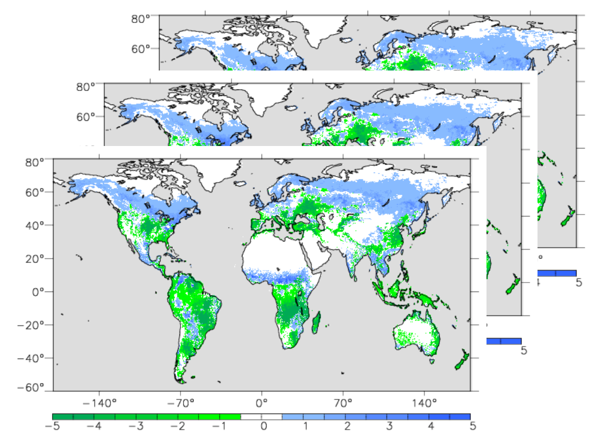
Data assimilation for large state vectors

Frédéric Chevallier

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



Objective

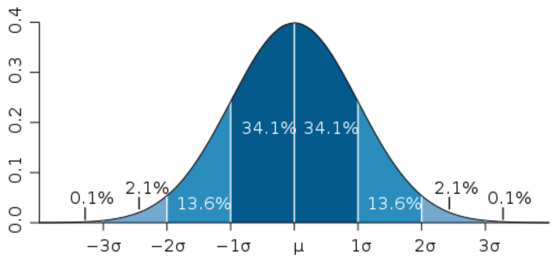
- Overview of data assimilation methods for large state vectors, from the point of view of CO_2 flux inversion
- Outline
  - Analytical formulation
  - Variational formulation
  - Monte Carlo formulation
  - Diagnostics
  - Prospects

Outline

- Analytical formulation
- Variational formulation
- Monte Carlo formulation
- Diagnostics
- Prospects

The linear problem with Gaussian error statistics and zero biases

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



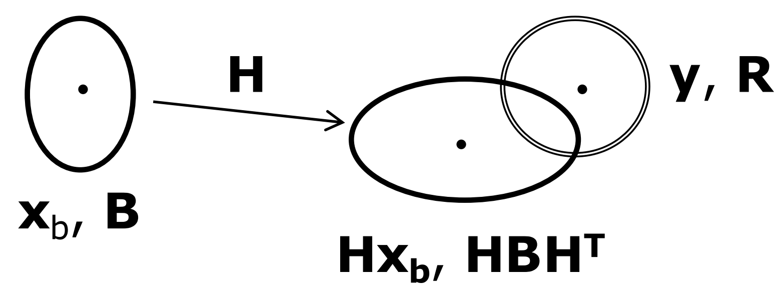
$n \in \mathbb{N}, \varphi \in \mathbb{R}^n, \mu \in \mathbb{R}^n, \mathbf{\Sigma} \in M(n, \mathbb{R})$ positive definite

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

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

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

x: state vector
$\mathbf{x_b}$: expected value of the prior pdf of x
y: observation vector
H: linear observation operator
B: Prior error covariance matrix
R: observation error covariance matrix



Analytical solution

- Expected value of the posterior PDF

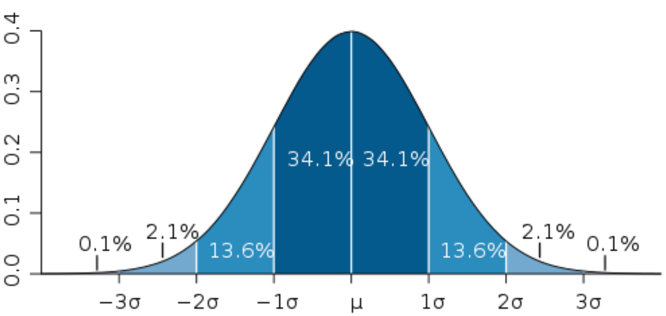
$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

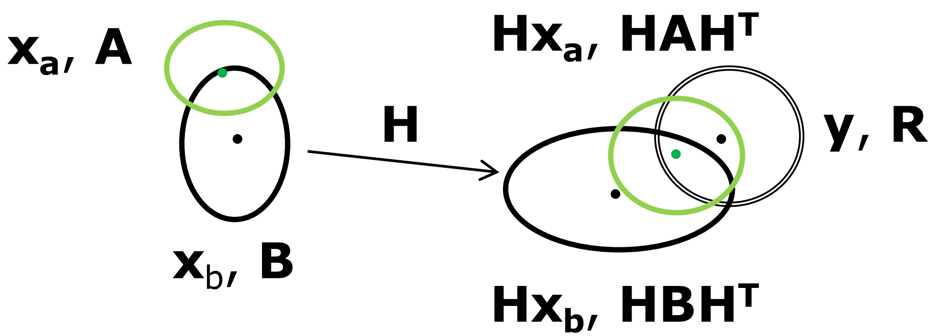
- Covariance of the posterior PDF

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

- Gain matrix

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





Example

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

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

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

Assigning errors

- Variances, correlations
- **B**:
  - Prior errors
- **R**:
  - Measurement errors
  - Representation errors
  - Errors of the observation operator **H**

