Approximate and efficient methods to assess error fields in spatial gridding with Diva (Data Interpolating Variational Analysis).

JEAN-MARIE BECKERS [*], ALEXANDER BARTH

University of Liège, Belgium

CHARLES TROUPIN,

IMEDEA, Spain

AIDA ALVERA-AZCÁRATE

University of Liège, Belgium

[*] *Corresponding author address:* Jean-Marie Beckers, University of Liège, GHER-AGO, Sart-Tilman B5, 4000 Liège, Belgium.

E-mail: JM.Beckers@ulg.ac.be

ABSTRACT

We present new approximate methods to provide error fields for the spatial analysis tool Diva. It is first shown how to replace the costly analysis of a large number of covariance functions by a single analysis for quick error computations. Then another method is presented where the error is only calculated in a small number of locations and from there the spatial error field itself interpolated by the analysis tool. The efficiency of the methods is illustrated on simple schematic test cases and a real application in the Mediterranean Sea. These examples show that with these methods one has the possibility for quick masking of regions void of sufficient data and the production of "exact" error fields at reasonable cost. The error-calculation methods can also be generalized for use with other analysis methods such as 3D-Var and are therefore potentially interesting for other implementations.

1. Introduction

Spatial analysis of observations, also called gridding, is a common task in oceanography or meteorology and a series of methods and implementations exist and are widely used. The N_d data points of values d_i, $i = 1, ..., N_d$ at location (x_i, y_i) are generally distributed unevenly in space. Furthermore the values d_i are affected by observational errors, including representativity errors. From this data set an analysis on a regular grid is often desired. It has been quickly recognized that it would be natural to define the best analysis as the one which has the lowest expected error. This definition has lead to Kriging and Optimal Interpolation (OI) methods (*e.g.* Gandin 1965; Delhomme 1978; Bretherton et al. 1976) and the Kalman-Bucy filter and data assimilation with adjoint models in the context of forcast models (*e.g.* Lorenc 1986).

These methods assume that statistics on observational errors and the spatial covariance of the field to be analyzed are available to infer the "best" analysis field. As these methods aim at minimizing the analysis error, it is not a surprise that they also provide the theoretical *a posteriori* error field for the analysis. The practical implementation of these methods can lead to very different performances, also when it is necessary to calculate the error fields (*e.g.* Bouttier and Courtier 2002).

The present paper will focus on a computationally efficient way to provide error fields for a gridding tool called Diva (Data Interpolating Variational Analysis) whose full description can be found elsewhere (Brasseur 1994; Brasseur et al. 1996; Troupin et al. 2012) and is not repeated here. Looking at Diva gridding is not restrictive as we can later exploit relationships with other formulations to allow for generalizations. In Diva, the gridded field φ over the two-dimensional domain D is searched as the field which minimizes J defined by

$$J[\varphi] = \sum_{i=1}^{N_d} \mu_i \left[d_i - \varphi(x_i, y_i)\right]^2 + \|\varphi\|^2 \tag{1}$$

where weights μ_i control how the analysis has to be close to the data d_i and where the norm $\|\varphi\|$ measuring spatial irregularity is defined

$$\|\varphi\| = \int_D \left(\alpha_2 \boldsymbol{\nabla}\boldsymbol{\nabla}\varphi : \boldsymbol{\nabla}\boldsymbol{\nabla}\varphi + \alpha_1 \boldsymbol{\nabla}\varphi \cdot \boldsymbol{\nabla}\varphi + \alpha_0 \varphi^2\right) dD \tag{2}$$

This term enforces the solution to be more or less regular via the use of the gradient operator $\boldsymbol{\nabla} = (\partial/\partial x, \partial/\partial y)$. Coefficients α_2, α_1 and α_0 control to which extend curvature, gradients and amplitudes of the fields are penalized[1]. The term penalizing the amplitude of solution ensures that in regions far away of data the analyzed anomalies tend to zero which avoids the extrapolation problems one would otherwise encounter (*e.g.* Seaman and Hutchinson 1985). The parameters of the formulation which can be translated into a correlation length scale and a signal to noise ratio (*e.g.* Brasseur et al. 1996), can be calibrated by cross-validation techniques such described in Wahba and Wendelberger (1980). The order two of the highest derivations in the regularization term remain however fixed as the other parameters allow for sufficient freedom.

This formulation is discretized on a finite-element mesh covering the domain with triangles. Each of the triangles is in fact subdivided in three sub-triangles on each of which the solution is expanded as a cubic polynomial. This rich function allow a sufficient degree of continuity so that the functional is well defined. The unknowns are then the coefficients of the polynomials, or in the finite-element vocabulary, the connectors. The functional is a quadratic function of these connectors and the minimization leads to a linear system to be solved for these connectors. In the present implementation this solution is done by a direct skyline solver exploiting the banded structure of the matrix to invert. For larger problems the recent Diva version also allows in iterative solution of this sparse linear system with a preconditioning.

Because of the finite-element grid covering only the real domain of interest D, disconnections, barriers, islands etc are naturally taken into account (*e.g.* Troupin et al. 2010). The solution can be shown to be equivalent to an optimal interpolation (*e.g.* McIntosh 1990; Barth et al. 2013) and to the solution of another minimization problem where the function to be minimized is defined as

$$2J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^{\mathsf{T}} \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{H}\mathbf{x} - \mathbf{d})^{\mathsf{T}} \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{d}) \tag{3}$$

where $\mathbf{x}$ is a column array storing the analyzed field on each grid point where the analysis is needed, $\mathbf{d}$ is an array containing the observations and $\mathbf{H}$ is a linear observation operator which extracts the gridded solution at the data locations, so that $\mathbf{d} - \mathbf{H}\mathbf{x}$ measures the misfit between the observations

[1] $\mathbf{a} \cdot \mathbf{a}$ stands for the standard scalar product $\sum_i a_i a_i$ of vectors and $\mathbf{A} : \mathbf{A}$ for its generalization $\sum_{i,j} A_{ij} A_{ij}$

and the field $\mathbf{x}$. $\mathbf{B}$ is the covariance matrix of the background field $\mathbf{x}_b$ and $\mathbf{R}$ is a covariance matrix holding the observational error covariances. The equivalence with Diva is ensured if $\mathbf{R}$ is diagonal, $\mathbf{B}$ is using the so-called kernel of the norm (2) as covariance function. The kernel is in fact nothing else than the correlation function one would use to create $\mathbf{B}$ yielding the same result in OI as with the variational approach (*e.g.* Wahba and Wendelberger 1980). Furthermore for an equivalent result the weights μ_j are scaled by the inverse of the signal-to-noise ratio defined by matrices $\mathbf{B}$ and $\mathbf{R}$ (Barth et al. 2013). The minimization formulation (3) is a special case of the so-called 3D-Var method (*e.g.* Fischer et al. 2005) with a linear observation operator. For simplicity we keep the name 3D-Var even if we use the equivalence with Diva in a 2D framework. The solution which minimizes the 3D-Var functional (3) is itself equivalent to the OI analysis step (*e.g.* Kalnay 2003):

$$\mathbf{x} - \mathbf{x}_b = \mathbf{K}\left(\mathbf{d} - \mathbf{H}\mathbf{x}_b\right) \tag{4}$$

with the Kalman gain matrix $\mathbf{K}$ defined by

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathsf{T}}\left(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathsf{T}} + \mathbf{R}\right)^{-1}. \tag{5}$$

The choice of the background field $\mathbf{x}_b$ depends on the application: for operational forecasts it is the modeled forecast, for oceanic cruise mapping it can be a climatological field, for the computation of climatologies it can be a constant reference value etc. For the simplicity of the presentation, we assume from here on that we work with anomalies with respect to this background field ($\mathbf{x} - \mathbf{x}_b$ is replaced by $\mathbf{x}$ and $\mathbf{d} - \mathbf{H}\mathbf{x}_b$ replaced by $\mathbf{d}$).

The analysis-error covariance $\mathbf{P}$ then reads with different equivalent formulations (*e.g.* Rabier and Courtier 1992; Courtier et al. 1994; Bouttier and Courtier 2002)

$$\mathbf{P} = \left(\mathbf{B}^{-1} + \mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1} = \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B}. \tag{6}$$

As easily seen, this matrix is also the inverse of the Hessian[2] matrix of J in (3). The diagonal terms of the error covariance matrix provide the error variance of the analysis on the grid defined by $\mathbf{x}$. This error variance in each point is the quantity we will focus on later.

[2]Derivatives are with respect to coordinates $\mathbf{x}$.

From (6), between the data locations, the error covariance $\mathbf{P}_d$ of the analysis reads, expressed in terms of the background covariance matrix between data locations $\mathbf{B}_d = \mathbf{HBH}^{\mathsf{T}}$

$$\begin{aligned} \mathbf{P}_d &= \mathbf{HPH}^{\mathsf{T}} = \mathbf{HBH}^{\mathsf{T}} - \mathbf{HKHBH}^{\mathsf{T}} \\ &= \mathbf{B}_d - \mathbf{B}_d(\mathbf{B}_d + \mathbf{R})^{-1}\mathbf{B}_d = \mathbf{B}_d(\mathbf{B}_d + \mathbf{R})^{-1}\mathbf{R}. \end{aligned} \tag{7}$$

In Diva or 3D-Var, matrices $\mathbf{K}$ and $\mathbf{B}$ are actually never formed but the application of $\mathbf{K}$ to a vector can be seen as the application of the analysis tool to a data set stored in this vector. Similarly $\mathbf{HK}$ applied to a vector consists of applying the tool to the data and then retrieving the analysis at the data locations.

For the linear observation operators used here, Diva, 3D-Var and OI provide the same results (under the hypotheses mentioned above), but the computational aspects are quite different, in particular when it comes to the error calculations.

For 3D-Var implementations, the calculation of the *a posteriori* error covariance requires the computation of the inverse Hessian matrix whereas the analysis itself only uses gradient calculations (*e.g.* Rabier and Courtier 1992). To some extent, the need to calculate the full Hessian matrix can be circumvented by the use of Lanczos vectors of the conjugate gradient approach (*e.g.* Moore et al. 2011, in the context of 4D-Var). In this case the need of more Lanczos vectors required to provide an accurate estimate of the Hessian matrix defeats however the purpose of the conjugate gradient approach to use as few iterations as possible. More recently, with approaches specifying the background covariance matrices by an ensemble (*e.g.* Hamill and Snyder 2000), error calculations can use the equivalence with OI to exploit the reduced rank of the covariance matrix.

For OI, in each point where the analysis is needed, an analysis of the covariance is requested for the *a posteriori* error calculation. This can lead to very high computational costs unless reduced rank approaches are possible (*e.g.* Kaplan et al. 2000; Beckers et al. 2006) or localization is used (*e.g.* Reynolds and Smith 1994). In the latter case, the error field can be calculated at the same time as the local analysis almost without additional cost. It also has the advantage to allow a highly parallel approach.

