Bayesian approach of pollen-based palaeoclimate reconstructions: *Toward the modelling of ecological processes*

Vincent Garreta

Université Aix-Marseille III
CEREGE UMR6635
Europôle de l'Arbois
Aix en Provence

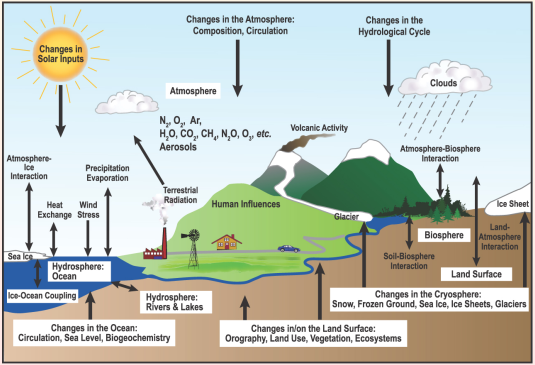
Thesis presentation, April 29, 2010

Part of the ESF EUROCORE programme DECVEG
Funded by a CNRS BDI grant

Climate research

Many questions, e.g human history and climate change issues require (deep) understanding of the 'climate system'

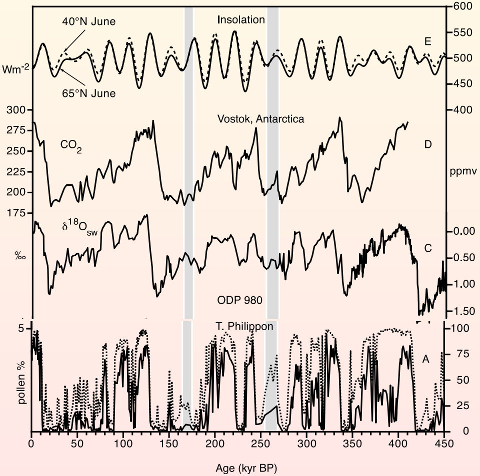
- earth-scale system
- interacting through complex exchange mechanisms
- reacting to forcings (e.g insolation, CO_2 concentration, solar activity)



Solomon et al. (2007)

Palaeoclimatology

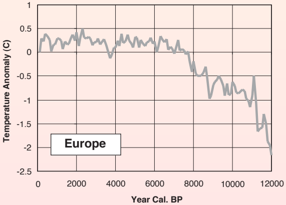
This non-reducible system has to be measured for a large range of conditions i.e forcings, having different characteristic time scales



- Climate reconstruction
- and uncertainties

In this presentation

- new method
- Holocene in Europe

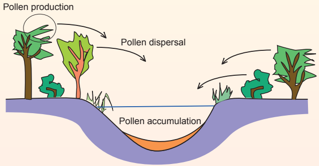


Tzedakis et al. (2003); Davis et al. (2003)

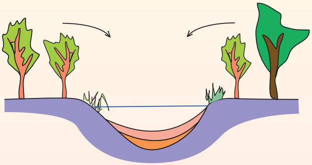
Pollen as a climate proxy

Pollen assemblages found in lake sediments provide an image of past vegetation partially controlled by climate.

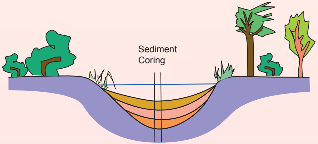
t_0: Mediterranean climate



t_1: Oceanic climate



t_{n+1}: Coring

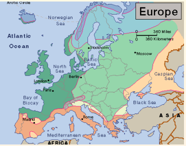


$\Rightarrow$ Vegetation (and climate) history recorded in sediment core

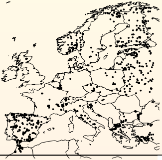
Pollen-based palaeoclimate reconstructions

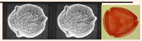
1. **Calibration**: Learning the links between modern climate and pollen

Modern climate



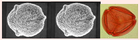
Pollen samples



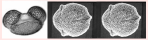
Mediterranean climate	(pinus, evergreen oak)	
Oceanic climate	(oak, beech)	

2. **Reconstruction**: inferring past climate based on pollen assemblages.

