Quantitative Design of Observational Networks

T. Kaminski (http://FastOpt.com)

R. Giering, E. Koffi

P. Rayner, M. Scholze, and M. Voßbeck

GEOTRACES WS, September 2007

SIXTH FRAMEWORK PROGRAMME

IMECC – INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE

University of BRISTOL

LSCE

FastOpt

Motivation

What is the question?

Motivation

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

•Which assumptions?

•Links to further information

•Discussion -> Any potential showstoppers for application in GEOTRACES?

Example Linear Model

- From Rayner et al. (Tellus, 1996)
- Inverse model based on atmospheric tracer model
- Extend given atmospheric network CO2
- Target quantity: Global Ocean uptake

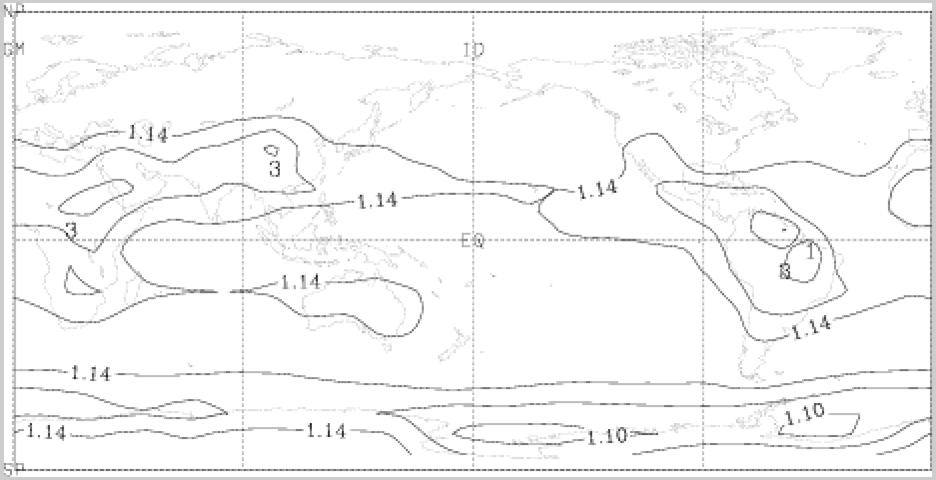


Figure shows additional station locations for two experiments,
“1” additional site allowed / “3” additional sites allows

Posterior Uncertainty

If the model was linear:

$$J(\tilde{\mathbf{x}}) = \frac{1}{2}\left[(\mathbf{M}\tilde{\mathbf{x}} - \mathbf{d})^T \mathbf{C}(d)^{-1}(\mathbf{M}\tilde{\mathbf{x}} - \mathbf{d}) + (\tilde{\mathbf{x}} - \mathbf{x_0})^T \mathbf{C}(x_0)^{-1}(\tilde{\mathbf{x}} - \mathbf{x_0})\right]$$

and data + priors have Gaussian PDF, then the posterior PDF is also Gaussian:

$$\rho(x) \sim e^{-J(x)}$$

with mean value:

$$\mathbf{x} = \mathbf{x_0} + [\mathbf{M}^T \mathbf{C}(d)^{-1}\mathbf{M} + \mathbf{C}(x_0)^{-1}]^{-1}\mathbf{M}^T \mathbf{C}(d)^{-1}(\mathbf{d} - \mathbf{M}\mathbf{x_0})$$

and uncertainty:

$$\mathbf{C}(x)^{-1} = \mathbf{M}^T \mathbf{C}(d)^{-1}\mathbf{M} + \mathbf{C}(x_0)^{-1}$$

which are related to the Hessian of the cost function:

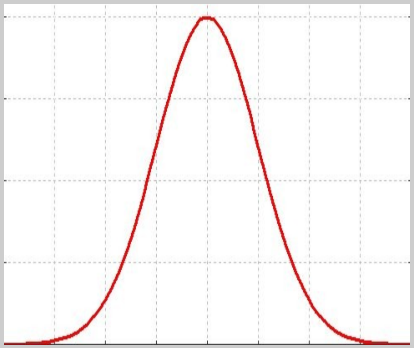
$$\mathbf{C}(x)^{-1} = \mathbf{H}$$

$$\frac{\partial^2 J}{\partial x_i \partial x_j}$$

For a non-linear model, this is an approximation

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.

