Relative Merits of 4D-Var and Ensemble Kalman Filter

Andrew Lorenc

Met Office, Exeter

Met Office

Outline

Consider 4D-Var & EnKF for NWP applications

Look at how they represent covariances, and hence their expected properties

- Incremental 4D-Var as a 4D covariance model
- EnKF sampling of covariances
- Compare assimilation characteristics
- Ease of implementation
- Two ways forward

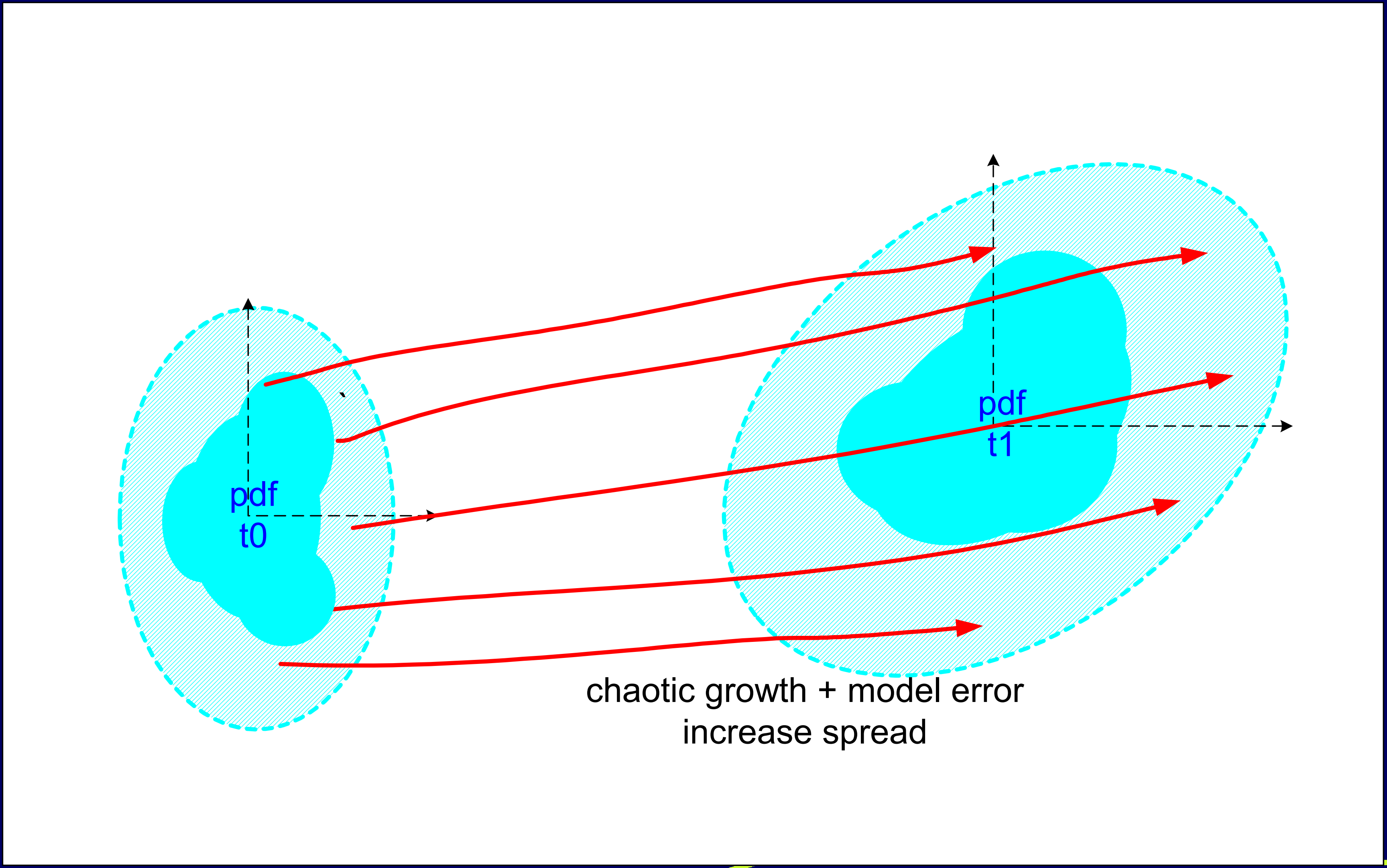
Data Assimilation

- Represent
  - in terms of NWP model variables
  - Uncertainty ⇒ PDF of all variables
- Evolve
  - using physically based NWP equations
  - should propagate uncertainty & allow for model error ⇒ Fokker-Planck equation
- Combine using Bayes theorem
  - characterise observation uncertainties
  - PDFs are unknowable!

Data Assimilation

- Represent
  - in terms of NWP model variables
  - Uncertainty ⇒ PDF of all variables
- Evolve
  - using physically based NWP equations
  - should propagate uncertainty & allow for model error ⇒ Fokker-Planck equation
- Combine using Bayes theorem
  - characterise observation uncertainties
  - PDFs are unknowable!

