Generation of analysis and consistent error fields using the Data Interpolating Variational Analysis (Diva)

C. Troupin[a,∗], A. Barth[a,b], D. Sirjacobs[c], M. Ouberdous[a], J.-M. Brankart[d], P. Brasseur[d], M. Rixen[e], A. Alvera-Azcárate[a], M. Belounis[a], A. Capet[a], F. Lenartz[a], M.-E Toussaint[a], J.-M. Beckers[a,f]

[a]*GeoHydrodynamics and Environment Research, MARE, University of Liège, BELGIUM*
[b]*Postdoctoral Researcher, F.R.S.-FNRS, BELGIUM*
[c]*Algology, Mycology and Experimental Systematics, Department of Life Sciences, University of Liège, BELGIUM*
[d]*LEGI - CNRS - Université de Grenoble, FRANCE*
[e]*NURC - NATO Undersea Research Centre, La Spezia, ITALY*
[f]*Honorary Research Associate, F.R.S.-FNRS, BELGIUM*

Abstract

The Data Interpolating Variational Analysis (Diva) is a method designed to interpolate irregularly-spaced, noisy data onto any desired location, in most cases on regular grids. It is the combination of a particular methodology, based on the minimisation of a cost function, and a numerically efficient method, based on a finite-element solver. The cost function penalises the misfit between the observations and the reconstructed field, as well as the regularity or smoothness of the field. The method bears similarities to the smoothing splines, where the second derivatives of the field are also penalised.

The intrinsic advantages of the method are its natural way to take into account topographic and dynamic constraints (coasts, advection, . . .) and its capacity to handle large data sets, frequently encountered in oceanography. The method provides gridded fields in two dimensions, usually in horizontal layers. Three-dimension fields are obtained by stacking horizontal layers.

In the present work, we summarize the background of the method and describe the possible methods to compute the error field associated to the analysis. In particular, we present new developments leading to a more consistent error estimation, by determining numerically the real covariance function in Diva, which is never formulated explicitly, contrarily to Optimal Interpolation. The real covariance function is obtained by two concurrent executions of Diva, the first providing the covariance for the second. With this improvement, the error field is now perfectly consistent with the inherent background covariance in all cases.

A two-dimension application using salinity measurements in the Mediterranean Sea is presented. Applied on these measurements, Optimal Interpolation and Diva provided very similar gridded fields (correlation: 98.6%, RMS of the difference: 0.02). The method using the real covariance produces an error field similar to the one of OI, except in the coastal areas.

Keywords: Variational analysis, Inverse method, Spatial interpolation, Mediterranean Sea, Optimal Interpolation
PACS: 92.05.Hj, 93.30.Rp, 93.85.Bc

1. Introduction

The generation of gridded fields from non-uniformly distributed observations (both in space and time) is a frequent concern in geosciences. Similarly to Ooyama (1987), we will refer to an *analysis* or *analysed field* as "the estimation of a continuous spatial field of a given variable from a set of discrete measurements". The range of applications is wide, going from model initializations to validation exercises or simple plotting purposes.

Optimal interpolation (noted OI from here on, e.g., von Storch and Zwiers, 1999; Chilès and Delfiner, 1999) is a popular analysis tool, owing to its ease of use and the error field associated to the analysis (e.g., Shen et al., 1998; Kaplan et al., 2000). The idea is to minimise the expected error variance. However, the method does not always produce this theoretical optimum, specially when the number of data is not sufficient and the covariances are not correctly specified (e.g., Rixen et al., 2000; Gomis et al., 2001).

The quality of OI and other gridding techniques relies on the correct specification of the covariances of the observational error and of the background field. The covariance functions used in OI are not restricted (except that the covariance matrices have to be positive-definite and symmetric), allowing for example correlated observational errors. Yet in most cases, covariances between two points are parametrized by simple expressions, such as a Gaussian function depending on the sole distance between the points. Adaptations of the OI scheme allow the use of anisotropic functions (e.g., Tandeo et al., 2011).

In ocean numerical modelling, most of the climatologies used for model initial or boundary conditions are still generated using basic interpolation methods. For instance, the various versions of the World Ocean Atlas (e.g., Locarnini et al., 2006, 2010) are based on distance-weighting methods (with dif-

---

∗Corresponding author. Tel.:+32 43662340
Email address: ctroupin@ulg.ac.be (C. Troupin)

ferent length scales) following the scheme of Barnes (1964), the Northwest European shelf climatology (Berx and Hughes, 2009) is based on the Delaunay triangulation, and the Generalized Digital Environmental Model (GDEM, Teague et al., 1990), is based on the determination of analytical curves that represent the mean vertical distributions of temperature and salinity for grid squares, with the help of individual profiles for the determination of the coefficients of the mathematical expressions for the curves. Only the Polar Science Center Hydrographic Climatology (PHC, Steele et al., 2001) employs OI for the creation of the gridded fields.

The method presented here, Diva (Data Interpolating Variational Analysis), generates analysed fields similar to those obtained with OI, but without formally expressing the covariance functions. It is based on the Variational Inverse Method (Brasseur and Haus, 1991; Brasseur et al., 1996), which consists in the minimisation of a cost function measuring the data-analysis misfit and the regularity (or roughness) of the reconstructed field. The variational analysis is also considered as a spline interpolation method (Wahba, 1975; Wahba and Wendelberger, 1980), as we will show in Section 2.

Diva can be defined as a specific algorithm that performs OI in a efficient way (due to its finite-element (FE) solver, Section 2.5) and for particular forms of the covariance functions, allowing the consideration of anisotropies and decorrelation introduced by coastlines or frontal structures. Particularly interesting is that, under some circumstances, OI and Diva methods are equivalent (e.g., McIntosh, 1990; Bennett, 1992).

While in OI, the covariance functions are generally parametrized, the covariances employed in Diva are not known explicitly, except in some particular cases. This can lead to error fields that are not consistent with the analysis. For this reason, we developed a new method to access the real covariance function, and thereby improve the error field.

The paper aims to describe the Diva method, present the techniques to produce the error fields and examine the differences with OI. The structure is as follows: in Section 2, we summarise the mathematical formulation of the method. The OI method and its theoretical equivalence with Diva is examined in Section 3. Section 4 is focused on the determination of the analysis parameters. The methods for computing the error field are described in Section 5. Finally, Section 6 presents a realistic application illustrating the theoretical aspects developed in the previous sections.

2. Diva method

In the next developments, we will consider that we are working with data anomalies: a reference (or background) field $\varphi_b(x, y)$ is subtracted from the data prior any analysis. The background field provides the value of the analysis in regions void of observations. Hence this field has to be properly defined by using the information contained in the set of observations, for example:

- the mean value of the data: $\varphi_b = $ constant.

- the linear fit of the data, obtained by linear regression: $\varphi_b = ax + by + c$, where (x,y) are the horizontal coordinates and a, b, c are constant coefficients.

Once the analysis is performed, the reference field is added in order to return to the original variable.

The idea behind Diva is to find an analysis field φ that will satisfy a set of constraints over a domain Ω, expressed in the form of a cost function. This cost function is made up of:

1. the distance between analysis and data (*observation constraint*),
2. the regularity of the analysis (*smoothness constraint*),
3. physical laws (*behaviour constraint*).

If only the observation constraint were considered, the result would be a pure interpolation, i.e., the analysis would contain all the data points. This is not suitable in the case of atmosphere or ocean observations because of the noise that affects them (measurement errors, but also *representativity* errors). In the case of hydrographic climatologies, none of the measurements is exactly representative climatic state. This is why other constraints have to be added. Such an approach was also used by Legler et al. (1989) for the analysis of pseudo-stress. Along with the three aforementioned constraints, they also included the divergence and the curl of the wind field.

2.1. Formulation

We will start by considering the observation and smoothness constraints. Considering a series of N data anomalies d_i at locations (x_i, y_i), the cost function reads, in Cartesian coordinates:

$$
\begin{aligned}
J[\varphi] \quad = \quad & \sum_{i=1}^{N} \mu_i \left[d_i - \varphi(x_i, y_i) \right]^2 \qquad (1) \\
& + \int_{\Omega} \left(\boldsymbol{\nabla\nabla}\varphi : \boldsymbol{\nabla\nabla}\varphi + \alpha_1 \boldsymbol{\nabla}\varphi \cdot \boldsymbol{\nabla}\varphi + \alpha_0 \varphi^2 \right) \mathrm{d}\Omega,
\end{aligned}
$$

where μ_i, α_0 and α_1 are coefficients that will be described in the next section. $\boldsymbol{\nabla}$ denotes horizontal gradients, and the symbol : stands for double summation[1].

The first term is a weighted sum of data-analysis misfits and is identified as the observation constraint, which brings the analysed field close to the data. The second term is a measure of the spatial variability (curvature, gradient and value) of the analysis and is identified as the smoothness constraint. Natural boundary conditions are adopted for the smoothing term (for details, see Brasseur, 1994).

The penalisation of second derivatives (in the smoothness constraint) is similar to the smoothing spline formulation (Wahba, 1975; Wahba and Wendelberger, 1980). The first term of (1) models "tensions" on a thin plate, hence the term *spline in tension* (Schweikert, 1966; Franke, 1985),

---

[1]$\boldsymbol{\nabla\nabla}\varphi : \boldsymbol{\nabla\nabla}\varphi = \sum_i \sum_j (\partial^2\varphi/\partial x_i \partial x_j)(\partial^2\varphi/\partial x_i \partial x_j)$, the generalisation of the scalar product of two vectors.