Implementation

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

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

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

- $\mathbf{B}$, $\mathbf{R}$, $\mathbf{x}_b$, $\mathbf{y}$ and H given
- Issues:
  - Compute **H**
    - from *H*
    - Exact derivatives
    - Finite differences
  - Matrix inversion
    - In practice, restricted to ranks < 10^5

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 $\mathbf{x}_b$ constant and updating $\mathbf{H}_i$

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

Outline

- Analytical formulation
- Variational formulation
- Monte Carlo formulation
- Diagnostics
- Prospects

Variational solution

- The linear problem with Gaussian error statistics and zero biases

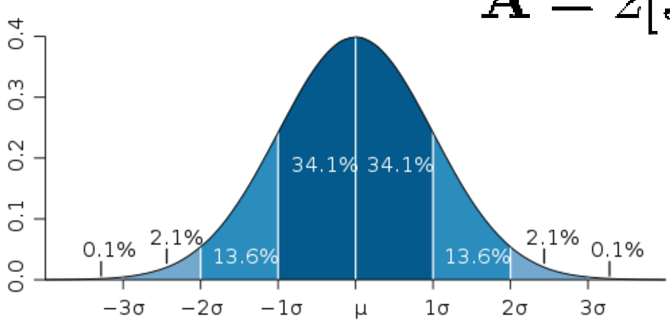
$$-2\ ln\ p(\mathbf{x}|\mathbf{y}) = (\mathbf{x}-\mathbf{x}_b)^{\mathrm{T}}\mathrm{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{Hx}-\mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{Hx}-\mathbf{y}) + c$$

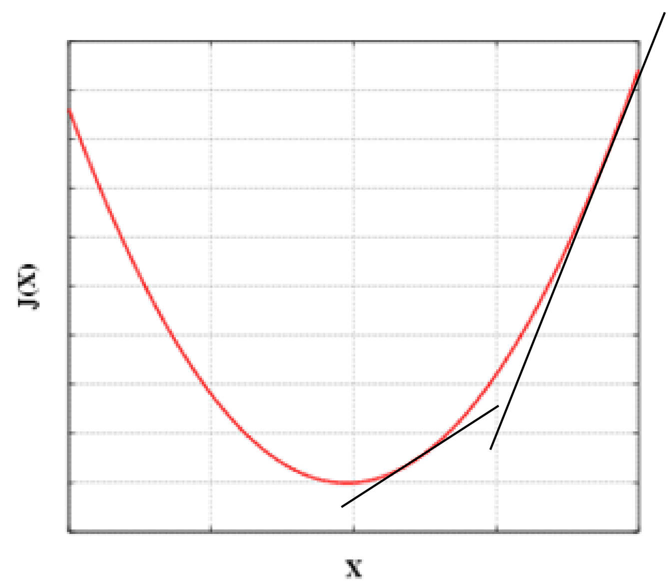
  - $\mathbf{x}_a$ minimises $J(\mathbf{x}) = -2\ ln\ p(\mathbf{x}|\mathbf{y})$

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

  - Covariance of the posterior PDF:

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





Implementation

- Inversion system:

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

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

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

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

Adjoint technique

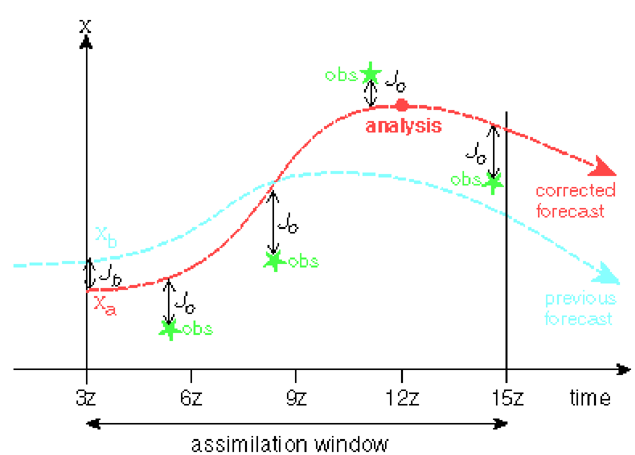
- Example:
  - Compute the adjoint instruction of the line:
    - $a = b^2$
- Forward statement
    - $a = b^2$
- Tangent-linear statement $\mathbf{y} = \mathbf{Hx}$
    - $\delta b = 0.\delta a + 1.\delta b$
    - $\delta a = 0.\delta a + 2b.\delta b$
- Adjoint statement $\mathbf{x}^* = \mathbf{H}^T\mathbf{y}^*$
    - $b^* = 2ba^* + 1.b^*$
    - $a^* = 0.a^* + 0.b^*$

