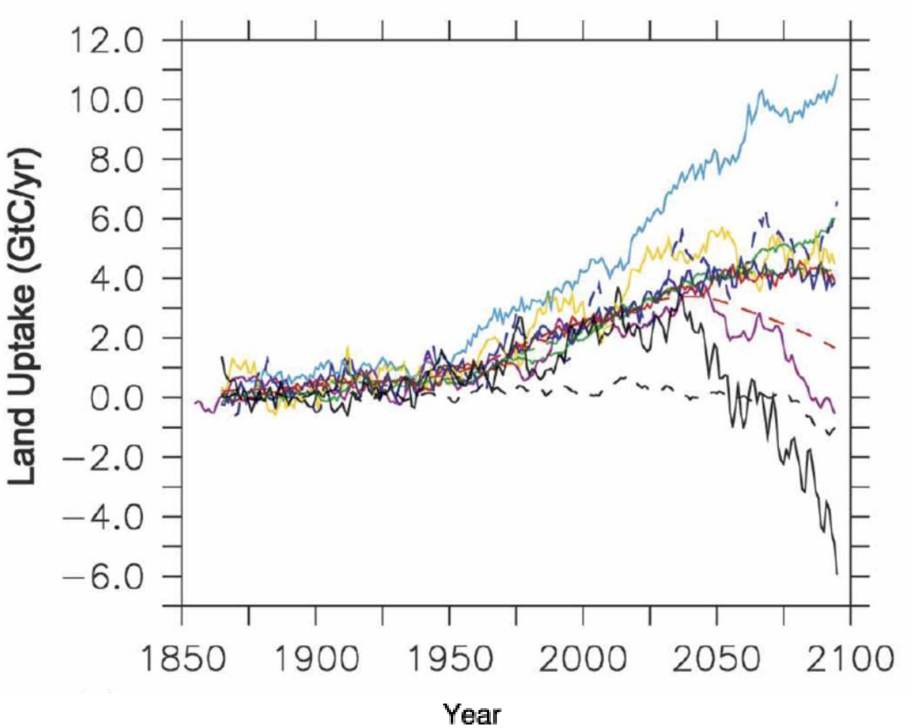
From observations to prediction through model-data integration: the importance of multiple constraints

Markus Reichstein, Nuno Carvalhais, Gregor Schürmann,
Thomas Wutzler, Matthias Forkel, Soenke Zaehle

Max-Planck Institute for Biogeochemistry, Jena
Department of Biogeochemical Integration

Motivation: carbon cycle uncertainty



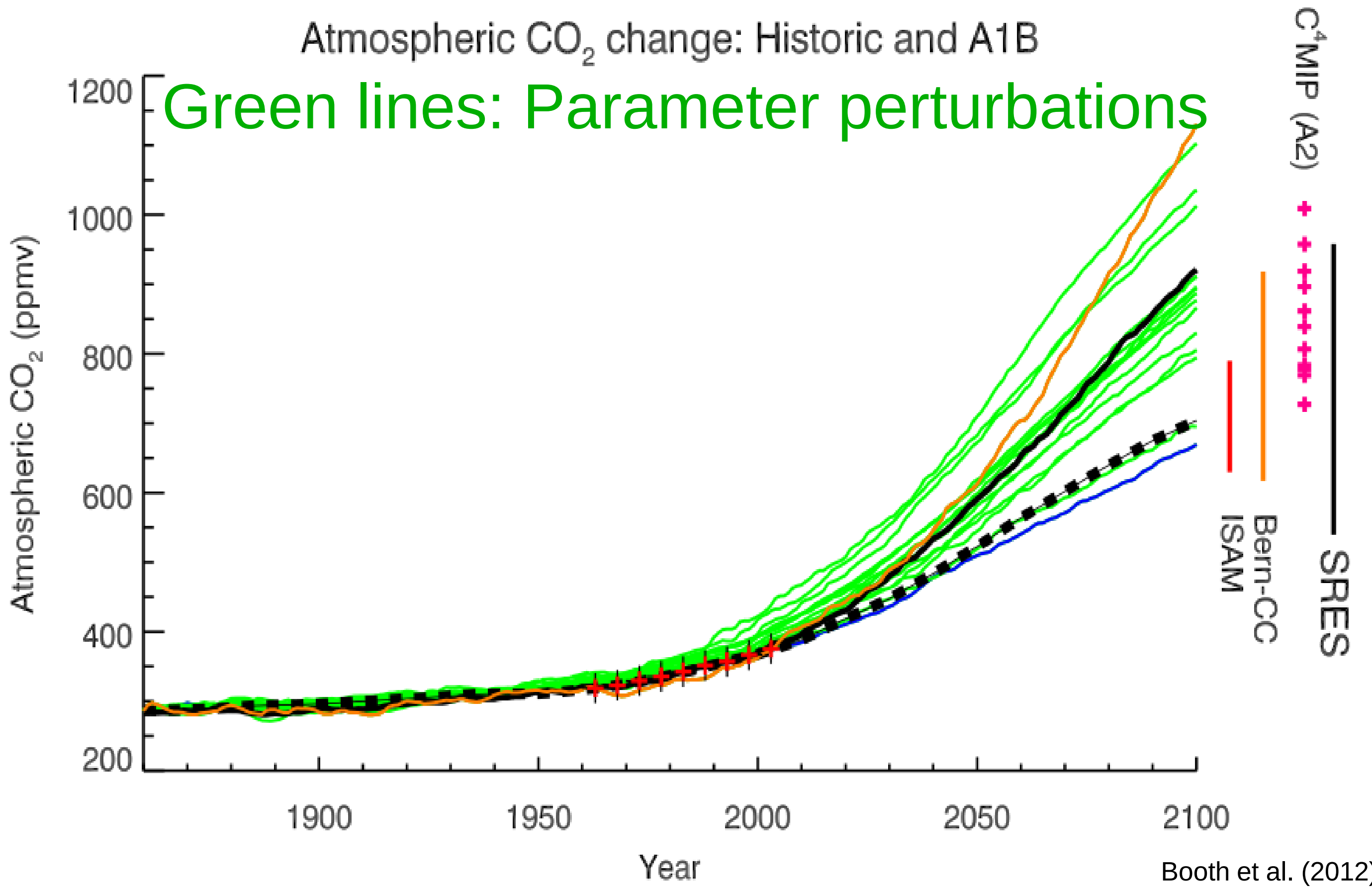
Friedlingstein et al. (2006)

~1 Pbyte/year
(~200 Billion bibles [2e8] or 50 Lib of Congress)

Potential sources of uncertainty

- model structure
  - missing processes
  - misrepresentation of states dynamics
- parameterizations
  - wrong sensitivities
- initial conditions
  - inappropriate characterization of ecosystem states

Previous tests on parameter uncertainties



Booth et al. (2012)

Model spread caused by parameters of the terrestrial component

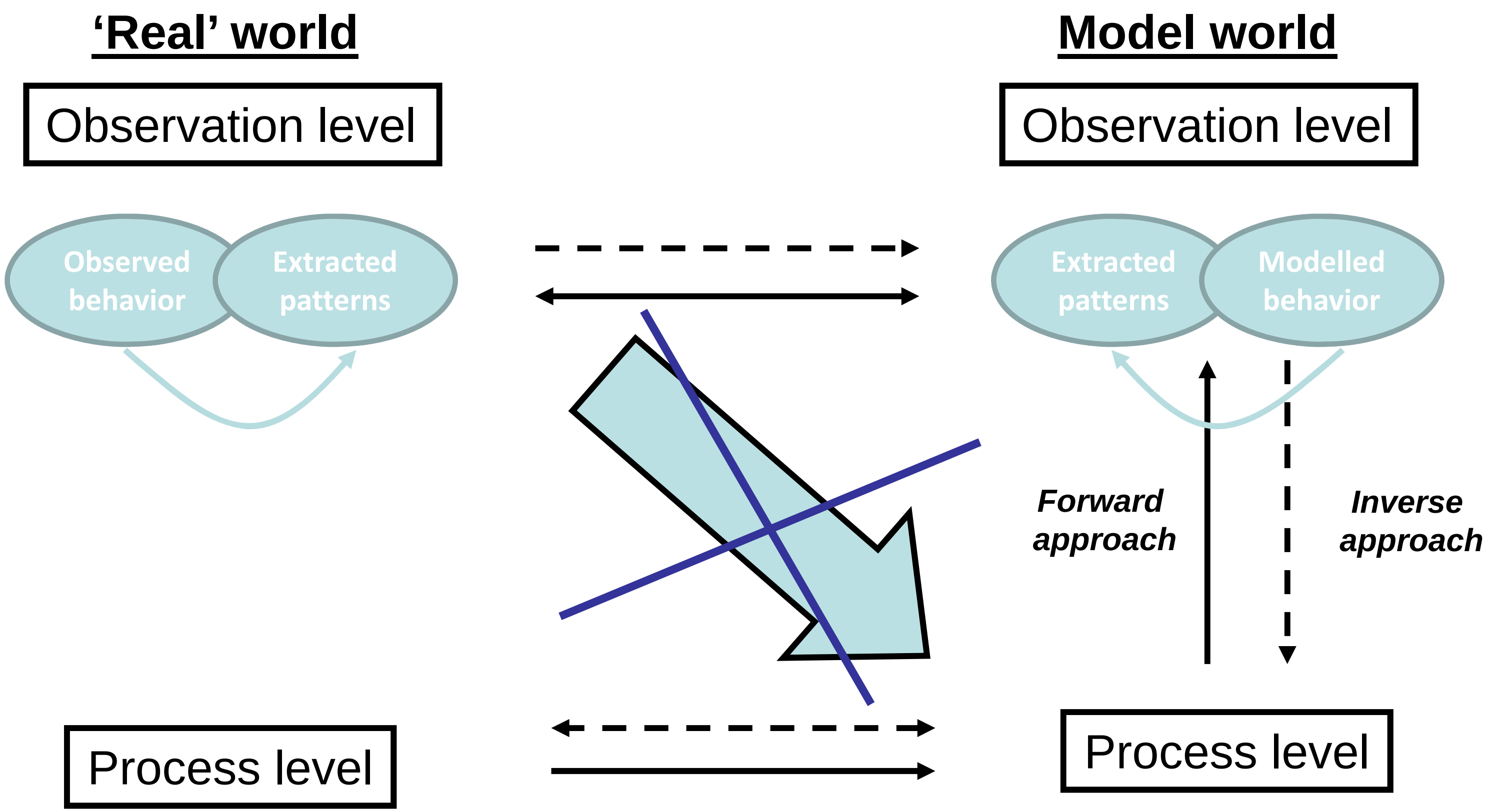
Highly parameterized formulations

Potential sources of uncertainty

- model structure
  - missing processes
  - misrepresentation of states dynamics
- parameterizations
  - wrong sensitivities
- initial conditions
  - inappropriate characterization of ecosystem states

Explore the information content of observations from ecosystem to regional/global scales

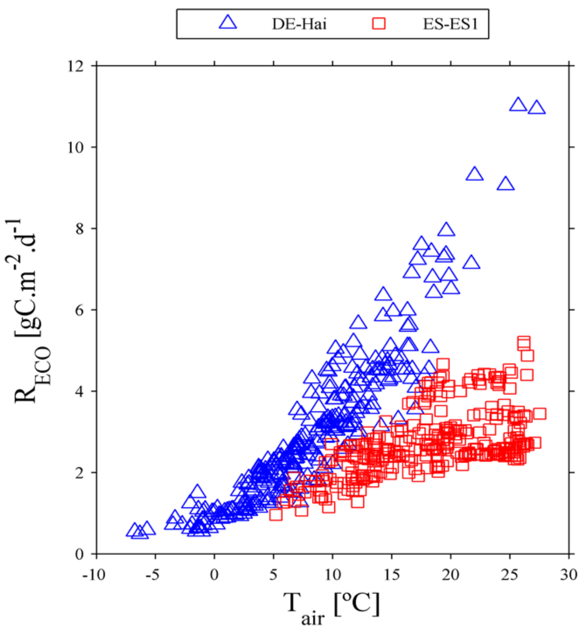
Linking models and observations



cf. Reichstein & Beer (2008), JPNSS

Linking models and observations

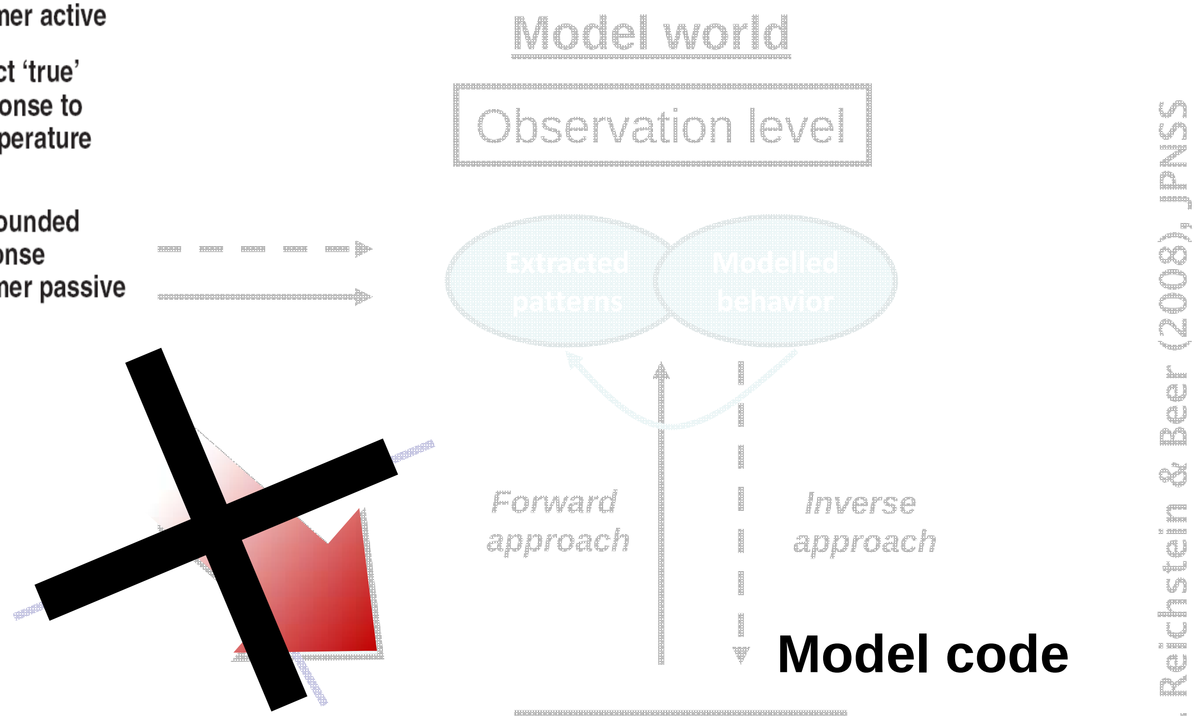
Reichstein et al. 2005



Confounded response summer active

Direct 'true' response to temperature

Confounded response summer passive

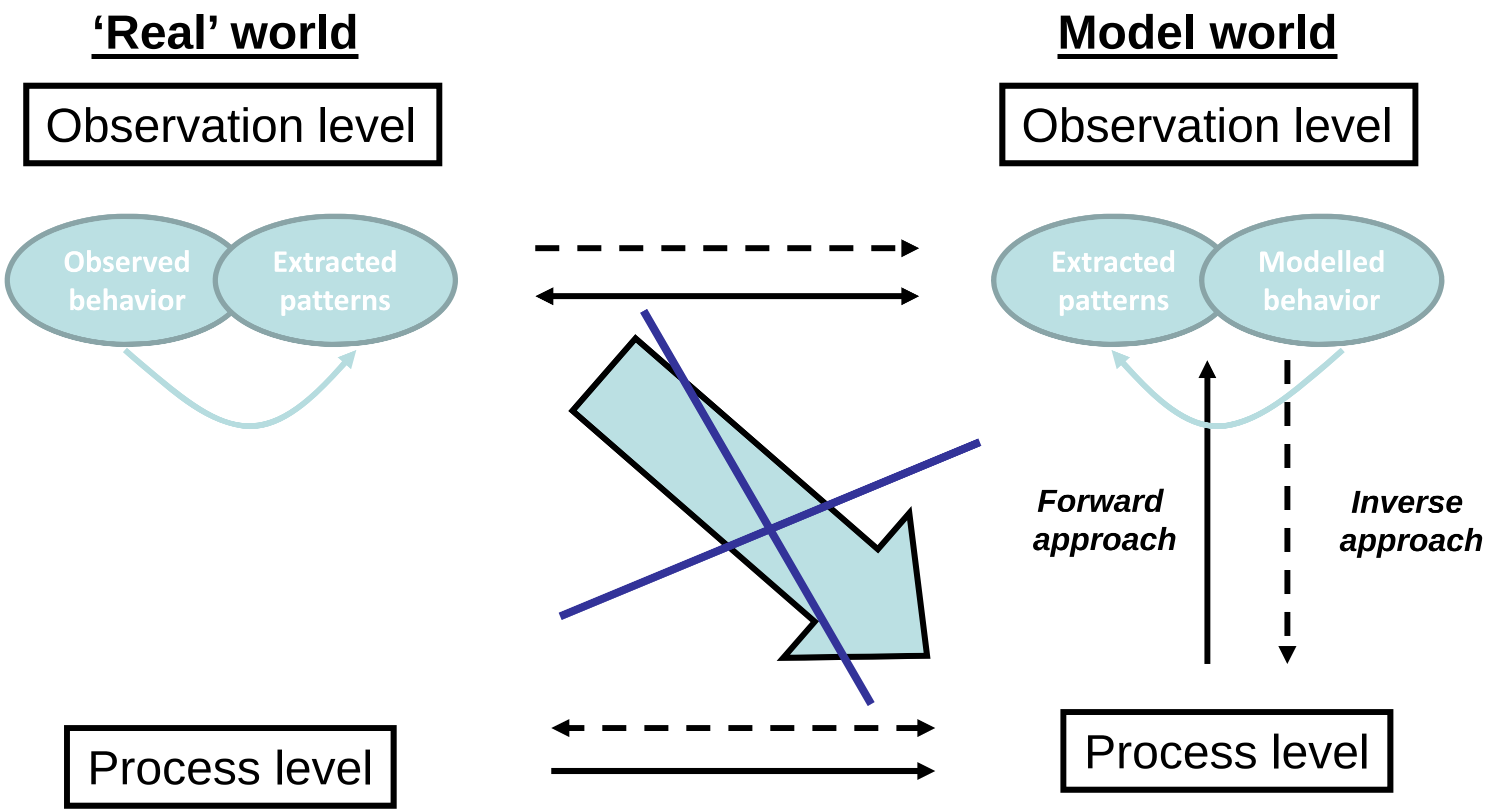


Process level

Model code

```
if (tsoil < -10.0)
{
     /* no decomp processes for tsoil < -10.0 C */
     t_scalar = 0.0;
}
else
{
     tk = tsoil + 273.15;
     t_scalar = exp(308.56*((1.0/71.02)-(1.0/(tk-227.13))));
}
```

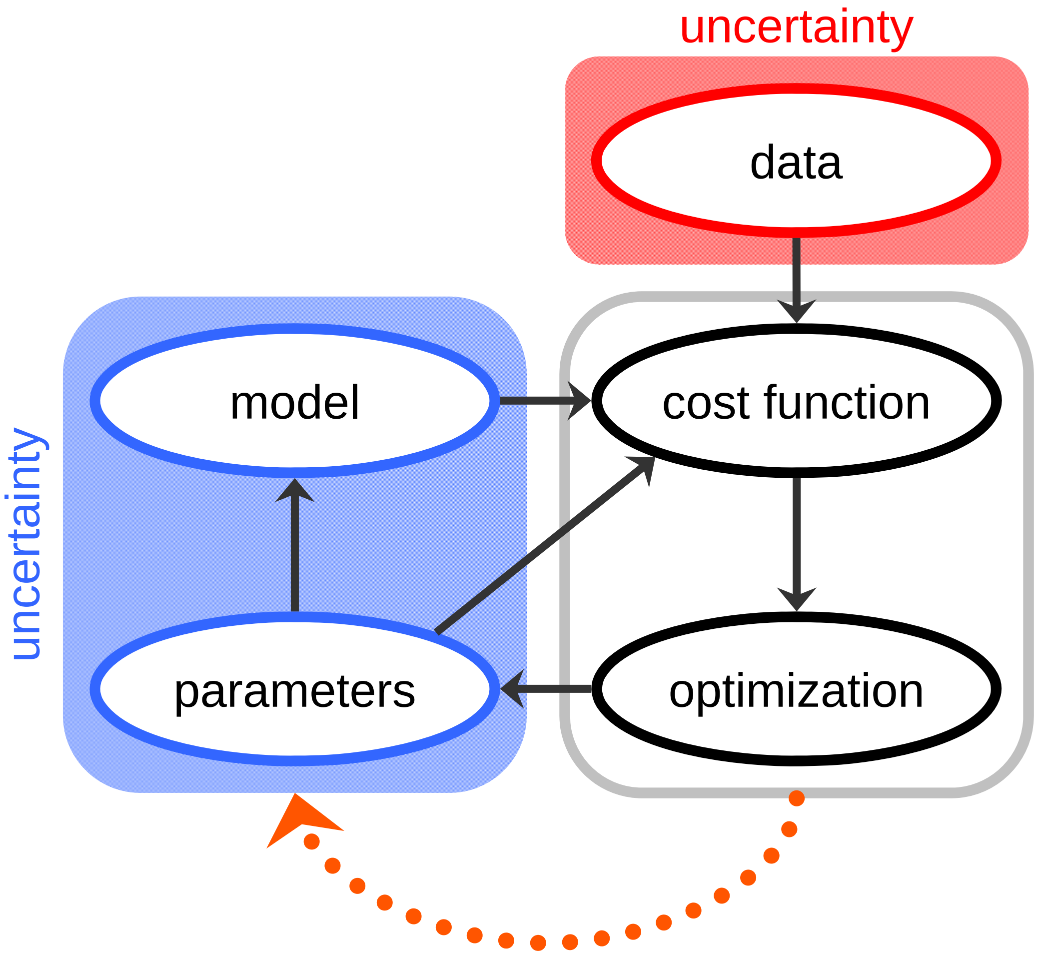
Linking models and observations



cf. Reichstein & Beer (2008), JPNSS

Challenge: Extract generalized information from data sets, confront these with model behavior and interpret differences in a system-oriented way

