Introduction to Data Assimilation, Subgrid-scale Parameterization and Predictability

Christian Franzke

Meteorological Institute

Center for Earth System Research and Sustainability

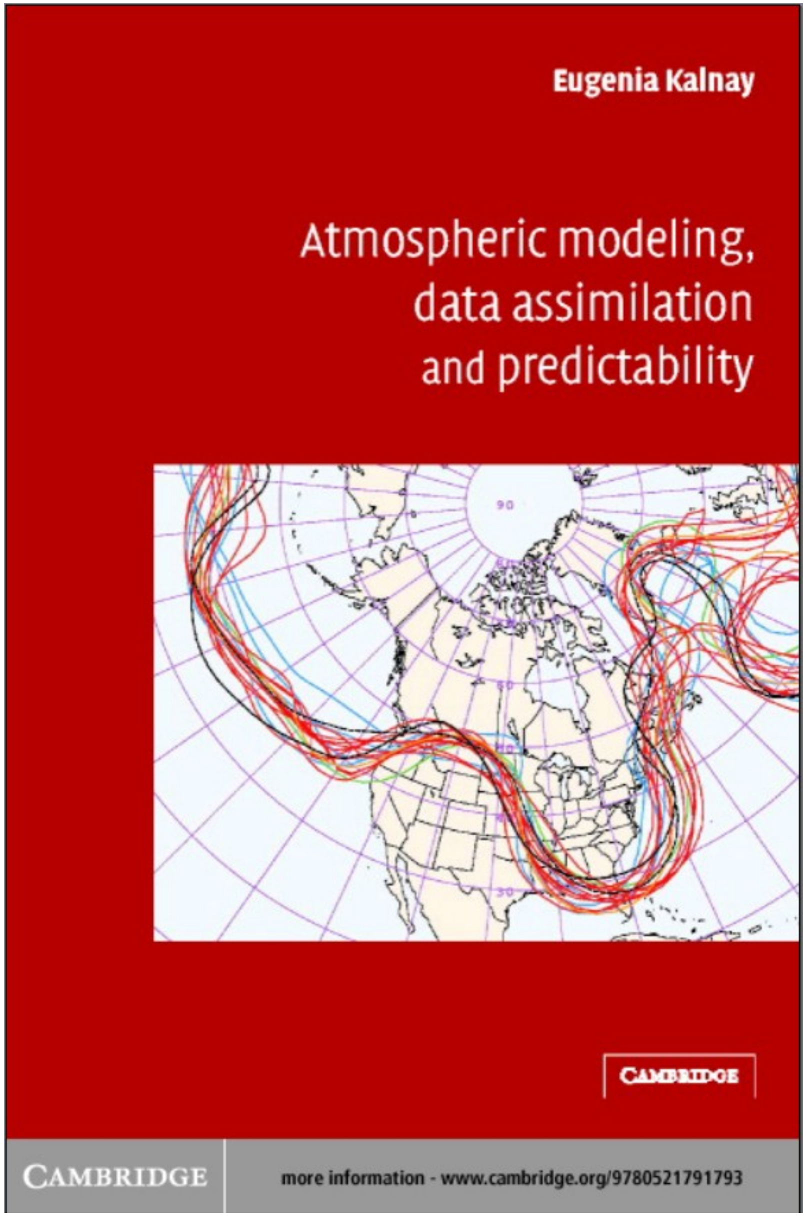
University of Hamburg

Email: christian.franzke@uni-hamburg.de

Outline

- Data Assimilation
  - Empirical Analysis Schemes
  - Least Squares Methods
  - Variational Methods
  - Kalman Filter

