



# **Carbon Cycle Data Assimilation System (CCDAS)**

**Philippe Peylin & several contributors...**

Laboratoire des Sciences du Climat et de l'Environnement  
Gif sur Yvette,  
France

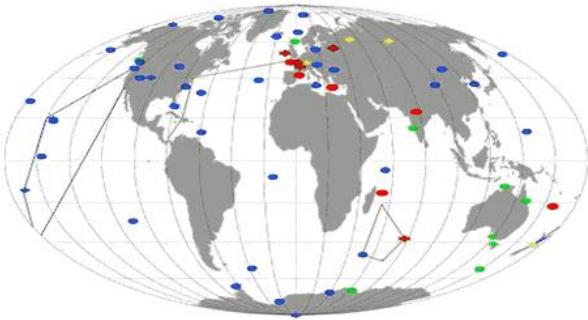
# Outline

---

- Current limitations of « standard » atmospheric flux inversions
- Multi-data streams assimilation: Basis for model parameters optimization (CCDAS)
- Potential of several land data streams
  - Fluxnet data
  - Satellite vegetation indexes
  - Biomass measurements
- Join multi-data assimilation
- Limitations & Prospects

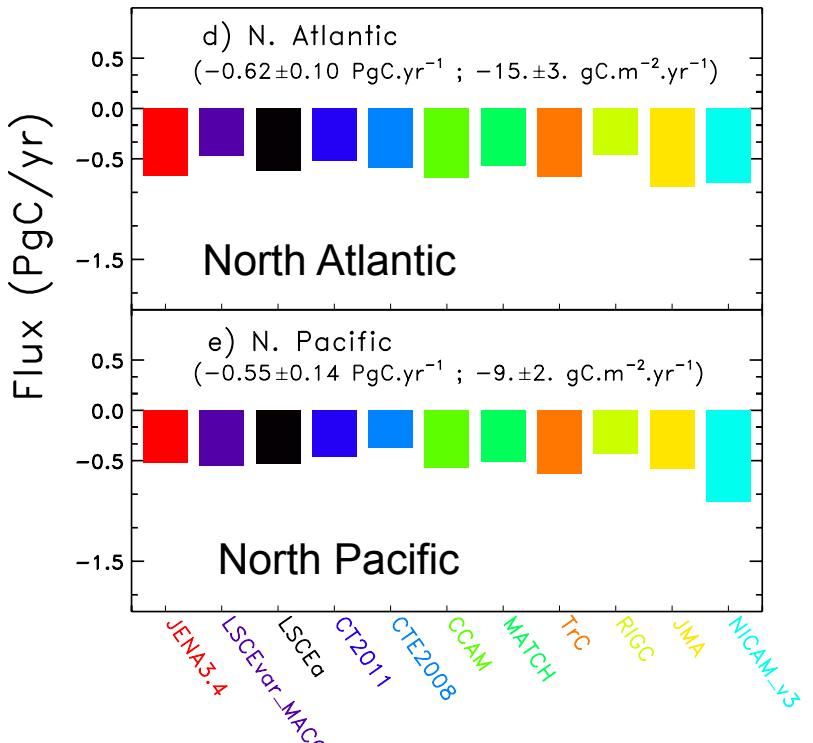
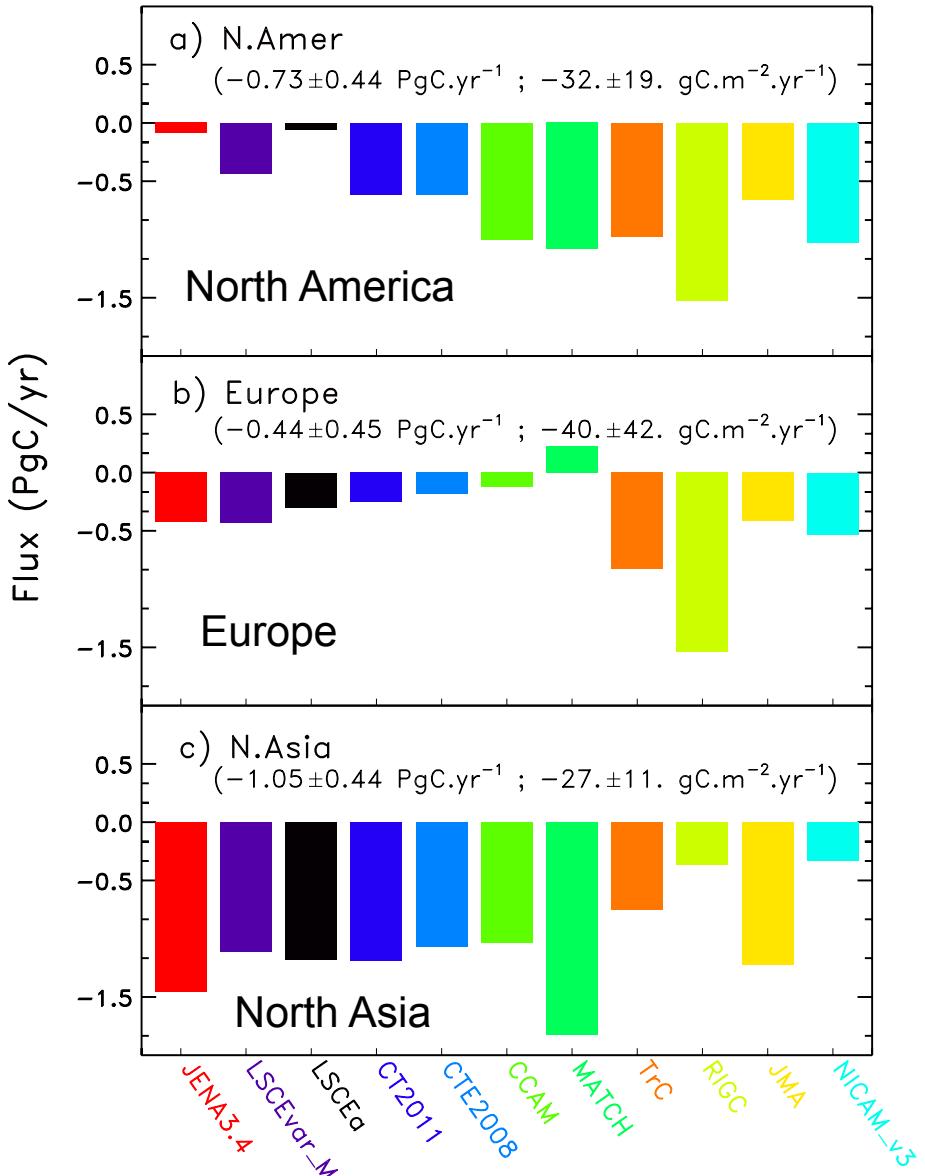
# Atmospheric CO<sub>2</sub> inversions....

---

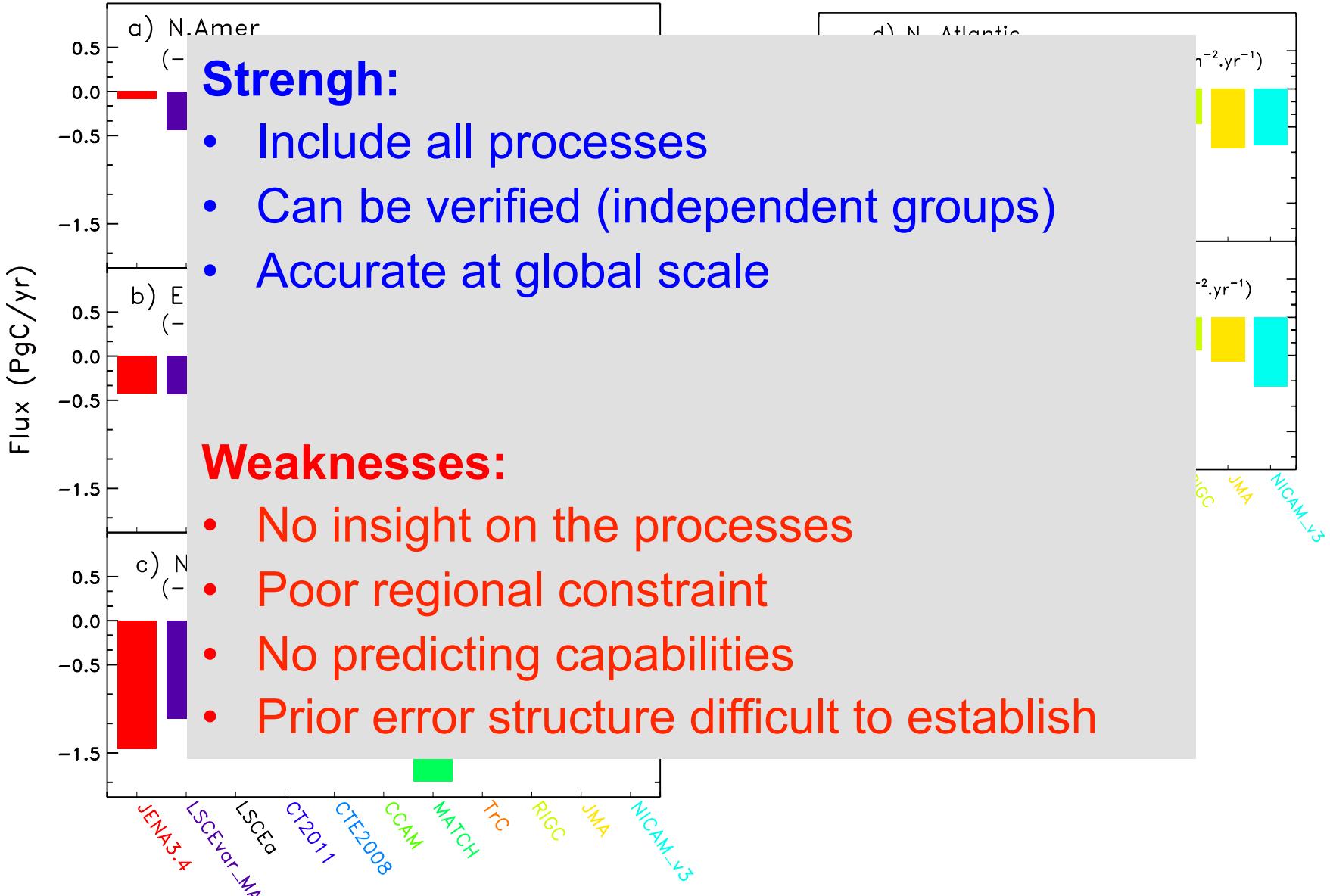


- Top-down approach :
  - ➔ Estimated fluxes account for all surface processes
- Verifiable by independent groups
- Several implementations applied so far...

# Atmospheric inversions: Continental scale long term mean



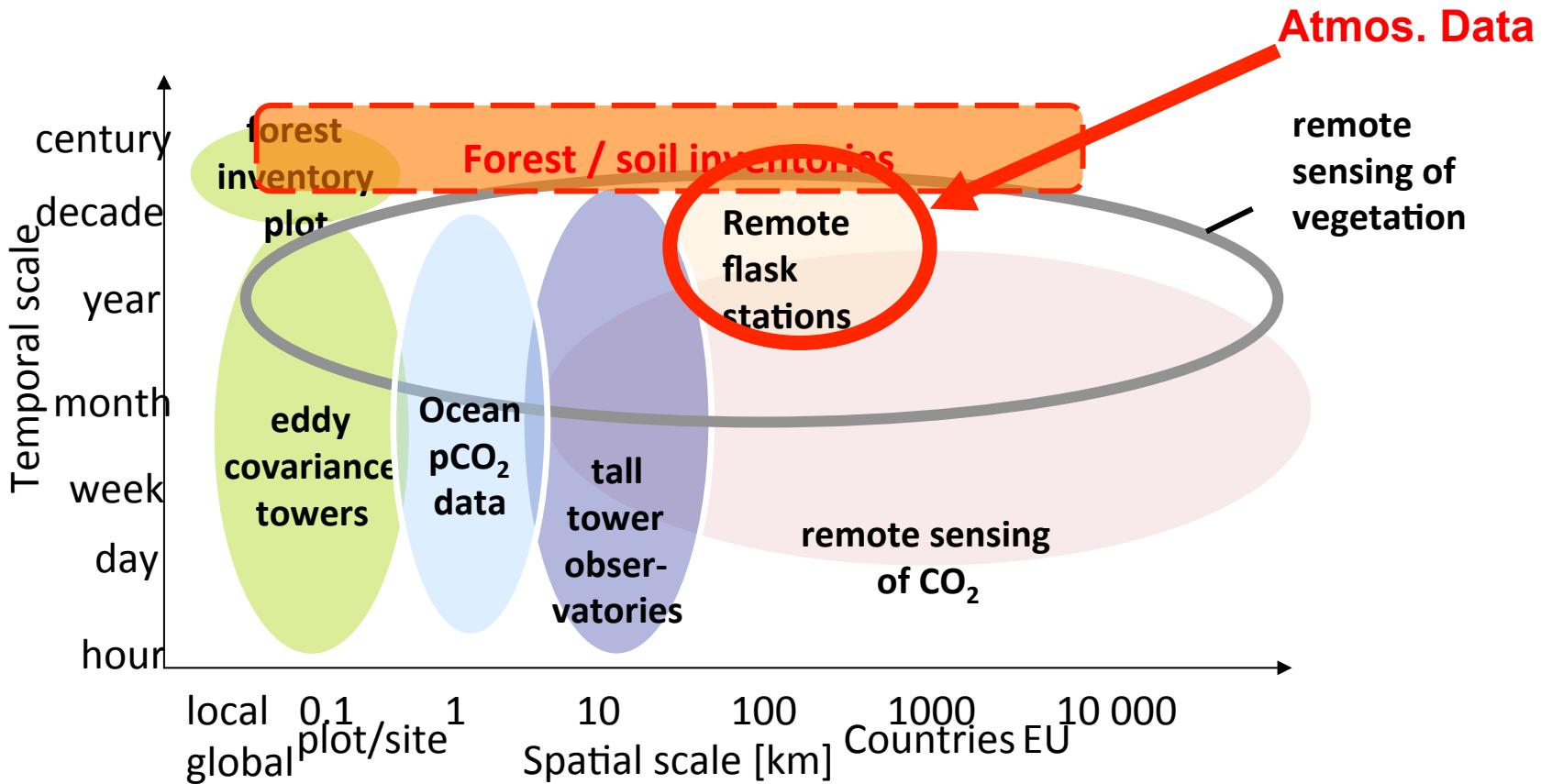
# Atmospheric inversions: Long term mean



# How to move forward ?

## Strong Need to :

- Combine the information from several data streams
- Attribute the net carbon flux variations to key processes



