State estimation using the ensemble Kalman filter, and its application for numerical weather prediction

Tom Hamill

NOAA Earth System Research Lab

Boulder, Colorado USA 80305

tom.hamill@noaa.gov

1

This will be a talk about the intersection of

"data assimilation"

and

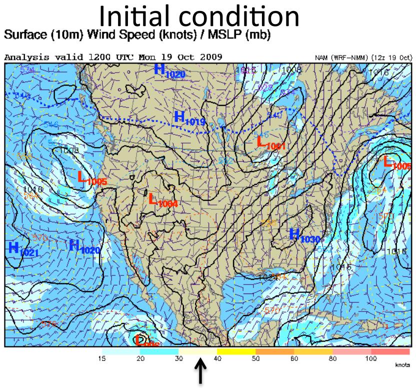
"ensemble prediction"

2

Data Assimilation:

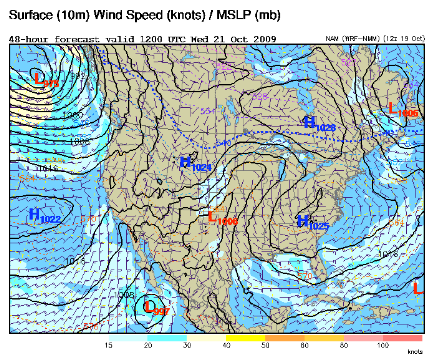
How we produce the initial condition(s) for weather forecasts

Initial condition



Weather forecast model

48-hour forecast



The initial conditions include the temperature, humidity, wind velocity, cloud cover, pressure, etc. on a regular grid at many vertical levels. Also: ocean temperatures, land temperatures, even soil moisture.

3

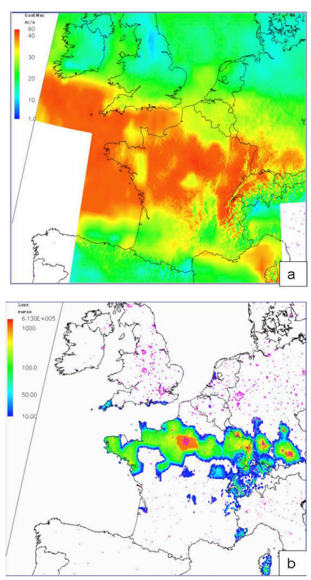


Figure 6. The modeled maximum wind-speed footprint (a), and the corresponding distribution of total industry-wide loss (b), for Lothar 12Z 25 - 00Z 27 December 1999.

from Keller et al. AMS Conf. preprint

Motivation for ensemble prediction: 1999 storm "Lothar"

Black Forest damage

from Neil Stringfellow (ETH) presentation

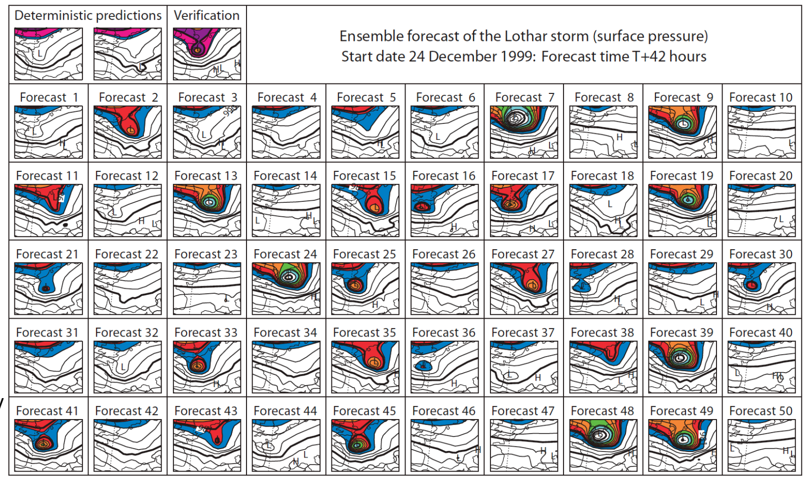
4

“Ensemble prediction” or “ensemble forecasting”

Multiple simulations of the weather from slightly different initial conditions, perhaps different forecast models

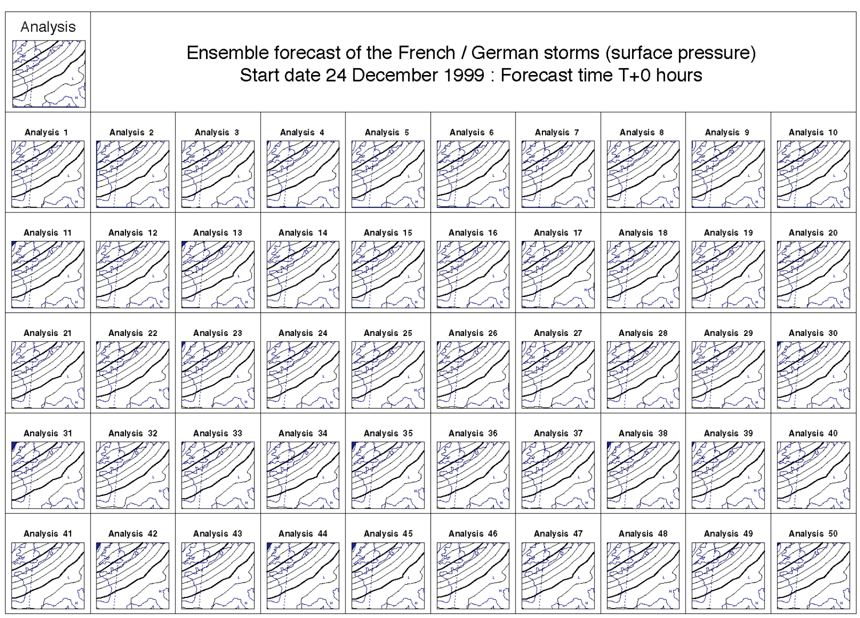
deterministic forecast totally misses damaging storm over France; some ensemble members forecast it well.

Probabilities commonly estimated from frequency of event in the ensemble.



from Tim Palmer’s book chapter, 2006.

