ESTIMATING THE STATE OF SPATIOTEMPORAL CHAOS: WEATHER FORECASTING ETC.

Edward Ott
University of Maryland

References:

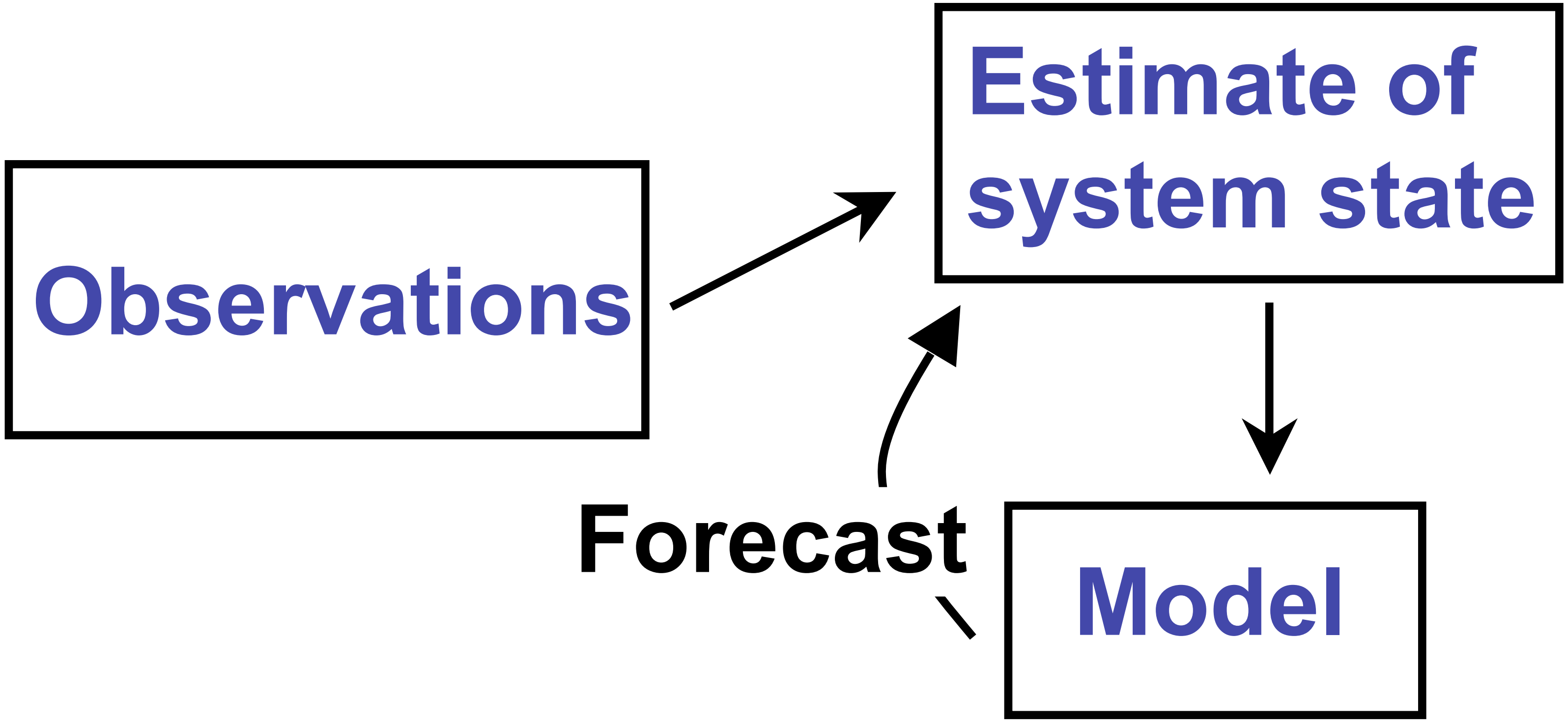
E. OTT, B. HUNT, I. SZUNYOGH, A.V.ZIMIN, E.KOSTELICH, M.CORAZZA, E. KALNAY, D.J. PATIL, & J. YORKE, TELLUS A (2004).

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

OUTLINE

- Review of some basic aspects of weather forecasting.
- Our method in brief.
- Tests of our method.

