A prototype of a carbon cycle data assimilation system (CCDAS) based on automatic differentiation

Heinrich Widmann[1], Marko Scholze[2], Peter Rayner[3], Wolfgang Knorr[1], Thomas Kaminski[4], Ralf Giering[4]

Copy of presentation at http://www.CCDAS.org

1 Max-Planck-Institut für Biogeochemie MPI-BGC Jena 2 Max-Planck-Institut für Meteorologie Max Planck Institute for Meteorology 3 CSIRO 4 FastOpt CAMELS

Overview

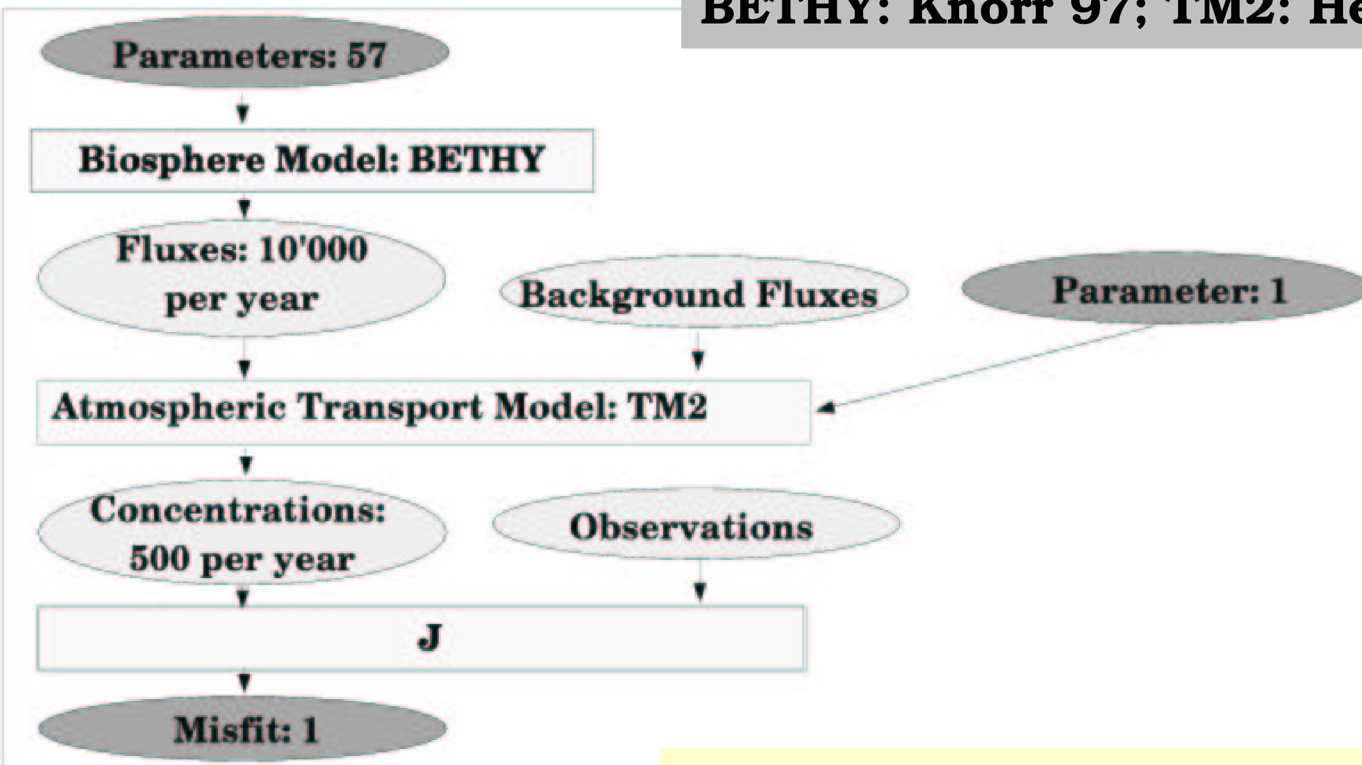
- Introduction
- Calibration step
- Prognostic step
- Model development within system
- Automatic Differentiation
- Summary