Data Assimilation

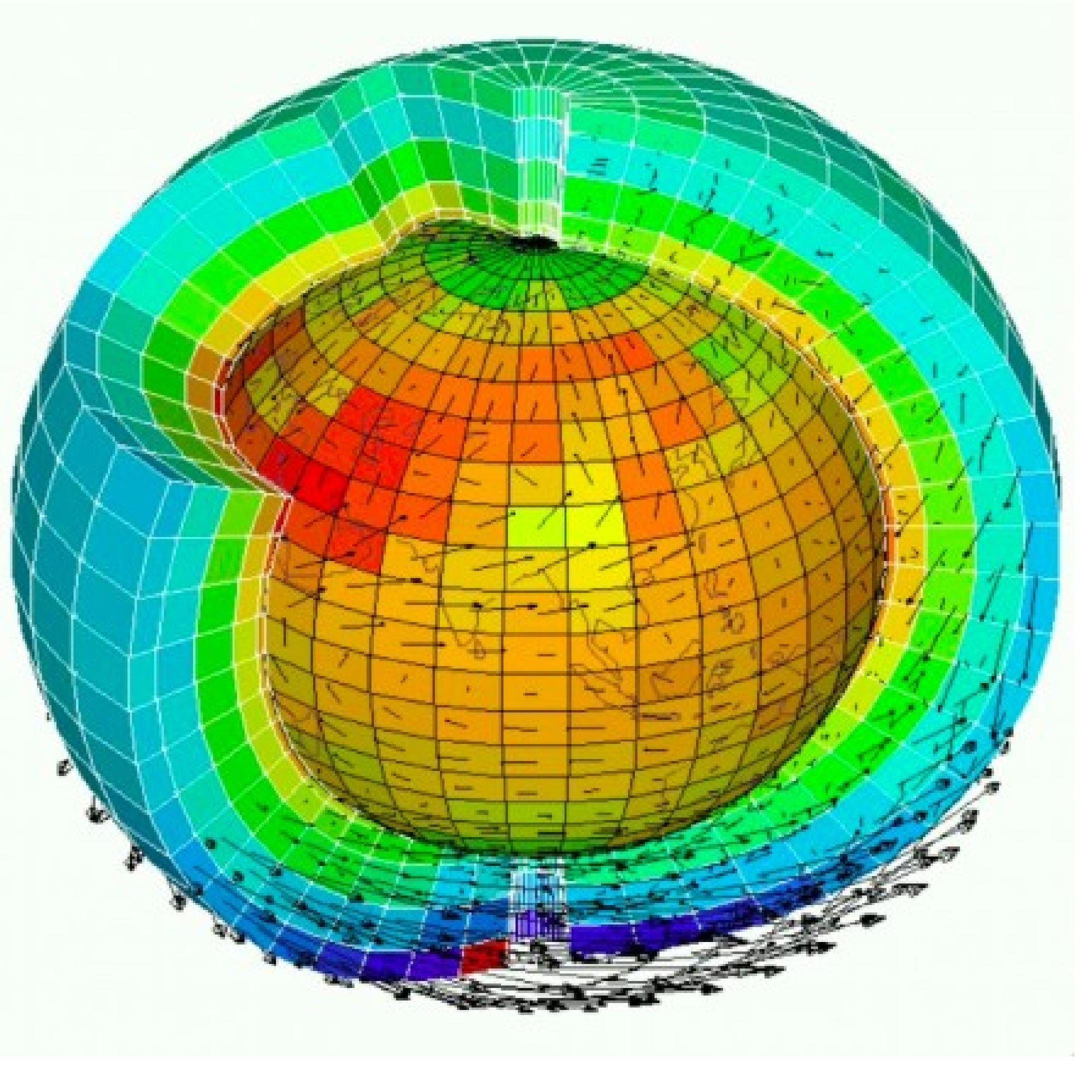


Data Assimilation

Assimilation of meteorological or oceanographical observations can be described as the process through which all the available information is used in order to estimate as accurately as possible the state of the atmospheric or oceanic flow. The available information essentially consists of the *observations* proper, and of the *physical laws* that govern the evolution of the flow. The latter are available in practice under the form of a *numerical model*. The existing assimilation algorithms can be described as either *sequential* or *variational*.

