Variational data assimilation

Background and methods

Lecturer: Ross Bannister, thanks: Amos Lawless

NCEO, Dept. of Meteorology, Univ. of Reading

5-8 March 2019, Univ. of Reading

Bayes' Theorem

Bayes' Theorem

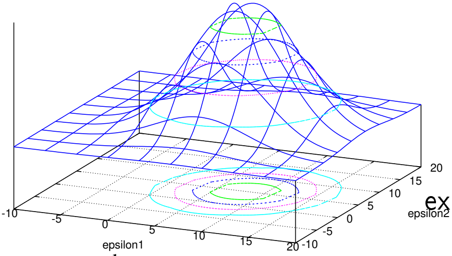
$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x}) \times p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$

$$\text{posterior distribution} = \frac{\text{prior distribution} \times \text{likelihood}}{\text{normalizing constant}}$$

- Prior distribution: PDF of the state before observations are considered (e.g. PDF of model forecast).
- Likelihood: PDF of observations given that the state is $\mathbf{x}$.
- Posterior: PDF of the state after the observations have been considered.

The Gaussian assumption

- PDFs are often described by Gaussians (normal distributions).
- Gaussian PDFs are described by a mean and covariance only.

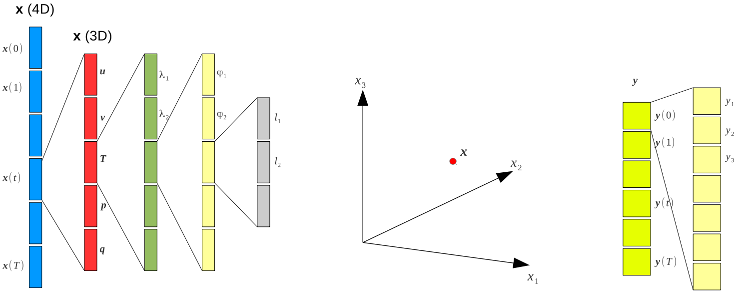


$$\varepsilon = \mathbf{x} - \mathbf{x}^{\mathrm{b}}$$

$$\mathbf{x} \sim N(\mathbf{x}^{\mathrm{b}}, \mathbf{B})$$

$$P(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{B})}} \times \exp -\frac{1}{2}\left(\mathbf{x} - \mathbf{x}^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}^{-1}\left(\mathbf{x} - \mathbf{x}^{\mathrm{b}}\right)$$

Meaning of x and y



- $\mathbf{x}^{\mathrm{a}}$ analysis; $\mathbf{x}^{\mathrm{b}}$ background state; $\delta\mathbf{x}$ increment (perturbation)
- $\mathbf{y}$ observations; $\mathbf{y}^{\mathrm{m}} = \mathscr{H}(\mathbf{x})$ model observations.
- $\mathscr{H}(\mathbf{x})$ is the observation operator / forward model.
- Sometimes $\mathbf{x}$ and $\mathbf{y}$ are for only one time (3DVar).
- $\mathbf{x}$-vectors have n elements; $\mathbf{y}$-vectors have p elements.

Back to the Gaussian assumption

Prior: mean $\mathbf{x}^{\mathrm{b}}$, covariance $\mathbf{B}$

$$P(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{B})}} \exp -\frac{1}{2}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right)$$

Likelihood: mean $\mathscr{H}(\mathbf{x})$, covariance $\mathbf{R}$

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^p \det(\mathbf{R})}} \exp -\frac{1}{2}\left(\mathbf{y}-\mathscr{H}(\mathbf{x})\right)^{\mathrm{T}} \mathbf{R}^{-1}\left(\mathbf{y}-\mathscr{H}(\mathbf{x})\right)$$

Posterior

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &= \frac{p(\mathbf{x}) \times p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})} \propto \exp -\frac{1}{2}\Big[\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right) \\ &\qquad + \left(\mathbf{y}-\mathscr{H}(\mathbf{x})\right)^{\mathrm{T}} \mathbf{R}^{-1}\left(\mathbf{y}-\mathscr{H}(\mathbf{x})\right)\Big] \end{aligned}$$

