SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

WP8 and WP9 developments

Data-Interpolating Variational Analysis (Diva) developments

C. Troupin, A. Barth, M. Ouberdous, A. Alvera-Azcárate & J.-M. Beckers

GHER GeoHydrodynamics and Environment Research

Université de Liège ULg

2nd Plenary Meeting, 26-27 September 2013, Lucca (Italy)

How to install and use the new tools?

Main principles for Diva developments

1. Backward compatibility, portability and independence on proprietary software

Main principles for Diva developments

1. Backward compatibility, portability and independence on proprietary software
2. Development & innovation: **user-driven**

Main principles for Diva developments

1. Backward compatibility, portability and independence on proprietary software
2. Development & innovation: **user-driven**
3. Strong scientific background

Releases: 4.5.1 – March 2013

New features: from user feedback during Diva workshop 2012 (Roumaillac)

Releases: 4.5.1 – March 2013

New features: from user feedback during Diva workshop 2012 (Roumaillac)

- Advection constraint with linear decay rate and local sources

Releases: 4.5.1 – March 2013

New features: from user feedback during Diva workshop 2012 (Roumaillac)

- Advection constraint with linear decay rate and local sources
- `divadetrend`: change in the detrending order

Releases: 4.5.1 – March 2013

New features: from user feedback during Diva workshop 2012 (Roumaillac)

- Advection constraint with linear decay rate and local sources
- `divadetrend`: change in the detrending order
- Two new error calculations
  - `divacpme`: quick & better than original poor man's error
  - `divaexerr`: almost exact error calculation, much faster than the exact calculation

Releases: 4.5.1 – March 2013

New features: from user feedback during Diva workshop 2012 (Roumaillac)

- Advection constraint with linear decay rate and local sources
- `divadetrend`: change in the detrending order
- Two new error calculations
- Simplified procedure for installation/compilation + tests

Releases: 4.5.1 – March 2013

New features: from user feedback during Diva workshop 2012 (Roumaillac)

- Advection constraint with linear decay rate and local sources
- `divadetrend`: change in the detrending order
- Two new error calculations
- Simplified procedure for installation/compilation + tests
- Housekeeping of the code
  (simplifications, error messages, cleaning up of code, further optimisations, elimination of depreciated tools)

Releases: 4.5.1 – March 2013

New features: from user feedback during Diva workshop 2012 (Roumaillac)

- Advection constraint with linear decay rate and local sources
- `divadetrend`: change in the detrending order
- Two new error calculations
- Simplified procedure for installation/compilation + tests
- Housekeeping of the code
- Updated user guide
  (augmented with examples and new tool descriptions)

Releases: 4.5.1 – March 2013

New features: from user feedback during Diva workshop 2012 (Roumaillac)

- Advection constraint with linear decay rate and local sources
- `divadetrend`: change in the detrending order
- Two new error calculations
- Simplified procedure for installation/compilation + tests
- Housekeeping of the code
- Updated user guide
- Possibilities to call Diva from other software via system calls

Releases: 4.5.1 – March 2013

New features: from user feedback during Diva workshop 2012 (Roumaillac)

- Advection constraint with linear decay rate and local sources
- `divadetrend`: change in the detrending order
- Two new error calculations
- Simplified procedure for installation/compilation + tests
- Housekeeping of the code
- Updated user guide
- Possibilities to call Diva from other software via system calls
- `divadoxml` adapted to new specifications from IFREMER

Releases: 4.6.1 – June 2013

Current official version

Releases: 4.6.1 – June 2013

Current official version

- Two additional solvers
  - parallel version
  - iterative version

Releases: 4.6.1 – June 2013

Current official version

- Two additional solvers
  - parallel version
  - iterative version
- Optimisations for large data sets

Releases: 4.6.1 – June 2013

Current official version

- Two additional solvers
  - parallel version
  - iterative version
- Optimisations for large data sets
- Optimisations of file exchanges for use with ODV