For Diva, several problems exist: a) neither covariance functions nor background matrices are explicitly formulated so that error calculations have been only made possible by exploiting the equivalence with OI and the discovery of a quick method to numerically calculate on the fly covariance functions (Troupin et al. 2012); b) the computational burden is still high as an analysis in each of the N points where the error is requested must be performed; c) localization could only be exploited at the inversion step of the finite-element formulation by exploiting the banded structure of the matrix to calculate the value of a connector. This has not been implemented as it would lead to suboptimal solutions and in any case would not allow the error calculation in parallel with the analysis (such as in OI implementations), as the error field is not formulated in terms of connectors.

So several methods are faced with high computational costs to retrieve error fields. Because generally covariances are estimated from data (*e.g.* Emery and Thomson 2001) and are not perfectly specified, we expect that error fields derived from the theoretical models are not "true" error fields in any case. Therefore it can be considered an overkill in computations trying to calculate errors with the full theoretical formulation in all locations and some relaxation can be accepted.

The present paper will present in Section 2 two "error calculations" which to various degrees mimic the "exact" error field but with reduced cost. The method will be illustrated in Section 3 with the 2D version of Diva, but generalizations to the other cases mentioned in the introduction will be discussed in Section 4.

2. Approximations for error analysis at reduced costs

The direct formulations for error covariances are rarely applied because matrices are too large and/or covariance matrices not explicitly formulated. Alternative ways to get information on the analysis error are desirable.

If we are only interested in the trace of $\mathbf{P}$, providing a global error estimate, randomized estimates (*e.g.* Girard 1998) apply the analysis tool to random vectors and provide trace estimates

(*e.g.* Troupin et al. 2013). These methods converge quite easily and are used in cross validation techniques (*e.g.* Xiang and Wahba 1996). The estimation of each individual term on the diagonal is however more challenging and convergence much slower. It is possible to use particular structures in the random vectors (*e.g.* Bekas et al. 2007), but convergence turns out to be still rather slow for our case, needing therefore a number of analyses of random vectors to be performed not significantly lower than N, the number of diagonal terms to be evaluated. Here we will exploit the idea of applying the analysis tool not to randomly chosen "data vectors" but to well designed ones. This has been implemented to probe the diagonal (Tang and Saad 2012), but here we will try to capture only some of the diagonal terms and then guessing the other terms by spatial coherence. Instead of trying to calculate the error covariance we can of course also rather focus on the error reduction term $\mathbf{\Delta}$ defined from (6) by $\mathbf{P} = \mathbf{B} - \mathbf{\Delta}$:

$$\mathbf{\Delta} = \underbrace{\mathbf{B}\mathbf{H}^{\mathsf{T}}\left(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathsf{T}} + \mathbf{R}\right)^{-1}}_{\mathbf{K}}\mathbf{H}\mathbf{B}. \tag{8}$$

This formulation shows that if we have a tool to analyze a data array (the under-braced terms are the formal equivalent of the tool), it is sufficient to analyze covariances (columns of $\mathbf{HB}$) to get the error field, but for each point in which the error is requested another covariance must be analyzed. For Diva, the main challenge in the past was the fact that covariances are never explicitly formulated, yet needed for the error computation. In first Diva versions (*e.g.* Rixen et al. 2000), this problem was circumvented by using as approximate covariance function an analytical solution for the minimum of (1) applied in an infinite isotropic domain (a method called *hybrid* in the following and used in Brankart and Brasseur. (1996); Troupin et al. (2012)). Recently it has been shown Troupin et al. (2012) how to use the Diva tool itself to numerically calculate the covariances in an optimized way (a method called *real covariance* in the following). Now we will aim at downgrading this method to make it more economical.

a. Clever poor man's error

If we replace covariances to be analyzed with a vector with all elements being a constant background variance, we generally overestimate the error reduction but we have a computational huge gain, because the same analysis is valid for ALL of the N points in which we want to calculate the error. Instead of N backward substitutions or iterative solutions, we only need one additional analysis to add an "error field" to the analysis. This was already implemented (Troupin et al. 2010) and was called *poor man's error*. In reality we can do better for a similar cost by looking at the situation of an isolated data point and focus on the error reduction (8).

With a single data of anomaly value d and isotropic covariances, the analysis x^a at a distance r from the data location reads according to (4)

$$x^a(r) \;=\; \frac{\sigma^2}{\sigma^2+\epsilon^2}\, c(r/L)\, d \tag{9}$$

where $c(r/L)$ is the correlation function of the background field (the kernel of the Divafunctional), ϵ^2 the observational noise and σ^2 the variance of the background field, defining the so-called signal-to-noise ratio σ^2/ϵ^2. The error reduction term (8), scaled by the background variance reads

$$\frac{\Delta(r)}{\sigma^2} \;=\; \frac{\sigma^2}{\sigma^2+\epsilon^2}\, c^2(r/L). \tag{10}$$

We see that when applying the idea of the poor man's error (putting $d = 1$ into the analysis) analysis (9) yields some resemblance with the actual non-dimensional error reduction (10), but overestimates the error reduction since it uses c instead of c^2.

To go further, we can notice that for the often used Gaussian correlation $c = \exp(-r^2/L^2)$, we have $c^2(r/L) = c(\sqrt{2}r/L)$. In other words, in this case we can obtain the exact error reduction by applying the poor man's error approach with a length scale divided by $\sqrt{2}$. For more general correlation functions, obviously it is rare to find a new length scale L' such that $c(r/L') = c^2(r/L)$, but one can try to optimize the value so that the functions are close to each other in a root mean square sense in a 2D domain:

$$\min_{L'} \int_0^\infty \left[c(r/L') \;-\; c^2(r/L)\right]^2 r\, dr. \tag{11}$$

This minimization can be done easily when the covariance function is known. For Diva, the covariance function in an infinite 2D domain can be expressed in terms of the modified Bessel function K_1 as (Brasseur 1994):

$$c(r/L) \;=\; \frac{r}{L}\mathrm{K}_1(\tfrac{r}{L}) \tag{12}$$

and the optimal value of L' is

$$L' = \frac{L}{1.70677}. \tag{13}$$

The quality of the approximation can be seen on Figure 1.

If we have several data points separated by distances much larger than the correlation length scale, the presence of other data points does not influence the analysis and error field around a data point and hence the poor man's error calculation replacing all data values by one and changing the correlation length scale will provide, with a single analysis, the error reduction term on the full grid.

For regions with higher data coverage, obviously the method provides a too optimistic view of the error, but the method can be easily used to mask gridded regions far away from the data (see error on mask on a 101×101 grid on Figure 2)

The recipe for the method, which we call *clever poor man's error* is thus straightforward: adapt the correlation length scale and then apply the analysis tool to a data vector with unit values to retrieve the complete error reduction field in a single analysis step.

b. Almost exact error fields

For a diagonal observational error covariance matrix, the error at data location i is easily reformulated from (7) as:

$$\mathbf{e_i}^{\mathsf{T}}\mathbf{P}_d\,\mathbf{e_i} = \mathrm{R_{ii}}\mathbf{e_i}^{\mathsf{T}}\mathbf{B}_d(\mathbf{B}_d \;+\; \mathbf{R})^{-1}\mathbf{e_i} \tag{14}$$

where $\mathbf{e_i}$ is a vector with zeros everywhere except 1 on element i. This makes appear an interesting parameter:

$$\mathrm{A_{ii}} = \mathbf{e_i}^{\mathsf{T}}\mathbf{B}_d(\mathbf{B}_d + \mathbf{R})^{-1}\mathbf{e_i} = \mathbf{e_i}^{\mathsf{T}}\mathbf{H}\underbrace{\mathbf{B}\mathbf{H}^{\mathsf{T}}\left(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathsf{T}} + \mathbf{R}\right)^{-1}}_{\mathbf{K}}\mathbf{e_i} \tag{15}$$

If we know the value of $\mathrm{A_{ii}}$ then the error is readily available at the data location as $\mathrm{R_{ii}A_{ii}}$. As before, the calculation of $\mathrm{A_{ii}}$ is suggested by the formulation read from right to left: It is sufficient to apply the analysis tool ($\mathbf{K}$) to a data vector with zeros everywhere and 1 at location i (vector $\mathbf{e_i}$) and then taking the value of the analysis at location i (operator $\mathbf{e_i}^{\mathsf{T}}\mathbf{H}$).

There is in fact another reason to calculate $\mathrm{A_{ii}}$: it provides a way for data quality check. Indeed (*e.g.* Bretherton et al. 1976; Troupin et al. 2013), the expected misfit between observation and analysis $\mathcal{E}\left[\mathbf{d} - \mathbf{H}\mathbf{K}\mathbf{d}\right]$ has the following variance

$$\mathcal{E}\left[\left(\mathbf{d} - \mathbf{H}\mathbf{K}\mathbf{d}\right)\left(\mathbf{d} - \mathbf{H}\mathbf{K}\mathbf{d}\right)^{\mathsf{T}}\right] = \mathbf{R}\left(\mathbf{I} - \mathbf{H}\mathbf{K}\right). \tag{16}$$

This variance can be exploited to check whether the actual analysis-data difference is significantly different from the expected difference, which then allows one to flag data as suspect. At data location i with diagonal observational error the data-analysis misfit according to (16) should have the following variance s_i^2

$$s_i^2 = \mathrm{R_{ii}}(1 - \mathrm{A_{ii}}) \tag{17}$$

so that comparing the actual analysis data misfit to the expected one, suspect data can be identified.

For this use and also because $\mathrm{A_{ii}}$ in needed in cross-validation techniques (*e.g.* Wahba and Wendelberger 1980; Brankart and Brasseur. 1996), the calculation of $\mathrm{A_{ii}}$ (via an analysis of a data vector with zeros everywhere except at data point i) has been optimized for Diva and is accessible at reasonable cost (Troupin et al. 2013). This means we can calculate the error estimates at data locations, which leaves only one problem: how to calculate the error in other locations ?

