Integration and network design with CCDAS

T. Kaminski[1], M. Scholze[2], P. Rayner[3], W. Knorr[2], R. Giering[1], and H. Widmann[4]

Copy of presentation at http://CCDAS.org

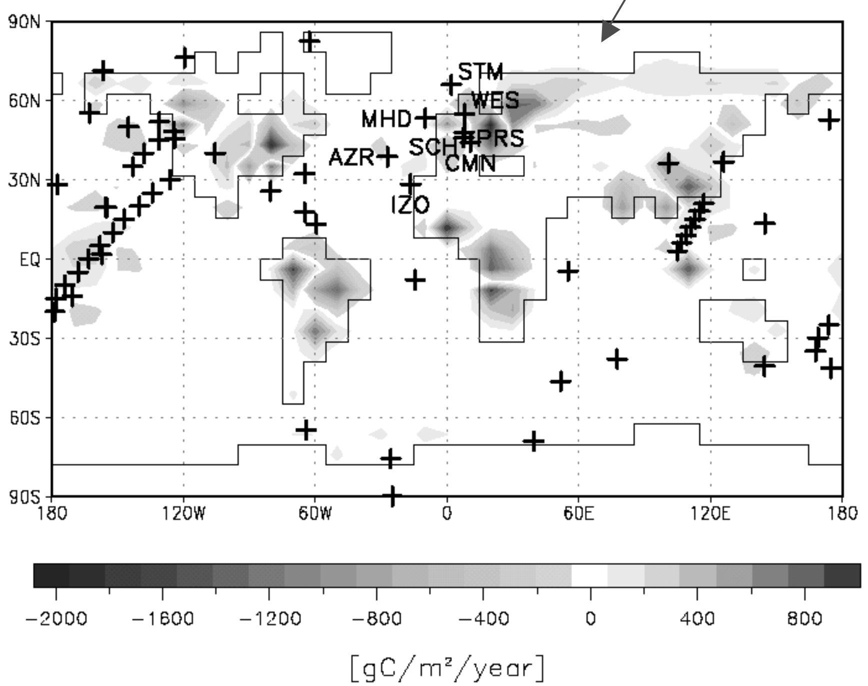
1 FastOpt 2 QUEST Quantifying and Understanding the Earth System 3 LSCE 4 Max-Planck-Institut für Biogeochemie

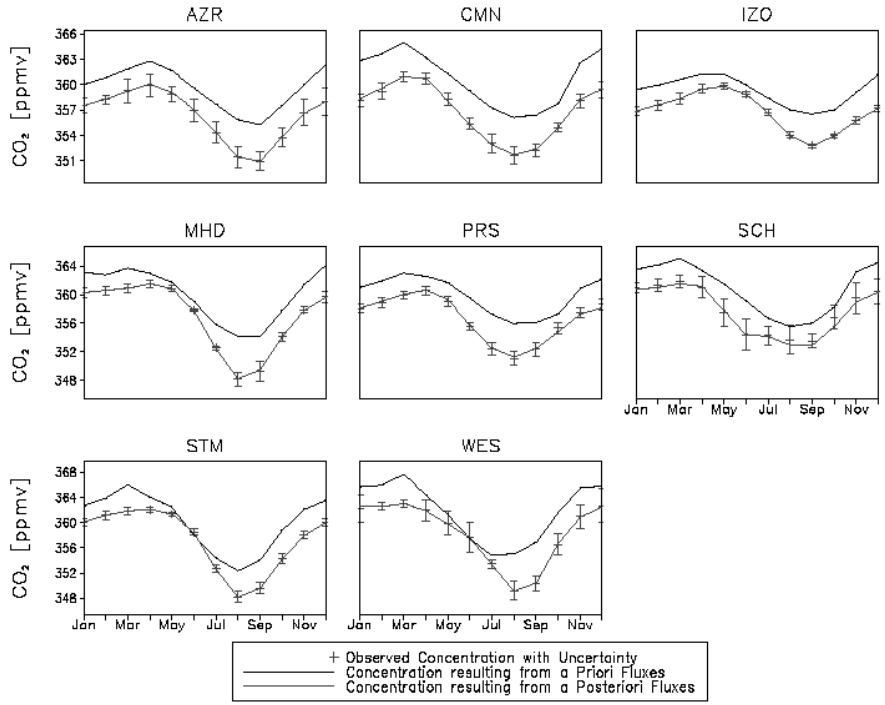
CCDAS

Motivation

some disadvantages of transport inversions

- **Not predictive**
- **Handles restricted types of observations**
- **Results depend *strongly* on additional information: Prescribed patterns, a priori fluxes and uncertainties**

