Current progress in data assimilation with the Local Ensemble Transform Kalman Filter

Eric Kostelich

ARIZONA STATE UNIVERSITY
SCHOOL OF MATHEMATICAL & STATISTICAL SCIENCES

June 23, 2010

Many Collaborators

- Istvan Szunyogh, U. Maryland and Texas A & M University
- Brian Hunt, Edward Ott, Eugenia Kalnay, U. Maryland
- Ross Hoffman, Atmospheric & Environmental Research, Inc., Lexington, MA
- Students & postdocs: D.J. Patil, David Kuhl, Junjie Liu, Dasa Merkova, Matt Hoffman, Steve Greybush, Liz Satterfield, Elena Fertig

P/reprints

- http://www.weatherchaos.umd.edu
- Mathematical details: B. R. Hunt, E. K., I. Szunyogh, Physica D **230** (2007), 112–126
- Review paper on GFS: I. Szunyogh, E. K., *et al.*, Tellus A **60** (2008), 113–130
- Bias estimation: S. J. Baek, B. R. Hunt, E. Kalnay, E. Ott, I. Szunyogh, Mon. Wea. Rev. **139** (2009), 2349–2364.
- ECOM model: R. N. Hoffman, *et al.*, J. Atmos. Ocean Tech. **25** (2008), 1638–1656.

Talk outline

- The basic approach
- Bias estimation
- Extensions: non-Gaussianity, nonlinear $\mathbf{H}$

Introduction: the estimation problem

- Observations: $\mathbf{y}_o = \mathbf{H}(\mathbf{x}_t) + \varepsilon$ with $E(\varepsilon) = \mathbf{0}$ and $E(\varepsilon\varepsilon^{\mathrm{T}}) = \mathbf{R}$
- Forecast: $\mathbf{x}_b = \mathbf{x}_t + \eta$ with $E(\eta) = \mathbf{0}$ and $E(\eta\eta^{\mathrm{T}}) = \mathbf{P}_b$
- Minimize the quadratic cost function: $J(\mathbf{x}) = \left(\mathbf{y}_o - \mathbf{H}(\mathbf{x})\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}_o - \mathbf{H}(\mathbf{x})\right) + (\mathbf{x} - \mathbf{x}_b)^{\mathrm{T}}\mathbf{P}_b^{-1}(\mathbf{x} - \mathbf{x}_b)$
- Current NCEP operations: $\sim$ 1.75 million obs assimilated into a $\sim$ 3 billion variable model every 6 hours

Minimization properties

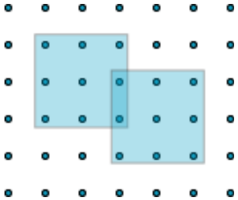
- When the errors are Gaussian and the underlying dynamics are linear, the minimizer of J is “optimal” (unbiased, minimum variance)
- To evaluate J, we must invert $\mathbf{R}$ and $\mathbf{P}_b$
- The observation covariance matrix $\mathbf{R}$ is a nearly diagonal $p \times p$ matrix with $p \sim 10^6 - 10^7$
- $\mathbf{P}_b$ is not diagonal and is $n \times n$, where $n \sim 10^7 - 10^9$ for typical weather models
- Ensemble Kalman filters attempt a low-rank empirical estimate of $\mathbf{P}_b$

Our approach: Use dynamics to reduce the dimensionality

- Key finding: Over typical synoptic regions, the forecast uncertainty evolves in a much lower-dimensional space than the phase space
- Example: The Global Forecast System at T62 resolution has ~ 3000 variables in a typical 1000×1000-km synoptic region
- An ensemble of 100–200 forecast vectors typically spans a ~ 40-dimensional subspace over short forecast periods

The LETKF algorithm: update overlapping local regions

- Consider all observations with a prespecified distance of each grid point
- Given k ensemble forecasts, change coordinates to the $(k-1)$-dimensional subspace that they span
- Compute the Kalman update in these coordinates
- Naturally parallel algorithm



Relation to 4DVar

- The LETKF and 4DVar attempt to minimize the same type of cost function
- LETKF uses a flow-dependent background covariance (determined empirically from an ensemble of forecasts)
- 4DVar requires integration of the nonlinear model and its linearization
- LETKF is model independent

