Data assimilation - concepts and methods

Alexandros Altis
Department of Mathematics
J. W. Goethe University Frankfurt

Main topics

- Basic concepts of data assimilation
- A simple illustration
- Mathematical modelling of the assimilation problem
- 3D-Var

Motivation

- Production of an accurate image of the true state of the atmosphere at a given time (analysis)
- Useful for:
  - Self-consistent diagnostic of atmosphere
  - Initial state for a numerical weather forecast

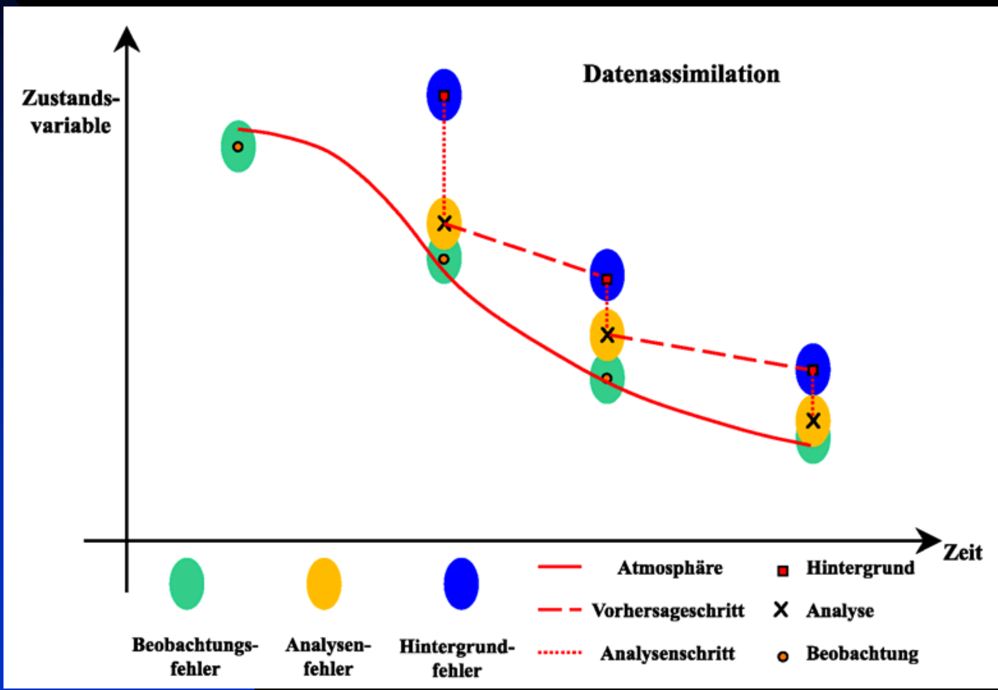
Basic concepts

- ~10^6 variable values (e.g. temperature, air-pressure) needed at gridpoints all over the globe
- Problem: Much less *observations* available
- Available: *Background knowledge* such as previous analysis (assume consistency in time)

Basic concepts

- **Data assimilation**: Combination of both, observations and background knowledge
- Unavoidable: Measurement errors, background errors

Basic concepts



A simple illustration

- Estimation of temperature in this room
  - T_o : observed temperature
  - T_b : background estimate
  - σ_o : standard deviation of measurement error
  - σ_b : standard deviation of background error
- Use *least-squares estimation* to calculate the analysis T_a
- Minimize the functional

$$J(T) = \frac{(T - T_b)^2}{\sigma_b^2} + \frac{(T - T_o)^2}{\sigma_o^2}$$

A simple illustration

- Minimize the functional

$$J(T) = \frac{(T - T_b)^2}{\sigma_b^2} + \frac{(T - T_o)^2}{\sigma_o^2}$$

- Optimal solution T_a:

$$T_a = kT_o + (1 - k)T_b, \quad \text{where } k = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

- For the error of the analysis σ_a we have:

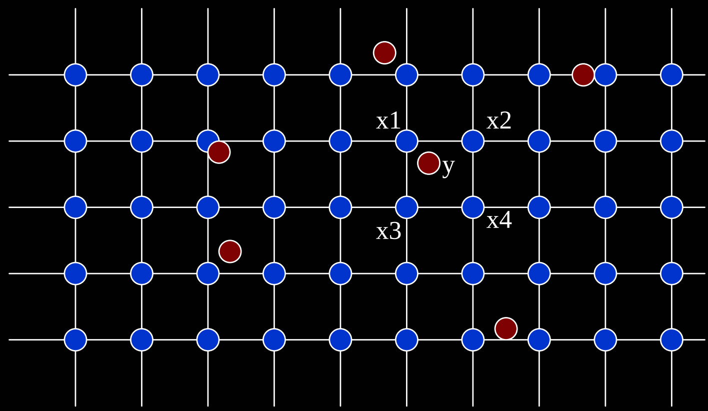
$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}$$

$\Rightarrow$ increase of exactness

Mathematical modelling

H(x1,x2,x3,x4)=y /* observation operator */

● : gridpoints
● : observations



Mathematical modelling

- General assimilation problem:

Minimize the functional

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

with

- x, x_b: vector of values at gridpoints (dimension n)
- y : observation vector (dimension p)
- H(x): forward or observation operator (n→p)
- B : covariance matrix of the background errors (n x n)
- R : covariance matrix of observation errors (p x p)

Mathematical modelling

- Assume certain hypotheses, e.g.:
  - Linearized observation operator:
    $\boxed{H(x) - H(x_b) = \mathrm{H}(x-x_b),}$
    where H is a linear operator (Jacobian matrix at x_b).
  - Uncorrelated errors: observation and background errors are mutually uncorrelated
  - Many more – leading to various methods

Mathematical modelling

- Theorem (least-squares analysis equations):

  The optimal least-squares estimator is given by the following equations:

  $x_a = x_b + K[y-H(x_b)]$, with
  $K = BH^T(HBH^T+R)^{-1}$

  Furthermore, if the background and observation error pdfs are Gaussian, then x_a is also the maximum likelihood estimator of the true state of x.

Outlook

Stochastic approach

$\Leftrightarrow$

Least-squares approach ?

Different numerical properties?

Necessary assumptions?

Three-Dimensional Variational Analysis (3D-Var)

- Instead of calculating the least-squares estimator

  $x_a = x_b + K[y-H(x_b)]$, with

  $K = BH^T(HBH^T+R)^{-1}$

- Minimize J(x) iteratively by performing several evaluations of the cost function

  $J(x)=(x-x_b)^T B^{-1}(x-x_b)+[y-H(x)]^T R^{-1}[y-H(x)]$

  and of its gradient

  $\mathrm{grad}[J(x)] = 2B^{-1}(x-x_b)-2H^T R^{-1}[y-H(x)]$

Three-Dimensional Variational Analysis (3D-Var)

- Stop minimization procedure
  - by limiting artificially the number of iterations

  or
  - by requiring that the norm of the gradient || grad[J(x)] || decreases by a predifined amount during the minimization

Thank you for listening!