Assessing the Efficiency of an Ensemble-Based Kalman Filter in Representing the Analysis Uncertainty

Istvan Szunyogh[1] Elizabeth Satterfield[2]

[1]Texas A&M University
Department of Atmospheric Sciences

[2]National Research Council

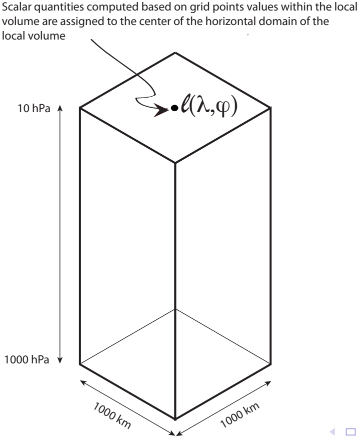
The 5th EnKF Workshop, May 21–24 2012, Rensselaerville, New York

The Problem

- One of the presumed advantages of EnKF over a conventional data assimilation system is that it generates an ensemble of analyses, which is consistent with the estimate of the analysis uncertainty provided by the data assimilation system, for the ensemble forecasting system.
- Do we have convincing evidence to claim success in this area? I would say “no”. (Not much has been said about the topic at this meeting)
- Potential sources of the difficulties:
  - Inherent limitations of EnKF, in particular, using a small ensemble
  - Sub-optimality of the particular ensemble system
- How can we diagnose these sources?

Local State Vector

We define a **local state** vector $\mathbf{x}_\ell$ with all N state variables of the model representation of the state within a local volume centered at location (grid point) ℓ



The Local Space of Ensemble Perturbations, $\mathbb{S}_\ell$

- Given is a K-member **ensemble of local forecasts**:

$$\{\mathbf{x}_\ell^{(k)}, k = 1 \dots K\}$$

- The **local ensemble mean**:

$$\bar{\mathbf{x}}_\ell = K^{-1} \sum_{k=1}^{K} \mathbf{x}_\ell^{e(k)}$$

- The **local ensemble perturbations**:

$$\{\mathbf{x}_\ell'^{(k)} = \mathbf{x}_\ell^{(k)} - \bar{\mathbf{x}}_\ell,\ k = 1 \dots K\}$$

- The **ensemble-based estimate of the local covariance matrix**:

$$\hat{\mathbf{P}}_\ell = (K-1)^{-1} \sum_{k=1}^{k} \mathbf{x}_\ell'^{(k)} \left(\mathbf{x}_\ell'^{(k)}\right)^T,$$

The Dimensionality of $\mathbb{S}_\ell$

- The range of $\hat{\mathbf{P}}_\ell$ (spanned by the K ensemble perturbations) defines a linear space $\mathbb{S}_\ell$ [$dim(\mathbb{S}_\ell) \leq K - 1$]
- The normalized eigenvectors associated with the first $K-1$ eigenvalues of $\hat{\mathbf{P}}_\ell$,

$$\{\mathbf{u}_k, k = 1, \ldots, K-1\}$$

  define an orthonormal basis in $\mathbb{S}_\ell$
- The basis vectors represent linearly independent patterns of uncertainty in the ensemble perturbations in the local region at ℓ.

Decomposition of the Local Forecast Uncertainty

- We define the **local forecast uncertainty** as

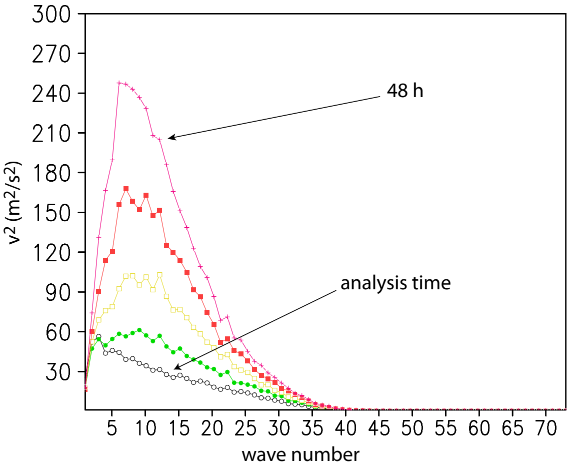
$$\delta \mathbf{x}_\ell = \mathbf{x}_\ell^t - \bar{\mathbf{x}}_\ell = \delta \mathbf{x}_\ell^{(\|)} + \delta \mathbf{x}_\ell^{(\perp)}$$