Variational DA – the idea

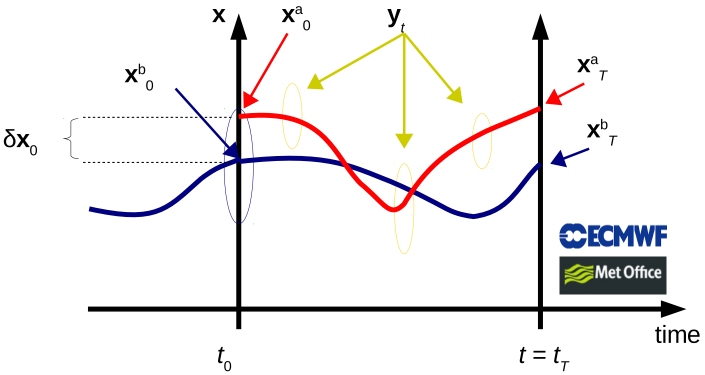
- In Var., we seek a solution that maximizes the posterior probability $p(\mathbf{x}|\mathbf{y})$ (maximum-a-posteriori).
- This is the most likely state given the observations (and the background), called the analysis, $\mathbf{x}^{\mathrm{a}}$.
- Maximizing $p(\mathbf{x}|\mathbf{y})$ is equivalent to minimizing $-\ln p(\mathbf{x}|\mathbf{y}) \equiv J(\mathbf{x})$ (a least-squares problem).

$$\begin{aligned}
p(\mathbf{x}|\mathbf{y}) &= C \exp -\frac{1}{2}\left[\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right)\right. \\
&\qquad \left. + \left(\mathbf{y}-\mathscr{H}(\mathbf{x})\right)^{\mathrm{T}} \mathbf{R}^{-1}\left(\mathbf{y}-\mathscr{H}(\mathbf{x})\right)\right] \\
J(\mathbf{x}) &= -\ln C + \frac{1}{2}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right) \\
&\qquad + \frac{1}{2}\left(\mathbf{y}-\mathscr{H}(\mathbf{x})\right)^{\mathrm{T}} \mathbf{R}^{-1}\left(\mathbf{y}-\mathscr{H}(\mathbf{x})\right) \\
&= \text{constant (ignored)} + J^{\mathrm{b}}(\mathbf{x}) + J^{\mathrm{o}}(\mathbf{x})
\end{aligned}$$

Four-dimensional Var (4DVar)

Aim

To find the 'best' estimate of the true state of the system (analysis), consistent with the observations, the background, and the system dynamics.



Towards a 4DVar cost function

Consider the observation operator in this case:

$$\mathscr{H}(\mathbf{x}) = \mathscr{H}\begin{pmatrix} \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_T \end{pmatrix} = \begin{pmatrix} \mathscr{H}_0(\mathbf{x}_0) \\ \vdots \\ \mathscr{H}_T(\mathbf{x}_T) \end{pmatrix}$$

So the J^{o} is (assume that $\mathbf{R}$ is block diagonal):

$$J^{\mathrm{o}} = \frac{1}{2}(\mathbf{y} - \mathscr{H}(\mathbf{x}))^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y} - \mathscr{H}(\mathbf{x})) =$$

$$\frac{1}{2}\begin{pmatrix} \mathbf{y}_0 - \mathscr{H}_0(\mathbf{x}_0) \\ \vdots \\ \mathbf{y}_T - \mathscr{H}_T(\mathbf{x}_T) \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \mathbf{R}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_T \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y}_0 - \mathscr{H}_0(\mathbf{x}_0) \\ \vdots \\ \mathbf{y}_T - \mathscr{H}_T(\mathbf{x}_T) \end{pmatrix}$$

$$= \frac{1}{2}\sum_{i=0}^{T} (\mathbf{y}_i - \mathscr{H}_i(\mathbf{x}_i))^{\mathrm{T}} \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathscr{H}_i(\mathbf{x}_i))$$

$$\text{where } \mathbf{x}_{i+1} = \mathscr{M}_i(\mathbf{x}_i)$$

The 4DVar cost function ('full 4DVar')

$$\text{Let } (\mathbf{a})^{\mathrm{T}} \mathbf{A}^{-1} (\mathbf{a}) \equiv (\mathbf{a})^{\mathrm{T}} \mathbf{A}^{-1} (\bullet)$$