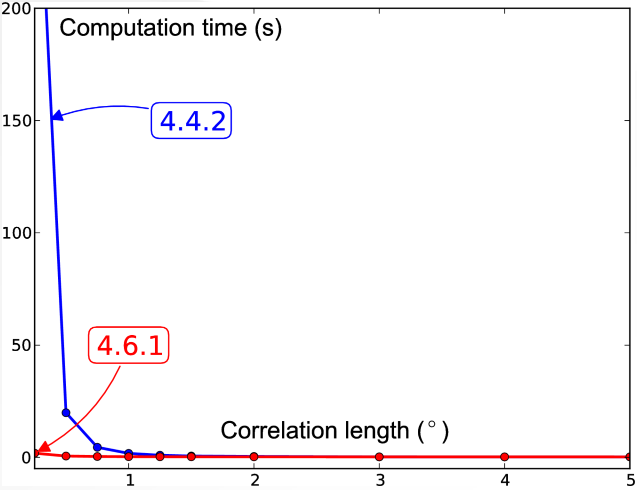
Releases: 4.6.1 – June 2013

Current official version

- Two additional solvers
  - parallel version
  - iterative version
- Optimisations for large data sets
- Optimisations of file exchanges for use with ODV
- Highly optimised new version of the grid generator

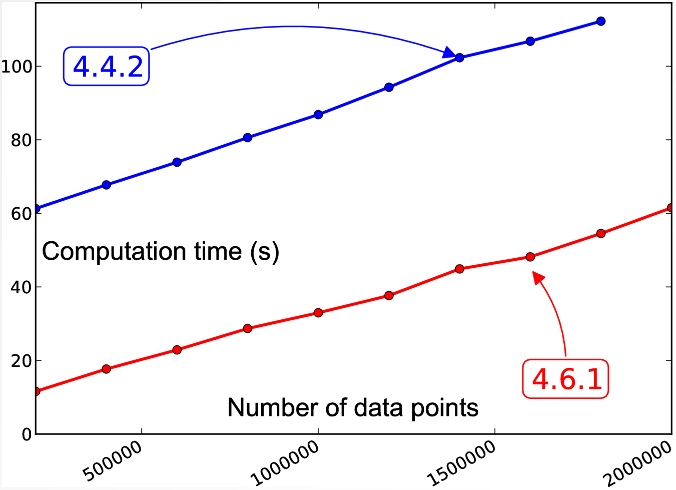
Better, faster, stronger …

Mesh:
very fine
meshes in a
few seconds



Better, faster, stronger …

Analysis:
2 million data
≈ 1 minute



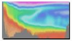
Better, faster, stronger ...

Solvers:

- Direct
- Parallel
- Iterative

Better, faster, stronger . . .

Mesh: $\approx 100 \times$ faster
Analysis: $\approx$ 5-10 $\times$ faster

$\rightarrow$ also quicker in ODV 

Releases: 4.7.1 – expected November 2013

Beta testers …

Releases: 4.7.1 – expected November 2013

Developed features

- Correlated observational errors

Releases: 4.7.1 – expected November 2013

Developed features

- Correlated observational errors
- Better file structures
  (input and driver better separated from command) in 4D loops

Releases: 4.7.1 – expected November 2013

Developed features

- Correlated observational errors
- Better file structures
- Automatic selection of solver (parallel, serial, iterative)
  depending on the problem type and size

Releases: 4.7.1 – expected November 2013

Developed features

- Correlated observational errors
- Better file structures
- Automatic selection of solver (parallel, serial, iterative)
- Retrieval of topographies from Diva-on-web

Releases: 4.7.1 – expected November 2013

Developed features

- Correlated observational errors
- Better file structures
- Automatic selection of solver (parallel, serial, iterative)
- Retrieval of topographies from Diva-on-web
- Improved version of the almost exact error calculation with boundary effects

