A Toolkit to Predict the Utility of Ensemble Based Forecasts of High-Dimensional Chaotic Systems

Istvan Szunyogh

Texas A&M University
Department of Atmospheric Sciences

Acknowledgements

The talk is mainly based on the pair of papers:

- **Satterfield** and **Szunyogh:** Predictability of the Performance of an Ensemble Forecast System: Predictability of the Space of Uncertainties. *Mon. Wea. Rev.*, in press.
- **Satterfield** and **Szunyogh:** Predictability of the Performance of an Ensemble Forecast System II: Predictability of the Magnitude and the Spectrum of Uncertainties. (to be submitted).

Many thanks to my former UMD Colleagues: Brian Hunt, Eugenia Kalnay, Eric Kostelich, Ed Ott and Jim Yorke

The Problem

- One extremely useful property of the tangent space is that it is linear?
- In the ensemble based analysis and forecast techniques, the ensemble perturbations have finite amplitude. Is it justified to assume that the linear space they span, $\mathbb{S}$, provides a good representation of the analysis and forecast uncertainties?
- **Remark:** Most ensemble based techniques (e.g. Ensemble Kalman Filters, Post-processing Techniques, Targeted Observation Methods) do assume that $\mathbb{S}$ provides a good representation of the uncertainties.

Disclaimer

Do not expect Weierstrassian rigor!

Local State Vectors

- 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. (For the rest of this presentation we now drop the argument ℓ from the notation of the local state vectors)
- We assume that the error in the state estimate $\mathbf{x}^e$ (analysis or forecast)

$$\xi = \mathbf{x}^e - \mathbf{x}^t,$$

is a random variable; $\mathbf{x}^t$ is the model representation of the (in practice unknown) true state of the atmosphere. The covariance between the different components of ξ is described by the error covariance matrix $\mathbf{P}_\ell$.

Ensemble-based Estimate of the Covariance Matrix

- The K-member ensemble of local state estimates:

$$\{\mathbf{x}^{(k)}, k = 1 \ldots K\}$$

- The ensemble-based estimate of the covariance matrix:

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

- The ensemble perturbations:

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

- The ensemble mean:

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

The Vector Space of Ensemble Perturbations

- 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 $\hat{\mathbf{P}}_\ell$,

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

  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 Vectors

- An arbitrary local state vector $\mathbf{x}$ can be decomposed as

$$\mathbf{x} = \bar{\mathbf{x}} + \delta\mathbf{x},$$

  where $\delta\mathbf{x}$ is the difference between $\mathbf{x}$ and the ensemble mean $\bar{\mathbf{x}}$.
- The perturbation vector $\delta\mathbf{x}$ can be further decomposed as,

$$\delta\mathbf{x} = \delta\mathbf{x}^{(\|)} + \delta\mathbf{x}^{(\perp)},$$

  where $\delta\mathbf{x}^{(\|)}$ is the component that projects into $\mathbb{S}_\ell$
- The vector $\delta\mathbf{x}^{(\perp)}$ is the component of $\delta\mathbf{x}$ that has no projection on $\mathbb{S}_\ell$.

Decomposition of the Error in the State Estimate

- The error ξ in the state estimate $\mathbf{x}^e$ can be decomposed as

$$\xi = \mathbf{x}^e - \mathbf{x}^t = \delta\xi^{(\parallel)} + \delta\xi^{(\perp)},$$

where,

$$\delta\xi^{(\parallel)} = \sum_{k=1}^{K-1} \left[(\delta\mathbf{x}^e - \delta\mathbf{x}^t)^T \mathbf{u}_k \right] \mathbf{u}_k, \quad \delta\xi^{(\perp)} = \delta\mathbf{x}^{e(\perp)} - \delta\mathbf{x}^{t(\perp)}.$$

- Explained Variance:

$$EV = \frac{\|\delta\xi^{(\parallel)}\|}{\|\xi\|} = \frac{\|\delta\xi^{(\parallel)}\|}{\|\delta\xi^{(\parallel)} + \delta\xi^{(\perp)}\|}.$$

E-dimension and Relative Nonlinearity

- The E-Dimension (a measure of the steepness of the spectrum)

$$E = \frac{\left[\sum_{i=1}^{K} \sqrt{\lambda_i}\right]^2}{\sum_{i=1}^{K} \lambda_i},$$

  Originally introduced as BV-dimension in Patil et al. (2001).