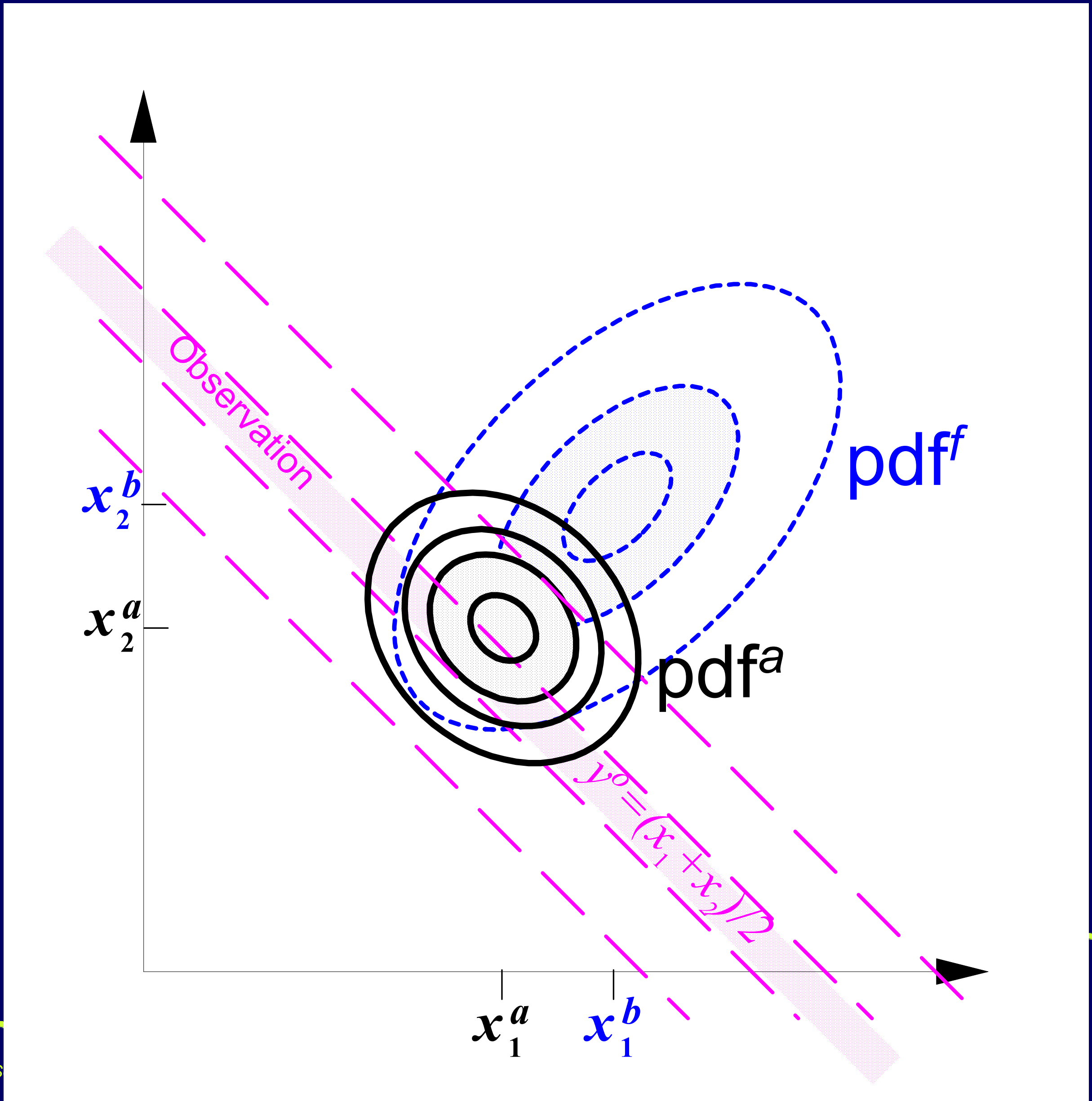
Fokker-Planck Equation



Data Assimilation

- Represent
  - in terms of NWP model variables
  - Uncertainty $\Rightarrow$ PDF of all variables
- Evolve
  - using physically based NWP equations
  - should propagate uncertainty & allow for model error $\Rightarrow$ Fokker-Planck equation
- Combine using Bayes theorem
  - characterise observation uncertainties
  - PDFs are unknowable!

Bayes' Theorem



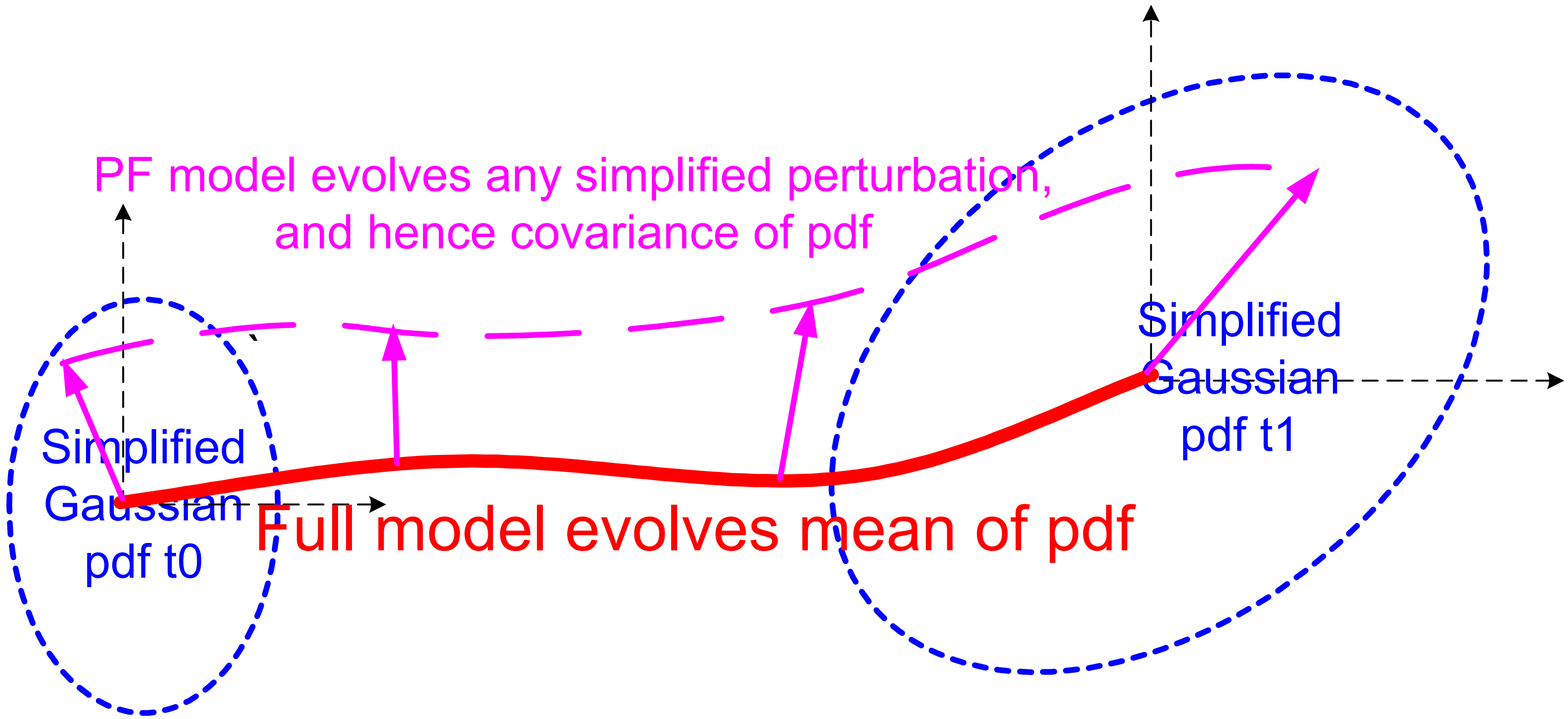
Flaws in “traditional” 4D-Var derivation

- The atmosphere can be chaotic at most scales, some with very short timescales.
- The 4D-Var penalty function for a “perfect” model will be fractal because of the chaotic scales represented, so a descent algorithm cannot work.
- There is no basis for saying the maximum of a complicated PDF is the best analysis.

Synoptic-scale Incremental 4D-Var

- Assume background PDFs are Gaussian
- Reduce dimensionality
- Linear evolution
- Builds on existing 3D-Var
- Can be thought of as a 4D (time & space) PDF describing uncertainty, for use in Bayesian fit to all observations in a 4D time-window.

Incremental 4D-Var



Incremental 4D-Var

- 3D-Var supplies 3D covariance at t=0, consistent with dynamical balance relationships
- PF model evolves this in time, to create a 4D covariance consistent with PF equations
- Do a 4D fit to observations in the time-window
- Covariances define relative weighting, interpolation and extrapolation of observations in space & time
- Covariances (thro' null-space) define classes of 4D analysis increments which are not allowed
  (e.g. unbalanced, or inconsistent with PF equations)

Incremental 4D-Var Equations

$\delta\underline{\mathbf{x}}$ is 4D increment to 4D guess: $$\underline{\mathbf{x}} = \underline{\mathbf{x}}^g + \delta\underline{\mathbf{x}}$$

Want 4D fit, minimising:

$$J(\delta\underline{\mathbf{x}}) = \tfrac{1}{2}\left(\delta\underline{\mathbf{x}} - \delta\underline{\mathbf{x}}^b\right)^T \mathbf{B}_{(\underline{\mathbf{x}})}^{-1}\left(\delta\underline{\mathbf{x}} - \delta\underline{\mathbf{x}}^b\right) + \tfrac{1}{2}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)^T \mathbf{R}^{-1}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)$$

$\underline{\mathbf{y}}$ is prediction of obs in time window: $$\underline{\mathbf{y}} = H(\underline{\mathbf{x}})$$

$\mathbf{v}$ is transformed control variable: $$\delta\underline{\mathbf{x}} = S^{-I}\underline{\mathbf{M}}\mathbf{U}\mathbf{v}$$