Inversion methods



Adapted from Lasslop 2010

Model-data integration

- Cost functions
  - Single versus multiple constraints approaches

$$\Omega = \frac{1}{2}\sum_{i=1}^{M}\frac{1}{N_i}\left\{\sum_{n=1}^{N_i}\frac{\left[\hat{y}_{i,n}(x,\mathbf{p}) - y_{i,n}\right]^2}{\sigma_{i,n}^2}\right\} + \frac{1}{2}\sum_{k=1}^{Q}\frac{\left(\hat{p}_k - p_k\right)^2}{\sigma_{p,k}^2}$$

Model-data integration

- Cost functions
  - Single versus multiple constraints approaches

data streams model estimates observations a priori parameter

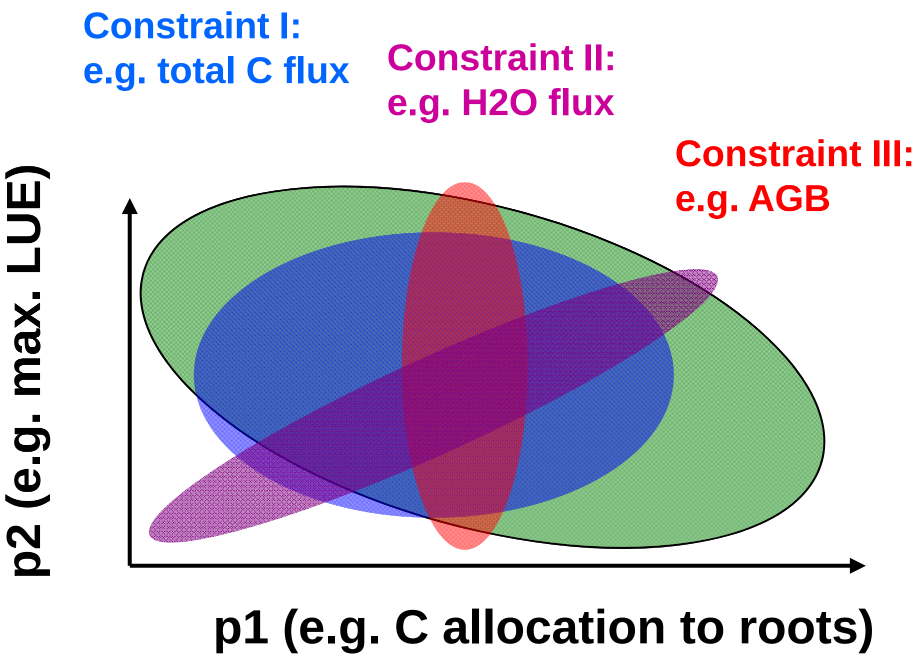
$$\Omega = \frac{1}{2}\sum_{i=1}^{M}\frac{1}{N_i}\left\{\sum_{n=1}^{N_i}\frac{\left[\hat{y}_{i,n}(x,\mathbf{p}) - y_{i,n}\right]^2}{\sigma_{i,n}^2}\right\} + \frac{1}{2}\sum_{k=1}^{Q}\frac{\left(\hat{p}_k - p_k\right)^2}{\sigma_{p,k}^2}$$

obs. uncertainty par. uncertainty

parameter vector proposed parameter

Multiple-constraint model identification & parameter estimation…

Challenges: Equifinality, Over-parameterisation
(e.g. Knorr et al. 2005, Reichstein et al. 2005)



- Bayesian model calibration against different constraints
- Extension of the approach to test different model structures and process representations (Model identification)

Addressing present and future variability in ecosystem carbon fluxes through modeling ensembles and model-data fusion approaches

INFUSION

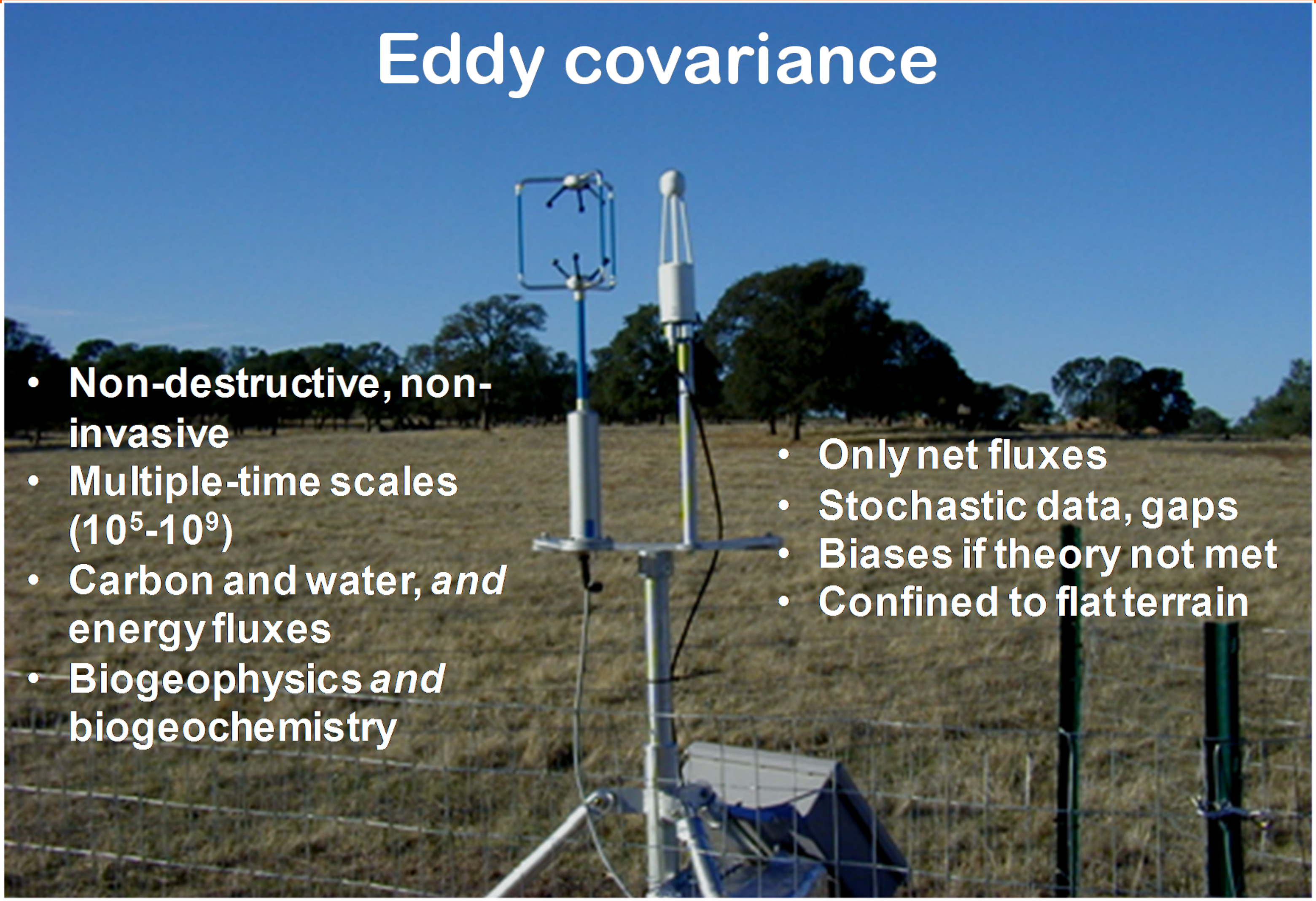
Nuno Carvalhais, Marcel van Oijen, Trevor Keenan, Natasha MacBean, Philippe Peylin, Anja Rammig, Susanne Rolinski, Tea Thum, André Granier, Dennis Loustau, Gregor Schuermann, Soenke Zaehle, Christian Beer, Miguel Mahecha, Jakob Zscheischler, Andrew Richardson and Markus Reichstein

models

- BASFOR [CEH]
- FoBAAR [Harvard]
- JSBACH [MPI BGC]
- LPJ [PIK]
- ORCHIDEE [LSCE]

Hesse: deciduous broadleaf forest, beech; Cfb - Warm temperate fully humid with warm summer

Quantifying ecosystem-atmosphere interactions

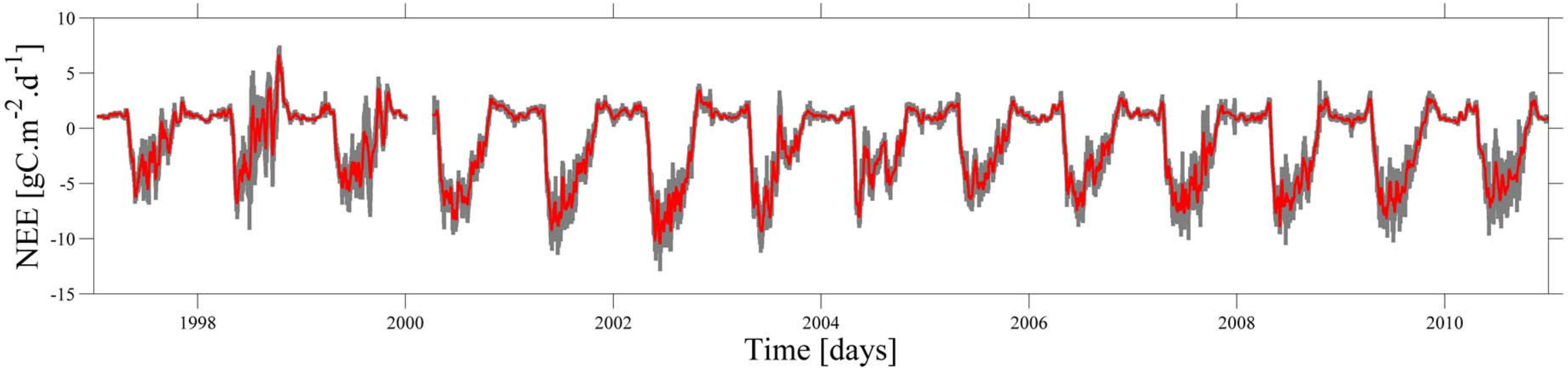


Data & uncertainties

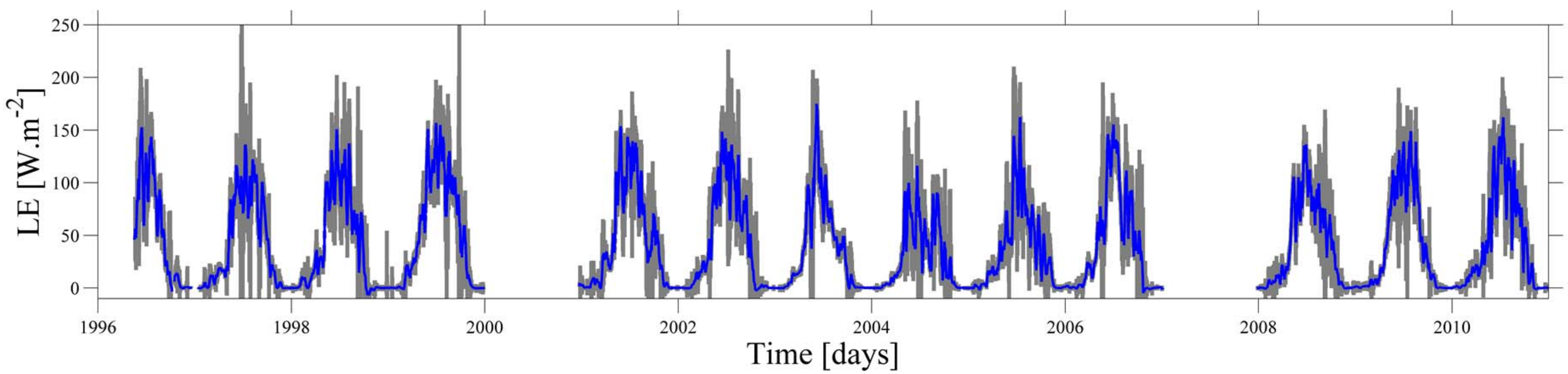
- Eddy covariance fluxes
  - Net Ecosystem Exchange (NEE)
    - random and *u** thresholds
  - Latent heat fluxes (LE)
    - random and EBC method

Data : flux measurements

NEE and NEE uncertainties



LE and LE uncertainties

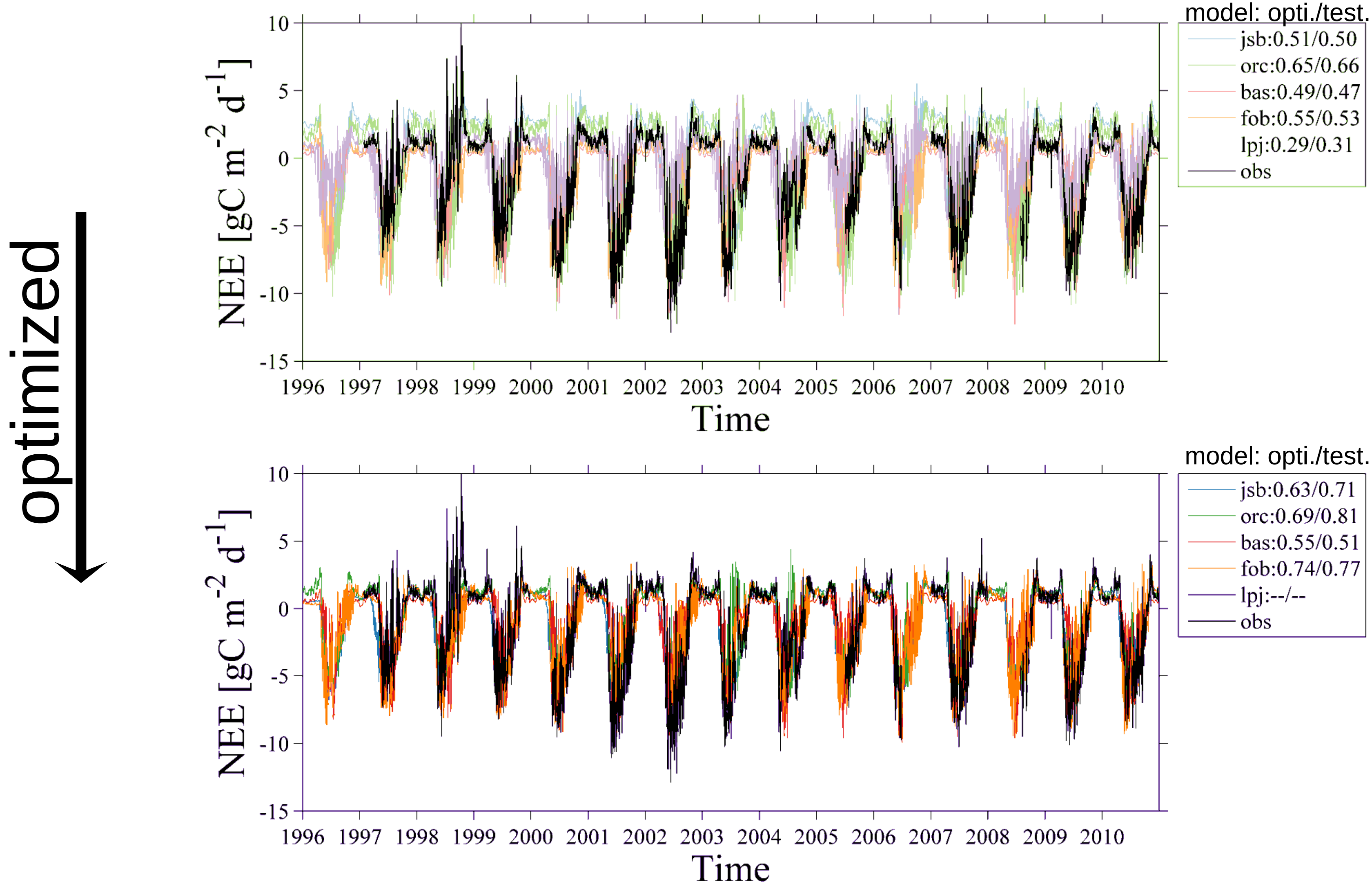


Data & uncertainties

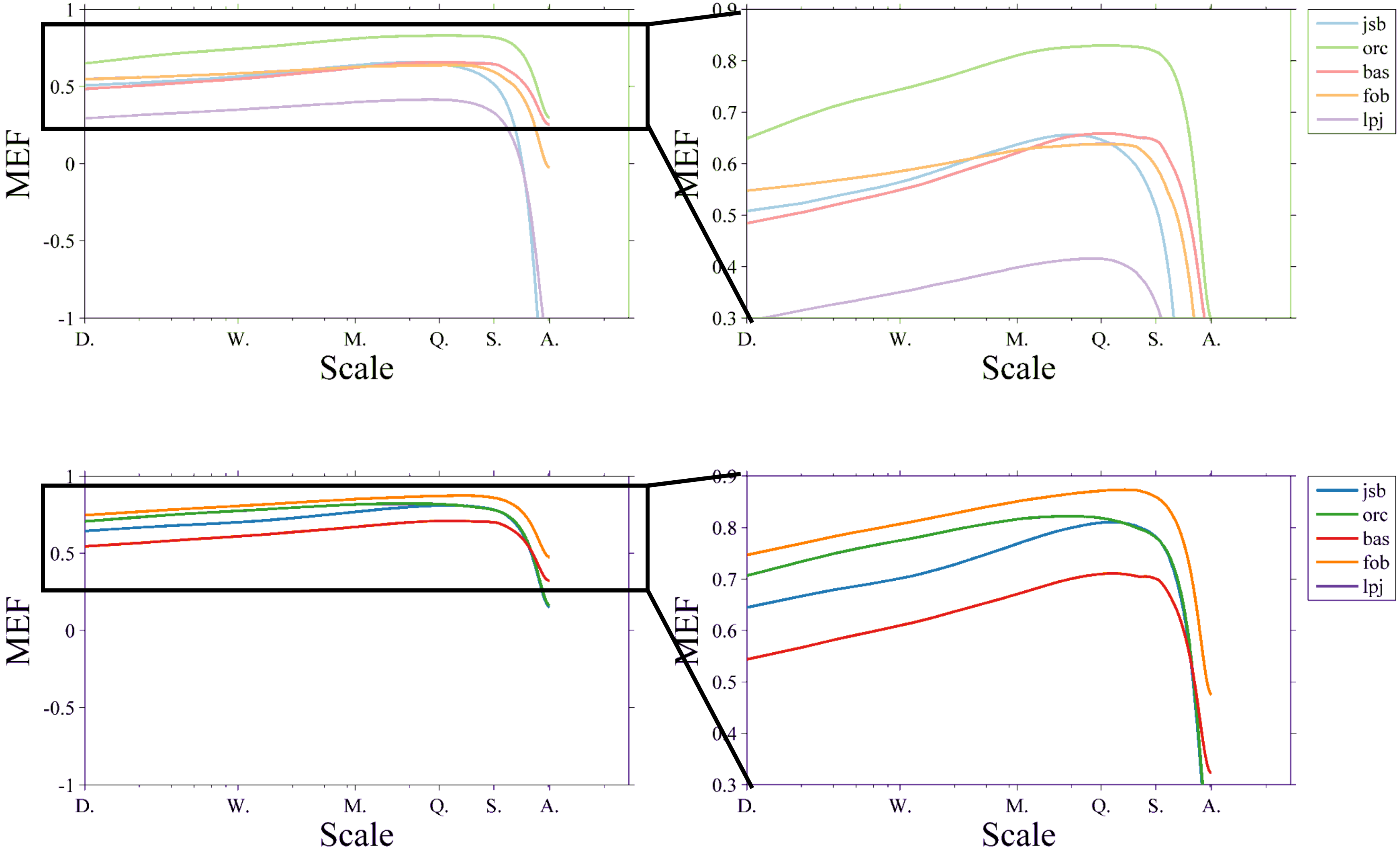
- Eddy covariance fluxes
  - Net Ecosystem Exchange (NEE)
    - random and *u** thresholds
  - Latent heat fluxes (LE)
    - random and EBC method
- Ancillary biometric data
  - AGB and AGB increments
    - natural variability, observational and parametric uncertainties in DBH curves
  - Total soil carbon stocks
    - Spatial variability and total profile representation

