



LSCE

LABORATOIRE DES SCIENCES DU CLIMAT
& DE L'ENVIRONNEMENT



Data Assimilation Methods for Parameter Calibration in the ORCHIDEE Land Surface Model

Simon Beylat (LSCE)

Philippe Peylin (LSCE) , Catherine Ottle (LSCE), Nina Raoult (ECMWF)

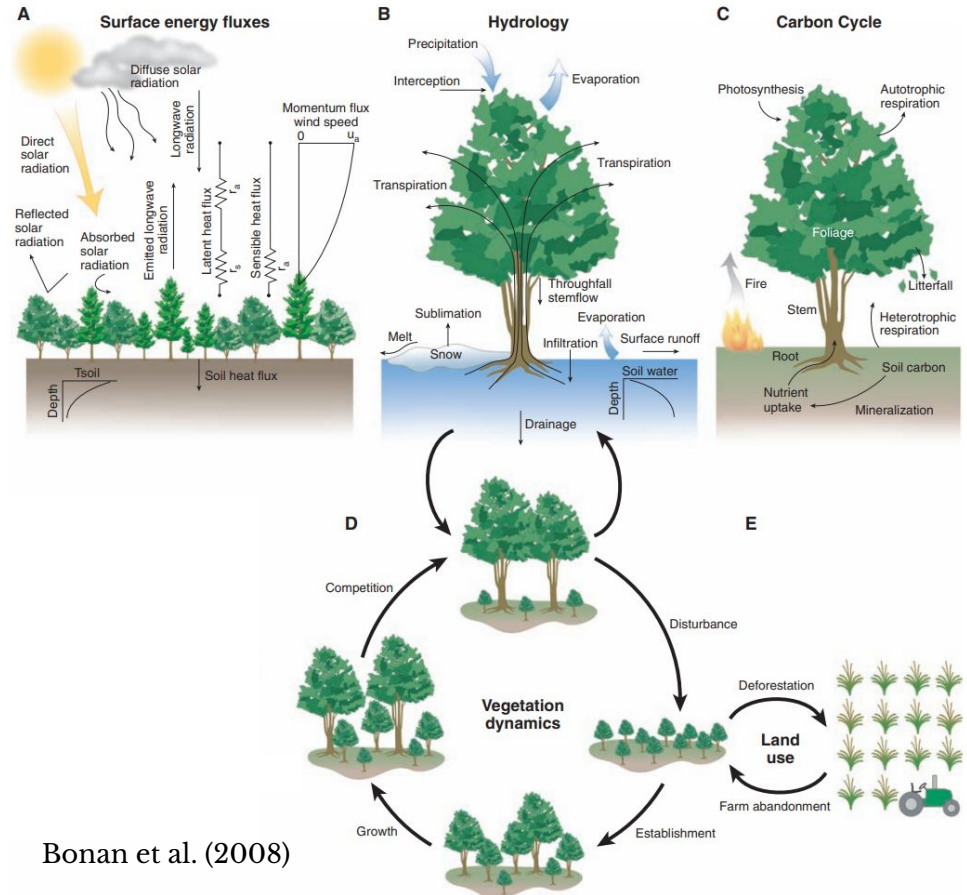
and the ORCHIDAS teams

simon.beylat@lsce.ipsl.fr

Land Surface Model

Principle of a land surface model:

- **Mathematically represents** water, heat and carbon exchanges between the land and the atmosphere.

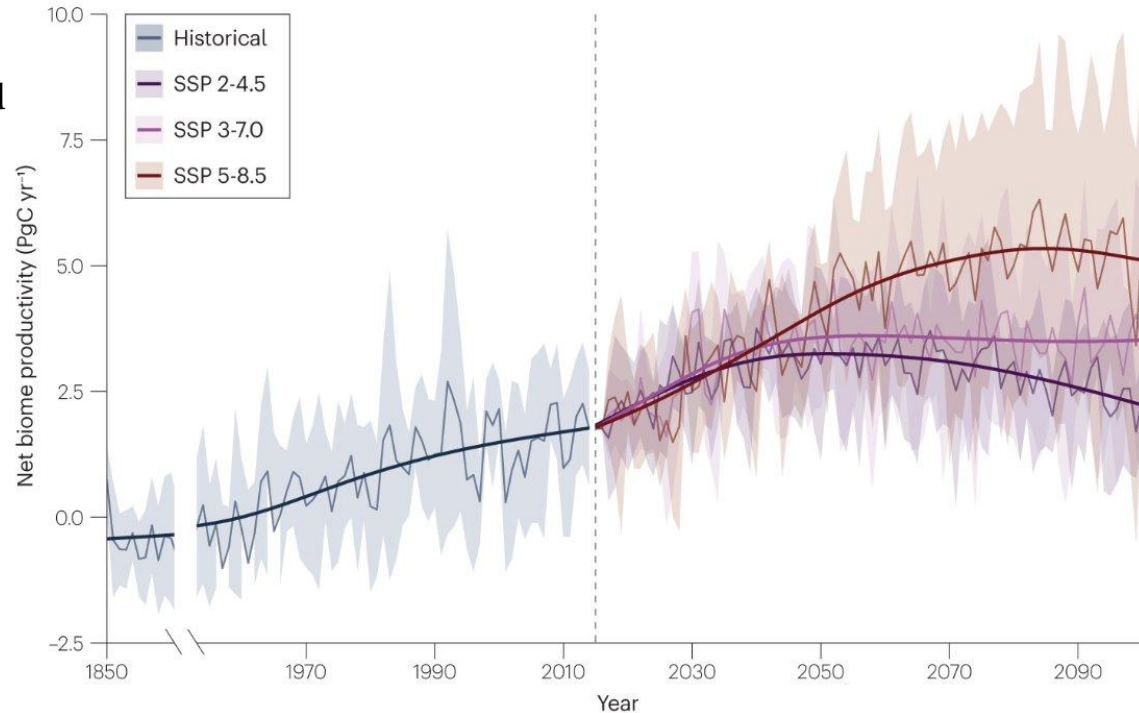


Bonan et al. (2008)

Land Surface Model

Principle of a land surface model:

- **Mathematically represents** water, heat and carbon exchanges between the land and the atmosphere.
- Can be **coupled** with a general circulation model or **climate** model for climate and weather simulations.



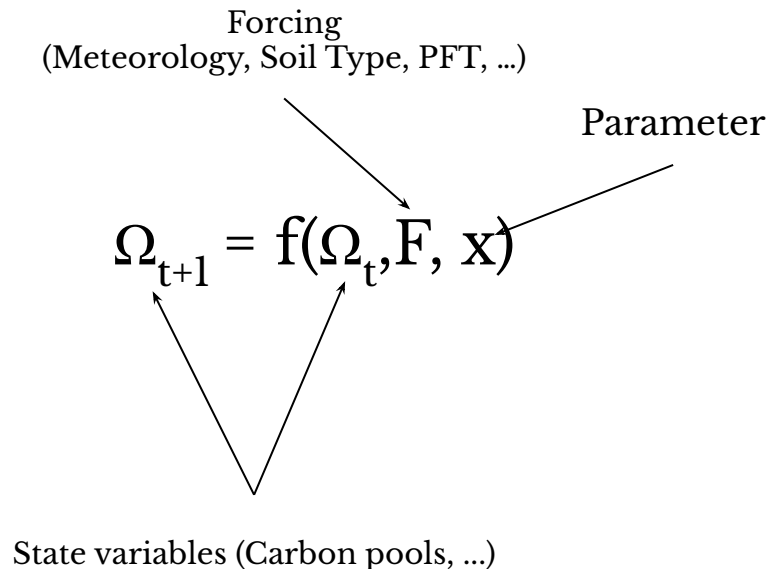
Ruehr et al. (2023)

Land Surface Model

Principle of a land surface model:

- **Mathematically represents** water, heat and **carbon exchanges** between the land and the atmosphere.
- Can be coupled with a general circulation model for climate and weather simulations.
- Uses **parameterisation**, as most processes are:
 - smaller in scale than typical model resolutions;
 - not well understood, so empirical representations are often used.

Accurate parameter calibration is essential for reliable LSM simulations.



