NOAA-CIRES Climate Diagnostics Center

Advancing understanding and predictions of climate variability

Ensemble-Based Data Assimilation

Thomas M. Hamill

NOAA-CIRES Climate Diagnostics Center

Boulder, Colorado USA

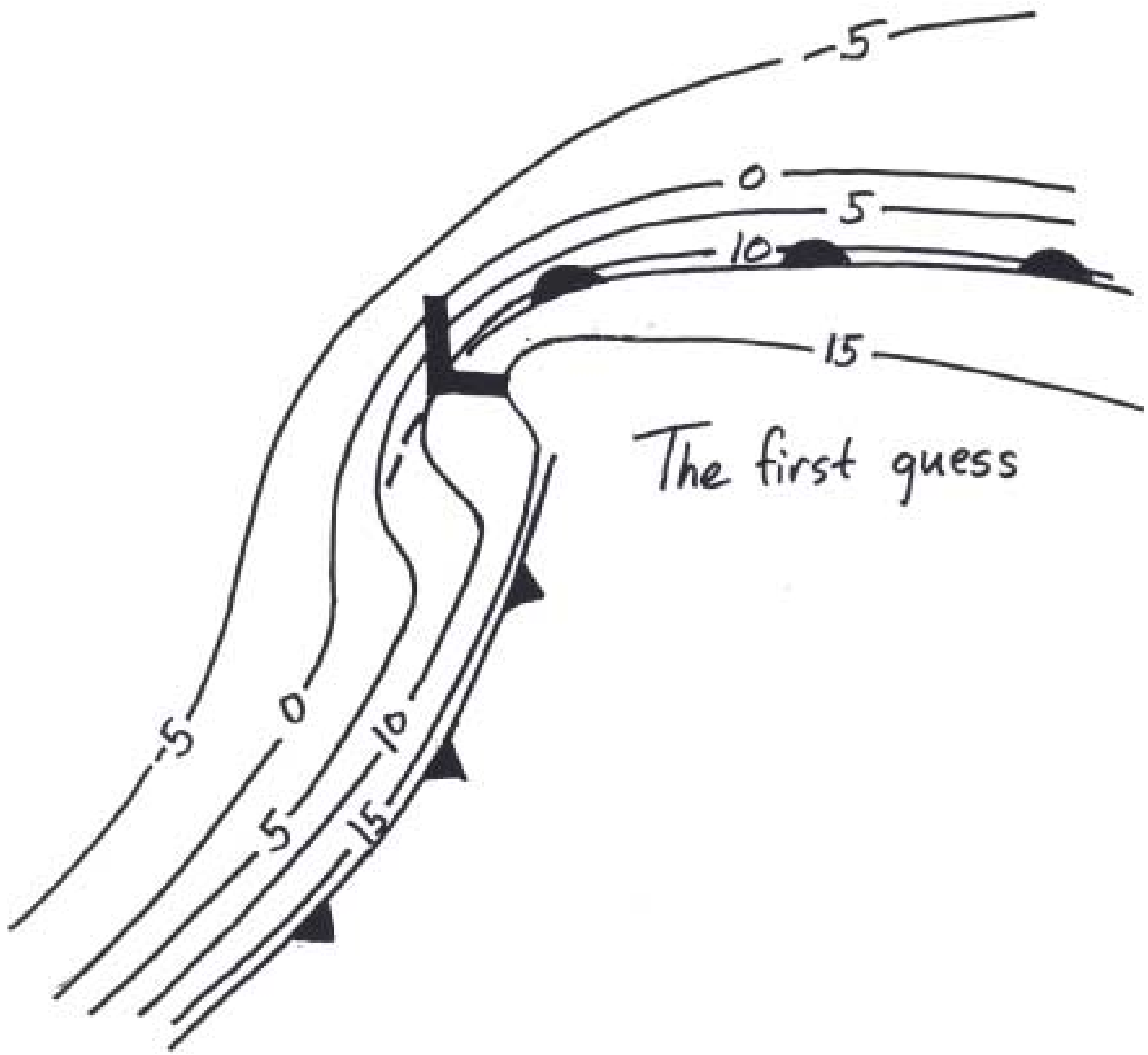
tom.hamill@noaa.gov

http://www.cdc.noaa.gov/~hamill/efda_review4.pdf

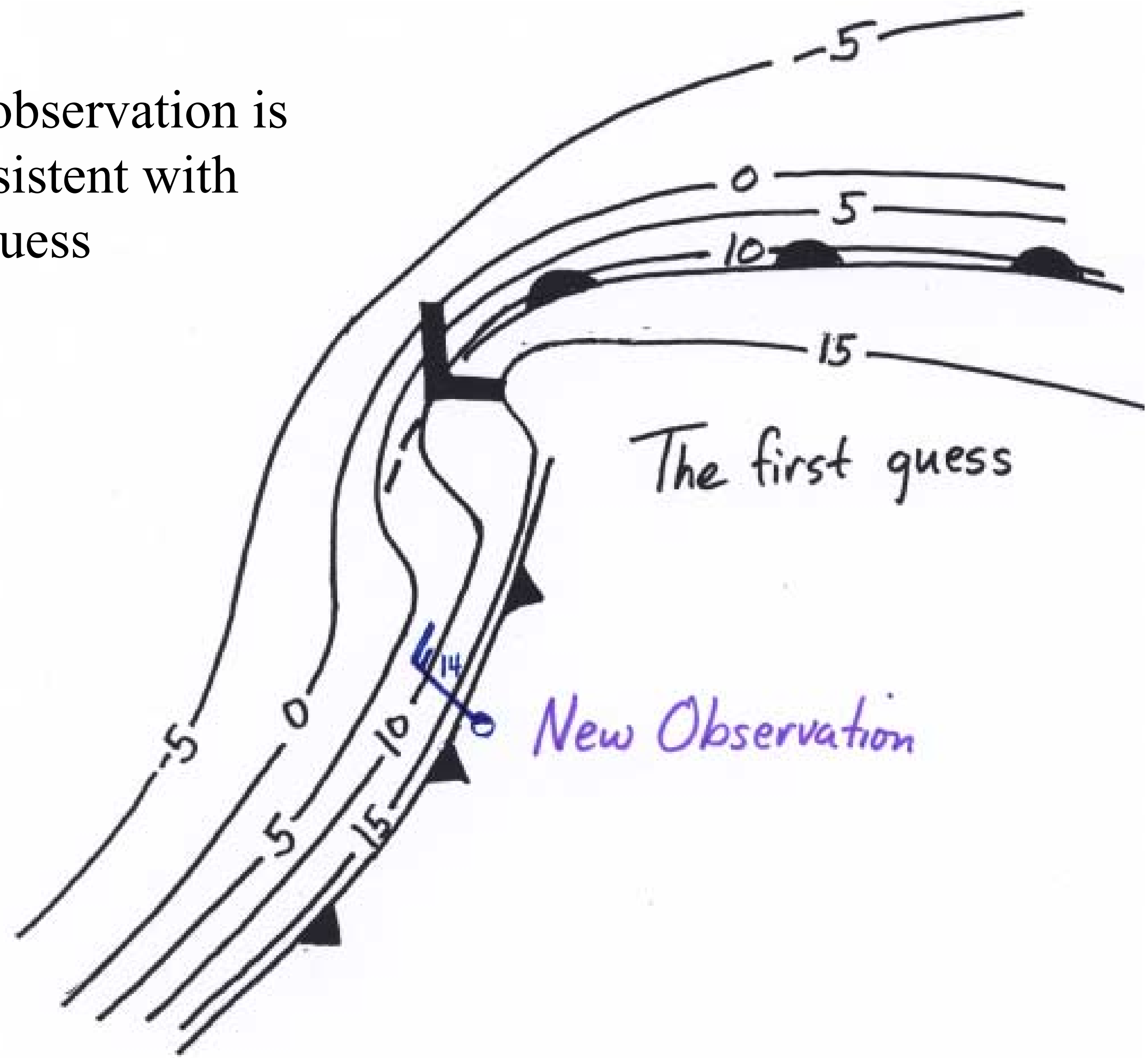
Motivation

- **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.

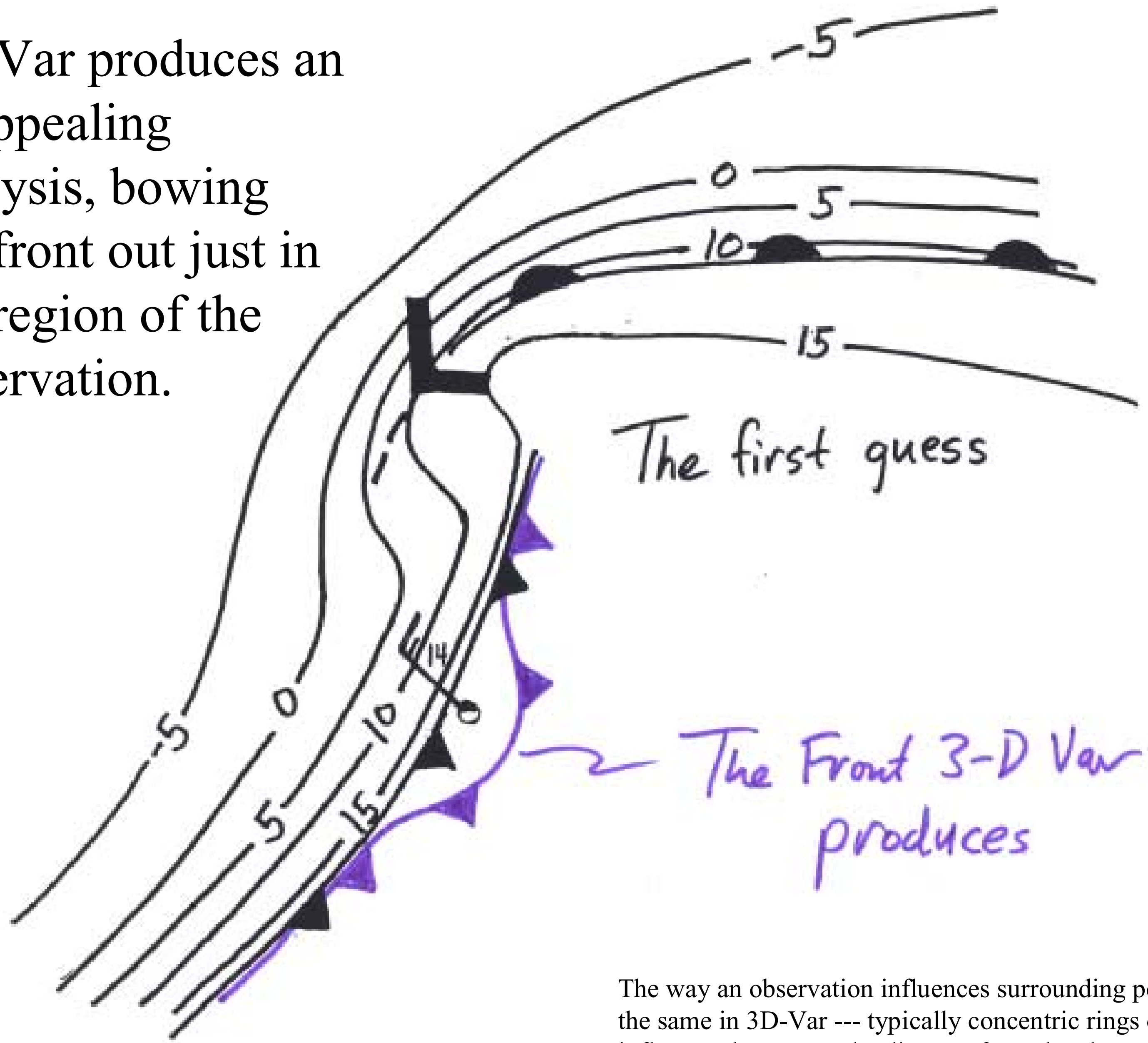
Example: where flow-dependent first-guess errors help