Kalnay 2003

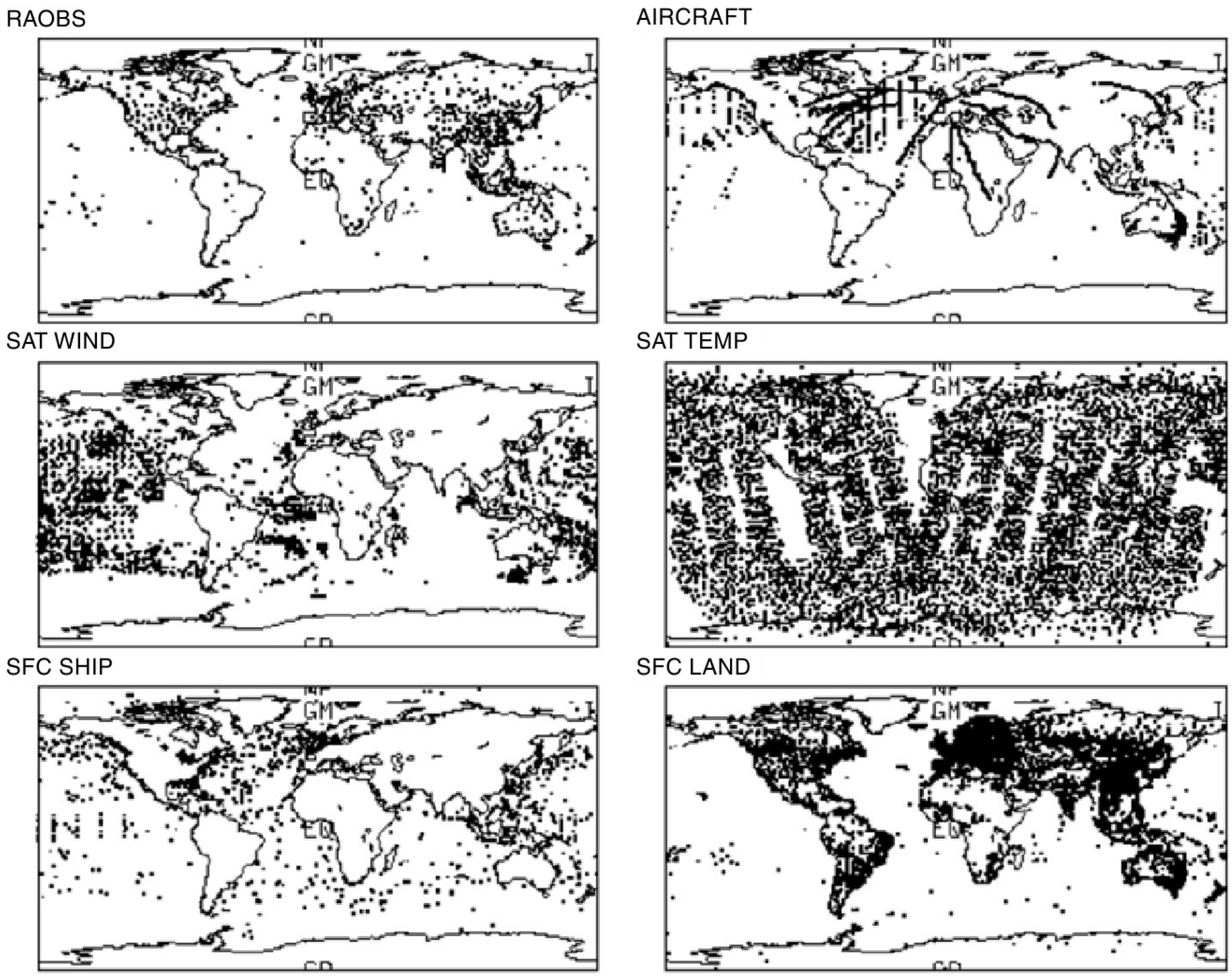
Data Assimilation



Olivier Thual

Data Assimilation

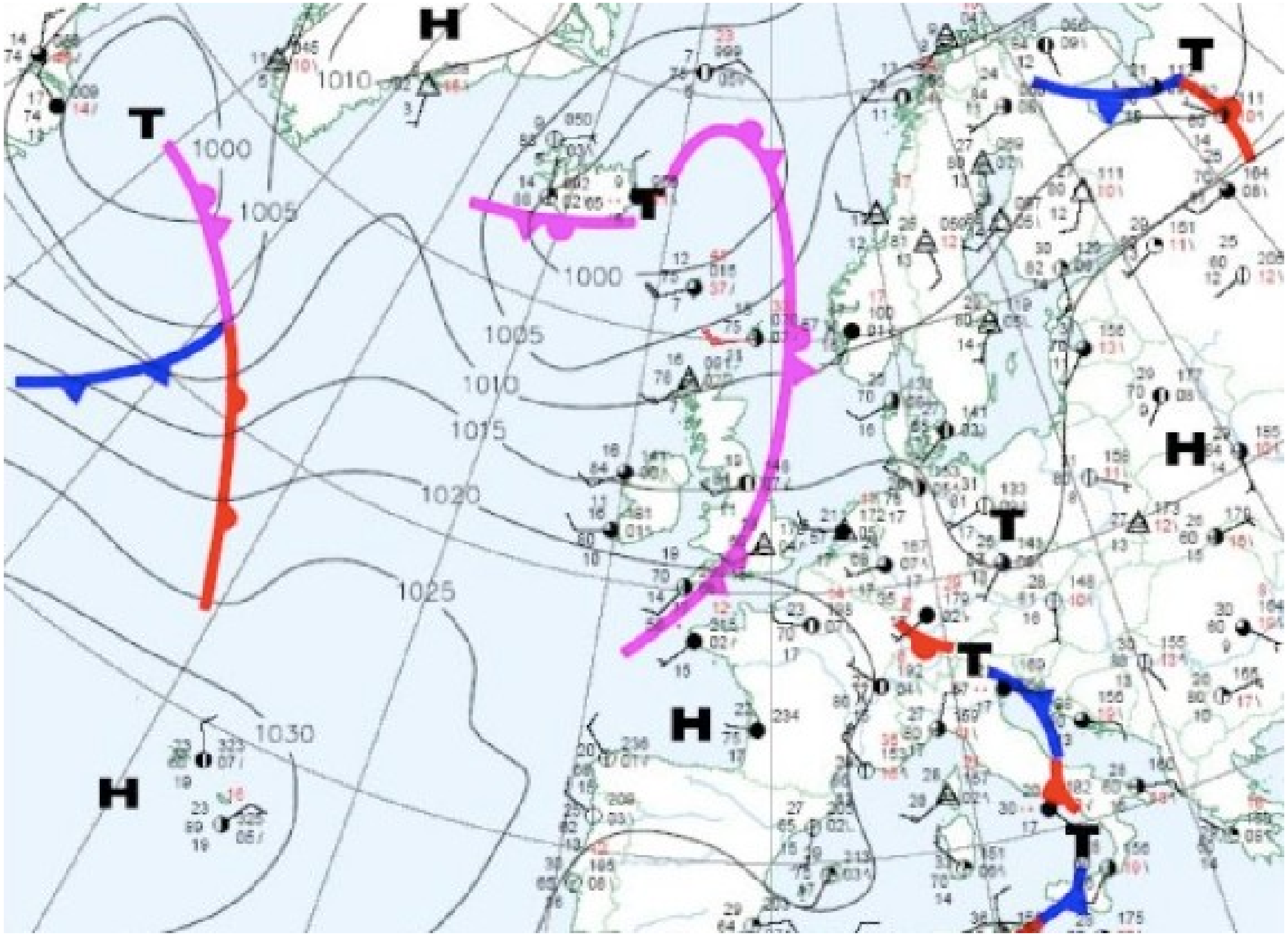
Data Assimilation



Typical distribution of observations in a ±3h window.