S^{-I} is incrementing operator,
$\underline{\mathbf{M}}$ is Perturbation Forecast model,
$\mathbf{U}$ is 3DVAR variable transform,
$S^{-I}\underline{\mathbf{M}}\mathbf{U}$ models 4D covariance $\mathbf{B}_{(\underline{\mathbf{x}})}$:

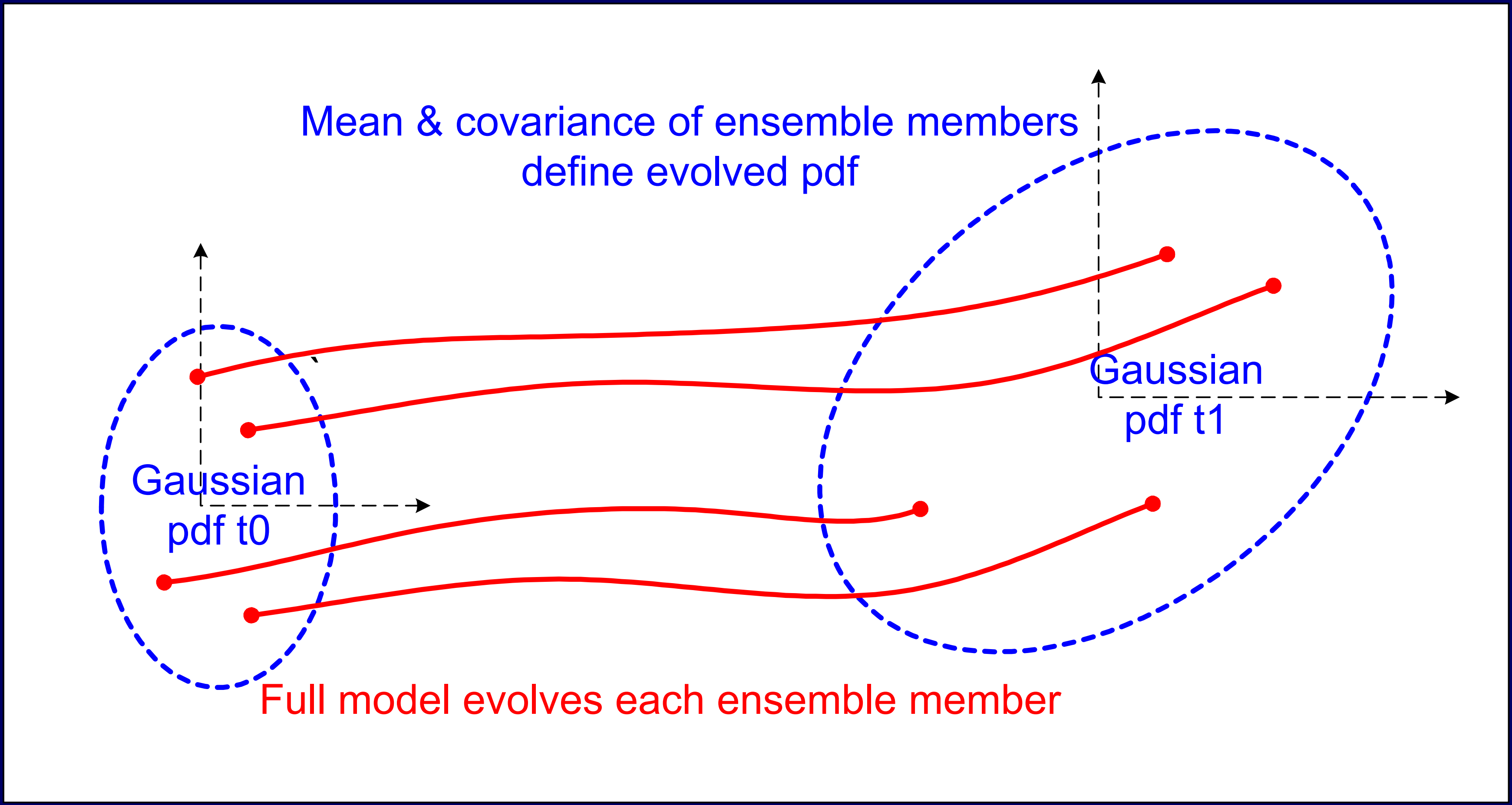
$$\mathbf{B}_{(\underline{\mathbf{x}})} = S^{-I}\underline{\mathbf{M}}\mathbf{U}\mathbf{U}^T\underline{\mathbf{M}}^T S^{-T}$$

Transformed minimisation: $$J(\mathbf{v}) = \tfrac{1}{2}\mathbf{v}^T\mathbf{v} + \tfrac{1}{2}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)^T \mathbf{R}^{-1}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)$$

Benefits of 4D-Var

- Retains benefits of 3D-Var:
  - assimilation of radiances, good balance, …
- Better than 3D-Var in using obs where there are tendencies represented by PF model
  - e.g. baroclinic developments $\Rightarrow$ severe weather
- Scope for better assimilation of cloud and ppn

Ensemble Kalman Filter



Extended Kalman Filter

$\mathbf{x}$ is mean of PDF, $\mathbf{P}$ is covariance.

Analysis step

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{P}^f(t_i)\mathbf{H}_i^T\left(\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right)^{-1}\left(\mathbf{y}_i^o - H_i\left(\mathbf{x}^f(t_i)\right)\right)$$

$$\mathbf{P}^a(t_i) = \mathbf{P}^f(t_i) - \mathbf{P}^f(t_i)\mathbf{H}_i^T\left(\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right)^{-1}\mathbf{H}_i\mathbf{P}^f(t_i)$$

Forecast step

$$\mathbf{x}^f(t_{i+1}) = M_i\left(\mathbf{x}^a(t_i)\right)$$

True discretised dynamics $\mathbf{x}^t$ assumed to differ by stochastic perturbations:

$$\mathbf{x}^t(t_{i+1}) = M_i\left(\mathbf{x}^t(t_i)\right) + \boldsymbol{\eta}(t_i)$$

where $\boldsymbol{\eta}$ is a noise process with zero mean and covariance matrix $\mathbf{Q}_i$.

$$\mathbf{P}^f(t_{i+1}) = \mathbf{M}_i\mathbf{P}^a(t_i)\mathbf{M}_i^T + \mathbf{Q}_i$$

Met Office

The Ensemble Kalman Filter (EnKF)

Construct an ensemble $\{\mathbf{x}_i^f\}, (i = 1, \_\_\_, N)$:

$$\mathbf{P}^f = \mathbf{P}_e^f = \overline{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)^T},$$

$$\mathbf{P}^f\mathbf{H}^T = \overline{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(H\left(\mathbf{x}^f\right) - \overline{H\left(\mathbf{x}^f\right)}\right)^T},$$

$$\mathbf{H}\mathbf{P}^f\mathbf{H}^T = \overline{\left(H\left(\mathbf{x}^f\right) - \overline{H\left(\mathbf{x}^f\right)}\right)\left(H\left(\mathbf{x}^f\right) - \overline{H\left(\mathbf{x}^f\right)}\right)^T}$$

Use these in the standard KF equation to update the best estimate (ensemble mean):

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - H\left(\overline{\mathbf{x}^f}\right)\right).$$

A form of square-root filter:

Matrix $\mathbf{X}$ has columns $\frac{1}{\sqrt{N-1}}\left(x_i - \overline{x}\right)$.

