Variational data assimilation and the ensemble Kalman filter

Amos S. Lawless

Data Assimilation Research Centre
University of Reading

a.s.lawless@reading.ac.uk

http://www.personal.reading.ac.uk/~sms00asl/

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Variational data assimilation – the idea

In variational data assimilation we seek the solution that maximises the *a posterior* probability $p(\mathbf{x}|\mathbf{y})$.

Since

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\{ -\frac{1}{2}\{(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + (H(\mathbf{x})-\mathbf{y})^T\mathbf{R}^{-1}(H(\mathbf{x})-\mathbf{y})\} \}$$

we will have the maximum probability when $\mathbf{x}$ minimises

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

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

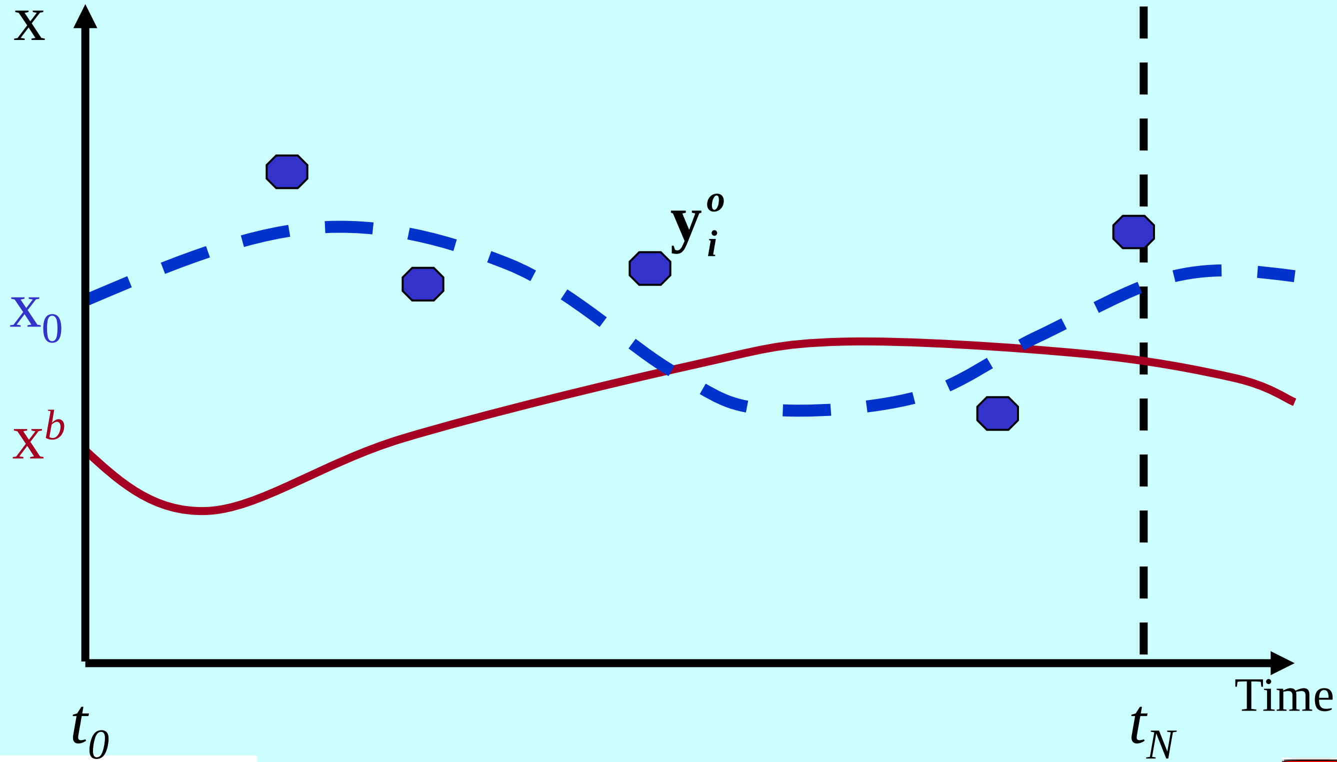
We consider two main algorithms

- Three-dimensional variational assimilation (3D-Var)
  - Where we consider 3 space dimensions.
- Four-dimensional variational assimilation (4D-Var)
  - Where we consider 3 space dimensions plus time as the 4^{th} dimension.
  - In this case we can consider the observation operator to include the dynamical model.

