



divand-1.0: n -dimensional variational data analysis for ocean observations

A. Barth^{1,*}, J.-M. Beckers¹, C. Troupin², A. Alvera-Azcárate¹, and L. Vandembulcke^{3,4}

¹GHER, University of Liège, Liège, Belgium

²IMEDEA, Esporles, Illes Balears, Spain

³seamod.ro/Jailoo srl, Sat Valeni, Com. Salatrucu, Jud. Arges, Romania

⁴CIIMAR, University of Porto, Porto, Portugal

* *Invited contribution by A. Barth, recipient of the EGU Arne Richter Award for Outstanding Young Scientists 2010.*

Correspondence to: A. Barth (a.barth@ulg.ac.be)

Received: 7 June 2013 – Published in Geosci. Model Dev. Discuss.: 23 July 2013

Revised: 18 October 2013 – Accepted: 12 December 2013 – Published: 29 January 2014

Abstract. A tool for multidimensional variational analysis (*divand*) is presented. It allows the interpolation and analysis of observations on curvilinear orthogonal grids in an arbitrary high dimensional space by minimizing a cost function. This cost function penalizes the deviation from the observations, the deviation from a first guess and abruptly varying fields based on a given correlation length (potentially varying in space and time). Additional constraints can be added to this cost function such as an advection constraint which forces the analysed field to align with the ocean current. The method decouples naturally disconnected areas based on topography and topology. This is useful in oceanography where disconnected water masses often have different physical properties. Individual elements of the a priori and a posteriori error covariance matrix can also be computed, in particular expected error variances of the analysis. A multidimensional approach (as opposed to stacking two-dimensional analysis) has the benefit of providing a smooth analysis in all dimensions, although the computational cost is increased.

Primal (problem solved in the grid space) and dual formulations (problem solved in the observational space) are implemented using either direct solvers (based on Cholesky factorization) or iterative solvers (conjugate gradient method). In most applications the primal formulation with the direct solver is the fastest, especially if an a posteriori error estimate is needed. However, for correlated observation errors the dual formulation with an iterative solver is more efficient.

The method is tested by using pseudo-observations from a global model. The distribution of the observations is based on the position of the Argo floats. The benefit of the three-dimensional analysis (longitude, latitude and time) compared to two-dimensional analysis (longitude and latitude) and the role of the advection constraint are highlighted. The tool *divand* is free software, and is distributed under the terms of the General Public Licence (GPL) (<http://modb.oce.ulg.ac.be/mediawiki/index.php/divand>).

1 Introduction

Deriving a complete gridded field based on a set of observations is a common problem in oceanography. In situ observations are generally sparse and inhomogeneously distributed. While satellite observations have typically a better spatial and temporal coverage (but measure only surface data) than in situ data, they present also gaps due to, e.g. the presence of clouds (in the case of thermal sea-surface temperature and optical surface properties of the ocean). Since the problem is generally under-determined, if the gridded field is to be derived from the observations alone, a first guess is introduced. The data analysis problem is also closely related to data assimilation where the observations are used in combination with a first guess coming from a model.

Several interpolation methods have been developed and presented in the scientific literature. Direct linear interpolation of the observations is rarely an option for ocean observations which are affected by noise and are not necessarily representative (e.g. a measurement at a specific time might not be representative for a monthly average). Also the gradient of the interpolated field is not necessarily continuous. Current interpolation methods take therefore, in one way or the other, the uncertainty of the observations into account. Most interpolation methods of uncertain observations can be classified as methods based on optimal interpolation (including kriging) and variational analysis.

For optimal interpolation methods (Gandin, 1965; Bretherton et al., 1976), the error covariance of the first guess is generally directly specified by analytical functions (Robinson, 1996; Guinehut et al., 2002; Roberts-Jones et al., 2012). When satellite or model data are used, this error covariance can also be specified by its eigenvalues/eigenvectors (Kaplan et al., 1997; Rayner et al., 2003; Beckers et al., 2006) or by an ensemble (Evensen, 2007). Applications to multiple spatial and/or temporal dimensions are common (Høyer and She, 2007; Buongiorno Nardelli et al., 2010) to ensure a continuity of the solution along those dimensions. Analytical functions for the error covariance are based generally on the distance between two given points. However, decoupling water masses separated by land and maintaining at the same time a spatially smooth field over the ocean is difficult in optimal interpolation.

Auxiliary variables can be used as additional dimensions in order to improve the realism of the covariance function. Buongiorno Nardelli (2012) used for instance temperature (based on satellite sea surface temperature) as an additional dimension to generate a sea surface analysis of salinity using optimal interpolation. This innovative application shows that dimensions are not necessarily restricted to space and time and that other related variables can be used to extend the notion of space and distance.

In variational analysis, a cost function is introduced which penalizes non-desirable properties of the analysed field, in particular the deviation from observations and from the background estimate and abrupt changes of the analysed fields. The variational approach is equivalent to the optimal interpolation formulation, but instead of specifying directly the error covariance of the first guess, the inverse of this matrix is parameterized: rather than imposing that two adjacent grid cells are correlated to each other, it is required that gradients (and higher-order derivatives) are small (Brasseur et al., 1996; Brankart and Brasseur, 1996). Decoupling water masses separated by land is natural in variational analysis as it can be included using boundary conditions on the spatial derivatives. Variational analysis in three or four dimensions is common in the context of data assimilation (Rabier et al., 2000; Dobricic and Pinardi, 2008; Moore et al., 2011b) where it is generally a multivariate analysis. In four-dimensional variational analysis, the analysed field has in general three dimensions (e.g.

longitude, latitude and depth for initial conditions or longitude, latitude and time for forcing fields) but they involve the numerical model and its adjoint to derive the relationship between the initial condition (or forcing fields) and observations at different time instances. However most of the data analysis applications to grid observations using variational methods are limited to two dimensions: either two horizontal dimensions (e.g. Troupin et al., 2010) or vertical transects (Yari et al., 2012). Three or four dimensional (space and time) fields are then obtained by assembling individual analyses. Inhomogeneous data distribution might then lead to spurious abrupt variations along the additional dimensions, which require ad hoc filtering of the assembled field. The presented tool implements a monovariate analysis. However the information from additional environmental variables can be incorporated as additional constraints.

The variational approach is also attractive for problems where it is easier to formulate physical properties of the underlying field in terms of constraints than in terms of correlation/covariance. For a two-dimensional surface current analysis for example, one can impose that the horizontal divergence is small (Legler and Navon, 1991; Bennett et al., 1997; Yaremchuk and Sentchev, 2009) by adding a corresponding term to the cost function. This kind of constraint would be more difficult to implement in an interpolation method.

On the other hand, in optimal interpolation one can quite easily derive the error variance of the analysed fields, which is more difficult but feasible for variational methods (Troupin et al., 2012). Optimal interpolation in the local approximation can also be quite efficiently applied to distributed-memory parallel computing architectures.

The aim of this manuscript is to implement and test a variational analysis program that can operate in an arbitrary high dimensional space and with a cost function that can be easily extended with additional constraints. The benefit of this method will be assessed in comparison to assembled two-dimensional analyses using an advection constraint forcing the gradients of an analysis to be aligned with a given vector field.

Different approaches can be used to model the background error covariance. In the smoothing norm splines approach (Wahba and Wendelberger, 1980; Wahba, 1990; McIntosh, 1990; Brasseur and Haus, 1991; Brasseur et al., 1996; Troupin et al., 2012) a norm is defined using squares of derivatives up to a certain degree. The background error covariance can also be specified by solving the diffusion equations iteratively over a pseudo-time dimension. The solution can be obtained using explicit (Bennett et al., 1996; Weaver et al., 2003; Lorenzo et al., 2007; Muccino et al., 2008; Moore et al., 2011b) or implicit stepping schemes (Bennett et al., 1997; Weaver and Courtier, 2001; Chua and Bennett, 2001; Zaron et al., 2009; Carrier and Ngodock, 2010). Another commonly used technique is based on a recursive filter (Lorenç, 1992; Hayden, 1995; Purser et al., 2003a, b; Martin et al., 2007; Dobricic and Pinardi, 2008). Mirouze

and Weaver (2010) have shown that this approach is actually equivalent to the method based on the implicit diffusion operator. For variational data assimilation the approach based on diffusion operators (or recursive filter) is commonly used, as it allows to define the background error covariance (or its inverse) in terms of square root matrices which is often used for preconditioning (e.g. Lorenc, 1997; Habén et al., 2011). As DIVA (Data-Interpolating Variational Analysis), the n -dimensional tool `divand` also uses norm splines to define the background error covariance in combination with a direct solver which does not require preconditioning. In the case of DIVA the skyline method (Dhatt and Touzot, 1984) is used, while `divand` uses the supernodal sparse Cholesky factorization (Chen et al., 2008; Davis and Hager, 2009). Other iterative methods (which require preconditioning) have also been implemented.

The variational inverse method (Brasseur and Haus, 1991) implemented in the tool DIVA computes the minimum of the cost function in two dimensions using a triangular finite-element mesh (Brasseur et al., 1996; Brankart and Brasseur, 1996, 1998; Troupin et al., 2012). A web interface has also been developed for this tool (Barth et al., 2010). We present an extension to n -dimensions which is called `divand`. To simplify the testing and implementation in an arbitrarily high dimensional space, a regular curvilinear mesh is used for the tool `divand`.

In Sect. 2, the formulation of the method in n -dimensional space and the derivation of the analytical kernel functions for an infinitely large domain are introduced. The relationship between the highest derivative needed in the formulation and the dimension of the domain is shown. Section 3 presents the different implemented algorithms. Simple numerical tests are performed in Sect. 4 to show the consistency of the numerical results with the analytical solutions of Sect. 2. Implementation details and capabilities of the tool are given in Sect. 5. The tool is also tested in a realistic configuration to reconstruct global temperature in Sect. 6.

2 Formulation

Variational inverse methods aim to derive a continuous field which is close to the observations and satisfies a series of a priori constraints. In particular, the field should be “smooth”. It is therefore important to quantify the “smoothness” of a field. While the interpolated field should be close to observations, it should not necessarily pass through all observations because observations have errors and often do not represent the same information. A cost function is formulated which includes these constraints:

$$J[\varphi] = \sum_{j=1}^{N_d} \mu_j [d_j - \varphi(\mathbf{x}_j)]^2 + \|\varphi - \varphi_b\|^2 + J_c(\varphi), \quad (1)$$

where d_j are the measurements at the location \mathbf{x}_j and μ_j is their weight, and φ_b is a background estimate of the field.

The term $J_c(\varphi)$ represents potentially additional constraints which will be specified later.

