



Carbon Cycle Data Assimilation System (CCDAS)

Philippe Peylin & several contributors...

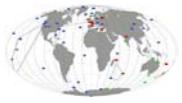
Laboratoire des Sciences du Climat et de l'Environnement
Gif sur Yvette,
France

Objectives

- Illustrate the potential of multi-data C Cycle assimilation systems
- Stress the risks of a model parameters optimizations...
- Using examples for the Land C cycle

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 CO₂ inversions....



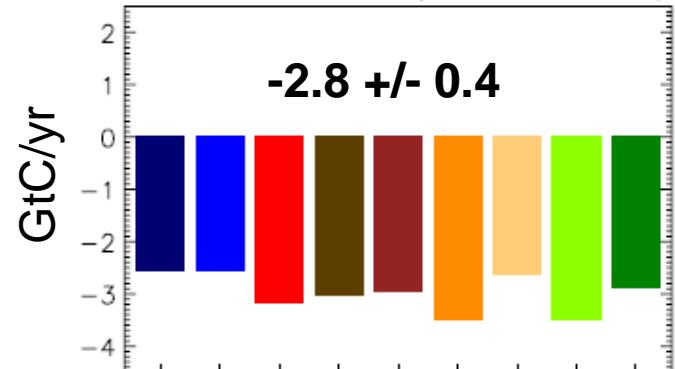
- Top-down approach :
 - ➔ Estimated fluxes account for all surface processes
- Verifiable by independent groups
- Several implementations applied so far...



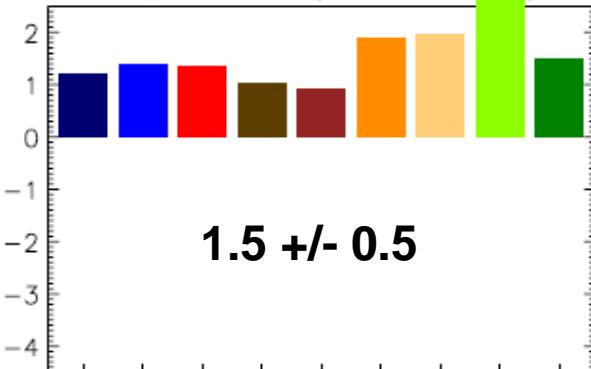
Atmospheric CO₂ inversions....

(mean fluxes 96-04)

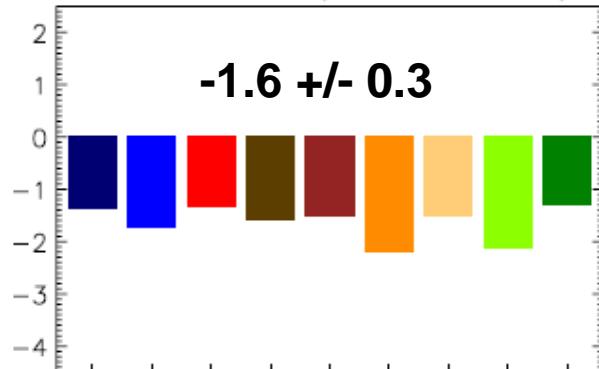
N. Hemisphere



Tropics



S. Hemisphere

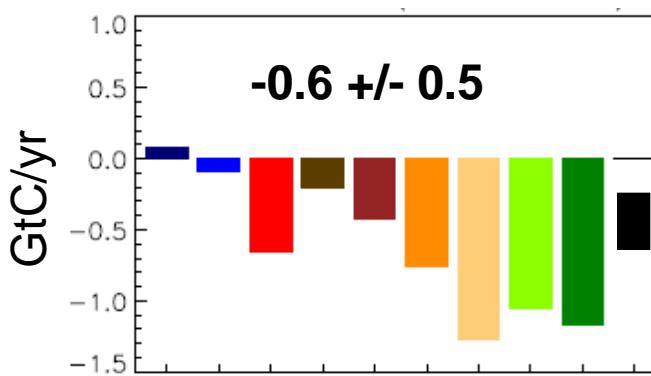


LSCE_v2.1_peylin
 LSCE_v3.0_chevallier
 JENA_s96v3.1_rodenbeck

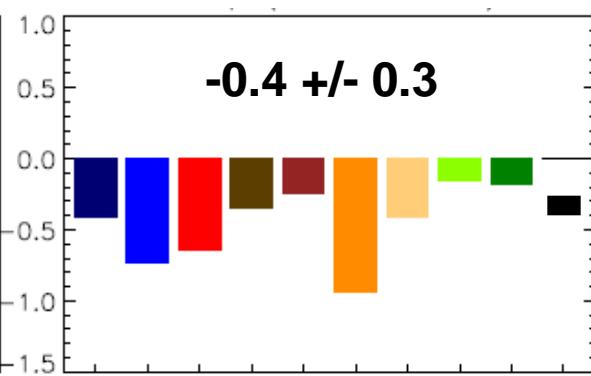
CTracker_US
 CTracker_EU_wouter
 TRANSCOM-3_mean

PSU_butler
 C13_Match_rayner
 C13_CCAM_law

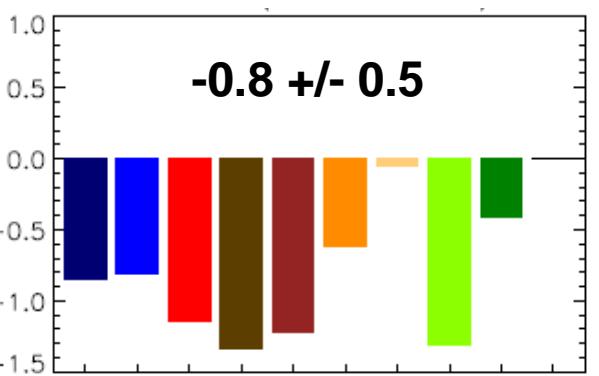
N. America



Europe



N. Asia





Atmospheric CO₂ inversions....

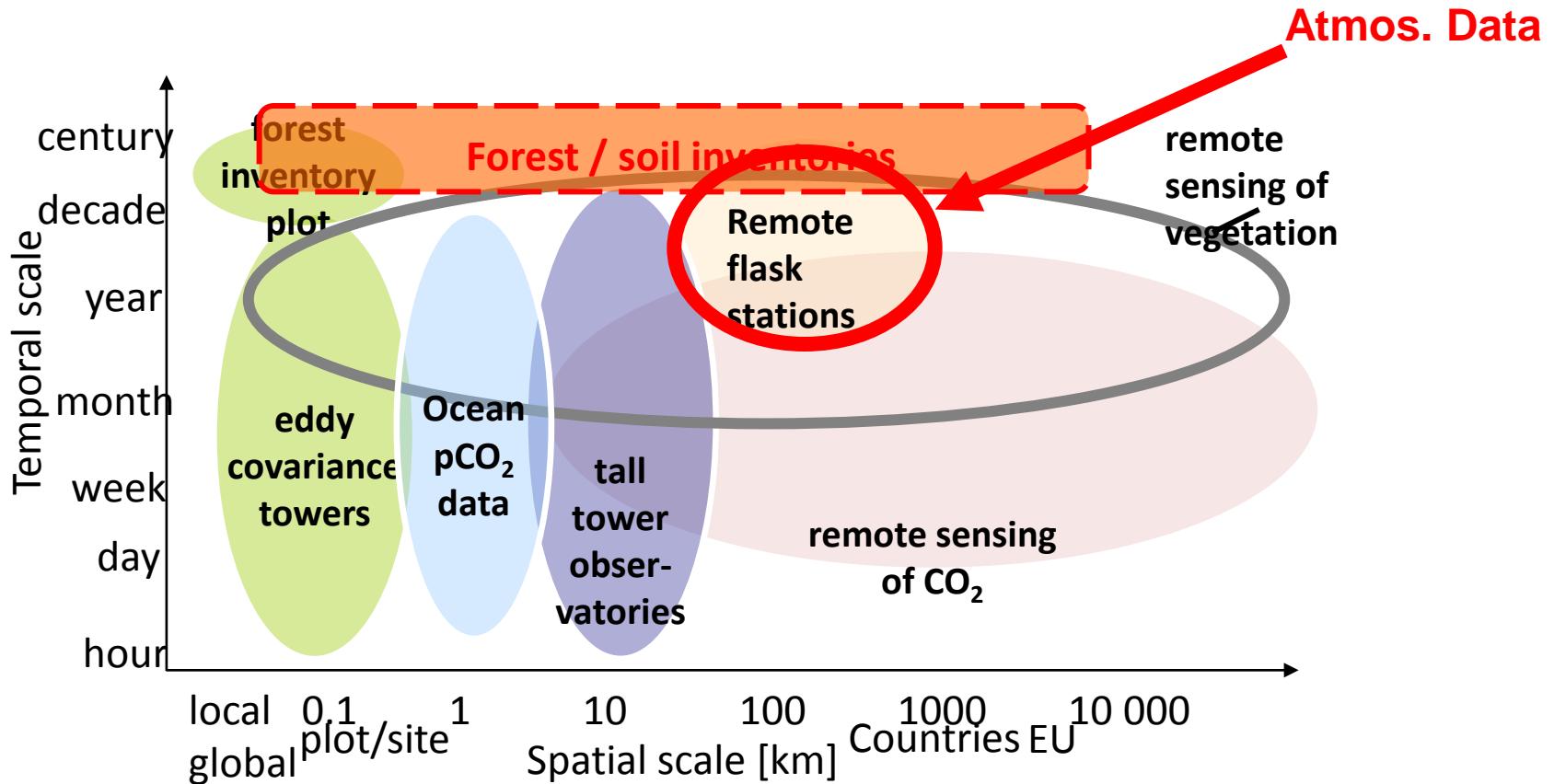
(mean fluxes 96-04)

- Strong constraint at large spatial scales
- But still poor regional constraints
- Depend critically on
 - Transport model errors
 - Prior flux information (var/covar error matrix)
- No insight on the processes..
- Difficulties to handle “network changes”
- No predicting capabilities

How to move forward ?

Strong Need to :

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

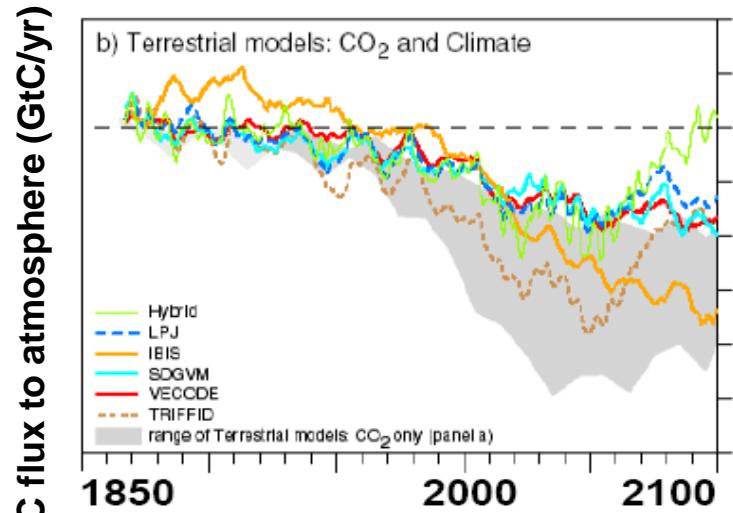


Outline

- 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

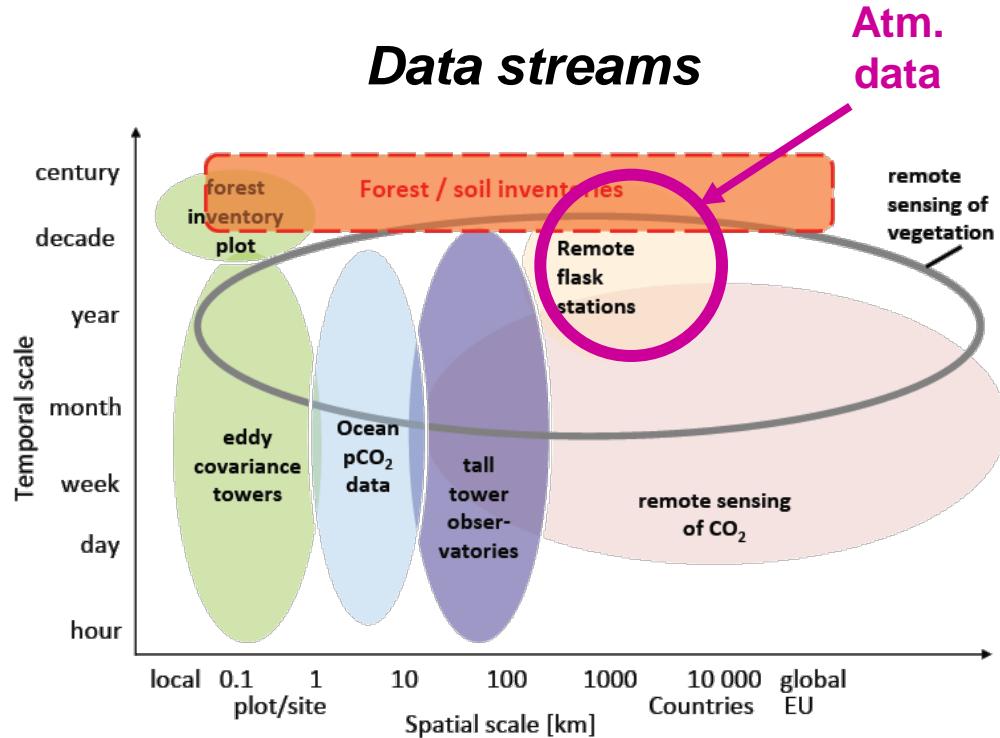
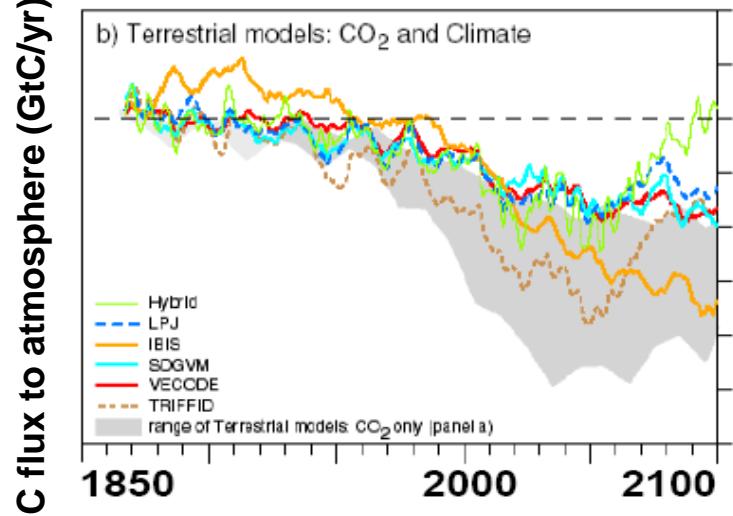
Needs for a Carbon Cycle Data Assimilation System

