Ensemble-Based Atmospheric Data Assimilation: A Tutorial

Thomas M. Hamill

NOAA-CIRES Climate Diagnostics Center

Boulder, Colorado USA 80301

tom.hamill@noaa.gov

www.cdc.noaa.gov/people/tom.hamill/efda_review.ppt

With minor modifications by Ming Xue for presentation at OU OBAN/DA class

Motivation

- **Ensemble forecasting** : Provides flow-dependent estimate of *uncertainty of the forecast.*

- **Data assimilation** : requires information about *uncertainty in prior forecast* and observations.

- More accurate estimate of uncertainty, less error in the analysis. Improved initial conditions, improved ensemble forecasts.

- Ensemble forecasting ⟷ data assimilation.

Example: where flow-dependent first-guess/background errors help
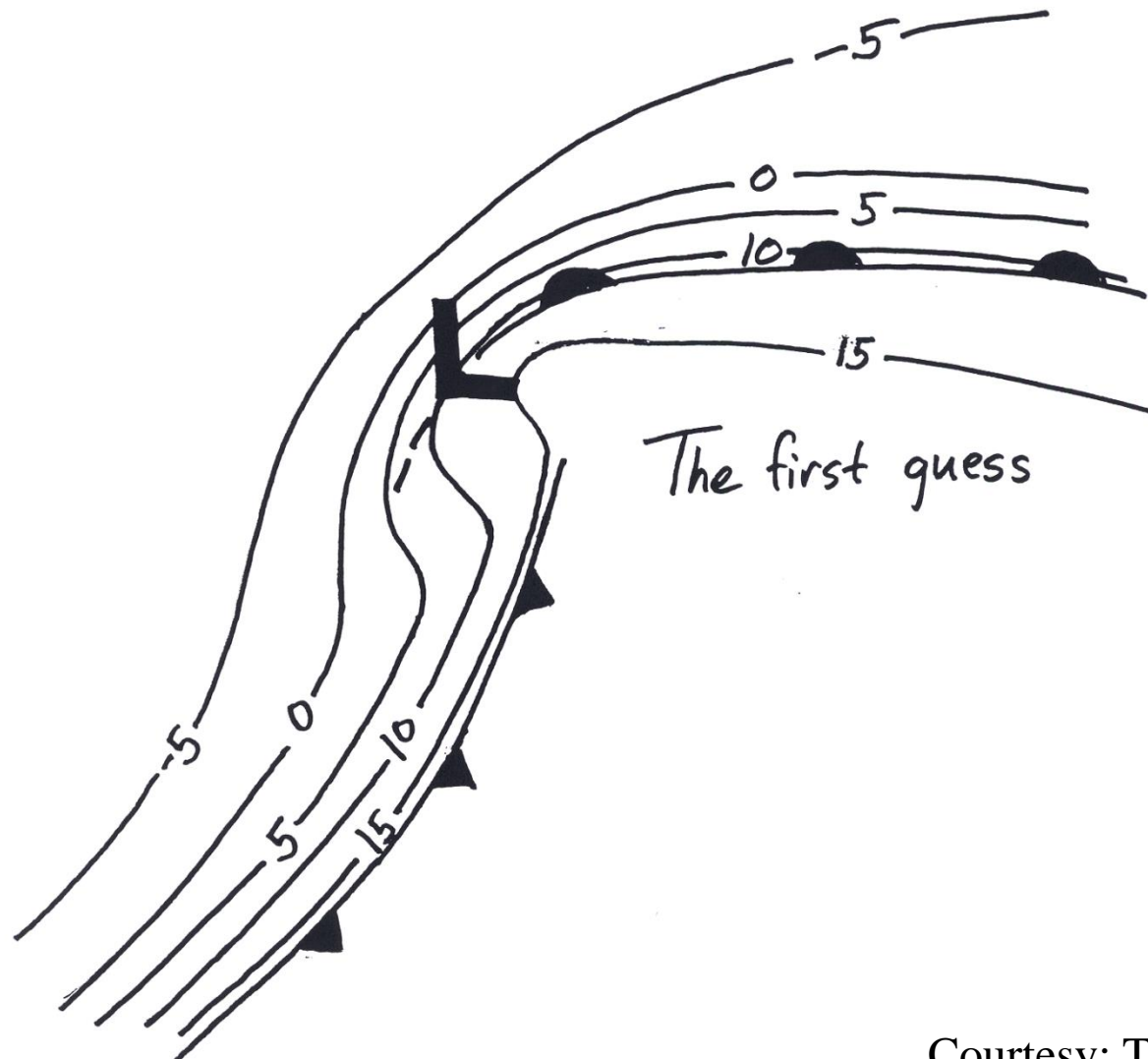
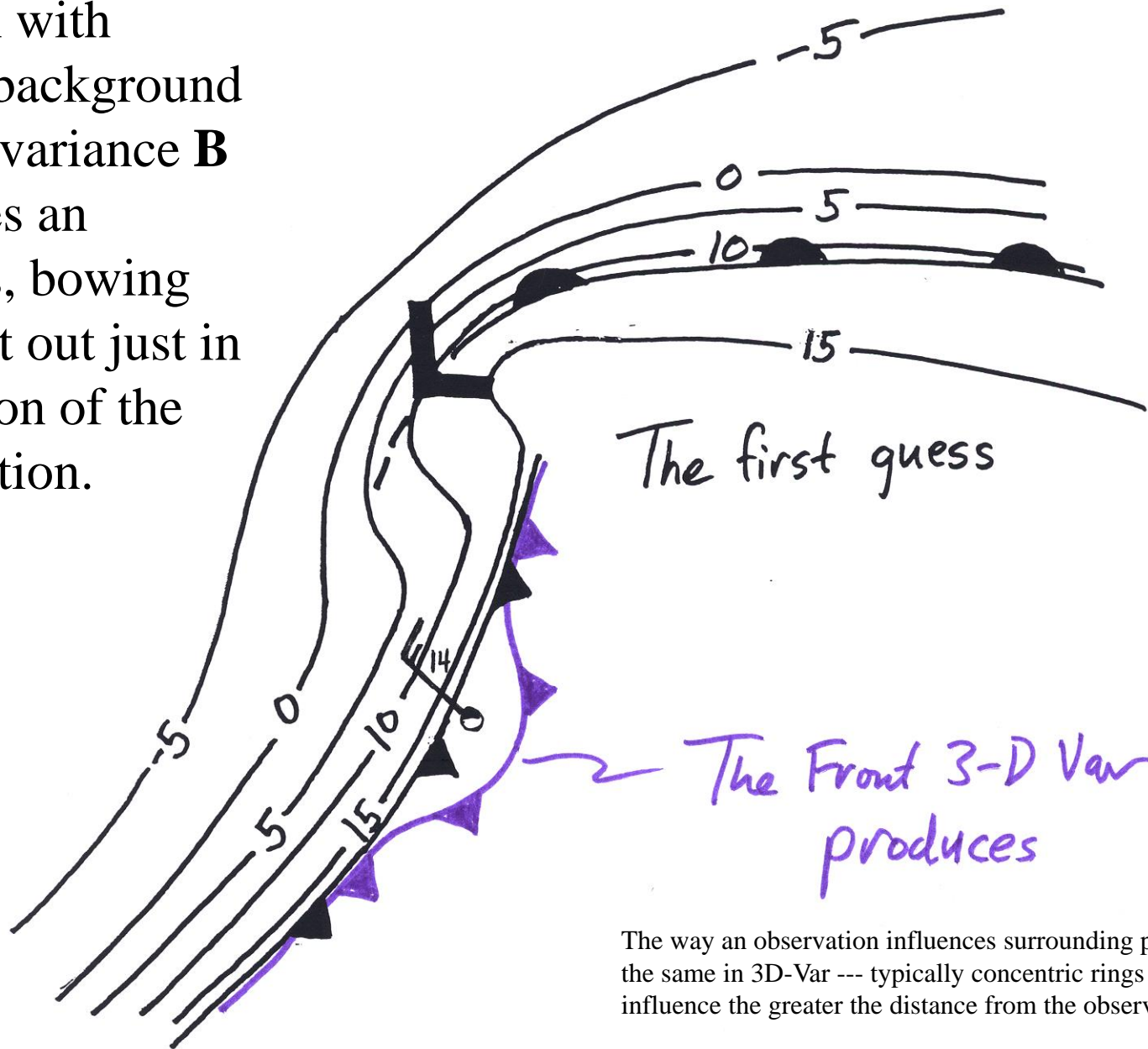Courtesy: Thomas Hamill

A new observation is taken near the front that is inconsistent with first guess

The first guess
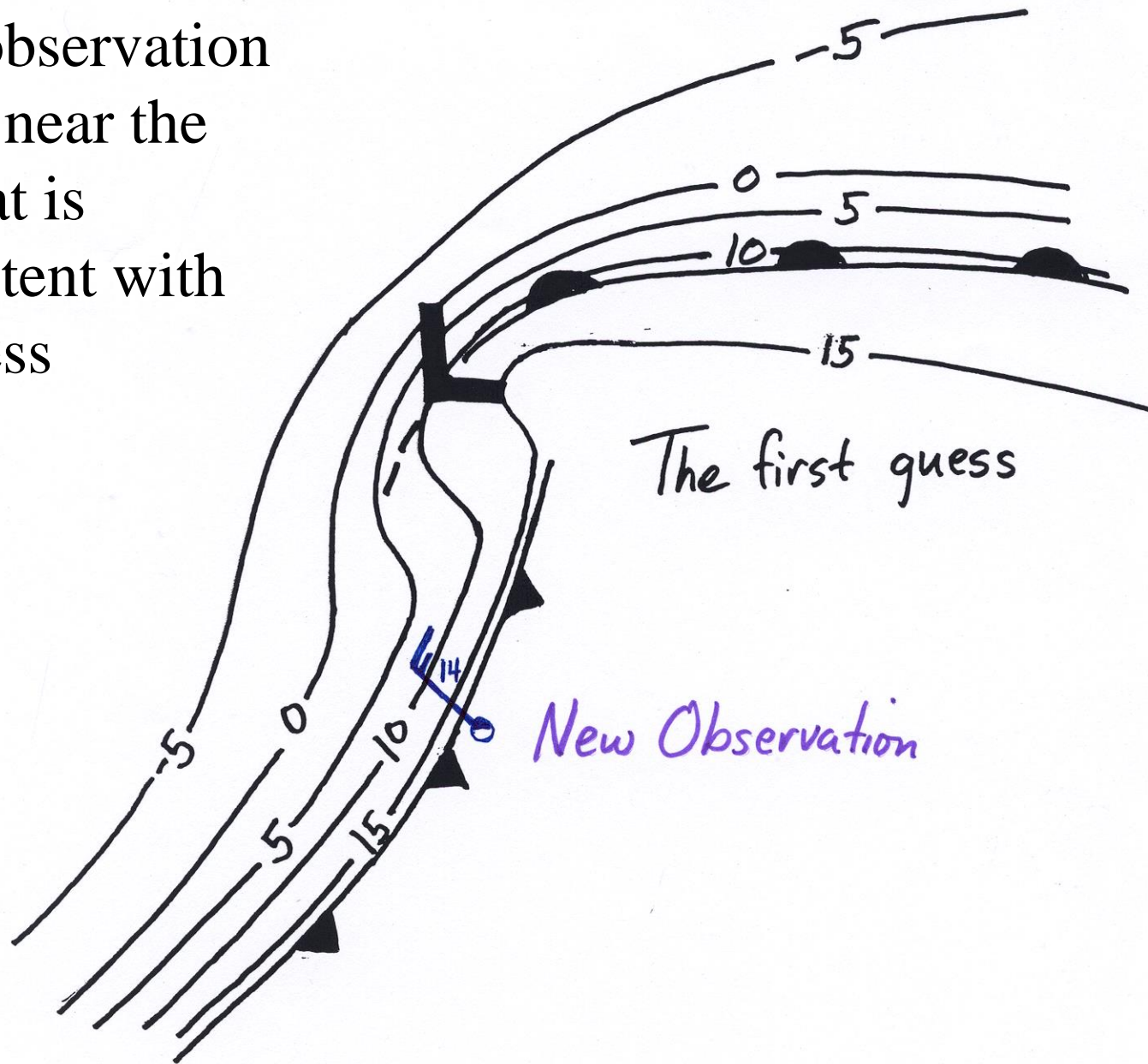
New Observation

3DVAR with typical background error covariance **B** produces an analysis, bowing the front out just in the region of the observation.

