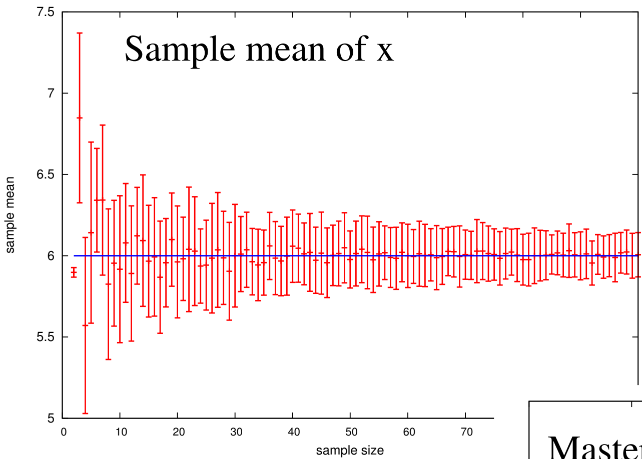
EnKF practical considerations (localization, inflation, etc.)

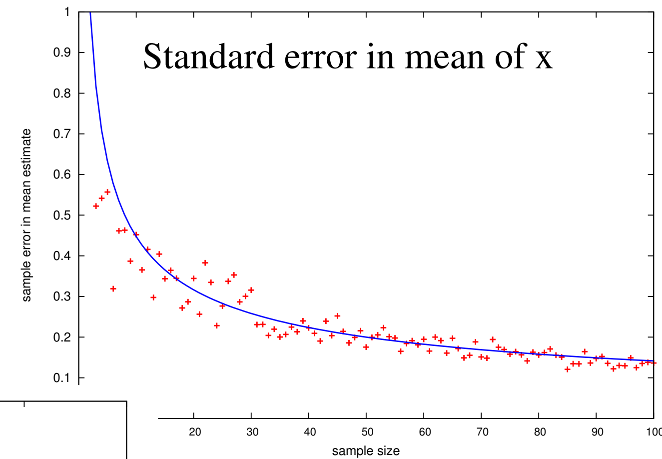
Vn 2.0, Ross Bannister, r.n.bannister@reading.ac.uk

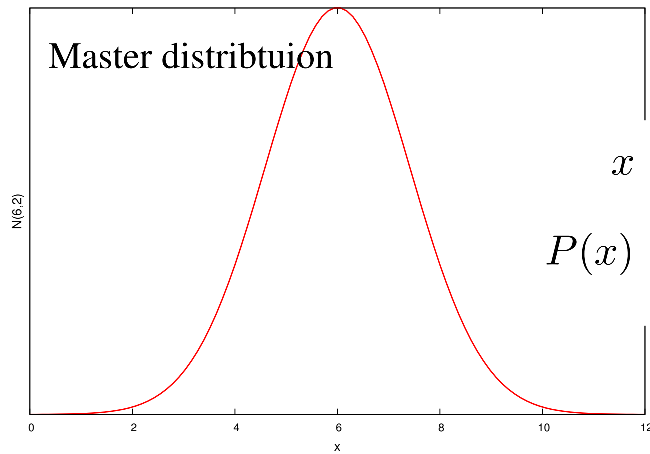
Variational DA vs. ensemble KF

	Variational	**Bare ensemble KF**
Description	Minimize a cost function (maximum a-posteriori)	N ens. members of poss. background and analysis. Analyses derived from KF eqs.
Flavours	1D-Var, 3-D Var, 4-D Var (strong/weak const.)	Basic EnKF, 'square-root' forms, localized filters
Uncertainty	Respect obs and background uncertainty	Respect obs and background uncertainty
Stats	Gaussian	Gaussian
Operators	Allows weakly non-linear $\mathcal{M}$, $\mathcal{H}$ Allows direct and indirect observations Need $\mathbf{M}$, $\mathbf{H}$, and $\mathbf{M}^{\mathrm{T}}$, $\mathbf{H}^{\mathrm{T}}$	Allows weakly non-linear $\mathcal{M}$, $\mathcal{H}$ Allows direct and indirect observations Does not need $\mathbf{M}$, $\mathbf{H}$, and $\mathbf{M}^{\mathrm{T}}$, $\mathbf{H}^{\mathrm{T}}$
Obs types	Direct and indirect observations	Direct and indirect observations
A-priori error stats	$\mathbf{P}_{\mathrm{f}} \rightarrow \mathbf{B}$ (prescribed) $\mathbf{B}$ difficult to determine	$\mathbf{P}_{\mathrm{f}}$ adapts with flow (approx. from ens.) Initial ensemble difficult to determine Ens. tends to be under-spread (filter divergence)
Analysis	Smooth and balanced (according to $\mathbf{B}$) Analysis err. stats can be est. with extra procedures	Sampling noise in ens. leads to noisy analyses Appropriate balance properties Est. of analysis err stats from analysis ens.

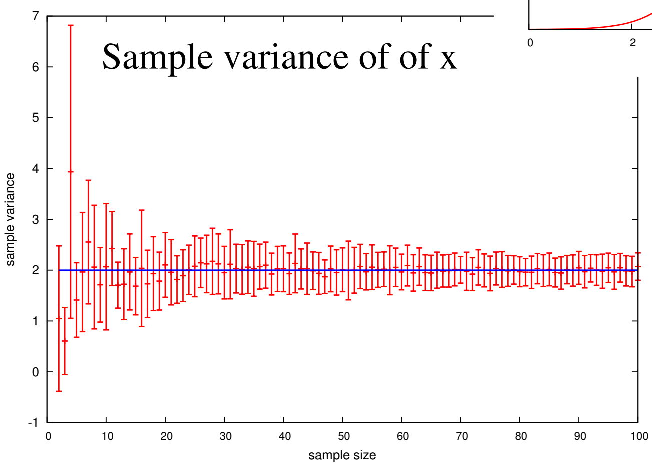
Sampling error for one variable

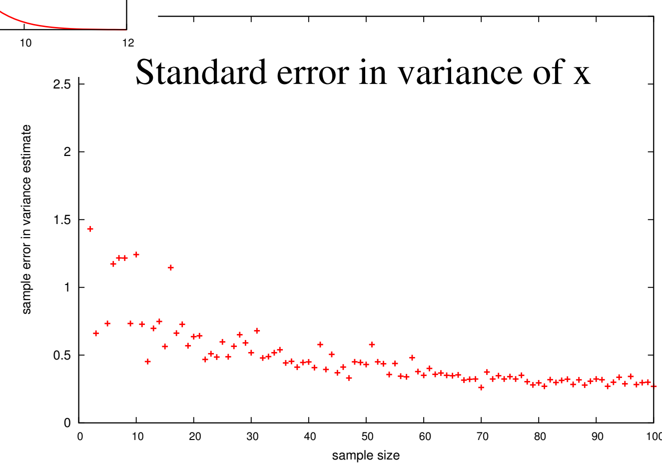






$$
\begin{aligned}
x &\sim N(6,2), \\
P(x) &= \frac{1}{4\pi} \exp -\frac{(x-6)^2}{4}.
\end{aligned}
$$





Sampling error and covariances

Basic ensemble estimate of the forecast error covariance matrix:

1. Take ensemble analysis at $t = -T$ (N ensemble members stored in an $n \times N$ matrix):

$$\mathbf{X}_\mathrm{A}(-T) = \begin{pmatrix} \uparrow & & \uparrow \\ \mathbf{x}_\mathrm{A}^{(1)}(-T) & \cdots & \mathbf{x}_\mathrm{A}^{(N)}(-T) \\ \downarrow & & \downarrow \end{pmatrix}.$$

2. Propagate all members to $t = 0$ (with added noise to represent model error):

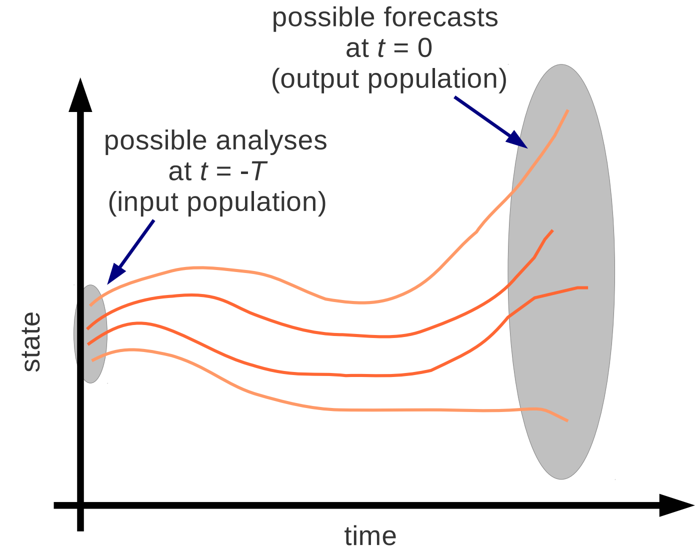
$$\mathbf{X}_\mathrm{f}(0) = \mathcal{M}\left[\mathbf{X}_\mathrm{A}(-T)\right] + \underline{\boldsymbol{\eta}} = \begin{pmatrix} \uparrow & & \uparrow \\ \mathbf{x}_\mathrm{f}^{(1)}(0) & \cdots & \mathbf{x}_\mathrm{f}^{(N)}(0) \\ \downarrow & & \downarrow \end{pmatrix}.$$

3. Calculate perturbations from the mean, $\mathbf{x}_\mathrm{f}'^{(i)} = \mathbf{x}_\mathrm{f}^{(i)} - \langle \mathbf{x}_\mathrm{f} \rangle$ (proxy for forecast errors):

$$\mathbf{X}_\mathrm{f}' = \mathbf{X}_\mathrm{f}'(0) = \begin{pmatrix} \uparrow & & \uparrow \\ \mathbf{x}_\mathrm{f}'^{(0)} & \cdots & \mathbf{x}_\mathrm{f}'^{(N)} \\ \downarrow & & \downarrow \end{pmatrix}.$$

4. Formula for the sample error covariance:

$$\mathbf{P}_\mathrm{f} \approx \mathbf{P}_\mathrm{f}^{(N)} = \frac{1}{N-1}\sum_{i=1}^{N} \mathbf{x}_\mathrm{f}'^{(i)} \mathbf{x}_\mathrm{f}'^{(i)\mathrm{T}}, \quad = \frac{1}{N-1}\mathbf{X}_\mathrm{f}'\mathbf{X}_\mathrm{f}'^{\mathrm{T}}.$$



This matrix is not calculated explicitly for large systems, but we can use the formula to explore the consequences of $N \ll n$.

Sampling error and covariances (continued)

Reminder

$$\text{Analysis increment formula for member } i: \mathbf{x}_\mathrm{A}^{(i)} - \mathbf{x}_\mathrm{f}^{(i)} = \mathbf{P}_\mathrm{f}^{(N)}\mathbf{H}^\mathrm{T}\left(\mathbf{R} + \mathbf{H}\mathbf{P}_\mathrm{f}^{(N)}\mathbf{H}^\mathrm{T}\right)^{-1}\left(\mathbf{y}_\mathrm{o}^{(i)} - \mathcal{H}(\mathbf{x}_\mathrm{f}^{(i)})\right),$$
$$= \mathbf{P}_\mathrm{f}^{(N)}\mathbf{v}^{(i)},$$
$$\text{Reminder: sample forecast error cov. matrix: } \mathbf{P}_\mathrm{f}^{(N)} = \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{x}_\mathrm{f}'^{(i)}\mathbf{x}_\mathrm{f}'^{(i)\mathrm{T}}.$$

1. Analysis increments $(\mathbf{x}_\mathrm{A}^{(i)} - \mathbf{x}_\mathrm{f}^{(i)})$ lie in the subspace of the forecast error ensemble

Approximate $\mathbf{P}_\mathrm{f}$ with $\mathbf{P}_\mathrm{f}^{(N)}$ in the analysis increment formula:

$$\mathbf{x}_\mathrm{A}^{(i)} - \mathbf{x}_\mathrm{B}^{(i)} \approx \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{x}_\mathrm{f}'^{(i)}\mathbf{x}_\mathrm{f}'^{(i)\mathrm{T}}\mathbf{v}^{(i)},$$
$$\approx \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{x}_\mathrm{f}'^{(i)}\alpha^{(i)},$$
$$\text{where } \alpha^{(i)} = \mathbf{x}_\mathrm{f}'^{(i)\mathrm{T}}\mathbf{v}^{(i)} = \mathbf{x}_\mathrm{f}'^{(i)}\cdot\mathbf{v}^{(i)}.$$

Even if the observations indicate otherwise, the analysis increments are restricted to be a linear combination of the forecast error ensemble.

2. The forecast error covariance matrix is rank deficient

The rank of $\mathbf{P}_{\mathrm{f}}^{(N)}$ is an indication of the size of the state space spanned by the forecast error ensemble.

$$\mathrm{rank}\left(\mathbf{P}_{\mathrm{f}}^{(N)}\right) \leq N - 1.$$

This is a guide to the severity of the sampling problem in point 1.

3. The forecast ensemble spread will be subject to sampling error

- If the spread is too large then the analysis ens. will over-fit the obs. - too little trust in the fc. ens.
- If the spread is too small then the analysis ens. will under-fit the obs. - too much trust in the fc. ens.
  - Once in this regime, it is difficult to escape as the ens. will (effectively) ignore the obs..
  - This is called "filter divergence" (because we diverge from reality).

Filter divergence means that each ensemble member will (effectively) be free running.