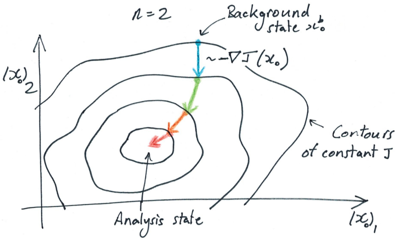
$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}_0^{-1} (\bullet) + \frac{1}{2}(\mathbf{y} - \mathscr{H}(\mathbf{x}))^{\mathrm{T}} \mathbf{R}^{-1} (\bullet) \\ &= \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}_0^{-1} (\bullet) + \frac{1}{2}\sum_{i=0}^{T}(\mathbf{y}_i - \mathscr{H}_i(\mathbf{x}_i))^{\mathrm{T}} \mathbf{R}_i^{-1} (\bullet) \end{aligned}$$

$$\text{subject to } \mathbf{x}_{i+1} = \mathscr{M}_i(\mathbf{x}_i)$$

- $\mathbf{x}_0^{\mathrm{b}}$ a-priori (background) state at t_0.
- $\mathbf{y}_i$ observations at t_i.
- $\mathscr{H}_i(\mathbf{x}_i)$ observation operator at t_i.
- $\mathbf{B}_0$ background error covariance matrix at t_0.
- $\mathbf{R}_i$ observation error covariance matrix at t_i.

How to minimize this cost function?

Minimize $J(\mathbf{x})$ iteratively



Use the gradient of J at each iteration:

$$\mathbf{x}_0^{k+1} = \mathbf{x}_0^k + \alpha \nabla J(\mathbf{x}_0^k)$$

The gradient of the cost function

$$\nabla J(\mathbf{x}_0) = \begin{pmatrix} \partial J / \partial (\mathbf{x}_0)_1 \\ \vdots \\ \partial J / \partial (\mathbf{x}_0)_n \end{pmatrix}$$

$-\nabla J$ points in the direction of steepest descent.

Methods: steepest descent (inefficient), conjugate gradient (more efficient), . . .

The gradient of the cost function (wrt $\mathbf{x}(t_0)$)

Either:

1. Diff. $J(\mathbf{x}_0)$ w.r.t. $\mathbf{x}_0$ with $\mathbf{x}_i = \mathscr{M}_{i-1}(\mathscr{M}_{i-2}(\cdots \mathscr{M}_0(\mathbf{x}_0)))$.
2. Diff. $J(\mathbf{x}) = J(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_T)$ w.r.t. $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_T$ subject to the constraint

$$\mathbf{x}_{i+1} - \mathscr{M}_i(\mathbf{x}_i) = 0$$

$$L(\mathbf{x}, \lambda) = J(\mathbf{x}) + \sum_{i=0}^{T-1} \lambda_{i+1}^{\mathrm{T}} (\mathbf{x}_{i+1} - \mathscr{M}_i(\mathbf{x}_i)).$$

Each approach leads to the **adjoint method**

- An efficient means of computing the gradient.
- Uses the linearized/adjoint of $\mathscr{M}_i$ and $\mathscr{H}_i$: $\mathbf{M}_i^{\mathrm{T}}$ and $\mathbf{H}_i^{\mathrm{T}}$.

The adjoint method

Equivalent gradient formula:

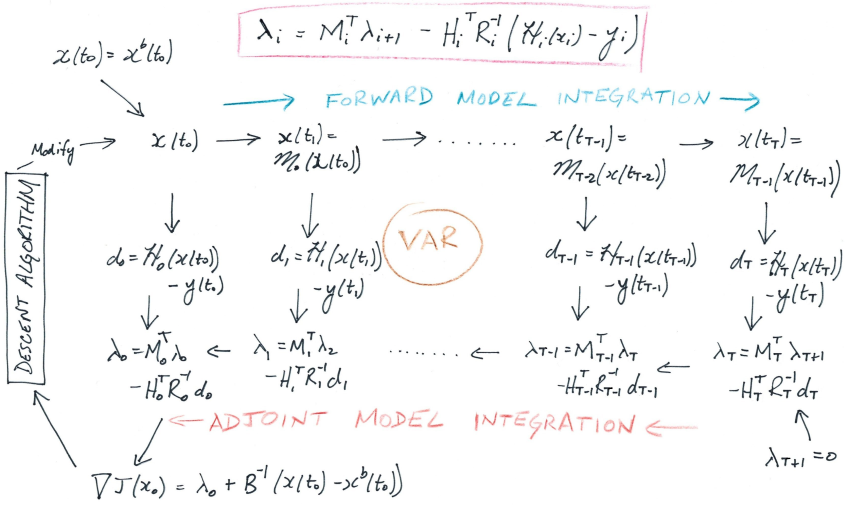
1.

