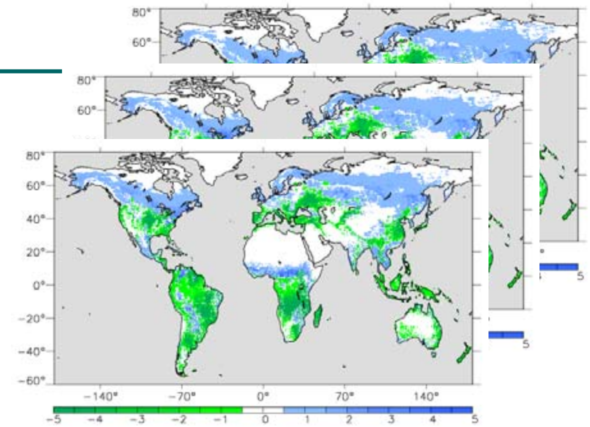
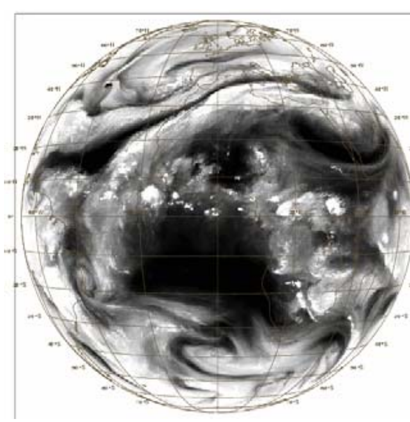
Data assimilation for large state vectors



Frédéric Chevallier

Laboratoire des Sciences du Climat et de l’Environnement
Gif-sur-Yvette
France

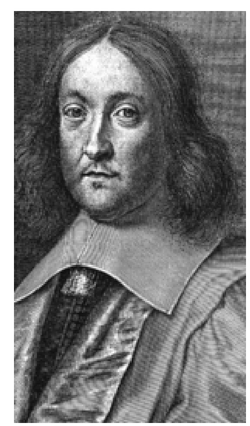


Outline

- Introduction
- Basic concepts
- Prior information
- Analytical solution
- Variational solution
- Monte Carlo solution
- Diagnostics

In the beginning…

- Summer 1654: correspondence between Blaise Pascal and Pierre Fermat
  - 2 players gamble 32 pistols each in a game of chance
  - The contest is best in five sets
  - The game stops before the end
  - How should the 64 pistols be shared?



In the beginning…

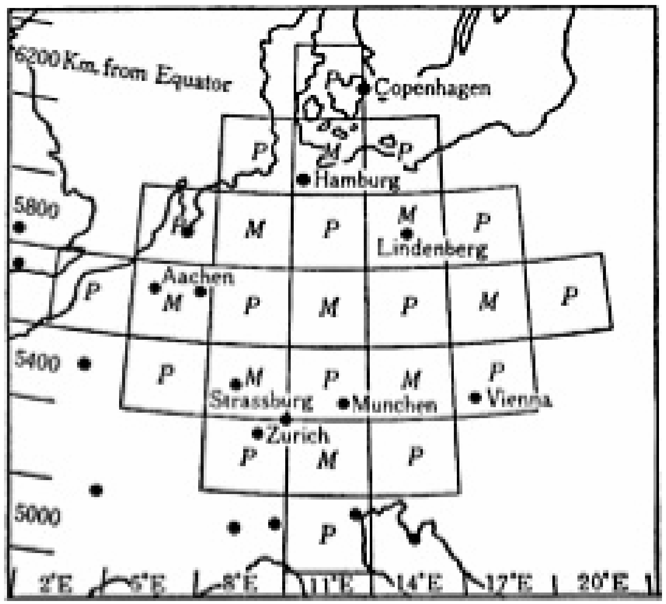
- Summer 1654: correspondence between Blaise Pascal and Pierre Fermat
  - 2 players gamble 32 pistols each in a game of chance
  - The contest is best in five sets
  - The game stops before the end
  - How should the 64 pistols be shared?
- Chance becomes a scientific topic
- Virtual realizations of a random variable

Two examples of application

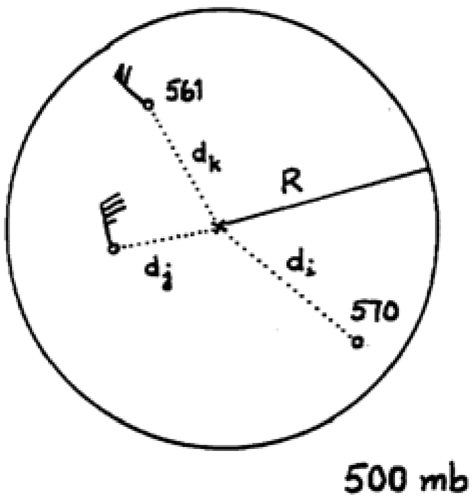
- Weather analyses
- Carbon flux estimation

Weather analyses

- How to start a forecast



Observations for 7 a.m., 20 May 1910
for a forecast at 1 p.m.
Six weeks of hand-computation
Richardson (1922)

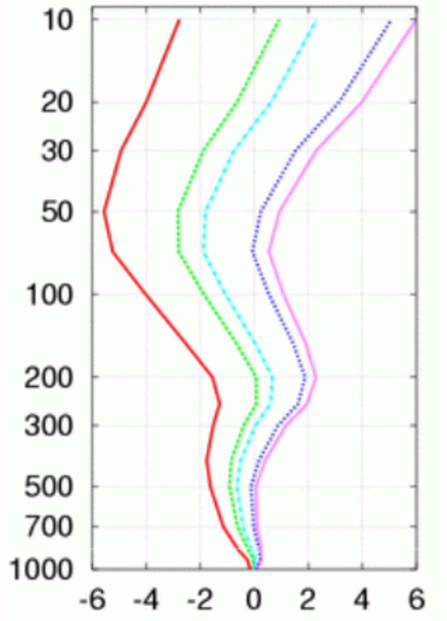


Local interpolation
Cressman (1959)

Weather analyses

- Uncertain observations
  - Need of a statistical approach

Model – Obs
Solar elevations
Low
High
Andrae et al. (2004)



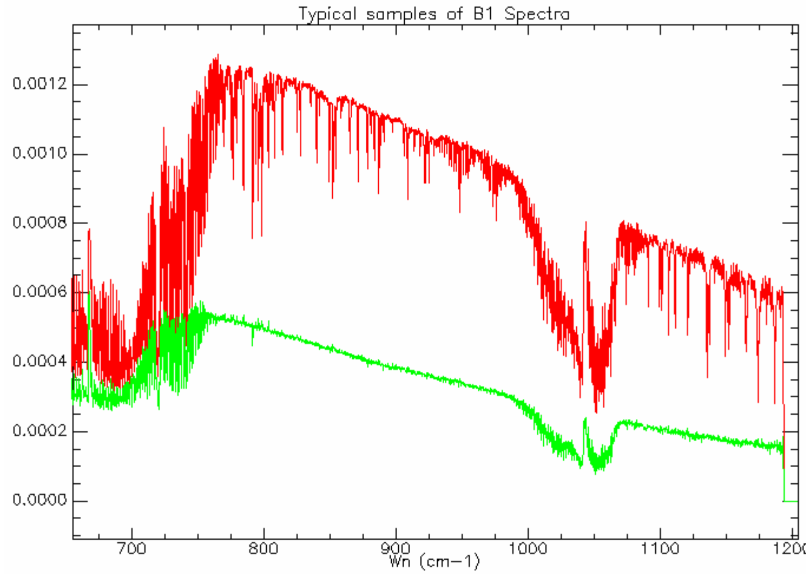
Pressure (hPa)

Temperature (K)

Weather analyses

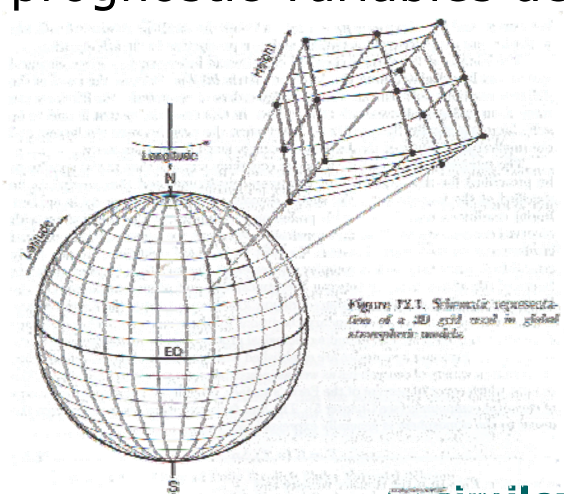
- Uncertain observations
- Indirect observations

IASI spectrum



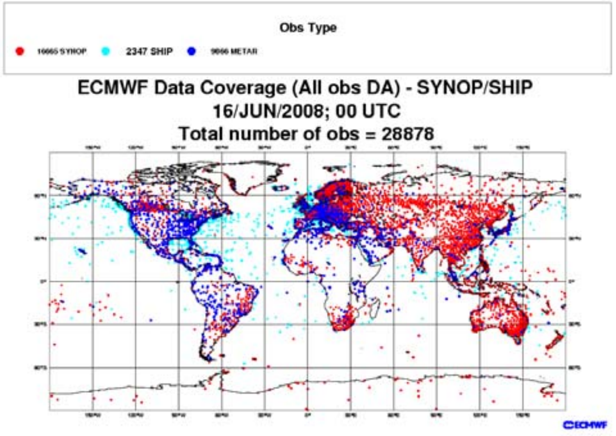
Weather analyses

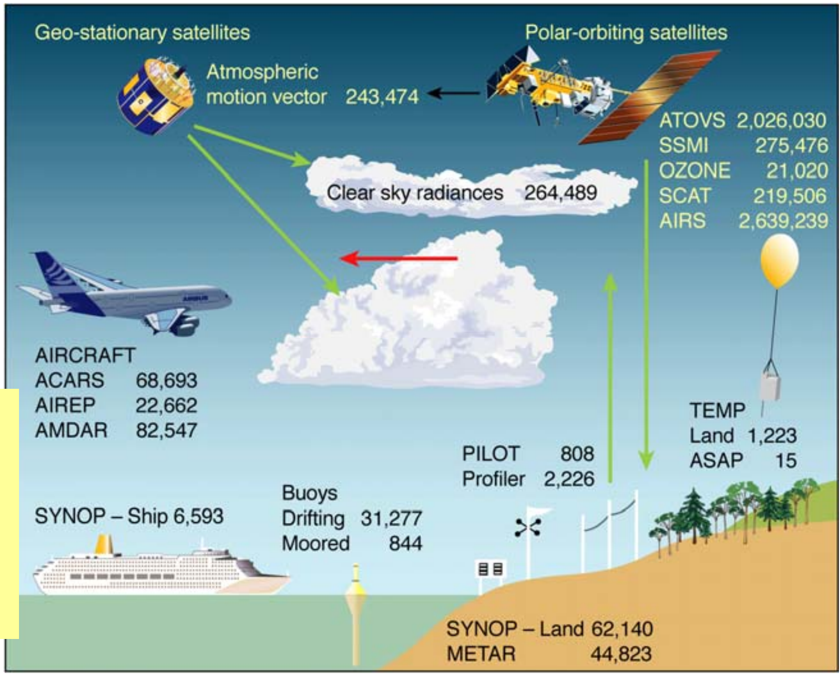
- Uncertain observations
- Indirect observations
- Very large state vector
  - ECMWF grid points:
    - 76,757,590 in upper air
    - 91 vertical levels, 25km horizontal
    - 6 prognostic variables defined at each grid point



Weather analyses

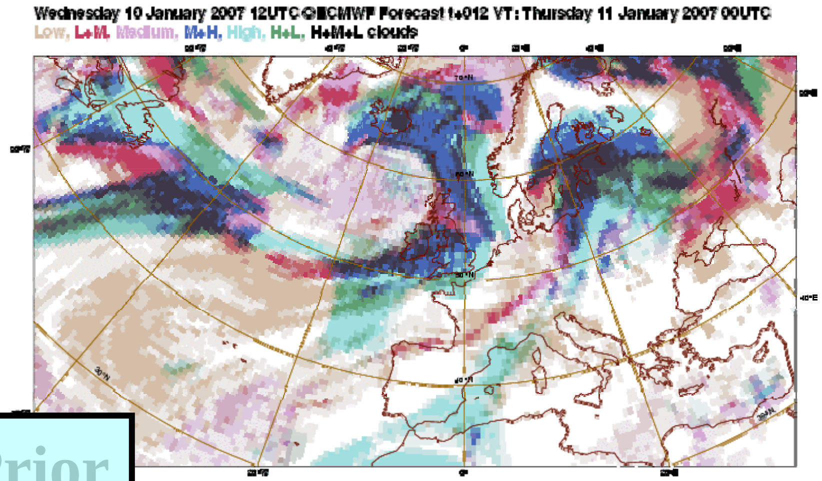
- Uncertain observations
- Indirect observations
- Very large state vector
- Very large observation vector
  - Irregular data coverage





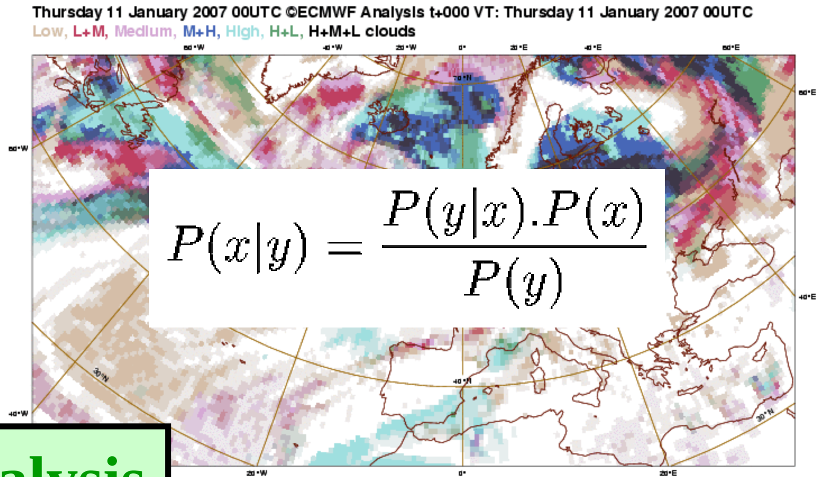
Data sources
at ECMWF
# of obs. assimilated
Over 24h on 13/02/2006

NWP analysis



Prior

Observations

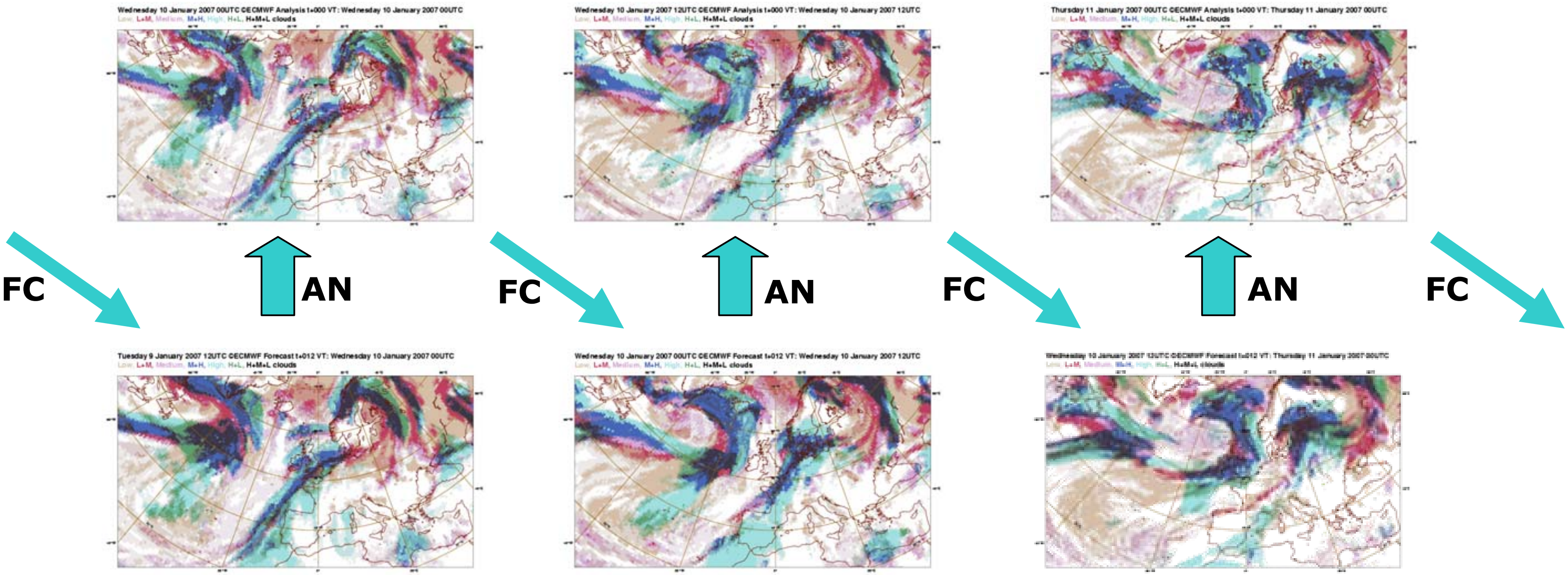


$$P(x|y) = \frac{P(y|x).P(x)}{P(y)}$$

Analysis

NWP sequential filter

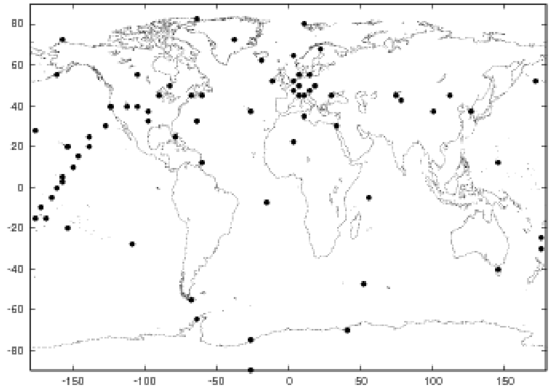
10 Jan. 00 UTC | 10 Jan. 12 UTC | 11 Jan. 00 UTC