2.2. Parameter meaning

Using a characteristic length scale L, we introduc dimensional space coordinates (denoted by $\tilde{}$) as follo

- for the gradients: $\tilde{\boldsymbol{\nabla}} = L\boldsymbol{\nabla}$,

- for the domain: $\Omega = L^2\tilde{\Omega}$.

Formula (1) becomes:

$$
\begin{aligned}
\tilde{J}[\varphi] \quad = \quad & \sum_{i=1}^{N} \mu_i L^2 [d_i - \varphi(x_i, y_i)]^2 \\
+ \quad & \int_{\tilde{\Omega}} \Big(\tilde{\boldsymbol{\nabla}}\tilde{\boldsymbol{\nabla}}\varphi : \tilde{\boldsymbol{\nabla}}\tilde{\boldsymbol{\nabla}}\varphi + \alpha_1 L^2 \tilde{\boldsymbol{\nabla}}\varphi \cdot \tilde{\boldsymbol{\nabla}}\varphi + \alpha_0 L^4
\end{aligned}
$$

α_0 fixes the length scale L the first and the last term of tl gral have a similar importance:

$$
\alpha_0 L^4 = 1. \tag{3}
$$

The terms $\mu_i L^2$ fix the weights on the individual observations. If the typical misfit is represented by the observational noise standard deviation ϵ_i^2 of data point i and the integral norm representative of the background field variance σ^2, then the data weights are given by

$$
\mu_i L^2 = 4\pi \frac{\sigma^2}{\epsilon_i^2}. \tag{4}
$$

The parameters L and σ^2/ϵ_i^2 are thus directly comparable to the corresponding parameters in OI.

Finally α_1 fixes the influence of gradients:

$$
\alpha_1 L^2 = 2\xi. \tag{5}
$$

With this scaling, ξ is a non-dimensional parameter close to 1 if the gradients are to be penalized with a similar weight than the second derivatives (the default value in Diva is $\xi = 1$) and underived terms.

With the overall signal-to-noise ratio defined as

$$
\lambda = \frac{\sigma^2}{\epsilon^2}, \tag{6}
$$

with a global noise level on data $\epsilon^2 = \sum_i \epsilon_i^2 / N$, we can define relative weights on the data with noise ϵ_i^2 as

$$
w_i = \frac{\mu_i L^2}{4\pi\lambda} \quad \text{with} \quad \sum_i \frac{1}{w_i} = N. \tag{7}
$$

Instead of the original parameters μ_i, α_0 and α_1, we can therefore work with relative weights w_i on data, an overall signal-to-noise ratio λ, a length scale L and a shape parameter ξ. Note that coefficients α_0, α_1 and μ are allowed to change spatially. In view of the previous relations, this amounts to change locally the length scale: $L(x,y) = \ell(x,y)\mathcal{L}$ where $\mathcal{L}$ is the global length-scale and $\ell(x,y)$. An example using a variable correlation length is presented in Fig. 1.
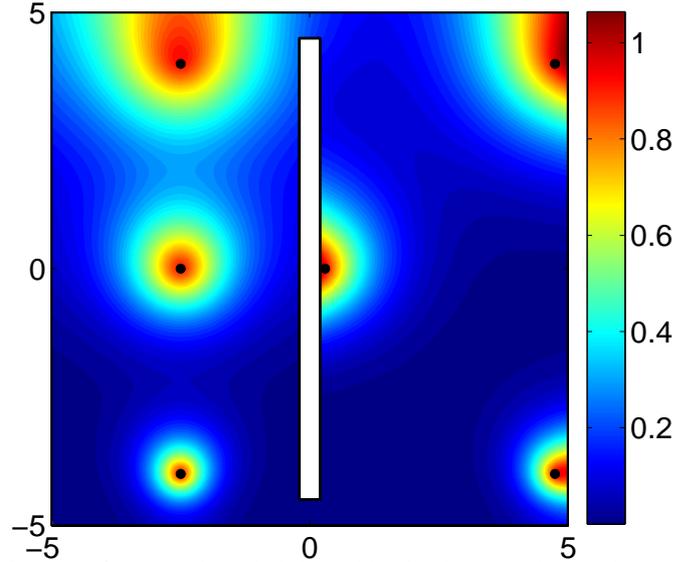

Figure 1: A few data points (black dots) with unit value are put into the domain of interest. Coastlines decouple sub-basins so that data influence does not cross land with the variational minimisation. Relative length scale increases from 0.5 to 2 in the *y*-direction.

2.3. Advection constraint

Until now, we did not consider any behaviour constrain in the cost function (1). If we assume that the field approximatively satisfies a stationary advection-diffusion equation with a velocity field $\boldsymbol{u}$, $\tilde{J}$ from (2) is augmented:

$$
\tilde{J}_a = \tilde{J}(\varphi) + \frac{\theta}{U^2} \int_{\tilde{\Omega}} \left[\boldsymbol{u} \cdot \tilde{\boldsymbol{\nabla}}\varphi - \frac{\mathcal{A}}{L} \tilde{\boldsymbol{\nabla}} \cdot \tilde{\boldsymbol{\nabla}}\varphi \right]^2 d\tilde{\Omega}, \tag{8}
$$

where U is a velocity scale computed from the field $\boldsymbol{u}$ and $\mathcal{A}$ the horizontal diffusion coefficient. At large scales, it is reasonable to assume that advection dominates diffusion: $\mathcal{A} \leq UL$ (relatively high Reynolds numbers). Otherwise for dominant diffusion, the term just adds another isotropic filtering effect, already included in the regularization term of (2). The scaling is such that for $\theta \sim 1$, the advection and smoothing constraints have a similar importance.

Zero diffusion leads to increased correlations in the direction of $\boldsymbol{u}$, which can be any vector indicating the direction in which correlation has to be increased, for example isobaths or isopycnals. Note that the vector field does not need to be divergence free. The geostrophic velocity field derived from temperature and salinity measurements can also serve as a constrain (Brasseur, 1994).

The advection constraint can be interpreted as another way to include non isotropic correlations into the analysis, again naturally and consistently with physics. Doing so in OI is also possible technically, but more difficult practically.

2.4. Kernel function

In the case of an infinite domain, the kernel of the integral term of (1) can be determined analytically for particular values of the parameters α_0 and α_1 (Brasseur, 1994).

- For $\alpha_0 L^4 = 1$, $\alpha_1 L^2 = 2$ (or $\xi = 1$), the kernel is

$$c(r) = \frac{r}{L} \mathsf{K}_1 \left(\frac{r}{L} \right), \qquad (9)$$

where r is the distance between the data point and the analysis point and K_1 the modified Bessel function of the second kind (Abramowitz and Stegun, 1964, p. 358). The existence of such an analytical solution is what motivated the choice of $\xi = 1$ in Section 2.2. This kernel is shown in Fig. 2.

- For $\alpha_0 L^4 = 1$, $\alpha_1 = 0$ (or $\xi = 0$), the kernel is

$$c(r) = -\frac{L^2}{2\pi} \mathsf{kei}_0 \left(\frac{r}{L} \right), \qquad (10)$$

where kei_0 is the Kelvin function of order zero (Abramowitz and Stegun, 1964, p. 379).

In realistic applications, the domain is bounded and the kernel cannot be expressed analytically any more.

2.5. Finite-element solution

The minimisation of (2) or (8) is solved by a FE approach (Brasseur, 1991): over each element e, the solution φ_e is expanded in terms of connector values q, which ensure the solution is continuously derivable (Brasseur, 1994), and shape functions s (3rd order polynomials), which serve to compute the field at any desired location:

$$\varphi_e = q_e^T s. \qquad (11)$$

The connector values are found by substituting (11) in the cost function (2), which yields the linear algebraic system (Brasseur (1994); Brasseur et al. (1996)):

$$\mathbf{K} q = g. \qquad (12)$$

where $\mathbf{K}$ and g are referred to as the *stiffness matrix* and the *charge vector*, respectively. In the FE theory, the stiffness matrix relates the nodal forces to the nodal displacements, while the charge vector represents the forces applied on the node of the elements (Logan, 2012). $\mathbf{K}$ is the sum of a component related to the smoothness constraint ($\mathbf{K}_s$) and a component related to the N data points:

$$\mathbf{K}_d = \sum_{i=1}^{N} \mu_i \mathbf{S}_i \mathbf{S}_i^{\mathrm{T}}, \qquad (13)$$

where $\mathbf{S}_i$ is a column vector containing the shape function values associated with a data point $d_i = d(x_i, y_i)$. $\mathbf{S}_i$ has zeros everywhere, except for the indexes of all connectors of the element in which the data point lies.

The charge vector g_i is obtained by projecting the data onto the connectors:

$$g_i = \mu_i \mathbf{S}_i d_i. \qquad (14)$$

The total charge vector g is the sum of the individual ones and provides the solution φ by inversion of (12). The size of the matrix $\mathbf{K}$ to be inverted is $N_e \times N_e$ where N_e is the number of connectors of the FE mesh and not the number of data N. Note that the data values and locations enter the formulation only through $\mathbf{K}_d$ and g. Once the connectors are computed, the solution can be computed at any location using (11).

Because the minimisation is performed only on the domain of interest Ω (the domain covered by water in the case of ocean data), sub-basins or regions separated by land are automatically decoupled (Fig. 1), since their respective elements are not physically connected. Hence covariances across obstacles are strongly decreased and an anisotropic background covariance is introduced into the analysis in a natural way, i.e., without any additional parametrisation.

3. Optimal Interpolation

Similarly to Section 2, we will consider data anomalies d with respect to a reference field.