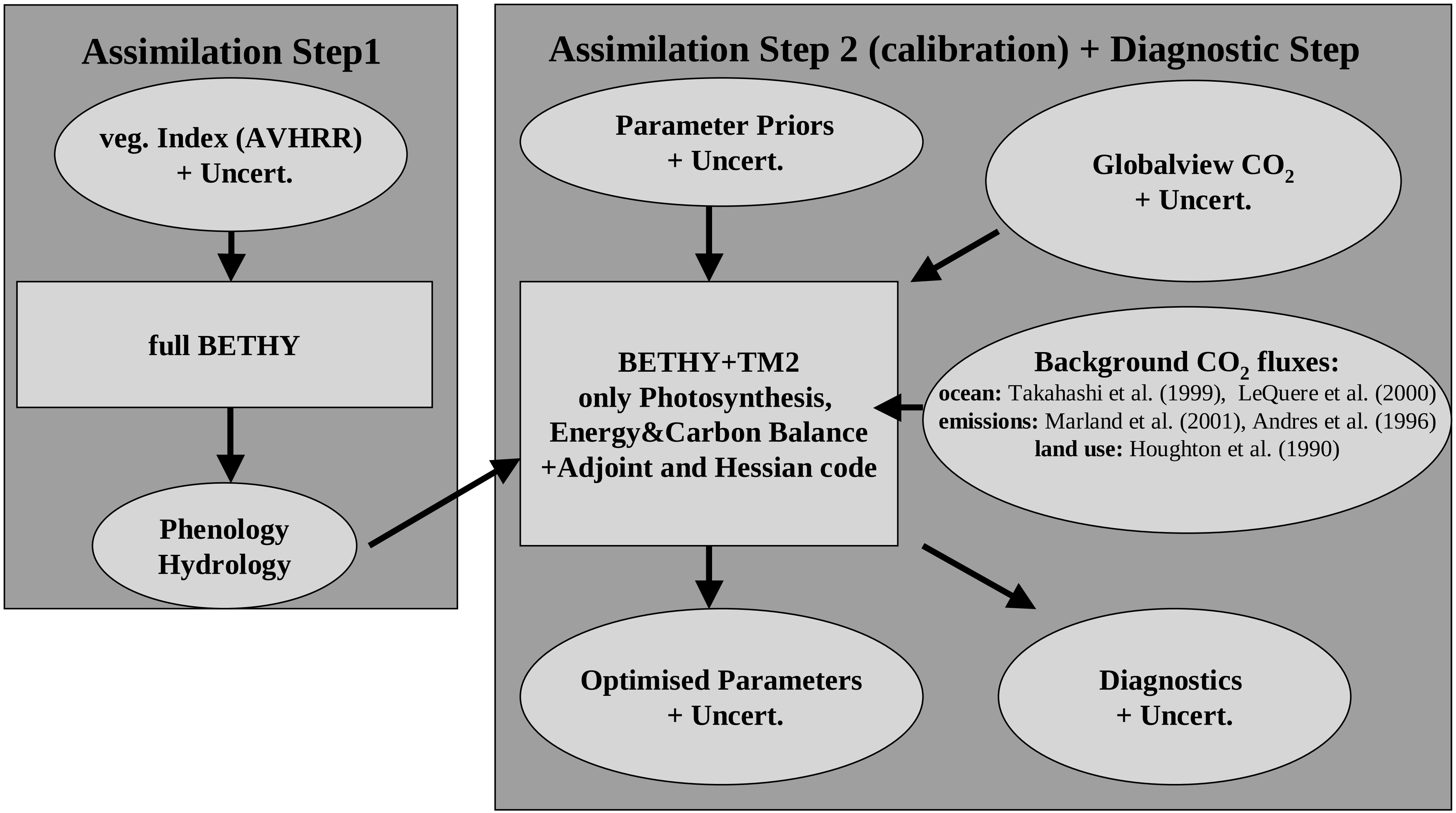




Overview

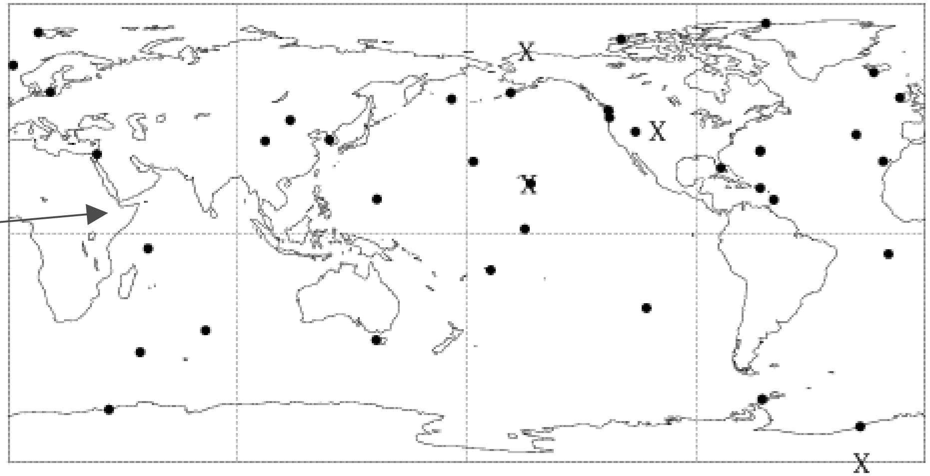
- Motivation
- CCDAS overview
- Improved processes
- More observations
- Network design
- Conclusions/Left-out issues

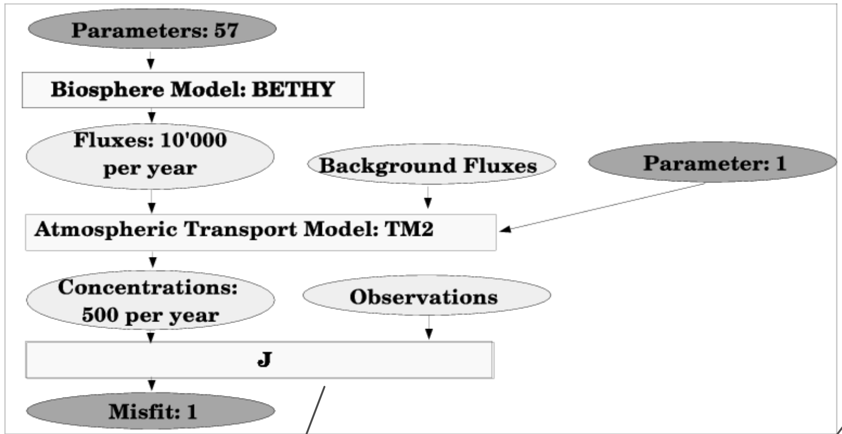
Carbon Cycle Data Assimilation System (CCDAS) current form



CCDAS calibration step

- Terrestrial biosphere model BETHY (Knorr 97) delivers CO2 fluxes to atmosphere
- Uncertainty in process parameters from laboratory measurements
- Global atmospheric network provides additional constraint





covariance of uncertainty in priors for parameters

covariance of uncertainty in measurements + model

priors for parameters

observed concentrations

$$J(m) = \frac{1}{2}(m - m_0)^T C_m^{-1}(m - m_0) + \frac{1}{2}(c(m) - d)^T C_d^{-1}(c(m) - d)$$

Minimisation and Parameter-Uncertainties

Gradient of *J*(*m*) provides

search directions for minimisation.