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



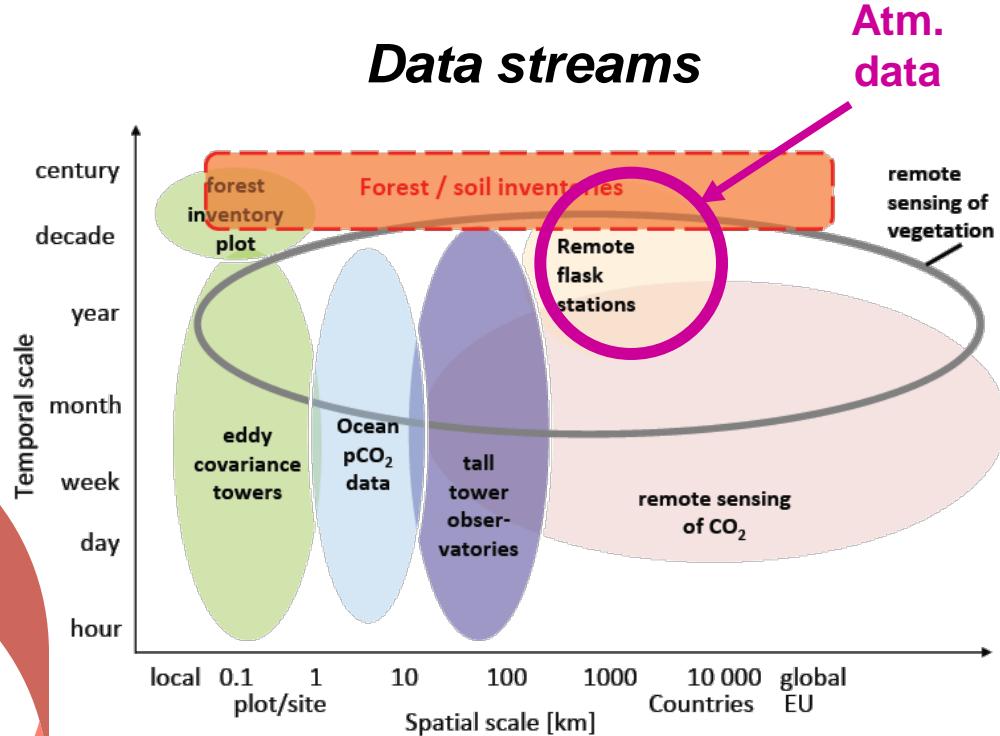
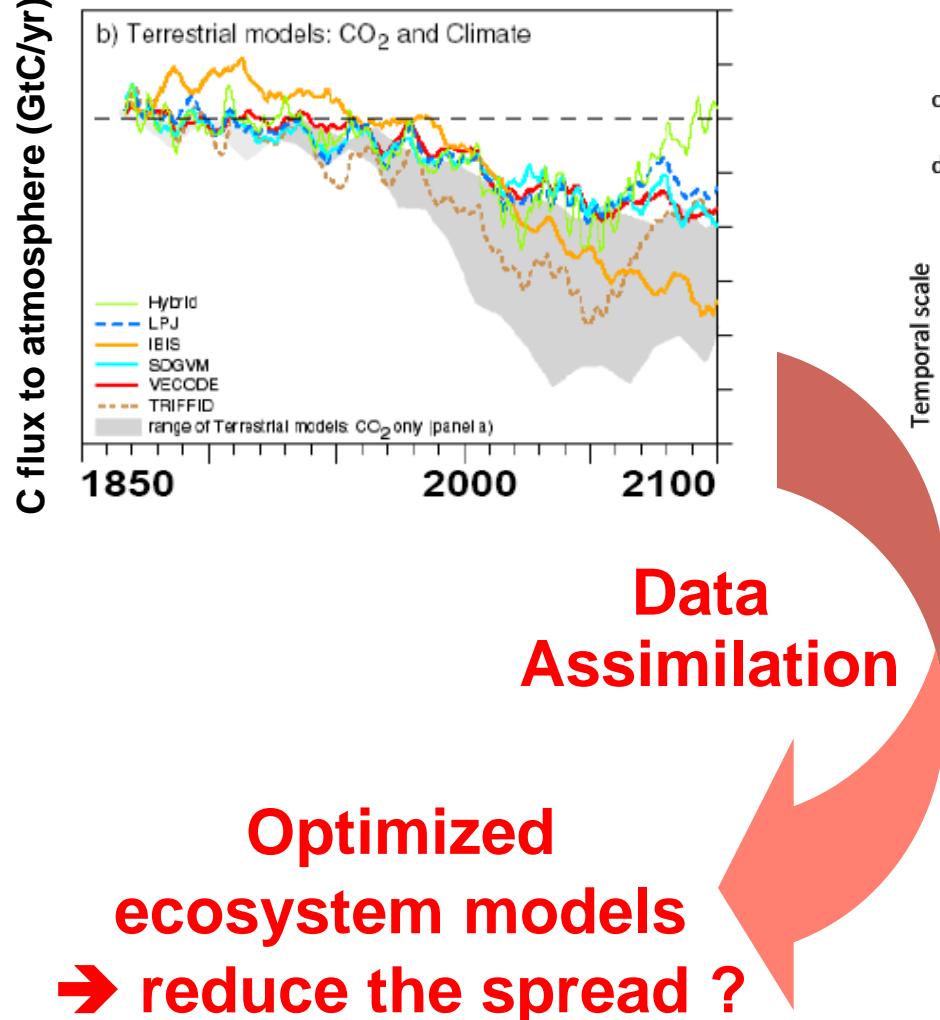
Needs for a Carbon Cycle Data Assimilation System

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

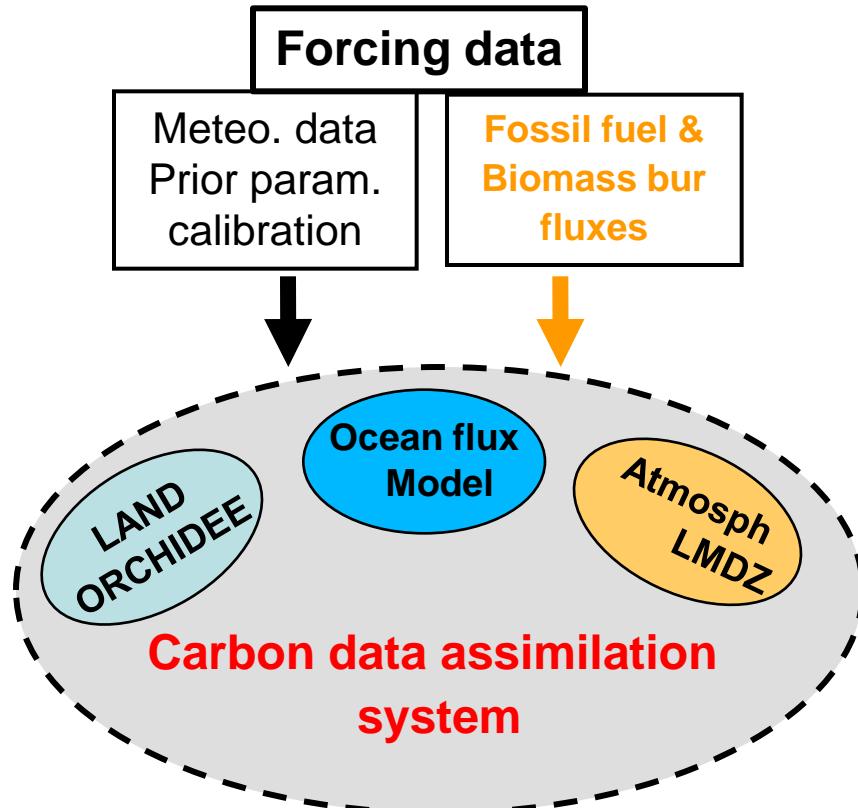


Needs for a Carbon Cycle Data Assimilation System

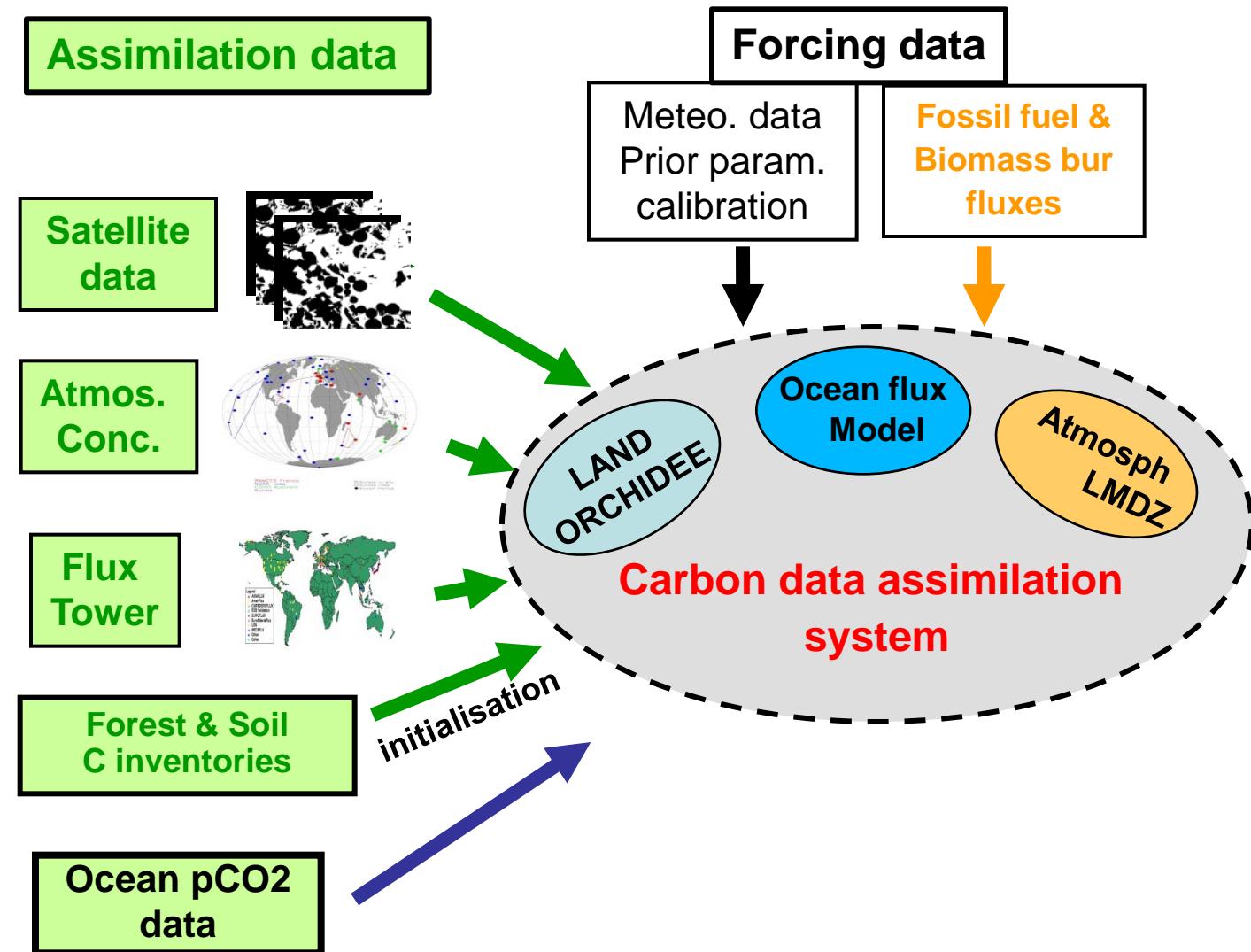
Large uncertainty from land to predict global C-balance (C4MIP)



