Recovery of sea level fields of the last decades from altimetry and tide gauge data

*Francisco M. **Calafat**, Damià Gomis, Ananda Pascual, Marta Marcos and Simón Ruiz*
Mediterranean Institute for Adavanced Studies, Palma de Mallorca

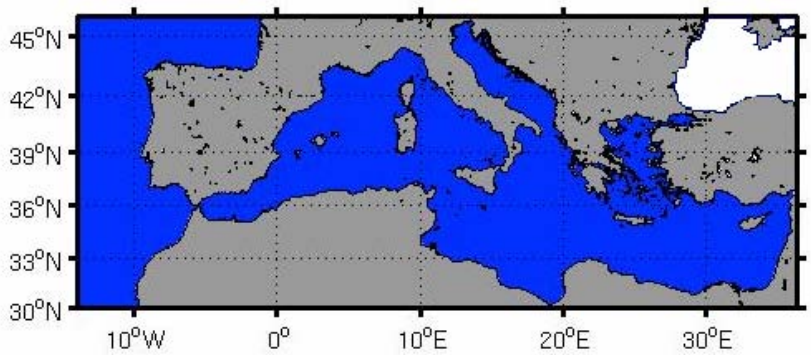
OUTLINE

1. **Objective of this work.**
2. **The dataset.**
3. **Methodology:**
   - **3.1. Principal components regression.**
   - **3.2. Substitution of leading PCs (principal components).**
4. **Comparison between both methodologies:**
   - **4.1. Prediction skill over the whole domain.**
   - **4.2. Prediction skill for past times.**
5. **Sensitivity.**
6. **The reconstruction: trends.**
7. **Conclusions.**

1. Objective of this work

WHAT IS THE OBJECTIVE OF THIS WORK?

- **to reconstruct the monthly distribution of sea level in the Mediterranean Sea and the northeastern sector of the Atlantic Ocean for the period 1950-2000.**



WHAT DATA WILL BE USED FOR THE RECONSTRUCTION?

- **The available long tide gauges series will be combined with the complete spatial coverage offered by satellite altimetry.**

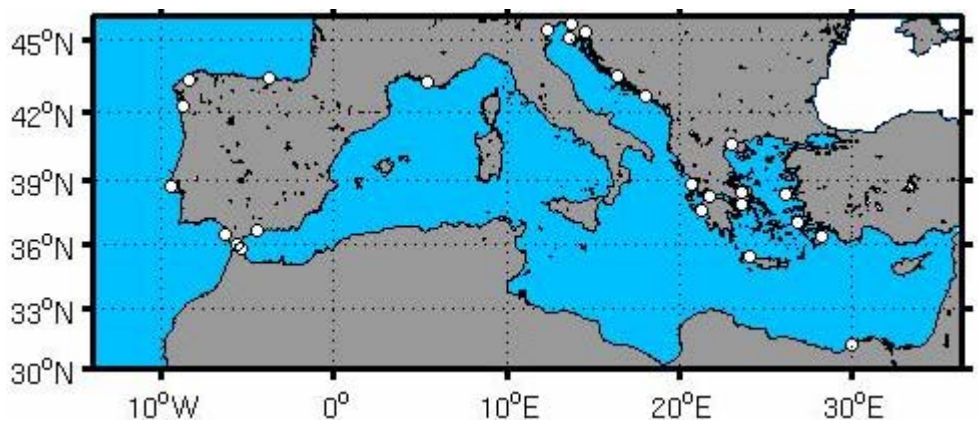
WHAT METHODOLOGIES WILL BE USED?

- **i) a Principal Components regression of the altimetry dataset on the tide gauge one and ii) a substitution of leading Principal Components from altimetry for the ones from tide gauges.**

2. The dataset

2.1. The tide Gauge dataset

- **Obtained from the Permanent Service for Mean Sea Level (PSMSL)**
- **Monthly mean sea level data. Seasonal cycle has been removed.**
- **Gaps smaller than 3 months were filled by using splines, while gaps larger than 2 months were filled by means of a multiple linear gression.**
- **27 tide gauges are selected out of a total of 68 stations, 7 of them span the period 1950-2000 and 20 span the period 1969-2000. Among those 27 tide gauges 4 were held aside for not being coherent to their neighbours.**
- **The reconstruction is carried out for the periods 1969-2000 by using the 23 tide gauges and 1950-2000 by using the 7 longest time series.**



