A consistent view on the terrestrial carbon cycle through simultaneous assimilation of multiple data streams into a model of the terrestrial carbon cycle

Marko Scholze[1], T. Kaminski[2], W. Knorr[1], Peter Rayner[3] and Gregor Schürmann[4]

1 Lund University, Sweden 2 The Inversion Lab, Hamburg, Germany 3 University of Melbourne, Australia 4 Max Planck Institute for Biogeochemisty, Jena, Germany

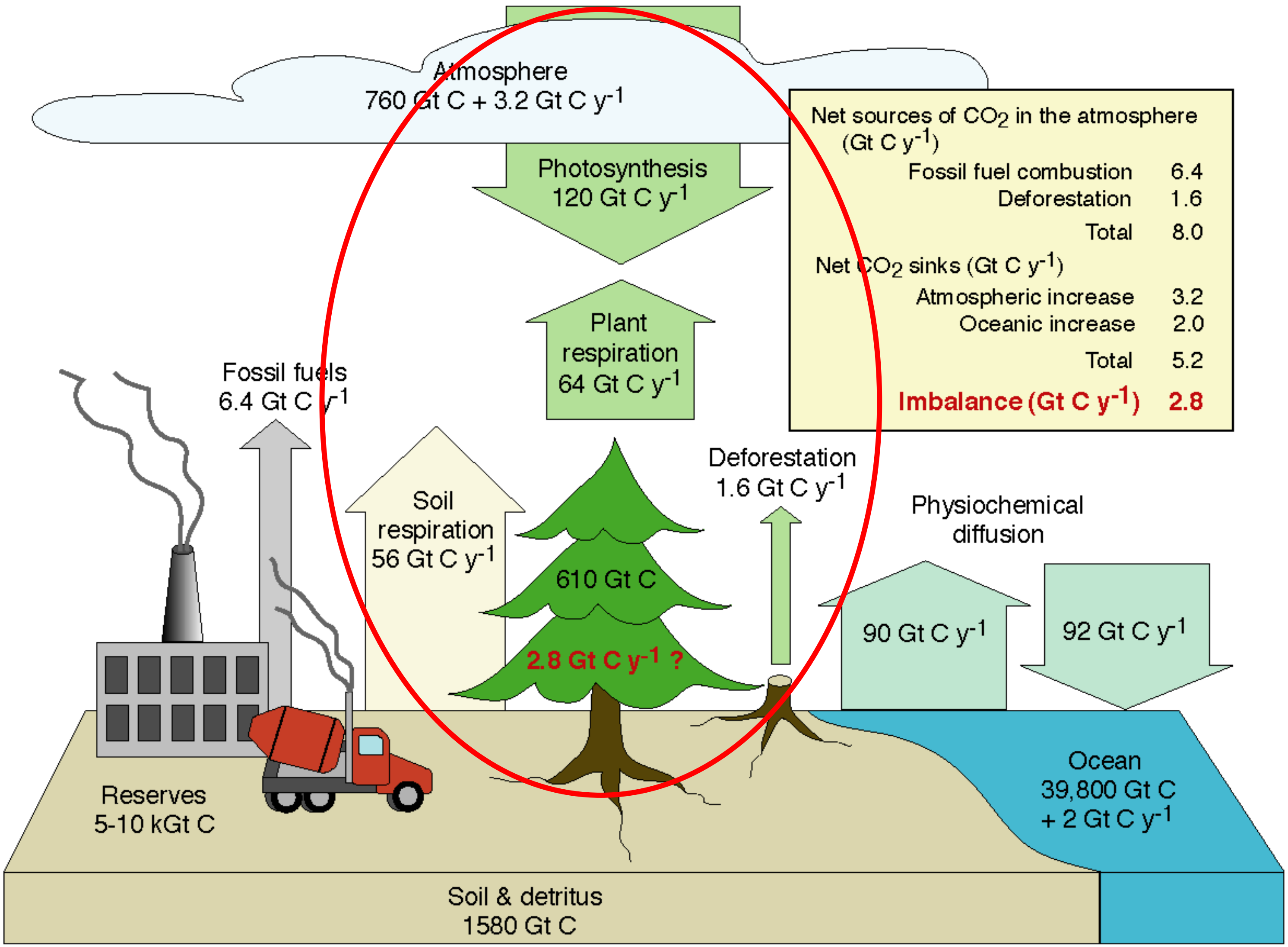
RIKEN International Symposium on Data Assimilation

27 Februar – 2 March 2017, Kobe, Japan

Outline

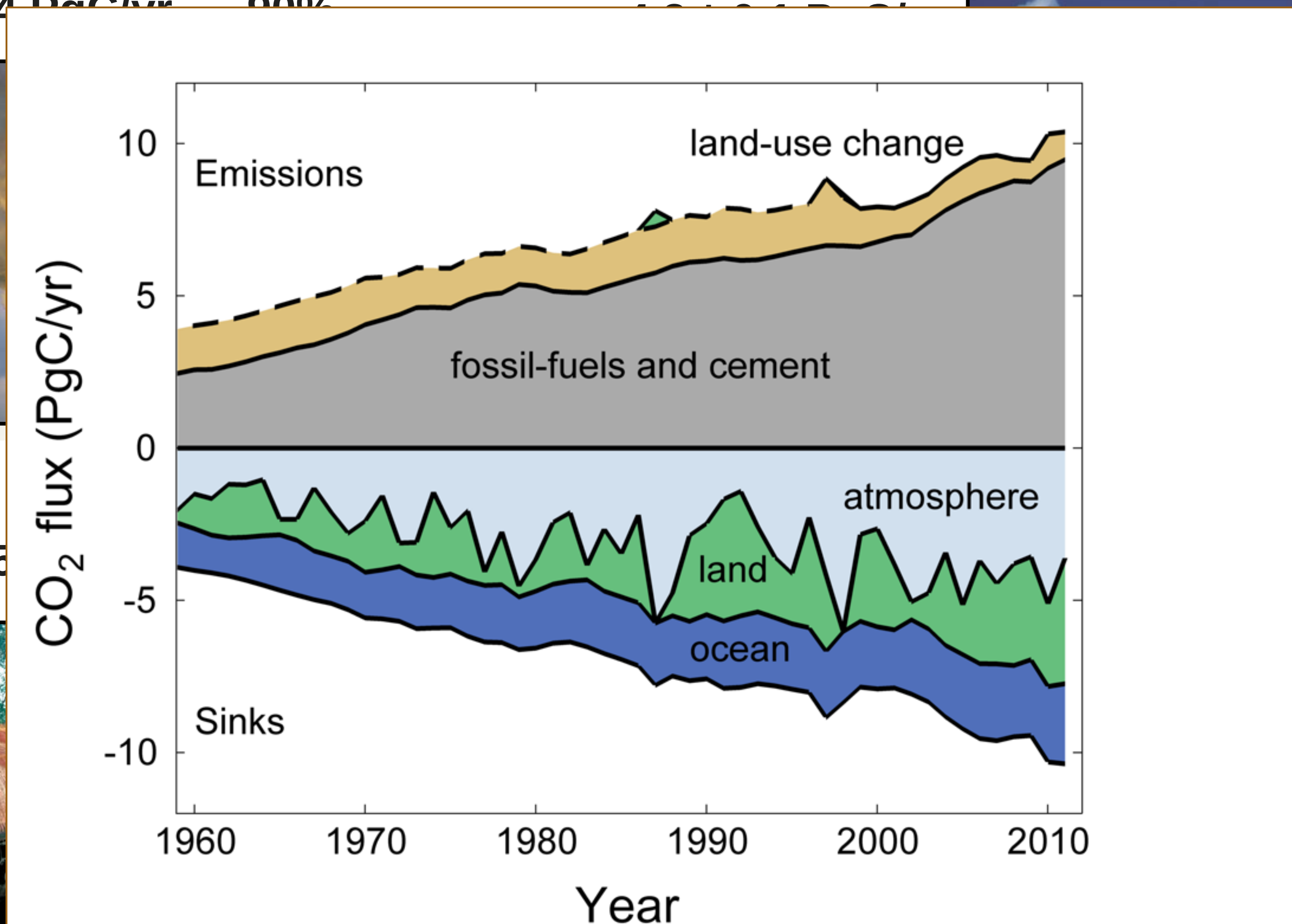
- Introduction
- Carbon Cycle Data Assimilation System
- Multiple constraints
- Non-convergence problem
- Conclusions

