Further introduction to Data assimilation - including error covariances

Vn 2.0, Ross Bannister, r.n.bannister@reading.ac.uk

What do we want DA to achieve?

To combine imperfect data from models, from observations distributed in time and space, exploiting any relevant physical constraints, to produce a more accurate and comprehensive picture of the system as it evolves in time.

"[The atmosphere] is a chaotic system in which errors introduced into the system can grow with time . . . As a consequence, data assimilation is a struggle between chaotic destruction of knowledge and its restoration by new observations." Leith (1993)

What sorts of things have errors?

"All models are wrong . . ." (George Box)

"All models are wrong and all observations are inaccurate." (a data assimilator)

$\mathbf{y}$ Observations (system dependent - e.g. temperature, pressure, humidity, wind, radiance, trace gas mixing ratio, time delay (GPS), radar reflectivity, salinity, optical reflectance, ...).

$\mathbf{x}_{\mathrm{B}}$ Forecast (background) state vector (system dependent).

$\mathcal{H}, \mathcal{M}$, etc. Operators used within the data assimilation itself (e.g. observation operator, model operator, etc.).

$\mathbf{x}_{\mathrm{A}}$ Assimilated state (analysis) state vector (system dependent).

Also: representivity error (due to finite representation of state vector), boundary condition error, ...

Lecture outline

- Representing uncertainty
  - Errors vs. error statistics.
  - PDFs.
  - Normal distributions in one and higher dimensions.
- Combining imperfect information with Bayes Theorem.
- Different ways of solving the same data assimilation problem.
  - Variational assimilation, Kalman filtering, particle filtering, and hybrid methods.
- The state vector and the observation vector.
- Covariance matrices.
  - Anatomy.
  - Importance in (Gaussian) data assimilation.
  - Correlation functions and structure functions.
  - Modelling covariance matrices for your application.

How do we represent uncertainty?

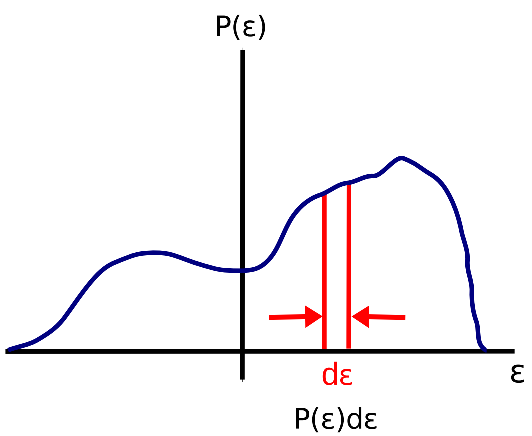
Errors:

- The difference between some estimated quantity and the truth. E.g.:
  - in a forecast $\boldsymbol{\epsilon}_{\mathrm{f}} = \mathbf{x}_{\mathrm{f}} - \mathbf{x}_{\mathrm{t}}$,
  - in an observation $\boldsymbol{\epsilon}_y = \mathbf{y} - \mathbf{y}_{\mathrm{t}}$.
- Errors are unknown and unknowable.

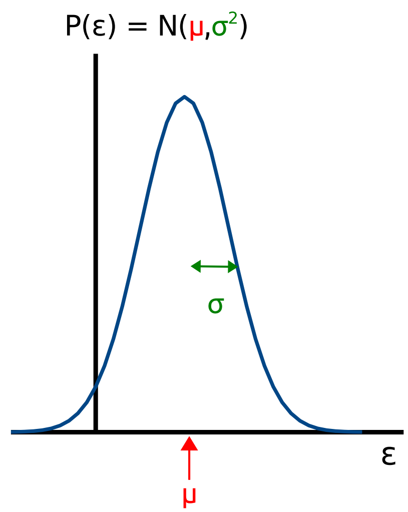
Error statistics:

- Some useful measures of the possible values that $\boldsymbol{\epsilon}$ could have (e.g. a PDF or quantities that describe a PDF).
- Error statistics are knowable (but can be difficult to determine).

GENERAL PDF



GAUSSIAN \ NORMAL PDF



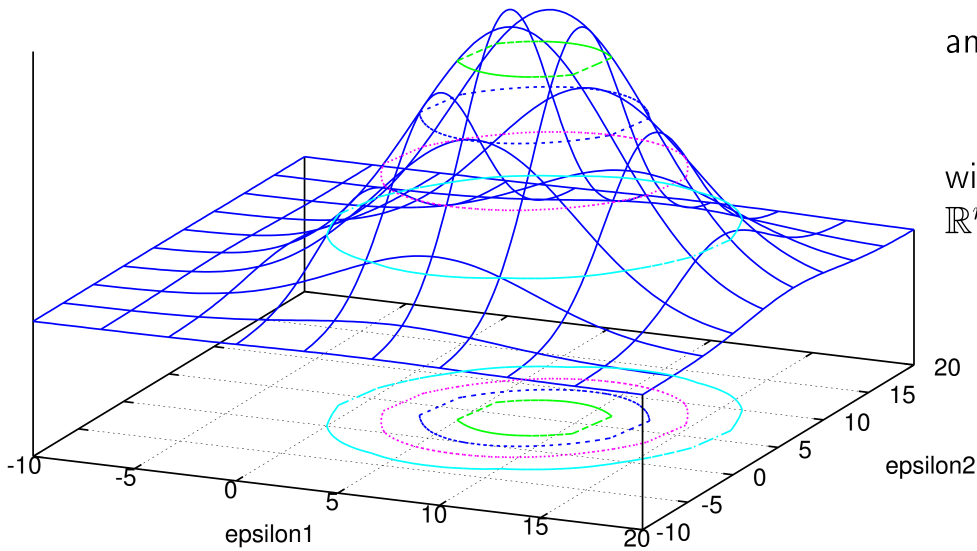
The probability that the error ϵ lies between ϵ and $\epsilon + d\epsilon$ is $P(\epsilon)d\epsilon$.

