

# The Local Ensemble Transform Kalman Filter (LETKF)

Eric Kostelich

Arizona State University

Co-workers:

Istvan Szunyogh, Brian Hunt,  
Ed Ott, Eugenia Kalnay, Jim Yorke,  
and many others

<http://www.weatherchaos.umd.edu>

<http://math.asu.edu/~eric>

## Main topics

- <http://www.weatherchaos.umd.edu>
- **Mathematical details:** B. R. Hunt, E.K., Szunyogh, 2007: Efficient Data Assimilation for Spatiotemporal Chaos: A Local Ensemble Transform Kalman Filter, *Physica D* **230**, 112-126.
- **Review paper:** I. Szunyogh, E.K., et al., 2008: A Local Ensemble Transform Kalman Filter Data Assimilation System for the NCEP Global Model, *Tellus A* **60**, 113-130.

## The estimation problem

- Observations:  $\mathbf{y}_o = \mathbf{H}(\mathbf{x}_t) + \boldsymbol{\varepsilon}$ ,  $\mathbf{E}(\boldsymbol{\varepsilon})=0$ ,  $\mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)=\mathbf{R}$
- Model forecast:  $\mathbf{x}_b = \mathbf{x}_t + \boldsymbol{\eta}$ ,  $\mathbf{E}(\boldsymbol{\eta})=0$ ,  $\mathbf{E}(\boldsymbol{\eta}\boldsymbol{\eta}^T)=\mathbf{P}_b$
- Minimize the objective function

$$J(\mathbf{x}) = (\mathbf{y}-\mathbf{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}-\mathbf{H}(\mathbf{x})) + (\mathbf{x}-\mathbf{x}_b)^T \mathbf{P}_b^{-1} (\mathbf{x}-\mathbf{x}_b)$$

- Current NCEP operations: ~1.75 million obs assimilated into ~3 billion variable model every 6 hours

## Minimization is expensive

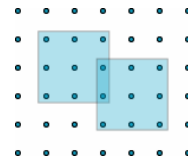
- When the errors are Gaussian and the underlying dynamics are linear, the minimizer of  $\mathbf{J}$  is “optimal” (unbiased, minimum variance).
- To evaluate  $\mathbf{J}$ , we must invert  $\mathbf{R}$  and  $\mathbf{P}_b$
- Observation covariance matrix  $\mathbf{R}$  is a nearly diagonal  $p \times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

## Key idea: Use dynamics to reduce the dimensionality

- Over typical synoptic regions, the forecast uncertainty evolves in a much lower-dimensional space than the phase space
- Appears to be a general property of many spatiotemporal models (weather, climate, ocean)
- A medium-resolution weather model has  $\sim 3000$  variables in a typical  $1000 \times 1000$  km synoptic region ( $\sim$ Texas)
- A typical ensemble of  $100\text{--}200$  forecast vectors over a Texas-sized region spans a  $\sim 40$  dimensional “unstable manifold”

## The LETKF algorithm

- Reduces the uncertainty in this low-dimensional subspace in any given synoptic region ( $\sim 1000 \times 1000$  km)
- Given  $k$  ensemble forecasts, each local  $\mathbf{P}_b$  is  $k \times k$
- The analysis is done independently in a local region around every model grid point using only local obs
- The set of observations used for each local analysis varies slowly in space (important for continuity)
- Naturally parallel algorithm



## The LETKF & 4DVar

- Given unlimited computational resources, each minimizes the same cost function
- 4DVar uses a constant covariance background matrix; LETKF uses time-dependent one from ensemble of forecasts
- 4DVar requires integration of nonlinear model and its linearization
- LETKF uses an ensemble of forecasts instead and permits efficient parallel implementations. **Model independent algorithm.**

## Current applications of the LETKF

- Work in progress:
  - NCEP's Global Forecast System
  - ECOM (Estuarine & Coastal Ocean Model)
  - NASA's Finite-volume General Circulation Model (replacing PSAS)
  - NCEP's Regional Spectral Model
- Beginning work on:
  - Carbon data assimilation for climate models
  - GFDL model of Martian atmosphere (Mars Microwave Sounder)

## Other applications

- CPTEC Brazil in in the process of operational implementation
- JMA has been considering LETKF for ensemble generation, Earth-simulator reanalysis
- Consensus system built at NCEP for testing

## Details of local analyses

- The assimilation is done using observations and a cylinder about each model grid point. *Each local region is processed independently* and the analyses at each center grid point are assembled into a global analysis.
- Given an ensemble of  $k$  background vectors  $\mathbf{x}_b^i$ , form their mean  $\underline{\mathbf{x}}_b$  and the background perturbation matrix  $\mathbf{X}_b$ , whose  $i$ th column is  $\mathbf{x}_b^i - \underline{\mathbf{x}}_b$
- The analysis determines the linear combination of ensemble solutions that best fits the observations (i.e., minimizes the quadratic cost function).