We will present 4D-Var first and 3D-Var as a variant of this.

Four-dimensional variational assimilation (4D-Var)

Aim: Find the best estimate of the true state of the system (*analysis*), consistent with both observations distributed in time and the system dynamics.



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

4D-Var cost function

Minimize

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{N}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)$$

with respect to x_0, subject to

$$\mathbf{x}_{i+1} = \mathcal{M}_i(\mathbf{x}_i),$$

x^b - *a priori* (background) state – Size of order 10^8 - 10^9

y_i - Observations – Size of order 10^6 - 10^7

H_i - Observation operator

B - Background error covariance matrix

R_i - Observation error covariance matrix

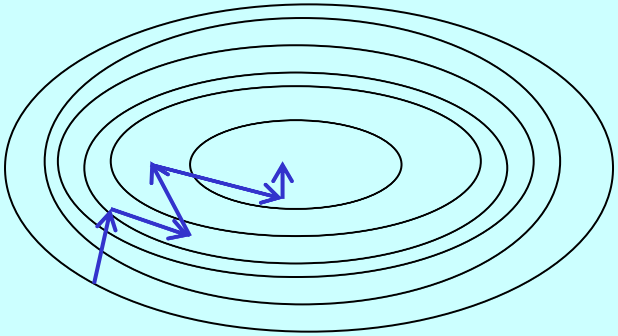
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

For N=0 we have 3D-Var

University of Reading

Numerical minimization - Gradient descent methods

Iterative methods, where each successive iteration is based on the value of the function and its gradient at the current iteration.



$$\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} - \alpha\, \varphi(\mathbf{x}_0^{(k)})$$

where α is a step length and φ is a direction that depends on $J(\mathbf{x}_0^{(k)})$ and its gradient.

Problem: How do we calculate the gradient of $J(\mathbf{x}_0^{(k)})$ with respect to $\mathbf{x}_0^{(k)}$?

Method of Lagrange multipliers

We introduce Lagrange multipliers $\boldsymbol{\lambda}_i$ at time t_i and define the Lagrangian

$$\mathcal{L}(\mathbf{x}_i, \boldsymbol{\lambda}_i) = \mathcal{J}(\mathbf{x}_0) + \sum_{i=0}^{N-1} \boldsymbol{\lambda}_{i+1}^{\mathrm{T}} (\mathbf{x}_{i+1} - \mathcal{M}_i(\mathbf{x}_i))$$

Then necessary conditions for a minimum of the cost function subject to the constraint are found by taking variations with respect to $\boldsymbol{\lambda}_i$ and $\mathbf{x}_i$.

Variations with respect to $\boldsymbol{\lambda}_i$ simply give the original constraint.

$$\mathcal{L}(\mathbf{x}_i, \boldsymbol{\lambda}_i) = \mathcal{J}(\mathbf{x}_0) + \sum_{i=0}^{N-1} \boldsymbol{\lambda}_{i+1}^{\mathrm{T}} (\mathbf{x}_{i+1} - \mathcal{M}_i(\mathbf{x}_i))$$

Variations with respect to $\mathbf{x}_i$ give the *adjoint* equations

$$\boldsymbol{\lambda}_i = \mathbf{M}_i^T \boldsymbol{\lambda}_{i+1} - \mathbf{H}_i^T \mathbf{R}_i^{-1} (\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)$$

with boundary condition $\boldsymbol{\lambda}_{N+1} = 0$.

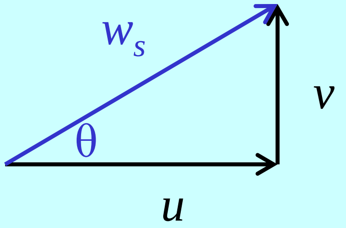
Then at initial time we have

$$\nabla \mathcal{J}(\mathbf{x}_0) = -\boldsymbol{\lambda}_0 + \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b)$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

An aside – What are the linear operators **H** & **M**?



Suppose we observe the wind speed w_s.