2. The dataset

2.1. The altimetry dataset

- **The fields were obtained by combining several altimeter missions, namely: Topex/Poseidon, Jason-1, ERS1/2 and ENVISAT.**
- **The resolution of altimetry fields is 1/4º, resulting in a total of 6983 grid points covering the selected domain.**
- **The period spanned by the altimetry data is 1993-2005.**
- **In order to recover the total sea level signal, we added back the atmospheric component of sea level (the MOG2D outputs) to SLA gridded fields.**
- **Afterwards, every grid-point time series was filtered out with a 30 day running filter and the final product was subsampled in order to have a monthly temporal resolution.**

3. Theory and implementation

3.1. Principal component regression.

- **Let Y be an *nxp* matrix containing the alimetry dataset.**
- **Let X be an *nxq* matrix containing the tide gauge dataset.**
- **We want to describe Y by the model :** $Y = X\Gamma + \mathrm{E}$

  where Γ **is a matrix containing the coefficients of the regression.**
- **By substituting the singular value decomposition of X into the model we can write** $Y = U\beta + \mathrm{E}$**, where** $\beta = LF^T\Gamma$**, and U and F are orthogonal matrices.**
- **Regressing Y on U we find** $\hat{\beta} = U^T Y$ **. Then the matrix of coefficients can be calculated as follows:**

$$\hat{\Gamma} = FL^{-1}\hat{\beta}$$

3. Theory and implementation

3.2. Substitution of leading PCs.

By using Principal Component Analysis the datasets can be writen as follows:

$$y(t,r) = \sum_{j=1}^{p} a_j(t) \cdot e_j(r)$$

$$x(t,r) = \sum_{j=1}^{p} b_j(t) \cdot f_j(r)$$

Where a, b are the PCs and e, f are the empirical orthogonal functions (EOFs).

• By substituting the leading PCs calculated from altimetry for the ones computed from tide gauges the field can be reconstructed:

$$\hat{y}(t,r) = b_1(t) \cdot e_1(r) + b_2(t) \cdot e_2(r)$$

4. Comparison between both methodologies

4.1. Prediction skill over the whole domain.

The altimetry dataset is broken into 2 subsets

Altimetry data in the period 1993-1997.
It will be used as the training period in the regression model and to calculate the EOFs.

Altimetry data in the period 1998-2000.

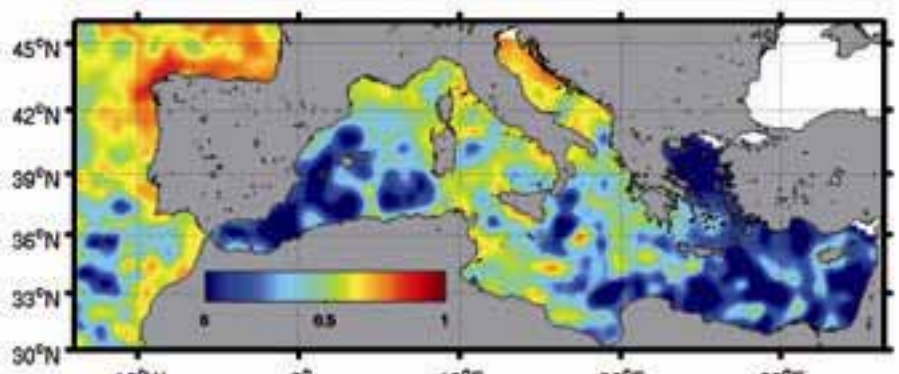
It will be used to test the goodness of the prediction all over the domain.

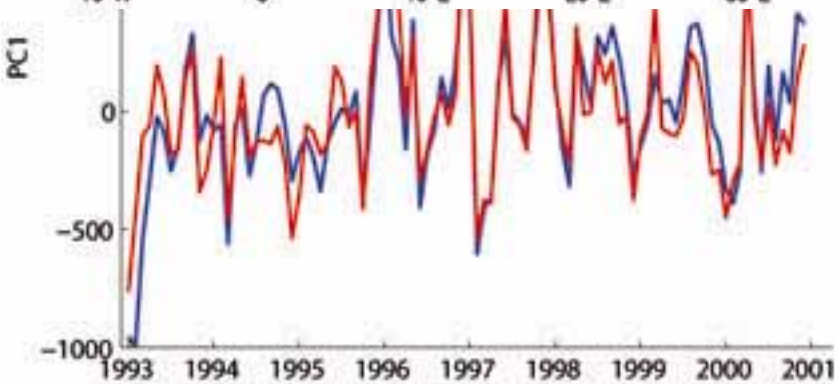
- **This test will be carried out by using all of the 23 tide gauges, either for the regression method and the substitution method.**
- **The period 1993-1997 will be used to train the models and afterwards the field will be reconstructed in the period 1998-2000 and compared with the altimetry observations.**

4. Comparison between both methodologies

Before carrying out the reconstruction we compared the leading PCs of both datasets.