Adjoint technique

- Example:
  - Compute the adjoint instruction of the line:
    - $a = a^2$
- Forward statement
    - $a = a^2$
- Tangent-linear statement $\mathbf{y} = \mathbf{H}\mathbf{x}$
    - $\delta a = 2a . \delta a$
- Adjoint statement $\mathbf{x}^* = \mathbf{H}^T \mathbf{y}^*$
    - $a^* = 2a . a^*$

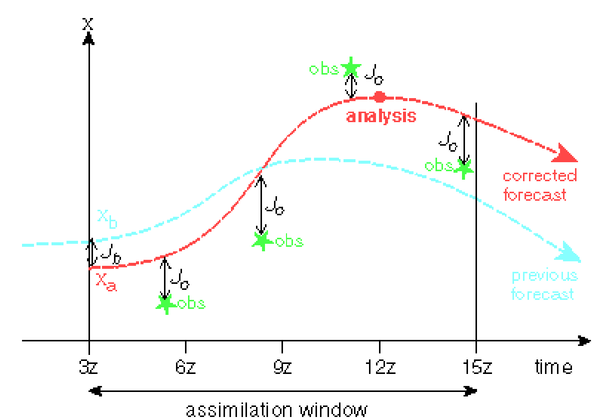
Handling the linearization points

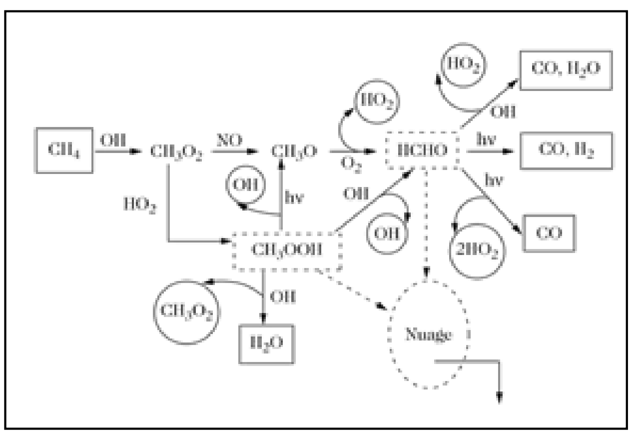
- Handling of trajectory
  - The derivatives in **H** are a function of **x**
  - $\mathbf{Hx} = \mathbf{H}_n\ \mathbf{H}_{n-1} \ldots \mathbf{H}_2\ \mathbf{H}_1\mathbf{x}$ (forward)
  - $\mathbf{H}^T\mathbf{y}^* = \mathbf{H}_!^T\ \mathbf{H}_2^T \ldots \mathbf{H}_{n-1}^T\ \mathbf{H}_n^T\mathbf{y}^*$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

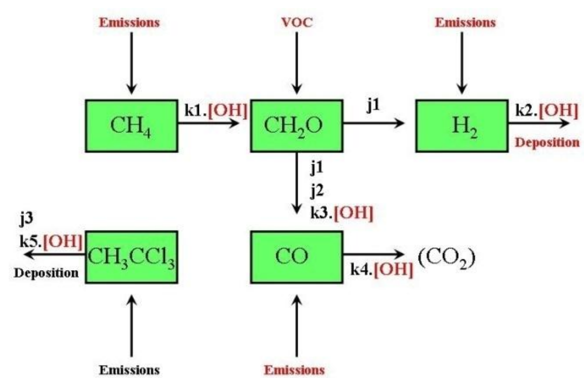


Which code?

- 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
  - 41 softwares currently listed on 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 towards the NL code
- Correctness of the AD…
  - Linearity
  - $(\mathbf{Hx})^T\mathbf{Hx} = \mathbf{x}^T\mathbf{H}^T(\mathbf{Hx})$
- … to the machine epsilon (relative error due to rounding in floating point arithmetic)

Invert **R** matrix

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

- Try to have **R** diagonal
  - Ignore correlations
  - Observation thinning
  - Increase variances and set correlations to zero
- Block-diagonal **R**
- Directly define the precision matrix $\mathbf{R}^{-1}$
  - Chevallier 2007, Mukherjee et al. 2011

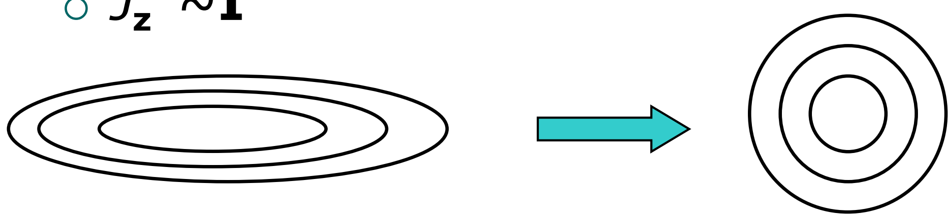
Invert **B** matrix

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

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

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 gradient methods
    - $J_{\mathbf{z}}'' \sim \mathbf{I}$