4. The correlations will be subject to sampling error

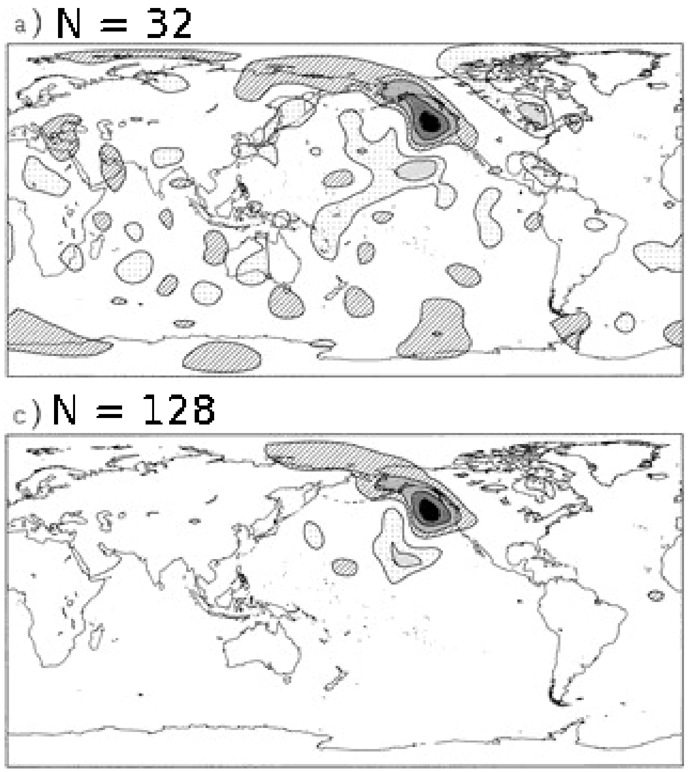
- The error in the sample correlation between errors at locations i and j has expectation:

$$[\mathcal{E}(\delta\mathbf{C}_{\mathrm{f}}^{(N)})]_{ij} \sim \frac{1}{\sqrt{N}}\left(1 - ([\mathbf{C}_{\mathrm{f}}]_{ij})^2\right),$$

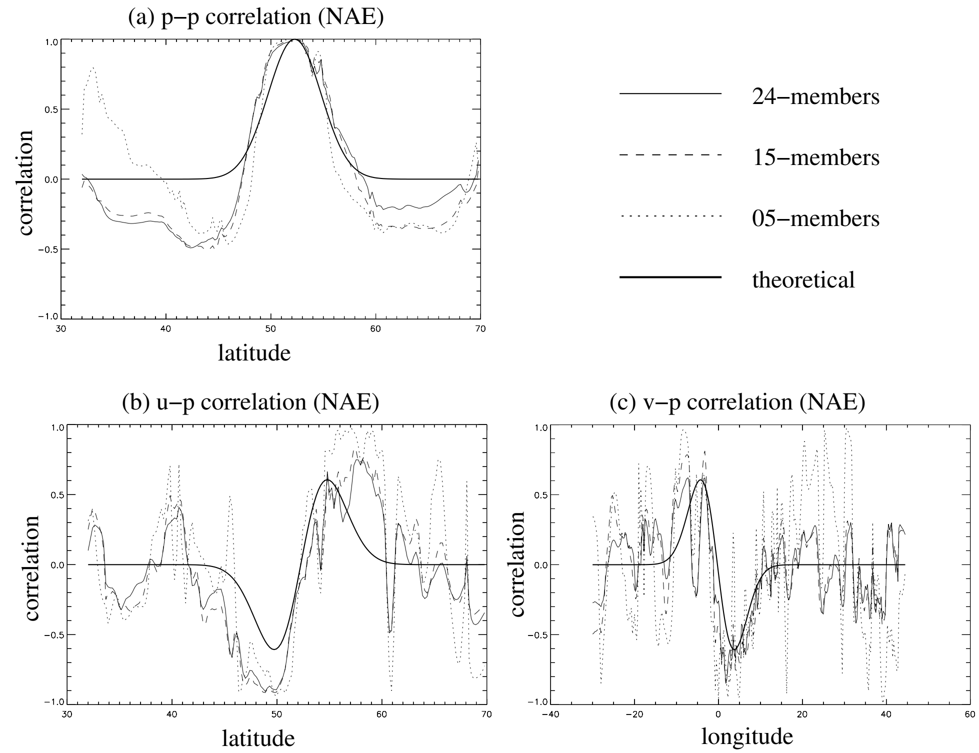
  (errors are expected to be large when N small and/or correlations are close to zero).

- Pairs of distant points would be expected to have correlations close to zero.

Sampling error means that we can't trust distant correlations. Left untreated this noise will destroy the benefits of DA (analysis increments will be influenced by distant observations).



From Houtekamer & Mitchell, 1998

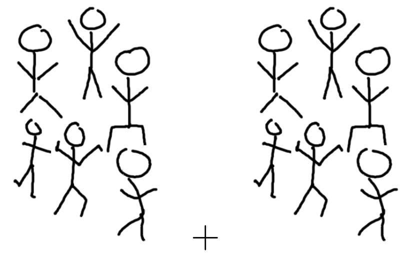


From Bannister, Migliorini & Dixon, 2011

Making progress

What can be done to reduce/mitigate this problem $N \ll n$?

- **Use more ensemble members**.



  - This is expensive.
  - How many is 'enough'?

- **Ensemble inflation**.

  - Artificially increase the size of each $\mathbf{x}_{\mathrm{f}}'^{(i)}$.
  - How do we know what the ensemble spread should be?

- **Localization**.

  - Eliminate far-field correlations.
  - How should this be done?
  - Does this have any other consequences?

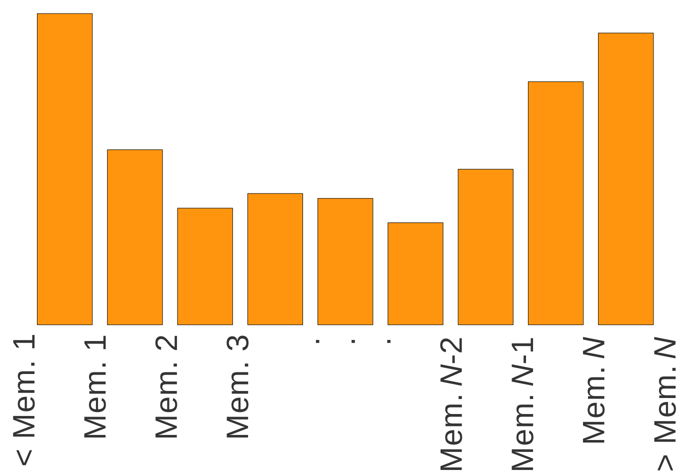
- **Combine ensemble with variational approaches**.

  - Adopt a hybrid method.
  - How to do this?

Ensemble inflation: how do we know what the ens spread should be?

