Process parameter optimisation in terrestrial carbon cycle models: the curse of the forecast

M. Scholze, Physical Geography and Ecosystem Analysis, Lund University, Sweden

Terrestrial ecosystem models (TEMs) contain the coupling of many biogeochemical processes with a large number of parameters involved. These parameters are often based on (semi-)empirical relationships derived from local scale or laboratory experiments. In many cases those parameters are often plant species specific but the TEMs lump together many species into a plant functional type and apply those parameters on a larger (normally global) scale. This upscaling process is highly uncertain and leaves many of those parameters highly uncertain. In order to reduce the uncertainties, parameter estimation methods can be applied, which allow the model to be constrained against observations.

The Carbon Cycle Data Assimilation System (CCDAS) is one of these parameter estimation frameworks mapping modelled terrestrial ecosystem fluxes of CO_2 to the atmosphere and also capable to predict the evolution of these fluxes into the future. The main feature of CCDAS is its capability of deriving an optimal set of parameters for the underlying process based terrestrial biosphere model BETHY from assimilating atmospheric CO_2 concentration observations as well as other observations of the terrestrial carbon cycle representative for different temporal and spatial scales and processes (such as remotely sense vegetation greenness and eddy-covariance observations of latent heat fluxes). As a variational data assimilation scheme, CCDAS relies on first and second derivatives of the underlying model for estimating process parameters with uncertainty ranges. In a subsequent step these parameter uncertainties are mapped forward onto uncertainty ranges for predicted land-atmosphere exchange fluxes.

The results obtained from the consistent assimilation of multiple data streams emphasize the importance of integrating multiple data streams, as this allows for a more comprehensive assessment of model structures. If the model is not able to integrate the observations simultaneously this hints to either deficiencies in the process formulation or observational biases. The need for a mass conserving system in order to allow the calculation of annual CO_2 budgets and the inclusion of parameter for slowly evolving processes in the assimilation system may not be compatible with a short-term (~days) forecasting system but likely also not needed for these forecasts.

Process parameter optimisation in terrestrial carbon cycle models: the curse of the forecast

Marko Scholze

Lund University, Sweden

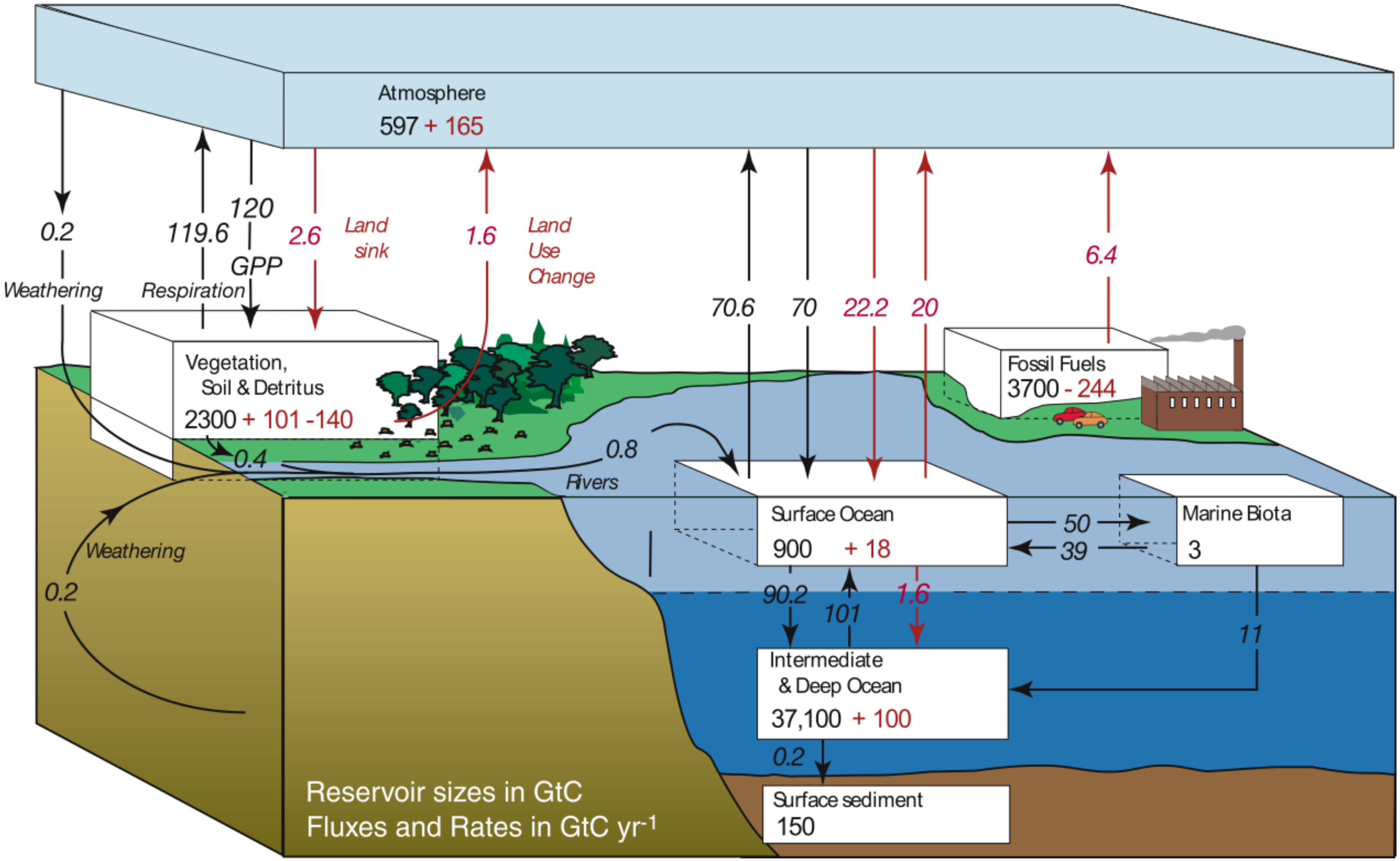
ECMWF workshop on parameter estimation and inverse modelling,

22 October 2013

Outline

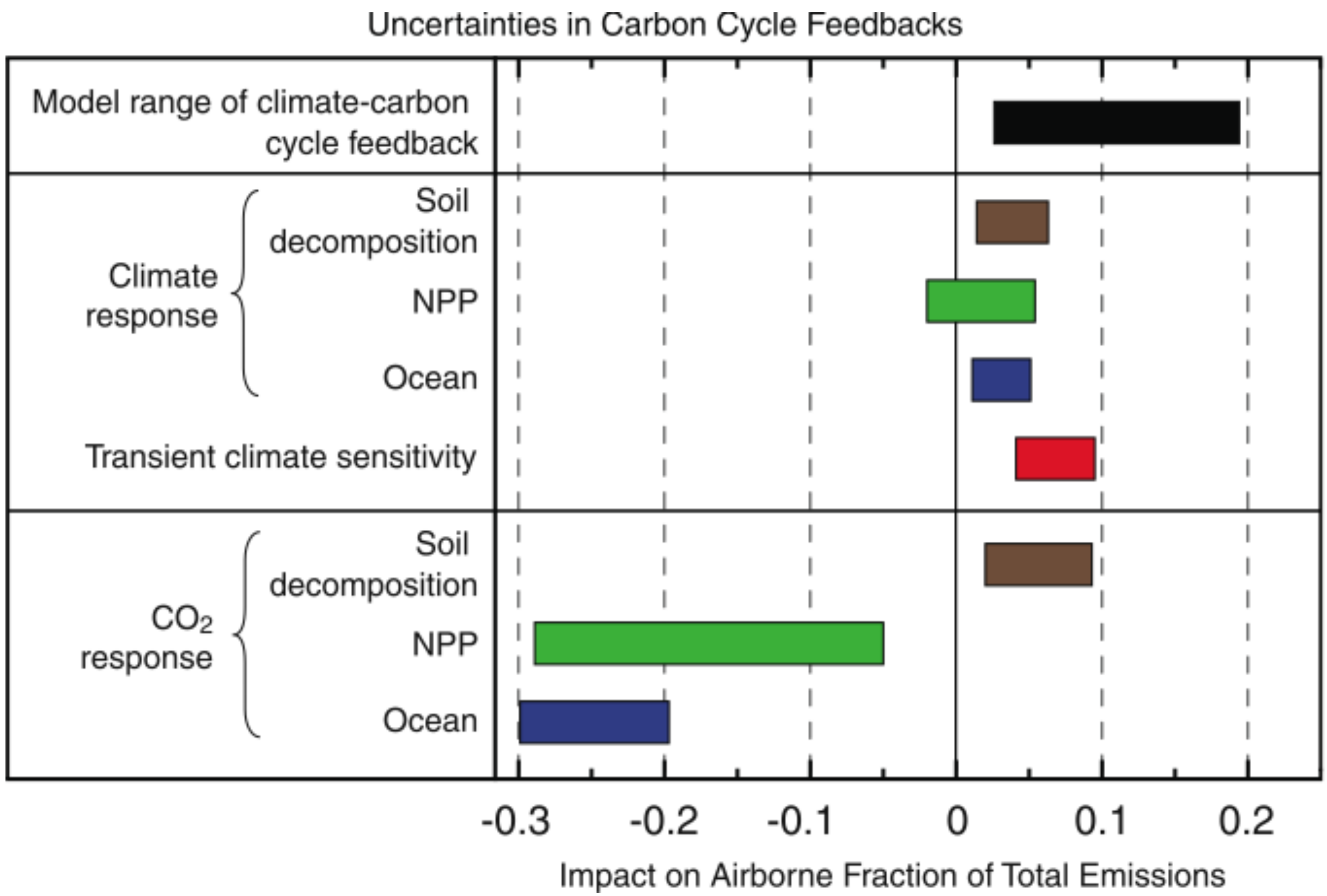
- Why parameter optimisation
- The Carbon Cycle Data Assimilation System
- Some thoughts on consistent parameter optimisation within an operational forecasting environment -> the curse of the forecast…