Form of a 1-D Gaussian:

$$\begin{aligned} \epsilon &\sim N(\mu, \sigma^2), \\ P(\epsilon) &= \frac{1}{2\pi\sigma^2} \exp - \frac{(\epsilon - \mu)^2}{2\sigma^2}. \end{aligned}$$

First moment (μ) and second moment (σ^2) only.

How do we represent uncertainty (continued)?



The probability that the error $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2)^{\mathrm{T}}$ lies between $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon} + d\boldsymbol{\epsilon}$ is $P(\boldsymbol{\epsilon})d\boldsymbol{\epsilon} = P(\boldsymbol{\epsilon})d\epsilon_1 d\epsilon_2$.

Form of an n-dimensional Gaussian for $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots \epsilon_n)^{\mathrm{T}}$ with mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots \mu_n)^{\mathrm{T}} \in \mathbb{R}^n$ and covariance $\mathbf{S} \in \mathbb{R}^{n \times n}$:

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

The Gaussian shown has

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} 8 \\ 5 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} 5 & 1 \\ 1 & 7.5 \end{pmatrix}.$$

General form of $\mathbf{S}$:

$$\mathbf{S} = \begin{pmatrix} \left\langle (\epsilon_1 - \mu_1)^2 \right\rangle & \left\langle (\epsilon_1 - \mu_1)(\epsilon_2 - \mu_2) \right\rangle & \cdots & \left\langle (\epsilon_1 - \mu_1)(\epsilon_n - \mu_n) \right\rangle \\ \left\langle (\epsilon_2 - \mu_2)(\epsilon_1 - \mu_1) \right\rangle & \left\langle (\epsilon_2 - \mu_2)^2 \right\rangle & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \left\langle (\epsilon_n - \mu_n)(\epsilon_1 - \mu_1) \right\rangle & \cdots & \cdots & \left\langle (\epsilon_n - \mu_n)^2 \right\rangle \end{pmatrix}^2.$$

Bayes Theorem (the root of all wisdom)

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}, \quad \propto P(B|A) \times P(A).$$

Let A be the event $\mathbf{x} = \mathbf{x}_\mathrm{t} \in \mathbb{R}^n$ and B be the event $\mathbf{y}_\mathrm{o} \in \mathbb{R}^p$:

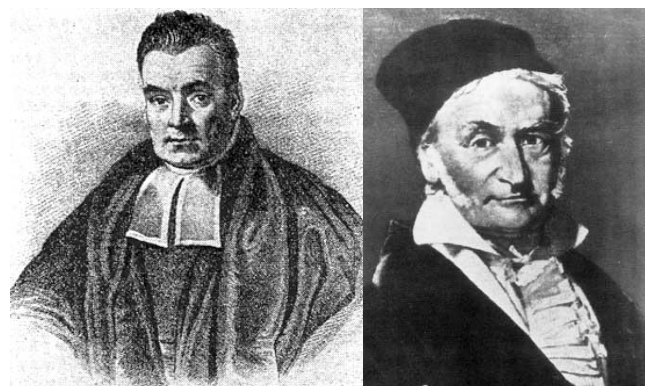
$$\underbrace{P(\mathbf{x} = \mathbf{x}_\mathrm{t}|\mathbf{y}_\mathrm{o})}_{\text{posterior}} \propto \underbrace{P(\mathbf{y}_\mathrm{o}|\mathbf{x} = \mathbf{x}_\mathrm{t})}_{\text{likelihood}} \times \underbrace{P(\mathbf{x} = \mathbf{x}_\mathrm{t})}_{\text{prior}}.$$

Approaches to DA:

- 1st moment: Find the mode of $P(\mathbf{x}|\mathbf{y}_\mathrm{o})$ (maximum likelihood est. - the most likely $\mathbf{x}$).
- 1st moment: Find the mean of $P(\mathbf{x}|\mathbf{y}_\mathrm{o})$, $\langle \mathbf{x} \rangle = \int P(\mathbf{x}|\mathbf{y}_\mathrm{o})\mathbf{x}\, d\mathbf{x}$ (minimum variance est.).
- 1st and 2nd moments: find the covariance of $P(\mathbf{x}|\mathbf{y}_\mathrm{o})$, $\mathrm{cov} = \int P(\mathbf{x}|\mathbf{y}_\mathrm{o})\,(\mathbf{x} - \langle \mathbf{x} \rangle)\,(\mathbf{x} - \langle \mathbf{x} \rangle)^\mathrm{T}\, d\mathbf{x}$.
- The whole PDF (or an approx. of).

Approximations: assume that each PDF is Gaussian.

- Likelihood: mean $\mathcal{H}(\mathbf{x}_\mathrm{t})$, covariance $\mathbf{R} \in \mathbb{R}^{p \times p}$.
- Prior: mean $\mathbf{x}_\mathrm{f}$, covariance $\mathbf{P}_\mathrm{f} \in \mathbb{R}^{n \times n}$.