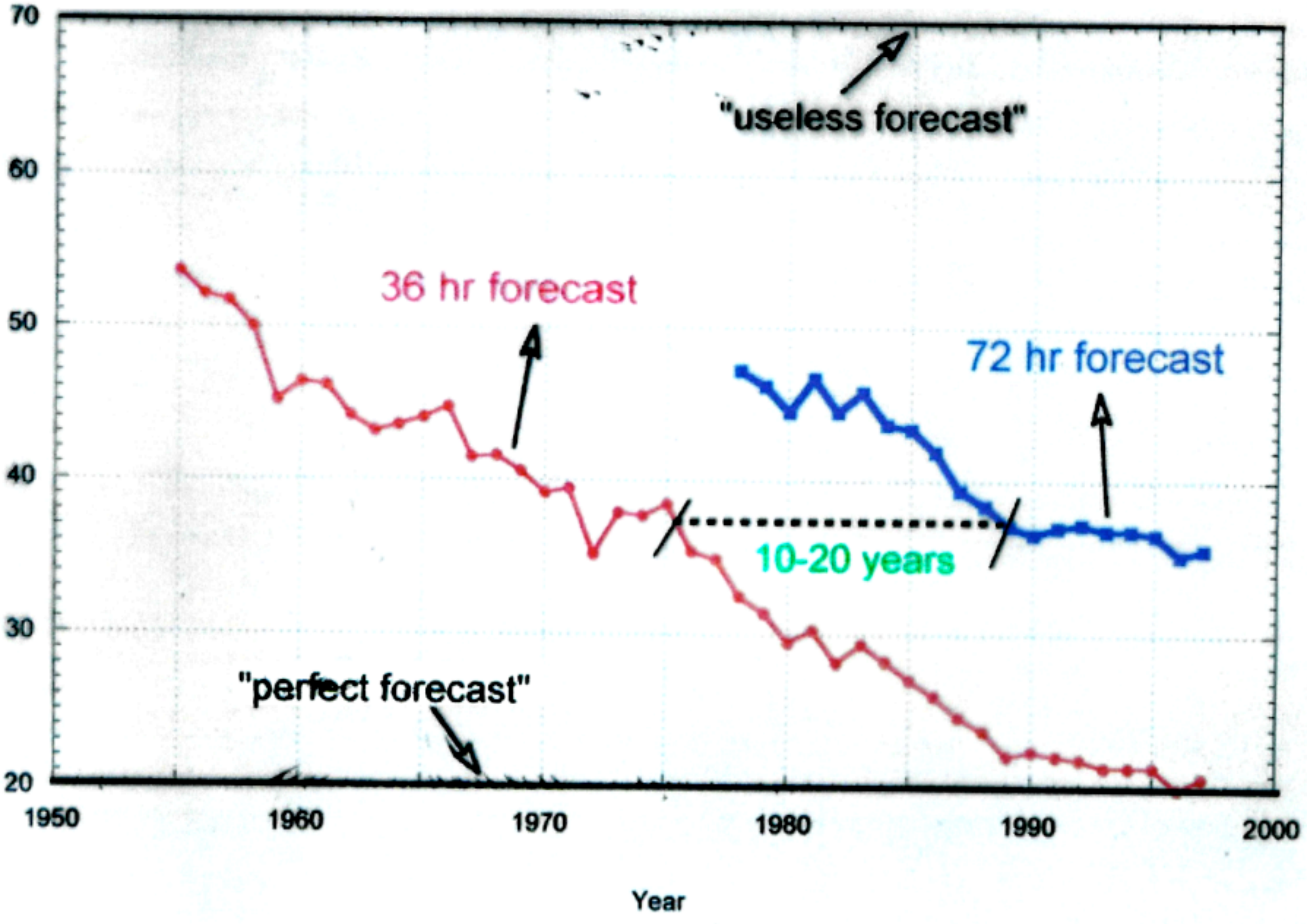
THE THREE COMPONENTS OF STATE ESTIMATION & FORECASTING

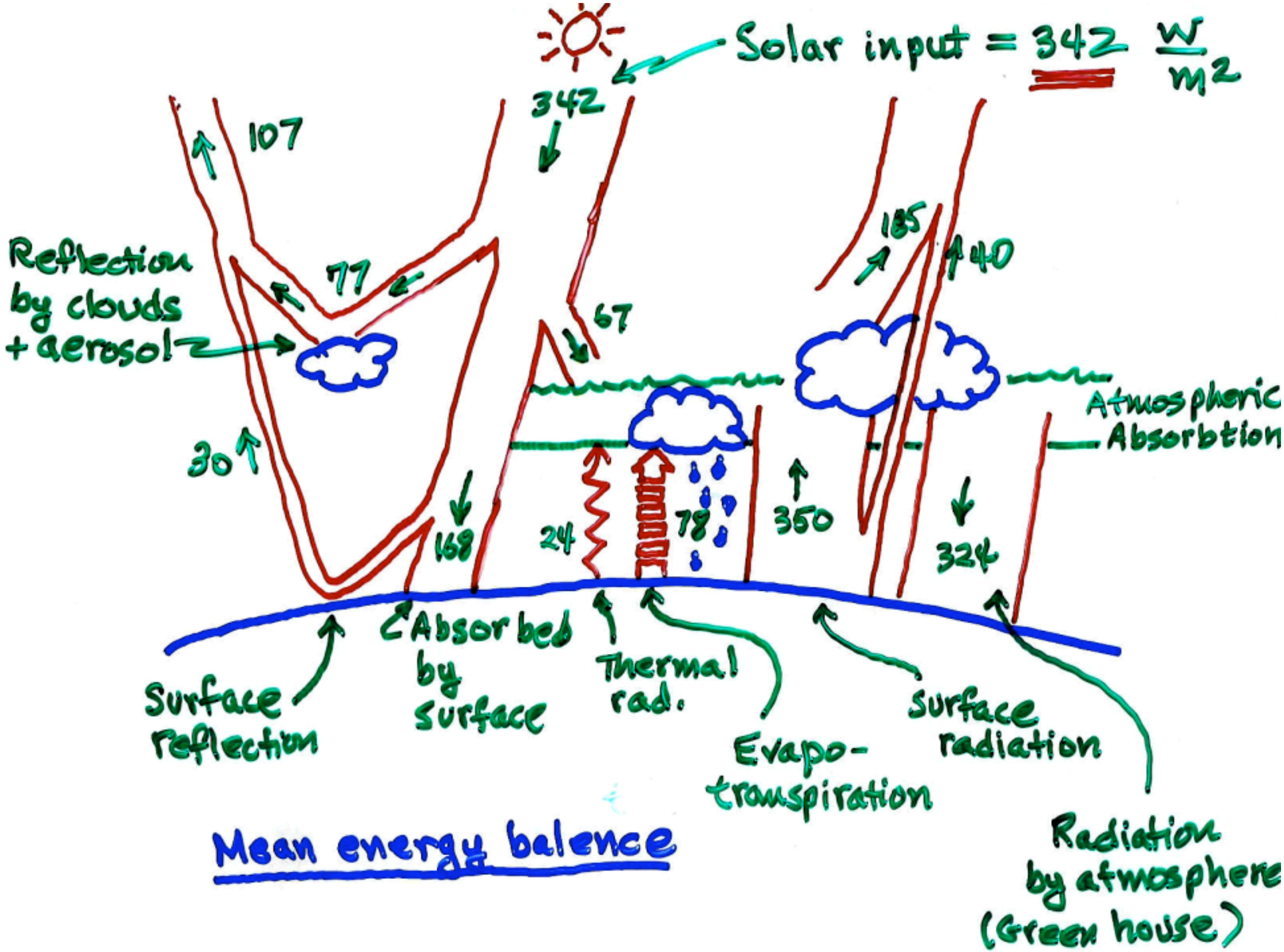


'Components' of this process:

- Observing
- Data Assimilation
- Model Evolution

NMC/NCEP operational S1 scores over North America (500 hPa)