Kalnay 2003

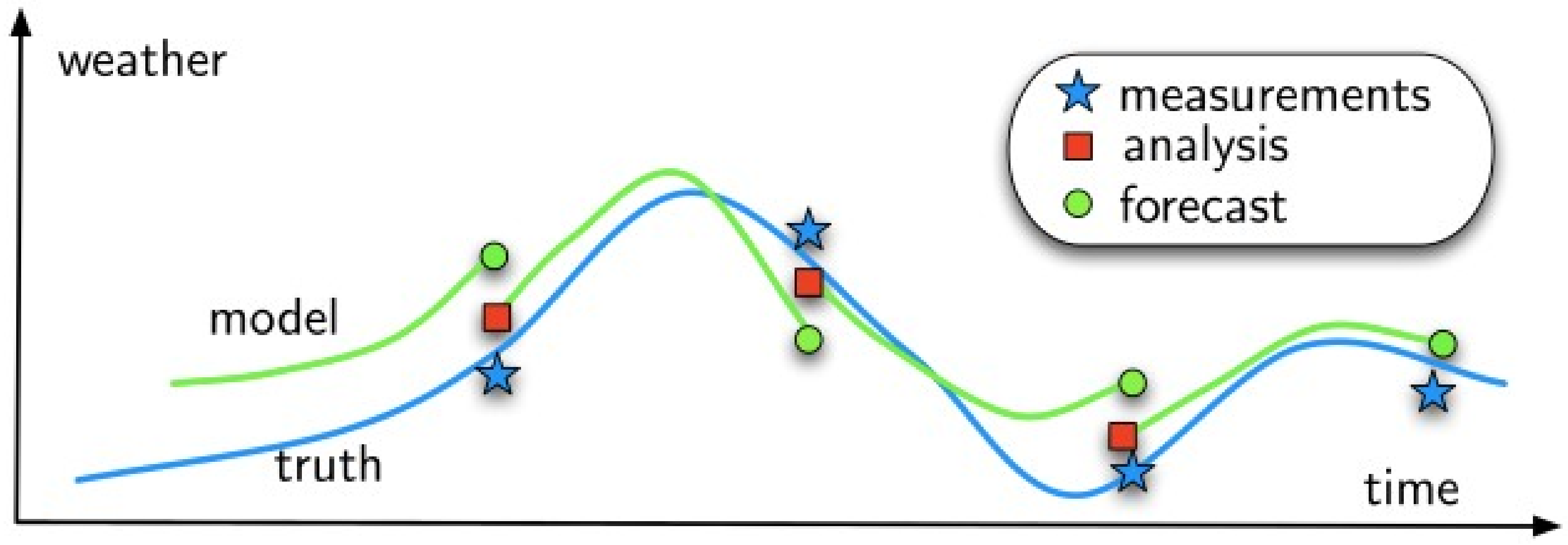
Data Assimilation



Analysis

Olivier Thual

Data Assimilation



Olivier Thual

Data Assimilation

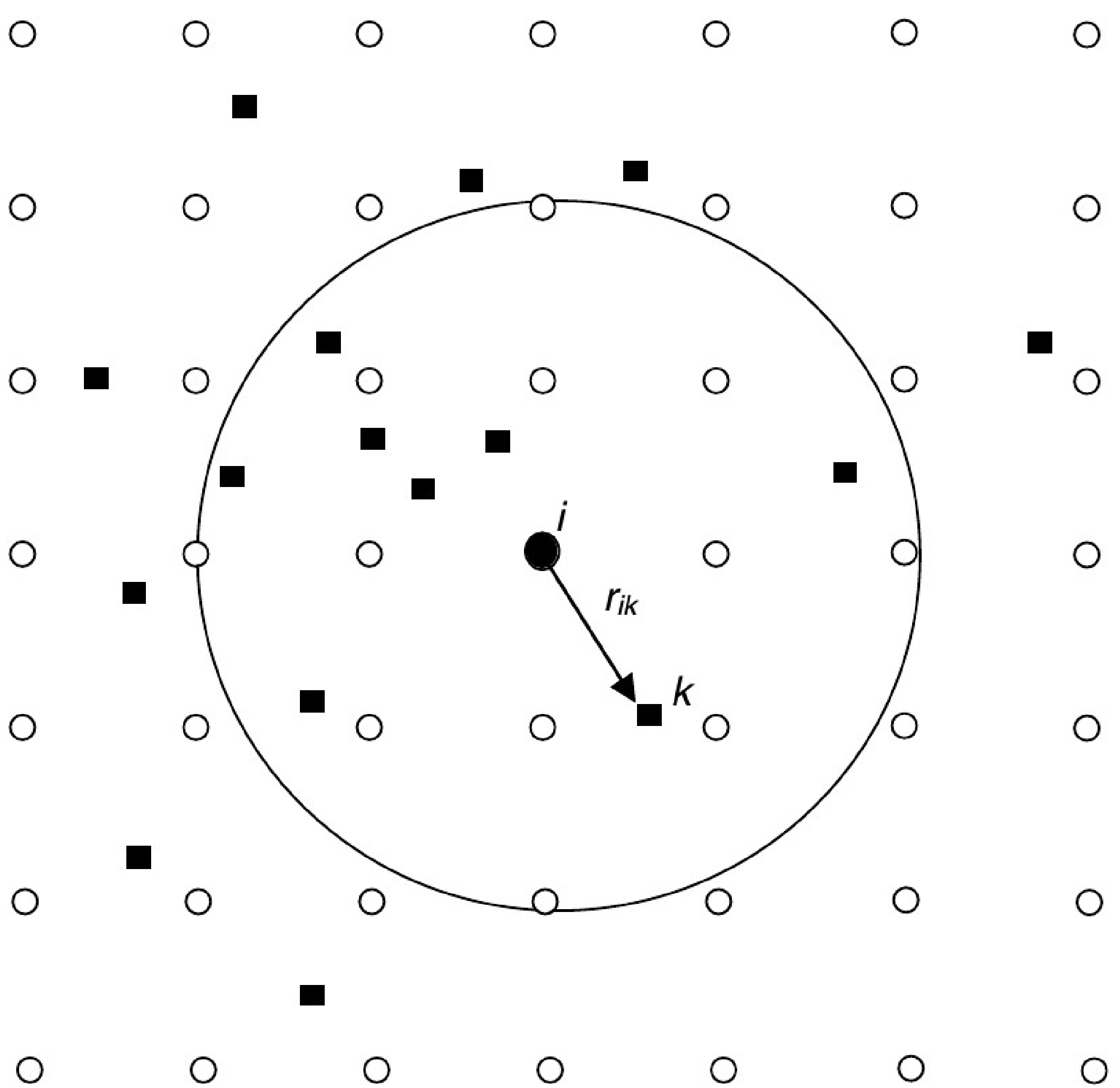


Figure 5.1.1: Schematic of grid points (circles), irregularly distributed observations (squares), and a radius of influence around a grid point i marked with a black circle. In 4DDA, the grid-point analysis is a combination of the forecast at the grid point (first guess) and the observational increments (observation minus first guess) computed at the observational points k. In certain analysis schemes, like SCM, only observations within the radius of influence, indicated by a circle, affect the analysis at the black grid point.

Kalnay 2003

Data Assimilation

Local polynomial interpolation

$$z(x,y)=a_{00}+a_{10}x+a_{01}y+a_{20}x^2+a_{11}xy+a_{02}y^2$$

The six coefficients are determined by minimizing the mean square difference between the polynomial and the observations:

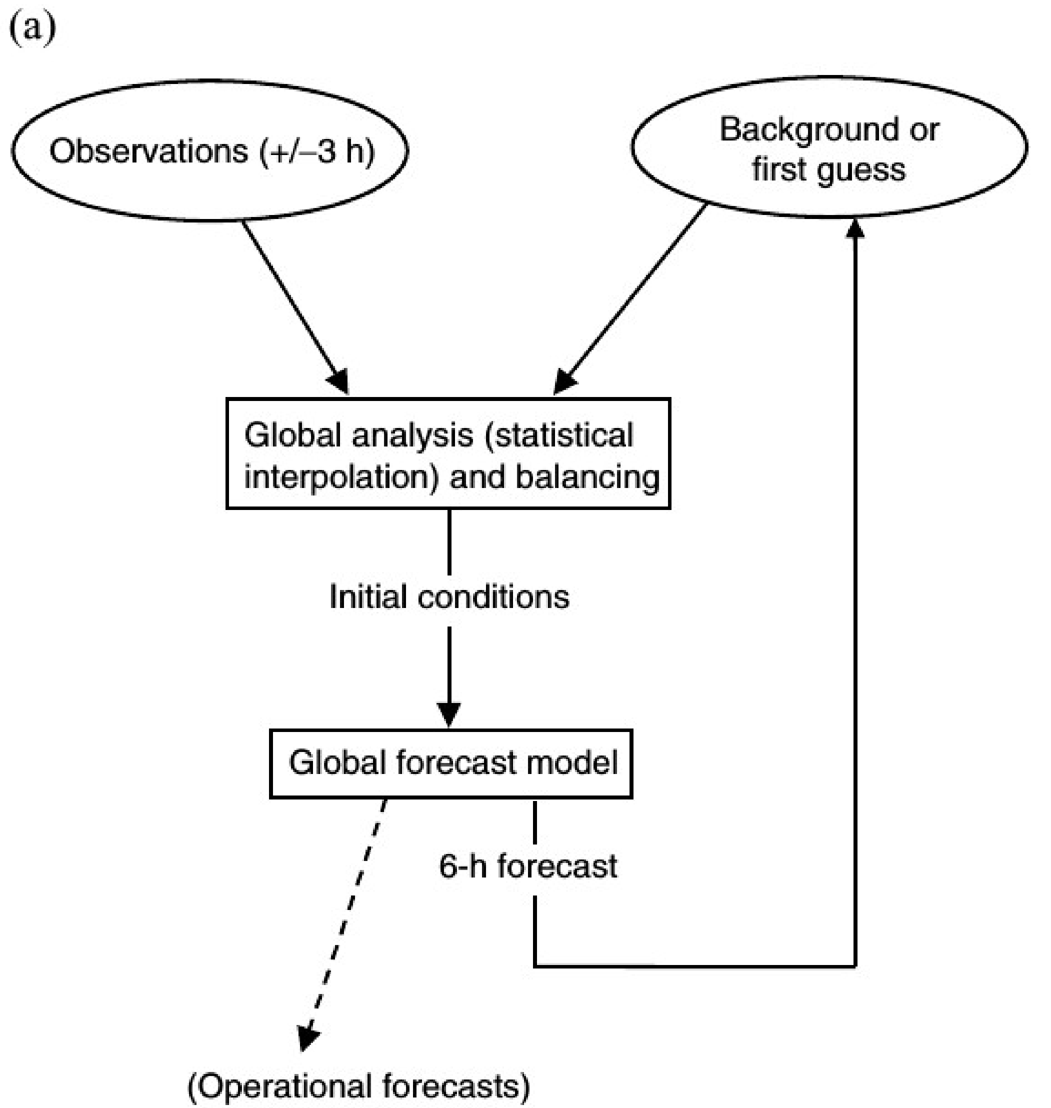
$$min_{a_{ij}}E=min_{a_{ij}}\sum_{k=1}^{K}p_k(z_k^o-z(x,y))^2+\sum_{k=1}^{K}q_k((u_k^o-u_g(x,y))^2+(v_k^o-v_g(x,y))^2)$$