^{14}C dating + pollen

$\leftarrow$ t_1: 

t_1: Oceanic

$\leftarrow$ t_0: 

t_0: Mediterranean ($\approx$)

Transfer Functions & notations

Pollen $Y = (Y^1, .., Y^k)$, counts per taxa relatively to their sum $\Sigma_{j=1}^k Y^j$

Climate $C = (C^1, .., C^l)$, vector of climate variables, e.g $(T_{\text{jan}},\ T_{\text{jul}},\ P_{\text{ann}})$

Example of TF noted $f()$: $$Y = f(C, \theta) + \epsilon$$

1. **Calibration**

   obtain $\hat{\theta}$ such that for all $s = 1..N$, $\quad Y_s \approx f(C_s, \hat{\theta})$

   based on modern data  and 

2. **Reconstruction**

   obtain C_t such that, $\quad Y_t \approx f(C_t, \hat{\theta})$

Two types of TF

CorrelativeTF are *statistical* relations *climate-pollen*

- Backward $C = f(Y, \theta, \epsilon)$

  Reconstruction = *prediction* C_t

- Direct $Y = f(C, \theta, \epsilon)$

  Reconstruction = *inversion* C_t

Mechanistic TF include a vegetation model in *climate-vegetation-pollen*

$$Y = g\left(\text{vegetation}\left[C, \mathrm{CO_2}, ...\right], \theta, \epsilon\right)$$

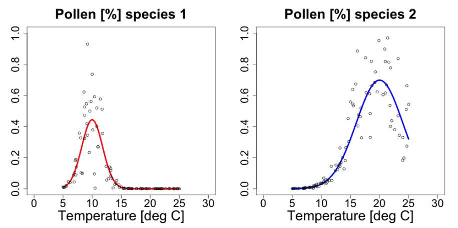
Reconstruction = *inversion mechanistic model*

Correlative paradigm

Correlative paradigm: interest lies in the realised correlation climate-pollen

+

$\Rightarrow$ learning $Y = f(C, \theta)$

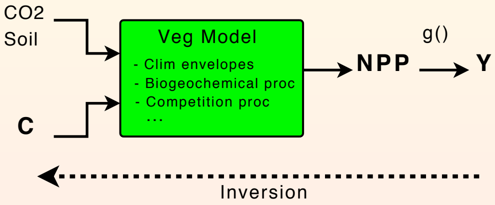


- Problem if species depend on, e.g CO_2
- Problem with extrapolation

Mechanistic approach

Guiot et al. (2000) propose

- use a vegetation model for (climate, CO_2, soil)-vegetation
- use a statistical model for vegetation-pollen



$$Y = g\left(\text{vegetation}\left[C, CO_2, ...\right], \theta, \epsilon\right)$$

- CO_2 and main factors included
- Extrapolation $\approx$ (depending on g)

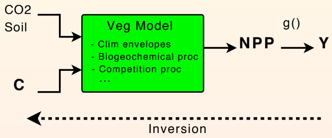
Bayesian inference for palaeoclimatology

Bayesian framework:

Prior + model(data) $\rightarrow$ Posterior

$$Y = g\left(\text{vegetation}\left[C, \mathrm{CO}_2, ...\right], \theta, \epsilon\right)$$

$$\Leftrightarrow p(Y|C, \theta)$$

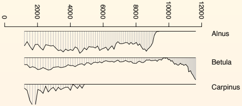


1. Calibration: $p(\theta|Y_s, C_s) \propto p(Y_s|C_s, \theta)\, p(\theta)$
2. Reconstruction: $p(C_t|Y_t) \propto \int p(Y_t|C_t, \theta)\, p(\theta|Y_s, C_s)\, p(C_t)\, d\theta$

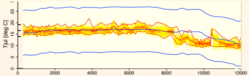
Modern inference algorithms (e.g MCMC) allow to obtain such posteriors

Outline

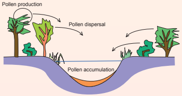
1. Develop the inversion of a dynamic vegetation model
   → temporal reconstructions



⇒



2. Propose a process-based modelling of vegetation-pollen
   → spatial calibration



⇒ consider spatialised

3. Combine both methods
   → toward full *process-based* and *spatio-temporal* reconstructions

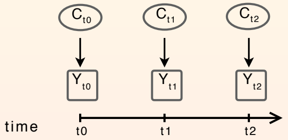
4. Conclusion and perspectives

Part 1: Inversion of a Dynamic Vegetation Model (Garreta, Miller et al. 2009)

