Forecast verification of a 3D model of the Mediterranean Sea.

The use of discrete wavelet transforms and EOFs in the skill assessment of spatial forecasts

A. Alvera-Azcárate [a], A. Barth [a], Z. Ben Bouallègue [b], M. Rixen [c], J.M. Beckers [b]

[a] *College of Marine Science, University of South Florida, 140 7th Avenue South, 33701 Saint Petersburg, Florida, USA* [b] *GHER-MARE, University of Liège, Allée du 6 Août 17, B5, Sart Tilman, 4000 Liège, Belgium* [c] *NATO Undersea Research Centre (NURC), Viale San Bartolomeo 400, 19138 La Spezia, Italy*

Abstract: The quality assessment of a nested model system of the Mediterranean Sea is realised. The model has two zooms in the Provençal Basin and in the Ligurian Sea, realised with a two-way nesting approach.
The experiment lasts for nine weeks, and at each week sea surface temperature (SST) and sea level anomaly are assimilated. The quality assessment of the surface temperature is done in a spatio-temporal approach, to take into account the high complexity of the SST distribution. We focus on the multi-scale nature of oceanic processes using two powerful tools for spatio-temporal analysis, wavelets and Empirical Orthogonal Functions (EOFs). We apply two-dimensional wavelets to decompose the high-resolution model and observed SST into different spatial scales. The Ligurian Sea model results are compared to observations at each of those spatial scales, with special attention on how the assimilation affects the model behaviour. We also use EOFs to assess the similarities between the Mediterranean Sea model and the observed SST. The results show that the assimilation mainly affects the model large-scale features, whereas the small scales show little or no improvement and sometimes, even a decrease in their skill. The multiresolution analysis reveals the connection between large- and small-scale errors, and how the choice of the maximum correlation length of the assimilation scheme affects the distribution of the model error among the different spatial scales.

Keywords: Forecast verification; Spatio-temporal analysis; Wavelets; EOFs; Mediterranean Sea

1. Introduction

The quality assessment of a three-dimensional model forecast is a difficult task, particularly when studying its spatio-temporal characteristics. For this kind of study we can use *in situ* data, but their coverage is usually limited in space and/or time, making difficult the comparison with a three-dimensional model. Another possibility is the use of satellite observations, which have very good coverage in space and time but are limited to the ocean surface. The comparison of the model even to such incomplete fields can however add valuable information to the understanding of the model behaviour.

The comparison between these complex fields, model and observations, is not straightforward because two fields describing the distribution of a variable can differ in many ways (*e.g.* an overall bias, missing processes, bad positioning of a given feature, etc.)- The first step to assess the quality of a forecast is thus to establish which processes are being studied. The definition of the spatio-temporal scales related to these processes will then fix the model characteristics that we need to study.

Forecast verification depends thus on what we want to verify, and at which spatio-temporal scales.
The verification process must be designed to find the answers to the questions that may arise when dealing with forecast results. In field forecast verification, these questions are related to the accuracy of the model with respect to observations, but also to the spatial distribution of the studied variable and the temporal evolution of this distribution. The study of the spatial distribution of a variable can help us to see if there are missing processes in the model (such as a recurrent gyre or a front) and to learn about the capacity of the model to represent the reality at different physical spatial scales.

The study of such complex error fields cannot be done with a unique error measure. To answer questions related to the spatial distribution of a variable, its spatial distribution averaged over time can be studied. However, in field forecast verification special attention must be paid to the fact that the number of grid points in a model and the spatial correlation between them makes very difficult to study the skill distribution in space (Livezey and Chen, 1983). It is very unlikely that two adjacent points in a model grid are completely independent, so the interpretation of a spatial skill (temporal averaged) becomes ambiguous (Wilks, 1995; Briggs and Levine, 1997; Jolliffe and Stephenson, 2003).

Time error evolution is thus preferred to avoid the correlation problem, as long as the time between analyses is longer than the temporal correlation scale. An average over all points in a grid at a given time is thus commonly used in the verification process. The temporal evolution of the error is very useful to obtain a general idea about the quality of the model. It is often the only way an error measure is applied. However, to answer some of the questions specified above, one needs to keep the spatial distribution of the studied variable. Spatio-temporal techniques can help us to study the evolution of a variable in time, keeping the spatial distribution information, and avoiding the ambiguity of correlation between adjacent points. Multi-scale techniques allow us to study the behaviour of a model at different spatial scales (Daubechies, 1992; Mallat, 1998), or even to focus on a specific scale of interest. Nested models, as the one we are working with, can also be considered as a multi-scale approach. Each nested level is a refinement of the parent model, so the verification at those two model grids can also give us an idea about the model behaviour at different spatial and/or temporal scales (Denis et al., 2003).

