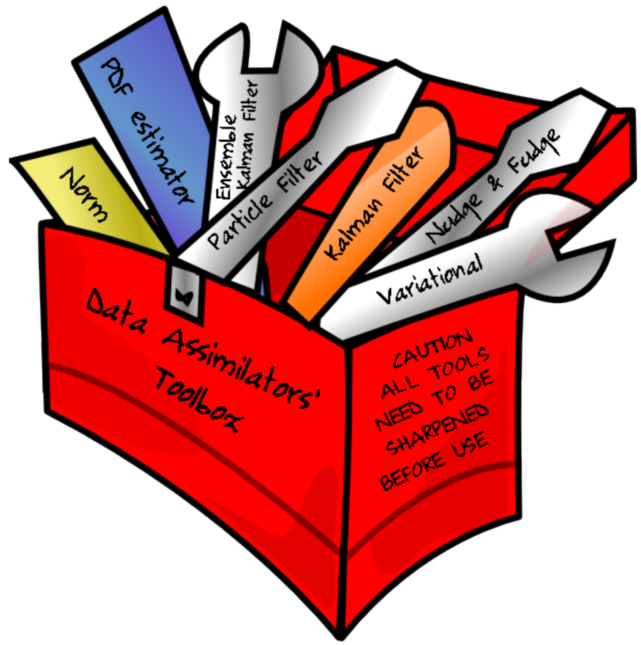
Covariance Matrices in Variational DA

Ross Bannister, r.n.bannister@reading.ac.uk

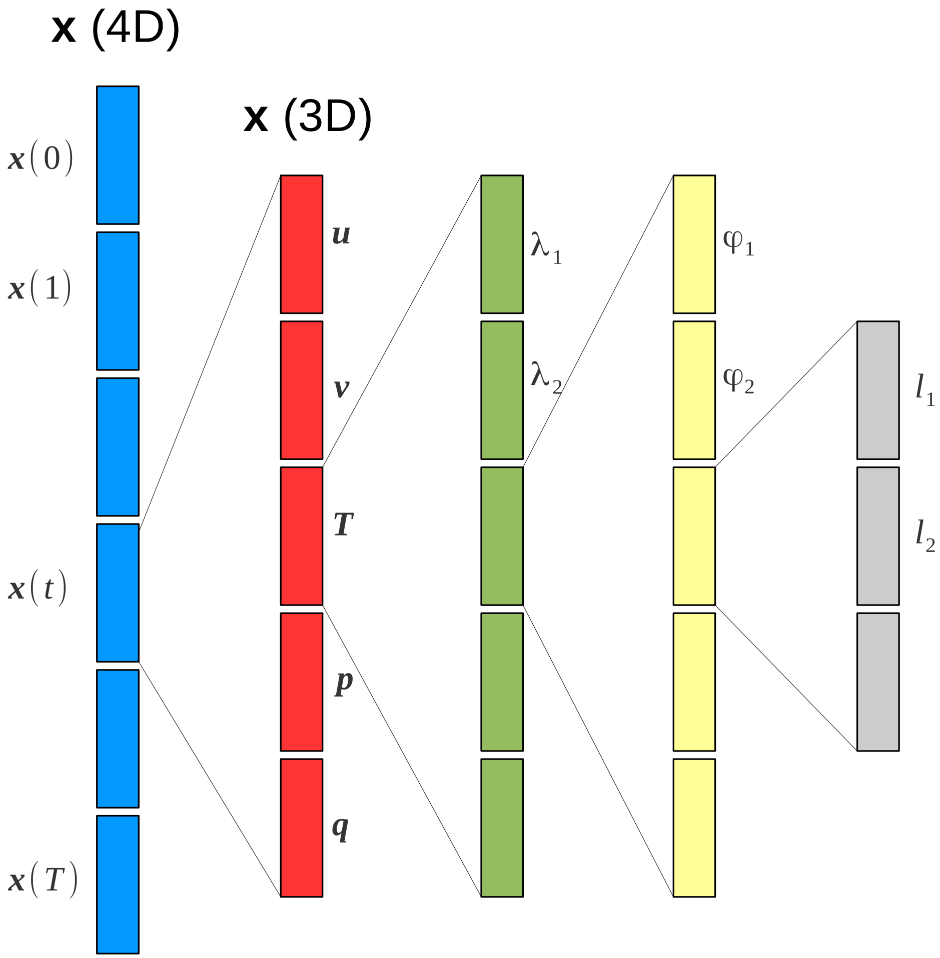


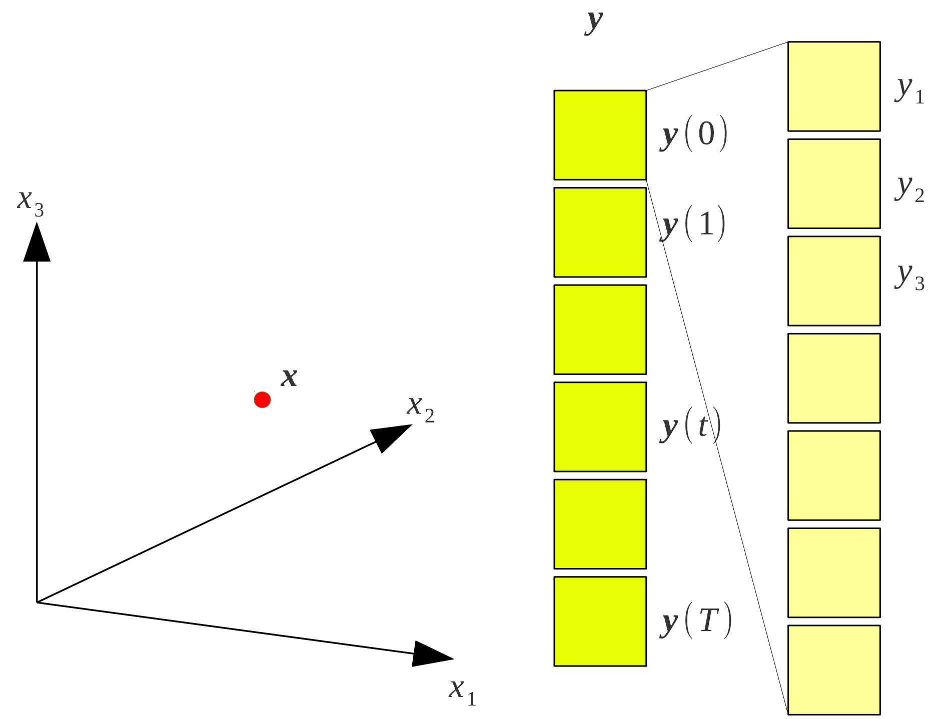
ECMWF Training Course, University of Reading, March 2016

Lecture outline

- Reminders
- Importance of the $\mathbf{B}$-matrix
- Modelling $\mathbf{B}$
- Measuring $\mathbf{B}$

Example state and observation vectors





$\mathbf{x}_{\mathrm{A}}$ analysis state
$\mathbf{x}_{\mathrm{B}}$ background state

Sometimes $\mathbf{x}$ and $\mathbf{y}$ are for only one time (3D-Var)

$\mathbf{x}$-vectors have n elements in total
$\mathbf{y}$-vectors have p elements in total

Reminder of the variational cost function

$$\mathcal{J}[\mathbf{x}] \sim \frac{1}{2}\left|\mathbf{x}-\mathbf{x}_{\mathrm{B}}\right|^2 + \frac{1}{2}\left|\mathbf{y}-\boldsymbol{\mathcal{H}}(\mathbf{x})\right|^2,$$

$$\sim \frac{1}{2}\left(\mathbf{x}-\mathbf{x}_{\mathrm{B}}\right)^{\mathrm{T}}\left(\mathbf{x}-\mathbf{x}_{\mathrm{B}}\right) + \frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mathcal{H}}(\mathbf{x})\right)^{\mathrm{T}}\left(\mathbf{y}-\boldsymbol{\mathcal{H}}(\mathbf{x})\right),$$