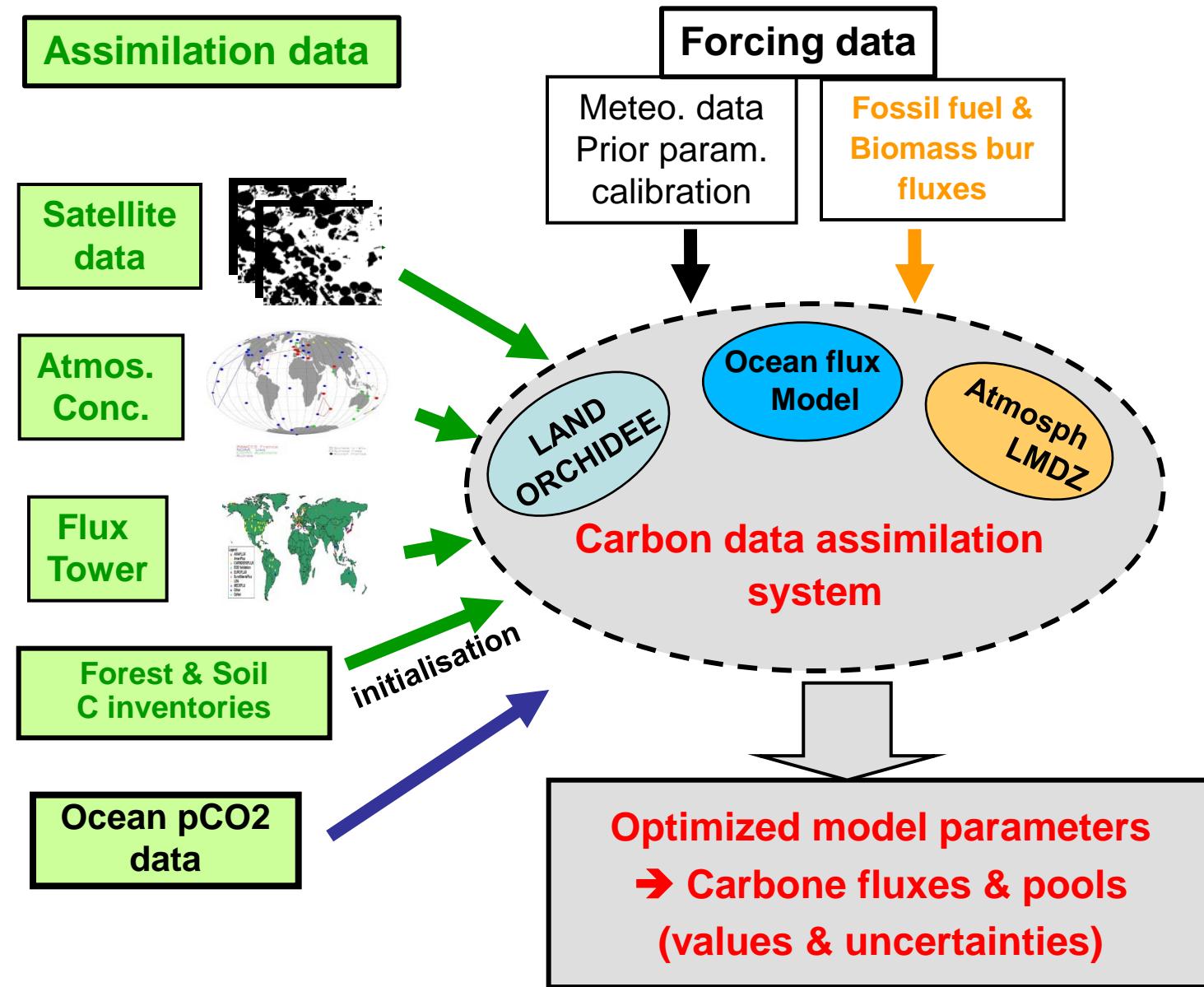
Structure of a global “CCDAS”



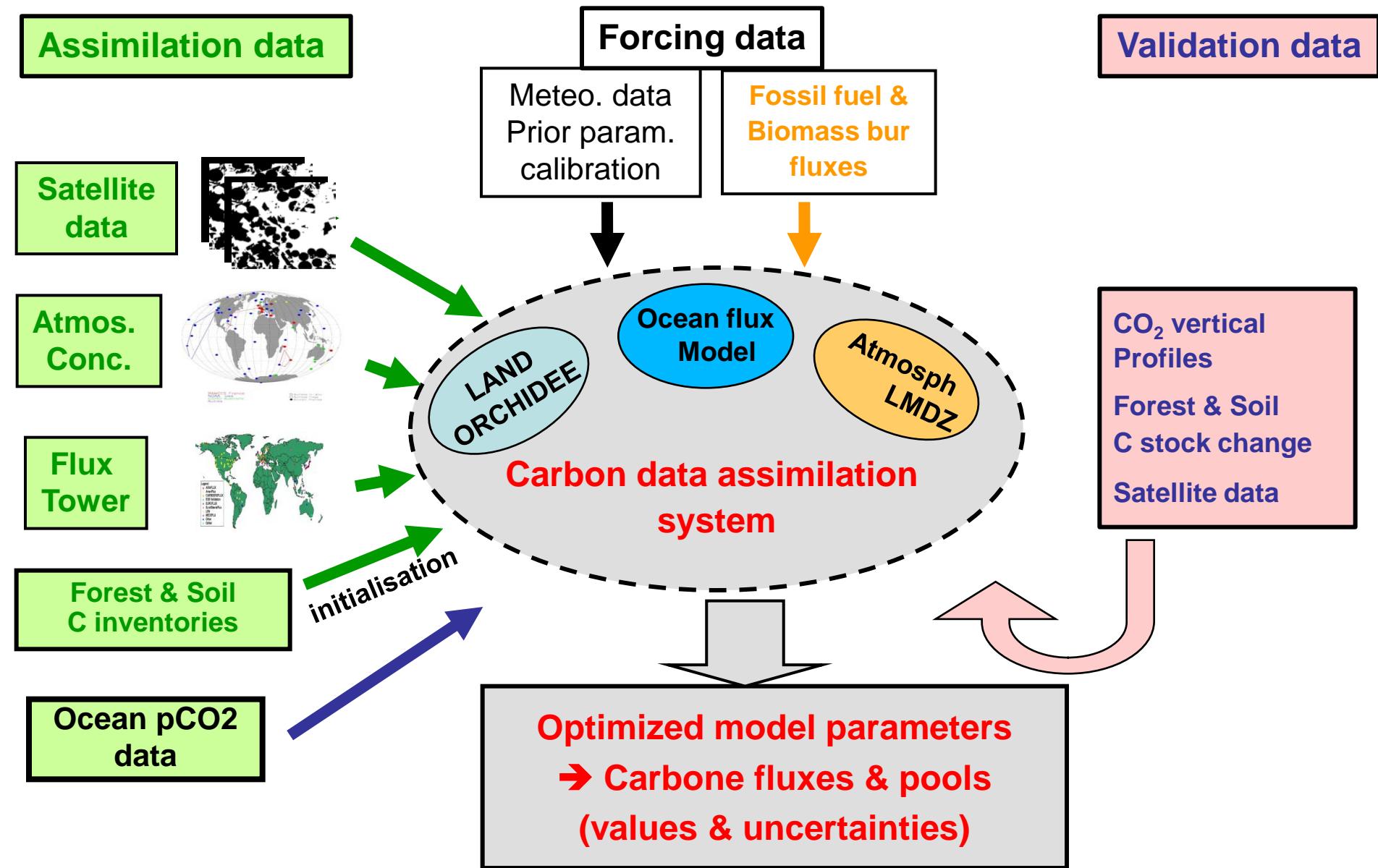
Structure of a global “CCDAS”



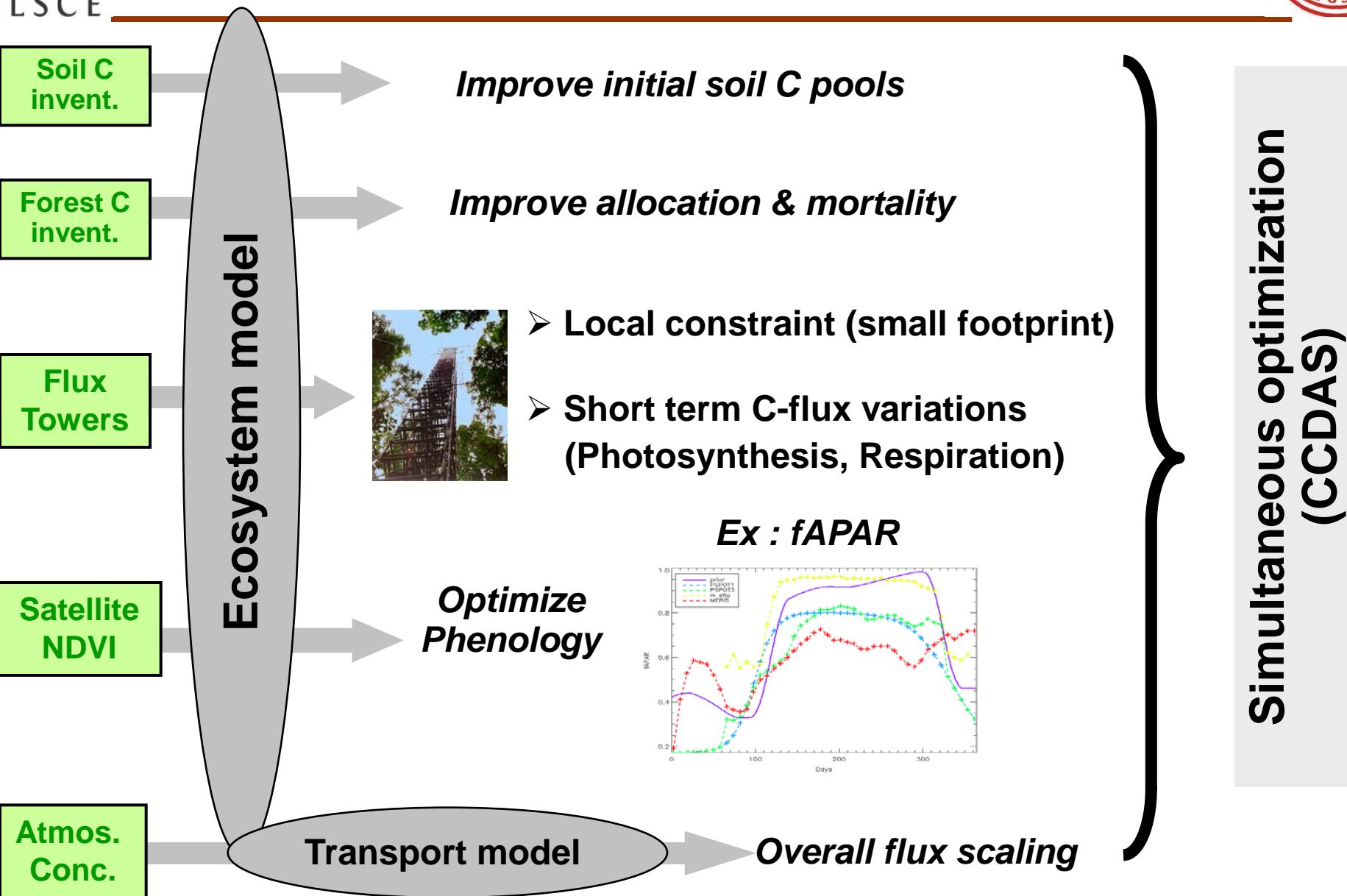
Structure of a global “CCDAS”



Structure of a global “CCDAS”



Land CCDAS components



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

Minimize the cost function $J(\mathbf{x})$ to obtain the mean of $p(\mathbf{x}|\mathbf{y})$