Second Derivative (Hessian) of *J*(*m*)

yields curvature of *J*,

provides estimated

uncertainty in m_{opt}

$$\mathbf{C}_m \approx \left\{ \frac{\partial^2 J(m_{opt})}{\partial m_{i,j}^2} \right\}^{-1}$$

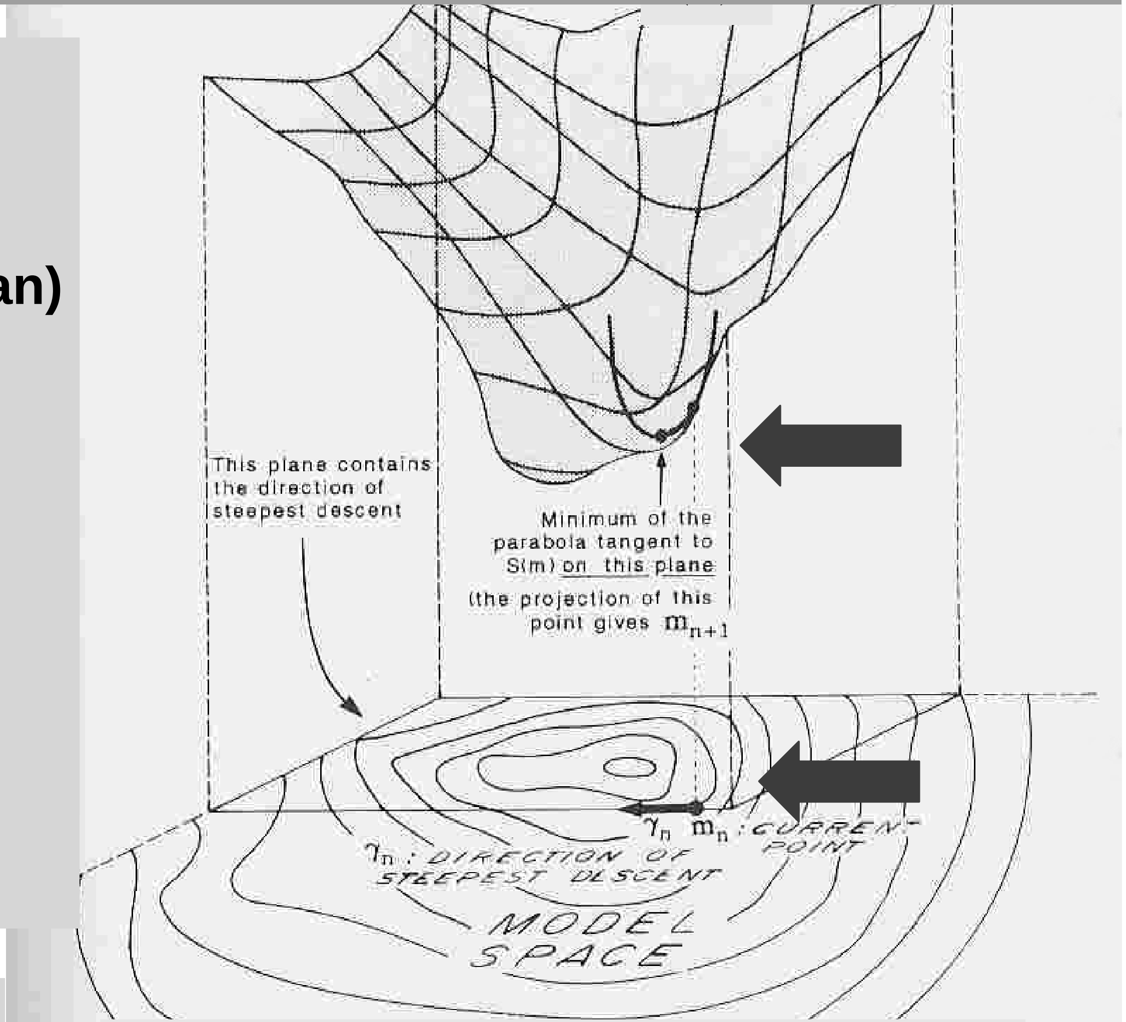
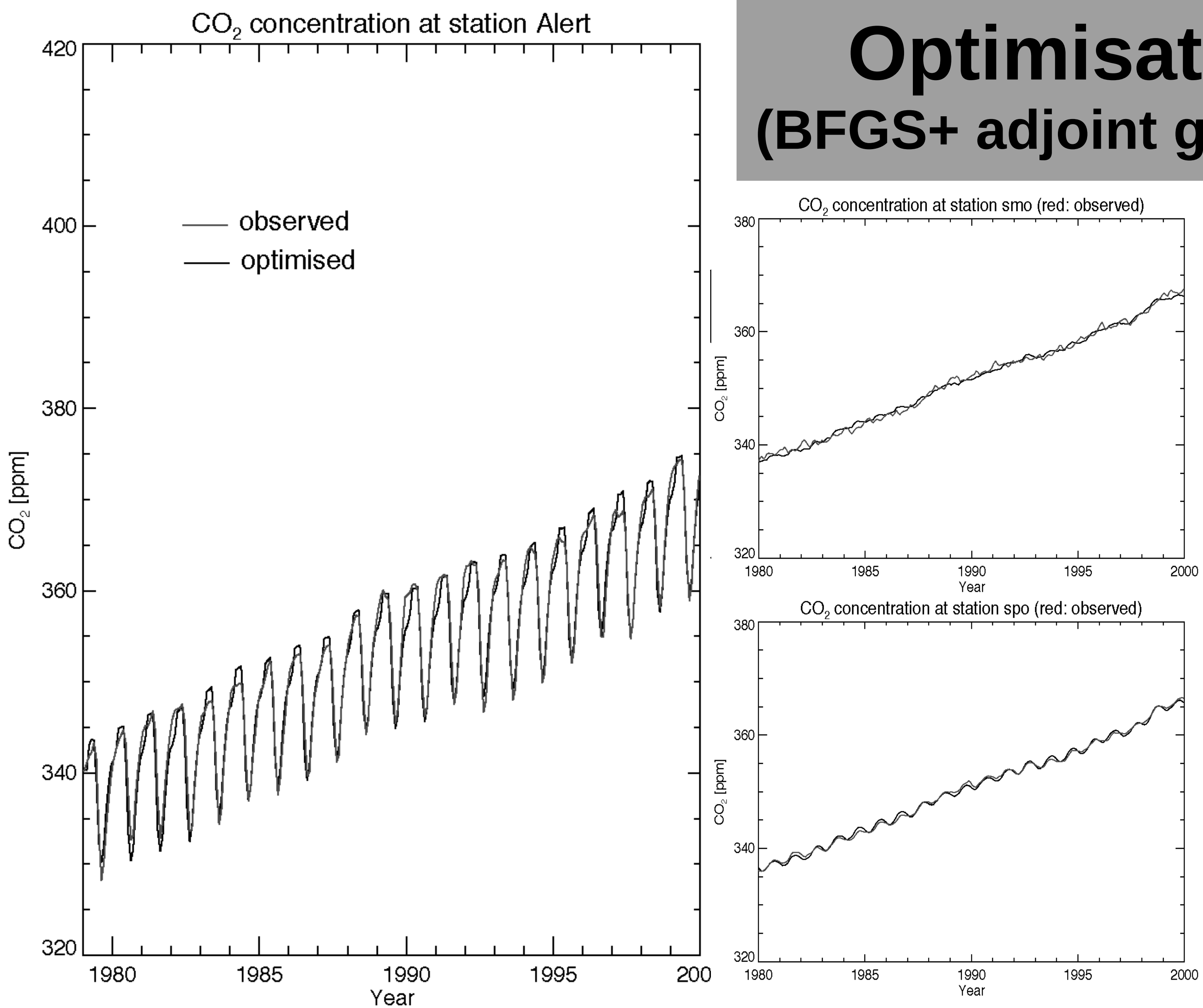