An easy way to achieve this is to add a pseudo-data point with a virtual huge observational error for any location where the error has to be calculated. For Diva, this high observational error translates into a very small data weight μ_i (1), which numerically does not cause any problem in

the data analysis step. It is then easy to calculate the error at any location. However this would still be costly if done everywhere, as $\mathrm{A_{ii}}$ needs to be calculated in this pseudo-data location without the benefit in terms of outliers detection or cross validation (as we know that the data are not real). We should therefore limit the number of additional pseudo-data points and still be able to calculate the error everywhere. In fact we can consider this again as a gridding problem: knowing the error "exactly" at a series of points, what is the value of the error field in other locations? We can thus use the gridding tool itself where the "observations" are the calculated errors and where the "observational" error is zero and hence the signal-to-noise ratio is infinity (or just very large in the numerical code). There remains to specify the correlation length scale for gridding the error field, but as shown in the analysis of the clever poor man's error, a good choice is the adapted length scale L' (13). Furthermore, it is easy to define the background field if we grid the error reduction: since the "data" locations are the places where we have the error exactly, in other locations we do not have data and the background error reduction is simply zero. Finally, because of the influence of "data" over a correlation length distance, it seems reasonable to add randomly $\alpha^2 D/L^2$ pseudo data over the surface D where $\alpha \sim 1$ defines the precision with which we want the error field.

For completeness, a discussion on the background covariance is needed. Up to now we have scaled the error reduction by σ^2, the overall background variance. However, with Diva, the background covariance varies spatially and increases near boundaries because of the variational formulation (Troupin et al. 2012). So the local background variance in location (x, y) has a value of $\sigma^2 \hat{B}(x, y)$ where $\hat{B}(x, y)$ is now a non-dimensional local background covariance. Sometimes it is interesting to present the relative errors in which case a scaling by this local background covariance is necessary. The calculation of local background variances can be done at some cost with the covariance module (Troupin et al. 2012) of Diva, but only applied in the data points in this case. So one has the choice to scale or not with this non-dimensional background field and the unscaled error field is referred to as *with boundary effect* or *bnd*. With the boundary effect, because of the less uniform behavior of the error field near the boundaries (see examples later), we generate a series of pseudo-data in each finite-element mesh forming the boundary (we can

add more points compared to the scaled error field because we do not need to calculate the local background variance in the unscaled version).

A final comment concerns the number of data and the cost to calculate $\mathrm{A_{ii}}$ for each data point: generally the number of data points is much lower than the number of grid points so that the computational burden to calculate these coefficients remains reasonable compared to the burden of a full error calculation. Should there be a very large number of observations, there is no problem to restrict the error calculation to a subset of the data points as together with the pseudo data points a nice coverage of the grid is easily achieved.

3. Test cases

To diagnose the quality of the error estimates, we will provide three indicators: a graphical representation and two numbers. The first metric is simply the relative error on the error field (root mean square of the difference in error variances between the true error field and the approximate one compared to the true error variance). The second one tries to check how well the error field can be used to mask regions with insufficient data coverage. Typically, when the error variance of the analysis is larger than 50% of the background variance, it means the data did not provide a significant amount of information and the analysis could be masked. Then we can compare the masks derived from the exact error and the approximate one and see how many grid points do not have the same mask.

a. A single data point

This case simply serves to check that the analysis we showed is valid and to see how the different methods compare in the situation with a single data point in the center of the domain with a unit signal-to-noise ratio and a unit background variance. In this case, the error variance at the origin is 0.5 and the standard deviation shown in Figure 2 is 0.707. The number of grid points for the gridded field is 101×101 to which we can compare the number of mask misses.

For all errors without taking into account the boundary effects, the visual inspection shows that the hybrid, clever poor man's error and almost exact error approach are indistinguishable from the exact solution. Only the poor man's error is significantly different, as expected. Quantitatively the relative errors on the error fields are less than a percent and no mask errors occur, except again for the poor man's error. The hybrid error estimate is very close to the exact one using real covariances. The slight difference is due to the fact that the analytical covariance function (12) is the one of an infinite domain, whereas the computation domain used here is finite. When boundary effects are taken into account, we observe the highest errors near the boundary (see Troupin et al. (2012) for details and explanations). But again, the approximate fields are of excellent quality though with higher rms error because of the stronger spatial variability of the error field. To capture this variability better, we can increase the number of pseudo data by increasing α. Indeed with a value of $\alpha = 3$ (Figure 3) the quality increases, whereas decreasing the value of α provides still acceptable results and the error mask is still excellent in this case.

The computational time is not yet shown here as a single data point is rarely encountered in practice and the CPU time of the present case is similar to the one of Section c (see Table 1).

b. Aligned data points

A slightly more complicated situation is one where ten points are aligned in $y = 0$ for $x \geq 0$ as shown on Figure 4.

The poor man's error is now clearly too optimistic, also at the data locations, because it overestimates the error reduction at each data point due to the other data points. The clever poor man's estimate clearly reduces the problem but the hybrid and almost exact error outperform it. We also see that the hybrid method degrades near data points close to the boundary, as to be expected.

c. *Points in part of the domain only*

The same conclusions as in the previous case hold if we now place 150 data points in the upper right part of the domain (Figure 5). The clever poor man's error improves the results from the poor man's error, but the hybrid and almost exact error perform better, with the best approximate method again the almost exact error version. For boundary effects, capturing the error field near the boundary is more problematic but the error field and mask are still of quality.

Up to now we only compared the quality of the fields but we can also compare the computational load. As seen in Table 1, the most expensive methods are those calculating the exact field (with scaled or unscaled background variances). The hybrid method consumes less time because it does not need the calculation of a covariance function by another Diva calculation but can use an analytical function instead. However, compared to the cost of the almost exact error version, the hybrid method is one order of magnitude more expensive, yet the almost exact error calculation provides error estimates of similar or better quality. Finally, the poor man's error calculations are clearly the fastest and therefore interesting for exploratory work.

d. *Realistic test case*

We finally test the methods with the same data set as the one used in Troupin et al. (2012) so that we can use the same statistical parameters and do not need to recalibrate the analysis. We use salinity measurements in the Mediterranean Sea at a depth of 30 m in July, for the 1980-1990 period and reconstruct the solution on a high resolution output grid with 500×250 grid points.

The analysis itself (Figure 6) shows the well known features such as the inflow of Atlantic waters in Gibraltar, the anticyclonic gyres in the Alboran Sea, the spreading of the Atlantic Waters off the North African coast, the high salinities of the eastern Levantine basin, a signature of Black Sea waters in the Aegean sea and the high salinity in the Northern part of the Western Mediterranean. Also the influence of the Po river in the Northern Adriatic is visible. This analysis itself is calculated within a few seconds and we focus now on the computationally more expensive error

fields.

The error fields are scaled by the global background variance and white crosses indicate real data locations and black dots pseudo-data locations. The real error field (upper panel of Figure 7) shows the effect of low data coverage in the southern parts and the lower errors near data locations. As before, the poor man's error is quite optimistic and quantitatively not reliable. The mask derived from the poor man's error with only 43 incorrectly masked points has some skills, but the clever poor man's error provides more acceptable quantitative results and masks. In particular, the regions void of data in the Southern part and around Sardaigna are now captured. The hybrid method and almost exact approach (Figure 8) have similar metrics, but if we look at the details, the "almost exact" error field clearly better resolves features such as the higher error fields around Sardinia and in the eastern Thyrrenian Sea. Also the error structure in the Alboran Sea is better recovered, despite the very low number of pseudo-data (black dots) used.

For the error fields with boundary effects (Figure 9), using the high pseudo-data coverage along the coast makes it possible to capture the variable background variance, but because of the fine mesh along the coast, probably too many pseudo-data have been added there. This results in excellent metrics, with only four incorrectly masked points and only one percent error on the error field. The relatively large number of pseudo-data is then reflected also in the CPU time. But even with this coverage, the computational gain of a factor 11 compared to the exact calculation is still significant. Comparing CPU times in this realistic case shows without doubt the usefulness of the new approaches (Table 1) which have been included in the Diva tool `http://modb.oce.ulg.ac.be/mediawiki/index.php/DIVA`. Indeed climatology productions generally requires gridding at several levels, month or seasons for several parameters so that already in the 2D case the computational efficiency matters. When it comes to generalizations of our methods to 3D-Var or OI in several dimensions, then the expected gain might be even more interesting as we will show now.

Here we presented some particular test cases and one may wonder how the computational efficiency behaves in other situations, in particular we want to know the computational gain we can

expect for the almost exact error calculation compared to the exact one based on real covariances. In the latter case, for each grid point we need to perform an analysis. For a n-dimensional domain[3] of size D and grid spacing Δ, the number of analysis for the exact error calculation is therefore D/Δ^n. For the almost exact error calculation we cover the domain with random points in which we need to make an analysis. This leads to D/L^n required analyses, to which we have to add the analyses needed at the N_d data locations. We evaluate this number as ϵN_d with $\epsilon = 1$ if we need to calculate A_{ii} in all data points and $\epsilon < 1$ if we use only a fraction ϵ of the observations to calculate the error exactly. We note that $\epsilon = 0$ when a quality check approach using (17) already provided the values of A_{ii}. The gain G we obtain then writes

$$G = \frac{D/\Delta^n}{D/L^n + \epsilon N_d} = \frac{1}{1 + \epsilon \frac{N_d}{\mathcal{N}}} \left(\frac{L}{\Delta}\right)^n \tag{18}$$

where $\mathcal{N} = D/L^n$ is a measure of the degrees of freedom of the background field.

Normally the numerical grids have a grid spacing which is much smaller than the physical length scales and the last term is therefore in favor of a very high efficiency. If we work with a forecast model, its numerical grid is typically recommended to be 8 times smaller than the scales of interest. With only a few data points we then reach gains of one to two order of magnitudes in 2D and almost three order of magnitudes in 3D. The gain decreases if the number of observations is high and allows to capture the degrees of freedom of the system. If the number of observations is much larger than $\mathcal{N}$ it is then advised to use a fraction $\epsilon \sim \mathcal{N}/N_d$ to retain efficiency and still capture the error field.

4. Generalizations

We have presented our ideas in the framework of Diva with a diagonal observational error covariance matrix and will now analyze how the methods can be applied in other frameworks.

A first problem which can be encountered is therefore a non-diagonal observational error covariance matrix $\mathbf{R}$. The clever poor man's error was designed by looking at an isolated data point

[3] We can assume that with a suitable change of variable the different dimensions have been made comparable.

and it was shown reliable in regions with isolated data and at sufficient distance from data clusters. In all these locations it does not matter whether $\mathbf{R}$ is diagonal or not and therefore the clever poor man's error should perform similarly in these regions. The application of the clever poor man's error by any analysis tool (be it OI, 3D-Var or Diva) technically also does not depend on whether or not $\mathbf{R}$ is diagonal since it just demands the application of the analysis tool to a special data vector.

