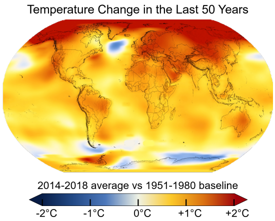
Quantifying the Impact of Environmental Parameters on Biodiversity

Clovis Galiez

Grenoble
Statistiques pour les sciences du Vivant et de l'Homme

May 25th 2020

Context: global warming



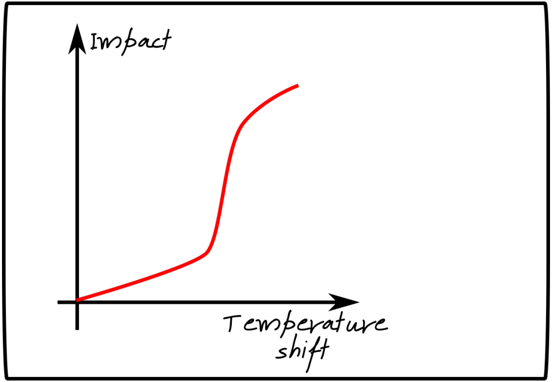
What impact? ⟶

Scientific question

How to quantify the impact on ecosystems of a change in environmental parameter (such as temperature)?

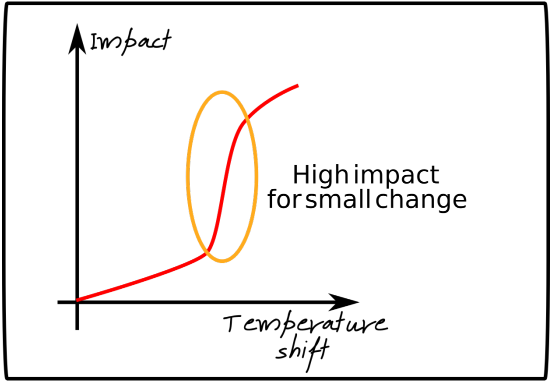
Goal: enable identification of critical ranges

The goal of this PhD project is to provide a **measure of impact of an environmental variable on ecosystems...**.



Goal: enable identification of critical ranges

The goal of this PhD project is to provide a **measure of impact of an environmental variable on ecosystems...**.



...and ultimately detect **tipping points**.

Approach

Sampling DNA directly in the environment

Technology now enables to measure abundance of species by DNA sequencing directly from the environment.

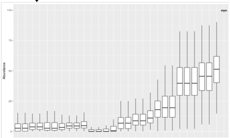
The *(very)* big picture:

Biological sample

DNA sequencing →

Metagenome

bioinformatics magic ↓

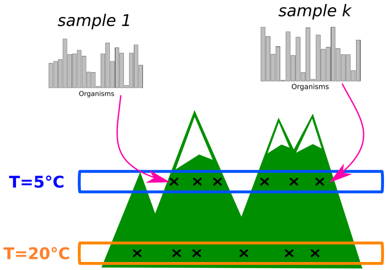


Abundance of species[1]

[1] *NB:* The data readily available for the project!

Data-driven approach

We consider existing ecosystems as possible optimal equilibrium given the environmental parameters (e.g. temperature).



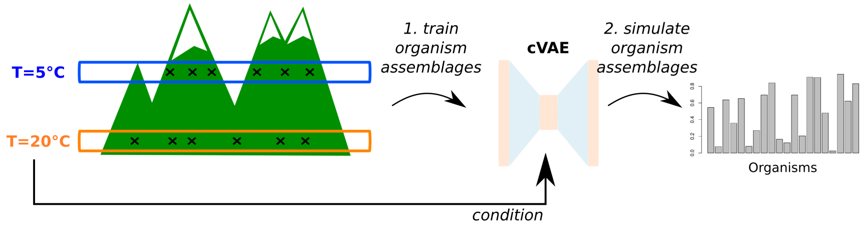
Our approach

We will devise a distance between sample distributions at various temperature to quantify the biodiversity shift.