-5

0

5

10

15

The first guess

-5

0

10

14

5

15

The Front 3-D Var produces

The way an observation influences surrounding points is always the same in 3D-Var --- typically concentric rings of decreasing influence the greater the distance from the observation.

Courtesy: Thomas Hamill

Here's the front you'd re-draw based on synoptic intuition, understanding that the errors are *flow dependent.*
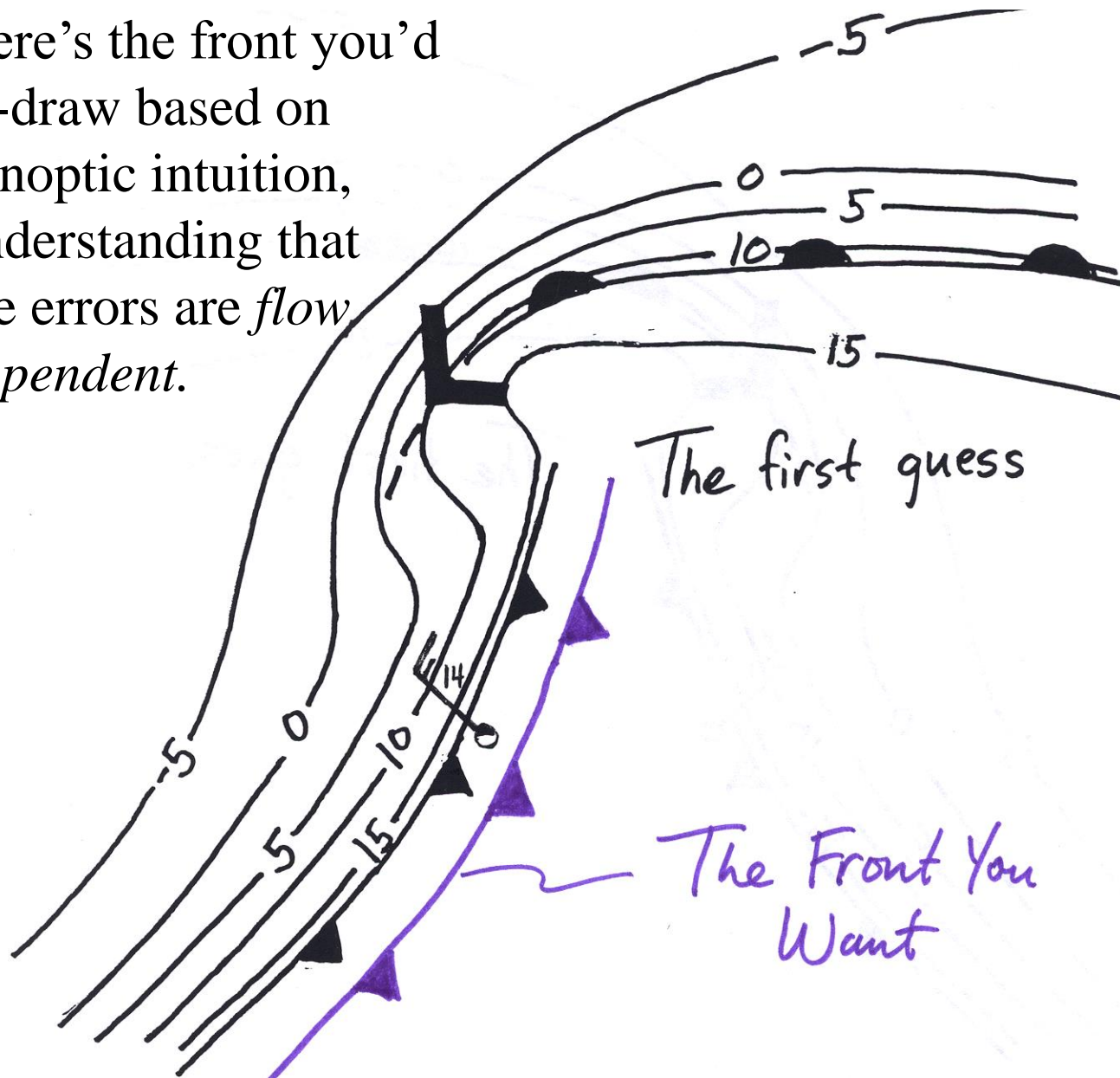
In this special case, your synoptic intuition would tell you that temperature errors ought to be correlated along the front, from northeast to southwest. However, you wouldn't expect that same NE-SW correlation if the observation were near the warm front. Analyses might be improved dramatically if the errors could have this *flow-dependency.*

How can we teach data assimilation algorithms to do this?

-5

0

5

10

15

The first guess

-5

0

10

14

5

15

The Front You Want

Courtesy: Thomas Hamill

Ensemble-based data assimilation algorithms

- Can use ensemble to model the statistics of the first guess ("background") errors.

- Tests in simple models show dramatically improved sets of objective analyses

- These sets of objective analyses could also be useful for initializing ensemble forecasts.

4D-Var: As far as data assimilation can go?

- Finds model trajectory that best fits observations over a time window.

- BUT:

  - Requires linear tangent and adjoint (difficult to code, and are linearity of error growth assumptions met?)

  - What if forecast model trajectory unlike real atmosphere's trajectory? (model error)

- Ensemble-based approaches avoid these problems (though they have their own, as we'll see).

Example: assimilation of a sparse network of surface-pressure observations
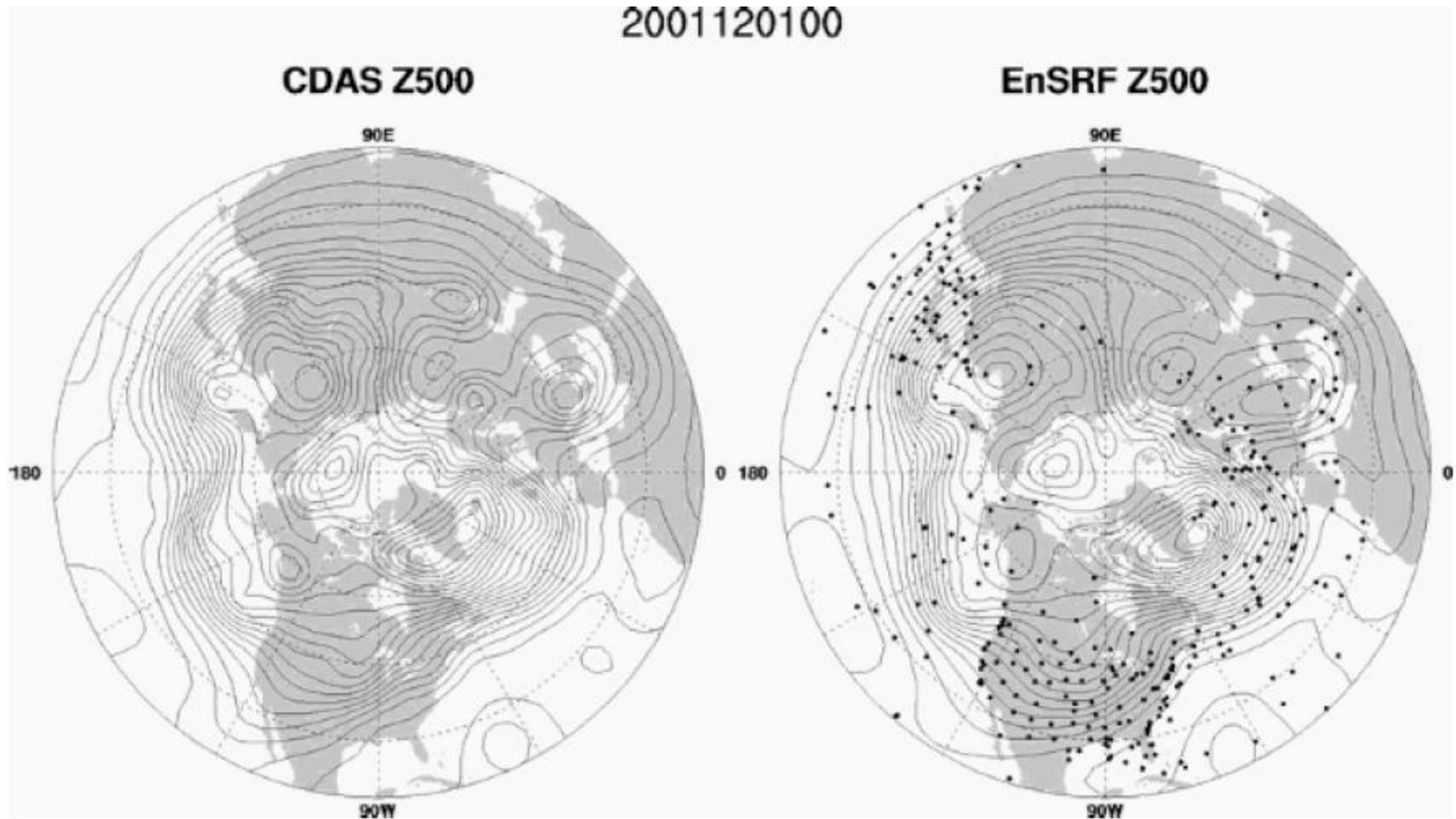
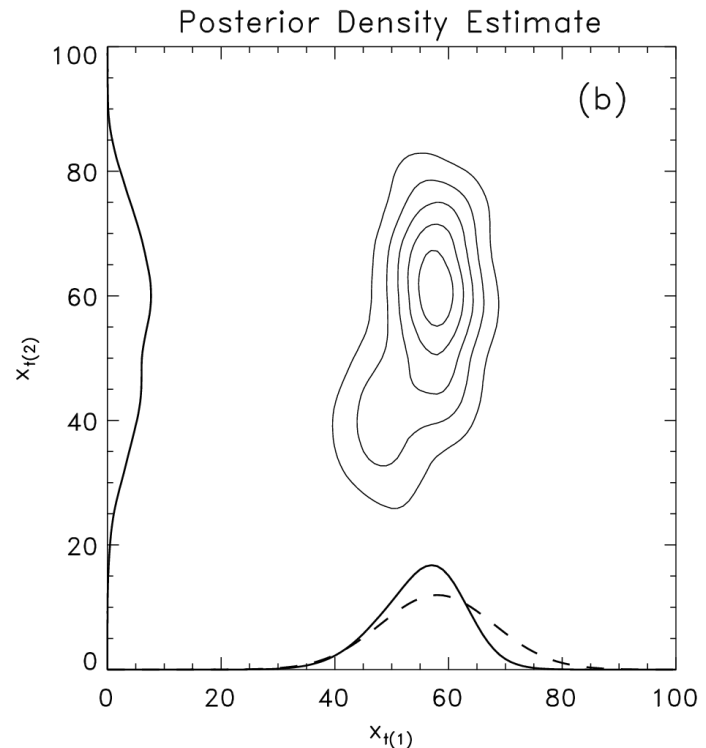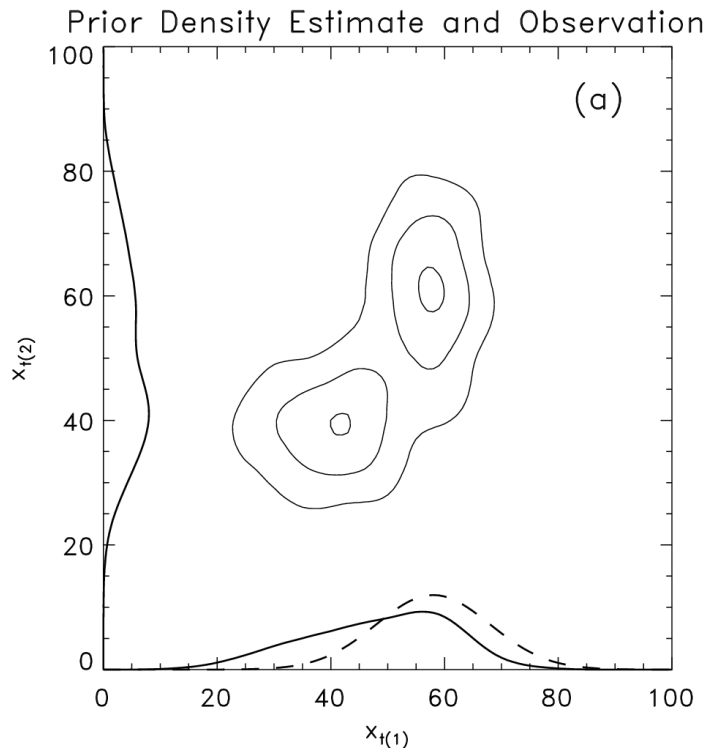# From first principles: Bayesian data assimilation

$\mathbf{x}_t$ = (unknown) true model state

$\psi_t = \left[\mathbf{y}_t, \psi_{t-1} \right]$ = observations = [Today's, all previous]

$$P\left(\mathbf{x}_t \mid \psi_t \right) \propto P\left(\mathbf{y}_t \middle| \mathbf{x}_t \right) P\left(\mathbf{x}_t \middle| \psi_{t-1} \right)$$

"posterior"   "prior"

A manipulation of Baye's Rule, assuming observation errors are independent in time.