Wavelet Transforms are widely used in multi-scale decomposition studies (Daubechies, 1992; Torrence and Compo, 1998). They overcome the localisation problem of Fourier Transforms. By using a variable window size that is translated and dilated over the studied domain, wavelets allow us to separate a signal into orthogonal components related to the position and scale of the signal (Mallat, 1989, 1998).

Several studies have applied a multi-scale approach to analyse the spatial behaviour of a variable. Liang and Robinson (2005) established a multi-scale Energy and Vorticity Analysis (MS-EVA), that uses wavelets for the multi-scale decomposition. Liang and Robinson (2004) used MS-EVA to study the energy and vorticity balances at different spatial and temporal scales of the Iceland-Faroe Front and considered the transfer and distribution of energy and vorticity between the large-scale, mesoscale and sub-mesoscale. A similar approach was used by Fournier (2002, 2003) for atmospheric fields. Yano et al. (2001) made a three-dimensional study of a convective cloud system, characterising preferred spatial orientations of the system. Briggs and Levine (1997) and Casati et al. (2004) worked with 500-hPa geopotential height fields and rain fields respectively.

Two-dimensional wavelets have been recently applied in the frame of forecast verification (Briggs and Levine, 1997; Casati et al., 2004). In these papers the authors decomposed a model into several spatial scales using wavelets, and assessed the quality of the model at each of these scales. As these scales have a physical meaning, we can identify the error with a physical process characteristic of each scale (Briggs and Levine, 1997).

This last approach is exploited in this work. Wavelets are a perfect tool for spatio-temporal analyses of two-dimensional fields. The aim is to decompose the model forecast sea surface temperature (SST) and the observations into different resolution levels or scales. Then, the comparison between model and observations can be done at each of these scales. This allows us to identify the scales that are mainly contributing to the global error, and thus to have a closer look into the behaviour of these fields.

Empirical Orthogonal Functions (EOFs) are also used in this work. They can decompose a matrix into orthogonal modes representing the major patterns of variability found in the data. EOFs are very useful in the comparison between a model and observations in a spatio-temporal frame. We can study whether the model presents the same spatial and temporal distribution than the observations at their major modes of variability. The study of the spatial modes allows us to see if there are missing features in the model, or if they are represented with the appropriate intensity. The temporal modes allow us to see if the temporal cycle (*e.g.* monthly, seasonal or annual cycle) is accurately represented by the model, or if there is a shift or a difference in intensity between model and observations. Several works use EOFs to study the spatial and temporal characteristics of the ocean. For example, Lermusiaux (2001) studied the three-dimensional variability of the Massachusetts Bay using EOFs. Lermusiaux (2002) made an analysis about the sensitivity of a mapping method to the size of the considered subspace, the studied scales, and the system dynamics. Molcard et al. (2002) realised an EOF analysis of the wind stress over the Mediterranean Sea, to study the relationship between wind stress and the ocean response. Beckers et al. (2002) used EOFs for the intercomparison of different models in the Mediterranean Sea.

In this work we present the quality assessment of a three-dimensional model in the Mediterranean Sea. The model has two zooms to the Provençal basin and to the Ligurian Sea (Barth et al., 2005). These three models are connected by a two-way nesting approach. The quality assessment of the different resolution models gives us an idea about how it behaves at different grid resolutions. We study the model error at the basin and sub-basin scales before applying the mentioned spatio-temporal techniques. We then apply wavelet decomposition to the Ligurian Sea model, and this decomposition will allow us to study the spatial characteristics from one to about a hundred kilometres.

The model is forced by data assimilation. Weekly SST and SLA are assimilated during nine weeks. In addition,

we dispose of daily satellite SST images for the quality assessment. Thus, we focus on time scales from one day to a month.

This work is organised as follows: first, a brief description of the Mediterranean and Ligurian Sea main characteristics is made in Section 2. Then the model and data used, and a brief description of discrete wavelet transforms are presented in Section 3. Section 4 presents the time evolution of the model SST error of the Mediterranean and Ligurian Sea models, which gives us an idea about the behaviour of the model at the basin and sub-basin scales. Then these models are analysed in a spatio-temporal frame: the Ligurian Sea model first, with two-dimensional discrete wavelet analysis in Section 5, and then the Mediterranean Sea with an EOF analysis in Section 6. We pay special attention to the impact of the assimilation in the model behaviour and in the distribution of the error at different scales. Finally, we present the conclusions of our work in Section 7 along with the recommendations, obtained through the error analyses, to improve the model skill.

