Introduction to Data Assimilation

Javier Amezcua
Data Assimilation Research Centre
University of Reading

University of Reading

Data Assimilation

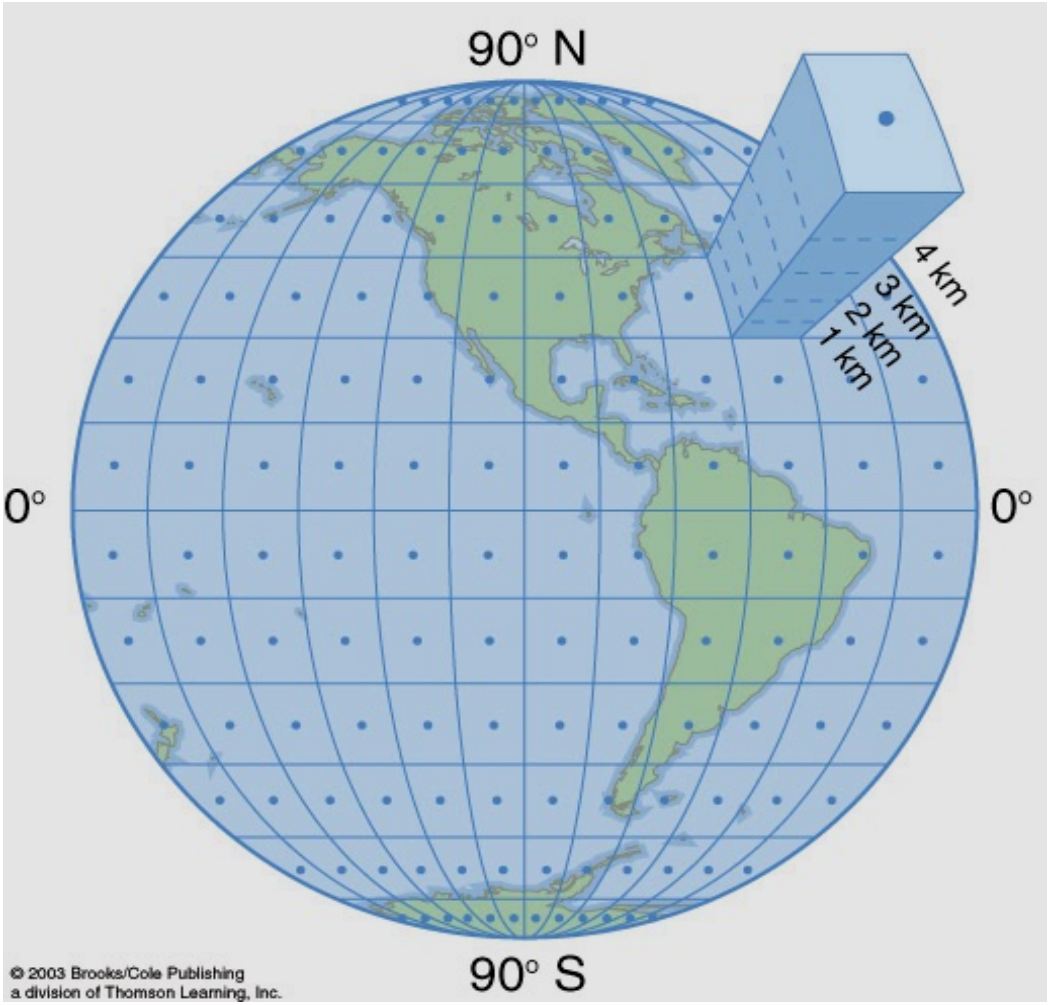
Consider we are interested in a (physical / dynamical) **system**.

Then, **DA** has two main **objectives**:

a. To find a **current estimate** that can be used to **produce forecasts**.

b. To quantify the **uncertainty of the estimate**, and to know the **time evolution** of this **uncertainty**.

Our system of interest

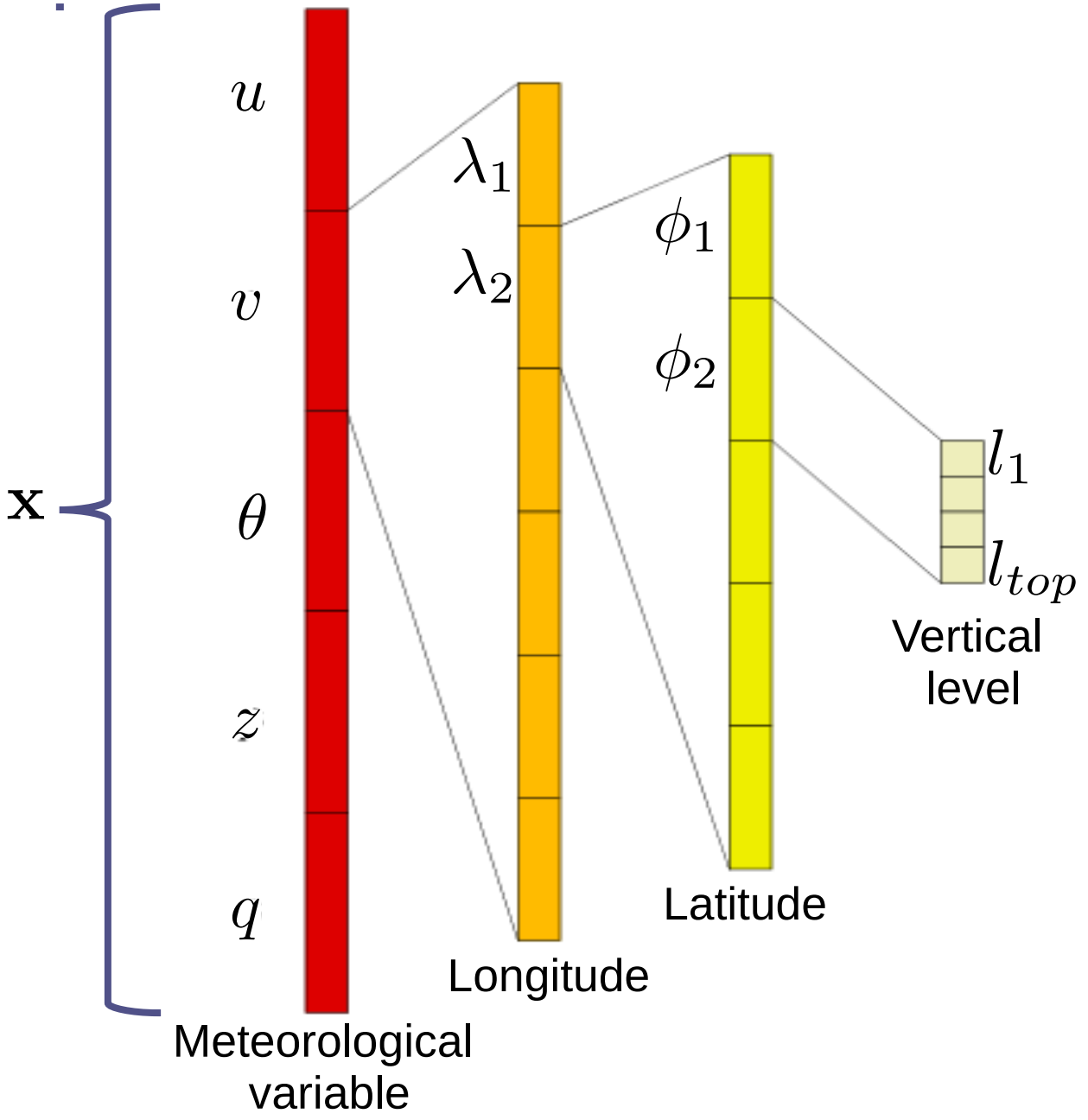


The **state variables** of the system are: **meteorological variables** (wind speed, temperature, etc) in **every** single **gridpoint**.

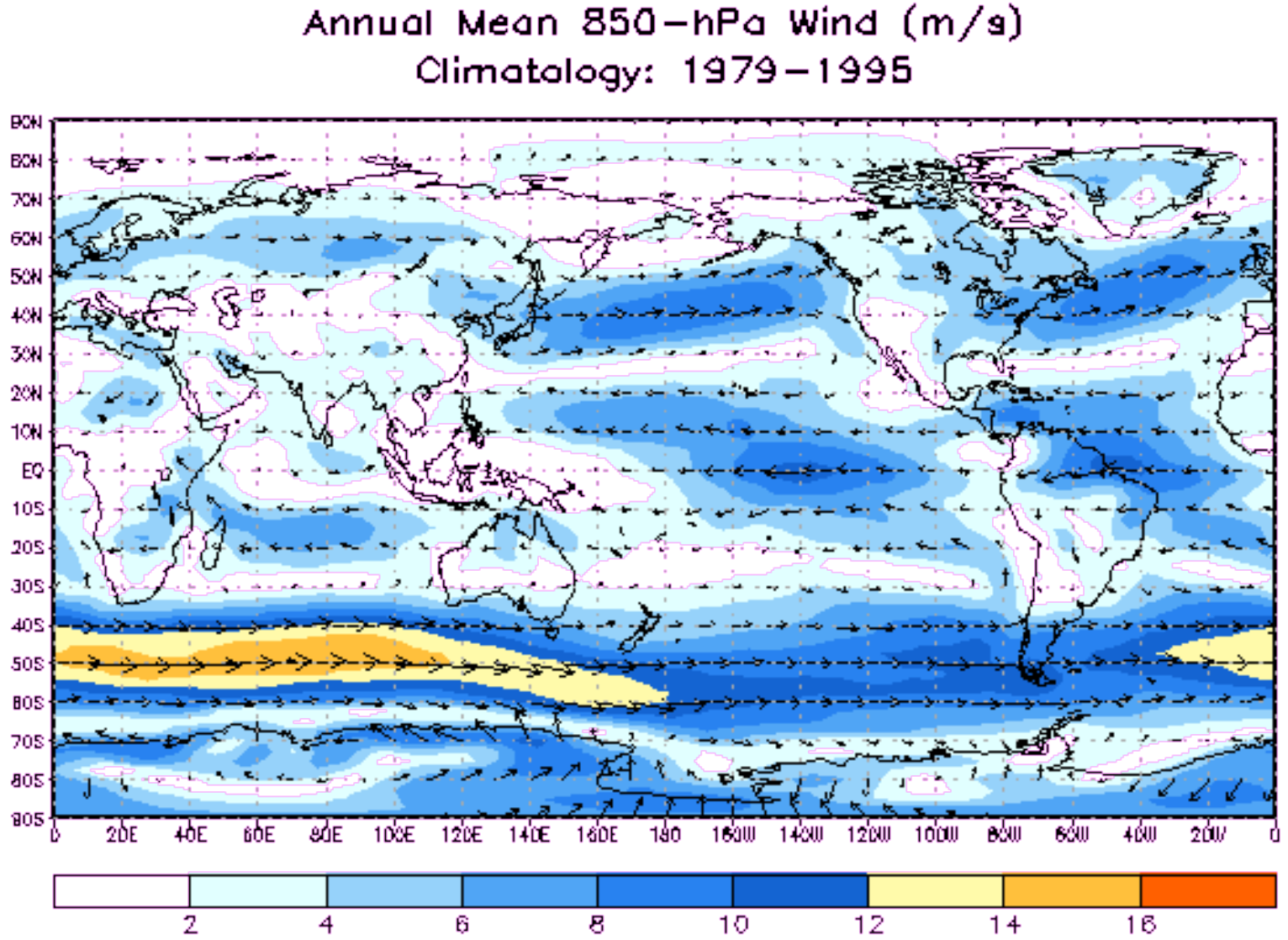
Our system of interest

State variables:

$$\mathbf{x} \in \mathcal{R}^{N_x}$$



Climatological information



Long-term **average information** of the system gives an idea of **'permissible' values**. But this is not what we are after.

Two challenges

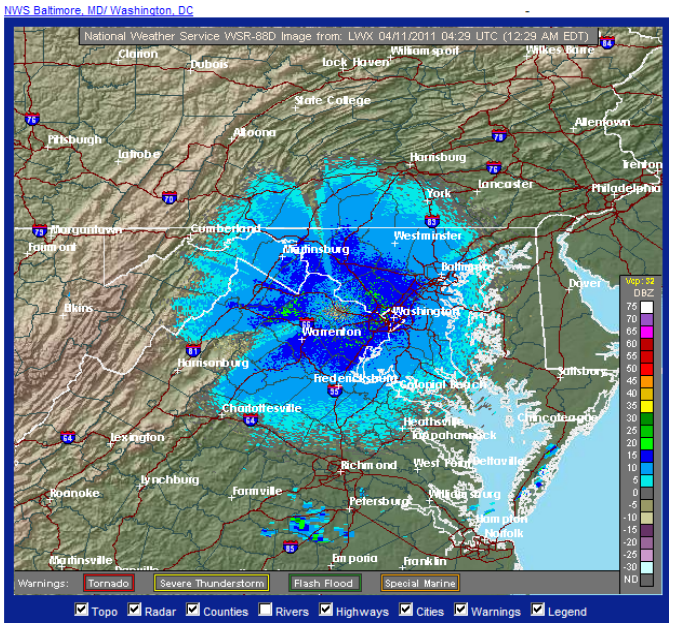
1. Determining the **current estate** of the system (all state variables) **at a given moment** of time. This is **estimation**.

2. Given some initial conditions, determine the **future state of the system** (all state variables). This is **prediction**.

Contrast these with the **aims of DA**!

Two sources of information

- **Observations**
  - How accurate?
  - How dense?
  - How do they relate to the state variables?



- **Models**
  - **Diagnostic** equations

$$p = \rho R T$$

$$\mathbf{v} = \frac{\hat{\mathbf{k}}}{f} \times \nabla_p \Phi$$

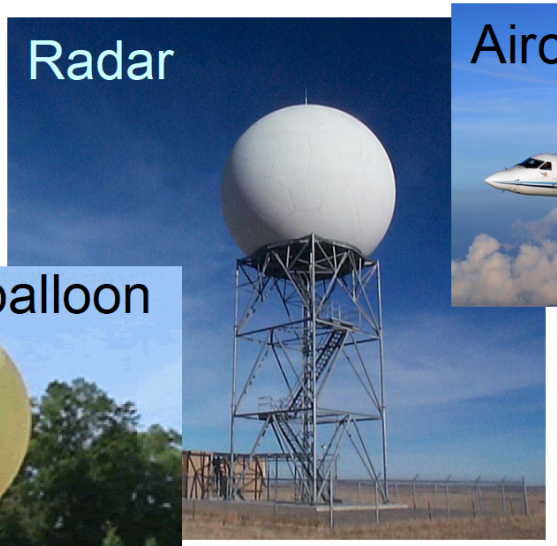
  - **Prognostic** equations (future)

$$\frac{\mathbf{D}}{Dt}\mathbf{v} = -\frac{1}{\rho}\nabla p - f\hat{\mathbf{k}} \times \mathbf{v} + \mathbf{F}$$

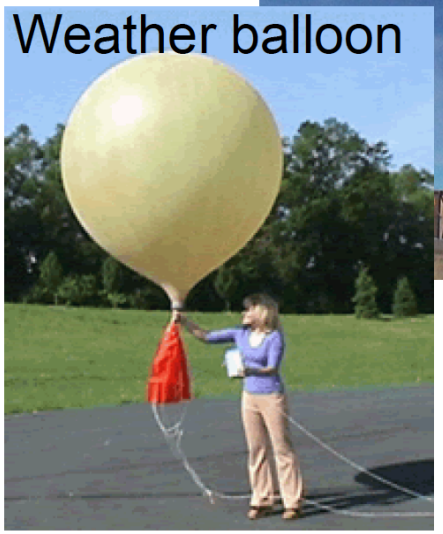
None of them **are perfect**! The **both have errors** and we must take them into consideration when combining them.

Observations

.



Satellite

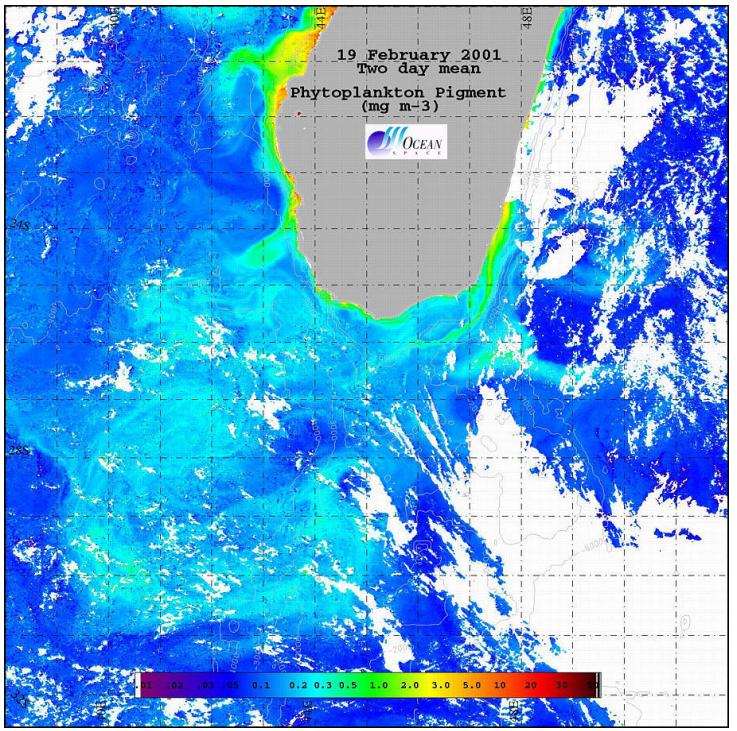


Observations

- **In situ** observations:

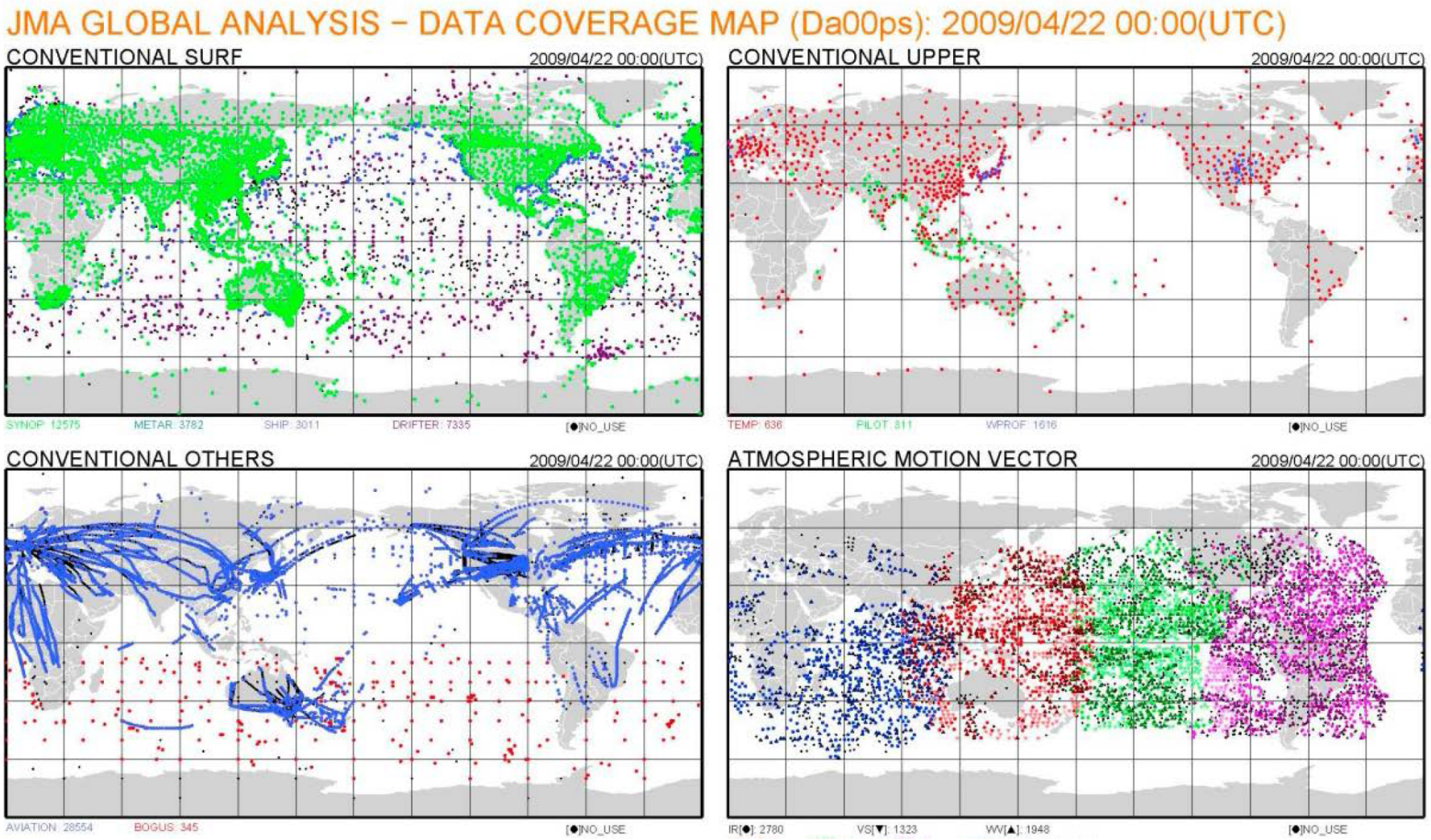
  They are **direct**, but they can be **irregular in space and time,** e.g. sparse hydrographic observations.

- **Remote sensing** observations: They are **indirect.** E.g. satellites measuring the sea-surface temperature.



