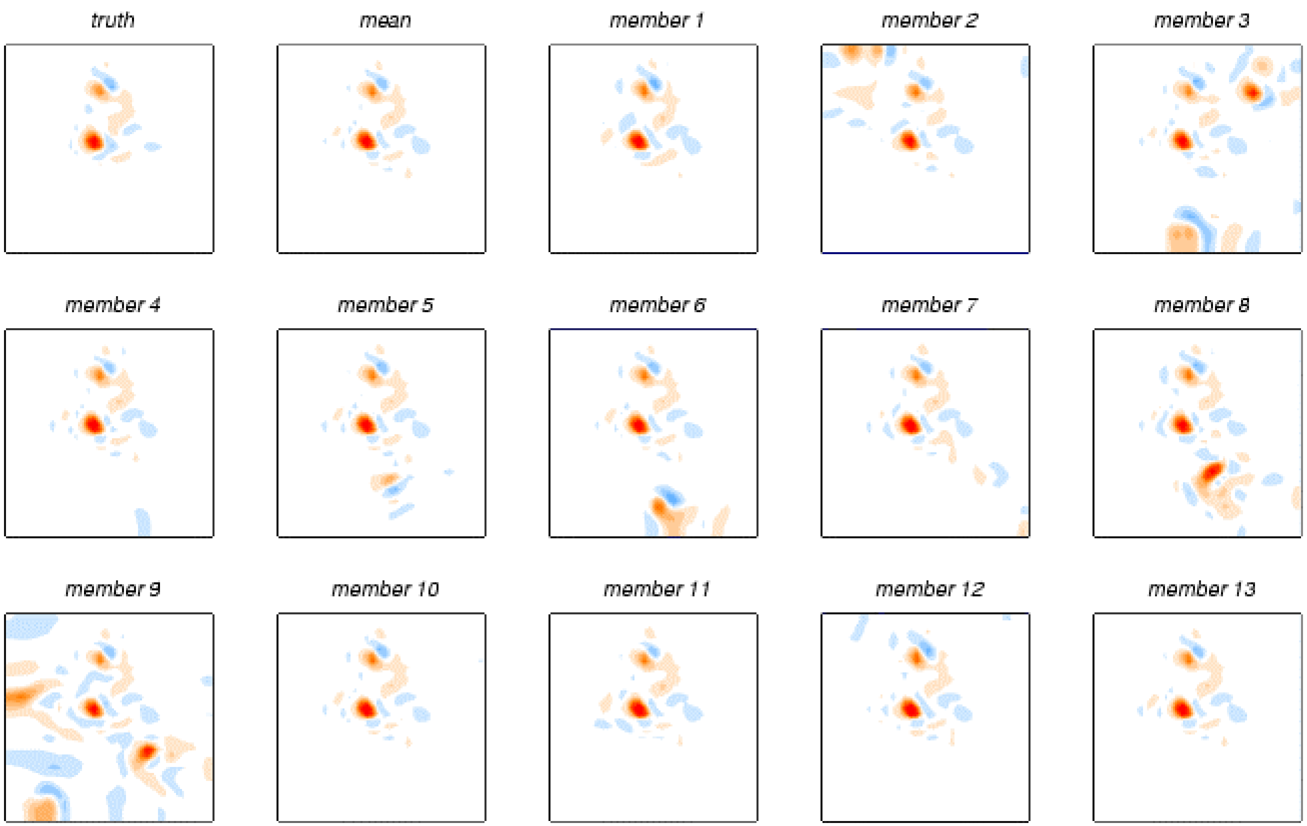
Introduction to Ensemble Filters



- a Bayesian view of data assimilation
- overview of ensemble filters

Data Assimilation via Bayes Rule

True state is unknown

- observations, models both have random errors
- probability density of x^t, $[x^t]$, is most that can be determined