Kalnay 2003

Data Assimilation

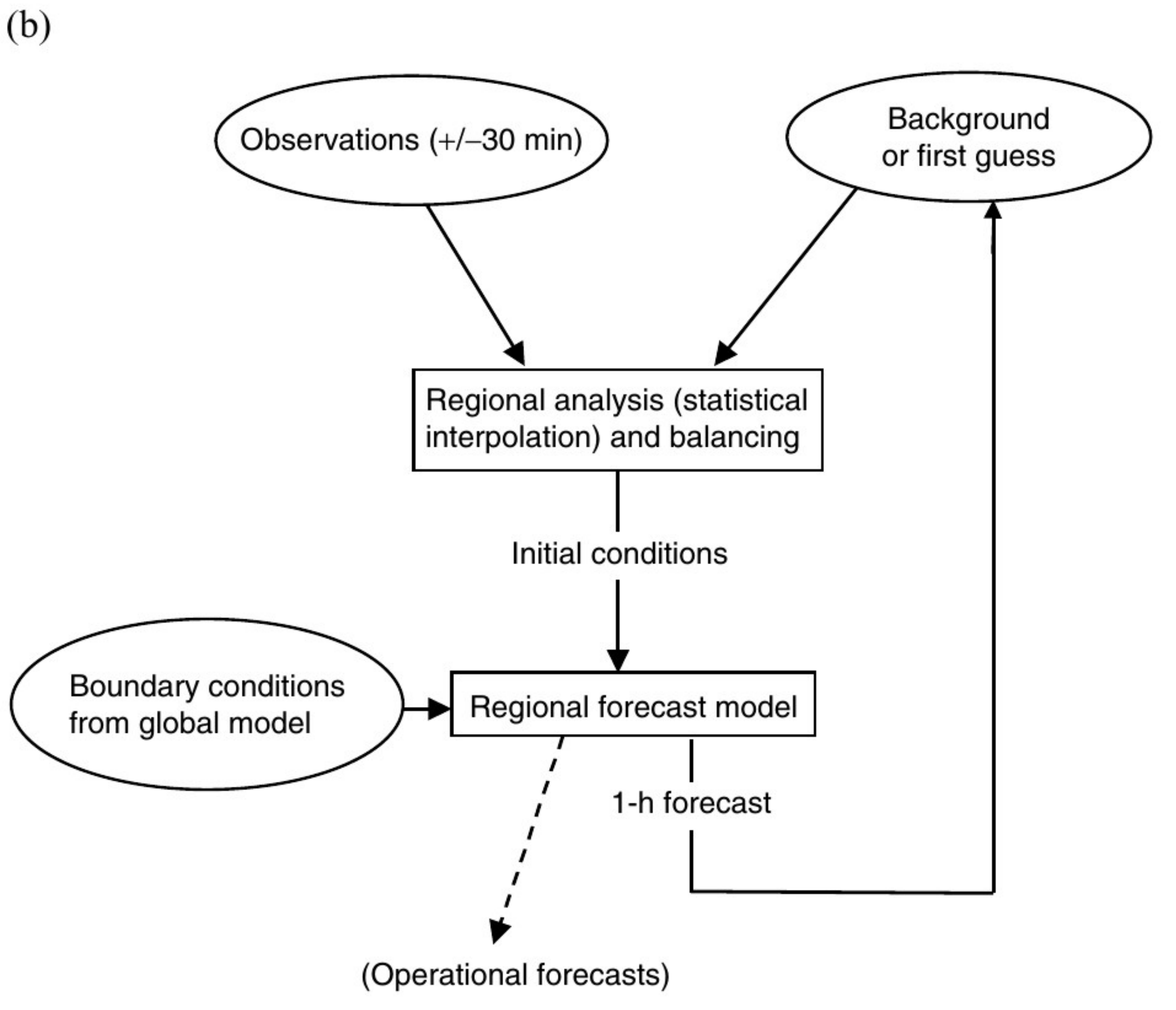
What is the problem with simple interpolation?