Releases: 4.7.1 – expected November 2013

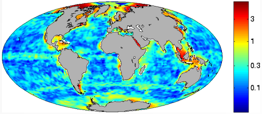
Developed features

- Correlated observational errors
- Better file structures
- Automatic selection of solver (parallel, serial, iterative)
- Retrieval of topographies from Diva-on-web
- Improved version of the almost exact error calculation with boundary effects
- Incorporation of metadata
  (EDMO-CDI identifier, space-time location)
  into 4D NetCDF files of climatologies

Scientific developments – innovations

4-dimensional generalisation: `divand`

RMS 3D analysis



- Derivation of the kernel for n dimensions
- Additional constraint
- Algorithms (primal and dual formulations)

Released code version available at:
http://modb.oce.ulg.ac.be/mediawiki/index.php/Divand

Scientific developments – innovations

Spatially correlated observations

Ideally:	observation errors not correlated
Reality:	clusters of observations (cruises, . . .)
Consequence:	observations error covariance matrix is not diagonal

Scientific developments – innovations

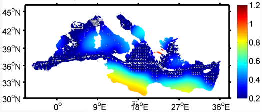
New error computation

Poor man's error: quick, but error underestimation

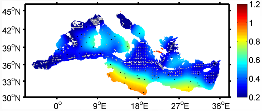
Real covariance: correct error estimation but very slow

Now: two quicker/more accurate methods

Clever poor man's estimate



Almost exact error



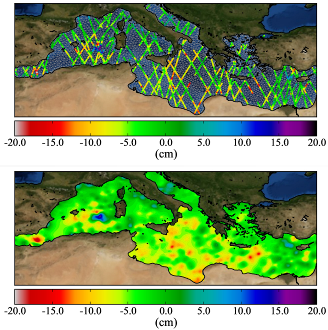
PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

SeaDataNet

Scientific developments – innovations

Adaptation to altimetry data

- Particular temporal/spatial coverage
- Input files: NetCDF
- Modified data weights according to time of measurement



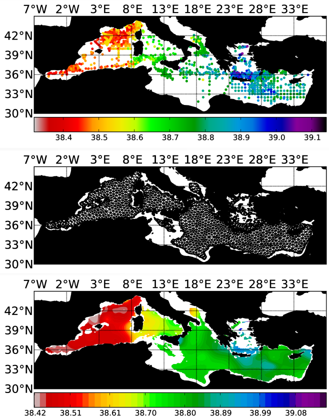
Scientific developments – innovations

Python plotting tools

→ python™

- Free alternative to matlab/octave
- Easily deals with NetCDF
- Publication quality figures with Matplotlib

http://modb.oce.ulg.ac.be/mediawiki/index.php/Diva_python



Publications

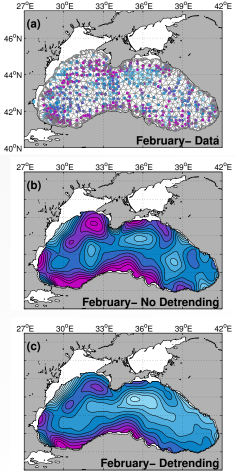
Detrending:

Recognizing temporal trends in spatial interpolation : an application to the Black Sea Cold Intermediate Layer and mixed layer depth

A. Capet, C. Troupin, J. Carstensen, M. Grégoire & J.-M. Beckers

Ocean Dynamics

Under revision

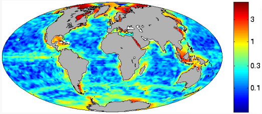


Publications

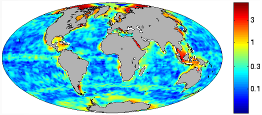
Diva-nd:

divand-1.0: n-dimensional variational data analysis for ocean observations
A. Barth, J.-M. Beckers, C. Troupin,
A. Alvera-Azcárate & L. Vandenbulcke
Geoscientific Model Development
Under revision

