OPTIMAL ESTIMATION IN METEOROLOGY

ALATNET Seminar
Gourdon
June 2001

Contents

1 Introduction

A general framework : **estimation theory**.

The basis of estimation theory can be found in Gauss (1809), in his attempt to determine the orbits of comets and planets from partial astronomical data:

"*If the astronomical observations and other quantities on which the computations of orbits is based, were absolutely correct, the elements also, whether deduced from three of four observations, would be strictly accurate (so far indeed as the motion is supposed to take place exactly according to the laws of Kepler), and therefore, if other observations were used, they might be confirmed, but not corrected. But since our measurements and observations are nothing more than approximations to truth, the same must be true of all calculations resting upon them, and the highest aim of all computation made concerning concrete phenomena must be to approximate, as nearly as practicable to the truth. But this can be accomplished in no other way than by a suitable combination of more observations than the number absolutely requisite for the determination of the unknown quantities. This problem can only be properly undertaken when an approximate knowledge of the orbit has been already obtained, which is afterward to be corrected, so as to satisfy all of the observations in the most accurate manner possible.*" (Theoria Motus Corporum Coelestium, 1809)

Key ideas :
- observations are not perfect,
- the estimate will also be approximate,
- the estimate must be an optimal combination of observations,
- it is better to overdetermine the problem,
- a model can be used to give a preliminary estimate,
- model are never perfect,
- the estimate should fit observations within their errors.

A simplified problem of estimation

Find the best estimation of a variable x knowing two measurements y_1^o et y_2^o.

These measurements can be written :

$$y_1^o = x^t + \epsilon_1^o$$

and

$$y_2^o = x^t + \epsilon_2^o,$$

where x^t stands for the "truth", that we aim to estimate.

ϵ_1^o et ϵ_2^o are the observation errors associated with these data. They are of course unknown, but we can consider that we know the specifications of the instruments that have produced these observations.

In particular, the following hypotheses can be assumed.

1) The instruments are un-biased, which can be expressed, from a statistical point of vue, as

$$E(\epsilon_1^o) = E(\epsilon_2^o) = 0.$$

2) The precisions of the instruments are known and related to the statistical variances of the respective errors :

$$E({\epsilon^{o2}_1}) = \sigma_1^{o2}$$

and

$$E(\epsilon_2^{o2}) = \sigma_2^{o2}.$$

3) One can also assume that, since the observations are given by different instruments, their errors are not correlated, that is

$$E(\epsilon_1^o \epsilon_2^o) = 0.$$

1st approach:

minimisation of the estimation error variance

One seeks an estimate x^a of x^t (the truth) under the form of a linear combination of observations y_1^o et y_2^o:

$$x^a = k_1 y_1^o + k_2 y_2^o.$$

Coefficients k_1 et k_2 are the unknowns of the problem.

One wants x^a to be the best statistical estimate of the truth x^t: in particular, it has to be un-biased when compared to x^t.

The statistical bias of x^a with respect to x^t writes:

$$\begin{aligned} E(x^a - x^t) &= E(k_1 y_1^o + k_2 y_2^o - x^t) \\ &= E(k_1(x^t + \epsilon_1^o) + k_2(x^t + \epsilon_2^o) - x^t) \\ &= k_1 E(\epsilon_1^o) + k_2 E(\epsilon_2^o) + (k_1 + k_2 - 1)E(x^t) \\ &= (k_1 + k_2 - 1)E(x^t) \end{aligned}$$

since

$$E(\epsilon_1^o) = E(\epsilon_2^o) = 0.$$

$\Rightarrow$ The estimate x^a is un-biased, if the the sum of the coefficients is equal to 1:

$$k_1 + k_2 = 1.$$

Among all the un-biased estimates, we seek the one that minimizes the variance of the estimation error, that is:

$$\sigma_a^2 = E(\ (x^a - x^t)^2\ ).$$

Since $k_1 + k_2 = 1$, it follows that $x^a = k_1 y_1^o + (1 - k_1) y_2^o$ and thus

$$\begin{aligned}
\sigma_a^2(k_1) &= E[\ (k_1(x^t + \epsilon_1^o) + (1 - k_1)(x^t + \epsilon_2^o) - x^t)^2\ ] \\
&= E[\ (k_1\epsilon_1^o + (1 - k_1)\epsilon_2^o)^2\ ] \\
&= k_1^2 E[(\epsilon_1^o)^2] + 2k_1(1 - k_1)E[\epsilon_1^o\epsilon_2^o] + (1 - k_1)^2 E[(\epsilon_2^o)^2] \\
&= k_1^2 {\sigma_1^o}^2 + (1 - k_1)^2 {\sigma_2^o}^2,
\end{aligned}$$

since $E(\epsilon_1^o\epsilon_2^o) = 0$.

The optimal value k_1 is obtained for $J'_{k_1} = 0$, that is

$$2k_1{\sigma_1^o}^2 - 2(1 - k_1){\sigma_2^o}^2 = 0,$$

or

$$2k_1({\sigma_1^o}^2 + {\sigma_2^o}^2) - 2{\sigma_2^o}^2 = 0,$$

which yields $k_1 = \dfrac{{\sigma_2^o}^2}{{\sigma_1^o}^2 + {\sigma_2^o}^2}$ and $k_2 = 1 - k_1 = \dfrac{{\sigma_1^o}^2}{{\sigma_1^o}^2 + {\sigma_2^o}^2}$.

Coefficients k_1 et k_2 can be re-written under a slightly different form :

$$k_1 = \left(\frac{{\sigma_1^o}^2{\sigma_2^o}^2}{{\sigma_1^o}^2 + {\sigma_2^o}^2}\right)\frac{1}{{\sigma_1^o}^2} \qquad k_2 = \left(\frac{{\sigma_1^o}^2{\sigma_2^o}^2}{{\sigma_1^o}^2 + {\sigma_2^o}^2}\right)\frac{1}{{\sigma_2^o}^2}.$$

Quality of the estimation

Substituting for the expressions of k_1 et k_2 in σ^{a2}, we get

$$\sigma^{a2} = \frac{\sigma_1^{o2}\sigma_2^{o2}}{\sigma_1^{o2} + \sigma_2^{o2}}$$

or

$$\frac{1}{\sigma^{a2}} = \frac{1}{\sigma_1^{o2}} + \frac{1}{\sigma_2^{o2}}.$$

Note that x^a can also be re-written

$$\frac{x^a}{\sigma^{a2}} = \frac{y_1^o}{\sigma_1^{o2}} + \frac{y_2^o}{\sigma_2^{o2}}.$$

If we call **precision** the inverse of the error variance, we find that the precision of the estimate is equal to the sum of the precisions of the observations.

This result can be generalized for any number of observations:

$$\frac{1}{\sigma^{a2}} = \sum_{i=1}^{p} \frac{1}{\sigma_i^2}.$$

This means that any new measurement, whatever its precision is, allows to reduce the estimation error variance.

In particular, if we have p observations with same precision, we get that

$$\frac{1}{\sigma^{a2}} = \frac{p}{\sigma_i^2}.$$

On the other hand, for one single observation y^o with error variance σ^{o2}, we have

$$x^a = y^o \text{ and } \sigma^{a2} = \sigma^{o2}.$$

2nd approach: Weighted Least Squares

Without any other hypothesis, we can also seek an estimate x^a of x^t that is close to the observations and that takes into account their respective precisions.

