



Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature

A. Alvera-Azcárate ^{a,*}, A. Barth ^a, M. Rixen ^b, J.M. Beckers ^a

^a GHER, Department of AGO, University of Liège, Allée du 6 Août 17, B5, Sart Tilman, 4000 Liège, Belgium

^b NATO/ISACLANT Undersea Research Centre, Viale San Bartolomeo 400, 19138, La Spezia, Italy

Received 30 March 2004; received in revised form 26 July 2004; accepted 4 August 2004

Available online 16 September 2004

Abstract

A method for the reconstruction of missing data based on an EOF decomposition has been applied to a large data set, a test case of Sea Surface Temperature satellite images of the Adriatic Sea. The EOF decomposition is realised with a Lanczos method, which allows optimising computational time for large matrices. The results show that the reconstruction method leads to accurate reconstructions as well as a low cpu time when dealing with realistic cases. The method has been tested with different amounts of missing data, artificially adding clouds ranging from 40% to 80% of data loss, and then compared to the same data set with no missing data. A comparison with *in situ* data has also been made. These validation studies show that results are robust, even when the amount of missing data is very high. The reconstruction of the data from the Adriatic Sea shows realistic features and a reliable temperature distribution. In addition, the method is compared to an Optimal Interpolation reconstruction. The results obtained with both methods are very similar. The main difference is the computational time, which is reduced nearly 30 times with the method presented here. Once the reconstruction has been performed, the EOF decomposition is analysed to show

* Corresponding author. Address: GHER, Department of AGO, University of Liège, Allée du 6 Août 17, B5, Sart Tilman, 4000 Liège, Belgium. Tel.: +32 4 3662340.

E-mail address: a.alveraazcarate@ulg.ac.be (A. Alvera-Azcárate).

URL: <http://modb.oce.ulg.ac.be/alvera>

the method's reliability, and a cold event on the Albanian coast is studied. The reconstructed data reflect the effect of wind on the Albanian coast, that led to a cold-water episode in this zone for a 6-day period.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: EOF decomposition; Missing data; SST and *in situ* validation; Adriatic Sea

1. Introduction

Satellite images are commonly used in oceanography. They allow realising general studies of the sea surface characteristics (e.g. Gacić et al., 1997; Borzelli et al., 1999) as well as studies at depth that present a signal at the surface, such as internal waves (e.g. da Silva et al., 1998; Liu et al., 1998). Satellite images are widely used due to their extensive coverage, in time and space. No other data acquisition method gives the coverage and precision of satellite images as quickly. Obtaining observational data in this way is therefore very useful in operational oceanography and in near real time studies.

There are several kinds of data measured by satellites, depending on the type of sensors used: Sea Surface Temperature (SST), chlorophyll or Sea Surface Height (SSH). The receptors working in the visible and infrared ranges present however one disadvantage: the clouds present in the atmosphere can totally or partially cover the studied area. The loss of data due to clouds can reach a high percentage in some periods. There can also be noise, or malfunctions in the satellite that reduce its coverage. A complete data set is crucial for many applications using satellite images, as in Empirical Orthogonal Function (EOF) analysis, when tracking features in the ocean or in the study of zones with high spatial and temporal variability. In addition, complete fields of sea surface temperature are used to force meteorological models. For other studies where it might not be crucial to have complete satellite coverage, it is always preferable to have a maximum amount of information.

Several methods have been used in the past when dealing with recovery of missing data. Spline interpolation (Emery and Thomson, 1998) has been used by, e.g., Everson et al. (1997). Inverse methods, such as optimal interpolation (OI) (Bennett, 2002), have also been widely used in the reconstruction of SST data sets (e.g. Chu et al., 1997; Fieguth et al., 1998; Houseago-Stokes, 2000; He et al., 2003), as well as in the reconstruction of Sea Level Anomaly (SLA) data sets (e.g. Fieguth et al., 1998; Le Traon et al., 1998; Le Traon and Didarboure, 1999; Le Traon et al., 2001). Many studies have reconstructed historical SST and *in situ* data sets (see e.g. Reynolds and Smith, 1994; Smith et al., 1996; Kaplan et al., 1997; Kaplan et al., 1998) using OI. Reynolds and Smith (1994) used OI to realise a global SST analysis. In Smith et al. (1996), the OI analyses are used to compute the most dominant EOFs. These EOFs are then fitted to *in situ* data to compute the reconstructed SST set. The work by Kaplan et al. (1997) first calculates the EOFs from the covariance matrix, and then a least-squares fit is done to estimate the best reconstruction of a 136-year SST set on the Atlantic Ocean. In Kaplan et al. (1998) the same technique is applied, and three different fitting methods are compared. One disadvantage of the optimal interpolation as a method for reconstructing missing data is the necessity for a priori information about the error statistics of the data, generally poorly known (Bennett, 2002). The computational cost of these techniques may be prohibitive when using large matrices (Kaplan et al., 1997).

The alternative methodology presented by Beckers and Rixen (2003) is a self-consistent, parameter-free technique for the reconstruction of gappy data that presents the advantage of not needing this kind of a priori information. The method allows calculating the missing data from an optimal number of EOFs determined by a cross-validation technique (e.g. Wilks, 1995; Brankart and Brasseur, 1996; von Storch and Zwiers, 1999). This cross-validation also gives an error estimate of the filled data. The method is based on the fact that an EOF analysis aims to extract a small number of significant degrees of freedom, present in the physical system, from a large data set. These reduced variables should represent a large fraction of the original variability of the data set (e.g. Wilks, 1995). The combination of the dominant EOFs and their amplitudes can therefore help recover missing data values.

The aim of this work is the application of the reconstruction method presented by Beckers and Rixen (2003), hereafter called DINEOF (Data INterpolating Empirical Orthogonal Functions) to a realistic case: a data set covering a whole subbasin at high resolution. The data chosen for this purpose is a series of Advanced Very High Resolution Radiometer (AVHRR) images covering the entire Adriatic Sea for a six-month period.

The paper is organised as follows: in Section 2, a description of the main characteristics of the Adriatic Sea is presented, followed by a review of applications of satellite images in this zone. The data set used in this work is described in Section 3. In Section 4 the method used for the reconstruction of missing data is briefly presented, as well as some estimates of the performance of the code and computational time. Reconstruction results of the Adriatic images are shown in Section 5, with additional validation studies, apart from the cross-validation itself. A comparison with a classical OI reconstruction is made in Section 6. To show the robustness of the results obtained, the EOF decomposition of the reconstructed data is presented in Section 7. Finally, we present our conclusions.

2. The Adriatic Sea

The Adriatic Sea is a subbasin of the Mediterranean Sea (see Fig. 1) of about 800 by 200 km, connected to the Ionian Sea by the Strait of Otranto. The northern and central parts of the basin are very shallow, with maximum depths of 270 m, and a mean depth of 35 m (Artefiani et al., 1997a). The southern part is deeper, reaching 1200 m, but nearby, in the Strait of Otranto, the depth decreases again, to 780 m (e.g. Cushman-Roisin et al., 2001). The circulation in the basin is cyclonic, with water coming from the Ionian basin and entering the Strait of Otranto to the east of the Adriatic Sea (Eastern Adriatic Current, EAC). This current is wide and weak, and brings, at depth, warm and salty modified Levantine Intermediate Water to the northern basin. A western current flowing southward closes the cyclonic circulation. This current is called WAC (Western Adriatic Current), and is thinner than the EAC. In winter, it brings at surface cold and fresh water to the southern basin. Salty and cold water coming with the WAC fills the depth at this southern part. Embedded in this main circulation path are three sub-basin cyclonic gyres, in the northern part of the Adriatic, between the Istrian Peninsula and the Jakuba Pit, and in the southern Adriatic Sea, respectively (e.g. Orlić et al., 1992; Poulain, 2001). The basin-wide cyclonic surface circulation is mainly produced by winds. There are two dominant winds: the steady southeasterly wind, called Sirocco, which enters the basin by the Strait of Otranto, and the northeasterly wind

