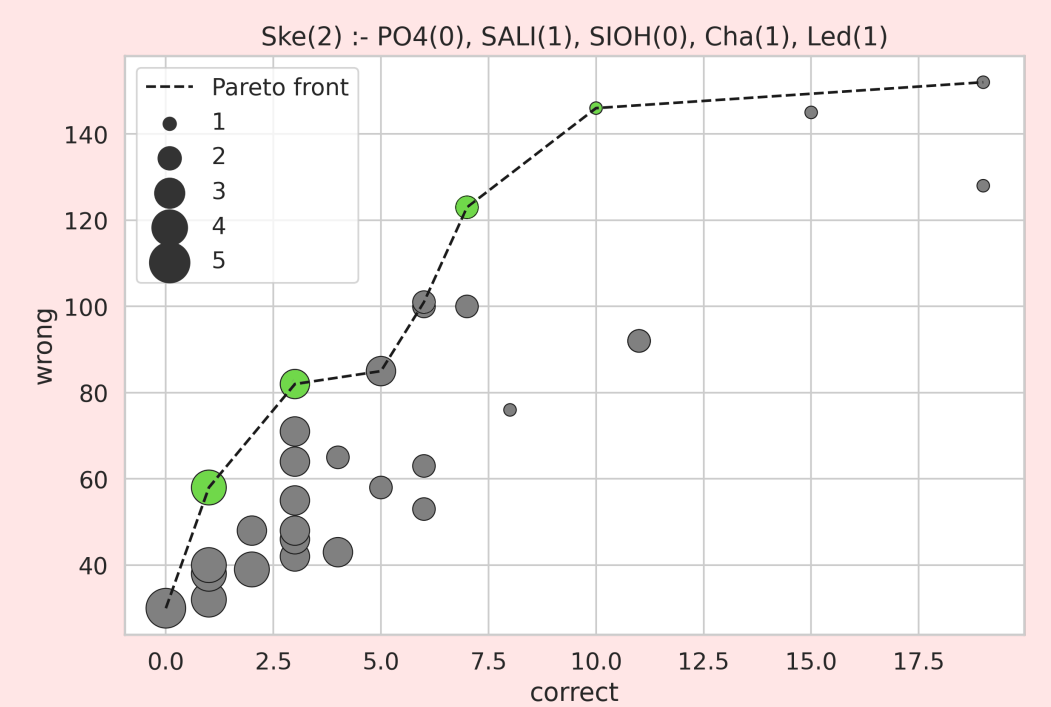
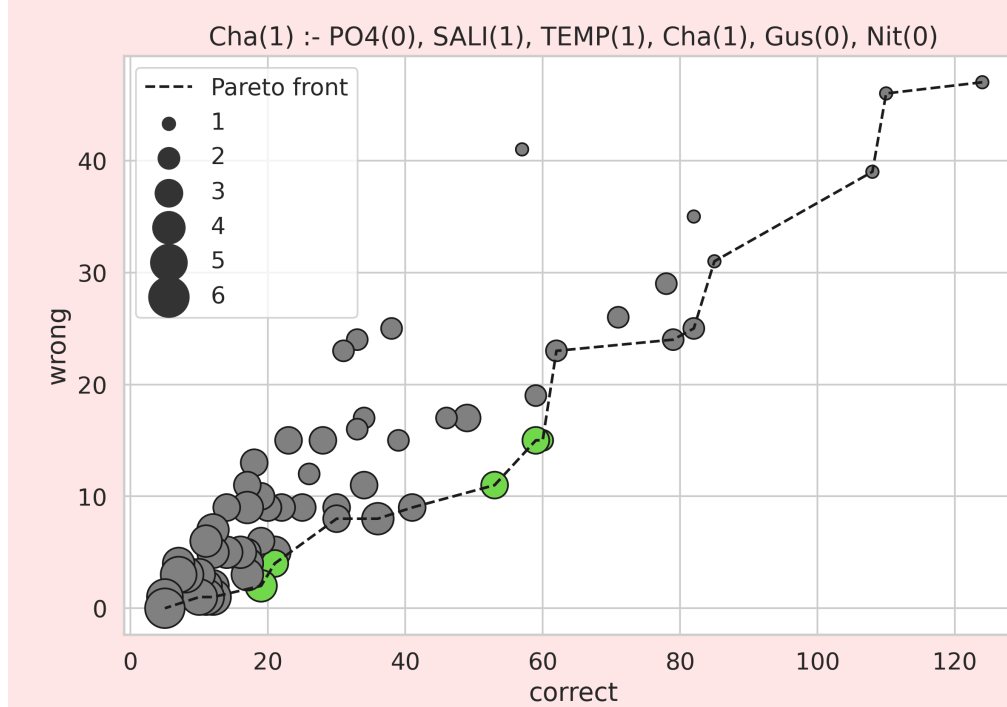


