INTRODUCTION TO DATA ASSIMILATION: STATISTICAL METHODS

Emmanuel Cosme
Université Grenoble Alpes
Institut des Géosciences de l'Environnement

Outline

- Intro: inverse problems and data assimilation (DA)
- DA concepts (1): Nudging, 4DVar, Sequential methods
- DA concepts (2): Ensemble DA
- Ensemble DA: Analysis schemes
- Observing System Experiments (OSE)
- Observing System Simulation Experiments (OSSE)

What is data assimilation?

- Data assimilation is a specific class of inverse problems…

What is an inverse problem?

- According to Tarantola (2005), the inverse problem consists of using the actual result of some measurements to infer the values of the variables that characterize the observed system.

$$y = h(x) \longrightarrow x = h^{-1}(y)$$

What is an inverse problem?

- But operator h is rarely invertible!

- In practice, we use a first guess x^b (a prior knowledge), and implement smart ways to correct this first guess so that it fits the observations better.

$$x^b \longrightarrow x^a, \quad \text{so that} \quad h(x^a) \approx y$$

What is an inverse problem?

- But operator h is rarely invertible!

- In practice, we use a first guess x^b (a prior knowledge), and implement smart ways to correct this first guess so that it fits the observations better.

$$x^b \longrightarrow x^a, \quad \text{so that} \quad h(x^a) \approx y$$

- Moreover, measurements are almost always contaminated by some errors, as well as the first guess. The solution should not fit the observations exactly, but rather be a trade-off between the first guess and the measurements, based on their errors.

What is an inverse problem?

- Example: you think that the room temperature is 20±1°C. A thermometer tells you it is 19°C, and thermometer precision is 0.5°C. What it the room temperature?

So, what is data assimilation?

- Data assimilation is a specific class of inverse problems where the system is dynamical, and the measurements are distributed in time.

So, what is data assimilation?

- As in many inverse problem, there often are many less observations than system variables to estimate.

- The system dynamics are represented by a model (almost always numerical). This model provides prior information (x^b …) and is an essential ingredient of a data assimilation system.

- This is why data assimilation is often presented (and implemented) as the set of methods designed to "constrain models with time-distributed observations".

Historical legacy

- Geophysical data assimilation was first introduced in meteorology to improve Numerical Weather Forecast (with a numerical model).

- This consolidates the viewpoint according to which data assimilation consists in “constraining models with observations”.

Important details

- In geophysics, the data assimilation challenge includes the dimension problem: the size of problems at stake is a strong constraint for the assimilation methods.
- In practice, many “elegant” methods classical in inverse problem theory are not applicable to data assimilation.
- The geoscience data assimilation community has developed its own set of methods: Ensemble Kalman filters, 3DVar, 4DVar, Optimal Interpolation, etc. These methods strongly rely on models dynamics.

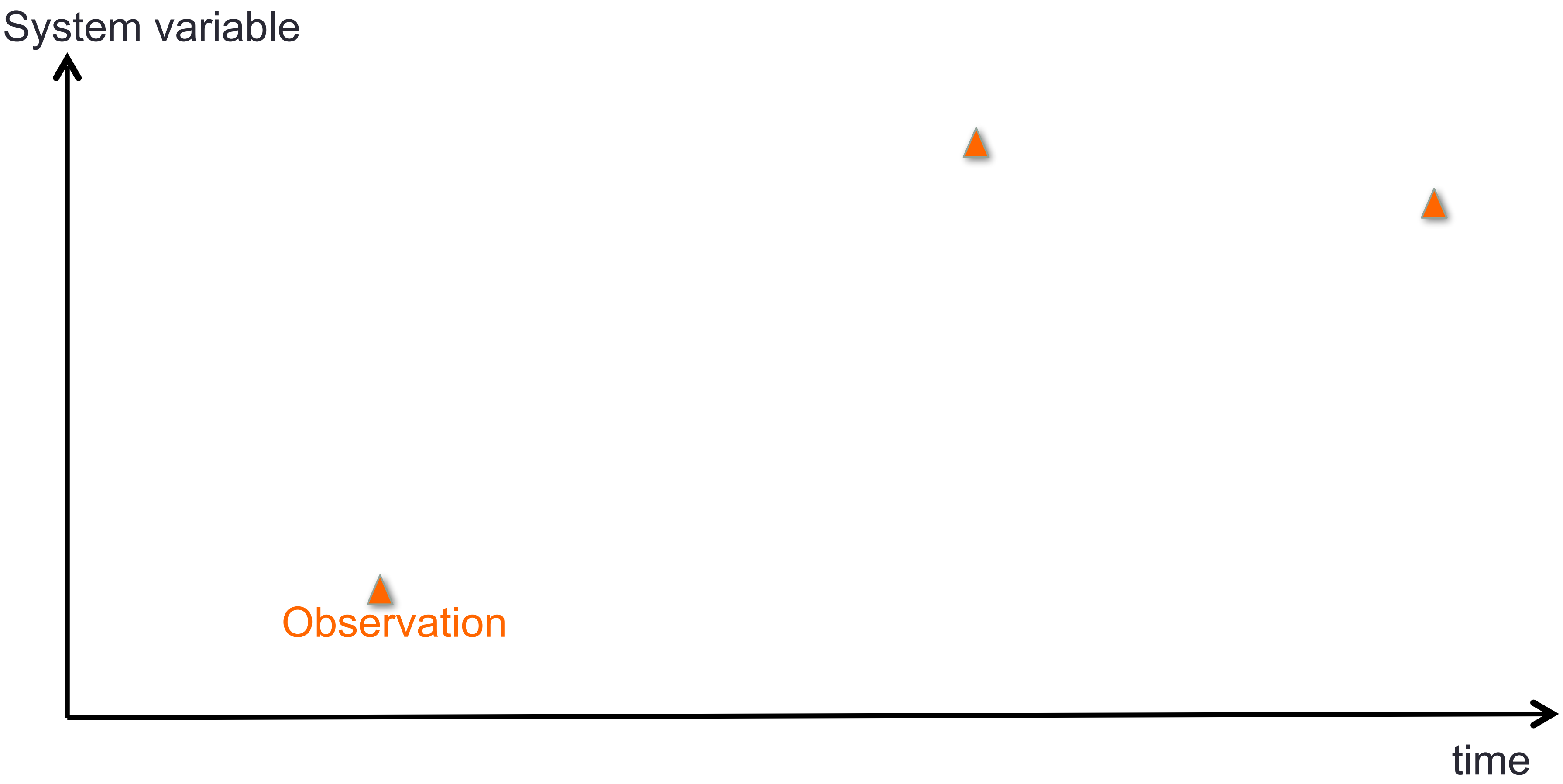
Possible applications of data assimilation

- Forecast initialization;
- estimation of the trajectory of a system to study its variability (reanalyses);
- identification of systematic errors in numerical models;
- optimization of observation networks;
- estimation of unobserved variables;
- estimation of parameters.

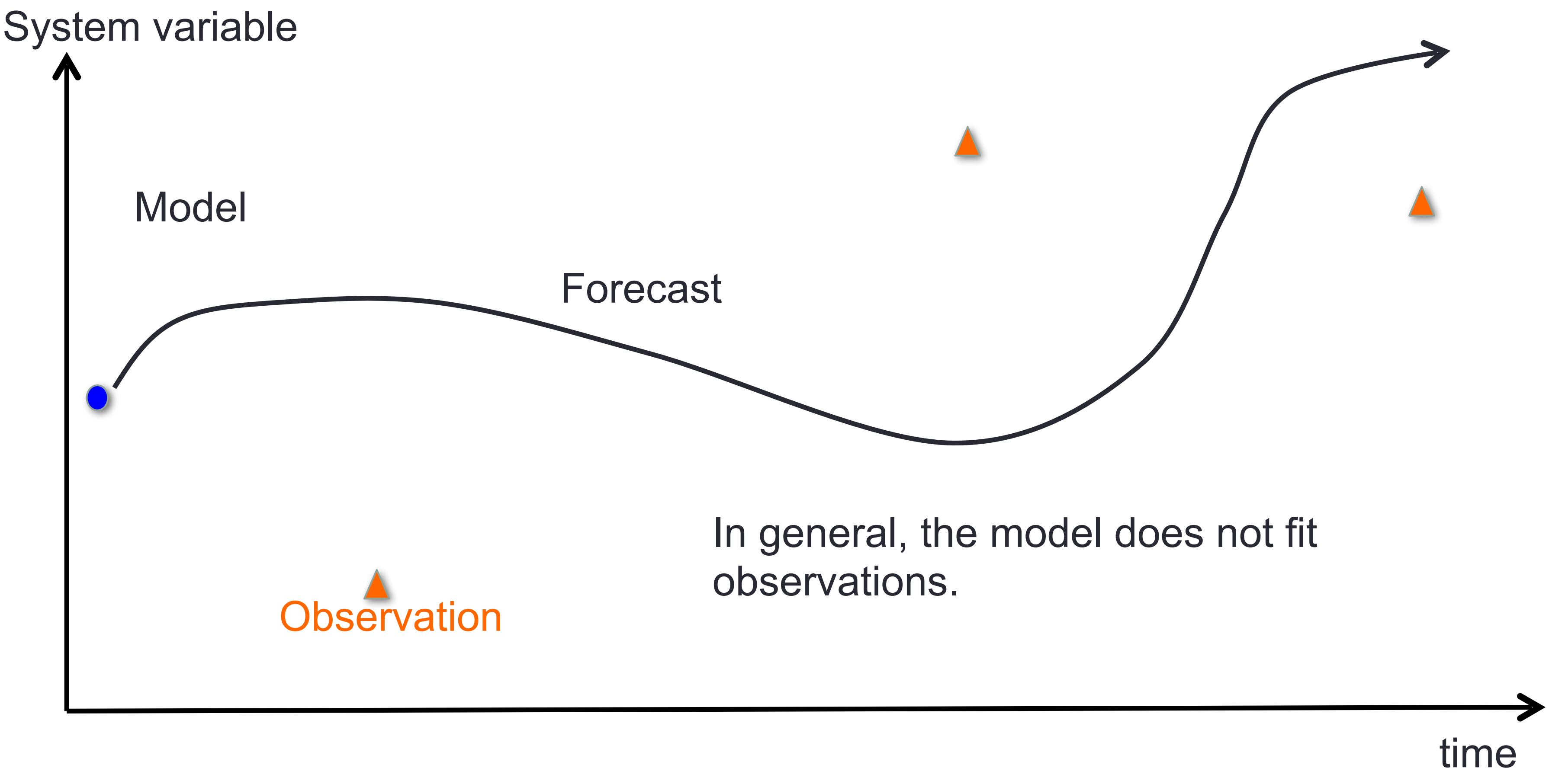
Data assimilation methods

- Three classes of methods: Variational, Statistical, and others… ;
- The first two have strong theoretical connections and equivalences but different implementations;
- They can be hybridized (and are, more and more);
- With high-dimensional systems, statistical methods are generally implemented using ensembles: ensemble data assimilation.

