Diva-4.2.1: presentation of the new features

C. Troupin[1,4,⋆], F. Machín[2], M. Ouberdous[1], M. Rixen[3], D. Sirjacobs[1] and J.-M. Beckers[1]

[1] GHER - MARE, Sart-Tilman B5, University of Liège, Allée du 6-Août 17, 4000 Liège, BELGIUM
[2] Institut de Ciències del Mar (CSIC), Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, SPAIN
[3] NURC - NATO Undersea Research Centre, Viale San Bartolomeo 400, 19126 La Spezia, ITALY
[4] PhD Student, FRS-FRIA, BELGIUM

1 Introduction

Diva is a software designed to create gridded fields from sparse *in situ* data and relies on a finite-element technique to solve a variational principle. We present an overview of the software capabilities as well as the latest modules developed in the frame of the SeaDataNet project:

- Semi-normed analysis;
- Ordinary cross validation;
- Advection constraint;
- Methods for error estimation.

2 Theory

2.1 Variational inverse method

The field φ reconstructed by Diva using N_d data d_j located at (x_j, y_j) is the solution of the variational principle

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

with

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

where α_i and μ are determined from the data themselves, through their *correlation length* L (tool `divafit`) and *signal-to-noise ration* λ (tools `divagcv`, `divacv` and `divagcvrand`).
Resolution of Eq. (1) relies on a highly optimized finite-element technique, which permits computational efficiency independent on the data number and the consideration of real boundaries (coastlines and bottom).

2.2 Advection constraint

The advection constraint aims at modeling the effects of velocity on the reconstructed field. In theory, this constraint is activated by adding a term to the norm (2), leading to

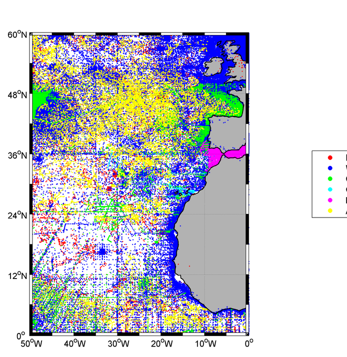
$$\tilde{J} = J(\varphi) + \frac{\theta}{U^2 L^2} \int_{\tilde{D}} \left[\vec{u} \cdot \tilde{\nabla}\varphi - \frac{\mathcal{A}}{L} \tilde{\nabla} \cdot \tilde{\nabla}\varphi\right]^2 \mathrm{d}\tilde{D} \tag{3}$$

where U is a velocity scale deduced from the provided (u, v) field;
L is a characteristic length;
θ is a parameter that controls the weight of the additional term;
$\mathcal{A}$ is a diffusion coefficient.
Parameters θ and $\mathcal{A}$ as well as a velocity field on a regular grid (Fig.5) have to be specified by the user.

3 Implementation

3.1 Data

Data were gathered from various databases in the region $0-60°$N $\times$ $0-50°$ W. We processed them to remove duplicates, detect outliers and perform vertical interpolation with *Weighted Parabolas* method [Reiniger and Ross (1968)]. We consider only surface temperatures during winter (Fig.1).



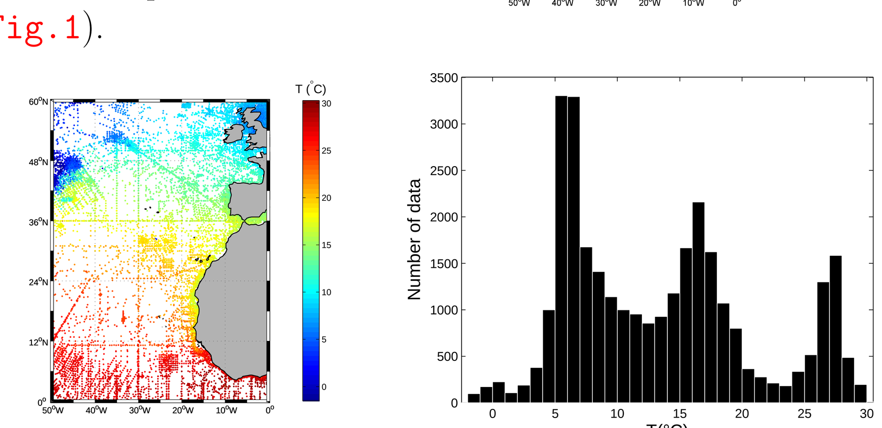
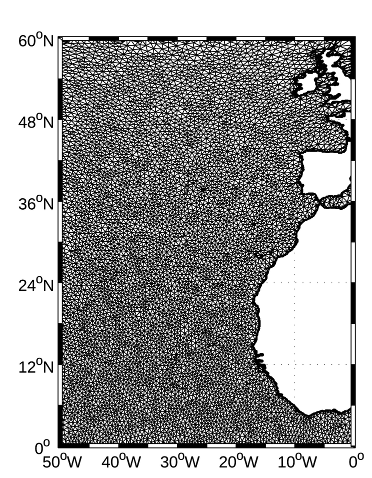