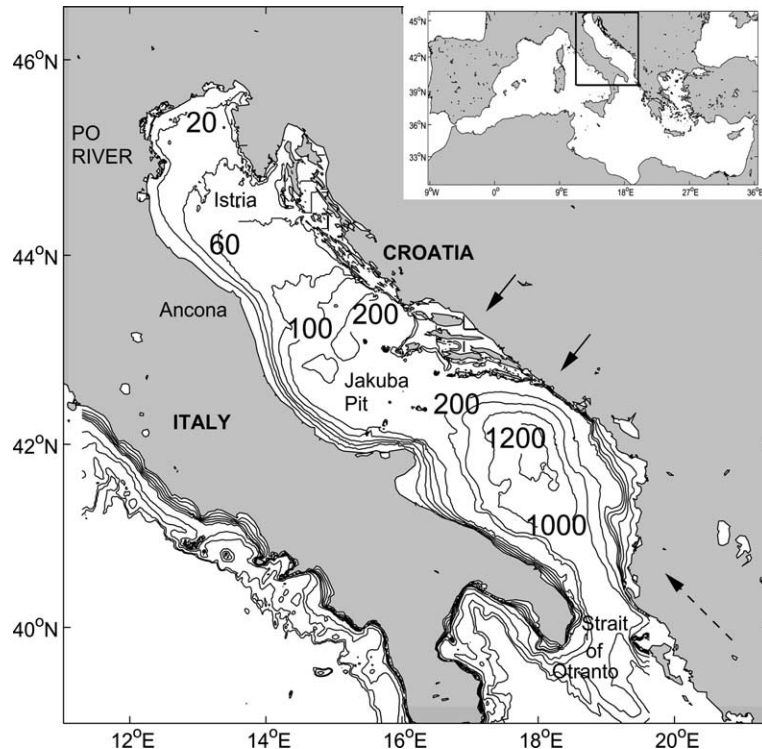


Fig. 1. The Adriatic Sea and its bathymetry (depth in meters). The arrows show the typical winds found in this area. In solid line, the Bora, and in dashed line, the Sirocco.

called Bora, which enters the Adriatic Sea from the Northeast (Bergamasco et al., 1999; Poulain, 1999). In Fig. 1 one can see the general direction of these winds. The fresh water input in the Adriatic Sea is mainly due to rivers in the northern and eastern coasts. The Po River accounts for 28% of this runoff (Cushman-Roisin et al., 2001). The mean surface temperature in winter is 14°C. In summer, the southern part is warmer (24°C) than the northern part (22°C) (values obtained from the MEDAR/Medatlas climatology, MEDAR-Group (2002)). In the zonal direction, the water is warmer in the open sea than in the eastern coastal waters. A coastal front can be found throughout the year, but its position varies between seasons. For a detailed description of the Adriatic's characteristics see e.g., Orlić et al. (1992), Artegiani et al. (1997a), Artegiani et al. (1997b), Cushman-Roisin et al. (2001).

2.1. Satellite observations in the Adriatic Sea

Remote sensed data has been used to study the Adriatic Sea in different works (e.g. Gacić et al., 1997; Borzelli et al., 1999; Cushman-Roisin et al., 2001; Mauri and Poulain, 2001; Alvarez, 2003). A large number of features can be identified from the SST field. For example, the signature of the Western Coastal Layer (WCL), formed to the south of the Po delta and related to the WAC. Filaments advecting cold waters from the eastern coast (detaching at the Istrian Peninsula and at southern positions along the Croatian coast, between 43°N and 44°N) to open seawaters have

also been detected from satellite images (Borzelli et al., 1999). These filaments can extend up to a hundred of kilometres, with a width of about 10–20 km, and have been observed in summer. Mauri and Poulain (2001) have also observed cold waters (21–23 °C) in the nearby the Croatian coast in September and October. The signal of the Po River can be identified from satellite images, as it can spread over the entire northern basin in the summer season. The Po River water is slightly warmer than seawater in summer, so other variables, such as chlorophyll, act as better tracers of the Po plume than temperature. In winter the plume is weaker, the water flows mainly southward and the Po River water is several degrees colder than seawater. The work by Gacíc et al. (1997) presents a complete analysis of the seasonal and interannual variability of the surface temperature of the Adriatic Sea. Studies of this kind could benefit from any method recovering missing information due to clouds.

3. Data Set

A set of 135 Advanced Very High Resolution Radiometer (AVHRR) images of the Adriatic Sea obtained from <http://radlab.soest.hawaii.edu/avhrr/adria/cdrom/html/> (Dousset et al., 1998) is used to test the DINEOF method. The images range from 9 May 1995 to 22 October 1995. Their size is 248×709 pixels, with a resolution of 1.25 km. Land points are not used in this method, so the final spatial size is 94 755 (out of 175 832). In order to minimise skin temperature effects due to diurnal heating, only nighttime images are used. The mean cloud coverage is 61%. Some of these images present extreme cloud coverage (more than 95%) and have been eliminated in order to obtain reliable results. Images containing less than 5% of data do not provide useful information, and might affect the final result. A subset of 105 images (hereafter called the ‘Complete Set’ for clarity) with less than 95% of cloud coverage is kept for the analysis. The mean cloud coverage

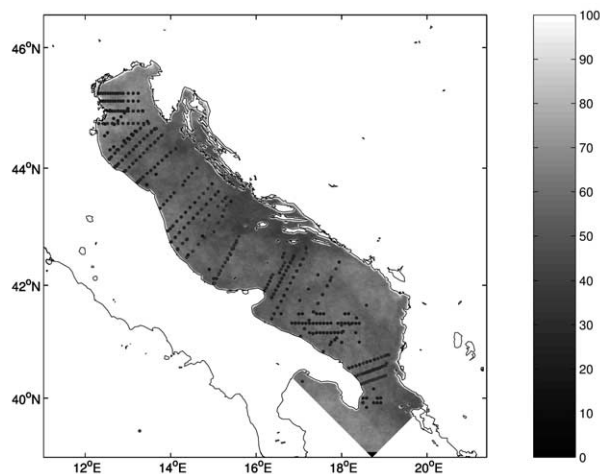


Fig. 2. Mean percentage of cloud coverage for the Complete Set of images. The points show the distribution of the *in situ* data obtained from the MEDAR/Medatlas database used for validation of the reconstruction. A total of 452 stations were extracted from this database.

of this data set is 52%. In Fig. 2 we can see that the distribution of the cloud coverage in the Complete Set is very homogeneous, with a slightly higher cloud concentration in the northern Adriatic and in the southwestern part of the basin. The coastline has extreme cloud coverage, probably due to slight errors that occurred when treating the images at these zones. This point highlights the importance of a correct image processing to avoid high data loss. Moreover, clouds in the satellite images must be precisely identified to obtain correct results in the reconstruction. Otherwise, artificial features may appear in the results. This is why a good cloud detection algorithm must be used. For the Adriatic data set, three cloud detection tests were conducted with nighttime images (see <http://radlab.soest.hawaii.edu/avhrr/adria/cdrom/html/index.htm> for detailed information on the cloud detection algorithm). These tests aimed to (1) detect temperature variations due to convective clouds (clear water pixels present very uniform brightness temperature, in contrast to clouded pixels); (2) compare to MODB climatology (Brasseur et al., 1996; Rixen et al., 2001); the maxima and minima of this climatology were increased and lowered by 1 °C, and pixels falling outside this range were eliminated; and (3) detect stratiform clouds (based on differences between the infrared channels). These tests followed the work by Saunders and Kriebel (1988). However, as we will show, the reconstruction results in a filtering of the original data, which can correct some of the artificial features that were not detected by these tests (see for example, Fig. 6).