3.1. Formulation

The OI method is formally different and is designed to lead to an analysed field of which the expected error variance is minimal. In this sense the method is optimal. The analysis φ^a at any location r of the OI is given by (e.g., Gandin, 1965; Bretherton et al., 1976; Chilès and Delfiner, 1999)

$$\varphi^a(r) = c^{\mathrm{T}}(\mathbf{B} + \mathbf{R})^{-1} \mathbf{d} = \mathbf{H} \mathbf{d}, \qquad (15)$$

where

- c is a vector containing the background covariance between the point at which the analysis is to be performed and all data points,

- $\mathbf{B}$ is the positive-definite covariance matrix of the background field between all data points,

- $\mathbf{R}$ is the positive-definite error-covariance matrix of the data.

These data-covariance matrices are statistical averages $\langle \ \rangle$ of data products:

$$\left\langle d\, d^{\mathrm{T}} \right\rangle = \mathbf{B} + \mathbf{R}, \qquad (16)$$

and are symmetric. The expected analysis error is minimal if these covariance matrices are specified correctly.

The analysis is a linear weighting of the data through the analysis matrix $\mathbf{H} = c^{\mathrm{T}}(\mathbf{B} + \mathbf{R})^{-1}$ and the inversion of a matrix of size $N \times N$ is required. Though adaptations have been made to the OI scheme to improve the numerical efficiency (e.g., Hartman and Hössjer, 2008; Zhang and Wang, 2010), here we will mostly consider the original scheme.

OI does not only provide the analysed field but also the associated error variance ϵ_a^2:

$$\begin{aligned} \epsilon_a^2(r) &= \sigma^2(r) - c^{\mathrm{T}}(\mathbf{B} + \mathbf{R})^{-1} c \\ &= \sigma^2 - \mathbf{H} c \end{aligned} \qquad (17)$$

where σ^2 is the local variance of the background field. The last equality highlights that the error is provided by the analysis of a

4

pseudo-data array containing covariances c. This interpretation will be helpful when discussing the error calculation Section 5.

The analysis $\tilde{d}$ at the original data locations is

$$\tilde{d} = \mathbf{B}(\mathbf{B} + \mathbf{R})^{-1}d = \mathbf{A}d, \qquad (18)$$

where $\mathbf{A}$ is the matrix used to perform the analysis at the data points.

3.2. Equivalence between Diva and OI

With both methods, the analysis at any point is computed as a linear combination of the data values. Diva performs OI for particular forms of the covariance functions. Though formally different, the two methods can lead to identical analyses using the same data, provided some conditions are fulfilled:

1. The observational errors are uncorrelated, leading to a diagonal matrix $\mathbf{R}$ for OI.
2. The kernel of the integral term of the Diva functional is the correlation function used by OI (e.g., McIntosh, 1990; Brankart and Brasseur, 1996).
3. The data-weights of Diva are calculated using (4).
4. OI uses the same background variance σ^2 and observational error variances ϵ_i^2.

The equivalence of the analysis allows the interpretation of Diva results from an OI point of view and provides some clear signification to the length scale L and the signal-to-noise ratio λ used in Diva. It also demonstrates that error fields for Diva can be obtained by performing an analysis of covariances according to (17), where $\mathbf{H}$ stands for the analysis operator, i.e., the Diva tool. In spite of this equivalence, the practical solution is not the same: with Diva the solution (and matrix inversion) is performed in the FE space, whereas in OI it is performed in the data space.

In realistic applications, the conditions described in Section 2.4 are not fulfilled, so the kernel of Diva is not known analytically. If the shape of the kernel is needed, it has to be estimated numerically. We will show in Section 5 that this can be done by using the Diva analysis tool itself. Intuitively since the covariance tells us how a point value is related to a remote location, we can "look" at covariances by placing a unit data point somewhere and plot the analysis value in other places (see Fig. 2).

4. Parameter estimation and statistics

In principle, OI does not need any calibration of parameters since the covariance matrices define the solution (15). However, in practice, these covariance matrices are imperfectly known and are in most cases prescribed by a priori guesses on the structure of the covariances (e.g., Bretherton et al., 1976). Typically, isotropic covariances depending on the distance between points are prescribed as functions of r/L where L is an appropriate correlation length scale, that is to say, a length scale characteristic of the considered problem. Also the level of noise is generally poorly known and enters the practical formulation as a calibration constant.

4.1. Signal-to-noise ratio

If we consider that the structure of the covariances can be fixed or prescribed in OI, but their error variance is imperfectly known, we can express:

$$\mathbf{B} = \sigma^2 \hat{\mathbf{B}}, \quad \mathbf{R} = \epsilon^2 \hat{\mathbf{R}}, \quad \frac{\langle d^{\mathrm{T}} d \rangle}{N} = \sigma^2 + \epsilon^2 \qquad (19)$$

where $\hat{\ }$ matrices are prescribed a priori and are non-dimensional, while the background variance σ^2 and the observational error variance ϵ^2 are imperfectly known. Combining relations (16) and (19), we have

$$\frac{1}{N} \mathrm{trace}\,(\mathbf{R}) = \epsilon^2 \quad \text{or} \quad \frac{1}{N} \mathrm{trace}\,(\hat{\mathbf{R}}) = 1, \qquad (20)$$

$$\frac{1}{N} \mathrm{trace}\,(\mathbf{B}) = \sigma^2 \quad \text{or} \quad \frac{1}{N} \mathrm{trace}\,(\hat{\mathbf{B}}) = 1. \qquad (21)$$

If the background and noise variances are imperfectly known, then the unknown parameter that controls the analysis is the signal-to-noise ratio λ defined in (6). Indeed, we see for example that matrix $\mathbf{A}$ (18) depends only on λ:

$$\mathbf{A}(\lambda) = \hat{\mathbf{B}}(\hat{\mathbf{B}} + \lambda^{-1}\hat{\mathbf{R}})^{-1}. \qquad (22)$$

We also note that

$$\sigma^2 = \frac{\lambda}{1+\lambda}\frac{\langle d^{\mathrm{T}} d \rangle}{N} \sim \frac{\lambda}{1+\lambda}\frac{d^{\mathrm{T}} d}{N}, \qquad (23)$$

$$\epsilon^2 = \frac{1}{1+\lambda}\frac{\langle d^{\mathrm{T}} d \rangle}{N} \sim \frac{1}{1+\lambda}\frac{d^{\mathrm{T}} d}{N}, \qquad (24)$$

so that knowing λ, we can estimate the signal and noise variances from the data values, assuming $d^{\mathrm{T}}d/N$ is a good estimate of the expected data variance. If there are enough data points with homogeneous statistics, we can replace the statistical expectations $\langle\,\rangle$ by a spatial average.

The error field (17) using the non-dimensional covariance matrices reads:

$$\epsilon_a^2(r) = \sigma^2 \hat{B}(r,r) - \sigma^2 \hat{b}^{\mathrm{T}}(\hat{\mathbf{B}} + \lambda^{-1}\hat{\mathbf{R}})^{-1}\hat{b}, \qquad (25)$$

where σ^2 is the overall background covariance and $\hat{b}$, a vector containing the non-dimensional covariance between the data points and the location for the error calculation. Note that the local variance of the background field is now $\sigma^2 \tilde{B}(r,r)$.

In Diva, the calibration parameters in the original formulation are $\mu_i, \alpha_0, \alpha_1, \xi$. They can be related to the signal-to-noise ratio and the correlation length by (3), (4) and (5) and we face the same problem as in OI regarding the calibration of the parameters.

4.2. Generalised cross-validation

If we had an independent error estimate of the OI or Diva analysis, we would optimize the value λ so as to reduce this analysis error estimate. It can be shown that the generalised cross-validator

$$\Theta^2 = \frac{\|d - \tilde{d}\|^2}{N\left(1 - \frac{1}{N}\mathrm{trace}\,(\mathbf{A})\right)^2} = \frac{\|(\mathbf{I} - \mathbf{A})d\|^2}{(1/N)\,(\mathrm{trace}\,(\mathbf{I} - \mathbf{A}))^2}, \qquad (26)$$

5

where matrix $\mathbf{A}$ is defined in (18), constitutes a global estimate of the analysis error variance. The role of the denominator is to penalize more strongly data points in which the analysis is forced to be close to the observations ($\mathbf{A}$ close to identity matrix $\mathbf{I}$) (Craven and Wahba, 1979; Golub et al., 1979) and therefore accounts for the self-influence of the data point.

Craven and Wahba (1979) showed that, at the minimum, Θ^2 is the sum of the average analysis error variance ϵ_a^2 and the data noise:

$$\Theta^2 = \epsilon_a^2 + \epsilon^2 \qquad (27)$$

When the observational errors are uncorrelated but vary in space, the residual measure $r = (\boldsymbol{d} - \tilde{\boldsymbol{d}})^{\mathrm{T}}(\boldsymbol{d} - \tilde{\boldsymbol{d}})$ should include the relative noise level and the measure r should be replaced by

$$r = (\boldsymbol{d} - \tilde{\boldsymbol{d}})^{\mathrm{T}}\,\hat{\mathbf{R}}^{-1}\,(\boldsymbol{d} - \tilde{\boldsymbol{d}}). \qquad (28)$$

For a diagonal matrix $\hat{\mathbf{R}}$, we define weights w_i such that

$$\epsilon_i^2 = \frac{\epsilon^2}{w_i} = \frac{\sigma^2}{\lambda}\frac{1}{w_i}. \qquad (29)$$

Defining the diagonal matrix $\mathbf{W} = \mathrm{diag}(w_i)$, the generalised cross-validator then reads:

$$\Theta^2 = \frac{\left(\boldsymbol{d} - \tilde{\boldsymbol{d}}\right)^{\mathrm{T}}\mathbf{W}\left(\boldsymbol{d} - \tilde{\boldsymbol{d}}\right)}{N\left(1 - \frac{1}{N}\mathrm{trace}\,(\mathbf{A})\right)^2}, \qquad (30)$$

where the weights should, according to (21), satisfy (7).