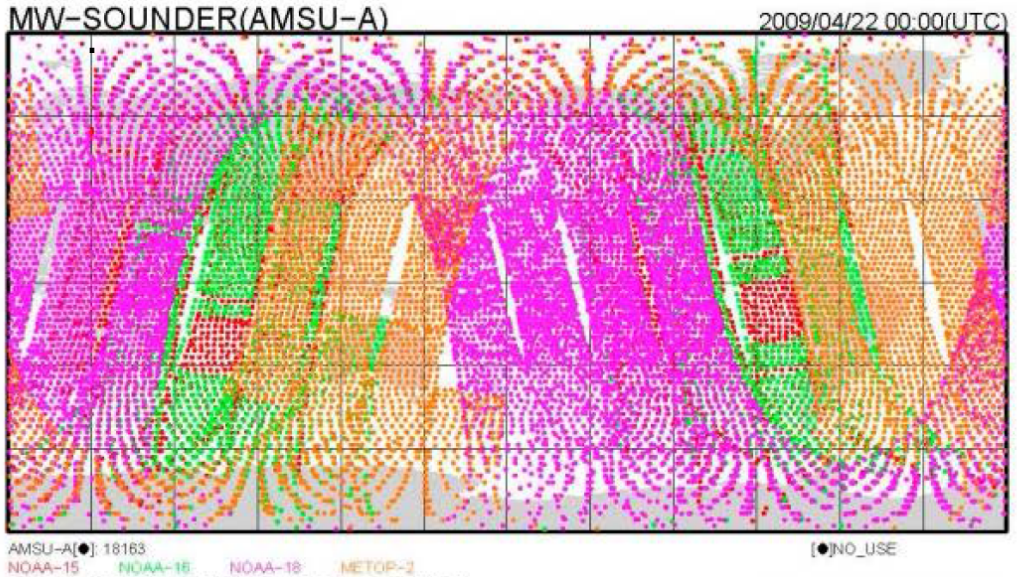
Observation coverage

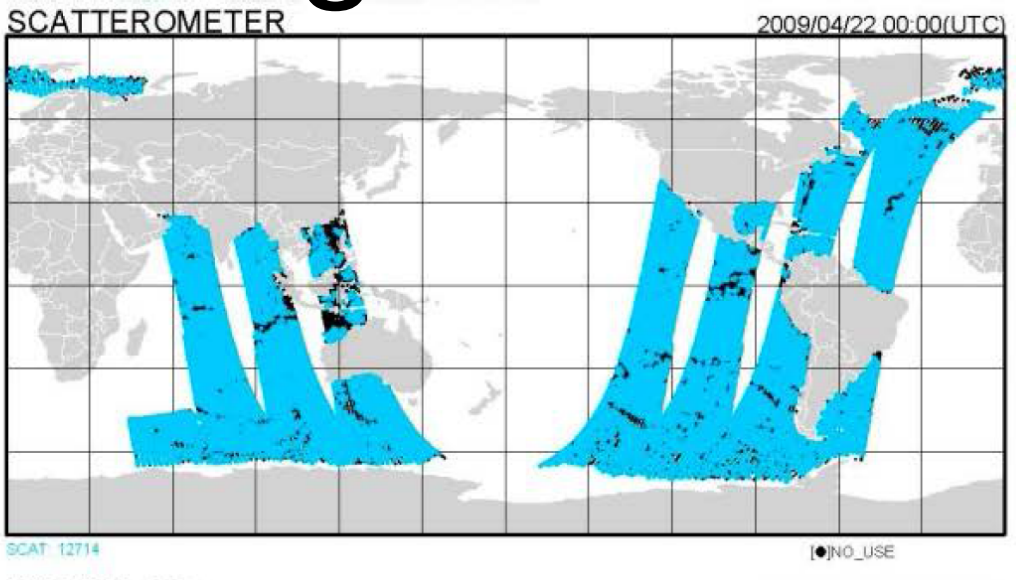
.

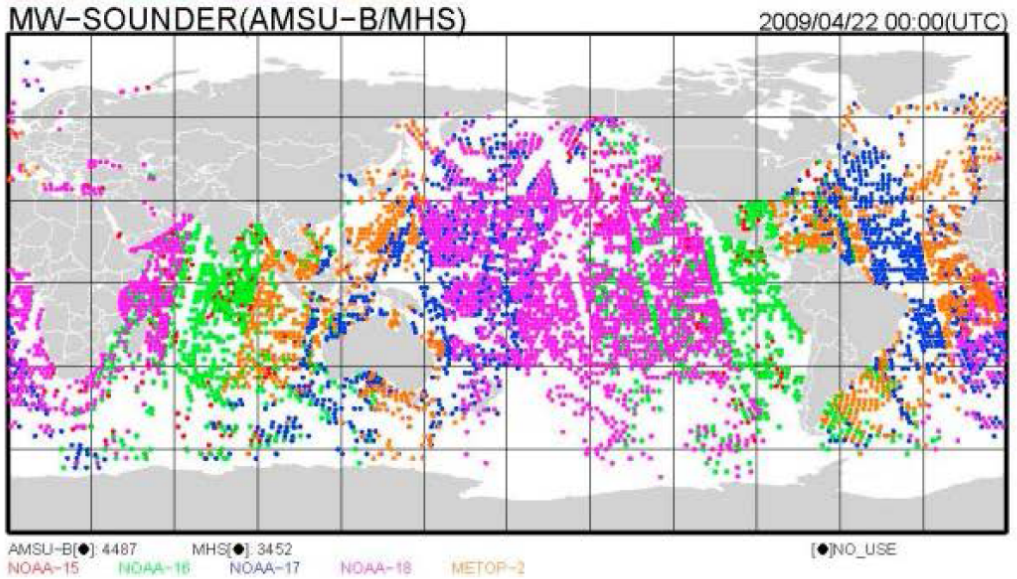


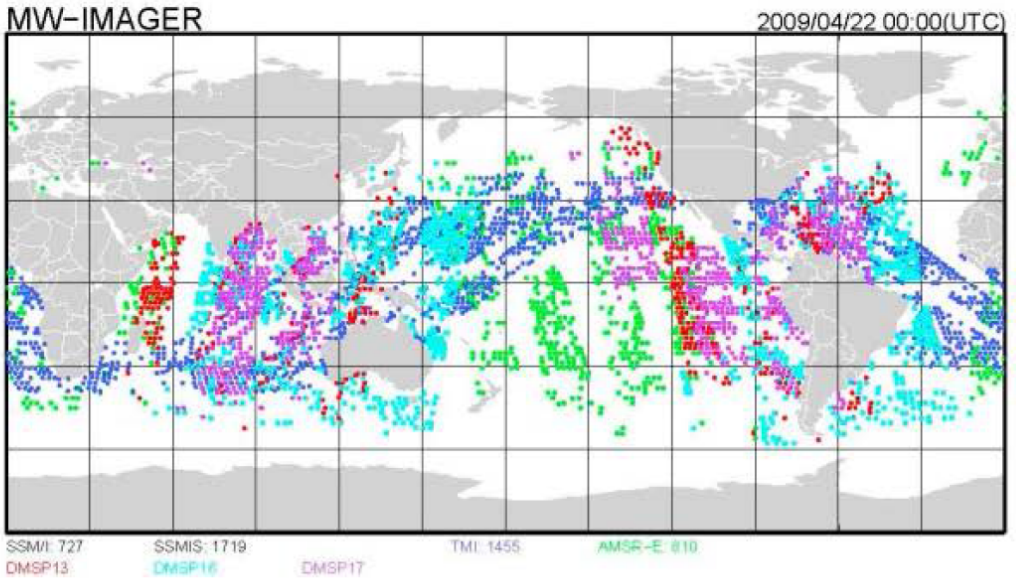
World's effort! (no border in the atmosphere)

Observation coverage









Great coverage nowadays. Nonetheless we do not observe every single variable at every single model gridpoint. The **system is partially observed**.

Observations

.

$$\mathbf{y} = h\left(\mathbf{x}\right) + \text{error}$$

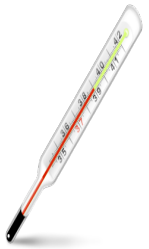
$\mathbf{y} \in \mathcal{R}^{N_y}$

Usually: $N_y \ll N_x$

Transformation of the state variables via an **observation operator.**

The **observations** are not perfect. **Errors** come from:
a. Instrument capabilities.
b. Representativity: i.e. observations and models may have a different resolution.
c. Characterising the observation operator incorrectly.
...

Observation operators $h(\mathbf{x})$

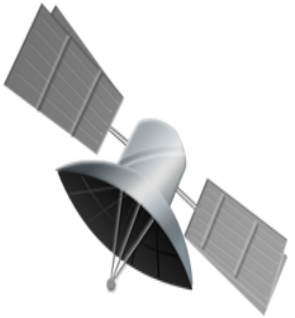


Variable:
Temperature at a point

Observation:
Temperature at a point

Operator: Identity

$$\mathbf{y} = \mathbf{x}$$

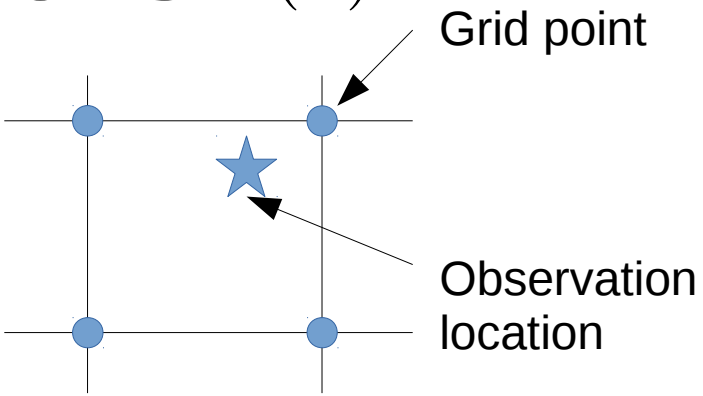


Variable:
Temperature at a vertical level

Observation: Total radiance coming from a vertical column

Operator: Integral transformation

$$\mathbf{y} = \int_0^{z_{top}} \sigma_{Boltz} \mathbf{x}(z)^4 dz$$



Variable: Temperature at gridpoints

Observation:
Temperature outside a gridpoint

Operator: Interpolator

$$\mathbf{y} = \mathbf{H}\mathbf{x}$$

$$\mathbf{H} \in \mathcal{R}^{N_y \times N_x}$$

Retrieving the value(s) of the state variable(s) from the observation(s) is called the **inverse problem**. This is a related problem.

Models

.

$$\mathbf{x}^t = m^{t-1 \to t}\left(\mathbf{x}^{t-1}\right) + \text{error}$$

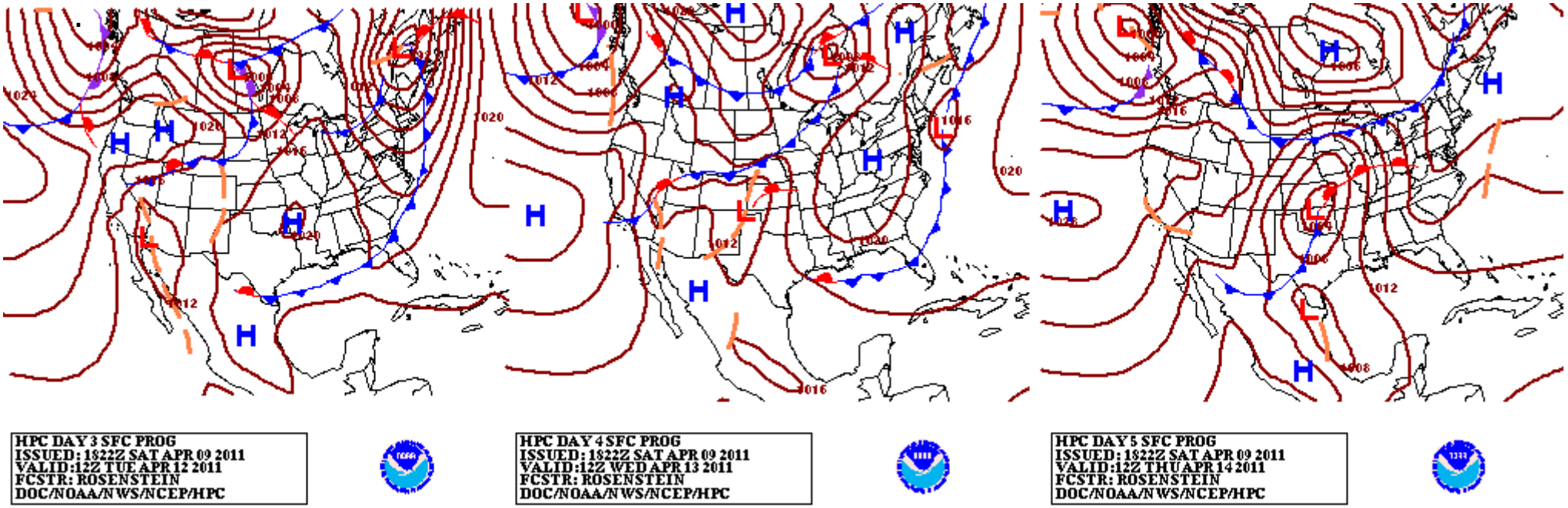
$\mathbf{x}^t \in \mathcal{R}^{N_x}$

Evolution operator

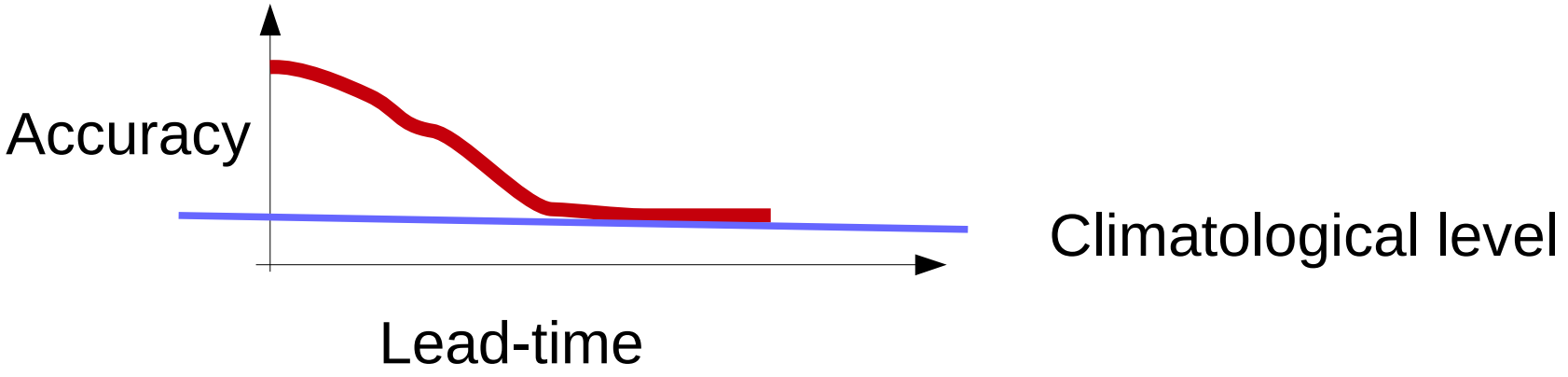
Previous value of the variable.

The models are not perfect.
Errors come from:
a. Unknown physics
b. Numerical error in the time/space discretisation of continuous equations.
c. Subgrid processes that need to be parameterised.
...

Forecast with different lead-times



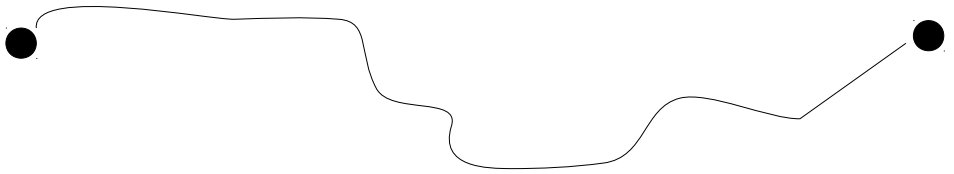
Should we consider the three of forecasts to have the same accuracy (different lead-times)?



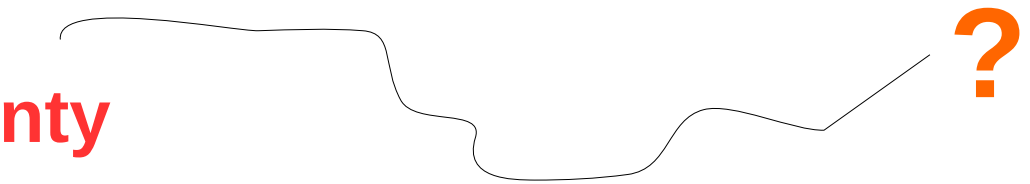
A perfect **model** with **uncertain initial conditions**

$$\mathbf{x}^t = m^{0 \to t}\left(\mathbf{x}^0\right)$$

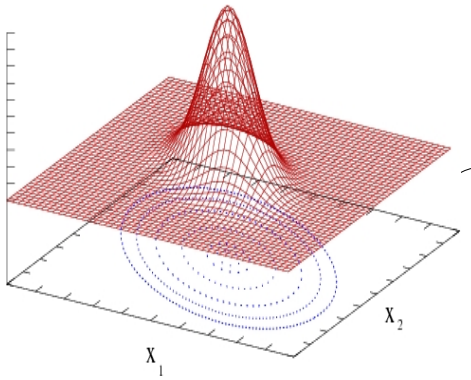
point estimate

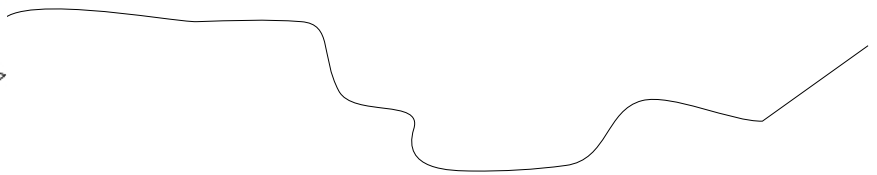


uncertainty



?





Deterministic **chaos**

.

$$\mathbf{x}^t = m^{0 \to t}\left(\mathbf{x}^t\right)$$

Consider the **model** to be **perfect**. Then the **state of the system –**at any time- **is completely determined by the initial conditions**.

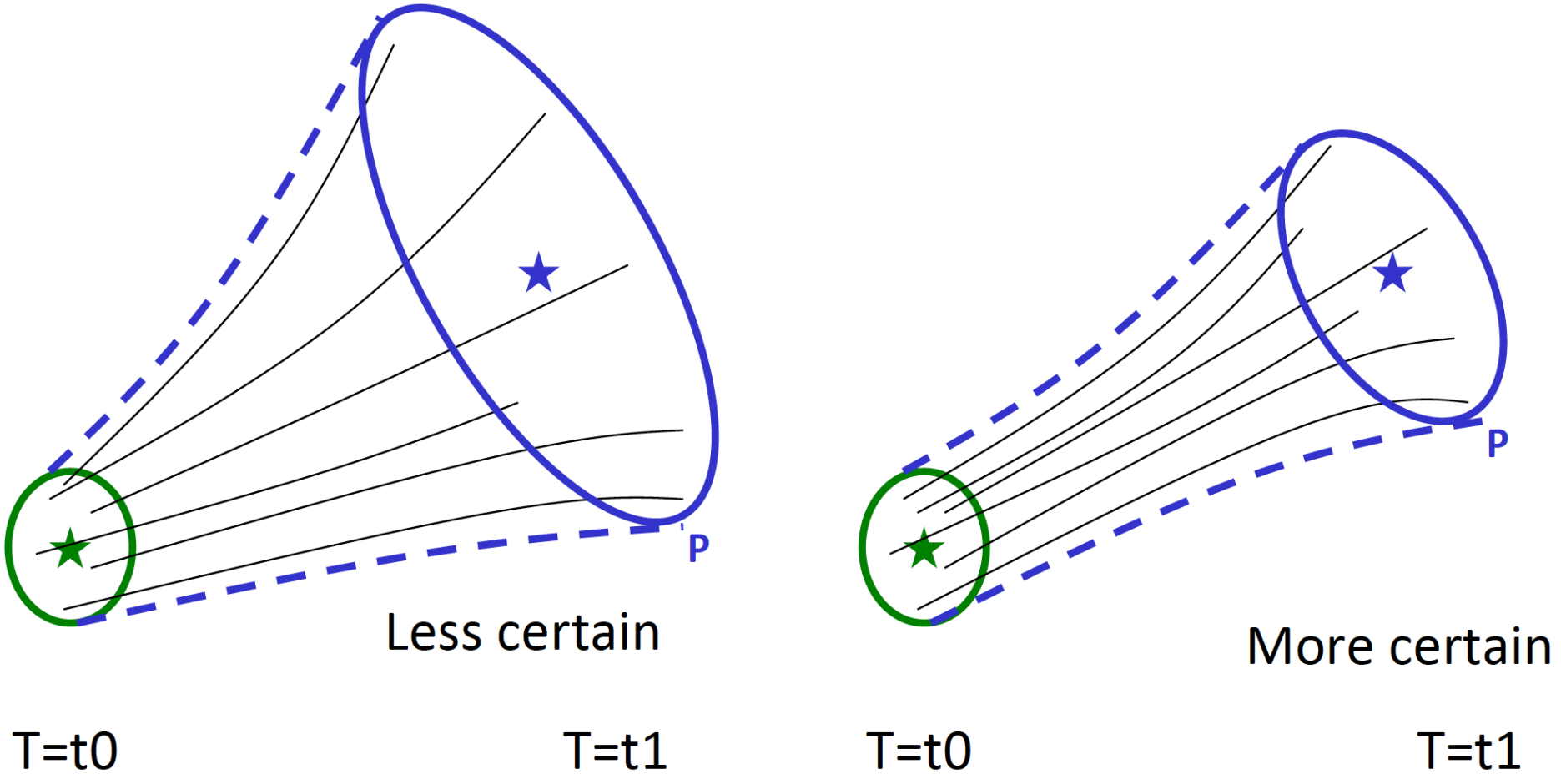
Can we determine with **absolute precision**? **No**.

