Statistical Methods for Numerical Weather Prediction.

Thomas Bengtsson
(Doug Nychka, Chris Snyder)
Geophysical Statistics Project – NCAR, Boulder, CO
`www.cgd.ucar.edu/~tocke/` *tocke@ucar.edu*

University of Missouri-Columbia
Department of Statistics
Oct 11, 2001.

GSP is supported by the National Science Foundation
under grants DMS 9815344 and DMS 9312686.

Outline

1. Numerical Weather Prediction
    - Ensemble Forecasting
    - Lorenz ODE's
2. Atmospheric Data Assimilation & State-Space Framework
    - Nonlinearities (Non-gaussian forecast distributions)
        - Simulations
    - High-dimensional (Computational issues)
3. Future Directions

Numerical Weather Prediction (NWP)

- A weather forecast is produced by integrating forward (in time) a system of nonlinear differential equations:

$$\mathbf{x}_{t+\delta_t} = \mathbf{x}_t + \int_t^{t+\delta_t} \Omega(u)du$$

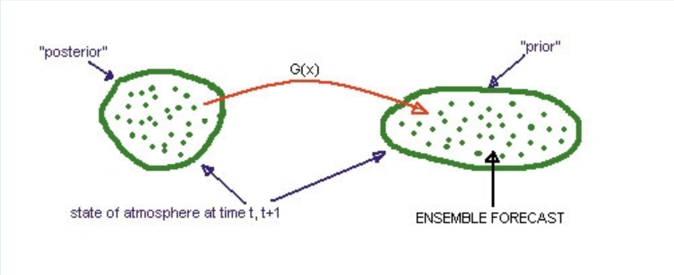
  Here, $\mathbf{x}_t$ is initial condition (current state of atmosphere) and $\Omega(t) = \dot{\mathbf{x}}_t$ defines the physics.

- A NWP model must be able to incorporate and combine:
  1. physical laws for atmosphere (classical mechanics, thermo dynamics).
  2. statistical and numerical techniques.

- Forecasting (weather) is an uncertain proposition
  - a matter of probability?

Ensemble Forecasting

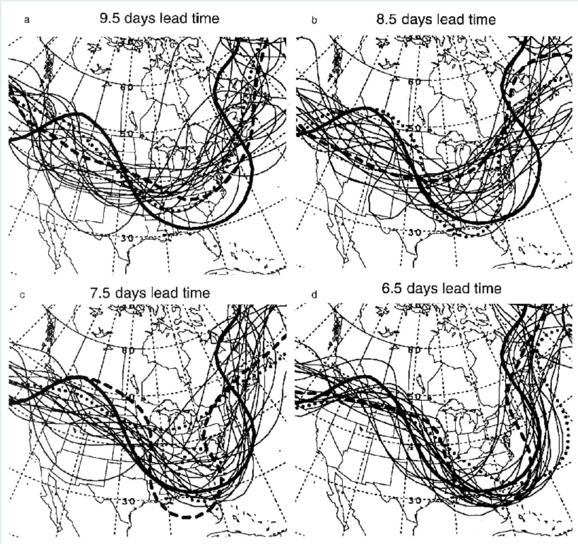
- A probabilistic view of prediction: $p(\mathbf{x}_t)$.
- Difficult to solve forward integration of $p(\mathbf{x}_t)$ analytically.
- An ensemble forecast is a (sample) collection of weather forecasts that verify at the same time.



- The ensemble is (generally) derived under the same dynamic model starting from different initial conditions.
- Issue: how to sample from posterior?

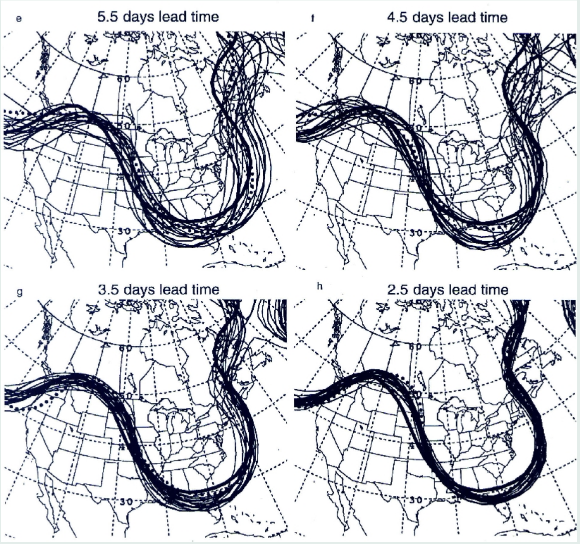
NCEP Ensemble (Sivillo *et al.*, 1997)

- 5640-m contour line of 500-hPa height field (15 November, 1995)



- Solid line ↔ actual weather, 15 Nov.