The distance between x^a and the observations can be measured by the following quadratic expression :

$$J(x^a) = \frac{1}{2}\frac{(x^a - y_1^o)^2}{\sigma_1^{o2}} + \frac{1}{2}\frac{(x^a - y_2^o)^2}{\sigma_2^{o2}}.$$

The minimum of J is reached for

$$\frac{\partial J(x^a)}{\partial x^a} = \frac{x^a - y_1^o}{\sigma_1^{o2}} + \frac{x^a - y_2^o}{\sigma_2^{o2}},$$

that is

$$x^a = \frac{\sigma_2^{o2}}{\sigma_1^{o2} + \sigma_2^{o2}} y_1^o + \frac{\sigma_1^{o2}}{\sigma_1^{o2} + \sigma_2^{o2}} y_2^o = (\frac{\sigma_1^{o2}\sigma_2^{o2}}{\sigma_1^{o2} + \sigma_2^{o2}})(\frac{y_1^o}{\sigma_1^{o2}} + \frac{y_2^o}{\sigma_2^{o2}}).$$

We recover the same expression as the one obtained by minimizing the variance of the estimation error.

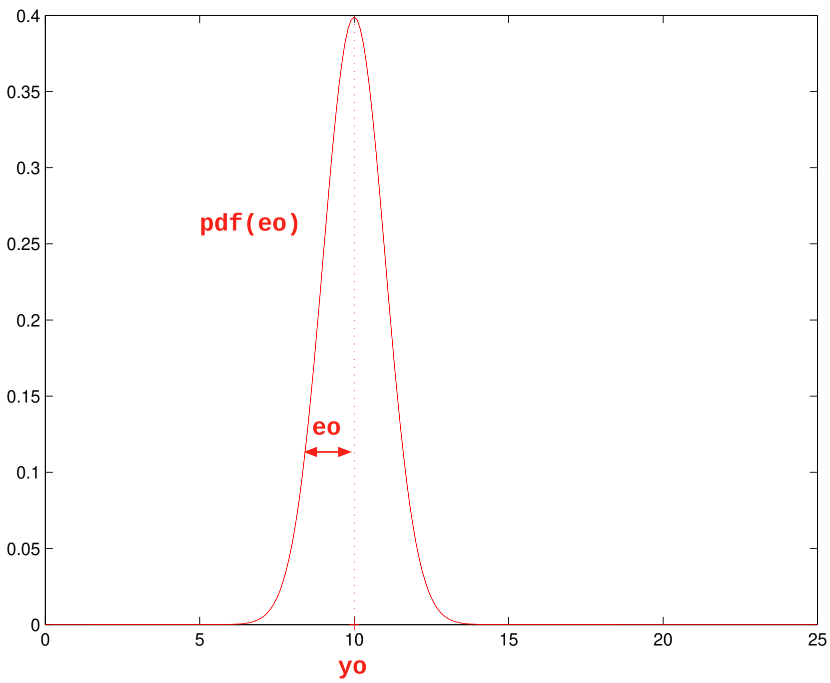
3rd approach: Maximum-likelihood

Now, we assume that each of the observation error, ϵ_1^o and ϵ_2^o, is distributed with a Gaussian probability density function (pdf). This means that the probability for the observation error to lie between ϵ_i^o et $\epsilon_i^o + d\epsilon_i^o$ is

$$p(\epsilon_i^o) = \frac{1}{\sigma_i\sqrt{2\pi}} exp(-\frac{\epsilon_i^{o2}}{2\sigma_i^2})$$

with

$$\sigma_i^2 = \int_{-\infty}^{\infty} \epsilon_i^{o2} p(\epsilon_i^o) d\epsilon_i^o.$$



If we only have a single observation, then the most likely value of x is the value for which $p(\epsilon)$ is maximum. From the expression of p, this happens for

$$\epsilon = 0$$

that is

$$x^a = y^o.$$

If we have 2 observations, then the joint probability that observation 1 error lies between ϵ_1^o and $\epsilon_1^o + d\epsilon_1^o$
and that observation 2 error lies between ϵ_2^o et $\epsilon_2^o + d\epsilon_2^o$
is given by the product of the individual probabilities :

$$\begin{aligned} p(\epsilon_1^o, \epsilon_2^o) &= p(\epsilon_1^o)p(\epsilon_2^o) \\ &\propto \prod_{n=1}^{2} exp(-\frac{\epsilon_i^{o2}}{2\sigma_i^2}) \\ &\propto exp(-(\frac{(x-y_1^o)^2}{2\sigma_1^{o2}} + \frac{(x-y_2^o)^2}{2\sigma_2^{o2}})). \end{aligned}$$

In this case, the most likely value of x is the one for which $p(\epsilon_1^o, \epsilon_2^o)$ is maximum, which happens when the expression inside the exponant is minimum.

This is the **Maximum-likelihood** criterion.

Thus, x^a should minimize

$$J(x) = \frac{1}{2}\frac{(x-y_1^o)^2}{\sigma_1^{o2}} + \frac{1}{2}\frac{(x-y_2^o)^2}{\sigma_2^{o2}}.$$

$\Rightarrow$ If the observation errors are distributed with Gaussian pdfs, then the Maximum-likelihood estimate is the same as the least squares estimate and thus also the same as the the estimate that minimizes the estimation error variance.

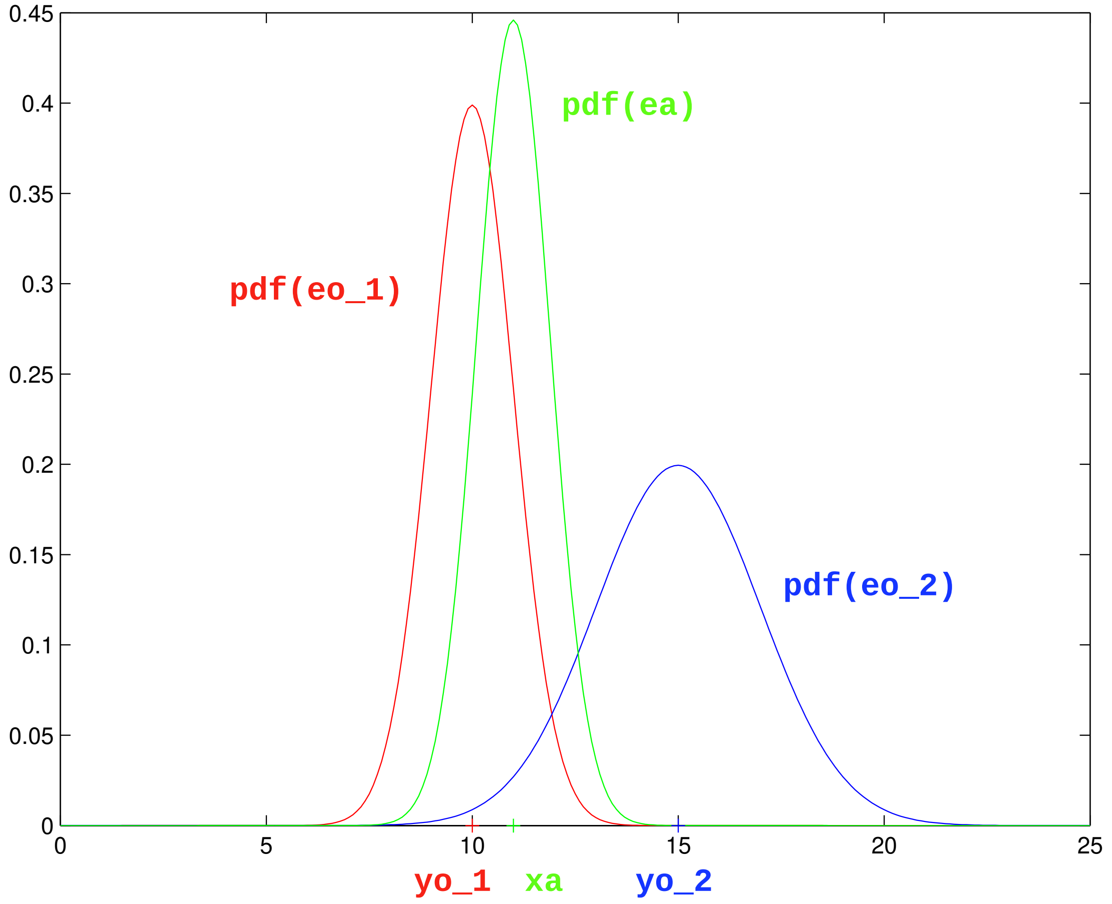


Figure 1: Gaussian pdfs for ϵ_1^o, ϵ_2^o and ϵ^a.