2. The Mediterranean sea

The Mediterranean Sea (see Fig. 1) is a semi-enclosed sea with an extension of approximately 3000 km in the zonal direction and 1500 km in the meridional direction. It is connected to the Atlantic Ocean by the Strait of Gibraltar, a narrow canal of 20 km width and 300 m depth.

The Mediterranean Sea freshwater input (precipitation and river discharges) does not balance the evaporation. This evaporation/precipitation balance mainly controls the large-scale circulation in the Mediterranean Sea: the evaporation makes the surface water more saline, originating a convective process that involves the entire water column and that leads to the formation of deep waters in zones like the Western Mediterranean and the Rhodes gyre. In the Strait of Gibraltar, the cooler and lighter Atlantic water enters the Mediterranean above the outflowing saltier and warmer Mediterranean Water (Viudez et al., 1996). Due to evaporation and mixing, the Atlantic Water gradually changes its physical and chemical properties, forming the Modified Atlantic Water (MAW). The MAW forms a surface layer of about 100-200 m depth all over the Mediterranean Sea (Millot, 1999). Circulation in the Mediterranean Sea follows a cyclonic path all over its basins and sub-basins. The MAW contributes to the formation of an intermediate water mass (Levantine Intermediate Water, LIW) in the Eastern Mediterranean, due to warming and evaporation processes. This newly formed LIW returns to the west and exits the Mediterranean through the Strait of Gibraltar below the MAW.

Fig. 1: *The Mediterranean Sea grid (22 × 28 km resolution) and its bathymetry. The two successive squares represent the next nested domains, the Provençal Basin in solid line and the Ligurian Sea in dashed line.*
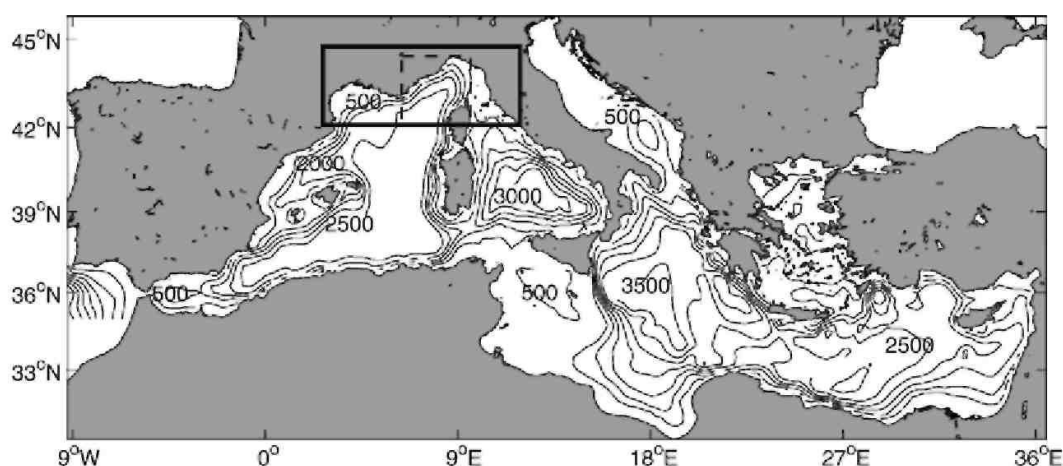
The circulation in the Mediterranean Sea occurs at mainly three spatial scales: basin scale, sub-basin scale and mesoscale. At a basin scale, the most important pattern is the zonal circulation of MAW. The sub-basin scale is characterised by gyres (sub-basinscale gyres) and permanent or quasi permanent cyclonic and anticyclonic structures *(e.g.* Rhodes gyre and Ierapetra Gyre) interconnected by intense jets and meandering currents (Malanote-Rizzoli et al., 1999). We can also think in terms of time scale, as interannual scale *(e.g.* deep water mass formation and variations in volume transport), seasonal scale *(e.g.* thermocline depth variations, gyres) and smaller scales (Brankart and Brasseur, 1998; Pinardi and Masetti, 2000).

2.1. Liguro-Provençal basin

The circulation in the Ligurian Sea (dashed box in Fig. 1) describes a cyclonic gyre, which is more intense in winter and is mainly due to wind curl stress (Larnicol et al., 1995). Two northward currents surrounding the coast of Corsica, the West Corsican Current and the East Corsican Current, feed the circulation pattern of this area. Both branches join north of Corsica, forming the Northern Current (NC), which flows south-westward following the French and the Spanish coasts along the continental slope. The signal of the NC extends from the north of Corsica to as far as the Catalan Sea (Albérola et al., 1995; Sammari et al., 1995; Astraldi et al., 1999; Millot, 1999).

