Hybrid Data Assimilation

Lecturer: Ross Bannister

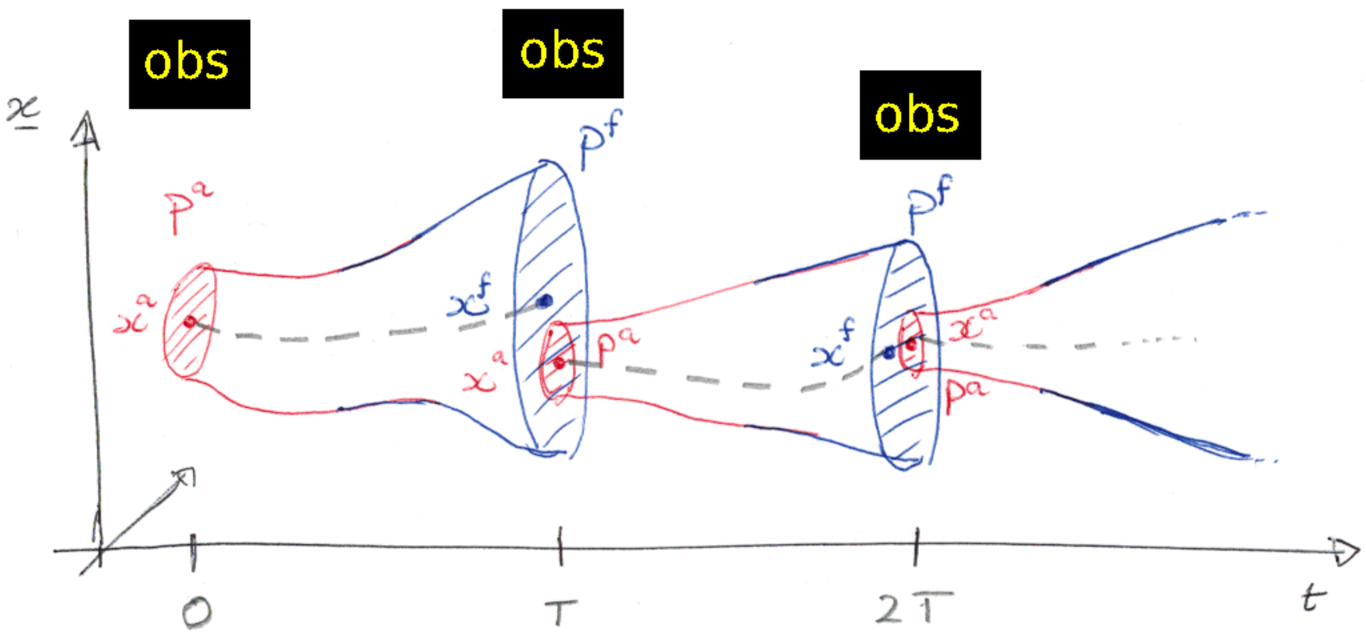
NCEO, Dept. of Meteorology, Univ. of Reading, UK

9-13 June 2025

(flow-dependent variational)

What do we have, and what do we want to improve?

1. Kalman filter



forecast state ... $\mathbf{x}^{\mathrm{f}}_{t+1} \;=\; \boldsymbol{\mathcal{M}}_t\left(\mathbf{x}^{\mathrm{a}}_t\right)$

... and covariance $\mathbf{P}^{\mathrm{f}}_{t+1} \;=\; \mathbf{M}_t \mathbf{P}^{\mathrm{a}}_t \mathbf{M}^{\intercal}_t + \mathbf{Q}_t$

update state ... $\mathbf{x}^{\mathrm{a}}_t \;=\; \mathbf{x}^{\mathrm{f}}_t + \mathbf{K}_t\left(\mathbf{y}_t - \mathbf{h}_t(\mathbf{x}^{\mathrm{f}}_t)\right)$

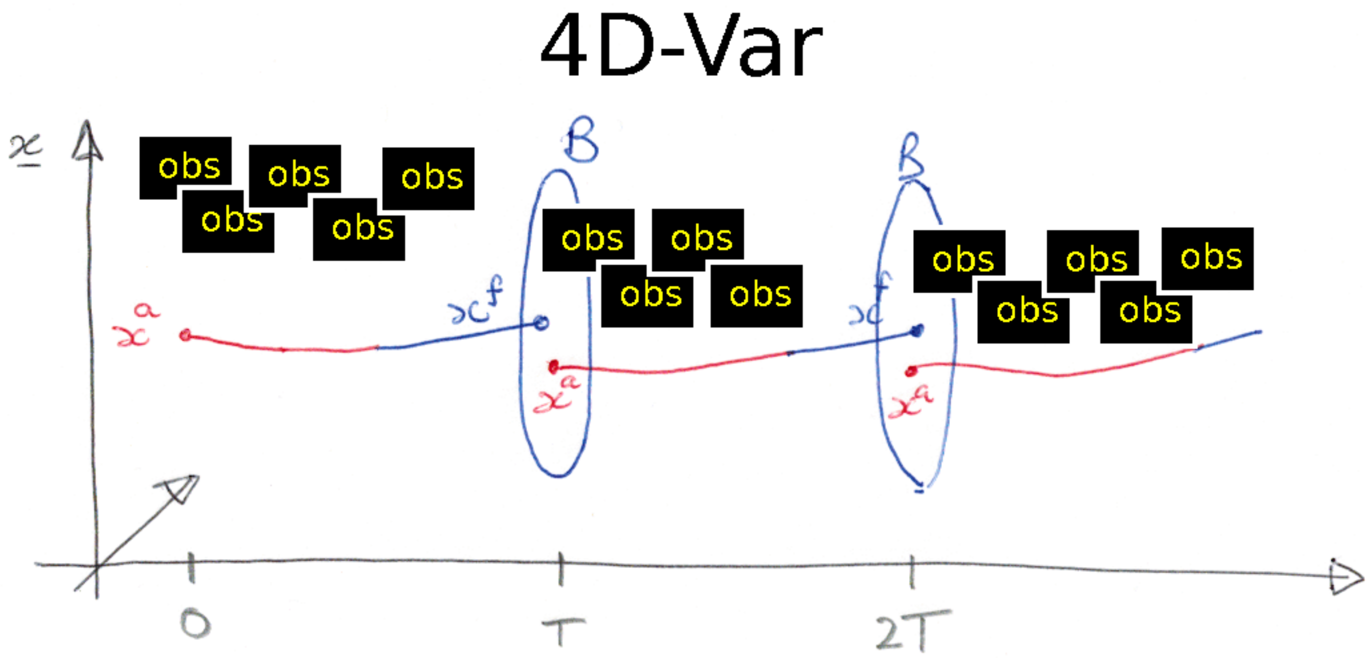
... and cov $\mathbf{P}^{\mathrm{a}}_t \;=\; \left(\mathbf{I} - \mathbf{K}_t \mathbf{H}_t\right)\mathbf{P}^{\mathrm{f}}_t$

where $\mathbf{K}_t \;=\; \mathbf{P}^{\mathrm{f}}_t \mathbf{H}^{\intercal}_t \left(\mathbf{H}_t \mathbf{P}^{\mathrm{f}}_t \mathbf{H}^{\intercal}_t + \mathbf{R}_t\right)^{-1}$

- State and covariances are updated
- Gold standard for linear systems
- Restricted to application to small state spaces, n

What do we have, and what do we want to improve?