The Global Carbon Cycle



Global Carbon Budget

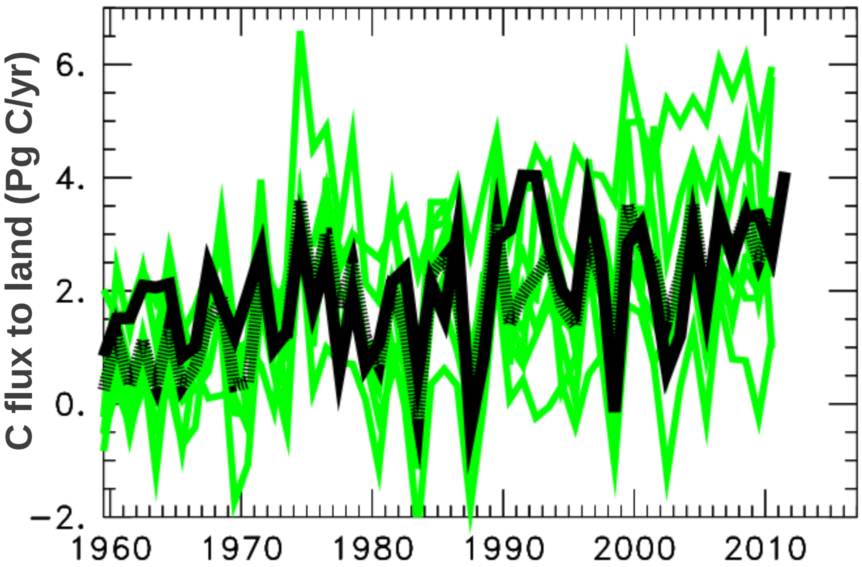
8.3±0.4



1.0±0.5

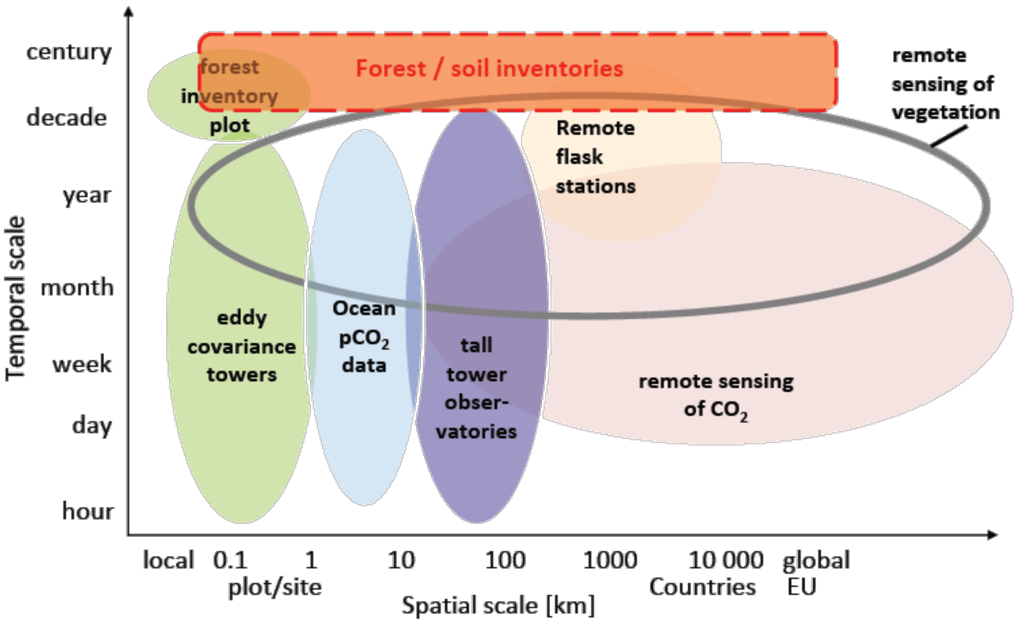
The case for data assimilation

Large uncertainty from land to predict C-balance (GCP)



Le Quéré et al. 2013

Available Observations



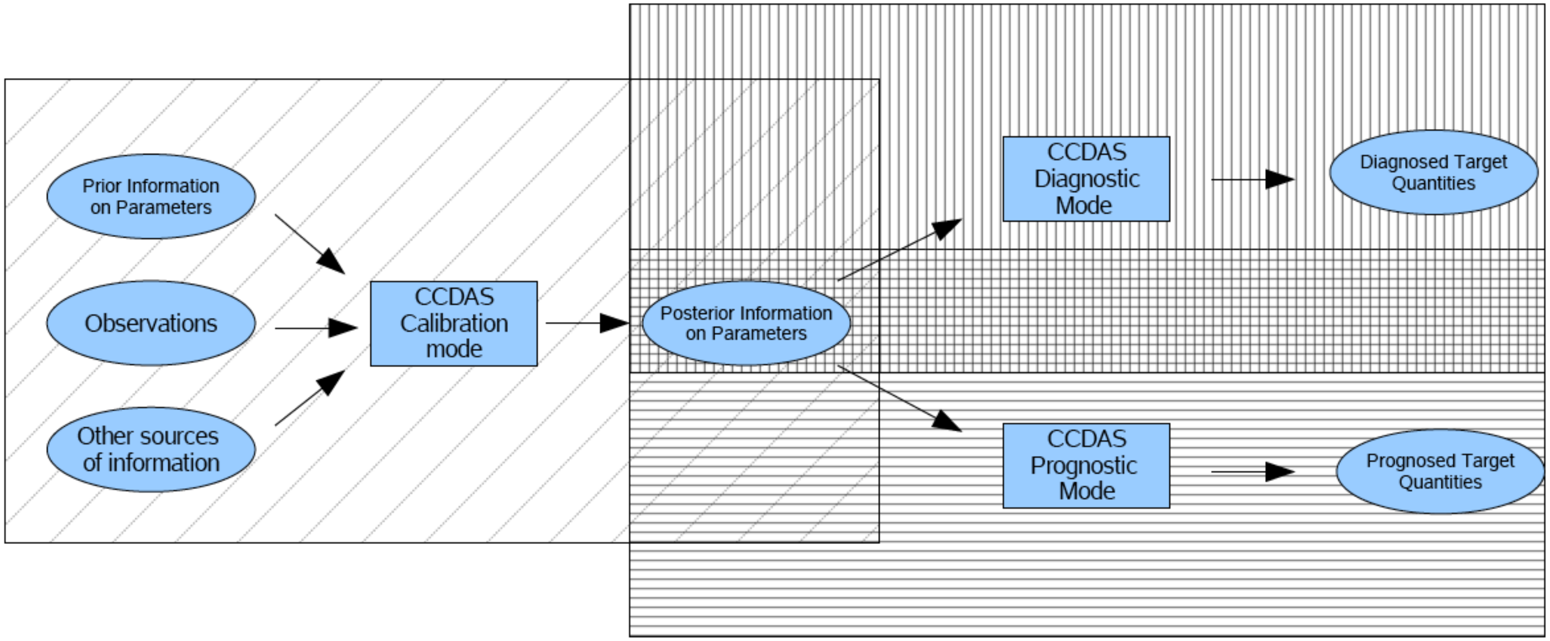
⇒ Carbon Cycle Data Assimilation System

= ecophysiological constraints from forward modelling

+ observational constraints from inverse modelling

CCDAS methodology

- Based on process-based terrestrial ecosystem model (BETHY)
- Optimizing parameter values (~100) based on gradient info
- Hessian (2^{nd} deriv.) to estimate posterior parameter uncertainty
- Error propagation by using linearised model

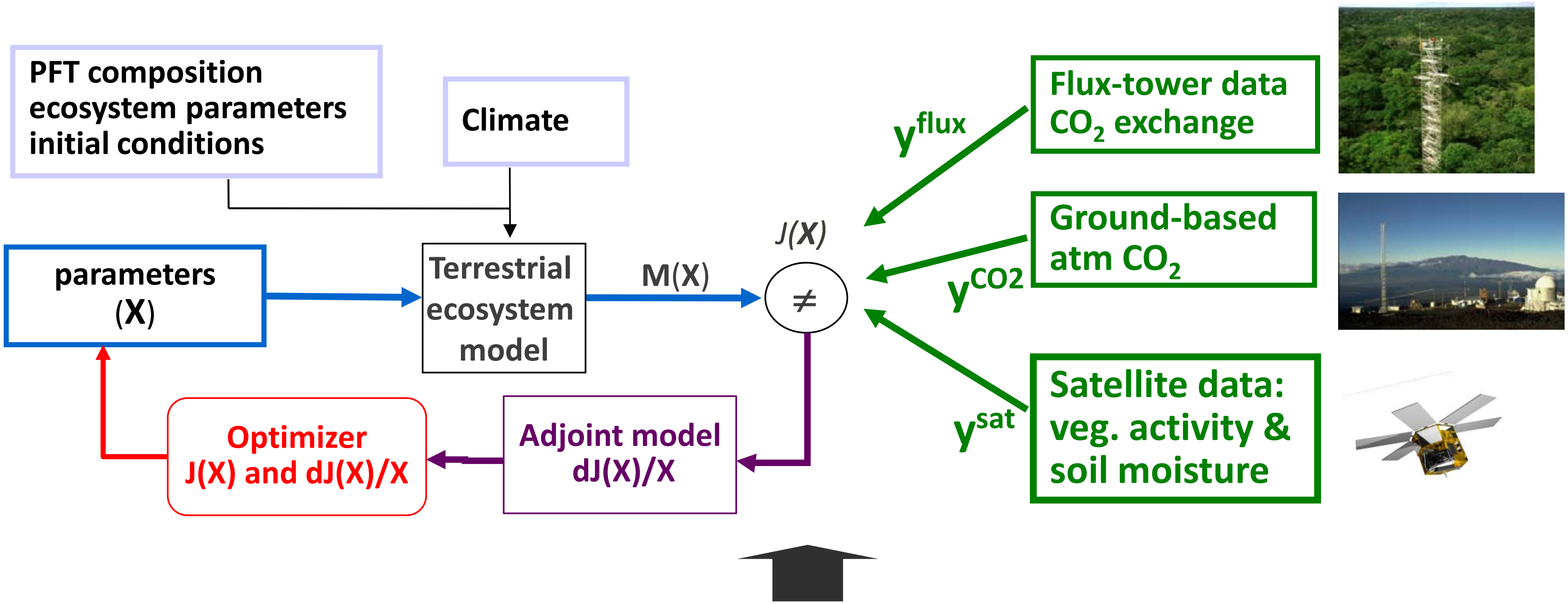


Scholze et al. (2007)

Process parameters

- Process parameters are invariant in time
- Parameterisations in biological systems are often based on (semi-)empirical relationships -> no universal/fundamental theory as in physical systems
- Parameters are often plant species specific but model lumps together many species into a plant functional and this upscaling process is highly uncertain