How **sensitive** is the forecast to these **errors in initial conditions**?

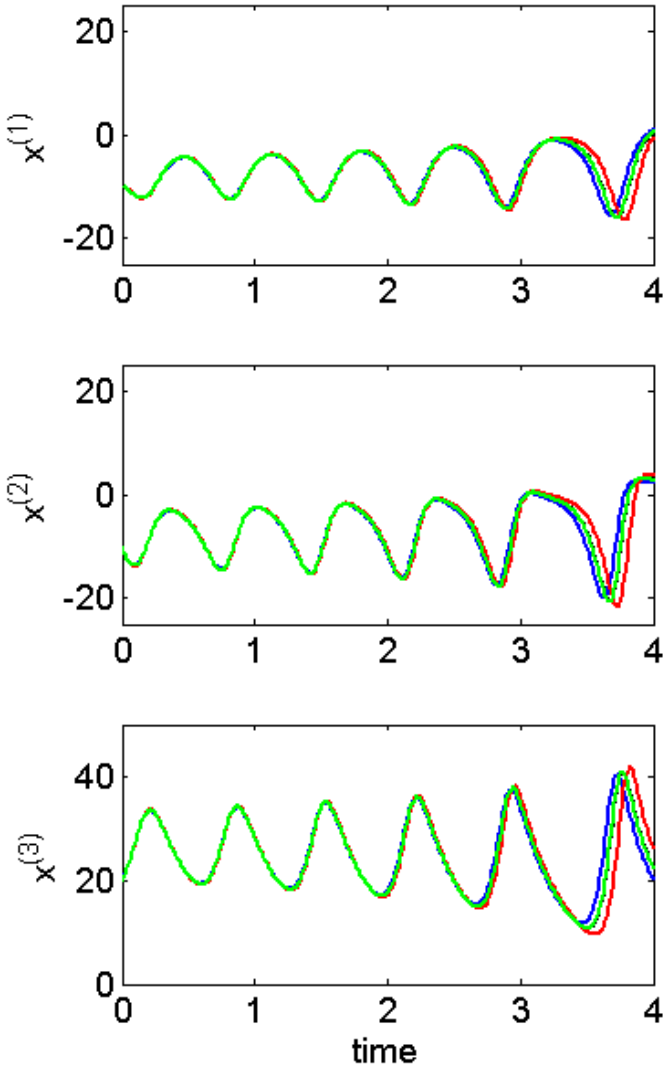
In **chaotic systems** –like the atmosphere- it matters a lot.

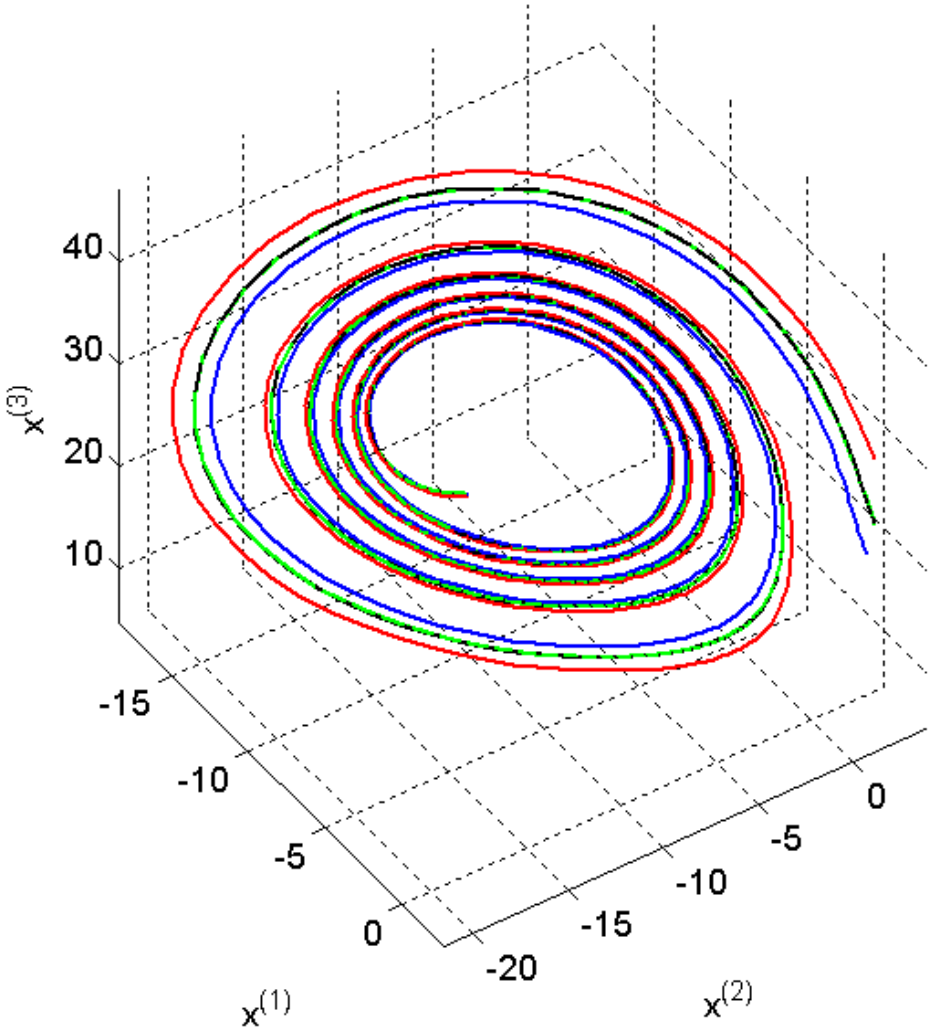
Sensitivity to initial conditions

- Perturb the initial conditions and run the multiple forecasts (a.k.a. ensemble forecasts)



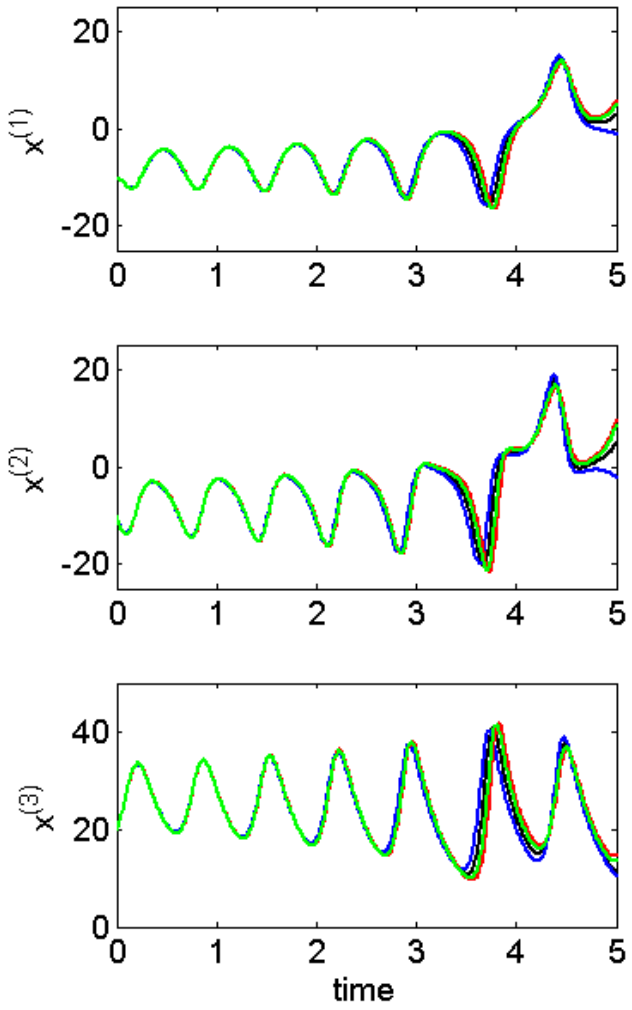
Example: Lorenz 1963 model

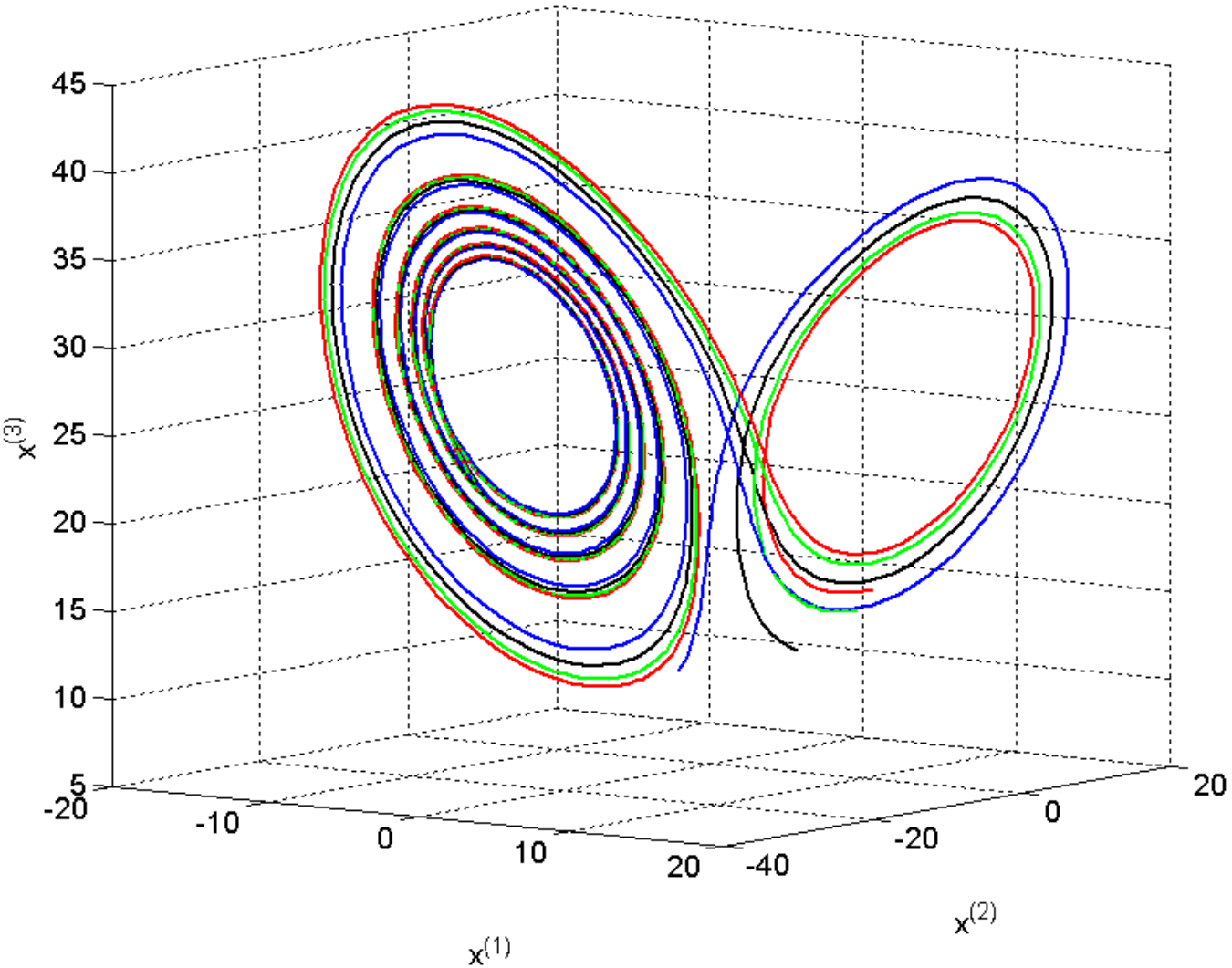




Example: Lorenz 1963 model

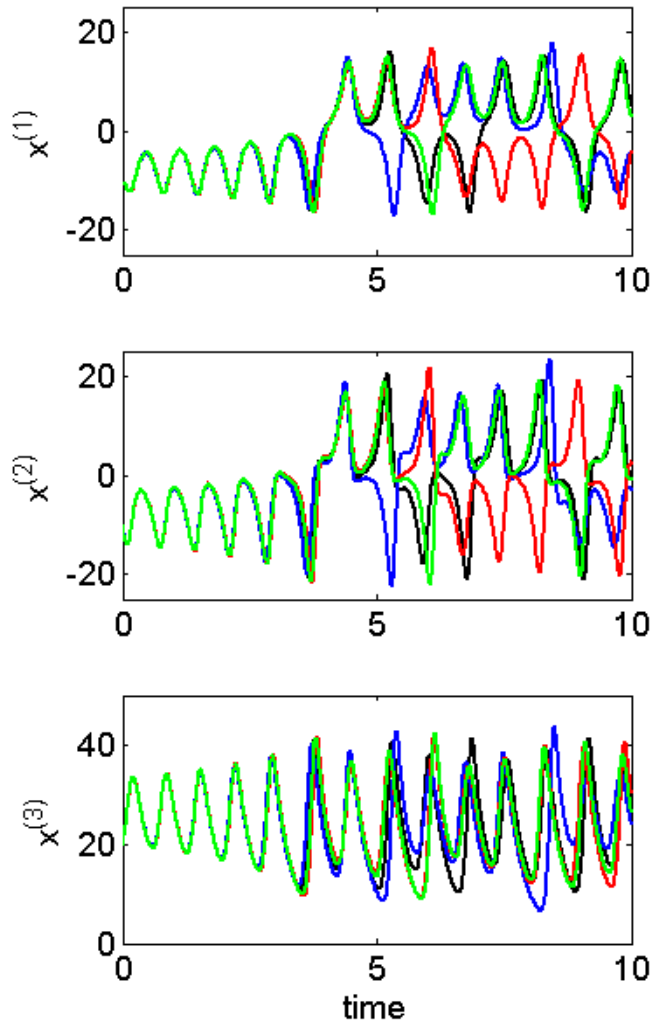
.

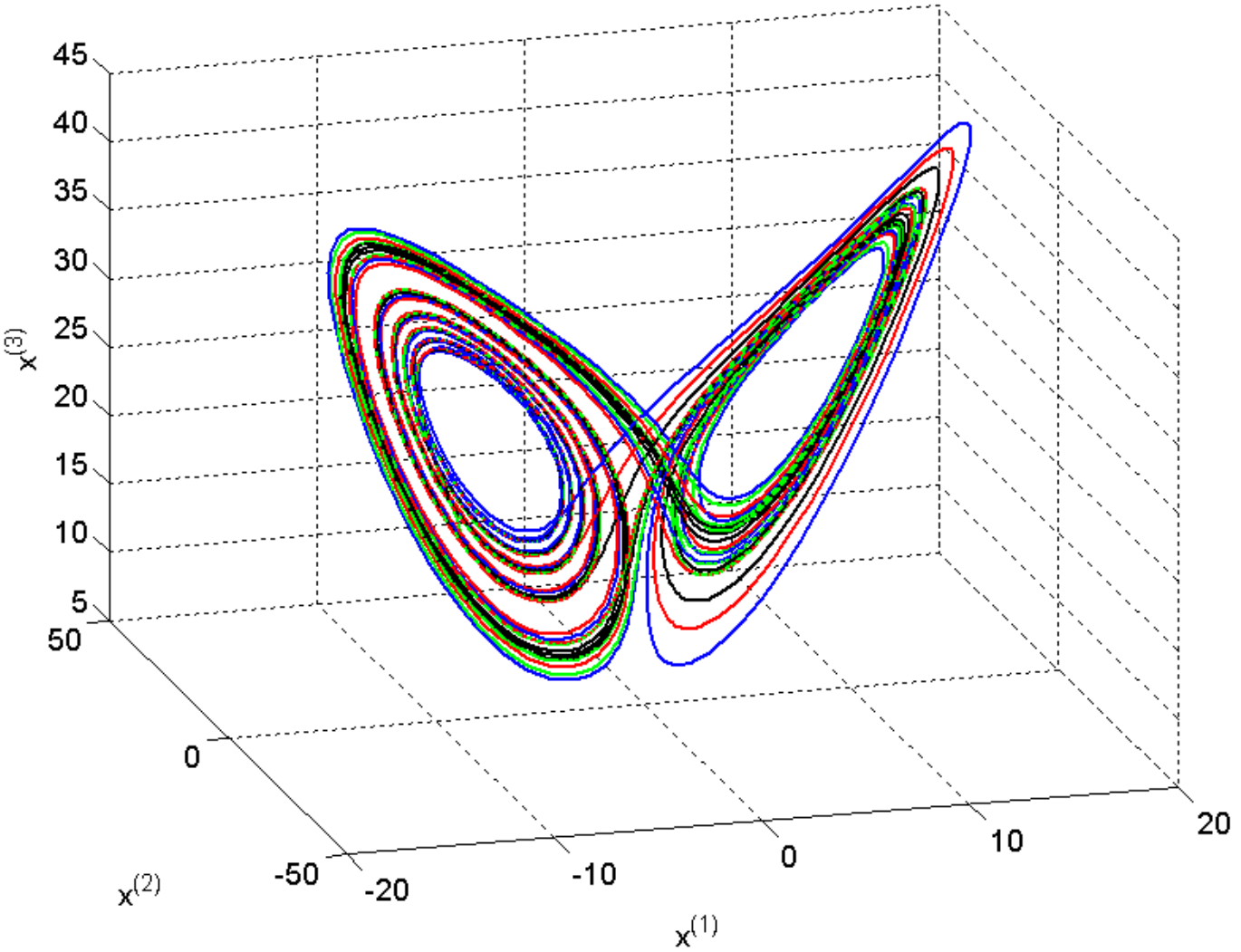




Example: Lorenz 1963 model

.

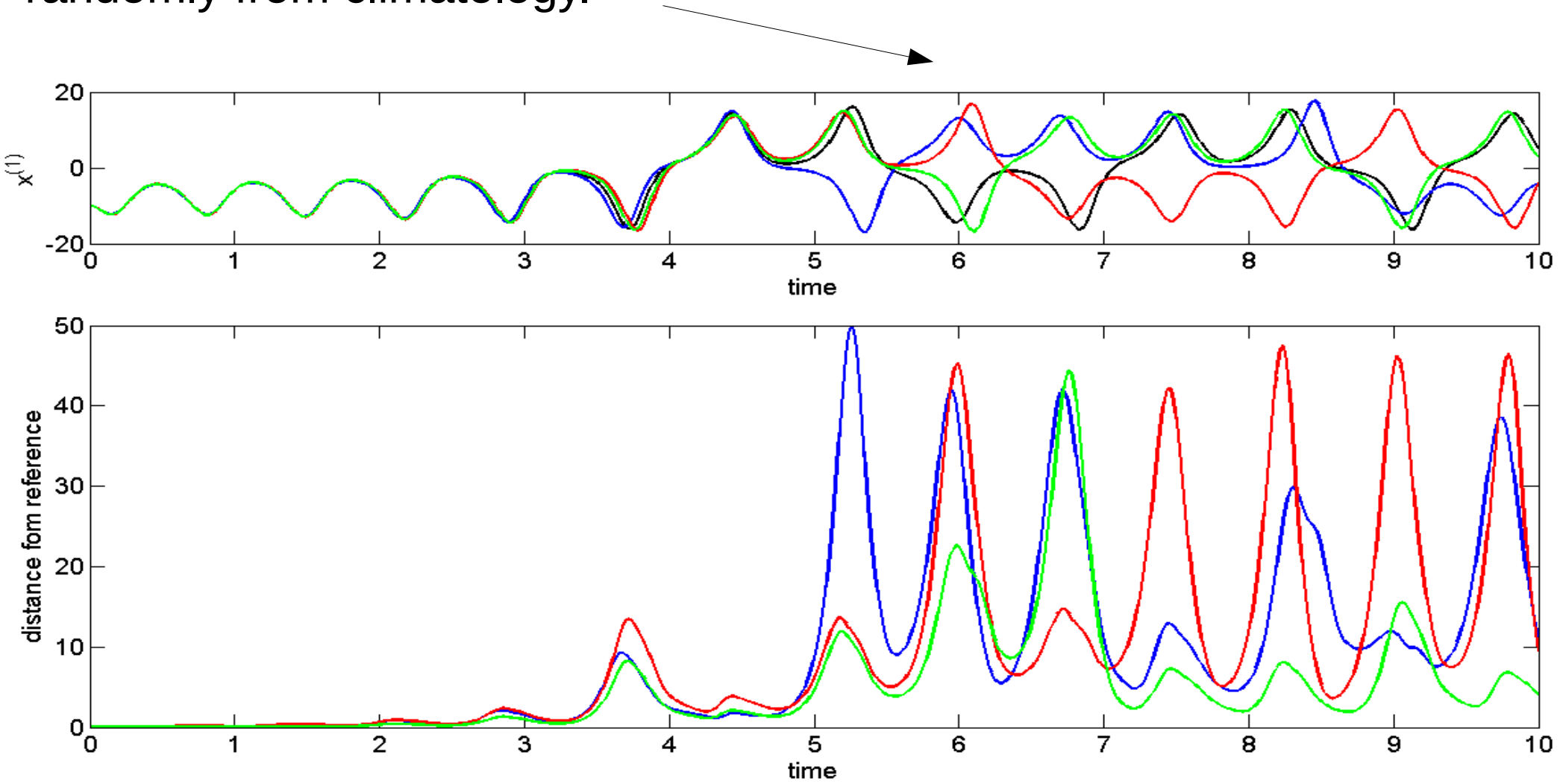




Example: Lorenz 1963 model

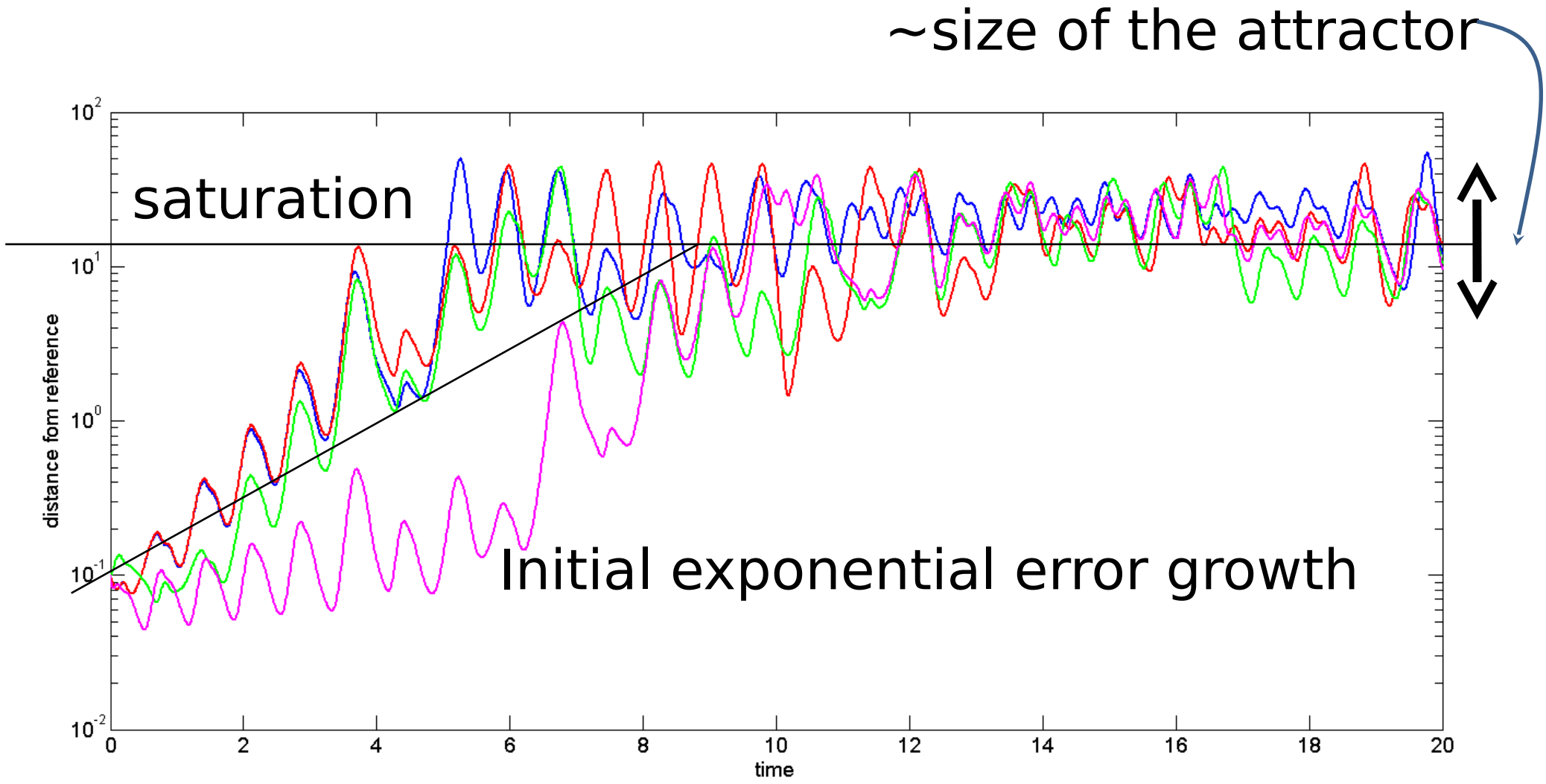
.