Introduction

- Carbon Dioxide is an important greenhouse gas
- Behaviour of carbon cycle is of interest:
  1. Prognose climate change (feed-backs)
  2. Diagnose quantities that are relevant in the context of the Kyoto-protocol (verification)
- Main Sources/Sinks: Fossil fuel emissions, Ocean, terrestrial biosphere/land use change
- There are large uncertainties in the parameters of terrestrial biosphere models. Hence the simulated fluxes to the atmosphere are uncertain, too.
- Terrestrial biosphere influences Kyoto accounting
- How can different pieces of information be combined to reduce uncertainties

Setup for Calibration Step

BETHY: Knorr 97; TM2: Heimann 95



$$J(m) = \tfrac{1}{2}\left[(m-m_0)C_m^{-1}(m-m_0)+(M(m)-d)C_d^{-1}(M(m)-d)\right]$$

Gradient Method

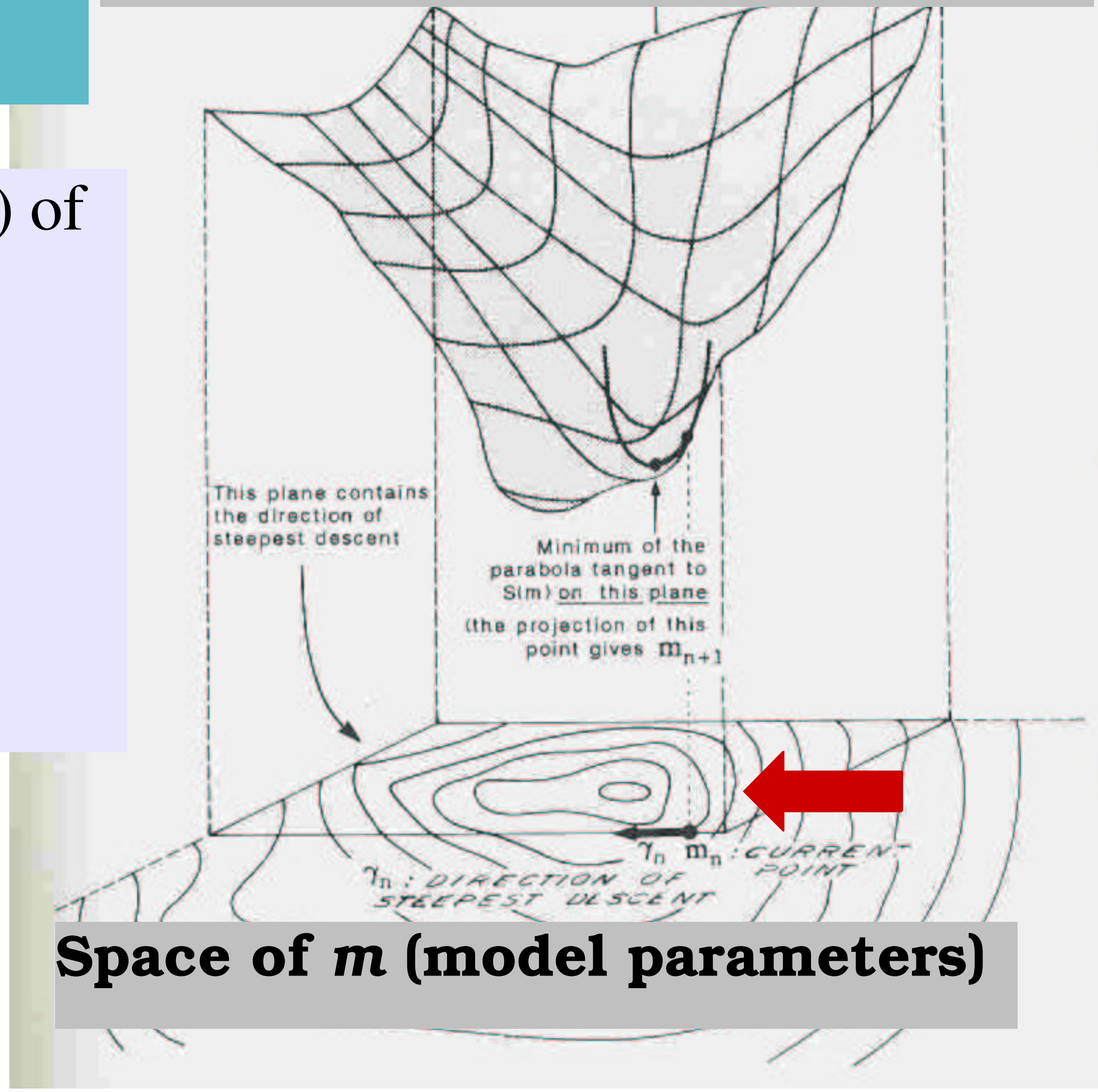
First derivative (Gradient) of $J(m)$ w.r.t. m (model parameters) :

