NEURAL METHODS AND MIXED DATA:

TOWARDS A BETTER SPATIAL AND TEMPORAL RESOLUTION OF MARINE ECOSYSTEMS AND PHYTOPLANKTON

Robin Fuchs

Supervisors:

- Denys Pommeret
- Melilotus Thyssen

Institut de Mathématiques de Marseille

I2M - UMR CNRS 7373

Aix-Marseille Université • CNRS • École Centrale de Marseille

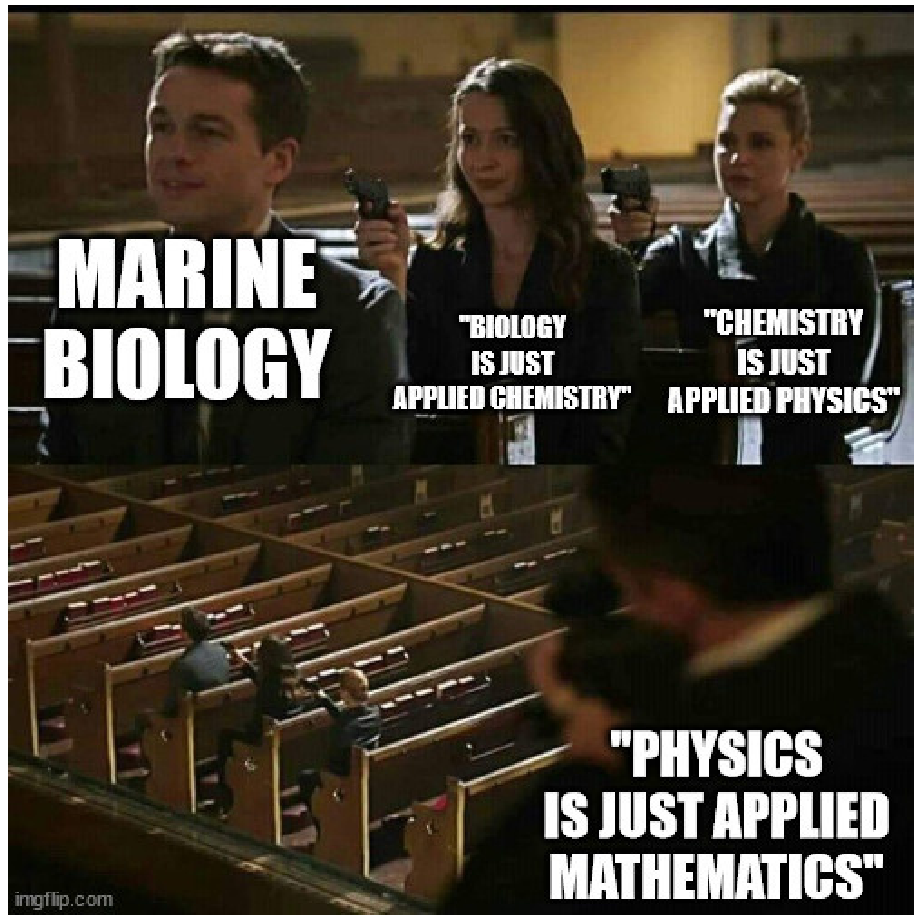
Institut Méditerranéen d'Océanologie

Introduction

Introduction

Ph.D. thesis in Statistics applied to phytoplankton and marine ecosystems

- Developing new statistical models suited for oceanography
- Apply innovative statistical methodologies to oceanographic issues



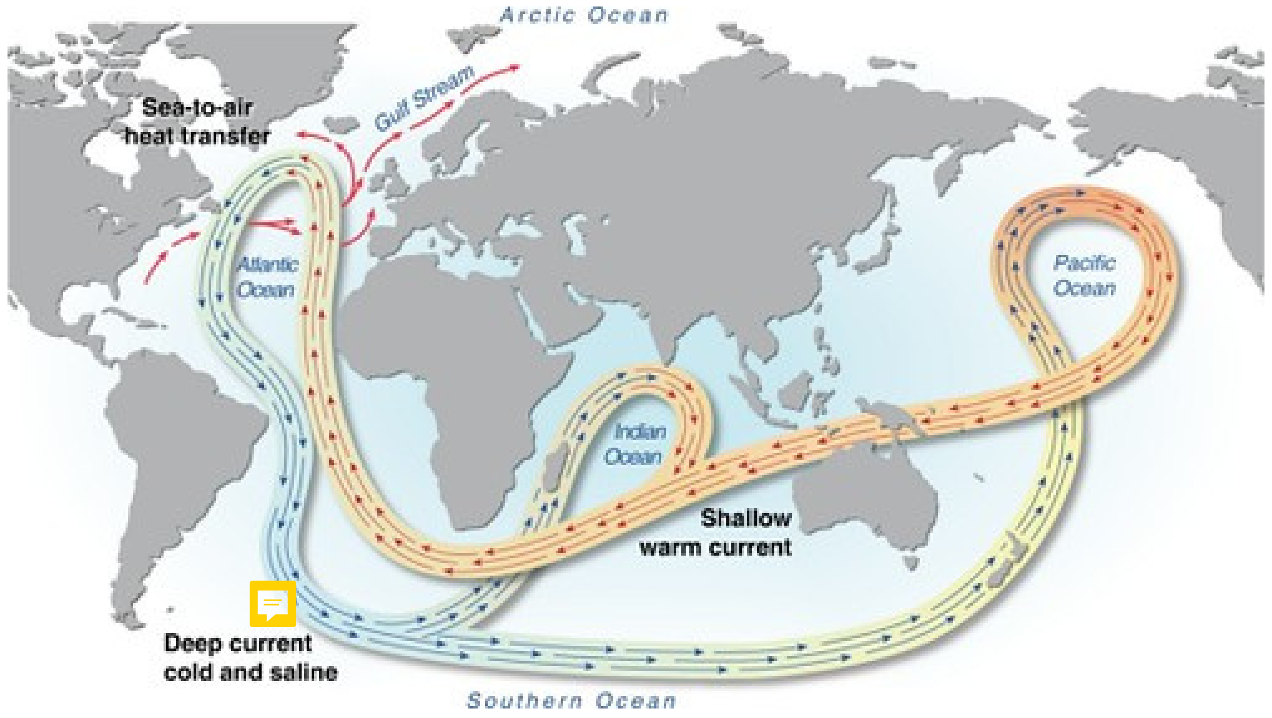
Why studying phytoplankton?

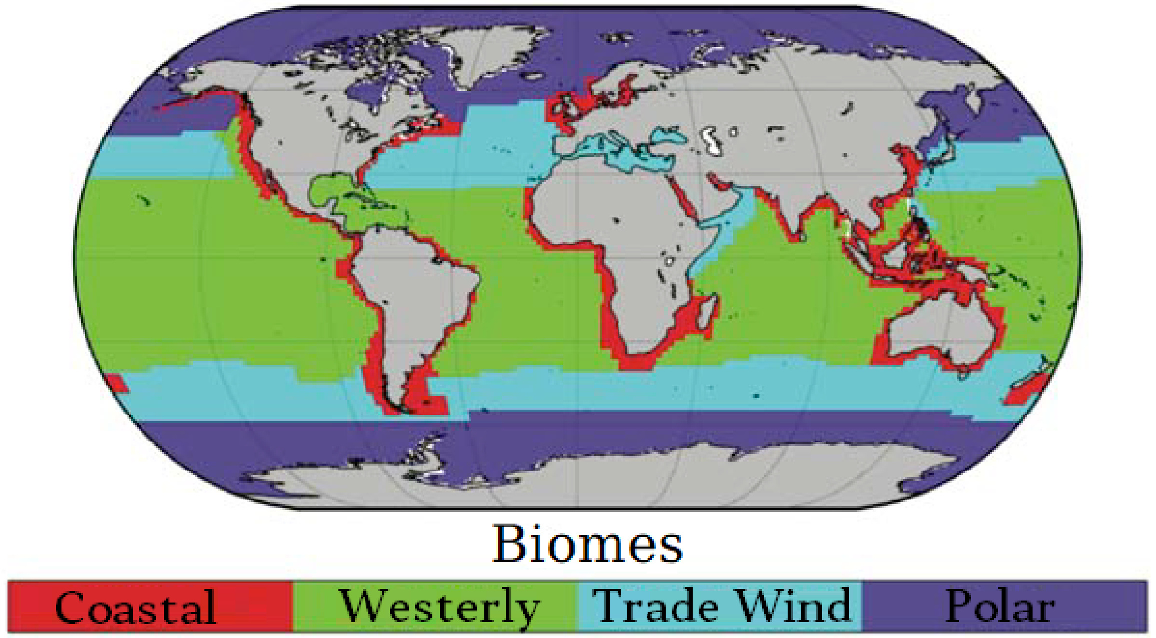
- Half the CO2 captured
- One of the first elements of the marine food web

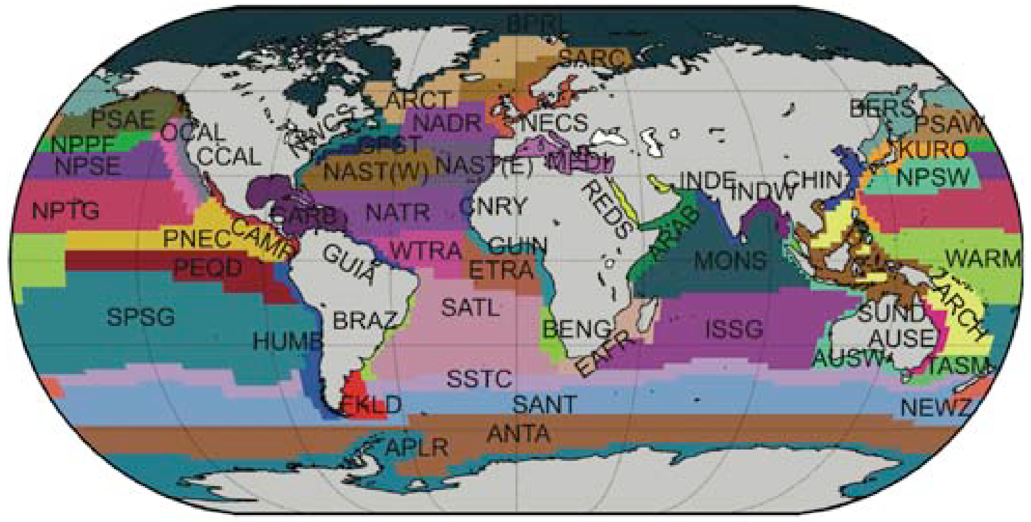
Introduction

Standard horizontal partition

- Phytoplankton is mostly non-motile
- Global circulation is important at a large temporal and spatial scale
- Creation of highly contrasted **regions** or **biomes** by Longhurst (1995)
- Physics strongly shape phytoplankton repartition







Biogeochemical provinces

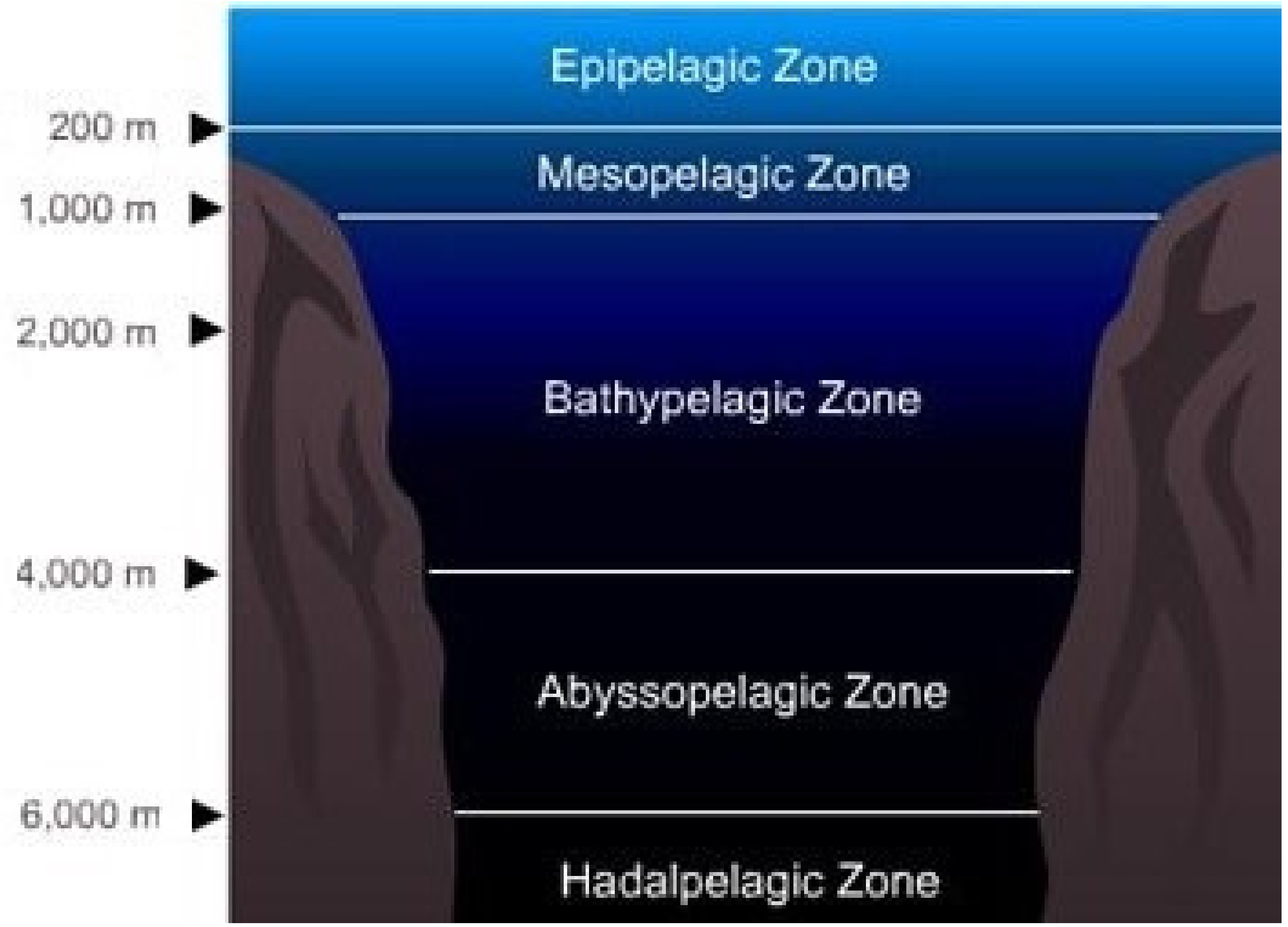
Reygondeau (2013)

Introduction

Standard vertical partition

- Phytoplankton lives in the upper layer of the ocean: the epipelagic zone
- Access to light needed for the photosynthesis
- Access to nutrients coming from deeper waters

Did you know ?
The epipelagic zone is also called the twilight zone?



Classical vertical partition of the water column, adapted from deepoceanfacts.com

Introduction

Yes, but...

- Fixed zone boundaries for moving environments
- Non-fractal phenomena: Local boundaries do not reflect global boundaries

Solutions provided:

Determining local fundamental ecological niches* and vertical epipelagic boundaries based on coupled physics/biology variables

* Hutchinson's ecological niches (1957):
  - **Fundamental niche**: all the conditions necessary for an organism or species to exist
  - Realized niche: part of the fundamental niche that the organism/species can occupy due to the competition with other species.

Introduction

Need for high frequency observation

Why?

- High division rate
- Nyquist theory: Infra-day observation
- Expected high reactivity to pulse events

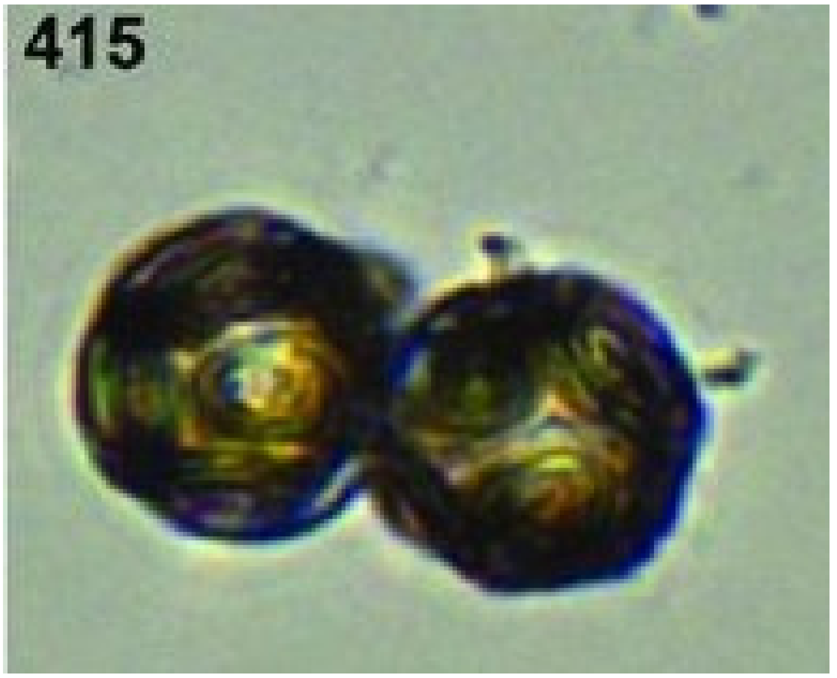
How?

Multiple observation methods:

- **Automated Flow cytometry**
- HPLC
- Satellite observation

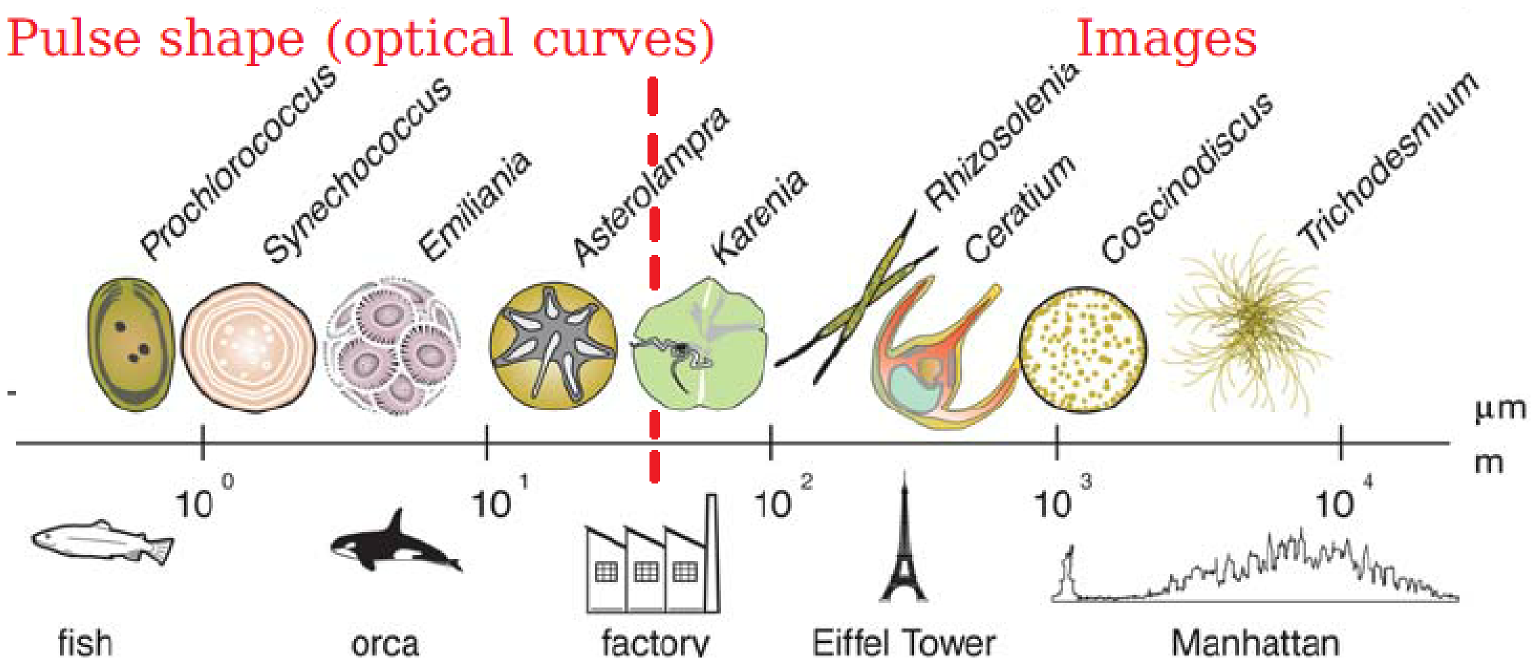
Pulse-shape recording Flow cytometry (AFCM)