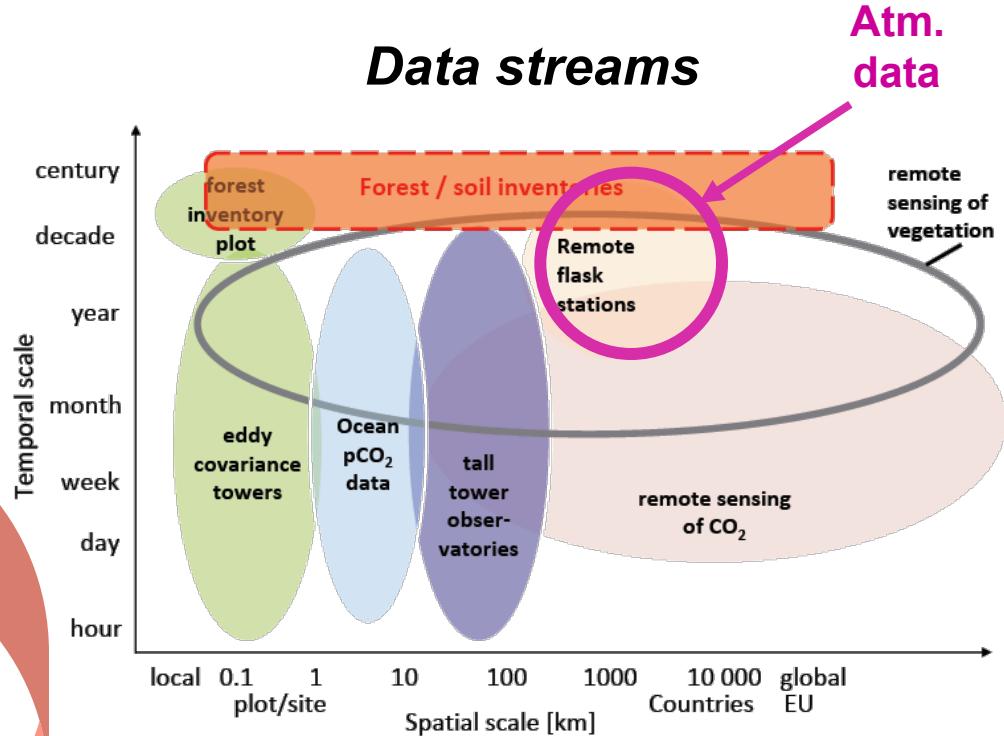
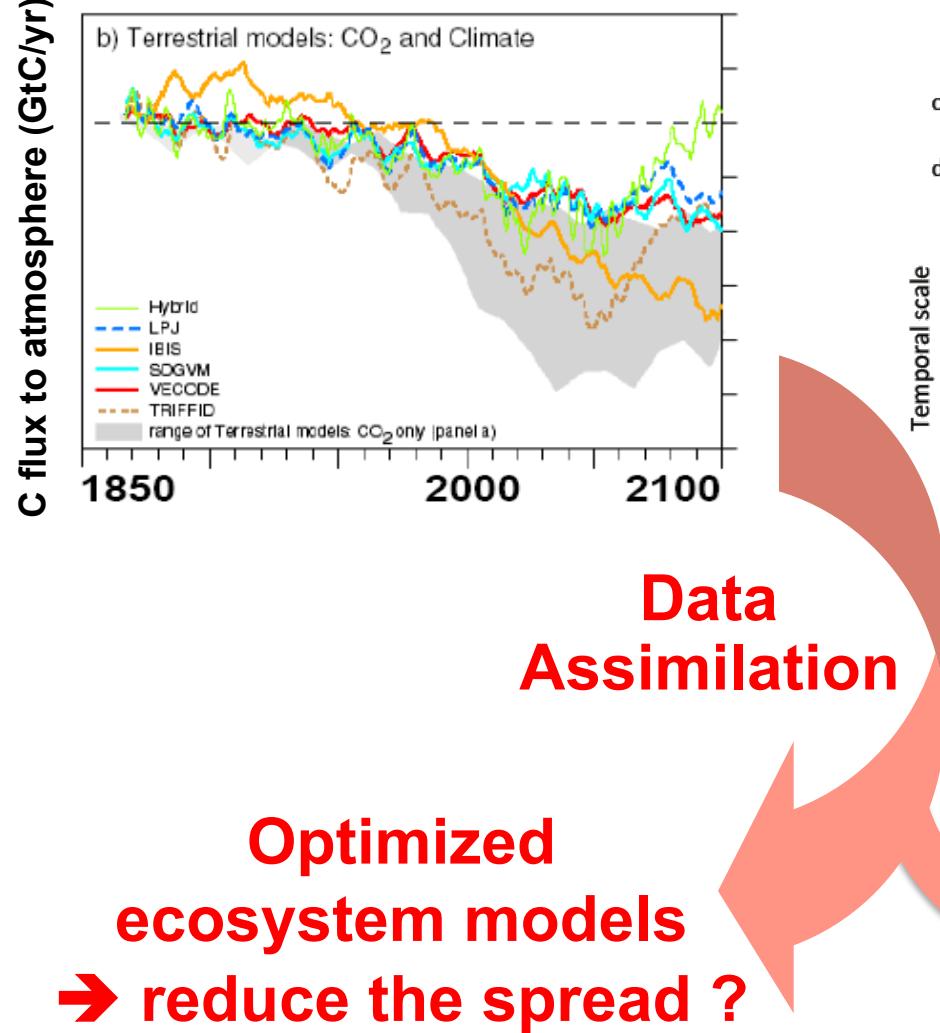
# Outline

---

- Current limitations of « standard » atmospheric flux inversions
- Multi-data streams assimilation: Basis for model parameters optimization (CCDAS)
- Potential of several land data streams
  - Fluxnet data
  - Satellite vegetation indexes
  - Biomass measurements
- Join multi-data assimilation
- Limitations & Prospects

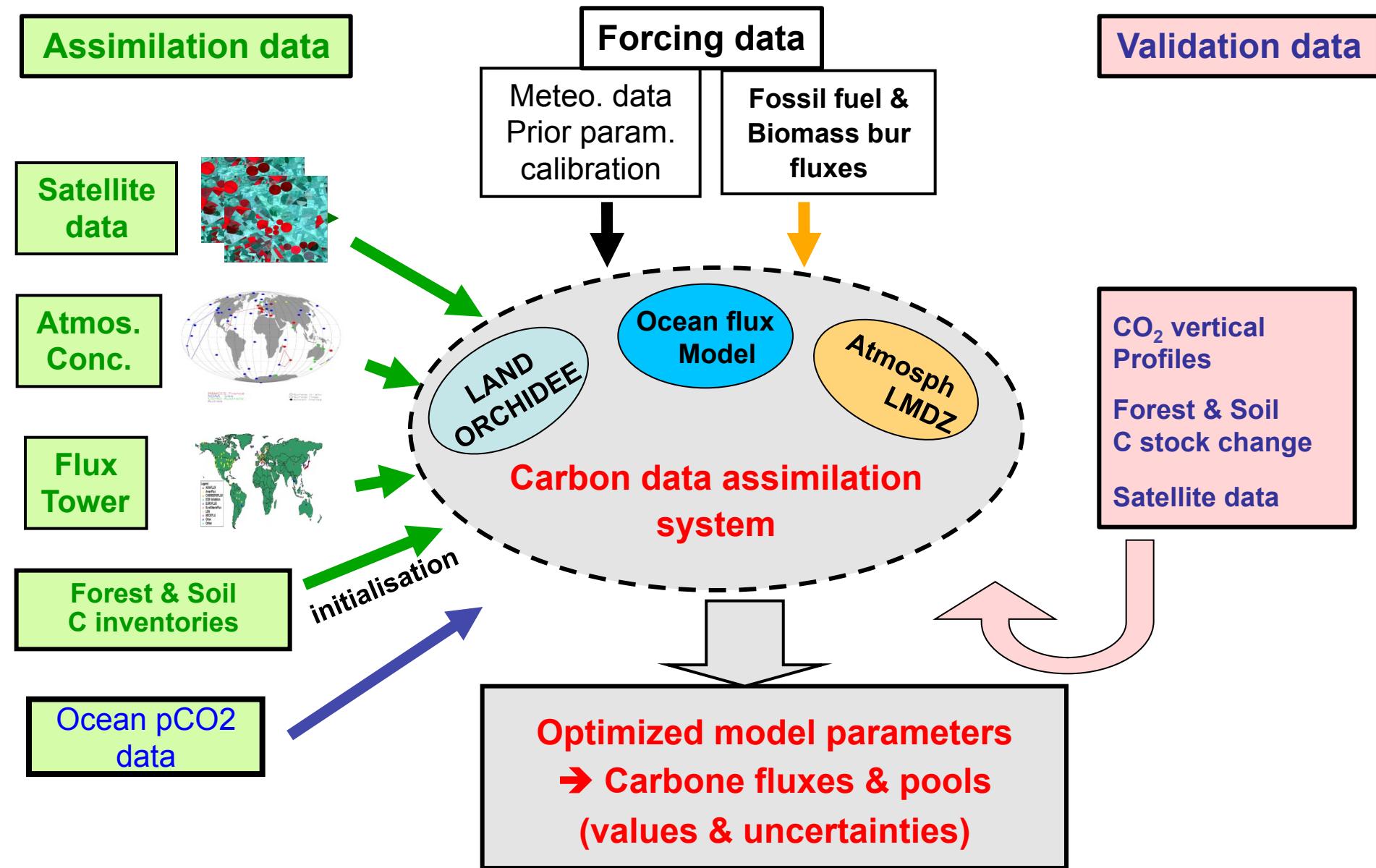
# Needs for a Carbon Cycle Data Assimilation System

Large uncertainty from land to predict global C-balance (C4MIP)



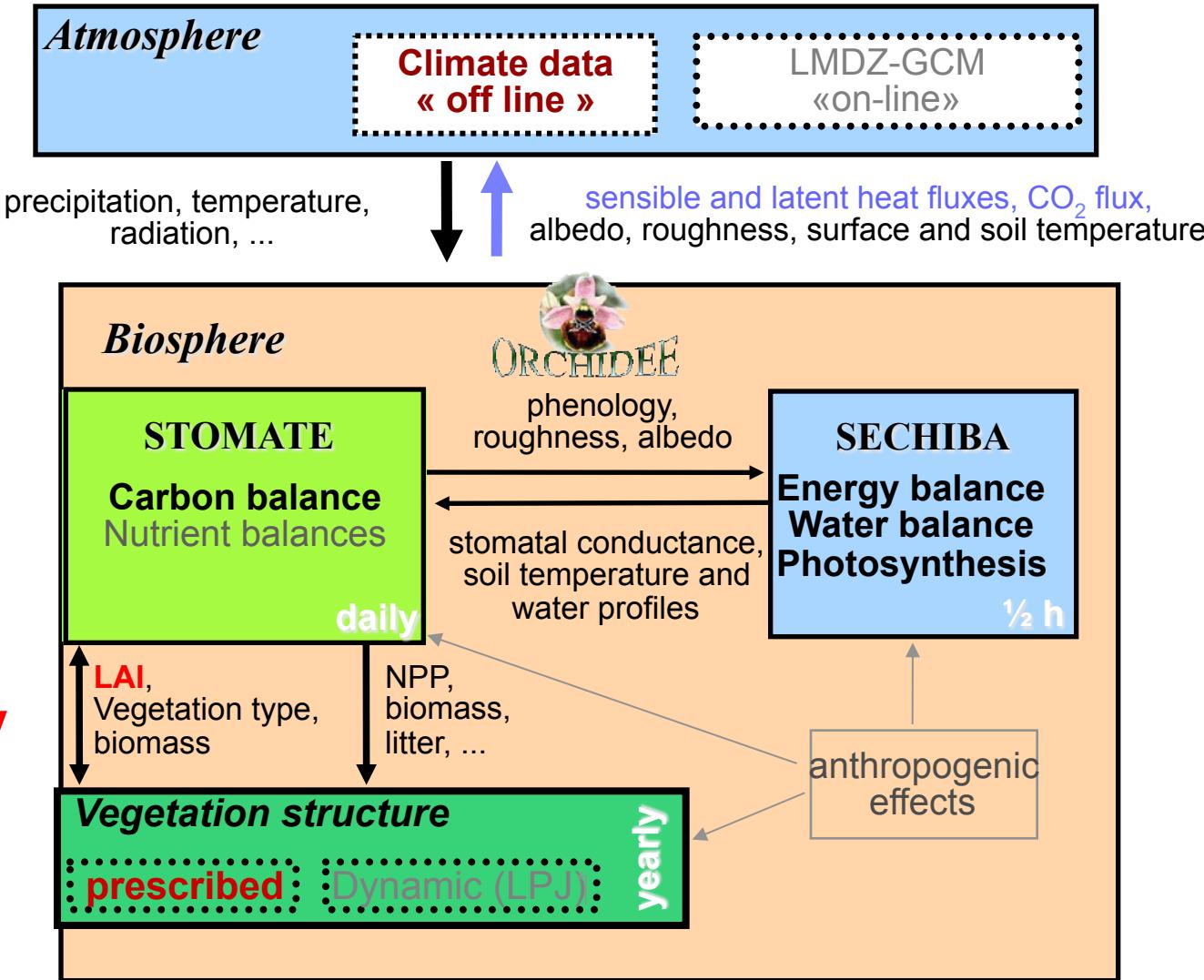
- Improve:**
- Process understanding
  - Uncertainty estimates
  - C land budget estimates
  - Future climate predictions

# Structure of a global “CCDAS”



# The ORCHIDEE ecosystem model

- Process driven model used for IPCC AR5 simulations
- Energy / Water / Carbon balances
- Global - Site level
- 13 PFT's
- Pronostic phenology
- $\frac{1}{2}$  hourly time step
- multiple C pools



# Formalism...

---

Baye's theorem:  $p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x}).p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$

Assuming Gaussian Error statistics

The mean of  $p(\mathbf{x}|\mathbf{y})$  is the minimum of the cost function  $J(\mathbf{x})$

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{Hx} - \mathbf{y})^T \mathbf{R}^{-1} (\mathbf{Hx} - \mathbf{y})$$

$\mathbf{x}$ : state vector ;

$\mathbf{x}_b$ : mean prior value of state vector

$\mathbf{y}$ : observation vector ;

$\mathbf{H}$ : linear observation operator

$\mathbf{B}$  /  $\mathbf{R}$ : Background / Observation error covariance matrix

# Formalism...

---

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{Hx} - \mathbf{y})^T \mathbf{R}^{-1} (\mathbf{Hx} - \mathbf{y})$$

- **Analytical solution**

- Need to linearize the model  $H(x)$
- **Sensitivities ( $H$ ) from tangent linear or Adjoint**

$$\mathbf{K} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1})^{-1} \mathbf{H}^T \mathbf{R}^{-1}$$

$$\mathbf{K} = \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1}$$

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{Hx}_b)$$

$$\mathbf{A} = \mathbf{B} - \mathbf{K} \mathbf{H} \mathbf{B}$$

- **Variational solution**

- Adapted to large size problems
- **Error estimation more difficult !**

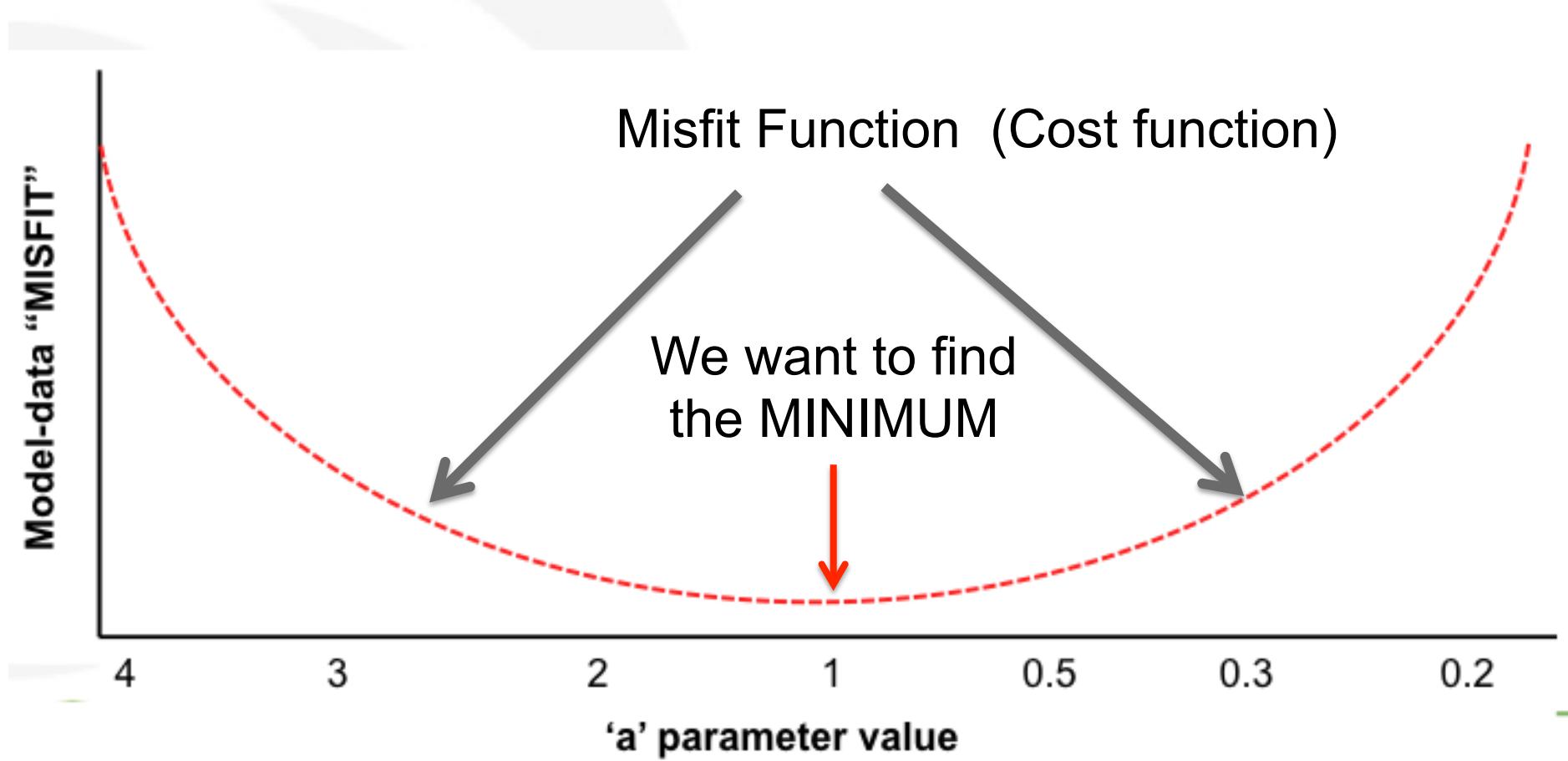
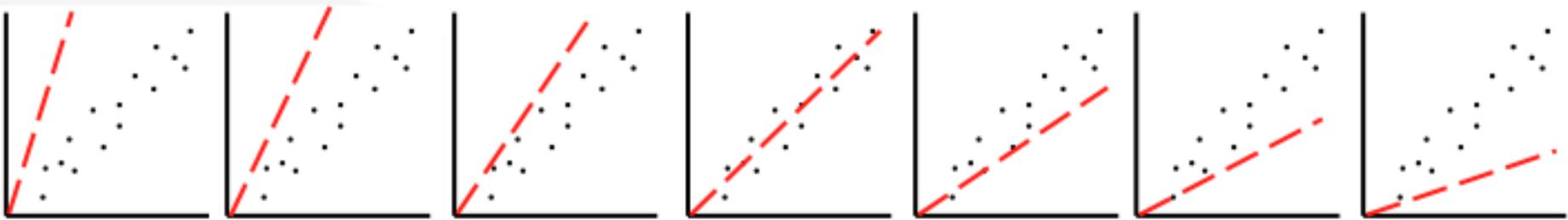
$$\nabla J(\mathbf{x}) = 2\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + 2\mathbf{H}^T \mathbf{R}^{-1}(\mathbf{Hx} - \mathbf{y})$$

