

Published 2016 in IGI Global:

<http://www.igi-global.com/book/oceanographic-marine-cross-domain-data/148510>

doi: 10.4018/978-1-5225-0700-0

Analysis of Ocean In Situ Observations and Web-Based Visualization From Individual Measurements to an Integrated View

Alexander Barth

University of Liège, Belgium

Sylvain Watelet

University of Liège, Belgium

Charles Troupin

Balearic Islands Coastal Ocean Observing and Forecasting System (SOCIB), Spain

Aida Alvera-Azcárate

University of Liège, Belgium

Jean-Marie Beckers

University of Liège, Belgium

ABSTRACT

The sparsity of observations poses a challenge common to various ocean disciplines. Even for physical parameters where the spatial and temporal coverage is higher, current observational networks undersample a broad spectrum of scales. This situation is generally more severe for chemical and biological parameters because such sensors are less widely deployed. The present chapter describes the analysis tool DIVA (Data-Interpolating Variational Analysis) which is designed to generate gridded fields from in situ observations. DIVA has been applied to various physical (temperature and salinity), chemical (concentration of nitrate, nitrite and phosphate) and biological parameters (abundance of a species). The chapter also shows the technologies used to visualize the gridded fields. Visualization of analyses from in situ observations provide a unique set of challenges since the accuracy of the analysed field is not spatially uniform as it strongly depends on the location of the observations. In addition, an adequate treatment of the depth and time dimensions is essential.

Keywords: Variational Inverse Method, Data-Interpolating Variational Analysis, Statistical Modelling, In Situ Measurements, Climatology, Data Products, Web-Based Visualization, Web Map Service

[ABSTRACT](#)

[INTRODUCTION](#)

[BACKGROUND](#)

[MAIN FOCUS OF THE CHAPTER](#)

[DATA-INTERPOLATING VARIATIONAL ANALYSIS](#)

[Basic formulation](#)

[Advection constraint](#)

[Anamorphosis transformation](#)

[Detrending](#)

[Analysis in more than two dimensions](#)

[APPLICATION OF DIVA](#)

[Interpolation of glider sections](#)

[Post processing of DIVA analyses](#)

[VISUALIZATION OF GRIDDED FIELDS](#)

[Scalar plots](#)

[Vector plots](#)

[Location markers](#)

[Deployment and configuration of OceanBrowser](#)

[VISUALIZATION OF IN SITU OBSERVATIONS](#)

[CONCLUSION](#)

[ACKNOWLEDGEMENTS](#)

[REFERENCES](#)

[KEY TERMS AND DEFINITIONS](#)

[REVIEWERS' COMMENTS](#)

[Reviewer 1](#)

[Reviewer 2](#)

INTRODUCTION

In situ measurements of ocean properties are generally sparsely distributed and thus undersample the ocean variability. A good knowledge of the scales and dynamics of the ocean is essential to make the best use of this limited amount of data. The data distribution is often very inhomogeneous, both in space and time. Accuracy is also inhomogeneous and point measurements (e.g., moorings, coastal

stations) are not necessarily representative of typical ocean conditions. Therefore it is necessary to take all these aspects into account when analysing in situ measurements.

Various applications and interpretations of marine data require a gridded field covering the complete domain as a first step. These applications include, for example:

- computation of budgets (e.g., heat, salt content),
- identification of long-term trends,
- identification and characterization of oceanographic features,
- determination of derived variables from parameters not necessarily measured at the same location (for example density which is a function of temperature and salinity),
- model initialization and validation.

In principle, some of these applications could be also performed directly on the measured observations. For example, the average ocean temperature could be naively computed as the mean of all observations. However in practice the data distribution is very inhomogeneous in space (with a higher density of observations near the surface and close to the coast) and in time (e.g., more ship-based observations are generally available during summer than during winter and the density of observations has increased substantially during the last decades). Hence computing the mean using directly the observations would result in a biased estimation of the average temperature.

BACKGROUND

Most analysis methods are linear techniques, which means that the analysed field (or its anomaly versus a reference state) at a given location is a weighted linear combination of the observations. Different methods exist to compute those weights. The most obvious way to compute them is to use a monotonic decreasing function of the distance between the location of the analysed field and the observations: observations farther away from a given location have less influence than observations nearby. This approach is implemented in the Cressman method (Cressman, 1959). The weighting function usually decreases over a given length-scale and is exactly zero beyond a given search radius. Various weighting schemes have been proposed (Barnes, 1964). The search radius can be made to vary in space depending on data coverage and/or dynamical scales. The Cressman weighting is a very simple and numerically quite efficient method. However, it suffers from some limitations, in particular no estimate can be obtained at locations where no observation is located within the search radius. In regions with very few observations, the method can therefore return a discontinuous field. In addition, the presence of topographic barriers disconnecting different water masses cannot be taken into account easily. Another limitation is that all observations are assumed to have a similar error variance since the weighting is based only on distance.

The optimal interpolation method (Gandin, 1965; Bretherton et al., 1976) solves some shortcomings of the Cressman method. A background field represents the first guess of the field. Observations are treated as anomalies relative to this background field. Optimal interpolation is a statistical approach in the sense that the result minimizes the total expected error provided the error covariance of the background field and of the observations are known. In practice, it is however difficult to estimate the error covariance of the observations, and even more difficult to estimate the error covariance of the background field. Most approaches use simple analytical correlation functions based on distance to parameterize the background error covariance (Robinson, 1996; Guinehut et al., 2002; Roberts-Jones et al., 2012), but more sophisticated approaches based on empirical orthogonal functions (Kaplan et al., 1997; Rayner et al., 2003; Beckers et al., 2006) or an ensemble of states derived from e.g. a numerical model (Evensen, 2007) have also been used. The analysis derived from optimal interpolation naturally provides results defined over the entire domain over which the first guess is defined. Far away from an observation, the analysed field gradually reverts back to the first guess. Unlike the Cressman weighting, optimal interpolation allows the computation of the relative error variance of the analysed field. The major difficulty of optimal interpolation is to derive a realistic error covariance and its parameterization is an active field of research nowadays (Hyer and She, 2007; Buongiorno Nardelli et al., 2010; Buongiorno Nardelli, 2012).

A different approach to derive a complete field based on in situ observations can be obtained by defining a cost function. Instead of minimizing the total expected error variance of the field as in optimal interpolation, the variational methods minimize the misfit relative to observations while imposing additional constraints. In particular, the resulting field must be smooth over a given length-scale. It has been shown that this approach can be seen as mathematically equivalent to optimal interpolation, with the peculiarity that the background error covariance is not defined explicitly as in optimal interpolation but implicitly in terms of a regularization constraint in the cost function being minimized. Boundary conditions of the regularization constraint naturally decouple basins which are geographically close but not connected due to the topography. While not impossible, it is quite difficult to include such constraints in optimal interpolation. Among the variational methods one can distinguish between techniques similar to 3D-Var where the regularization constraint is defined by a diffusion operator (Bennett et al., 1996; Weaver et al., 2003; Lorenzo et al., 2007; Moore et al., 2011) or spline interpolation where the cost function minimizes a series of derivatives of the field to ensure a smooth result (Wahba and Wendelberger, 1980; Wahba, 1990; McIntosh, 1990; Brasseur and Haus, 1991; Brasseur et al., 1996; Troupin et al., 2012). While the first class of methods generally uses an iterative solver, in particular the conjugate gradient method, the second class of formulations requires an efficient sparse matrix solver (Chen et al., 2008; Davis and Hager, 2009). Because the mathematical formulation of the variational approach can be seen as a special case of optimal interpolation, it is not possible to give a clear-cut distinction between these approaches.

Other related data analysis techniques consist of expressing the analysed field as a combination of a set of normal basis functions. These basis functions can be for example Fourier modes, the eigenvectors of the Laplacian operator (Lipphardt et al., 2000; Kaplan and Lekien, 2007) or empirical orthogonal functions (Beckers and Rixen, 2003; Alvera-Azcárate et al., 2009). These techniques are in fact related to optimal interpolation and variational analysis, since the solution of the latter methods can also be seen as a linear combination of the eigenvectors of the underlying background error covariance matrix (Beckers et al., 2006).

As ocean observations are sparse, gridding these observations by the aforementioned techniques is often a first step to visualize fields of ocean parameters. The DIVA tool has been integrated in the visualization tool Ocean Data View (Schlitzer, 2002) for this purpose. In the following we limit ourselves to web-based visualization. The Reading e-Science center developed ncWMS which is a primarily Web Map Service (de la Beaujardiere, 2006) to visualize netCDF files (Blower et al., 2013). ncWMS is a Java application which is deployed using a standard servlet container such as Apache Tomcat. On top of ncWMS, the browser-based client GODIVA2 has been developed (Blower et al., 2009). Other web-based clients include Lw4nc2 (Lightweight for netCDF viewer) and Leaflet.TimeDimension (<https://github.com/socib/Leaflet.TimeDimension>) both developed by SOCIB, Spain.

The present chapter describes the web-based visualization tool OceanBrowser. The server component is written in Python using the matplotlib library (Hunter, 2007).

These web-based visualization systems have different strengths and weaknesses. The focus of the development of OceanBrowser has been the ability to visualize along arbitrary vertical sections. Those systems are also compatible to a certain degree thanks to the Web Map Service protocol published by the Open Geospatial Consortium. However, some extensions to the protocol are necessary to cover the specific case of analysing ocean data.

MAIN FOCUS OF THE CHAPTER

This chapter aims to address the following issues in the context of managing marine observations:

- It introduces the problem of the data analysis of in situ observations.
- The challenges in deriving gridded fields and the visualization of the data fields will be discussed.

- Based on a concrete implementation, it will show how web-based tools and open standards can be used to integrate datasets from different oceanographic disciplines and to facilitate the interpretation of the dataset across different domains.

DATA-INTERPOLATING VARIATIONAL ANALYSIS

An analysis tool called DIVA (Data-Interpolating Variational Analysis) has been developed to create gridded fields from sparse in situ observations. The purpose of the analysis is to give the best estimation of a continuous spatial field of a given variable from a set of discrete measurements, taking into account the errors attached to the measurements.

Basic formulation

The Variational Inverse Method (VIM) implemented in DIVA aims to find an analysed field that satisfies a set of constraints over a given domain, expressed in the form of a cost function. Before elaborating its mathematical formulation, we consider that we are working with data anomalies: a reference (first guess) field is subtracted from the data before any analysis. In DIVA, this reference field can be either the null guess (i.e., data are already considered as anomalies), the data average, a regression plane derived from the data set or even a previous DIVA analysis (on a longer period for example). Once the analysis is performed, the reference field is added in order to return to the original variable. For a two-dimensional problem, the cost function J to minimize is defined as follows (Troupin et al., 2012):

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

with

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

where the coefficient α_0 penalizes the field itself, α_1 penalizes gradients (no trends), α_2 penalizes variability (regularization), and μ_j coefficients penalize data-analysis misfits (objective analysis). Thus, this cost function takes into account the deviation between analysis $\varphi(x_j, y_j)$ and data d_j

(observation constraint) and the smoothness of the analysis (regularization constraint), while it is also possible to take the advection by currents into account (physical constraint; see next subsection).