The trajectories are so different they may as well have been chosen randomly from climatology.



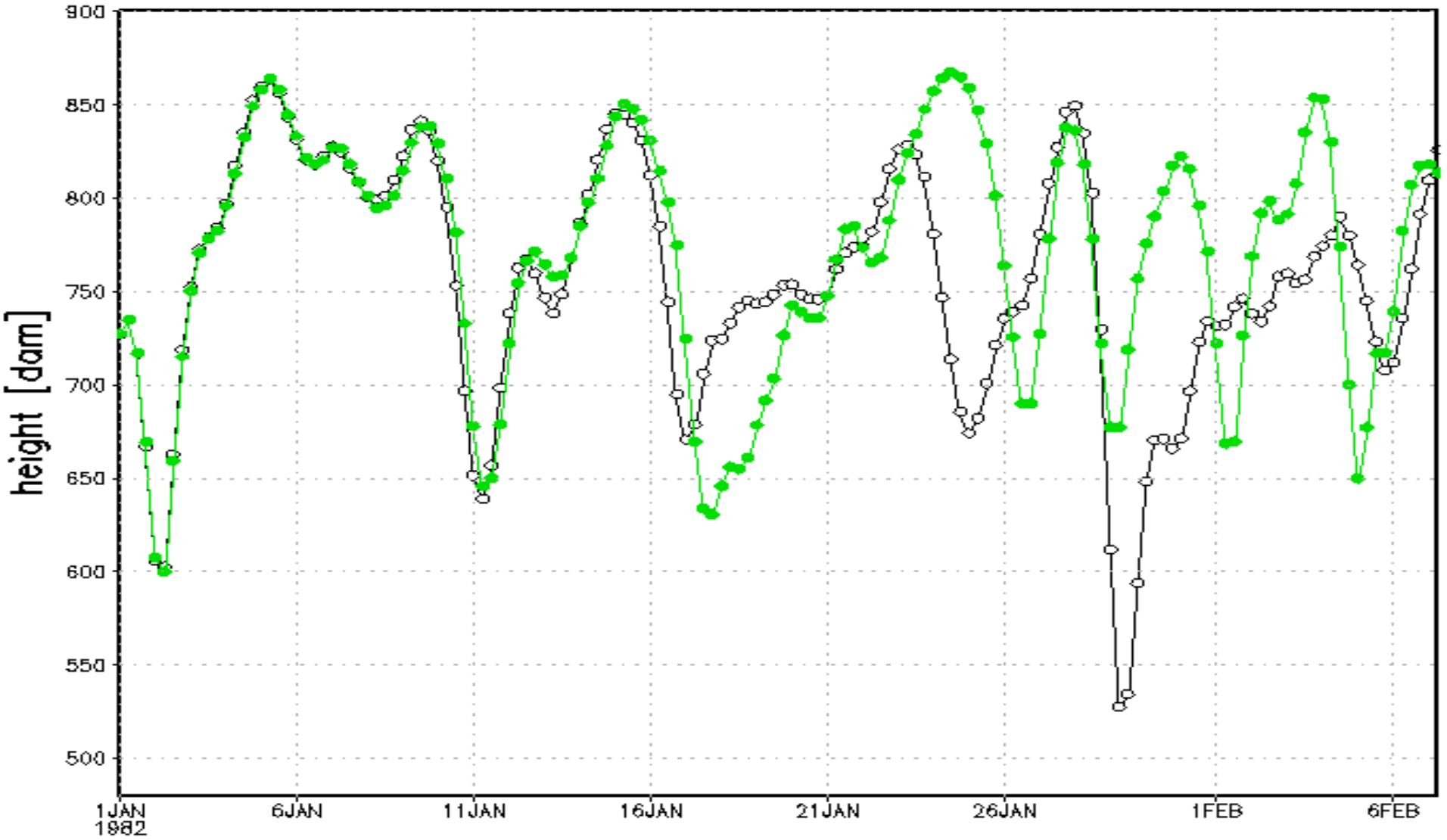
Example: Lorenz 1963 model

.

