Reconstruction of SMOS data over the North Atlantic Ocean using DINEOF

Aida Alvera-Azcárate
Alexander Barth
Jean-Marie Beckers

Contact: **a.alvera@ulg.ac.be**

AGO-GHER, University of Liege, Belgium

In this poster
- Study of Sea surface salinity (SSS) data obtained from SMOS over the North Atlantic ocean

- Use of DINEOF (Data Interpolating Empirical Orthogonal Functions) to reconstruct missing data, detect outliers and reduce noise

- Study of spatial and temporal distribution of reconstructed dataset, outlier distribution and EOF basis

1

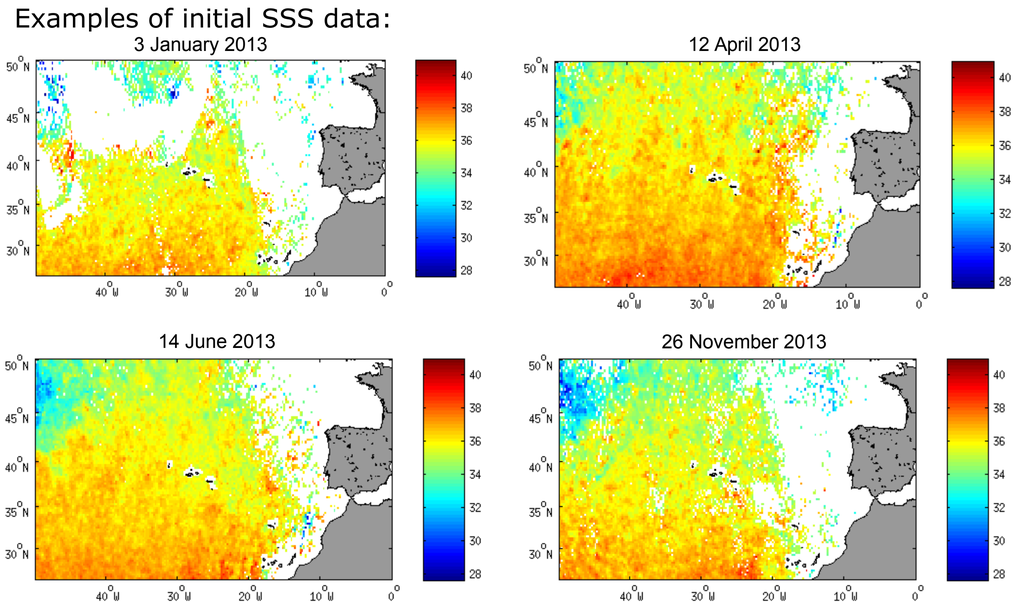
Data used

SSS from SMOS-BEC, 3-day average fields
Zone: North-East Atlantic Ocean
Period: January – December 2013
No quality flags were used

Problems to solve

- Outliers - Noise - Missing data

Examples of initial SSS data:

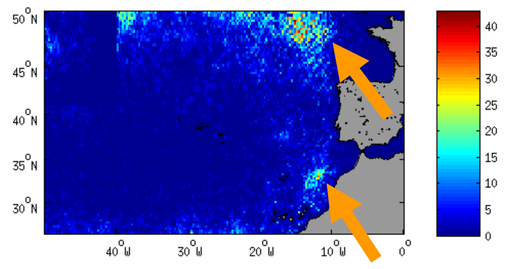


2

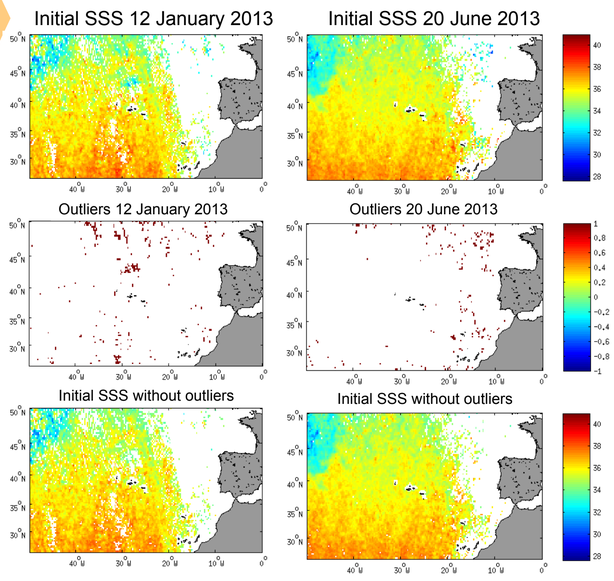
Outlier detection

- A first DINEOF analysis is performed on the initial data
- Three tests are applied to classify pixels as suspect:
  - Departure from the DINEOF truncated EOF basis
  - Departure from a local median
  - Proximity to clouds and land
- A weighted sum of these 3 tests allows to determine which pixels will be finally classified as outliers
- For this particular configuration:
  - Weights: EOF test (0.7), local median (0.2), proximity to clouds/land (0.1)
  - Threshold level to classify a pixel as outlier: 2
- Data at the Gulf Stream: smaller weight (1/3 of the outlier test result) to allow for low salinity values

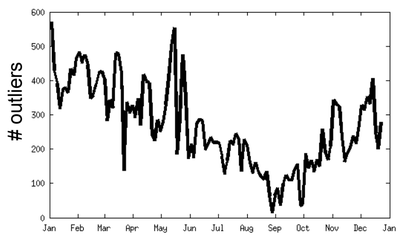
The sum of the pixels classified as outlier for all 2013 reveals some **"hot zones"** for outlier occurrence:



Examples:



It also appears there is some seasonality in the number of outliers detected:



DINEOF

Data Interpolating Empirical Orthogonal Functions

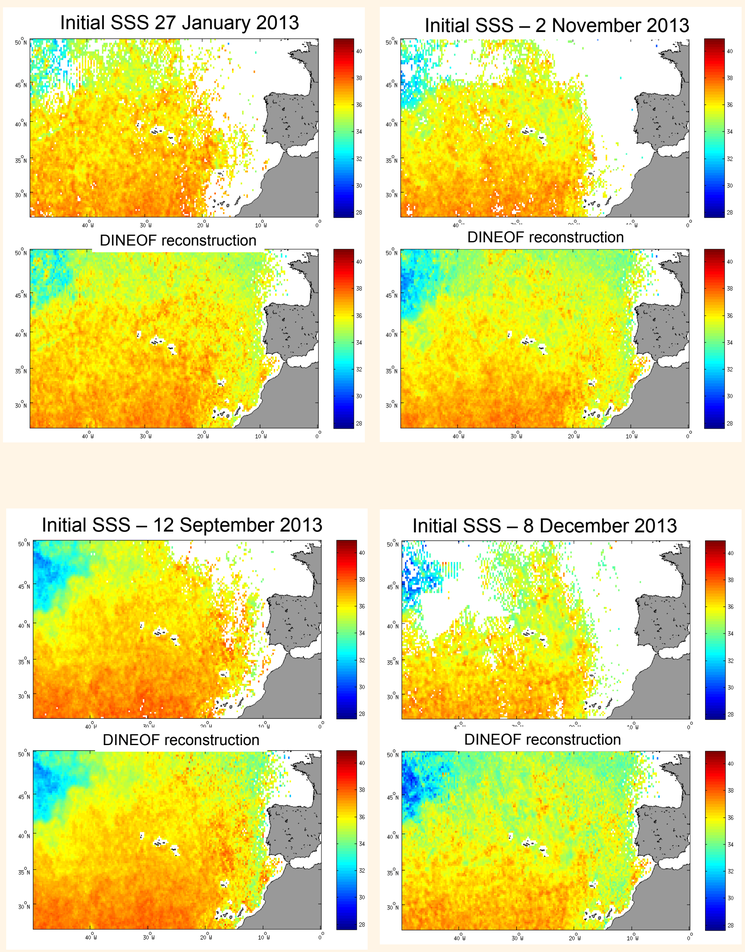
- Technique to **fill in missing data** in geophysical data sets
- **Truncated EOF** basis to calculate missing data (iterative method)
- Optimal number of EOFs?: reconstruction error by **cross-validation**
- Uses EOF basis to infer missing data: **non-parametric** in its basic form
- No need of a priori information (correlation length, covariance function...)
- Spatio-temporal coherence exploited to calculate missing values
- EOFs extract main patterns of variability