Introducing the characteristic length L , we can define dimensionless space coordinates: $\widehat{\nabla} = L \nabla$ and $D = L^2 \widehat{D}$, and rewrite the equations (1) and (2) as

$$J[\varphi] = \sum_{j=1}^{N_d} \mu_j L^2 [d_j - \varphi(x_j, y_j)]^2 + \int_{\widehat{D}} (\alpha_2 \widehat{\nabla} \widehat{\nabla} \varphi : \widehat{\nabla} \widehat{\nabla} \varphi + \alpha_1 L^2 \widehat{\nabla} \varphi \cdot \widehat{\nabla} \varphi + \alpha_0 L^4 \varphi^2) d\widehat{D}. \quad (3)$$

Without loss of generality, we can choose $\alpha_2 = 1$. α_0 fixes the characteristic length L for which the first and the last terms are equivalent:

$$\alpha_0 L^4 = 1. \quad (4)$$

The term $\mu_j L^2$ accounts for the weight on each individual observation, depending on the signal (σ^2 , the variance of the background field) to noise (ε_j^2) ratio (see also Barth et al., 2014),

$$\mu_j L^2 = 4\pi \frac{\sigma^2}{\varepsilon_j^2}. \quad (5)$$

Finally, the term $\alpha_1 L^2$ fixes the relative importance of gradients. In order to penalize the gradients with a similar weight than the second derivatives, we set

$$\alpha_1 L^2 = 2\zeta \quad (6)$$

where the adimensional shape factor ζ is set to 1 by default in DIVA.

The overall signal-to-noise ratio λ and the overall error variance of the observations ε^2 are defined as follows:

$$\lambda = \frac{\sigma^2}{\varepsilon^2} \quad ; \quad \varepsilon^2 = \frac{\sum_{j=1}^N \varepsilon_j^2}{N} \quad (7)$$

This allows one to specify the following relative weights:

$$w_j = \frac{\mu_j L^2}{4\pi\lambda} \quad \text{with} \quad \sum_{j=1}^N \frac{1}{w_j} = N. \quad (8)$$

Given ζ and w_j , the relative importance of each constraint (μ , α_0 , α_1 , α_2) is thus linked to two main control parameters: the signal-to-noise ratio λ and the correlation length L . Both can be obtained automatically from the datasets using DIVA tools or can be left to the user's choice.

To ensure that the analysis problem is well-posed, derivatives of sufficiently high order have to be included in the cost function. In fact for an analysis in n dimension, the cost-function should include at least a derivative of the order of $1+n/2$ rounded up for an odd number of dimensions. This guarantees that the analysis problem is well-posed and has a continuous derivative (Barth et al, 2014).

If the covariances are chosen accordingly, the VIM is equivalent to the optimal interpolation method. However, the number of operations in the VIM is almost independent of the number of data, leading to a potential gain of computational time for large problems. And since the minimization problem of the VIM is solved by a finite-element method with a mesh (Figure 1) only defined on the ocean, the spurious propagation of information through topographic barriers is easily avoided, unlike most implementations of optimal interpolation.

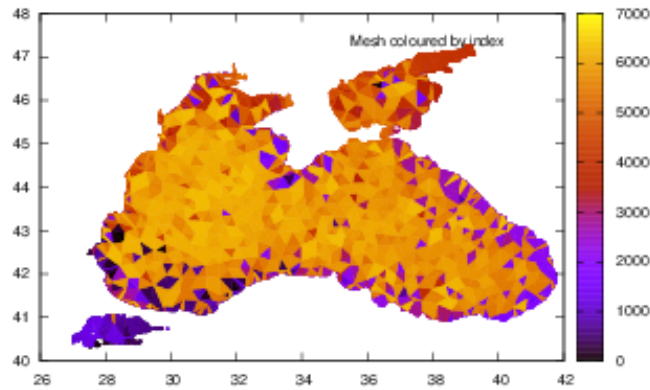


Figure 1: A finite element mesh computed by DIVA when using a correlation length of 1.5° over the Black Sea at the surface. Each color corresponds to a different triangular element mesh. Note the smaller mesh size near the coastlines.

Advection constraint

The cost function J can be augmented with an optional behaviour constraint, in order to use the advection field as a complementary source of information for the interpolation. Let us consider the general equation for a scalar oceanic quantity φ (Brasseur, 1995):

$$\frac{\delta\varphi}{\delta t} + \mathbf{u} \cdot \nabla \varphi + v_3 \frac{\delta\varphi}{\delta x_3} = Q^\varphi + \lambda_h \nabla \nabla \varphi + \frac{\delta}{\delta x_3} (\lambda_v \frac{\delta\varphi}{\delta x_3}) \quad (9)$$

where \mathbf{u} represents the horizontal speed, v_3 the vertical component of the speed, λ_h and λ_v are horizontal and vertical coefficients of diffusion, and Q^φ is the term corresponding to the source/decay of φ .

In equation (9), we can invoke certain simplifying assumptions. In most cases, we can neglect Q^o , for instance when analysing temperature or salinity for which the sources are confined at the air-water interface. Further, we ignore the temporal derivative in equation (9) since DIVA aims to deal with slowly evolving situations, such as climatologies. Besides, DIVA performs its 3D analyses via superimposed 2D analyses, so the vertical movement cannot be considered in equation (9). With these assumptions in mind, we get:

$$\mathbf{u} \cdot \nabla \phi \approx \lambda_h \nabla \nabla \phi \quad (10)$$

and the resulting advection-diffusion constraint to add to equation (1) :

$$C[\phi] = \gamma \{ \mathbf{u} \cdot \nabla \phi - \lambda_h \nabla \nabla \phi \}^2 \quad (11)$$

where γ weights the relative importance of this advection term with respect to the observations and regularization ones, in order to be in the same order of magnitude. In the dimensionless version (equation 3) of the cost function $J[\phi]$, the advection constraint is added as follows (Troupin et al., 2012):

$$J_a[\phi] = J[\phi] + \theta U^2 \int_{\hat{D}} [\mathbf{u} \cdot \nabla \phi - \frac{A}{L} \nabla \nabla \phi]^2 d\hat{D} \quad (12)$$

where U is the characteristic velocity scale computed from \mathbf{u} , A is the horizontal diffusion coefficient, and θ a scaling factor ensuring the same weight as the other constraint when set to 1.

This advection constraint tends to increase the correlation in the direction of \mathbf{u} (Figure 2), which can be derived from isobaths, isopycnals, can be the geostrophic velocity computed from temperature and salinity, or even provided from the output of a hydrodynamic model (Figure 3).

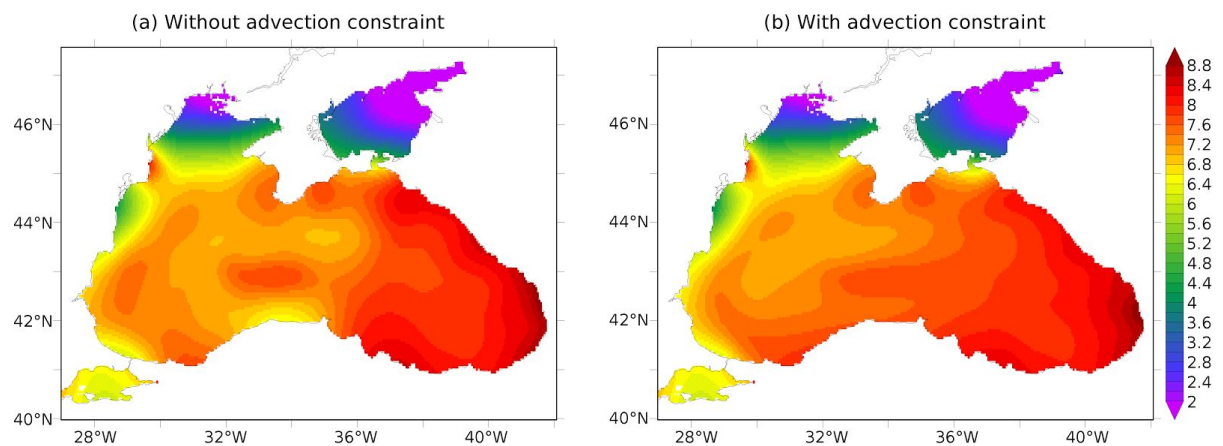


Figure 2: Example of SST analysis over the Black Sea without advection constraint (panel a). The World Ocean Database 2013 was used for data extraction. The analysis was performed between 1980 and 2010 on the month of March. Panel (b) shows the same SST analysis, with the advection

constraint enabled. The used velocity field is shown in Figure 3. This advection constraint tends to smooth the short scales gradients where the currents are strong, especially in the central part of the Black Sea (thanks to the enhanced propagation of the information).

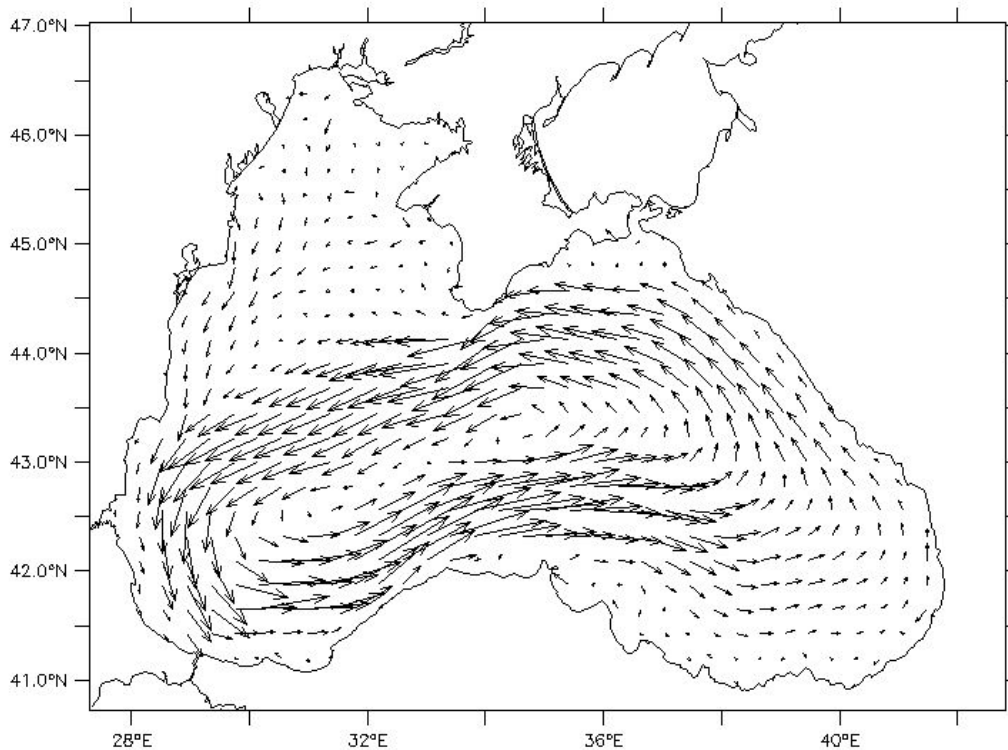


Figure 3: The velocity field we used for this experiment is described in Capet et al. (2012). It consists of a 1980-2010 average of the surface currents output of an eddy-resolving 3D model, for the month of March.

