# Explorer le potentiel de History Matching pour calibrer les paramètres d'ORCHIDEE

Simon Beylat, Nina Raoult, James M. Salter, Frédéric Hourdin, Vladislav Bastrikov, Catherine Ottlé, and Philippe Peylin

# ORCHIDEE

ORCHIDEE
LAND SURFACE MODEL

- Interaction between land and atmosphere.
- Energy and Water Balance
- Carbon Budget

**! A lot of empirical equation = Many parameters !**

For example :
Vegetation describe using 15 plant functional types :

| PFT | Climate | Vegetation type | Phenology class |
|---|---|---|---|
| 1 | Global | NA | Bare soil |
| 2 | Tropical | Woody | Broadleaf evergreen |
| 3 | Tropical | Woody | Broadleaf deciduous |
| 4 | Temperate | Woody | Needleleaf evergreen |
| 5 | Temperate | Woody | Broadleaf evergreen |
| 6 | Temperate | Woody | Broadleaf summer green |
| 7 | Boreal | Woody | Needleleaf evergreen |
| 8 | Boreal | Woody | Broadleaf summer green |
| 9 | Boreal | Woody | Needleleaf deciduous |
| 10 | Temperate | Herbaceous | Natural ($C_3$) |
| 11 | Global | Herbaceous | Natural ($C_4$) |
| 12 | Global | Herbaceous | Managed ($C_3$) |
| 13 | Global | Herbaceous | Managed ($C_4$) |
| 14 | Tropical | Herbaceous | Natural ($C_3$) |
| 15 | Boreal | Herbaceous | Natural ($C_3$) |

**ORCHIDEE use many free parameters to describe vegetation.**

**A parameter has a value for each PFTs -> Very large number of parameters.**

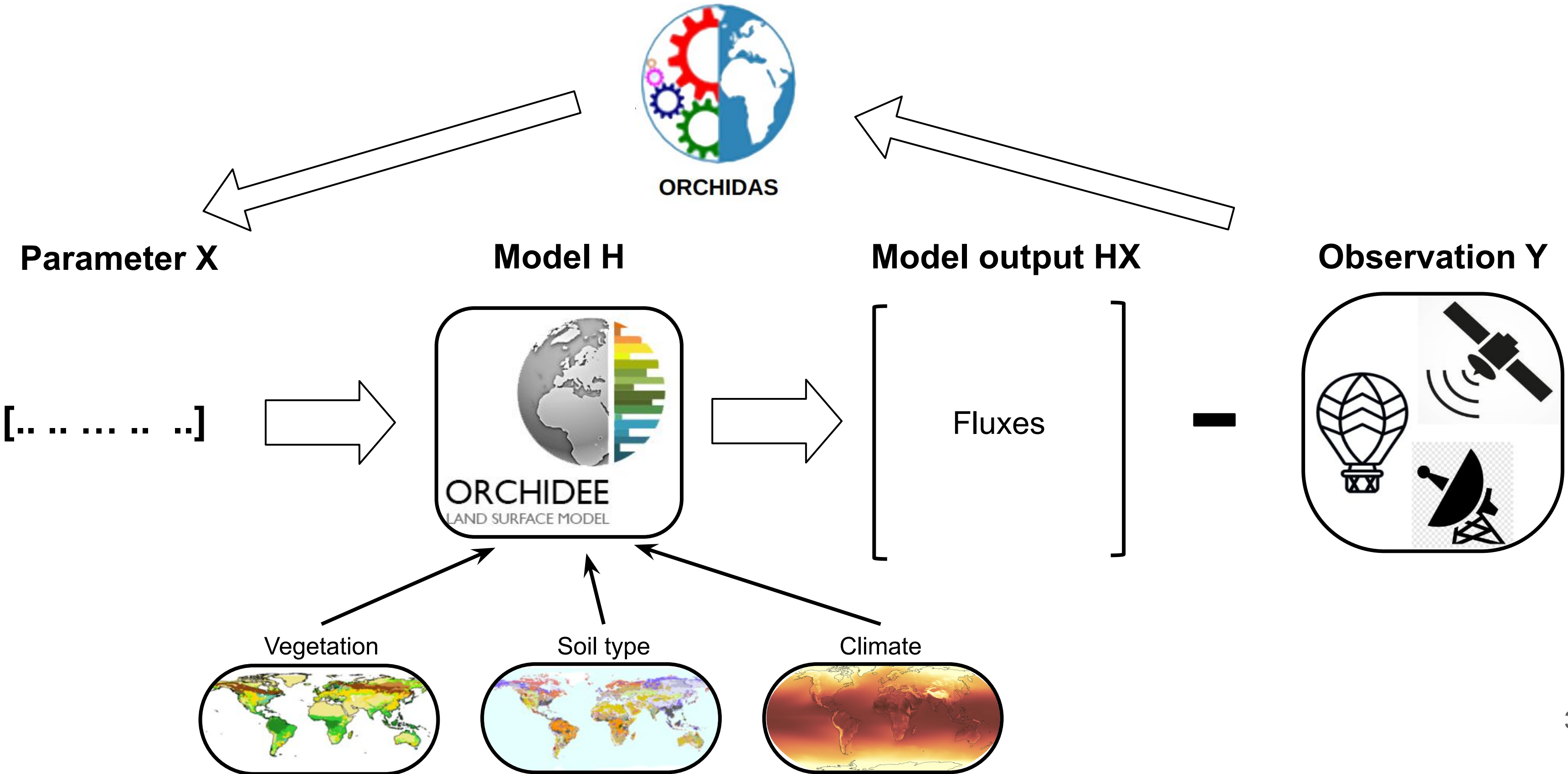
**We need to calibrate them rigorously.**

# ORCHIDAS : ORCHIDEE Data Assimilation Systems

**A Bayesian setup :**

*(Tarantola 1987/2005)*

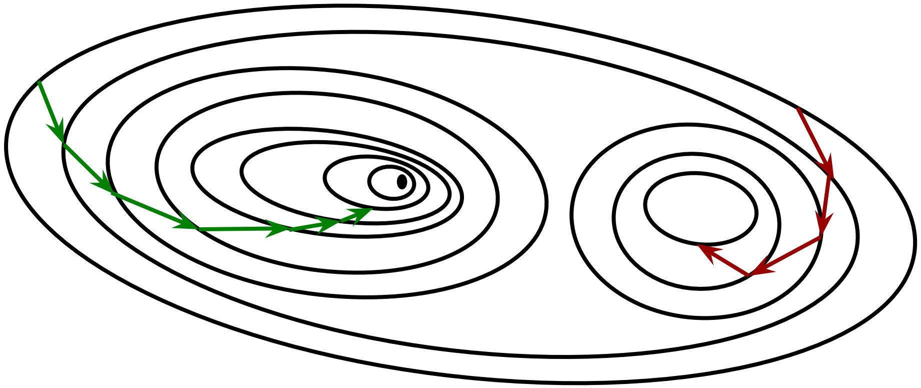
$$J(X) = \tfrac{1}{2}(HX-Y)^T R^{-1}(HX-Y) + \tfrac{1}{2}(X_b-X)^T B^{-1}(X_b-X)$$