$$-\partial J(m)/\partial m$$

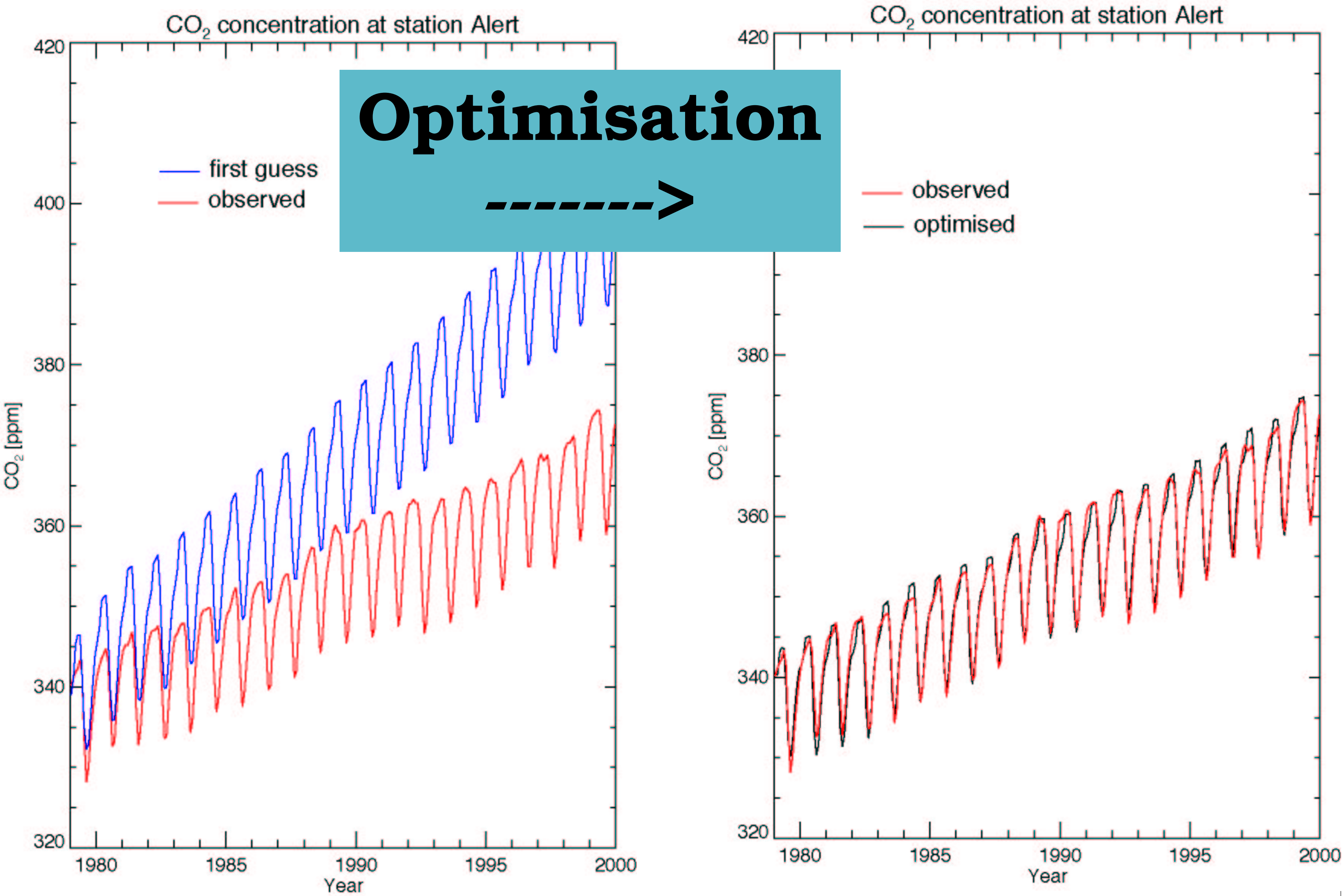
yields direction of steepest descent

Figure taken from **Tarantola** '87

Cost Funtion J(m)



Space of *m* (model parameters)


CO$_2$ concentration at station Alert
first guess