Anamorphosis transformation

Using a non-gaussian distributed variable can lead to unrealistic values when performing the DIVA analysis. Indeed, DIVA tries to interpolate at best where no data are available, which can imply negative values if the nearest observations are close to zero. Obviously, we have to reject such a result for variables such as salinity, chlorophyll-a concentration, or other concentration values.

An easy way to overcome this issue is to apply a transformation on the data prior to the DIVA analysis. After the analysis, the inverse transformation is applied on the analyzed field. Several transformations have been implemented in DIVA: logarithm, logit and anamorphosis. While the two

first transformations are predefined, the last one is directly based on the shape of the original data histogram in view of creating a gaussian-like distribution which favours more realistic analyzed values. An example is shown in Figure 4.

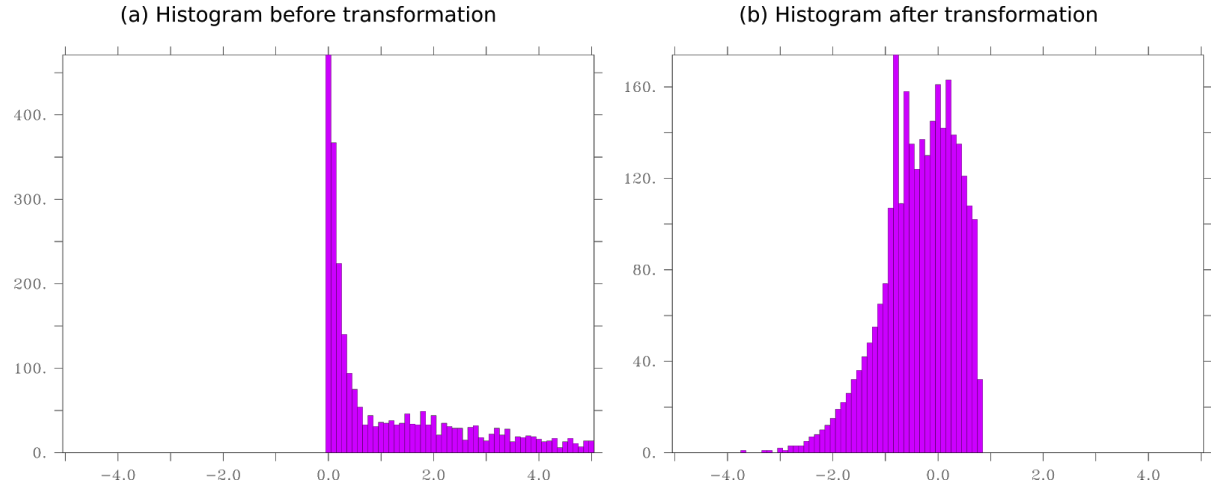


Figure 4: Histogram of nitrate concentration [$\mu\text{mol/l}$] at the surface of the Black Sea between 1900 and 2015 (panel a) and histogram of nitrate concentration after the anamorphosis transformation (panel b). Its shape is as close as possible to a gaussian distribution.

Detrending

One of the most challenging issues in data analysis is the uneven temporal distribution of data, which causes representativity errors. For example, the observation frequency is very often higher in summer than in winter, leading to a positive bias in the temperature analysis.

A detrending method, described in Duchon (1977), has been implemented in DIVA and applied by Capet et al. (2014) in order to recognize and eliminate this bias. Let us define the groups as the various time intervals which can be considered for any trend detection: years, seasons, months... In one of these groups (e.g. the years), each observation corresponds to a unique class c_j (e.g. 1952). The data analysis misfit term in equation (1) can be rewritten by including an unknown trend value for each class (dC_1, dC_2, \dots):

$$\sum_{j \in C_1} \mu_j [d_j - dC_1 - \varphi(x_j, y_j)] + \sum_{j \in C_2} \mu_j [d_j - dC_2 - \varphi(x_j, y_j)] + \dots \quad (13)$$

If the function $\varphi(x_j, y_j)$ were known, we could easily derive the dC_i from a minimization on each class:

$$dC_1 = \frac{\sum_{j \in C_1} \mu_j [d_j - \varphi(x_j, y_j)]}{\sum_{j \in C_1} \mu_j} \quad (14)$$

and similarly for the other classes. But since $\varphi(x,y)$ is unknown too, we have to solve the issue by an iterative process, starting by an analysis without detrending. Using this field $\varphi(x,y)$, we compute a first guess of the trends in each group and subtract it from the original data. After that, a new analysis is performed, the trends are recalculated, and the iterations continue until a convergence criterion is fulfilled. This method has been successfully used in Capet et al. (2014) for the cold content of the Black Sea cold intermediate layer as well as for its mixed layer depth. In both cases, the method allowed the reduction of the bias caused by the uneven data availability between summer and winter or between cold and warm years.

Analysis in more than two dimensions

DIVA is implemented to work on a two-dimensional space. This is generally a horizontal plane at a given depth and for a given time period. But DIVA can also be applied to a vertical section. For multidimensional applications, for instance deriving an ocean property as a function of longitude, latitude, depth and time, horizontal analyses are performed independently for various depth and time instances and then assembled in a 4-dimensional analyses.

Maintaining coherency and avoiding abrupt variations in the depth and time dimension is difficult in this case. Sometimes it is necessary to filter the parameters (in particular the correlation length and signal-to-noise ratio) along the depth and time dimensions and to use data from adjacent layers in the vertical. DIVA has been implemented to work in two dimensions only because it uses a triangular mesh and a finite element method which are increasingly difficult to implement in 3 or 4 dimensions. However, the variational inverse method on structured grids can be easily extended to higher dimensions. The tool DIVAnd uses structured grids and can in fact work with an arbitrary high number of dimensions. The limiting factor are the available resources (in particular memory and cpu time). DIVAnd solves the equations (1) and (2) where the gradient and Laplacian are the n -dimensional operators. The grid must be structured, but it can be any orthogonal curvilinear grid. For instance one could use a cartesian or a spherical grid.

To assess the skill of the analysis method, Barth et al. (2014) used pseudo-observations (extracted from a global NEMO model at about $\frac{1}{4}$ degree resolution) at the location of the ARGO profiles to reconstruct monthly global surface temperatures. The use of a 3D-analysis versus a 2D-analysis (without advection constraint) resulted in an improvement of 27% while the best results were achieved with a 3D-analysis with advection constraint. This resulted in an RMS error of 0.85°C , a skill improvement of 44% relative to a 2D-analysis without advection constraint.

The more limited the data coverage is, the more it becomes important to efficiently use observations from close-by depth and time instances. One can therefore expect that the benefit of a multidimensional analysis would be even larger for observations from biological or chemical sensors where the data coverage is poorer.

DIVAnd is written in the MATLAB programming language, but it also runs efficiently under GNU Octave which implements a large part of the MATLAB syntax and functions. It can therefore be used running only free software which is an important aspect to facilitate the adoption of this software by new users.

APPLICATION OF DIVA

DIVA supports the ODV4 spreadsheet format (Ocean Data View, version 4) or simple text files as inputs, which can contain either profiles or time series. These time series can be handled by a weighting option, in order to correct the bias caused by the over-abundance of data on a particular zone (temporal and/or spatial). DIVA was originally designed to work in 2D, although it is now possible to solve pseudo 4D problems. In that case, the problem is divided into 2D layers associated with each other by a vertical filtering of two statistical parameters: the signal-to-noise ratio and the correlation length. Among other parameters, the user can choose the periods (years and months), the depths of the 2D layers and the variables. The results are then stored in a combined netCDF file.

For each 2D analysis, DIVA requires a contour usually representing the interface ocean/land at the considered depth. The topography used to create these contours is left to the user's choice, and several possibilities of extraction are proposed. The user has the possibility to set a variable correlation length over the domain, for instance in the case of smaller structures near the coast.

DIVA is mainly released as a command-line tool, although alternative (lighter) distributions include: DIVA for VirtualBox, DIVA for Matlab, Diva-on-Web and DIVA in Ocean Data View. It can be run on various operating systems (Linux, Mac OS X and Windows).

In order to lower the barrier to use the DIVA products, a web-based viewer has been developed which aims to provide an easy access to the interpolated fields as well as to auxiliary variables such as the error field and the number of observations per grid cell.

Interpolation of glider sections

The gridding of data along vertical sections (horizontal coordinate vs. depth) is a particular case of DIVA. The contour that delimits the domain of analysis is easily constructed from the bathymetry. The main difference to take into account is the difference of magnitude between the vertical (10-100 m) and horizontal scale (10-100 km). An analysis performed without considering this difference, i.e. assuming that the domain is isotropic, would lead to a gridded field that does not realistically represent the characteristics of the ocean. The solution adopted and implemented in DIVA consists in applying a coordinate change to the data, based on the ratio between the vertical scale and the horizontal scale.

The following example describes an application gridding glider observations in a vertical section located in the Ibiza Channel, between the Ibiza Island and Spain mainland (Figure 5). The vertical

resolution of the data collected by a glider can be very high, thus the use of a gridding tool able to cope with this is often desired.