- High frequency
- Low monitoring cost
- Resolve the whole size range



Walker et al. (2018)

CytoSense Automated Flow Cytometer

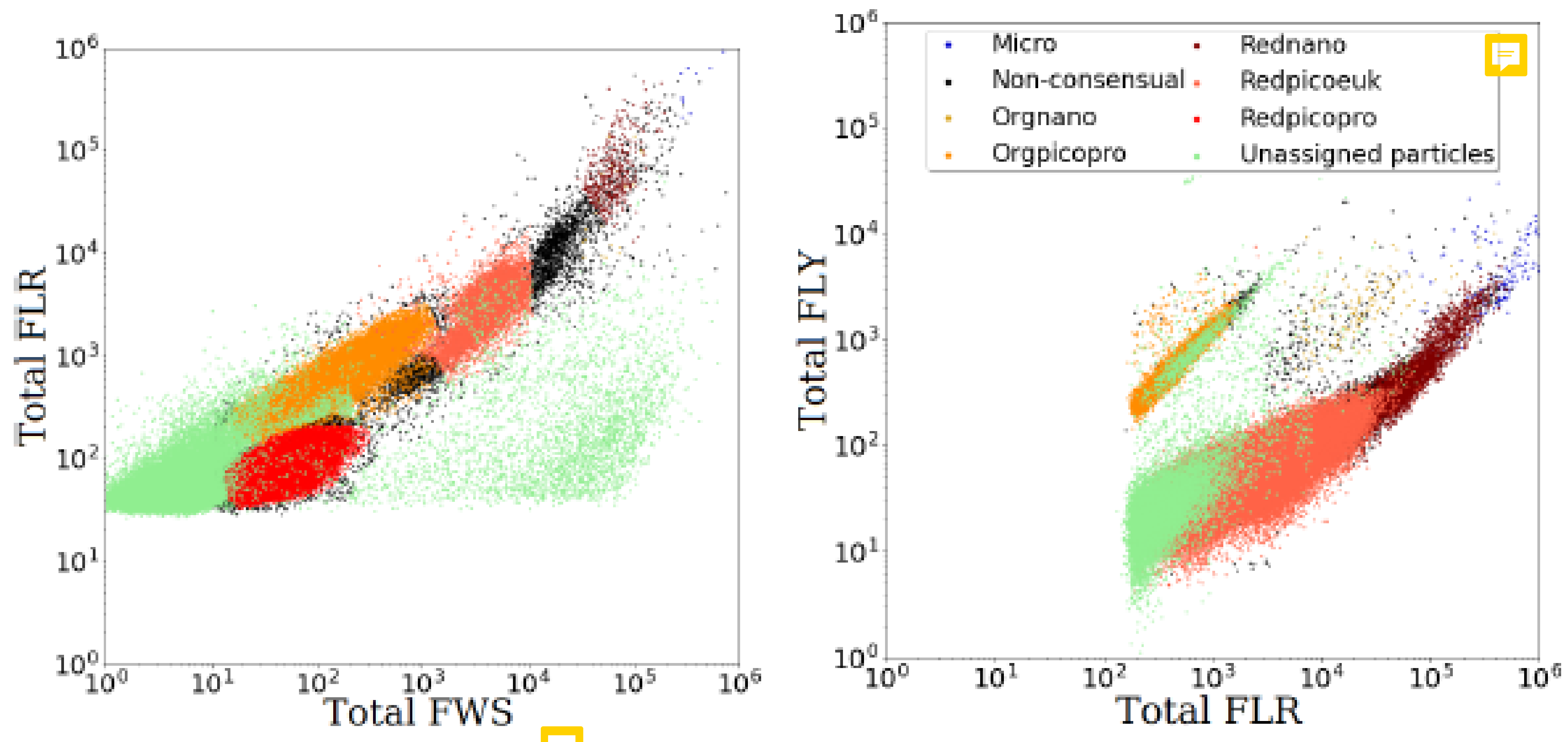


Finkel et al. (2010)

Introduction

Flow cytometric functional groups (PFGs)

→ Focus on piconano-phytoplankton resolved by Automated pulse-shape recording Flow Cytometry (AFCM)



Fuchs et al. (submitted)

Interoperable nomenclature	Expert suggested nomenclature
Micro	Microphytoplankton
Orgnano	Cryptophytes-like
Orgpicopro	*Synechococcus*
Rednano	Nanoeukaryotes
Redpicoeuk	Picoeukaryotes
Redpicopro	*Prochlorococcus*

Table: Correspondence table between the SeaDataCloud Flow Cytometry Standardised Cluster Names, identified as the interoperable nomenclature and published by the Natural Environmental Research council, and the correspondence with an expert denomination.

Thyssen et al. (submitted)

Size: Redpicopro < Orgpicopro < Redpico < Rednano < Orgnano < Redmicro

Introduction

Sporadic wind events

Potential high impact on phytoplankton

- Water masses replacements: Induced upwelling
- Nutrient enrichements
- Temperature drop
- Perturbation of zooplankton predation and viral lysis

Evidences in the literature

- Thyssen et al. (2008): All groups do not react in the same way (based on two events)
- Dugenne et al. (2014): Increase in the cell growth rates after the event (based on one event)
- Martin-Platero (2018): Sensibility to the temporal frequency (one event)
- ...

Yes, but...

AFCM lacks standardization

- Until Thyssen et al. (submitted): High diversity of the nomenclature used between studies
- The data treatment (gating) is done manually: but **few error assessments** performed

Standardized gating method

- Estimate the manual error
- Introduce a automatic treatment method based on convolutional neural network

Low result representativeness

- Studies limited to one or two wind events
- Manual treatment of phytoplankton data
- Low temporal frequencies
- Focus only on the biggest phytoplankton cells

Long high frequency study

- Twenty wind events over two years
- Rupture detection methods to estimate causal effects

OUTLINE

Characterizing local ecological niches and vertical boundaries

I/ Clustering Mixed data

MDGMM presentation and example

Application to phytoplankton ecological niches

II/ Generating Mixed data

MIAMI presentation and example

Application to phytoplankton future climate

III/ Local vertical boundaries for the epipelagic layer

Rupture detection though the water column

Estimation of the high-frequency response to intense wind-events

I/ Example of phytoplankton response to a storm in the Ligurian Sea

FUMSECK cruise May 2019

II/ Completing AFCM automation

Introducing Convolutional neural methods for AFCM pulse shapes

III/ Generalization over two years of data

Causal relationships to sporadic wind-events

Part I:
Ecological niches and vertical boundaries

Mixed data

What is it?

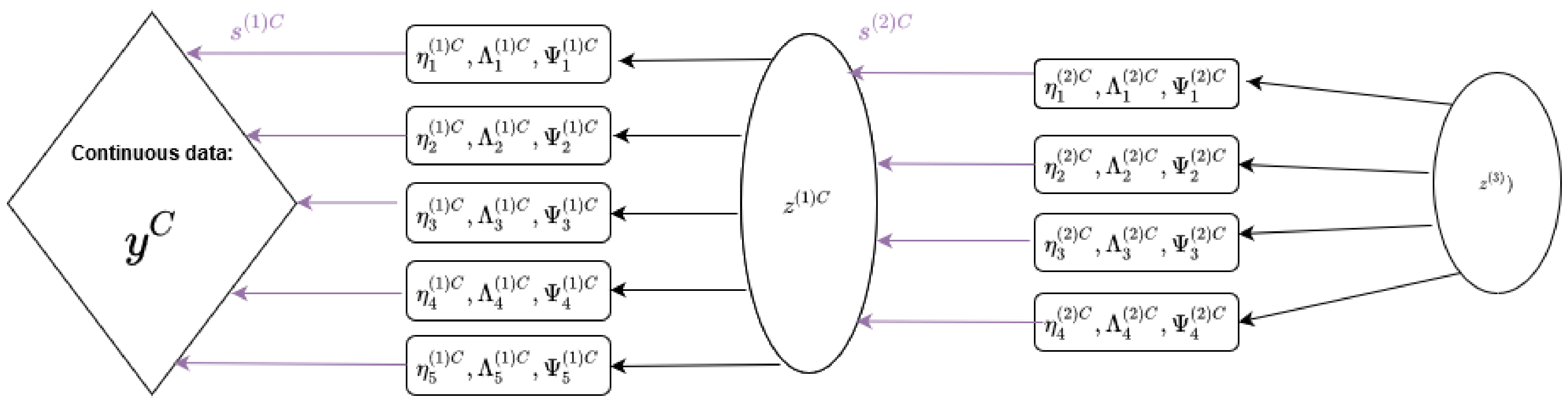
- Categorical variables: That exhibit a finite and non-ordered number of modalities
- Ordinal variables presenting a finite and ordered number of modalities
- Binary variables taking only two modalities
- Count variables having a finite countable and ≥ 2 number of modalities
- Continuous variables: Infinite and uncountable number of modalities

Potential issues

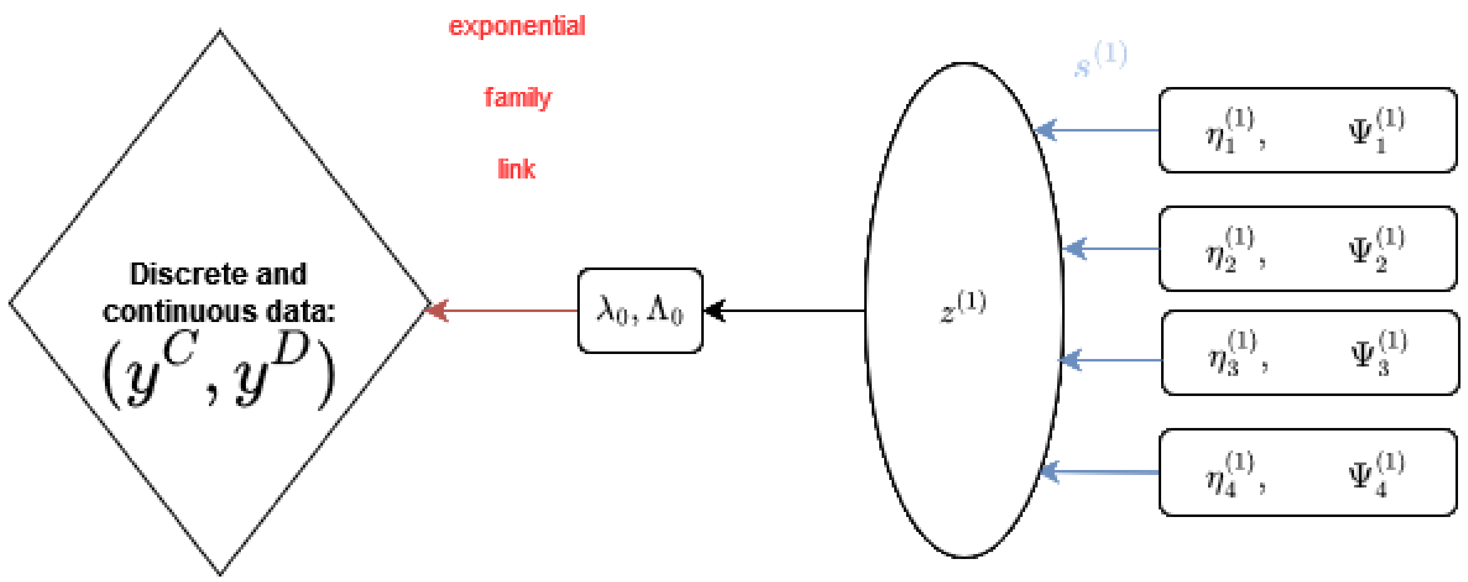
- Heterogeneous data types: complicated variable space
- Difficult to define what similar observations are
- How to model the variable types?

MDGMM

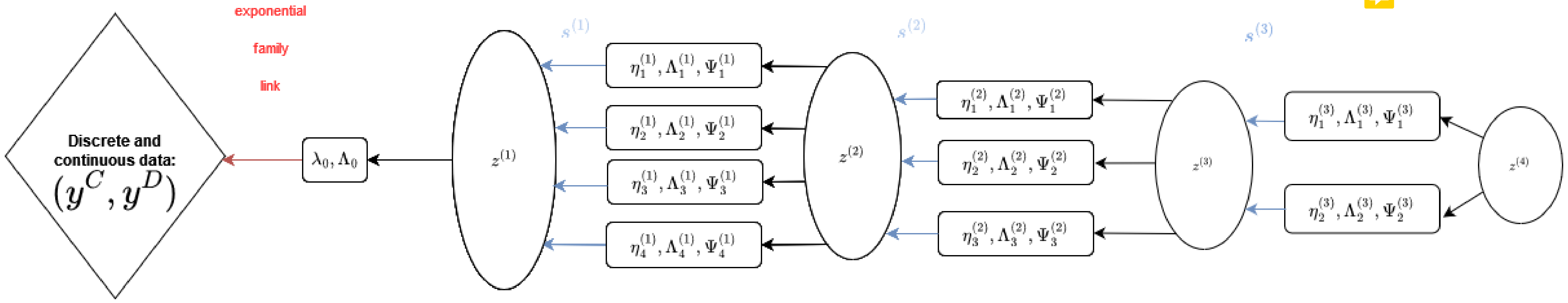
DGMM (Viroli and McLachlan, 2019)



GLLVM (Cagnone et Viroli, 2014)



MDGMM (Fuchs et al., 2021)



DGMM Genesis

Factor Analyzer (FA)

Goal of the model:
Compress the signal held by each observation y_i from dimension p to dimension r with $r << p$.

Model expression:

$$\begin{aligned} y_i = \eta \quad & + \quad \Lambda z_i \quad && + u_i \\ (p, 1) = (p, 1) + \quad & (p, r) \times (r, 1) && + (p, 1) \end{aligned}$$

with $i \in [1, n]$ the observation index, η a vector of constants, $z_i \sim \mathrm{N}(0, I_p)$, $u_i \sim \mathrm{N}(0, \Psi)$ and Λ the loading matrix.

DGMM Genesis

Mixture of Factor Analyzers (MFA)

Generalisation of the Factor Model:

The signal held by each observation can be compressed in K_1 possible ways:

$$y_i = \eta_{k_1} + \Lambda_{k_1} z_i + u_{i,k_1} \text{ with probability } \pi_{k_1}$$

For $k_1 \in [1, K_1]$ and with $z_i \sim \mathrm{N}(0, I_p)$

DGMM Genesis

Two-layer DGMM

A two layers DGMM corresponds to two nested MFAs:
$\Rightarrow$ we assume now that z_i is itself a MFA with K_2 factors of dimensions r_2, with $r_2 < r_1 < p$. Hence:

$$\begin{cases} y_i = \eta_{k_1}^{(1)} + \Lambda_{k_1}^{(1)} z_i^{(1)} + u_{i,k_1}^{(1)} \text{ with probability } \pi_{k_1}^{(1)} \\ z_i^{(1)} = \eta_{k_2}^{(2)} + \Lambda_{k_2}^{(2)} z_i^{(2)} + u_{i,k_2}^{(2)} \text{ with probability } \pi_{k_2}^{(2)} \end{cases}$$

with $z_i^{(2)} \sim \mathrm{N}(0, I_p)$, $k_1 \in [1, K_1]$ and $k_2 \in [1, K_2]$

DGMM Genesis

L-layer DGMM

A DGMM(L) is therefore a succession of L nested MFAs, and can be written as :

$$\begin{cases} y_i = \eta_{k_1}^{(1)} + \Lambda_{k_1}^{(1)} z_i^{(1)} + u_{i,k_1}^{(1)} \text{ with probability } \pi_{k_1}^{(1)} \\ z_i^{(1)} = \eta_{k_2}^{(2)} + \Lambda_{k_2}^{(2)} z_i^{(2)} + u_{i,k_2}^{(2)} \text{ with probability } \pi_{k_2}^{(2)} \\ \ldots \\ z_i^{(L-1)} = \eta_{k_L}^{(L)} + \Lambda_{k_L}^{(L)} z_i^{(L)} + u_{i,k_L}^{(L)} \text{ with probability } \pi_{k_L}^{(L)} \\ z_i^{(L)} \sim \mathcal{N}(0, I_{r_L}) \end{cases}$$

For $k_1 \in [1, K_1]$ and $k_2 \in [1, K_2]$, ...

Extended GLLVM

Cagnone and Viroli (2014)

Original model

$$f(y^D|\Theta_D, \Theta_{L_0+1:}) = \int_{z^{(1)D}} \prod_{j=1}^{p_D} f(y_j^D|z^{(1)D}, \Theta_D, \Theta_{L_0+1:}) f(z^{(1)D}|\Theta_D, \Theta_{L_0+1:}) dz^{(1)D},$$

where y_j^D is the jth variable of y^D and $z^{(1)D}$ is drawn from a standard Gaussian. The density of $f(y_j^D|z^{(1)D}, \Theta_D, \Theta_{L_0+1:})$ is called the link function and belongs to an exponential family.

Examples of link function:

If y_j^D is a binary variable, $f(y_j^D|z^{(1)D}, \Theta_D, \Theta_{L_0+1:}) = f(z^{(1)D})^{y_j^D}(1 - f(z^{(1)D}))^{n-y_j^D}$

If y_j^D is a categorical variable, $y_j^D|z^{(1)D}, \Theta_D, \Theta_{L_0+1:} \sim M(f(z^{(1)D}))$

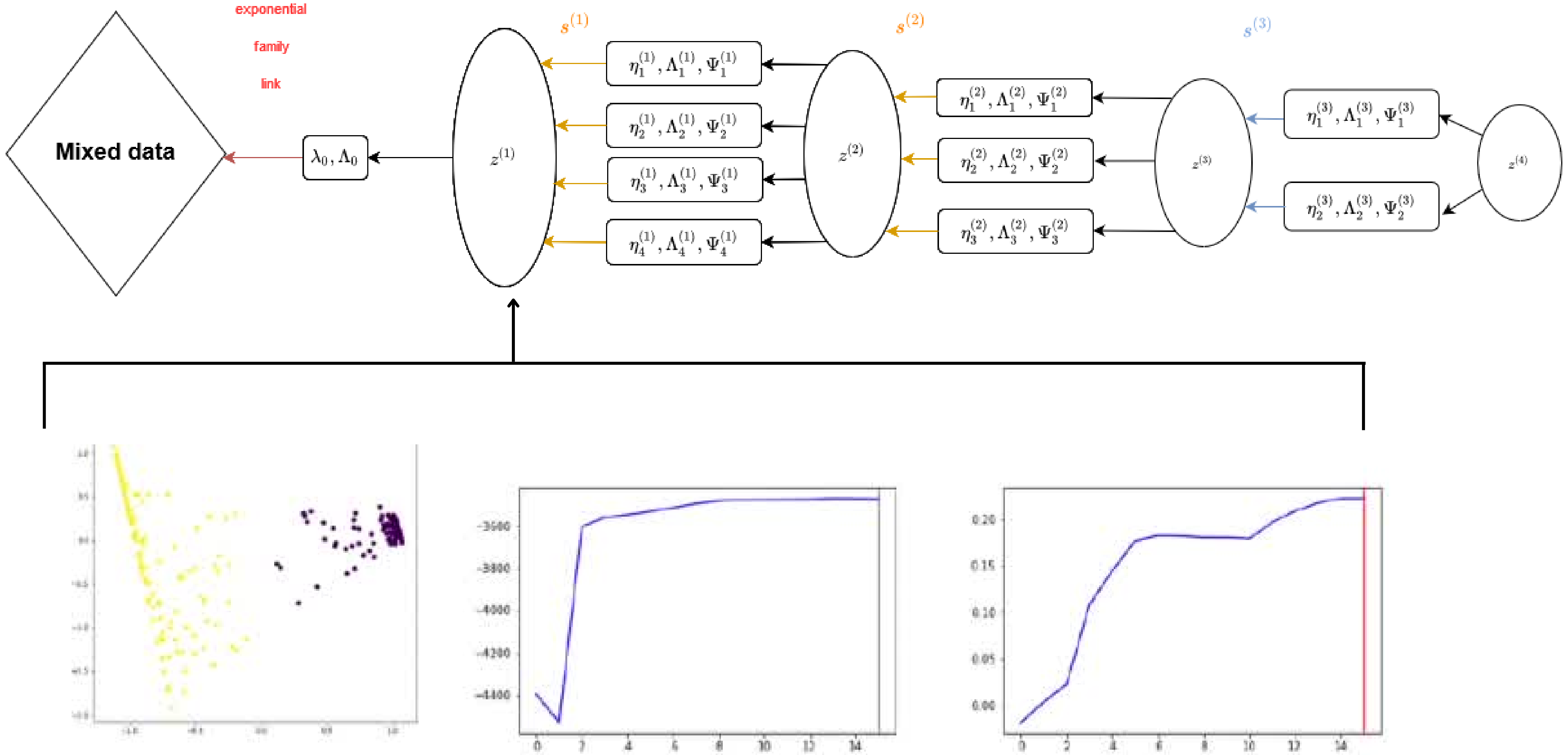
Generalization

- $z^{(1)}$ is a Gaussian
- $z^{(1)}$ is a Mixture of Gaussians
- $z^{(1)}$ is a Mixture of Factor Analyzers
- $z^{(1)}$ is a DGMM