4. The method

Beckers and Rixen (2003) have presented a self-consistent method for the reconstruction of missing data in oceanographic data sets, DINEOF. Consider that \mathbf{X} is the initial matrix of dimensions $m \times n$, $m > n$ (with m the spatial dimension and n the temporal dimension), containing the observations. It may also contain some unknown values corresponding to the missing data. For the reconstruction of these data, a Singular Value Decomposition (SVD) technique is used to compute the EOFs of the matrix, in which a first guess has been introduced for the missing data. The equation:

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (1)$$

allows calculating the spatial EOFs, \mathbf{U} , with dimension $m \times r$, the temporal EOFs, \mathbf{V} , with dimension $n \times r$, and their singular values \mathbf{S} , with dimension $r \times r$. The value r is the rank of the matrix, with $r \leq \min(m, n)$. Only the most significant spatial and temporal EOFs are necessary for the reconstruction method. The k largest singular values and singular vectors can also be calculated by eigenvector decomposition:

$$\mathbf{X}\mathbf{X}^T\mathbf{u}_i = \rho_i^2\mathbf{u}_i \quad (2)$$

with \mathbf{u}_i the i th column of \mathbf{U} and ρ_i the corresponding singular value, $i = 1, \dots, k$. To avoid using Eq. (2), which implies the $m \times m$ matrix $\mathbf{X}\mathbf{X}^T$, we can rather use:

$$\mathbf{A}\mathbf{v}_i = \rho_i^2\mathbf{v}_i \quad (3)$$

$$\mathbf{u}_i = \frac{\mathbf{X}\mathbf{v}_i}{\rho_i} \quad (4)$$

where $\mathbf{A} = \mathbf{X}^T\mathbf{X}$ is a real symmetric $n \times n$ matrix.

DINEOF can be explained as follows:

- The average value of the matrix is subtracted once for the entire procedure and the missing data points are initialised to zero in order to have an unbiased first guess. The missing data, however, are ‘flagged’ to differentiate them from those existing points on the mean. This demeaned matrix is used throughout the whole procedure.
- Two steps are then repeated for a given k until convergence:
 - An EOF decomposition of the matrix is realised, with only the first k EOFs, to obtain a first estimate of the singular values and singular vectors.
 - The elements $\mathbf{X}_{i,j}$ corresponding to the flagged missing data are now replaced by the value obtained with the EOF series:

$$\mathbf{X}_{i,j} = \sum_{p=1}^k \rho_p(\mathbf{u}_p)_i (\mathbf{v}_p^T)_j \quad (5)$$

An improved guess has thus been introduced for the missing data, so we recompute the EOFs and obtain a new value for the missing data.

- Once the convergence is reached, the number of computed EOFs is increased, from $k = 1, \dots, k_{\max}$, so at the end we have k_{\max} estimates for the missing data reconstructed with $1, 2, \dots, k_{\max}$ EOFs. But which estimate is the best? The answer is obtained by cross-validation.
- We calculate the optimal number of EOFs from the series of k_{\max} EOFs. To do so, a random set of data is initially set aside from the valid data to apply a cross-validation technique, as described in, e.g., Wilks (1995, Chapter 6), Brankart and Brasseur (1996), von Storch and Zwiers (1999, Chapter 18), Beckers and Rixen (2003). This data set has a size of $\min(0.01 \times m \times n + 40, 0.03 \times m \times n)$, and for this particular case, 99532 points are retained for the cross-validation. The optimal number of EOFs is the one that minimises the error between the data set aside and the values obtained at these points with the reconstruction method.
- Once the optimal number N of EOFs is known, the whole procedure is repeated, now including the data set aside for cross-validation, but only with the N first EOFs considered as optimal. Final values for the missing data are then computed.

This is a general description of how the method works. For a more detailed description, see Beckers and Rixen (2003).

In the present work, we apply the reconstruction method for missing data DINEOF to a large matrix. We use a Lanczos method (see, e.g. Chatelin, 1993; Toumazou and Cretaux, 2001) for the EOF decomposition phase to make the application of DINEOF effective when working with large matrices. The desired characteristic of the EOF decomposition algorithm is the possibility to compute only the k largest EOFs at a small computational cost, since it must be used several times during the DINEOF iterations.

Toumazou and Cretaux (2001) have compared three different EOF decomposition methods, one based on the SVD algorithm, and two that express Eq. (1) as an eigenvalue problem: the QR strategy and the Lanczos method. They have shown that a Krylov-type method, called the Lanczos method, is a good choice when using large matrices. The EOF analysis performed by

the mentioned methods gives similar results, although the Lanczos method requires half the storage memory than the others, and up to 22 times less computational time for large matrices. Another attractive characteristic of the Lanczos method is that it does not need to compute all the singular values, only the k largest ones. For these reasons, the Lanczos solver provided by Toumazou and Cretaux (2001), which uses the ARPACK freeware (Lehoucq et al., 1997), has been implemented in DINEOF.

The main characteristic that makes the Lanczos method suitable when dealing with a large matrix is that, instead of working with the $n \times n$ matrix \mathbf{A} of Eq. (3), a $p \times p$ ($p \ll n$) tridiagonal matrix is used, obtained by the projection of \mathbf{A} onto the sub-space Krylov $K_p(\mathbf{A}, q)$ (Chatelin, 1993). The eigenvalues are calculated in this reduced matrix, in an iterative way until a convergence criterion is satisfied. This stopping criterion is based on the backward error (Bennani and Braconnier, 1994). For a more detailed explanation of the Lanczos method, see, e.g., Chatelin (1993) and (Toumazou and Cretaux, 2001).

4.1. Code performance

A comparison between several codes that calculate EOF decomposition has been already made by Toumazou and Cretaux (2001) and is not the aim of this work. However, several tests were carried out to test the EOF reconstruction code performance. A Silicon Graphics MIPS R12000, 400MHz is used to run DINEOF. In Table 1 we show the computation time of the

Table 1
Time (in s) for EOF decomposition with the Lanczos method, and for the whole reconstruction process

m	n	Time (SVD decomposition)	Time (reconstruction)
94755	135	27.96	7148.72
94755	105	20.31	5389.77
94755	75	15.19	3992.92
94755	50	13.46	3589.61
94755	20	1.72	822.3
39616	135	15.94	3846.17
39616	105	10.53	2182.37
39616	75	8.11	1941.81
39616	50	3.5	1018.5
39616	20	0.62	307.83
23205	135	6.85	1597.34
23205	105	5.0	1026.96
23205	75	2.68	668.25
23205	50	1.91	385.32
23205	20	0.34	138.2
8281	135	1.39	371.7
8281	105	1.17	283.51
8281	75	0.84	209.42
8281	50	0.47	133.69
8281	20	0.12	49.12

The size of the matrix is $m \times n$, with m the spatial dimension and n the temporal dimension.

Lanczos method for EOF decomposition (i.e., Lanczos method called once), calculated with several subsets obtained from the initial data. For a matrix of dimensions $m \times n$, the cost of the Lanczos method is proportional to $m^{1.23}n^{1.5}$. The total computational time for DINEOF is also presented in Table 1, being the cost of the total reconstruction method $m^{1.25}n^{1.17}$, with, always, $m > n$. As can be seen, for the larger matrix studied (94755×135) a total time of 119 min is necessary, and for the smallest case (8281×20), the whole procedure is finished in less than a minute. In an attempt to further improve computational time, interesting for very large matrices reconstruction, we have introduced different convergence criteria in the calculation of the singular values, as well as different criteria for the iterations made when searching the convergence of the singular values. We tested different values for missing data initialisation, stopping criteria to find the optimal number of EOFs, and the control of the iterations made by the Lanczos method. Only slight changes in computational time were attained, accounting for up to 10% of saved time.

5. Results

The first test we conducted is the reconstruction of the Complete Set. DINEOF keeps 10 EOFs as optimal for reconstruction (number of EOFs that minimises the error in cross-validation), as can be seen in Fig. 3, which minimise the expected error to 0.6°C . These 10 EOFs explain 98.97% of the initial variance, which has been calculated as the variance of those points not covered by clouds.

In Figs. 4–6 we can see three examples of the quality of DINEOF results. They show three original images, with blanks where there are no data, and their reconstruction. Fig. 4 is of 23 July, and we can see the warm plume of the Po River, which reaches 32°C and spreads over a large surface in the northern Adriatic and flows southward following the west coast, forming the WCL. We can also see a weak cold filament going from the Istrian Peninsula (in the east coast) to open sea. Finally, at the southeast, on the Albanian coast, there is a strong cold signal. This kind of cooling

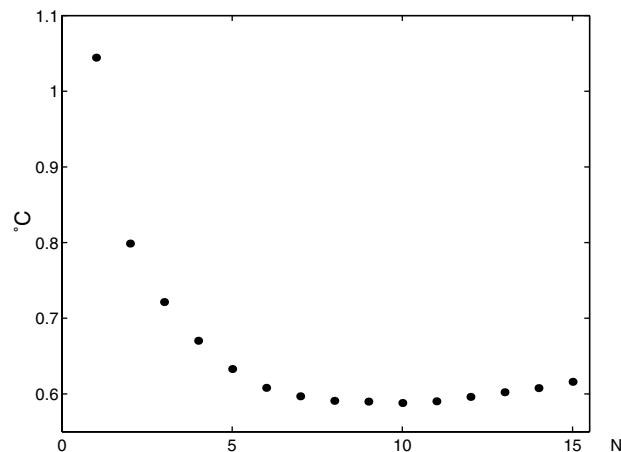


Fig. 3. Error obtained with cross-validation for reconstruction of the Complete Set. N is the number of EOFs.