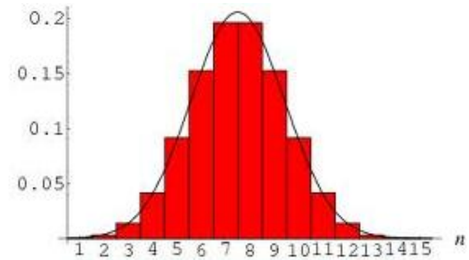
  - $\mathbf{z} = \mathbf{B}^{-1/2}(\mathbf{x}-\mathbf{x}_b)$ is a simple approximation
    - J unchanged
    - $\mathrm{grad}_z(J) = \mathbf{B}^{+1/2}\,\mathrm{grad}_{\mathbf{x}}(J)$

Outline

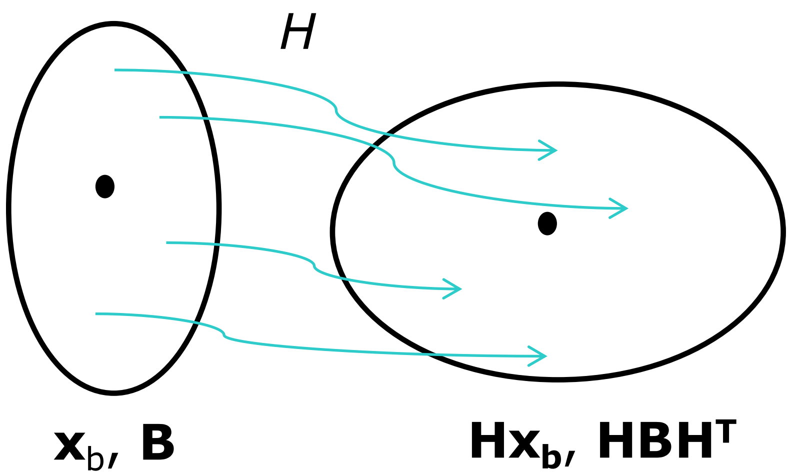
- Analytical formulation
- Variational formulation
- Monte Carlo formulation
- Diagnostics
- Prospects

Ensemble methods

- Principle: replace some of the pdf computations using finite-size ensembles



- Ex:

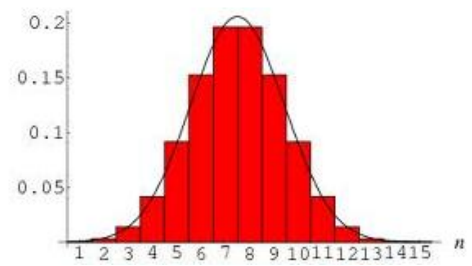


$$\mathrm{K} = \mathrm{BH^T}(\mathrm{HBH^T} + \mathrm{R})^{-1}$$

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

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

Ensemble methods



- Particle filters (Le Doucet et al. 2001)
- 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
  - ex: http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/
- Maximum likelihood ensemble filter (Zupanski 2005)
  - Ensemble forecast of error statistics
  - Minimize cost function in ensemble subspace
- ...

Ensemble methods

- Less limitation wrt linearity or pdf model
- No adjoint model
- Parallel hardware