Then we have $\mathbf{x} = \begin{pmatrix} u \\ v \end{pmatrix}$, $\mathbf{y} = w_s$ and $\mathbf{y} = H(\mathbf{x})$

with

$$H(\mathbf{x}) = \sqrt{u^2 + v^2}$$

Then

$$\mathbf{H} = \begin{pmatrix} \frac{\partial H}{\partial u} & \frac{\partial H}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{u}{\sqrt{u^2+v^2}} & \frac{v}{\sqrt{u^2+v^2}} \end{pmatrix}$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

So where have we got to?

We wish to minimize

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{N}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)$$

with respect to x_0, subject to

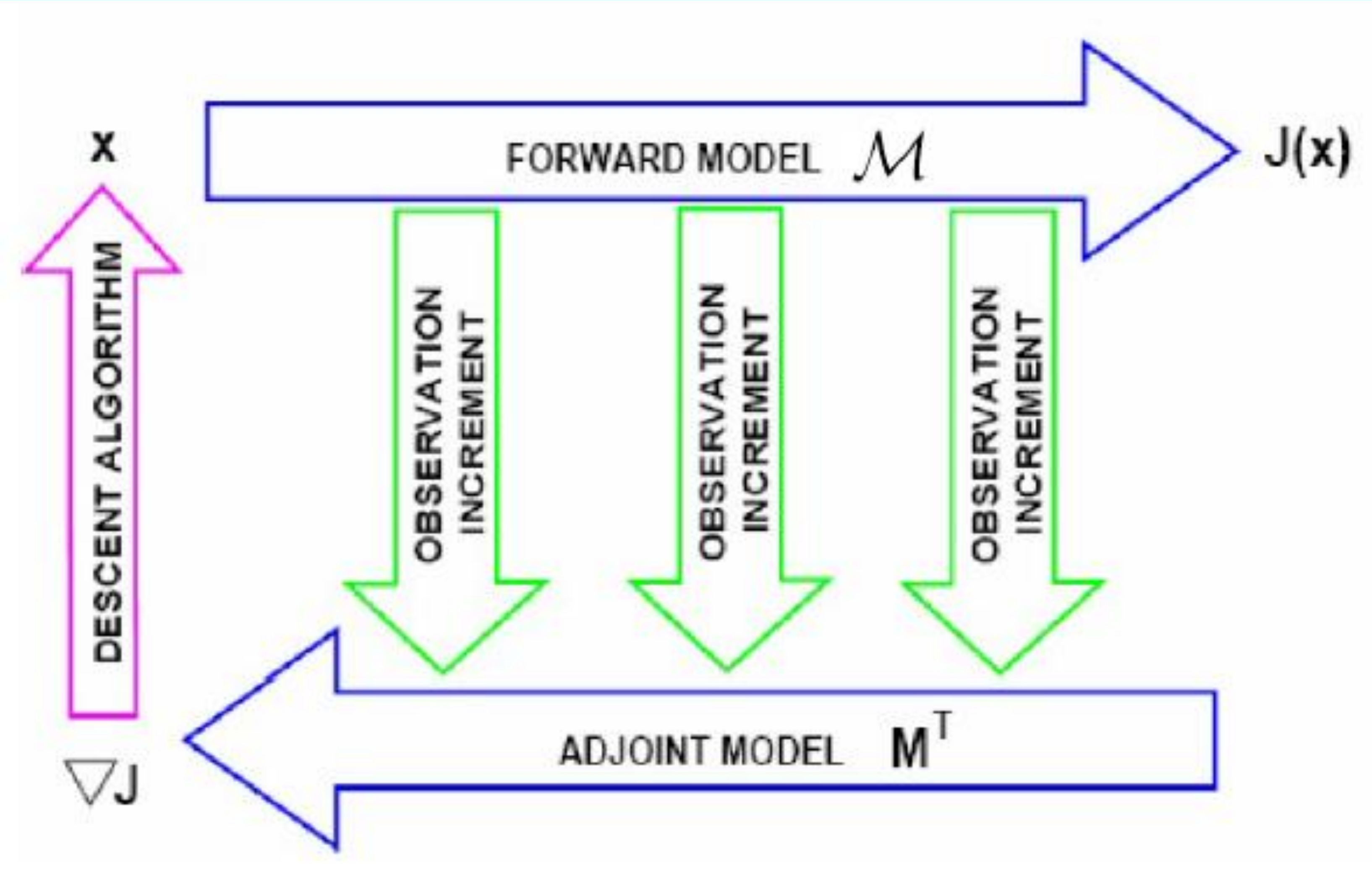
$$\mathbf{x}_{i+1} = \mathcal{M}_i(\mathbf{x}_i),$$

