Supporting the improvement of the carbon observing system by quantitative network design

M. Voßbeck (http://FastOpt.com)

T. Kaminski, P. Rayner, M. Scholze, E. Koffi, and R. Giering

5. Scientific Computing Seminar, Kiel 2010

SIXTH FRAMEWORK PROGRAMME

INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE

University of BRISTOL

FastOpt

Motivation

What is the question?

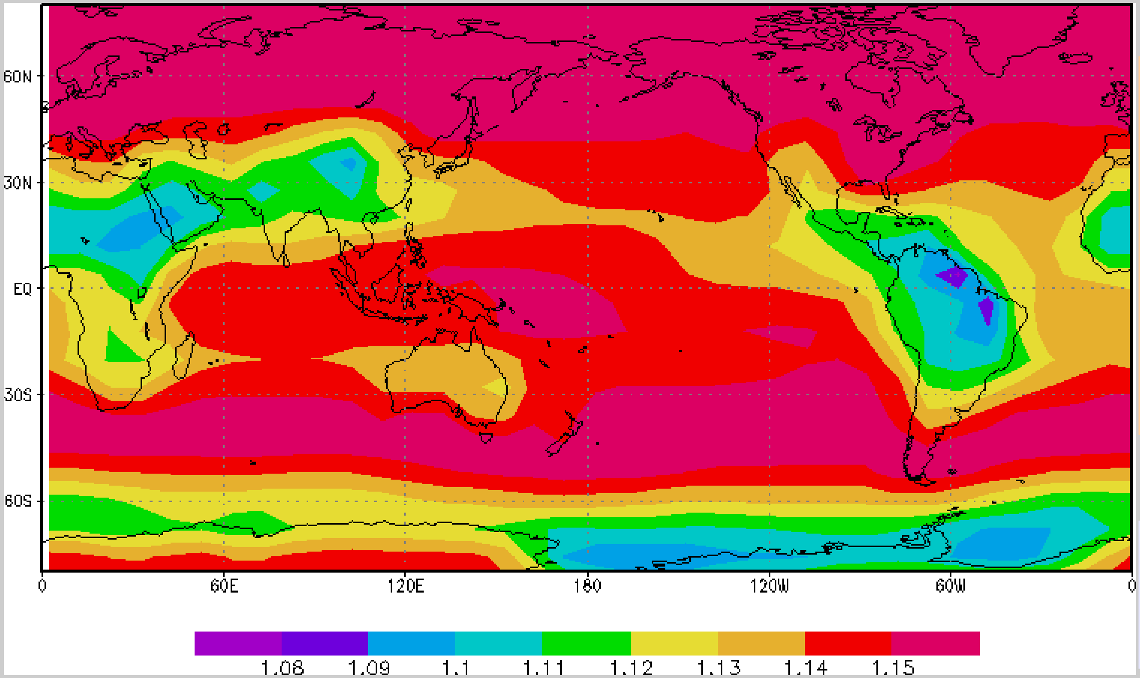
Can construct a machinery that,
for a given network and a given target quantity,
can approximate the uncertainty
with which the value of the target quantity is
constrained by the observations

Outline

- Motivation
- Example
- Method
- Demo
- Summary
- Links to further information

Example Linear Model

Target quantity:
Standard deviation of Global Ocean Uptake (GtC/yr)