  where $\mathbf{x}_\ell^t$ is (an estimate of) the **true local state**
- The **local ensemble spread**, $VS_\ell = trace(\hat{\mathbf{P}}_\ell)$, is an estimate of the TV_ℓ expected value of $\|\delta \mathbf{x}_\ell\|^2$ => the expected value, VS, of VS_ℓ over all locations and verification times, should be equal to the expected value, TV, of TV_ℓ over all locations and verification times.
- The projection of $\delta \mathbf{x}_\ell$ into $\mathbb{S}_\ell$ is $\delta \mathbf{x}_\ell^{(\|)}$. We introduce the notation TVS for the expected value of $\left(\delta \mathbf{x}_\ell^{(\|)}\right)^2$
- When $\delta \mathbf{x}_\ell$ can be expressed as a linear combination of the ensemble perturbations, $\delta \mathbf{x}_\ell^{(\|)} = \delta \mathbf{x}_\ell$ and $TVS = TV$

Analysis-Forecast System

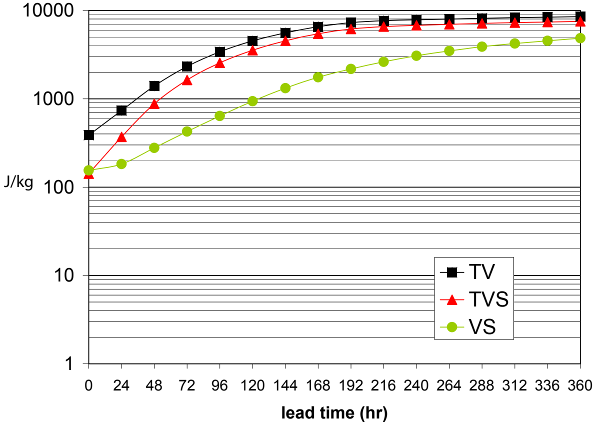
- **Data Assmilation:** Local Ensemble Transform Kalman Filter with 40 ensemble members. (Szunyogh et al. 2008)
- **Model:** 2004 version of NCEP GFS at resolution T62 (about 210 km) and 28-levels
- **Statistics:** Collected for 45 days (January and February 2004), all results shown are for NH extratropics
- **Observations:** (Non-radiance) observations of the atmosphere
- **Variance Inflation:** Was tuned to satisfy $VS \approx TVS$ (larger VS was found to degrade the analyses and ensuing forecasts)
- **Most Results** were summarized in Satterfield and Szunyogh (2010, 2011) and Szunyogh and Satterfield (2011)

Time Evolution of the Power Spectrum of the Forecast Error (Meridional Wind at 500 hPa)



The Evolution of *VS*, *TV*, and *TVS* with Forecast Time

For forecast times longer than about 3 days, $\mathbb{S}_\ell$ provides a good representation of the state $\mathbf{x}_\ell^t$, but the ensemble underestimates the magnitude of $\delta\mathbf{x}_\ell^t$



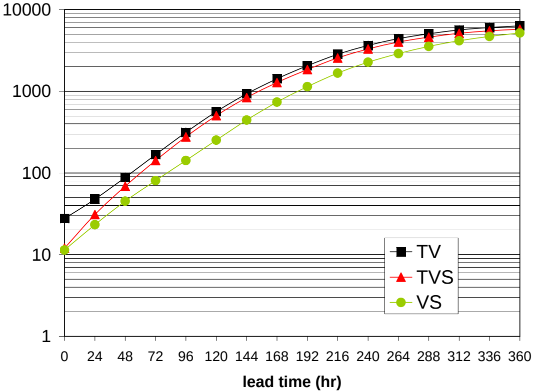
Why Does the Spread Underestimate the Uncertainty (Even at Times When $\mathbb{S}_\ell$ Is Captured Well?)