In order to define the norm $\|\cdot\|$, the length scale L_i in every domain dimension is introduced. These length scales form the diagonal elements of the matrix \mathbf{L} :

$$\mathbf{L} = \begin{pmatrix} L_1 & 0 & & \\ 0 & L_2 & & \\ & & \ddots & \\ & & & \ddots \end{pmatrix}. \quad (2)$$

Based on these length-scales, we define the following scaled differential operators for the gradient and Laplacian:

$$\tilde{\nabla} = \mathbf{L}\nabla, \quad (3)$$

$$\tilde{\nabla}^2 = \nabla \cdot (\mathbf{L}^2 \nabla). \quad (4)$$

A scalar product $\langle f, g \rangle$ of two functions, f and g , is defined using the scaled gradient and Laplacian.

$$\begin{aligned} \langle f, g \rangle = & \frac{1}{c} \int_D \alpha_0 f g + \alpha_1 (\tilde{\nabla} f) \cdot (\tilde{\nabla} g) + \alpha_2 (\tilde{\nabla}^2 f) (\tilde{\nabla}^2 g) \\ & + \alpha_3 (\tilde{\nabla} \tilde{\nabla}^2 f) \cdot (\tilde{\nabla} \tilde{\nabla}^2 g) + \dots d\mathbf{x} \end{aligned} \quad (5)$$

This scalar product is closely related to the spline norm defined by Wahba and Wendelberger (1980) and McIntosh (1990) for two dimensions.

Equation (5) corresponds to the covariance inner product $\mathbf{x}^T \mathbf{B}^{-1} \mathbf{y}$ where the vectors \mathbf{x} and \mathbf{y} represent the fields f and g and the matrix \mathbf{B} is the background error covariance (Gaspari and Cohn, 1999).

We note m the highest derivative in this scalar product. The parameter c is a normalization coefficient that will be chosen later. The coefficients α_i are generally considered positive so that the cost function has certainly a finite minimum.

The norm used in the background constraint of Eq. (1) is defined using this scalar product by

$$\|\varphi\|^2 = \langle \varphi, \varphi \rangle. \quad (6)$$

As the scalar product (5) is symmetric and as the norm is positive for all fields φ , the scalar product defines a covariance function (Gaspari and Cohn, 1999). If the field φ is discretized on a grid and all elements are grouped into the vector \mathbf{x} , the cost function can be written as

$$\begin{aligned} J(\mathbf{x}) = & (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) \\ & + (\mathbf{H}\mathbf{x} - \mathbf{y}^o)^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y}^o) + J_c(\mathbf{x}), \end{aligned} \quad (7)$$

where we also regrouped all observations into vector \mathbf{y}^o and the discretized background field in vector \mathbf{x}_b . \mathbf{H} is a discretized local interpolation operator allowing to compare the gridded field with the observations at the data locations. Therefore the observation operator is linear and can be represented efficiently by a sparse matrix.

This cost function is commonly used in optimal interpolation where the matrices \mathbf{B} and \mathbf{R} are the error covariance of the background estimate and of the observations respectively. The scalar product in Eq. (5) defines the matrix \mathbf{B} and the diagonal matrix \mathbf{R} is composed by the inverse of the data weight $\frac{1}{\mu_j}$.

2.1 Kernel

The so-called reproducing kernel $K(\mathbf{x}, \mathbf{y})$ associated with Eq. (5) is defined by (McIntosh, 1990)

$$\langle f, K \rangle = f \quad (8)$$

and will be helpful in understanding the covariance structure of \mathbf{B} .

If the domain is infinitely large ($D = \mathbb{R}^n$) and the correlation lengths L_i are constant in all dimensions, we can analytically derive the function K . First we assume that the correlation lengths L_i are all equal to one, and later the more general case with arbitrary (but constant) values of L_i will be derived. The derivation follows the PhD thesis of Brasseur (1994), where the kernel is derived for two-dimensional problems. Substituting the definition of scalar product from Eq. (5) in Eq. (8) and by integrating by parts one obtains

$$\begin{aligned} \langle f, K \rangle &= \frac{1}{c} \int_{\mathbb{R}^n} \alpha_0 f(\mathbf{x}) K(\mathbf{x}, \mathbf{y}) + \alpha_1 \left(\tilde{\nabla} f(\mathbf{x}) \right) \cdot \left(\tilde{\nabla} K(\mathbf{x}, \mathbf{y}) \right) \\ &\quad + \alpha_2 \left(\tilde{\nabla}^2 f(\mathbf{x}) \right) \left(\tilde{\nabla}^2 K(\mathbf{x}, \mathbf{y}) \right) + \dots d\mathbf{x} \\ &= \frac{1}{c} \int_{\mathbb{R}^n} f(\mathbf{x}) \left[\alpha_0 K(\mathbf{x}, \mathbf{y}) - \alpha_1 \tilde{\nabla}^2 K(\mathbf{x}, \mathbf{y}) \right. \\ &\quad \left. + \alpha_2 \tilde{\nabla}^4 K(\mathbf{x}, \mathbf{y}) + \dots \right] d\mathbf{x} \\ &= f(\mathbf{y}). \end{aligned}$$

As this last equation must be true for any function $f(\mathbf{x})$, the expression in brackets must be equal to the Dirac function (times c):

$$\alpha_0 K(\mathbf{x}, \mathbf{y}) - \alpha_1 \tilde{\nabla}^2 K(\mathbf{x}, \mathbf{y}) + \alpha_2 \tilde{\nabla}^4 K(\mathbf{x}, \mathbf{y}) + \dots = c \delta(\mathbf{x} - \mathbf{y}).$$

Since the kernel is translation invariant, we can set $\mathbf{y} = 0$ without loss of generality. By applying the Fourier transform, we obtain

$$\hat{K}(\mathbf{k}) = \frac{c}{\alpha_0 + \alpha_1 k^2 + \alpha_2 k^4 + \dots + \alpha_m k^{2m}},$$

where $\hat{K}(\mathbf{k})$ is the Fourier transform of kernel $K(\mathbf{x})$:

$$\hat{K}(\mathbf{k}) = \int_{\mathbb{R}^n} K(\mathbf{x}) e^{-i\mathbf{x} \cdot \mathbf{k}} d\mathbf{x}.$$

The kernel $K(\mathbf{x})$ can thus be found by using the inverse Fourier transform:

$$K(\mathbf{x}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{K}(\mathbf{k}) e^{i\mathbf{x} \cdot \mathbf{k}} d\mathbf{k}.$$

In particular, the value of the kernel at $\mathbf{x} = 0$ corresponds to the integral:

$$K(0) = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \frac{c}{(2\pi)^n} \int_0^\infty \frac{k^{n-1}}{\alpha_0 + \alpha_1 k^2 + \alpha_2 k^4 + \dots + \alpha_m k^{2m}} dk,$$

where integration variables were transformed to n -dimensional polar coordinates and integration was performed over all angles. The kernel is in fact nothing else than the correlation function one would use to create \mathbf{B} yielding the same result in optimal interpolation as with the variational approach (Wahba and Wendelberger, 1980). We naturally choose the value of c such that $K(0) = 1$:

$$\frac{1}{c} = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \frac{1}{(2\pi)^n} \int_0^\infty \frac{k^{n-1}}{\alpha_0 + \alpha_1 k^2 + \alpha_2 k^4 + \dots + \alpha_m k^{2m}} dk.$$

Assuming all $\alpha_i \geq 0$, the integral at the right-hand side is defined if $m > \frac{n}{2}$. This condition links thus the number of dimensions n and the order of the highest derivative needed in the formulation. The Fourier transform of the kernel \hat{K} is a radial function depending on the norm of the wave number \mathbf{k} . The inverse Fourier transform of a radial function is also a radial function which can be derived with the Hankel transform (Appendix A):

$$K(r) = (2\pi)^{-\frac{n}{2}} r^{\frac{2-n}{2}} \int_0^\infty J_{\frac{n-2}{2}}(kr) k^{\frac{n-2}{2}} \hat{K}(k) k dk, \quad (9)$$

where $J_\nu(r)$ is the Bessel function of first kind of order ν . To continue the analytical derivations we must make assumptions about the coefficients α_i . We assume that the coefficients α_i are chosen as binomial coefficients.

$$\alpha_i = \frac{m!}{i!(m-i)!} \quad 1 \leq i \leq m.$$

In this case, the norm can also be expressed as the inverse of an implicit diffusion operator (Weaver and Mirouze, 2013) which is commonly used in variational data analysis in one dimension (Bennett et al., 1997; Chua and Bennett, 2001) or two or three dimensions (Weaver and Courtier, 2001; Chua and Bennett, 2001; Zaron et al., 2009; Carrier and Ngodock, 2010).

The Fourier transform of the kernel ($\hat{K}^{n,m}(k)$) for the highest derivative m and dimension n can be written as

$$\hat{K}^{n,m}(k) = \frac{c}{(1+k^2)^m}.$$

Using this expression in Eq. (9), the radial part of the kernel $K^{n,m}(r)$ becomes

$$K^{n,m}(r) = c^{n,m} (2\pi)^{-\frac{n}{2}} r^{\frac{2-n}{2}} \int_0^\infty J_{\frac{n-2}{2}}(kr) k^{\frac{n-2}{2}} \frac{k}{(1+k^2)^m} dk. \quad (10)$$

Table 1. Kernel as a function of non-dimensional radius $\rho = |\mathbf{L}^{-1}\mathbf{x}| \geq 0$ for different values of the dimension n and the highest derivative m .

	$m = 1$	$m = 2$	$m = 3$
$n = 1$	$e^{-\rho}$	$(1 + \rho)e^{-\rho}$	$(1 + \rho(1 + \rho/3))e^{-\rho}$
$n = 2$	-	$\rho K_1(\rho)$	$\frac{\rho^2}{2} K_2(\rho)$
$n = 3$	-	$e^{-\rho}$	$(1 + \rho)e^{-\rho}$
$n = 4$	-	-	$\rho K_1(\rho)$
$n = 5$	-	-	$e^{-\rho}$

The normalization coefficient is now noted $c^{n,m}$ as it depends on the dimension n and the order of the highest derivative m . By integrating by parts, we can derive a recursion relationship relating the kernels with different values of n and m .

$$\begin{aligned}
 K^{n,m}(r) &= c^{n,m} \frac{(2\pi)^{-\frac{n}{2}}}{2(1-m)} r^{\frac{2-n}{2}} \\
 &\int_0^\infty J_{\frac{n-2}{2}}(kr) k^{\frac{n-2}{2}} \frac{d}{dk} \left(\frac{1}{(1+k^2)^{m-1}} \right) dk \\
 &= c^{n,m} \frac{(2\pi)^{-\frac{n}{2}}}{2(m-1)} r^{\frac{4-n}{2}} \\
 &\int_0^\infty J_{\frac{n-4}{2}}(kr) k^{\frac{n-4}{2}} \frac{k}{(1+k^2)^{m-1}} dk \quad (11) \\
 &= \frac{1}{4\pi(m-1)} \frac{c^{n,m}}{c^{n-2,m-1}} K^{n-2,m-1}(r), \quad (12)
 \end{aligned}$$

where in step (11), we used the following equation relating Bessel functions of first kind of different order:

$$\frac{d}{dx} (x^p J_p(x)) = x^p J_{p-1}(x)$$

or

$$\frac{d}{dk} \left(J_{\frac{n-2}{2}}(kr) k^{\frac{n-2}{2}} \right) = r J_{\frac{n-4}{2}}(kr) k^{\frac{n-2}{2}}.$$

Since $K^{n,m}(\mathbf{x})$ is one for $\mathbf{x} = 0$, the normalization coefficients are thus related by

$$c^{n,m} = 4\pi(m-1)c^{n-2,m-1}. \quad (13)$$

The recurrence relationship therefore shows that it is sufficient to calculate the kernel ($K^{n,m}$) and the normalization coefficients for $n = 1$ and $n = 2$. For $n = 1$, we find the following solution for the integral in Eq. (10):

$$\begin{aligned}
 K^{1,m}(r) &= \frac{2}{\Gamma(m-1/2)} \left(\frac{r}{2} \right)^{(m-1/2)} K_{m-1/2}(r) \\
 c^{1,m} &= \frac{2\sqrt{\pi} \Gamma(m)}{\Gamma(m-1/2)},
 \end{aligned}$$

Table 2. Normalization coefficient $c^{n,m}$ for different values of the dimension n and the highest derivative m .