Bayesian data assimilation: 2-D example

Prior Density Estimate and Observation (a)

Posterior Density Estimate (b)

Computationally expensive! Here, probabilities explicitly updated on 100x100 grid; costs multiply geometrically with the number of dimensions of model state.

Data assimilation terminology

- $\mathbf{y}$: Observation vector (raobs, satellite, etc.)
- $\mathbf{x}^b$: Background state vector (1st guess)
- $\mathbf{x}^a$: Analysis state vector
- $\mathbf{H}$: (Linear) operator to convert model state $\rightarrow$ obs
- $\mathbf{R}$: Observation - error covariance matrix
- $\mathbf{P}^b$: Background - error covariance matrix
- $\mathbf{P}^a$: Analysis - error covariance matrix

Simplifying Bayesian D.A.: toward the Kalman Filter

$$P\left(\mathbf{x}_t \middle| \psi_t\right) \propto P\left(\mathbf{y}_t \middle| \mathbf{x}_t\right) P\left(\mathbf{x}_t \middle| \psi_{t-1}\right) \quad \text{(manipulation of Baye's rule)}$$

Assume

$$P(\mathbf{x}_t \middle| \psi_{t-1}) \sim N\left(\mathbf{x}^b, \mathbf{P}^b\right) \propto \exp\left(-\frac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^{\mathrm{T}} \mathbf{P}^{b^{-1}} \left(\mathbf{x}-\mathbf{x}^b\right)\right)$$

$$P(\mathbf{y}_t \middle| \mathbf{x}_t) \sim N(\mathbf{y}_t, \mathbf{R}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{Hx}-\mathbf{y}_t\right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\mathbf{Hx}-\mathbf{y}_t\right)\right)$$

so

$$P(\mathbf{x}_t \middle| \psi_t) \propto \exp\left(-\frac{1}{2}\left[\left(\mathbf{x}-\mathbf{x}^b\right)^{\mathrm{T}} \mathbf{P}^{b-1} \left(\mathbf{x}-\mathbf{x}^b\right)+\left(\mathbf{Hx}-\mathbf{y}\right)^{\mathrm{T}} \mathbf{R}^{-1}\left(\mathbf{Hx}-\mathbf{y}\right)\right]\right) \quad (@)$$

Maximizing (@) equivalent to minimizing –ln(@), i.e., minimizing the functional

Kalman filter derivation continued

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

After much math, plus other assumptions about linearity, we end up with the "Kalman filter" equations (see Lorenc, *QJRMS*, 1986).

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^b\right)$$

This "update" equation tells how to estimate the analysis state. A weighted correction of the difference between the observation and the background is added to the background.

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

K is the "Kalman Gain Matrix." It indicates how much to weight the observations relative to the background and how to spread their influence to other grid points

$$\mathbf{P}^a = \left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)\mathbf{P}^b$$

P^a is the "analysis-error covariance. The Kalman filter indicates not only the most likely state but also quantifies the uncertainty in the analysis state.

$$\mathbf{P}^b{}_{t+1} = \mathbf{M}\mathbf{P}^a\mathbf{M}^T + \mathbf{Q}$$

How the background errors at the next data assimilation time are estimated. **M** is the "tangent linear" of the forecast model *M*
Assumed . $\mathbf{x}_{t+1} = \mathbf{M}\mathbf{x}_t + \eta_t, \qquad \left\langle \eta_t\eta_t^T \right\rangle = \mathbf{Q}$

$$\mathbf{x}^b_{t+1} = \mathbf{M}\mathbf{x}^a$$

How the forecast is propagated. In "extended Kalman filter" **M** is replaced by fully nonlinear *M*

Kalman filter update equation: example in 1-D

$$x^a = x^b - \frac{P^b}{P^b + R}\left(y - x^b\right) = \frac{R}{P^b + R}\, x^b + \frac{P^b}{P^b + R}\, y$$

$$P^a = P^b - P^b\, \frac{P^b}{P^b + R} = P^b\, \frac{R}{P^b + R}$$

Kalman filter problems

- Covariance propagation step $\mathbf{P}_t^a \rightarrow \mathbf{P}^b{}_{t+1}$ very expensive
- Still need tangent linear/adjoint models for evolving covariances, linear error growth assumption questionable.

From the Kalman filter to the ensemble Kalman Filter

- What if we estimate $\mathbf{P}^b$ from a random ensemble of forecasts? (Evensen, *JGR* 1994)
- Let's design a procedure so if error growth is linear and ensemble size infinite, gives same result as Kalman filter.

Canonical ensemble Kalman filter update equations

$$\mathbf{x}_\mathbf{i}^\mathbf{a} = \mathbf{x}_\mathbf{i}^\mathbf{b} + \mathbf{K}\left(\mathbf{y}_\mathbf{i} - H\mathbf{x}_\mathbf{i}^\mathbf{b}\right)$$

H = (possibly nonlinear) operator from model to observation space

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

$\mathbf{x}$ = state vector
($_\mathbf{i}$ for i*th* member)

$$\mathbf{P}^b = \mathbf{X}\mathbf{X}^\mathbf{T}$$

$$\mathbf{y}_\mathbf{i} = \mathbf{y} + \mathbf{y}_\mathbf{i}^{'}$$

$$\mathbf{y}_\mathbf{i}^{'} \sim N(0, \mathbf{R})$$

$$\mathbf{X} = \left(\mathbf{x}_\mathbf{1}^\mathbf{b} - \overline{\mathbf{x}^\mathbf{b}}, \ldots, \mathbf{x}_\mathbf{n}^\mathbf{b} - \overline{\mathbf{x}^\mathbf{b}}\right)$$

Notes: (1) An ensemble of parallel data assimilation cycles is conducted, assimilating *perturbed observations* .

(2) Background-error covariances are estimated using the ensemble.

# The ensemble Kalman filter: a schematic
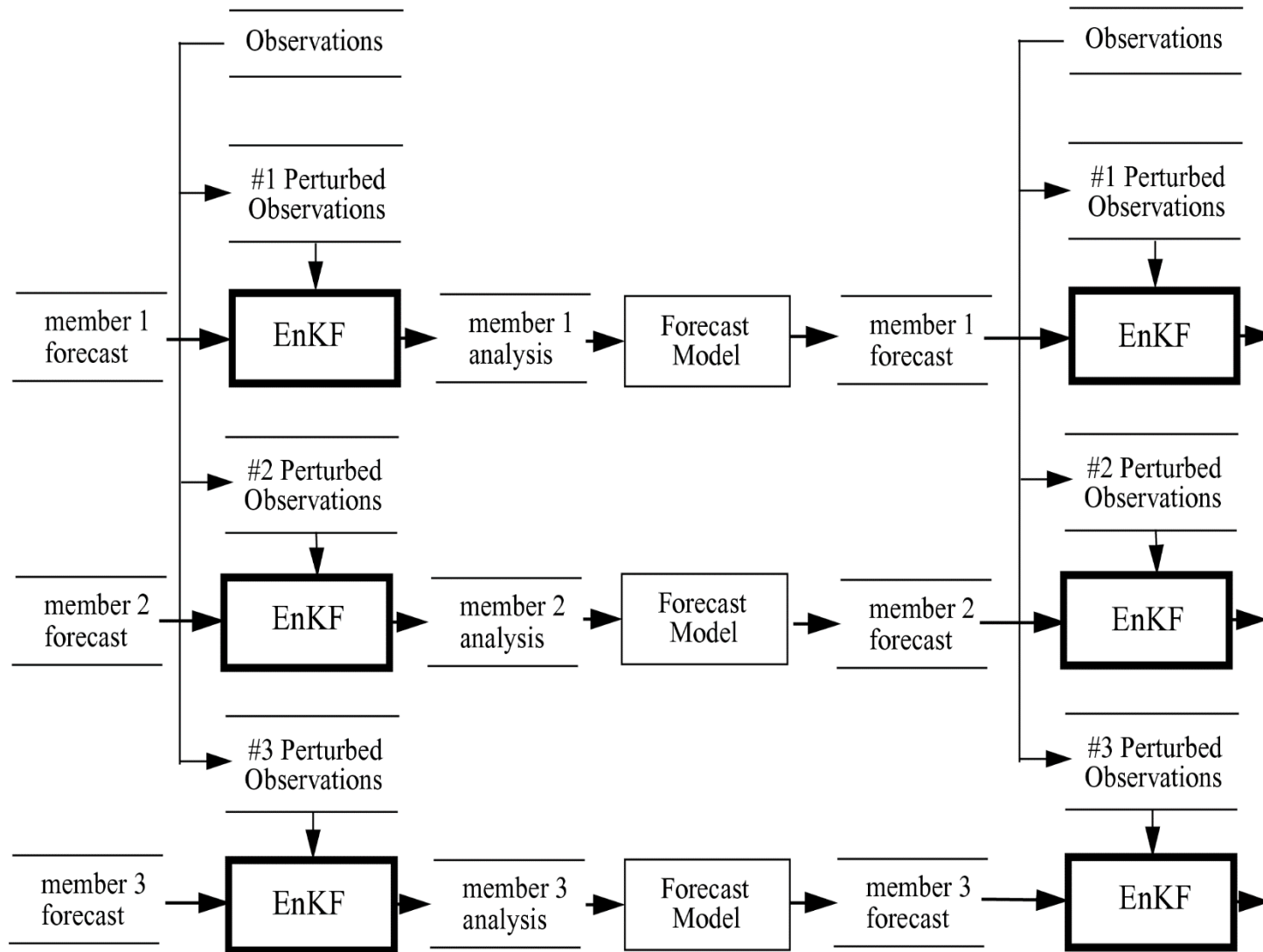
(This schematic is a bit of an inappropriate simplification, for EnKF uses every member to estimate background-error covariances)