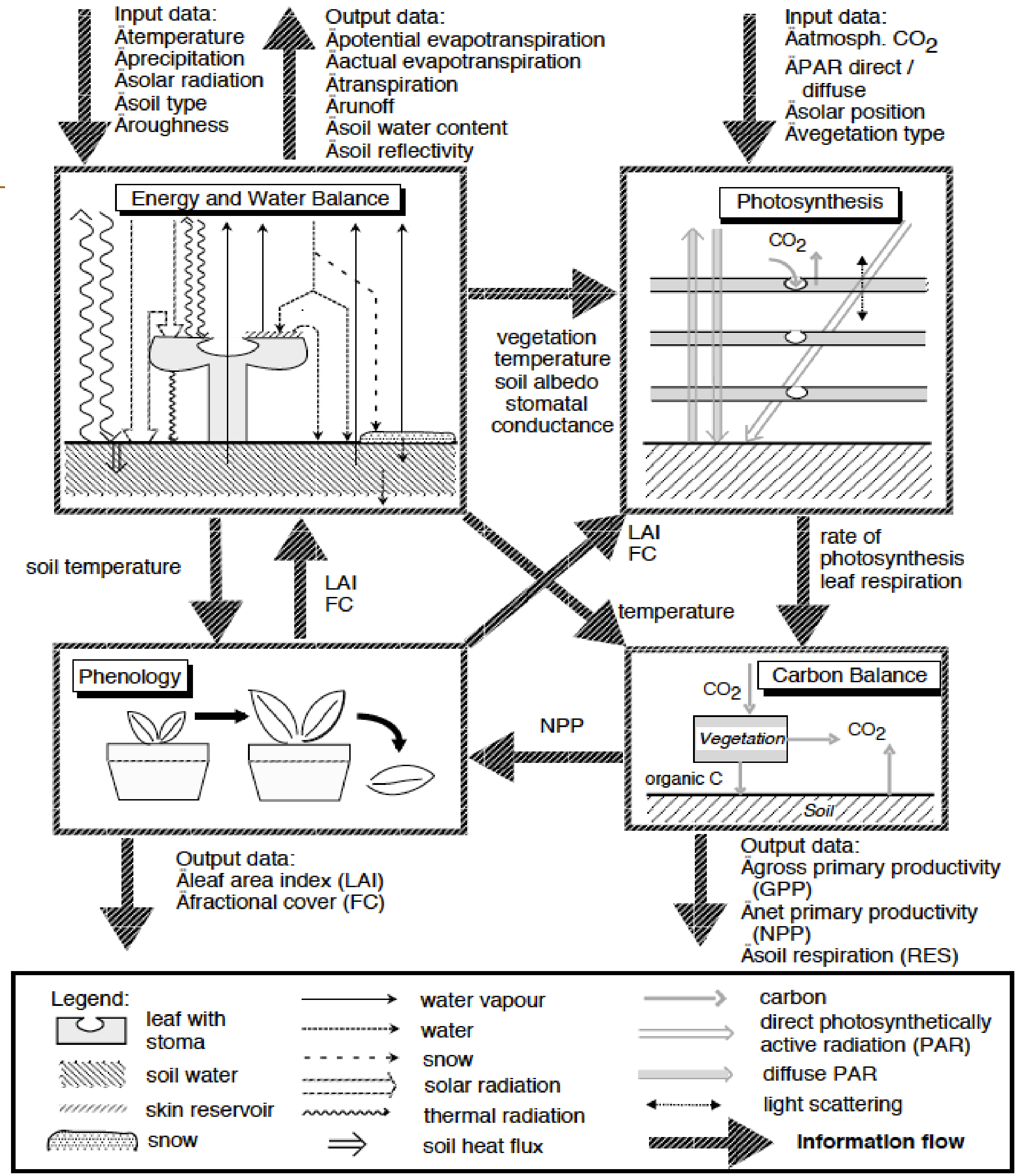
CCDAS approach



Cost function: $J(x) = \frac{1}{2}\left[\sum (y - M(x))^t C_y^{-1} (y - M(x)) + (x - x_p)^t C_p^{-1} (x - x_p)\right]$

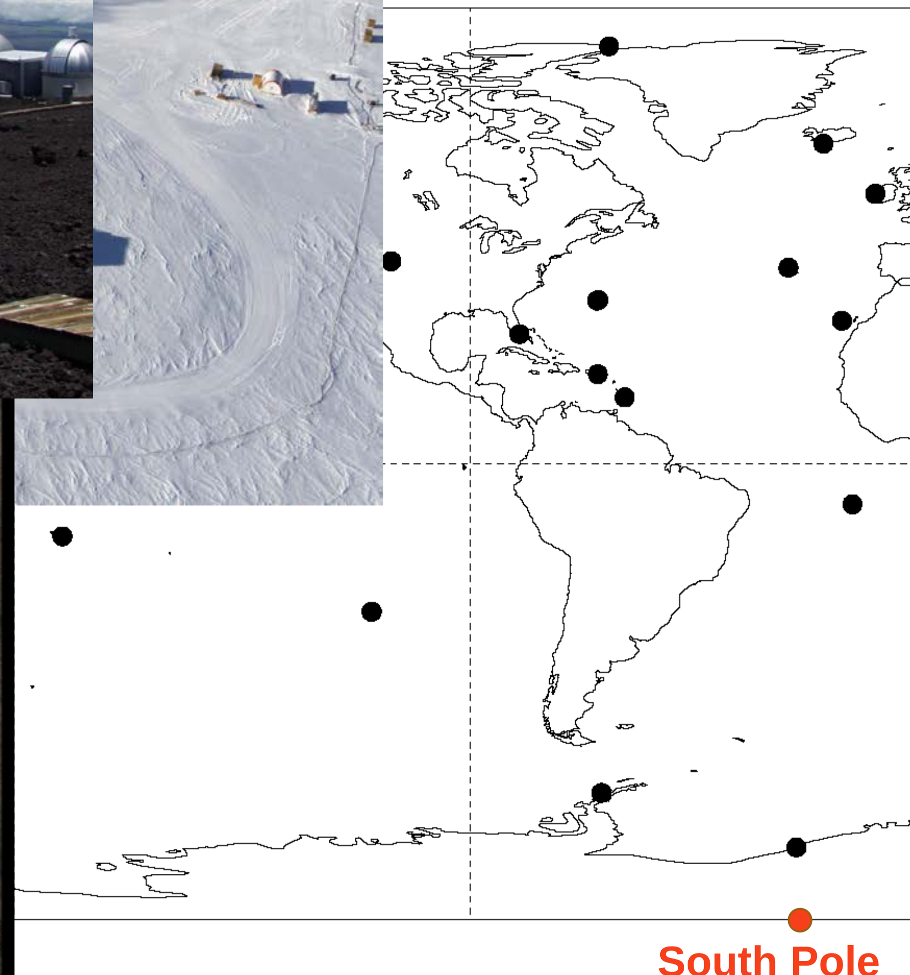
Need to define the error matrices C_y^{-1}, C_p^{-1}

BETHY

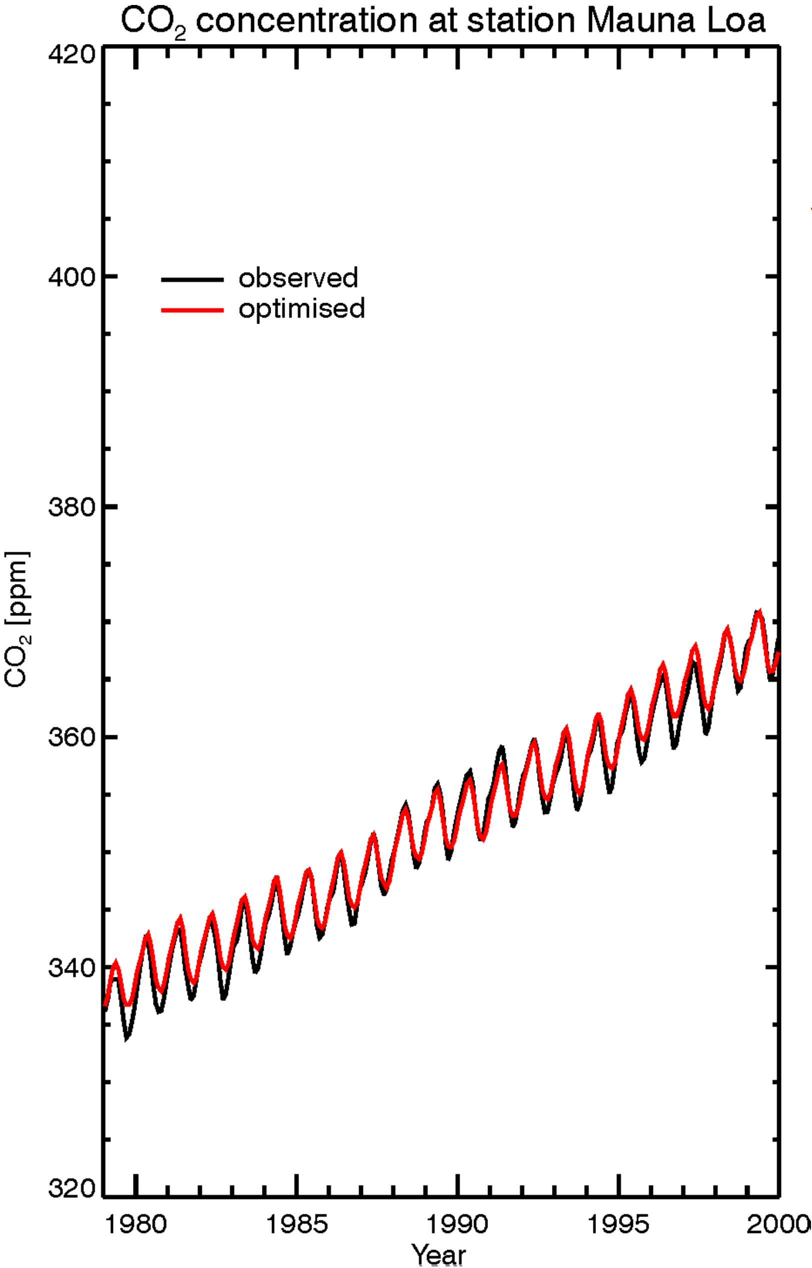


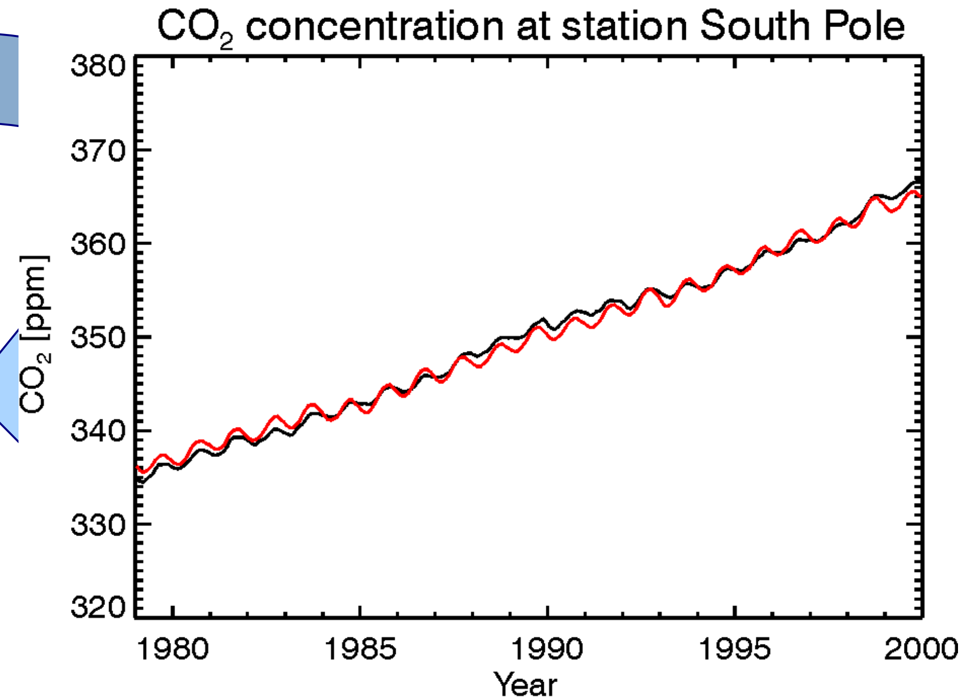
Knorr (2000)

