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 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

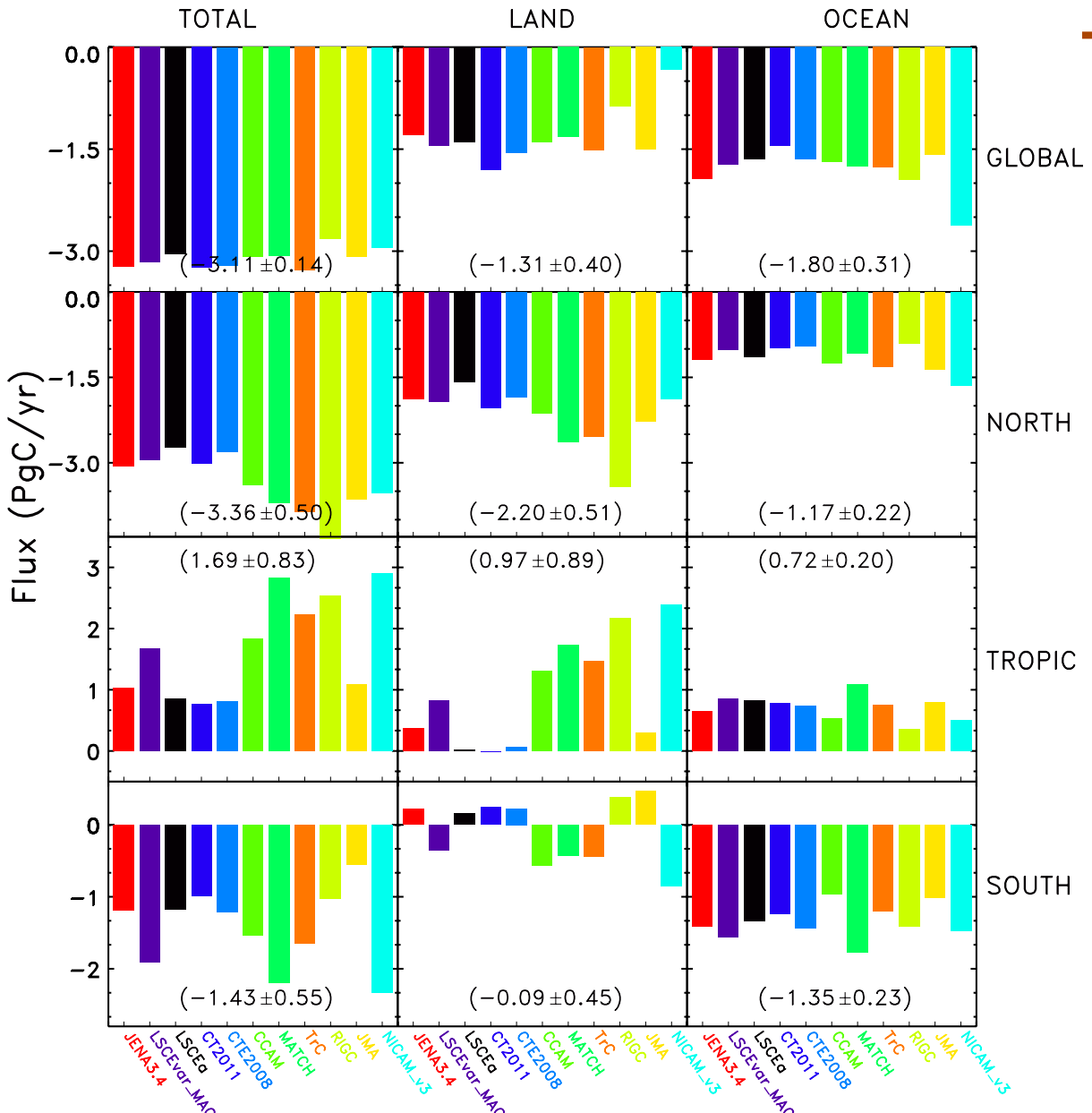
Atmospheric CO2 inversions….

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

Atmospheric inversions

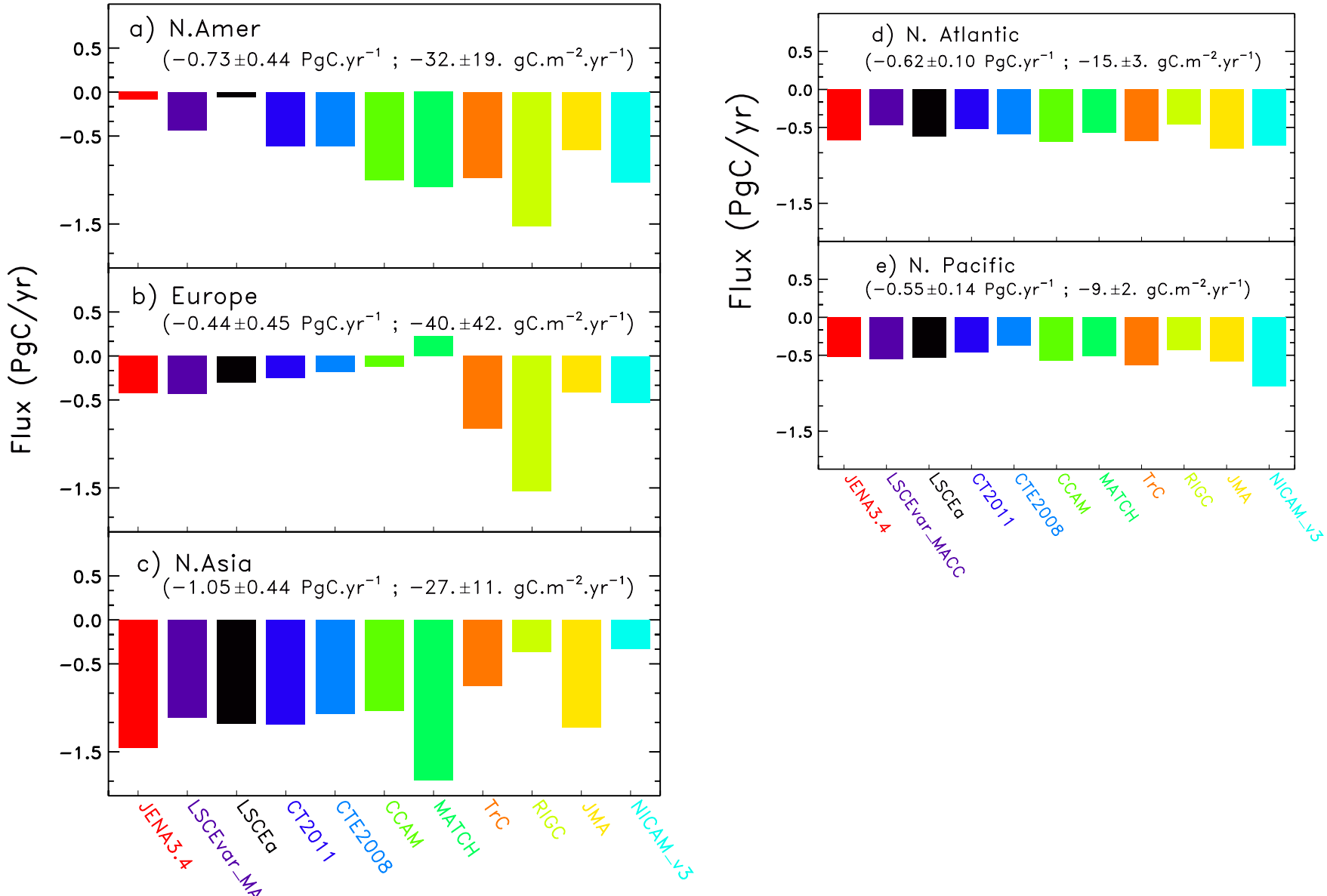
	Inverse System	No regions	Contact	Time Period	Obs	observing stations	IAV
LSCEa	Lsce_an_v2.1	Grid-cell (96x72)	Philippe Peylin	1996-2004	MM	76	Yes
LSCEv	Lsce_var_v1.0	Grid-cell (96x72)	Frederic Chevallier	1988-2008	Raw	128	Yes
CCAM	C13_CCAM_LAW	146	Rachel Law	1992-2008	MM	73 CO_2, 7 C13	No
MATCH	C13_MATCH_Rayner	116	Peter Rayner	1992-2008	MM	73 CO_2, 7 C13	No
CTrUS	Carbontracker_US	156	Andy Jacobson Wouter Peters	2000-2008	Raw	94	Yes
CTrEU	Carbontracker_EU	156	Wouter Peters	2000-2008	Raw	117	Yes
JENA	Jena_s96_v3.3	Grid-cell (72x48)	C. Roedenbeck	1996-2008	Raw	53	Yes
RIGC	Rigc_Patra	64	Prabir Patra	1993-2007	MM	74	Yes
JMA	JMA_2010	22	K. Yamada	1985-2008	MM	146	Yes
TrC	TRCOM_mean	22	Kevin Gurney	1995-2008	MM	103	No
NICAM	Nicam_Niwa	40	Yosuke Niwa	1988-2007	MM	94	Yes

Atmospheric inversions: Long term mean

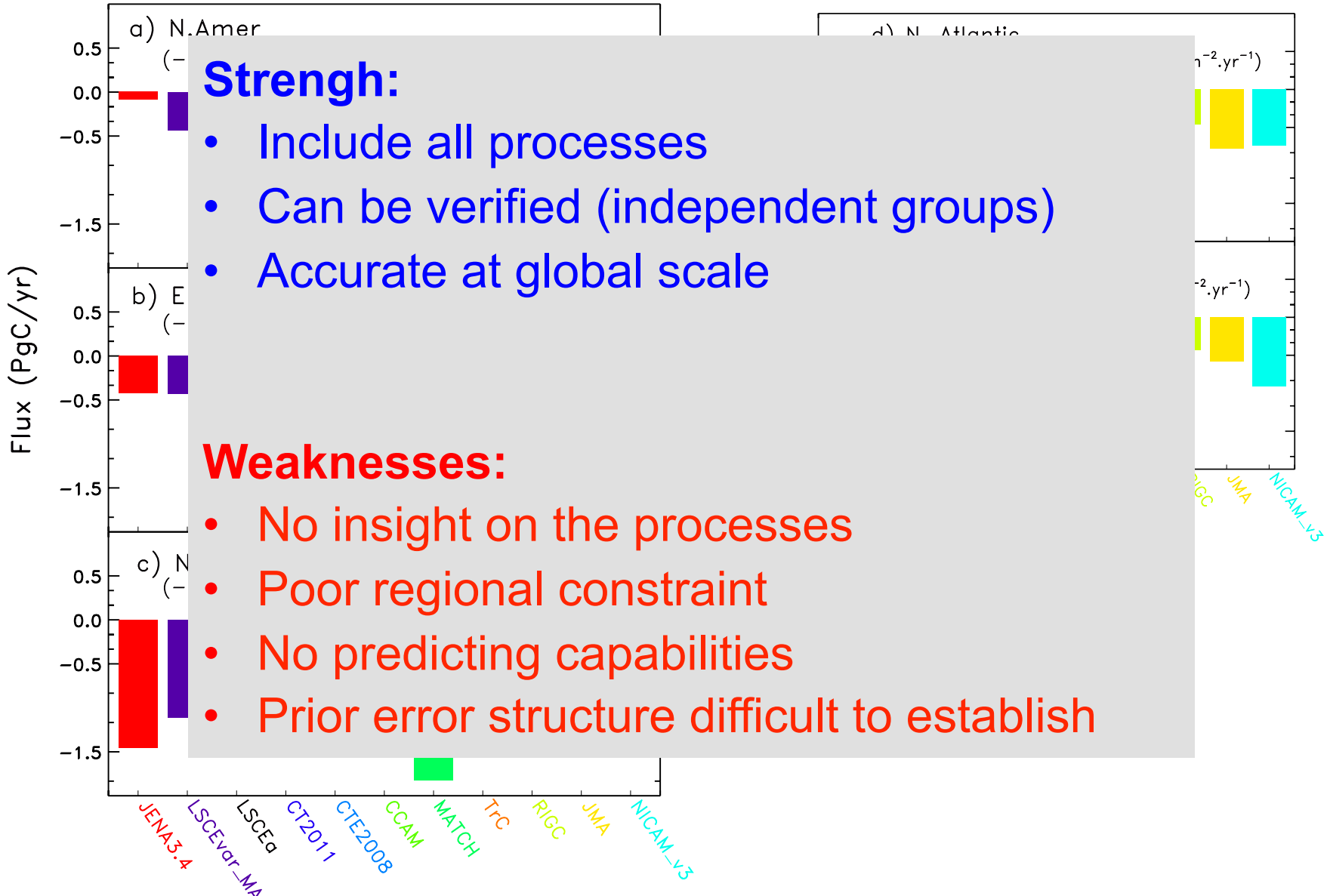


JENA_s96
LSCE_var
LSCE_an
CTrac_US
CTrac_EU
C13CCAM
C13MATCH
TRCOM
RIGC
JMA
NCAM

Atmospheric inversions: Long term mean



Atmospheric inversions: Long term mean



Strengh:

- Include all processes
- Can be verified (independent groups)
- Accurate at global scale

Weaknesses:

- No insight on the processes
- Poor regional constraint
- No predicting capabilities
- Prior error structure difficult to establish

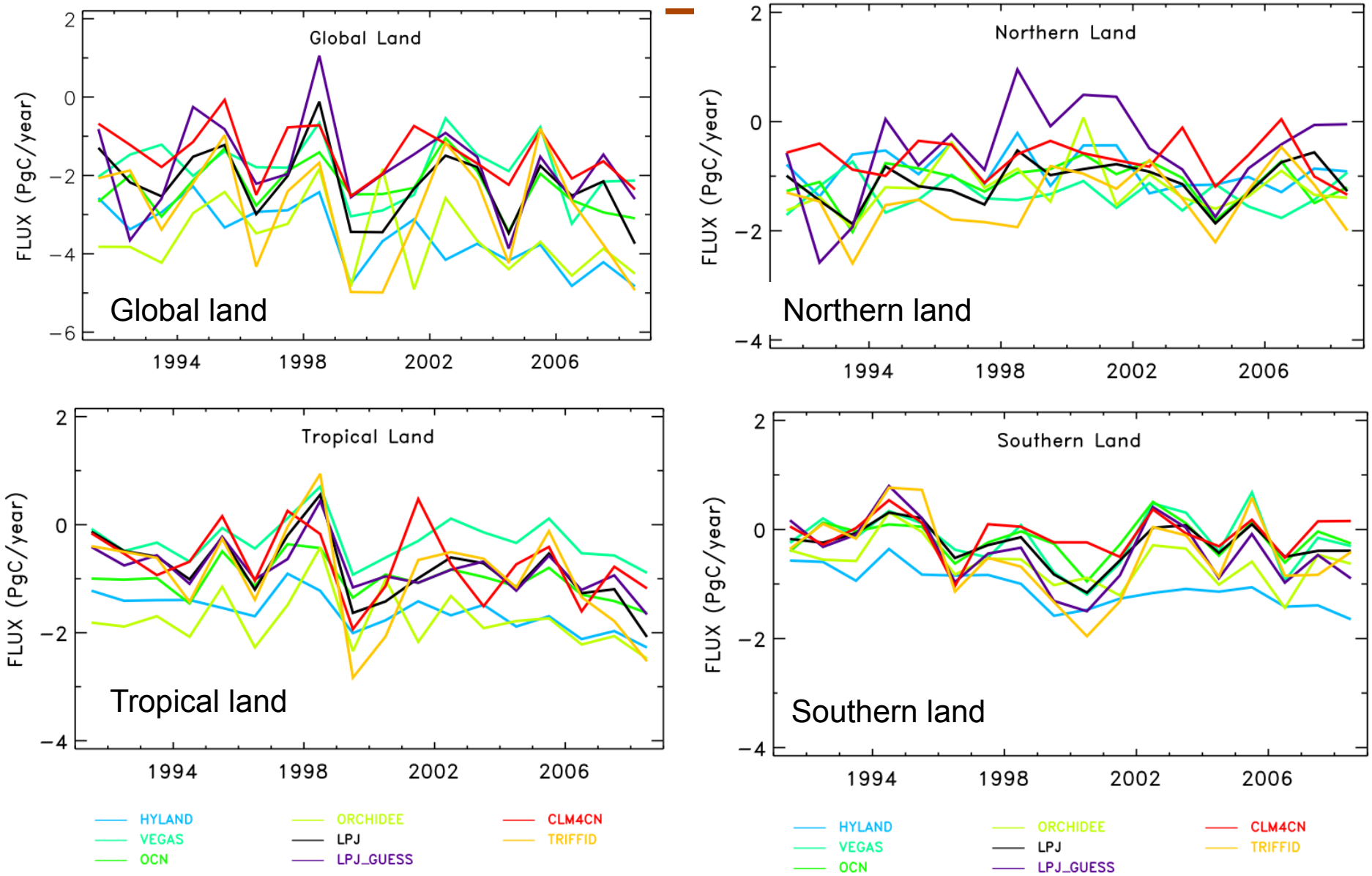
Dynamic global vegetation models

➔ DGVM models used in Trendy intercomparison

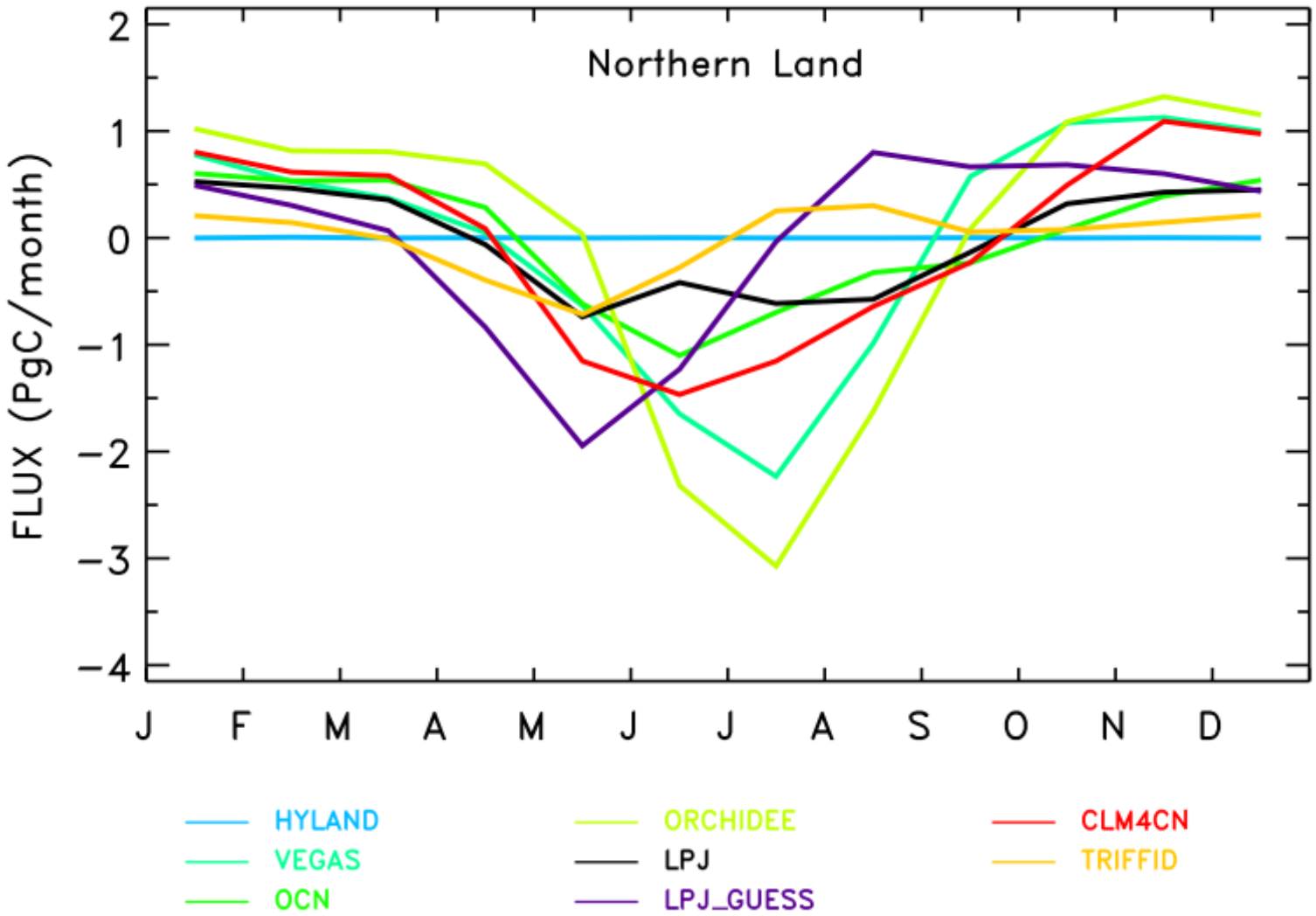
Model Name	Abbreviation	Spatial resolution	Land Surface Model	Full Nitrogen Cycle	River Export Flux	Fire simulation	Source
Community Land Model 4CN	CLM4CN	0.5°×0.5°	Yes	Yes	No	Yes	Olcson et al., 2010; Lawrence et al., 2011
Hyland	HYL	3.75°×2.5°	No	No	No	No	Friend et al., 1997; Levy et al., 2004
Lund-Potsdam-Jena	LPJ	0.5°×0.5°	No	No	No	Yes	Sitch et al., 2003
LPJ-GUESS	LPJ-GUESS	0.5°×0.5°	No	No	No	Yes	Smith et al., 2001
ORCHIDEE-CN	OCN	3.75°×2.5°	Yes	Yes	No	No	Zachle& Friend, 2010; Zachle et al., 2010
ORCHIDEE	ORC	0.5°×0.5°	Yes	No	No	No	Krinner et al., 2005
Sheffield-DGVM	SDGVM	3.75°×2.5°	No	No	Yes	Yes	Woodward et al., 1995
TRIFFID	TRI	3.75°×2.5°	Yes	No	No	No	Cox, 2001
VEGAS	VEGAS	2.5°×2.5°		No		No	Zeng et al., 2005