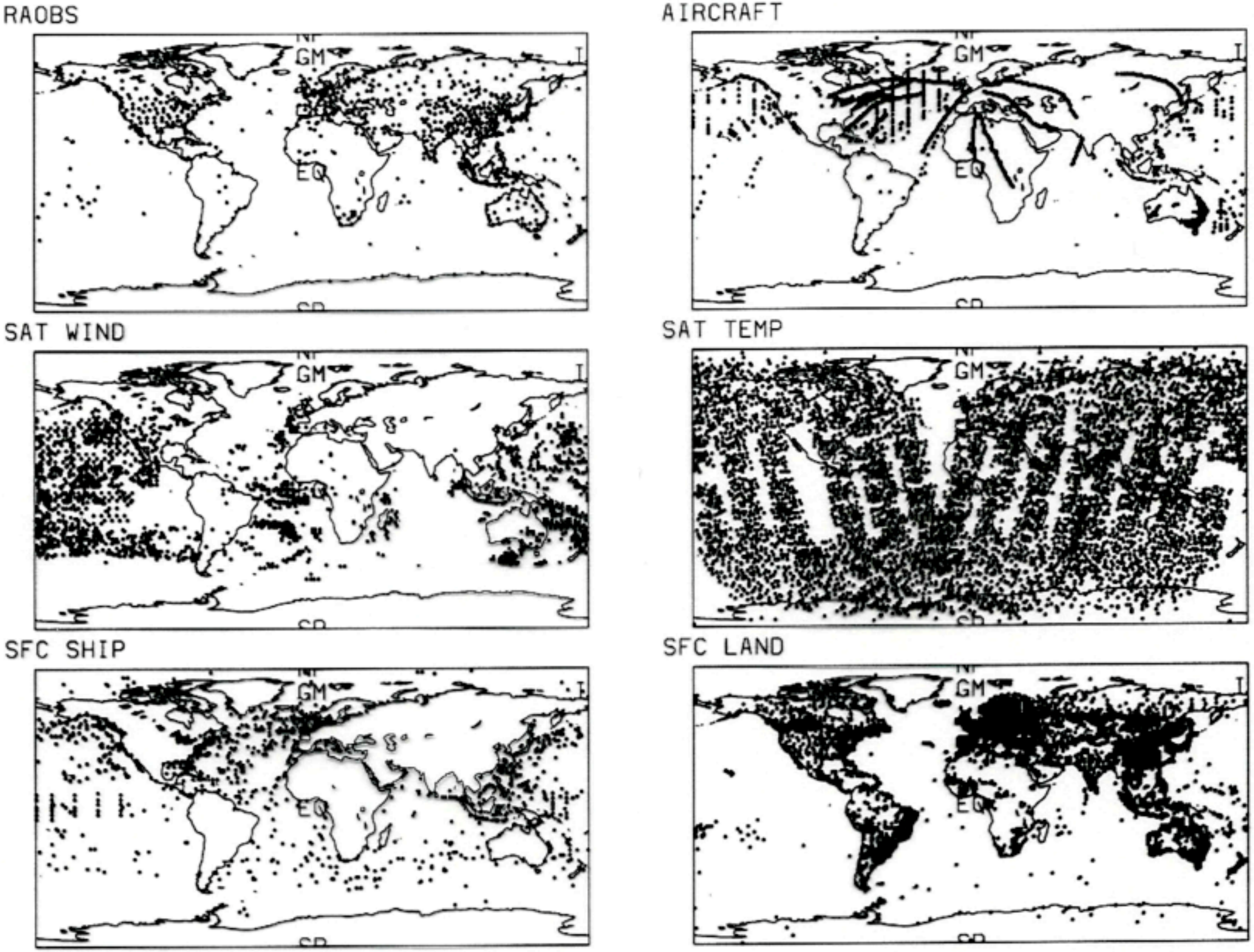


Mean energy balence

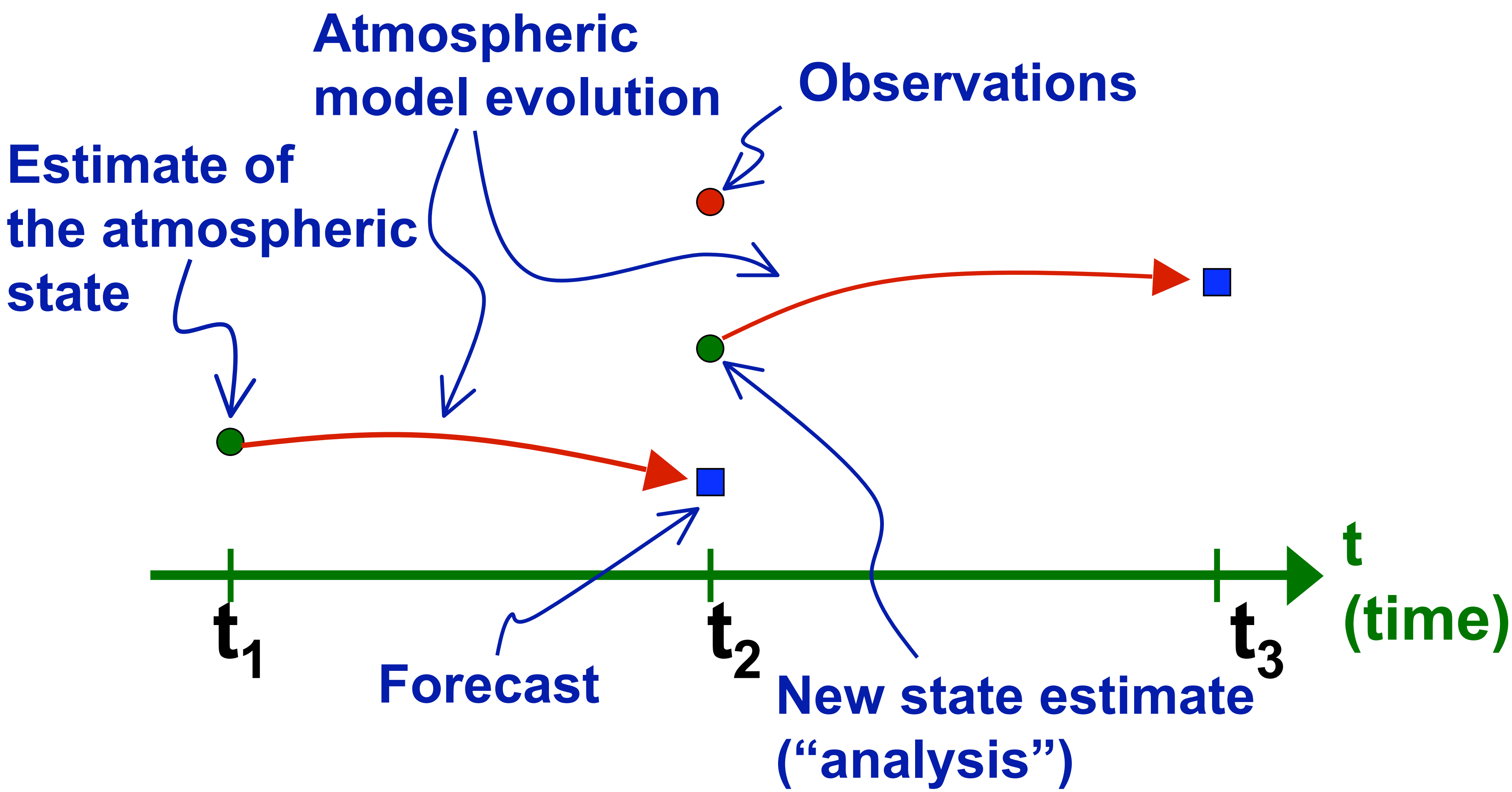
FACTORS INFLUENCING WEATHER

Changes in solar input
Ocean-air interaction
Air-ice coupling
Precipitation
Evaporation
Clouds
Forests
Mountains
Deserts
Subgrid scale modeling
Etc.

DATA DISTRIBUTION Ø1SEP97ØØZ-Ø1SEP97ØØZ

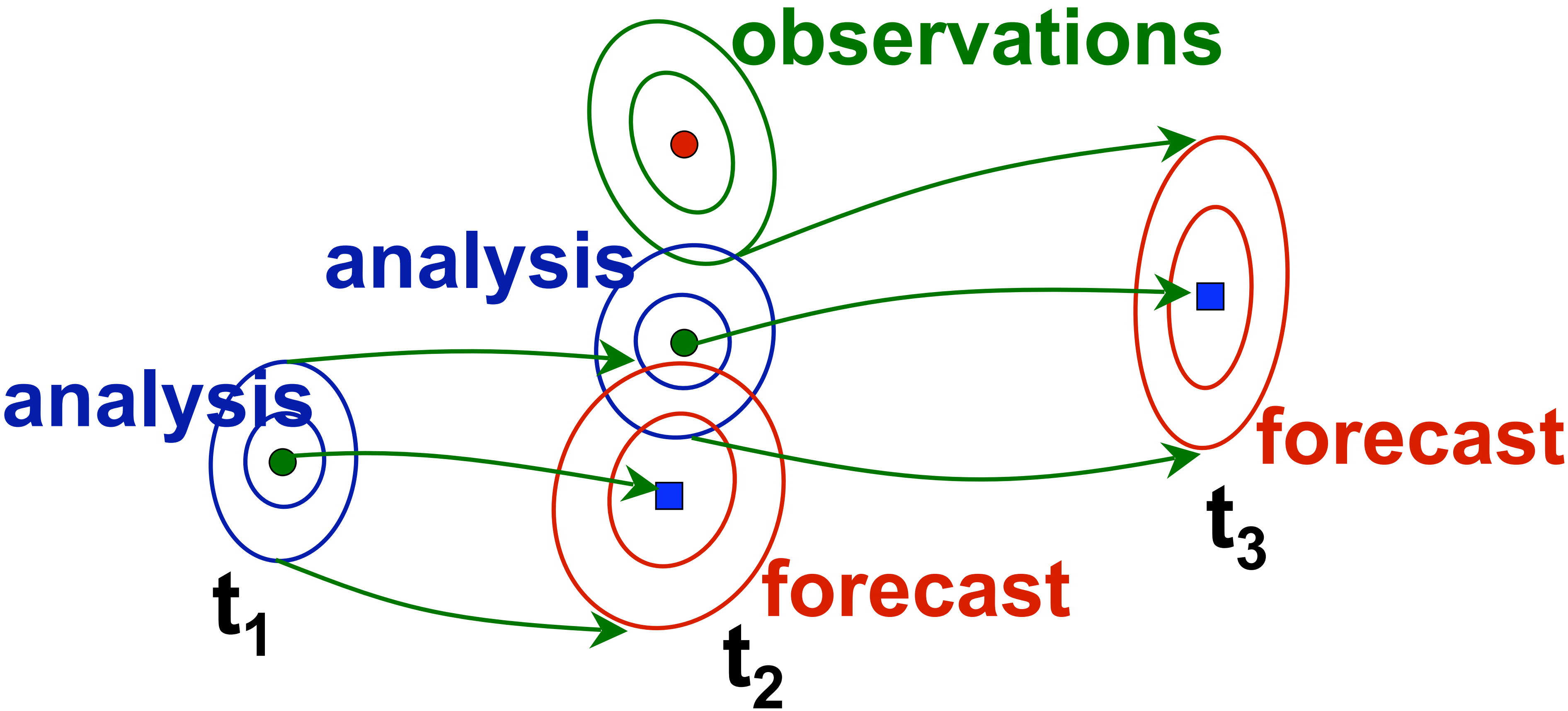


DATA ASSIMILATION



- Obs. are scattered in **location** and have errors.
- Forecasts (as we all know) have uncertainties.

