ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Introduction to Data Assimilation

Reima Eresmaa

Finnish Meteorological Institute

Outline

1) The purpose of data assimilation

2) The inputs for data assimilation

3) Analysis methods

Theoretical basis for data assimilation

Statistical approach

4) Data assimilation algorithms

Static assimilation schemes

Four-dimensional variational assimilation

The purpose of data assimilation

Essentially, data assimilation is the procedure for importing the information content of observations into the numerical modelling system

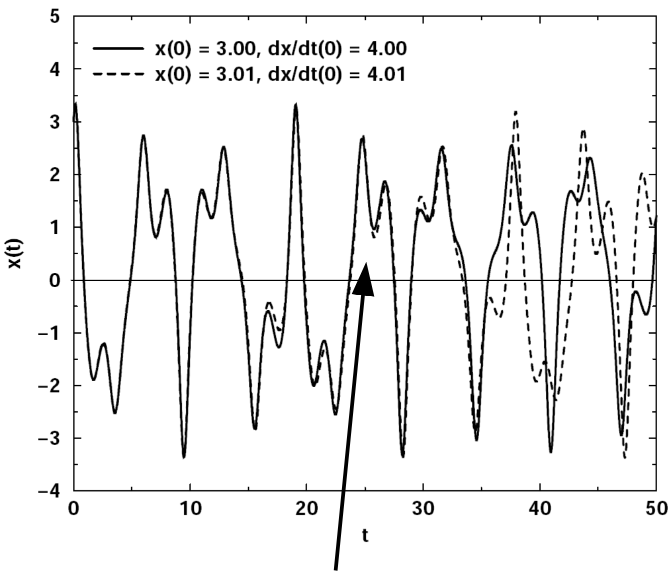
The output of data assimilation

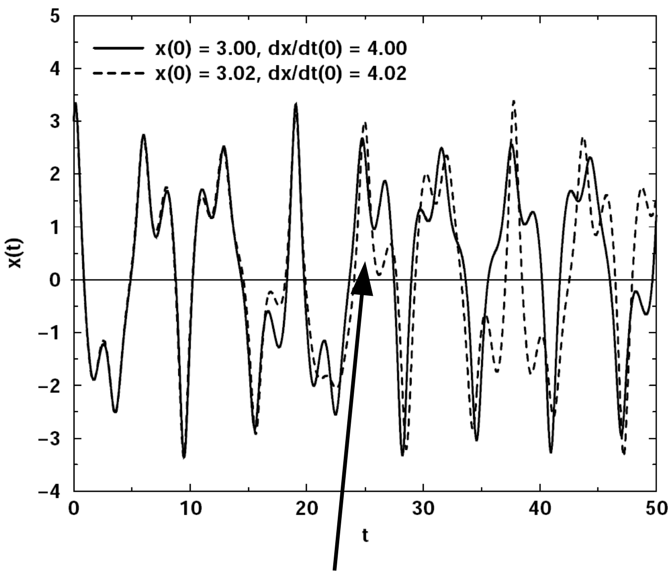
- is *a maximum likelihood estimate* of atmospheric state
- is called *analysis*
- serves as *the initial condition* for the deterministic forecast

The purpose of data assimilation

A future state of *a chaotic system* cannot be successfully predicted without an accurate initial state information

