Introduction to Data Assimilation

.

Javier Amezcua
Data Assimilation Research Centre
University of Reading
March 2018

Data Assimilation Research Centre

University of Reading

Data Assimilation

DA is the 'art' of **combining information** from different sources in an '**optimal**' way. Generally, these sources are **models** and **observations**.

This has the aim of getting a **better estimate** of the state of a **system**.

Optimal includes –among other things- **considering the uncertainty** (or conversely, the precision) of the sources.

Data Assimilation

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

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

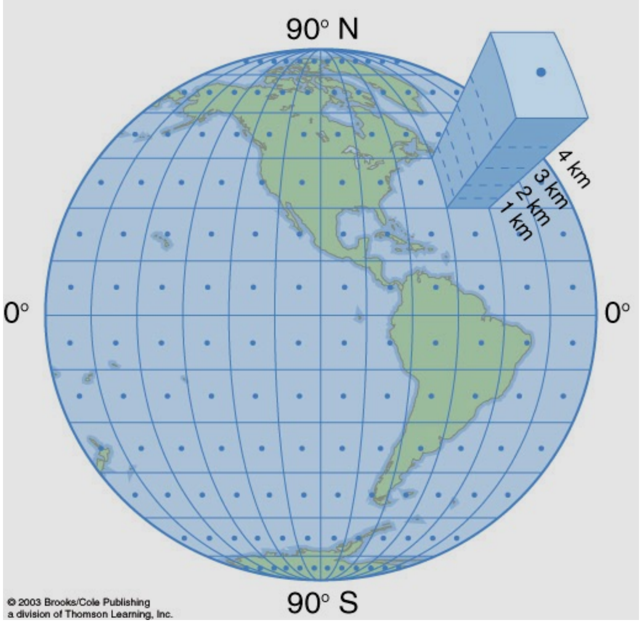
a. To find a **current estimate** that can be used to **produce forecasts**.

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

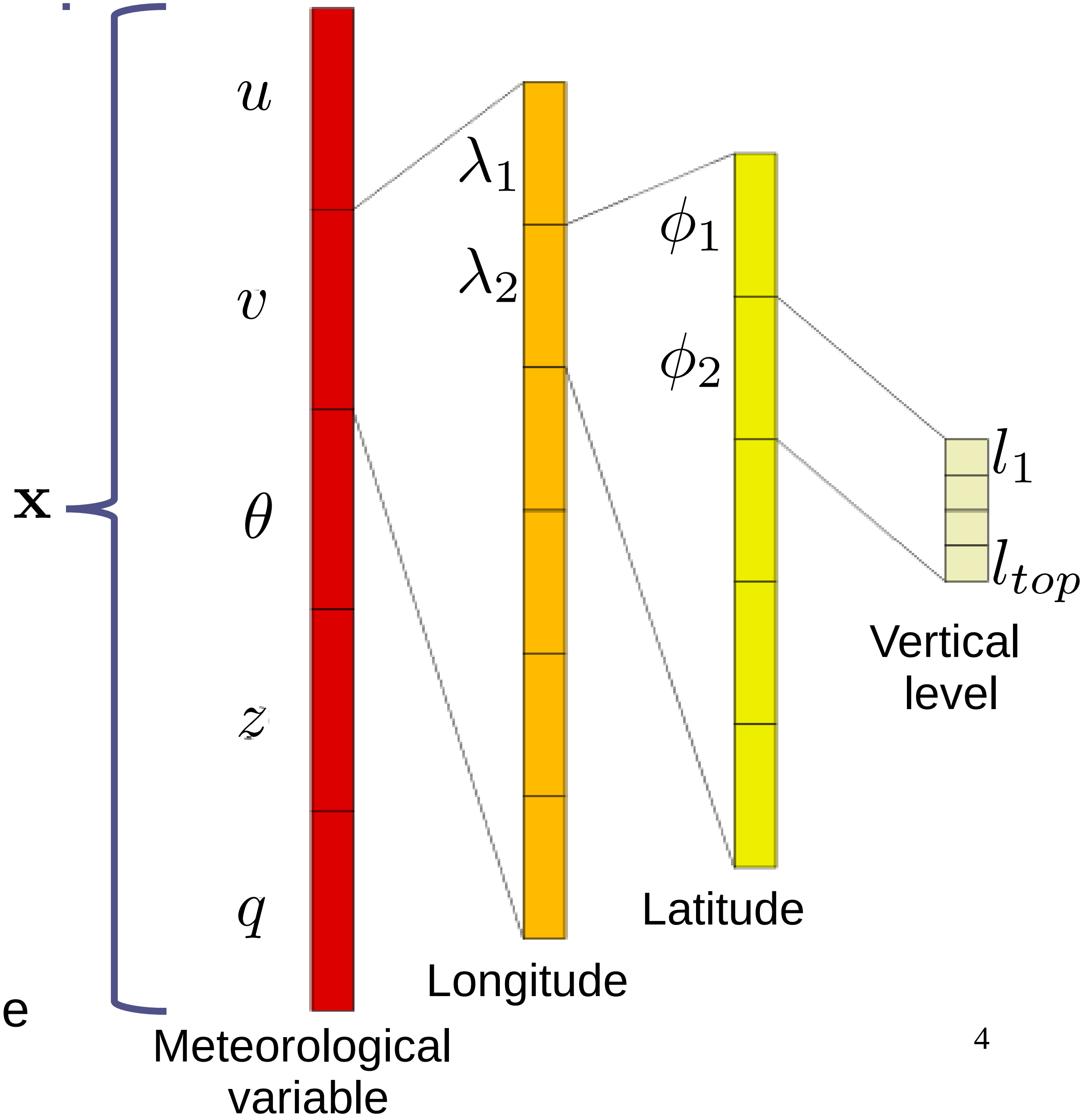
Our system of interest

State variables:

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



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



Two challenges

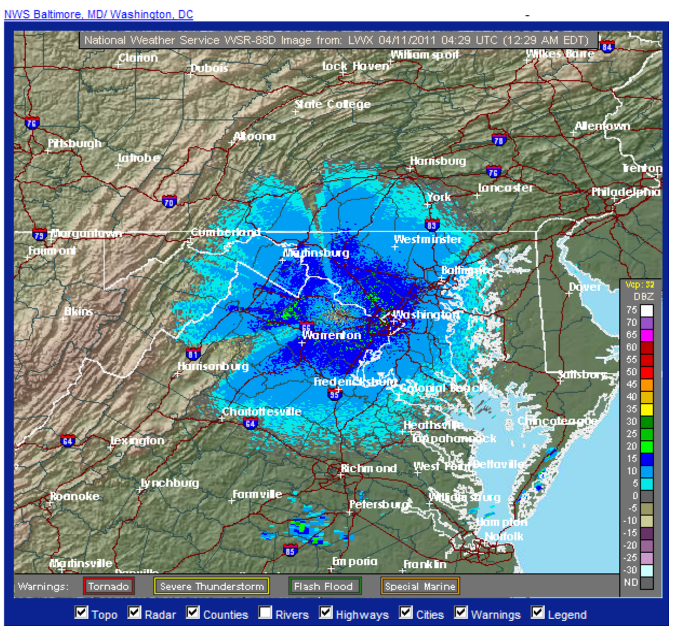
1. Determining the **current estate** of the system (all state variables) **at a given moment** of time. This is **estimation**.

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)

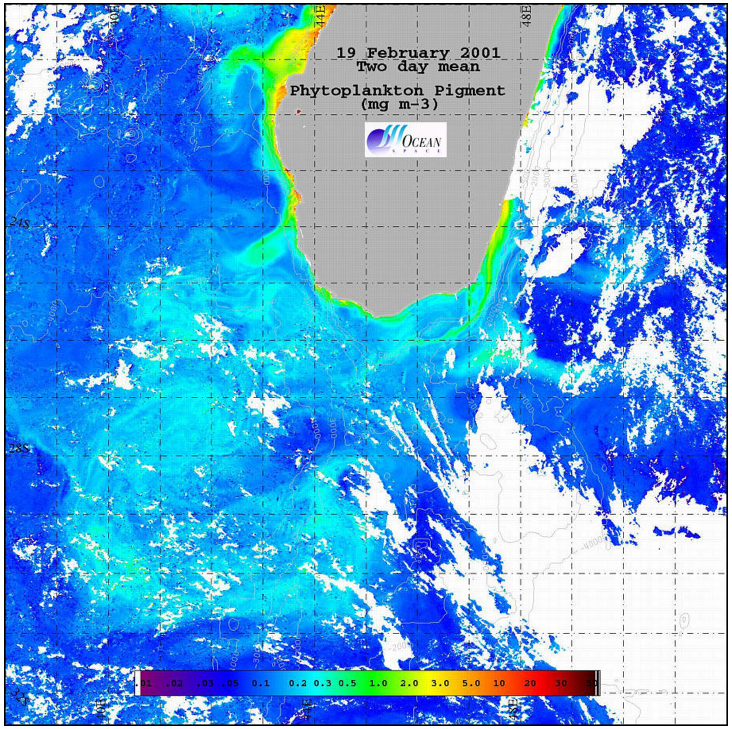
$$\frac{\mathbf{D}}{Dt}\mathbf{v} = -\frac{1}{\rho}\nabla p - f\hat{\mathbf{k}} \times \mathbf{v} + \mathbf{F}$$

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

Observations

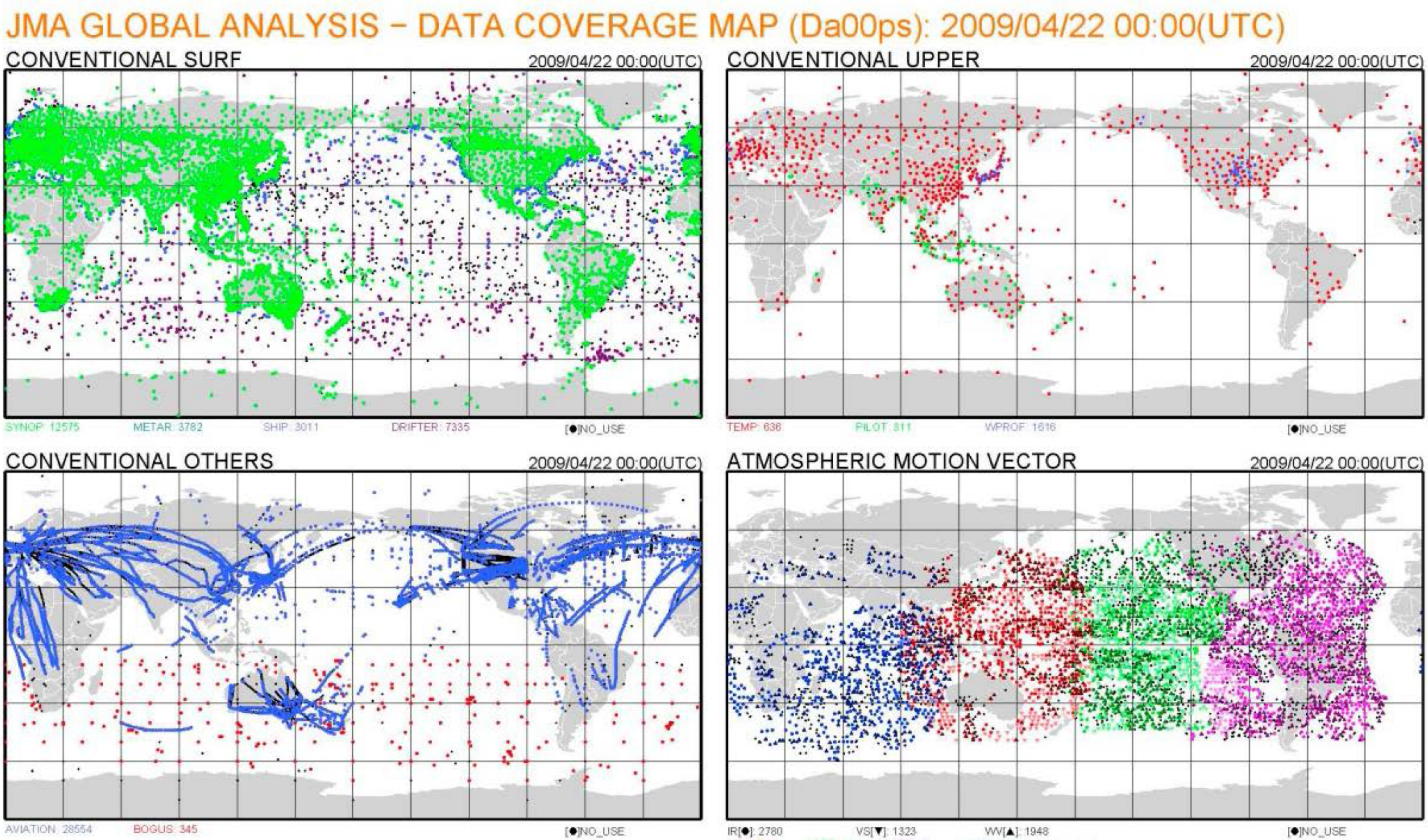
- **In situ** observations:
  They are **direct**, but they can be **irregular in space and time,** e.g. sparse hydrographic observations.