Data assimilation concept

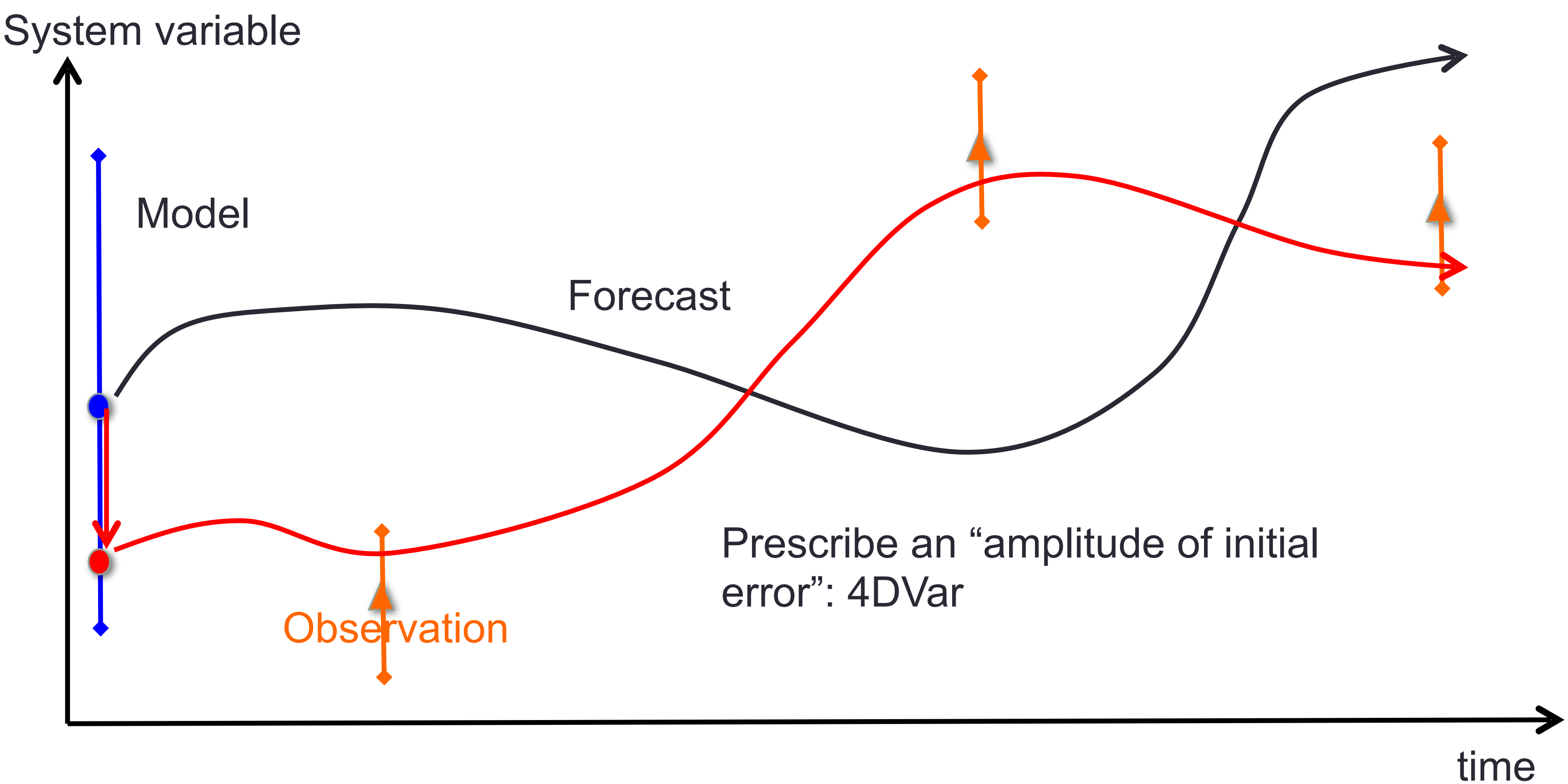


Data assimilation concept



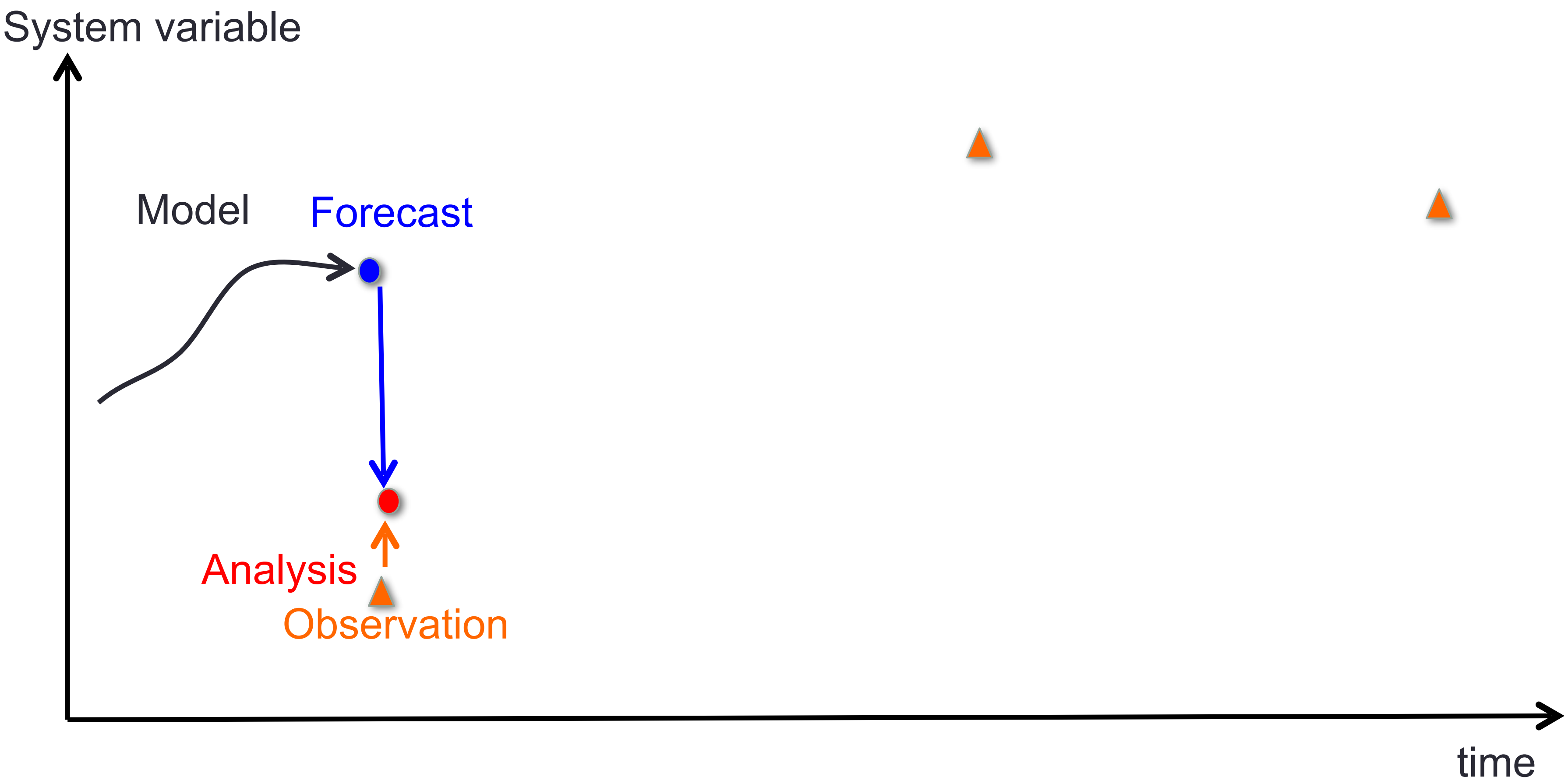
The "poor man's" DA method: nudging

4DVar data assimilation

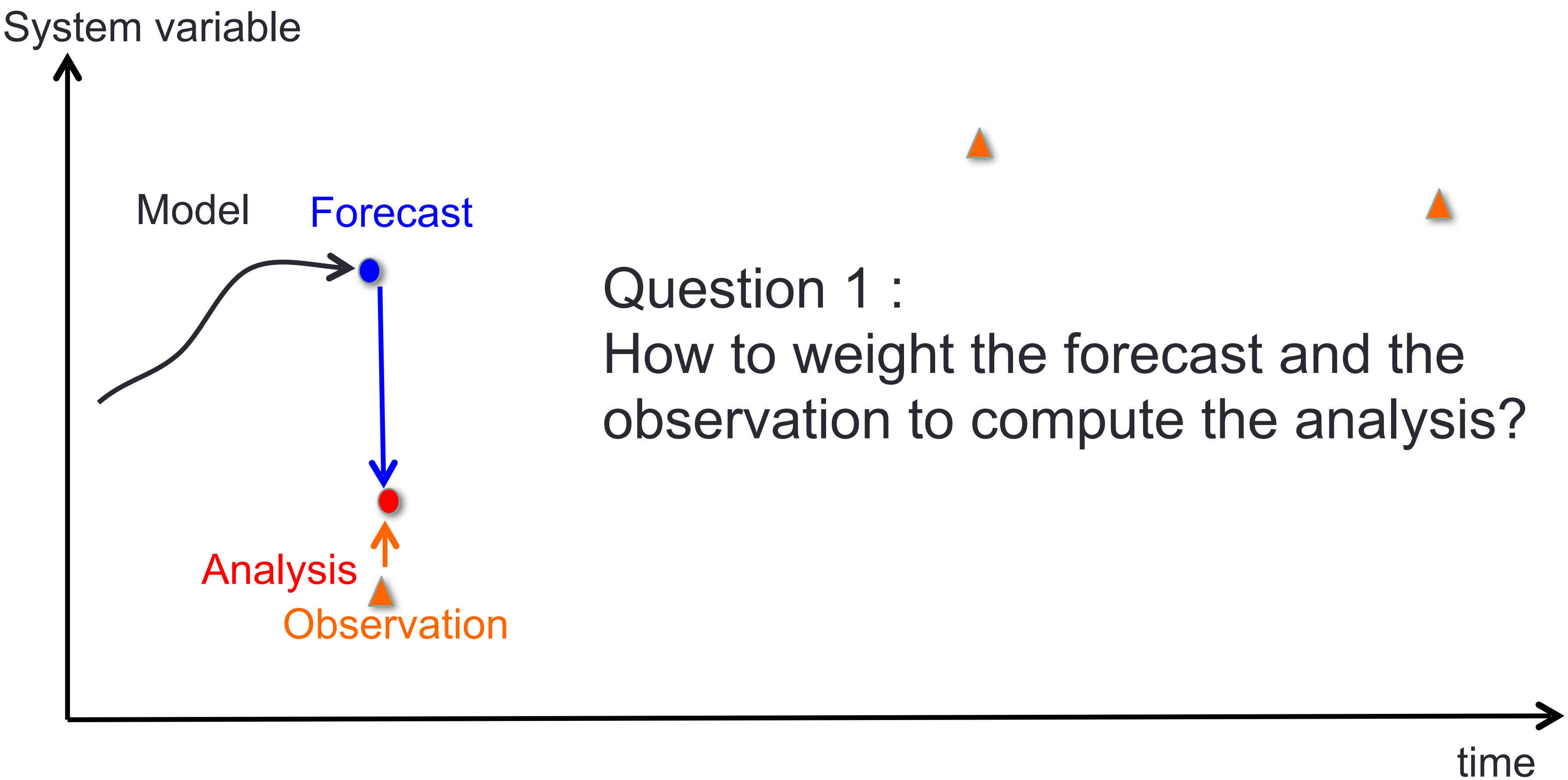


Sequential data assimilation

Sequential data assimilation

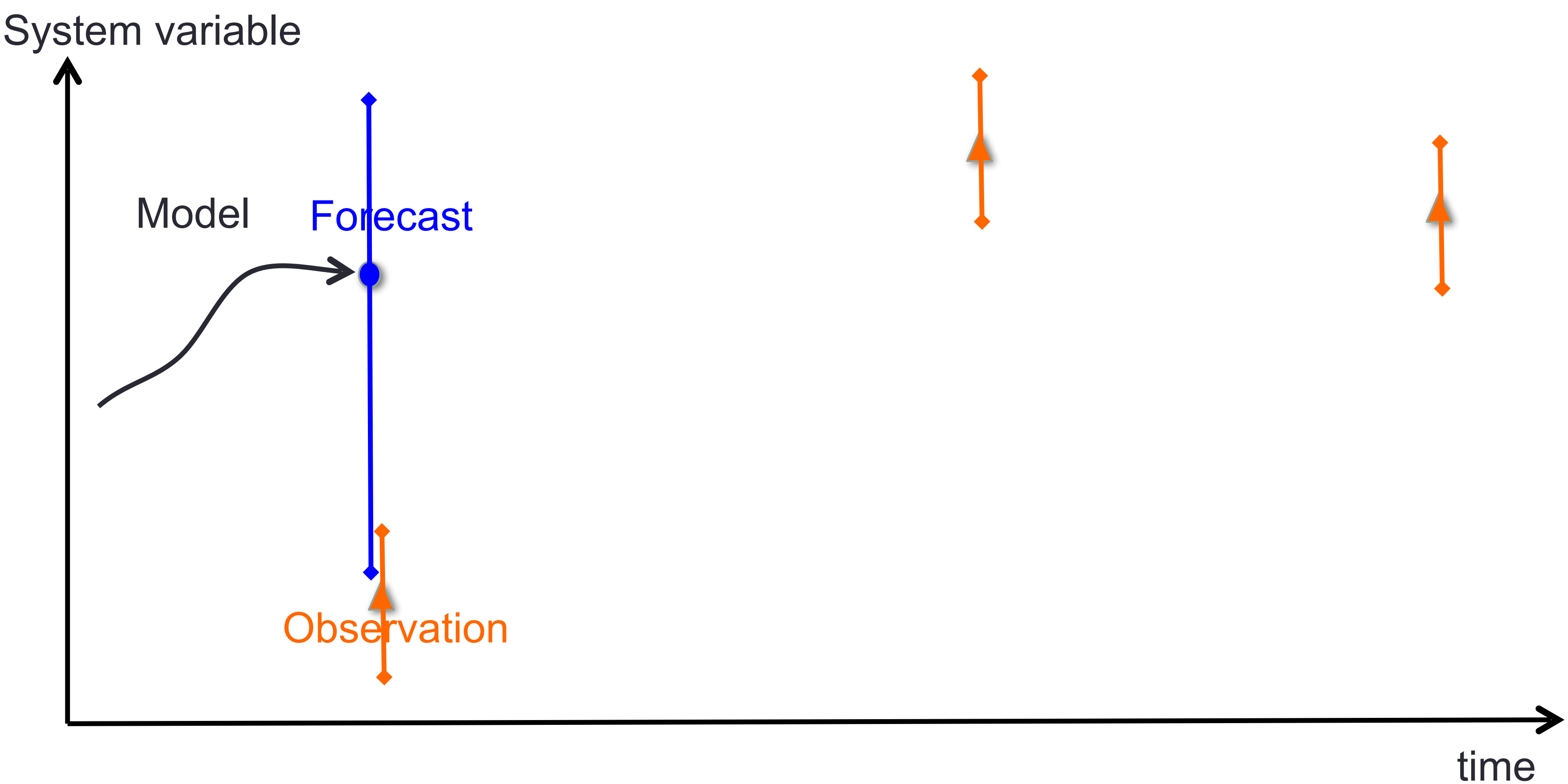


Sequential data assimilation

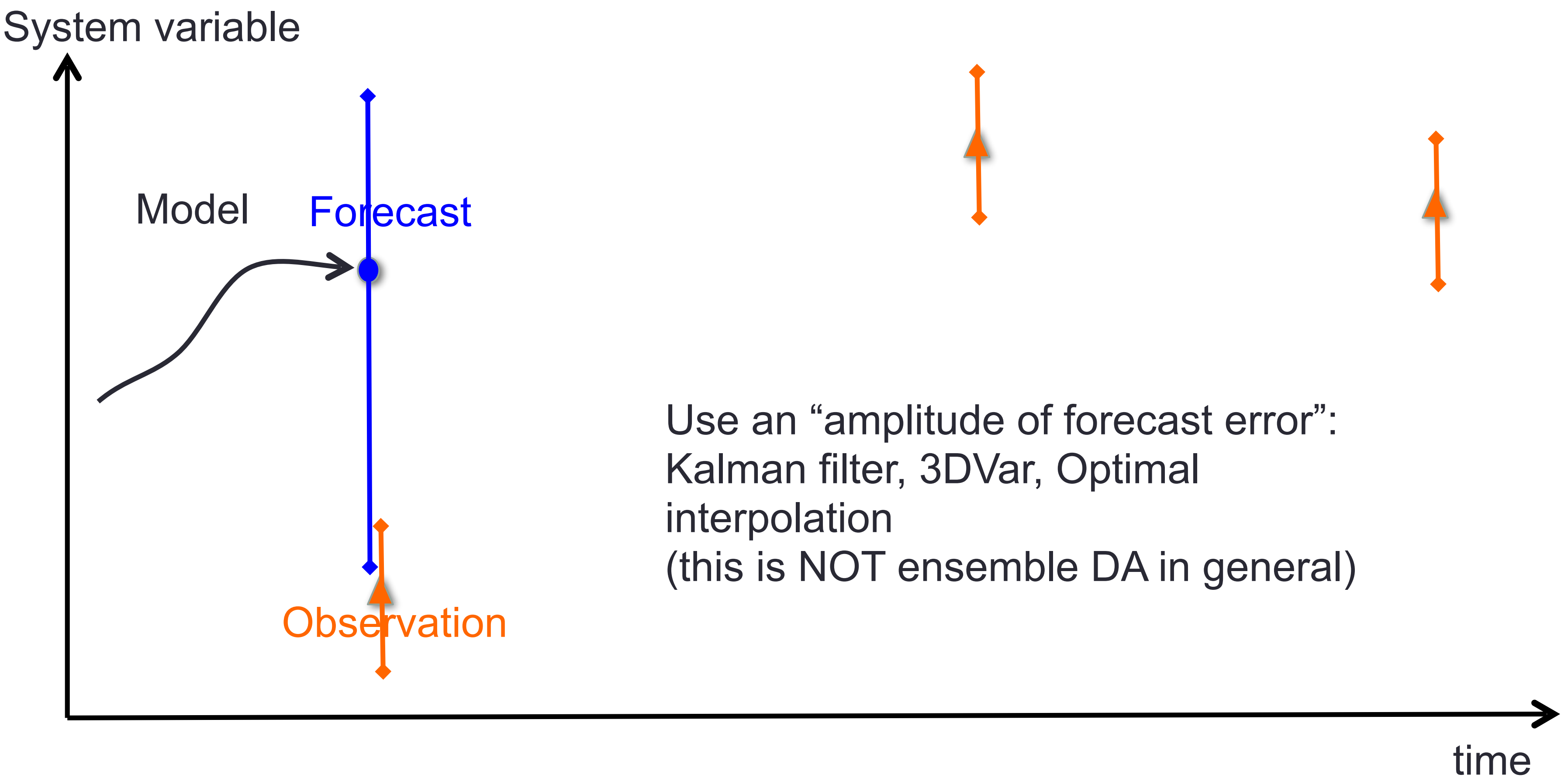


Question 1 :
How to weight the forecast and the observation to compute the analysis?

Sequential data assimilation

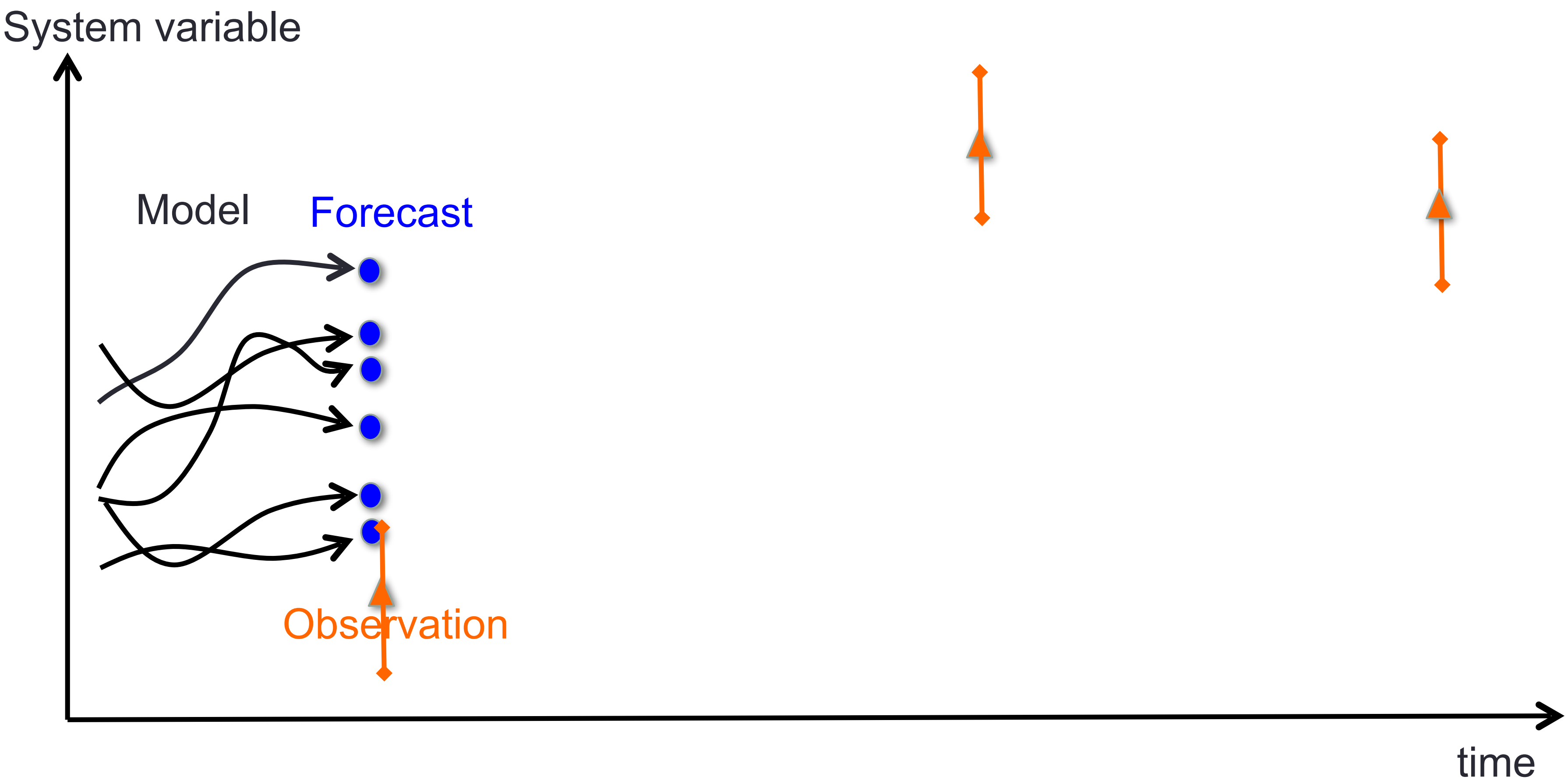