$$\begin{aligned}
P(\mathbf{x} = \mathbf{x}_\mathrm{t}|\mathbf{y}_\mathrm{o}) &\propto \exp -\frac{1}{2}(\mathbf{y}_\mathrm{o} - \mathcal{H}(\mathbf{x}_\mathrm{t}))^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}_\mathrm{o} - \mathcal{H}(\mathbf{x}_\mathrm{t})) \times \exp -\frac{1}{2}(\mathbf{x} - \mathbf{x}_\mathrm{f})^\mathrm{T}\mathbf{P}_\mathrm{f}^{-1}(\mathbf{x} - \mathbf{x}_\mathrm{f}), \\
&\propto \exp -\frac{1}{2}(\mathbf{y}_\mathrm{o} - \mathcal{H}(\mathbf{x}))^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}_\mathrm{o} - \mathcal{H}(\mathbf{x})) \times \exp -\frac{1}{2}(\mathbf{x} - \mathbf{x}_\mathrm{f})^\mathrm{T}\mathbf{P}_\mathrm{f}^{-1}(\mathbf{x} - \mathbf{x}_\mathrm{f}), \\
&\propto \exp -\frac{1}{2}\left[(\mathbf{y}_\mathrm{o} - \mathcal{H}(\mathbf{x}))^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}_\mathrm{o} - \mathcal{H}(\mathbf{x})) + (\mathbf{x} - \mathbf{x}_\mathrm{f})^\mathrm{T}\mathbf{P}_\mathrm{f}^{-1}(\mathbf{x} - \mathbf{x}_\mathrm{f})\right], \\
&\propto \exp -J[\mathbf{x}], \\
\text{cost function: } J[\mathbf{x}] &\equiv \frac{1}{2}(\mathbf{x} - \mathbf{x}_\mathrm{f})^\mathrm{T}\mathbf{P}_\mathrm{f}^{-1}(\mathbf{x} - \mathbf{x}_\mathrm{f}) + \frac{1}{2}(\mathbf{y}_\mathrm{o} - \mathcal{H}(\mathbf{x}))^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}_\mathrm{o} - \mathcal{H}(\mathbf{x})).
\end{aligned}$$

Data assimilation approaches used in practice

Variational

- Assume Gaussian statistics.
- Solve a variational problem that minimizes the cost function:

$$J[\mathbf{x}] = \frac{1}{2}(\mathbf{x}-\mathbf{x}_{\mathrm{f}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_{\mathrm{f}}) + \frac{1}{2}(\mathbf{y}_{\mathrm{o}}-\mathcal{H}(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}_{\mathrm{o}}-\mathcal{H}(\mathbf{x})).$$

- 1st moment of prior: a-priori is evolved from a previous variational analysis.
- 2nd moment of prior: $\mathbf{P}_{\mathrm{f}} \to \mathbf{B}$ (prescribed) background error covariance matrix.
- Prescribed $\mathbf{R}$.
- Variants: 1D-Var (e.g. for atmospheric profile), 3D-Var (no consideration of time), strong constraint 4D-Var (considers observations over a time window assuming perfect model), weak constraint 4D-Var (account for imperfect model), variational bias estimation, ...
- *Lectures and practicals on Tuesday.*

Kalman Filter

- Assume Gaussian statistics.
- Use a formula that gives the mean (or mode) of the posterior $P(\mathbf{x}|\mathbf{y}_{\mathrm{o}})$:

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

and its error covariance:

$$\mathbf{P}_{\mathrm{A}} = \left[\mathbf{I}-\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R}+\mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\right)^{-1}\mathbf{H}\right]\mathbf{P}_{\mathrm{f}}.$$

- 1st moment of prior: a-priori is evolved from a previous KF analysis:

$$\mathbf{x}_{\mathrm{B}} = \mathcal{M}(\mathbf{x}_{\mathrm{A}}^{\mathrm{prev}}).$$

- 2nd moment of prior: $\mathbf{P}_{\mathrm{f}}$ is evolved from a previous KF analysis.

$$\mathbf{P}_{\mathrm{f}} = \mathbf{M}\mathbf{P}_{\mathrm{a}}^{\mathrm{prev}}\mathbf{M}^{\mathrm{T}} + \mathbf{Q}.$$

- Prescribed $\mathbf{R}$.
- Variants: Optimal Interpolation (not really considered a KF as it does not evolve covariance), ensemble KF (estimate 1st and 2nd moments of prior and posterior PDFs with an ensemble).
- *Lectures and practicals on Wednesday|Thursday.*

Data assimilation approaches used in practice (continued)

Particle Filter

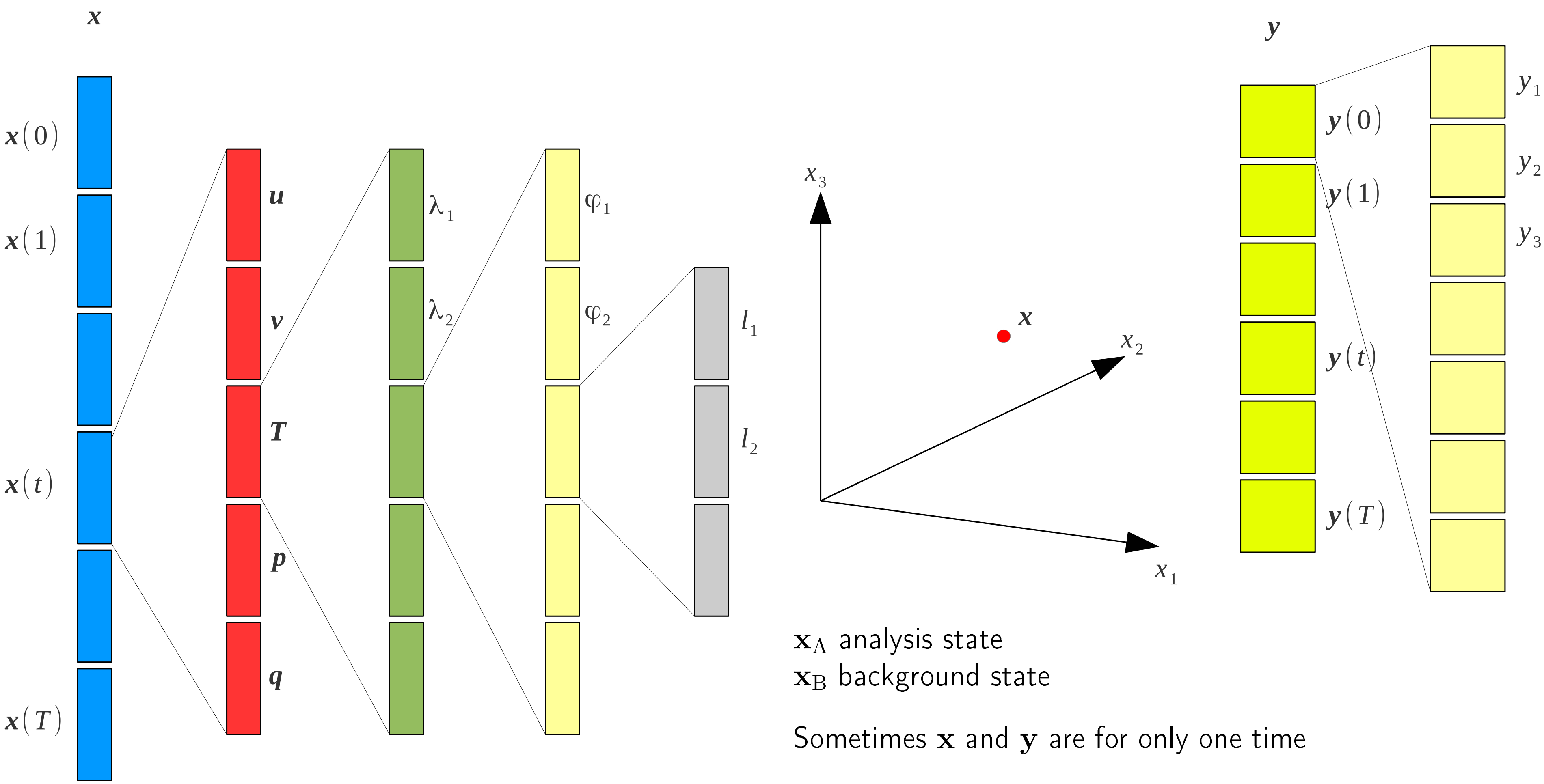
- Approximate PDFs that describe prior and posterior states with a weighted combination of states (the 'particles').
- Non-Gaussian and fully non-linear.
- The "curse of dimensionality".
- *Lectures and practicals on Thursday.*