Figure taken from **Tarantola** '87

Space of *m* (model parameters)

Optimisation
(BFGS+ adjoint gradient)



Posterior uncertainties on parameters

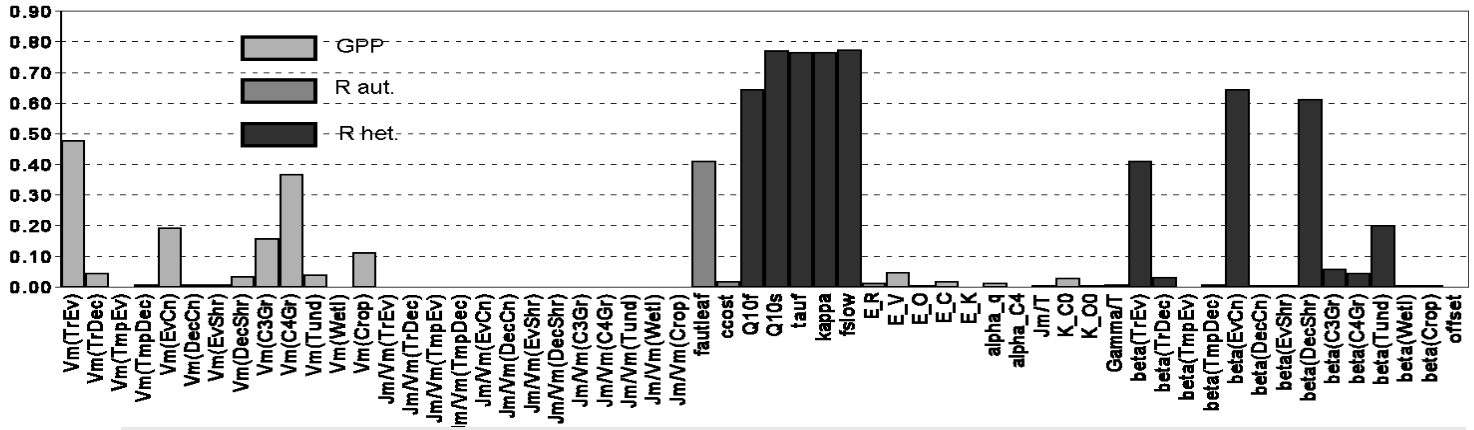
Use inverse Hessian of objective function to approximate posterior uncertainties

$$C_m \approx \left(\frac{\partial^2 J(m_{opt})}{\partial m_{i,j}^2} \right)^{-1}$$

examples:

	first guess	optimized	prior unc.	opt.unc.	Vm(TrEv)	Vm(EvCn)	Vm(C3Gr)	Vm(Crop)
	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**

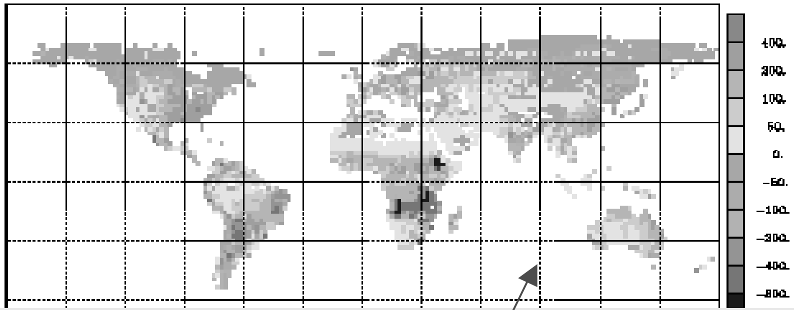
Relative reduction of uncertainties



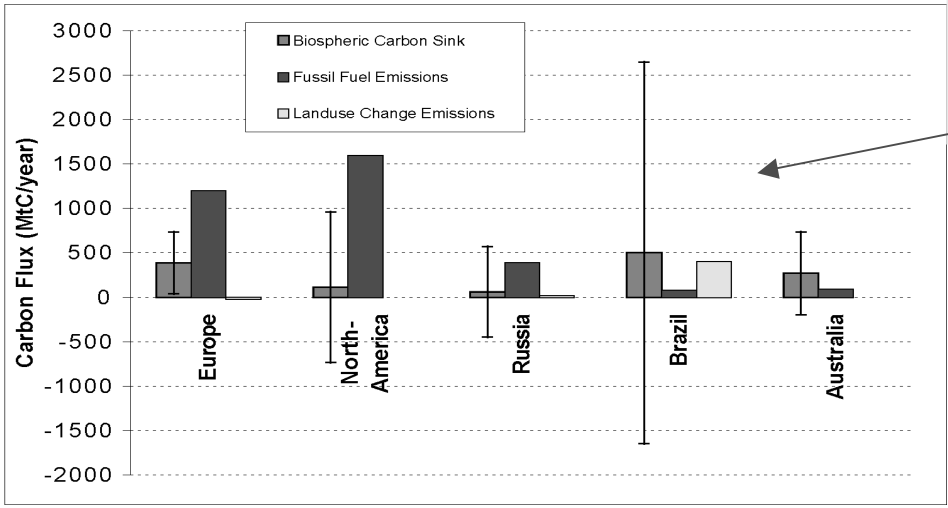
Observations resolve about 10-15 directions in parameter space

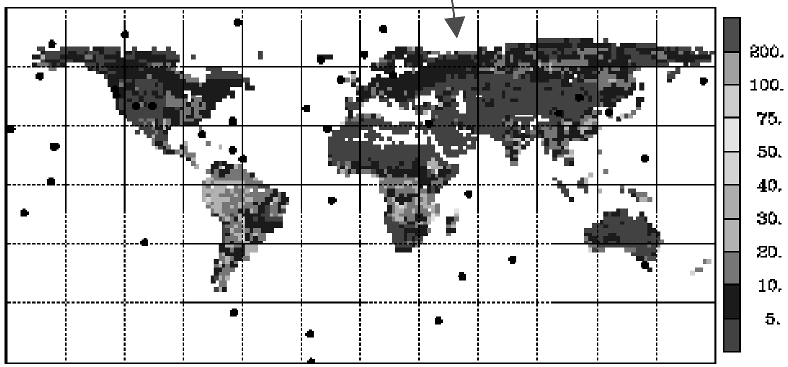
CCDAS diagnostic step
Global fluxes and their uncertainties

$$C_y = \left(\frac{\partial y_i(m_{opt})}{\partial m_j} \right) C_m \left(\frac{\partial y_i(m_{opt})}{\partial m_j} \right)^T$$