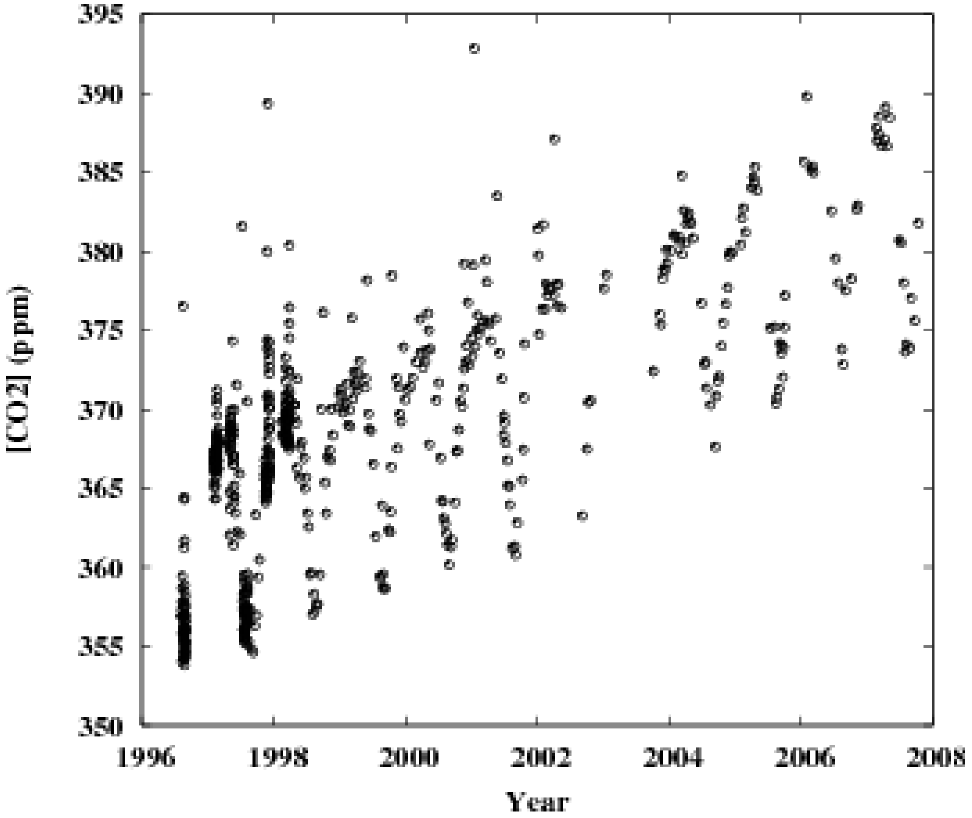


+ Obs. | + Obs. | + Obs.

Estimation of carbon fluxes

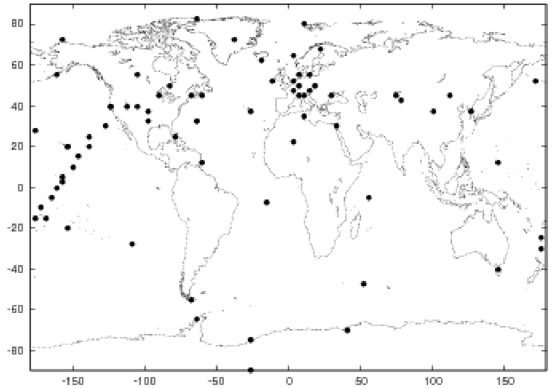
- CO_2 concentrations measured at Mace Head (IRL)

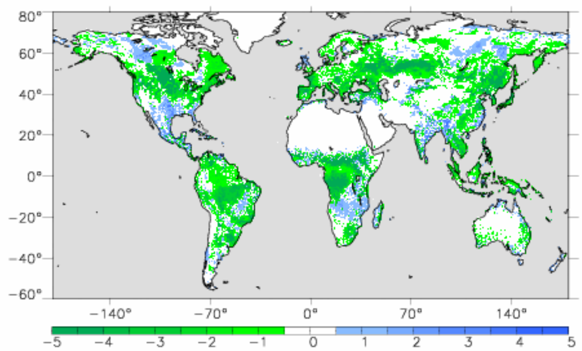


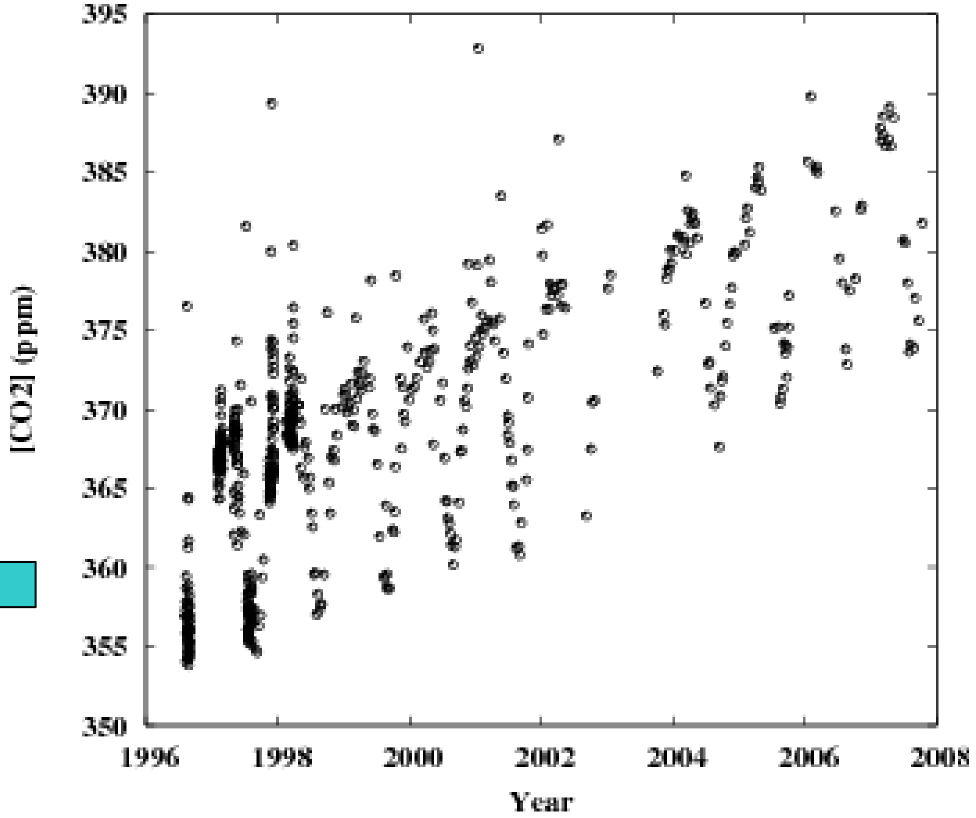


Estimation of carbon fluxes

- CO_2 concentrations measured at Mace Head (IRL)





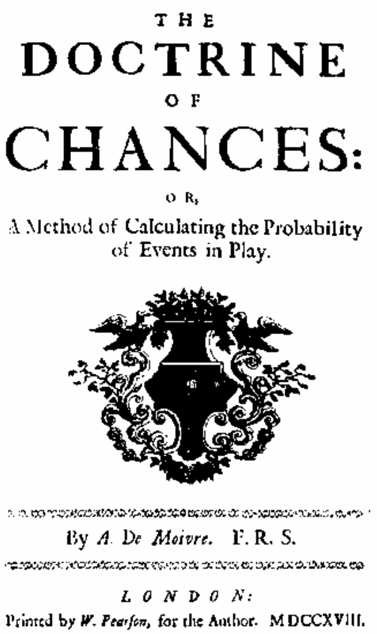


Estimation of carbon fluxes

- Uncertain observations
- Indirect observations
  - Quasi-linear observation operator
- Very large observation vector
  - Forthcoming satellite observations
- Very large state vector

From conditional probabilities…

- “The Probability of the happening of two Events dependent, is the product of the Probability of the happening of one of them, by the Probability which the other will have of happening, when the first is considered as having happened” (de Moivre 1718)
- $p(A \wedge B) = p(A)\, p(B|A)$


THE

DOCTRINE

OF

CHANCES:

OR,

A Method of Calculating the Probability of Events in Play.

By *A. De Moivre.* F. R. S.

LONDON:

Printed by *W. Pearſon*, for the Author. MDCCXVIII.

… to Bayes' theorem

- $p(A \land B) = p(B \land A)$
- $p(A)\, p(B|A) = p(B)\, p(A|B)$
- $p(A|B) = p(A)\, p(B|A) \,/\, p(B)$

… to Bayes' theorem

- $p(A \wedge B) = p(B \wedge A)$
- $p(A)\, p(B|A) = p(B)\, p(A|B)$
- $p(A|B) = p(A)\, p(B|A) \,/\, p(B)$

"If an event can be produced by a number *n* of different causes, then the probabilities of these causes given the event are to each other as the probabilities of the event given the causes, and the probability of the existence of each of these is equal to the probability of the event given that cause, divided by the sum of all of the probabilities of the event given each of these causes"

(Laplace 1774)

… to Bayes' theorem (cont')

- Monovariate discrete:

$$P(A|B) = P(A).\frac{P(B|A)}{P(B)}$$

- Monovariate continuous

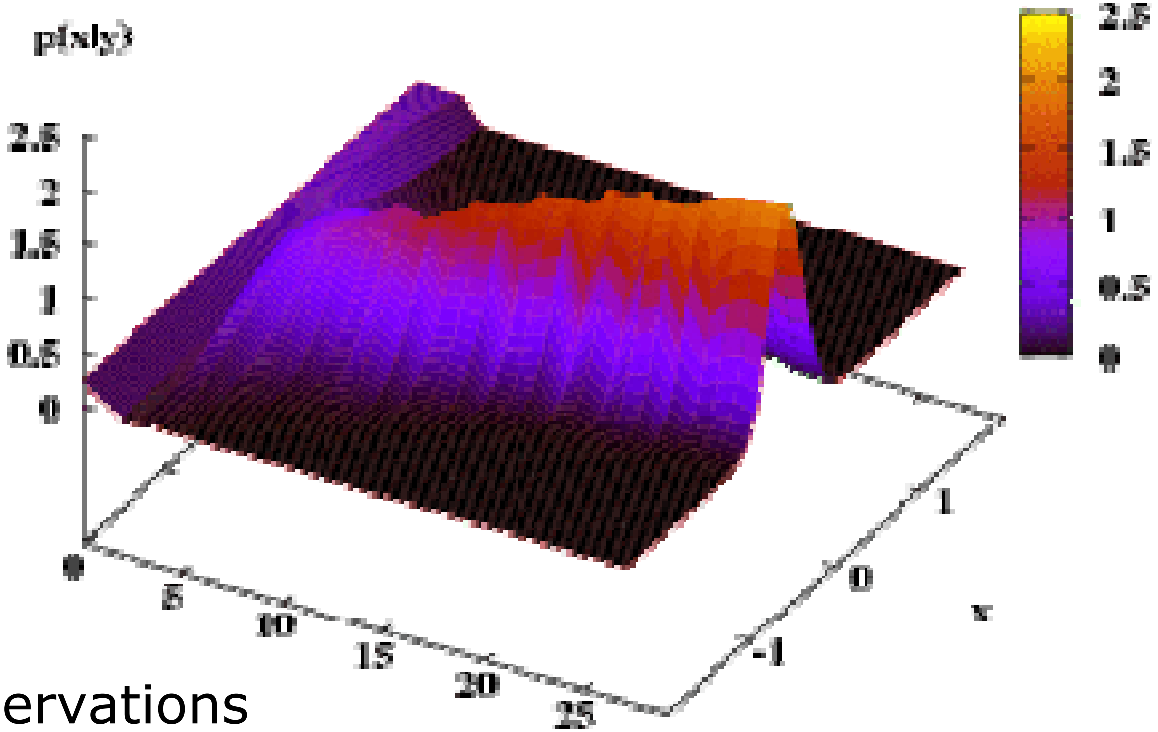
$$P(x|y) = P(x).\frac{P(y|x)}{P(y)}$$

- Multivariate continuous

$$P(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}).\frac{P(\mathbf{y}|\mathbf{x})}{P(\mathbf{y})}$$

Model of learning

$$P(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}).\frac{P(\mathbf{y}_1|\mathbf{x})}{P(\mathbf{y}_1)}.\frac{P(\mathbf{y}_2|\mathbf{x})}{P(\mathbf{y}_2)} \ \ldots$$



Ex:
monovariate x,
up to 25 observations

Number of observations

Some vocabulary

- Community-dependent
- **x**: state vector or control vector
- P(**x**): prior or background pdf
- P(**y**|**x**): observation pdf
- P(**x**|**y**): posterior pdf or analysis or inversion

$$P(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}).\frac{P(\mathbf{y}|\mathbf{x})}{P(\mathbf{y})}$$

Prior pdf

- The prior has the same weight than an observation

$$P(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}).\frac{P(\mathbf{y}|\mathbf{x})}{P(\mathbf{y})}$$

- “It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so” (Mark Twain)

- Accurate observations make an uncertain prior vanish
- pdf model

Prior pdf (cont')

- Uninformative prior?
  - Equiprobability (principle of insufficient reason or of indifference, Bernoulli, Laplace)
  - The unbounded flat pdf does not integrate to 1. (improper prior)

Example 1

- What is the pdf $p(x)$ of a uniform prior for $x \in [0,1]$?

Example 1

- What is the pdf $p(x)$ of a uniform prior for $x \in [0,1]$?
  - $\int_0^1 p(x)\, dx = 1.0$
  - $p(x) = k$
  - $\int_0^1 k\, dx = 1.0$
  - $k\, [x]_0^1 = 1.0$
  - $k = p(x) = 1.0$

Example 2

- What is the pdf $p(x)$ of a uniform prior for $y=x^2$, $x \in [0,1]$?

Example 2

- What is the pdf $p(x)$ of a uniform prior for $y=x^2$, $x \in [0,1]$?
  - $\int_0^1 p(y)\,dy = 1.0$
  - $p(y) = k = 1.0$
  - $a \in [0,1]$, $b \in [0,1]$
  - $\int_{x=a}^{x=b} p(x)\,dx = \int_{x=a}^{x=b} p(y)\,dy = \int_{x=a}^{x=b} 2x\,dx$
  - $p(x) = 2x$

Example 2

- What is the pdf $p(x)$ of a uniform prior for $y=x^2$, $x \in [0,1]$?
  - $\int_0^1 p(y)\, dy = 1.0$
  - $p(y) = k = 1.0$
  - $a \in [0,1]$, $b \in [0,1]$
  - $\int_{x=a}^{x=b} p(x)\, dx = \int_{x=a}^{x=b} p(y)\, dy = \int_{x=a}^{x=b} 2x\, dx$
  - $p(x) = 2x$

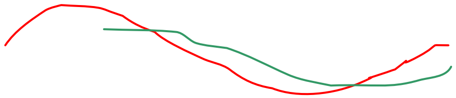
    - $\int_0^1 p(x)\, dx = [x^2]_0^1 = 1.0$
    - Note the apparent inconsistency between $p(x)$ and $p(y)$ for $x = 0$
    - Can we reconcile $p(x)$ and $p(x^2)$?