The Ligurian Sea presents a marked seasonal variability, mainly noticed in the variations in volume and position of the NC all along the year (Albérola et al., 1995; Albérola and Millot, 2003). The modulation of the NC is determined by changes in the two currents that feed it, the Western and Eastern Corsican currents (Astraldi and Gasparini, 1992; Sammari et al., 1995; Astraldi et al., 1999). The NC is also affected by a high mesoscale activity (Albérola et al., 1995), mainly reflected in the formation of meandering structures all along the NC path. The meandering structures can be as wide as the NC and they can advance at the same velocity (Sammari et al., 1995). These seasonal and mesoscale variations can be seen all along the NC south-westward flow (Font et al., 1995).

3. Material and methods

3.1. The GHER model

The GHER (GeoHydrodynamics and Environmental Research) model, a three-dimensional model (Beckers, 1991) has been applied to the Mediterranean Sea. The model is governed by the primitive equations and it has a free surface. In the vertical, the model uses a double-sigma coordinate with 31 levels. The numerical scheme conserves mass, heat and salt. Other details of the GHER model can be found in Beckers (1991) and Barth et al. (2005).

The Smith and Sandwell (1997) bathymetry has been used in the model. The initial salinity and temperature are computed from MODB (Mediterranean Ocean DataBase, Brasseur et al. (1996)). Two river freshwater inputs are also included: the Rhone (in France) and the Arno (in Italy) rivers, obtained from Tusseau and Mouchel (1994) and from Rinaldi (pers. comm.), respectively. Daily heat and momentum fluxes are calculated from bulk formulae. Temperature of the air at two meters, cloud coverage, wind speed at 10 m, dew point temperature and pressure fields are obtained from the European Center for Medium-Range Weather Forecasting (ECMWF), and they are used for the computation of the heat fluxes. The ECMWF analysis fields have a resolution of 1/2° and are produced each 6 h. The heat fluxes used in the model are interactive, *i.e.,* they depend of the surface temperature provided by the GHER model.

This model has been applied in a two-way nesting approach to the Ligurian Sea (Barth et al., 2005). The model consists in three nested sub-domains: the coarsest grid represents the whole Mediterranean, with 22 × 28 km resolution, the intermediate grid covers the Liguro-Provençal basin (4x6 km), and the finest grid is centred in the Ligurian Sea, with 1.5 × 1.9 km resolution. These three domains can be seen in Fig. 1, where the bathymetry is also represented.

From 6 July to 1 September 2000, an assimilation experiment has been carried out. Each week SST and SLA satellite data are assimilated into the model via a Singular Evolutive Extended Kalman filter (Pham and Verron, 1998; Brasseur et al., 1999; Brankart et al., 2003), which is a reduced order Kalman filter. For the computation of the error covariance of the model, an ensemble of 200 members was generated. Initial conditions and atmospheric forcings are perturbed and then the ensemble of 200 members runs for 2 weeks, beginning on 5 July. The final ensemble is used to calculate the error covariance of the model. For further information on the assimilation experiment, see (Barth et al., 2007-this issue).

The aim of this study is to learn how the SST assimilation influences the model skill. We refer to the model with SST assimilation as *Forecast.* This simulation is studied each week before the assimilation of SST and SLA, so the *Forecast* is still independent from the observations that will be assimilated into the model. To assess the contribution of the assimilation of SST and SLA into the model, we use a free run (*i.e.* a run without assimilation, but with the same parameterisation and forcings as the *Forecast*). This data set is called *Free run.* Finally, the result of the assimilation of observations is used. This data set is used to study the impact of the assimilation step on the model, and it is called *Analysis.* This data set is then dependent of observations.

3.2. Observations

Advanced Very High Resolution Radiometer (AVHRR) SST data from the German Aerospace Research Center (DLR, http://eoweb.dlr.de:8080/servlets/ template/welcome/entryPage.vm) are assimilated into the model, and they will be used in the skill assessment. They have a resolution of approximately 1 km. To avoid artificial skill when comparing the model with assimilated data, the comparison is always done before the data is assimilated. DLR data are mean weekly composite temperature fields. The composite value for every pixel is computed from the daily maximum temperature images, and clouded pixels are excluded from the data. We refer to this data set as DLR SST.

Daily AVHRR Pathfinder v5 SST data (hereafter Pathfinder SST) are also used. Pathfinder SST data are available via anonymous ftp at the Jet Propulsion Laboratory web site (ftp://podaac.jpl.nasa.gov). They consist on daily averaged global SST maps. A subset ranging from 6 July to 1 September 2000 and covering the Mediterranean Sea is used. Only descending (nighttime) passes are selected, to minimise skin temperature effects.