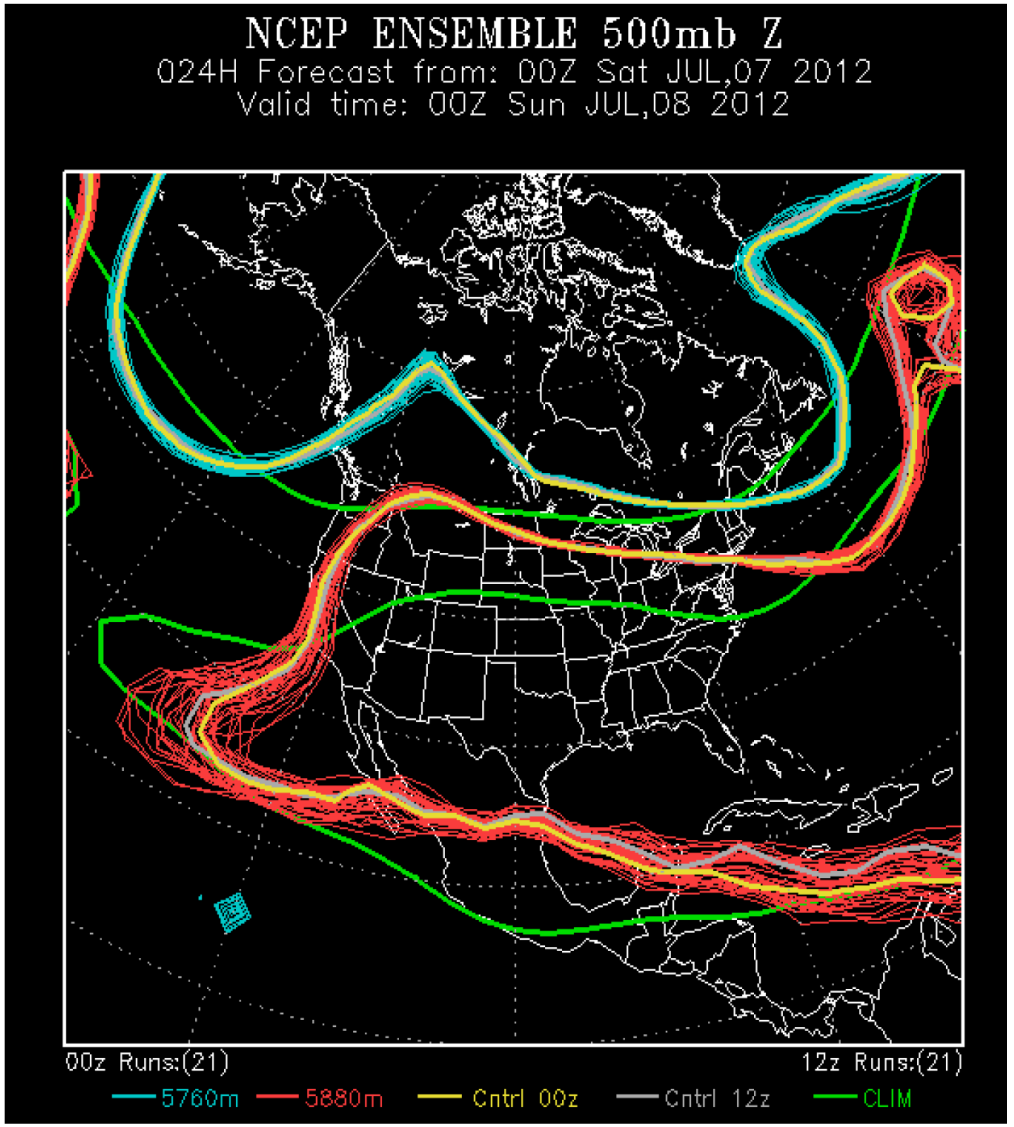


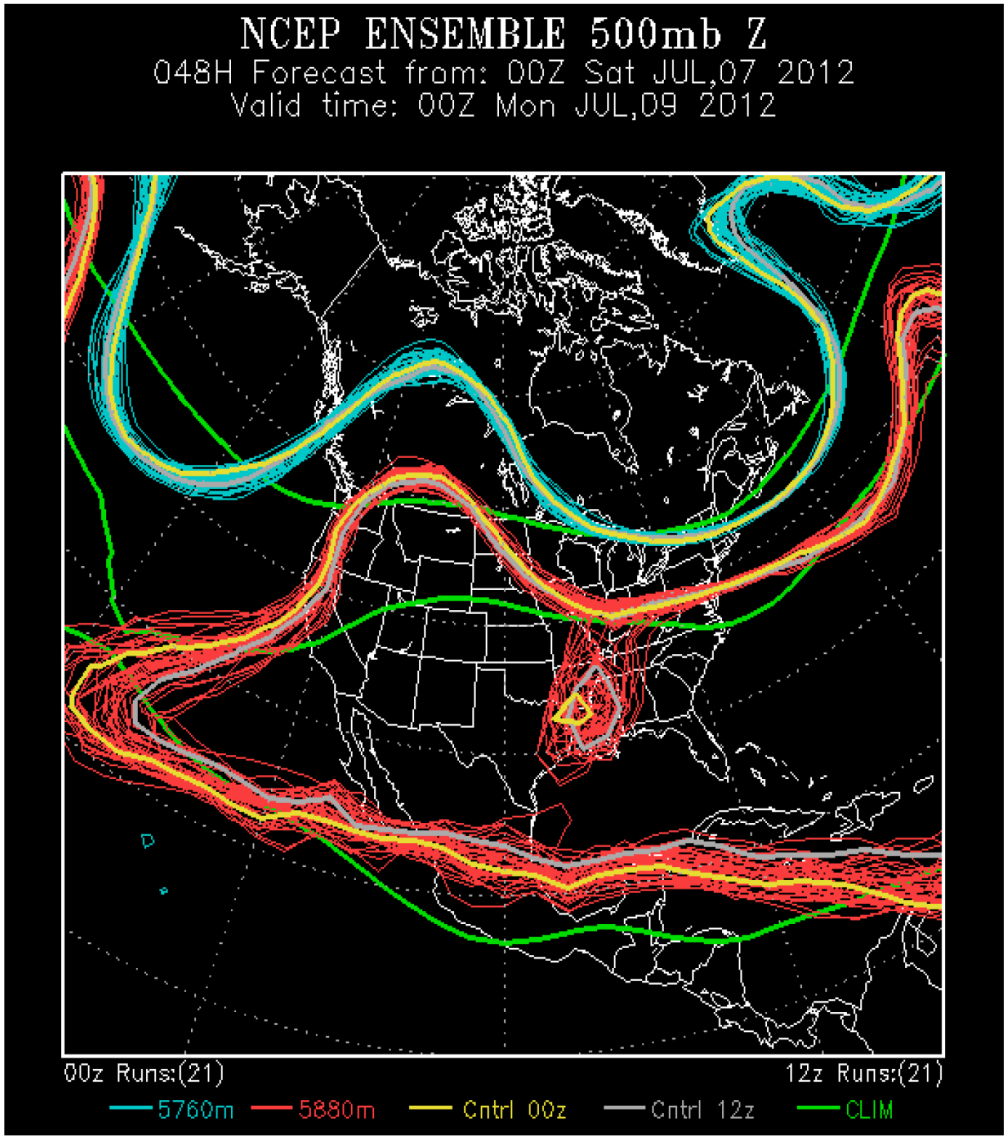
The atmosphere is chaotic

Evolution of the 500-hPa geopotential height in CP

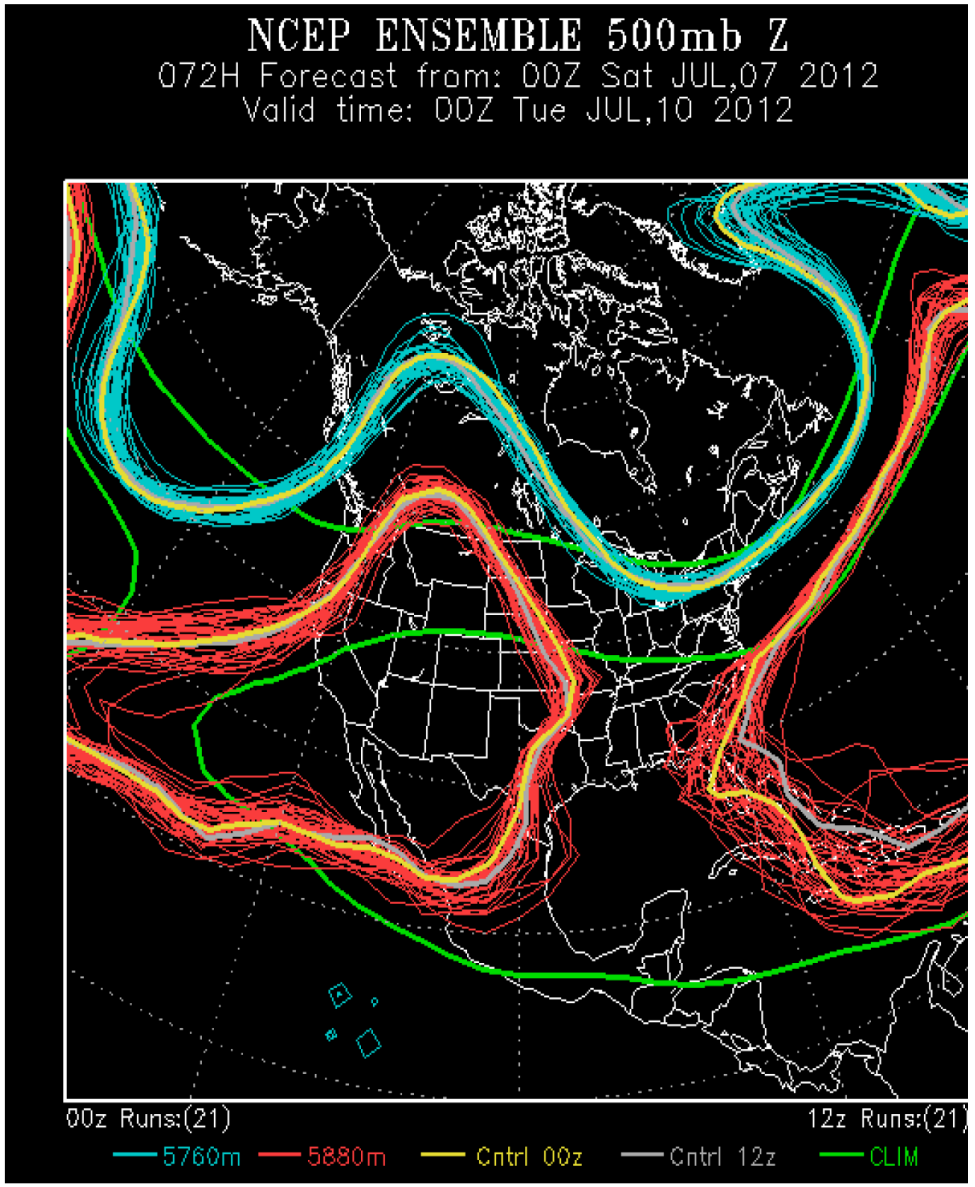


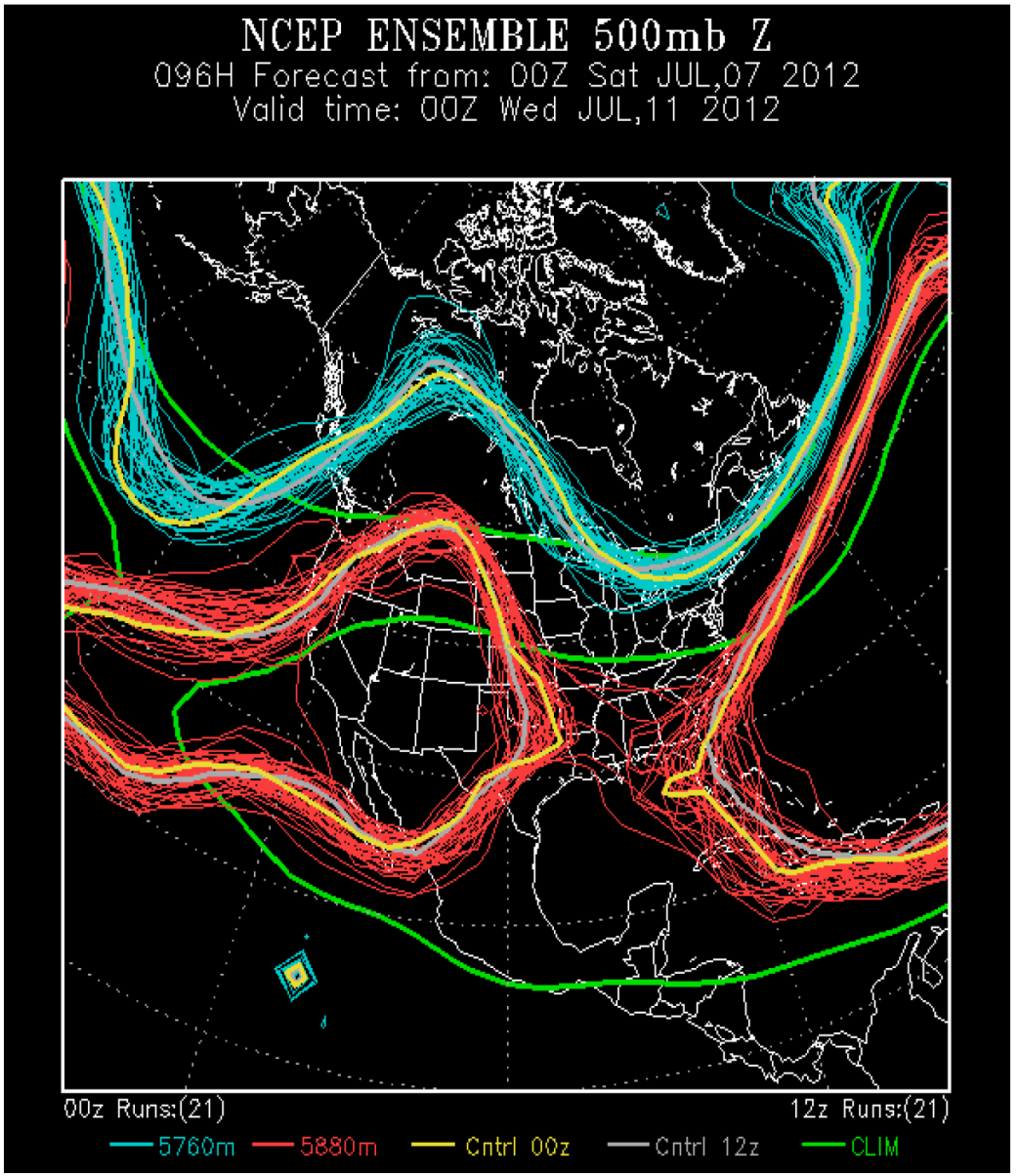
Weather is chaotic



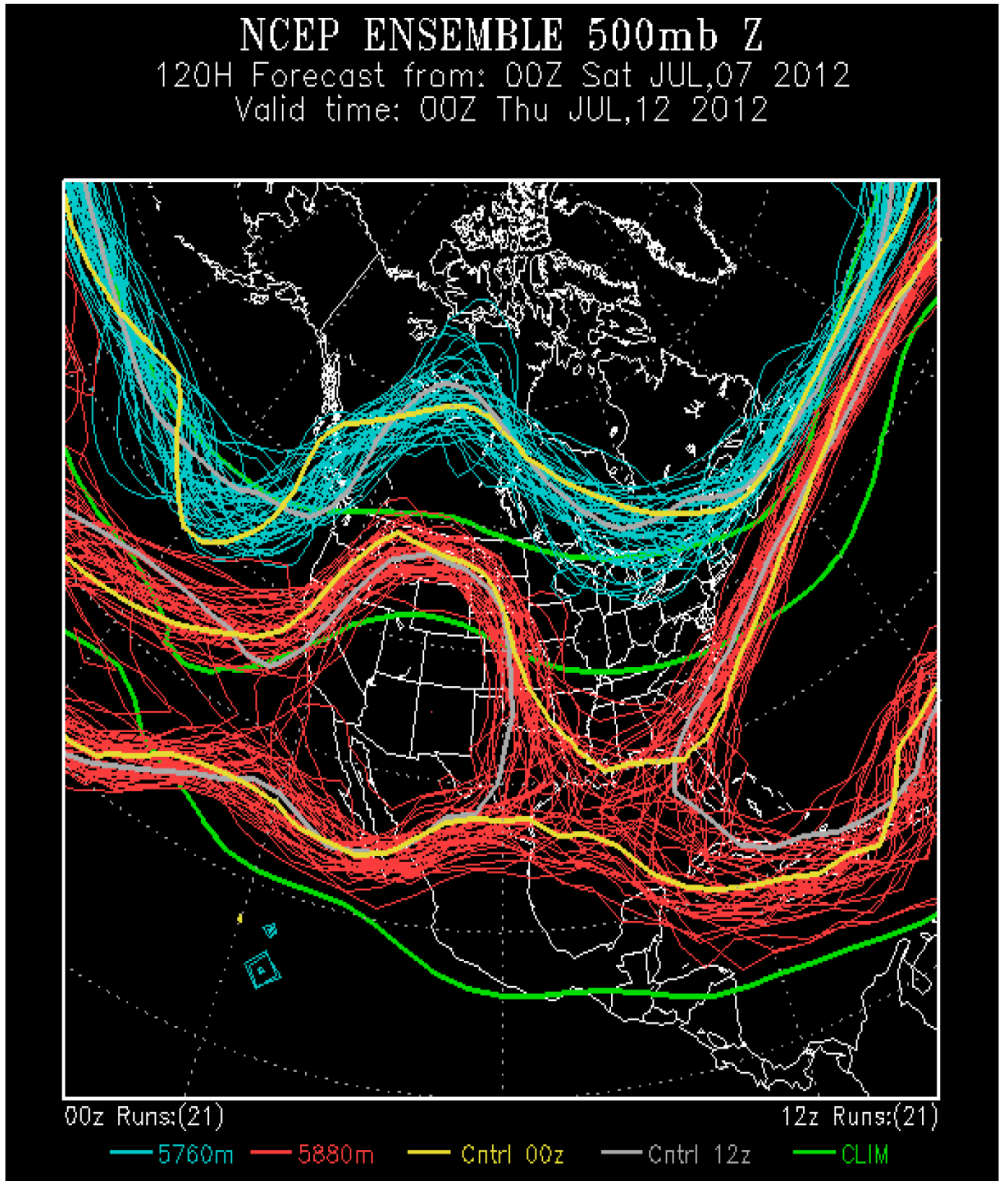


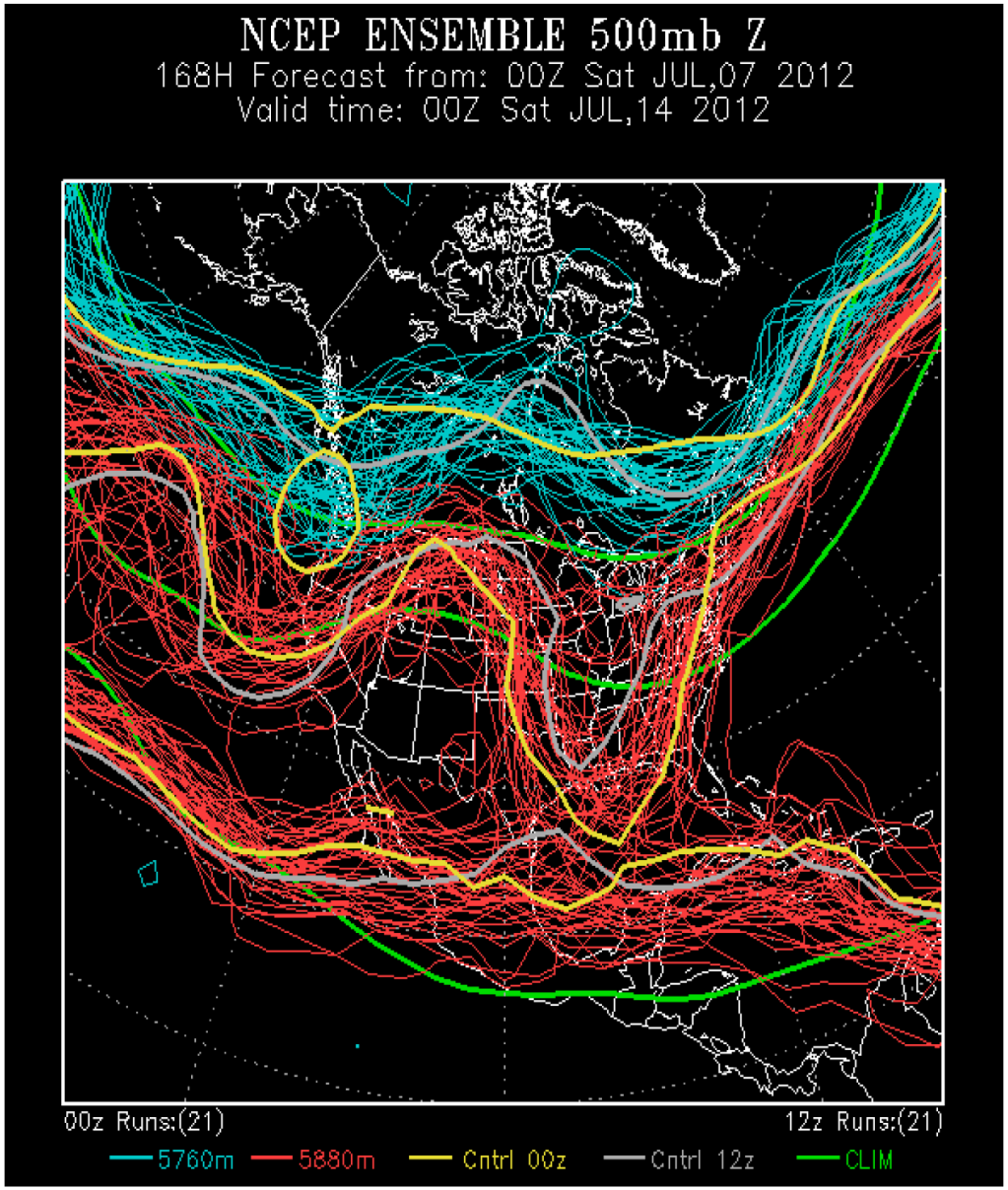
Weather is chaotic





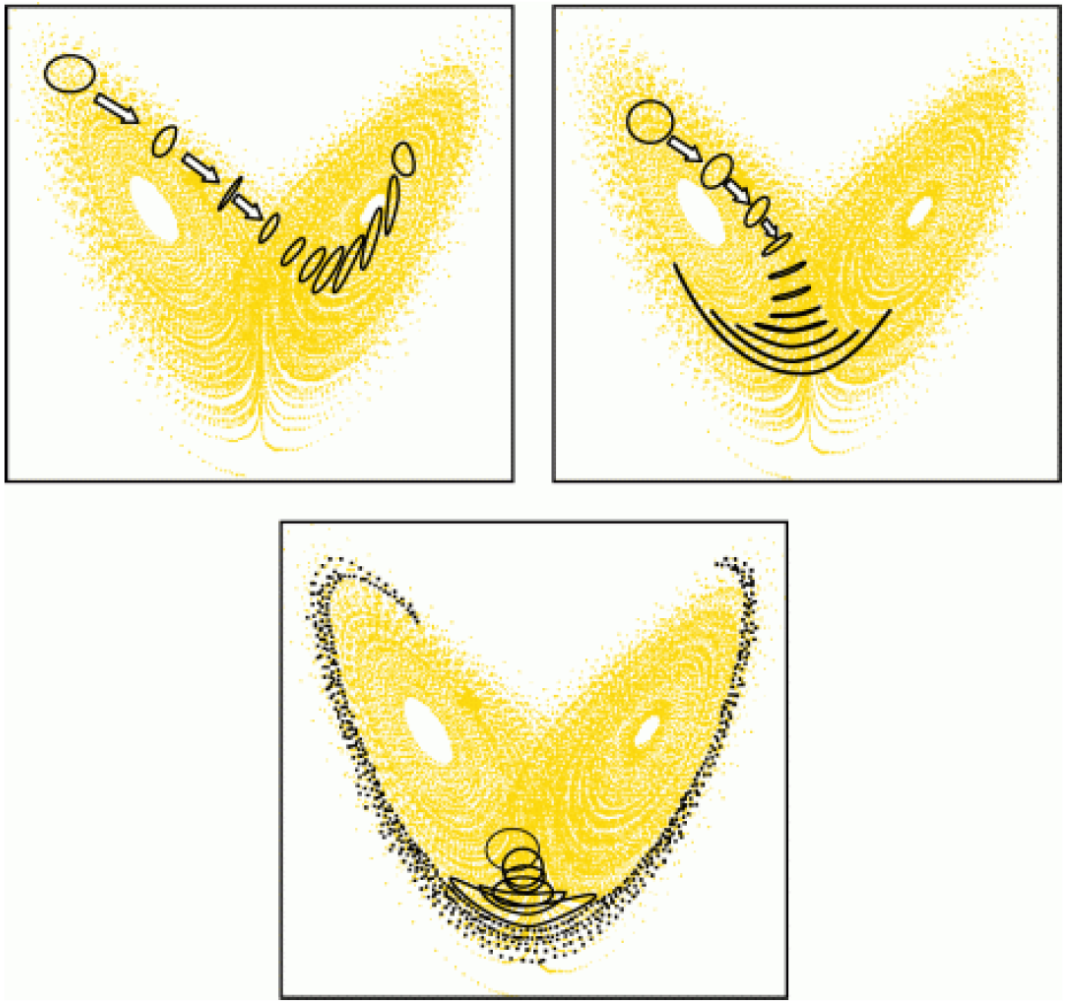
Weather is chaotic





Sensitivity to initial conditions can depend on the situation

More predictable

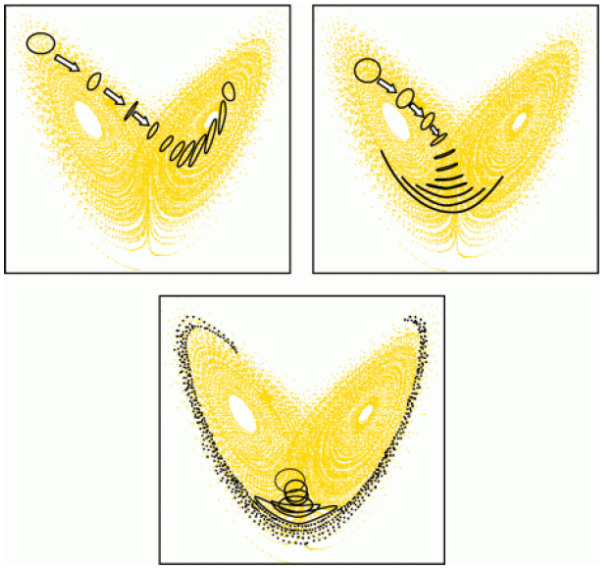


Less predictable

Very unpredictable

What can we do?

- Obtain more accurate initial conditions
  - More observations
  - Better data assimilation methods

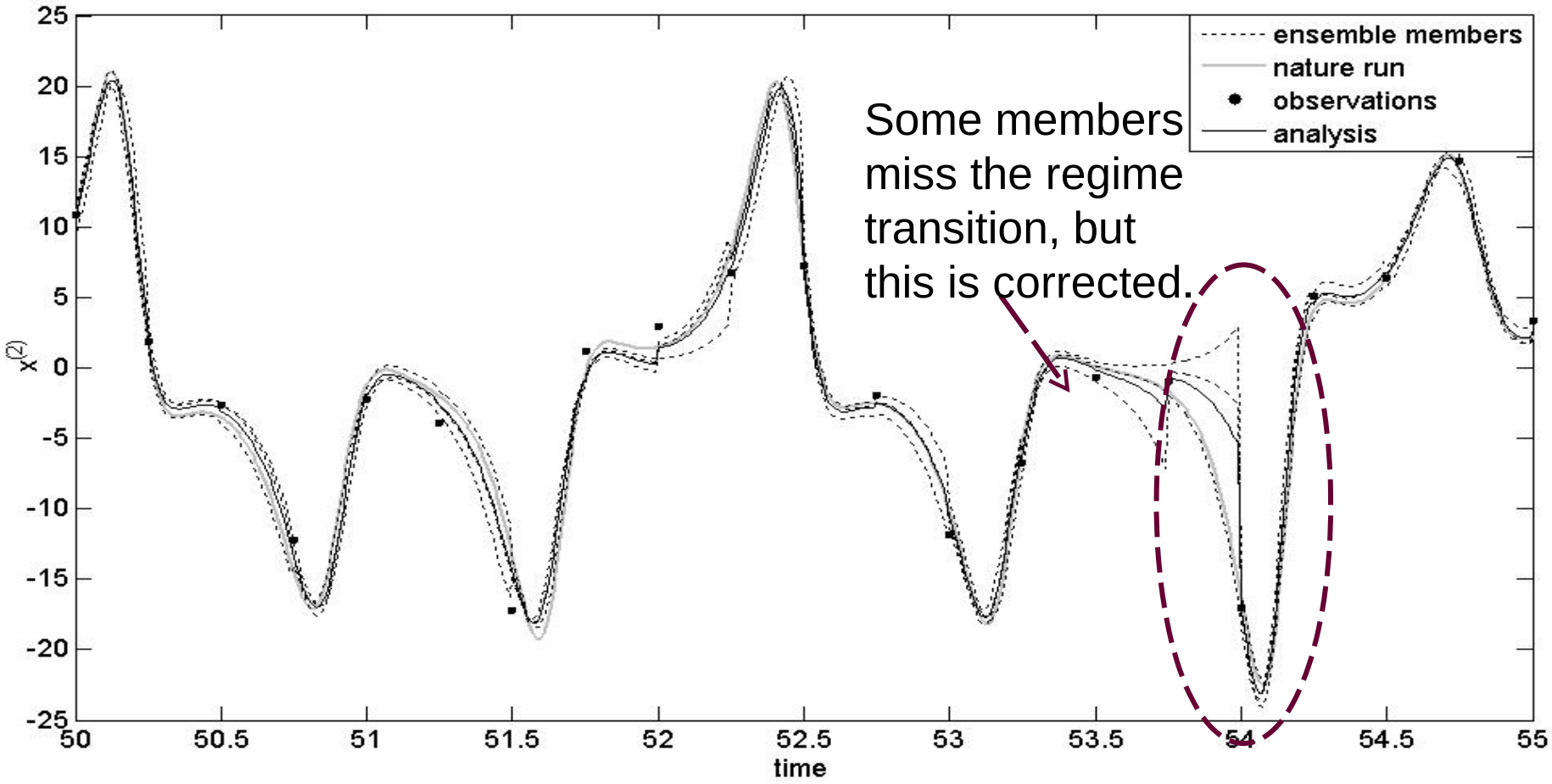


- Understand the error growth
  - Better understand the dynamics and physics

- Predict the predictability
  - Let users know how (un)certain the forecasts.

DA gives the tools to achieve these.

Revisiting **forecasts**



This example uses a 3-member ETKF in the Lorenz 1963 model. You will learn about this later.

DA: Combining **models** and **observations**

We need to develop a general theory on how to combine observations and models.

That framework does exist: **Bayes' Theorem!**

We will derive **Bayes' Theorem** and show how all existing methods can be shown to be approximations of Bayes' Theorem.

But first, let us start with intuitive ideas.

How do we process new data?

A process description

- **Prior knowledge**, from a model, **a cat.**
- **Observations**, the **dog.**
- **Posterior** knowledge, improvement of the model, **the dog that has eaten the cat.**

What is missing?

Uncertainty !!!

Bayes' theorem

Relationship between **joint** and **marginal** pdf:

$$pdf(u) = \int_{-\infty}^{\infty} pdf(u,v)dv \qquad pdf(v) = \int_{-\infty}^{\infty} pdf(u,v)du$$

Also:

$$\begin{aligned} pdf(u,v) &= pdf(v|u)pdf(u) \\ &= pdf(u|v)pdf(u) \end{aligned}$$

Using the two equalities for the joint pdf we get:

$$pdf(u|v) = \frac{pdf(v|u)pdf(u)}{pdf(v)}$$

This is **Bayes' theorem**, a really powerful result. It can be considered the **basis of DA**. Let us do a simple example to understand it before moving on.

Bayes' theorem in DA

Likelihood. Pdf of the observations given a value of the state variable.

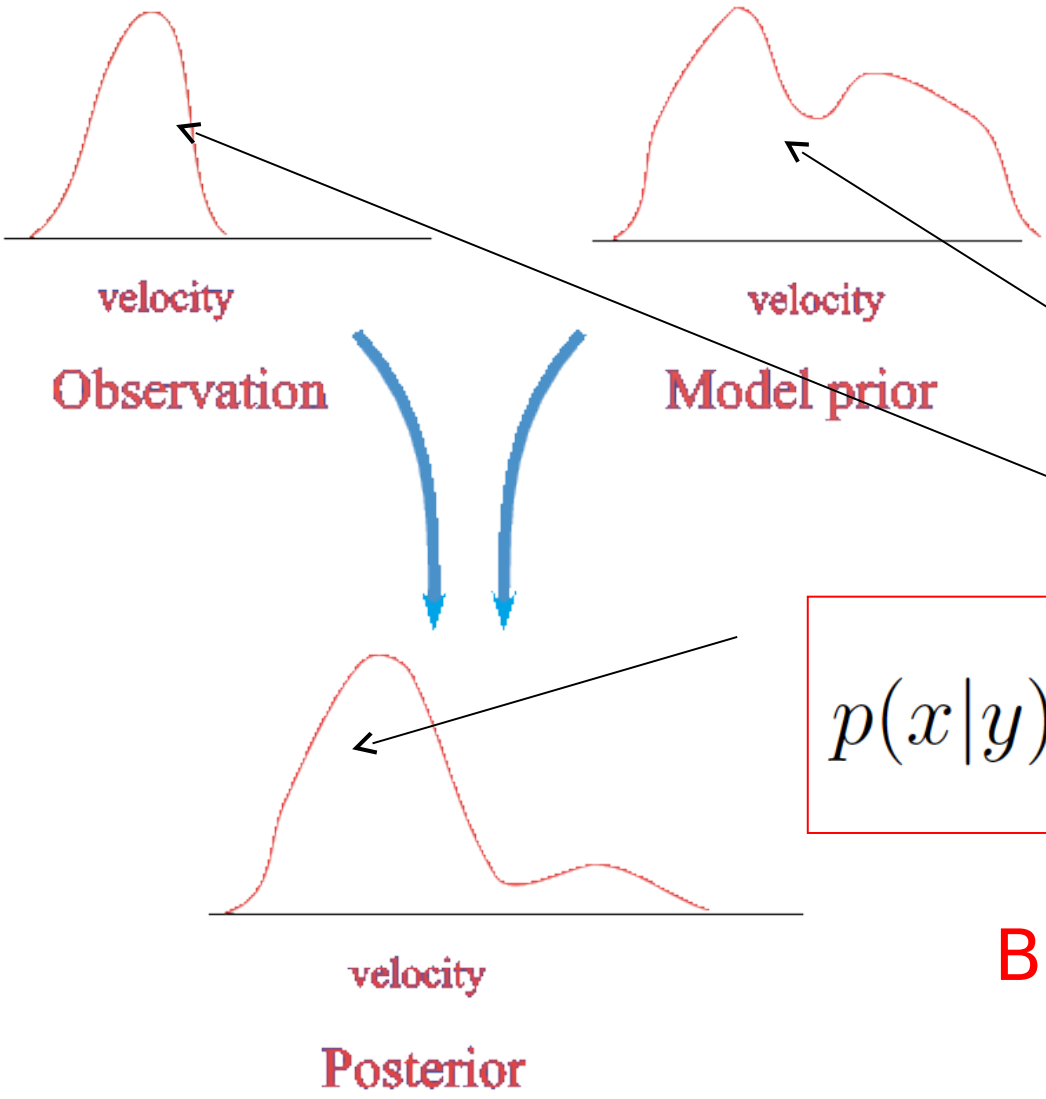
Prior pdf. Pdf of the state variables coming from the model

$$pdf(\mathbf{x}|\mathbf{y}) = \frac{pdf(\mathbf{y}|\mathbf{x})pdf(\mathbf{x})}{p(\mathbf{y})}$$

Posterior pdf. Pdf of the state variables given the observations.