New observation is inconsistent with first guess

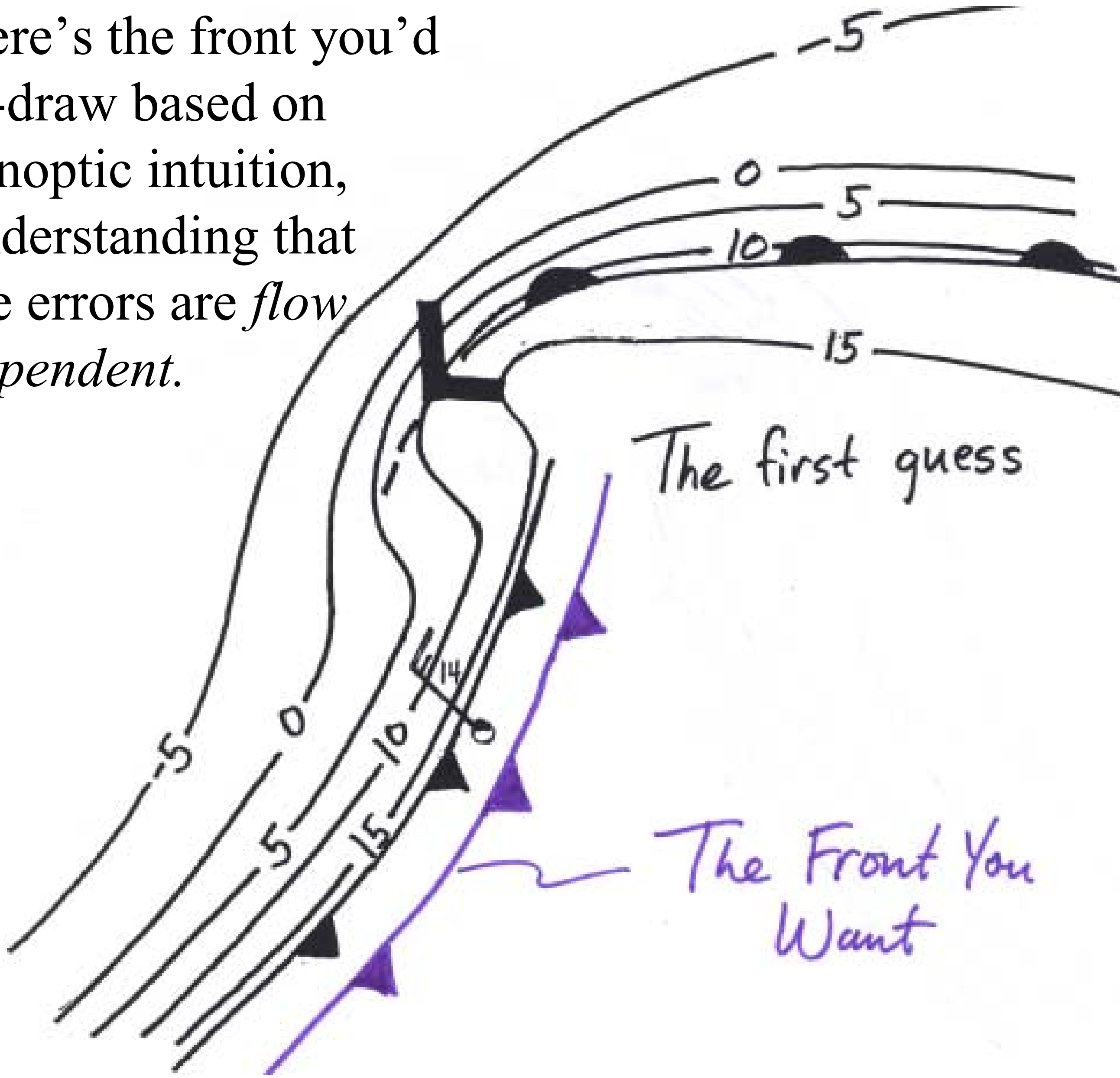


3D-Var produces an unappealing analysis, bowing the front out just in the region of the observation.



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.

Here's the front you'd re-draw based on synoptic intuition, understanding that the errors are *flow dependent*.



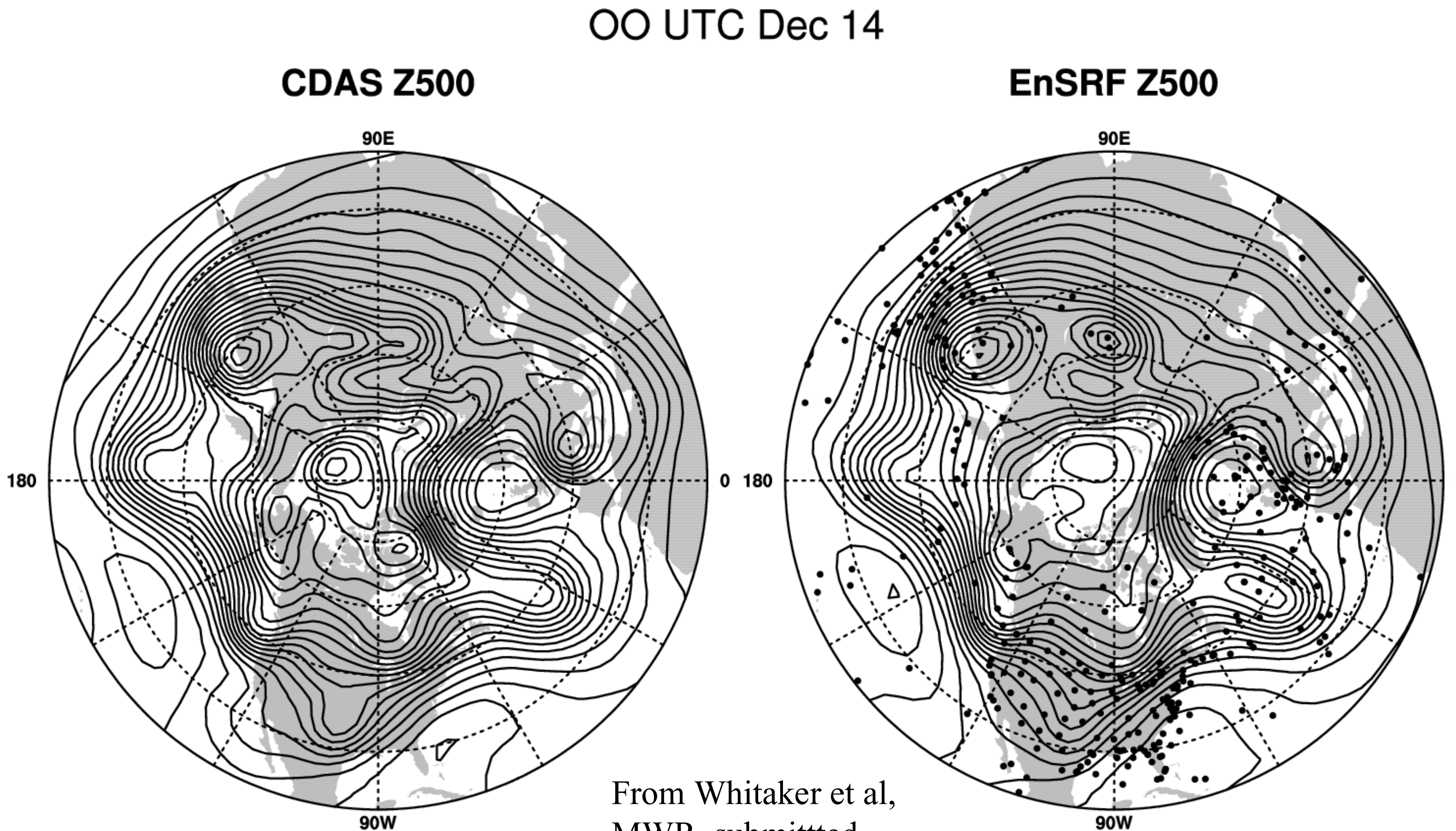
In this special case, your synoptic intuition would tell you that temperature errors ought to be correlated along the front, from northeast to southwest. However, you wouldn't expect that same NE-SW correlation if the observation were near the warm front. Analyses might be improved dramatically if the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

How can we teach data assimilation algorithms to do this?

Ensemble-based data assimilation algorithms

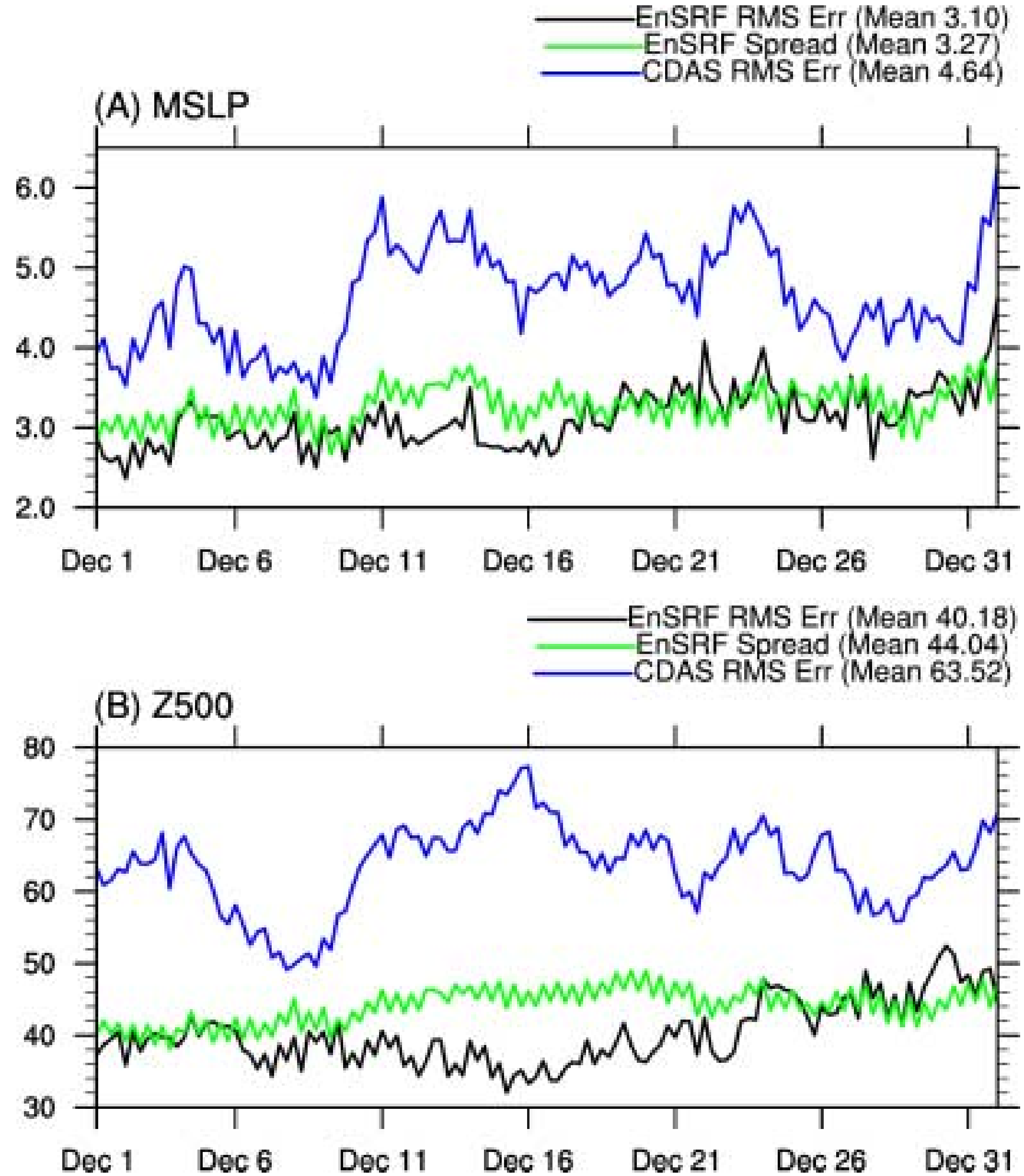
- Can use ensemble to model the statistics of the first guess ("background") errors.
- Better blending of observations and first guess.
- Initial tests show dramatically improved sets of objective analyses.
- These sets of objective analyses are exactly the sort of initial conditions we need to initialize ensemble forecasts.

Example: Assimilation of sparse sea-level pressure data into T62 GCM



From Whitaker et al,
MWR, submittted

Example: assimilation of sparse sea-level pressure data into T62 GCM



From Whitaker et al, MWR, submittted

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 get around these problems.