2 Generalization to the multi-dimensional case

We now consider that we seek a vector $\boldsymbol{x}$ with dimension n.
$\boldsymbol{x}$ can be composed of
- the value(s) at a single point (previous case),
- or at the points of a 1D, 2D, 3D grid,
- of one or many parameters (temperature, wind components, ...).

In the most general case, $\boldsymbol{x}$ aims to describe the atmospheric flow: **state vector**.

Similarly we note $\boldsymbol{y}^o$ the vector, with dimension p, of the set of available observations.

We can write $\boldsymbol{y}^o$ under the form

$$\boldsymbol{y}^o = H(\boldsymbol{x}^t) + \boldsymbol{\epsilon}^H + \boldsymbol{\epsilon}^o = H(\boldsymbol{x}^t) + \boldsymbol{\epsilon}.$$

H is the operator that allows to establish a relationship between the parameters to be estimated and the observations.

If the **observation operator** H is linear, then there is a matrix $\boldsymbol{H}$, with dimension $p \times n$, such as $H(\boldsymbol{x}) = \boldsymbol{H}\boldsymbol{x}$. We call **observation matrix** the corresponding matrix $\boldsymbol{H}$.

$\boldsymbol{\epsilon}^H$ is the error related to the form of H (**representativeness error**). $\boldsymbol{\epsilon}^o$ is the instrument error. Both errors are often mixed in a single term $\boldsymbol{\epsilon}$.

Form of H

In the previous case of 2 co-localized observations, we can check that

$$H = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\boldsymbol{y}^o = \boldsymbol{H}\boldsymbol{x}^t + \boldsymbol{\epsilon}^o$$

$$\begin{pmatrix} y_1^o \\ y_2^o \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \boldsymbol{x}^t + \begin{pmatrix} \epsilon_1^o \\ \epsilon_2^o \end{pmatrix}.$$

In a 1D, 2D, 3D space, H involves an interpolation of the state vector at the observation points.

In the general case, in which we have to deal with observations distributed in time, H can include the integration of a forecast model.

$\boldsymbol{H}$ can correspond
- to the linearization of the non-linear relation between an indirect observation and the state variable (use of satellite radiances in 3D/4D-Var, for example),
- or to the linearization of the forecast model (4D-Var).

Expression of H in the Arpège/Aladin 3D-Var

Case of a single observation at given point.

H includes:

- a transformation from spectral to physical space,
- an interpolation of the 4 (or 12 points) surrounding the observation,
- a vertical interpolation at the level of the observation,
- the computation of the model equivalent of the observation.

1st approach: minimization of the estimation error variance

As in the 2 observations case, we consider that the errors associated with the observations are un-biased:

$$E(\boldsymbol{\epsilon}^o) = \mathbf{0}.$$

$$E\begin{pmatrix} \epsilon_1^o \\ \epsilon_2^o \\ . \\ . \\ \epsilon_p^o \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ 0 \end{pmatrix}.$$

We note $\boldsymbol{R}$ the matrix of variance-covariance of the observation errors.

$$E(\boldsymbol{\epsilon}^o \boldsymbol{\epsilon}^{oT}) = \boldsymbol{R}.$$

$$\boldsymbol{R} = \begin{pmatrix} \sigma_1^{o2} & . & . & . & . \\ . & \sigma_2^{o2} & . & . & . \\ . & . & \sigma_i^{o2} & E(\epsilon_i^o \epsilon_j^o) & . \\ . & . & . & . & . \\ . & . & . & . & \sigma_p^{o2} \end{pmatrix}.$$

Thus, in comparison, with the previous simplified case we introduce the possibility to have cross-correlated observation errors.

We seek the estimate $\boldsymbol{x}^a$ under the form of a linear combination of observations:

$$\boldsymbol{x}^a = \boldsymbol{K}\boldsymbol{y}^o,$$

where $\boldsymbol{K}$ is a $n \times p$ matrix to be determined.

$\boldsymbol{x}^a$ must be an un-biased estimate, that is

$$E(\boldsymbol{x}^a - \boldsymbol{x}^t) = E(\boldsymbol{K}(\boldsymbol{H}\boldsymbol{x}^t + \boldsymbol{\epsilon}^o) - \boldsymbol{x}^t) = \boldsymbol{K}E(\boldsymbol{\epsilon}^o) + (\boldsymbol{K}\boldsymbol{H} - \boldsymbol{I}_{n\times n})E(\boldsymbol{x}^t) = \mathbf{0},$$

where $\boldsymbol{I}_{n\times n}$ is the identity matrix of order n.

This condition is fulfilled if

$$\boldsymbol{K}\boldsymbol{H} = \boldsymbol{I}_{n\times n}.$$

Among all the estimates verifying this relation, we seek the one that minimizes the estimation error variance, that is which minimizes the trace of the $n \times n$ matrice of error variance-covariance

$$\boldsymbol{A} = E((\boldsymbol{x}^a - \boldsymbol{x}^t)(\boldsymbol{x}^a - \boldsymbol{x}^t)^T)$$

It can be shown that the solution of this problem is given by

$$\boldsymbol{K} = (\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1},$$

and thus

$$\boldsymbol{x}^a = \boldsymbol{K}\boldsymbol{y}^o = (\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{y}^o.$$

The vector $\boldsymbol{x}^a$, defined by this expression, is called **Best Linear Unbiased Estimate (BLUE)**.

Quality of the estimation

Substituting for the expression of $\boldsymbol{x}^a$ in $\boldsymbol{A}$, we get

$$\boldsymbol{A} = (\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1},$$

Then $\boldsymbol{x}^a$ also writes

$$\boldsymbol{x}^a = \boldsymbol{K}\boldsymbol{y}^o = (\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{y}^o = \boldsymbol{A}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{y}^o.$$

or alternatively

$$\boldsymbol{A}^{-1}\boldsymbol{x}^a = \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{y}^o.$$

$\boldsymbol{A}$ describes the variances-covariances of the estimation errors corresponding to the components of $\boldsymbol{x}^a$.

For $\boldsymbol{H} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\boldsymbol{R} = \begin{pmatrix} \sigma_1^{o2} & 0 \\ 0 & \sigma_2^{o2} \end{pmatrix}$, corresponding to the previous example of 2 observations, matrix $\boldsymbol{A}$ is reduced to a scalar and we simply get

$$\boldsymbol{A} = \sigma^{a2} = (\frac{1}{\sigma_1^{o2}} + \frac{1}{\sigma_1^{o2}})^{-1}$$

and

$$\boldsymbol{K} = \begin{pmatrix} \dfrac{\sigma_2^{o2}}{\sigma_1^{o2} + \sigma_2^{o2}} & \dfrac{\sigma_1^{o2}}{\sigma_1^{o2} + \sigma_2^{o2}} \end{pmatrix}.$$

2nd approach : weighted least squares

As in the the case of the 2 observations, we can define the following quadratic expression of $\boldsymbol{x}$:

$$J(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}^o)^T \boldsymbol{R}^{-1} (\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}^o).$$

Under this form, the best estimate of $\boldsymbol{x}^t$ is the one that minimizes the distance to observations $\boldsymbol{y}^o$, weighted by our knowledge of the respective measurement error variances.

The derivative of J with respect to $\boldsymbol{x}$ writes :

$$\nabla_{\boldsymbol{x}} J = \boldsymbol{H}^T \boldsymbol{R}^{-1} (\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}^o).$$

It is a vector composed of the derivatives of J with respect to each component of $\boldsymbol{x}$ (gradient vector).