	$m = 1$	$m = 2$	$m = 3$
$n = 1$	2	4	$\frac{16}{3}$
$n = 2$	-	4π	8π
$n = 3$	-	8π	32π
$n = 4$	-	-	$32\pi^2$
$n = 5$	-	-	$64\pi^2$

where $K_\nu(r)$ is the modified Bessel function of second kind of order ν . For $n = 2$, the solution of Eq. (10) is:

$$\begin{aligned}
 K^{2,m}(r) &= \frac{2}{\Gamma(m-1)} \left(\frac{r}{2} \right)^{(m-1)} K_{m-1}(r) \\
 c^{2,m} &= 4\pi(m-1).
 \end{aligned}$$

Using the recursion relationship from Eqs. (12) and (13) with the solution for $n = 1$ ($n = 2$) one can derive the kernel and $c^{n,m}$ for every odd (even) value of n . After simplifications, it follows that for any n (odd and even) and m , the normalized kernel $K^{n,m}$ and the corresponding normalization factor $c^{n,m}$ can be written as

$$\begin{aligned}
 K^{n,m}(r) &= \frac{2}{\Gamma(\nu)} \left(\frac{r}{2} \right)^\nu K_\nu(r) \\
 c^{n,m} &= \frac{(4\pi)^{n/2} \Gamma(m)}{\Gamma(\nu)}
 \end{aligned}$$

with $\nu = m - n/2$. Non-isotropic covariance functions are often introduced by using coordinate stretching (Derber and Rosati, 1989; Carton and Hackert, 1990). We can finally derive the case for a correlation length different from one, using the change of variables $\mathbf{x} \rightarrow \mathbf{L}^{-1}\mathbf{x}$:

$$\begin{aligned}
 K^{n,m}(\mathbf{x}) &= \frac{2}{\Gamma(\nu)} \left(\frac{|\mathbf{L}^{-1}\mathbf{x}|}{2} \right)^\nu K_\nu(|\mathbf{L}^{-1}\mathbf{x}|) \\
 c^{n,m} &= \frac{(4\pi)^{n/2} \Gamma(m) |\mathbf{L}|}{\Gamma(\nu)},
 \end{aligned}$$

where $|\mathbf{L}|$ is the determinant of the diagonal matrix \mathbf{L} :

$$|\mathbf{L}| = \prod_{i=1}^n L_i.$$

The kernels and normalization coefficients can be expanded further for particular values of n and m leading to the lines of Tables 1 and 2. Our results agree with the solution derived by Brasseur et al. (1996) for the case of two dimensions $n = 2$ and $m = 2$.

The correlation function is the class of the Matérn family of spatial correlations (Matérn, 1986; Guttorp and Gneiting, 2006; Weaver and Mirouze, 2013). These functions have also been previously derived to define a smoothing operator $\tilde{\phi}(\mathbf{x})$ based on the Laplacian:

$$\gamma_n (1 - L^2 \nabla^2)^m \phi(\mathbf{x}) = \tilde{\phi}(\mathbf{x}), \quad (14)$$

where γ_n is a normalization constant, L the smoothing length-scale and $m = \nu + n/2$ as before and ν a smoothness parameter. For $\nu = 1/2$ and $\nu = 5/2$, one recovers the well-known second-order and third-order auto-regressive functions (Daley, 1991; Gaspari and Cohn, 1999; Gneiting, 1999). A formulation based on the diffusion operator allows for a straight-forward square root decomposition if m is even (Weaver and Courtier, 2001):

$$(1 - L^2 \nabla^2)^m = (1 - L^2 \nabla^2)^{m/2} (1 - L^2 \nabla^2)^{m/2}. \quad (15)$$

This square root decomposition is mainly used in connection with a conjugate gradient solver as a preconditioner (e.g. Haben et al., 2011).

In the *divand* tool a distinction is made between the actual dimension n and the effective dimension. The effective dimension is the number of dimensions with a non-zero correlation length. Setting a correlation length to zero decouples the different dimensions; this is used to emulate the results that one obtains by “stacking” results of two-dimensional analyses as it has been done previously (e.g. Troupin et al., 2010). The normalization coefficient used in this case is based on the effective dimension. This ensures that one obtains exactly the same results of a stacked 2-D interpolation by analysing data in a 3-D domain with a zero correlation length in the third dimension.

2.2 Additional constraints

In addition to the observation constraint and smoothness constraint, an arbitrary number of other constraints can be included in *divand*. Those constraints are characterized by the symmetric positive-defined matrix \mathbf{Q}_i , the matrix \mathbf{C}_i and the vectors \mathbf{z}_i .

$$J_c(\mathbf{x}) = \sum_i (\mathbf{C}_i \mathbf{x} - \mathbf{z}_i)^T \mathbf{Q}_i^{-1} (\mathbf{C}_i \mathbf{x} - \mathbf{z}_i).$$

Those additional constraints can be re-absorbed in the definition of the operator \mathbf{H} , \mathbf{R} and \mathbf{y}^o :

$$\begin{pmatrix} \mathbf{y}^o \\ \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \end{pmatrix} \rightarrow \mathbf{y}^o \quad \begin{pmatrix} \mathbf{H} \\ \mathbf{C}_1 \\ \mathbf{C}_2 \\ \vdots \end{pmatrix} \rightarrow \mathbf{H} \quad \begin{pmatrix} \mathbf{R} & & 0 \\ & \mathbf{Q}_1 & \\ 0 & & \mathbf{Q}_2 \\ & & & \ddots \end{pmatrix} \rightarrow \mathbf{R}. \quad (16)$$

With these definitions the cost function has again the familiar form:

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{H}\mathbf{x} - \mathbf{y}^o)^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y}^o). \quad (17)$$

A common example for an additional constraint is the so-called advection constraint (Wahba and Wendelberger, 1980; Brasseur and Haus, 1991; Troupin et al., 2012). Anisotropic background error covariance can be achieved by requiring that the gradient of the analysis is aligned to a given vector field.

$$J_a(\phi) = \int_D (\mathbf{v} \cdot \nabla \phi)^2 dD, \quad (18)$$

where \mathbf{v} is a vector field with n components. Such a constraint is useful in a geophysical context to force a field to be close to a stationary (or time dependent) solution of the advection equation, in which case the vector field \mathbf{v} is related to the ocean current. The diffusion term is not included as it is generally small for geophysical applications and since the background constraint acts similar to a diffusion.

3 Minimization and algorithms

As the cost function is quadratic, one can obtain its minimum analytically. The derivation of its minimum is well known (Courtier et al., 1998) and is included here for completeness. If \mathbf{x}^a is the minimum of the cost function J , a small variation $\delta\mathbf{x}$ of \mathbf{x}^a would not change the cost function in the first order of $\delta\mathbf{x}$. Noting T a transposed matrix or vector,

$$\begin{aligned} \delta J &= J(\mathbf{x}^a + \delta\mathbf{x}) - J(\mathbf{x}^a) \\ &= 2\delta\mathbf{x}^T \mathbf{B}^{-1} (\mathbf{x}^a - \mathbf{x}_b) + 2\delta\mathbf{x}^T \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{x}^a - \mathbf{y}^o) = 0. \end{aligned}$$

As $\delta\mathbf{x}$ is arbitrary, the expression multiplying $\delta\mathbf{x}^T$ must be zero. The optimal state \mathbf{x}^a is thus given by

$$\mathbf{x}^a = \mathbf{x}_b + \mathbf{P}\mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y}^o - \mathbf{H}\mathbf{x}_b), \quad (19)$$

where we have introduced the matrix \mathbf{P} .

$$\mathbf{P}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}. \quad (20)$$

The interpretation of this matrix becomes clear since we can rearrange the cost function as

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^a)^T \mathbf{P}^{-1} (\mathbf{x} - \mathbf{x}^a) + \text{constant}. \quad (21)$$

The matrix \mathbf{P} represents thus the error covariance of the analysis \mathbf{x}^a (Rabier and Courtier, 1992; Courtier et al., 1994).

3.1 Primal formulation

The primal formulation of the algorithm follows directly from Eqs. (19) and (20). The matrices \mathbf{P} and \mathbf{B} are never formed explicitly and the tool works only with the inverse of these matrices noted \mathbf{P}_{inv} and \mathbf{B}_{inv} respectively. In order to give the algorithm in a compact form close to the mathematical equations, we introduce the backslash operator by

$$\mathbf{X} = \mathbf{A} \setminus \mathbf{B}, \quad (22)$$

which is equivalent to solving the system $\mathbf{A}\mathbf{X} = \mathbf{B}$ for the matrix (or vector) \mathbf{X} . With this notation, the primal algorithm reads

$$\mathbf{P}_{\text{inv}} = \mathbf{B}_{\text{inv}} + \mathbf{H}^T (\mathbf{R} \setminus \mathbf{H}), \quad (23)$$

$$\mathbf{x}^a = \mathbf{P}_{\text{inv}} \setminus (\mathbf{H}^T (\mathbf{R} \setminus \mathbf{y}^o)). \quad (24)$$

Different approaches have been implemented here to solve the analysis equation in its primal formulation involving either a direct solver, a factorization or the conjugate gradient method.

3.1.1 Direct solver

The solver based on SuiteSparse (Chen et al., 2008; Davis and Hager, 2009) can be used directly if the matrix $\mathbf{R} \setminus \mathbf{H}$ is sparse. This is in particular the case when \mathbf{R} is diagonal. \mathbf{P}_{inv} will then also be a sparse matrix which can be efficiently stored. This approach is useful when no error field needs to be computed. The direct solver can also be applied for a non-diagonal \mathbf{R} but this approach is prohibitive in terms of computational cost and memory.

3.1.2 Factorization

The inverse of the a posteriori error covariance matrix \mathbf{P}_{inv} is factorized in the following products:

$$\mathbf{R}_P^T \mathbf{R}_P = \mathbf{Q}_P^T \mathbf{P}_{\text{inv}} \mathbf{Q}_P, \quad (25)$$

where \mathbf{R}_P is an upper triangular matrix and \mathbf{Q}_P is a permutation matrix (chosen to preserve the sparse character of \mathbf{R}_P). The factorization is performed by CHOLMOD (CHOLesky MODification) (Chen et al., 2008; Davis and Hager, 2009). Once the matrix \mathbf{P}_{inv} is factorized, the product between \mathbf{P} and any vector \mathbf{x} can be computed efficiently by

$$\mathbf{P}\mathbf{x} = \mathbf{P}_{\text{inv}} \setminus \mathbf{x} = \mathbf{Q}_P (\mathbf{R}_P \setminus (\mathbf{R}_P^T \setminus (\mathbf{Q}_P^T \mathbf{x}))). \quad (26)$$

This approach is useful if the error field is required, since a large number of products between \mathbf{P} and a given vector must be computed. Determining all elements of \mathbf{P} would be prohibitive, but individual elements (such as the diagonal elements) are computed by

$$\mathbf{P}_{ij} = \mathbf{e}_i (\mathbf{P}_{\text{inv}} \setminus \mathbf{e}_j), \quad (27)$$

where \mathbf{e}_i is the i th basis vector. The tool `divand` returns a matrix-like object allowing to compute any element of \mathbf{P} or the product of \mathbf{P} with a given vector. This latter product is useful if one wants to derive the expected error of an integrated quantity such as a transport along a section or any other weighted sum.