Hybrid

- Virtually all methods make practical approximations.
- Can combine different methods.
- E.g. variational and ensemble:
  - Variational methods do not have adequate flow dependence in $\mathbf{B}$.
  - Ensemble KF methods suffer from low rank (No. of ensemble members $\ll$ No. of elements in state).
- *Lecture on Thursday.*

All methods derive from Bayes Theorem (the root of all wisdom).

Example state and observation vectors

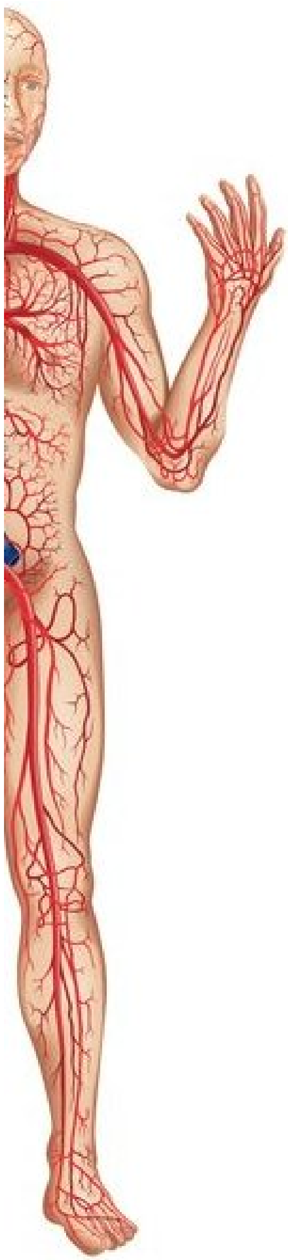


$\mathbf{x}_{\mathrm{A}}$ analysis state
$\mathbf{x}_{\mathrm{B}}$ background state

Sometimes $\mathbf{x}$ and $\mathbf{y}$ are for only one time

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

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

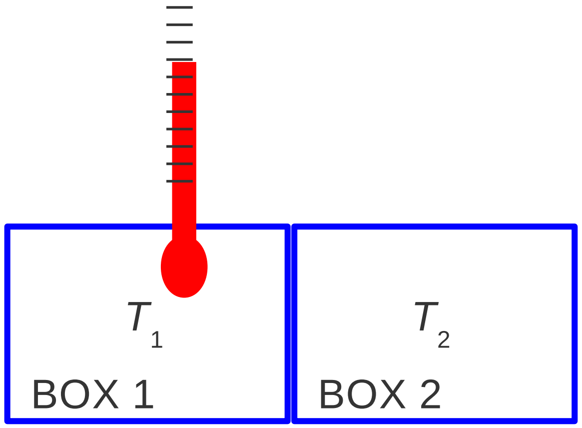
$$\mathbf{x} = \begin{pmatrix} \mathbf{p} \\ \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} p_1 \\ \vdots \\ p_{n/3} \\ u_1 \\ \vdots \\ u_{n/3} \\ v_1 \\ \vdots \\ v_{n/3} \end{pmatrix}, \qquad \mathrm{cov}(\mathbf{x}') = \langle \mathbf{x}' \mathbf{x}'^{\mathrm{T}} \rangle = \begin{pmatrix}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

multivariate covariance sub–matrix

These covariances are symmetrix matrices.