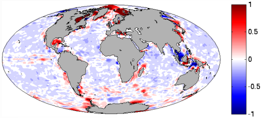
RMS 2D analysis



RMS 3D analysis



RMS(2D) – RMS(3D)



Publications

Error field:

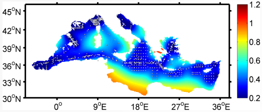
Approximate and efficient methods to assess error fields in spatial gridding with Diva (Data Interpolating Variational Analysis)

J.-M. Beckers, A. Barth, C. Troupin & A. Alvera-Azcárate

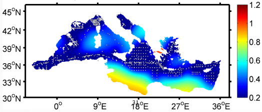
Journal of Atmospheric and Oceanic Technology

Under revision

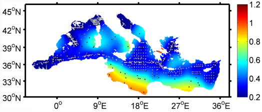
Real covariance



Clever poor man's estimate



Almost exact error



Diva workshop

Where?	STARESO station (ULg), Calvi, FRANCE
When?	Monday 4 (arrival) – Friday 8 (departure) November 2013
Who?	SeaDataNet / EMODnet partners, all levels
What?	Installation, test cases, 2D, 3D, 4D, . . .
✉	ctroupin@ulg.ac.be

More details: http://modb.oce.ulg.ac.be/mediawiki/index.php/Diva_workshop_2013_Stareso

→ Conclusions

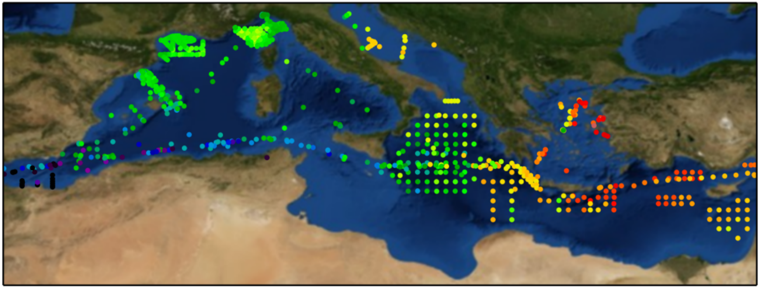
Impact of correlated observation errors

$$\begin{aligned} \text{observation error} &= \text{measurement error} \\ &+ \textit{representativity}\ \text{error} \end{aligned}$$

$$\begin{aligned} \text{field(x,y,t)} &= \text{mean state(x,y)} \\ &+ \text{mesoscale variability(x,y,}\mathbf{t}\text{)} \end{aligned}$$

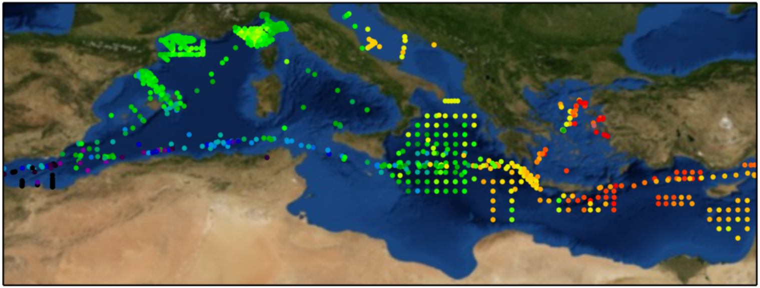
Impact of correlated observation errors

How can we derive the *mean state*,
based on point measurements of a *field*?