- Local relative nonlinearity

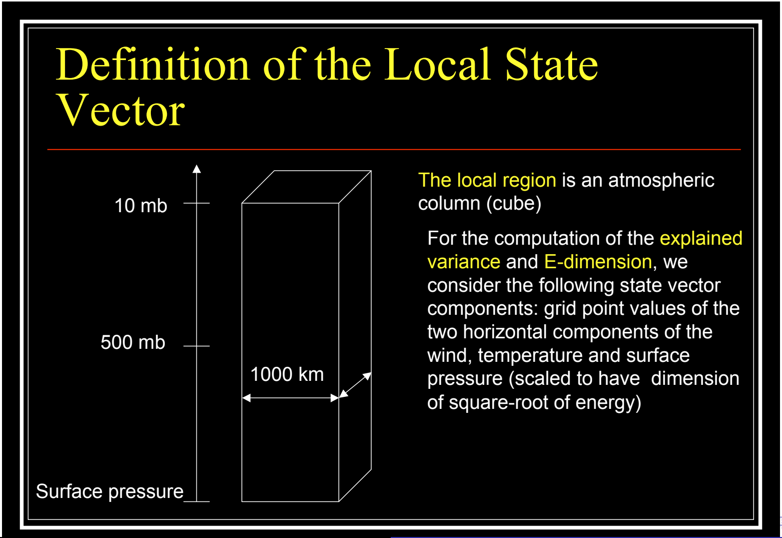
$$\rho_\ell = \frac{\|\bar{\mathbf{x}} - \mathbf{x}^e\|}{\frac{1}{K}\sum_{k=1}^{K} \|\delta \mathbf{x}^{(k)}\|},$$

  where $\mathbf{x}^e$ is forecast from the ensemble mean initial condition. Inspired by Gilmour et al. (2001)

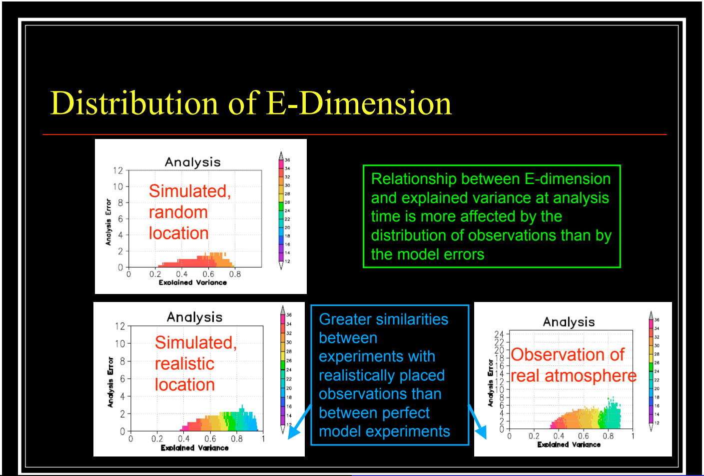
Analysis-Forecast System

- Data Assmilation: Local Ensemble Transform Kalman Filter with 40 ensemble members. (Ott et al. 2004; Szunyogh et al. 2005; Hunt et al., 2007; Szunyogh et al. 2008)
- Model: 2004 version of NCEP GFS at resolution T62 (about 150 km) and 28-levels
- Deterministic State Estimate ($\mathbf{x}^e$): forecast from ensemble mean analysis
- Error Statistics: Collected for 45 days (January and February 2004), all results shown are for NH extratropics
- Observations:
  - **Experiment 1:** Simulated observations at random locations (approximately uniformly distributed)
  - **Experiment 2:** Simulated observations at realistic locations
  - **Experiment 3:** Observations of the real atmosphere

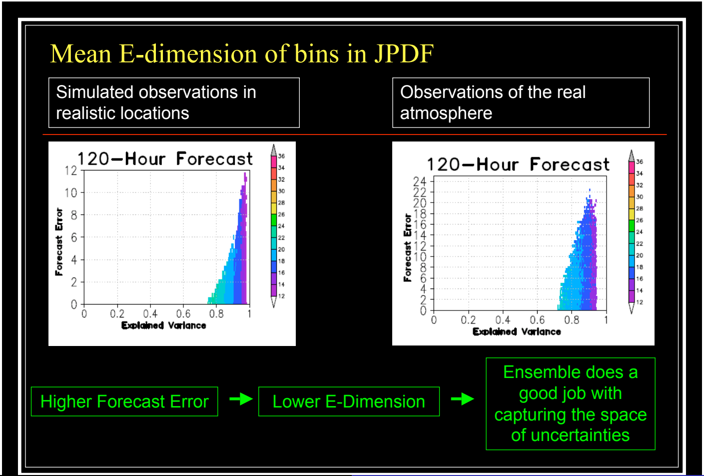
Definition of Local Volume



The Relationship Between Forecast Errors, Explained Variance and E-dimension: Analysis Time

