Anatomy of a covariance matrix

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

\[
\mathbf{x} = \mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}, \quad \text{cov}(\mathbf{p}') = \langle \mathbf{p}' \mathbf{p}'^T \rangle = \begin{pmatrix} \langle p'_1^2 \rangle & \langle p'_1 p'_2 \rangle & \cdots & \langle p'_1 p'_n \rangle \\ \langle p'_2 p'_1 \rangle & \langle p'_2^2 \rangle & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \langle p'_n p'_1 \rangle & \langle p'_n p'_2 \rangle & \cdots & \langle p'_n^2 \rangle \end{pmatrix},
\]

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):

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

These covariances are symmetric matrices.

- If \( \mathbf{x}' \) are forecast errors, \( \mathbf{e}_h \), then above is \( \mathbf{B} \)-matrix.

- Observation error covariance: \( \mathbf{R} = \langle \mathbf{y}' \mathbf{y}'^T \rangle \), \( \mathbf{y}' \) is observation error.
Importance of covariance matrices (demo with $n = n$, $p = 1$)

The analysis formula for the analysis increment is:

$$
x^a = x^b + B H^T (R + H B H^T)^{-1} (y - \mathcal{H}(x^b)).$$

$$
x = \begin{pmatrix} T_1 \\ \vdots \\ T_i \\ \vdots \\ T_n \end{pmatrix}, \quad x^b = \begin{pmatrix} T^b_1 \\ \vdots \\ T^b_i \\ \vdots \\ T^b_n \end{pmatrix}, \quad y = (y), \quad \mathcal{H}(x) = T_i,$$

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

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

$$
BH^T = \begin{pmatrix} B_{1i} \\ \vdots \\ B_{i1} \\ \vdots \\ B_{ni} \end{pmatrix} = (B_{ii}) = (\sigma^2_{B_{ii}}),
$$

$$
HBH^T = \begin{pmatrix} \sigma^2_o \\ \sigma^2_{B_{ii}} \\ \vdots \\ \sigma^2_{B_{ni}} \end{pmatrix}.
$$

The analysis increment is a vector $\propto$ the $i$th column of $B$ (called a structure function or covariance function).
Structure functions for flow in the mid-latitude atmosphere

Structure function $i$ (in this case $i$ is the pressure field at this position)

Zonal wind increment (long/lat)

Merid. wind increment (lat/height)

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

- **Observation error covariance matrices** \( (R) \):
  - Describes errors in the observing system (e.g. the instrument), errors in the observation operator, and representativity error.
  - Often taken to be diagonal for independent obs.
  - If obs. errors are not independent, then there are off-diagonal elements.
    If measurements are not independent (e.g. if they are derived using some procedure) then \( R \) should not be diagonal.

- **Background error covariance matrices** \( (B) \):
  - Describes errors in the background state (forecast from previous analysis).
  - Depends on the analysis errors of the previous assimilation, and on forecast model error.
  - Can be rarely represented explicitly \( (x \in \mathbb{R}^n [n \sim 10^9], B \in \mathbb{R}^{n \times n} [n \times n \sim 10^{18}]) \).
  - Difficult to measure (need a large sample of (unknown) 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 \( B \) is simpler - e.g. diagonal).

- **Model error covariance matrices** \( (Q) \):
  - Describes errors in the forecast model used within 4D-Var.
  - Often completely neglected operationally.
Making variational DA work – control variable transforms (CVTs)

- Key to success of 3D/4D-Var in NWP is the $B$-matrix. Incremental 3dVar cost fn:

$$J[δx] = δx^T B^{-1} δx + \left[ y - \mathcal{H}(x^b) - Hδx \right]^T R^{-1} \left[ y - \mathcal{H}(x^b) - Hδx \right]$$

$$x = x^b + δx$$

- $B$ can be modelled, e.g., via (linear) change of variables - a CVT:
  - $δx = Uδv$.
  - Background errors in the $δv$-representation are assumed to be mutually uncorrelated:

$$\langle ε^b_b ε^b_b^T \rangle_B \approx B, \quad \langle δvδv^T \rangle_B = I, \quad \langle [U^{-1}_S B] [U^{-1}_S B]^T \rangle_B \approx I, \quad \therefore UU^T \approx B.$$ 

- This problem is minimized now w.r.t. $δv$:

$$J[δv] = \frac{1}{2} δv^T δv + \frac{1}{2} \left[ y - \mathcal{H}(x^b) - HUδv \right]^T R^{-1} \left[ y - \mathcal{H}(x^b) - HUδv \right],$$

$$∇_{δv} J = δv - U^T H^T R^{-1} \left[ y - \mathcal{H}(x^b) - HUδv \right].$$
Simple example of Control Variable Transform (CVT)

System (two correlated variables)

- State vector ($\vec{T}$ in K, $\Delta z$ in dam):
  \[
  \delta x = \begin{pmatrix}
  \delta \vec{T} \\
  \delta \Delta z
  \end{pmatrix}.
  \]

- Constraint applies (weakly applied hypsometric equation):
  \[
  \delta \Delta z = \frac{L}{\text{balanced contribution}} \delta \vec{T} + \frac{\delta \Delta z_{\text{unbal}}}{\text{unbalanced contribution}},
  \]
  where \( L = \frac{R}{10g} \ln \frac{1000\text{hPa}}{500\text{hPa}} \).