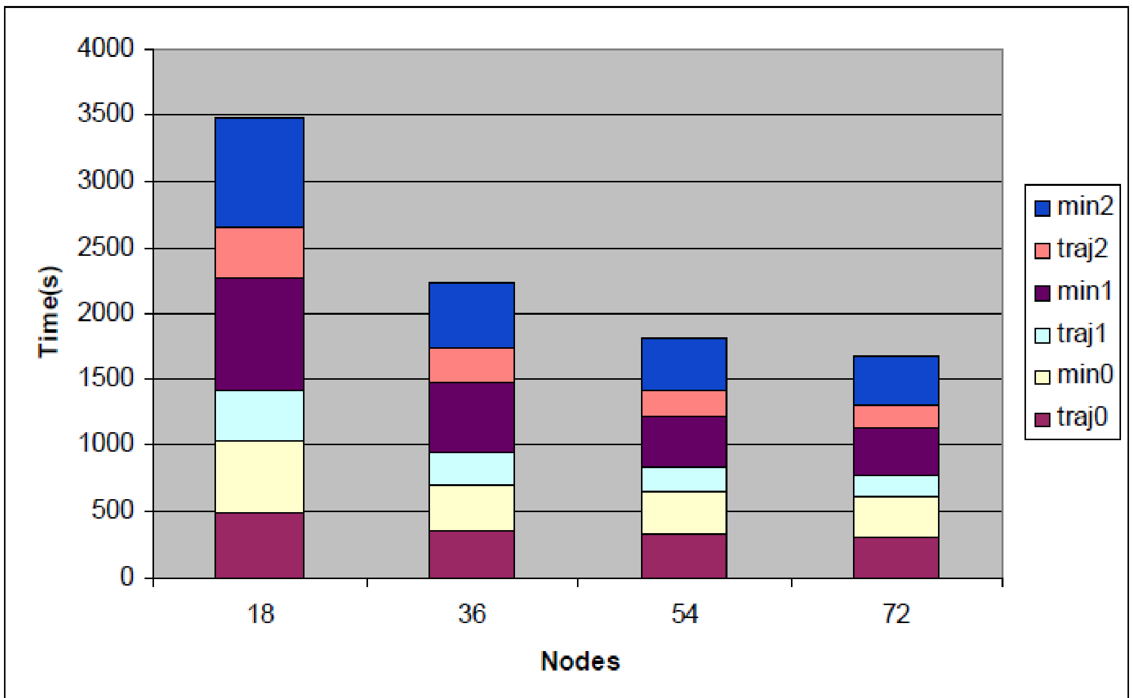


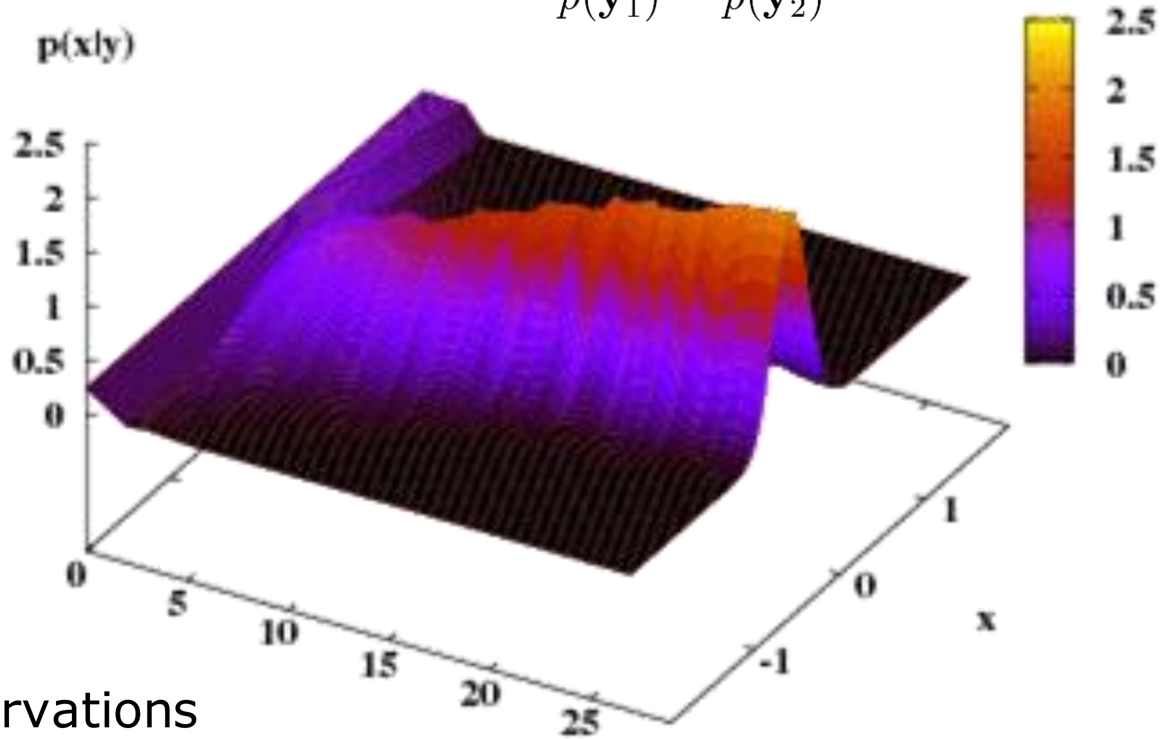
Figure 2. Run-times of 4D-Var for different node counts with 4D-Var broken down into its constituents.