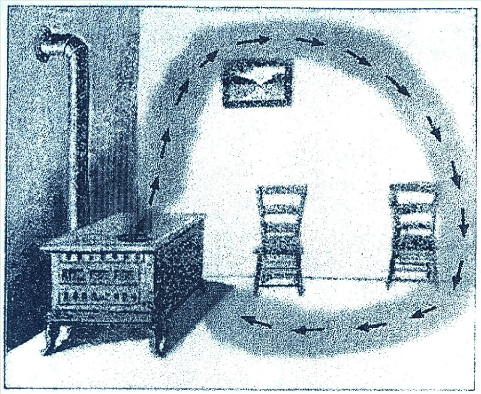
NCEP Ensemble (Sivillo *et al.*, 1997)



- Ensemble spread (variance) decreased $\leftrightarrow$ forecast more accurate.

Lorenz System (Lorenz, 1963)

- Simplification of motion of a fluid heated below in a gravitational field.

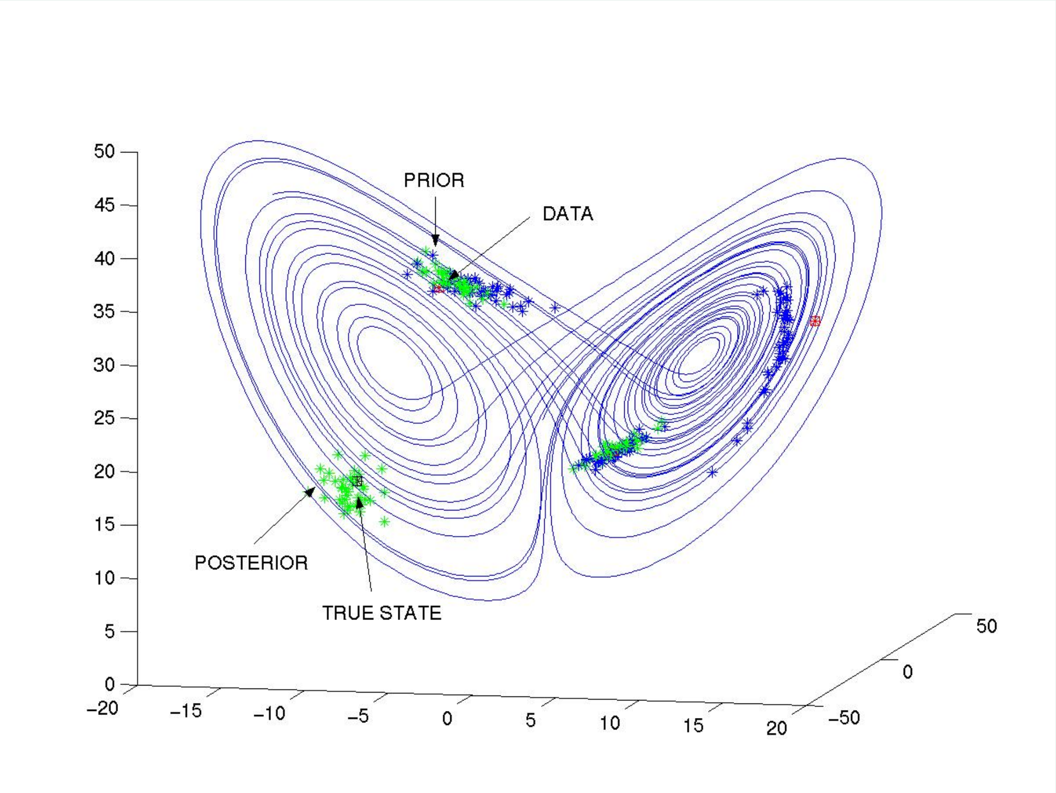


- Equations:

$$\begin{aligned}\dot{x}_t &= -\sigma(x_t + y_t) \\ \dot{y}_t &= r x_t - y_t - x_t y_t \\ \dot{z}_t &= x_t y_t - b z_t\end{aligned}$$

- $x_t \propto$ intensity of fluid flow
  y_t represents ΔT between ascending/descending currents
  $z_t \propto$ temperature gradient.