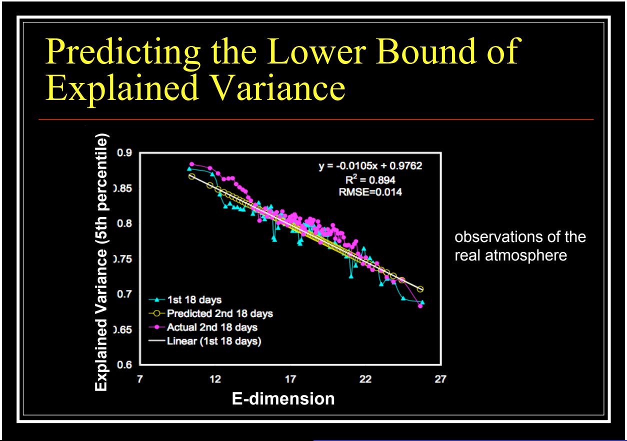


The Relationship Between Forecast Errors, Explained Variance and E-dimension: 120-hour forecast time



Predictability of the Lower Bound of the Explained Variance at 120-Hour Forecast Time

Experiment 3

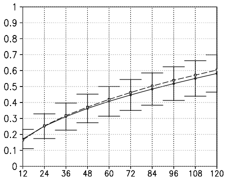


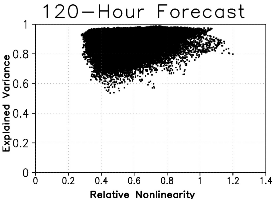
The Relationship Between Forecast Errors, Explained Variance and E-dimension: Summary

- The qualitative differences between the results with simulated observations and observations of the real atmosphere are surprisingly small.
- Larger errors lead to lower E-dimension and higher explained variance: faster error growth leads to a steeper spectrum, which improves the performance of the "small" ensemble in capturing the errors. These features are present at both analysis and the later forecast times, but are much more pronounced at longer forecast times.
- How is this possible? One would expect that finite amplitude perturbations would behave more linearly at shorter lead times.

Relative nonlinearity: Experiment 3

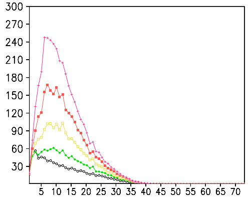
The relative nonlinearity clearly indicates that linearity is breaking down with increasing forecast time: (Relative nonlinearity vs. forecast time [hr])

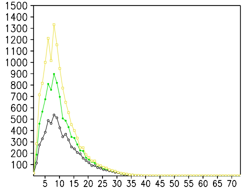




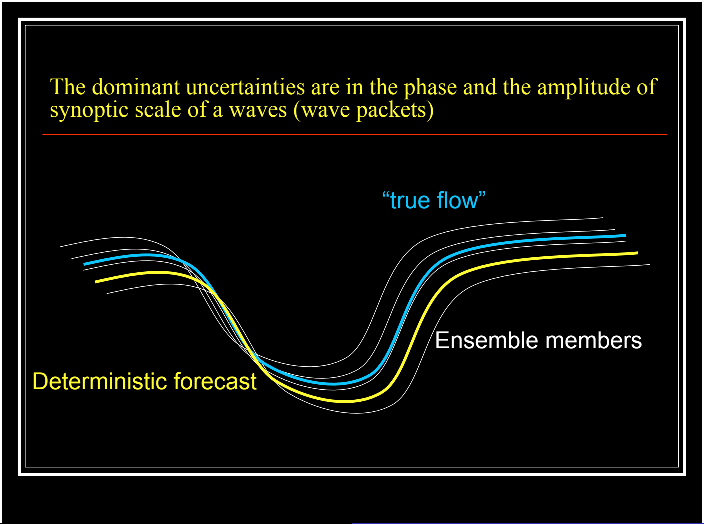
Spectral Analysis: Power Spectrum of Meridional Wind at 500 hPa in Experiment 3

Power spectrum every 12 hours between 00-hr and 48-hr forecast times and between 72-hr and 120-hr forecast times:



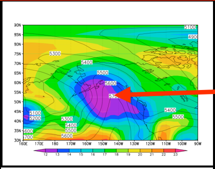


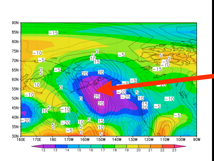
Schematic Illustration in Physical Space