Example: two obs with known error distributions

- $y_i = x^t + \epsilon_i$, with $[\epsilon_i]$ known

Wish to calculate $[x^t | y_1 = y_1^o, y_2 = y_2^o]$

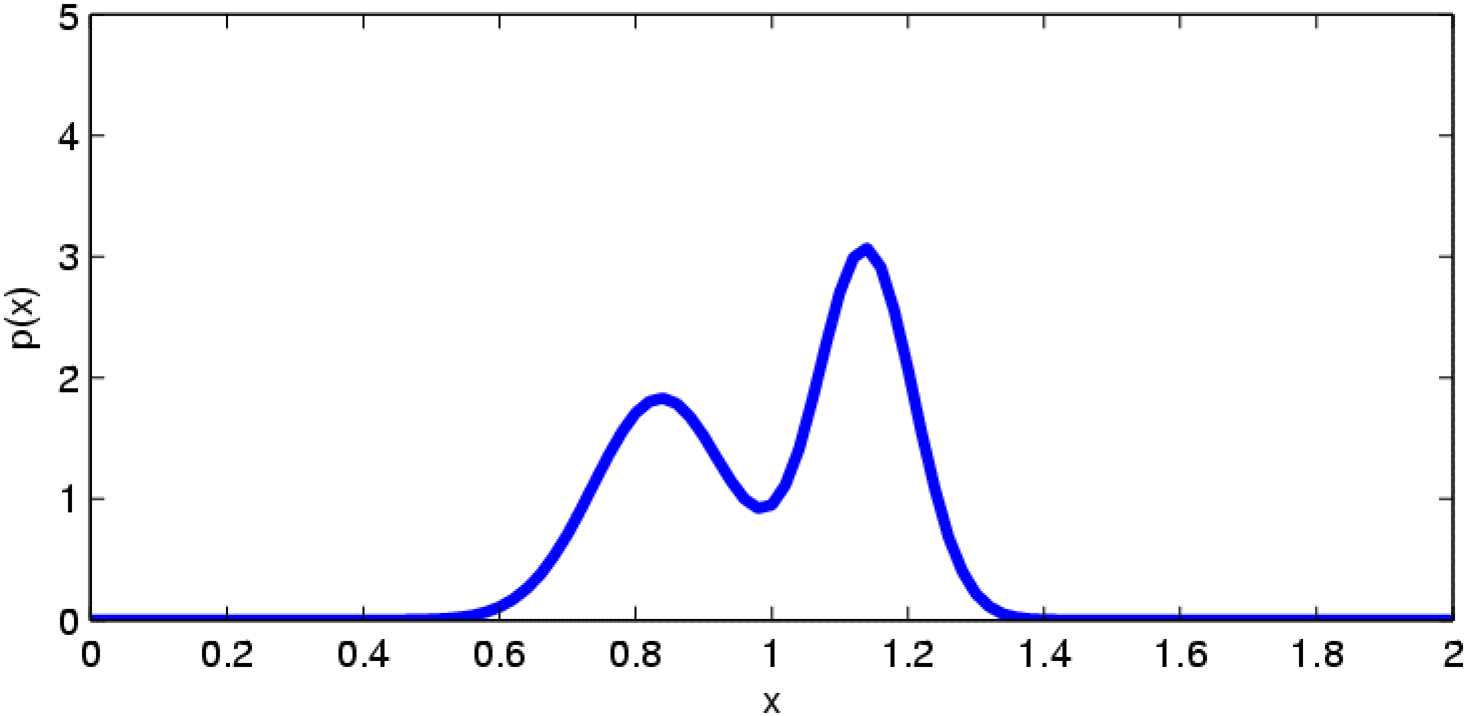
- Bayes rule

