A web interface for gridding arbitrarily distributed in situ data based on Data-Interpolating Variational Analysis (DIVA)

A. Barth\textsuperscript{1,2}, A. Alvera-Azcárate\textsuperscript{1,2}, C. Troupin\textsuperscript{1}, M. Ouberdous\textsuperscript{1}, and J.-M. Beckers\textsuperscript{1,2}

\textsuperscript{1}GeoHydrodynamics and Environment Research (GHER), MARE, AGO, University of Liège, Liège, Belgium
\textsuperscript{2}National Fund for Scientific Research, Belgium

Received: 15 December 2009 – Revised: 15 May 2010 – Accepted: 1 June 2010 – Published: 27 September 2010

Abstract. Spatial interpolation of observations on a regular grid is a common task in many oceanographic disciplines (and geosciences in general). It is often used to create climatological maps for physical, biological or chemical parameters representing e.g. monthly or seasonally averaged fields. Since instantaneous observations can not be directly related to a field representing an average, simple spatial interpolation of observations is in general not acceptable. DIVA (Data-Interpolating Variational Analysis) is an analysis tool which takes the error in the observations and the typical spatial scale of the underlying field into account. Barriers due to the coastline and the topography in general and also currents estimates (if available) are used to propagate the information of a given observation spatially.

DIVA is a command-line driven application written in Fortran and Shell Scripts. To make DIVA easier to use, a web interface has been developed (http://gher-diva.phys.ulg.ac.be). Installation and compilation of DIVA is therefore not required. The user can directly upload the data in ASCII format and enter several parameters for the analysis. The analyzed field, location of the observations, and the error mask are presented as different layers using the Web Map Service protocol. They are visualized in the browser using the Javascript library OpenLayers allowing the user to interact with layers (for example zooming and panning). Finally, the results can be downloaded as a NetCDF file, Matlab/Octave file and Keyhole Markup Language (KML) file for visualization in applications such as Google Earth.

1 Introduction

Most oceanographic applications (and geoscience problems in general) deal with large spatial scales which are generally under-sampled by current in situ observation networks. To make optimal use of the limited amount of in situ observations, sophisticated analysis tools are necessary.

Several methods are commonly used to interpolate in situ observations to obtain gridded fields. Examples are optimal interpolation (Gandin, 1965; Bretherton et al., 1976), kriging (e.g. Chilès and Delfiner, 1999) or variational analysis (e.g. Brasseur et al., 1996; Yaremchuk and Sentchev, 2009). Those methods can create continuous fields based on observations at discrete locations (possibly contaminated with errors) and some a priori statistical information on the field. Tools have been developed that these interpolation schemes. Examples are the optimal interpolation package from the Harvard Ocean Prediction System (HOPS; Robinson, 1996) and DIVA (Data Interpolating Variational Analysis; Brasseur et al., 1996), based on variational analysis. Those tools have been applied successfully in various oceans and at different scales but they are difficult to use for less technical users, as they usually require programming skills.

Building a web-application around a gridding program can decrease the complexity and technical skills required to use those analysis tools. Web-based programs are in fact becoming increasingly popular in geosciences and are used for various applications such as hydrological forecast verification (Kruger et al., 2007), the seismic model E3D (You et al., 2008), atmospheric ray-tracing (Ghoddousi-Fard et al., 2009) and landform simulation (Luo et al., 2006). Integrated efforts to give access to databases and analysis tools through on-line portals are also increasingly demanded (e.g. CHRONOS portal for Earth history data and tools, Fils et al. (2009)). It has also been shown that web-based applications can be used very efficiently in teaching (e.g. Dong et al., 2009).
The use of web-applications for geosciences data analysis allows the users to have rapid access to the software without needing to install it on a desktop computer. In addition, updates, improvements and new features of the software are directly available to the users without re-installation. Some web-based applications allow the user also to easily share the results with colleagues, via the Internet. Limitations of these web-based applications are the need of a fast Internet connexion and the possibility that not all web-browsers are supported by the application (because some web-browsers implement web standards only partially). In some cases, not all features of the software are available through the web application, which might limit the analyzes that can be made through it.

In this work we present a web-based application of DIVA called Diva-on-web. It can be applied to any geophysical variable, and one of its most common applications is the generation of climatologies (e.g. Rixen et al., 2005; Troupin et al., 2010), that can be used to study the spatial and temporal variability of a given parameter, or initialize and validate numerical ocean models, for example.