SOCIB CANALES Mission, September 2013

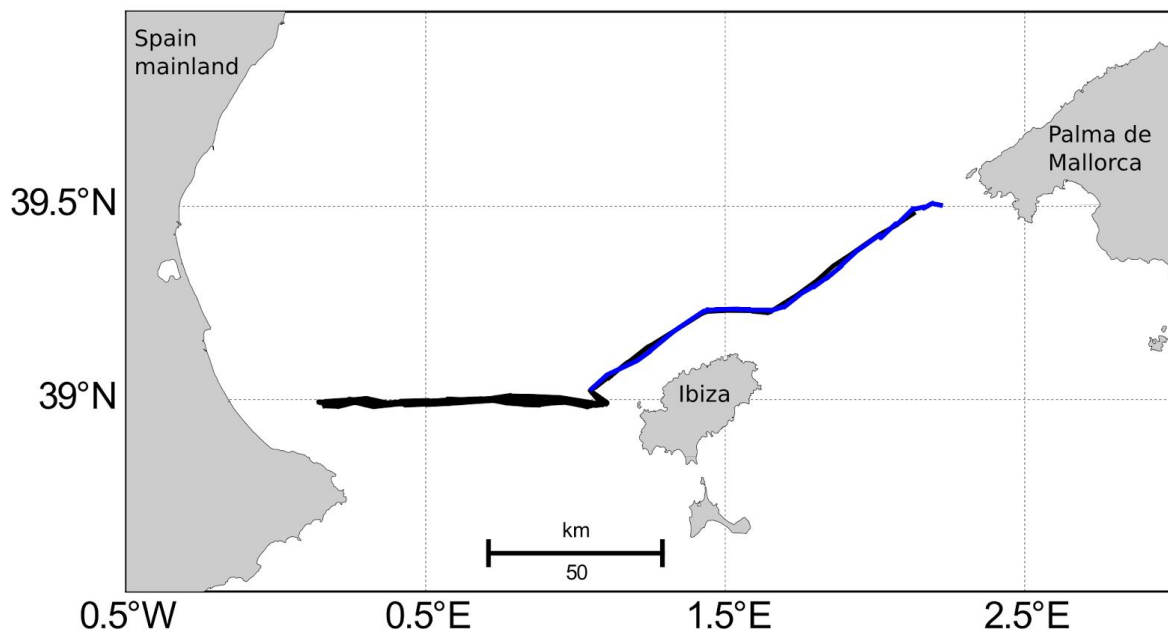


Figure 5: track of the CANALES mission in the Balearic Sea in September 2013. The blue line is the section that will be interpolated. The data are open and freely available from www.socib.es

For this mission, the sampling strategy in the Ibiza Channel was to try to close the section with profiles going from the surface to the bottom (approximately 950 m). The transect (blue line) was repeated 4 times in September-October 2013. Temperature and salinity were acquired for all the transects, while chlorophyll-a concentration was only measured during the first of the four transects, and until a depth of 300 m. The number of data points for the studied transect is as follows: 39575 for salinity and 8452 for chlorophyll-a concentration.

The gridded fields are shown between the surface and 325 m depth for both variables. In this layer, the salinity increases with depth, with the lowest values in the surface layer at the two ends of the Ibiza Channel (Figure 6). Although the interpolation with DIVA has a smoothing effect on the data, the gridded field still displays small scale features characteristic of the studied area.

The chlorophyll-a concentration exhibits a deep chlorophyll-a maximum located between 50 and 100 m depth, with values exceeding 0.7 mg/m^3 . Below 150 m, the values are very close to zero (Figure 7).

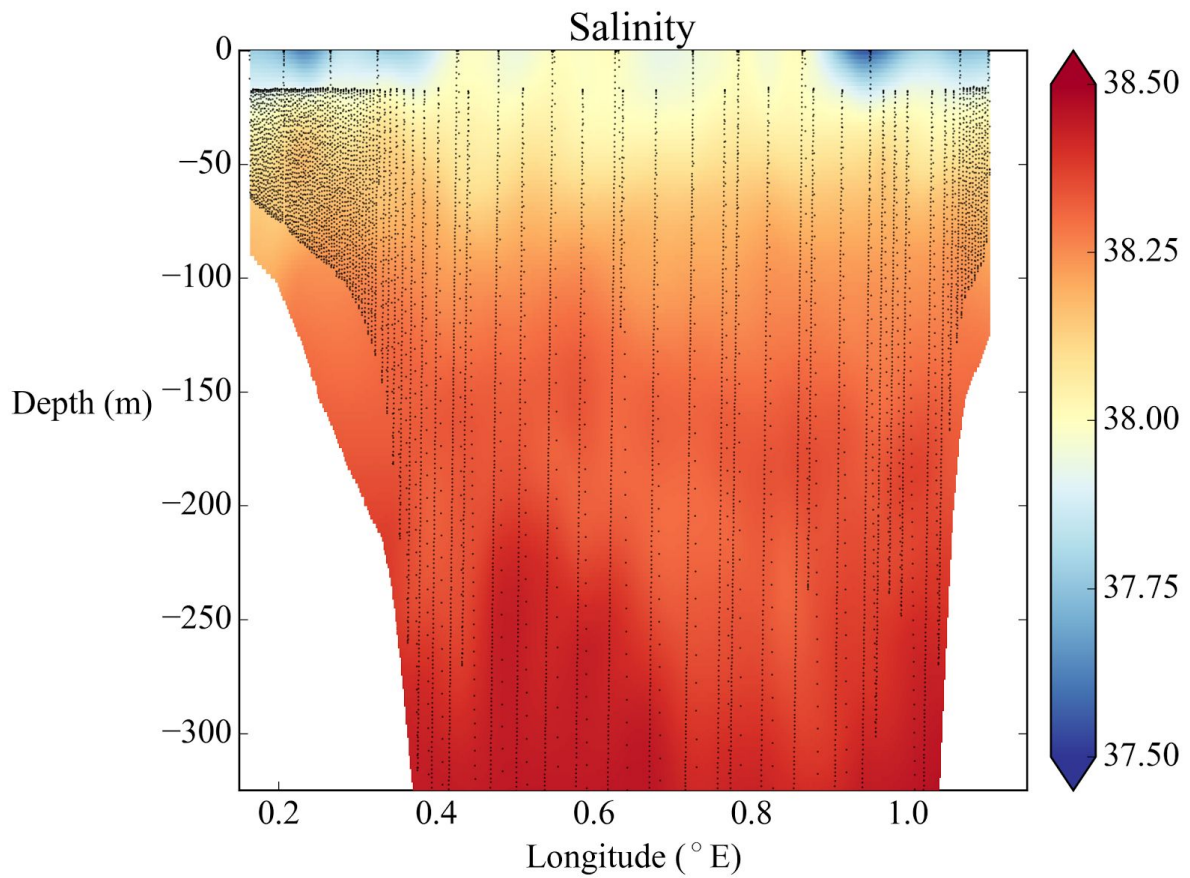


Figure 6: vertical section of salinity obtained with DIVA, between the surface and 300 m depth. Black dots indicate data location (only one out of 10 points represented, for the sake of clarity). The bathymetry is extracted from the GEBCO bathymetry with a 30-minutes resolution.

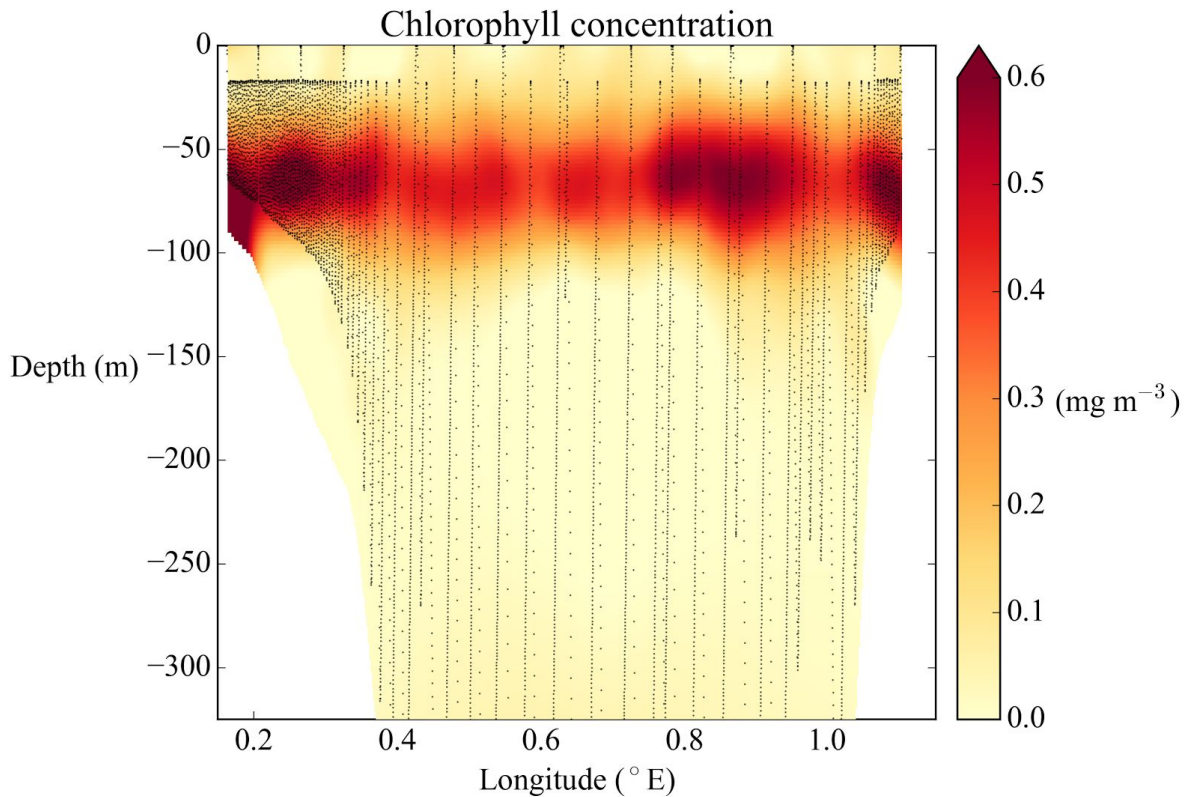


Figure 7: vertical section of chlorophyll-a concentration obtained with DIVA, between the surface and 300 m depth.

Post processing of DIVA analyses

The final 4D netCDF file includes all useful information about the performed analysis. Eight 4D-variables are available: the analysed field, the absolute error, the relative error, the analyzed field masked by two thresholds (0.3 and 0.5) of relative error (Beckers et al., 2014), the data bins (number of observations in each grid cell over the domain), the outliers bins and the (variable or constant) correlation length field. Three 3D-variables are also available: the analyzed field at the deepest analyzed level, with or without a masking from the relative error field (same two thresholds). The correlation length, the signal-to-noise ratio and the variance of the background field are also available to make the analysis reproducible. Finally, some 1D-variables related to the original observations have been added to this combined netCDF file in order to keep a record of all observations used in an analysis.

On top of that, several global attributes give more metadata information about the creation of the file itself :

- the netCDF conventions followed for its creation,

- the name of the funding project,
- the name of the institution,
- the name of the production group,
- a web link towards the downloading of the products,
- a web link towards the visualization of the products (e.g. OceanBrowser),
- the author email,
- the creation date,
- the title of the netCDF,
- the filename and path,
- the data source(s),
- some additional comments.

The major part of these global attributes can be specified by the user before running the analysis.

In the event the user wishes to cut the spatial domain of the final netCDF, a script has been implemented in order to avoid reperforming the whole analysis. The user has simply to provide a contour line specifying the new smaller domain. For a quick visualization of the results one can use gnuplot, a command-line utility designed for the visualisation of many different types of data. The command-line implementation makes it perfectly suitable for a combined use with DIVA. A single script generates a set of figures including the contour, the finite-element mesh, the data locations and the analysed field, depending on the files present in the input and output directories. As some users are more familiar with MATLAB or Python, toolkits for plotting the results in these programming environments are also provided

(http://modb.oce.ulg.ac.be/mediawiki/index.php/DIVA#Plotting_tools).

