DIVA

Overview

Jean-Marie Beckers and GHER-DIVA group Aida Alvera, Alexander Barth, Charles Troupin, Mohamed Ouberdous, Sylvain Watelet

http://modb.oce.ulg.ac.be/GHER

Université de Liège
MARE-GHER Sart-Tilman B5
4000 Liège, Belgique
Outline

- Gridding
- Diva
- Implementations
- And for data managers?
- Summary
• **Gridding**

• **Diva**

• **Implementations**

• **And for data managers?**

• **Summary**
Common problem

Appears when trying to produce maps, calculate volume averages, prepare initial conditions for models, quality control of data ...
Solutions
Solutions
Solutions
Solutions
Estimation of today’s temperature in Ostende

- Observer 1: 14°C
- Observer 2: 16°C

Your best guess?
Estimation of today’s temperature in Ostende

- Observer 1: 14°
- Observer 2: 16°

Your best guess?
15°
Estimation of today’s temperature in Ostende

- Observer 1: 14°
- Observer 2: 16°

Your best guess?
15°

But what if observer 1 uses digital thermometer and observer 2 his finger?
Estimation of today’s temperature in Ostende

- Observer 1: 14°
- Observer 2: 16°

Your best guess?
15°

But what if observer 1 in Bruges uses digital thermometer and observer 2 in Ostende his finger?
Estimation of today’s temperature in Ostende

- Observer 1: 14°
- Observer 2: 16°

Your best guess?
15°

But what if observer 1 in Sidney uses digital thermometer and observer 2 in Ostende his finger?
Estimation of today’s temperature in Ostende

- Observer 1: 14°
- Observer 2: 16°

Your best guess?
15°

But what if observer 1 in Sidney uses digital thermometer and observer 2 in Ostende his finger?
Best guess probably 16° but with a higher error bar.

Exploit knowledge of errors and distance both for the estimate itself but also the error bars.
Optimal interpolation

Exploit knowledge of errors and distance both for the estimate itself but also the error bars

When done mathematically searching for the estimate with lowest expected error: OPTIMAL INTERPOLATION
Optimal, but needs some statistical information
Optimal Interpolation

- Need for covariance of the background field between data points: each element $i, j$ of $B$ provides the covariance between points in location $i$ and $j$. Covariance between a given point on the analysis grid and all data points is also needed.

- Need for observational error information (not only instrumental). Equivalently signal-to-noise ratio.

Analysis leads to spatial analysis and error field at any desired location if covariance between any two points is known. Practical problems in addition to covariance specification: huge matrix inversions for large data sets.
Background covariance

Problem, how to specify background covariances (between all data points and between data points and the desired analysis location).

- Normally obtained via statistics on data. Seldom possible (noticeable exception: satellite images).
- Standard OI: via functions $B_{ij} = f(r/L)$ where $r$ is the distance between points $i$ and $j$, but still function $f$ needs to be determined. $L$ is the so-called correlation length. Here statistics on all data couples as a function of distance. Example: $f = \sigma^2 \exp(-r^2/L^2)$.
- Via functionals (see Kernel of DIVA later)

Note that real covariance are reduced by obstacles or increased along current.
● Gridding

● Diva

● Implementations

● And for data managers?

● Summary
**DIVA Basics**

Variational Inverse Method, (Brasseur et al., 1996). Knowing data $d_j$ at location $(x_j, y_j)$, search the field $\varphi$ which minimizes

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

$$\|\varphi\| = \int_D (\alpha_2 \nabla\nabla \varphi : \nabla\nabla \varphi + \alpha_1 \nabla \varphi \cdot \nabla \varphi + \alpha_0 \varphi^2) dD$$  \hspace{1cm} (2)

The background field $\varphi_b$ is typically the data average value.

- $\alpha_0$ penalizes the field itself (anomalies),
- $\alpha_1$ penalizes gradients $\nabla \varphi$ (no spatial trends),
- $\alpha_2$ penalizes variability (regularization of second derivatives $\nabla\nabla \varphi$),
- $\alpha_\ast$ can be related to a length scale $L$ of the analysis,
- $\mu_j$ penalizes data-analysis misfits (objective).
Basics

\[ \mu = \frac{\sigma^2}{\epsilon^2} \frac{4\pi}{L^2} \]  

(3)

where the \( \sigma^2 / \epsilon^2 \) is known as a signal to noise ratio \( S/N \).