Many studies of Data Assimilation for Parameter Calibration

Review Papers :

Parameter Estimation in Land Surface Models: Challenges and Opportunities With Data Assimilation and Machine Learning

Nina Raoult^{1,2}, Natalie Douglas³, Natasha MacBean^{4,5}, Jana Kolassa^{6,7}, Tristan Quaife³, Andrew G. Roberts⁸, Rosie Fisher⁹, Istem Fer¹⁰, Cédric Bacour¹¹, Katherine Dagon¹², Linnia Hawkins¹³, Nuno Carvalhais^{14,15,16}, Elizabeth Cooper¹⁷, Michael C. Dietze¹⁸, Pierre Gentile¹³, Thomas Kaminski¹⁹, Daniel Kennedy¹², Hannah M. Liddy^{20,21}, David J. P. Moore²², Philippe Peylin¹¹, Ewan Pinnington², Benjamin Sanderson⁹, Marko Scholze²³, Christian Seiler²⁴, T. Luke Smallman²⁵, Noemi Vergopolan^{26,27}, Toni Viskari²⁸, Mathew Williams²⁵, and John Zobitz²⁹

More than 80 studies

ORCHIDAS :

ORCHIDEE Data Assimilation Systems

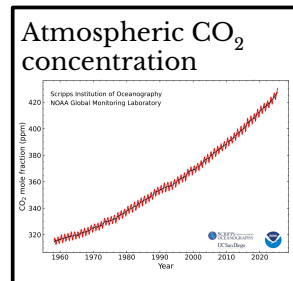
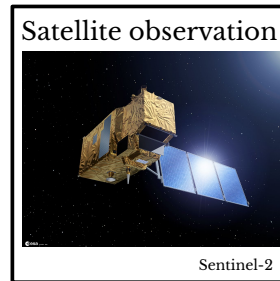


ORCHIDEE
LAND SURFACE MODEL

Santaren et al. (2007, 2014)
Kuppel et al. (2012, 2013, 2014)
Bacour et al. (2015, 2019a, 2019b)
MacBean et al. (2015, 2016, 2018)
Raoult et al. (2018, 2021, 2023,
2024a, 2024b, 2024c)

...

Typical observations :



And more (Above ground biomass, snow albedo, ...)

Other Land Surface Models:

CLM (US)
JULES (UK)
JSBACH (German)
LPJ-GUESS (Sweden)
Many others...

Bayesian inversion framework

$$p(\mathbf{x}|\mathbf{y}) \propto \boxed{p(\mathbf{y}|\mathbf{x})}, \boxed{p(\mathbf{x})}$$

$\mathbf{x}$ Parameters vector

$\mathbf{y}$ Observations vector

$$J(\mathbf{x}) = \underbrace{\frac{1}{2}(\mathcal{H}(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}) - \mathbf{y})}_{\text{Observation term}} + \underbrace{\frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b)}_{\text{Background term}}$$

Model operator

Prior knowledge

Minimisation of the cost function

$$J(\mathbf{x}) = \frac{1}{2}(\mathcal{H}(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}) - \mathbf{y}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b)$$

Variational method

$$\nabla J(\mathbf{x}_i) = \mathbf{H}^T \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}) + \mathbf{B}^{-1}(\mathbf{x}_i - \mathbf{x}_b)$$

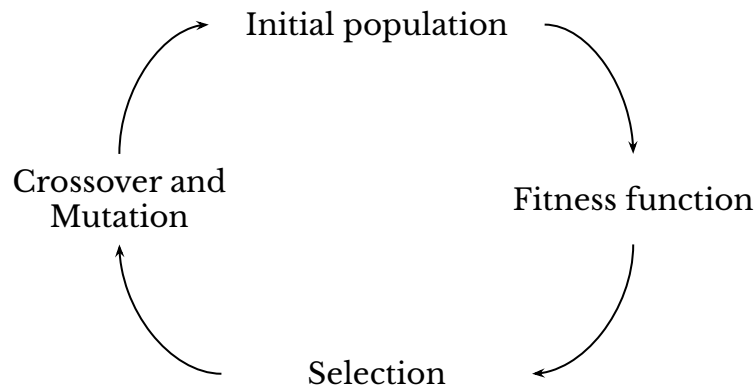
Adjoint model

Finite difference estimation :

$$\mathbf{H} \approx \frac{\mathcal{H}(\mathbf{x} + \Delta \mathbf{x}) - \mathcal{H}(\mathbf{x})}{\Delta \mathbf{x}}$$

Tarantola (1987, 2005); Santaren et al. (2007); Kuppel et al. (2012)

Monte Carlo (Genetic algorithm)



Goldberg (1989); Haupt and Haupt (2004); Santaren et al. (2014)

ORCHIDAS : ORCHIDEE Data Assimilation Systems

Variational method

Local Vision



- Fast;
 - Gives one solution;
 - Can be stuck in local minima;
 - Provides Gaussian uncertainty.
- + Highly effective for simple problems
- Needs TL/Adjoint.



ORCHIDAS

Monte Carlo - Genetic Algorithm

Global Vision



- Expensive;
 - Gives one solution;
 - Not sensible to local minima;
 - Does not provide uncertainty.
- + Explores all the parameter space
- Can overfit.

History Matching

Global Vision



- Expensive;
 - Provides regions of parameter space;
 - Not sensible to local minima;
 - Quantifies well uncertainty.
- + Can easily use multi-criteria.
- Can not find the best parameter set.

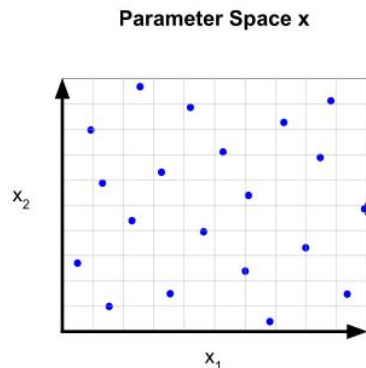


Raoult, Beylat et al. (2024)

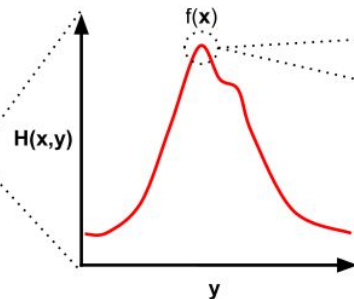
History Matching

1

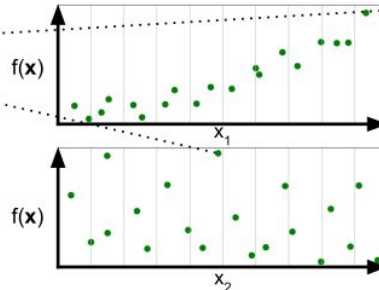
Run the model N times with parameters sampled from a Latin hypercube sampling.



Model Output Space $H(x)$



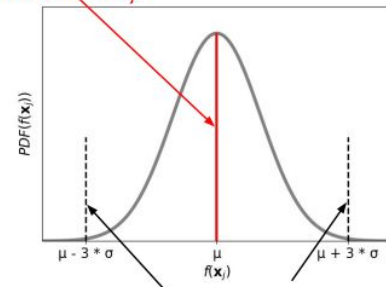
Metric f against parameter



3

Build an emulator that links the parameter to the metric using the sample.

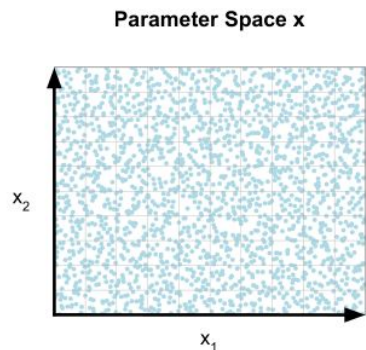
The mean μ is the prediction of $f(x_i)$



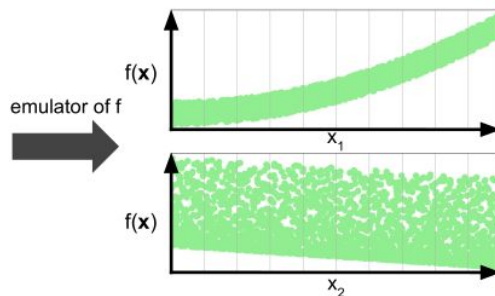
The standard deviation σ is the uncertainty of our prediction.

4

Run the emulator a large number of times in order to completely fulfil the parameter space.



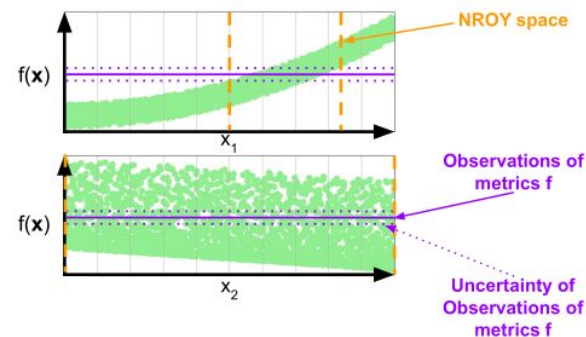
Metric f against parameter



5

Rule out region of parameter using observations of metrics f .