Sequential DA without ensemble: Kalman filter, 3DVar, OI

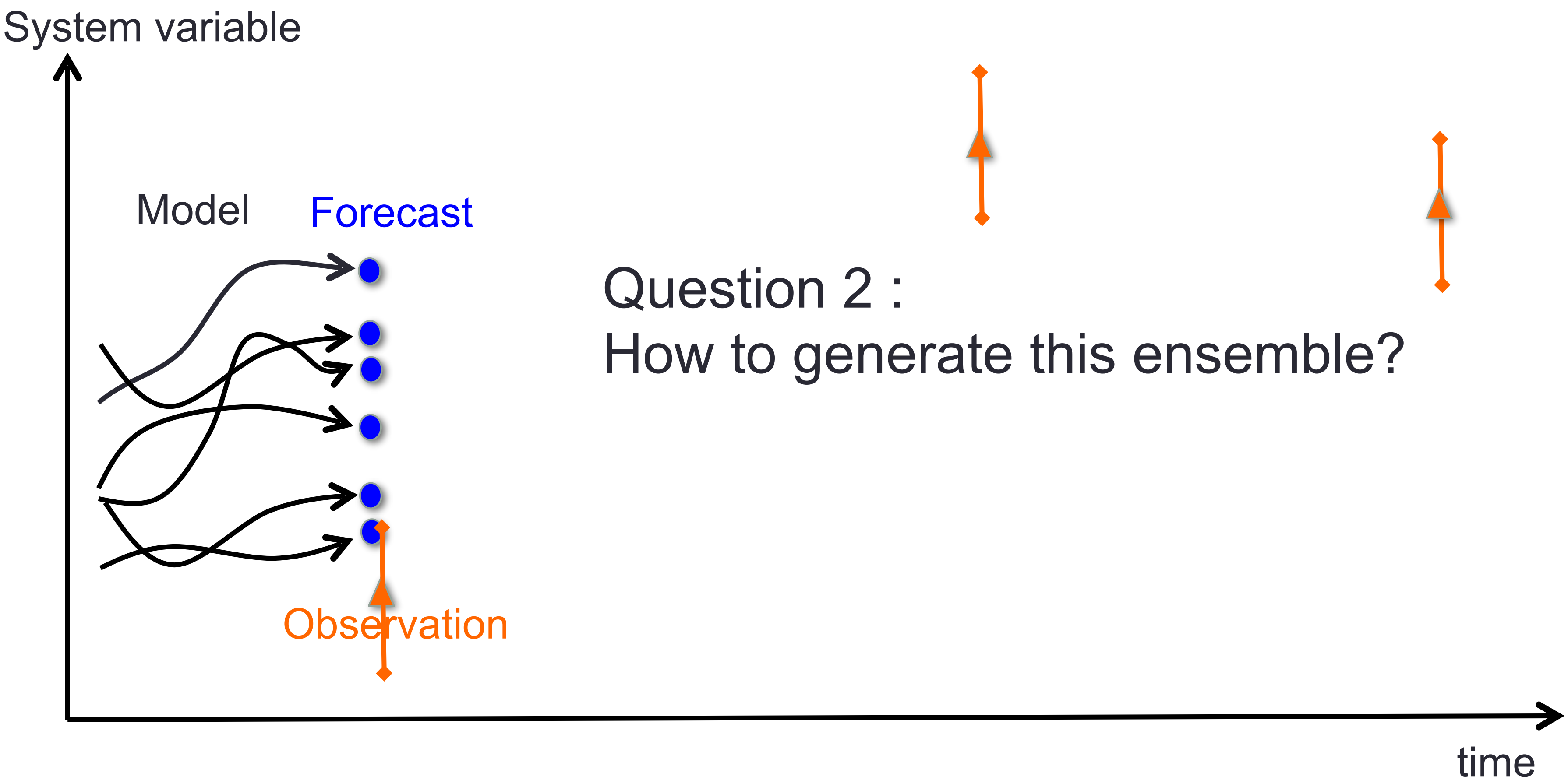


Use an “amplitude of forecast error”:
Kalman filter, 3DVar, Optimal interpolation
(this is NOT ensemble DA in general)

Ensemble data assimilation



Ensemble data assimilation



Question 2 :
How to generate this ensemble?

How to generate the ensemble?

Dynamical model:

$$\mathbf{x}_k = \mathcal{M}(\mathbf{x}_{k-1}, \mathbf{F}_{k-1}, \mathbf{w}_{k-1})$$

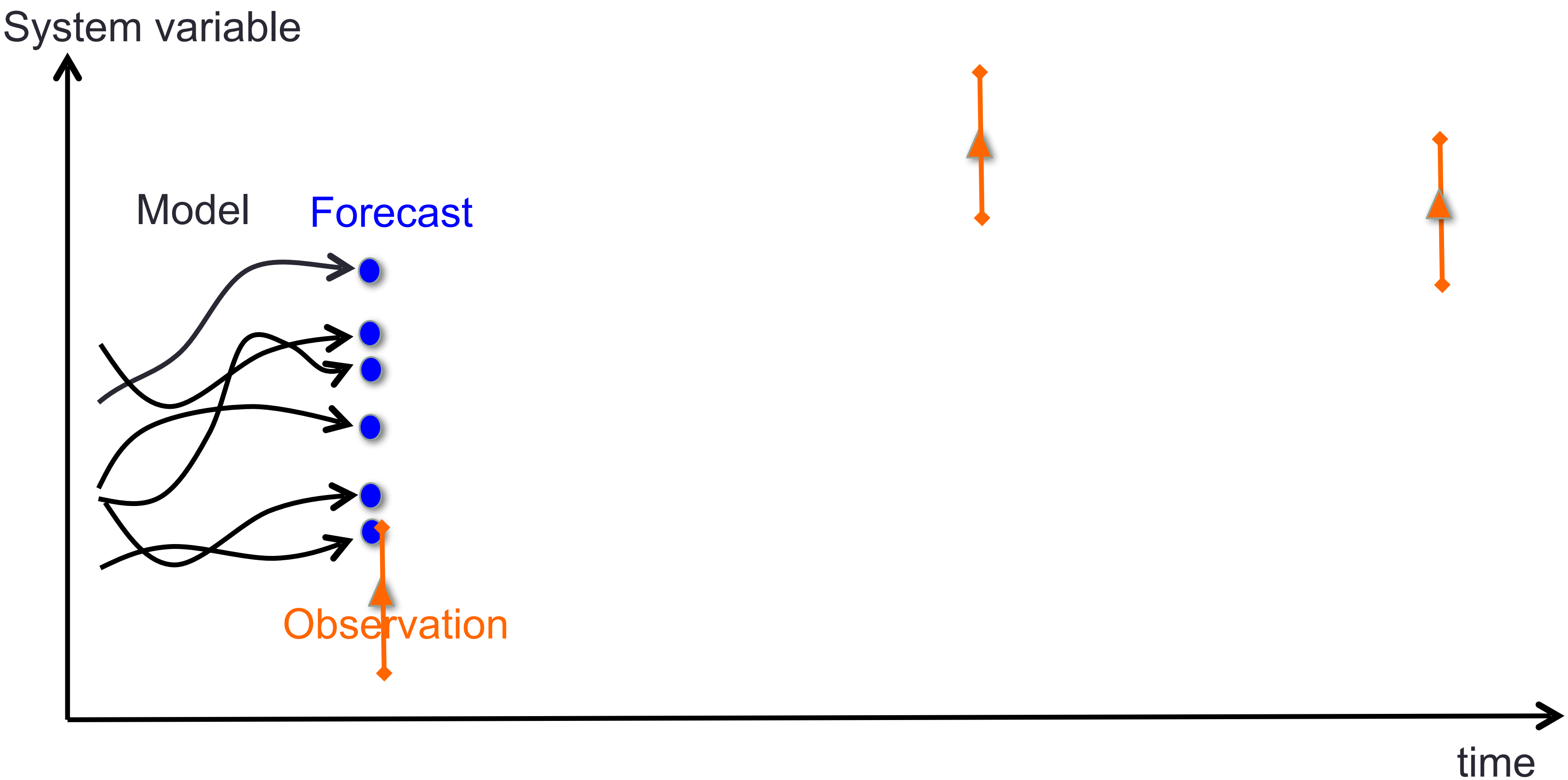
System state vector at time k

External forcings (potentially uncertain)