For the almost exact error, we notice that at data locations (7) still holds for non-diagonal $\mathbf{R}$ and that the value of the error at data location i reads

$$\mathbf{e_i}^\mathsf{T}\mathbf{P}_d\,\mathbf{e_i} = \mathbf{e_i}^\mathsf{T}\mathbf{H}\underbrace{\mathbf{B}\mathbf{H}^\mathsf{T}\left(\mathbf{H}\mathbf{B}\mathbf{H}^\mathsf{T} + \mathbf{R}\right)^{-1}}_{\mathbf{K}}\mathbf{R}\mathbf{e_i}. \tag{19}$$

This can be read again from right to left to design the recipe for calculating the errors at the data locations: Extract column i of the observational error covariance matrix (or in other words, fill a vector with observational error covariances with respect to point i), apply your analysis tool ($\mathbf{K}$) and extract the solution at your data point i (observation operator $\mathbf{e_i}^\mathsf{T}\mathbf{H}$). If the tool does not rely on $\mathbf{R}$ but on its inverse, an inversion is needed first, but as observational error covariance matrices are generally block diagonal with narrow bands, this is a feasible operation. Alternatively one can calculate the error reduction term at location i

$$\mathbf{e_i}^\mathsf{T}\mathbf{\Delta}\,\mathbf{e_i} = \mathbf{e_i}^\mathsf{T}\mathbf{H}\underbrace{\mathbf{B}\mathbf{H}^\mathsf{T}\left(\mathbf{H}\mathbf{B}\mathbf{H}^\mathsf{T} + \mathbf{R}\right)^{-1}}_{\mathbf{K}}\mathbf{B}_d\mathbf{e_i} \tag{20}$$

if it is easier to work with the background covariance matrix at the data points. We note that to apply (19), we actually do not even have to know how exactly the background error covariance is expressed, all we have to use is the analysis tool $\mathbf{K}$ applied to a series of "data". One can therefore assess the exact error in a series of points and for the final gridding of the error, the tool can be used again, here even without the need to maintain the correlated observational error since the "field" to be gridded has no observational error anymore.

The presence of a non-diagonal $\mathbf{R}$ therefore still allows the application of the new methods by any analysis tool.

Another problem which can be encountered with other tools than Diva, is to find a way to adapt the correlation length during the clever poor man's error calculation or the final gridding step of the almost exact error approach. If the methods use a length scale in their formulation, then it should simply be adapted according to (11) for the specific correlation function of the method, so that the correlation function with the new scale mimics the squared correlation function. If the method uses an explicitly formulated correlation function which can be changed by the user, then it is suggested to replace the correlation function by its square. This is even simpler and further should improve the quality of the error field. This interpretation also paves the way for situations in which the background error covariance matrix is specified by numerical correlations. During the clever poor man's error calculation or the final gridding step of the almost exact error approach one simply needs to use the squares of the correlations. In other cases, the background covariance can be formulated by recursive filters (*e.g.* Hayden and Purser 1995). Since these filters contain parameters determining the filter width, one can adapt the filter parameters to change the correlation length scales. Some other models work in spectral space and the analysis is also performed in spectral case. In these situations, the spectral representations of covariance functions have a specific signature of the correlation scales. Tampering with the spectrum can therefore be used to change the scales of the underlying covariance function: For example: a Gaussian correlation function in n dimensions which only depends on distance r

$$c(r) = \exp\left(-r^2/L^2\right) \tag{21}$$

has a spectral density $a(k)$ in wave number space k given by

$$a(k) = (\pi L^2)^{n/2} \exp\left(-\pi^2 L^2 k^2\right). \tag{22}$$

To divide the length scale L by a factor $\sqrt{2}$ in two dimensions, it is therefore sufficient to change amplitudes $a(k)$ according to

$$a(k) \leftarrow \frac{a(0)}{2}\sqrt{\frac{a(k)}{a(0)}}. \tag{23}$$

This shows that in order to reduce the correlation length scale, amplitudes of the higher modes get more importance. For spectral models on spheres, the coefficients of the spherical harmonics

also define an underlying correlation function and can be modified to change the correlation length scale.

Still other background covariance specifications rely on projections on empirical orthogonal functions (EOFs). Such EOFs decompositions are to some extend similar to a spectral decompositions, but the base functions are calculated from the data instead of being defined by analytical functions given *a priori*. The equivalent of the spectral density such as (22) is captured in the singular values of the SVD decomposition leading to the EOFs. These coefficients or singular values can therefore be tampered with when a change in correlation length scale is to be obtained.

There are thus several possibilities to change the correlation function of the analysis tool so that it can be optimized to mimic its own square. In complicated implementations the approach should of course be tested, and possibly calibrated, by looking at a covariance function generated by an analysis with a single data point and comparing it to the one obtained when using the tampered version. One should retrieve a correlation function for the tampered version which is close to the square of the original one.

We see that there are many ways to adapt the length scale or correlations for the clever poor man' error calculation and the final gridding step of the almost exact error approach. Should this adaptation be difficult or not efficient, the almost exact error approach can still be applied by covering the domain with more pseudo data and making the final gridding step using the original covariances or a simpler gridding tool. Indeed, the error is already calculated exactly with a fine resolution so that ANY gridding method, even with a poorly specified correlation structure, when applied to these exact values of the error, should work fine. This is however then at the expense of more analyses to get the exact error in more locations.

To illustrate these ideas on an example, we can look at a typical 3D-Var approach used in operational mode, using the so-called NMC method (*e.g.* Parrish and Derber 1992; Fisher 2003), presented here assuming we are still working with anomalies with respect to the background field.

It starts with a definition of a change of variables, with

$$\mathbf{x} = \mathbf{U}\mathbf{v}, \quad \mathbf{B} = \mathbf{U}\mathbf{U}^{\mathsf{T}} \tag{24}$$

so that (3) becomes,

$$2J(\mathbf{v}) = \mathbf{v}^{\mathsf{T}}\mathbf{v} + (\mathbf{HUv} - \mathbf{d})^{\mathsf{T}}\mathbf{R}^{-1}(\mathbf{HUv} - \mathbf{d}). \tag{25}$$

This can be interpreted as a change of variables in which the new state vector $\mathbf{v}$ has uncorrelated background errors. The minimization of (25) is then nicely conditioned if the observational error is large, since the quadratic function is then dominated by the background contribution which is now isotropic and leading to good convergence. The fact that the solution is done via a new variable $\mathbf{v}$ is not essential for our purpose, because after the minimization, the solution $\mathbf{x} = \mathbf{Uv}$ is still the minimum of (25) and hence defines the analysis tool $\mathbf{K}$ exactly as before. Also the presence of a non-diagonal $\mathbf{R}$ causes no problem as shown earlier.

Therefore the only problem we have to deal with is the problem of the correlation length scales or correlation functions which need to be adapted. To do so, we now have to look at how $\mathbf{U}$ is designed. Written as

$$\mathbf{v} = \mathbf{U}^{-1}\mathbf{x} \tag{26}$$

we see that $\mathbf{U}^{-1}$ is supposed to transform the original state vector into one in which the background errors are uncorrelated. This supposes as a first step that physical balances are used to eliminate some variables as a function of others to avoid keeping the associated correlation. Formally $\mathbf{U}_p^{-1}\mathbf{x}$ provides this state vector in which balances have been taken into account. At this step we do not need an adaptation for our methods. From there, the variables are scaled by the local standard deviation of the background field. For spectral methods this requires a transformation to real space, division by the local standard deviation and a transformation back to spectral space, formally by applying $\mathbf{\Sigma}^{-1}$. For a grid method, the operation simply divides by the local standard deviation. Here again our methods do not introduce any change. Then the horizontal and vertical correlations need to be taken into account by successively trying to take out correlations in the horizontal (formally operation $\mathbf{U}_h^{-1}$) and the vertical directions (operation $\mathbf{U}_v^{-1}$). This will involve the correlations which we will have to tamper with. The operations read now

$$\mathbf{v} = \mathbf{U}_v^{-1}\mathbf{U}_h^{-1}\mathbf{\Sigma}^{-1}\mathbf{U}_p^{-1}\mathbf{x} \tag{27}$$

which defines $\mathbf{U} = \mathbf{U}_p\mathbf{\Sigma}\mathbf{U}_h\mathbf{U}_v$ and finally $\mathbf{B} = \mathbf{U}_p\mathbf{\Sigma}\mathbf{U}_h\mathbf{U}_v\mathbf{U}_v^{\mathsf{T}}\mathbf{U}_h^{\mathsf{T}}\mathbf{\Sigma}^{\mathsf{T}}\mathbf{U}_p^{\mathsf{T}}$. Obviously, in practice $\mathbf{B}$ is never formed but the succession of operations described above are applied to the state vector $\mathbf{x}$, then the minimization is performed on $\mathbf{v}$ and finally the optimal $\mathbf{x}$ retrieved by the inverse operations in reverse order:

$$\mathbf{x} = \mathbf{U}_p\mathbf{\Sigma}\mathbf{U}_h\mathbf{U}_v\mathbf{v}. \tag{28}$$

There remains to see how to adapt $\mathbf{U}_h$ and $\mathbf{U}_v$ to accommodate changes in correlations. When the model works in spectral space, it is generally assumed that the modes are independent and matrix $\mathbf{U}_h$ is diagonal. The spectral coefficients found on the diagonal of $\mathbf{U}_h$ define in this case the underlying horizontal correlation function in physical space and by changing their values we can change the correlation function as shown in example (23). If the model works in grid space, then $\mathbf{U}_h$ can be specified by recursive filters or covariance functions, which can also be changed to meet our requirements. Finally on the vertical, $\mathbf{U}_v\mathbf{U}_v^{\mathsf{T}}$ is in fact a vertical correlation matrix. It is generally considered a block diagonal matrix (one block for each variable and each spectral mode or spatial grid point) and is composed therefore by a series of small $N_z \times N_z$ correlation matrices, where N_z is the number of vertical levels. These individual matrices must be adapted in our case to change the correlation functions. This can be done by taking for example the square of the correlations. If the matrices are already decomposed by a singular value decomposition to work with EOFs, as stated above, one can tamper the singular values to change the correlations.

It is now clear that the adaptations to change the correlations are quite localized and therefore it should be possible to implement the poor man's error and the almost exact error calculations in operational 3D-Var implementations. We can finally note that in the NMC version, the parameters involved in $\mathbf{U}$ are fitted by assuming that statistics on differences in forecasts for the same moment but of different length (24h and 48h) are a good proxy for background errors. This calibration does not affect the possibility to readjust later the correlation during the poor man's error calculation or the final gridding of the almost exact error.

5. Conclusion

The preparation of error fields is generally much more expensive than the preparation of an analysis. We proposed two new ideas to provide some practical and economic ways to provide such error fields. The first method only needs a second analysis with modified correlation length scale and is particularly well suited for exploratory analysis or masking of gridded fields in regions insufficiently covered by data (such as done in the web version (Barth et al. 2010) or within ODV (Schlitzer 2013). The second method on the other hand can be used for cases in which sufficient confidence in the covariance matrices justifies the use of the full error calculation. In this case, the new method we presented drastically reduces the computational burden without sacrifying on the quality of the error field. The method is particularly useful when employed in parallel with outliers detection methods and cross validation as the same computations can be reused.