$$[x^t | y_1, y_2] \propto [y_2 | x^t, y_1][x^t | y_1]$$

- if obs errors independent, $[y_2 | x^t, y_1] = [y_2 | x^t]$

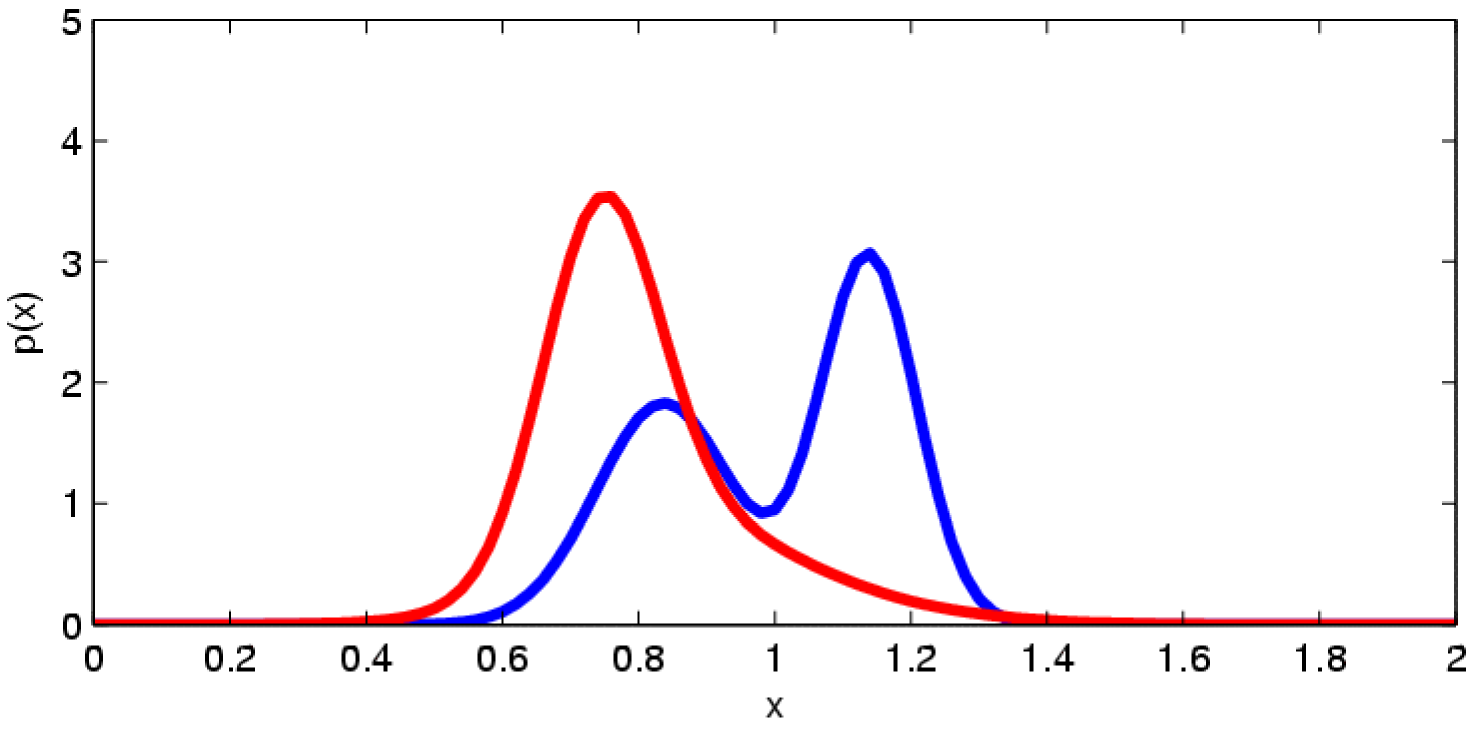
Bayes Illustrated

- $[x^t|y_1]$ for $y_1^o = 1.1$ (blue)