To have these tools working, the user simply needs a library for reading of netCDF files and a library for the plotting of data on a map (m_map in Matlab, Basemap in Python). The examples are easily configurable and automatised if one wants to produce a large amount of high-quality figures, suitable for publication in a peer-reviewed journal.

VISUALIZATION OF GRIDDED FIELDS

This section discusses the on-line visualization of DIVA analyses. Contrary to most standard geographic information systems, where there is a strong focus on time-invariant information on the Earth surface (thus no depth and time dimension), the visualization of ocean data sets as generated by DIVA poses several challenges.

The ocean is generally a highly stratified fluid, where the seawater properties vary strongly with depth. While the ocean can be seen as a system made up of individual layers, the vertical movements (upwelling due to winds for instance) and mixing couple the different layers. For instance a vertical section of nutrients is often used to characterize and assess the effect of wind-induced upwelling at the coast. The ocean is also non-stationary as it varies on a broad range of temporal scales. Observational networks generally capture a limited range of these temporal scales. Not all temporal

scales are present, but the ocean state is clearly non-stationary. It is necessary to integrate the depth and time dimensions to provide a full 4-dimensional view of the ocean.

A web-based visualization allows the integration with other datasets, which facilitates the interpretation of the dataset in conjunction with results from different domains (for instance physical and biogeochemical variables are strongly related due to the impact of ocean currents). The web interface should allow users to easily find a useful dataset and provide a mechanism to download it (entirely or a subset) and to be interoperable with other visualization platforms.

The web-interface OceanBrowser aims to respond to those challenges and requirements by allowing horizontal and vertical sections. Horizontal sections of the 4-dimensional fields (longitude, latitude, depth and time) can be visualized at a selected depth layer and time slice. The available depth and time slices correspond to the values included in the netCDF file. No additional time and depth interpolation is required for the horizontal sections.

The climatological fields can also be extracted and visualized on arbitrary vertical sections. In this case, the parameters are actually interpolated along a section defined by the user (unlike the horizontal sections). The vertical section can either be drawn by the mouse in the overview map (Figure 8) or generated automatically by using a criteria like a fixed distance from the coast or a fixed ocean depth.

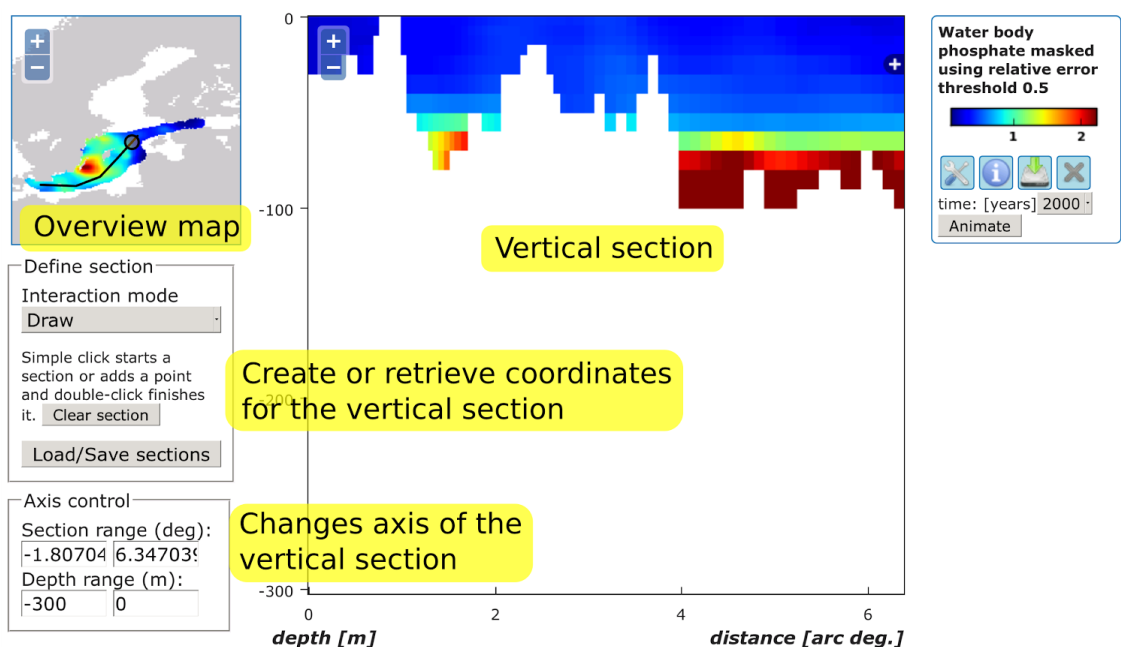


Figure 8: To extract the gridded data along a vertical section, the user draws the section in the overview map with the mouse. The section can also be generated automatically by choosing a fixed distance from the coast or isobath level.

The maps displayed in the browser are created dynamically and therefore several options are made available to the user to customize the graphical rendering of those layers.

Scalar plots

Layers representing scalar variables can be displayed either using flat shading, filled contours or simple contours. For flat shading, the color corresponding to a value of a grid cell is uniformly applied to this grid cell (Figure 9). This is the default option. The gridded field can also be displayed using contour lines (with and without filling the color between the contours). While the option flat shading and filled contour result in an opaque layer, the option of simple contours can be used to display different fields from the same domain at the same time and to assess to correspondence between two fields. As the ocean currents and (to a lesser degree) river inputs affect several variables in a similar way, gradients in various fields often align. The similarity of the structures in e.g. salinity and nutrients can for example be easily assessed by combining these variables in a single plot. Several options controlling the color-map are also available.

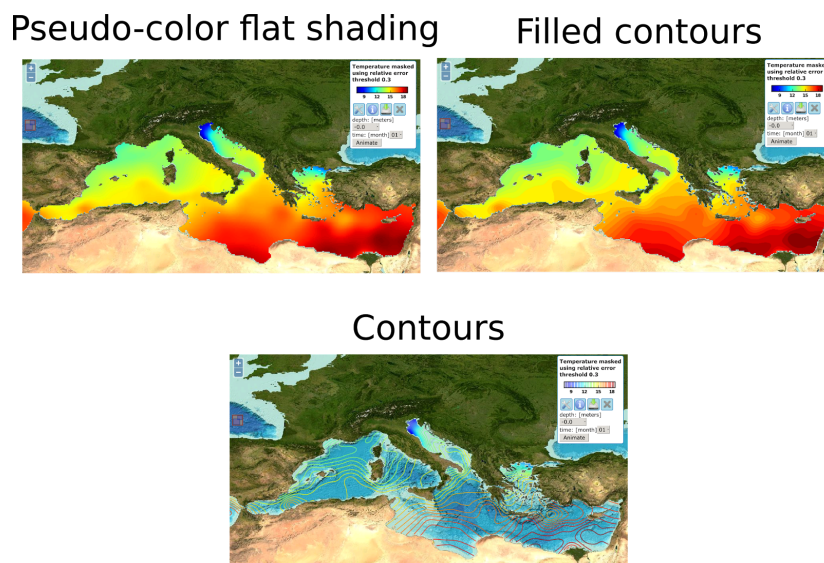


Figure 9: The visualization of layer can be customized by using flat shading, filled contours or simple contours

Vector plots

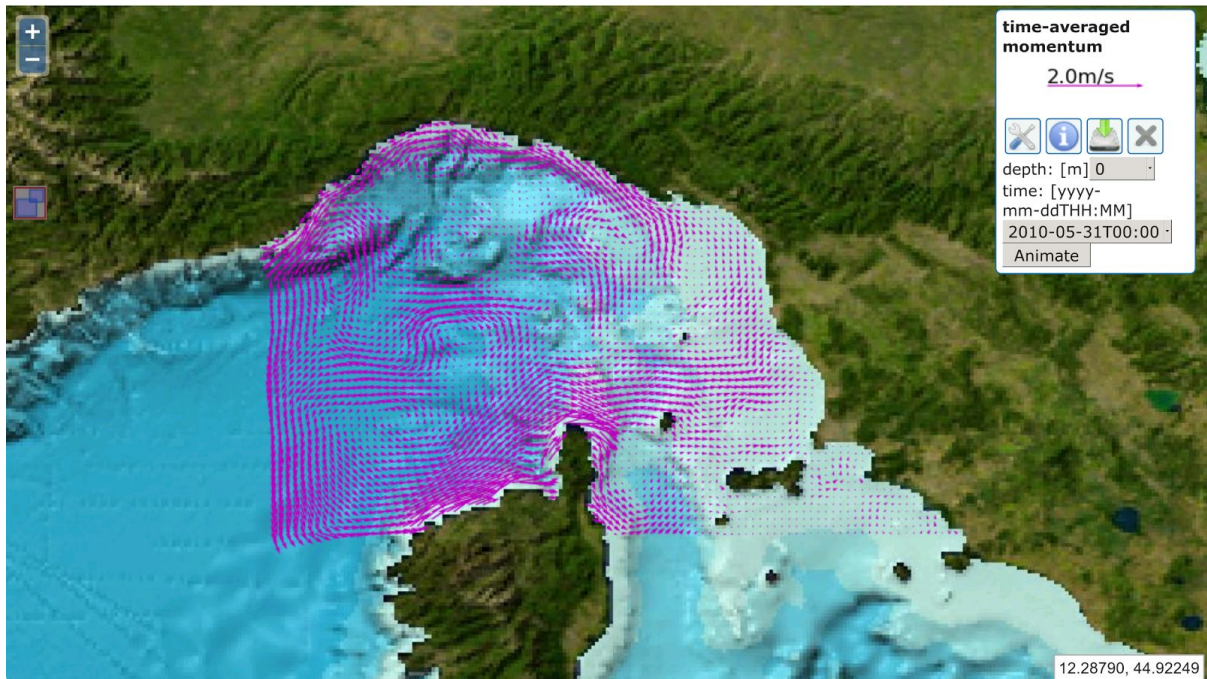


Figure 10: Surface currents generated from ROMS visualized in OceanBrowser

Most data files represent scalar fields, but it is also possible to import vector fields into OceanBrowser. The layer can be customized by changing the density of the arrows and by adapting the scaling of the vectors. As an illustration Figure 10 shows the surface currents in the Ligurian Sea derived from the ROMS model (Shchepetkin and McWilliams, 2003, 2005). The spatial distribution of several ocean parameters is often related to ocean currents, which is indeed the original motivation to include the advection constraint in the DIVA analysis. A combined view of the ocean currents and the derived analysis thus allows one to visually assess how the information from the observations is spread horizontally.