Example: *Duffing's equation* $\ddot{x} + 0.05\dot{x} + x^3 = 7.5\cos t$



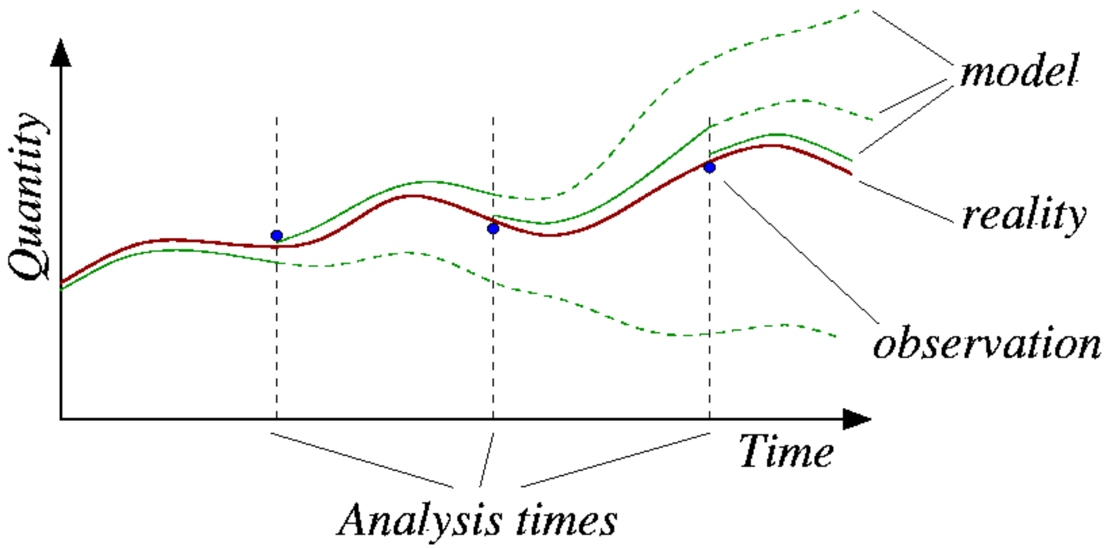


(from the ECMWF training material / Elias Holm)

Data assimilation inputs: observations

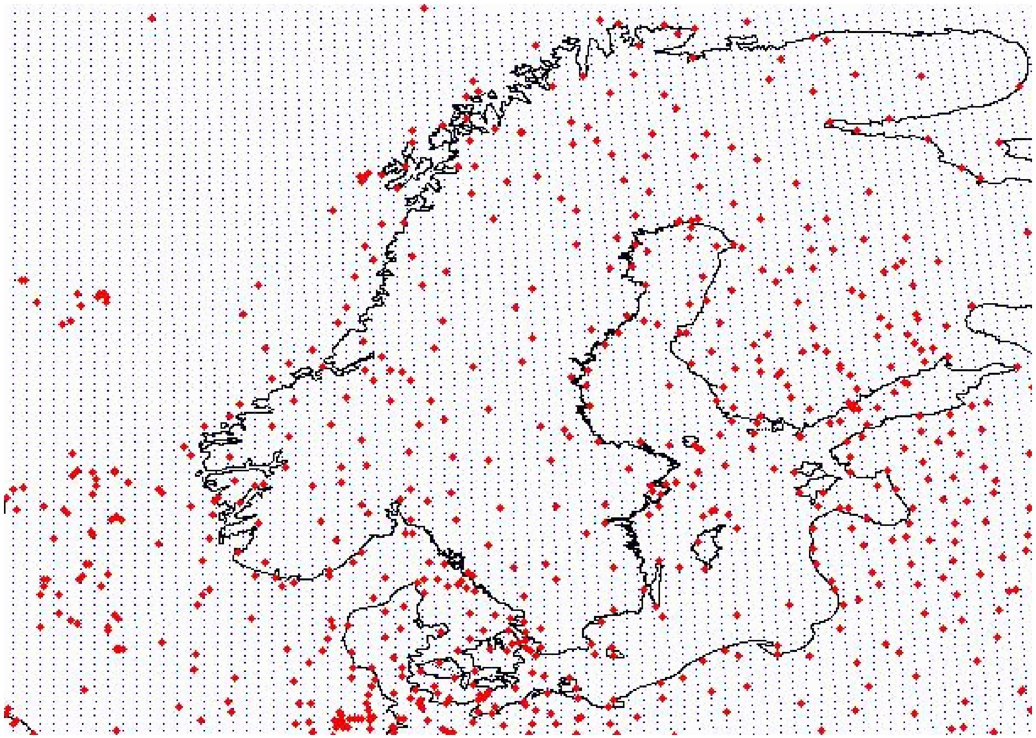
The new information to the forecast model is imported in the form of *observations*

If there were no observational input, the forecast model would drift away from the actual atmospheric state



Data assimilation inputs: background field

Problem: there are fewer observations than parameters to be analysed



Data assimilation inputs: background field

Problem: there are fewer observations than parameters to be analysed

Solution: *background field*

- contains first guess for all analysis variables
- makes the analysis problem over-determined
- is usually a short-range forecast from the previous analysis cycle

Data assimilation inputs: error covariances

Two sources of information are used in the analysis:

- observations ($\boldsymbol{y}$)
- background field ($\boldsymbol{x}^b$)