On each iteration we have to calculate *J* and its gradient

- To calculate *J* we need to run the nonlinear model
- To calculate the gradient of *J* we need one run of the adjoint model (backward in time)

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

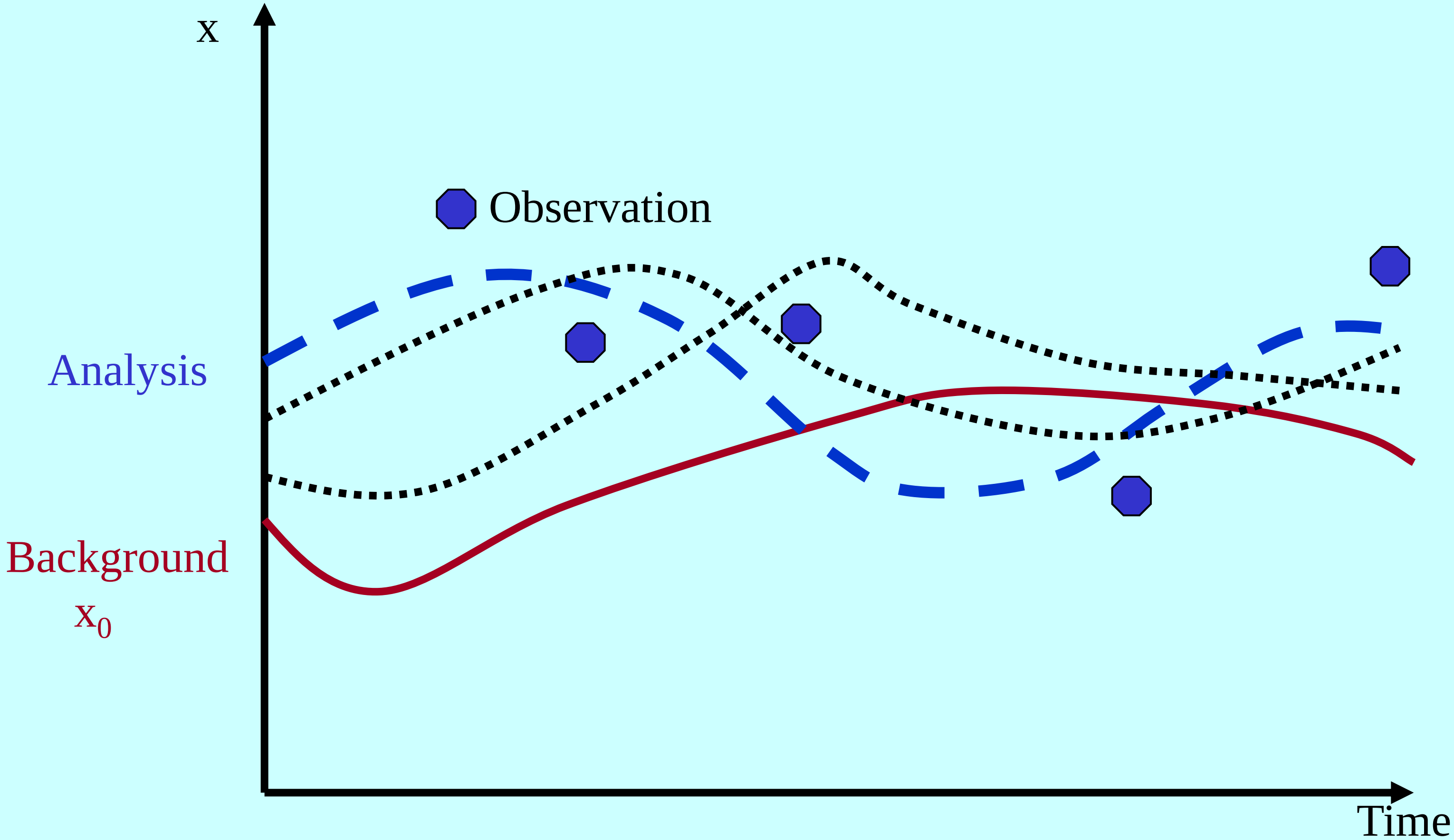


BUT this can be computationally expensive!