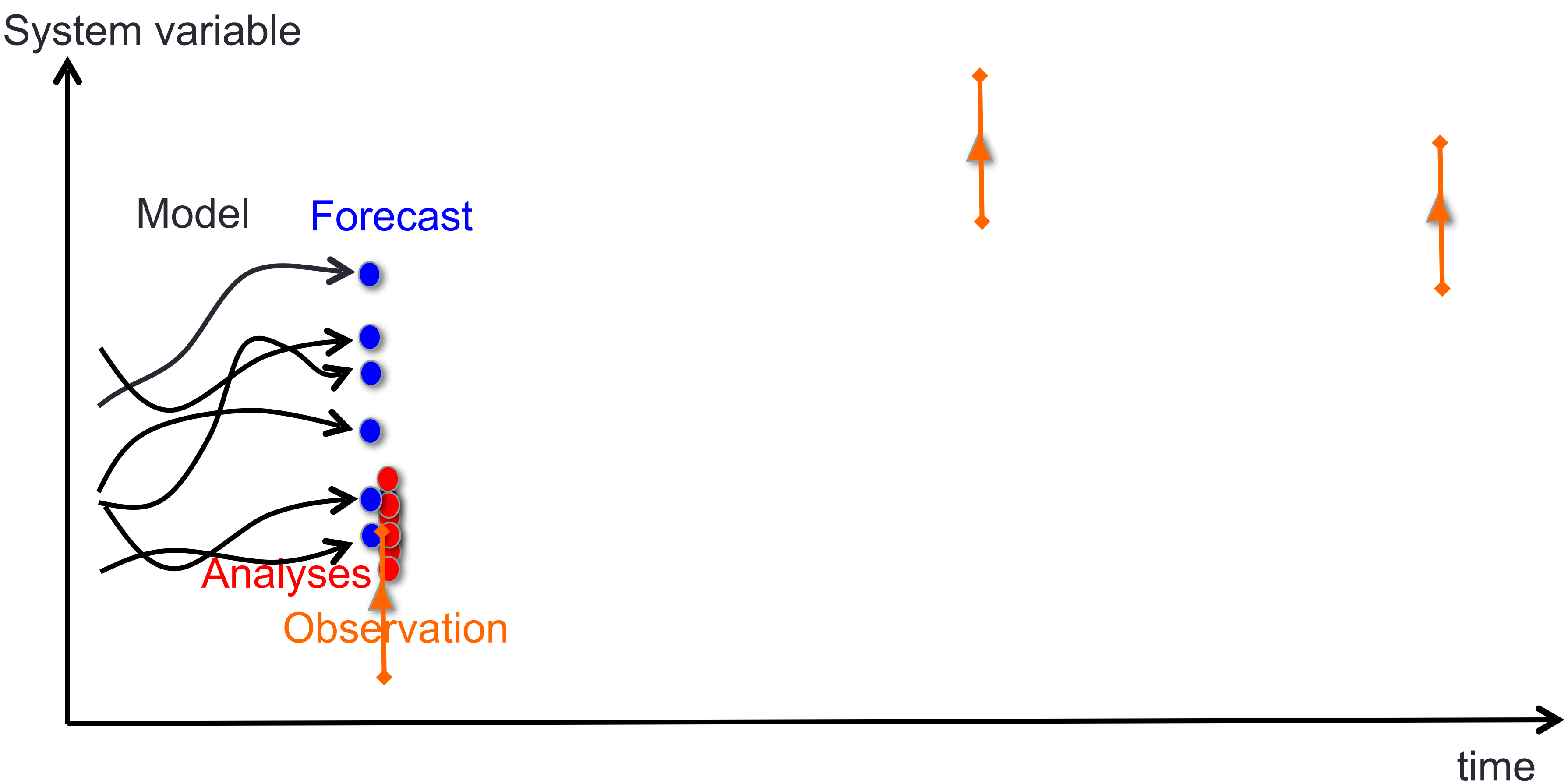
System state vector at time k-1 (potentially uncertain)

Model's intrinsic uncertainties

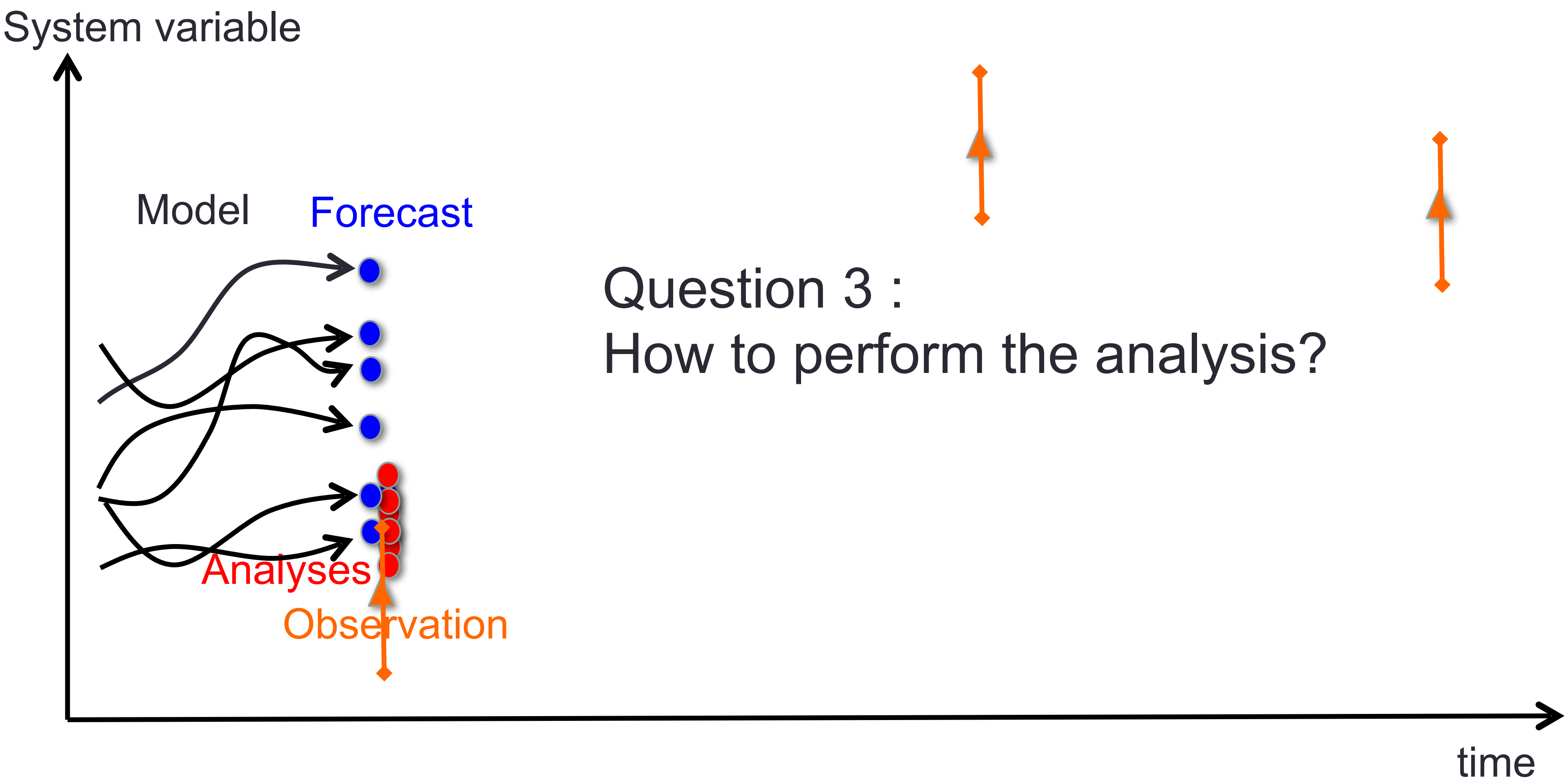
Ensemble data assimilation



Ensemble data assimilation

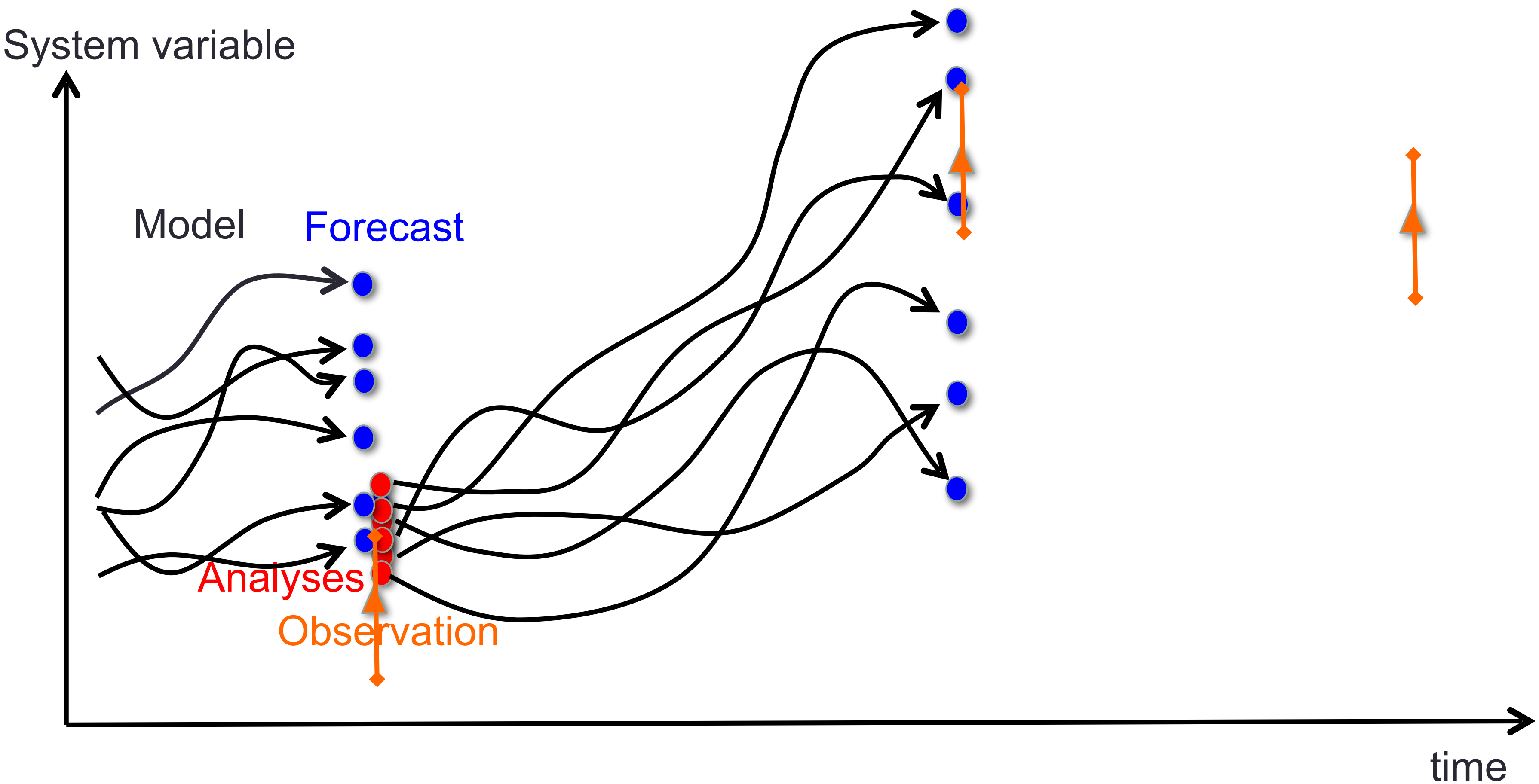


Ensemble data assimilation

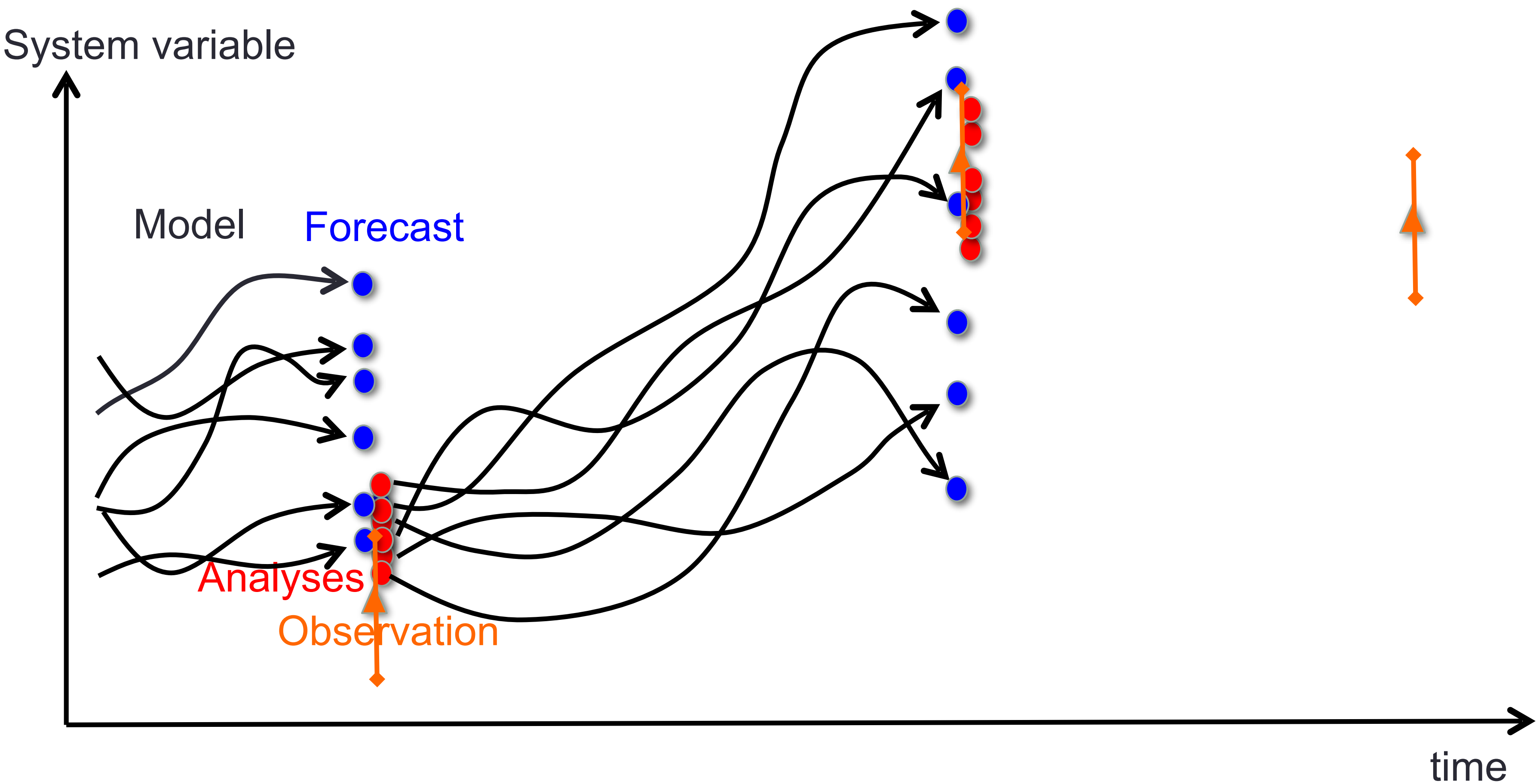


Question 3 :
How to perform the analysis?