Location markers

The primary use of these markers is to show the observation locations used in DIVA. Within the SeaDataNet and EMODNET Chemistry projects, every observation has an identifier which is composed by an institution EDMO code (European Directory of Marine Organisations) and a local identifier. The combination of the EDMO code and local identifier guarantees to uniquely identify a set of observations (usually a vertical profile). Such a layer gives access to the underlying data by linking it to the data download service (Figure 11).

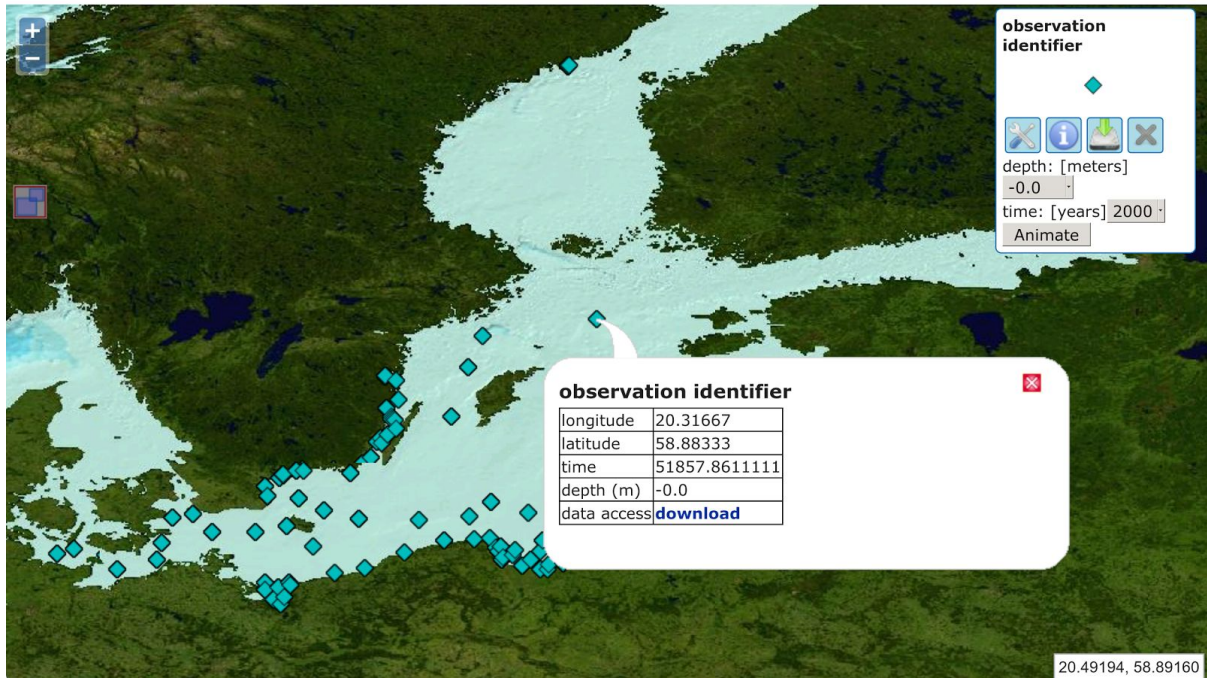


Figure 11: Location marker of the observations used in DIVA for a given depth and time slice

The horizontal and vertical sections can be animated in order to examine the temporal evolution. The web interface can also import third-party layers by using standard Web Map Service (WMS) requests. The user needs only to specify the URL of the WMS server and its supported version.

Interoperability of OceanBrowser with other web-services and clients depends critically on the use of open standards. Beside the W3C web-standards, OceanBrowser uses OGC WMS and netCDF. The implementation of the WMS protocol for the DIVA data products supports the following requests:

- GetCapabilities

This request is used to provide all layers of the map server. To every parameter and to every region corresponds a different WMS layer. An example of such a request would be:

<http://ec.oceanbrowser.net/emodnet/Python/web/wms?request=GetCapabilities&service=WMS&version=1.3.0>

- GetMap

This request allows the extraction of a horizontal section of the 4-dimensional netCDF file at the specified depth and time for a selected latitude/longitude bounding box. By default, the axis are not displayed on a map. This can be activated by setting the non-standard parameter DECORATED to true.

The GetMap can also be used to extract a vertical section. The path of the section is encoded in the SECTION parameter: the longitude and latitude are separated by a comma and the coordinates by the

pipe-symbol (|). The x-axis corresponds to the distance in arc degrees along the section (the first point is the origin) and the y-axis is the depth in meters. The parameter `RATIO` defines the aspect ratio of the vertical section.

Images can be returned in raster (PNG) and vector image formats (SVG, EPS and PDF). They can also be saved as a KML file so that the current layer can be visualized in programs like Google Earth and combined with other information imported in such programs. By providing multiple time instances, the web map server can also return animations in the WebM or MP4 format using this GetMap request. As the animations are generated dynamically, it usually takes a couple of minutes to create them. The frame rate of the animation is controlled through the parameter `RATE`.

- `GetFeatureInfo`

This request returns a simple XML file with the underlying value of the analysed field. This information is usually requested when the user clicks on a map.

However, the WMS standards (in version 1.1.1 and 1.3.0) are not completely adequate for ocean analyses. A WMS allows one to represent a data set according to a list of different styles. A legend is attributed to each style which is a colorbar for scalars. The legend for a given style is represented by a link to an image.

A single legend or colorbar is used for entire data set (for all depth layers and time instances in particular). However, the ocean is strongly stratified and a unique legend does not provide sufficient contrast because the ocean properties at depth are often very different from the properties near the surface. The solution is to make the legend dynamic so that it can be adjusted based on a range of values at a specified depth and time. Instead of using a limited number of styles, the style parameter in OceanBrowser uses a list for key-values pairs. It can contain the information as detailed in Table 1. Key and values are separated by a colon and key-value pairs by a plus sign. The order of the key/values is not significant.

Other WMS servers used for ocean data have developed similar extensions (Blower et al., 2013). For instance the parameters `vmin` and `vmax` for the colorbar are called `COLORRANGE` in ncWMS. Support of this parameter syntax is also implemented in OceanBrowser.

Table 1: Parameters to control the plotting style of WMS layers

Key	Value
<code>method</code>	plotting method (pseudo-color, contours or filled contour)
<code>ncontours</code>	number of contours (ignored for pcolor)
<code>inverted</code>	true if color-map is inverted

cmap	color-map name (http://www.scipy.org/Cookbook/Matplotlib/Show_colormaps)
vmin	minimum of color-map range
vmax	maximum of color-map range

Deployment and configuration of OceanBrowser

The basic idea was to make the incorporation of data in OceanBrowser as easy as adding a static web page to a web server. Similarly to most web servers, OceanBrowser uses also a “document root” directory under which all netCDF files can be placed, possibly using subdirectories in order to structure the data fields according to e.g. domain, parameters or time range. Internally OceanBrowser holds an inventory of all data fields and the inventory is regenerated when the modification time of the root directory changes (which can be forced using the Linux `touch` command). For adding a new file to OceanBrowser it is thus not necessary to restart the server application.

Every netCDF file is treated as a virtual folder containing various netCDF variables. If the number of netCDF variables is high, the individual variables can also be grouped in virtual subfolders. This is extensively used for DIVA analysis as several auxiliary netCDF variables are placed in a virtual folder called “Additional fields” which includes for instance the expected error standard deviation of the analysis. The underlying 4-dimensional netCDF file can be either downloaded as a whole from the interface or only as a subset using the OPeNDAP protocol.

While OceanBrowser has been primarily developed to visualize gridded datasets from DIVA it can also be used for satellite datasets which can have a more complex horizontal grid, as shown in Figure 12 which can have a more complex horizontal coordinate as it is sometimes related to the trajectory of the satellite.

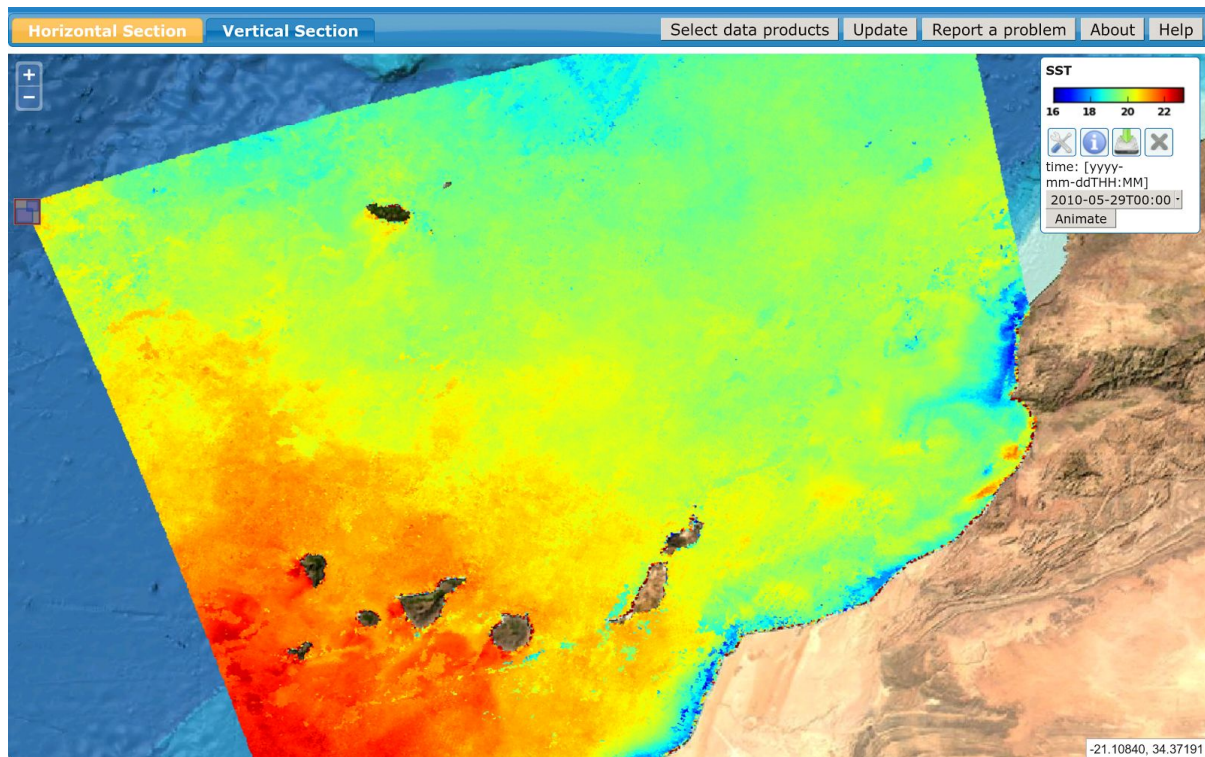


Figure 12: L3 Satellite Sea Surface Temperature around the Canary Islands analysed by DINEOF using Medspiration Level 3 data.