on



Single constraint atm. CO_2: data fit





Rayner et al., 2005

Posterior uncertainties on parameters

Inverse Hessian of cost function approximates posterior uncertainties

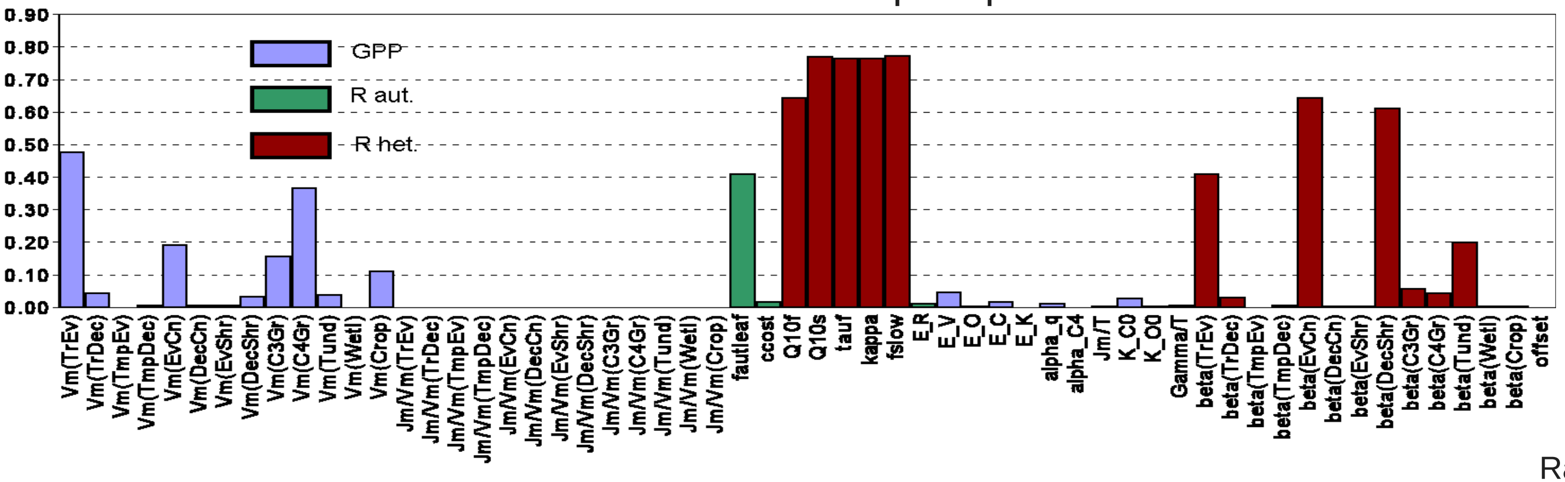
$$\mathbf{C}_p \approx \left\{ \frac{\partial^2 J\left(\vec{p}_{\text{opt}}\right)}{\partial p_{i,j}^2} \right\}^{-1}$$

examples:

μmol/m ^{2}s μmol/m ^{2}s % % error covariance

0.31 0.80

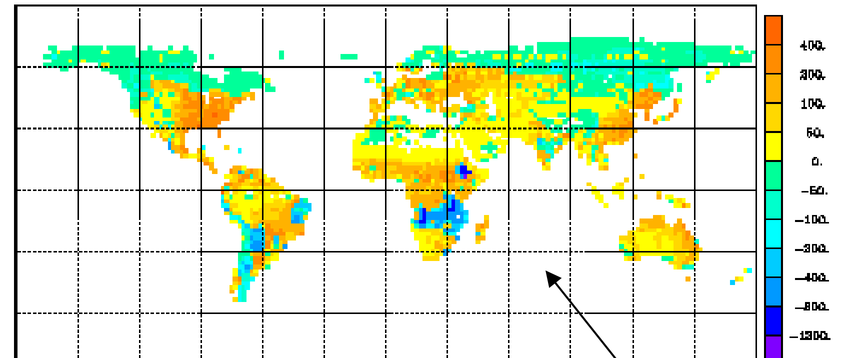
Relative Error Reduction $1-\sigma_{opt}/\sigma_{prior}$



Rayner et al., 2005

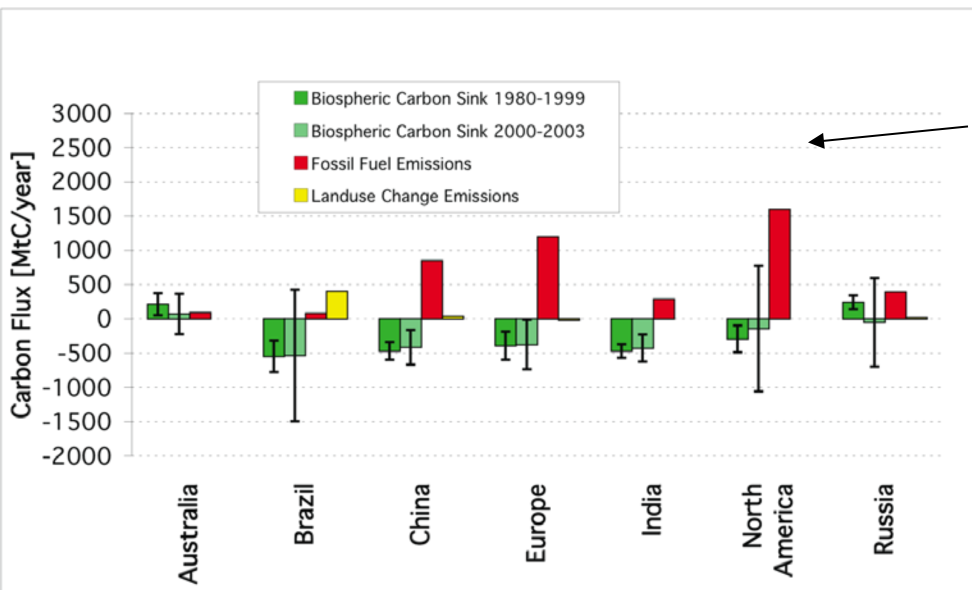
Net C fluxes and their uncertainties

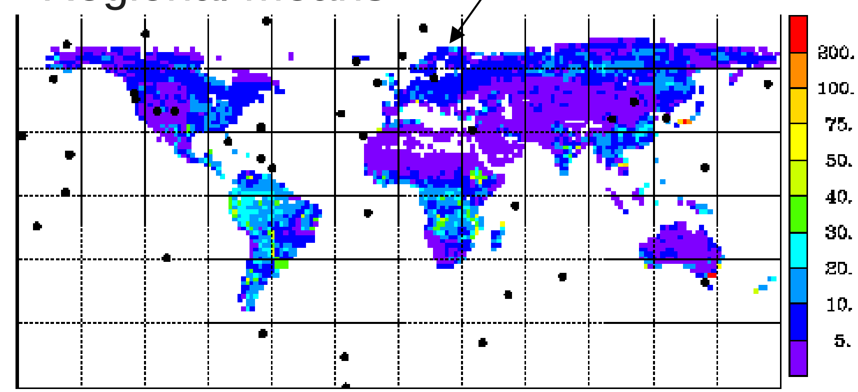
$$\mathbf{C}_y = \left(\frac{\partial y_i(\vec{p}_{\text{opt}})}{\partial p_j}\right)\mathbf{C}_p\left(\frac{\partial y_i(\vec{p}_{\text{opt}})}{\partial p_j}\right)^T$$



Examples for diagnostics:

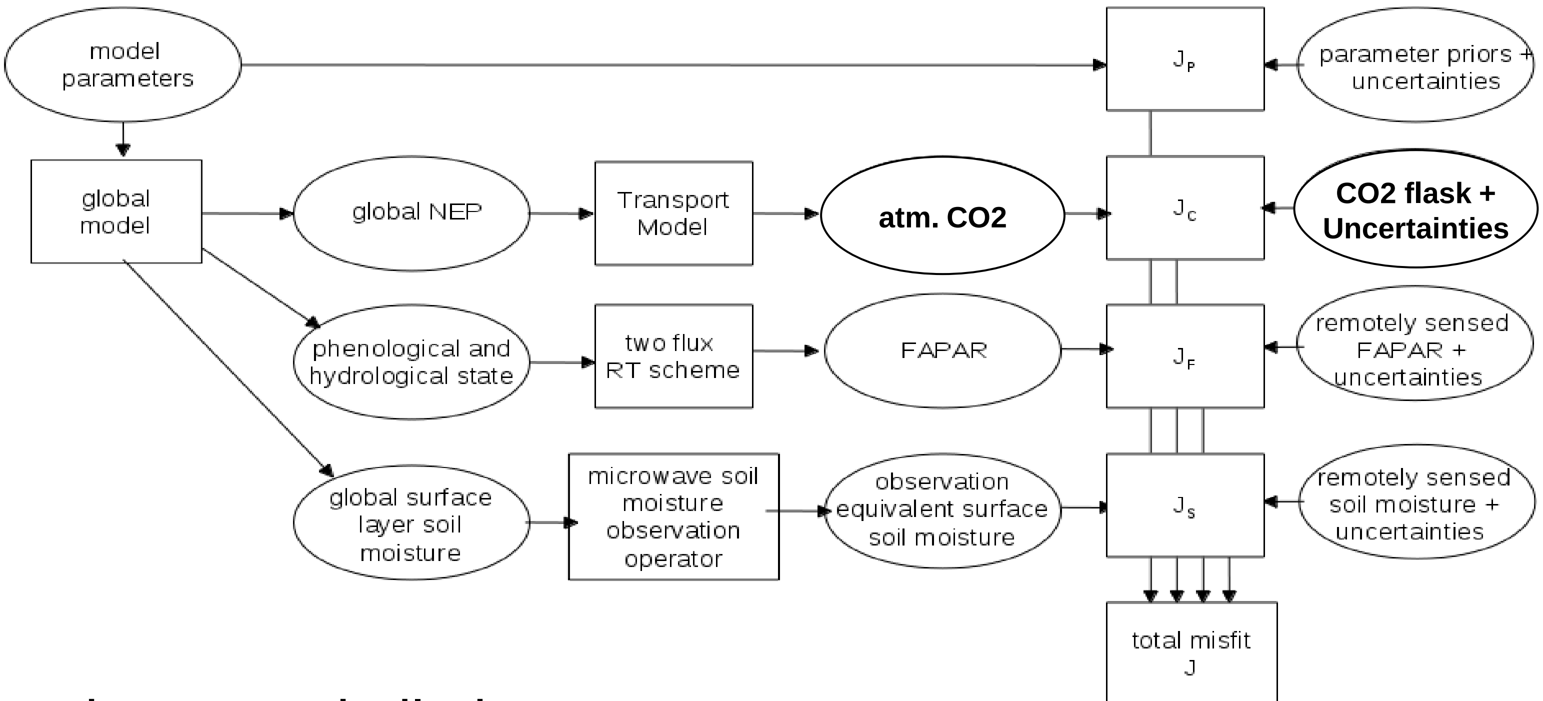
- Long term mean fluxes to atmosphere ($gC/m^2/year$) and uncertainties
- Regional means