DIVA is used by partners of the European SeaDataNet project to create climatologies using the hydrographic observations. Several European national data centers have adopted this software. The web interface was also created specifically to ease the learning curve of new users to apply DIVA. In the frame of the EMODNET (chemical lot) project it will also be used to interpolate chemical observations. Recently, DIVA has been integrated in Ocean Data View (Schlitzer, 2002) as a plug-in for gridding. Ocean Data View is a program allowing visualization and analysis of various oceanographic data sets.

In Sect. 2 the mathematical foundation of DIVA and its advantages and limitations are presented. Programming aspects are discussed in Sect. 3. The usage of the web interface is then presented in Sect. 4. A conclusion and summary are given in Sect. 5.

2 Mathematical formulation of Diva

The variational inverse method or variational analysis (Brasseur et al., 1996) aims to find a smooth field which is close to a set of observations. A cost function \( J \) is defined which penalizes deviations from both requirements. The optimal field \( \phi \) is defined as the field which minimizes this cost function.

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

(1)

The first term is similar to an RMS error between the field and the observations \( d_j \) (measured at the location \( (x_j, y_j) \)). \( \phi_b \) is a background estimate or a first-guess of the field \( \phi \). Each observation is weighted by a coefficient \( \mu_j \). Those coefficients can be constant if all observations are assumed to be of comparable accuracy. But if for some reason, some observations are expected to be more accurate than others, it can be taken into account by the weights \( \mu_j \). The second term is a measure of non-smoothness and is defined over the domain \( D \) as:

\[
\| \phi \|^2 = \int_D \alpha_2 \nabla \nabla \phi : \nabla \nabla \phi + \alpha_1 \nabla \phi \cdot \nabla \phi + \alpha_0 \phi^2 dD
\]  

(2)

Without loss of generality, we can set \( \alpha_2 \) to 1. The variational inverse method is equivalent to the general formulation of optimal interpolation, but the background covariance is implicitly defined by Eq. (2). The parameter \( \alpha_1 \) is usually set to \( 2/L^2 \) since for this case, an analytical solution has been found for the underlying background covariance functions in an infinite domain (Brasseur et al., 1996). The parameters \( \alpha_0 \) and \( \mu_i \) can be related to the correlation length \( L \) and signal-to-noise ratio \( \sigma_i^2 / \epsilon_i^2 \):

\[
\alpha_0 L^4 = 1
\]  

(3)

\[
\mu_i L^2 = 4\pi \sigma_i^2 / \epsilon_i^2
\]  

(4)

The analytical solution is used to estimate the correlation length and signal-to-noise ratio based on the observations.

In DIVA, Eq. (1) is discretized on a triangular grid using the finite element method. This provides the flexibility to accurately represent a complex coastline and isobath.

The advantages and limitations of the DIVA can be illustrated by trying to reconstructed a field sampled at discrete locations. Panel (a) of Fig. 1 shows an idealized square domain with a barrier (e.g. a peninsula or a dike) of a geophysical parameter, say temperature. The barrier suppresses the exchanges between each side of the barrier and the temperature on each side is indeed quite different. In this example, the field varies smoothly over some length-scale. The field is sampled in the locations shown in panel (b) and the observations are also affected by errors. In regions where a measurement campaign has been carried out, a higher spatial coverage is achieved. However, some large gaps are also present. Obviously some information about the position of the structures and fronts is lost by having observations only at a limited set of locations. One cannot expect therefore to obtain exactly the true field.

Panel (c) shows what would happen if the observations would have been interpolated linearly. The domain is decomposed into triangles where the vertices are the location of the data points based on the Delaunay triangulation. Within each triangle, the value is interpolated linearly. The noise on the observations is in general amplified when higher order schemes, such as cubic interpolation, are used. This method is for example implemented in the function griddata of Matlab which is widely used for numerical analysis of ocean observations. Since the coastline is not taken into account, the
information on one side of the barrier, affects the interpolated fields on the other side.

The analyzed field computed by DIVA is shown in panel (d). Noise in the observations has been (partially) filtered out. It can also be seen that the presence of the land-sea boundary is taken into account and that the observations on one side of the peninsula do not affect the temperature on the other side.