3

Results of DINEOF reconstruction

- 75 EOFs are retained by DINEOF
- 96.3 % of explained variance
(1st EOF: 76.6%; 2nd EOF: 4.27%; 3rd EOF: 1.85)
- Cross-validation error: 0.62

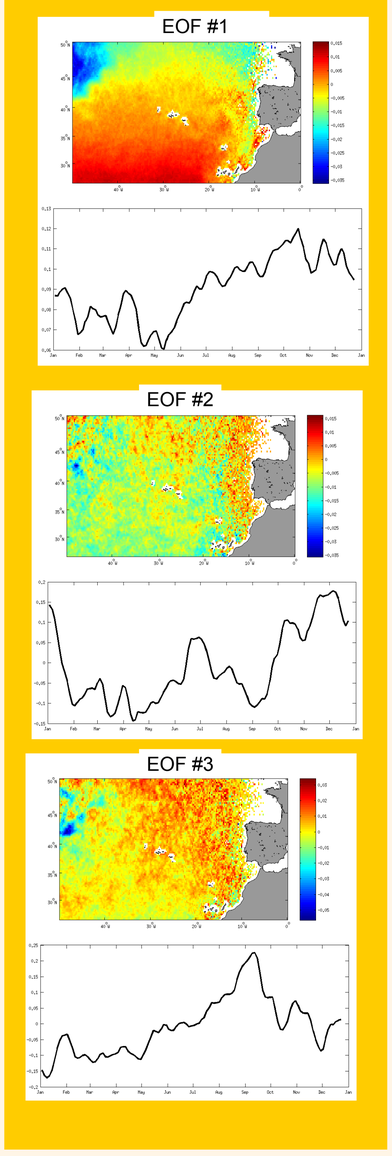
After removal of the outliers, this second DINEOF step is mainly to fill in missing data:



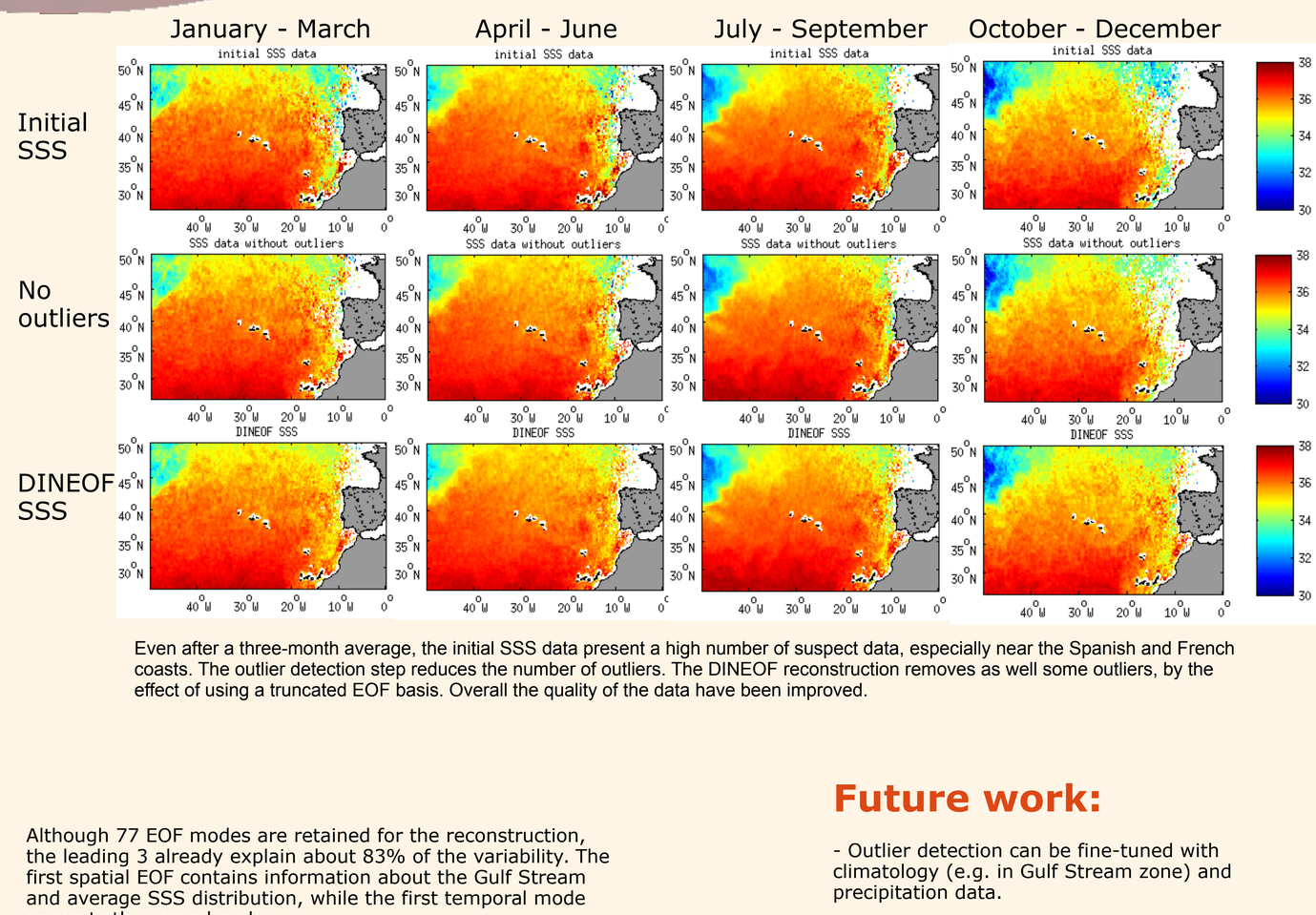
The reconstructed images shown here demonstrate the capabilities of DINEOF to reconstruct SSS even in zones with very high percentage of missing data.