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



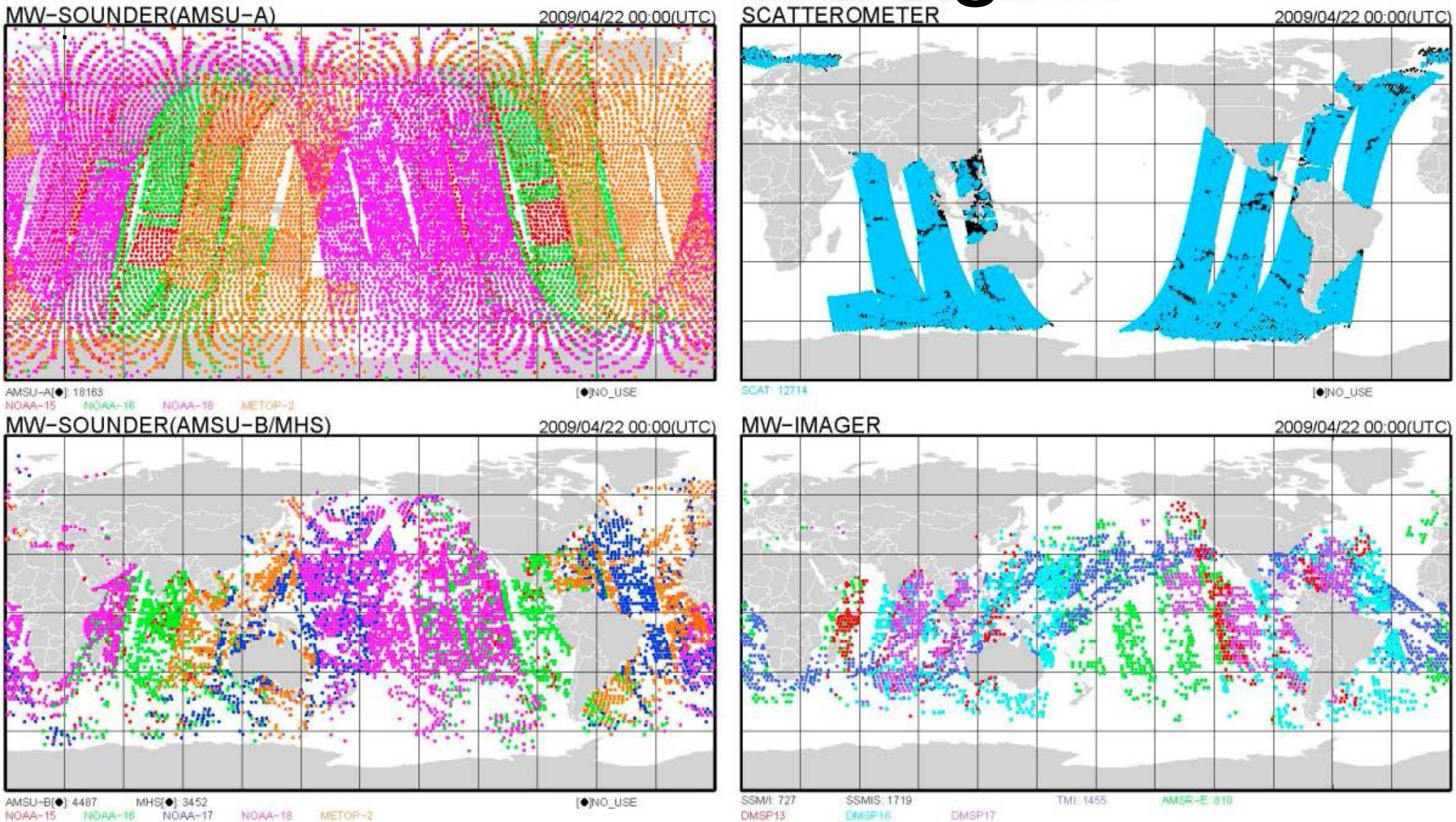
Observation coverage

.



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

.

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

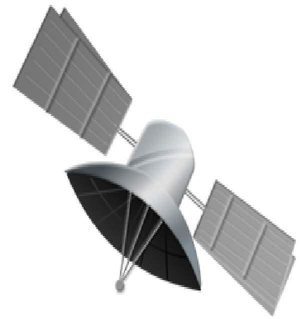
$\mathbf{y} \in \mathcal{R}^{N_y}$

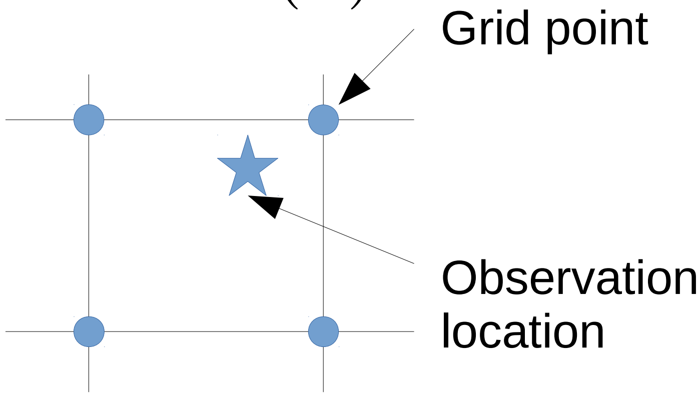
Usually: $N_y \ll N_x$

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

.

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

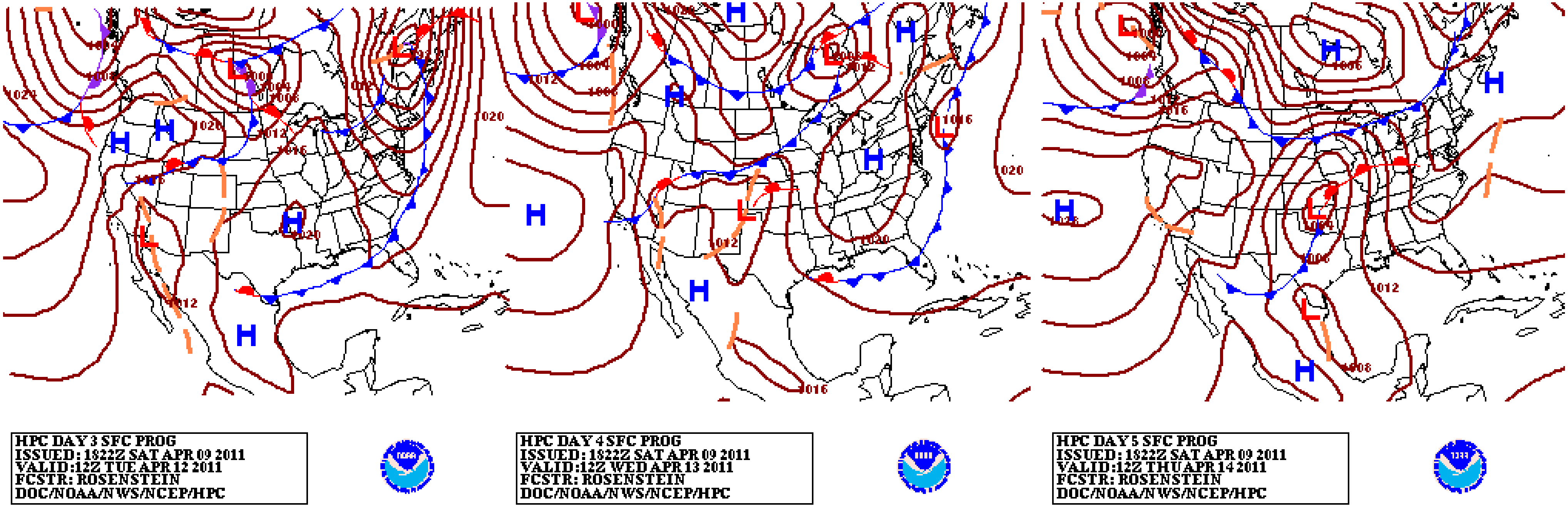
$$\mathbf{x}^t \in \mathcal{R}^{N_x}$$

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.
...

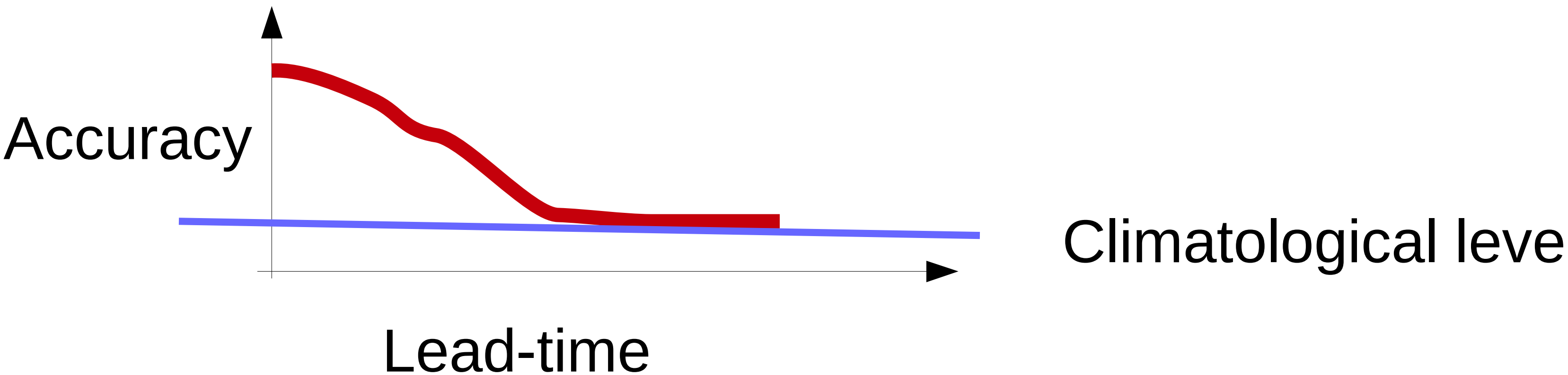
Evolution operator

Previous value of the variable.

Forecast with different lead-times

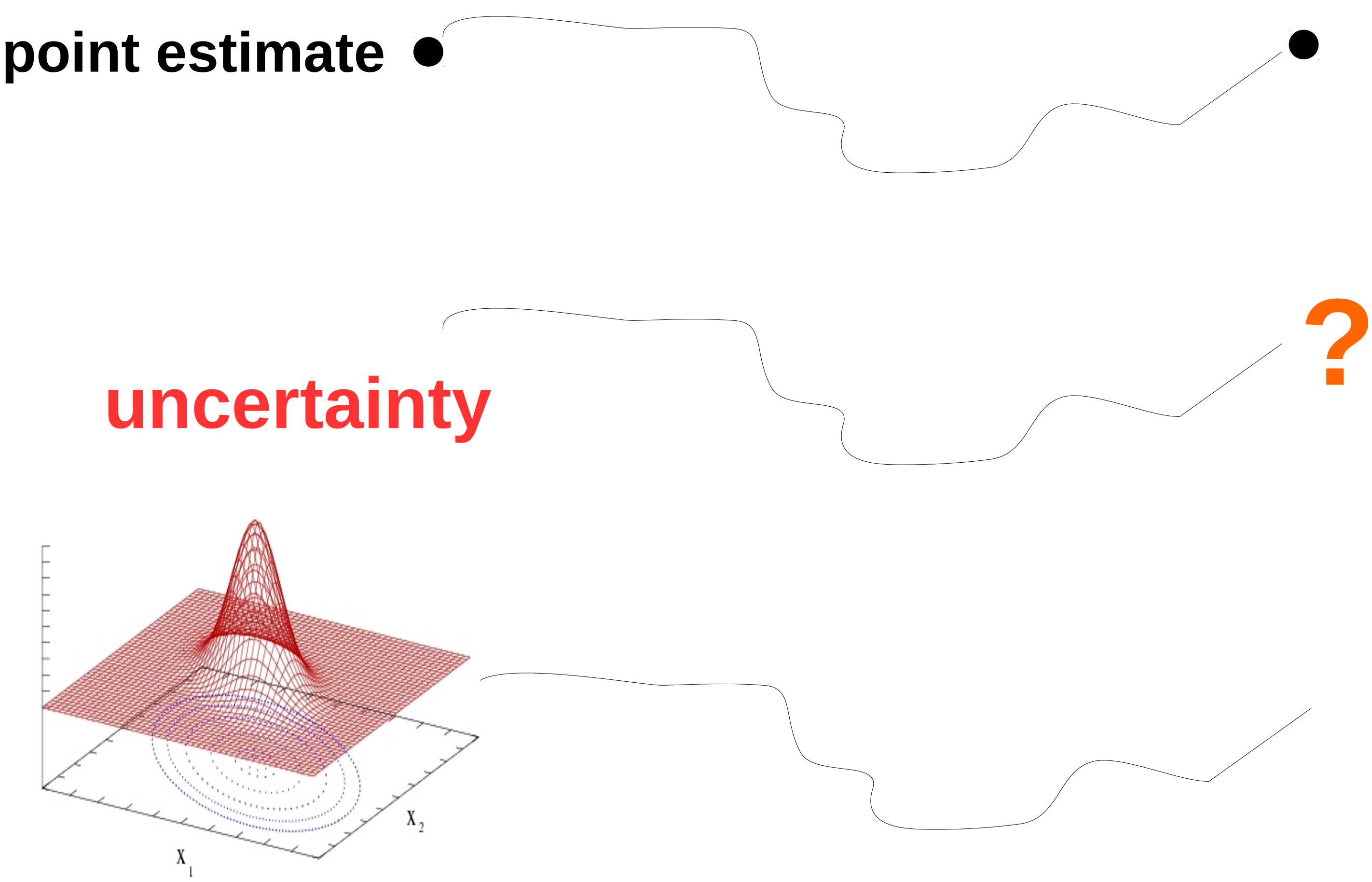


Should we consider the three of forecasts to have the same accuracy (different lead-times)?



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

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



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**.

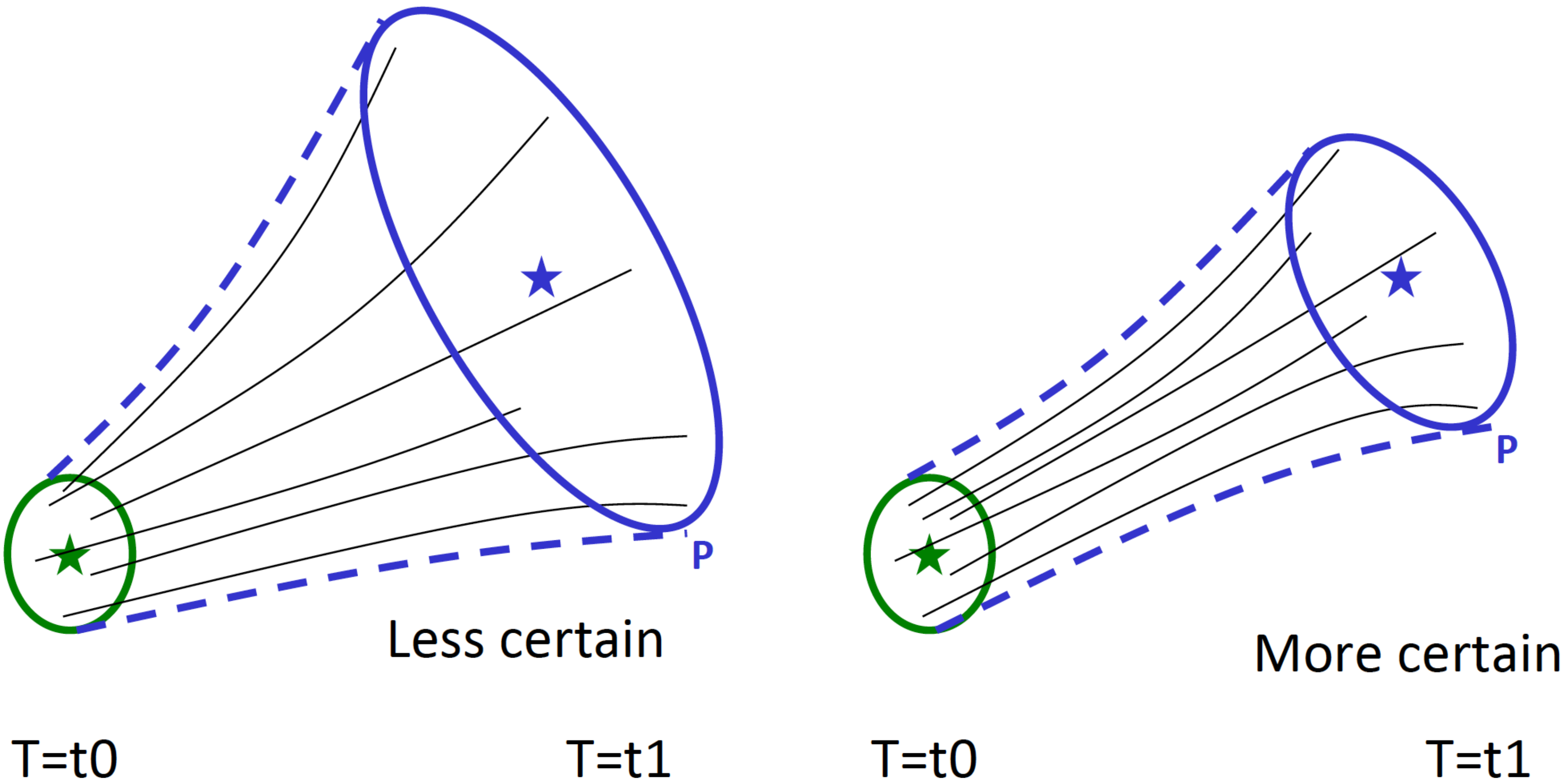
Can we determine with **absolute precision**? **No**.