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

\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**

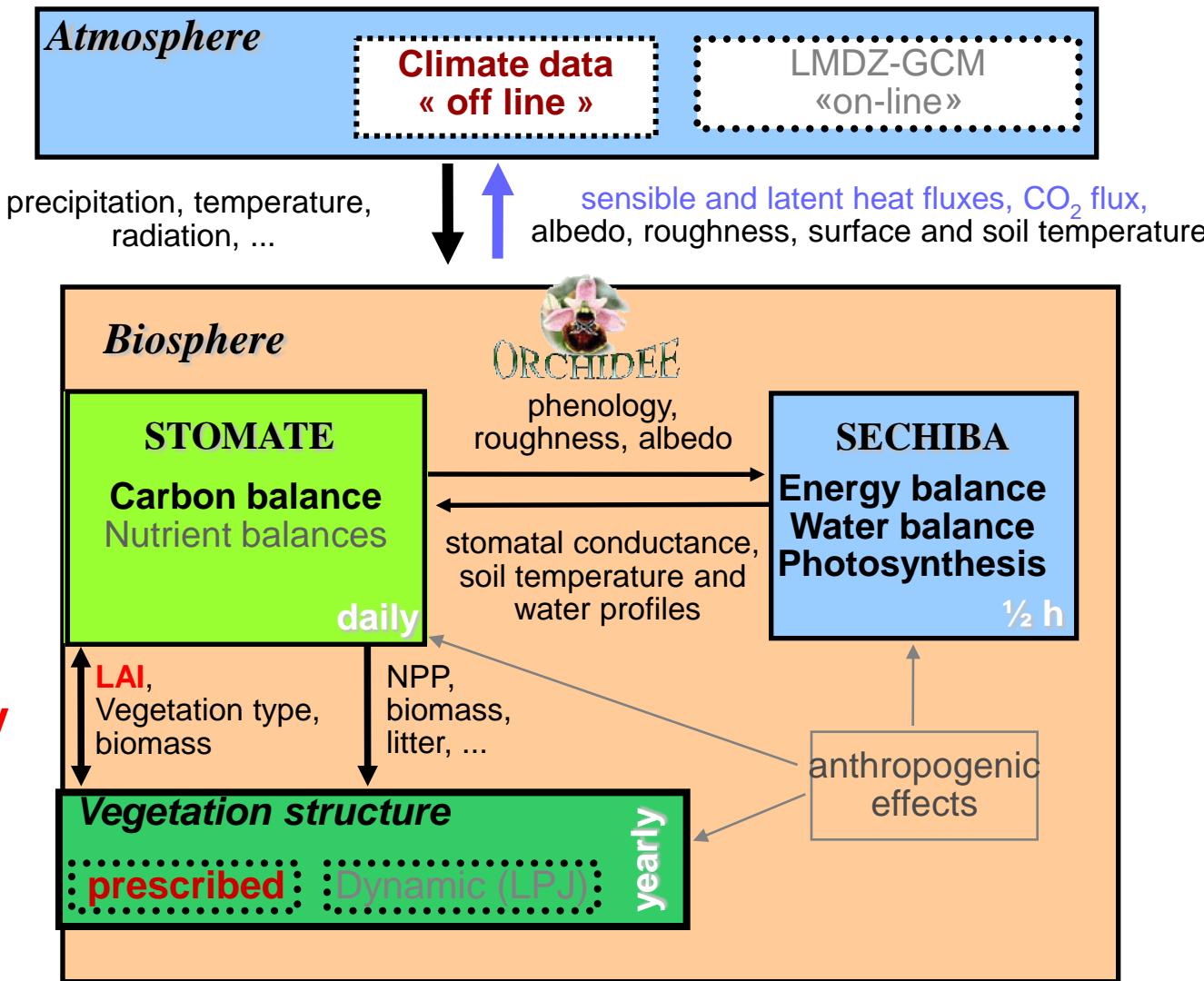
- Used mostly for site-level studies
- Required time usually prohibitive with “complex model”
- No limitations wrt LINEARITY & parameter PDF

Outline

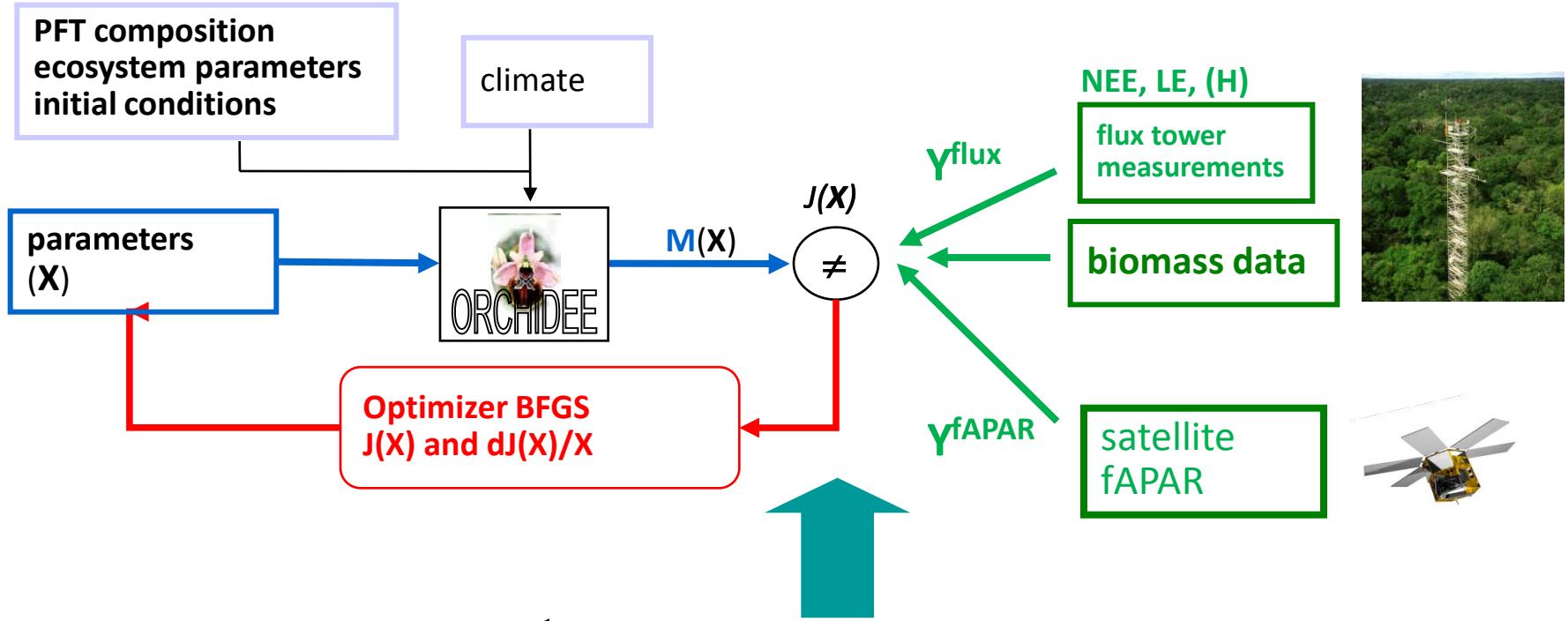
- Current limitations of « standard » atmospheric flux inversions
- Multi-data streams assimilation: Basis for model parameters optimization (CCDAS)
- Potential of several land data streams
 - Fluxnet data
 - Satellite vegetation indexes
 - Biomass measurements
- Join multi-data assimilation
- Limitations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

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



Implementation..



Cost function:
$$J(x) = \frac{1}{2} \left[(y - M(x))^t R^{-1} (y - M(x)) + (x - x_b)^t P_b^{-1} (x - x_b) \right]$$

- Iterative minimization using either:

- Variational approach (with Tangent Linear model for DJ/dx)
- Monte Carlo approach

Assimilation of Flux data

- **Several recent studies :**
 - ➔ synthesis of the different approaches in Williams et al. 2009.
- **Single site studies:**
 - Wang et al. 2001, 2007
 - Braswell et al. 2005
 - Knorr et al., 2005
 - Santaren et al. 2007
- **« Accross site » studies**
 - Verbeeck et al. 2009
 - Medvigy et al. 2011



Assimilation of Flux data

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

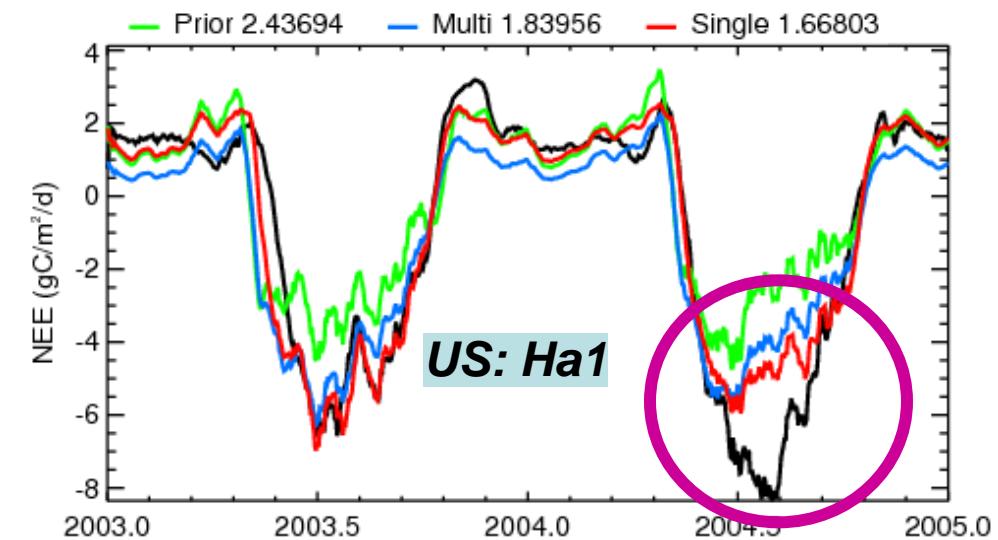
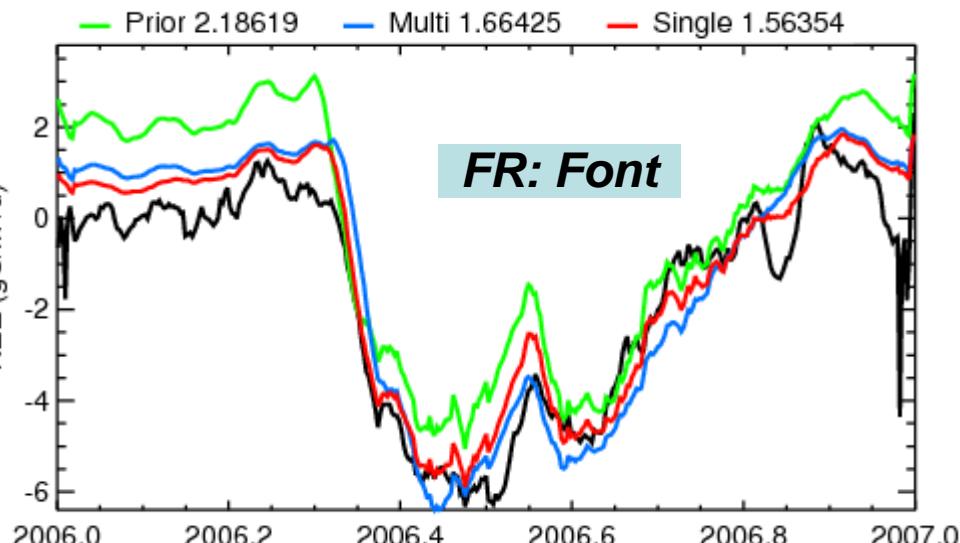
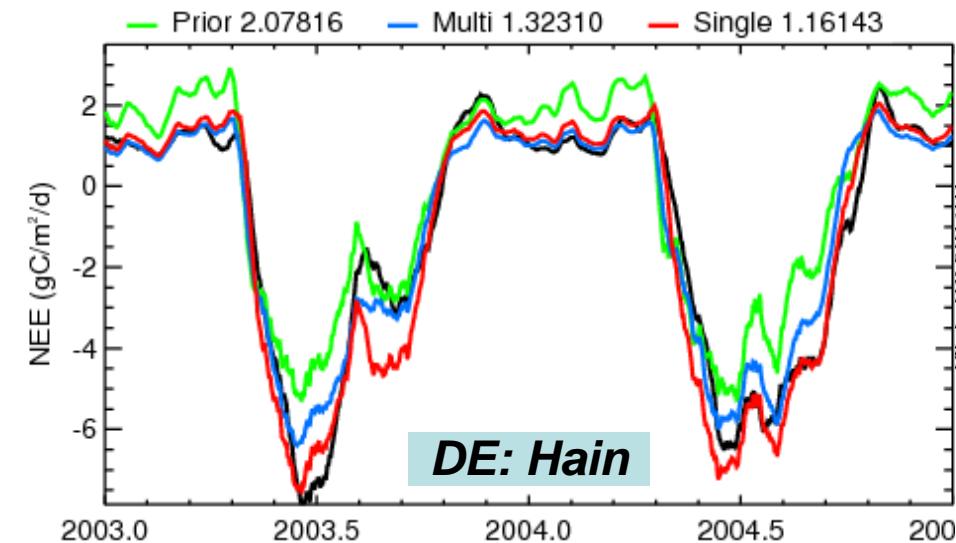


- **Obs type : NEE & Latent heat**
- **Resolution : daily data**
- **period : 3 to 4 years per site**