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Incremental 4D-Var



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Incremental 4D-Var

We solve a series of linearized minimization problems

$$\begin{aligned}\tilde{\mathcal{J}}^{(k)}[\delta\mathbf{x}_0^{(k)}] &= \frac{1}{2}(\delta\mathbf{x}_0^{(k)} - [\mathbf{x}^b - \mathbf{x}_0{}^{(k)}])^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{x}_0^{(k)} - [\mathbf{x}^b - \mathbf{x}_0{}^{(k)}]) \\ &+ \frac{1}{2}\sum_{i=0}^{N}(\mathbf{H}_i\delta\mathbf{x}_i^{(k)} - \mathbf{d}_i^{(k)})^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{H}_i\delta\mathbf{x}_i^{(k)} - \mathbf{d}_i^{(k)})\end{aligned}$$

with

$$\begin{aligned}\mathbf{d}_i &= \mathbf{y}_i - \mathcal{H}_i[\mathbf{x}_i^{(k)}] \\ \delta\mathbf{x}_{i+1} &= \mathbf{M}_i\delta\mathbf{x}_i\end{aligned}$$

and update using

$$\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} + \delta\mathbf{x}_0^{(k)}$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Comments on incremental formulation

- Inner loop cost function is linear quadratic, so has a unique minimum.
- Can simplify the linear model (low resolution, simplified physics) in order to save computational time.
- Equivalent to an approximate Gauss-Newton procedure – Convergence results proved by *Lawless, Gratton & Nichols, QJRMS, 2005; Gratton, Lawless & Nichols, SIAM J. on Optimization, 2007.*
- Used in several operational centres, including ECMWF and Met Office.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

3D-FGAT (First guess at appropriate time)

We solve a series of linearized minimization problems

$$\begin{aligned}\tilde{\mathcal{J}}^{(k)}[\delta\mathbf{x}_0^{(k)}] &= \frac{1}{2}(\delta\mathbf{x}_0^{(k)} - [\mathbf{x}^b - \mathbf{x}_0{}^{(k)}])^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{x}_0^{(k)} - [\mathbf{x}^b - \mathbf{x}_0{}^{(k)}]) \\ &+ \frac{1}{2}\sum_{i=0}^{N}(\mathbf{H}_i\delta\mathbf{x}_i^{(k)} - \mathbf{d}_i^{(k)})^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{H}_i\delta\mathbf{x}_i^{(k)} - \mathbf{d}_i^{(k)})\end{aligned}$$

with

$$\begin{aligned}\mathbf{d}_i &= \mathbf{y}_i - \mathcal{H}_i[\mathbf{x}_i^{(k)}] \\ \delta\mathbf{x}_{i+1} &= \mathbf{M}_i\delta\mathbf{x}_i\end{aligned}$$

Replace this equation with
$\delta\mathbf{x}_{i+1} = \delta\mathbf{x}_i$

and update using

$$\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} + \delta\mathbf{x}_0^{(k)}$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Properties of 4D-Var

- Observations are treated at correct time.
- Use of dynamics means that more information can be obtained from observations.
- Covariances are implicitly evolved.
- Standard formulation assumes model is perfect. Weak-constraint 4D-Var being developed to relax this assumption.
- In practice development of linear and adjoint models may be complex, but can be done at level of code.

Ensemble Kalman filter

The basic idea

- In the Kalman filter we assimilate the observations sequentially,making use of the equation we found in the first lecture.