Impact of correlated observation errors

Oceanographic data sets are generally *clustered* in space and time (measurement campaigns)



Impact of correlated observation errors

Diva currently does not take into account the correlation of the observation errors

Synthetic experiments

- Generate a random mean field (longitude and latitude)
- Generate a random mean mesoscale field
  (longitude, latitude and **time**)
  given a length-scale (space and time) and variance
- Sum both
- Extract observations

Synthetic experiments

- Try to determine the mean field based on observations using:
  1. correct observations error covariance $\mathbf{R}$ — non-diagonal R
  2. only the diagonal part $\mathbf{R}$ — diagonal R
  3. bin all close-by observations together within the correlation length-scale and use corresponding diagonal $\mathbf{R}$ — binning
  4. use all observations, but inflate error variance of observations in cluster — inflation
  5. use all observations, but inflate error variance of observations by the product $\mathbf{R}\vec{v}$ ($\vec{v} = 1\ 1\ 1 \ldots 1\ 1$) — inflation 2

Synthetic experiments

- Mean state, mesoscale field and observation locations randomly chosen.
  Experiments are repeated 100 times and average are shown.
- Location of 100 observations is controlled by the number of clusters and the *width* of each cluster.

Synthetic experiments

number of clusters: 1, 3, 5, 7, 9, 20, 40, 60, 80, 100

cluster width (relative to the domain size) =
0.031623, 0.046416, 0.068129, 0.1, 0.146780,
0.215443, 0.316228, 0.464159, 0.681292, 1

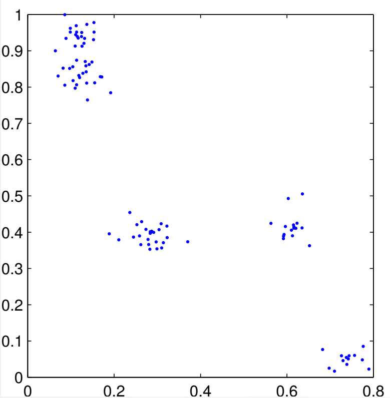
variance of background = 1

variance of observation = 2

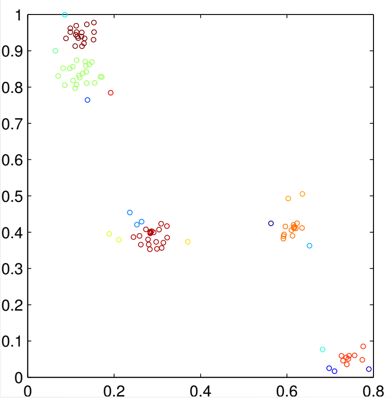
domain: $100 \times 100 \times 100$ grid points
$[0, 1] \times [0, 1] \times [0, 1]$

correlation lengths (space and time) = 10

Distribution of data points



Distribution of data points



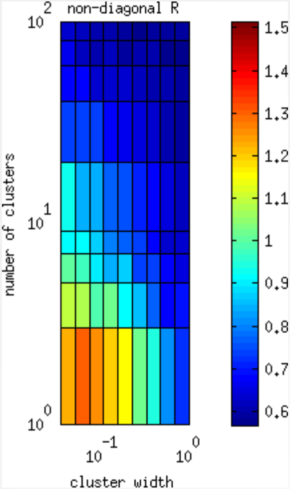
Observations ’close-by’

Cluster analysis: determine highly-correlated observations.

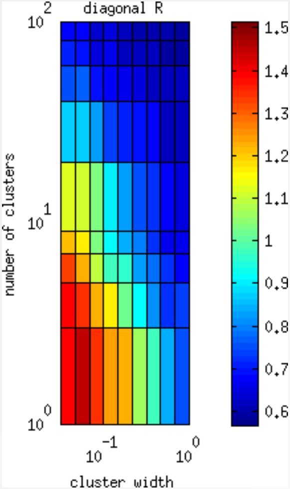
Observation error correlation length < cluster size (typically).

Such observation clusters used for binning and covariance inflation.