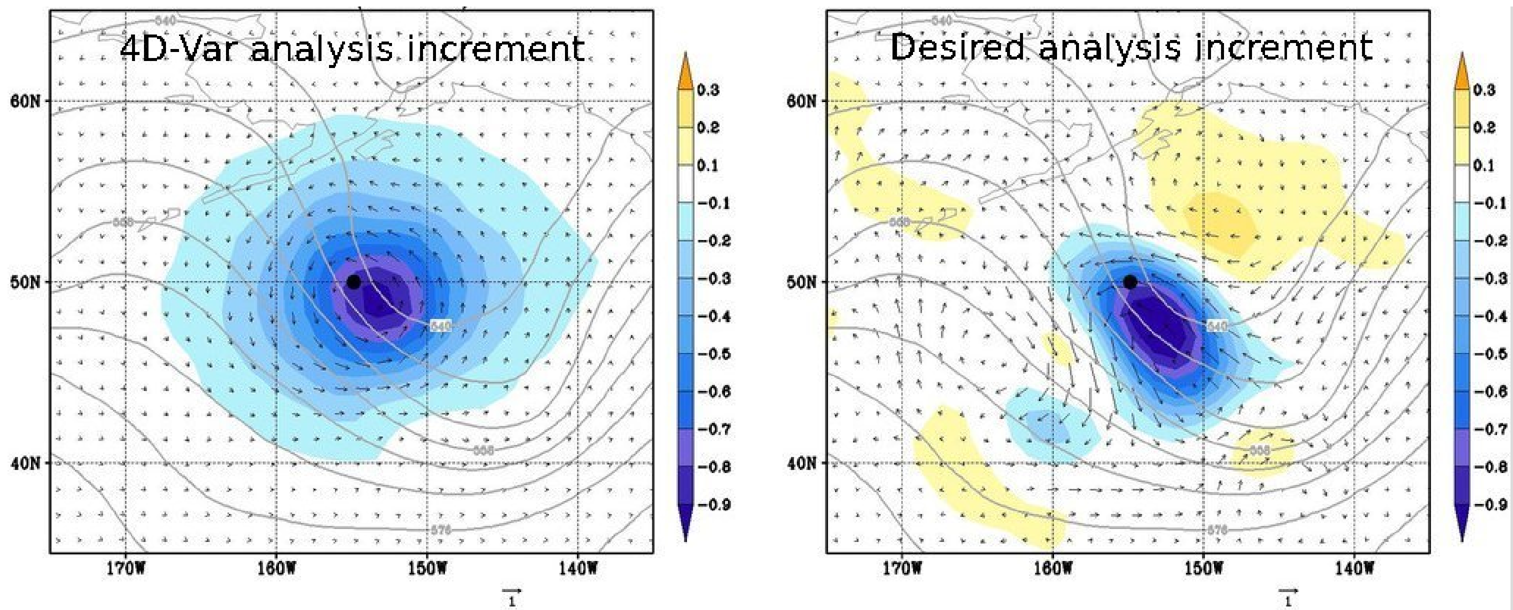
2. Variational data assimilation (e.g. strong constraint inc. 4D-Var)



- Approximation $\mathbf{P}^{\mathrm{f}} \sim \mathbf{B}$ is made
- 4D-Var does *implicitly* evolve the covariances $\mathbf{B}_t = \mathbf{M}_{t-1} \ldots \mathbf{M}_0 \mathbf{B} \mathbf{M}_0^\intercal \ldots \mathbf{M}_{t-1}^\intercal$ for $0 \le t \le T$
- Covariances reset to $\mathbf{B}$ at the start of each cycle
- $\mathbf{P}_t^{\mathrm{a}}$ is not normally available *explicitly*
- Need to have the *tangent linear*, $\mathbf{M}_t$, $\mathbf{H}_t$ and *adjoints*, $\mathbf{M}_t^\intercal$, $\mathbf{H}_t^\intercal$
- Is efficient for application to systems with large state spaces, n

What do we have, and what do we want to improve?

Reminder – why is P^{f} or B so important?



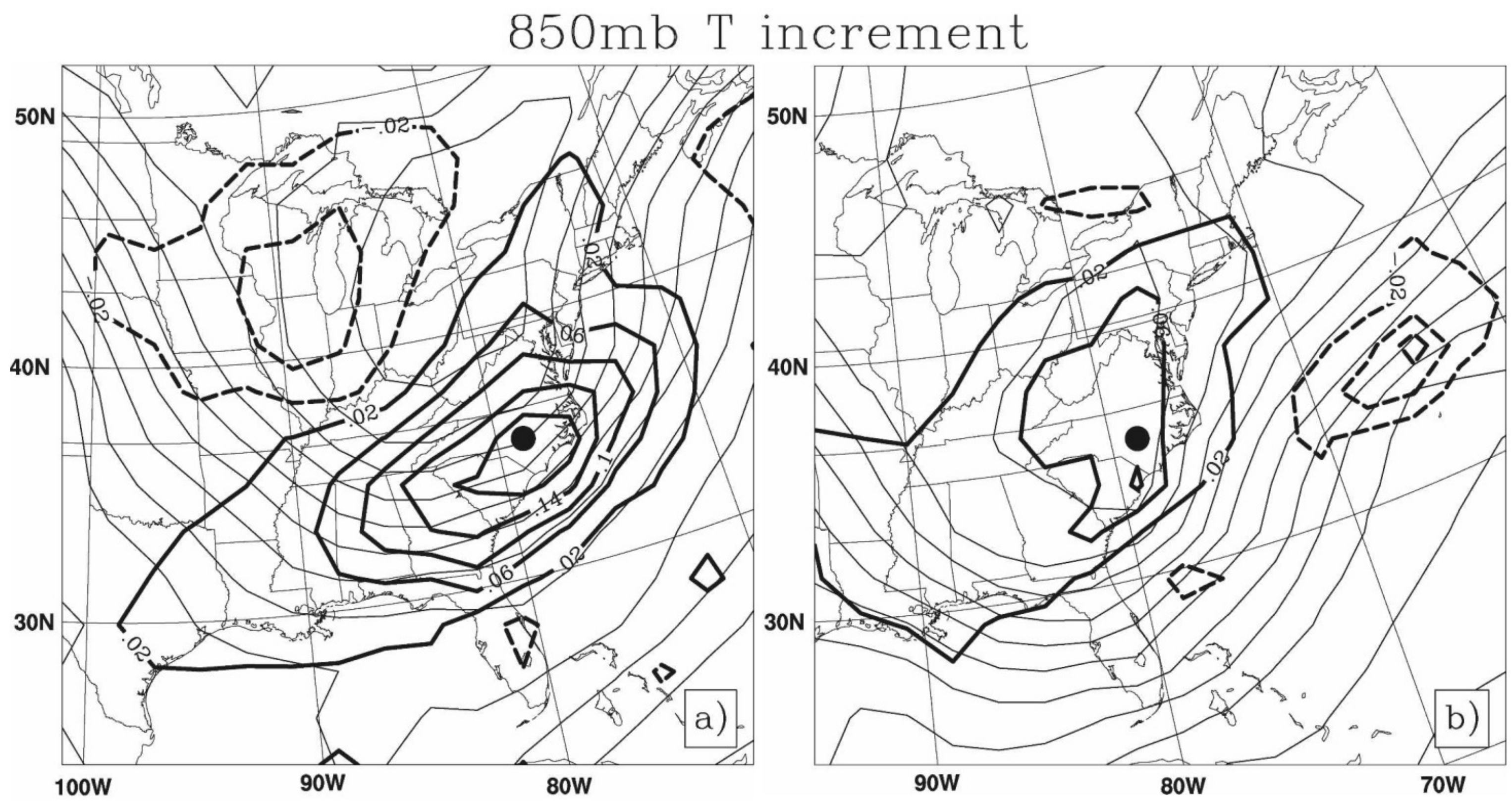
From Kleist and Ide, MWR 143, pp 452-470 (2015)

Colours: analysis increments of T, arrows: analysis increments of (u, v), contours: background geopotential height. All data are at 500 hPa [8].

$$\text{Analysis:} \quad \mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \begin{pmatrix} \mathbf{B}_{1j} \\ \vdots \\ \mathbf{B}_{jj} \\ \vdots \\ \mathbf{B}_{nj} \end{pmatrix} \frac{\mathbf{y}_1 - \mathbf{x}_j^{\mathrm{b}}}{\mathbf{R}_{11} + \mathbf{B}_{jj}}$$

What do we have, and what do we want to improve?