formal consideration of uncertainties in model-data integration

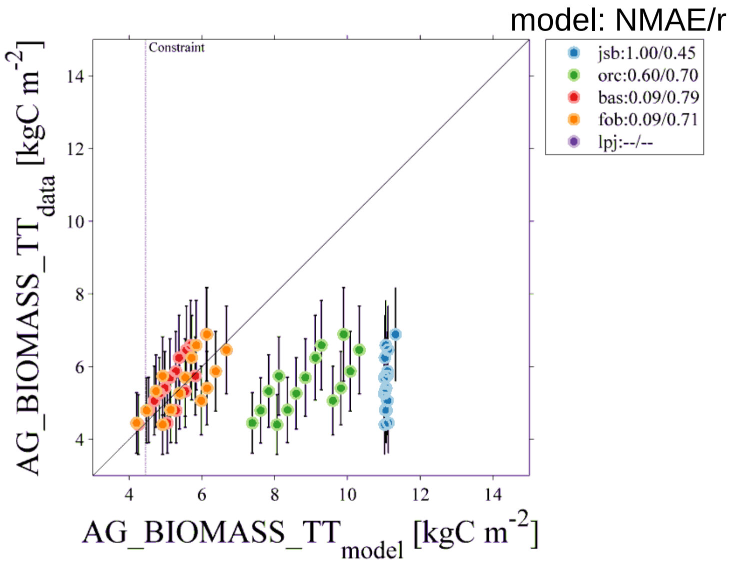
Multi-model MDF : Hesse : NEE

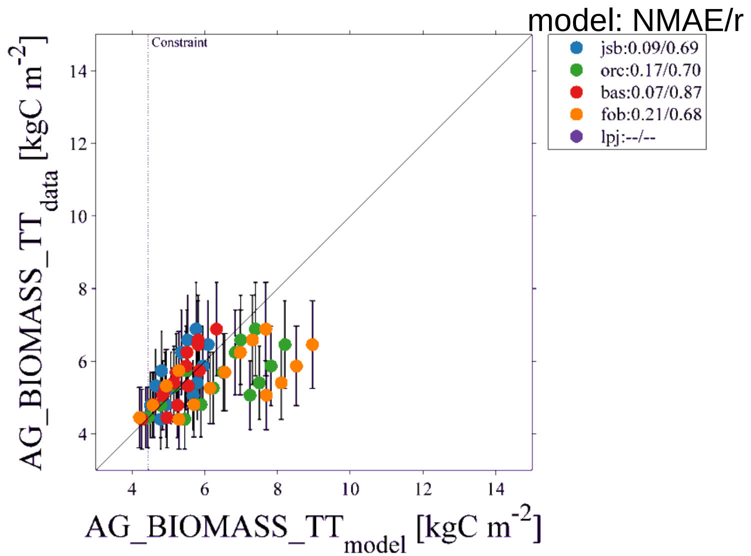


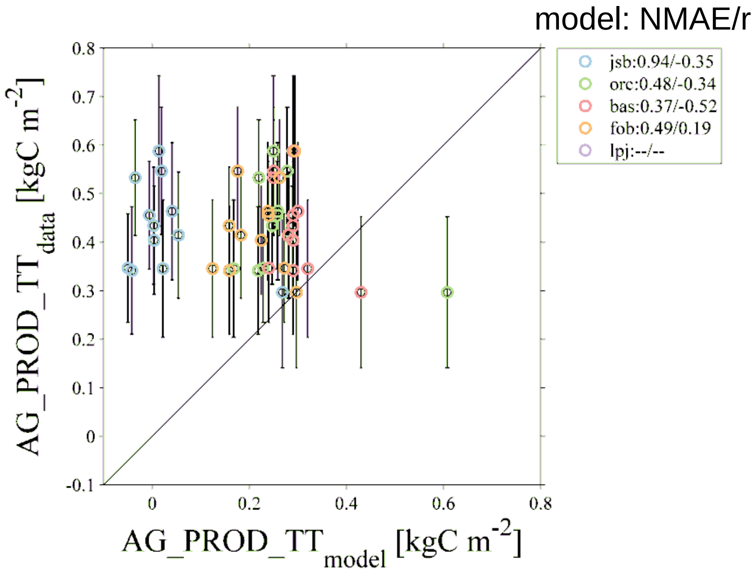
Hesse : misfits vs time scales

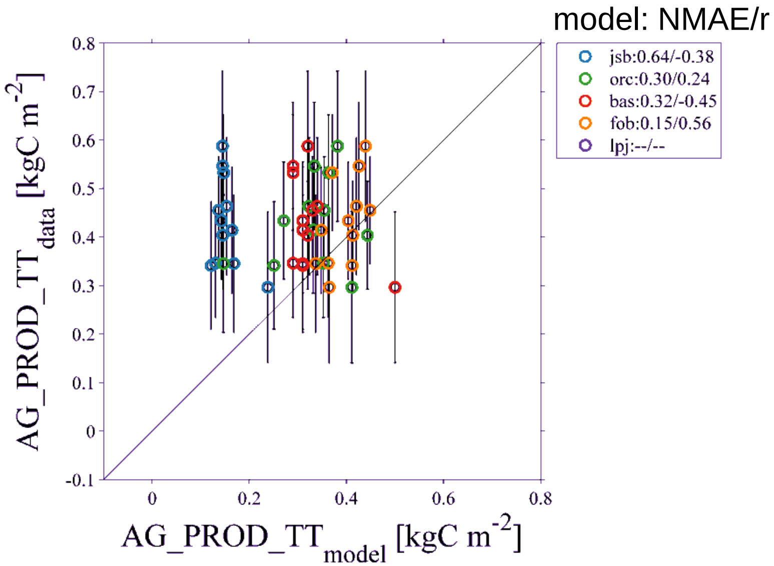


mdf : Hesse : vegetation stocks

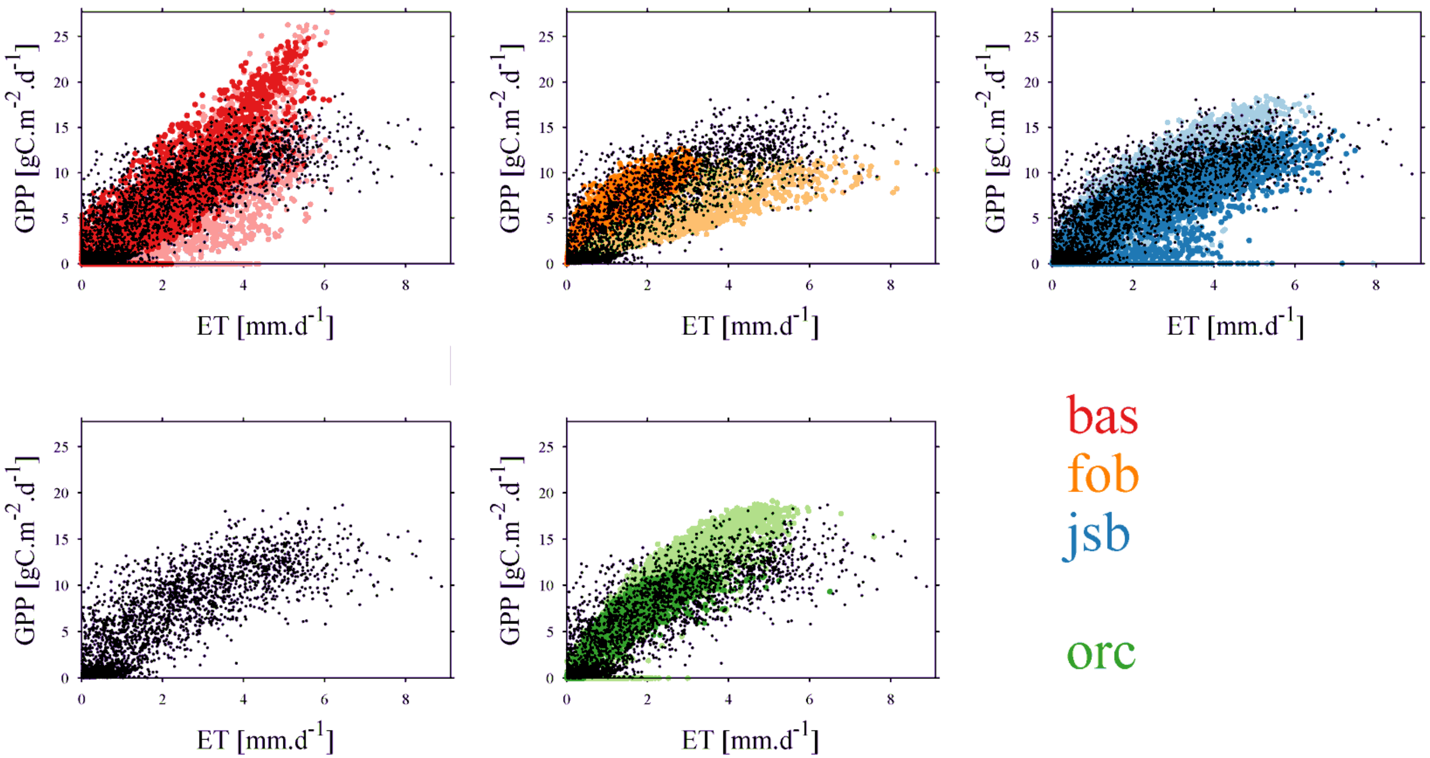




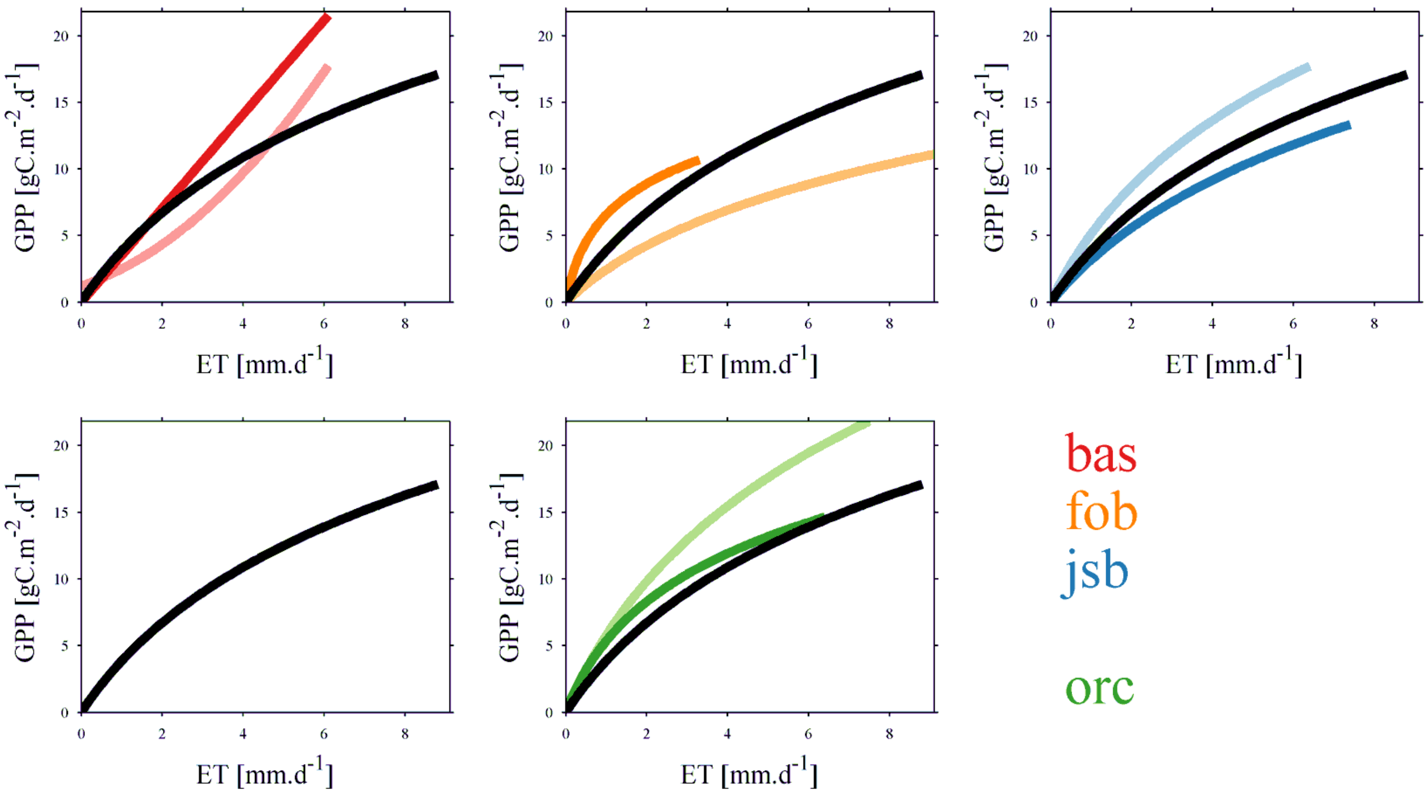




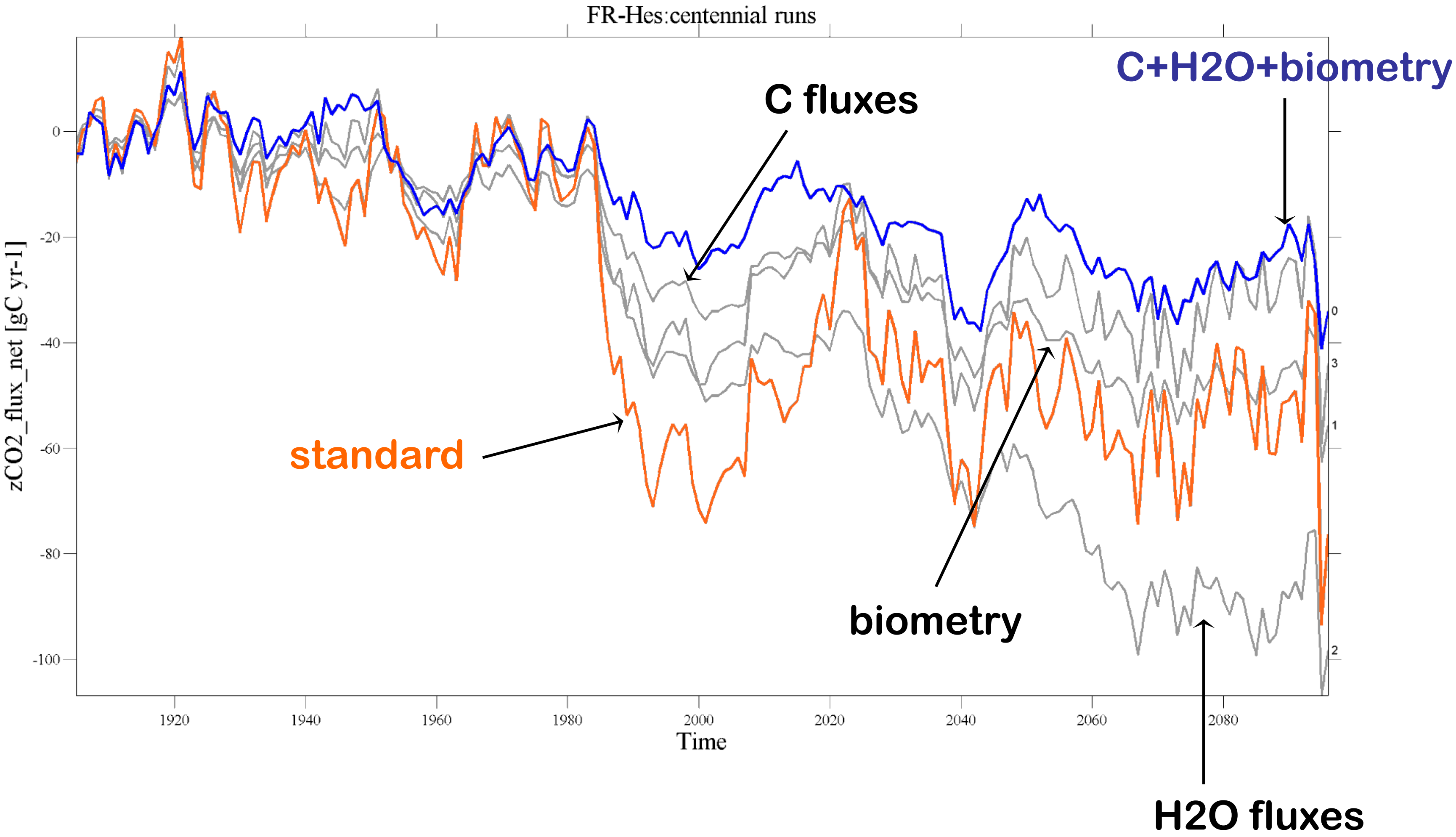
Description of water use efficiency



Description of water use efficiency



JSBACH : implications of multiple constraints

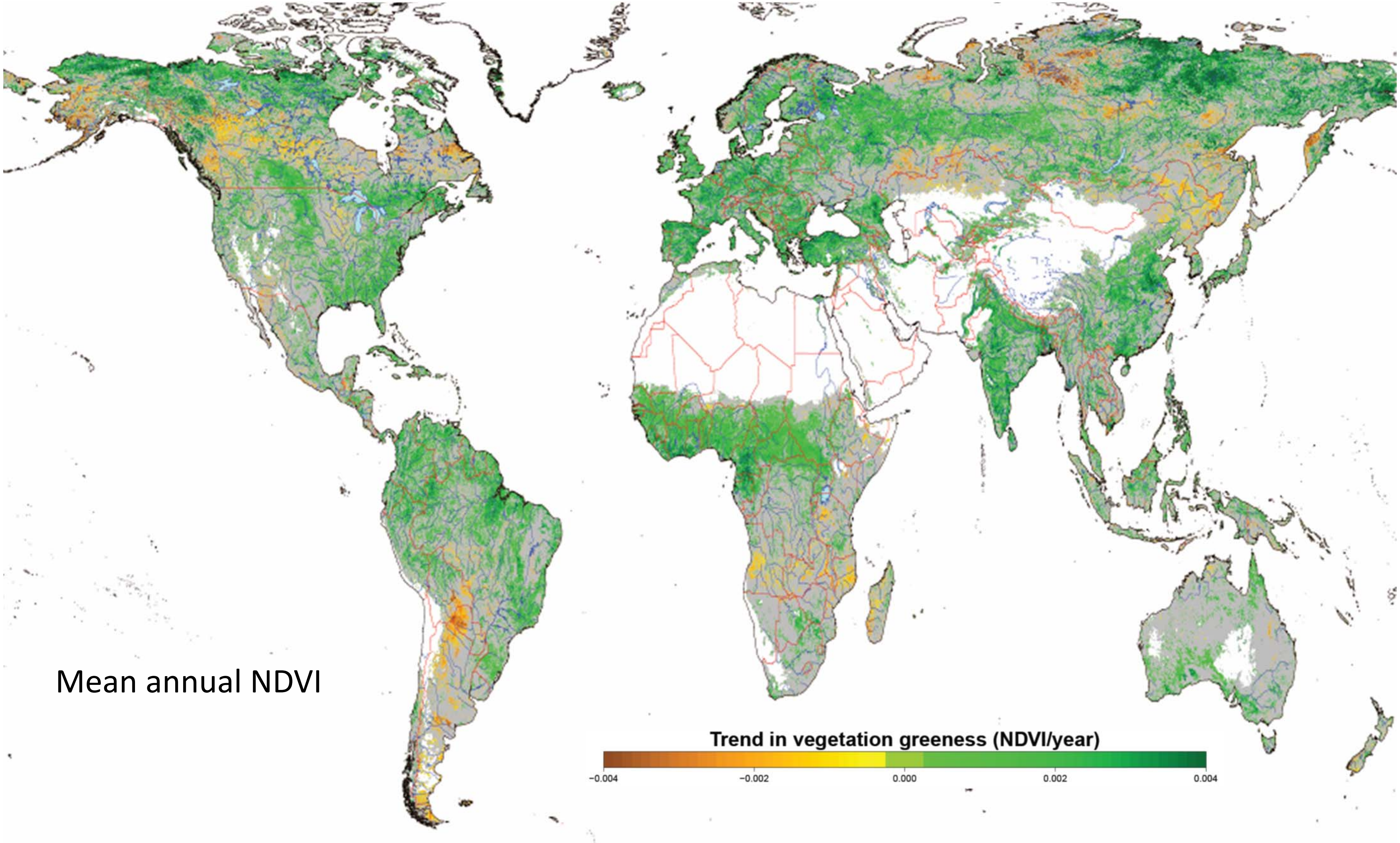


Exploring seasonal and decadal dynamics of phenology

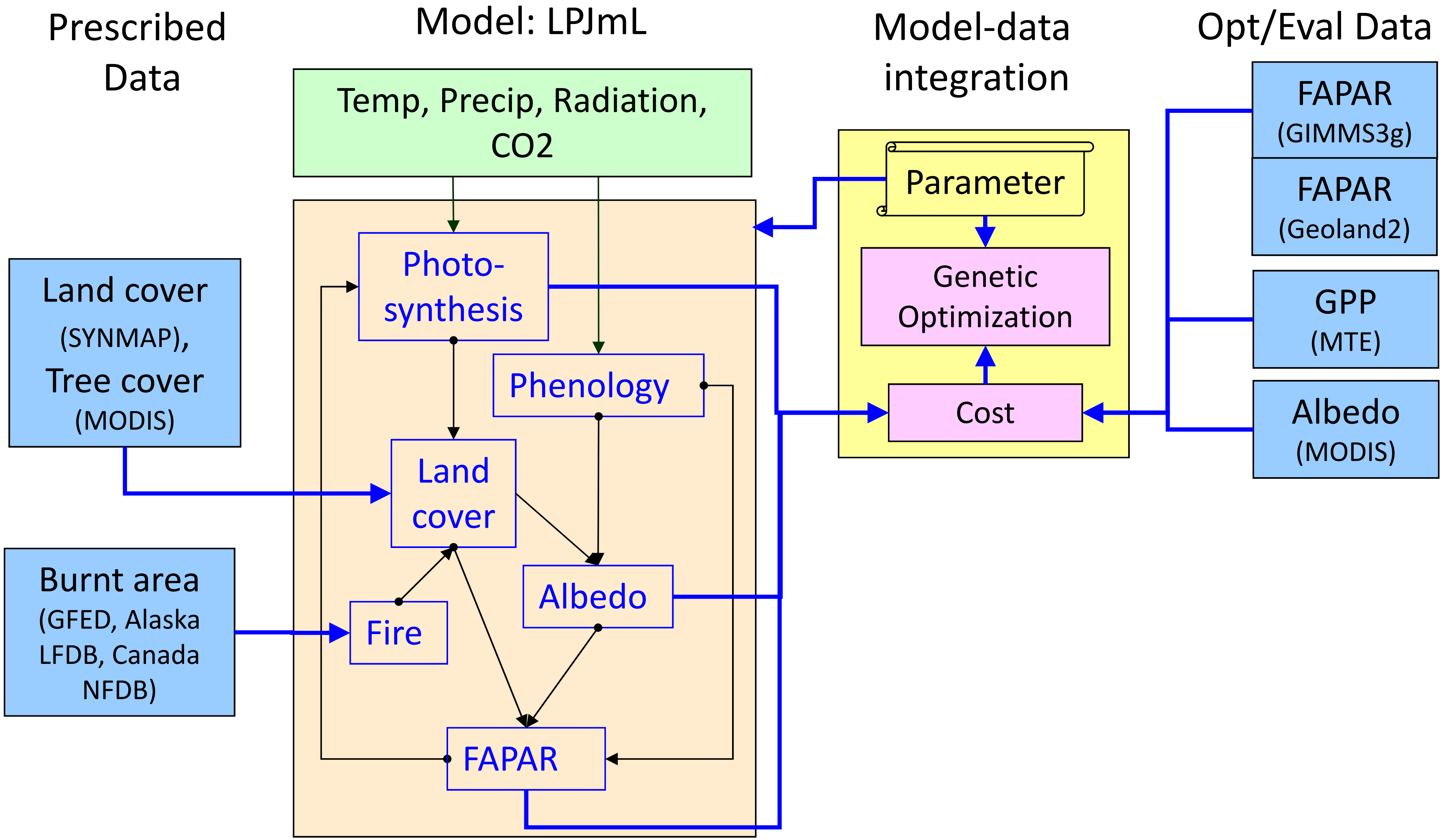
LPJML-MDI

Forkel et al., BGD, 2014; Forkel et al, in prep.

Trends in vegetation greenness 1982-2011

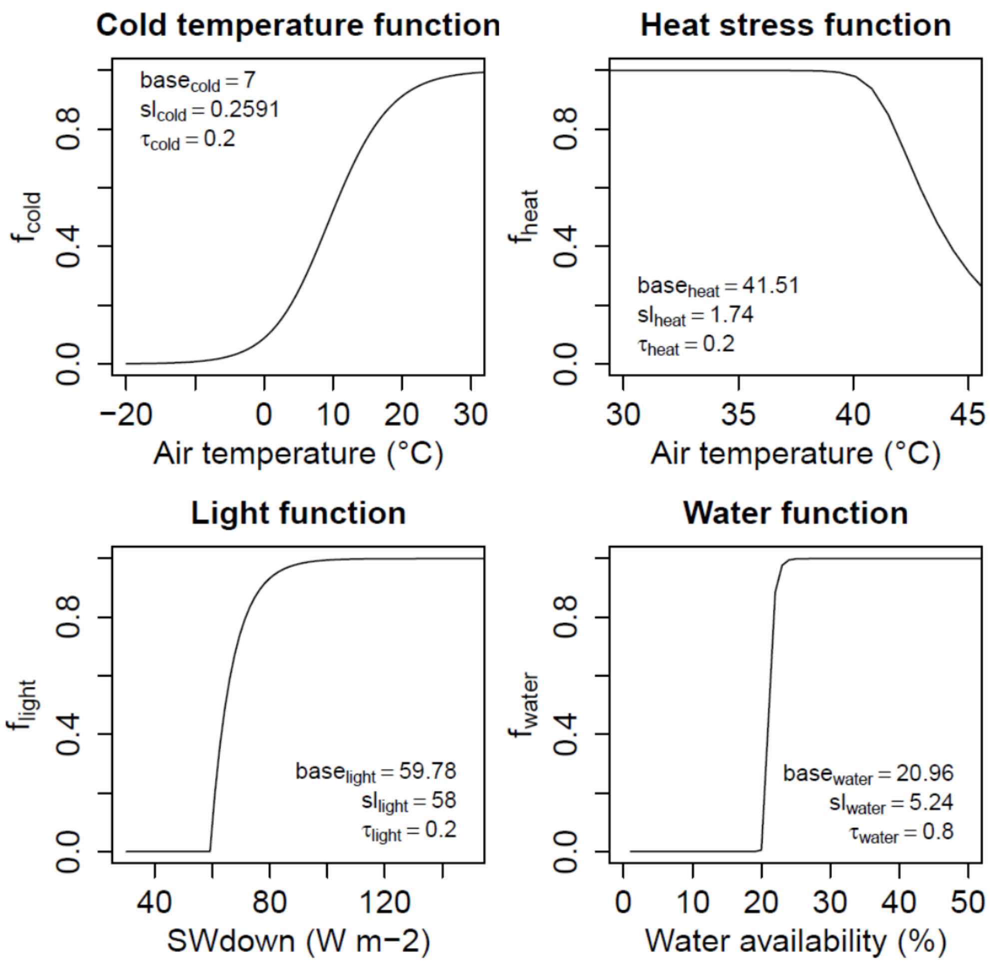


LPJmL-MDI setup



New phenology scheme based on GSI

