Introduction to Optimal Interpolation and Variational Analysis

Alexander Barth, Aida Alvera Azcárate, Pascal Joassin, Jean-Marie Beckers, Charles Troupin

a.barth@ulg.ac.be

Revision 1.1

November 26, 2008
# Contents

1 Theory of optimal interpolation and variational analysis ................................................................. 2
   1.1 The gridding problem .................................................................................................................. 2
   1.2 Observation errors ...................................................................................................................... 5
   1.3 Data analysis and gridding methodologies .................................................................................. 6
   1.4 Objective analysis ...................................................................................................................... 6
   1.5 Cressman method ....................................................................................................................... 8
   1.6 Optimal interpolation .................................................................................................................. 10
       1.6.1 Error covariances .............................................................................................................. 11
       1.6.2 Assumptions ..................................................................................................................... 12
       1.6.3 Analysis .......................................................................................................................... 12
   1.7 Variational Inverse Method ....................................................................................................... 14
       1.7.1 Equivalence between OI and variational analysis .............................................................. 16
       1.7.2 Weights on data .............................................................................................................. 16
       1.7.3 Correlation function ........................................................................................................ 17
       1.7.4 Error fields by analogy with optimal interpolation ........................................................... 17
       1.7.5 Diva .................................................................................................................................. 17
       1.7.6 Example ........................................................................................................................... 17
   1.8 Comparison .................................................................................................................................. 19
   1.9 Summary ....................................................................................................................................... 19

2 Modelling of error covariances ........................................................................................................... 21
   2.1 Background error covariance matrix .......................................................................................... 21
       2.1.1 Parametric error covariance .............................................................................................. 21
       2.1.2 Ensemble error covariance ............................................................................................... 23
       2.1.3 Error covariance based on differential operators ............................................................. 23
   2.2 Observation error covariance ..................................................................................................... 24
   2.3 Summary ....................................................................................................................................... 24

3 Parameter optimization and cross-validation ................................................................................. 26
   3.1 A priori parameter optimization ............................................................................................... 26
   3.2 Validation ....................................................................................................................................... 27
   3.3 Cross-validation ........................................................................................................................... 27
   3.4 Generalized cross-validation ....................................................................................................... 28
   3.5 Summary ....................................................................................................................................... 29

4 Real world example .......................................................................................................................... 30
   4.1 Data .............................................................................................................................................. 30
   4.2 Finite-element mesh ..................................................................................................................... 31
   4.3 Results .......................................................................................................................................... 32

A Some matrix identities ....................................................................................................................... 34

References ............................................................................................................................................... 35
Chapter 1

Theory of optimal interpolation and variational analysis

Contents

1.1 The gridding problem ........................................... 2
1.2 Observation errors ........................................... 5
1.3 Data analysis and gridding methodologies ................... 6
1.4 Objective analysis ........................................... 6
1.5 Cressman method ........................................... 8
1.6 Optimal interpolation ......................................... 10
  1.6.1 Error covariances ........................................ 11
  1.6.2 Assumptions ........................................... 12
  1.6.3 Analysis ............................................... 12
1.7 Variational Inverse Method .................................. 14
  1.7.1 Equivalence between OI and variational analysis ....... 16
  1.7.2 Weights on data ........................................ 16
  1.7.3 Correlation function ................................... 17
  1.7.4 Error fields by analogy with optimal interpolation ..... 17
  1.7.5 Diva .................................................. 17
  1.7.6 Example ............................................... 17
1.8 Comparison ................................................ 19
1.9 Summary ................................................ 19

1.1 The gridding problem

Gridding is the determination of a field $\phi(r)$, on regular grid of positions $r$ based on arbitrarily located observations. Often the vector $r$ is on a 2D horizontal plane in which case longitude and latitude are used to specify a location. But the vector $r$ can also represent a point in 3D and even 4D (longitude, latitude, depth and time).

The fewer observations are available, the harder the gridding problem is. In oceanography, in situ observations are sparse. Not only is the quantity of the observations a problem, but also their distribution. Most of the time, the observations are also inhomogeneously distributed in space and time. For example, there are more observations in the coastal zones in summer than in the open ocean in winter.

Moreover, the variability of the ocean is the sum of various processes occurring at different spatial and temporal scales. The available observations might be suitable to detect the large-scale variations, but not the small-scale processes. Even if we limit our analysis to the large-scale processes, the observations
contain all scales at the same time.

In light of all those difficulties, several approaches have been developed to make an “optimal” use of the limited amount of in situ data available.

Figure 1.1 shows an idealized square domain with a barrier (e.g. a peninsula or a dike). Let’s assume that the field represents temperature. The barrier suppresses the exchanges between each side of the barrier. The temperature on each side is indeed quite different. In this example, the field varies smoothly over some length-scale. As mentioned previously, in the real ocean there might be several length-scales present at the same time.

![Figure 1.1: Example of oceanographic field. This field is the true field that we want to reconstruct based on observations.

The field is sampled in the locations shown in figure 1.2. In regions where a measurement campaign has been carried out, a higher spatial coverage is achieved. However, some large gaps are also present. The distribution of observations is inhomogeneous. Based on the value of the field at the shown location, we will estimate the true field.](image)
Figure 1.2: Sampling locations within the domain

Figure 1.3: Value of the true field extract at the location of the observations.

The figure 1.3 shows only the value of the observations. Obviously some information about the
position of the structures and fronts is lost by having observations only at a limited set of locations. While the different gridding methods examined here aim to reproduce the original field as realistically as possible, one cannot expect to obtain exactly the true field. But the more information about its structure and evolution we include in the analysis, to close we can get to the true field. For example, in the context of data assimilation the equations governing the evolution of a field are include in the analysis.