Reminder – why is P^{f} or B so important (continued)?



Wang et al., MWR 136, 5116-5131 (2008)

Thick contours: temperature increments after assimilating a single temperature ob. Thin contours: background temperature [9].

(a) 0000 UTC 14 Jan 2003, (b) 0000 UTC 24 Jan 2003

Important aside: control variable transforms to model the $\mathbb{B}$-matrix in Var

$$\boxed{\delta\mathbf{x}_0 = \mathbf{U}\mathbf{v}_\mathrm{B}}$$

$$\left.\begin{array}{r} \text{if } \langle \delta\mathbf{x}_0 \delta\mathbf{x}_0^\intercal \rangle_\mathrm{f} = \mathbf{B}_0 \\ \text{and } \langle \mathbf{v}_\mathrm{B} \mathbf{v}_\mathrm{B}^\intercal \rangle_\mathrm{f} = \mathbf{I} \end{array}\right\} \quad \text{then} \quad \begin{aligned} \langle \delta\mathbf{x}_0 \delta\mathbf{x}_0^\intercal \rangle_\mathrm{f} &= \langle \mathbf{U}\mathbf{v}_\mathrm{B}\mathbf{v}_\mathrm{B}^\intercal \mathbf{U}^\intercal \rangle_\mathrm{f} \\ &= \mathbf{U} \langle \mathbf{v}_\mathrm{B}\mathbf{v}_\mathrm{B}^\intercal \rangle_\mathrm{f} \mathbf{U}^\intercal \\ &= \mathbf{U}\mathbf{U}^\intercal \end{aligned}$$

- Incremental 4D-Var cost function (minimised with respect to $\delta\mathbf{x}_0$):

$$J(\delta\mathbf{x}_0) = \frac{1}{2}\delta\mathbf{x}_0^\intercal \mathbf{B}_0^{-1} \delta\mathbf{x}_0 + \frac{1}{2}\sum_{t=0}^{T} \left(\mathbf{y}_t - \boldsymbol{\mathcal{H}}_t[\boldsymbol{\mathcal{M}}_{0\to t}(\mathbf{x}_t^\mathrm{b})] - \mathbf{H}_t \mathbf{M}_{0\to t}\delta\mathbf{x}_0\right)^\intercal \mathbf{R}_t^{-1} \left(\bullet\right)$$

- Minimise the variational cost function with respect to $\mathbf{v}_\mathrm{B}$ instead of with respect to $\delta\mathbf{x}_0$:

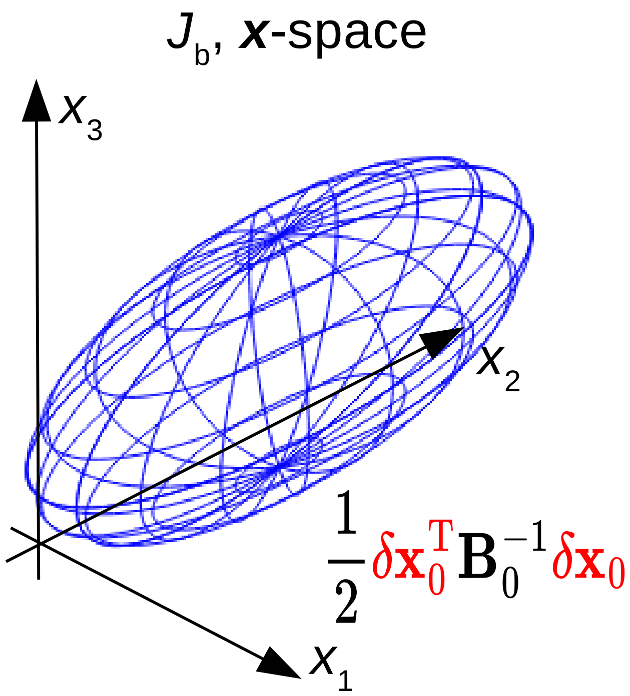
$$\text{e.g. } J(\mathbf{v}_\mathrm{B}) = \frac{1}{2}\mathbf{v}_\mathrm{B}^\intercal \mathbf{v}_\mathrm{B} + \frac{1}{2}\sum_{t=0}^{T} \left(\mathbf{y}_t - \boldsymbol{\mathcal{H}}_t[\boldsymbol{\mathcal{M}}_{0\to t}(\mathbf{x}_t^\mathrm{b})] - \mathbf{H}_t \mathbf{M}_{0\to t}\mathbf{U}\mathbf{v}_\mathrm{B}\right)^\intercal \mathbf{R}^{-1} \left(\bullet\right)$$

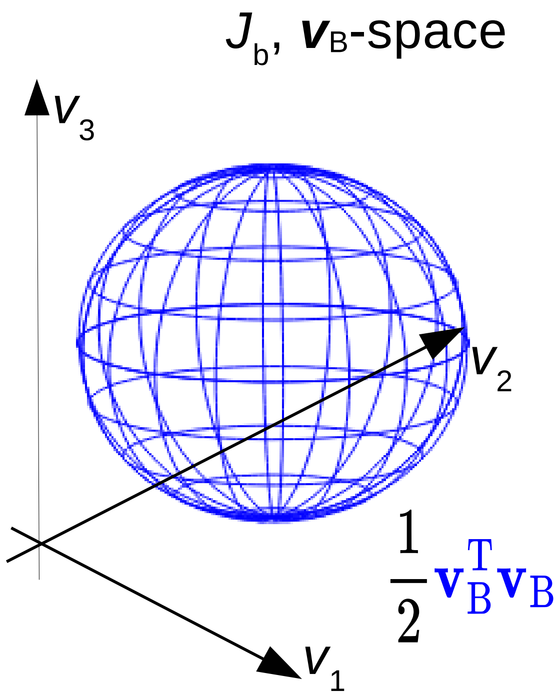
- Analysis is: $\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{U}\left(\underset{\mathbf{v}_{\mathrm{B}}}{\operatorname{argmin}} J(\mathbf{v}_{\mathrm{B}})\right)$