FIGURE 1: *Localisation and histogram of the data.*

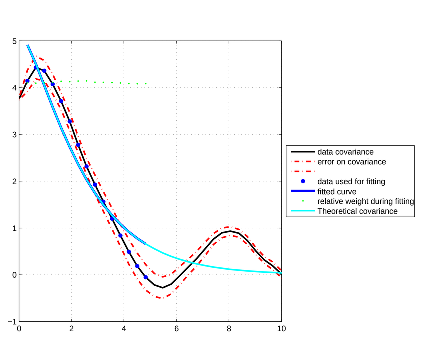
3.2 Contours and mesh

Contours are created from Naval Oceanographic Office 5-min resolution topography on the standard depth levels (tool `divacont`) Mesh is generated with a characteristic length $L = 3$ (command `divamesh`).



3.3 Parameters

Correlation length L is determined through a fit of the data covariance function with the execution of `divafit`, yielding a value of $L = 1.65$, equivalent to $184.1507\,km$. Signal-to-noise ration λ is estimated with one of the following tools:



1. *Ordinary Cross Validation* (CV) applied on the whole dataset (`divacv`) or on a sample containing n values (`divacvrand`).
2. *Generalized Cross Validation* (GCV, tool `divagcv`).

Both methods are designed to provide the value of λ that minimise an estimator Θ (Fig.2). We get of value of $\lambda = 2.971$.

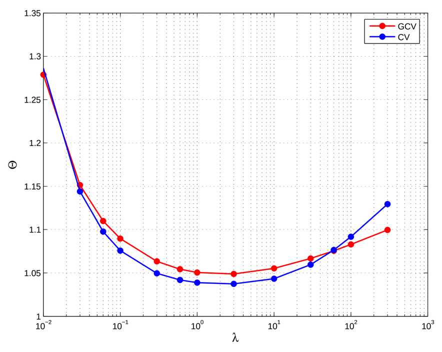


FIGURE 2: *Results of Generalized Cross Validation (red curve) and Cross Validation with a sample of n = 1000 data.*