1.2 Observation errors

Observations are in general affected by different error sources and other “problems” that need to be taken into account:

- Instrumental errors (limited precision or possible bias of the sensor)
- Representative errors: the observations do not necessarily corresponds to the field we want to obtain. For example, we want to have a monthly average, but the observations are instantaneous (or averages over a very short period of time).
- Synopticity errors: all observations are not taken at the same time.
- Other errors sources: human errors (e.g. permutation of longitude and latitude), transmission errors, malfunctioning of the instrument, wrong decimal separators....

While the three first source of error are taken into account during the analysis (since they are essentially of random nature), the error of the forth point should be either corrected beforehand if this is possible or the observations should not be used in the analysis.

Quality control is an important step to exclude suspicious data from the analysis. But since this is a subjective decision, the data should never be deleted but flagged as suspicious or bad data.

In figure 1.4, a random perturbation was added to the observation shown in figure 1.3. This simulates the impact of the different error sources. To simplify matters, each observation was perturbed independently.
1.3 Data analysis and gridding methodologies

Because observations have errors, it is always better to produce a field approximation and never a strict interpolation. Figure 1.5 shows what would happen if the observations would have been interpolated linearly. The noise on the observations is in general amplified when higher order schemes, such as cubic interpolation, are used. Hence in the following we will concentrate on approximating fields.

There are a number of methods that have been traditionally used to solve the gridding problem:

- Subjective: drawing by hand (1.6)
- Objective: predefined mathematical operations
- Data Assimilation: use of physical/biochemical dynamic governing equations

Choice:

- Subjective method is not sufficiently ... objective.
- Data Assimilation: region and model dependent.

⇒ Objective analysis of data that are anomalies with respect to a background field.

1.4 Objective analysis

As opposed to the subjective method, objective analysis techniques aim to use mathematical formulations to infer the value of the field at unobserved locations based on the observation $d_j$. Most objective methods can be expressed as a linear combination of data anomalies $d_j$ (with respect to the background field $\varphi_b(r)$) using weights $w_j$:

$$\varphi(r) = \varphi_b(r) + \sum_{j=1}^{N_d} w_j d_j$$  \hspace{1cm} (1.1)
Figure 1.5: Gridded field using linear interpolation. The domain is decomposed into triangles where the vertices are the location of the data points based on the Delaunay triangulation. Within each triangle, the value is interpolated linearly. This method is implemented in the function `griddata` of Matlab and GNU Octave.

Figure 1.6: Isohyet (lines of constant precipitation) drawn by hand (from http://www.srh.noaa.gov/hgx/hurricanes/1970s.htm)
The field $\varphi(r)$ can be evaluated in any position $r$, hence gridding is possible. The background field (or first guess) $\varphi_b$ is defined \textit{a priori} and anomalies calculated with respect to this reference field (for example a climatological average). There are several ways to define the weighting function $w_j$, which result in different gridding techniques. We will review in the following the most used gridding methods.

1.5 Cressman method

Cressman weights depend only on the distance $r$ between the location $r$ where the value of the field should be estimated and the location of the observation $r_j$:

$$ r = |r - r_j| $$

(1.2)

The weights are then parameterized according to,

$$ \tilde{w}_j(r) = \frac{R^2 - r^2}{R^2 + r^2} \quad \text{for} \quad r < R $$

$$ = 0 \quad \text{for} \quad r \geq R $$

(1.3)

The weights as a function of distance are shown in figure 1.7. Weights must be scaled by their sum to ensure no bias.

$$ w_j = \frac{\tilde{w}_j}{\sum_j \tilde{w}_j} $$

(1.4)

The search radius $R$ is the typical control parameter and defines the length-scale over which an observation is used. This length scale can be made to vary in space depending on data coverage and/or physical scales. This parameter is chosen by the users based on their knowledge of the domain and the problem.

![Figure 1.7: Cressman weights for $R = 2$ (blue) and Barnes weights for $R = 1$ (red).](image)
The Cressman weighting is a very simple and numerically quite efficient method. However, it suffers from some limitations which are apparent in figure 1.8.

- No estimate can be obtained at locations when no observation is located within the $R$.
- In regions with very few observations, the method can return a discontinuous field.
- The presence of barriers disconnecting different waters masses cannot be taken into account easily.
- Also all observations are assumed to have a similar error variance since the weighting is based only on distance.

As a variant of the Cressman weights, other weighting functions can be defined. In the Barnes scheme, the weights are defined using a Gaussian function:

$$
\tilde{w}(d) = e^{-\frac{d^2}{2\sigma^2}}
$$

(1.5)
Since the Barnes weights are never zero, in principle all observations are used for the gridding. An estimation can be obtained everywhere (which can be accurate or not). Artificial discontinuities are avoided using the Barnes weights (figure 1.9).

1.6 Optimal interpolation

To derive the optimal interpolation method (Gandin, 1965), it is convenient to introduce the concept of “state vector”. The state vector $x$ is a column vector containing all unknowns that we want to estimate from the observations. For our gridding problem, the state vector contains the value of the field at the discrete locations of the grid.

The observation vector $y^o$ contains all observations. Given a state vector $x$, one can extract, or interpolate, the field at the location of the observations. Since the gridded field is on a regular grid, one can use a method such as bilinear interpolation. This operation is expressed by the matrix $H$. When this matrix $H$ is applied (or multiplied to the left) to a state vector $x$, it returns the field interpolated at the location of the observations $Hx$ (see figure 1.10).
Figure 1.10: The operator $H(\cdot)$ interpolates gridded field to the locations of the observations. This figure shows schematically the position of measurements (crosses) and the gridded field (dots). In its simplest form, $H(\cdot)$ performs a bilinear interpolation.