Sitch et al. GBC–Discussion

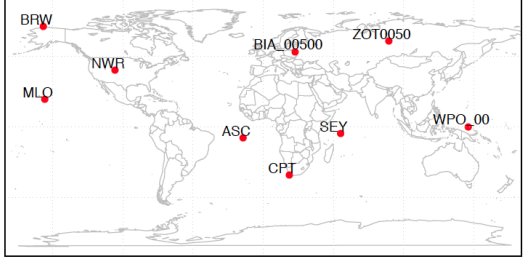
DGVM: Annual flux temporal variations

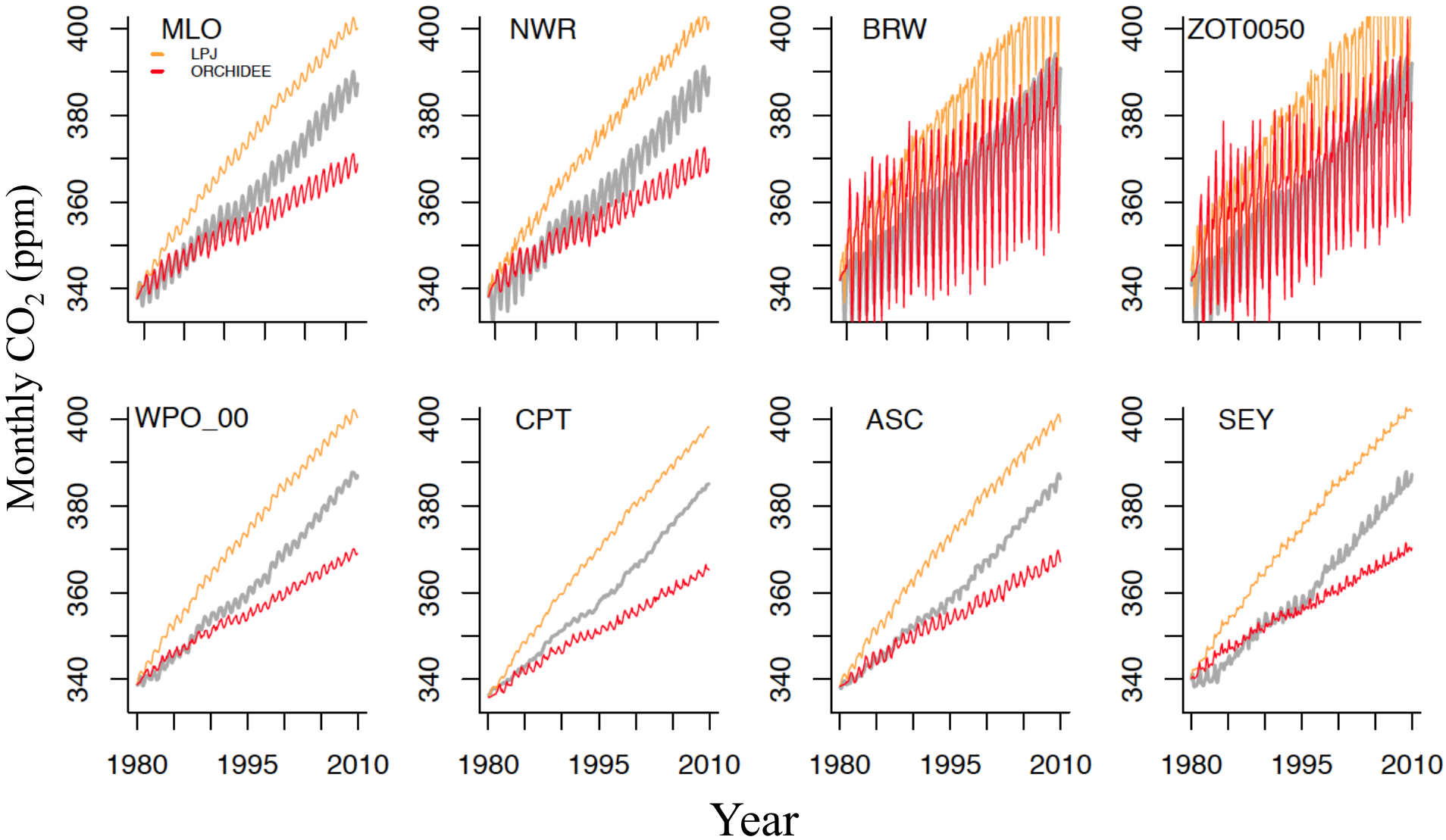


Mean seasonal cycle : Nothern land



S3: CO_2 concentrations (obs. In grey)

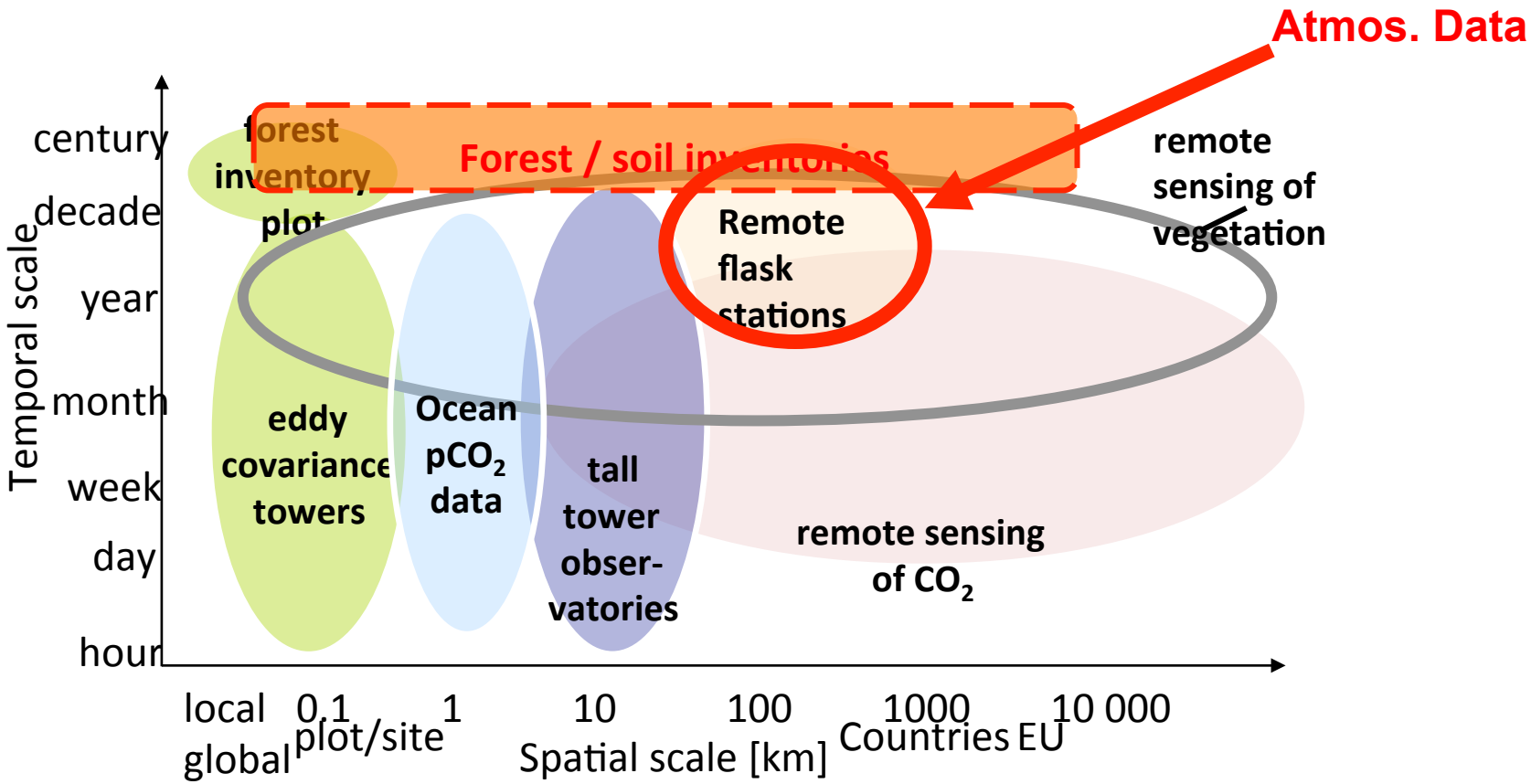




How to move forward ?

Strong Need to :

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

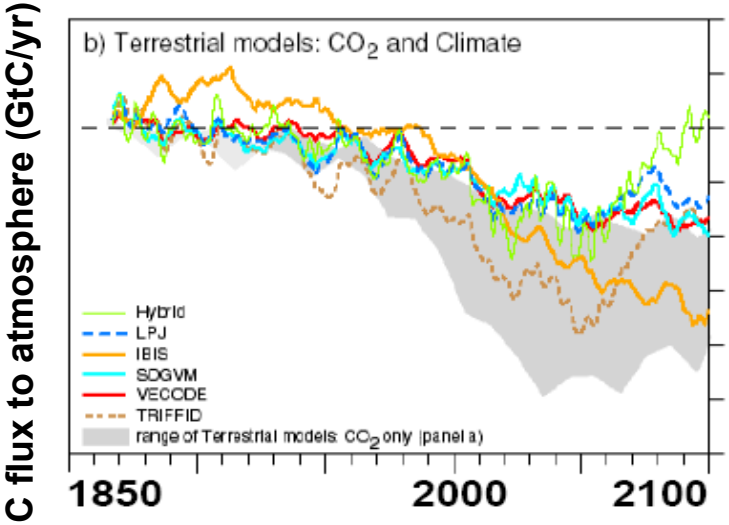


Outline

- 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

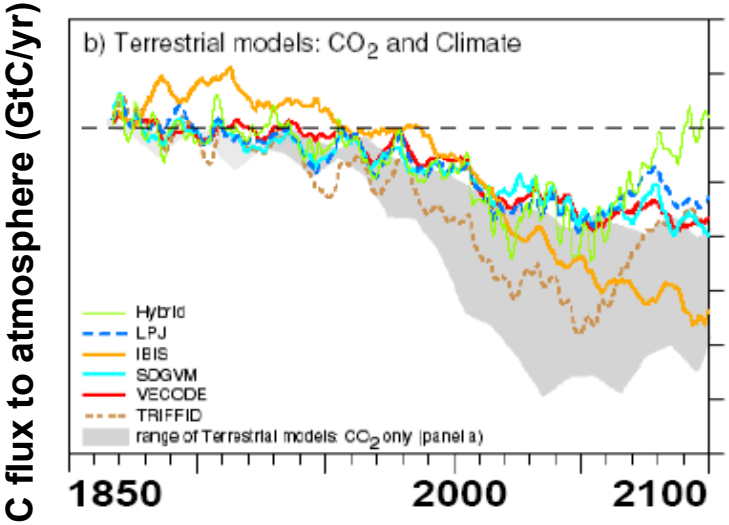
Needs for a Carbon Cycle Data Assimilation System

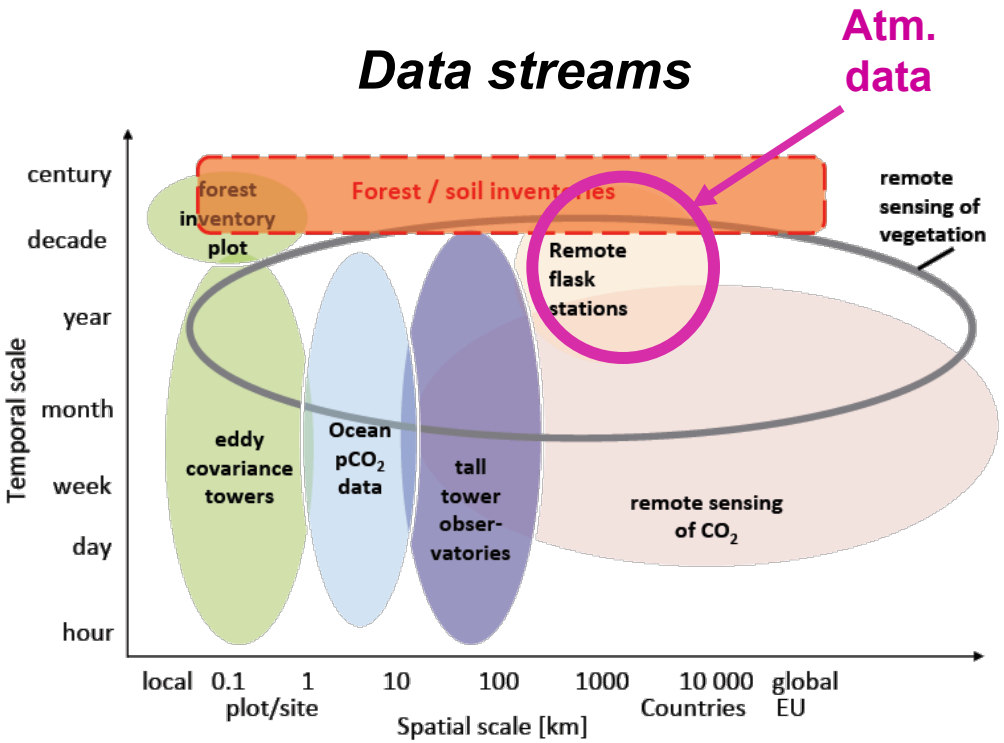
Large uncertainty from land to predict global C-balance (C4MIP)



Needs for a Carbon Cycle Data Assimilation System

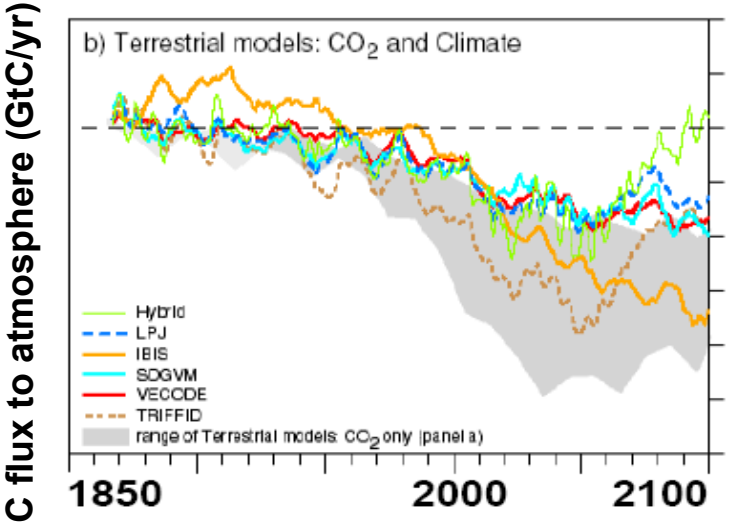
Large uncertainty from land to predict global C-balance (C4MIP)



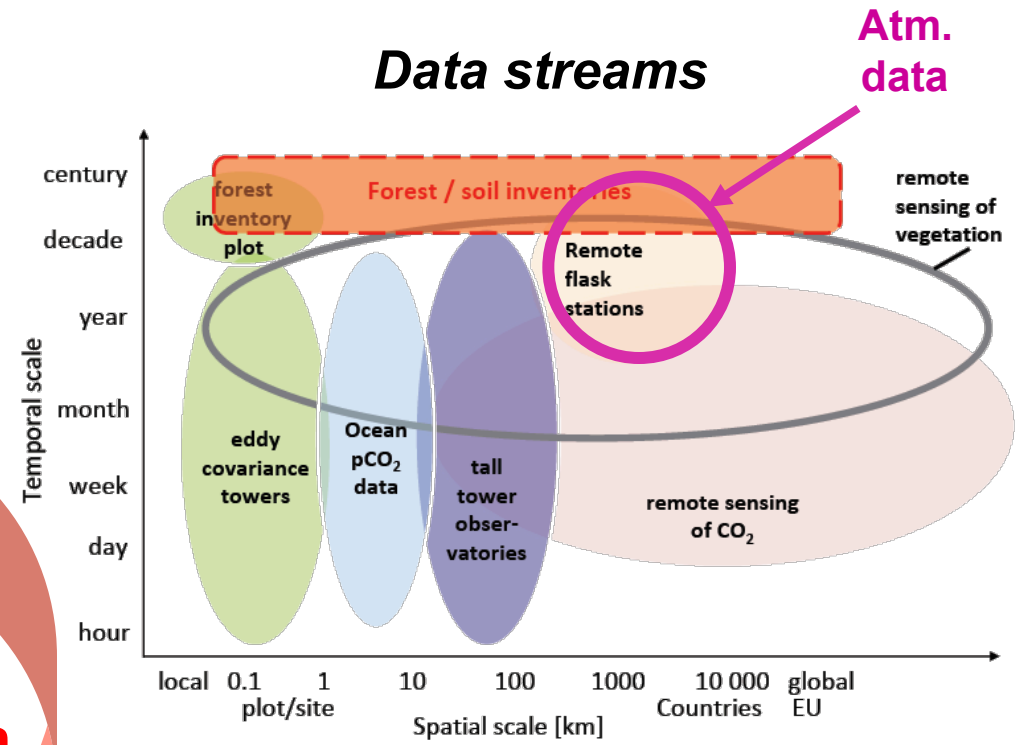


Needs for a Carbon Cycle Data Assimilation System

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



Data streams

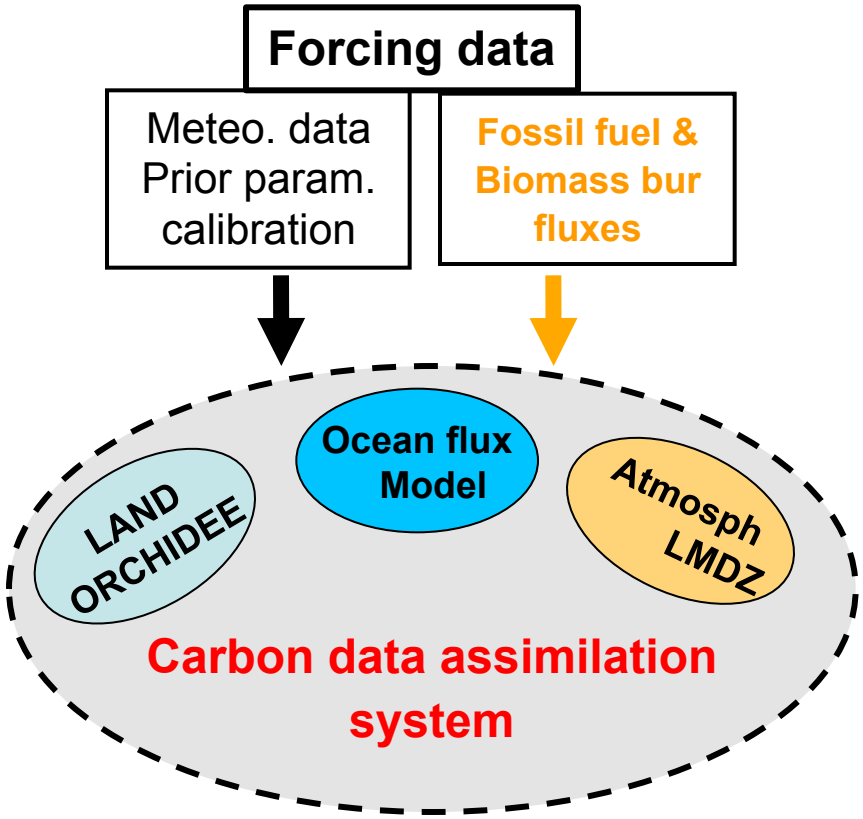


Data Assimilation

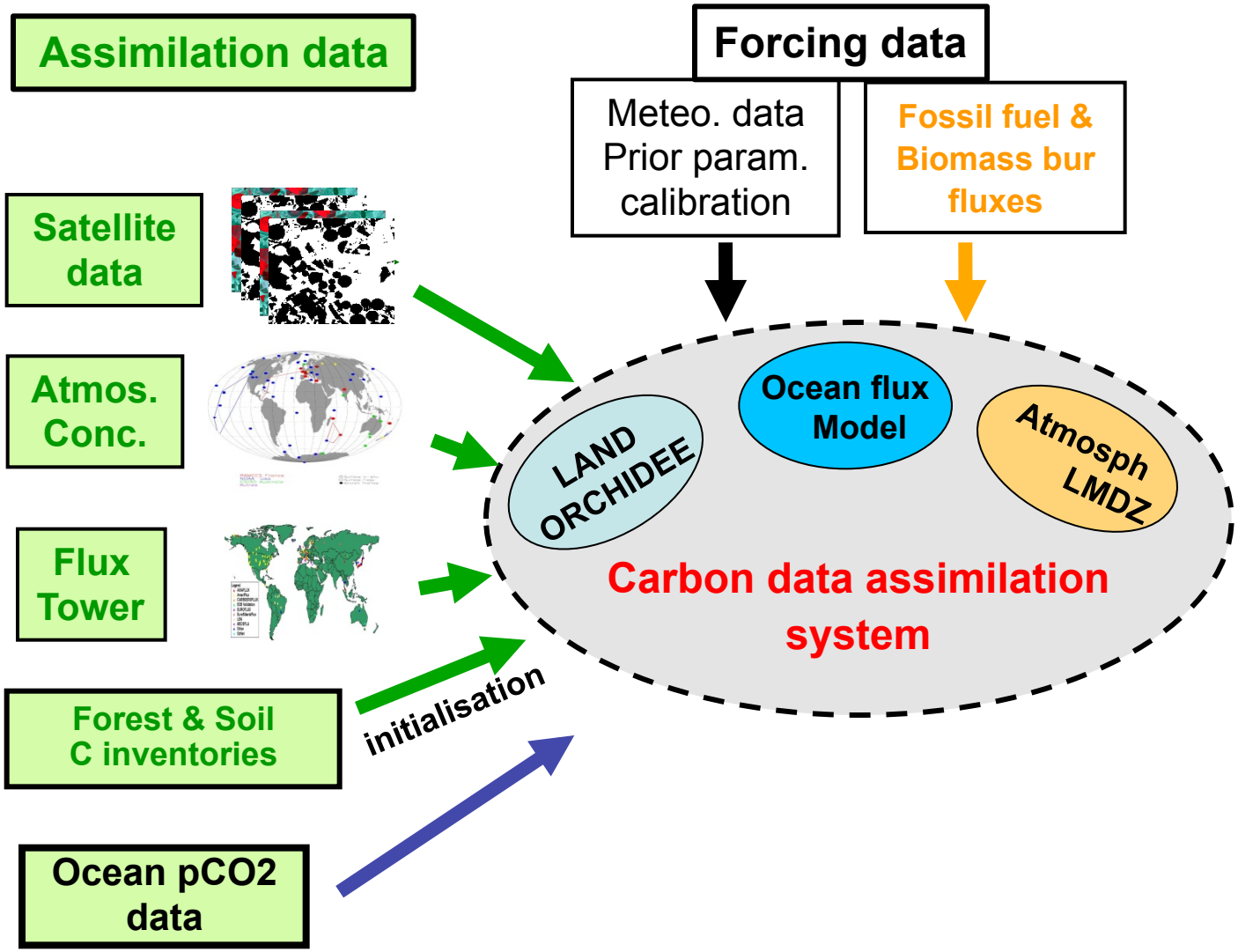
Optimized ecosystem models

➔ reduce the spread ?