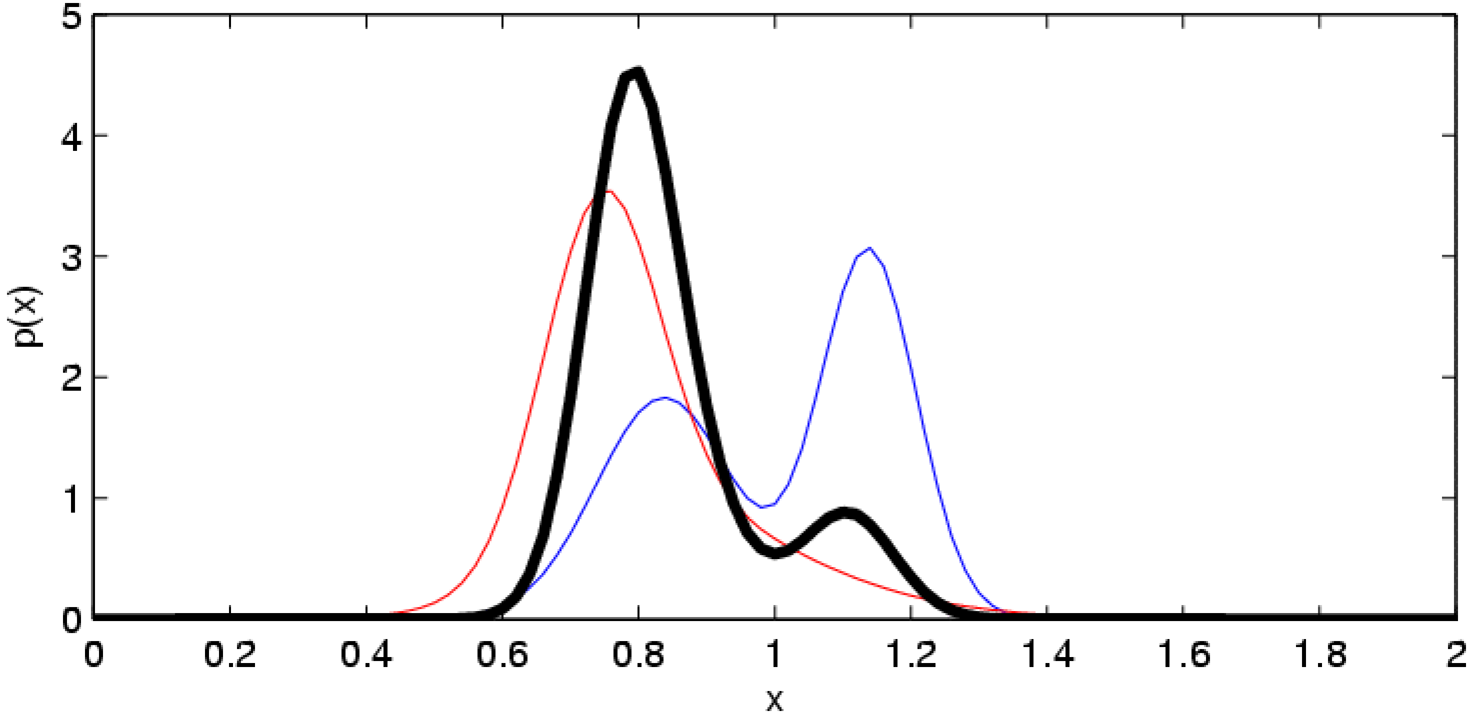
Bayes Illustrated (cont.)

- $[x^t|y_1]$ for $y_1^o = 1.1$ (blue)
- $[y_2|x^t]$ for $y_2^o = 0.75$ (red)



Bayes Illustrated (cont.)