$$\nabla J \equiv \nabla J(\mathbf{x}_0) = \mathbf{B}_0^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right) - \\ - \sum_{i=0}^{T} \mathbf{M}_0^{\mathrm{T}} \ldots \mathbf{M}_{i-1}^{\mathrm{T}} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \left(\mathbf{y}_i - \mathscr{H}_i(\mathbf{x}_i)\right)$$

2.

$$\begin{aligned} \lambda_{T+1} &= 0 \\ \lambda_i &= \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \left(\mathbf{y}_i - \mathscr{H}_i(\mathbf{x}_i)\right) + \mathbf{M}_i^{\mathrm{T}} \lambda_{i+1} \\ \lambda_0 &= \nabla J_{\mathrm{o}} \\ \therefore \nabla J &= \nabla J_{\mathrm{b}} + \nabla J_{\mathrm{o}} = \mathbf{B}_0^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right) + \lambda_0 \end{aligned}$$

The adjoint method



Simplifications and complications

- The full 4DVar method is expensive and difficult to solve.
- Model $\mathscr{M}_i$ is non-linear.
- Observation operators, $\mathscr{H}_i$ can be non-linear.
- Linear $\mathscr{H} \rightarrow$ quadratic cost function – easy(er) to minimize, $J^{\mathrm{o}} \sim \frac{1}{2}(y - ax)^2/\sigma_{\mathrm{o}}^2$.
- Non-linear $\mathscr{H} \rightarrow$ non-quadratic cost function – hard to minimize, $J^{\mathrm{o}} \sim \frac{1}{2}(y - f(x))^2/\sigma_{\mathrm{o}}^2$.
- Later will recognise that models are 'wrong'!

Look for simplifications:

Incremental 4DVar (linearized 4DVar)
3D-FGAT
3DVar

Complications:

Weak constraint
(imperfect model)

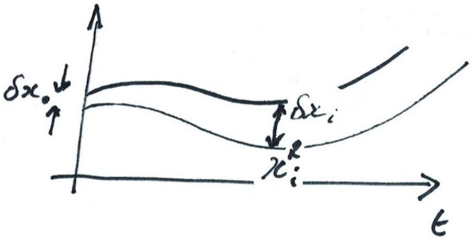
$$\text{definitions: } \mathbf{x}^{\mathrm{R}}_{i+1(k)} = \mathscr{M}_i\left(\mathbf{x}^{\mathrm{R}}_{i(k)}\right)$$

$$\mathbf{x}_i = \mathbf{x}^{\mathrm{R}}_{i(k)} + \delta\mathbf{x}_i \qquad \mathbf{x}^{\mathrm{b}}_0 = \mathbf{x}^{\mathrm{R}}_{0(k)} + \delta\mathbf{x}^{\mathrm{b}}_0$$

$$\mathbf{x}_{i+1} = \mathscr{M}_i(\mathbf{x}_i) \qquad \delta\mathbf{x}_{i+1} \approx \mathbf{M}_{i(k)}\delta\mathbf{x}_i$$

$$\mathscr{H}_i(\mathbf{x}_i) \approx \mathscr{H}_i\left(\mathbf{x}^{\mathrm{R}}_{i(k)}\right) + \mathbf{H}_{i(k)}\delta\mathbf{x}_i$$

$$\delta\mathbf{x}_i \approx \mathbf{M}_{i-1(k)}\mathbf{M}_{i-2(k)}\ldots\mathbf{M}_{0(k)}\delta\mathbf{x}_0$$



Incremental 4DVar 2

$$J_{(k)}(\delta\mathbf{x}_0) = \frac{1}{2}\left(\delta\mathbf{x}_0 - \delta\mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}}\mathbf{B}_0^{-1}\left(\bullet\right) + \frac{1}{2}\sum_{i=0}^{T}\left(\mathbf{y}_i - \mathscr{H}_i(\mathbf{x}_{i(k)}^{\mathrm{R}}) - \mathbf{H}_{i(k)}\delta\mathbf{x}_i\right)^{\mathrm{T}}\mathbf{R}_i^{-1}\left(\bullet\right)$$

- 'Inner loop': iterations to find $\delta\mathbf{x}_0$ (as adjoint method).
- 'Outer loop' (k): iterate $\mathbf{x}_{0(k+1)}^{\mathrm{R}} = \mathbf{x}_{0(k)}^{\mathrm{R}} + \delta\mathbf{x}_0$
- Inner loop is exactly quadratic (e.g. has a unique minimum).
- Inner loop can be simplified (lower res., simplified physics).