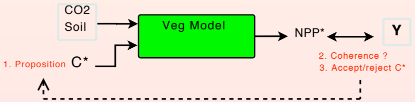
Guiot et al. with BIOME3

- Static & Deterministic model

  $\Rightarrow$ 1 by 1 inversion

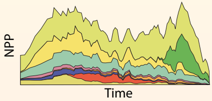


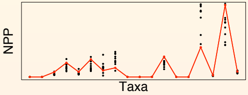
- MCMC algorithm



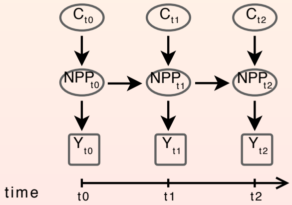
Using LPJ-GUESS (Smith et al. 2001)

- Dynamic & Stochastic model



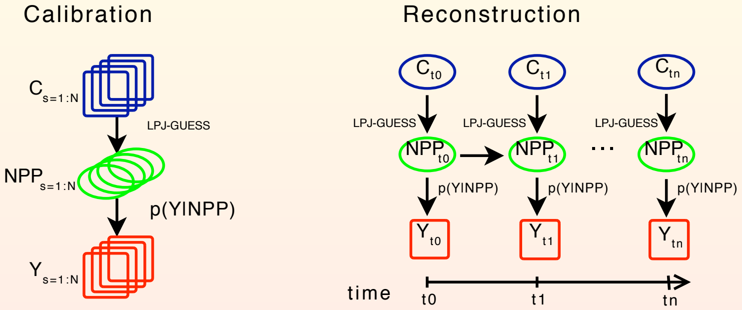


$\Rightarrow$ Joint inversion !



- High-dimensional problem...

Statistical model embedding LPJ-GUESS

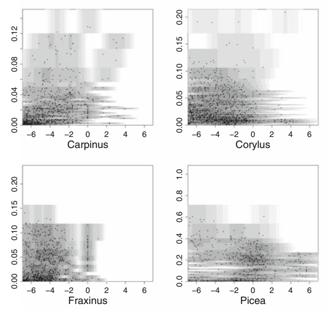


- $p(Y_t|\mathrm{NPP}_t)$ pollen/vegetation distribution $\rightarrow$ to be constructed

$p(Y_t|\mathrm{NPP}_t)$ modelling

It translates the relations between $k = 15$ NPP and pollen Y

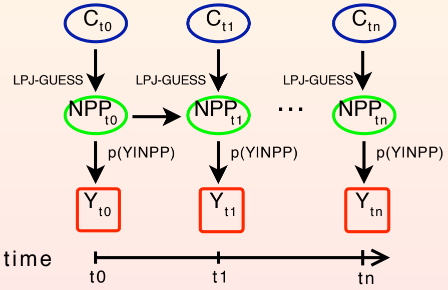
- Simplifications based on independence hypotheses
- Finally, statistical smoothing of
  $\longrightarrow$ 1 by 1 pollen response to vegetation abundance



The inference problem

Reconstruction equivalent to obtaining $p(C_{t_0:t_n}, \mathrm{NPP}_{t_0:t_n} | Y_{t_0:t_n})$ from

- a hidden Markov model
- defined through an implicit (transition) distribution, LPJ-GUESS



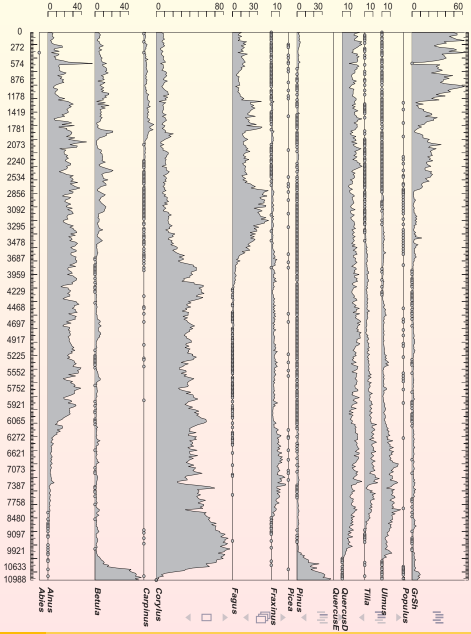
Use of a Sequential Monte Carlo (SMC or Particle Filter) (Doucet et al. 2001)

