A quick introduction to data assimilation

Alexandre Fournier

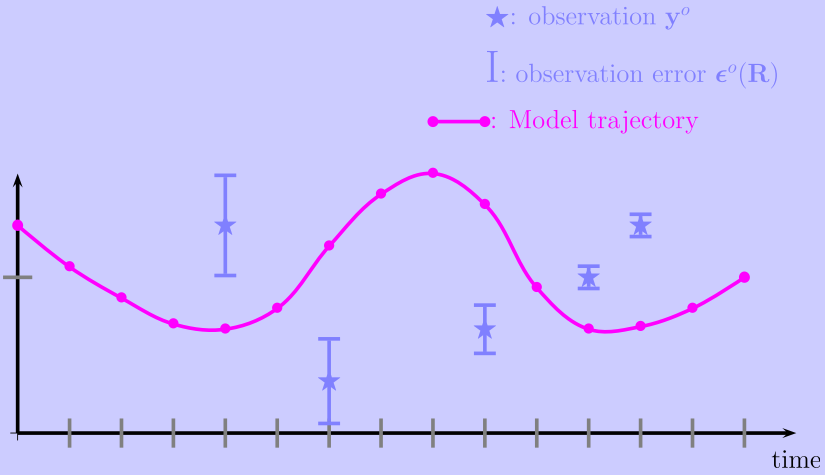
Institut de Physique du Globe de Paris, Paris, France

LabEx UnivEarthS Fall School, October 2015

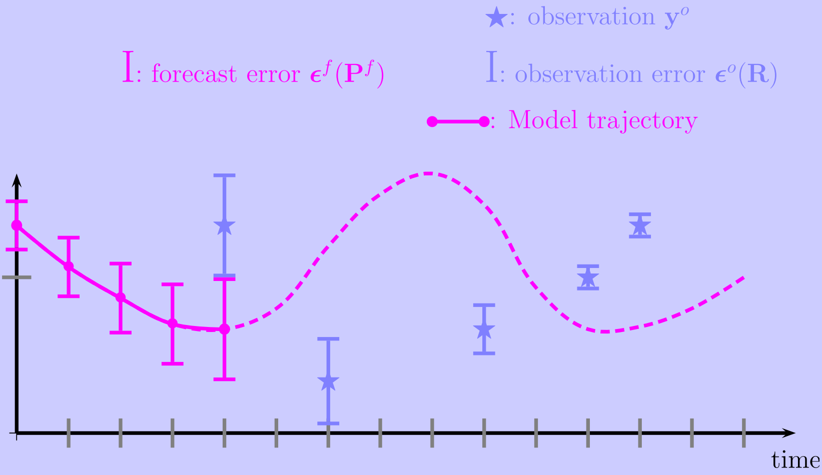
Tentative plan

1. **Introduction**
2. **Stochastic Estimation. The BLUE**
3. **The Kalman filter**
4. **Variational assimilation**
5. **Numerical Weather Prediction**
6. **Other examples**