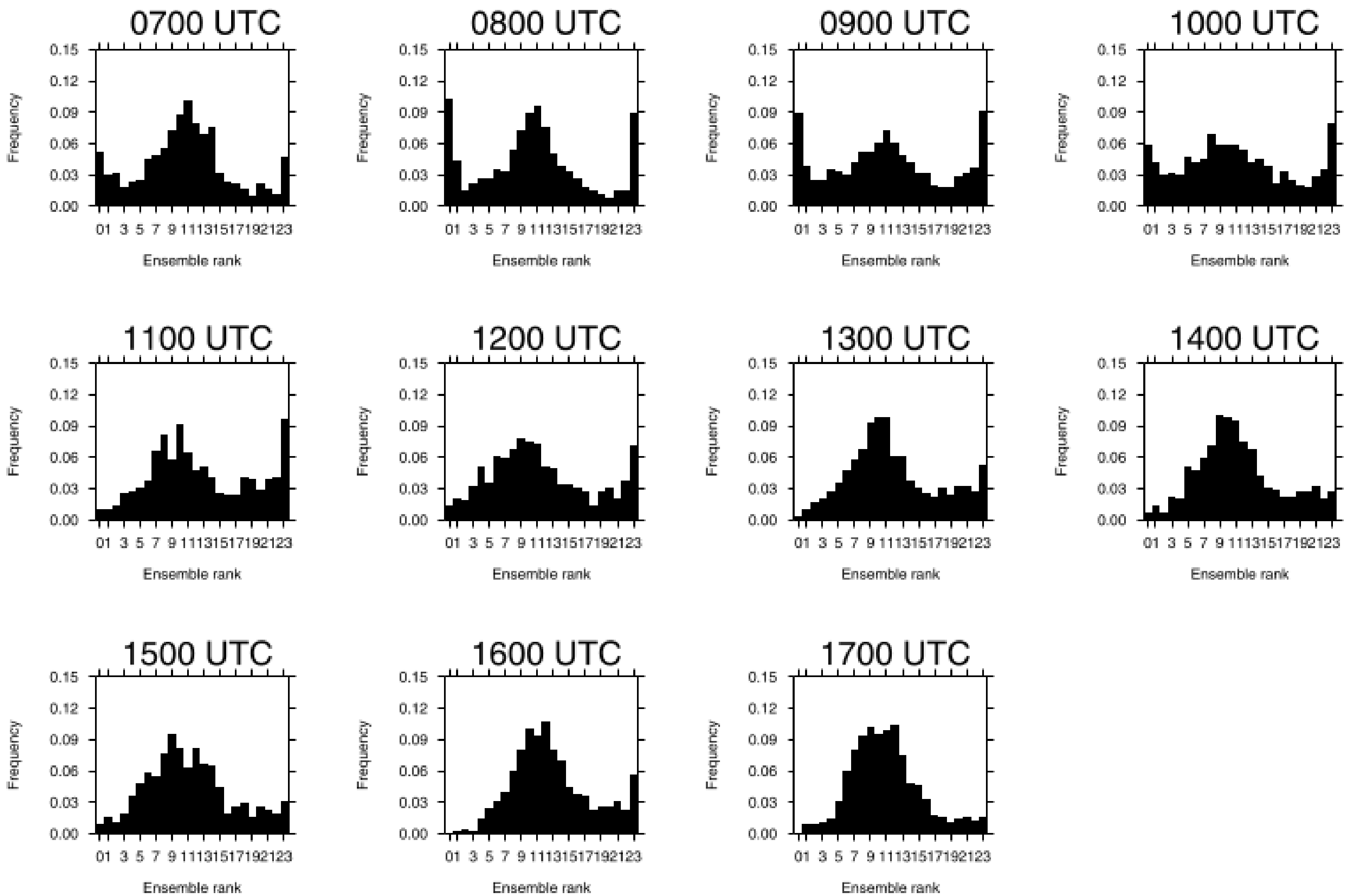
Rank histograms (Talagrand diagrams)

- Each ensemble member should be equally likely.
- Consider a point in space that has many observations:
  - Rank values of ensemble members at that point from lowest to highest ($N-1$ bins). Add an extra bin at each end to give $N+1$ bins.
  - Bin each observation to give a frequency histogram.
- Interpretation:
  - $\cup$-shaped: the ensemble is under-spread.
  - $\cap$-shaped: the ensemble is over-spread.
  - Flat: the ensemble is correctly spread.
  - Asymmetric: the ensemble is biased.



Ensemble inflation: how do we know what the ens spread should be?

Example rank histograms



Rank histograms for surface precipitation rate rate. From Migliorini et al., 2011.

Ensemble inflation: how do we know what the ens spread should be?

Spread/skill diagrams

Suppose that we have an ensemble of N forecasts, $\mathbf{x}^{(i)}$ spread around a background state, $\mathbf{x}_{\mathrm{b}}$.

1. For each ob (ob index j):

   (a) Calculate ens variance at the ob location [where $y^{\mathrm{m}}_{ij} = H_j(\mathbf{x}^{(i)})$ and $\overline{y^{\mathrm{m}}_j} = \frac{1}{N}\sum_{i=1}^{N} H_j(\mathbf{x}^{(i)})$]:

   $$\sigma^2_{\mathrm{ens},j} = \frac{1}{N-1}\sum_{i=1}^{N}\left(y^{\mathrm{m}}_{ij} - \overline{y^{\mathrm{m}}_j}\right)^2 .$$

   (b) Calculate the innovation [where $y^{\mathrm{m}}_{\mathrm{b}j} = H_j(\mathbf{x}_{\mathrm{b}})$]:

   $$d_j = y_j - y^{\mathrm{m}}_{\mathrm{b}j}.$$

2. Bin the results according to the ens var (let there be M obs per bin).

3. For each bin (bin index k):

   (a) Calculate the mean ens var:

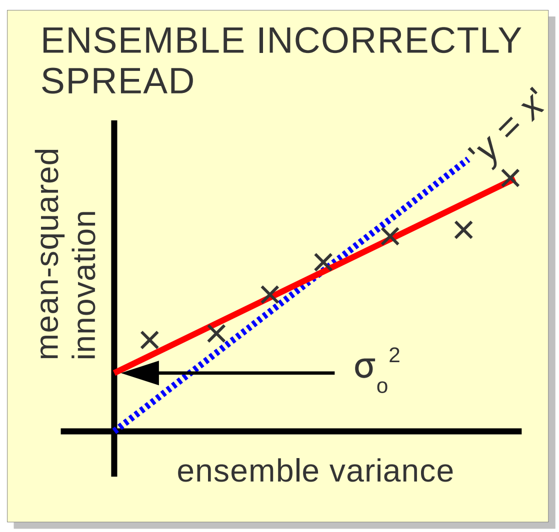
   $$\sigma^2_{\mathrm{ens}}(k) = \left\langle \sigma^2_{\mathrm{ens},j}\right\rangle_{\mathrm{bin}\ k} .$$

   (b) Calculate the innovation mean-square:

   $$\sigma^2_{\mathrm{innov}}(k) = \left\langle d^2_j\right\rangle_{\mathrm{bin}\ k} .$$