The vector $\mathbf{x}^t$ is the true field (see figure 1.1) and we want to obtain an analysis $\mathbf{x}^a$ which is as close as possible to the true state given the first guess $\mathbf{x}^b$ (or background) and the observations $\mathbf{y}^o$ (see figure 1.4). However, the first guess and the observations may have errors:

$$\mathbf{x}^b = \mathbf{x}^t + \mathbf{\eta}^b$$

$$\mathbf{y}^o = H\mathbf{x}^t + \mathbf{\varepsilon}$$

where $\mathbf{\eta}^b$ represents the error of the first guess and $\mathbf{\varepsilon}$ the error of the observations.

### 1.6.1 Error covariances

Error covariances describe how different variables or the same variable at different locations are related in average. The covariance between the random variable $x$ and $y$ is defined as:

$$\text{cov}(x,y) = E[(x - E[x])(y - E[y])]$$

In practice, the expected value $E[\cdot]$ can be estimated by averaging over repeated realizations of the random variables $x_i$ and $y_i$ for $i = 1,\ldots,N$. The covariance of a variable with itself is the variance.

$$\text{var}(x) = \text{cov}(x,x) = E[(x - E[x])^2]$$

From the covariance one can compute the correlation,

$$\text{corr}(x,y) = \frac{\text{cov}(x,y)}{\sqrt{\text{var}(x) \cdot \text{var}(y)}}$$

The correlation is always bounded between -1 and 1. For vectors, the concept of covariance is extended to a covariance matrix. The elements $i,j$ of the covariance matrix of the vector $\mathbf{x}$ are the covariance between the elements $x_i$ and $x_j$.

$$\mathbf{P} = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T]$$

The covariance matrix is per construction symmetric and positive-semidefinite (its eigenvalues are positive). The diagonal elements of the covariance matrix are the variance of each element. The covariance matrix can also be decomposed in variance and correlation,

$$\mathbf{P} = \mathbf{D}\mathbf{C}\mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix. Its diagonal elements are the variance of the vector $\mathbf{x}$. The matrix $\mathbf{C}$ is the correlation matrix. All elements of the correlation matrix are contained between -1 and 1. The
diagonal elements of $C$ are 1 since they represent the correlation of a variable with itself.

### 1.6.2 Assumptions

In order to proceed we must make some hypotheses concerning those errors. We assume that the observations and the first guess are unbiased, *i.e.* that the error is zero in average:

$$
E[\eta^b] = 0 \quad (1.13) \\
E[\epsilon] = 0 \quad (1.14)
$$

It is also necessary to have an *a priori* knowledge of the magnitude of the observation and the first guess errors. Those are expressed as error covariance matrices:

$$
E[\eta^b \eta^b^T] = P^b \quad (1.15) \\
E[\epsilon \epsilon^T] = R \quad (1.16) \\
E[\eta^b \epsilon^T] = 0 \quad (1.17)
$$

The last equation states that the error of the first guess is independent of the observation errors.

### 1.6.3 Analysis

The optimal interpolation (OI) scheme can be derived as the Best Linear Unbiased Estimator (BLUE) of the true state $x^t$ which has the following properties:

- The estimator is linear in $x^b$ and $y^o$
- The estimator is not biased: $E[x^a] = x^t$ (1.18)
- This estimate has a minimal total variance *i.e.* no other estimator would have an error variance lower than the BLUE estimator.

The only unbiased linear combination between $x^b$ and $y^o$ is the following:

$$
x^a = x^b + K (y^o - H x^b)
$$

where the matrix $K$ is the called the “Kalman gain”. This matrix represents the gridding operation or “analysis” since it returns a gridded field when it is applied to a vector of observations minus the first guess. By subtracting the true state $x^t$ from this equation, one can derive how an error on the first guess or on the observations affects the analysis:

$$
\eta^a = \eta^b + K (\epsilon - H \eta^b) = (I - KH) \eta^b + K \epsilon
$$

The error covariance of the analysis which we want to minimize if given by:

$$
P^a (K) = E[\eta^a \eta^a^T] = (I - K H) P^b (I - K H)^T + KRK^T
$$

The error covariance of the analysis depends on the Kalman gain $K$. The total error variance of the analysis is the trace of matrix $P^a$.

$$
\text{tr} (P^a (K)) = \text{tr} (P^b) + \text{tr} (K H P^b H^T K^T) - 2 \text{tr} (P^b H^T K^T) + \text{tr} (KRK^T)
$$

If the total variance is at its minimum, then a small increment of the gain $\delta K$ does not modify the value of the total variance in the first order of $\delta K$. 