OceanBrowser is composed of a client part running in the web browser written in JavaScript and a server part written in Python. It uses several Python modules and in particular matplotlib and its Basemap plugin for the visualization. The Python server runs on a Linux server with at least 8GB of memory. In total, OceanBrowser requires 13 software packages to be installed on the server. To simplify the deployment, OceanBrowser can be installed using the programme docker. OceanBrowser and all its software dependencies are made available in a software container on the web site called “docker hub” (<https://hub.docker.com/r/abarth/oceanbrowser/>). Installing and running OceanBrowser can be achieved by running the single Linux command (provided that docker is installed):

```
docker run -p 8080:80 --name my-oceanbrowser-container \
-v /some/netcdf/files:/var/www/data:ro abarth/oceanbrowser
```

Docker automatically downloads and runs OceanBrowser. OceanBrowser will be available at port 8080 on localhost and servers files in the specified directory. Additional configuration can be done using environment variables (including public URL, name and contact information).

To improve performance, OceanBrowser implements cache control headers for GetCapabilities and GetMap requests. The use of these cache control headers has a significant impact, improving the responsiveness of the web server. HTTP cache control gives a fine grain control about reusing server responses while minimizing the risk to using an out-of-date content.

For example if the client asks for the first time a list of all available layers (GetCapabilities), the server includes the following HTTP headers:

```
Cache-Control: "must-revalidate, max-age=5"  
Last-Modified: "Tue, 10 Apr 2015 12:30:18 GMT"
```

This tells the web browser that it must check with the server if a newer version exists (after 5 seconds) and the time-stamp in the “Last-Modified” header reflects the time when the resource was last modified. Any future web requests of the web browser of the same resource are conditional request and include the following header:

```
If-Modified-Since:"Tue, 10 Apr 2015 12:30:18 GMT"
```

The server can either inform the web browser that its cached version is the current version (cache revalidation) or responding with a latest version of the corresponding request (and including the timestamp of the latest version in the Last-Modified header). On OceanBrowser, potential cache revalidation is fast because it requires only to compare the timestamp provided by the web browser and the timestamp of the directory containing all netCDF files.

The caching system has been extended by using in addition server side caching (Figure 13). A caching server (nginx) is placed between the web-browser and OceanBrowser. The web-browser first contacts the caching server which then contacts OceanBrowser. OceanBrowser needs only to generate the response for a given request, if the cached version at the browser and the caching server is outdated (or non-existent). In this case, the new version will be cached at the client (web-browser) and the caching server. Improving the caching system has resulted in a significant improvement of the responsiveness and the download speed to display the list of all layers and map tiles.

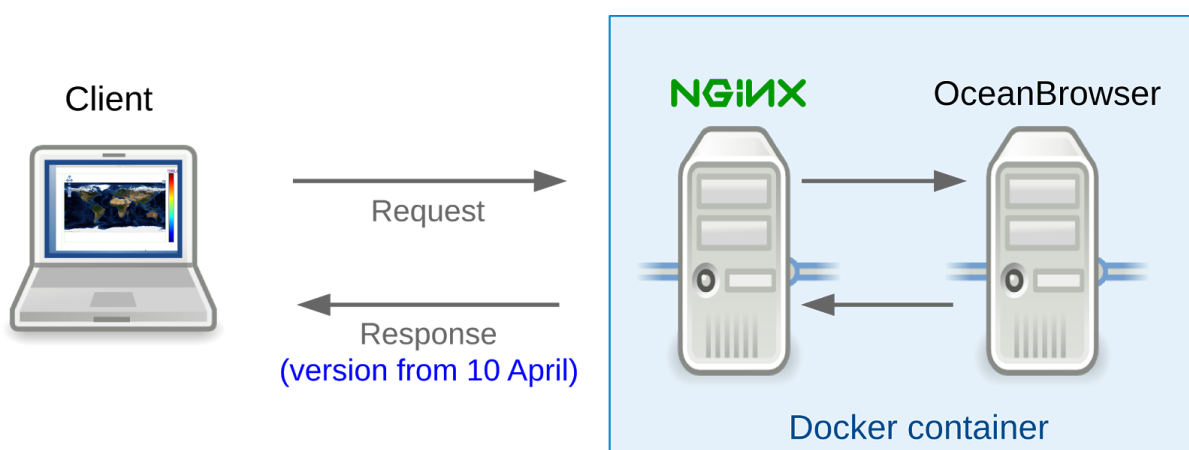


Figure 13: To enable server-side caching the nginx server is placed between the client browser and the server software OceanBrowser.

VISUALIZATION OF IN SITU OBSERVATIONS

Third-party Web Map Services can be included in OceanBrowser. This is particularly useful to view a climatology in combination with the corresponding in situ observations. Special consideration was given to the integration of the Oceanotron web service developed by Ifremer which provides quality-controlled temperature and salinity measurements of the Arctic Ocean, North Atlantic Ocean, Baltic Sea, Black Sea, Mediterranean Sea and North Sea. Observations within a specified depth and time range can be queried and displayed on a map. Such observations can be visualized in conjunction with the DIVA analysis (Figure 14). This is useful to validate the climatology in a first step, to identify outliers in the observations, but also to assess anomalous conditions where some observations are different from average conditions.

Oceanotron is based on ncWMS and to realize this integration some of the extensions of the WMS protocol used by ncWMS have been implemented in OceanBrowser in order to allow a customization of the legend style (in particular the colorbar mapping and colorbar range). The way that the depth and the time dimensions are handled for serving in situ observations is also different from the way that those dimensions are represented for gridded climatologies. The depth and time are discrete values for the gridded climatologies, while they must be ranges for in situ observations.

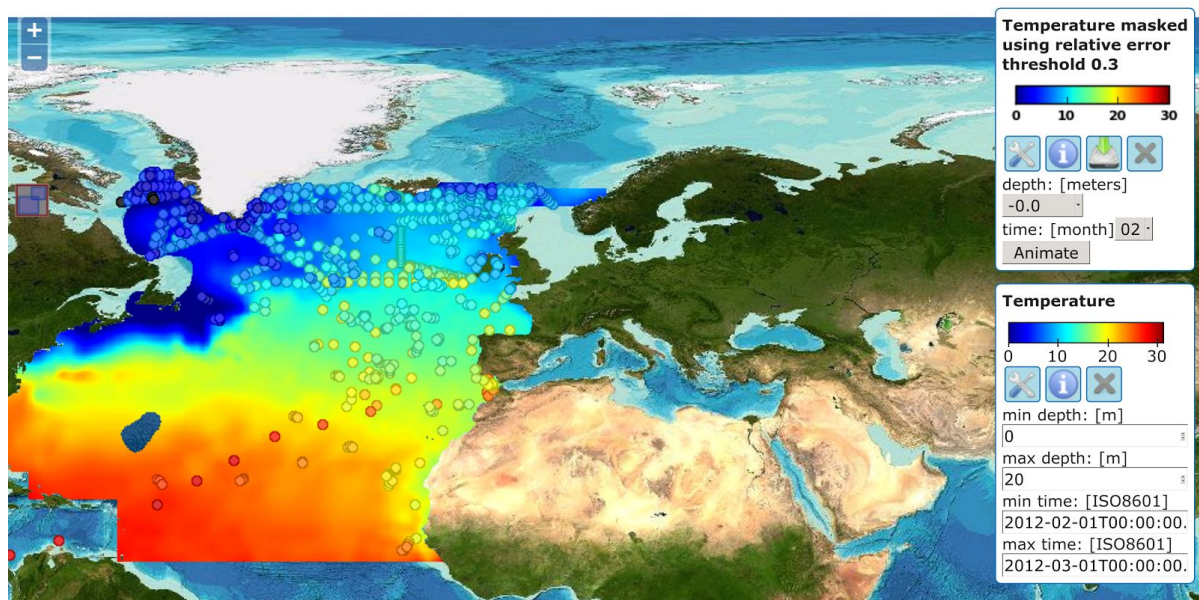


Figure 14: Climatological temperature of the month February and in situ observation of February 2012

CONCLUSION

DIVA allows to aggregate in situ observations and to derive a coherent spatial view of ocean observations. It is generally used to derive monthly, seasonal or yearly climatological mean estimates. The accuracy of the analysis can be assessed by taking into consideration the error field which depends essentially on the data distribution, the accuracy of the underlying observations and the correlation length-scale. Originally DIVA was developed to analyse physical parameters like temperature and salinity, but over the time special adaptations (like the anamorphosis transform) have been implemented to make DIVA also suitable for chemical and biological parameters.

Those data fields are generally four-dimensional (longitude, latitude, depth and time) and a web interface was developed to integrate the depth and time dimensions. OceanBrowser allows the visualization of the climatologies either along horizontal sections or arbitrary vertical sections. The time evolution can also be assessed by producing an animation.

DIVA and OceanBrowser are open source and released under the terms of the GNU General Public License and Affero General Public License to promote its use and adoption. Easy deployment of OceanBrowser is possible using Linux containers which allows to bundle all software dependencies and to ease the installation process.

Currently, DIVA does not take into account the possible relationship between variables. In future versions of DIVA we plan to implement a multivariate version. Such analyses can take the advantage of the relationship between e.g. physical parameters and biological parameters as the spatial coverage of biological parameters is much lower than the coverage of physical parameters. The information of the physical parameters would then be used to improve the analysis of the biological parameter. A multivariate extension of DIVA can also be used to make analysis of surface currents as the individual velocity components are strongly coupled through the model dynamics.

ACKNOWLEDGEMENTS

The development of DIVA and OceanBrowser was possible thanks to funding from the FP7 EU projects SeaDataNet 2 (contract number 283607) and ODIP (contract number 312492), EMODnet Chemistry (contract number SI2.656742) and EMODnet Biology (contract number SI2.657705) from DG MARE. The F.R.S.-FNRS is acknowledged to fund the position of Alexander Barth. We also thank Ifremer for developing Oceanotron (in particular Thomas Loubrieu and Mickael Treguer) and for DIVA products of the Atlantic Ocean (in particular Aurélie Briand). GEBCO project is acknowledged for providing the GEBCO world map, <http://www.gebco.net>. This is a MARE publication.