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

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

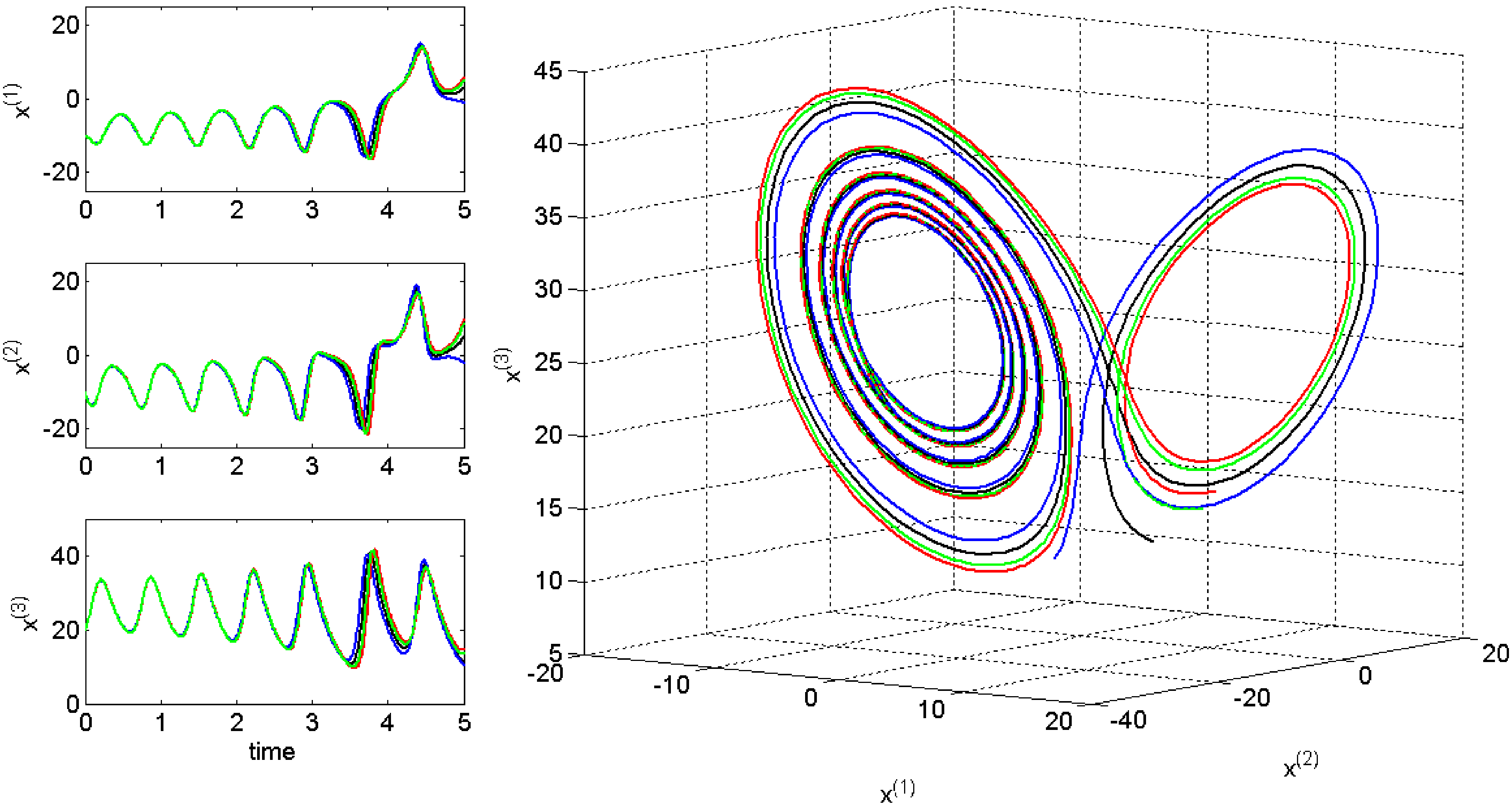
Sensitivity to initial conditions

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



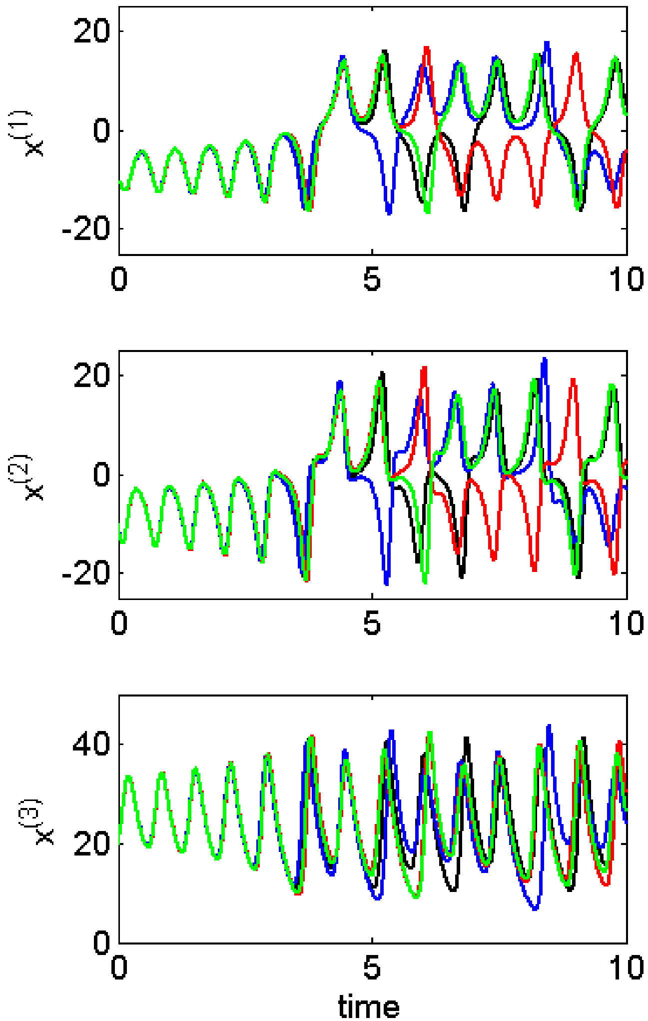
Example: Lorenz 1963 model

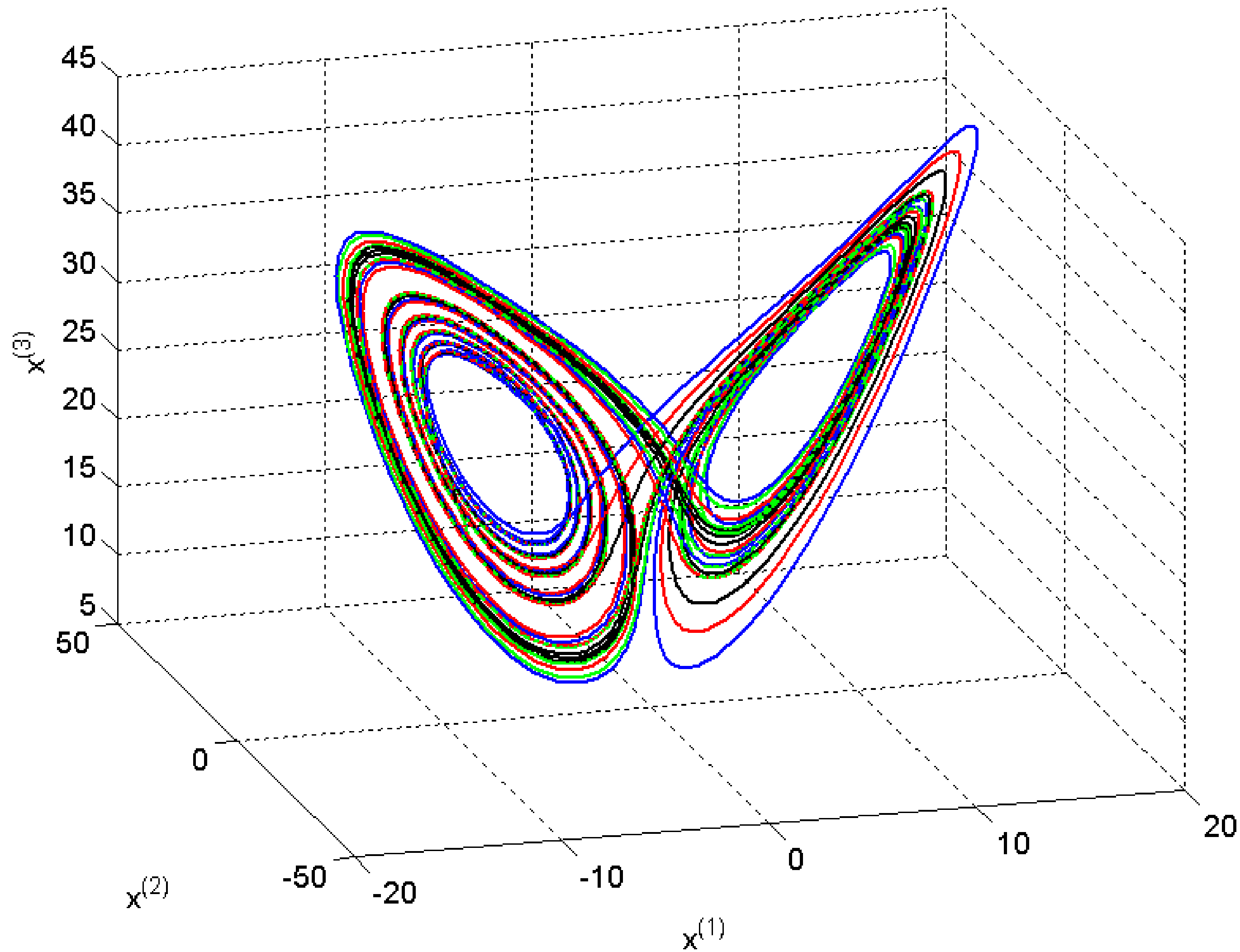
.



Example: Lorenz 1963 model

.