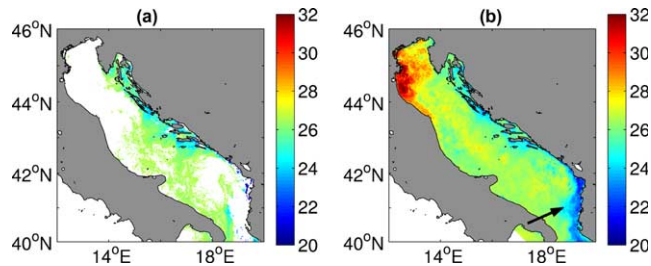


Fig. 4. Original cloudy image (a) from the Complete Set for 23 July and its reconstruction (b). A warm Po plume is clearly seen, as well as cold waters along the Albanian coast, marked with an arrow.

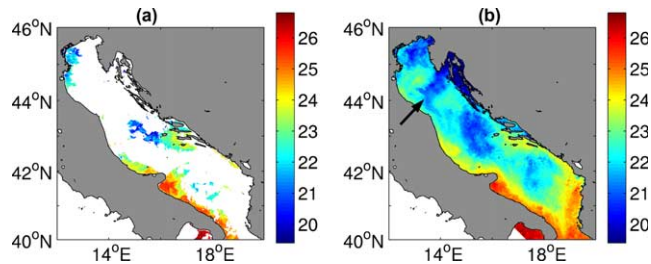


Fig. 5. Original cloudy image (a) from the Complete Set for 3 September and its reconstruction (b). The arrow shows a strong cold filament detaching from the Istrian Peninsula.

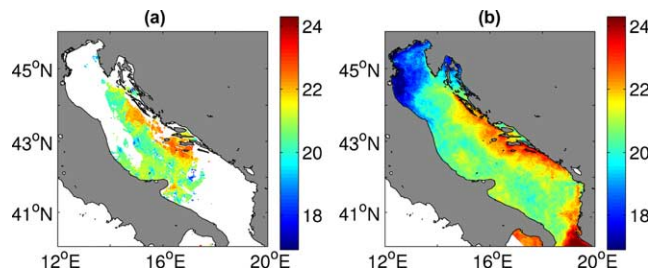


Fig. 6. Original cloudy image (a) from the Complete Set for 5 October and its reconstruction (b).

event has been reported by e.g. Bergamasco and Gacić (1996) and Gacić et al. (1997), and the cause may be the cooling and mixing due to the Bora wind that blows parallel to this coast. This image will be further commented later in this work. Fig. 5 corresponds to 3 September. In this image, a strong cold filament can be observed at the same location as in Fig. 4. This filament has a temperature of about 20°C, which corresponds to the situation described by Mauri and Poulain (2001) for September–October. In Fig. 6, of 5 October, a strong cooling event occurs in the northern Adriatic, where temperature decreases to 18°C. This situation is typical of autumn, when temperatures begin to decrease in this zone (Gacić et al., 1997). In this image, we clearly see the effect of filtering of the reconstruction method, due to the rejection of some EOFs by cross-validation: in the cloudy image, some cold pixels are present at the limits of clouds in the

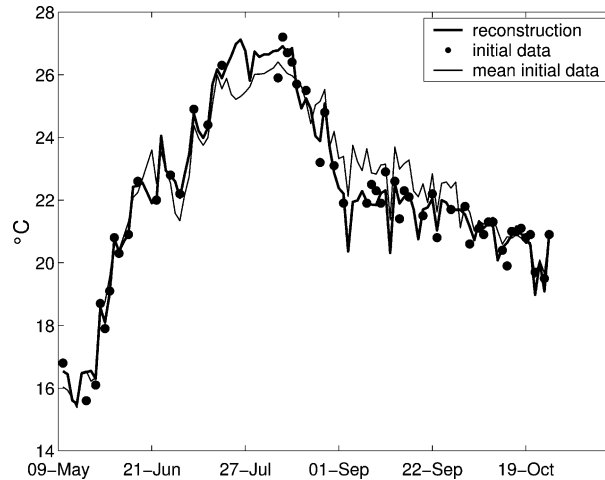


Fig. 7. Time Series of a random point (point at [42N,17.3E]) in the Adriatic Sea. Solid coarse line represents the reconstructed field, • represents the original data, with gaps where there are no data, and thin solid line represents the mean initial temperature.

centre of the Adriatic. The gradient with surrounding waters suggests that it is not a real feature, and that some cloud pixels may have been interpreted in the initial data set as sea pixels. In the reconstruction these features have been eliminated, and the result is a more realistic representation of surface temperature. Also real features, such as currents or meanders, are smoothed in the reconstructed data. Preventing the smoothing of real features by DINEOF is currently under study. The whole series of 105 images reconstructed by DINEOF, as well as the other results presented later in this work can be seen at <http://modb.oce.ulg.ac.be/alvera>.

In Fig. 7 a time series of a random point (point located at a latitude of 42N and a longitude of 17.3E, situated in the centre Adriatic, south of the Croatian Islands) in the Adriatic Sea is shown, with the reconstructed and the original values, including the gaps. As we can see, the reconstruction process smooths the original data. The mean temperature of the initial matrix, calculated over existing points, has been also plotted. The reconstructed series shows some spikes that are also present in the initial data, as for example, around the 10 September. This may indicate that there are some days that show smaller temperature in the whole basin, and this effect is also reflected in the reconstruction.

5.1. Validation

To establish the efficiency of DINEOF, a third data set extracted from the previous one is used. In this data set, only the cleanest images are taken into account. The result is a sequence of 15 images with mean cloud coverage of 18%. Extra cloud coverage is then added to this data set, so that we can then compare the reconstruction to the original images. Clouds are extracted from other images from the Complete data set, so that the coverage presents a realistic aspect. The extra clouds added increase mean cloud coverage up to 40%, 60% and 80%, in order to construct three different data sets, each with different amounts of cloud coverage.

Continuity in time of the images is not necessary to effectuate the reconstruction. Good results can be obtained with irregularly time-distributed data. However, if the set of images is too sparse in some periods (e.g., only one or two images in summer and the rest in autumn), the reconstruction will be deteriorated by irrelevant EOFs describing the variability of a different period of the year, since the information for the former period is insufficient. To avoid this problem, the set of 15 images contains only data from September and October.

An example of reconstruction of these three sets can be found in Fig. 8 which shows the reconstruction of the three data sets the 16 October 1995. The initial, almost clean image of 16 October is shown in Fig. 8a. The images with an additional 40%, 60% and 80% are shown in Fig. 8 c.1, d.1 and e.1 respectively. Their reconstruction is shown in Fig. 8 c.2, d.2 and e.2 respectively. The reconstructed image from the 40% of added clouds shows a good agreement with the original field, free of clouds, as well as the image that was covered with 60% of clouds. The main physical features are realistically represented, such as cold temperatures south of the Po River and the warm current entering by the Strait of Otranto and following the east coast. Temperature

