Variational data assimilation II
Background and methods

Ross Bannister

21–24 June 2022, Univ. of Reading

Quiz

Consider the minimum point of the strong constraint cost function (the 'analysis')

What should we expect $J_{\min}$ to be?

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

- State vector size n, number of observations p.
- Assume **B**, **R** correct, $\mathscr{H}$, $\mathscr{M}$ perfect and linear, Gaussian error statistics.

1. $(n+p)/2$
2. $n/2$
3. $p/2$
4. No theoretical value

Hmmm ...

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.

Making variational DA work – control variable transforms

- $\mathbf{B}_0$ is an $n \times n$ matrix.
  - In operational problems $\mathbf{B}_0$ is too large to store, let alone invert.
  - An unknowable matrix.
- Can model the essential features of $\mathbf{B}_0$ with a change of variable, $\delta\mathbf{x} = \mathbf{U}\delta\mathbf{v}$ (a control variable transform).
  - Hypothesise that the problem is much simpler when posed in terms of $\delta\mathbf{v}$ rather than $\delta\mathbf{x}$.

Cost function in terms of $\delta\mathbf{v}$ → Minimise w.r.t. $\delta\mathbf{v}$ (trivial B-matrix) ↖ iterate ↙ → Convert back to $\delta\mathbf{x}$

- Equivalent to solving original problem w.r.t. $\delta\mathbf{x}$ with $\mathbf{B}_0 = \mathbf{U}\mathbf{U}^{\mathrm{T}}$. See Bannister (2008).

CVTs in more detail

- Illustrate in simplest case: 3DVar with $\mathbf{x}_0^{\mathrm{R}} = \mathbf{x}_0^{\mathrm{b}}$ (and drop time index)

$$J^{\mathrm{3DVar}}(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}\left(\mathbf{y} - \mathscr{H}(\mathbf{x}^{\mathrm{b}}) - \mathbf{H}\delta\mathbf{x}\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\bullet\right).$$

- Make a change of variable: $\delta\mathbf{x} = \mathbf{U}\delta\mathbf{v}$ and assume B-matrix of errors in the $\delta\mathbf{v}$ representation is identity.

$$\mathbf{B} = \left\langle \delta\mathbf{x}\delta\mathbf{x}^{\mathrm{T}} \right\rangle_{\mathrm{b}} = \mathbf{U}\left\langle \delta\mathbf{v}\delta\mathbf{v}^{\mathrm{T}} \right\rangle_{\mathrm{b}}\mathbf{U}^{\mathrm{T}} = \mathbf{U}\mathbf{I}\mathbf{U}^{\mathrm{T}} = \mathbf{U}\mathbf{U}^{\mathrm{T}}.$$

- Substitute into $J^{\mathrm{3DVar}}(\delta\mathbf{x})$: gives a cost function w.r.t. $\delta\mathbf{v}$

$$\begin{aligned}
J^{\mathrm{3DVar}}(\delta\mathbf{v}) &= \frac{1}{2}\delta\mathbf{v}^{\mathrm{T}}\delta\mathbf{v} + \frac{1}{2}\left(\mathbf{y} - \mathscr{H}(\mathbf{x}^{\mathrm{b}}) - \mathbf{H}\mathbf{U}\delta\mathbf{v}\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\bullet\right) \\
\nabla_{\delta\mathbf{v}} J^{\mathrm{3DVar}} &= \delta\mathbf{v} - \mathbf{U}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{y} - \mathscr{H}(\mathbf{x}^{\mathrm{b}}) - \mathbf{H}\mathbf{U}\delta\mathbf{v}\right) \\
\mathbf{x}^{\mathrm{a}} &= \mathbf{x}^{\mathrm{b}} + \mathbf{U}\left(\mathrm{argmin}\left[J^{\mathrm{3DVar}}(\delta\mathbf{v})\right]\right).
\end{aligned}$$

Estimating a B-matrix

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \mathbf{B} = \left\langle \left(\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}}\right)\left(\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}}\right)^{\mathrm{T}} \right\rangle_{\mathrm{b}}$$

$$= \begin{pmatrix} \left\langle (x_1^{\mathrm{b}} - x_1^{\mathrm{t}})^2 \right\rangle_{\mathrm{b}} & \cdots & \left\langle (x_1^{\mathrm{b}} - x_1^{\mathrm{t}})(x_n^{\mathrm{b}} - x_n^{\mathrm{t}}) \right\rangle_{\mathrm{b}} \\ \vdots & \ddots & \vdots \\ \left\langle (x_n^{\mathrm{b}} - x_n^{\mathrm{t}})(x_1^{\mathrm{b}} - x_1^{\mathrm{t}}) \right\rangle_{\mathrm{b}} & \cdots & \left\langle (x_n^{\mathrm{b}} - x_n^{\mathrm{t}})^2 \right\rangle_{\mathrm{b}} \end{pmatrix}$$

$\langle \bullet \rangle_{\mathrm{b}}$: average over population of possible backgrounds.

Problem

The 'truth', $\mathbf{x}^{\mathrm{t}}$, that appears in the definition of error, $\varepsilon = \mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}}$, is unknowable, so need a proxy for this quantity.