12
\[
\text{tr} \left( \mathbf{P}^a (\mathbf{K} + \delta \mathbf{K}) \right) - \text{tr} \left( \mathbf{P}^a (\mathbf{K}) \right) = 0
\]
\[
= 2 \text{tr} \left( \mathbf{K} \mathbf{H} \mathbf{P}^b \mathbf{H}^T \delta \mathbf{K}^T \right) - 2 \text{tr} \left( \mathbf{P}^b \mathbf{H}^T \delta \mathbf{K}^T \right) + 2 \text{tr} \left( \mathbf{K} \mathbf{R} \delta \mathbf{K}^T \right)
\]
\[
= 2 \text{tr} \left( \left[ \mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R} \right] - \mathbf{P}^b \mathbf{H}^T \right) \delta \mathbf{K}^T \]

Since the perturbation of the Kalman gain \( \delta \mathbf{K} \) is arbitrary, the expression in brackets must be zero. The optimal gain, or the Kalman gain, is thus:

\[
\mathbf{K} = \mathbf{P}^b \mathbf{H}^T \left( \mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R} \right)^{-1}
\]

The error covariance of the BLUE estimator is thus:

\[
\mathbf{P}^a = \mathbf{P}^b - \mathbf{K} \mathbf{H} \mathbf{P}^b
\]
\[
= \mathbf{P}^b - \mathbf{P}^b \mathbf{H}^T \left( \mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R} \right)^{-1} \mathbf{H} \mathbf{P}^b
\]

In practice, often only the error variance of the analysis (\( i.e. \) the diagonal elements of \( \mathbf{P}^a \)) is computed. As an illustration, the field obtained by optimal interpolation is shown in figure 1.11. Unlike Cressman weighting, optimal interpolation allows to compute the relative error variance (figure 1.12) of the analysed field. This relative error is bounded between 0 and 1 because it is scaled by the error variance of the first guess.
1.7 Variational Inverse Method

The variational inverse method (VIM) or variational analysis (Brasseur et al., 1996) aims to find a smooth field which is close to the observed values. A cost function $J$ is defined which penalizes a field which is far away from the observations and which is not smooth. The optimal field is the one which minimizes this cost function.

$$J[\varphi] = N_d \sum_{j=1}^{Nd} \mu_j [d_j - \varphi(x_j, y_j)]^2 + \|\varphi - \varphi_b\|^2$$

(1.28)

The first term is similar to an RMS error between the field and the observations. Each observation is weighted by a coefficient $\mu_j$. Those coefficients can be constant if all observations are believed to be of comparable accuracy. But if for some reason, some observations are expected to be more accurate than others, it can be taken into account by using a variable $\mu_j$. The second term is a measure of non-smoothness.

**Exercise 1**

*Why is it not possible to consider only the first term of equation (1.28). Why is it necessary to include a constraint about the smoothness of the field?*

The degree of smoothness is quantified with the first and second derivatives of the field. For a slowly varying field those derivative are small.

$$\|\varphi\| = \int_D (\alpha_2 \nabla \nabla \varphi : \nabla \nabla \varphi + \alpha_1 \nabla \varphi \cdot \nabla \varphi + \alpha_0 \varphi^2) dD$$

(1.29)
The operator $\mathbf{A} : \mathbf{B}$ represents the sum over the element-wise product between $\mathbf{A}$ and $\mathbf{B}$.

$$\mathbf{A} : \mathbf{B} = \sum_{ij} A_{ij} B_{ij}$$  \hspace{1cm} (1.30)

The term $\nabla \nabla \varphi : \nabla \nabla \varphi$ of the smoothness constraint has thus the following form in e.g. two dimensions:

$$\nabla \nabla \varphi : \nabla \nabla \varphi = \sum_{i,j=1}^{2} \left( \frac{\partial^2 \varphi}{\partial x_i \partial x_j} \right)^2$$ \hspace{1cm} (1.31)

$$= \left( \frac{\partial^2 \varphi}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 \varphi}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 \varphi}{\partial y^2} \right)^2$$ \hspace{1cm} (1.32)

The parameters used in the cost functions are attributed to different constraints:

- $\alpha_0$ penalizes the field itself (anomalies)
- $\alpha_1$ penalizes the gradients (no trends)
- $\alpha_2$ penalizes the variability (regularization)
- $\mu$ penalizes the data-analysis misfits (objective)

Without loss of generality we can choose $\alpha_2 = 1$. The minimization is actually performed by a Finite-Element Method, hence the need for generating a finite-element grid (figure 1.13).

![Figure 1.13: Example of the triangular mesh](image)

Because the field is only defined in the sea, the minimization also works only within the contours defining the "coastline".
1.7.1 Equivalence between OI and variational analysis

Using the notation introduced previously, the minimization problem can be rewritten using the state vector \( \mathbf{x} \). The elements of the vector \( \mathbf{x} \) are the values \( \varphi_{ij} \) of the field at discrete locations of the grid with a resolution of \( \Delta x \) and \( \Delta y \).

\[
\varphi_{ij} = \varphi(i \Delta x, j \Delta y) \tag{1.33}
\]

The derivatives are replaced by finite differences. For example the derivative in \( x \) is approximated by:

\[
\frac{\partial \varphi}{\partial x} \left( (i + \frac{1}{2}) \Delta x, j \Delta y \right) \sim \frac{\varphi_{i+1,j} - \varphi_{ij}}{\Delta x} \tag{1.34}
\]

This derivative can be also expressed as a matrix containing mostly zeros and \( \pm \frac{1}{\Delta x} \). In fact all derivatives in 1.29 can be written as matrices applied to the state vector \( \mathbf{x} \).

For a field given at discrete locations, the integral in 1.29 can be approximated by a sum over all elements. For example, the last term becomes,

\[
\int_D \varphi^2 \, dD \sim \frac{1}{\Delta x \Delta y} \sum_{ij} \varphi_{ij}^2 = \frac{1}{\Delta x \Delta y} \mathbf{x}^T \mathbf{x} \tag{1.35}
\]

After discretizing equations 1.28 and 1.29, the cost function \( J \) can be written as:

\[
J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^T \mathbf{W}^b (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y}^o - \mathbf{H} \mathbf{x})^T \mathbf{W}^o (\mathbf{y}^o - \mathbf{H} \mathbf{x}) \tag{1.36}
\]

where the matrix \( \mathbf{W}^b \) is directly related to equation 1.29 and to the parameters \( \alpha_0, \alpha_1 \) and \( \alpha_2 \) and where \( \mathbf{W}^o \) is a diagonal matrix whose diagonal elements are \( \mu_j \).