Ensemble Forecast



Atmospheric Data Assimilation & State-Space Framework

Data Assimilation

Updating our knowledge of the state of the atmosphere once new weather data is available.

Atmospheric Model

Weather observations $\longrightarrow$ $\mathbf{y}_t = H_t \mathbf{x}_t + \boldsymbol{\epsilon}_t$

Atmospheric State $\longrightarrow$ $\mathbf{x}_t = G(\mathbf{x}_{t-1})$

$\mathbf{y}_t \in \Re^{10^5}$, data

$\mathbf{x}_t \in \Re^{10^7}$, unobserved

H_t maps state to observation (linear or non-linear)

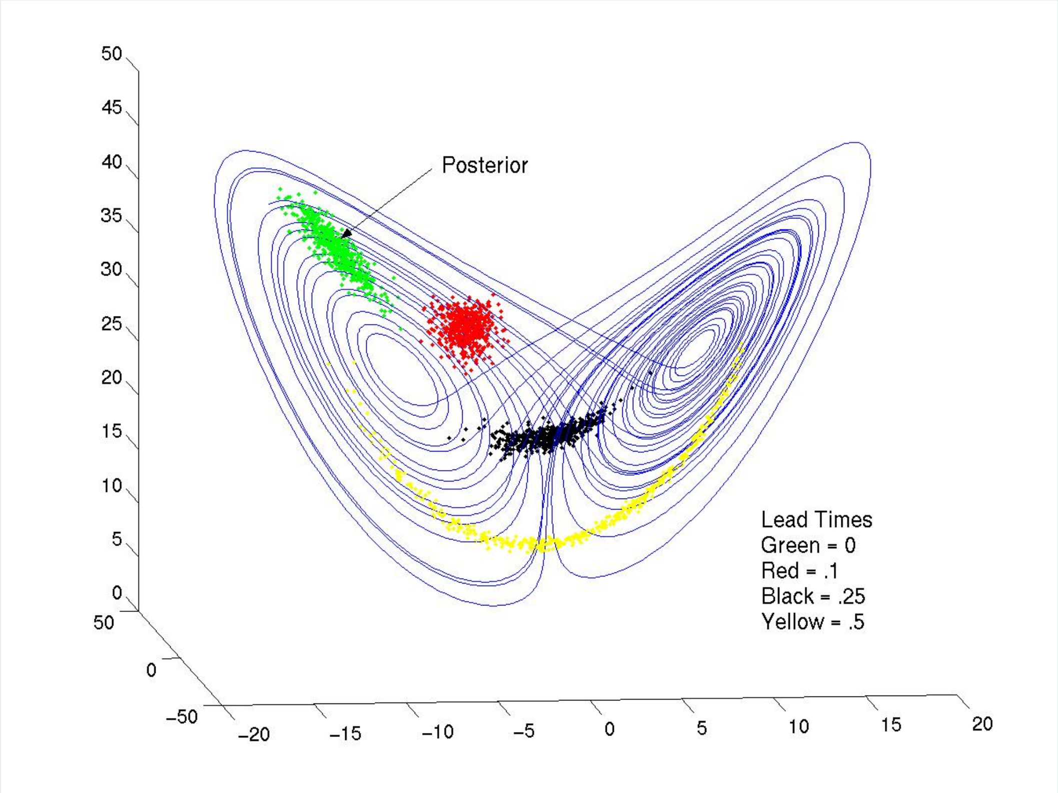
G highly nonlinear, chaotic (known or approximate)

$\boldsymbol{\epsilon}_t$ (gaussian) observation error, $cov(\boldsymbol{\epsilon}_t) = \mathbf{R}$