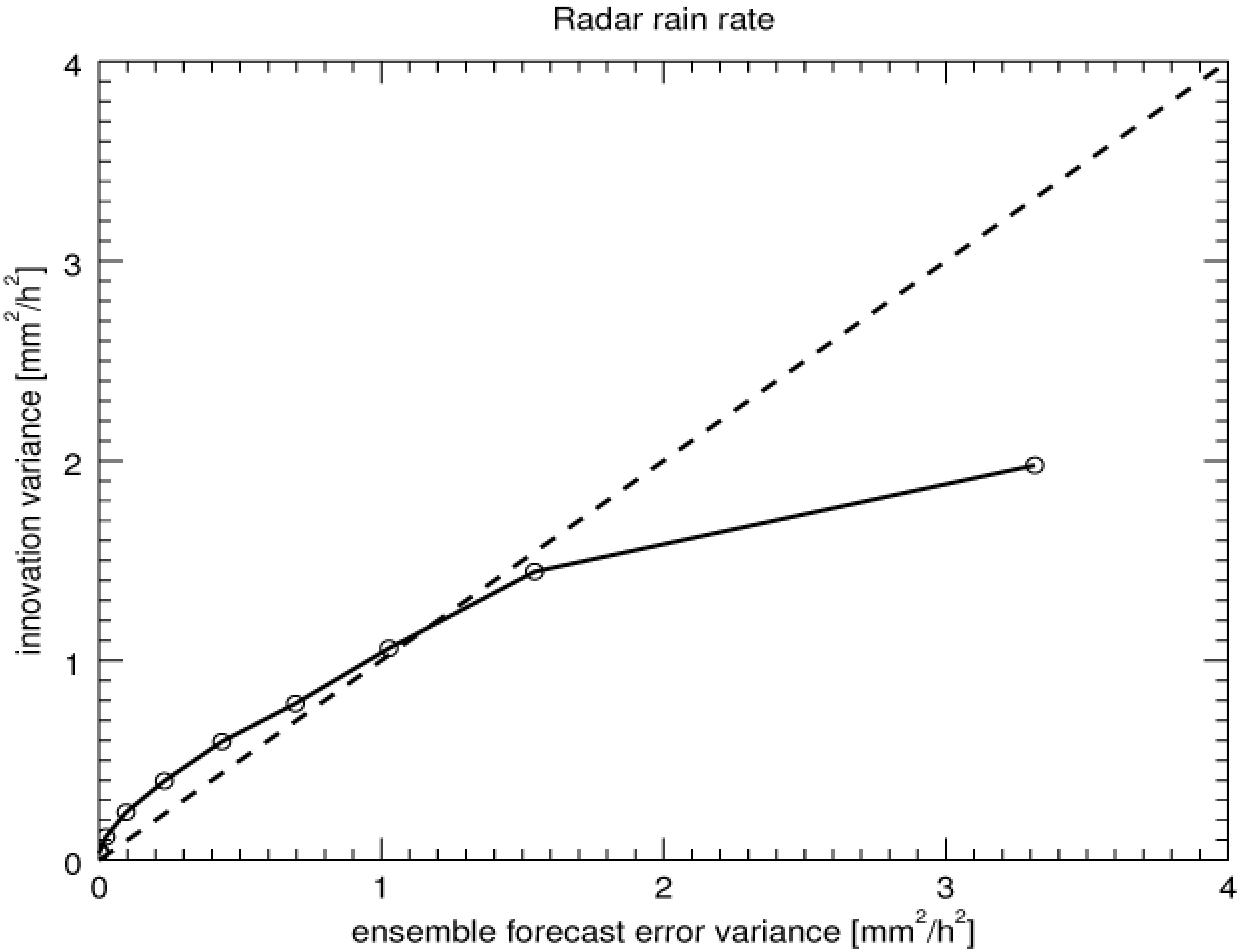
4. Plot $\sigma^2_{\mathrm{innov}}(k)$ (the 'skill') against $\sigma^2_{\mathrm{ens}}(k)$ (the spread).

5. Can use this information to derive an *inflation factor*.



Ensemble inflation: how do we know what the ens spread should be?

Example spread/skill diagram



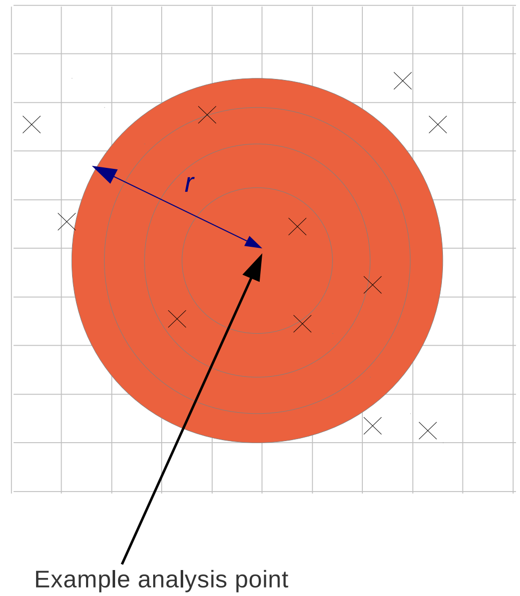
From Baker et al., 2014

Localisation

Many ways of doing localisation, e.g.:

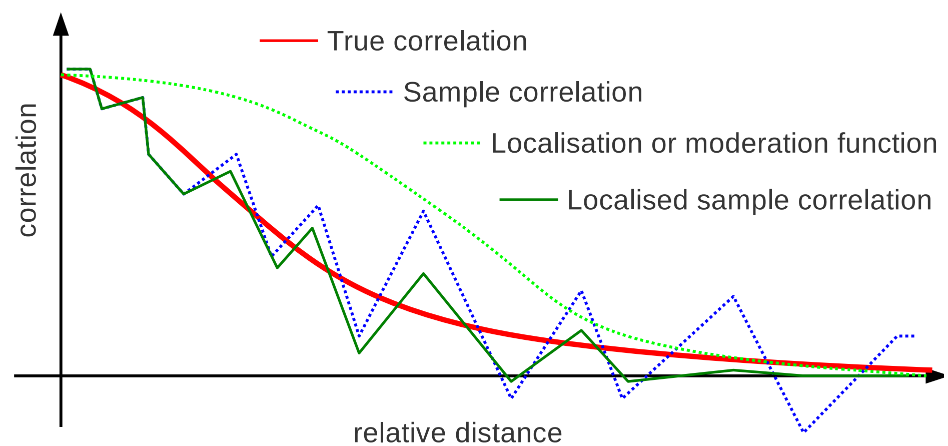
$\mathbf{R}$-localisation:

- Perform a separate ens analysis at each grid point.
- Include obs inside a defined radius. Divide obs error variance by a weight, ρ (decreases with distance).
- Used in the ETKF.
- Difficult to use for non-local observations.

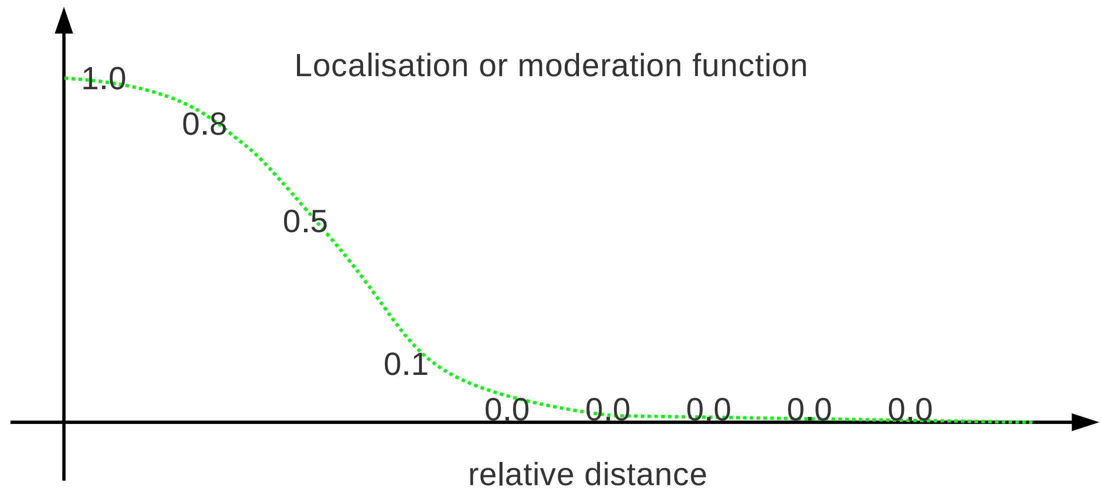


$\mathbf{P}_\mathrm{f}$-localisation:

- Modify $\mathbf{P}_\mathrm{f}^{(N)}$ with a localisation/moderation function that decreases with separation.
- What length-scale? How to do multivariate aspects?
- Has side effects (e.g. affects length-scales, affects balance).