- Scan climate time after time

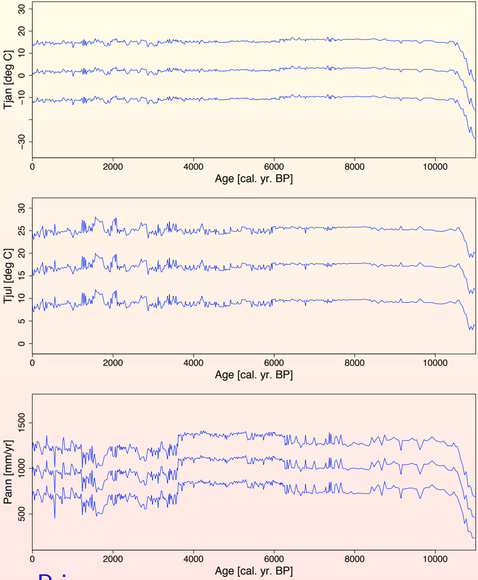
Meerfelder Maar Holocene reconstruction

Data from Litt et al. (2009)

- Maar in the Eifel (SW Germany)
- 406 samples between 11kaBP and 0BP
  $\approx$ 30 yr between samples
- Dates are varved & ^{14}C

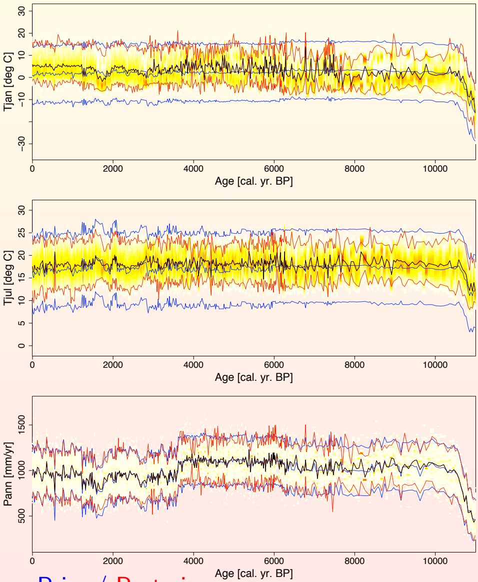


Reconstruction



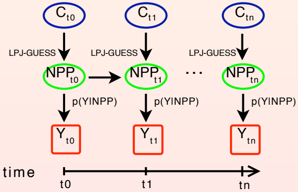
Prior

Reconstruction



Prior / Posterior

- No constraint on precipitation (Prior=Posterior)
- Coherent with pollen diagram
- Confidence intervals
  - Not process-based
- Noisy reconstruction
  - No climate correlation

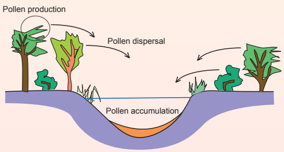


Part 2: Process-based modelling of $p(Y|\mathrm{NPP})$ with F. Mortier and J. Chadœuf

Mechanistic TF → full process-based TF

- causality, coherent quantification of uncertainties
- extrapolation (no-analogue problem)

Vegetation-pollen study has a long history in palaeo-ecology (*)



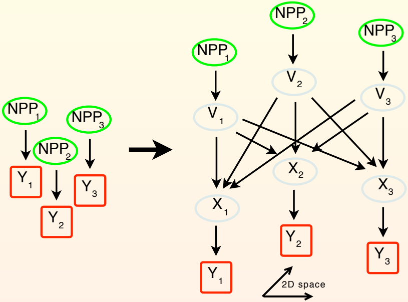
Main processes include

- Pollen production
- dispersal
- capture and accumulation

(*) including Von Post 1916; Davis 1963; Tauber, 1965; Kabailiene 1969; Webb 1974; Prentice 1985; Sugita 1994 ; Pacioreck and McLachlan 2009

Modelling Approach

We use a hierarchical approach to expand $p(Y|\text{NPP})$ and represent main processes



Shares structure with classical model (ERV) but

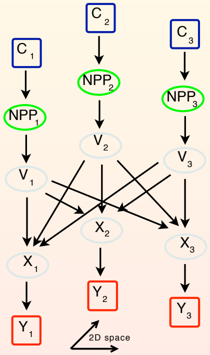
- Global scale (Europe)
- Parameter inference (no need exp. values)
- Hierarchical + bayesian (coherent uncert. quantification)