From first principles: Bayesian data assimilation

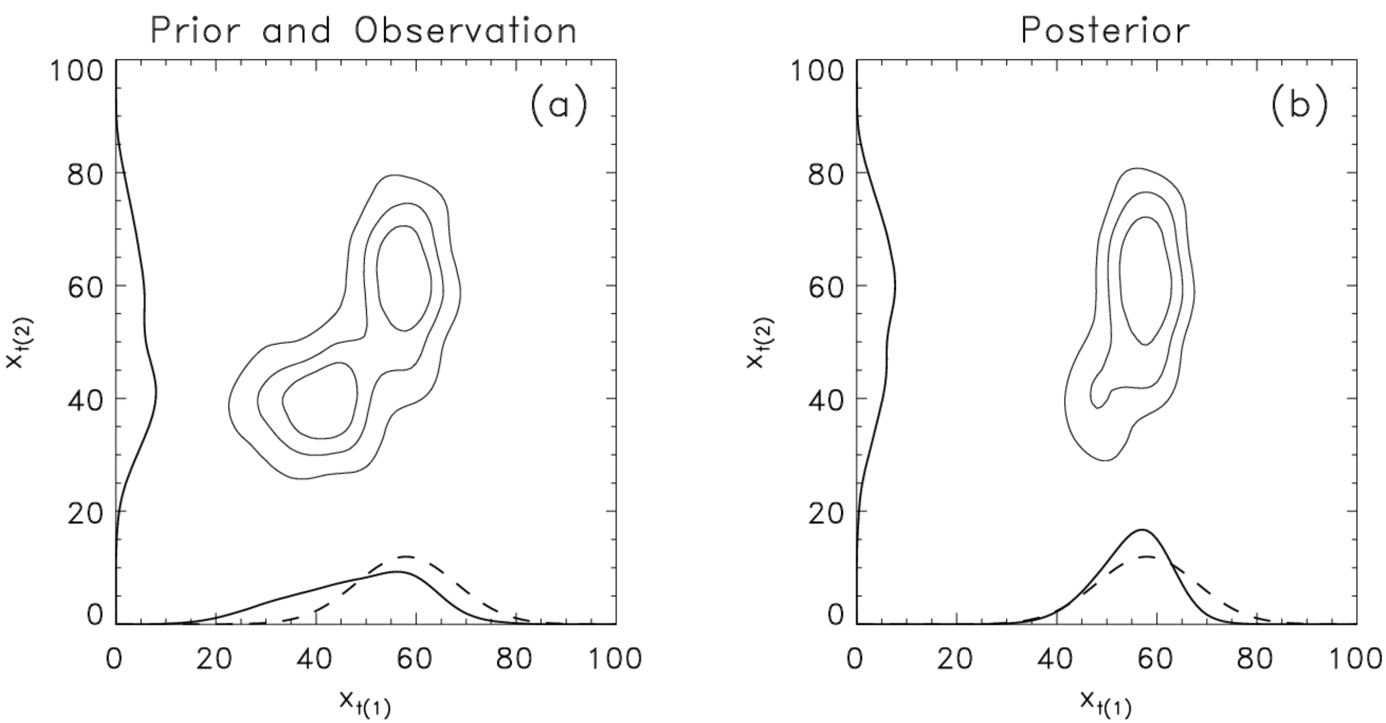
$\mathbf{x}_t$ = unknown true value of model state

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

$P\left(\mathbf{x}_t \mid \psi_t\right) \propto P\left(\mathbf{y}_t \middle| \mathbf{x}_t\right) P\left(\mathbf{x}_t \middle| \psi_{t-1}\right)$ Baye's Rule

"posterior" "prior"

Bayesian data assimilation: 2-D example



Computationally expensive!

Data assimilation terminology

- $\mathbf{y}$: Observation vector (raobs, satellite, etc.)
- $\mathbf{x}^b$: Background state vector (1st guess)
- $\mathbf{x}^a$: Analysis state vector
- $\mathbf{H}$: Operator to convert model state 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

Assume

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

$$P(\mathbf{y}_\mathrm{t}\mid\mathbf{x}_\mathrm{t}) \sim N(\mathbf{y}_\mathrm{t},\mathbf{R}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{H}\mathbf{x}_\mathrm{t}-\mathbf{y}_\mathrm{t}\right)^\mathrm{T}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{x}_\mathrm{t}-\mathbf{y}_\mathrm{t}\right)\right)$$

Apply Baye’s rule;

$$P\left(\mathbf{x}_\mathrm{t}\left|\psi_t\right.\right) \propto P\left(\mathbf{y}_\mathrm{t}\left|\mathbf{x}_\mathrm{t}\right.\right)P\left(\mathbf{x}_\mathrm{t}\left|\psi_{\mathrm{t-1}}\right.\right)$$

$$P(\mathbf{x}_\mathrm{t}\mid\psi_\mathrm{t}) \propto \exp\left(-\frac{1}{2}\left[\left(\mathbf{x}_\mathrm{t}-\mathbf{x}^\mathrm{b}\right)^\mathrm{T}\mathbf{P}^{\mathrm{b}-1}\left(\mathbf{x}_\mathrm{t}-\mathbf{x}^\mathbf{b}\right)+\left(\mathbf{H}\mathbf{x}_\mathrm{t}-\mathbf{y}\right)^\mathrm{T}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{x}_\mathrm{t}-\mathbf{y}\right)\right]\right) \qquad (@)$$

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

Kalman filter derivation continued

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

After much math, 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 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

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

$\mathbf{P}^a$ 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 Assumed . $\mathbf{x}_{t+1} = \mathbf{M}(\mathbf{x}_t) + \eta_t, \quad \langle \eta_t \eta_t^T \rangle = \mathbf{Q}$

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

Kalman filter update equation: example in 1-D

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

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

Kalman filter problems

- Covariance propagation step very expensive
- Still need tangent linear/adjoint models for evolving covariances, linear error growth assumption still 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.

Ensemble Kalman filter equations

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

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

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

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

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

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

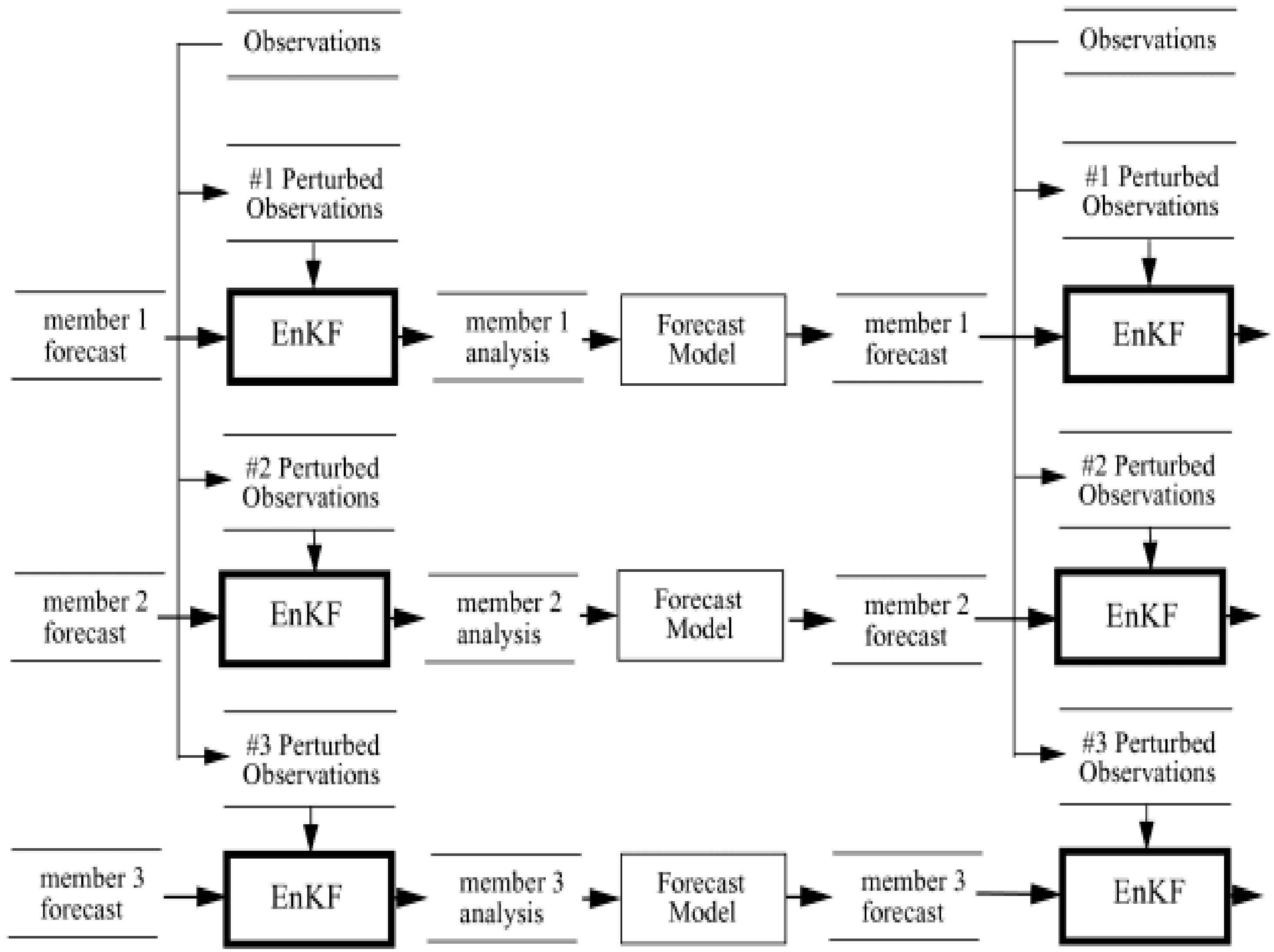
$$\mathbf{y}_\mathrm{i} = \mathbf{y} + \mathbf{y}_\mathrm{i}^{'}$$

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

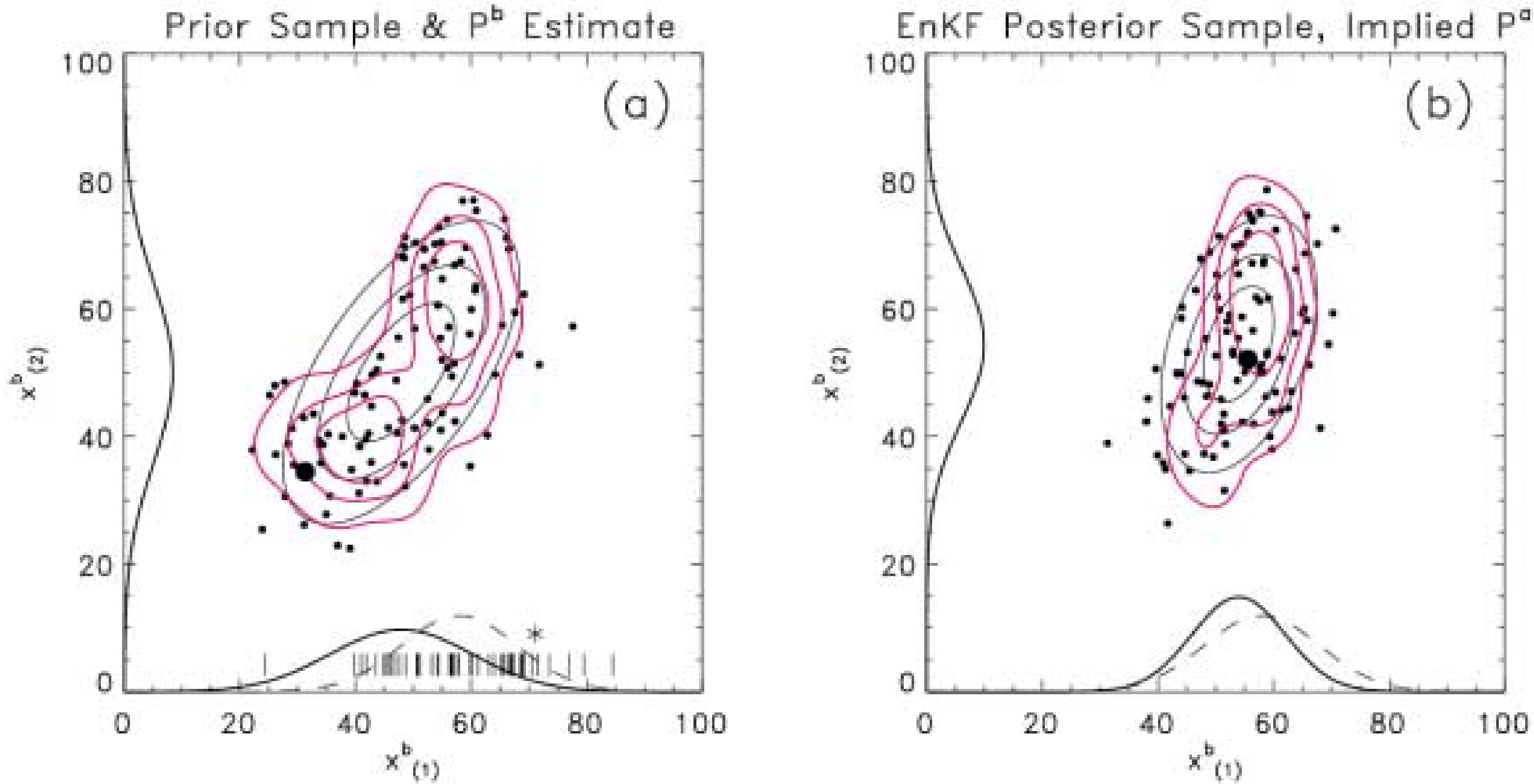
(1) An ensemble of parallel data assimilation cycles is conducted, assimilating perturbed observations .