Project steps

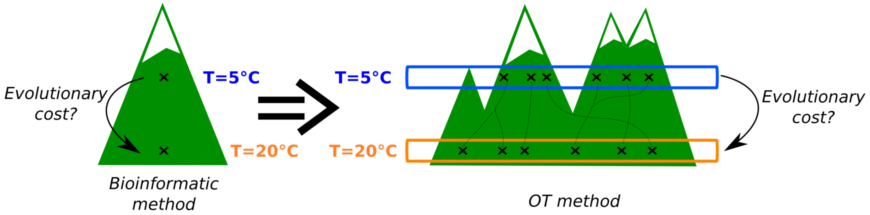
see also supplementary slides for Gantt and milestones

WP1: Simulation of assemblages



We will simulate unseen assemblages by interpolating data with conditional variational autoencoders (**cVAE**).
cVAEs will be learned on real data samples and benchmarking will be done using synthetic data generated with user-defined biotic and abiotic rules.

WP2: ecosystem sensitivity to environmental changes



In the example of a temperature increase, the hypothesis is that an organism assemblage at T_1 will shift to the closest (in terms of a given dissimilarity D) assemblage at T_2.
Optimal Transport (**OT**) theory provides a good framework[2] to evaluate the cost of an environmental parameter change on the ecosystem.

[2] see also supplementary slides for more details

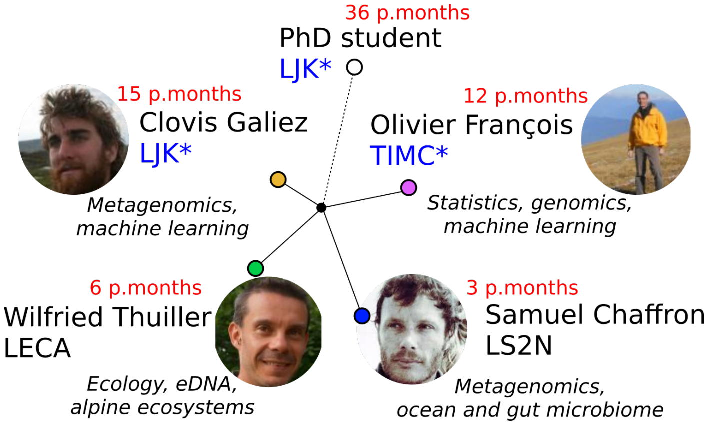
WP3: Application to real data

Data is readily available, with expertize among the consortium:

- Alpine ecosystem (LECA): Orchamp eDNA data
  Goal: measure 1. the impact of temperature change using eDNA samples on altitude gradients, and 2. ecosystem adaptation to brutal shifts
- Gut microbiome (TIMC): amplicon DNA (16S barcodes)
  Goal: assess the impact of environmental conditions of humans on their gut microbiome in term of shift in biodiversity and biological functions
- Marine (LS2N): Tara Oceans shotgun metagemomics
  Goal: measure the impact of gloabl temperature change in the ocean in terms of ecological services and functions

Consortium

Consortium composition



*: In PersyvalLab

Summary of the QIEP-B project

Highlights of the QIEP-B project:

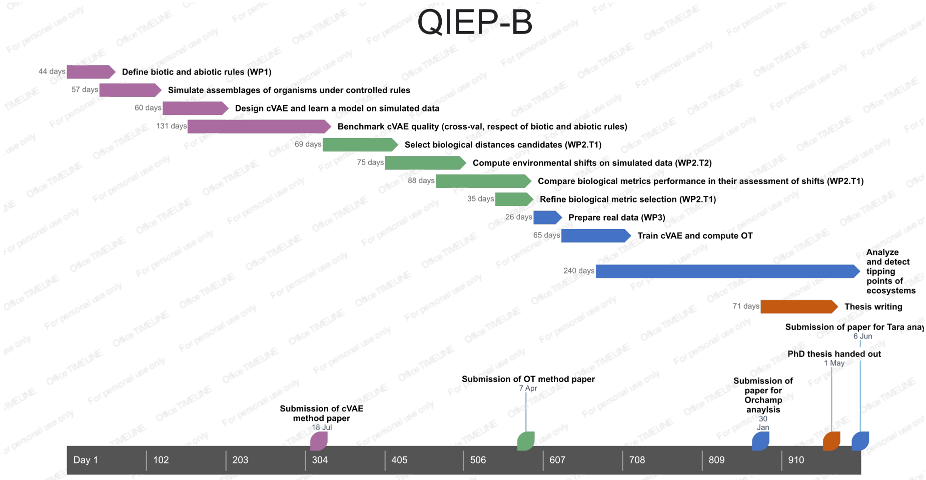
- We devise a data-driven and model-free method for tackling global change monitoring and forecasting of biodiversity
- This project will contribute to strangthen the links between Grenoble labs (LJK, TIMC, LECA) and open up to a new collaboration in Nantes (LS2N).
- This project widens the scope of the PersyvalLab to data-driven research applied to ecology.