MDGMM Genesis

Fuchs, Viroli and Pommeret (2021)

- Trained by Monte Carlo EM-algorithm (MCEM)
- Initialization based on regressions and MFA fitting
- Architecture selection by post-pruning



(a) Latent representation of the clustering layer (b) Likelihood of the model (c) silhouette Score

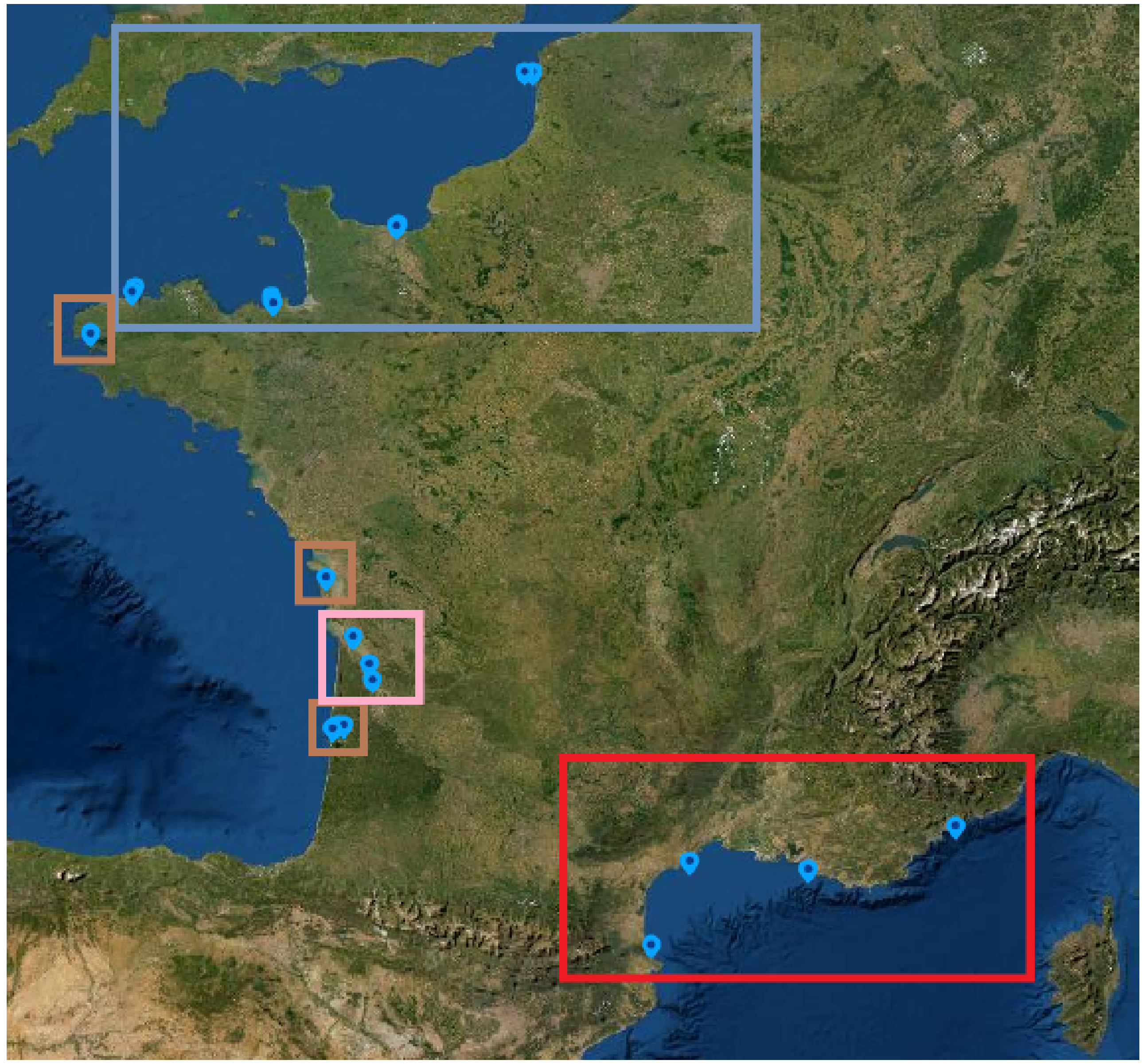
View of $z^{(1)}$ during the training process

MDGMM: Application example

I captured the people in this room
and fed my model with you

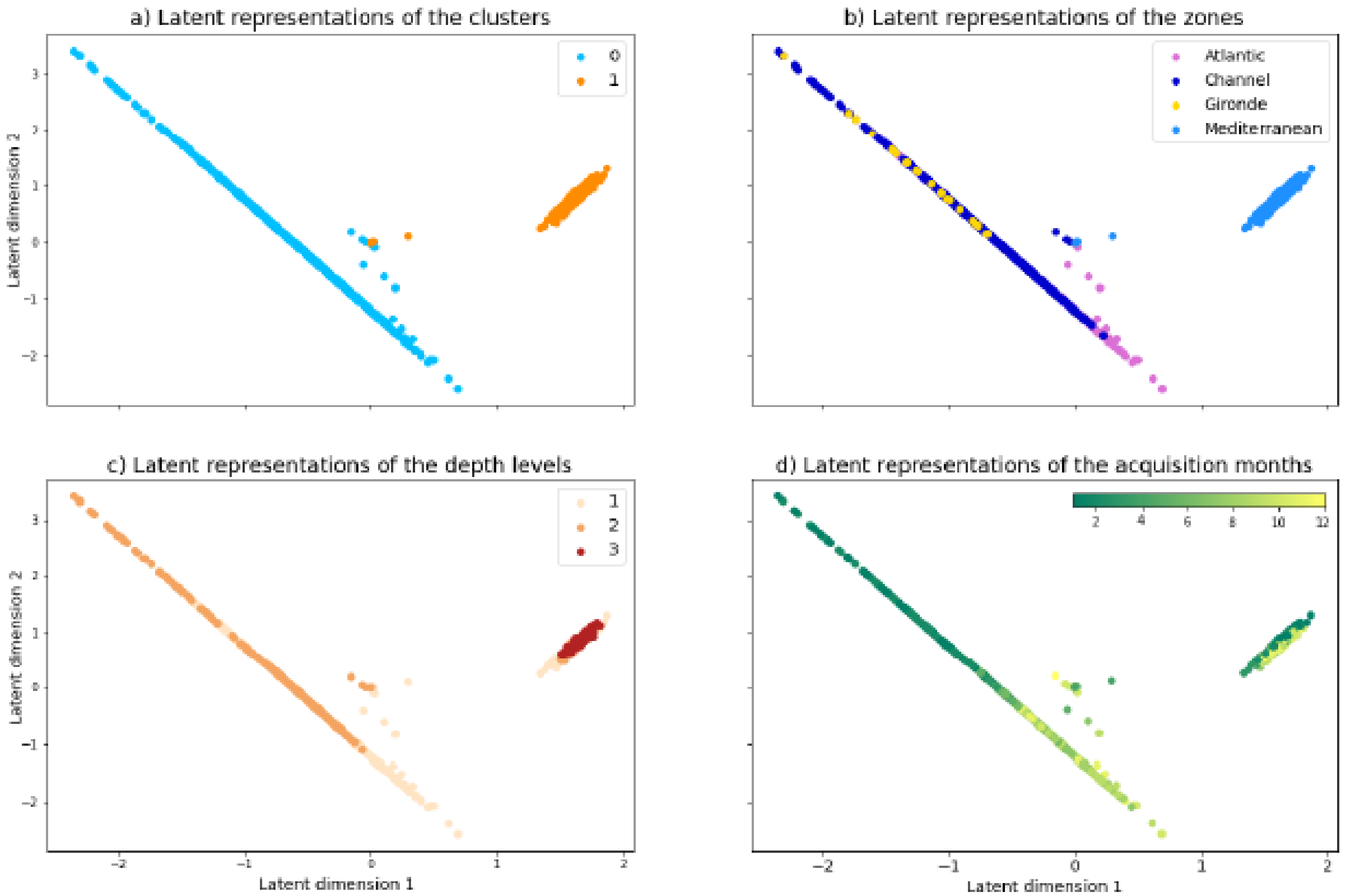
MDGMM: Analyzing ecological niches

Data



Maps of the eleven SOMLIT stations and the associated zones: The Mediterranean Sea stations are denoted by a red rectangle, the Atlantic stations are in brown, the Gironde River stations in pink and the Channel-related stations in blue (based on the Leaflet map library).

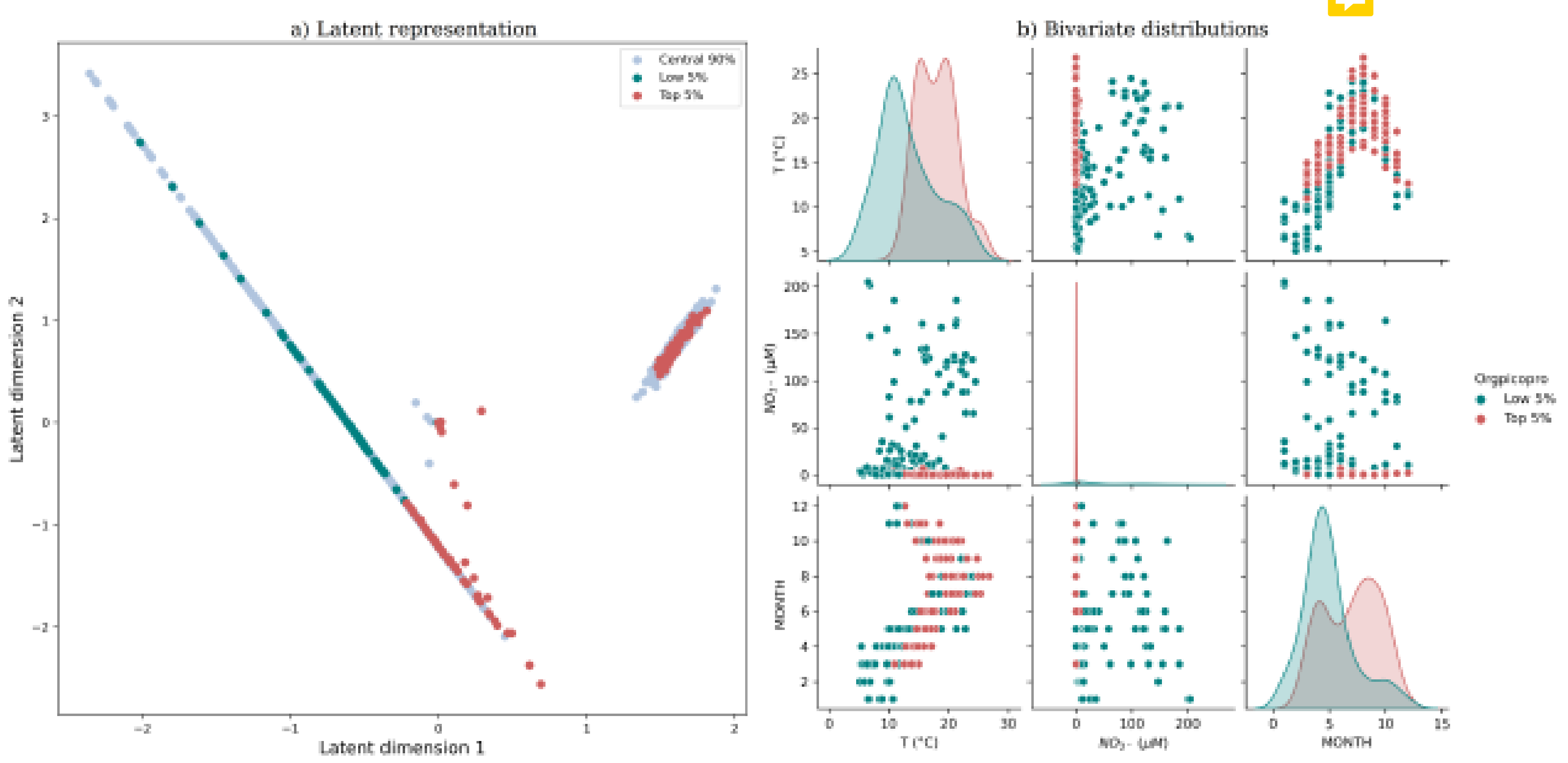
$z^{(1)}$ vizualization



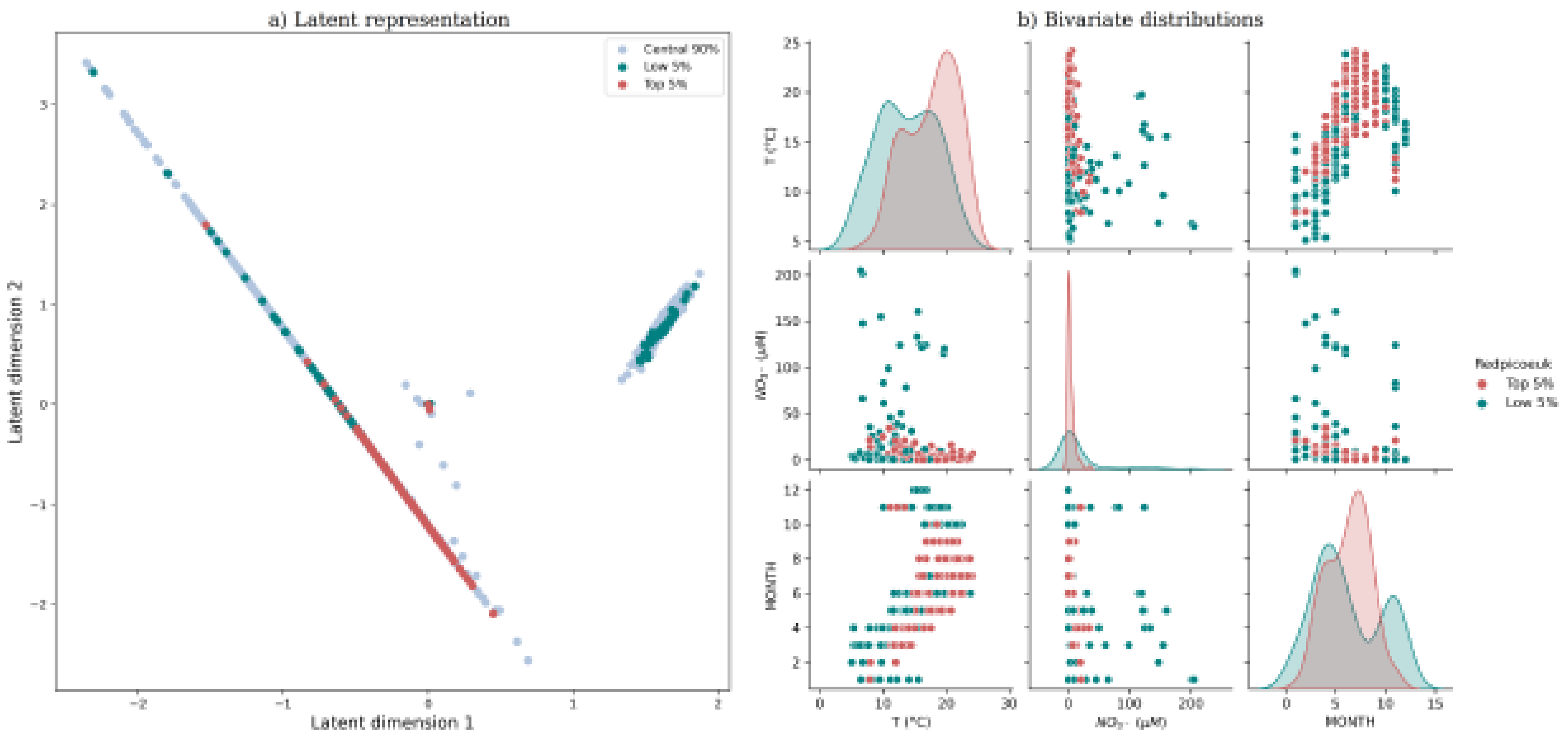
Latent representation of the SOMLIT data. a) Latent representation colored by MDGMM cluster number (the model identifies two clusters here, numbered 0 and 1). b) Latent representation of the data colored by the zone of belonging ("ZONE" variable). c) Latent representation of the observations colored by sampling depth ("DEPTH" variable). d) Latent representation of the data colored by sampling month ("MONTH" variable), 1 corresponds to January and 12 to December.

Ecological niches

Mediterranean Sea: Opposite ecological niches



Orgpicopro distribution representations. a) Representation in the latent space of the lowest 5% abundances, central 90% abundances and top 5% abundances. b) Bivariate distribution of the temperature, nitrate concentration and month broken down between the lowest 5% and top 5% Orgpicopro abundances. The diagonal plots correspond to the marginal distributions of each "environmental" variable for the top 5% (red distribution) and lowest 5% (blue distribution) Orgpicopro abundances.



Redpicoeuk distribution representations. a) Representation in the latent space of the lowest 5% abundances, central 90% abundances and top 5% abundances. b) Bivariate distribution of the temperature, nitrate concentration and month broken down between the lowest 5% and top 5% Redpicoeuk abundances. The diagonal plots correspond to the marginal distributions of each "environmental" variable for the top 5% (red distribution) and lowest 5% (blue distribution) Redpicoeuk abundances.

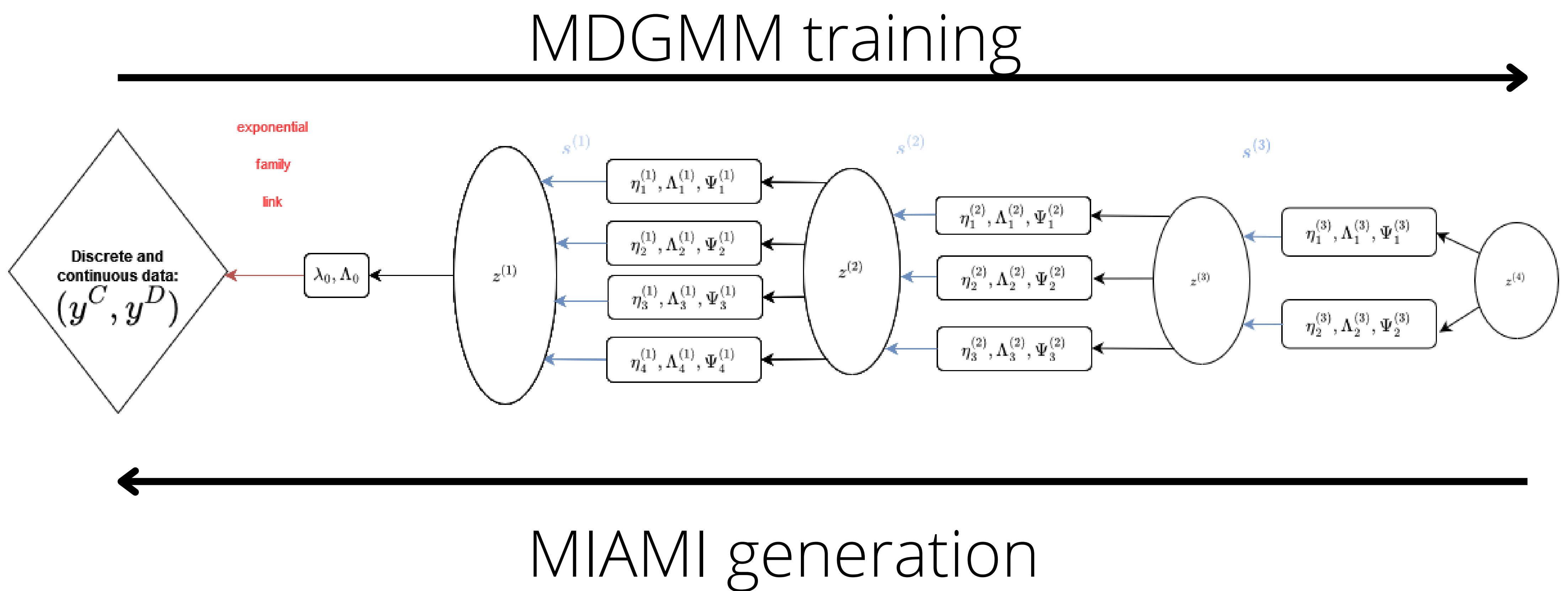
MIAMI: MIxed data Augmentation MIxture

Create synthetic data

The tilded variables being the variables estimated during the MDGMM training and using Bayes rule, one obtains:

$$f(y^*|\tilde{\Theta}) = \frac{f(\tilde{z}^{(1)}|\tilde{\Theta})f(y^*|\tilde{z}^{(1)},\tilde{\Theta})}{f(\tilde{z}^{(1)}|y^*,\tilde{\Theta})}$$
$$\propto f(\tilde{z}^{(1)}|\tilde{\Theta})\prod_{j=1}^{p} f(y_j^*|\tilde{z}^{(1)},\tilde{\Theta}),$$

$f(\tilde{z}^{(1)}|\tilde{\Theta})$ follows a DGMM distribution and $f(y_j|\tilde{z}^{(1)},\tilde{\Theta})$ belongs to an exponential family:
One can sample synthetic observations y^* by sampling values according to weights $f(\tilde{z}^{(1)}|\tilde{\Theta})$.