The global carbon cycle



IPCC Climate Change 2007: The Physical Science Basis

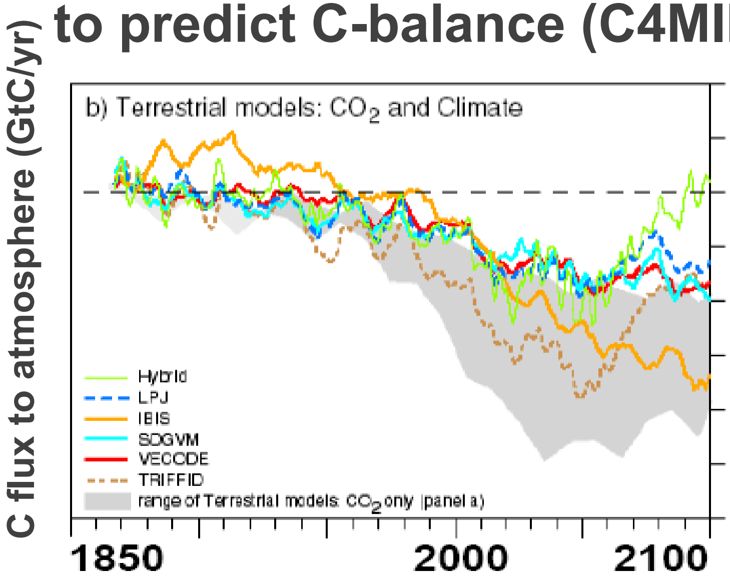
Carbon Cycle-Climate feedback: breakdown of uncertainties



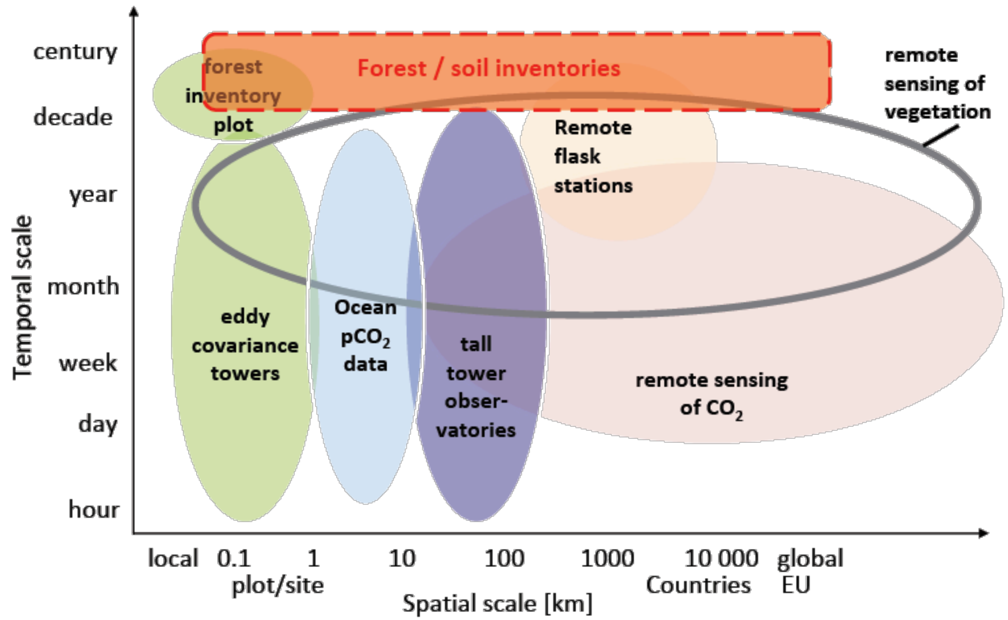
IPCC Climate Change 2007: The Physical Science Basis

The case for data assimilation

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



Available Observations

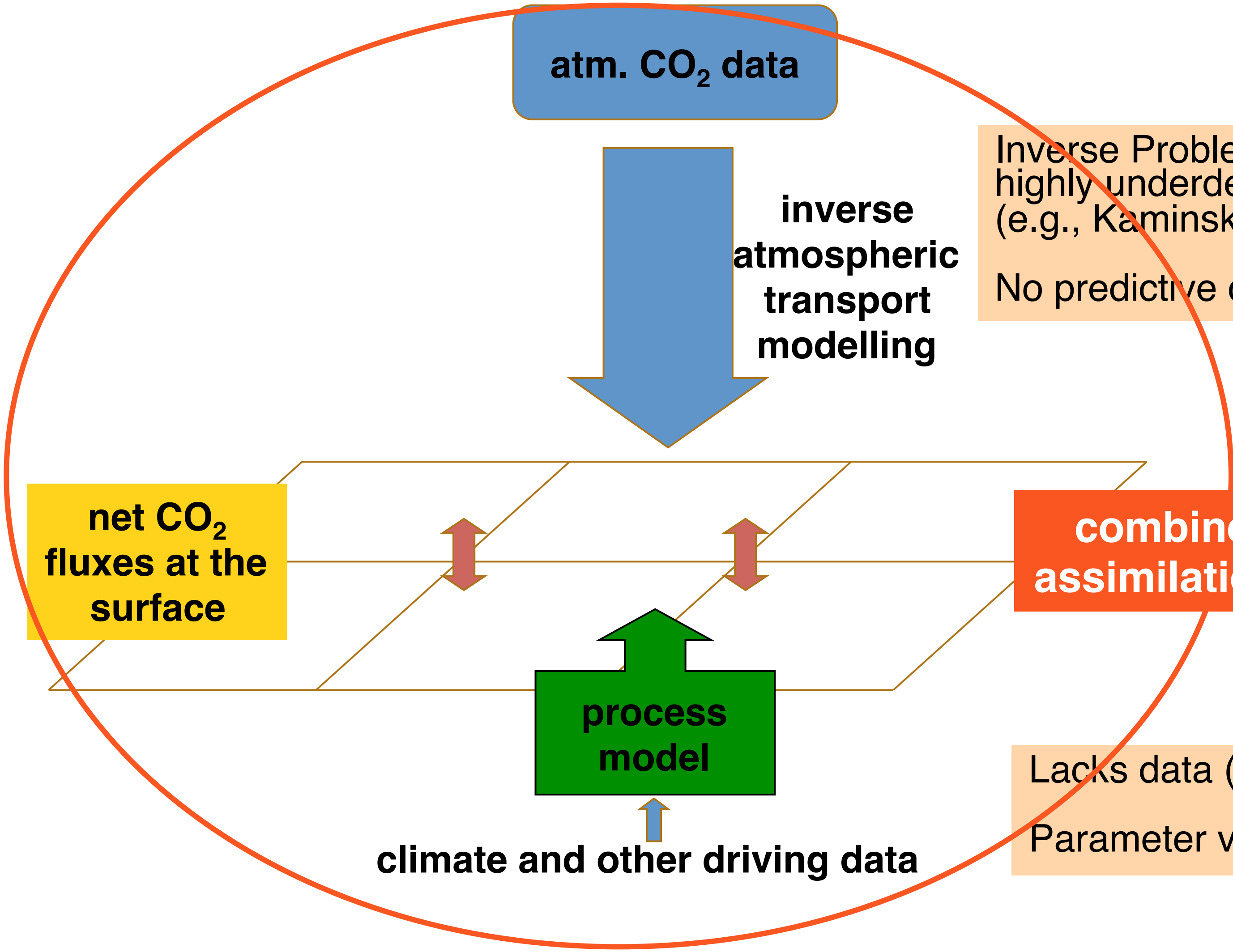


$\Rightarrow$ Carbon Cycle Data Assimilation System

= ecophysiological constraints from forward modelling

+ observational constraints from inverse modelling

Top down / Bottom up



Inverse Problem
highly underdetermined
(e.g., Kaminski and Heimann, 2001)