Sequential assimilation and forecasting:

$$p(\mathbf{x}_t),\ \mathbf{y}_t \xrightarrow{Bayes} p(\mathbf{x}_t|\mathbf{y}_t) \quad \xrightarrow{G(\cdot)} \quad p(\mathbf{x}_{t+1}),\ \mathbf{y}_{t+1} \xrightarrow{Bayes} p(\mathbf{x}_{t+1}|\mathbf{y}_{t+1})$$

Forecast Chaos



Linear Filtering

- Assuming $p(\mathbf{x_t})$ and $p(\mathbf{y}_t|\mathbf{x_t})$ both normal, assimilate $\mathbf{y}_t$ and $p(\mathbf{x}_t)$ using the Kalman Filter (KF) (Kalman, 1960):

$$E(\mathbf{x}_t|\mathbf{Y}_t) = E(\mathbf{x}_t|\mathbf{Y}_{t-1}) + \mathbf{K}_t[\mathbf{y}_t - \mathbf{H}_t E(\mathbf{x}_t|\mathbf{Y}_{t-1})]$$

$$\mathbf{P}_t^u = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_t^f,$$

where

$$\mathbf{K}_t = \mathbf{P}_t^f\mathbf{H}_t'(\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t' + \mathbf{R})^{-1}, \text{ and}$$
$$\mathbf{P}_t^f = E\{[\mathbf{x}_t - E(\mathbf{x}_t|\mathbf{Y}_{t-1})][\mathbf{x}_t - E(\mathbf{x}_t|\mathbf{Y}_{t-1})]'\}.$$

- Easy to implement sequentially in systems with linear dynamics.
- Covariance recursion expensive for high-dimensional systems.

The Ensemble Kalman Filter (EnsKF)

- EnsKF proceeds by estimating the first two moments of the forecast distribution using an ensemble (sample) of state vectors.

$$\hat{E}(\mathbf{x}_t|\mathbf{Y}_t) = \hat{E}(\mathbf{x}_t|\mathbf{Y}_{t-1}) + \hat{\mathbf{K}}_t[\mathbf{y}_t - \mathbf{H}_t\hat{E}(\mathbf{x}_t|\mathbf{Y}_{t-1})]$$

  Algorithm:

  i. sample $\mathbf{x}_i \sim p(\mathbf{x}_{t-1}|\mathbf{Y}_{t-1})$, for i=1,...,m.
  ii. propagate $\mathbf{x}_i^f = G(\mathbf{x}_i)$, for i = 1,...,m.
  iii. calculate $\hat{E}(\mathbf{x}_t|\mathbf{Y}_{t-1}) = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i^f$, and $\hat{\mathbf{P}}_t^f$.
  iv. update $\hat{E}(\mathbf{x}_t|\mathbf{Y}_{t-1})$ using the sample moments from iii).

- Advantage: covariance information is propagated in a compact and reduced dimensional form, real-time efficiency, feasible to implement in high-dimensional systems, performs well in low-order systems with unstable dynamics

Nonlinear Filtering

- No universal analytical solution exists.
- Use extended KF (Jazwinski, 1970); or, Sequential Monte Carlo (SMC) filters (Doucet *et al.*, 2001).
- Expect problems with the extended KF if:
  - $G(\cdot)$ strongly nonlinear (Evensen, 1994; Miller *at al.* 1994).

    use sample based filter, e.g. ensemble Kalman filter (Evensen, 1994).
  - $p(\mathbf{x}_t)$ non-gaussian

    use mixture (Gaussian sum) Kalman filter (Alspach & Sorenson, 1972; Chen & Liu, 2000).
- Expect problems with SMC filters if:
  - $\dim(\mathbf{x}_t)$ is large (Gilks *et al.*, 1996; Robert & Casella, 1999)

A Mixture Ensemble Kalman Filter

- Suppose $p(\mathbf{x}_t)$ is non-gaussian.
- We approximate $p(\mathbf{x}_t)$ by a mixture of Gaussian distributions:

$$p(\mathbf{x}_t) = \sum_{i=1}^{k} p_i \text{MN}(\boldsymbol{\mu}_i, \mathbf{P}_i)$$

By Bayes theorem,

$$p(\mathbf{x}_t | \mathbf{y}_t) = \sum_{i=1}^{k} p_i^{\star} \text{MN}(\boldsymbol{\mu}_i^{\star}, \mathbf{P}_i^{\star})$$