5. Evaluation of the analysis error

The analysis performed with the optimised parameters should have the lowest global error, as measured by (30). Nevertheless, the spatial distribution of the error is also of interest. The latter is very easy to calculate with OI: according to (17) we need to apply the analysis (operator $\mathbf{H}$) on a vector containing the covariances of the data locations with the location where the error has to be calculated. Hence, in principle, a new analysis has to be performed for each point in which the error is requested. On the contrary, the error calculation in Diva is not trivial since the covariance functions are never explicitly specified. The object of this section is to define methods to generate error fields associated to the analysis.

5.1. The poor man's estimate

To circumvent the main problems (unknown covariance function and repeated analysis), Brasseur (1994) estimated the error by analysing a vector of "covariances" with constant σ^2. As all "covariances" are identical, the error can be assessed in all locations with the same analysis. The advantage is the fast calculation, but the drawback is a systematic underestimation of the actual errors, since the error reduction by the overestimated covariances (17) is also overestimated. The poor man's error field is a very efficient way to assess data-coverage and determine the regions where the analysis cannot be trusted.

5.2. The hybrid approach

In this approach, the covariance vector to be analysed is calculated using the covariance function of an infinite domain (Brankart and Brasseur, 1998; Rixen et al., 2000). In an infinite domain, the error calculation is then "exact", while for more complicated domains, it is nearly exact far away from boundaries, provided no anisotropic constraint is activated.

The problem of the approximate hybrid error calculation relies on the fact that instead of the actual covariance, only the kernel computed under the assumption of an infinite domain (without dynamic constrain) is analysed with Diva to produce the error field. Hence, for anisotropic cases (as found with advection constraints, variable length scale or near boundaries), the hybrid error field can be incoherent. This is what motivated the evaluation of the real covariance function in Diva.

5.3. The real covariance method

If we look back at the OI interpretation, we can place a data of value 1 at location $\boldsymbol{r}$ and compute the analysis $\varphi_1(\boldsymbol{r}')$ at a location $\boldsymbol{r}'$:

$$\varphi_1(\boldsymbol{r}') = \frac{\lambda\hat{B}(\boldsymbol{r}, \boldsymbol{r}')}{\lambda\hat{B}(\boldsymbol{r}, \boldsymbol{r}) + \hat{R}(\boldsymbol{r})}, \qquad (31)$$

where $\hat{B}(\boldsymbol{r}, \boldsymbol{r}')$ is the non-dimensional covariance function between points $\boldsymbol{r}$ and $\boldsymbol{r}'$, whereas $\hat{R}$ is the normalized observational error variance. Normalization was done respectively by the background variance σ^2 and noise ϵ^2, yielding the signal-to-noise ratio λ previously defined. At the data location itself, we get the analysis

$$\varphi_1(\boldsymbol{r}) = \frac{\lambda\hat{B}(\boldsymbol{r}, \boldsymbol{r})}{\lambda\hat{B}(\boldsymbol{r}, \boldsymbol{r}) + \hat{R}(\boldsymbol{r})}. \qquad (32)$$

In terms of interpretation of the covariance function as the kernel of the norm (the second term of (1)), it is the background covariance that is modified by the anisotropy and not the noise level. Hence, if we put the unit data value with a unit signal-to-noise ratio in $\boldsymbol{r}$, we directly have

$$\varphi_1(\boldsymbol{r}') = \frac{\hat{B}(\boldsymbol{r}, \boldsymbol{r}')}{\hat{B}(\boldsymbol{r}, \boldsymbol{r}) + 1} \qquad \text{and} \qquad \varphi_1(\boldsymbol{r}) = \frac{\hat{B}(\boldsymbol{r}, \boldsymbol{r})}{\hat{B}(\boldsymbol{r}, \boldsymbol{r}) + 1}. \qquad (33)$$

The left-hand sides are provided by the Diva application to the unit data point in location $\boldsymbol{r}$ with unit signal-to-noise ratio and analysed in any desired location $\boldsymbol{r}'$ and $\boldsymbol{r}$. From these two values, it is therefore easy to calculate the covariance function $\hat{B}$ inherently used in Diva.

For the error calculation at a point $\boldsymbol{r}$, we have the following procedure:

1. Put a unit value at $\boldsymbol{r}$ and perform an analysis with $\lambda = 1$.
2. Save the result at the locations of the original data and of the error-calculation, where the analysis value is

$$\varphi_0 = \frac{\hat{B}(\boldsymbol{r}, \boldsymbol{r})}{\hat{B}(\boldsymbol{r}, \boldsymbol{r}) + 1}, \qquad (34)$$

3. Calculate the background variance $\hat{B}(\boldsymbol{r}, \boldsymbol{r})$ at the error-field location:

$$\hat{B}(\boldsymbol{r}, \boldsymbol{r}) = \frac{\varphi_0}{1 - \varphi_0}. \tag{35}$$

4. Calculate the covariance $\hat{B}(\boldsymbol{r}, \boldsymbol{r}_i)$ between the error location and data locations. Since at the data points i located at $\boldsymbol{r}_i$, Diva application provides

$$\varphi_i = \frac{\hat{B}(\boldsymbol{r}, \boldsymbol{r}_i)}{\hat{B}(\boldsymbol{r}, \boldsymbol{r}) + 1}, \tag{36}$$

the covariance $\hat{B}(\boldsymbol{r}, \boldsymbol{r}_i)$ is obtained as

$$\hat{B}(\boldsymbol{r}, \boldsymbol{r}_i) = \frac{\varphi_i}{(1 - \varphi_0)}. \tag{37}$$

Up to the multiplication constant $1/(1 - \varphi_0)$, the non-dimensional covariance of a point in position $\boldsymbol{r}$ with a list of other points can therefore be obtained by putting a unit data value in $\boldsymbol{r}$ and taking the value of the analysis at the coordinates of the list of points.

To illustrate the procedure, we take a simple case with one point in the center of the domain, and another point near the boundary. Near the boundary, data points influence more easily the analysis because rigidity is reduced (Fig. 2). This translates into a larger background variance. Error fields will therefore be larger near boundaries when there are no nearby data.
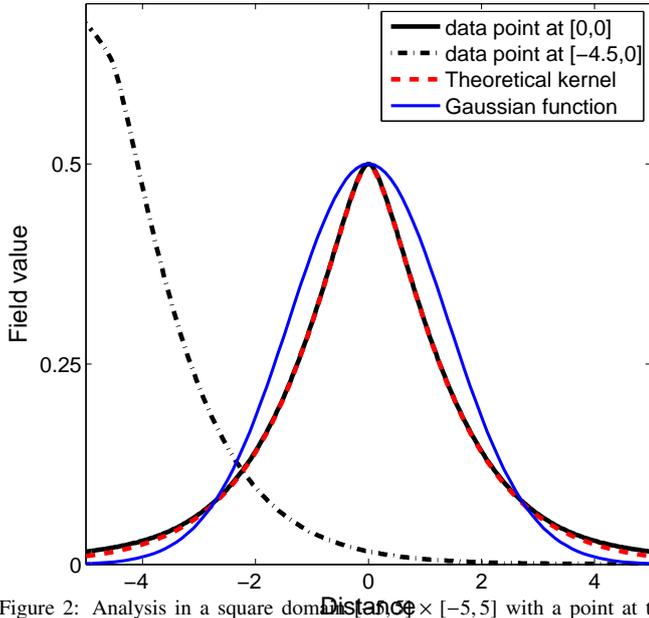
Figure 2: Analysis in a square domain $[-5, 5] \times [-5, 5]$ with a point at the centre and another at $(-4.5, 0)$. The signal-to-noise ratio $\lambda = 1$ for both cases. Note the larger analysis value near the boundary, indicating a larger background variance.

From there, applying the correction $1/(1 - \varphi_0)$, we can use the covariances as input vector for a second Diva execution, so as to perform an analysis of the covariance and getting access to the error of the analysis at the desired location. Indeed, using the equivalence of Diva and OI, if the analysis step applied to a data vector $\boldsymbol{d}$ is formally written

$$\varphi_a = \mathbf{H}\boldsymbol{d}, \tag{38}$$

then the error is

$$\epsilon_a^2(\boldsymbol{r}) = \sigma^2 \hat{B}(\boldsymbol{r}, \boldsymbol{r}) - \sigma^2 \mathbf{H}\hat{\boldsymbol{b}}, \tag{39}$$

where $\hat{B}(\boldsymbol{r}, \boldsymbol{r})$ is the local relative background variance (calculated by (35)) and $\hat{\boldsymbol{b}}$ is a vector filled according to (37).

5.3.1. Error on integral of a field

The calculation of covariances can also be exploited to derive the error committed in the evaluation of the integral of the analysed field (for instance heat or salt content) over the whole domain or a sub-domain. The theory and implementation, which are not in the scope of the present paper, are detailed in (Troupin et al., 2010a, additional document `divaintegral.pdf`).

5.4. Numerical cost

Let us assume that the error is requested at N_c locations (sparse points or regular grid). The error computation with OI (in its original formulation) requires the inversion of a matrix of size $N_d \times N_d$ and the projection of onto the N_c locations. We will now see how it can be done with Diva.