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



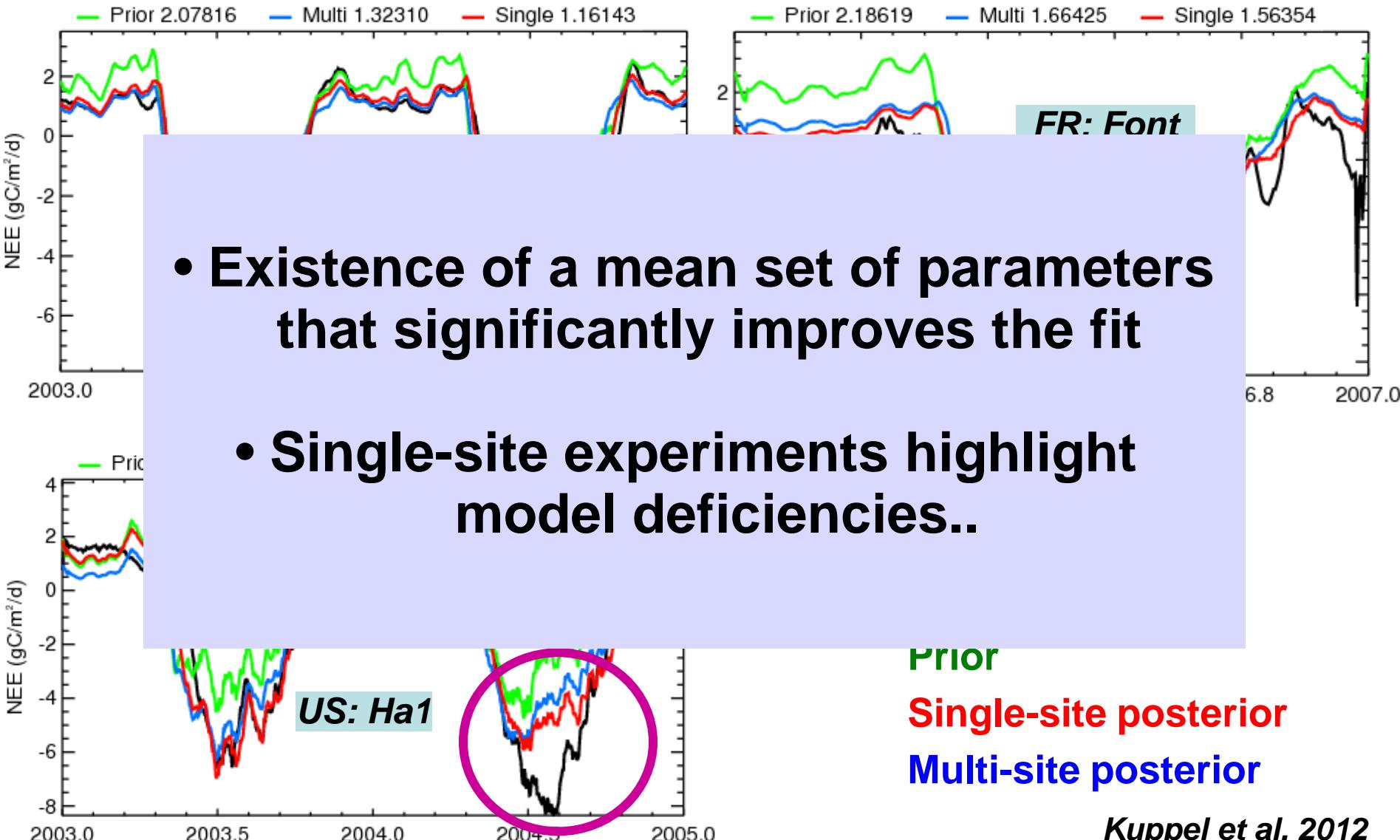
NEE ($\text{gC}/\text{m}^2/\text{d}$)

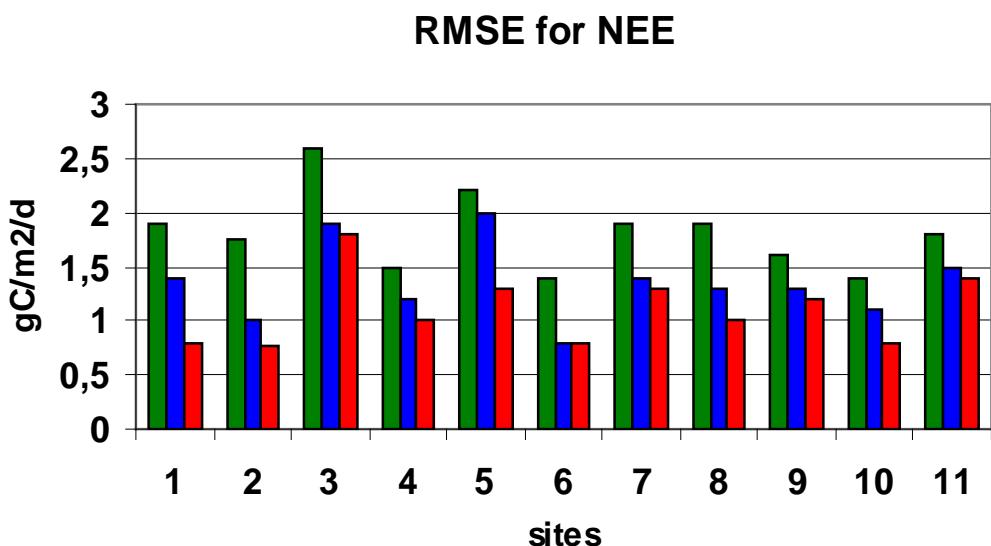
Data

Prior

Single-site posterior

Multi-site posterior



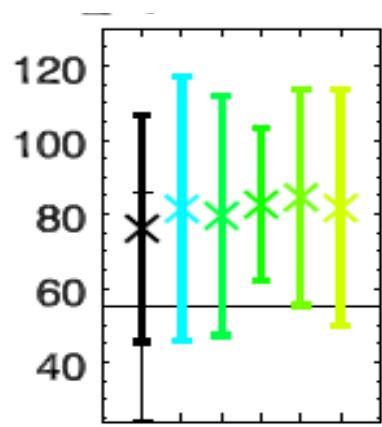


Prior Single-site posterior
Multi-site posterior

→ RMSE reduce by
- 30% for NEE
- 15 % for LE

Retrieved parameter values

Maximum photosynthetic capacity

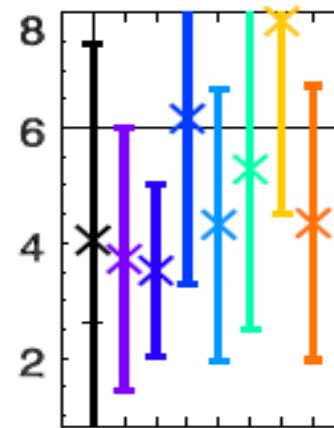


Black: Multi-site

Colors: Single-site

→ *Agreement
betw all sites*

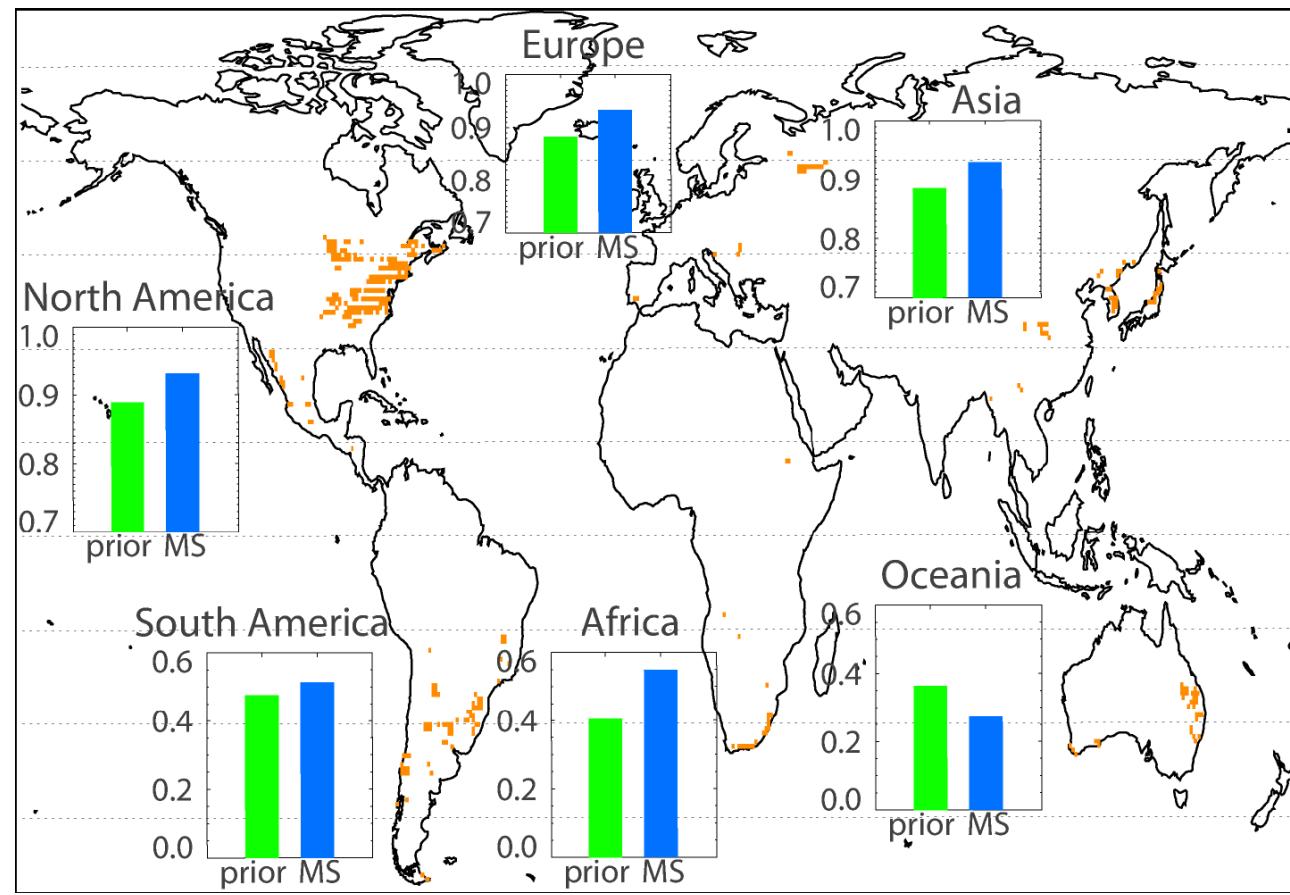
Water stress function (slope)



→ *Need to
improve model !*

Evaluation of the new parameter set

- Use GPP & TER from FluxNet partition as first check
- Global simulation compared with **MODIS NDVI** : increase correlation btw model-fAPAR and NDVI for DBF pixels



Prior model
Posterior model

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

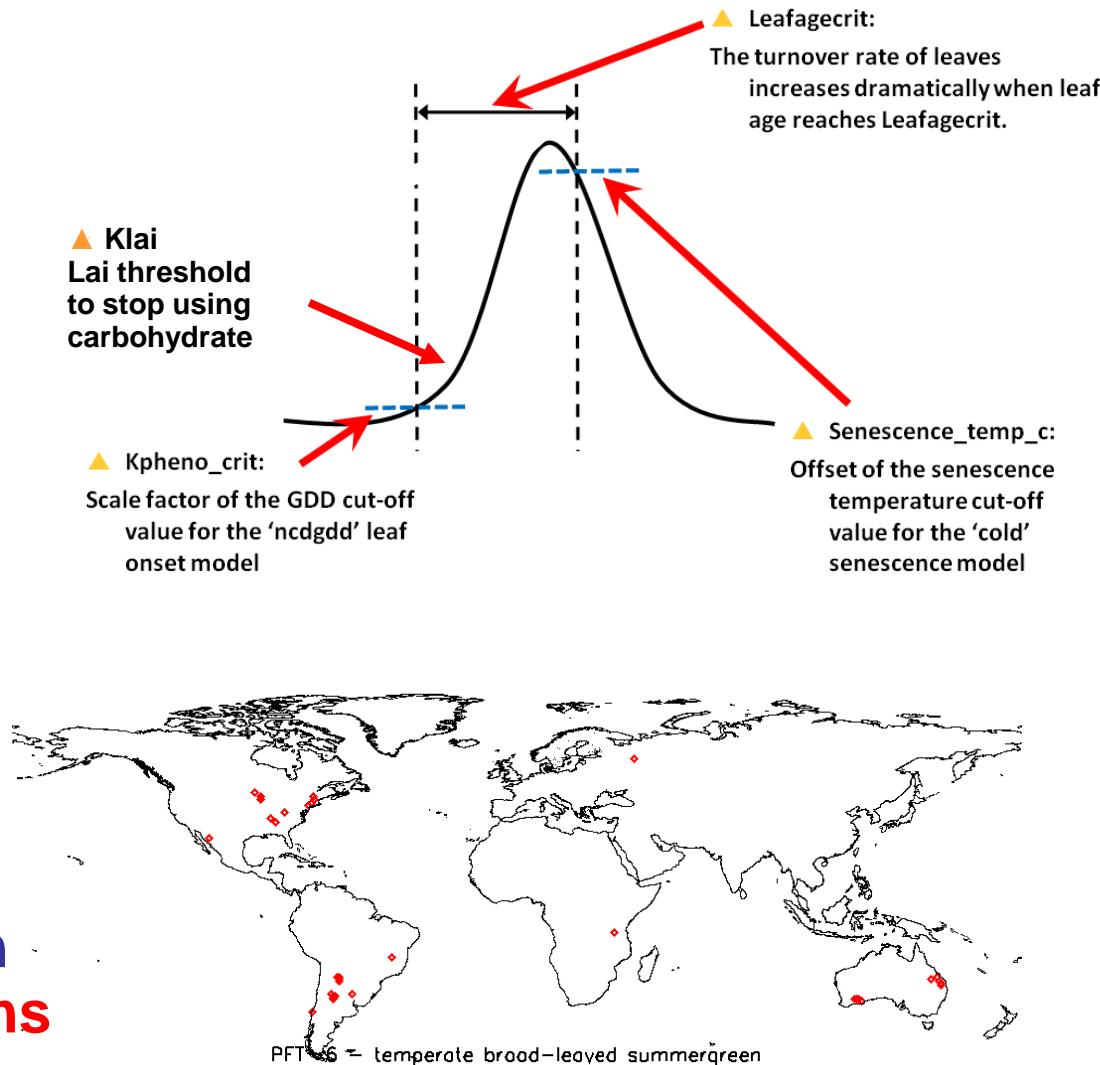
- Used initially by all ecosystem modelers to “manually” adjust their phenology model

- Recent formal studies with a complete statistical approach:
 - Stockli et al. 2008, 2011
 - Knorr et al. 2010