$$\mathbf{A} = 2[J''(\mathbf{x}_a)]^{-1}$$

- **Monte Carlo approaches**

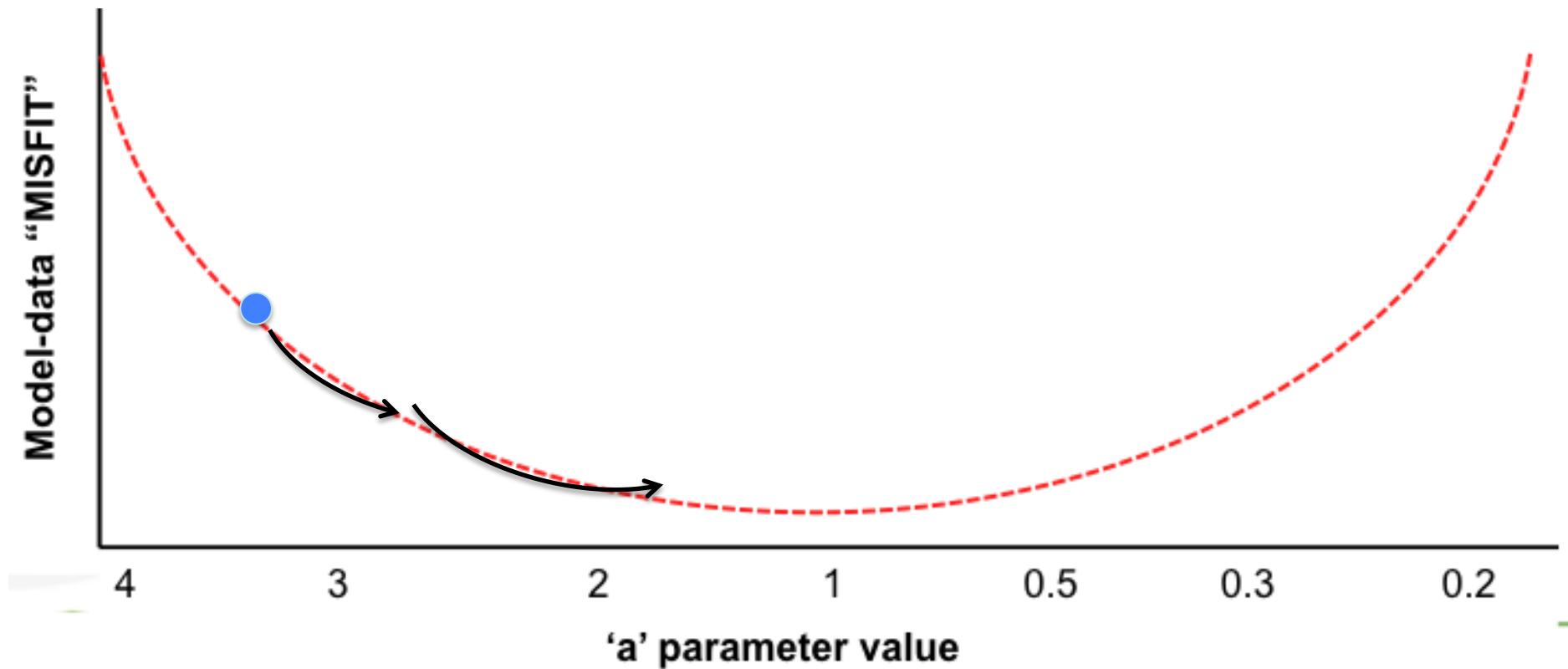
- Used mostly for site-level studies
- **Required time usually prohibitive with “complex model”**
- No limitations wrt LINEARITY & parameter PDF

# Simplest case (linear model) !



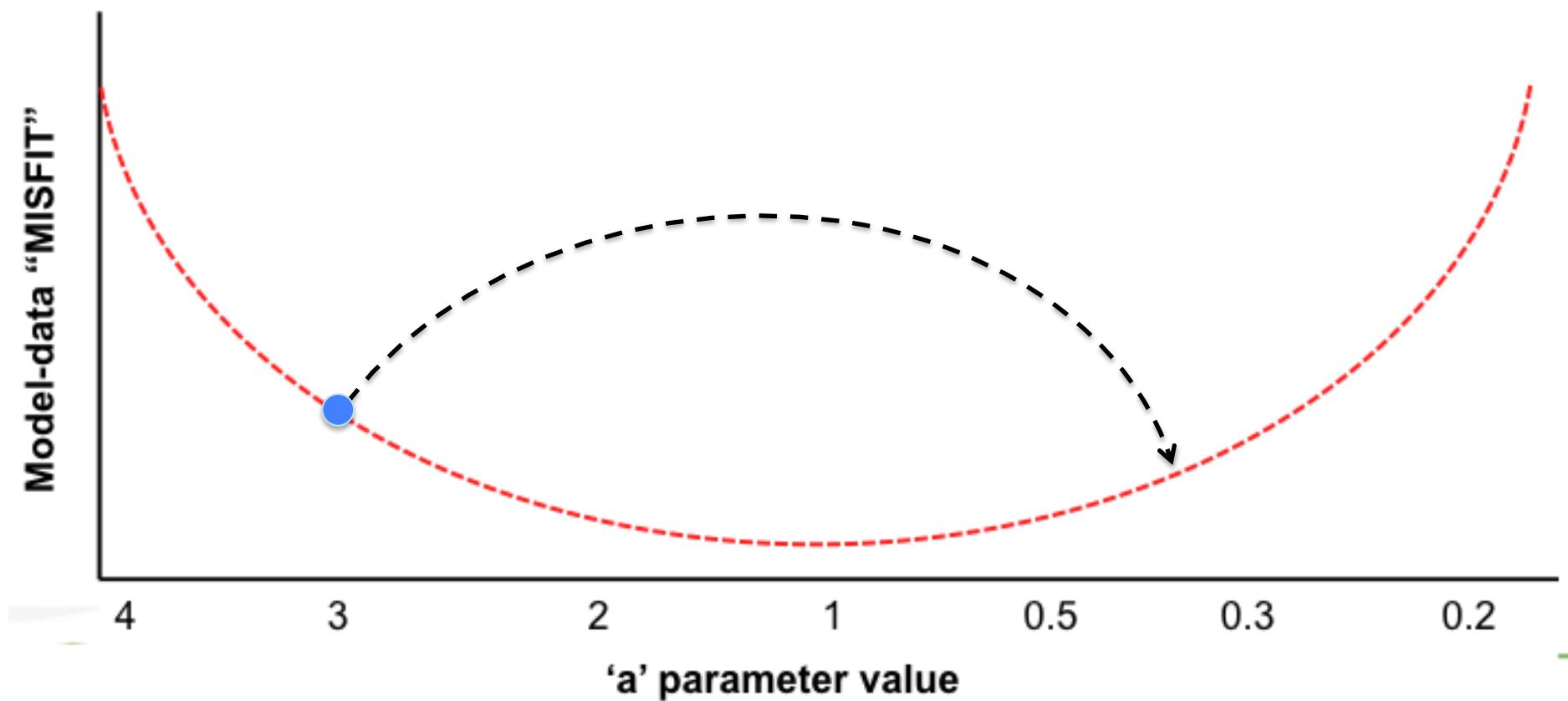
# Simplest case (linear model) !

- “Gradient-descent” methods
- Calculate the first derivative of the cost function in order to calculate the gradient...



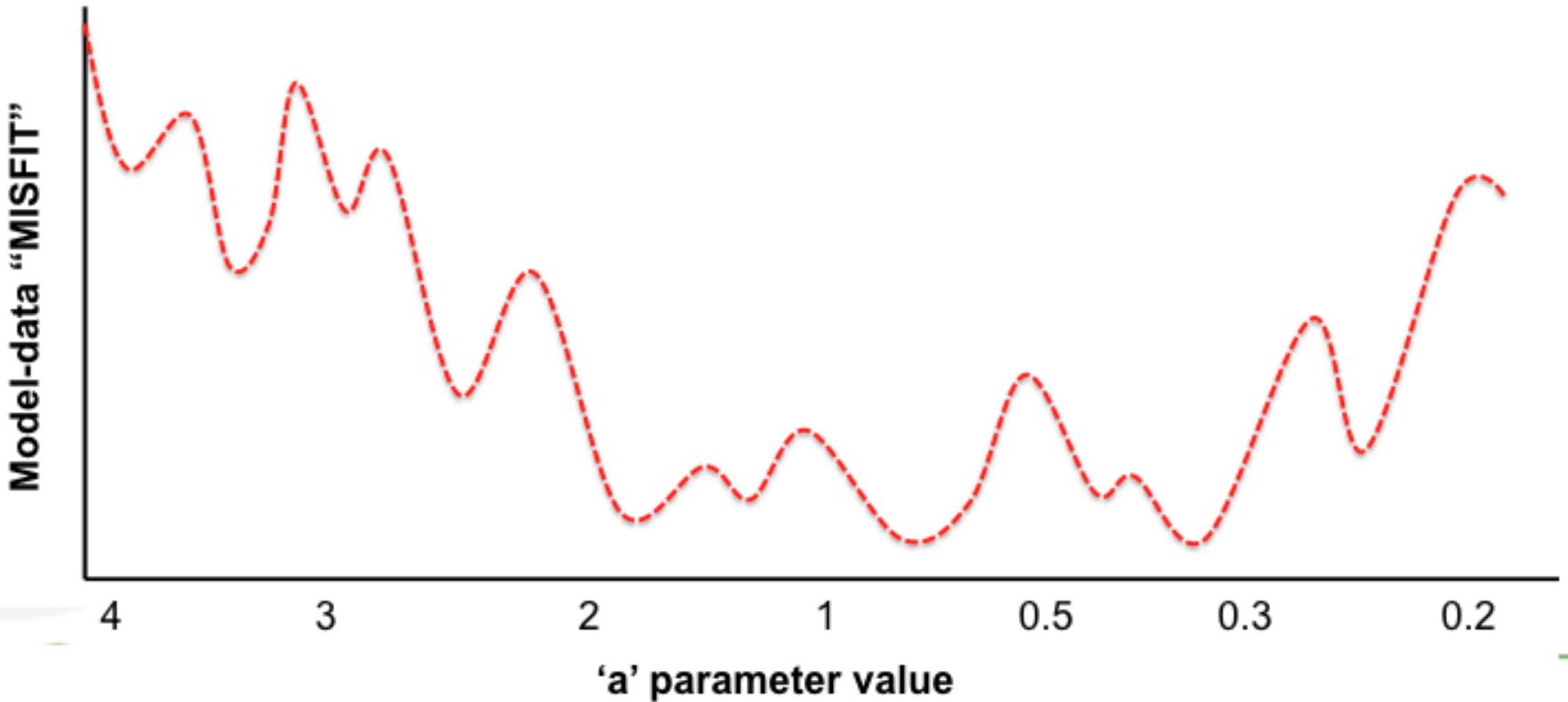
# Simplest case (linear model) !

- “Global search” methods (Genetic algorithm, Metropolis Hastings MCMC)
- Search parameter space...
- At each iteration calculate the misfit and accept or reject parameter



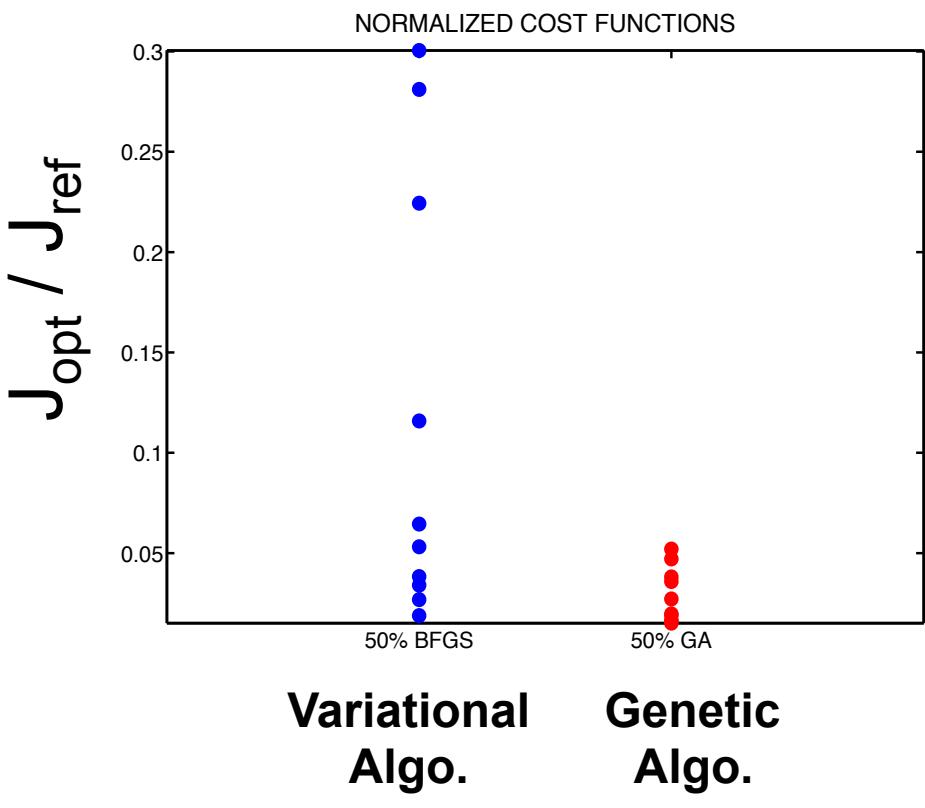
# Complex case!

- We want to find the MINIMUM of the misfit function...
- BUT! Your misfit function may look like this...!!



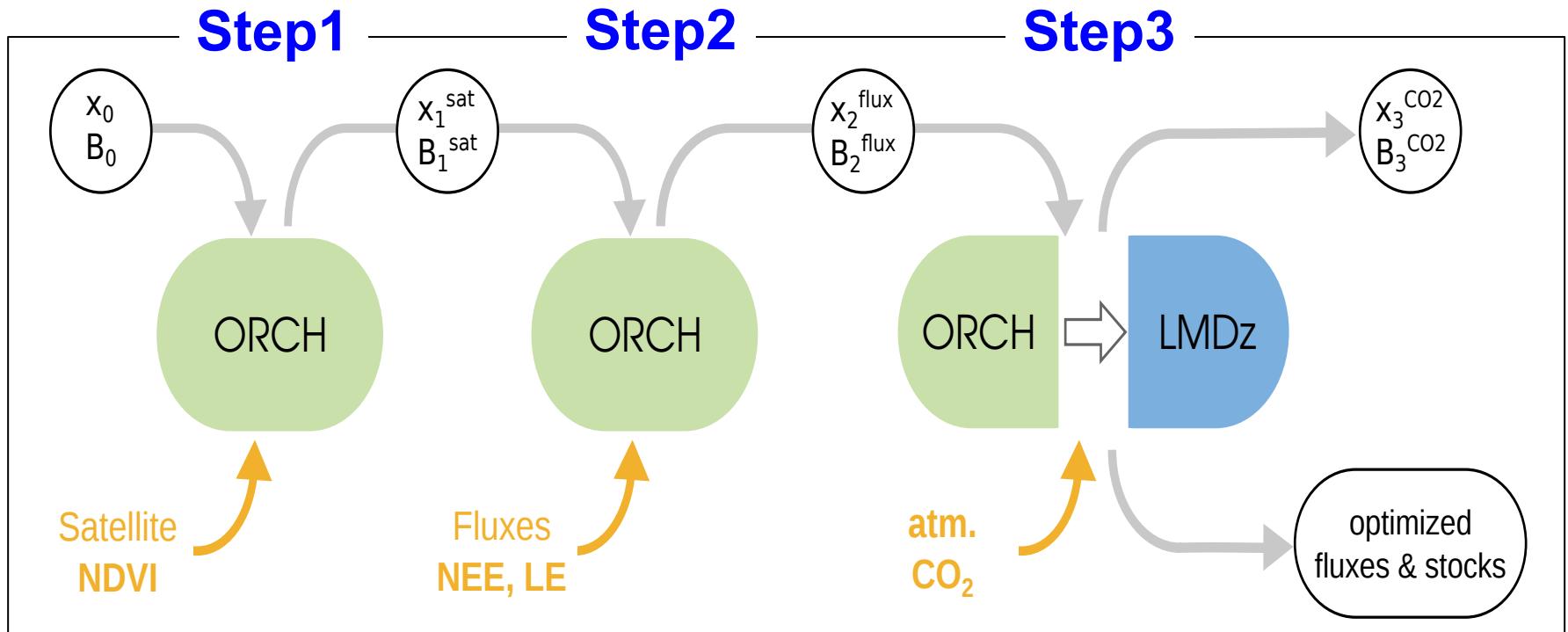
# Genetic vs Variational algorithm

- Pseudo-obs tests
- One site: Hesse  
(Beach forest)
- 20 parameters
- NEE/LE daily ; 1 year
- 10 tests with different random priors