Example 2

- Can we reconcile $p(x)$ and $p(x^2)$?
  - $p(x) \sim p(x^n)$ for uniform $p(\ln(x))$

Prior pdf (cont')

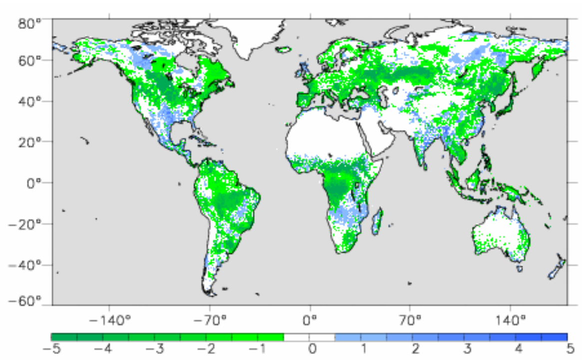
- The illusion of uninformative priors
- Invariance properties
  - Translation $p(x) = p(x+\delta x)$
- Maximum entropy
- pdf model parameters part of the estimation problem
- Evaluate prior pdf with accurate observations
  - Once and for all
  - Situation dependent?
- Use a proxy of the prior pdf
  - NMC method: use spread of forecasts valid at the same time but at different ranges

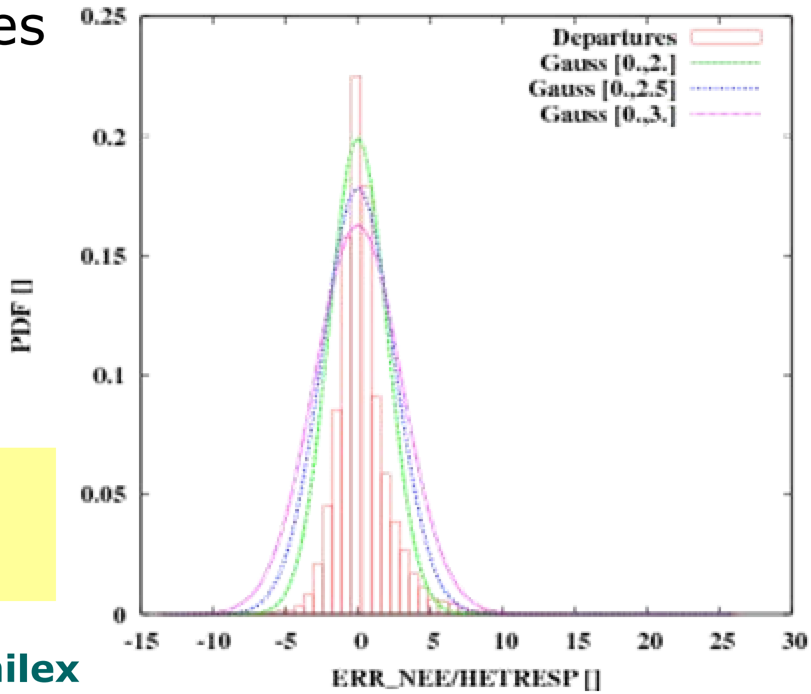


- Choice of state variables
  - z, z^2, $\log(z)$, …

Prior flux uncertainties

- ORCHIDEE model of the terrestrial biosphere: pdf from model-obs statistics
  - Statistics of ORCHIDEE errors wrt CarboEurope flux database
    - 200 sites, daily fluxes

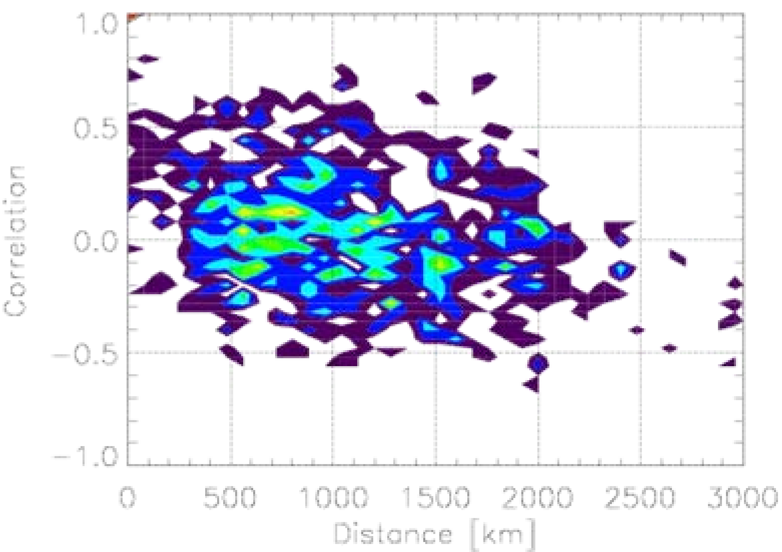


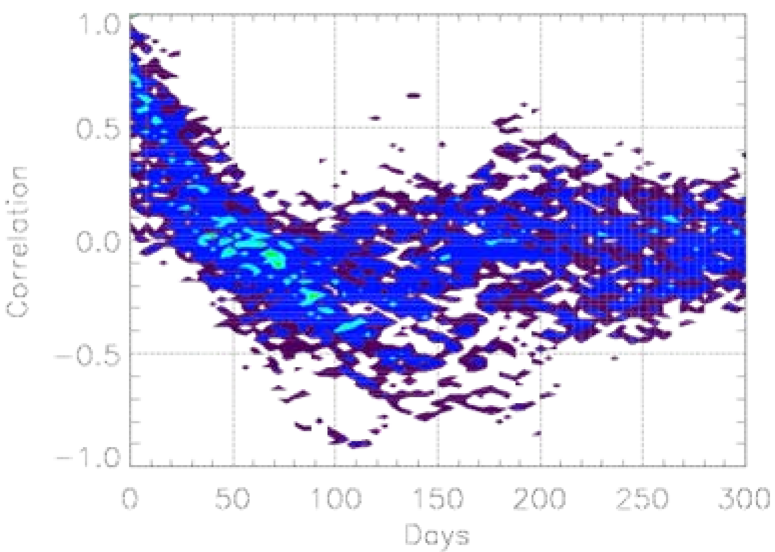


Distribution of the errors of simulated daily CO_2 fluxes ($gC/m^2/day$)

Prior flux uncertainties

- ORCHIDEE model of the terrestrial biosphere: pdf from model-obs statistics
  - Statistics of ORCHIDEE errors wrt CarboEurope flux database
    - 200 sites, daily fluxes





Computing the posterior pdf

$$P(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}).\frac{P(\mathbf{y}|\mathbf{x})}{P(\mathbf{y})}$$

$$P(\mathbf{y}) = \int P(\mathbf{x})P(\mathbf{y}|\mathbf{x})d\mathbf{x}$$

- Analytical solution
- Variational solution
- Monte Carlo solution

The linear problem with Gaussian statistics and zero biases

$$P(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}).\frac{P(\mathbf{y}|\mathbf{x})}{P(\mathbf{y})}$$

$n \in \mathbb{N}, \varphi \in \mathbb{R}^n, \mu \in \mathbb{R}^n, \Sigma \in M(n, \mathbb{R})$ positive definite

$$P(\varphi) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\varphi-\mu)^T \mathbf{\Sigma}^{-1}(\varphi-\mu)}$$

$$\mathbf{y} = \mathrm{H}\mathbf{x} + \varepsilon$$

$$-2\ lnP(\mathbf{x}|\mathbf{y}) = (\mathbf{x} - \mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + (\mathbf{H}\mathbf{x} - \mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y})$$

x: state vector
x_b: mean prior value of state vector
y: observation vector
H: linear observation operator
B: background error covariance matrix
R: observation error covariance matrix

Analytical solution

- Expected value of the posterior PDF

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

- Covariance of the posterior PDF

$$\mathbf{A} = \mathbf{B} - \mathbf{KHB}$$

- Gain matrix

$$\mathbf{K} = \mathbf{BH}^{\mathrm{T}}(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R})^{-1}$$

Implementation

- Inversion system:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

$$\mathbf{A} = \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B}$$

- **B**, **R**, $\mathbf{x}_b$, **y** and *H* given
- *H* : operator from state vector space to observation space
- Issues:
  - Compute **H**
    - Exact derivatives
    - Finite differences
    - From *H*
  - Matrix inversion

Non-linear observation operator

- In the tangent-linear hypothesis, the non-linear operators are linearized in the vicinity of some state of $\mathbf{x}$
  - $H[\mathbf{x}] \sim H[\mathbf{x}_b] + \mathbf{H}(\mathbf{x}-\mathbf{x}_b)$
  - Loss of optimality
  - Statistics less Gaussian
  - The degree of linearity is relative to $\mathbf{x}-\mathbf{x}_b$

Non-linear observation operator

- In the tangent-linear hypothesis, the non-linear operators are linearized in the vicinity of some state of **x**
  - $H[\mathbf{x}] \sim H[\mathbf{x}_b] + \mathbf{H}(\mathbf{x}-\mathbf{x}_b)$
- Possible inner loop/ outer loop system
  - $H[\mathbf{x}] \sim H[\mathbf{x}_a^i] + \mathbf{H}_i\,(\mathbf{x}-\mathbf{x}_a^i)$
  - Repeat the inversion keeping the same $\mathbf{x}_b$

$$\mathbf{K}_i = \mathbf{B}\mathbf{H}_i^T(\mathbf{H}_i\mathbf{B}\mathbf{H}_i^T + \mathbf{R})^{-1}$$

$$\mathbf{x}_a^i = \mathbf{x}_b + \mathbf{K}_i(\mathbf{y} - \mathbf{H}_i\mathbf{x}_b)$$

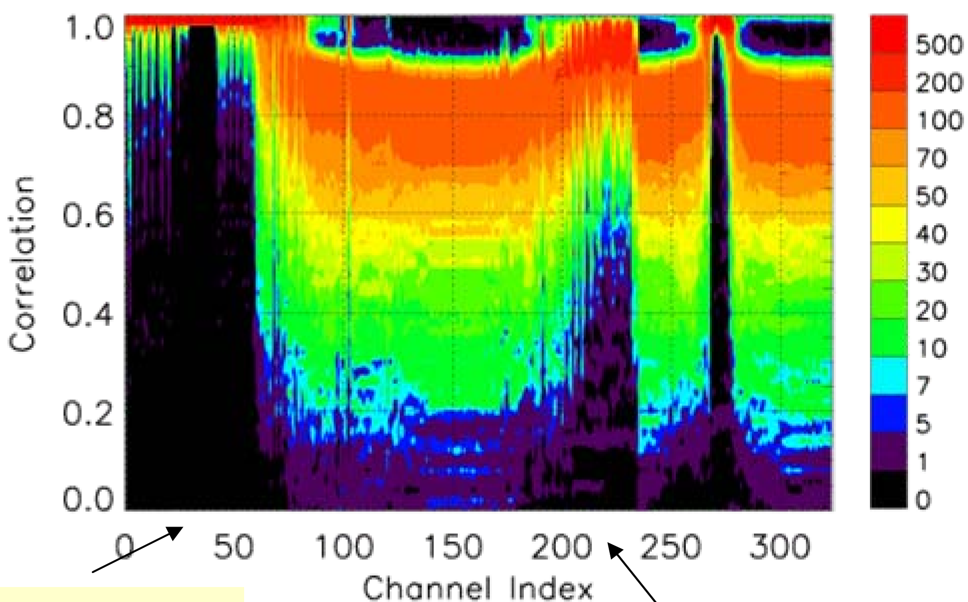
Observation screening

- Only use data with accurate linearized observation operator
  - Or possibility of weak constraint

4.5 micron

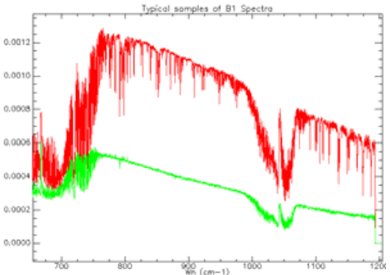
PDF of correlations between **H**.δx and *H*[x+ δx] – *H*[x]
Hemispheric data

x = T, q profile
H = simulate AIRS in the presence of clouds
δx = perturbation from **B**



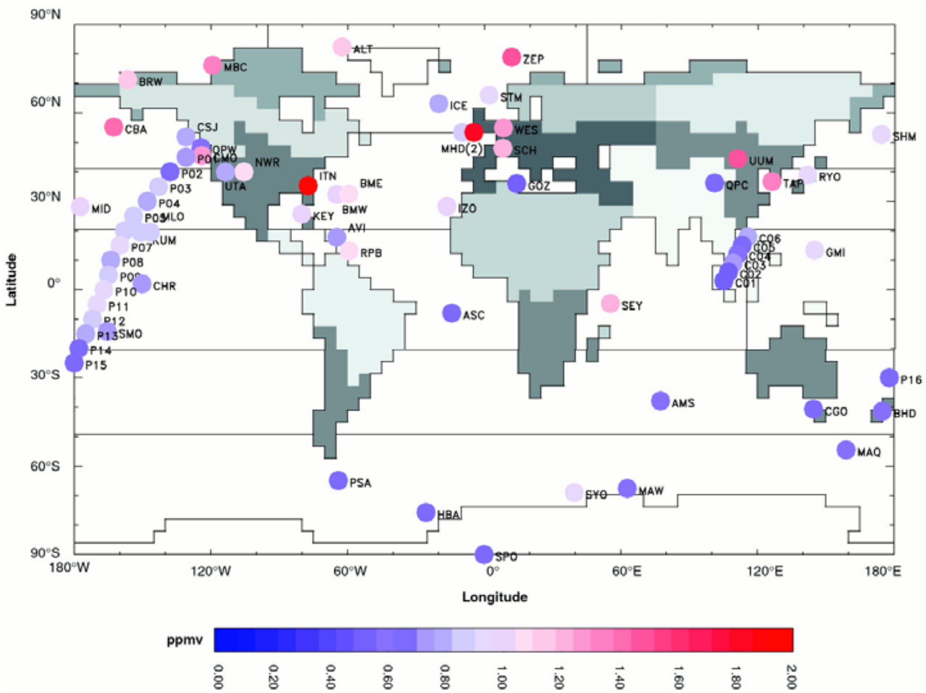
14.3 micron

6.3 micron

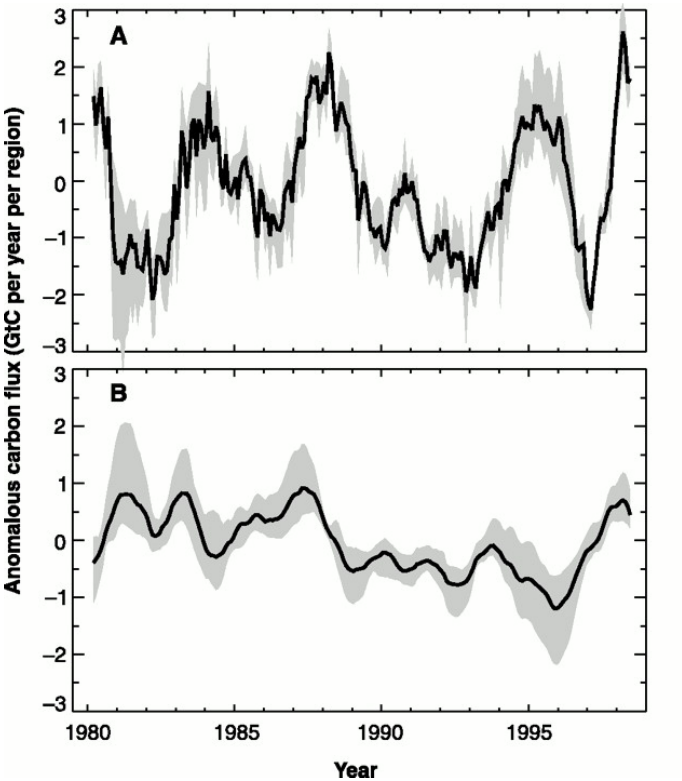


Assimilex

Analytical solution: CO_2



Year-to-year variability of the CO2 trend at the monitoring sites used in the inverse procedure over the period 1980-1998.



Inferred anomalous changes in the global land (**A**) and ocean (**B**) carbon fluxes

Bousquet et al., Science, 2000

Example

- $x_b = 15.0$, $\sigma_b = 1.0$
- $y = 15.5$, $\sigma_y = 0.5$
- $h = 1$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

$$\mathbf{A} = \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B}$$

Example

- $x_b = 15.0$, $\sigma_b = 1.0$
- $y = 15.5$, $\sigma_y = 0.5$
- $h = 1$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$
$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$
$$\mathbf{A} = \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B}$$

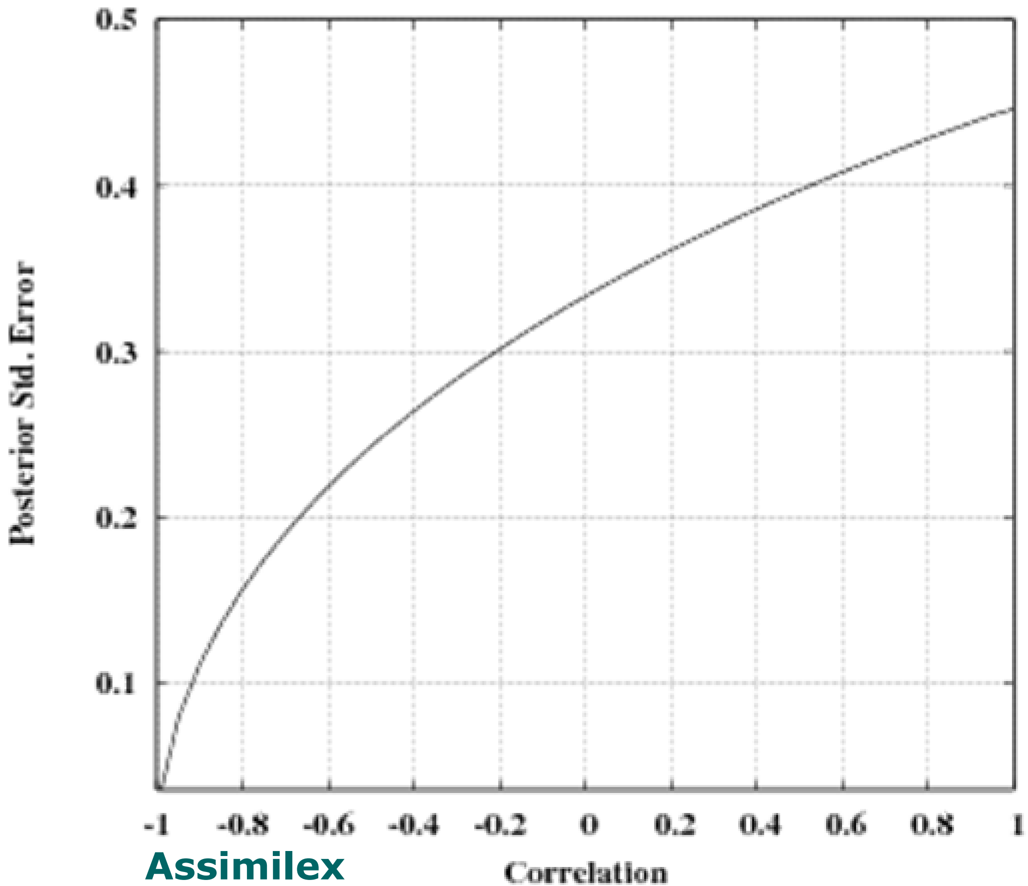
- $k = 1.0^2/(0.5^2+1.0^2) = 0.8$
- $x_a = 15.0 + 0.8\ (15.5 - 15.0) = 15.4$
- $\sigma_a = \sqrt{(1.0^2(1\text{-}k))} \approx 0.45$

