MISSING DATA RECONSTRUCTION OF A SST CLOUDY DATA SET OF THE ADRIATIC SEA USING EMPIRICAL ORTHOGONAL FUNCTIONS.

Alvera-Azcárate A. [1]*, Barth A. [1], Rixen M. [2] and Beckers J.M [1]

[1] G.H.E.R. University of Liège, Belgium - * A.AlveraAzcarate@ulg.ac.be

[2] SACLANT Undersea Research Centre, Italy

Abstract

A method for the reconstruction of missing data in oceanographic data sets is presented. The method, DINEOF (Data INterpolating Empirical Orthogonal Functions) is based on an EOF decomposition. It has been applied to a realistic case, a data set on the Adriatic Sea. In order to optimize the computational time, a Lanczos method has been used for the EOF decomposition. The results are accurate, as can be seen from different tests realized with the data. The error of the reconstruction in relation with the original data is of about 0.6°C, as depicted for a cross validation analysis. The temperature distribution is reliable, and realistic physical features are obtained.

Keywords: missing data, EOF decomposition, Adriatic Sea

Introduction

DINEOF (Data INterpolating Empirical Orthogonal Functions) is a method for the reconstruction of gappy data. When dealing with satellite images there are often gaps in the data, due to cloud coverage, technical malfunctions, noise and so on. Also when dealing with other oceanographic data sets there can be missing data due to some of these reasons. For many applications using these data sets, a complete set is necessary, with no missing data. Examples are an EOF analysis, a wavelet analysis, or feature tracking in the ocean. For recovering these data, DINEOF has been applied. This method has the advantage of no needing any *a priori* information, so no additional calibration is needed.

How the method works

DINEOF has been presented by Beckers [1], and it is based on an EOF decomposition. The mean value of the data is subtracted from the set and the missing values are initialized to zero. Then the EOF decomposition is done, with only one EOF, and a new value for the missing data is calculated from a truncated series of the EOF obtained. These two steps are repeated until convergence of the value for the missing data. Now the number of EOFs requested is increased to two, and the procedure is repeated. Finally, we have a series of missing data obtained with $1 \ldots k$ EOFs, and the optimal number of EOFs for the reconstruction of the data set must be calculated. This is done by a cross validation technique [1, 2]: a random data set is set aside from the initial valid data, and they are considered as missing. The optimal number of EOFs minimizes the error between the data set aside and the values obtained at these points with the reconstruction method.

The EOF decomposition itself has been carried out with a Lanczos method, presented by Toumazou [3]. This method allows to calculate a given number of EOFs in a small cpu time, so the computational time required by DINEOF has been optimized.

An application to the Adriatic Sea

The authors have applied the method DINEOF to a test case in the Adriatic Sea. A total of 105 Sea Surface Temperature (SST) AVHRR images have been treated, ranging from 09 May 1995 to 22 October 1995. The mean cloud coverage of this data set is 52%. In the reconstruction of this data set DINEOF keeps 10 EOFs as the optimal number for the reconstruction of the missing data. The cross validation gives an expected error of 0.6°C for this reconstruction. In Figure 1 one can see the reconstruction of one of the images, corresponding to September 3, 1995. Figure 1a is the original image, with blanks where there are initially no data. Figure 1b is the reconstruction of this image. As can be seen, the reconstruction gives realistic results. We can appreciate a cold filament detaching from the east coast. The signal of this kind of filaments has been studied by, e.g. [4].

Other tests have been carried out with this data set [5]. For establishing the capacity of DINEOF to reconstruct data sets with different amounts of cloud coverage, a subset of 15 images with a mean cloud coverage of 18% is used. Then, extra cloud coverage has been added, up to 40%, 60% and 80%. The reconstruction can be thus compared to the original data. The Root Mean Square (RMS) error obtained between the reconstruction of the 40%, 60% and 80% extra cloud coverage sets and the original data is of 0.89°C, 0.78°C and 1.25°C respectively. Also a validation with *in situ* data has been made. Data from the MEDAR/Medatlas database [6] have been extracted. The error between the reconstruction and those data is of 0.95°C. All the results and tests presented here are available at http://modb.oce.ulg.ac.be/alvera.

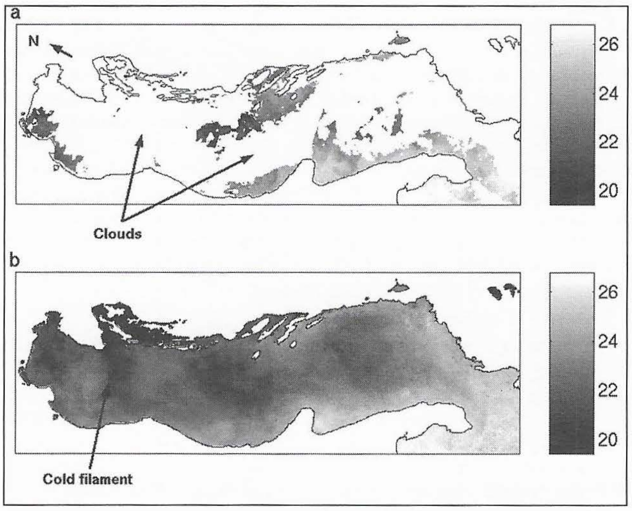

Fig. 1. Reconstruction of the 3 September 1995. (a) is the original image with blanks where there is no data. (b) is the reconstruction. We can see a cold filament detaching from the east coast and the warm temperatures at the south Adriatic.

References

1 - Beckers, J.M. and Rixen, M. 2003. EOF calculations and data filling from incomplete oceanographic data sets. Journal of Atmospheric and Oceanic Technology. In press.

2 - Brankart, J.M. and Brasseur, P. 1996. Optimal analysis of in situ data in the Western Mediterranean using statistics and cross validation. *Journal of Atmospheric and Oceanic Technology,* 16:477-491.

3 - Toumazou, V. and Cretaux, J.F. 2001. Using a Lanczos eigensolver in the computation of Empirical Orthogonal Functions. *Monthly Weather Review.* 125,5:1243-1250.

4 - Borzelli, G., Manzanella, G. Marullo, S. and Santoleri, R. 1999. Observations of coastal filaments in the Adriatic Sea. *Journal of Marine Systems.* 20:187-203.

5 - Alvera-Azcárate, A., Barth, A. Rixen, M. and Beckers, J.M. 2003. Reconstruction of incomplete oceanographic data sets using Empirical Orthogonal Functions. Application to the Adriatic Sea. Submitted to *Journal of Physical Oceanography.*

6 - MEDAR Group. 2002. Medatlas/2002 database. Mediterranean and Blach Sea database of temperature, salinity and bio-chemical parameters. Climatological Atlas. Ifremer Edition, 4 CD-ROM.