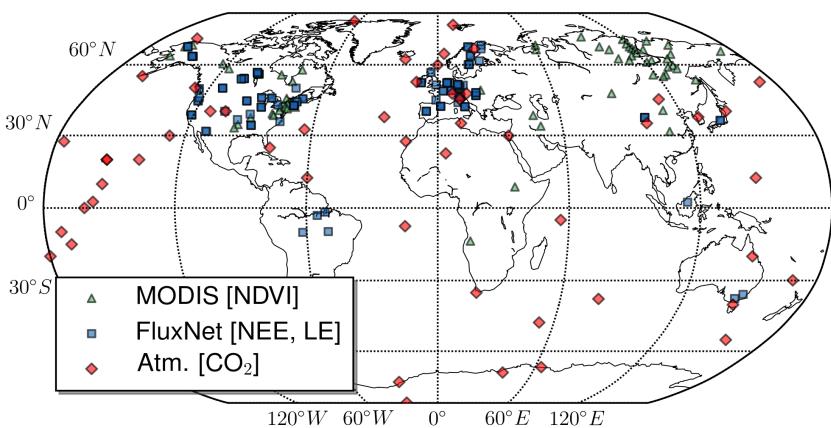


(Santaren et al, 2014)

# Exemple of sequential assimilation...



Site location  
for each  
data streams



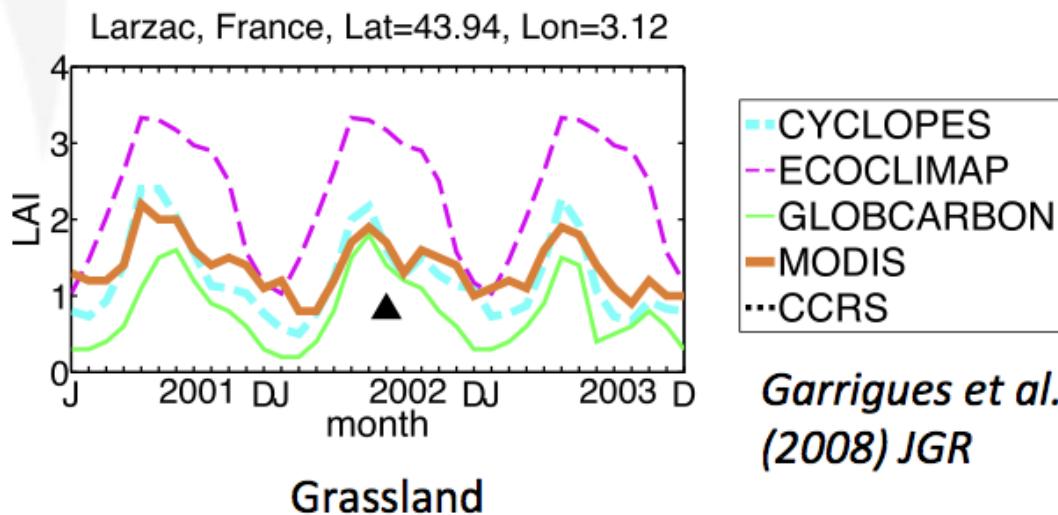
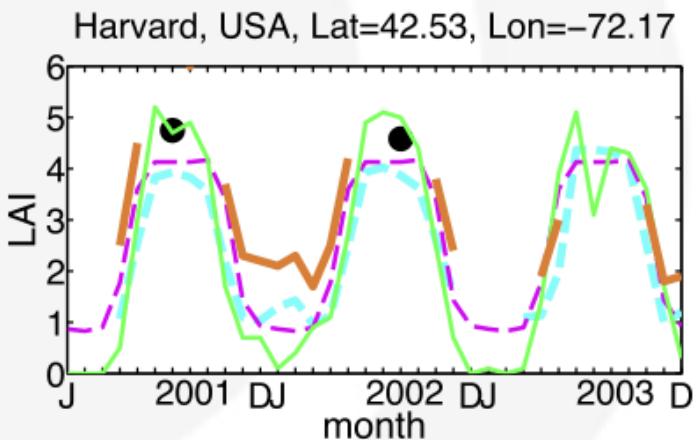
# Outline

---

- Current limitations of « standard » atmospheric flux inversions
- Multi-data streams assimilation: Basis for model parameters optimization (CCDAS)
- Potential of several land data streams
  - Satellite vegetation indexes
  - Fluxnet data
  - Biomass measurements
- Join multi-data assimilation
- Limitations & Prospects

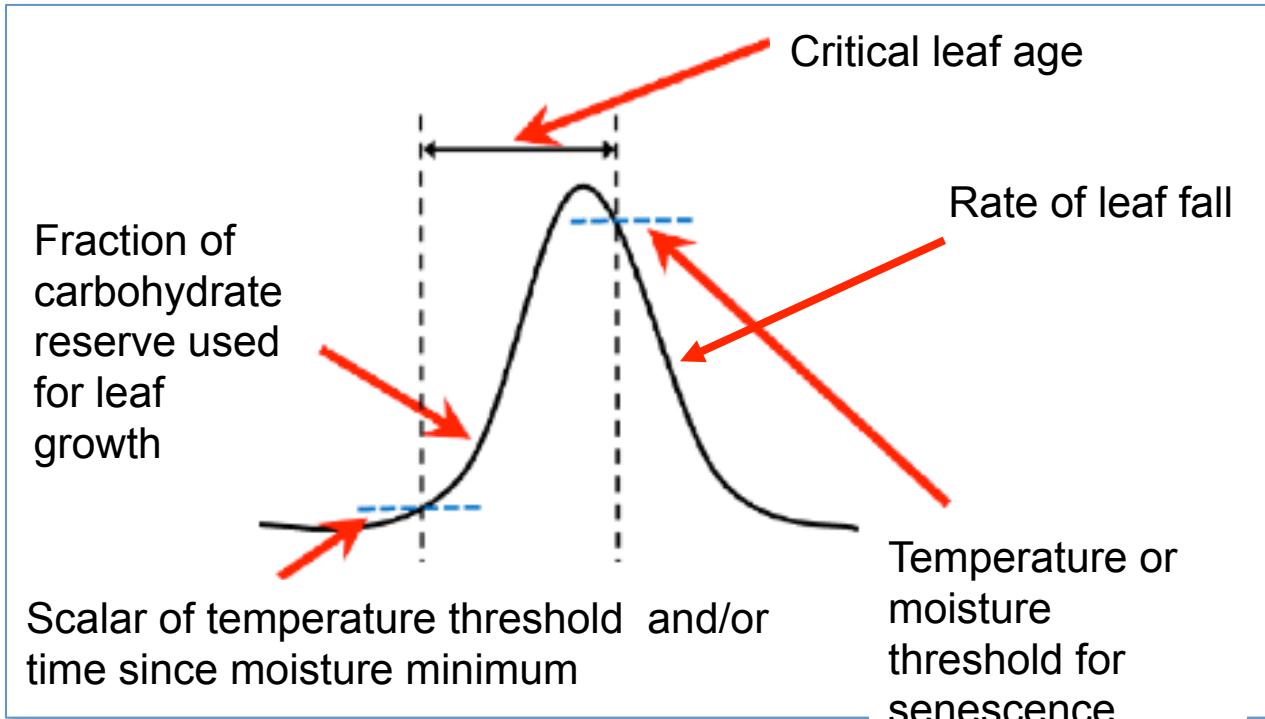
# Satellite data to optimize phenology...

- Used initially to “manually” adjust model phenology
- Recent formal studies with a complete statistical approach:
  - ✓ Stockli et al. 2008, 2011 ; Knorr et al. 2010
- Only the temporal variations are robust across products...



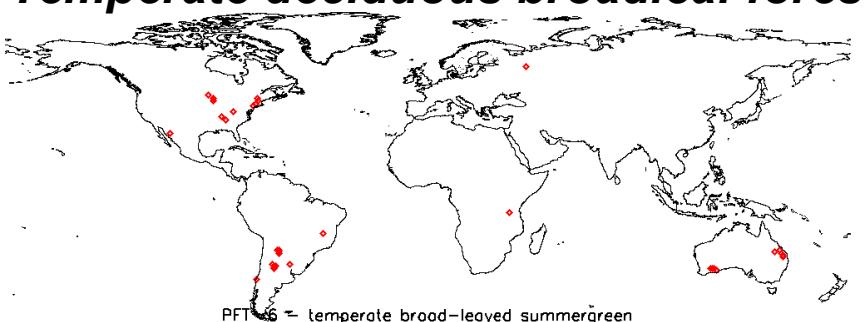
# MODIS NDVI to optimize phenology...

- 4 – 6 parameters per PFT
- 15 random grid points with available obs.
- PFT vegetation cover > 0.6

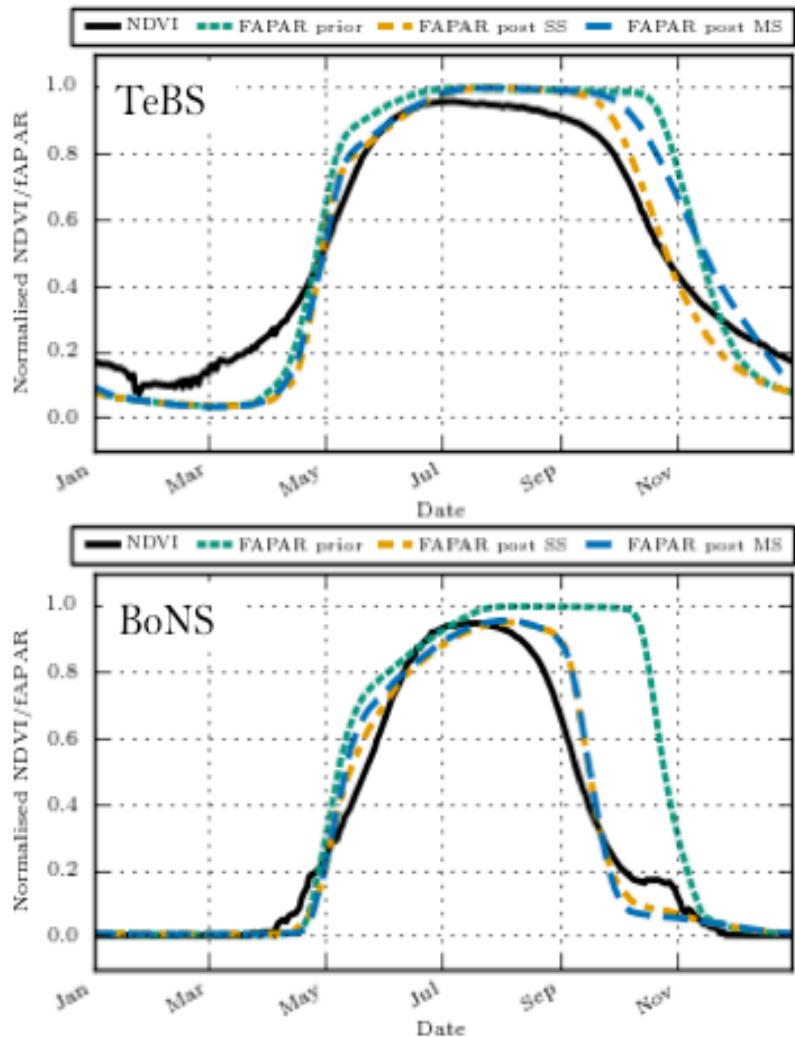


**multi-sites optimization**  
 &  
**single-site optimizations**

*Ex: Temperate deciduous broadleaf forest*



# Mean seasonal cycle

**DATA**
**PRIOR model**
**Posterior SS**
**Posterior MS**


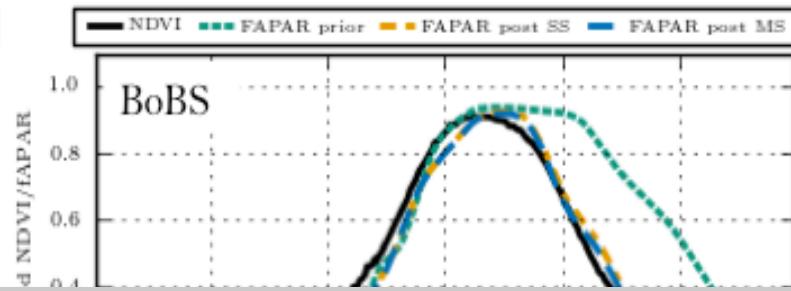
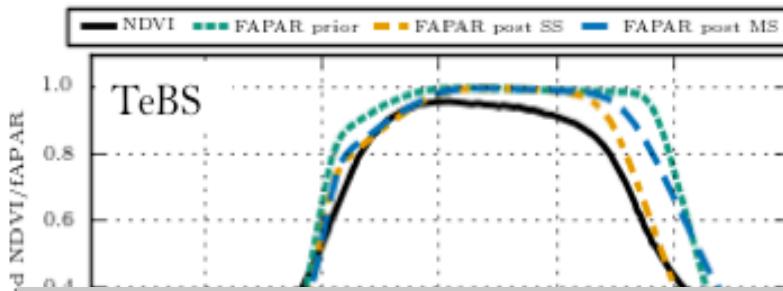
# Mean seasonal cycle

**DATA**

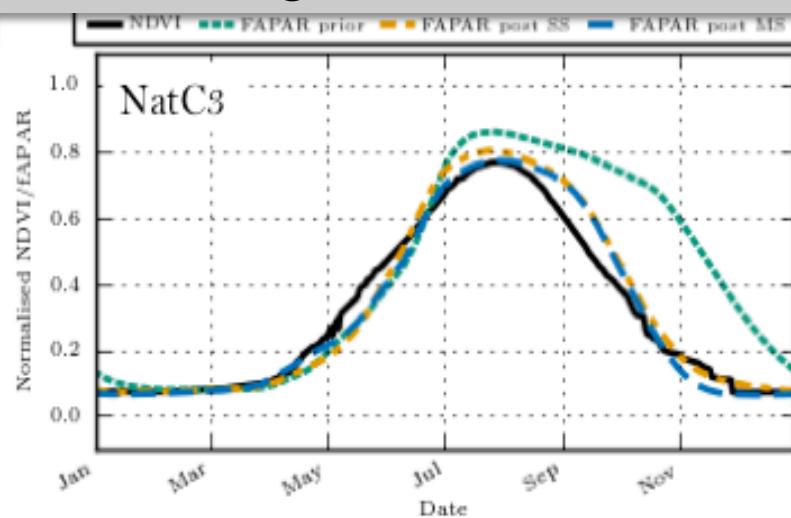
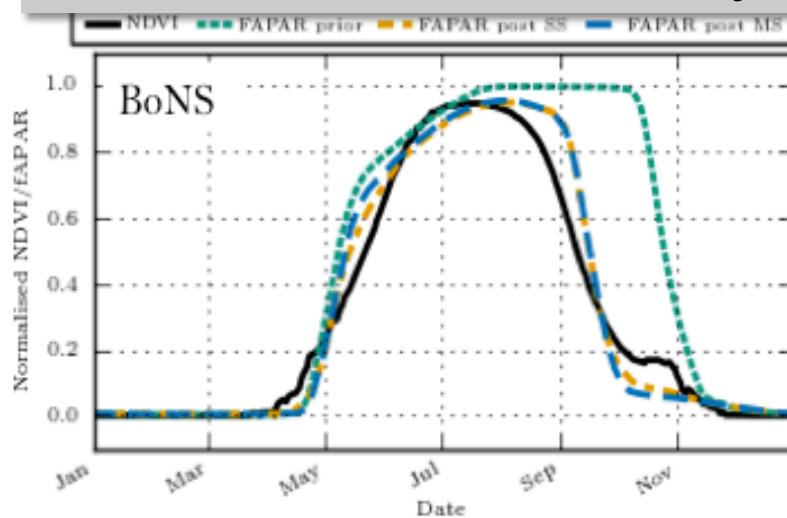
**PRIOR model**