No predictive capability

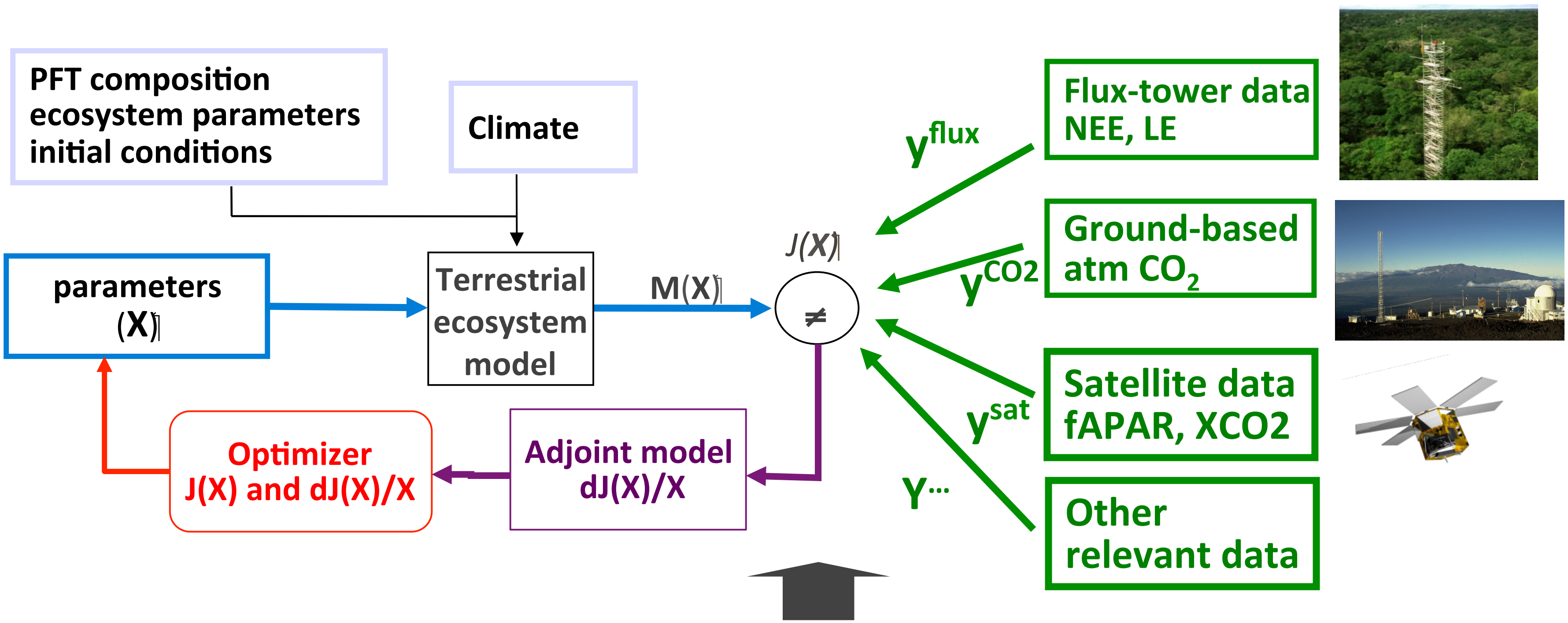
Lacks data (esp. atm.) constraint

Parameter validity

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 type
- Upscaling process is highly uncertain

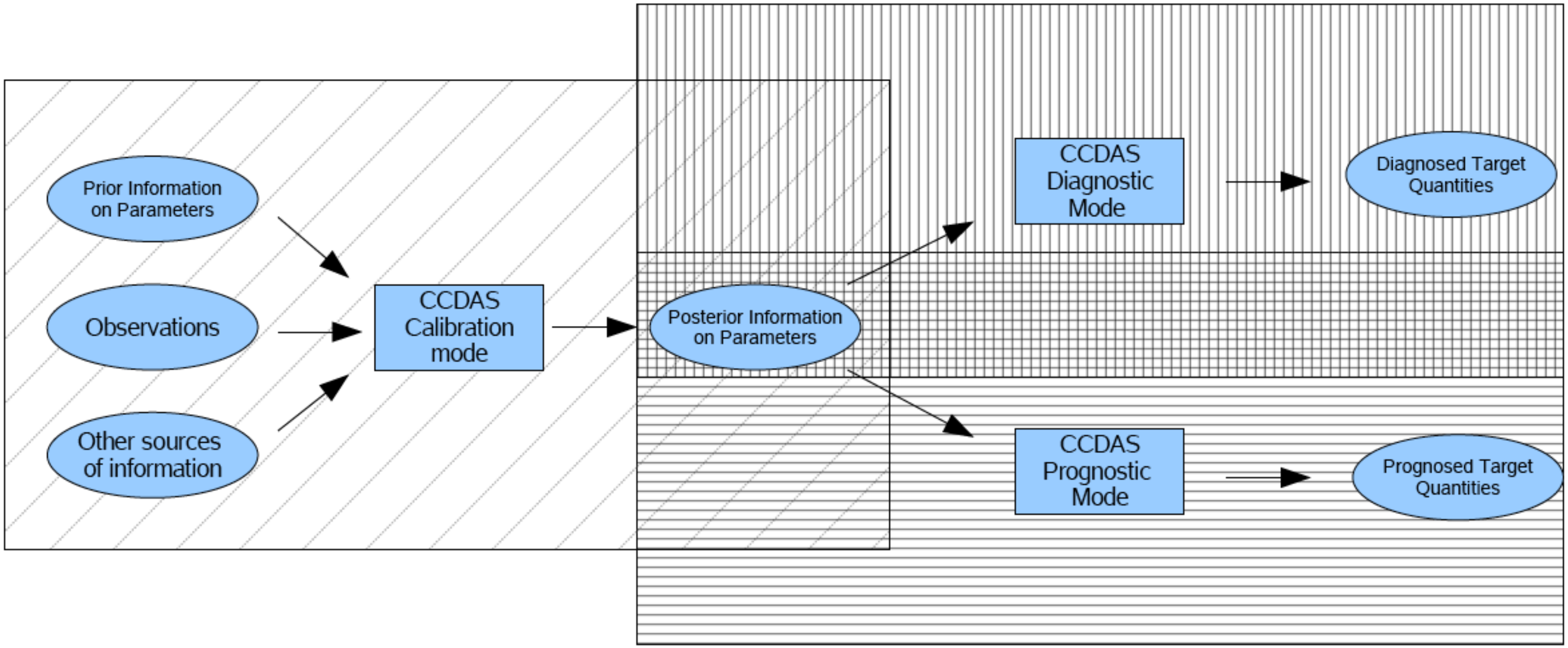
C-cycle data assimilation system



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

- Need to define the error matrices *R, P*
- Iterative minimization algortihm

CCDAS two-step procedure for inferring diagnostics and prognostics



CCDAS

- Iterative minimisation of the cost function J(**x**)
- Optimisation uses the gradient of J(**x**) with respect to the parameters
- Second order derivatives (Hessian) at minimum provide approximation of parameter uncertainties (a posteriori): $\mathbf{C}_{po}{}^{-1} = \partial^2 J(\mathbf{x}_{po}) / \partial \mathbf{x}^2$
- Uncertainties on target quantities (e.g. net flux, NEP) via linearisation of model (Jacobian matrix):

$$\mathbf{C}_{NEP} = \partial \mathbf{M} / \partial \mathbf{x}\ \mathbf{C}_{po}\ \partial \mathbf{M} / \partial \mathbf{x}^T$$

- All derivatives provided via automatic differentiation of model code (TAF)

BETHY

GPP:

C3 photosynthesis – *Farquhar et al. (1980)*

C4 photosynthesis – *Collatz et al. (1992)*

stomata – *Knorr (1997)*

$$GPP = \min\left[J_C, J_E\right]$$

$$J_C = V_{\max}\left[\frac{C_i - \Gamma *}{C_i + K_C\left(1 + {}^{O_x}\!/_{K_0}\right)}\right]$$

$$J_E = \left[\frac{\alpha_q I J_{\max}}{\sqrt{J_{\max}^2 + \alpha_q^2 I^2}}\right]\left[\frac{C_i - \Gamma *}{4\left(C_i + 2\Gamma *\right)}\right]$$

Δt=1h

Plant respiration:

maintenance resp. = f(N_{leaf}, T) – *Farquhar, Ryan (1991)*

growth resp. ~ NPP – *Ryan (1991)*

Δt=1h

Soil respiration:

fast/slow pool resp., temperature and soil moisture dependant

$$R_{S,f} = C_f(1 - f_s)(\omega^\kappa Q_{10f}^{Ta/10}) / \tau_f$$

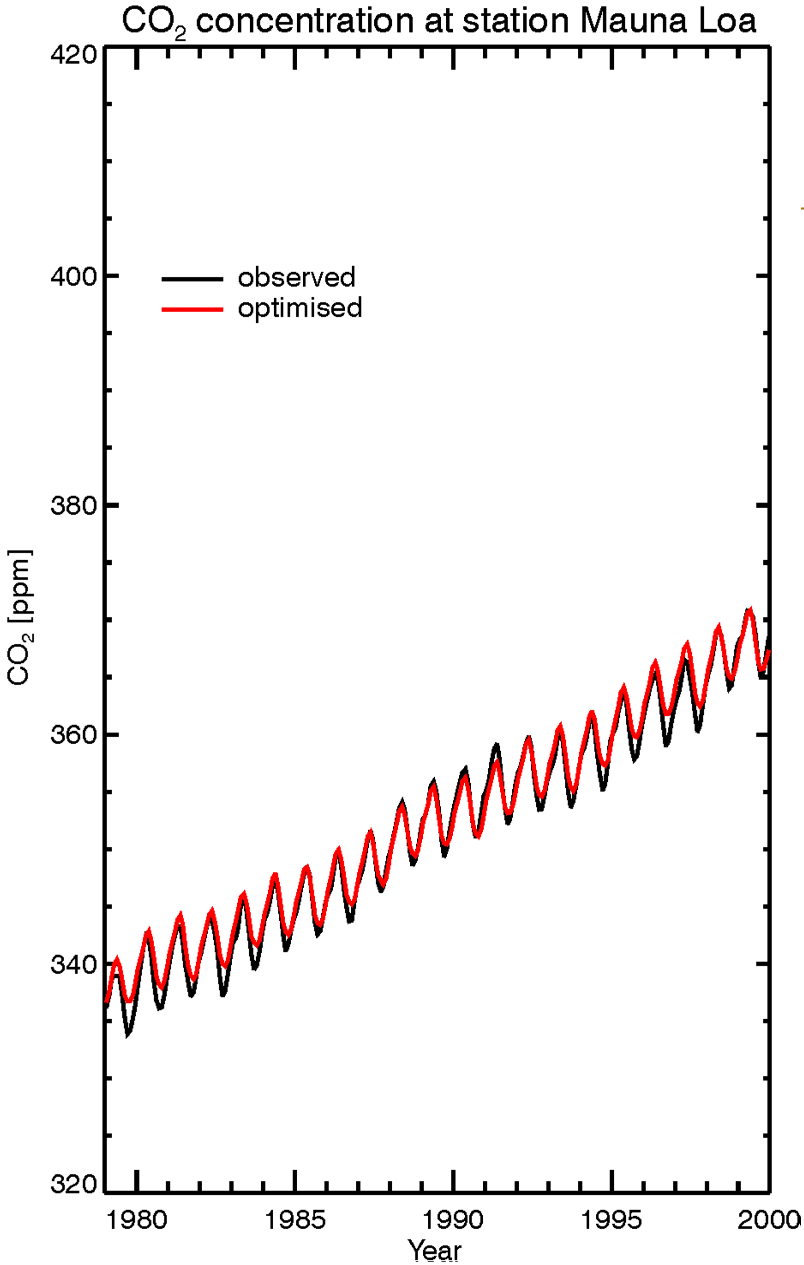
Δt=1day

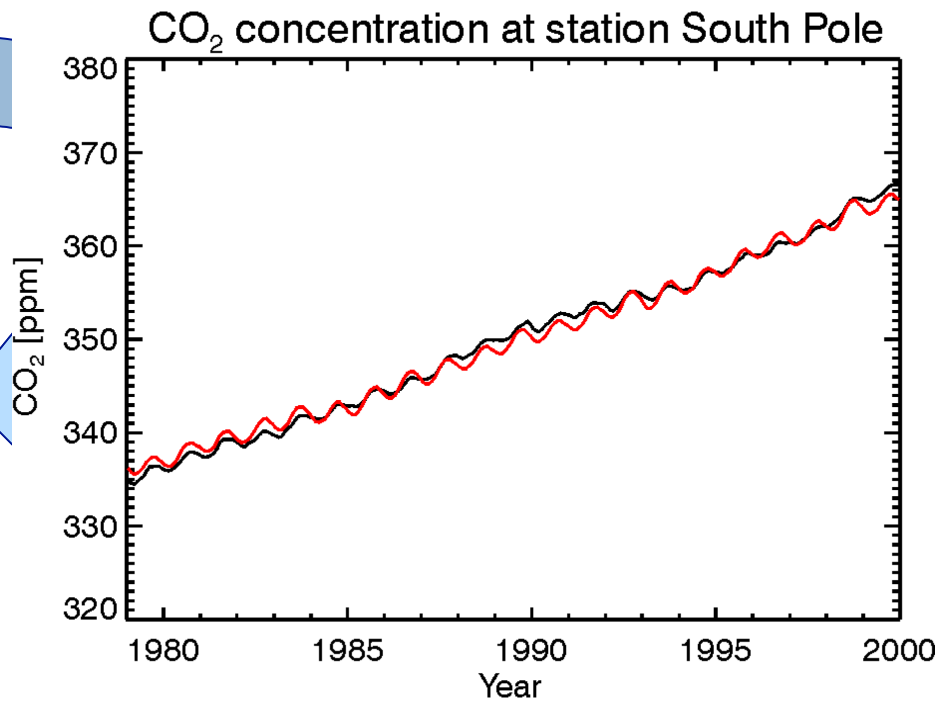
Carbon balance:

average NPP = b average soil resp. (at each grid point)

β<1: source
β>1: sink

Data fit





Rayner et al., 2005

Posterior uncertainties on parameters

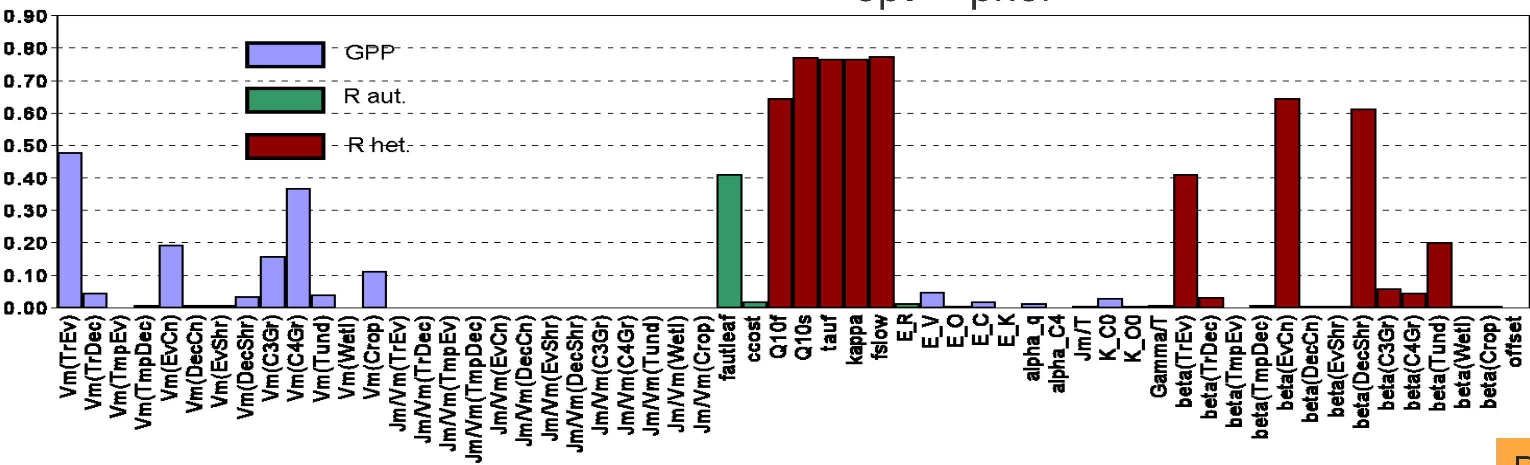
Inverse Hessian of cost function approximates posterior uncertainties

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

examples:

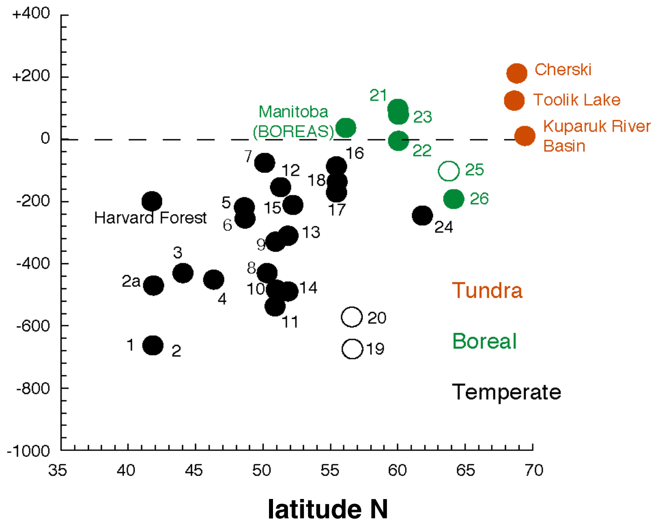
	μmol/m ^{2}s	μmol/m ^{2}s	%	%	error covariance			
Vm(TrEv)	60.0	43.2	20.0	10.5	**0.28**	0.02	-0.02	0.05
Vm(EvCn)	29.0	32.6	20.0	16.2	0.02	**0.65**	-0.10	0.08
Vm(C3Gr)	42.0	18.0	20.0	16.9	-0.02	-0.10	**0.71**	-0.31
Vm(Crop)	117.0	45.4	20.0	17.8	0.05	0.08	-0.31	**0.80**