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

- The background state comes from the forecast of the previous analysis.
- In the Kalman filter, the uncertainty on the background comes from a forecast of the uncertainty on the analysis.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Framework

- We assume a linear model and observation operator.
- The model may be imperfect.

$$\mathbf{x}_{i+1}^{t} = \mathbf{M}_i \mathbf{x}_i^{t} + \boldsymbol{\eta}_i$$

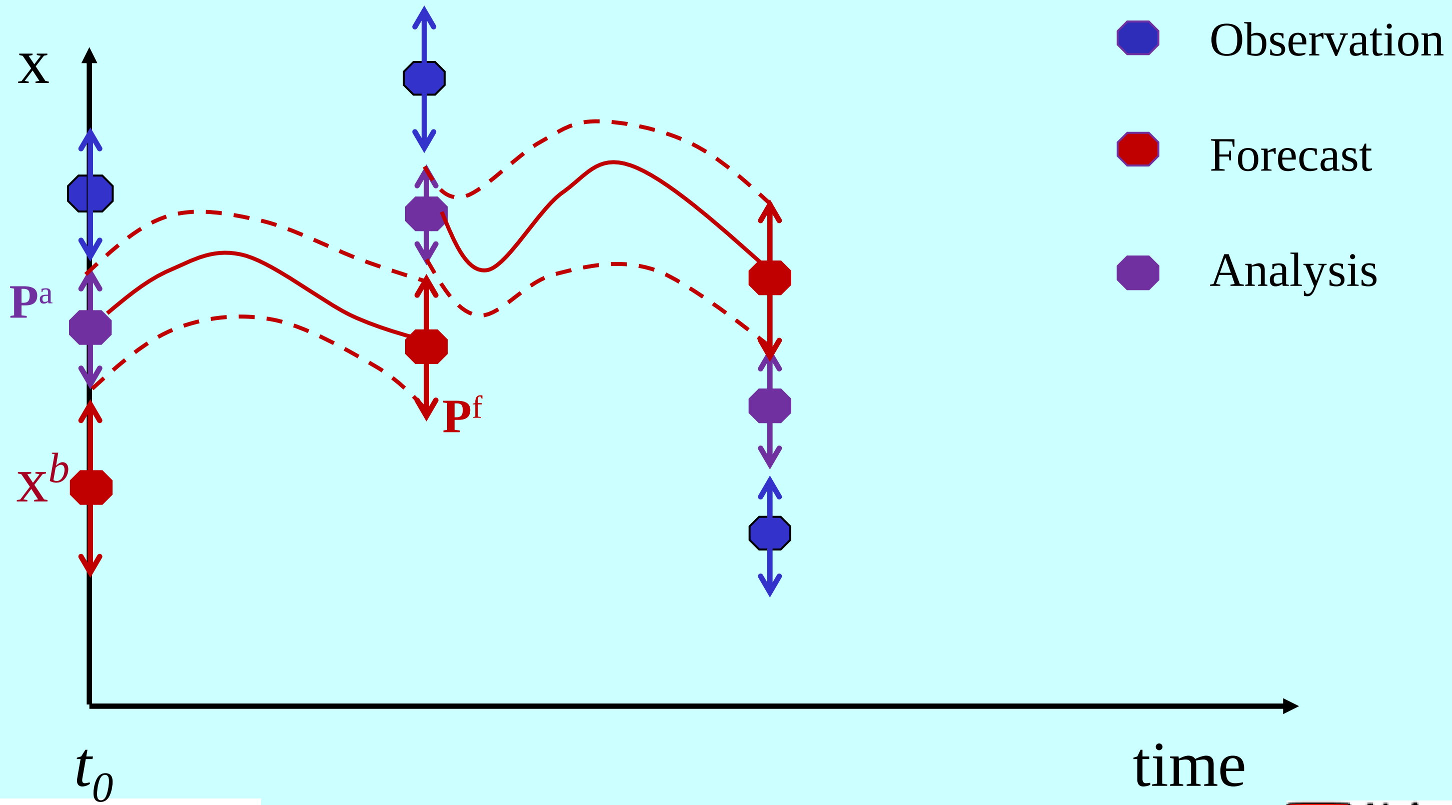
with

$$< \boldsymbol{\eta}_i >= 0$$

and

$$< \boldsymbol{\eta}_i \boldsymbol{\eta}_i^{T} >= \mathbf{Q}_i$$

Kalman filter - Illustration



We have the following steps:

- Kalman gain computation

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

- State analysis

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}_i (\mathbf{y}_i - \mathbf{H}_i \mathbf{x}_i^f)$$

- Error covariance of analysis

$$\mathbf{P}_i^a = (\mathbf{I} - \mathbf{K}_i \mathbf{H}_i) \mathbf{P}_i^f$$

- State forecast

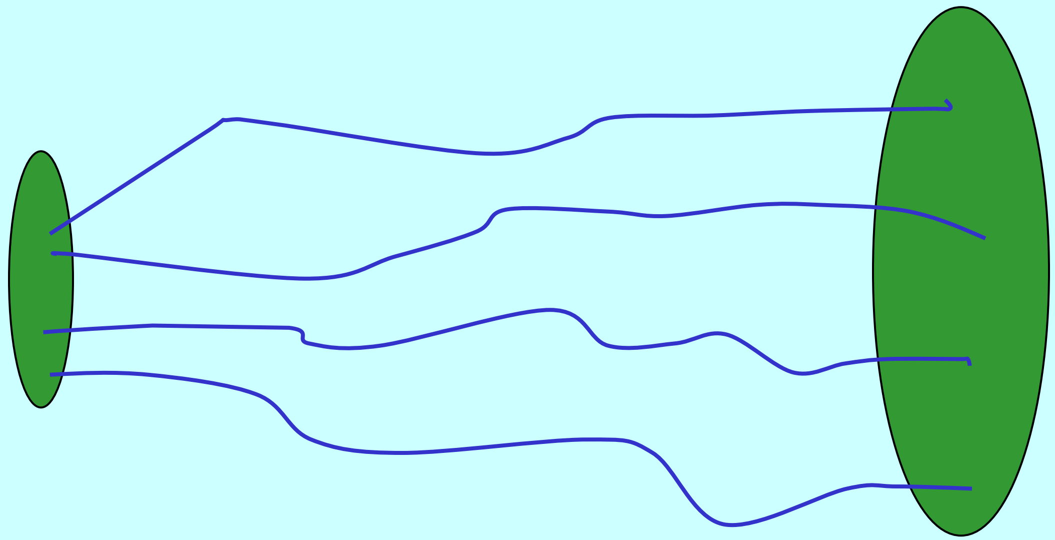
$$\mathbf{x}_i^f = \mathbf{M}_{i-1}\mathbf{x}_{i-1}^a$$

- Error covariance forecast

$$\mathbf{P}_i^f = \mathbf{M}_{i-1}\mathbf{P}_{i-1}^a\mathbf{M}_{i-1}^T + \mathbf{Q}_{i-1}$$

Ensemble Kalman filter

In the ensemble Kalman filter (EnKF) the error covariance forecast is approximated by an ensemble of model runs



Uncertainty at analysis time

Uncertainty at forecast time with covariance **P** (Gaussian)

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

There are 2 main types of EnKF:

I. Perturbed observation EnKF
II. Deterministic EnKF

We consider both of them.

I. Perturbed observation EnKF

Prediction step:

1. Evolve each ensemble member using the nonlinear model

$$\mathbf{x}^{(i),f} = \mathcal{M}(\mathbf{x}^{(i),a}) + \boldsymbol{\eta}$$

$$\boldsymbol{\eta} \sim \mathcal{N}(0, \mathbf{Q})$$

2. Form the ensemble mean

$$\overline{\mathbf{x}^f} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}^{(i),f}$$

3. Form the perturbation matrix and reconstruct the forecast error covariance matrix

$$\mathbf{X}^f = \frac{1}{\sqrt{N-1}}[\mathbf{x}^{(1),f} - \overline{\mathbf{x}^f}, \mathbf{x}^{(2),f} - \overline{\mathbf{x}^f}, \ldots, \mathbf{x}^{(N),f} - \overline{\mathbf{x}^f}]$$

$$\mathbf{P}^f = \mathbf{X}^f (\mathbf{X}^f)^T$$

Analysis step:

Update each ensemble member, perturbing the observations

$$\mathbf{x}^{(i),a} = \mathbf{x}^{(i),f} + \mathbf{K}(\mathbf{y} + \epsilon_y + \mathbf{H}\mathbf{x}^{(i),f})$$

with

$$\epsilon_y \sim \mathcal{N}(0, \mathbf{R})$$

and

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R}_e)^{-1}$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Notes

- Perturbing the observations is necessary to ensure

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f$$

- However this introduces extra sampling noise.