The minimum is then reached for

$$\boldsymbol{x}^a = (\boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{H})^{-1} \boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{y}^o.$$

$\Rightarrow$ We find again the "BLUE" expression.

3rd approach: Maximum-likelihood

Now, we assume that the observation error vector follows a multi-dimensional Gaussian pdf. This writes

$$\begin{aligned} p(\boldsymbol{\epsilon^o}) &\propto exp(-\frac{1}{2}(\boldsymbol{\epsilon^o})^T \boldsymbol{R}^{-1} \boldsymbol{\epsilon^o}) \\ &\propto exp(-\frac{1}{2}(\boldsymbol{Hx} - \boldsymbol{y}^o)^T \boldsymbol{R}^{-1} (\boldsymbol{Hx} - \boldsymbol{y}^o)). \end{aligned}$$

In this case, the most likely value of $\boldsymbol{x}$ is the one for which $p(\boldsymbol{\epsilon^o})$ is maximum, which happens when the expression inside the exponant is minimum.

Thus, $\boldsymbol{x^a}$ should minimize

$$J(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{Hx} - \boldsymbol{y}^o)^T \boldsymbol{R}^{-1} (\boldsymbol{Hx} - \boldsymbol{y}^o).$$

We get the same result as in the 2 observations case.

$\Rightarrow$ If the observation errors are distributed with multi-dimensional Gaussian pdf, then the Maximum-likelihood estimate is the same as the least squares estimate and thus also the same as the the estimate that minimizes the estimation error variance.

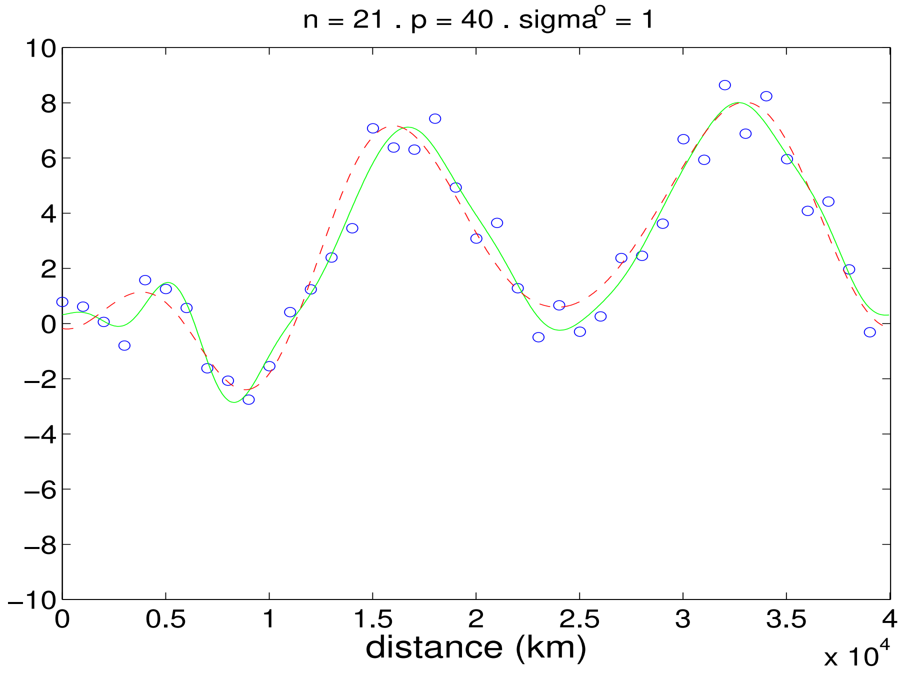


Figure 1: **Over-determined case**; dashed line: truth; circles: observations; solid line: estimation.

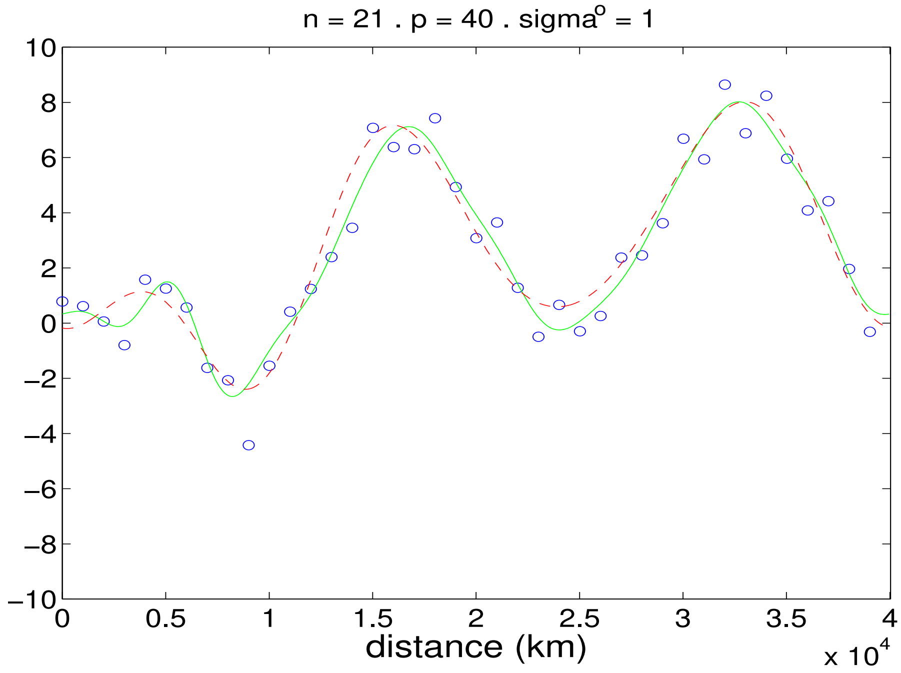


Figure 2: Idem, but 10th observation has a 3 times larger error variance.

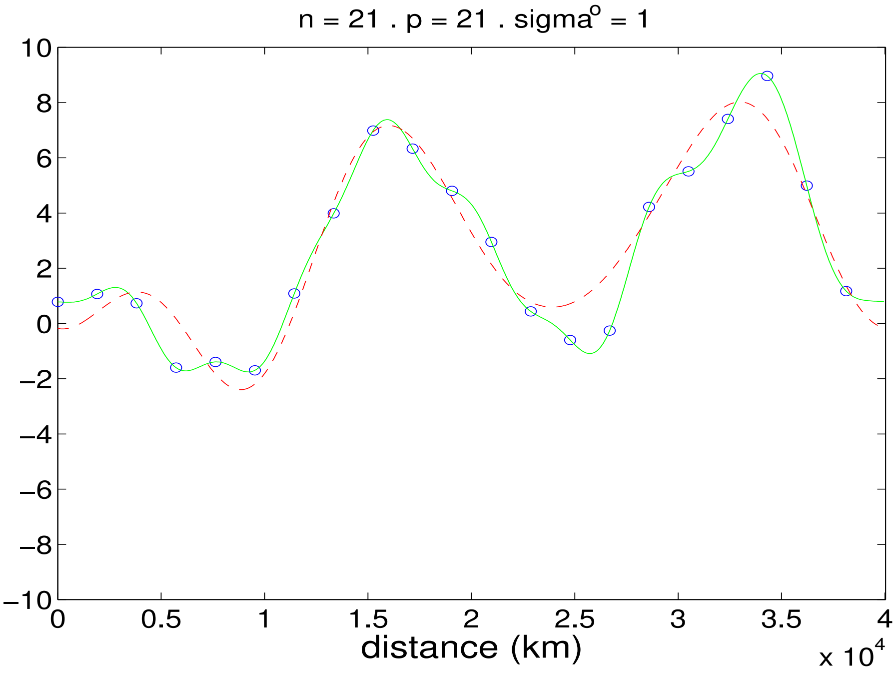


Figure 3: **No over-determination**: p = 21 and n = 21.