How the EnKF works:
2-D example
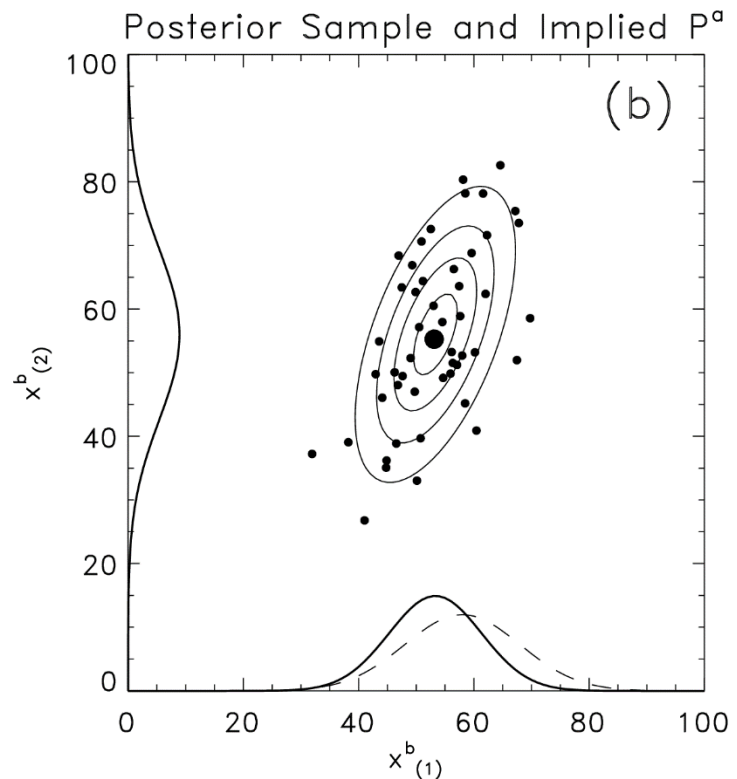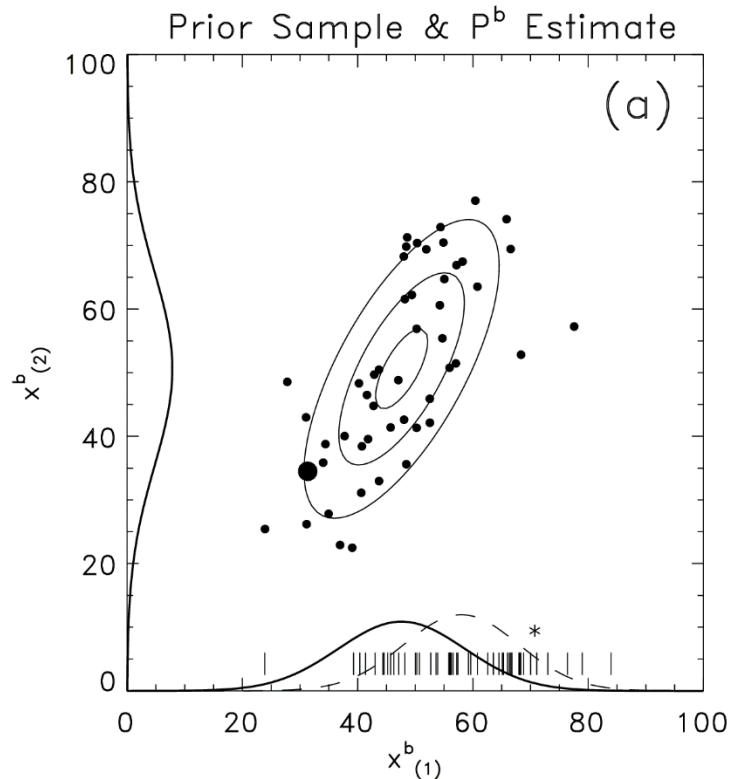
Start with a random sample from bimodal distribution used
in previous Bayesian data assimilation example. Contours reflect
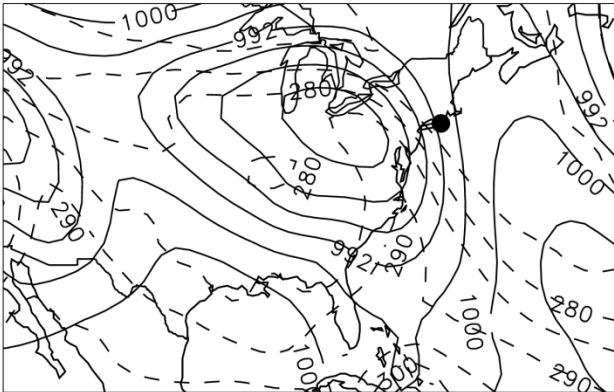the Gaussian distribution fitted to ensemble data.
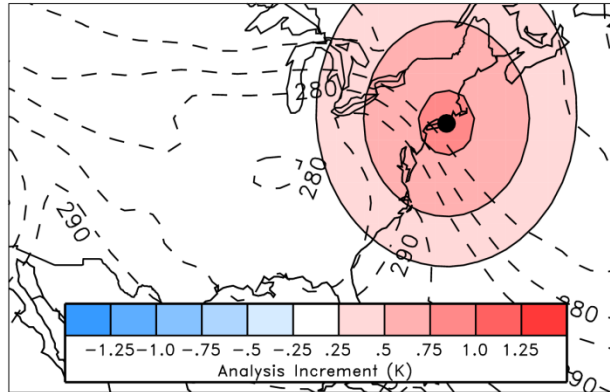
Why perturb the observations?
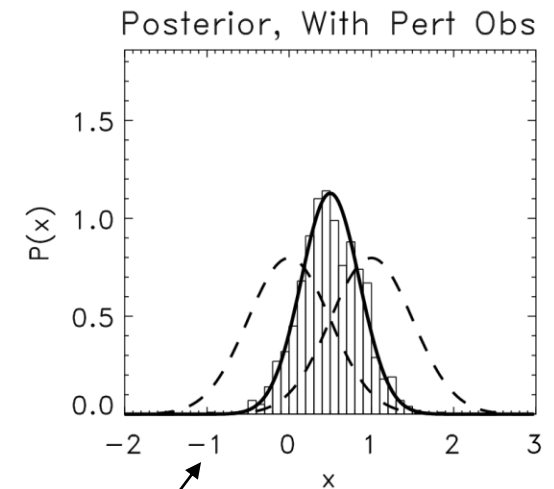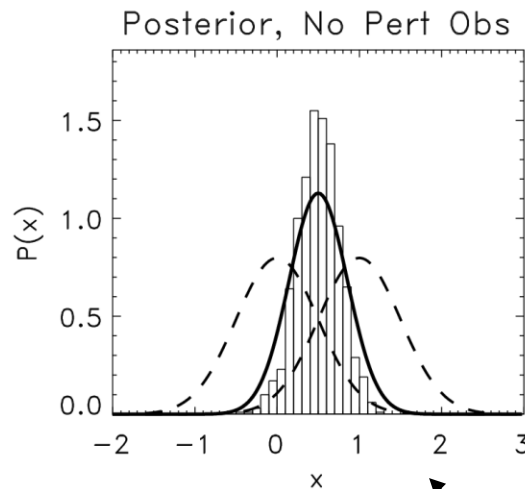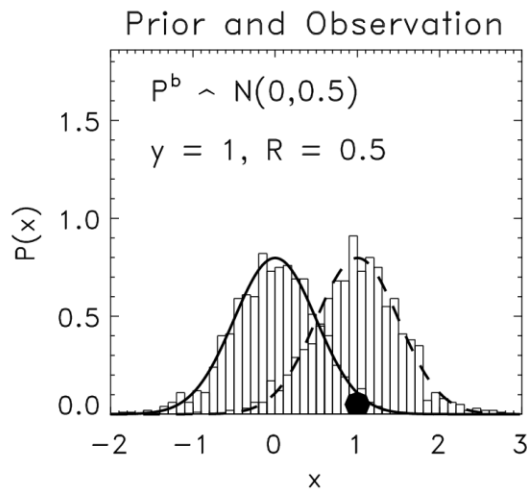
histograms denote the ensemble values; heavy black line denotes theoretical expected analysis-error covariance

How the EnKF may achieve its improvement relative to 3D-Var: better background-error covariances
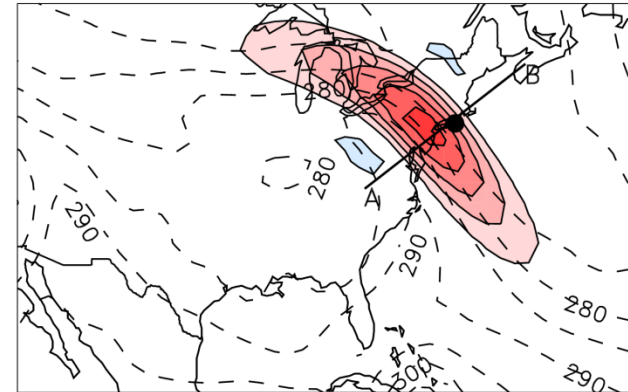

1000 hPa temperature (K) and surface pressure (hPa)   3D-Var increment   Ensemble Filter Increment

Analysis Increment (K): −1.25 −1.0 −.75 −.5 −.25 .25 .5 .75 1.0 1.25

Output from a "single-observation" experiment. The EnKF is cycled for a long time in a dry, primitive equation model. The cycle is interrupted and a single observation 1K greater than first guess is assimilated. Maps of the analysis minus first guess are plotted. These "analysis increments" are proportional to the background-error covariances between every other model grid point and the observation location.
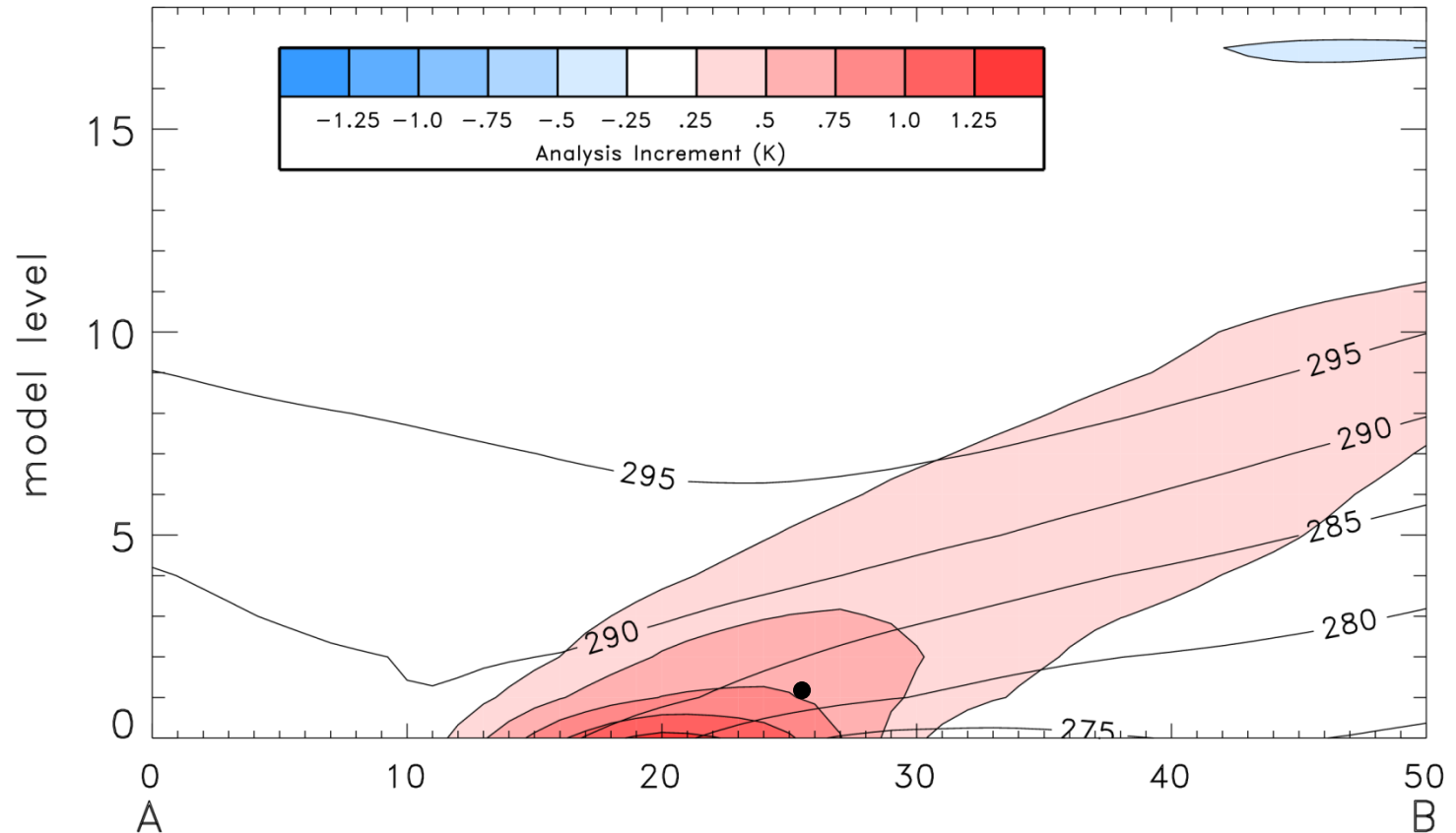
Why are the biggest increments not located at the observation?