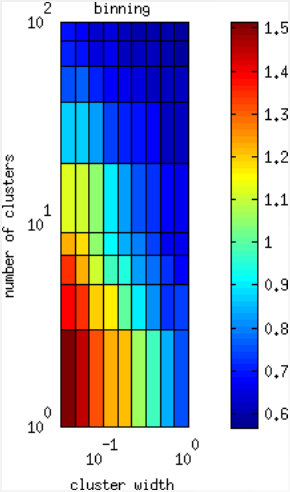
RMS between the true and reconstructed fields



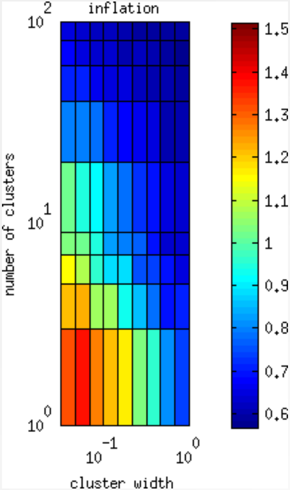
RMS between the true and reconstructed fields



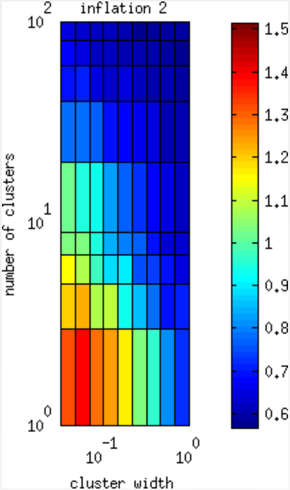
RMS between the true and reconstructed fields



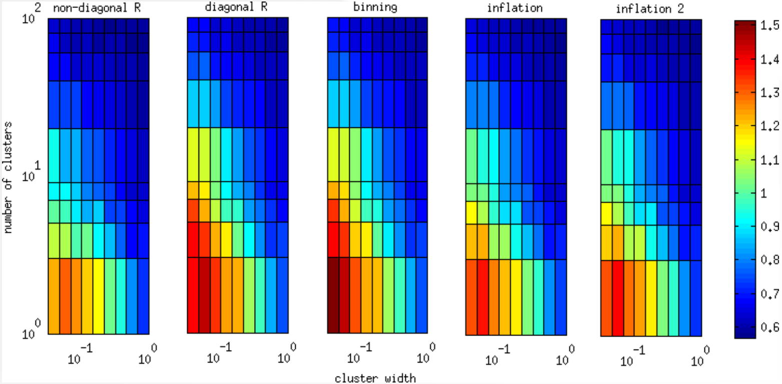
RMS between the true and reconstructed fields



RMS between the true and reconstructed fields

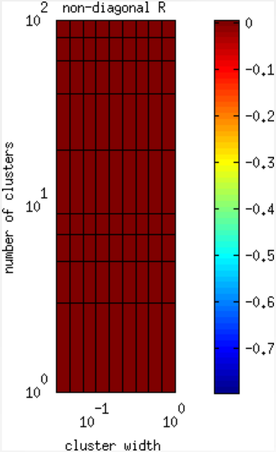


RMS between the true and reconstructed fields



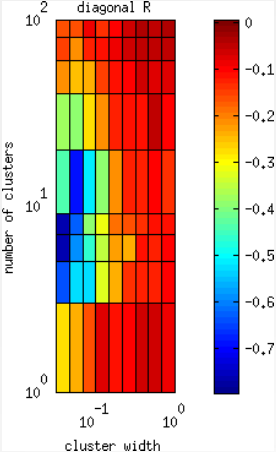
- Data is more clustered → reconstruction degrades
- More severe degradation with diagonal R and binning

Degradation relative to non-diagonal R



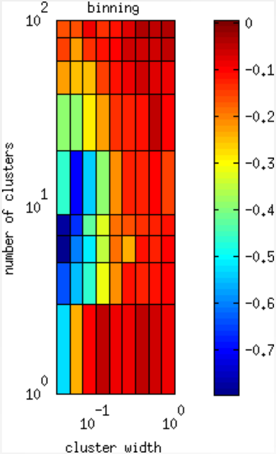
- Skill score using non-diagonal R approach
  as reference.
- Skill score is negative since all other approaches are approximations.