$$\mathcal{J}[\mathbf{x}] = \frac{1}{2}\left|\mathbf{x}-\mathbf{x}_{\mathrm{B}}\right|^2_{\mathbf{B}} + \frac{1}{2}\left|\mathbf{y}-\boldsymbol{\mathcal{H}}(\mathbf{x})\right|^2_{\mathbf{R}},$$

$$= \frac{1}{2}\left(\mathbf{x}-\mathbf{x}_{\mathrm{B}}\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}_{\mathrm{B}}\right) + \frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mathcal{H}}(\mathbf{x})\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}-\boldsymbol{\mathcal{H}}(\mathbf{x})\right), \qquad \text{3DVar}$$

$$\mathcal{J}[\mathbf{x}] = \underbrace{\frac{1}{2}\left(\mathbf{x}-\mathbf{x}_{\mathrm{B}}\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}_{\mathrm{B}}\right)}_{\text{measures departure from } \mathbf{x}_{\mathrm{B}}} + \underbrace{\frac{1}{2}\sum_{i=0}^{T}\left(\mathbf{y}(t_i)-\boldsymbol{\mathcal{H}}(\mathbf{x}(t_i))\right)^{\mathrm{T}}\left[\mathbf{R}(t_i)\right]^{-1}\left(\mathbf{y}(t_i)-\boldsymbol{\mathcal{H}}(\mathbf{x}(t_i))\right),}_{\text{measures departure from } \mathbf{y}} \qquad \text{4DVar}$$