Here $(\boldsymbol{\mu}_i^{\star}, \mathbf{P}_i^{\star})$ are found by KF, and $p_i^{\star} \propto p_i \times p(\mathbf{y}_t | \boldsymbol{\mu}_i, \mathbf{P}_i)$

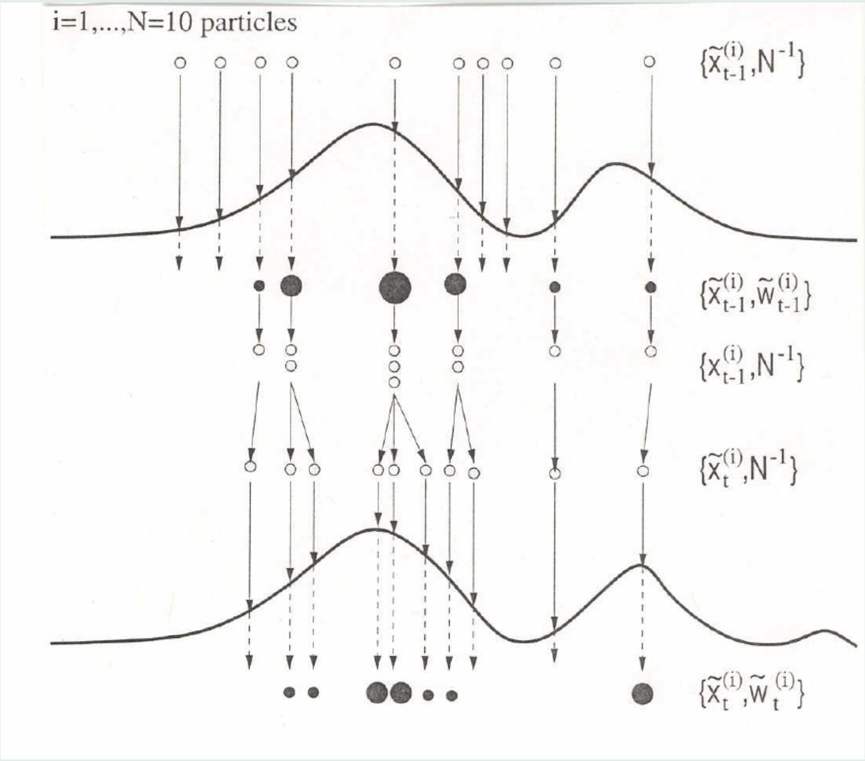
- Need to choose k, $\text{MN}(\boldsymbol{\mu}_i, \mathbf{P}_i)$

  Gaussian prior/posterior $\rightarrow$ k = 1

  Kernel density estimate $\rightarrow$ k = ensemble size

A (Bootstrap) Particle Filter (Doucet *et al.*, 2001)

$$\hat{p}(\mathbf{x}_{t-1}),\ \mathbf{y}_{t-1} \xrightarrow{\text{Bayes}} \hat{p}(\mathbf{x}_{t-1}|\mathbf{y}_{t-1}) \quad \xrightarrow{\text{G(sample)}} \quad \hat{p}(\mathbf{x}_{t}),\ \mathbf{y}_{t} \xrightarrow{\text{Bayes}} \hat{p}(\mathbf{x}_{t}|\mathbf{y}_{t})$$



A Sampling Scheme

1. Calculate $\boldsymbol{\mu}_i^\star$ using KF, and find $p_i^\star$.
2. Generate the following random quantities:

$$\mathbf{x}^\star \sim \mathrm{MN}(\mathbf{0}, \mathbf{P}_i) \;, \quad \mathbf{y}^\star = \mathbf{H}\mathbf{x}^\star + \mathbf{e}$$

   where $\mathbf{e} \sim \mathrm{MN}(\mathbf{0}, \mathbf{R})$.

3. Find $\mathbf{u} = \mathbf{x}^\star - \mathbf{K}_i\mathbf{y}^\star$, and let $\mathbf{z}_i = \boldsymbol{\mu}_i^\star + \mathbf{u}$. Use $\mathbf{z}_i$ with probability $p_i^\star$.