MIAMI: Example

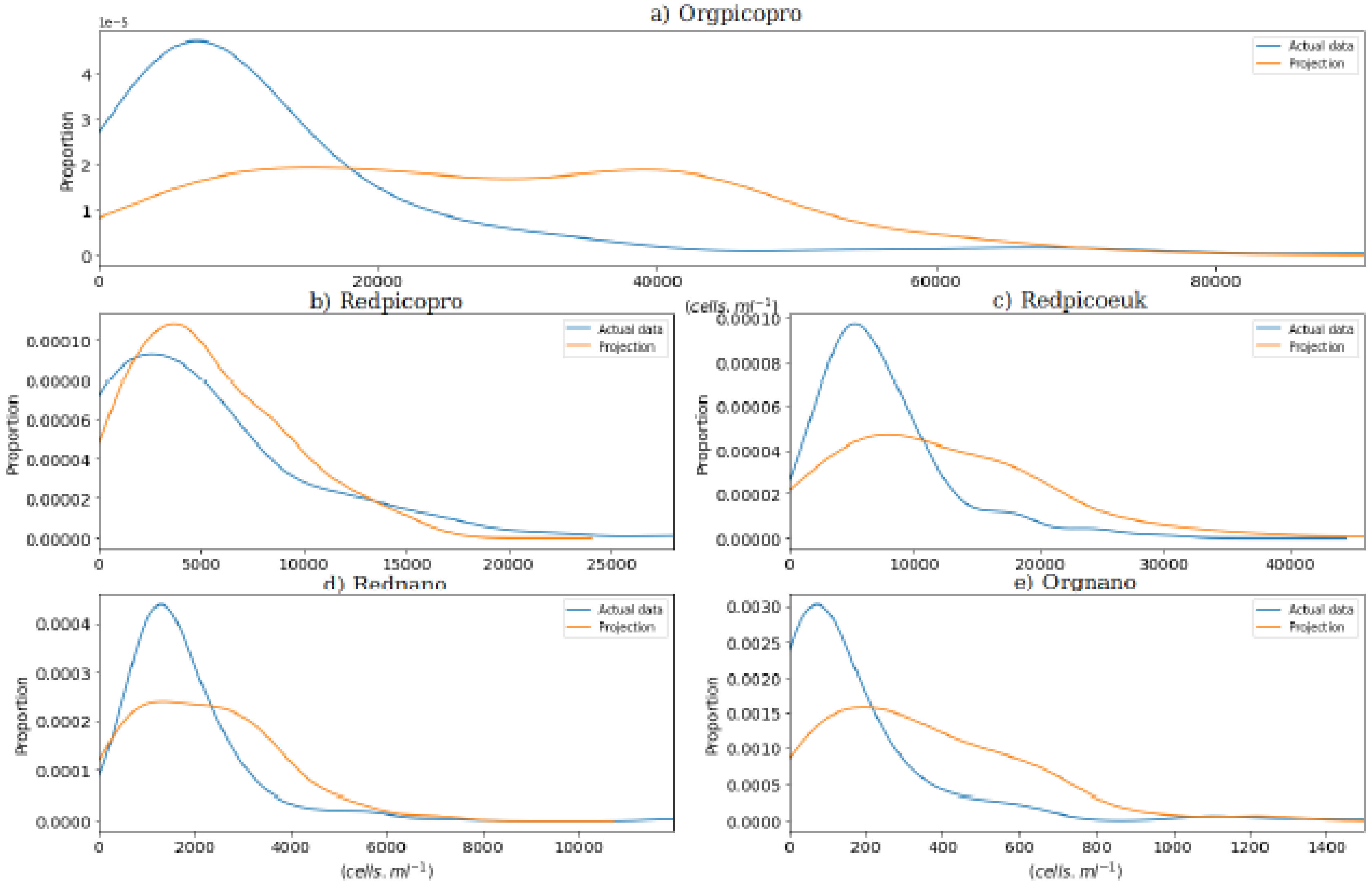
Recreating the missing people

MIAMI: Prospective

Future Mediterranean temperature rising +2°C

READING

- Siginificant increase in abundances for all phytoplankton functional groups (PFGs), except Redpicopro (non-significant).
- Orgpicopro and the Orgnano: +52%
- Redpicoeuk abundance: +39%.
- In general, simulated distributions flatter than for the actual data.



Distribution of the functional group abundances in the actual SOMLIT data and for a simulated increase in water temperature by 2°C in winter ($n = 180$ in both cases). The distribution of the data is shown for the Orgpicopro (a), Redpicopro (b), Redpicoeuk (c), Rednano (d), and Orgnano (e). The mean of each cPFG actual and simulated distributions are significantly different (Bonferroni-corrected Student-Welch test, $p<0.01$).

Yes, but...

Depths of the different layers not always known beforehand

Contrary to SOMLIT data, during cruises one need to know the depth of the epipelagic layer

Existing methods

are usually based on a simple variable thresholds:
Ex: 1% of surface PAR, -0.5°C w.r.t. surface temperature, change in density of 0.125 kg m-3 wrt to the surface

Current method limitations

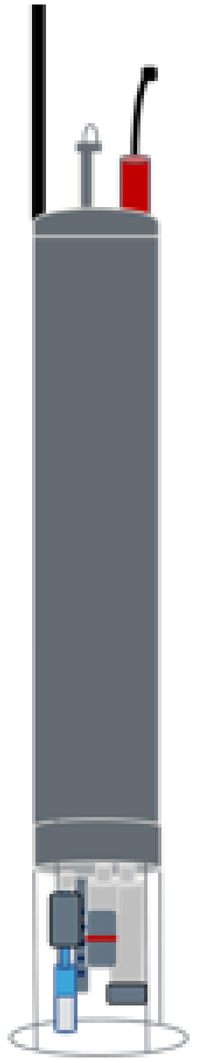
- Sensitive to outliers
- The value of the thresholds lacks foundations
- Choice of the variable (temperature, density, PAR): capture different patterns

RUBALIZ: Epipelagic/Mesopelagic boundaries

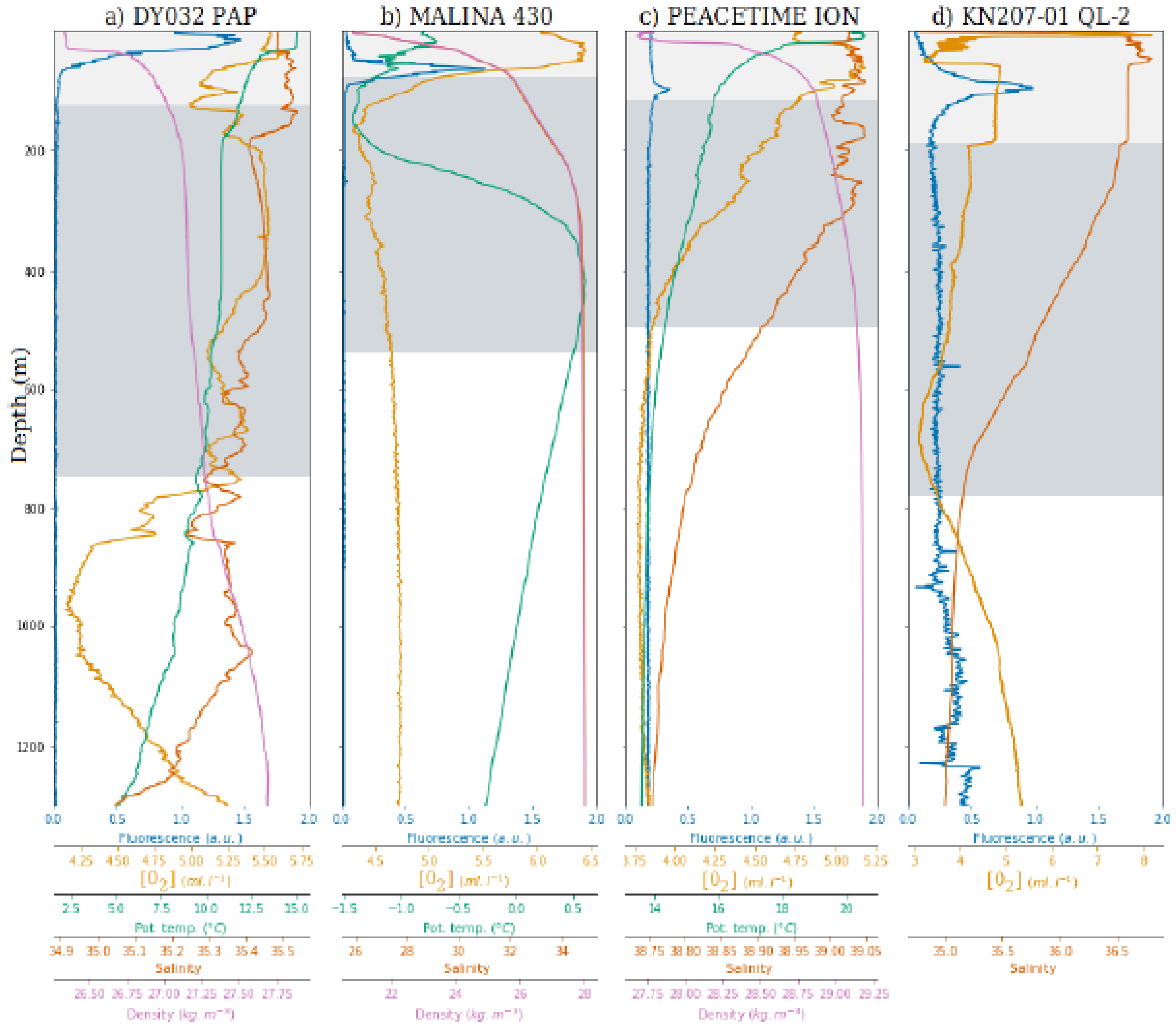
RUBALIZ: A RUpture-Based detection method for the Active mesopeLagIc Zone

Main features

- Use 5 variables characterizing the water column: fluorescence, oxygen, potential temperature, salinity and density collected by CTD
- Indicate the contribution of each variable
- Look for rupture in the signal
- Can take several CTD-cast to avoid individual sensor failure/noise



CTD



Epipelagic and mesopelagic layers found by RUBALIZ for four stations

Fuchs, Baumas et al. (accepted)

RUBALIZ: Epipelagic/Mesopelagic boundaries

RUBALIZ: A RUpture-Based detection method for the Active mesopeLagIc Zone

How does it works?

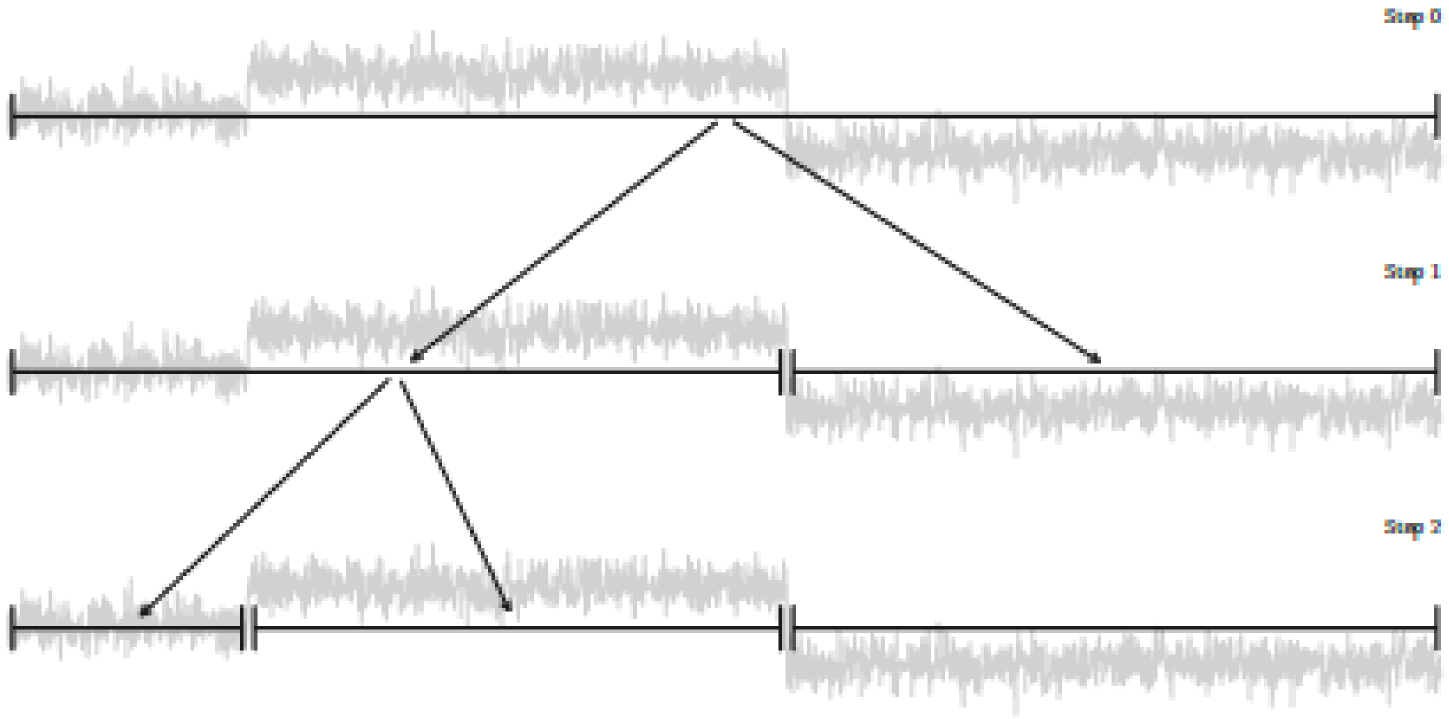
- Each variable is normalized
- Epipelagic boundary is determined, then mesopelagic boundary is estimated.
- Cost function: reproducing kernel Hilbert space (rkhs)
- Optimization method: Binary segmentation (Binseg)

Cost function (Fuchs, Baumas et al., accepted)

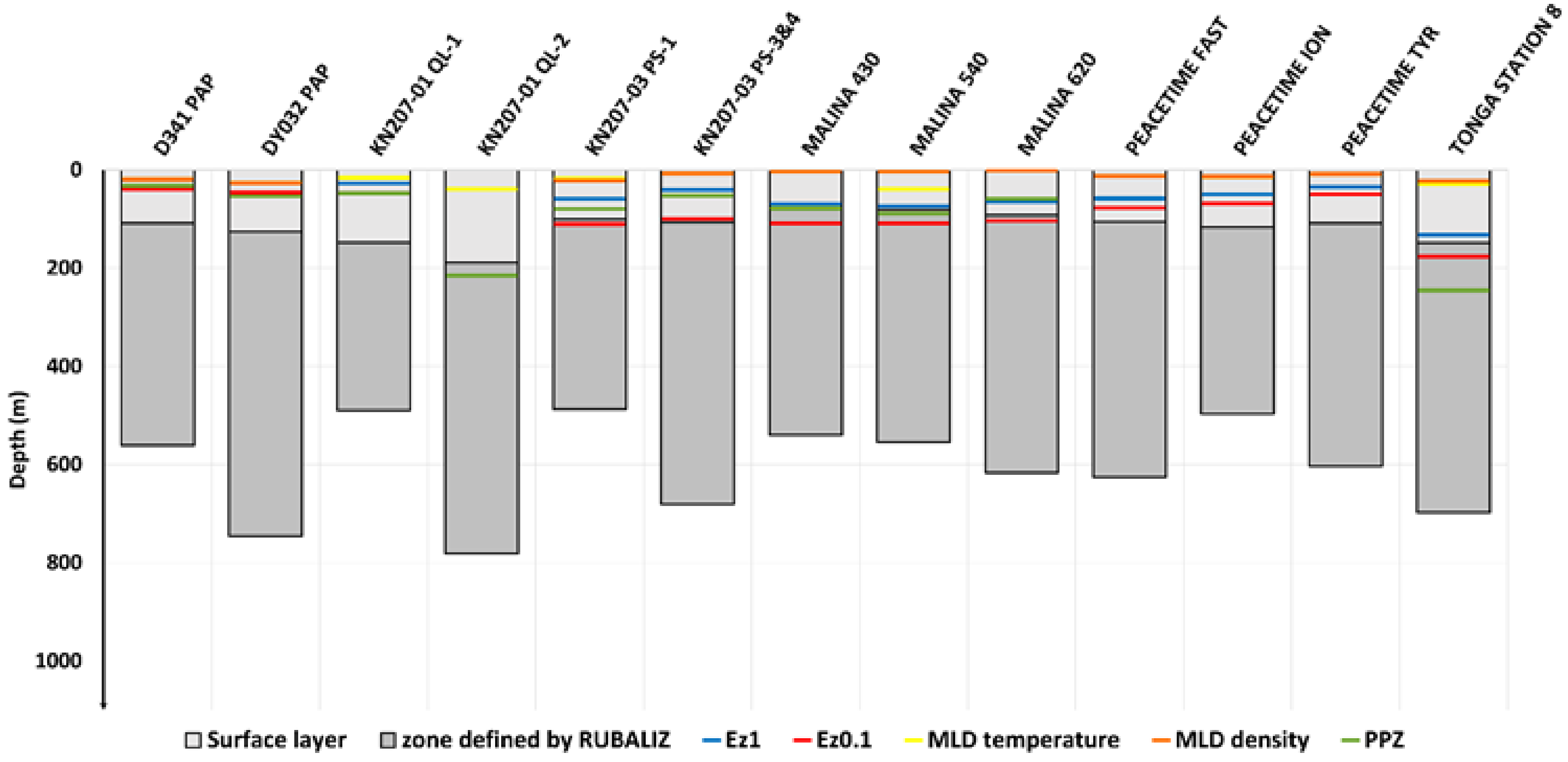
$$c_{kernel}(y_{a..b}) := \sum_{z=a}^{b} \| \phi(y_z) - \bar{\mu}_{a..b} \|_{H}^{2}{}_{,H}$$

with $y_{a..b}$ the subsignal between depths a and b, $\bar{\mu}_{a..b}$ the mean of the embedded subsignal $\{\phi(y_z)\}_{z \in [a,b], \underline{z} \leq a \leq b \leq \bar{z}}$, and $\| . \|_H^2$ as defined in (1).