Impact of error correlations (1/2)

- Correlated errors for observations of the same variable
- $\sigma_b = 1.0$
- $\sigma_{y1} = 0.5$
- $\sigma_{y2} = 0.5$
- $Cor(\varepsilon_{y1} , \varepsilon_{y2})$
- $h = [1,1]^T$

Impact of error correlations (1/2)

- Correlated errors for observations of the same variable
- $\sigma_b = 1.0$
- $\sigma_{y1} = 0.5$
- $\sigma_{y2} = 0.5$
- $Cor(\varepsilon_{y1}, \varepsilon_{y2})$
- $h = [1,1]^T$

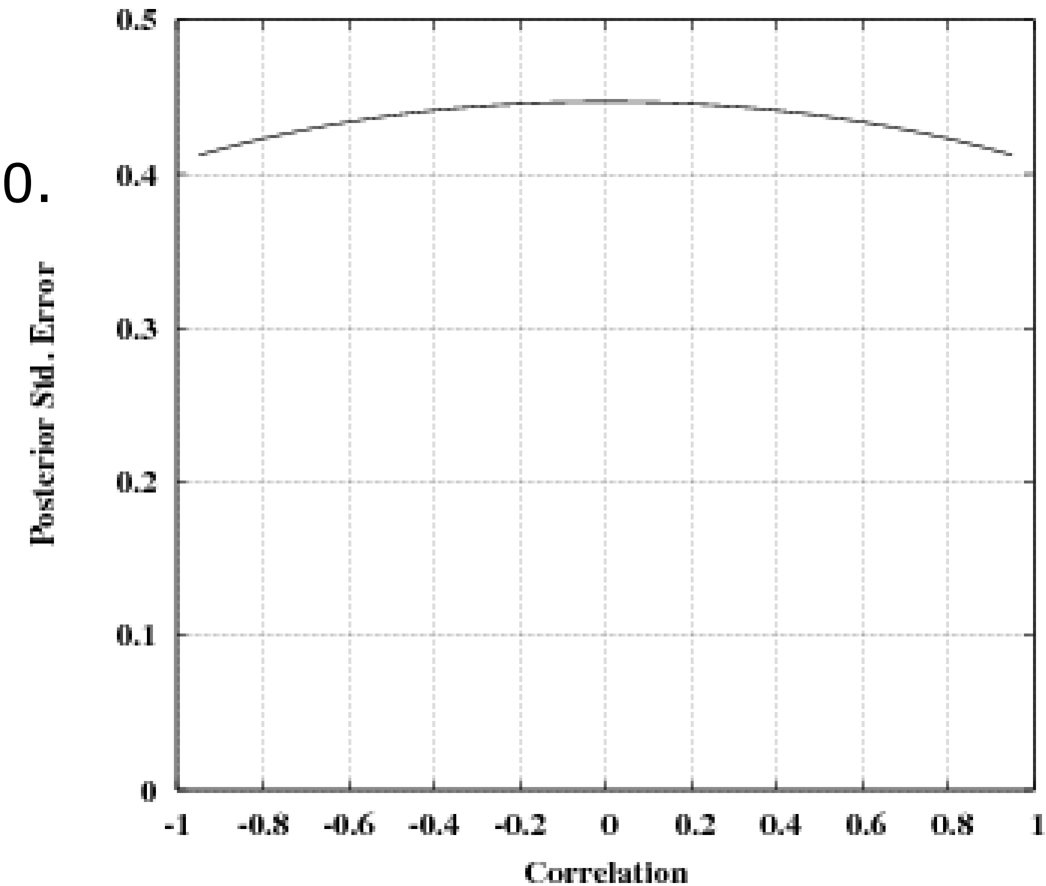


Impact of error correlations (2/2)

- Correlated errors for observations of a different variable
- $\sigma_{b1} = 1.0$
- $\sigma_{b2} = 1.0$
- $\mathrm{Cor}(\varepsilon_{b1}, \varepsilon_{b2}) = 0.$
- $\sigma_{y1} = 0.5$
- $\sigma_{y2} = 0.5$
- $\mathrm{Cor}(\varepsilon_{y1}, \varepsilon_{y2})$
- $\mathbf{H} = I_2$

Impact of error correlations (2/2)

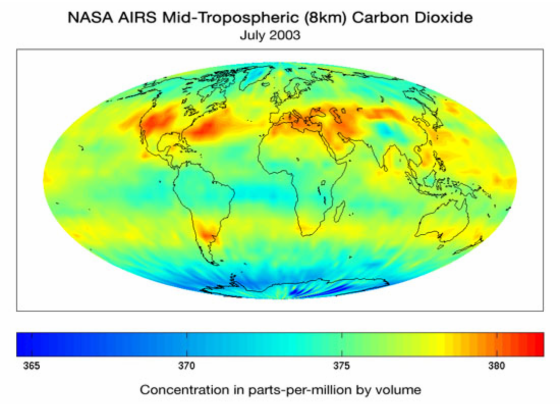
- Correlated errors for observations of a different variable
- $\sigma_{b1} = 1.0$
- $\sigma_{b2} = 1.0$
- $Cor(\varepsilon_{b1} , \varepsilon_{b2}) = 0.$
- $\sigma_{y1} = 0.5$
- $\sigma_{y2} = 0.5$
- $Cor(\varepsilon_{y1} , \varepsilon_{y2})$
- $\mathbf{H} = I_2$

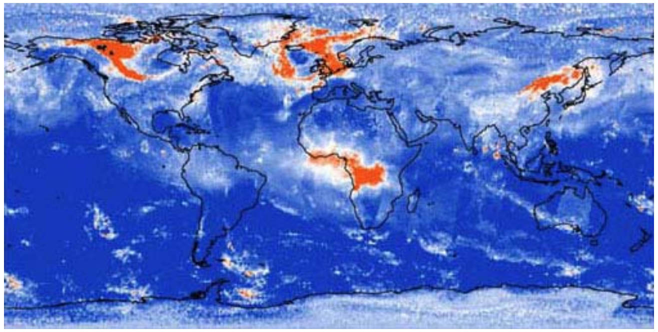


Impact of error correlations

- Possible applications
  - Multitracer inversion from satellite products
    - Dense data: case 1/2
    - Distinct variables: case 2/2

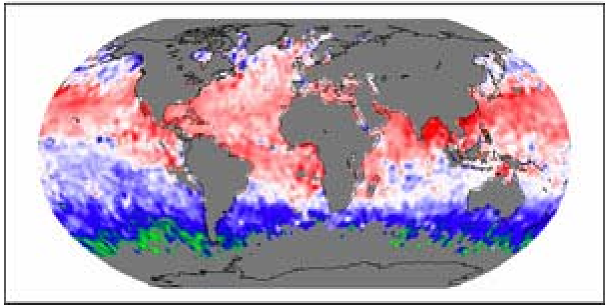
AIRS CO_2





AIRS CO

AIRS CH_4



Implementation

- Inversion system:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

$$\mathbf{A} = \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B}$$

- Issues:
  - Compute **H**
  - Matrix inversion

Computing the posterior pdf

$$P(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}).\frac{P(\mathbf{y}|\mathbf{x})}{P(\mathbf{y})}$$

$$P(\mathbf{y}) = \int P(\mathbf{x})P(\mathbf{y}|\mathbf{x})d\mathbf{x}$$

- Analytical solution
- Variational solution
- Monte Carlo solution

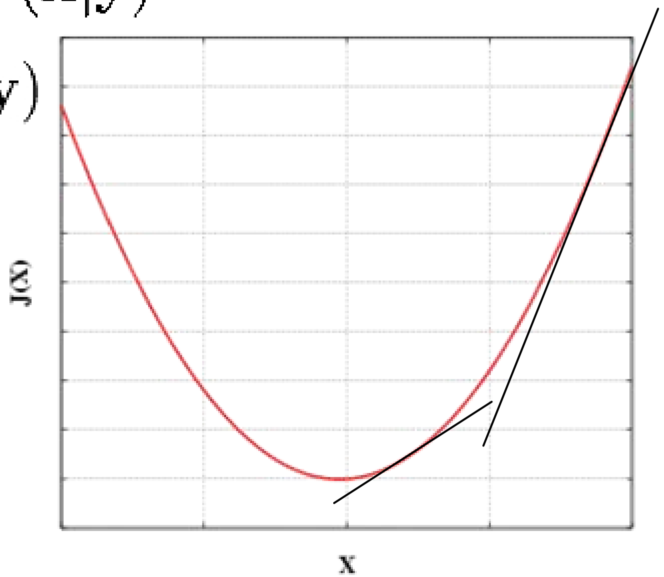
Variational solution

- The linear problem with Gaussian statistics and zero biases

$$-2\,\ln P(\mathbf{x}|\mathbf{y}) = (\mathbf{x}-\mathbf{x}_b)^{\mathbf{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{H}\mathbf{x}-\mathbf{y})^{\mathbf{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}-\mathbf{y})$$

  - X_a minimises $J(\mathbf{x}) = -2\,\ln P(\mathbf{x}|\mathbf{y})$

$$\nabla J(\mathbf{x}) = 2\mathrm{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + 2\mathrm{H}^T\mathrm{R}^{-1}(\mathrm{H}\mathbf{x}-\mathbf{y})$$



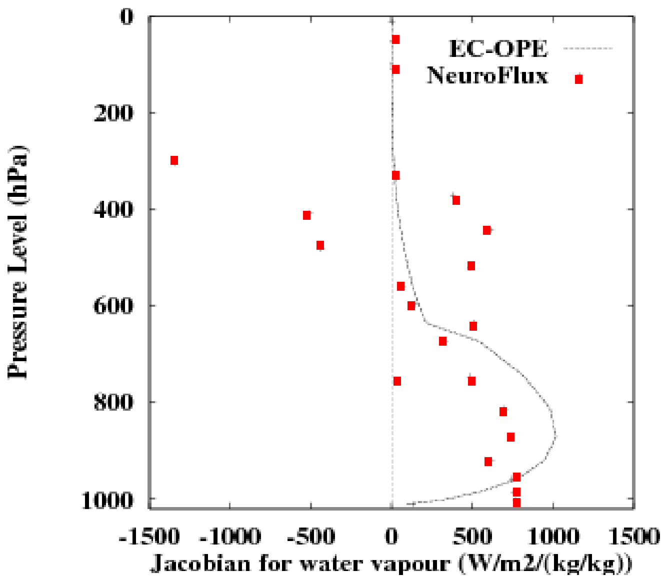
  - Covariance of the posterior

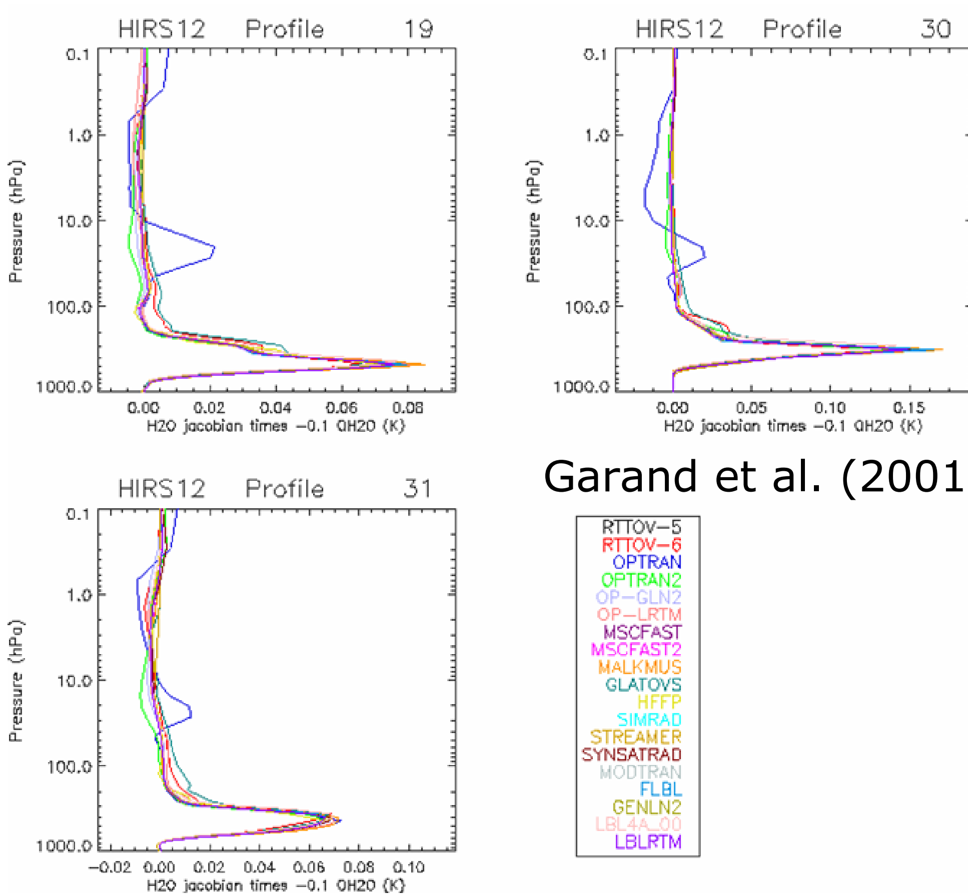
$$\mathbf{A} = 2[J''(\mathbf{x}_a)]^{-1}$$

Accuracy of Jacobians

- New types of model validation
  - Tough requirement for statistics-based models

Chevallier and Mahfouf (2001)



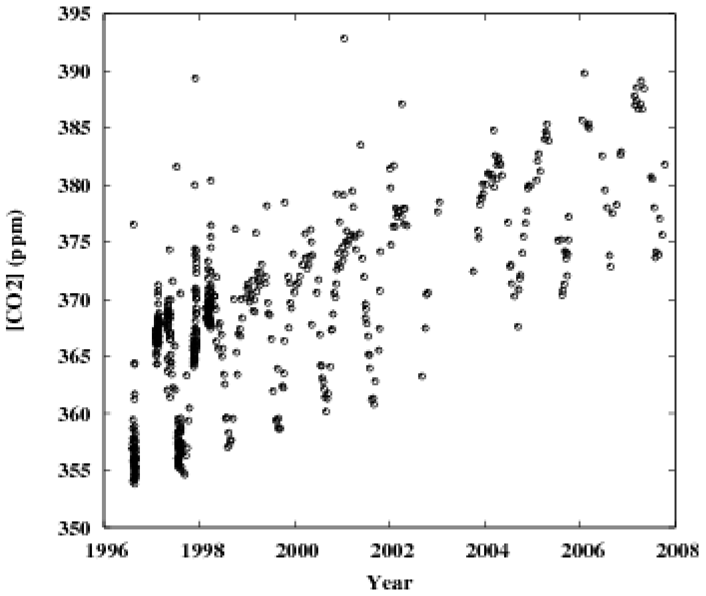


Garand et al. (2001)

Outliers

$$J(\mathbf{x}) = (\mathbf{x}-\mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{H}\mathbf{x}-\mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}-\mathbf{y})$$

- Outliers may drive the quadratic function
  - Preliminary screening
  - Ex: remove observations for which $|Hx_b - y| > 3\ \sigma_o$ …



Implementation

- Inversion system:

$$J(\mathbf{x}) = (\mathbf{x}-\mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{H}\mathbf{x}-\mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}-\mathbf{y})$$

$$\nabla J(\mathbf{x}) = 2\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + 2\mathbf{H}^{T}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}-\mathbf{y})$$

$$\mathbf{A} = 2[J''(\mathbf{x}_a)]^{-1}$$

- **B**, **R**, $\mathbf{x}_b$, **y** and *H* given
- Issues:
  - Compute **H** and $\mathbf{H}^T$
  - Matrix inversions
  - Minimisation method (grad($J(\mathbf{x}_a)$) ~ 0)
  - Compute J''

Compute **H** and **H**^{T}

$$\nabla J(\mathbf{x}) = 2\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x_b}) + 2\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}])$$

- **H**: Tangent-linear (Jacobian) matrix **H**
- **H**^{T}: Adjoint matrix of **H**
- Chain rule:
  - $\mathbf{Hx} = \mathbf{H}_n\ \mathbf{H}_{n-1} \ldots \mathbf{H}_2\ \mathbf{H}_1\mathbf{x}$
  - $\mathbf{H^Ty^*} = \mathbf{H}_!^\mathbf{T}\ \mathbf{H}_2^\mathbf{T} \ldots \mathbf{H}_{n-1}^\mathbf{T}\ \mathbf{H}_n^\mathbf{T}\mathbf{y^*}$
  - First order Taylor development of each individual line of code
- Example:
  - Compute the tangent-linear and adjoint operators of the following lines of code:
    - $a = b^2$
    - $a = a^2$