$$\boxed{\mathbf{B}_0 = \mathbf{U}\mathbf{U}^{\intercal}}$$

- Equivalent to minimising original incremental cost function with

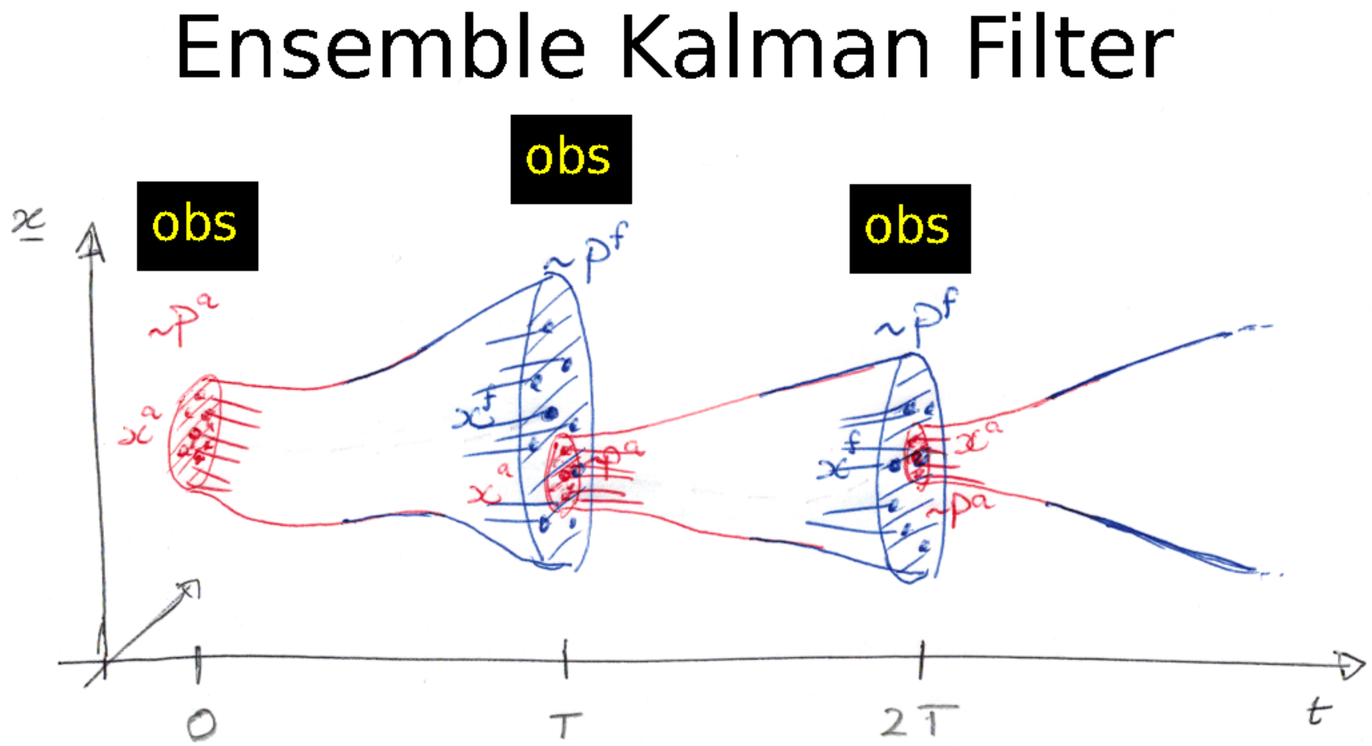
- $\mathbf{U}\mathbf{U}^{\intercal}$ is the *implied covariance*, $\mathbf{U} = \mathbf{B}_0^{1/2}$
- $J(\mathbf{v}_{\mathrm{B}})$ is numerically better conditioned than $J(\delta\mathbf{x})$





What do we have, and what do we want to improve?

3. Ensemble data assimilation



mean: $\overline{\mathbf{x}_t^{\mathrm{f}}} \approx \frac{1}{N}\sum_{\ell=1}^{N}\mathbf{x}_t^{\mathrm{f}(\ell)}$

perturbation: $\mathbf{x}_t^{\mathrm{f}(\ell)} - \overline{\mathbf{x}_t^{\mathrm{f}}}$

covariance: $\left[\mathbf{P}_t^{\mathrm{f}}\right]_{ij} \approx \frac{1}{N-1}\sum_{\ell=1}^{N}\left(\left[\mathbf{x}_t^{\mathrm{f}(\ell)}\right]_i - \overline{\left[\mathbf{x}_t^{\mathrm{f}}\right]_i}\right)\left(\left[\mathbf{x}_t^{\mathrm{f}(\ell)}\right]_j - \overline{\left[\mathbf{x}_t^{\mathrm{f}}\right]_j}\right)$

$$\mathbf{P}_t^{\mathrm{f}} \approx \frac{1}{N-1}\sum_{\ell=1}^{N}\left(\mathbf{x}_t^{\mathrm{f}(\ell)} - \overline{\mathbf{x}_t^{\mathrm{f}}}\right)\left(\mathbf{x}_t^{\mathrm{f}(\ell)} - \overline{\mathbf{x}_t^{\mathrm{f}}}\right)^{\mathsf{T}}$$

matrix of ens perts: $\mathbf{X}_t'^{\mathrm{f}} = \frac{1}{\sqrt{N-1}}\begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ \mathbf{x}_t^{\mathrm{f}(1)} - \overline{\mathbf{x}_t^{\mathrm{f}}} & \cdots & \mathbf{x}_t^{\mathrm{f}(\ell)} - \overline{\mathbf{x}_t^{\mathrm{f}}} & \cdots & \mathbf{x}_t^{\mathrm{f}(N)} - \overline{\mathbf{x}_t^{\mathrm{f}}} \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix}$

$$\left[\mathbf{X}_t'^{\mathrm{f}}\right]_{i\ell} = \frac{\left[\mathbf{x}_t^{\mathrm{f}(\ell)}\right]_i - \left[\overline{\mathbf{x}_t^{\mathrm{f}}}\right]_i}{\sqrt{N-1}}$$

$$\boxed{\mathbf{P}_t^{\mathrm{f}} \approx \mathbf{X}_t'^{\mathrm{f}}\mathbf{X}_t'^{\mathrm{f}\mathsf{T}}}$$

What do we have, and what do we want to improve?

3. Ensemble data assimilation (continued)

- Comes in many flavours (stochastic, square-root, etc.), all derived from the Kalman update equation
- Ensemble of states are updated: flow-dependent covariances $\mathbf{P}^{\mathrm{f}}$ and $\mathbf{P}^{\mathrm{a}}$ are implied from the ensemble are approximated (and are not computed explicitly)
- Is efficient for application to systems with large state spaces, n
- $\mathbf{P}^{\mathrm{f}}$ and $\mathbf{P}^{\mathrm{a}}$ are rank deficient (hence sampling error present)

How do we combine the properties of ‘flow-dependentness’ of ensemble methods with the ‘full-rankness’ of variational methods?

Possible definitions of a hybrid data assimilation method

A. A purely variational scheme, with a separate EnKF system to give extra estimates of analysis error

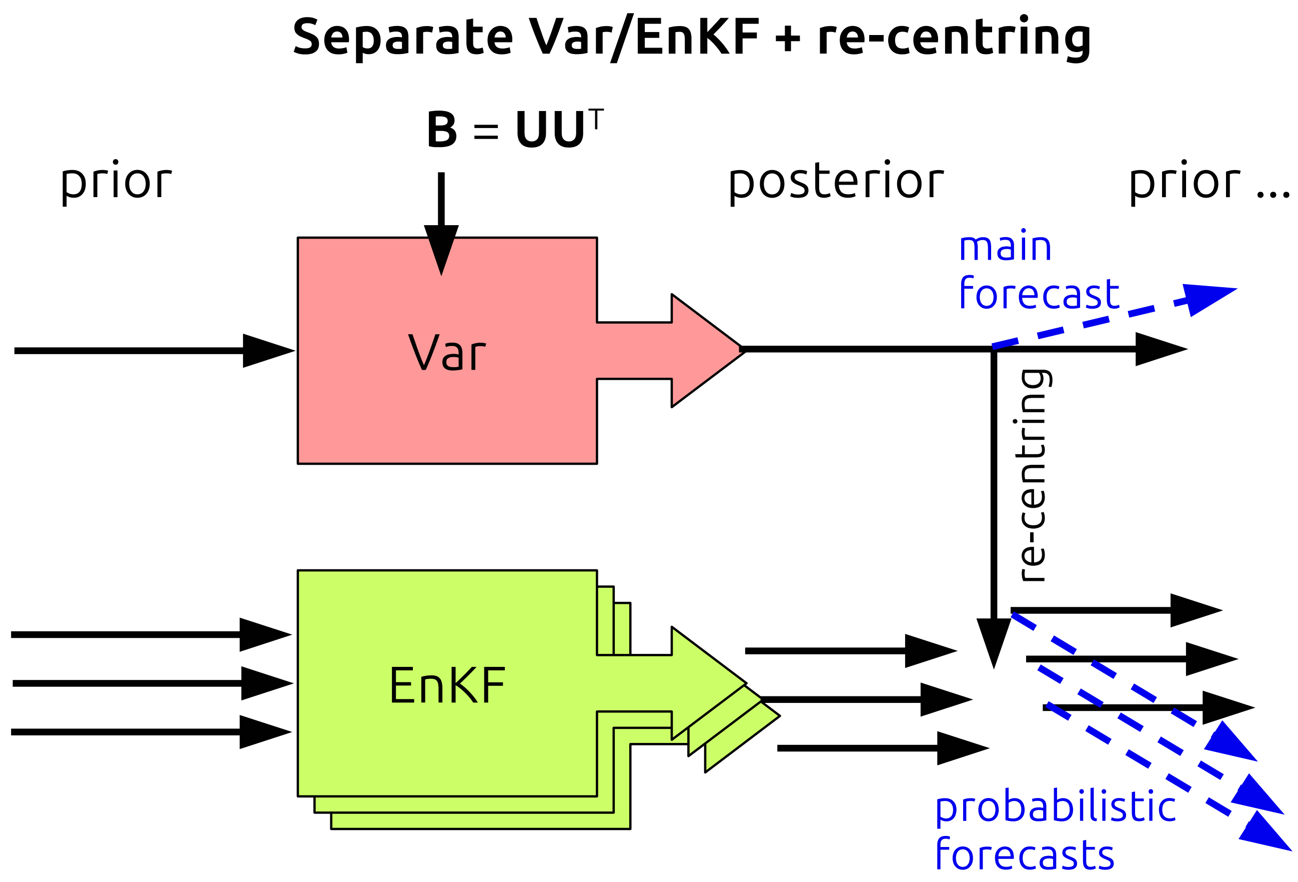
B. A purely variational scheme that re-calibrates the $\mathbf{B}$-matrix model ($\mathbf{U}\mathbf{U}^\intercal$) using ensemble information