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





FastOpt
QUEST Quantifying and Understanding the Earth System
Max-Planck-Institut für Biogeochemie
MPI-BGC Jena
CCDAS
CAMELS

CCDAS
remarks, difficulties/problems

- remarks:
  - CCDAS tests a given combination of observational data plus model formulation with uncertain parameters
  - CCDAS delivers optimal parameters, diagnostics/prognostics, and their a posteriori uncertainties
  - all derivative code (adjoint, Hessian, Jacobian) generated automatically from model code by compiler tool TAF-> quick updates of CCDAS after change of model formulation
  - does not directly estimate quantities of interest, but indirectly via parameters (similar to transport inversion on the full TM grid with later aggregation)
  - CCDAS posterior flux field consistent with trajectory of process model rather than linear combination of prescribed flux patterns (classical transport inversion)

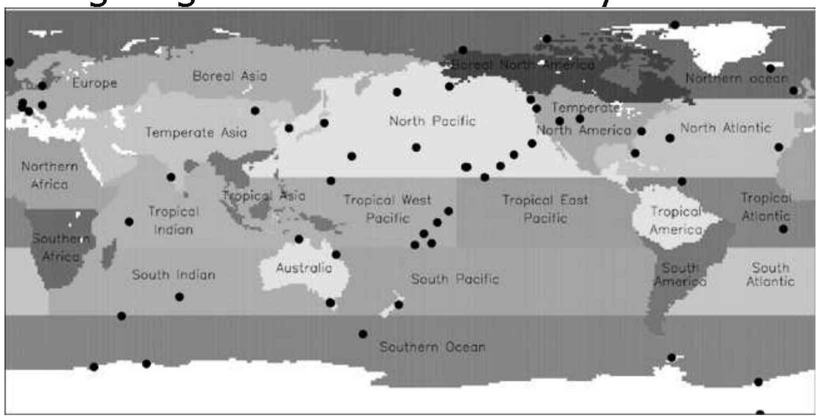
- some of the difficulties/problems:
  - Prognostic uncertainty (error bars) only reflect parameter uncertainty What about uncertainty in model formulation, driving fields...
  - Uncertainty propagation only for means and covariances (specific PDFs), and only with a linearised model
  - Result depends on a priori information on parameters
  - Two step assimilation procedure sub optimal
  - Only fAPAR and GLOBALVIEW observations included
  - Background fluxes are prescribed
  - Process models could be improved (soil, transport, balance constraint ...)
  - Processes missing: fire ...
  - lots of technical issues (bounds on parameters, driving data, Eigenvalues of Hessian ...)

Overview

- Motivation
- CCDAS overview
- Improved processes
- More observations
- Network design
- Conclusions/Left-out issues

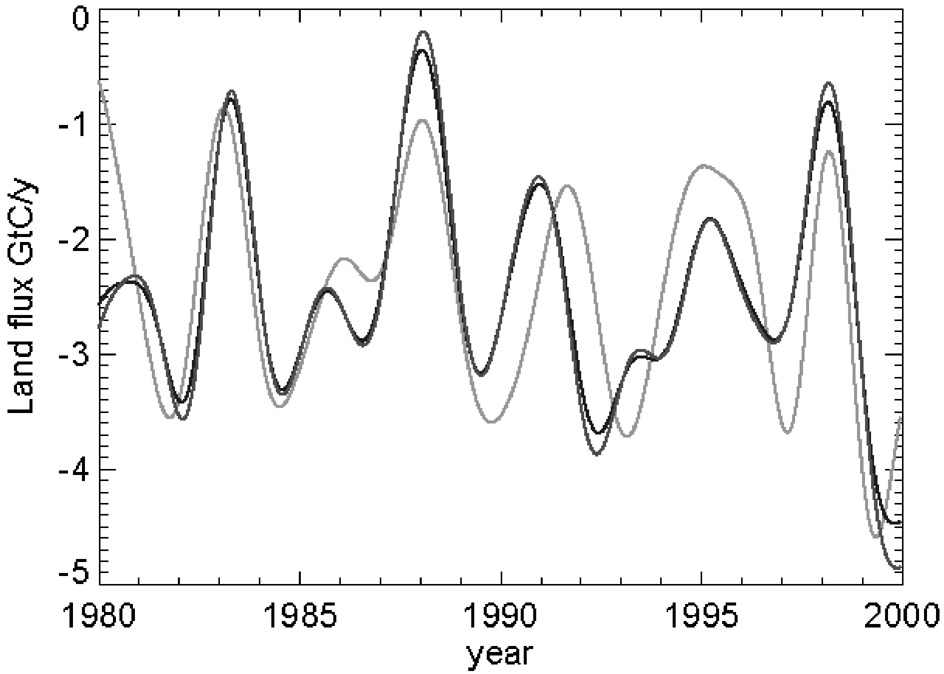
Including the ocean

- A 1 GtC/month pulse lasting for three months is used as a basis function for the optimisation
- Oceans are divided into the 11 TransCom-3 regions
- That means: 11 regions * 12 months * 21 yr / 3 months = 924 additional parameters
- Test case:
  - all 924 parameters have a prior of 0. (assuming that our background ocean flux is correct)
  - each pulse has an uncertainty of 0.1 GtC/month giving an annual uncertainty of ~2 GtC for the total ocean flux