A MORE REFINED SCENARIO



<u>Note</u>: Analysis pdf at t_1 is dynamically evolved to obtain the forecast pdf at t_2.

GOALS OF DATA ASSIMILATION

- Determine the most likely current system state & pdf given:
  (a) a model for the system dynamics,
  (b) observations.

- Use this info (the "analysis") to forecast the most likely system state and its uncertainty (i.e., obtain the forecast pdf).

KALMAN FILTER

- **For the case of linear dynamics, all pdfs are Gaussian, and there is a known rigorous solution to the state estimation problem: the Kalman filter.**

  (pdf of obs.) + (pdf of forecast) → (pdf of state)

- **In the nonlinear case one can often still approximate the pdfs as Gaussian, and, in principle, the Kalman filter could then be applied.**

- **A key input is the forecast pdf.**

DETERMINING THE ANALYSIS PDF $F_a(\underline{x})$

$\Phi_f(\underline{x})$ = **forecasted state PDF**

$\Phi_{obs}(\underline{y}|\underline{x})$ = **PDF of expected obs. given true system state $\underline{x}$**

- **Bayes' theorem:** $\Phi_a(\underline{x}) = \dfrac{F_f(\underline{x})\, F_{obs}(\underline{y}|\underline{x})}{\oint d\underline{x}\, F_f F_{obs}}$
- **Assume Gaussian statistics:**

$$\Phi_f(\underline{x}) \sim \exp\left\{-\frac{1}{2}(\underline{x}-\underline{x}_f)^T \underline{\underline{R}}_f^{-1} (\underline{x}-\underline{x}_f)\right\}$$

$$\Phi_{obs}(\underline{y}|\underline{x}) \sim \exp\left\{-\frac{1}{2}(\underline{y}-\underline{\underline{H}}\,\underline{x})^T \underline{\underline{R}}_{obs}^{-1} (\underline{y}-\underline{\underline{H}}\,\underline{x})\right\}$$

- **Analysis PDF:**

$$p_a(\underline{x}) \sim \exp\left\{-\frac{1}{2}(\underline{x}-\underline{x}_a)^T (\underline{\underline{P}}_a)^{-1} (\underline{x}-\underline{x}_a)\right\}$$

$$\underline{\underline{P}}_a = \left[\underline{\underline{P}}_f^{-1} + \underline{\underline{H}}^T \underline{\underline{P}}_{obs}^{-1} \underline{\underline{H}}\right]^{-1}$$

$$\underline{x}_a = \underline{x}_f + \underline{\underline{P}}_a \underline{\underline{H}}^T \underline{\underline{P}}_{obs} (\underline{y} - \underline{\underline{H}}\,\underline{x}_f)$$

CURRENT NCEP OPERATIONAL APPROACH (3DVAR)

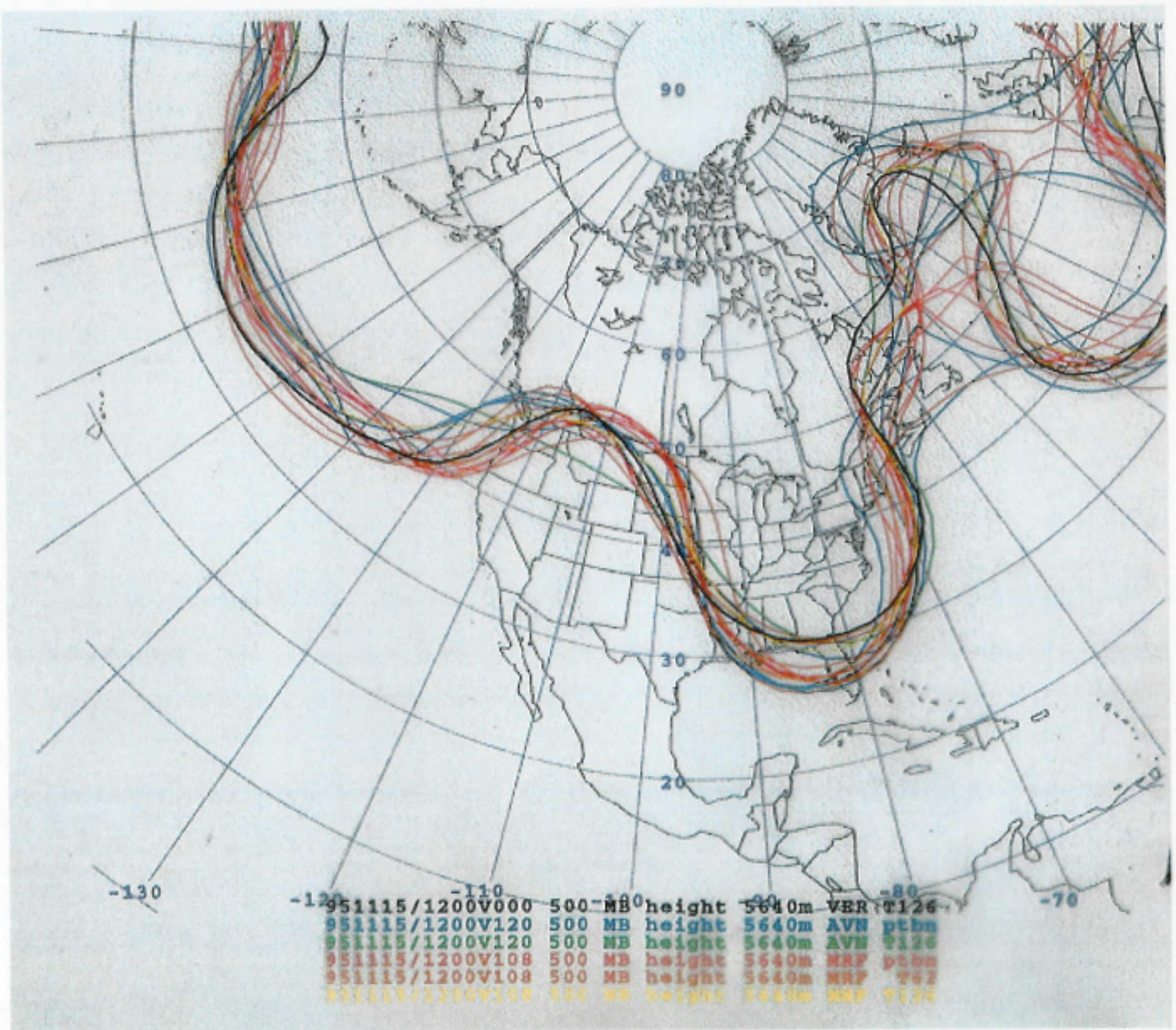
- A constant, time-independent forecast error covariance, $\underline{\underline{P}}_f$, is assumed.

  forecast pdf ~ $\exp\{-\frac{1}{2}(\underline{x}-\underline{x}_f)^T(\underline{\underline{P}}_f)^{-1}(\underline{x}-\underline{x}_f)\}$