C. A purely variational scheme that uses an ensemble to imply the flow-dependent $\mathbf{B}$-matrix ($\mathbf{X}_t'^{\mathrm{f}}\mathbf{X}_t'^{\mathrm{f}\intercal}$)

D. A method that combines the $\mathbf{B}$-matrix of Var with the $\mathbf{P}^{\mathrm{f}}$-matrix of the EnKF

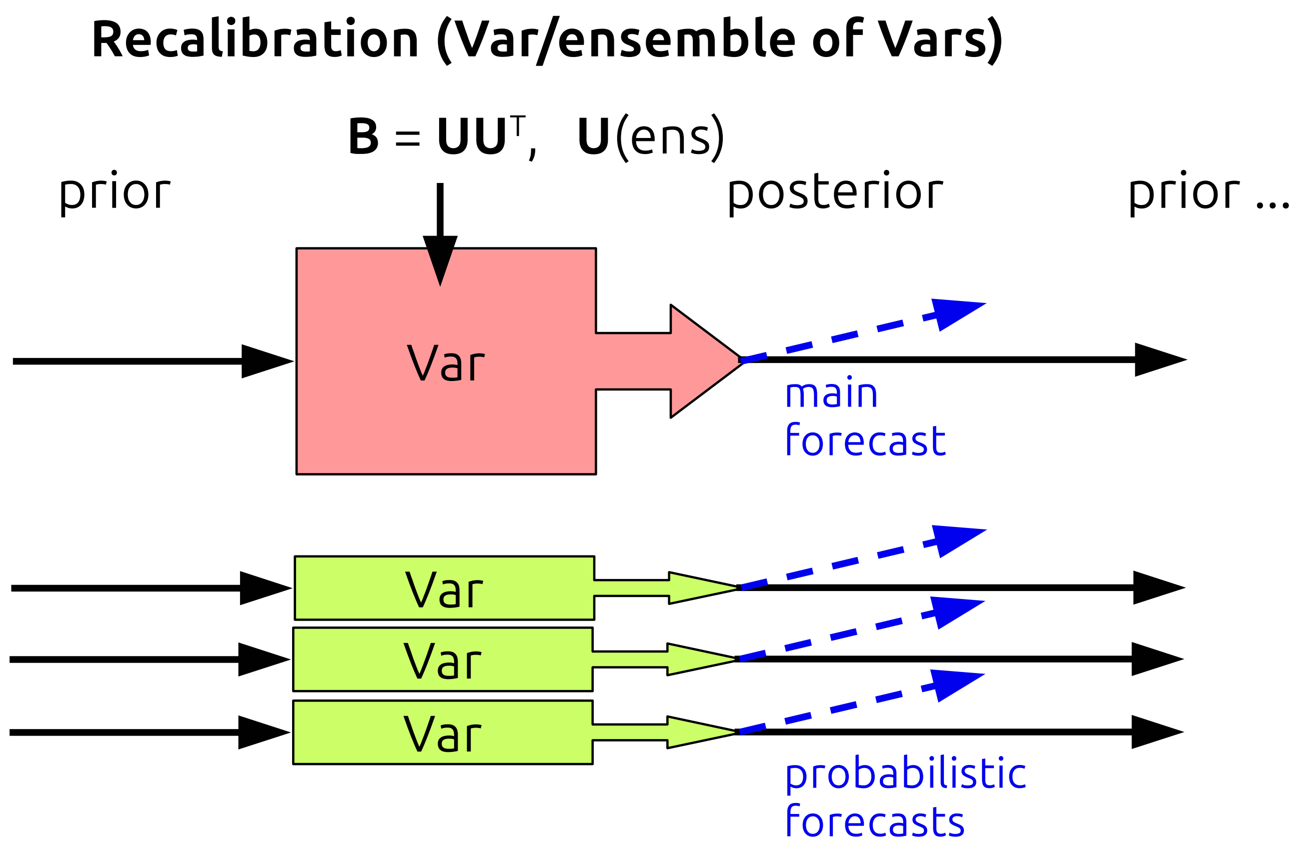
E. A method that delegates part of the covariance estimate to a machine learning algorithm

A. A purely variational scheme, with a separate EnKF system to give estimates of analysis error



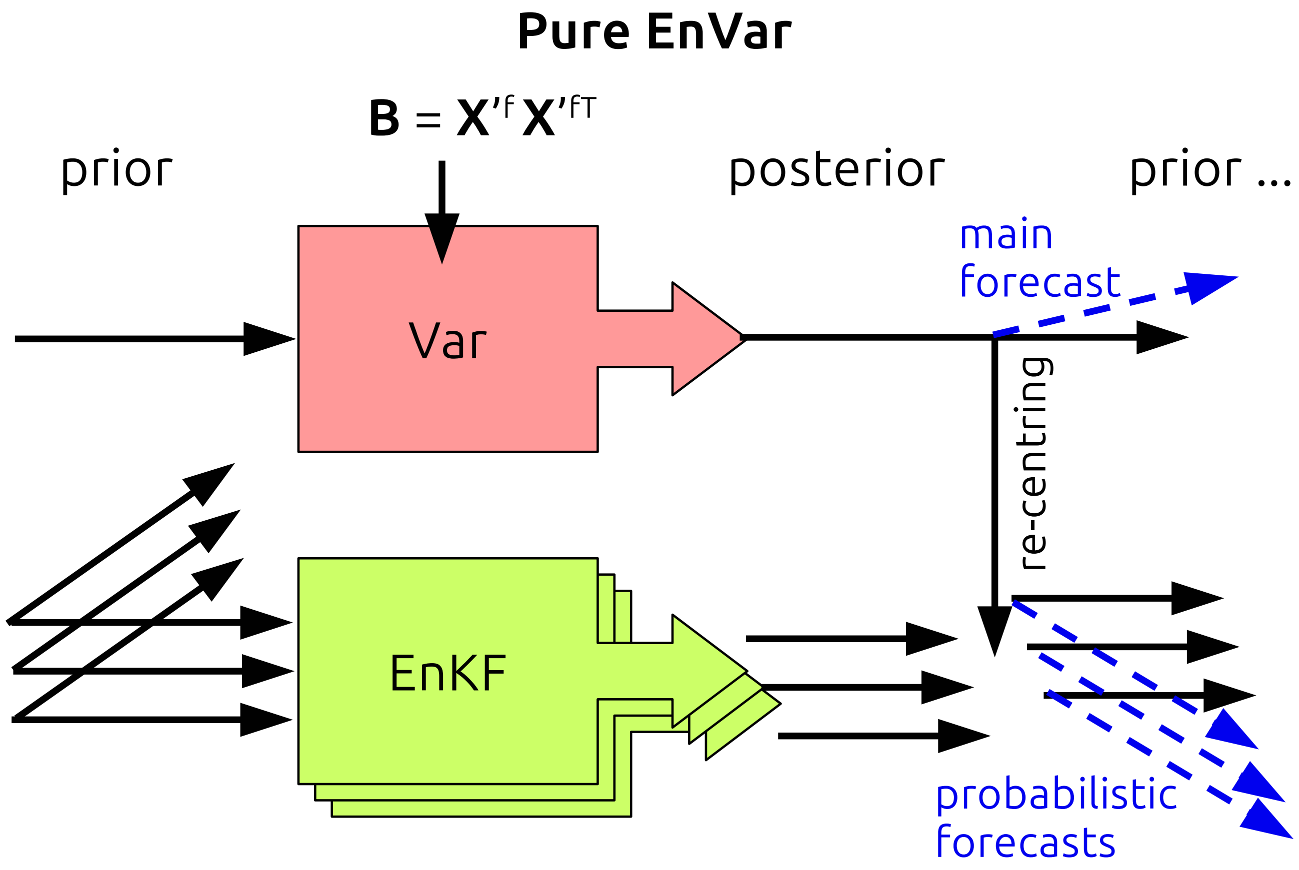
E.g. Bowler et al. (2008) [3]