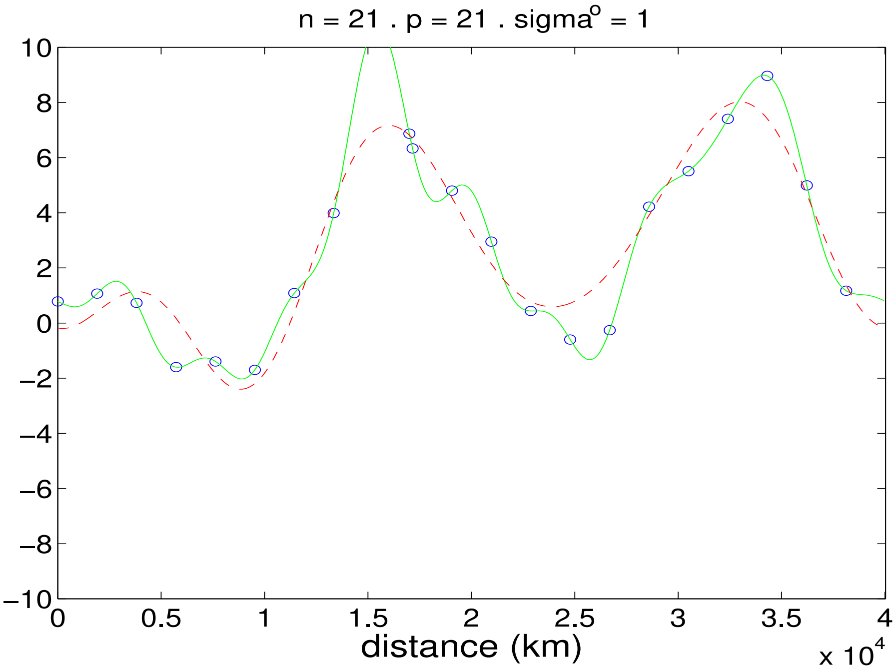


Figure 4: Idem, but observations 9 et 10 are closer.

3 Using a background

In order to "over-determine" the estimation problem, we will now make use of 2 sources of information :

- a set of measurements $\boldsymbol{y}^o$, with dimension p.

- a **background** $\boldsymbol{x}^b$ with same dimension as the vector $\boldsymbol{x}$ that we want to estimate (*prior* information).

The background can be written

$$\boldsymbol{x}^b = \boldsymbol{x}^t + \boldsymbol{\epsilon}^b$$

(in this case, the observation operator id reduced to the identity matrix $\boldsymbol{I}_n$),

$\boldsymbol{\epsilon}^b$ is the vector of background errors.

As previously, we can write

$$\boldsymbol{y}^o = \boldsymbol{H}\boldsymbol{x}^t + \boldsymbol{\epsilon}^o.$$

The ensemble of available "observations" is thus composed by the vector $\boldsymbol{z}^o = \begin{pmatrix} \boldsymbol{x}^b \\ \boldsymbol{y}^o \end{pmatrix}$, with dimension $n + p$.

We have

$$\begin{aligned} \boldsymbol{z}^o &= \begin{pmatrix} \boldsymbol{I}_{n\times n} \\ \boldsymbol{H} \end{pmatrix} \boldsymbol{x}^t + \begin{pmatrix} \boldsymbol{\epsilon}^b \\ \boldsymbol{\epsilon}^o \end{pmatrix} \\ &= \boldsymbol{H}^{z^o} \boldsymbol{x}^t + \boldsymbol{\epsilon}^{z^o}. \end{aligned}$$

Hypothèses

We will assume that the errors $\boldsymbol{\epsilon}^{z^o}$ of $\boldsymbol{z}^o$, i.e. the background and observation errors, respectively $\boldsymbol{\epsilon}^b$ and $\boldsymbol{\epsilon}^o$, are un-biased:

$$E(\boldsymbol{\epsilon}^{z^o}) = \begin{pmatrix} E(\boldsymbol{\epsilon}^b) \\ E(\boldsymbol{\epsilon}^o) \end{pmatrix} = \mathbf{0}.$$

We also suppose that we know the form of the covariances between these errors:

$$\boldsymbol{R}^{z^o} = \begin{pmatrix} E(\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^b)^T) & E(\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^o)^T) \\ E(\boldsymbol{\epsilon}^o(\boldsymbol{\epsilon}^b)^T) & E(\boldsymbol{\epsilon}^o(\boldsymbol{\epsilon}^o)^T) \end{pmatrix} = \begin{pmatrix} \boldsymbol{B} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{R} \end{pmatrix}.$$

(we postulate that background and observations errors are not correlated: $E(\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^o)^T) = \mathbf{0}_{n\times p}$; $E(\boldsymbol{\epsilon}^o(\boldsymbol{\epsilon}^b)^T) = \mathbf{0}_{p\times n}$).

$\boldsymbol{B}$ is the background error covariance matrix.

$\boldsymbol{R}$ is the observations errors covariance matrix, related to the precision of the various observations.

Expression of the solution

We seek a solution under the form

$$\boldsymbol{x}^a = \boldsymbol{K}^{z^o}\boldsymbol{z}^o = \boldsymbol{K}^{z^o}\begin{pmatrix} \boldsymbol{x}^b \\ \boldsymbol{y}^o \end{pmatrix}.$$

By applying the above expression for $\boldsymbol{K}$, the estimate writes

$$\begin{aligned}
\boldsymbol{x}^a &= (\boldsymbol{H}^{z^oT}\boldsymbol{R}^{z^o-1}\boldsymbol{H}^{z^o})^{-1}\boldsymbol{H}^{z^oT}\boldsymbol{R}^{z^o-1}\begin{pmatrix} \boldsymbol{x}^b \\ \boldsymbol{y}^o \end{pmatrix} \\
&= (\begin{pmatrix} \boldsymbol{I}_{n\times n} \\ \boldsymbol{H} \end{pmatrix}^T \begin{pmatrix} \boldsymbol{B} & 0 \\ 0 & \boldsymbol{R} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{H} \end{pmatrix})^{-1} \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{H} \end{pmatrix}^T \begin{pmatrix} \boldsymbol{B} & 0 \\ 0 & \boldsymbol{R} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{x}^b \\ \boldsymbol{y}^o \end{pmatrix} \\
&= (\boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}(\boldsymbol{B}^{-1}\boldsymbol{x}^b + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{y}^o).
\end{aligned}$$

The estimation error covariance matrix writes

$$\boldsymbol{A} = (\boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}$$

or

$$\boldsymbol{A}^{-1} = \boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}.$$

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

Then we have

$$\boldsymbol{x}^a = (\boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}(\boldsymbol{B}^{-1}\boldsymbol{x}^b + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{y}^o)$$

or alternatively

$$\boldsymbol{A}^{-1}\boldsymbol{x}^a = \boldsymbol{B}^{-1}\boldsymbol{x}^b + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{y}^o.$$

2nd approach : weighted least squares

We can also simply define the following quadratic expression of $\boldsymbol{x}$:

$$J(\boldsymbol{x}^a) = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^b)^T\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}^b) + \frac{1}{2}(\boldsymbol{H}\boldsymbol{x}-\boldsymbol{y}^o)^T\boldsymbol{R}^{-1}(\boldsymbol{H}\boldsymbol{x}-\boldsymbol{y}^o).$$

This scalar function (or **cost function**) measures the distance between the estimate and the available information, i.e. the background and the measurements. It also takes into account the respective precisions accorded to each source of information.

The gradient of J with respect to $\boldsymbol{x}$ writes

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

The expression of the solution is thus

$$\boldsymbol{x}^a = (\boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}(\boldsymbol{B}^{-1}\boldsymbol{x}^b + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{y}^o),$$

i.e. the expression of the "BLUE" previously derived.