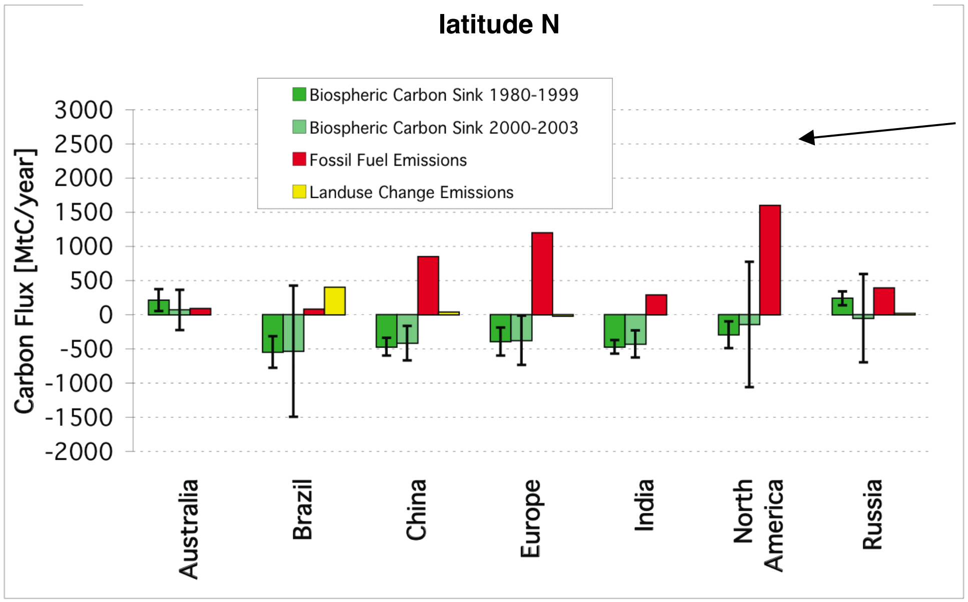
Relative Error Reduction $1-\sigma_{opt}/\sigma_{prior}$

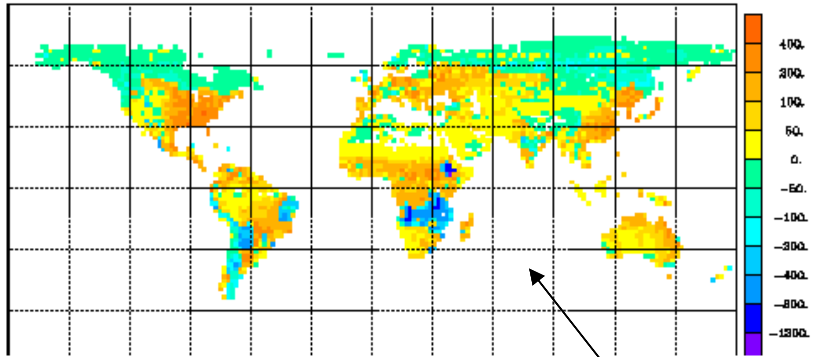


Rayner et al., 2005

Net C fluxes and their uncertainties

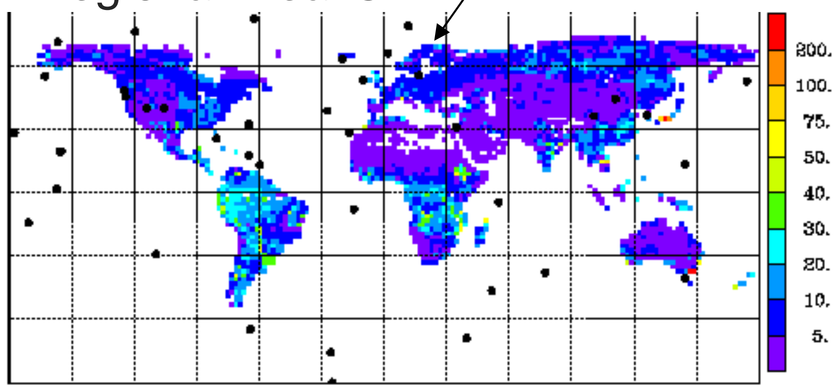






Examples for diagnostics:

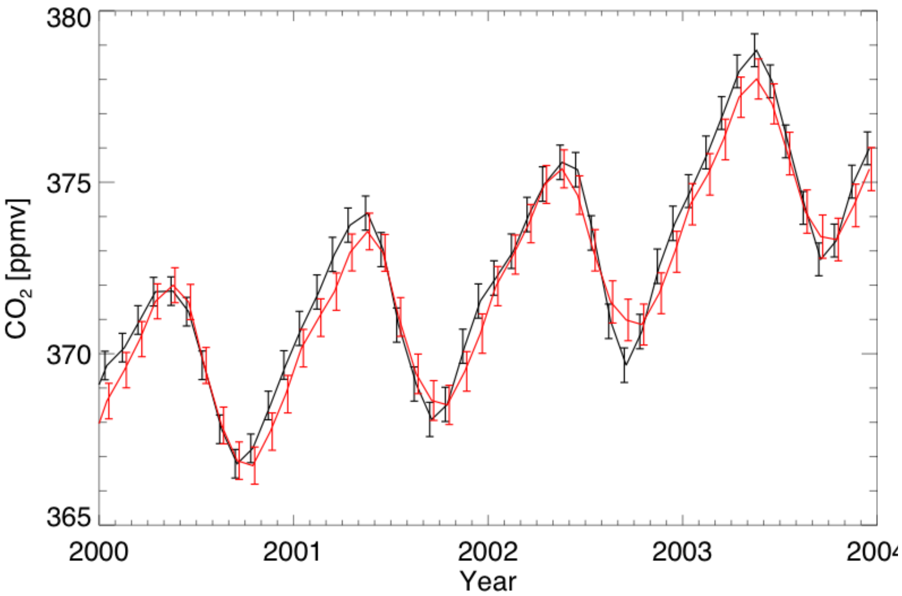
- Long term mean fluxes to atmosphere (gC/m²/year) and uncertainties
- Regional means



Rayner et al., 2005

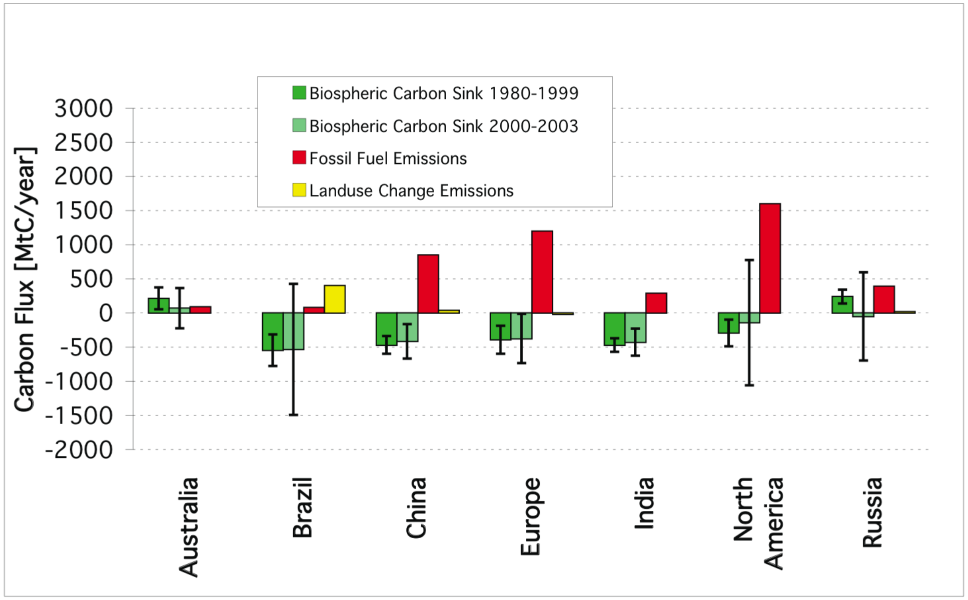
CCDAS prognostic mode
hindasting 2000-2003

CO$_2$ concentration at Mauna Loa



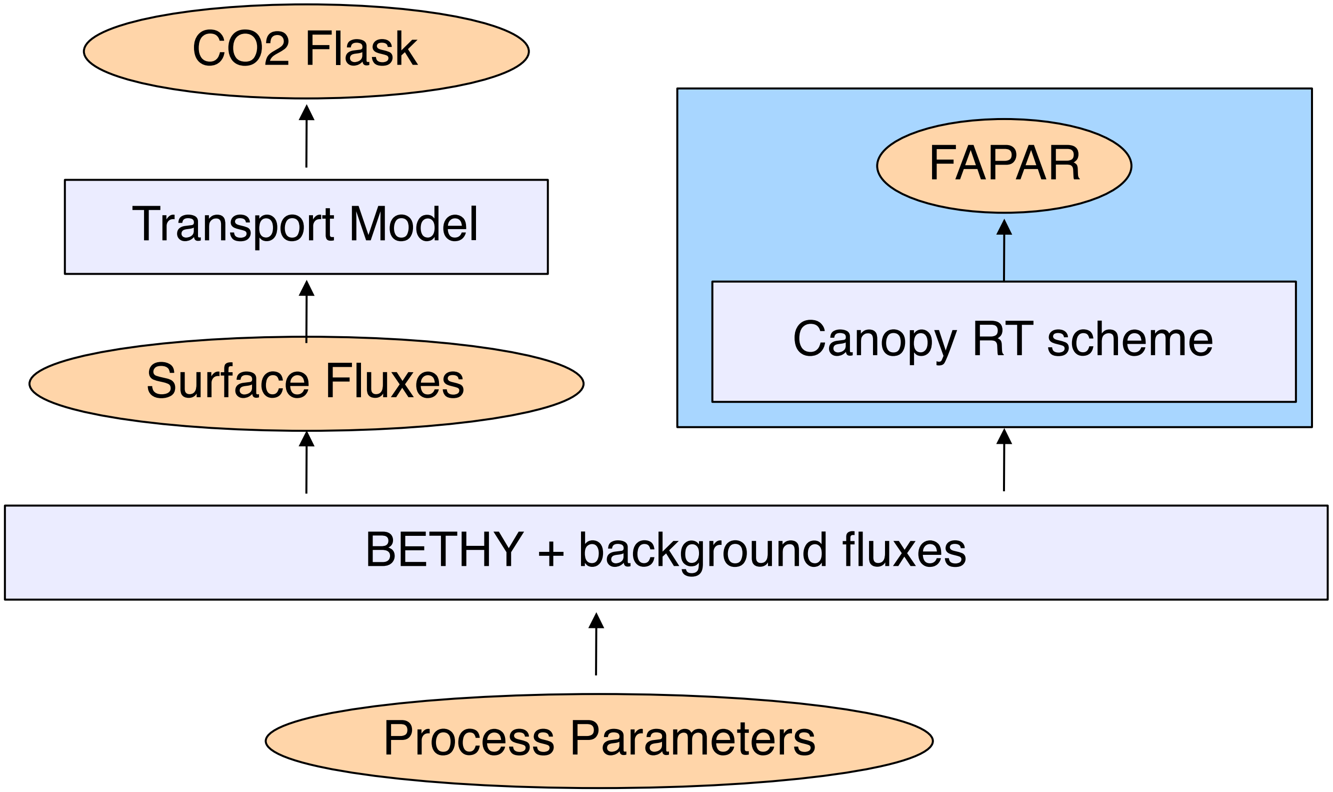
black: observations
red: modelled concentrations