5.4.1. Poor man's estimate

An analysis is performed on a vector filled with constant covariance σ^2. As the vector of covariances is filled with identical values σ^2, the error is assessed for all the N_c locations with the same analysis. Since the matrix to be inverted for the analysis (the stiffness matrix $\mathbf{K}$, (12)) is already factorized[2], the additional cost is almost zero because a single analysis is needed.

5.4.2. Hybrid method

Again, the cost is kept low by exploiting the already existing factorization. For each of the N_c locations, matrix-vector operations are needed. Roughly, if the cost of the first analysis is M, the error field calculation requests $MN_c/\sqrt{N_e}$ operations, where N_e is the number of degrees of freedom of the finite-element mesh.

5.4.3. Real covariance function

For each of the N_c points in which the error is to be evaluated, an additional analyse providing the covariance function is needed. If done naively, this would be prohibitive since a new analyse with another data set (the unit data in different locations) normally requests a new matrix inversion.

To save computing time, a nice mathematical property was discovered and exploited: performing an analysis that differs only by one data point from a previous analysis can be performed using an already existing matrix factorization. In this case, the generation of all covariances (always changing only one data point) can be performed with a cost similar to a single error analysis with prescribed mathematical covariance functions.

---

[2]LU decomposition or factorization: the matrix is written as the product of a lower triangular matrix L and an upper triangular matrix U.

To explain the method, let us suppose that we have constructed and inverted the stiffness matrix $\mathbf{K_0}$ for the same problem without any data point (as explained in Section 2.5, $\mathbf{K_0}$ has a component that is related to the smoothness constraint, not to the data). Then, adding a single data point with unit value at location i would demand the solution of

$$\left(\mathbf{K}_0 + \mu_i \mathbf{S}_i \mathbf{S}_i^{\mathrm{T}}\right) \boldsymbol{q} \ = \ \mu_i \mathbf{S}_i. \tag{40}$$

Here, the Woodbury-Sherman formula yields

$$\boldsymbol{q} = \left(\frac{1}{1 + \mu_i \mathbf{S}_i^{\mathrm{T}} \mathbf{K}_0^{-1} \mathbf{S}_i}\right) \mu_i \mathbf{K}_0^{-1} \mathbf{S}_i. \tag{41}$$

Since the term in parenthesis is a scalar multiplicative factor, the solution with the data is obtained by analysing this data point with the stiffness matrix $\mathbf{K}_0$ and then multiplying all analyses by $1/(1 + \mu_i \mathbf{S}_i^{\mathrm{T}} \mathbf{K}_0^{-1} \mathbf{S}_i)$. The value of

$$\varphi_0 = \mu_i \mathbf{S}_i^{\mathrm{T}} \mathbf{K}_0^{-1} \mathbf{S}_i \tag{42}$$

is actually nothing else than the analysis at the new data location, again using the stiffness matrix $\mathbf{K}_0$. Hence the recipe for calculation of the covariances is now clear:

1. Create and invert once the stiffness matrix $\mathbf{K}_0$ constructed without any data points.
2. For each point for which the covariance is needed:
   (a) Create the elementary charge vector $\mu_i \mathbf{S}_i$.
   (b) Apply the already inverted stiffness matrix $\mathbf{K}_0^{-1}$ and the multiplicative factor of (41) to derive the values of φ_0 and φ_i
   (c) Compute the covariances using (35) and (37).

The efficiency of the method is due to the fact that we remain within the same Diva execution, where the matrix inversion $\mathbf{K}_0^{-1}$ is much less expensive than the initial factorization. Indeed the cost is reduced by a factor $\sqrt{N_e}$, with N_e typically around 10^4-10^5, thus gain is again significant. Each covariance can be stored on disk for later use by the error calculation.

Overall the cost for the full error calculation is now roughly equivalent to twice the hybrid approach, which is a substantial reduction compared to a brute force approach. The only unsolved problem is the storage of the covariance functions if they are calculated before the actual Diva run for the error calculations. This storage will take $N \times N_c$ words when N_c points for error calculations are requested. The intermediate storage can however be avoided by using Unix-pipes between concurrent execution of two Diva cores, one providing the covariance for the other on demand.

5.5. Usage

Calculation of the full covariance function for error calculation is recommended when at least one of the following conditions is true:

- Advection constraint is strong.

- Signal-to-noise ratio is high and few data are available near the boundaries.

- ξ (penalizing gradients) is different from 1 and the kernel is not (10) any more.

- The correlation length is variable over the domain.

In the aforementioned cases, the assumptions for the hybrid approach are not fulfilled and the use of (10) to express the covariance will provide an error field that is not coherent with the analysis. For instance, lower errors can be obtained in regions void of observations.

6. Realistic application

Regional climatologies (i.e., four-dimension fields: longitude, latitude, depth, time) were generated using Diva, for instance in the Mediterranean Sea (Brasseur et al., 1996; Brankart and Brasseur, 1998; Rixen et al., 2005a,b) and in the northeastern Atlantic Ocean (Troupin et al., 2010b). Global analysis at the ocean surface were generated by Tyberghein et al. (2012) for various physical and chemical variables.

Here our focus is on 2-D (horizontal) fields and the corresponding error estimates. The restriction to 2-D analysis is justified by the fact that 3-D analysis are obtained by stacking horizontal analysis corresponding to different horizontal layers (isobaths or isobars). The connection between adjacent layers is addressed by the consistence of the parameters (L and λ) from one layer to its neighbours.

In this application, we retrieved salinity measurements in the Mediterranean Sea at a depth of 30 m in July, for the 1980-1990 period (Fig. 3). The data set is built up by exploiting the SeaDataNet portal (http://www.seadatanet.org) and the World Ocean Database 2009 (WOD09, Boyer et al., 2009) and contains 1061 data points. Basic quality controls (range, vertical gradient values, etc.) are applied prior to the extraction. The decision to perform the analysis only on a subset (1980-1990) of the available data was made with the goal of better revealing the differences between OI and Diva. Tests performed with a much a better data coverage provided very similar results, as the analysis was more constrained by the data than by the method.

The regular grid for the analysis extends from 7°W to 36°E and from 30°15'N to 45°45'N, with a horizontal resolution of about 10 km. Measurements in the Black Sea and the Atlantic Ocean were removed in order to concentrate only on the Mediterranean Sea properties. The land-sea contours are created from the GEBCO bathymetry.

The reference field φ_b is chosen as a linear fit of the measurements, allowing us to represent the eastward increase of salinity in the Mediterranean Sea. For the selected period, the data distribution is inhomogeneous, the southern part of the sea, off Libya and Egypt, having almost no measurements. The most sampled regions are the Alboran Sea, the Ligurian Sea and the Ionan Sea. The less saline waters ($S \leq 36.5$) are located in the Alboran Sea and along the coasts of Morocco and Algeria (Fig. 4). They constitute the Modified Atlantic Water (MAW, Gascard and Richez, 1985), of which the salinity increases eastward. The largest values of salinity ($S \geq 39$) are observed in the eastern part of the domain, close to the Cyprus island. 52% of the measurements fall in the interval 38–38.5.
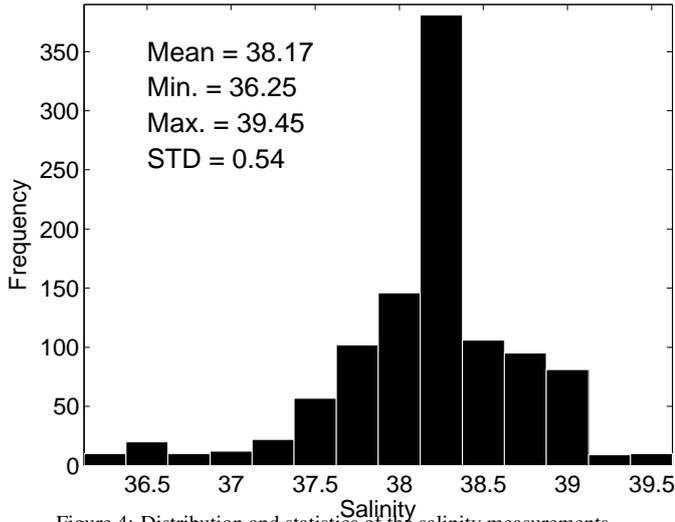
Figure 4: Distribution and statistics of the salinity measurements.

6.1. Determination of the analysis parameters

6.1.1. Correlation length

The correlation length L translates the distance over which a given data point influences its neighbourhood (2.2). It is evaluated by fitting the theoretical kernel of the second term of (1) (see Fig. 2) to the correlation between data, or in other words, it is the value of L in (10) that provides the best fit to the correlation presented in Fig. 5. The fit yields the value $L = 1.42°$ ($\simeq 160$ km). In the same region, Brasseur et al. (1996) obtained $L = 80$ km, but they used measurements covering a longer period (1900-1986) and for the whole year.
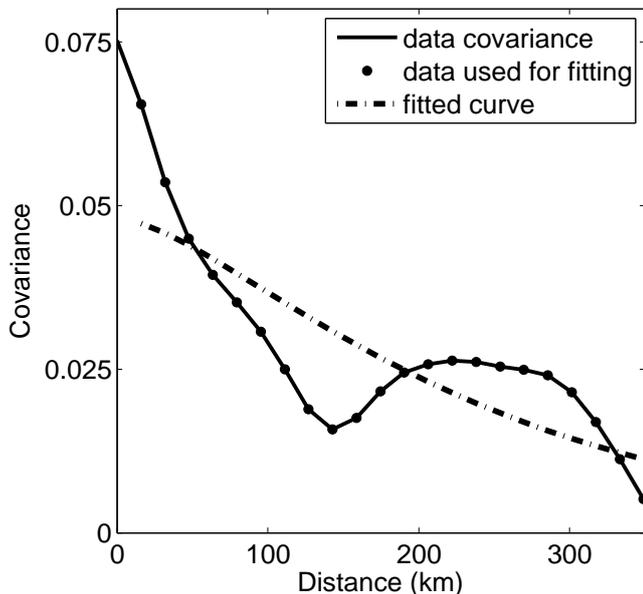
Figure 5: Fit of the data correlation to the theoretical kernel (dashed line).