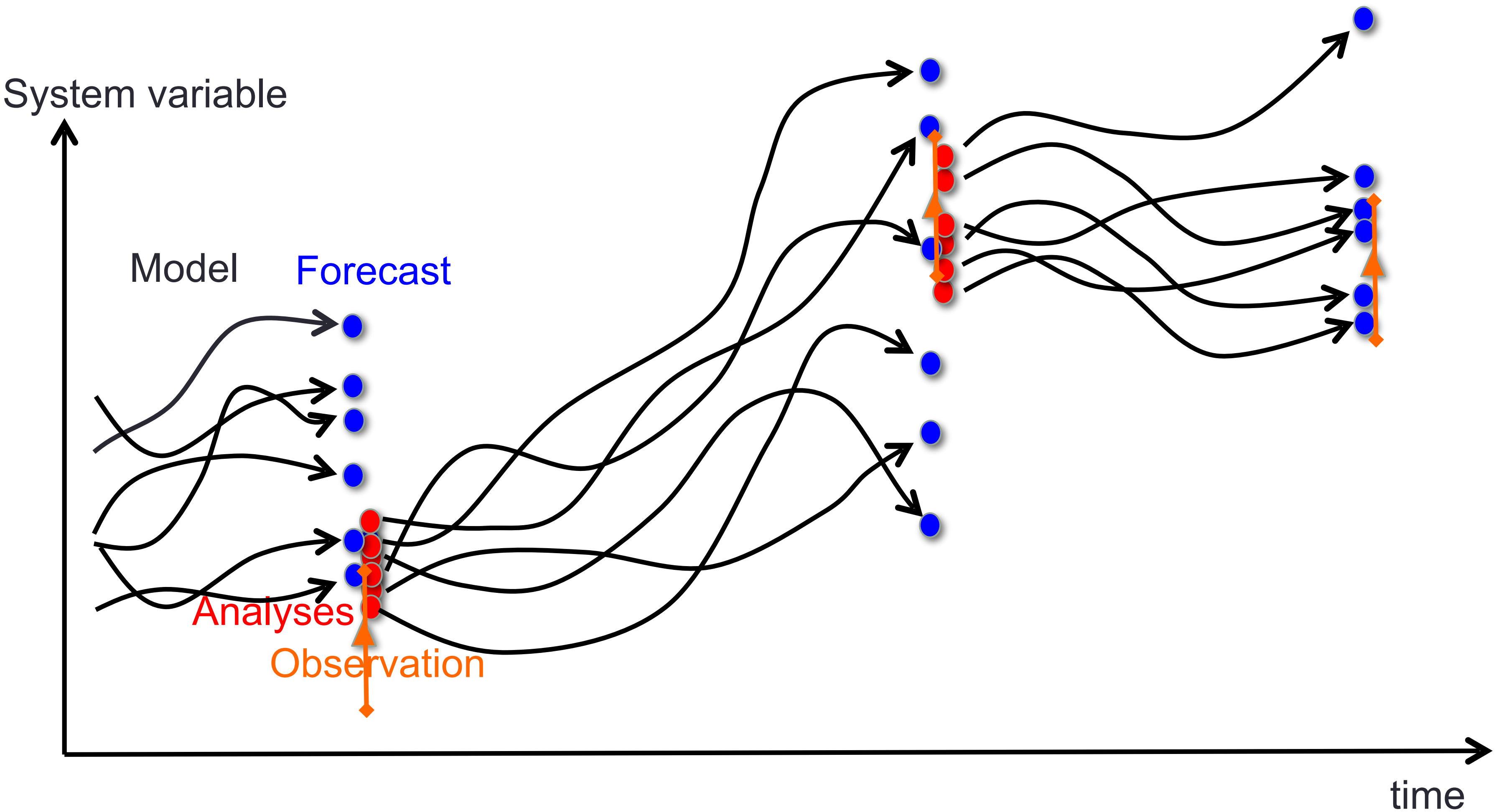
Ensemble data assimilation



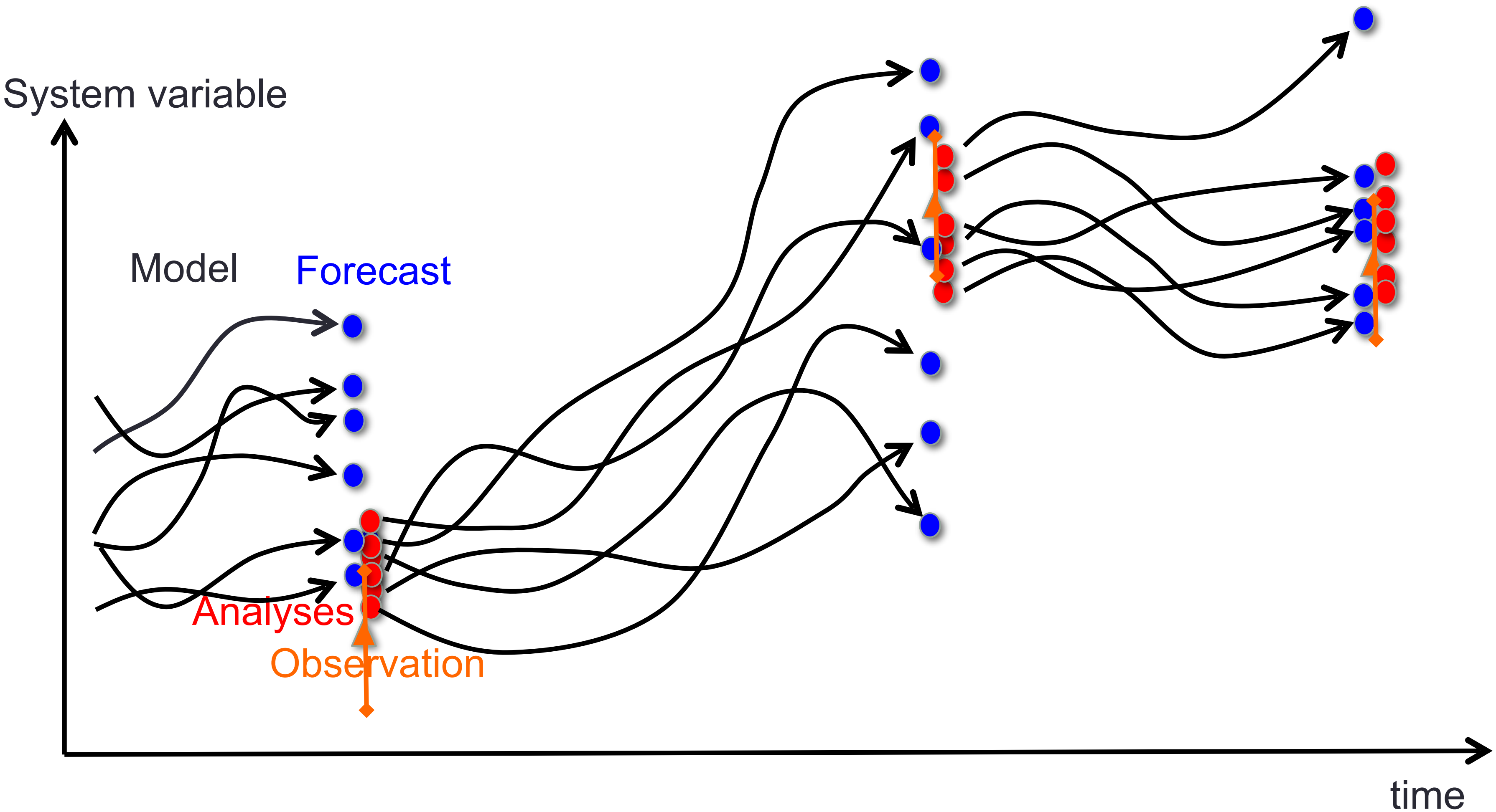
Ensemble data assimilation



Ensemble data assimilation



Ensemble data assimilation



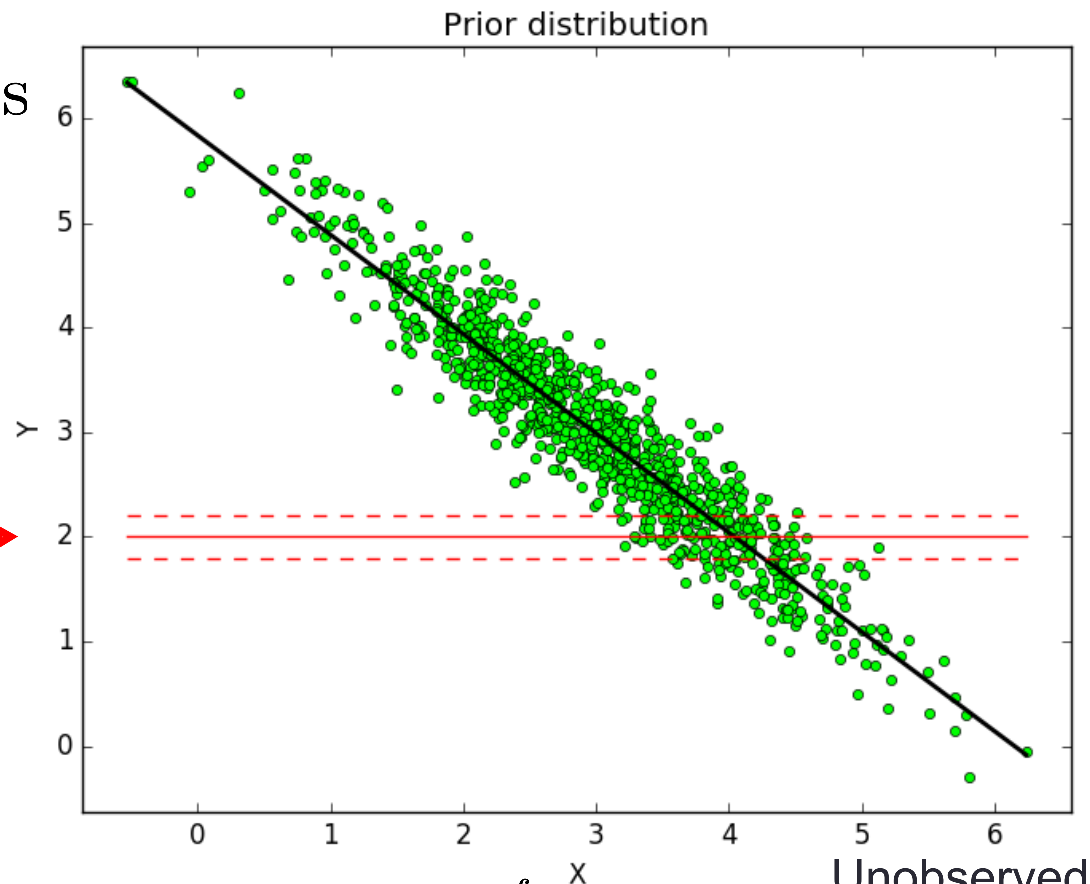
Ensemble analysis

- Ensemble methods really differ from each other by the analysis scheme;
- Two families of method:
  - Sampling methods (e.g. particle filter);
  - Transformation methods (e.g. Ensemble Kalman filter);