They are derived from the 5-channel AVHRR on board the NOAA-7, 9, 11, 14, 16 and 17 polar orbiting satellites. Together with the Pathfinder SST data, quality flags can also be obtained. Clouds are identified from these quality flags, so each user can decide the mask that will be applied to the data. The possibility of choosing the level of quality of the data that we want to keep, and the fact that nighttime passes are separated from daytime passes, makes the Pathfinder SST a product of higher quality than the daily DLR SST. Quality flags classify the Pathfinder SST from 0 (the worst) to 7 (the best), but there is no linear relation between those levels. The quality flags are given pixel-by-pixel, and they are the result of a series of 'pass' or 'fails' of different quality tests. These tests penalise pixels outside the Reynolds' climatology (Reynolds and Smith, 1994), or those pixels situated at the edges of the scan line. Other tests include the brightness temperature, zenith angle and glint tests. A complete explanation about the Pathfinder products and how these tests were performed can be found in Kilpatrick et al. (2001).

We have classified quality flags under 5 as clouds or missing data. The Pathfinder SST data set used in this work has thus some missing data. Before using them we have reconstructed the missing data with an EOF-based method, called DINEOF (Data INterpolating Empirical Orthogonal Functions, Beckers and Rixen, 2003; Alvera-Azcárate et al., 2005). DINEOF is a parameter-free method that allows to reconstruct missing data in an accurate and fast way. The main advantages of this method is that it does not need *a priori* knowledge about the error statistics of the data, and that it can reconstruct big matrices with up to 30 times less computational time than other classically used methods, as optimal interpolation (Alvera-Azcárate et al., 2005), leading to the same accuracy of the results. The method uses a cross-validation technique to establish the optimal number of EOFs necessary for the reconstruction. DINEOF is thus inherently self-contained but will work well only as long as EOF analysis on a complete data set would work efficiently. In particular, propagating features are generally not well captured by the method and the orthogonality of the modes does not always allow interpreting them physically.

To validate the reconstruction of missing data in the Pathfinder SST data set, we compared them to 256 *in situ* stations from the MEDAR/Medatlas database (MEDAR-Group, 2002). The Root Mean Square (RMS) error between the reconstruction and *in situ* data was 0.81 °C for initially missing points and 0.93 °C for initially existing points. This last error can be considered as the error between the satellite and *in situ* data. Note that the error of the initially missing points is even smaller than the error of the initially existing data.

3.3. Discrete wavelet transforms

In this section we give a brief presentation of the one-dimensional wavelet transform and its main characteristics, used afterwards in the model results analysis. For a comprehensive explanation on wavelets the reader is referred to Daubechies (1992), Mallat (1998) and references therein.

Wavelets are a set of functions generated by a mother wavelet $\psi(x)$. This real function has, among others, the following properties (Kumar and Foufoula-Georgiou, 1997; Torrence and Compo, 1998):

- Unit norm: $\int \psi(x)^2 \mathrm{d}x = 1$.
- Zero mean: $\int \psi(x) \mathrm{d}x = 0$. This characteristic allows a wavelet to extract high-frequency variations of the studied function.
- Most wavelets have a compact support, or a exponential decay to zero. Wavelets allow therefore a localisation

in space in contrast to Fourier modes, which have an infinite support.

From this mother wavelet, a set of orthogonal wavelets $\{\psi_{m,n} \forall m, n \in \mathbb{Z}\}$ is obtained by (Daubechies, 1992):

$$\psi_{m,n}(x) = 2^{-m/2}\psi(2^{-m}x-n) = \frac{1}{\sqrt{2^m}}\psi\left(\frac{x-n2^m}{2^m}\right) \qquad (1)$$

where $\psi(x)$ is dilated and translated by means of parameters m and n. In order to complete this set of functions, we must introduce the scaling function $\phi(x)$ (also called father wavelet) with the following properties:

• Unit norm: $\int \phi(x)^2 dx = 1$.
• Unit mean: $\int \phi(x)dx = 1$. This characteristic allows a scaling function to extract low-frequency variations of the studied function.
• The mother wavelet and the scaling function are orthogonal: $\int \psi(x)\phi(x)dx = 1$.