# Optimisation Algorithm

June - 2023

## Gradient Descent : L-BFGS-B



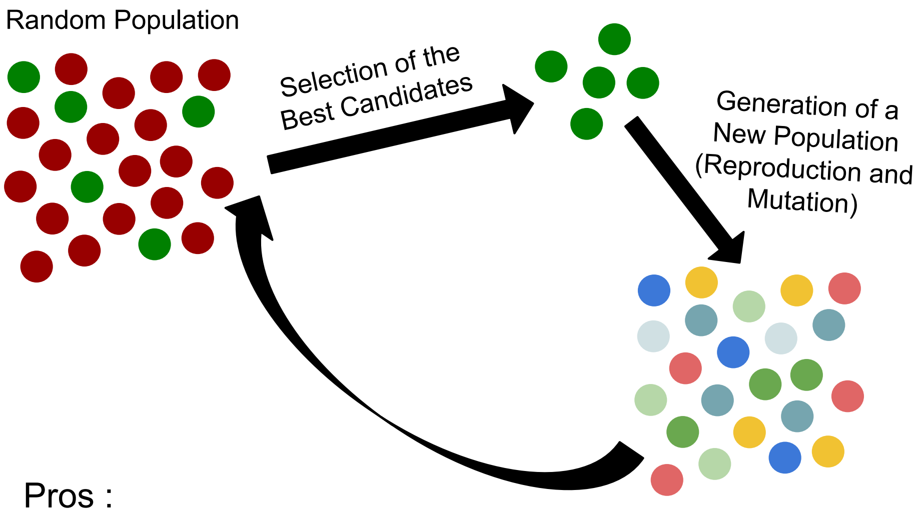
Pros :

- Fast
- A direct path leading to the minimum

Cons:

- Can be stuck in local minima
- Need to calculate the Jacobian and Hessian

## Genetic Algorithm




Pros :

- Less chance to be stuck in local minima
- Exploring a higher proportion of the loss function

Cons:

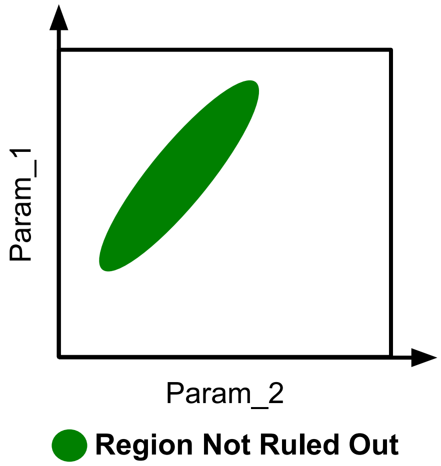
- Expensive and slow

**In the end, you have one set of parameters and reduced posterior uncertainty!**

# History Matching

"**History matching** uses **emulators** to **rule out regions of the parameter space** of the climate model **that are inconsistent with physical observations**" *(Williamson et al. 2013)*

Output of History Matching :




What could we do with this subset of parameters?

Several possibilities :

- **Exploration** : Having more information about the parameter space.
- **Uncertainty** : Simulation of a set of parameters to quantify uncertainty.
- **Optimisation** :
    - Using optimisation (minimisation) techniques on the subset to find the optimum.
        - Reduces the possibility of falling into local minima.
        - **Reduces the computing time.**
    - Continuing the iteration with History Matching.

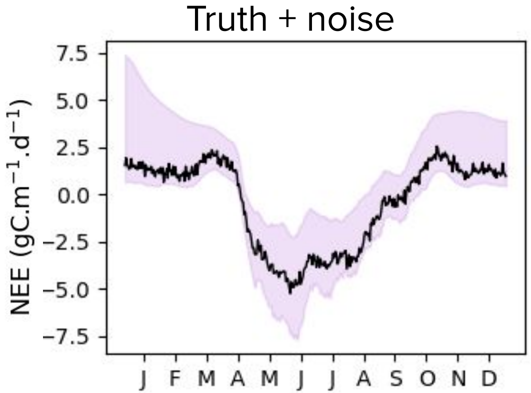
# To investigate the potential of History Matching, we first set up a twin experiment

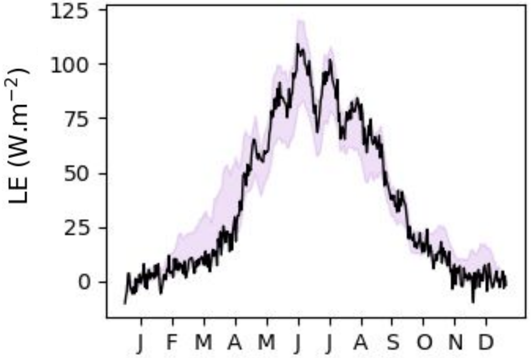
Twin experiment :
Pseudo data created by ORCHIDEE.
We know the true parameter