Principle of the particle filter

$x^f_{k,\text{obs}}$

Observed variable



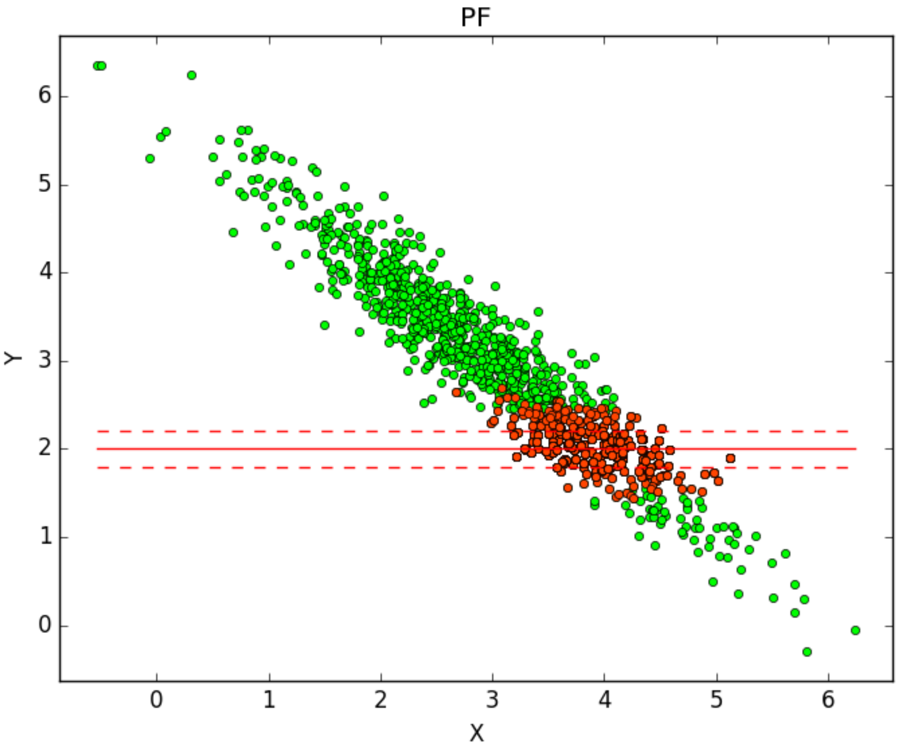
y_k

Observation

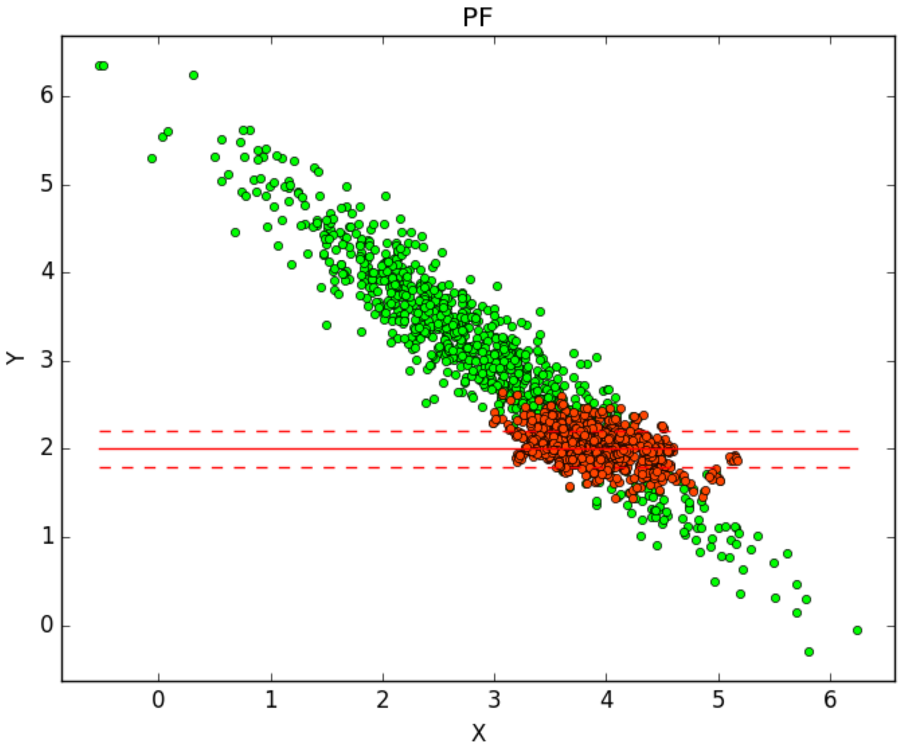
$x^f_{k,\text{non obs}}$

Unobserved variable

Principle of the particle filter



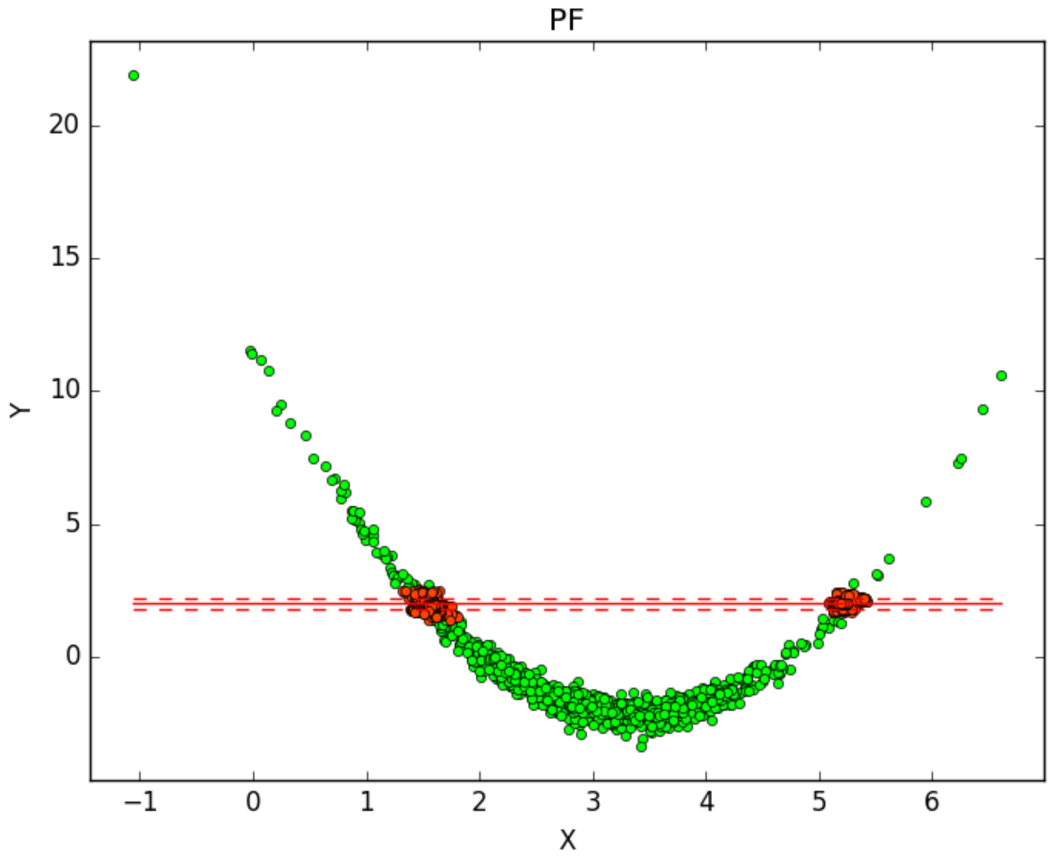
1. Selection



2. Resampling

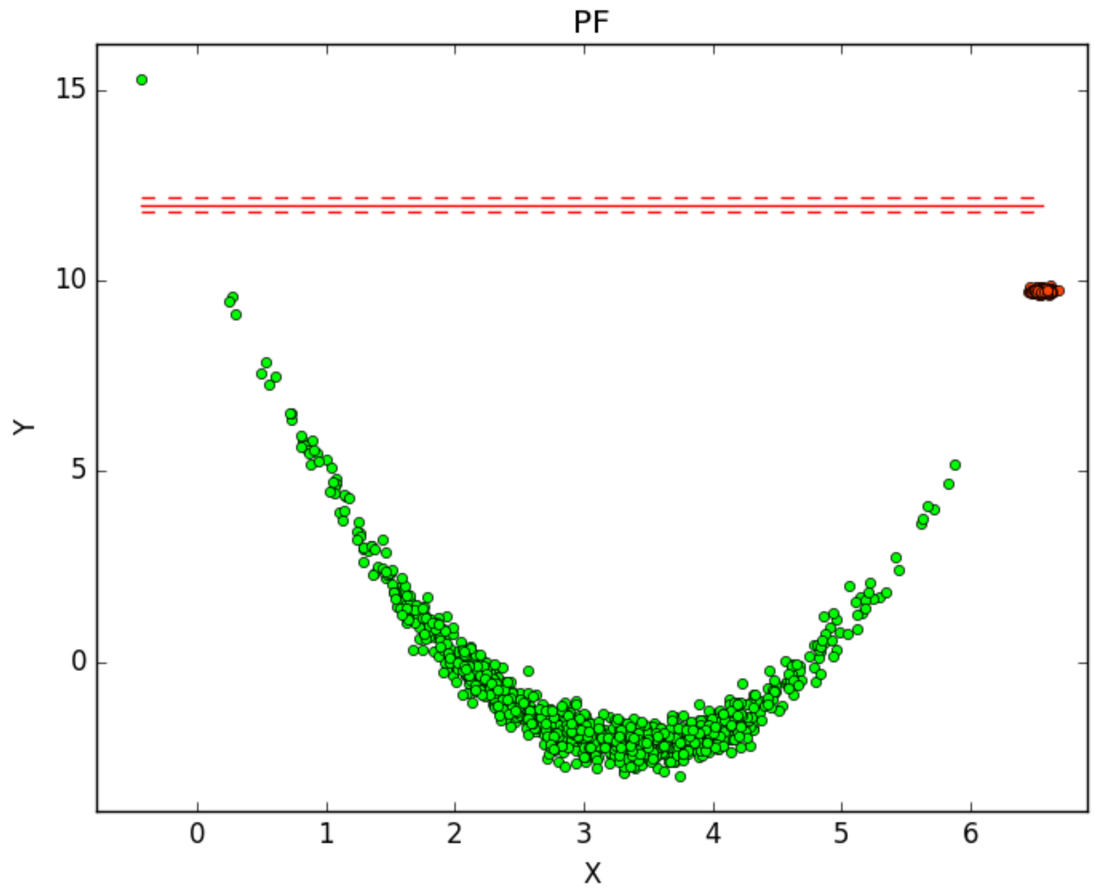
Advantage of the particle filter

- Valid for nonlinear models, complex distributions…
- Easy and quick to code



Drawback of the particle filter

- Highly subject to the curse of dimensionality.

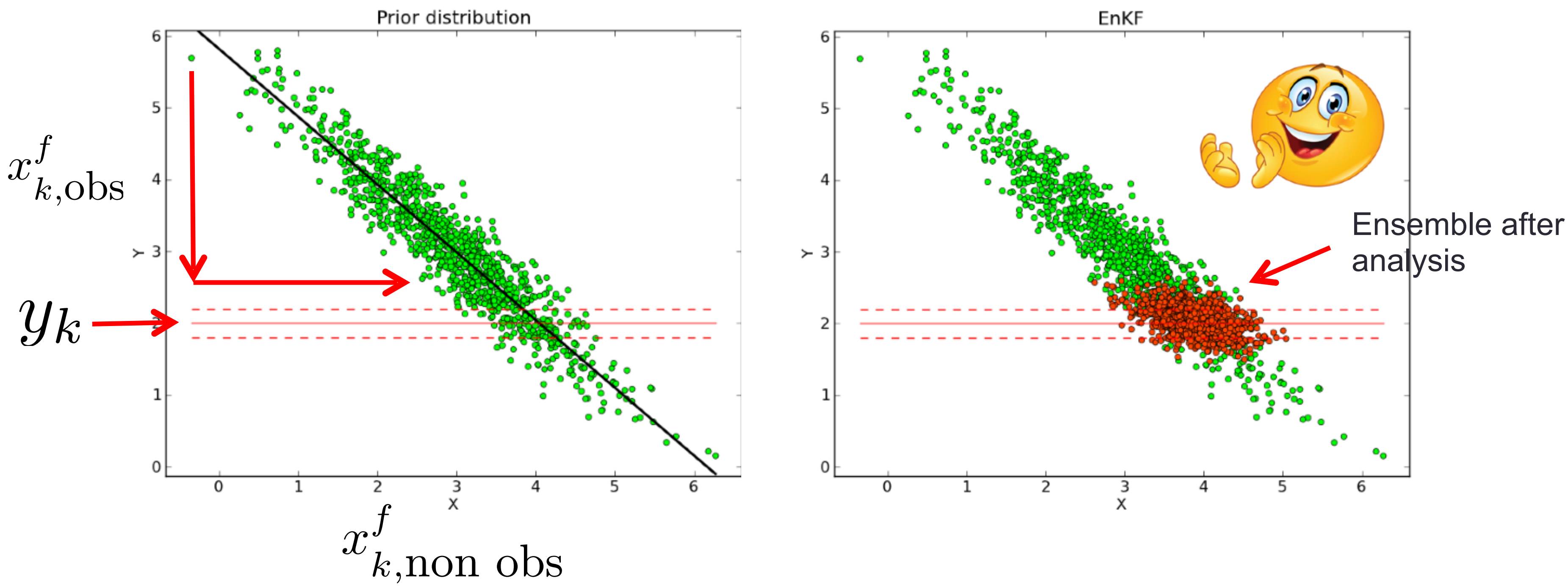


Ensemble analysis

- Ensemble methods really differ from each other by the analysis scheme;
- Two families of method:
  - Sampling methods (e.g. particle filter);
  - Transformation methods (e.g. Ensemble Kalman filter);