Marginal pdf of the observations. It is often the case we do not need to compute this, since it acts as a normalisation constant.

Bayes' theorem in DA

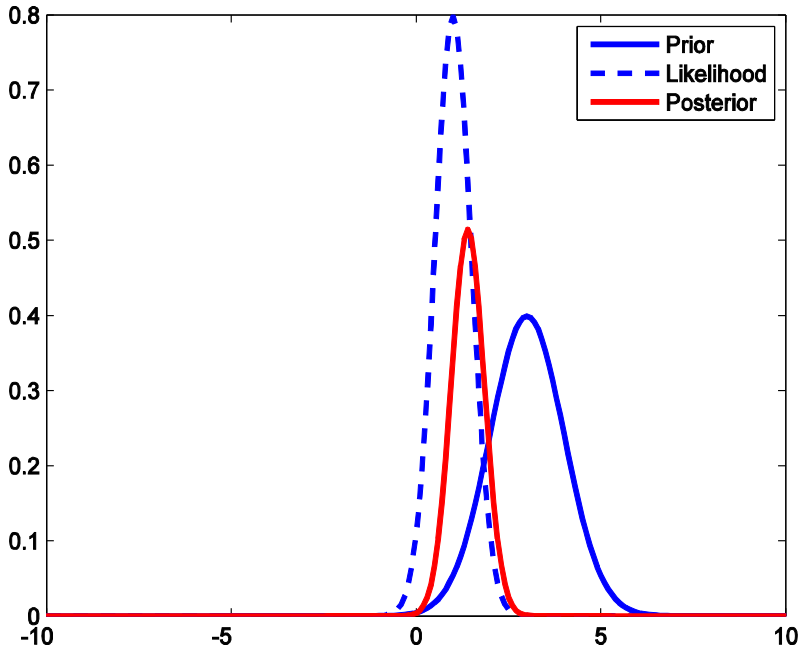


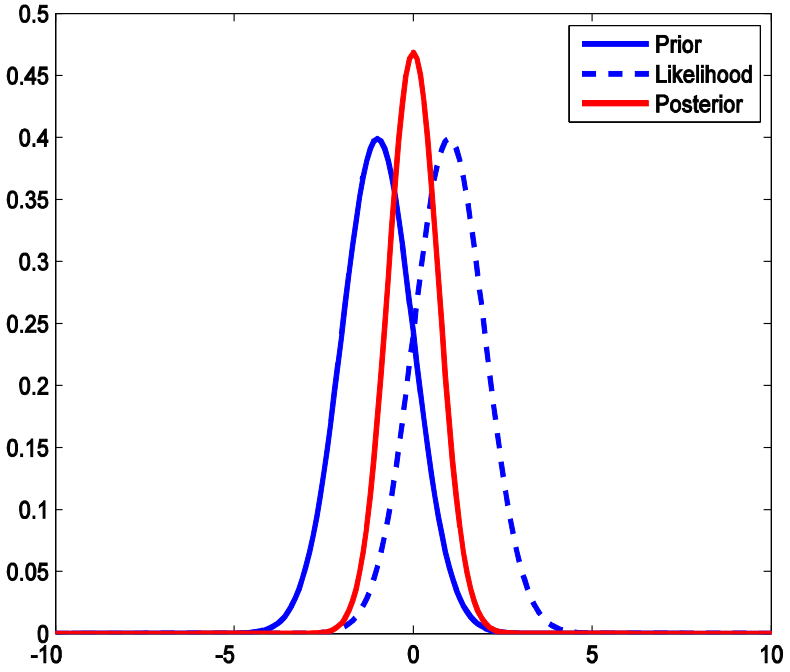
$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)\ dx}$$

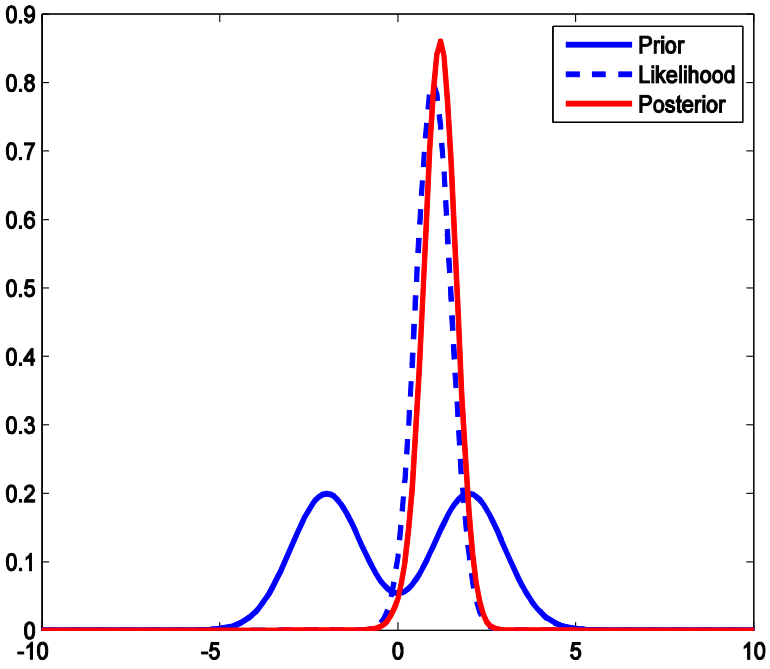
Bayes' theorem

Examples of Bayes' theorem in action

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$







Reality bites

$$pdf(\mathbf{x}|\mathbf{y}) = \frac{pdf(\mathbf{y}|\mathbf{x})pdf(\mathbf{x})}{p(\mathbf{y})}$$

Estimating these pdf's in large dimensional systems is virtually impossible. **Approximate solutions** lead to DA methods:

- **Variational** methods: solves for the **mode** of the posterior.

- **Kalman-based** methods: solve for the **mean** and **covariance** of the posterior.

- **Particle filters**: find a weak **(sample) representation** of the posterior pdf.

The Gaussian world

Considering errors to be Gaussian can be quite convenient.
The pdf is completely determined by the **mean** and **covariance**.

Prior

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}|^{n/2}} \exp\{ -\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T \mathbf{P}^{-1}(\mathbf{x}-\mathbf{x}_b)\}$$

Likelihood

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{R}|^{p/2}} \exp\{ -\frac{1}{2}(\mathbf{y}-H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))\}$$

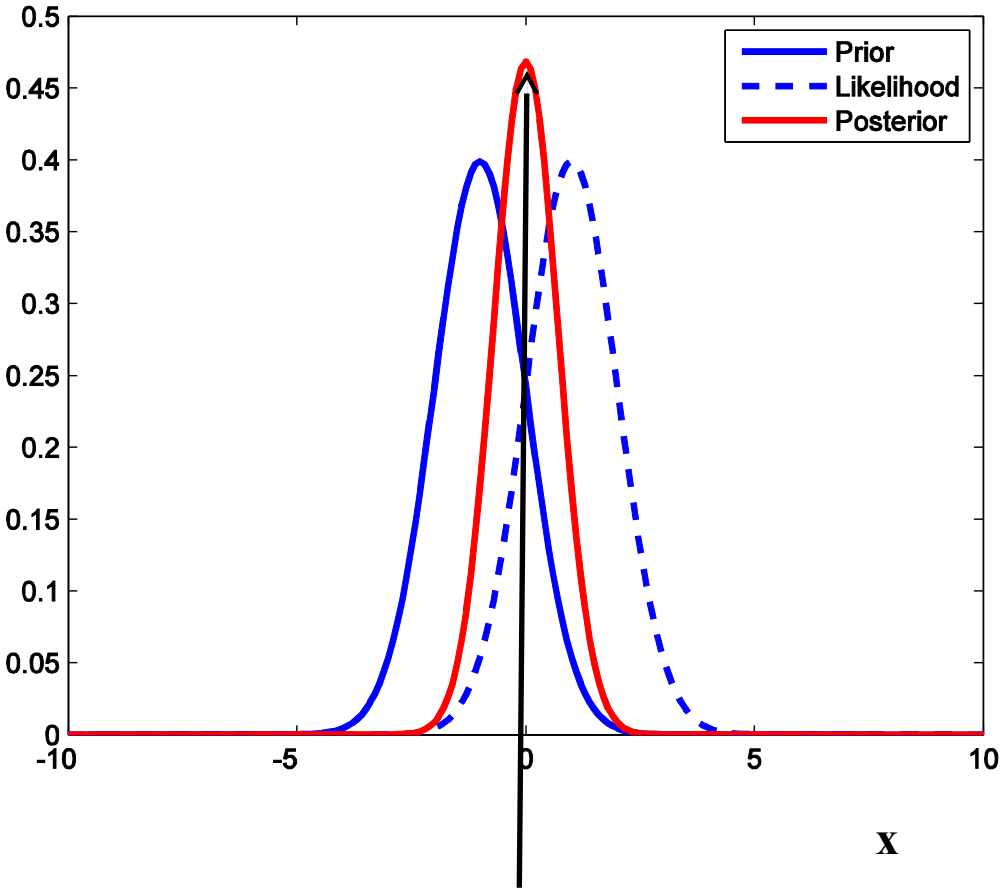
Posterior

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\{ -\frac{1}{2}\{(\mathbf{x}-\mathbf{x}_b)^T \mathbf{P}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{y}-H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))\}\}$$

Maximum a-posteriori estimator (MAP)

For a Gaussian distribution the mean and mode coincide.

$p(\mathbf{x}|\mathbf{y})$



The Gaussian world

Recalling the posterior in this case.

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\{ -\frac{1}{2}\{(\mathbf{x}-\mathbf{x}_b)^T\mathbf{P}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))\}\}$$

We need the minimiser of the exponent (which we call cost-function)

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

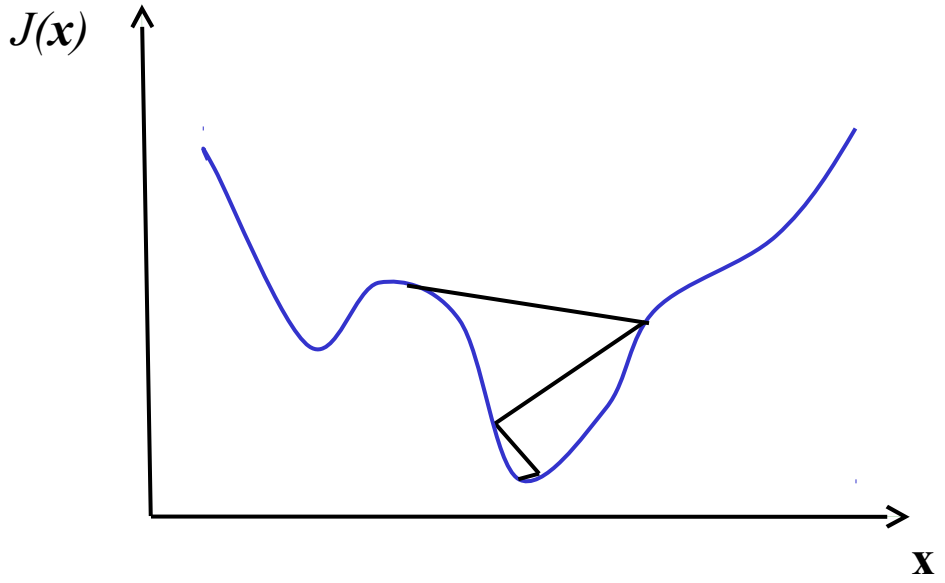
Which for linear **H** is:

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

The matrices are huge! How to solve in practice?

1. Variational methods

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



Find the minimum of the cost function via (iterative) optimisation techniques. One needs the gradient of the cost function.

The background error covariance is static.

2. Kalman filter

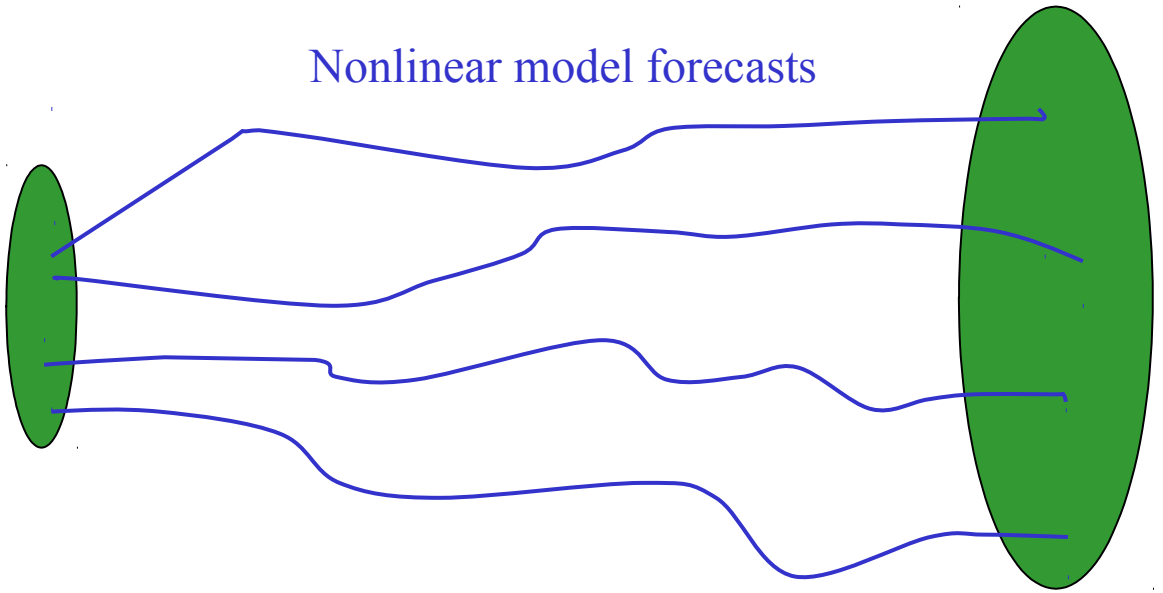
Solve directly.

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

- It is exact in the linear case.
- The covariance is updated.
- It can be extended to non-linear case via linearisation.

3. Ensemble Kalman filter

Use sample estimators with the KF equations.



Uncertainty at analysis time

Uncertainty at forecast time with covariance **P** (Gaussian)

3. Hybrid methods

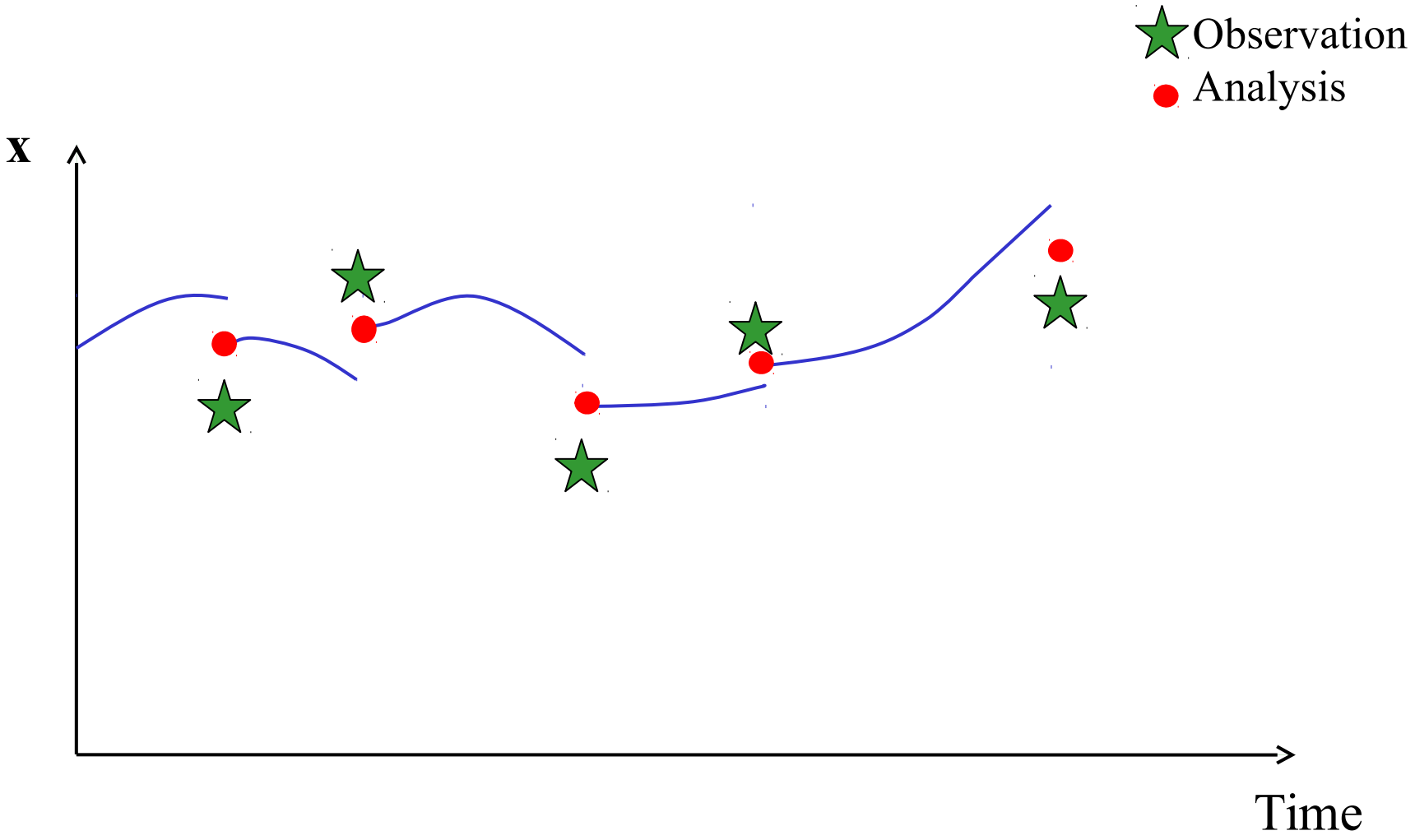
- Different flavours.
- For example, use sample covariances within the variational framework.
- Use 4D (space-time) covariances.

4. Particle filters

- Generate samples from the posterior (using tricks like importance sampling).
- Does not require the Gaussian assumption.

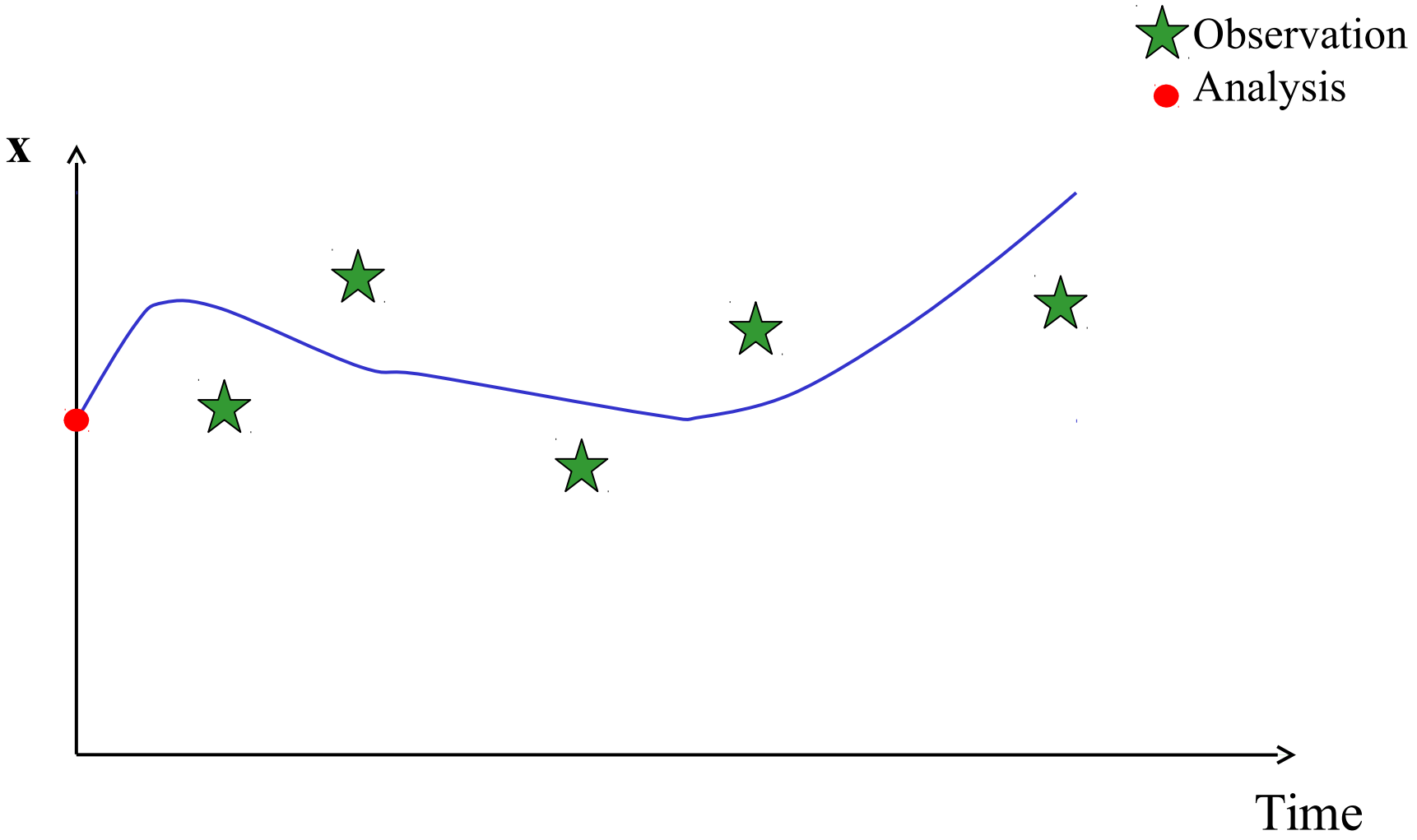
Filters

Assimilate every time observations are available.