- $\mathbf{P}_b = (k-1)^{-1} \mathbf{X}_b \mathbf{X}_b^T$  is the ensemble covariance matrix of rank  $k-1$ .
- Since  $\mathbf{X}_b$  is 1-1 on its column space  $S$ , we minimize  $\mathbf{J}$  on  $S$ , where  $\mathbf{P}_b^{-1}$  is defined.
- Computing the singular vectors of  $\mathbf{X}_b$  is expensive
- Instead treat  $\mathbf{X}_b$  as a linear transformation from an abstract  $k$ -dimensional space  $S'$  to  $S$
- If  $\mathbf{w} \in S'$  is Gaussian with mean  $\mathbf{0}$  and covariance  $(k-1)^{-1} \mathbf{I}$  then  $\mathbf{x} = \underline{\mathbf{x}}_b + \mathbf{X}_b \mathbf{w}$  is Gaussian with mean  $\underline{\mathbf{x}}_b$  and covariance  $\mathbf{P}_b$

## Treatment of observations

- Linearize  $\mathbf{H}$  about the ensemble mean as  $\mathbf{H}(\underline{\mathbf{x}}_b + \mathbf{X}_b \mathbf{w}) \approx \underline{\mathbf{y}}_b + \mathbf{Y}_b \mathbf{w}$   
where  $\underline{\mathbf{y}}_b$  is the mean of  $\mathbf{H}(\mathbf{x}_b^i)$  and  $\mathbf{Y}_b$  is the matrix whose  $i$ th column is  $\mathbf{H}(\mathbf{x}_b^i) - \underline{\mathbf{y}}_b$
- Only the components of  $\mathbf{H}(\mathbf{x}_b^i)$  belonging to the given local region are selected to form  $\underline{\mathbf{y}}_b$  and  $\mathbf{Y}_b$
- Minimize the cost function 
$$\mathbf{J}'(\mathbf{w}) = (k-1)^{-1} \mathbf{w}^T \mathbf{w} + [\mathbf{y}_o - \underline{\mathbf{y}}_b - \mathbf{Y}_b \mathbf{w}]^T \mathbf{R}^{-1} [\mathbf{y}_o - \underline{\mathbf{y}}_b - \mathbf{Y}_b \mathbf{w}]$$
- The minimizer  $\underline{\mathbf{w}}_a$  of  $\mathbf{J}'$  is perpendicular to the null space of  $\mathbf{X}_b$  and  $\underline{\mathbf{x}}_a = \underline{\mathbf{x}}_b + \mathbf{X}_b \underline{\mathbf{w}}_a$  minimizes the original  $\mathbf{J}$ .

## Net Result

- Minimizer is  $\underline{\mathbf{w}}_a = \mathbf{Q}\mathbf{Y}_b^T\mathbf{R}^{-1}(\mathbf{y}_o - \underline{\mathbf{y}}_b)$  where  
 $\mathbf{Q} = [(k-1)\mathbf{I} + \mathbf{Y}_b^T\mathbf{R}^{-1}\mathbf{Y}_b]^{-1}$
- In model space, the analysis mean becomes  
 $\underline{\mathbf{x}}_a = \underline{\mathbf{x}}_b + \mathbf{X}_b\underline{\mathbf{w}}_a$  with covariance matrix  
 $\mathbf{P}_a = \mathbf{X}_b\mathbf{Q}\mathbf{X}_b^T$
- The analysis perturbations are given by  
 $\mathbf{X}_a = \mathbf{X}_b\mathbf{W}_a$  where  $\mathbf{W}_a = [(k-1)\mathbf{Q}]^{1/2}$  and we take the symmetric square root
- This choice assures that  $\mathbf{W}_a$  depends continuously on  $\mathbf{Q}$  and that the columns of  $\mathbf{X}_a$  sum to zero (for the correct sample mean)

## Computational Efficiency

- Given ensemble of size  $k$  and  $s$  observations in a particular local region
- Most expensive step (> 90% of cpu cycles): computing  $\mathbf{Y}_b^T\mathbf{R}^{-1}\mathbf{Y}_b$  which is proportional to  $k^2 s$
- 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

## Important properties of the LETKF

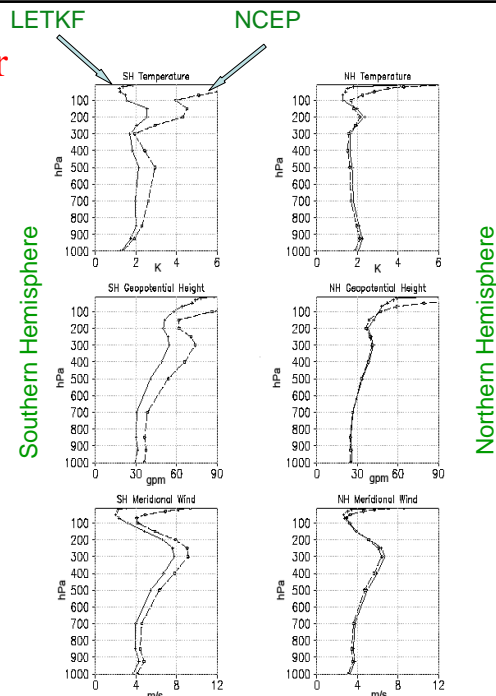
- The estimation problem is solved independently on each local region.
- Updates the estimate of the current state and also its uncertainty
- Assimilates all data at once
- Observations can be nonlinear functions of the state vector
- Can interpolate in time as well as space (full 4-d assimilation scheme)
- The local region size and ensemble size are the only free parameters
- **Model independent** (no adjoints!)
- Can estimate model parameters/errors as well as initial conditions