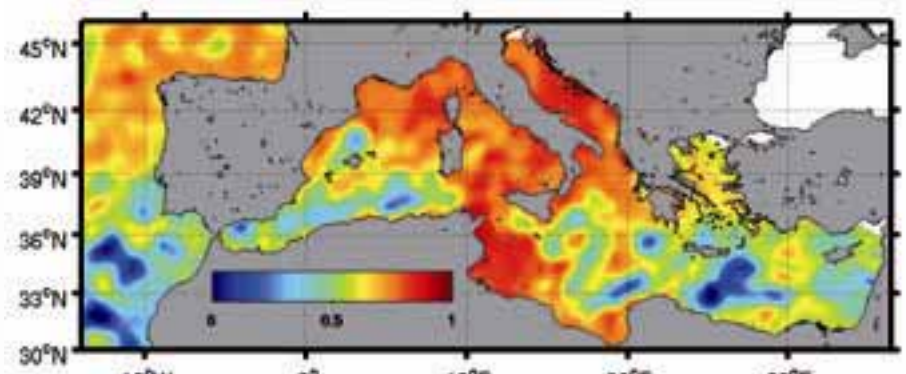
Map of the correlation between observations and reconstruction for the case of the regression.

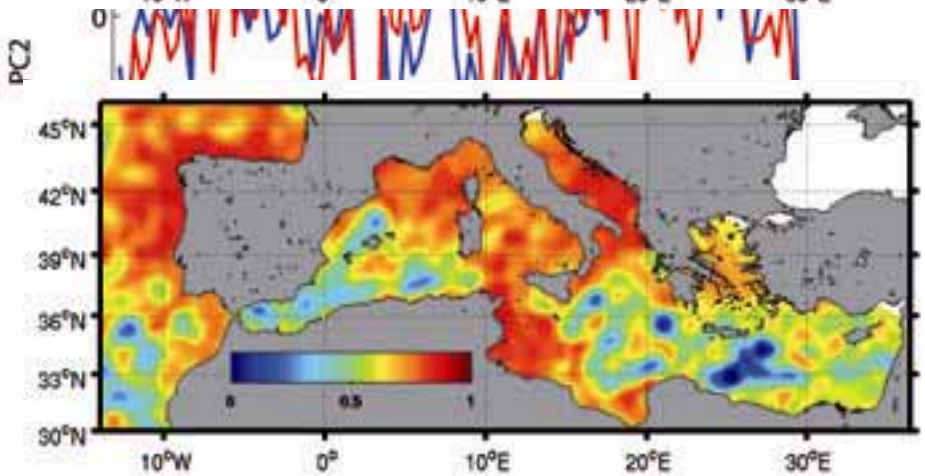




Comparison between the PC 1 from altimetry and tide gauge datasets. Correlation=0.84

Map of the correlation between observations and reconstruction for the case of the substitution method when using only the first PC.





Map of the correlation between observations and reconstruction for the case of the substitution method when using the first 2 PCs.

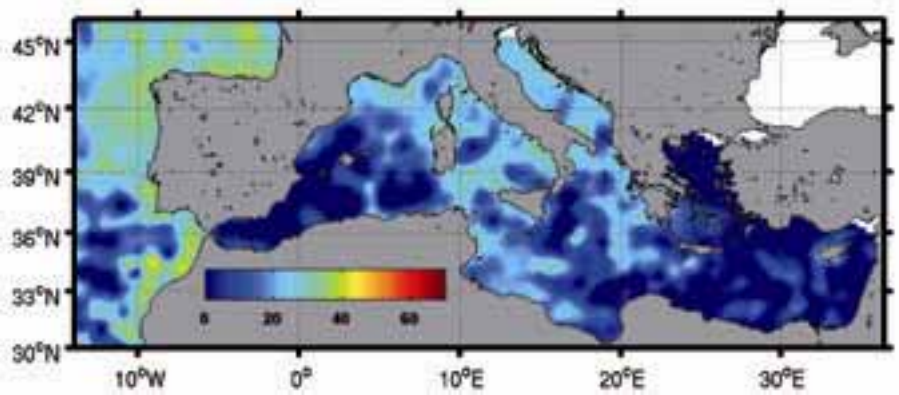
4. Comparison between both methodologies