$$\mathbf{x}_{\mathrm{A}} = \mathrm{argmin}(\mathcal{J}[\mathbf{x}]).$$

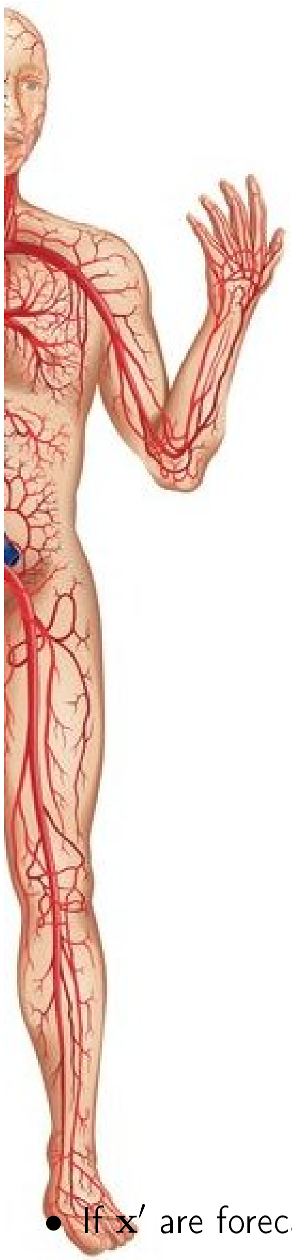
- The covariance matrices $\mathbf{B}$ and $\mathbf{R}$ influence the analysis profoundly.
- If $\boldsymbol{\mathcal{H}}(\mathbf{x})$ is a linear or weakly non-linear function then we can write (3D-Var for simplicity):

$$\mathbf{y}-\boldsymbol{\mathcal{H}}(\mathbf{x}_{\mathrm{B}}+\delta\mathbf{x}) \approx \mathbf{y}-\left(\boldsymbol{\mathcal{H}}(\mathbf{x}_{\mathrm{B}})+\mathbf{H}\delta\mathbf{x}\right),$$

and $\mathbf{x}_{\mathrm{A}}$ can be written explicitly:

$$\mathbf{x}_{\mathrm{A}} = \mathbf{x}_{\mathrm{B}} + \mathbf{B}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\right)^{-1}\left(\mathbf{y}-\boldsymbol{\mathcal{H}}(\mathbf{x}_{\mathrm{B}})\right).$$

Anatomy of a covariance matrix



Univariate background error covariance matrix (e.g. if $\mathbf{x}$ represents a pressure field only):

variance

$$\mathbf{x} = \mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}, \qquad \mathrm{cov}(\mathbf{p}') = \langle \mathbf{p}'\mathbf{p}'^{\mathrm{T}} \rangle = \begin{pmatrix} \langle p_1'^2 \rangle & \langle p_1' p_2' \rangle & \cdots & \langle p_1' p_n' \rangle \\ \langle p_2' p_1' \rangle & \langle p_2'^2 \rangle & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \langle p_n' p_1' \rangle & \cdots & \cdots & \langle p_n'^2 \rangle \end{pmatrix}.$$

outer product

covariance
(univariate)

where $\mathbf{p}' = \mathbf{p} - \langle \mathbf{p} \rangle$.

Multivariate background error covariance matrix (e.g. if $\mathbf{x}$ represents pressure, zonal wind and meridional wind):

autocovariance sub–matrix

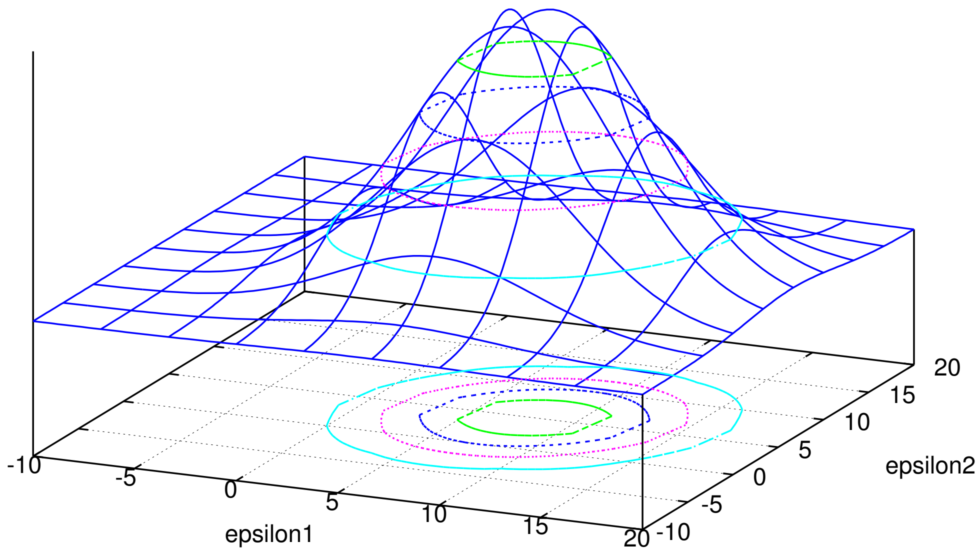
$$\mathbf{x} = \begin{pmatrix} \mathbf{p} \\ \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} p_1 \\ \vdots \\ p_{n/3} \\ u_1 \\ \vdots \\ u_{n/3} \\ v_1 \\ \vdots \\ v_{n/3} \end{pmatrix}, \qquad \mathrm{cov}(\mathbf{x}') = \langle \mathbf{x}'\mathbf{x}'^{\mathrm{T}} \rangle = \begin{pmatrix} \langle \mathbf{p}'\mathbf{p}'^{\mathrm{T}} \rangle & \langle \mathbf{p}'\mathbf{u}'^{\mathrm{T}} \rangle & \langle \mathbf{p}'\mathbf{v}'^{\mathrm{T}} \rangle \\ \langle \mathbf{u}'\mathbf{p}'^{\mathrm{T}} \rangle & \langle \mathbf{u}'\mathbf{u}'^{\mathrm{T}} \rangle & \langle \mathbf{u}'\mathbf{v}'^{\mathrm{T}} \rangle \\ \langle \mathbf{v}'\mathbf{p}'^{\mathrm{T}} \rangle & \langle \mathbf{v}'\mathbf{u}'^{\mathrm{T}} \rangle & \langle \mathbf{v}'\mathbf{v}'^{\mathrm{T}} \rangle \end{pmatrix}.$$