3 Programming aspects

DIVA is built in a modular way, where the different tasks in the analysis process are realized with a set of tools. The analysis process is composed at least of the following tasks:

- extraction of a contour line from a bathymetry
- creation of a finite element mesh
- analysis of the observations and interpolation of the results onto a given set of locations

The mesh generation is based on a simple Delaunay triangulation. The analysis uses a direct solver based on the skyline method (Dhatt and Touzot, 1984). A direct solver allows repeating the analysis steps with different observations (at the same location) at a very low numerical cost since the matrix involved in the analysis has to be factorized only once.

All these steps are implemented as separate Fortran programs which are usually wrapped around shell scripts managing command line arguments and placing files where the diva tools expect them. DIVA requires a Unix-like environment. In particular, it runs on Linux and MS Windows with Cygwin. DIVA is distributed as source code, and binaries for those two platforms are also provided.

Text files are used to specify parameters of the analysis and to activate and deactivate various options. The output of DIVA is a text file or a NetCDF file containing the analyzed field interpolated at some locations determined by the user. Usually the analysis field is interpolated to a regular grid which is easier to manipulate than the field on the original triangular grid.

The command line driver interface of DIVA is similar to other high performance codes such as hydrodynamic ocean models, e.g. GHER (Beckers, 1991) or ROMS (Shchepetkin and McWilliams, 2005), and is thus familiar to those users. Scripts are easily build upon the command line tools allowing batch processing and combining features in a new way. The DIVA source code distribution for instance already contains scripts which perform an analysis on several layers and several moments in time and of various parameters. Command-line driven software are also more easily deployed on high performance computers.

Fig. 1. Reconstruction of a field based on scattered observations using linear interpolation and DIVA.
However, the command line interface is perceived as difficult to learn. Sometimes commercial software with a graphical user interface but simpler gridding algorithms are preferred over freely available software such as DIVA or HOPS using state-of-the art gridding and analysis methods but with a steeper learning curve.

Web browsers offer the possibility of scripting HTTP requests (a technique commonly referred as asynchronous Javascript and XML or AJAX). This allows the development of Web applications having a similar mode of interaction than standard graphical desktop applications. Creating a web interface for DIVA has several advantages specially for new users. They do not have to install DIVA and its software dependencies. Unlike the command-line interface, the web interface gives directly visual feedback of the different steps involved in the analysis. Performing sensitivity tests of the analysis relative to various parameters can thus be easily carried out. Another benefit of the web interface that was explored in the development of the DIVA web application, is the possibility of sharing the analysis. The analysis can be accessed by a direct link which can be, for example, sent by email to colleagues. A web-based viewer of the analysis can also be embedded in a web page. However, it is not possible to analyze a large amount of data and to carry out batch-processing. For such applications, the command-line interface is still the best approach.

The web interface contains server-side and client-side components (Fig. 2). The server components are written in Python and embedded in Apache with mod_python calling the underlying DIVA scripts and Fortran executables. The steps involved in an analysis with DIVA are wrapped as a method of a Python object. The instances of these objects keep track which operations have already been performed. In some conditions, a given step does not have to be repeated if it was executed previously. For example, if the finite element mesh was already created for a given correlation length, it does not have to be recreated if the user requests a new analysis with the same or slightly different correlation length (the threshold here is 10%). The object oriented approach allows to hide those caching mechanism form the calling level. Instances of this object are saved on the disk and are persistent across different user requests.

4 DIVA web interface

This section explains the usage of the web interface from a user point of view and explains the underlying exchanges between the server and the client. The procedure is illustrated with near-surface in situ temperature data (from 5m depth to the surface) of the western Mediterranean. All data in June 1985 and 2005 from the databases WOD05, Coriolis, Hydrobase and MedAtlasII are used in this example (Troupin et al., 2010) to compute a climatological average of the surface temperature in June.

4.1 Observations

Only observations in a 2-D horizontal section can be processed at a time in Diva-on-web. The desired time and depth level must be first extracted from all available observations. The observations can be uploaded as a 3-column text file (corresponding to longitude, latitude and value of the observations respectively). If an optional forth column is present, it represents the relative weight $\mu_i$ that should be given to the observations. The user can specify the decimal and column separator since these separator depend on the program used for text export.

After the text file is uploaded to the server, the observations are loaded by the server to verify its format and to determine the bounding box which includes all observations. This information is returned in XML format to the browser.