In order to maintain statistical optimality, *the error statistics* need to be specified for both of these:

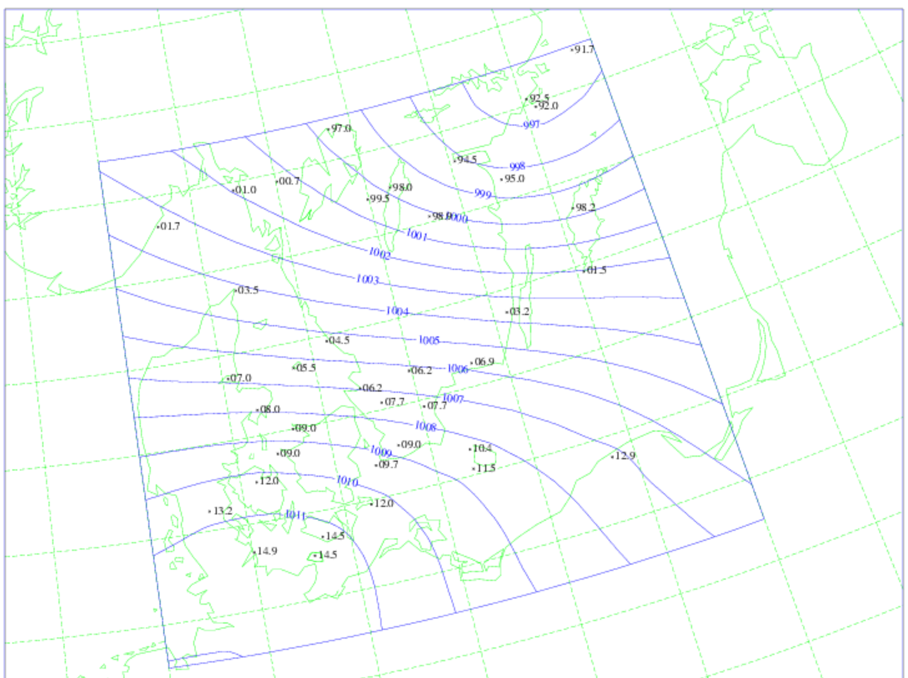
- *the observation error covariance matrix* ($\mathbf{R}$)
- *the background error covariance matrix* ($\mathbf{B}$)

The importance of realistic error statistics

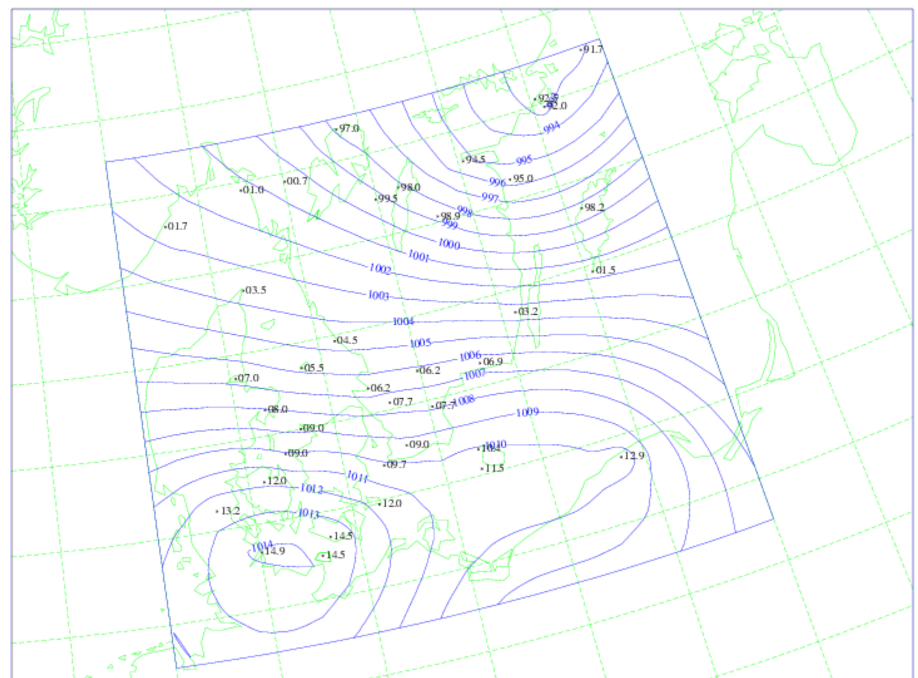
Surface pressure analysed using the ANALAB program

