# Flow-Dependence of the Performance of an Ensemble-Based Analysis-Forecast System

**Istvan Szunyogh**

**University of Maryland, College Park**

Institute for Physical Science and Technology

& Department of Atmospheric and Oceanic Science

Mathematical Advancement in Geophysical Data Assimilation, BIRS, Banff, Canada, February 3-8, 2008

## Acknowledgement

- Special thanks for many of the results in this talk to
  - David Kuhl
  - Elizabeth Satterfield
- and for inspiring discussions, algorithmic and code development to
  - Eric Kostelich, Gyorgyi Gyarmati, Brian Hunt, Eugenia Kalnay, Edward Ott, Jim Yorke, Michael Oczkowski, Elana Fertig, DJ Patil, Aleksey Zimin,

## Outline

- A conceptual mathematical framework to study the dynamics of the atmosphere (ocean, planetary atmospheres, etc.)
- Applications to data assimilation and predictability with the model component of the NCEP GFS at T62L28 resolution

## The Challenge

- Mathematical foundation of tools to study the asymptotic behavior of low-dimensional dynamical (physical) systems is solid
  - The original equations derived from first principles of physics does not have to be low dimensional, but there must exist a low-dimensional underlying system
  - Most rigorous mathematical results are summarized in an influential paper by Eckmann and Ruelle (1985), which was introduced to the atmospheric science literature by Legras and Vautard (1996)
- The systems we study are inherently high-dimensional:

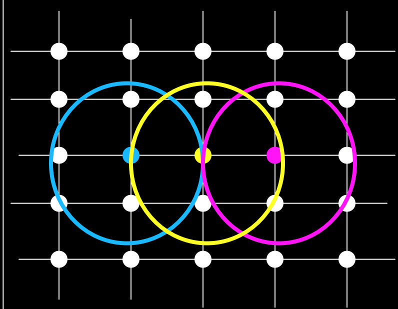
# Some concepts borrowed from low-dimensional chaos

- Differentiable dynamics (tangent space, mapping between tangent spaces)
- Dimensions (number of excited degrees of freedom)
- Invariant Manifolds (e.g., Unstable Manifold)
- Entropy (Production of Information)
- Characteristic exponents (Sensitivity to initial uncertainty)
- Problem: We often use this terminology to motivate our arguments, but is there a way to introduce similar concepts in a more formal way to our high-dimensional systems?

# DESCLAIMER!!!

- Do not expect Weierstrassian rigor from this talk
  - I am an atmospheric scientist
  - I do not believe that rigorous mathematics is available: frameworks exist to solve problems, but these frameworks are often motivated by a mixture of intuition and results for low-dimensional systems
  - To put it into context, it took two centuries for some of the greats of mathematics to get from Newton and Leibnitz to Weierstrass (and some serious beer drinking and sword fighting by Weierstrass before he was ready to start developing his rigorous approach to calculus at the age of 30)

# One Potential Approach
illustration for a 2D model grid



Local state vector: components of the global state vector in the local region

- Given is an ensemble of global state vectors
- A local region is assigned to each grid point
- Local ensemble perturbations are defined

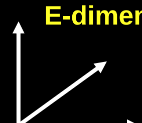
- Collection of local ensemble perturbations provide a high-dimensional estimate of the tangent space based on a small ensemble
- Linearity can be valid for longer times in local regions

# E-dimension: a measure of complexity in the local region

- E-dimension: A measure of the steepness of the spectrum of the ensemble-based error covariance matrix in the local region
- The smaller the E-dimension the steeper the spectrum (introduced in Patil et al. 2001, *PRL*; discussed in details an illustrated on complex meteorological examples in Oczkowski et al., 2005, *JAS*)

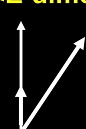
**Three orthogonal perturbations**

**E-dimension=3**



**All three perturbations in one plane**

**1<E-dimension<2**



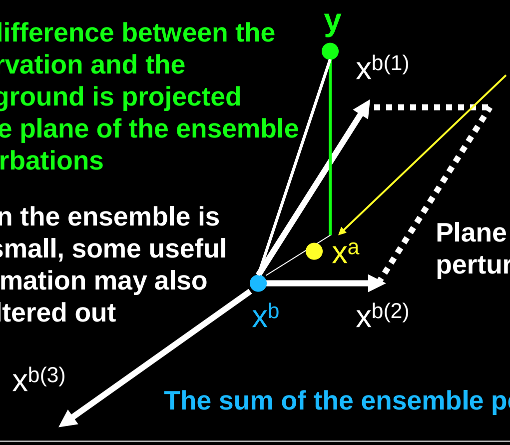
**E-dimension=1**

# Motivated the LETKF

3d state space, 3-member ensemble on a plane

The difference between the observation and the background is projected on the plane of the ensemble perturbations