Uncertainty for target in 2 steps

x: Parameters

x_{pr}: Priors

C_{pr}: Uncertainties

$M(x)$: Model

d: Observations

C_d: Their uncertainties

σ_{d_i}: Uncorrelated!

$J(x)$: Cost function

$\frac{d^2 J(x)}{dx^2}$: Hessian

x_{po}: Posterior parameters

C_{po}: Posterior uncertainties

$y(x)$: Target quantity

σ_y: Its uncertainty

$$J(x) = \frac{1}{2}(x - x_{pr})^T {C_{pr}}^{-1} (x - x_{pr}) + \frac{1}{2} \sum_{i=1,nd} \left(\frac{M_i(x) - d_i}{\sigma_{d_i}} \right)^2$$

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

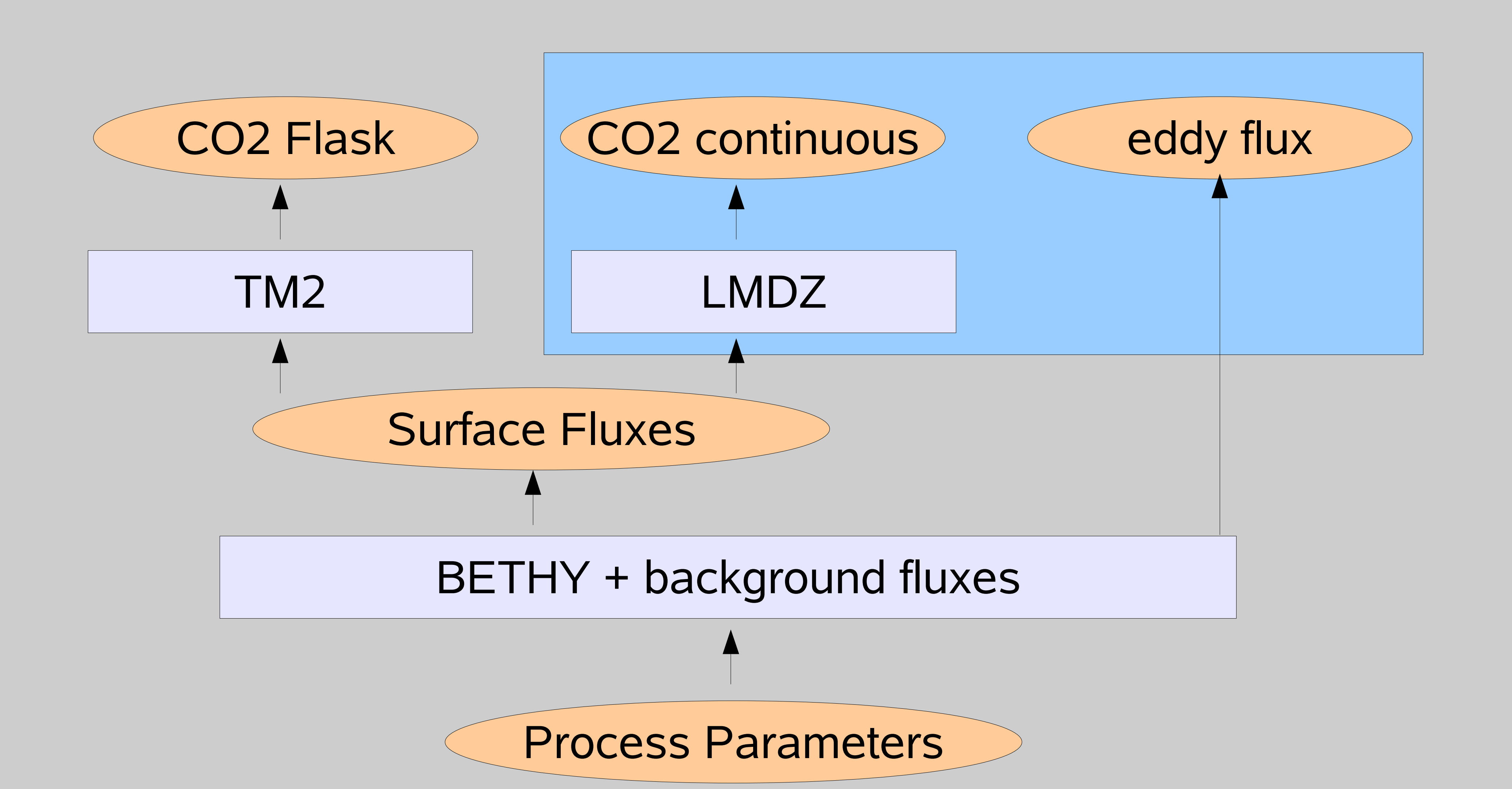
- Hessian independent of x for linear model
- For synthetic data use $d = M(x)$.
- Decomposes nicely, can precompute model contribution