Metric f against parameter



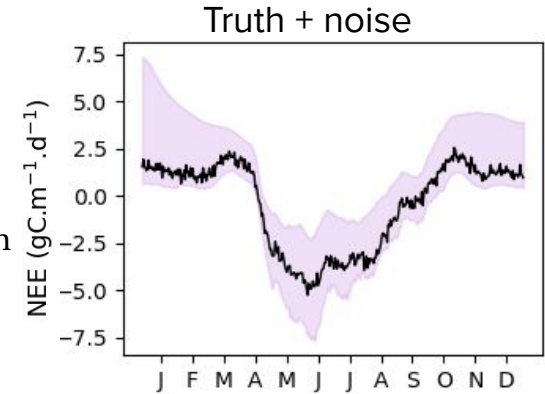
Experiment to compare methods



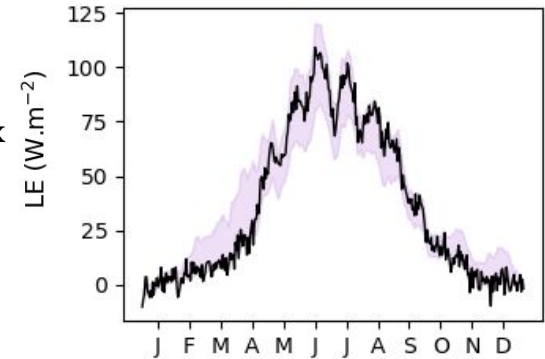
Site: FR-Fon (2005)



NEE :
Net Ecosystem
Exchange Net Carbon
flux



LE : Latent heat flux

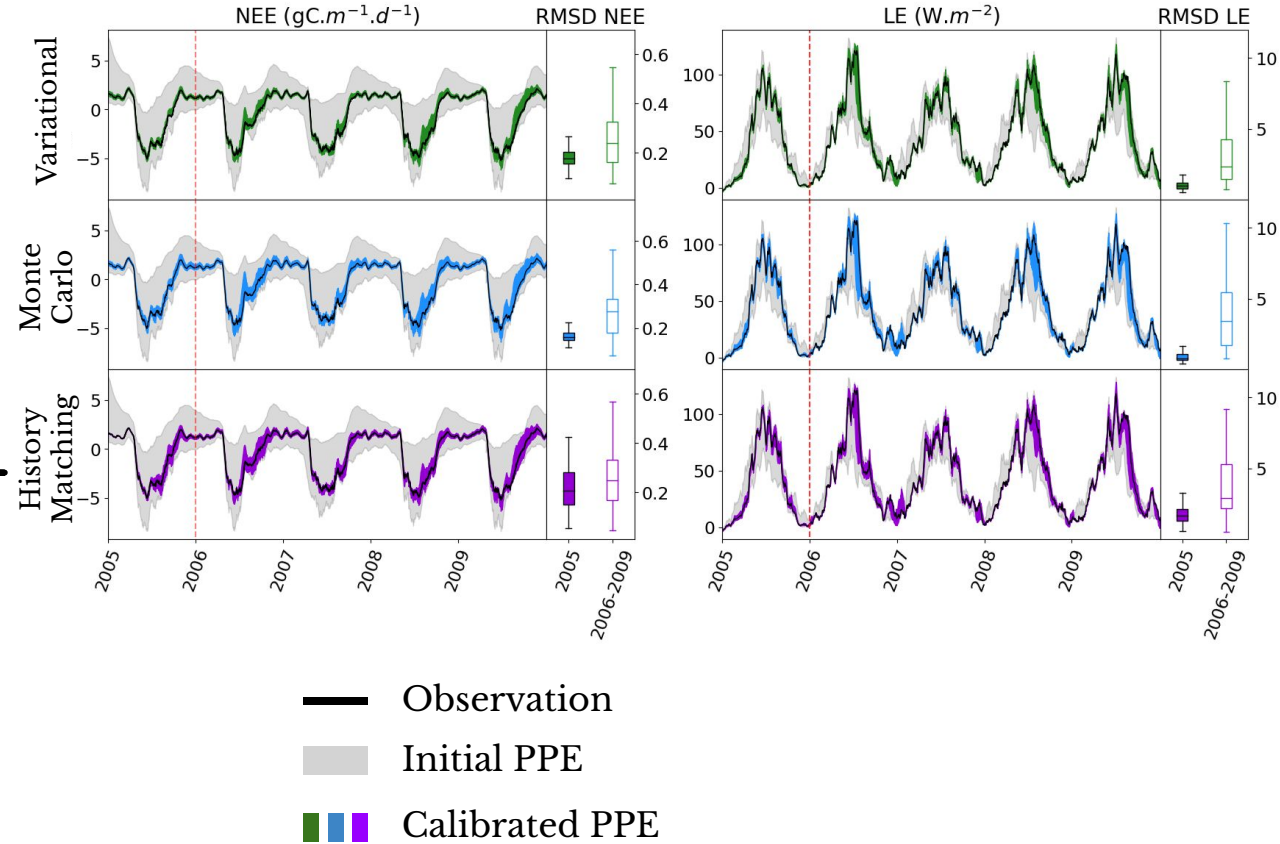
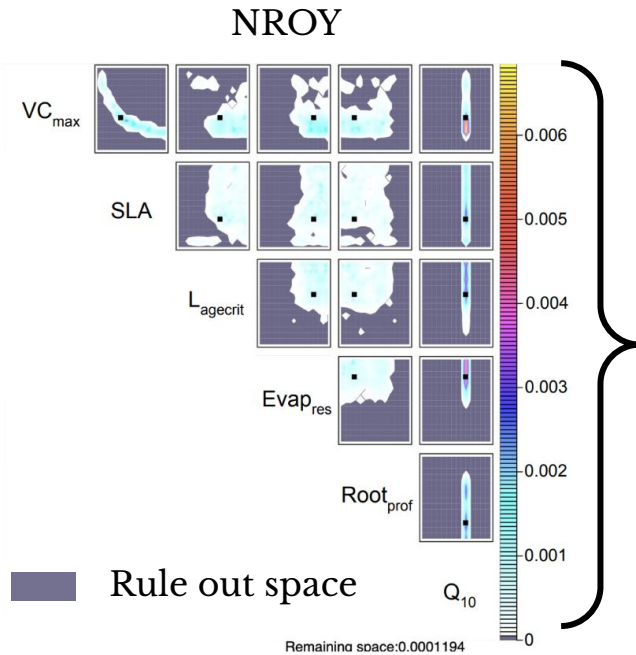


6 parameters used :

	Description
VC _{max}	Maximum carboxylation rate ($\mu\text{molm}^{-1}\text{s}^{-1}$)
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 (-)

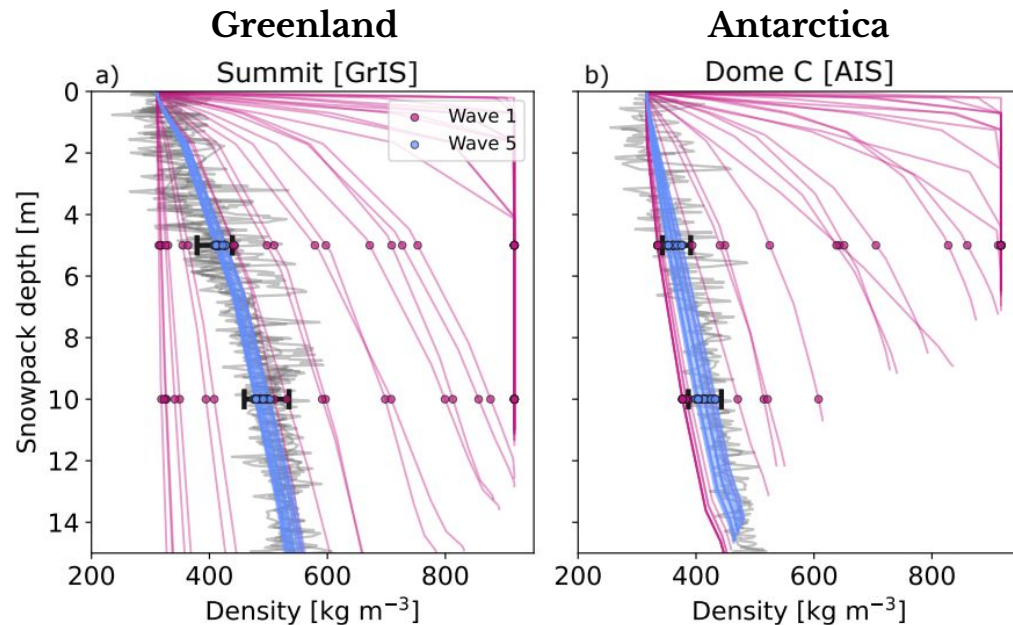
Need to compare Ensemble :
VarDA or GA : 200 optimisations using
random priors.
1 ensemble of History Matching (HM)