Adjoint technique

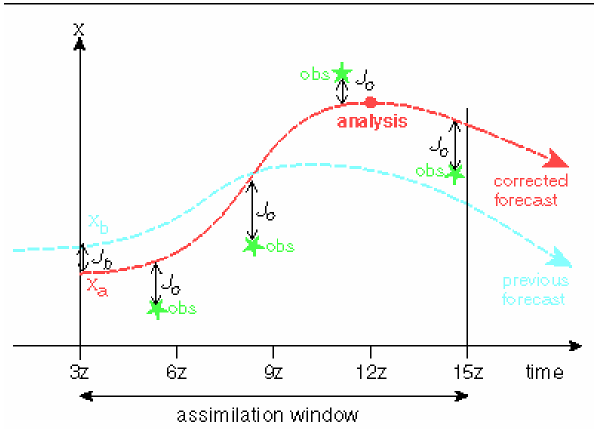
- Example:
  - Compute the adjoint instruction of the line:
    - $a = b^2$
- Forward statement
    - $a = b^2$
- Tangent-linear statement
    - $\delta b = 0.\delta a + 1.\delta b$
    - $\delta a = 0.\delta a + 2b.\delta b$
- Adjoint statement
    - $b^* = 2b a^* + 1.b^*$
    - $a^* = 0.a^* + 0.b^*$

Adjoint technique

- Example:
  - Compute the adjoint instruction of the line:
    - $a = a^2$
- Forward statement
    - $a = a^2$
- Tangent-linear statement
    - $\delta a = 2a.\delta a$
- Adjoint statement
    - $a^* = 2a.a^*$

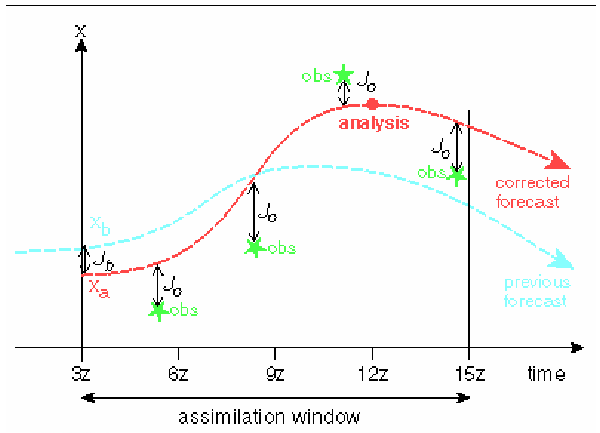
Handling the linearization points

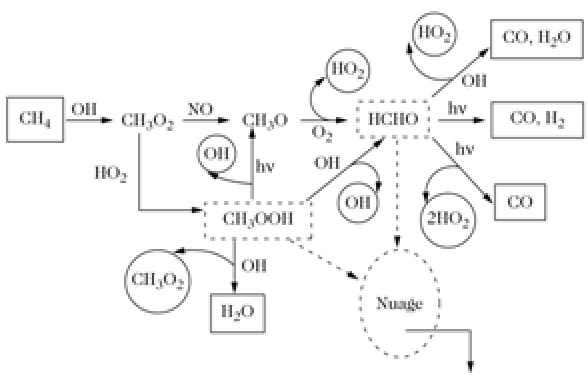
- Handling of trajectory
  - $\mathbf{Hx} = \mathbf{H}_n \, \mathbf{H}_{n-1} \ldots \mathbf{H}_2 \, \mathbf{H}_1\mathbf{x}$ (forward)
  - $\mathbf{H}^T\mathbf{y}^* = \mathbf{H}_!^T \, \mathbf{H}_2^T \ldots \mathbf{H}_{n-1}^T \, \mathbf{H}_n^T\mathbf{y}^*$ (backward)
  - Linearization points for the adjoint
    - Stored in computer memory
    - Stored on disk
    - Recomputed on the fly
    - Some mixture of the above

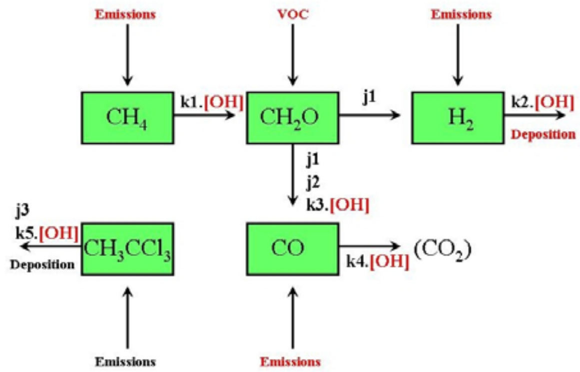


Which code?

- Adjoint of full code or of simplified version?
  - Time handling
    - $\mathbf{H}_t(x) \sim \mathbf{H}(x)$
  - Spatial resolution
    - $\mathbf{H}_{HR}(x) \sim \mathbf{H}_{LR}(x)$
  - Sophistication of physics







Adjoint coding

- Manual coding
- Automatic differentiation
  - 29 softwares listed in http://www.autodiff.org
  - Source code transformation
    - From the original code
    - From a recoded version
  - Operator overloading
  - Freeware or not

- Correctness of the TL
  - Linearity
  - Convergence of the Taylor development
- Correctness of the AD
  - Linearity
  - $(\mathbf{Hx})^T\mathbf{Hx} = \mathbf{x}^T\mathbf{H}^T(\mathbf{Hx})$
- … to the accuracy of the computer

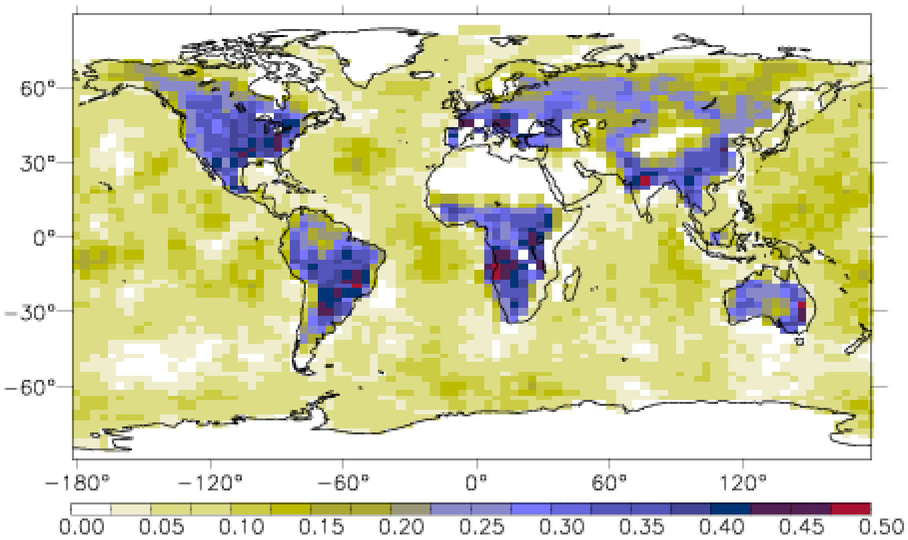
Invert **R** matrix

$$\nabla J(\mathbf{x}) = 2\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x_b}) + 2\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}])$$

- $\mathbf{R}^{-1}$:
  - Try to have it diagonal
    - Ignore correlations
    - Data thinning
    - Increase variances and set correlations to zero
  - Example:
    - Invert CO_2 fluxes from forthcoming OCO satellite observations
    - Hypothesised correlations of 0.5 from one observation to the next

Reference case

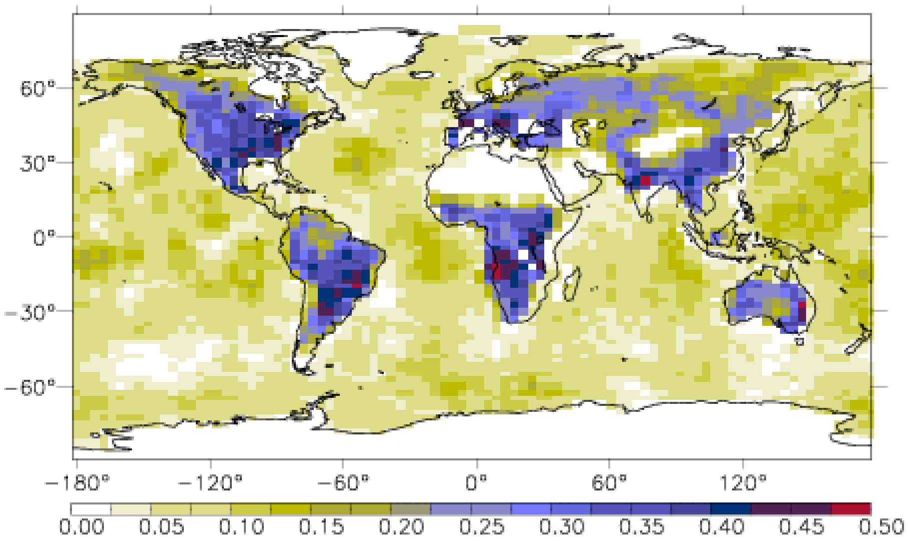
- No correlations
- error reduction 1-sig(post)/sig(prior)



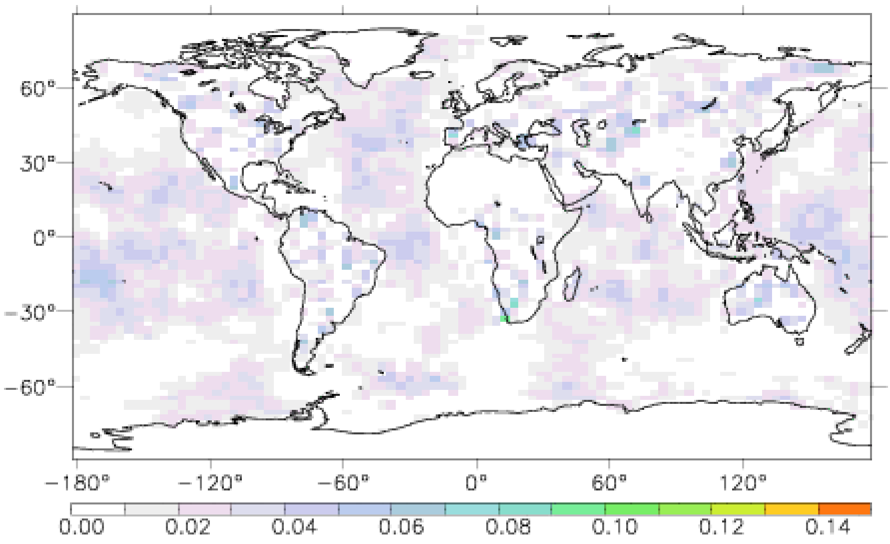
Reference error reduction

Test case 1

- Correlations properly accounted for
- error reduction 1-sig(post)/sig(prior)



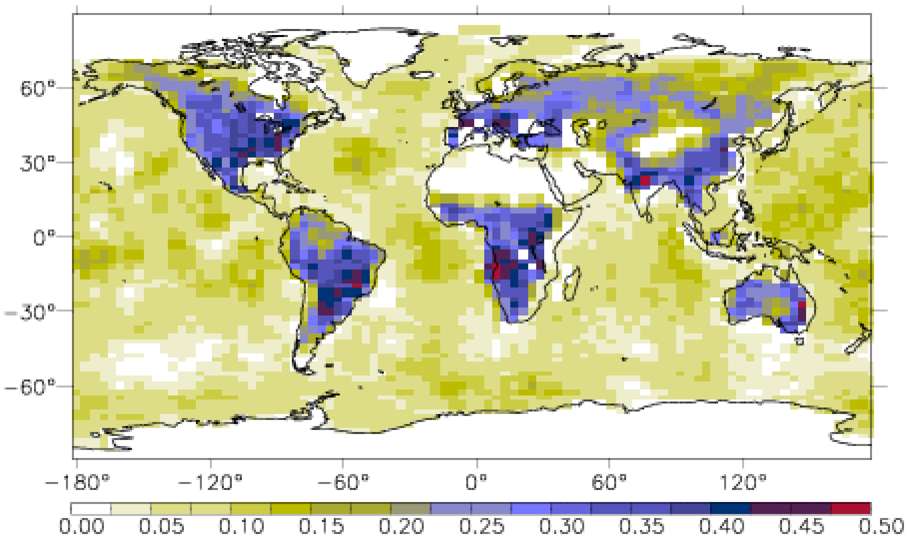
Reference error reduction



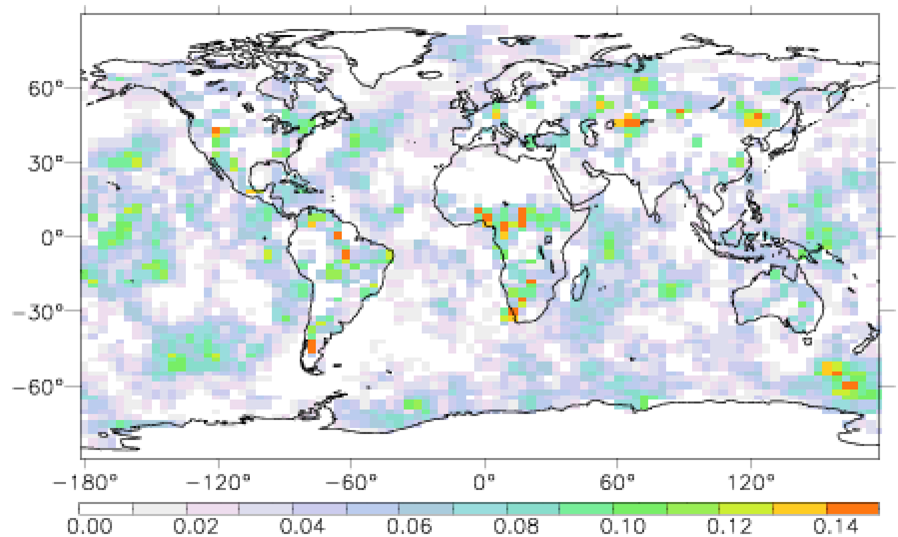
Change in error reduction

Test case 2

- Correlations simply ignored in the inversion
- error reduction 1-sig(post)/sig(prior)



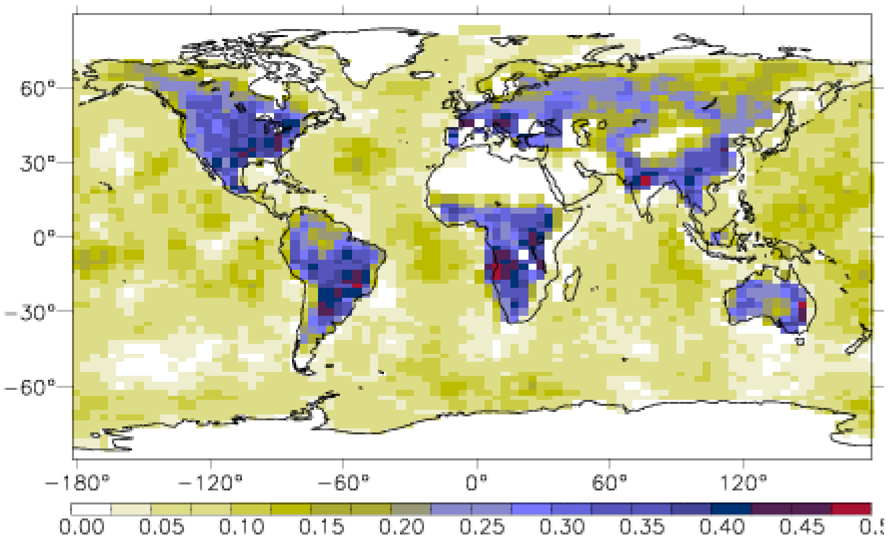
Reference error reduction



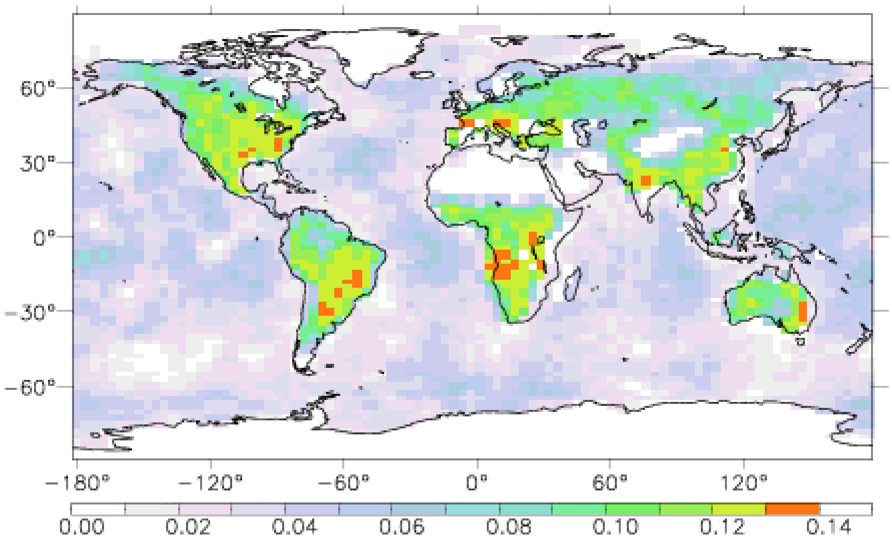
Change in error reduction

Test case 3

- Data thinning (remove one obs. every two)
- error reduction 1-sig(post)/sig(prior)