Background and True State,
Cross Section through Front
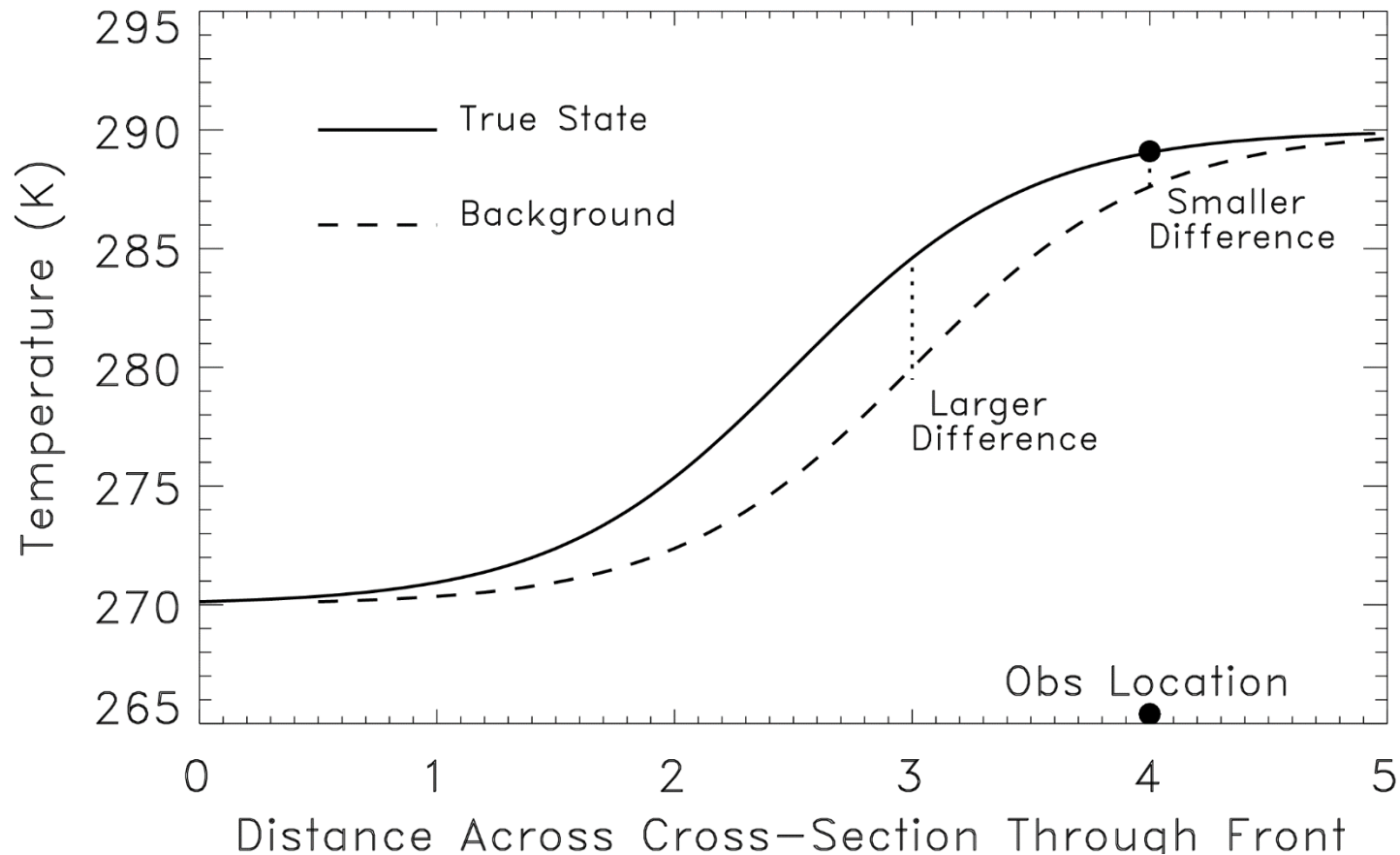
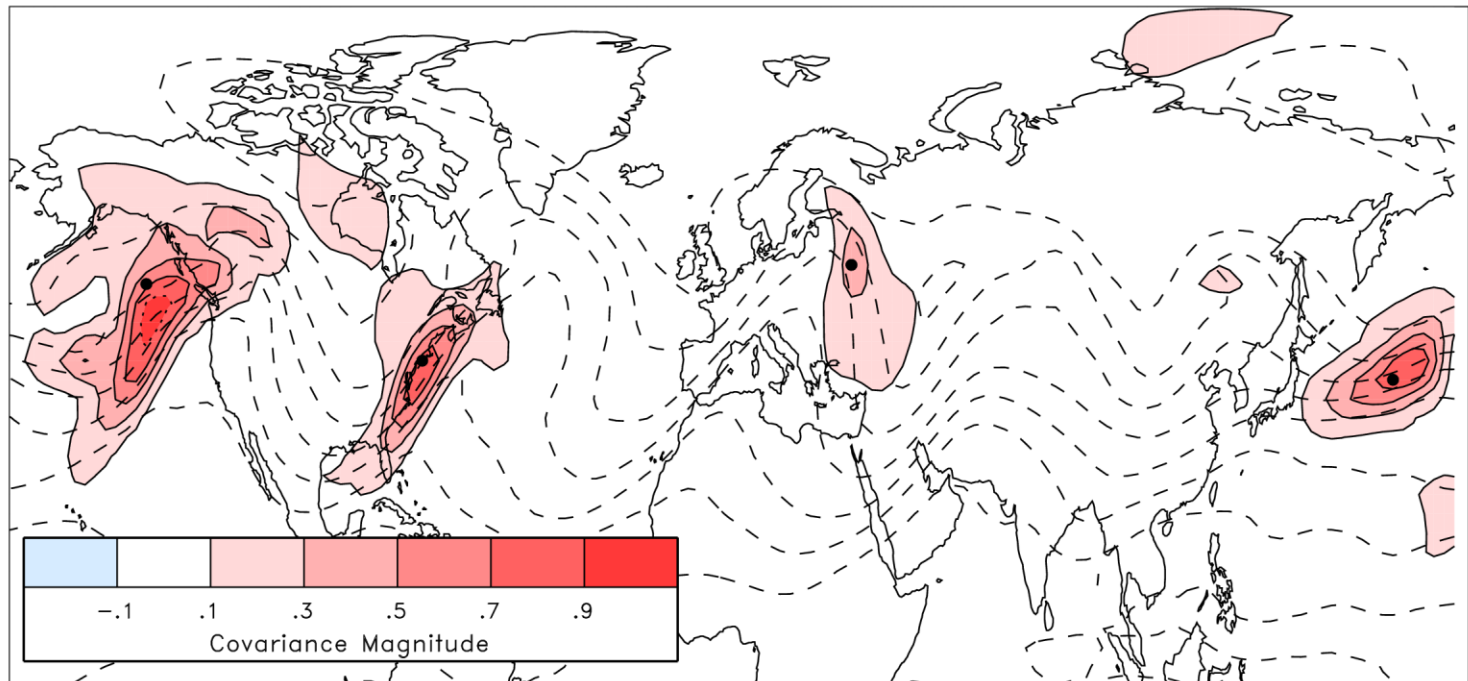# Analysis increment: vertical cross section through front

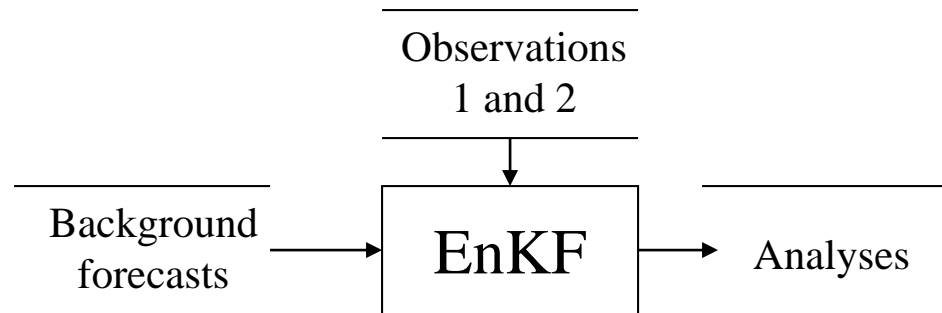Cross section of temperature increments, ensemble mean analyzed θ (K)

More examples of flow-dependent background-error covariances

Covariances of background forecasts, 900 hPa Temp

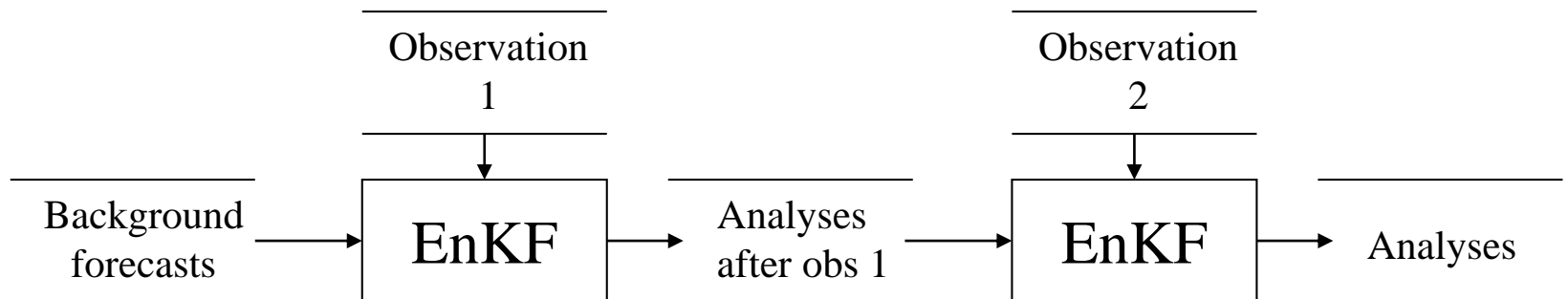# Computational trickery in EnKF:
(1) serial processing of observations

Method 1

Observations
1 and 2

Background
forecasts → EnKF → Analyses

Method 2

Observation
1

Observation
2

Background
forecasts → EnKF → Analyses
after obs 1 → EnKF → Analyses

Computational trickery in EnKF: (2) Simplifying Kalman gain calculation

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

$$define \quad \overline{H\mathbf{x}^b} = \frac{1}{m}\sum_{i=1}^{m} H\mathbf{x}_i^b$$

$$\mathbf{P}^b H^T = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathbf{x}_i^b - \overline{\mathbf{x}^b} \right)\left(H\mathbf{x}_i^b - \overline{H\mathbf{x}^b} \right)^T$$

$$H\mathbf{P}^b H^T = \frac{1}{m-1}\sum_{i=1}^{m}\left(H\mathbf{x}_i^b - \overline{H\mathbf{x}^b} \right)\left(H\mathbf{x}_i^b - \overline{H\mathbf{x}^b} \right)^T$$

The key here is that the huge matrix $\mathbf{P}^b$ is never explicitly formed

Different implementations of ensemble filters

- Double EnKF (Houtekamer and Mitchell, *MWR*, March 1998)
- Ensemble adjustment filter (EAKF; Anderson, *MWR*, Dec 2001)
- Ensemble square-root filter (EnSRF; Whitaker and Hamill, *MWR*, July 2002)
- Ensemble transform Kalman filter (ETKF; Bishop et al, *MWR*, March 2001)
- Local ETKF or LETKF (Hunt et al. 2007, Phys. D)
- Others as well (Lermusiaux, Pham, Keppenne, Heemink, etc.)

Why different implementations?

- Save computational expense
- Problems with perturbed obs. Suppose

$\mathbf{P}^b \sim N(0,1), \quad \mathbf{R} \sim N(0,1), \quad E[\text{corr}(\mathbf{x}^b, \mathbf{y}')] = 0$

Random sample $\mathbf{x}^b$: [0.19, 0.06, 1.29, 0.36, -0.61]

Random sample $\mathbf{y}'$: [-1.04, 0.46, -0.12, -0.65, 2.10]

Sample corr $(\mathbf{x}^b, \mathbf{y}') = $ - 0.61 !

Noise added through perturbed observations can introduce spurious correlations between background, observation errors. However, there may be some advantages to perturbed observations in situations where the prior is highly nonlinear. See Lawson and Hansen, MWR, Aug 2004.