**Site:** FR-Fon (2005)

NEE :
Net Ecosystem Exchange
Net Carbon flux

LE : Latent heat flux



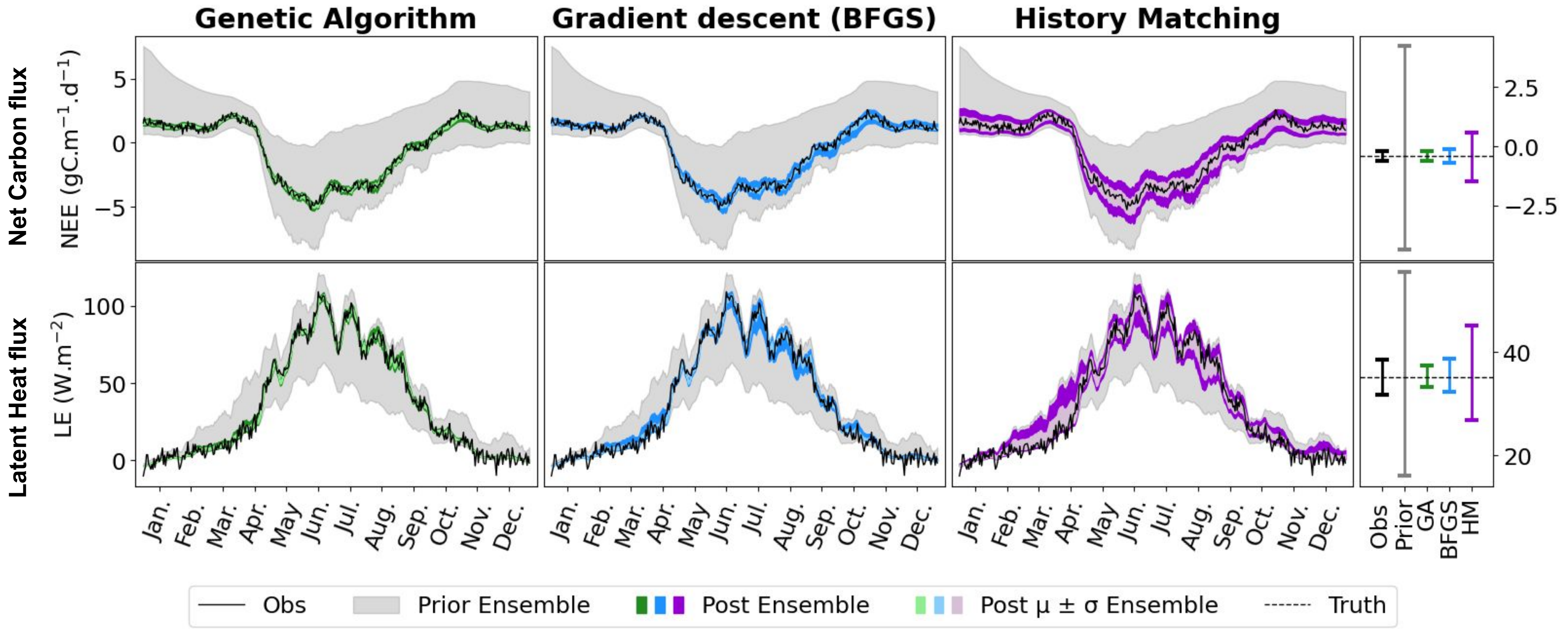

**6 parameters used :**

| | Description |
|---|---|
| $VC_{max}$ | Maximum carboxylation rate ($\mu$molm$^{-1}$s$^{-1}$) |
| SLA | Specific leaf area (m$^2$) |
| $L_{agecrit}$ | Critical leaf age for starting leaf senescence (days) |
| $Evap_{res}$ | Factor controlling bare soil resistance to evapotranspiration (-) |
| $Root_{prof}$ | Root profile parameter of an exponential function that describes the decrease of root density as a function of depth (m) |
| $Q_{10}$ | Parameter determining the temperature dependency of the heterotrophic respiration (-) |