Then

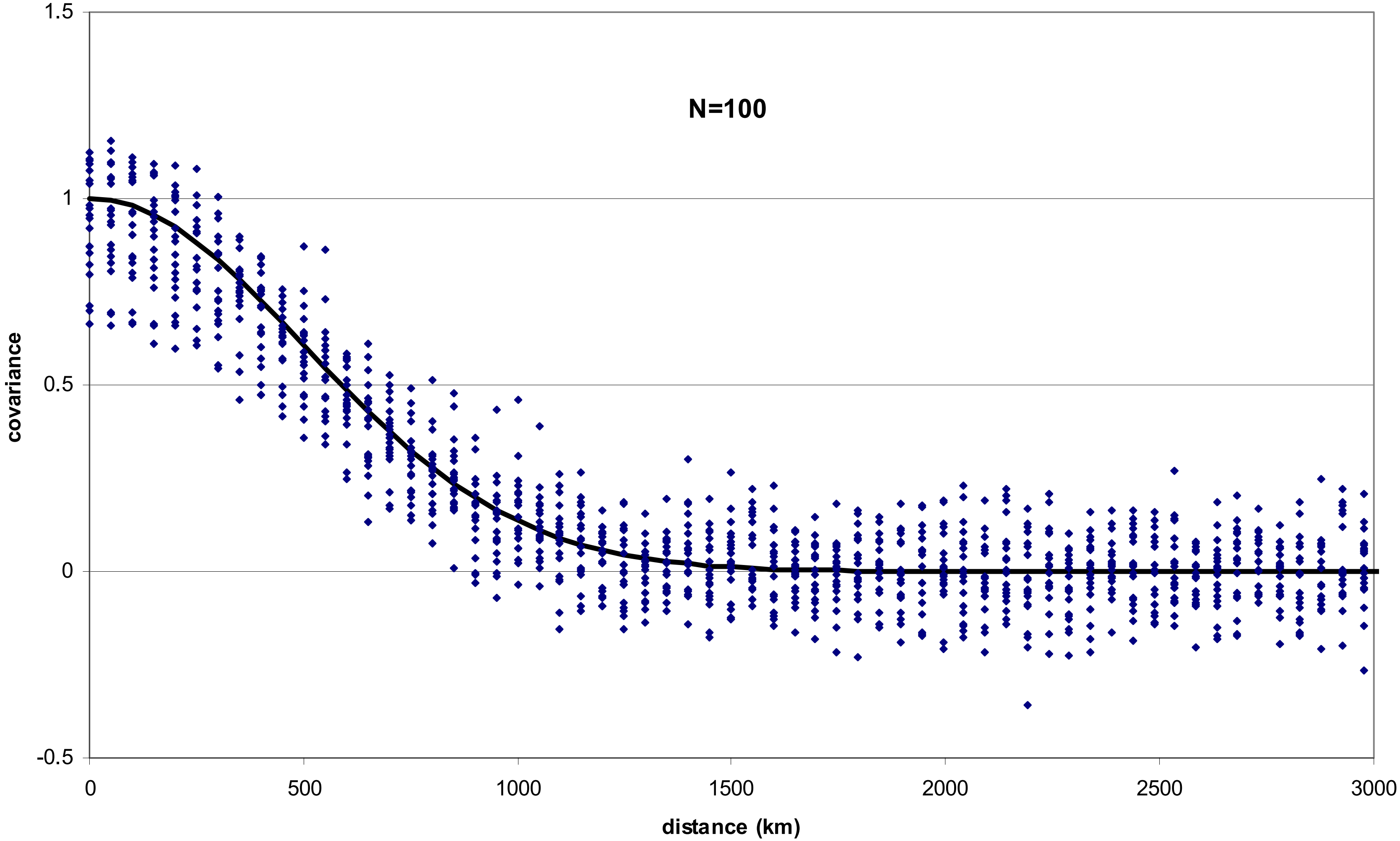
$$\mathbf{P}_e^f = \mathbf{X}^f \mathbf{X}^{f^T},$$

Analysis step has to construct $\mathbf{X}^a$ such that

$$\mathbf{P}_e^a = \mathbf{X}^a \mathbf{X}^{a^T},$$

Forecast step has to get $\mathbf{X}_{j+1}^f$ from $\mathbf{X}_j^a$.

Errors in sampled EnKF covariances



Errors in sampled EnKF covariances (2)

A sub-optimal Kalman gain calculated using the estimated covariances:

$$\mathbf{K}_e = \mathbf{P}_e^f \mathbf{H}^T \left(\mathbf{H} \mathbf{P}_e^f \mathbf{H}^T + \mathbf{R} \right)^{-1},$$

gives larger analysis errors, which can be calculated if one knows the true covariance:

$$\mathbf{P}^a = \mathbf{P}^f - 2\mathbf{K}_e \mathbf{H} \mathbf{P}^f + \mathbf{K}_e \left(\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R} \right) \mathbf{K}_e^T .$$

Global average variance, using a perfect ob:

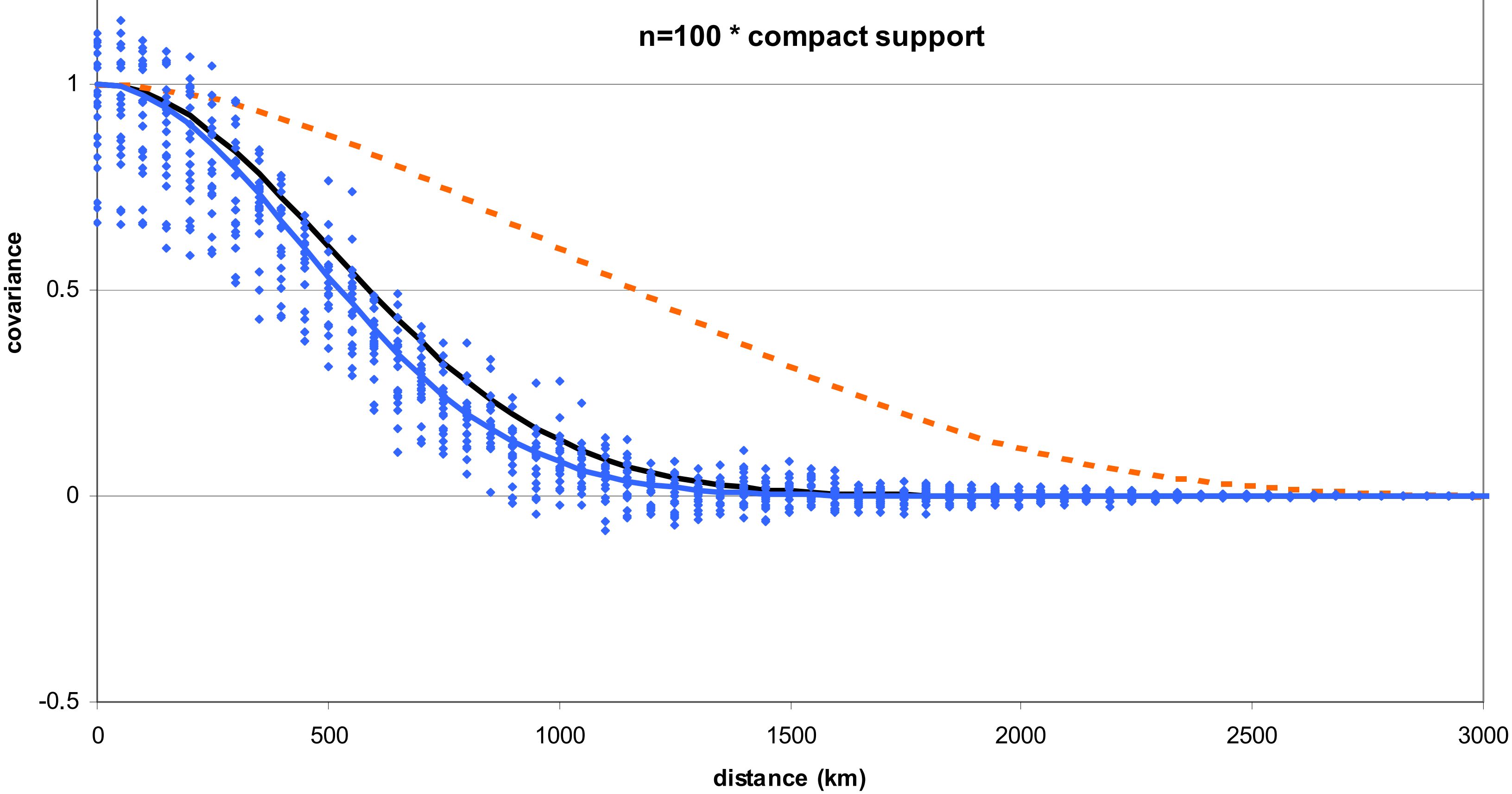
- Correct $\mathbf{K}$ gives 1–0.0015
- Sampled $\mathbf{K}_e$ gives 1+0.0087