Rayner et al., 2005

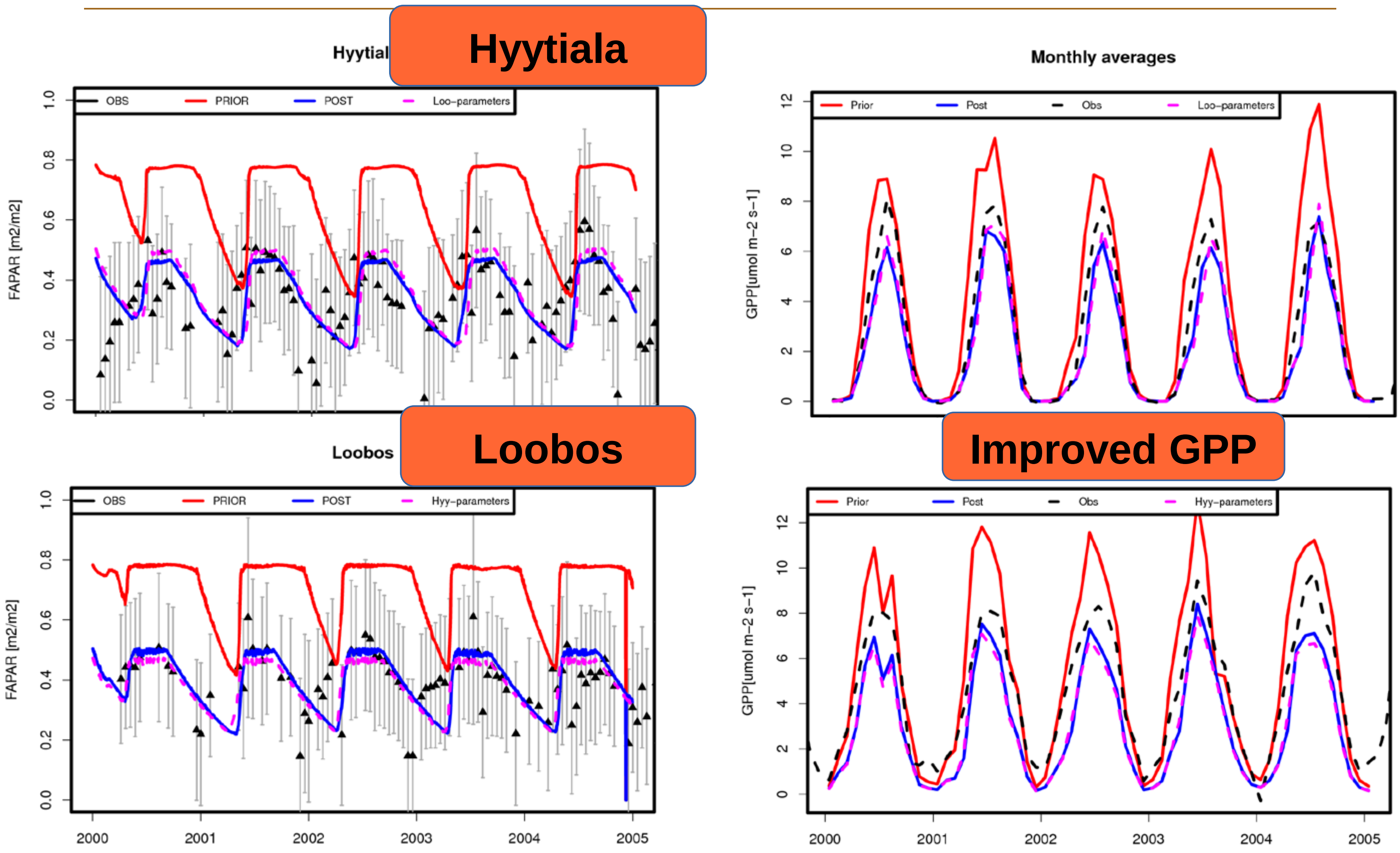
Multiple constraints, 1st example



Consistent Assimilation :

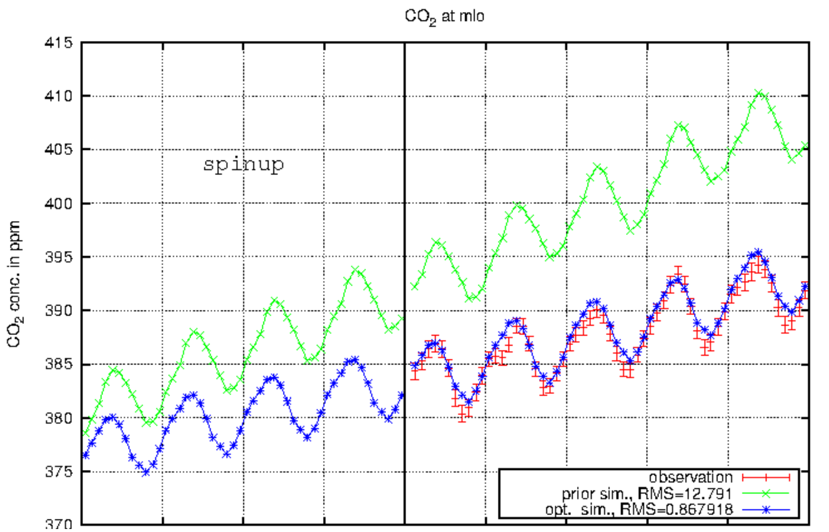
- all data streams jointly
- in a single long assimilation window

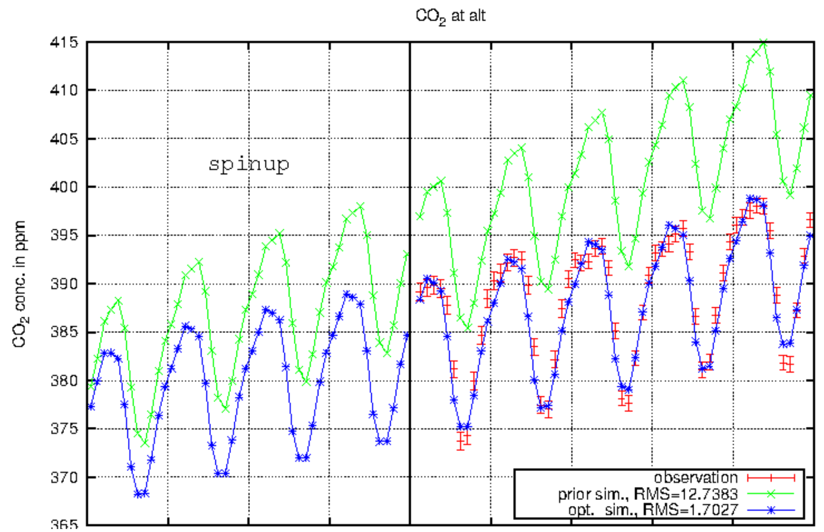
Transfer of information in space and between variables

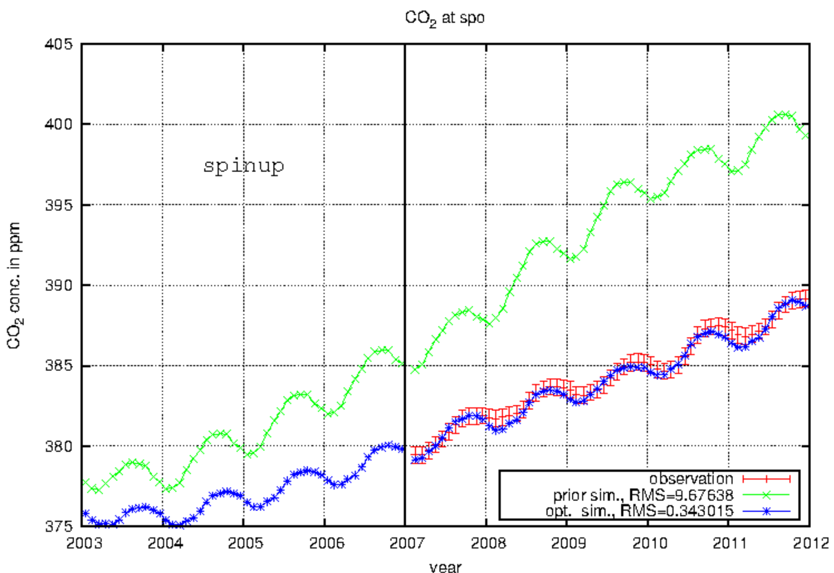


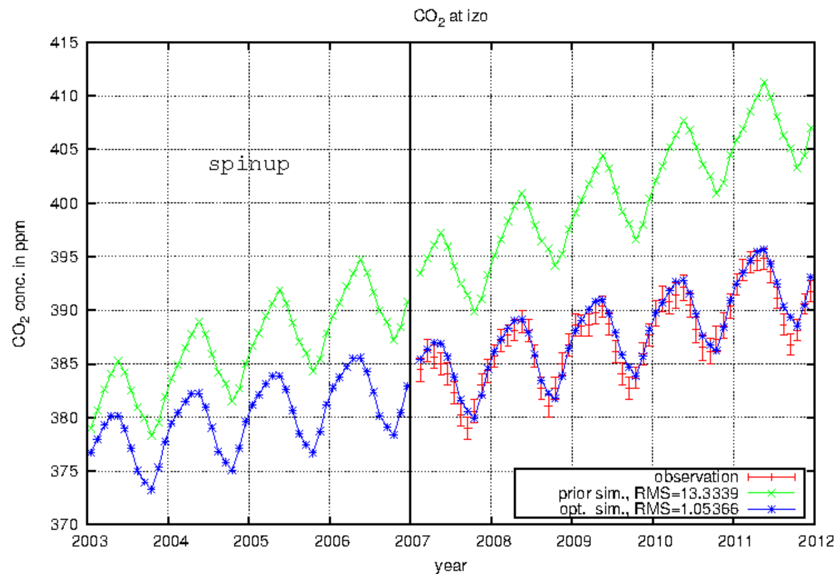
Schuermann et al. (2016)