One can show that this cost function \( J \) is at its minimum for \( \mathbf{x} = \mathbf{x}^a \) given by

\[
\mathbf{x}^a = \mathbf{x}^b + (\mathbf{W}^b + \mathbf{H}^T \mathbf{W}^o \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W}^o (\mathbf{y}^o - \mathbf{H} \mathbf{x}^b) \tag{1.37}
\]

Using the Sherman-Morrison-Woodbury formula (see appendix A), one obtains

\[
\mathbf{x}^a = \mathbf{x}^b + \mathbf{W}^b^{-1} \mathbf{H}^T \left( \mathbf{H} \mathbf{W}^b^{-1} \mathbf{H}^T + \mathbf{W}^o^{-1} \right)^{-1} (\mathbf{y}^o - \mathbf{H} \mathbf{x}^b) \tag{1.38}
\]

which is identical to the optimal interpolation analysis if the weights are defined as the inverse of the covariance matrix used in the optimal interpolation analysis.

\[
\mathbf{W}^b^{-1} = \mathbf{P}^b \tag{1.39}
\]

\[
\mathbf{W}^o^{-1} = \mathbf{R} \tag{1.40}
\]

These results allow us to interpret the inverse of the weights in terms of error covariances. An advantage of the variational inverse method over optimal interpolation is that boundary effects can be easily taken into account. On the other hand, the users have more flexibility in choosing the background error covariance with optimal interpolation.

1.7.2 Weights on data

The weights \( \mu \) on the data are related to the signal-to-noise ratio \( \sigma^2/\epsilon^2 \) and the correlation length \( L \) (Brankart and Brasseur, 1996):

\[
\mu = \frac{\sigma^2 4\pi}{\epsilon^2 L^2} \tag{1.41}
\]
1.7.3 Correlation function

In variational analysis, the correlation function is not specified \textit{a priori}, but its form is determined by the differential in equation (1.28). For an infinite domain, one can show that for $\alpha_0 = L^{-4}$ and $\alpha_1 = 2L^{-2}$ the correlation function is given by (Brasseur \textit{et al.}, 1996):

$$C(r) = \frac{r}{L} K_1 \left( \frac{r}{L} \right)$$

where $r$ is the Euclidean distance, $L$ is the correlation length and $K_1$ is the modified Bessel function (Abramowitz and Stegun, 1964).

1.7.4 Error fields by analogy with optimal interpolation

Error fields can be derived since:

- The optimal interpolation analysis is equivalent to VIM if the weights used by VIM are the inverse of the error covariances used by optimal interpolation.

- In the context of optimal interpolation, we showed that the error field equals analysis of the covariance fields:

$$P^a = P^b - KHP^b$$

- $\Rightarrow$: error field of VIM equals analysis (by VIM) of covariance fields (the input of the analysis tool is a vector containing the covariance of data points with the point in which the error estimate is to be calculated)

Hybrid approach: provide covariance field from analytical kernel function rather than actual covariance. Error calculation benefits from the already performed matrix inversion.

1.7.5 Diva

\textsc{Diva} is a program written in Fortran and Shell scripts performing a variational analysis. This program is available at http://modb.oce.ulg.ac.be/modb/diva.html. Its features include:

- graphical interface or command line
- 3D field are computed by performing a succession of 2D analyses
- generalized cross-validation for parameter estimation
- Land and topography are taken into account.

1.7.6 Example

The analysed field computed by \textsc{Diva} for the observations of figure 1.4 is shown in figure 1.14. The results are similar to the field obtained by optimal interpolation. But now, it can be seen that the presence of the land-sea boundary is taken into account and that the observations on one side of the peninsula do not affect the temperature on the other side.
An estimation of the relative error is shown in figure 1.15. As expected, the error is lowest near the observations and tends to 1 far away from them. Note that the presence of the barrier also impacts the estimation of error variance. The error estimate provided by optimal interpolation and variational inverse method should be interpreted cautiously since we supposed that the used error covariances are correct.

Figure 1.14: Field reconstructed using Diva implementing the variational inverse method
1.8 Comparison

<table>
<thead>
<tr>
<th>Method</th>
<th>min($\epsilon^2$)</th>
<th>3D</th>
<th>Multivar</th>
<th>Ops/image</th>
<th>$\epsilon(r)$</th>
<th>a priori</th>
<th>C.V.</th>
<th>anisotropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cressman</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>$N_d N_a$</td>
<td>$w(r/L)$</td>
<td>(L)</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>O.I.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>$N_a^2 + N_d N_a$</td>
<td>$c(r/L)$</td>
<td>$L, \sigma^2/\mu^2$</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>V.I.M.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>$N_a^{3/2}$</td>
<td>*</td>
<td>$K(r/L)$</td>
<td>$L, \sigma^2/\mu^2$</td>
<td>*</td>
</tr>
<tr>
<td>DINEOF</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>$N_a^{5/4}$</td>
<td>*</td>
<td>stat.</td>
<td>$N$</td>
<td>*</td>
</tr>
</tbody>
</table>

$N_d$ : number of data points  
$N_a$ : number of grid points for analysis  
$N$ : number of EOFs  
$L$ : correlation length  
$\sigma^2/\epsilon^2$ : signal-to-noise ratio  
* : available feature  
(•) : available with some adaptations