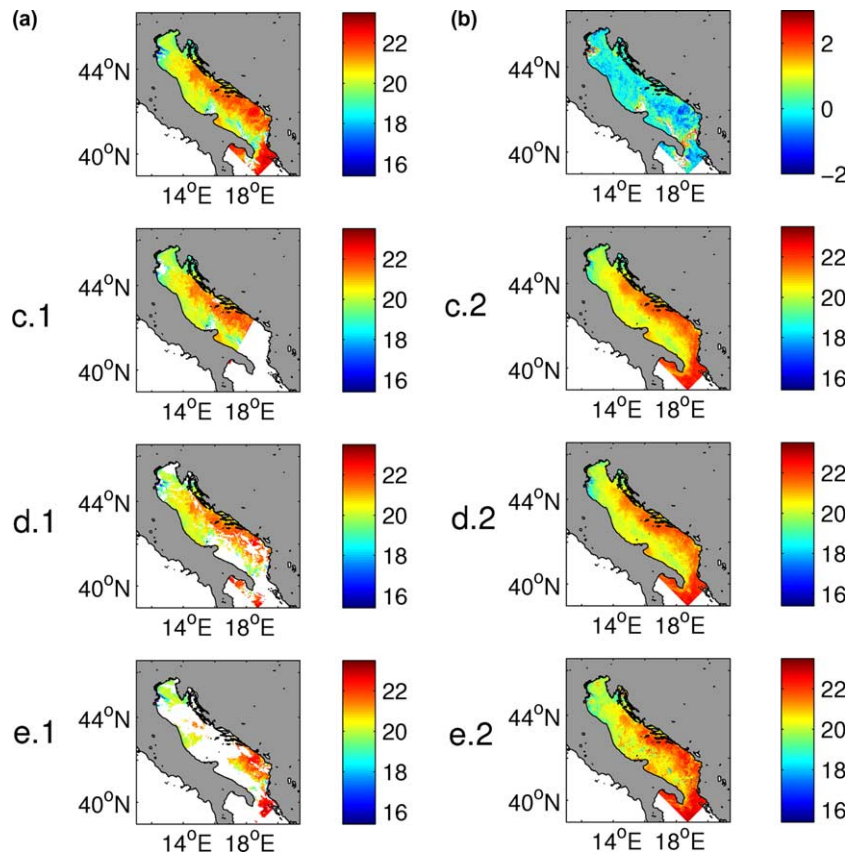


Fig. 8. Image of 16 October. (a) is the original image, with low cloud coverage; (b) is the difference between the reconstruction of 40% added clouds and the initial image; (c.1) is the initial image plus 40% of added clouds; (c.2) shows the reconstruction of (c.1); (d.1) is the initial image plus 60% of added clouds; (d.2) shows the reconstruction of (d.1); (e.1) is the initial image plus 80% of added clouds; (e.2) shows the reconstruction of (e.1).

distribution is maintained. Fig. 8 e.2 is a little bit noisier, due to the extreme cloud coverage of this data set, 80%. In this last figure, the maximal temperatures are slightly weaker in the reconstruction than in the original field, although the temperature distribution is well represented (again, the warm current entering the Adriatic Sea by the Strait of Otranto is maintained). In Fig. 8b the difference between the reconstruction of the 40% added clouds 8c.2 and the initial clean image 8a is shown. The difference between them is very little, with the higher values, as it would be expected, in the zone where clouds were added, although there are differences of the same magnitude in the centre Adriatic.

In these tests, only 2 EOFs were retained for the reconstruction (minimum error obtained by cross-validation, see Fig. 9). The data loss due to added clouds results in a decrease of variance on the reconstructed images. The variance of the three data sets (with 40%, 60% and 80% of cloud coverage) is 94%, 90%, and 80% of the original data variance respectively. Note that the variance has been calculated in relation to the images where extra clouds were added.

Fig. 10 shows the Root Mean Square (RMS) error between the three reconstructed subsets of 15 images and the original one, related to cloud coverage. Error increases with cloud coverage, but not very much. As can be seen, there are points of near 90% coverage that show only errors of 0.7°C, while other points show higher errors and smaller cloud coverage. The RMS error for each data set is about 0.89°C, 0.78°C and 1.25°C for the 40%, 60%, and 80% of added cloud coverage respectively, which is comparable to the error estimate obtained with cross-validation (Fig. 9).

It may appear surprising that the set of 40% added cloud coverage presents a slightly higher error than the set of 60% added clouds. This is simply due to the fact that the added clouds are ‘real’ clouds, i.e., they are taken from other images of the data set, so they are not homogeneously distributed, as would be obtained with an artificial random coverage. As a result, the set with 40% extra cloud coverage has zones where cloud coverage is very high, while clouds in the 60% added clouds set are more homogeneously distributed. As this situation is certainly possible in reality, we decided not to modify the cloud distribution, although it does have to be considered when looking at the obtained RMS error.

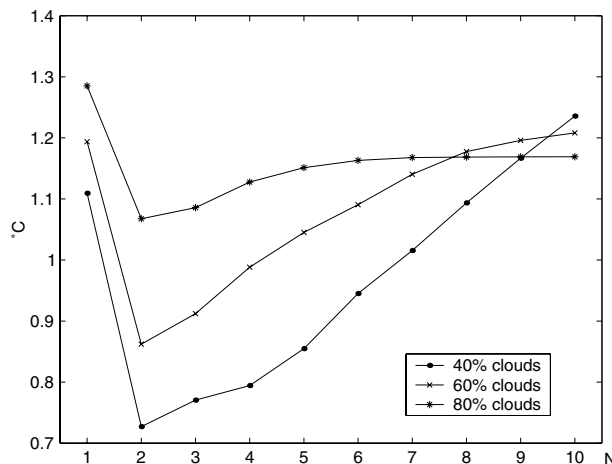


Fig. 9. Error obtained with cross-validation for the validation test. N is the number of EOFs.

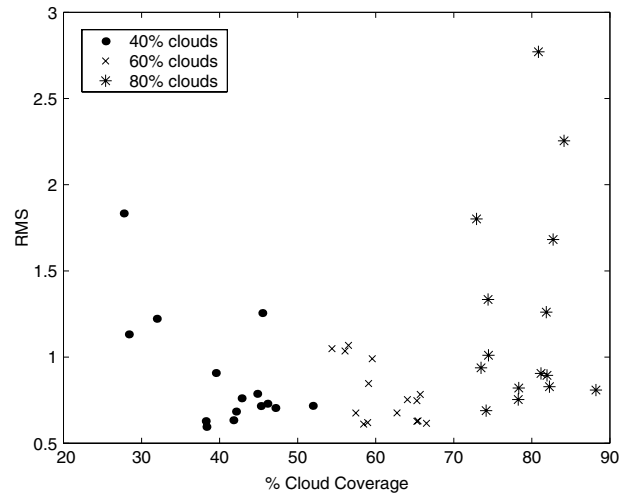


Fig. 10. RMS error related to cloud coverage.

5.2. Independent data

In situ data from the MEDAR/Medatlas database (MEDAR-Group, 2002), covering the period of the Complete Set, have been used as a source of independent data for the validation. In Fig. 2 we can see the distribution of the stations extracted from MEDAR/Medatlas. A total of 452 observations taken at these stations are considered. Surface points from MEDAR/Medatlas database are warmer than the satellite images used here, mostly in summer, due to daily reheating. As the SST images used in this work are night images, the data from MEDAR/Medatlas have been taken at 5 m depth in order to avoid skin temperature effects. The error between *in situ* and reconstructed data is minimised at this depth. The difference between MEDAR/Medatlas data and the reconstruction of the Complete Set is presented in Fig. 11. In this figure we can see the difference between MEDAR/Medatlas data and: (a) the original points (i.e., not covered by clouds) on the SST data set before applying DINEOF; (b) these same points after the reconstruction; and (c)

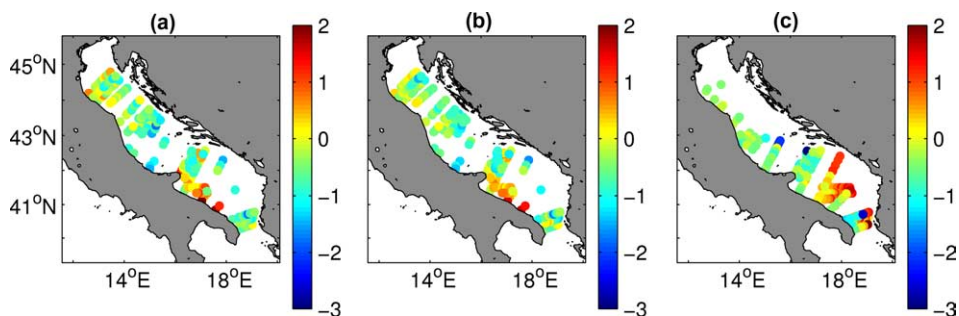


Fig. 11. Difference between the satellite images and *in situ* data from MEDAR/Medatlas database. (a) Shows this difference for initial non-clouded points, before reconstruction. (b) Shows this difference for initial non-clouded points, after reconstruction. (c) Shows this difference for initially clouded points, reconstructed by DINEOF.

those points that were missing in the initial data and whose value have been obtained using DINEOF. The total RMS error for the original satellite data (before DINEOF) is 0.71°C , which we can consider as the error of the satellite measures in relation with *in situ* measures. Other works have found similar errors in satellite data (e.g. Wick et al., 1992). After the reconstruction, the error of these points is 0.67°C . This reduction in error is due to the fact that the truncate series of EOFs used for the reconstruction does not contain noise that may exist in the initial data. For the clouded points that have been reconstructed, the RMS error is 0.95°C , i.e., only 0.28°C higher than the real initial points. The difference between the satellite data and *in situ* data is higher in southern stations, with the satellite data generally warmer than *in situ* data.