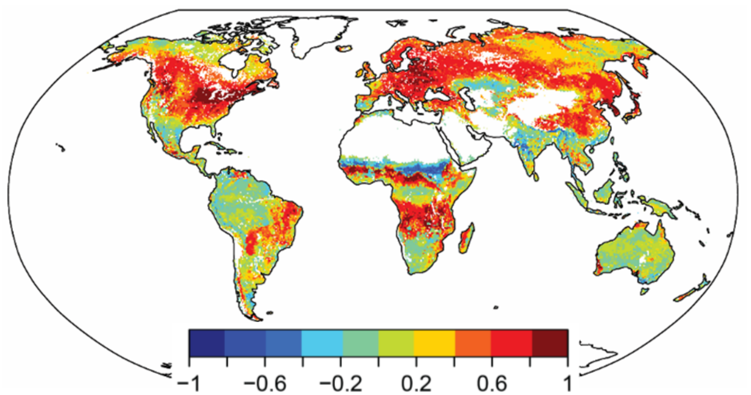
[Jolly et al., 2005]



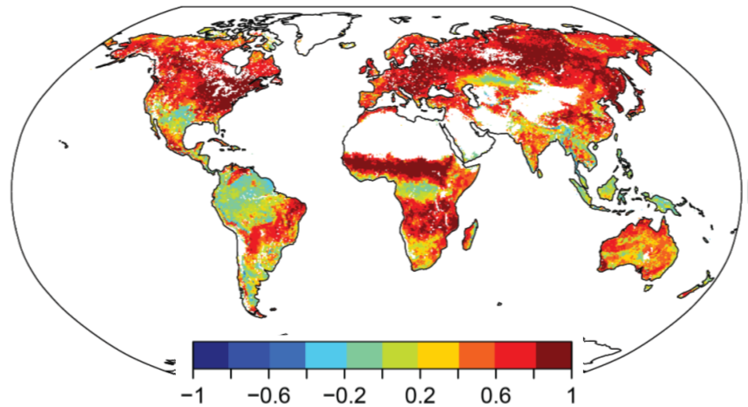
New phenology scheme based on GSI

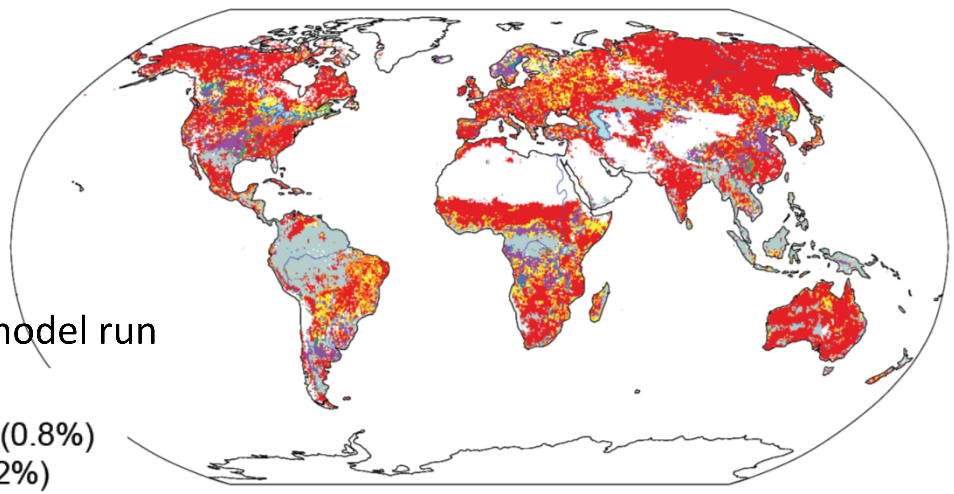
[Jolly et al., 2005]

a) Cor LPJmL-OP-prior ~ GIMMS3g



b) Cor LPJmL-GSI ~ GIMMS3g

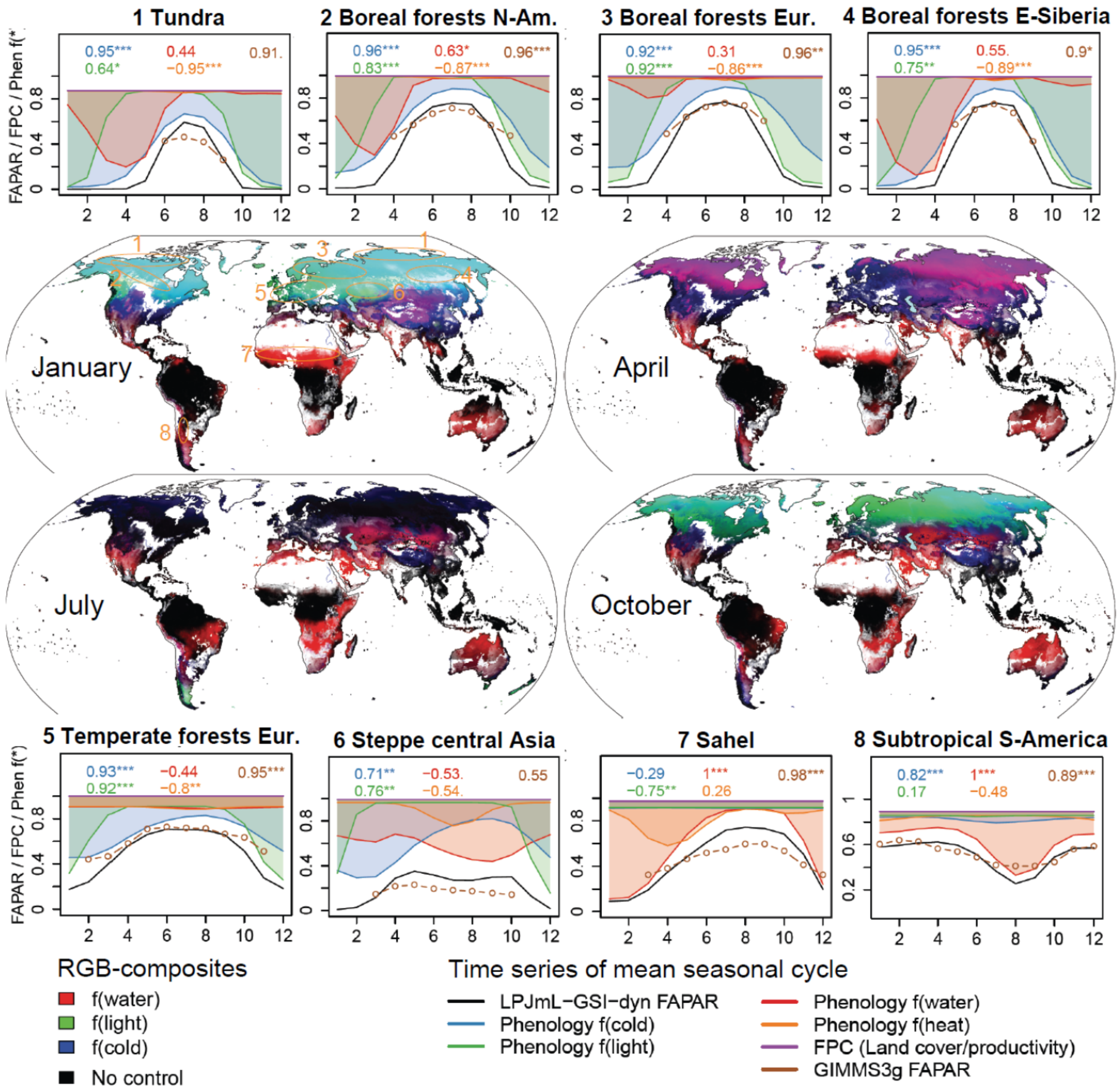




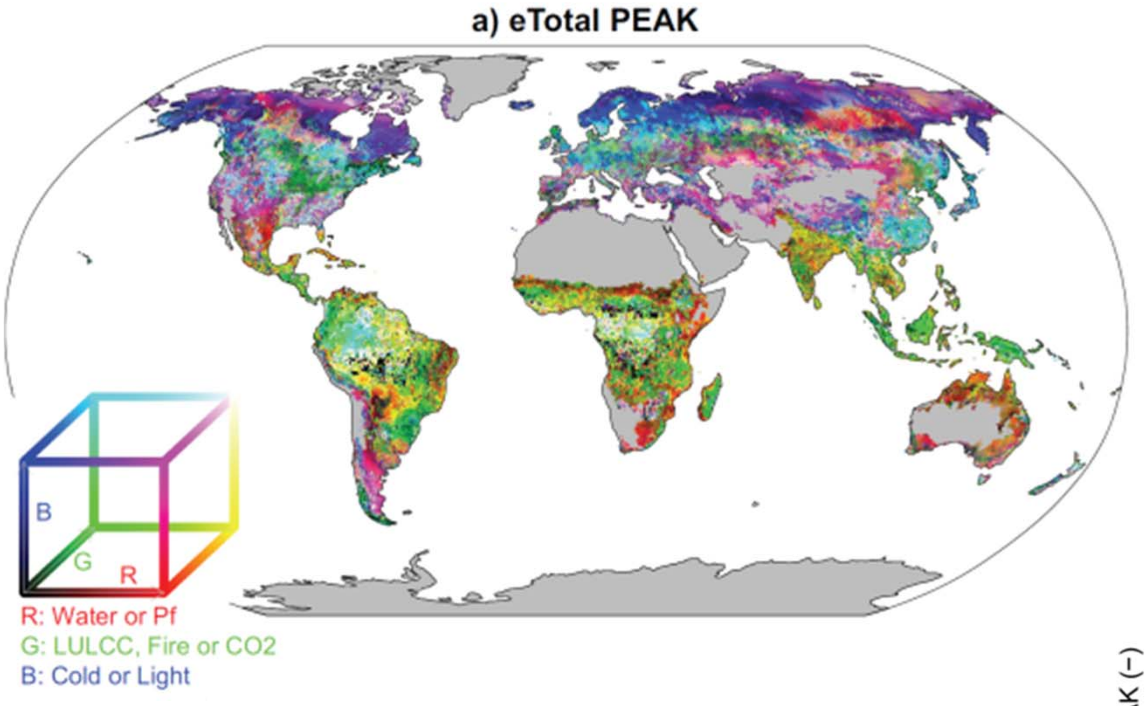
c) Best LPJmL model run

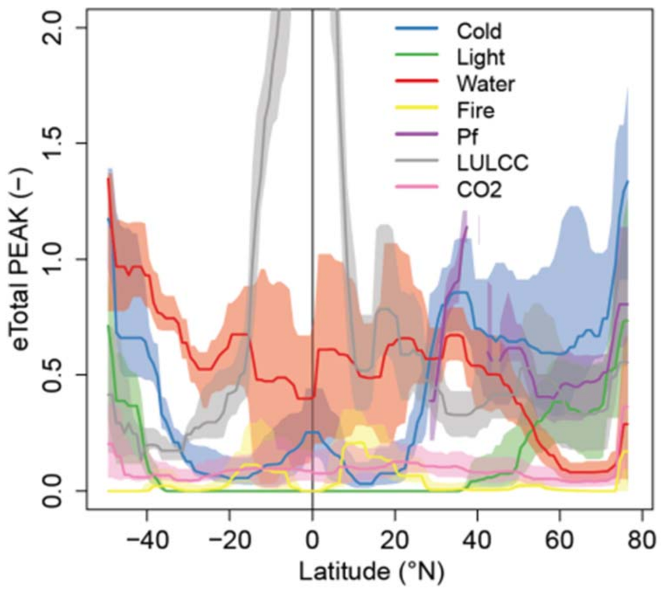
- Cor < 0.2 (10%)
- LPJmL−OP−dyn (0.8%)
- LPJmL−OP−gc (2%)
- LPJmL−GSI (60%)
- LPJmL−OP−dyn + LPJmL−OP−gc (8%)
- LPJmL−OP−dyn + LPJmL−GSI (5%)
- LPJmL−OP−gc + LPJmL−GSI (10%)

Seasonal controls on phenological development



Drivers of annual and decadal variability





Improving the Modelled Global Terrestrial Carbon Cycle by
Assimilating CO_2 Mole Fractions and FAPAR with the
MPI – Carbon Cycle Data Assimilation System

MPI-CCDAS

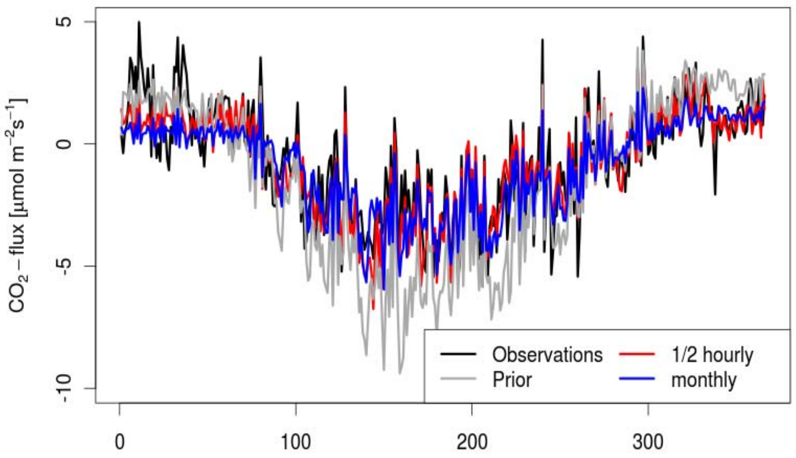
Schuermann et al., in prep.