B. A purely variational scheme that re-calibrates the B-matrix model (UU^T) using ensemble information



E.g. Bonavita et al. (2016) [2]

C. A purely variational scheme that uses an ensemble to imply the flow-dependent B-matrix ($\mathbf{X}_t'^{\mathrm{f}}\mathbf{X}_t'^{\mathrm{f}\mathsf{T}}$)



Flavours En4DVar and 4DEnVar

See references in Bannister (2017) [1]

C(i) Pure ensemble-variational methods (EnVar)

En4DVar, no localisation

$$\delta \mathbf{x}_0 = \mathbf{X}'^{\mathrm{f}}_0 \mathbf{v}_{\mathrm{ens}} \qquad\qquad \mathbf{v}_{\mathrm{ens}} \in \mathbb{R}^N$$

Start with the incremental formulation of 4DVar (reference state is the background)

$$J^{\mathrm{4DVar}}(\delta \mathbf{x}_0) = \frac{1}{2}\delta \mathbf{x}_0^{\mathsf{T}} \mathbf{B}_0^{-1} \delta \mathbf{x}_0 + \frac{1}{2}\sum_{t=0}^{T} \left(\mathbf{y}_t - \boldsymbol{\mathcal{H}}_t(\mathbf{x}_t^{\mathrm{b}}) - \mathbf{H}_t \mathbf{M}_{t-1} \mathbf{M}_{t-2} \ldots \mathbf{M}_0 \delta \mathbf{x}_0\right)^{\mathsf{T}} \mathbf{R}_t^{-1} (\bullet)$$

In control variable $\mathbf{v}_{\mathrm{ens}} \in \mathbb{R}^N$ space

$$J^{\mathrm{En4DVar}}(\mathbf{v}_{\mathrm{ens}}) = \frac{1}{2}\mathbf{v}_{\mathrm{ens}}^{\mathsf{T}} \mathbf{v}_{\mathrm{ens}} + \frac{1}{2}\sum_{t=0}^{T} \left(\mathbf{y}_t - \boldsymbol{\mathcal{H}}_t(\mathbf{x}_t^{\mathrm{b}}) - \mathbf{H}_t \mathbf{M}_{t-1} \mathbf{M}_{t-2} \ldots \mathbf{M}_0 \mathbf{X}'^{\mathrm{f}}_0 \mathbf{v}_{\mathrm{ens}}\right)^{\mathsf{T}} \mathbf{R}_t^{-1} (\bullet)$$

$$\mathbf{x}_0^{\mathrm{a}} = \mathbf{x}_0^{\mathrm{b}} + \mathbf{X}'^{\mathrm{f}}_0 \operatorname{argmin}\left(J^{\mathrm{En4DVar}}(\mathbf{v}_{\mathrm{ens}})\right)$$

- Still need the tangent linear model (and adjoint for the gradient w.r.t. $\mathbf{v}_{\mathrm{ens}}$). This is similar to ordinary 4DVar, but with a different control variable transform.

C(ii) Pure ensemble-variational methods (EnVar)

4DEnVar, no localisation

Start with the En4DVar cost function:

$$J^{\text{En4DVar}}(\mathbf{v}_{\text{ens}}) = \frac{1}{2}\mathbf{v}_{\text{ens}}^{\mathsf{T}}\mathbf{v}_{\text{ens}} + \frac{1}{2}\sum_{t=0}^{T}\left(\mathbf{y}_t - \boldsymbol{\mathcal{H}}_t(\mathbf{x}_t^{\text{b}}) - \mathbf{H}_t \underbrace{\overbrace{\mathbf{M}_{t-1}\mathbf{M}_{t-2}\ldots\mathbf{M}_0}^{\mathbf{M}_{0\to t}} \mathbf{X}'^{\text{f}}_0}_{\mathbf{X}'^{\text{f}}_t} \mathbf{v}_{\text{ens}}\right)^{\mathsf{T}} \mathbf{R}_t^{-1}\left(\bullet\right)$$

Consider the ℓth column of $\mathbf{X}_0^{\text{f}}$ (call $\mathbf{x}_0^{\text{f}(\ell)}$) and do a Taylor expansion of $\boldsymbol{\mathcal{M}}_{0\to t}$:

$$\begin{aligned}
\boldsymbol{\mathcal{M}}_{0\to t}\left(\underbrace{\mathbf{x}_0^{\text{f}(\ell)}}_{\substack{\ell\text{th column}\\ \text{of } \mathbf{X}_0^{\text{f}}}}\right) &\approx \boldsymbol{\mathcal{M}}_{0\to t}\left(\overline{\mathbf{x}_0^{\text{f}}}\right) + \underbrace{\mathbf{M}_{0\to t}\left(\underbrace{\mathbf{x}_0^{\text{f}(\ell)} - \overline{\mathbf{x}_0^{\text{f}}}}_{\ell\text{th column of } \mathbf{X}'^{\text{f}}_0}\right)}_{\ell\text{th column of } \mathbf{X}'^{\text{f}}_t} \\
&\approx \overline{\boldsymbol{\mathcal{M}}_{0\to t}\left(\mathbf{x}_0^{\text{f}}\right)} + \mathbf{M}_{0\to t}\left(\mathbf{x}_0^{\text{f}(\ell)} - \overline{\mathbf{x}_0^{\text{f}}}\right)
\end{aligned}$$