Binseg (Truong et al. 2020)

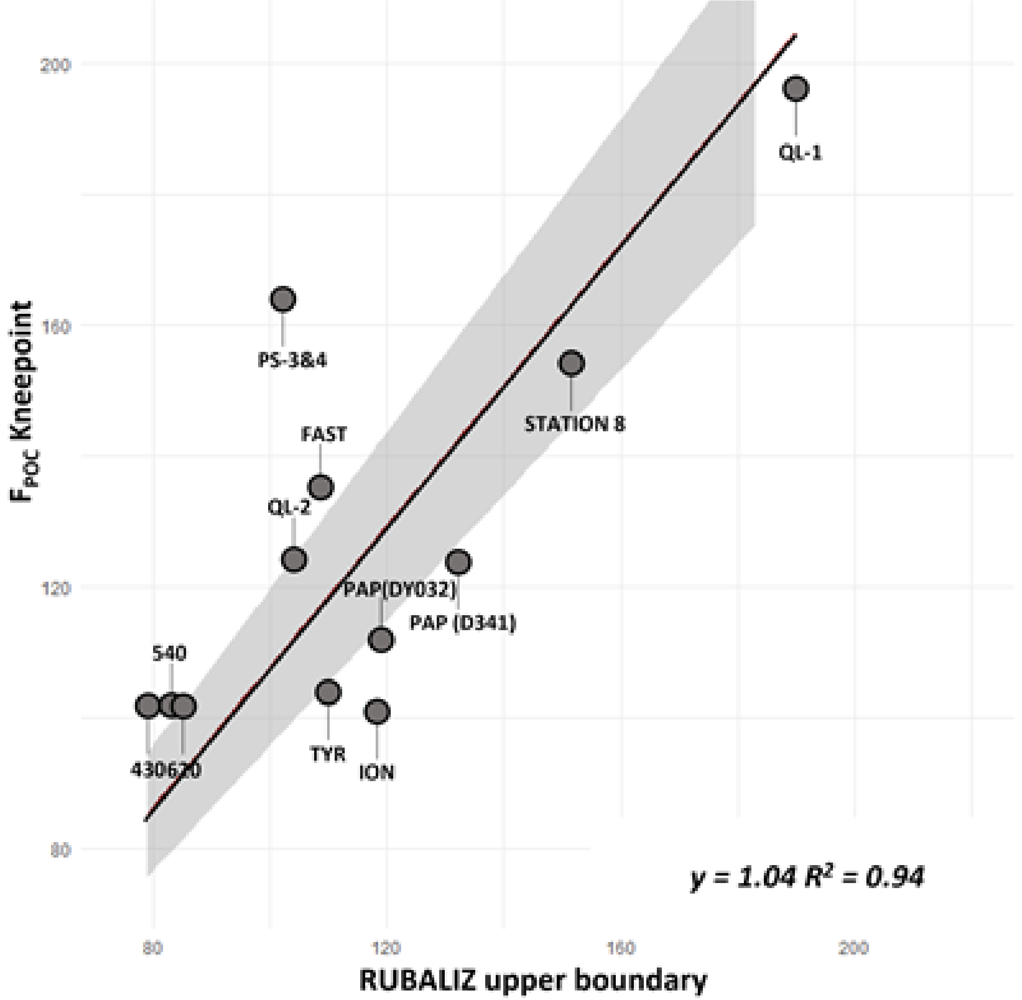


Epipelagic boundaries shallower than 200m deep



Capture well the carbon supply flux

Inflexion in the Particular Organic Carbon match the epipelagic boundary found by RUBALIZ



Part I/ Wrap up

Vertical boundaries

- Shallower than 200m
- Strong physics/biology coupling

Ecological niches

- Most variability comes from temporal and spatial variability
- Opposite ecological niches of Orgpicopro and Redpicoeuk in the North-Western Mediterranean Sea

Effect of water warming

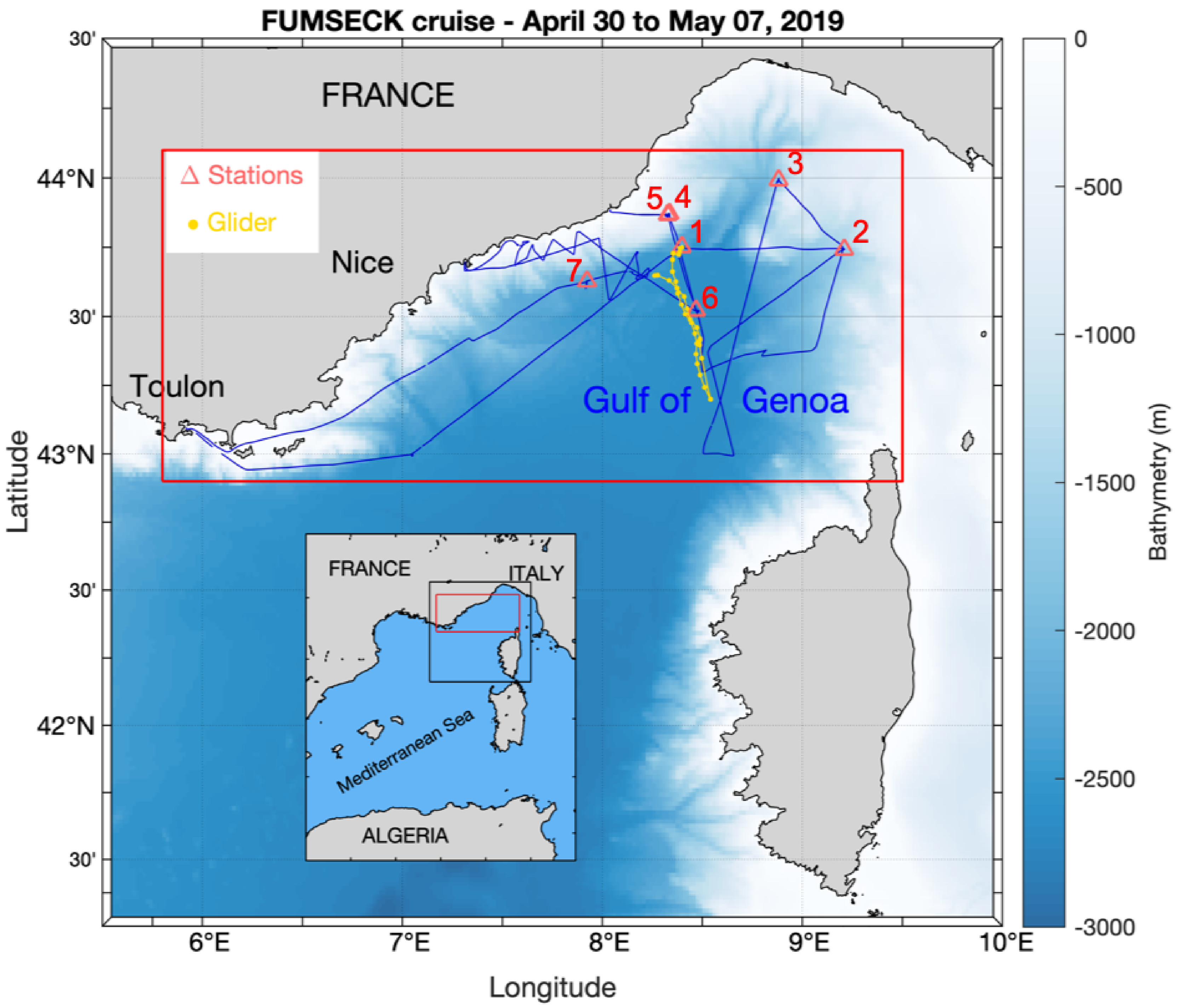
- Most phytoplankton groups might take advantage of higher temperature
- Yet, global change effects are not limited to increasing temperature

Part II:
High-frequency impact of wind-induced events

FUMSECK cruise: Example of wind-induced event effects

May 2019

- In the Ligurian Sea
- Coupling Physics/Biology with multiple *in situ* sensors (e.g. ADCP, flow cytometer, MVP, glider), satellite data and 3D modeling
- A storm happened on May 5, 2019 at night



Cruise map (Barrillon et al., submitted)

Example of wind-induced effects

Physics

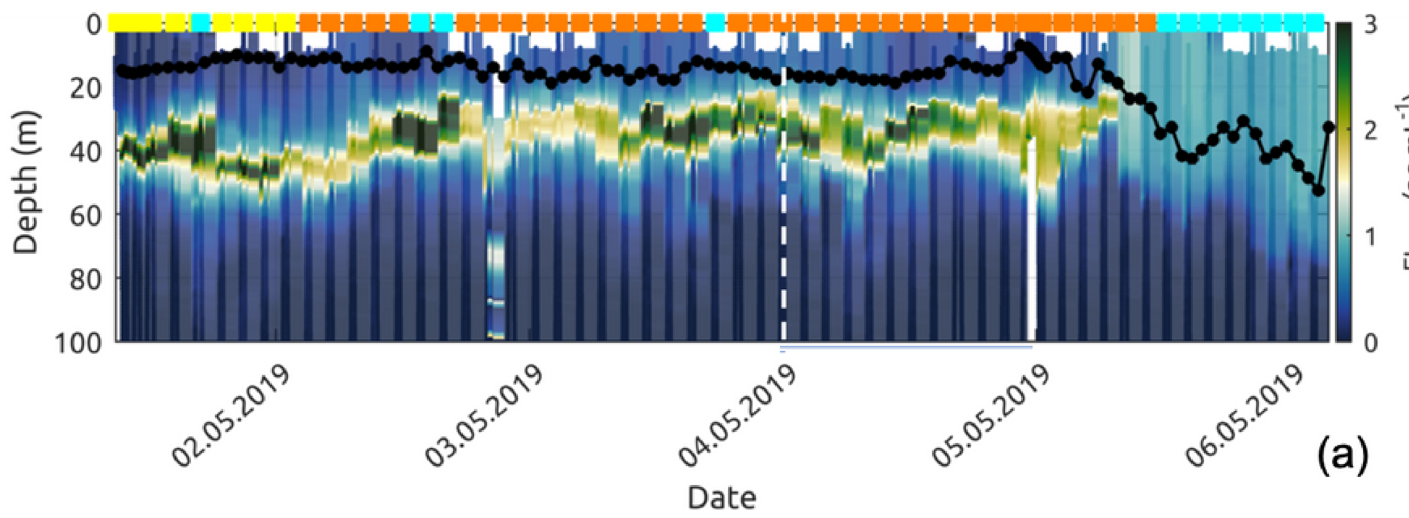
- Upwelling of deep and colder waters
- Deep Chlorophyll Maxima Dilution
- Deepening of the mixed layer depth from 15m to 50m

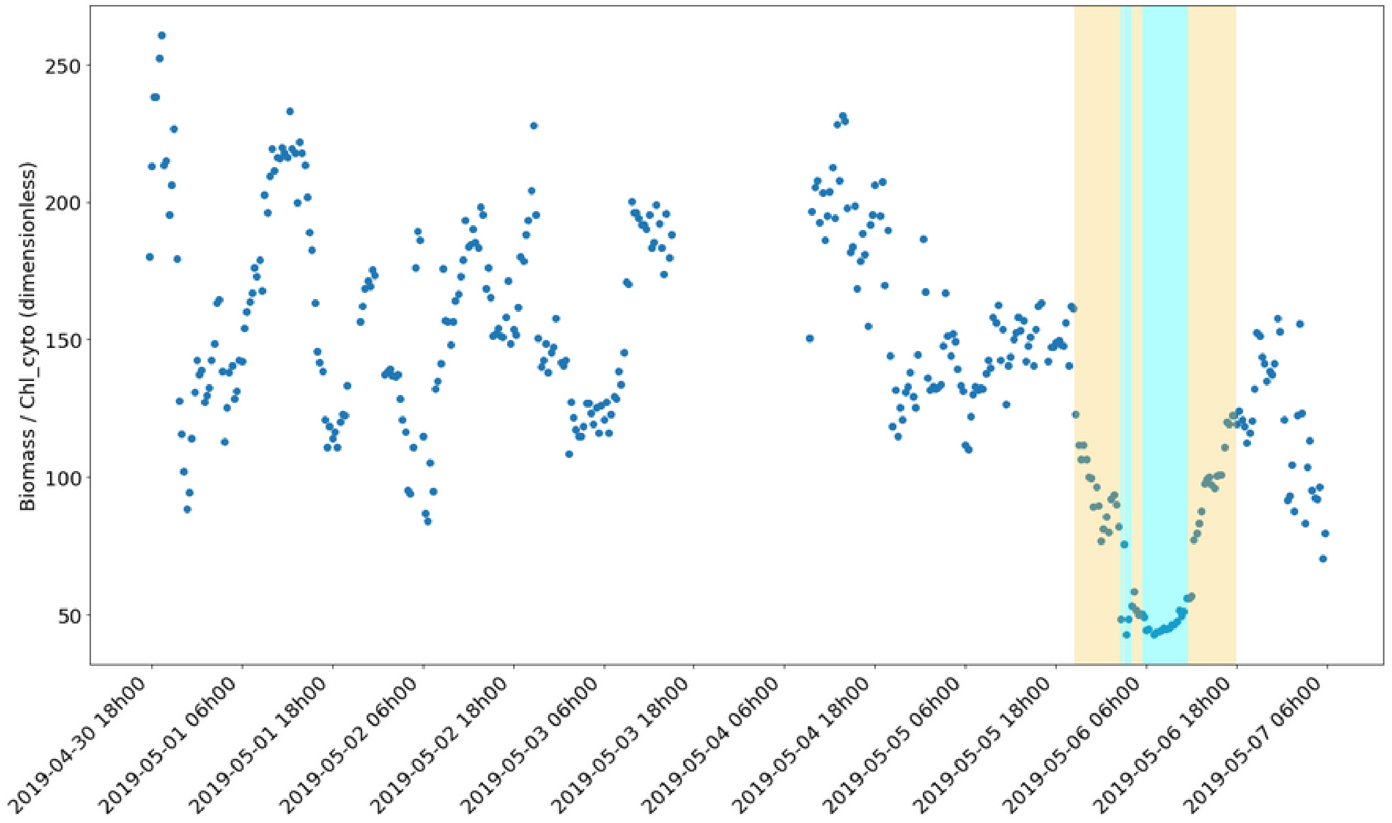
Biochimics

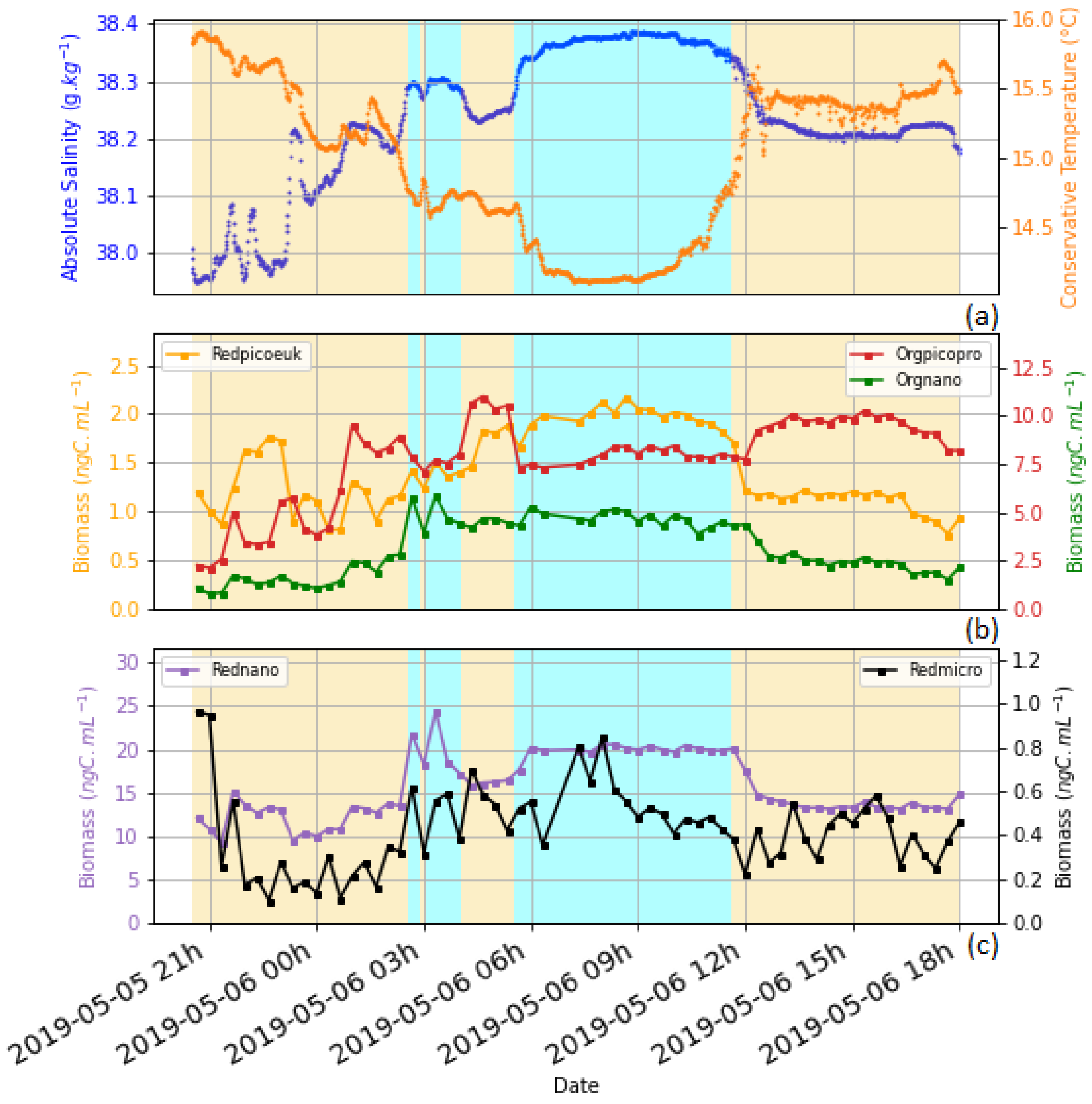
- Nitrate concentration x2
- Surface chlorophyll-a concentration x2

Phytoplankton

- Most groups abundance x2
- Cell Carbon/Chlorophyll ratio/2







High-frequency impact

AFCM useful for long-term high-frequency observation

Yes, but...

Standardization of the nomenclature

Thyssen et al. (submitted): In progress

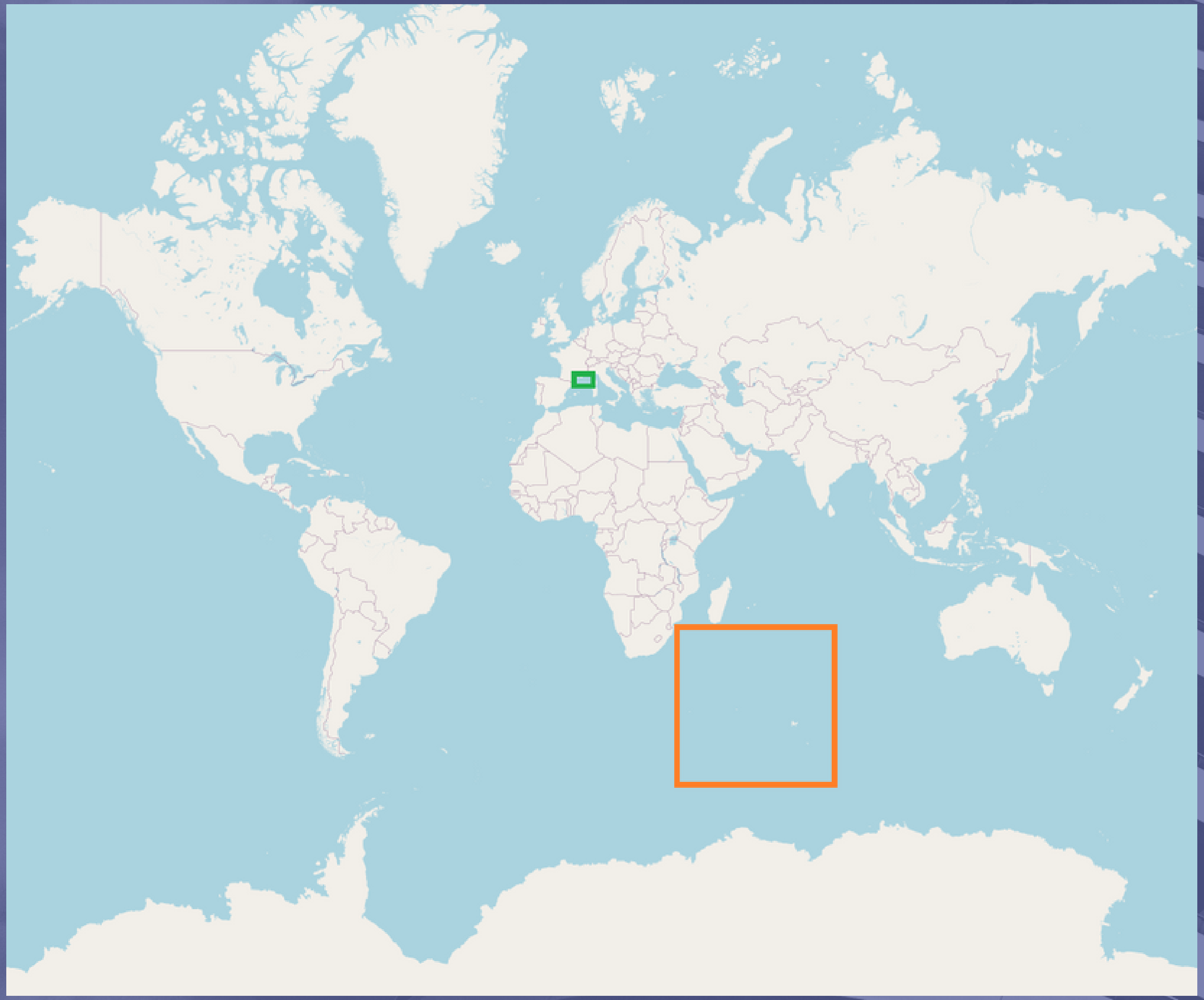
Manual gating errors

Need to be better assessed.
Small literature:
Garcia et al. (2014)
Wacquet et al. (unpublished)