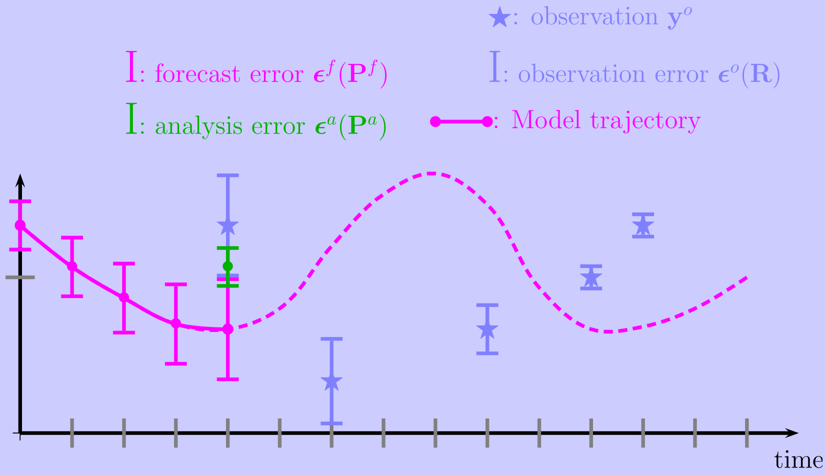
Assimilation in a dynamical framework



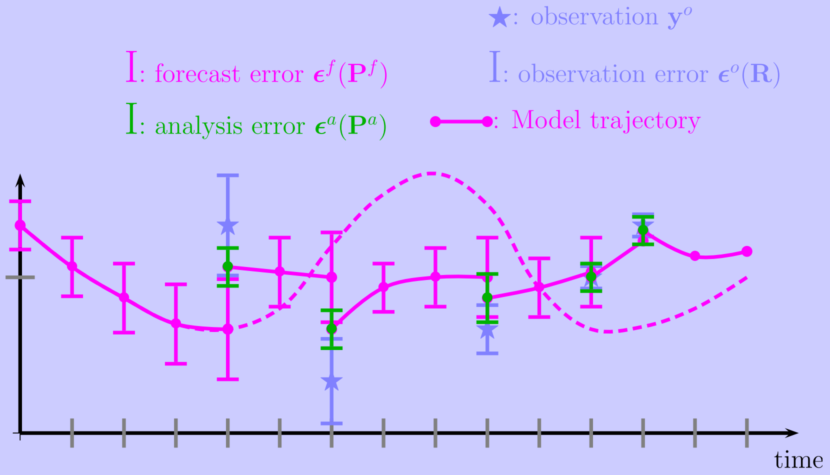
Sequential assimilation



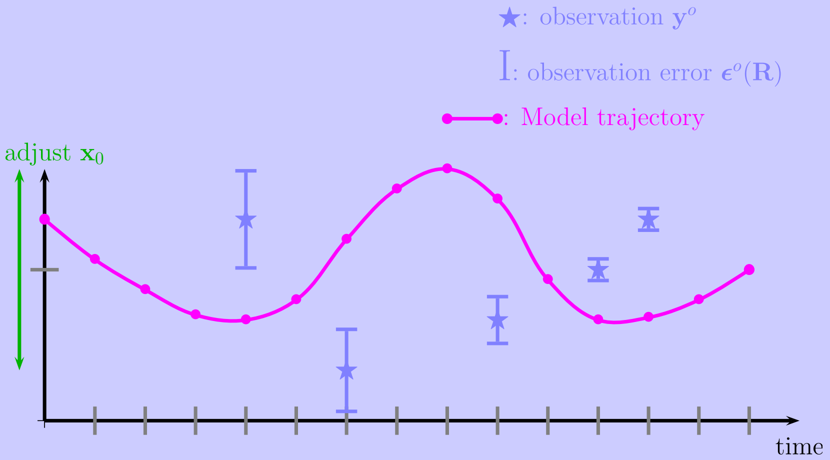
Sequential assimilation



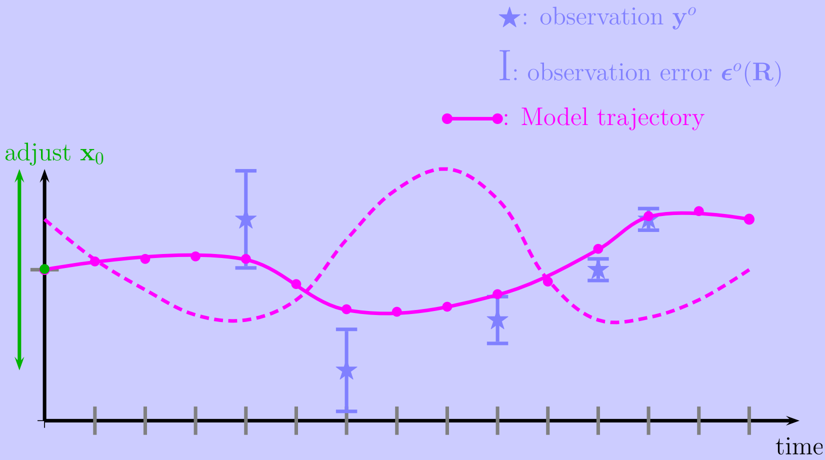
Sequential assimilation



Variational assimilation



Variational assimilation, 4D-Var (adjoint)