Site & global scale optimizations

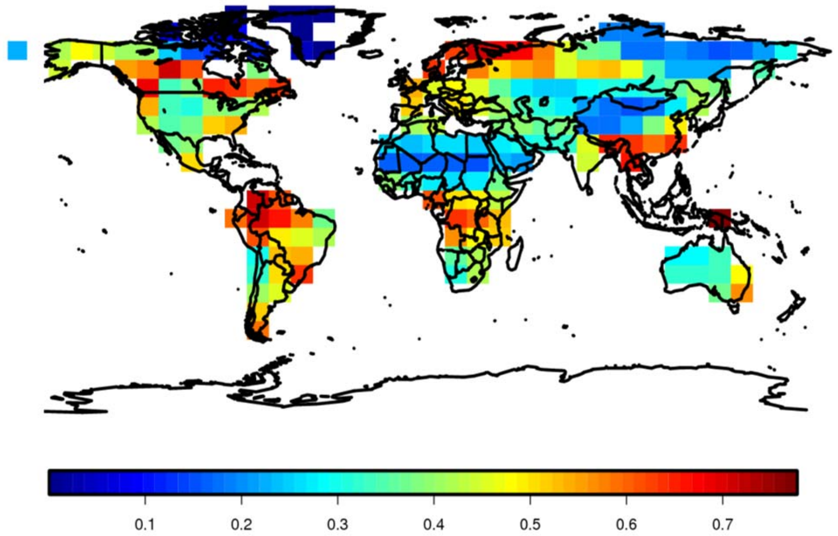
Site scale

Global scale

Net Ecosystem Exchange



FAPAR



Fluxnet NEE

CO2 mole fractions

Satellite observations for both

Details of the exercise

Spatial resolution of 8° x10°

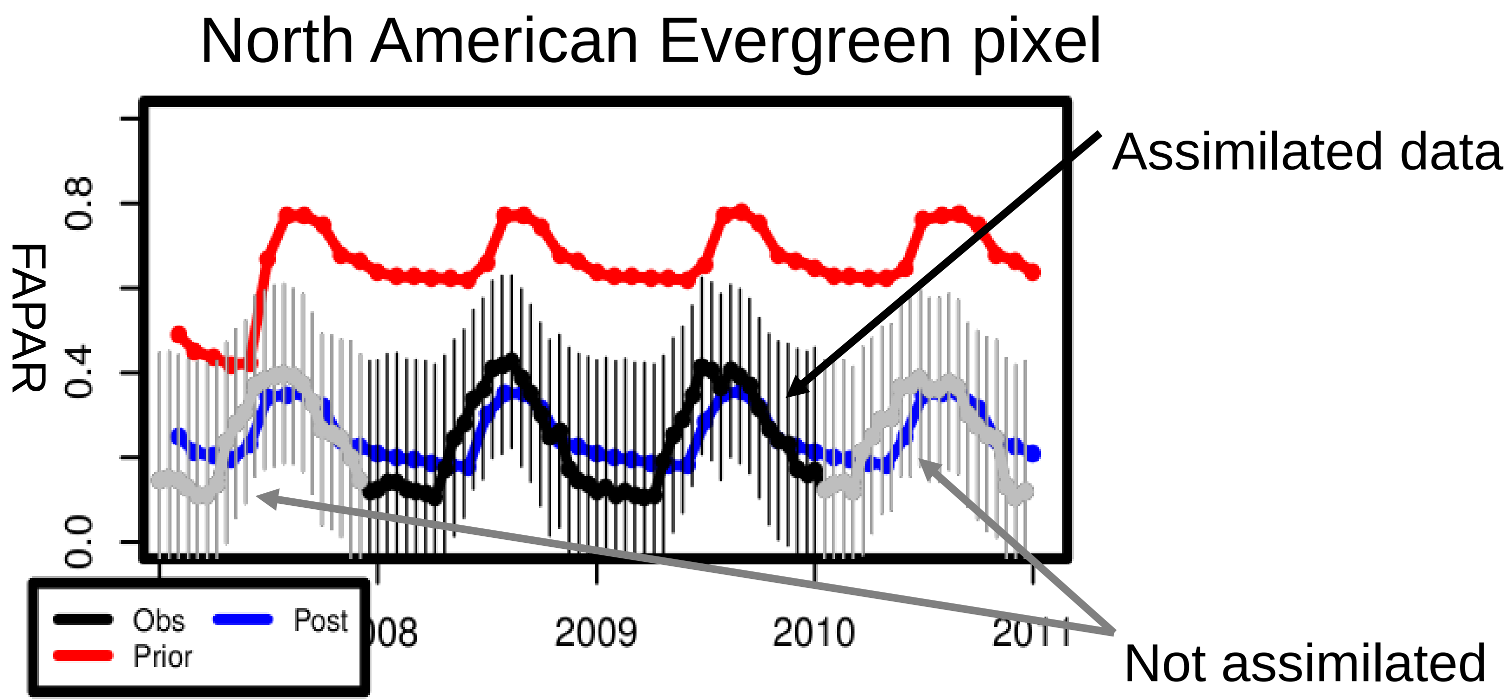
Assimilating 2 years (2008 & 2009)

1 year (2007) as spin-up

1 year (2010) as "evaluation"

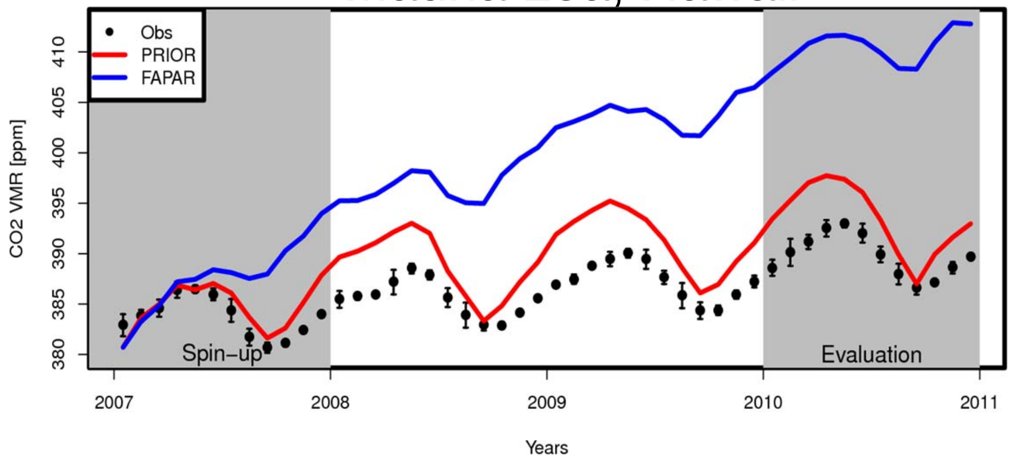
Prior: JSBACH without having seen observations

Post(erior): JSBACH with improved parameters/initial conditions after having seen observations

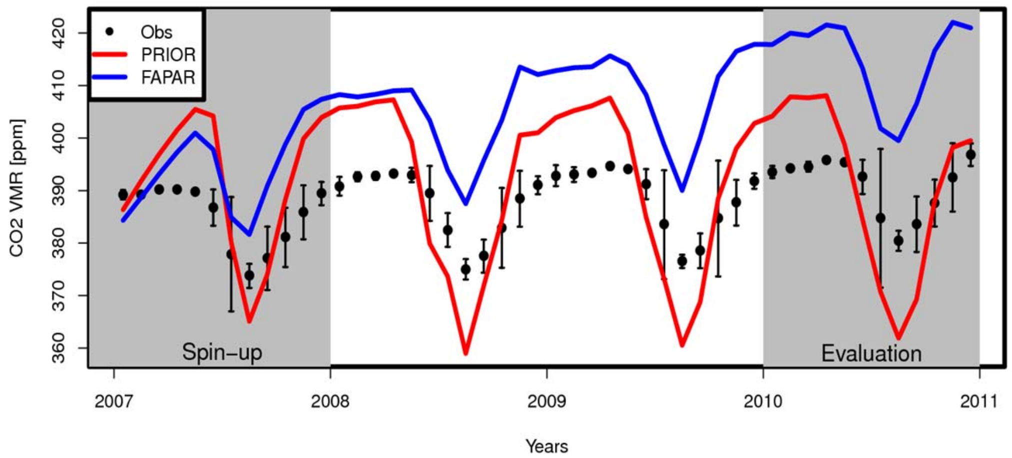


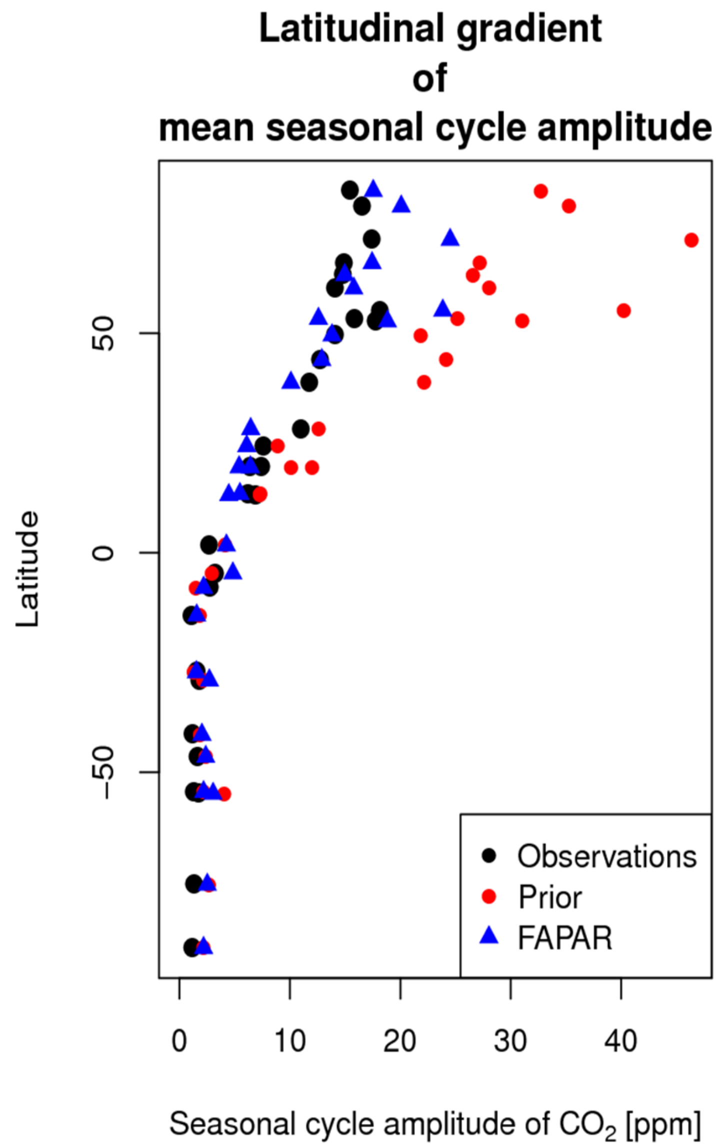
CO2 (FAPAR assimilation)

Mauna Loa, Hawaii

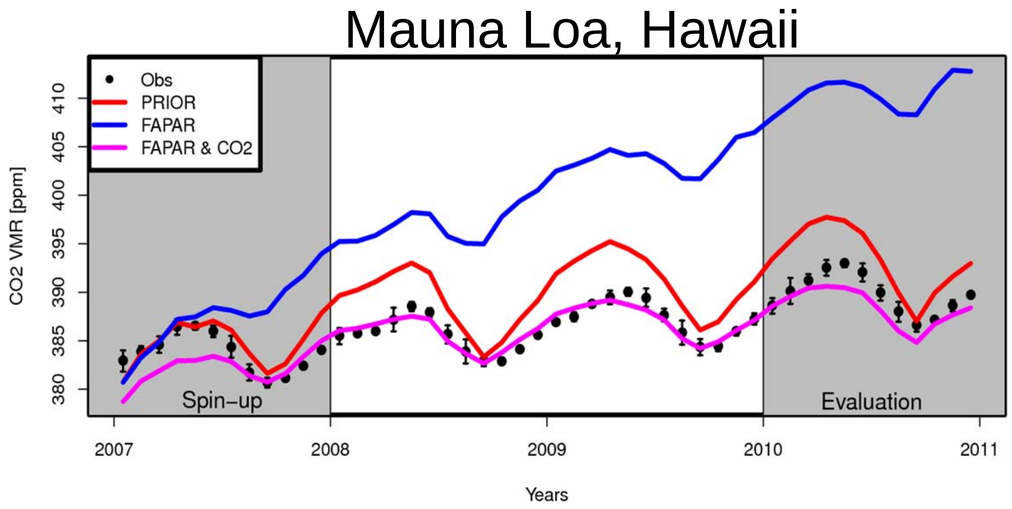


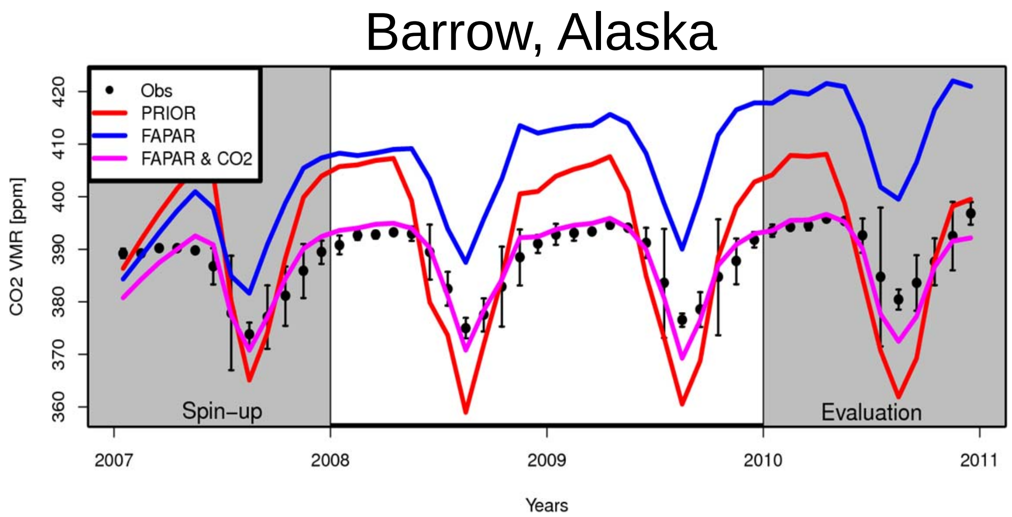
Barrow, Alaska

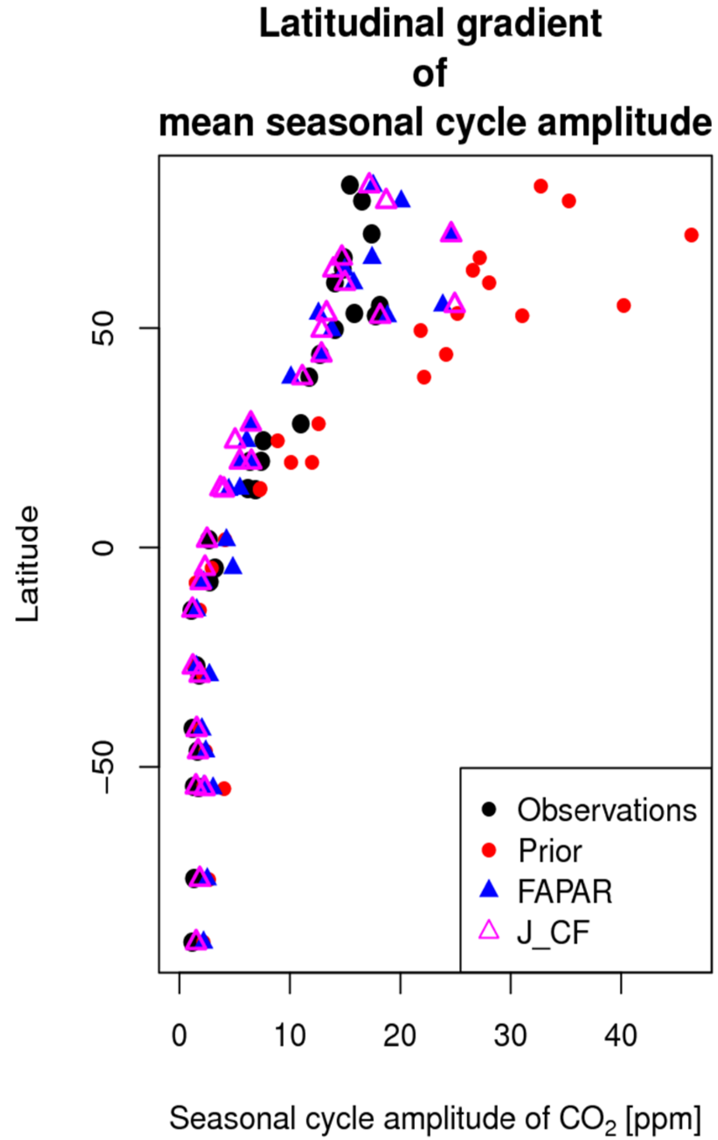




CO2 (FAPAR & CO2 assimilation)