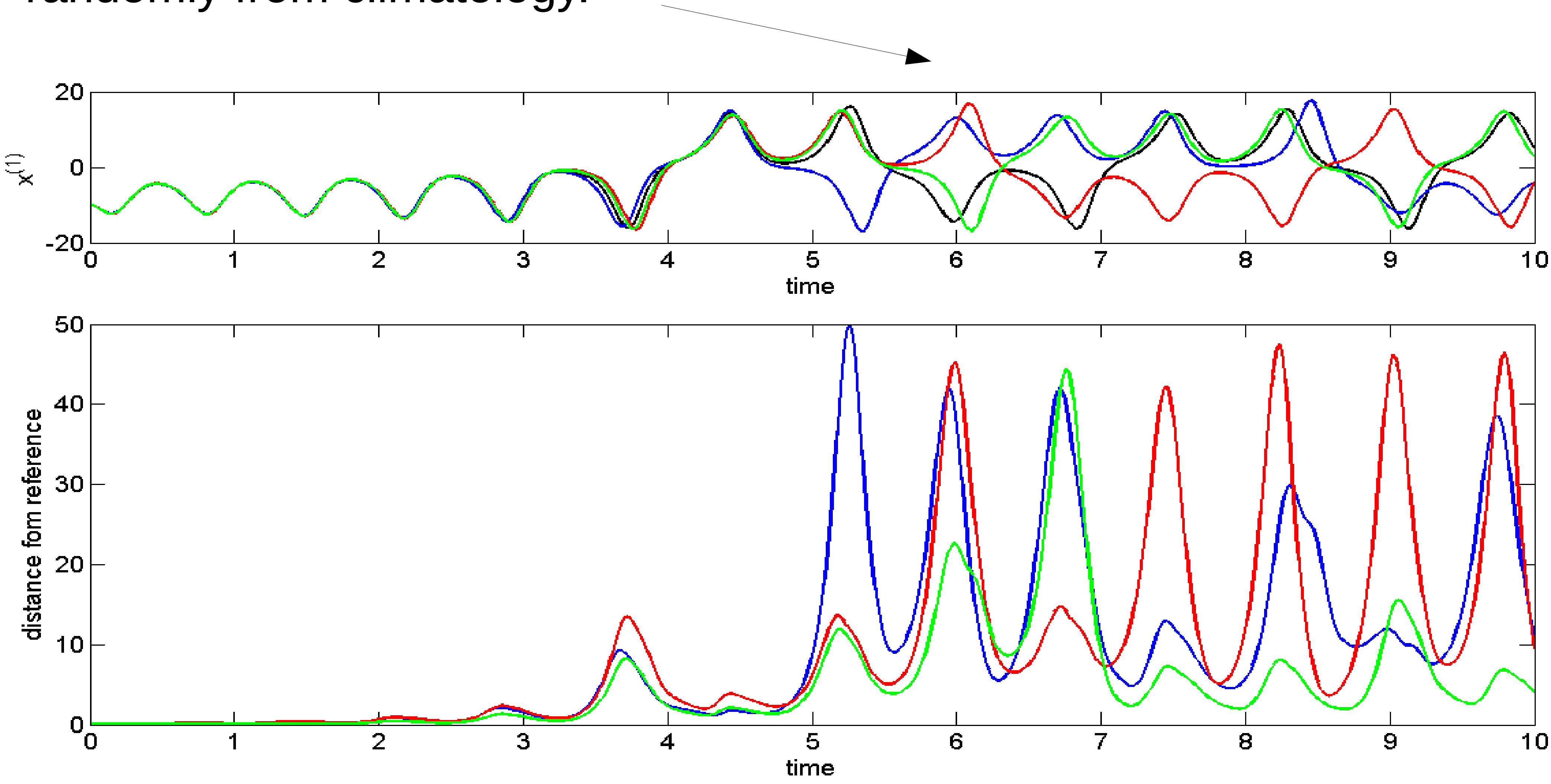




Example: Lorenz 1963 model

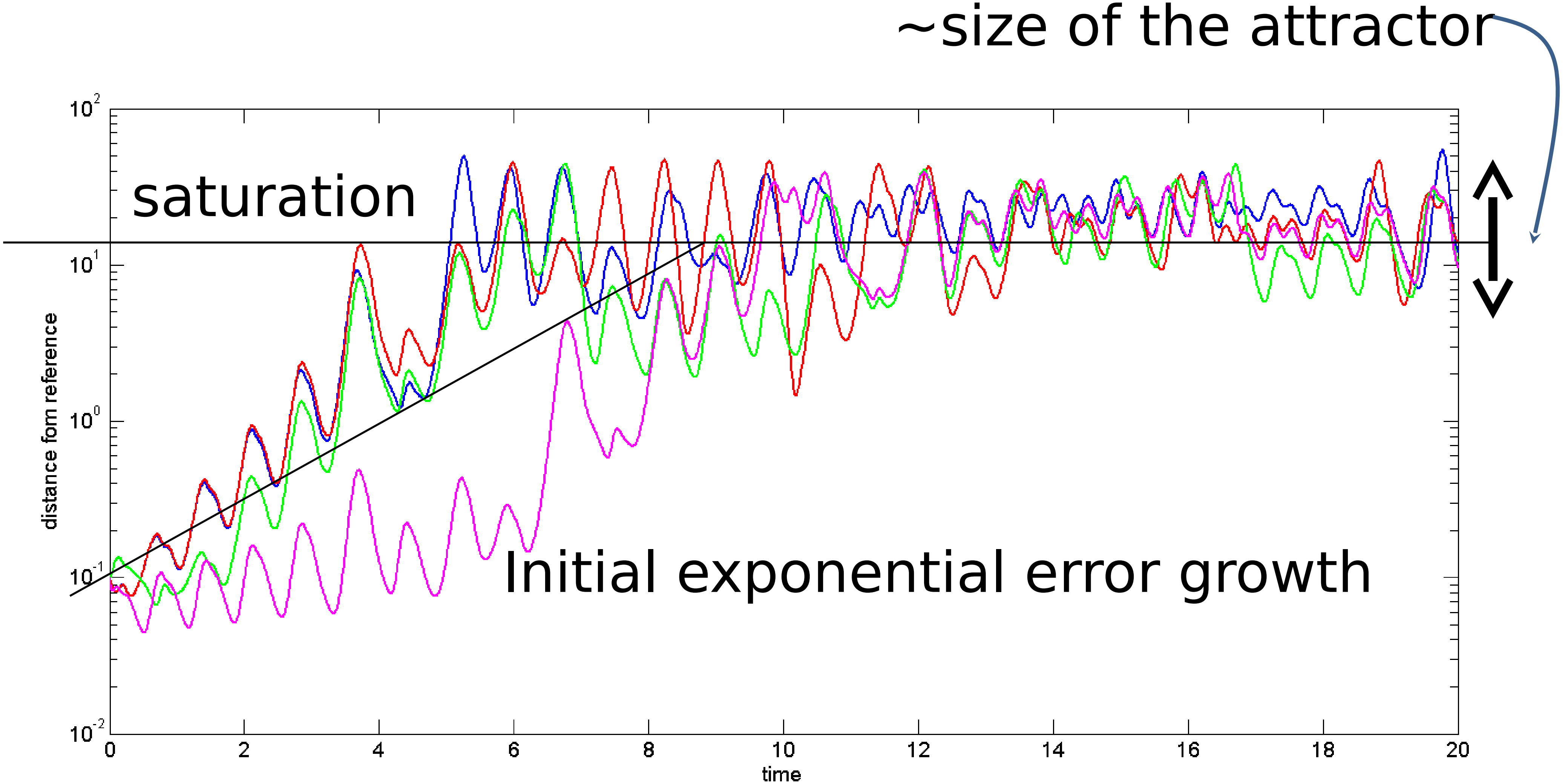
.

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

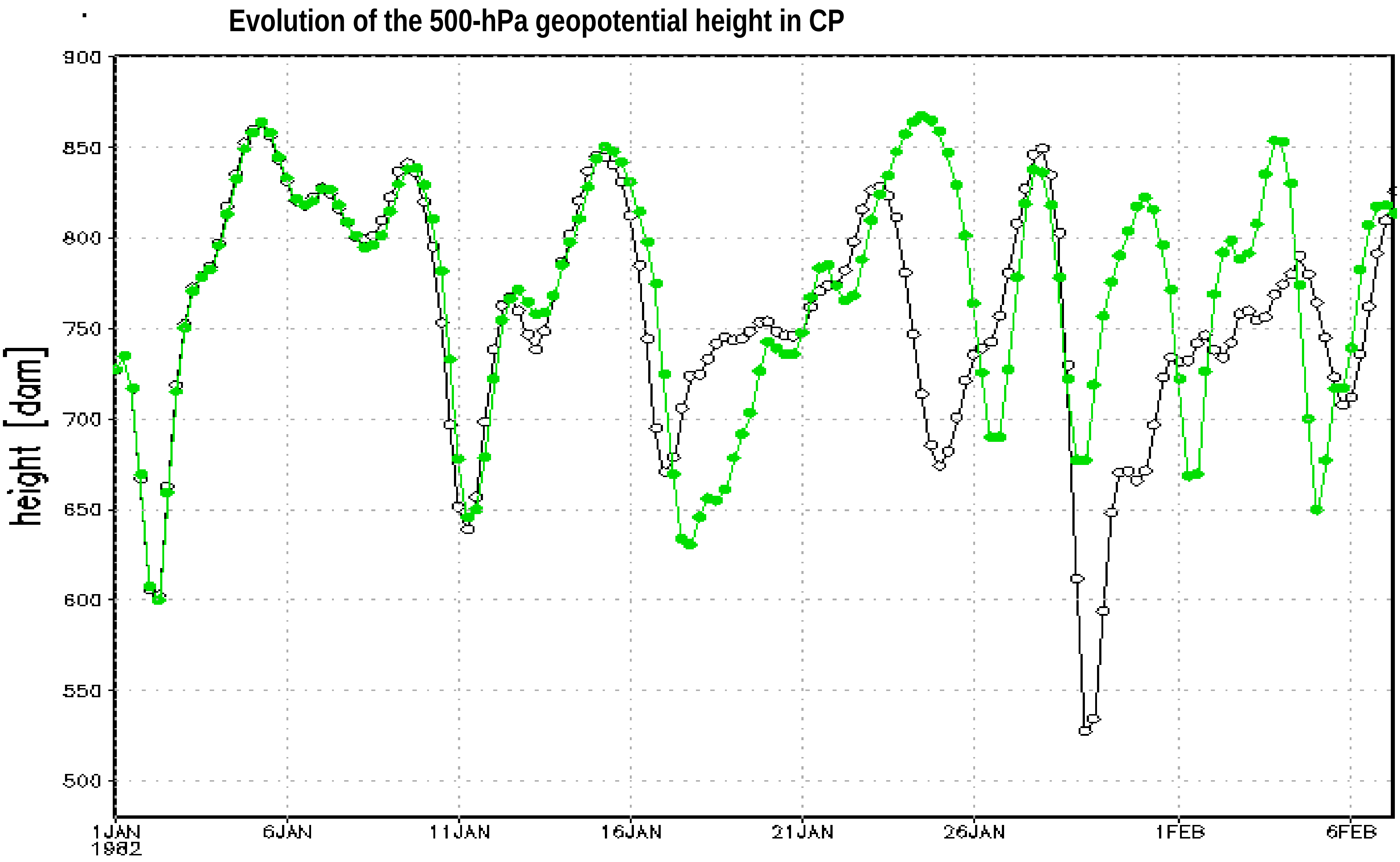


Example: Lorenz 1963 model

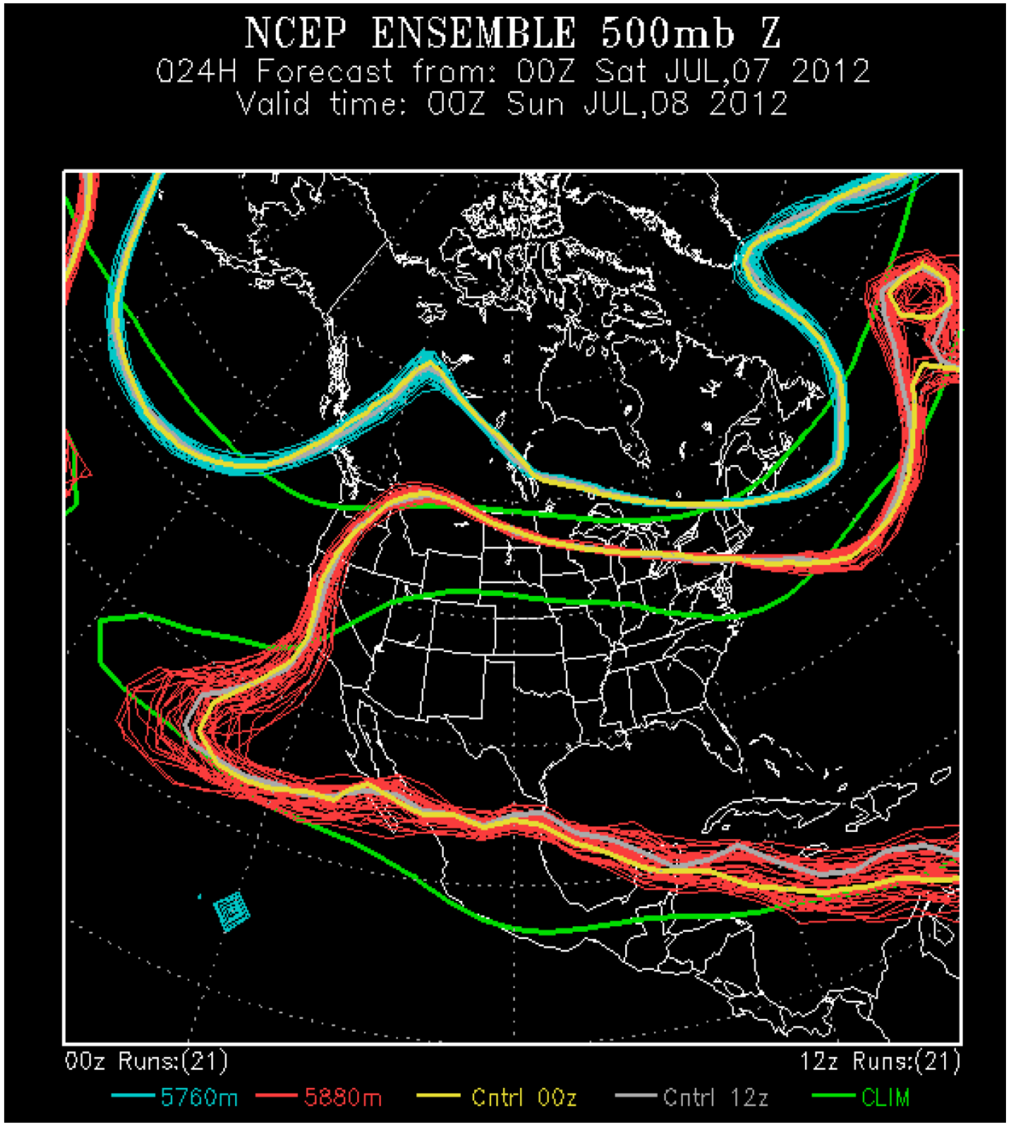
.