3.1.3 Conjugate gradient method

The conjugate gradient method (Golub and Van Loan, 1996) is commonly used in variational data assimilation (Moore et al., 2011a). This method provides an iterative solution to linear equations:

$$\mathbf{A}\mathbf{x}^a = \mathbf{b}, \quad (28)$$

where the vector \mathbf{b} and the symmetric and positive-defined matrix \mathbf{A} are given here by

$$\mathbf{b} = \mathbf{H}^T (\mathbf{R} \setminus \mathbf{y}^o) \quad (29)$$

$$\mathbf{A}\mathbf{x}^a = \mathbf{B}_{\text{inv}} \mathbf{x}^a + \mathbf{H}^T (\mathbf{R} \setminus (\mathbf{H}\mathbf{x}^a)). \quad (30)$$

The conjugate gradient algorithm is applied to solve for \mathbf{x}^a . For large domain sizes, the matrix \mathbf{A} tends to be ill-conditioned (e.g. Haben et al., 2011) and the use of a preconditioner is necessary. In variational data assimilation it is common to define the covariance matrix implicitly by $\mathbf{B} = \mathbf{V}\mathbf{V}^T$ where \mathbf{V} is a sequence of operators requiring physical consistency, e.g. using balance relationships and empirical orthogonal functions, and spatial smoothness, e.g. using a projection in spectral space (Dobricic and Pinardi, 2008; Bannister, 2008). A preconditioning using \mathbf{V}^{-1} amounts to transforming the background error covariance matrix into the identity matrix (Lorenc et al., 2000; Cullen, 2003; Haben et al., 2011). The matrix \mathbf{V} is not necessarily square because some modes of variability might be excluded (Lorenc et al., 2000). In this case the inverse of \mathbf{V} is interpreted as the pseudo-inverse. Theoretical bounds for the condition number in idealized conditions have been derived for this type of preconditioning (Haben et al., 2011).

A similar preconditioner is also implemented in `divand` using a Cholesky decomposition of the matrix \mathbf{B}_{inv} into the square root matrix $\mathbf{B}^{-1/2}$:

$$\mathbf{B}_{\text{inv}} = \mathbf{B}^{-1/2} \mathbf{B}^{-1/2 T}. \quad (31)$$

The matrix $\mathbf{B}^{-1/2}$ is an upper triangular matrix and the inverse of this matrix times a vector can be efficiently computed efficiently by back substitution. If the observation error covariance matrix \mathbf{R} is diagonal and if the observation operator \mathbf{H} is a sparse matrix, then the inverse of the analysis error covariance matrix \mathbf{P}_{inv} is also sparse. In this case, the matrix \mathbf{P}_{inv} can also be decomposed using a modified incomplete Cholesky decomposition (Jones and Plassmann, 1995; Benzi, 2002).

These two preconditioner have been tested for a unit square domain in two dimensions discretized on a regular $N \times N$ grid. Observations are evenly distributed on a 10×10 grid. The signal-to-noise ratio is set to 1 and the correlation length-scale is set to 0.1.

Figure 3 shows the number of iterations needed to reach a residual lower than 10^{-6} for different domain sizes N (ranging from 10 to 100) using either no preconditioner, or a preconditioning based on the square root decomposition of \mathbf{B}_{inv} , or the modified incomplete Cholesky decomposition with a drop tolerance of zero (MIC(0)). Even for the largest domain, the computation of the preconditioner matrices themselves takes less than 0.15 s and its contribution to the total CPU time is negligible.

The increased convergence using the preconditioner based on $\mathbf{B}^{-1/2}$ compared to no-preconditioner is well expected (Lorenc, 1997; Gauthier et al., 1999; Haben et al., 2011). The preconditioning using the modified incomplete Cholesky decomposition leads to the smallest number of iterations in this test case. This is attributed to the fact that the MIC(0) preconditioner takes also the error reduction due to the observations into account (the term $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$ in the Hessian

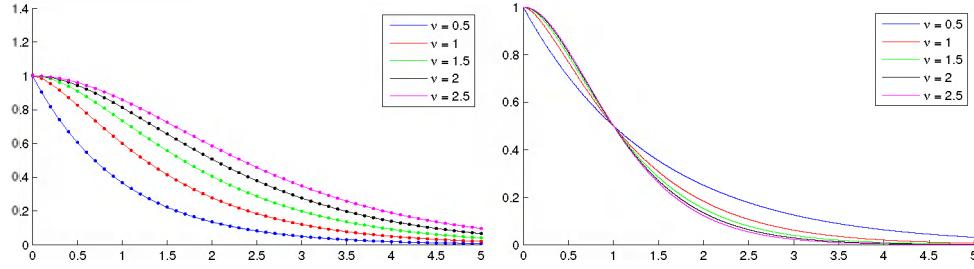


Fig. 1. Solid lines show the analytical kernels for different values of ν and the dots show the numerical kernel (left) and analytical kernels with scaled r_h (right).

matrix of the cost function) while this is not the case by using the square root matrix $\mathbf{B}^{-1/2}$ as preconditioner. It should be noted that these results depend thus also on the data distribution and the signal-to-noise ratio of the observations. In particular, the fewer observations are available or the higher the error variance of the observation is, the better becomes the conditioning based on $\mathbf{B}^{-1/2}$ (Haben et al., 2011). The tool *divand* has been written such that the user can provide a custom preconditioner based on the background error covariance \mathbf{B} , observation error covariance \mathbf{R} and observation operator \mathbf{H} .

3.2 Dual formulation

Using the Sherman–Morrison–Woodbury formula (Golub and Van Loan, 1996), the solution can also be written in the dual formulation, also called Physical-Space Statistical Analysis System (Courtier, 1997; Cohn et al., 1998; Štajner et al., 2001; Dee and da Silva, 2003):

$$\mathbf{x}^a = \mathbf{B}\mathbf{H}^T (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \mathbf{y}^o, \quad (32)$$

$$\mathbf{P} = \mathbf{B} - \mathbf{B}\mathbf{H}^T (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H}\mathbf{B}. \quad (33)$$

In this formulation, all implemented methods in *divand* are based on the conjugate gradient method, to solve iteratively the following equation for \mathbf{y}' :

$$\mathbf{C}\mathbf{y}' = \mathbf{y}^o, \quad (34)$$

where \mathbf{C} represents a symmetric and positively defined matrix specified in operator form:

$$\mathbf{C}\mathbf{y}' = \mathbf{H}(\mathbf{P}_{\text{inv}} \setminus (\mathbf{H}^T \mathbf{y}')) + \mathbf{R}\mathbf{y}'. \quad (35)$$

Once the vector \mathbf{y}' is known, the analysis \mathbf{x}^a is obtained by

$$\mathbf{x}^a = \mathbf{P}_{\text{inv}} \setminus (\mathbf{H}^T \mathbf{y}'). \quad (36)$$

If \mathbf{R} is a non-diagonal matrix, one can use a preconditioner based on the diagonal elements of \mathbf{R} noted \mathbf{R}' .

$$\mathbf{M} = \mathbf{H}(\mathbf{P}_{\text{inv}} \setminus \mathbf{H}^T) + \mathbf{R}' \quad (37)$$

The matrix \mathbf{M} can be efficiently factorized in sparse matrices:

$$\mathbf{R}_M^T \mathbf{R}_M = \mathbf{Q}_M^T \mathbf{M}_{\text{inv}} \mathbf{Q}_M, \quad (38)$$

where \mathbf{R}_M is an upper triangular matrix and \mathbf{Q}_M is a permutation matrix.

3.2.1 Factorization

If the conjugate algorithm requires a large number of iterations, it is useful to factorize the matrix \mathbf{P}_{inv} to accelerate the product of \mathbf{P} and a vector, which requires solving a linear system.

$$\mathbf{R}_P^T \mathbf{R}_P = \mathbf{Q}_P^T \mathbf{P}_{\text{inv}} \mathbf{Q}_P, \quad (39)$$

where, as before, \mathbf{R}_P is an upper triangular matrix and \mathbf{Q}_P is a permutation matrix. All products of \mathbf{P} times a vector \mathbf{x} are computed efficiently by

$$\mathbf{P}\mathbf{x} = \mathbf{P}_{\text{inv}} \setminus \mathbf{x} = \mathbf{Q}_P(\mathbf{R}_P \setminus (\mathbf{R}_P^T \setminus (\mathbf{Q}_P^T \mathbf{x}))). \quad (40)$$

4 Numerical tests

4.1 Numerical kernel

It has been verified in numerical tests that the code reproduces well the analytical kernels. The domain is one-dimensional for $\nu = 1/2, 3/2$ and $5/2$ and two-dimensional for $\nu = 1$ and 2 . Every direction ranges from -10 to 10 and is discretized with 201 grid points. A hypothetical observation located at the centre of the domain is analysed with a signal-to-noise ratio of 1 and with a correlation length scale of 1. Theoretically the value of the analysed field should be $1/2$ at the centre. The radial solution multiplied by two is compared to the analytical solution of an infinite domain (Fig. 1). In all tests, the scaling and structure of the numerical kernels correspond well to the analytical solutions. To obtain this correspondence, it is necessary that the grid resolution resolves well the correlation length. Qualitative tests have shown that the numerical kernels match well the analytical functions as

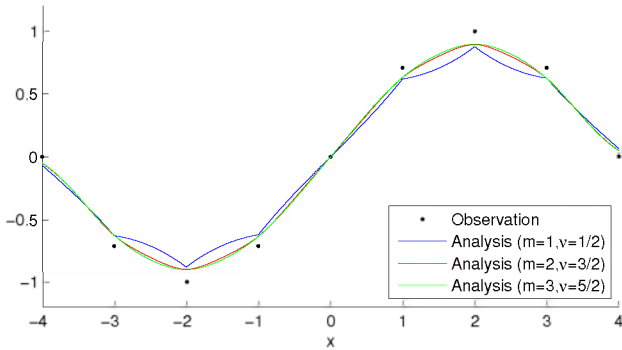


Fig. 2. Impact of higher-order derivatives.

Table 3. Radial distance where the kernel is 1/2 for different values of ν .

ν	r_h
1/2	0.69315
1	1.25715
3/2	1.67835
2	2.02700
5/2	2.33026

long as the grid spacing is one fourth (or less) of the correlation length.

The kernels differ by the rate at which they decrease to zero. It is important to consider this aspect when comparing the correlation length from analyses using a different number of derivatives. Table 3 shows the value of r for which the analytical kernels are 1/2. They can be used as proportionality coefficients to make the kernels more comparable (Fig. 1).