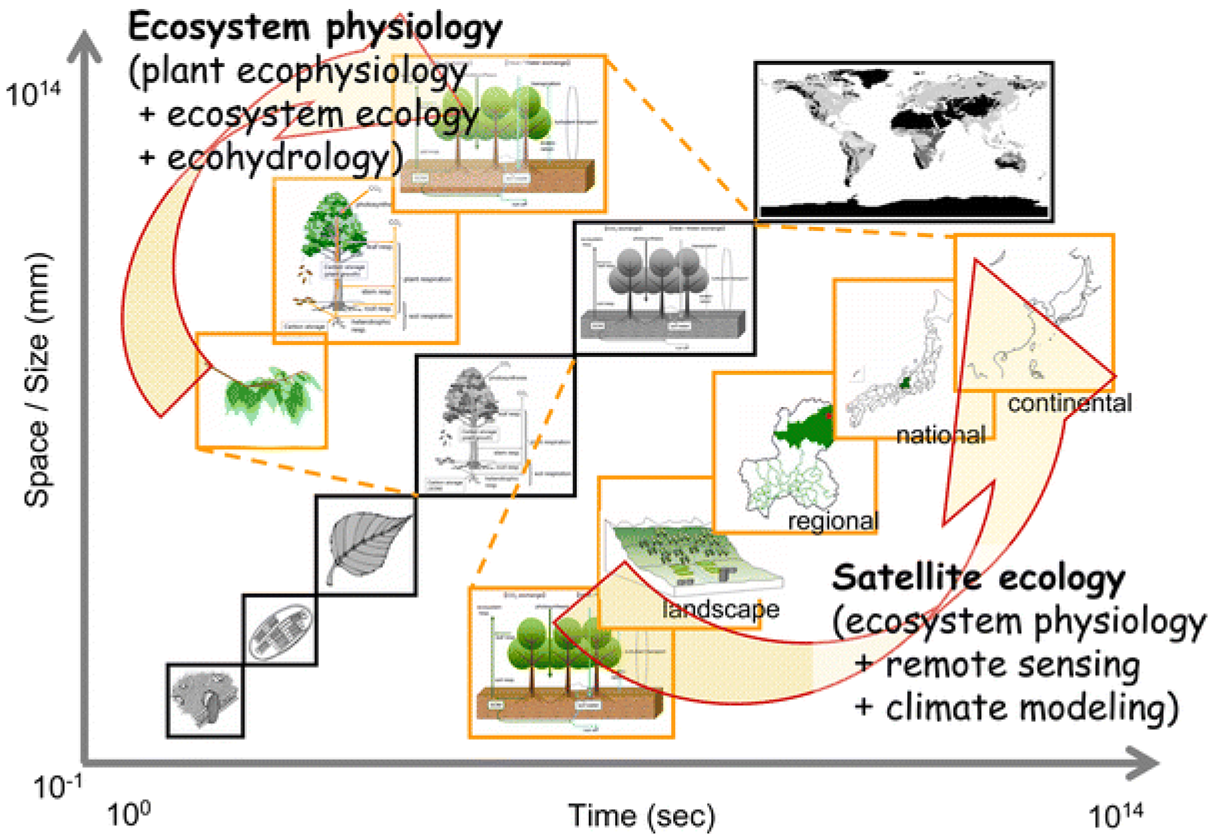




Conclusions

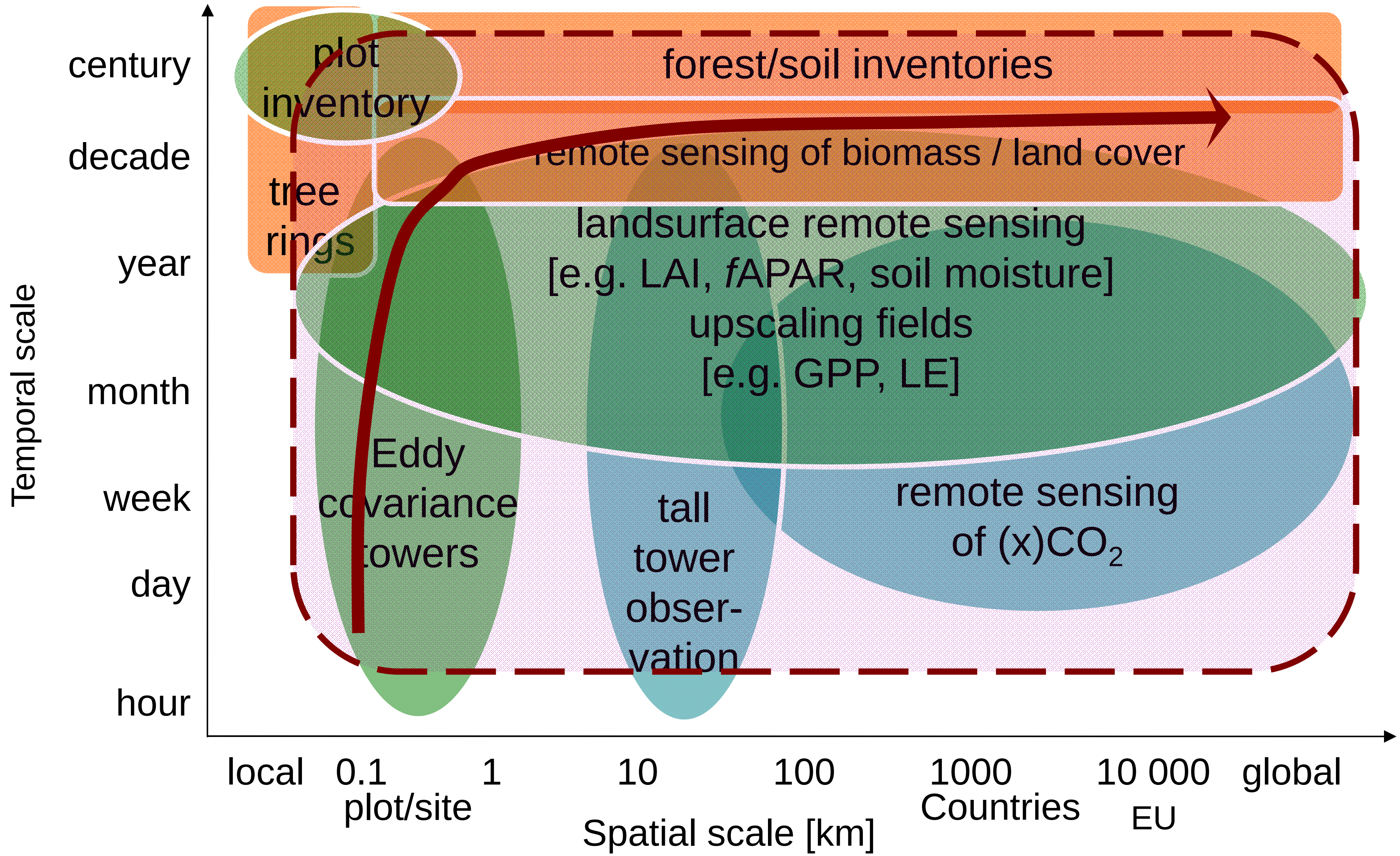
- importance of multiple data streams in model-data integration exercises
  - further constrained parameterizations
  - consistency with observed states
  - addressing equifinality
  - predictive uncertainty
- significant implications for diagnostic and prognostic model runs
- remote sensing provide unique constraints to integrate site-level and regional to global scales dynamics of responses of terrestrial ecosystem to climate variability
- allocation / lag effects and the carbon-hydrological cycle

processes and observations spanning from wide scales



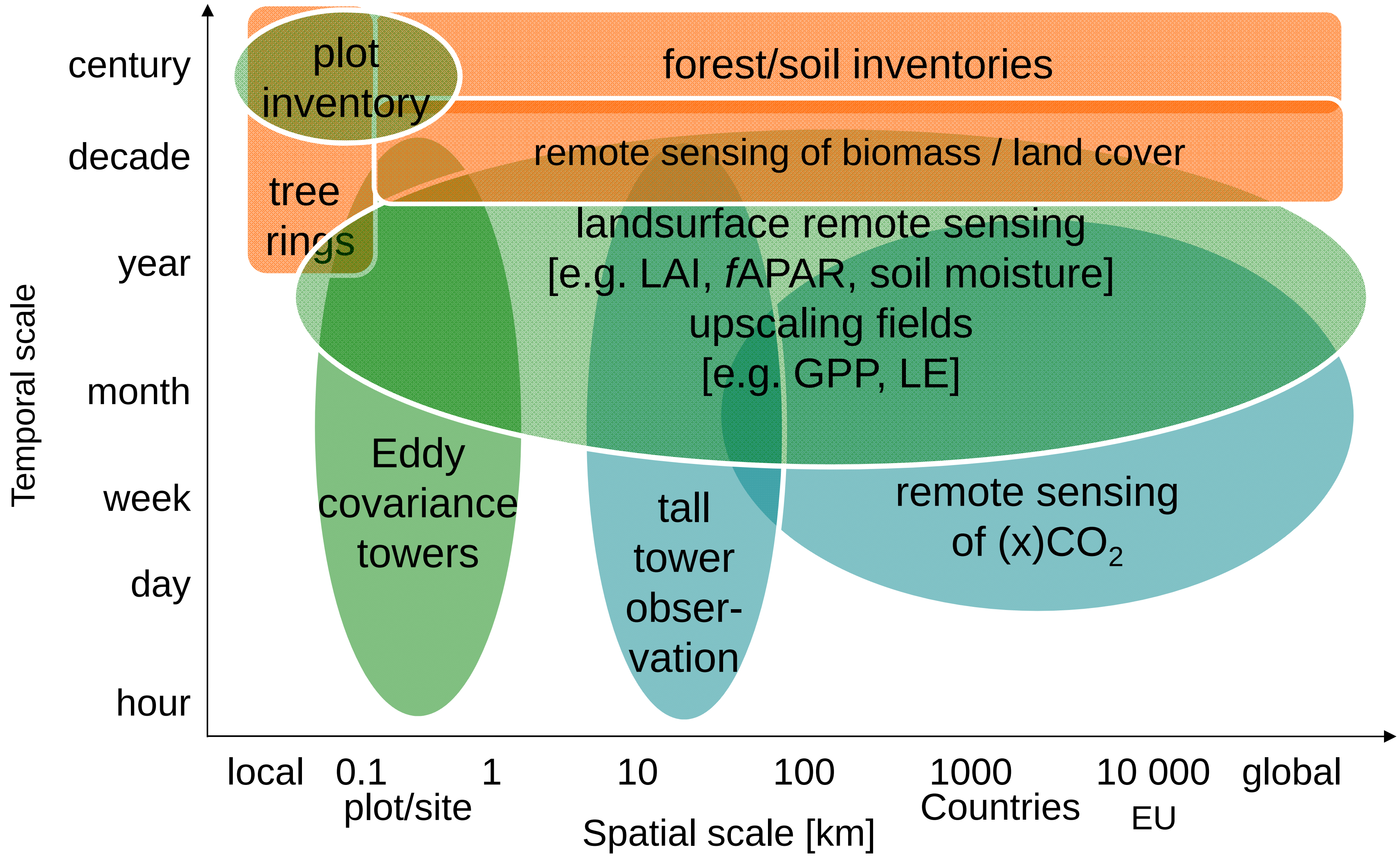
http://link.springer.com/article/10.1007%2Fs10265-008-0188-2/fulltext.html

Observational scales



THANK YOU!

observational scales



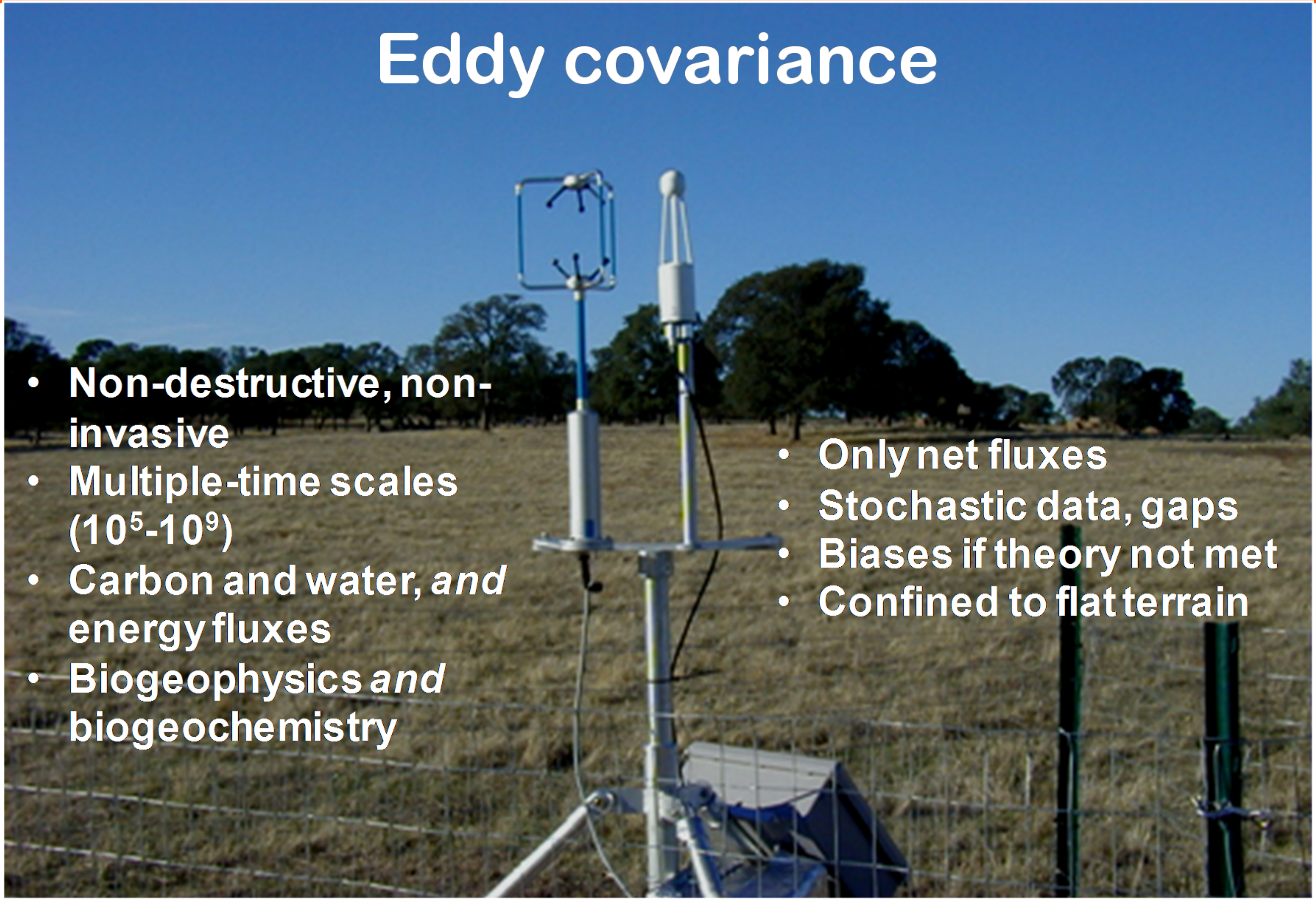
From ecosystem level to regional/global scales

- Parameterizations
  - Based on biotic and abiotic covariates (e.g. Carvalhais et al., 2010; Horn and Shulz, 2011)
  - Based on spatial/temporal distributions of plant functional types

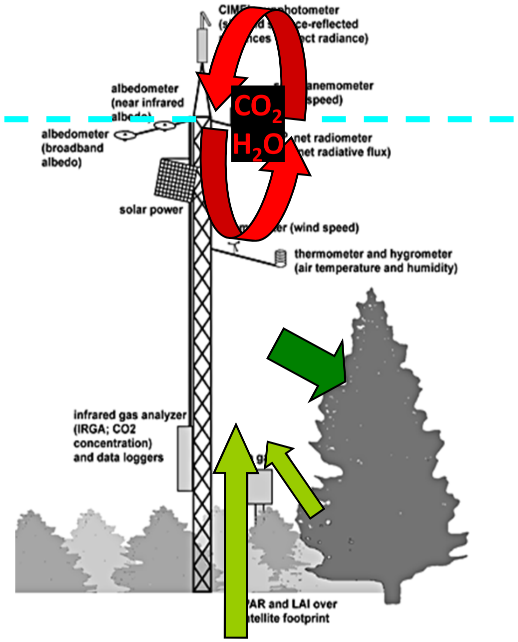
Acknowledge:

- Site particularities (e.g. ground water access, disturbance history/initial conditions, …)
- Determination of site representativeness