Smoothers

Assimilate observations over a time window.



Characteristics of traditional DA methods

	Method		Observations		Covariance	
	Variational	Kalman	Sequential	Smoother	Static	Dynamic
3DVar	✓		✓		✓	
4DVar	✓			✓	(✓)	✓
Optimal Interpolation		✓	✓		✓	
Kalman Filters		✓	✓			✓
Kalman Smoother		✓		✓		✓

Solution is got using (iterative) **minimisation** techniques.

Solution is got using explicit **linear algebra**.

Estimation is done for an **instant**.

Estimation is done within a **time window**.

Uncertainty is considered **fixed** in time.

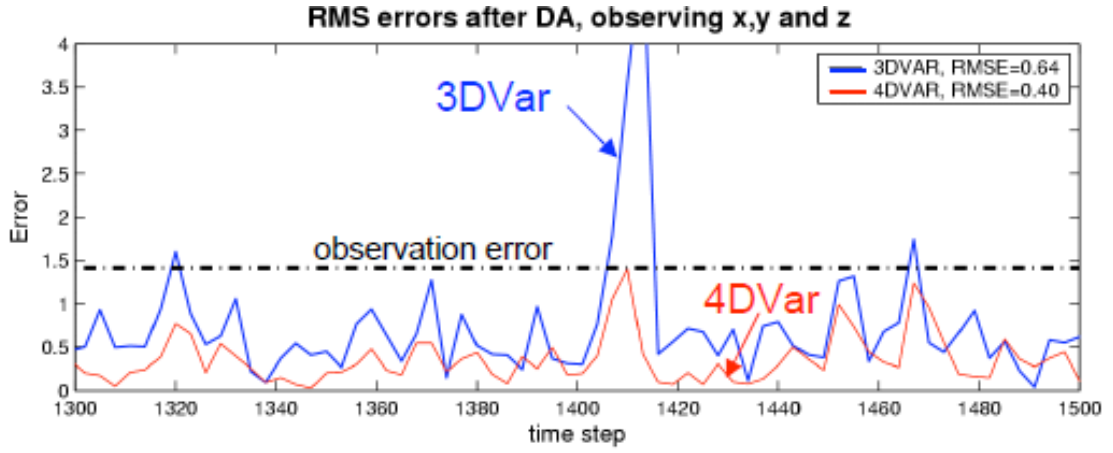
Uncertainty evolves in time.

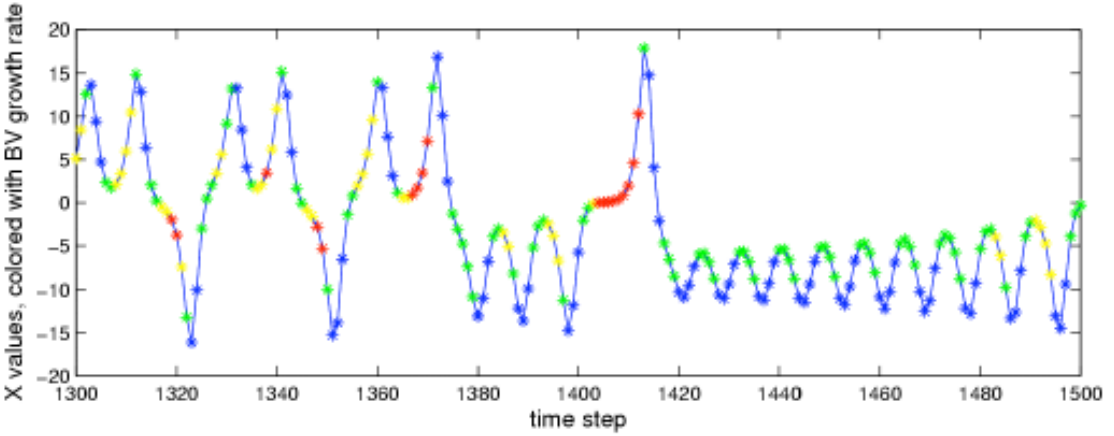
Var

3D vs 4DVar

4DVar has important information from the future (after all, it is a smoother), 3DVar does not.

The figure shows a comparison of the performance of the two methods. Taken from Evans et al, 2005.





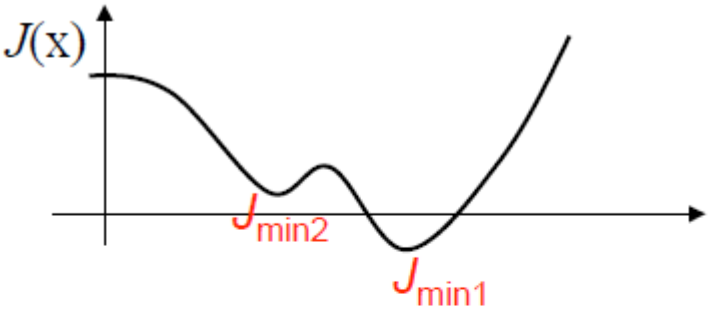
DA cycle and observations: $8\Delta t$, $\mathbf{R}=2*\mathbf{I}$
4D-Var assimilation window: $24\Delta t$

How long should the assimilation window be?

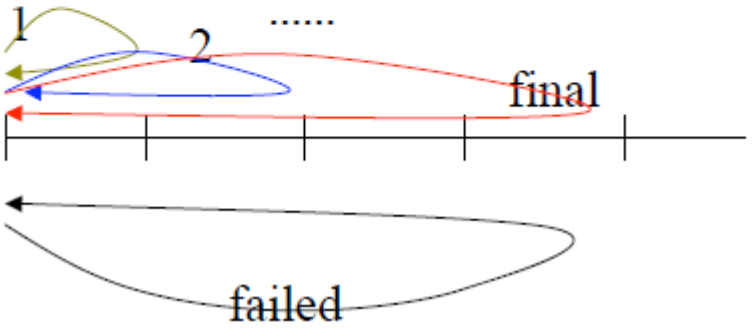
The longer the 4D assimilation window the more observations we'll have… but also the more nonlinear the forecast will be. The best should be somewhere in the middle.

	Win=8	16	24	32	40	48	56	64	72
Fixed window	0.59	0.59	0.47	0.43	0.62	0.95	0.96	0.91	0.98
Start with short window	0.59	0.51	0.47	0.43	0.42	0.39	0.44	0.38	0.43

Performance of 4DVar using the Lorenz 1963 and different lengths of assimilation window (Kalnay *et al.*, 2007).



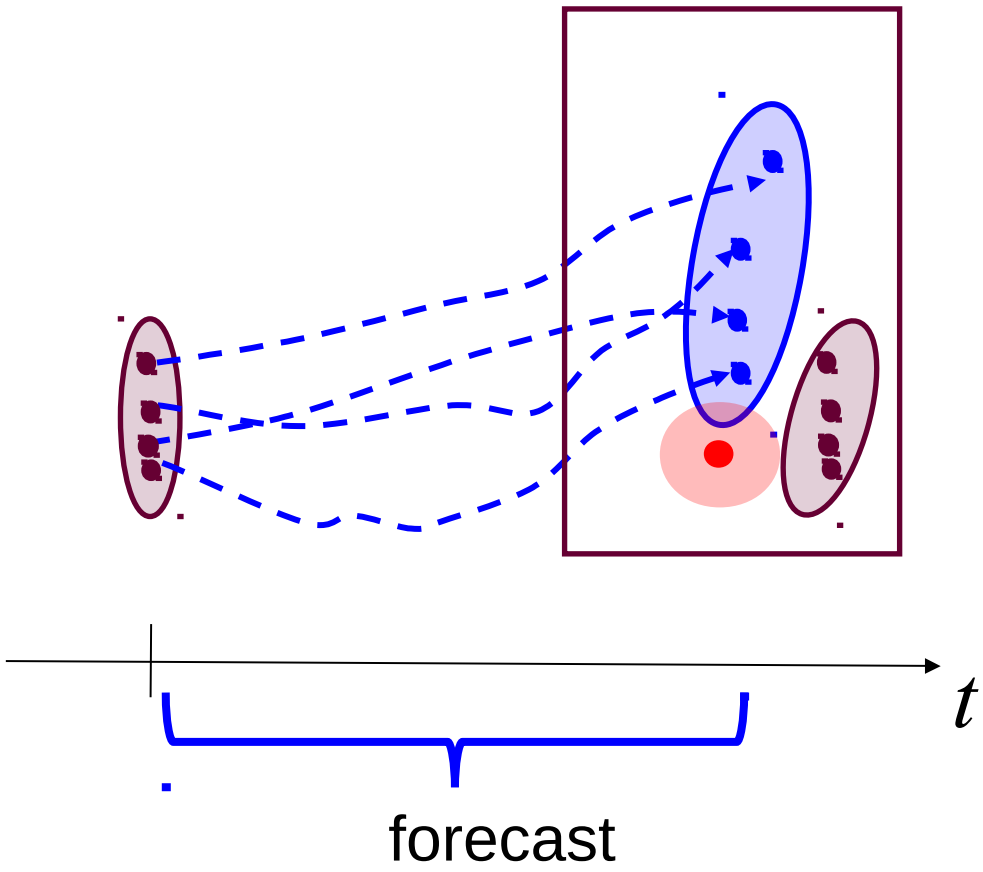
It is recommendable to do the minimization progressively while increasing the assimilation window (Pires *et al.*, 1996).



Ensemble Kalman Filter

Ensemble Kalman Filter

The Ensemble Kalman Filter is a Monte-Carlo implementation of the KF. Study a family of *M* solutions, which we call ensemble (Evensen, 1994).



assimilation

$$\mathbf{X} = \left[\mathbf{x}_1 \mid \mathbf{x}_2 \mid \cdots \mid \mathbf{x}_M \right] \in \Re^{N \times M}$$

Ensemble Kalman Filter

$$\mathbf{X} = \left[\mathbf{x}_1 \,|\, \mathbf{x}_2 \,|\, \cdots \,|\, \mathbf{x}_M \right] \in \Re^{N \times M}$$

We use the **sample estimators** of the mean and covariance for the KF analysis step.

$$\bar{\mathbf{x}} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{x}_m \qquad \mathbf{P} = \frac{1}{M-1} \hat{\mathbf{X}} \hat{\mathbf{X}}^T$$

Where we define an ensemble of perturbations:

$$\hat{\mathbf{X}} = \left[\mathbf{x}_1 - \bar{\mathbf{x}} \,|\, \mathbf{x}_2 - \bar{\mathbf{x}} \,|\, \cdots \,|\, \mathbf{x}_M - \bar{\mathbf{x}} \right] \in \Re^{N \times M}$$

EnKF features

- Unlike the EKF, the EnKF **makes use of the full nonlinear model**: more accurate determination of error growth (and of error saturation).

- The Kalman gain **K** is computed **without the need to calculate $\mathbf{P}^b$**.

- No need to linearize the observation operator *h.*

- It is a **highly parallel algorithm**.

- Model error term can be included as a perturbation of the deterministic forecast.

Ensemble Kalman Filter

To assimilate the mean and covariance, we just follow the KF equations.

background analysis

$\mathbf{P}^b \rightarrow \mathbf{P}^a$

$\bar{\mathbf{x}}^b \rightarrow \bar{\mathbf{x}}^a$

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^b\right)$$

$$\mathbf{P}^a = \left(\mathbf{I} - \mathbf{K}\ \mathbf{H}^T\right)\mathbf{P}^b$$

However, how to assimilate the individual ensemble members is not trivial.

$\mathbf{X}^b \rightarrow \mathbf{X}^a$

Not unique since the square root of a matrix is not unique.

Ensemble Kalman Filter: families

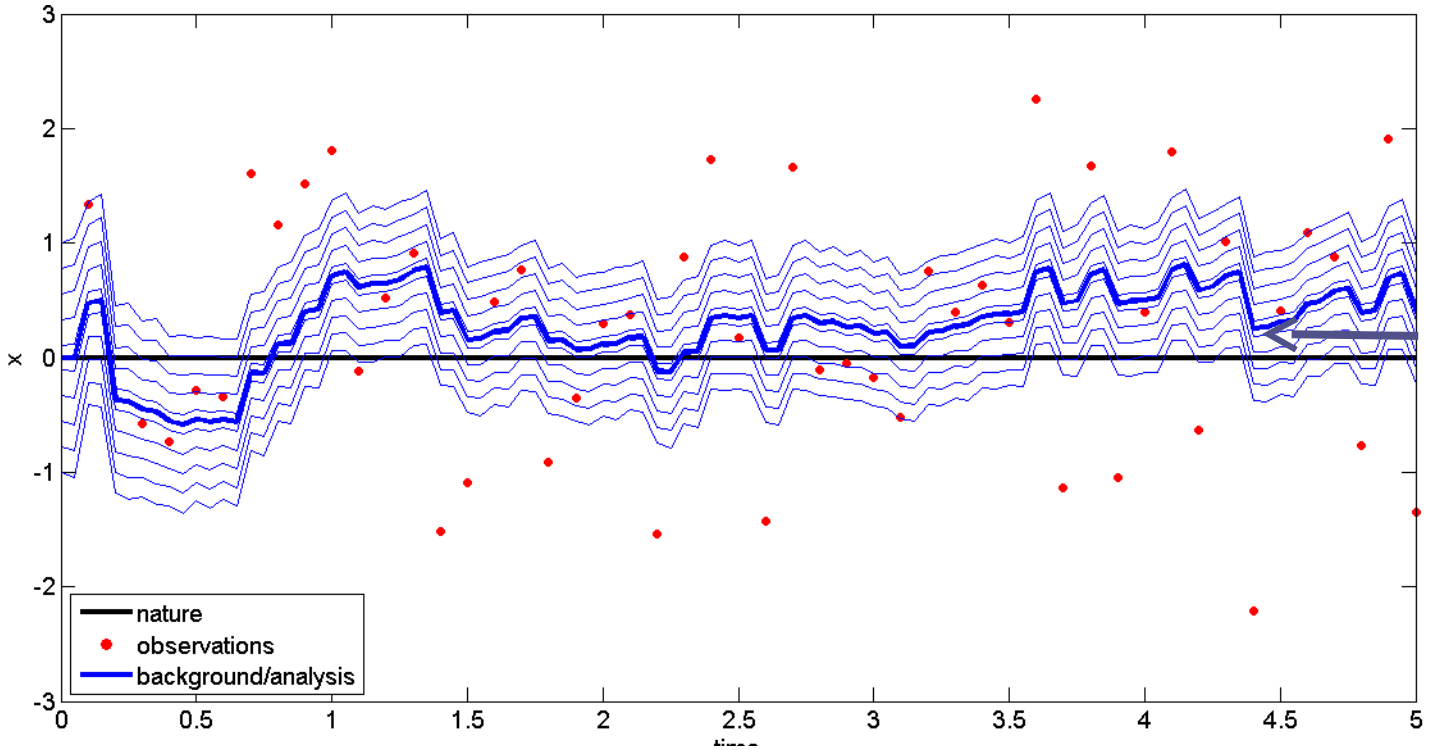
Stochastic (Burgers *et al.,* 1996; Houtekamer and Mitchell, 1998):

- Apply the KF equations to each ensemble member.
- Requires perturbed observations for each member.

Deterministic implementation: ensemble square root filters (Tippet *et al.*, 2003).

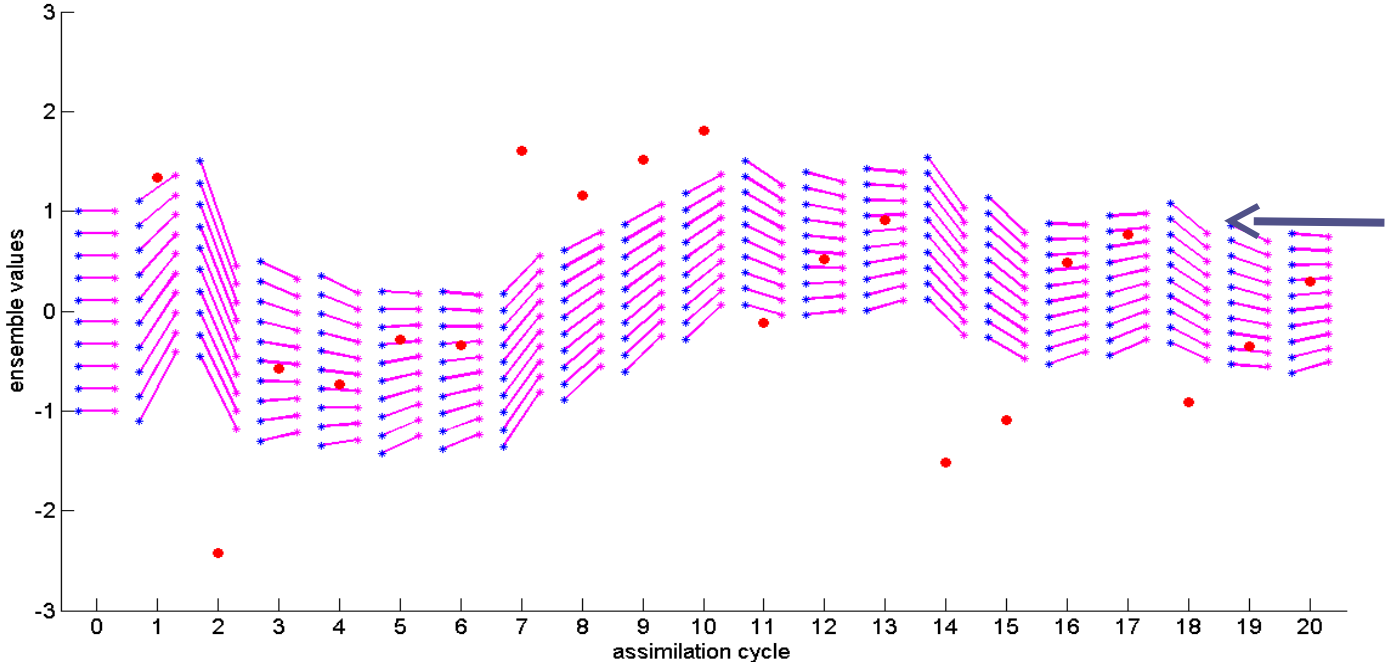
- Serial EnSRF (Whitaker and Hamill, 2002)
- Ensemble Adjustment Kalman Filter (Anderson, 2001)
- Ensemble Transform Kalman Filter (Bishop et al, 2001), LETKF (Hunt et al, 2007)

ETKF in a simple univariate model



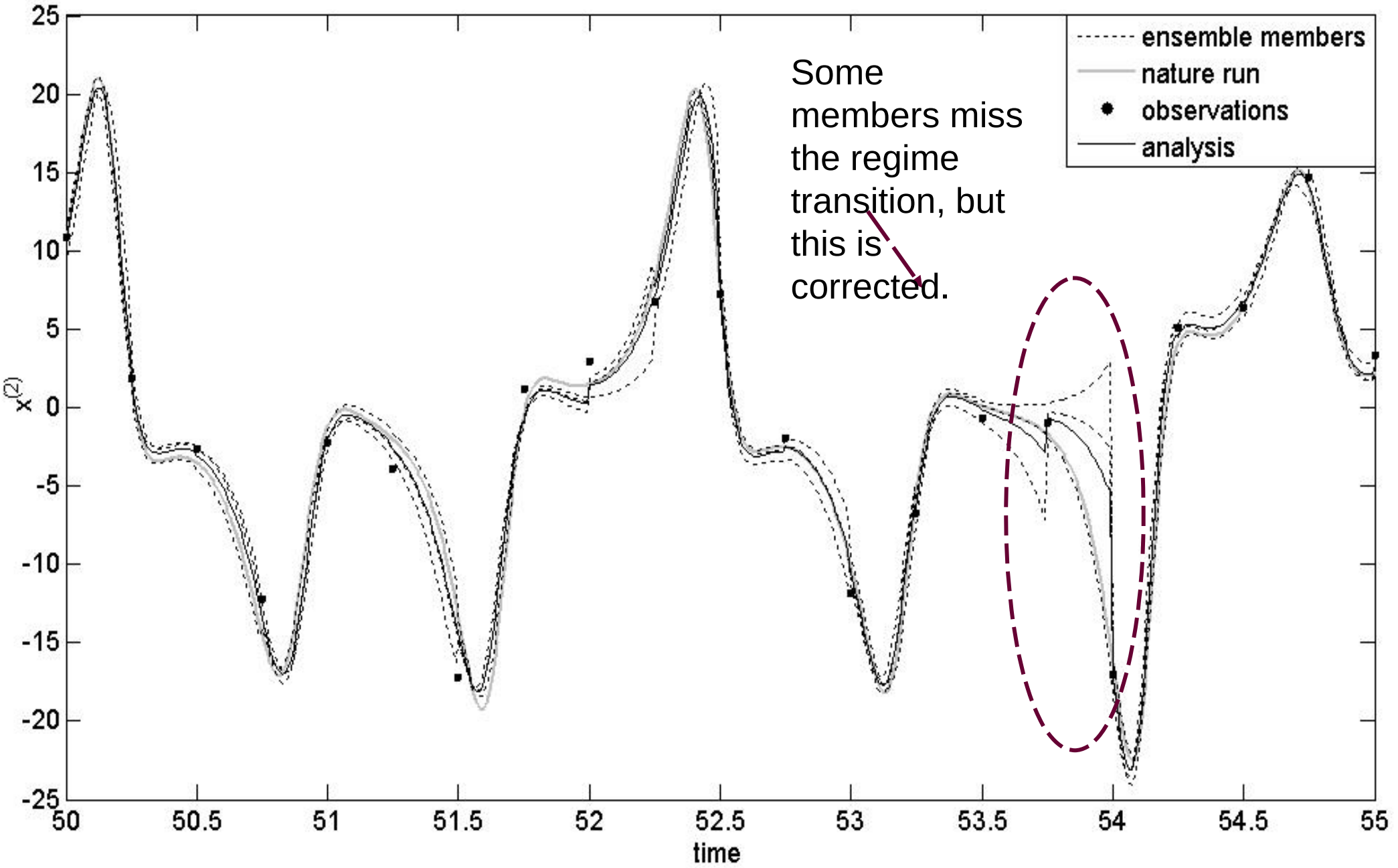
$$x_{t+1} = (1+\Delta)x_t;\ x_{true} = 0$$

Time evolution



Only analysis

ETKF in Lorenz 1963 with 3 members



Sampling

- Two effects of **finite sample size**:
- **- Underestimation** of sample **covariance**.
- **- Spurious** long-range **correlation**
- **- In-breeding** effects.