**Posterior SS**

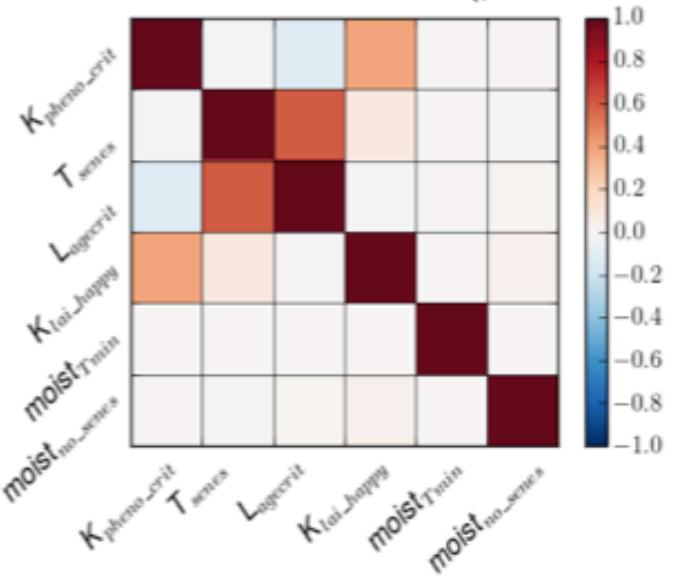
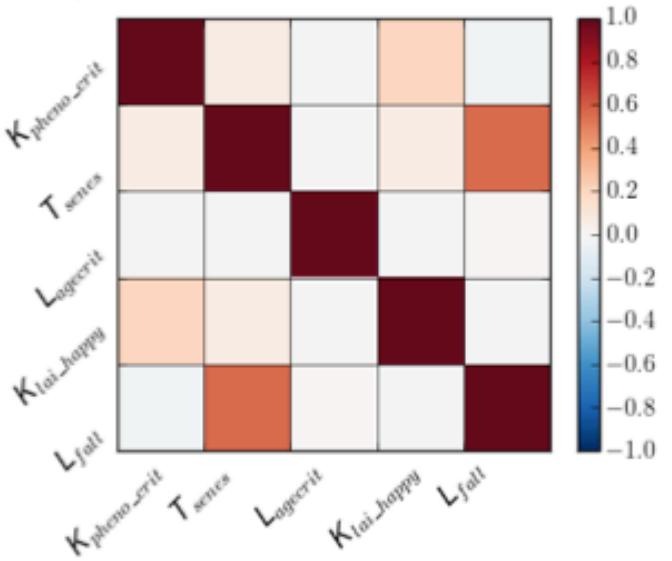
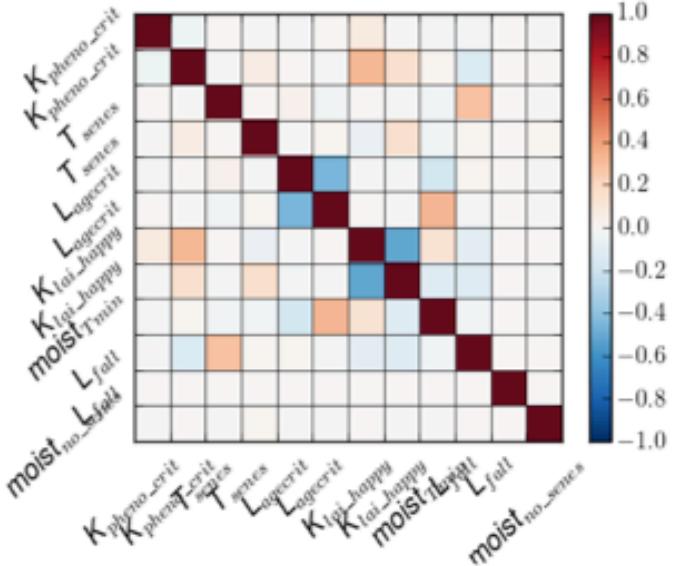
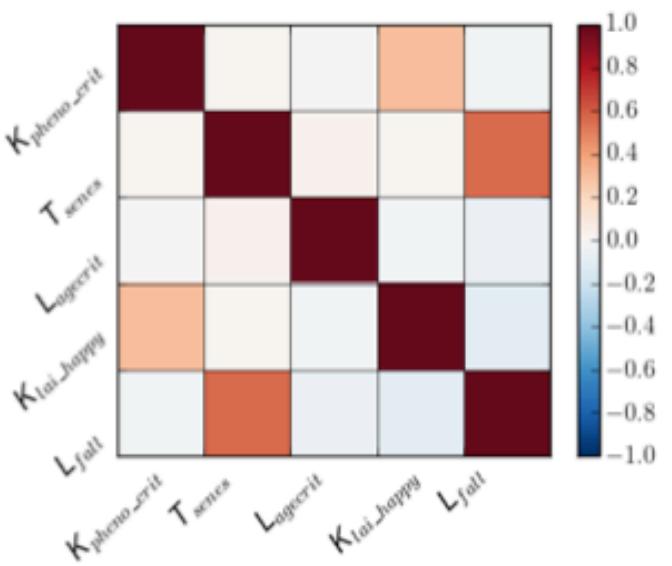
**Posterior MS**



- Reduction in GSL → earlier end of growing season
- Multi-site does similarly well as single-site

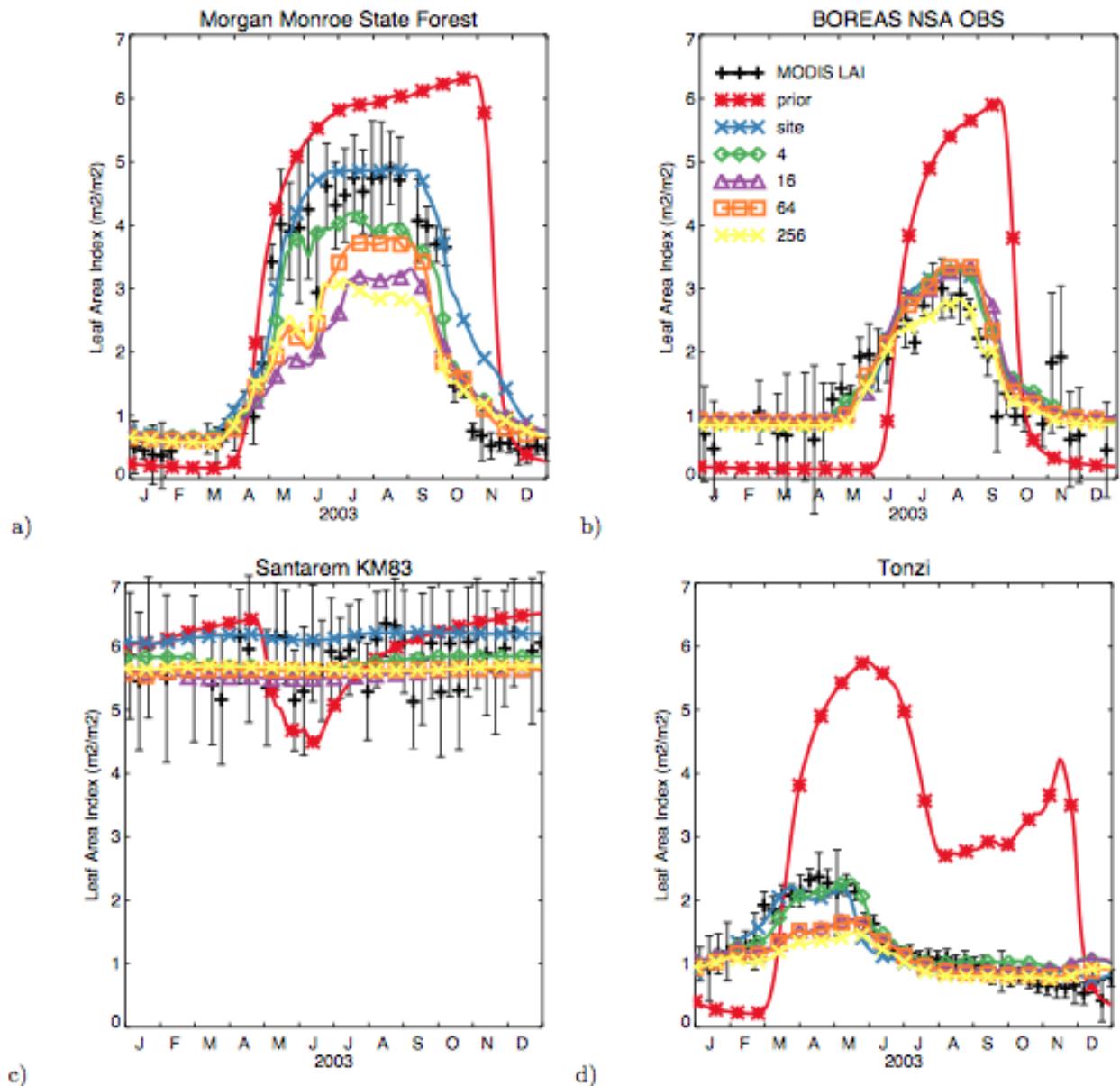


# Posterior parameters – covariance



# Results from Stockli et al. 2011

- Assimilate MODIS LAI or FAPAR
- Calibrate a specific Phenology model
- Use different nb of regions in the optimization
- → compare at few sites with in situ observations.



# Outline

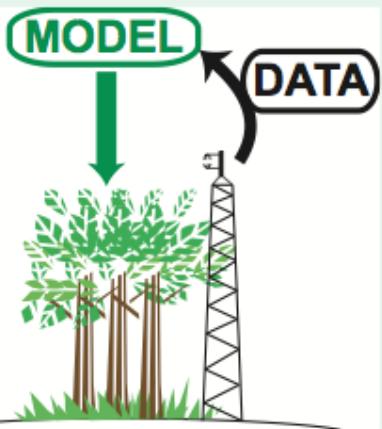
---

- Current limitations of « standard » atmospheric flux inversions
- Multi-data streams assimilation: Basis for model parameters optimization (CCDAS)
- Potential of several land data streams
  - Satellite vegetation indexes
  - Fluxnet data
  - Biomass measurements
- Join multi-data assimilation
- Limitations & Prospects

# History of Flux data assimilation



## Site-specific optimization...

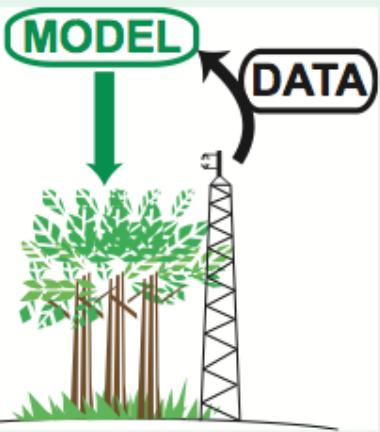


- (Wang et al., 2001)
- (Braswell et al., 2005)
- (Knorr and Kattge, 2005)
- (Santaren et al., 2007)
- (Wang et al., 2007)
- (Williams et al., 2009)
- (Carvalhais et al., 2010)
- (Verbeeck et al., 2011)
- ...

# History of Flux data assimilation

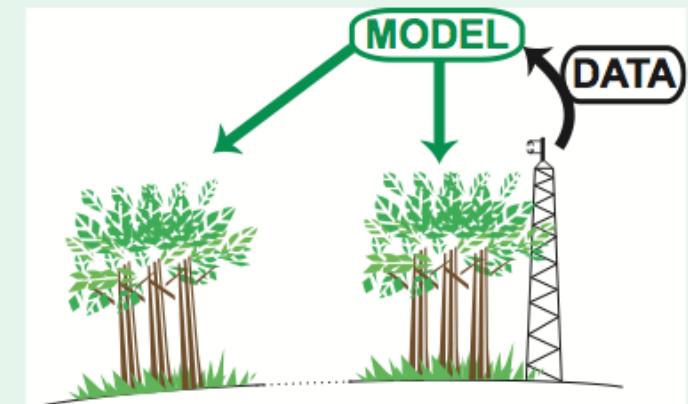


## Site-specific optimization...



- (Wang et al., 2001)  
(Braswell et al., 2005)  
(Knorr and Kattge, 2005)  
(Santaren et al., 2007)  
(Wang et al., 2007)  
(Williams et al., 2009)  
(Carvalhais et al., 2010)  
(Verbeeck et al., 2011)  
...

## ...and evaluation at other sites

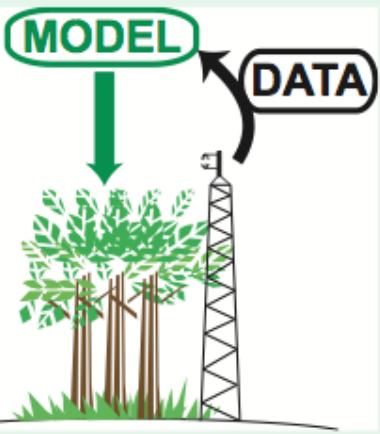


- (Medvigy et al., 2009; Verbeeck et al., 2011)

# History of Flux data assimilation

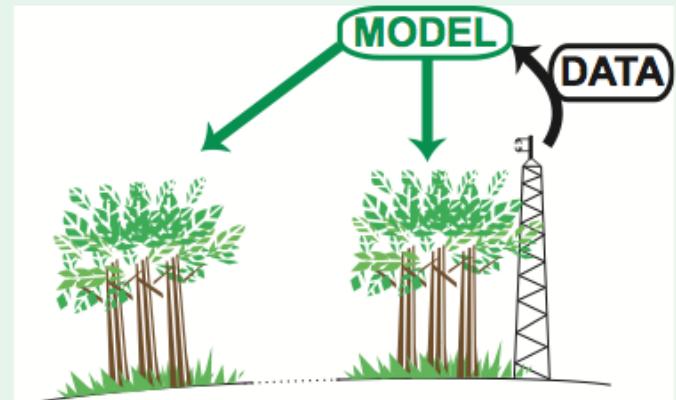


## Site-specific optimization...



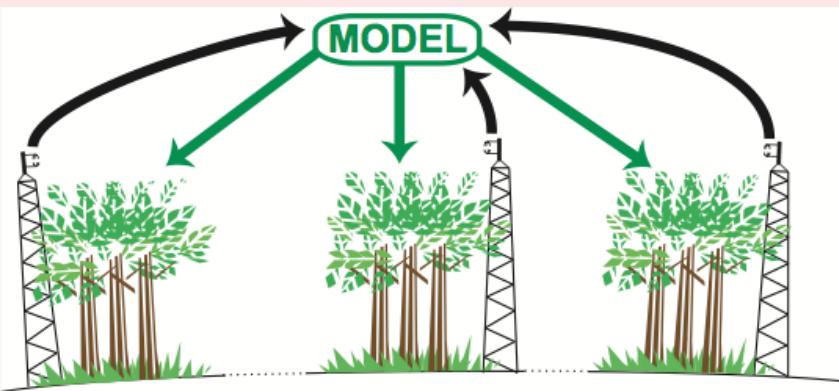
(Wang et al., 2001)  
(Braswell et al., 2005)  
(Knorr and Kattge, 2005)  
(Santaren et al., 2007)  
(Wang et al., 2007)  
(Williams et al., 2009)  
(Carvalhais et al., 2010)  
(Verbeeck et al., 2011)  
...

## ...and evaluation at other sites



(Medvigy et al., 2009; Verbeeck et al., 2011)

## Multi-site optimization



- Starting....
- Kuppel et al. 2012 (ORCHIDEE)
- Groendijk et al. 2011 (simple model)

# Assimilation of Flux data

Ex: temperate Deciduous Broadleaf Forest  
use 12 sites with > 70 % DBF coverage



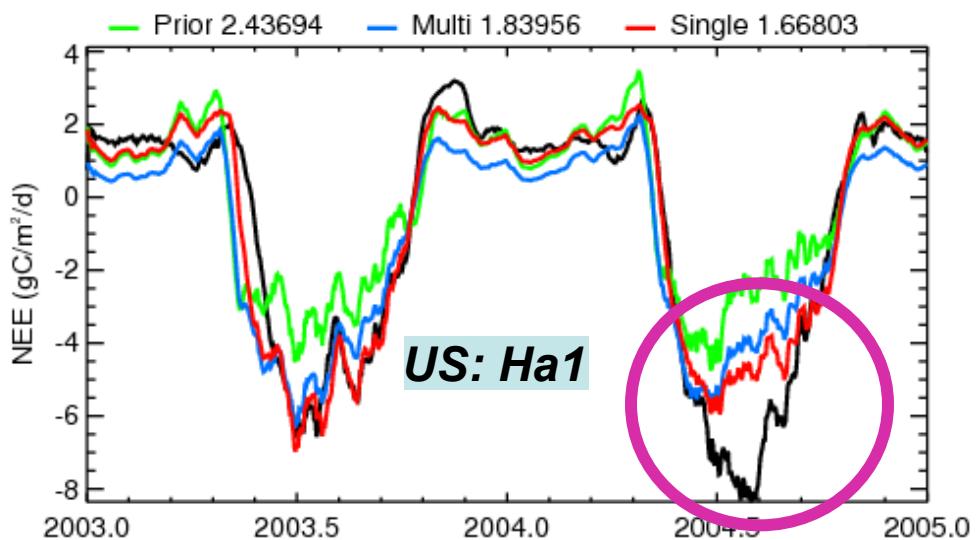
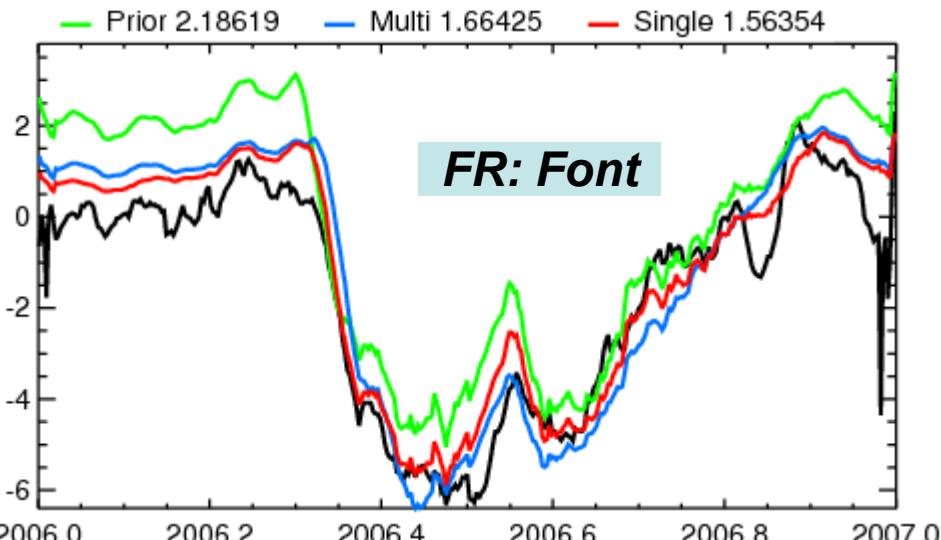
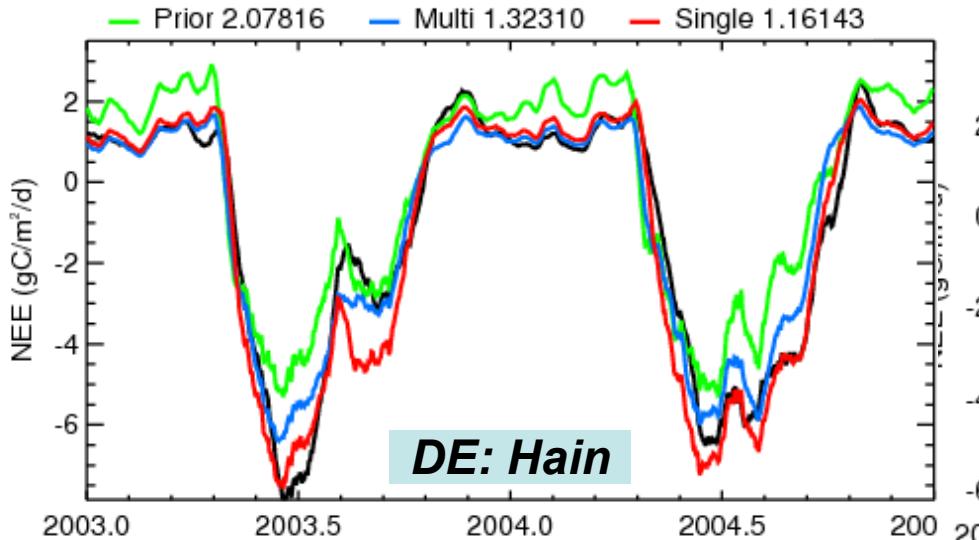
- **Obs type : NEE & Latent heat**
- **Resolution : daily data**
- **period : 3 to 4 years per site**