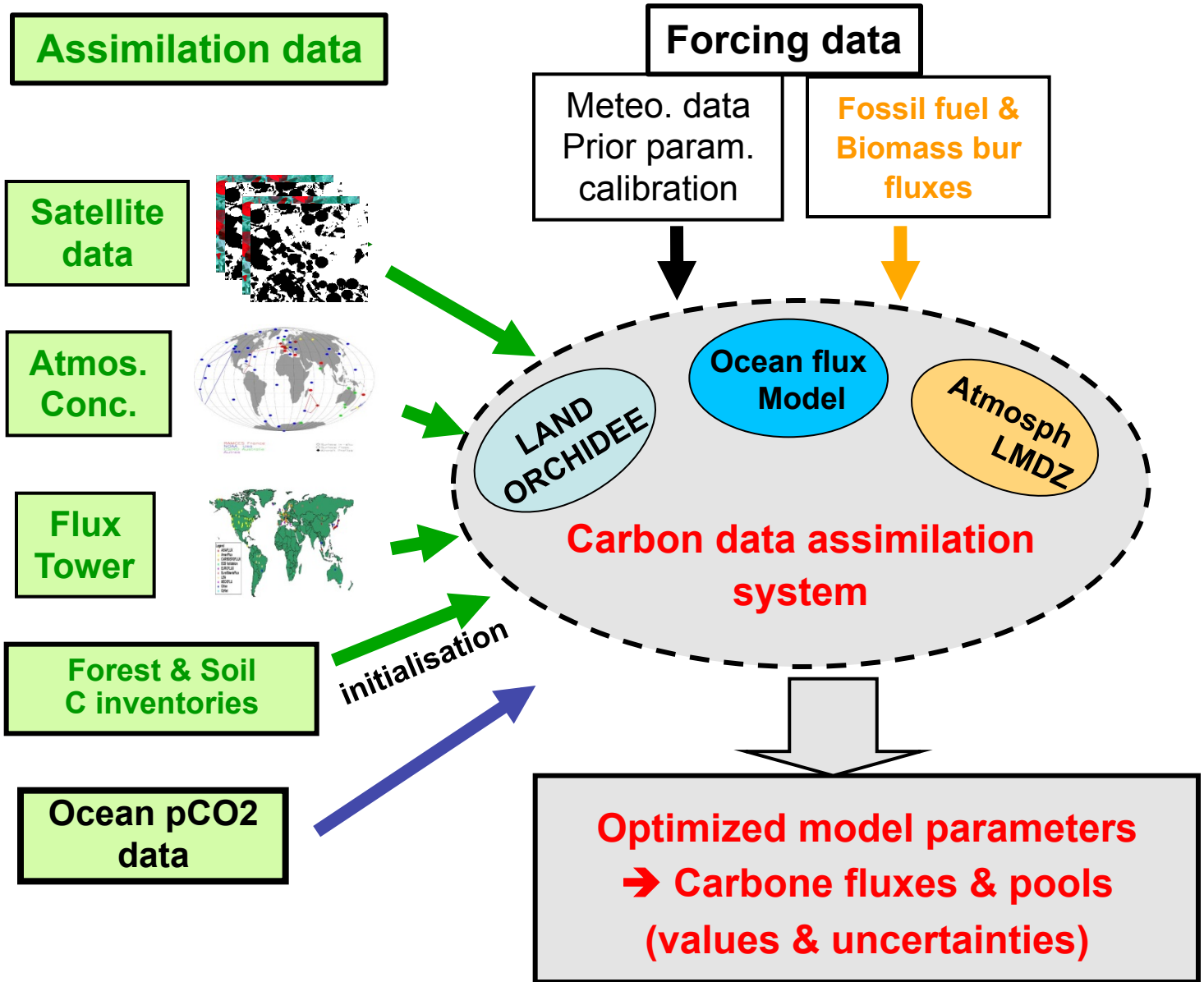
Structure of a global "CCDAS"



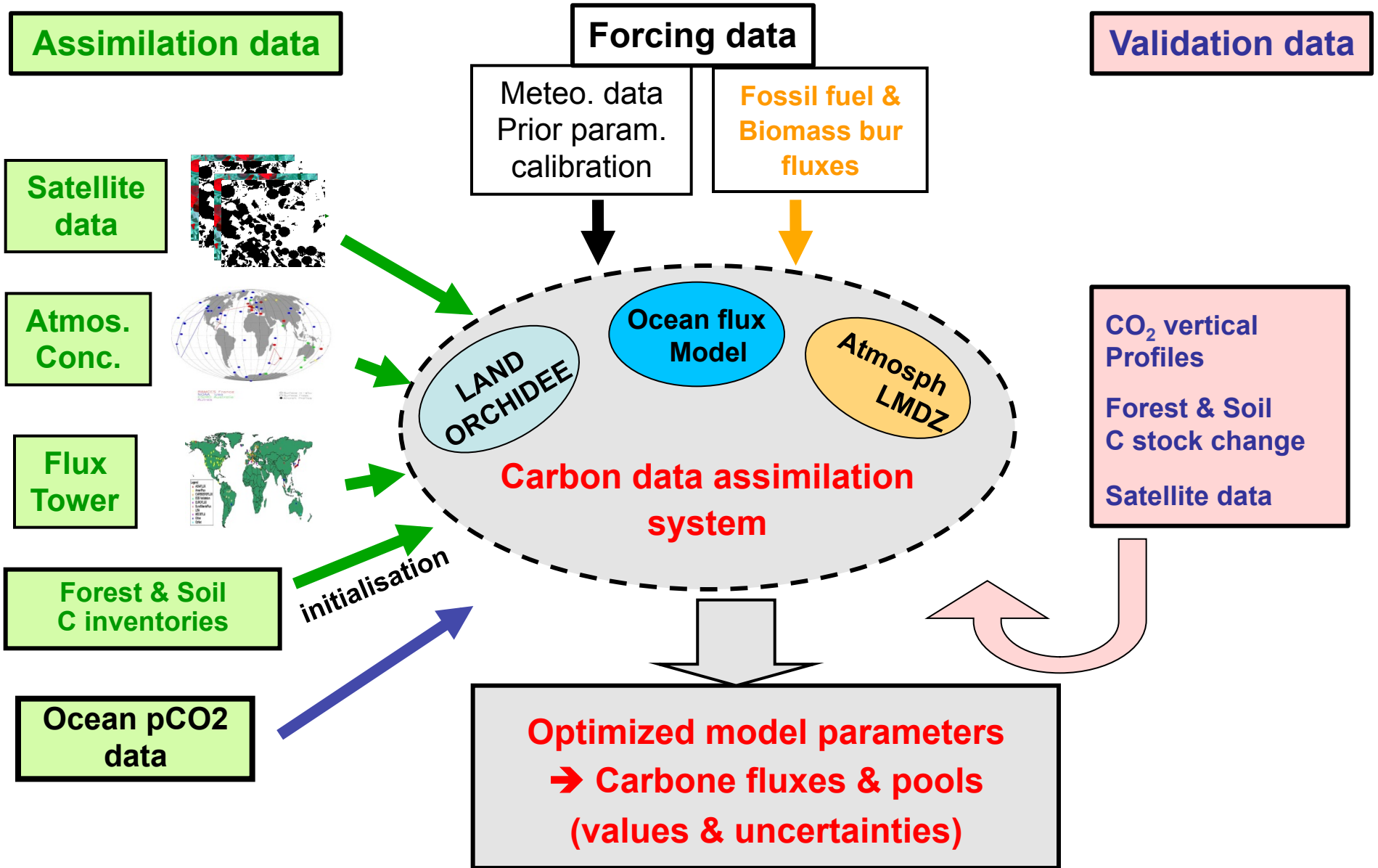
Structure of a global "CCDAS"



Structure of a global "CCDAS"



Structure of a global "CCDAS"



Land CCDAS components

Soil C invent.

Forest C invent.

Flux Towers

Satellite NDVI

Atmos. Conc.

Ecosystem model

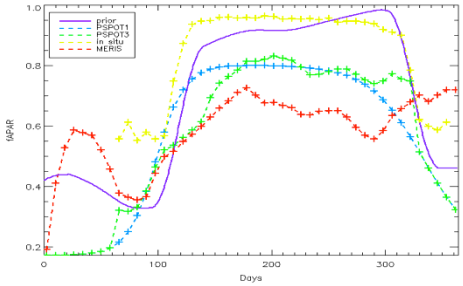
Improve initial soil C pools

Improve allocation & mortality

- Local constraint (small footprint)
- Short term C-flux variations (Photosynthesis, Respiration)

Ex : fAPAR

Optimize Phenology



Transport model

Overall flux scaling

Simultaneous optimization (CCDAS)

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(x) to obtain the mean of p(x|y)

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

x: state vector ;
xb: mean prior value of state vector
y: observation vector ;
H: linear observation operator
B / R: Background / Observation error covariance matrix

Formalism…

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

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

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

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

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

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

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

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

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

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

Parameters versus Flux optimization

	Flux	Parameters
Atm data	• Dependent on record length • Error correlation difficult to assess	• Dependent on ecosystem model structural errors • over-tuning of parameters always possible
FluxNet NEE	• Scale missmatch	• constraint “fast” processes params (1/2h to seasonal)
NDVI	• Not possible	• constraint Phenology params.

➔ **Approach may depend on “scientific objective”**

➔ “Best flux re-analysis” : Atm. data to constrain fluxes

➔ “Best model for future simul.” : Atm. data to constraint parameters

Stepwise vs Simultaneous Assimilation

Simultaneous

Stepwise

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

- All data streams assimilated consistently at once
- Consistency of the cost function
- "Fit to all" data stream ensured

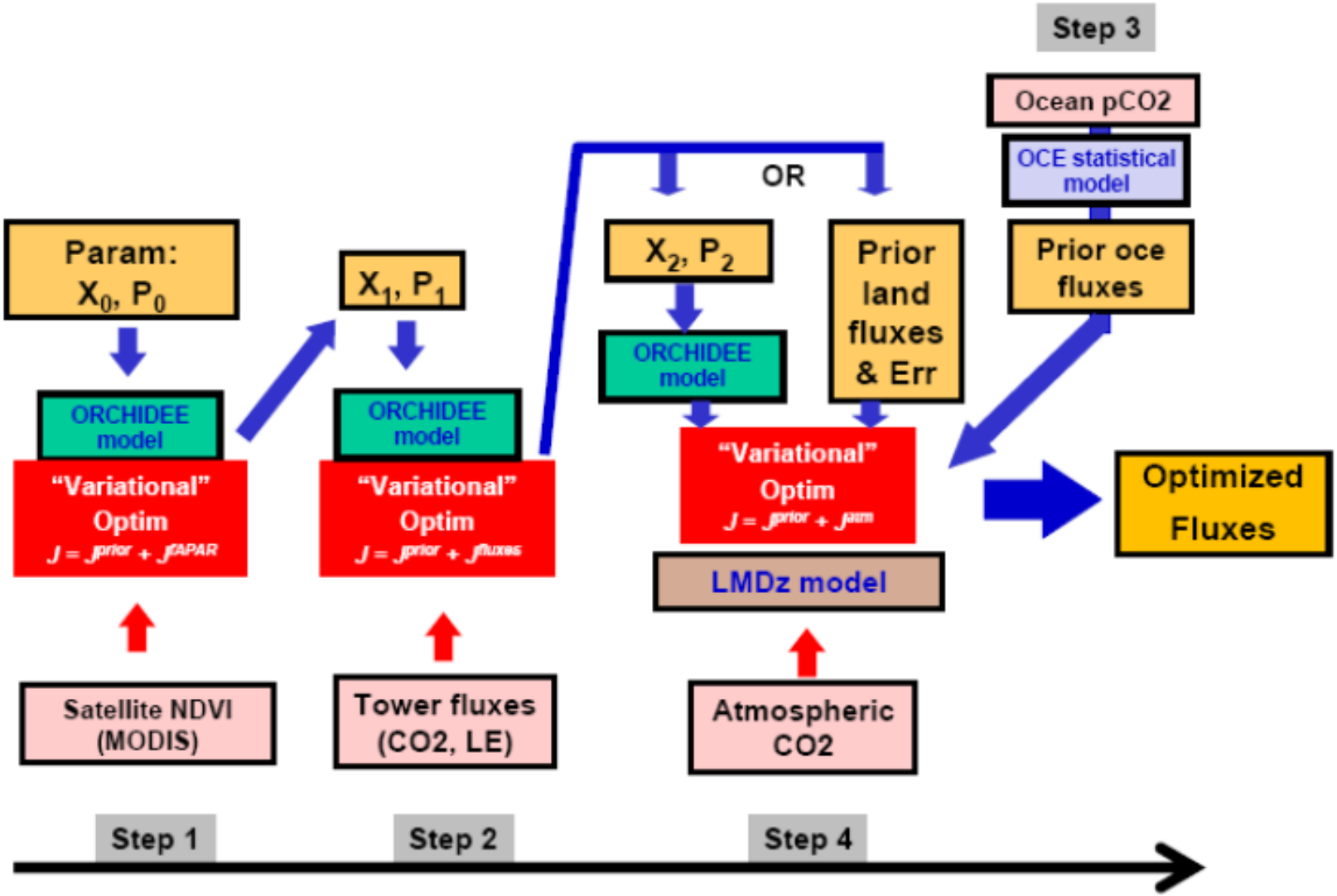
- Huge computing cost (around 8 weeks for 20 yr)
- Difficult to set the relative errors between data streams
- Problem of Equi-finality..

- Each step is more easy to handle
- Easier to control errors per data-stre
- Expose only specific params to specific observations
- Assessment of the contribution of each data stream

- Possible missfit of data stream from "previous steps"
- Difficulties to carry over the error from one step to the other

➔ Stepwise is less rigorous but easier to implement

Exemple of sequential assimilation…

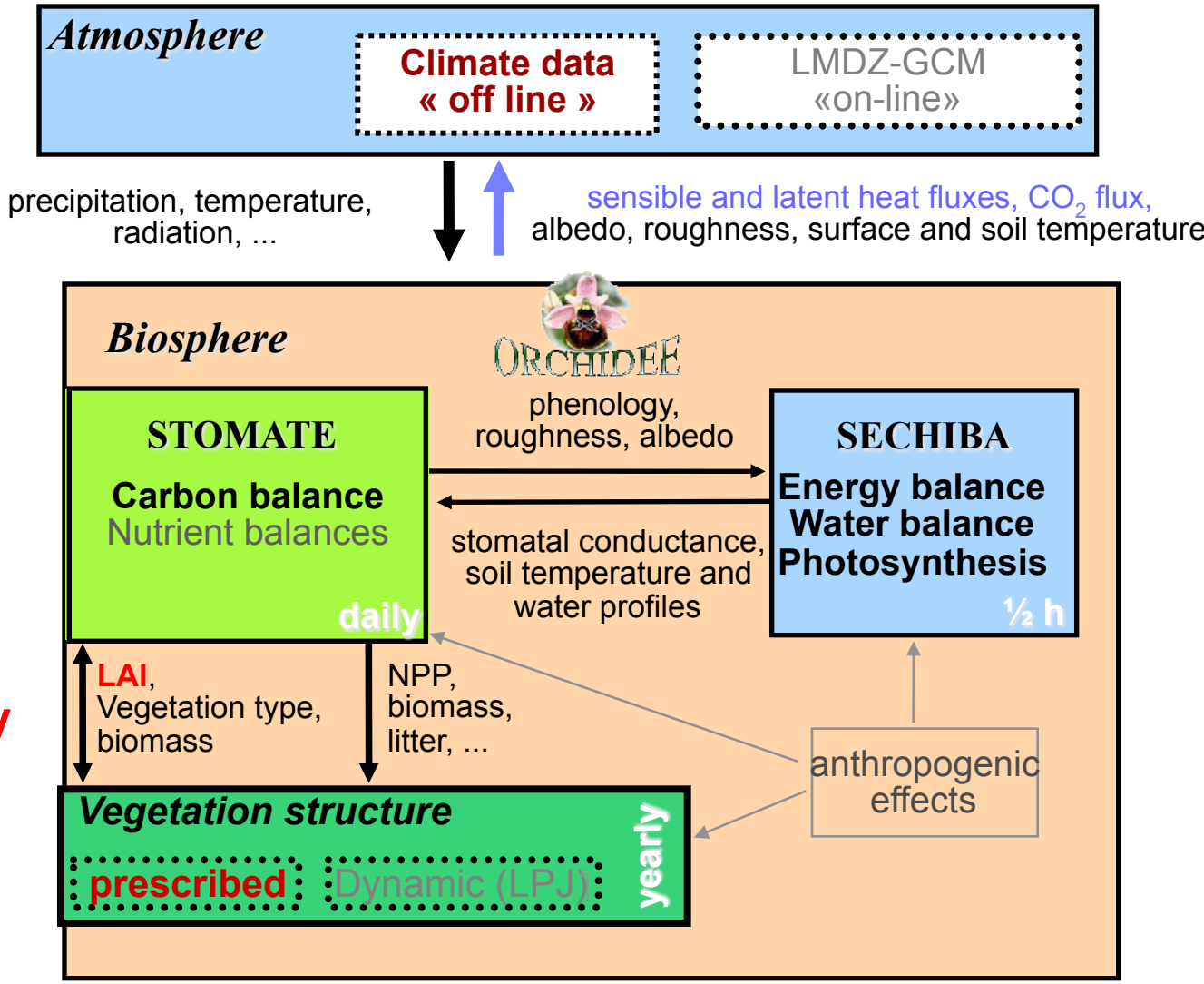


Outline

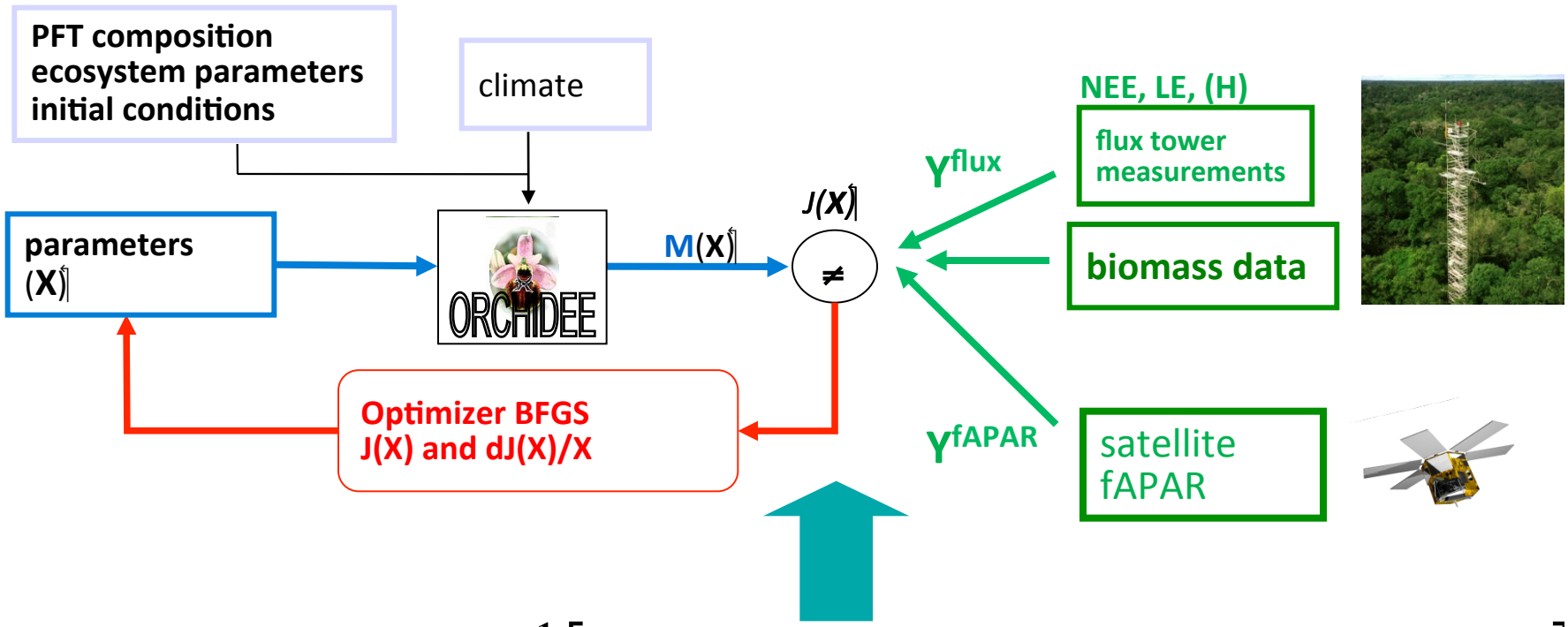
- Current limitations of « standard » atmospheric flux inversions
- Multi-data streams assimilation: Basis for model parameters optimization (CCDAS)
- Potential of several land data streams
  - Satellite vegetation indexes
  - Fluxnet data
  - Biomass measurements
- Join multi-data assimilation
- Limitations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

The ORCHIDEE ecosystem model

- **Process driven model used for IPCC AR5 simulations**
- **Energy / Water / Carbon balances**
- **Global - Site level**
- **13 PFT's**
- **½ hourly time step**
- **multiple C pools**



Implementation..



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

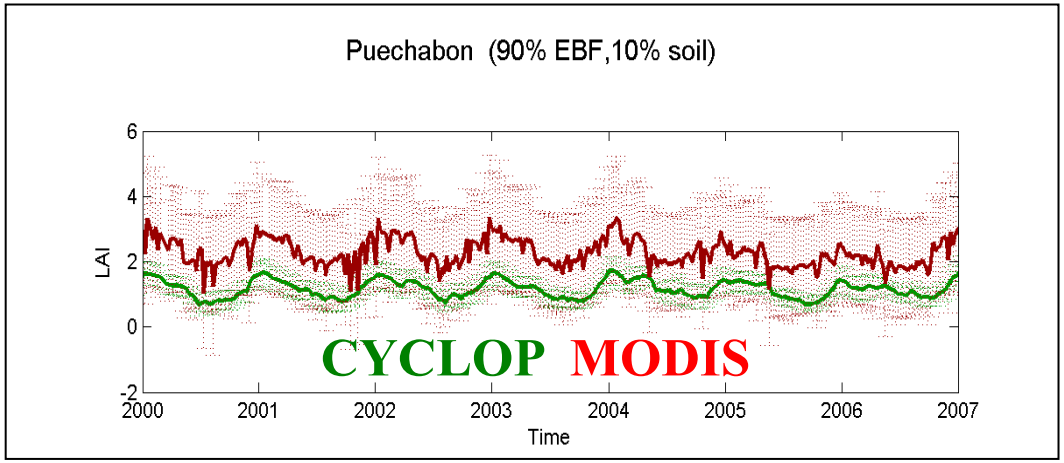
- **Iterative minimization using either:**
  - **Variational approach (with Tangent Linear model for DJ/dx)**
  - **Monte Carlo approach**

Satellite data to optimize phenology…

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

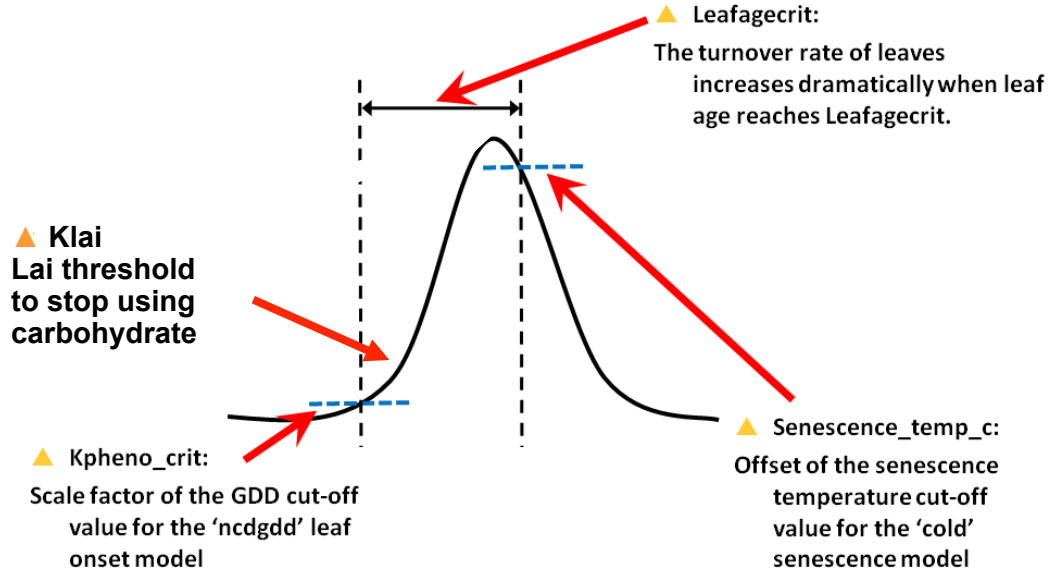
- Recent formal studies with a complete statistical approach:
  - Stockli et al. 2008, 2011
  - Knorr et al. 2010

- Only the temporal variations are robust across products…



MODIS NDVI to optimize phenology…

- Optimize 4 parameters using normalised NDVI (2000-2008) to optimize model fAPAR

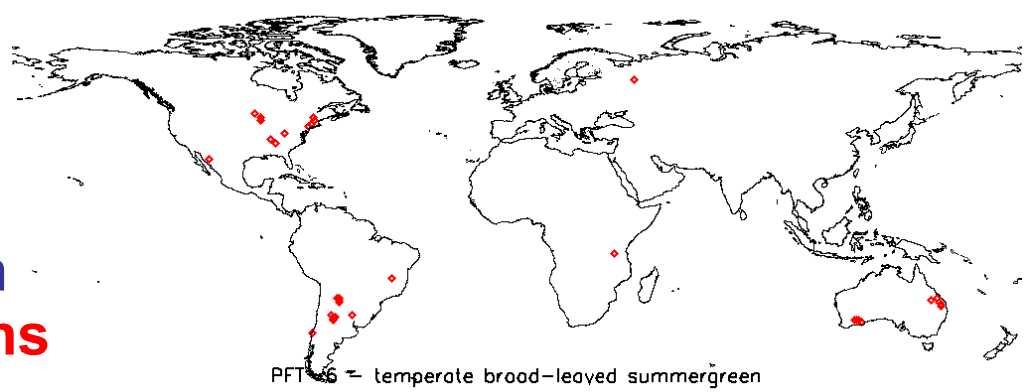


- ORCHIDEE run with IERA Meteo (0.7°)