REFERENCES

- Alvera-Azcárate, A., Barth, A., Sirjacobs, D., & Beckers, J.-M. (2009). Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF, *Ocean Science*, 5, 475–485, doi:10.5194/os-5-475-2009.
- Barnes, S. (1964). A technique for maximizing details in numerical map analysis, *Journal of Applied Meteorology*, 3, 395–409, doi:10.1175/1520-0450(1964)003<0396:ATFMDI>2.0.CO;2.
- Barth, A. (2015, September 11). *OceanBrowser*, Retrieved from <https://hub.docker.com/r/abarth/oceanbrowser/>.
- Barth, A., Beckers, J.-M., Troupin, C., Alvera-Azcárate, A., & Vandenbulcke, L. (2014). divand-1.0: n-dimensional variational data analysis for ocean observations, *Geoscientific Model Development*, 7(1), 225–241, doi:10.5194/gmd-7-225-2014.
- Beckers, J.-M., & Rixen, M. (2003). EOF calculation and data filling from incomplete oceanographic datasets, *Journal of Atmospheric and Oceanic Technology*, 20, 1839–1856.
- Beckers, J.-M., Barth, A., & Alvera-Azcárate, A. (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.
- Beckers, J.-M., Barth, A., Troupin, C., & Alvera-Azcárate, A. (2014). Approximate and efficient methods to assess error fields in spatial gridding with data interpolating variational analysis (DIVA), *Journal of Atmospheric and Oceanic Technology*, 31, 515–530, doi:10.1175/JTECH-D-13-00130.1.
- Bennett, A. F., Chua, B. S., & Leslie, L. M. (1996). Generalized inversion of a global numerical weather prediction model, *Meteorology and Atmospheric Physics*, 60, 165–178, doi:10.1007/BF01029793.
- Blower, J., Haines, K., Santokhee, A., & Liu, C. (2009). GODIVA2: interactive visualization of environmental data on the Web, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 367(1890), 1035–1039, doi:10.1098/rsta.2008.0180.
- Blower, J., Gemmell, A., Griffiths, G., Haines, K., Santokhee, A., & Yang, X. (2013). A Web Map Service implementation for the visualization of multidimensional gridded environmental data, *Environmental Modelling & Software*, 47, 218 – 224, doi:10.1016/j.envsoft.2013.04.002.
- Brasseur, P., & Haus, J. (1991). Application of a 3–D variational inverse model to the analysis of ecohydrodynamic data in the Northern Bering and Southern Chuckchi seas, *Journal of Marine Systems*, 1, 383–401, doi:10.1016/0924-7963(91)90006-G.

- Brasseur, P., Beckers, J.-M., Brankart, J.-M., & Schoenauen, R. (1996). Seasonal temperature and salinity fields in the Mediterranean Sea: Climatological analyses of an historical data set, *Deep-Sea Research*, 43(2), 159–192, doi:10.1016/0967-0637(96)00012-X.
- Bretherton, F. P., Davis, R. E., & Fandry, C. B. (1976). A technique for objective analysis and design of oceanographic experiment applied to MODE-73, *Deep-Sea Research*, 23, 559–582, doi:10.1016/0011-7471(76)90001-2.
- Buongiorno Nardelli, B. (2012). A novel approach for the high-resolution interpolation of in situ sea surface salinity, *Journal of Atmospheric and Oceanic Technology*, 29, 867–879, doi:10.1175/JTECH-D-11-00099.1.
- Buongiorno Nardelli, B., Colella, S., Santoleri, R., Guarracino, M., & Kholod, A. (2010). A re-analysis of black sea surface temperature, *Journal of Marine Systems*, 79(1–2), 50–64, doi:10.1016/j.jmarsys.2009.07.001.
- Capet, A., Barth, A., Beckers, J.-M., & Grégoire, M. (2012). Interannual variability of black sea's hydrodynamics and connection to atmospheric patterns, *Deep Sea Research Part II: Topical Studies in Oceanography*, 77, 128–142, doi:10.1016/j.dsr2.2012.04.010.
- Capet, A., Troupin, C., Carstensen, J., Grégoire, M., & Beckers, J.-M. (2014). Untangling spatial and temporal trends in the variability of the Black Sea Cold Intermediate Layer and mixed Layer Depth using the DIVA detrending procedure, *Ocean Dynamics*, 64(3), 315–324, doi:10.1007/s10236-013-0683-4.
- Chen, Y., Davis, T. A., Hager, W. W., & Rajamanickam, S. (2008). Algorithm 887: Cholmod, supernodal sparse cholesky factorization and update/downdate, *ACM Transactions on Mathematical Software*, 35(3), 22:1–22:14, doi:10.1145/1391989.1391995.
- Cressman, G. (1959). An operational objective analysis system, *Monthly Weather Review*, 88, 327–342, doi:10.1175/1520-0493(1959)087<0367:AOOAS>2.0.CO;2.
- Davis, T. A., & Hager, W. W. (2009). Dynamic supernodes in sparse Cholesky update/downdate and triangular solves, *ACM Transactions on Mathematical Software*, 35(4), 27:1–27:23, doi:10.1145/1462173.1462176.
- de la Beaujardiere, J. (Ed.) (2006). OpenGIS® Web Map Server Implementation Specification Version 1.3.0, Open Geospatial Consortium Inc.
- Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in sobolev spaces, in *Constructive theory of functions of several variables*, pp. 85–100, Springer, doi:10.1007/BFb0086566.

- Evensen, G. (2007). *Data assimilation: the Ensemble Kalman Filter*, Springer, 279pp.
- Gandin, L. S. (1965). Objective analysis of meteorological fields, Israel Program for Scientific Translation, Jerusalem, 242 pp.
- Guinehut, S., Larnicol, G., & Traon, P. L. (2002). Design of an array of profiling floats in the north atlantic from model simulations, *Journal of Marine Systems*, 35(1-2), 1–9, doi:10.1016/S0924-7963(02)00042-8.
- Hunter, J. (2007). Matplotlib: A 2d graphics environment, *Computing in Science Engineering*, 9(3), 90–95, doi:10.1109/MCSE.2007.55.
- Hyer, J. L., & She, J. (2007). Optimal interpolation of sea surface temperature for the North Sea and Baltic Sea, *Journal of Marine Systems*, 65(1-4), 176–189, doi:10.1016/j.jmarsys.2005.03.008.
- Kaplan, A., Kushnir, Y., Cane, M., & Blumenthal, M. (1997). Reduced Space Optimal Analysis for Historical Datasets: 136 Years of Atlantic Sea Surface Temperatures, *Journal of Geophysical Research*, 102(C3), 27,835–27,860, doi:10.1029/97JC01734.
- Kaplan, D. M., & Lekien, F. (2007). Spatial interpolation and filtering of surface current data based on open-boundary modal analysis, *Journal of Geophysical Research: Oceans*, 112(C12), doi:10.1029/2006JC003984.
- Lipphardt, B. L., Kirwan, A. D., Grosch, C. E., Lewis, J. K., & Paduan, J. D. (2000). Blending HF radar and model velocities in Monterey Bay through normal mode analysis, *Journal of Geophysical Research: Oceans*, 105(C2), 3425–3450, doi:10.1029/1999JC900295.
- Lorenzo, E. D., Moore, A. M., Arango, H. G., Cornuelle, B. D., Miller, A. J., ... Bennett, A. F. (2007). Weak and strong constraint data assimilation in the inverse Regional Ocean Modeling System (ROMS): Development and application for a baroclinic coastal upwelling system, *Ocean Modelling*, 16(3-4), 160–187, doi:10.1016/j.ocemod.2006.08.002.
- McIntosh, P. C. (1990). Oceanographic data interpolation: Objective analysis and splines, *Journal of Geophysical Research*, 95(C8), 13,529–13,541, doi:10.1029/JC095iC08p13529.
- Moore, A. M., Arango, H. G., Broquet, G., Powell, B. S., Weaver, A. T., & Zavala-Garay, J. (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.
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., ... Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *Journal of Geophysical Research*, 108(D14), 4407, doi:10.1029/2002JD002670.

Roberts-Jones, J., Fiedler, E. K., & Martin, M. J. (2012). Daily, Global, High-Resolution SST and Sea Ice Reanalysis for 1985-2007 Using the OSTIA System, *Journal of Climate*, 25, 6215–6232, doi:10.1175/JCLI-D-11-00648.1.

Robinson, A. R. (1996). Physical processes, field estimation and an approach to interdisciplinary ocean modeling, *Earth-Science Reviews*, 40, 3–54, doi:10.1016/0012-8252(95)00030-5.

Schlitzer, R. (2002). Interactive analysis and visualization of geoscience data with Ocean Data View, *Computers & Geosciences*, 28(10), 1211 – 1218, doi:10.1016/S0098-3004(02)00040-7.

Shchepetkin, A., & McWilliams, J. (2003). A Method for Computing Horizontal Pressure-Gradient Force in an Oceanic Model with a Non-Aligned Vertical Coordinate, *Journal of Geophysical Research*, 108, 1–34, doi:10.1029/2001JC001047.

Shchepetkin, A., & McWilliams, J. (2005). The Regional Oceanic Modeling System: A split-explicit, free-surface, topography-following-coordinate ocean model, *Ocean Modelling*, 9, 347–404, doi:10.1016/j.ocemod.2004.08.002.

Troupin, C., Barth, A., Sirjacobs, D., Ouberdous, M., Brankart, J.-M., Brasseur, P., ... Beckers, J.-M. (2012). Generation of analysis and consistent error fields using the Data Interpolating Variational Analysis (DIVA), *Ocean Modelling*, 52–53, 90–101, doi:10.1016/j.ocemod.2012.05.002.

Wahba, G. (1990). *Spline models for observational data*, CBM-NSF Regional Conference Series in Applied Mathematics, vol. 59, SIAM Journal of Applied Mathematics, Philadelphia, PA, 169 pp.

Wahba, G., & Wendelberger, J. (1980). Some new mathematical methods for variational objective analysis using splines and cross validation., *Monthly Weather Review*, 108, 1122–1143, doi:10.1175/1520-0493(1980)108<1122:SNMMFV>2.0.CO;2.

Weaver, A. T., Vialard, J., & Anderson, D. L. T. (2003). Three- and Four-Dimensional Variational Assimilation with a General Circulation Model of the Tropical Pacific Ocean. Part I: Formulation, Internal Diagnostics, and Consistency Checks, *Monthly Weather Review*, 131, 1360–1378, doi:10.1175/1520-0493(2003)131<1360:TAFVAW>2.0.CO;2.

KEY TERMS AND DEFINITIONS

Variational Inverse Method (VIM): data analysis method based on the formulation of a cost-function to be minimized. While interpolating a gridded field on arbitrary location is generally

considered the “direct” or “forward” problem, the inverse problem tries to deduce from observation at arbitrary location a gridded field.

Error covariance: measure how much the error (of the first guess estimate or of the observation) is related to the error at a different location (or time). The error covariance can be decomposed in variance and decorrelation.

Correlation length: the length-scale over which the error covariance is significantly different from zero.

Signal-to-noise ratio: a measure of the variance of the background error compared to the variance of the observation errors.

Web Map Service (WMS): the OGC server-client protocol to advertise all available layers and to transfer maps, usually as images.

Web Feature Service (WFS): the OGC server-client protocol for discrete features.

netCDF (Network Common Data Form): a binary data format for multidimensional data sets and widely adopted in oceanography and meteorology.