Hamrud, ECMWF TM 616, 2010

Particle filter

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

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

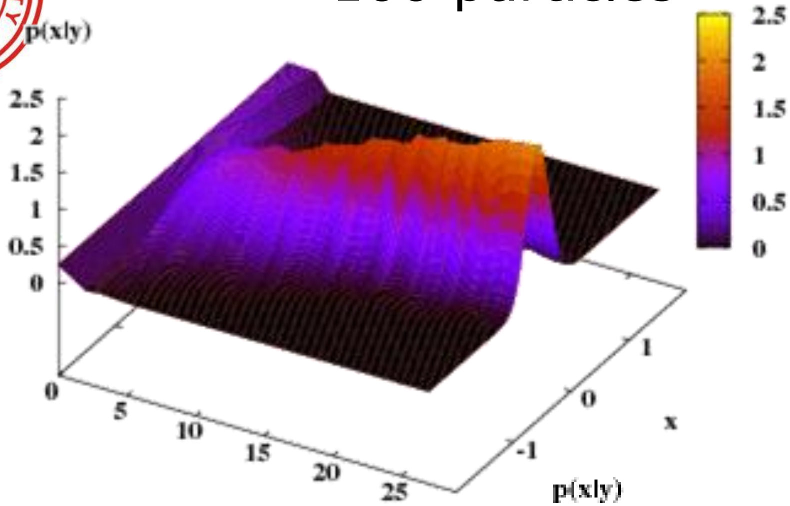


Ex: 100 particles
monovariate x,
Gaussian pdfs,
up to 25 observations

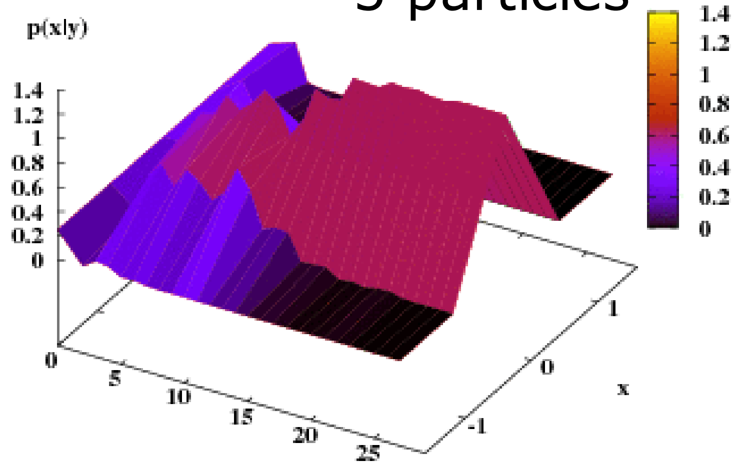
Number of observations

Particle filter

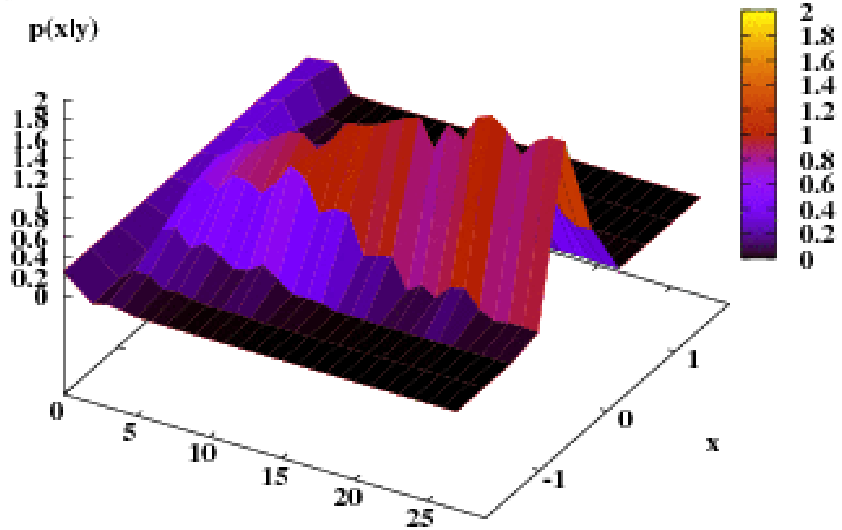
100 particles



5 particles



10 particles



Particle filter

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

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

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

Effective ensemble methods

- Fight against the curse of dimensionality
- Localization
  - Restrict the radius of influence of the observations
- Add hard constraints to reduce the size of the state vector
  - From flux estimation to model parameter estimation
- Split the problem into pieces
  - Sequential

- Trick or treat?

Ensemble methods for diagnostics

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

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

- Define truth $\mathbf{x}_t$
- Sample $\mathbf{x}_b$ from N($\mathbf{x}_t$, $\mathbf{B}$)
- Sample $\mathbf{y}$ from N($\mathbf{H}\mathbf{x}_t$, $\mathbf{R}$)

- The distribution of $\mathbf{x}_a$ follows $\mathbf{A}$

Outline

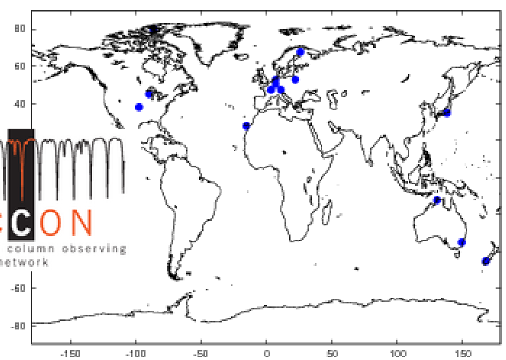
- Analytical formulation
- Variational formulation
- Monte Carlo formulation
- Diagnostics
- Prospects