# Optimized model parameters : 21 per PFTs

---

Name	Description	Associated processes	Genericity
$V_{cmax}$	Maximum carboxylation rate	Photosynthesis	PFT
$G_{s,slope}$	Slope of assimilation in stomatic conductance	Photosynthesis	PFT
$c_{Tmin}, c_{Topt}$	Offset for minimum/optimal photosynthesis temperature	Photosynthesis	PFT
SLA	Specific leaf area (LAI per dry matter content)	Photosynthesis, Respiration	PFT
$K_{pheno,crit}$	Multiplicative factor for growing season start threshold	Phenology	PFT
$c_{T, senescence}$	Offset for temperature threshold for senescence	Phenology	PFT
$LAI_{MAX}$	Maximum LAI per PFT	Photosynthesis, Phenology, Energy balance	PFT
$L_{agecrit}$	Average critical age for leaves	Phenology	PFT
$K_{lai,happy}$	LAI threshold to stop carbohydrate use	Photosynthesis, Phenology	PFT
$Hum_{cste}$	Root profile	Photosynthesis, Water stress	PFT
$Dpu_{cste}$	Total depth of soil water pool	Water stress, Energy balance	Global
Q10	Temperature dependence of heterotrophic respiration	Heterotrophic respiration	Global
$K_{soilC}$	Multiplicative factor of initial carbon pools	Heterotrophic respiration	Site
$b_H, c_H$	Humidity dependence of heterotrophic respiration	Heterotrophic respiration	Global
$MR_b, MR_a$	Offset and first-degree coefficient for temperature dependence of maintenance respiration	Maintenance respiration	PFT
$GR_{frac}$	Fraction of biomass allocated to growth respiration	Growth respiration	PFT
$Z0_{overheight}$	Characteristic rugosity length	Energy balance	Global
$K_{albedo,veg}$	Multiplying factor for surface albedo	Energy balance	Global

# Model – FluxNet data fit : ex. for 3 sites



**NEE ( $\text{gC}/\text{m}^2/\text{d}$ )**

**Data**

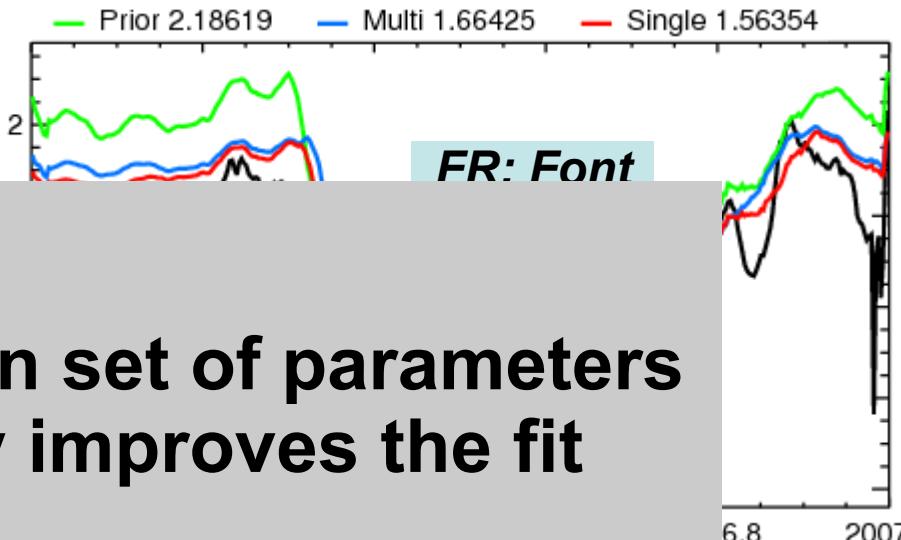
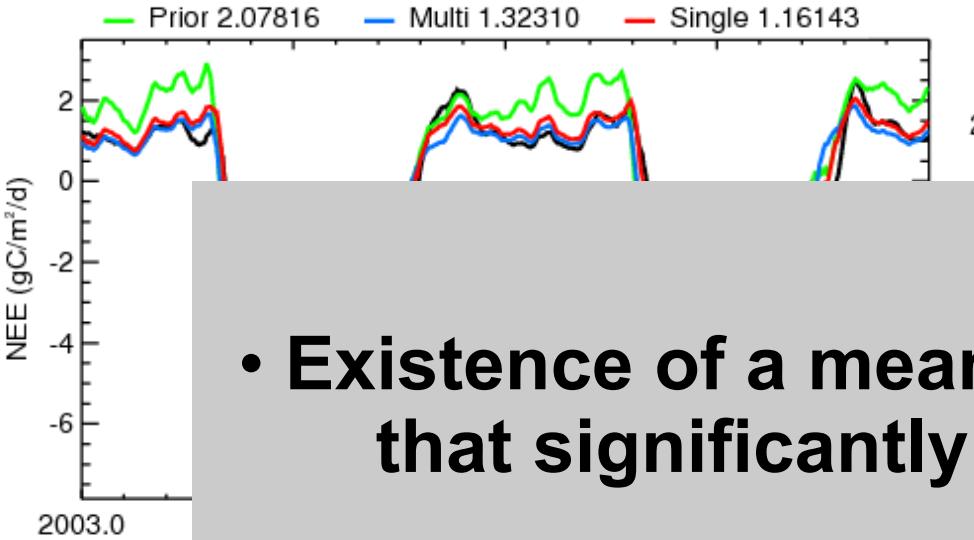
**Prior**

**Single-site posterior**

**Multi-site posterior**

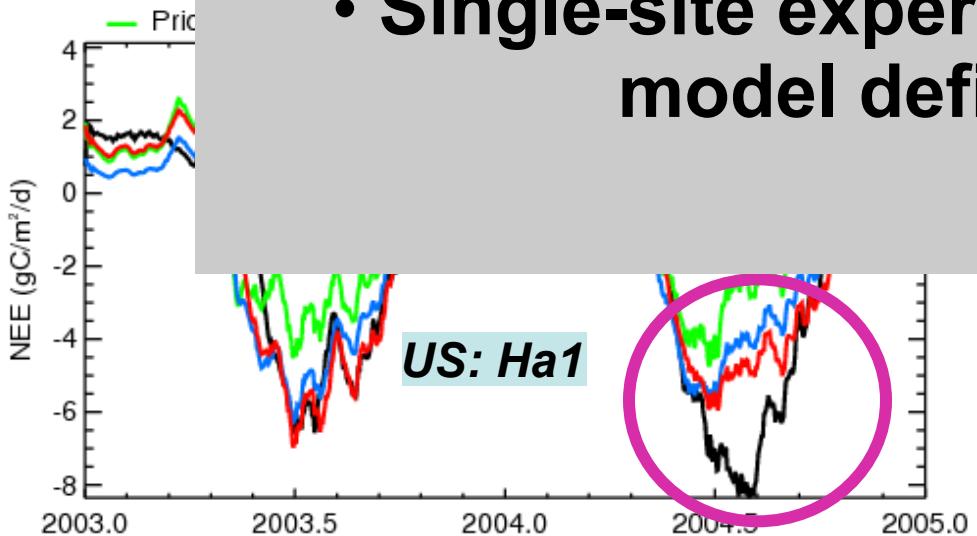
*Kuppel et al. 2012*

# Model – FluxNet data fit : ex. for 3 sites



- Existence of a mean set of parameters that significantly improves the fit**

- Single-site experiments highlight model deficiencies..**



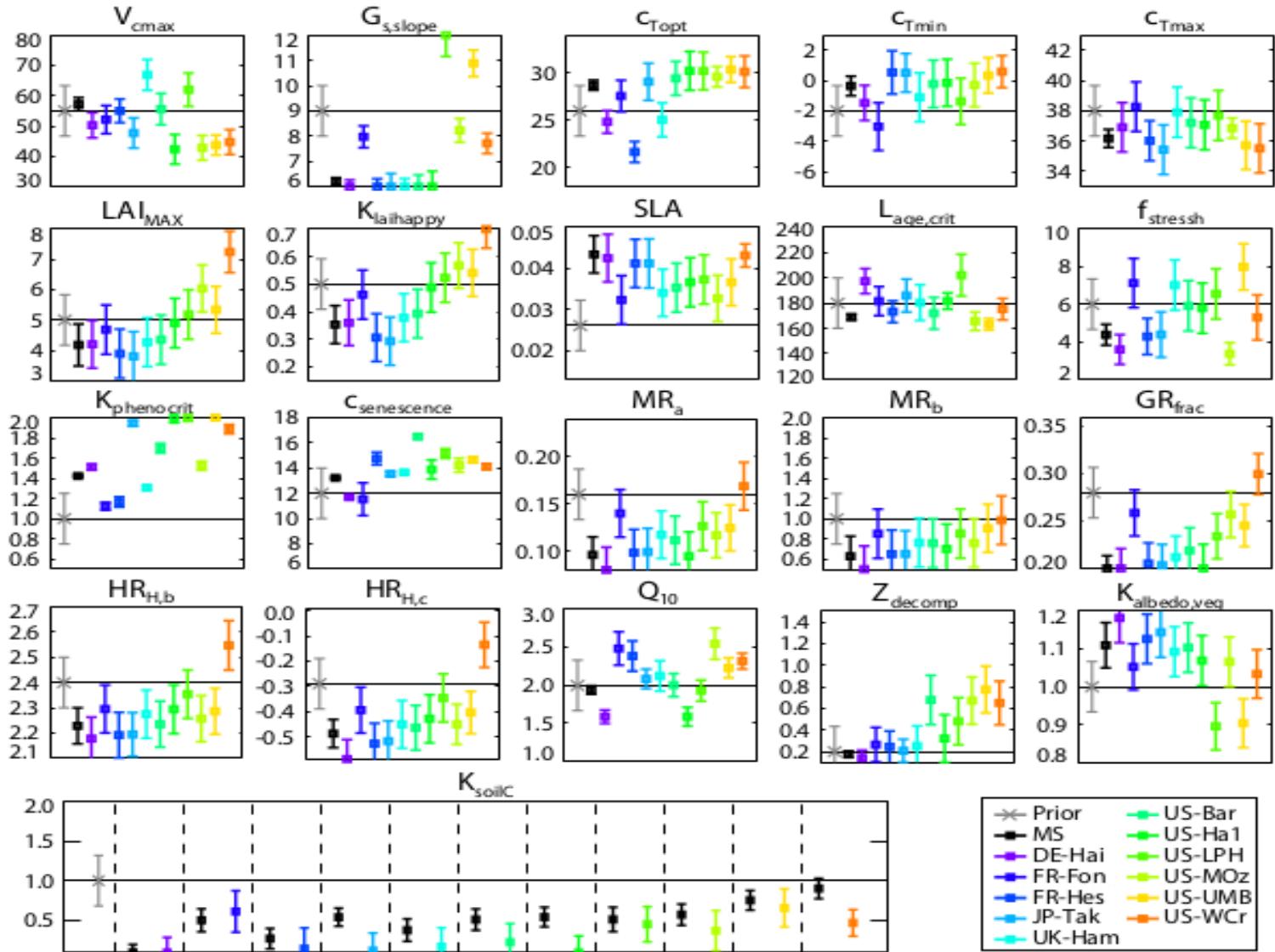
Prior

Single-site posterior

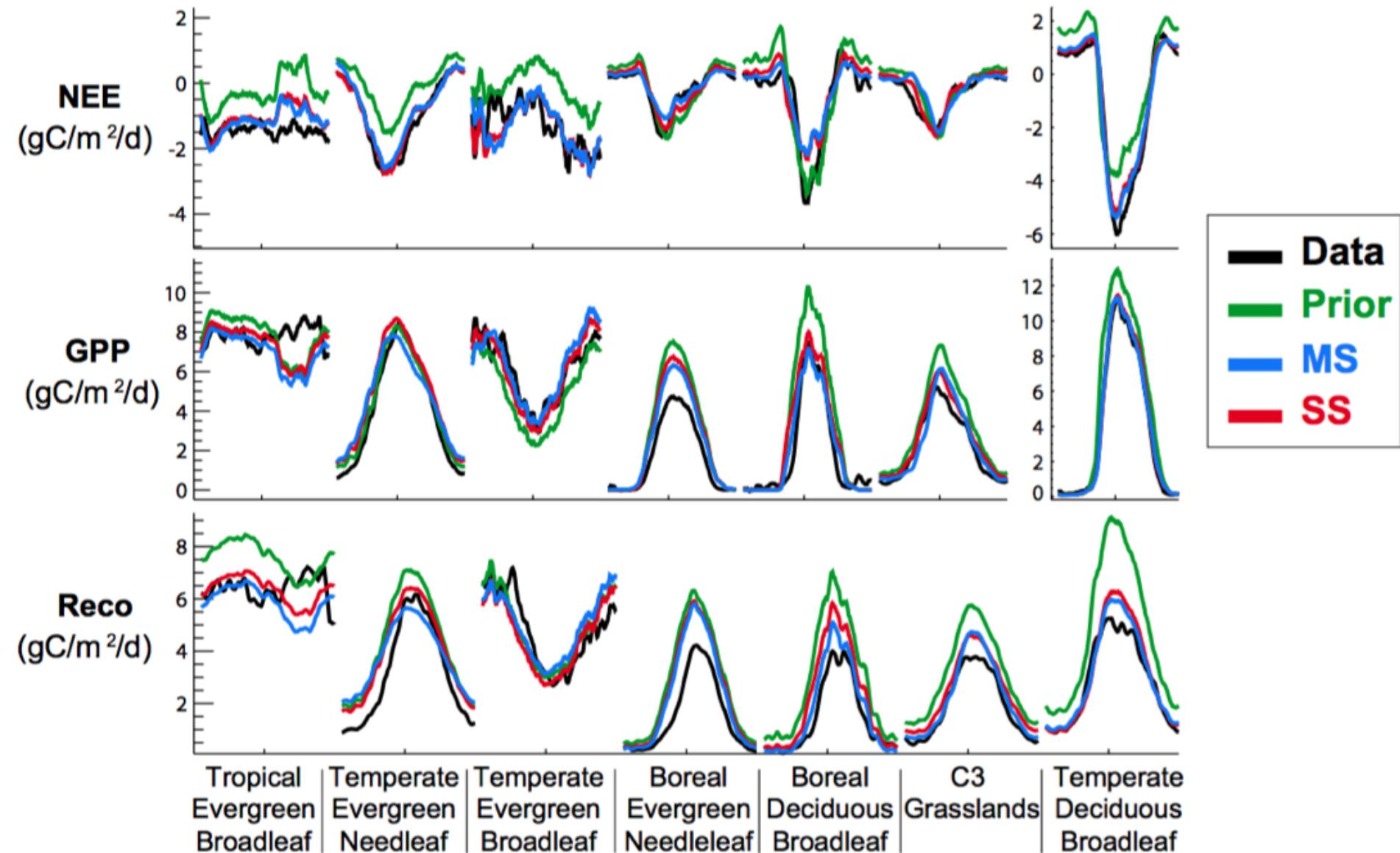
Multi-site posterior

# Parameters errors

**Black: Multi-site Colors: Single-site**



# Results for all Plant Functional Types...



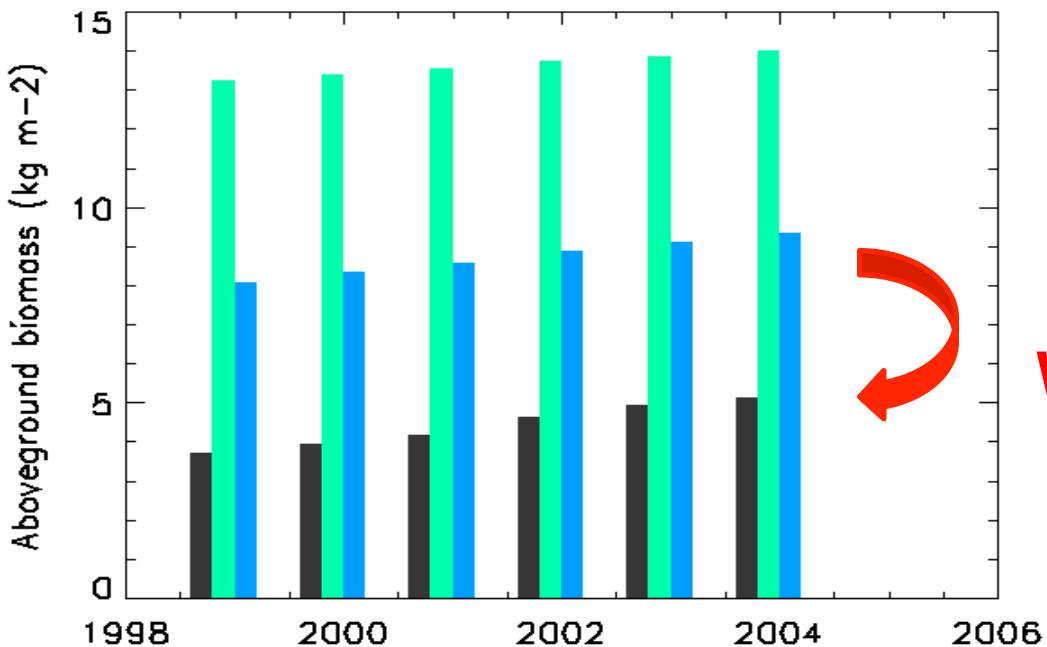
# Outline

---

- Current limitations of « standard » atmospheric flux inversions
- Multi-data streams assimilation: Basis for model parameters optimization (CCDAS)
- Potential of several land data streams
  - Satellite vegetation indexes
  - Fluxnet data
  - Biomass measurements
- Join multi-data assimilation
- Limitations & Prospects