**Need to compare Ensemble :**
DA : 200 optimisations using random priors
1 ensemble of HM

# Results : Model-data fit

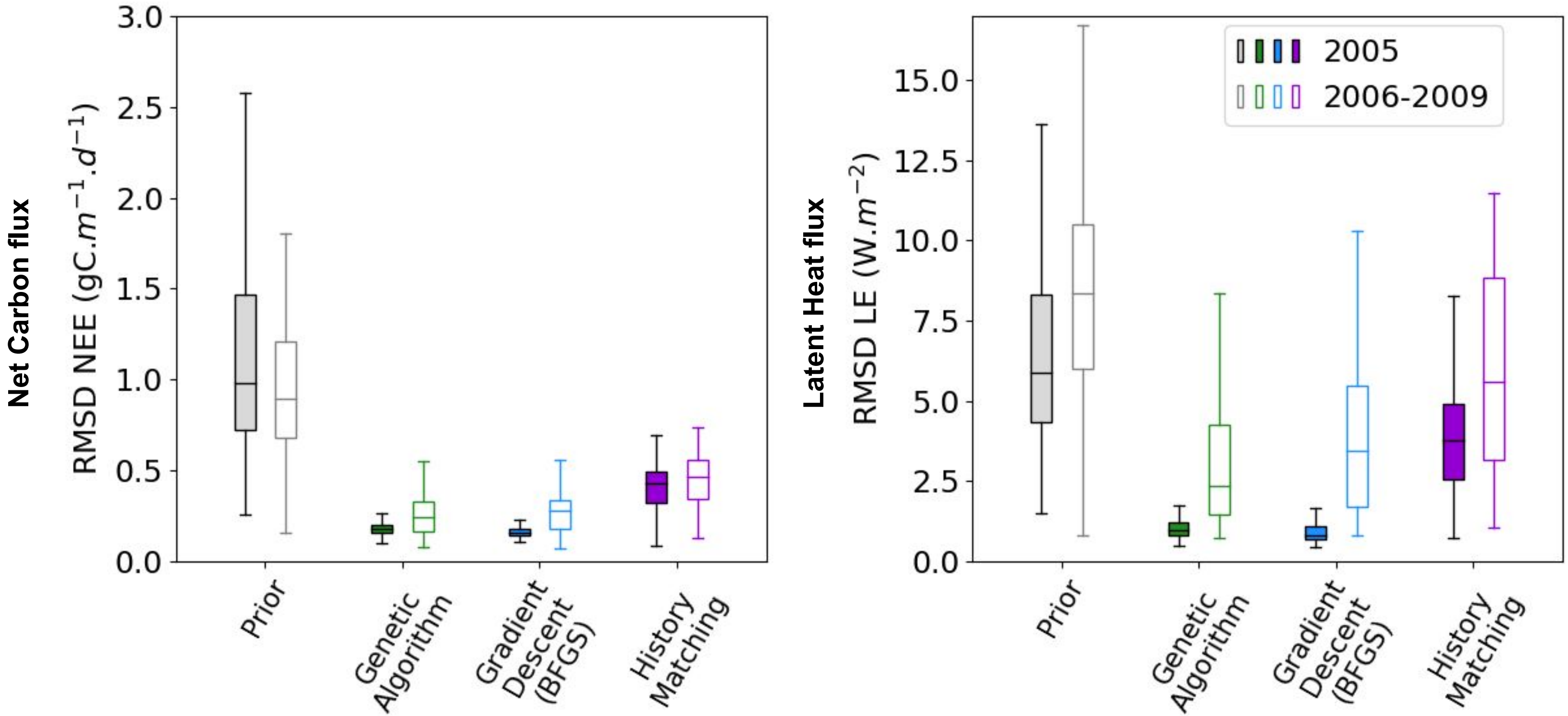
Metric: RMSD



# Results : Model-data fit

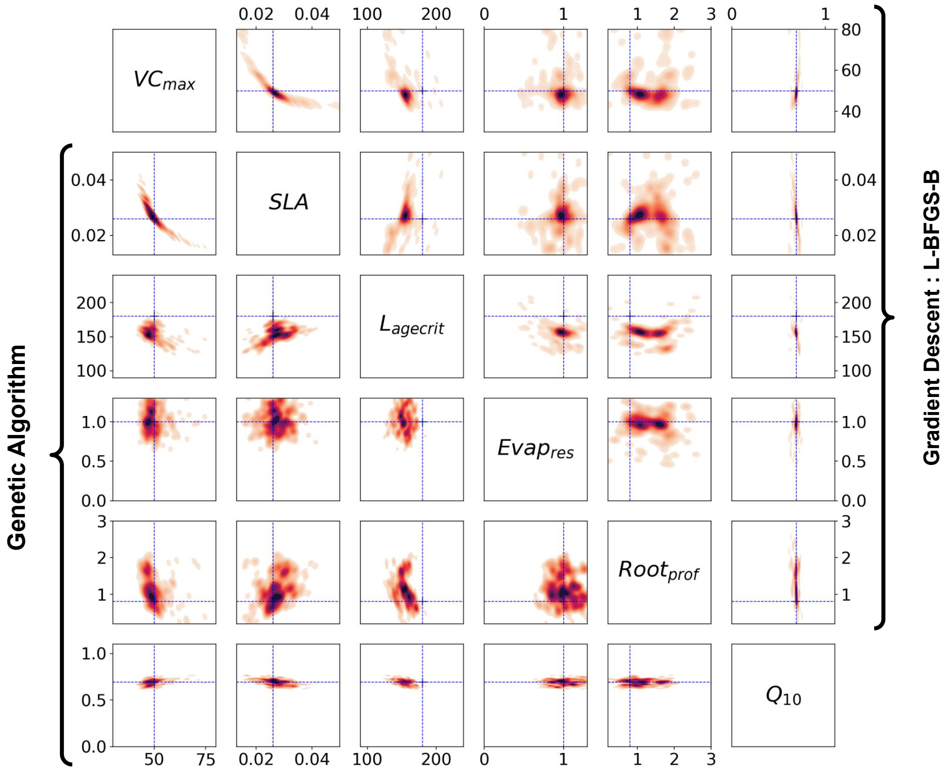
HM overfit less than DA

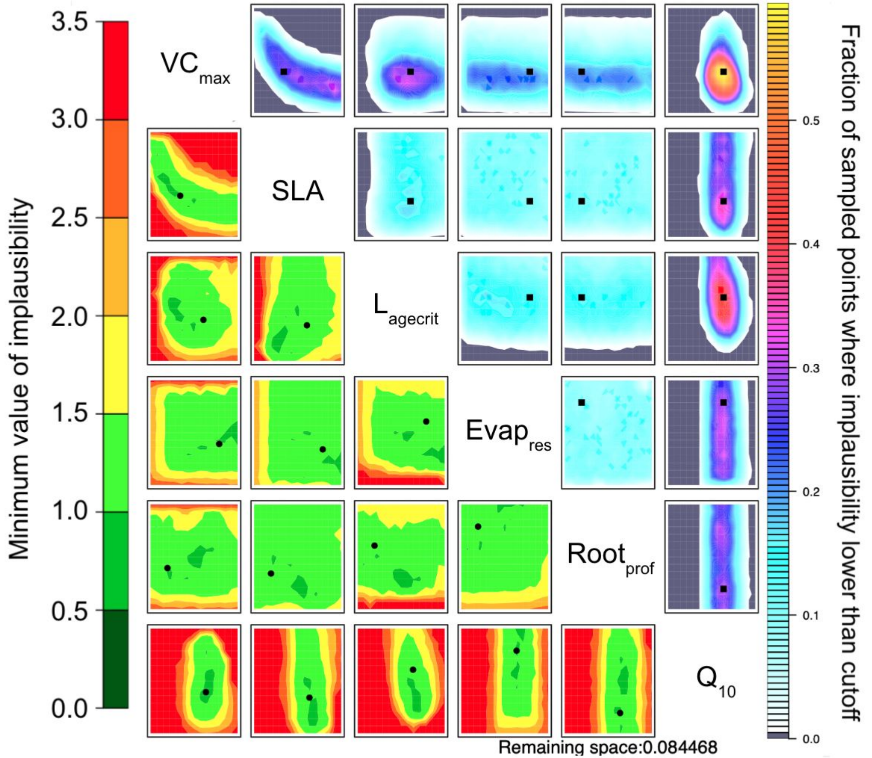
Calibration over 2005
Evaluation over 2006-2009