Simplification 1: 3D-FGAT

- Three dimensional variational data assimilation with first guess (i.e. $\mathbf{x}^{\mathrm{R}}_{i(k)}$) is computed at the appropriate time.
- Simplification is that $\mathbf{M}_{i(k)} \rightarrow \mathbf{I}$, i.e. $\delta\mathbf{x}_i = \mathbf{M}_{i-1(k)} \ldots \mathbf{M}_{0(k)} \delta\mathbf{x}_0 \rightarrow \delta\mathbf{x}_0$.

$$
\begin{aligned}
J^{\mathrm{3DFGAT}}_{(k)}(\delta\mathbf{x}_0) &= \frac{1}{2}\left(\delta\mathbf{x}_0 - \delta\mathbf{x}^{\mathrm{b}}_0\right)^{\mathrm{T}} \mathbf{B}_0^{-1} (\bullet) + \\
& \frac{1}{2}\sum_{i=0}^{T}\left(\mathbf{y}_i - \mathscr{H}_i(\mathbf{x}^{\mathrm{R}}_{i(k)}) - \mathbf{H}_{i(k)}\delta\mathbf{x}_0\right)^{\mathrm{T}} \mathbf{R}_i^{-1} \left(\bullet\right) \\
& \qquad\qquad\qquad\qquad\qquad\qquad \uparrow
\end{aligned}
$$

Simplification 2: 3DVar

- This has no time dependence within the assimilation window.
- Not used (these days "3D-Var" really means 3D-FGAT).

$$
\begin{aligned}
J^{\mathrm{3DVar}}_{(k)}(\delta \mathbf{x}_0) &= \frac{1}{2}\left(\delta \mathbf{x}_0 - \delta \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}_0^{-1}\left(\bullet\right) + \\
&\frac{1}{2}\sum_{i=0}^{T}\left(\mathbf{y}_i - \mathscr{H}_i(\mathbf{x}^{\mathrm{R}}_{0(k)}) - \mathbf{H}_{i(k)}\delta \mathbf{x}_0\right)^{\mathrm{T}} \mathbf{R}_i^{-1}\left(\bullet\right) \\
&\qquad\qquad\qquad \uparrow
\end{aligned}
$$

Properties of 4DVar

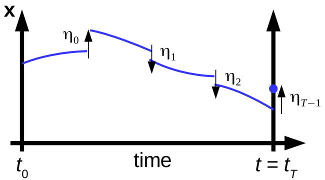
- Observations are treated at the correct time.
- Use of dynamics means that more information can be obtained from observations.
- Covariance $\mathbf{B}_0$ is implicitly evolved, $\mathbf{B}_i = \left(\mathbf{M}_{i-1(k)} \ldots \mathbf{M}_{0(k)}\right) \mathbf{B}_0 \left(\mathbf{M}_{i-1(k)} \ldots \mathbf{M}_{0(k)}\right)^{\mathrm{T}}$.
- In practice development of linear and adjoint models is complex.
  - $\mathscr{M}_i$, $\mathscr{H}_i$, $\mathbf{M}_i$, $\mathbf{H}_i$, $\mathbf{M}_i^{\mathrm{T}}$, and $\mathbf{H}_i^{\mathrm{T}}$ are subroutines, and so 'matrices' are usually not in explicit matrix form.

But note

- Standard 4DVar assumes the model is perfect.
- This can lead to sub-optimalities.
- Weak-constraint 4DVar relaxes this assumption.