From this scaling function, a set of functions is obtained by transforming the scaling function $\phi(x)$:

$$\phi_{m,n}(x) = 2^{-m/2}\phi(2^{-m}x-n) \qquad (2)$$

For any M, the set of functions formed by the wavelets $\{\psi_{m,n} \forall m \leq M, n \in \mathbb{Z}\}$ and the scaling function

$\{\phi_{M,n} \forall n \in \mathbb{Z}\}$ is a complete set of orthogonal functions. The discrete wavelet transform consists in projecting a function $g(x)$ onto these basis functions. The term "discrete" refers to the fact that the scale parameter m and the translation parameter n take only integer values. The scalar product between the function $g(x)$ and all the basis functions are the scaling function coefficients $a_{m,n}$ and the wavelets coefficients $d_{m,n}$:

$$a_{m,n} = \int \phi_{m,n}(x)g(x)dx \qquad (3)$$

$$d_{m,n} = \int \psi_{m,n}(x)g(x)dx \qquad (4)$$

for all m, $n \in \mathbb{Z}$ (with $\mathbb{Z}$ the ensemble of integer numbers) and $m \leq M$. Each of these coefficients is related to a scale and a location given by parameters m and n. This property makes wavelets very appropriate for multi-resolution analysis (Mallat, 1998).

From the coefficients $d_{m,n}$ and $a_{m,n}$, the original function can be reconstructed by:

$$g(x) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{M} d_{m,n}\psi_{m,n}(x)$$
$$+ \sum_{n=-\infty}^{\infty} a_{M,n}\phi_{M,n}(x) \qquad (5)$$

In particular if $M \to \infty$, for a function with zero mean:

$$g(x) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} d_{m,n}\psi_{m,n}(x) \qquad (6)$$

The approximation $A_m(x)$ and the details $D_m(x)$ are defined for all $m \leq M$ by:

$$A_m(x) = \sum_{n=-\infty}^{\infty} a_{m,n}\phi_{m,n}(x) \tag{7}$$

$$D_m(x) = \sum_{n=-\infty}^{\infty} d_{m,n}\psi_{m,n}(x) \tag{8}$$

The functions A_m represent the variations larger or equal than the scale m, and functions D_m represent the variations smaller than m. Rewriting Eq. (5), a function g(x) can be decomposed into separate scales or resolution levels:

$$g(x) = D_1(x) + D_2(x) + ... + D_M(x) + A_M(x) \tag{9}$$

A discrete wavelet transform can be viewed as a hierarchical decomposition that iteratively filters the high-frequency signal out of the large-scale signal. The approximation of the previous scale is decomposed into approximation and details of the current scale,

$$A_0(x) = g(x) \tag{10}$$

$$A_m(x) = A_{m+1}(x) + D_{m+1}(x) \tag{11}$$

The wavelet decomposition can also be applied to a vector g(x_i) with $i = 1 .. N$, $N = 2^M$ (i. e. a vector of dyadic length). The vector g(x_i) can be decomposed into M scales. The wavelet coefficients are then obtained from the discrete wavelet $\psi_{m,n}(x_i)$ and scaling function $\Phi_{m,n}(x_i)$ As before, indexes m and n determine the dilation and translation of the wavelet and scaling function. The approximation and detail coefficients are represented by:

$$a_{m,n} = \sum_{i=1}^{N} \phi_{m,n}(x_i)g(x_i) \tag{12}$$

$$d_{m,n} = \sum_{i=1}^{N} \psi_{m,n}(x_i)g(x_i) \tag{13}$$

where $m = 1...M$ and $n = 1...2^{M-m}$. For the largest scale of the decomposition M, the index n is dropped for the approximation coefficient (a_M) and the scaling function ($\Phi M(x_i)$) in order to simplify the notation. The approximation coefficient a_M is related to the mean value of the vector $\bar{g}$, since:

$$\phi_M(x_i) = \frac{1}{\sqrt{N}} \tag{14}$$

$$a_M = \sqrt{N}\bar{g} \tag{15}$$

The decomposition of the vector g(x_i) into M scales is given by:

$$g(x_i) = \sum_{m=1}^{M} \sum_{n=1}^{2^{M-m}} d_{m,n}\psi_{m,n}(x_i) + a_M\phi_M(x_i) \tag{16}$$

We will work with the detail coefficients at all scales and the approximation at the largest scale, as they retain the necessary information at each scale to reconstruct the initial function. There are different interpretations of the length scale associated to the detail coefficients. Here we follow the convention given by Lindsay et al. (1996) and Casati et al. (2004), which interprets the details resolution as the grid resolution. The reader should be aware that some authors (*e.g.* Yano et al., 2001) define the details resolution as the shortest wavelength that can be represented by the wavelet on a grid. Both conventions differ by a factor of two. If Δx is the initial grid

spacing of the function $g(x)$, then we define the resolution of the details coefficients at scale m as $\Delta x 2^{m-1}$.

The explanation on wavelet characteristics given previously can be extended to two-dimensional wavelets, which are obtained by the orthogonal product of one-dimensional wavelets and scaling function. The detail coefficients, however, are divided into horizontal, vertical and diagonal details that account for different spatially oriented frequency channels (Mallat, 1989; Kumar and Foufoula-Georgiou, 1993).