When the ensemble is too small, some useful information may also be filtered out




$x^b$-$x^a$ is obtained in the plane of the ensemble perturbations: potentially an efficient filter of observational noise

Plane of the ensemble perturbations

The sum of the ensemble perturbations is zero

# Remarks on LETKF

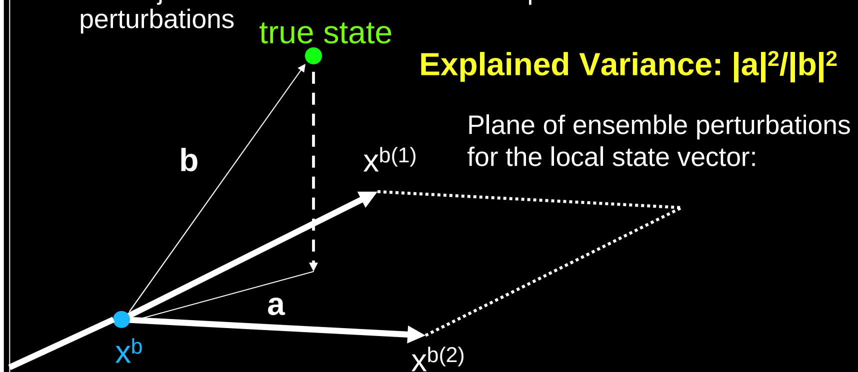
- The local approach motivated the development of the LETKF, but in the current formulation of the algorithm the definition of local regions is not a formal requirement
- Most importantly, H(x) computed globally and any observation can be chosen to affect the analysis of any state vector component

# Experimental design
of Szunyogh et al. 2005 (Tellus A)

- Observations: Noisy observations of a time series of true states (generated by a long model integration), full vertical soundings are located at randomly selected model grid point location (10% coverage for the results shown here, but the scheme is still stable at 2.5% coverage)
- Data Assimilation: LETKF with 40 ensemble members
- Model: NCEP GFS at resolution T62 (about 150 km) and 28-levels
- Error Statistic collected for 45 days (January-February)

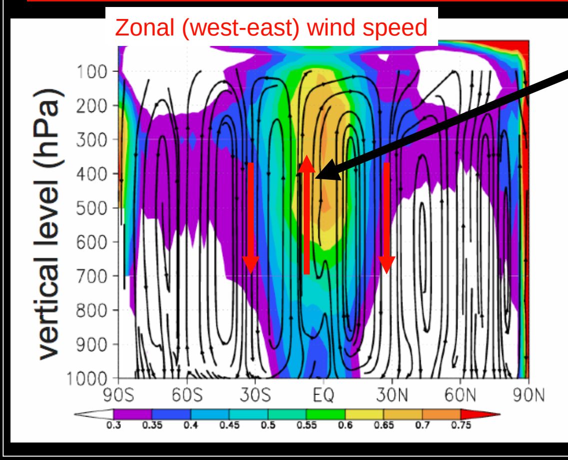
# Explained Variance: a measure of ensemble performance in the local region

- b: True error
- a: Projection of the true error on the space of the ensemble perturbations



# Vertical Distribution of RMS Error

averaged over time and along latitudes



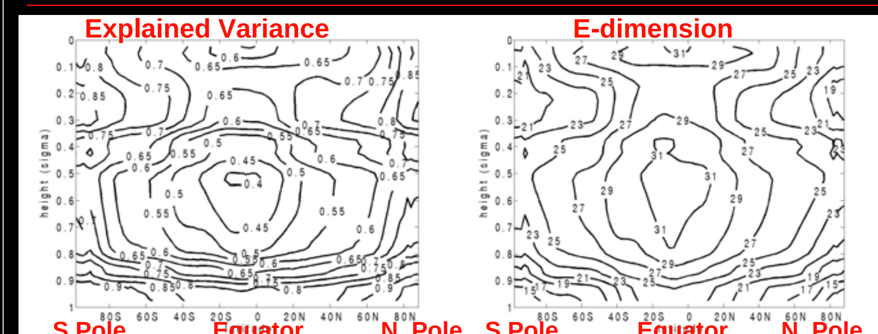

**The error is the largest in the region of upward motions in the Tropics (parameterized deep convection)**

Reminder: the model is perfect, observation coverage homogeneous!!!