	Regression	1 EOF	2 EOF
Correlation in the Mediterranean	0.35	0.64	0.65
Correlation in the Atlantic	0.57	0.60	0.66

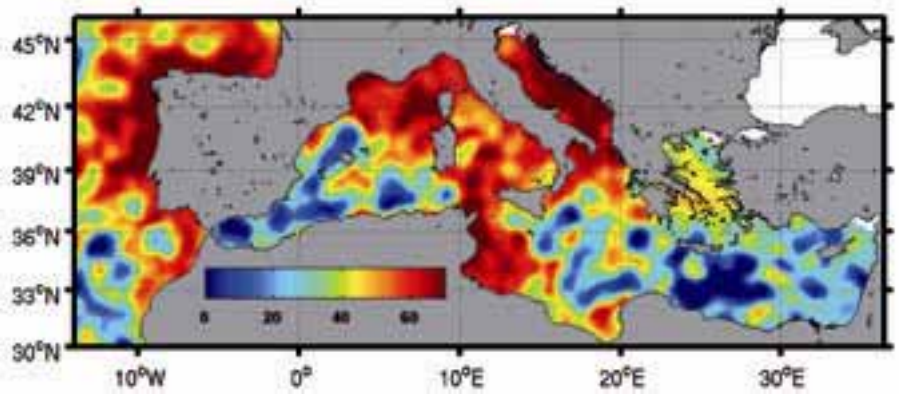
- The substitution method gives the best reconstruction.
- Using 2 PCs gives the best reconstruction, especially in the Atlantic.
- Correlations less than 0.3 would not be significantly different from zero.

4. Comparison between both methodologies

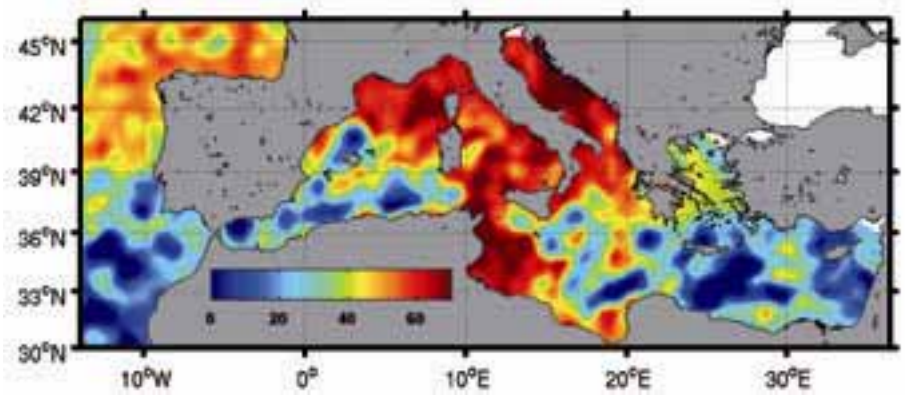
Map of the variance explained by the reconstruction for the case of the regression



Map of the variance explained by the reconstruction for the case of the substitution method when using the first 2 PCs.



Map of the variance explained by the reconstruction for the case of the substitution method when using only the first PCs.



- **The substitution method gives the best reconstruction.**
- **Using 2 PCs gives the best reconstruction, especially in the Atlantic.**