History Matching results

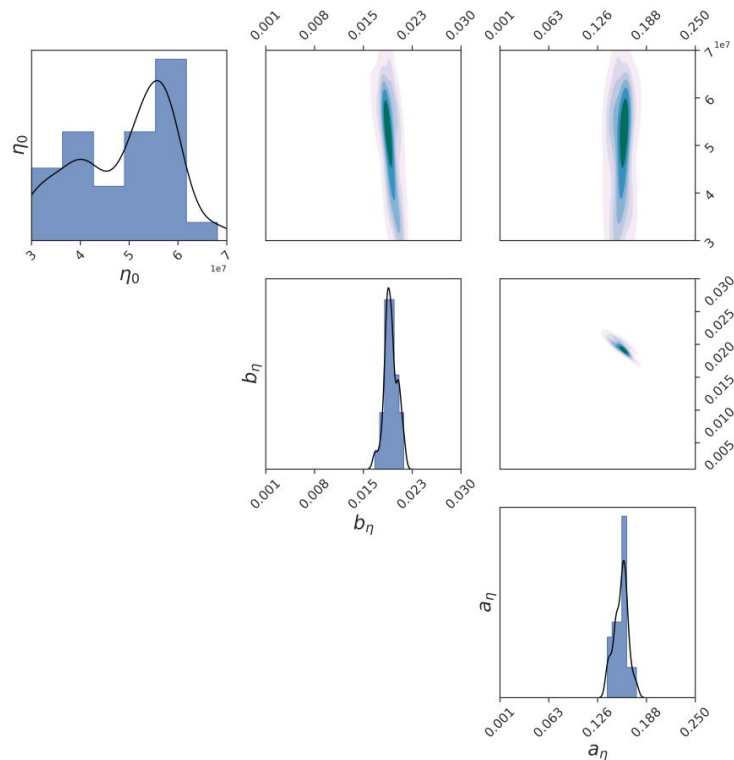


Application Case : Dry snow densification over the Greenland and Antarctic ice sheets in the ORCHIDEE land surface model

Results:



Snow Viscosity Parameter
uncertainty:



ORCHIDAS : ORCHIDEE Data Assimilation Systems

Variational method

Local Vision

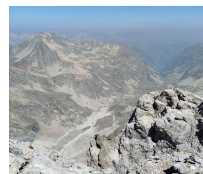


- Fast;
 - Gives one solution;
 - Can be stuck in local minima;
 - Provides Gaussian uncertainty.
- + Highly effective for simple problems
- Needs TL/Adjoint.

Ensemble Variational method



Hybrid vision



- Fast;
 - Gives one solution;
 - Less sensible to local minima;
 - Provides Gaussian uncertainty.
- + Good with many parameters and costly simulation
- Strong linear assumption.

Beylat et al. (2025); Douglas, Beylat et al. (In Review)



ORCHIDAS

Monte Carlo - Genetic Algorithm

Global Vision



- Expensive;
 - Gives one solution;
 - Not sensible to local minima;
 - Does not provide uncertainty.
- + Explores all the parameter space
- Can overfit.

History Matching



Global Vision



- Expensive;
 - Provides regions of parameter space;
 - Not sensible to local minima;
 - Quantifies well uncertainty.
- + Can easily use multi-criteria.
- Can not find the best parameter set.

Raoult, Beylat et al. (2024)

Ensemble Variational method

Variational Cost Function $J(\mathbf{x}) = \frac{1}{2}(\mathcal{H}(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}) - \mathbf{y}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b)$

Generating a ensemble:

$$\mathbf{X}'_b = \frac{1}{\sqrt{N-1}}(\mathbf{x}_1 - \mathbf{x}_b; \mathbf{x}_2 - \mathbf{x}_b; \dots; \mathbf{x}_N - \mathbf{x}_b)$$

$$\mathbf{x}_i = \mathbf{x}_b + \mathbf{X}'_b \mathbf{w}$$

Ensemble Var Cost Function :

$$J(\mathbf{w}) = \frac{1}{2}(\mathbf{H}\mathbf{X}'_b \mathbf{w} + \mathcal{H}(\mathbf{x}_b) - \mathbf{y})^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{X}'_b \mathbf{w} + \mathcal{H}(\mathbf{x}_b) - \mathbf{y}) + \frac{1}{2}\mathbf{w}^T \mathbf{w}$$

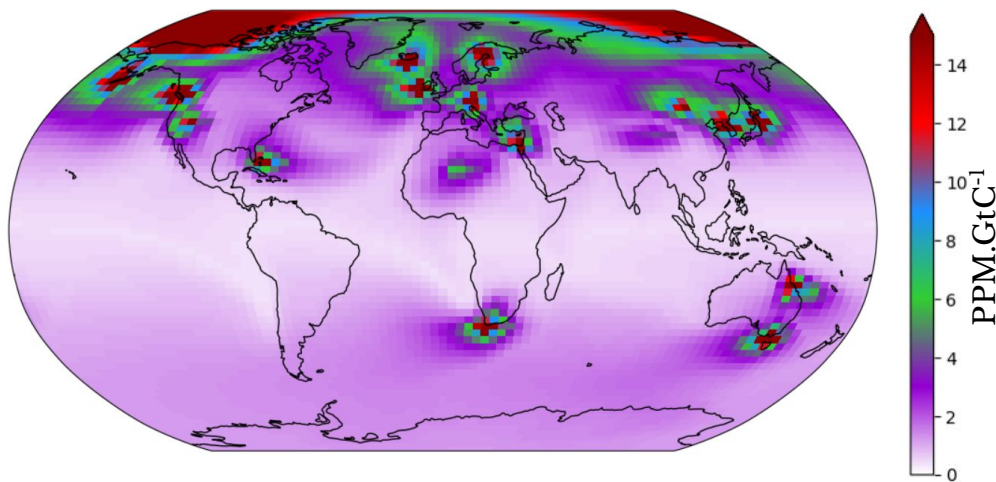
$$\nabla J(\mathbf{w}) = (\mathbf{H}\mathbf{X}'_b)^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{X}'_b \mathbf{w} + \mathcal{H}(\mathbf{x}_b) - \mathbf{y}) + \mathbf{w}$$

$$\mathbf{H}\mathbf{X}'_b \approx \frac{1}{\sqrt{N-1}}(\mathcal{H}(\mathbf{x}_1) - \mathcal{H}(\mathbf{x}_b); \mathcal{H}(\mathbf{x}_2) - \mathcal{H}(\mathbf{x}_b); \dots; \mathcal{H}(\mathbf{x}_N) - \mathcal{H}(\mathbf{x}_b))$$

Comparison of the Variational Method and Ensemble Variational Method

Twin experiment : Pseudo observation created using ORCHIDEE.
We know the true parameters.

Atmospheric stations:
21 stations across the continent

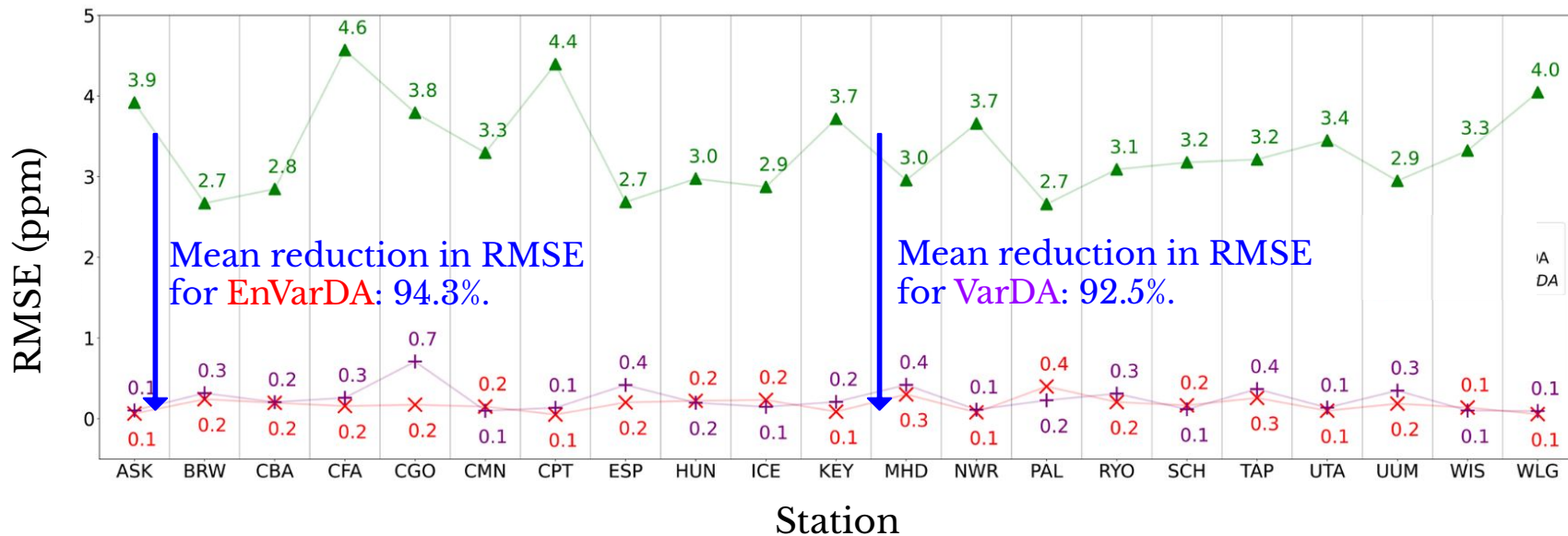


Set-up:
Assimilation of two years of data
Pre-computed transport model

57 Parameters :

- Photosynthesis
 - V_{cmax} for each PFT
 - SLA for each PFT
- Phenology
 - LAI for each PFT
- Autotrophic respiration
 - $m_{maint.resp}$ for each PFT
- Heterotrophic respiration
 - Q10