Degradation relative to non-diagonal R



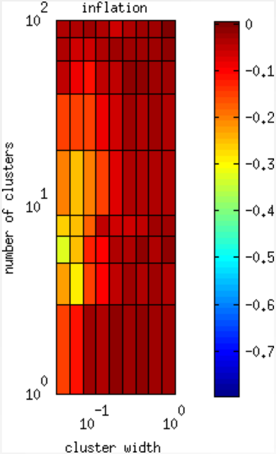
- Skill score using non-diagonal R approach
  as reference.
- Skill score is negative since all other approaches are approximations.
- Degradation is quite large for 10 clusters and 0.05 cluster width using the diagonal R and binning approaches.

Degradation relative to non-diagonal R



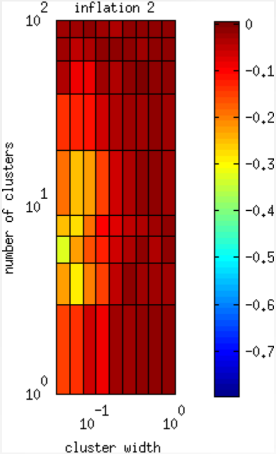
- Skill score using non-diagonal R approach
  as reference.
- Skill score is negative since all other approaches are approximations.
- Degradation is quite large for 10 clusters and 0.05 cluster width using the diagonal R and binning approaches.

Degradation relative to non-diagonal R



- Skill score using non-diagonal R approach
  as reference.
- Skill score is negative since all other approaches are approximations.
- Degradation is quite large for 10 clusters and 0.05 cluster width using the diagonal R and binning approaches.
- Degradation is much less using inflation

Degradation relative to non-diagonal R



- Skill score using non-diagonal R approach
  as reference.
- Skill score is negative since all other approaches are approximations.
- Degradation is quite large for 10 clusters and 0.05 cluster width using the diagonal R and binning approaches.
- Degradation is much less using inflation

Impact on error variance of the analysis

- Using a diagonal observation error covariance also affects the expected error variance of the analysis
- 5 clusters with a width of 0.05
- RMS of error variance relative to non-diagonal R

Experiment	RMS of error variance
diagonal R	0.144947
binning	0.126911
inflation	0.134404
inflation 2	0.128268

Summary

- Best method: inflation 2
  (based on the sum of all elements of $\mathbf{R}$ in a given row)

Summary

- Best method: inflation 2
  (based on the sum of all elements of $\mathbf{R}$ in a given row)
- In comparison the binning approach leads to a large degradation,
  but to a relatively smaller degradation of the error estimate.

Summary

- Best method: inflation 2
  (based on the sum of all elements of $\mathbf{R}$ in a given row)
- In comparison the binning approach leads to a large degradation,
  but to a relatively smaller degradation of the error estimate.
- In practice, the question remains: *how to estimate the correlation length of the observation error covariance?*

Conclusions

- Improved documentation, installation, log files, . . .
- Enhanced mesh generation, solver (direct, parallel, iterative) and error computation
- Cutting-edge developments in spatial interpolation theory
- Adapted version of OceanBrowser

Conclusions

- Improved documentation, installation, log files, …
- Enhanced mesh generation, solver (direct, parallel, iterative) and error computation
- Cutting-edge developments in spatial interpolation theory
- Adapted version of OceanBrowser

Conclusions

- Improved documentation, installation, log files, . . .
- Enhanced mesh generation, solver (direct, parallel, iterative) and error computation
- Cutting-edge developments in spatial interpolation theory
- Adapted version of OceanBrowser

Conclusions

- Improved documentation, installation, log files, . . .
- Enhanced mesh generation, solver (direct, parallel, iterative) and error computation
- Cutting-edge developments in spatial interpolation theory
- Adapted version of OceanBrowser

Thanks to all
for your attention

Thanks to Giuseppe & his team
for everything