Note that the perturbation $\mathbf{u}$ has the correct (posterior) covariance:

$$\begin{aligned}
\mathrm{Cov}(\mathbf{u}) &= \mathrm{Cov}(\mathbf{x}^*) + \mathrm{Cov}(\mathbf{K}_i\mathbf{y}^\star) - 2\mathrm{Cov}(\mathbf{x}^*, \mathbf{K}_i\mathbf{y}^\star) \\
&= \mathbf{P}_i + \mathbf{P}_i\mathbf{H}^T(\mathbf{H}\mathbf{P}_i\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_i - 2\mathbf{P}_i\mathbf{H}^T(\mathbf{H}\mathbf{P}_i\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_i \\
&= \mathbf{P}_i - \mathbf{K}_i\mathbf{H}\mathbf{P}_i = \mathbf{P}_i^u
\end{aligned}$$

- No need to factor $\mathbf{P}_i$ if $\mathbf{x}^\star$ is a sample perturbation from the prior ensemble.

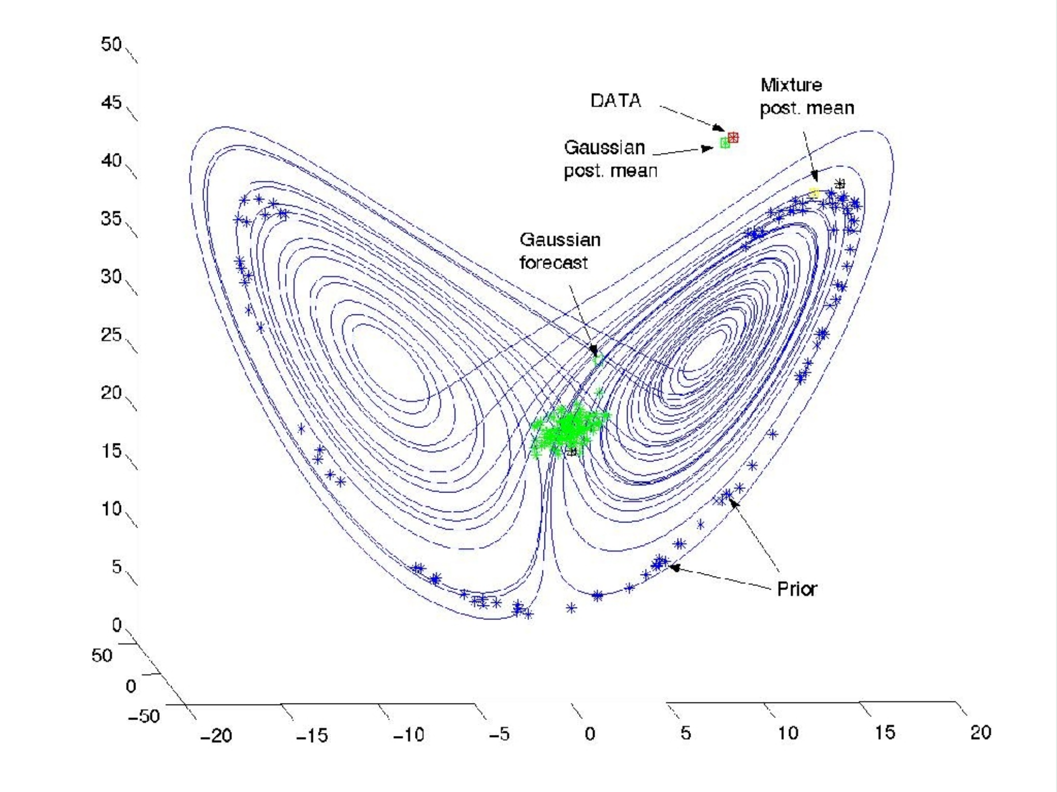
Form of prior	$\hat{p}(\mathbf{x_t})$	Statistics
Gaussian	$\mathrm{MN}(\hat{\boldsymbol{\mu}}, \hat{\mathbf{P}})$	$(\hat{\boldsymbol{\mu}}, \hat{\mathbf{P}})$ = ensemble mean, cov.
Mixture	$\sum_{i=1}^{m} \frac{1}{m}\mathrm{MN}(\hat{\boldsymbol{\mu}}_i, \hat{\mathbf{P}}_i)$	$\hat{\boldsymbol{\mu}}_i$ = i:th ensemble member
		$\hat{\mathbf{P}}_i$ = cov. in neighborhood of $\hat{\boldsymbol{\mu}}_i$