Solution by finite element method. Note decoupling of subbasins. (Each element is in fact composed by three sub-elements, each one with cubic functions)
Bad news 😞

- No error estimate comes with the method, only an indicator of data-coverage.
Bad news 😞

- No error estimate comes with the method, only an indicator of data-coverage.

- VIM brand already taken:
**Bad news**

- No error estimate comes with the method, only an indicator of data-coverage.

- VIM brand already taken:

- Method is equivalent to well established existing optimal interpolation (OI)
Good news 😊

- Name easily changed: DIVA (Data-Interpolating Variational Analysis)
- Optimal interpolation (OI) provides error estimates, so DIVA can also provide it via equivalence
- DIVA has some **practical advantages** over OI
**DIVA as OI**

DIVA is identical to the well known Optimal Interpolation

![Graph](image)

- if so-called reproducing kernel of the norm = **covariance function** of OI,
- if the noise is random, spatially uncorrelated and the signal/noise ratio parameter is identical with OI.

In this case, the OI solution = DIVA solution.

- Advantages of DIVA: regularization, fast finite-element solution, boundary effects taken into account.
- Difficulties: generalizations to 3D and multivariate versions are "hybrid".

Major direct advantage of DIVA: matrix to invert is related to the finite-element mesh, NOT the number of data. Useful for large data sets (Rixen et al., 2000). Equivalence allows to calculate error fields with DIVA even if formulation does not rely on error minimisation.
Illustration of covariance functions
Standard ODV analysis

(R. Schlitzer ODV example)
ODV-DIVA analysis

(R. Schlitzer ODV example)
Additions to basic tool

Advection constraint: Augmented cost function to deal with preferred correlation directions, e.g., via advection with velocity $u$ and diffusion $A$

$$J_a = J(\varphi) + \theta \int_D \left[ u \cdot \nabla \varphi - A \nabla \cdot \nabla \varphi \right]^2 dD$$  \hspace{1cm} (4)

Other features

- Error fields taking data distribution into account.
- Toolbox approach allowing to design own versions.
- 3D and 4D modes by looping, hydrostatic constraint in 3D mode.
- Cross validation tools to infer statistical parameters and error estimates.
- Climatology production version with heterogeneous data distributions (detrending).
- Outlier detection.
- ...
Covariances with advection
Spatial coherence of parameters: here correlation length obtained with covariance fitting (Troupin et al., 2010).
Signal to noise ratio

The most elusive parameter.

- Noise is not only instrumental error:
- Very hard problem to decide on value with dependent data (cross-validation approaches fail).

A series of estimation tools are provided with DIVA, but here the experience of oceanographers is critical. A posteriori analysis of residuals allows to verify coherence. With reasonable amount of data, parameter not critical for analysis but for error estimates.
Huge problems

LDEO data base with $4.5 \times 10^6$ measurements (Takahashi et al., 2009). Running on a laptop within a few minutes. Shown here, temperature fields.
Huge problems, outliers and relative error field

Outliers detected via comparison of statistically expected residuals (value provided by the DIVA analysis) and actual residuals (note that expected residuals decrease for large S/N). Error field (on the right) can be used to mask regions with large uncertainties (low data coverage and/or large errors on data).
Outlier detection to detect encoding problems

Add value of 27 in 29°E, 35°N (21788th data point): corresponds to a displaced profile within a subregion
Outlier with divaqc
Detrending

Heterogeneous data distribution:

First analysis show a bias for each year’s data with respect to first analysis. Subtract the bias estimate and redo the analysis, accumulating the bias. After convergence, detrended analysis + bias of the year.
Example without detrending

DIVA analysis of sin-cosine spatial structure with superimposed decadal, seasonal and daily cycles and noise.
Example with detrending

DIVA analysis of sin-cosine spatial structure with superimposed decadal, seasonal and daily cycles and noise.
Trends can also be retrieved
Heterogeneous case

Baltic Salinity Climatology (Bassompierre et al., 2010)
**Heterogeneous case, trends**

- Nutrient and climatic trends from Kattegat to Bothnian Bay (SDN products)

  - Phosphate is released by the sediment, when hypoxia occurs.