4 Analysis

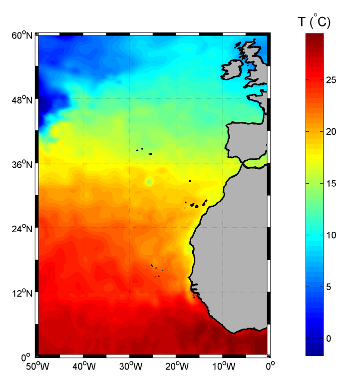


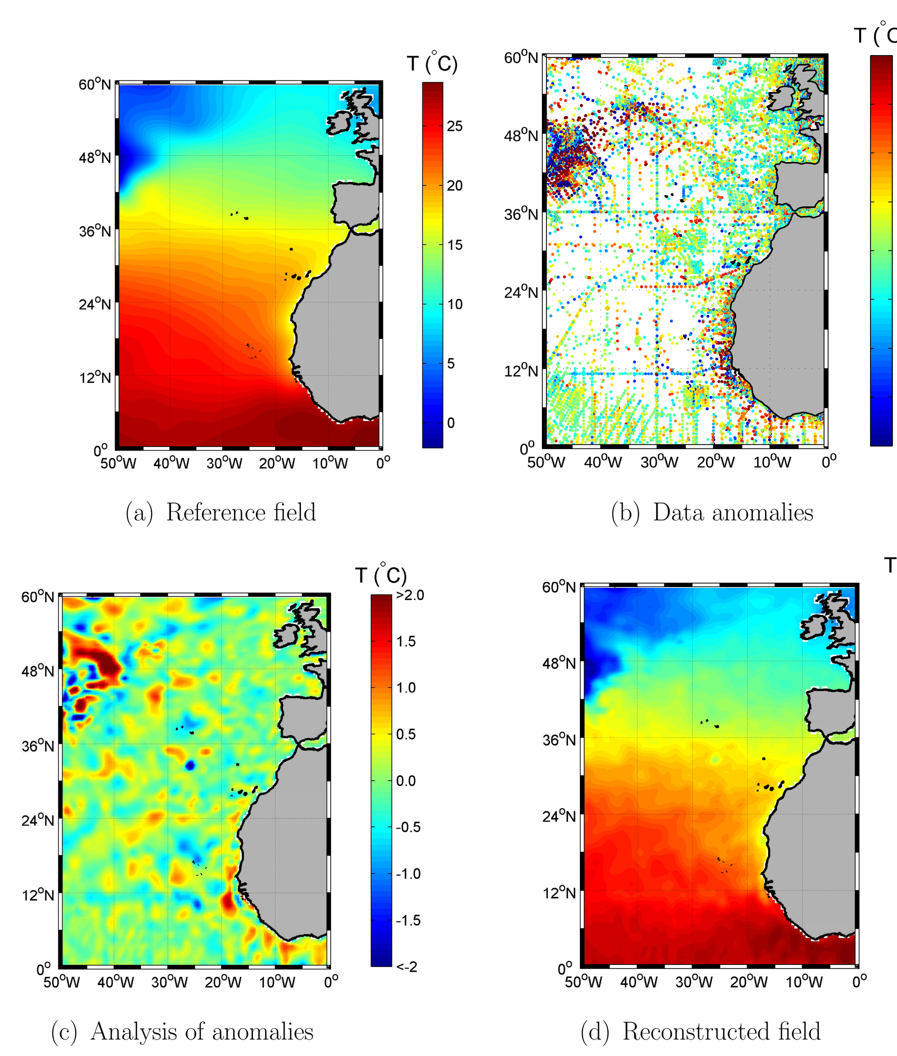
FIGURE 3: *Analysed field obtained with parameters from Sec. 3.3*

4.1 Semi-normed analysis

Semi-normed analysis consists in four steps:

1. `divarefe` → creates a so-called *reference field* (Fig.4(a)), using large correlation length and small signal-to-noise ratio;
2. `divaanom` → subtracts the reference field from the data values in order to work with anomalies (Fig.4(b));
3. `divacalc` → performs an analysis on the anomalies (Fig.4(c));
4. `divasumup` → adding the analysed anomaly field to the background field (Fig.4(d)).

The four steps are regrouped in script `divaseminorm`.