Approaches to estimating a B-matrix (1)

"Canadian quick" method

$$\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}} \sim \left(\mathbf{x}^{\mathrm{b}}(t+T) - \mathbf{x}^{\mathrm{b}}(T)\right)/\sqrt{2}.$$

Take population from one long time run, Polavarapu et al. (2005).

Approaches to estimating a B-matrix (2)

Analysis of innovations

Choose a pair of direct/independent obs locations separated by Δr:

$$\begin{aligned} & \left[y_r - x^{\mathrm{b}}_r\right]\left[y_{r+\Delta r} - x^{\mathrm{b}}_{r+\Delta r}\right] = \\ \left[\left\{y_r - x^{\mathrm{t}}_r\right\} - \left\{x^{\mathrm{b}}_r - x^{\mathrm{t}}_r\right\}\right] & \left[\left\{y_{r+\Delta r} - x^{\mathrm{t}}_{r+\Delta r}\right\} - \left\{x^{\mathrm{b}}_{r+\Delta r} - x^{\mathrm{t}}_{r+\Delta r}\right\}\right] = \\ & \left[\varepsilon^y_r - \varepsilon^{\mathrm{b}}_r\right]\left[\varepsilon^y_{r+\Delta r} - \varepsilon^{\mathrm{b}}_{r+\Delta r}\right]. \end{aligned}$$

Take the expectation:

$$\begin{aligned} \left\langle\left[\varepsilon^y_r - \varepsilon^{\mathrm{b}}_r\right]\left[\varepsilon^y_{r+\Delta r} - \varepsilon^{\mathrm{b}}_{r+\Delta r}\right]\right\rangle &= \left\langle \varepsilon^y_r \varepsilon^y_{r+\Delta r}\right\rangle + \left\langle \varepsilon^{\mathrm{b}}_r \varepsilon^{\mathrm{b}}_{r+\Delta r}\right\rangle \\ &= \sigma^2_{\mathrm{o}} \delta_{\Delta r,0} + \sigma^2_{\mathrm{b}} \mathrm{cor}_{\mathrm{b}}(\Delta r). \end{aligned}$$

Above assumes obs and bg errors, as are errors between obs at different locations. Take population from many pairs with same Δr. Rutherford (1972), Hollingsworth and Lönnberg (1986), Järvinen (2001).

Approaches to estimating a B-matrix (3)

National Meteorological Center (NMC) method

Choose pairs of lagged forecasts valid at the same time, e.g.:

$$\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}} \sim \left(\mathbf{x}^{\mathrm{b}}_{48}(t) - \mathbf{x}^{\mathrm{b}}_{24}(t) \right) / \sqrt{2}.$$

Take population from difference at many times. Parrish and Derber (1992), Berre et al. (2006).

Approaches to estimating a B-matrix (4)

Ensemble method

If you have an ensemble that is correctly spread:

$$\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}} \sim \mathbf{x}^{\mathrm{b}}_{(i)} - \langle \mathbf{x}^{\mathrm{b}} \rangle$$

or

$$\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}} \sim \left(\mathbf{x}^{\mathrm{b}}_{(i)} - \mathbf{x}^{\mathrm{b}}_{(j)} \right) / \sqrt{2}.$$

Take population from ensemble members and over many times. Houtekamer et al. (1996), Buehner (2005), Bonavita et al. (2015).

- *Control variable transforms*: Bannister R.N., A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics., Quarterly Journal of the Royal Meteorological Society, 134, 1971–1996 (2008).
- *Canadian 'quick' method*: Polavarapu S., Ren S., Rochon Y., Sankey D., Ek N., Koshyk J., Tarasick D., Data assimilation with the Canadian middle atmosphere model. Atmos.-Ocean 43: 77–100 (2005).
- *Analysis of innovations*: Hollingsworth A., Lönnberg P., The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. Tellus 38A: 111–136 (1986).
- *Analysis of innovations*: Järvinen H., Temporal evolution of innovation and residual statistics in the ECMWF variational data assimilation systems. Tellus 53A: 333–347 (2001).
- *NMC method*: Parrish D.F., Derber J.C., The National Meteorological Center's spectral statistical interpolation analysis system. Mon. Wea. Rev. 120 1747–1763 (1992).
- *NMC method*: Berre L., Ştefănescu S.E., Pereira M.B., The representation of the analysis effect in three error simulation techniques. Tellus 58A 196–209 (2006).
- *Ensemble method*: Houtekamer P.L., Lefaivre L., Derome J., Ritchie H., Mitchell H.L., A system simulation approach to ensemble prediction. Mon. Wea. Rev. 124, 1225–1242 (1996).
- *Ensemble method*: Buehner M., Ensemble derived stationary and flow dependent background error covariances: Evaluation in a quasi-operational NWP setting. Q.J.R. Meteorol. Soc. 131, 1013–1043 (2005).
- *Ensemble method*: Bonavita M., Holm E., Isaksen L., Fisher M., The evolution of the ECMWF hybrid data assimilation system, Q.J.R. Meteor. Soc. (2015).