Initial conditions for Lothar ensemble forecasts

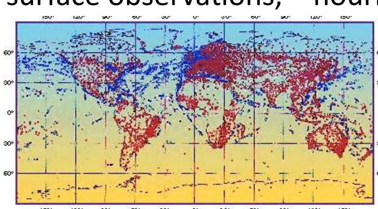


Data assimilation, from first principles

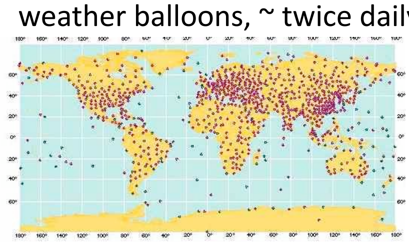
7

What “observations” (measurements) are available to estimate the state?

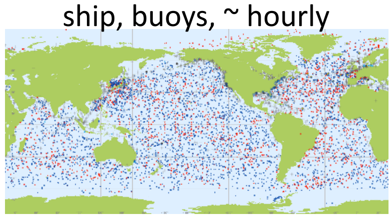
surface observations, ~ hourly



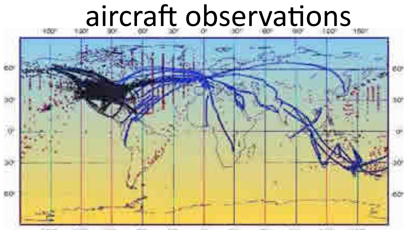
weather balloons, ~ twice daily



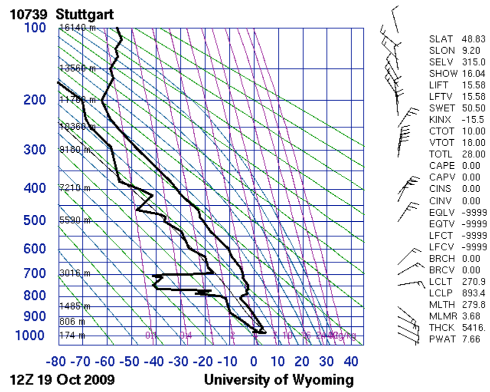
ship, buoys, ~ hourly



aircraft observations

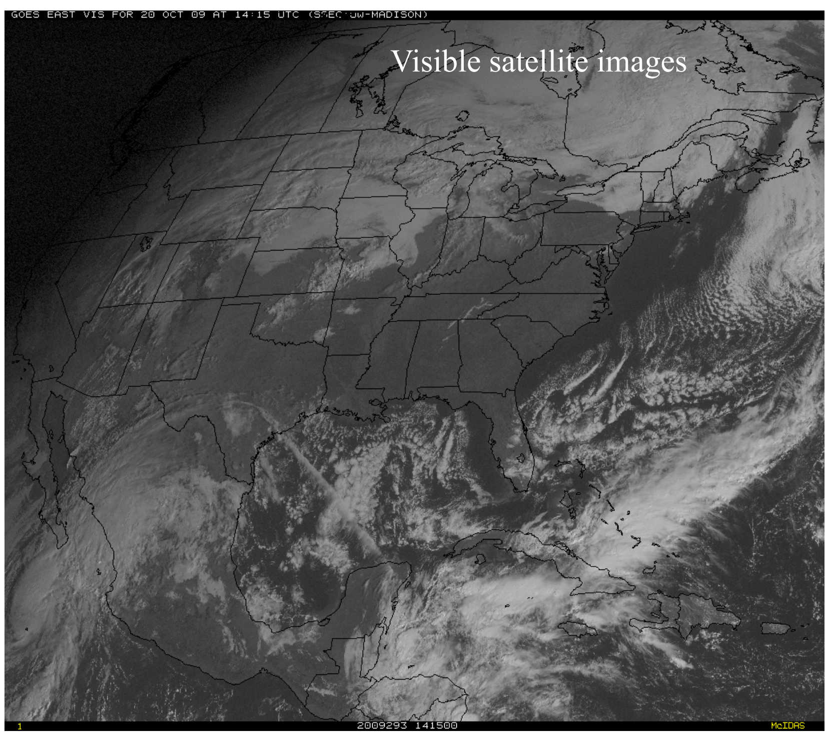


Weather balloons, i.e., rawinsondes

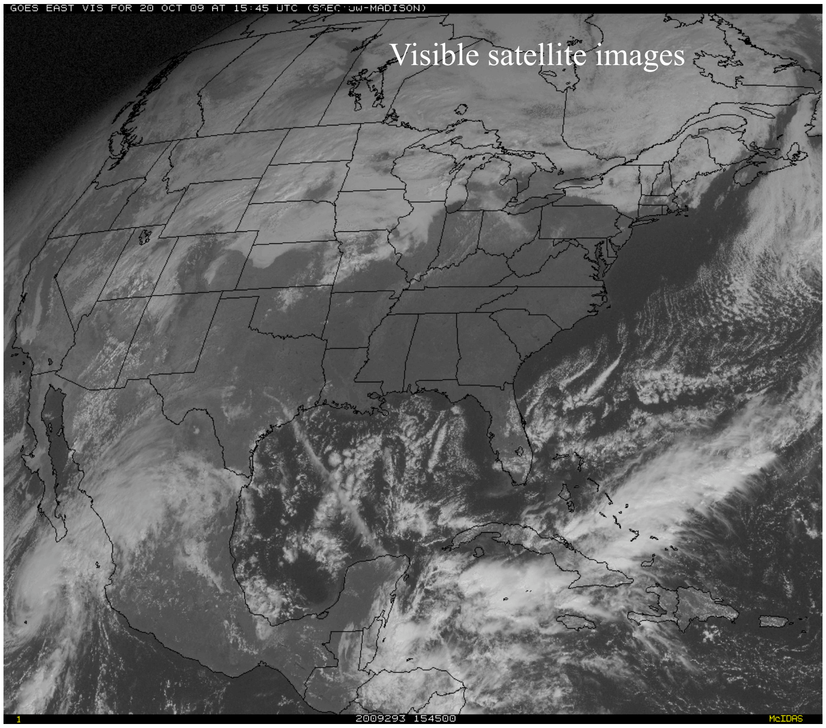


They're a very useful source of information, but they're not available uniformly around the world.