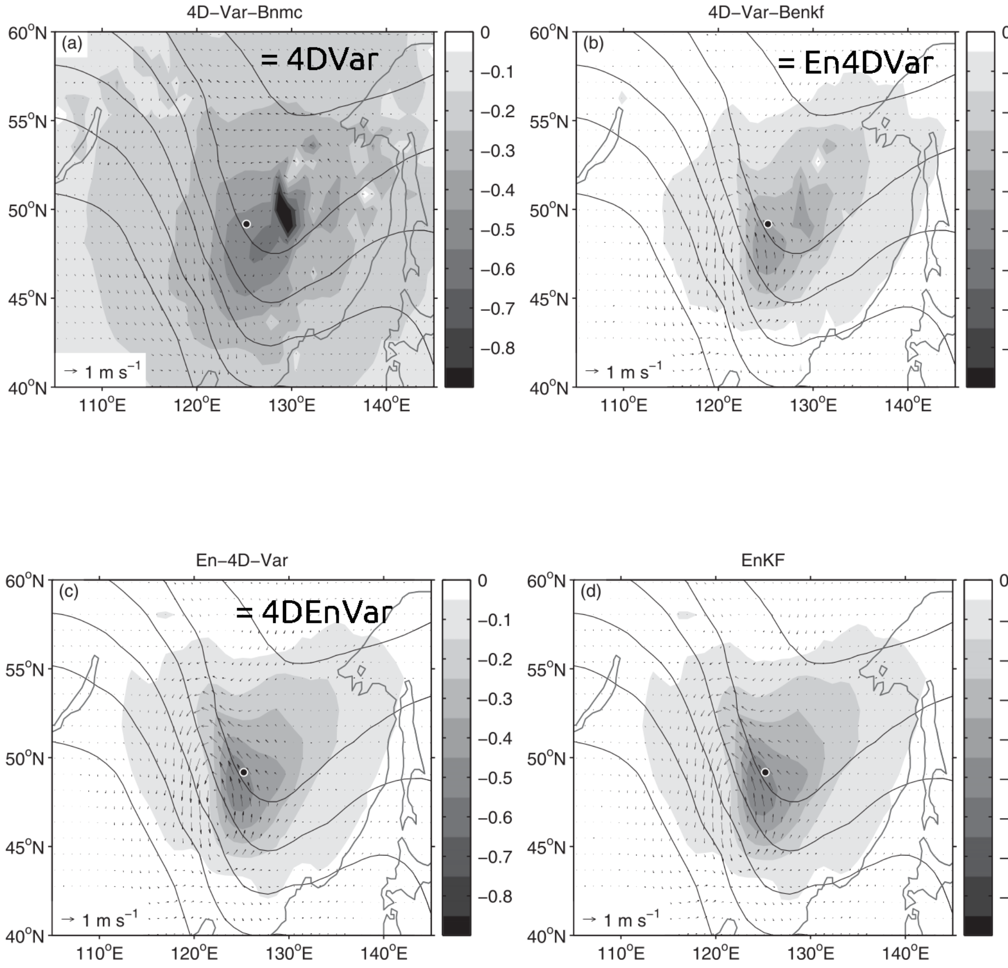
$$\therefore \underbrace{\mathbf{M}_{0\to t}\left(\mathbf{x}_0^{\text{f}(\ell)} - \overline{\mathbf{x}_0^{\text{f}}}\right)}_{\ell\text{th column of } \mathbf{X}'^{\text{f}}_t} \approx \boldsymbol{\mathcal{M}}_{0\to t}\left(\mathbf{x}_0^{\text{f}(\ell)}\right) - \overline{\boldsymbol{\mathcal{M}}_{0\to t}\left(\mathbf{x}_0^{\text{f}}\right)}$$

- Use the above to build columns of $\mathbf{M}_{0\to t}\mathbf{X}'^{\mathrm{f}}_0 \equiv \mathbf{X}'^{\mathrm{f}}_t$ $\qquad$ (N runs of the model).
- This eliminates the need for a TLM (and its adjoint for the gradient).
- The method is called 4DEnVar.

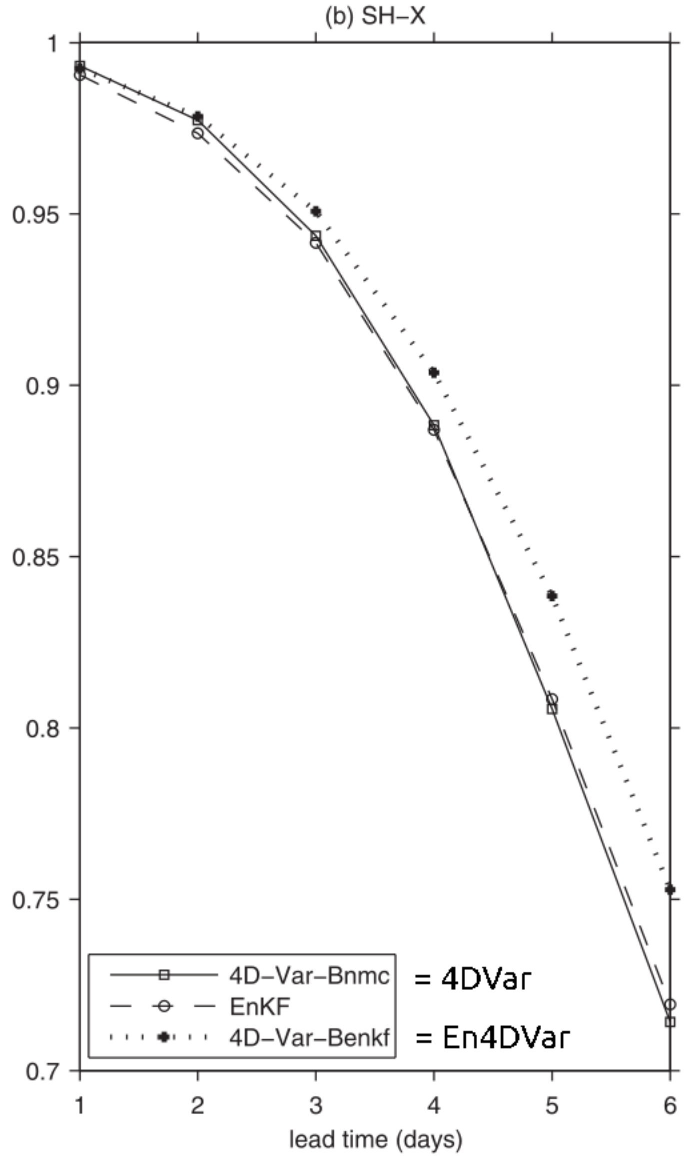
$$J^{\mathrm{4DEnVar}}(\mathbf{v}_{\mathrm{ens}}) = \frac{1}{2}\mathbf{v}_{\mathrm{ens}}^{\mathsf{T}}\mathbf{v}_{\mathrm{ens}} + \frac{1}{2}\sum_{t=0}^{T}\left(\mathbf{y}_t - \boldsymbol{\mathcal{H}}_t(\mathbf{x}_t^{\mathrm{b}}) - \mathbf{H}_t\mathbf{X}'^{\mathrm{f}}_t\mathbf{v}_{\mathrm{ens}}\right)^{\mathsf{T}}\mathbf{R}_t^{-1}\left(\bullet\right)$$

$$\mathbf{x}_t^{\mathrm{a}} = \boldsymbol{\mathcal{M}}_{0\to t}\left(\mathbf{x}_t^{\mathrm{b}}\right) + \mathbf{X}'^{\mathrm{f}}_t\,\mathrm{argmin}\left(J^{\mathrm{4DEnVar}}(\mathbf{v}_{\mathrm{ens}})\right)$$

Single ob experiments and performance [4, 5]