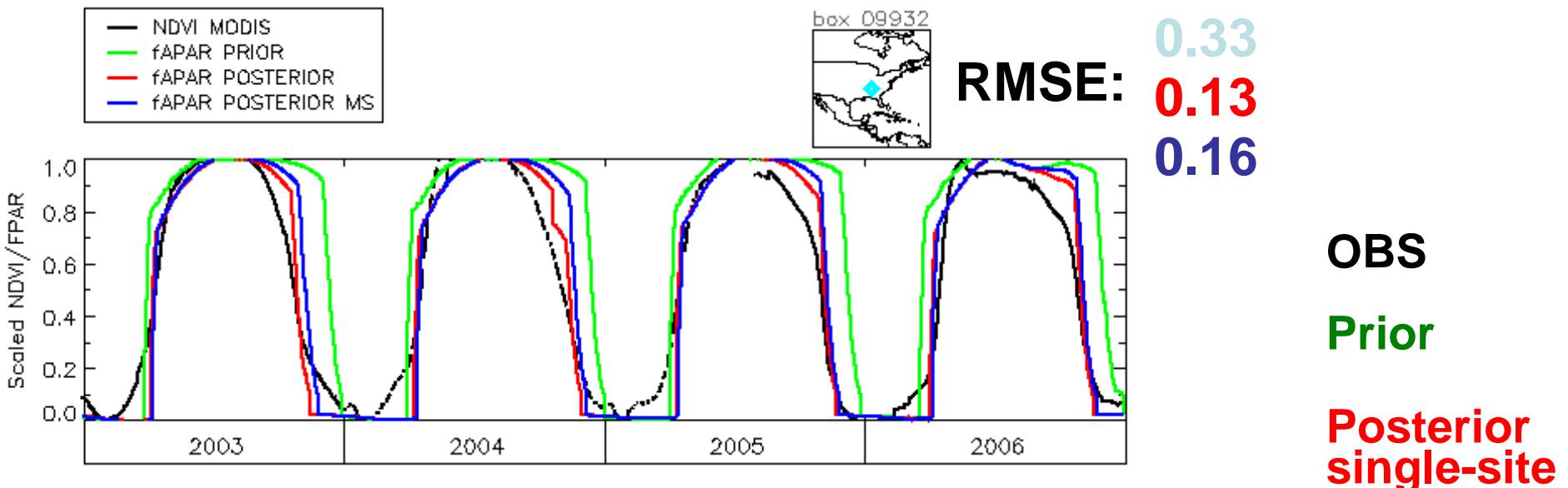
- Optimize 4 parameters using normalised NDVI (2000-2008) to optimize model fAPAR
- ORCHIDEE run with IERA Meteo (0.7°)
- For each PFT use 30 points with $>70\%$ PFT cover

Ex: Temperate deciduous broadleaf forest

1 multi-sites optimization
 & 30 single-site optimizations



PFT : 'temperate broad-leaved summergreen'



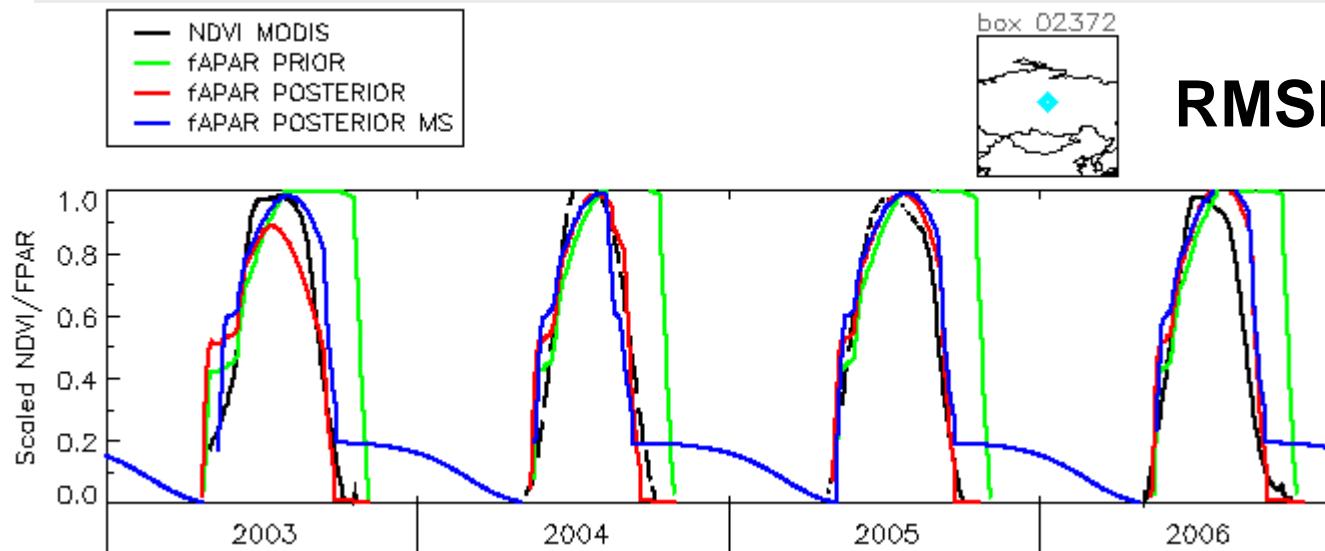
OBS

Prior

Posterior
single-site

Posterior
Multi-site

PFT : 'boreal needleleaf summergreen'



Evaluation of the new model parameters

- ▲ New **ORCHIDEE global simulation** with optimized parameters for 4 PFTs out of 12

- ▲ Global correlations between satellite NDVI and modeled fAPAR time-series:
→ **significant improvement..**

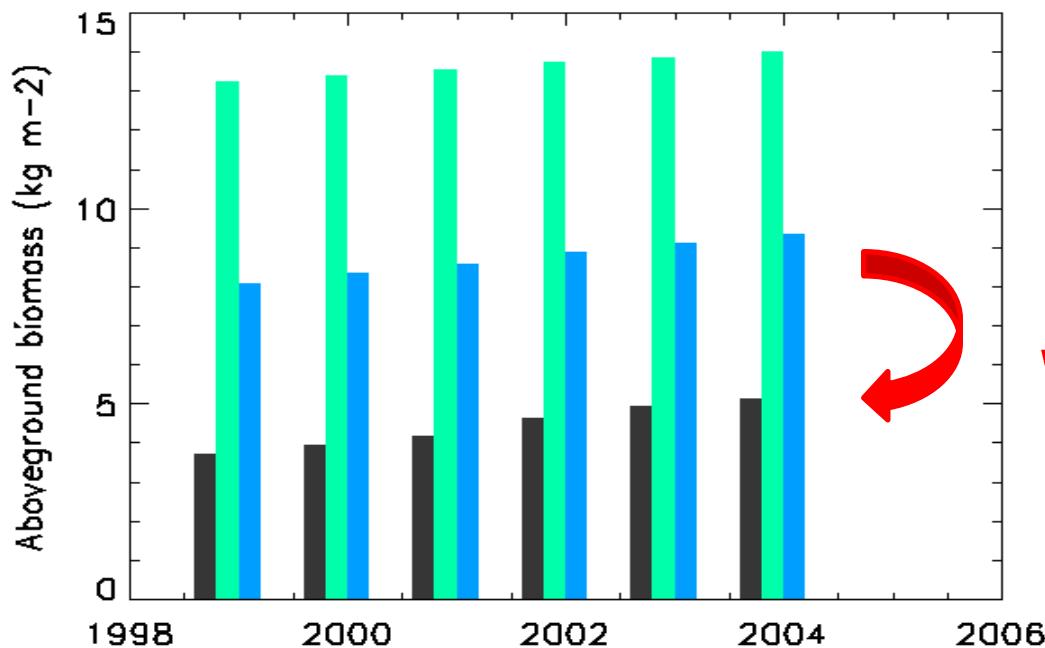
| Mean correlation value | prior | posterior |
|---|-------|-----------|
| PFT 6: temperate broad-leaved summergreen | 0.70 | 0.73 |
| PFT 8: boreal broad-leaved summergreen | 0.72 | 0.86 |
| PFT 9: boreal needleleaf summergreen | 0.39 | 0.89 |
| PFT 10: C3 grass | 0.46 | 0.56 |

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

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

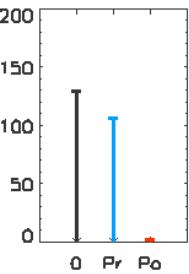
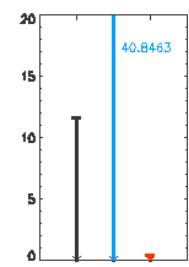
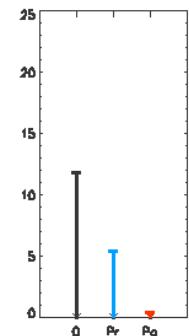
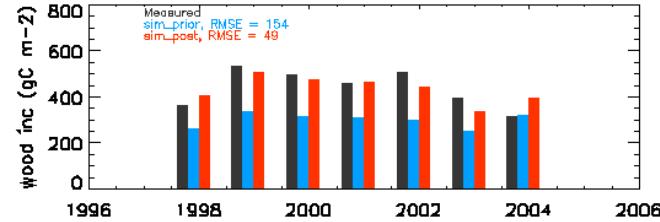
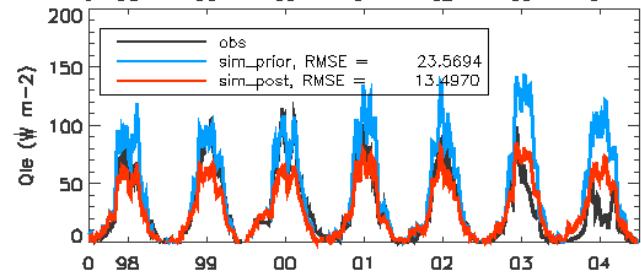
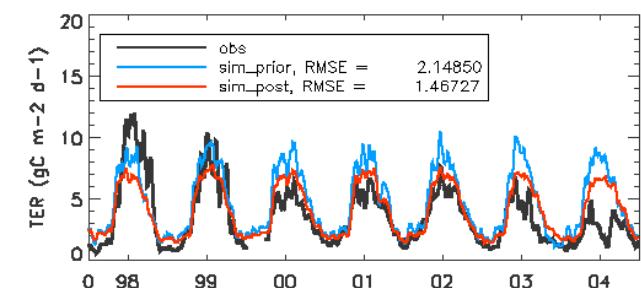
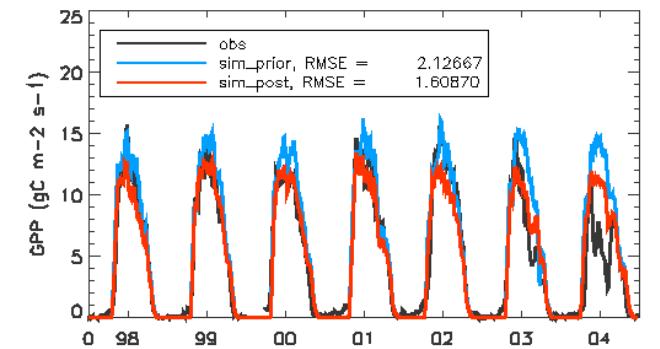
Above ground biomass (Hesse site) : Prior model output



Wrong input
and / or
Wrong mortality

Measurement
Model: Steady state
Model: Realistic age

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



FluxNet data (daily) & Yearly Biomass increments

(25 flux related params)

Measurement

Prior model

GPP_{RMSE} = 2.1

TER_{RMSE} = 2.1

Qle_{RMSE} = 24

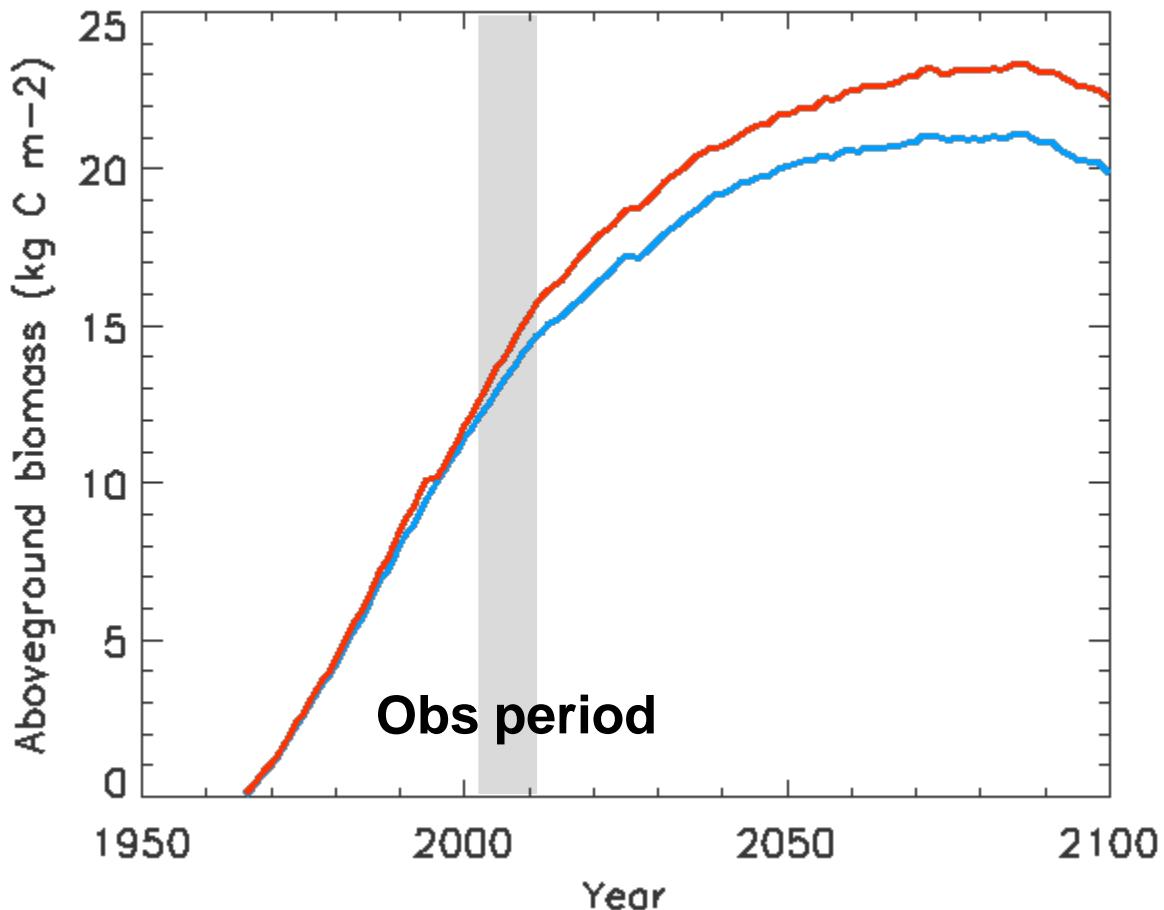
Posterior model

GPP_{RMSE} = 1.6

TER_{RMSE} = 1.4

Qle_{RMSE} = 13

Above ground biomass (Hesse site)