The kernel for $\nu = 1/2$ has a discontinuous derivative for $r = 0$ which makes this function unfit for practical use. A simple one-dimensional analysis with $m = 1$ (i.e. $\nu = 1/2$) illustrates the problem (Fig. 2). As there is no penalty on the second derivative, the analysis has a discontinuous derivative at every observation location (black dots). Using higher order derivatives solves this problem. This example shows also that the analyses with higher order kernels ($\nu = 3/2$ and $\nu = 5/2$) are very similar. In these numerical experiments, the correlation length L is the inverse of the values in Table 3. This problem appears in all configurations where $\nu = 1/2$. In particular also in three-dimensional analyses when the highest derivative in the cost function is a Laplacian. This is surprising because the first derivative is discontinuous despite the fact that the cost function penalizes the second derivative. By default the cost function in *divand* therefore includes the derivatives up to $1 + n/2$ (rounded upwards) to ensure that the analysis has a continuous derivative.

Table 4. Run time in seconds for different domain sizes for MATLAB (R2012b) and Octave version 3.6.4 (using GotoBLAS2 or MKL) for the primal algorithm with Cholesky factorization.

	MATLAB-R2012b	Octave-GotoBLAS2	Octave-MKL
100	0.304	0.415	0.451
200	1.694	1.808	1.692
300	5.235	5.155	5.088
400	7.527	8.104	8.388
500	15.501	13.905	14.115
600	31.457	24.906	25.156

4.2 Benchmark

The tool *divand* is written such that it can run on MATLAB and GNU Octave. In general, interpreted language tends to be slow on explicit loops. Interpreters like MATLAB can reduce this performance penalty by using just-in-time compilation (which is currently not available in Octave). Therefore MATLAB tends to be faster on performance benchmarks (e.g. Chaves et al., 2006; Leros et al., 2010) than GNU Octave. Explicit loops are avoided in the tool *divand* except for the number of dimensions which is a quite short loop (typically 1–4 dimensions).

The run-time performance of *divand* for MATLAB (version R2012b) and GNU Octave (version 3.6.4) with two implementations of the BLAS library (either GotoBLAS2 version 1.13 or Intel’s Math Kernel Library (MKL) version 10.1) were tested. The benchmark was performed on an Intel Xeon L5420 CPU (using a single core) with 16 GB memory. The domain is a square and the used correlation length is inversely proportional to the number of grid points in one dimension.

Observations come from an analytical function and are located at every fifth grid point (in the two dimensions). In all cases the correctness of the analysed field was verified by comparing the interpolated field to the original analytical function. The benchmark was repeated five times and the median value is shown in Table 4 for the primal algorithm with Cholesky factorization and for the dual algorithm with conjugate gradient minimization (Table 5). For the primal algorithm, all tested versions perform equally well with a slight advantage of Octave for domain sizes of 500×500 and 600×600 . Profiling of the code shows that for the primal algorithm most of the time is spent in the Cholesky factorization using the library CHOLMOD included in SuiteSparse in both MATLAB and Octave which explains the similar results.

In general, the dual algorithm is much slower than the primal algorithm with Cholesky factorization. However, it should be reminded that the latter would be unpractical in some cases (in particular with spatially correlated observation errors). The difference between the different interpreters is more pronounced for the dual case. The benchmarks show

that MATLAB is faster for small domain sizes (30×30 and 50×50), while GNU Octave (with GotoBLAS2 and MKL) outperforms MATLAB for larger domain sizes. For small domain sizes, the preparation of the matrices to invert represents a significant fraction of the total run time where GNU Octave tends to be slower due to the lack of a just-in-time compiler. For larger domains, the cost is dominated by matrix operations which are faster in Octave for our case. In our benchmark, the Math Kernel Library is slightly faster in Octave than the GotoBLAS2 library.

5 Implementation

The *divand* tool is implemented such that it allows analysis in an arbitrary high dimensional space. The grid must be an orthogonal curvilinear grid and some dimensions might be cyclic. Internally the n -dimensional arrays, e.g. for the analysed field, are represented as a “flat” vector. To implement the background constraint, the following basic operations are implemented as sparse matrices:

- differentiation along a given dimension;
- grid staggering along a given dimension;
- trimming the first and last elements of a given dimension.

All other differential operators are derived as a product and sum of these basic operations. For instance, the advection constraint is implemented by performing successively the following operations: differentiation along the i th dimension for $i = 1, \dots, n$ (operator \mathbf{D}_i), staggering the result back to the centre of the grid cell (operator \mathbf{S}_i), removing land points (operator \mathbf{P}) and multiplying with a diagonal matrix formed by the i th velocity component (grid points corresponding to land are removed; operator $\text{diag}(\mathbf{v}_i)$). All of these operations are represented as a sparse matrix and the resulting matrix \mathbf{A} is thus also sparse:

$$\mathbf{A} = \sum_{i=1}^n \text{diag}(\mathbf{v}_i) \mathbf{P} \mathbf{S}_i \mathbf{D}_i.$$

The additional constraint corresponding to the advection constraint has thus the following form:

$$J_c(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^T \mathbf{A}^T \mathbf{A} (\mathbf{x} - \mathbf{x}^b).$$

This example illustrates that once the basic operators have been implemented, more complex operations can be derived on a relatively high level.

Consistent error calculations are also possible with the tool *divand* to estimate the error variance of the analysed field. This error variance reflects among others the distribution of the observations, the correlation length and the background variance error. However, the accuracy of the error estimate

of the analysed field depends crucially on the validity of the background and observation error covariance.

The tool introduces new matrix objects which implement several matrix operations (such as multiplication, multiplication by its inverse, extraction of diagonal elements). These new matrix objects include:

- a matrix specified by its inverse and potentially factorized (for \mathbf{B} and \mathbf{P});
- a matrix specified by an analytical function (for \mathbf{R} and \mathbf{C}_i);
- a matrix of the form $\mathbf{C} + \mathbf{B}\mathbf{B}^T$, which can be inverted using the Sherman–Morrison–Woodbury formula (for \mathbf{R});
- a matrix composed by block matrices (for additional constraints).

By adding these new matrix objects one can code the algorithm in a compact way which is close to the original mathematical formulation. For instance, the product of the analysis error covariance matrix \mathbf{P} and vector \mathbf{x} can just be coded as $\mathbf{P} * \mathbf{x}$ and the matrix multiplication method of matrix object \mathbf{P} implements the multiplication using the factorized form of Eq. (26).

6 Realistic test case

The interpolation was tested using pseudo-observations of temperature coming from a global ocean model. The results of the year 2007 from the ORCA2 model (Mathiot et al., 2011) with a spatial varying resolution (generally close to 2° resolution) are used. Making a reconstruction with model data allows one to compare the analysed field to the original model data and to assess the quality of the analysis. The position of the pseudo-observations are the real positions of the Argo observations from year 2007. Observations are extracted from the daily model results and the analysis targets monthly means. This mimics the common setup with real observations where measurements are instantaneous while the analyses represent a mean over a given time period. Only surface data are reconstructed for every month separately (2-D analysis) or all 12 months are considered together (3-D analysis). The analysis is compared to a monthly model climatology for the year 2007 and the RMS (root mean squared) difference is calculated. This approach is similar to twin experiments used to test data assimilation schemes (e.g. Nerger et al., 2012). This procedure has also been used before in the context of data analysis, e.g. by Guinehut et al. (2002) and Buongiorno Nardelli (2012). In the following test cases, the solver based on Cholesky factorization is used (Chen et al., 2008; Davis and Hager, 2009). The accuracy of the result is tested by verifying that gradient of the cost function at the analysis is close to the floating point precision.

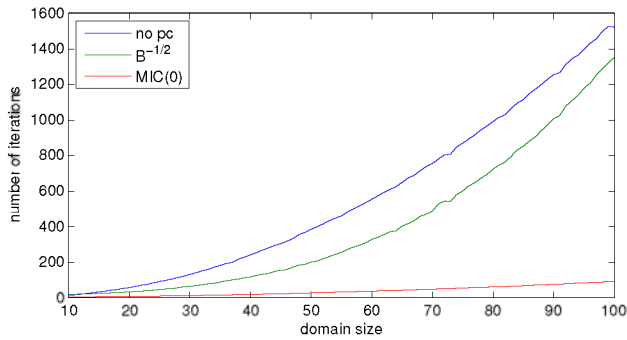


Fig. 3. Number of iteration as a function of domain size (in one dimension). The number of grid points if the domain size is squared.

Table 5. Run time in seconds for different domain sizes for MATLAB (R2012b) and Octave version 3.6.4 (using GotoBLAS2 or MKL) for the dual algorithm.

	MATLAB-R2012b	Octave-GotoBLAS2	Octave-MKL
30	0.039	0.111	0.145
50	0.122	0.161	0.203
100	3.021	2.194	1.879
200	60.685	42.194	35.367
300	443.830	275.641	230.988

The central question of this test case is to assess the benefit of a 3-D analysis (longitude, latitude and time) compared to a 2-D analysis (longitude and latitude). Afterward different variants of the analysis are also tested, in particular the advection constraint. Parameters in the analysis (signal-to-noise ratio, spatial and temporal correlation length and strength of the advection constraint) are optimized to ensure the comparison of every approach in its best possible configuration. Analyses are compared to a model climatology obtained by averaging the daily model output.

The following experiments use the direct solver with a sparse Cholesky factorization.

6.1 2-D analysis

All observations from the same month are considered as data from the same time instance. The correlation length is chosen here to be identical in both horizontal dimensions. Signal-to-noise ratio and correlation length are optimized by an “exhaustive search” of those parameters. This search is implemented by varying the signal-to-noise ratio between 1 and 30 (60 values equally distributed) and correlation length of 100–3000 km (also 60 values equally distributed). In total, 3600 analyses have thus been carried out for this case. The RMS difference (space and time average) between the analysis and the reference model climatology is minimum for a signal-to-noise ratio of 14 and a correlation length of 1072 km (Fig. 4). The global RMS error of this analysis is 1.1501 °C (Table 6). This experiment serves as the baseline for other experiments

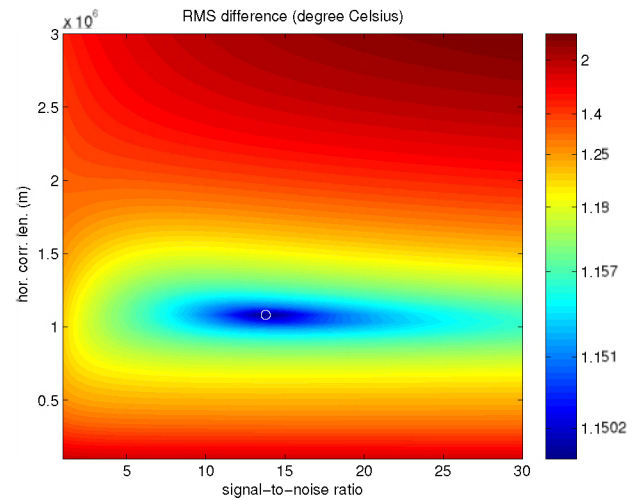


Fig. 4. RMS difference between the reference climatology and the analysis for different values of signal-to-noise ratio and correlation length. A non-linear colour-map is used to enhance detail near the minimum.

and improvements which will be expressed as using the mean square skill score using this experiment as a reference.

Figure 5 (top left panel) shows the RMS error averaged over time for every spatial grid point. This RMS field reflects essentially the data coverage and areas with poor coverage that can have RMS errors of 3 °C and more. Due to the sparsity of Argo data in the coastal area (relative to the near-shore scales of variability), the RMS error is generally highest near the coast.