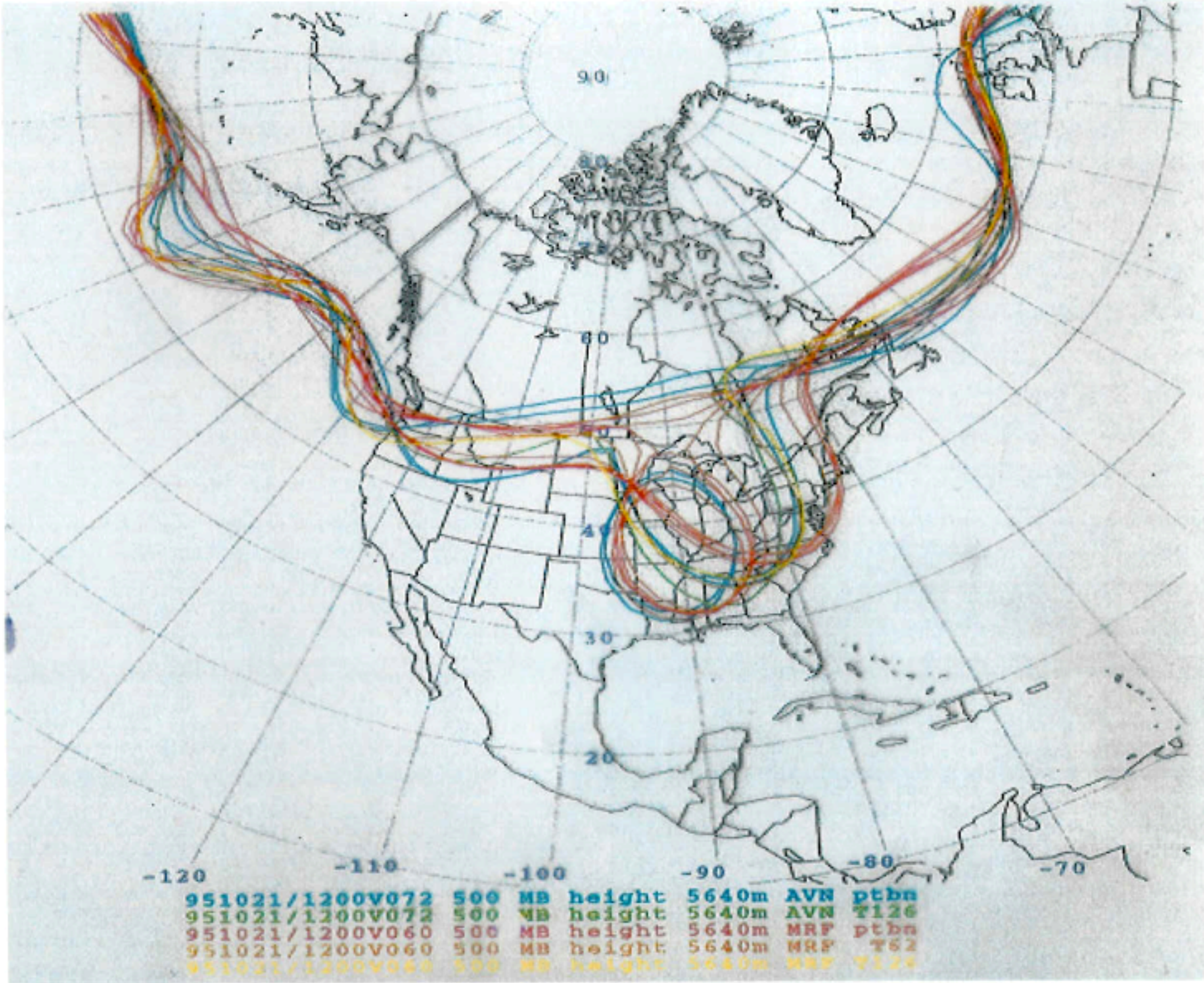
- The Kalman filter equations for the system state pdf are then applied treating the assumed $\underline{\underline{P}}_f$ as if it were correct.

  ECMWF : 4DVAR

These approaches ignore the time variability of $\underline{\underline{P}}_f$.



17 5 DAY FORECASTS



17 2.5 DAY FORECASTS

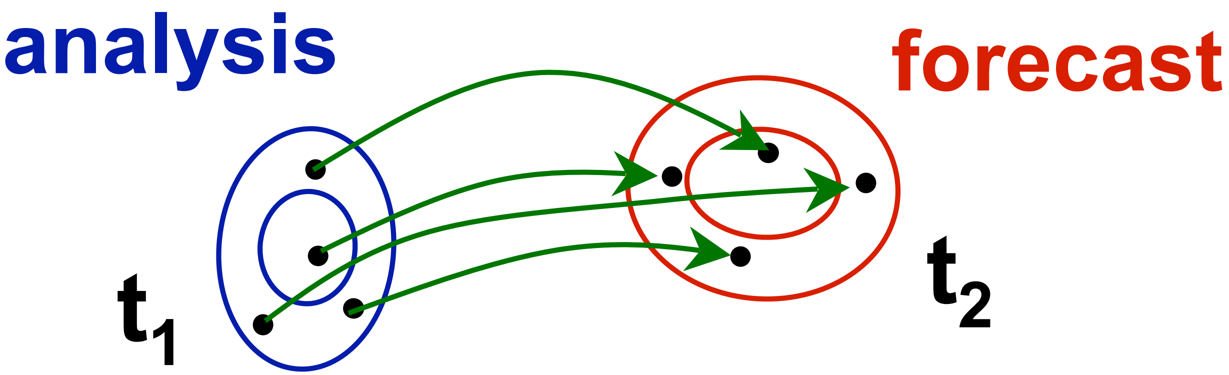
PROBLEM

- Currently data assimilation is already a very computationally costly part of operational numerical weather prediction.
- Implementation of a full Kalman filter would be many many times more costly, and is impractical for the foreseeable future.

REDUCED KALMAN FILTERS

- We seek a practical method that accounts for dynamical evolution of atmospheric forecast uncertainties at relatively low computational cost.

- Ensemble Kalman filters:



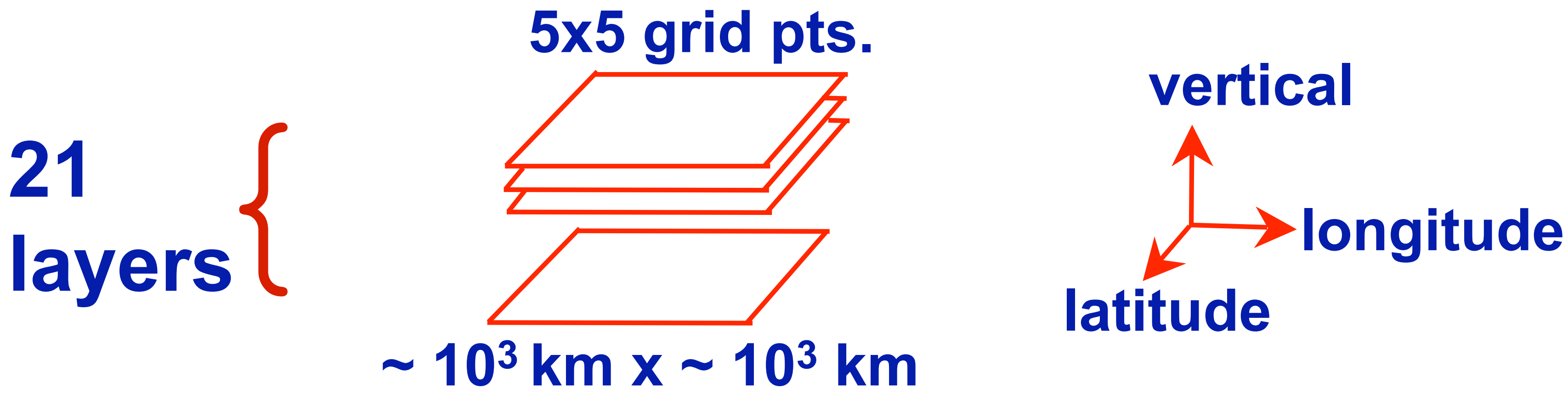
Evansen, 1994
Bishop et. al., 2001
Whitaker and Hamill, 2002