6. Comparison with an Optimal Interpolation method

The Harvard Ocean Descriptive Predictive System (HOPS) OI package (Davis, 1985; Robinson and Leslie, 1985; Carter and Robinson, 1987) has been used to reconstruct the Complete Set in order to compare this reconstruction to DINEOF. The HOPS system is based on a minimisation of a pre-selected error norm, chosen to be the mean square difference between the estimate and the true value of the field. A multivariate space-time objective analysis scheme is used to produce regularly gridded fields in time and space (analyses) from the initial data (i.e., irregularly distributed observational data). In this case, the clouded images are used as initial data. The main parameters used in this reconstruction are summarised in Table 2.

The results of reconstruction by the OI method have been compared to MEDAR/Medatlas *in situ* data. The OI approach has a high computational cost, and to obtain results similar to DINEOF, a total computational time of 40h 30min is necessary. It must be said that for this OI reconstruction, only one point in 100 was used, because the reconstruction of the total Complete Set with this method would have taken too long. The RMS error between initially non-clouded points and *in situ* data is 1.78°C . For clouded points, the RMS error is 2.4°C . The OI reconstruction thus presents a higher error than DINEOF (errors given in Section 5.2). Of course, a higher number of points could have been used for the reconstruction, thus obtaining a smaller error. As the test realised here was at the limit of the computer resources, we were unable to take a higher number of points.

Table 2
Main parameters of the OI reconstruction

Number of influential points	20
Radius of influence	25 km
Influential time window	6 days
Observation error	0.1°C
Correlation, zonal zero crossing	60 km
Correlation, meridional zero crossing	60 km
Zonal decorrelation (decay) scale	40 km
Meridional decorrelation (decay) scale	40 km
Temporal decorrelation (decay) scale	6 days

The OI results can be improved by a fit to the initial data. The procedure explained by Smith et al. (1996) is followed. First, the 15 most dominant EOFs are calculated from the OI reconstruction and then the amplitudes are fitted, in the least-squares sense, to the initial existent data. The RMS error is now 0.89°C for initially clouded points, and 0.65°C for non-clouded points. This improvement in the OI reconstruction is easy to obtain and entails only small additional computational cost. The errors of the fitted reconstruction are thus similar to those obtained by the DINEOF reconstruction, with a difference in time required to calculate them.

An additional test was carried out, to determine what the behaviour of the OI procedure would be if the same computational time as used for DINEOF were allotted, i.e., one hour and a half. To stay within this time, only one point out of 1600 could be taken, otherwise the allowed computational time would have been exceeded. The reconstruction obtained in this case with the OI procedure is very poor, and the mean value of the initial existing points is usually recovered as the reconstruction, due to the lack of influential points.

7. An application: EOF analysis

To conclude this work, we realised an EOF decomposition. We aim to show the quality of the reconstructed images with this example. Any kind of study can be done with the reconstructed data set, as there is no longer a coverage limitation. The first three EOFs are presented, and we pay special attention to a cold event that occurred on the Albanian coast in July (Fig. 4).

Fig. 12 shows the first three spatial modes. These three modes account for 94% of the variability (84.6%, 7.2% and 2.2% respectively). The first mode (Fig. 12a) is positive almost everywhere in the basin, indicating a general warming or cooling of the Adriatic Sea. The temperature distribution corresponds to that described as the typical situation in the Adriatic Sea. Indeed the east coast presents smaller values than the west coast. The zone surrounding the Po River has the strongest values, indicating that the influence of the Po River is very high. When looking at the first temporal mode (Fig. 13a) we see a seasonal cycle, with a warming trend when approaching summer and a cooling when approaching autumn. The first mode thus represents the seasonal cycle ranging from May to October.

In the first and second modes, there is a clear signal of the EAC entering the Ionian Sea. The different evolution in time of both modes contributes to the modulation of this current. The first

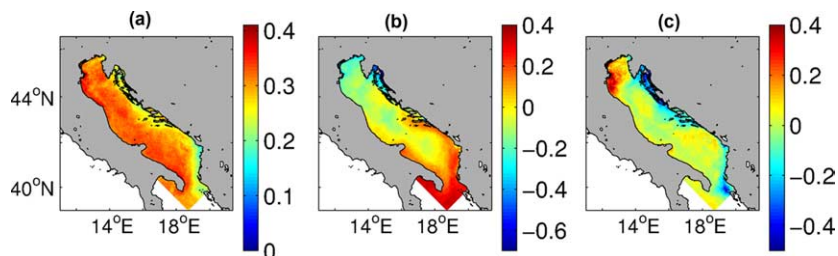


Fig. 12. Spatial EOFs, modes 1–3 (a)–(c). Units are $^{\circ}\text{C}^{1/2}$.

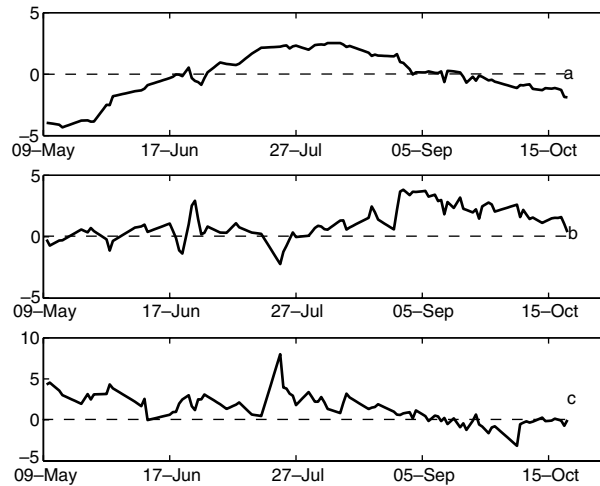


Fig. 13. Temporal EOFs, modes 1–3 (a)–(c). Units are $^{\circ}\text{C}^{1/2}$.

mode reaches a maximum when the second one has its smallest values, i.e., at the end of July. Then the second mode presents its highest values while the first one decreases towards zero.

The second spatial mode (Fig. 12b) also presents the WAC exiting the Adriatic Sea. We can see the signal of the filament detaching from the Istrian Peninsula. In the second temporal mode (Fig. 13b), we can see that the fluctuations reach their highest values from 1 September to the end of the record. The EAC is more intense in autumn (Artegiani et al., 1997b; Cushman-Roisin et al., 2001; Poulain, 2001), and when looking at the Complete Set, we can see that both currents, EAC and WAC, increase their intensity in this period. We can also observe a north–south division of the basin temperature in the second spatial mode, with the largest negative values in the Croatian coast, south of the Istrian peninsula. In summer, the temperature in the northern basin is higher than the temperature in the southern basin. The opposite occurs in winter (Cushman-Roisin et al., 2001). As can be seen when looking at the second temporal mode, the values from May to August are near zero or negative, which indicates that the northern basin is warmer than the southern one in the summer season, as the second spatial mode presents negative values in the north. In autumn, the second temporal mode reaches its highest values, indicating that the northern basin is then colder than the southern basin.