3rd approach: Maximum-likelihood

We assume that the observation error vector but also the background error vector follow multi-dimensional Gaussian pdfs. This writes

$$\begin{aligned} p(\boldsymbol{\epsilon^o}) &\propto exp(-\frac{1}{2}(\boldsymbol{\epsilon^o})^T \boldsymbol{R}^{-1}\boldsymbol{\epsilon^o}) \\ &\propto exp(-\frac{1}{2}(\boldsymbol{Hx}-\boldsymbol{y}^o)^T \boldsymbol{R}^{-1}(\boldsymbol{Hx}-\boldsymbol{y}^o)) \end{aligned}$$

and

$$\begin{aligned} p(\boldsymbol{\epsilon^b}) &\propto exp(-\frac{1}{2}(\boldsymbol{\epsilon^b})^T \boldsymbol{B}^{-1}\boldsymbol{\epsilon^b}) \\ &\propto exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^b)^T \boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}^b)). \end{aligned}$$

with their joint probabilities given by

$$p(\boldsymbol{\epsilon^o},\boldsymbol{\epsilon^b}) \propto exp(-\frac{1}{2}((\boldsymbol{Hx}-\boldsymbol{y}^o)^T \boldsymbol{R}^{-1}(\boldsymbol{Hx}-\boldsymbol{y}^o)+(\boldsymbol{x}-\boldsymbol{x}^b)^T \boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}^b))).$$

The most likely value of $\boldsymbol{x}$ is the one for which $p(\boldsymbol{\epsilon^o},\boldsymbol{\epsilon^b})$ is maximum, which happens when the expression inside the exponant is minimum.

Thus, $\boldsymbol{x^a}$ should minimize

$$J(\boldsymbol{x}) = \frac{1}{2}((\boldsymbol{Hx}-\boldsymbol{y}^o)^T \boldsymbol{R}^{-1}(\boldsymbol{Hx}-\boldsymbol{y}^o)+(\boldsymbol{x}-\boldsymbol{x}^b)^T \boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}^b)).$$

$\Rightarrow$ If the observation and background errors are distributed with multi-dimensional Gaussian pdfs, then the Maximum-likelihood estimate is the same as the least squares estimate and thus also the same as the the estimate that minimizes the estimation error variance.

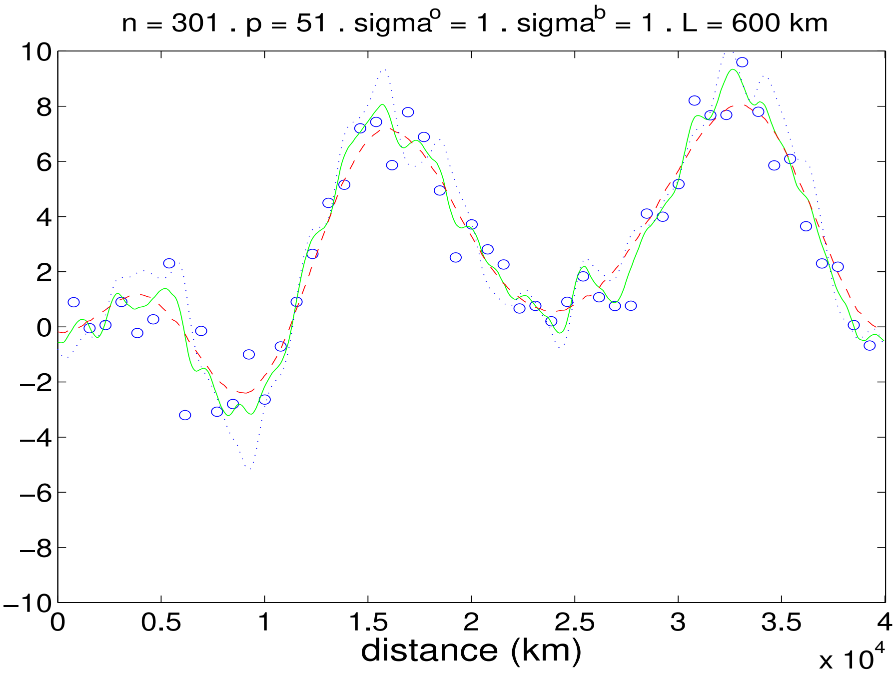


Figure 1: Dashed line: truth; dotted line: background; circles: observations ; solid line: estimate.

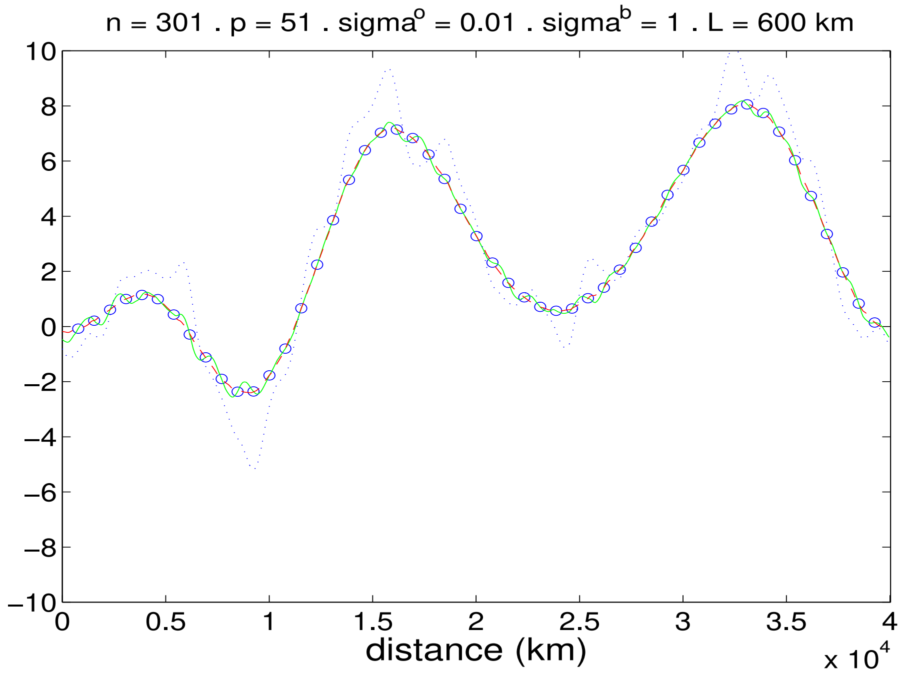


Figure 2: Idem, but with $\boldsymbol{R} \ll \boldsymbol{B}$.

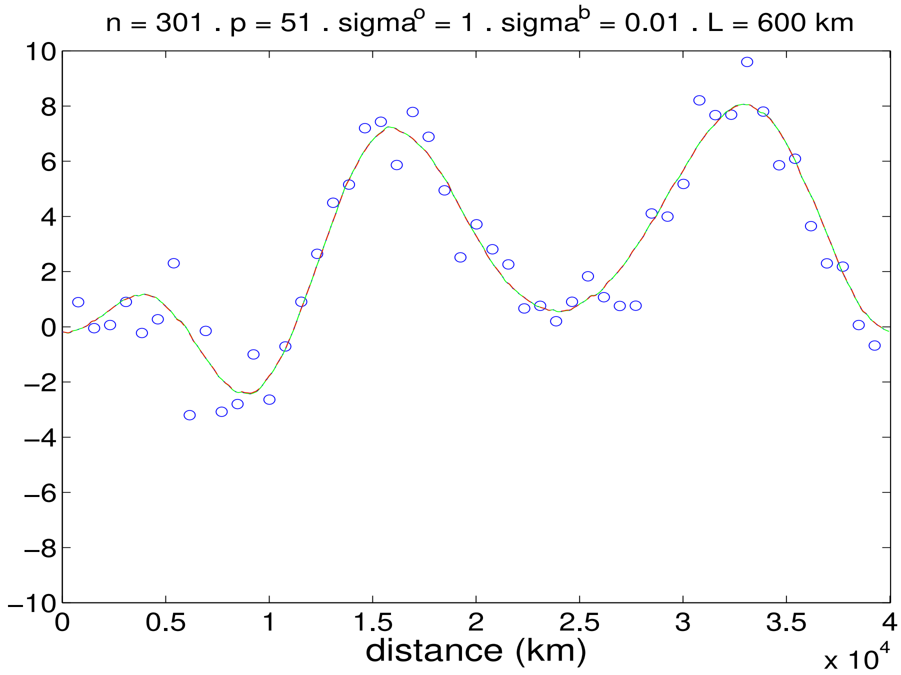


Figure 3: Idem, but with $\boldsymbol{R} \gg \boldsymbol{B}$.