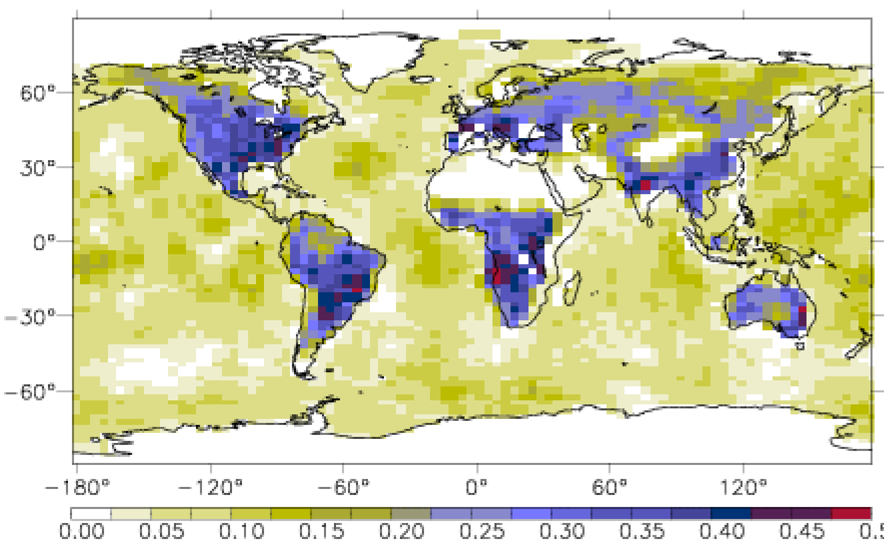
Reference error reduction



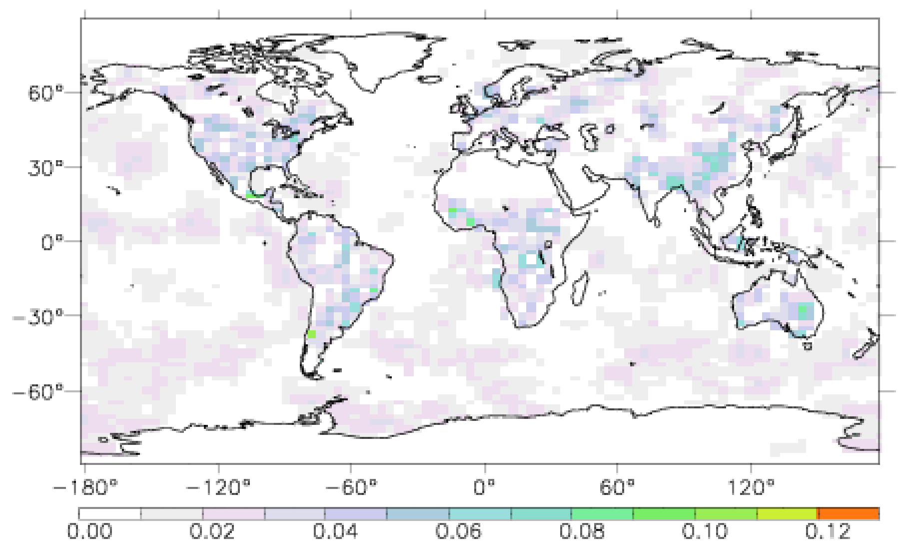
Change in error reduction

Test case 4

- Obs. variances multiplied by 2, no correlations
- error reduction 1-sig(post)/sig(prior)



Reference error reduction



Change in error reduction

Invert **B** matrix

$$\nabla J(\mathbf{x}) = 2\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x_b}) + 2\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}])$$

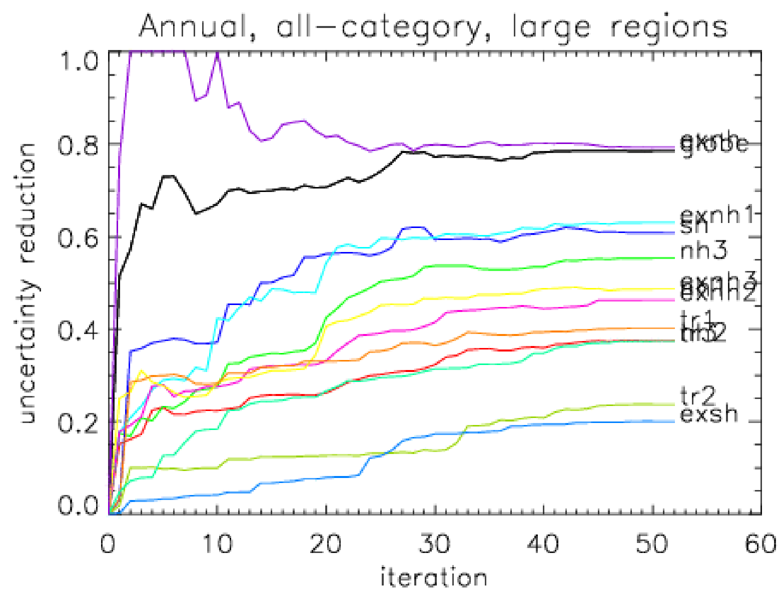
- $\mathbf{B}^{-1}$:
  - **B** sparse
  - Ex:
    - $\mathbf{B} = \mathbf{S}^T\mathbf{C}\mathbf{S}$ with **S** vector of standard deviations, **C** eigenvalue-decomposed $\mathbf{C} = \mathbf{V}^T\mathbf{v}\mathbf{V}$
    - **C** block-diagonal, or product of block-diagonal matrices
    - $\mathbf{B}^{-1} = \mathbf{S}\,\mathbf{V}\mathbf{v}^{-1}\mathbf{V}^T\,\mathbf{S}^T$

Minimisation algorithm

- Many optimization methods available
- Conjugate gradient
  - Lanczos algorithm provides the leading eigenvalues of the Hessian of the cost function as by-product of the minimisation

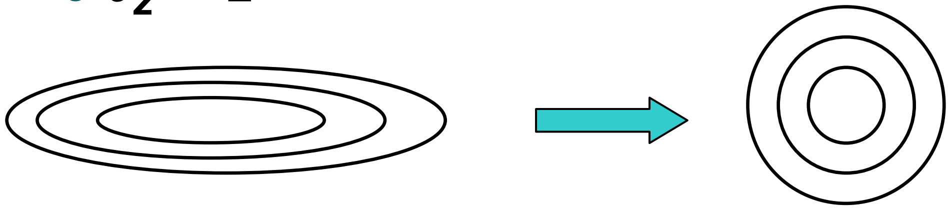
$$\mathbf{A} = 2[J''(\mathbf{x}_a)]^{-1}$$

CH_4 flux inversion
Error reduction
for a series of large regions
Meirink et al. (2008)



Conditioning

- Many optimization methods available
- More efficient with preconditioning
  - State vector ≠ physical vector
  - $\mathbf{z} = \mathbf{A}^{-1/2}(\mathbf{x}-\mathbf{x}_b)$ reduces the minimisation to one iteration with conjugate gradients
    - $J_{\mathbf{z}}'' \sim \mathbf{I}$



  - $\mathbf{z} = \mathbf{B}^{-1/2}(\mathbf{x}-\mathbf{x}_b)$ is a simple approximation
    - J unchanged
    - $\mathrm{grad}_z(J) = \mathbf{B}^{+1/2}\,\mathrm{grad}_{\mathbf{x}}(J)$

Implementation

- Inversion system:

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{H}\mathbf{x} - \mathbf{y})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y})$$

$$\nabla J(\mathbf{x}) = 2\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + 2\mathbf{H}^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y})$$

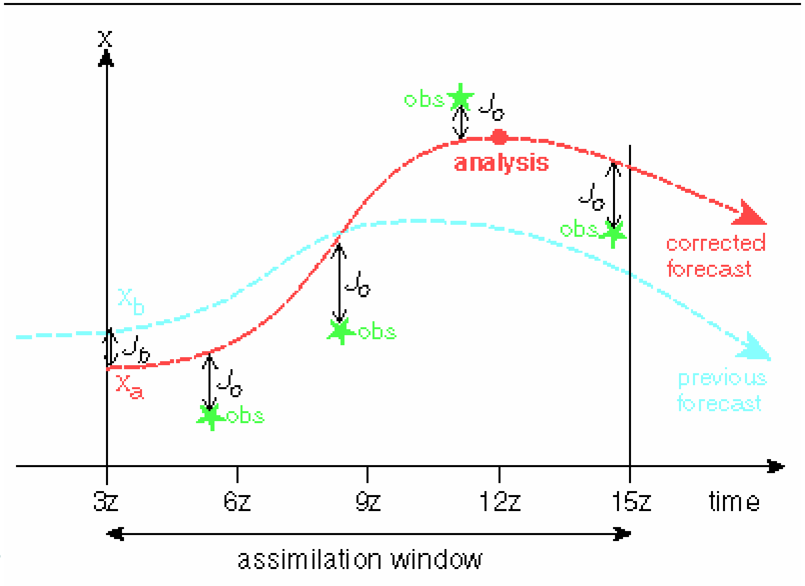
$$\mathbf{A} = 2[J''(\mathbf{x}_a)]^{-1}$$

- **B**, **R**, $\mathbf{x}_b$, **y** and *H* given
- Issues:
  - Compute **H** and $\mathbf{H}^T$
  - Matrix inversions
  - Minimisation method (grad($J(\mathbf{x}_a)$) ~ 0)
  - Compute **A**

NWP 4D-Var

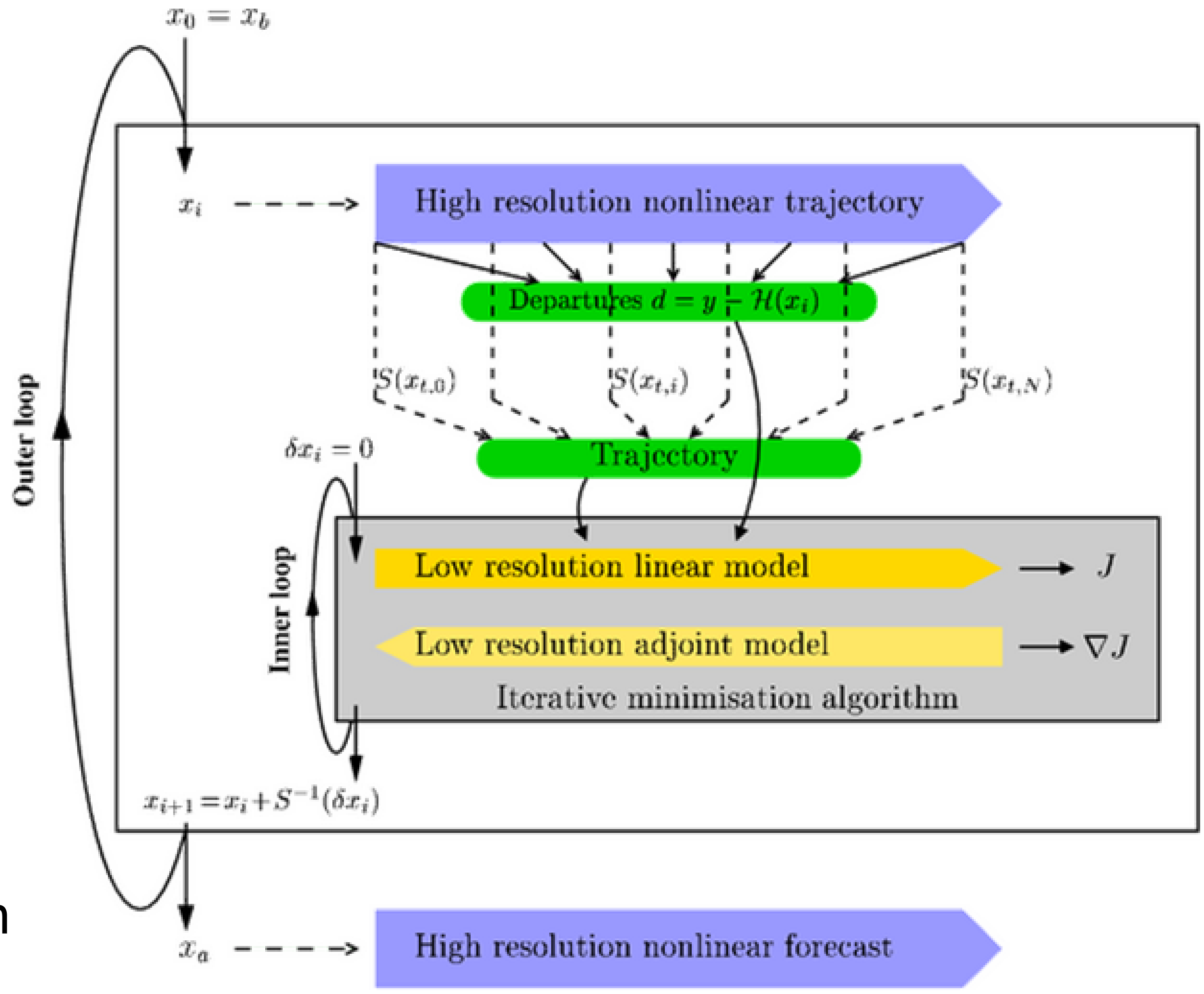
$$J(\mathbf{x}) = (\mathbf{x}-\mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{H}\mathbf{x}-\mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}-\mathbf{y})$$

x: atmospheric state at initial time step
H: linearized observation operator
inc. time evolution of x



The ECMWF 12h 4D-Var system (1997, 2000)

Incremental 4D-Var



ECMWF system

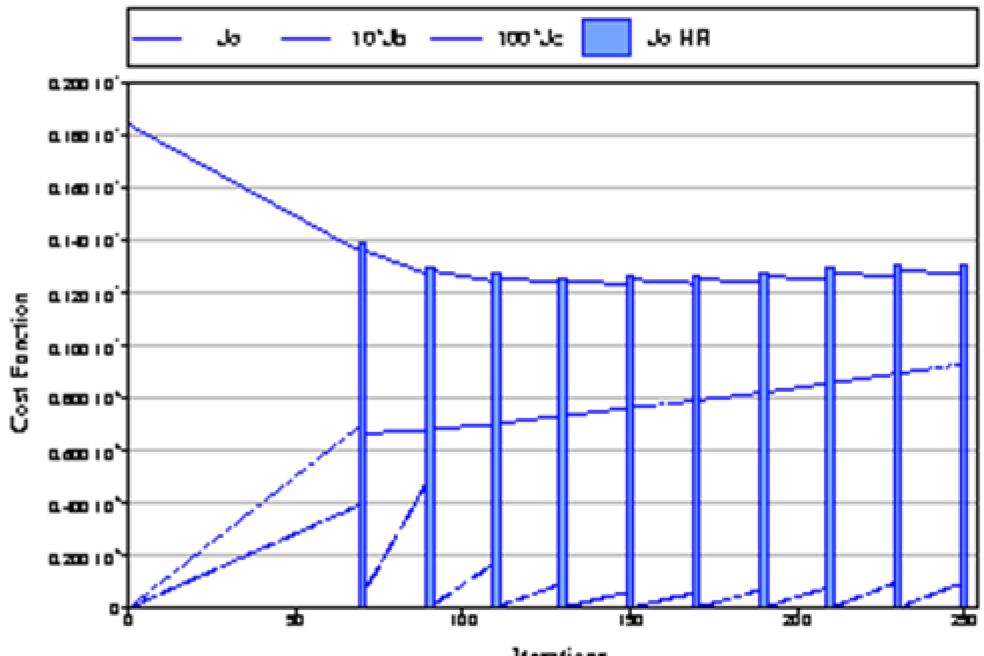
Inner loop/ outer loop system

$$\nabla J(\mathbf{x}) = 2\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x_b}) + 2\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}])$$