Met Office

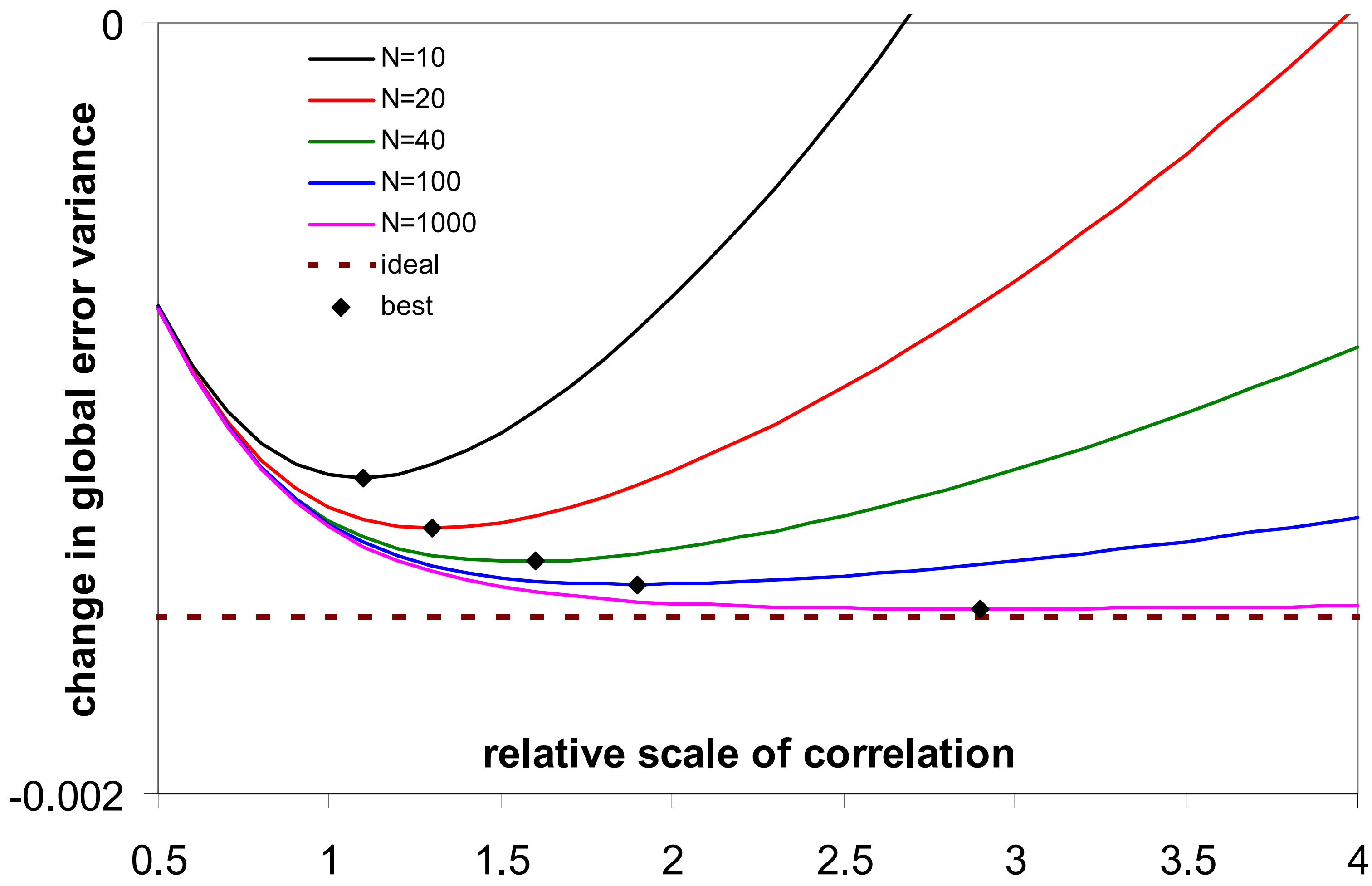
The Schur or Hadamard Product

$\mathbf{A} = \mathbf{B} \circ \mathbf{C}$ such that $A_{i,j} = B_{i,j} C_{i,j}$.

If B and C are valid covariances, then so is A.



Best scale for C depends on ensemble size N:



Choices in EnKF

- Treat pdfs as Gaussian
  - represented by mean & covariance
  - excludes Ens DA methods for small nonlinear problems
- Localise covariances
  - excludes ETKF
- *How to generate analysis ensemble?*
  - ? Perturbed observations
  - ? Transform methods (ETKF EAKF EnSRF)

Met Office

Perturbed model & observation ensemble

Add stochastic perturbations into the forecast of each ensemble member:

$$\mathbf{x}_i^f(t_{k+1}) = M\left(\mathbf{x}_i^a(t_k)\right) + \boldsymbol{\eta}_i(t_k),$$

$\boldsymbol{\eta}$ is noise with zero mean and covariance $\mathbf{Q}$.

$$\mathbf{P}_e^a(t_k) = \overline{\left(\mathbf{x}^a(t_k) - \overline{\mathbf{x}^a(t_k)}\right)\left(\mathbf{x}^a(t_k) - \overline{\mathbf{x}^a(t_k)}\right)^T}$$

$$\mathbf{P}_e^f(t_{k+1}) = \overline{\left(\mathbf{x}^f(t_{k+1}) - \overline{\mathbf{x}^f(t_{k+1})}\right)\left(\mathbf{x}^f(t_{k+1}) - \overline{\mathbf{x}^f(t_{k+1})}\right)^T}$$

$$\simeq \mathbf{M}\mathbf{P}_e^a(t_k)\mathbf{M}^T + \mathbf{Q}$$

Add stochastic perturbations into the analysis of each ensemble member:

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - H\left(\mathbf{x}_i^f\right) + \boldsymbol{\varepsilon}_i\right)$$

$\boldsymbol{\varepsilon}$ is noise with zero mean and covariance $\mathbf{R}$.

Met Office

Transform methods avoiding perturbed obs

Using SVD, transform $\left\{\mathbf{x}_i^f\right\}$ with covariance $\mathbf{P}^f = \mathbf{X}^f \mathbf{X}^{f\,T}$

into $\left\{\mathbf{x}_i^a\right\}$ with $\mathbf{P}^a = \mathbf{X}^a \mathbf{X}^{a\,T} = \mathbf{P}^f - \mathbf{P}^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \mathbf{H}\mathbf{P}^f$:

1) Update all ensemble by mean increment.