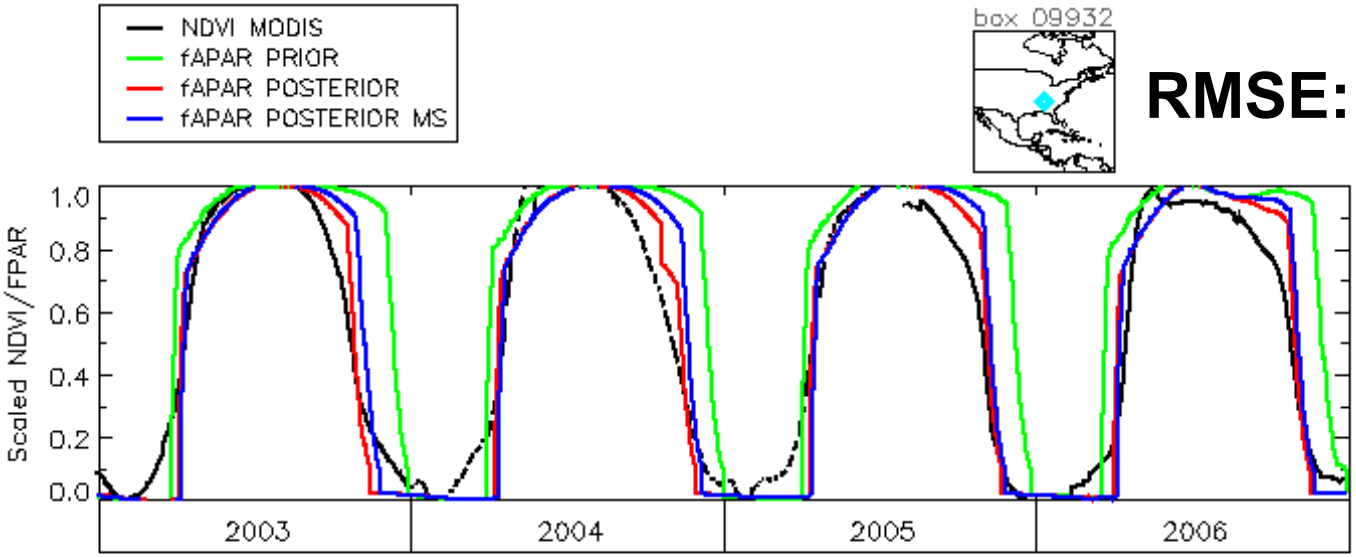
- For each PFT use 30 points with >70% PFT cover

Ex: Temperate deciduous broadleaf forest

1 multi-sites optimization
& 30 single-site optimizatins

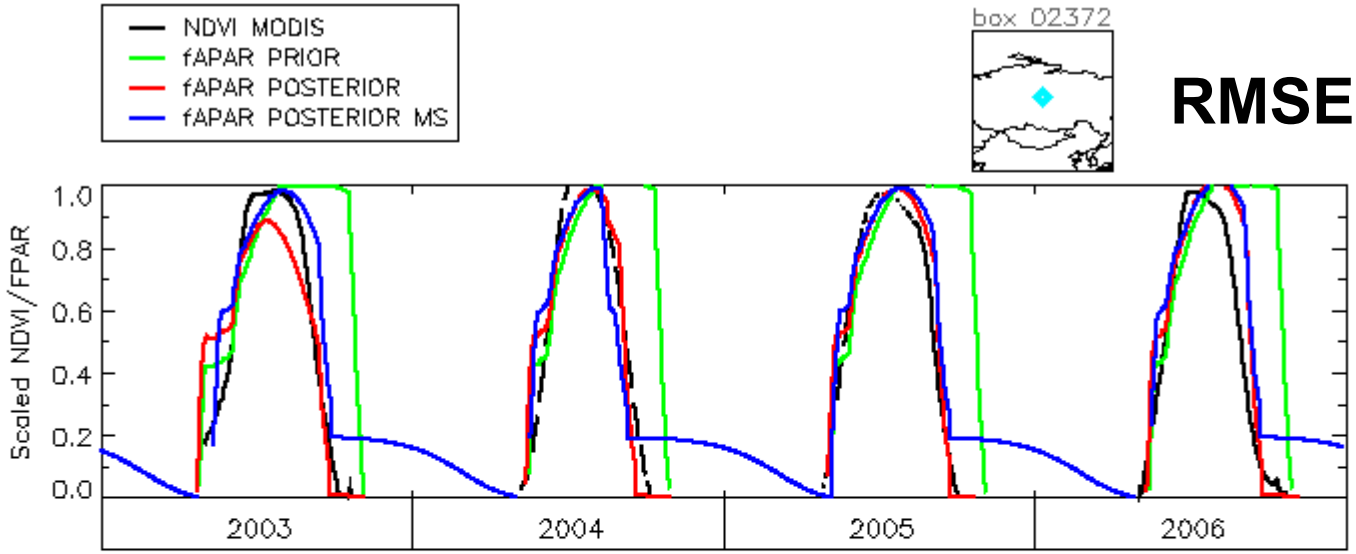


PFT : 'temperate broad-leaved summergreen'



RMSE: 0.33 / 0.13 / 0.16

PFT : 'boreal needleleaf summergreen'



RMSE: 0.43 / 0.13 / 0.16

Prior

Posterior single-site

Posterior Multi-site

Maignan et al. 2012

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:

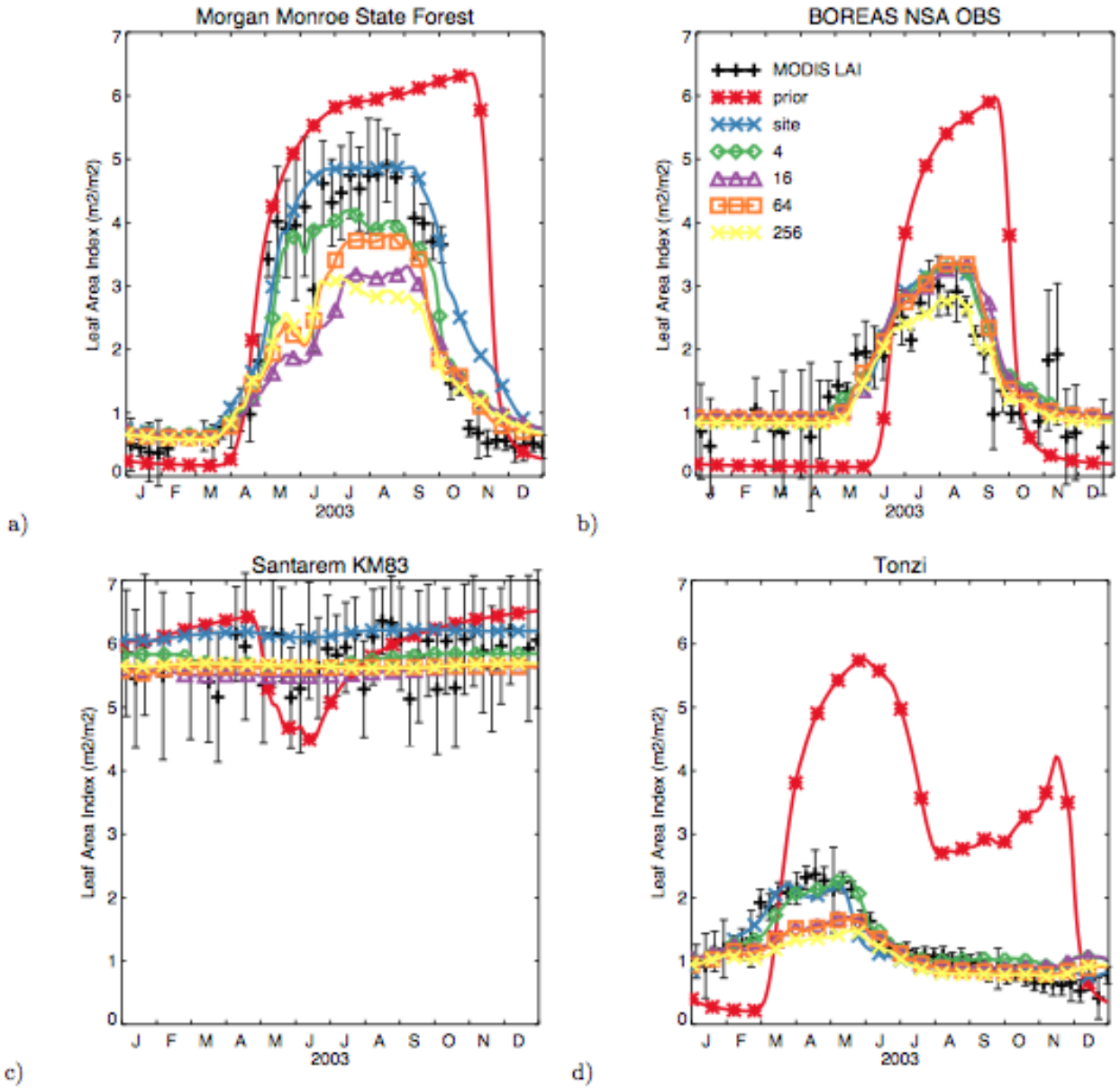
  ➔ **significant improvement..**

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 needeleaf summergreen	**0.39**	**0.89**
PFT 10: C3 grass	**0.46**	**0.56**

Maignan et al. 2012

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.



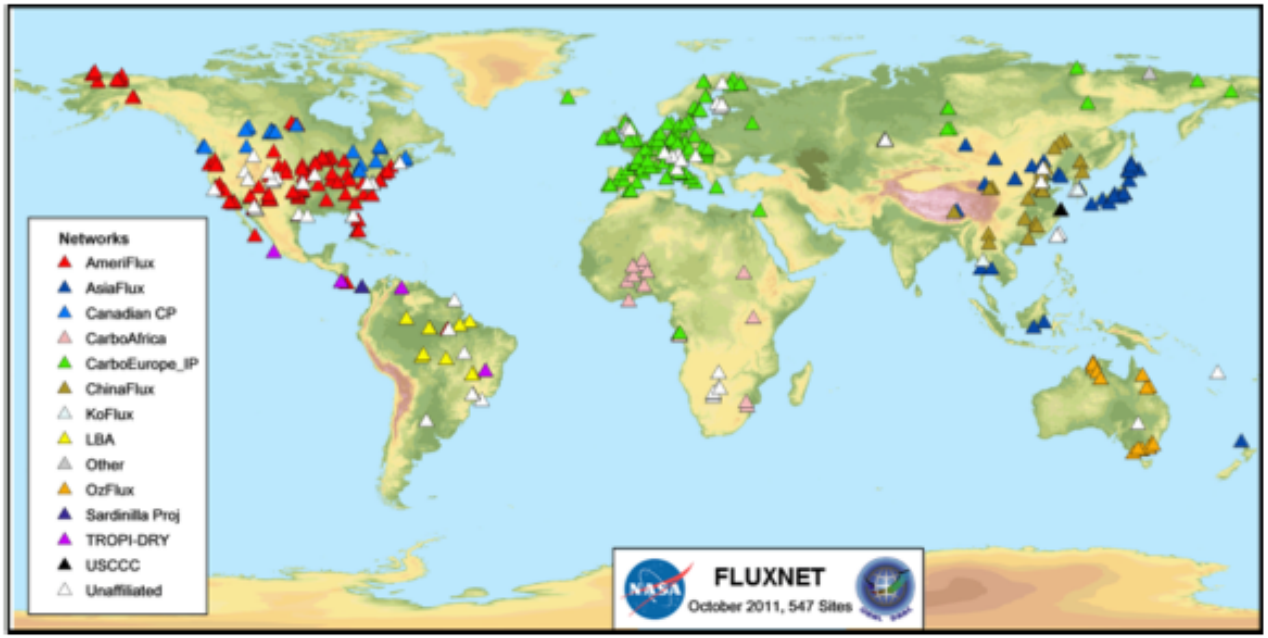
Outline

- 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

Assimilation of Flux data

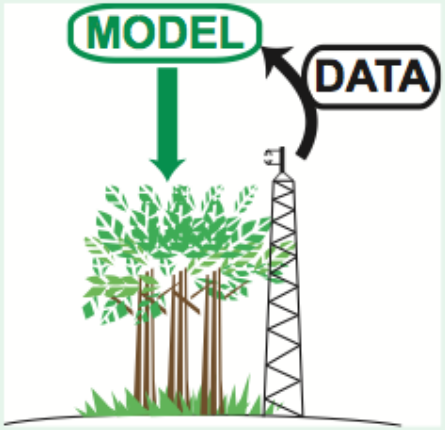
➔ Half hourly measurements of :

- NEE
- LE & SH



History of Flux data assimilation

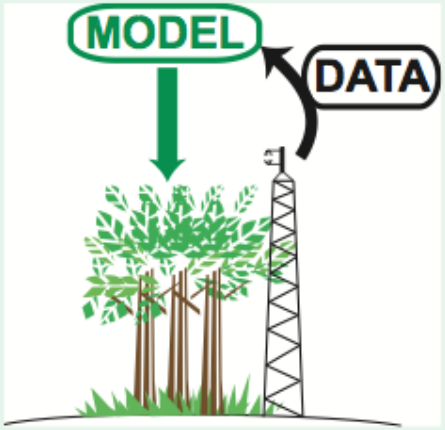
Site-specific optimization...



(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)
...

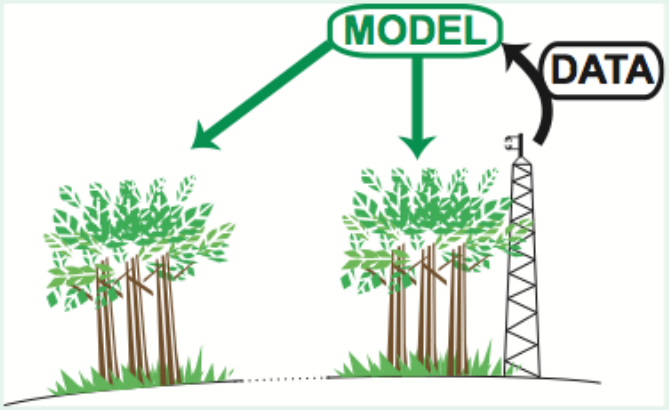
History of Flux data assimilation

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)
...

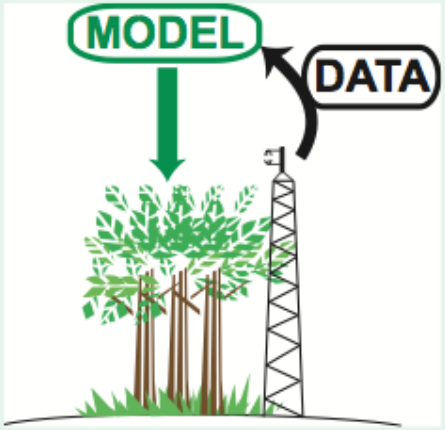
...and evaluation at other sites



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

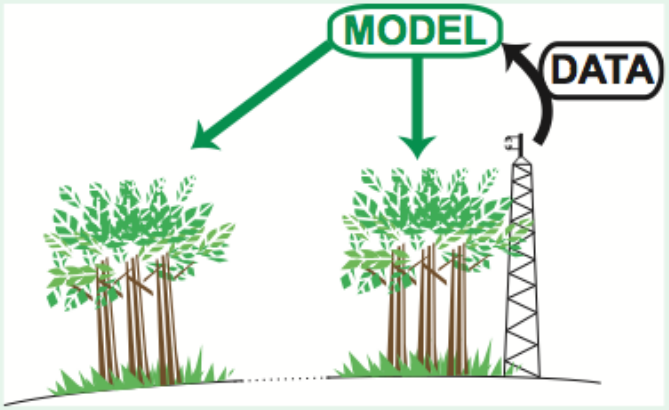
History of Flux data assimilation

Site-specific optimization...



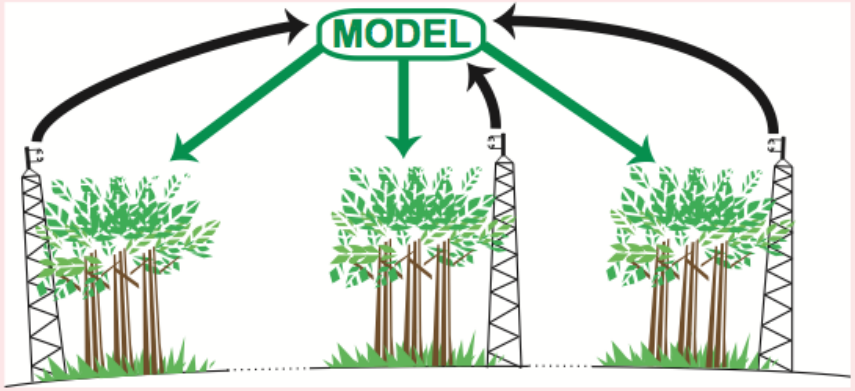
(Wang et al., 2001)
(Braswell et al., 2005)
(Knorr and Kattge, 2005)
(Santaren et al., 2007)
(Wang et al., 2007)
(Williams et al., 2009)
(Carvalhais et al., 2010)
(Verbeeck et al., 2011)
...

...and evaluation at other sites



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

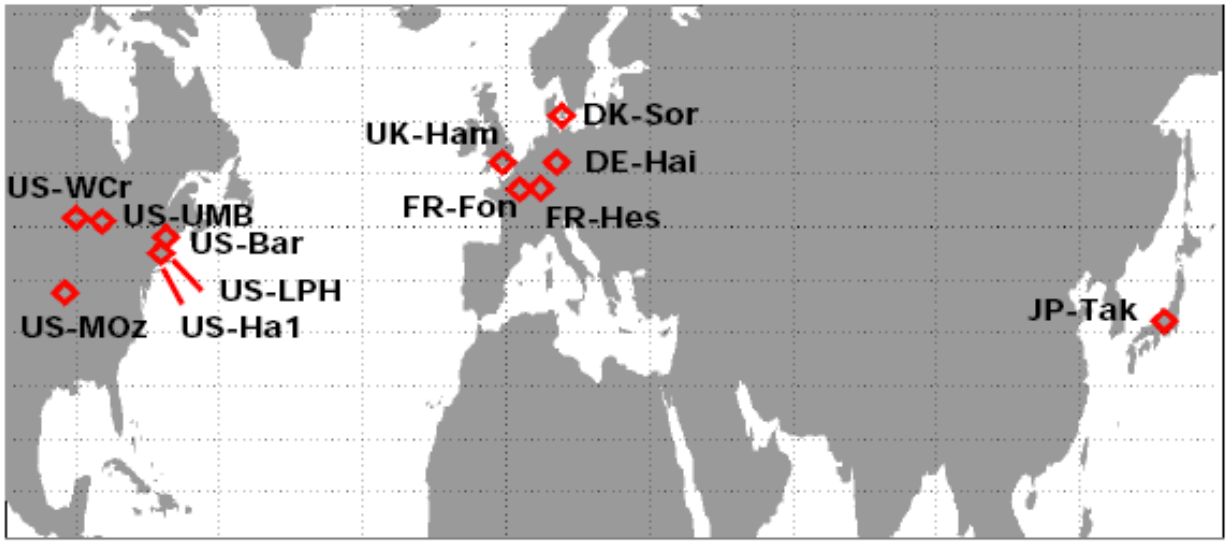
Multi-site optimization



- ➔ 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**

Kuppel et al. 2012

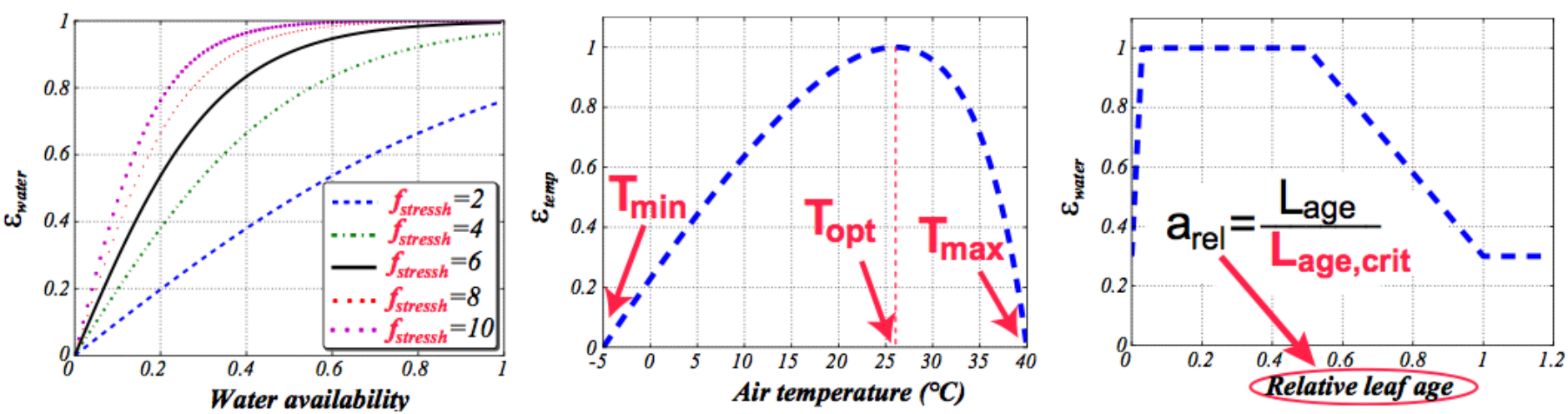
Example of parameter equations

C3 photosynthesis

Photosynthetic capacity	Stomatal conductance	CO_2 diffusion
$A = \min(V_c, V_j)\left(1 - \frac{\Gamma^*}{C_i}\right) - R_d$	$g_s = \mathbf{G_{s,slope}} \frac{h_r}{C_a} A + G_{s,offset}$	$A = g_s\,(C_a - C_i)\,/1.6$

PFT-dependent parameters / Fixed equations:

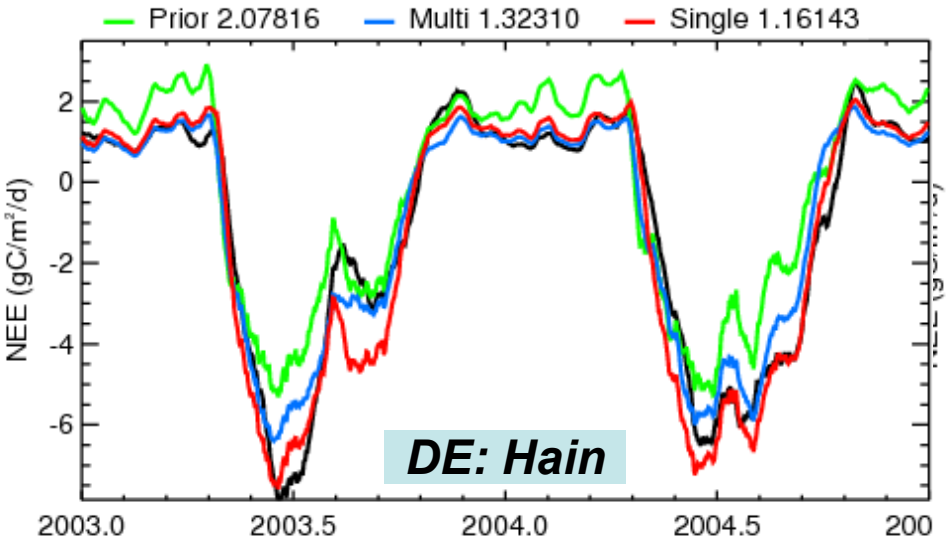
$$V_{\{c,j\}max} = \mathbf{V_{\{c,j\}max,opt}} \cdot \varepsilon_{water}(\mathbf{f_{stressh}}) \cdot \varepsilon_{temp}(\mathbf{c_{Tmin}}, \mathbf{c_{Topt}}, \mathbf{c_{Tmax}}) \cdot \varepsilon_{leaf}(\mathbf{L_{age,crit}})$$