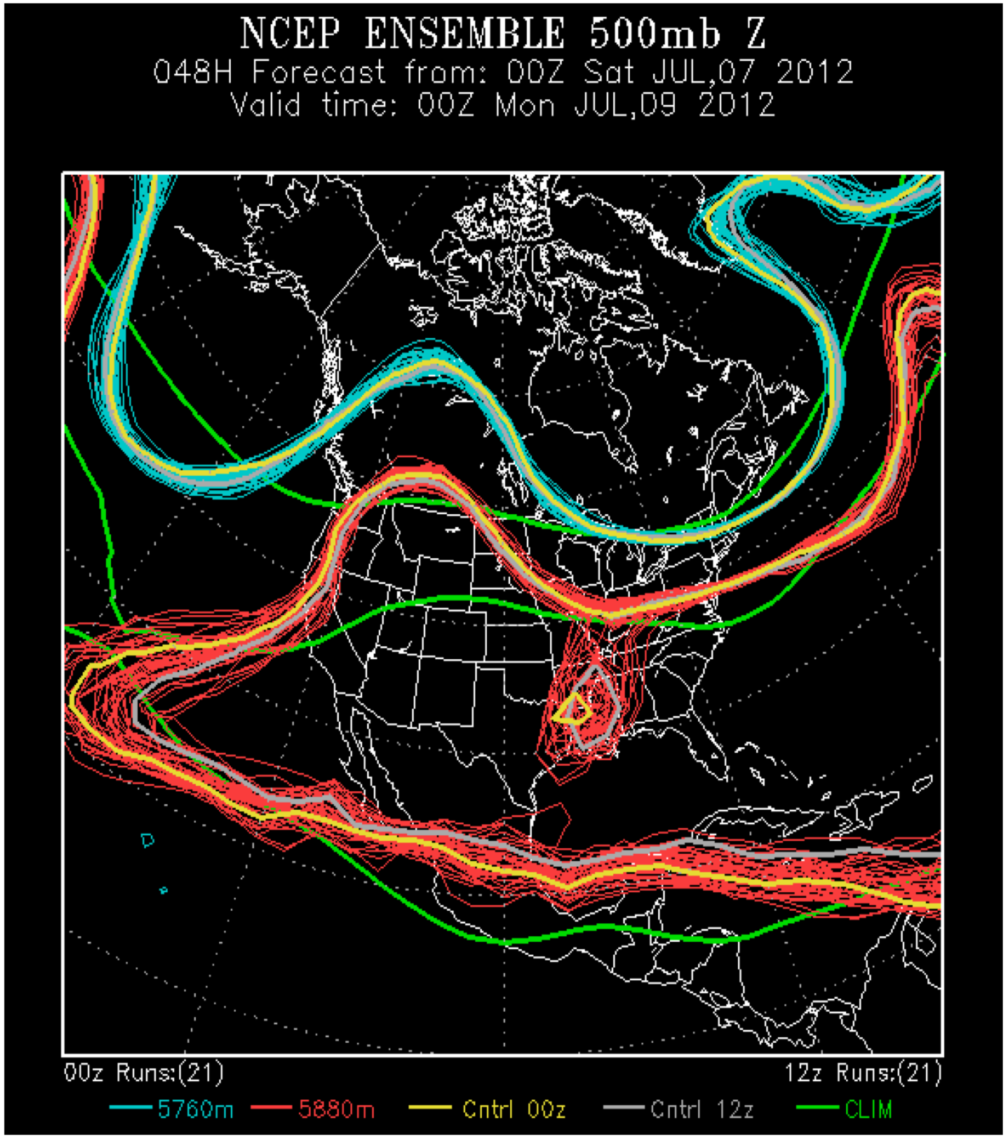


The atmosphere is chaotic

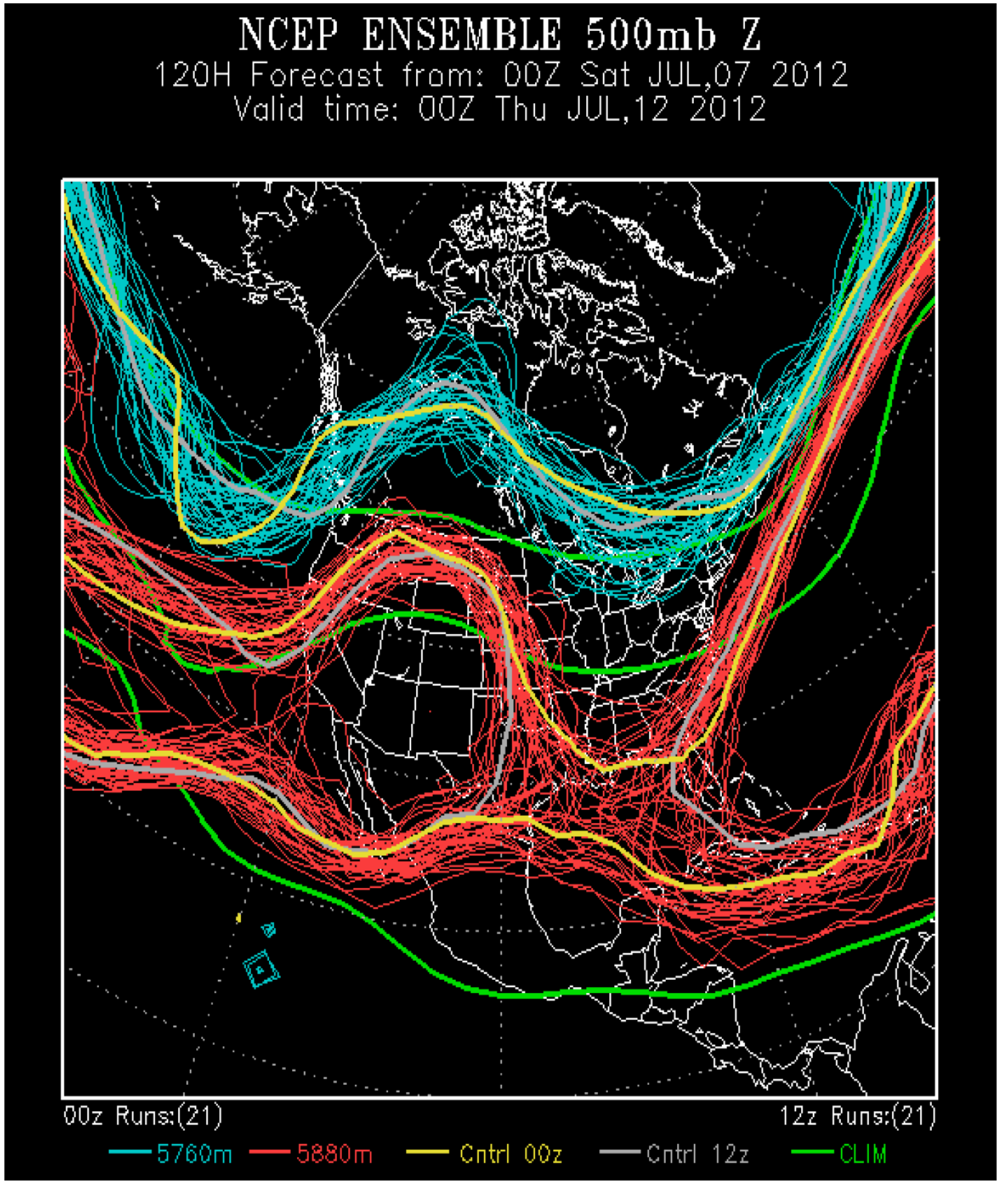


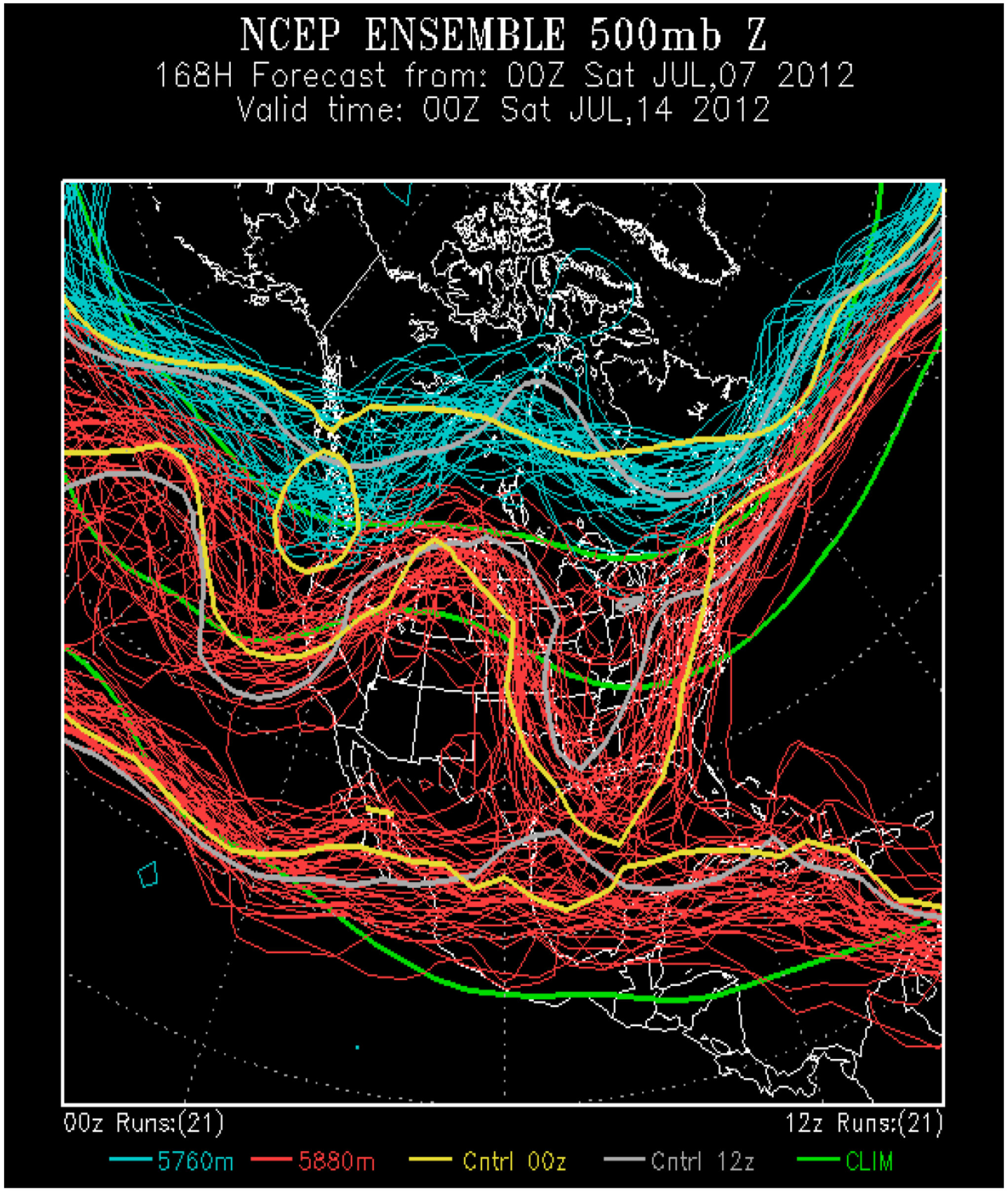
Weather is chaotic





Weather is chaotic

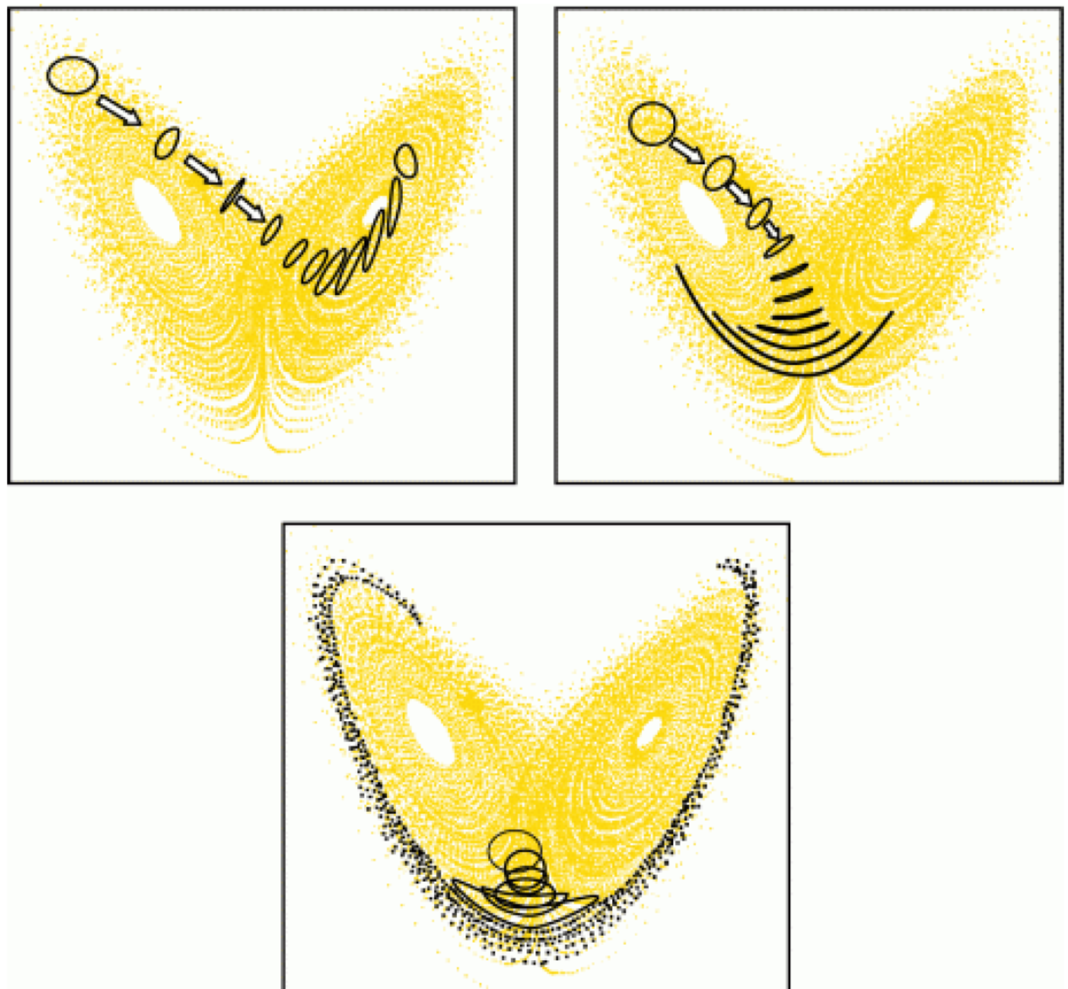




Sensitivity to initial conditions can depend on the situation

More predictable

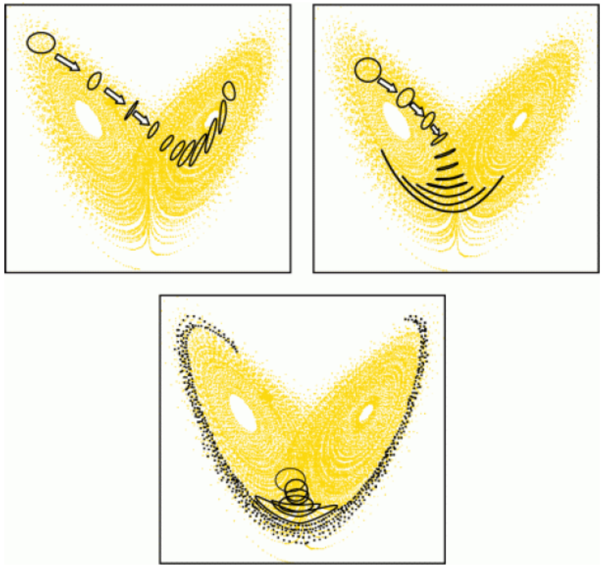
Less predictable



Very unpredictable

What can we do?