Current applications

- NCEP Global Forecast System (GFS)
- ECOM (Estuarine & Coastal Ocean Model)—with Ross Hoffman
- NCEP Regional Spectral Model—Dasa Merkova
- CO and O_3 assimilation—Dave Kuhl
- GFDL model (Mars Microwave Sounder)—John Wilson, Matt Hoffman
- CAM/CASA' model (OCO)—Junjie Liu, Inez Fung
- CPTEC and JMA working towards operational implementation

Other advantages of the LETKF

- Uses flow-dependent covariances, including the analysis uncertainty
- Assimilates all data at once—full 4-d scheme
- Nonlinear observation operators are readily accommodated
- The local region and ensemble size are the only free parameters
- No adjoint necessary!
- Model parameters can be estimated as part of the state vector

Details of the local analyses

- Goal: Find the linear combination of ensemble solutions that best fits the observations
- Start with background ensemble $\{\mathbf{x}_b^i\}_{i=1}^k$ with mean $\bar{\mathbf{x}}_b$ and perturbation matrix

$$\mathbf{X}_b = \left(\mathbf{x}_b^1 - \bar{\mathbf{x}}_b \;\; \mathbf{x}_b^2 - \bar{\mathbf{x}}_b \;\; \cdots \;\; \mathbf{x}_b^k - \bar{\mathbf{x}}_b \right)$$

- Compute the analysis mean $\bar{\mathbf{x}}_a$ and analysis ensemble perturbations $\mathbf{X}_a = \mathbf{X}_b \mathbf{W}_a$
- The analysis is $\mathbf{X}_a + \bar{\mathbf{x}}_a$

Change of basis to the background perturbations

- The ensemble covariance matrix $\mathbf{P}_b = (k-1)^{-1}\mathbf{X}_b\mathbf{X}_b^{\mathrm{T}}$ has rank $k-1$
- Since $\mathbf{X}_b$ is 1–1 on its column space S, we minimize J on S, where $\mathbf{P}_b^{-1}$ is defined
- Treat $\mathbf{X}_b$ as a linear transformation from an abstract k-dimensional space $\tilde{S}$ to S
- If $\mathbf{w} \in \tilde{S}$ is Gaussian with mean $\mathbf{0}$ and covariance $(k-1)^{-1}\mathbf{I}$, then $\mathbf{x} = \bar{\mathbf{x}}_b + \mathbf{X}_b\mathbf{w}$ is Gaussian with mean $\bar{\mathbf{x}}_b$ and covariance $\mathbf{P}_b$

The objective function

- The cost function

$$
\begin{aligned}
J(\mathbf{x}) &= (\mathbf{x}-\mathbf{x}_b)^{\mathrm{T}}\mathbf{P}_b^{-1}(\mathbf{x}-\mathbf{x}_b) \\
&\quad + \left(\mathbf{y}_o - \mathbf{H}(\mathbf{x})\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}_o - \mathbf{H}(\mathbf{x})\right)
\end{aligned}
$$

becomes

$$
\begin{aligned}
\widetilde{J}(\mathbf{w}) &= (k-1)^{-1}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \\
&\quad [\mathbf{y}_0 - \mathbf{H}(\bar{\mathbf{x}}_b + \mathbf{X}_b\mathbf{w})]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y}_0 - \mathbf{H}(\bar{\mathbf{x}}_b + \mathbf{X}_b\mathbf{w})]
\end{aligned}
$$

and we linearize $\mathbf{H}(\bar{\mathbf{x}}_b + \mathbf{X}_b\mathbf{w}) \approx \left(\bar{\mathbf{y}}_b - \mathbf{Y}_b\mathbf{w}\right)$ where

$$
\bar{\mathbf{y}}_b = \overline{\mathbf{H}(\mathbf{x}_b^i)} \quad \text{and} \quad \mathbf{Y}_b = \left(\, \mathbf{H}(\mathbf{x}_b^1) - \bar{\mathbf{y}}_b \;\cdots\; \mathbf{H}(\mathbf{x}_b^k) - \bar{\mathbf{y}}_b \,\right)
$$