Figure: Pareto frontier of weighted combinations for two examples rules

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

$$a^0 \leftarrow a^0 \wedge b^1 \wedge c^2 \implies \left\{ \begin{array}{l} a \xrightarrow{+} a \\ b \xrightarrow{?} a \\ c \xrightarrow{-} a \end{array} \right.$$

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

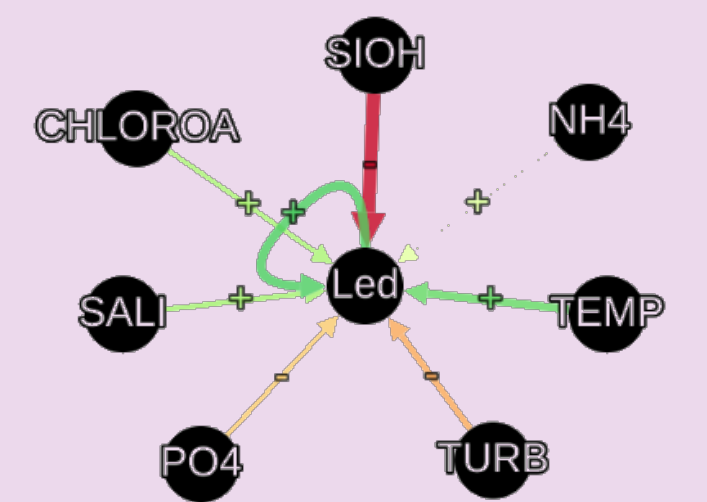


Figure: Influence extraction and influence graph of the species Led.

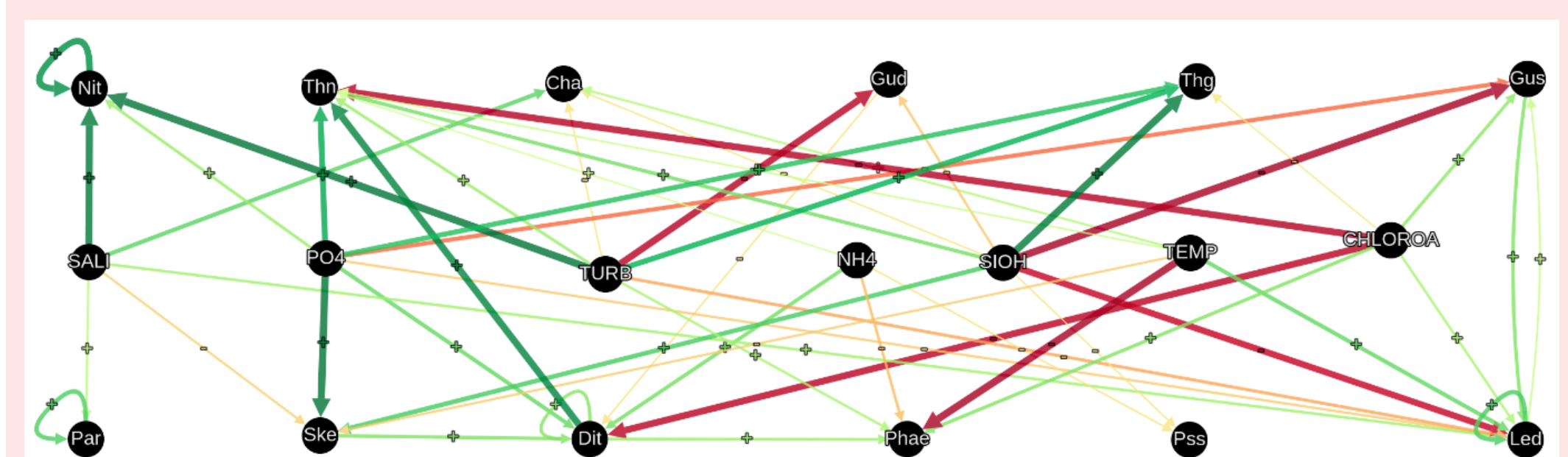


Figure: Influence graph between hydrological factors and phytoplankton species.