We illustrated the approach using the specific analysis tool Diva, but also paved the way for generalizations for a variety of situations when background covariances are formulated differently or when the observational error covariance matrix is non-diagonal. The ideas presented here can therefore be implemented in various versions of analysis tools.

In particular we detailed how both methods can be adapted to 3D-Var approaches used in operational systems. They could then provide an alternative to the Lanczos-vector based estimates of the Hessian matrix. The new approach is particularly interesting if the background covariance is factorized or a very efficient preconditioning was applied so that the calculation of several minimizations to get error estimates in selected locations can be tackled.

Concerning future work in the context of Diva, in the present paper we limited ourselves to the implementation of the case of uncorrelated observational error, i.e. a diagonal $\mathbf{R}$. Dealing with non diagonal $\mathbf{R}$ is already more problematic with Diva for the analysis itself. When data are provided with regular spatial patterns (such as along altimeter tracks or on satellite images), augmented data arrays can be used to account for correlated observational errors in methods that only deal with diagonal matrices for the observation error covariance (Brankart et al. 2009). This problem will be looked at in the future. Finally there is also some room for improvement in Diva in case one

514 is interested in the unscaled error fields showing the boundary effects by reducing the number of
515 pseudo points near the boundaries if the computational load is too high and meshes very fine. The
516 choice of the location of the additional pseudo data could also be further optimized when other
517 constraints are used, such as the advection constraint already included in Diva.

518 *Acknowledgments.*

519 Diva has received funding from the European Union Seventh Framework Programme (FP7/2007-
520 2013) under grant agreement No. 283607, SeaDataNet 2, and from project EMODNET (MARE/2008/03
521 - Lot 3 Chemistry - SI2.531432) from the Directorate-General for Maritime Affairs and Fisheries.
522 This research was also supported by the SANGOMA Project (European FP7-SPACE-2011 project,
523 Grant 283580). The F.R.S.-FNRS is acknowledged for providing supercomputing access. This is
524 a MARE publication.

REFERENCES

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

Barth, A., J.-M. Beckers, C. Troupin, and A. Alvera-Azcárate, 2013: Divand: n-dimensional variational data analysis for ocean climatologies. *submitted*.

Beckers, J.-M., A. Barth, and A. Alvera-Azcárate, 2006: DINEOF reconstruction of clouded images including error maps. Application to the sea-surface temperature around Corsican island. *Ocean Science*, **2**, 183–199, doi:10.5194/os-2-183-2006.

Bekas, C., E. Kokiopoulou, and Y. Saad, 2007: An estimator for the diagonal of a matrix. *Applied numerical mathematics*, **57 (11)**, 1214–1229.

Bouttier, F. and P. Courtier, 2002: Data assimilation concepts and methods March 1999. *Meteorological training course lecture series. ECMWF*.

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

Brankart, J.-M., C. Ubelmann, C.-E. Testut, E. Cosme, P. Brasseur, and J. Verron, 2009: Efficient parameterization of the observation error covariance matrix for square root or ensemble Kalman filters: application to ocean altimetry. *Monthly Weather Review*, **137 (6)**, 1908–1927.

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., J.-M. Beckers, J.-M. Brankart, and R. Schoenauen, 1996: Seasonal temperature and salinity fields in the Mediterranean Sea: Climatological analyses of a historical data set. *Deep-Sea Research I*, **43 (2)**, 159–192, doi:10.1016/0967-0637(96)00012-X.

Bretherton, F. P., R. E. Davis, and C. Fandry, 1976: A technique for objective analysis and design of oceanographic instruments applied to MODE-73. *Deep-Sea Research*, **23**, 559–582, doi:10.1016/0011-7471(76)90001-2.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, **120**, 1367–1387, doi:10.1002/qj.49712051912.

Delhomme, J., 1978: Kriging in the hydrosciences. *Advances in Water Resources*, **1 (5)**, 251–266, doi:10.1016/0309-1708(78)90039-8.

Emery, W. J. and R. E. Thomson, 2001: *Data analysis methods in physical oceanography*. Elsevier.

Fischer, C., T. Montmerle, L. Berre, L. Auger, and S. E. STEFANESCU, 2005: An overview of the variational assimilation in the ALADIN/France numerical weather-prediction system. *Quarterly Journal of the Royal Meteorological Society*, **131 (613)**, 3477–3492, doi:10.1256/qj.05.115.

Fisher, M., 2003: Background error covariance modelling. *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*, 45–63.

Gandin, L. S., 1965: Objective analysis of meteorological fields. Tech. rep., Israel Program for Scientific Translations, Jerusalem.

Girard, D. A., 1998: Asymptotic comparison of (partial) cross-validation, GCV and randomized GCV in nonparametric regression. *The Annals of Statistics*, **26 (1)**, 315–334.

Hamill, T. M. and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Monthly Weather Review*, **128 (8)**, 2905–2919.

Hayden, C. M. and R. J. Purser, 1995: Recursive filter objective analysis of meteorological fields: Applications to nesdis operational processing. *Journal of Applied Meteorology*, **34 (1)**, 3–15.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation and predictability*. Cambridge university press.

Kaplan, A., Y. Kushnir, and M. A. Cane, 2000: Reduced space optimal interpolation of historical marine sea level pressure: 1854-1992. *Journal of Climate*, **13 (16)**, 2987–3002, doi:10.1175/1520-0442(2000)013⟨2987:RSOIOH⟩2.0.CO;2.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, **112**, 1177–1194, doi:10.1002/qj.49711247414.

McIntosh, P. C., 1990: Oceanographic data interpolation: objective analysis and splines. *Journal of Geophysical Research*, **95**, 13 529–13 541.

Moore, A. M., H. G. Arango, G. Broquet, B. S. Powell, A. T. Weaver, and J. Zavala-Garay, 2011: The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems: Part I - System overview and formulation. *Progress in Oceanography*, **91 (1)**, 34 – 49, doi:10.1016/j.pocean.2011.05.004.

Parrish, D. and J. Derber, 1992: The national meteorological center's spectral statistical interpolation analysis system. *Monthly Weather Review*, **120**, 1747–1763.

Rabier, F. and P. Courtier, 1992: Four-dimensional assimilation in the presence of baroclinic instability. *Quarterly Journal of the Royal Meteorological Society*, **118**, 649–672, doi:10.1002/qj.49711850604.

Reynolds, R. W. and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *Journal of Climate*, **7 (6)**, 929–948.

Rixen, M., J.-M. Beckers, J.-M. Brankart, and P. Brasseur, 2000: A numerically efficient data

593 analysis method with error map generation. *Ocean Modelling*, **2 (1-2)**, 45–60, doi:10.1016/
594 S1463-5003(00)00009-3.

595 Schlitzer, R., 2013: Ocean Data View. URL `http://odv.awi.de`.

596 Seaman, R. S. and M. Hutchinson, 1985: Comparative real data tests of some objective analysis
597 methods by withholding observations. *Aust. Meteor. Mag.*, **33**, 37–46.

598 Tang, J. M. and Y. Saad, 2012: A probing method for computing the diagonal of a matrix inverse.
599 *Numerical Linear Algebra with Applications*, **19 (3)**, 485–501.

600 Troupin, C., F. Machín, M. Ouberdous, D. Sirjacobs, A. Barth, and J.-M. Beckers, 2010: High-
601 resolution climatology of the North-East Atlantic using Data-Interpolating Variational Analysis
602 (Diva). *Journal of Geophysical Research*, **115**, C08 005, doi:10.1029/2009JC005512.

603 Troupin, C., et al., 2012: Generation of analysis and consistent error fields using the Data Inter-
604 polating Variational Analysis (Diva). *Ocean Modelling*, **52-53**, 90–101, doi:10.1016/j.ocemod.
605 2012.05.002.

606 Troupin, C., et al., 2013: Diva: a gridding software for geophysical data. Part I: determination of
607 analysis parameters. *submitted*.

608 Wahba, G. and J. Wendelberger, 1980: Some new mathematical methods for variational objective
609 analysis using splines and cross validation. *Monthly Weather Review*, **108**, 1122–1143.

610 Xiang, D. and G. Wahba, 1996: A generalized approximate cross validation for smoothing splines
611 with non-Gaussian data. *Statistica Sinica*, **6**, 675–692.

List of Tables

Method	Schematic case	Realistic case
Real covariance	572.4	1944.9
Poor man	3.8	6.1
Clever poor man	8.3	13.4
Hybrid	345.3	1256.6
Almost exact	37.2	81.4
Real covariance bnd	568.3	1951.9
Almost exact bnd	33.7	175.1

TABLE 1. CPU time (in seconds) for the test case with 150 data points distributed randomly in part of the domain (schematic case) and a realistic case of the Mediterranean Sea