Treatment of the observation operator $\mathbf{H}$

- Only the components of $\mathbf{H}(\mathbf{x}_b^i)$ belonging to the current local region are selected to form $\bar{\mathbf{y}}_b$ and $\mathbf{Y}_b$
- If $\mathbf{w}_a$ minimizes $\widetilde{J}$, then $\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_b + \mathbf{X}_b\mathbf{w}_a$ minimizes the original J
- Modifications are required for satellite radiance observations

The analysis ensemble

- The minimizer $\mathbf{w}_a = \mathbf{Q}\mathbf{Y}_b^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}_o - \bar{\mathbf{y}}_b)$, where $\mathbf{Q} = [(k-1)\mathbf{I} + \mathbf{Y}_b^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}_b]^{-1}$
- The analysis mean is $\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_b + \mathbf{X}_b\mathbf{w}_a$ with covariance matrix $\mathbf{P}_a = \mathbf{X}_b\mathbf{Q}\mathbf{X}_b^{\mathrm{T}}$
- The analysis perturbations are $\mathbf{X}_a = \mathbf{X}_b\mathbf{W}_a$ where $\mathbf{W}_a = [(k-1)\mathbf{Q}]^{1/2}$ and we take the symmetric square root
- $\mathbf{W}_a$ then depends continuously on $\mathbf{Q}$ and assures that the columns of $\mathbf{X}_a$ sum to $\mathbf{0}$ (for the correct sample mean)

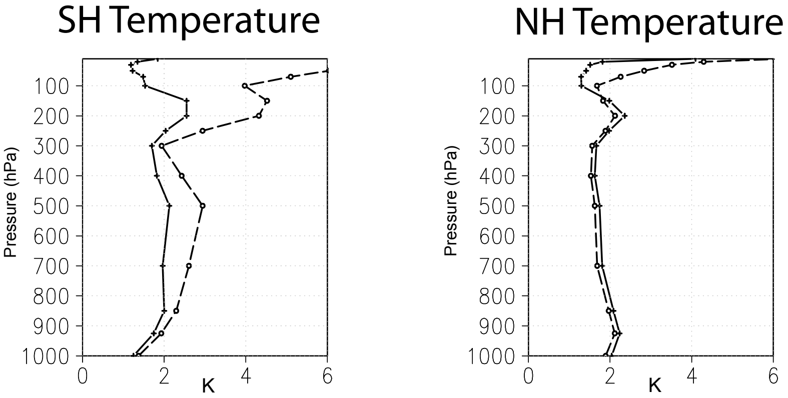
Computational efficiency

- The forecasts are the most expensive part!
- Given an ensemble of size k and s observations in a local region
- Most expensive step ($> 90\%$ of cpu cycles): computing $\mathbf{Y}_b^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}_b$, which is $O(k^2 s)$
- Second most expensive step: computing the symmetric square root, which is $O(k^3)$
- Observation lookup is $O(\log L)$ where L is the size of the total observation set
- Wall-clock time: 40 sec on 16 quad-core processors for 500,000 obs and T62 GFS with a 40-member ensemble

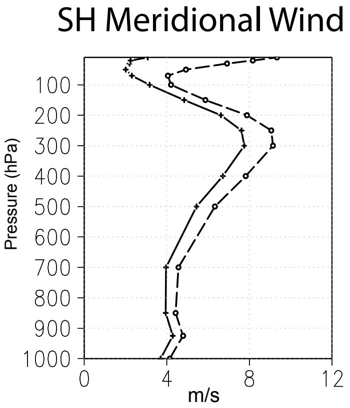
Representative results

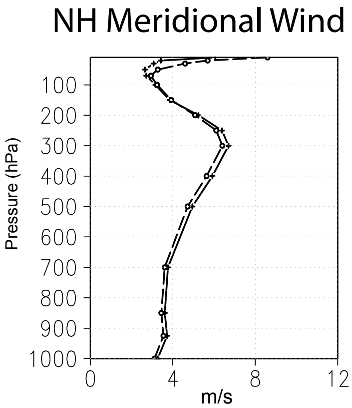
- Reference system ("truth"): NCEP operational analyses (T254L64) using the SSI (Parris & Derber, 1992) with all available data (including satellite radiances), truncated to T62L28
- Benchmark system: NCEP operational system, omitting radiance data, with forecasts run at T62L28
- Comparison: $\langle \text{48-h forecast} - \text{truth} \rangle^2$ with forecasts started from the LETKF analysis and from the Benchmark system