Results

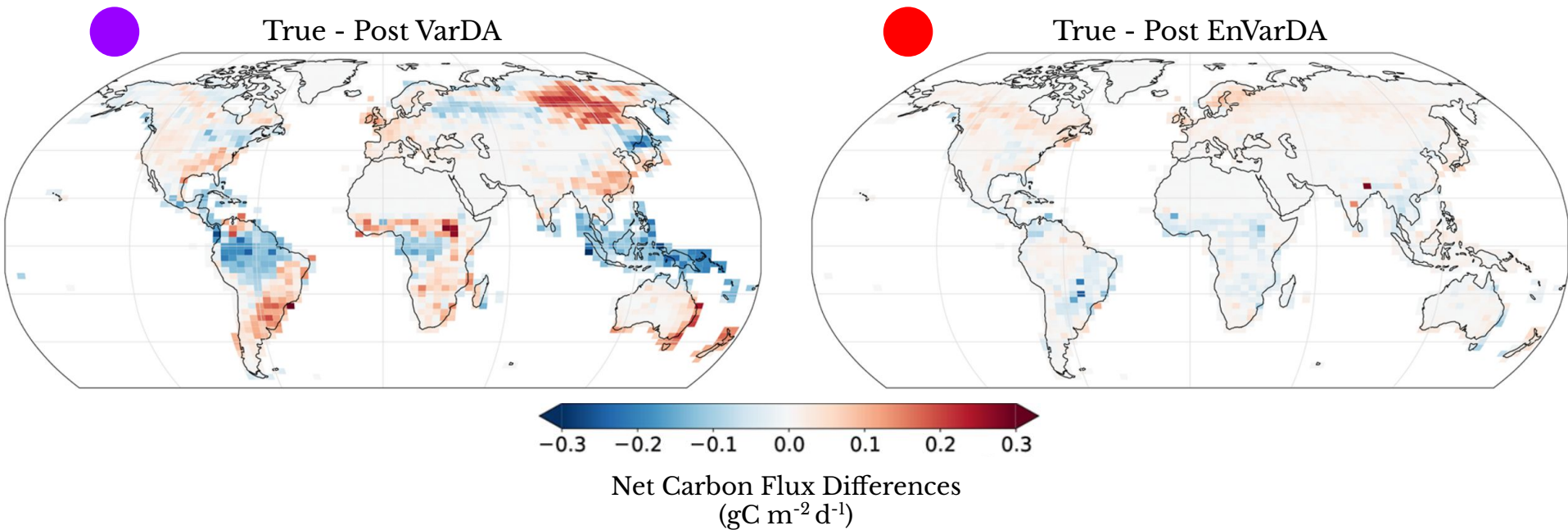


▲ Prior

✗ Post EnVarDA (Ensemble Variational method)

✚ Post VarDA (Variational method)

Results



25 iterations =
1450 simulations $\approx$ 265 hours

-80%

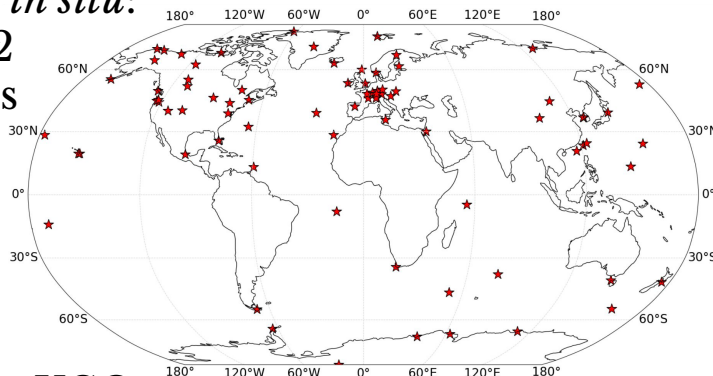
300 simulations $\approx$ 55 hours

Application Case:

Assimilation of atmospheric CO₂ Concentration data to calibrate the surface net carbon fluxes

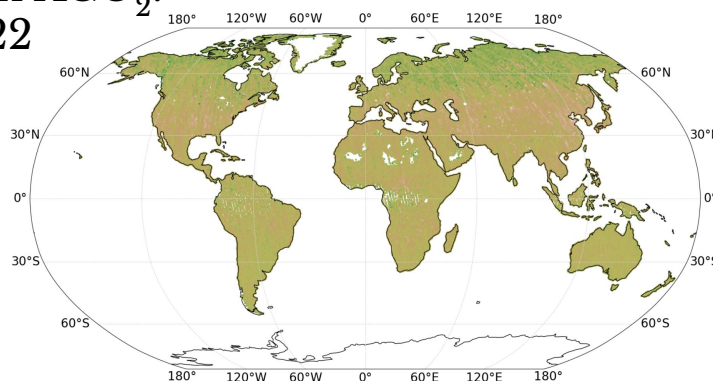
1) Assimilation *in situ*:

- 2018-2022
- 80 stations

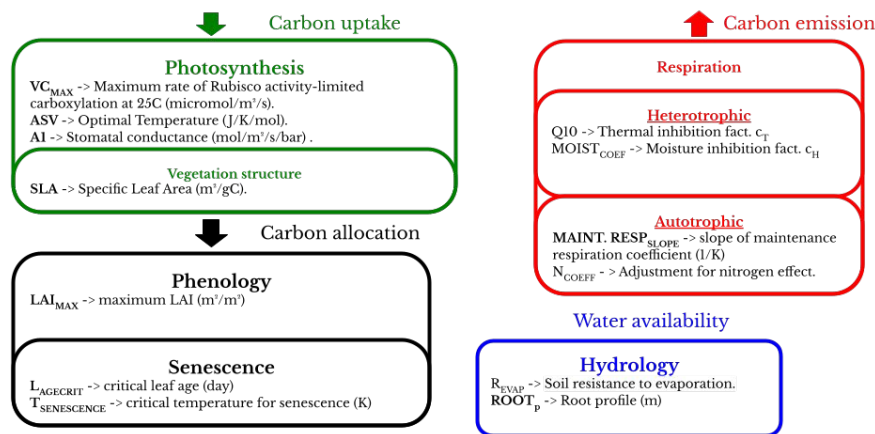


2) Assimilation XCO₂:

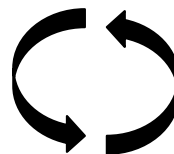
- 2018-2022
- OCO-2



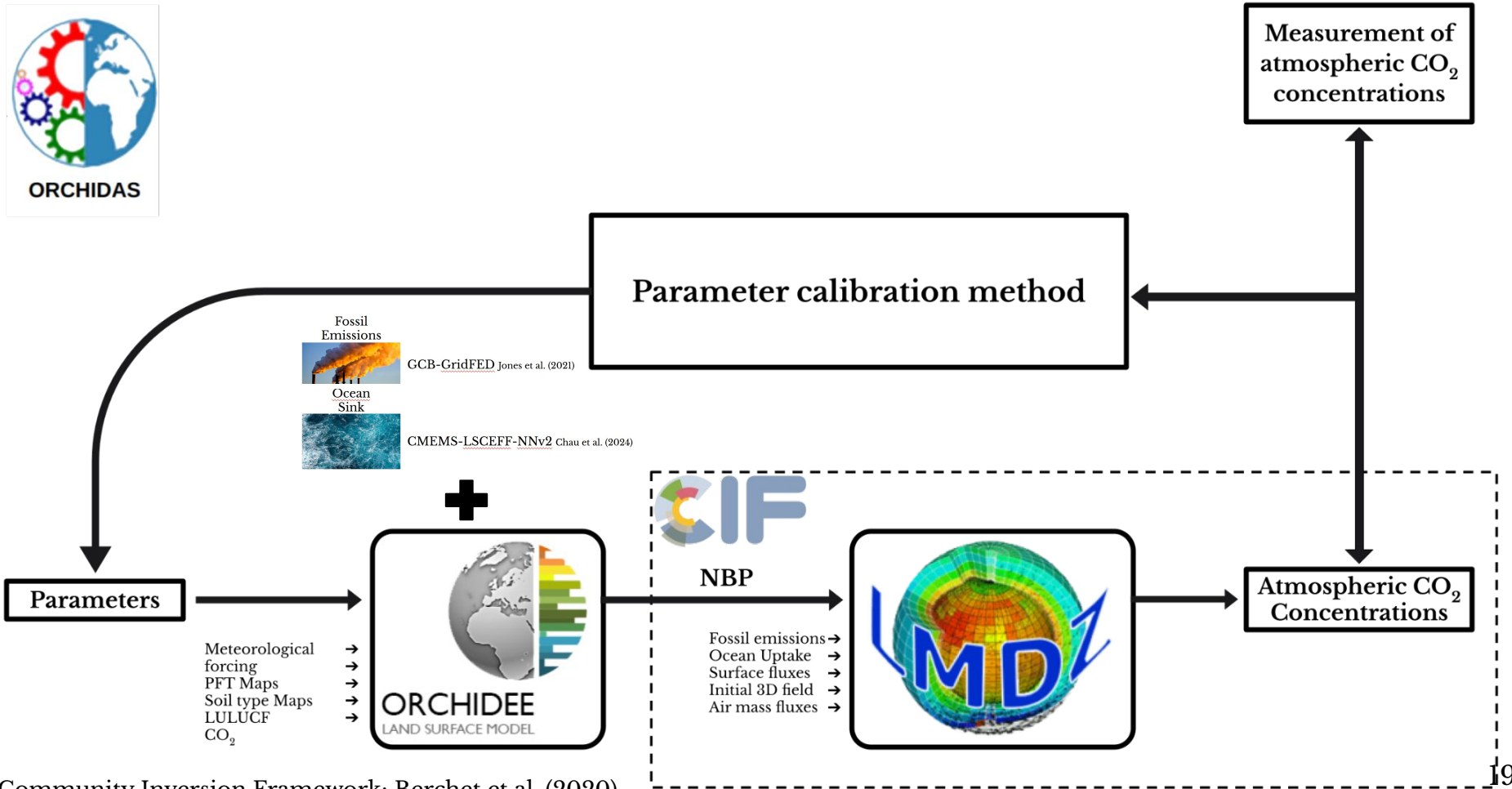
Calibration of the 125 ORCHIDEE parameters:



Pre-spin up of 5 years

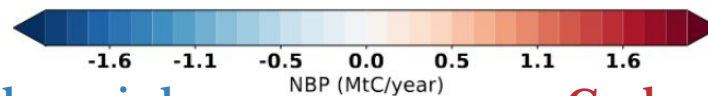
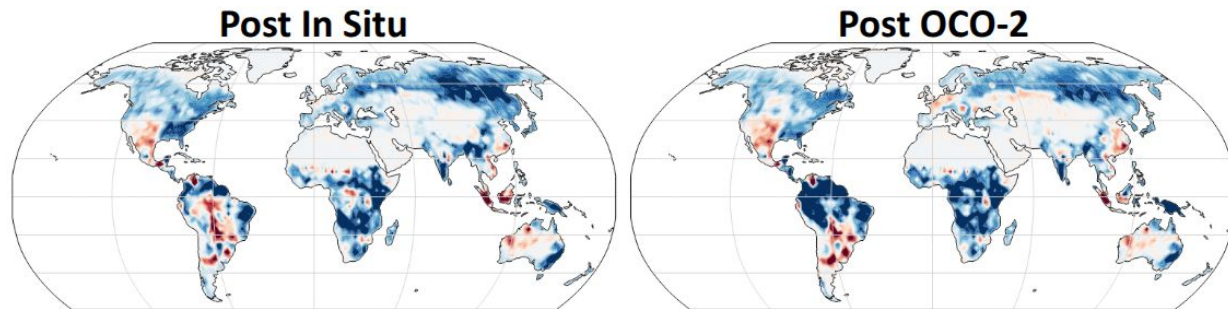


Set-up



Results

Map of mean net carbon fluxes between 2018 and 2022:

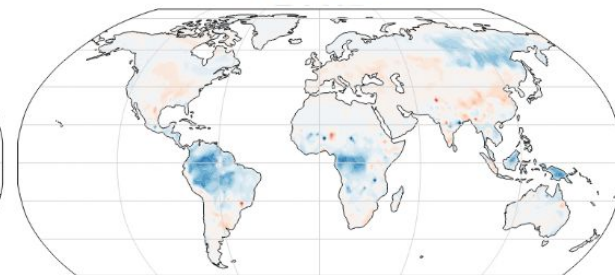
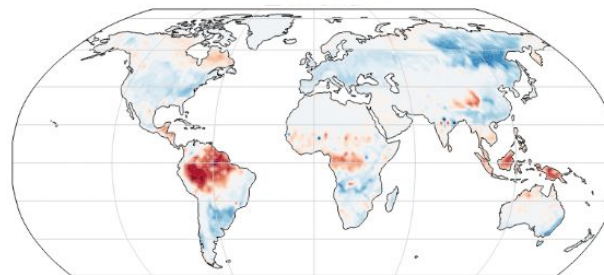


Carbon sink

Carbon source

Δ In Situ

Δ OCO-2



Increase in
Carbon sink

Decrease in
Carbon sink

Reducing average
model-observation biases:

In Situ: 0.99 ppm $\rightarrow$ 0.1 ppm
OCO-2: 1.0 ppm $\rightarrow$ 0.3 ppm

Conclusion

- Parameter calibration remains central to land surface modelling for improving performance and quantifying uncertainty;
- Most existing methods are either computationally expensive or technically demanding (requiring tangent linear and adjoint models);
- Emulators as a promising approach
 - + Well suited for global sensitivity analyses, uncertainty quantification,
 - Lose efficiency and scalability when the parameter space becomes very high-dimensional

Perspectives

- Future directions:
 - Working on emulators: for example Deep Gaussian Processes (hybrid GP + neural networks).
 - Rewriting land surface models in modern programming languages (e.g. Julia with automatic differentiation and GPU support) to auto-generate adjoint models;
 - Hybrid modelling (replacing parameterisations with ML/deep learning): attractive but raises concerns around physical interpretability, uncertainty quantification, and robustness under unseen climate conditions.