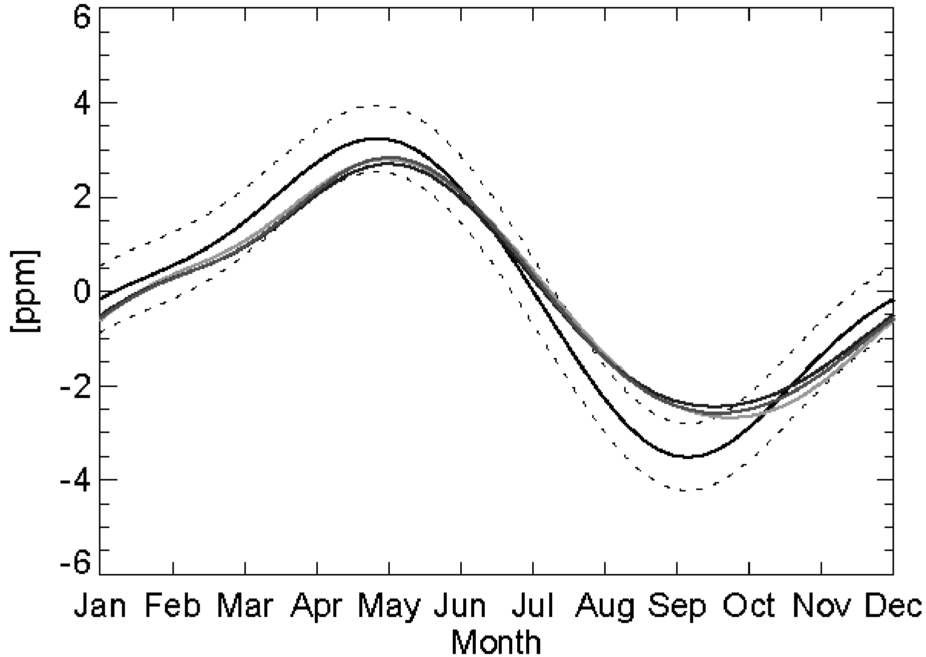


Including the ocean

Global land flux



Seasonality at MLO



Observations
High resolution standard model
Low resolution model
Low-res incl. ocean basis functions

Extending the model

- **Study uses extremely simplified form of an ocean model:**

  flux(x,t) = Σ coefficient(i) * pattern(i,x,t)

- **Optimising coefficients for biosphere patterns**
  allows the optimisation to compensate for errors (missing processes) in BETHY
  (weak constraint 4DVar, see ,e.g., Zupanski (1993))

- **It is preferable to include a process model.**
- **Candidates: fire, marine biogeochemistry, ...**

- **Also: Improvement of BETHY: More sophisticated soil model**
  or Transport Model: TM2 -> TM3, interannual winds, higher resolution...

Extending the model

Efficient handling of spin-up

- CCDAS uses a ß-factor as PFT-specific parameter; determines net flux:

  average NPP = ß (average soil respiration)

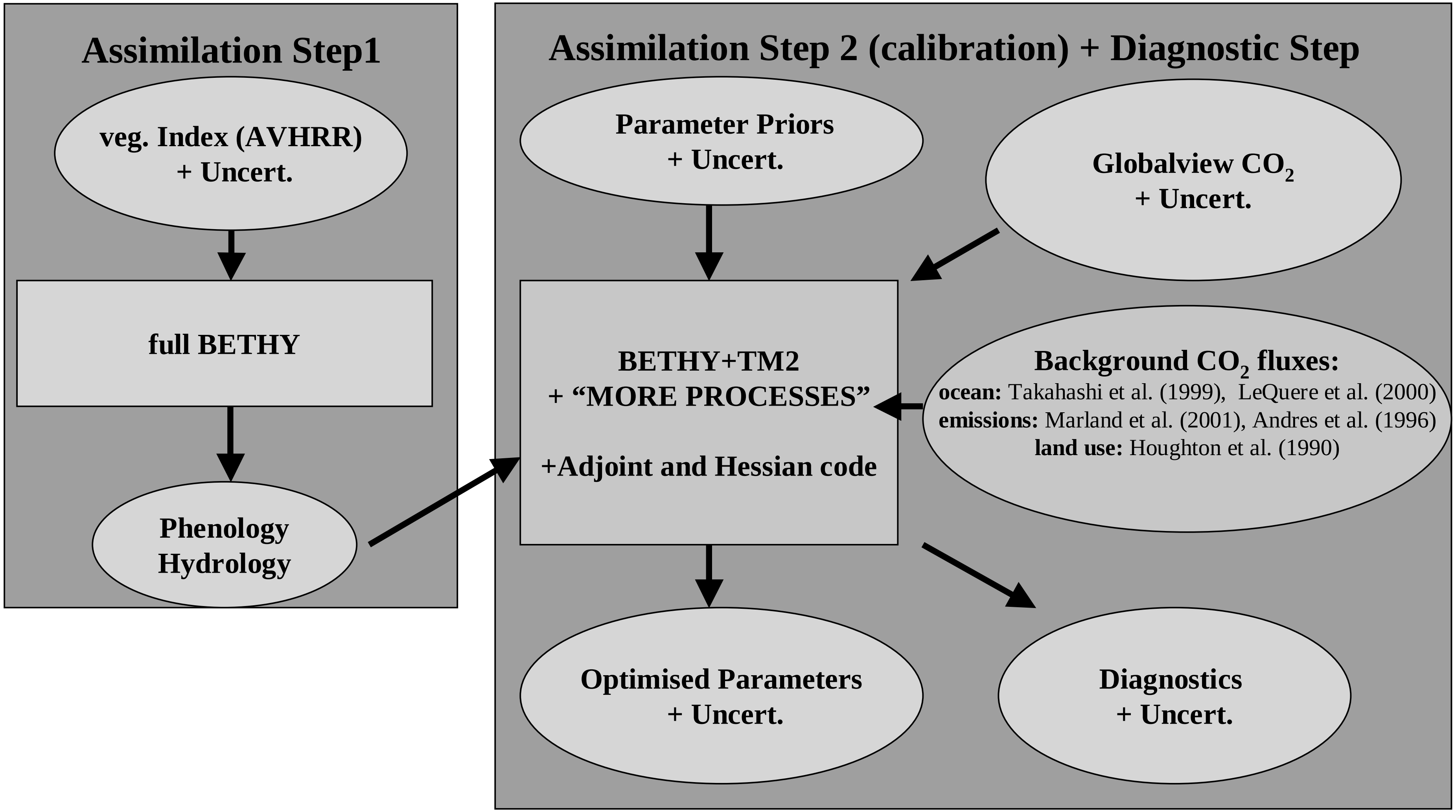
- ß avoids spin-up of slow carbon pool, with all the complications involved
- Alternative model-formulation may require a spin-up of several 1000 years
- Parameter sensitivities need to take account of spin-up period
- Simplest way is to run the adjoint through the spin-up
- Alternative way via implicit function theorem for final year of spin up:

  s = model (s, x)   s: equilibrium state; x: parameters

  ds/dx = d (model)/ds ds/dx + d(model)/dx

- Need to compute d(model)/ds and d(model)/dx only for final iteration
- Concept demonstrated for spin-up of box model of atmospheric transport (accepted by LNCS, see http://FastOpt.com)