## Validation Experiments

- Operational (T254, ~50 km) analyses with full observing network are taken as “truth”
- LETKF tested with full observing network (minus satellite radiances but including satellite-derived wind estimates) and 60-member ensemble
- Analysis/forecast cycle run from Jan. 1 to Mar. 1, 2004 using operational GFS at T62 resolution
- **“Benchmark”** analysis: NCEP operational system at T62 (~200 km) with reduced dataset
- **“LETKF”** analysis: Our data assimilation system using reduced dataset at T62 resolution
- Thanks to Zoltan Toth and NCEP

## 48-hour forecast error

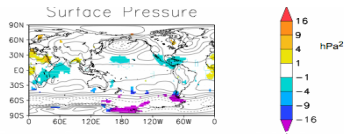
- NCEP operational analyses taken as truth
- Compute  $\langle \text{forecast} - \text{truth} \rangle$  using forecasts started from LETKF and NCEP T62 analyses from Jan. 11 to Feb. 27, 2004
- NCEP surface analyses are copied into LETKF



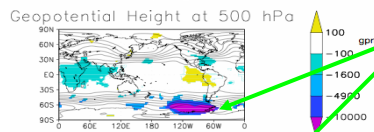
## Notes on Observations

- The LETKF and the Benchmark SSI system use different  $\mathbf{H}$  operators.
- The LETKF does not (yet) assimilate surface data (other than surface pressure) and assumes independent observational errors even in regions of high observational density.
- Handling of Quikscat and other near-surface data is not optimal.
- Benchmark SSI data provided by NCEP (Y. Song and Z. Toth)

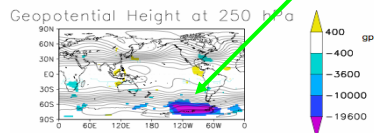
## Geographical comparison of LETKF and the NCEP SSI analysis errors



48-hour forecasts with real observations (no radiances)



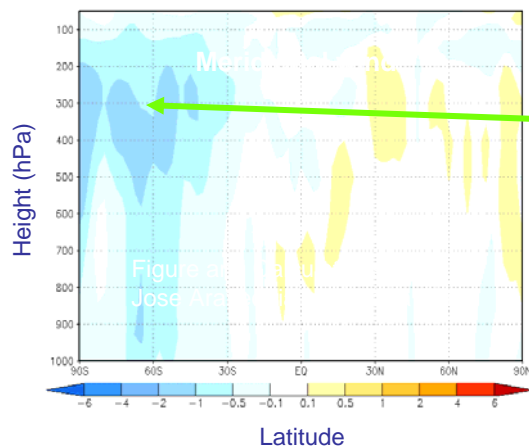
The advantage of the LETKF is the largest where the observation density is the lowest



Results with simulated obs and known true states show a similar geographical distribution

## Assimilation of satellite radiances (José Aravéquia)

Effects of AMSU-A data on 48-h forecasts



- The large improvement in the SH suggests that there is a lot of useful information in the estimated background error covariance matrix between the temperature (most closely related to the radiances) and the wind

## Scaling and memory

- All background data needed to analyze a given model grid point is distributed only once.
- Likewise for the observations, except for observations that are assimilated in adjacent local regions which are assigned to separate processors.
- Load balancing algorithm is expected to provide nearly linear scaling to 8192 cpus.

## Potential optimization

- The LETKF computes the analysis mean and finds the linear combination of ensemble perturbations that best fits the data in a least-squares sense
- The set of observations used to compute it varies slowly with location
- The weight matrix  $\mathbf{W}_a$  can be computed only at a subset of model grid points and interpolated in between.

## Conclusions

- LETKF merits serious consideration for research and operational applications
- LETKF has the largest advantage where the observations are sparse (ideal for ocean and planetary atmosphere DA, if the model errors are sufficiently small)
- Model independence makes it adaptable
- LETKF is one of the most mature, flexible, computationally efficient and well-tested, ensemble-based Kalman filter systems

## Thanks to:

- National Science Foundation
- Keck Foundation
- McDonnell Foundation
- NASA
- Army Research Office
- National Centers for Environmental Prediction
- ASU College of Liberal Arts & Sciences