Fixes:

- Covariance inflation
- Covariance localization

- Also, the sample covariance matrix is singular for $N>M$…

  How many members would we need? At least as many as the **number of unstable directions of error growth**?

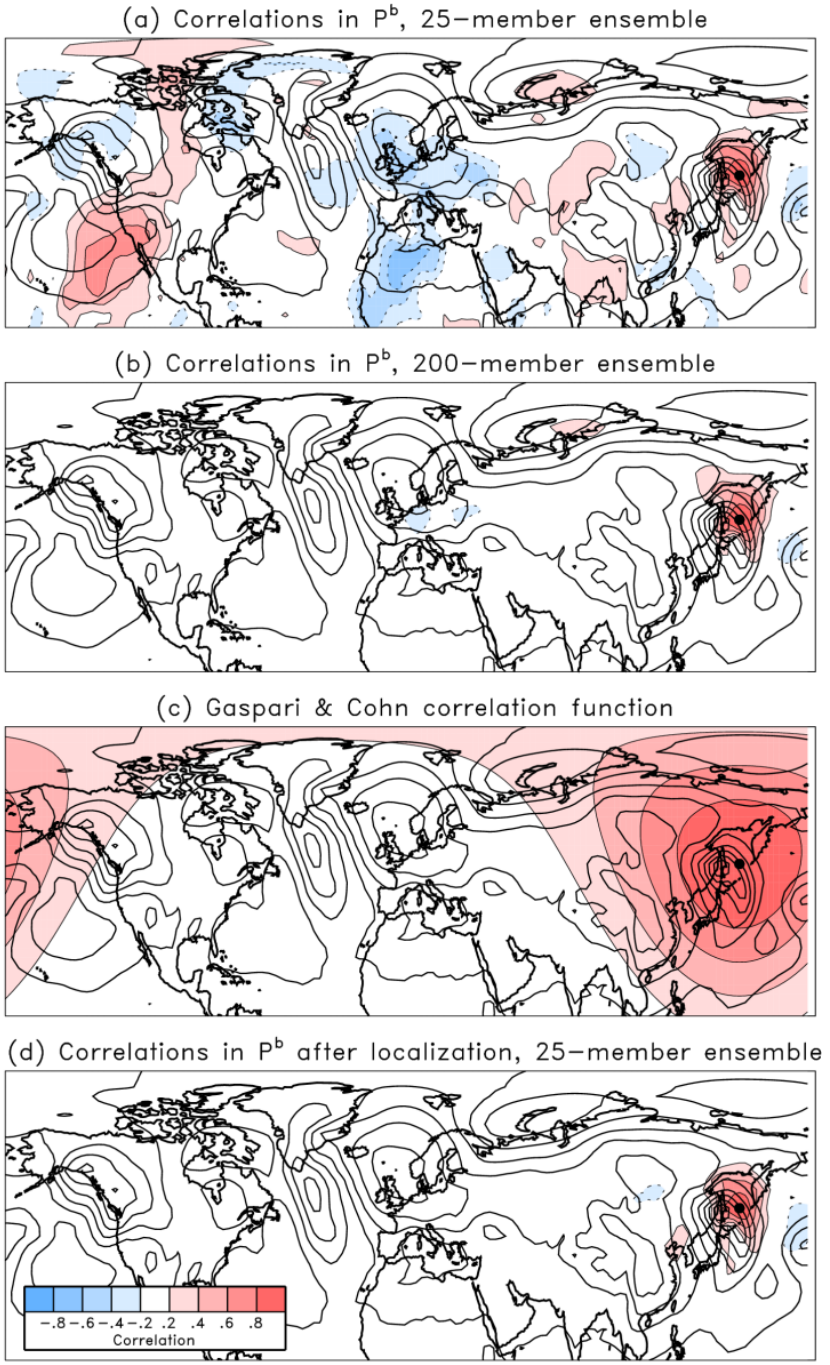
Covariance localization

(a) $\mathbf{P}^b_e$ (N=25)

(b) $\mathbf{P}^b_e$ (N=200)

(c) Correlation function with compact support

(d) localized $\mathbf{P}^f_e$ (N=25)



From Fig. 6.4 of Hamill, 2006

Parameter estimation

Extending the state vector

The state vector can be extended with the parameters of the model. We can use the covariance to update values of poorly known parameters.

$$\mathbf{x}_t = f\left(\mathbf{x}_{t-1};\boldsymbol{\theta}\right)$$

$$\widetilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \cdots \\ \boldsymbol{\theta} \end{bmatrix}$$

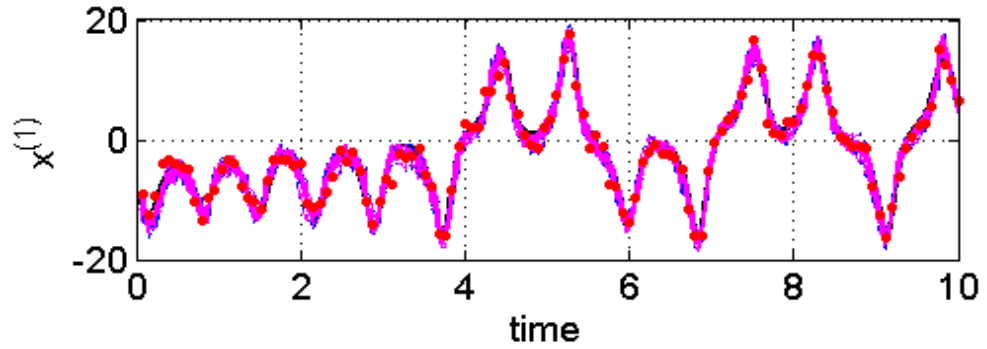
$$P = \begin{bmatrix} Cov\left(\mathbf{x},\mathbf{x}\right) & Cov\left(\mathbf{x},\boldsymbol{\theta}\right) \\ Cov\left(\mathbf{x},\boldsymbol{\theta}\right) & Cov\left(\boldsymbol{\theta},\boldsymbol{\theta}\right) \end{bmatrix}$$

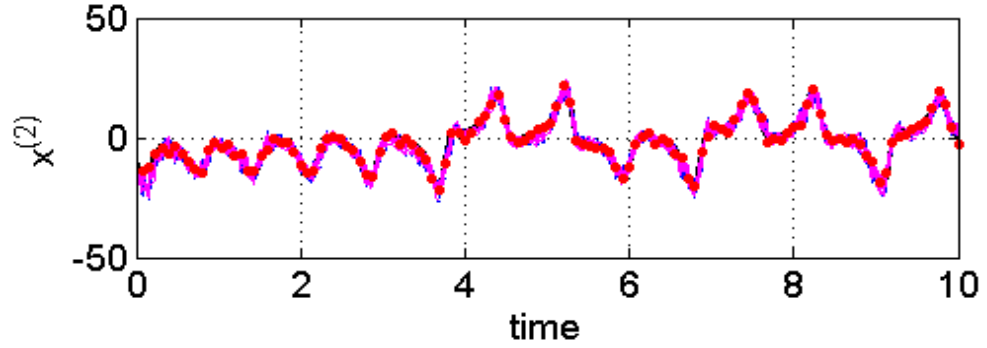
This ‘cross-covariance’ carries information from state variables to parameters. Remember: **parameters are not observables.**

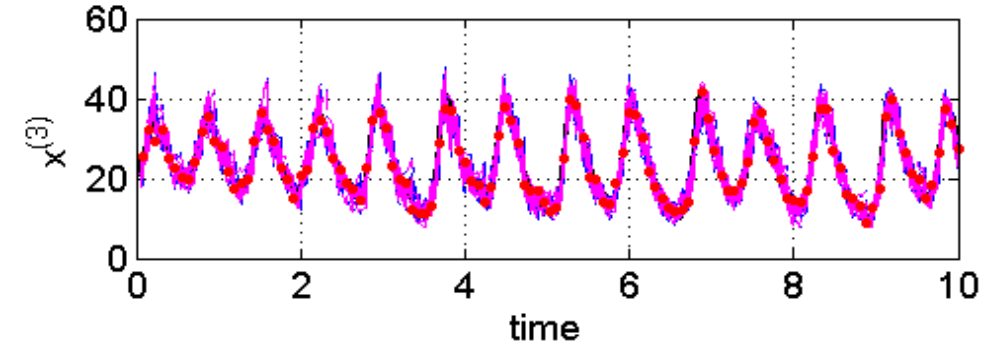
There are no dynamics for the parameters, one can perturb them during the forecast.

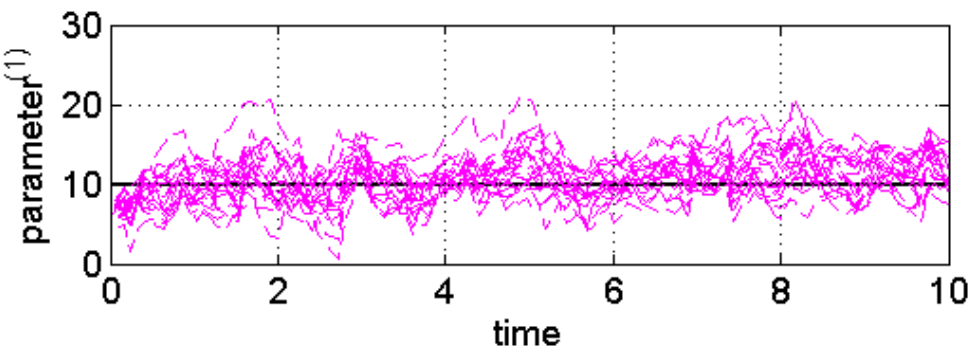
Example

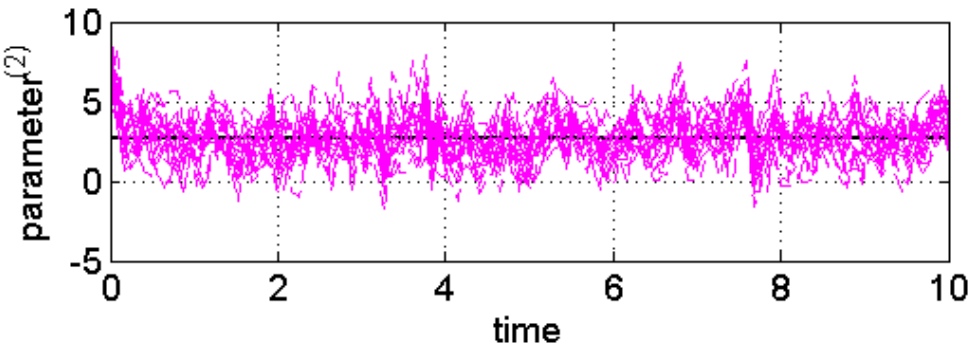
Using Lorenz 1963, estimate the values of the state variables and the parameters.

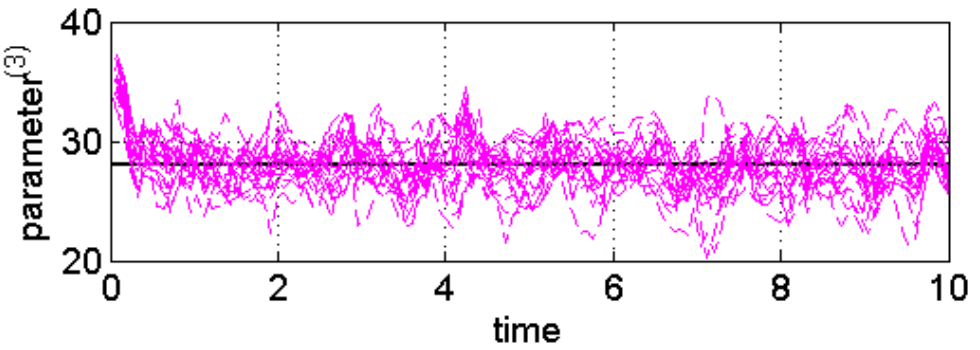












Some words on ensembles

Nowadays, NWP does ensemble forecasts to quantify uncertainty. They are readily available for DA.

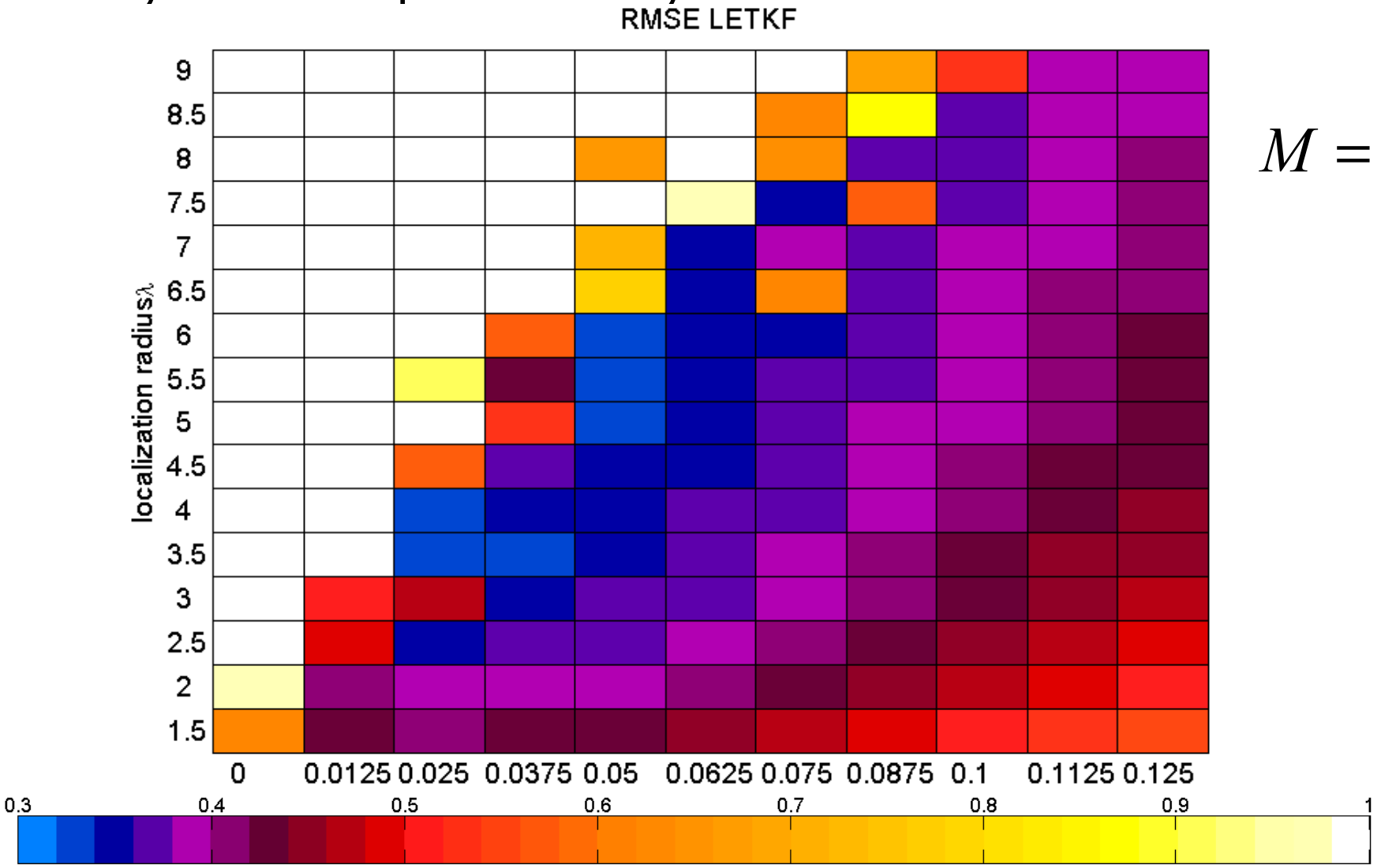
Sample covariance has information about the errors of the day, it ‘knows’ about the flow. Nonetheless, it has sampling error and can be singular.

Parameter estimation can be implemented in a straightforward fashion.

EnKF's do not require adjoints.

Combined effects of inflation and localization

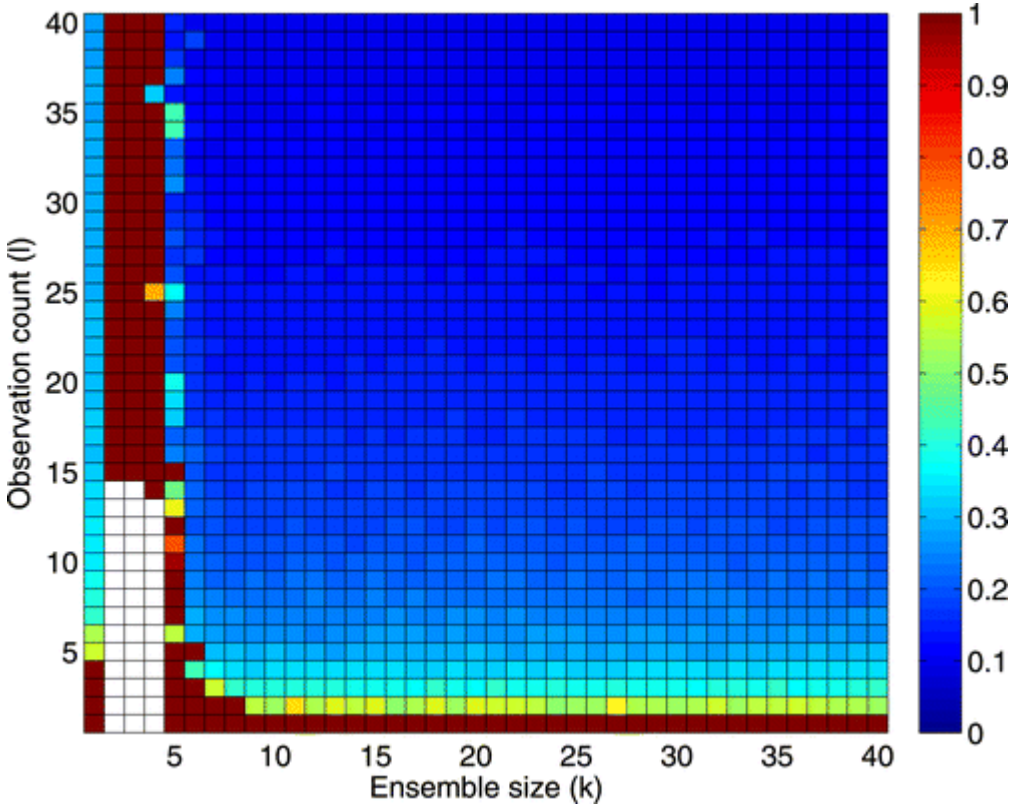
Experiments with Lorenz 1996 and 40 variables, observing every 2 time steps and every other variable.

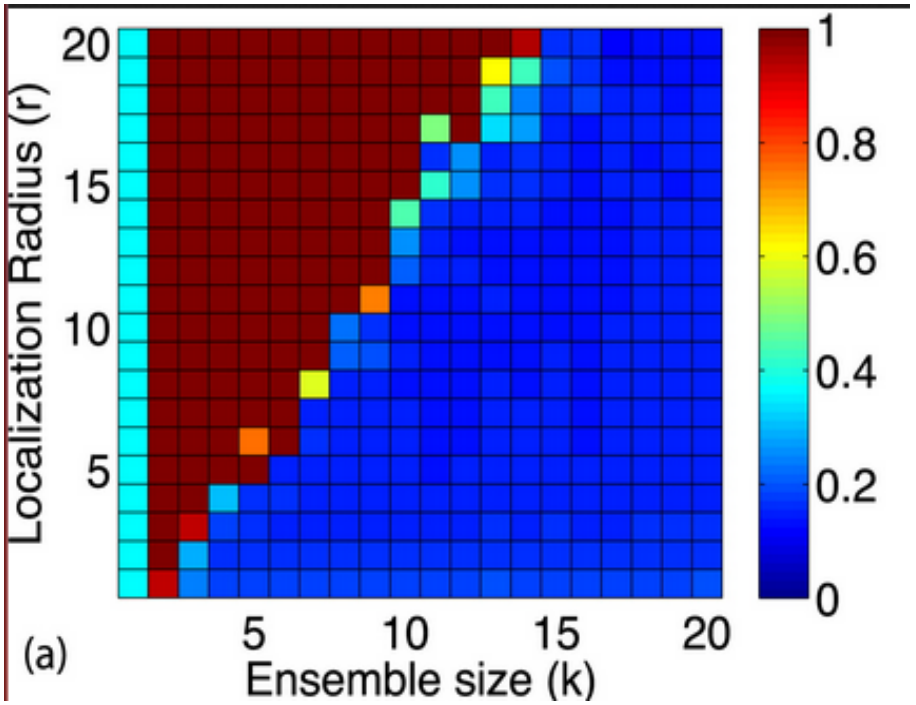


$M = 10$

Amezcua et al, 2012.

Interactions of different parameters in the EnKF





Penny, 2014

Combining the best of 2 worlds?

A static covariance is full rank, it is invertible, it gives idea of the climatology.

Climatology

An ensemble covariance has information of the flow, but it can be singular and contains sampling errors.

Flow/State Dependence

$$\mathbf{B} = \alpha \mathbf{B}_{static} + (1-\alpha)\mathbf{B}_{ensemble}$$ → Compromise?

There are several ways to implement this.