Forecast Cycle



Kalnay 2003

Forecast Cycle



Kalnay 2003

Successive correction method

First estimate → background (or first guess) field

$$f_i^0 = f_i^b$$

Successive corrections:

$$f_i^{n+1} = f_i^n + \frac{\sum_{k=1}^{K_i^n} w_{ik}^n \left(f_k^{Obs} - f_k^n\right)}{\sum_{k=1}^{K_i^n} w_{ik}^n + \varepsilon^2}$$

$$w_{ik}^n = \frac{R_n^2 - r_{ik}^2}{R_n^2 + r_{ik}^2} \quad \text{for} \quad r_{ik}^2 \leqslant R_n^2$$

otherwise 0

R can change with iteration:
e.g. R_1=1500km, R_2=1200km, R_3=750km, R_4=300km

Kalnay 2003

Nudging

Newtonian relaxation or nudging:

$$\frac{\partial u}{\partial t} = \vec{u} \cdot \nabla u + fv - \frac{\partial \Phi}{\partial x} + \frac{u_{obs} - u}{\tau_u}$$

and similar for the other equations.

Kalnay 2003

Least Squares Method

Two independent observations:

$$T_1 = T_t + \varepsilon_1$$

$$T_2 = T_t + \varepsilon_2$$

T_1, T_2: Observations

T_t: Truth

ε: observation errors

We assume that measurements are unbiased:

$E(T_1 - T_t) = E(T_2 - T_t) = 0$

$\leftrightarrow$ $E(\varepsilon_1) = E(\varepsilon_2) = 0$

Furthermore: $E(\varepsilon_1^2) = \sigma_1^2$ and $E(\varepsilon_2^2) = \sigma_2^2$

$$E(\varepsilon_1 \varepsilon_2) = 0$$

Kalnay 2003

Least Squares Method

Estimate T_t from linear combination

$$T_a = a_1 T_1 + a_2 T_2$$

T_a: analysis → should be unbiased

→ $E(T_a) = E(T_t)$

which implies $a_1 + a_2 = 1$

Kalnay 2003

Least Squares Method

Best Estimate T_t: Minimizing mean squared error

$\sigma_a^2 = E[(T_a - T_t)^2] = E[(a_1(T_1 - T_t) + a_2(T_2 - T_t))^2]$

subject to constraint $a_1 + a_2 = 1$

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad \text{and} \quad a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

Kalnay 2003

Least Squares Method

How to minimize a function?

Lagrange multipliers

Optimization problem:
maximize f(x,y)
subject to g(x,y)=0

L(x,y,λ)=f(x,y)+λg(x,y)

Solve $\nabla_{x,y,\lambda} L(x,y,\lambda)=0$

Least Squares Method

Relationship between analysis and observations variances

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

If the coefficients are optimal, and the statistics of the errors are exact, then the “precision” of the analysis (defined as the inverse of the variance) is the sum of the precisions of the measurements.

Kalnay 2003

Variational approach

Kalnay 2003

Variational approach

Cost function

$$J(T) = \frac{1}{2}\left[\frac{(T-T_1)^2}{\sigma_1^2} + \frac{(T-T_2)^2}{\sigma_2^2}\right]$$

Minimum of J is obtained for $T=T_a$

J can be found via Maximum Likelihood approach

Has same weights as Least Squares approach
(Show as homework!)