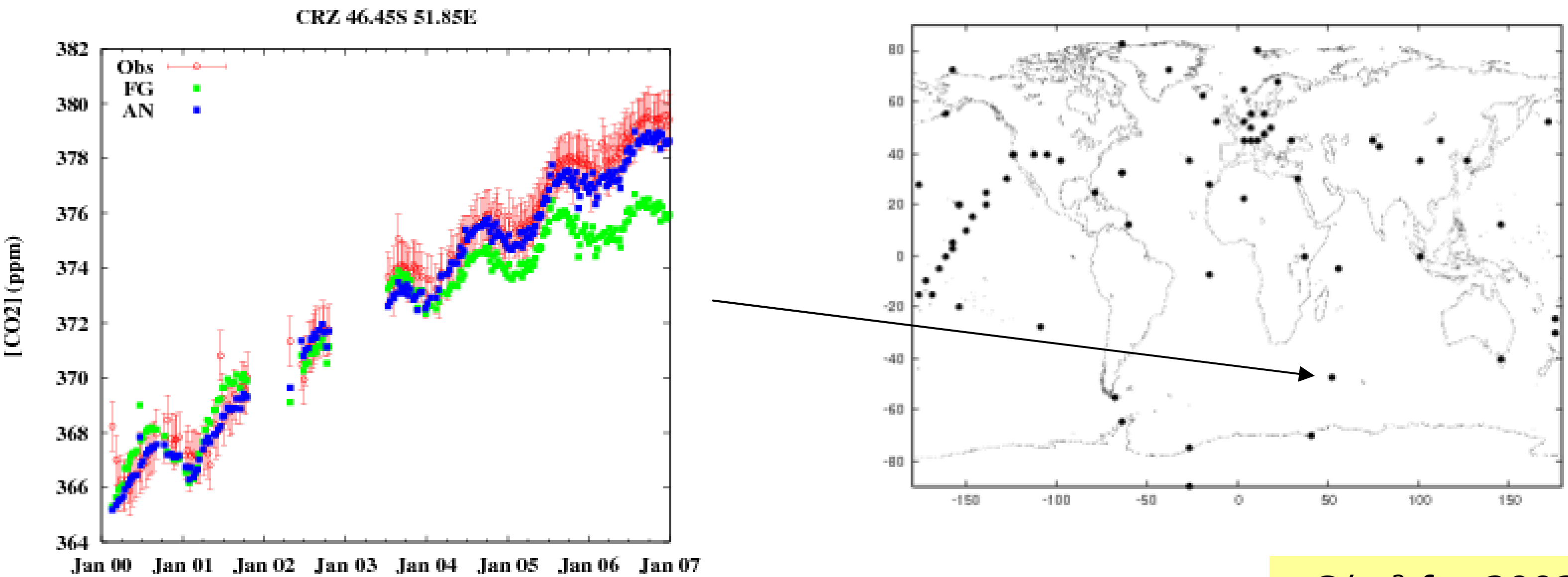
- Inner loop strictly linear (TL)
- Non-linear updates in outer loop

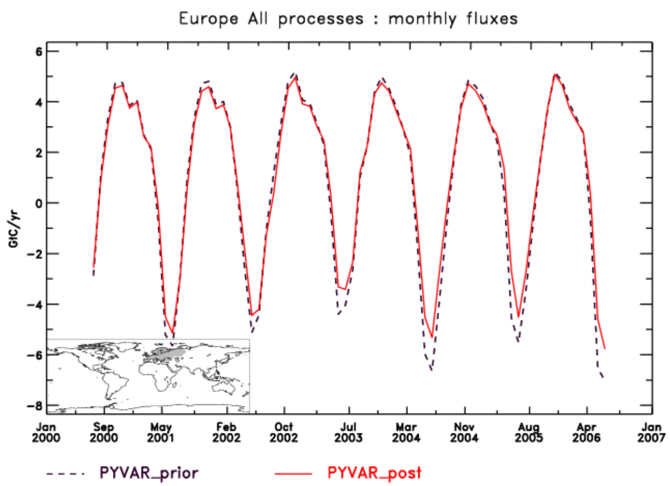
J = Jo + Jb + Jc

Trémolet (2005)

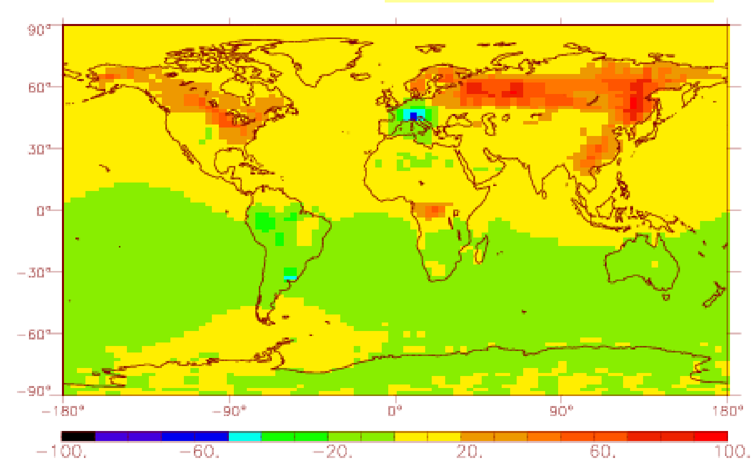


Variational inversion of CO_2 fluxes

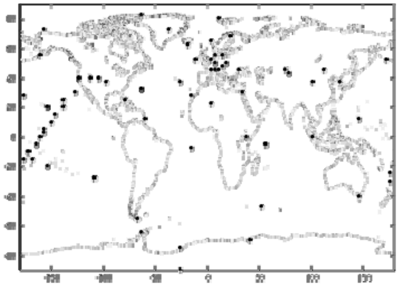




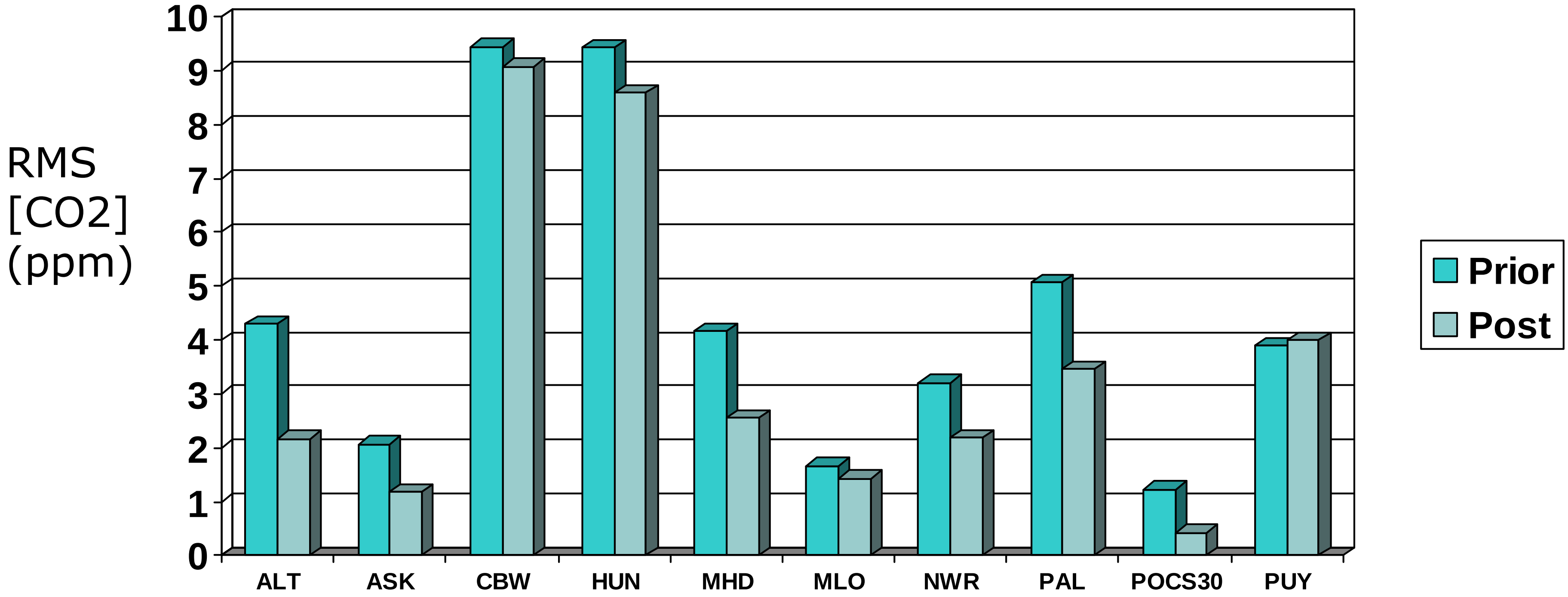
gC/m² for 2003



Fit to dependent data



Example: 10 out of 69 stations



Computing the posterior pdf

$$P(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}).\frac{P(\mathbf{y}|\mathbf{x})}{P(\mathbf{y})}$$

$$P(\mathbf{y}) = \int P(\mathbf{x})P(\mathbf{y}|\mathbf{x})d\mathbf{x}$$

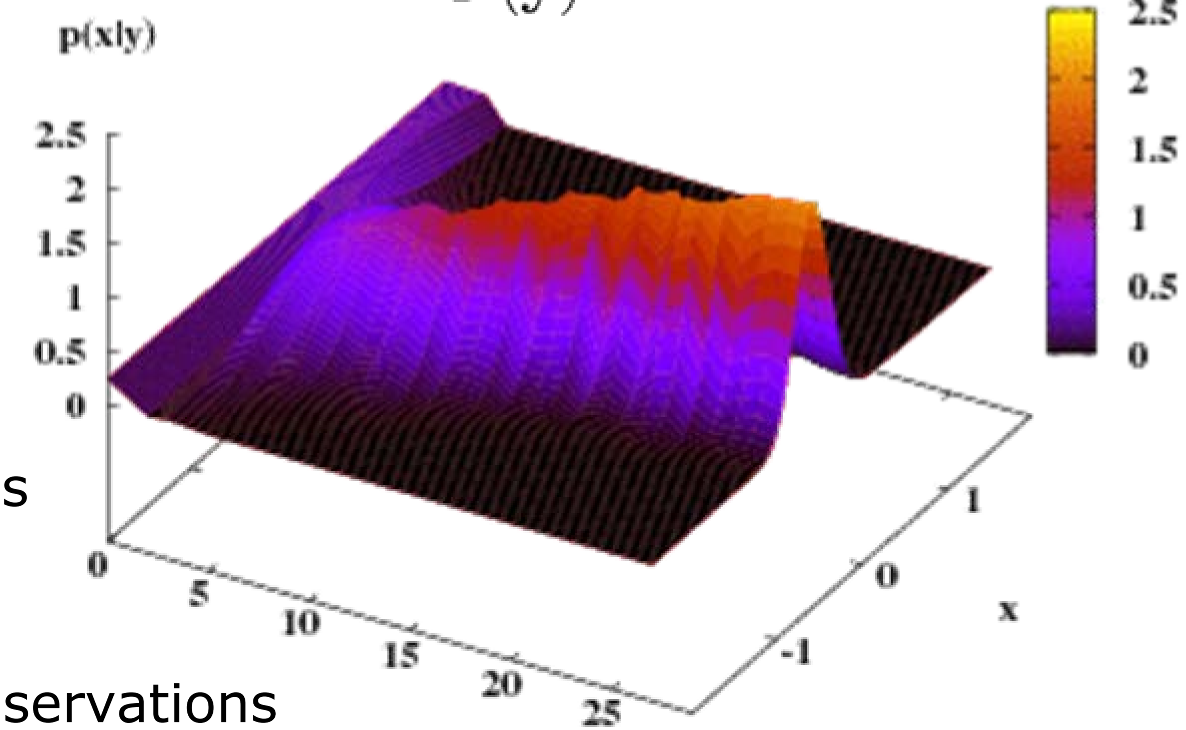
- Analytical solution
- Variational solution
- Monte Carlo solution

Particle filter

- Apply Bayes' formula to a discrete ensemble of **x**'s

$$P(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}).\frac{P(\mathbf{y}|\mathbf{x})}{P(\mathbf{y})}$$

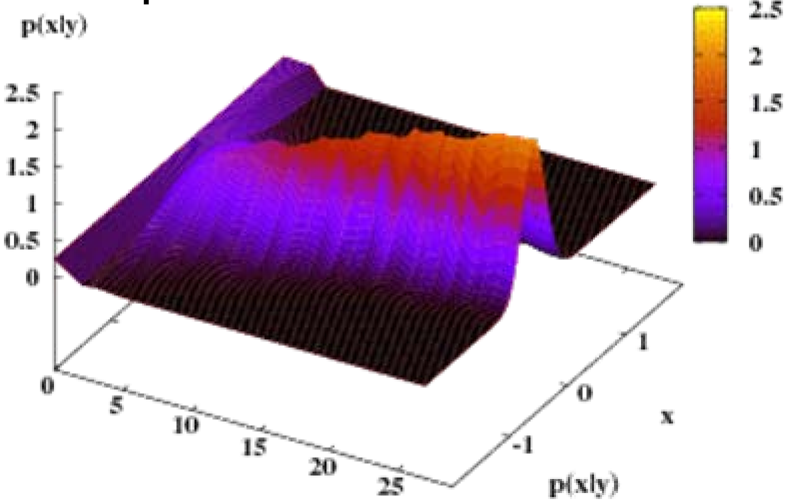
Ex: 100 points
monovariate x,
Gaussian pdfs,
up to 25 observations



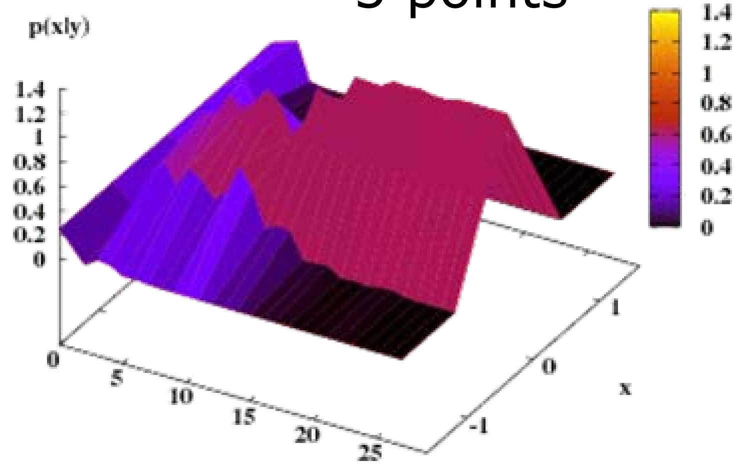
Number of observations

Particle filter

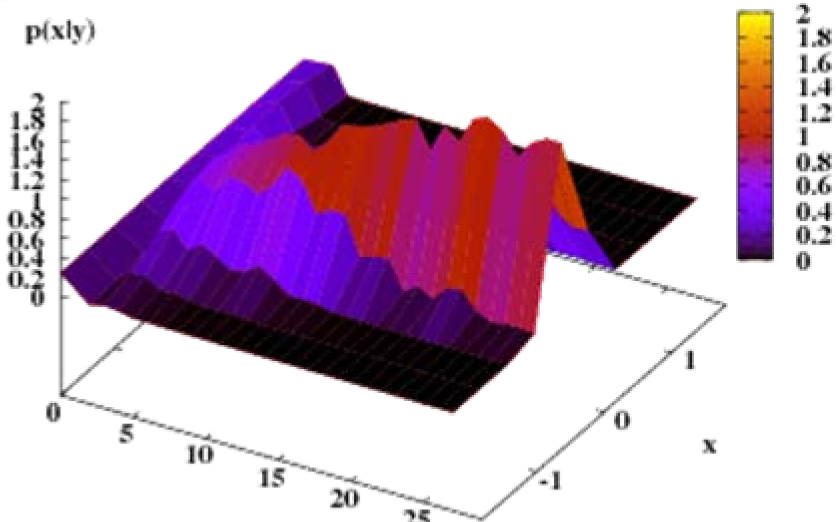
100 points



5 points



10 points



Particle filter

- Curse of dimensionality
  - Sampling high-dimensional spaces
  - Exponential increase of ensemble size to maintain a given sampling accuracy
  - + Numerical issues

$$P(\mathbf{y}) = \int P(\mathbf{x})P(\mathbf{y}|\mathbf{x})d\mathbf{x}$$

- Localization

Ensemble methods

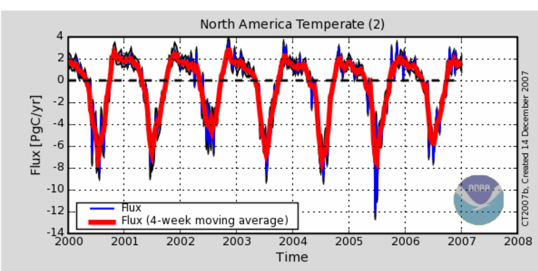
- No limitation wrt linearity or pdf model
- No adjoint model
- Parallel hardwares

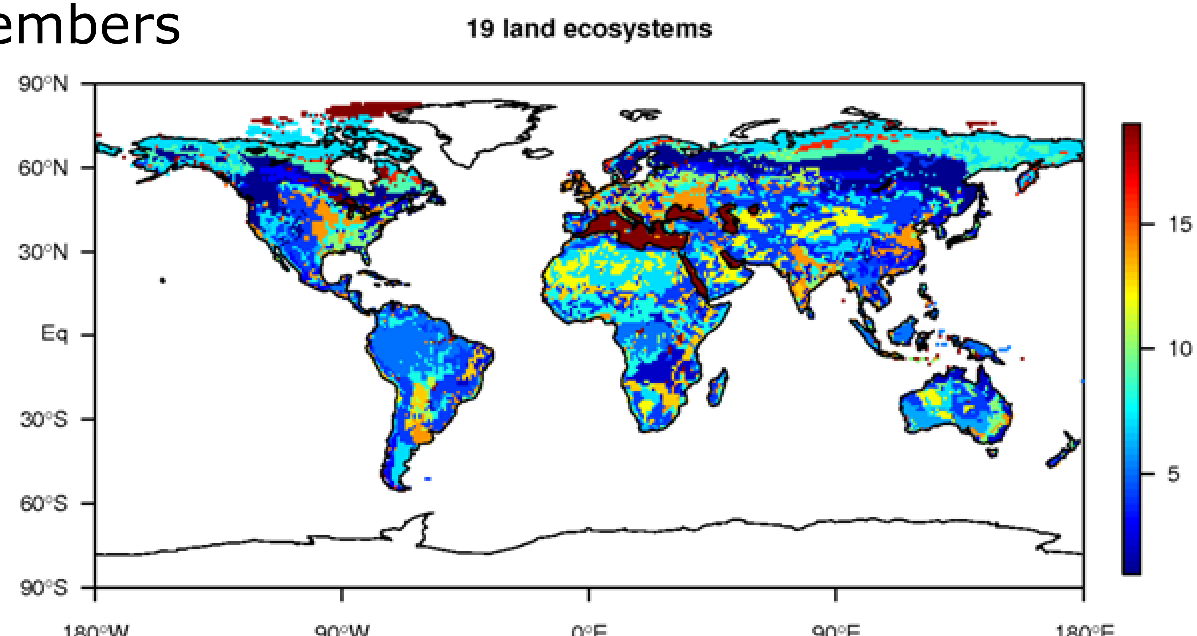
Ensemble methods

- Ensemble Kalman filter (Evensen 1994)
  - Ensemble forecast of error statistics
  - Full-rank analytical analysis
- Ensemble square root filter (Whitaker and Hamill 2002)
  - Ensemble forecast of error statistics
  - Reduced rank analytical analysis
- Maximum likelihood ensemble filter (Zupanski 2005)
  - Ensemble forecast of error statistics
  - Minimize cost function in ensemble subspace
- …