Transfer of information in space

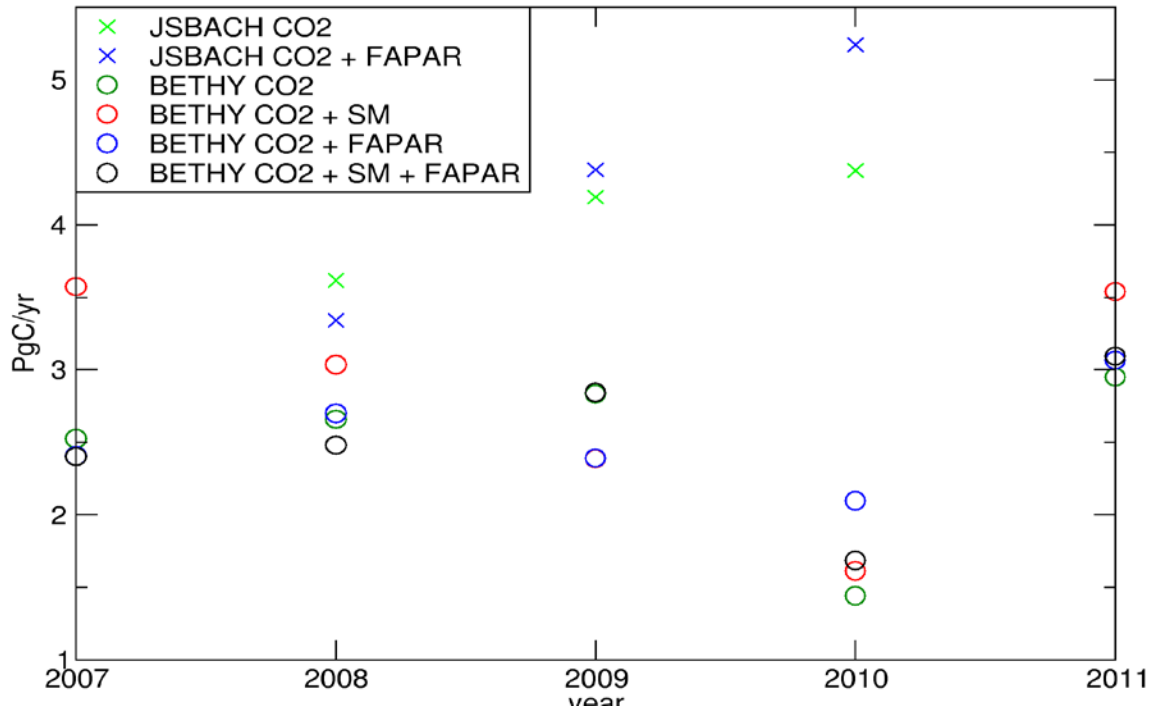








Posterior land uptake



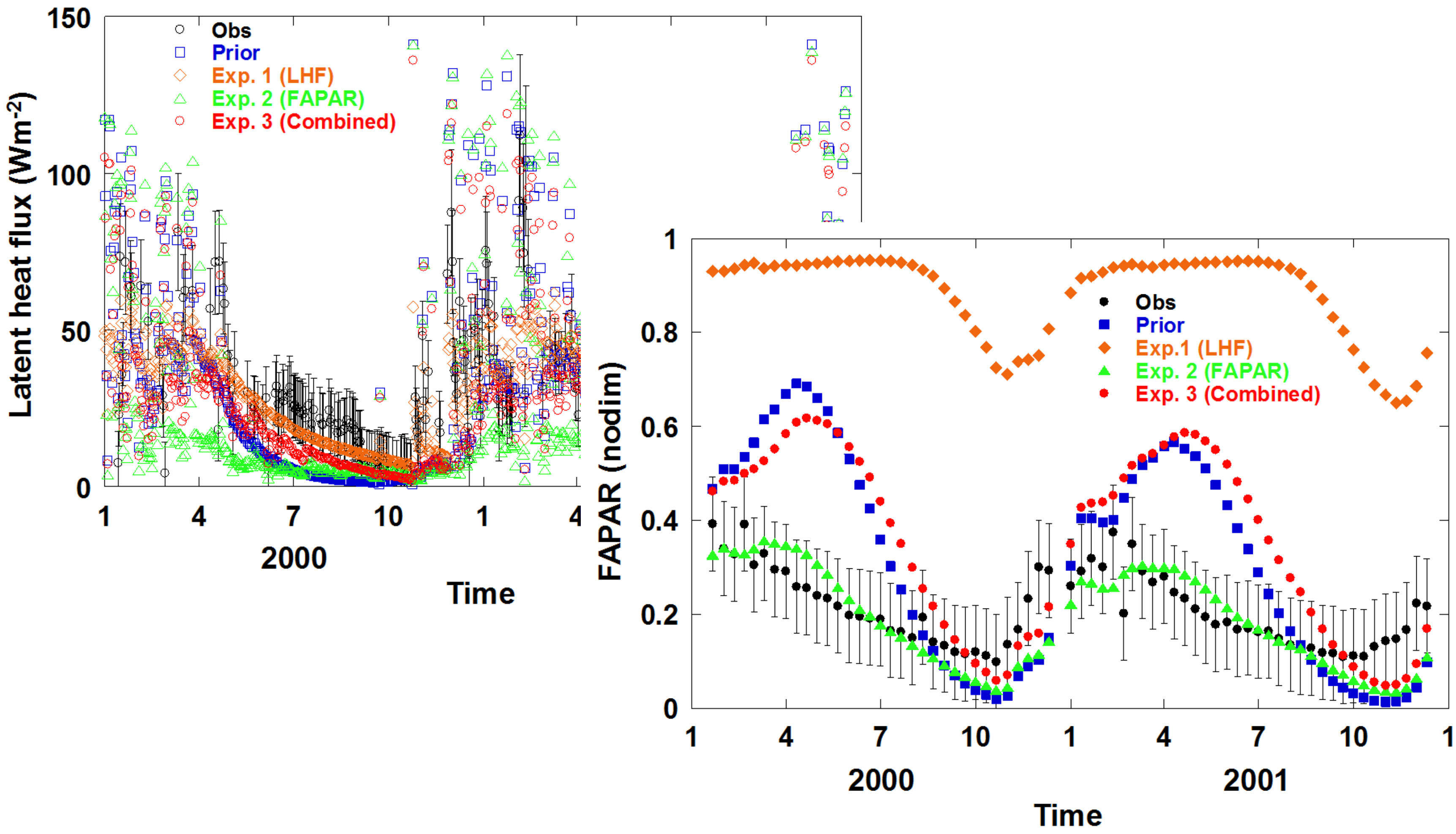
Annual land uptake from 2007 to after assimilation of different combinations of data streams (SM: soil moisture, FAPAR: fraction of absorbed photosynthetic active radiation)

Multiple constraints, 2nd example

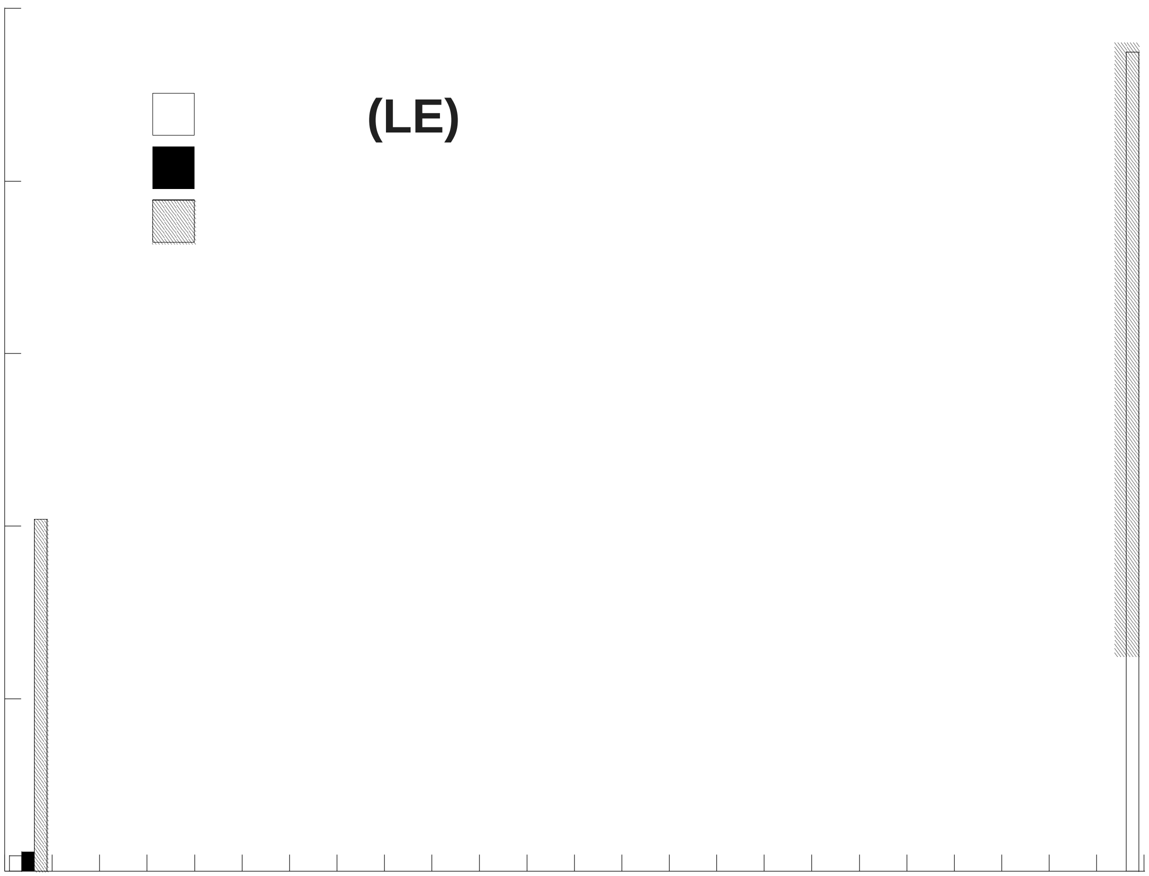
- Simultaneous assimilation of two data streams at site level Maun, Botswana over 2 years (2000-2001)
- Daily LE fluxes, no gap-filled data (464 observations)
- Satellite FAPAR observations, 10-daily (70 observations)
- Optimization of 24 model parameters
- 2 Plant Functional Types: tropical broadleaf deciduous tree and C4 grass