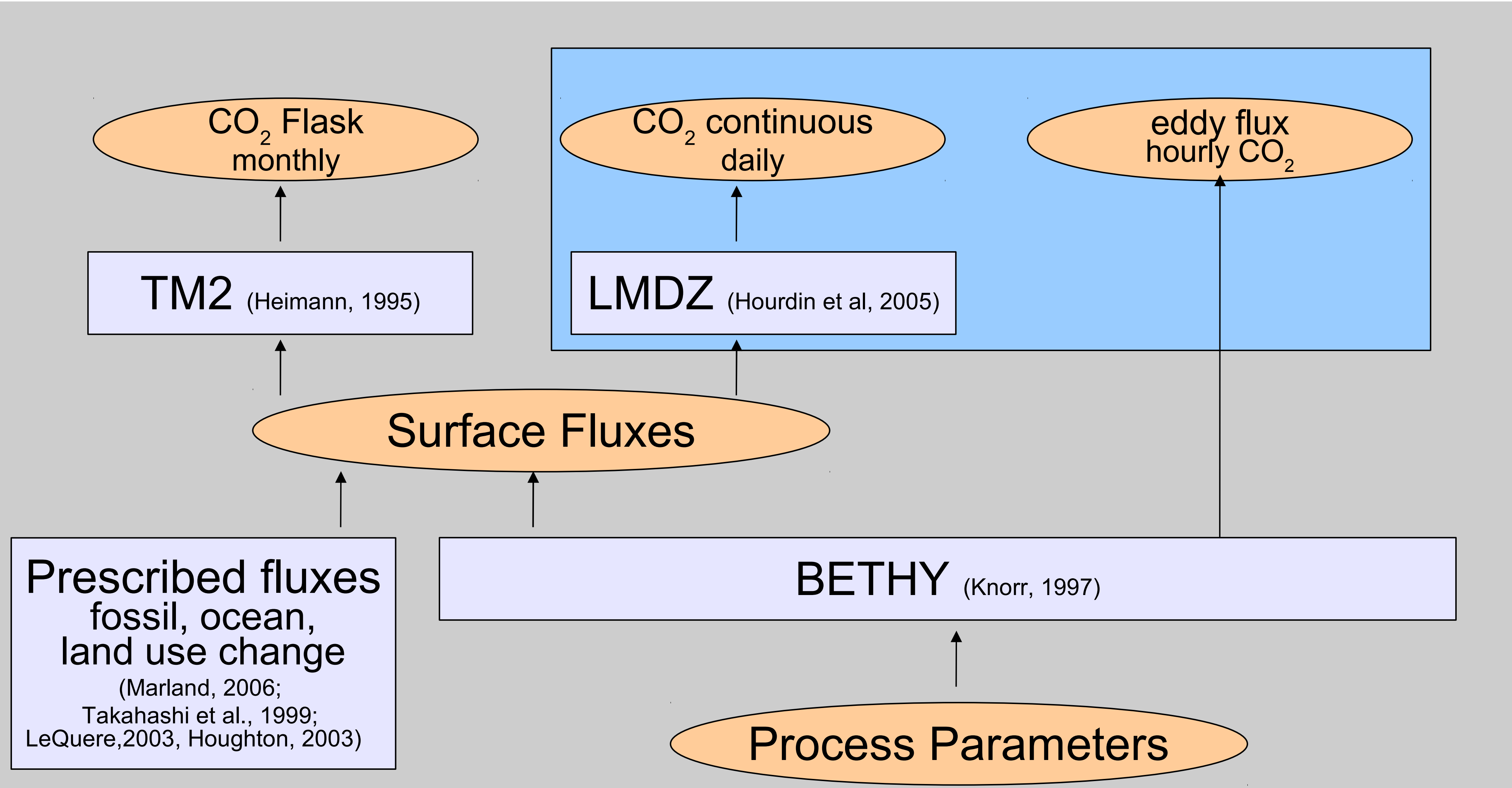
Rayner et al.
(Tellus, 1996)

Inverse model based on atmospheric tracer model

Existing network yields 1.16 GtC/yr

- Global constraints with narrow uncertainty on "ocean uptake + terrestrial uptake" couples information from disaparate regions:
  Improvement in knowledge of Amazonian fluxes improves knowledge of both total land and ocean fluxes!

Carbon Cycle Data Assimilation System (CCDAS) Forward Modelling Chain



SIXTH FRAMEWORK PROGRAMME

IMECC INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE

University of BRISTOL

LSCE

FastOpt

Model and Observational Uncertainties

- No observation/no model is perfect.
- It is convenient to quantify observations and their model counterpart by probability density functions (PDFs).
- The simplest assumption is that they are Gaussian.

$$\sigma_d^2 = \sigma_{obs}^2 + \sigma_{mod}^2$$

total uncertainty

uncertainty from model error

uncertainty from observational error

- If the observation refers to a point in space and time, there is a representation error because the counterpart simulated by the model refers to a box in space and time:
  the corresponding uncertainty must be accounted for either by the observational or by the model contribution to total uncertainty.