9

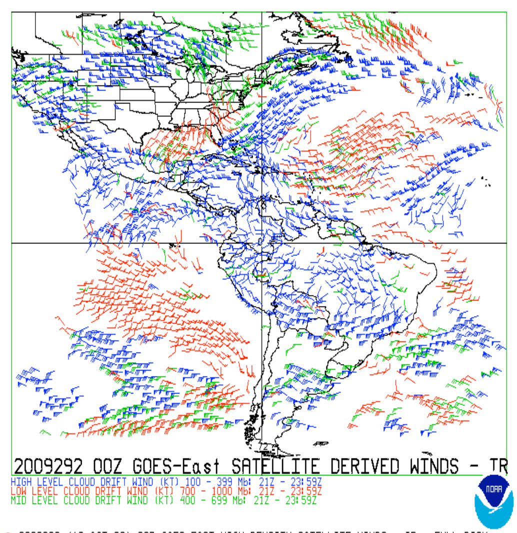


10



11

By watching the movement in successive images, one can estimate cloud-top wind motion



Where there are no clouds, no wind vectors can be estimated.

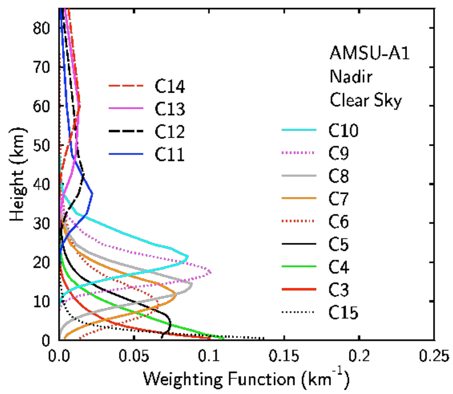
Winds can only be determined for one level at a given location.

12

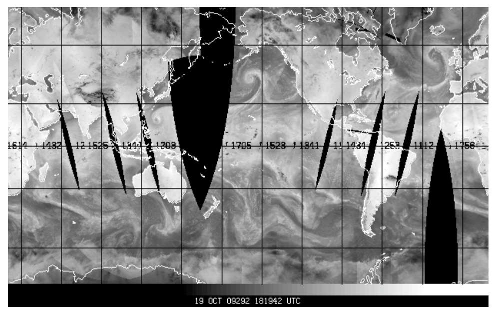
Passive “sounder” instruments on satellites

AMSU-A microwave sounder on US NOAA polar-orbiting satellite

Each channel’s approximate weight for the temperature at a given height

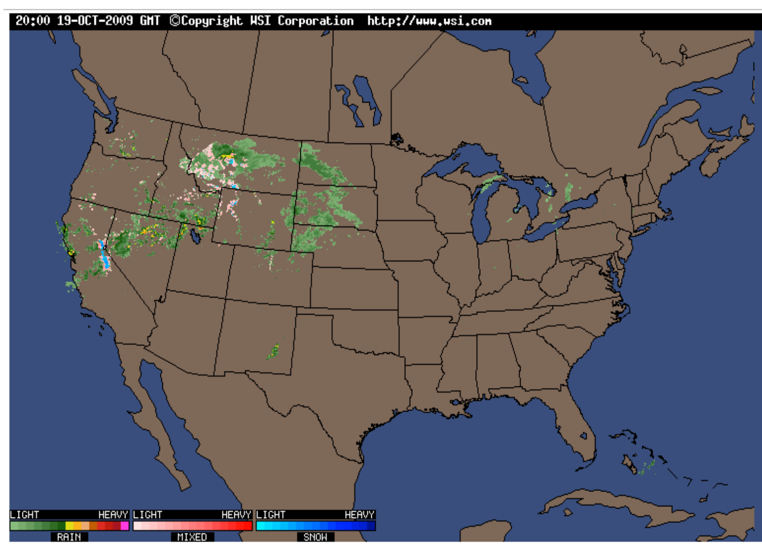


Channel 15 radiance



13

Radars



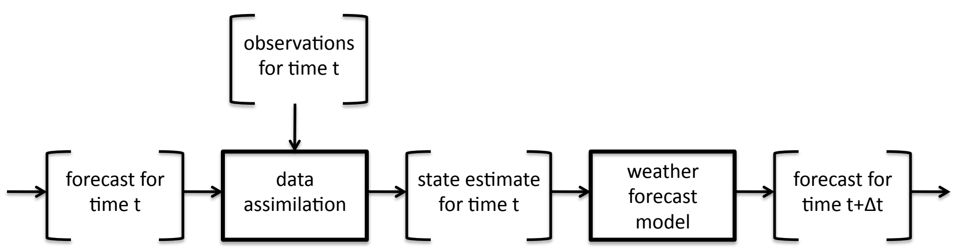
14

Information available to estimate the atmospheric state at time *t*

- Observations (measurements)
  - Many over US, Europe, China
  - Fewer over the oceans
  - May measure something else (radiance at a particular frequency) than what we need to initialize a model forecast.
- A prior forecast of the state at time *t*, initialized at $t - \Delta t$
  - important for defining the state where there are few or no observations.

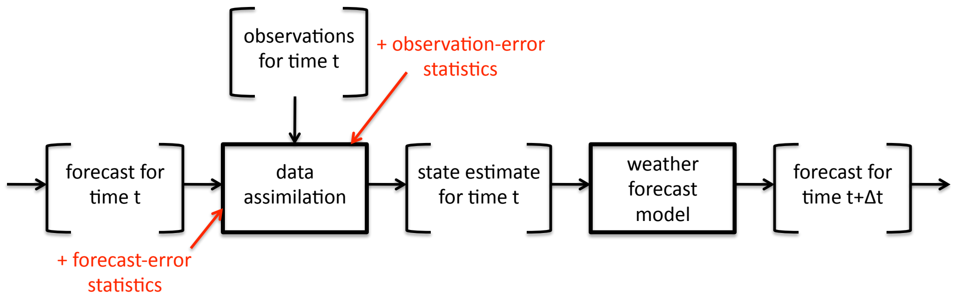
15

State estimation ("data assimilation")