Regional carbon balances

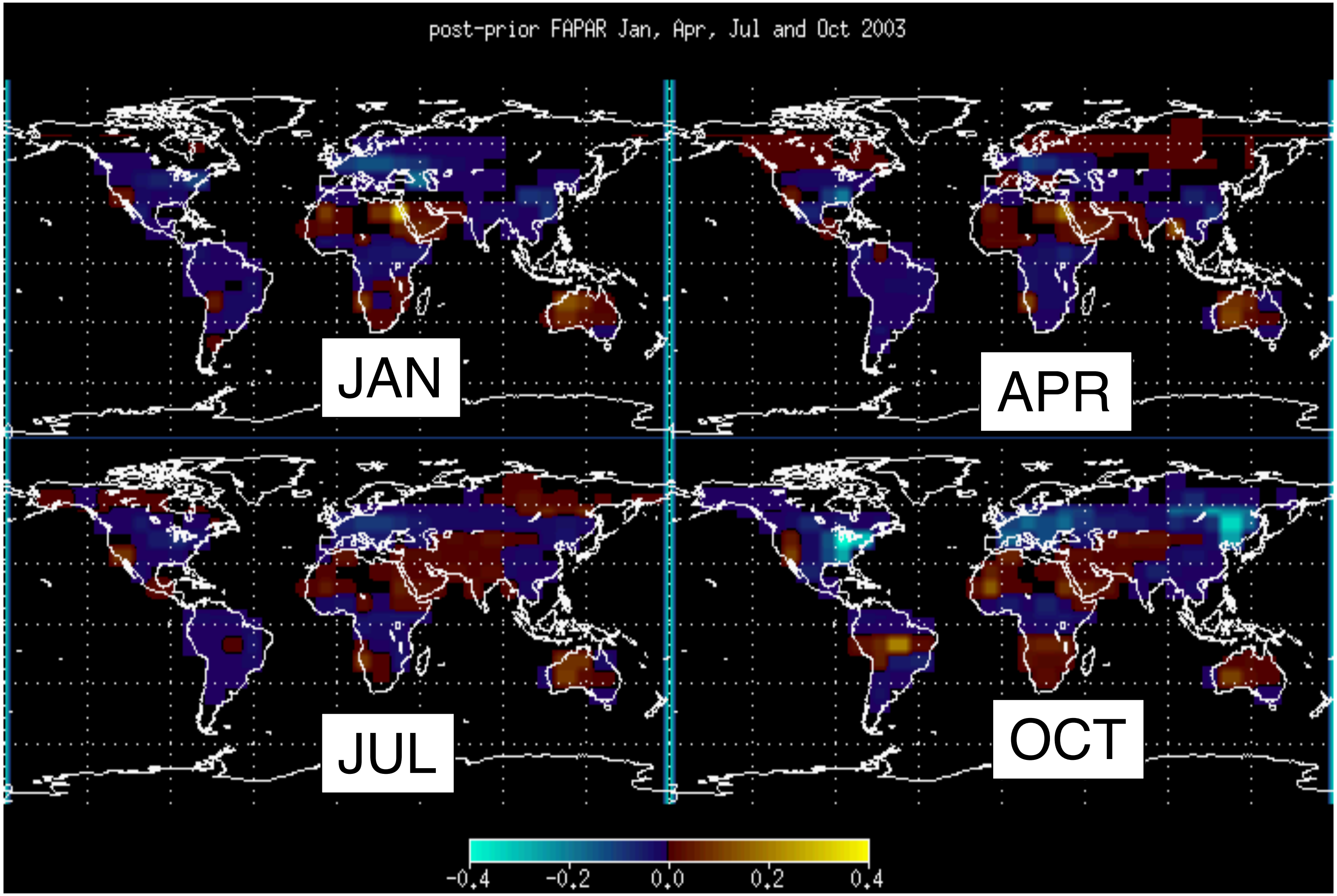


Scholze et al., 2007

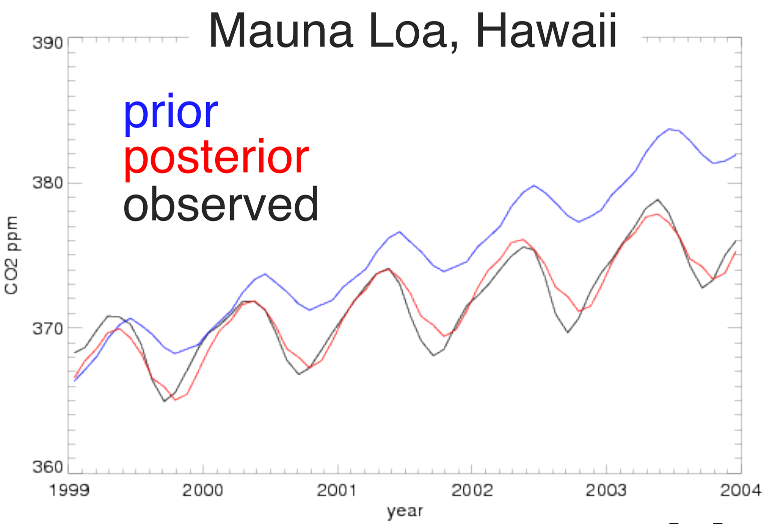
Simultaneous assimilation of MERIS FAPAR and atmospheric CO_2