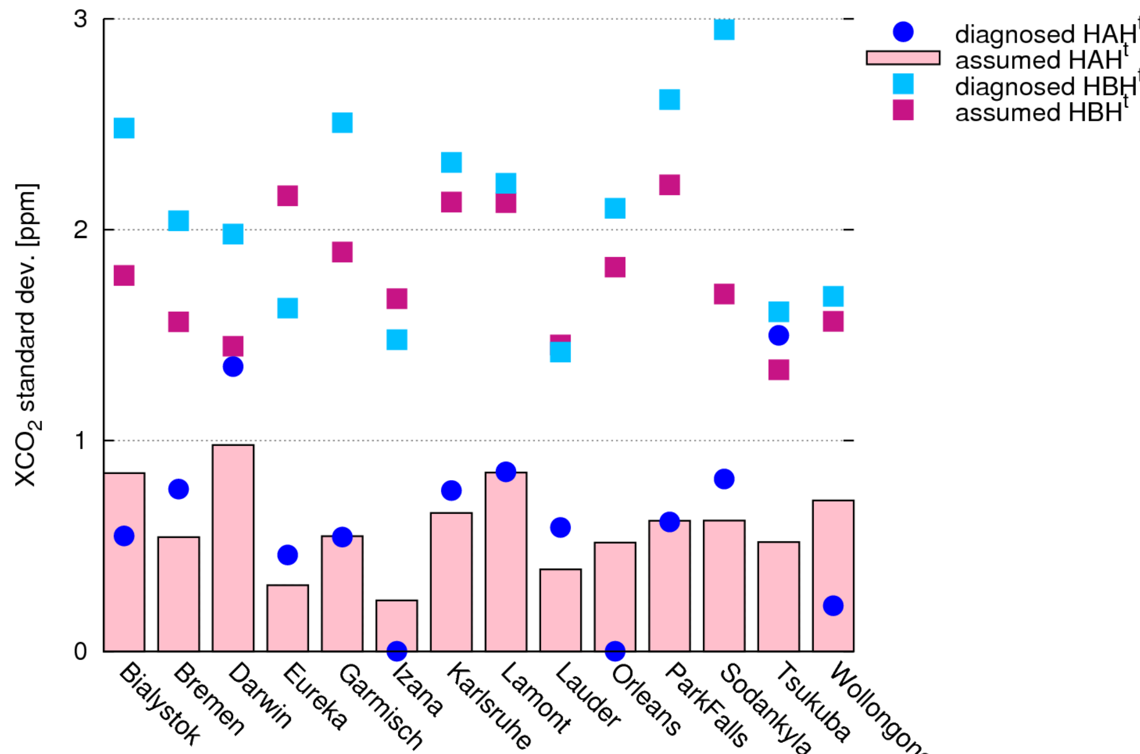
Evaluation

- Diagnosed error bars and error correlations
- $J(\mathbf{x}_a) < J(\mathbf{x}_b)$
- $J(\mathbf{x}_a)$ 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





Outline

- Analytical formulation
- Variational formulation
- Monte Carlo formulation
- Diagnostics
- Prospects

Inversion methods

- Analytical formulation
  - Matrix size limiting
- Ensemble methods
  - Ensemble size limiting
- Variational method
  - Iteration number limiting

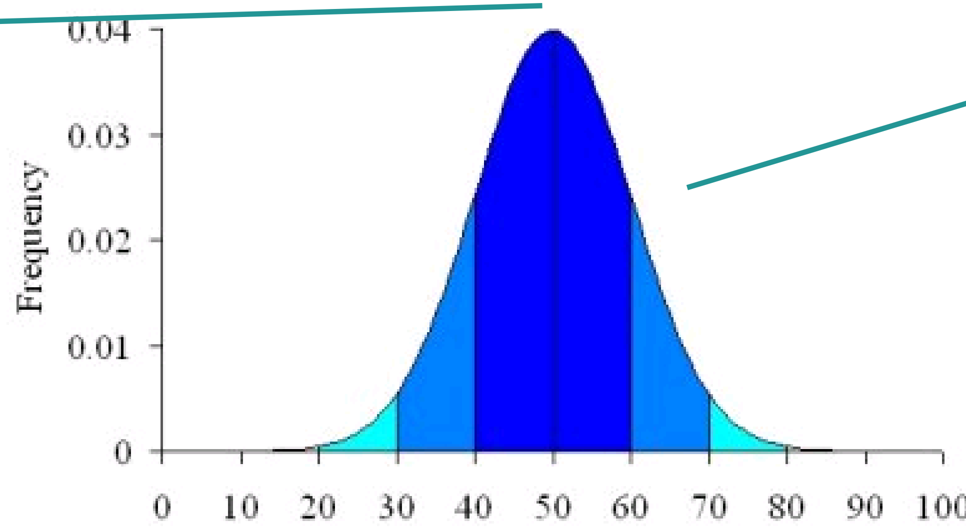
- Hybrid approaches

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x}).p(\mathbf{y}|\mathbf{x})$$

LSCE inversion system (PYVAR)