multivariate covariance sub–matrix

These covariances are symmetric matrices.

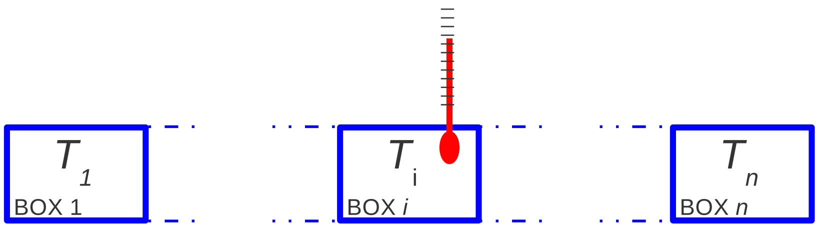
- If $\mathbf{x}'$ are forecast errors, $\boldsymbol{\epsilon}_{\mathrm{B}}$, then above is $\mathbf{B}$-matrix.
- Observation error covariance: $\mathbf{R} = \langle \mathbf{y}'\mathbf{y}'^{\mathrm{T}} \rangle$, $\mathbf{y}'$ is observation error.

Link to Gaussian PDFs



$$
\begin{aligned}
\mathbf{x} &= \mathbf{x}_{\mathrm{B}} - \boldsymbol{\epsilon}_{\mathrm{B}}, \\
\mathbf{x} &\sim N(\mathbf{x}_{\mathrm{B}}, \mathbf{B}), \\
\boldsymbol{\epsilon}_{\mathrm{B}} &\sim N(0, \mathbf{B}), \\
P(\boldsymbol{\epsilon}_{\mathrm{B}}) &= \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{B})}} = \exp -\frac{1}{2}\boldsymbol{\epsilon}_{\mathrm{B}}^{\mathrm{T}}\mathbf{B}^{-1}\boldsymbol{\epsilon}_{\mathrm{B}}.
\end{aligned}
$$

Importance of covariance matrices (demo with $n = n$, $p = 1$)



$$\mathbf{x} = \begin{pmatrix} T_1 \\ \vdots \\ T_i \\ \vdots \\ T_n \end{pmatrix}, \quad \mathbf{x}_\mathrm{B} = \begin{pmatrix} T_\mathrm{B1} \\ \vdots \\ T_{\mathrm{B}i} \\ \vdots \\ T_{\mathrm{B}n} \end{pmatrix}, \quad \mathbf{y} = (y)\,, \quad \boldsymbol{\mathcal{H}}(\mathbf{x}) = T_i,$$

$$\mathbf{H} = \begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} B_{11} & \cdots & B_{1i} & \cdots & B_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ B_{i1} & \cdots & B_{ii} & \cdots & B_{in} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{n1} & \cdots & B_{ni} & \cdots & B_{nn} \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} \sigma_\mathrm{o}^2 \end{pmatrix}.$$

The analysis formula for the analysis increment is:

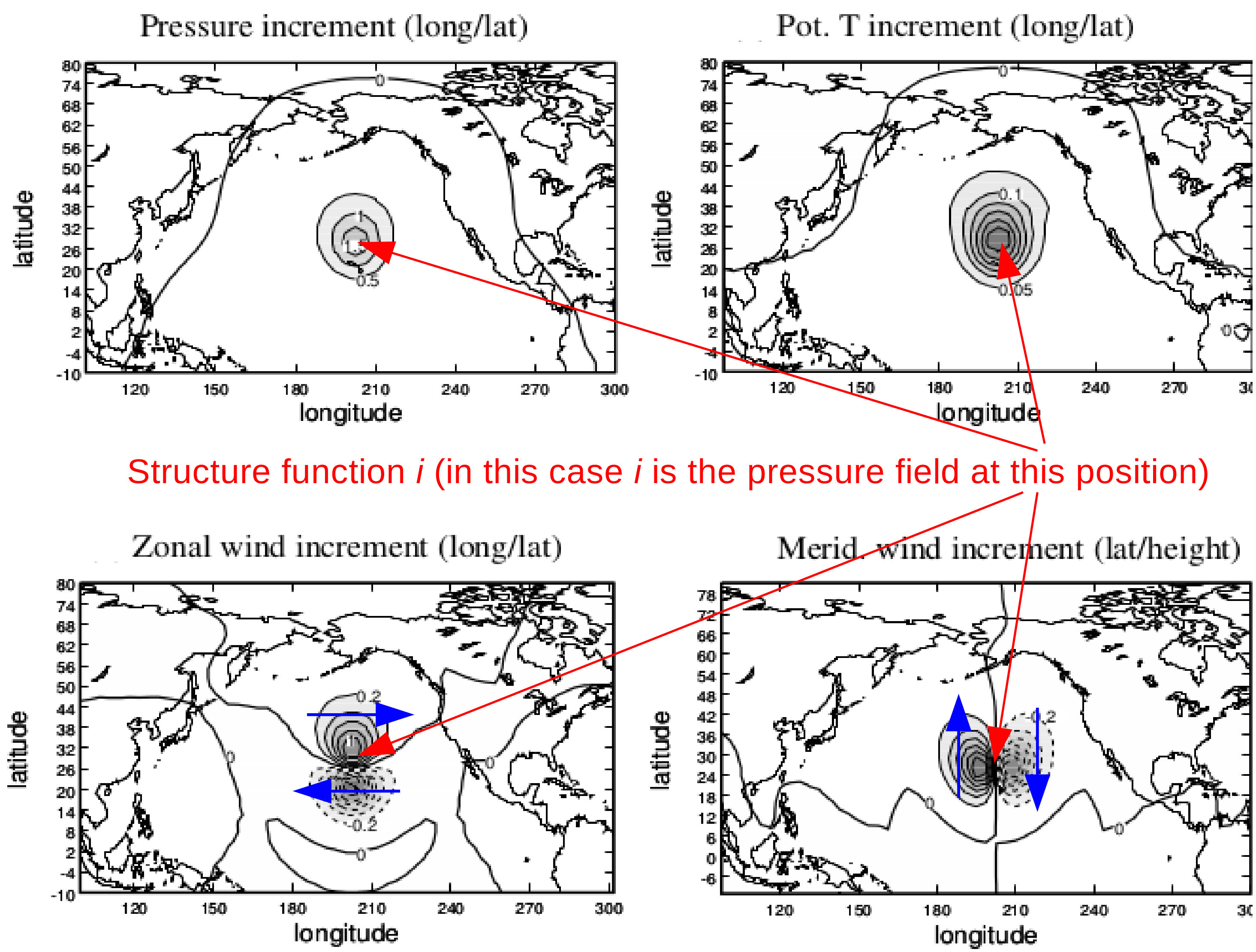
$$\mathbf{x}_\mathrm{A} = \mathbf{x}_\mathrm{B} + \mathbf{B}\mathbf{H}^\mathrm{T}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^\mathrm{T}\right)^{-1}\left(\mathbf{y} - \boldsymbol{\mathcal{H}}(\mathbf{x}_\mathrm{B})\right).$$

$$\mathbf{B}\mathbf{H}^\mathrm{T} = \begin{pmatrix} B_{11} & \cdots & B_{1i} & \cdots & B_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ B_{i1} & \cdots & B_{ii} & \cdots & B_{in} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{n1} & \cdots & B_{ni} & \cdots & B_{nn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} B_{1i} \\ \vdots \\ B_{ii} \\ \vdots \\ B_{ni} \end{pmatrix}, \qquad \mathbf{H}\mathbf{B}\mathbf{H}^\mathrm{T} = \begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \end{pmatrix} \begin{pmatrix} B_{1i} \\ \vdots \\ B_{ii} \\ \vdots \\ B_{ni} \end{pmatrix} = (B_{ii}) = \left(\sigma_{\mathrm{B}i}^2\right),$$

$$\mathbf{x}_\mathrm{A} = \begin{pmatrix} T_\mathrm{B1} \\ \vdots \\ T_{\mathrm{B}i} \\ \vdots \\ T_{\mathrm{B}n} \end{pmatrix} + \begin{pmatrix} B_{1i} \\ \vdots \\ B_{ii} \\ \vdots \\ B_{ni} \end{pmatrix} \frac{1}{\sigma_\mathrm{o}^2 + \sigma_{\mathrm{B}i}^2}\left(y - T_{\mathrm{B}i}\right).$$

The analysis increment is a vector $\propto$ the ith column of $\mathbf{B}$ (called a structure function or covariance function).