Challenge in statistics

- Model zero-inflation and over-dispersion for Multinomial data
- Infer a spatial structure on large dataset

Hidden levels

1. Potential $\longrightarrow$ actual vegetation $p(V|\mathrm{NPP}, \theta_1)$
   - Additive noise
   - Stationary parameters
2. Pollen production and dispersal θ_2
   - Linear production per taxa $(b^j\, V^j)$
   - Gaussian dispersion with dispersal range per taxa
3. Pollen accumulation $p(X|V, \theta_2)$
   - Modelled following a Poisson distribution
4. Sampling $p(Y|X)$
   - Modelled following a Multinomial distribution centred on $p_i^j = X_i^j / \Sigma_j X_i^j$



European dataset results

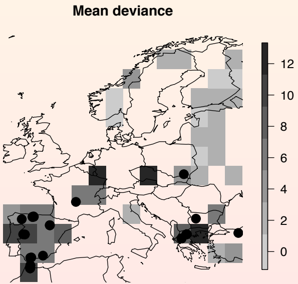
1300 sites and 15 taxa ($\approx$ 75 parameters and 39 000 latent variables)

Dispersal parameters per taxa

Taxa (j)	Pinus	Quercus Ever.	Populus	Grass & Shrubs
$2\gamma^j$ (in km)	160	40	900	100
b^j/b^{15}	0.44	0.05	0.01	1

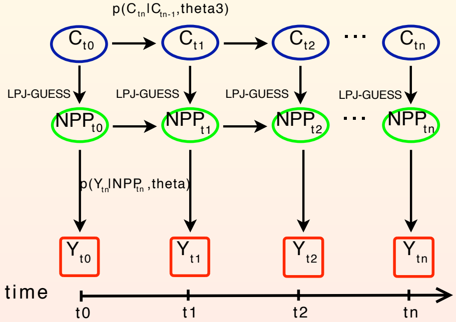
Adequation testing indicates problems

- Overdispersion not sufficient
  $\Rightarrow$ NB model proposed
- Discrepancy not homogeneous in space
  $\Rightarrow$ LPJ-GUESS quantitative validation?



Part 3: Merging process-based and temporal approaches

- Insert process-based model $p(Y|\text{NPP})$ inside SMC
- Model and infer temporal correlation in climate



$\Rightarrow$ Full-process based: straightforward in theory! But

- $p(Y|\text{NPP})$ is now defined as

$$\int p(Y|X)\, p(X|V,\theta_2)\, p(V|\text{NPP},\theta_1)\, p(\theta_1,\theta_2|Y_s,C_s)\; d(X,V,\theta_1,\theta_2)$$

Modelling and inferring climate correlation

Simple model for climate temporal correlation, i.e climate inertia

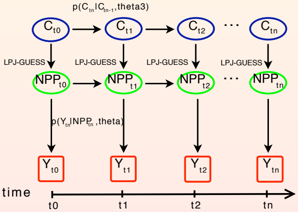
$$p(C_{t_n}|C_{t_{n-1}}, \theta_3) = \mathcal{N}\left(C_{t_{n-1}}, (t_n - t_{n-1})\theta_3\Sigma\right)$$

$$\Rightarrow \quad Var(C_{t_n} - C_{t_{n-1}}) \propto (t_n - t_{n-1})\theta_3$$

May be hard to infer sequentially a parameter 'static' in time

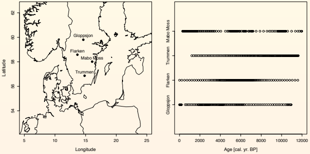
We use conjugated distributions (Storvik 2002 ; Fearnhead 2002) to update

$$p(\theta_3|C_{t_0:t}, \mathrm{NPP}_{t_0:t}, Y_{t_0:t})$$

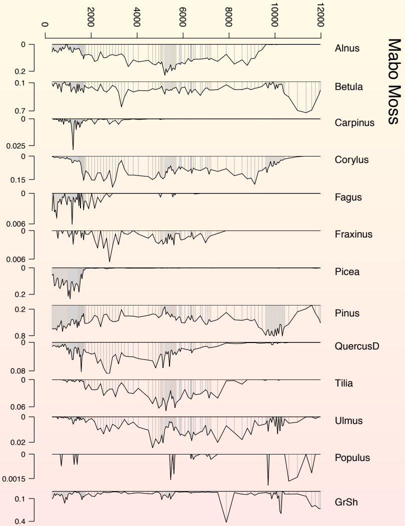


$\Rightarrow$ Imposes a 2-passes algorithm !