Quantifying ecosystem-atmosphere interactions

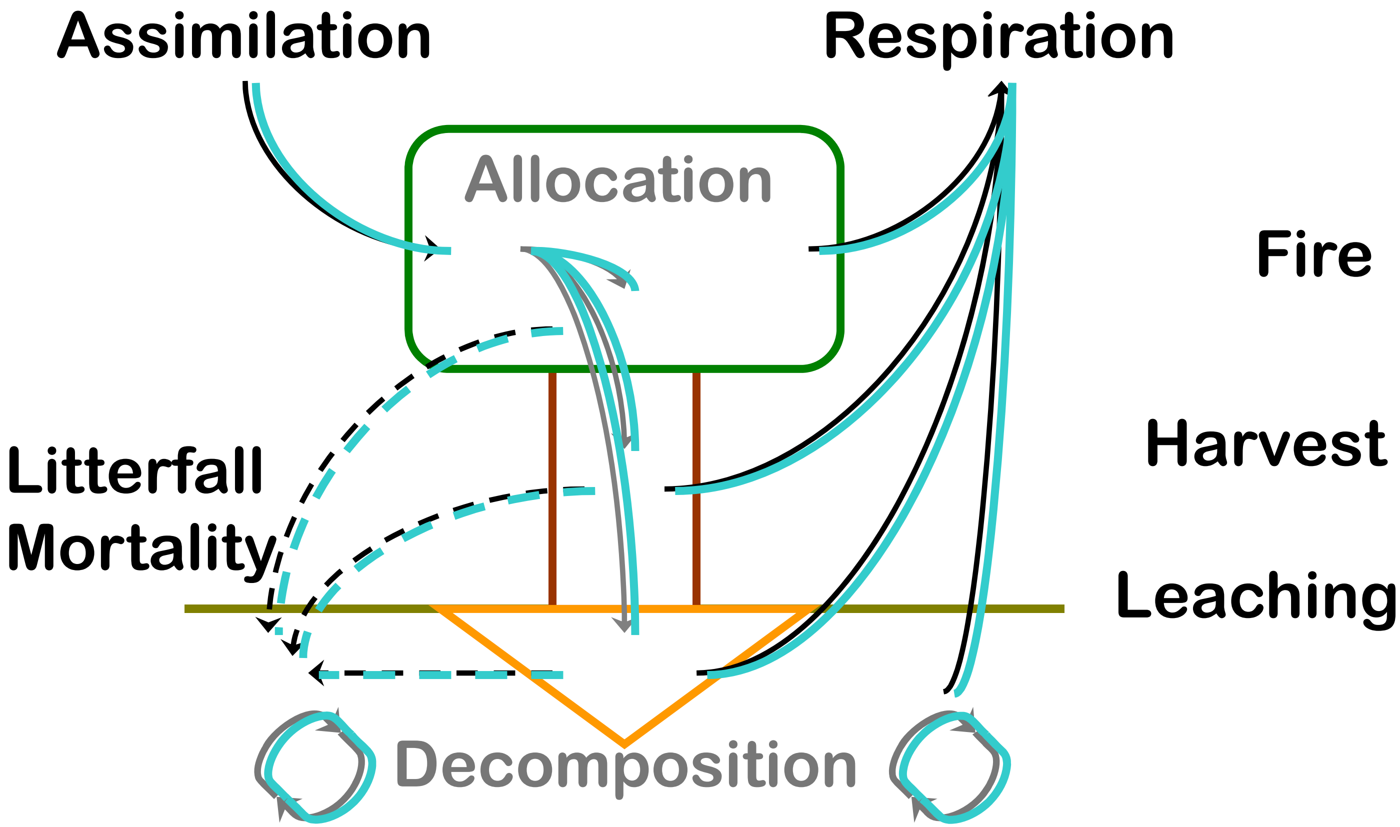


ecosystem fluxes



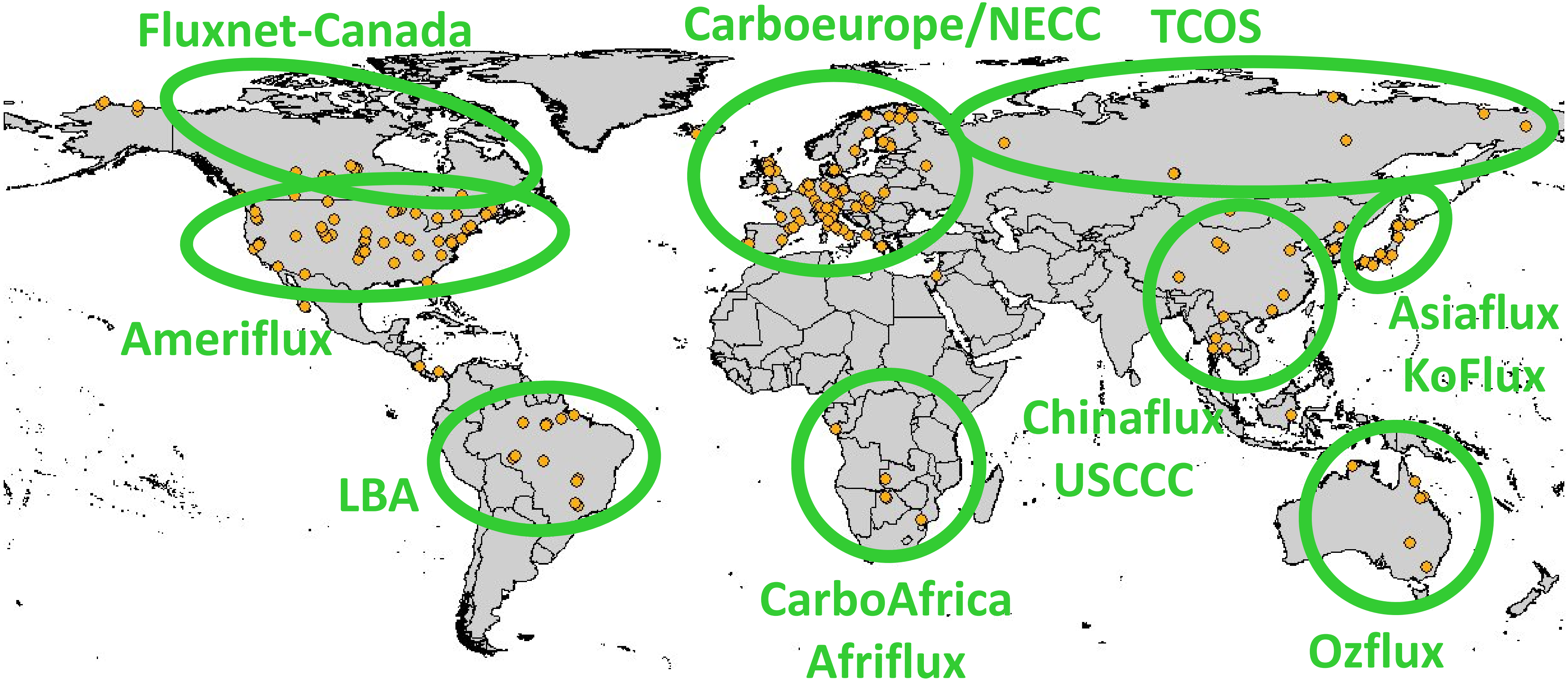
$$NEP = (GPP - R_A) - R_H$$

Ecosystem C cycling and fluxes



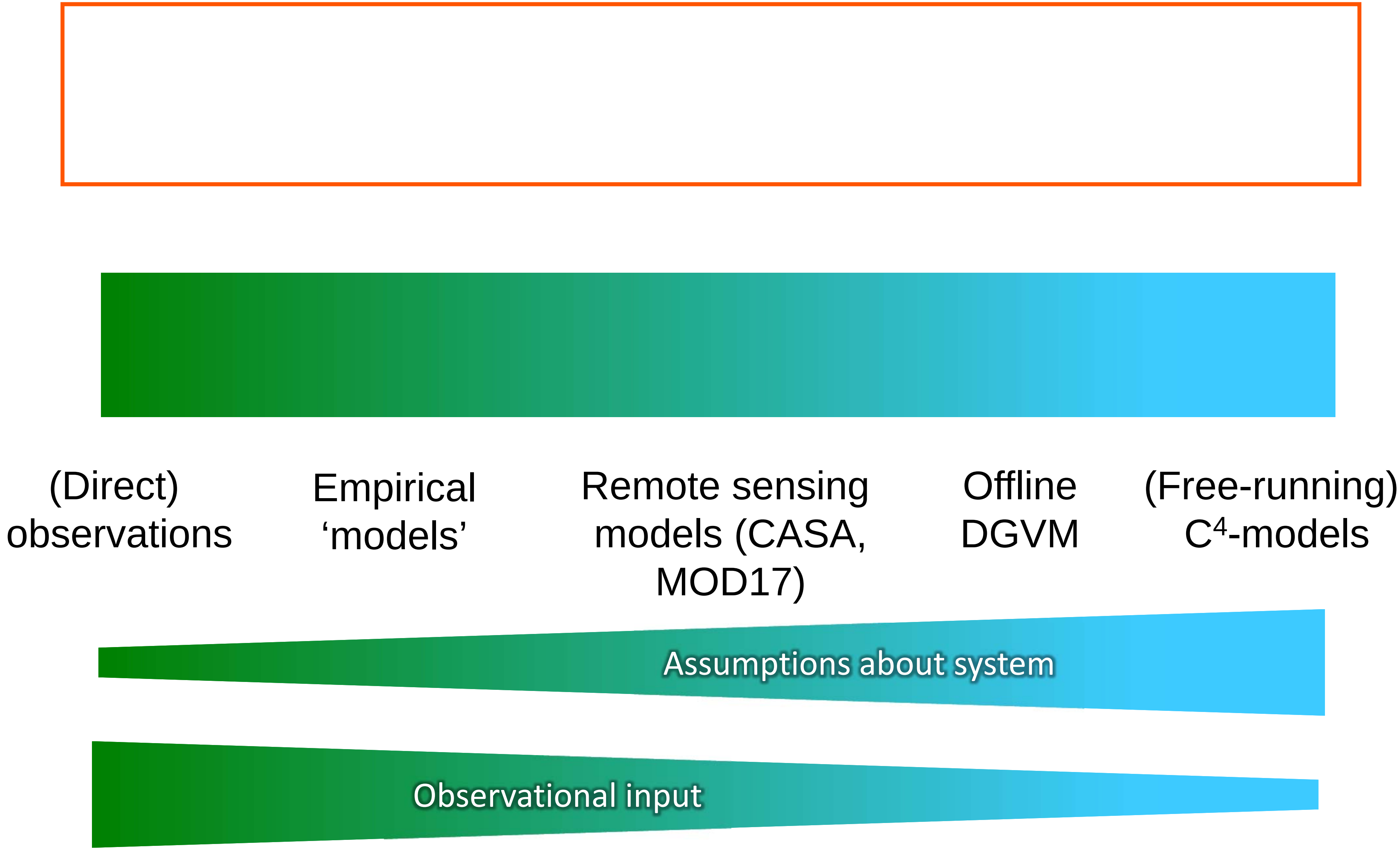
[Adapted from Lasslop 2010]

FLUXNET: a network of network of eddy covariance sites



<u>**La Thuille data set:**</u>

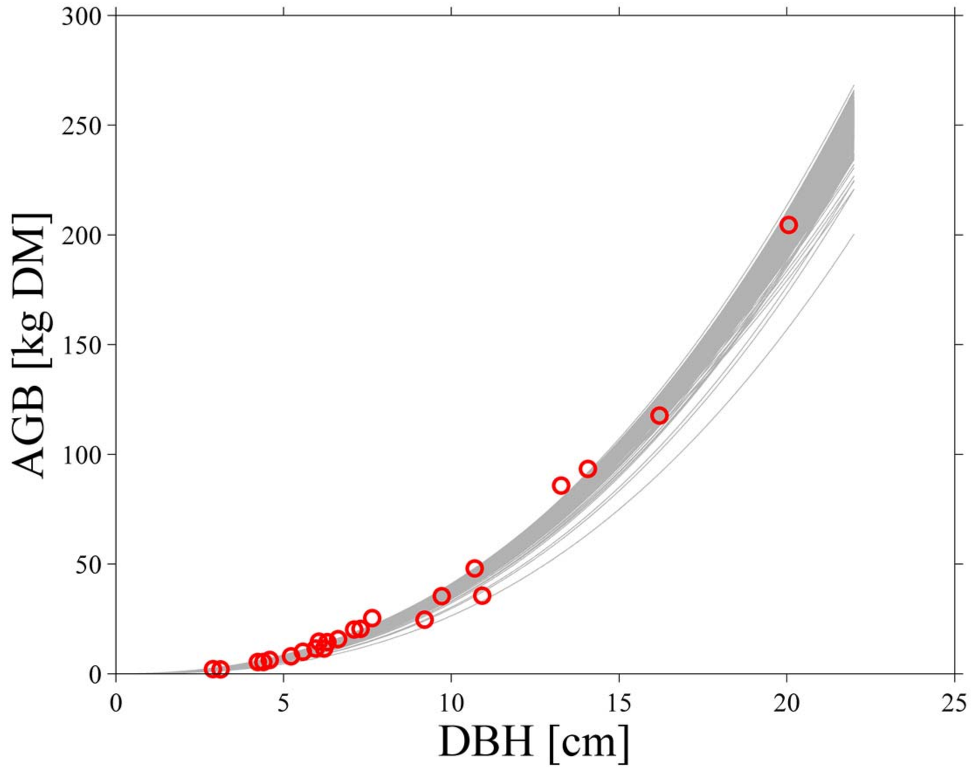
- **>950 site-years from >250 sites**
- **Standardized u*-filtering, gap-filling, flux-partitioning and uncertainties (Aubinet et al. 2001, Foken et al. 2003, Reichstein et al. 2005, Richardson et al. 2006, Papale et al. 2006, Moffat et al. 2007, Desai et al. 2008, Lasslop et al. 2008)**



(Direct) observations
Empirical 'models'
Remote sensing models (CASA, MOD17)
Offline DGVM
(Free-running) C4-models
Assumptions about system
Observational input

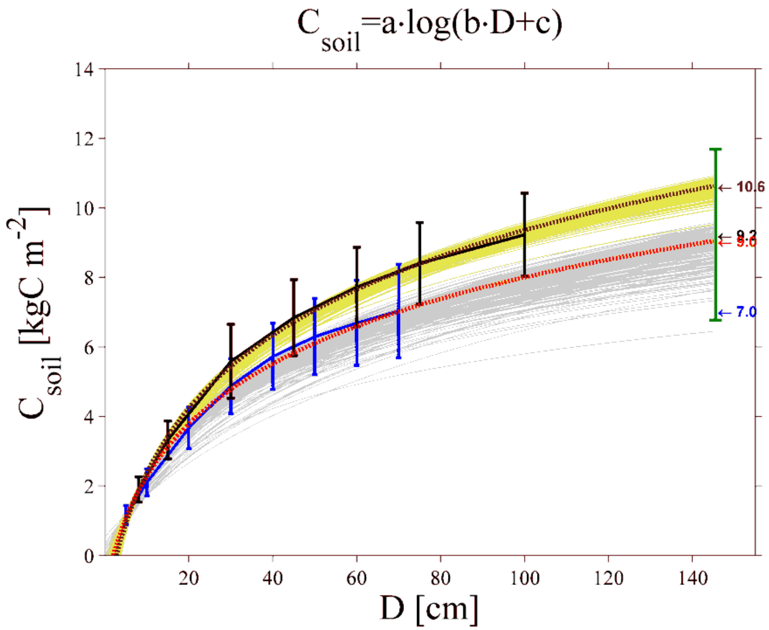
Data : vegetation and soil C stocks

AGB and AGB increments



[Wutzler et al., 2008]

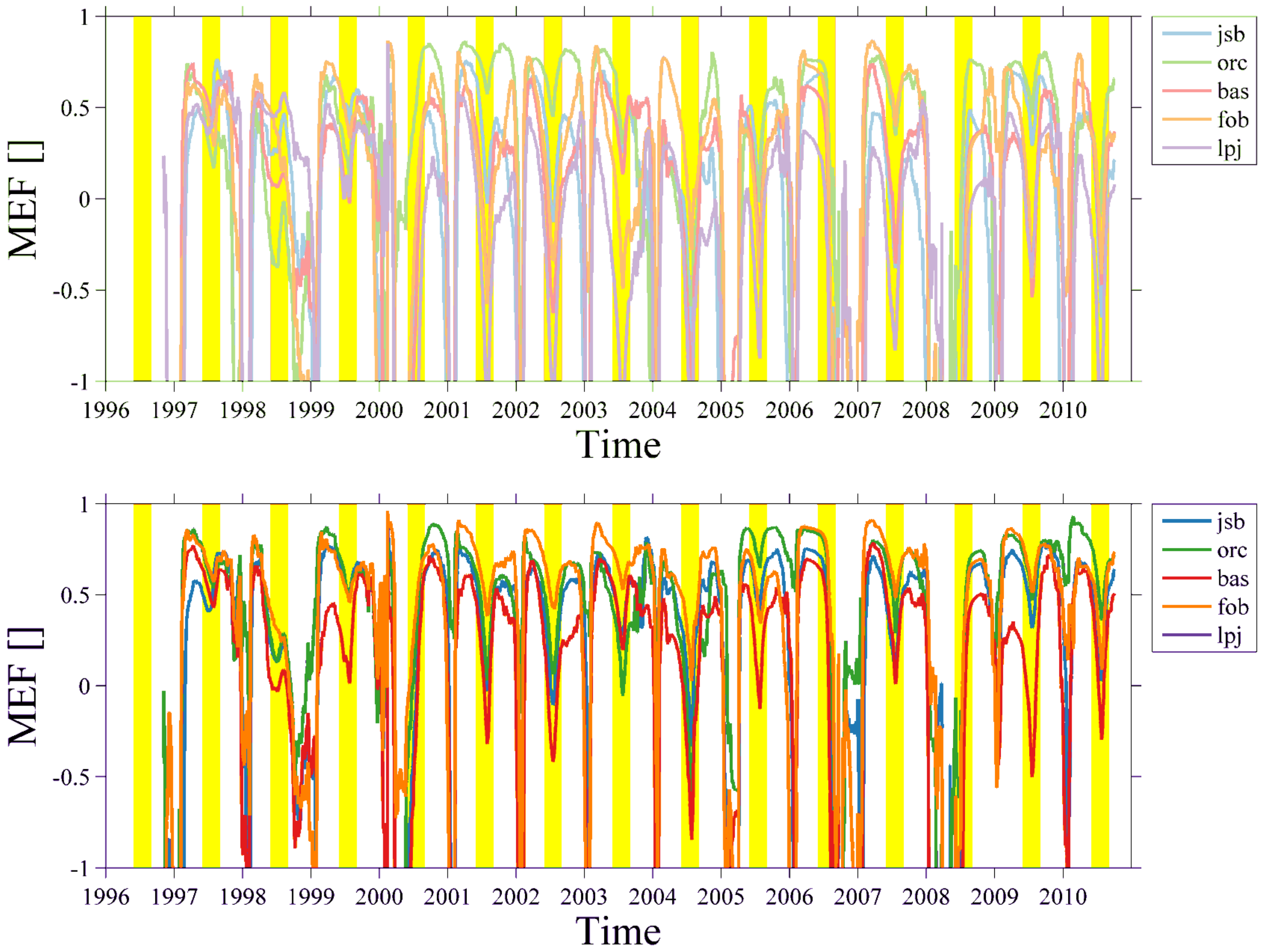
Soil C stocks



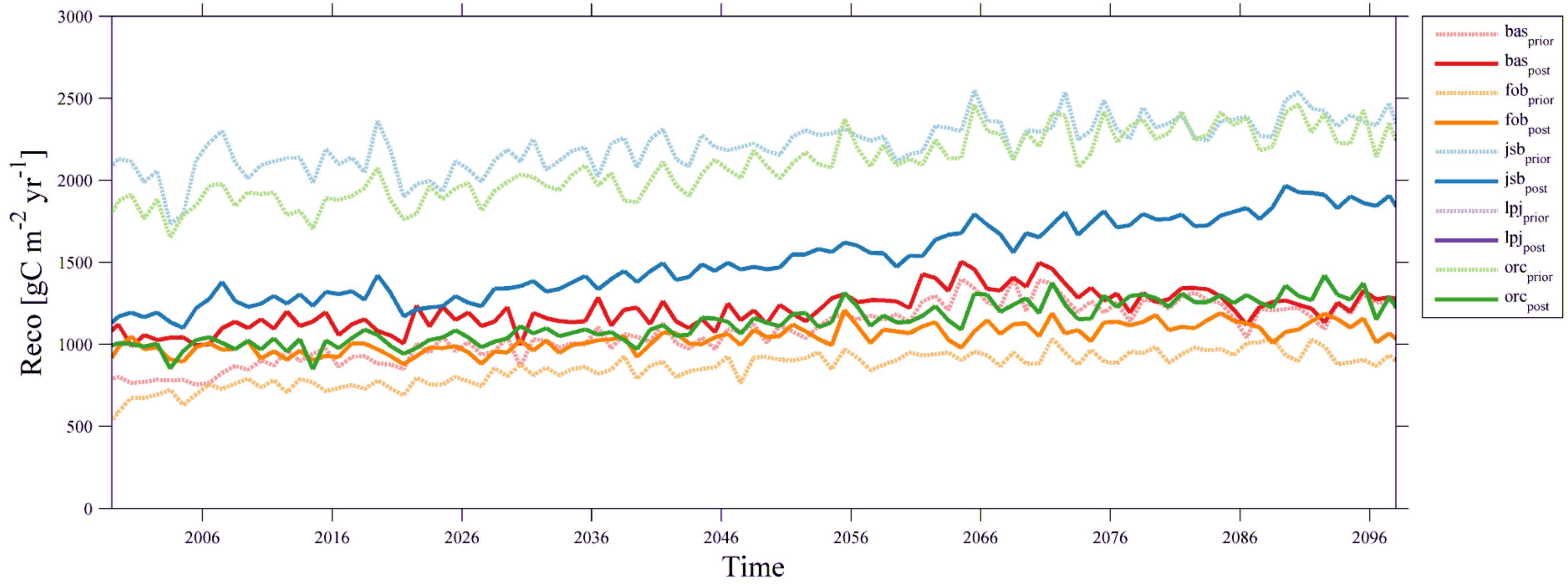
[Carvalhais et al., in prep.]

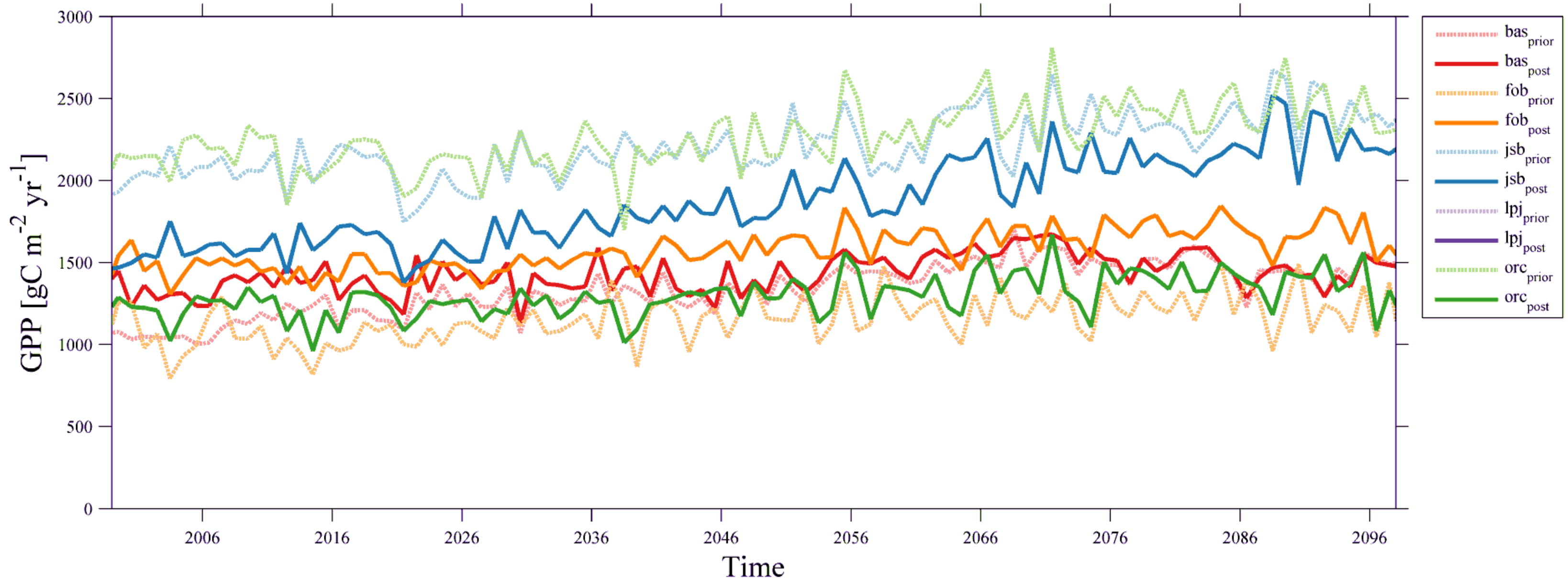
Hesse : seasonal misfits : window180days