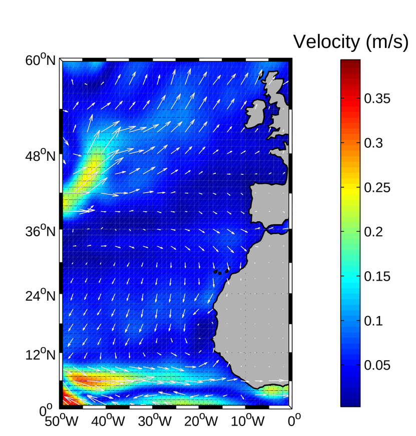


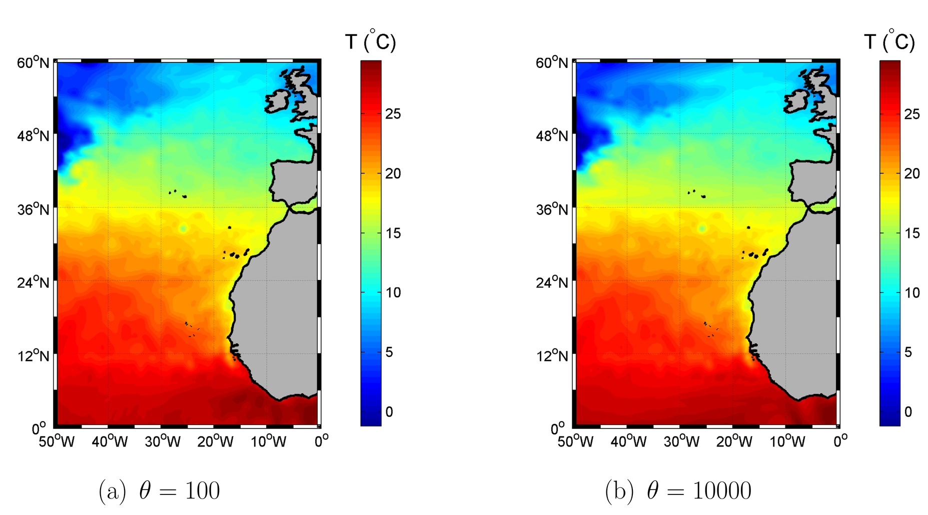
(a) Reference field
(b) Data anomalies
(c) Analysis of anomalies
(d) Reconstructed field

FIGURE 4: *The four steps of a semi-normed analysis.*

4.2 Advection constraint

The mean velocity field used for advection modeling is obtained from drifter measurements. Several values of parameter θ were tried (Fig.5) to observe the effects of weak, moderate or strong advection constraints.





(a) $\theta = 100$
(b) $\theta = 10000$

FIGURE 5: *Analysed field with different values of θ for the advection constraint.*

4.3 Outlier detection

Diva offers three methods for detecting outliers, implemented in `divaqc`, `divaqcbis` and `divaqcter`, providing a list of suspect data points. The advantage of these tools is that they only use the data and analysis, not any ***a priori*** information. In the application we found 628 suspect values out of 30259.

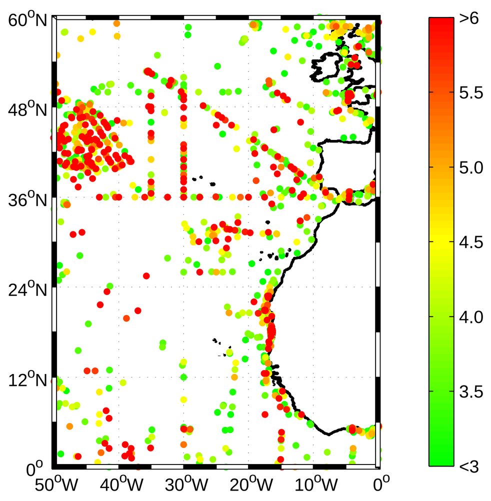
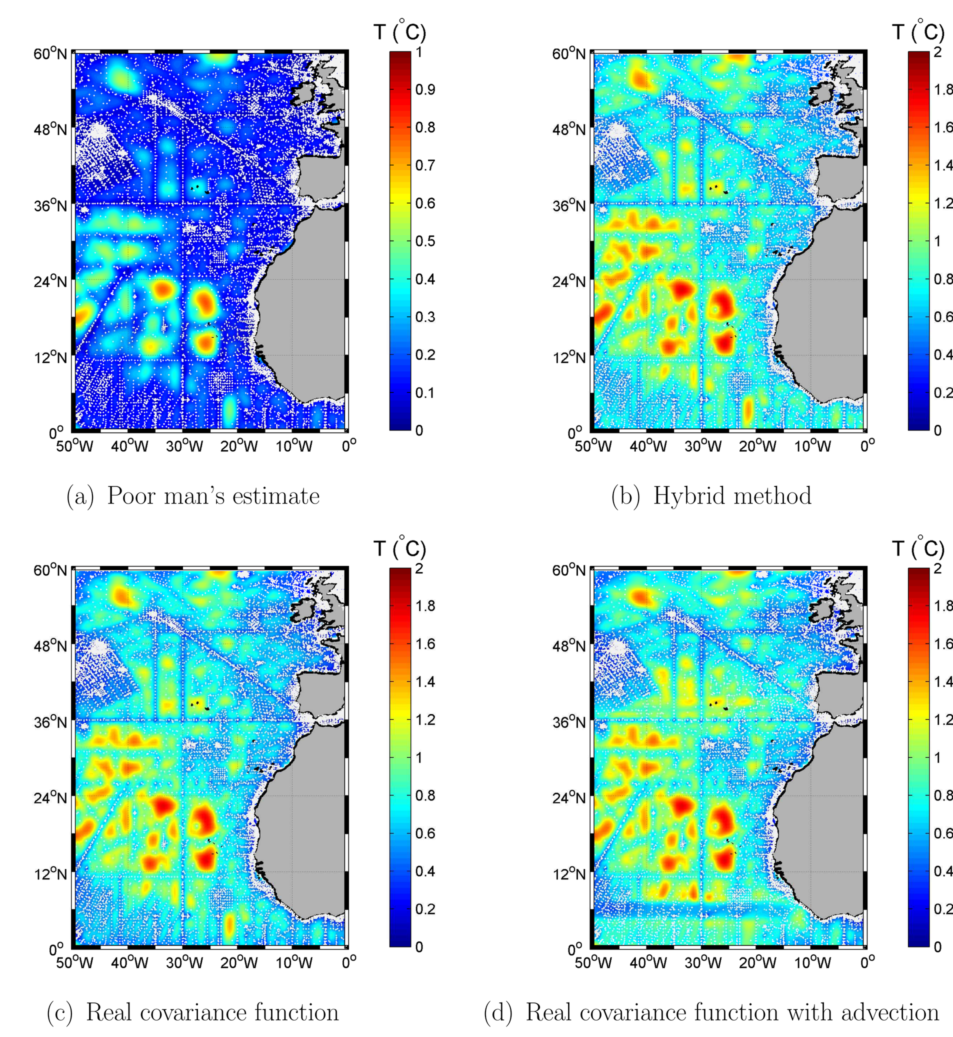