SIXTH FRAMEWORK PROGRAMME

IMECC INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE

University of BRISTOL

FastOpt

Computing posterior parameter uncertainty

- M: x → d maps the process parameters onto the observations (CO_2 concentrations, flask and continuous, CO_2 hourly fluxes)
- Introduce cost function:

$$J(x) = \tfrac{1}{2} \left[(Mx-d)^T C_d^{-1} (Mx-d) + (x - x_{pr})^T C_{pr}^{-1} (x - x_{pr}) \right]$$

- x_{pr} : priori information on the process parameter
  C_d, C_{pr} : covariance matrices of uncertainties in observations and prior parameters respecively.

- If model M is **linear**, and if data d and priors x_{pr} have Gaussian PDF, then posterior PDF is also Gaussian (Tarantola 87):

$$\rho (x) = K * \exp(-\tfrac{1}{2}(x - x_{po})^T C_{po}^{-1}(x - x_{po}))$$

  with mean value:

$$x_{po} = x_{pr} + [M^T C_d^{-1} M + C_{pr}^{-1}]^{-1} M^T C_d^{-1} (d - M x_{pr})$$

- x_{po} minimises J(x)

- Uncertainty given by:

$$C_{po}^{-1} = M^T C_d^{-1} M + C_{pr}^{-1} = d^2 J(x_{po}) / dx^2$$

Computing posterior parameter uncertainty

- For **non-linear, differentiable** model M posterior distribution of parameters is still approximately Gaussian (Tarantola 87)
$$\rho (x) \approx K * \exp(-\tfrac{1}{2}(x - x_{po})^T C_{po}^{-1}(x - x_{po}))$$
- Equation for posterior uncertainties holds as approximation:
$$C_{po} \approx (d^2J(x_{po}) / dx^2)^{-1}$$
- x_{po} now determined by minimising cost function J
- Use gradient based optimisation to solve, x_{po} fullfills:
$$dJ(x_{po}) / dx = 0$$
- Efficient first and second order derivatives required, use adjoint code for frist order derivatives
- All derivative code generated from model code by automatic differentiation tool **TAF** (Giering and Kaminski, 1998)

SIXTH FRAMEWORK PROGRAMME

INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE

University of BRISTOL

FastOpt

Network Desgin Tool: Simplifications of implementation

- For a number of i = 1,...,n_d of **uncorrelated** observations (C_d diagonal matrix) with (user provided) standard deviations σ_{d_i} the cost function can be written:

$$J(x) = ½ (x – x_{pr}) C_{pr}^{-1} (x – x_{pr}) + ½ \sum_{i=1}^{nd} ((M_i(x) – d_i)/\sigma_{d_i})^2$$

with:

$$C_{po}^{-1} = d^2J(x_{po}) / dx^2 = C_{pr}^{-1} + \sum_{i=1}^{nd} (1/\sigma_{d_i}^{2}) d^2/dx^2(M_i(x) – d_i)^2$$

uncertainies in observations **AND** model

- Using synthetic data d = M(x), model contribution in formula for C_{po} can be precomputed (do not depend on user input!)