- $[x^t|y_1]$ for $y_1^o = 1.1$ (blue)
- $[y_2|x^t]$ for $y_2^o = 0.75$ (red)
- $[x^t|y_1, y_2] \propto [y_2|x^t][x^t|y_1]$ (black)



Many variables

Notation

- $\mathbf{x} =$ state w.r.t. discrete basis, e.g. grid-point values or Fourier coefficients
- wish to estimate $\mathbf{x}^t$, true state *projected* onto discrete basis

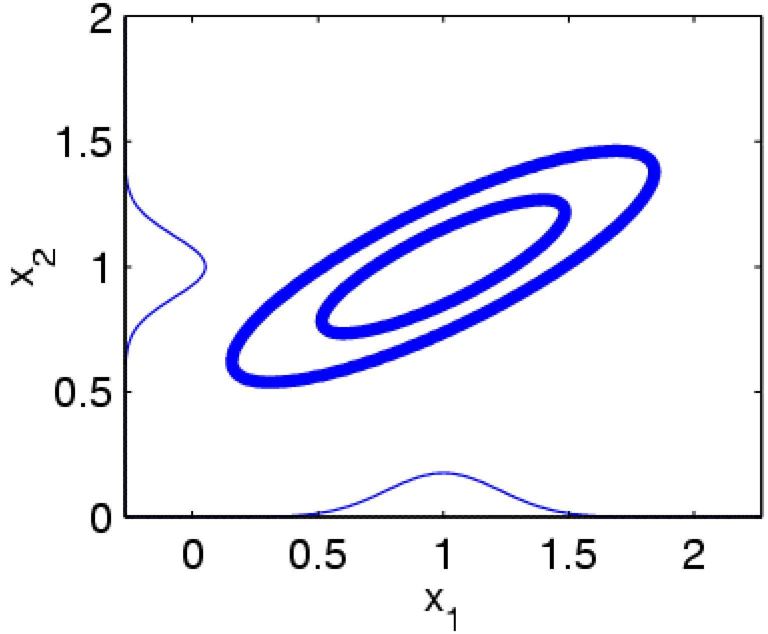
Available information, typically

- new observations
- recent forecast of $\mathbf{x}^t$; for NWP, often as accurate as obs
- approximate balance relations, such as geostrophy

Need covariance matrices for forecast ($\mathbf{P}^f$) and obs ($\mathbf{R}$)