In the third spatial mode (Fig. 12c) we can clearly see the signal of the Po plume, which extends over a large zone. This is the situation mainly found in summer, when winds decrease and the Po River water can spread over the entire northern basin. In winter, the plume is reduced, and winds force it to follow the west coast of the Adriatic Sea (Cushman-Roisin et al., 2001; Mauri and Poulain, 2001). This situation can be seen in the third temporal mode (Fig. 13a), where we see a gentle decreasing trend all along the time series as autumn approaches. The water coming from the Po River is warmer than the surrounding waters from May to August. In September–October, when the temporal mode becomes negative, the Po River water is colder than that in open sea. This situation has been described by Gacić et al. (1997) and Cushman-Roisin et al. (2001). Again, the largest negative values are found on the Croatian coast. These waters are cold in summer (positive values of the temporal mode), but in September, when the EAC inflow is increased, the hot waters

also affect this zone, which become warmer than in summer (negative values of the temporal mode). South of the Istrian peninsula, there is the signal of a filament with negative values.

One feature that attracted our attention is a peak that is present in the second and third temporal modes, occurring 23 July. This abrupt change is negative in the second mode, and positive in the third one. This peak occurs on 23 July; this image has already been shown (Fig. 4). In this figure, we can see a cold tongue of water along the Albanian coast, in the southeast of the Adriatic, which is also partially visible in the initial cloudy data set (Fig. 5). This feature can be observed from 22 to 26 July. Bergamasco and Gacić (1996) and Gacić et al. (1997) describe the occurrence of cold waters in this zone, but its cause is not well documented. It may be an upwelling due to the action of the northeasterly Bora wind, when it blows parallel to the coast, or simply

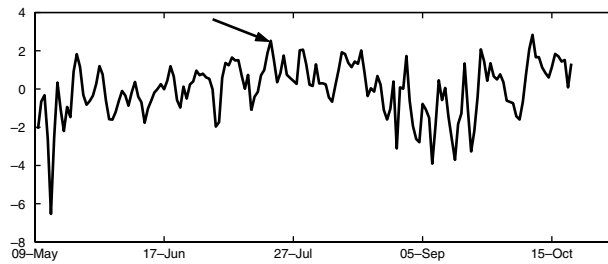


Fig. 14. ECMWF wind series projected to NE direction. The arrow shows the peak of wind speed on 20 July 1995.

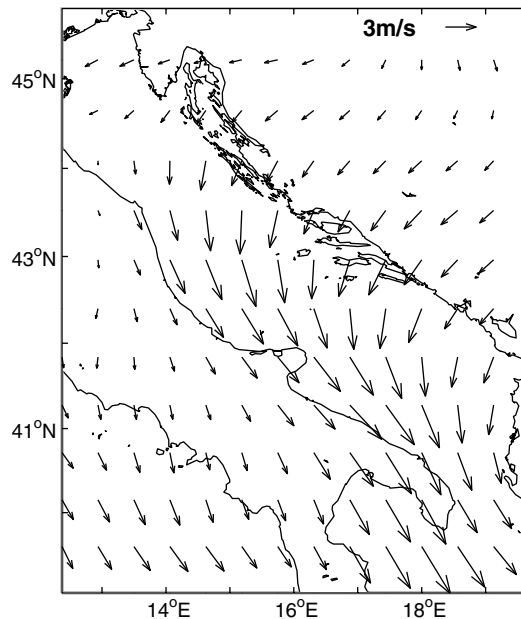


Fig. 15. Wind mean distribution for the period from 17 July to 22 July 1995.

the vertical mixing induced by the wind. The study of the wind series for this period is thus important to understand this event. We used the ECMWF Re-Analysis (ERA) wind series from May to October. This data set covers the Adriatic Sea with a half-degree resolution in both latitude and longitude. As a result, a total of 14×13 points over the Adriatic are used. In Fig. 14 the projection of this time series in the NE direction is shown in order to identify Bora wind. As can be seen, an arrow marks a peak occurring on 20 July, when the wind reaches high values in Bora direction. Looking at the wind distribution these days (Fig. 15), it can be seen that for a period of six days, from 17 to 22 July, the wind blows constantly parallel to the Albanian Coast, with the maximum speed reached on 20 July. The speed ranges from 3 to 5 m/s, which is not considered as a strong Bora event. However, this speed is only exceeded on 9 October, so the event of 20 July is one of the strongest of the period May–October 1995. The figures shown suggest that the cold tongue observed at the Albanian coast is caused by the action of the Bora. The reconstructed data set very accurately reflects this event.

8. Conclusion

A reconstruction method, called DINEOF, has been successfully applied to a large matrix. The method is robust, simple to use and does not need any a priori information about the error statistics of the data. The results obtained have been analysed, giving an example of their reliability and usefulness.

The aim of this work was the application of the method to a realistic case, a data set covering the whole Adriatic Sea for a six-month period. To do so, a Lanczos solver was combined with DINEOF, in order to calculate the EOFs in an optimised way. Computational times are good for the examples given in this paper. However, for the application of the methodology to very large matrices some additional optimisation work could be done, such as stop criteria for the calculation of the optimal number of EOFs, which would indicate when the minimum error calculated by cross-validation has been attained.

The results of the reconstruction in the Adriatic Sea are accurate and smoothly included in the final result, as seen in the visual examples given and in the validation studies made with a small set of the matrix. The validation was carried out in a data set with increasing amounts of missing data, where 40%, 60% and 80% of data loss were added. The comparison of the reconstructed fields with the original one reveals that the error is small, about 0.89°C , 0.78°C and 1.25°C for the 40%, 60% and 80% of missing data respectively. The comparison with *in situ* data from the MEDAR/Medatlas database also reveals an RMS error of the same magnitude, of about 0.95°C for the points that are covered by clouds in the Complete Set. When visually checking the reconstruction results, we can see that the main physical features are recovered in the final result, such as the Po River Plume, cold filaments generated in the eastern coast, or the warm water current entering the Adriatic Sea by the Strait of Otranto in autumn.

Some artificial features in the initial data are also filtered out of the final result using the reconstruction method. This is highly interesting when dealing with data sets that have not been properly treated and present abrupt temperature changes in the vicinity of clouds.

A comparison with an OI method was realised. The clouded data are reconstructed with OI, and then the EOFs are calculated and their amplitudes are fitted to the initial data. The results

show that both methods, fitted OI and DINEOF, perform very similarly. The main difference is in the computational time, which is nearly 30 times greater for the OI reconstruction. In many situations, such as an operational frame, this time difference is very important.

An EOF analysis of the reconstructed field shows the utility of the reconstructed images. The first mode shows the seasonal cycle, from spring to autumn. The second mode presents the modulation of two general currents, the EAC and the WAC. Finally, the third mode shows the modulation of the Po plume. A cold event on the Albanian coast has been studied, and compared to a time series of wind obtained from ECMWF. This cold event is shown to be related to Bora winds that blow over the Albanian coast in the studied period.

In this study, DINEOF has been applied to a temperature field. Further work may show its utility when reconstructing other variables, such as salinity or chlorophyll. When dealing with chlorophyll fields, the patchiness that usually characterises their distribution (Martin, 2003; Strass, 1992) makes the reconstruction of these data a difficult task. Chlorophyll fields are very decorrelated, and techniques like optimal interpolation may have difficulty reconstructing this data. The characteristics of DINEOF make it suitable for reconstruction of such data. Any other field obtained from satellites (from passive receptors that cover a large area and thus provide a good data coverage) or from dense oceanographic data sets is suitable to be reconstructed.

Acknowledgements

The authors wish to thank Vincent Toumazou for providing the code for the EOF decomposition, IterativeEOF, which allowed reducing the computational cost of this work, and two anonymous reviewers for their helpful comments. The National Fund for Scientific Research, Belgium, is acknowledged for the financing of a supercomputer. This work was realised in the framework of the European SOFT project (EVK3-CT-2000-00028). This is MARE publication MARE051.