Houtekamer & Mitchell 1998, 2001
Hamill et. al., 2001
Anderson, 2002

MOTIVATION FOR OUR METHOD

- Patil et al. (Phys. Rev. Lett. 2001)

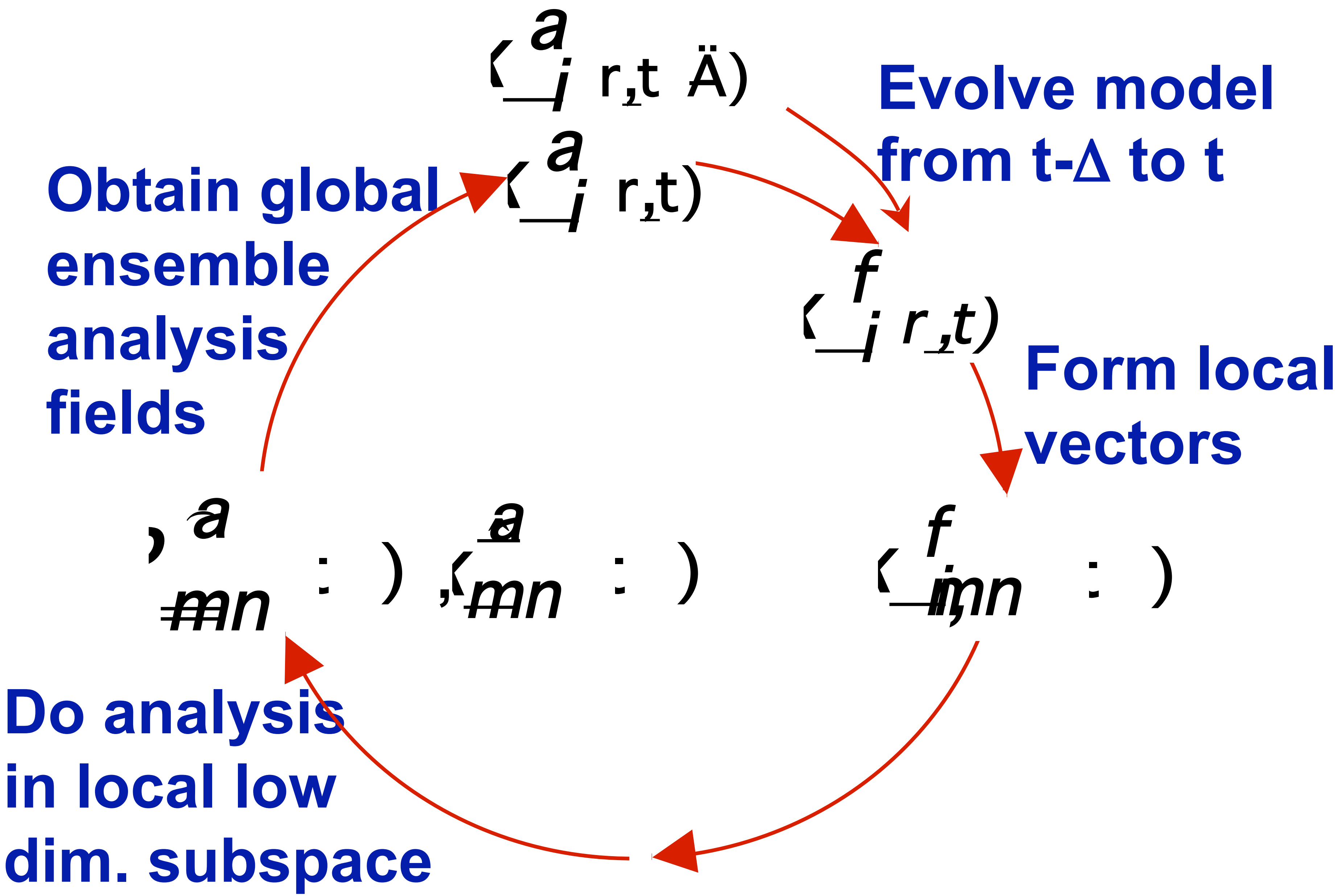
'Local Region' labeled by its central grid point



- It was found that in each local region the ensemble members approximately tend to lie in a surprisingly low dimensional subspace.

→ Take the estimated state in the local region to lie in this subspace.

SUMMARY OF STEPS IN OUR METHOD



PROPERTIES OF OUR METHOD

- Only operations on relatively small matrices are needed in the analyses. (We work in the local low dimensional subspaces.)
- The analyses in each local region are independent.

  → Fast parallel computations are possible.

TESTING OUR METHOD ON A ‘TOY’ MODEL

- “Truth run”: Run the model obtaining the true time series:

  $$\underline{x}^{true}(p, t_n) \quad (p = \text{grid point})$$

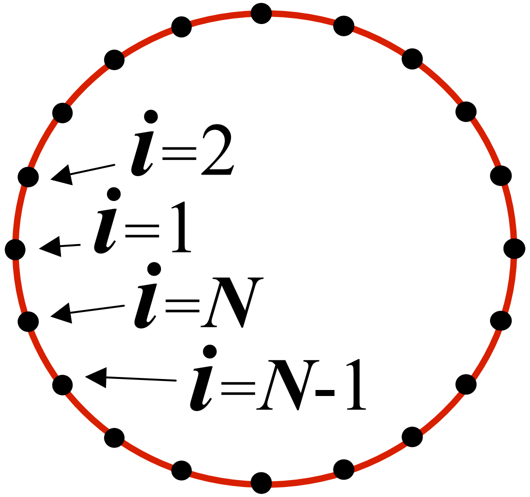
- Simulate obs.: $\underline{y}(p, t_n) = \underline{x}^{true}(p, t_n) + (\text{noise})$

  for some set of observing locations, p.

- Run our ‘local ensemble Kalman filter’ (LEKF) using the same model (perfect model scenario) and these observations to estimate the most probable state and pdf at each analysis time.
- Compare the estimated most probable system state with the true state.