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

The ensemble Kalman filter: a schematic

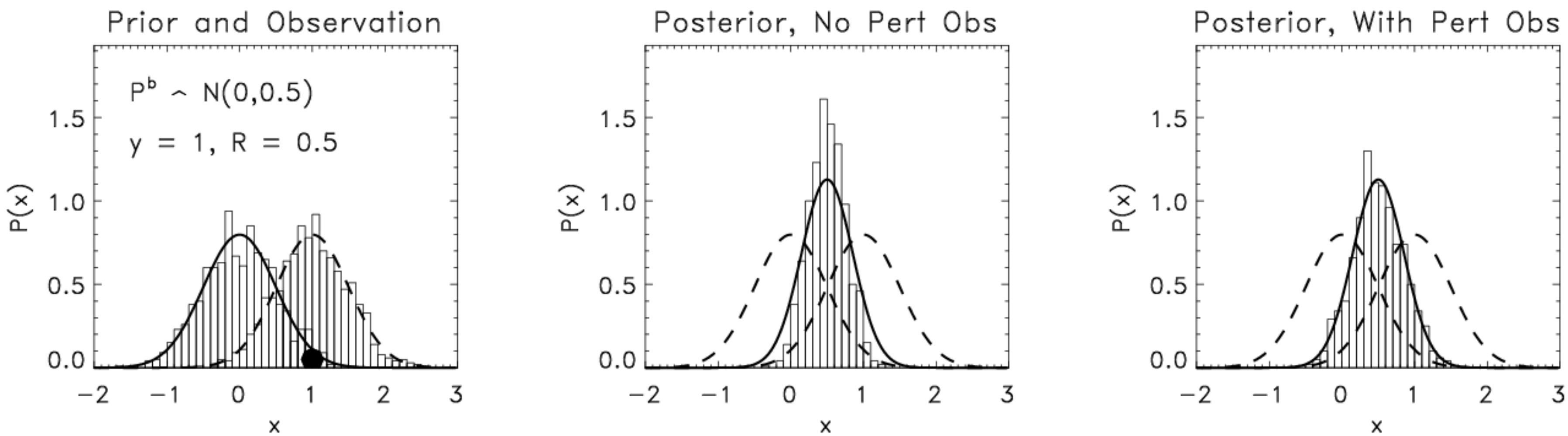


How the EnKF works: 2-D example



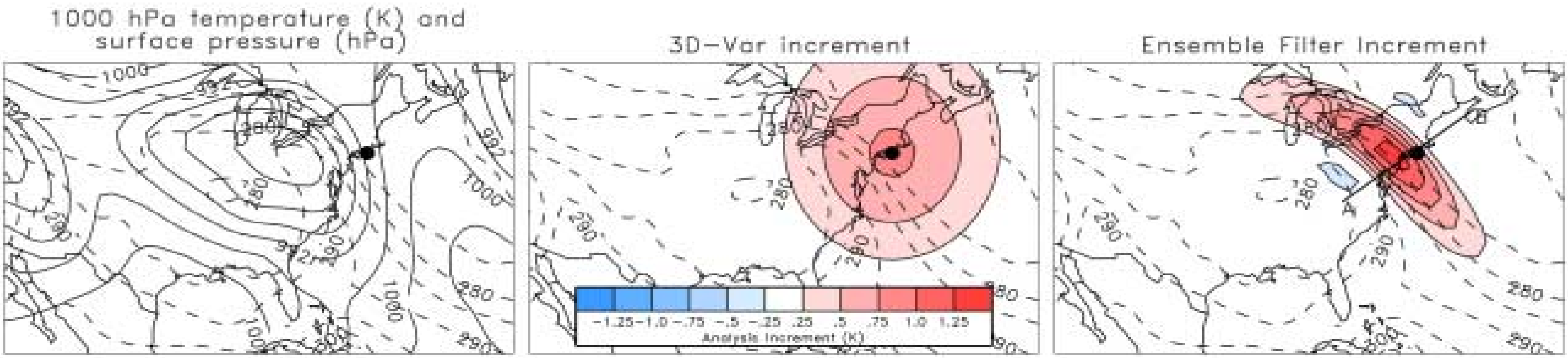
Start with a random sample from bimodal distribution used in previous Bayesian data assimilation example. Contours reflect the Gaussian distribution fitted to ensemble data.

Why perturb the observations?



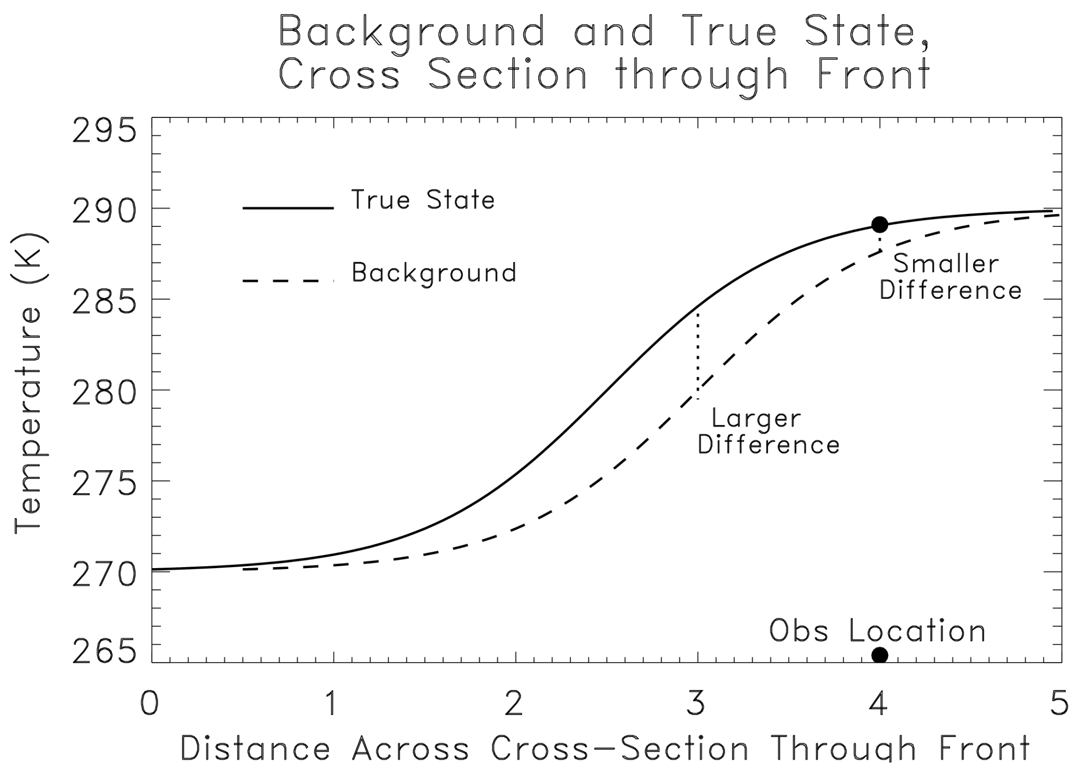
Perturbing the observations provides an easy way to ensure that the analysis-error covariances match those expected from the Kalman filter. There are other ensemble assimilation algorithms that get around perturbed obs (e.g., Whitaker and Hamill, MWR, July 2002)

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

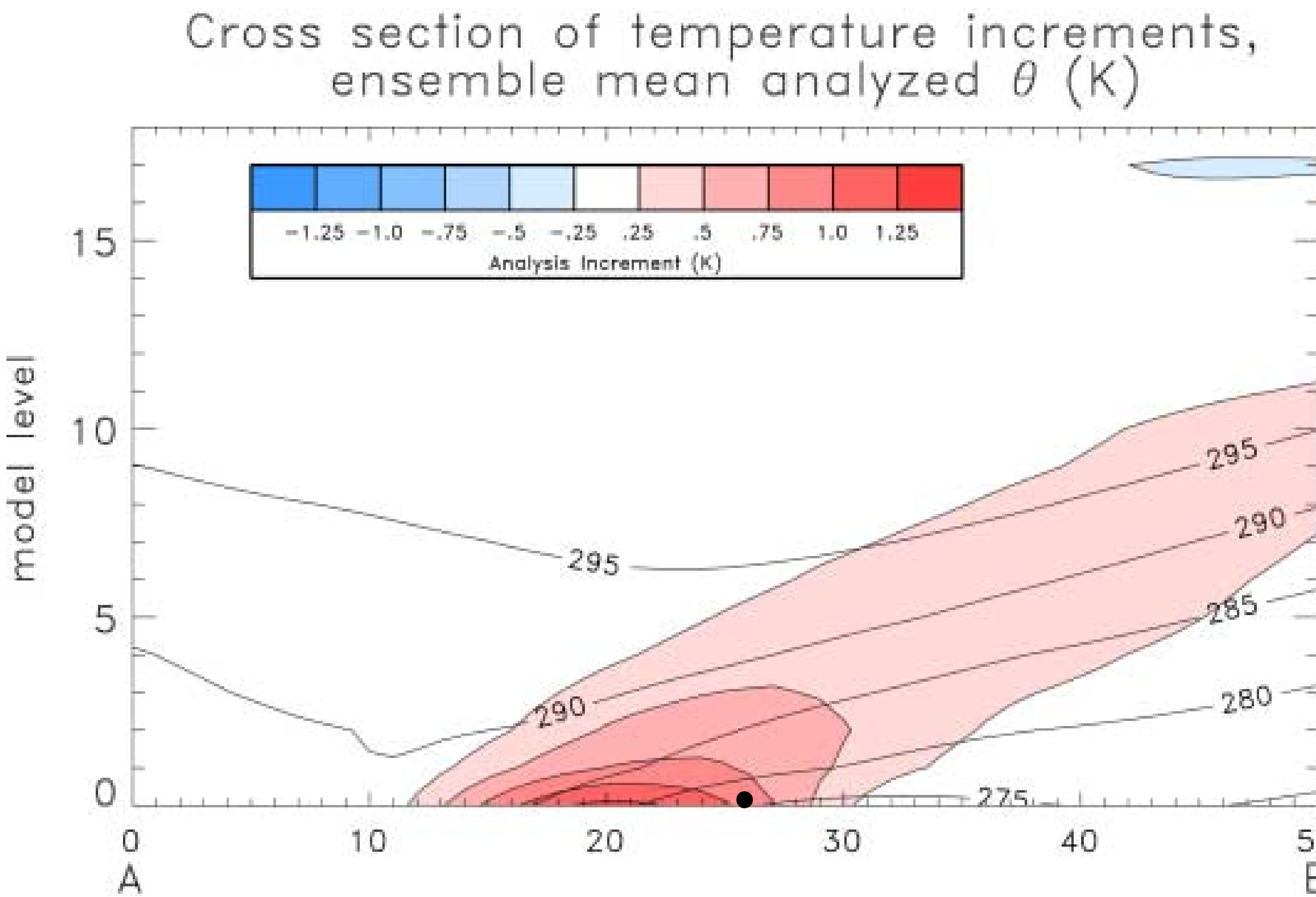


Output from a “single-observation” experiment. EnKF is cycled for a long time in a dry, primitive equation model. The cycle is interrupted and a single observation 1K greater than first guess is assimilated. Maps of the analysis minus first guess are plotted. These “analysis increments” are proportional to the background-error covariances between every other model grid point and the observation location.