left: ob error sdev twice the bg error sdev

right: bg error sdev twice the ob error sdev



Pressure analysis Tuesday 3 mar 1992 03 utc



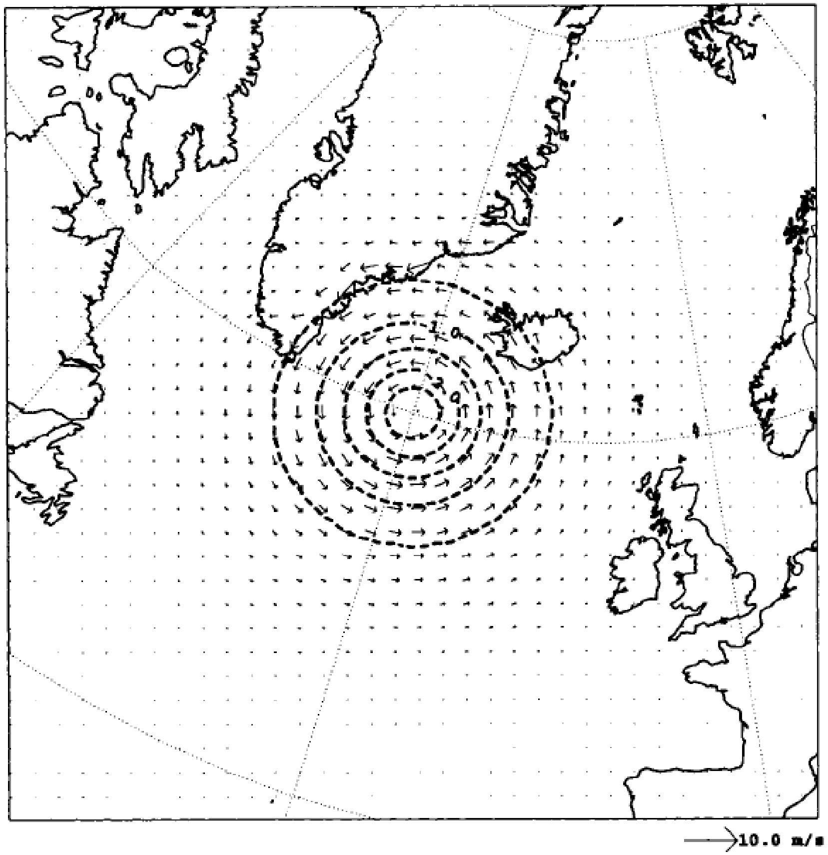
Pressure analysis Tuesday 3 mar 1992 03 utc

Atmospheric balance conditions

A single temperature observation modifies not only model temperature but also model wind field

Geostrophic relation is built in the background error covariance matrix ($\boldsymbol{B}$)

- *Analytical* balance
- *Statistical* balance



Gustafsson et al. (2001), *Tellus* **53A**, 425-446

Data assimilation inputs: observation operator

There is a need for an algorithm *transforming the model variables into observation space*

- the observation locations do not coincide with the model grid points
- the observed quantities may differ from the modelled ones

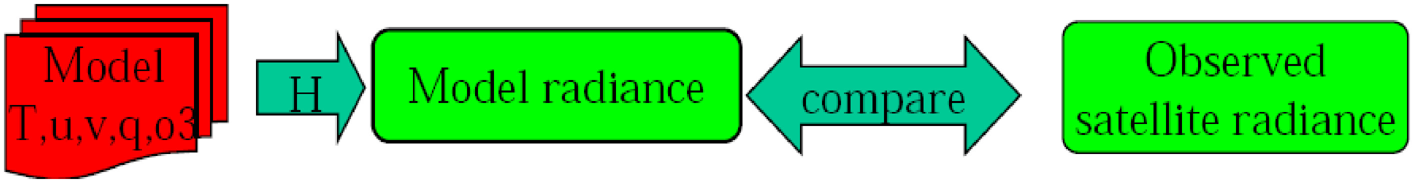
This algorithm is called observation operator H

Observation modelling

Observation modelling enables utilization of a wide range of observation types in data assimilation

Each assimilated observation type requires *a specific observation operator*

Observation operator can be non-linear and it can make use of several model variables