List of Figures

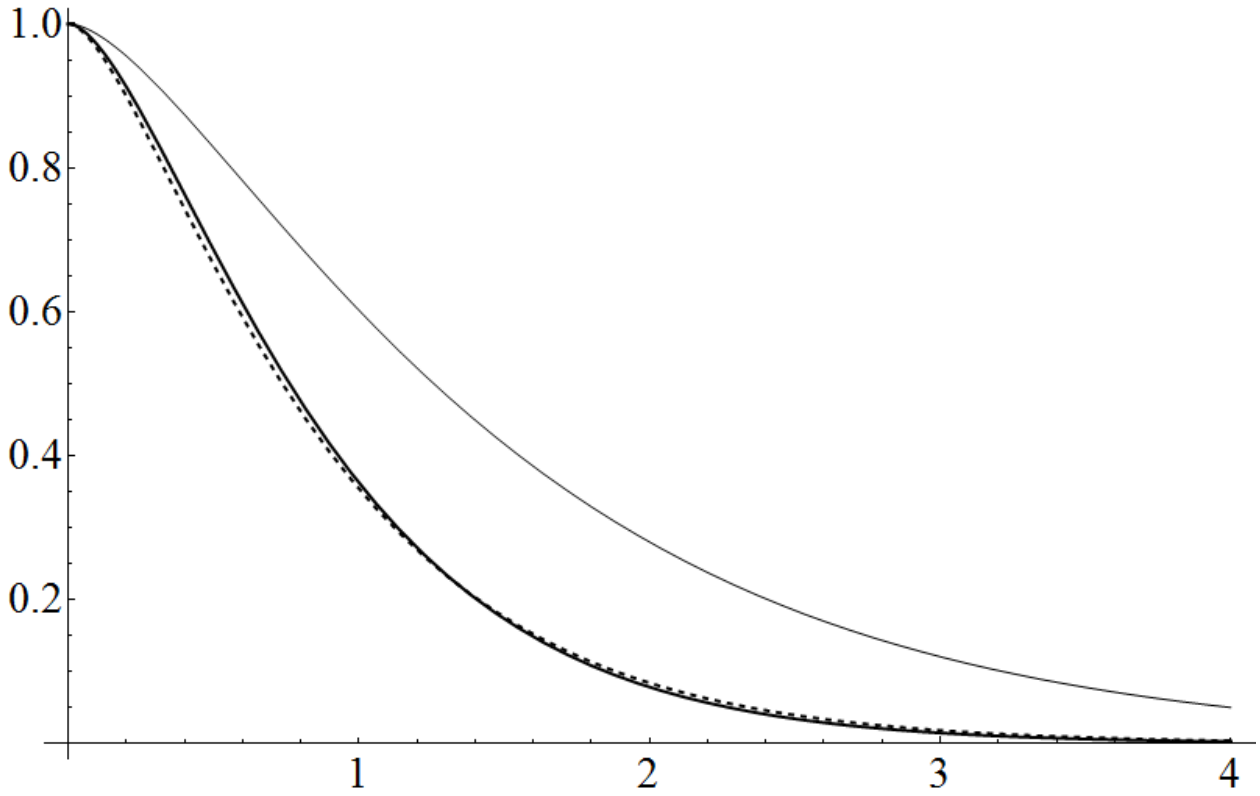

FIG. 1. Diva correlation function (the kernel of the Diva functional) in an infinite domain as a function of r/L (thin line). The squared correlation function which leads to the exact error reduction for one data point (thick line) shows how strongly the poor man's error using the thin line overestimates the error reduction. Adapting the correlation length scale $c(r/L')$ (dashed line) in the poor man's error (called clever poor man's error) shows how one can mimic the exact squared correlation by comparing the thick line (exact error reduction) and the dashed line (clever poor man's error reduction).

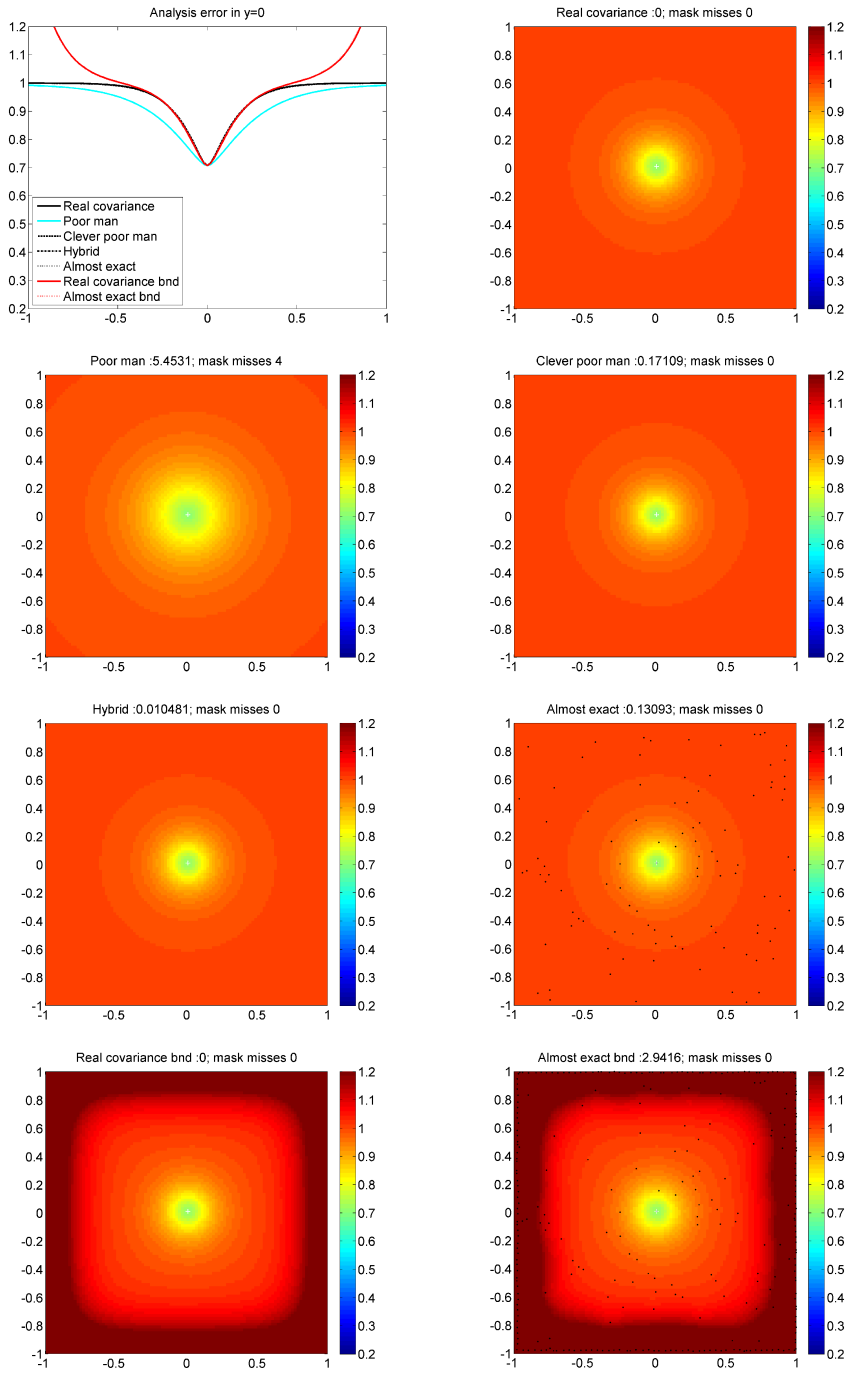

FIG. 2. Test case with a single point in the center of the domain. The error standard deviation is shown for the different methods. The upper-left panel is a section along $y = 0$. The title for each 2D plot identifies the method and includes two indicators of the quality of the error field. The first number is the relative error on the error field as a percentage, where the true field is the field *real covariance* when the error is scaled by the local background variance. For the case where boundary effects are taken into account the reference solutions is *real covariance bnd*. The second indicator gives the number of grid points where a mask derived from the error field is not the same as the exact one. White crosses indicate real data locations and black dots indicate pseudo-data locations.

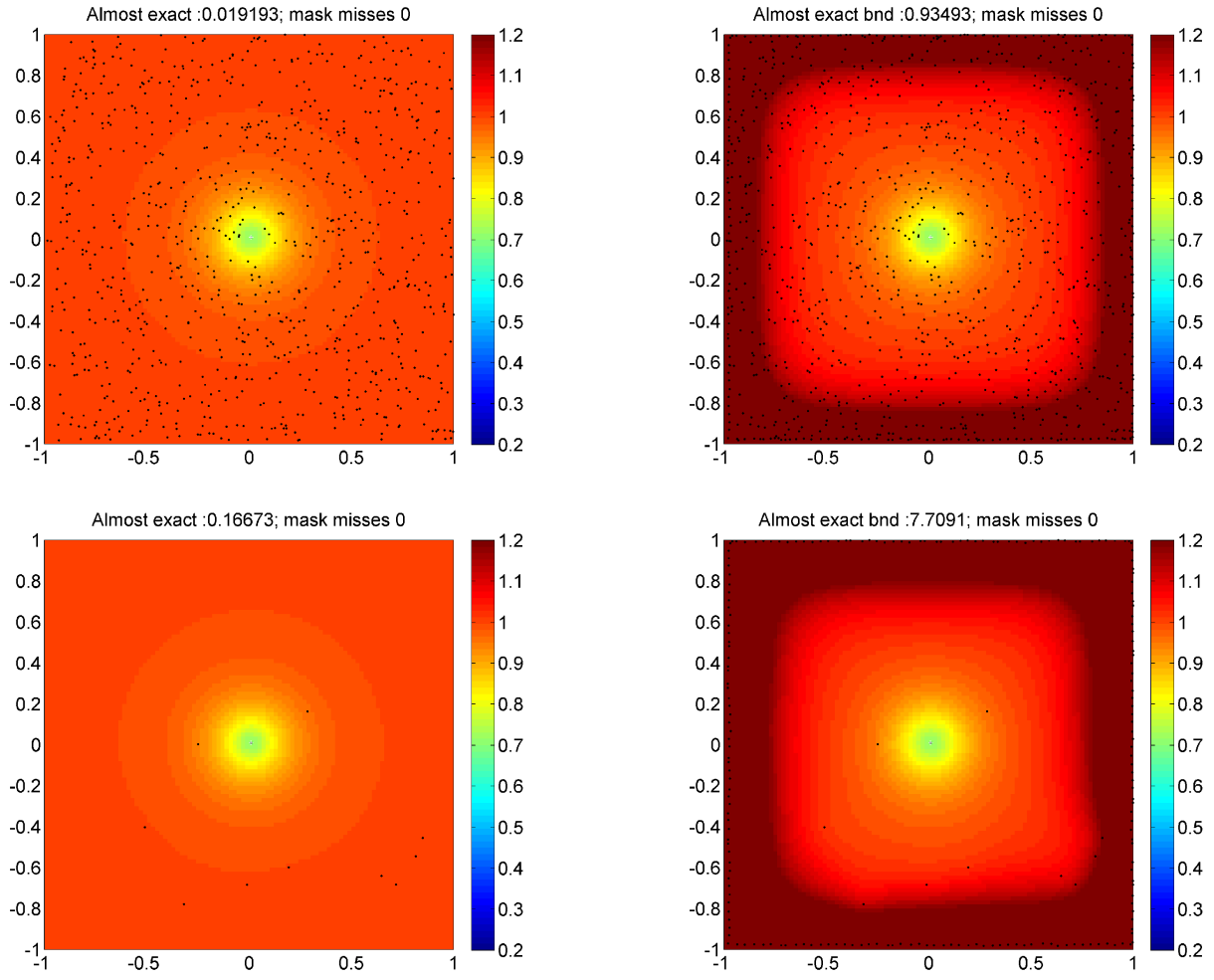

FIG. 3. Error fields for a single point in center with fine sampling $\alpha = 3$ of pseudo data (upper two plots) and coarse sampling $\alpha = 0.3$ (lower two plots). White crosses indicate real data locations and black dots pseudo-data locations.