SIXTH FRAMEWORK PROGRAMME

IMECC INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE

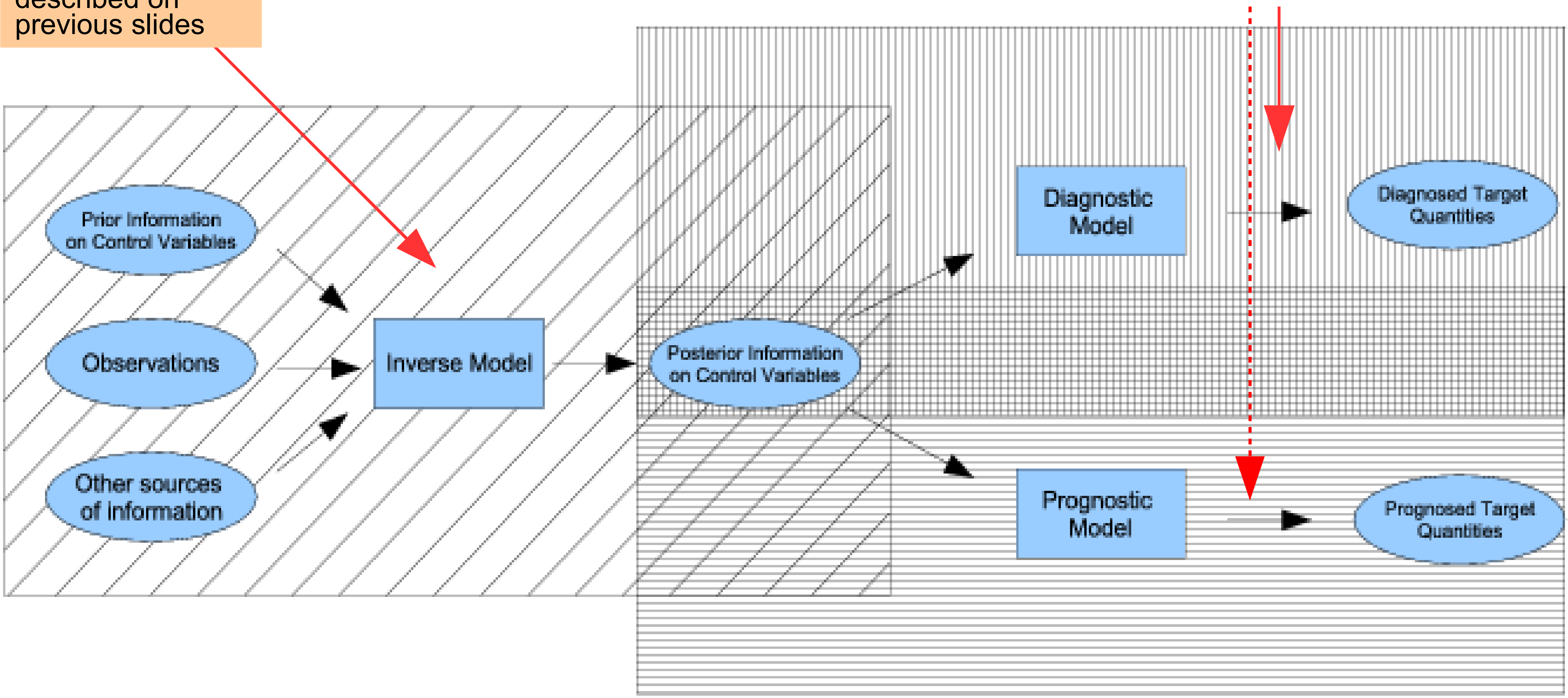
University of BRISTOL

FastOpt

CCDAS scheme

Inversion step: described on previous slides

Propagation step: shown next



Propagation step

- N: x → y maps process parameters onto the target quanity
  (diagnostic or prognostic)
- Diagnostic run means that it is based on same domain (spatial and temporal) as used in the inversion step, whereas prognostic run is (at least in parts) outside this domain
- PDF of y can again be approximated by a Gaussian with mean
  $$y_{po} = N(x_{po})$$
  and covariance
  $$C(y_{po}) = D(N)\,C_{po}\,D(N)^T + C(y_{mod})$$
- $D(N) = dN(x_{po})/dx$ is the linearisation of N around x_{po} (Jacobian), $C(y_{mod})$ accounts for the uncertainty of the target quantity for errors in N
- Recall $C_{po}^{-1} = C_{pr}^{-1} + \sum_{i=1}^{nd} (1/\sigma_{d_i}^{2})\, d^2/dx^2(M_i(x) - d_i)^2$
- Answers the initial question and eventually yields mapping from uncertainties of (potential) observations to uncertainty of the target quantity:
  $C(y_{po}) = C(y_{po}; \sigma_{d_1}, \ldots, \sigma_{d_{ND}})$

Propagation step also requires derivatives!