Pros & cons

Pros

- KF: error estimate after analysis
- 4D-Var: all observations in one sweep

Cons

- KF: Nonlinearities & problem size. EnKF
- 4D-Var: Implementation of adjoint model (Automatic differentiation)
- 4D-Var: Nonlinearities (tangent linear model) & perfect model assumption

The archetypal example: Numerical Weather Prediction

REVIEW

doi:10.1038/nature14956

The quiet revolution of numerical weather prediction

Peter Bauer[1], Alan Thorpe[1] & Gilbert Brunet[2]

Advances in numerical weather prediction represent a quiet revolution because they have resulted from a steady accumulation of scientific knowledge and technological advances over many years that, with only a few exceptions, have not been associated with the aura of fundamental physics breakthroughs. Nonetheless, the impact of numerical weather prediction is among the greatest of any area of physical science. As a computational problem, global weather prediction is comparable to the simulation of the human brain and of the evolution of the early Universe, and it is performed every day at major operational centres across the world.

Bauer et al., Nature (2015)

Short History

- 1900ies: prognostic set equations that one could solve (in principle)
- 1922: Richardson's forecast (computer = bunch of humans)
- 1950ies: first integation of simplified equations (Princeton, Sweden)
- 1963 (Lorenz, deterministic chaos): finite horizon of predictability
- Since 1970: numerical integration of relevant equations (NWP)

Set of equations

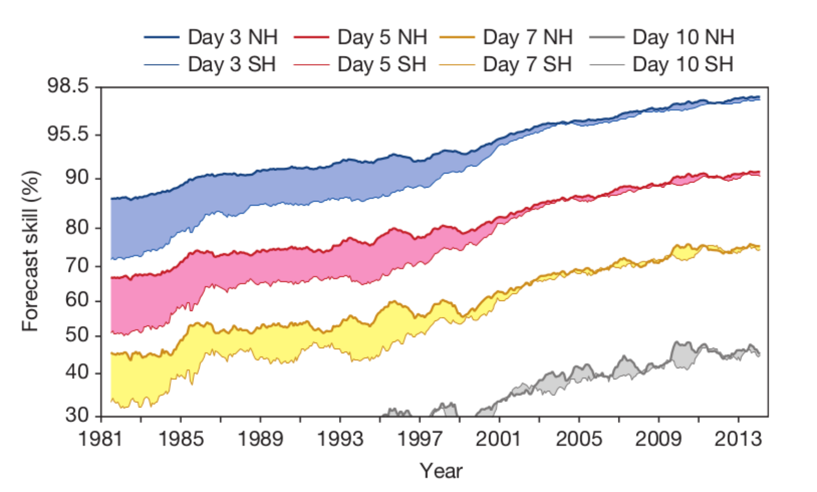
- Conservation of Mass
- Conservation of Momentum (Navier–Stokes, with background rotation)
- Conservation of Energy
- Constitutive relationships (ideal gas)
- + BC and IC (nonlinear mixed initial–boundary value problem)

changes in space and time of wind, pressure, temperature, density

Integration on a computer

- grid size for GCM of $\sim$10 km.
- resolved scales vs unresolved scales
- Subgrid scale modelling (parameterization): source terms for mass, momentum and heat

Measure of success



Why success?

1 day / per decade since the 1970ies.