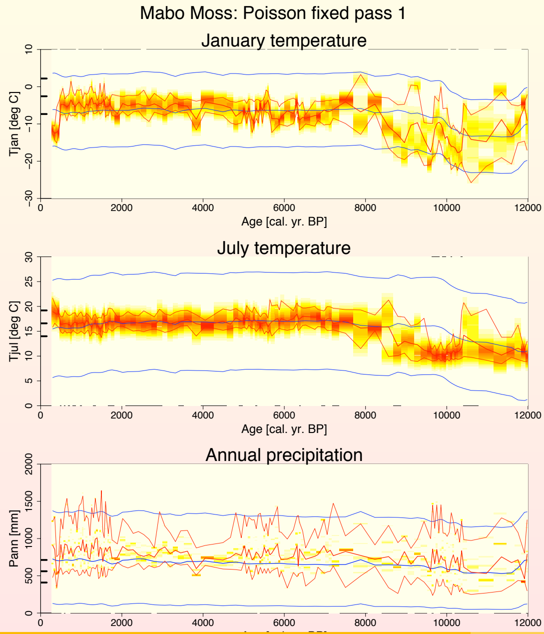
Palaeoclimate from 4 Swedish cores



- 4 close cores ($< 400km$)
- $\approx$ covering Holocene
- 1 reconstruction per site (50h computing each)

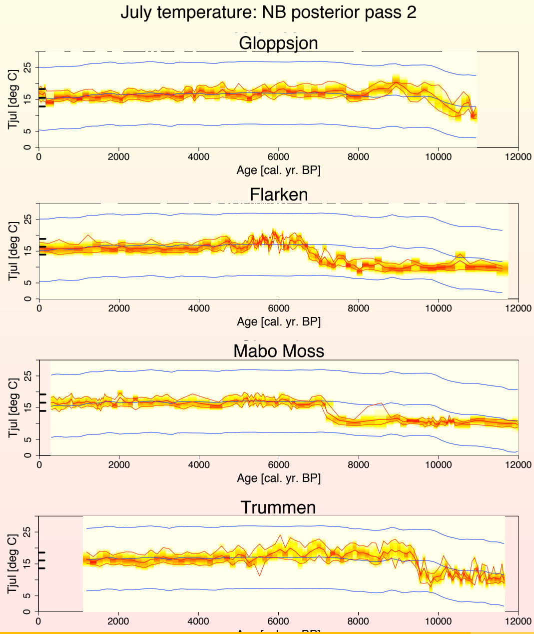


Mabo Moss reconstruction



- Smoother reconstructions
- Confidence interval process-based
- No more precipitation constraint

Four sites comparison



Inconsistent differences inter-sites (timing increase)

- Spatial differences in vegetation (migrational)
  - not accounted for in LPJ-GUESS nor $p(Y|\mathrm{NPP})$
- Over-sensitivity inversion
  - high dimensional integration
  - filtering algorithm

Conclusion

Palaeoclimatology

- We proposed a method for the inversion of the last vegetation model generation
- We proposed a process-based model to represent pollen-vegetation link at continental scale

Statistics

- We developed an inference algorithm for a dynamical system made of
  - an implicit model (LPJ-GUESS) and
  - a highly layered structure
- We proposed a model for spatial multinomial data showing zero-inflation and over-dispersion

Applications indicate

- A fast and robust smoothing algorithm for LPJ-GUESS is required
- Vegetation models such as LPJ-GUESS need
  - to be more strongly data-based (calibration and validation)
  - include vegetation spatial dynamics (migration)

Perpectives

Inference for dynamical systems defined through implicit models (parameters and states, smoothing, fast and robust)

Ideas include

- Local smoothing, model emulation, reified inference

Applied to

- Reconstruction of past climate
  - including ^{14}C dating uncertainties
  - spatio-temporal
- Calibration of parameters 'inside' a vegetation model
  - re-calibration using large, modern, dataset
  - using past vegetation dynamics to calibrate migration processes
- Inference problems requiring to merge mechanistic models and real data

Merci à tous !