2) Scale perturbations to have correct covariance:

- Ensemble Adjustment Kalman Filter (Anderson):

$$\mathbf{X}^a = \mathbf{A}^T \mathbf{X}^f.$$

- Ensemble Transform Kalman Filter (Bishop):

$$\mathbf{X}^a = \mathbf{X}^f \mathbf{T}.$$

- Ensemble square root filter (Tippett et al):

sequential processing of obs

Properties of transform methods

- Mathematically equivalent for Gaussian pdfs
- Reduces errors due to noisy estimation of covariances (Whitaker & Hamill)
- If covariance localisation is wanted, then only practicable with sequential processing of obs
- Localised sequential processing EnSRF is simple to code and implement

Degenerate covariances

- The basic EnKF (without the Schur product) only has *N* degrees of freedom available to fit the observations.
- A perfect observation removes a degree of freedom from the ensemble:
- So the EnKF can only fit *N* pieces of information (in an area whose size depends on the Schur product correlation scale).

Met Office

Degenerate covariances

$$\mathbf{x}_i^a - \overline{\mathbf{x}^a} = \sum_{j=1}^{N} \left(\mathbf{x}_j^f - \overline{\mathbf{x}^f}\right) w_{j,i} \quad for\ i = 1, N \; .$$

All the analysed ensemble members, and the mean, will fit a perfect observation. Applying (linear) *H* gives

$$0 = \sum_{j=1}^{N} \left(H\mathbf{x}_j^f - H\overline{\mathbf{x}^f}\right) w_{j,i} \quad for\ i = 1, N \; .$$

So, as long as the $H\mathbf{x}_j^f$ are different, the matrix of $w_{j,i}$ is degenerate, and adding an error-free observation removes a degree of freedom from the ensemble.

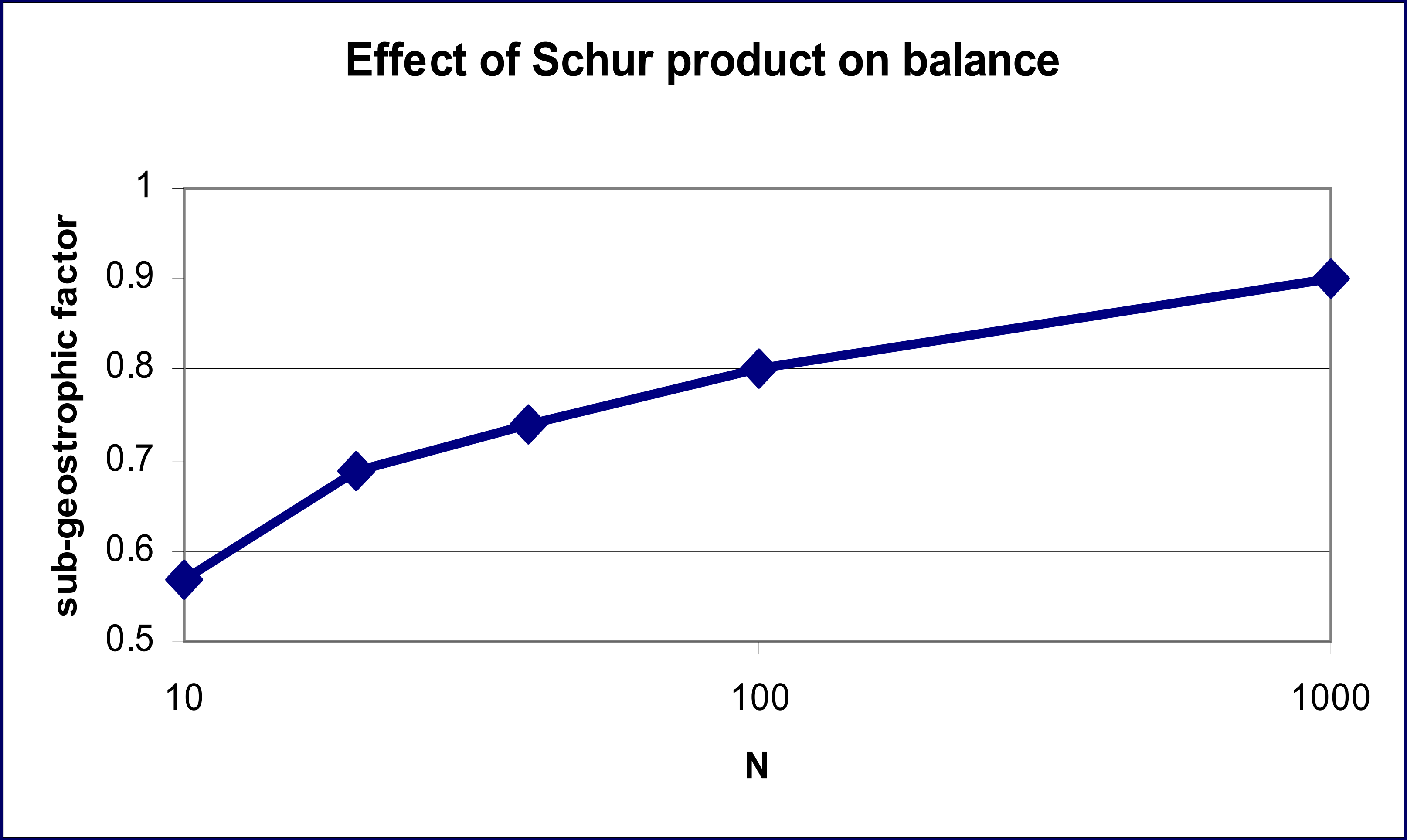
Balance

In the basic EnKF, linear balance in $\left\{\mathbf{x}_i^f - \overline{\mathbf{x}^f}\right\}$ will cause linear balance in $\left\{\mathbf{x}_i^a - \overline{\mathbf{x}^a}\right\}$.

This comes through the null-space of $\mathbf{P}^f$.

But $\mathbf{P}_e^f$ has more null-space due to finite *N*; in relaxing this using the Schur product we must also relax balance.

Effect of Schur product on geostrophic balance



Non-Gaussian Analysis Equations (1)