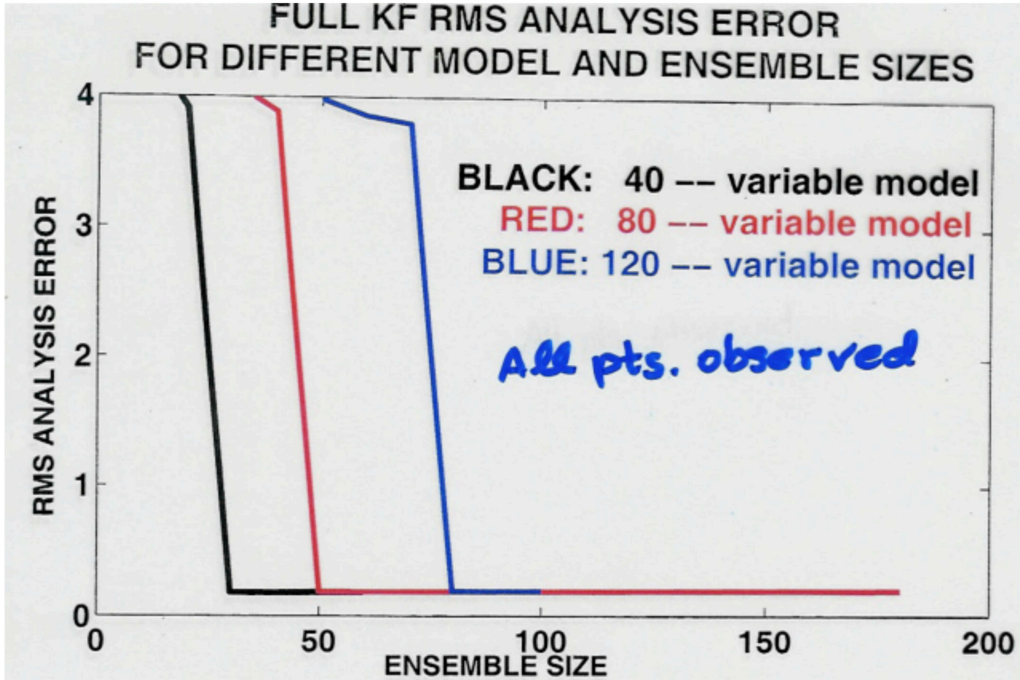
NUMERICAL EXPS. WITH A TOY MODEL

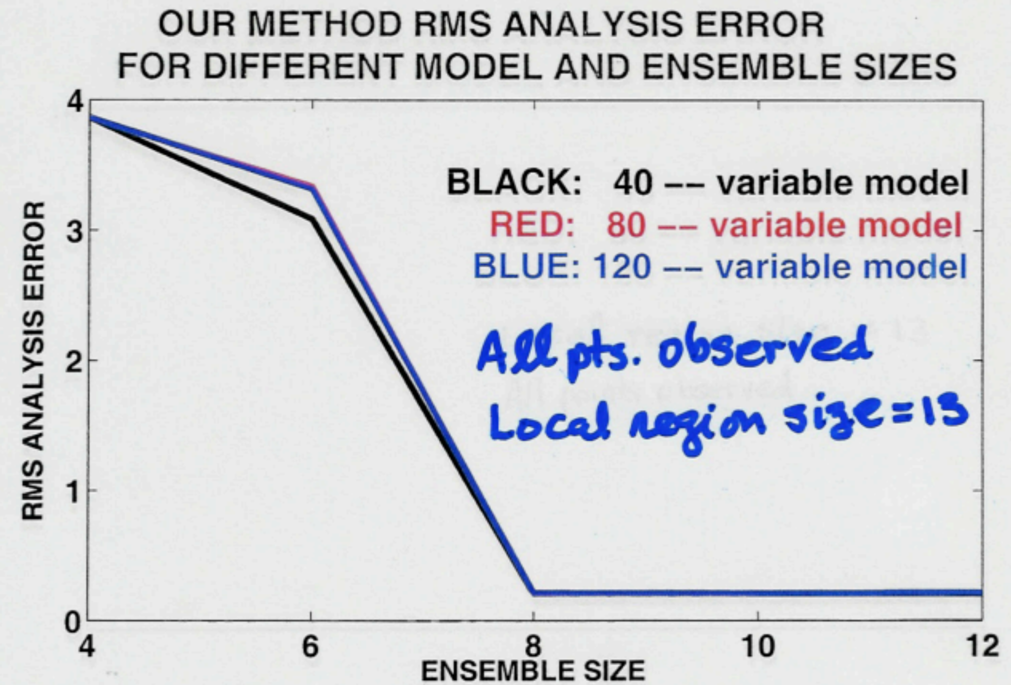
- **Lorenz (1996) :** $\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + 8$

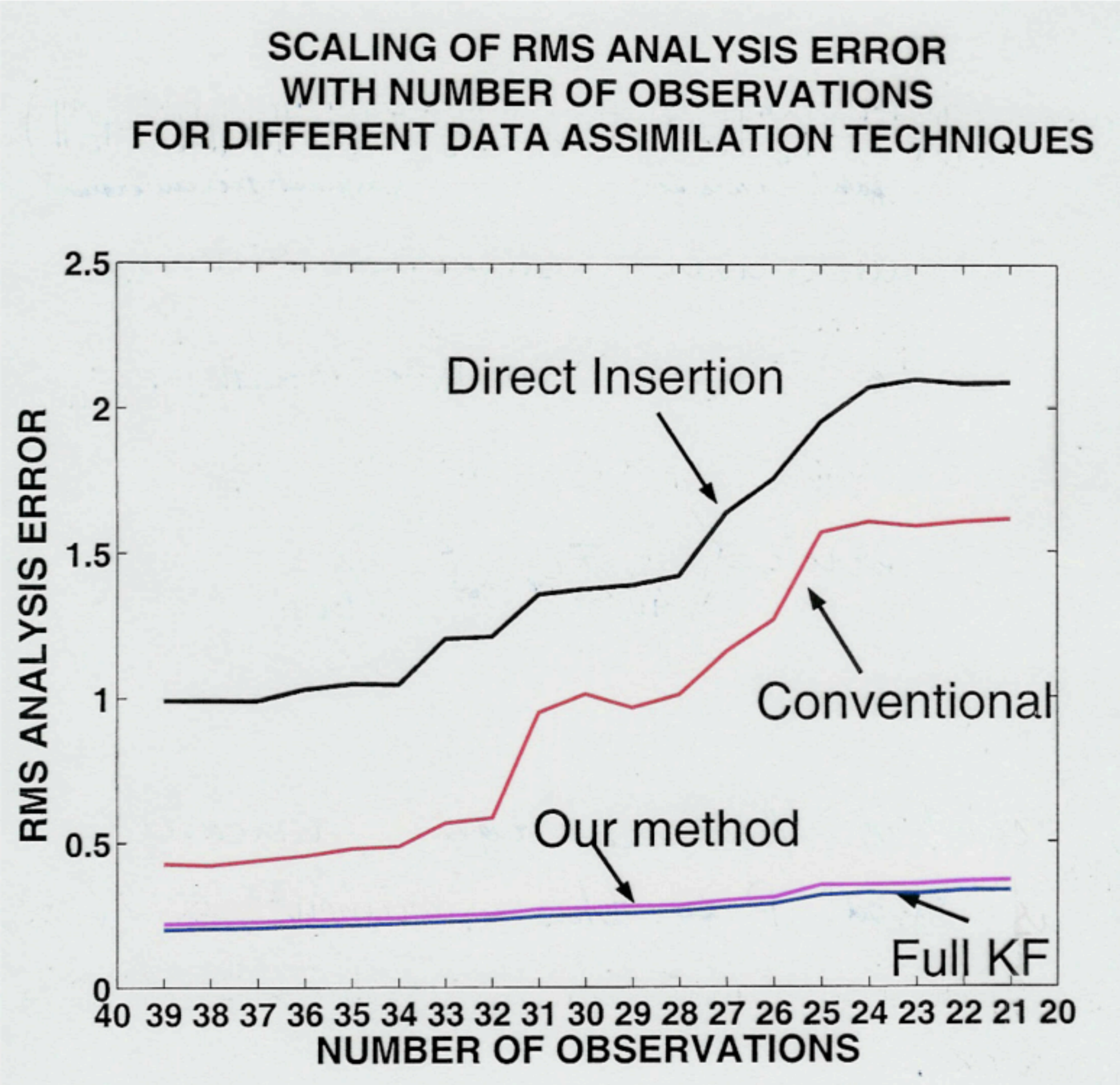


"Latitude Circle"
For N=40
13 positive Lyap. Exponents
Fractal dim. = 27.1

- **We compare results from our method with:**
  - **Global Kalman filter.**
  - **A method mimicking current data assimilation methods (i.e. a fixed forecast error covariance).**
  - **A naïve method called 'direct insertion'.**