Principle of the Ensemble Kalman filter

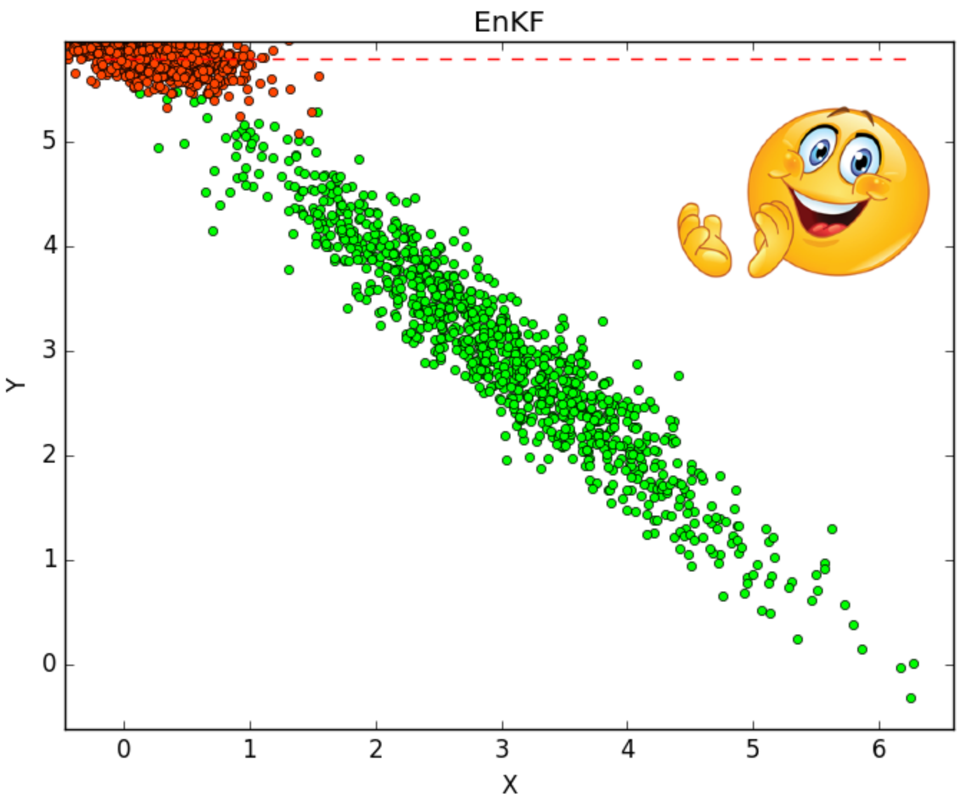
- Each particle is transformed, i.e. drawn closer to the observation



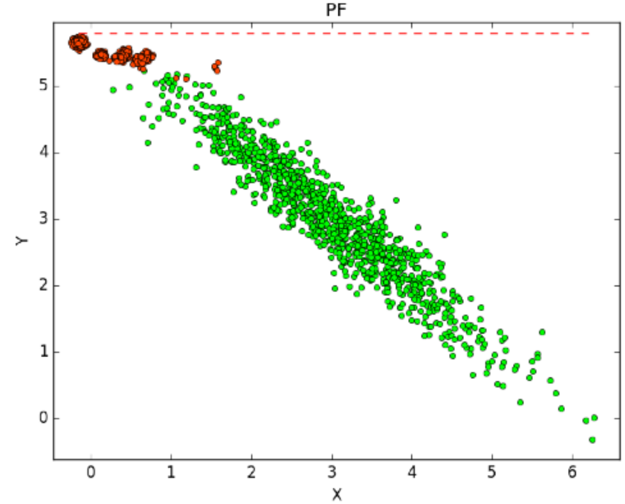
(Anderson, 2003)

Advantage of the Ensemble Kalman filter

- Resilient to the curse of dimensionality

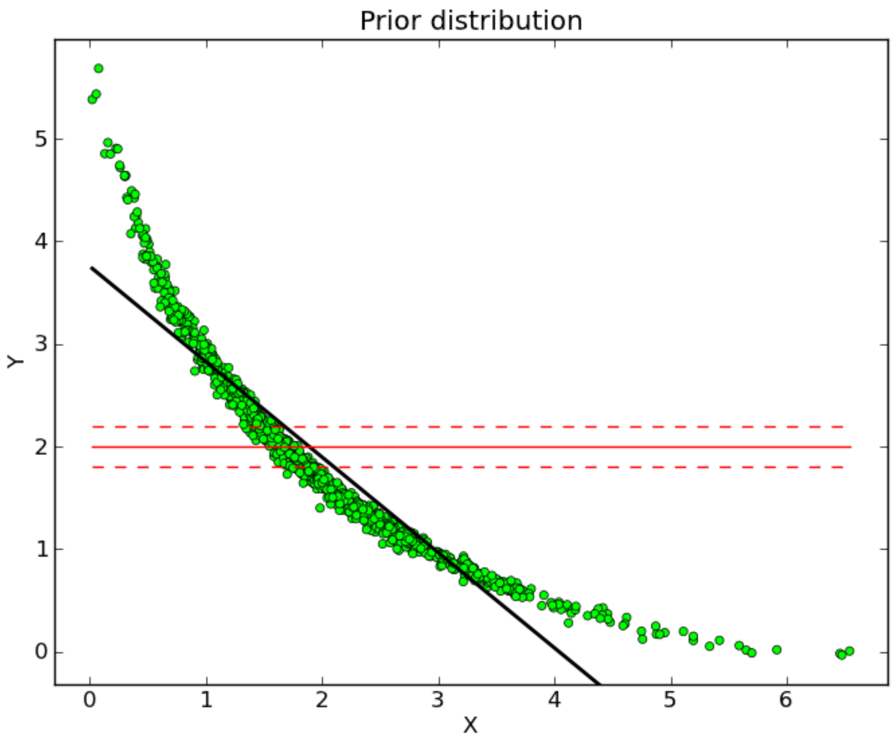


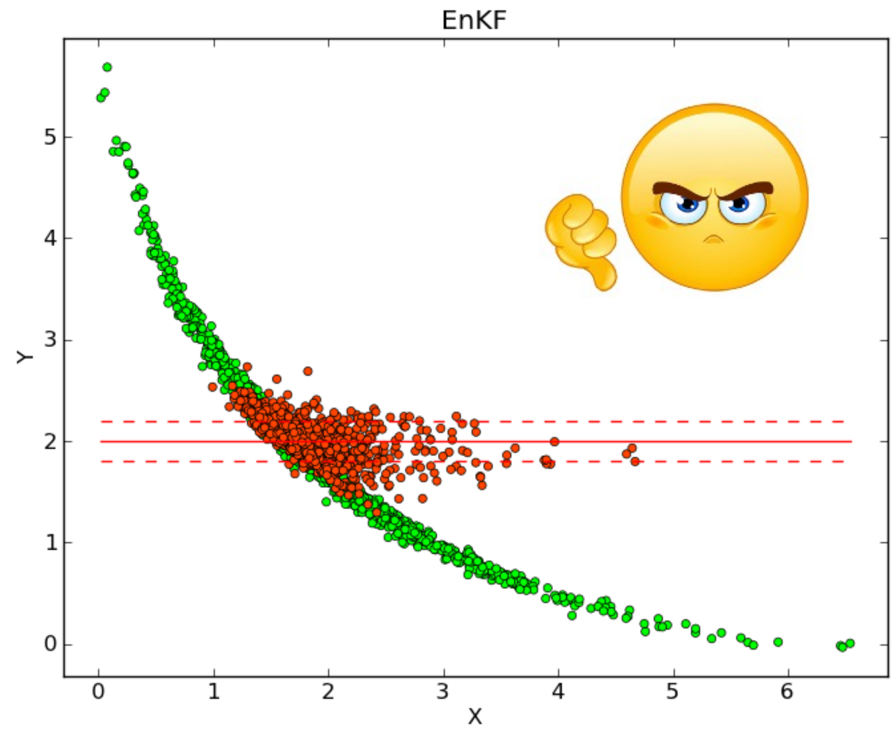
Particle filter



Drawbacks of the Ensemble Kalman filter

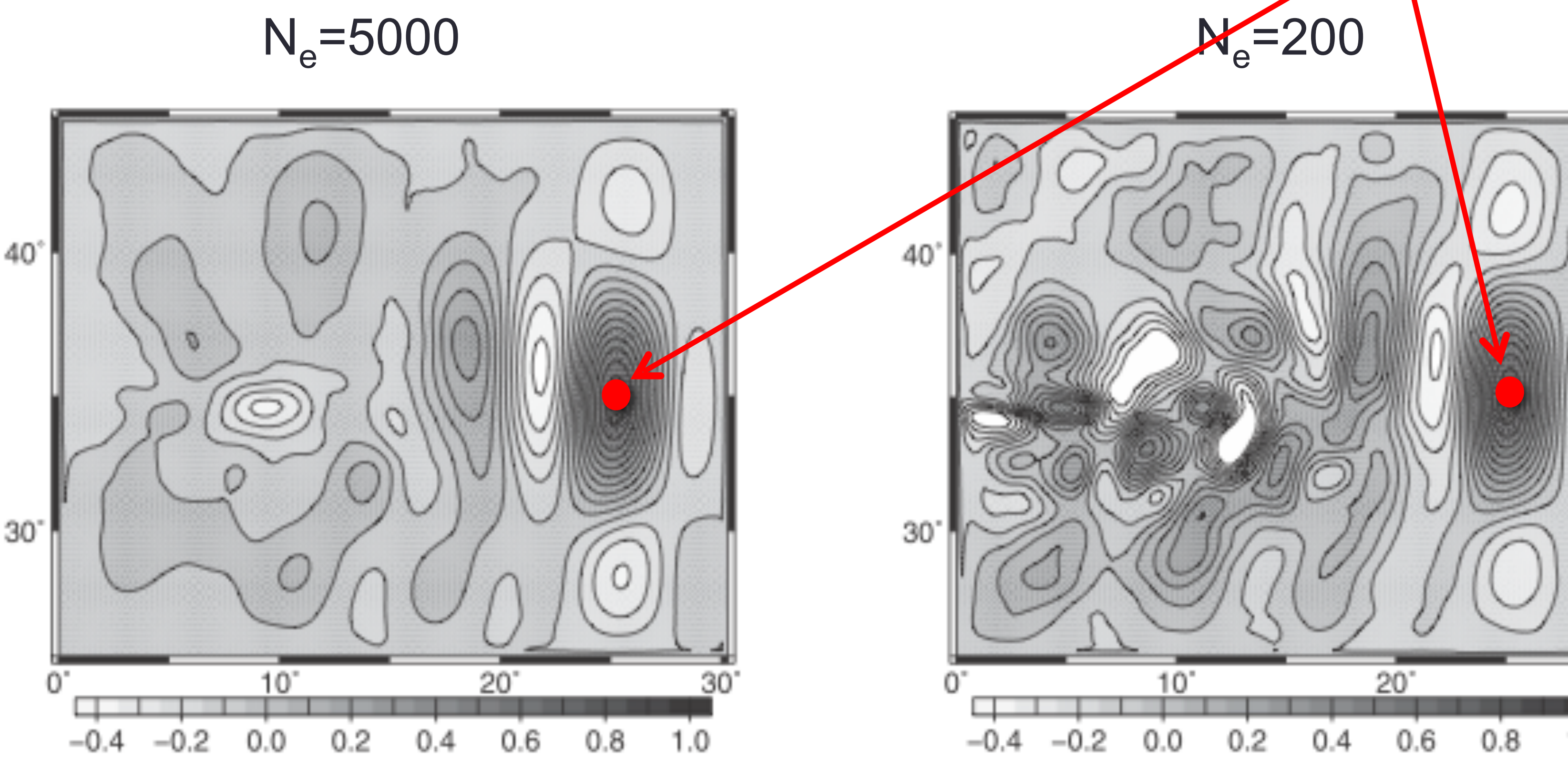
- Based on a linear regression





Drawbacks of the Ensemble Kalman filter

Increment of correction due to an observation here:



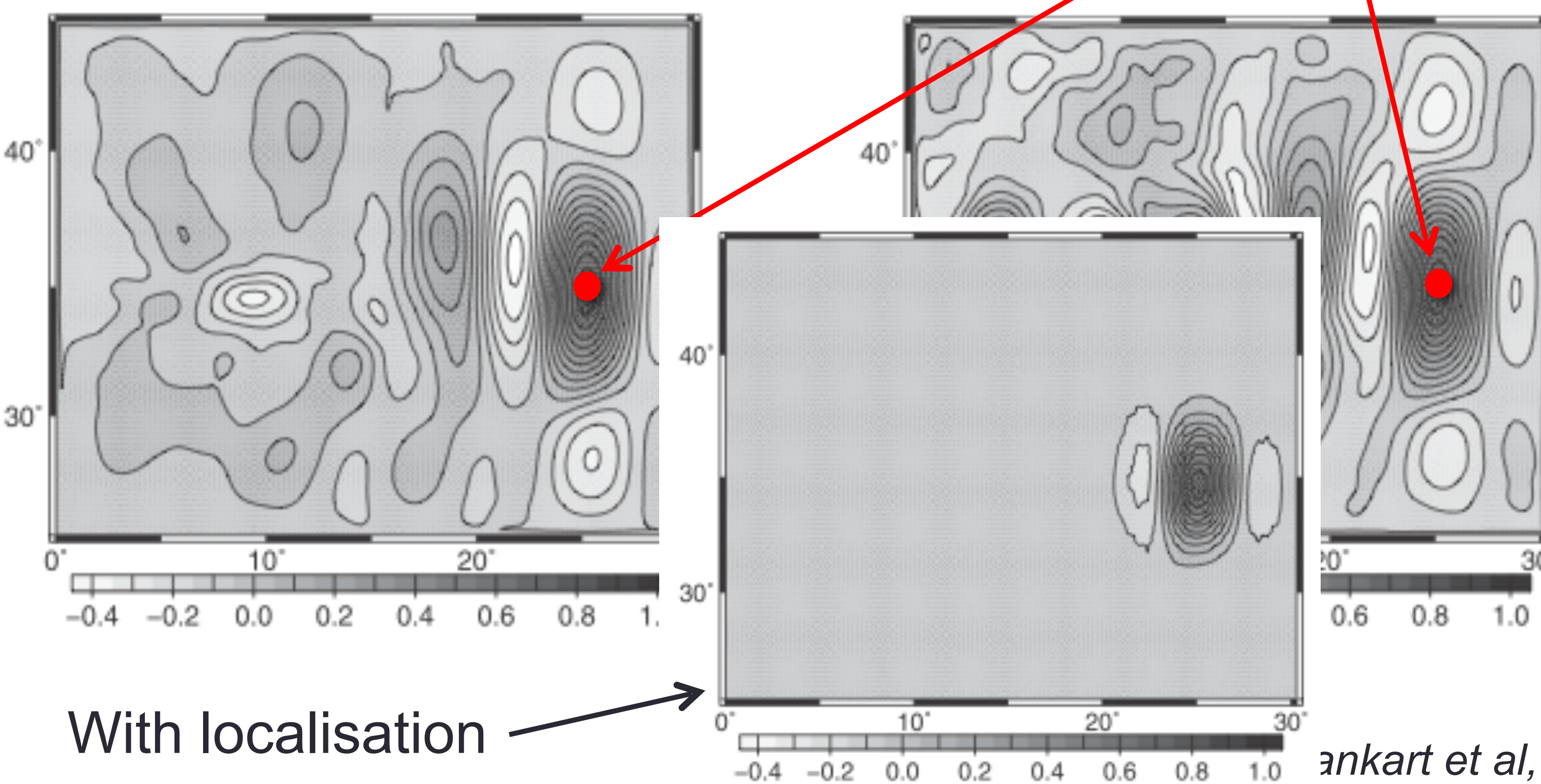
(Brankart et al, 2011)

Drawbacks of the Ensemble Kalman filter

Increment of correction due to an observation here:

N_e=5000

N_e=200



With localisation

ankart et al, 2011)

Drawbacks of the Ensemble Kalman filter

- Based on a linear regression;
  - To some extent, the detrimental effects can be mitigated using smart approaches (e.g. anamorphosis)

- Is also affected by a small ensemble size (subsampling);

- Localization:
  - Limits the unfortunate consequences of subsampling
  - But is quite heavy to implement
  - And annihilates the possible large-scale signature of observations

Summary

- Two families of method:
  - Sampling methods (e.g. particle filter);
  - Transformation methods (e.g. Ensemble Kalman filter);
- Both have pros and cons but only the EnKF works in practice with high-dimension problems.