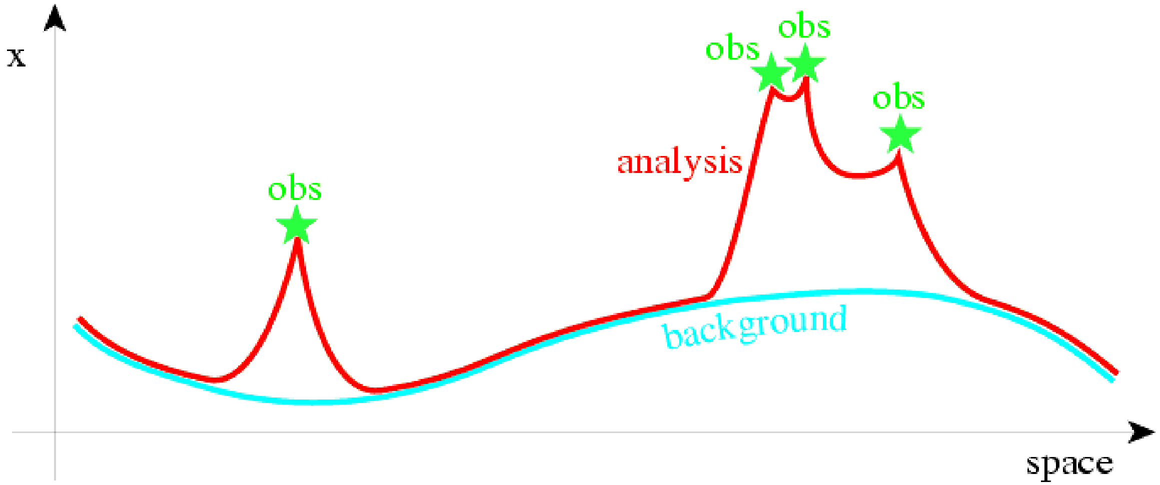


(from the ECMWF training material / Erik Andersson)

Analysis methods: Cressman analysis

Cressman analysis is a simple analysis method *treating the available observations as the truth*.

The observations affect the analysis within an arbitrary influence distance.



(from the ECMWF lecture notes / Bouttier & Courtier)

Statistical approach

How to deal with the observation and background errors?

- the errors itself are not known
- their *statistical properties* are known

Common assumptions:

- *normally distributed errors*
  - zero expectation; no biases
  - standard deviations are known
- *background error is uncorrelated with observation error*

Least-squares estimation

Statistical estimation theory provides the basis for novel data assimilation methods.

The least-squares analysis is obtained as

$$\mathbf{x} = \mathbf{x}^{\mathbf{b}} + \mathbf{K}(\mathbf{y} - H\mathbf{x}^{\mathbf{b}})$$

with *the gain (weighting) matrix* as

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$$

However, the huge dimension of the analysis problem necessitates some simplifications in practical applications

Optimal Interpolation (OI)

The fundamental hypothesis:

For each model variable, only a few observations are needed for determining the analysis increment

- The dimension of the gain matrix is greatly reduced

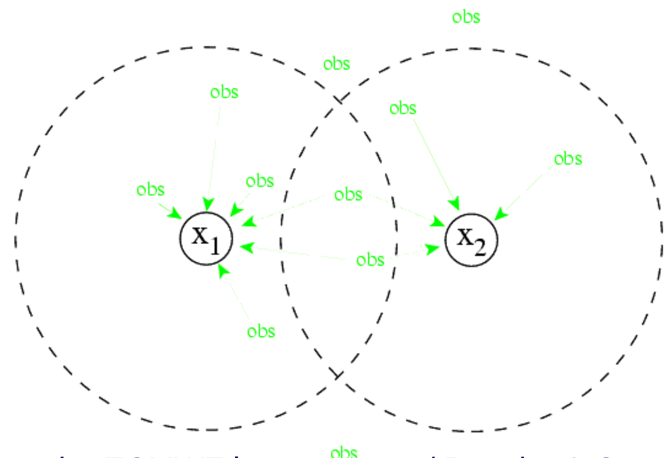
This raises a new question: *How to decide which observations one should use for each analysed variable?*

Observation selection strategies (OI scheme)