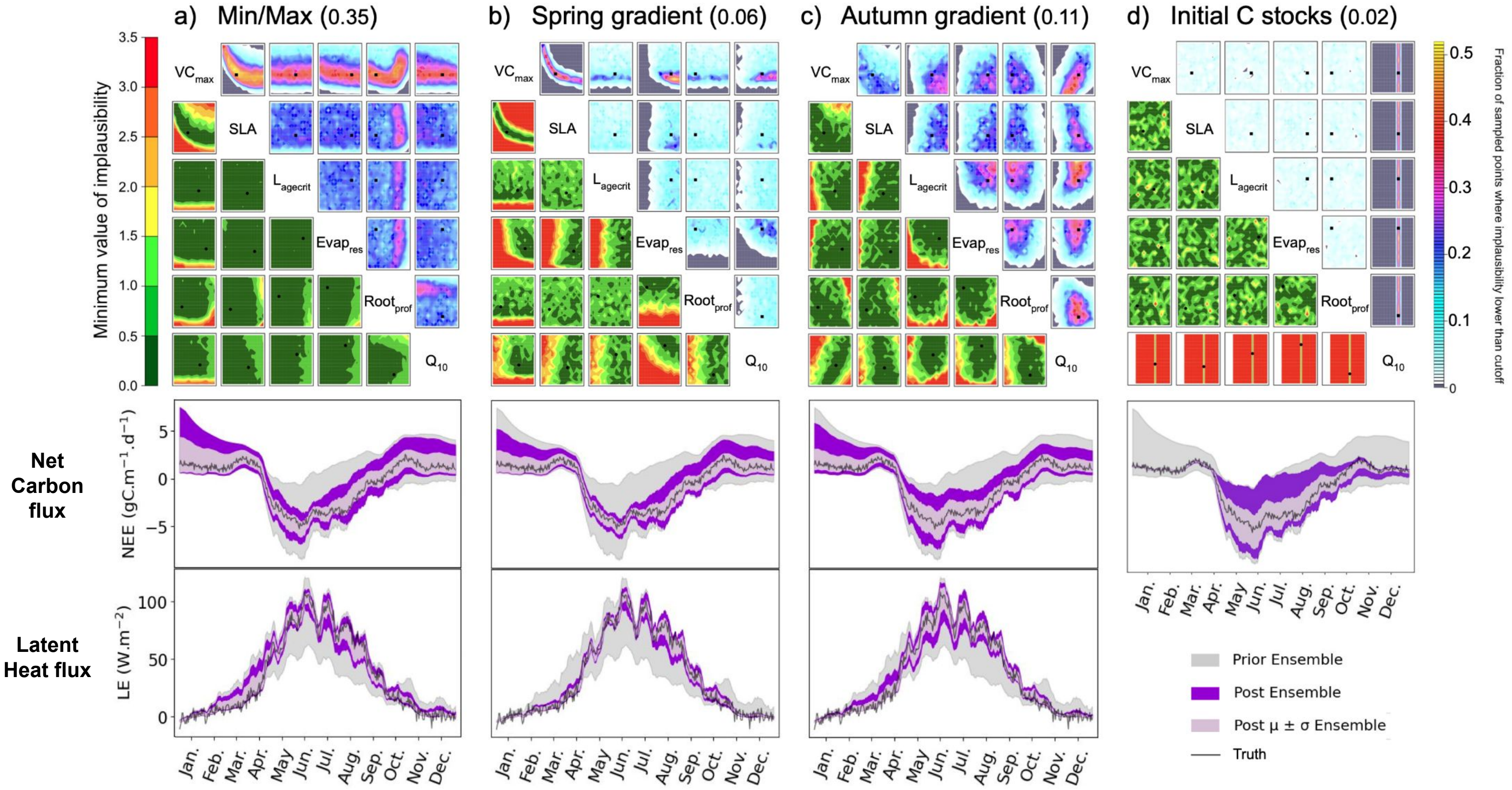
# Results : Parameter uncertainty

Spread of 200 optimisations

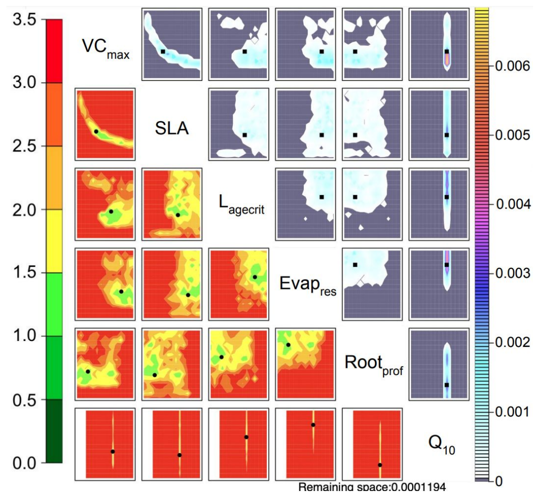


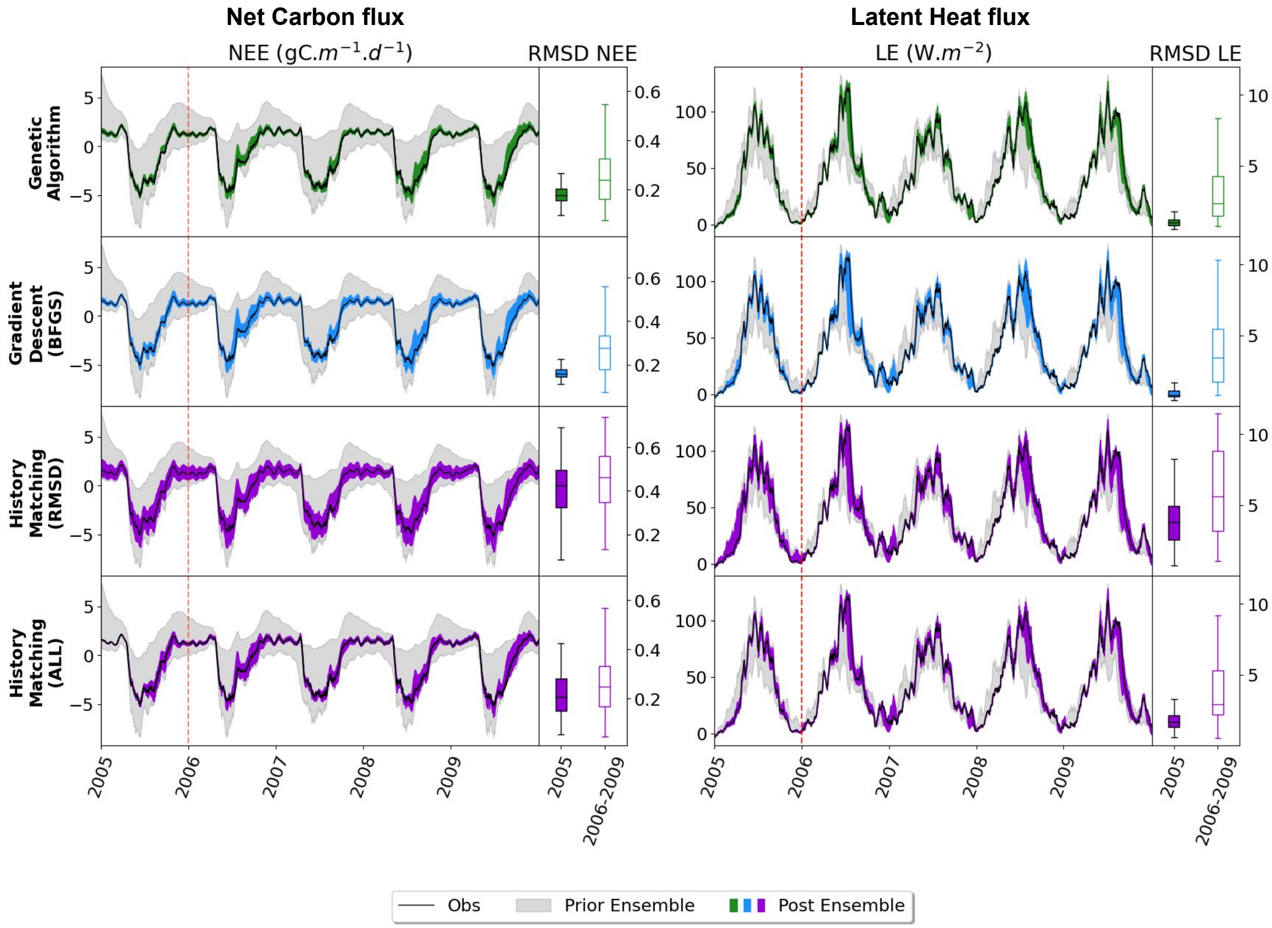



# Results : Multi metrics



# Results : Multi metrics



# Summary

- HM provides more information about parameter uncertainties
- HM does not overfit as much as DA
- Multi metric is very powerful when using HM