Modify evolution equation:

$$\mathbf{x}_{i+1} = \mathscr{M}_i(\mathbf{x}_i) + \eta_i$$
$$\text{where } \eta_i \sim N(0, \mathbf{Q}_i)$$



'State formulation' of WC4DVar

$$J^{\mathrm{wc}}(\mathbf{x}_0, \ldots, \mathbf{x}_T) = J^{\mathrm{b}} + J^{\mathrm{o}} + \frac{1}{2}\sum_{i=0}^{T-1}(\mathbf{x}_{i+1} - \mathscr{M}_i(\mathbf{x}_i))^{\mathrm{T}}\mathbf{Q}_i^{-1}(\bullet)$$

'Error formulation' of WC4DVar

$$J^{\mathrm{wc}}(\mathbf{x}_o, \eta_0 \ldots, \eta_{T-1}) = J^{\mathrm{b}} + J^{\mathrm{o}} + \frac{1}{2}\sum_{i=0}^{T-1}\eta_i^{\mathrm{T}}\mathbf{Q}_i^{-1}\eta_i$$

Implementation of weak constraint 4DVar

- Vector to be determined ('control vector') increases from n in 4DVar to $n + n(T - 1)$ in WC4DVar.
- The model error covariance matrices, $\mathbf{Q}_i$, need to be estimated. How?
- The 'state' formulation (determine $\mathbf{x}_0, \ldots, \mathbf{x}_T$) and the 'error' formulation (determine $\mathbf{x}_0, \eta_0 \ldots, \eta_{T-1}$) are mathematically equivalent, but can behave differently in practice.
- There is an incremental form of WC4DVar.

Summary of 4DVar

- The variational method forms the basis of many operational weather and ocean forecasting systems, including at ECMWF, the Met Office, Météo-France, etc.
- It allows complicated observation operators to be used (e.g. for assimilation of satellite data).
- It has been very successful.
- Incremental (quasi-linear) versions are usually implemented.
- It requires specification of $\mathbf{B}_0$, the background error cov. matrix, and $\mathbf{R}_i$, the observation error cov. matrix.
- 4DVar requires the development of linear and adjoint models – not a simple task!
- Weak constraint formulations require the additional specification of $\mathbf{Q}_i$.

Some challenges ahead

- Methods assume that error cov. matrices are correctly known.
- Representing $\mathbf{B}_0$.
  - Better models of $\mathbf{B}_0$.
  - Flow dependency (e.g. Ensemble-Var or hybrid methods).
- Representing $\mathbf{R}_i$.
  - Allowing for observation error covariances.
- Representing $\mathbf{Q}_i$.
- Numerical conditioning of the problem.
- Application to more complicated systems (e.g. high-resolution models, coupled atmosphere-ocean DA, chemical DA).
- Variational bias correction.
- Moist processes, inc. clouds.
- Effective use on massively parallel computer architectures.

Selected References

- *Original application of 4DVar*: Talagrand O, Courtier P, Variational assimilation of meteorological observations with the adjoint vorticity equation I: Theory, Q. J. R. Meteorol. Soc. 113, 1311–1328 (1987).
- *Excellent tutorial on Var*: Schlatter TW, Variational assimilation of meteorological observations in the lower atmosphere: A tutorial on how it works, J. Atmos. Sol. Terr. Phys. 62, 1057–1070 (2000).
- *Incremental 4DVar*: Courtier P, Thepaut J-N, Hollingsworth A, A strategy for operational implementation of 4D-Var, using an incremental approach, Q. J. R. Meteorol. Soc. 120, 1367–1387 (1994).
- *High-resolution application of 4DVar*: Park SK, Zupanski D, Four-dimensional variational data assimilation for mesoscale and storm scale applications, Meteorol. Atmos. Phys. 82, 173–208 (2003).
- *Met Office 4DVar*: Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne TJ, The Met Office global four-dimensional variational data assimilation scheme, Q. J. R. Meteorol. Soc. 133, 347–362 (2007).
- *Weak constraint 4DVar*: Tremolet Y, Model-error estimation in 4D-Var, Q. J. R. Meteorol. Soc. 133, 1267–1280 (2007).
- *Inner and outer loops*: Lawless, Gratton & Nichols, QJRMS, 2005; Gratton, Lawless & Nichols, SIAM J. on Optimization (2007).
- *More detailed survey of variational methods than can be done in this lecture (plus ensemble-variational, hybrid methods)*: Bannister R.N., A review of operational methods of variational and ensemble-variational data assimilation, Q.J.R. Meteor. Soc. 143, 607-633 (2017).