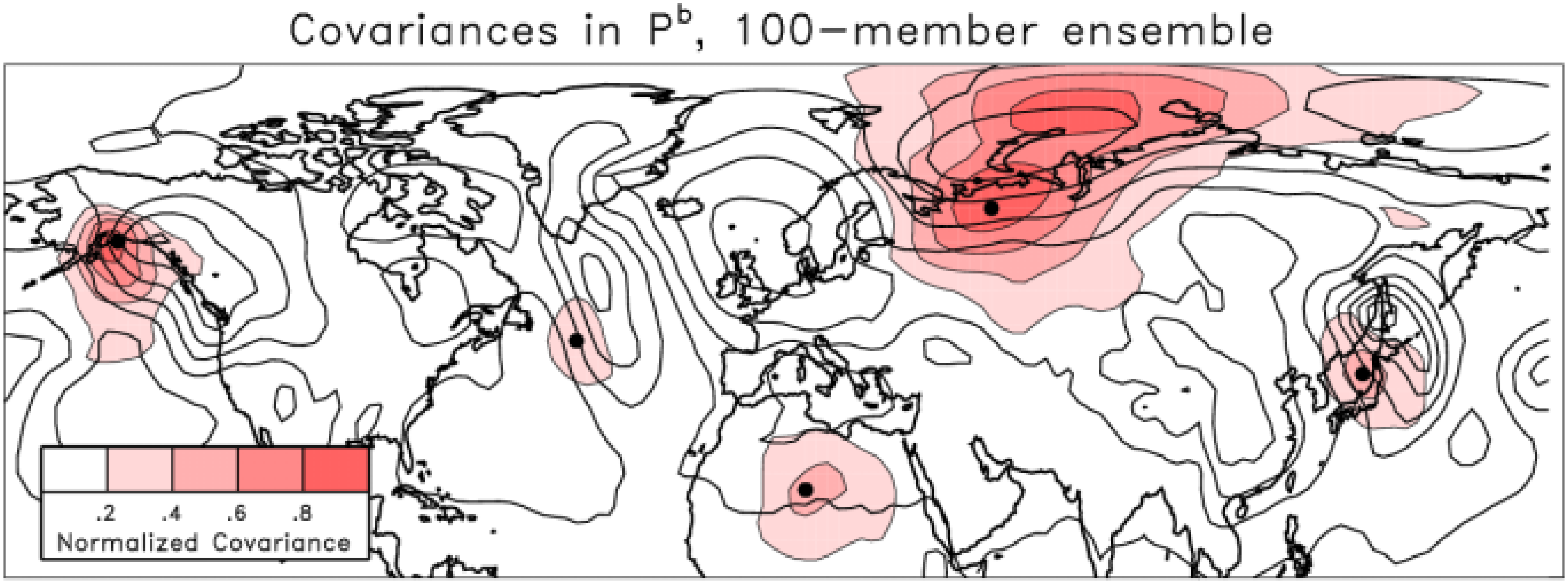
Why are the biggest increments not located at the observation?



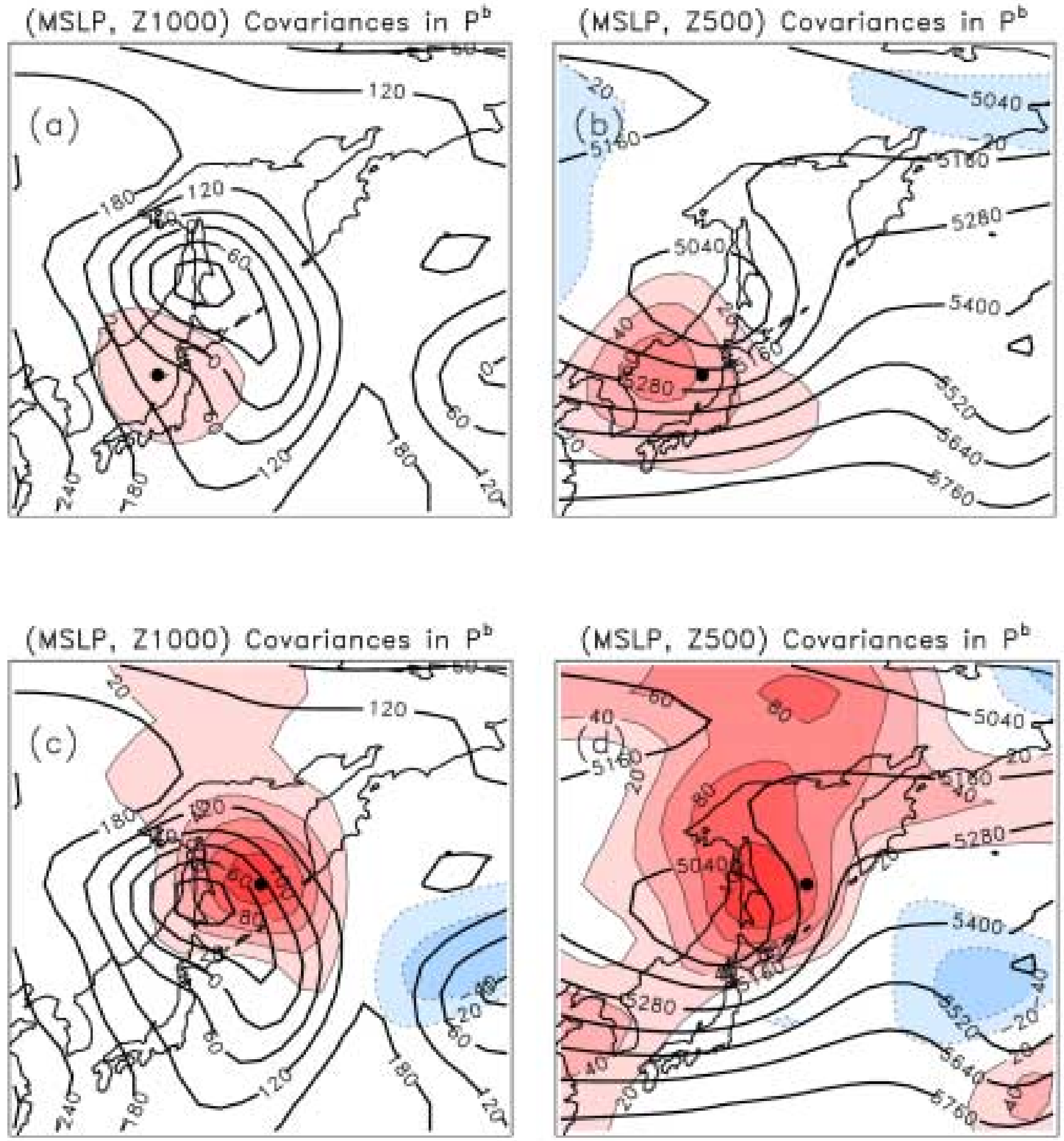
Analysis increment: vertical cross section through front



More examples of flow-dependent background-error covariances

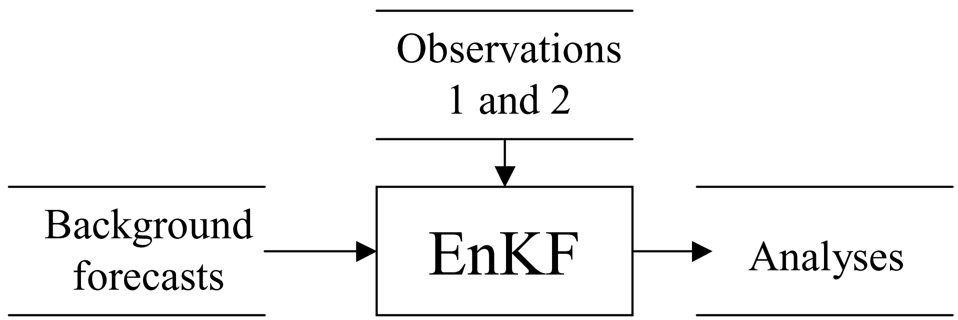


More examples of flow-dependent background-error covariances

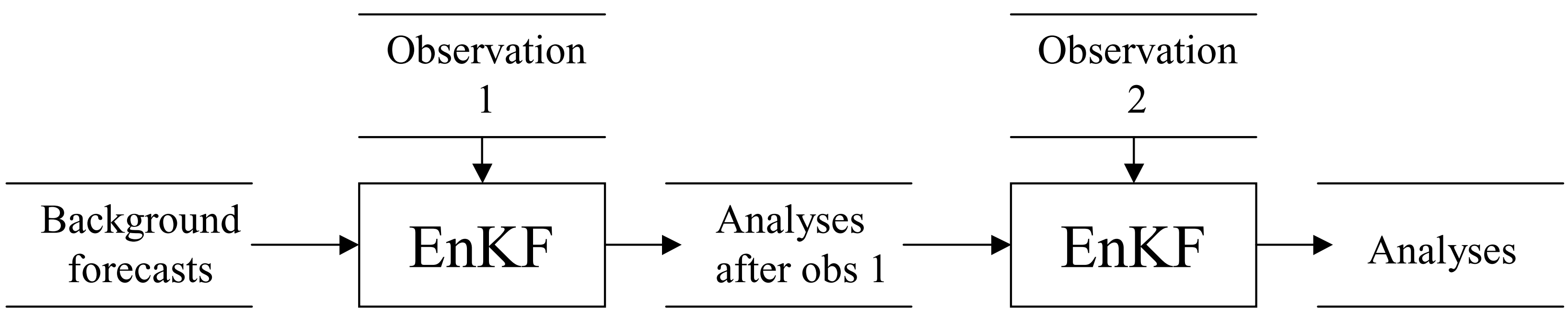


Computational trickery in EnKF: (1) serial processing of observations

Method 1



Method 2



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

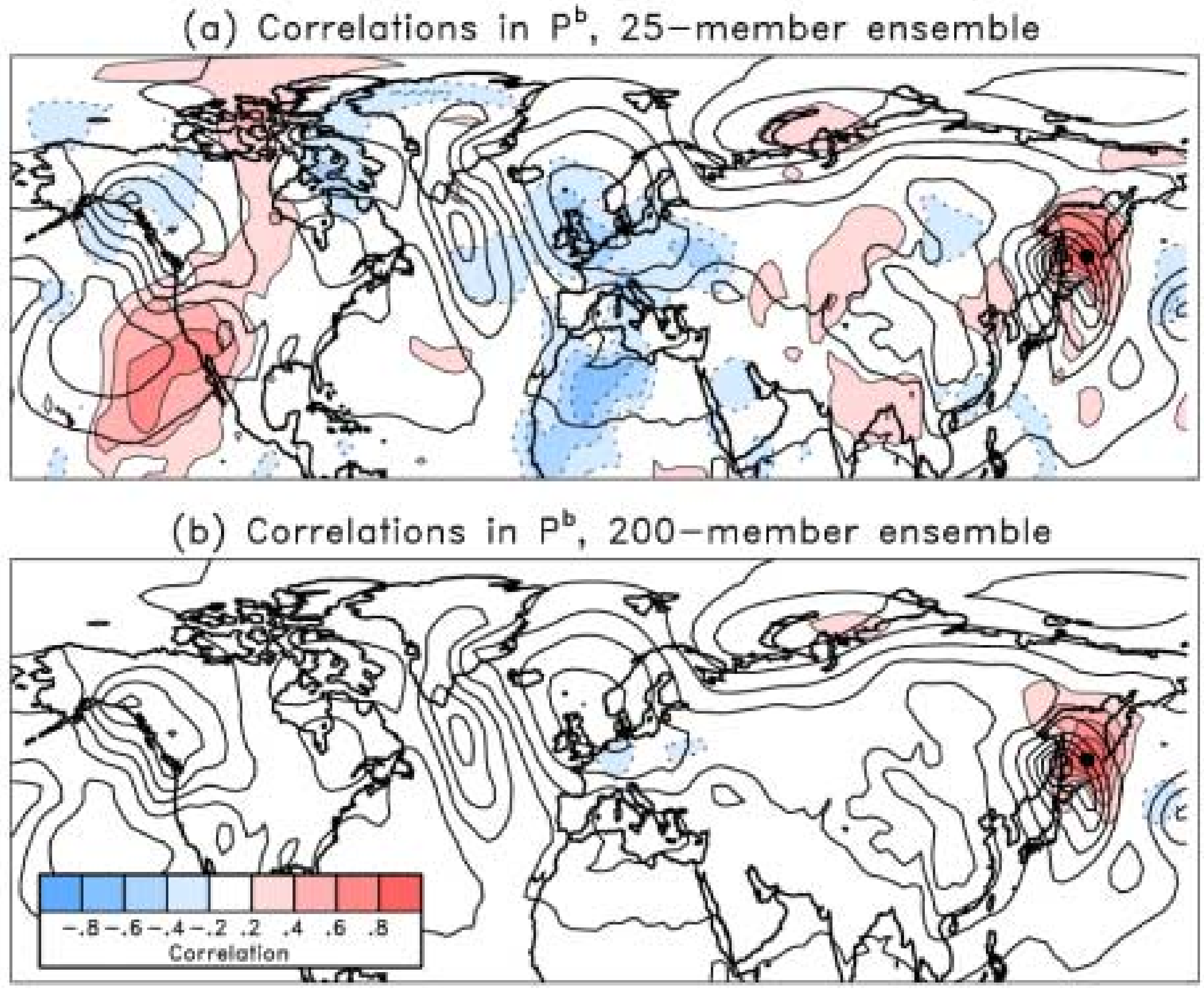
$$\mathbf{K} = \mathbf{P}^{\mathrm{b}} H^{\mathrm{T}} \left(H \mathbf{P}^{\mathrm{b}} H^{\mathrm{T}} + \mathbf{R} \right)^{-1}$$