1.9 Summary

- A first guess is needed  
- Observations have errors (in particular the observations do not necessarily represent what we want)  
- Error covariances are used to quantify the magnitude of the errors
• gridded field of **minimum expected error** → optimal interpolation

• **smooth** gridded field close to observations → variational inverse method

• **Both approaches are actually the same** (but some things are more easy to do in one approach than in the other)
Chapter 2

Modelling of error covariances

“Garbage in, garbage out”

2.1 Background error covariance matrix

The error covariance needs to be specified for optimal interpolation and variational analysis. It plays a crucial role in the accuracy of the output. A common test in the gridding problem is to apply the method for a single isolated data point. The resulting field is proportional to the error covariance of the background:

\[ x_a - x_b = P^b H^T (H P^b H^T + R)^{-1} (y^o - H x^b) \]  

simple scalar coefficient  

(2.1)

2.1.1 Parametric error covariance

In most optimal interpolation approaches, the error covariance is explicitly modelled based on the decomposition between variance and correlation:

\[ P = D C D \]  

(2.2)

For simplicity, the variance of the background field is often assumed constant. In this case, the matrix \( D \) is proportional to the identity matrix. The correlation matrix \( C \) is parameterized using analytical functions. Some basic characteristics are common to most correlation functions:

- The correlation is essentially a function of distance
- The correlation is one for zero distance (these are the diagonal elements of \( C \))
- Generally, the correlation decreases if distance increases

The correlation function is often a Gaussian function or a Gaussian function multiplied by a polynome. The length-scale of the function is an important parameter since it determines over which distance a data point can be extrapolated. This length-scale is the typical scale of the processes in the field. In the ocean, mesoscale processes have a length-scale of the order of the radius of deformation (figure 2.1),

\[ R_d = \frac{N H}{f} \]  

(2.3)

where \( N \) is the scale of Brunt-Väisälä frequency (a measure of the stability of the water column), \( H \) is the height scale over which the Brunt-Väisälä frequency is significantly different from zero, and \( f \) is the Coriolis parameter at some latitude. Biochemical tracers advected by ocean currents will have also this typical length-scale.
Figure 2.1: Gulf Stream and eddy field as measured by satellite altimetry the 28 January 2008 (units are m). The diameter of these eddies are related to the radius of deformation.

Since this length-scale does depend on stratification, it can vary from one location to another.

Modelling directly the error covariance matrix offers numerous “tweaks” such that the error covariance reflects better our expectations on how variables are correlated. The effect of island and barriers can be taken into account by imposing that the error covariance between two location is zero if a straight line connecting them intersects a land point. Additional fields can be used model to the error covariance. For example, satellite altimetry gives an indication about which water masses are likely to have similar properties.

\[
P(r_1, r_2) = e^{-\left(\zeta(r_1)-\zeta(r_2)\right)^2/L^2} e^{-|r_1-r_2|^2/L^2}
\]  

(2.4)

The impact of an observation is thus limited to the structures of similar sea surface height. The impact of the sea surface height on the covariance for a realistic example is shown in figure 2.2.
2.1.2 Ensemble error covariance

If we have a large amount of data, we can try to estimate the error covariance from a ensemble of realizations of the data, in particular, from a long time-series of the data. If the columns of $X$ contain $N$ different realizations (minus the mean state), the covariance can be estimated by

$$P = \frac{1}{N}XX^T$$

(2.5)

Also used in data assimilation, since different realizations of the model (with slightly different parameters and forcings) can be used to construct an ensemble. The columns of $X$ can also be empirical orthogonal functions (EOFs) scaled by their singular value. This is related to the way the technique DINEOF fills missing data (Beckers and Rixen, 2003; Alvera-Azcárate et al., 2005).

2.1.3 Error covariance based on differential operators

For variational inverse methods, the concept of error covariance is implicit. Instead of modelling the error covariance directly, the variational approach models its inverse with the help of a differential operator. Additional constraints can be included to modify the implicit covariance function. Sometimes an advection constraint is used which imposes that the field is a stationary solution to the advection diffusion equation.

$$\frac{\partial \varphi}{\partial t} = \mathbf{u} \cdot \nabla \varphi - \mathbf{A} \nabla \cdot \nabla \varphi = 0$$

(2.6)

where $\mathbf{A}$ is a diffusion coefficient and $\mathbf{u}$ is the velocity field.

The advection constraint aims thus at modeling the effects of velocity field on the reconstructed field. Such as constraint is introduced by adding a term to the cost function (1.28), leading to

$$\tilde{J} = J(\varphi) + \theta \int_D [\mathbf{u} \cdot \nabla \varphi - \mathbf{A} \nabla \cdot \nabla \varphi]^2 dD$$

(2.7)
The effect of this constraint is to propagate the information along the direction of the flow (figure 2.3). This option is available in Diva.

Figure 2.3: (a) Flow field. (b) Correlation modified by flow field.

2.2 Observation error covariance

The observations are often assumed to be uncorrelated. The observation error covariance $R$ is in this case a diagonal matrix. This maybe a valid assumption for the instrumental error, but it is not in general for the representativity error since this error is spatially correlated. However, is it complicated to use a non-diagonal matrix for the observations error covariance since it must be inverted. Its inverse is also ill-conditioned for long-spatial correlation. There are some procedure to deal with correlated observations:

- Binning of observations: instead of working with all observations, those which are close are averaged into bins and they are treated as a single observation. Correlation between binned observations is thus effectively reduced and can be ignored. The downside of this procedure is that the spatial resolution of the data is also reduced and that small-scale structures visible in the original data are not present in the binned observations.