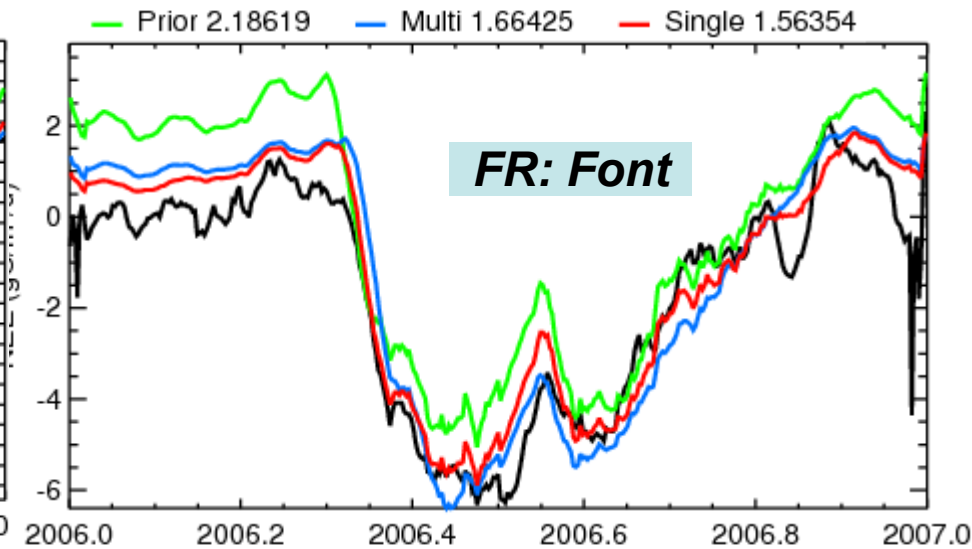


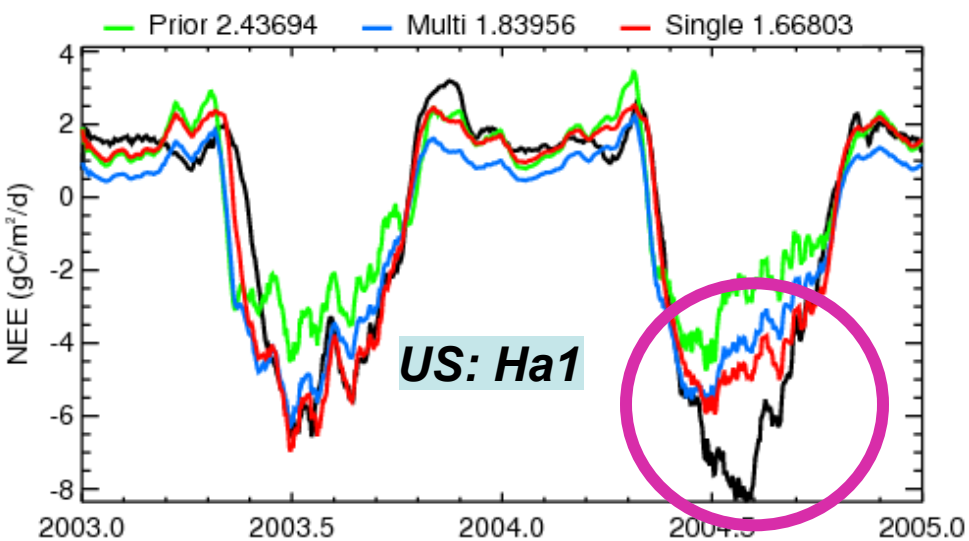
Optimized model parameters : 21 per PFTs

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 (gC/m2/d)

Data

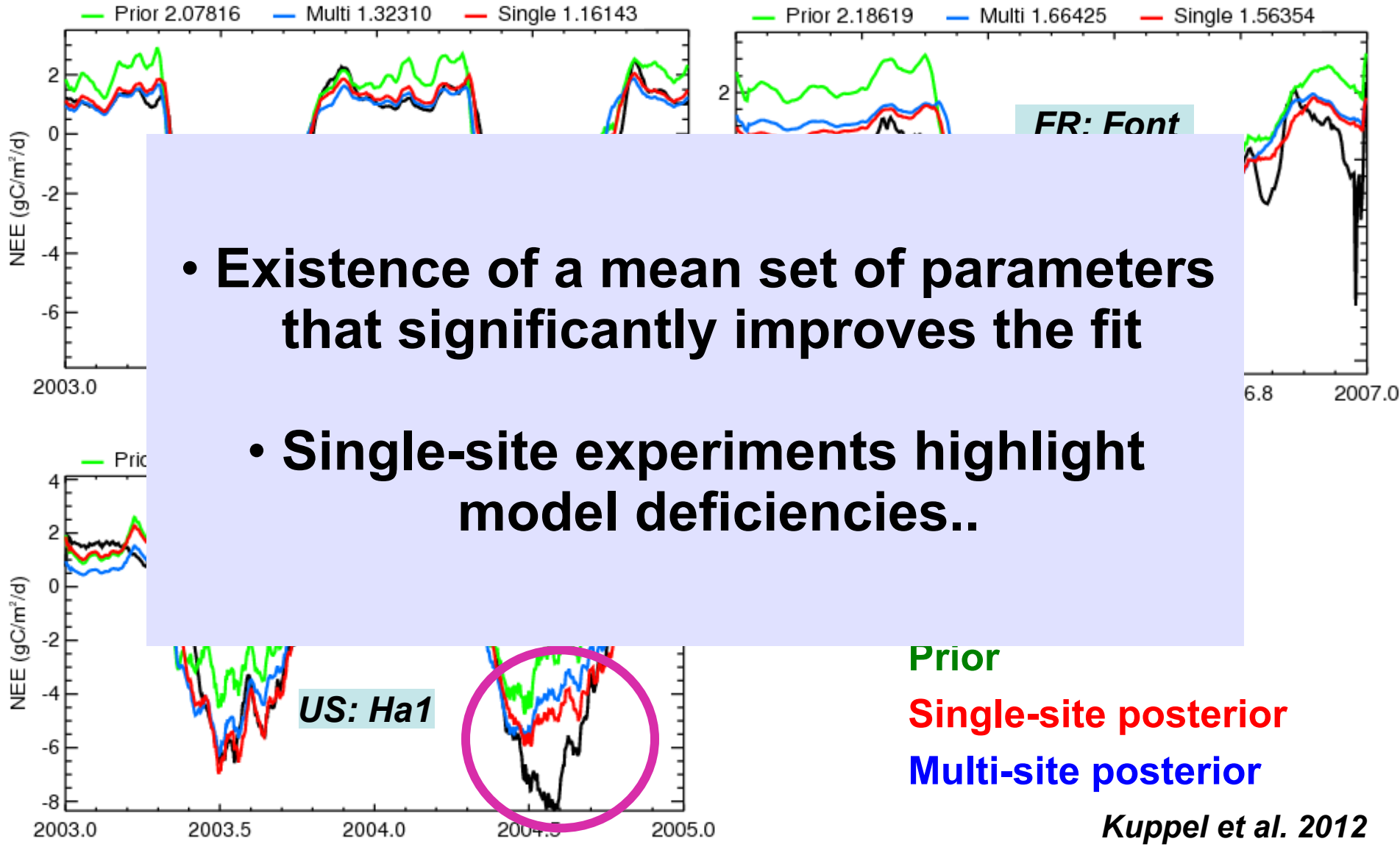
Prior

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

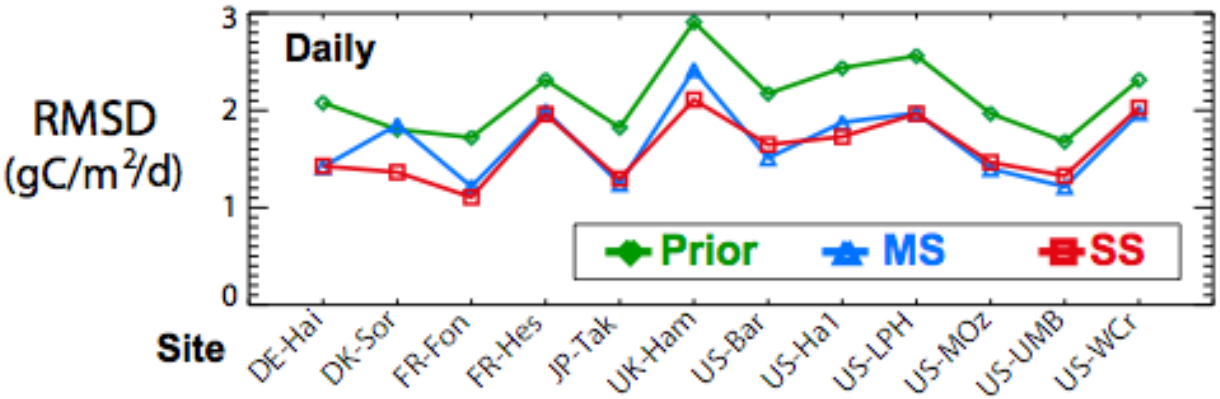
Single-site posterior

Multi-site posterior

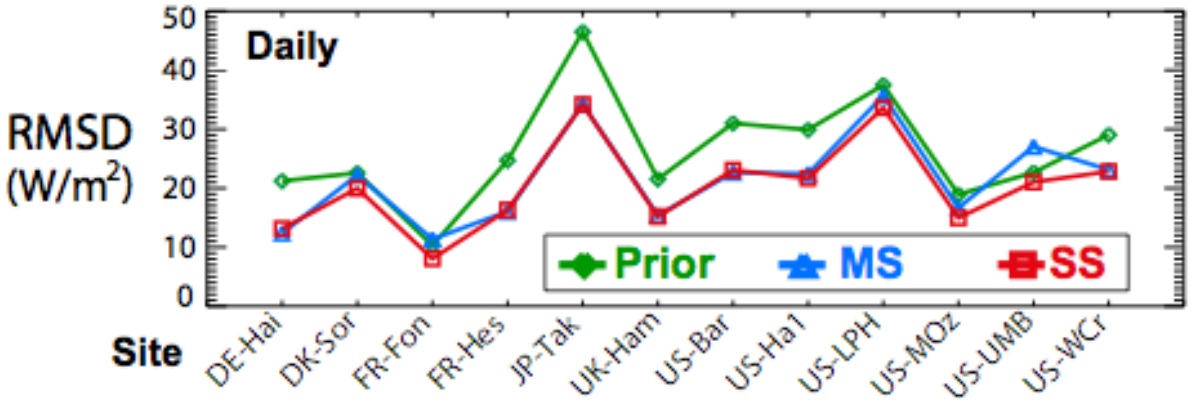
Kuppel et al. 2012

Model-data fit for NEE (RMSE)

Carbon flux (NEE)



Latent heat flux (LE)



Prior

Single-site posterior

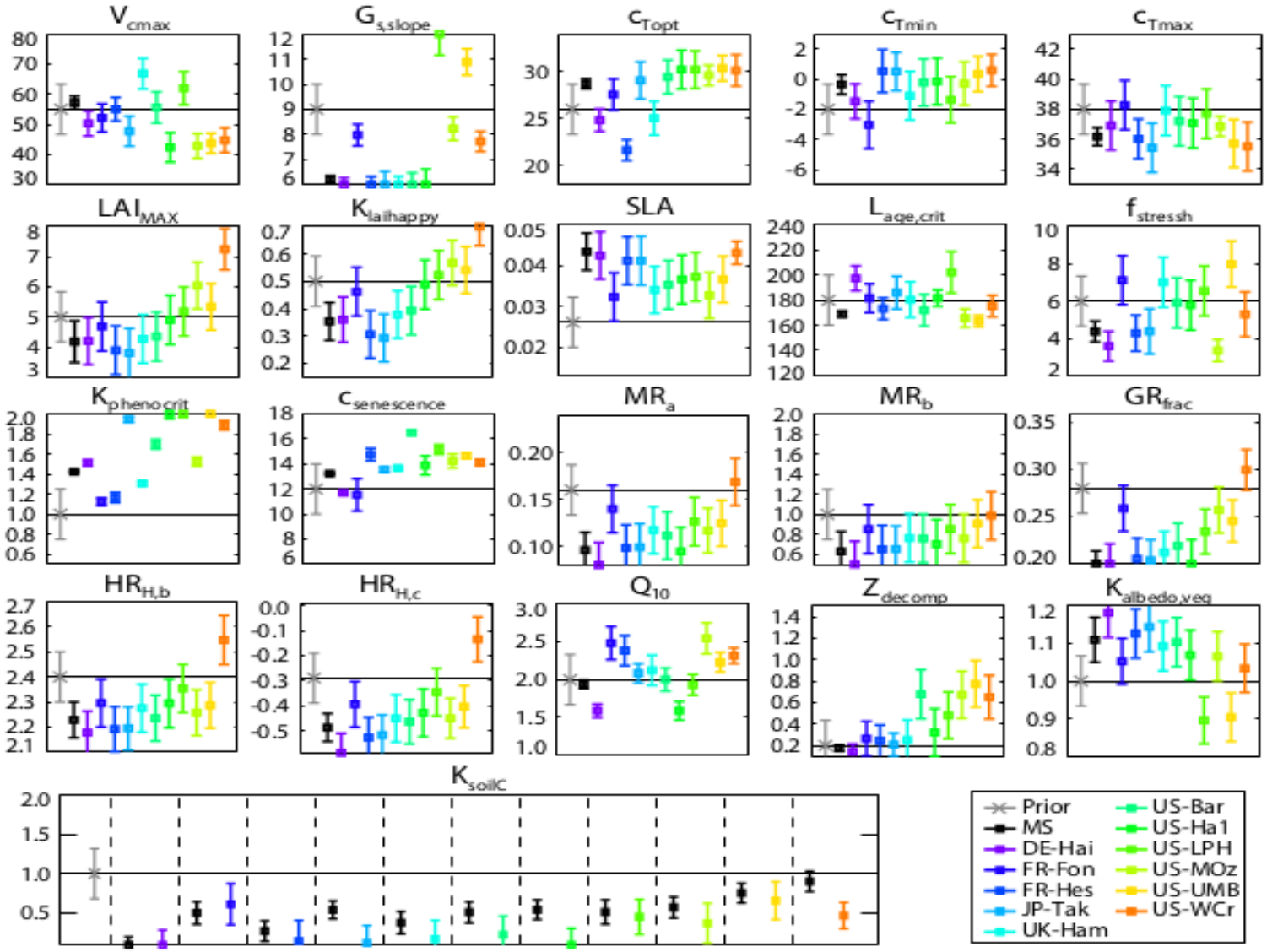
Multi-site posterior

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

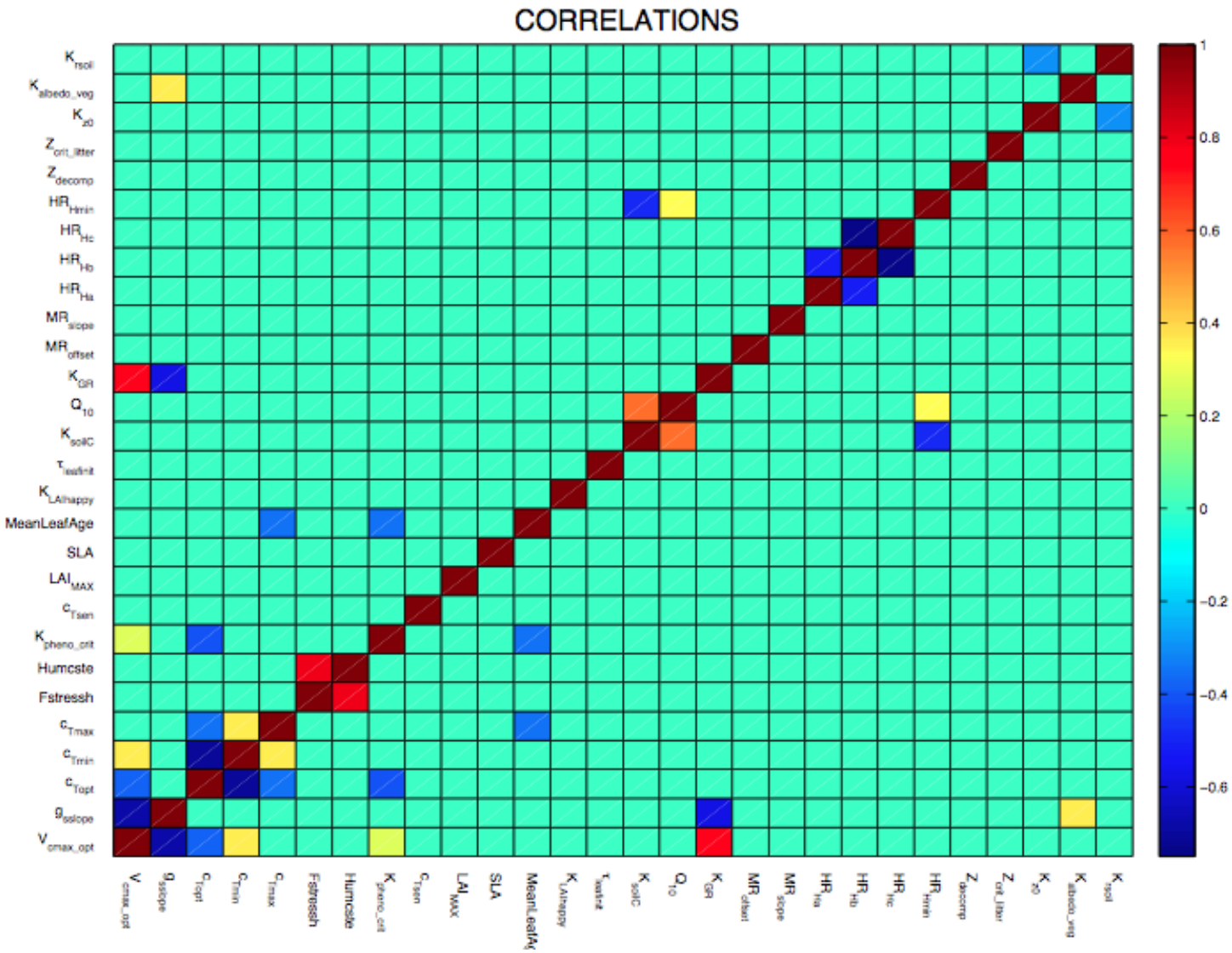
Kuppel et al. 2012

Parameters errors

Black: Multi-site Colors: Single-site

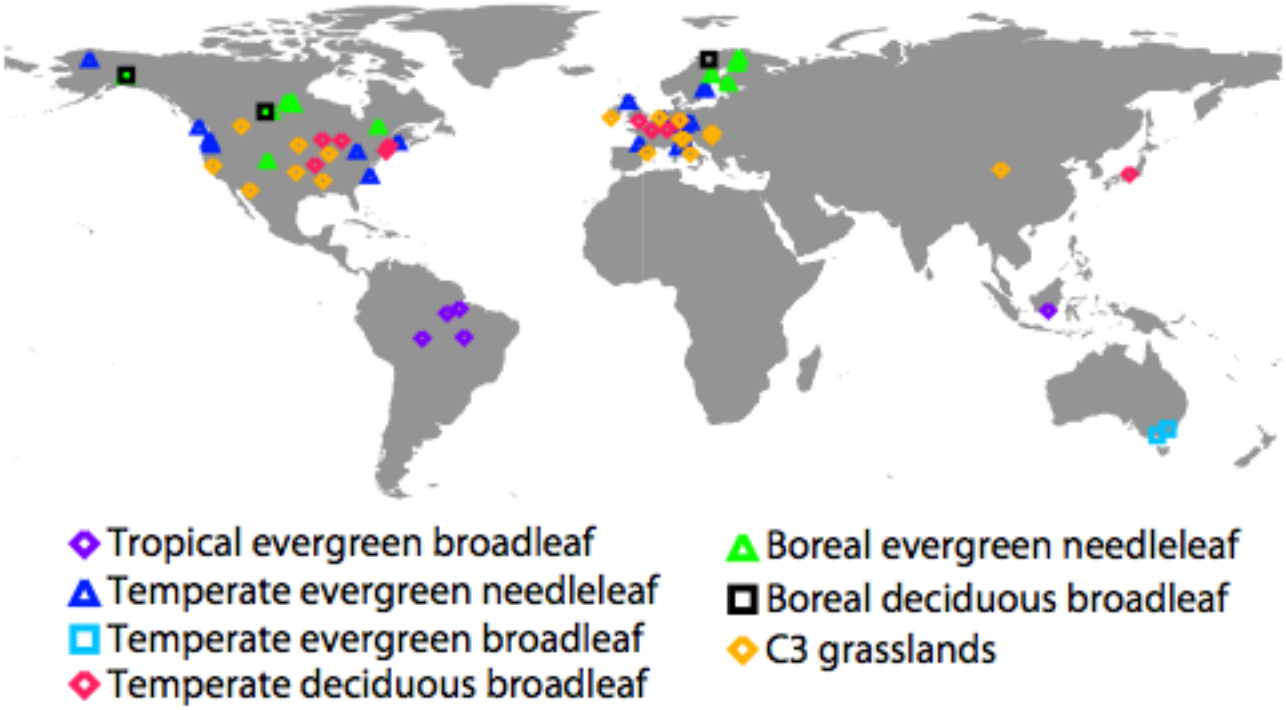


Posterior error covariance matrix



Assimilation of FluxNet data for all PFTs

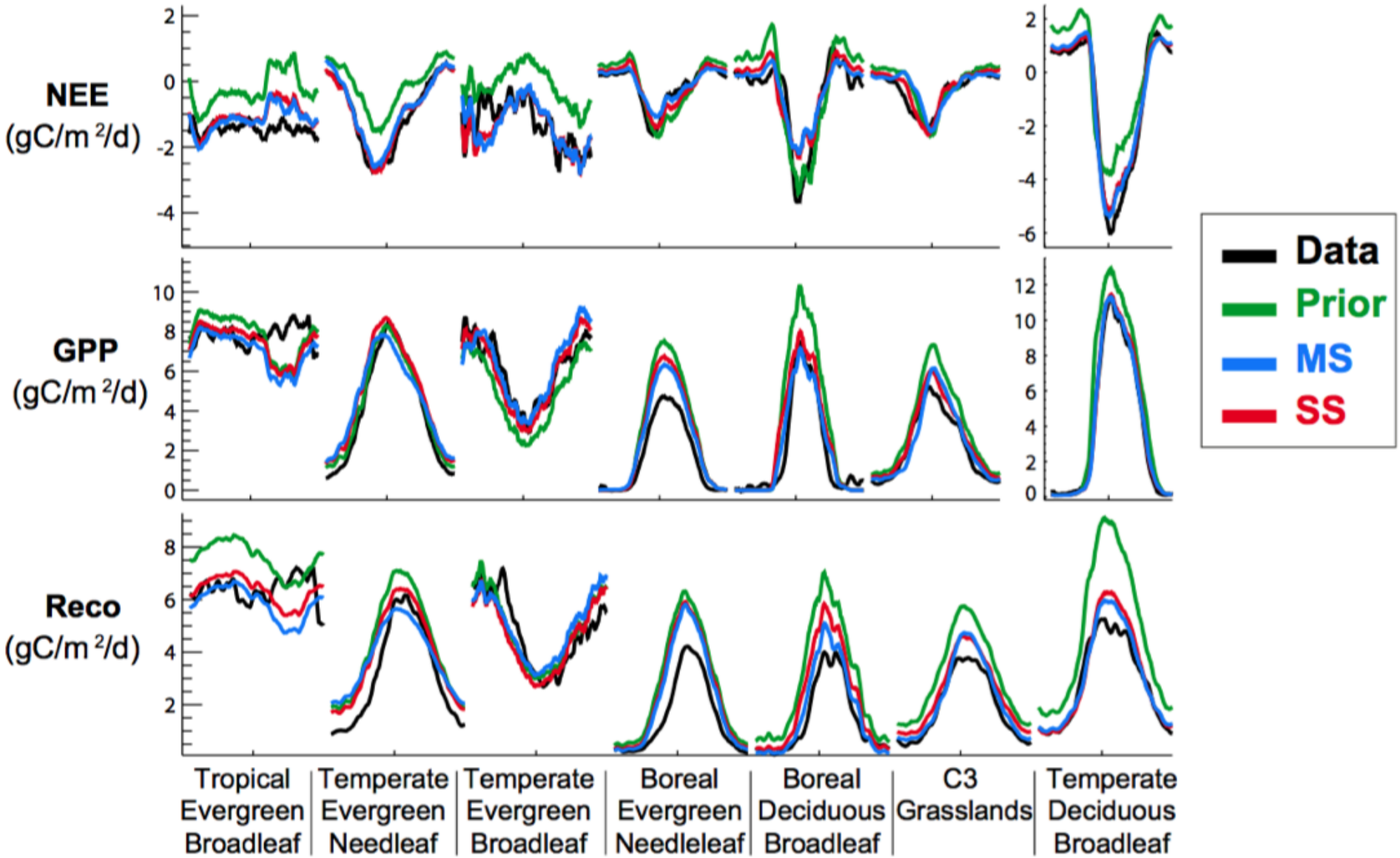
Localisation of all sites used in the Oprimization