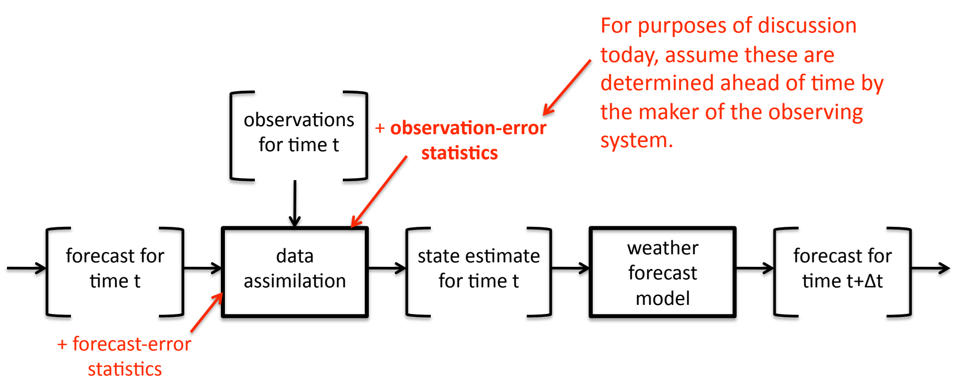


16

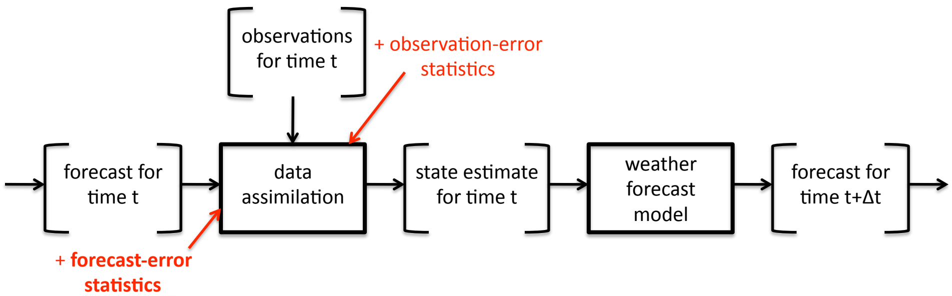
State estimation ("data assimilation")



State estimation ("data assimilation")



State estimation ("data assimilation")



Let's spend some time to understand the statistics behind data assimilation and then the forecasts and the characteristics of forecast-error statistics.

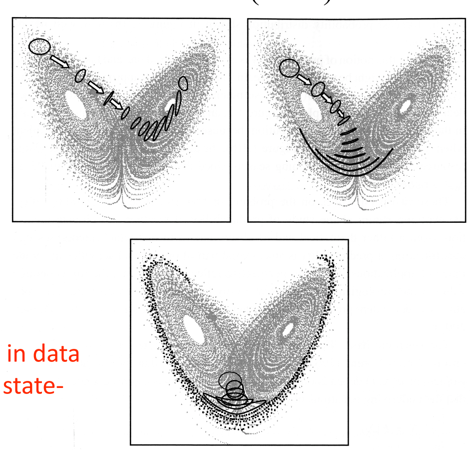
19

Forecast-errors statistics may be very different from one day to the next because of "chaos"

$$\frac{dx}{dt} = \sigma(y - x)$$
$$\frac{dy}{dt} = x(\rho - z) - y$$
$$\frac{dz}{dt} = xy - \beta z$$

σ, ρ, β are fixed.

The Lorenz (1963) model



Errors grow more quickly for some initial conditions than others. Hence, in data assimilation, would be useful to have state-dependent forecast error statistics

from Tim Palmer's 2006 book chapter

From first principles: Bayesian data assimilation

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

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

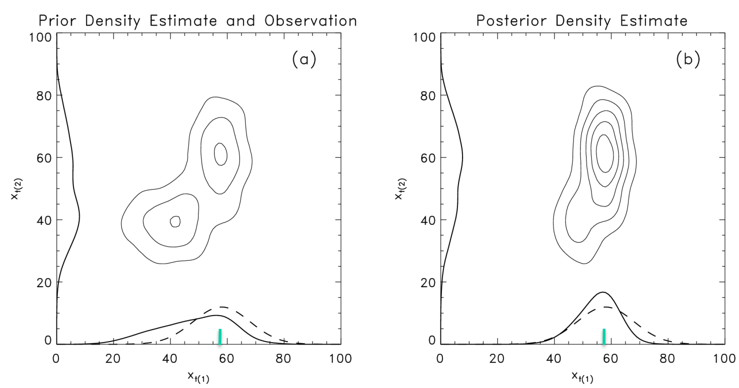
$$P(\mathbf{x}_t \mid \psi_t) \propto P(\mathbf{y}_t \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \psi_{t-1})$$

"posterior" "prior"

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

[derivation]

Bayesian data assimilation: 2-D example



Computationally expensive when highly dimensional! Here, probabilities explicitly updated on 100x100 grid; costs multiply geometrically with the number of dimensions of model state. Also: "curse of dimensionality"

Data assimilation terminology

- **y** : Observation vector (weather balloons, satellite radiances, etc.)
- $\mathbf{x}^b$: Background state vector ("prior")
- $\mathbf{x}^a$: Analysis state vector ("posterior")
- **H** : (hopefully linear) operator to convert model state → observation location & type
- **R** : Observation - error covariance matrix
- $\mathbf{P}^b$: Background - error covariance matrix
- $\mathbf{P}^a$: Analysis - error covariance matrix

23

Simplifying Bayesian data assimilation: toward the Kalman Filter

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

(manipulation of Baye's rule)

Assume

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

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

assumed Gaussian

so

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

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

24

Kalman filter update equations

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

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

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

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.

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

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

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

$\mathbf{P}^{\mathrm{a}}$ is the “analysis-error covariance. The Kalman filter indicates not only the most likely state but also quantifies the uncertainty in the analysis state.

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

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

$$\mathbf{x}^{\mathrm{b}}_{t+1} = \mathbf{Mx}^{\mathrm{a}}$$

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

25

Kalman filter update equation: example in 1-D, H=identity operator

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

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

26

Other roles of $\mathbf{P}^b$, the background-error covariance matrix

$$\mathbf{P}^b = \begin{bmatrix} \sigma^2(x_1^b) & Cov(x_1^b, x_2) & . & . & . & Cov(x_1^b, x_n^b) \\ Cov(x_1^b, x_2) & \sigma^2(x_2^b) & . & . & . & Cov(x_2^b, x_n^b) \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ Cov(x_1^b, x_n^b) & Cov(x_2^b, x_n^b) & . & . & . & \sigma^2(x_n^b) \end{bmatrix}$$

x_1 is the model state at grid point 1, x_2 the state at grid point 2, and so on. *n* dimensions to the model state.

Say x_1 is the surface temperature at Heidelberg, x_2 is the surface temperature at Hamburg. $Cov(x_1, x_2)$ indicates how the forecast errors are related between the two locations. If the forecast is too cold at Heidelberg, will it tend to be too cold at Hamburg, too?