optimized

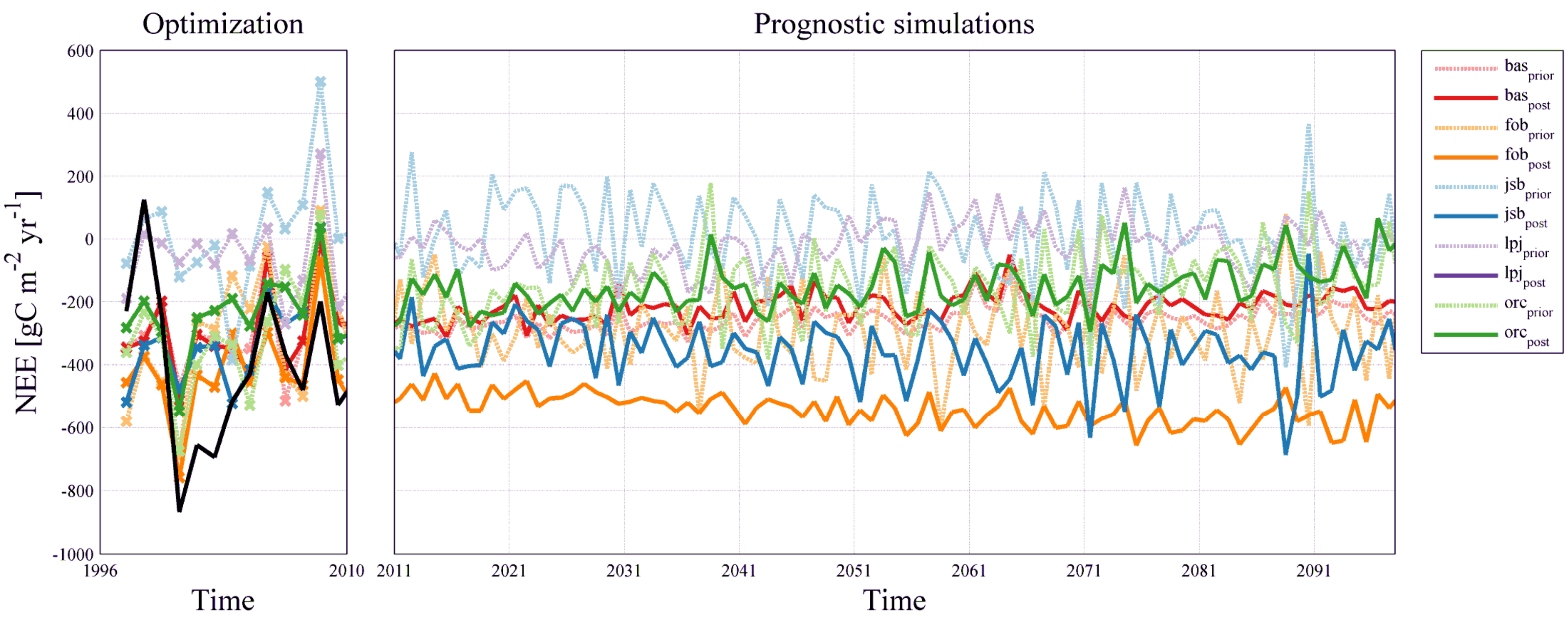


Changes in prognostic gross fluxes





Changes in net ecosystem fluxes

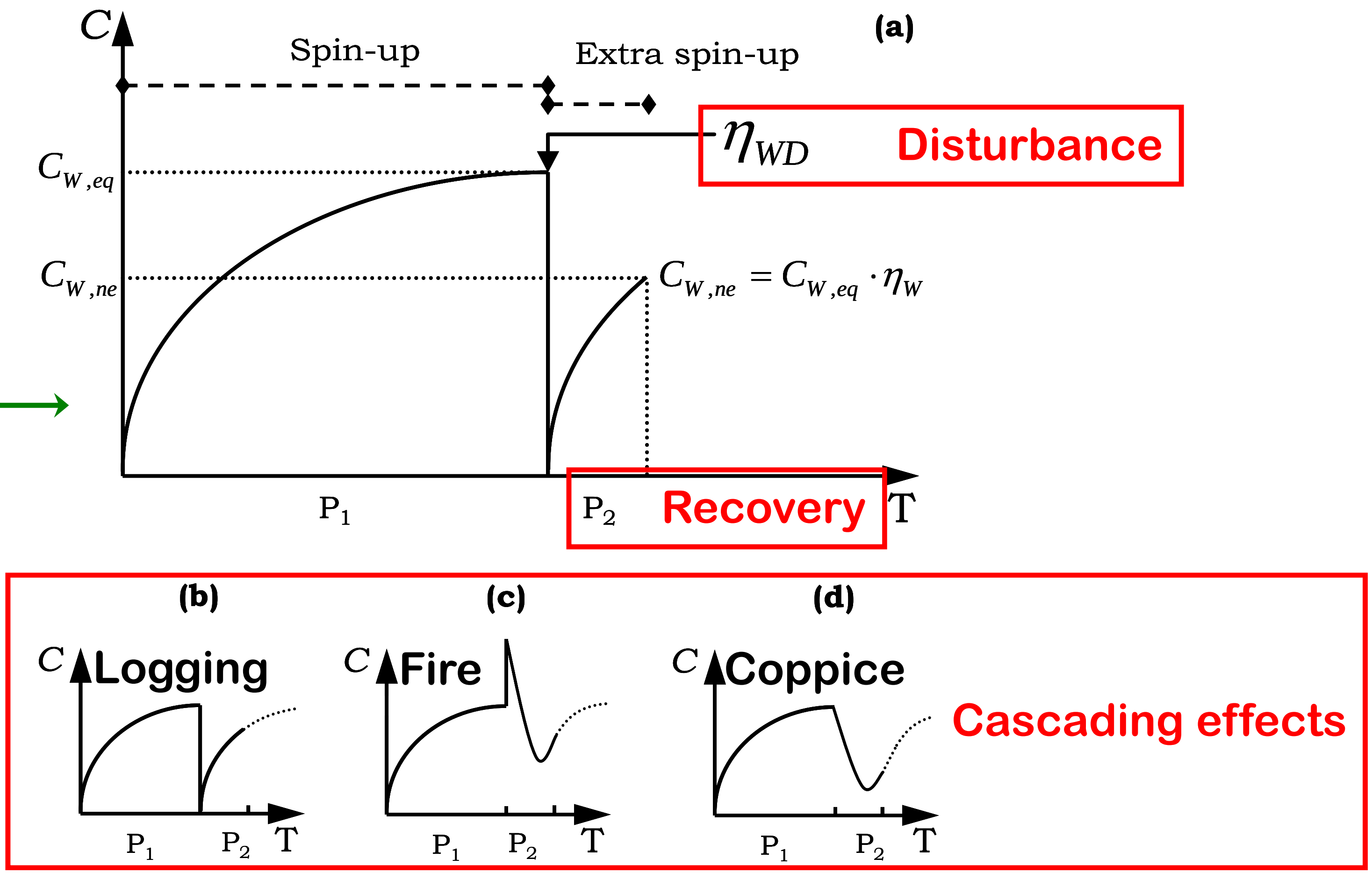


Not a clear sign of spread reduction

ADDRESSING DIFFERENT RECOVERY DYNAMICS WITH FLUX AND BIOMETRIC CONSTRAINTS

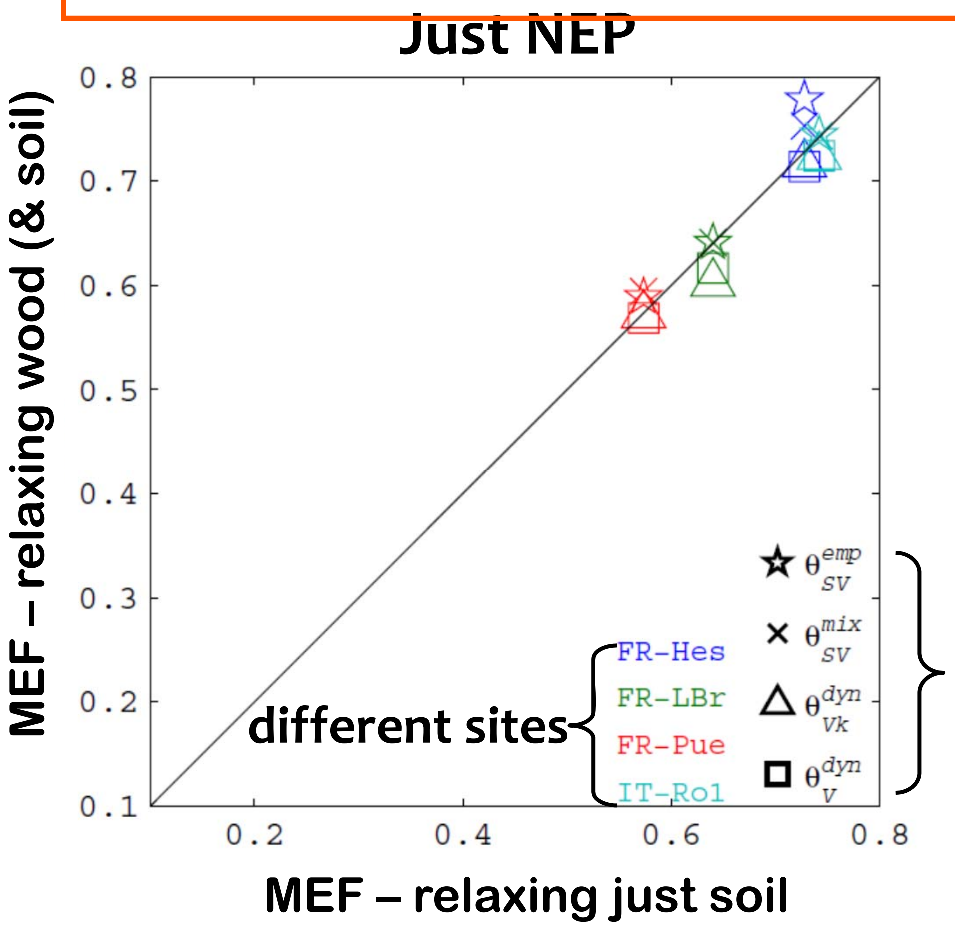
Challenging dynamics

But also empirical approaches



[Carvalhais et al., 2010]

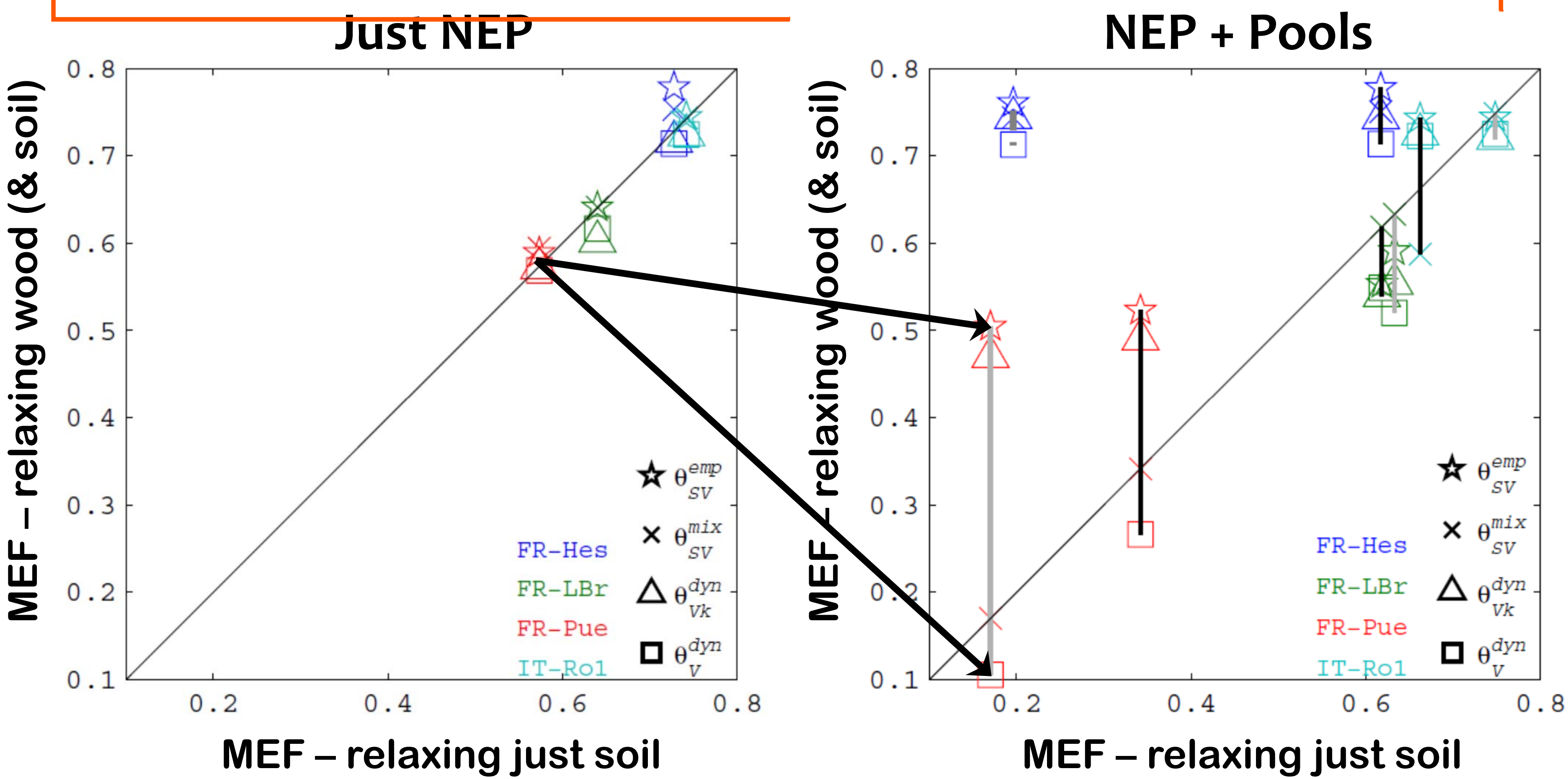
Scenario differentiation



- **Despite differences in the initialization routines it is not possible to distinguish between the different "prescribed dynamics"**

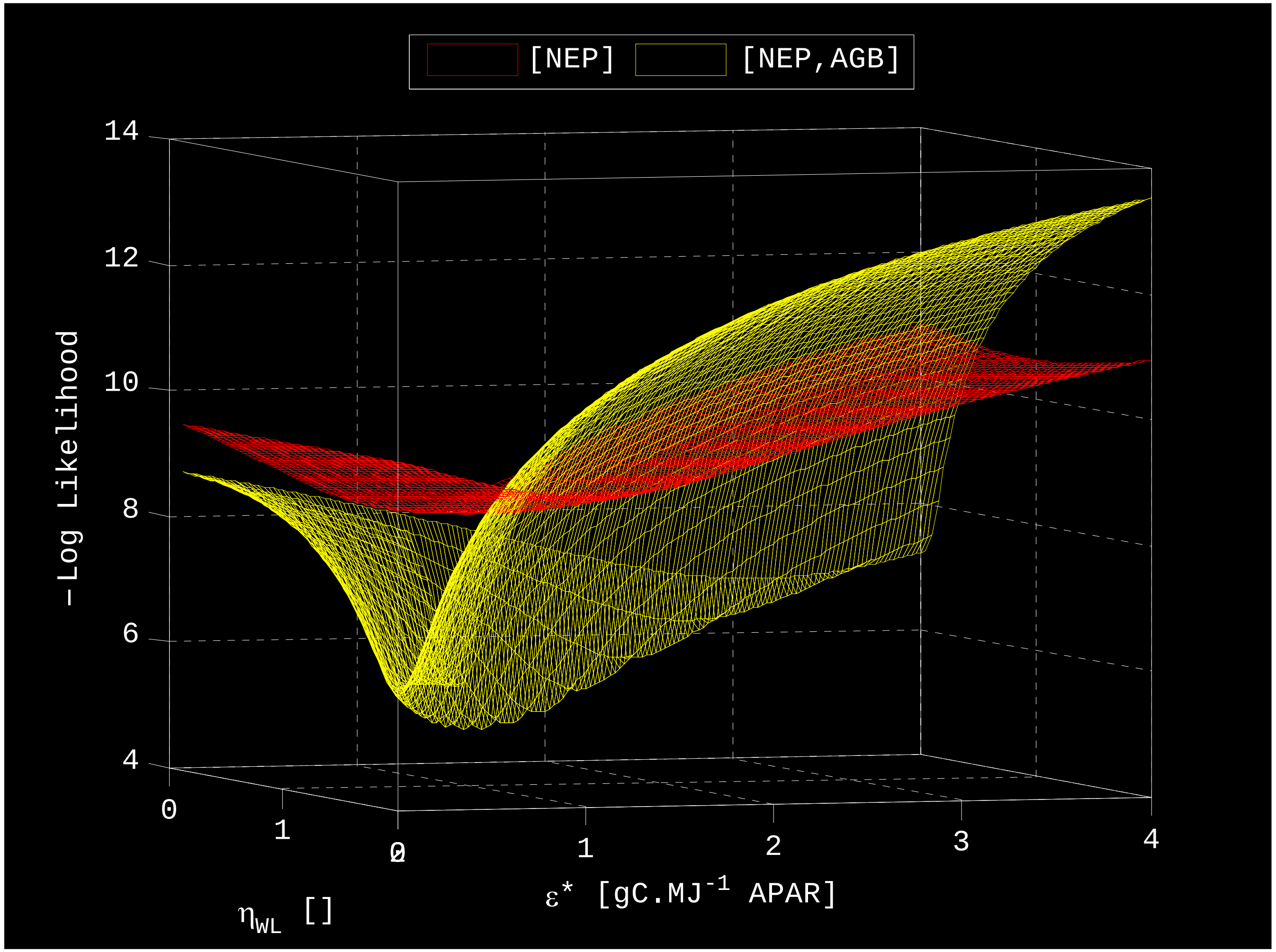
[Carvalhais et al., 2010]

Scenario differentiation



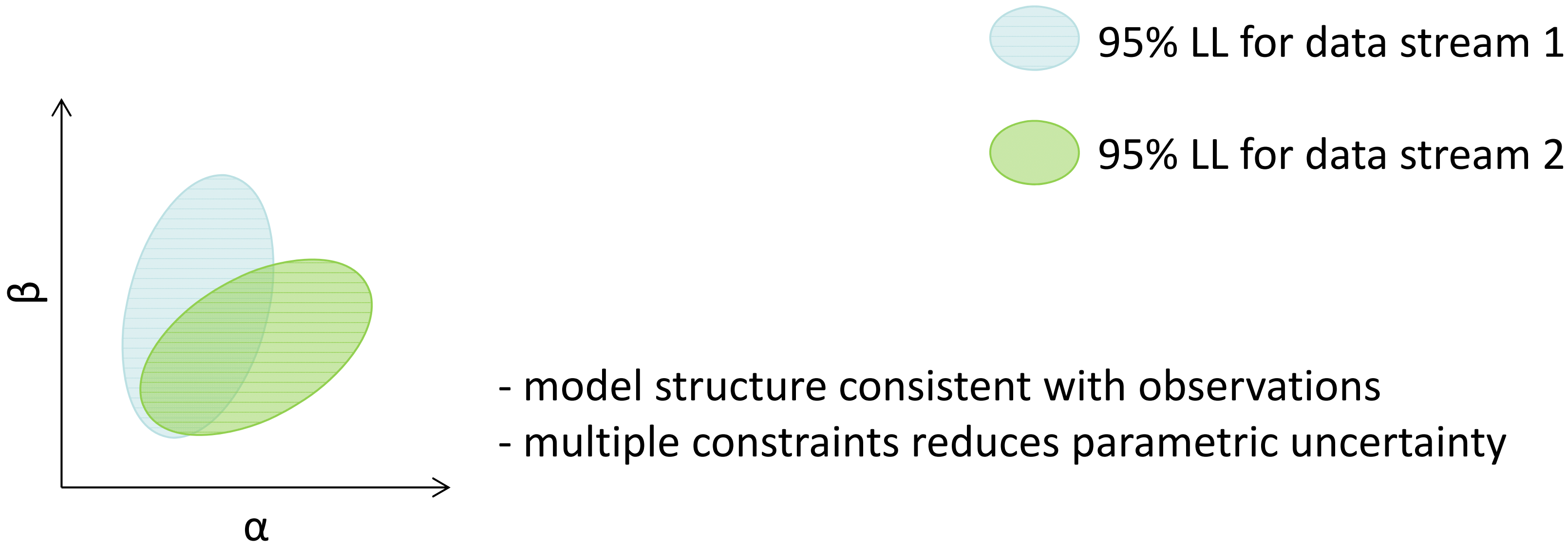
[Carvalhais et al., 2010]

different convergence / stronger constraints



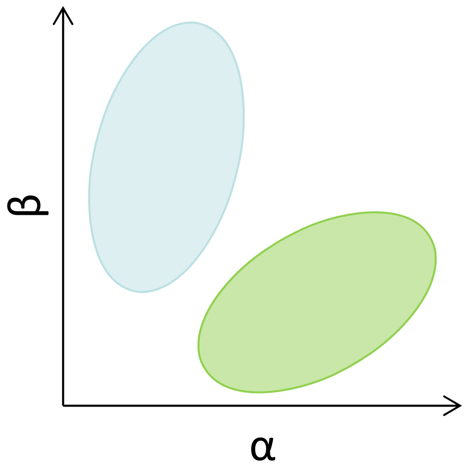
[Carvalhais et al., 2010]

Relevance of multiple constraints



- model structure consistent with observations
- multiple constraints reduces parametric uncertainty

<u>Challenges:</u> Equifinality, Over-parameterisation (e.g. Knorr et al. 2005, Reichstein et al. 2005)



- model structure inconsistent with both datastreams
- (due) inflation of parameter uncertainty/multimodality

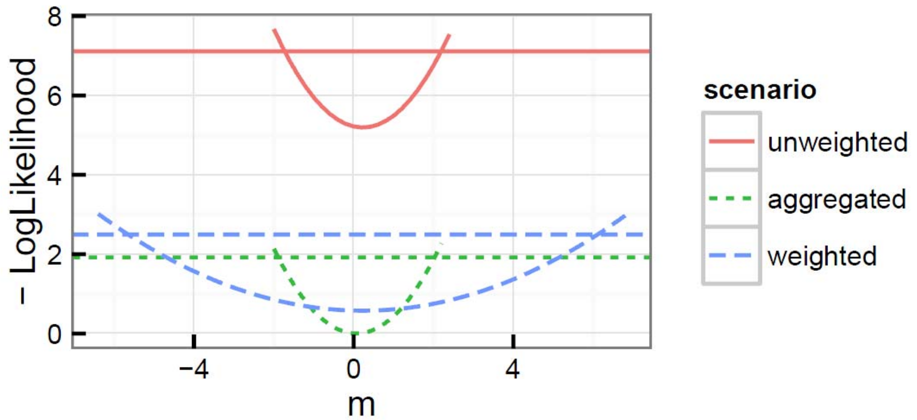
COMBINING DATASETS WITH SUBSTANTIALLY DIFFERENT STATISTICAL PROPERTIES

Wutzler and Carvalhais, in rev.

Particular challenge of multiple constraints approaches

Highly imbalanced dimensions in data streams:

- for a perfect model, different alternative cost functions do not affect the achievement of optimum, but weighted approaches inflate posterior uncertainties.



Particular challenge of multiple constraints approaches

$$\hat{y}_{i,\mathrm{rich}}(a,b,c) = a\,x_{1,\mathrm{sparse}} + b\,(x_{i,\mathrm{rich}} - \mathrm{c})$$

$$\hat{y}_{i,\mathrm{sparse}}(a,b) = a\,x_{i,\mathrm{sparse}} + b\,\overline{x}_{\mathrm{rich}}/10$$