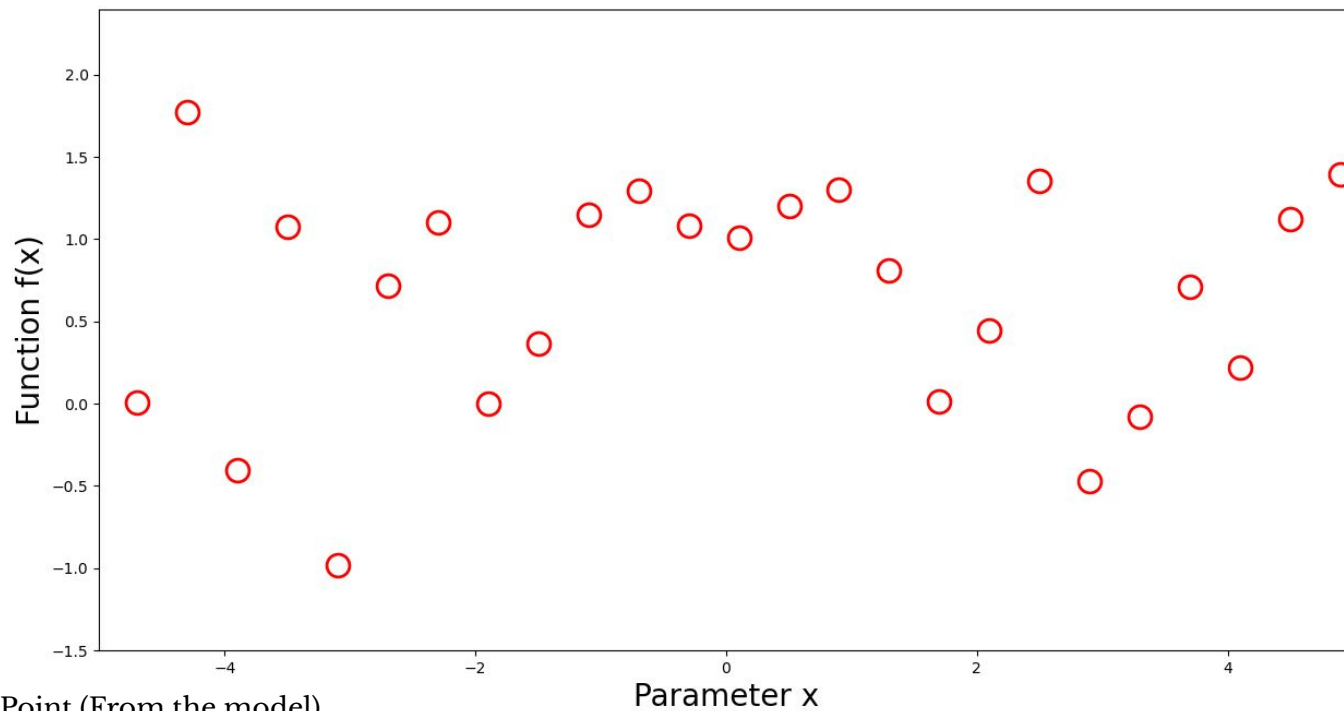


LABORATOIRE DES SCIENCES DU CLIMAT
& DE L'ENVIRONNEMENT

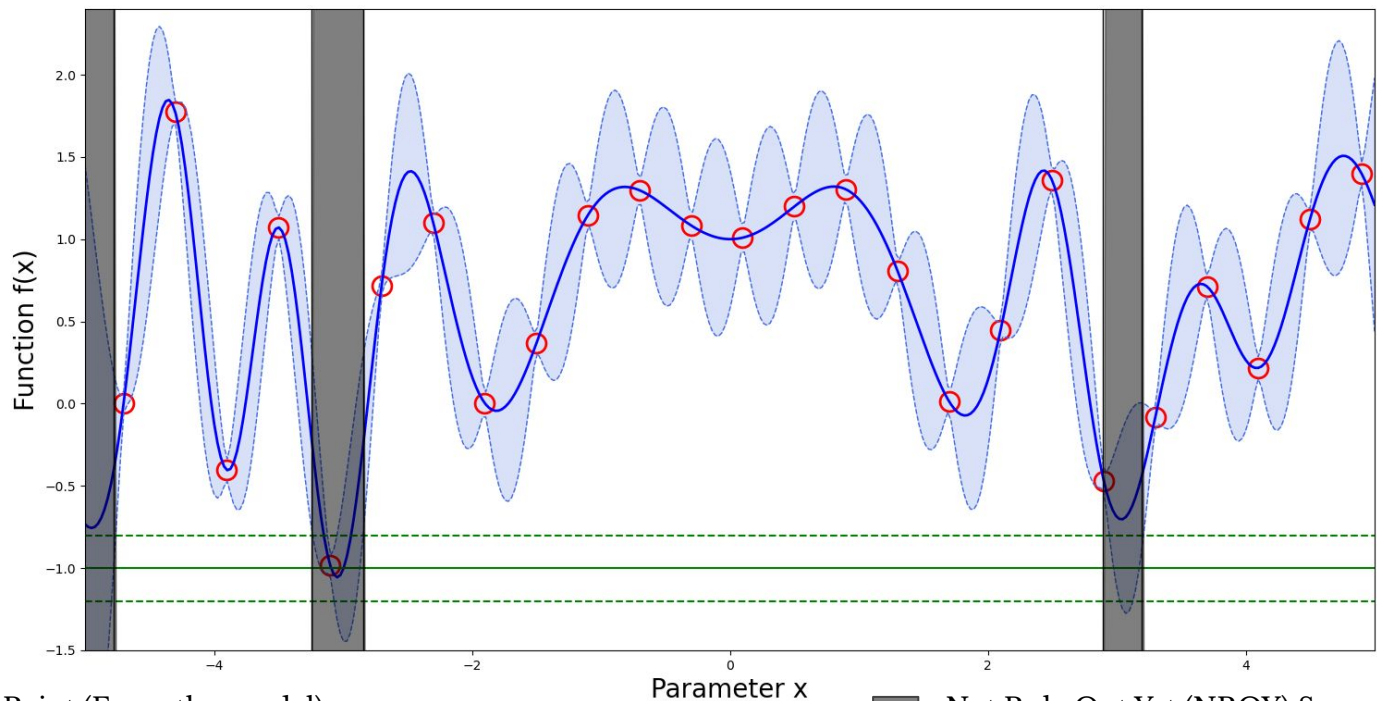


Thank you for your attention

History Matching



History Matching



Training Point (From the model)



Prediction of the Emulator



Uncertainty Prediction of the Emulator



Not Rule Out Yet (NROY) Space



Observation



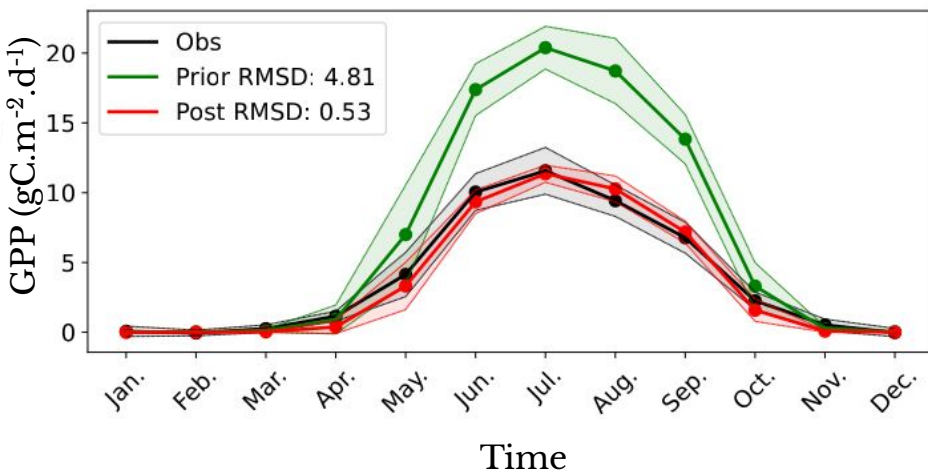
Uncertainty of the Observation

Regularisation of the method

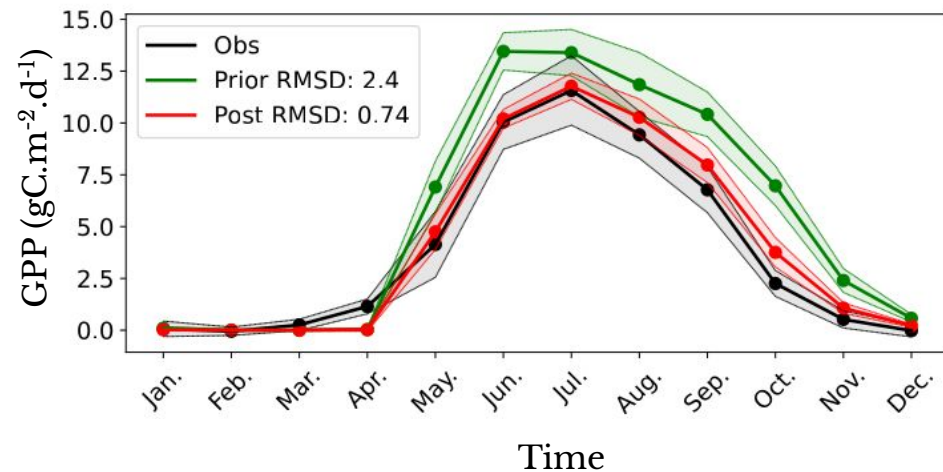
$$J(\mathbf{w}) = \frac{1}{2} (\mathbf{H}\mathbf{X}'_b \mathbf{w} + \mathcal{H}(x_b) - \mathbf{y})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{X}'_b \mathbf{w} + \mathcal{H}(x_b) - \mathbf{y}) + \frac{1}{2} \mathbf{w}^T \mathbf{w} \gamma$$