27

Other roles of $\mathbf{P}^b$, the background-error covariance matrix

$$\mathbf{P}^b = \begin{bmatrix} \sigma^2(x_1^b) & Cov(x_1^b, x_2) & . & . & . & Cov(x_1^b, x_n^b) \\ Cov(x_1^b, x_2) & \sigma^2(x_2^b) & . & . & . & Cov(x_2^b, x_n^b) \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ Cov(x_1^b, x_n^b) & Cov(x_2^b, x_n^b) & . & . & . & \sigma^2(x_n^b) \end{bmatrix}$$

x_1 is the model state at grid point 1, x_2 the state at grid point 2, and so on. *n* dimensions to the model state.

Say x_1 is the surface temperature at Heidelberg, x_2 is the surface east-west wind velocity at Heidelberg. $Cov(x_1, x_2)$ indicates how the temperature error may be related to the wind error. Perhaps when it's cooler at night, it tends to be less windy. These "cross-covariances" very important in atmospheric data assimilation.

28

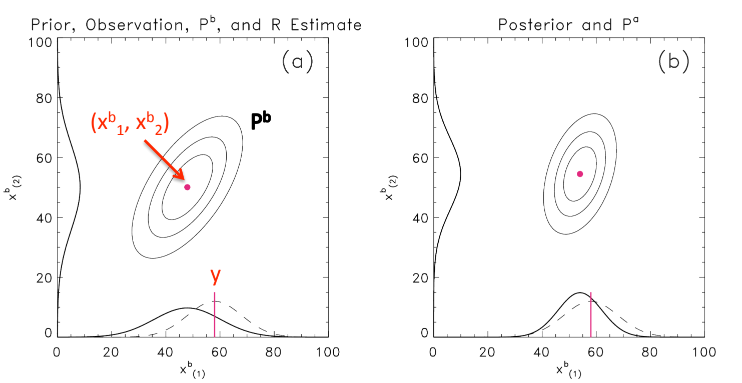
Other roles of $\mathbf{P}^b$, the background-error covariance matrix

$$\mathbf{P}^b = \begin{bmatrix} \sigma^2(x_1^b) & Cov(x_1^b,x_2) & \cdots & Cov(x_1^b,x_n^b) \\ Cov(x_1^b,x_2) & \sigma^2(x_2^b) & \cdots & Cov(x_2^b,x_n^b) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(x_1^b,x_n^b) & Cov(x_2^b,x_n^b) & \cdots & \sigma^2(x_n^b) \end{bmatrix}$$

x_1 is the model state at grid point 1, x_2 the state at grid point 2, and so on. *n* dimensions to the model state.

Another point: for numerical weather prediction this matrix is very large, ~ $10^7 \times 10^7$. Thinning and lower-rank approximations are necessary to make this computationally affordable.

Kalman filter update, 2-dimensional example



$\mathbf{P}^b$ covariance matrix controls the amount x^b_2 is adjusted from the observation y; if Cov(x^b_1, x^b_2) = 0, no adjustment to x^b_2

Kalman filter update equations

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

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 . $\mathbf{x}_{t+1} = \mathbf{M}\mathbf{x}_{t} + \eta_{t}, \quad \left\langle \eta_{t}\eta_{t}^{T}\right\rangle = \mathbf{Q}$

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

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

31

Kalman filter limitations when applied to atmospheric data assimilation

- Future forecast model states are estimated with linear operator. Atmosphere sometimes behaves very non-linearly.
- Covariance propagation step $\mathbf{P}_{t}^{a} \rightarrow \mathbf{P}^{b}_{t+1}$ very computationally expensive.
- Still need tangent linear (**M**) & adjoint model (**M**^{T}) for evolving covariances, and linear error growth assumption questionable.

32

Can we improve on the Kalman filter?

- Utilize a non-linear forecast model
- Improve computational efficiency
- Remove some restrictive linear & Gaussian assumptions.

33

From the Kalman filter to the "ensemble" Kalman filter (EnKF)

- 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.

34

Canonical EnKF update equations (for time *t*)

$$\mathbf{x_i^a} = \mathbf{x^b_i} + \mathbf{K}\left(\mathbf{y_i} - H\mathbf{x^b_i}\right)$$

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

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

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

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

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

$$\mathbf{X} = \left(\mathbf{x_1^b} - \overline{\mathbf{x^b}}, \ldots, \mathbf{x_n^b} - \overline{\mathbf{x^b}}\right)$$

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

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

35

Propagation of state and error covariances in EnKF

$$\mathbf{P}^a(t) = \left\langle \left[\mathbf{x}_i^a(t) - \overline{\mathbf{x}}_i^a(t)\right]\left[\mathbf{x}_i^a(t) - \overline{\mathbf{x}}_i^a(t)\right]^T \right\rangle$$

(**P**^{a} never explicitly formed)

$$\mathbf{x}_i^b(t+1) = M\mathbf{x}_i^a(t)$$

if forecast model is “perfect”

- or -

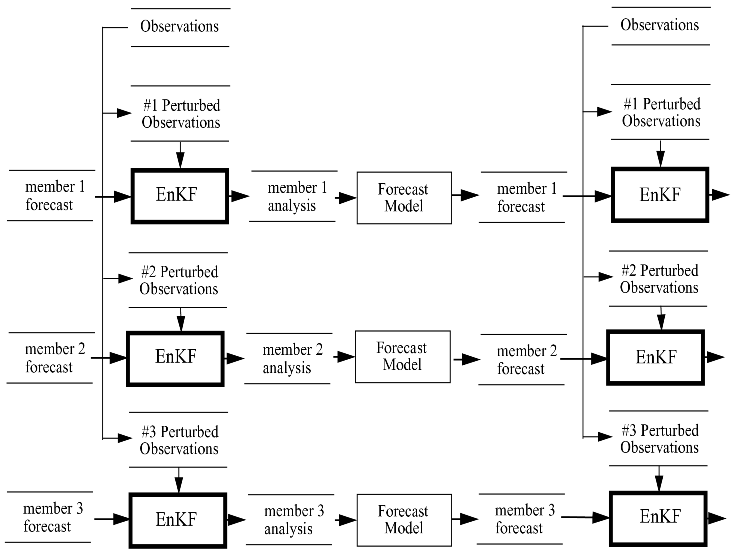
$$\mathbf{x}_i^b(t+1) = M\mathbf{x}_i^a(t) + \eta_i$$

$$\left\langle \eta_i \eta_i^T \right\rangle = \mathbf{Q}$$

if forecast model has model error.