4.2 Contour generation

At the next step, the user can choose the coordinates of the domain over which the data will be gridded. The default is the bounding box of the observations. The zonal and meridional resolution of the regular grid on which the finite element solution is interpolated has to be also specified. Since the variational analysis takes the bottom topography into account, the depth of the horizontal sections containing the observations has to be specified as well. A contour line is generated for the given depth using a digital bathymetry.

Currently, the web interface uses a global bathymetry based on GEBCO_08 version 20090202 (Monahan, 2007) which has a resolution of 30”. In addition to the global bathymetry, a coastal bathymetry of North West Corsica from the RACE project with a resolution of 1.8” is also included. To improve performance, 6 global bathymetries at degraded resolution (1’, 2’, 4’, 8’, 16’ and 32”) are pre-computed since analysis for ocean basins have a typical resolution ranging from 1/4” to 1” and do not thus require the depth at the original resolution of the GEBCO bathymetry. Creating contour lines directly from the original bathymetry would lead to a contour with prohibitively large amount of small features.
Fig. 3. Detail of the finite element grid of the Balearic Islands. The finite element grid can be visualized in the browser as a WMS (Web Map Service) layer.

Fig. 4. Empirical covariance based on observations used to estimate correlation length and signal-to-noise ratio.

As a first step, only the bathymetric data with the requested boundary box is loaded. As the bathymetric data is stored in NetCDF format on the server, this sub-setting is easily performed. As a next step, the bathymetry is filtered by a 2-D-convolution with a filtering length-scale equal to the requested resolution of the analysis. Finally, the bathymetry is interpolated onto the user requested domain and resolution. The procedure will be essentially an interpolation if the requested resolution is less than 30" and it will be an averaging if the resolution is larger than 30".

From this re-gridded bathymetry a contour line is computed assuming that the bathymetry varies linearly at scales equal to the resolution of the remapped grid. From the contour a triangular mesh is formed using Delaunay triangulation. The mesh resolution is about 5 times smaller than the correlation length. As an illustration the finite element grid for the western Mediterranean in shown in Fig. 3. Internally, every triangle is divided into 3 sub-triangles. The solution is a cubic polynomial over each sub-triangle. The elements are connected by the constraints that the solution is continuous and derivable. Taking those constrains into account, there are 12 degrees of freedom for every triangle shown in Fig. 3 (Brasseur, 1994).

The contour generation ensures that only spatial scales of topographic features than can be resolved by the user-chosen regular grid are actually included in contour and triangular mesh.

At some places, land lies under the sea levels (e.g. an area behind a dike). If the coastline and isobath are computed on the bathymetry alone, those places are misrepresented as ocean. Therefore the land-sea mask available by GTOPO30 (from the US Geological Survey) is used to raise every grid point on land which is below sea level to one meter above sea level.

4.3 Estimation of correlation length and signal-to-noise ratio

Correlation length and signal-to-noise ratio can be estimated a priori based on the observations (Brasseur et al., 1996). A large number of observation pairs (currently 1000 pairs if sufficiently observations are available, otherwise all possible pairs) are chosen randomly. Those pairs are grouped into bins depending on the distance separating them. The covariance is computed for each bin which provides an empirical
estimation of the covariance as a function of distance. Since
the observation error is assumed uncorrelated in space, it af-
fects only the bin representing the shortest distance. A mod-
ified Bessel (corresponding to the analytical covariance far
from boundaries (Brasseur et al., 1996)) is fitted to the em-
pirical covariance to determine the correlation length and the
variance of the field. From the difference between the fitted
variance and the variance for the shortest distance, the signal-
to-noise ratio can be estimated. In practice, the fit is only
performed over distances shorter than the first zero-crossing
of the covariance.

Figure 4 illustrates this procedure for in situ surface tem-
perature observations of the western Mediterranean Sea in
June. In this case the fitted correlation length scale is 0.99
degrees and the signal-to-noise ratio is 0.64. The empirical
covariances beyond 2.5 arc degrees (where the covariance
is zero) are not used during the fitting. The fitted covari-
ance function approximates the empirical covariance func-
tion within the range of its uncertainty. When a user requests
the parameter estimation by covariance fitting, an image like
Fig. 4 will be shown in the user’s web browser. The user can
therefore easily decide if the fit is meaningful or not.