Structure functions for flow in the mid-latitude atmosphere



In this case the wind part of the structure function is in geostrophic balance with the pressure

Modelling a covariance matrix

- Observation error covariance matrices ($\mathbf{R}$):
  - Often taken to be diagonal for independent obs. Observation error variances (diagonal elements) depend on characteristics of the instrument.
  - Another contribution is representivity error which will have diagonal (and possibly off-diagonal) elements.
  - If measurements are not independent (e.g. if they are derived using some procedure) then $\mathbf{R}$ should not be diagonal.
- Background error covariance matrices ($\mathbf{B}$):
  - Can be rarely represented explicitly ($\mathbf{x} \in \mathbb{R}^n$ $[n \sim 10^9]$, $\mathbf{B} \in \mathbb{R}^{n\times n}$ $[n \times n \sim 10^{18}]$).
  - Difficult to measure (need a large sample of (unknowable) forecast errors).
  - Can be modelled using a variety of methods:
    - 'Inverse Laplacians'.
    - Diffusion operators (used e.g. in Ocean DA).
    - Recursive filters.
    - Spectral methods, wavelet methods.
    - Exploit physics (e.g. geophysical balance).
    - Control variable transforms (transform to a space where $\mathbf{B}$ is simpler - e.g. diagonal).
- Model error covariance matrices ($\mathbf{Q}$).

Making variational DA work - control variable transforms (CVTs)

- Key to success of 3D/4D-Var in NWP is the $\mathbf{B}$-matrix.
- This is modelled, e.g., via (linear) change of variables - a CVT:
  - $\delta\mathbf{x} = \mathbf{U}\delta\mathbf{v}$.
  - Background errors in the $\delta\mathbf{v}$-representation are assumed to be mutually uncorrelated:

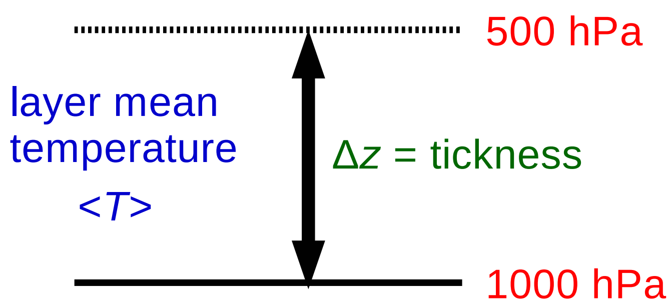
$$\begin{aligned}
\left\langle \boldsymbol{\epsilon}_{\mathrm{B}} \boldsymbol{\epsilon}_{\mathrm{B}}^{\mathrm{T}} \right\rangle_{\mathrm{B}} &\approx \mathbf{B}, \\
\left\langle \delta\mathbf{v}\delta\mathbf{v}^{\mathrm{T}} \right\rangle_{\mathrm{B}} &= \mathbf{I}, \\
\left\langle \left[\mathbf{U}^{-1}\boldsymbol{\epsilon}_{\mathrm{B}}\right] \left[\mathbf{U}^{-1}\boldsymbol{\epsilon}_{\mathrm{B}}\right]^{\mathrm{T}} \right\rangle_{\mathrm{B}} &\approx \mathbf{I}, \\
\therefore \mathbf{U}\mathbf{U}^{\mathrm{T}} &\approx \mathbf{B}.
\end{aligned}$$

  - This problem is minimized now w.r.t. $\delta\mathbf{v}$:

$$\begin{aligned}
\mathcal{J}[\delta\mathbf{v}] &= \frac{1}{2}\delta\mathbf{v}^{\mathrm{T}}\delta\mathbf{v} + \frac{1}{2}\left[\mathbf{y} - \boldsymbol{\mathcal{H}}(\mathbf{x}_{\mathrm{B}}) - \mathbf{H}\underbrace{\mathbf{U}\delta\mathbf{v}}_{\delta\mathbf{x}}\right]^{\mathrm{T}} \mathbf{R}^{-1} \left[\mathbf{y} - \boldsymbol{\mathcal{H}}(\mathbf{x}_{\mathrm{B}}) - \mathbf{H}\underbrace{\mathbf{U}\delta\mathbf{v}}_{\delta\mathbf{x}}\right], \\
\nabla_{\delta\mathbf{v}}\mathcal{J} &= \delta\mathbf{v} - \mathbf{U}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{y} - \boldsymbol{\mathcal{H}}(\mathbf{x}_{\mathrm{B}}) - \mathbf{H}\mathbf{U}\delta\mathbf{v}\right].
\end{aligned}$$

Simple example of Control Variable Transform (CVT)

System (two correlated variables)



- State vector ($\langle T\rangle$ in K, Δz in dam):

$$\delta\mathbf{x} = \begin{pmatrix} \delta\langle T\rangle \\ \delta\Delta z \end{pmatrix}.$$