4. Comparison between both methodologies

4.2. Prediction skill for past times.

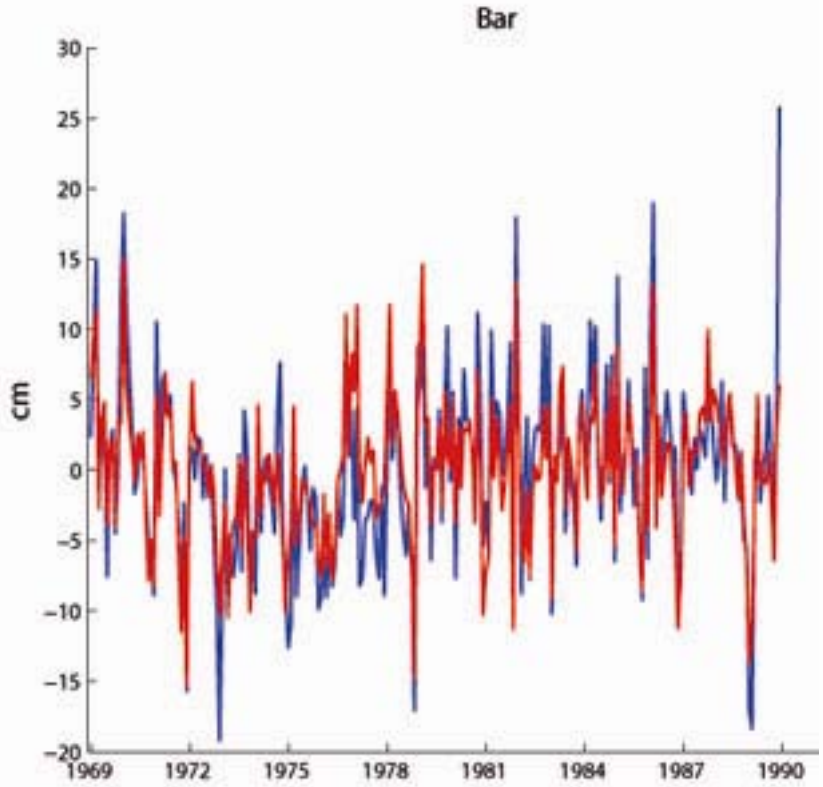
- **In this test we want to check the reconstruction beyond the period covered by altimetry.**
- **The reconstruction will be checked at tide gauge locations. The tide gauges used in this test are independent, they were not used in the reconstruction.**
- **The tide gauge locations are: Bar (42 05 N 19 05 E), S. Jean de Luz (43 24 N 01 41 W), Genova (44 24 N 08 54 E), Alicante (38 20 N 00 29 W).**
- **This test will be carried out by using all of the 23 tide gauges, either for the regression method and the substitution method.**
- **The period 1993-2000 will be used to train the models and afterwards the field will be reconstructed in the period 1969-2000.**
- **The period used for each station is: Bar (1969-1990), S. Jean de Luz (1969-1992), Genova (1977-1992) and Alicante (1969-1992).**

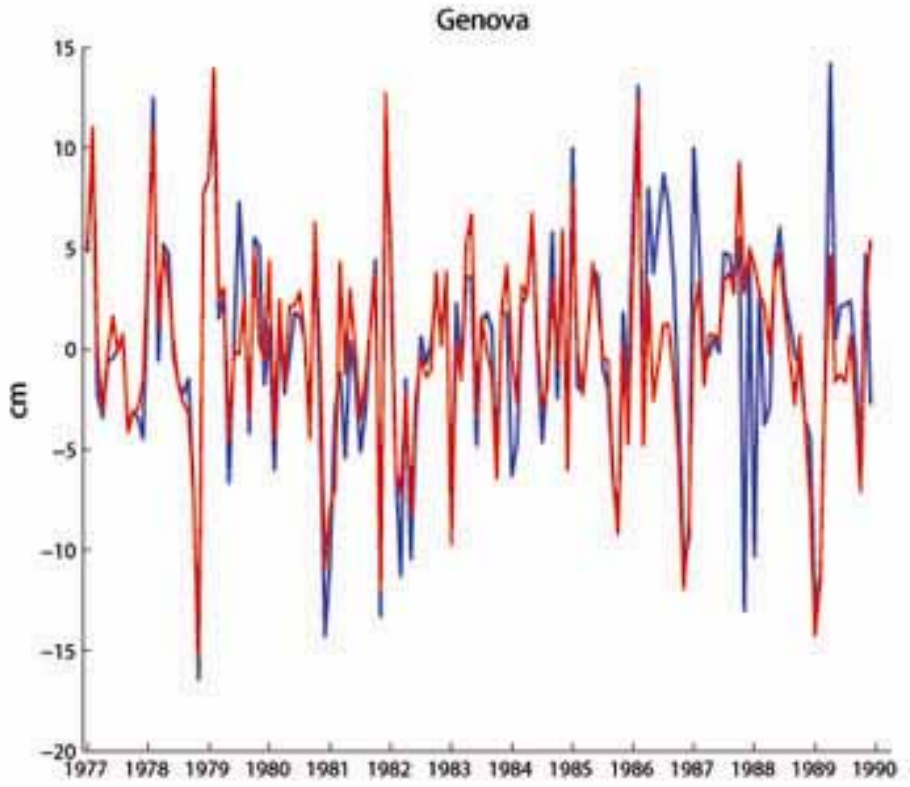
4. Comparison between both methodologies