Adjusting noisy background-error covariances: "covariance localization"

Ensemble Est. of Variance/Covariance

Correlation Function

Covariance after Localization
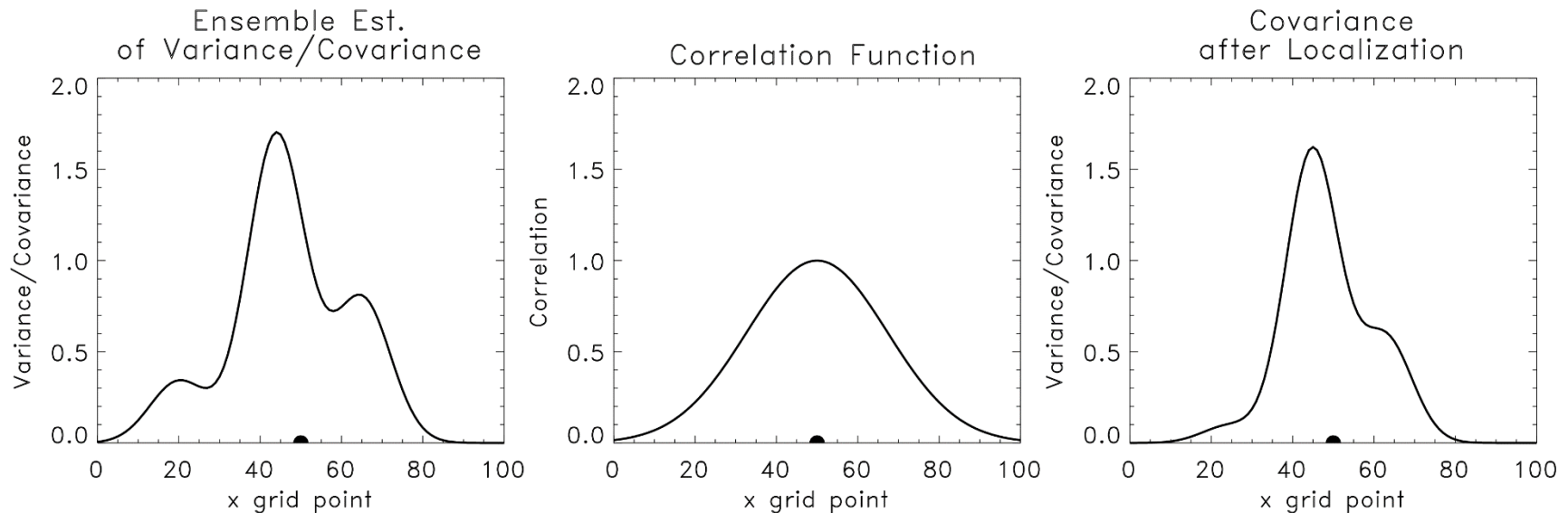
# The commonly used localization 'Gaspari-Cohn' function

Compactly supported 5th-order piecewise rational functions

$$C_0(z, 1/2, c) = \begin{cases} -\frac{1}{4}(|z|/c)^5 + \frac{1}{2}(z/c)^4 + \frac{5}{8}(|z|/c)^3 - \frac{5}{3}(z/c)^2 + 1, & 0 \leqslant |z| \leqslant c, \\ \frac{1}{12}(|z|/c)^5 - \frac{1}{2}(z/c)^4 + \frac{5}{8}(|z|/c)^3 \\ \quad + \frac{5}{3}(z/c)^2 - 5(|z|/c) + 4 - \frac{2}{3}c/|z|, & c \leqslant |z| \leqslant 2c, \\ 0, & 2c \leqslant |z|. \end{cases}$$

$$(4.10)$$

Gaussian function

$$G_0(z, L) = \exp\left(-\frac{z^2}{2L^2}\right)$$
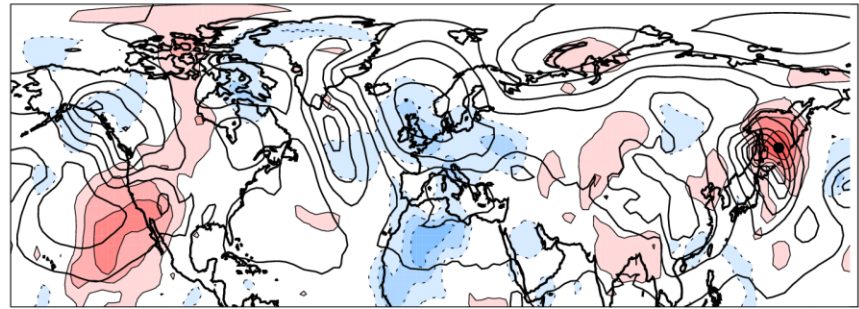

Legend:
— C0 (z,1/2,c), Eq. (4.10)
−·−· G0 (z,L), Eq. (4.12)

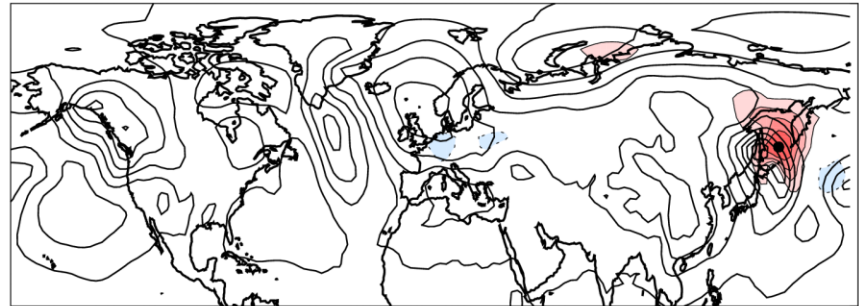The ensemble covariances is multiplied by this distance-z-based function. 2c is the cut off radius
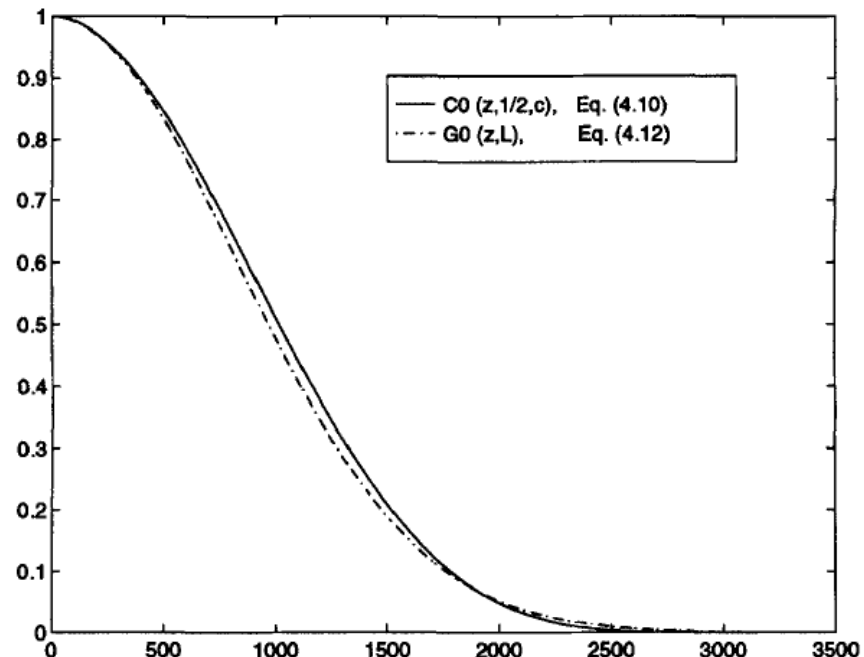
The piecewise rational function $C_0(z, 1/2, c)$ of the example in section 4(c) and the Gaussian function $G_0(z, L)$ in Eq. (4.12), for $c = 1500$ km and $L = c\,(0.3)^{1/2}$ km. See text for explanation.

Gaspari and Cohn (QJ 1999)

Covariance localization in practice

from Hamill review paper, to appear
in upcoming Cambridge Press book.
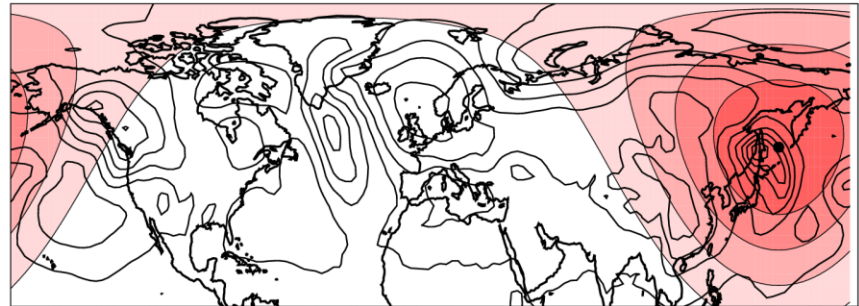See also Houtekamer and Mitchell, MWR,
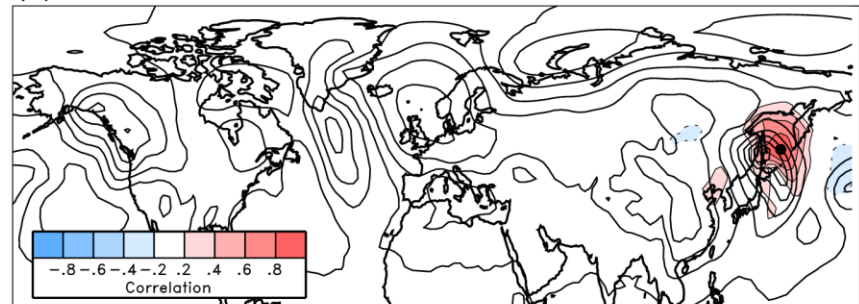March 1998

(a) Correlations in P^b, 25-member ensemble

(b) Correlations in P^b, 200-member ensemble

(c) Gaspari & Cohn correlation function

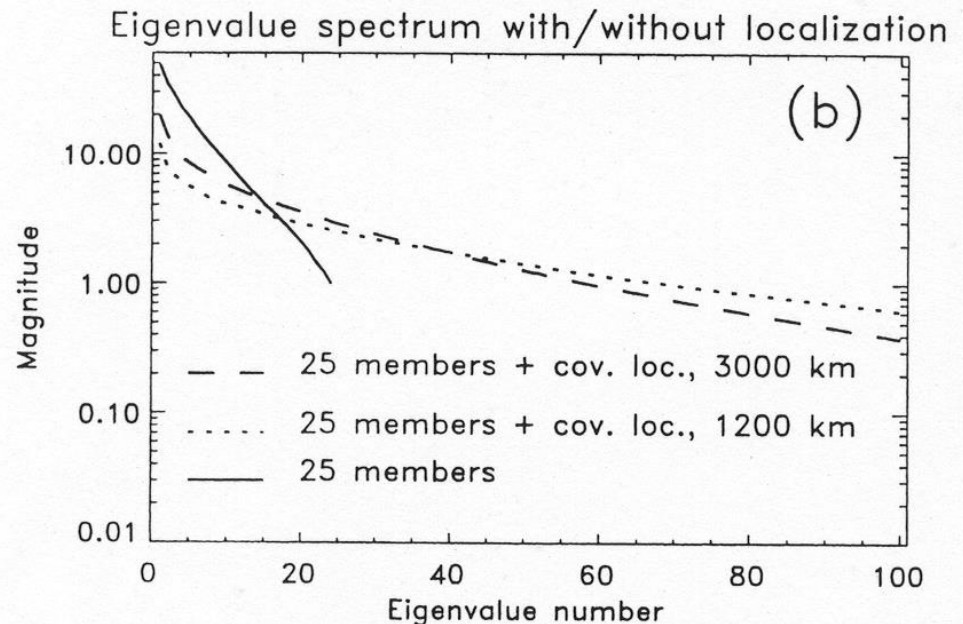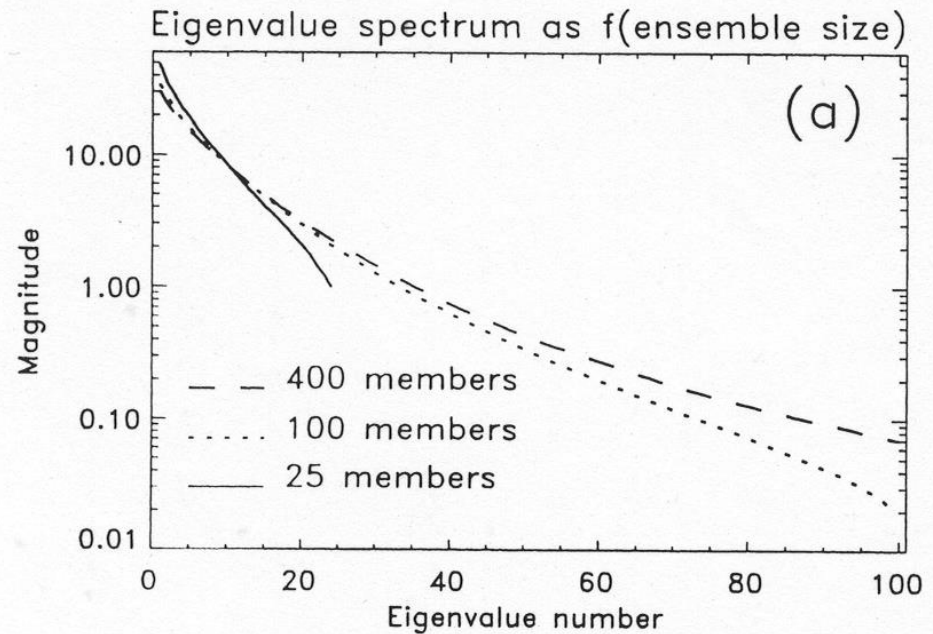(d) Correlations in P^b after localization, 25-member ensemble

−.8 −.6 −.4 −.2 .2 .4 .6 .8
Correlation

How does covariance localization make up for larger ensemble?