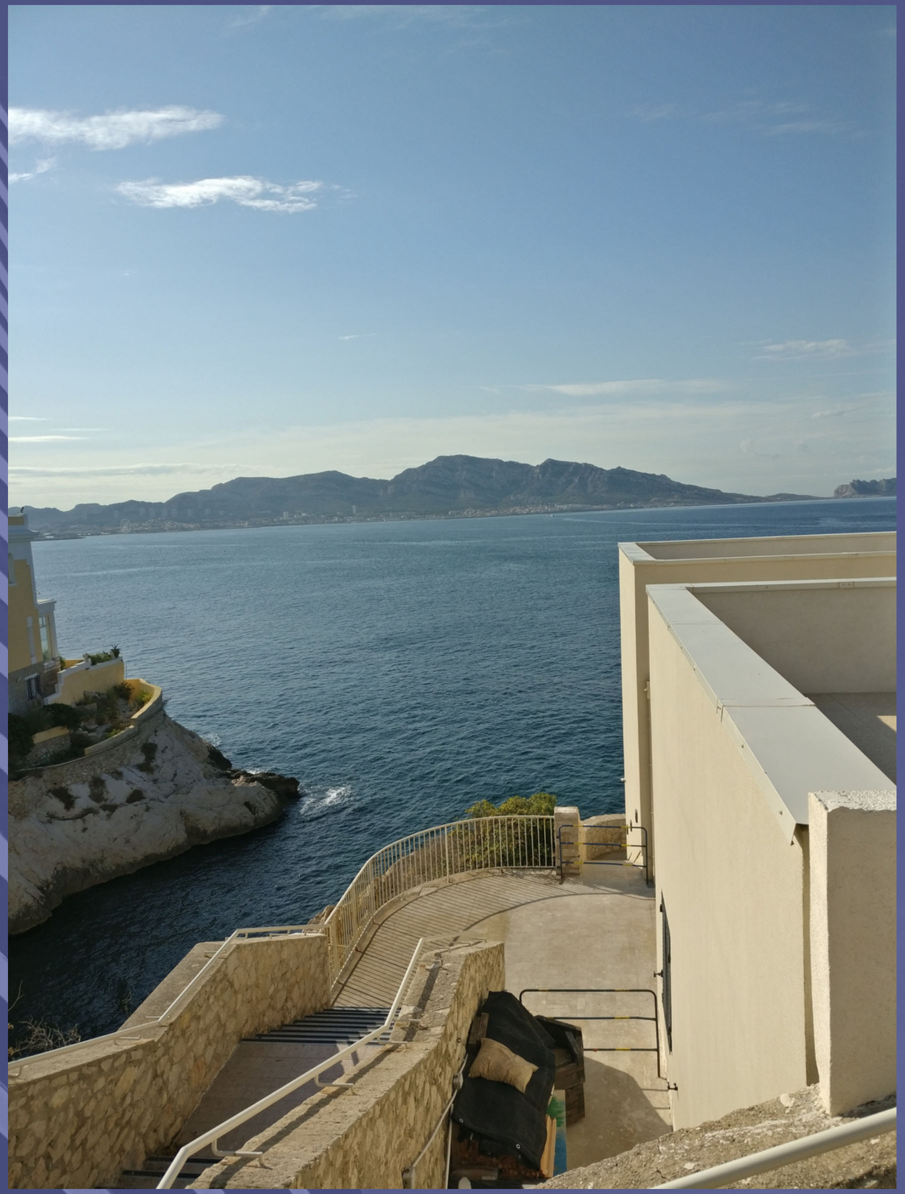
Need for reliable automatic methods

Convolutional neural networks compared with existing methods

Dataset locations



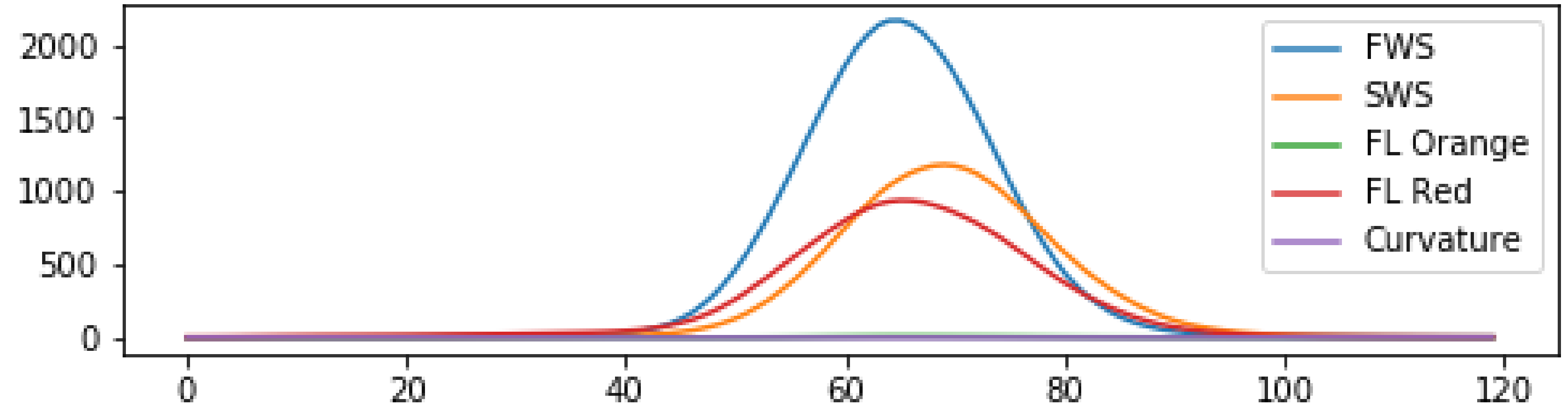
Caption:
SSL@MM station (Green)
GEOTRACES SWINGS (Orange)



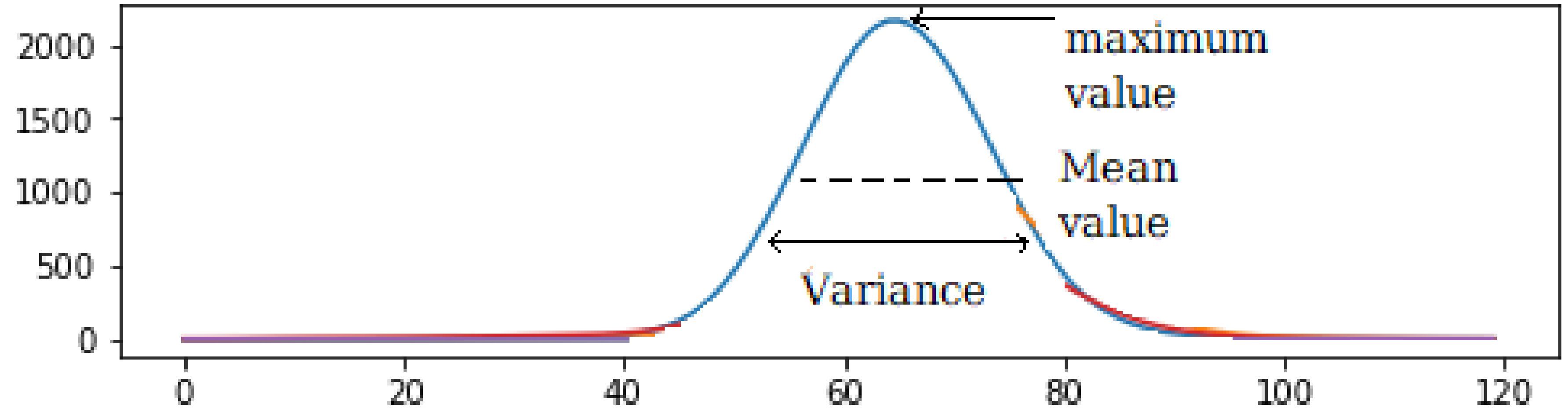
Caption:
SSL@MM station

Until now: Manual gating

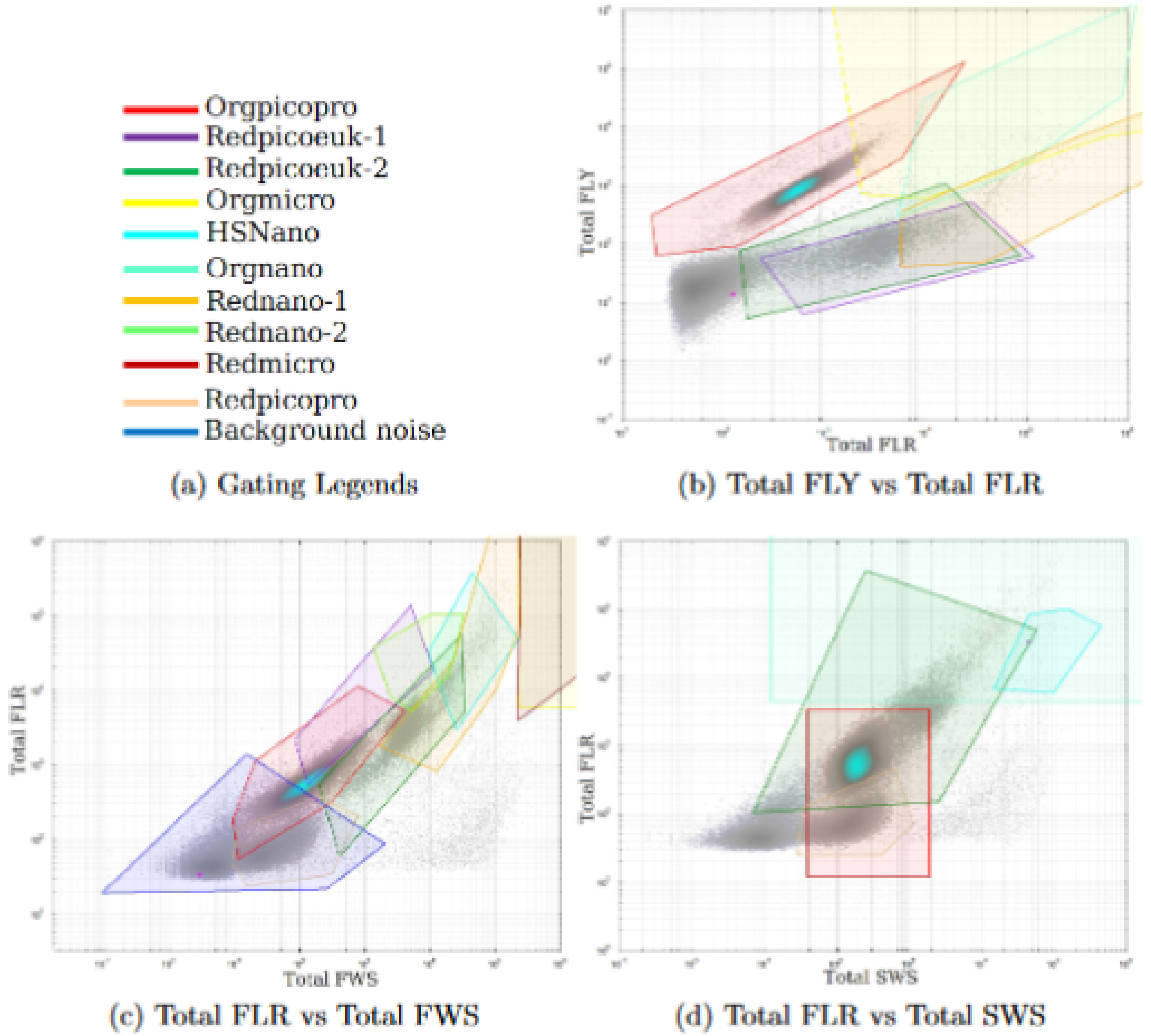
1) Five pulse shapes per cell



2) Compute pulse shapes descriptors



3) Plot the descriptors and draw group borders (gates)



(a) Gating Legends

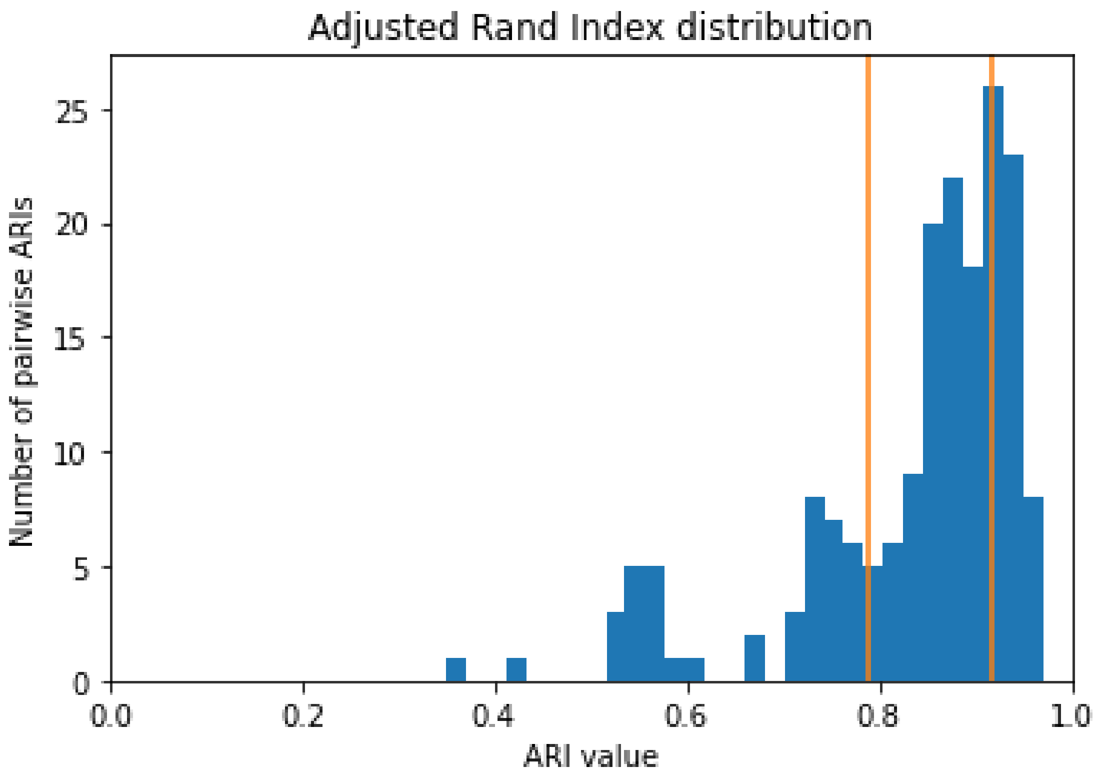
(b) Total FLY vs Total FLR

(c) Total FLR vs Total FWS

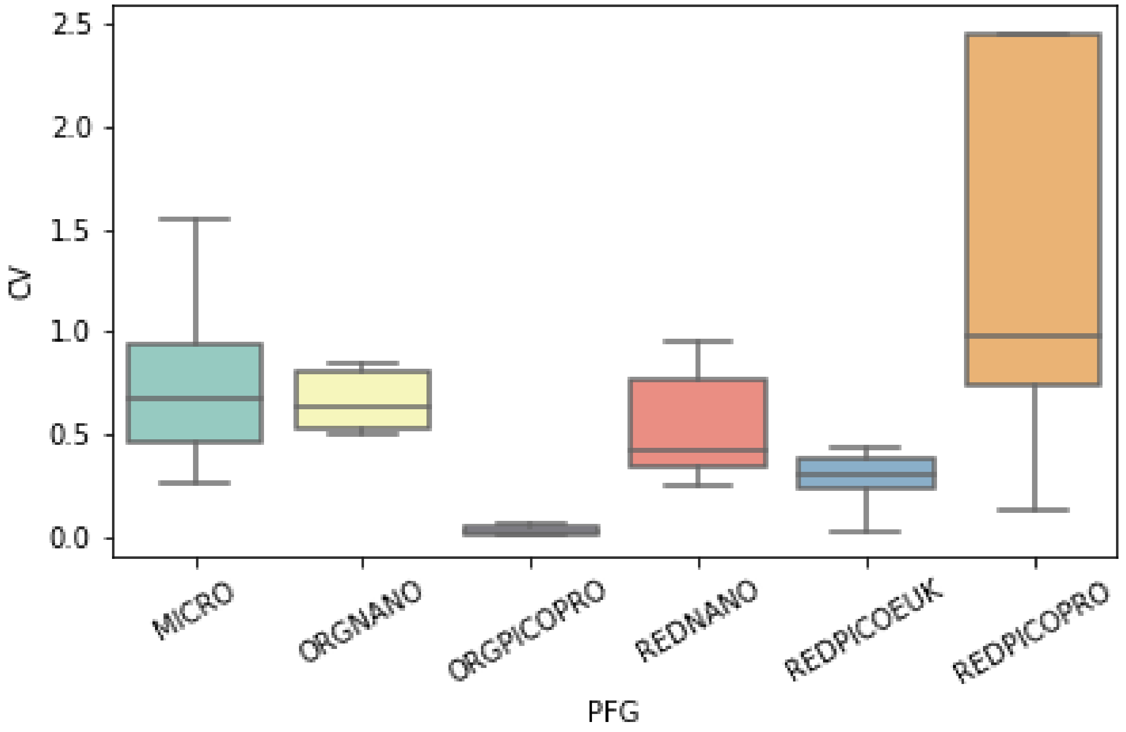
(d) Total FLR vs Total SWS

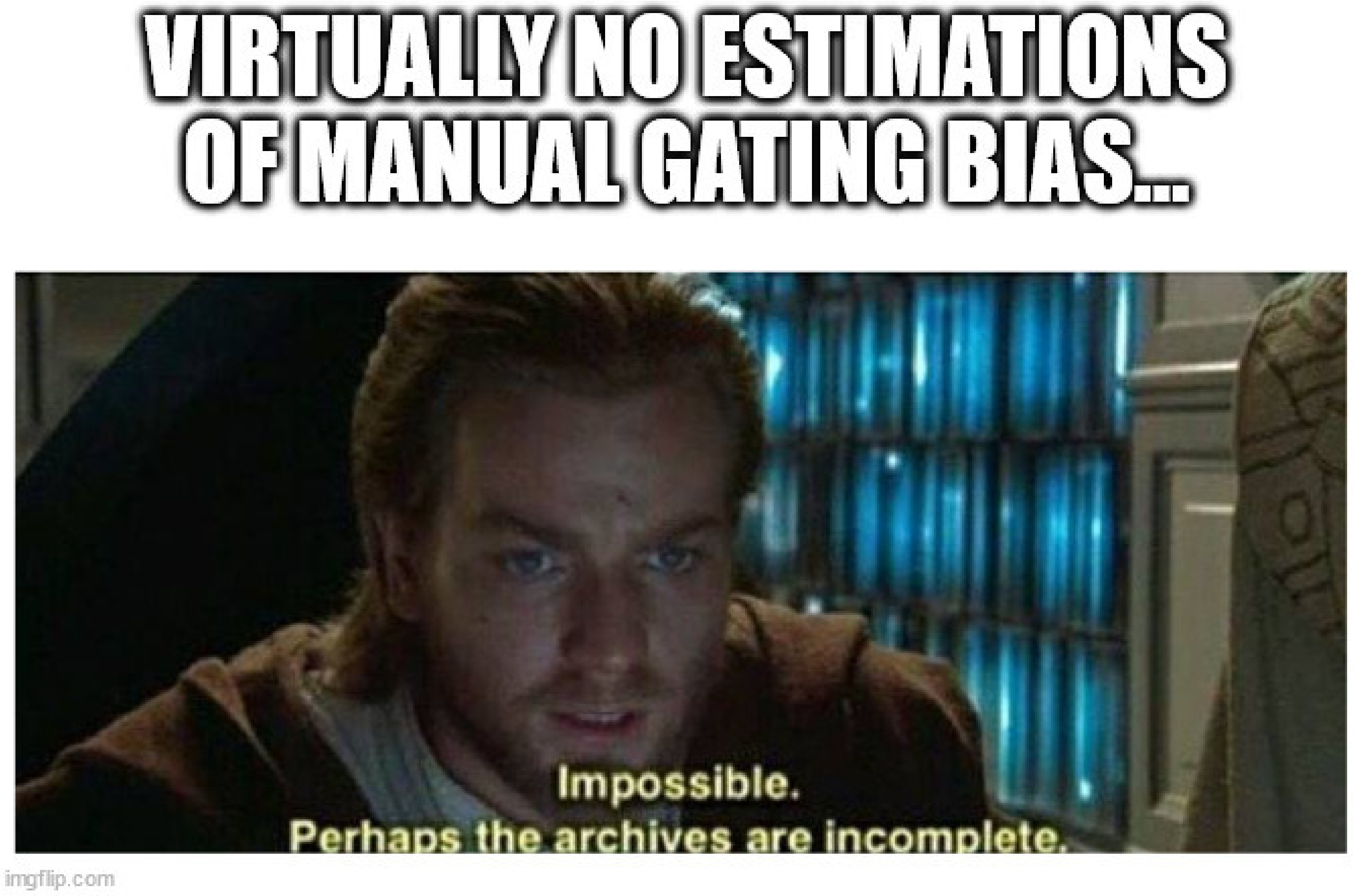
Manual bias estimation

Reading:
The closest the ARI is to 1 the more the experts agree on the gating process