$$C_{po} \approx \frac{d^2 J(x_{po})}{dx^2}^{-1}$$

$$\sigma_y \approx \frac{dy(x_{po})}{dx} C_{po} \frac{dy(x_{po})}{dx}^T \approx \frac{dy(x_{po})}{dx} \frac{d^2 J(x_{po})}{dx^2}^{-1} \frac{dy(x_{po})}{dx}^T$$

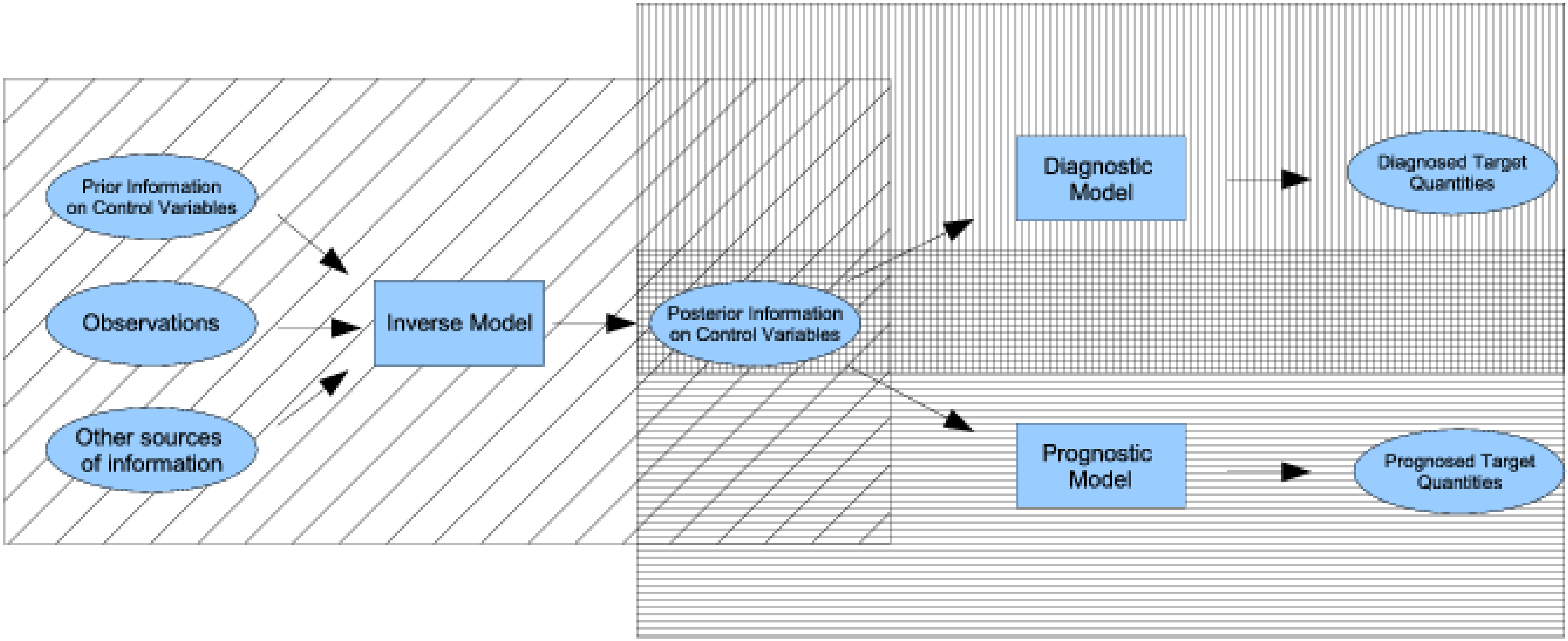
Derivative information can be efficiently provided by compiler tool TAF