An Example: Case of Largest Error in Experiment 3

An Example: Part I

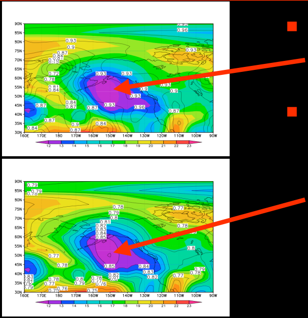




- Local low-dimensional area coincides with a ridge over the NE Pacific
- The largest local error in the 5-day meridional wind forecast at 500 hPa coincides with the lowest dimension

An Example: Case of Largest Error in Experiment 3

An Example: Part II



- Local low-dimensional area coincides with are of highest explained variance
- Linear regression provides a good prediction of the location of highest explained variance and the quantitative value of explained variance

A Couple of Remarks on Scale Interactions

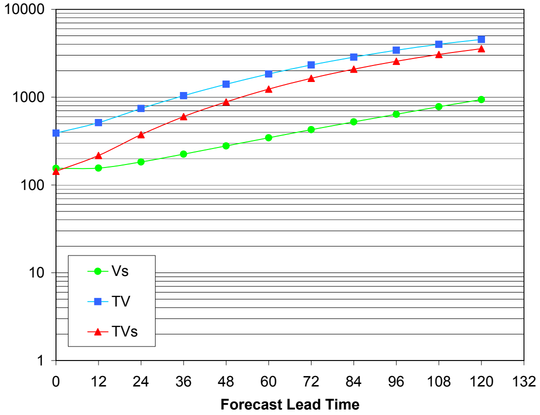
- Our findings on the spectral behavior of the uncertainties are in good agreement with those of Tribbia and Baumhefner, 2004: Scale Interactions and Atmospheric Predictability: An Updates Perspective. *Mon. Wea. Rev.*, **132,** 703-713: *"...error growth eventually asymptotes to an exponential growth of baroclinically active scales."*
- The baroclinically active scales sit on top of the inertial range of 2-D turbulence, characterized by an inverse cascade of energy.

Predictability of the Magnitude of the Error

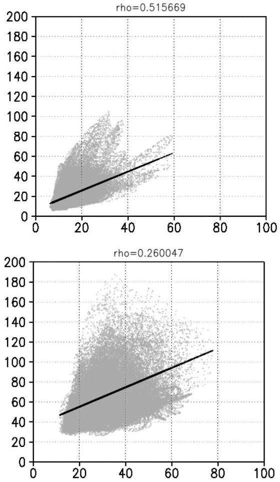
Diagnostic quantities

- Ensemble Variance: $V_S = trace(\hat{\mathbf{P}}_\ell)$
- Forecast Error Variance: $TV = E[\delta^2 \mathbf{x}^t]$ (expected value of the difference between the ensemble mean and the truth)
- For a perfect ensemble $E[V_S] = E[TV]$
- Projection of Forecast Error Variance on $\mathbb{S}_l$: $TV_S = E[\delta^2 \mathbf{x}^{t(\parallel)}]$

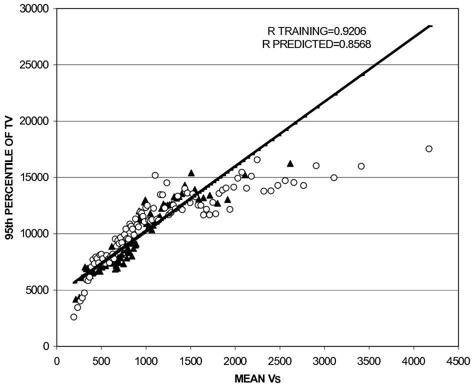
Results for Experiment 3



$\delta^2 \mathbf{x}^t$ vs. V_S, Experiments 1 and 2, Not the right relationship to look at!



95th Percentile of $\delta^2\mathbf{x}^t$ vs. V_S



Concluding Remarks

- A diagnostic not discussed today is the spectrum of d-ratio: a diagnostic to measure the performance of the ensemble in distinguishing between the importance of the different directions within $\mathbb{S}_\ell$
- While the linear space $\mathbb{S}_\ell$ captures the dominant errors after a few days, nonlinear effects play a very important role in the evolution of the errors. Lyapunov terminology should not be evoked to explain these results.
- Our results are good news for linear post-processing techniques.
- We found a couple of interesting predictive linear relationships. Many more may exist.
- Results would probably be very different using variables dominated by smaller scale variability (e.g., precipitation)