(a) eigenvalue spectrum from small ensemble too steep, not enough variance in trailing directions. (b) Covariance localization adds directions, flattens eigenvalue spectrum)

Eigenvalue spectrum as f(ensemble size)

(a)

- − − 400 members
- ..... 100 members
- —— 25 members

Magnitude — Eigenvalue number

Eigenvalue spectrum with/without localization

(b)

- − − 25 members + cov. loc., 3000 km
- ..... 25 members + cov. loc., 1200 km
- —— 25 members

Magnitude — Eigenvalue number

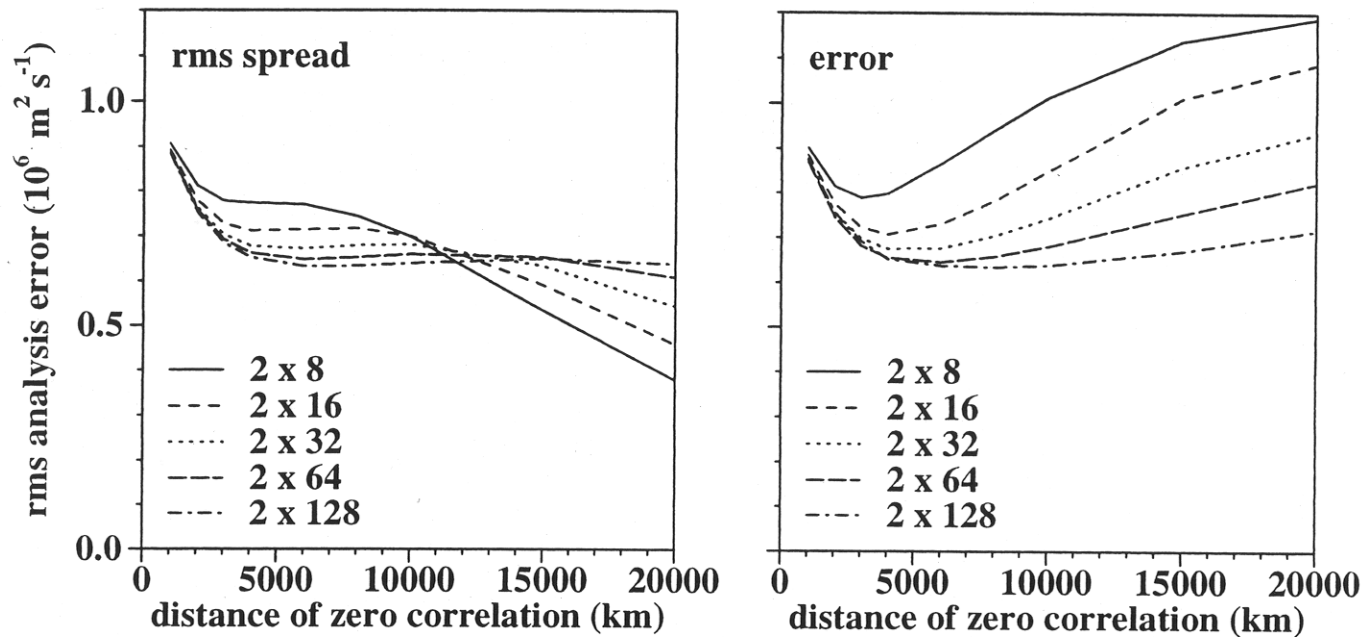# Covariance localization and size of the ensemble

FIG. 4. Analysis error as a function of r_1 (km), the distance beyond which ρ is zero, for various ensemble sizes. (left) The rms spread in the ensemble, and (right) the rms error of the ensemble mean.

Smaller ensembles achieve lowest error and comparable spread/error with a tighter localization function

(from Houtekamer and Mitchell, MWR, Jan 2001)
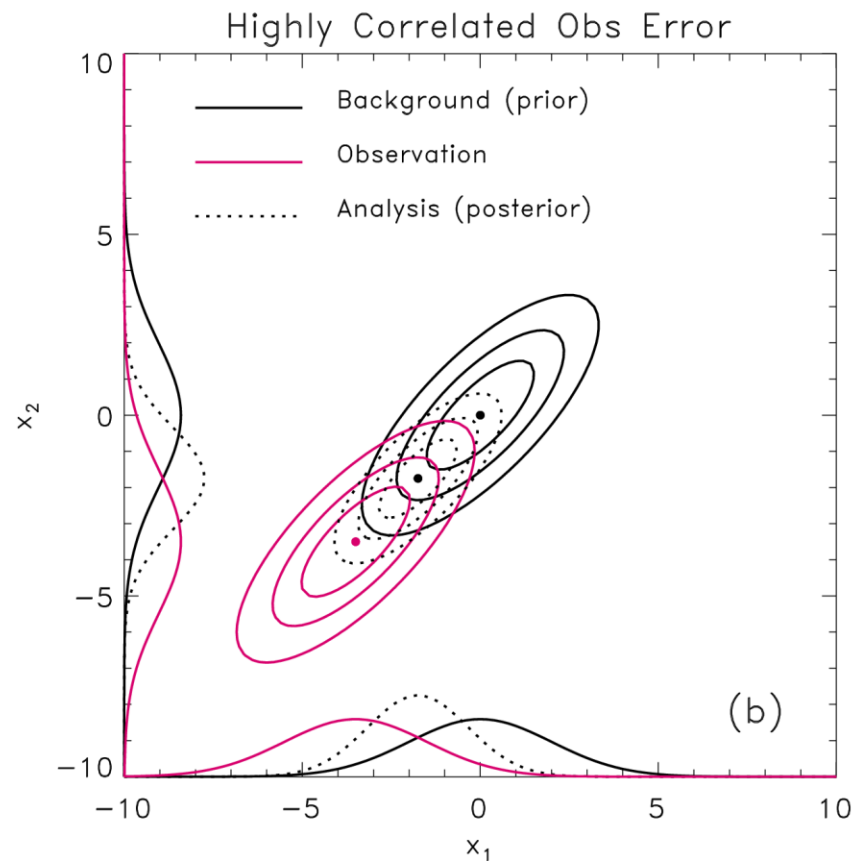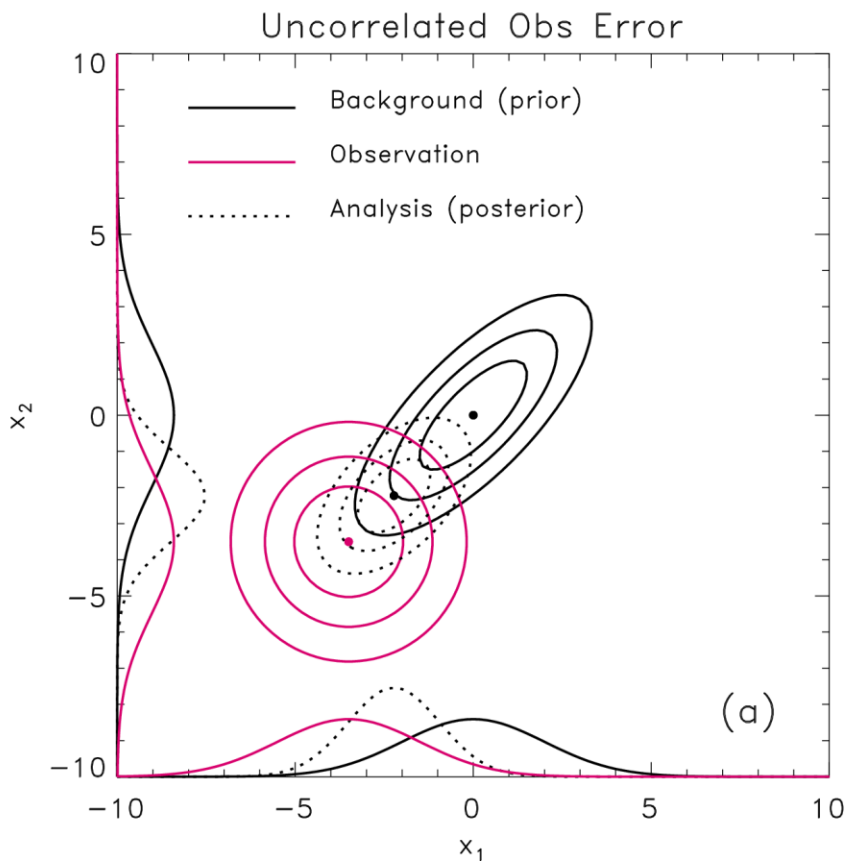
Problem: "filter divergence"

Ensemble of solutions drifts away from true solution. During data assimilation, small variance in background forecasts causes data assimilation to ignore influence of new observations.
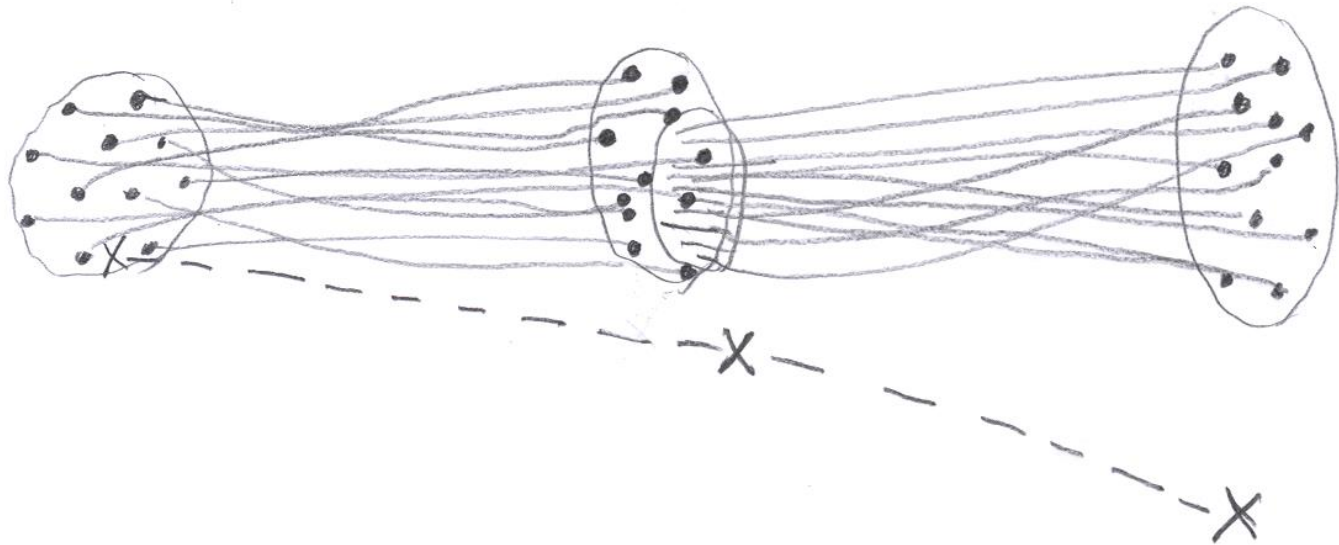
Filter divergence: some causes

- *Observation error statistics incorrect.*
- *Too small an ensemble.*
  - Poor covariance estimates just due to random sampling.
  - Not enough ensemble members. If M members and G > M growing directions, no variance in some directions.
- *Model error.*
  - Not enough resolution. Interaction of small scales with larger scales impossible.
  - Deterministic sub-gridscale parameterizations.
  - Other model aspects unperturbed (e.g., land surface condition)
  - Others

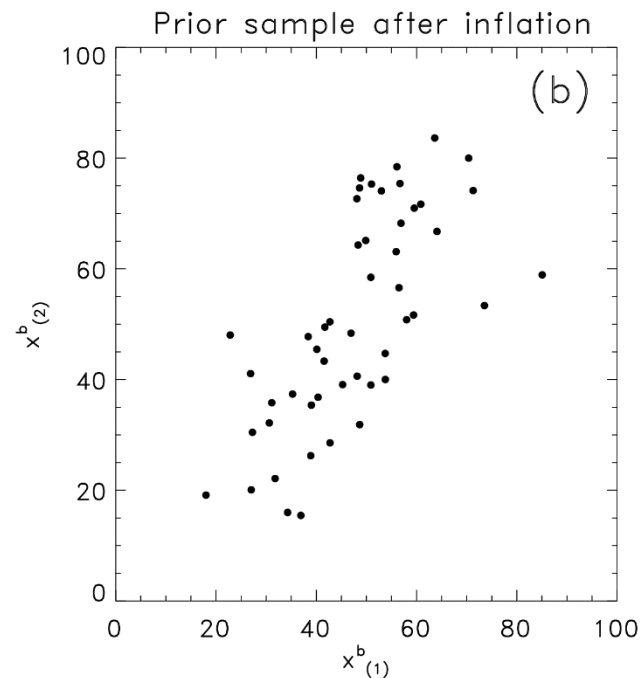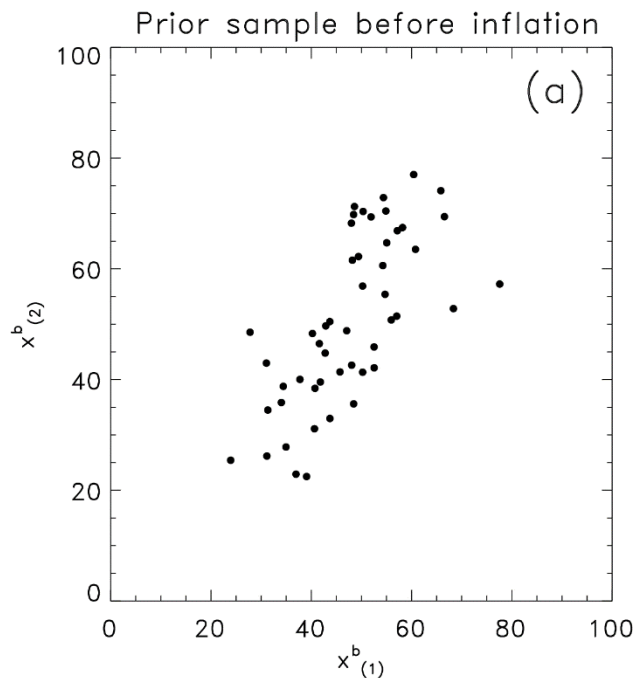# Effects of misspecifying observation error characteristics