**All these results are in the article of Nina Raoult currently under review in GMD :**

## Exploring the Potential of History Matching for Land Surface Model Calibration

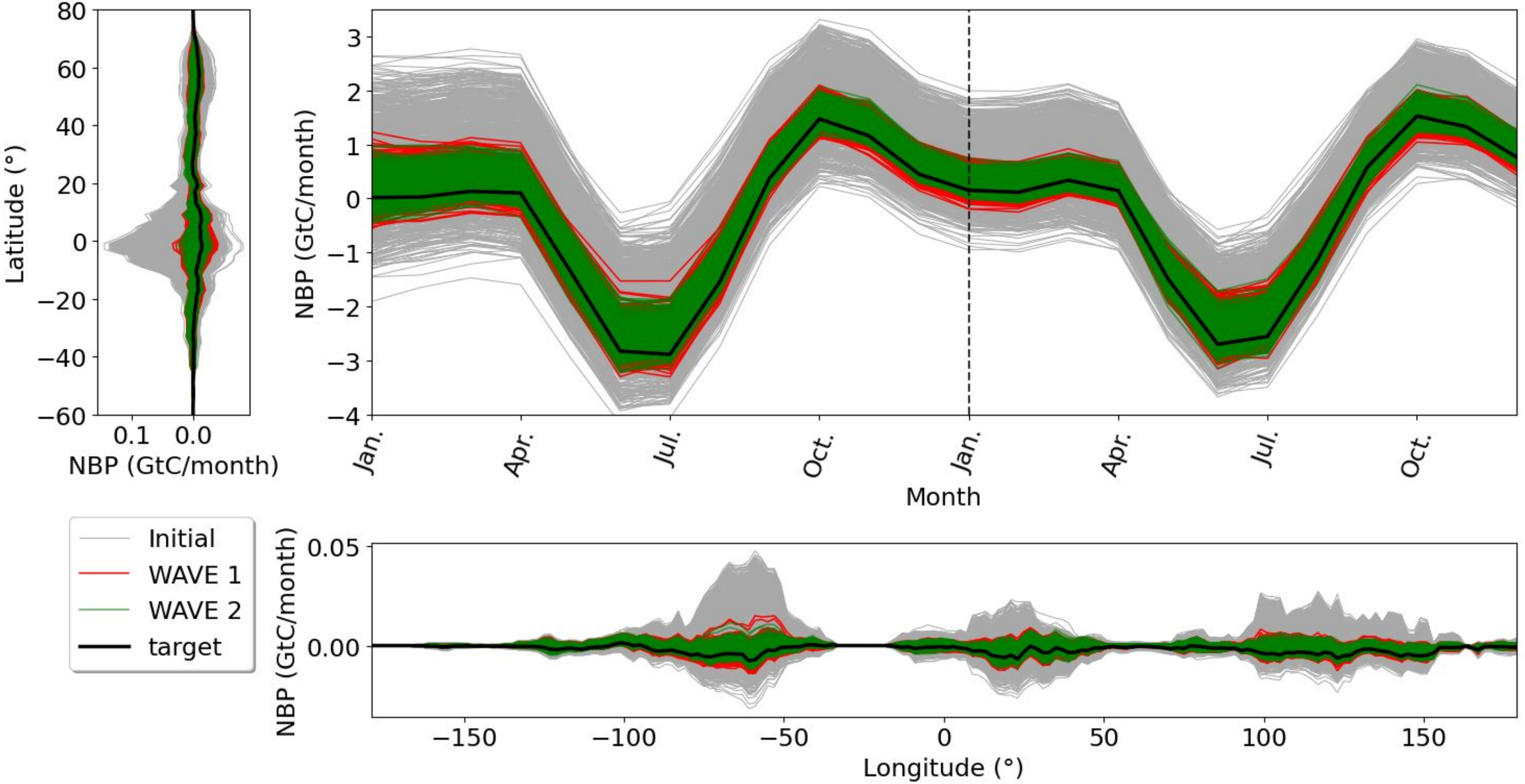
Nina Raoult[1], Simon Beylat[2,3], James M. Salter[1], Frédéric Hourdin[4], Vladislav Bastrikov[5], Catherine Ottlé[2], and Philippe Peylin[2]

[1]Department of Mathematics and Statistics, Faculty of Environment, Science and Economy, University of Exeter, Laver Building, North Park Road, Exeter, EX4 4QE, UK
[2]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, 91191, France
[3]School of Geography, Earth and Atmospheric Sciences, University of Melbourne, Victoria, Australia
[4]Laboratoire de Météorologie Dynamique, LMD/IPSL, Sorbonne Université, CNRS, École Polytechnique, ENS, Paris, 75005, France
[5]Science Partners, Paris, France

**Correspondence:** Nina Raoult (n.m.raoult2@exeter.ac.uk)

# work in progress : adjust the global carbon balance of ORCHIDEE

# First results : first wave with 120 parameters

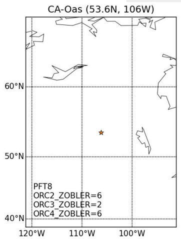



**WAVE 1 : 0.19%**
**WAVE 2 : 0.0459%**

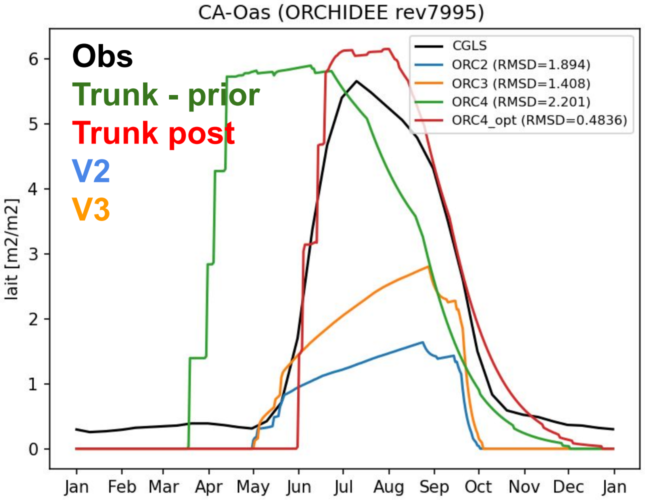
# Offline Calibration des paramètres ORCHIDEE-Trunk pour CMIP7

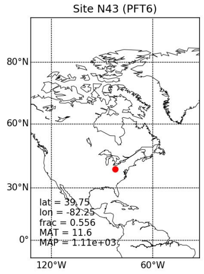
- Optimisation sur sites (FluxNet) : Multi-site calibration par PFT
- Obs: Flux C / LE / SH / ; Hauteur arbres ; LAI satellite ; Ratio biomasse C ; …
- Paramètres: Photosynthèse, Allocation C, Mortalité,...
- Approche standard avec Algorithme Génétique !

Ex: optimisation phenologie (LAI)



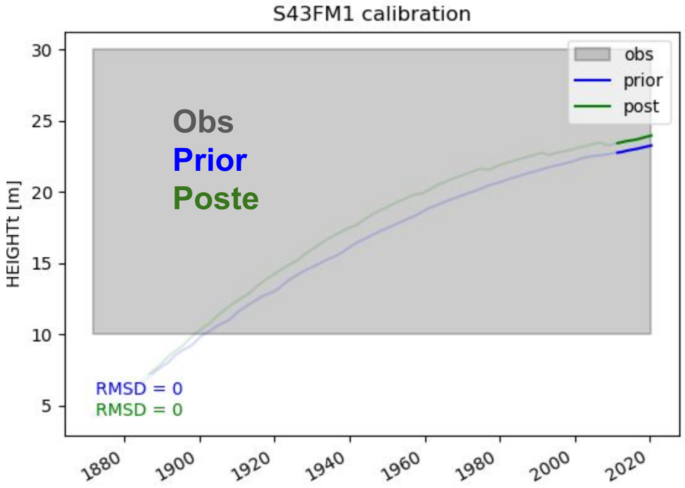

*Canada*




*USA*

Ex: optimisation hauteur arbres



# Offline Calibration des paramètres ORCHIDEE-Trunk pour CMIP7

- Optimisation Albedo (MODIS) et Couverture de Neige (ESA-CCI)