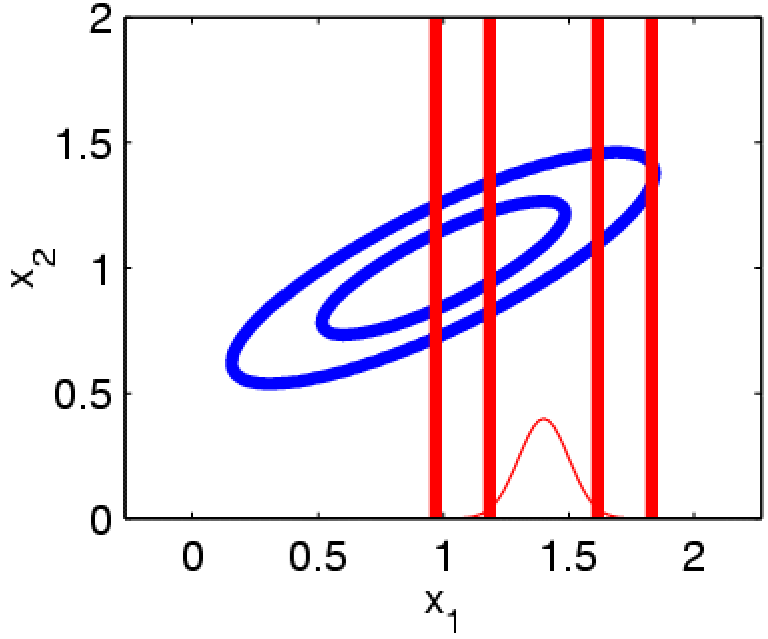
Importance of Forecast Covariances

- 2D: $\mathbf{x} = (x_1, x_2)$
- forecast



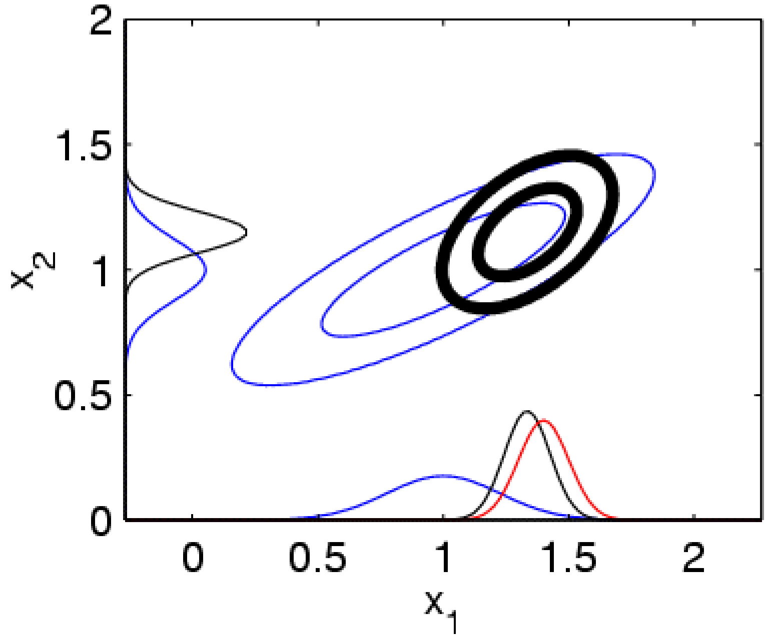
Importance of Forecast Covariances

- observation, $y = x_1 + \text{noise} = 1.4$
- observation likelihood independent of x_2



Importance of Forecast Covariances

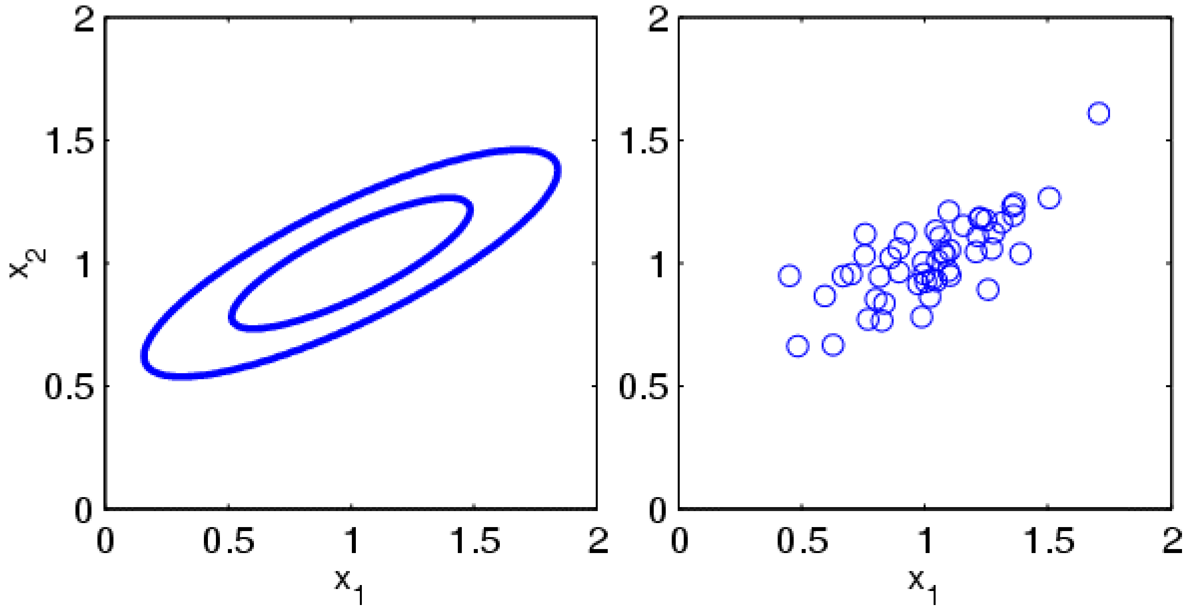
- analysis
- $\mathrm{Cov}(x_1, x_2)$ provides information on unobserved variable



EnKF Sketch

Monte-Carlo approach (samples vs. distributions)

- estimate covariances from ensemble of forecasts



Given ensemble at $t = t_k$,

- forecast each member to t_{k+1}, time of next observations
- update each member at t_{k+1}
- continue as above, from $t = t_{k+1}$

Ensemble Size

Doesn't ensemble need to be huge?