Potential Answers:

- Lack of accounting for the effects of model uncertainties
- Inherent limitations of EnKF using a small ensemble (e.g., lack of accounting for patterns of uncertainty in the initial conditions, which later pay an important role in the evolution of the forecast uncertainty; assumption of linear error dynamics)
- Sub-optimality of the data assimilation system, which generates the initial perturbations

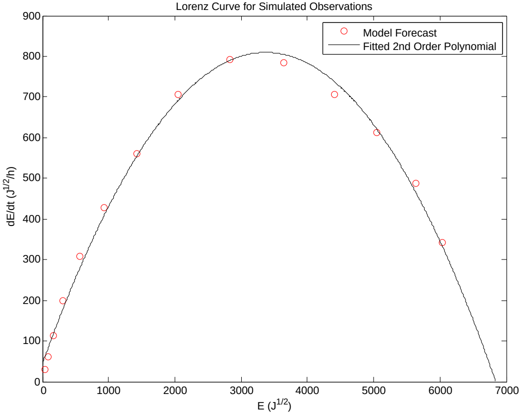
Results with Simulated Observations

The general problem remains, although the magnitude of the underestimation of the uncertainty is less severe

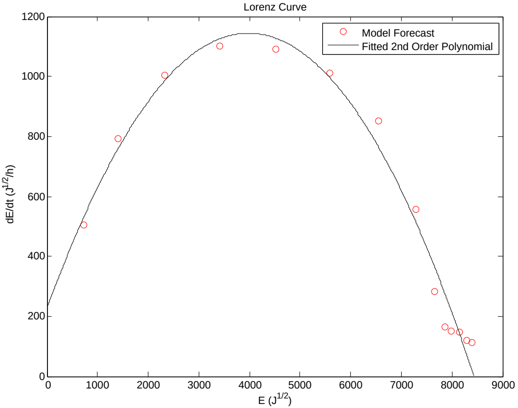


Lorenz Curve for Uncertainty with Simulated Observations

$$\frac{dE}{dt} = \alpha E \left(1 - \frac{E}{E_\infty}\right) = -\frac{\alpha}{E_\infty} E^2 + \alpha E \text{ (Lorenz 1982)}$$



Lorenz Curve for Real Observations



$$\frac{dE}{dt} = (\alpha E + \beta)\left(1 - \frac{E}{E_\infty}\right)$$ (Dalcher and Kalnay 1986)

Comparison of Estimated Parameters

Parameter	Sim. Obs.	Real Obs.	Sim. Obs. Sp.	Real Obs. Sp.
α	0.46	0.49	0.43	0.31
e^{α}	1.58	1.63	1.54	1.37
β	43	226	36	60
E_{∞}	6843	8437	6109	5558

- As expected, β and E_{∞} are larger when model errors are present
- The parameter α tends to be larger for the uncertainty than the spread (only slightly when model errors are not present, and by a large margin where model errors are present)
- When model errors are present, α is much smaller for the spread than the uncertainty (balance issues?)

Concluding Remarks

- We proposed a linear diagnostic tool to investigate the performance of the ensemble even at forecast times, where the evolution of the ensemble is highly nonlinear
- The Lorenz curve for the uncertainty and its time derivative provides a good approximation for the spatio-temporally averaged evolution of the magnitude of the forecast errors
- One can take advantage of the above properties when tuning such components of EnKF, as the representation of the effect of model uncertainty