Importance of covariance matrices - graphical demonstration

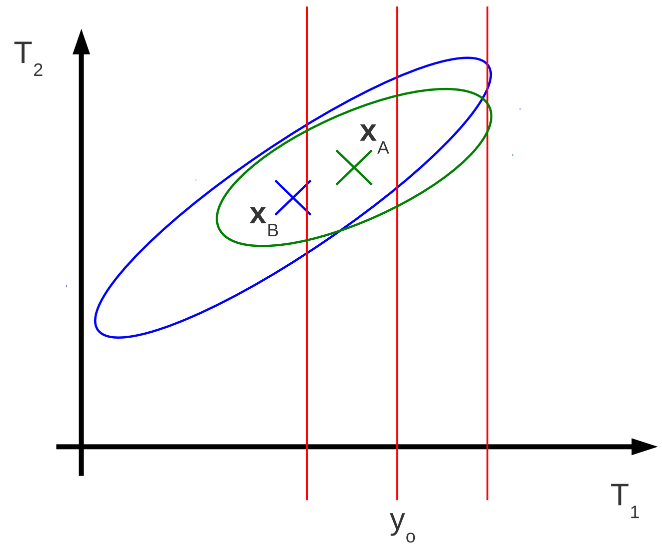
A single observation in a 2-element system $(n = 2,\, p = 1)$.



$$\mathbf{x} = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}, \quad \mathbf{x}_\mathrm{B} = \begin{pmatrix} T_{\mathrm{B}1} \\ T_{\mathrm{B}2} \end{pmatrix}, \quad \mathbf{y}_\mathrm{o} = (y)\,,$$

$$\mathcal{H}(\mathbf{x}) = T_1, \quad \mathbf{H} = \begin{pmatrix} 1 & 0 \end{pmatrix},$$

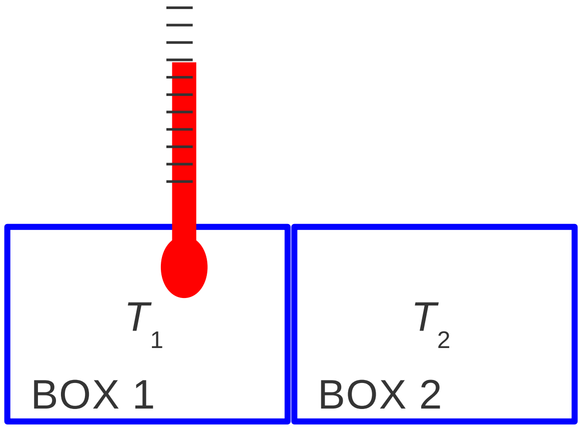
$$\mathbf{P}_\mathrm{f} = \begin{pmatrix} \sigma^2_{\mathrm{B}1} & \alpha \\ \alpha & \sigma^2_{\mathrm{B}2} \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} \sigma^2_\mathrm{o} \end{pmatrix}.$$



$$\begin{aligned} \underbrace{P(\mathbf{x} = \mathbf{x}_\mathrm{t} | \mathbf{y}_\mathrm{o})}_{\text{posterior}} &\propto \underbrace{P(\mathbf{y}_\mathrm{o} | \mathbf{x} = \mathbf{x}_\mathrm{t})}_{\text{likelihood}} \times \underbrace{P(\mathbf{x} = \mathbf{x}_\mathrm{t})}_{\text{prior}}, \\ &\propto \exp -\frac{1}{2} \frac{(y_\mathrm{o} - T_1)^2}{\sigma^2_\mathrm{o}} \times \\ &\quad \exp -\frac{1}{2} \frac{\sigma^2_{\mathrm{B}2}(T_1 - T_{\mathrm{B}1})^2 + \sigma^2_{\mathrm{B}1}(T_2 - T_{\mathrm{B}2})^2 - 2\alpha(T_1 - T_{\mathrm{B}1})(T_2 - T_{\mathrm{B}2})}{2(\sigma^2_{\mathrm{B}1}\sigma^2_{\mathrm{B}2} - \alpha^2)}. \end{aligned}$$

Importance of covariance matrices - mathematical demonstration with a 2-element state vector

A single observation in a 2-element system ($n = 2$, $p = 1$).



The KF formula for the analysis increment is:

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

$$\mathbf{x} = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}, \quad \mathbf{x}_\mathrm{B} = \begin{pmatrix} T_\mathrm{B1} \\ T_\mathrm{B2} \end{pmatrix}, \quad \mathbf{y}_\mathrm{o} = (y)\,, \quad \mathcal{H}(\mathbf{x}) = T_1,$$

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad \mathbf{P}_\mathrm{f} = \begin{pmatrix} \sigma_\mathrm{B1}^2 & \alpha \\ \alpha & \sigma_\mathrm{B2}^2 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} \sigma_\mathrm{o}^2 \end{pmatrix}.$$

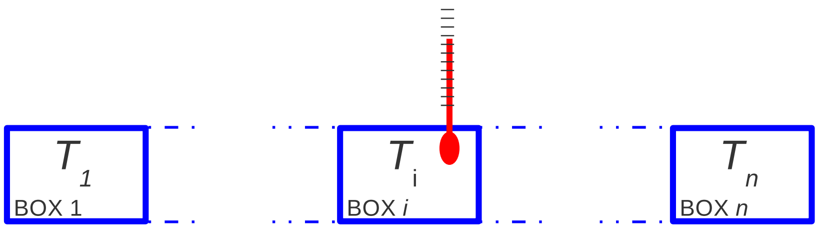
$$\mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T} = \begin{pmatrix} \sigma_\mathrm{B1}^2 & \alpha \\ \alpha & \sigma_\mathrm{B2}^2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \sigma_\mathrm{B1}^2 \\ \alpha \end{pmatrix}, \qquad \mathbf{H}\mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T} = \begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} \sigma_\mathrm{B1}^2 \\ \alpha \end{pmatrix} = (\sigma_\mathrm{B1}^2)\,,$$

$$\mathbf{x}_\mathrm{A} = \begin{pmatrix} T_\mathrm{B1} \\ T_\mathrm{B2} \end{pmatrix} + \begin{pmatrix} \sigma_\mathrm{B1}^2 \\ \alpha \end{pmatrix}\frac{1}{\sigma_\mathrm{o}^2 + \sigma_\mathrm{B1}^2}(y - T_\mathrm{B1})\,.$$

- The analysis increment is a vector $\propto$ the first column of $\mathbf{P}_\mathrm{f}$ (called a structure function or covariance function).
  - The observation of box 1 influences analysis in box 2 because the a-priori errors are correlated (α).
- It is also $\propto$ the innovation $y - T_\mathrm{B1}$.
- If $\sigma_\mathrm{o}^2 \gg \sigma_\mathrm{B1}^2$ then the analysis innovation vanishes.
- If $\sigma_\mathrm{o}^2 \ll \sigma_\mathrm{B1}^2$ then box 1 will be set to the observation value and box 2 will be set to $T_\mathrm{B2} + \alpha(y - T_\mathrm{B1})/\sigma_\mathrm{B1}^2$.