References

- Alvarez, A., 2003. Performance of satellite-based ocean forecasting (SOFT) systems: a study in the Adriatic Sea. *Journal of Atmospheric and Oceanic Technology* 20 (5), 717–729.
- Artegiani, A., Bregant, D., Paschini, E., Pinardi, N., Raicich, F., Russo, A., 1997a. The Adriatic Sea general circulation. Part I: Air–Sea interactions and water mass structure. *Journal of Physical Oceanography* 27, 1492–1514.
- Artegiani, A., Bregant, D., Paschini, E., Pinardi, N., Raicich, F., Russo, A., 1997b. The Adriatic Sea general circulation. Part II: Baroclinic circulation structure. *Journal of Physical Oceanography* 27, 1515–1532.
- Beckers, J., Rixen, M., 2003. EOF calculations and data filling from incomplete oceanographic data sets. *Journal of Atmospheric and Oceanic Technology* 20, 1839–1856.
- Bennani, M., Braconnier, T., 1994. Stopping criteria for eigensolvers. Technical Report TR/PA/94/22, CERFACS, Toulouse, France, pp. 1–20.
- Bennett, A., 2002. *Inverse modeling of the ocean and the atmosphere*, Cambridge.
- Bergamasco, A., Gacić, M., 1996. Baroclinic response of the Adriatic Sea to an episode of Bora wind. *Journal of Physical Oceanography* 26, 1354–1369.
- Bergamasco, A., Oguz, T., Malanotte-Rizzoli, P., 1999. Modelling dense water mass formation and winter circulation in the northern and central Adriatic Sea. *Journal of Marine Systems* 20, 279–300.

- Borzelli, G., Manzella, G., Marullo, S., Santoleri, R., 1999. Observations of coastal filaments in the Adriatic Sea. *Journal of Marine Systems* 20, 187–203.
- Brankart, J., 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.
- Brasseur, P., Beckers, J., Brankart, J., Schoenauen, R., 1996. Seasonal temperature and salinity fields in the Mediterranean Sea: Climatological analyses of a historical data set. *Deep-Sea Research I* 43, 159–192.
- Carter, E., Robinson, A., 1987. Analysis models for the estimation of oceanic fields. *Journal of Atmospheric and Oceanic Technology* 4 (1), 49–74.
- Chatelin, F., 1993. *Eigenvalues of Matrices*. Wiley.
- Chu, P., Hsing-Chia, T., Chang, C., Chen, J., 1997. South China warm pool detected in spring from the Navy's Master Oceanographic Observational Data Set (MOODS). *Journal of Geophysical Research* 102 (C7), 15761–15771.
- Cushman-Roisin, B., Gacić, M., Poulain, P., Artegiani, A., 2001. *Physical Oceanography of the Adriatic Sea*. Kluwer Academic Publishers.
- da Silva, J., Ermakov, S., Robinson, I., Jeans, D., Kijashko, S., 1998. Role of surface films in ERS SAR signatures of internal waves on the shelf. 1. Short-period internal waves. *Journal of Geophysical Research* 103 (C4), 8009–8031.
- Davis, R., 1985. Objective mapping by least squares fitting. *Journal of Geophysical Research* 90 (C3), 4773–4777.
- Dousset, B., Firing, J., Flament, P., Jackson, H., Lumpkin, C., Maroni, C., Nacini, E., Poulain, P., Pouliquen, S., Sawyer, M., Young, D., 1998. Adriatic sea surface temperature images from the NOAA-AVHRR, 1995. Département d'Océanographie Spatiale IFREMER, Brest CD-ROM.
- Emery, W., Thomson, R., 1998. *Data Analysis Methods in Physical Oceanography*. Pergamon.
- Everson, R., Cornillon, P., Sirovich, L., Webber, A., 1997. An empirical eigenfunction analysis of sea surface temperatures in the western North Atlantic. *Journal of Physical Oceanography* 27, 468–479.
- Fieguth, P., Menemenlis, D., Ho, T., Willsky, A., Wunsch, C., 1998. Mapping mediterranean altimeter data with a multiresolution optimal interpolation algorithm. *Journal of Atmospheric and Oceanic Technology* 15, 535–546.
- Gacić, M., Marullo, S., Santoleri, R., Bergamasco, A., 1997. Analysis of the seasonal and interannual variability of the sea surface temperature field in the Adriatic Sea from AVHRR data (1984–1992). *Journal of Geophysical Research* 102, 22937–22946.
- He, R., Weisberg, R., Zhang, H., Muller-Karger, F., R.W., H., 2003. A cloud-free, satellite-derived, sea surface temperature analysis for the West Florida Shelf. *Geophysical Research Letters* 30 (15).
- Housego-Stokes, R., 2000. Using optimal interpolation and EOF analysis on North Atlantic satellite data. *International WOCE Newsletter* 38, 26–28.
- Kaplan, A., Cane, M.A., Kushnir, Y., Clement, A.C., Blumenthal, B., Rajagolapan, B., 1998. Analyses of global sea surface temperature 1856–1991. *Journal of Geophysical Research* 103 (C9), 18576–18589.
- Kaplan, A., Kushnir, Y., Cane, M., Blumenthal, M., 1997. Reduced space optimal analysis for historical datasets: 136 years of Atlantic Sea surface temperatures. *Journal of Geophysical Research* 102 (C3), 27835–27860.
- Le Traon, P., Didarboure, G., 1999. Mesoscale mapping capabilities of multiple-satellite altimeter missions. *Journal of Atmospheric and Oceanic Technology* 16 (9), 1208–1223.
- Le Traon, P., Didarboure, G., Ducet, N., 2001. Use of a high-resolution model to analyze the mapping capabilities of multiple-altimeter missions. *Journal of Atmospheric and Oceanic Technology* 18 (7), 1277–1288.
- Le Traon, P., Nadal, F., Ducet, N., 1998. An improved mapping method of multisatellite altimeter data. *Journal of Atmospheric and Oceanic Technology* 15 (2), 522–534.
- Lehoucq, R., Sorensen, D., Yang, C., 1997. ARPACK user's guide: solution of large scale eigenvalue problems with implicitly restarted Arnoldi methods. pp. 1–152. URL Available online at <http://www.caam.rice.edu/software/ARPACK/>.
- Liu, A., Chang, Y., Hsu, M., Liang, N., 1998. Evolution of nonlinear internal waves in the East and South China seas. *Journal of Geophysical Research* 103 (C4), 7995–8008.
- Martin, A., 2003. Phytoplankton patchiness: the role of lateral stirring and mixing. *Progress in Oceanography* 57, 125–174.
- Mauri, E., Poulain, P., 2001. Northern Adriatic Sea surface circulation and temperature/pigment fields in September and October 1997. *Journal of Marine Systems* 29, 51–67.

- MEDAR-Group, 2002. MEDATLAS/2002 database. Mediterranean and Black Sea database of temperature salinity and bio-chemical parameters. Climatological Atlas. Ifremer edition 4 CD-ROM.
- Orlić, M., Gacić, M., La Violette, P., 1992. The currents and circulation of the Adriatic Sea. *Oceanologica Acta* 15 (2), 109–124.
- Poulain, P., 1999. Drifter observations of surface circulation in the Adriatic Sea between December 1994 and March 1996. *Journal of Marine Systems* 20, 231–253.
- Poulain, P., 2001. Adriatic Sea circulation as derived from drifter data between 1990 and 1999. *Journal of Marine Systems* 29, 3–32.
- Reynolds, R.W., Smith, T., 1994. Improved global sea surface temperature analyses using optimum interpolation. *Journal of Climate* 7, 929–948.
- Rixen, M., Beckers, J.-M., Brankart, J.-M., Brasseur, P., 2001. A numerically efficient data analysis method with error map generation. *Ocean Modelling* 2 (1–2), 45–60.
- Robinson, A., Leslie, W., 1985. Estimation and prediction of oceanic eddy fields. *Progress in Oceanography* 14, 485–510.
- Saunders, R., Kriebel, K.T., 1988. An improved method for detecting clear sky and cloudy radiances from AVHRR data. *International Journal of Remote Sensing* 9, 123–150.
- Smith, T.M., Reynolds, R.W., Levezy, R.E., Stokes, D.C., 1996. Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *Journal of Climate* 9, 1403–1420.
- Strass, V., 1992. Chlorophyll patchiness caused by mesoscale upwelling at fronts. *Deep-Sea Research* 39, 75–96.
- Toumazou, V., Cretaux, J., 2001. Using a Lanczos eigensolver in the computation of empirical orthogonal functions. *Monthly Weather Review* 125 (5), 1243–1250.
- von Storch, H., Zwiers, W., 1999. *Statistical Analysis in Climate Research*. Cambridge University Press.
- Wick, G., Emery, W., Schluessel, P., 1992. A comprehensive comparison between satellite-measured skin and multichannel sea surface temperature. *Journal of Geophysical Research* 97 (C4), 5569–5595.
- Wilks, D., 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press.