48-hour forecast error: temperature in extratropics

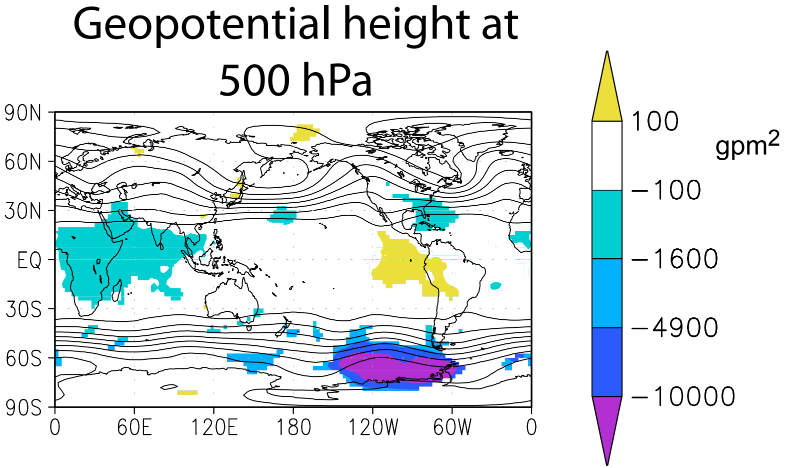


48-hour forecast error: wind in extratropics





Geographical comparison of 48-h forecast error



Model bias

- Many sources of model bias: finite resolution, parametrizations of subgrid processes, errors in boundary conditions, etc.
- Baek *et al.* (2006 and 2009) proposed methods for doing local ensemble Kalman filtering in presence of model bias

Mathematical setup

- Atmospheric state: $\mathbf{u}(\mathbf{r}, t_n) = \mathbf{u}_n$
- Atmospheric dynamics: $\mathbf{u}_n = \mathscr{F}(\mathbf{u}_{n-1})$
- Forecast model: $\mathbf{x}_n = \mathbf{f}(\mathbf{x}_{n-1})$ approximates $\mathbf{u}_n$
- Assumption: There exists a projection $\mathscr{P}$ from the infinite-dimensional space of the atmospheric dynamics to the finite-dimensional model space
- Data assimilation computes an estimate $\mathbf{x}_n^a$ of $\mathscr{P}(\mathbf{u}_n)$ using the forecast $\mathbf{x}_n^b$ and observations $\mathbf{y}_n^o$

Kalman filter data assimilation scheme

- Forecast: $\mathbf{x}_n^b = \mathbf{f}(\mathbf{x}_{n-1}^a)$
- Analysis: $\mathbf{x}_n^a = \mathbf{x}_n^b + \mathbf{K}_n \left[\mathbf{y}_n^o - \mathbf{H}(\mathbf{x}_n^b)\right]$
- Observation operator: $\mathbf{y}_n^o = \mathbf{H}(\mathbf{x}_n^b) + \varepsilon_n$
- "Noise:" $\varepsilon_n \sim N(\mathbf{0}, \mathbf{R}_n)$, assuming $\mathbf{x}_n^b \approx \mathscr{P}(\mathbf{u}_n)$

Kalman filter data assimilation scheme

- Forecast: $\mathbf{x}_n^b = \mathbf{f}(\mathbf{x}_{n-1}^a)$
- Analysis: $\mathbf{x}_n^a = \mathbf{x}_n^b + \mathbf{K}_n \left[\mathbf{y}_n^o - \mathbf{H}(\mathbf{x}_n^b)\right]$
- Observation operator: $\mathbf{y}_n^o = \mathbf{H}(\mathbf{x}_n^b) + \varepsilon_n$
- "Noise:" $\varepsilon_n \sim N(\mathbf{0}, \mathbf{R}_n)$, assuming $\mathbf{x}_n^b \approx \mathscr{P}(\mathbf{u}_n)$
- Model error: $\mathscr{P}(\mathbf{u}_n) = \mathbf{f}(\mathscr{P}(\mathbf{u}_{n-1})) + \mathbf{b}_n$
- Model error evolution: $\mathbf{b}_n = \mathbf{G}(\mathbf{b}_{n-1})$