- m = (40,400), $\mathbf{H}_t = \mathbf{I}$, var($\boldsymbol{\epsilon}_t$) = 4$\mathbf{I}$, T = 5000.
- Error measure:
  median of $RMSE_t = \sqrt{(\mathbf{x}_t - \hat{E}(\mathbf{x}_t|\mathbf{y}_t))'(\mathbf{x}_t - \hat{E}(\mathbf{x}_t|\mathbf{y}_t))/3}$

δ_t	Gaussian, k=1		Mixture, k=m	
	$m = 40$	$m = 400$	$m = 40$	$m = 400$
.1	.51	.35	.54	?
.25	.75	.68	.56	.52
.5	1.06	1.05	.76	.71

- Improvement is 25-30%.

Unstable Dynamics



Conditional Simulation Results

- Condition on posterior means from gaussian assimilation located in saddle (T = 250).
- m = (40,400), $\mathbf{H}_t = \mathbf{I}$, $\text{var}(\boldsymbol{\epsilon}_t) = 4\mathbf{I}$.
- Error measure: median of $RMSE_t$

δ_t	Gaussian	Mixture	
	$m = 400$	$m = 40$	$m = 400$
.5	1.64	.94	.73

- Improvement following unstable (saddle) area is 45-55%.

Lemma

For uncorrelated (independent) measurement errors, sequential assimilation of observations yields the same result as simultaneous assimilation.

- Implication: The inverse of the Kalman gain matrix does not have to be explicitly calculated.

- The j^{th} observation at time t is related to the state by $y_t^j = h_t^j \mathbf{x}_t + \epsilon_t^j$. Assimilation of y_t^j is given by

$$E(\mathbf{x}_t|\mathbf{y}_t^j, \mathbf{Y}_{t-1}) = E(\mathbf{x}_t|\mathbf{y}_t^{(j-1)}, \mathbf{Y}_{t-1}) + \mathbf{K}_t^{(f,j-1)}[y^j - h_t^j E(\mathbf{x}_t|\mathbf{y}_t^{(j-1)}, \mathbf{Y}_{t-1})],$$

where $\mathbf{y}_t^k = (y_t^1, y_t^2, \ldots, y_t^k)'$, and $\mathbf{K}_t^{(f,k)} = \frac{\mathbf{P}_t^{(f,k)} h_t^{j\prime}}{(h_t^j \mathbf{P}_t^{(f,k)} h_t^{j\prime} + R)}$.

- To obtain $E(\mathbf{x}_t|\mathbf{Y}_t)$ iterate above for all observations in $\mathbf{y}_t$.

Computational Issues: Limiting Impact of Observations.

- Observations are (fairly) local in space $\rightarrow h_t$ is sparse.
- (From sequential update) Need to compute $\mathbf{P}_t^f h_t'$.

$$\hat{\mathbf{P}}_t^f h_t' = \frac{1}{m-1} \sum_{i=1}^{m} [\mathbf{x}_i^f - \hat{E}(\mathbf{x}_t | \mathbf{Y}_{t-1})] \{ h_t [\mathbf{x}_i^f - \hat{E}(\mathbf{x}_t | \mathbf{Y}_{t-1})] \}'$$
$$= \frac{1}{m-1} \sum_{i=1}^{m} \alpha_i [\mathbf{x}_i^f - \hat{E}(\mathbf{x}_t | \mathbf{Y}_{t-1})]$$

- In terms of error structure, physically remote state variables *should* be uncorrelated. Observations over Boulder should not update the atmospheric state over London.
- By considering local information in the updates, effects of sampling variability of $\mathbf{x}_i^f$ is decreased.

MSE Properties of EnsKF

- What are the effects of sampling variability on EnsKF update?
- Suppose linear dynamics: $\hat{E}(\mathbf{x}_t|\mathbf{Y}_{t-1}) \sim \text{MN}(E(\mathbf{x}_t|\mathbf{Y}_{t-1}), \frac{1}{m}\mathbf{P}_t^f)$.