- Control vector ($\langle \delta \vec{v} \delta \vec{v}^T \rangle_B = I$):
  \[
  \delta v = \begin{pmatrix}
  \delta v_{\text{bal}} \\
  \delta v_{\text{unbal}}
  \end{pmatrix}.
  \]

- Scale by background error standard deviations, $\delta \vec{T} = \sigma_{\text{bal}} \delta v_{\text{bal}}$, $\delta \Delta z_{\text{unbal}} = \sigma_{\text{unbal}} \delta v_{\text{unbal}}$:
  \[
  \begin{pmatrix}
  \delta \vec{T} \\
  \delta \Delta z_{\text{unbal}}
  \end{pmatrix} = \begin{pmatrix}
  \sigma_{\text{bal}} & 0 \\
  0 & \sigma_{\text{unbal}}
  \end{pmatrix} \begin{pmatrix}
  \delta v_{\text{bal}} \\
  \delta v_{\text{unbal}}
  \end{pmatrix}.
  \]

- The complete CVT ($\delta x = U \delta v$):
  \[
  \begin{pmatrix}
  \delta \vec{T} \\
  \delta \Delta z
  \end{pmatrix}_{\delta x} = \begin{pmatrix}
  1 & 0 \\
  L & 1
  \end{pmatrix} \begin{pmatrix}
  \sigma_{\text{bal}} & 0 \\
  0 & \sigma_{\text{unbal}}
  \end{pmatrix} \begin{pmatrix}
  \delta v_{\text{bal}} \\
  \delta v_{\text{unbal}}
  \end{pmatrix}_{\delta v}.
  \]

- Implied covariances ($B = U \Sigma U^T$):
  \[
  B = \begin{pmatrix}
  \sigma_{\text{bal}}^2 & \sigma_{\text{bal}} \sigma_{\text{unbal}} L \\
  \sigma_{\text{bal}} \sigma_{\text{unbal}} L & \sigma_{\text{unbal}}^2 \sigma_{\text{bal}}^2 + \sigma_{\text{unbal}}^2
  \end{pmatrix}.
  \]

- Observation of $\vec{T}$ then gives information about $\Delta z$ (and vice-versa) in a physically consistent way.
Methods to estimate $B$

Reminder

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad B = \left\langle (x^b - x^t)(x^b - x^t)^T \right\rangle_B,$$

$$= \begin{pmatrix} \langle (x^b_1 - x^t_1)^2 \rangle_B & \langle (x^b_1 - x^t_1)(x^b_2 - x^t_2) \rangle_B & \cdots & \langle (x^b_1 - x^t_1)(x^b_n - x^t_n) \rangle_B \\ \langle (x^b_2 - x^t_2)(x^b_1 - x^t_1) \rangle_B & \langle (x^b_2 - x^t_2)^2 \rangle_B & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \langle (x^b_n - x^t_n)(x^b_1 - x^t_1) \rangle_B & \cdots & \cdots & \langle (x^b_n - x^t_n)^2 \rangle_B \end{pmatrix}.$$ 

$\langle \bullet \rangle_B$: average over population of possible backgrounds.

Problem

$x^t$ is unknowable so need a proxy for forecast error $x^b - x^t$. 
## Popular approaches

<table>
<thead>
<tr>
<th>Method</th>
<th>Description and references</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Canadian quick&quot; method</td>
<td>$x^b - x^t \sim \left( x^b(t + T) - x^b(T) \right) / \sqrt{2}$. Take population from one long time run. Polavarapu et al. (2005)</td>
</tr>
</tbody>
</table>
| Analysis of innovations | Choose a pair of direct and independent obs separated by $r$:

$$
\begin{align*}
[y(r) - x^b(r)] \left[ y(r + \Delta r) - x^b(r + \Delta r) \right] &= \left[ \{y(r) - x^b(r)\} - \{x^b(r) - x^t(r)\} \right] \left[ \{y(r + \Delta r) - x^t(r + \Delta r)\} - \{x^b(r + \Delta r) - x^t(r + \Delta r)\} \right] \\
\left\langle \left[ e^y(r) - e^{x^b}(r) \right] \left[ e^y(r + \Delta r) - e^{x^b}(r + \Delta r) \right] \right\rangle &= \langle e^y(r)e^y(r + \Delta r) \rangle + \langle e^{x^b}(r)e^{x^b}(r + \Delta r) \rangle,
\end{align*}
$$

(above assumes obs and bg errors are uncorrelated). Take population from many pairs with same $\Delta r$. Furthermore suppose that $\Delta r > 0$: $\langle e^y(r)e^y(r + \Delta r) \rangle = 0$. Rutherford (1972), Hollingsworth and Lönnberg (1986), Järvinen (2001) |
| NMC method              | Choose pairs of lagged forecasts valid at the same time, e.g.: $x^b - x^t \sim \left( x^b_{48}(t) - x^b_{24}(t) \right) / \sqrt{2}$. 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:

$$
\begin{align*}
x^b - x^t \sim \langle x^b \rangle \text{ or } x^b - x^t &\sim \left( x^b_{(i)} - x^b_{(j)} \right) / \sqrt{2}.
\end{align*}
$$

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.
  - B and R appear in variational cost function (and Q in weak constraint formulations).

- Covariance matrices are important.
  - E.g. B specifies how precise x^b is, and how to give smooth analysis increments between positions in space and between different variables.

- B is too large to be known (and there is too little information to know it anyway!)
  - B needs to be modelled based on reasonable ideas.
  - The method of “control variable transforms” is a leading method.
  - Minimize $J$ in “control variable space” (easy) which is related to model space via the control variable transform.

- It is impossible to measure 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).