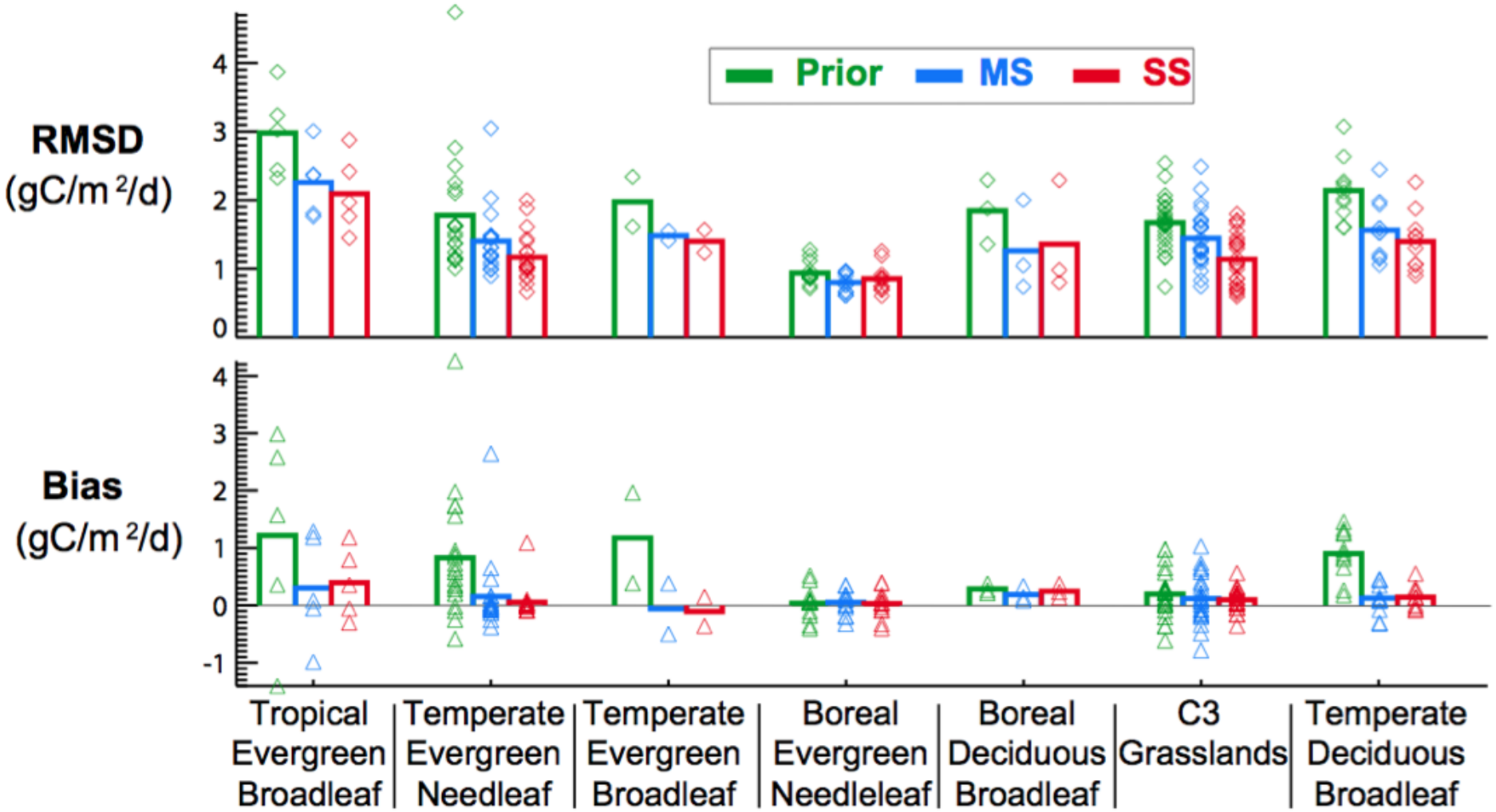
➔ Between 60 and 80 sites depending on the tests

➔ NEE & LE : Correction for the Energy budget

Results for all Plant Functional Types…

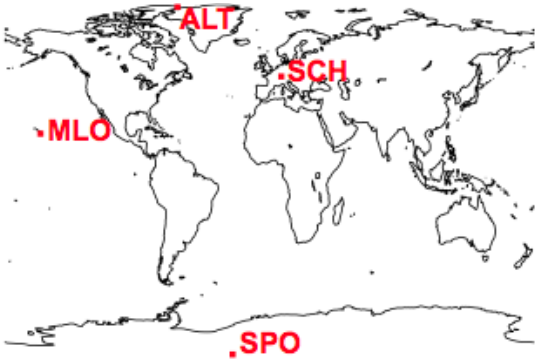


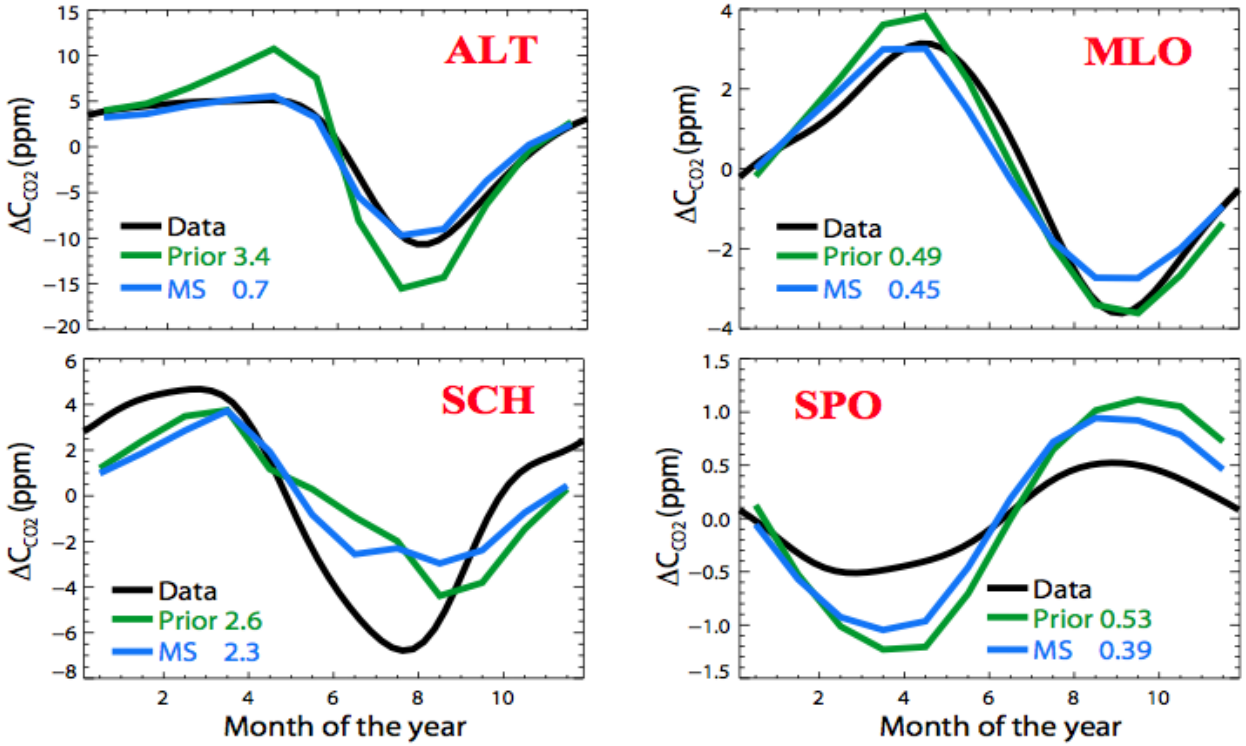
Results for all Plant Functional Types…



Evaluation against atmospheric CO2

Mean seasonal cycle at ground-based observatories (1989-2009)





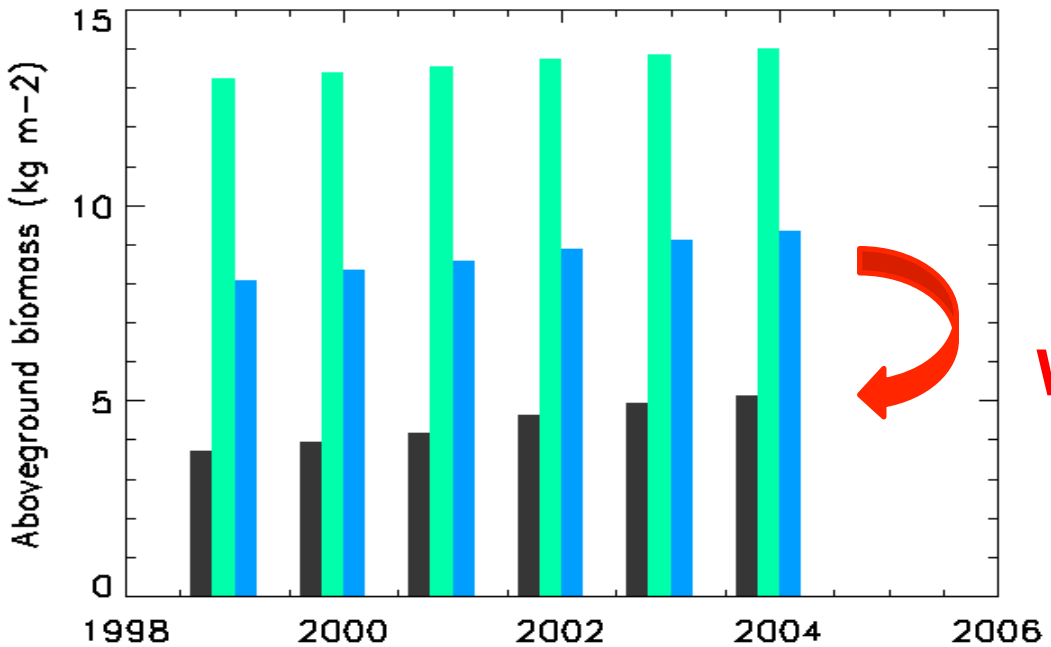
- Major improvement from NH boreal ecosystems' fluxes
- Need for more constraints elsewhere ?

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

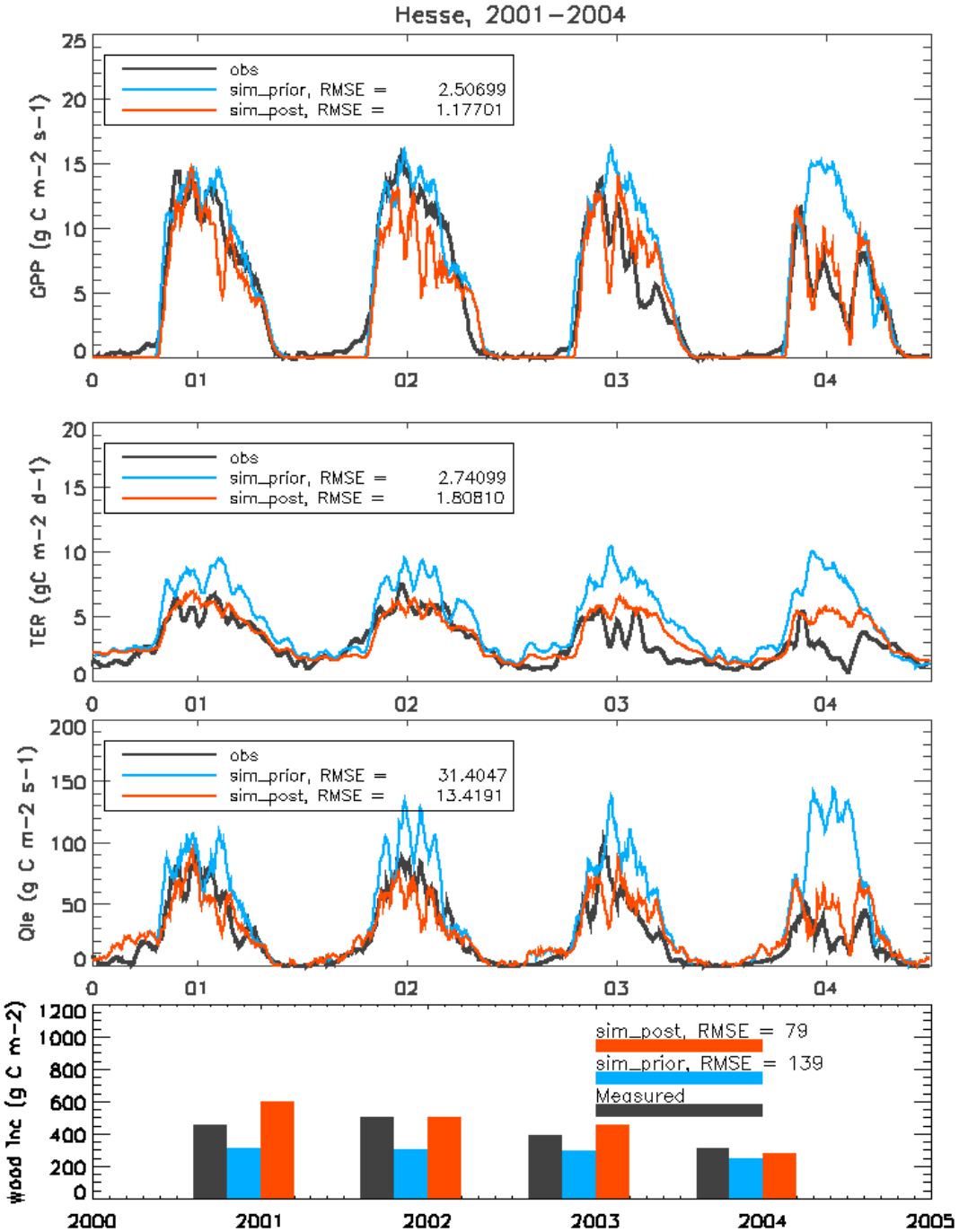
Measurement

Model: Steady state

Model: Realistic age

Thum et al. 2012

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



(25 flux related params)

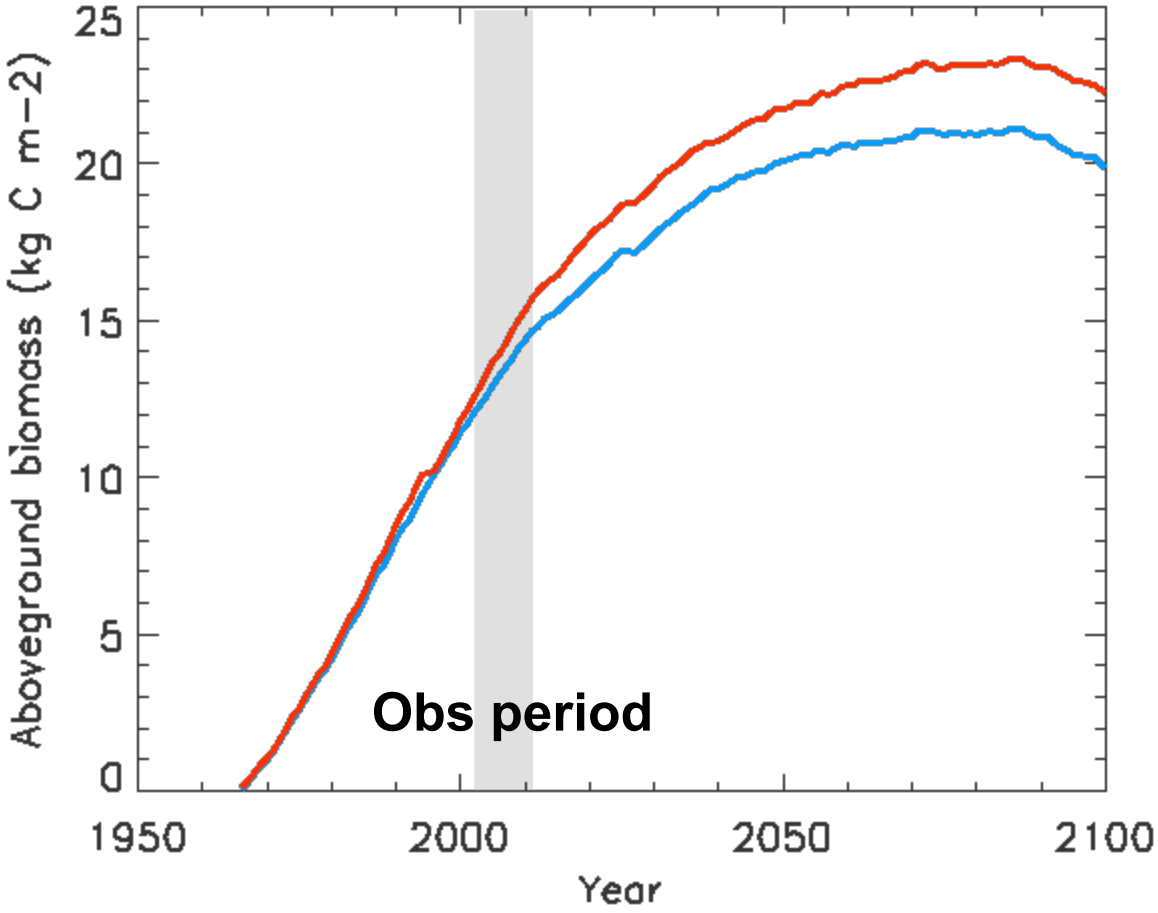
Measurement

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

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

Impact on futur run: 1965-2100: aboveground biomass

Above ground biomass (Hesse site)



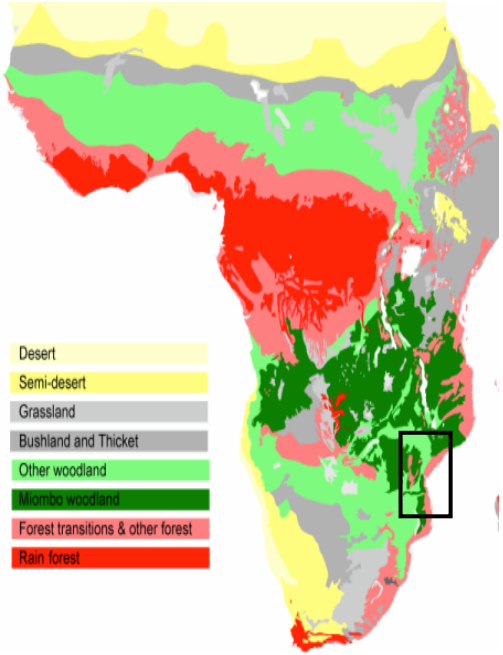
Default parameter values

Optimized parameter values

Thum et al. 2012

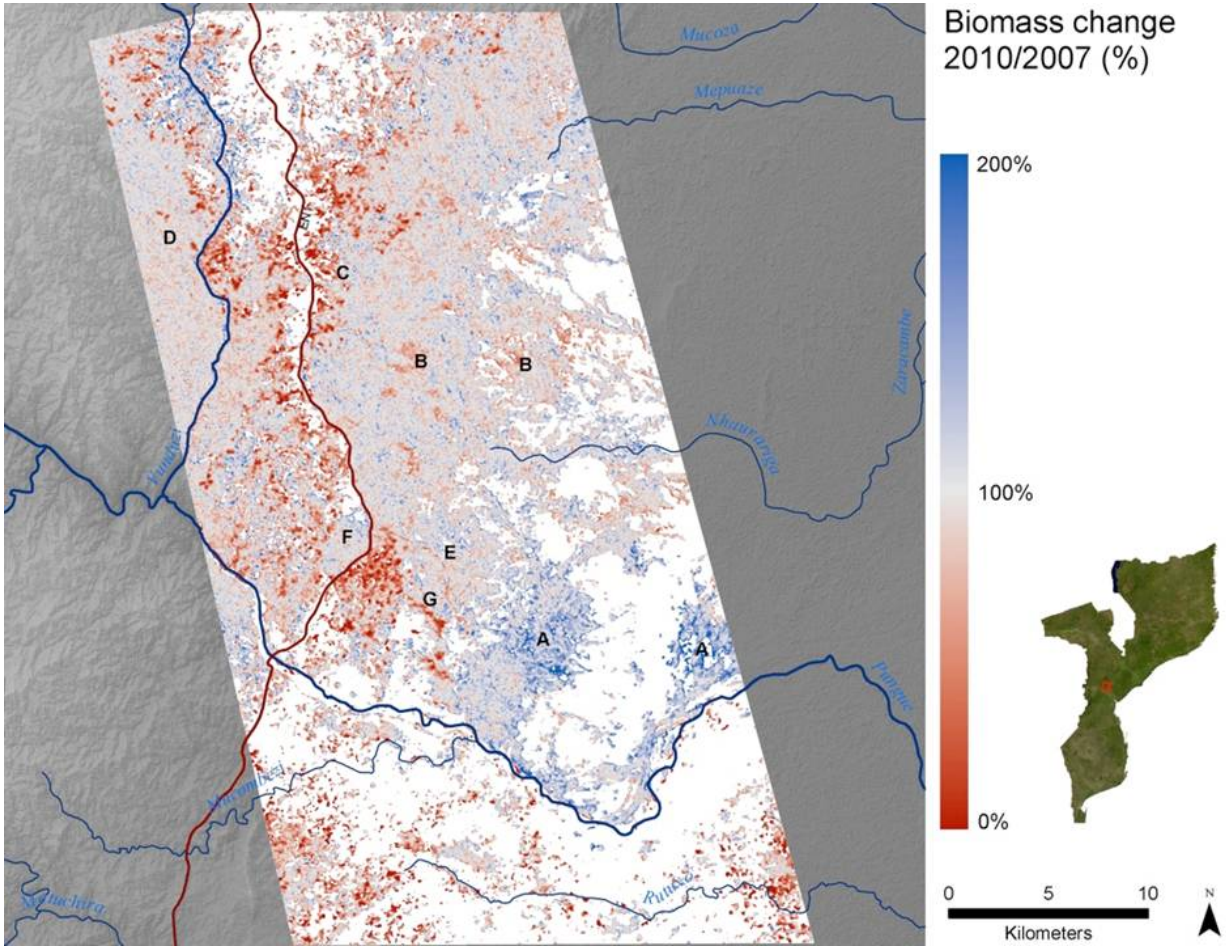
Using radar retrievals of forest biomass data Over tropical woodlands

Biome demography is critical for African woodlands

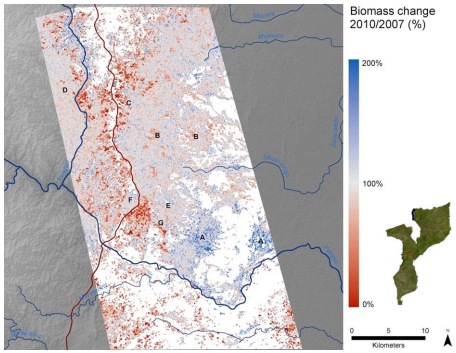


Courtesy of Mathew Williams et al.

Biomass change from 2007 to 2010

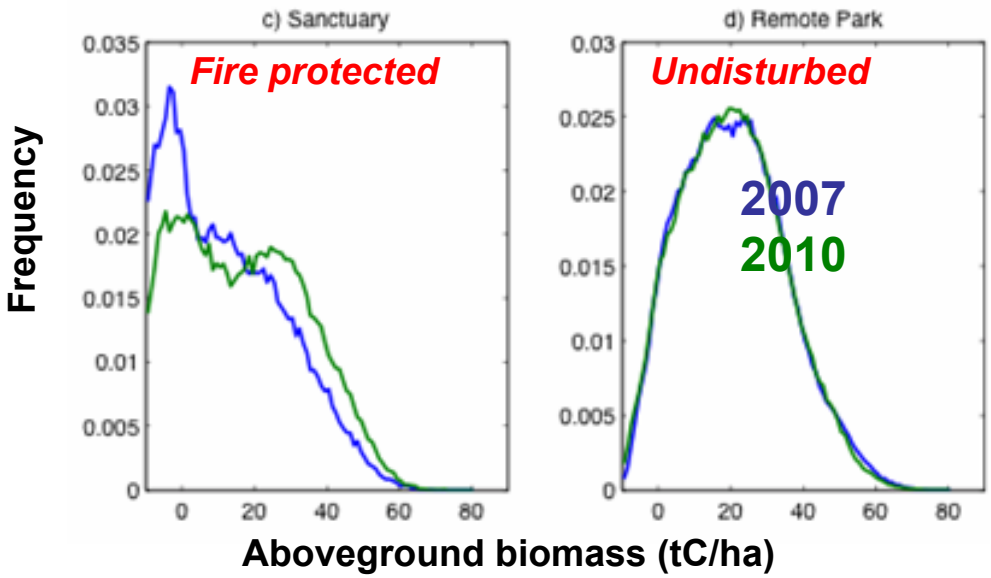


Using radar retrievals of forest biomass data Over tropical woodlands



$$\Delta C_w = a_w \, NPP - t_w \, C_w - P \, F \, C_w$$

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.