- **Daily check**
- Better data
- Better representation of unresolved processes

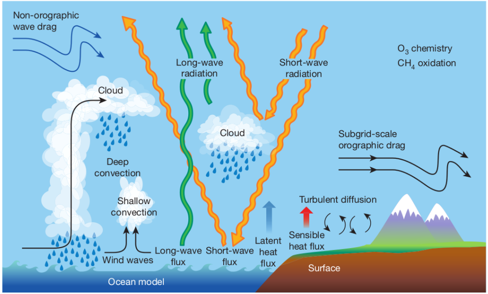
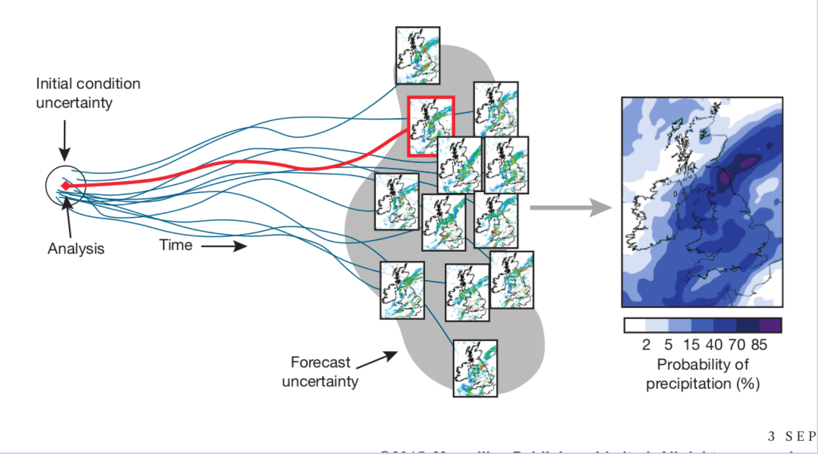


Figure 2 | Physical processes of importance to weather prediction. These are not explicitly resolved in current NWP models but they are represented via parameterizations describing their contributions to the resolved scales in terms of mass, momentum and heat transfers.

- Better assimilation technology (uncertainties & objective analysis)
- ...

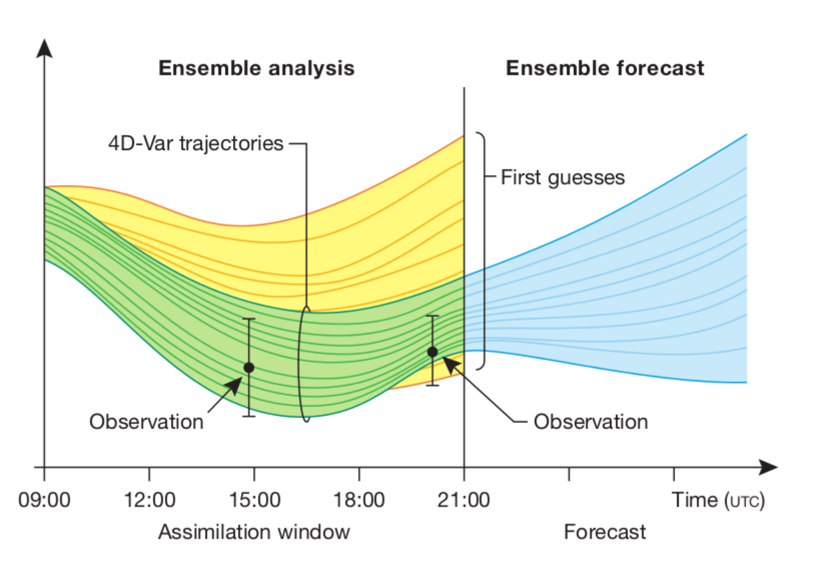
The size of the problem at hand: 50 million observations are ingested every 6 hr in numerical models possessing 10^9 field variables.