Differences are due to differences in the dynamics

# Relationship Between Explained Variance and E-dimension: Correlation:-0.93

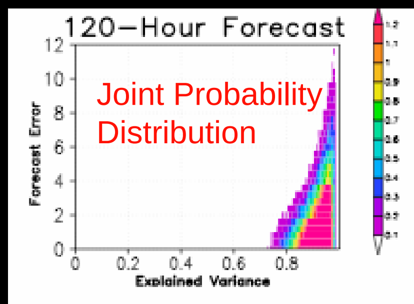
averaged in time and along latitudes

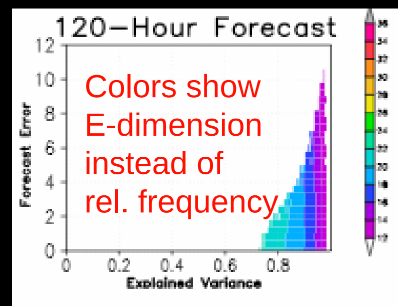



**When # of ensemble members >20, the explained variance changes little in time and the filter remains stable ("unstable" manifold is well captured), beyond 40, the improvement is small**

# Predictability of Predictability

Kuhl et al. (2007 JAS)




Rapid Error Growth ⟶ Low E-dimension ⟶ Good Representation of Uncertainties

Low predictability ⟶ High Predictability of Predictability

# Main Conclusion of the Study

Lower E-dimension ⟵ Fast Error Growth is typically confined to few phase space directions

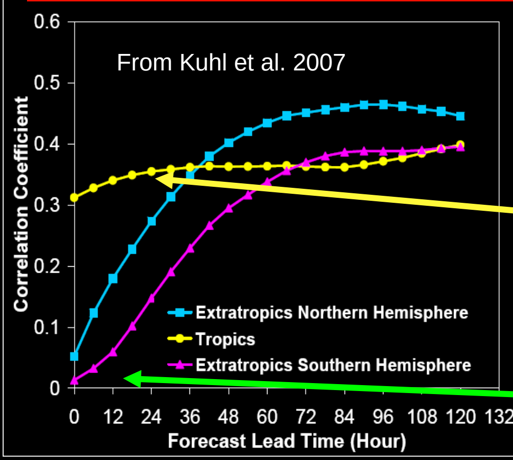
↓

Higher Explained Variance

↓ Analysis expects the right background errors and few observations can make a big correction

Lower analysis error

## Spread-Skill Correlation for randomly distributed simulated vertical soundings




Spread: Measured by the ensemble standard deviation

Skill: Absolute error of the ensemble mean forecast

Insufficient data coverage to suppress small scale errors in the Tropics at analysis time

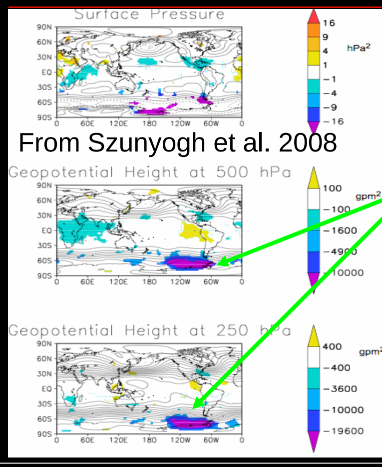
In the extratropics, 10% observational coverage is sufficient to remove errors with well defined structures at analysis time

## Experiments with Observations of the Real Atmosphere

- Observations of the real atmosphere, except for radiances (Szunyogh, Kostelich, Gyarmati et al. 2007, Tellus, in press)
- The LETKF and the Benchmark SSI system use different **H** operators; the one used with the LETKF is less sophisticated.
- Benchmark SSI analyses and forecasts provided by NCEP (Y. Song and Z. Toth)
- 60-member ensemble

# Comparison of the LETKF and the SSI

48-hour forecasts with real observations (no radiances)