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. CO2
- Join multi-data assimilation
- Limitations & Prospects

First CCDAS with Atmospheric CO2 : Rayner et al. 2005

- Optimizing 57 paramerters of BETHY
- Using TM2 transport model with 41 stations

Flux IAV North / Tropic partition

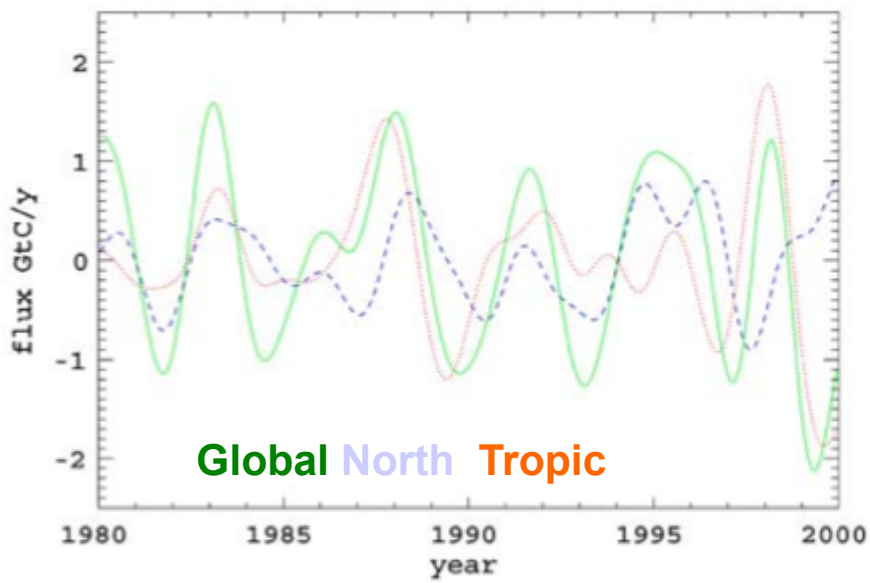


Figure 11. Global (solid line), tropical (20°S–20°N) (dotted line), and northern extratropical (20°N–90°N) (dashed line) anomalies in flux to the atmosphere from the optimized model.

Flux IAV NPP / Resp partition

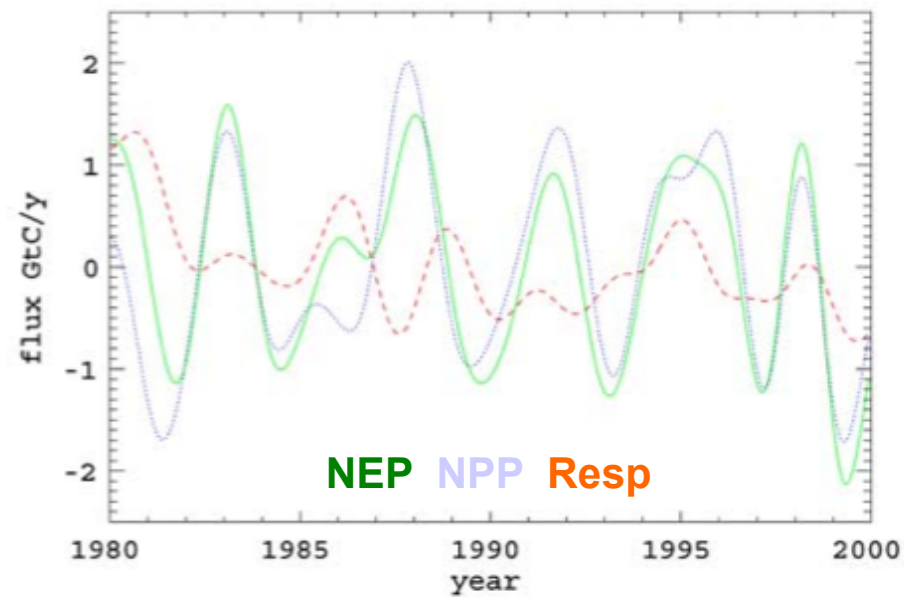
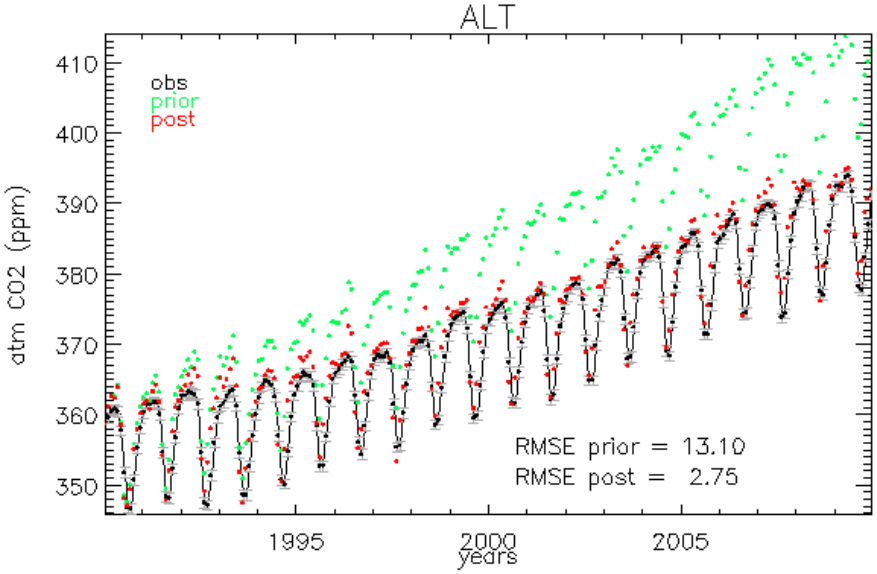


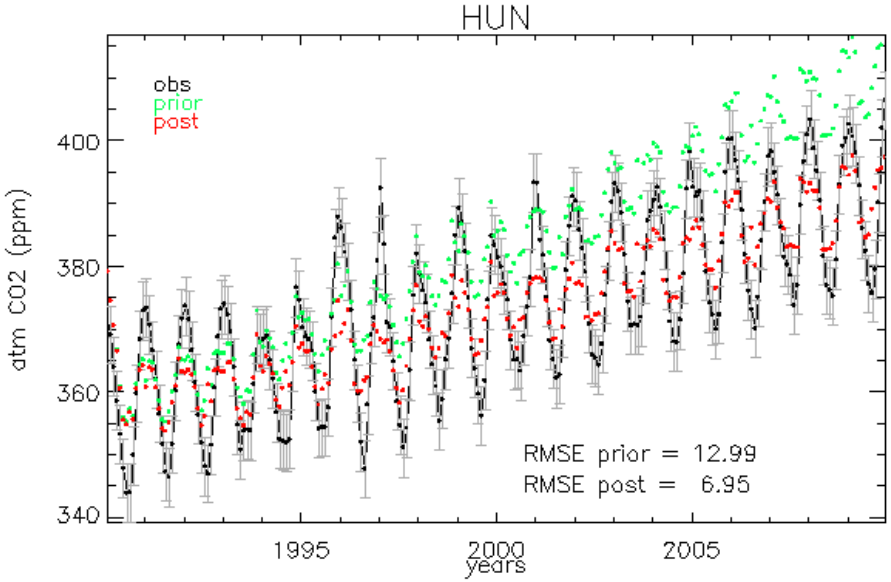
Figure 12. Global anomalies in negative NEP (solid line), negative NPP (dotted line), and fast respiration (dashed line) on interannual timescales. See text for details of time filtering.

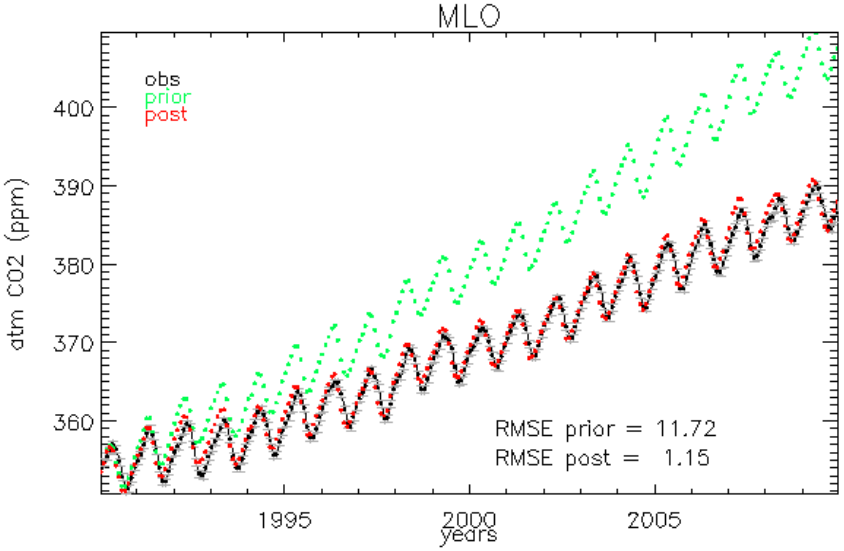
LSCE-CCDAS : Assimiation of Atm. CO2 data

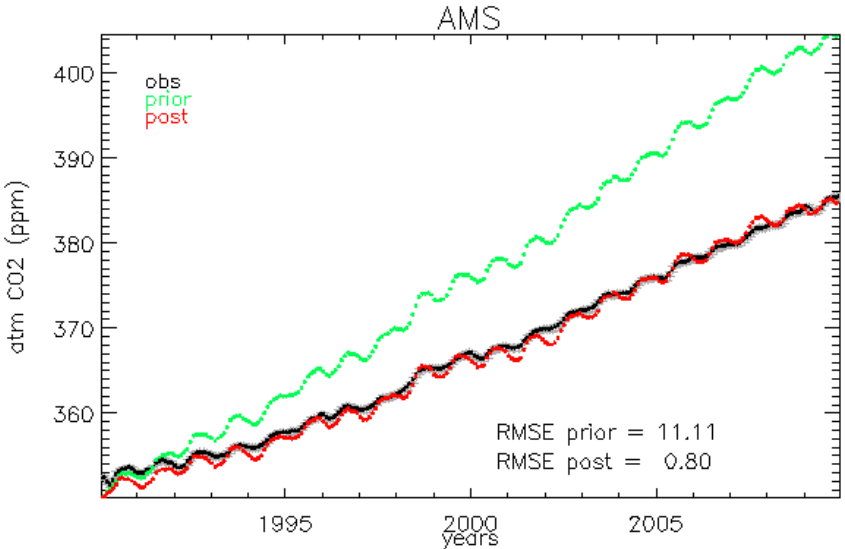
MODIS NDVI	FluNet NEE / LE	Atmospheric CO2
4 phenology params per PFTs	4 + (≈ 15) ≈ 20 params per PFTs (photosynthesis, respiration)	Initial soil C = 50 params + 3 params per PFTs (from previous set)
40 params	≈ 100 params	≈ 80 params

LSCE-CCDAS : Fit to the data

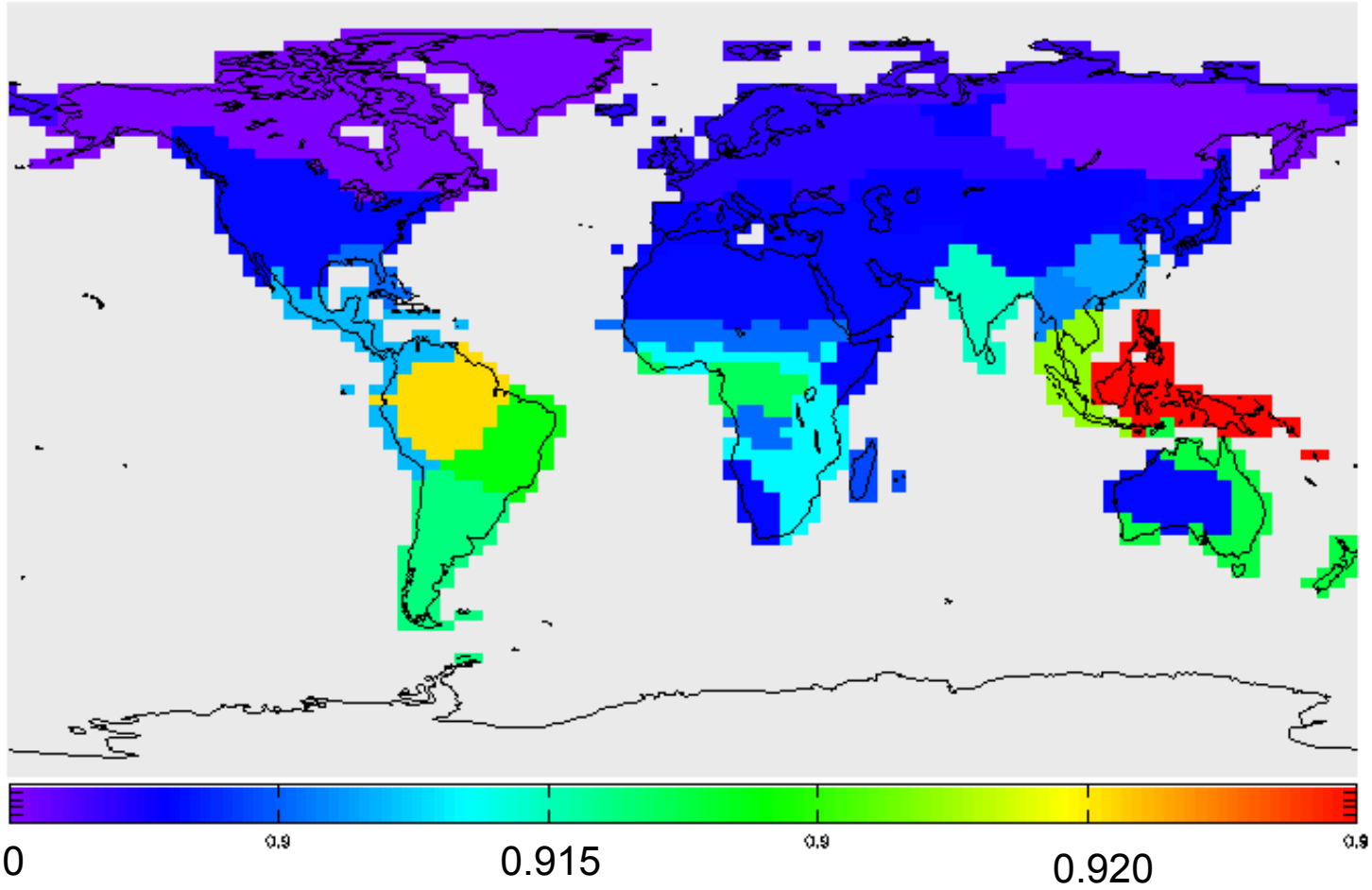








Optimized « scalar » of initial soil carbon pools sizes…



Multi-data simultaneous Assimilation

- Very few emerging studies…

- Results with BETHY ecosystem model
  + TM2 transport model

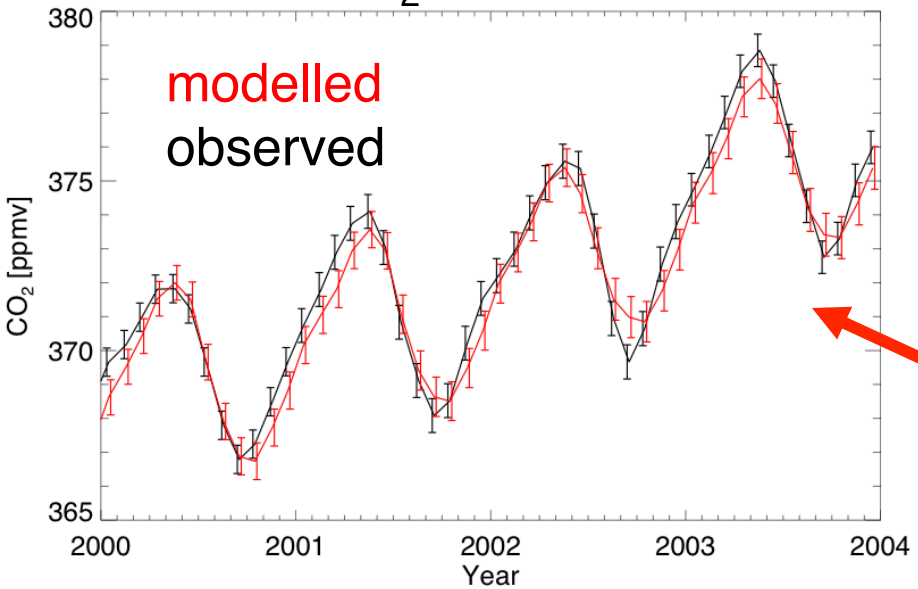
  - ✓ Assimilation of Atm CO_2 and satellite fAPAR
    (*Scholze et al. 2007*)

- Preliminary results with ORCHIDEE + LMDz

  - ✓ MODIS-NDVI + FluxNET + Atm CO_2
  - ✓ 3 years

Join Assimilation: “BETHY” CCDAS

Atm. CO_2 at Mauna Loa

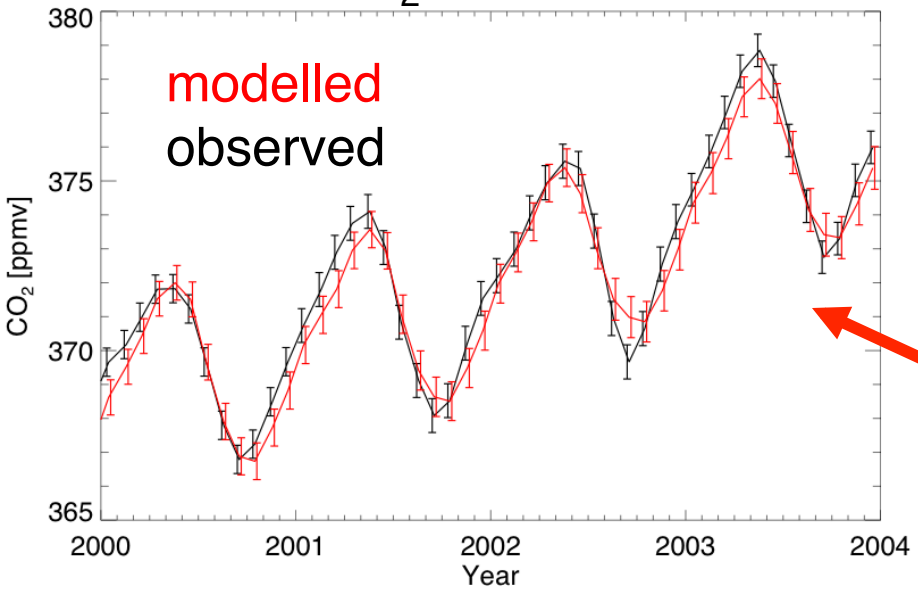


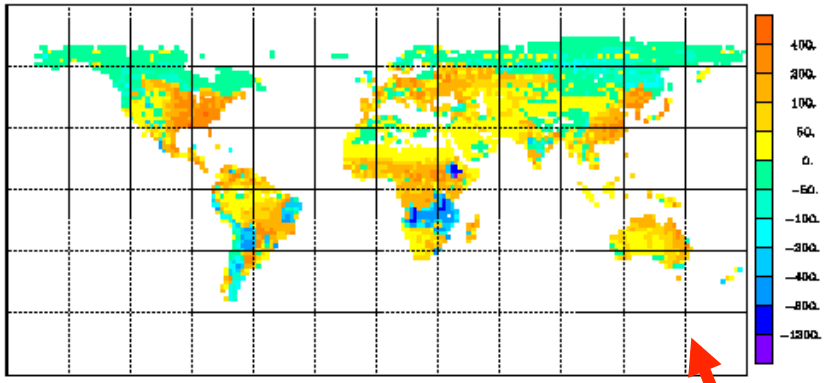
- Predicted atm. CO_2

Rayner et al., 2005; Scholze et al., 2007

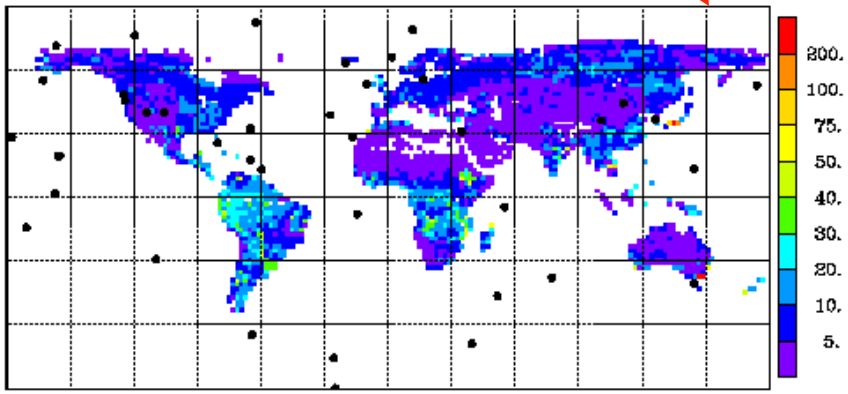
Join Assimilation: "BETHY" CCDAS

Atm. CO_2 at Mauna Loa



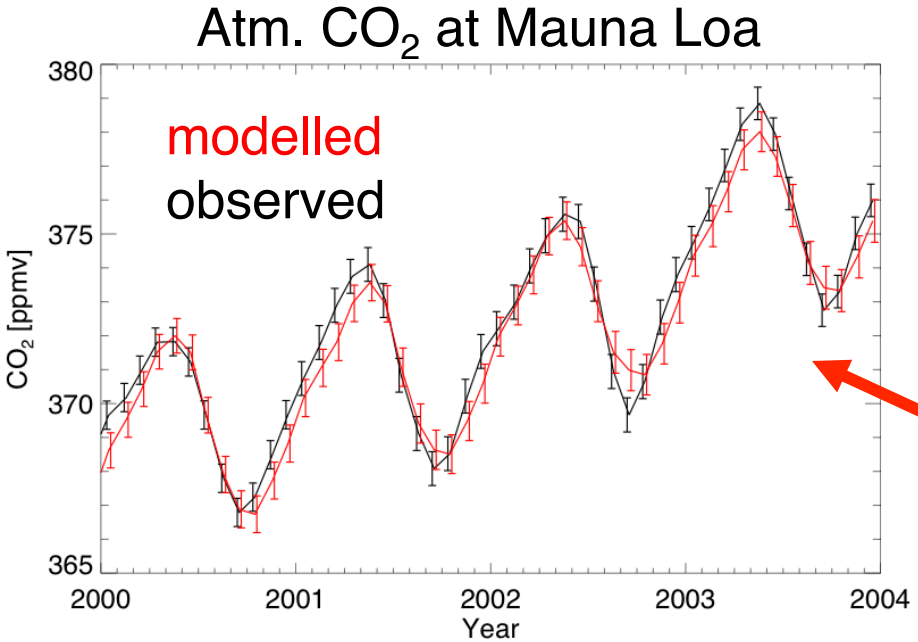


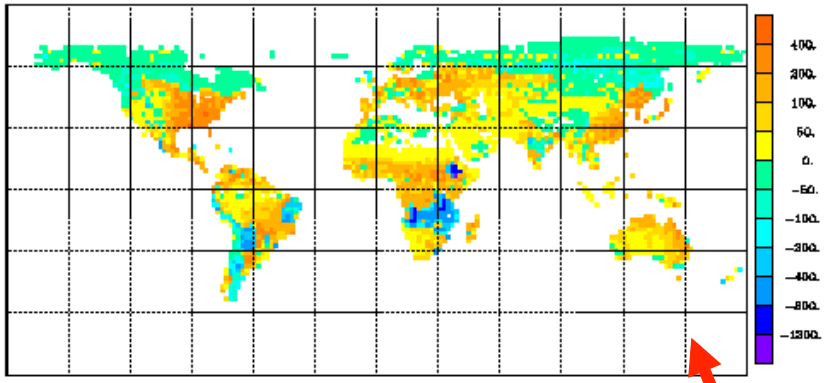
- Predicted atm. CO_2
- Long term mean fluxes to atmosphere ($gC/m^2/year$) and uncertainties



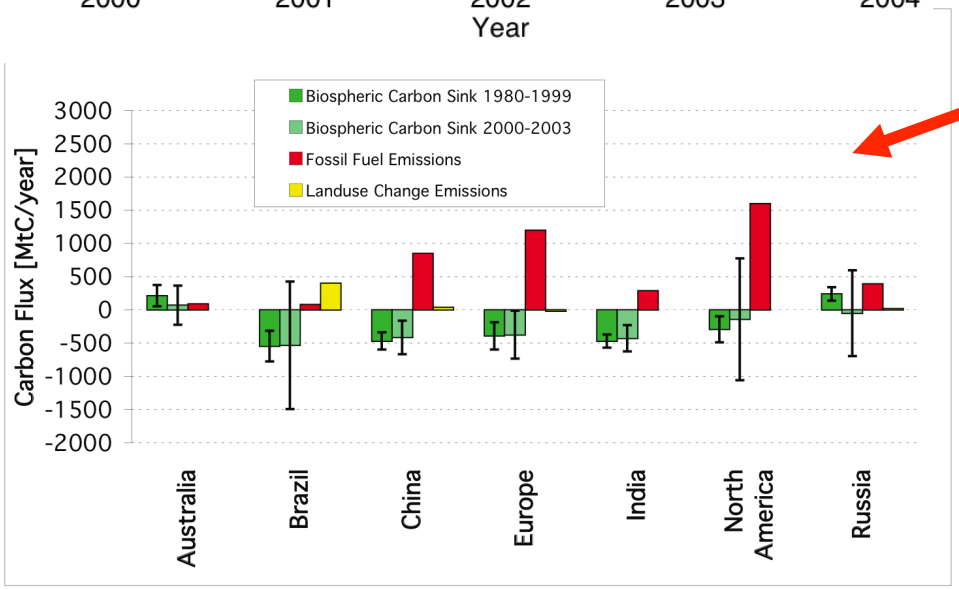
Rayner et al., 2005; Scholze et al., 2007

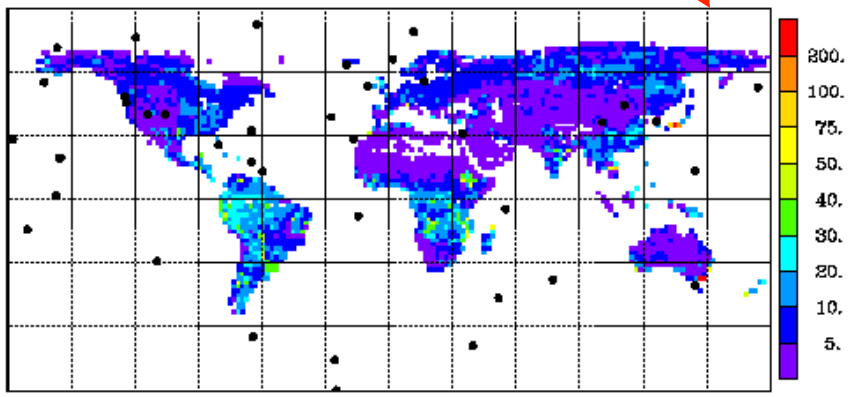
Join Assimilation: "BETHY" CCDAS





- Predicted atm. CO_2
- Long term mean fluxes to atmosphere ($gC/m^2/year$) and uncertainties
- Regional means diag./prog.