- Constraint applies (weakly applied hypsometric equation):

$$\delta\Delta z = \underbrace{L\delta\langle T\rangle}_{\text{balanced contribution}} + \underbrace{\delta\Delta z_{\text{unbal}}}_{\text{unbalanced contribution}},$$

$$\text{where } L = \frac{R}{10g}\ln\frac{1000\text{hPa}}{500\text{hPa}}.$$

- Control vector ($\langle\delta\mathbf{v}\delta\mathbf{v}^{\mathrm{T}}\rangle_{\mathrm{B}} = \mathbf{I}$):

$$\delta\mathbf{v} = \begin{pmatrix} \delta v_{\text{bal}} \\ \delta v_{\text{unbal}} \end{pmatrix}.$$

- Scale by background error standard deviations, $\delta\langle T\rangle = \sigma_{\text{bal}}\delta v_{\text{bal}}$, $\delta\Delta z_{\text{unbal}} = \sigma_{\text{unbal}}\delta v_{\text{unbal}}$:

$$\begin{pmatrix} \delta\langle T\rangle \\ \delta\Delta z_{\text{unbal}} \end{pmatrix} = \begin{pmatrix} \sigma_{\text{bal}} & 0 \\ 0 & \sigma_{\text{unbal}} \end{pmatrix} \begin{pmatrix} \delta v_{\text{bal}} \\ \delta v_{\text{unbal}} \end{pmatrix}.$$

- The complete CVT ($\delta\mathbf{x} = \mathbf{U}\delta\mathbf{v}$):

$$\underbrace{\begin{pmatrix} \delta\langle T\rangle \\ \delta\Delta z \end{pmatrix}}_{\delta\mathbf{x}} = \underbrace{\begin{pmatrix} 1 & 0 \\ L & 1 \end{pmatrix}\begin{pmatrix} \sigma_{\text{bal}} & 0 \\ 0 & \sigma_{\text{unbal}} \end{pmatrix}}_{\mathbf{U}} \underbrace{\begin{pmatrix} \delta v_{\text{bal}} \\ \delta v_{\text{unbal}} \end{pmatrix}}_{\delta\mathbf{v}}.$$

- Implied covariances ($\mathbf{B} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$):

$$\mathbf{B} = \begin{pmatrix} \sigma^2_{\text{bal}} & \sigma^2_{\text{bal}}L \\ \sigma^2_{\text{bal}}L & \sigma^2_{\text{bal}}L^2 + \sigma^2_{\text{unbal}} \end{pmatrix}.$$

- Observation of $\langle T\rangle$ then gives information about Δz (and vice-versa) in a physically consistent way.

Methods to estimate B

Reminder

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

$$= \begin{pmatrix} \left\langle (x_{\mathrm{B}1} - x_{\mathrm{t}1})^2 \right\rangle_{\mathrm{B}} & \langle (x_{\mathrm{B}1} - x_{\mathrm{t}1})(x_{\mathrm{B}2} - x_{\mathrm{t}2}) \rangle_{\mathrm{B}} & \cdots & \langle (x_{\mathrm{B}1} - x_{\mathrm{t}1})(x_{\mathrm{B}n} - x_{\mathrm{t}n}) \rangle_{\mathrm{B}} \\ \langle (x_{\mathrm{B}2} - x_{\mathrm{t}2})(x_{\mathrm{B}1} - x_{\mathrm{t}1}) \rangle_{\mathrm{B}} & \left\langle (x_{\mathrm{B}2} - x_{\mathrm{t}2})^2 \right\rangle_{\mathrm{B}} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \langle (x_{\mathrm{B}n} - x_{\mathrm{t}n})(x_{\mathrm{B}1} - x_{\mathrm{t}1}) \rangle_{\mathrm{B}} & \cdots & \cdots & \left\langle (x_{\mathrm{B}n} - x_{\mathrm{t}n})^2 \right\rangle_{\mathrm{B}} \end{pmatrix}.$$

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

Problem

$\mathbf{x}_{\mathrm{t}}$ is unknowable so need a proxy for forecast error $\mathbf{x}_{\mathrm{B}} - \mathbf{x}_{\mathrm{t}}$.