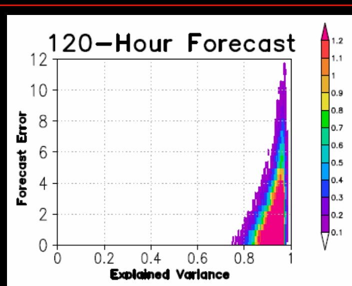

From Szunyogh et al. 2008

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

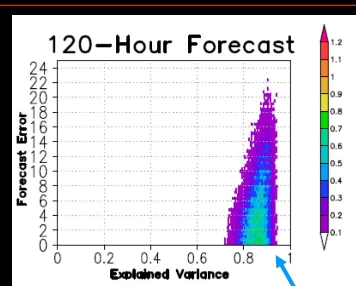
Results are shown only where The difference is statistically Significant at the 99% level

# Joint Probability Distribution Function (JPDF) for Explained Variance and Forecast Error

Simulated observations in realistic locations




Observations of the real atmosphere

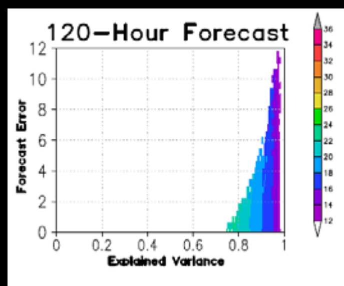
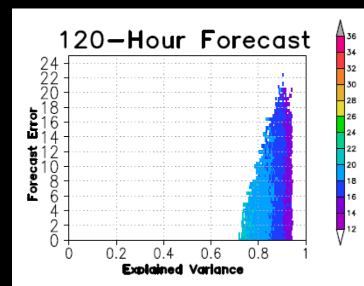



For real atmosphere, explained variance never reaches 1

For both perfect model and real atmosphere:

High Forecast Error → Increased Likelihood that Explained Variance is High

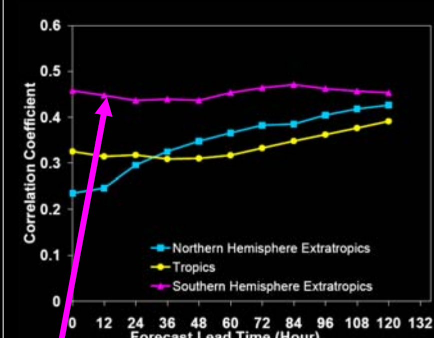
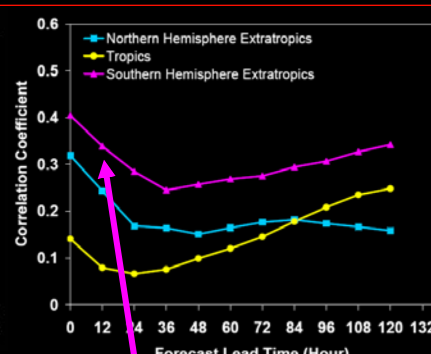
# Mean E-dimension of bins in JPDF

| Simulated observations in realistic locations | Observations of the real atmosphere |
|---|---|
|  |  |



Higher Forecast Error → Lower E-Dimension → Ensemble does a good job of capturing the space of uncertainties

# Spread-Skill Correlation

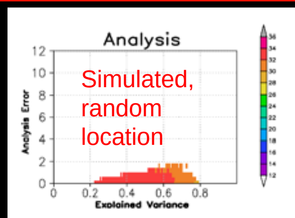
| Simulated observations in realistic locations | Observations of the real atmosphere |
|---|---|
|  |  |



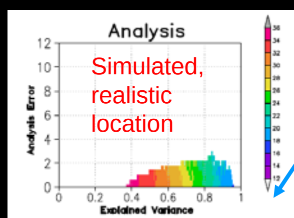
Data coverage is not sufficient to remove all errors correctly identified by the ensemble

Model errors have little impact on initially high correlations in SH XT

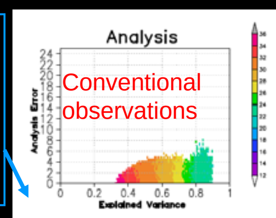
# Distribution of E-Dimension




Relationship between E-dimension and explained variance at analysis time is more affected by the distribution of observations than by the model errors




Greater similarities between experiments with realistically placed observations than between perfect model experiments




# Conclusions

- Introducing local state vectors may be a way to introduce formal tools to study high dimensional systems
- Applying simple diagnostics to the local state vectors, we were able to explain some aspects of the behavior of an ensemble based analysis-forecast system
- Our results suggest that the performance of the ensemble (both in analysis and forecast mode) is strongly flow dependent
- Fortunately, the ensemble performs best when it is the most important, in cases of fast error growth
- All papers available at http://weatherchaos.umd.edu