  - Big storm in January 1993, caused massive water through the Danish straights inflows with high salinity and oxygenated the Baltic.

  - Increased temperature increases risk of hypoxia.

  - $tphs0707005 =$ total phosphate July -100 m
  - $psal01011008 =$ salinity January level -40 m
  - $temp101010010 =$ temperature October -20 m

  (Bassompierre *et al.*, 2010)
- Gridding
- Diva
- Implementations
- And for data managers?
- Summary
How to use DIVA?
Try parameters of gridding in ODV

On Display Style

- change analysis method (quick, VG, DIVA)
- change X scale-length and Y scale-length
Try parameters of gridding in ODV

On DIVA Settings
- Change Signal-to-Noise Ratio
Observed changes in analyses due to changes in parameters (here signal-to-noise ratio from 1 to 100)
Diva-on-web

Diva-on-web

Data upload (3 column ascii file or ODV4) and output grid definition
Diva-on-web

Analysis parameter definition (or fit)

**Analysis with Diva**

Correlation length [deg]: 10.3516
Signal to noise ratio: 1.08879

**Optional parameters**

Maximum rel. error (from 0 to 1): 0.3
Analysis (and numerical grid)

**Analysis with Diva**

- **Correlation length [deg]:** 10.3516
- **Signal to noise ratio:** 1.08679

**Quality of the fit (0: bad 1: good):** 0.963967

**Optional parameters**

- **Maximum rel. error (from 0 to 1):** 0.3
Diva-on-web

Plotting options
Diva-on-web

Download options

Analysis with Diva

Correlation length [deg]: 10.3516
Signal to noise ratio: 1.08679

Quality of the fit (0: bad 1: good): 0.963967

Optional parameters

Maximum rel. error (from 0 to 1): 0.3

Available formats:

- NetCDF file (.nc)
- Octave or Matlab (.mat)
- Google Earth (.kml)
- Image Format: PNG, PDF, EPS, SVG

-11.25000, -105.00000
Exploitation of WMS-OGC layering techniques

Analysis with Diva

Correlation length [deg]: 10.3516
Signal to noise ratio: 1.08679

Quality of the fit (0: bad 1: good): 0.963967

Optional parameters

Maximum rel. error (from 0 to 1):

Copy and paste this line of code in your web page. To keep your results on our web-server, you must click on save below. We would appreciate if you leave your name and email address [more].

Name: 
Email: 

Save
Help

More information

Software download for advanced users: GODIVA 4D

http://modb.oce.ulg.ac.be/mediawiki/index.php/DIVA
● Gridding
● Diva
● Implementations
● And for data managers?
● Summary
**DIVA and data-managers as providers**

Scientist rarely use single profiles

- Need for standardisations to simplify automated processing. If not strict adherence
  - Risk of missing data without user noting the profile was not used
  - Risk of crashing a complex calculation because a single profile causes problems
- Need for correct metainformation (units, names etc): otherwise again incorrect selections, mixing of different units etc
- QUALITY information: standardised flags for data selection AND weighting of information

Users do NOT want to go through checking of all profiles again and if they want to merge with other data they want to automate transformations.
DIVA as a tool for data-managers

DIVA can also be used by data-managers themselves

- Checking data flow and conformity to standards by trying to use DIVA (diva-on-web or GODIVA)
- Visual representation and verification of data set (not only profiles but spatial structures)
- Exploit outlier detection using spatial information on top of standard QC
- For interested centers: diva analysis as exploratory tool for data bases (mapping on the fly with diva-on-web engine or batch mode of ODV)
- Gridding
- Diva
- Implementations
- And for data managers?
- Summary
Possible uses for you

DIVA can be used in several ways. As data managers you can work with DIVA as format checking tool and gridding tool (gridding is not plotting). When using DIVA

- Report non-standard ODV4 files retrieved from SDN interface to SDN-TTT
- Report diva runtime errors (be it in ODV or diva-on-web or GODIVA) to GHER
- For installation of GODIVA (4D batch mode), contact GHER
- For bugs on diva-on-web, use web interface for reporting
- Do not hesitate to add things to a wishlist
- Strong interactions during DIVA workshops
  

but you can also point scientists to DIVA if they use your data base.


http://orbi.ulg.ac.be/handle/2268/4299

http://orbi.ulg.ac.be/handle/2268/125731