$\mathrm{P_f}$-localisation (univariate)



$$
\begin{aligned}
\mathbf{P}_\mathrm{f}^{\mathrm{Loc}} &= \mathbf{P}_\mathrm{f} \circ \mathbf{\Omega}, \\
&= \begin{pmatrix}
P_{\mathrm{f}11}^{(N)} & P_{\mathrm{f}12}^{(N)} & \cdots & \cdots & P_{\mathrm{f}15}^{(N)} & \cdots & \cdots & P_{\mathrm{f}18}^{(N)} & P_{\mathrm{f}19}^{(N)} \\
P_{\mathrm{f}21}^{(N)} & P_{\mathrm{f}22}^{(N)} & \cdots & \cdots & P_{\mathrm{f}25}^{(N)} & \cdots & \cdots & P_{\mathrm{f}28}^{(N)} & P_{\mathrm{f}29}^{(N)} \\
\vdots & \vdots & \ddots & \cdots & P_{\mathrm{f}35}^{(N)} & \cdots & \cdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & P_{\mathrm{f}45}^{(N)} & \vdots & \vdots & \vdots & \vdots \\
P_{\mathrm{f}51}^{(N)} & P_{\mathrm{f}52}^{(N)} & P_{\mathrm{f}53}^{(N)} & P_{\mathrm{f}54}^{(N)} & P_{\mathrm{f}55}^{(N)} & P_{\mathrm{f}56}^{(N)} & P_{\mathrm{f}57}^{(N)} & P_{\mathrm{f}58}^{(N)} & P_{\mathrm{f}59}^{(N)} \\
\vdots & \vdots & \vdots & \vdots & P_{\mathrm{f}65}^{(N)} & \ddots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \cdots & \cdots & P_{\mathrm{f}75}^{(N)} & \cdots & \ddots & \vdots & \vdots \\
P_{\mathrm{f}81}^{(N)} & P_{\mathrm{f}82}^{(N)} & \cdots & \cdots & P_{\mathrm{f}85}^{(N)} & \cdots & \cdots & P_{\mathrm{f}88}^{(N)} & P_{\mathrm{f}89}^{(N)} \\
P_{\mathrm{f}91}^{(N)} & P_{\mathrm{f}92}^{(N)} & \cdots & \cdots & P_{\mathrm{f}95}^{(N)} & \cdots & \cdots & P_{\mathrm{f}98}^{(N)} & P_{\mathrm{f}99}^{(N)}
\end{pmatrix}
\circ
\begin{pmatrix}
1.0 & 0.8 & 0.5 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.8 & 1.0 & 0.8 & 0.5 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.5 & 0.8 & 1.0 & 0.8 & 0.5 & 0.1 & 0.0 & 0.0 & 0.0 \\
0.1 & 0.5 & 0.5 & 1.0 & 0.8 & 0.5 & 0.1 & 0.0 & 0.0 \\
0.0 & 0.1 & 0.1 & 0.8 & 1.0 & 0.8 & 0.5 & 0.1 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.5 & 0.8 & 1.0 & 0.8 & 0.5 & 0.1 \\
0.0 & 0.0 & 0.0 & 0.1 & 0.5 & 0.8 & 1.0 & 0.8 & 0.5 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.5 & 0.8 & 1.0 & 0.8 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.5 & 0.8 & 1.0
\end{pmatrix}, \\
P_{\mathrm{f}ij}^{\mathrm{Loc}} &= P_{\mathrm{f}ij}\Omega_{ij}.
\end{aligned}
$$

Can be extended to multivariate localisation. But - we rarely have access to explicit $\mathbf{P}_\mathrm{f}$ or $\mathbf{\Omega}$ matrices $(n \times n)$.

Localisation without explicit $\mathbf{P}_\mathrm{f}$ and $\mathbf{\Omega}$ matrices

$$\text{Sample forecast error cov. matrix from } N \text{ members}: \mathbf{P}_\mathrm{f}^{(N)} = \frac{1}{N-1}\sum_{l=1}^{N} \mathbf{x}_\mathrm{f}^{\prime(l)} \mathbf{x}_\mathrm{f}^{\prime(l)\mathrm{T}},$$

$$\text{Sample localisation/moderation matrix from } K \text{ members}: \mathbf{\Omega}^{(K)} = \frac{1}{K-1}\sum_{k=1}^{K} \boldsymbol{\omega}^{(k)} \boldsymbol{\omega}^{(k)\mathrm{T}},$$

$$\text{One matrix element: } P_{\mathrm{f}ij}^{(N)} = \frac{1}{N-1}\sum_{l=1}^{N} x_{\mathrm{f}i}^{\prime(l)} x_{\mathrm{f}j}^{\prime(l)},$$

$$\text{One matrix element: } \Omega_{ij}^{(K)} = \frac{1}{K-1}\sum_{k=1}^{K} \omega_i^{(k)} \omega_j^{(k)}.$$

$$\begin{aligned}
P_{\mathrm{f}ij}^{\mathrm{Loc}} &= P_{\mathrm{f}ij}^{(N)} \Omega_{ij}^{(K)}, \\
&= \left[\frac{1}{N-1}\sum_{l=1}^{N} x_{\mathrm{f}i}^{\prime(l)} x_{\mathrm{f}j}^{\prime(l)}\right] \left[\frac{1}{K-1}\sum_{k=1}^{K} \omega_i^{(k)} \omega_j^{(k)}\right], \\
&= \frac{1}{N-1}\frac{1}{K-1}\sum_{l=1}^{N}\sum_{k=1}^{K} \underbrace{x_{\mathrm{f}i}^{\prime(l)} \omega_i^{(k)}}_{\substack{\text{element } i \\ \text{of } \tilde{\mathbf{x}}^{\prime(m)}}} \underbrace{x_{\mathrm{f}j}^{\prime(l)} \omega_j^{(k)}}_{\substack{\text{element } j \\ \text{of } \tilde{\mathbf{x}}^{\prime(m)}}}, \\
&= \frac{1}{N-1}\frac{1}{K-1}\sum_{m=1}^{M} \tilde{x}_i^{\prime(m)} \tilde{x}_j^{\prime(m)}, \qquad M = NK, \\
\tilde{\mathbf{x}}^{\prime(m)} &= \mathbf{x}_\mathrm{f}^{\prime(l)} \circ \boldsymbol{\omega}^{(k)} = \begin{pmatrix} x_{\mathrm{f}1}^{\prime(l)} \\ \vdots \\ x_{\mathrm{f}n}^{\prime(l)} \end{pmatrix} \circ \begin{pmatrix} \omega_1^{(k)} \\ \vdots \\ \omega_n^{(k)} \end{pmatrix} = \begin{pmatrix} x_{\mathrm{f}1}^{\prime(l)} \omega_1^{(k)} \\ \vdots \\ x_{\mathrm{f}n}^{\prime(l)} \omega_n^{(k)} \end{pmatrix},
\end{aligned}$$