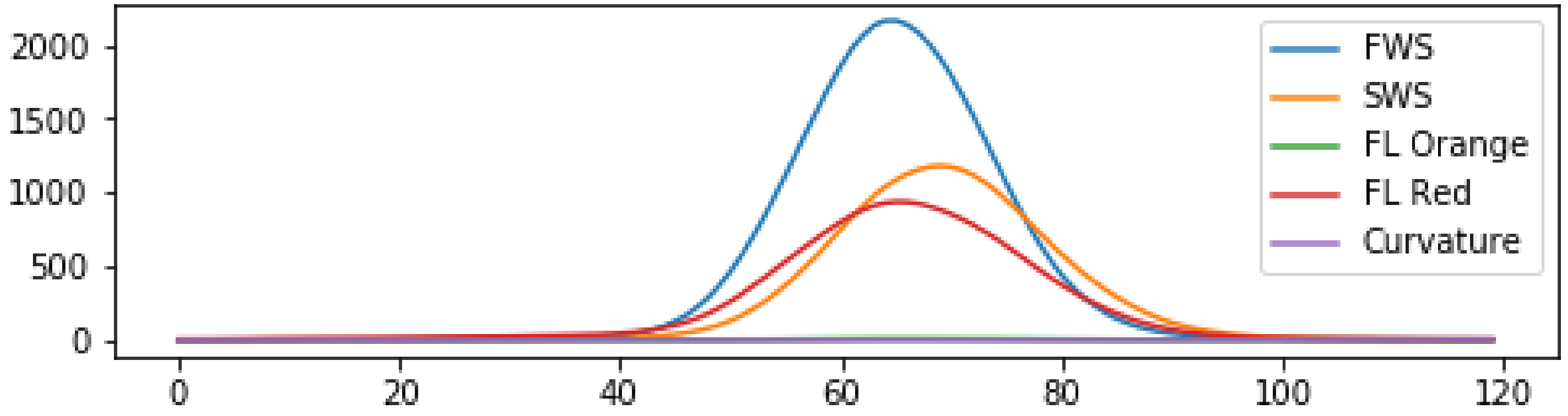


Reading:
CV>1
<=>
Count standard errors between experts > Mean expert counts

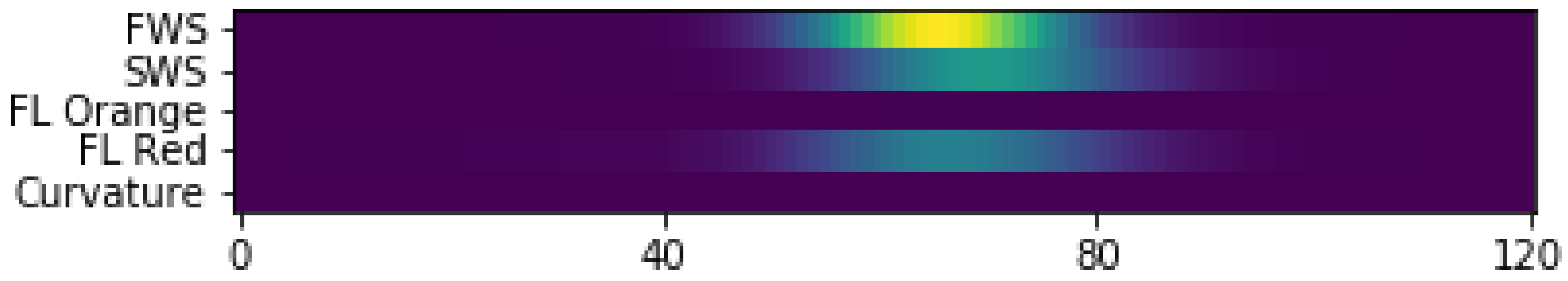




Five curves as an image



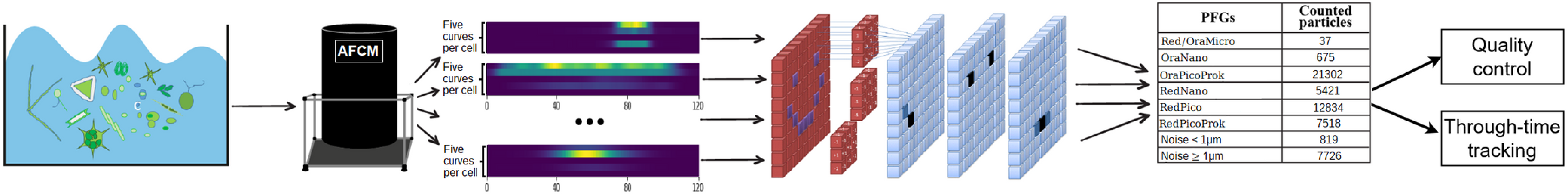
FWS	0	0	⋯	1870,1	1950,4	2103,6	1972,2	⋯	0
SWS	0	0	⋯	1023,7	1165,8	1094,1	1013,9	⋯	0
FL Orange	0	0	⋯	898,1	1007,5	879,0	838,2	⋯	0
FL Red	0	0	⋯	⋯	⋯	⋯	⋯	⋯	0
Curvature	0	0	⋯	⋯	⋯	⋯	⋯	⋯	0



$:=$

From now:

Automatic gating with Deep Convolutional neural networks



PFGs	Counted particles
Red/OraMicro	37
OraNano	675
OraPicoProk	21302
RedNano	5421
RedPico	12834
RedPicoProk	7518
Noise < 1µm	819
Noise ≥ 1µm	7726

Prediction workflow from seawater to final data (Fuchs et al. 2022)

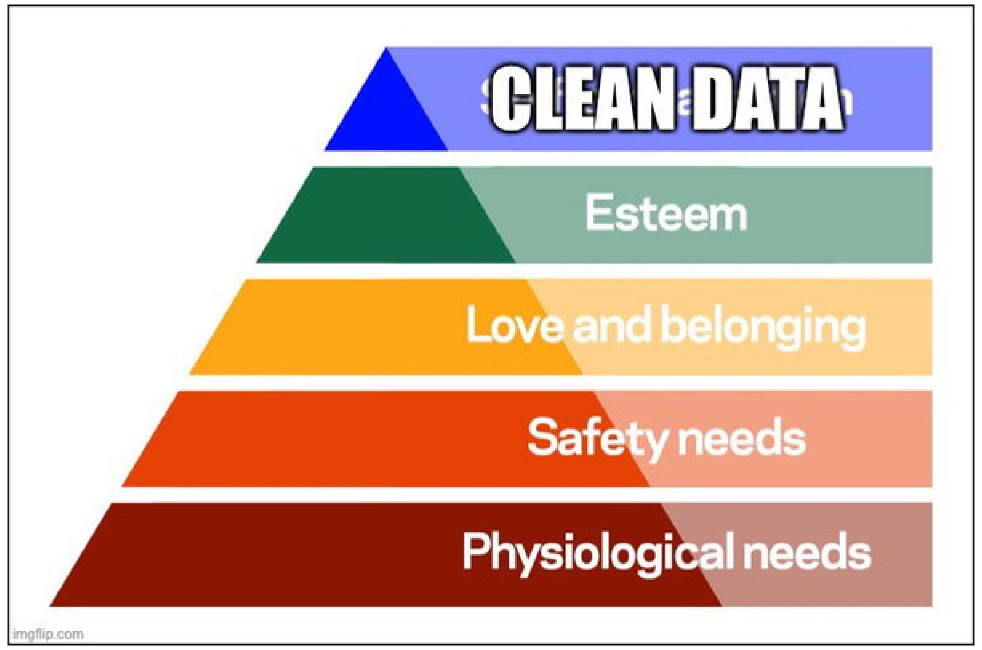
Do you need Deep Learning for this?

CNN

Neural networks for image classification

Data

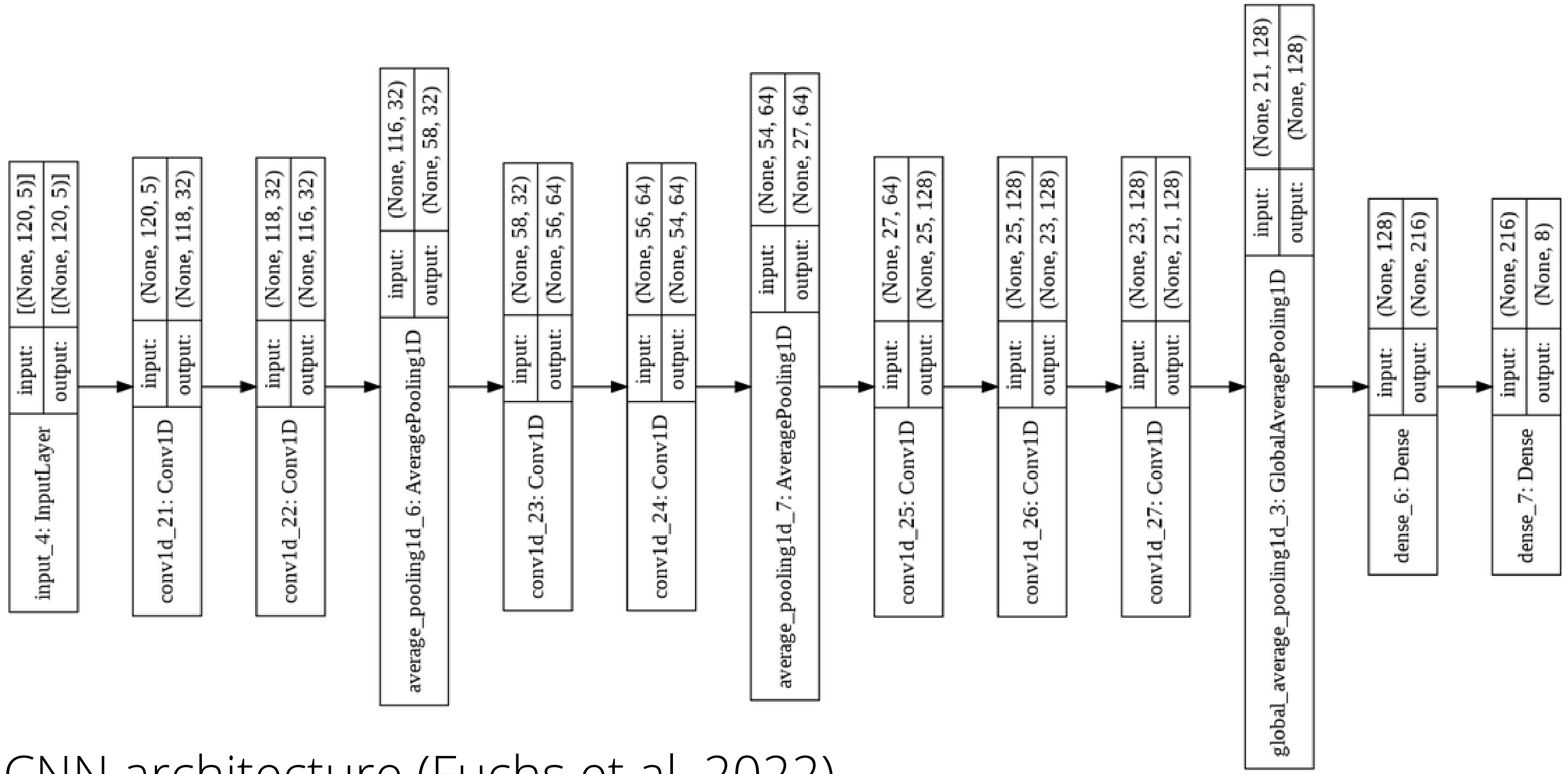
- 56 acquisition files
- Keep only inter-expert consensual particles.
- Get a similar number of observations for each group
- ~50 000 observations in the training set: medium size dataset



Maslow revisited

Model

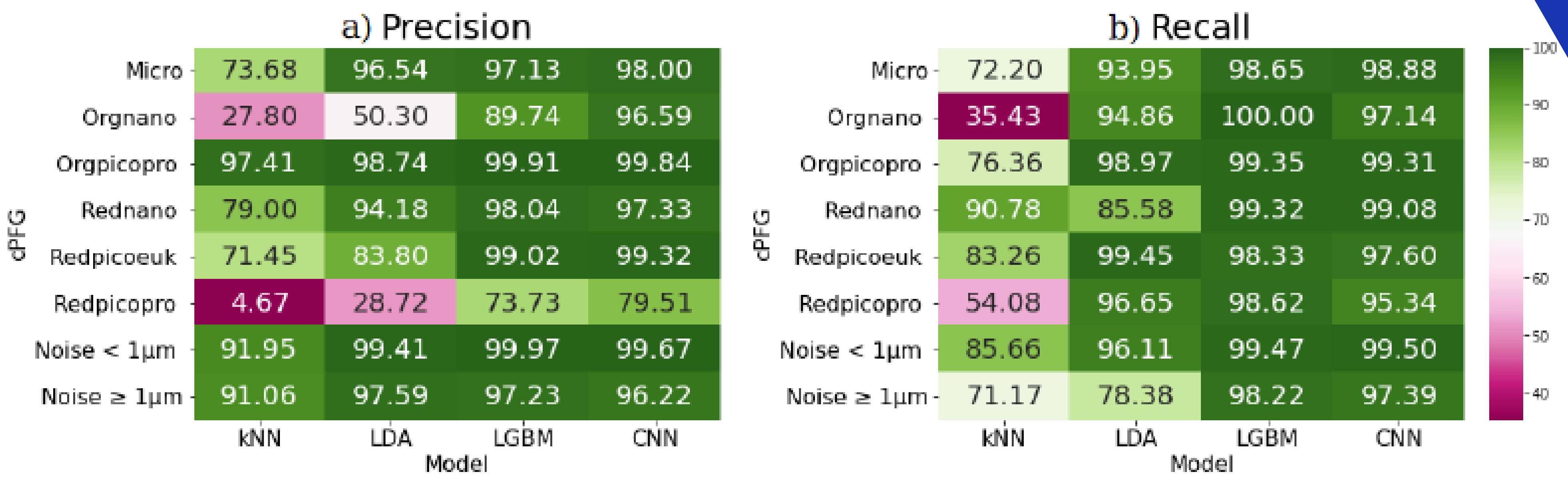
- VGG-inspired architecture (Kingma et al. 2014)
- Ranger optimizer (Yong et al. 2020)
- Tuning of hyper-parameters using Bayesian hyperoptimisation (Bergstra et al. 2013)
- No other losses seem to beat the categorial cross-entropy (Focal loss, class balanced loss)



CNN architecture (Fuchs et al. 2022)

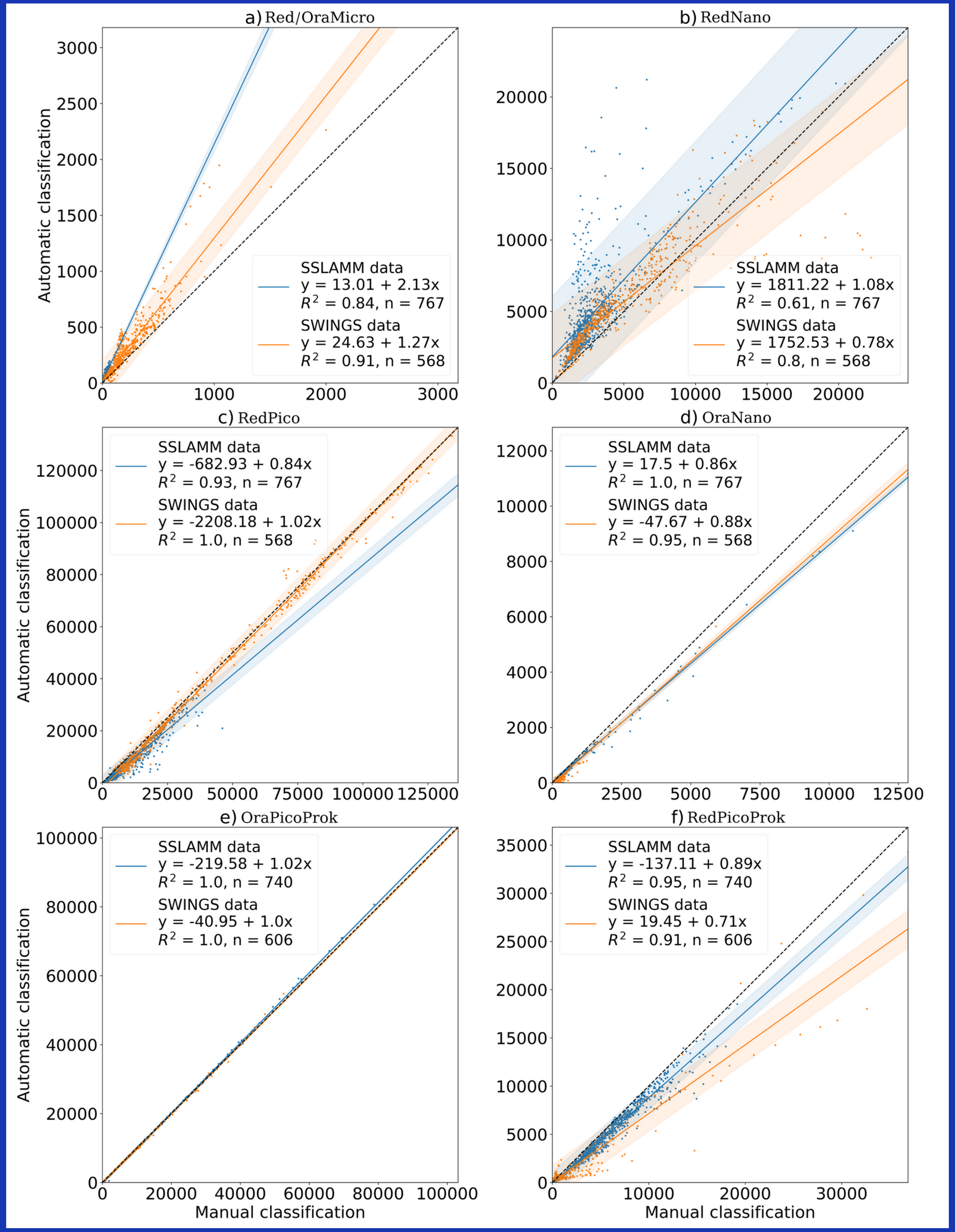
Performances

- State-of-the-art performance (main challenger: LGBM)
- Manual and automatic (CNN) gatings match for most groups on both SSL@MM and SWINGS data



Precision and recall of benchmark models for each functional group (SSL@MM data) Fuchs et al. (2022)