6.1.2. Signal-to-noise ratio

The signal-to-noise ratio λ can be estimated using ordinary or generalized cross-validations (see Section 4.2). Recent exper-

iments showed that the value of λ provided by these methods tend to be too large, leading to analysed fields displaying unrealistic features (e.g., tracks of oceanographic cruises). This problem in the estimation of λ is attributed to the inhomogeneous spatial distribution of data and to the dependence between close measurements. Thus for the sake of simplicity, the next examples are generated using a unit value for λ, similarly to Troupin et al. (2010b).

6.2. Results

6.2.1. Analysed field

The analysis is performed on the anomalies obtained by subtracting the background field φ_b (linear fit of the data). The latter is then added in order to obtain the salinity field (Fig. 6) The maps exhibits known features in the Mediterranean Sea, such as the eastward propagation Atlantic water in the Algerian Current (e.g., Millot, 1999) or the decrease of salinity due to river discharge (e.g., Po river, northern Adriatic; Rhone river, Gulf of Lyon).
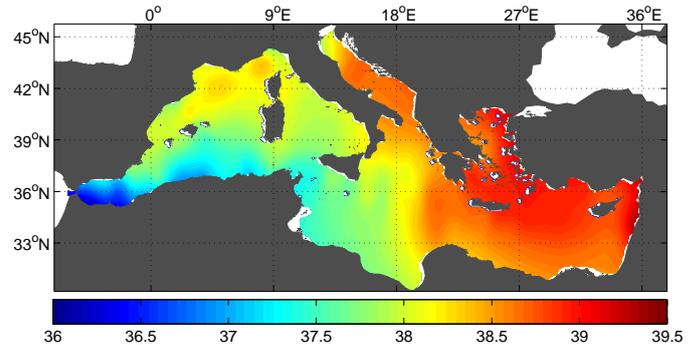

Figure 6: Salinity field obtained with $L = 1.42°$ and $\lambda = 1$.

6.2.2. Comparison with OI

Under certain circumstances, OI and Diva are equivalent (Section 3.2). An OI of the same data anomalies was performed using a Fortran module in which the covariance function was set to a Bessel function (10), instead of a Gaussian (Fig. 2). The OI performed using the same analysis parameters as for Diva ($L = 1.42°$, $\lambda = 1$) is very close to those shown in Fig. 6. The difference between the fields (Fig. 7) demonstrate the divergence of the two techniques: away from the coast, the differences are lower than 0.05 (salinity units) in absolute value. The slight discrepancies observed out of the coastal areas may be due to the particular implementation of OI used for this comparison: the interpolation is performed by using only the m observations closest to the point where the analysed field has to be computed (with $m = 30$ in the present case). Also, due to the size of the Mediterranean Sea, it is difficult to be sufficiently distant from the coast.

The most significant differences occur close to the coasts, especially along the Iberian and Italian Peninsulas. This was expected, since the covariance functions used in Diva are not isotropic in the vicinity of physical boundaries. The role of physical boundaries (land, islands) can also be included in OI,

9

namely the technique called *kriging with faults* (Delhomme, 1978; Chilès and Delfiner, 1999). Another option is described by Ridgway et al. (2002), who use locally weighted regression (also called loess mapping, Cleveland and Devlin, 1988) for producing gridded fields around Australia. The method allowed for anisotropy and topographical effects, but required a large number of adjustable parameters, hence making the analysis more difficult to adapt to a particular region.
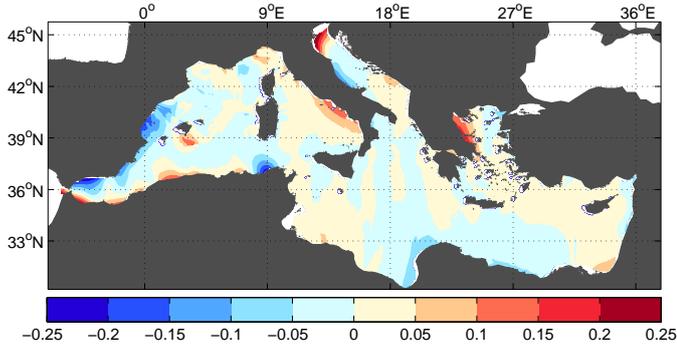

Figure 7: Difference between OI and Diva salinity fields obtained with $L = 1.42°$ and $\lambda = 1$.

Looking closer at the difference between the two fields (Fig. 8), we observe a high correlation (> 98.5%), with most of the differences lower than 0.025 in absolute value.
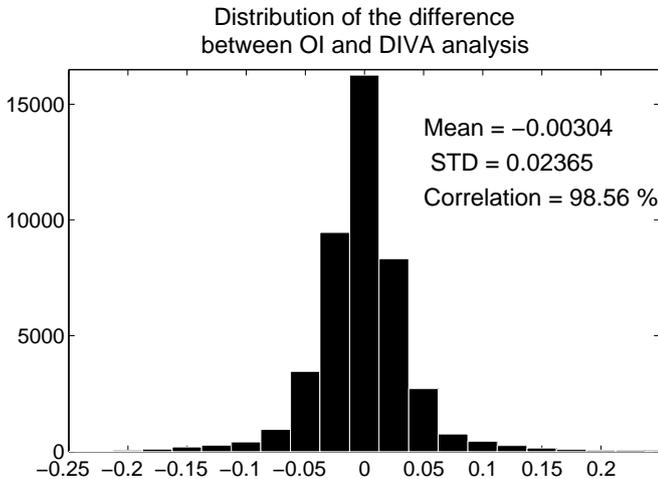
Figure 8: Distribution and statistics of the difference between the salinity fields obtained by Diva and by OI.

6.3. Error fields

The error field corresponding to the data set used for the previous analyses is calculated according to the methods described in Section 5. In this example the error is scaled locally by the local variance of the background field $\hat{B}$; yielding relative errors.

The OI field (Fig. 9a) shows the effect of the data coverage on the error magnitude. The relative error lies between 40 and 60% around the regions with a sufficient amount of observations. The largest values (> 80%) occur along the south coasts

of the Mediterranean Sea, where almost no data are available for the considered period. This means that the analysis obtained in these areas cannot be taken with much confidence.

The poor man's estimate (Fig. 9b) provides an error field with lower values over the whole domain. Where data are available, the error is below 20%, whereas the 80-100% error region is limited to a small zone close to the coast of Libya. The relation (25) explains why this method underestimate the error: the error variance (ϵ^2) is obtained by subtracting the analysis applied to a vector containing the data covariances ($\mathbf{H}c$) to the variance of the background field (σ^2). When we replace the covariance vector c by a vector containing only 1, the covariances are overestimated, so is the error reduction by the analysis.

The hybrid method was build by analogy with the OI error estimate (Brankart and Brasseur, 1998). Thus it is expected that the two methods provide comparable results. Indeed, the error field of Figs. 9(a) and (c) exhibit a similar spatial distribution. Some discrepancies appear in certain regions: along the Italian coasts (on both side of the Peninsula), in the Alboran Sea, and around Cyprus. These differences are related to the presence of the coasts:

1. As with the analysis, the FE method prevent the information to cross land. Hence the error reduction due to the analysis is lower with the hybrid method than with OI.

2. Close to the coasts, the variance of the background field in Diva is increased, due to the specified boundary conditions.

Finally, the error using the real covariance function (Figs. 9d) is also close to the hybrid results. The main differences between the two methods occur in the coastal areas, for instance in the Adriatic Sea or around Cyprus. In these regions, the error is lower when the real covariance is employed, because it allows for the consideration of coastline effects.

The choice of one particular method depends on several factors:

- The size of the output grid.

- The number of analyses to be performed.

- The sources of anisotropies (advection, coastlines).

If the objective is limited to having an indication of the area where the analysed field cannot be trusted, then the poor man's estimate is sufficient. If a more complex error field is to be constructed, a bypass is the reduction of the output grid resolution. This solution is particularly welcome when a large number of analyses ($O(10^2)$) is required, as it is the case for a climatology.

7. Conclusions

Diva is a method to generate gridded fields from sparse observations by minimising a cost function. The method performs analysis in a similar way to OI, but for specific and relevant forms of the covariance functions, for which geometric (e.g., coastlines) and dynamic (e.g., currents) constraints can be taken into account. The minimisation of the cost function is carried

out with a finite-element solver, which guarantees the numerical performance of the method and allows one to work with large data sets ($> 10^6$ data points).

We presented novel features implemented in the software, allowing for the improvement of the gridding method. The most remarkable feature is the construction of the error field associated to the analysis has been improved: in addition to the poor man's estimate and to the hybrid method, new developments give access to the real covariance function, which is not explicitly formulated in Diva. Using this covariance function helps to obtain a more accurate estimate of the error. The algorithms implemented for these calculations have been optimised, so that the numerical cost is not prohibitive for realistic applications.

An application using salinity measurements in the Mediterranean Sea at 30 m for the month of July has been presented. The analyse provided by Diva has been compared with the results coming from OI. A good agreement is found between the two methods, the largest deviations occurring in the coastal areas, where the covariance functions have a different behaviour (isotropic in OI, not in Diva).