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

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

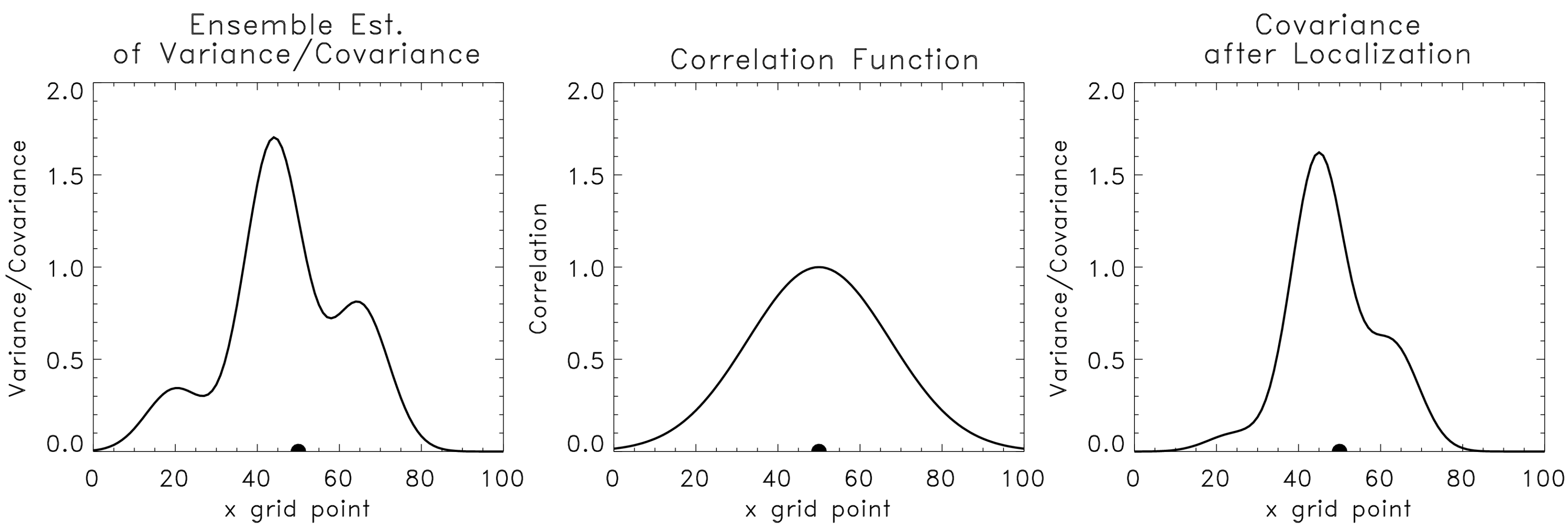
$$H\mathbf{P}^{\mathrm{b}} H^{\mathrm{T}} = \frac{1}{m-1} \sum_{i=1}^{m} \left(H\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{H\mathbf{x}^{\mathrm{b}}} \right) \left(H\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{H\mathbf{x}^{\mathrm{b}}} \right)^{\mathrm{T}}$$

Issues with EnKF: noisy covariances

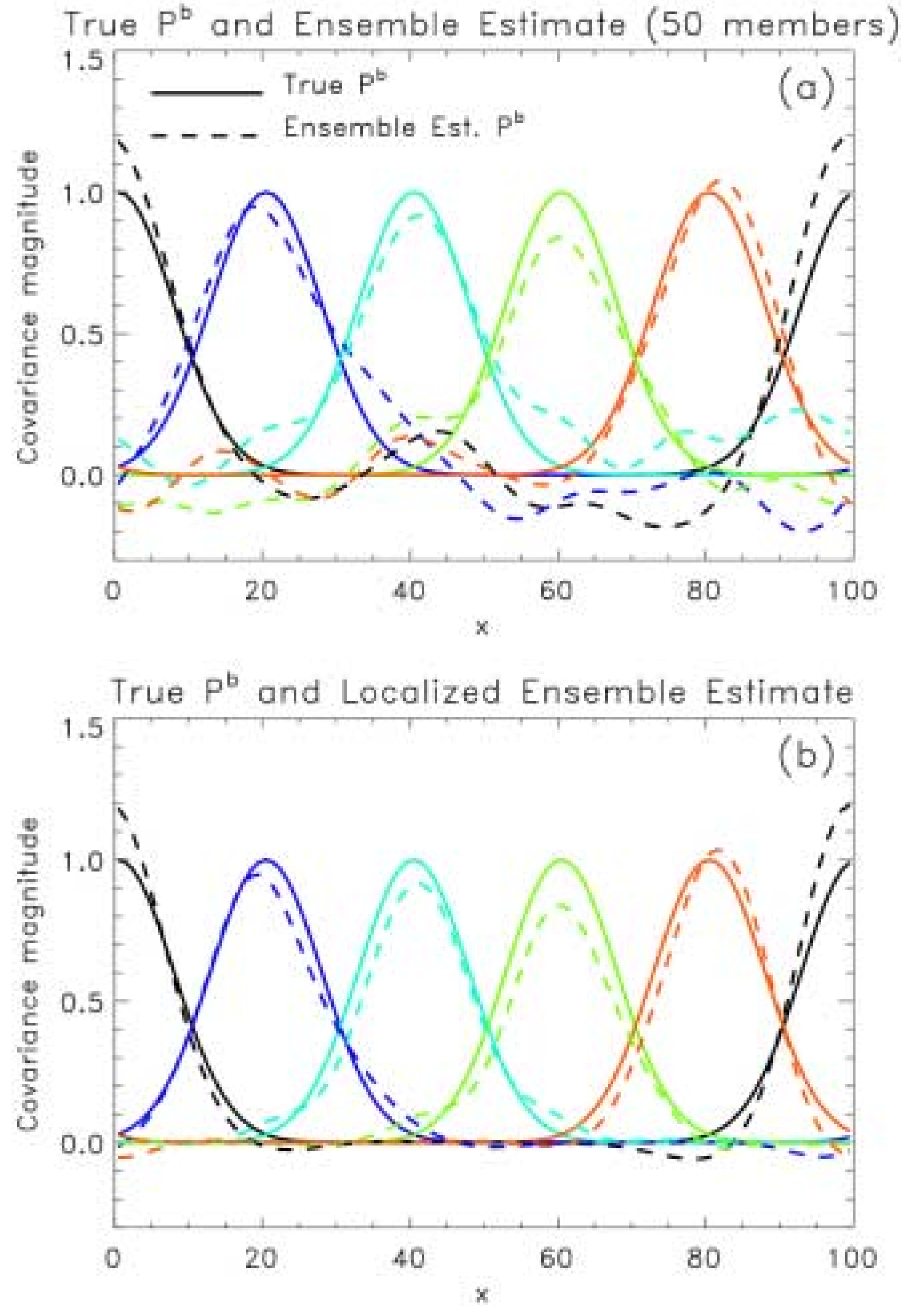


(Analysis variable is sea-level pressure)

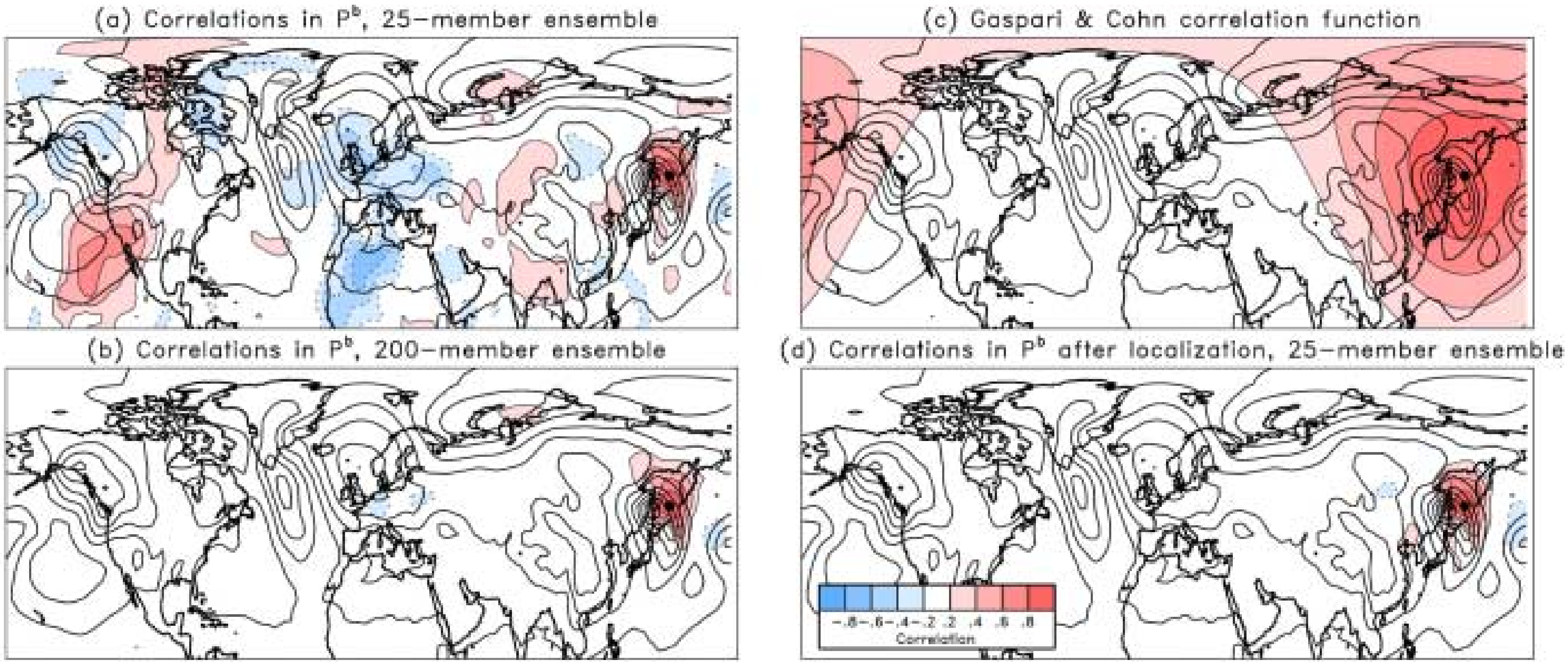
A solution to noisy covariances: “covariance localization”



Another localization example



Covariance model before / after localization



(Analysis variable is sea-level pressure)

Question: how many ensemble members needed?

- Depends …
  - How many directions in model phase space where error grows?
  - Covariance “localization” can ameliorate problems with too small an ensemble
- Generally, experiments with 25 to a few hundred members have shown success in a variety of models.