Possible filter divergence remedies

- Higher-resolution model, more members (but costly!)
- Covariance localization (discussed before)
- Possible ways to parameterize model error
  - Apply bias correction
  - Covariance inflation
  - Integrate stochastic noise
  - Add simulated model error noise at data assimilation time.
  - Multi-model ensembles

Remedy: covariance inflation

$$\mathbf{x}_i^b \leftarrow r\left[\mathbf{x}_i^b - \overline{\mathbf{x}^b}\right] + \overline{\mathbf{x}^b}$$

r is inflation factor

Prior sample before inflation (a)

Prior sample after inflation (b)

Disadvantage: what if model error in altogether different subspace?

source: Anderson and Anderson, MWR, Dec 1999

Remedy: integrating stochastic noise

$$d\mathbf{x} = M(\mathbf{x})dt + S(\mathbf{x}, t)\,d\mathbf{W}$$

———— Deterministic

- - - - Deterministic + Random Walk
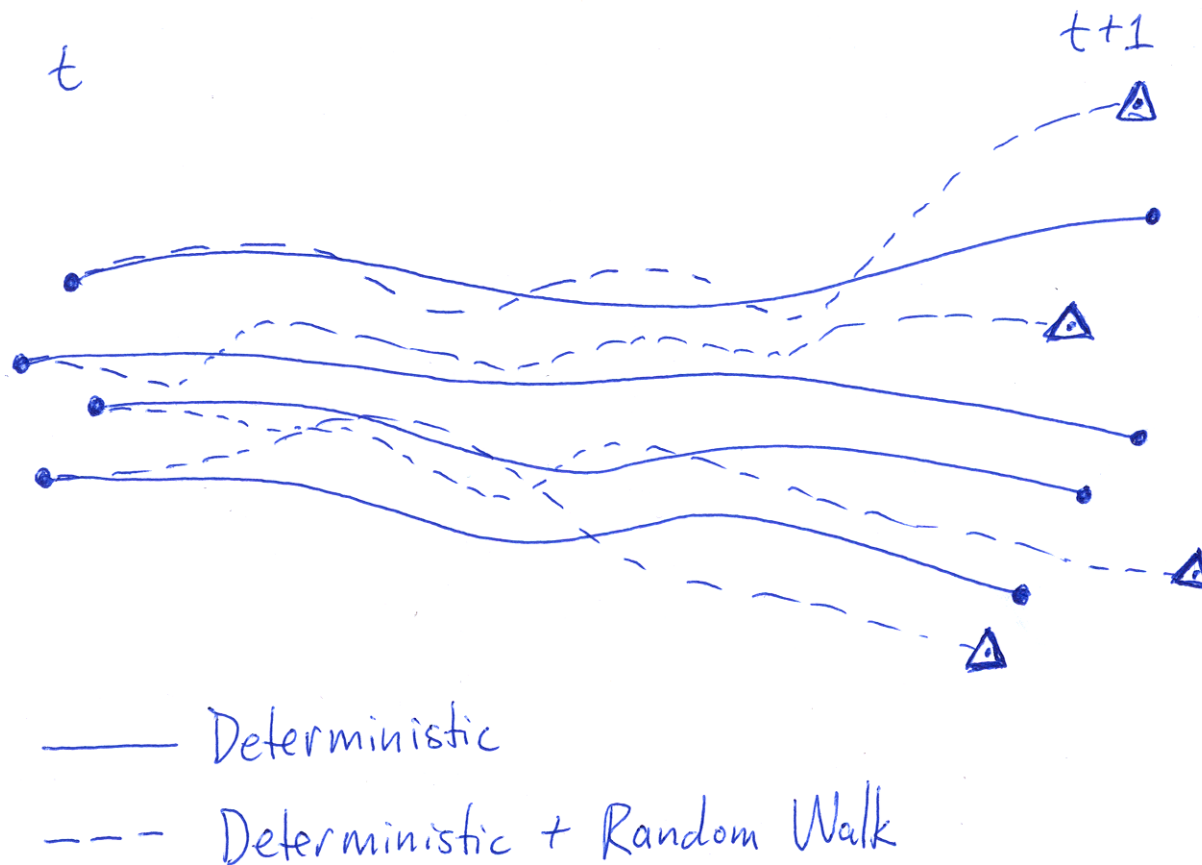
Remedy : integrating stochastic noise, continued

- Questions
  - What is structure of $S(\mathbf{x},t)$?
  - Integration methodology for noise?
  - Will this produce ~ balanced covariances?
  - Will noise project upon growing structures and increase overall variance?

- Early experiments in ECMWF ensemble to simulate stochastic effects of sub-gridscale in parameterizations (Buizza et al., *QJRMS*, 1999).

Remedy : adding noise
at data assimilation time

Idea follows Dee (Apr 1995 *MWR*) and Mitchell and
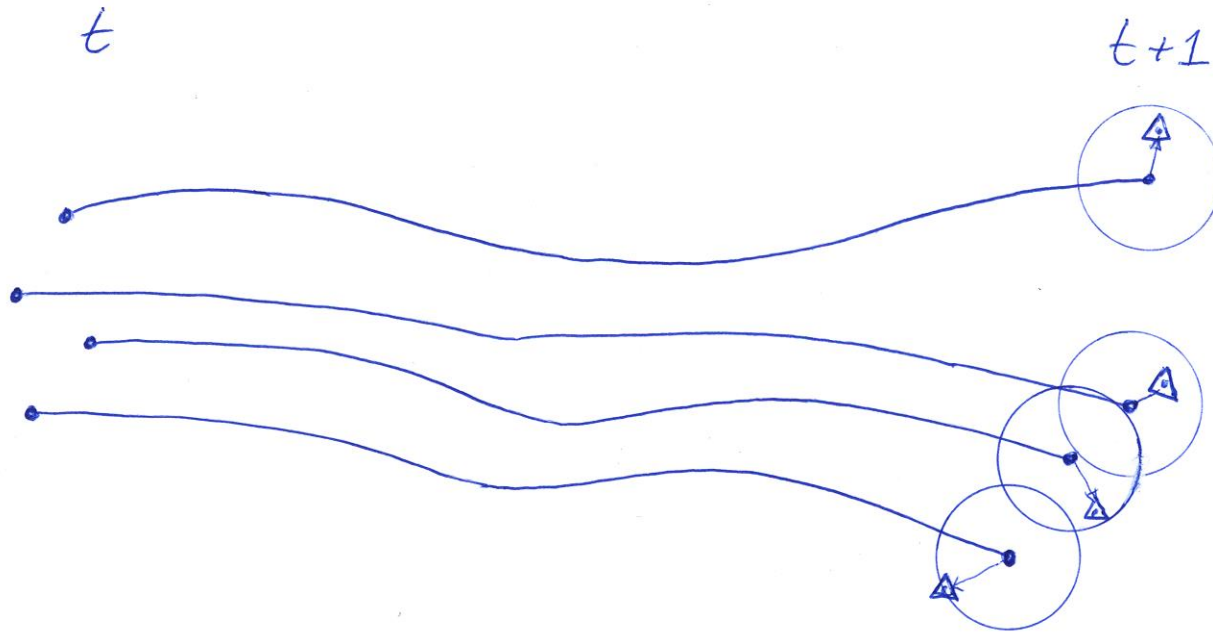Houtekamer (Feb 2000 *MWR*)

$$\mathbf{x}_{t+1} = M\mathbf{x}_t + \eta_t \qquad \left\langle \eta_t \eta_t^{\mathrm{T}} \right\rangle = \mathbf{Q} \qquad \left\langle \eta_t \right\rangle = 0$$

$$\mathbf{x}_{t+1}^{b} = M\mathbf{x}_t^{a} \qquad \left\langle \left(\mathbf{x}_{t+1}^{b} - \overline{\mathbf{x}_{t+1}^{b}} \right) \left(\mathbf{x}_{t+1}^{b} - \overline{\mathbf{x}_{t+1}^{b}} \right)^{\mathrm{T}} \right\rangle \approx \mathbf{MP}^{a}\mathbf{M}^{\mathrm{T}}$$

$$\textit{define ens. of } x_{t+1}^{b} \textit{ such that} \quad \left\langle \left(x_{t+1}^{b} - \overline{x_{t+1}^{b}} \right) \left(x_{t+1}^{b} - \overline{x_{t+1}^{b}} \right)^{\mathrm{T}} \right\rangle = \mathbf{MP}^{a}\mathbf{M}^{\mathrm{T}} + \mathbf{Q}$$

$$\textit{solution}: x_{t+1}^{b} = \mathbf{x}_{t+1}^{b} + \xi_{t+1} \qquad \left\langle \xi_{t+1} \xi_{t+1}^{\mathrm{T}} \right\rangle = \mathbf{Q} \qquad \left\langle \xi_{t+1} \right\rangle = 0$$

Remedy : adding noise
at data assimilation time

Integrate deterministic model forward to next analysis time. Then add noise to deterministic forecasts consistent with **Q.**

Remedy : adding noise
at data assimilation time (cont'd)

- Forming proper covariance model **Q** important

- Mitchell and Houtekamer: estimate parameters of **Q** from data assimilation innovation statistics. (MWR, Feb 2000)

- Hamill and Whitaker: estimate from differences between lo-res, hi-res simulations (www.cdc.noaa.gov/people/tom.hamill/modelerr.pdf)

Remedy : multi-model ensemble?

- Integrate different members with different models/different parameterizations.
- Initial testing shows covariances from such an ensemble are highly unbalanced (M. Buehner, RPN Canada, personal communication).

Other Covariance Inflation Methods

- Multiplicative inflation – multiply perturbation by a factor
- Additive inflation – add noise or perturbations to perturbations
- Relaxation to prior perturbation (RTPP) Zhang et al. (MWR 2004)

$$\mathbf{x}\,'^{a}_{new} = (1-\alpha)\mathbf{x}\,'^{a} + \alpha\mathbf{x}\,'^{f}$$

$$0 \leq \alpha < 1$$

- Relaxation to prior spread (RTPS) Anderson (Tellus 2009); Whitaker and Hamill (2010)

$$\mathbf{x}\,'^{a}_{new} = \alpha\mathbf{x}\,'^{a}, \qquad \alpha = \sqrt{c\,\frac{\sigma_b^2 - \sigma_a^2}{\sigma_a^2} + 1}$$

When c=1, σ_b is fully restored.

When c < 1, σ_b is partially restored

- Other adaptive inflation (e.g., El Gharamti MWR 2018; Gharamti et al. MWR 2019)

Conclusions

- Ensemble data assimilation literature growing rapidly because of
  - Great results in simple models
  - Coding ease (no tangent linear/adjoint)
  - Conceptual appeal (more gracefully handles some nonlinearity)
- However:
  - Computationally expensive
  - Requires careful modeling of observation, forecast-error covariances to exploit benefits. Exploration of these issues relatively new.
- Acknowledgments: Chris Snyder, Jeff Whitaker, Jeff Anderson, Peter Houtekamer
- For more information
  - www.cdc.noaa.gov/people/tom.hamill/efda_review.ppt
  - www.cdc.noaa.gov/people/tom.hamill/efda_review5.pdf

- Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation. *Predictability of Weather and Climate, T. Palmer and R. Hagedorn, Eds., Cambridge Press,* **124-156.**