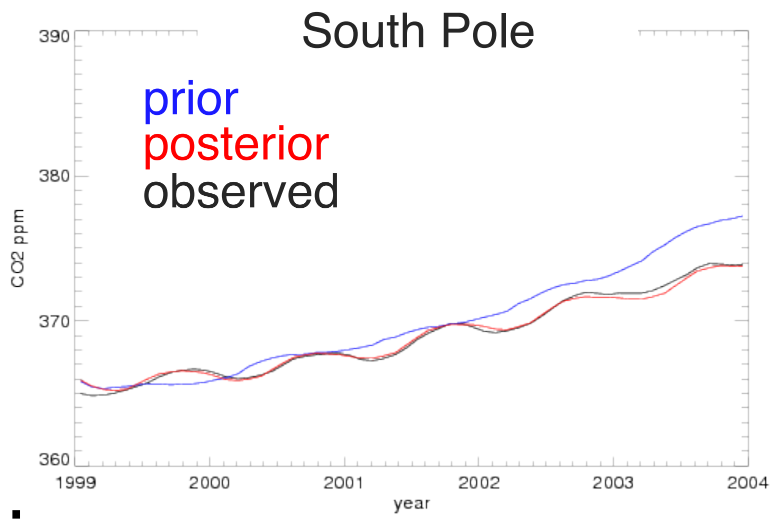


Results FAPAR
Difference posterior - prior

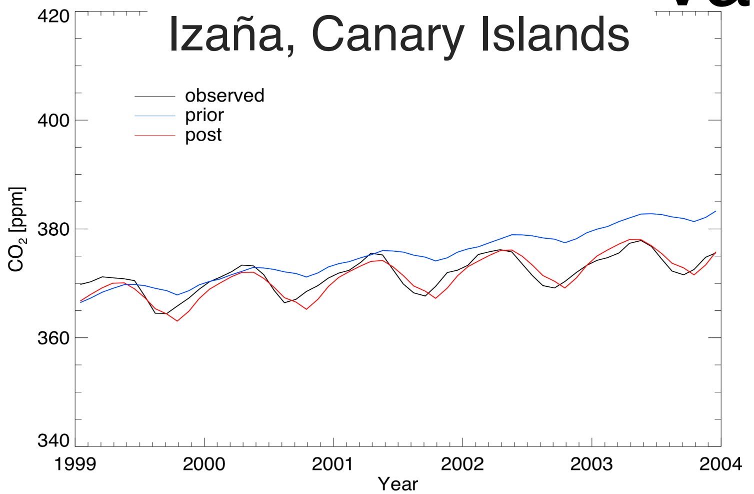


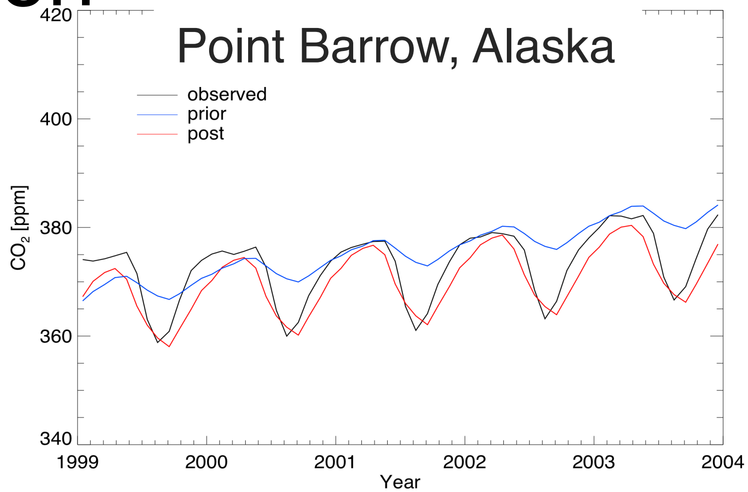
Fit to atmospheric CO_2





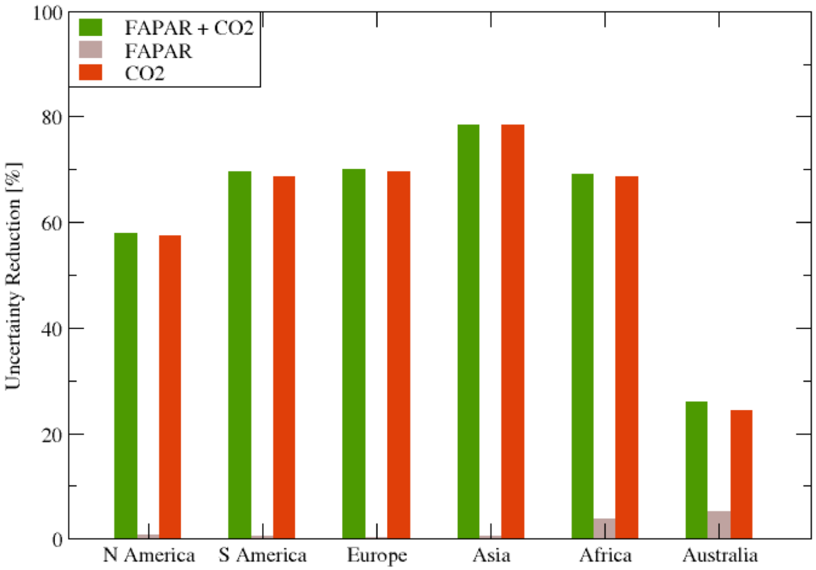
Validation



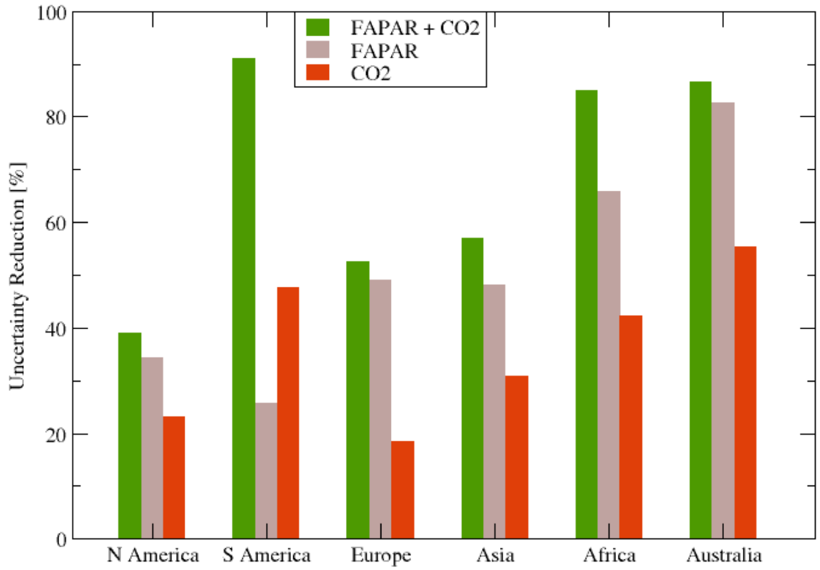


Uncertainty reduction on simulated fluxes

Regional NEP



Regional Evapotranspiration



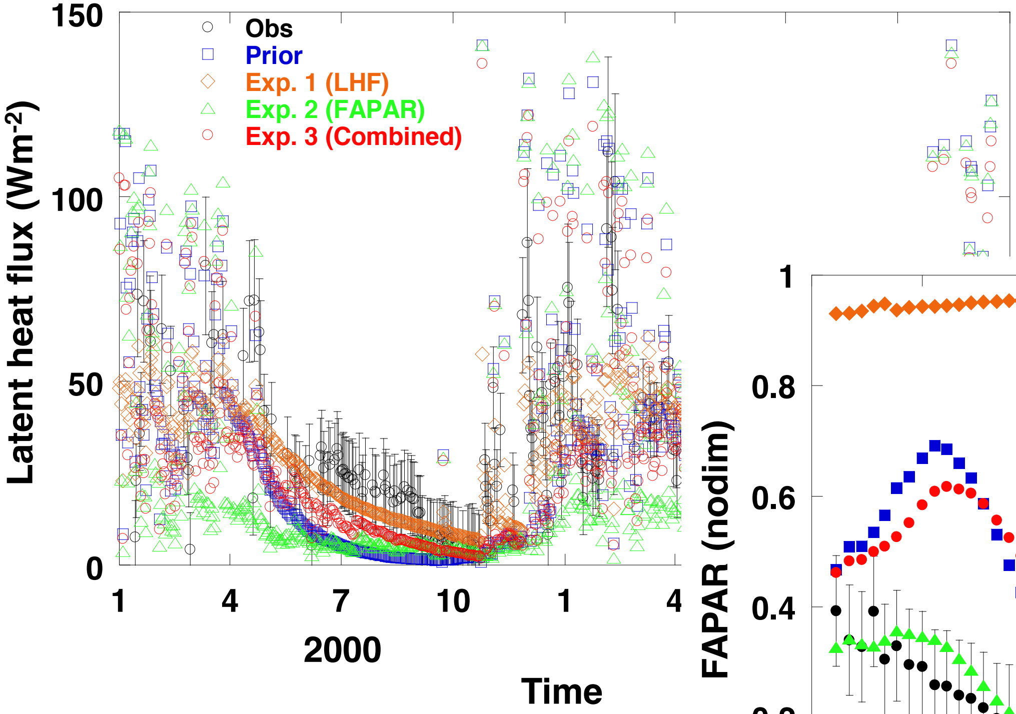
Kaminski et al., 2010

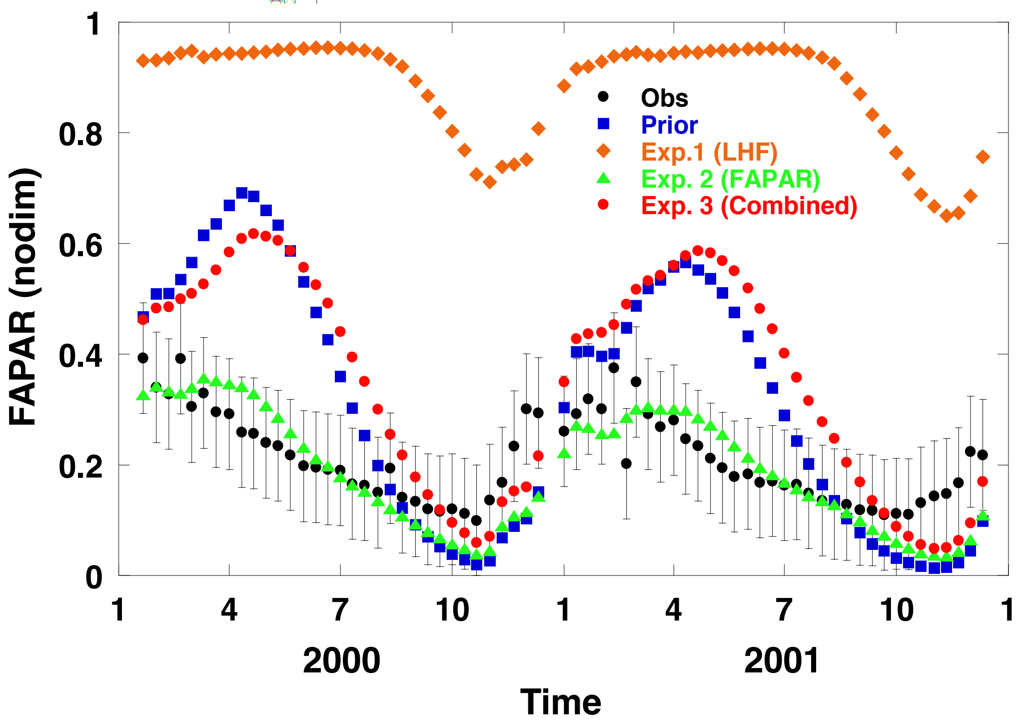
Assimilation of MERIS FAPAR and latent heat flux