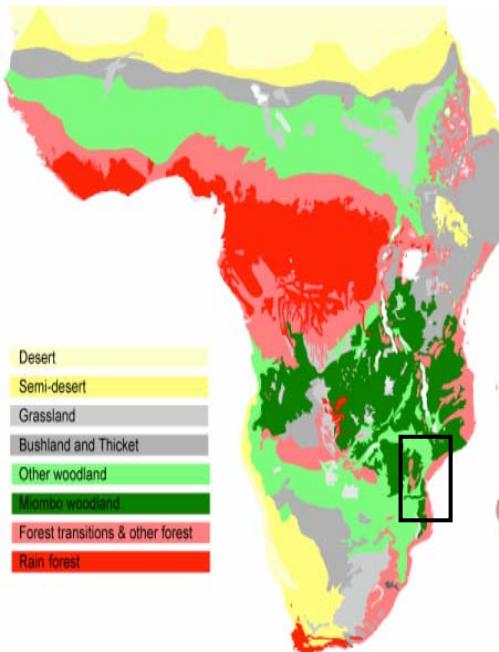


Default
parameter
values

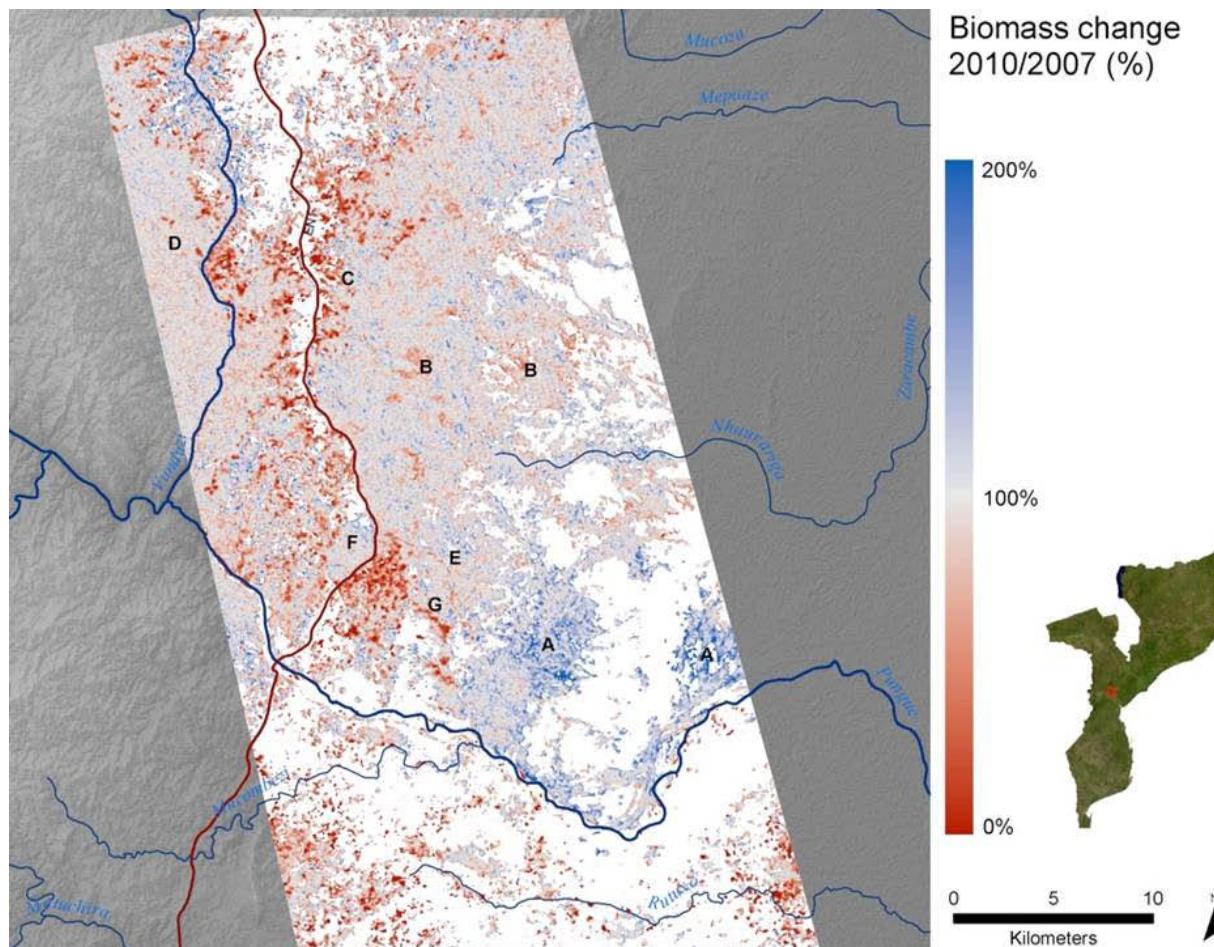
Optimized
parameter
values

Using radar retrievals of forest biomass data Over tropical woodlands

Biome demography
is critical for
African woodlands



Biomass change
from 2007 to 2010

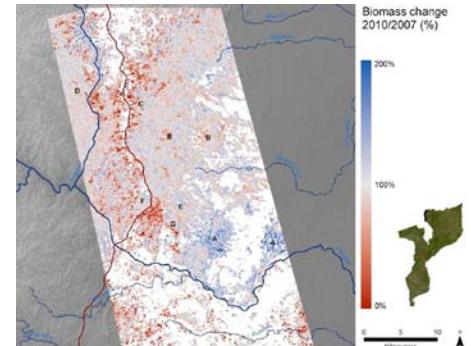


Courtesy of
Mathew Williams et al.

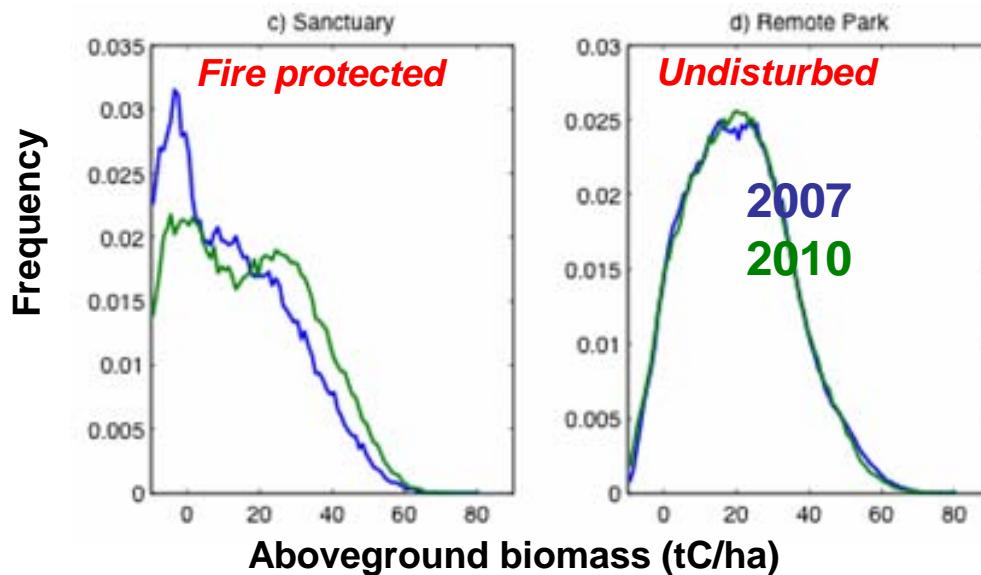
Using radar retrievals of forest biomass data Over tropical woodlands



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



Biomass change (ΔC_w) is determined by growth (NPP), tree lifespan (t_w) and by the probability (P) and intensity (F) of disturbance



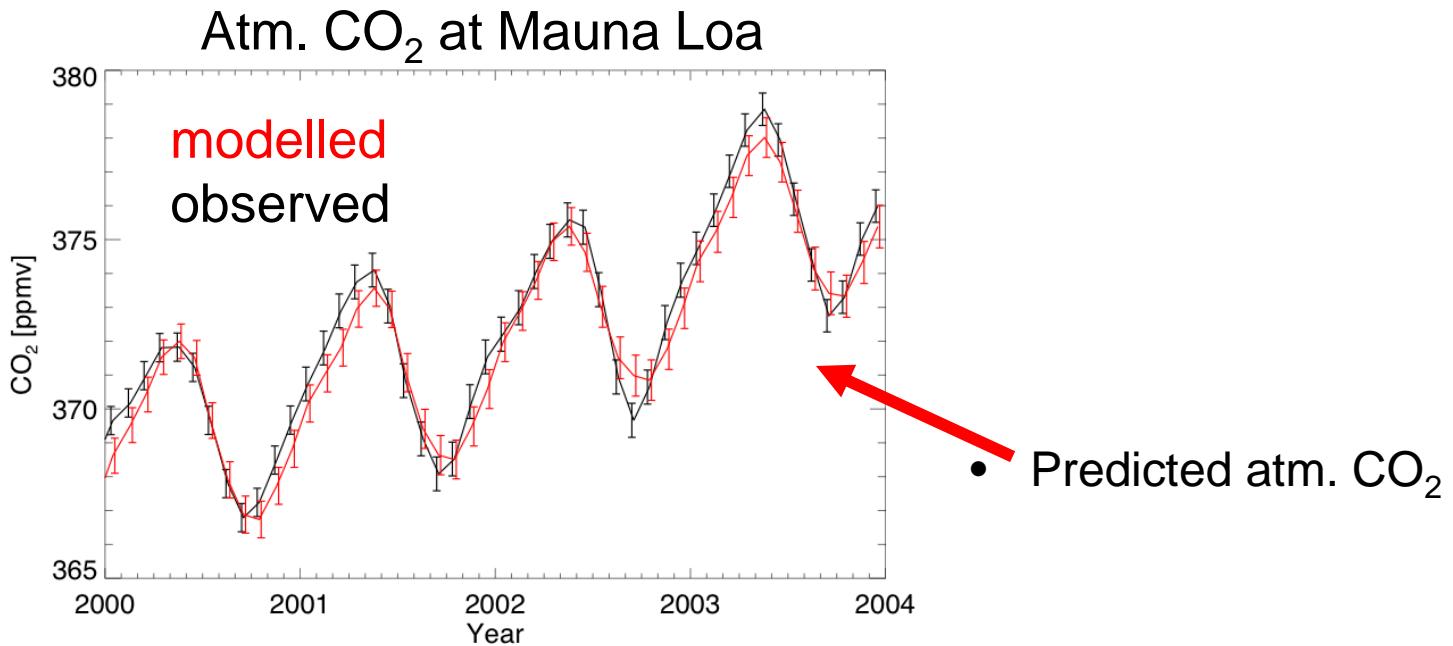
→ Assimilation scheme
to determine parameters
 P and F