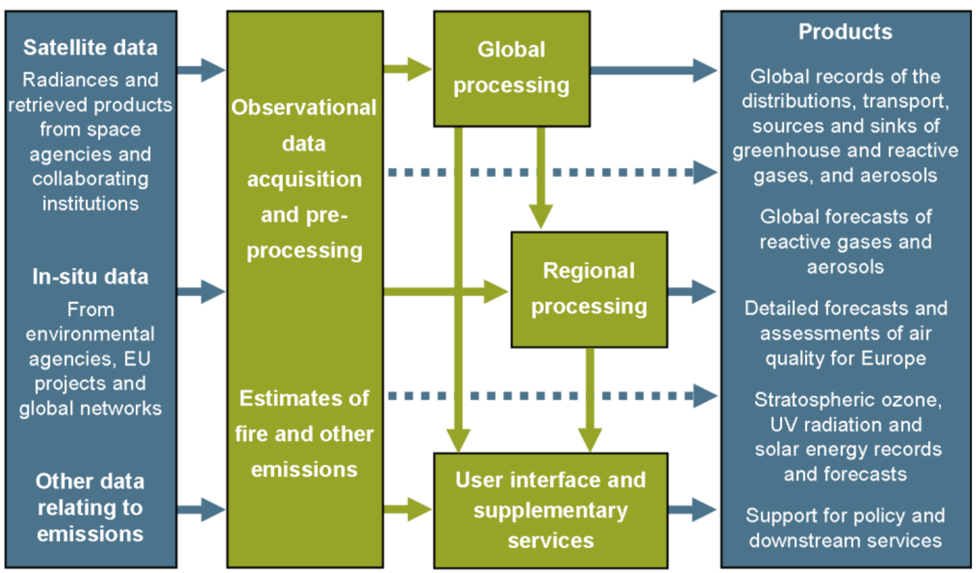
$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x}).p(\mathbf{y}|\mathbf{x})$$

- Variational approach for high-resolution information
  - Weekly fluxes at 3.75x2.5 deg^2 global
  - or hourly fluxes at ~ 10 – 100 km^2 regional
- Ensemble approach for coarse resolution information
  - Mean variance of the flux errors over long periods of time

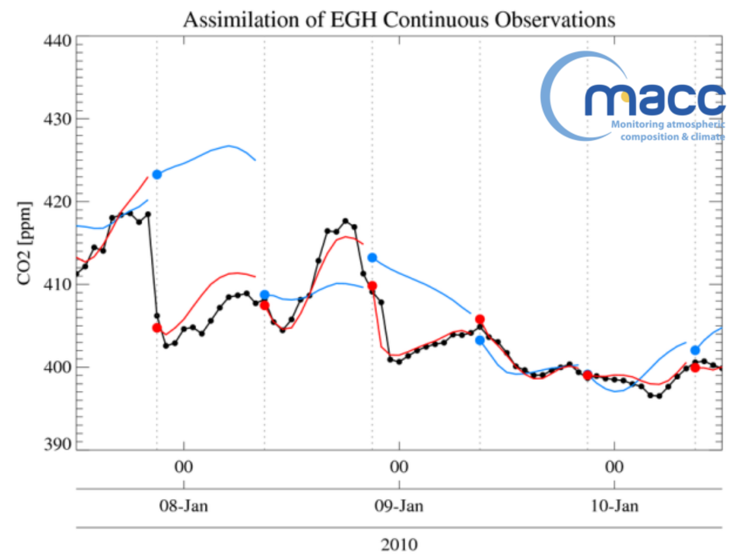


Towards an operational processing by dedicated centres

- European *Global Monitoring for Environment and Security*
  - Suite of projects GEMS, MACC and MACC-II
- NRT needs
- traceability



MACC service infrastructure



Assimilation of IMECC data within MACC (R. Engelen)

In ten years

- Dense regional networks and sparse international networks combined with satellite instruments
- High spatial resolution (<50km), even at global scale, very high resolution for specific areas, like cities or plants
- Inform policy at regional, national and international levels
  - Dense network needed (mesh < 100km)
- First space-borne lidar CO_2 measurements and CO_2 imagery
- Coupling with other carbon related observations within models of the carbon cycle
  - Comprehensive carbon information systems

Some references on-line

- F. Bouttier and P. Courtier: *Data assimilation concepts and methods*
  - www.ecmwf.int/newsevents/training/lecture_notes/pdf_files/ASSIM/Ass_cons.pdf
- D. Jacob: Lectures on inverse modeling
  - acmg.seas.harvard.edu/education/jacob_lectures_inverse_modeling.pdf
- E.T. Jaynes: *Probability theory: the logic of Science*
  - omega.albany.edu:8008/JaynesBook.html
- A. Tarantola: *Inverse problem theory*
  - www.ipgp.jussieu.fr/~tarantola/Files/Professional/Books/index.html

- Application to CO_2 flux inversion
  - www.esrl.noaa.gov/gmd/ccgg/carbontracker/
  - www.carboscope.eu