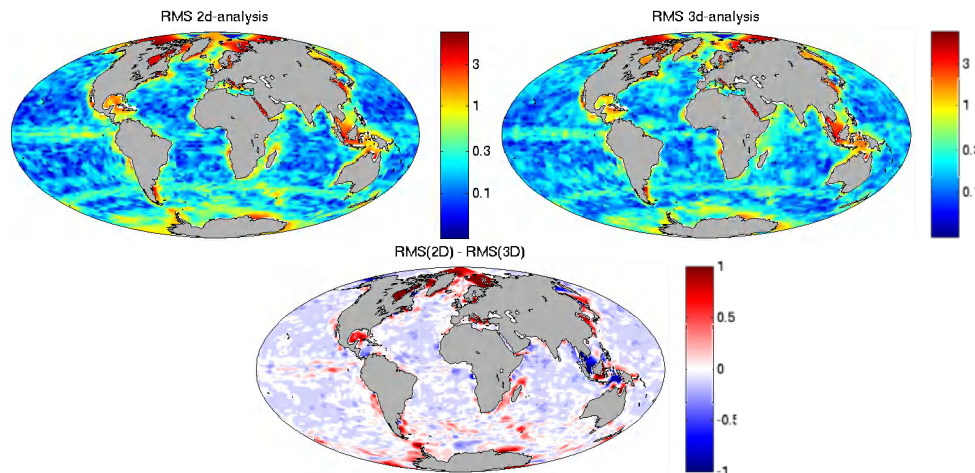
As a variant of the 2-D analysis, the stationary advection is added to the cost function.

$$J_a[\varphi] = \int_D (\mathbf{v} \cdot \nabla \varphi)^2 dD = a_s^2 \int_D \left(u \frac{\partial \varphi}{\partial x} + v \frac{\partial \varphi}{\partial y} \right)^2 dD, \quad (41)$$

where $\mathbf{v} = (a_s u, a_s v)$. The vector (u, v) represents the monthly-averaged model currents. The coefficient a_s determines how strong the advection constraint should be enforced. It is instructive to visualize the impact of a point observation without and with advection constraint. The correction by a point observation is in fact directly related to the background error covariance. Figure 6 shows the impact of an observation located at 72° W, 36.9° N (white cross). Without advection constraint, the covariance is mostly isotropic. The slight deviation from isotropy is due to the proximity of the coastline. The location with the largest impact is marked by a white circle. One could expect that the location with the largest impact coincides with the location of the observation. This is actually the case for an observation far away from the coastline. However, the background error variance is not spatially uniform. In the interior of the domain, every grid point is constrained by the four direct neighbours (or more if higher derivatives than the Laplacian are used). This is not the case

Table 6. Summary of all experiments with the optimal parameter values.

Experiment	λ	L_X (km)	L_T (month)	Advection	RMS (°C)	Skill score (%)
2-D	14.0	1072	–	–	1.1501	0
2-D, advection	71.2	1171	–	7.22	0.9696	29
3-D	27.0	1397	4.9	–	0.9822	27
3-D, fractional time	29.5	1373	4.7	–	0.9820	27
3-D, fractional time, advection	53.8	1477	4.6	1.19	0.8589	44

**Fig. 5.** Top left: RMS difference (averaged over time) between the 2-D analysis and the model reference climatology. Top right: idem for 3-D. Bottom: difference of RMS error of the 2-D analysis minus the RMS error of the 3-D analysis.

at the boundary of the domain and boundary points are thus less constrained. Therefore the background error variance is higher near the coast.

With the advection constraint, the covariance is elongated along the path of the Gulf Stream (downstream and upstream). This is a desirable effect since tracers in the ocean tend to be uniform along ocean currents. The variance with advection constraint is relatively spatially uniform near the location of the observation and thus the location of maximum impact coincides with the position of the observation.

The 2-D analysis with advection constraint has thus in total three parameters: signal-to-noise ratio, spatial correlation and strength of the advection constraint. These parameters are optimized by the Nelder–Mead algorithm (Nelder and Mead, 1965) by minimizing the RMS difference between the analysis and the reference climatology.

6.2 3-D analysis

In a first test, all observations from the same month are again considered coming from the same time instance. Although this is not necessary for a 3-D analysis, it simplifies the comparison with the previous 2-D case where the information of the actual day is not taken into account. Signal-to-noise ratio,

spatial correlation and temporal correlation are optimized using the non-linear Nelder–Mead minimization as before. The RMS is minimum with 0.9822, for a signal-to-noise ratio of 27, spatial correlation length of 1373 km and a temporal correlation length of 4.9 months. Extending the analysis from 2-D to 3-D improves the skill (mean square skill score) by 27%.

For every spatial grid point the time-averaged RMS error is computed (Fig. 5). To facilitate the comparison with the 2-D case, the difference of these RMS values is also shown (Fig. 5, bottom panel). Red shows areas where the 3-D analysis is better and in blue areas the 2-D analysis is more accurate. The RMS error is generally reduced by the 3-D analysis in coastal areas where few observations are present. However also a small degradation is observed in the open ocean. This is attributed to the fact that signal-to-noise ratio and correlation length are optimized globally. It is probable that space-dependent parameters (distinguishing for example between open ocean and near shore conditions) will improve the analysis even more.

In a second test, the actual days of the pseudo-measurements are used in the analysis (noted fractional time in Table 6) which slightly improves the analysis. The small

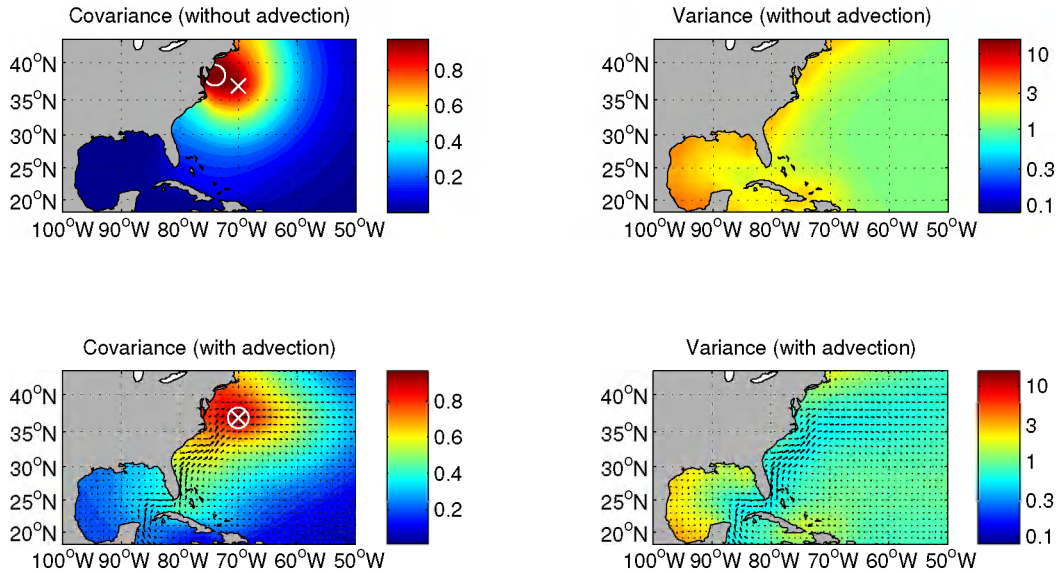


Fig. 6. 2-D case: background error covariance (left panels) relative to the location marked by a cross, and surrounding grid points and background variance (right panels). The upper (lower) panels correspond to the case without (with) advection constraint. The circle indicates the grid point with the highest covariance.

increase of the signal-to-noise ratio is consistent with the fact that by providing the exact date (instead of the month), the observations are more coherent. The relative small improvement is related to the fact that the optimal temporal correlation is 4.9 months and larger than one month (the time resolution of the original 3-D experiment).

The analysis is also performed using the advection constraint based on model currents. Since the time dimension is included, it is possible to use the non-stationary advection constraint:

$$J_a[\varphi] = \int_D (\mathbf{v} \cdot \nabla \varphi)^2 dD = a_s^2 \int_D \left(u \frac{\partial \varphi}{\partial x} + v \frac{\partial \varphi}{\partial y} + \frac{\partial \varphi}{\partial t} \right)^2 dD, \quad (42)$$

where $\mathbf{v} = (a_s u, a_s v, a_s)$ and a_s determines the strength of the advection constraint. Figure 7 shows the impact of a point observation in the 3-D case. Without the advection constraint, the covariance is essentially uniform with a small modulation due to the proximity of the coastline. With a time-dependent advection constraint a distinction between upstream and downstream is made if two different time steps are considered. The location of the observation is more strongly connected to the upstream area of previous time instances and more strongly related to the downstream area of the following time instances. The time-dependent covariances with the advection constraint can thus relate different time instances taking the advection into account. As in the 2-D case, the advection constraint is introduced with a proportionality coefficient a_s allowing to tune the strength of this effect. The calibration of this parameter is related to, among others, the overall significance of advection compared to other processes and the accuracy of the current field. In this test case, the

advection parameter is determined, together with the spatial and temporal correlation scale and the signal-to-noise ratio, by minimizing the difference between the analysis and the reference climatology as in Sect. 6.1.

The optimal values of the analysis parameters are shown in Table 6. To compare the different variants, a skill score relative to the 2-D case (without advection) has been computed:

$$\text{skill score} = 1 - \frac{\text{RMS}^2}{\text{RMS}_{2-D}^2}. \quad (43)$$

It follows that inclusion of the advection constraint in the 2-D analysis improves the skill by 29%. It is surprising to see that this improvement is of similar amplitude as the improvement obtained by including the time dimension as the later requires the solution of a 12 times larger system (the number of months). Including the exact date of a measurement instead of its month leads to only a small improvement. The best analysis is obtained when using a 3-D domain in combination with the advection constraint leading to an improvement of 44%. Including the advection constraint has again the beneficial effect of increasing the optimal signal-to-noise ratio as the observations are more coherent along flow lines.