	Correlation			Variance explained (%)		
	Regression	1 PC	2 PCs	Regression	1 PC	2 PCs
Bar	0.88	0.91	0.91	71	72	72
S. Jean de Luz	0.64	0.68	0.72	34	37	45
Genova	0.73	0.77	0.75	50	57	56
Alicante	0.59	0.65	0.69	28	37	44

- The best results are given by the substitution method, particularly when 2 PCs are used.
- The prediction given by the regression is better than appeared to be in the last section. That is because the training period is larger and therefore overfitting is minimized.
- Correlations less than 0.17 would not be significantly different from zero.

4. Comparison between both methodologies





- **In these plots, the blue curve represents the observation at the tide gauge location and the red one represents the prediction.**

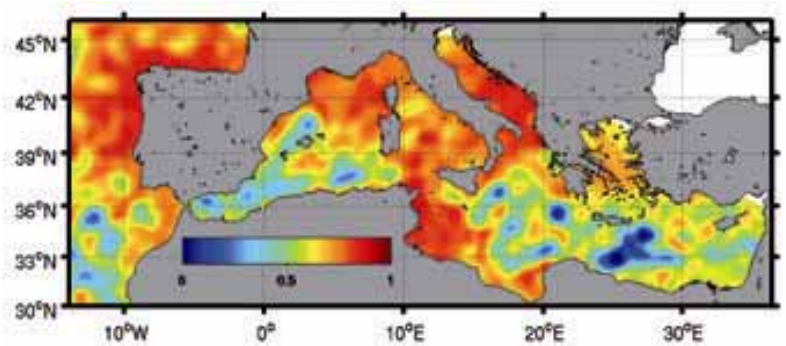
- **The correlation between the observation and the prediction remains constant for the whole period.**

5. Sensitivity

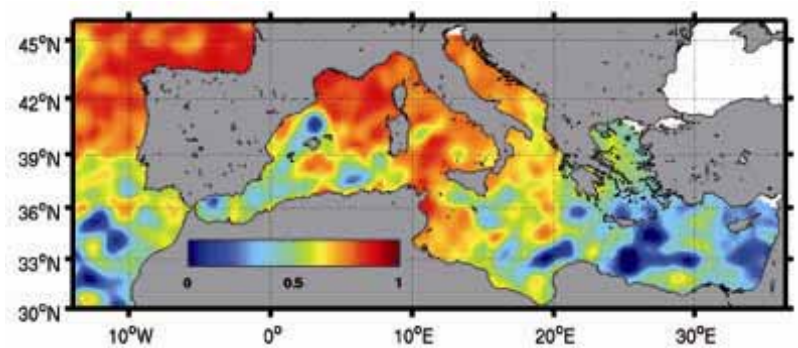
5.1. Substitution method.

• To check the sensitivity of the substitution method, the same tests as before are carried out for the case of using the 7 tide gauges that span the period 1950-2000.

Prediction skill all over the domain (1998-2000)



23 tide gauges



7 tide gauges

• There are no significant differences between both analysis. The method seems not to be much sensitive to the number of tide gauges used.

5. Sensitivity

Prediction skill for past times (1950-2000)

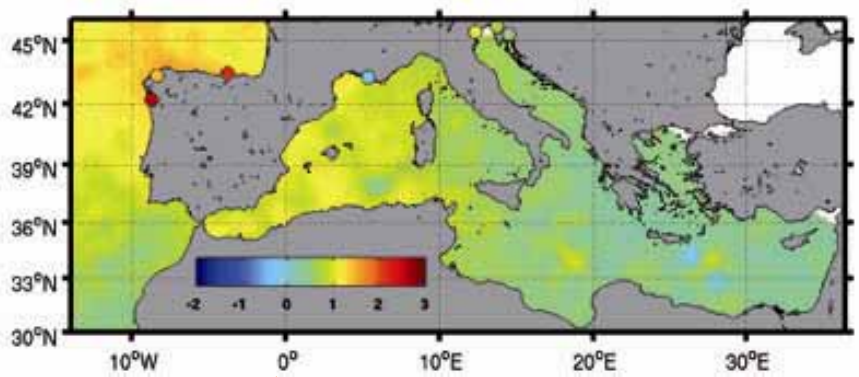
• The table shows the results of the reconstruction when using 7 tide gauges in the substitution method.