Carbon Cycle Data Assimilation System (CCDAS) with more processes



Adding more observations

within Carbon Cycle Data Assimilation System (CCDAS)

$$J(m) = \tfrac{1}{2}\,(m-m_0)^T\, C_m^{-1}(m-m_0)$$
$$+\ \tfrac{1}{2}\,(c_{mod}(m)-c_{obs})^T\, C_c^{-1}(c_{mod}(m)-c_{obs})$$
$$+\ \tfrac{1}{2}\,(f_{mod}(m)-f_{obs})^T\, C_f^{-1}(f_{mod}(m)-f_{obs})$$
$$+\ \tfrac{1}{2}\,(I_{mod}(m)-I_{obs})^T\, C_I^{-1}(I_{mod}(m)-I_{obs})$$
$$+\ \tfrac{1}{2}\,(R_{mod}(m)-R_{obs})^T\, C_R^{-1}(R_{mod}(m)-R_{obs})$$
$$+\ \text{etc} \ldots$$

Flux Data

Inventories

Atmospheric Isotope Ratios

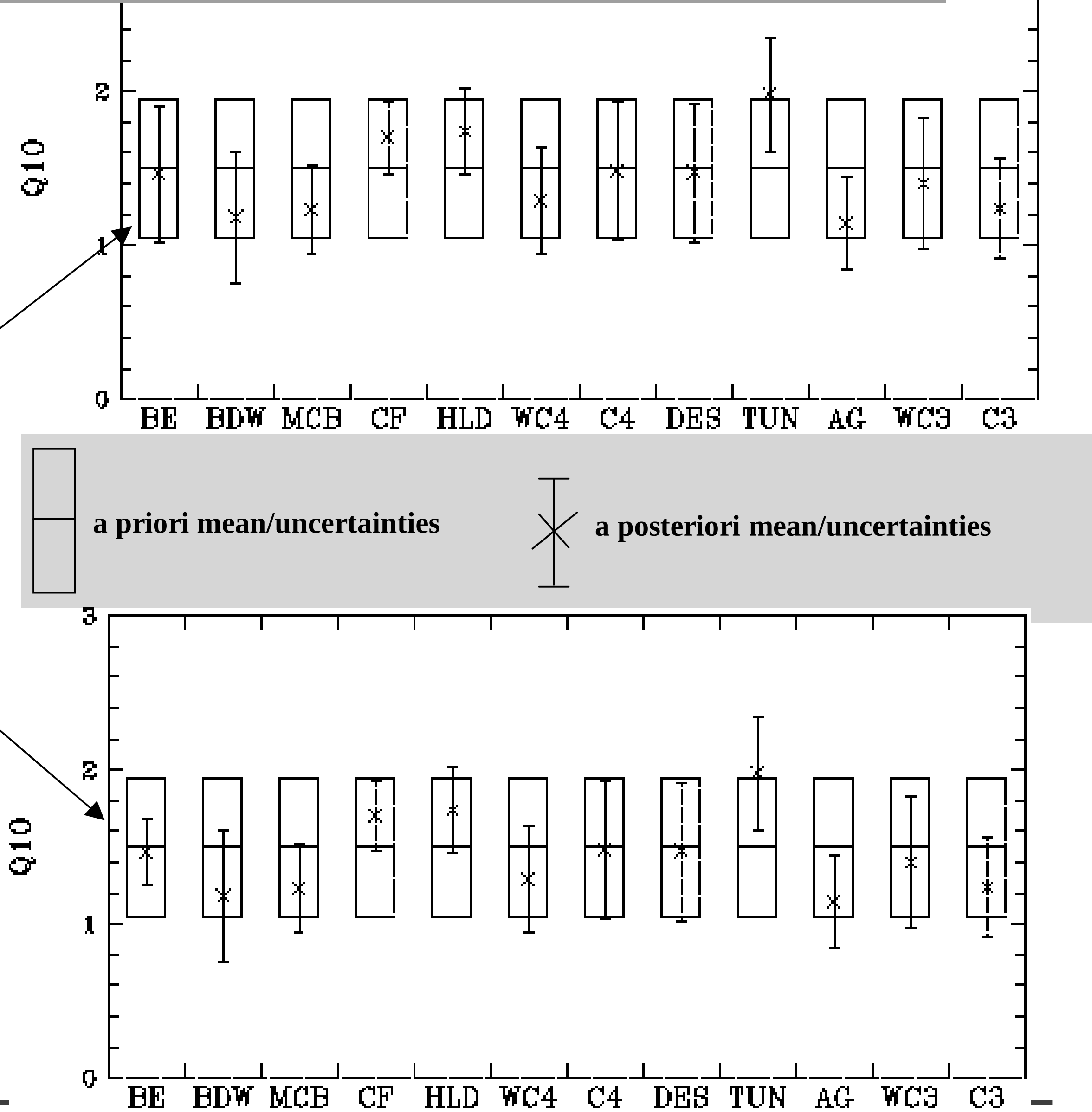
- Can add further constraints on any quantity that can be extracted from the model (possibly after extensions)
- Covariance matrices are crucial: Determine relative weights!
- Uses Gaussian assumption; can also use logarithm of quantity (lognormal distribution), ...