The wavelet chosen to decompose a data set will undoubtedly affect its decomposition. This is why a careful choice of the mother wavelet is very important. Torrence and Compo (1998) name several choices that should be taken into account when choosing the wavelet: *i*) orthogonal or non-orthogonal wavelets; *ii*) complex or real wavelets (complex wavelets are also largely used in geophysics, *e.g.* Spedding et al. (1993), Chapa et al. (1998), Rao and Murthy (2001); *iii*) the width of the wavelet (or compact support); and *iv*) the shape of the wavelet; in fact, the shape of the wavelet should be chosen accordingly to the characteristics of the data to analyse, because the results change when using different wavelet forms. There are other methods to choose the more adequate waveform. In Briggs and Levine (1997), the wavelet that minimises the entropy of the wavelet transform is chosen. Casati et al. (2004) worked with binary rain fields, and the square-shaped Haar wavelet fitted the best to the data. In this work, two-dimensional Haar wavelets are used.

In Oceanography, there is an additional degree of difficulty when using two-dimensional wavelets. Oceanographic domains often contain irregular boundaries at the limit between land and sea, which is not the case in the atmosphere, for example. We have treated land-sea boundaries as an extension of the one-dimensional case: when the analysed time series does not have a dyadic length, it is normally padded with extra values up to the nearest dyadic length. These added points can be constructed in many different ways, *e.g.* with zeros, or a periodic extension of the original data, a gently decrease to zero, the mean value of the variable, etc. The same can be done in two dimensions to reach the dyadic length trying to minimise the perturbation of the wavelet coefficients at the land-sea boundary. Land points have been replaced by the mean value of the field, and discontinuities at the land-sea boundary have been reduced by the following iterative procedure:

$$g_{i,j}^{(s+1)} = g_{i,j}^{(s)} + \alpha(4g_{i,j}^{(s)} - g_{i+1,j}^{(s)} - g_{i-1,j}^{(s)}$$
$$-g_{i,j+1}^{(s)} - g_{i,j-1}^{(s)}) \qquad \text{if } i,j \text{ is a land point} \qquad (17)$$

$$g_{i,j}^{(s+1)} = g_{i,j}^{(s)} \quad \text{elsewhere} \qquad (18)$$

with $\alpha = 0.05$, and $N_s = 500$ iterations performed. The width of the transition zone is approximately $\sqrt{\alpha N_s} = 5$ grid points. This is a very simple approach, but it has proven useful in our case. Future work should include more specific methods for irregular boundary treatment, as the approach by Oh et al. (2001) and Naveau and Oh (2004), based on a polynomial wavelet regression, specially designed to deal with irregular boundaries. Second generation wavelets (Sweldens, 1997; Daubechies et al., 1999) can be very helpful when working on irregular grids, and can also deal with irregular boundaries.

As the wavelet is translated along the domain, it crosses land-sea boundaries. The transition from sea to land points results in perturbed wavelet coefficients, even after applying Eq. (17). For this reason, the use of a small support wavelet will help to reduce this effect. Wavelets have a determined compact support, or effectively non-zero zone. A wide wavelet affects a higher number of data points than a narrow wavelet. For example, Daubechies Wavelet family has a compact support of $2N_o-1$, with N_o the order of the wavelet. The Haar Wavelet has a support width of 2, the smallest support width of all the wavelet families. For this reason, the Haar Wavelet has been chosen for this work to reduce the boundary effects.

3.3.1. Wavelets and RMS error decomposition

As previously shown, wavelets form a complete basis of orthonormal functions. Therefore the L_2 norm of a given function $g(x_i)$ can be decomposed into the sum of the square of its wavelets coefficients. In a discrete form, this relationship can be written as:

$$\sum_{i=1}^{N} g(x_i)^2 = \sum_{m=1}^{M} \sum_{n=1}^{2^{M-m}} d_{m,n}^2 + a_M^2 \qquad (19)$$

where $g(x_i)$ is the value of the function at point i; N is the total number of spatial points; $d_{m,n}$ are the detail coefficients at scale m and point n; and a_M is the approximation coefficient at scale M, the coarsest of the decomposition. Given any size of the initial decomposed matrix, the term a_M is a single number, which is related to the mean of the initial vector.