observed
CO$_2$ [ppm]
Year
Optimisation
-------->
CO$_2$ concentration at station Alert
observed
optimised
CO$_2$ [ppm]
Year

Covariances in Parameter Uncertainties

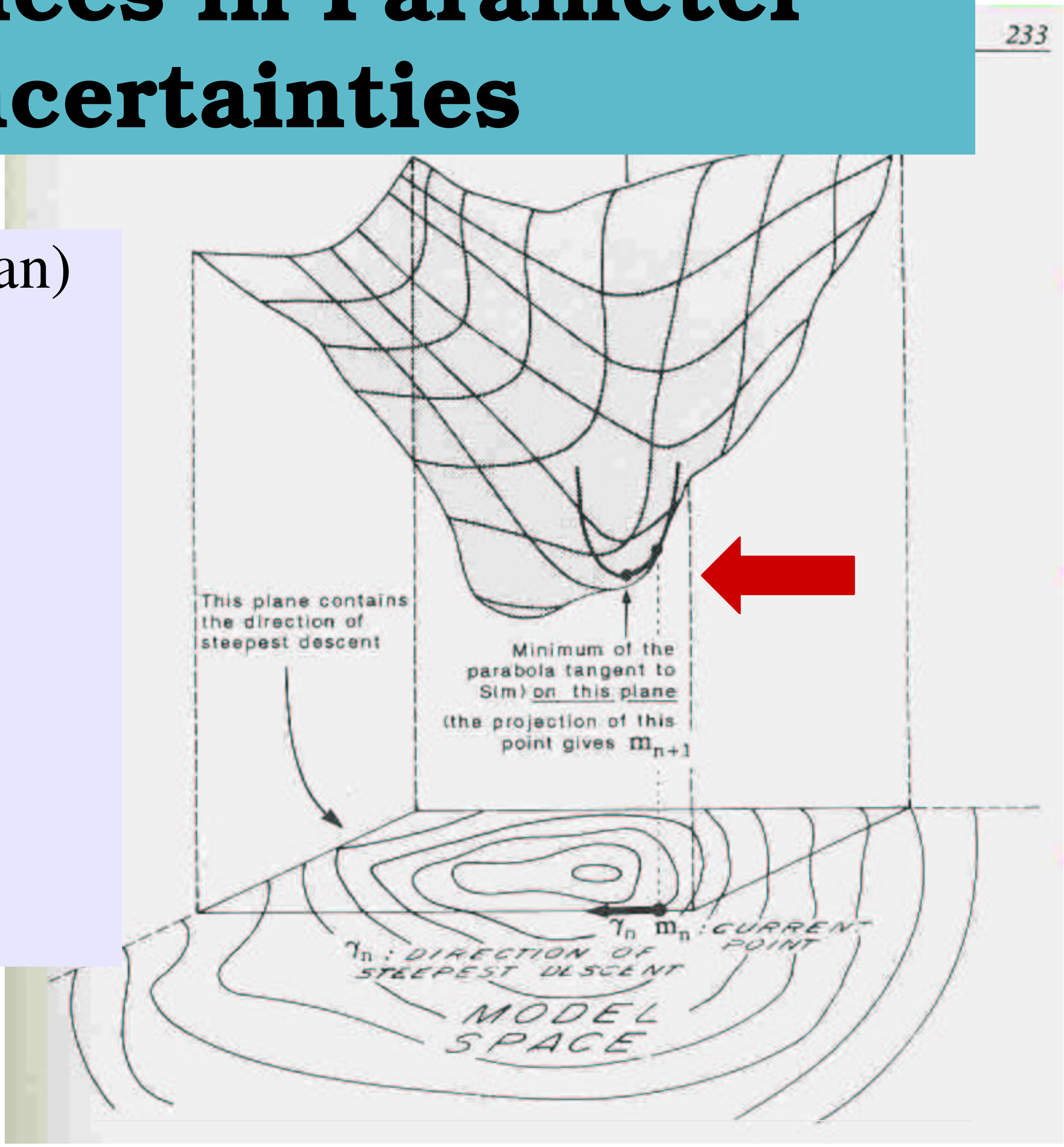
Second Derivative (Hessian) of $J(m)$:

$$H = \partial^2 J(m)/\partial m^2$$

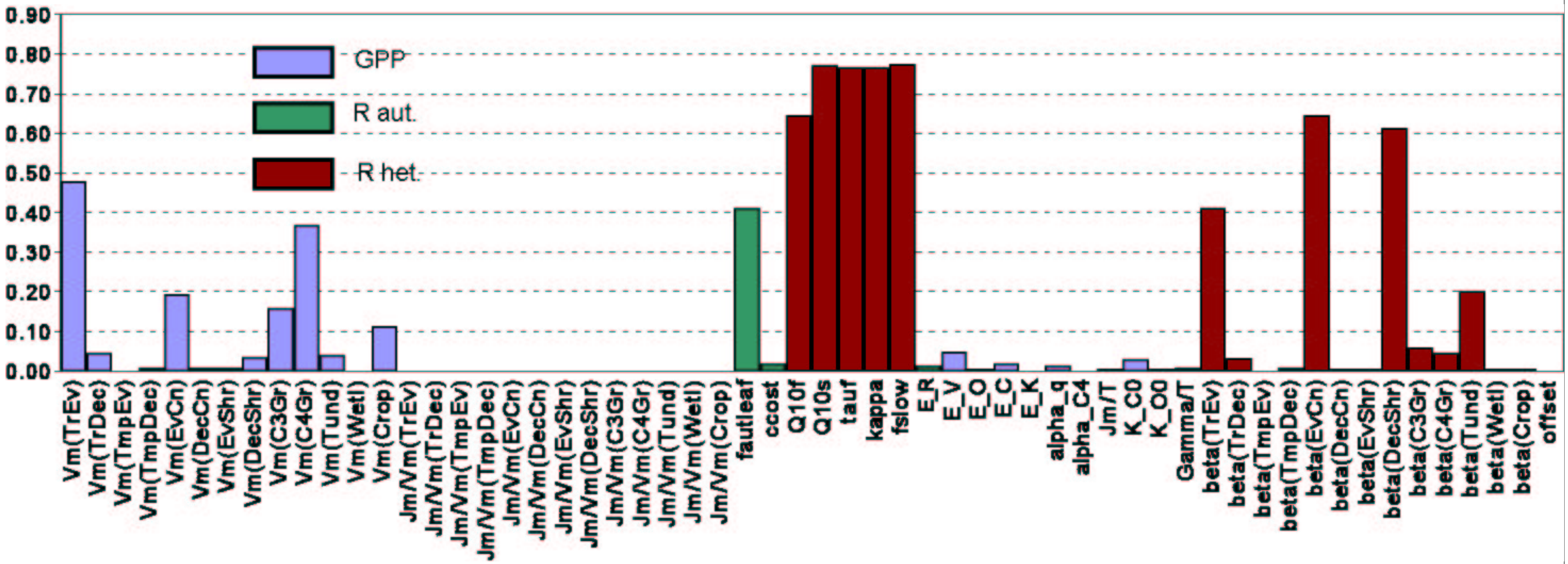
yields curvature of J, provides estimated uncertainty in m_{opt}