Alternative expressions of the solution

The expression of the estimate

$$\boldsymbol{x}^a = (\boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}(\boldsymbol{B}^{-1}\boldsymbol{x}^b + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{y}^o)$$

can be re-written as

$$\begin{aligned}\boldsymbol{x}^a &= (\boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}(\boldsymbol{B}^{-1}\boldsymbol{x}^b + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{x}^b)\\ &+ (\boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}(\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{x}^b)\end{aligned}$$

that is

$$\boldsymbol{x}^a = \boldsymbol{x}^b + (\boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}(\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{x}^b).$$

It can be shown that this can be again re-written as

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{B}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T + \boldsymbol{R})^{-1}(\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{x}^b)$$

and the estimation covariance matrix as

$$\boldsymbol{A} = \boldsymbol{B} - \boldsymbol{B}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T + \boldsymbol{R})^{-1}\boldsymbol{H}\boldsymbol{B}.$$

The vector $\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{x}^b$ is called **innovation** vector.

The matrix $\boldsymbol{K} = \boldsymbol{B}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T + \boldsymbol{R})^{-1}$ is called **gain matrix**.

4 Optimal Interpolation

The optimal estimate, previously derived, is given by

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{B}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T + \boldsymbol{R})^{-1}(\boldsymbol{y}^o - \boldsymbol{H}\boldsymbol{x}^b).$$

In the Optimal Interpolation (OI) formalism (Gandin, 1963), the notion of observation operator is not explicitely introduced. One assumes that a continuous representation of the atmosphere is available (which is of course wrong). Then the notion of representativeness error does not clearly appear.

In OI, the expression of the estimate is rather given by

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{B}^{xy}(\boldsymbol{B}^{yy} + \boldsymbol{R})^{-1}(\boldsymbol{y}^o - \boldsymbol{x}^{b,o}),$$

where $\boldsymbol{x}^{b,o}$ is the evaluation of the background at observation points. Matrices $\boldsymbol{B}^{xy}$ and $\boldsymbol{B}^{yy}$ respectively describe background error covariances between grid-points and observation points, and between observation points.

It is assumed that it exists a continuous correlation model that allows to compute matrices $\boldsymbol{B}^{xy}$ and $\boldsymbol{B}^{yy}$.

The estimation covariance matrix can also be determined with

$$\boldsymbol{A} = \boldsymbol{B} - \boldsymbol{B}^{xy}(\boldsymbol{B}^{yy} + \boldsymbol{R})^{-1}\boldsymbol{B}^{xy},$$

but in practice only estimation error variances (the diagonal elemenents of $\boldsymbol{A}$) are most often calculated.

Selection of observations

The main limitation comes from the dimension $p \times p$ of the matrix $(\boldsymbol{B}^{yy} + \boldsymbol{R})$ to be inversed. In practice, p is of the order of 10^4 to 10^5 !

Then to make the problem tractable, the number of observations is artificially reduced.

Two kinds of selections of the observations are performed.

- By "box" (ECMWF/ old Optimal Interpolation scheme, for example): the domain is split in sub-domains and a matrix inversion is performed for each sub-domain.

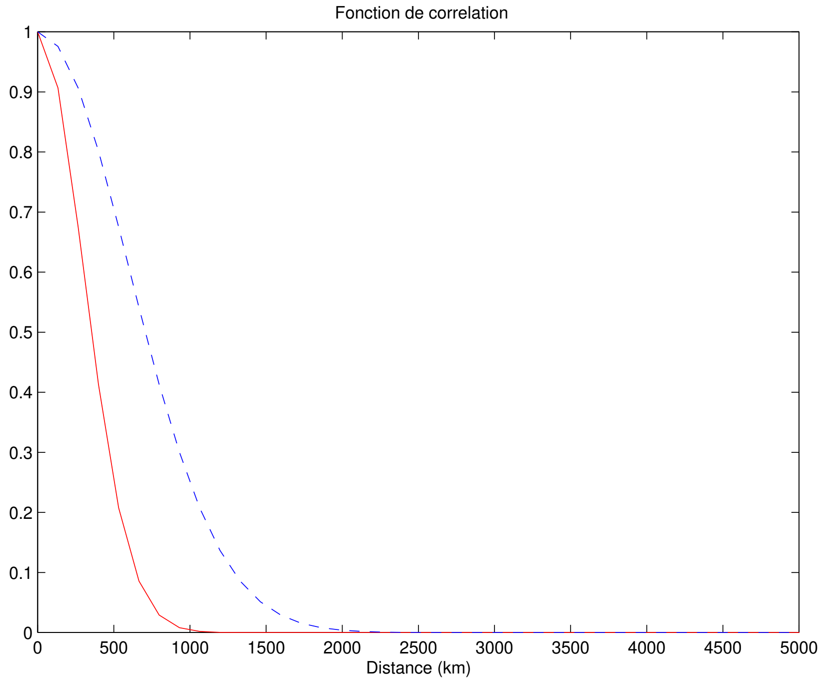
- By grid-point (CANARI/DIAGPACK): each grid-point is successively analyzed with only a few observations in its neighbourhood.

5 Structure functions

Example of correlation function

Gaussian model: $\rho_{\epsilon^b \epsilon^b}(r) = e^{-\frac{r^2}{2L^2}}$.

L is the **length-scale** of the correlation function (**structure function**).



Influence of the length-scale of a structure function

For a single observation, $\boldsymbol{H} = \boldsymbol{h}$ with only one line in the matrix. Then

$$\begin{aligned}\boldsymbol{x}^a - \boldsymbol{x}^b &= \boldsymbol{B}\boldsymbol{h}^T(\boldsymbol{h}\boldsymbol{B}\boldsymbol{h}^T + \boldsymbol{R})^{-1}(y^o - H(\boldsymbol{x^b})) \\ &= \boldsymbol{B}\boldsymbol{h}^T((\sigma^b)^2 + (\sigma^o)^2)^{-1}(y^o - H(\boldsymbol{x^b})).\end{aligned}$$

If the background error variance is uniform, then

$$\boldsymbol{x}^a - \boldsymbol{x}^b = \boldsymbol{\rho}\frac{(\sigma^b)^2}{(\sigma^b)^2 + (\sigma^o)^2}(y^o - H(\boldsymbol{x^b})),$$

where $\boldsymbol{\rho}$ is the vector given by the correlation function based at the observation point.

The correlation introduced in $\boldsymbol{B}$ controls the way the discrepancies between observations and background will be spread.