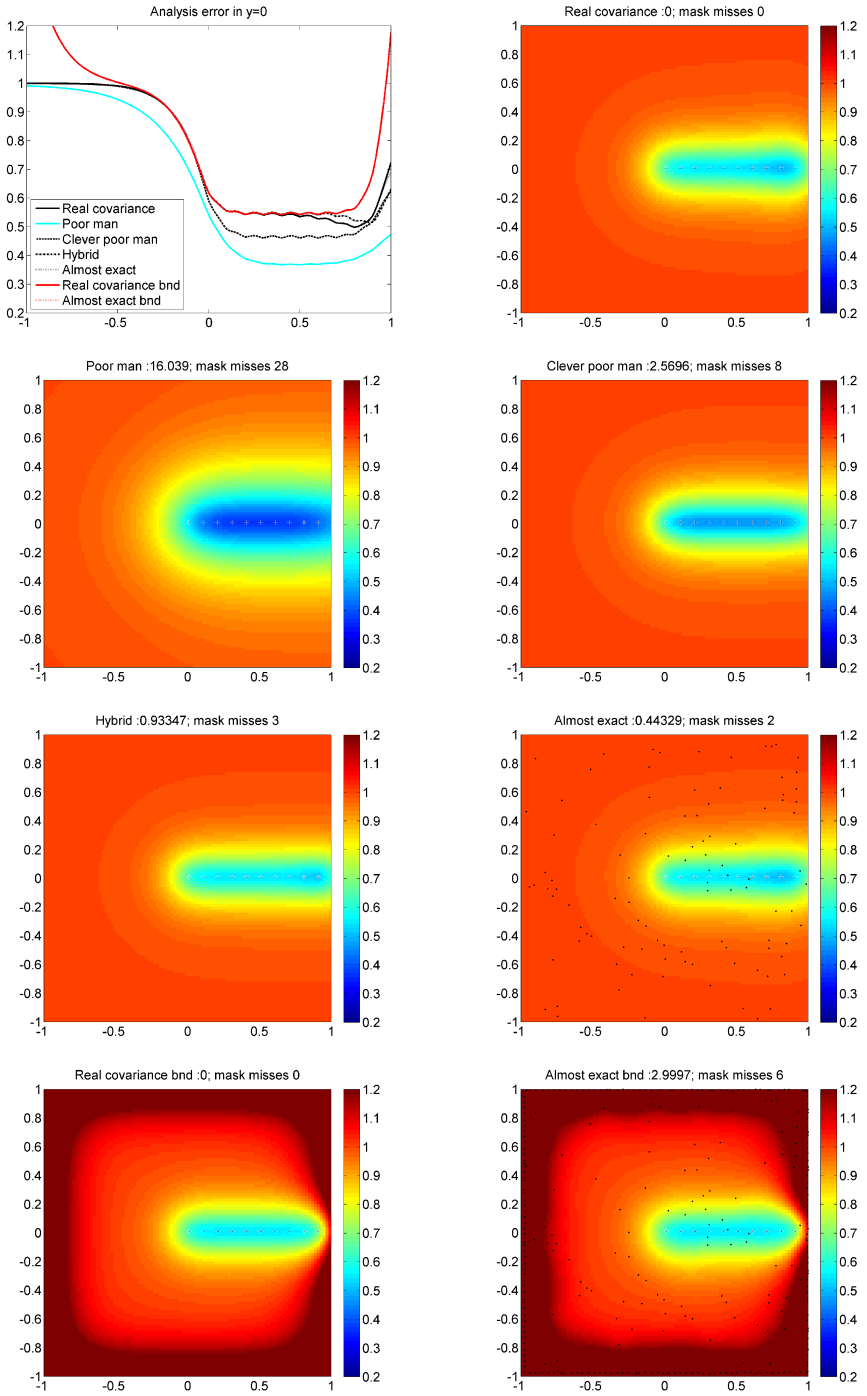

FIG. 4. Error fields for ten data points in $y = 0, x \geq 0$. White crosses indicate real data locations and black dots pseudo-data locations.

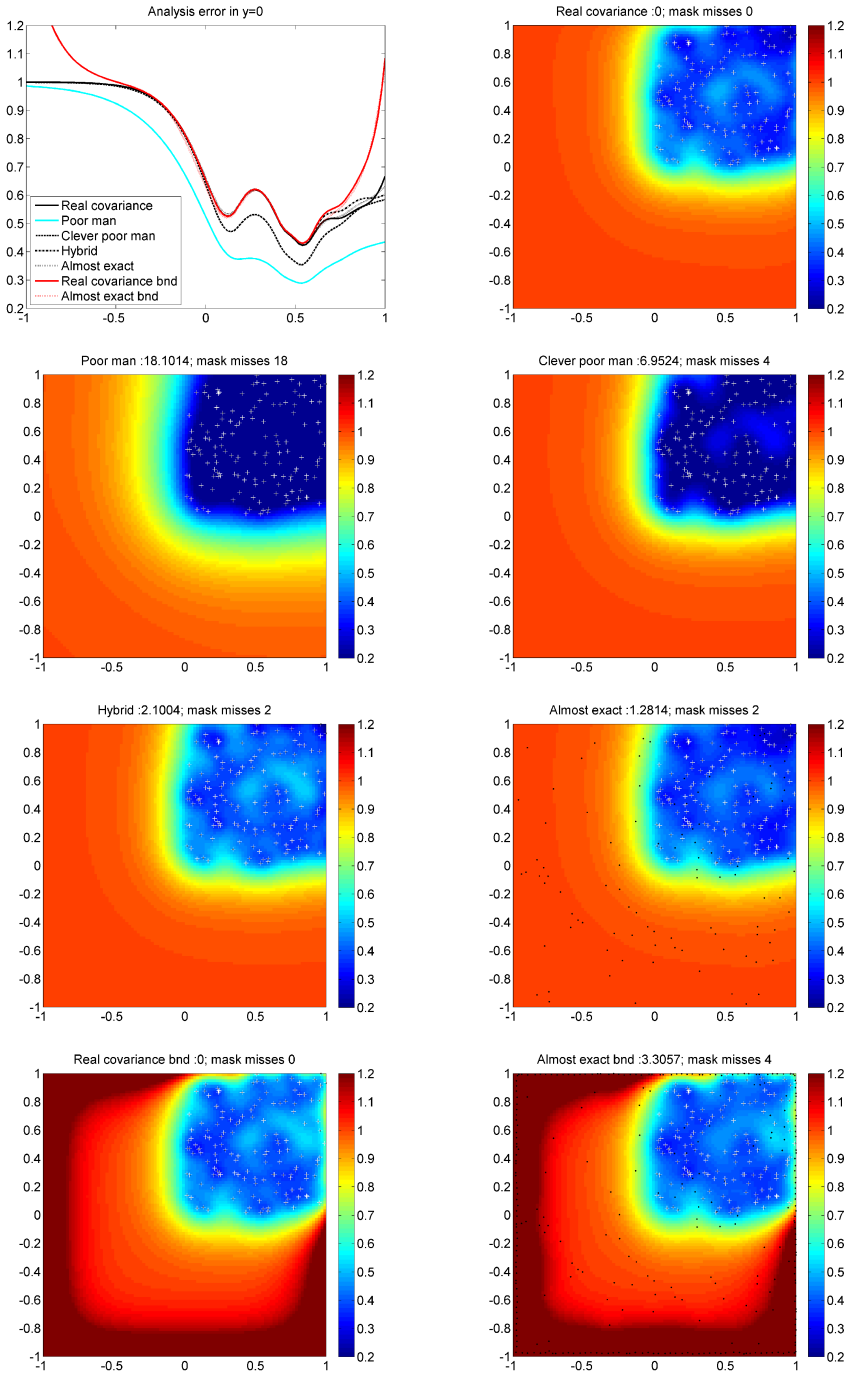

FIG. 5. Error fields for 150 random points in one quadrant. White crosses indicate real data locations and black dots pseudo-data locations.

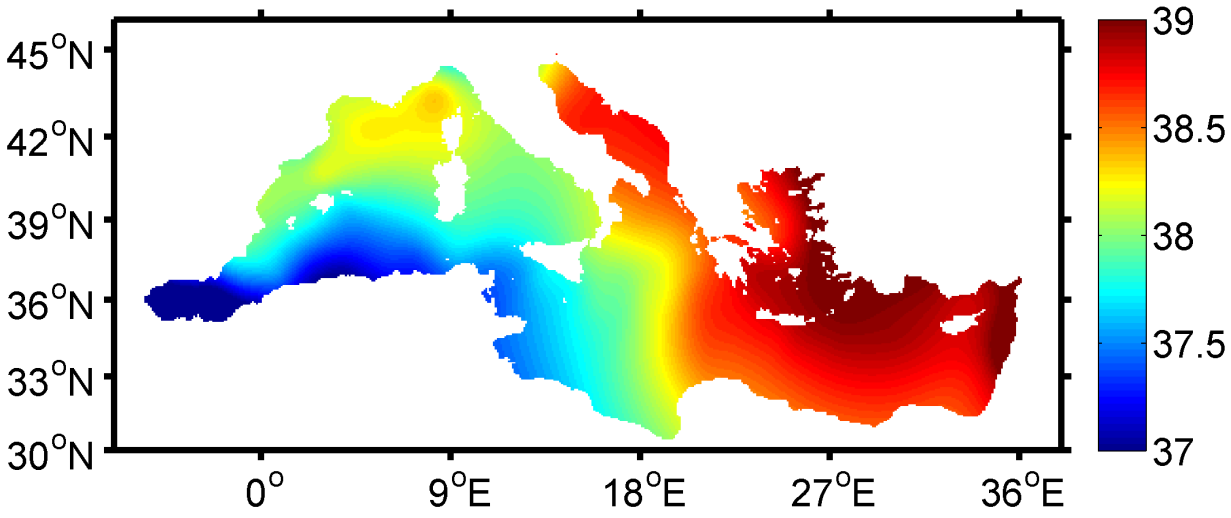

FIG. 6. Analysis of salinity measurements in the Mediterranean Sea at a depth of 30 m in July, for the 1980-1990 period.