Covariance localization and size of the ensemble

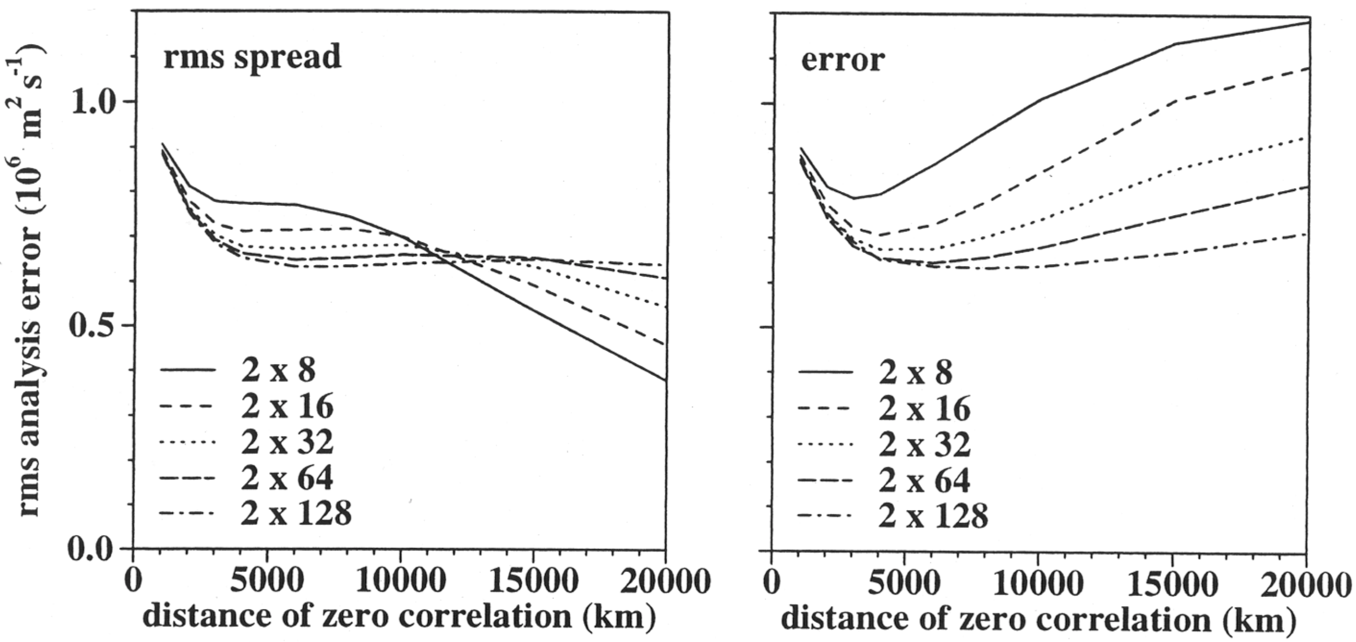
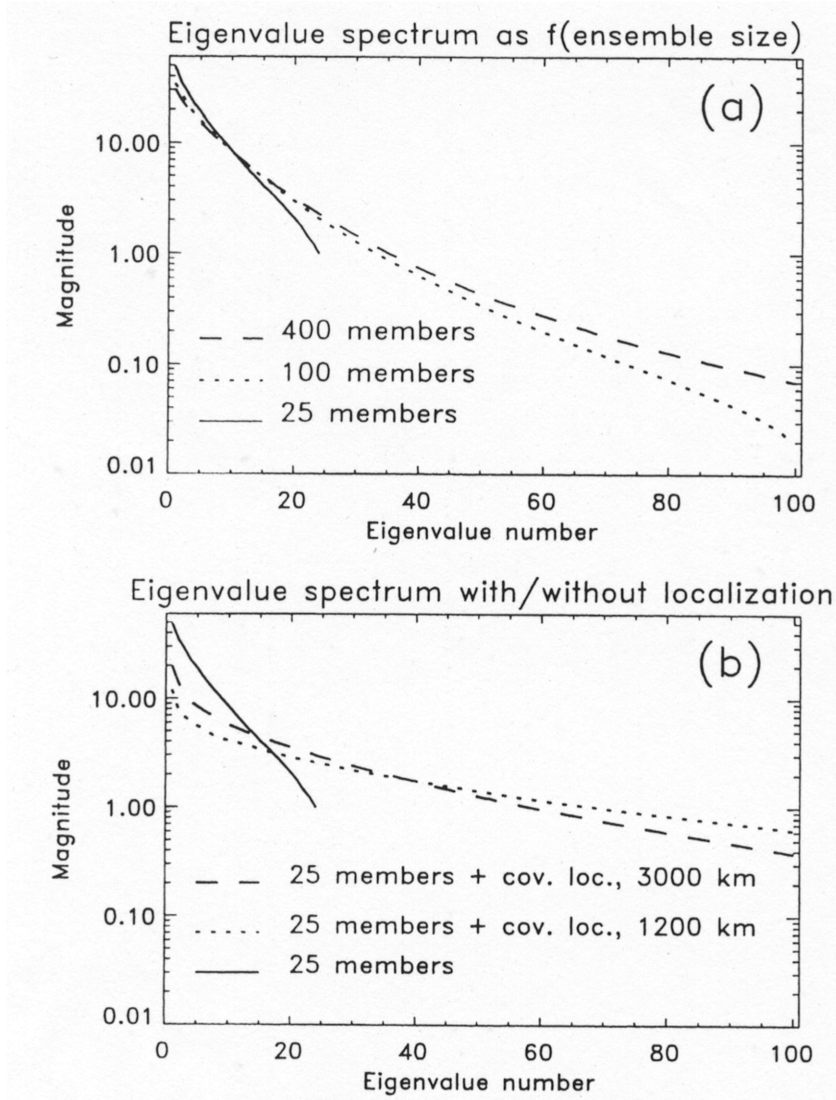


FIG. 4. Analysis error as a function of r_1 (km), the distance beyond which ρ is zero, for various ensemble sizes. (left) The rms spread in the ensemble, and (right) the rms error of the ensemble mean.

From Houtekamer & Mitchell, MWR, Jan. 2001

Smaller ensembles require tighter localization function

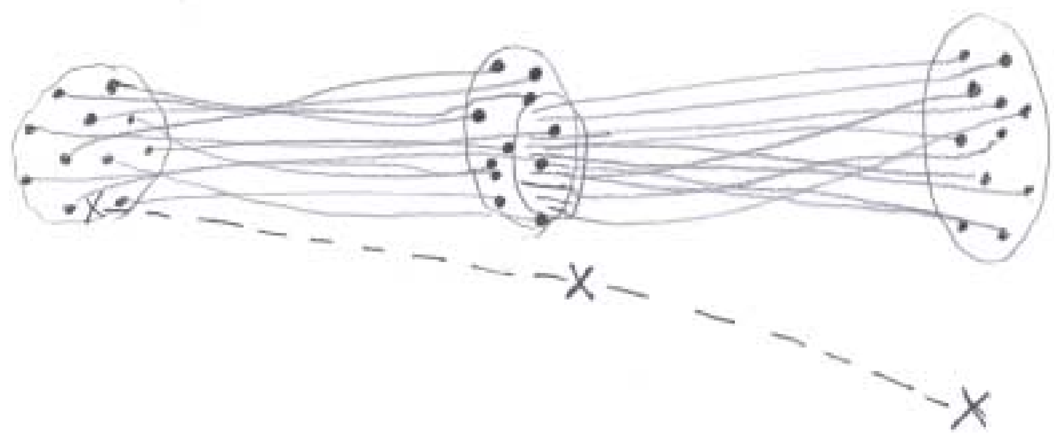
How does covariance localization make up for larger ensemble?



From Hamill et al., MWR, Nov 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. One major cause: **model error.**



Filter divergence causes

- *Too small an ensemble*.
  - Insufficient variance in certain directions 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
- *Other errors (e.g., mis-specified observation errors)*

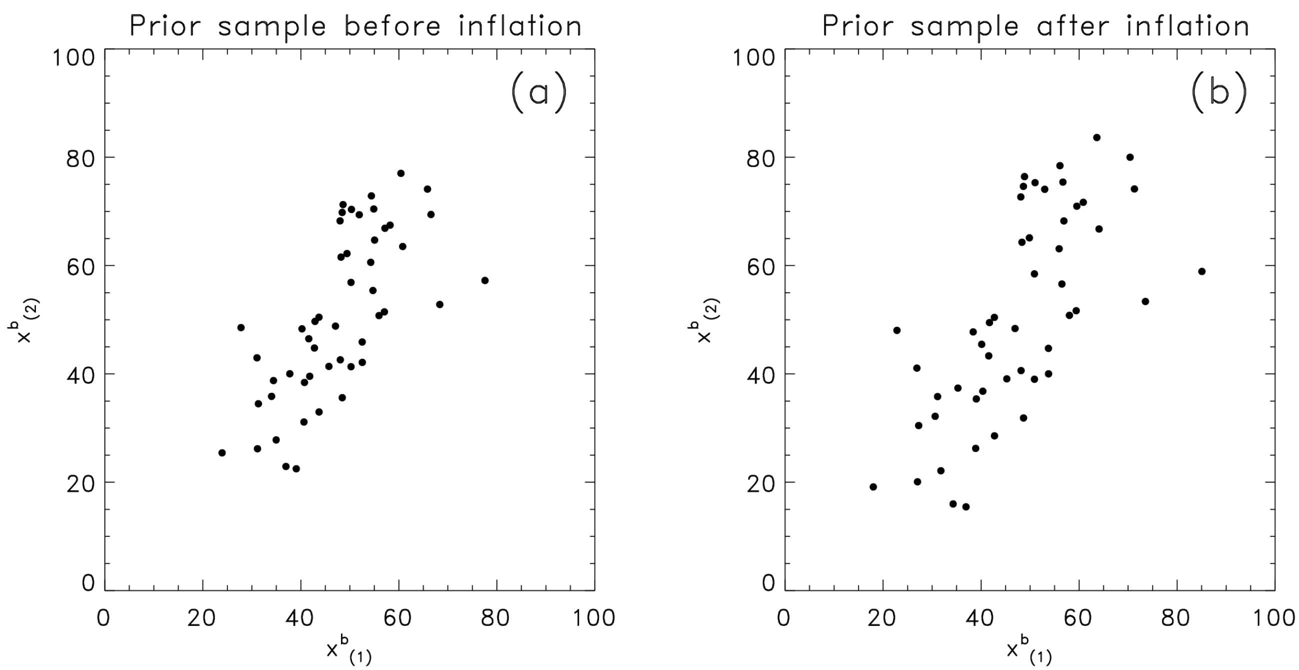
Possible filter divergence remedies

- Parameterizing model error.
  - Covariance inflation
  - Integrating stochastic noise
  - Adding simulated model error noise at data assimilation time.
  - Multi-model ensembles?
- Covariance localization (discussed before)
- Better specification of $\mathbf{R}$

Remedy 1: covariance inflation

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

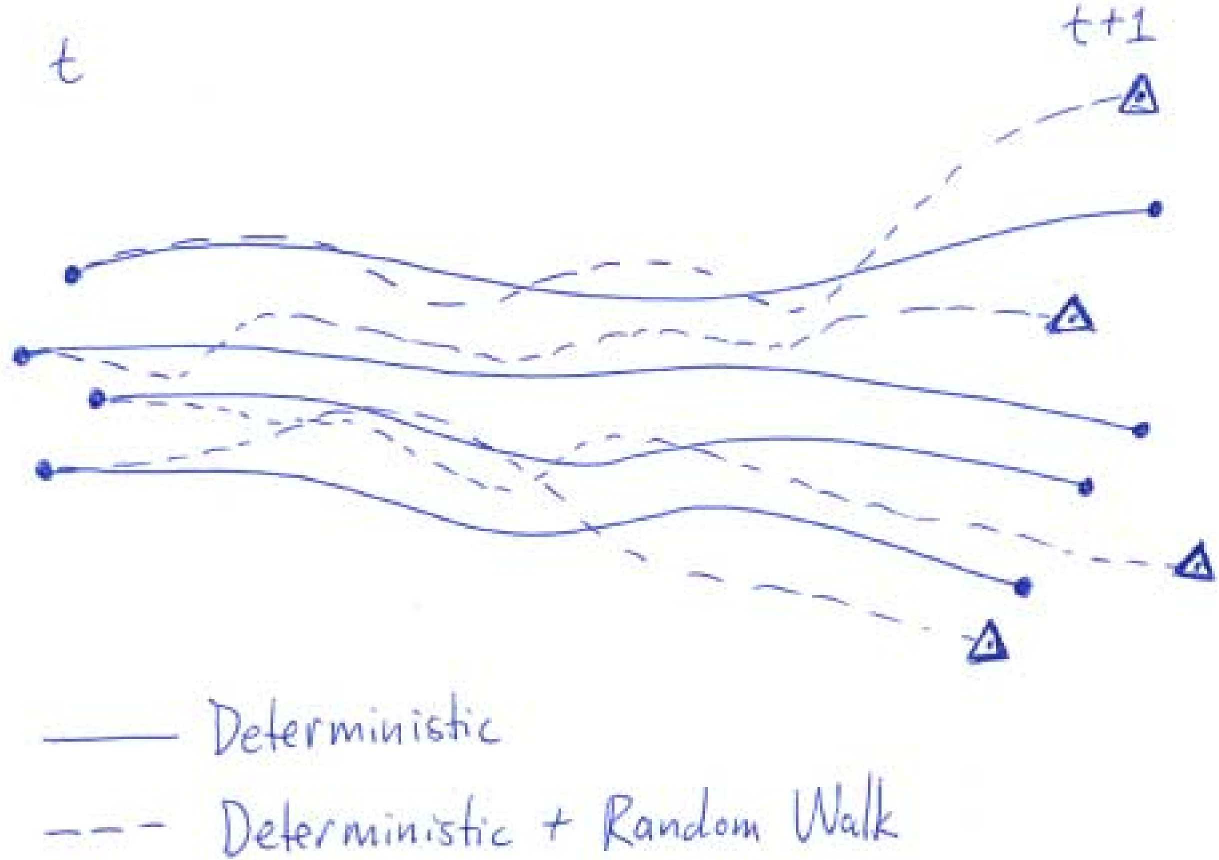
r is inflation factor



Disadvantage: model error in different subspace?

Remedy 2: Integrating stochastic noise

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



Remedy 2: Integrating stochastic noise

- Questions upon 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, QJ, 1999).