36

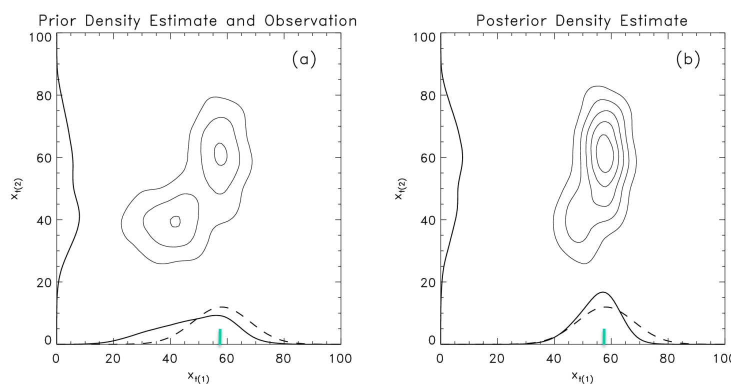
The ensemble Kalman filter: a schematic



(This schematic is a bit of an inappropriate simplification, for EnKF uses every member to estimate background-error covariances)

37

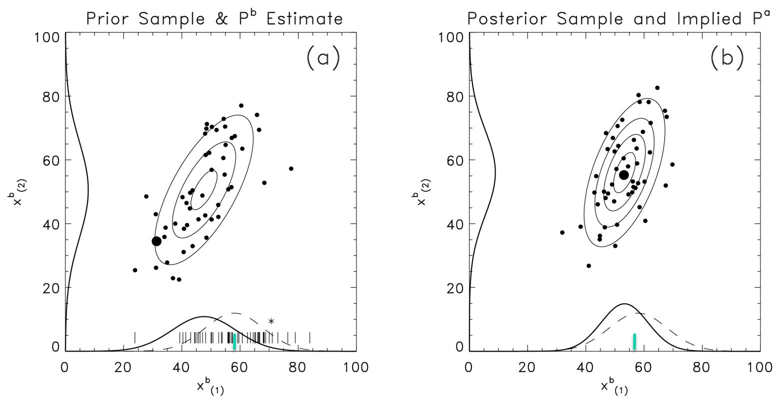
Bayesian data assimilation: 2-D example



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

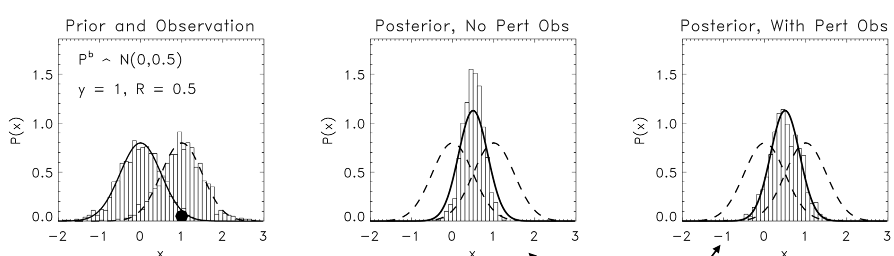
38

How the EnKF works: 2-D example



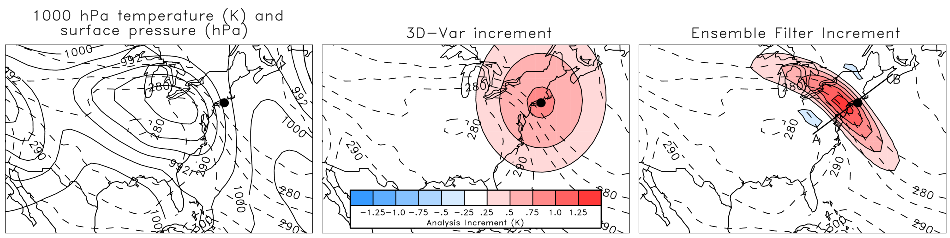
Start with a random sample from bimodal distribution used in previous Bayesian data assimilation example. Contours reflect the Gaussian distribution fitted to ensemble data.

Why perturb the observations?



histograms denote the ensemble values; heavy black line denotes theoretical expected analysis-error covariance

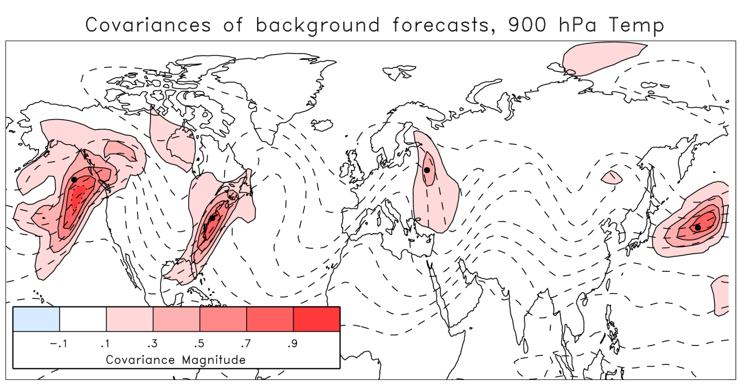
How the EnKF may achieve its improvement relative to previous methods: better background-error covariances



Output from a “single-observation” experiment. The EnKF is cycled for a long time. The cycle is interrupted and a single observation 1K greater than the mean prior 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 background at the observation location.