- Inflating the error variance. The spatial correlation observation errors reduce then the weight of the observations in the analysis. Observations with strongly correlation errors are indeed redundant. This can also be taken into account by using a diagonal observation error covariance matrix with increased error variance.

- Lower the dimensionality of the problem. In oceanography, most in situ observations have a high vertical and/or temporal resolution (e.g. vertical profiles, moorings) and a low horizontal resolution. Thus the error correlation of the observations is important in time and vertical dimension. However, by performing 2D-analysis of vertical levels and time instances independently of vertically and temporally interpolated (or averaged similar to the binning approach), the correlation in those dimensions does not play a role in this case.

2.3 Summary

- the more realistic the error covariances are the better the analysis will be
- in optimal interpolation, the error covariances are often analytical functions
• in variational inverse method, the error covariance are implicitly defined. The cost function penalize abrupt changes of the field and large deviation from the first guess

• Key parameters:
  • correlation length of the error of the first guess
  • error variance of the observation and error variance of the first guess
  • their ratio is called signal-to-noise ratio
Chapter 3

Parameter optimization and cross-validation

Optimal interpolation and the variational inverse method are only optimal if their underlying error covariance matrices are known. Even after one particular class of error covariance matrices has been adopted, several parameters have to be determined such as the correlation length and the signal-to-noise ratio. The optimality of any method is obviously lost if some parameters are wrong.

3.1 A priori parameter optimization

By analysing the difference between the background estimate and the observations, useful information on their statistics can be derived. Since the errors of the background estimate and the errors of the observations are supposed to be independent, one can show that the covariance of their difference is the sum of two covariance matrices:

$$E[(y^o - Hx^b)(y^o - Hx^b)^T] = HPH^T + R$$

This relationship can be used to estimate some parameters of the error statistics if the following assumptions are true:

- The error variance of the background is everywhere the same ($\sigma^2$)
- The correlation of the background error between the points $r_i$ and $r_j$ depends only on the distance between them.
- The observation errors have all the same variance ($\epsilon^2$) and are uncorrelated.

The matrix $HPH^T$ is the error covariance matrix of the background at the location of the observations. This term can be calculated by evaluating the covariance at the location of the observations $r_i$:

$$(HPH^T)_{ij} = \sigma^2 C(|r_i - r_j|; L)$$

where $C$ is the correlation function depending only on the distance between two observations and on the correlation length-scale $L$.

The operator $E[\cdot]$ in equation (3.1) is the expected value. Here we approximate the expected value by averaging over a sample of couples of observations (or all possible couples of observations). The couples are binned according to their distance. The covariance is computed by averaging over all couples in a given bin. If this procedure is repeated for all bins, one obtains the covariance as a function of distance. In figure 3.1, the analytical function is fitted to obtain the correlation length-scale $L$ and the error variance $\sigma^2$. The first data bin with zero radial distance is not taken into account for the fitting since it includes the observation error variance $\epsilon^2$. The variance in the first bin can be used to estimate the error variance $\epsilon^2$. 
Figure 3.1: Determination of the correlation length-scale and error covariance of the background. This approach is realized with the tool \textit{divafit}.

### 3.2 Validation

The validation consists in using an independent data set to determine the accuracy of the results. The data sets maybe some observations explicitly set aside for the validation purposes or some newly available data. One can use for example satellite SST to validate a climatology obtained by \textit{in situ} data.

Some poorly known parameters can be varied to minimize the RMS error between the analysis and the validation data set. This provides a mean to objectively calibrate some parameters. It is crucial for this step that the validation data set is independent from the observation used in the analysis.

#### Exercise 2

What would be the “optimal” signal-to-noise ratio if the analysed observations would include the validation data set (in the context of optimal interpolation or variational inverse method and all other parameters are kept constant)?

### 3.3 Cross-validation

In the cross-validation approach, one data point $d_i$ is excluded from the data-set and the analysis is performed without it. The analyzed field is interpolated at the location of the excluded point. This interpolated value $\tilde{d}_i$ is thus not influenced by the excluded data point $d_i$. Their difference $\theta_i$ give thus an indication of the accuracy of the field at the location $i$:

$$\theta_i = d_i - \tilde{d}_i$$  

This procedure is repeated for all observations $d_i$ for $1 \leq i \leq N_d$. The RMS error between $d_i$ and $\tilde{d}_i$ is called the cross-validator $\Theta$:  

\[27\]
\[ \Theta^2 = \frac{1}{N_d} \sum_{i=1}^{N_d} \hat{\theta}_i^2 \]  

(3.4)

\( \Theta \) is a global measure of the accuracy of the analysed field. However, this procedure would require \( N_d \) additional analysis. For most applications, this is too prohibitive in terms of calculation. Craven and Wahba (1979) showed that it is not necessary to discard the data points during the analysis if we modify the error estimate as follows

\[ \hat{\theta}_i^2 = \frac{(d_i - \tilde{d}_i)^2}{(1 - A_{ii})^2} \]  

(3.5)

where \( \tilde{d}_i \) is the analysis field at location \( i \) using all data and where \( A_{ii} \) is the diagonal term \( i \) of matrix \( A \) defined by:

\[ A = HK = HP^T (HP^T + R)^{-1} \]  

(3.6)

This result is surprising because it allows to compute the value of a field obtained without a particular observation without actually discarding it. This result is also know as the missing data lemma.