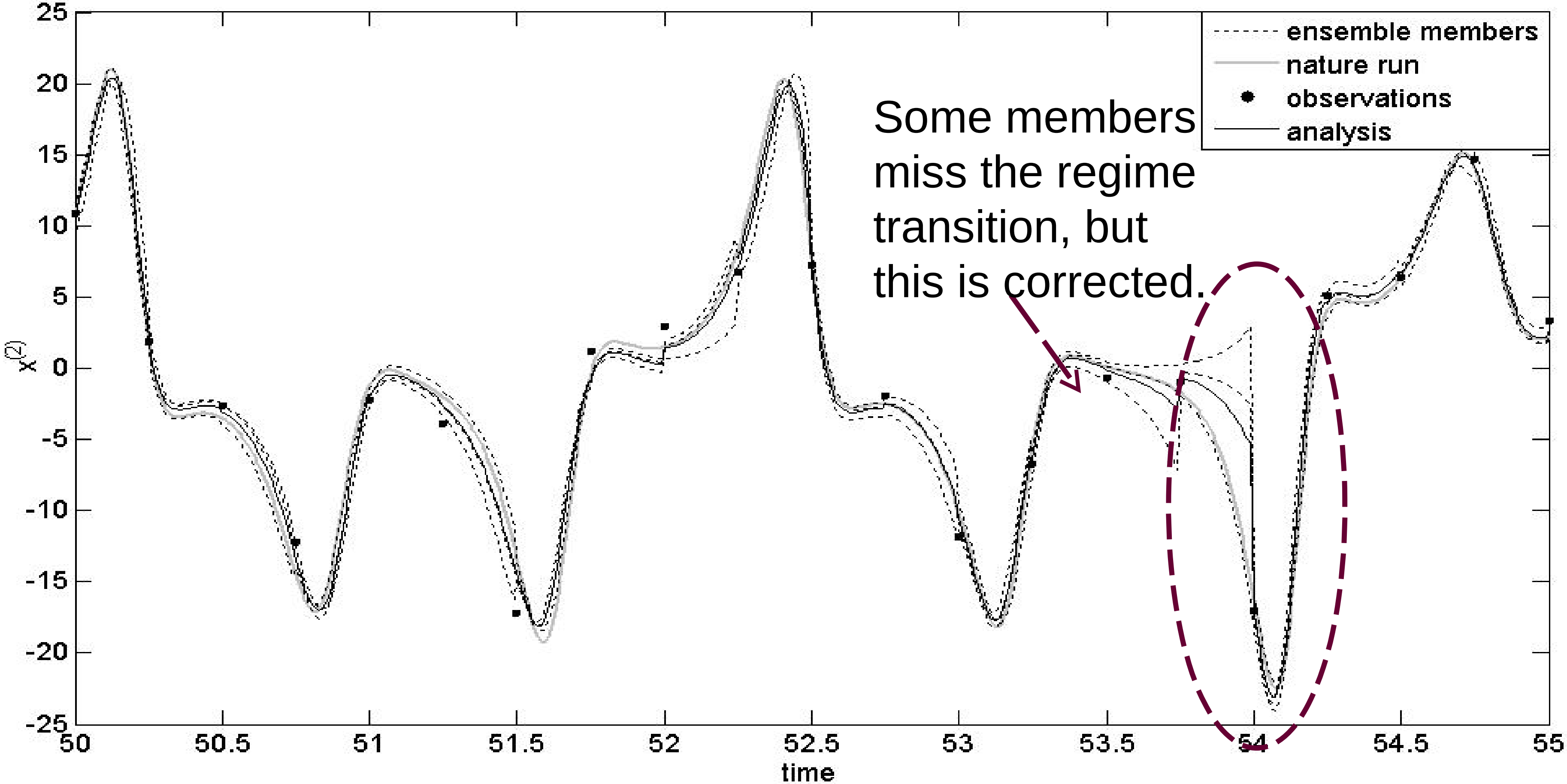
- Obtain more accurate initial conditions
  - More observations
  - Better data assimilation methods



- 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 !!!

Basics on **probability** and **statistics**

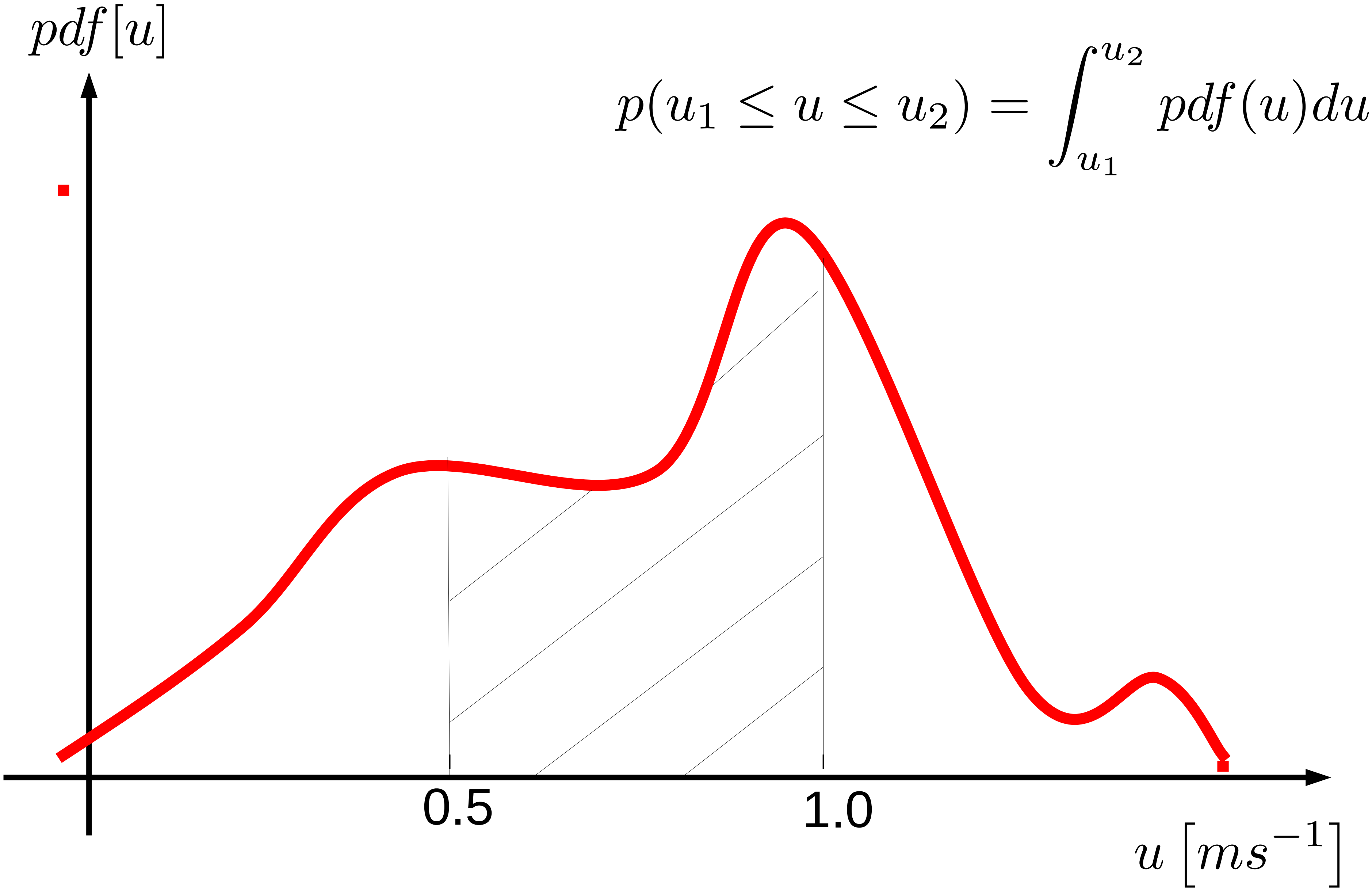
- Deterministic experiment: We know the result before it happens.

- Random experiment: We do not know the result, but we know the set to which it belongs (sample space), and we know something about the chances of different outcomes.

- Random variable: mapping from the sample space into the real numbers. Described by a probability mass function (discrete case) or a probability density function (continuous case).

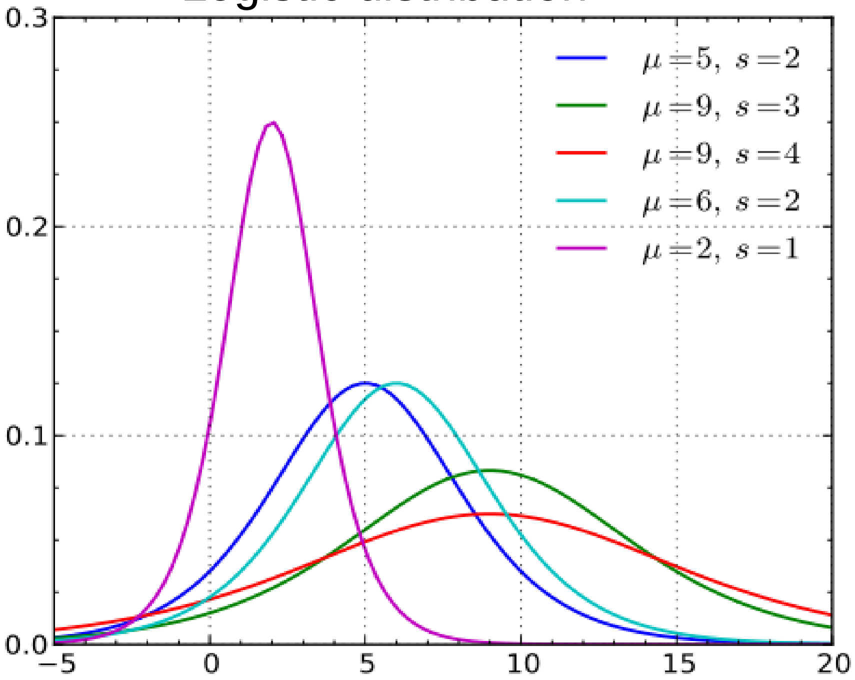
- Stochastic process: a repetition of random experiments through time.

Probability density functions. Univariate case

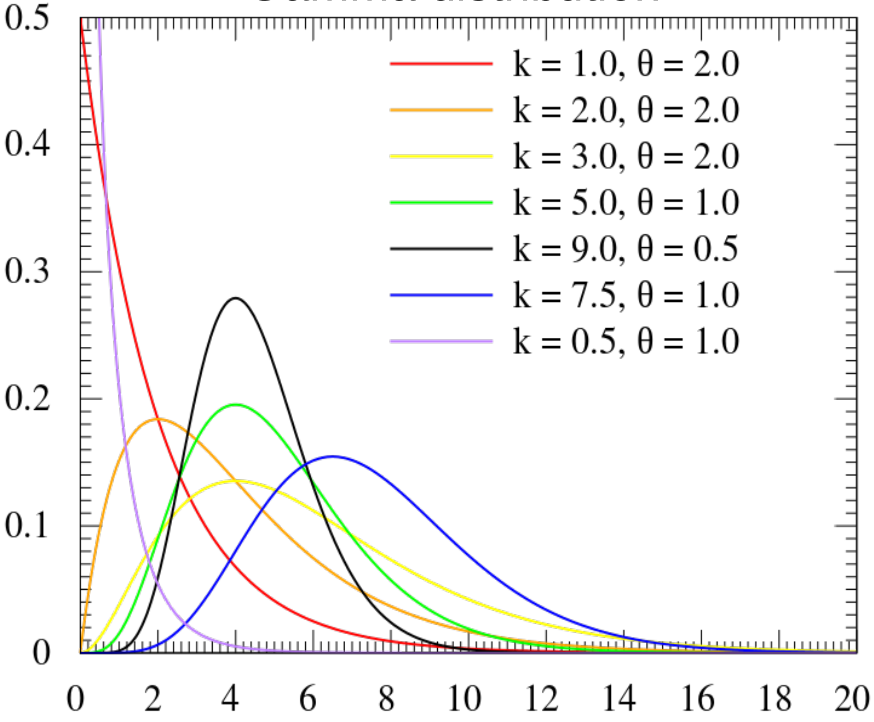


Parametric distributions (1D)

Logistic distribution



Gamma distribution



$$f(x;\mu,s) = \frac{1}{4s} sech^2\left(\frac{x-\mu}{2s}\right)$$

$$f(x;k,\theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} exp\left(-\frac{x}{\theta}\right)$$

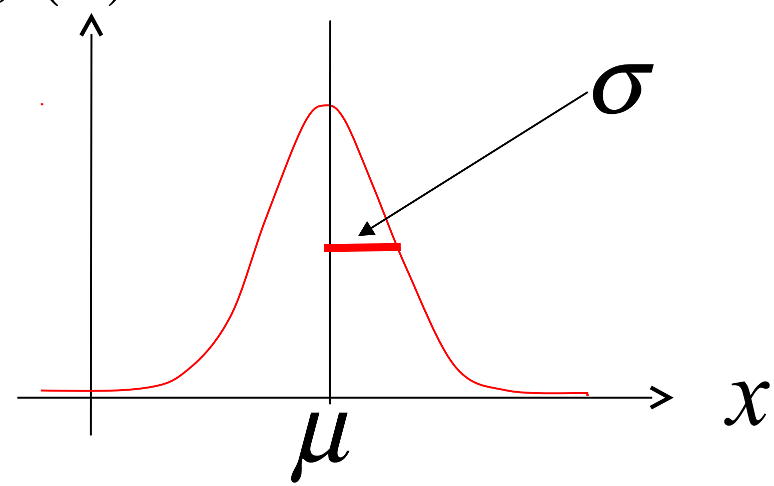
Statistics:
- **Central tendency**: mean, median, mode
- **Dispersion / variability**: variance, range
- **Shape**: skeweness, kurtosis

The Gaussian distribution

Errors are often considered to be Gaussian.