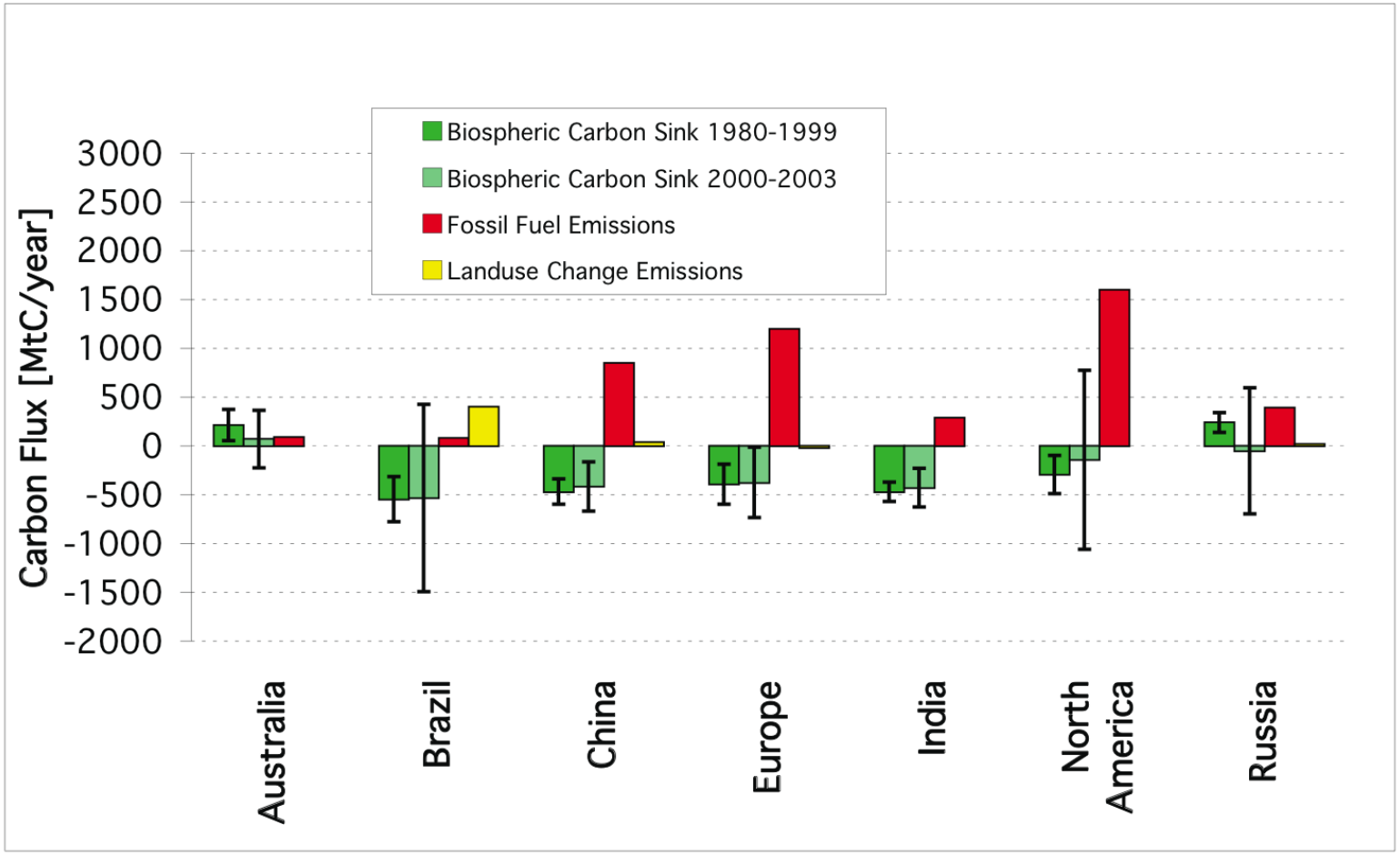




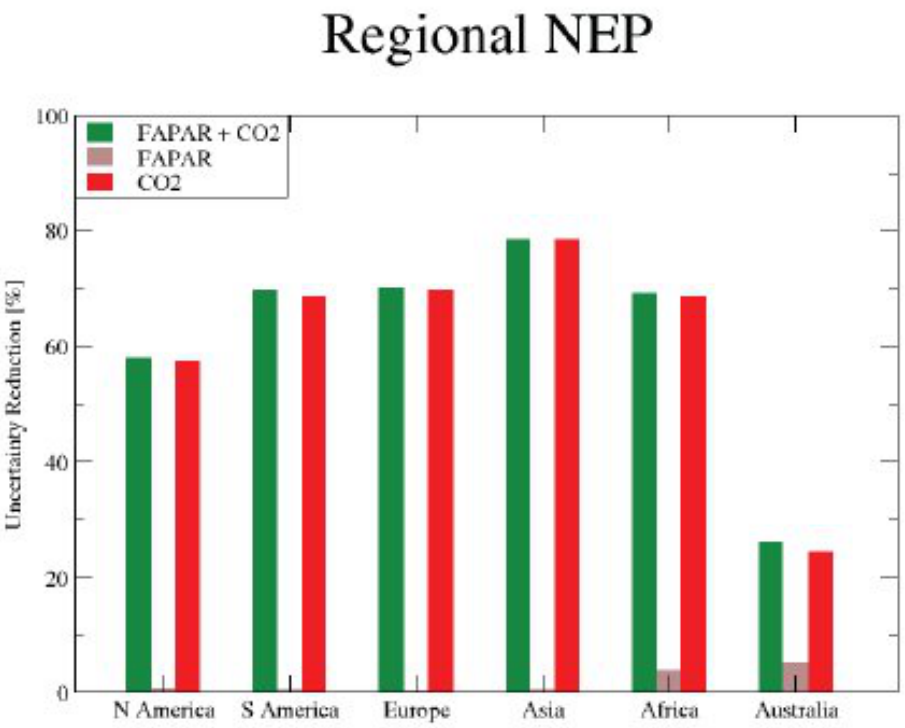
Rayner et al., 2005; Scholze et al., 2007

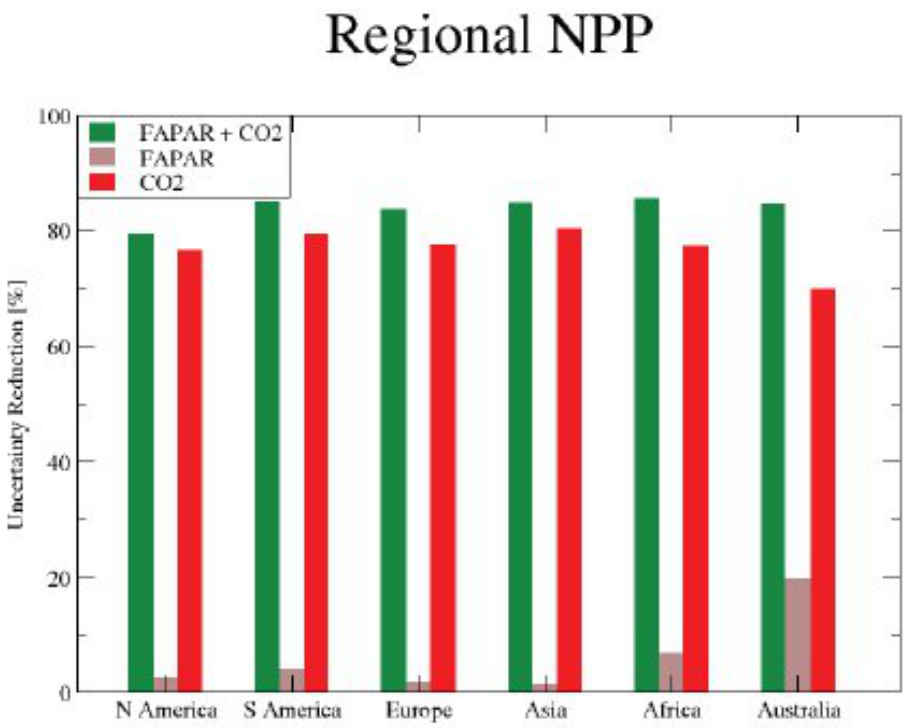
Join Assimilation: "BETHY" CCDAS

- **Regional means diagnostics**



Benefit per data stream: Carbon 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

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

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-linearities might complicate the parameter optimization

➔ Need to :
- keep independent data for model output validation
- Keep classical Atmospheric inversion

CCDAS error assessment

- **Estimation of the parameter errors**

$$P_b' = \left[H^t . R^{-1} . H + P_b^{-1} \right]^{-1}$$

P_b – prior parameter error P_b' – posterior parameter error
R – prior flux error R' – posterior flux error
H – Jacobian matrix

- **Propagation of parameter- errors on fluxes**

$$R' = H . P_b' . H^t$$

- **Evaluation against other products**

Evaluation of Model error: Rationale

$$J(\mathbf{x}) = \frac{1}{2}\left[\underbrace{(\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x}))}_{\text{observations}} + \underbrace{(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b)}_{\text{parameters}}\right]$$

$$\underbrace{\mathbf{R}}_{\text{observation error}} = \underbrace{\mathbf{R}_{\text{meas}}}_{\text{measurement error}} + \underbrace{\mathbf{R}_{\text{model}}}_{\text{inadequate/missing model equations}}$$

- $\mathbf{R}_{\text{meas}}$ from experimentalists
- $\mathbf{R}_{\text{model}}$?

Evaluation of Model error: Principle

Notations

H: Observation operator (model)	$\mathbf{H}$: Linearized H
$\mathbf{x}$: State vector	$\mathbf{x}_b$: Background state (prior parameters)
$\mathbf{y}$: Vector of observation data	$\mathbf{d}_o^b = \mathbf{y} - H(\mathbf{x}_b)$: Prior residuals

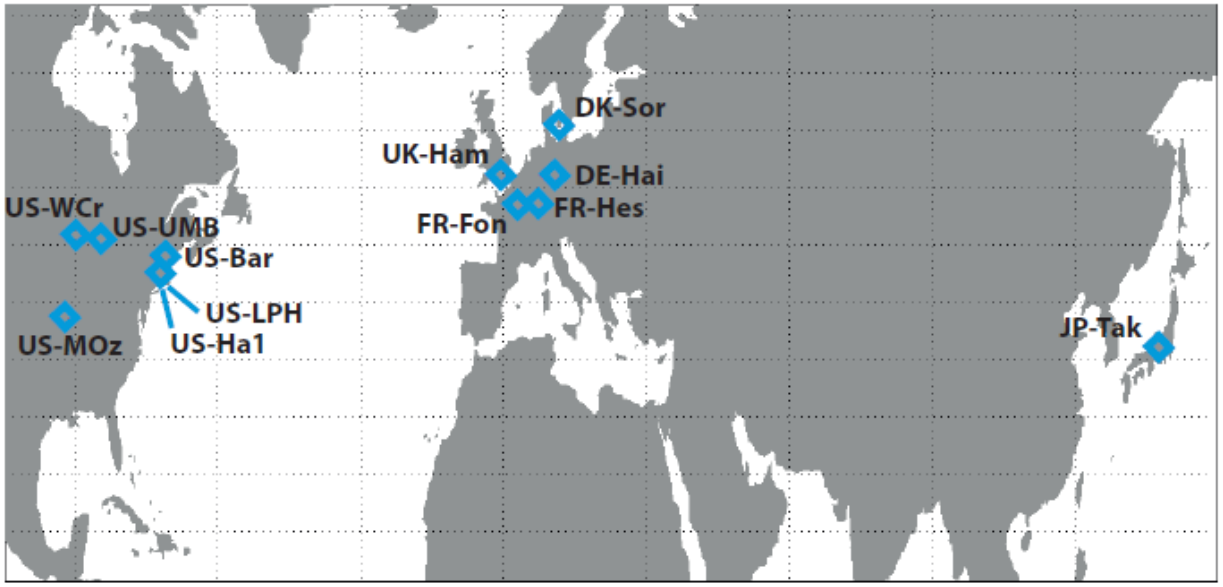
Diagnostics in observation space

$$\underbrace{E\left[\mathbf{d}_o^b\left(\mathbf{d}_o^b\right)^T\right]}_{\text{Cross-product of prior residuals}} = \underbrace{\mathbf{HBH}^T}_{\text{Background error}} + \underbrace{\mathbf{R}}_{\text{Observation error}}$$

(Desroziers et al., 2005)

Daily NEE : diagnosis set-up

Temperate deciduous broadleaf forest sites



Parameters
$V_{cmax,opt}$, $c_{T,min/opt,max}$, $L_{age,crit}$, $f_{stressh}$
$G_{s,slope}$
LAI_{MAX}, SLA
$K_{lai,alloc}$
$K_{phenocrit}$, c_{senes}
Hum_{cste}, Dpu_{cste}
MR_a, MR_b, GR_{frac}
K_{soilC}, Q_{10}, HR_a, HR_b, HR_c, HR_{min}, Z_{decomp}
$K_{albedo,veg}$, $Z0_{overheight}$

Prescribed background error

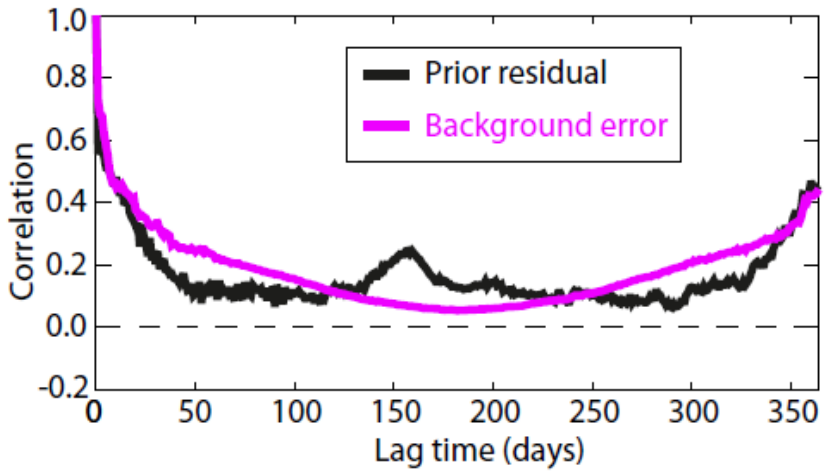
- Diagonal **B**
- $\sigma_b = f\,(\min_{par}, \max_{par})$

Hypothesis

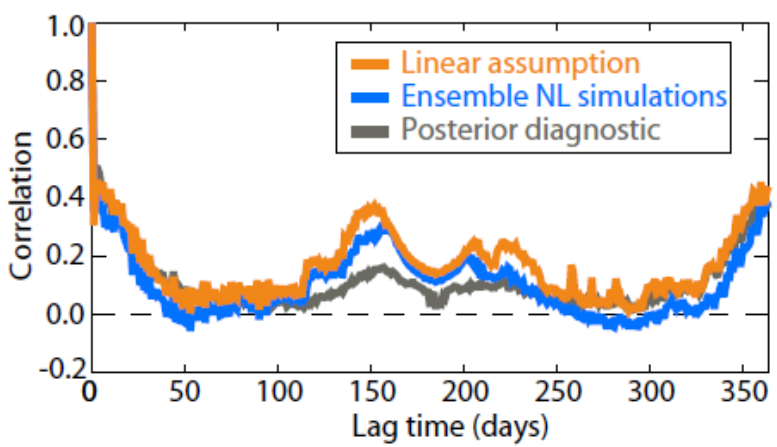
Errors stationary in time

Temporal error structure

$E\left[\mathbf{d}_o^b\left(\mathbf{d}_o^b\right)^T\right]$ and $\mathbf{HBH}^T$



$\mathbf{R} = E\left[\mathbf{d}_o^b\left(\mathbf{d}_o^b\right)^T\right] - \mathbf{HBH}^T$



Median prior residuals

- SD = 2.1 $gC.m^{-2}.d^{-1}$
- Seasonal structure
- Bump due to overestimated GSL ?

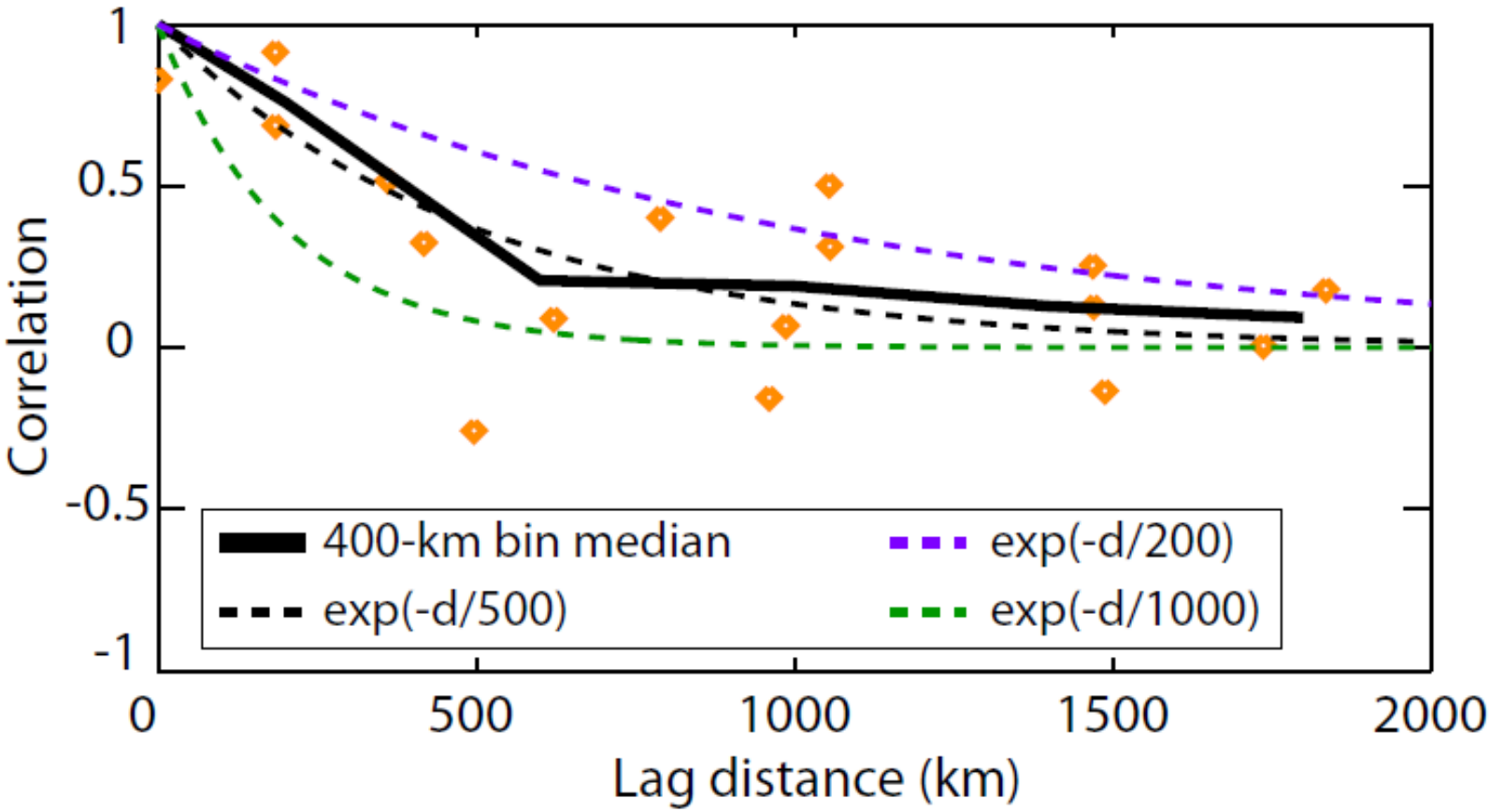
Median background error

- SD = 1.3 $gC.m^{-2}.d^{-1}$
- Smooth seasonal structure

Median observation error

- SD = 1.7 $gC.m^{-2}.d^{-1}$
- Rapid decrease after 1 day
- Robust without linearity near $\mathbf{x}_b$
- Consistent with posterior diagnostics

Spatial structure of R



- Nearly exponential median decay
- e-folding length of ~500 km

The model error

$$\mathbf{R}_{\text{model}} = \mathbf{R} - \mathbf{R}_{\text{meas}}$$

$$\mathbf{R}_{\text{meas}} \begin{cases} 0.2 \leq \sigma_{\text{meas}} \leq 0.8 \text{ gC.m}^2\text{.d}^{-1} \text{(Richardson et al., 2008)} \\ \text{No significant spatial error structure} \\ \text{Negligible daily error autocorrelation (Lasslop et al., 2008)} \end{cases}$$

$$\Rightarrow \mathbf{R}_{\text{model}} \begin{cases} \sigma_{\text{model}} \approx 1.6 \text{ gC.m}^2\text{.d}^{-1} \\ \text{similar to } \mathbf{R} \text{ in structure} \end{cases}$$

NEE error budget in other ecosystems

	Observation error (measure + model Structure)	Prior-parameter error
Tropical EBF	2.78	1.23
Temperate ENF	1.28	0.81
Temperate EBF	1.45	1.39
Temperate DBF	1.65	1.42
Boreal ENF	0.70	0.61
Boreal DBF	1.35	1.17
C3 grasslands	1.36	1.01

For daily averaged fluxes (gC/m²/d)