Basically the denominator penalizes more heavily data points in which the analysis is forced to be close to the data and accounts therefore for the self-influence of the data points (which is absent in the case of pure cross-validation).

If a large number of observations are available, then a different approach can be adopted to estimate the error of the analysis:

- A small subset of randomly chosen observations are set aside (the cross-validation data set).
- The analysis is performed without the cross-validation data set.
- The result is compared to the observations initially set aside and an RMS error is computed.
- The two last steps are repeated to optimize certain parameters of the analysis.
- Once the optimal parameters are derived a final analysis is carried out using all data.

The size of the cross-validation data set has to be sufficiently large so that the error estimate is robust but at the same time its size has to be a negligible fraction of the total data set (typically 1%). This explains why a large amount of data is necessary. This approach is used in DINEOF to determine the optimal number of EOFs.

### 3.4 Generalized cross-validation

In the case of optimal interpolation, the gain matrix \( K \) is formed explicitly. Thus, the diagonal elements of \( A = HK \) can be easily computed. However, in variational analysis, the gain matrix or the matrix \( A \) is never explicitly calculated. This means that we can compute product between \( A \) and any vector \( z \), but we cannot directly extract any element of the matrix \( A \). If one would want to compute the diagonal elements of this matrix explicitly, one would need to perform \( N_d \) additional analysis (and thus negating the benefit of the missing data lemma). But since the aim of the general cross-validator is to provide an overall accuracy of the field, the elements \( A_{ii} \) are approximated by their spatial average (Brankart and Brasseur, 1996):

\[ A_{ii} \sim \bar{A} = \frac{1}{N_d} \sum_{i=1}^{N_d} A_{ii} \]  

(3.7)

In Diva \( \bar{A} \) is estimated using an approach similar to the one of Girard (1989),

\[ \bar{A} = \frac{z^T A z}{z^T z} \]  

(3.8)
where \( z \) is a vector of random variables of zero mean. The formula can be understood as follows: for random components of \( z \), the off-diagonal components of \( A \) will be summed up with random signs (because of terms \( z_i A_{ij} z_j \)), with an expected mean value of zero. On the other hand, the diagonal terms will always add up with positive weights \( z_i^2 \), so that when dividing by the sum of the weights, we retrieve a good estimate of the trace. For robustness, the trace estimate can be repeated several times (with different random vectors), averaging the different estimates.

Finally, the cross-validator is computed using \( \bar{A} \), by

\[
\hat{\theta}_i^2 = \frac{(d_i - \bar{d}_i)^2}{(1 - \bar{A})^2} \Theta^2 = \frac{1}{N_d} \sum_{i=1}^{N_d} \theta_i^2
\]  

(3.9)

An estimate of the global accuracy of an analysis is already an useful information in itself. But the ability to calculate this value allows to optimize the parameters such as the signal-to-noise ratio. Indeed, by performing the analysis with several set of parameters, one can compute the cross-validator for each analysis. The set of parameters producing the lowest cross-validator is thus the optimal one.

3.5 Summary

- Don’t skip the validation of your results
- If you don’t have an additional dataset, Cross-validation allows you to estimate the error of your analysis
  - set some date aside
  - do the analysis
  - compute the RMS error between the analysis and the data set aside
- By minimizing the RMS error, you can estimate some parameters of your analysis
Chapter 4

Real world example

The following shows an application of Diva to compute three-dimensional fields of temperature and salinity in the North Atlantic by Troupin et al. (2008).

4.1 Data

Data were gathered from various databases in the region $0 - 60^\circ N \times 0 - 50^\circ W$. We processed them to remove duplicates, detect outliers and perform vertical interpolation with Weighted Parabolas method (Reiniger and Ross, 1968).
Figure 4.1: Localisation of temperature observations at 200 m and 3500 m in winter and summer.

4.2 Finite-element mesh

Resolution of the minimization problem relies on a highly optimized finite-element technique, which permits computational efficiency independent on the data number and the consideration of real boundaries (coastlines and bottom).
4.3 Results

Our results are compared with 1/4° climatology from the World Ocean Atlas 2001 (WOA01 in the following, (Boyer et al., 2005))
Figure 4.3: Comparison of analysed temperature field with Diva (left) and WOA01.
Appendix A

Some matrix identities

For all symmetric matrices $A$ and $C$ and for all matrices $B$ of conforming dimension, we have:

$$C^{-1} - C^{-1}B^T (BC^{-1}B^T - A)^{-1} BC^{-1} = (C - B^T A^{-1}B)^{-1} \tag{A.1}$$

This is a special case of the Sherman-Morrison-Woodbury formula. From this last equation, one can show that:

$$(A^{-1} + B^T C^{-1}B)^{-1} B^T C^{-1} = AB^T (C + BAB^T)^{-1} \tag{A.2}$$

Instead of performing the inverse in space of matrix $A$ the inverse is done in space of matrix $C$. This equation was used to prove that the variational inverse method is equivalent to the optimal interpolation analysis. The Kalman gain can be computed by:

$$K = P^b H^T (HP^b H^T + R)^{-1} \tag{A.3}$$

$$= \left( P^{b^{-1}} + H^T R^{-1} H \right)^{-1} H^T R^{-1} \tag{A.4}$$

Optimal interpolation uses the first expression and variational methods use the second approach, but both are mathematically equivalent.

Acknowledgements

- The National Fund of Scientific Research (F.R.S.-FNRS, Belgium)
- SeaDataNet project
- SESAME, Southern European Seas: Assessing and Modelling Ecosystem changes
- ECOOP, European COastal-shelf sea OPerational observing and forecasting system
Bibliography