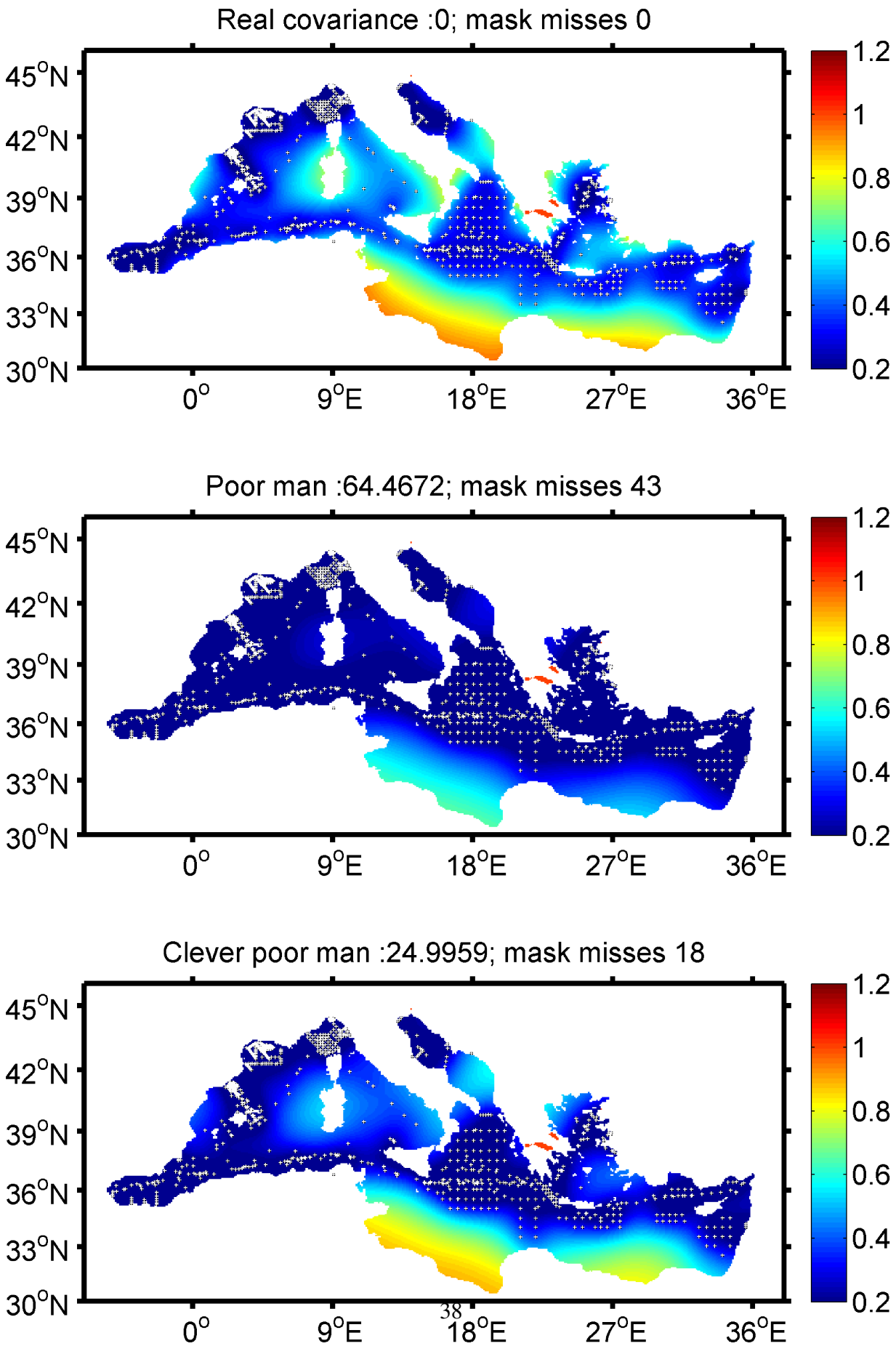

FIG. 7. Real error field, poor man's and clever poor man's error.

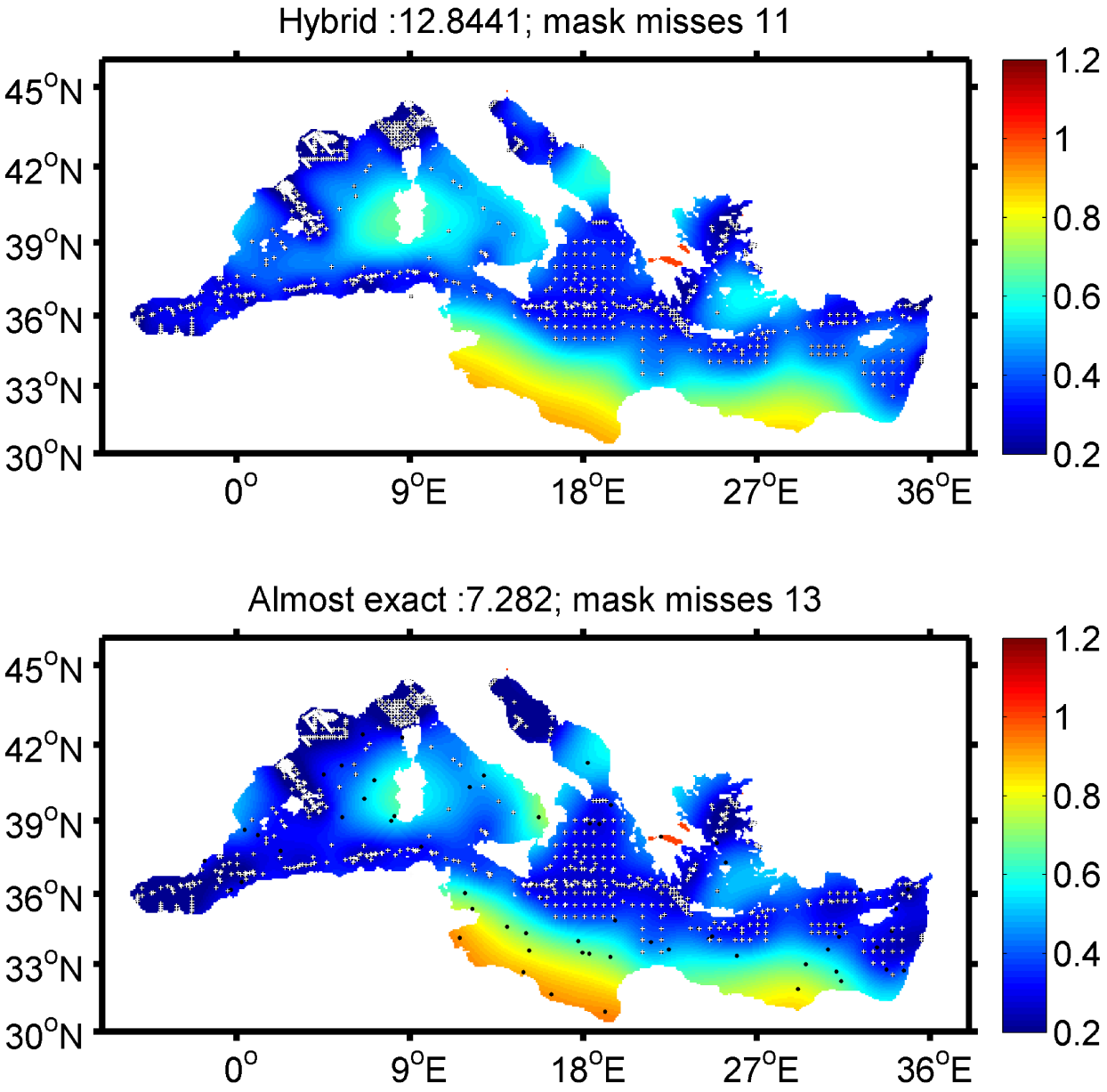

FIG. 8. Hybrid and almost exact approach.

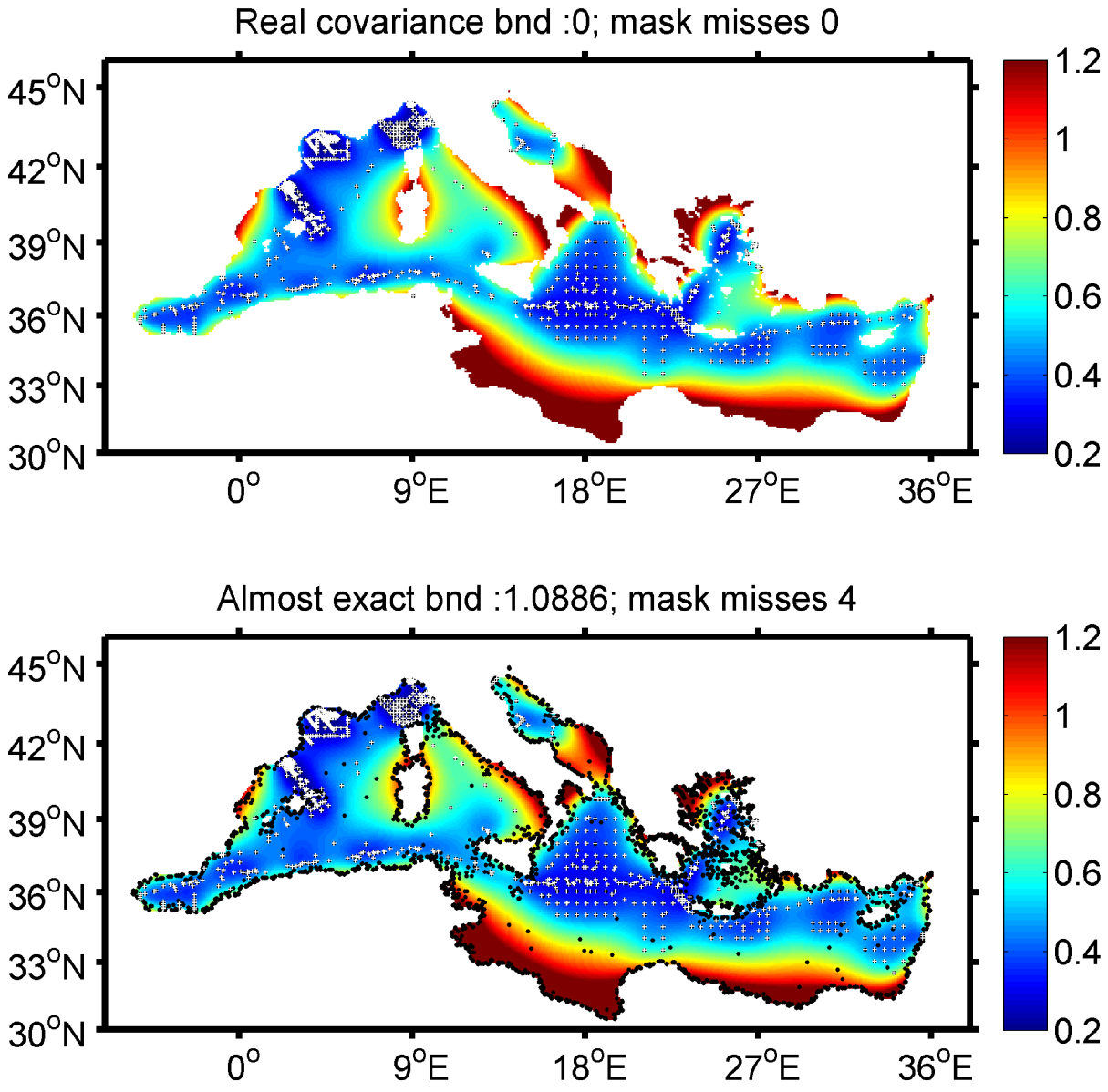

FIG. 9. Real error with boundary effects and almost exact approach.