41

More examples of flow-dependent background-error covariances



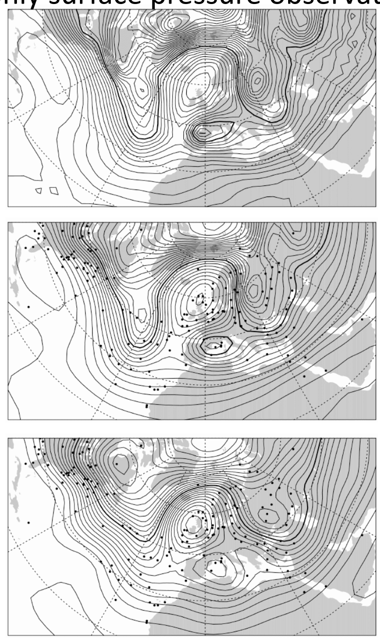
42

Example : 500 hPa height analyses (~5500 m elevation) assimilating only surface pressure observations

Full NCEP-NCAR Reanalysis (3D-Var) (120,000+ obs)

Ensemble Kalman Filter (214 surface pressure obs)

Older method, similar to 3D-Var (214 surface pressure obs)



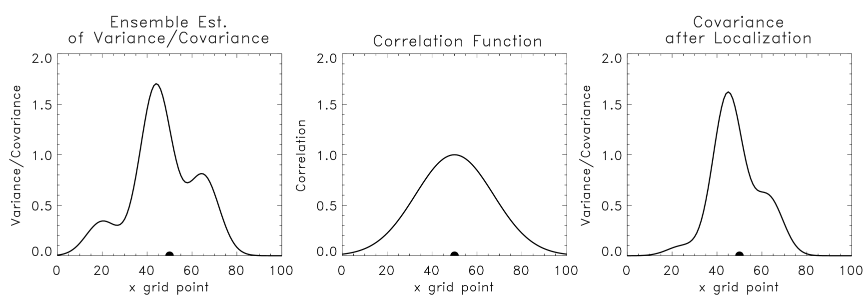
Black dots show surface pressure ob observation locations

RMS = 39.8 m

RMS = 82.4 m

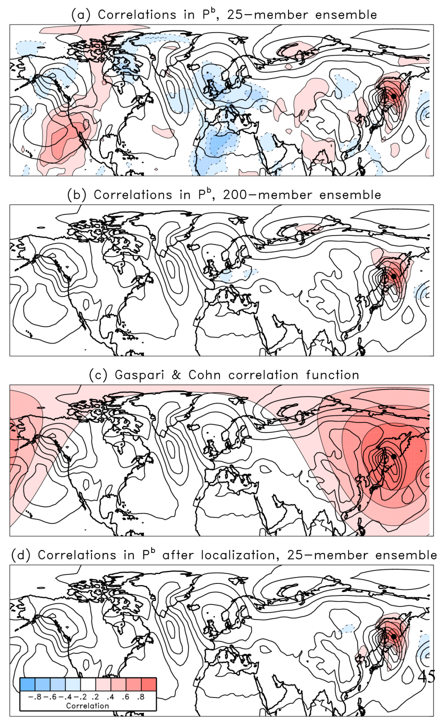
fig. from Jeff Whitaker

A modification to the basic EnKF: "covariance localization"



Estimates of covariances from a small ensemble will be noisy, with signal-to-noise small especially when covariance is small

Covariance localization in practice

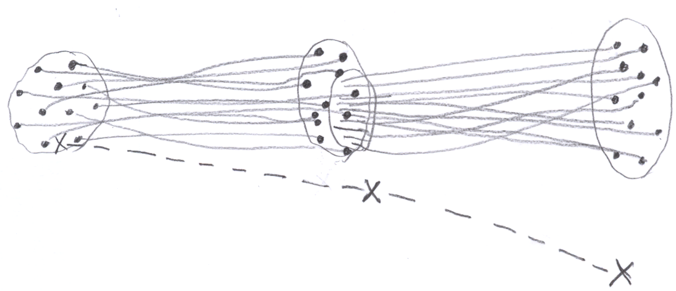


from Hamill review paper in "Predictability of Weather and Climate" (Cambridge Press), 2006

45

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.



46

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, so growth of differences between ensemble members happens too slowly.
  - Imperfections of how we treat "sub-gridscale" phenomena like thunderstorms.
  - Other model aspects unperturbed (e.g., land surface condition)
  - Others

47

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

48

Challenges

- How to deal with non-normally distributed errors.
  - "particle filters" sound promising but haven't found been demonstrated successfully with high-dimensional systems yet.
- Covariance localization and the problems it introduces (imbalances, difficult application to non-point observations).
- Reduction of (or an accurate quantification of) "model error."

49

Conclusions

- Ensemble Kalman filter used more and more because of
  - Great results in simple models and more recently, NWP models
  - Coding ease
  - Conceptual appeal (more gracefully handles some nonlinearity and treatment of model error)
- However:
  - Somewhat computationally expensive
  - Requires careful modeling of observation, forecast-error covariances to exploit benefits. Exploration of these issues relatively new.
  - Still many statistical challenges ahead.
- Acknowledgments: Jeff Whitaker, Chris Snyder, Jeff Anderson, Peter Houtekamer
- For more information, a review article:
  - www.esrl.noaa.gov/psd/people/tom.hamill/ensda_review.pdf

50

Bayesian data assimilation background

$$P(\mathbf{x}_t|\psi_t) \propto P(\psi_t|\mathbf{x}_t)P(\mathbf{x}_t)$$ Bayes rule

$$P(\psi_t|\mathbf{x}_t) = P(\mathbf{y}_t|\mathbf{x}_t)P(\psi_{t-1}|\mathbf{x}_t)$$ assuming independence of errors in time.

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

$$P(\psi_{t-1}|\mathbf{x}_t)P(\mathbf{x}_t) = P(\mathbf{x}_t|\psi_{t-1})$$ Bayes rule

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

[back]

51

Different implementations of ensemble filters

- Double EnKF (Houtekamer and Mitchell, *MWR*, March 1998)
- Ensemble adjustment filter (EnAF; 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 EnKF (Ott et al, U. Maryland; *Tellus*, in press)
- Others as well (Lermusiaux, Pham, Keppenne, Heemink, etc.)

52

Why different implementations?

- Save computational expense
- Problems with perturbed obs. Suppose

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

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

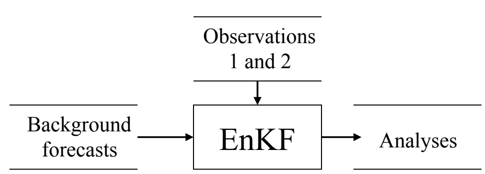
  Random sample $\mathbf{y}'$: [-1.04, 0.46, -0.12, -0.65, 2.10]

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

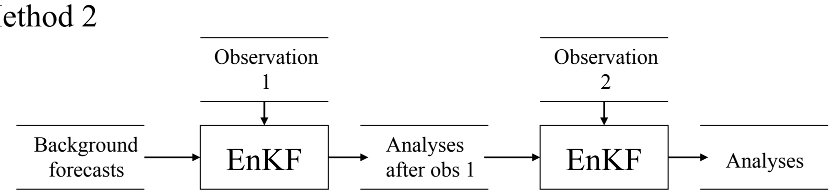
Noise added through perturbed observations can introduce spurious correlations between background, observation errors. However, there may be some advantages to perturbed observations in situations where the prior is highly nonlinear. See Lawson and Hansen, MWR, Aug 2004.