- Paramètres: 2 paramètres pour couverture neige, 2 albedo neige, 1 albedo veg

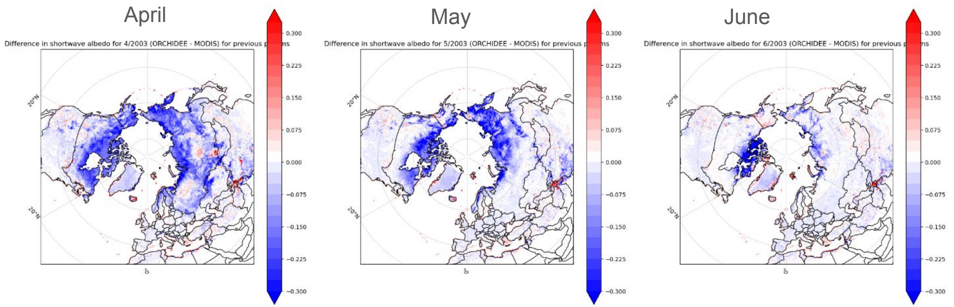
$$albedo = frac_{veg}[(1 - frac_{snow,veg})alb_{veg} + frac_{snow,veg}alb_{snow,veg}] + frac_{nobio}[(1 - frac_{snow,nobio})alb_{nobio} + frac_{snow,nobio}alb_{snow,nobio}]$$

$$alb_{snow,veg} = \frac{\sum_{pft=1}^{15} frac_{max,pft} \times alb_{snow,pft}}{\sum_{pft=1}^{15} frac_{max,pft}}$$

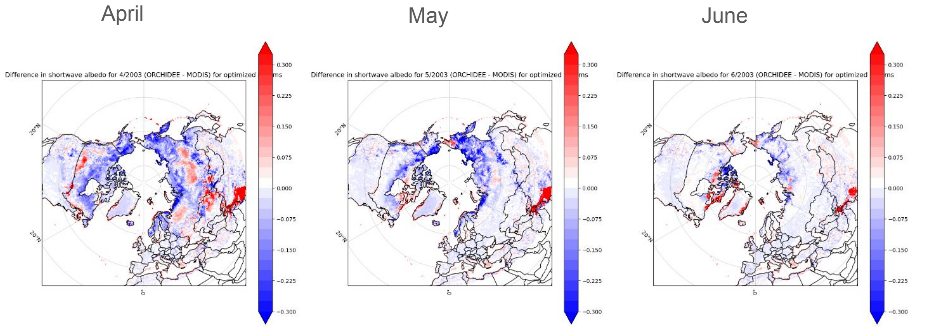
$$alb_{snow,pft} = alb_{snow,aged,pft} + alb_{snow,dec,pft} \times e^{-age_{snow}/tcst_{snow}}$$

- Approche standard avec Algorithme Génétique !

## Albedo : Prior difference with MODIS



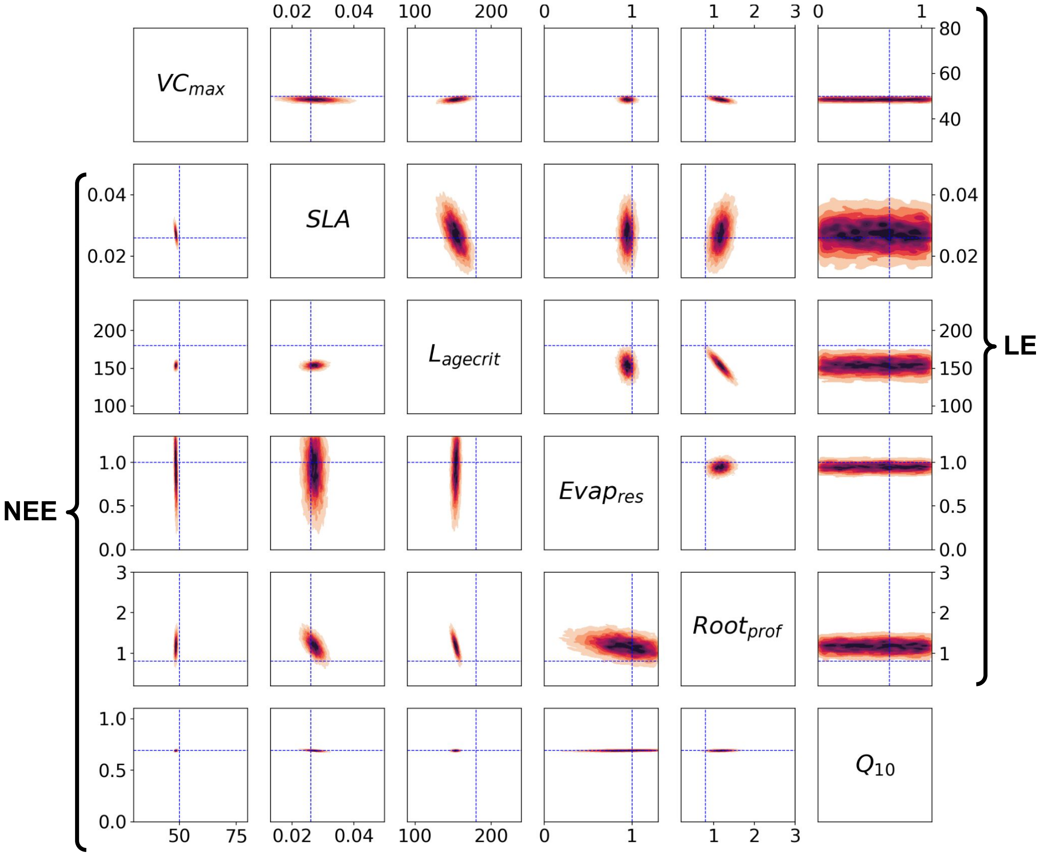

## Albedo : Posterior difference with MODIS



**Thank you for your attention !**

# Results : Parameter uncertainty

## Using Bpost matrix




Posterior Covariance operator :

$$B_{post} = [H^T R^{-1} H + B^{-1}]^{-1}$$

$B_{post}$ diagonal elements are variances, the root square of diagonal elements can be seen as “**uncertainty bars**”.

$B_{post}$ off-diagonal elements (which represent the covariances), it is easier to check correlations:

$$Corr^{\alpha,\beta} = B_{post}^{\alpha,\beta} / (\sigma^{\alpha}\sigma^{\beta})$$

If $Corr^{\alpha,\beta}$ close to 0, the uncertainty of $m^{\alpha}$ and $m^{\beta}$ is uncorrelated.
If $Corr^{\alpha,\beta}$ close to ±1, the uncertainty of $m^{\alpha}$ and $m^{\beta}$ is correlated.