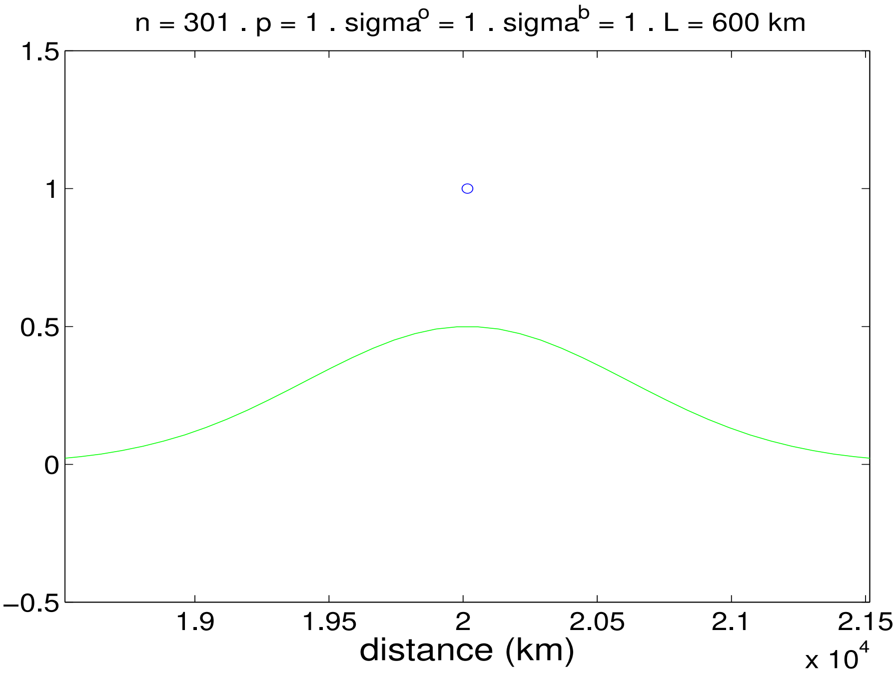


Figure 1: 1 single observation and L = 600 km (circle : $y^o - H(\boldsymbol{x}^b)$.

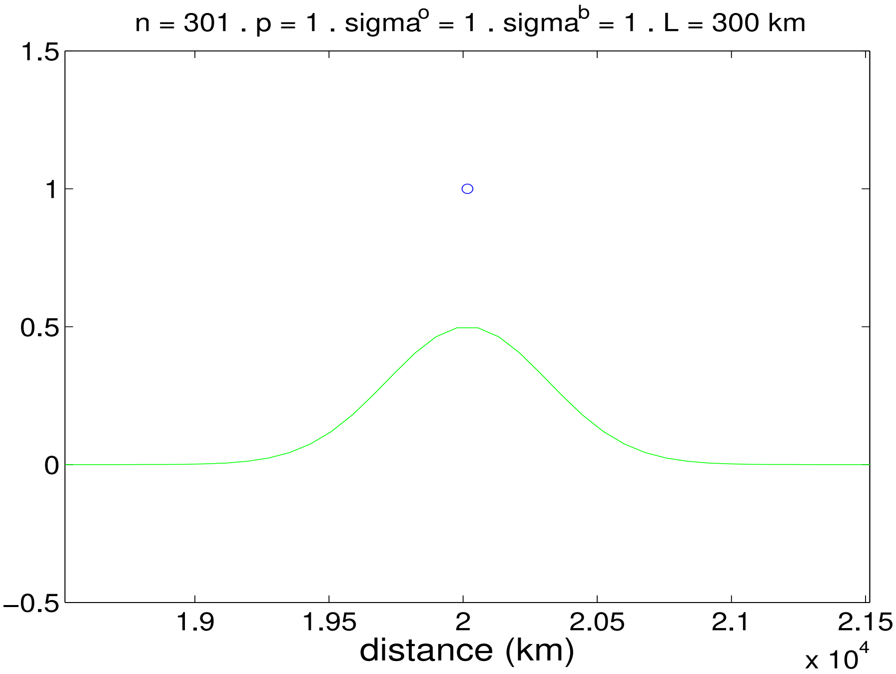


Figure 2: Idem, but with L = 300 km.

6 Using a model: the Kalman Filter

Principle of a 3D-Var
or Optimal Interpolation scheme

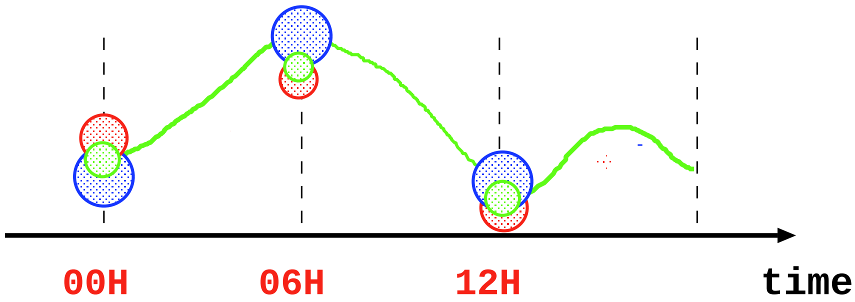


Figure 1: Circles symbolyzed the uncertainty around background $\boldsymbol{x}^b$, observations $\boldsymbol{y}^o$ and analysis $\boldsymbol{x}^a$.

In a 3D-Var or Optimal Interpolation scheme, the background error covariances are kept unchanged along time.

Matrix $\boldsymbol{B}$ is modelized from statistics and by relying on strong hypotheses: homogeneity, isotropy.

Evolution of the background error covariances: the Kalman Filter

We explicitely introduce the forecast model that allows to build the background:

$$\boldsymbol{x}^t_{i+1} = M_{(i,i+1)}\boldsymbol{x}^t_i + \boldsymbol{\epsilon}^m_i,$$

where $\boldsymbol{x}^t_{i+1}$ and $\boldsymbol{x}^t_i$ are respectively the true states of the atmospheric flow at times t_{i+1} et t_i and $M_{(i,i+1)}$ is the "forecast model" operator from time t_i to time t_{i+1}.

$\boldsymbol{\epsilon}^m_i$ is the error introduced by using this forecast model to make evolve the true states of the flow: it is the **model error**.

Hypotheses

The model error at time t_i is un-biased:

$$E(\boldsymbol{\epsilon}^m_i) = \mathbf{0}.$$

The covariance matrix of the errors $\boldsymbol{\epsilon}^m_i$ is known:

$$E(\boldsymbol{\epsilon}^m_i(\boldsymbol{\epsilon}^m_i)^T) = \boldsymbol{Q}_i.$$

We also assume that analysis error $\boldsymbol{\epsilon}^a_i$ at time t_i and model error $\boldsymbol{\epsilon}^m_i$ are not correlated:

$$E(\boldsymbol{\epsilon}^a_i(\boldsymbol{\epsilon}^m_i)^T) = \mathbf{0}.$$

The Kalman Filter steps

At each time t_i, we have the following steps.

- 1. Computation of the innovation vector:

$$\boldsymbol{d}_i = \boldsymbol{y}_i^o - H_i(\boldsymbol{x}_i^f).$$

- 2. Computation of the gain matrix:

$$\boldsymbol{K}_i = \boldsymbol{P}_i^f \boldsymbol{H}_i{}^T (\boldsymbol{H}_i \boldsymbol{P}_i^f \boldsymbol{H}_i{}^T + \boldsymbol{R}_i)^{-1}.$$

- 3. Determination of the analysis:

$$\boldsymbol{x}_i^a = \boldsymbol{x}_i^f + \boldsymbol{K}_i \boldsymbol{d}_i.$$

- 4. Updating of the estimation error covariances:

$$(\boldsymbol{A} =)\boldsymbol{P}_i^a = \boldsymbol{P}_i^f - \boldsymbol{K}_i \boldsymbol{H}_i{}^T \boldsymbol{P}_i^f.$$

- 5. Evolution of the state vector:

$$\boldsymbol{x}_{i+1}^f = M_{(i,i+1)}(\boldsymbol{x}_i^a).$$

- 6. Evolution of the estimation error covariances:

$$\boldsymbol{P}_{i+1}^f = \boldsymbol{M} \boldsymbol{P}_i^a \boldsymbol{M}^T + \boldsymbol{Q}_i.$$

In an Optimal Interpolation (or a 3D-Var) scheme, the last step is omitted.

For a 4D-Var scheme, it can be shown that there is an implicit evolution of $\boldsymbol{P}$ during the assimilation period.