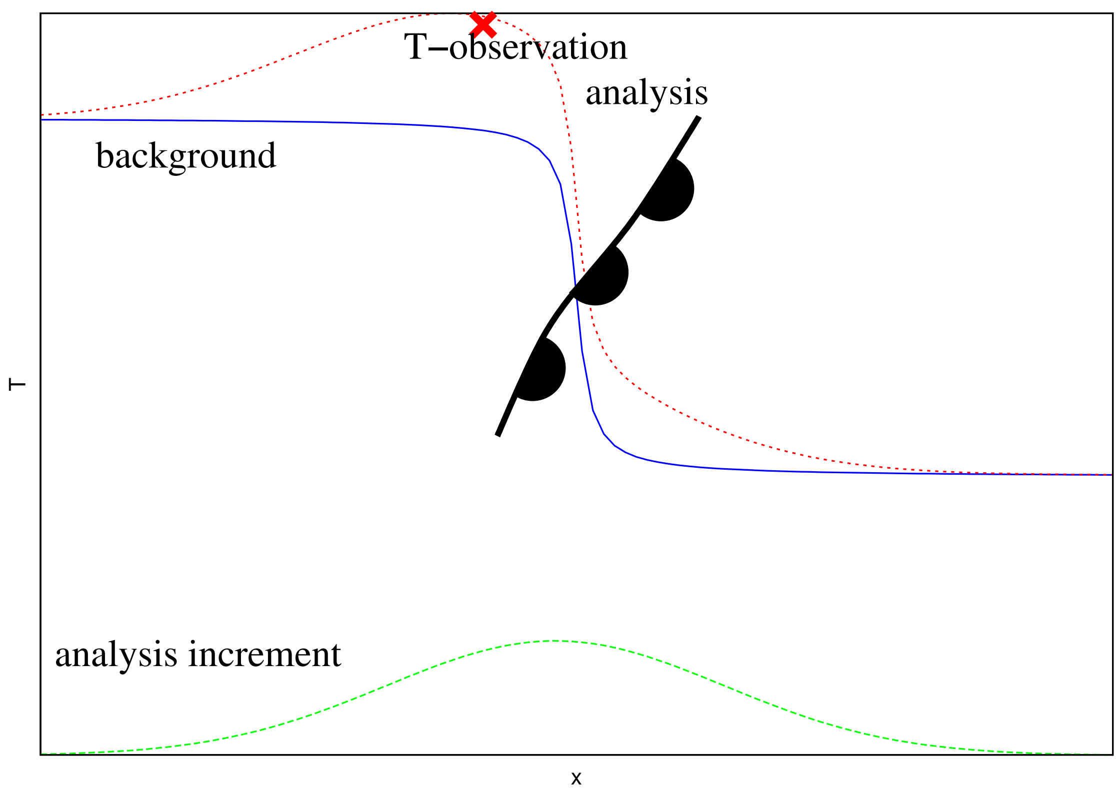
$$m = 1 \Rightarrow (l=1, k=1) \quad m = 2 \Rightarrow (l=1, k=2) \quad \ldots \quad m = K \Rightarrow (l=1, k=K) \quad m = K+1 \Rightarrow (l=2, k=1) \quad \ldots$$

Hybrid (Var and ensemble) data assimilation

Key differences between variational and ensemble methods (background error statistics)

- **Variational**: $\mathbf{P}_\mathrm{f} \to \mathbf{B}$: $\mathbf{B}$ is full rank but is static
- **Ensemble KF**: $\mathbf{P}_\mathrm{f} \to \mathbf{P}_\mathrm{f}^{(N)}$: $\mathbf{P}_\mathrm{f}^{(N)}$ is low rank, but is flow-dependent.

Flow-dependence is important!

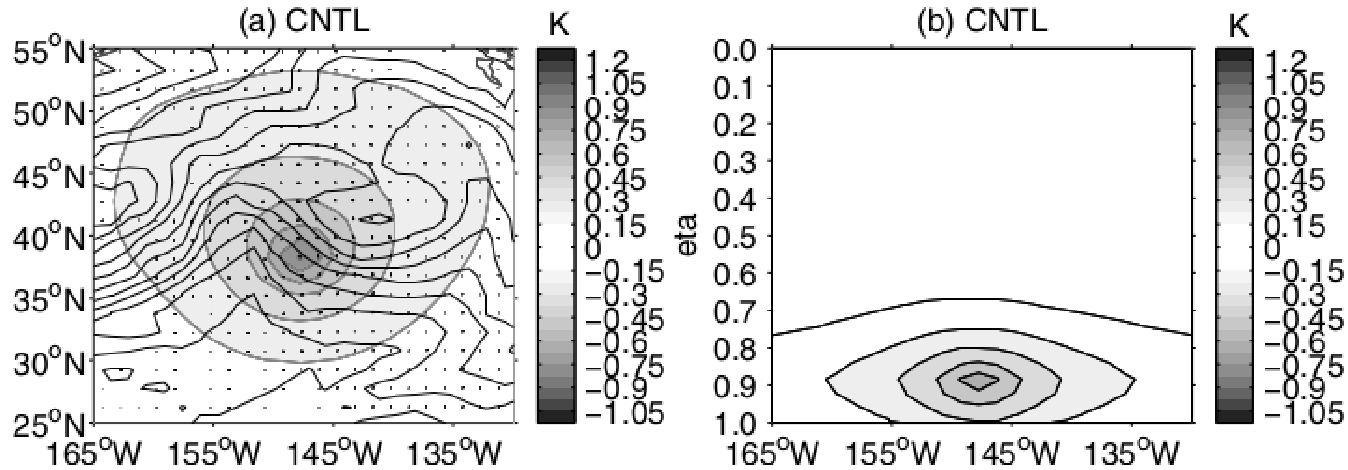


The $\mathbf{B}$-matrix doesn't know about the front in basic variational schemes. The $\mathbf{P}_\mathrm{f}^{(N)}$-matrix does know about the front.

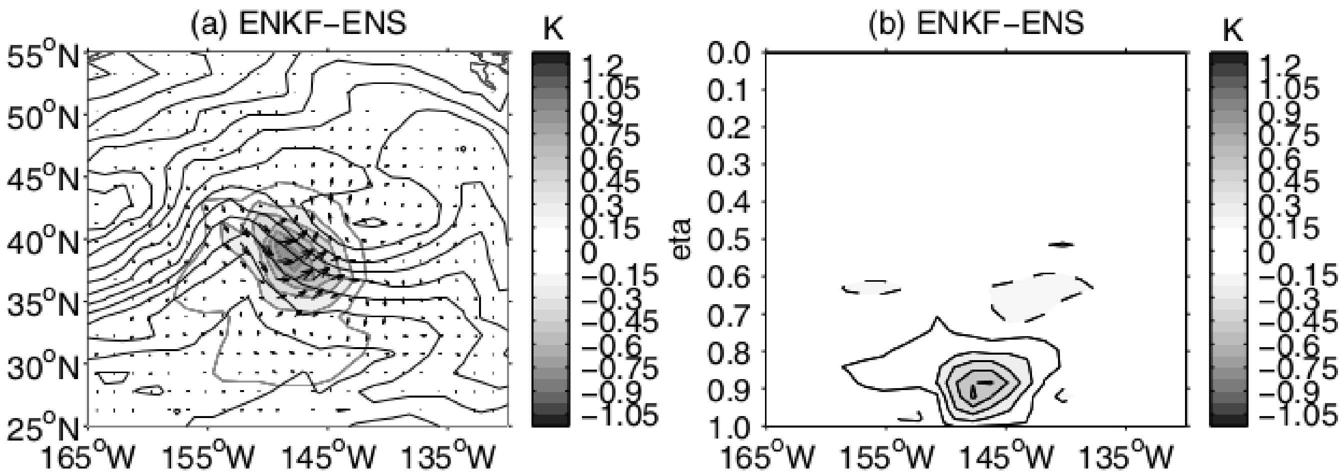
Flow-dependent structure functions

Longitude/latitude and longitude/pressure analysis increments of a single T observation