FIGURE 6: *Outlier detection with `divaqc`. Color scale indicates the normalized misfit.*

5 Error field

One of Diva major asset is the possibility of an error field computation. Previously the method was based on analogies with Optimal Interpolation and the error field was estimated as the analysis of covariance fields. Now three methods are available, depending on the data considered and type of analysis performed:

1. a *poor man's* error indicator, where the covariances to be analysed are replaced by 1; this method provides underestimated error field (Fig.7(a));
2. an *hybrid* error calculation [Rixen et al., (2000)], based on an additional analysis per point in which the error is to be calculated, which takes profit of the already performed LU decomposition (Fig.7(b));
3. an error calculation with the *Diva covariance function*: it demands two analyses per point in which the error is needed: one is done with the already existing LU decomposition of one Diva execution;
the other with an existing LU decomposition of another Diva execution (Fig.7(c)).
This method is recommended when working with variable correlation length or with advection (Fig.7(d)).



(a) Poor man's estimate
(b) Hybrid method
(c) Real covariance function
(d) Real covariance function with advection

FIGURE 7: *Error fields computed with different methods; white dots indicate data positions. Note the differences between color scales.*

6 Conclusion

We considered a large dataset covering the North Atlantic to illustrate the efficiency of various Diva software tools for producing realistic gridded fields.
Analysis parameters (correlation length and signal-to-noise ratio) were determined in an objective way using tools provided with the software.
Various error computation were tried their results underline the influence of data coverage.
Further work will be concentrated on multi-level analysis in order to create a complete climatology for the region of interest.

Acknowledgments

Diva was developed during E.U. MODB project and improved within the working package "Geostatistic analysis tools" (JRA4) of SeaDataNet project, an Integrated Infrastructure Initiative of the E.U. Sixth Framework Programme.
We thank the participants to the Diva workshops in Liège (November 2006) and in Calvi (November 2007) for their numerous valuable comments which help us to improve the software.

⋆ **Contact:** ctroupin@ulg.ac.be
http://modb.oce.ulg.ac.be/projects/1/diva