Ensemble methods: CO_2

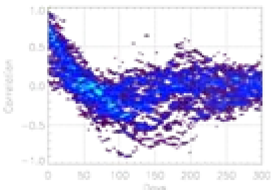
- CarbonTracker
  http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/
  - Ensemble square root filter
  - 135 parameters
  - 150 members



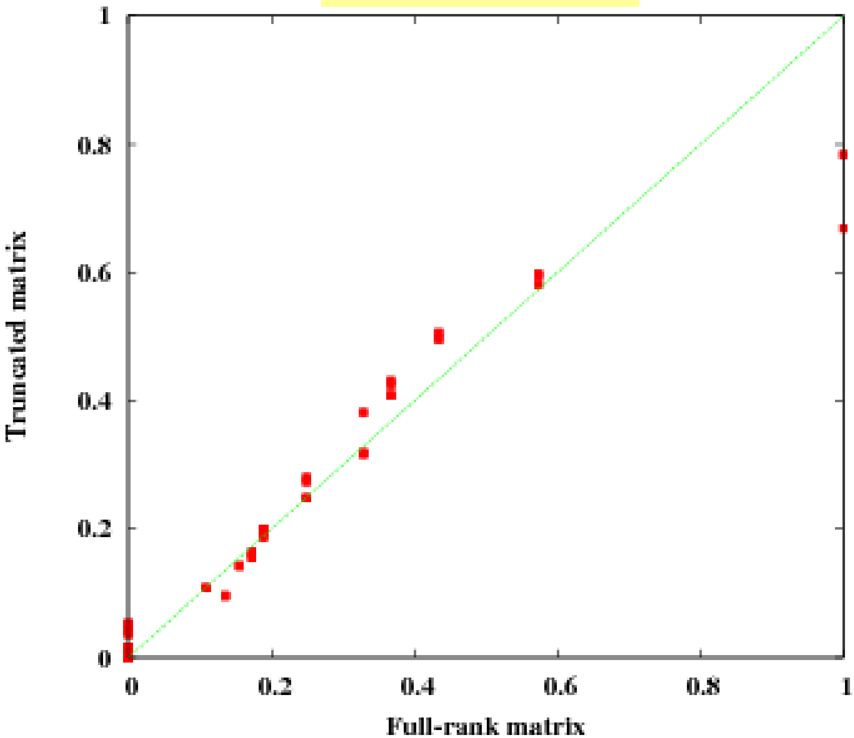


Replace **B** by an ensemble?

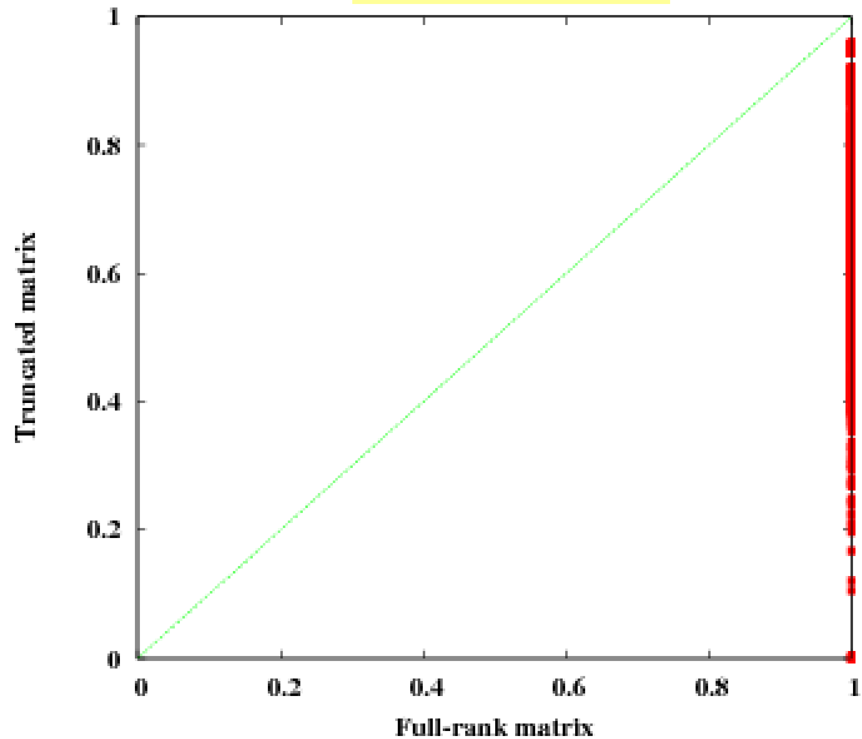
- Correlation matrix defined with 500km e-folding lengths over land and 1000km over ocean
  - Perform PCA – truncate to 500 PCs – validate



Anti-diagonal



Diagonal



Effective ensemble methods

- Localization
- Add hard constraints to reduce the size of the state vector
  - From flux estimation to model parameter estimation
- Split problem into pieces
  - Sequential
  - Pre-processing
- Trick or treat?
- Diagnostics only?

Ensemble methods for diagnostics

- Ensembles of inversions with consistent statistics make it possible to reconstruct the posterior pdf

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$

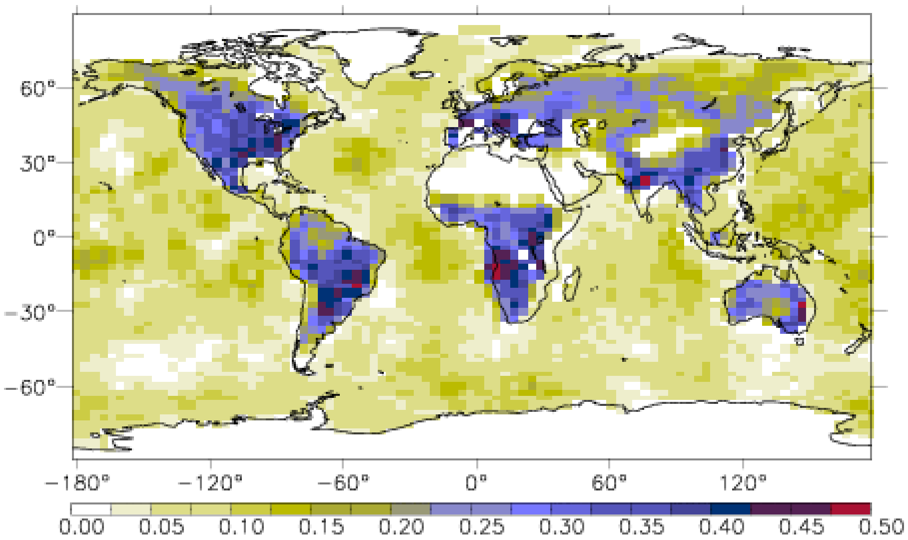
$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

$$\mathbf{A} = \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B}$$

- Sample $\mathbf{x}_b$ from **B**
- Sample **y** from **R**
- $\mathbf{x}_a$ follows **A**

OCO error reduction

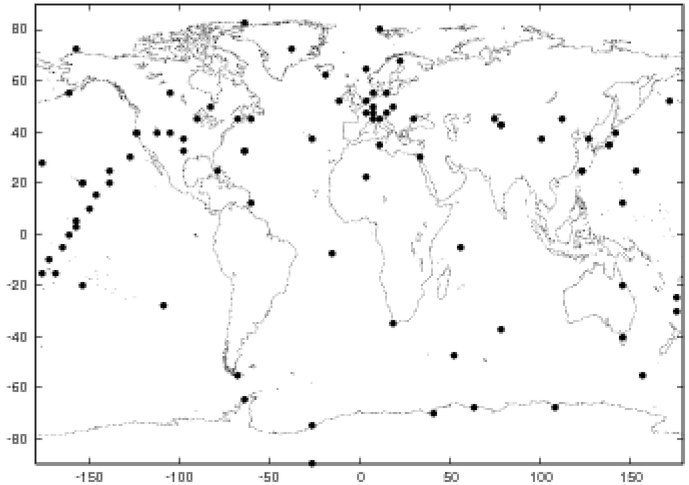
- No correlations
- error reduction 1-sig(post)/sig(prior)
- 4 years of 8-day segments (180 fluxes)

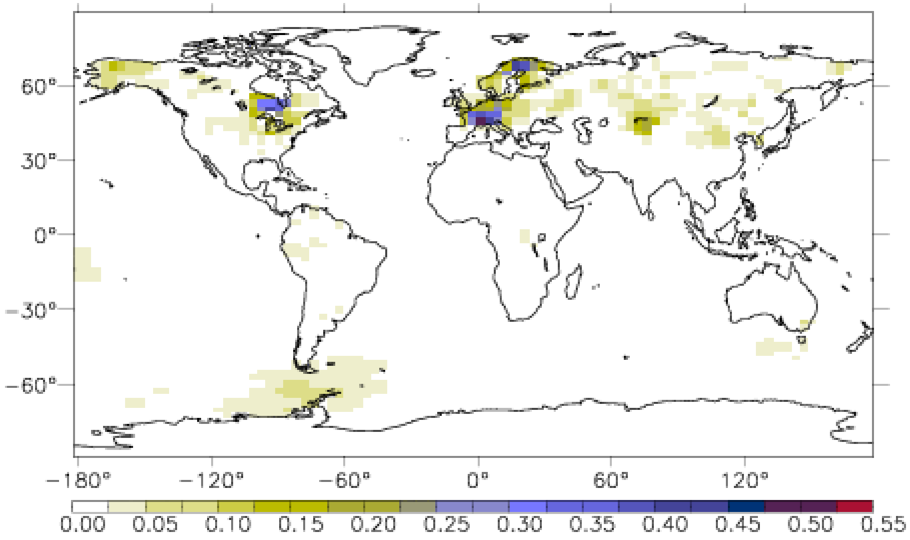


Reference error reduction

Expected uncertainty reduction

- Monte Carlo approach with 5 iterations
- Random errors

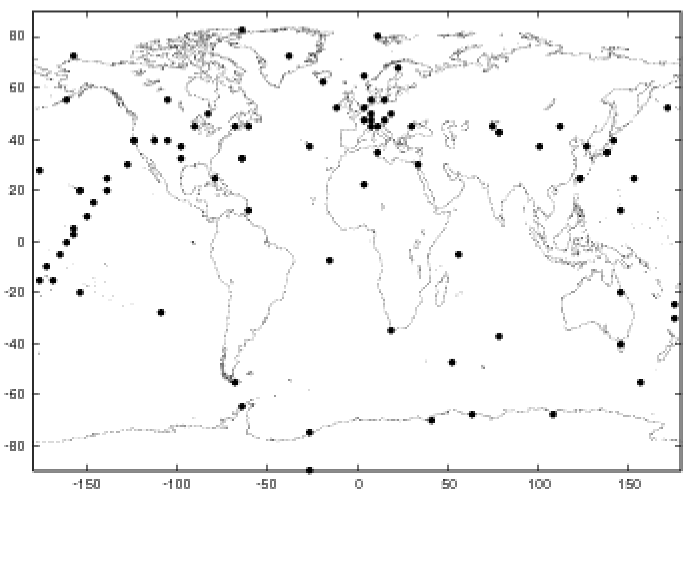


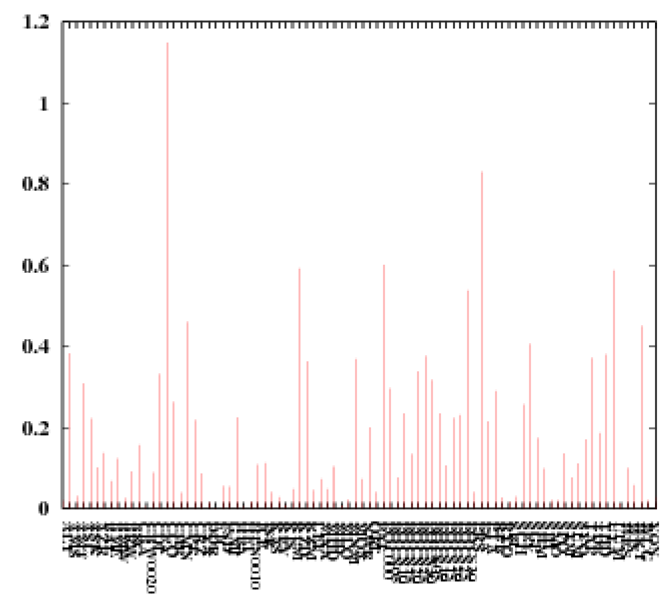


1 – sig_a/sig_b

Ensemble methods for diagnostics

- Similarly, one may compute the influence matrix
  - $\mathbf{S} = d\hat{\mathbf{y}}/d\mathbf{y} = \mathbf{R}^{-1}\mathbf{HAH}^{T}$, $\hat{\mathbf{y}} = \mathbf{H}x_a$



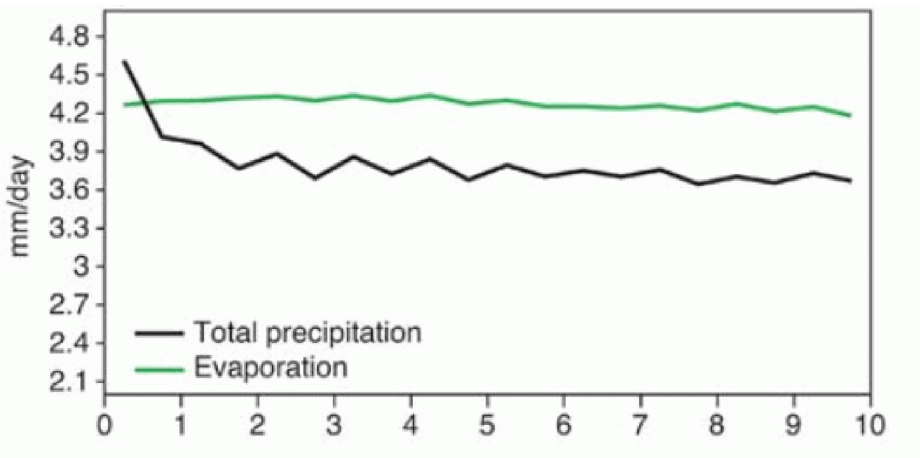


Evaluation

- $J(\mathbf{x}_a) < J(\mathbf{x}_b)$
- $J(\mathbf{x}_a)$ follows a chi-square pdf centered on p with std. dev. $\sqrt{p}$
  - p: number of observations
- The sum of two normal distributions is a normal distribution
  - $H(\mathbf{x}_b) - \mathbf{y}$: zero bias, covariance $\mathbf{HBH}^T+\mathbf{R}$
- Real world vs. theory

Evaluation (cont')

- Use independent (new) observations $\mathbf{y}_n$ unbiased with covariance $\mathbf{R}_n$
  - $H(\mathbf{x}_a) - \mathbf{y}_n$, unbiased, covariance $\mathbf{HAH}^T+\mathbf{R}_n$
  - $H(\mathbf{x}_a) - \mathbf{y}_n$ uncorrelated with $H(\mathbf{x}_b) - \mathbf{y}$ and unbiased
- Recycle $\mathbf{x}_a$
  - Spin-down or spin-up



Evolution of precipitation (mm day^{-1}) and evaporation (mm day^{-1}) in the tropical band between 30°N and 30°S during the 10-day forecasts, averaged over Apr 2002.
Andersson et al. (2005)

Assimilex

Summary

- Data assimilation for large state vectors
- Bayes' theorem as a paradigm
- Assignment of errors
- Analytical formulation
  - Gaussian framework, weakly-non-linear
  - Inversion of large matrices
  - Handling of a large observation operator
- Variational formulation
  - Gaussian framework, weakly-non-linear
- Monte Carlo formulation
  - Curse of dimensionality
- Reduce problem size
  - Localization
  - Add hard constraints
  - Split problem into pieces
- Hybrid approaches

Some references on-line

- F. Bouttier and P. Courtier: *Data assimilation concepts and methods*
  - http://www.ecmwf.int/newsevents/training/lecture_notes/pdf_files/ASSIM/Ass_cons.pdf

- E.T. Jaynes: *Probability theory: the logic of Science*
  - http://omega.albany.edu:8008/JaynesBook.html

- A. Tarantola: *Inverse problem theory*
  - http://www.ipgp.jussieu.fr/~tarantola/Files/Professional/Books/index.html