Example: A priori info + atmospheric CO2 + (pseudo) flux measurement

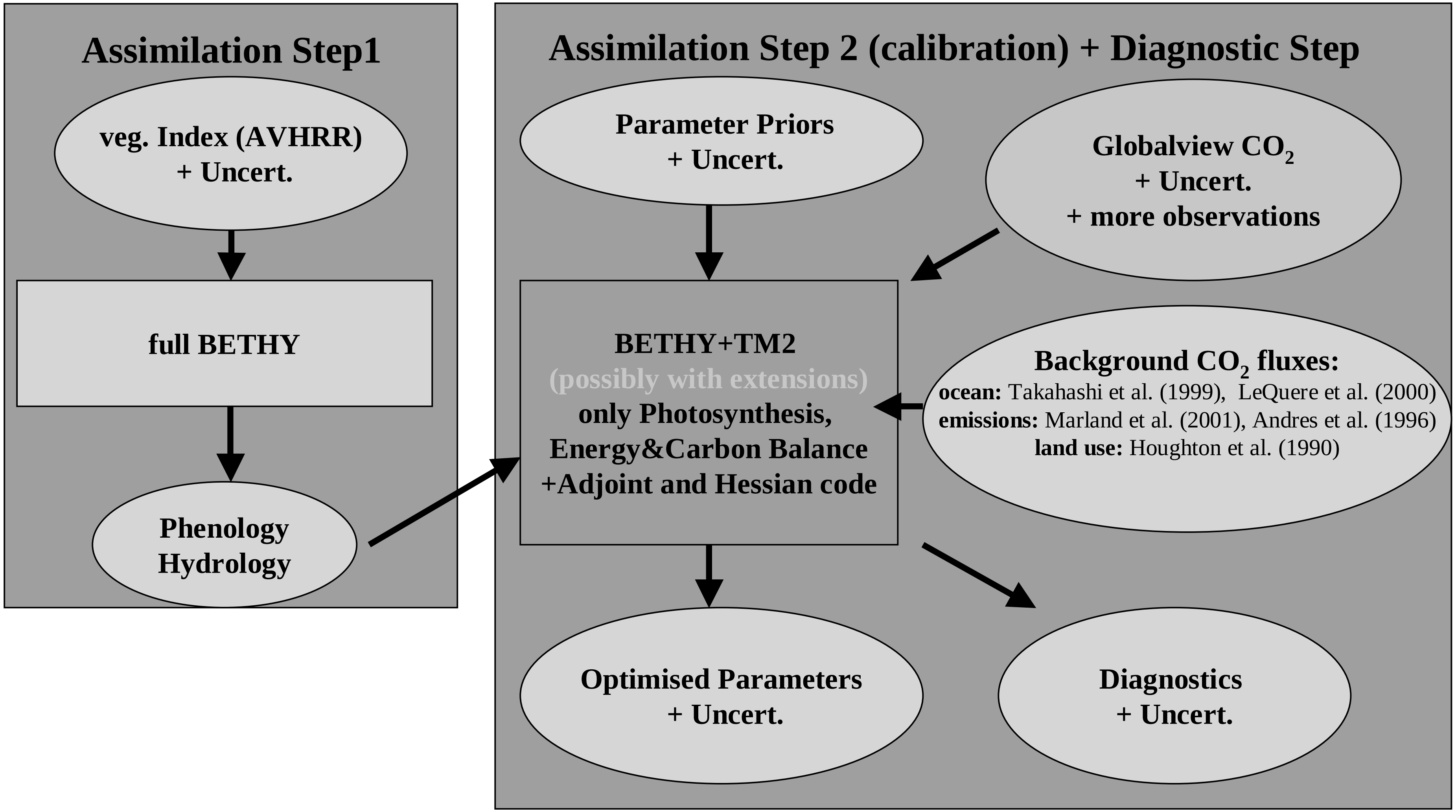
Comparison shows impact of a (pseudo) flux measurement in the broadleaf evergreen biome on Q10 estimated by an inversion of SDBM:

***Upper panel*:**
only concentration data

***Lower panel*:**
concentration data + pseudo flux measurement (mean: as predicted sigma: 10gC/m^2/year)



Carbon Cycle Data Assimilation System (CCDAS) with more observations



Network design with CCDAS

- *Rayner et al.* (1996) approach for observing system design can be adapted to CCDAS (propagate uncertainty independently from value of measurement)
  - Define target quantity (e.g. European balance for 1990 or 2010)
  - Repeat following procedure for each candidate observing system:
    - Prescribe uncertainty of potential observations
    - Prescribe prior parameter uncertainties
    - Backpropagate uncertainties to parameter uncertainties (like pseudo flux measurement for SDBM)
    - Propagate parameter uncertainties forward to target quantity
- Best candidate is the one with smallest uncertainty on target, it depends on:
  - Target quantity
  - Process models
  - prescribed uncertainties
- Candidate observing systems may be generated automatically by an optimisation algorithm, as in *Rayner et al.* (1996)
- Best network depends on network design tool
  -> Do network design with the system you want to use for assimilation later
- Construction of CCDAS-based tool for observing system design (FastOpt/LSCE/QUEST) part of proposed IMECC I3 FP6 Instrument (co-ordinated by PJ Rayner)

Supporting design of ocean observing system

- How sensitive is the target quantity (say European budget) to changes in a fixed input quantity (say ocean background flux)?

- Standard approach would be a new CCDAS run with modified input
- Optimisation is iterative procedure using, say, 100 runs of model and adjoint

- Efficient alternative for modified parameters via implicit function theorem (second order adjoint, Le Dimet et al. 2002)
  Optimisation for input field b yields optimal parameters x satisfying:

$$(d/dx)\, J(x,b) = 0$$

  This defines x as an implicit function of input field b, i.e. x(b) AND the sensitivity of the optimal parameters w.r.t. input field, dx/db is:

$$(d/dx)\, [(d/dx)\, J(x(b),b)]\, dx/db + (d/db)\, [(d/dx)\, J(x,b)] = 0$$

Supporting design of ocean observing system

- The sensitivity of the optimal parameters w.r.t. input field, db/dx is:

  (d/dx) [(d/dx) J(x(b),b)] dx/db + (d/db) [(d/dx) J(x(b),b)] = 0

- Sensitivity dx/db takes observational constraint into account

- To solve for dx/db, second derivative code required for
  (d/dx) [(d/dx) J(x(b),b)] : Hessian (is computed by CCDAS anyway)
  (d/db) [(d/dx) J(x(b),b)] : Has to be generated and evaluated

- Parameter sensitivity dx/db to be multiplied by diagnostic sensitivity df/dx:
  df/db = df/dx dx/db

- Approach to be demonstrated within FP6 CarboOcean IP (co-ordinated by C. Heinze) with:
  f : European budget
  b: ocean fluxes on 8 by 10 global grid

  Deliver sensitivity maps to support design of observation system;
  indicate ocean regions with high impact on European balance

Conclusions

- **CCDAS**
  - is a prototype with many limitations
  - can be extended to assimilate more observations (and quantify their impact)
  - can be extended by more processes (and deliver their optimal parameters)
  - can be used in prediction mode
  - can support observing-system design
  - is based on modern software (Fortran 95), automatic differentiation with TAF
  - looks well-suited as tool for integration and network design

- **Some of the issues not addressed:**
  - Prior estimates for parameters and uncertainties crucial -> Jens Kattge @ Jena
  - Two-step procedure sub-optimal (information from the second step missing in the first)
  - Relies on a single TEM: do test and compare different formulations but cannot not quantify uncertainty via differences among TEMs

- **More info, papers, etc:**
  http://CCDAS.org, http://FastOpt.com