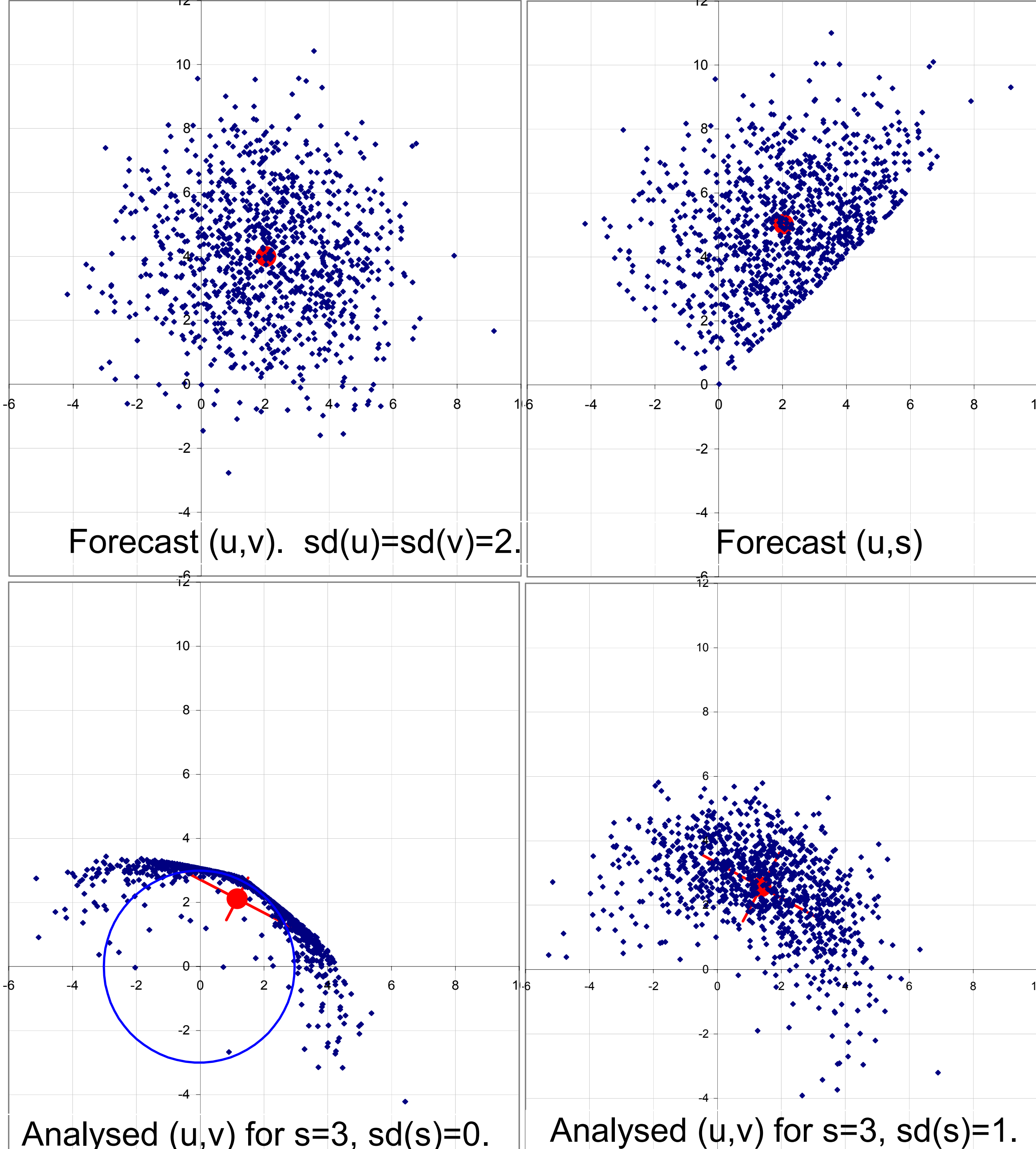
Nonlinear observation operators

Even if forecast error pdf is Gaussian, described by $\mathbf{P}_e^f$, the pdf of $H\left(\mathbf{x}^f\right)$ is only approximated by a Gaussian with covariance

$$\mathbf{HP}^f\mathbf{H}^T \simeq \overline{\left(H\left(\mathbf{x}^f\right)-\overline{H\left(\mathbf{x}^f\right)}\right)\left(H\left(\mathbf{x}^f\right)-\overline{H\left(\mathbf{x}^f\right)}\right)^T}$$

So for situations where information from nearby observations helps (e.g. de-aliassing scatterometer observations), Var with a non-quadratic penalty function should do better than the (sequential Gaussian) EnKF.

Example EnKF of a wind speed ob (N=1000).



Non-Gaussian Analysis Equations (2)
Quality Control

- For situations where information from nearby observations helps (e.g. extreme obs corroborating each other), Var with a non-quadratic penalty function should do better than the (sequential Gaussian) EnKF.
- But many QC decisions are in data-sparse areas, where the principle source of corroborative information is the forecast. If the error variance "of the day" from the ensemble, despite sampling noise, is more accurate than that assumed in Var, then the EnKF will do better.

Assimilation characteristics

	Incremental 4D-Var	EnKF
Forecast covariances	Modelled at t_0 (usually isotropic), time evolution for a finite time-window represented by linear and adjoint models.	Sampled by ensemble (flow-dependent). Noisy: must be modified to have compact support using Schur product.
Ability to fit detailed observations.	Limited by resolution of simplified model. Tendencies fitted within time-window.	Fewer data (in a region) than ensemble members. Tendency information only extracted if obs properly fitted.
Balance constraints	Can be imposed through a dynamical design to the variable transform, or a separate balance penalty.	Only imposed if each forecast in the ensemble is balanced. Lost slightly in Schur product.
Nonlinear observation operators	Allowed if differentiable. (Results uncertain if pdf is bimodal in range of interest.)	Allowed, but resulting pdf modelled by Gaussian.
Non-Gaussian observational errors	Allowed if differentiable. (Results uncertain if pdf is bimodal in range of interest.)	Not allowed. Prior QC step is needed.

Practical characteristics

	Incremental 4D-Var	EnKF
Forecast model	Predict evolution of mean. No switches.	Predict typical state. May have stochastic physics and switches.
Linear model	Predict average evolution of finite perturbations from the mean. May be simplified.	Not needed.
Adjoint model	Needed for (simplified) linear model.	Not needed.
Covariance model	Significant effort for covariance model. Adjoint code needed.	Simple correlation in Schur product. Covariance inflation to keep the right spread.
Observation operators	Linear and adjoint operators needed (not usually difficult).	Only uses forward operators.
Analysis algorithm	Descent algorithm available as "off the shelf" software.	EnSRF very easy. Simultaneous box algorithms are more complicated (like OI).
Suitability for parallel computers	Require parallel simplified and adjoint models.	Forecasts can run in parallel. Covariances require a transposition. Sequential proc of obs difficult.
Limited-area modelling	Error covariance models OK to specify boundary value errors	Ensemble of global forecasts to provide boundary conditions.

Two possible ways forward

- Hybrid 4D-Var - EnKF for mainstream NWP
- Nested EnKF for special applications

Mainstream NWP Systems - requirements

- High resolution to reduced errors of representativeness
- Analyse all scales with significant errors
- Quality control observations
- Use nonlinear observations
- Update satellite bias corrections
- Better (mean) short-period forecast
  more important than better error estimates

VAR can address all of these simultaneously

$\Rightarrow$ *Seek to enhance it by adding benefits of EnKF*

Variational use of EnKF covariance (1)

- Can use the ensemble generated covariances in a variational algorithm
- This should give identical results to the mean from an ideal EnKF algorithm
- As usual with VAR, the analysis error covariance is ***not*** automatically obtained
- So the VAR method cannot easily generate an ensemble, but it can use one made by another system.

Variational use of EnKF covariance (1)

Basic EnKF

$$\mathbf{P}_e^f = \mathbf{X}^f \mathbf{X}^{f^T}.$$

Define transform from $\boldsymbol{\alpha}$: $\mathbf{x} = \overline{\mathbf{x}^f} + \mathbf{X}^f \boldsymbol{\alpha}$.

Let covariance $\mathbf{B}_{(\alpha)} = \langle \boldsymbol{\alpha}\boldsymbol{\alpha}^T \rangle = \mathbf{I}$.

Then $\left\langle \left(\mathbf{x} - \overline{\mathbf{x}^f}\right)\left(\mathbf{x} - \overline{\mathbf{x}^f}\right)^T \right\rangle = \mathbf{P}_e^f$.