Popular approaches

Method	Description and references
"Canadian quick" method	$\mathbf{x}_\mathrm{B} - \mathbf{x}_\mathrm{t} \sim (\mathbf{x}_\mathrm{B}(t+T) - \mathbf{x}_\mathrm{B}(T))/\sqrt{2}$. Take population from one long time run. Polavarapu et al. (2005)
Analysis of innovations $\mathbf{d} = \mathbf{y} - \mathbf{H}\mathbf{x}_\mathrm{B}$	Choose a pair of direct and independent obs separated by r: $[y(r) - x_\mathrm{B}(r)]\,[y(r+\Delta r) - x_\mathrm{B}(r+\Delta r)] =$ $[\{y(r) - x_\mathrm{t}(r)\} - \{x_\mathrm{B}(r) - x_\mathrm{t}(r)\}]\,[\{y(r+\Delta r) - x_\mathrm{t}(r+\Delta r)\} - \{x_\mathrm{B}(r+\Delta r) - x_\mathrm{t}(r+\Delta r)\}]$ $\langle[\epsilon^y(r) - \epsilon^{x_\mathrm{B}}(r)]\,[\epsilon^y(r+\Delta r) - \epsilon^{x_\mathrm{B}}(r+\Delta r)]\rangle = \langle\epsilon^y(r)\epsilon^y(r+\Delta r)\rangle + \langle\epsilon^{x_\mathrm{B}}(r)\epsilon^{x_\mathrm{B}}(r+\Delta r)\rangle$, (above assumes obs and bg errors are uncorrelated). Take population from many pairs with same Δr. Furthermore if $\Delta r > 0$: $\langle\epsilon^y(r)\epsilon^y(r+\Delta r)\rangle = 0$. Rutherford (1972), Hollingsworth and Lönnberg (1986), Järvinen (2001)
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)
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)

Summary

- Covariance matrices appear in many DA methods (especially variational DA).
  - A covariance matrix describes the shape of a Gaussian distribution.
  - $\mathbf{B}$ and $\mathbf{R}$ appear in variational cost function (and $\mathbf{Q}$ in weak constraint formulations).
- Covariance matrices are important.
  - E.g. $\mathbf{B}$ specifies how precise $\mathbf{x}_{\mathrm{B}}$ is, and how to give smooth analysis increments between positions in space and between different variables.
- $\mathbf{B}$ is too large to be known (and there is too little information to know it anyway!)
  - $\mathbf{B}$ needs to be modelled based on reasonable ideas.
  - The method of "control variable transforms" is a leading method.
  - Minimize $\mathcal{J}$ in "control variable space" (easy) which is related to model space via the control variable transform.
- It is impossible to measure $\mathbf{B}$ exactly.
  - Use a proxy method.

Further reading - selected books and papers

- **Barlow, R.J.**, Statistics - A guide to the use of statistical methods in the physical sciences, John Wiley and Sons (1989). *This is an elementary, readable book on statistics for the scientist (e.g. it derives the Gaussian distribution from first principles). It also covers the least squares problem.*

- **Rodgers C.D.**, Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific Publishing (2000). *This is a very readable book. Even though it focuses on satellite retrieval theory (mathematically a similar problem to data assimilation), this is a good book for virtually everything that you need to know about covariances. It also contains a summary of basic data assimilation methods and has a useful appendix on linear algebra.*

- **Lewis J.M., Lakshmivarahan S., Dhall S.**, Dynamic Data Assimilation: A Least Squares Approach, Cambridge University Press (2006). *This huge book covers a lot of material with a lot of repetition. It has some good introductory chapters and some useful results if you know where to look. (Unfortunately there are LOADS of typos.)*

- **Kalnay E.**, Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press (2002). *A large section of this book covers data assimilation, and there is also a lot of basic material for the budding dynamic modeller. The data assimilation part is introductory, but covers most key ideas. It will leave you wanting to know more!*

- **Schlatter T.W.**, Variational assimilation of meteorological observations in the lower atmosphere: a tutorial on how it works, J. Atmos. and Solar-Terr. Phys. 62 pp.1057-1070 (2000). *It is worth getting hold of this paper as it is an excellent description of variational data assimilation (relevant to lectures later in the course).*

- **Bannister R.N.**, A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances., Q.J. Roy. Met. Soc. 134, 1951-1970 (2008) and **Bannister R.N.**, A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics., Q.J. Roy. Met. Soc. 134, 1971-1996 (2008). *What can I say - blatant self publicity! A source of information about background error covariances and how they can be modelled.*

- **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). *"Canadian quick" method.*

- **Rutherford I.D.** 1972. Data assimilation by statistical interpolation of forecast error fields. J. Atmos. Sci. 29: 809–815. *Original reference to the analysis of innovations method.*

- **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). *The most famous work on the analysis of innovations method.*

- **Järvinen H.**, Temporal evolution of innovation and residual statistics in the ECMWF variational data assimilation systems. Tellus 53A: 333–347 (2001). *More recent work on the analysis of innovations method.*

- **Parrish D.F., Derber J.C.**, The National Meteorological Center's spectral statistical interpolation analysis system. Mon. Wea. Rev. 120 1747–1763 (1992). *Original reference for the 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). *In-depth analysis of the NMC 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). *Explains the ideas behind the generation of an ensemble.*

- **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). *Example background error covariances derived from an ensemble.*

- **Bonavita M., Holm E., Isaksen L., Fisher M.**, The evolution of the ECMWF hybrid data assimilation system, Q.J.R. Meteor. Soc. (2015). *Latest paper documenting the ensemble-based calibration of the ECMWF* $\mathbf{B}$*-matrix.*