Observing Systems Experiments (OSEs)

- Goal of an OSE: Measuring the impact of some observing system on the analysis (and forecast).

Masutani et al (2010)

Observing Systems Experiments (OSEs)

- Goal of an OSE: Measuring the impact of some observing system on the analysis (and forecast).

- Steps:
  - Control run assimilating all available observations;
  - Perturbation run assimilating all observations excluding the type under evaluation;
  - Comparison of analysis and/or forecast skills between control and perturbation run.

- Note: easy to perform with an operational system.

Masutani et al (2010)

Example of OSEs

- With the Met Office NWP system, Dumelow (2003)

- Tested systems:
  - 1. NO SONDE: ‘in-situ’ profile observations.
  - 2. NO STRAD: satellite radiance data.
  - 3. NO AMV: AMV data.
  - 4. NO SAT: satellite data in the NO STRAD and NO AMV runs and SSM/I winds.
  - 5. NO AIRCRAFT: aircraft data.
  - 6. NO SURF: observations from the surface network.

Example of OSEs

- With the Met Office NWP system, Dumelow (2003)

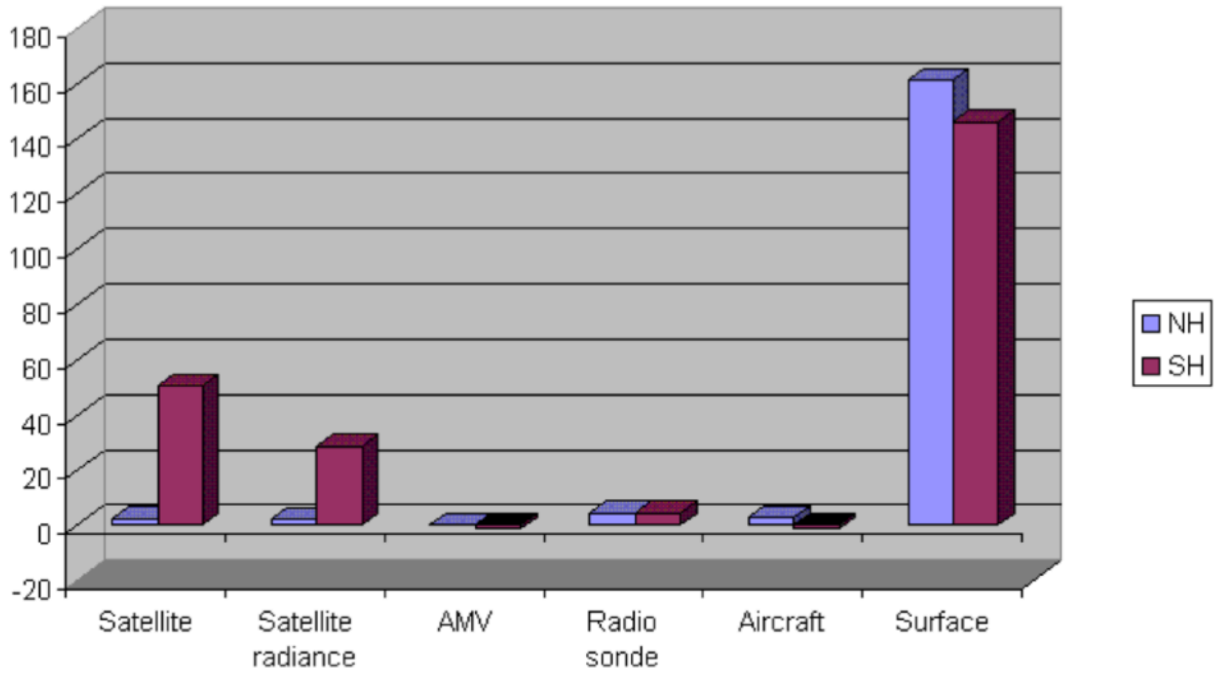


Figure 6 (d) Percentage change in RMS error of 24-hr forecast of mean sea level pressure meaned over the northern and southern hemisphere for all observation types.

Large positive impact of surface observation on mean SLP forecast.

Example of OSEs

- With the Met Office NWP system, Dumelow (2003)

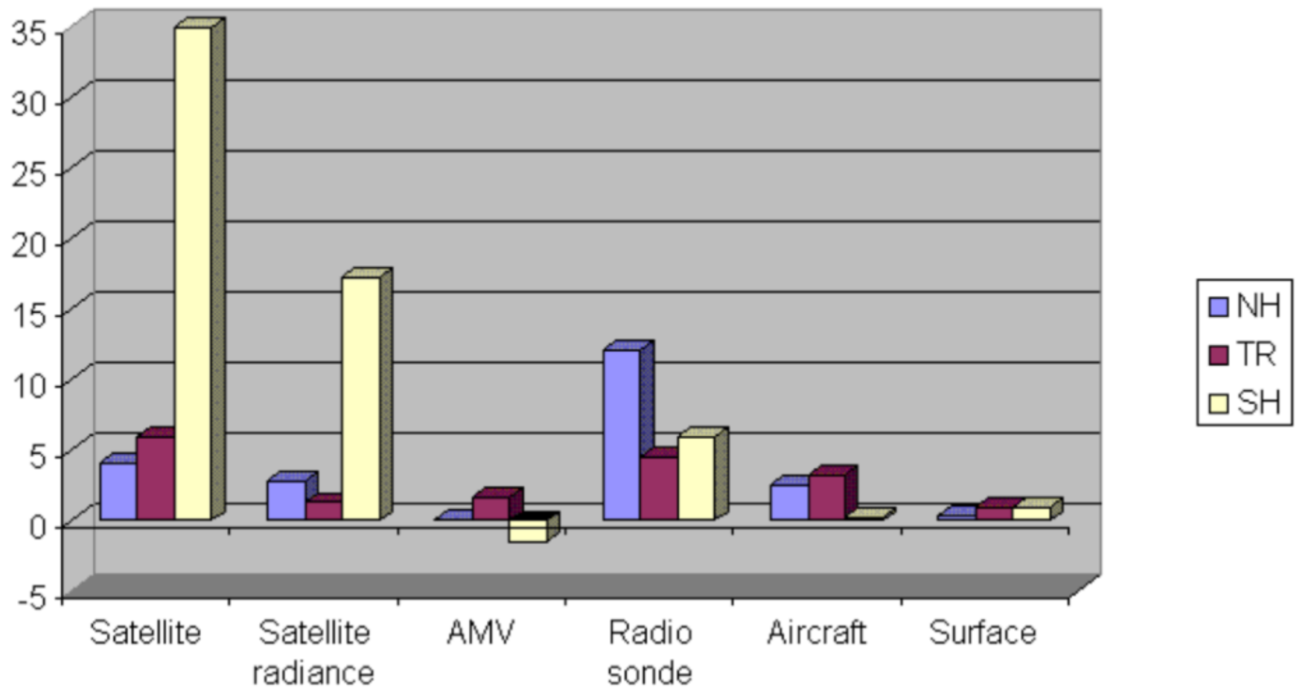


Figure 8. Percentage change in RMS errors (vs observations) for 24-hr forecasts of 250 hPa wind.

Small impact of surface observation on mean 250 hPa wind. Large impact of satellite data in the Southern hemisphere.

Observing System Simulation Experiments (OSSEs)

- Goals of an OSSE:
  - Measuring the impact of a future observing system
  - Design an observation network (to be implemented)
  - Test data assimilation methods

Masutani et al (2010)

Observing System Simulation Experiments (OSSEs)

- Goals of an OSSE:
  - Measuring the impact of a future observing system
  - Design an observation network (to be implemented)
  - Test data assimilation methods

- Steps:
  - Reference run without data assimilation (provides a “truth”);
  - Simulate observations from the truth, including realistic errors;
  - Control run, either:
    - Without data assimilation, but setting different from reference run;
    - Or assimilating the simulated observations of already existing systems;
  - Perturbation run assimilating the simulated observations from the system under study;
  - Comparison of analysis and/or forecast skills between control and perturbation run.

Masutani et al (2010)

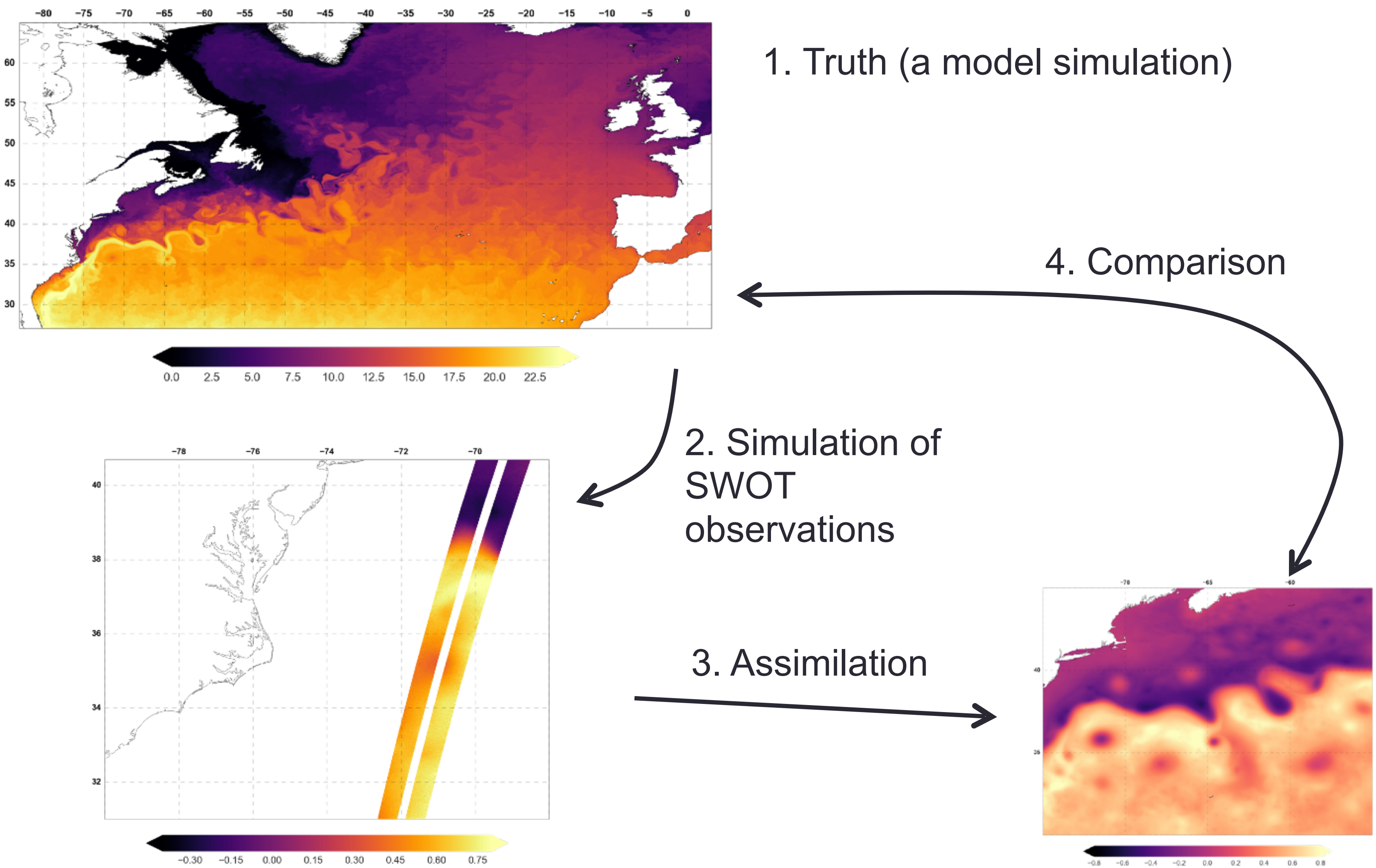
Observing System Simulation Experiments (OSSEs)

- Steps:
  - Reference run without data assimilation (provides a “truth”);
  - Simulate observations from the truth, including realistic errors;
  - Control run, either:
    - Without data assimilation, but setting different from reference run;
    - Or assimilating the simulated observations of already existing systems;
  - Perturbation run assimilating the simulated observations from the system under study;
  - Comparison of analysis and/or forecast skills between control and perturbation run.

- Important note: here the comparison can be made using the truth.

Masutani et al (2010)

Example of an OSSE: SWOT



End of introduction