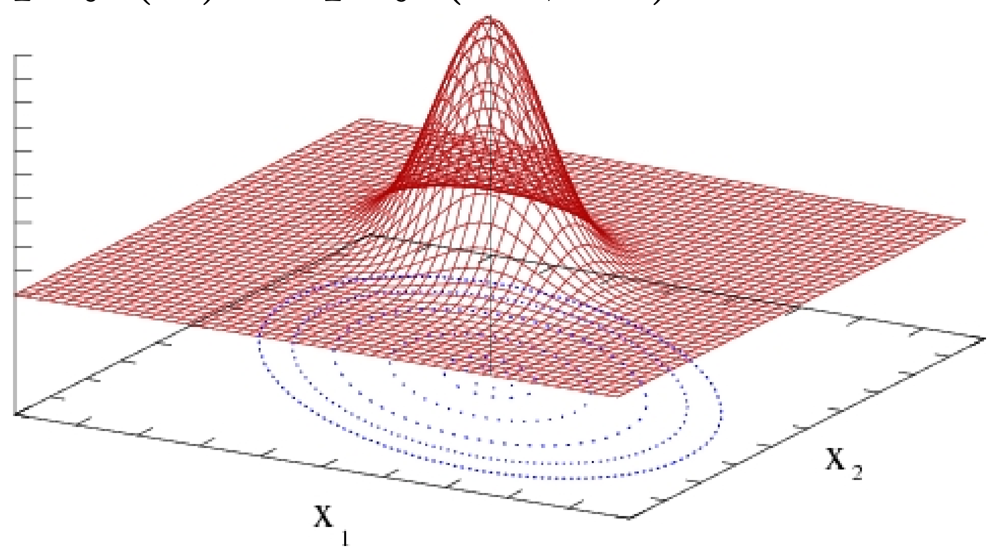
$$pdf(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$pdf(x)$$



$$pdf(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{S}|}} \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \mathbf{S}^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

$$pdf(\mathbf{x}) = pdf(x^1, x^2)$$



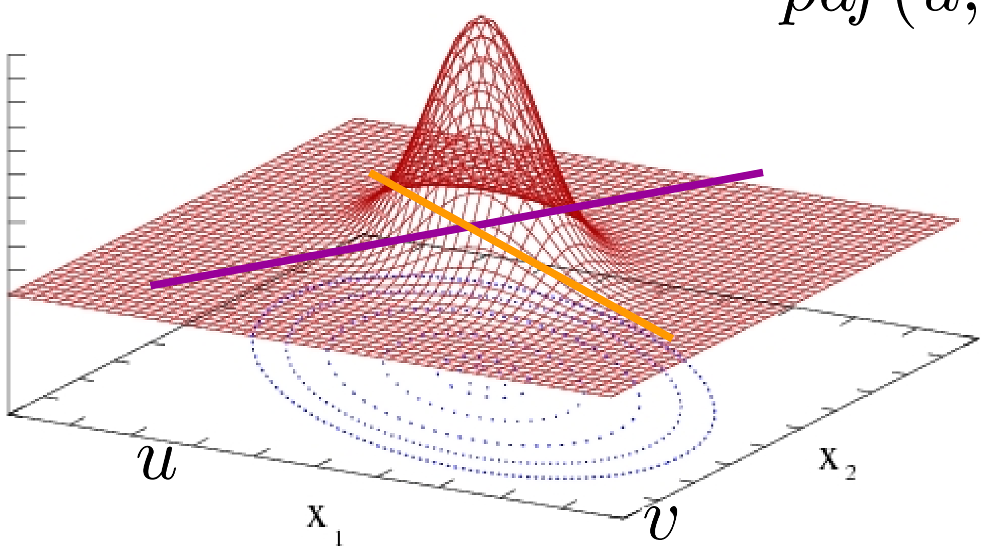
μ mean

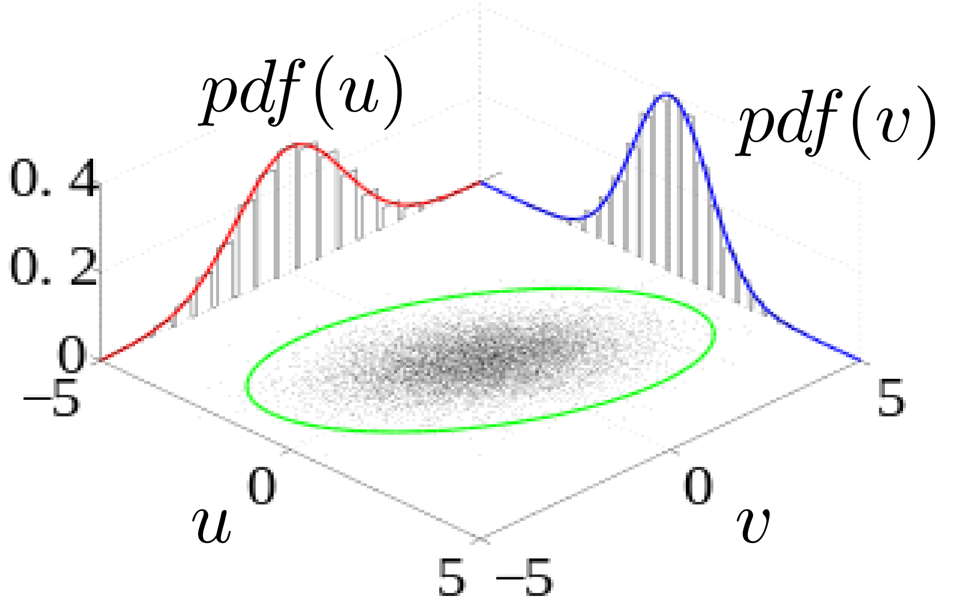
σ^2 variance

$\mathbf{S}$ covariance matrix

More on PDFs

Consider two variables: $\{u, v\}$





$pdf(u),\ pdf(v)$ Marginal pdf's

$pdf(u, v)$ Joint pdf

$pdf(v|u = u^*)$

Conditional pdf's

$pdf(u|v = v^*)$

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.

A (really) simple example on conditional probabilities

DA conference:
- 20 attendees, 12 female and 8 male.
- 4 females wear glasses, 6 males wear glasses.
- If a person is picked at random and this person wears glasses, what is the probability of the person being a male?

Variables		**Permissible values**
x	Gender	$\Omega_x = \{\text{male},\ \text{female}\}$
y	Wearing glasses	$\Omega_y = \{\text{yes},\ \text{no}\}$

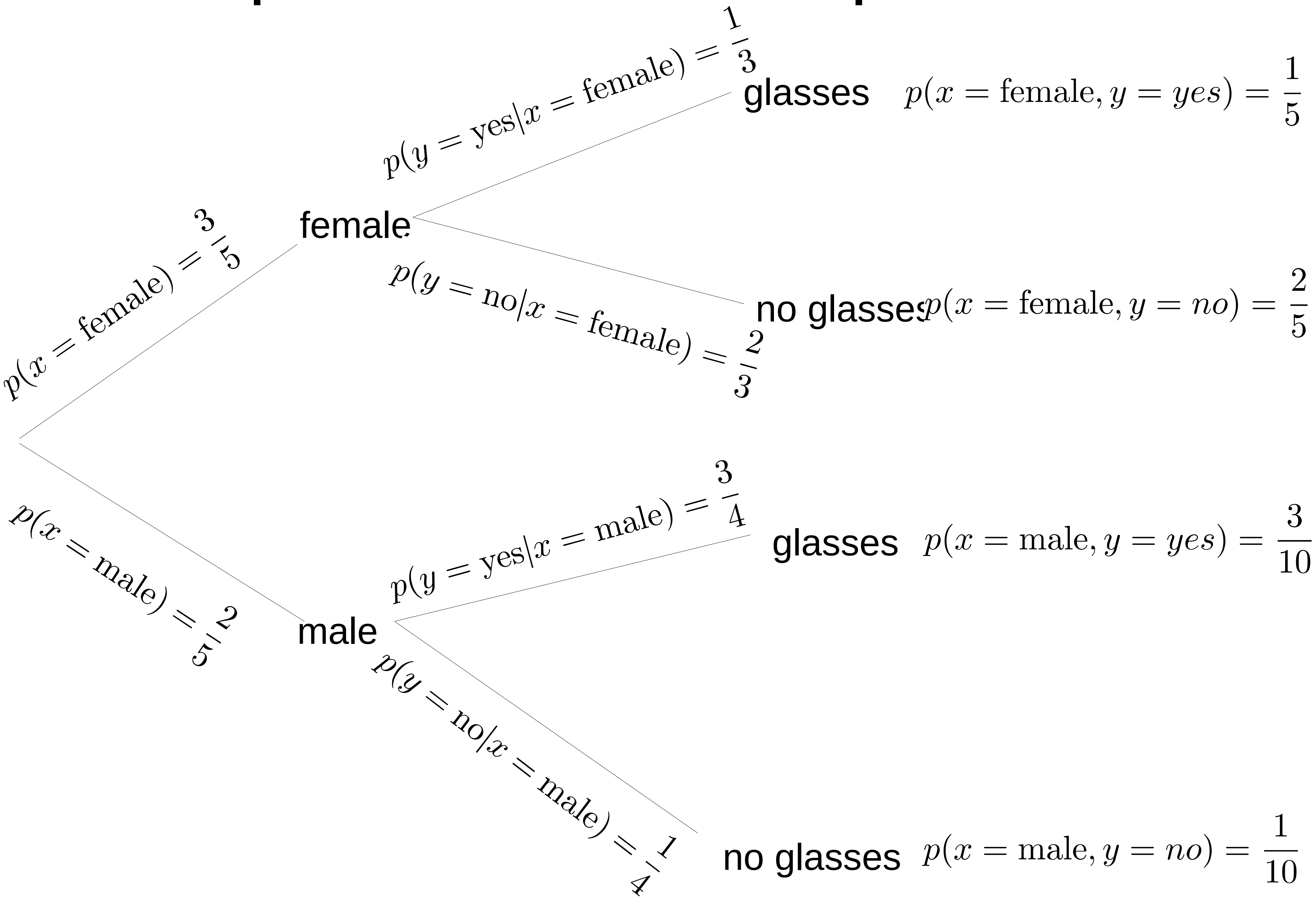
$$pmf(x = \text{Female}) = \frac{12}{20} = \frac{3}{5}$$

$$pmf(x = \text{Male}) = \frac{8}{20} = \frac{2}{5}$$

$$pmf(y = \text{Yes}) = \frac{10}{20} = \frac{1}{2}$$

$$pmf(y = \text{No}) = \frac{10}{20} = \frac{1}{2}$$

Example on conditional probabilities



Example on conditional probabilities

$$p(x = \text{male}|y = \text{yes}) = \frac{p(y = \text{yes}|x = \text{male})p(x = \text{male})}{p(y = \text{yes})}$$

$$= \frac{\frac{3}{4}\frac{2}{5}}{\frac{1}{2}} = \frac{3}{5}$$

Original probability $p(x = \text{male}) = \frac{2}{5}$

Observation: the person wears glasses!

New probability $p(x = \text{male}|y = \text{yes}) = \frac{3}{5}$

I have updated my knowledge!

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

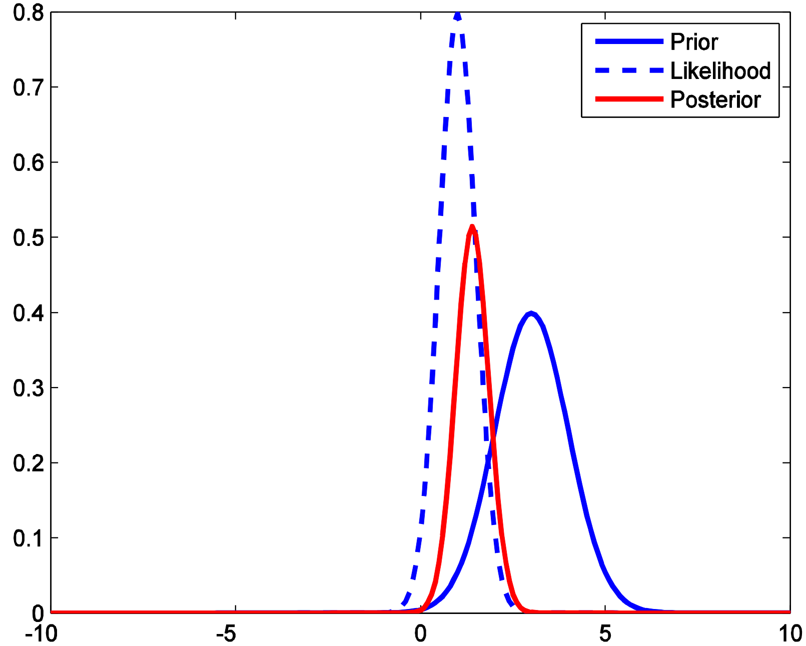
$$pdf(\mathbf{x}|\mathbf{y}) = \frac{pdf(\mathbf{y}|\mathbf{x})pdf(\mathbf{x})}{p(\mathbf{y})}$$

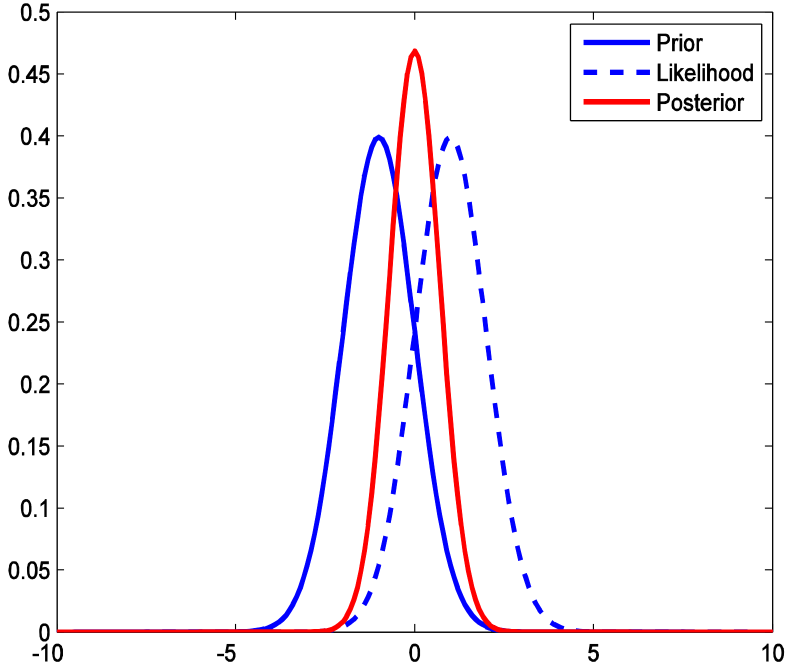
Posterior pdf. Pdf of the state variables given the observations.

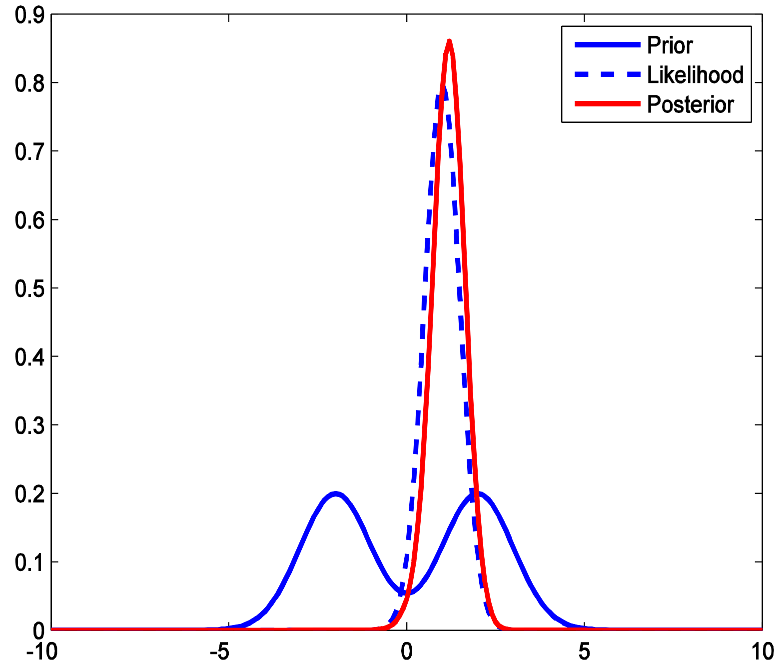
Marginal pdf of the observations. It is often the case we do not need to compute this, since it acts as a normalisation constant.