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

Method 1



Method 2



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

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

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

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

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

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

55

Covariance localization and size of the ensemble

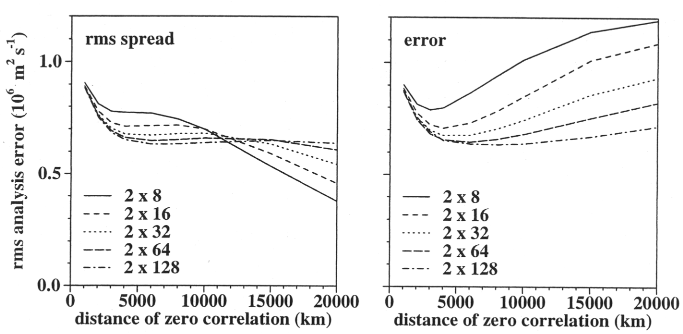
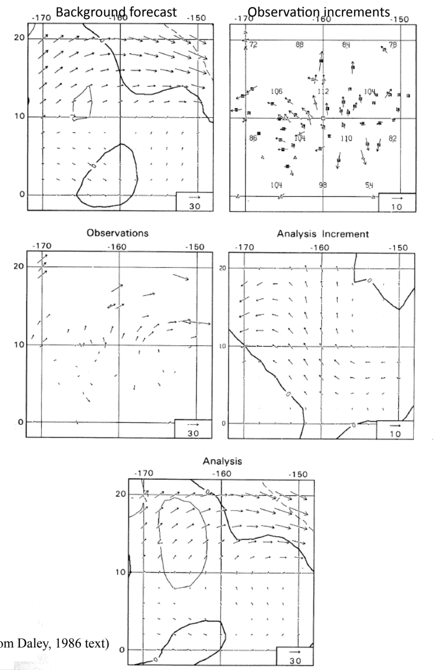


FIG. 4. Analysis error as a function of r_1 (km), the distance beyond which ρ is zero, for various ensemble sizes. (left) The rms spread in the ensemble, and (right) the rms error of the ensemble mean.

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

56

(from Houtekamer and Mitchell, MWR, Jan 2001)



(from Daley, 1986 text)

Figure 1.12 Illustration of data analysis of observation increments for the 200 mb wind field. (From Daley, *Mon. Wea. Rev.* **113**: 1066, 1985. The American Meteorological Society.)

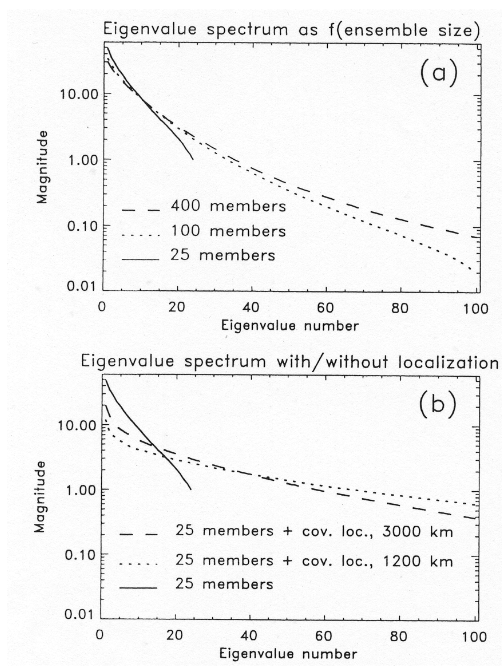
More on the process of data assimilation

The discrepancies between the observations and the 6-hour forecast are the "observation increment." The analysis increment is an estimate of the change to be applied to the 6-hour forecast, to form the analysis.

The statistics are hidden here, but:
If your forecast is very accurate
and your observations very inaccurate,
then the analysis increment will be small…and vice versa.

57

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)

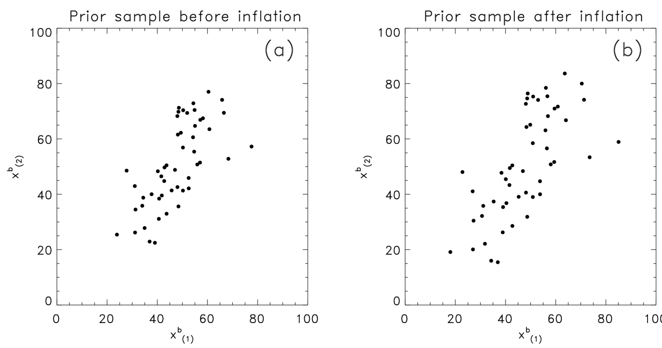
58

source: Hamill et al, MWR, Nov 2001

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



Disadvantage: what if model error in altogether different subspace?

59

source: Anderson and Anderson, MWR, Dec 1999

Remedy: integrating stochastic noise

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

60

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).

61

Remedy : adding noise at data assimilation time

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

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

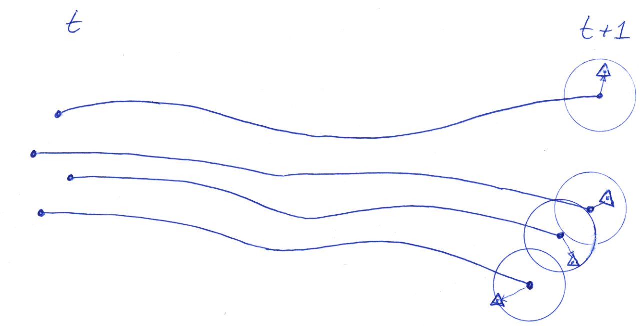
$$\mathbf{x}^{b}_{t+1} = M\mathbf{x}^{a}_{t} \qquad \left\langle \left(\mathbf{x}^{b}_{t+1} - \overline{\mathbf{x}^{b}_{t+1}}\right)\left(\mathbf{x}^{b}_{t+1} - \overline{\mathbf{x}^{b}_{t+1}}\right)^{T} \right\rangle \approx \mathbf{M}\mathbf{P}^{a}\mathbf{M}^{T}$$

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

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

62

Remedy : adding noise at data assimilation time



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

63

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)

64

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).

65