Kato et al. (2012), *Biogeosciences*

Fit to LE and FAPAR data



Posterior parameter uncertainty



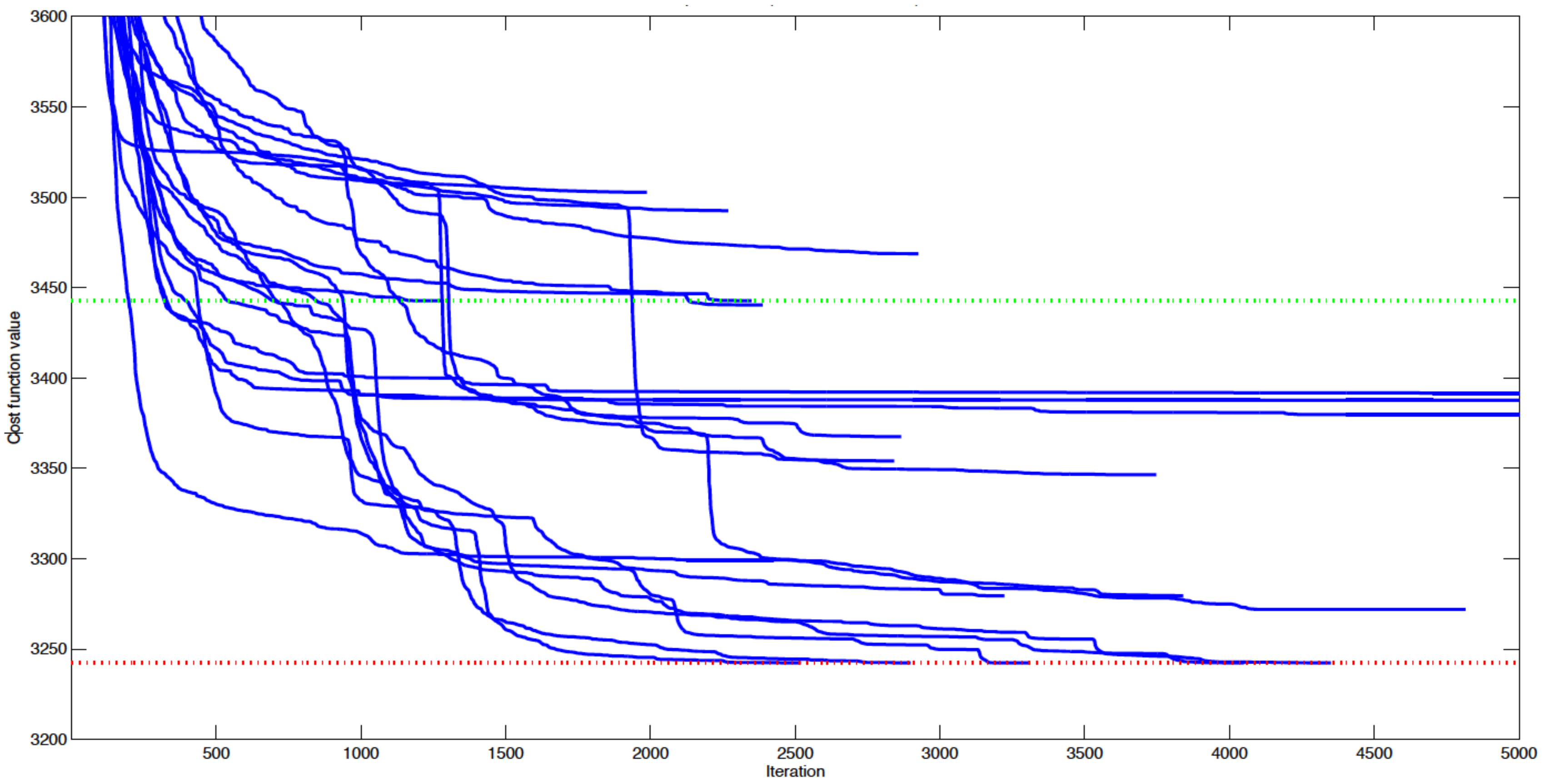
Parameter

Relative reduction:

$$1-\frac{\sigma_{post}}{\sigma_{prior}}$$

Robustness of optimal solution

- Different starting point -> same minimum?
- Local minimum vs global minimum
- Non-convergence problems



Global minimum?

Comparison of two Minima

1. $c = 3242.32 = 2740.16(\text{observations}) + 502.16(\text{parameter})$
2. $c = 3442.64 = 3213.68(\text{observations}) + 228.96(\text{parameter})$

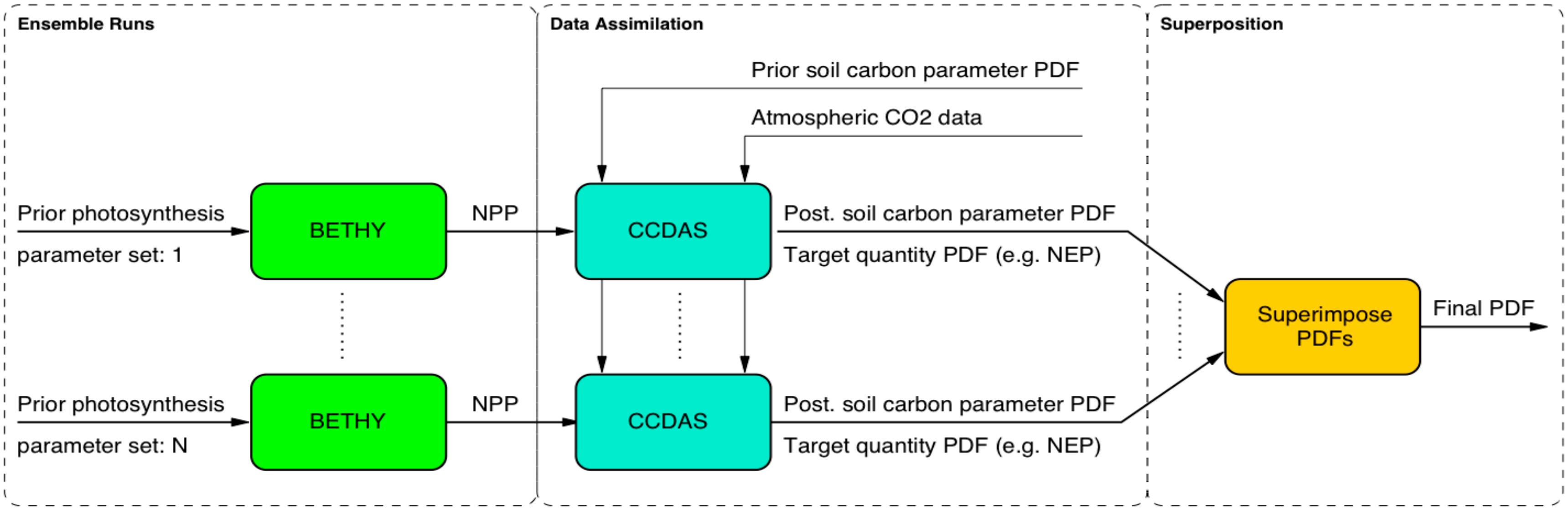
- Parameter values are totally different for both minima

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}
1	0.96	0.42	1.40	0.69	0.47	0.89	1.26	0.23	2.44	0.57	1.02	1.47	-0.26
2	0.99	0.33	-0.17	1.05	1.09	0.26	0.93	1.96	2.20	0.97	0.63	0.93	-0.41

Find sub-set of model which converges using 4D-var, use ensemble for remaining model