Remedy 3: adding noise at data assimilation time

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

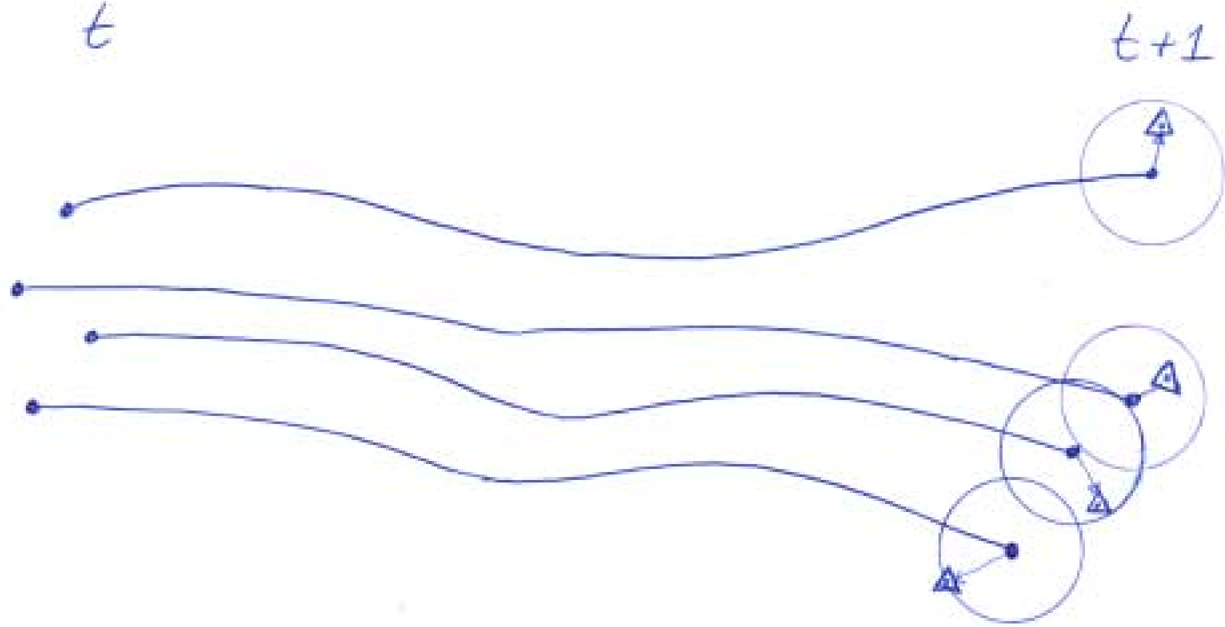
$$\mathbf{x}_{t+1} = M\mathbf{x}_t + \eta_t \qquad \left\langle \eta_t \eta_t^{T} \right\rangle = \mathbf{Q} \qquad \left\langle \eta_t \right\rangle = 0$$

$$\mathbf{x}^b_{t+1} = M\mathbf{x}^a_t \qquad \left\langle \left(\mathbf{x}^b_{t+1} - \overline{\mathbf{x}^b_{t+1}}\right)\left(\mathbf{x}^b_{t+1} - \overline{\mathbf{x}^b_{t+1}}\right)^T \right\rangle \approx \mathbf{M}\mathbf{P}^a\mathbf{M}^T$$

$$\textit{define ens. of } x^b_{t+1} \textit{ such that } \left\langle \left(x^b_{t+1} - \overline{x^b_{t+1}}\right)\left(x^b_{t+1} - \overline{x^b_{t+1}}\right)^T \right\rangle = \mathbf{M}\mathbf{P}^a\mathbf{M}^T + \mathbf{Q}$$

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

Remedy 3: adding noise at data assimilation time



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

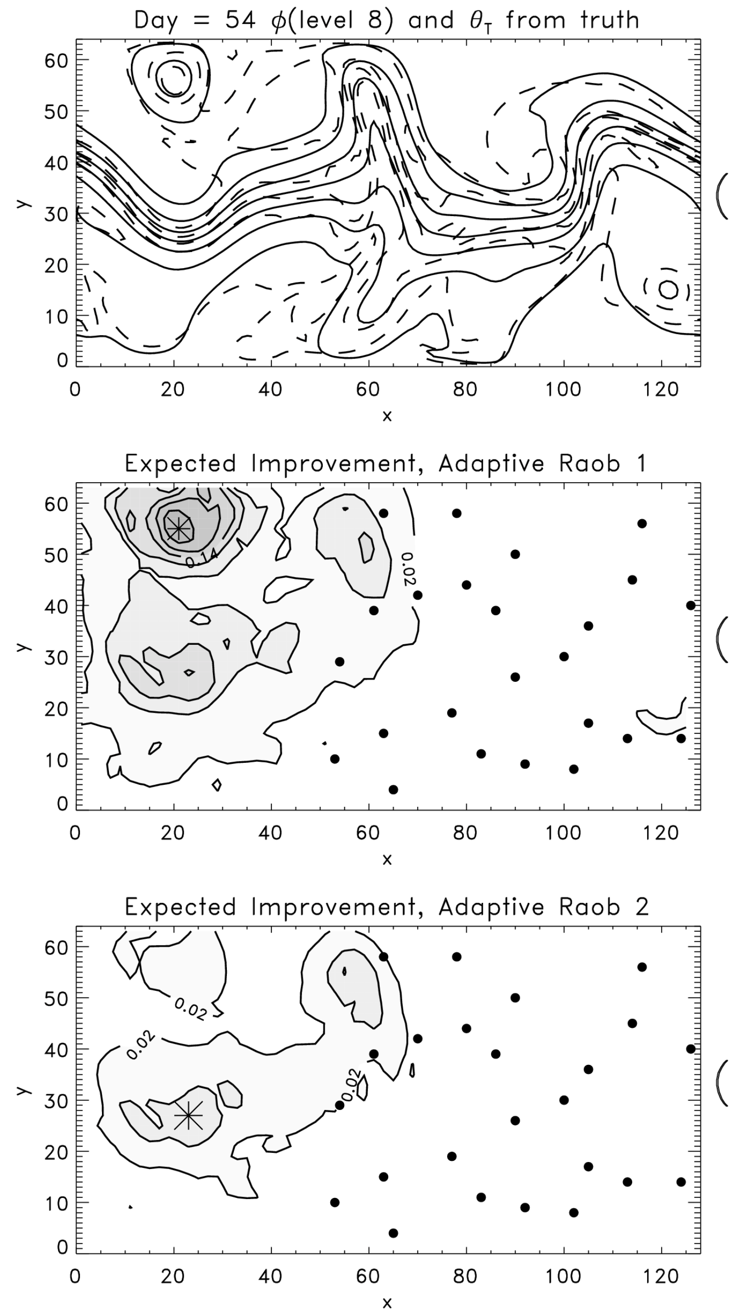
Remedy 3: adding noise at data assimilation time (cont'd)

- Knowing statistics of **Q** crucial
- Mitchell and Houtekamer: estimate parameters of **Q** from data assimilation innovation statistics.
- DIS: only simple model of **Q** can be fitted.

Remedy 4: 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).

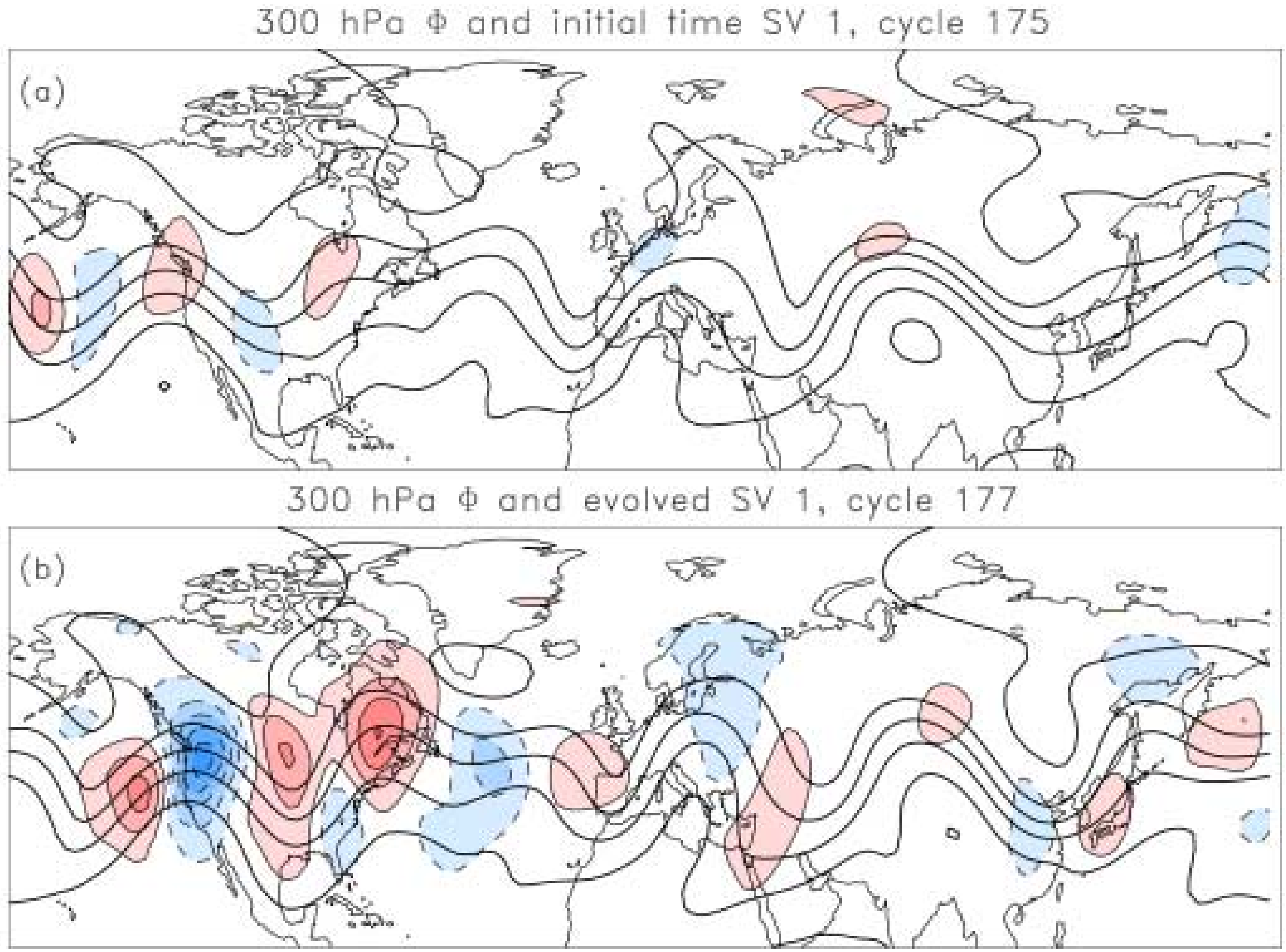
Applications of ensemble filter: adaptive observations



Can predict the expected reduction in analysis or forecast error from inserting a rawinsonde at a particular location.

From Hamill and Snyder, MWR, June 2002

Applications of ensemble filter: analysis-error covariance singular vectors



From Hamill et al., MWR, August 2003

4D-Var vs. ensemble assimilation

- Produce the same mean state IF
  - Forecast model perfect
  - Linear evolution of forecast perturbations
  - Gaussian statistics
  - 4D-Var cycles background-error covariances
  - Ensemble size infinite
  - *H* operators approximately linear

4D-Var vs. ensemble assimilation

- 4D-Var ADV:
  - Established
  - More efficient with complex *H*?
  - Handles nonlinear *H*
- 4D-Var DIS:
  - Adjoint / TLM coding
  - Linearity assumptions
  - No automatically generated ensemble
  - Tough to treat model error

- Ensemble ADV:
  - Model error treatable
  - Cycles covariances
  - Automatically generates ensemble
- Ensemble DIS:
  - Expensive for complex *H*
  - Nonlinear *H* problematic?
  - Unproven in real NWP models.
  - Finnecky; must get **R**, **Q** right

Kalman filter vs. ensemble assimilation

- Accuracy tradeoff: sampling errors (ensemble) versus falsely assuming linearity (Kalman filter).

- No adjoint / tangent linear code required for ensemble assimilation.

- Whichever way, both permit treatment of model error.

Conclusions

- Ensemble data assimilation literature growing rapidly because of
  - Great results in simple models
  - Coding ease
- Unites data assimilation and ensemble forecasting
- Keep an eye on the literature, or join in the exploration.
- www.cdc.noaa.gov/~hamill/efda_review4.pdf