7 Conclusions

A variational analysis tool has been developed and tested with a realistic data distribution from Argo, but with pseudo-observations extracted from a model. This allows to compare

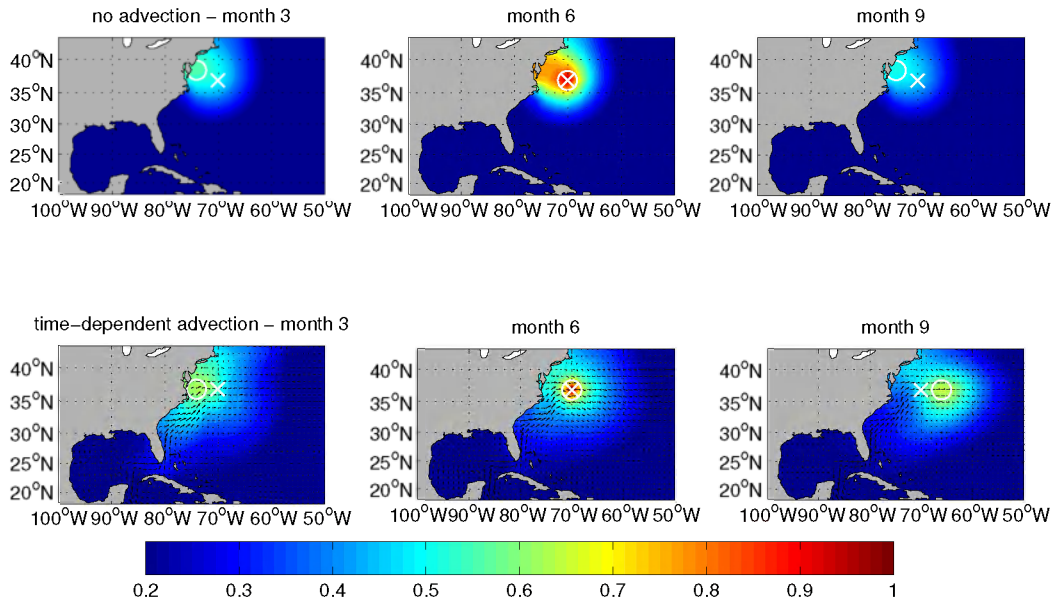


Fig. 7. 3-D case: background error covariance without (upper row) and with advection constraint (lower row) for a data point located at the cross and at month 6.

the analysis to model climatology data and to quantitatively compare different analyses. Parameters are optimized by minimizing the difference between the analysis and the model climatology. However in practice, a cross-validation data set is needed for such optimization, which ideally should be homogeneously distributed. An improvement with 3-D (longitude, latitude and time) versus 2-D analysis (horizontal only) was shown. A relatively larger reduction of the RMS error was also observed by including the advection constraint (stationary in the 2-D case and time-dependent in the 3-D case). However, it should be noted that the current fields used here are dynamically coherent with the tracer fields as they come from the same model. In a realistic setup with real observations, an improvement similar to the one reported here will require quite accurate current fields.

The source code of the tool *divand* is available at <http://modb.oce.ulg.ac.be/mediawiki/index.php/divand> and distributed under the terms of the General Public License. The toolbox will also be made available through to the Octave extensions repository Octave-Forge.

Appendix A

Hankel transform

We note $\hat{f}(\mathbf{k})$ the Fourier transform of a function $f(\mathbf{x})$:

$$\hat{f}(\mathbf{k}) = \int_{\mathbb{R}^n} e^{-i\mathbf{k}\mathbf{x}} f(\mathbf{x}) d^n \mathbf{x}. \quad (\text{A1})$$

If the function $f(\mathbf{x})$ is a radial function, $f(\mathbf{x}) = F(r)$, then its Fourier transform also depends only on the module of the wave-number vector, $\hat{f}(\mathbf{k}) = \hat{F}(k)$, where $r = |\mathbf{x}|$ and $k = |\mathbf{k}|$.

The Fourier transform of a radial function in \mathbb{R}^n is given in terms of the Hankel transform by (Arfken, 1985)

$$k^{\frac{n-2}{2}} \hat{F}(k) = (2\pi)^{\frac{n}{2}} \int_0^{\infty} J_{\frac{n-2}{2}}(kr) r^{\frac{n-2}{2}} F(r) r dr, \quad (\text{A2})$$

where $J_\nu(r)$ is the Bessel function of first kind of order ν . The inverse Fourier transform is given by a similar relationship:

$$r^{\frac{n-2}{2}} F(r) = (2\pi)^{-\frac{n}{2}} \int_0^{\infty} J_{\frac{n-2}{2}}(kr) k^{\frac{n-2}{2}} \hat{F}(k) k dk. \quad (\text{A3})$$

Acknowledgements. This work was funded by the SeaDataNet FP7 EU project (grant 283607), by the EU project EMODnet Chemistry, by the project PREDANTAR (SD/CA/04A) from the federal Belgian Science policy and the National Fund for Scientific Research, Belgium (F.R.S.-FNRS). The authors would like to thank two anonymous reviewers for their helpful comments and constructive criticism. This is a MARE publication.

Edited by: D. Ham

References

- Arfken, G.: *Mathematical Methods for Physicists*, Academic Press, Orlando, FL, 3rd Edn., 795 pp., 1985.
- Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics, *Q. J. R. Meteorol. Soc.*, 134, 1971–1996, doi:10.1002/qj.340, 2008.
- Barth, A., Alvera-Azcárate, A., Troupin, C., Ouberdous, M., and Beckers, J.-M.: A web interface for gridding arbitrarily distributed in situ data based on Data-Interpolating Variational Analysis (DIVA), *Adv. Geosci.*, 28, 29–37, doi:10.5194/adgeo-28-29-2010, 2010.
- Beckers, J.-M., Barth, A., and Alvera-Azcárate, A.: DINEOF reconstruction of clouded images including error maps – application to the Sea-Surface Temperature around Corsican Island, *Ocean Sci.*, 2, 183–199, doi:10.5194/os-2-183-2006, 2006.
- Bennett, A. F., Chua, B. S., and Leslie, L. M.: Generalized inversion of a global numerical weather prediction model, *Meteor. Atmos. Phys.*, 60, 165–178, 1996.
- Bennett, A. F., Chua, B., and Leslie, L.: Generalized inversion of a global numerical weather prediction model, II: Analysis and implementation, *Meteor. Atmos. Phys.*, 62, 129–140, doi:10.1007/BF01029698, 1997.
- Benzi, M.: *Preconditioning Techniques for Large Linear Systems: A Survey*, *J. Comput. Phys.*, 182, 418–477, doi:10.1006/jcph.2002.7176, 2002.
- Brankart, J.-M. and Brasseur, P.: The general circulation in the Mediterranean Sea: a climatological approach., *J. Marine Sys.*, 18, 41–70, doi:10.1016/S0924-7963(98)00005-0, 1998.
- Brankart, J.-M. and Brasseur, P.: Optimal analysis of in situ data in the Western Mediterranean using statistics and cross-validation., *J. Atmos. Ocean. Tech.*, 13, 477–491, doi:10.1175/1520-0426(1996)013<0477:OAOISD>2.0.CO;2, 1996.
- Brasseur, P.: *Reconstitution de champs d'observations océanographiques par le Modèle Variationnel Inverse: Méthodologie et Applications.*, Ph.D. thesis, Université de Liège, Collection des publications, Sciences appliquées, Liège, Belgium, 1994.
- Brasseur, P. and Haus, J.: Application of a 3-D variational inverse model to the analysis of echydrodynamic data in the Northern Bering and Southern Chuckchi Seas., *J. Marine Sys.*, 1, 383–401, doi:10.1016/0924-7963(91)90006-G, 1991.
- Brasseur, P., Beckers, J.-M., Brankart, J.-M., and Schoenauen, R.: *Seasonal Temperature and Salinity Fields in the Mediterranean Sea: Climatological Analyses of an Historical Data Set.*, *Deep-Sea Res.*, 43, 159–192, doi:10.1016/0967-0637(96)00012-X, 1996.
- Bretherton, F. P., Davis, R. E., and Fandry, C. B.: A technique for objective analysis and design of oceanographic experiment applied to MODE-73., *Deep-Sea Res.*, 23, 559–582, doi:10.1016/0011-7471(76)90001-2, 1976.
- Buongiorno Nardelli, B.: A novel approach for the high-resolution interpolation of in situ sea surface salinity, *J. Atmos. Ocean. Tech.*, 29, 867–879, doi:10.1175/JTECH-D-11-00099.1, 2012.
- Buongiorno Nardelli, B., Colella, S., Santoleri, R., Guarracino, M., and Kholod, A.: A re-analysis of Black Sea surface temperature, *J. Marine Sys.*, 79, 50–64, doi:10.1016/j.jmarsys.2009.07.001, 2010.
- Carrier, M. J. and Ngodock, H.: Background-error correlation model based on the implicit solution of a diffusion equation, *Ocean Modell.*, 35, 45 – 53, doi:10.1016/j.ocemod.2010.06.003, 2010.
- Carton, J. and Hackert, E.: Data assimilation applied to the temperature and circulation in the tropical Atlantic, 1983–84, *J. Phys. Oceanogr.*, 20, 1150–1165, doi:10.1175/1520-0485(1990)020<1150:DAATTT>2.0.CO;2, 1990.
- Chaves, J. C., Nehrbass, J., Guilfoos, B., Gardiner, J., Ahalt, S., Krishnamurthy, A., Unpingco, J., Chalker, A., Warnock, A., and Samsi, S.: Octave and Python: High-level scripting languages productivity and performance evaluation, in: *HPCMP Users Group Conference*, 2006, 429–434, IEEE, 2006.
- Chen, Y., Davis, T. A., Hager, W. W., and Rajamanickam, S.: Algorithm 887: CHOLMOD, Supernodal Sparse Cholesky Factorization and Update/Downdate, *ACM Trans. Math. Soft.*, 35, 22:1–22:14, doi:10.1145/1391989.1391995, 2008.
- Chua, B. S. and Bennett, A. F.: An inverse ocean modeling system, *Ocean Modell.*, 3, 137–165, doi:10.1016/S1463-5003(01)00006-3, 2001.
- Cohn, S. E., da Silva, A., Guo, J., Sienkiewicz, M., and Lamich, D.: Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System, *Mon. Weather Rev.*, 126, 2913–2926, 1998.
- Courtier, P.: Dual formulation of four-dimensional variational assimilation, *Q. J. R. Meteorol. Soc.*, 123, 2449–2461, doi:10.1002/qj.49712354414, 1997.
- Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. R. Meteorol. Soc.*, 120, 1367–1387, doi:10.1002/qj.49712051912, 1994.
- Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., and Fisher, M.: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation, *Q. J. R. Meteorol. Soc.*, 124, 1783–1808, doi:10.1002/qj.49712455002, 1998.
- Cullen, M. J. P.: Four-dimensional variational data assimilation: A new formulation of the background-error covariance matrix based on a potential-vorticity representation, *Q. J. R. Meteorol. Soc.*, 129, 2777–2796, doi:10.1256/qj.02.10, 2003.
- Daley, R.: *Atmospheric Data Analysis*, Cambridge University Press, New York, 457 pp., 1991.
- Davis, T. A. and Hager, W. W.: Dynamic supernodes in sparse Cholesky update/downdate and triangular solves, *ACM Trans. Math. Soft.*, 35, 27:1–27:23, doi:10.1145/1462173.1462176, 2009.
- Dee, D. P. and da Silva, A. M.: The Choice of Variable for Atmospheric Moisture Analysis, *Mon. Weather Rev.*, 131, 155–171, doi:10.1175/1520-0493(2003)131<0155:TCOVFA>2.0.CO;2, 2003.
- Derber, J. and Rosati, A.: A global oceanic data assimilation system, *J. Phys. Oceanogr.*, 19, 1333–1347, doi:10.1175/1520-0485(1989)019<1333:AGODAS>2.0.CO;2, 1989.
- Dhatt, G. and Touzot, G.: Une présentation de la méthode des éléments finis., in: *Collection Université de Compiègne*, Paris, S. A. Maloine, editor, 1984.
- Dobricic, S. and Pinardi, N.: An oceanographic three-dimensional variational data assimilation scheme, *Ocean Modell.*, 22, 89–105, doi:10.1016/j.ocemod.2008.01.004, 2008.
- Evensen, G.: *Data assimilation: the Ensemble Kalman Filter*, Springer, 279 pp., 2007.