Comparing automatic and manual gating

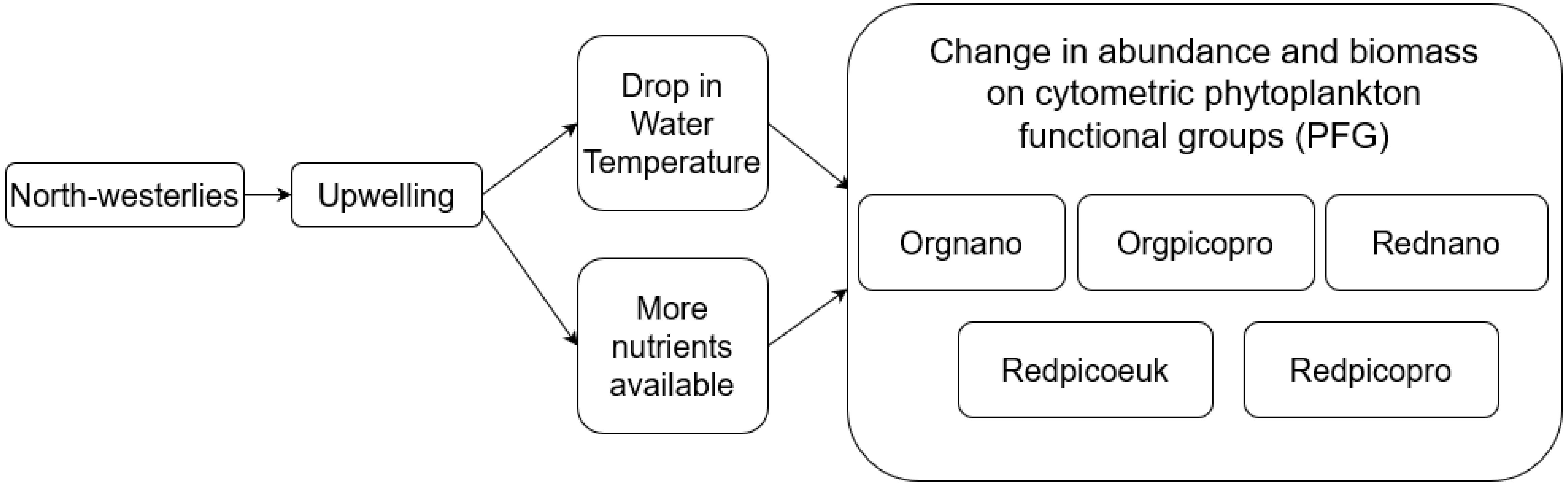


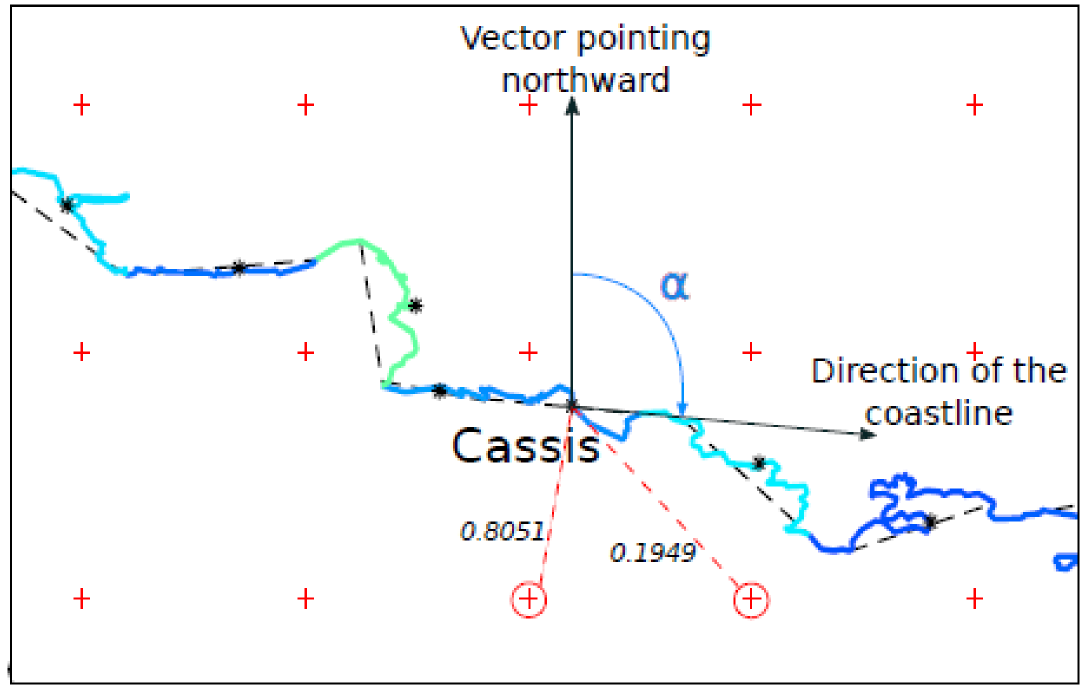
And now:

Estimating reproducible wind-induced phytoplankton changes

SSLAMM station

- From September 2019 to November 2021
- Focus on stratified periods: End of May to early November
- Rupture detection method
- Twenty events identified by temperature anomaly

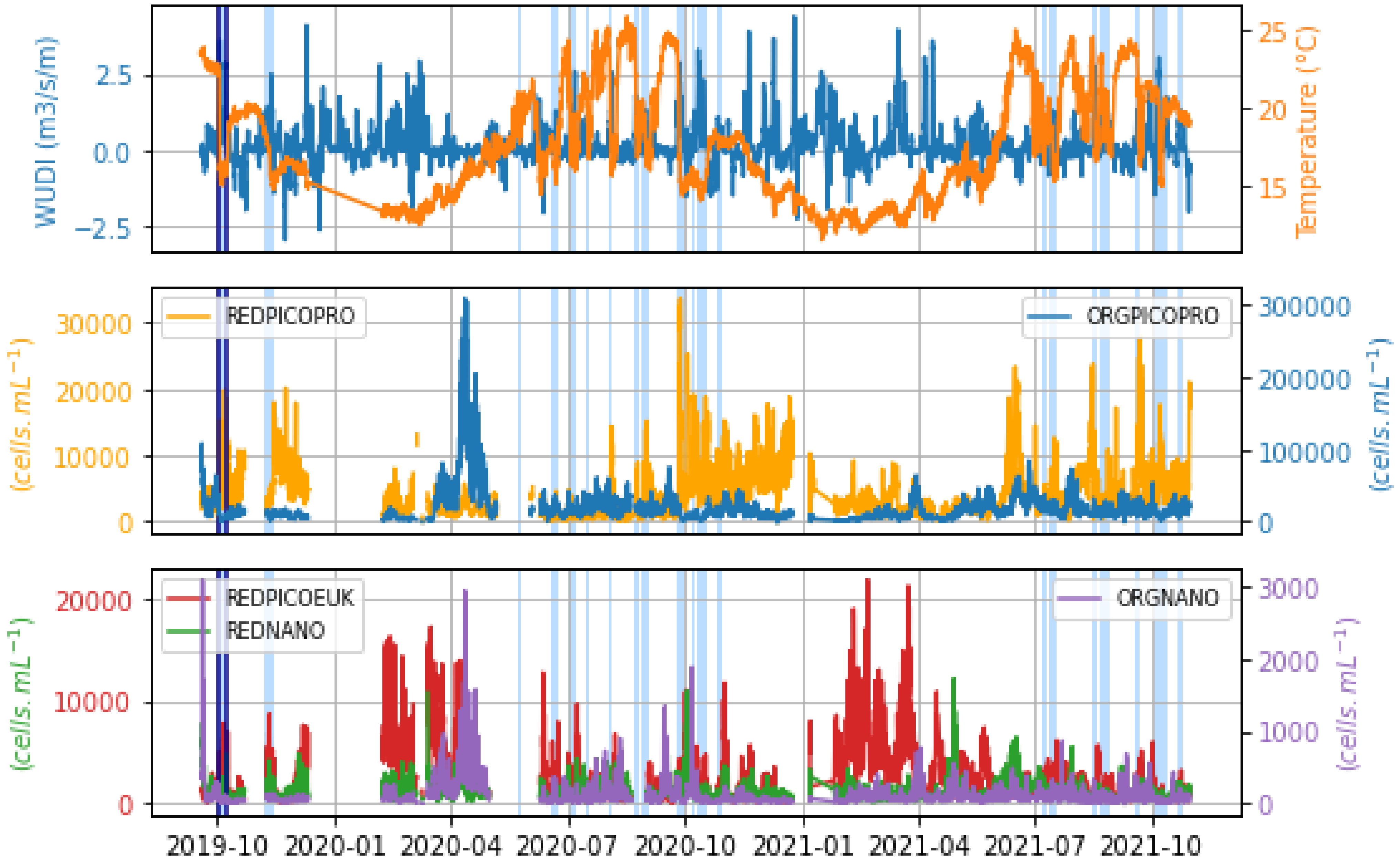




Wind Upwelling/ Downwelling Index (WUDI), Odic et al. (accepted)

Method outline

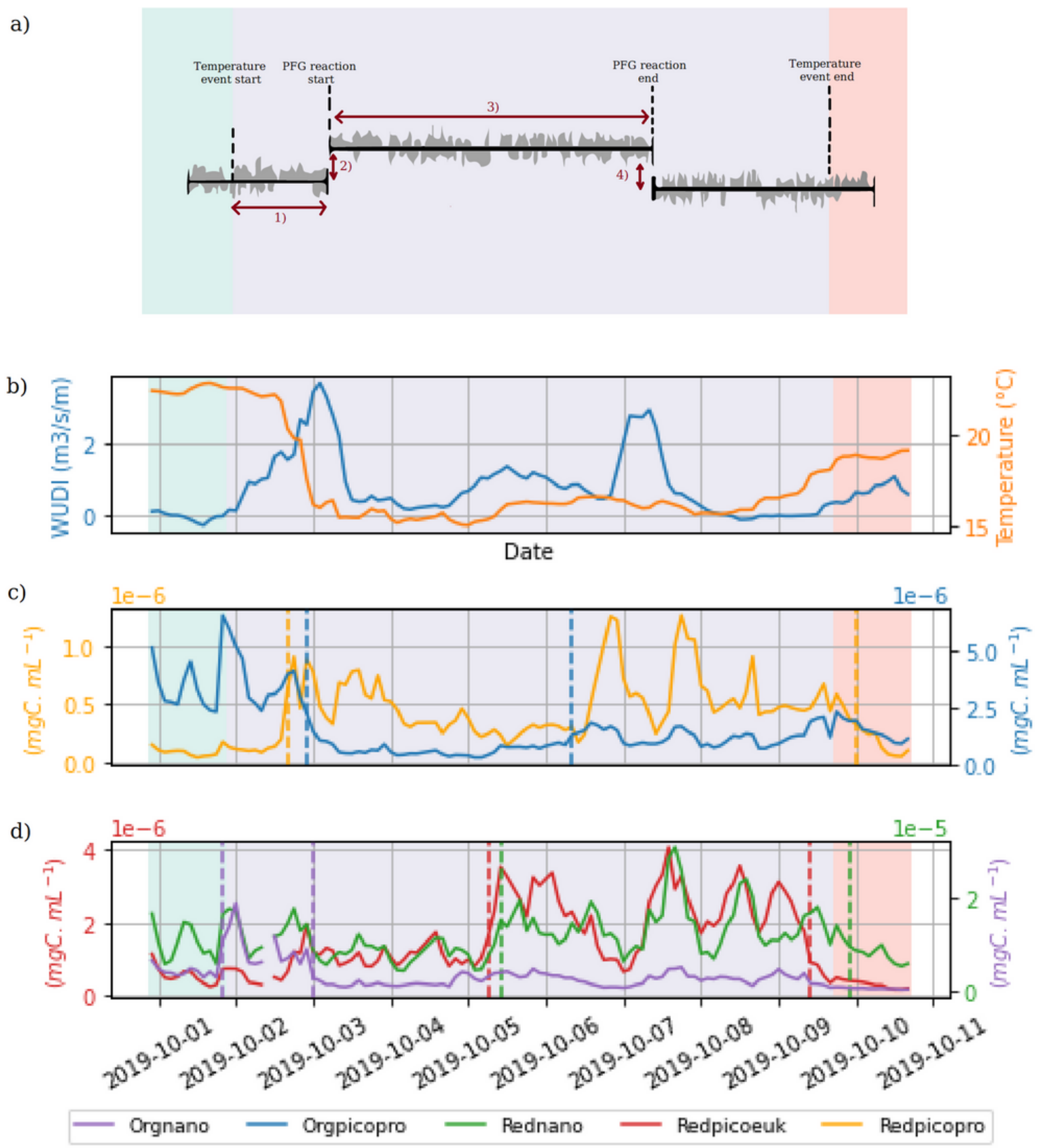
Full time series



- Blue spans: upwelling in stratified period
- First subplot: WUDI index, water temperature
- 2nd et 3rd subplots: Phytoplankton functional groups (PFG) abundances (concentrations)

Fuchs et al. (submitted)

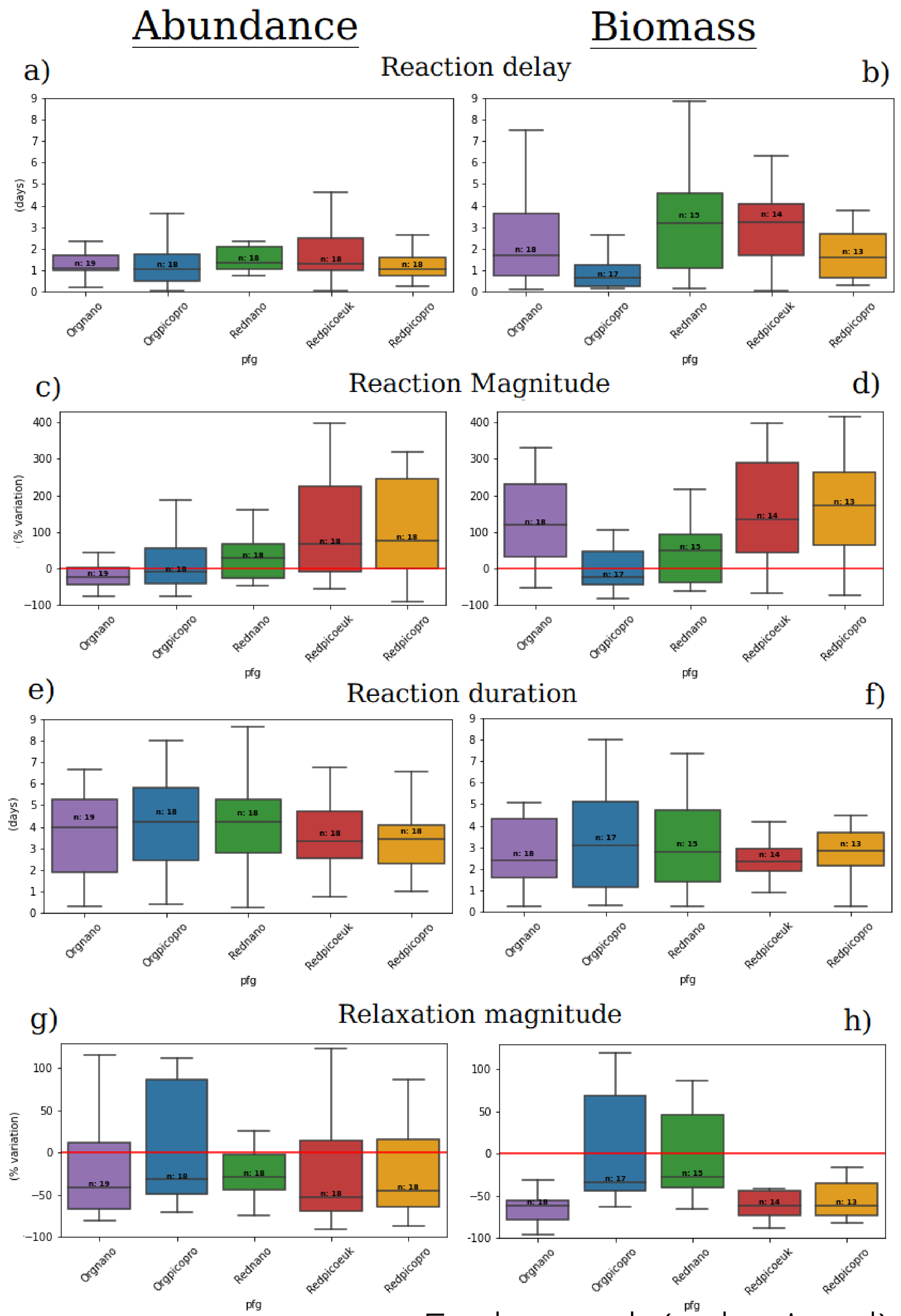
Zoom on an event



- Identify: Delay before reaction, reaction magnitude and length, and relaxation magnitude
- Dashed lines: Identified reaction of each PFG

Results

- Biomass peaks and daily rates of increase induced by the most **extreme upwellings** are of the **same magnitude** as the **spring bloom** ones.
- Phytoplankton abundance/biomass **reactions start less than 2 days/4 days** after the upwelling onset and **last 2 to 5 days.**
- During upwelling events **all biomasses** (but Orgpicopro) median/maximum **increases range** 50-173/100-400%, **then** sharply **drop back** to normal.



Fuchs et al. (submitted)

Conclusion

Main contributions

MDGMM

- **Flexible** model thanks to its mixture structure
- Simple link functions + graphical utilities: **remains interpretable**
- Model selection and initialization procedure were designed
- Define horizontal and vertical local boundaries of phytoplankton ecological niches: spatio-temporal dependence matters
- Orgpicopro and Redpicoeuk have opposite ecological niches in the North-Western Mediterranean Sea

MIAMI

- Model-based data generation with the desired properties
- Temperature rise seems to favour most phytoplankton function groups

RUBALIZ

- The method is robust to outliers and defectuous CTD-casts
- It matches the variation of the Carbon supply
- It gives foundations to the lower boundary of the epipelagic zone: the classical 200m boundary was too deep for the thirteen station tested

Main contributions

FUMSECK

- The sensors well described the physics/biology coupling: Rare and insightful observations
- Wind-induced events triggered drop in temperature, higher nitrate and higher phytoplankton groups (except for the Orgpicopro)

Automatic gating

- Manual gating errors are significant, especially for less represented groups (with CVs>100%)
- CNN obtained state-of-the-art performance and gated a file in less than a minute (faster than humans)
- Present good generalization properties from one place to another: Representative and consensual datasets (not shown)

20 wind events at the SSL@MM

- PFGs react in less than a week, and during less than a week: very fast changes
- The most extreme events have similar effects as bloom per unit of time
- Gives heuristics/function forms to integrate wind-induced events in biogeochemical models

Main limitations and axes of improvements

MDGMM

- GLLVM is rigid and costly in training time => Genetic Programming?
- Deep MDGMM architectures are instable without real performance gains
- Trade-off between the dimension reduction and clustering tasks

MIAMI

- Inherits from MDGMM limitations
- Brute force selection of interesting synthetic observations: could use Bayesian optimization

RUBALIZ

- Maximal ranges for the epipelagic zone and mesopelagic zone are defined beforehand by the user: these ranges could be specified in a Bayesian fashion
- The RUBALIZ approach is not sufficient to resolve the mesopelagic carbon budgets: over determinants exist such as Prokaryotic Growth Efficiencies or Leu-to-C conversion factors

FUMSECK

- Only one event for which the boat came back after the storm
- The boat left less than one day after the end of the storm: impossible to observe come-back-to-normal forces.
- Did not observe virus and zooplankton activity

Automatic gating

- Medium-size training set: Data augmentation to implement
- Do not use images taken by AFCM: valuable for biggest cells
- Prediction at the individual cell and not at the functional group level
- Successive acquisitions treated as independent
- Could use the CNN for biovolume and biomass predictions

Wind-induced events on PFGs

- Nutrients are collected at a low frequency: could use Ultraviolet Optical Sensors (nitrate) and Electrochemical Sensors (phosphate).
- Do not observed virus and zooplankton activity
- No history of where the water masses originated: could use HFRs or modeling

Core Team

Robin Fuchs

Denys Pommeret

Melilotus Thyssen

Special thanks

Add photos here

Submitted or published papers

1. Fuchs R., Pommeret D., Viroli C., "Mixed Deep Gaussian Mixture Model: A clustering model for mixed datasets", Advances in Data Analysis and Classification, 2021 (published)

2. Fuchs R., Pommeret D., Stocksieker S., "MIAMI: MIxed data Augmentation MIxture", 22nd International Conference on Computational Science and Its Applications, 2022 (published)

3. Fuchs R. , Thyssen M., Creach V., Dugenne M., Izard L., Latimier M., Louchart A., Marrec P., Rijkeboer M., Grégori G., Pommeret D., "Automatic recognition of flow cytometric phytoplankton functional groups using Convolutional Neural Networks", Limnology and Oceanography: Methods, 2022 (published)

4. Fuchs R., Baumas C.M.J., Garel M., Nerini D., Le Moigne F.A.C., Tamburini C., "A RUpture-Based detection method for the Active mesopeLagIc Zone (RUBALIZ): a crucial step towards rigorous carbon budget assessments", Limnology and Oceanography: Methods, 2022 (accepted)

5. Fuchs R., Rossi V., Caille C., Bensoussan N., Pinazo C., Grosso O., Thyssen M., "Intermittent upwelling events trigger delayed, major, and reproducible pico-nanophytoplankton responses in coastal oligotrophic waters", Geophysical Research Letters, 2022 (submitted)

6. Barrillon S., Fuchs R., Petrenko A., Comby C., Bosse A., Yohia C., Fuda J.-L., Bhairy N., Berline L., Cyr F., Doglioli A., Grégori G., Tzortzis R., d'Ovidio F., and Thyssen M., "Intense storm in the north-western Mediterranean Sea strongly shaped local physics and generated significant phytoplankton reaction", Biogeosciences, 2022 (submitted)

Thank you!

My 3 years of PhD