Ensemble forecasting



Met Office

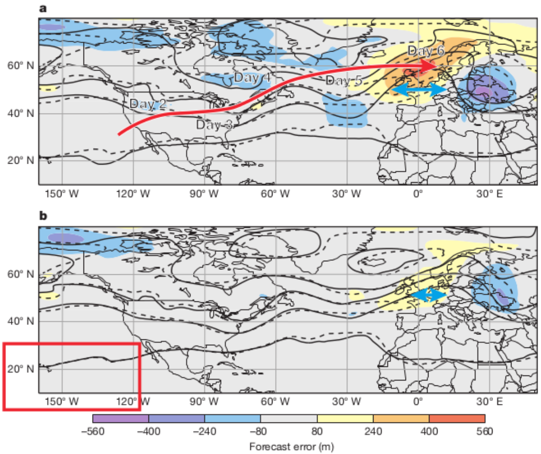
Ensemble 4DVar



Finite horizon of predictability

BOX 1

Sensitivity of forecasts to initial conditions and error propagation

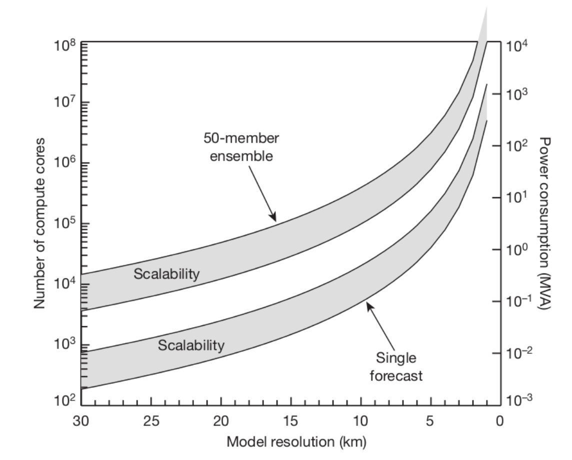


Box 1 Figure | Maps showing the long-range impact of model initialization on the European forecast. Panel **a** shows the day-6 mean forecast error (the height of the 500 hPa pressure level in metres) of the flow at around 5 km height (colour-coded shading), the forecast itself (solid isolines) and the verifying analysis (dashed isolines) valid on 15 February 2014. Over the western US, the jet stream extended far to the south, aligned with a lower level trough. The identifies the tropical East Pacific (boxed in **b**) as a likely location of a possible forecast error source. This area was characterized by very large 24-h forecast errors of upper-level winds because of the paucity of wind observations there. When running an experiment where the area in the box in **b** is relaxed towards the analysis rather than evolving in the forecast, the strong initial growth of forecast errors is reduced and, six days later, the lag of the wave patterns between forecast and

The future

- Even more observations (eg upper level wind with Doppler-radar technology)
- More physics (coupling with ocean, land surface, sea-ice models)
- More chemistry (aerosol, trace gases)
- Global resolution of ~ 1 km. Energy concern.

Technological challenge



Codes need to evolve to meet hardware requirements (exascale: computing / data processing)

The future

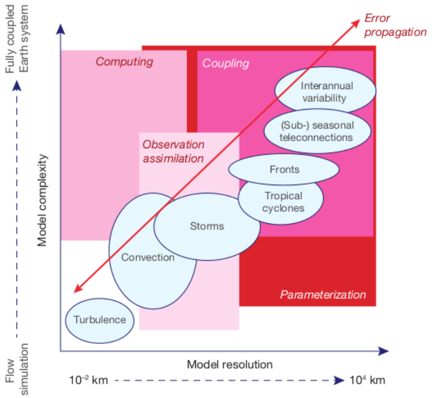


Figure 6 | Key challenge areas for NWP in the future. Advances in forecast skill will come from scientific and technological innovation in computing, the representation of physical processes in parameterizations, coupling of Earth-system components, the use of observations with advanced data assimilation algorithms, and the consistent description of uncertainties through ensemble methods and how they interact across scales. The ellipses show key phenomena relevant for NWP as a function of scales between 10^{-2} and 10^4 km resolved in numerical models and the modelled complexity of processes characterizing the small-scale flow up to the fully coupled Earth system. The boxes represent scale-complexity regions where the most significant challenges for future predictive skill improvement exist. The arrow highlights the importance of error propagation across resolution range and Earth-system components.

Bibliography I