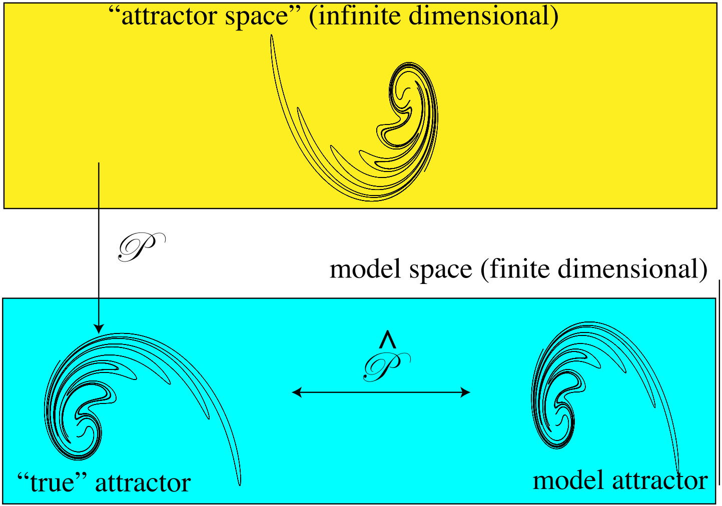
Two potential approaches to handle model bias

- Bias model I: Correct the background using the estimated bias and compute the analysis using the bias-corrected background and latest observations
- Assumption: The model is subject to the same error whether it is started from the projection $\mathscr{P}(\mathbf{u}_{n-1})$ or from the best estimate thereof, $\mathbf{x}^a_{n-1}$

Two potential approaches to handle model bias

- Bias model I: Correct the background using the estimated bias and compute the analysis using the bias-corrected background and latest observations
- Assumption: The model is subject to the same error whether it is started from the projection $\mathscr{P}(\mathbf{u}_{n-1})$ or from the best estimate thereof, $\mathbf{x}_{n-1}^a$
- Bias model II: Assume the forecast dynamics evolve on a model attractor that is shifted from the "true" one
- Assumption: For every finite-dimensional projection $\mathscr{P}(\mathbf{u}_n)$ of the "true" atmospheric state, there is a corresponding point $\widehat{\mathscr{P}}$ on the model attractor

Schematic illustration



Revised Kalman filter implementation

- Bias model I:

$$
\begin{aligned}
\mathbf{x}_n^b &= \mathbf{f}(\mathbf{x}_{n-1}^a) + \mathbf{b}_n \\
\mathbf{x}_n^a &= \mathbf{x}_n^b + \mathbf{K}_n \left[\mathbf{y}_n^o - \mathbf{H}\left(\mathbf{x}_n^b\right) \right]
\end{aligned}
$$

- Bias model II:

$$
\begin{aligned}
\mathbf{x}_n^b &= \mathbf{f}(\mathbf{x}_{n-1}^a) \\
\mathbf{x}_n^a &= \mathbf{x}_n^b + \mathbf{K}_n \left[\mathbf{y}_n^o - \mathbf{H}\left(\mathbf{x}_n^b + \mathbf{b}_n\right) \right]
\end{aligned}
$$