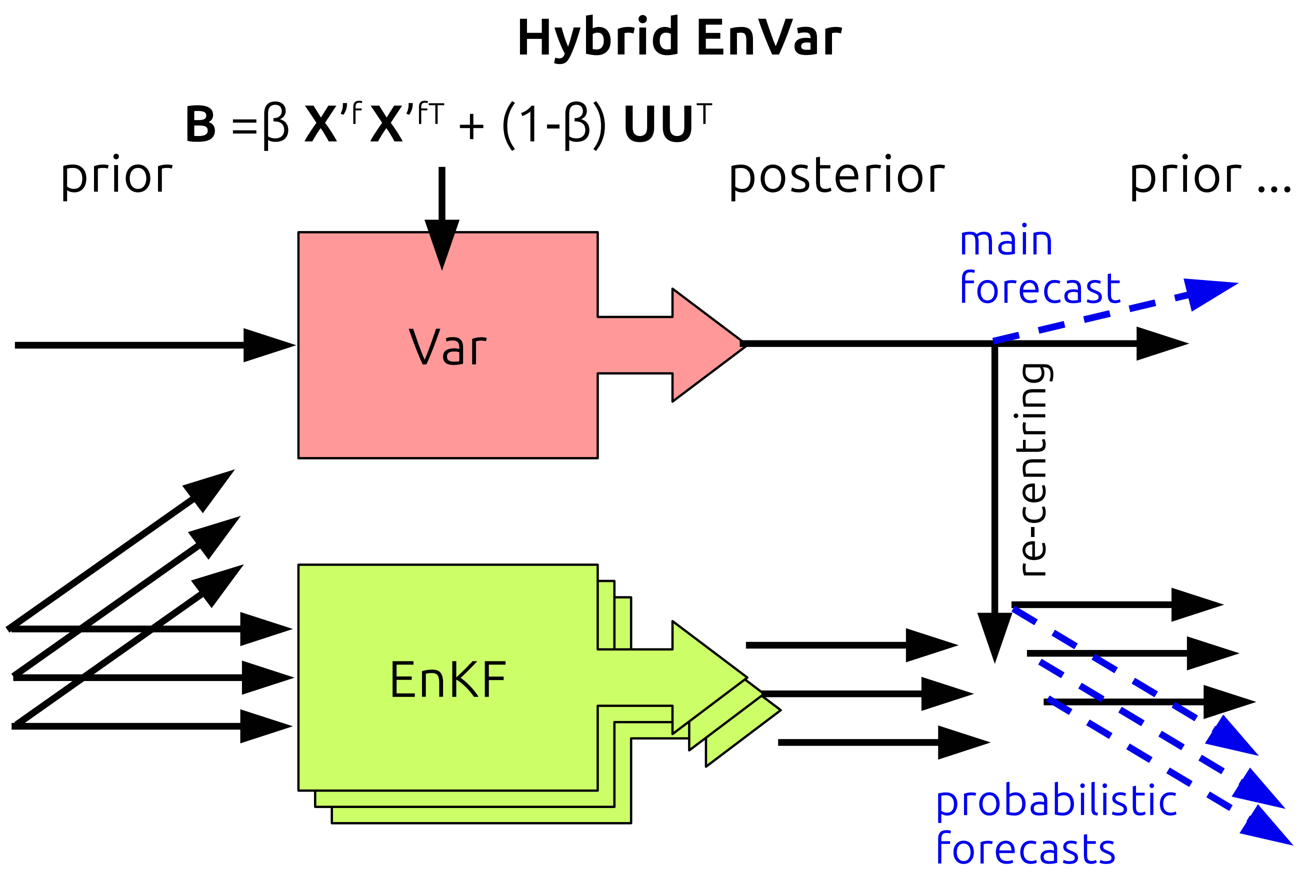


500hPa analysis increments of T (shading) and (u, v) due to T ob $y - H(\mathbf{x}^{\mathrm{b}}) = -1\mathrm{K}$; contours are bg geopotential height.



Southern hemisphere anomaly correlations for 500 hPa geopotential height.

D. A method that combines the B-matrix of Var with the P^{f}-matrix of the EnKF



See references in Bannister (2017) [1]

D. A method that combines the B-matrix of Var with the $\mathrm{P^f}$-matrix of the EnKF

- $\mathbf{P}^{\mathrm{f}}$ is flow-dependent (good), but rank deficient, etc. (bad) and not completely mitigated for with localisation.
- Remember we also have the original $\mathbf{B}$-matrix from traditional variational assimilation:
- $\mathbf{B}$ is not fully flow-dependent (bad), but can be full-rank (good), and can have some useful properties (e.g. produces nearly balanced increments)
- Propose to combine them [7]:

$$\mathbf{P}_{\mathrm{h}} = (1-\beta)\mathbf{B} + \beta\mathbf{P}^{\mathrm{f}} \qquad 0 \leq \beta \leq 1$$

- Trick is to represent this as a CVT

$$\delta\mathbf{x} = \mathbf{U}_{\mathrm{hy}} \, \mathbf{v}_{\mathrm{hy}}$$

$$\underbrace{\begin{pmatrix}\sqrt{1-\beta}\,\mathbf{U} & \sqrt{\beta}\,\mathbf{X}'^{\mathrm{f}}\end{pmatrix}}_{\text{hybrid CVT}} \underbrace{\begin{pmatrix}\mathbf{v}_{\mathrm{B}} \\ \mathbf{v}_{\mathrm{ens}}\end{pmatrix}}_{\substack{\text{hybrid}\\ \text{control}\\ \text{variable}}} \qquad \mathbf{v}_{\mathrm{h}} \in \mathbb{R}^{n+N}$$

- Implied covariances

$$\begin{pmatrix}\sqrt{1-\beta}\mathbf{U} & \sqrt{\beta}\mathbf{X}'^{\mathrm{f}}\end{pmatrix}\begin{pmatrix}\sqrt{1-\beta}\mathbf{U}^{\intercal} \\ \sqrt{\beta}\mathbf{X}'^{\mathrm{f}\intercal}\end{pmatrix} = (1-\beta)\mathbf{U}\mathbf{U}^{\mathrm{T}} + \beta\mathbf{X}'^{\mathrm{f}}\mathbf{X}'^{\mathrm{f}\intercal}$$

$$J^{\mathrm{Hybrid-En4DVar}}(\mathbf{v}_{\mathrm{hy}}) = \frac{1}{2}\mathbf{v}_{\mathrm{hy}}^{\intercal}\mathbf{v}_{\mathrm{hy}} + \frac{1}{2}\sum_{t=0}^{T}\left(\mathbf{y}_t - \boldsymbol{\mathcal{H}}_t(\mathbf{x}_t^{\mathrm{b}}) - \mathbf{H}_t\mathbf{M}_{t-1}\mathbf{M}_{t-2}\dots\mathbf{M}_0\mathbf{U}_{\mathrm{hy}}\mathbf{v}_{\mathrm{hy}}\right)^{\intercal}\mathbf{R}_t^{-1}\left(\bullet\right)$$

$$\mathbf{x}_0^{\mathrm{a}} = \mathbf{x}_0^{\mathrm{b}} + \mathbf{U}_{\mathrm{hy}}\,\mathrm{argmin}\left(J^{\mathrm{Hybrid-En4DVar}}(\mathbf{v}_{\mathrm{hy}})\right)$$

A note on localisation

- The above schemes do not use localisation
- This may be OK for some systems, e.g. [6] uses 4DEnVar in an ecosystem carbon model
- The EnVar schemes may be modified to include localisation
- A more complex control variable transform
- Details in [1]

Bibliography

[1] RN Bannister. A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143:607–633, 2017.

[2] Massimo Bonavita, Elias Hólm, Lars Isaksen, and Mike Fisher. The evolution of the ECMWF hybrid data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 142(694):287–303, 2016.

[3] Neill E Bowler, Alberto Arribas, Kenneth R Mylne, Kelvyn B Robertson, and Sarah E Beare. The MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 134(632):703–722, 2008.

[4] Mark Buehner, PL Houtekamer, Cecilien Charette, Herschel L Mitchell, and Bin He. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Monthly Weather Review*, 138(5):1550–1566, 2010.

[5] Mark Buehner, PL Houtekamer, Cecilien Charette, Herschel L Mitchell, and Bin He. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Monthly Weather Review*, 138(5):1567–1586, 2010.

[6] Natalie Douglas, Tristan Quaife, and Ross Bannister. Exploring a hybrid ensemble–variational data assimilation technique (4DEnVar) with a simple ecosystem carbon model. *Environmental Modelling & Software*, page 106361, 2025.

[7] Thomas M Hamill and Chris Snyder. A hybrid ensemble Kalman filter-3D variational analysis scheme. *Monthly Weather Review*, 128(8):2905–2919, 2000.

[8] Daryl T Kleist and Kayo Ide. An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Monthly Weather Review*, 143(2):452–470, 2015.

[9] Xuguang Wang, Dale M Barker, Chris Snyder, and Thomas M Hamill. A hybrid ETKF-3DVar data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Monthly Weather Review*, 136(12):5116–5131, 2008.