$$C_{y,\,post} = H^{-1}$$

Figure taken from **Tarantola** '87

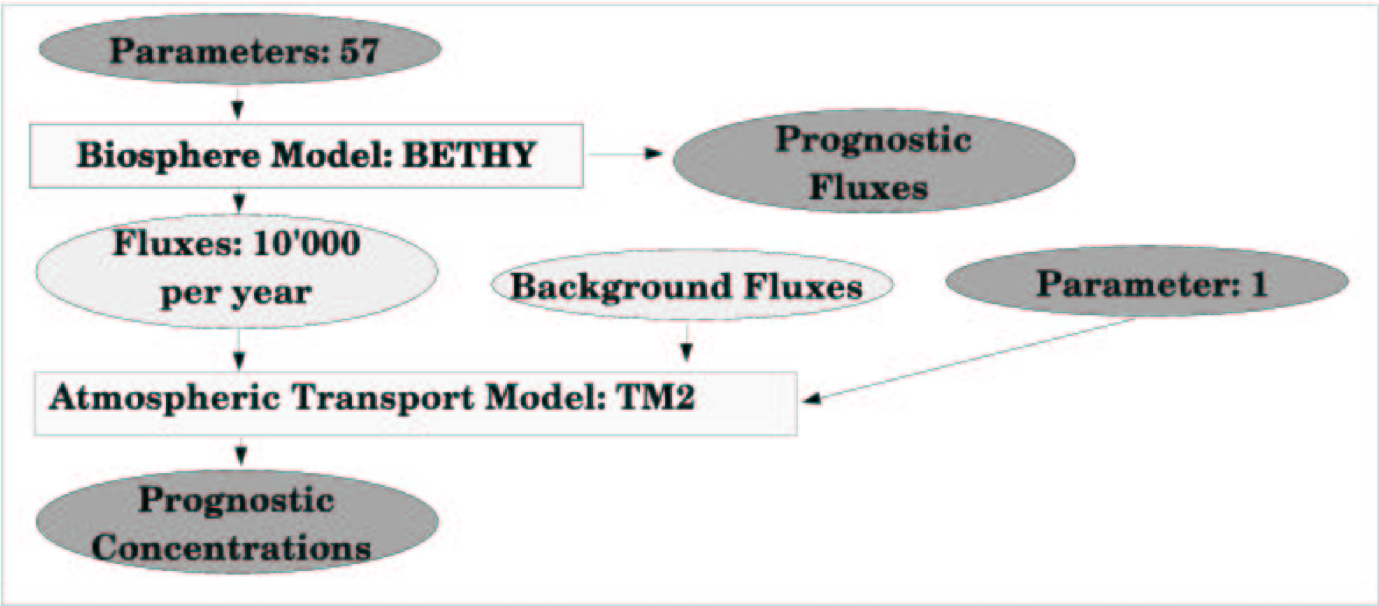


Relative reduction of uncertainties



Observations resolve about 10-15 directions in parameter space

Setup for prognostic step



BETHY: Knorr 97; TM2: Heimann 95

Covariances in Uncertainties of Prognostics

Prognostics, e.g. CO_2 fluxes, are a function of model parameters

$$Y = M(\boldsymbol{m}_{opt})$$

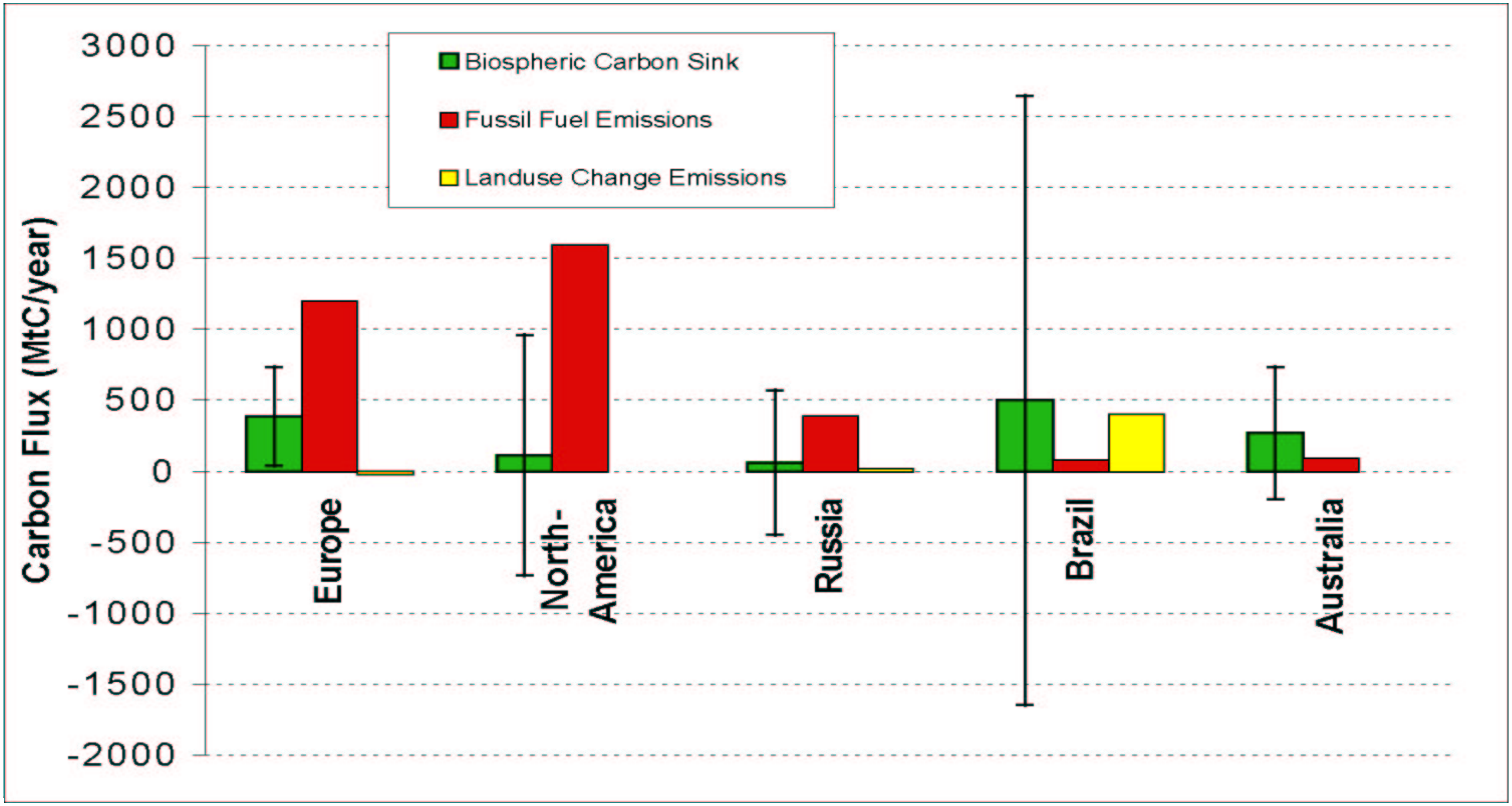
Posterior (after optimisation) covariance in uncertainties of prognostics

Jacobian matrix
(adjoint or tangent linear model)

$$C_{y,post} = (\mathrm{D}M)\, C_{m,post}\, (\mathrm{D}M)^T$$

Error covariance
of parameters

Regional Net Carbon Balance and Uncertainties



Max-Planck-Institut für Biogeochemie
Max-Planck-Institut für Meteorologie
Max Planck Institute for Meteorology
CSIRO
FastOpt
CAMELS

Model development within System

- System can test a given combination observational data + model formulation with uncertain parameters, and deliver optimal parameters, prognostics, and their a posteriori uncertainties
- Initial and boundary conditions can be included as unknowns/control variables (we have 1 already)
- Model is developed further within system
- Model development benefits from sensitivity information and comparison with data
- Work is ongoing, numbers are from model formulation we are not yet happy with…

Automatic Differentiation

- Uses adjoint, tangent linear and Hessian code
- All derivative code generated from Fortran-90 source code of model (~5500 lines) by AD tool TAF
- Process fully automatic: no hand coding

Automatic Differentiation

Performance: CPU time

CPU time in multiples of CPU time for model (Linux, XEON 2GHz, Lahey lf95: -tpp -O):

- **tangent linear model: 1.5**
- **adjoint model: 3.4**
- **Hessian (12 columns): 50**