Observations : Harvard FluxNet

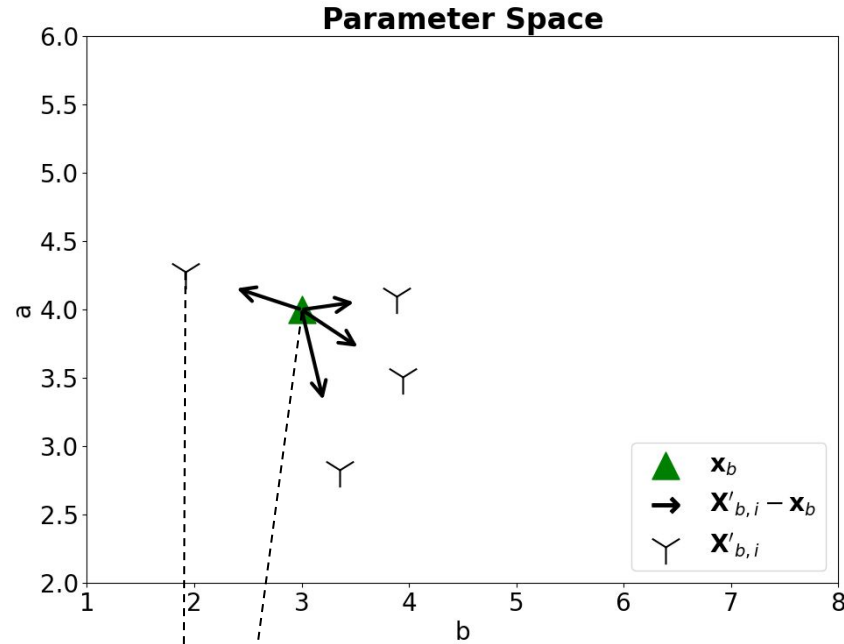
JULES



ORCHIDEE



Ensemble Variational method

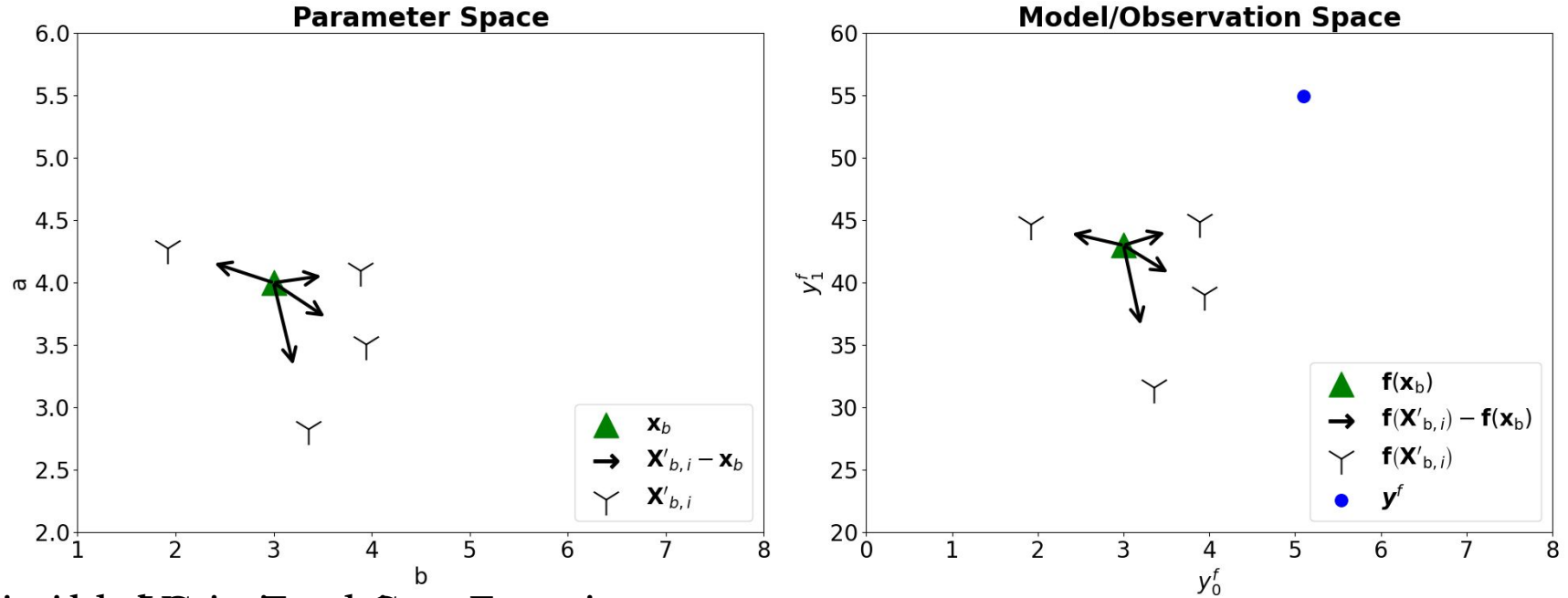


$$f(a,b) = y$$

$$\mathbf{X}'_b = \frac{1}{\sqrt{N-1}} (\underbrace{\mathbf{x}_1}_{\text{dashed}} - \underbrace{\mathbf{x}_b}_{\text{dashed}}; \mathbf{x}_2 - \mathbf{x}_b; \dots; \mathbf{x}_N - \mathbf{x}_b)$$

$$\mathbf{H}\mathbf{X}'_b \approx \frac{1}{\sqrt{N-1}} (\underbrace{\mathcal{H}(\mathbf{x}_1)}_{\text{dashed}} - \underbrace{\mathcal{H}(\mathbf{x}_b)}_{\text{dashed}}; \mathcal{H}(\mathbf{x}_2) - \mathcal{H}(\mathbf{x}_b); \dots; \mathcal{H}(\mathbf{x}_N) - \mathcal{H}(\mathbf{x}_b))$$

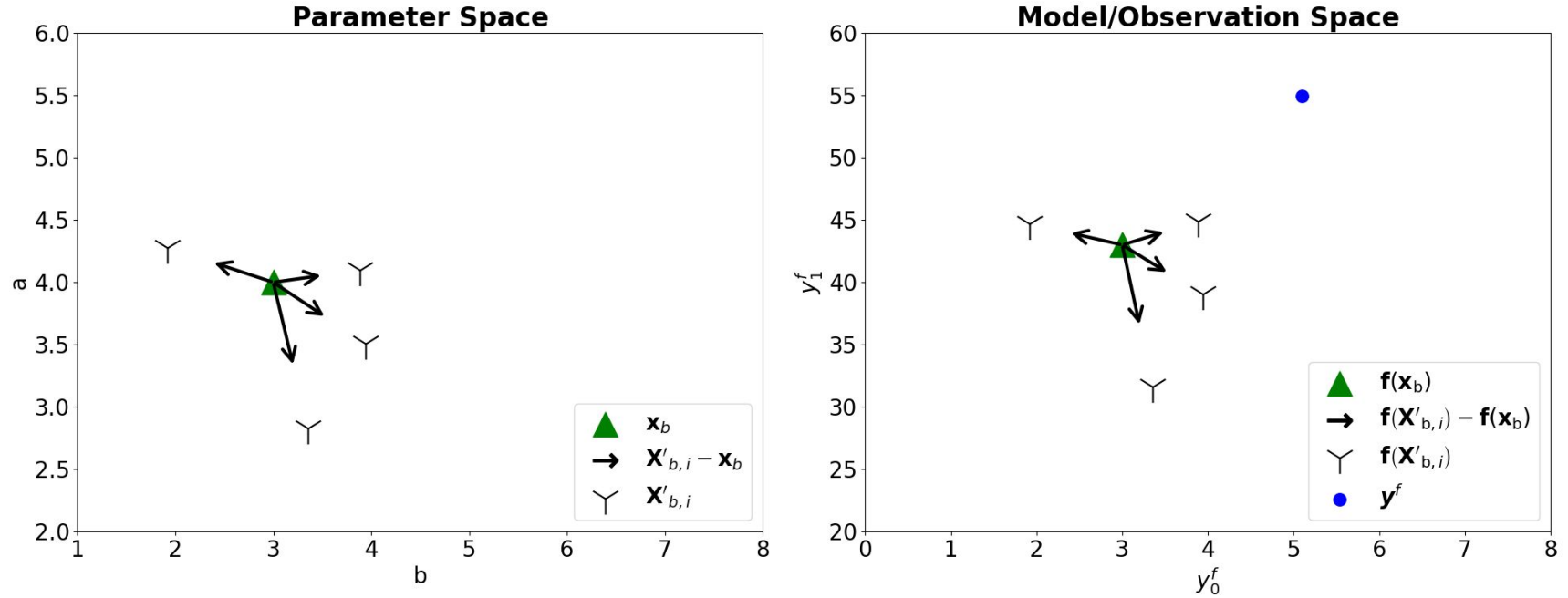
Ensemble Variational method



Ensemble Variational Method Cost Function :

$$J(\mathbf{w}) = \frac{1}{2}(\mathbf{H}\mathbf{X}'_b\mathbf{w} + \mathcal{H}(\mathbf{x}_b) - \mathbf{y})^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{X}'_b\mathbf{w} + \mathcal{H}(\mathbf{x}_b) - \mathbf{y}) + \frac{1}{2}\mathbf{w}^T \mathbf{w}$$

Ensemble Variational method



$$\mathbf{w}^* = (\mathbf{I} + (\mathbf{H}\mathbf{X}'_b)^T \mathbf{R}^{-1} \mathbf{H}\mathbf{X}'_b)^{-1} (\mathbf{H}\mathbf{X}'_b)^T \mathbf{R}^{-1} (\mathbf{y} - \mathcal{H}(\mathbf{x}_b))$$

Ensemble Variational method