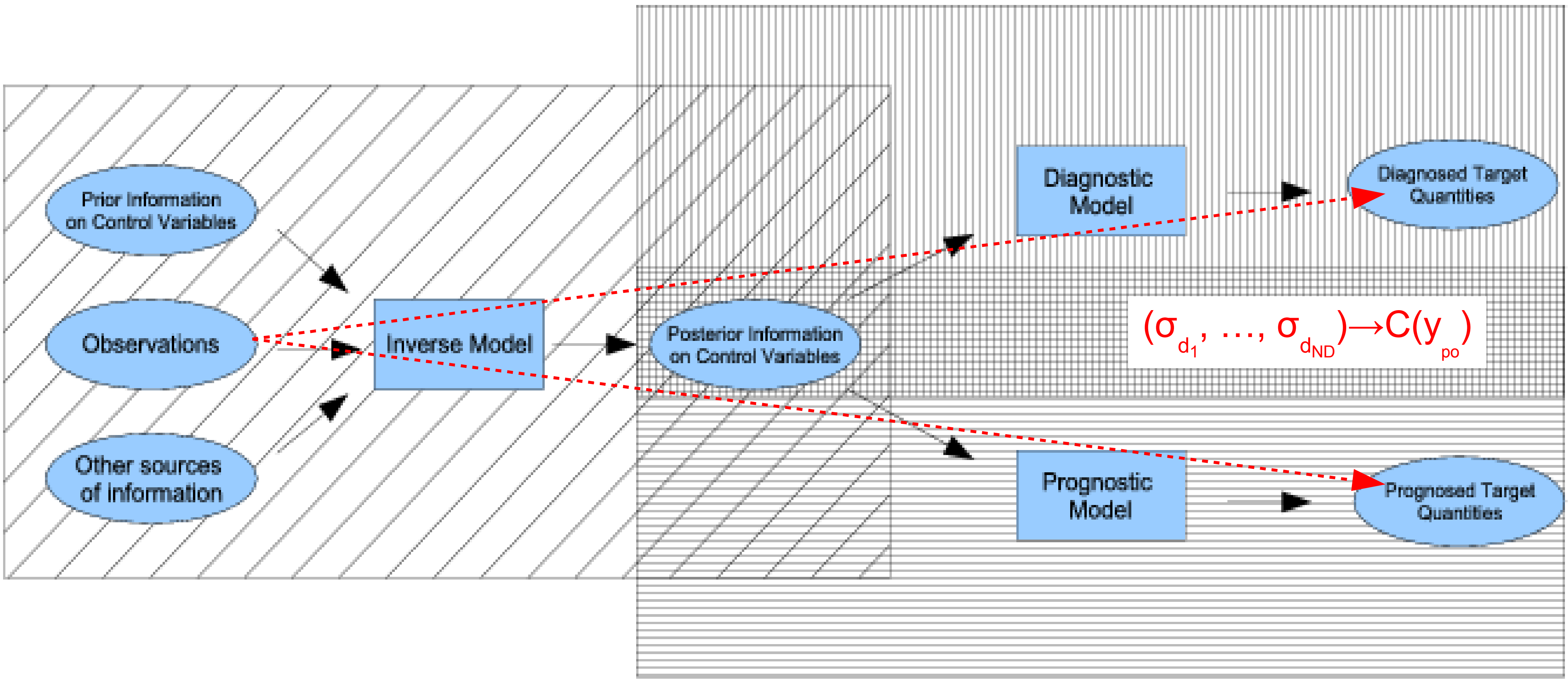
SIXTH FRAMEWORK PROGRAMME

INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE

University of BRISTOL

FastOpt

Uncertainties of Target Quantities



SIXTH FRAMEWORK PROGRAMME

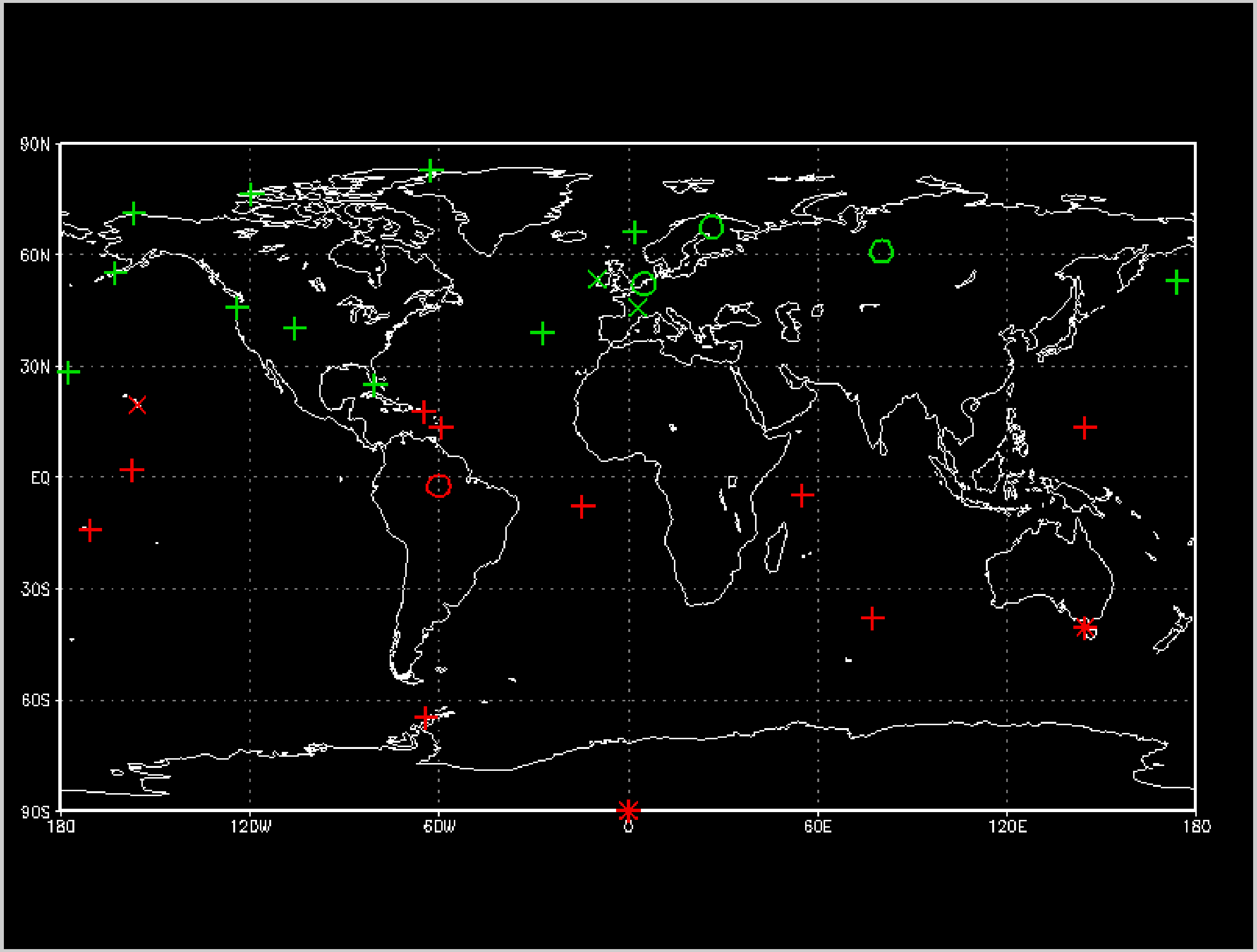
IMECC INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE

University of BRISTOL

LSCE

FastOpt

Sketch of Network DesignTool



Observations [sigma]

+ Flask [enter]
x Continuous [enter]
o Eddy Flux [enter]

|Compute|

Targets [sigma]

European Uptake [ ]
Global Uptake [ ]

Demo: Network Desgin Tool

SIXTH FRAMEWORK PROGRAMME

IMECC INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE

University of BRISTOL

LSCE

FastOpt

Summary: Assumptions and Ingredients

Assumptions:

- Gaussian uncertainties on priors, observations, and from model error (or function of Gaussian, e.g. lognormal)
- Model not too non linear

Ingredients:

- Assimilation system that can (efficiently) propagate uncertainties; helpful: Adjoint, Hessian, and Jacobian codes
- Ability to estimate uncertainties for priors, observations and due to model error; requires expertise of observationalists and modellers
- Need to take logistic constraints into account

Further Information

Terrestrial assimilation system applications and papers:
http://CCDAS.org

The corresponding network design project:

http://IMECC.CCDAS.org

with links to papers on network design

More assimilation systems, applications and papers:

http://FastOpt.com

SIXTH FRAMEWORK PROGRAMME

INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE

University of BRISTOL

FastOpt