Various alternative methods have been proposed and used
(e.g. Brankart and Brasseur, 1996, 1998) to estimate those
parameters based on cross-validation. These methods pro-
vide often more robust results but are much more CPU time
intensive since several analysis have to be performed with
different values of those correlation-length and signal-to-
noise ratio. Thus, we decided here to use the simpler and
much faster method to estimate them.
4.4 Analysis

After specifying the correlation length and the signal-to-noise ratio, the user can request the analysis. Since this can be a long process, especially for a fine finite element mesh, this operation can be canceled. The user has then the opportunity to revise some parameters of the analysis.

Along with the analysis, the expected error variance of the analysis can also be computed. DIVA has several methods implemented to estimate the relative error of the analysis: full error calculation, hybrid scheme (Brankart and Brasseur, 1998) and a scheme referred as “poor-man’s” method which consist in performing an analysis with a zero background and all observations set to one but at the same location as the observations. Since the error estimate is here only used to mask the results of the analysis where too few data is present, the latter method is used.

When the analysis is finished, the server notifies the client application and sends the range of values of the analysis. This information is used during the visualization as the default range of the color scale.

An example of the analysis for the western Mediterranean Sea is shown in Fig. 5 with the data and optimized parameters mentioned in Sect. 4.3. Various well-known oceanographic features such as the meandering Algerian Front and parts of the Liguro-Provençal Front can be clearly seen from this analysis.
4.5 Visualization and data access

The result is saved as a NetCDF file and transferred to the WMS server. The WMS server is written in Python and uses Matplotlib and basemap for graphics. The WMS server produces the images requested by the OpenLayers client library.

A threshold of the maximum allowable relative error can be chosen to mask the analysis where its error exceeds this threshold. This mask is overlaid to the analysis and can thus be easily deactivated. Figure 6 shows an overlay of the analysis, location of the observations and error mask. The error mask and the location of observations are important information to the user to decide how well structures seen in the analysis are supported by the observations.

The results of the analysis can be either downloaded as NetCDF file or Matlab file (which can also be loaded in GNU Octave and Python with scipy). The results can also be downloaded in graphical form either as a KML file for applications such as Google Earth (Fig. 8), as rasterized image (in PNG) or as a vector image (SVG, PDF and EPS are currently supported).

4.6 Local applications

For local applications at the scale of e.g. a small bay, the resolution of the global GEBCO bathymetry might be insufficient. For this reason additional local bathymetries can be added to the web interface and selected by the user during the setup of the domain. Currently, a high resolution bathymetry for northwest Corsica is the only local bathymetry included in the web interface. The aim of this local application (Fig. 7) is to serve as an educational tool to visualize observations carried out from the research station Stareso at the Bay of Calvi (Corsica, France). Based on user needs, other high-resolution bathymetries can be added in the future.

5 Conclusions

Diva-on-web is a web interface for gridding ocean data using Data-Interpolating Variational Analysis (DIVA). It is implemented with the purpose to be easy to use, yet using a state-of-the-art analysis scheme. Within the frame of the SeaDataNet project, several European Oceanographic Data Centers have adopted DIVA to use climatological averages based on in situ data. The web interface introduces the user to various parameters of DIVA (the correlation length and the signal-to-noise ratio), covariance fitting and the underlying finite element mesh. Those aspects should ease the learning curve for using DIVA efficiently.

The web interface is designed to perform only an analysis in the two-dimensional horizontal plane. Even if a user plans to make an analysis for several depths and several months, the user has the possibility to try Diva-on-web with a subset of their own data before installing DIVA. It provides thus an simple way to evaluate if DIVA is appropriate for a given problem. The results obtained by DIVA can also be easily compared to the interpolated fields obtained from another gridding algorithm.

The DIVA interface is implemented as a web service and the visualization of the results is based on the Web Map Service protocol. The gridding capability can therefore be easily integrated in 3rd-party web-pages. This can be particularly useful for an oceanographic data center, where available data is presented through a web-interface. By requesting gridded fields from Diva-on-web, an interpolated field can be shown along with the original data.

Acknowledgements. This work was supported by the SeaDataNet and the EMODNET (Chemical Lot) project of the European Union. The National Fund for Scientific Research, Belgium is acknowledged for funding the post-doctoral positions of Alexander Barth and Aida Alvera Azcárate. This is a MARE publication.

Edited by: G. M. R. Manzella and S. Nativi
Reviewed by: two anonymous referees

References