<u>*Pointwise selection:*</u>

Each analysis point is sensitive to observations within small distance

Different sets of observations are used for two neighbouring points

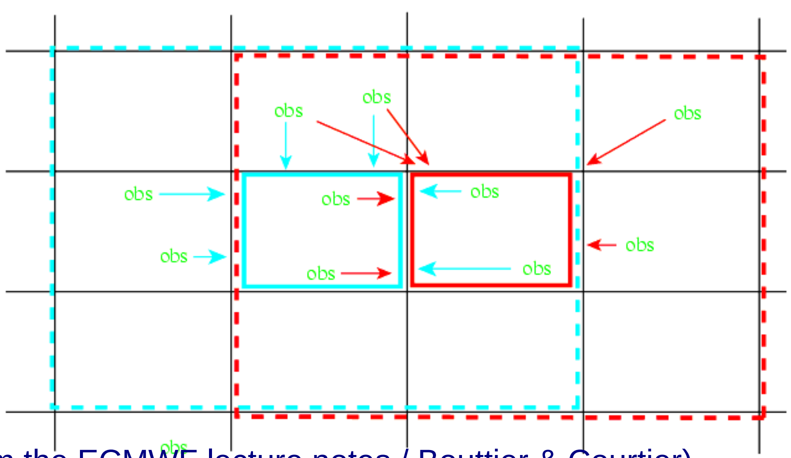


(from the ECMWF lecture notes / Bouttier & Courtier)

<u>*Box selection:*</u>

All points in an analysis box use all observations in a larger selection box

Two neighbouring boxes use mostly the same observations



(from the ECMWF lecture notes / Bouttier & Courtier)

3D Variational assimilation (3D-Var)

The analysis is found as the model state $\boldsymbol{x}$ corresponding to the minimum of the (scalar) cost function

$$J(\mathbf{x}) = \underbrace{\frac{1}{2}(\mathbf{x}-\mathbf{x^b})^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x^b})}_{J_b} + \underbrace{\frac{1}{2}(H\mathbf{x}-\mathbf{y})^T\mathbf{R}^{-1}(H\mathbf{x}-\mathbf{y})}_{J_o}$$

where

J_b measures the departure from the background field $\boldsymbol{x}^{\mathrm{b}}$

J_o measures the departure from the observations $\boldsymbol{y}$

Incremental formulation of 3D-Var

Let us define the analysis increment as $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^\mathbf{b}$

Making use of *the tangent-linear hypothesis*

$$H(\mathbf{x}^\mathbf{b} + \delta\mathbf{x}) - H\mathbf{x}^\mathbf{b} \approx \mathbf{H}\delta\mathbf{x}$$

the cost function becomes

$$J(\delta\mathbf{x}) = \underbrace{\frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x}}_{J_b} + \underbrace{\frac{1}{2}(H\mathbf{x}^\mathbf{b} + \mathbf{H}\delta\mathbf{x} - \mathbf{y})^T\mathbf{R}^{-1}(H\mathbf{x}^\mathbf{b} + \mathbf{H}\delta\mathbf{x} - \mathbf{y})}_{J_o}$$

The analysis increment is transformed into observation space using *the tangent-linear* observation operator $\mathbf{H}$.

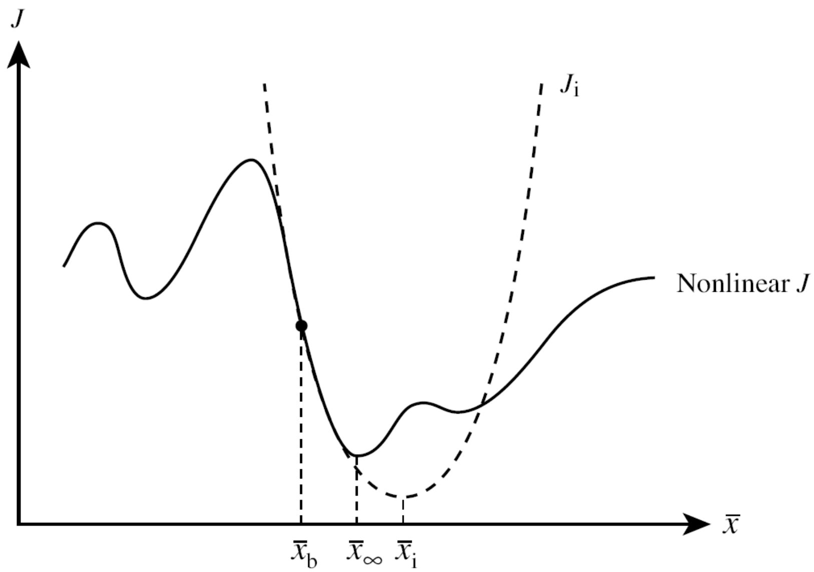
Iterative minimization of 3D-Var cost function