# Assimilation of Biomass measurements (ex: Site level ; Beech Forest ; France)

Above ground biomass (Hesse site) : Prior model output



Wrong input  
and / or  
Wrong mortality

Measurement  
Model: Steady state  
Model: Realistic age

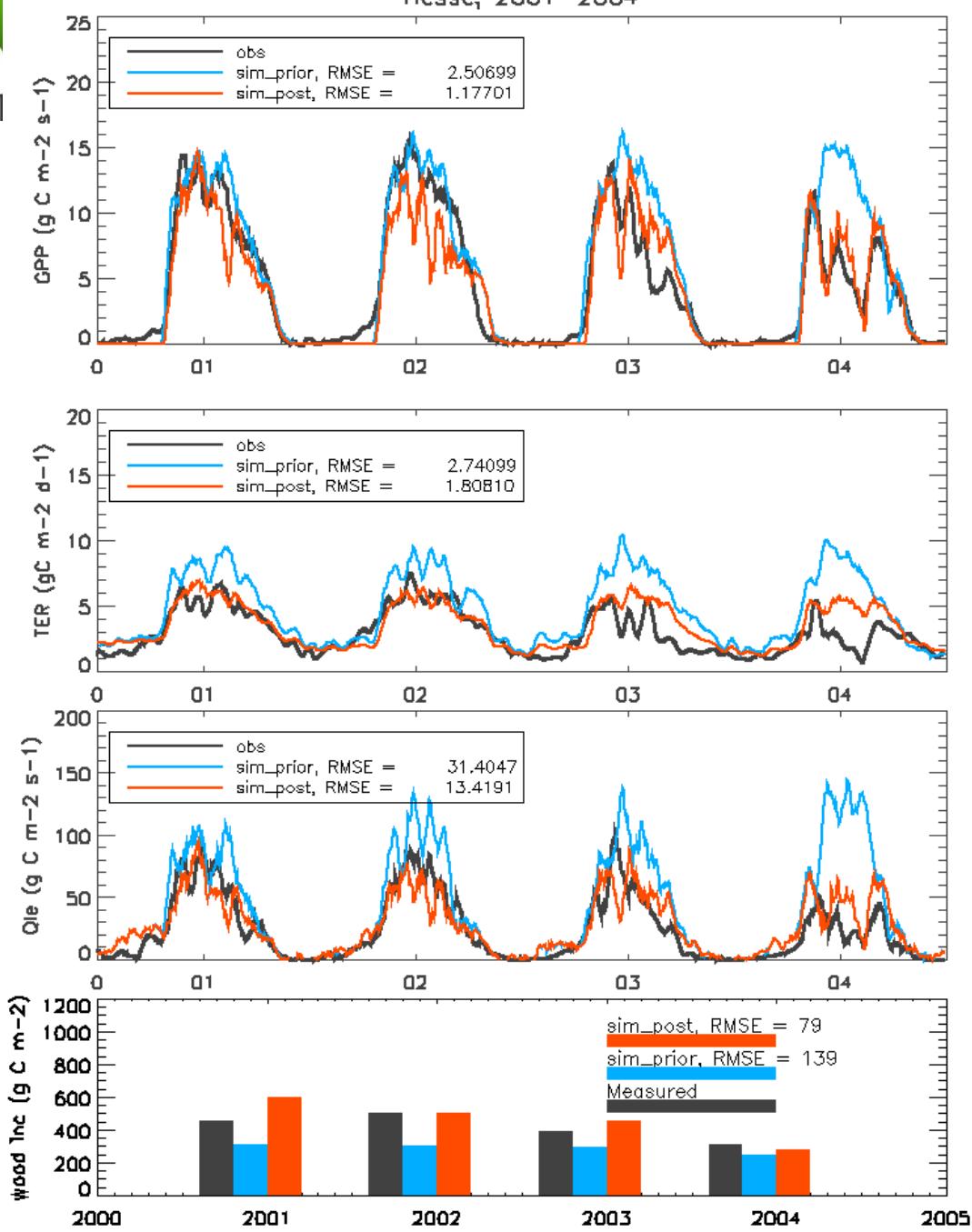
# Hesse site: Assimilation of Flux data & Yearly Biomass incr.

(25 flux related params)

Measurement

Prior model  
 $GPP_{RMSE} = 2.5$   
 $TER_{RMSE} = 2.7$   
 $Qle_{RMSE} = 31$

Posterior model  
 $GPP_{RMSE} = 1.1$   
 $TER_{RMSE} = 1.8$   
 $Qle_{RMSE} = 13$



# Outline

---

- Current limitations of « standard » atmospheric flux inversions
- Multi-data streams assimilation: Basis for model parameters optimization (CCDAS)
- Potential of several land data streams
  - Fluxnet data
  - Satellite vegetation indexes
  - Biomass measurements
- Assimilation of Atm. CO<sub>2</sub>
- Join multi-data assimilation
- Limitations & Prospects

MODIS  
NDVI

FluNet  
NEE / LE

Atmospheric  
CO<sub>2</sub>

4 phenology  
params  
per PFTs

4 + ( $\approx$  15)  $\approx$  20  
params per PFTs  
(photosynthesis,  
respiration)

Initial soil C = 50 params  
+ 3 params per PFTs  
(from previous set)

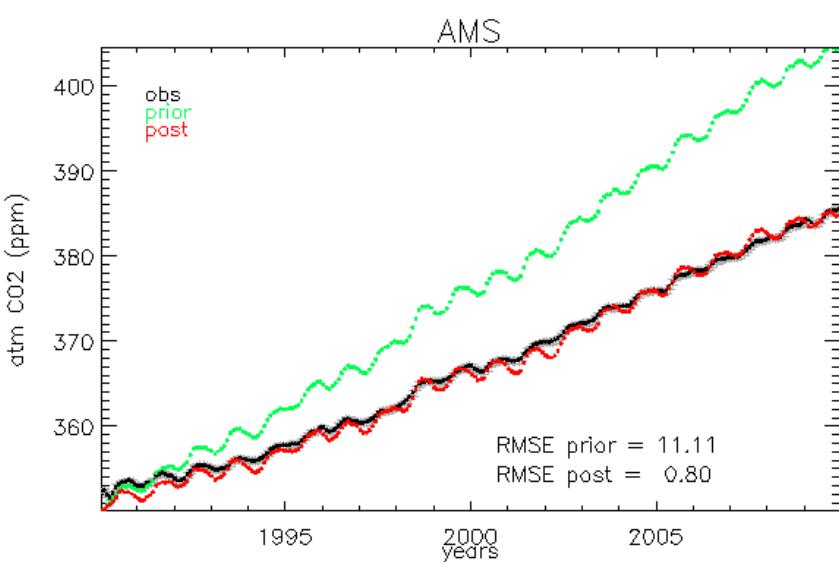
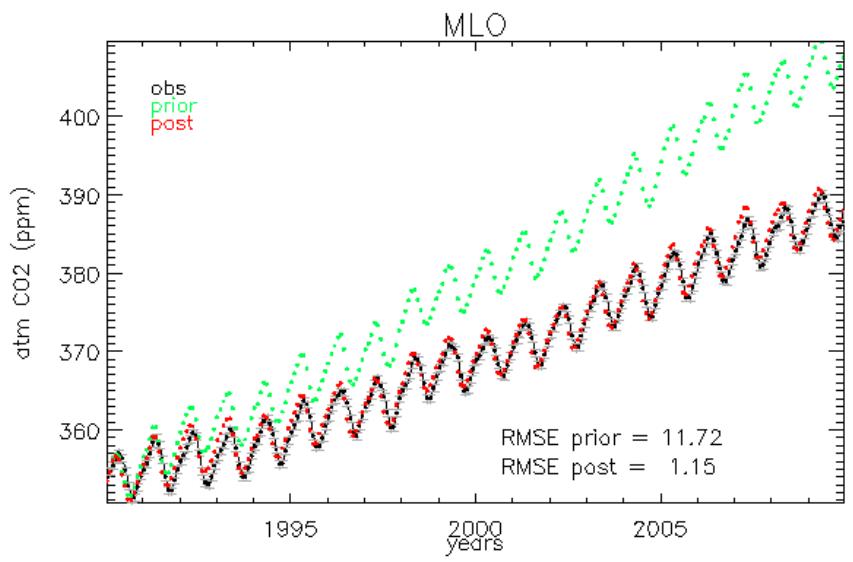
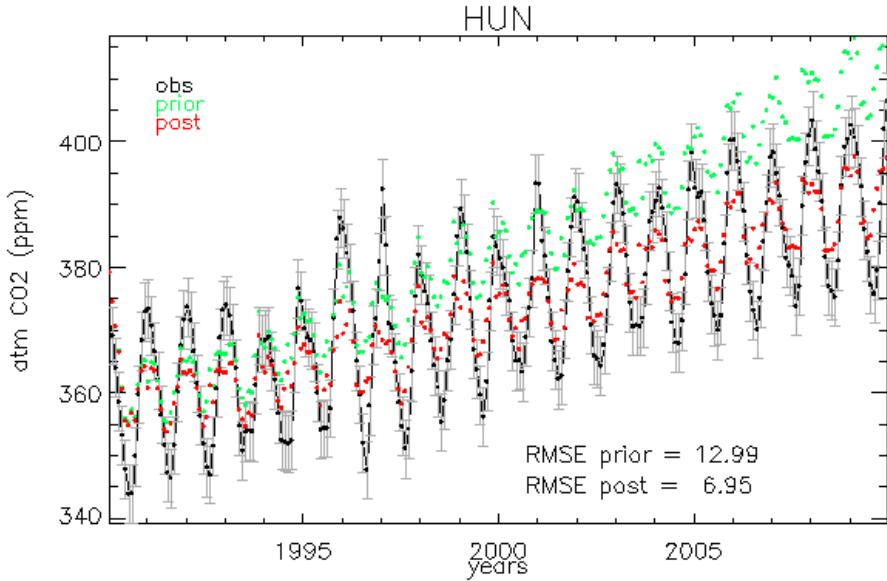
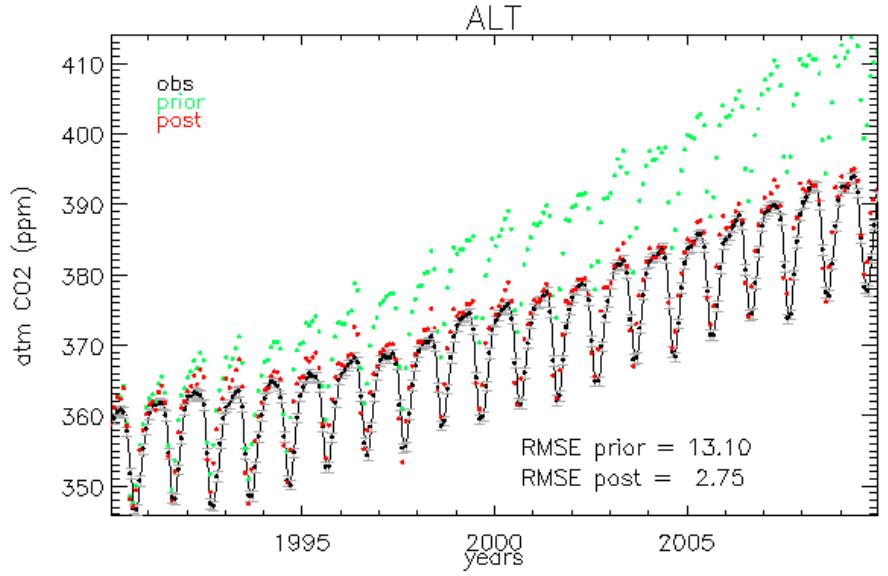
40 params

$\approx$  100 params

$\approx$  80 params

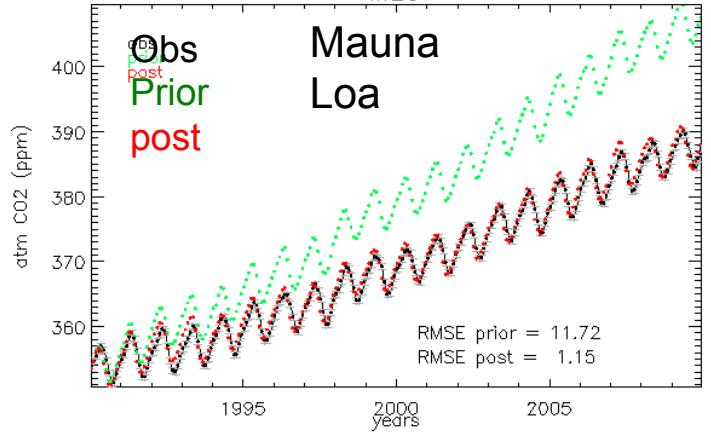


# LSCE-CCDAS : Fit to the data



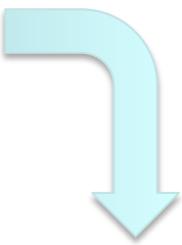
# Assimilation of atmospheric [CO<sub>2</sub>] data

## Optimization of the CO<sub>2</sub> trend

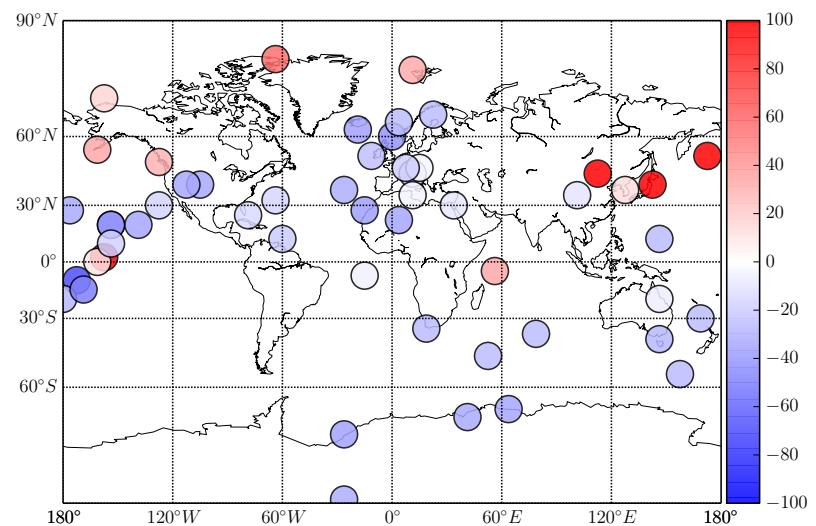


Signal decomposition:

- Amplitude : max – min
- Phase : CPU



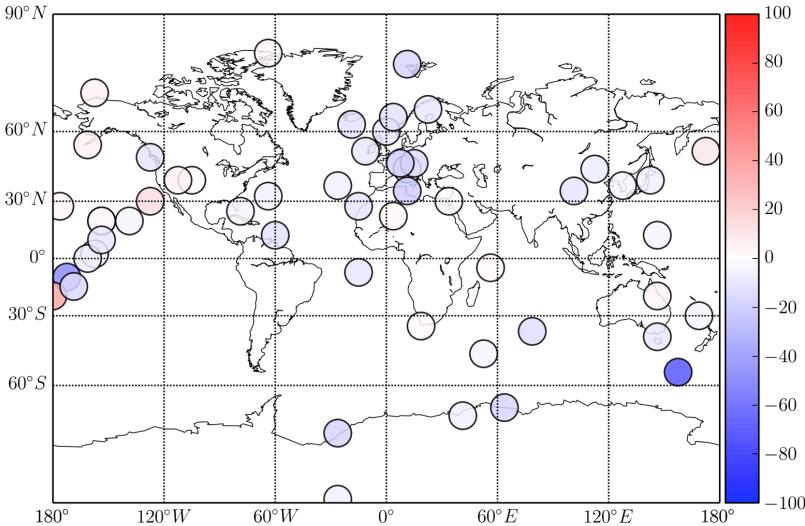
## Seasonal amplitude



Degradation

Improvement

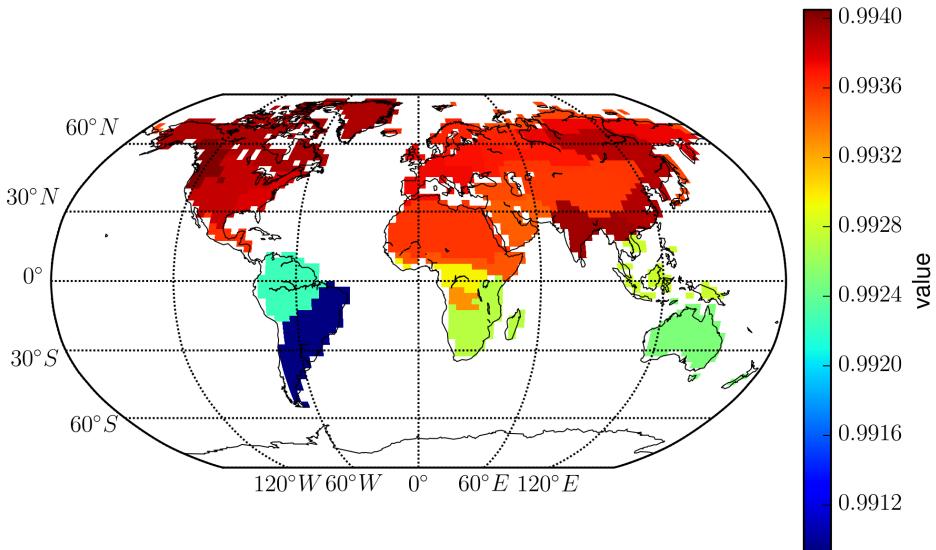
## Carbon uptake period length



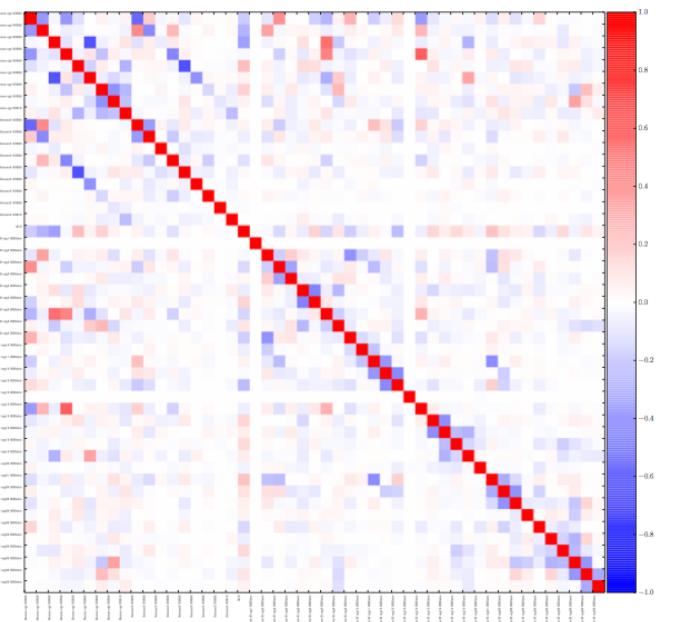
# Assimilation of atmospheric $\text{CO}_2$ data

→ Primary constraint on:

- Soil initial carbon pools..

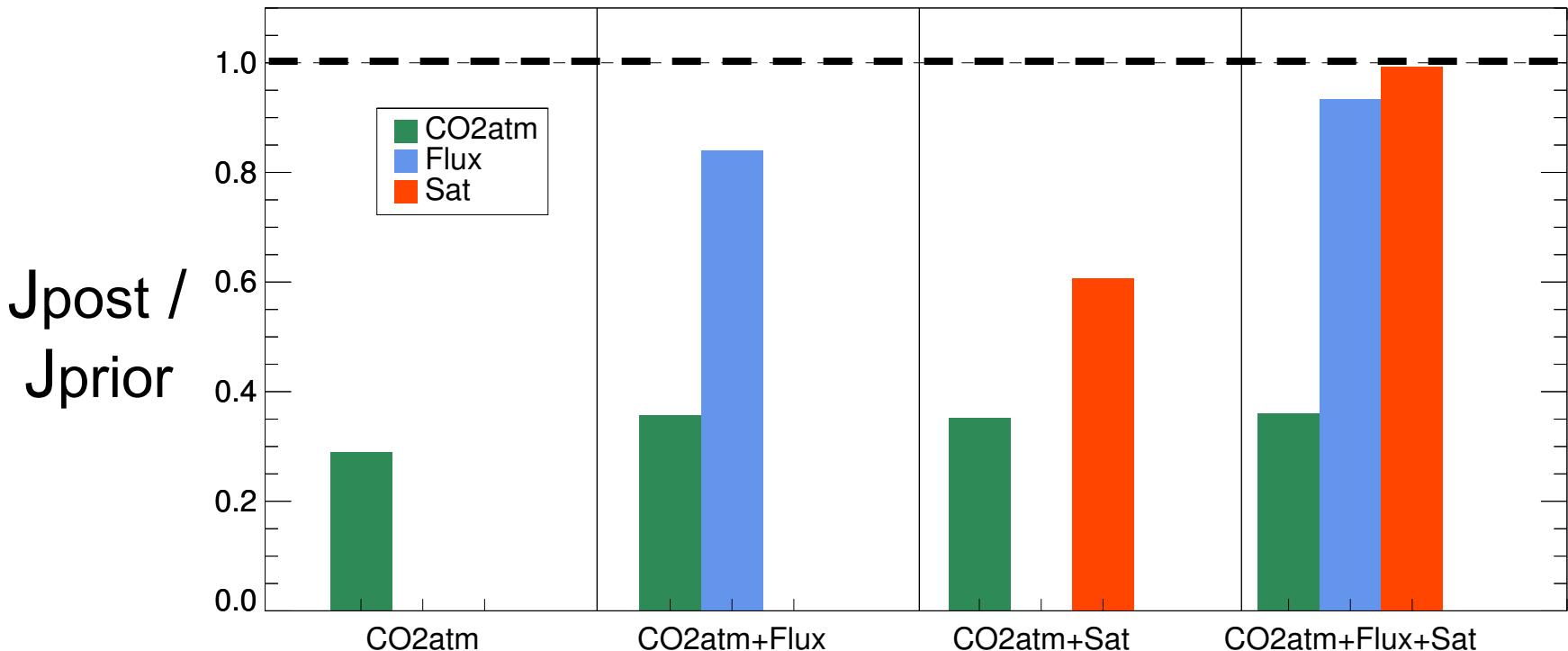


→ And significant error correlations between parameters



- Very few emerging studies...
- Results with BETHY ecosystem model
  - + TM2 transport model
- ✓ Assimilation of Atm CO<sub>2</sub> and satellite fAPAR  
(Scholze *et al.* 2007)
- Preliminary results with ORCHIDEE + LMDz
  - ✓ MODIS-NDVI + FluxNET + Atm CO<sub>2</sub>
  - ✓ 3 years

# LSCE CCDAS - Join assimilation : [CO2] & other data streams



Atm. [CO2]  
FLUXNET  
MODIS-NDVI

# Outline

---

- Current limitations of « standard » atmospheric flux inversions
- Multi-data streams assimilation: Basis for model parameters optimization (CCDAS)
- Potential of several land data streams
  - Fluxnet data
  - Satellite vegetation indexes
  - Biomass measurements
- Join multi-data assimilation
- Limitations & Prospects

# Summary: Potential of a CCDAS..

---

- Promizing approach to account for multi-data streams
- Helps to identify model deficiencies !
- Relative Error characterization bw data stream becomes crucial for a proper assimilation
- Anticipated data streams to become crucial:
  - soil-C observations...
  - data from Ecosystem Manipulative Experiments
- Recent large community effort :
  - GeoCarbon EU-project (5 land CCDAS)
  - Existing inter-comparison of Model-Data fusion exercise

# Limitations of a CCDAS...

---

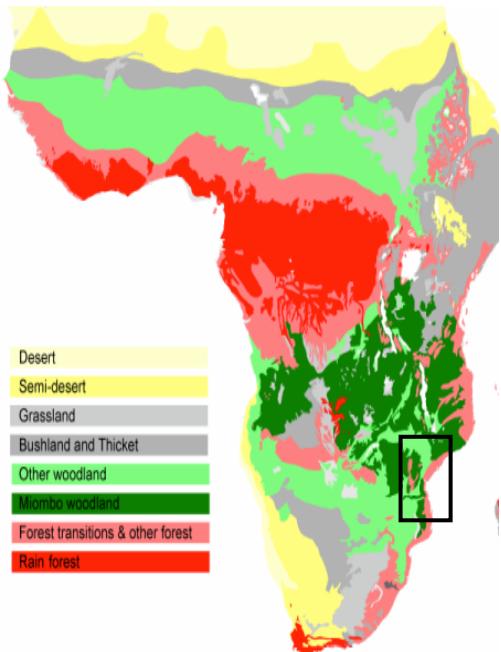
- Strongly rely on a given model structure!
- Missing processes in the ecosystem model might lead to
  - Wrong parameter estimates
  - Poor model predictability (Strong biases)
- Non-linearity might complicate the parameter optimization
- Need to :
  - keep independent data for model output validation
  - Keep classical Atmospheric inversion

---

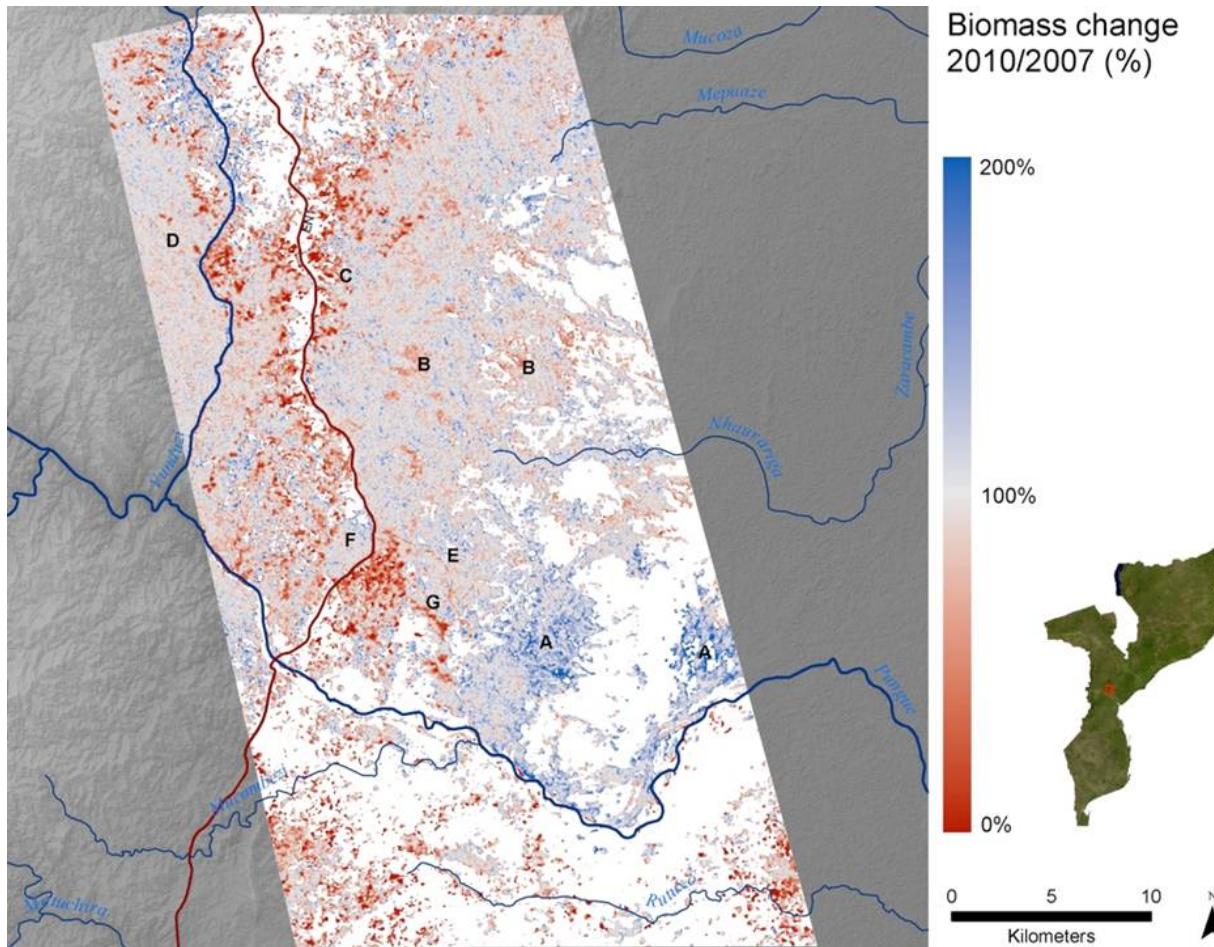
# Aditional slides

# Using radar retrievals of forest biomass data Over tropical woodlands

Biome demography  
is critical for  
African woodlands



Biomass change  
from 2007 to 2010

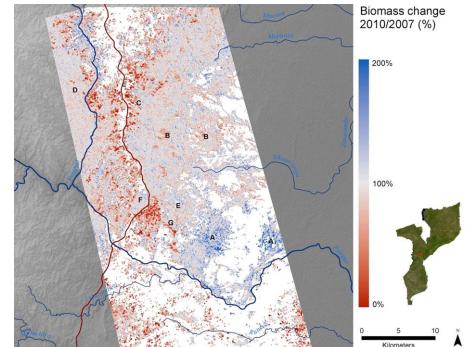


Courtesy of  
Mathew Williams et al.

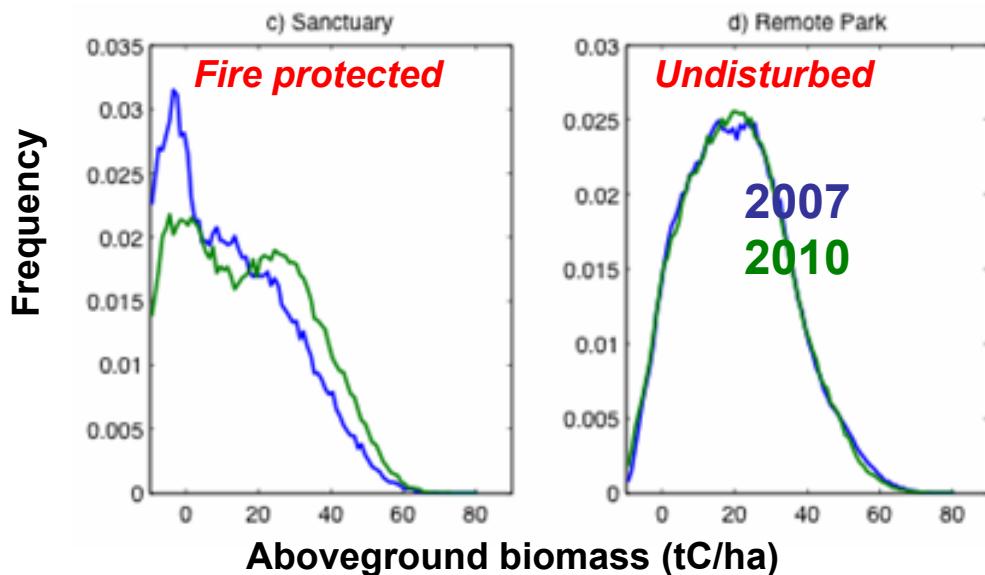
# Using radar retrievals of forest biomass data Over tropical woodlands



$$\Delta C_w = a_w \text{NPP} - t_w C_w - P F C_w$$



Biomass change ( $\Delta C_w$ ) is determined by growth (NPP), tree lifespan ( $t_w$ ) and by the **probability ( $P$ ) and intensity ( $F$ ) of disturbance**



→ Assimilation scheme  
to determine parameters  
 $P$  and  $F$

Courtesy of  
Mathew Williams et al.

# Objectives

---

- Illustrate the potential of multi-data C Cycle assimilation systems
- Stress the difficulties of model parameters optimizations...
- Using examples from ORCHIDEE-CCDAS