- There is also the problem of needing to invert a low rank matrix, for example using the pseudo-inverse.

II. Deterministic EnKF

The idea here is to create an ensemble consistent with

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f$$

The prediction step is the same as for the perturbed observation EnKF.

The analysis step then proceeds as follows:

Analysis step:

1. Transform the forecast ensemble to observation space

$$\mathbf{y}^{(i),f} = \mathbf{H}\mathbf{x}^{(i),f}$$

2. Compute the mean $\overline{\mathbf{y}^f}$ and a perturbation matrix $\mathbf{Y}^f$
3. Compute the analysis

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{K}(\mathbf{y} - \overline{\mathbf{y}^f})$$

with

$$\mathbf{K} = \mathbf{X}^f (\mathbf{Y}^f)^T (\mathbf{Y}^f (\mathbf{Y}^f)^T + \mathbf{R})^{-1}$$

and

$$\mathbf{X}^a = \mathbf{X}^f \mathbf{T}$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

The matrix **T** is chosen such that

$$\begin{aligned}\mathbf{P}^a &= \mathbf{X}^a(\mathbf{X}^a)^T = (\mathbf{X}^f\mathbf{T})(\mathbf{X}^f\mathbf{T})^T \\ &\approx (\mathbf{I} - \mathbf{KH})\mathbf{P}^f\end{aligned}$$

Different choices of **T** lead to different versions of the EnKF.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

EnKF issues

The small ensemble size relative to the size of the system leads to 2 problems that must be faced:

1. The ensemble collapses, i.e. the matrix $\mathbf{P}^f$ does not contain enough spread.

   Solution: Covariance inflation

$$\mathbf{P}^f = (1+\rho)\mathbf{P}_e^f$$

EnKF issues

2. The ensemble covariance matrix $\mathbf{P}^f$ is low rank, which leads to spurious long-range correlations

Solution: Covariance localization

$$\mathbf{P}^f = \mathbf{L} \circ \mathbf{P}_e^f$$

where **L** is a matrix that ensures long-range correlations are zero.

References – Variational methods

P. Courtier, J-N. Thepaut and A. Hollingsworth (1994), A strategy for operational implementation of 4D-Var, using an incremental approach, *Quart. J. Roy. Meteor. Soc.*, 120, 1367–1387.

Gratton, S., Lawless, A.S. and Nichols, N.K. (2007), Approximate Gauss-Newton methods for nonlinear least squares problems, SIAM J. on Optimization, 18, 106-132.

Lawless, A.S., Gratton, S. and Nichols, N.K. (2005), An investigation of incremental 4D-Var using non-tangent linear models, Quart. J. Royal Met. Soc., 131, 459-476.

Lawless, A.S. (2013), Variational data assimilation for very large environmental problems. In *Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences* (2013), Eds. Cullen, M.J.P., Freitag, M. A., Kindermann, S., Scheichl, R., Radon Series on Computational and Applied Mathematics 13. De Gruyter, pp. 55-90.

O. Talagrand and P. Courtier (1987), Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Quart. J. Roy. Meteor. Soc.*, 113, 1311-1328.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

References - EnKF

J. L. Anderson. An ensemble adjustment Kalman filter for data assimilation. Mon. Wea. Rev., 129:2884–2903, 2001.

G. Burgers, P. J. van Leeuwen, and G. Evensen. Analysis scheme in the ensemble Kalman filter. Mon. Wea. Rev., 126: 1719–1724, 1998.

G. Evensen. Data Assimilation: The Ensemble Kalman Filter. Springer, 2nd edition edition, 2009.

D. M. Livings, S. L. Dance, and N. K. Nichols. Unbiased ensemble square root filters. Physica D: Nonlinear Phenomena, 237(8):1021–1028, 2008.

M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, and J.S. Whitaker. Ensemble square root filters. Mon. Wea. Rev., 131:1485–1490, 2003.