-> combined ensemble-adjoint optimisation

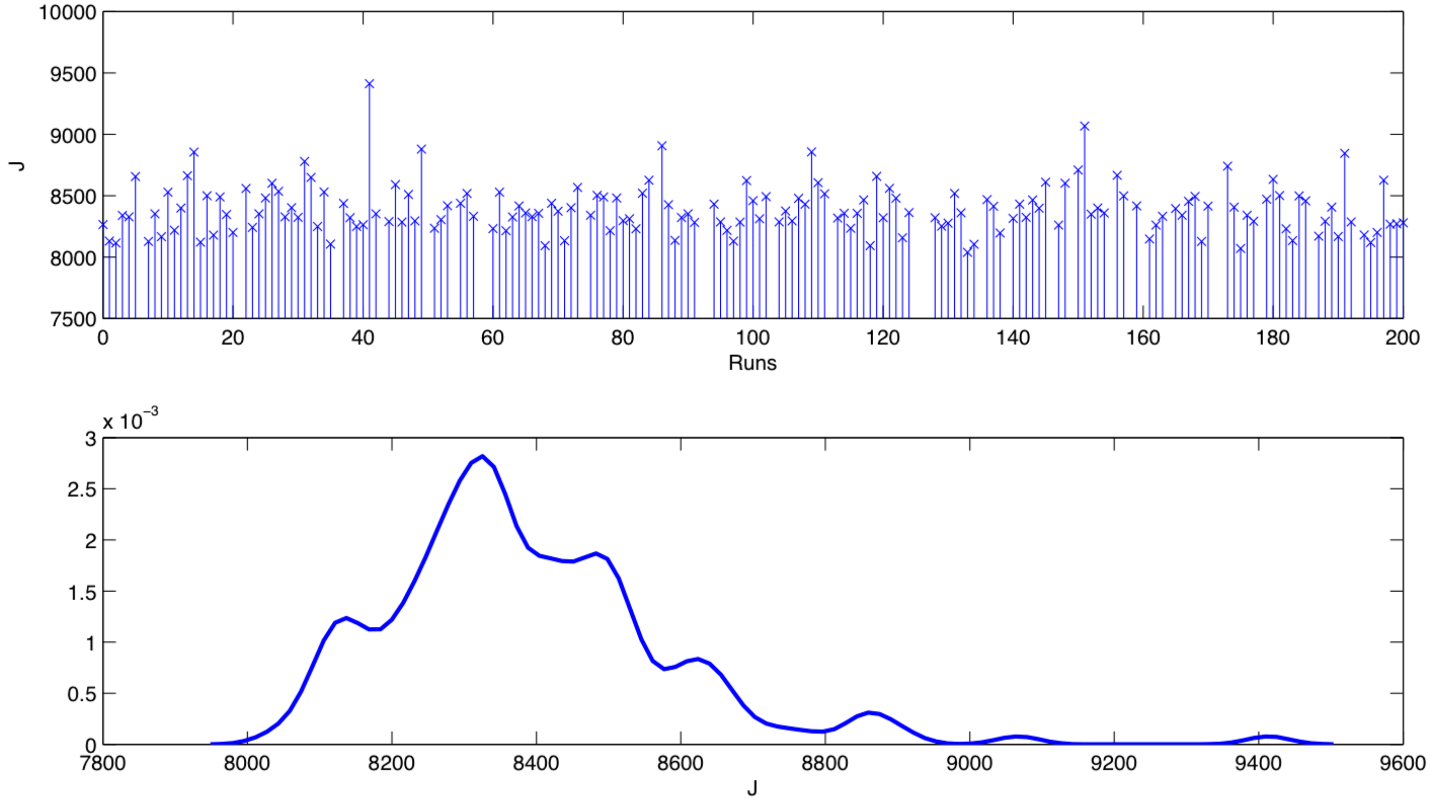
Combined ensemble-adjoint optimisation



- Combination of ensemble runs with CCDAS
- Photosynthesis (NPP), hydrology (soil moisture) and phenology (LAI) from stand-alone BETHY ensemble runs (200)
- Optimisation of 19 soil carbon parameters for each run over 25 years
- Propagation of the posterior soil carbon parameter uncertainties provides PDF for NEP for each run

Results for a test case: individual cost functions

170 out of 200 member ensemble kept, 30 runs are discarded due to non-convergence or non-physical posterior parameter values.



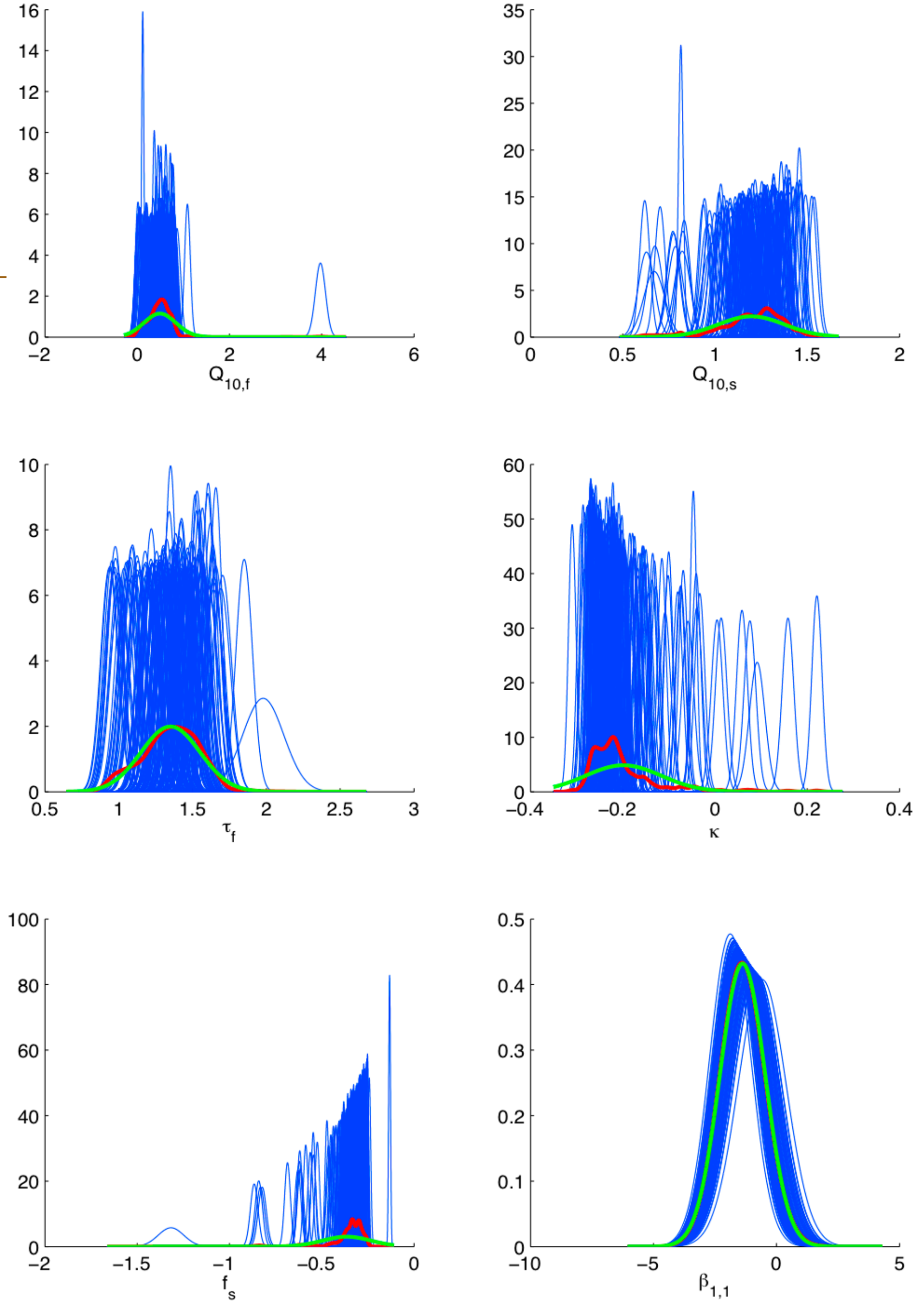
Test case: posterior parameter unc.

Blue:
Individual PDFs obtained from CCDAS using input from ensemble

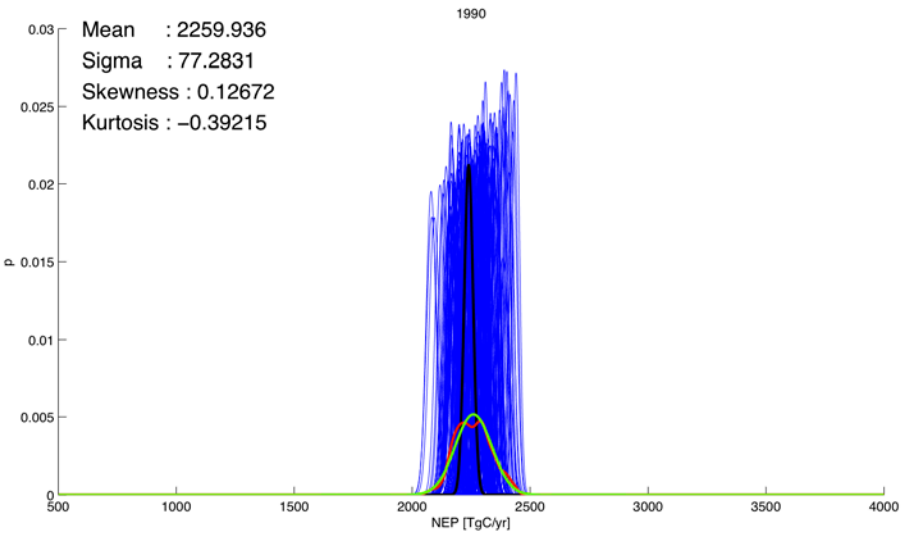
Red:
Superimposed PDF

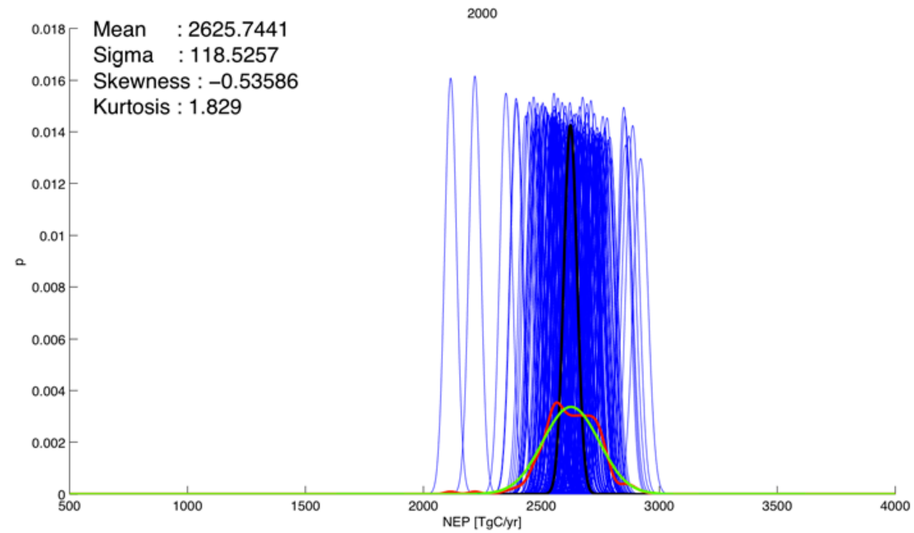
$$PDF^s = \frac{1}{N}\int_1^N PDF^i$$

Green:
Gaussian approximation



Results for a test case: global net C flux





Blue: Individual PDFs obtained from CCDAS using input from ensemble runs

Red: Superimposed PDF

Green: Gaussian Approximation

Black: Base case PDF

Summary

- CCDAS: Mathematically rigorous combination of process understanding and observations for carbon cycling
- Provides integrated view on global carbon cycle on all variables that can be simulated by the model at any time and place
- Regional scale carbon budgets based on combination of multiple data streams and process-based simulations
- Added value of data streams quantified through uncertainty reduction
- Can be extended to include further data streams, either for assimilation or validation
- Hierarchical Parameter Estimation
  - Combination of ensemble runs and 4D-Var in a data assimilation system
  - Superimpose individual PDFs for parameters and target quantities to obtain final PDF

Fit against GPP