At each iteration step, the the model state is updated using the cost function gradient

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

Note: *the adjoint* of the linearized observation operator $\mathbf{H}^T$

Needed for transformation from the observation space to the model space



(from the ECMWF training material / Elias Holm)

Algebraic tricks

Evaluation of cost function and its gradient involves inversions of covariance matrices with huge dimensions

- **Choice of control variables** – need not to be the same as model variables
- **Fourier transform** from grid point space to spectral space
- **Projection with eigenvectors** – makes the variables vertically uncorrelated
- Observation errors are often assumed uncorrelated

Some properties of OI and 3D-Var schemes

Iterative solution of 3D-Var is equivalent to the solution obtained through the gain matrix in the OI scheme!

However:

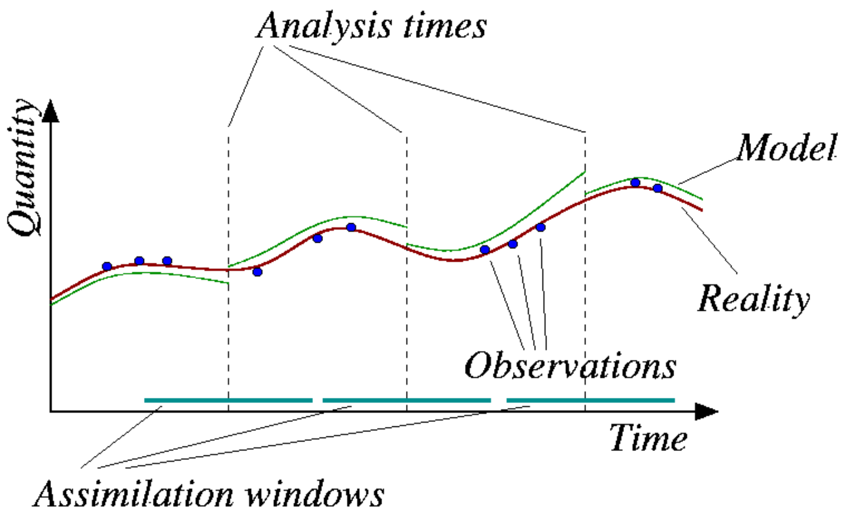
- global treatment of observations in 3D-Var *reduces noise* compared to the selective treatment of OI
- 3D-Var is capable of *exploiting wider range of observation types* than OI

Temporally distributed observations: FGAT

Due to time shift, an observation may be a poor indicator of the analysed atmospheric state.

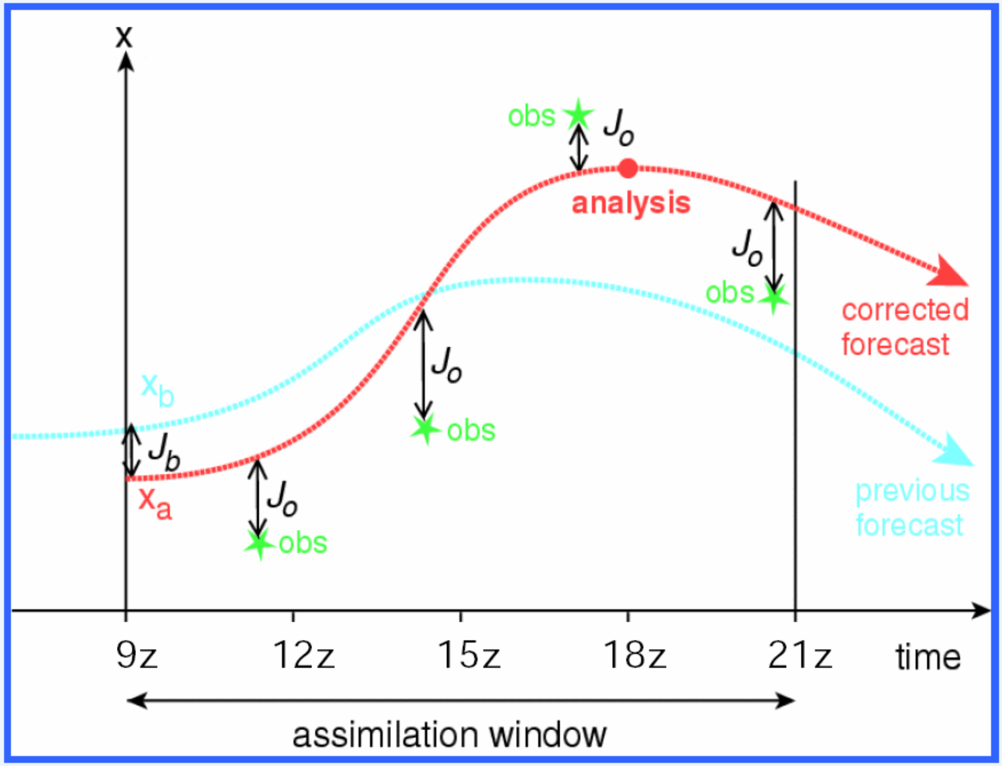
<u>*First Guess at Appropriate Time:*</u>