Carbon Cycle Data Assimilation System (CCDAS) Forward Modelling Chain

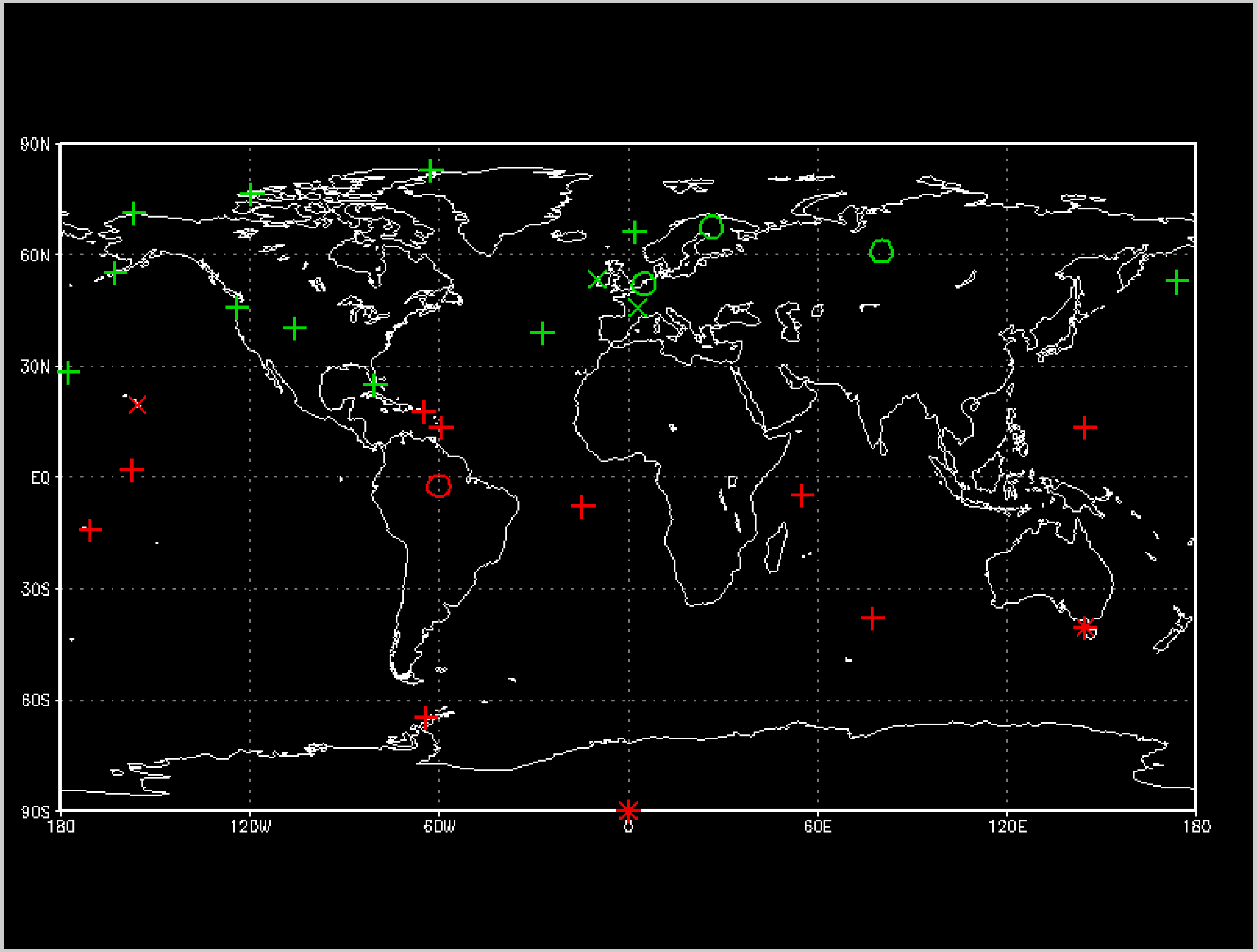


SIXTH FRAMEWORK PROGRAMME
IMECC
INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE
University of BRISTOL
LSCE
FastOpt

CCDAS scheme



Sketch of Network Designer



Observations [sigma]

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

|Compute|

Targets [sigma]

European Uptake [ ]
Global Uptake [ ]

SIXTH FRAMEWORK PROGRAMME

IMECC INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE

University of BRISTOL

FastOpt

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
- What else?

Ingredients:

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

Further Information – The commercial!

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

The corresponding network design project:

http://IMECC.CCDAS.org

with link to paper on network design
(Kaminski and Rayner, in press)

More assimilation systems, applications and papers:

http://FastOpt.com

Thomas.Kaminski@FastOpt.com

SIXTH FRAMEWORK PROGRAMME

IMECC INFRASTRUCTURE FOR MEASUREMENT OF THE EUROPEAN CARBON CYCLE

University of BRISTOL

LSCE

FastOpt