So the variational analysis for the transformed variable $\boldsymbol{\alpha}$ is obtained by minimising

$$J(\alpha) = \tfrac{1}{2}\alpha^T \alpha + \tfrac{1}{2}\left(\mathbf{y} - \mathbf{y}^o\right)^T \mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{y}^o\right)$$

$$\mathbf{y} = H\left(\overline{\mathbf{x}^f} + \mathbf{X}^f \boldsymbol{\alpha}\right)$$

(2)

Similarly, VAR can use ensemble covariances modified by a Schur product:

EnKF with Schur product

Use as control variable a vector $\boldsymbol{\alpha}$ of N 2D fields $\boldsymbol{\alpha}_i$, each with covariance C.
The variational problem is then to minimise

$$J(\alpha)=\tfrac{1}{2}\alpha^T\begin{pmatrix}\mathbf{C} & & 0\\ & \ddots & \\ 0 & & \mathbf{C}\end{pmatrix}^{-1}\alpha+\tfrac{1}{2}\left(\mathbf{y}-\mathbf{y}^o\right)^T\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{y}^o\right)$$

$$\mathbf{y}=H\left(\overline{\mathbf{x}^f}+\left(\mathbf{X}^f\circ\boldsymbol{\alpha}\right)\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\right)$$

As in standard 3D-Var methods, the inversion of the block diagonal matrix is avoided by a further horizontal transform of each field $\boldsymbol{\alpha}_i$ into spectral space.

Hybrid Var-EnKF with Schur product

(3)

VAR can use the ensemble to augment the "traditional" covariance model with some *Errors Of The Day*.

Use the traditional variation control variable v supplemented by α, so that we minimise

$$J(\mathbf{v},\boldsymbol{\alpha}) = \frac{\beta}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}\boldsymbol{\alpha}^T \begin{pmatrix} \mathbf{C} & & 0 \\ & \ddots & \\ 0 & & \mathbf{C} \end{pmatrix}^{-1} \boldsymbol{\alpha}$$

$$+\frac{1}{2}\left(\mathbf{y}-\mathbf{y}^o\right)^T \mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{y}^o\right)$$

$$\mathbf{y} = H\left(\mathbf{x}^g + \mathbf{U}\mathbf{v} + \left(\mathbf{X}^f \circ \boldsymbol{\alpha}\right)\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\right)$$

Should reduce "traditional" error covariances to compensate for those represented by the ensemble (β>1).

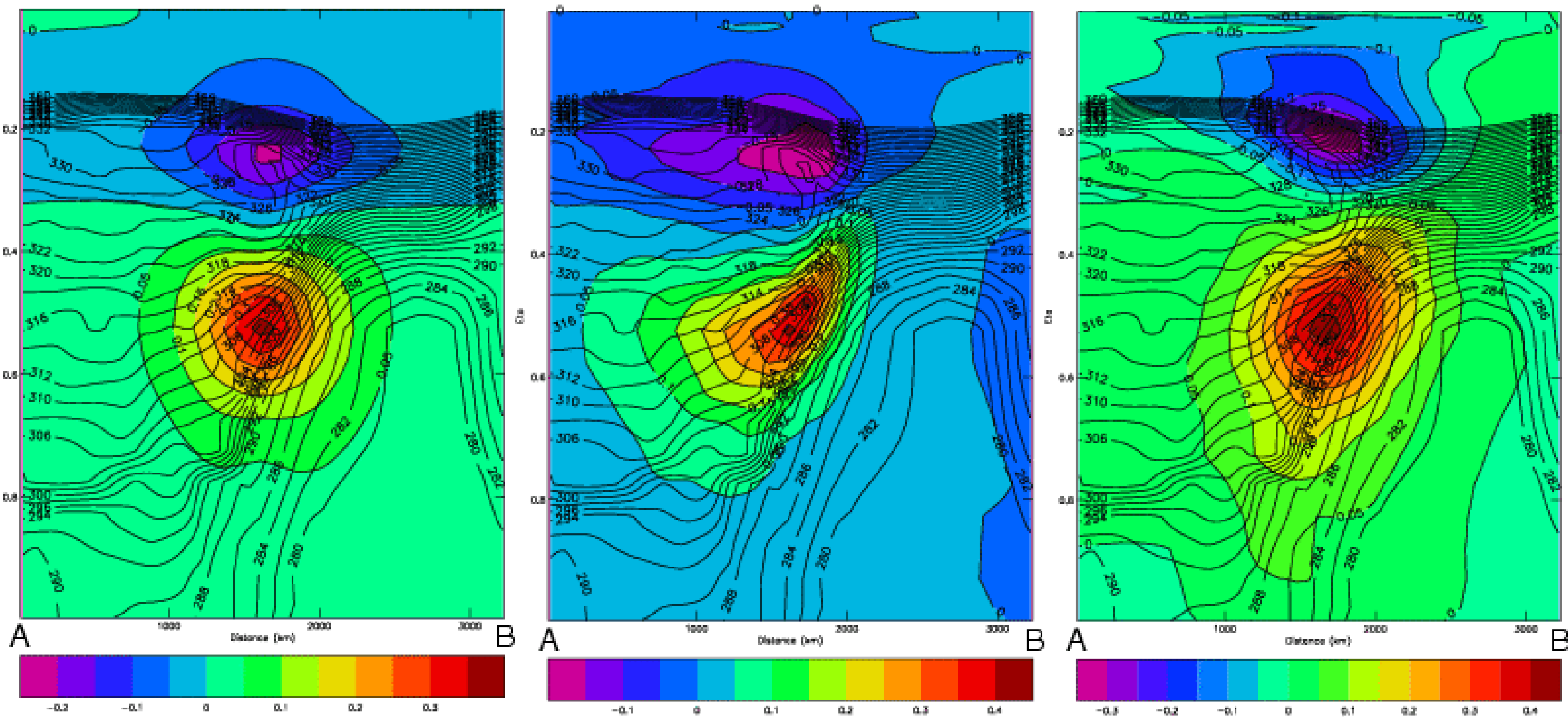
Dale Barker & Adrian Semple at Met Office.

Hamill & Snyder (2000).

Met Office

Response to a single T ob

Contours - UM analysis, thetal



Basic 3D-Var

VAR+GCT

3D-Var + 1 bred mode

Dale Barker EOTD expts.
Mark Dubal GCT expts.
Adrian Semple, 2001: A Meteorological Assessment of the Geostrophic Co-ordinate Transform and Error Breeding System When used in 3D Variational Data Assimilation. NWP Tech Rep 357.

Hybrid 4D-Var EnKF - a possible scenario

- High resolution "control" NWP model, for the best possible prediction of the "best" estimate.
- Incremental 4D-Var, using all relevant observations in a time-window (including nonlinear *H*, and QC).
- Final outer-iteration at sufficient resolution to extract information from the tendencies between consecutive observed fields.
- Background Covariances in 4D-Var enhanced by synoptic-scale *Errors of The Day* from an ensemble of perturbations.
- Ensemble propagated using an ETKF
  centred on the high-res 4D-Var control.

Met Office

EnKF for special applications

- Large uncertainty ⇒ difficulty with 4D-Var:
  - Deterministic NWP model for ensemble mean
  - Evolution of perturbations nonlinear
- EnKF can work with large uncertainty:
  - Sparse observations (Whitaker *et al.*)
  - Idealized convective-scale, with no large-scale errors (Snyder & Zhang)
- EnKF is easy to develop, even if model is also developing

➢ Use EnKF for R&D of new NWP applications with large uncertainty (e.g. convective-scale)

Nested in mainstream NWP bcs to give synoptic-scales

+ Var	Summary of (dis-)advantages	EnKF +

Simple to design & code.

Needs smooth forecast model.
Needs PF & Adjoint models.
Needs a covariance model.

Generates an ensemble forecast.

Sampled covariances noisy.
Can only fit *N* data.

Can extract info from tracers.

Nonlinear obs operators
& non-Gaussian errors modelled.

Complex obs operators (eg rain)
coped with automatically,
but sample is then fitted by Gaussian.

Incremental balance easy

External initialisation
of each forecast needed?

Accurate modelling of time-covariances
only within 4D-Var window.

Covariances evolved indefinitely
only if represented in ensemble.

4D-Var v EnKF

Summary

Met Office