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

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

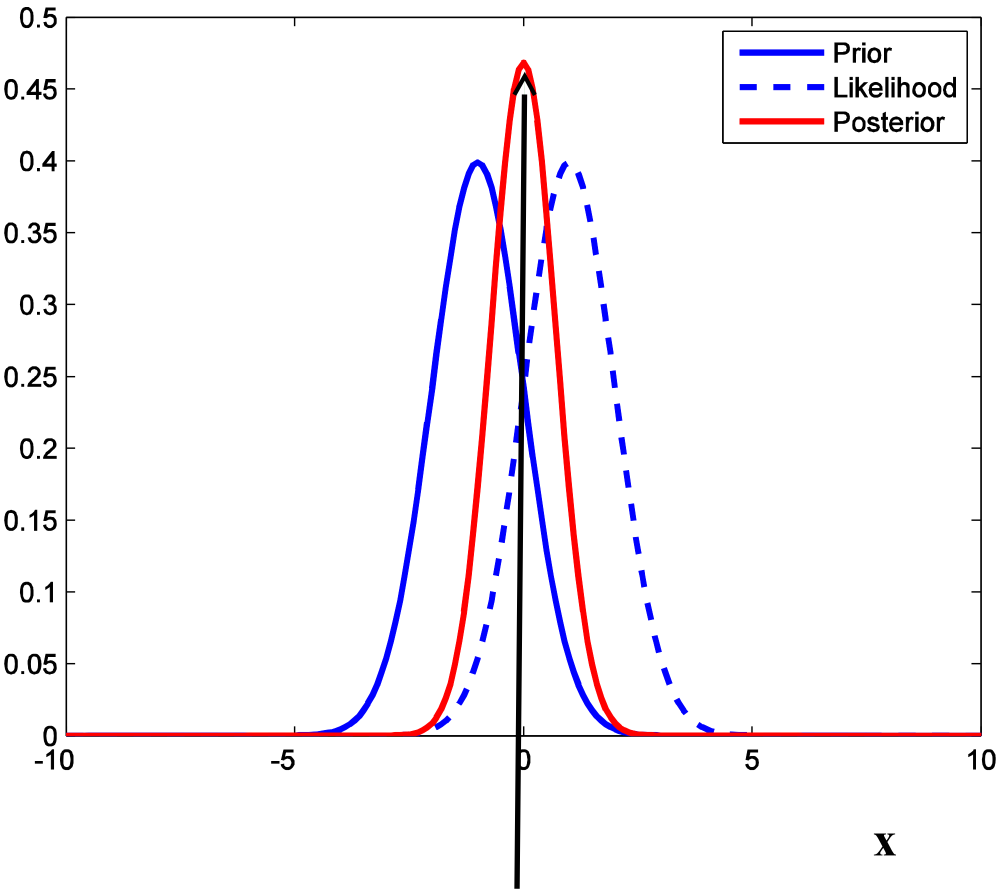
Posterior

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

Maximum a-posteriori estimator (MAP)

For a Gaussian distribution the mean and mode coincide.

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



$\mathbf{x}$

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)

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

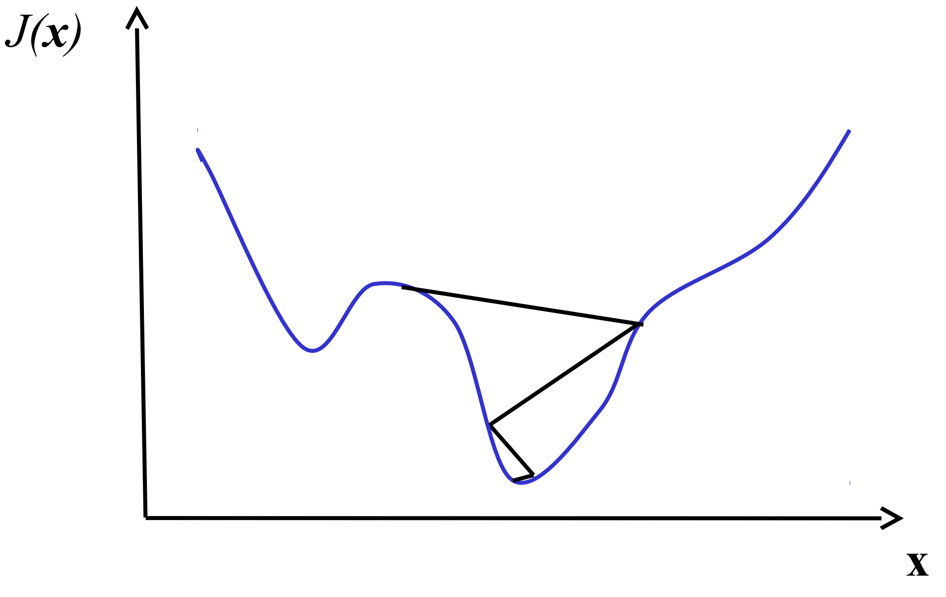
Which for linear **H** is:

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

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

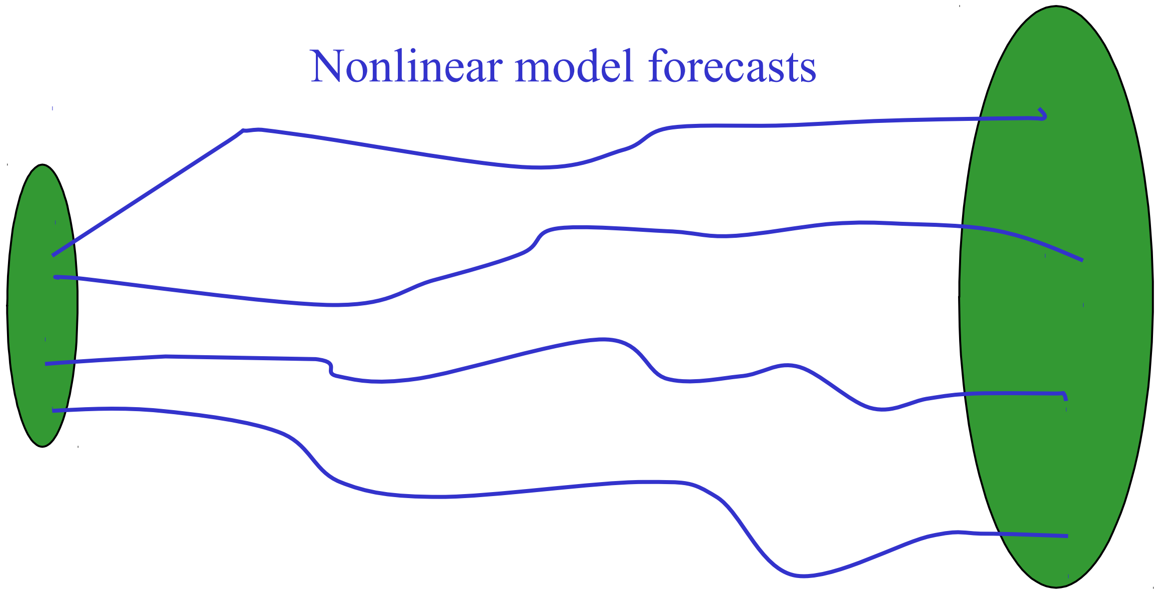
Solve directly.

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

- 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.



3. Hybrid methods

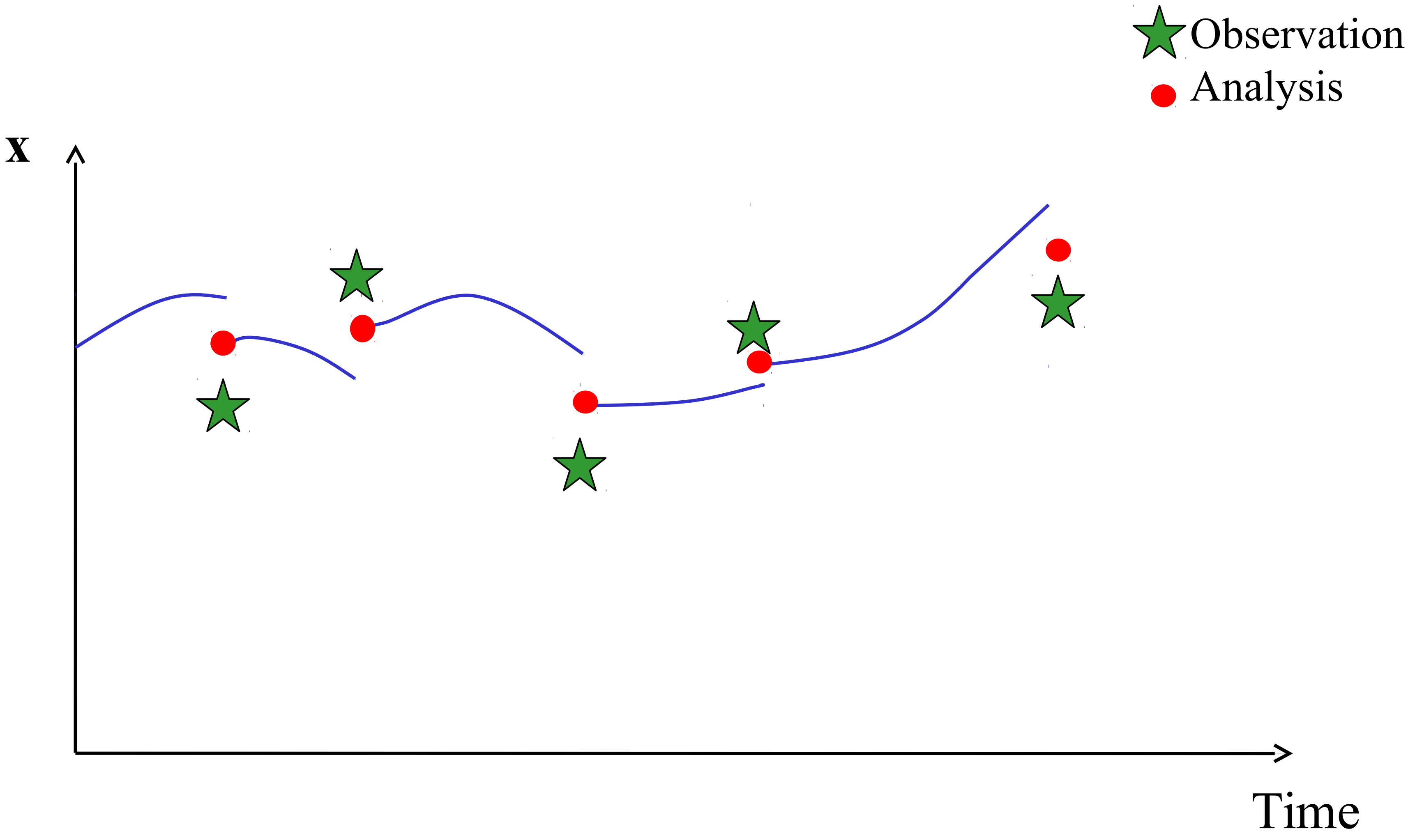
- Different flavours.
- For example, use sample covariances within the variational framework.
- Use 4D (space-time) covariances.

4. Particle filters

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

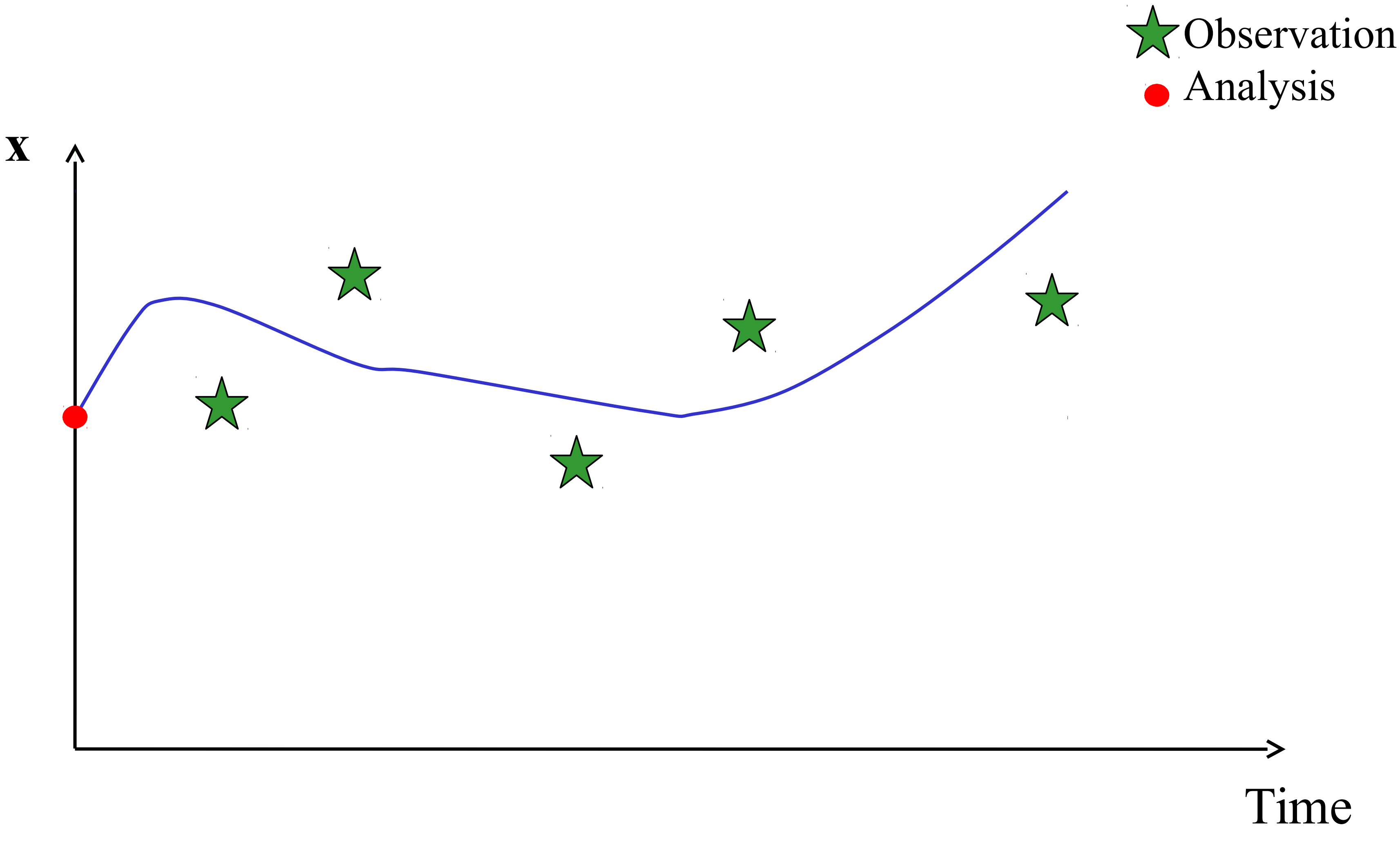
Filters

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.