FULL KF RMS ANALYSIS ERROR
FOR DIFFERENT MODEL AND ENSEMBLE SIZES
BLACK: 40 -- variable model
RED: 80 -- variable model
BLUE: 120 -- variable model
All pts. observed
RMS ANALYSIS ERROR
ENSEMBLE SIZE
0
1
2
3
4
0
50
100
150
200


OUR METHOD RMS ANALYSIS ERROR
FOR DIFFERENT MODEL AND ENSEMBLE SIZES
BLACK: 40 -- variable model
RED: 80 -- variable model
BLUE: 120 -- variable model
All pts. observed
Local region size=13
RMS ANALYSIS ERROR
ENSEMBLE SIZE
0
1
2
3
4
4
6
8
10
12



SCALING OF RMS ANALYSIS ERROR
WITH NUMBER OF OBSERVATIONS
FOR DIFFERENT DATA ASSIMILATION TECHNIQUES
RMS ANALYSIS ERROR
NUMBER OF OBSERVATIONS
Direct Insertion
Conventional
Our method
Full KF
0
0.5
1
1.5
2
2.5
40 39 38 37 36 35 34 33 32 31 30 29 28 27 26 25 24 23 22 21 20

MAIN RESULTS OF TOY MODEL TESTS

- Both the full KF and our LEKF give about the same accuracy which is substantially better than the ‘conventional method’ and direct insertion.
- Using our method the number of ensemble members needed to obtain good results is independent of the system size, N, while the full Kalman filter requires a number of ensemble members that scales as ~N.

TESTS ON REAL WEATHER MODELS

Our group*

* Ref. Szunyogh et al. Tellus A (2004)

NCEP model:

192 (E-W) x 94 (N-S) x 28 (vertical) x 4 (variables) = 2×10^6

Variables: surface pressure, horizontal wind, temperature, humidity

NASA model:

In the "perfect model scenario" our scheme can yield an over 50% improvement on the current NASA data assimilation system.

NOAA Colorado: (Whittaker and Hamill) NCEP model

Japan: (T. Miyoshi) High resolution code

Results so far:

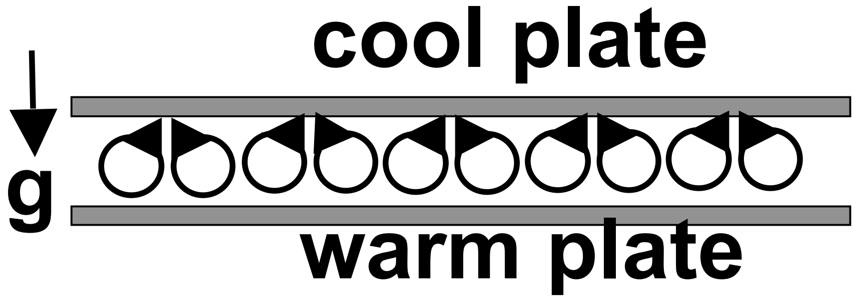
- **Local ensemble Kalman filter does better than current NCEP and NASA assimilation systems**
- **Fast**

OTHER APPLICATIONS

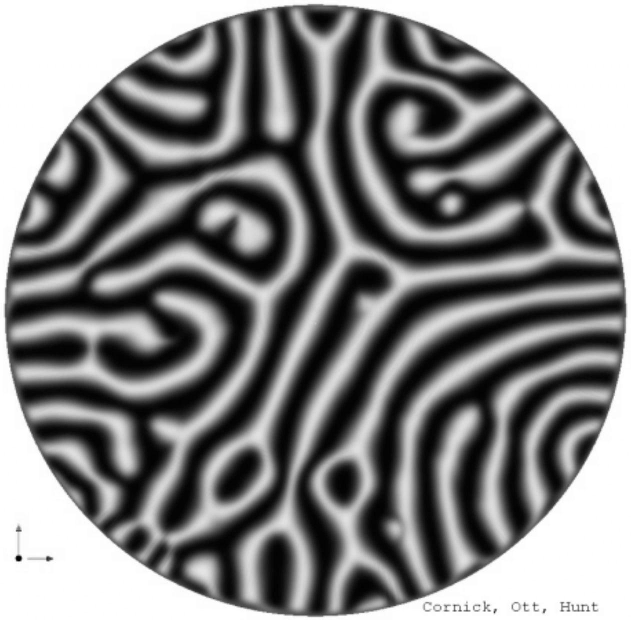
This work is potentially applicable to estimating the state of a large class of spatio-temporally chaotic systems (e.g., lab experiments).

Example:

Rayleigh-Benard convection



Top View:



M. Cornick, E. Ott, and B. Hunt in collaboration with the experimental group of Mike Schatz at Georgia Tech.

Rayleigh-Benard Data Assimilation Tests

Both perfect model numerical experiments and tests using data from the lab experiments were performed.

- **Shadowgraph observation model:** $I = \dfrac{I_0}{1 - \epsilon \nabla^2 \bar{T}(x, y)}$
- **Dynamical model: Boussinesq equations**

NOTE: $(\underline{x}, y, z)$ **not measured.**

"mean flow" $\equiv \dfrac{1}{d}\int_0^d u(\underline{x}, y, z)\,dz \equiv u(\bar{x}, y)$

Some preliminary results:

- **Works well in perfect model and with lab experiment data.**
- **Forecasts indicate that** $(\bar{x}, y)$ **is reasonably accurate.**
- **Parameter estimation of Ra, Pr, C.**

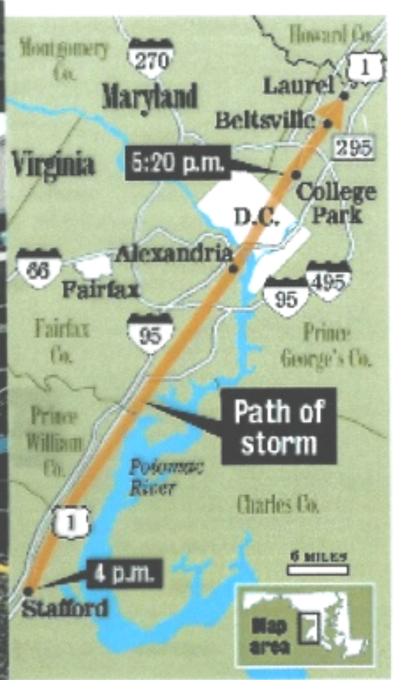
PROPERTIES OF THE METHOD

- Only low dimensional matrix operations are used in the analysis.
- Local analyses are independent and hence parallelizable.
- Potentially fast and accurate.

http://www.weatherchaos.umd.edu/publications.php

Tornado Hits University of Maryland

Chronology of Monday's deadly tornado



SOURCE: National Weather Service

4 p.m. Rotating winds spotted on radar in Stafford, Va.

4:09 p.m. National Weather Service issues tornado warning for northern Stafford County and eastern Prince William County.

4:53 p.m. Weather service issues warning for eastern Fairfax County, D.C. metro area and Alexandria.

5:03 p.m. Storm passes over the Pentagon.

5:10 p.m. Weather service issues warning for Prince George's County and eastern Montgomery County.

5:20 p.m. Storm strengthens, strikes College Park, moves on to Beltsville and Laurel.

5:31 p.m. Weather service issues tornado warning for Howard County.

5:45 p.m. Tornado warning in Prince George's County expires.

6:00 p.m. Howard tornado warning expires.

DENISE MURRAY : SUN STAFF

OUTLINE OF OUR METHOD

- Consider the global atmospheric state restricted to many local regions covering the surface of the Earth.
- Project the local states to their local low dim. subspace determined by the forecast ensemble.
- Do data assimilations for each local region in that region's low dim. subspace.
- Put together the local analyses to form a new ensemble of global states.
- Use the system model to advance each new ensemble member to the next analysis time.