- $N_x \geq O(10^6) \gg N_e \leq O(10^3)$
- $\widehat{\mathbf{P}}^f$ extremely rank deficient
- but, errors only $O(N_e^{-1/2})$ in each element.

Dynamics reduce effective dimensionality

- variance grows in some directions, decays in most
- $\mathbf{P}^f$ has a few large eigenvalues

Covariance localization

- typically, distant pts have small covariance, so hard to estimate
- allow each observation to influence analysis only locally
- increases rank of $\widehat{\mathbf{P}}^f$ and decreases detrimental effects of sampling errors, yet decreases computational cost

Attractions of the EnKF

1. Covariances incorporate information from dynamics

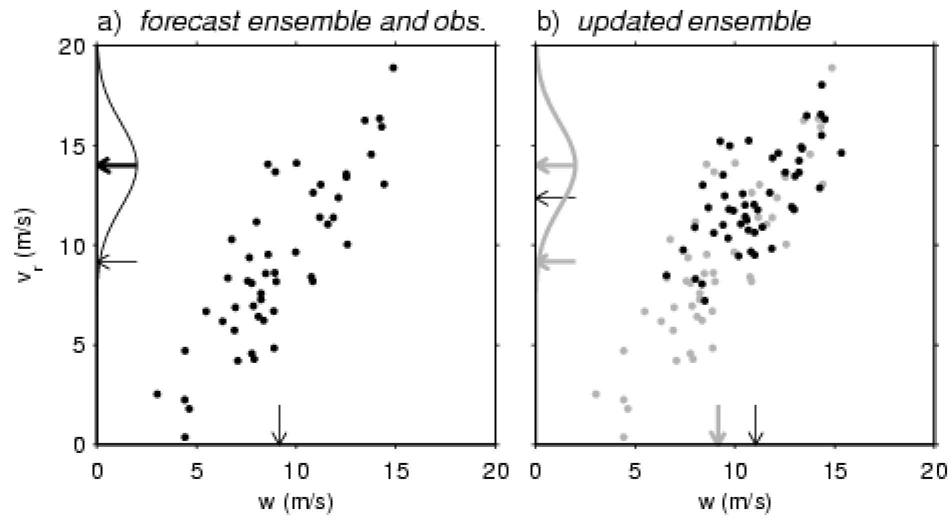
2. Ease of implementation
   - simple algorithm, relatively independent of model
   - no linearized or adjoint models
   - forecasts, at least, are highly parallel

3. Provides estimates of f/c and analysis uncertainty
   - natural foundation for ensemble forecasting system

4. Parameter estimation is convenient
   - include models parameters in extended state vector

Update Step: Schematic Radar Example

Update w given an observation of v_r



- let $v_r^f = \mathbf{Hx}^f$; compute for each member
- For each grid point i, estimate from ensemble

$$c_i = (\mathbf{P}^f \mathbf{H}^T)_i = \mathrm{Cov}(w_i^f, v_r^f), \quad d = \mathbf{HP}^f\mathbf{H}^T + \mathbf{R} = \mathrm{Var}(v_r^f) + \mathbf{R}$$

- update each member at ith grid point,

$$w_i^a = w_i^f + (\hat{c}_i/\hat{d})(v_r - v_r^f + \epsilon) \quad \epsilon \sim N(0, \mathbf{R})$$