	Correlacion		Variance explained (%)	
	1 PC	2 PCs	1 PC	2 PCs
Bar	0.82	0.88	70	72
S. Jean de Luz	0.77	0.76	56	56
Genova	0.83	0.84	65	66
Alicante	0.72	0.74	47	50
Dubrovnik (1956-2000)	0.84	0.91	66	77

• Again the results for this test shows that the substitution method is not very sensitive to the number of tide gauges used.

6. The reconstruction: trends

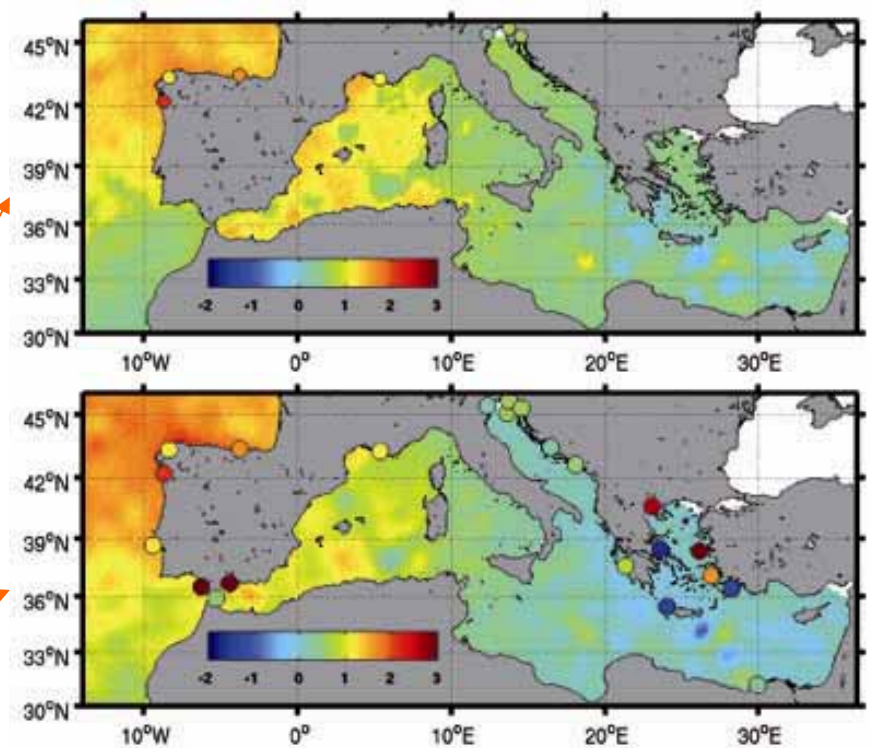
- The trends for the period 1950-2000 are calculated from the reconstruction.
- The trends for the period 1969-2000 are also calculated in order to find out whether the trends depend on the tide gauges used for the reconstruction. That will be done by comparing the trends calculated from the reconstruction using 23 (1969-2000) tide gauges and the one using 7 tide gauges (1950-2000).
- The trends calculated from tide gauge records are also shown on the map.



1950-2000

1969-2000 (7 tide gauges)

1969-2000 (23 tide gauges)



5. Conclusions

- **The method of the substitution gives better results than the regression.**
- **Using 2 EOFs gives better results than using only the first one.**
- **Correlations higher than 0.7 and variances explained higher than 60% are found in the Northeastern Mediterranean, the Adriatic Sea and in the Atlantic.**
- **The substitution method turns out not to be much dependent on the number of tide gauges used, while the regression method is rather sensitive.**
- **The reconstruction remains stable for the whole period of reconstruction.**
- **A map of trends for the period 1950-2000 shows higher trends in the Atlantic than in the Mediterranean and whithin the Mediterranean, trends are higher in the Western region. This is in agreement with the observed increase of the atmospheric pressure (more markes to the east).**

- **Future work: comparison with models**

THE END

THANK YOU FOR YOUR ATTENTION