*Courtesy of
Mathew Williams et al.*

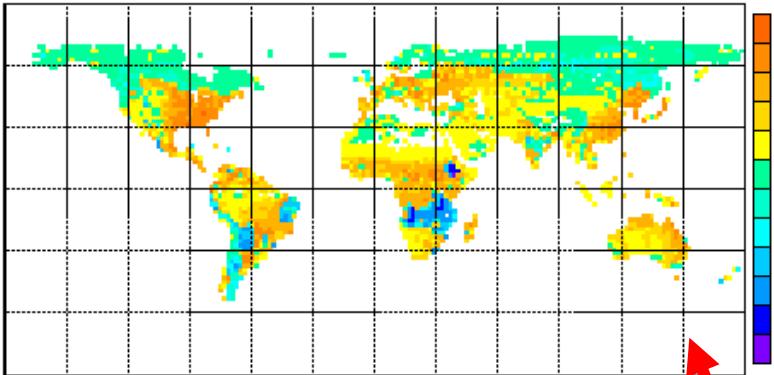
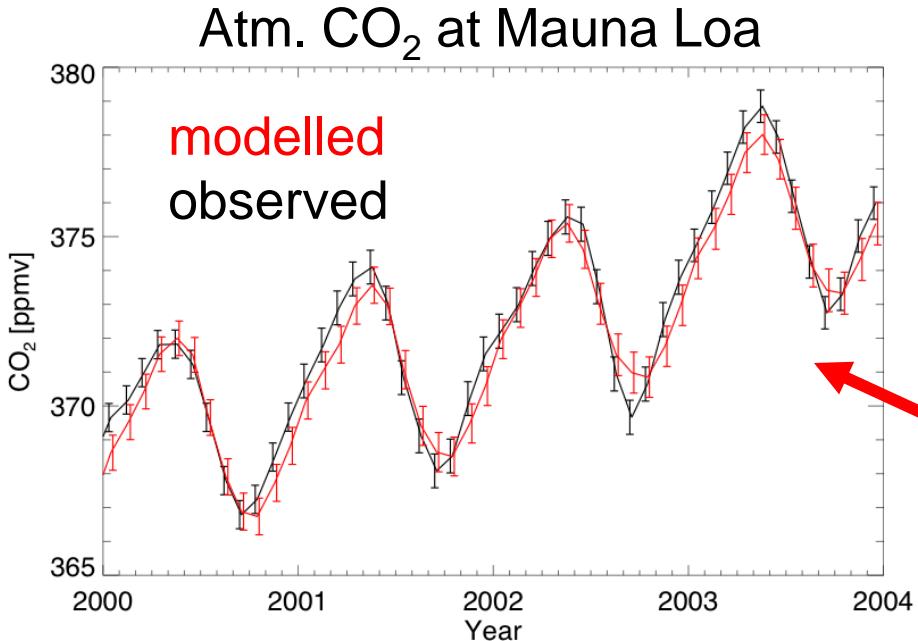
Multi-data Join Assimilation

- Very few published studies so far...
- Results with BETHY ecosystem model
 - + TM2 atmospheric transport model
- ✓ Assimilation of Atm CO₂ and satellite fAPAR
(Rayner P. et al. 2005, Scholze et al. 2007)
- ✓ Variational approach

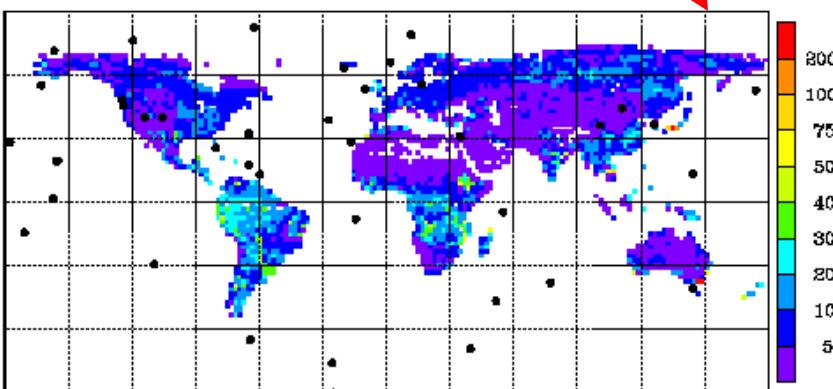
Join Assimilation: “BETHY” CCDAS



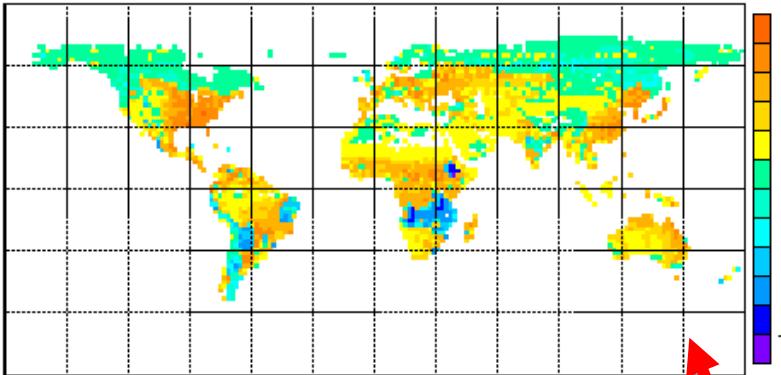
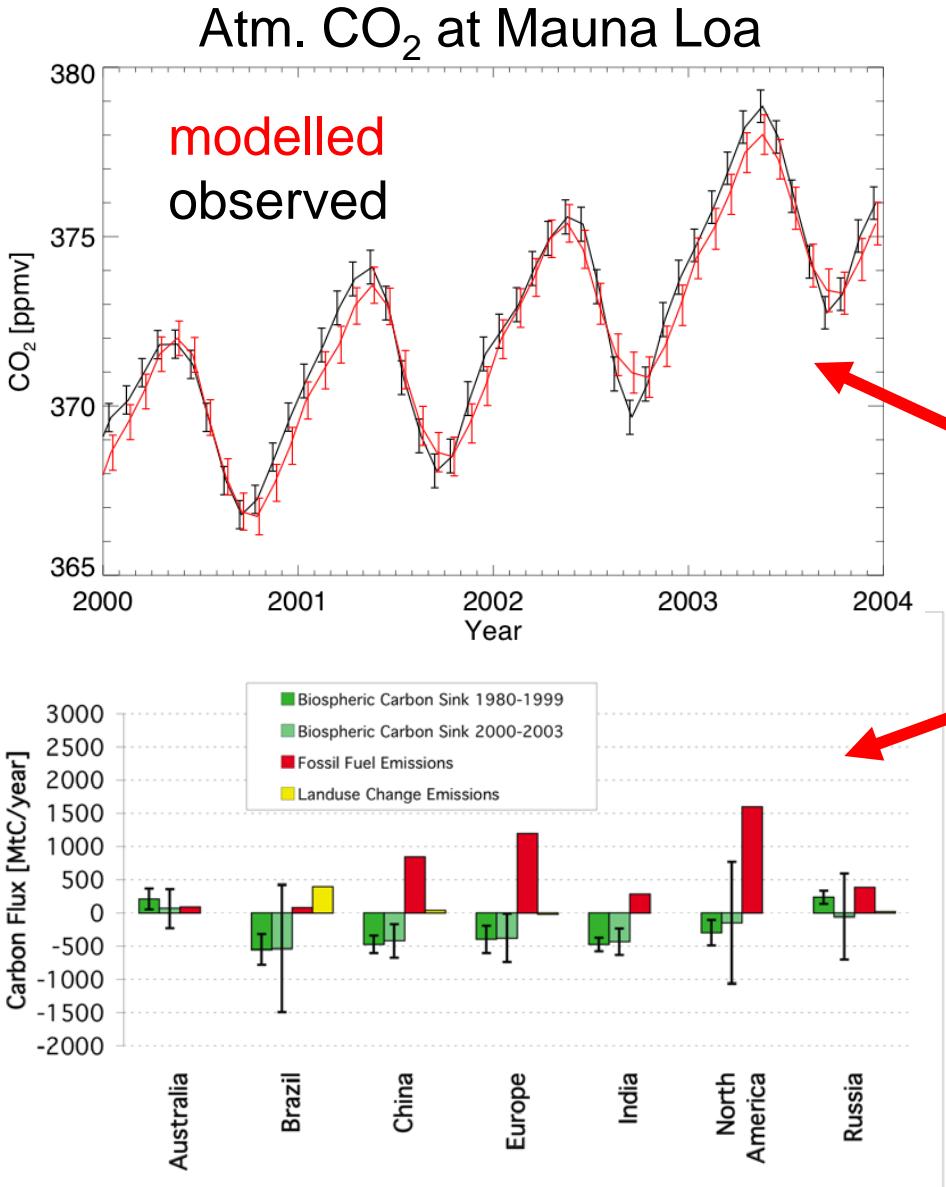
Join Assimilation: “BETHY” CCDAS



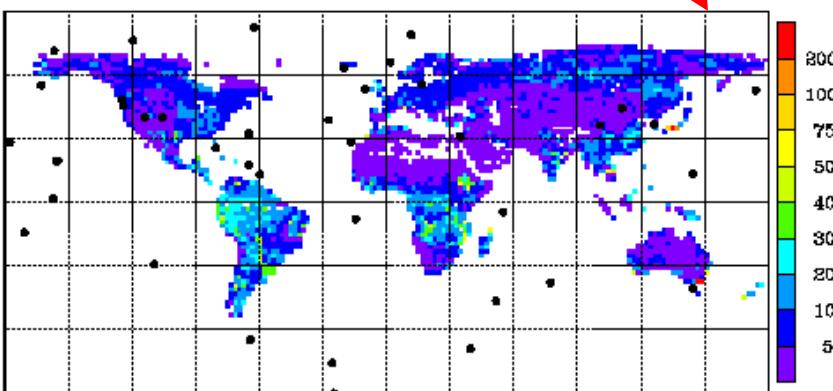
- Predicted atm. CO₂
- Long term mean fluxes to atmosphere (gC/m²/year) and uncertainties



Join Assimilation: “BETHY” CCDAS

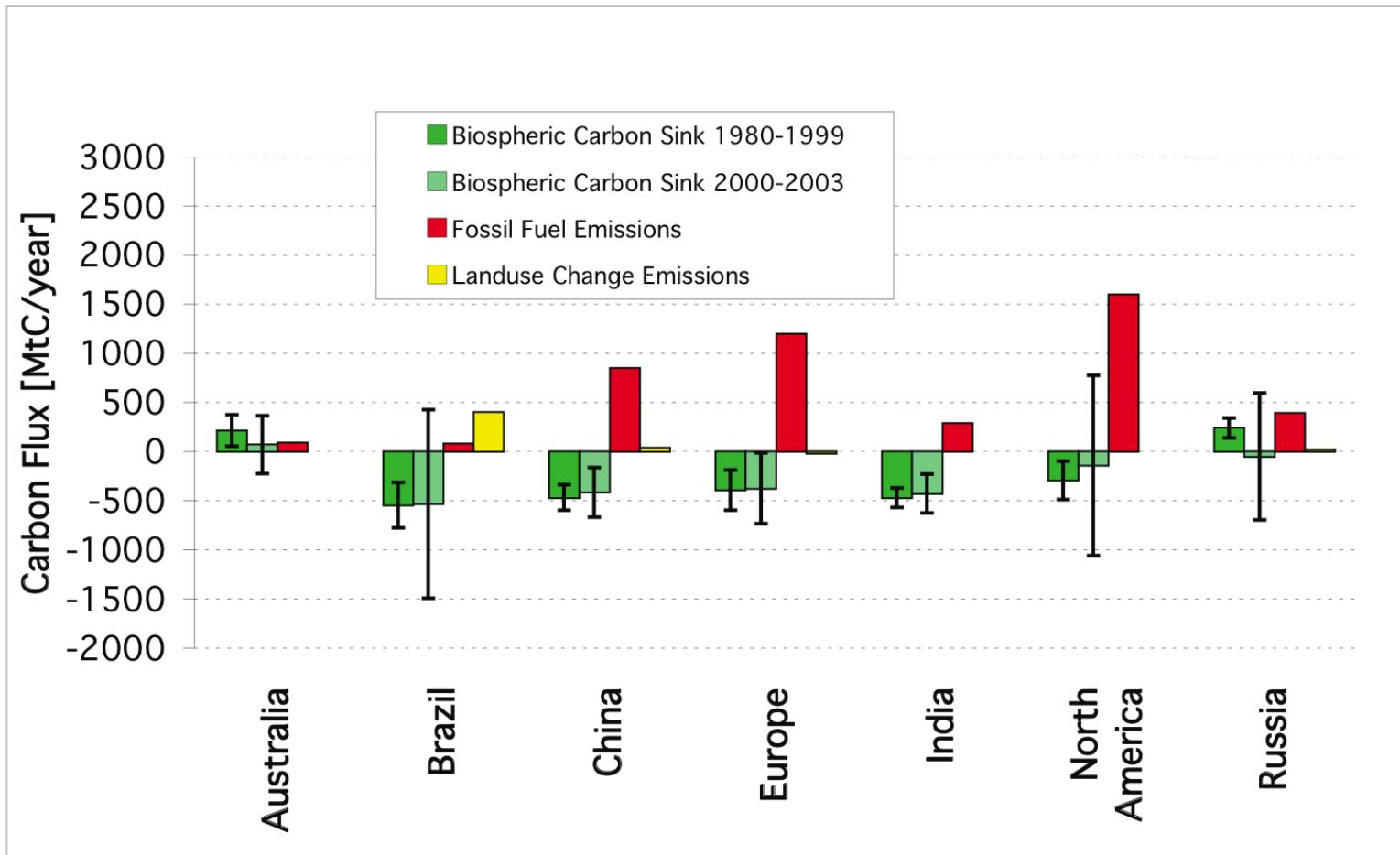


- Predicted atm. CO₂
- Long term mean fluxes to atmosphere (gC/m²/year) and uncertainties
- Regional means diag./prog.



Join Assimilation: “BETHY” CCDAS

- Regional means diagnostics



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
- Ongoing large community effort :
 - GeoCarbon EU-project (5 land CCDAS)
 - Existing inter-comparison of Model-Data fusion exercise

- 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-linearities might complicate the parameter optimization
- Need to :
 - keep independent data for model output validation
 - Keep classical Atmospheric inversion

How to account for soil complexity?

Current soil models may not capture the processes leading to global change response

- Litter quality and OM quality determines pools, τ
- No microbial functionality
- No priming
- Missing
 - Organo-mineral interactions
 - Physical structure and aggregation
- Parameterized for surface soil