Questions?

clovis.galiez@univ-grenoble-alpes.fr

Timeline

Gantt chart



QIEP-B
44 days
Define biotic and abiotic rules (WP1)
57 days
Simulate assemblages of organisms under controlled rules
60 days
Design cVAE and learn a model on simulated data
131 days
Benchmark cVAE quality (cross-val, respect of biotic and abiotic rules)
69 days
Select biological distances candidates (WP2.T1)
75 days
Compute environmental shifts on simulated data (WP2.T2)
88 days
Compare biological metrics performance in their assessment of shifts (WP2.T1)
35 days
Refine biological metric selection (WP2.T1)
26 days
Prepare real data (WP3)
65 days
Train cVAE and compute OT
240 days
Analyze and detect tipping points of ecosystems
71 days
Thesis writing
Submission of paper for Tara anay
6 Jun
PhD thesis handed out
1 May
Submission of OT method paper
7 Apr
Submission of paper for Orchamp anaylsis
30 Jan
Submission of cVAE method paper
18 Jul
Day 1
102
203
304
405
506
607
708
809
910

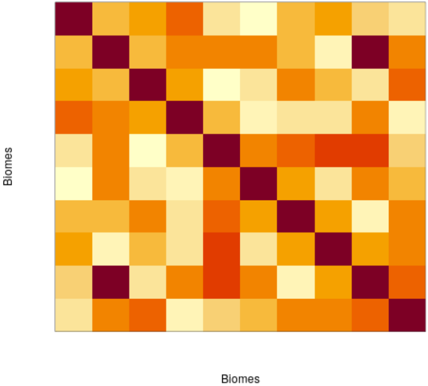
Milestones

- WP1 Develop simulation of assemblages
- WP1 Conditional Variational Autoencoders (cVAE) for learning organisms assemblages
- WP2 Use Optimal Transport throery to compute a distance between environmental conditions
- WP3 Apply on available data in the consortium (Alpine, ocean and gut ecosystems)

How OT will be used for assessing impact of environmental parameters on ecosystems?

Define a similarity between biomes

We can fix a disimilarity $D_{i,j}$ (bioinformatics methods, e.g. Bray-Curtis) matrix between biomes:



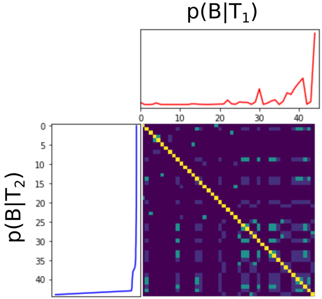
Wasserstein metric

Having a ground disimilarity $D_{i,j}$ between N samples, we lift the metric to the distribution of samples.

$$\mathcal{W}(B|T_1, B|T_2) = \min_{P \in A(B|T_1, B|T_2)} \sum_{i,j} D_{i,j} P_{i,j}$$

where

$$A(B|T_1, B|T_2) = \{P \in \mathbb{R}^{N \times N} | P \mathbb{1}_N = B|T_1 \text{ and } P^\top \mathbb{1}_N = B|T_2\}$$



Quantification of impact

Given a disimilarity between biomes, we will define for instance:

Impact of a change of temperature from T_0 to T_1

$$\iota(T_0, T_1) = \mathcal{W}(B|T_0; B|T_1)$$

Hopefully this can help to address questions such as:

- Detect the ranges of temperature that are the most sensitive to change:

$$s(T) = \frac{\iota(T, T+\delta T)}{\delta T}$$

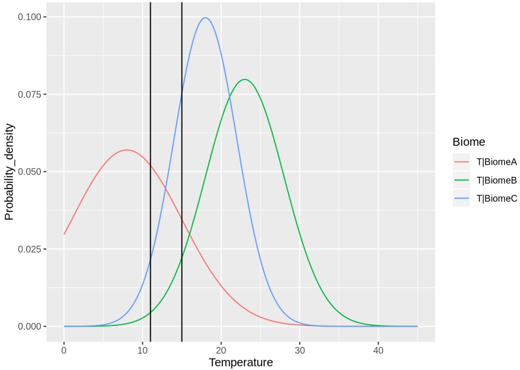
- Quantify the impact of a trajectory of evolution of temperature:

$$\int_a^b \iota(T(x), T(x+dx))^2 . f'(x) dx$$

If WP1 fails to provide good simulation?

Sample niche

If WP1 fails, instead of using enriched data by simulated assemblages, we will use only available data. We need to obtain a distribution of samples at a given temperature:



To this end...

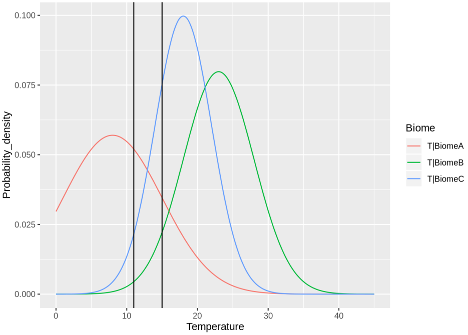
Bayes: reverting sample niche

...we use a simple Bayes rule.

$$p(B_k|T) = \frac{p(T|B)k)p(B_k)}{\sum_i p(T|B_i)p(B_i)}$$

$p(T|B_i)$

$p(B_i|T)$



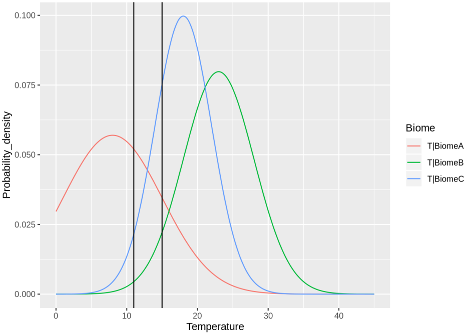
$\longrightarrow$

Bayes: reverting sample niche

...we use a simple Bayes rule.

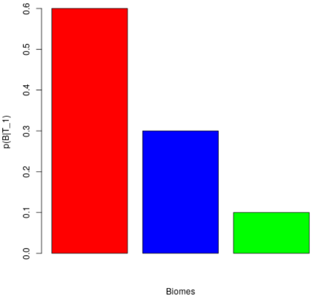
$$p(B_k|T) = \frac{p(T|B)k)p(B_k)}{\sum_i p(T|B_i)p(B_i)}$$

$p(T|B_i)$



$\rightarrow$

$p(B_i|T)$

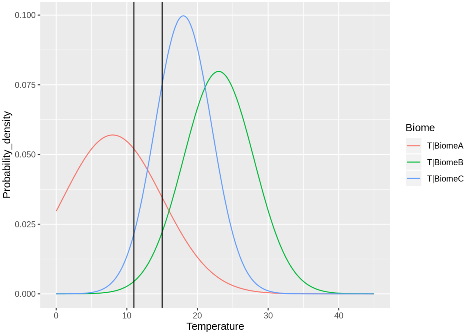


Bayes: reverting sample niche

...we use a simple Bayes rule.

$$p(B_k|T) = \frac{p(T|B)k)p(B_k)}{\sum_i p(T|B_i)p(B_i)}$$

$p(T|B_i)$



$\rightarrow$

$p(B_i|T)$