Short-range forecasts of *different lead times* are used for calculation of $\boldsymbol{y} - H\boldsymbol{x}^{\mathrm{b}}$



4D Variational assimilation (4D-Var)

4D-Var searches *the model trajectory* with best fit to the temporally distributed observations and the background field



(from the ECMWF training material / Lars Isaksen)

4D-Var cost function

The observation contribution to the cost function consists of *a term for each time slot* i *within the analysis time window:*

$$J(\mathbf{x}) = \underbrace{\frac{1}{2}(\mathbf{x}-\mathbf{x^b})^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x^b})}_{J_b} + \underbrace{\frac{1}{2}\sum_{i=0}^{n}(H_i\mathbf{x}_i-\mathbf{y}_i)^T\mathbf{R}_i^{-1}(H_i\mathbf{x}_i-\mathbf{y}_i)}_{J_o}$$

The time sequence of the model states ($\boldsymbol{x}_0$, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, ..., $\boldsymbol{x}_n$) is *required to be a solution to the dynamical forecast model equations*

Some remarks on 4D-Var

The observation operator includes also model forecast operator M:

$$
\begin{aligned}
H\mathbf{x} - \mathbf{y} &\rightarrow HM\mathbf{x} - \mathbf{y} \\
H\mathbf{x}^{\mathbf{b}} + \mathbf{H}\delta\mathbf{x} - \mathbf{y} &\rightarrow HM\mathbf{x}^{\mathbf{b}} + \mathbf{HM}\delta\mathbf{x} - \mathbf{y} \\
\mathbf{H}^T\mathbf{R}^{-1}(H\mathbf{x}^{\mathbf{b}} + \mathbf{H}\delta\mathbf{x} - \mathbf{y}) &\rightarrow \mathbf{M}^T\mathbf{H}^T\mathbf{R}^{-1}(HM\mathbf{x}^{\mathbf{b}} + \mathbf{HM}\delta\mathbf{x} - \mathbf{y})
\end{aligned}
$$

Efficient minimization of the cost function makes use of

- linearized model forecast operator $\mathbf{M}$
- adjoint of the linearized model forecast operator $\mathbf{M}^T$

Some remarks on 4D-Var

It is sufficient to determine the analysis $\boldsymbol{x}$ at the beginning of the assimilation time window ($\boldsymbol{x}_0$);

- the analysis at any later time step follows from $\boldsymbol{x}_0$ and the model forecast operator M

$$
\begin{aligned}
\mathbf{x}_0 &= \mathbf{x} \\
\mathbf{x}_1 &= M_1\mathbf{x}_0 = M_1\mathbf{x} \\
\mathbf{x}_2 &= M_2\mathbf{x}_1 = M_2M_1\mathbf{x} \\
&\vdots \\
\mathbf{x}_n &= M_n\mathbf{x}_{n-1} = M_nM_{n-1}\cdots M_1\mathbf{x}
\end{aligned}
$$

Further reading

Daley, 1991: *Atmospheric data analysis*. Cambridge University Press, 457 s.

ECMWF training material; available at http://www.ecmwf.int/newsevents/training/

Kalnay, 2003: *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press, 341 s.