- 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)
- SeaWiFS FAPAR observations, 10-daily temporal and 1.5km spatial resolution (70 observations)
- Optimization of 24 model parameters
- 2 Plant Functional Types: tropical broadleaf deciduous tree and C4 grass

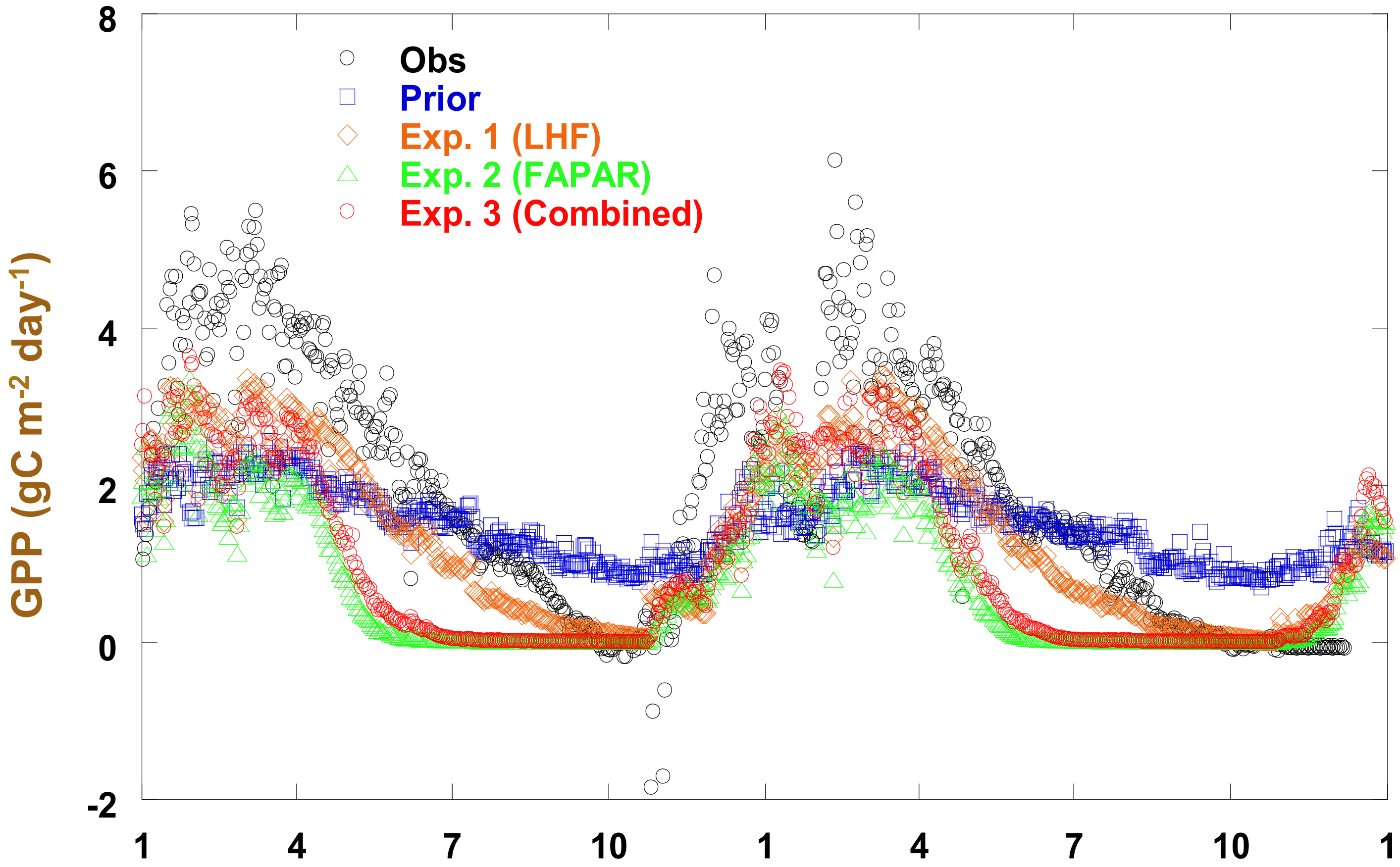
Kato et al. (2012), *Biogeosciences*

Fit to LE and FAPAR data

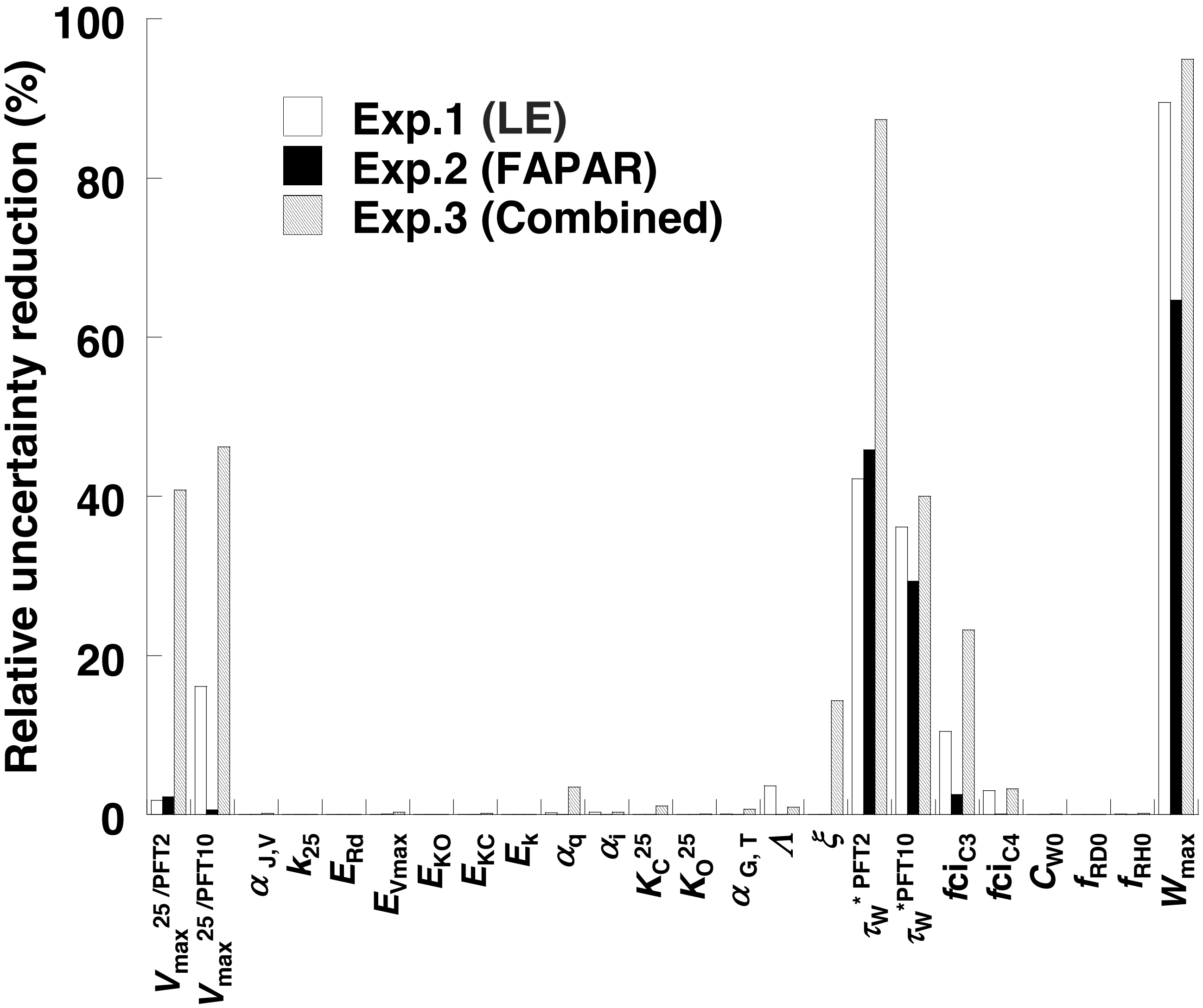




Fit against GPP



Posterior parameter uncertainty



Relative reduction:

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

Some thoughts…

- Parameters are invariant in time
- Sequential approach inconsistent over time (changes model trajectory) and not mass conserving
- Data update requires to re-calibrate the parameters over the entire time period and not only the update period (assimilation window)
- Long assimilation window to capture slow processes
- Probably not feasible in forecasting systems, but
  – Do parameter calibration less frequently (annually?)
  – Run forecasts with current calibrated parameter vector

Summary

- CCDAS tests a given combination of observational data + model formulation with uncertain parameters. It delivers optimal parameters, diagnostics/prognostics and their uncertainties.
- Methodology has been picked up by major modelling centres in Europe (MPI, LSCE, Met-Office)
- Multiple-data constraint can be significantly larger than each single data constraint together
- Method identifies mismatches between model and datasets, i.e. consistency between model and data
- Method may not be directly applicable for operational forcasting