Importance of covariance matrices - mathematical demonstration with a n-element state vector

A single observation in an n-element system ($n = n$, $p = 1$).



$$\mathbf{x} = \begin{pmatrix} T_1 \\ \vdots \\ T_i \\ \vdots \\ T_n \end{pmatrix}, \quad \mathbf{x}_\mathrm{B} = \begin{pmatrix} T_{\mathrm{B}1} \\ \vdots \\ T_{\mathrm{B}i} \\ \vdots \\ T_{\mathrm{B}n} \end{pmatrix}, \quad \mathbf{y}_\mathrm{o} = (y), \quad \mathcal{H}(\mathbf{x}) = T_i,$$

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

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

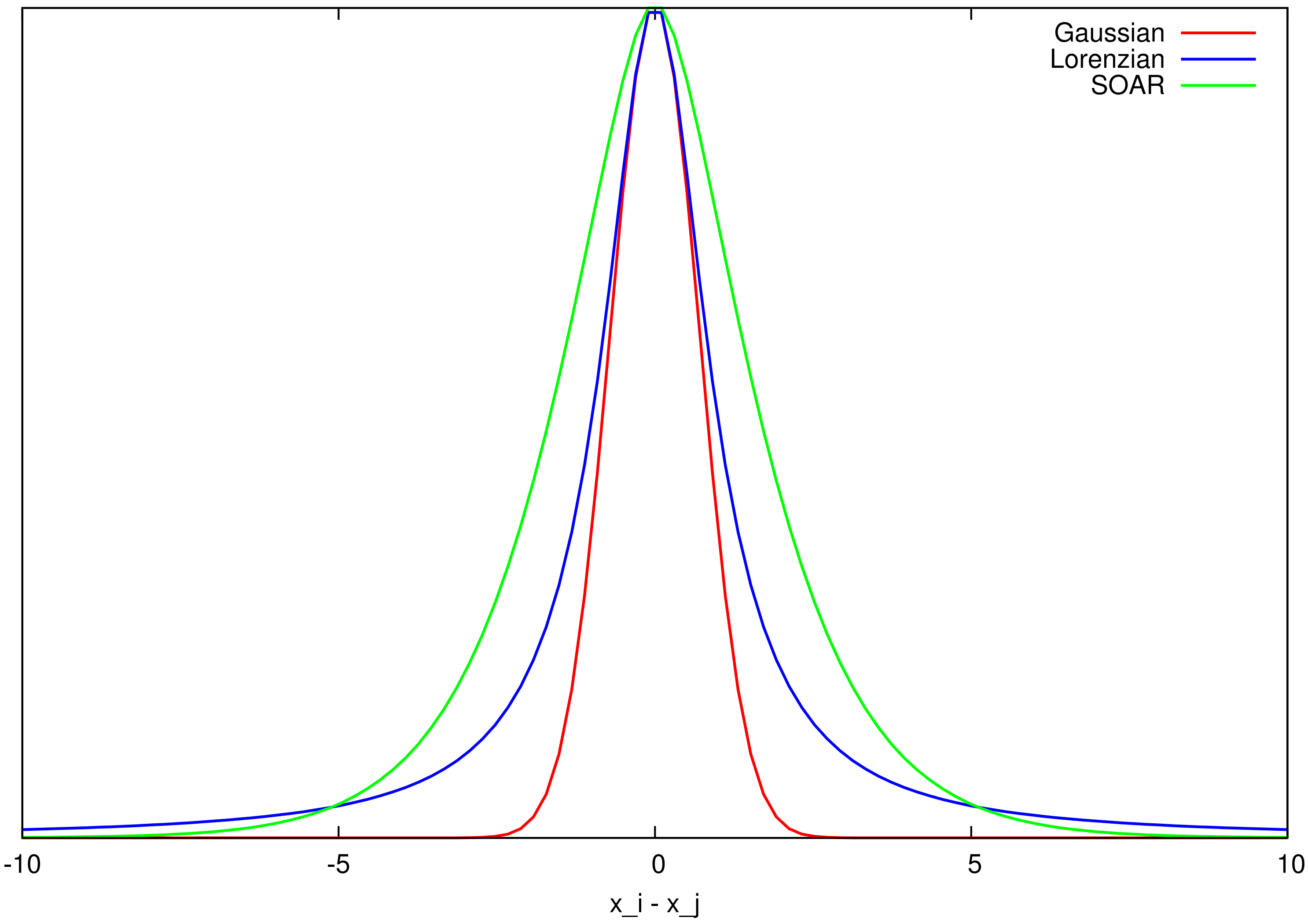
The KF formula for the analysis increment is:

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

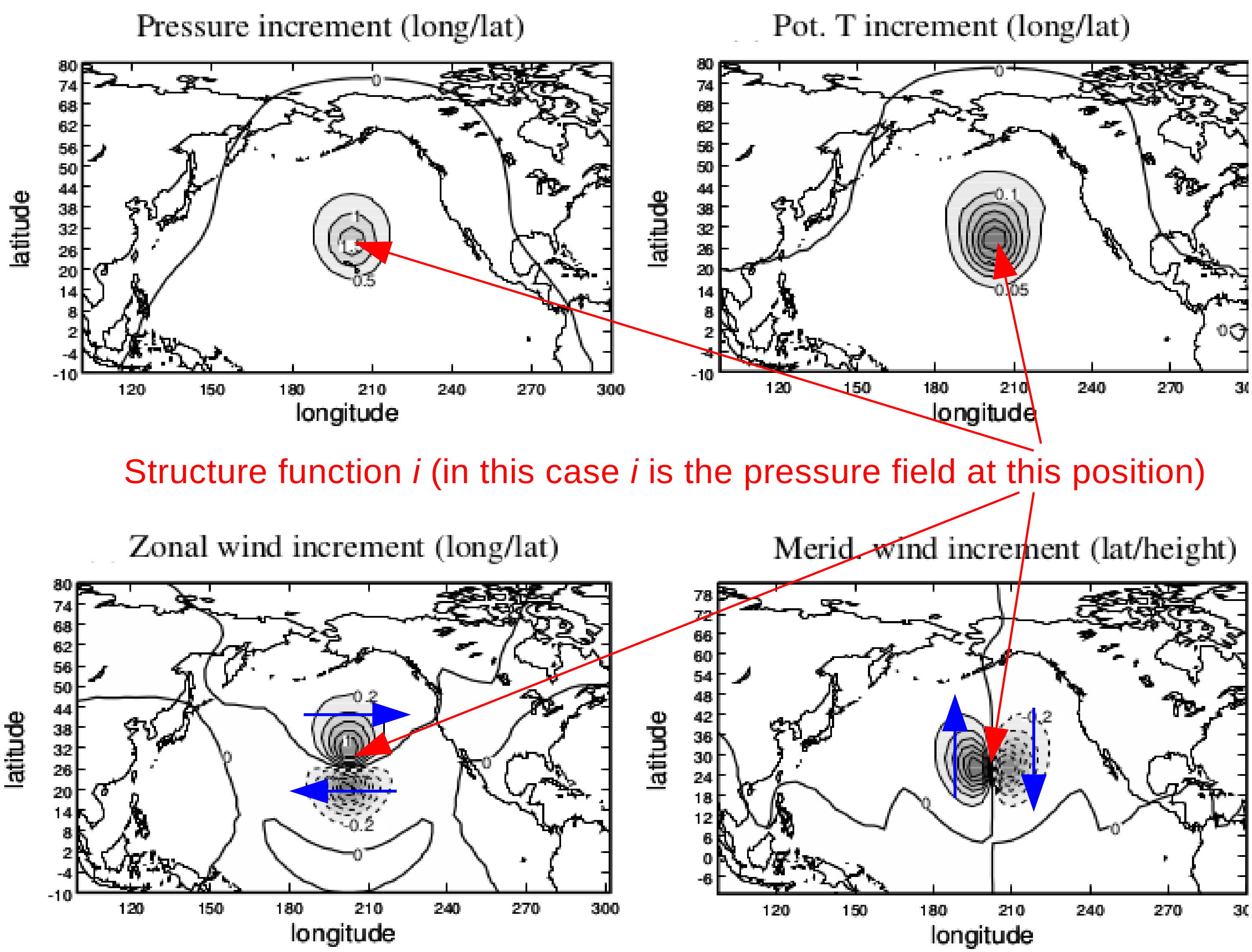
$$\mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T} = \begin{pmatrix} P_{\mathrm{f}11} & \cdots & P_{\mathrm{f}1i} & \cdots & P_{\mathrm{f}1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ P_{\mathrm{f}i1} & \cdots & P_{\mathrm{f}ii} & \cdots & P_{\mathrm{f}in} \\ \vdots & \vdots & \vdots & \ddots & \vdots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

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

- The analysis increment is a vector $\propto$ the ith column of $\mathbf{P}_\mathrm{f}$ (called a structure function or covariance function).
- Structure functions are often parametrised with a particular length-scale L. E.g.:
  - Gaussian shape $P_{\mathrm{f}ij} = \sigma_{\mathrm{B}i}\sigma_{\mathrm{B}j} \exp\left[-(x_i - x_j)^2/L^2\right]$,
  - Lorenzian shape $P_{\mathrm{f}ij} = \sigma_{\mathrm{B}i}\sigma_{\mathrm{B}j} / \left\{1 + \left[(x_i - x_j)^2/L^2\right]\right\}$,
  - SOAR (second order auto-regressive) function $P_{\mathrm{f}ij} = \sigma_{\mathrm{B}i}\sigma_{\mathrm{B}j} \left(1 + |x_i - x_j|/L\right) \exp\left(-|x_i - x_j|/L\right)$.



Structure functions for flow in the mid-latitude atmosphere



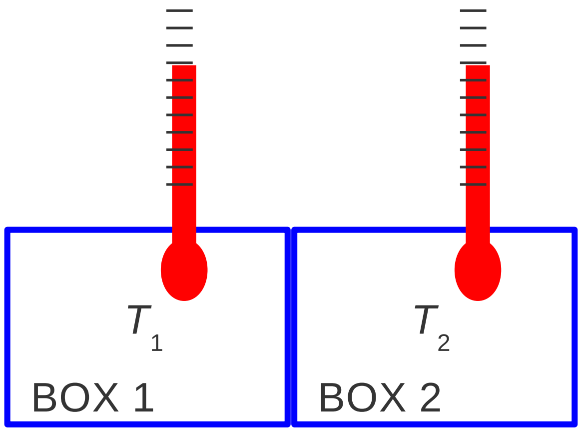
In this case the wind part of the structure function is in geostrophic balance with the pressure

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 ($\mathbf{P}_\mathrm{f}$):
  - Can be rarely represented explicitly.
  - 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 $\mathbf{P}_\mathrm{f}$ is simpler - e.g. diagonal).
- Model error covariance matrices ($\mathbf{Q}$).