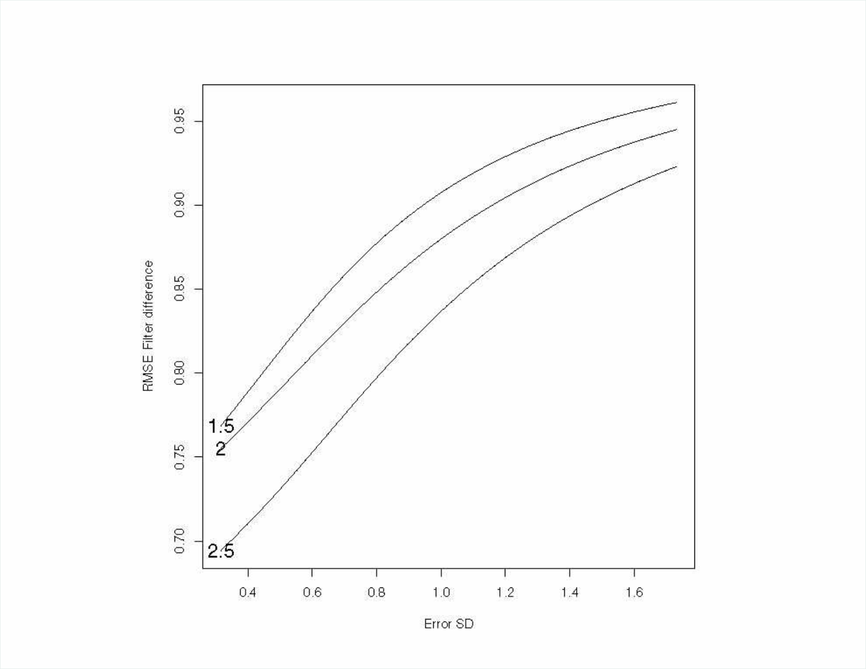
  We have the following orthogonal decomposition:

$$
\begin{aligned}
\hat{E}(\mathbf{x}_t|\mathbf{Y}_t) =& E(\mathbf{x}_t|\mathbf{Y}_{t-1}) + \mathbf{K}_t[\mathbf{y}_t - \mathbf{H}_t E(\mathbf{x}_t|\mathbf{Y}_{t-1})] \\
&+ (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)[\hat{E}(\mathbf{x}_t|\mathbf{Y}_{t-1}) - E(\mathbf{x}_t|\mathbf{Y}_{t-1})] \\
&+ (\hat{\mathbf{K}}_t - \mathbf{K}_t)[\mathbf{y}_t - \mathbf{H}_t E(\mathbf{x}_t|\mathbf{Y}_{t-1})] \\
&+ (\mathbf{K}_t - \hat{\mathbf{K}}_t)\{\mathbf{H}_t[\hat{E}(\mathbf{x}_t|\mathbf{Y}_{t-1}) - E(\mathbf{x}_t|\mathbf{Y}_{t-1})]\}
\end{aligned}
$$

- Let $\eta_t = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)[\hat{E}(\mathbf{x}_t|\mathbf{Y}_{t-1}) - E(\mathbf{x}_t|\mathbf{Y}_{t-1})]$.
- We wish to study $E({\eta_t}'\eta_t)$ as a function of the eigenstructure (values) of $\mathbf{P}_t^f$.

Effects of Sampling Variability on EnsKF

- Let i^{th} eigenvalue of $\mathbf{P}_t^f \propto i^{-\theta}, \theta > 1$. Set $tr(\mathbf{P}_t^f) = 1$.



- Here, $\mathbf{H}_t = \mathbf{I}$, and $\mathbf{R} = \mathbf{I}$. Y-axis: $\sqrt{m}\frac{E(\eta_t'\eta_t)}{tr(\mathbf{P}_t^u)}$, X-axis: $\frac{tr(\mathbf{R})}{tr(\mathbf{P}_t^u)}$.

Summary

- We have presented a bootstrap mixture Kalman filter for recursively tracking atmospheric states.
- But, ... what we really have is a updating procedure which is locally linear.
- Maybe, ... the weighting (represented by p_i's) can resolve non-gaussian structures.

Future Directions

- How to construct mixture? (order, kernels)
- Sequential parameter estimation; incorporate model (parameter) uncertainty.
- Validate mixture approach on higher dimensional system, e.g. Lorenz (1996).
- Formalize algorithms for rank deficient cases, i.e., when $m \lll dim(\mathbf{x}_t)$.