Automatic Sparsity Detection (ASD) in TAF

- TAF generates code to evaluate the sparsity pattern of a Jacobian (ASD)
- ASD-code propagates bit vectors -> is very efficient
- Can do forward mode and reverse mode

Example: DM for M: Paramters -> Net CO2 Flux in latitude bands

Parameters --->

latitude - - ->

```
0000000000000000000000000000000000000000000000000000000000
0000000000100000000000000001111111111100001110000000000000
0000001100110000000110011011111111111010111100000011001100
0001111110110000111111011011111111111010111100011111101100
0001111111111000111111111111111111111111111110001111111110
0001111111111000111111111111111111111111111110001111111110
0111101111110111101111111111111111111111111110111101111110
0111101111110111101111111111111111111111111110111101111110
1111101111111111110111111111111111111111111111111011111110
```

Max-Planck-Institut für Biogeochemie

Max-Planck-Institut für Meteorologie
Max Planck Institute for Meteorology

CSIRO

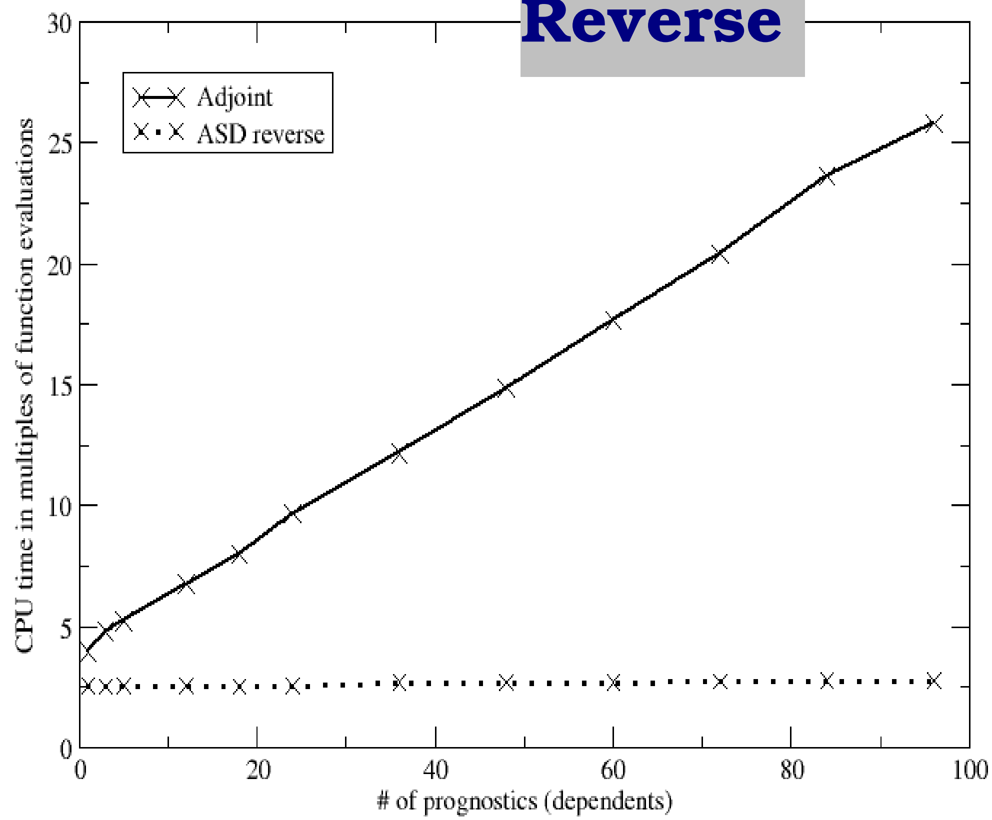
FastOpt

CAMELS

Jacobian and ASD Performance

Forward
58 independents
Jacobian: 12
ASD: 1.3

Reverse



CPU time quantified in multiples of function evaluations

Summary

- AD is **key technology** in CCDAS:
  Allows efficient optimisation
  and propagation of uncertainties
- Concept can be **generalised to other modelling** systems
- Automatic differentiation is essential
  to reduce the delay from model development
  to data assimilation