Modelling a background error covariance matrix (simple example - related variables and control variables)

System (two grid boxes)



- Same system as before $\mathbf{x} = \begin{pmatrix} T_1 & T_2 \end{pmatrix}^{\mathrm{T}}$.
- Suppose that constraint applies: $T_2 \approx \tau_0 + \mu_0 T_1$ (τ_0 and μ_0 are known constants).
- We have (e.g.) one observation of each temperature, $\mathbf{y} = (y_1, y_2)^{\mathrm{T}}$:

$$\mathbf{y} = \mathbf{H}\mathbf{x}, \qquad \mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Strategy A (assimilate with respect to x directly)

$$\begin{aligned} J[\mathbf{x}] &= \frac{1}{2} \begin{pmatrix} T_1 - T_{\mathrm{B1}} & T_2 - T_{\mathrm{B2}} \end{pmatrix} \mathbf{P}_{\mathrm{f}}^{(T)-1} \begin{pmatrix} T_1 - T_{\mathrm{B1}} \\ T_2 - T_{\mathrm{B2}} \end{pmatrix} + \frac{1}{2} \begin{pmatrix} y_1 - T_1 & y_2 - T_2 \end{pmatrix} \mathbf{R}^{-1} \begin{pmatrix} y_1 - T_1 \\ y_2 - T_2 \end{pmatrix}, \\ &= \frac{1}{2} (\mathbf{x} - \mathbf{x}_{\mathrm{B}})^{\mathrm{T}} \mathbf{P}_{\mathrm{f}}^{(T)-1} (\mathbf{x} - \mathbf{x}_{\mathrm{B}}) + \frac{1}{2} (\mathbf{y} - \mathbf{H}\mathbf{x})^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}). \end{aligned}$$

- Need to know the covariances explicitly.
- This approach does not exploit the constraint.

Simple example (continued)

Strategy A (assimilate with respect to x directly)

$$\text{copied from previous slide: } J[\mathbf{x}] = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}_{\mathrm{B}}\right)^{\mathrm{T}}\mathbf{P}_{\mathrm{f}}^{(T)-1}\left(\mathbf{x}-\mathbf{x}_{\mathrm{B}}\right) + \frac{1}{2}\left(\mathbf{y}-\mathbf{H}\mathbf{x}\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{H}\mathbf{x}\right).$$

Strategy B (use control variables)

- Reminder of the constraint: $T_2 \approx \tau_0 + \mu_0 T_1$.
- Let T_1 have background error variance $\sigma_{T_1}^2$ and let the constraint be written $T_2 = \tau_0 + \mu_0 T_1 + c$ where c has variance σ_c^2.
- Write the problem in terms of control variables χ_1 and χ_2:

$$T_1 = T_{\mathrm{B}1} + \sigma_{T_1}\chi_1, \qquad T_2 = \tau_0 + \mu_0\left(T_{\mathrm{B}1} + \sigma_{T_1}\chi_1\right) + \sigma_c^2\chi_2.$$

- Introduce a control vector: $\boldsymbol{\chi} = \begin{pmatrix} \chi_1 & \chi_2 \end{pmatrix}^{\mathrm{T}}$:

$$\mathbf{x} = \mathcal{U}(\boldsymbol{\chi}), \qquad \mathbf{x} = \mathbf{U}\boldsymbol{\chi} + \boldsymbol{\gamma} = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = \begin{pmatrix} \sigma_{T_1} & 0 \\ \mu_0\sigma_{T_1} & \sigma_c \end{pmatrix}\begin{pmatrix} \chi_1 \\ \chi_2 \end{pmatrix} + \begin{pmatrix} T_{\mathrm{B}1} \\ \tau_0 + \mu_0 T_{\mathrm{B}1} \end{pmatrix}.$$

- $\boldsymbol{\chi}$ has background $\mathbf{0} = \begin{pmatrix} 0 & 0 \end{pmatrix}^{\mathrm{T}}$, and background error covariance $\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$:

$$J[\boldsymbol{\chi}] = \frac{1}{2}\boldsymbol{\chi}^{\mathrm{T}}\boldsymbol{\chi} + \frac{1}{2}\left(\mathbf{y} - \mathbf{H}\mathcal{U}(\boldsymbol{\chi})\right)^{\mathrm{T}}\begin{pmatrix} R_{11} & 0 \\ 0 & R_{22} \end{pmatrix}^{-1}\left(\mathbf{y} - \mathbf{H}\mathcal{U}(\boldsymbol{\chi})\right).$$

- Minimise with respect to $\boldsymbol{\chi} \longrightarrow \boldsymbol{\chi}_{\mathrm{A}}$ giving $\mathbf{x}_{\mathrm{A}} = \mathcal{U}(\boldsymbol{\chi}_{\mathrm{A}})$.

Simple example (continued)

- All we need to know:
  - Background for T_1 (T_{B1}).
  - Background error variance of T_1 ($\sigma^2_{T_1}$).
  - Variance of c (σ^2_c).
- The constraint is applied with a strength specified by σ^2_c.
- The implied background error covariance of $\mathbf{x}$ is:

$$\mathbf{P}_{\mathrm{f}}^{(T,\mathrm{implied})} = \mathbf{U}\mathbf{U}^{\mathrm{T}} = \begin{pmatrix} \sigma_{T_1} & 0 \\ \mu_0\sigma_{T_1} & \sigma_c \end{pmatrix} \begin{pmatrix} \sigma_{T_1} & \mu_0\sigma_{T_1} \\ 0 & \sigma_c \end{pmatrix} = \begin{pmatrix} \sigma^2_{T_1} & \mu_0\sigma^2_{T_1} \\ \mu_0\sigma^2_{T_1} & \mu_0^2\sigma^2_{T_1} + \sigma^2_c \end{pmatrix}.$$

- Meteorological/oceanic data assimilation define a control variable transform with balance conditions.

Summary

- Uncertainty is in everything.
- Uncertainty is described by probabilities.
- All proper data assimilation problems need PDFs.
  - Related via Bayes Theorem.
- The normal distribution is often used to describe PDFs.
  - Mean and (co)variance.
  - Leads to Kalman Filter and variational cost functions.
  - (Co)variances describe the precision of the data (and hence the weight given to the data in DA).
- Have seen that background error covariances have a profound impact on the analysis.
  - Often influenced by physical constraints.
  - Explicit matrix size $n \times n$.
  - Can be modelled.
- Pointers to further information ...

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.*