The OI error field has been compared to the three methods implemented in Diva.

1. The poor man's estimate yielded an error field with lower values with respect to OI error over the whole domain.

2. The hybrid method provided an error field very similar to the one of OI.

3. The use of the real covariance also provided a similar error field, except in the coastal areas.

Future work on the method will be concentrated on:

- Spatially correlated observation errors: climatologies may be biased towards years with a larger number of data and parameters estimated by generalised cross-validation may not be reliable. Explicitly using a spatially correlated observation error would dramatically increase the numerical cost of the algorithm, hence simpler solutions have to be investigated.

  Brankart et al. (2009) propose to augment the observation vector with differences of the original observations in order to simulate correlation structures. Unfortunately this method requires the data to be located at the nodes of two-dimensional grids, then for sparse data, adaptations are necessary.

- The generalisation to 3 dimensions: currently only two-dimensional analyses can be performed. The generalisation to more than two dimensions would be useful for better representing moving observation platforms (such as gliders) and for dealing with constraints due to vertical advection and mixing.

- The multivariate approach: spatial patterns in physical and biochemical data are often related. A multivariate approach where two (or more) variables are analysed together would improve the overall analysis. For example, the statistical analysis of biochemical parameters found

mostly in rivers can benefit from salinity information to get a more realistic extension of river plumes.

- Transformation applied on data: for non-Gaussian distributed variables (e.g., concentrations), a direct analysis with Diva can lead to problems, such as negative concentrations or overshoots. A non-linear data transformation based on Gaussian anamorphosis should help solving these issues.

References

Abramowitz, M., Stegun, I.A. (Eds.), 1964. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover, New York. ISBN 0-486-61272-4.

Barnes, S.L., 1964. A technique for maximizing details in numerical weather map analysis. J. App. Meteor. 3, 396–409. doi:10.1175/1520-0450(1964)003¡0396:ATFMDI¿2.0.CO;2.

Barth, A., Alvera-Azcárate, A., Troupin, C., Ouberdous, M., Beckers, J.M., 2010. A web interface for griding arbitrarily distributed in situ data based on Data-Interpolating Variational Analysis (DIVA). Adv. Geophys. 28, 29–37. doi:10.5194/adgeo-28-29-2010.

Bennett, A., 1992. Inverse methods in physical oceanography. Cambridge Monographs on Mechanics and Applied Mathematics. ISBN 0-521-38568-7.

Berx, B., Hughes, S., 2009. Climatology of surface and near-bed temperature and salinity on the north-west European continental shelf for 1971-2000. Cont. Shelf Res. 29, 2286–2292. doi:10.1016/j.csr.2009.09.006.

Boyer, T.P., Antonov, J.I., Baranova, O.K., Garcia, H.E., Johnson, D.R., Locarnini, R.A., Mishonov, A.V., O'Brien, T.D., Seidov, D., Smolyar, I.V., Zweng, M.M., 2009. World Ocean Database 2009, Chapter 1: Introduction. Technical Report. National Oceanographic Data Center, Ocean Climate Laboratory. Washington, D.C. 216 pp.

Brankart, J.M., Brasseur, P., 1996. Optimal analysis of in situ data in the Western Mediterranean using statistics and cross-validation. J. Atmos. Oceanic Tech. 13, 477–491. doi:10.1175/1520-0426(1996)013¡0477:OAOISD¿2.0.CO;2.

Brankart, J.M., Brasseur, P., 1998. The general circulation in the Mediterranean Sea: a climatological approach. J. Mar. Syst. 18, 41–70. doi:10.1016/S0924-7963(98)00005-0.

Brankart, J.M., Ubelmann, C., Testut, C.E., Cosme, E., Brasseur, P., Verron, J., 2009. Efficient parameterization of the observation error covariance matrix for square root or Ensemble Kalman Filters: Application to ocean altimetry. Mon. Weather Rev. 137, 1908–1927. doi:10.1175/2008MWR2693.1.

Brasseur, P., 1994. Reconstruction de champs d'observations océanographiques par le Modèle Variationnel Inverse: Méthodologie et Applications. Ph.D. thesis. University of Liège.

Brasseur, P., Beckers, J.M., Brankart, J.M., Schoenauen, R., 1996. Seasonal temperature and salinity fields in the Mediterranean Sea: Climatological analyses of a historical data set. Deep-Sea Res. I 43, 159–192. doi:10.1016/0967-0637(96)00012-X.

Brasseur, P., Haus, J., 1991. Application of a 3-D variational inverse model to the analysis of ecohydrodynamic data in the Northern Bering and Southern Chukchi Seas. J. Mar. Syst. 1, 383–401. doi:10.1016/0924-7963(91)90006-G.

Brasseur, P.P., 1991. A variational inverse method for the reconstruction of general circulation fields in the northern Bering Sea. J. Geophys. Res. 96, 4891–4907. doi:10.1029/90JC02387.

Bretherton, F., Davis, R., Fandry, C., 1976. A technique for objective analysis and design of oceanic experiments applied to Mode-73. Deep-Sea Res. 23, 559–582. doi:10.1016/0011-7471(76)90001-2.

Chilès, J.P., Delfiner, P., 1999. Geostatistics: Modeling spatial uncertainty. Wiley-Interscience. 1st edition. ISBN 0-471-08315-1.

Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: An approach to regression analysis by local fitting. J. Amer. Stat. Assoc. 83, 596–610. doi:10.2307/2289282.

Craven, P., Wahba, G., 1979. Smoothing noisy data with spline functions. Numer. Math. 31, 377–403. doi:10.1007/BF01404567.

Delhomme, J., 1978. Kriging in the hydrosciences. Adv. Water Resour. 1, 251–266. doi:10.1016/0309-1708(78)90039-8.

Franke, R., 1985. Thin plate splines with tension. Comput. Aided Geom. D. 2, 87–95. doi:10.1016/0167-8396(85)90011-1.

Gandin, L.S., 1965. Objective Analysis of Meteorological Fields. Israel Program for Scientific Translations, Jerusalem.

Gascard, J., Richez, C., 1985. Water masses and circulation in the Western Alboran sea and in the Straits of Gibraltar. Prog. Oceanogr. 15, 157–216. doi:10.1016/0079-6611(85)90031-X.

Golub, G.H., Heath, M., Wahba, G., 1979. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21, 215–223. URL: http://www.jstor.org/stable/1268518

Gomis, D., Ruiz, S., Pedder, M., 2001. Diagnostic analysis of the 3D ageostrophic circulation from a multivariate spatial interpolation of CTD and ADCP data. Deep-Sea Res. 48, 269–295. doi:10.1016/S0967-0637(00)00060-1.

Hartman, L., Hössjer, O., 2008. Fast kriging of large data sets with Gaussian Markov random fields. Comput Stat. Data An. 52, 2331–2349. doi:10.1016/j.csda.2007.09.018.

Kaplan, A., Kushnir, Y., Cane, M.A., 2000. Reduced space optimal interpolation of historical marine sea level pressure: 1854-1992*. J. Clim. 13, 2987–3002. doi:10.1175/1520-0442(2000)013¡2987:RSOIOH¿2.0.CO;2.

Legler, D.M., Navon, I.M., O'Brien, J.J., 1989. Objective analysis of pseudostress over the indian ocean using a direct-minimization approach. Mon. Weather Rev. 117, 709–720. doi:10.1175/1520-0493(1989)117¡0709:OAOPOT¿2.0.CO;2.

Locarnini, R.A., Mishonov, A.V., Antonov, J.I., Boyer, T.P., Garcia, H.E., 2006. World Ocean Atlas 2005, Volume 1: Temperature. Technical Report. U.S. Government Printing Office. Washington, D.C. 182pp.

Locarnini, R.A., Mishonov, A.V., Antonov, J.I., Boyer, T.P., Garcia, H.E., Baranova, O.K., Zweng, M.M., Johnson, D.R., 2010. World Ocean Atlas 2009, Volume 1: Temperature. S. Levitus, Ed., NOAA Atlas NESDIS 61. Technical Report. U.S. Government Printing Office. Washington, D.C.

Logan, D.L., 2012. A first course in the Finite Element Method. Global Engineering. Fifth edition. ISBN 978-0-495-66825-1.

McIntosh, P.C., 1990. Oceanographic data interpolation: Objective analysis and splines. J. Geophys. Res. 95, 13529–13541. doi:10.1029/JC095iC08p13529.

Millot, C., 1999. Circulation in the western mediterranean sea. J. Mar. Syst. 20, 423–442. doi:10.1016/S0924-7963(98)00078-5.

Ooyama, K.V., 1987. Scale-controlled objective analysis. Mon. Weather Rev. 115, 2479–2506. doi:10.1175/1520-0493(1987)115¡2479:SCOA¿2.0.CO;2.

Ridgway, K.R., Dunn, J.R., Wilkin, J.L., 2002. Ocean interpolation by four-dimensional weighted least squares – Application to the waters around Australasia. J. Atmos. Oceanic Tech. 19, 1357–1375. doi:10.1175/1520-0426(2002)019¡1357:OIBFDW¿2.0.CO;2.

Rixen, M., Beckers, J.M., Brankart, J.M., Brasseur, P., 2000. A numerically efficient data analysis method with error map generation. Ocean Model. 2, 45–60. doi:10.1016/S1463-5003(00)00009-3.