**To compute uncertainty**, let's generate pseudorandom gaussian sample of m (set of parameters) using $m_{post}$ as mean and $B_{post}$ as covariance matrix.

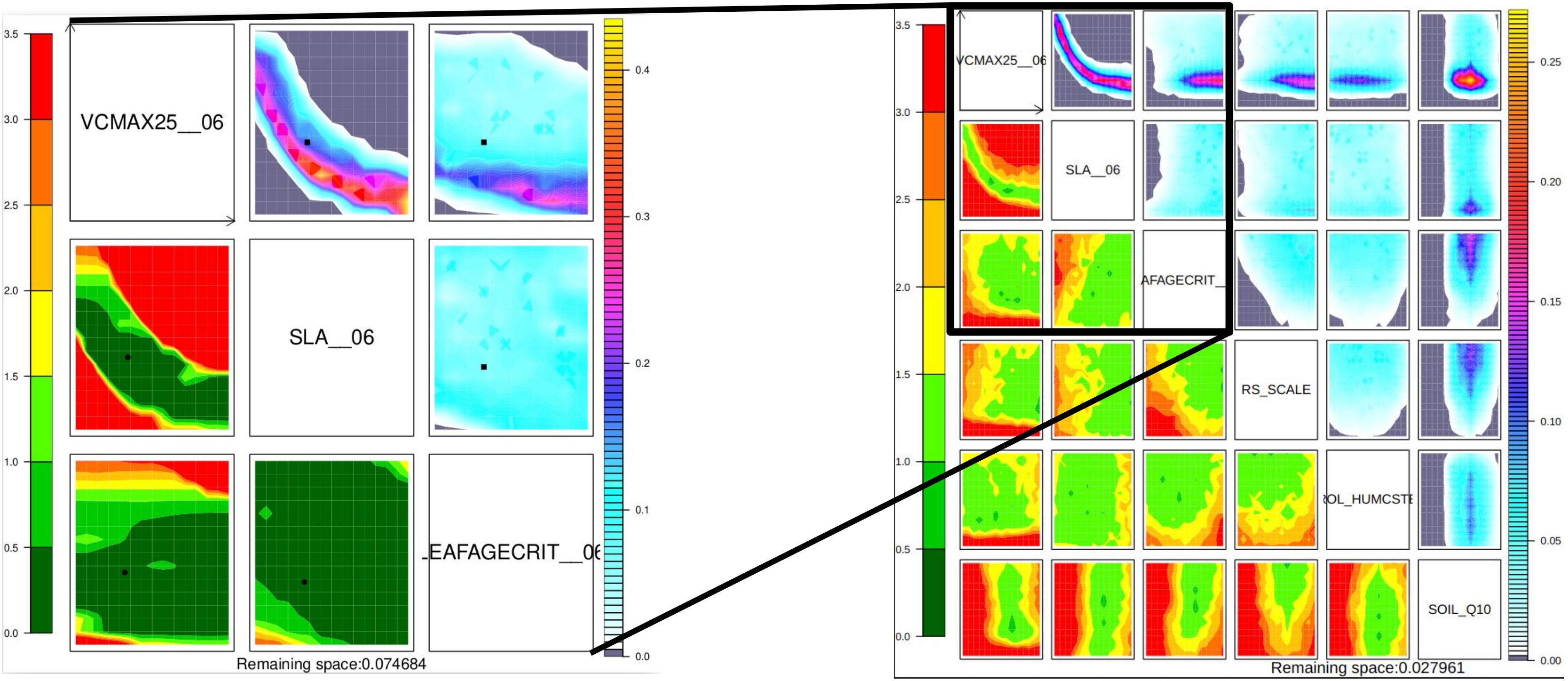
**Bpost decomposition:**

$$B_{post} = \left[\sum_{i=0}^{D} H_i^T \sigma_i^{-1} H_i + B^{-1}\right]^{-1} = \left[H_0^T \sigma_0^{-1} H_0 + H_1^T \sigma_1^{-1} H_1 + \ldots + H_D^T \sigma_D^{-1} H_D + B^{-1}\right]^{-1}$$

# Final word : Sensitive analysis + Structure of the model

Global - metrics Global net flux

FR-FON - metrics (9 metrics used)

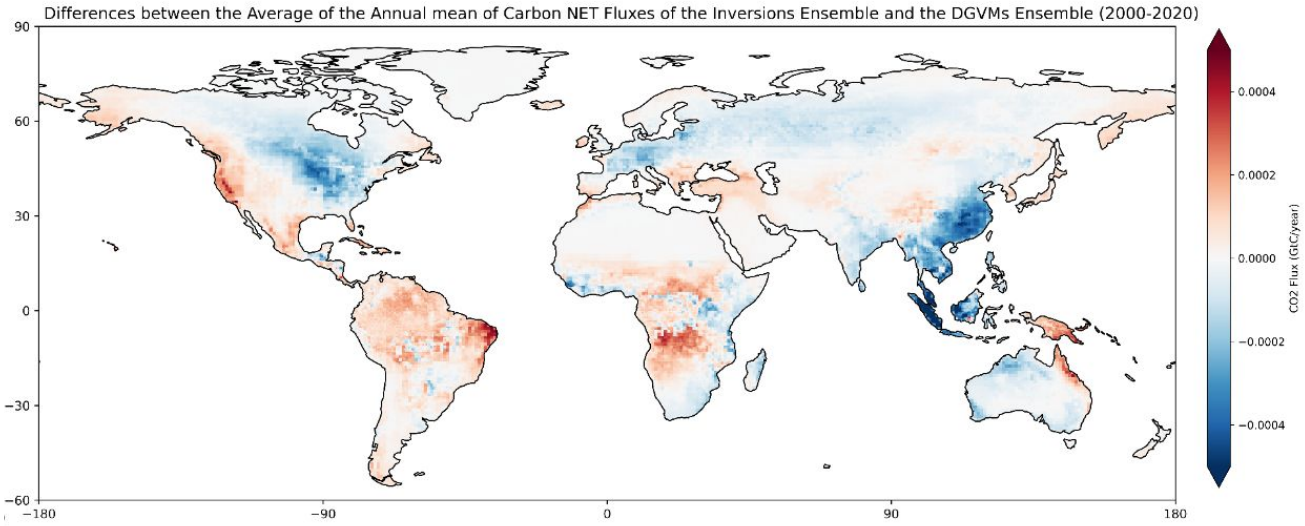


# work in progress : adjust the carbon balance.

**Understand the Land Carbon sink:**

**Top down : Atmospheric Inversion**
**Bottom up : Land Surface model**

Using HM to calibrate ORCHIDEE off-line on Atmospheric Inversion first. *(and then used LMDz to transport CO2 concentration)*




Data collected from TrendyLand V10 and GCP2021

**Difficulty :**

**Calibrating all the pft means a large number of parameters (between 100 and 200).**

# Final word : Sensitive analysis

## Heterotrophe Respiration