- Gandin, L. S.: Objective analysis of meteorological fields, Israel Program for Scientific Translation, Jerusalem, 242 pp., 1965.
- Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, *Q. J. R. Meteorol. Soc.*, 125, 723–757, doi:10.1002/qj.49712555417, 1999.
- Gauthier, P., Charette, C., Fillion, L., Koclas, P., and Laroche, S.: Implementation of a 3d variational data assimilation system at the Canadian Meteorological Centre, Part I: The global analysis, *Atmos.–Ocean*, 37, 103–156, doi:10.1080/07055900.1999.9649623, 1999.
- Gneiting, T.: Correlation functions for atmospheric data analysis, *Q. J. R. Meteorol. Soc.*, 125, 2449–2464, doi:10.1002/qj.49712555906, 1999.
- Golub, G. H. and Van Loan, C. F.: *Matrix Computations*, Johns Hopkins University Press, Baltimore, 3rd Edn., 1996.
- Guinehut, S., Larnicol, G., and Traon, P. L.: Design of an array of profiling floats in the North Atlantic from model simulations, *J. Marine Sys.*, 35, 1–9, doi:10.1016/S0924-7963(02)00042-8, 2002.
- Guttorp, P. and Gneiting, T.: Studies in the history of probability and statistics XLIX On the Matérn correlation family, *Biometrika*, 93, 989–995, doi:10.1093/biomet/93.4.989, 2006.
- Haben, S. A., Lawless, A. S., and Nichols, N. K.: Conditioning of incremental variational data assimilation, with application to the Met Office system, *Tellus A*, 63, 782–792, doi:10.1111/j.1600-0870.2011.00527.x, 2011.
- Hayden, C. M., R. J. P.: Recursive Filter Objective Analysis of Meteorological Fields: Applications to NESDIS Operational Processing, *J. Appl. Meteor.*, 34, 3–15, doi:10.1175/1520-0450-34.1.3, 1995.
- Høyer, J. L. and She, J.: Optimal interpolation of sea surface temperature for the North Sea and Baltic Sea, *J. Marine Sys.*, 65, 176–189, doi:10.1016/j.jmarsys.2005.03.008, 2007.
- Jones, M. T. and Plassmann, P. E.: An improved incomplete Cholesky factorization, *ACM Trans. Math. Softw.*, 21, 5–17, doi:10.1145/200979.200981, 1995.
- Kaplan, A., Kushnir, Y., Cane, M., and Blumenthal, M.: Reduced Space Optimal Analysis for Historical Datasets: 136 Years of Atlantic Sea Surface Temperatures, *J. Geophys. Res.*, 102, 27835–27860, doi:10.1029/97JC01734, 1997.
- Legler, D. M. and Navon, I. M.: VARIATM-A FORTRAN program for objective analysis of pseudostress wind fields using large-scale conjugate-gradient minimization, *Comput. Geosci.*, 17, 1–21, doi:10.1016/0098-3004(91)90077-Q, 1991.
- Leros, A., Andreatos, A., and Zagorianos, A.: Matlab-Octave science and engineering benchmarking and comparison, in: Proceedings of the 14th WSEAS international conference on Computers: part of the 14th WSEAS CSCC multiconference, 746–754, 2010.
- Lorenc, A. C.: Iterative analysis using covariance functions and filters, *Q. J. R. Meteorol. Soc.*, 118, 569–591, doi:10.1002/qj.49711850509, 1992.
- Lorenc, A. C.: Development of an operational variational assimilation scheme, *J. Meteor. Soc. Japan*, 75, 339–346, 1997.
- Lorenc, A. C., Ballard, S. P., Bell, R. S., Ingleby, N. B., Andrews, P. L. F., Barker, D. M., Bray, J. R., Clayton, A. M., Dalby, T., Li, D., Payne, T. J., and Saunders, F. W.: The Met Office global 3-dimensional variational data assimilation scheme, *Q. J. R. Meteorol. Soc.*, 126, 2991–3012, doi:10.1002/qj.49712657002, 2000.
- Lorenzo, E. D., Moore, A. M., Arango, H. G., Cornuelle, B. D., Miller, A. J., Powell, B., Chua, B. S., and Bennett, A. F.: Weak and strong constraint data assimilation in the inverse Regional Ocean Modeling System (ROMS): Development and application for a baroclinic coastal upwelling system, *Ocean Modell.*, 16, 160–187, doi:10.1016/j.ocemod.2006.08.002, 2007.
- Martin, M. J., Hines, A., and Bell, M. J.: Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact, *Q. J. R. Meteorol. Soc.*, 133, 981–995, doi:10.1002/qj.74, 2007.
- Matérn, B.: *Spatial Variation*, Springer-Verlag, Berlin, 2nd Edn., 1986.
- Mathiot, P., Goosse, H., Fichefet, T., Barnier, B., and Gallée, H.: Modelling the seasonal variability of the Antarctic Slope Current, *Ocean Sci.*, 7, 455–470, doi:10.5194/os-7-455-2011, 2011.
- McIntosh, P. C.: Oceanographic data interpolation: Objective analysis and splines, *J. Geophys. Res.*, 95, 13529–13541, doi:10.1029/JC095iC08p13529, 1990.
- Mirouze, I. and Weaver, A. T.: Representation of correlation functions in variational assimilation using an implicit diffusion operator, *Q. J. R. Meteorol. Soc.*, 136, 1421–1443, doi:10.1002/qj.643, 2010.
- Moore, A. M., Arango, H. G., Broquet, G., Edwards, C., Veneziani, M., Powell, B., Foley, D., Doyle, J. D., Costa, D., and Robinson, P.: The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems: Part II – Performance and application to the California Current System, *Prog. Oceanog.*, 91, 50–73, doi:10.1016/j.pocean.2011.05.003, 2011a.
- Moore, A. M., Arango, H. G., Broquet, G., Powell, B. S., Weaver, A. T., and Zavala-Garay, J.: The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems: Part I – System overview and formulation, *Prog. Oceanog.*, 91, 34–49, doi:10.1016/j.pocean.2011.05.004, 2011b.
- Muccino, J. C., Luo, H., Arango, H. G., Haidvogel, D., Levin, J. C., Bennett, A. F., Chua, B. S., Egbert, G. D., Cornuelle, B. D., Miller, A. J., Lorenzo, E. D., Moore, A. M., and Zaron, E. D.: The Inverse Ocean Modeling System. Part II: Applications, *J. Atmos. Ocean. Tech.*, 25, 1623–1637, doi:10.1175/2008JTECHO522.1, 2008.
- Nelder, J. A. and Mead, R.: A simplex method for function minimization, *Comput. J.*, 7, 308–313, doi:10.1093/comjnl/7.4.308, 1965.
- Nerger, L., Janjić, T., Schröter, J., and Hiller, W.: A regulated localization scheme for ensemble-based Kalman filters, *Q. J. R. Meteorol. Soc.*, 138, 802–812, doi:10.1002/qj.945, 2012.
- Purser, R., Wu, W.-S., Parish, D., and Roberts, N.: Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: spatially inhomogeneous and anisotropic covariances, *Mon. Weather Rev.*, 131, 1524–1535, doi:10.1175//2543.1, 2003a.
- Purser, R. J., Wu, W.-S., Parrish, D. F., and Roberts, N. M.: Numerical Aspects of the Application of Recursive Filters to Variational Statistical Analysis. Part I: Spatially Homogeneous and Isotropic Gaussian Covariances, *Mon. Weather Rev.*, 131, 1524–1535, doi:10.1175/1520-0493(2003)131<1524:NAOTAO>2.0.CO;2, 2003b.
- Rabier, F. and Courtier, P.: Four-dimensional assimilation in the presence of baroclinic instability, *Q. J. R. Meteorol. Soc.*, 118, 649–672, doi:10.1002/qj.49711850604, 1992.

- Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A.: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics, *Q. J. R. Meteorol. Soc.*, 126, 1143–1170, doi:10.1002/qj.49712656415, 2000.
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, 108, 4407, doi:10.1029/2002JD002670, 2003.
- Roberts-Jones, J., Fiedler, E. K., and Martin, M. J.: Daily, Global, High-Resolution SST and Sea Ice Reanalysis for 1985–2007 Using the OSTIA System, *J. Climate*, 25, 6215–6232, doi:10.1175/JCLI-D-11-00648.1, 2012.
- Robinson, A. R.: Physical processes, field estimation and an approach to interdisciplinary ocean modeling, *Earth-Sci. Rev.*, 40, 3–54, doi:10.1016/0012-8252(95)00030-5, 1996.
- Troupin, C., Machin, F., Ouberdous, M., Sirjacobs, D., Barth, A., and Beckers, J.-M.: High-resolution Climatology of the North-East Atlantic using Data-Interpolating Variational Analysis (Diva), *J. Geophys. Res.*, 115, C08005, doi:10.1029/2009JC005512, 2010.
- Troupin, C., Barth, A., Sirjacobs, D., Ouberdous, M., Brankart, J.-M., Brasseur, P., Rixen, M., Alvera-Azcárate, A., Belounis, M., Capet, A., Lenartz, F., Toussaint, M.-E., and Beckers, J.-M.: Generation of analysis and consistent error fields using the Data Interpolating Variational Analysis (DIVA), *Ocean Modell.*, 52–53, 90–101, doi:10.1016/j.ocemod.2012.05.002, 2012.
- Štajner, I., Riishøjgaard, L. P., and Rood, R. B.: The GEOS ozone data assimilation system: Specification of error statistics, *Q. J. R. Meteorol. Soc.*, 127, 1069–1094, doi:10.1002/qj.49712757320, 2001.
- Wahba, G.: Spline models for observational data, vol. 59 of CBM-NSF Regional Conference Series in Applied Mathematics, SIAM J. Appl. Math., Philadelphia, PA, 169 pp., 1990.
- Wahba, G. and Wendelberger, J.: 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, 1980.
- Weaver, A. T. and Courtier, P.: Correlation modelling on the sphere using a generalized diffusion equation, *Q. J. R. Meteorol. Soc.*, 127, 1815–1842, doi:10.1002/qj.49712757518, 2001.
- Weaver, A. T. and Mirouze, I.: On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation, *Q. J. R. Meteorol. Soc.*, 139, 242–260, doi:10.1002/qj.1955, 2013.
- Weaver, A. T., Vialard, J., and Anderson, D. L. T.: Three- and Four-Dimensional Variational Assimilation with a General Circulation Model of the Tropical Pacific Ocean. Part I: Formulation, Internal Diagnostics, and Consistency Checks, *Mon. Weather Rev.*, 131, 1360–1378, doi:10.1175/1520-0493(2003)131<1360:TAFVAW>2.0.CO;2, 2003.
- Yaremchuk, M. and Sentchev, A.: Mapping radar-derived sea surface currents with a variational method, *Cont. Shelf Res.*, 29, 1711–1722, doi:10.1016/j.csr.2009.05.016, c09009, 2009.
- Yari, S., Kovačević, V., Cardin, V., Gačić, M., and Bryden, H. L.: Direct estimate of water, heat, and salt transport through the Strait of Otranto, *J. Geophys. Res.*, 117, C09009 doi:10.1029/2012JC007936, 2012.
- Zaron, E. D., Chavanne, C., Egbert, G. D., and Flament, P.: Baroclinic tidal generation in the Kauai Channel inferred from high-frequency radio Doppler current meters, *Dyn. Atmos. Oceans*, 48, 93–120, doi:10.1016/j.dynatmoce.2009.03.002, 2009.