Rixen, M., Beckers, J.M., Levitus, S., Antonov, J., Boyer, T., Maillard, C., Fichaut, M., Balopoulos, E., Iona, S., Dooley, H., Garcia, M.J., Manca, B., Giorgetti, A., Manzella, G., Mikhailov, N., Pinardi, N., Zavatarelli, M., the Medar Consortium, 2005a. The Western Mediterranean Deep Water: a proxy for global climate change. Geophys. Res. Lett. 32, L12608. doi:10.1029/2005GL022702.

Rixen, M., Beckers, J.M., Maillard, C., the MEDAR Group, 2005b. A hydrographic and bio-chemical climatology of the Mediterranean and the Black Sea: a technical note on the use of coastal data. B. Geofis. Teor. Appl. 46, 319–327.

Schweikert, D.G., 1966. An interpolation curve using a spline in tension. J. Math. and Phys. 45, 312–317.

Shen, S., Smith, T., Ropelewski, C., Livezey, R., 1998. An optimal regional averaging method with error estimates and a test using tropical pacific SST data. J. Clim. 11, 2340–2350. doi:10.1175/1520-0442(1998)011¡2340:AORAMW¿2.0.CO;2.

Steele, M., Morley, R., Ermold, W., 2001. PHC: A global ocean hydrography with a high-quality Arctic Ocean. J. Clim. 14, 2079–2087. doi:10.1175/1520-0442(2001)014¡2079:PAGOHW¿2.0.CO;2.

von Storch, H., Zwiers, F., 1999. Statistical analysis in climate research. Cambdridge University Press, Cambridge. ISBN 0-521-45071-3.

Tandeo, P., Ailliot, P., Autret, E., 2011. Linear Gaussian state-space model with irregular sampling: application to sea surface temperature. Stoch. Env. Res. Risk. A. 25, 793–804. doi:10.1007/s00477-010-0442-8.

Teague, W.J., Carron, M.J., Hogan, P.J., 1990. A comparison between the Generalized Digital Environmental Model and Levitus climatologies. J. Geophys. Res. 95, 7167–7183. doi::10.1029/JC095iC05p07167.

Troupin, C., Beckers, J.M., Ouberdous, M., Sirjacobs, D., 2010a. DIVA User's Guide. GeoHydrodynamics and Environment Research. Available at http://modb.oce.ulg.ac.be/projects/1/diva.

Troupin, C., Machín, F., Ouberdous, M., Sirjacobs, D., Barth, A., Beckers, J.M., 2010b. High-resolution climatology of the north-east atlantic using data-interpolating variational analysis (Diva). J. Geophys. Res. 115, C08005. doi:10.1029/2009JC005512.

Tyberghein, L., Verbruggen, H., Klaas, P., Troupin, C., Mineur, F., De Clerck, O., 2012. ORACLE: a global environmental dataset for marine species distribution modeling. Global Ecology and Biogeography 21, 272–281. doi:10.1111/j.1466-8238.2011.00656.x.

Wahba, G., 1975. Smoothing noisy data with spline functions. Numer. Math. 24, 383–393. doi:10.1007/BF01437407.

Wahba, G., Wendelberger, J., 1980. Some new mathematical methods for variational objective analysis using splines and cross validation. Mon. Weather Rev. 108, 1122–1143. doi:10.1175/1520-0493(1980)108¡1122:SNMMFV¿2.0.CO;2.

Zhang, H., Wang, Y., 2010. Kriging and cross-validation for massive spatial data. Environmetrics 21, 290–304. doi:10.1002/env.1023.
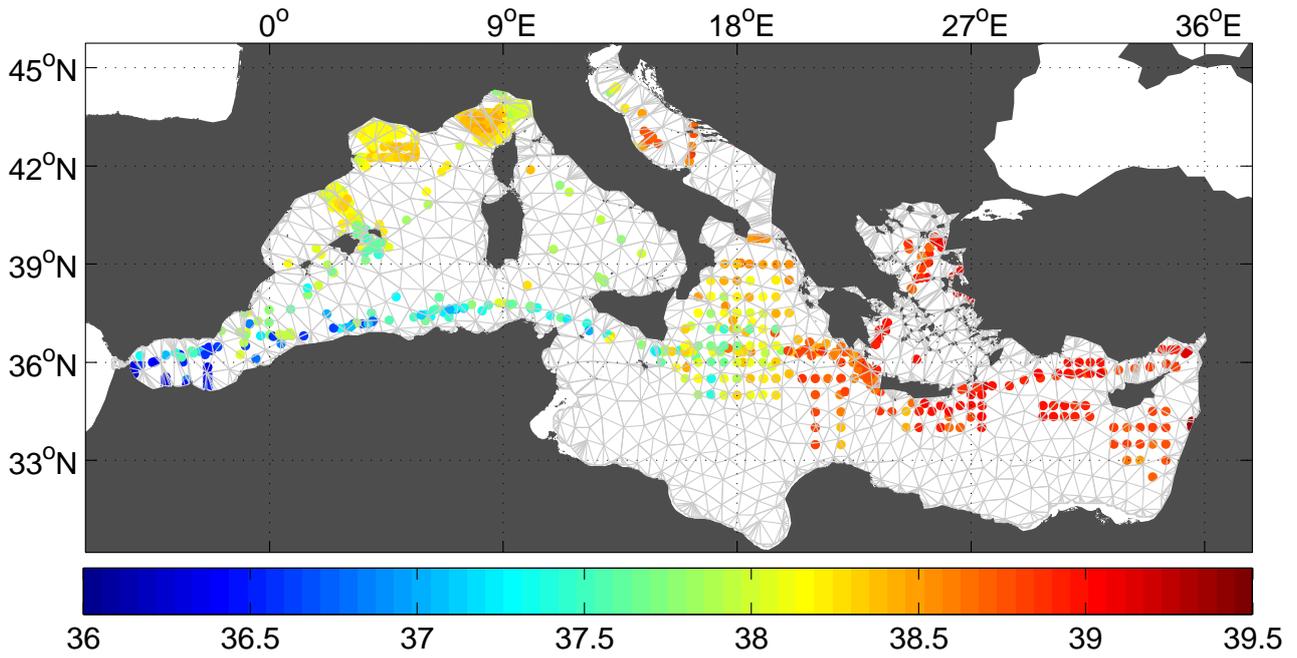
Figure 3: Finite-element mesh and salinity measurements used for the application.
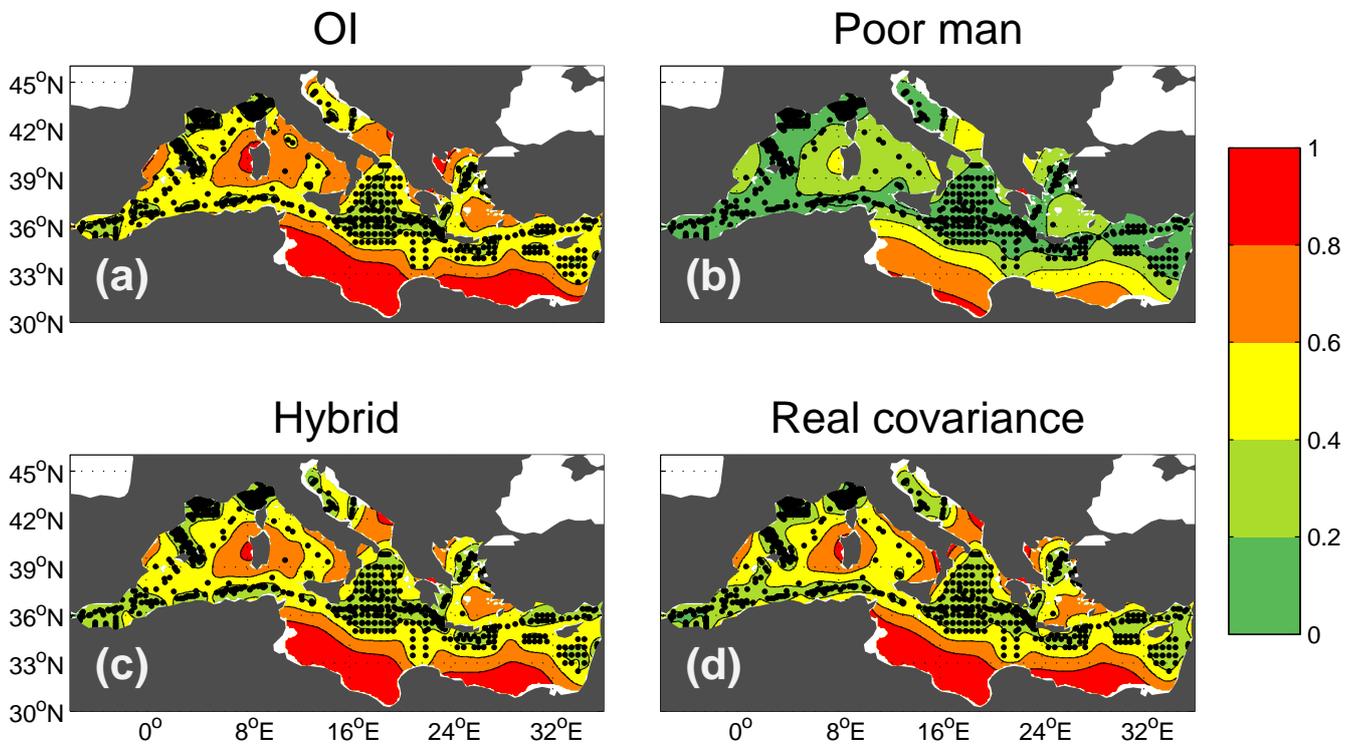

Figure 9: Error fields computed using four different methods: (a) OI, (b) poor man's estimate, (c) hybrid and (d) real covariance methods.