Kalnay 2003

Variational approach

Given two independent observations T_1 and T_2, which are assumed to have normally distributed errors with σ_1 and σ_2, what is the most likely value of T_t?

Kalnay 2003

Variational approach

PDF of T_1 given T_t and σ_1 is given by

$$p_{\sigma_1}(T_1|T_t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(T_1-T_t)^2}{2\sigma_1^2}}$$

Conversely, the likelihood of T_t given T_1 and σ_1 is given by

$$L_{\sigma_1}(T_t|T_1)=p_{\sigma_1}(T_1|T_t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(T_1-T_t)^2}{2\sigma_1^2}}$$

Kalnay 2003

Variational approach

Similarly, likelihood of T given T_2 and σ_2 is given by

$$L_{\sigma_2}(T_t|T_2)=p_{\sigma_2}(T_2|T_t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(T_2-T_t)^2}{2\sigma_2^2}}$$

Kalnay 2003

Variational approach

Most likely value of T given T_1 and T_2 is the one that maximizes the joint PDF (i.e. their product):

$$max_T L_{\sigma_1,\sigma_2}(T|T_1,T_2) = p_{\sigma_1}(T_1|T)\, p_{\sigma_2}(T_2|T) = \frac{1}{2\pi\sigma_1\sigma_2} e^{\frac{-(T_1-T)^2}{2\sigma_1^2} - \frac{(T_2-T)^2}{2\sigma_2^2}}$$

Since logarithm is a monotonic function:

$$max_T \ln L_{\sigma_1,\sigma_2}(T|T_1,T_2) = max_T\left[const. - \frac{(T_1-T)^2}{2\sigma_1^2} - \frac{(T_2-T)^2}{2\sigma_2^2}\right]$$

Corresponds to minimum of cost function J.

Kalnay 2003

Kalman Filter

Kalnay 2003

Kalman Filter

Assume that $T_1=T_b$ is the forecast (or background)
and the other is an observation $T_2=T_o$

then we can write the analysis as:

$$T_a=T_b+W(T_o-T_b)$$

Kalman Filter

$T_a = T_b + W(T_o - T_b)$

where $(T_o - T_b)$ is the *observational innovation*

i.e. the new information brought by the new observation

Kalnay 2003

Kalman Filter

$T_a = T_b + W(T_o - T_b)$

W is the optimal weight given by

$$W = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

$$\sigma_a^2 = \frac{1}{\sigma_b^{-2} + \sigma_o^{-2}} = (1 - W)\sigma_b^2$$

Kalnay 2003

Kalman Filter

$$T_a = T_b + W(T_o - T_b)$$

The analysis is obtained by adding to the first guess (background) the innovation (difference between the observation and first guess) weighted by the optimal weight.

Kalnay 2003

Kalman Filter

$$W = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

The optimal weight is the background error variance multiplied by the inverse of the total error variance.
Note that the larger the background error variance, the larger the correction to the first guess.

Kalman Filter

$$\sigma_a^2 = \frac{1}{\sigma_b^{-2} + \sigma_o^{-2}} = (1 - W)\sigma_b^2$$

The precision of the analysis is the sum of the precisions of the background and the observation.

The error variance of the analysis is the error variance of the background, reduced by a factor equal to one minus the optimal weight.

Kalnay 2003

Kalman Filter

In multidimensional case you have to replace the variances by covariance matrices.

Kalman Filter

If background is a forecast
→ simple sequential analysis cycle
observation is used at the time it appears
and is then discarded

Assume we have completed analysis at time
t_i (12UTC) and we want to proceed to the
next cycle t_{i+1} (18UTC)

Kalman Filter

Analysis cycle has two phases

- *Forecast phase* to update the background T_b and σ_b^2
- *Analysis phase* to update the analysis T_a and σ_a^2

Kalman Filter

In the forecast phase of the analysis cycle the background is first obtained through a forecast

$$T_b(t_{i+1}) = M[T_a(t_i)]$$

M: Forecast model (e.g. ICON-DWD)

Kalnay 2003

Kalman Filter

We also need to estimate the error variance of the background. We compute this using the forecast model.

If we apply $T_b(t_{i+1}) = M[T_a(t_i)]$

to update T_t there would be an error

$T_t(t_{i+1}) = M[T_t(t_i)] - \epsilon_M$

Assumed to be unbiased with error variance Q

Kalnay 2003

Kalman Filter

$$\varepsilon_{b,i+1}=(T_b-T_t)_{i+1}=M(T_a)_i-M(T_t)_i+\varepsilon_M=\mathbf{M}\varepsilon_{a,i}+\varepsilon_M$$

where **M** is the linearized or tangent linear model operator

Forecast of the background error covariance is

$$\sigma_{b,i+1}{}^2=E(\varepsilon_{b,i+1}{}^2)=\mathbf{M}^2\sigma_{a,i}{}^2+Q^2$$

Kalnay 2003