Surface pressure correction schemes

1. Orography correction:

$$\Delta z = \text{model orography} - \text{true orography}$$

   at the observing location

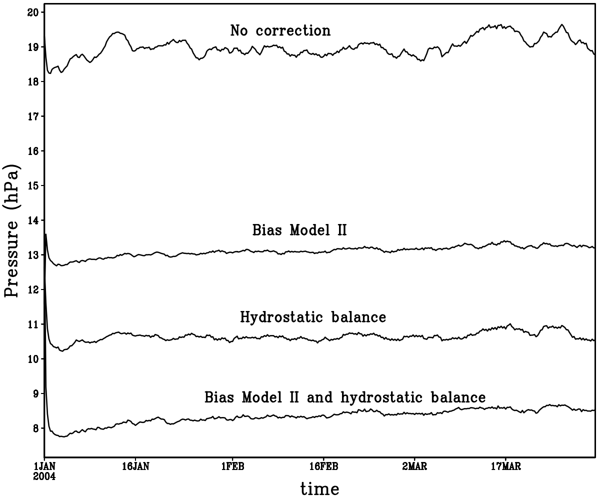
2. Hydrostatic correction: $p_s = p_s^{\text{model}} \times \exp(g\,\Delta z / R\bar{T})$
3. Bias model II: $\mathbf{H}(\mathbf{x}_n^b) = \widehat{\mathbf{H}}(\mathbf{x}_n^b) - \mathbf{c}_n$
4. Bias model II plus hydrostatic correction:

$$\mathbf{H}(\mathbf{x}_n^b) = \widehat{\mathbf{H}}(\mathbf{x}_n^b) \cdot \exp(g\,\Delta z / R\bar{T}) - \mathbf{c}_n$$

Numerical experiment (S. J. Baek)

- "Truth:" Generate surface pressure data using a free run of the T62L28 GFS ($144 \times 73 \times 28$ grid) for Jan.–Feb. 2004 at model grid points approximating a real observing network
- Model: Assimilate synthetic obs (surface pressure, temperature, wind) using a 60-member emsemble with the T30L7 Simplified Parametrization Primitive Equation (SPEEDY) model
- Compute RMS averages of 6-hour surface pressure forecast errors using the different bias correction schemes

Results (verified against background)



Work in progress

- Bias estimation in a radiative transfer model in assimilation of satellite radiance data (with José Aravéquia, CPTEC)
- Apply Baek's scheme to surface pressure measurements in the GFS
- Quantify effects on forecast and analysis accuracy using high-resolution NCEP analyses

Extensions to non-Gaussian data

- Nonlinearities in model evolution cause the ensemble eventually to underestimate the true uncertainty
- Compensate by variance inflation
- In ensemble coordinates, inflation gives (for $\rho \geq 0$)

$$\frac{(k-1)\mathbf{w}^{\mathrm{T}}\mathbf{w}}{1+\rho}$$

- Harlim and Hunt (2007) suggested the modification

$$\frac{(k-1)\mathbf{w}^{\mathrm{T}}\mathbf{w}}{1+\alpha\sqrt{\mathbf{w}^{\mathrm{T}}\mathbf{w}}}$$

Digital filtering with the LETKF (Kostelich & Szunyogh)

- Sometimes the analyzed fields excite spurious dynamics in the model forecasts due to balance issues between atmospheric variables
- Add an extra constraint to penalize (say) divergence or frequency of changes in the analysis window

$$
\begin{aligned}
\widetilde{J}(\mathbf{w}) \;=\;& (k-1)^{-1}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \\
& [\mathbf{y}_0 - \mathbf{H}(\bar{\mathbf{x}}_b + \mathbf{X}_b\mathbf{w})]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y}_0 - \mathbf{H}(\bar{\mathbf{x}}_b + \mathbf{X}_b\mathbf{w})] \\
& + \mathbf{w}^{\mathrm{T}}(\mathbf{X}_b - \mathbf{F})^{\mathrm{T}}\mathbf{Q}(\mathbf{X}_b - \mathbf{F})\mathbf{w}
\end{aligned}
$$

where $\mathbf{F}$ is a suitable linear combination of background forecast perturbations.

Conclusions

- The LETKF is a fast and accurate data assimilation scheme for meteorological applications
- Lends itself to potential applications to nonlinear models and digital filtering (though issues like multiple minima will arise)
- Local low dimensionality of the dynamics is key
- Requires a reasonably dense observing network for continuity
- The LETKF and ensemble diagnostics for predictability may be useful for climate simulations on decadal scales

Thanks to:

- National Science Foundation
- Keck Foundation
- McDonnell Foundation
- Office of Naval Research
- Preprints: www.weatherchaos.umd.edu