Only pixels that had data more than 2% of the time have been reconstructed, which explains that there are still missing data along the coast. This percentage should probably be incremented to 5% to decrease the noise in these zones.

EOF modes



Impact of outlier detection and DINEOF reconstruction on seasonal averages



Even after a three-month average, the initial SSS data present a high number of suspect data, especially near the Spanish and French coasts. The outlier detection step reduces the number of outliers. The DINEOF reconstruction removes as well some outliers, by the effect of using a truncated EOF basis. Overall the quality of the data have been improved.

Although 77 EOF modes are retained for the reconstruction, the leading 3 already explain about 83% of the variability. The first spatial EOF contains information about the Gulf Stream and average SSS distribution, while the first temporal mode presents the annual cycle.

The second and third temporal EOFs already contain a high amount of noise near the coast, although there is also some information about the general north-south gradient and the Gulf Stream.

Future work:

- Outlier detection can be fine-tuned with climatology (e.g. in Gulf Stream zone) and precipitation data.

- Use of in situ and Aquarius SSS to improve the maps of SSS

- An approach to decrease the high amount of noise could be a division of the domain in zones, allowing DINEOF to find a different number of optimal EOFs for the reconstruction. Typically one could expect that fewer EOFs are retained near the coast.

Acknowledgements
This work is financed by the ESA-STSE Pathfinders project "Improving Sea Surface Salinity Estimates through Multivariate and Multisensor Analyses"

SMOS salinity data were produced by the Barcelona Expert Centre (www.smos-bec.icm.csic.es), a joint initiative of the Spanish Research Council (CSIC) and Technical (University of Catalonia (UPC), mainly founded by the Spanish National Program on Space.

In situ salinity data were obtained from the Coriolis database (http://www.coriolis.eu.org) and World Ocean Database (http://www.nodc.noaa.gov/)

More information: **http://modb.oce.ulg.ac.be/DINEOF**

Some DINEOF references:

Development of DINEOF:
-Beckers and Rixen, 2003
JAOT, 20(12):1839-1856.
- Alvera-Azcarate et al, 2005
Ocean Model. 9:325-346.

Multivariate application:
Alvera-Azcarate et al, 2007
JGR, 112:C03008

Error maps:
Beckers et al 2006
Ocean Sci., 2(2):183-199

Outlier detection:
A. Alvera-Azcárate et al. 2012
Remote Sens. Environ. 119:84–91

Temporal correlation in EOFs
Alvera-Azcarate et al, 2009.
Ocean Sci., 5, 475-485.