Variational



Ensemble



From Buehner 2005

Hybrid (Var and ensemble) data assimilation (continued)

Propose doing a variational assimilation, but:

$$\mathbf{B} \rightarrow \alpha\mathbf{B} + (1-\alpha)\mathbf{P}_{\mathrm{f}}^{(N)} \text{ (unlocalised)}, \qquad \mathbf{B} \rightarrow \alpha\mathbf{B} + (1-\alpha)\mathbf{P}_{\mathrm{f}}^{\mathrm{Loc}} \text{ (localised)}, \qquad 0 \leq \alpha \leq 1.$$

A basic variational hybrid scheme (example for unlocalised case)

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

$$\underbrace{J[\boldsymbol{\chi}_{\mathrm{var}},\boldsymbol{\chi}_{\mathrm{ens}}]}_{\text{hybrid}} = \frac{1}{2}\underbrace{\boldsymbol{\chi}_{\mathrm{var}}^{\mathrm{T}}\boldsymbol{\chi}_{\mathrm{var}}}_{\text{static term}} + \frac{1}{2}\underbrace{\boldsymbol{\chi}_{\mathrm{ens}}^{\mathrm{T}}\boldsymbol{\chi}_{\mathrm{ens}}}_{\text{ensemble}} + \frac{1}{2}(\mathbf{y}_{\mathrm{o}}-\mathcal{H}\{\mathcal{U}[\boldsymbol{\chi}_{\mathrm{var}},\boldsymbol{\chi}_{\mathrm{ens}}]\})\mathbf{R}^{-1}(\mathbf{y}_{\mathrm{o}}-\mathcal{H}\{\mathcal{U}[\boldsymbol{\chi}_{\mathrm{var}},\boldsymbol{\chi}_{\mathrm{ens}}]\}).$$

- $\boldsymbol{\chi}_{\mathrm{var}}$ is part of the hybrid variational control vector associated with $\mathbf{B}$ (n elements, b/g variance $\mathbf{I}$).
- $\boldsymbol{\chi}_{\mathrm{ens}}$ is another part associated with the ensemble (N elements, b/g variance $\mathbf{I}$).

$$\begin{aligned}
\mathbf{x} = \mathcal{U}[\boldsymbol{\chi}_{\mathrm{var}},\boldsymbol{\chi}_{\mathrm{ens}}] &= \mathbf{x}_{\mathrm{f}} + \sqrt{\alpha}\mathbf{B}^{1/2}\boldsymbol{\chi}_{\mathrm{var}} + \sqrt{\frac{1-\alpha}{N-1}}\mathbf{X}_{\mathrm{f}}'\boldsymbol{\chi}_{\mathrm{ens}}, \qquad \mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{\mathrm{T}/2} \\
&= \mathbf{x}_{\mathrm{f}} + \begin{pmatrix} \sqrt{\alpha}\mathbf{B}^{1/2} & \sqrt{\frac{1-\alpha}{N-1}}\mathbf{X}_{\mathrm{f}}' \end{pmatrix} \begin{pmatrix} \boldsymbol{\chi}_{\mathrm{var}} \\ \boldsymbol{\chi}_{\mathrm{ens}} \end{pmatrix}, \\
&= \mathbf{x}_{\mathrm{f}} + \mathbf{U} \begin{pmatrix} \boldsymbol{\chi}_{\mathrm{var}} \\ \boldsymbol{\chi}_{\mathrm{ens}} \end{pmatrix}.
\end{aligned}$$

The cost function $J[\boldsymbol{\chi}_{\mathrm{var}},\boldsymbol{\chi}_{\mathrm{ens}}]$ is minimised with respect to $\boldsymbol{\chi}_{\mathrm{var}}$ and $\boldsymbol{\chi}_{\mathrm{ens}}$ simultaneously ($n+N$ elements). The implied background error covariance matrix is:

$$\mathbf{B}^{\mathrm{implied}} = \mathbf{U}\mathbf{U}^{\mathrm{T}} = \begin{pmatrix} \sqrt{\alpha}\mathbf{B}^{1/2} & \sqrt{\frac{1-\alpha}{N-1}}\mathbf{X}_{\mathrm{f}}' \end{pmatrix} \begin{pmatrix} \sqrt{\alpha}\mathbf{B}^{\mathrm{T}/2} \\ \sqrt{\frac{1-\alpha}{N-1}}\mathbf{X}_{\mathrm{f}}'^{\mathrm{T}} \end{pmatrix} = \alpha\mathbf{B} + (1-\alpha)\mathbf{P}_{\mathrm{f}}^{(N)}.$$

Summary

- Ensemble data assimilation schemes suffer from sampling error as $N \ll n$:
  - Analysis increments lie in subspace of ensemble.
  - Rank deficiency.
  - Filter divergence.
  - Anomalous far-field correlations.
- To make ensemble DA practical:
  - Ensemble inflation.
  - Localization.
  - Use with other schemes (hybrid).

Further Reading- selected papers and websites

Ensemble KF vs variational data assimilation, Schur product localisation

- Lorenc A.C., The potential of the ensemble Kalman filter for NWP - a comparison with 4d-Var, Q.J.R.Meteor.Soc. 129 3183-3203 (2003).
- Ehrendorfer M., A review of issues in ensemble-based Kalman filtering, Meteorol. Z. 16, 795-818 (2007).

Impact of sampling error

- Houtekamer P.L., Mitchell H.L., Data assimilation using an ensemble Kalman Filter technique, Mon. Wea. Rev. 126 796-811 (1998).

Inflation

- Bowler N.E., Arribas A., Mylne K.R., Robertson K.B., Beare S.E., The MOGREPS short-range ensemble prediction system, Q.J.R.Meteor.Soc. 134, 703-722 (2008).

Hybrid formulations

- Clayton A.M., Lorenc A.C., Barker D.M., Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office, Q.J.R.Meteor.Soc. DOI:10.1002/qj.2054.
- Wang X., Snyder C., Hamill T.M., On the theoretical equivalence of differently proposed ensemble-3D-Var hybrid analysis schemes, Mon. Wea. Rev. 135., 222-227 (2007).
- Buehner M., Ensemble derived stationary and flow dependent background error covariances: Evaluation in a quasi-operational NWP setting, Q.J.R.Meteor.Soc. 131, 1013-1043 (2005).

Forecast verification

- www.cawcr.gov.au/projects/verification

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., Q.J. Roy. Met. Soc. 134, 1971-1996 (2008).