If the vector $g(x_i)$ is the difference between a forecast $F(x_i)$ and observations $O(x_i)$, we have:

$$\sum_{i=1}^{N} (F(x_i) - O(x_i))^2 = \sum_{m=1}^{M} \sum_{n=1}^{2^{M-m}} (d_{m,n}^F - d_{m,n}^O)^2 + (a_M^F - a_M^O)^2 \qquad (20)$$

with $d_{m,n}^F$ the detail coefficients of the forecast at scale m and point n, $d_{m,n}^O$ the detail coefficients of the observations, and a_M^F and α_M^O the approximation of the forecast and the observations respectively. Dividing Eq. (20) by the total number of points N of the vector we obtain:

$$\frac{1}{N} \sum_{i=1}^{N} (F(x_i) - O(x_i))^2 = \frac{1}{N} \sum_{m=1}^{M} \sum_{n=1}^{2^{M-m}} (d_{m,n}^F - d_{m,n}^O)^2 + \frac{1}{N} (a_M^F - a_M^O)^2 \qquad (21)$$

We can define the squared RMS of the wavelet decomposition at scale m as

$$\text{RMS}_m^2 = \frac{1}{N} \sum_{n=1}^{2^{M-m}} (d_{m,n}^F - d_{m,n}^O)^2 \qquad (22)$$

The term $\frac{1}{N} (a_M^F - a_M^O)^2$ of Eq. (21) is the square of the bias. Eq. (21) can be rewritten as the sum of the squared centred RMS and the squared bias (B):

$$\text{RMS}^2 = \sum_{m=1}^{M} \text{RMS}_m^2 + B^2 \qquad (23)$$

We have divided the RMS error of the forecast $F(x_i)$ into the sum of the RMS error of all spatial scales of the wavelet decomposition plus a bias term $\left(B = \frac{1}{\sqrt{N}} (a_M^F - a_M^O)^2 \right)$. This decomposition is an extension to the largely known RMS error decomposition into centred RMS and bias (see, *e.g.* Murphy and Epstein, 1989; Taylor, 2001). It is apparent from Eq. (23) that the sum of RMS_m^2 over all scales m is equal to the centred RMS.

3.3.2. Taylor diagram

A wavelet decomposition is very useful to study the model results at different resolution levels. However, the results can be complex to analyse, because there are many factors that influence each scale, a different behaviour for each scale and error measure, and the evolution with time of those scores. A Taylor diagram (Taylor, 2001) can be very useful in this situation, because it groups several error measures into one graphic, allowing a joint interpretation of the results. The Taylor diagram has been already used in the study of atmospheric models (Gates et al., 1999; Boer and Lambert, 2001; Denis et al., 2003) and in the ocean (*e.g.* Raick et al., 2007-this issue). The Taylor diagram is based on the relation between the RMS error, the centred RMS error and bias *(B)*:

$$\text{RMS}^2 = B^2 + \text{RMS}'^2 \qquad (24)$$

then the centred RMS (RMS') can be expressed as:

$$RMS'^2 = S_f^2 + S_o^2 - 2S_f S_o r \qquad (25)$$

where r is the correlation and S_f and S_0 are the variance of the forecast and the observations respectively NoNote the relation between Eq. (25) and $\left(S_f = \frac{1}{N-1}\sum_{n-1}^{N}(x_n^f - \bar{x})^2\right)$, the law of the cosines:

$$c^2 = a^2 + b^2 - 2ab \cos\phi \qquad (26)$$

where a, b and c are the sides of a triangle and φ is the angle opposite to c. With this relation, we can easily plot into one diagram several statistics of a field (their standard deviations, the correlation and the RMS'). The mentioned statistics are plotted as follows:

• The standard deviation is plotted as the radial distance to the graphic origin.

• The correlation between both fields is given by the azimutal position.

• The RMS' is the linear distance between the observations and the model points.

The standard deviation and the RMS' can be normalised by the standard deviation of the observations $(R\hat{M}S' = RMS'/S_o;\ \hat{S}_f = S_f/S_o\ \text{and}\ \hat{S}_o = 1)$, so that the observations lay in the abscissa axis, with a normalised standard deviation of 1. The other variables are plotted relative to the observations. A point represented this way into the graphic has the same standard deviation as the observations if it is situated at a distance of one from the origin. The standard deviation is higher if this distance is higher, and the opposite if the distance is smaller than one. A large angle separating the model from the observations indicates a bad correlation.

4. Model skill at the basin and sub-basin scales

The first test we will perform consists in studying the impact of the assimilation of DLR SST data at each assimilation cycle, to identify weaknesses or problems that may appear because of the assimilation itself. We compare the DLR SST data (with a resolution of ~ 1 km) to the Ligurian Sea model (with a resolution of 1.5 × 1.9 km) and in the Mediterranean Sea model (with a resolution of 22 × 28 km). For the comparison with the Mediterranean Sea model, the DLR SST resolution is degraded to 10 km. This study will give us an idea about how the model behaves at the basin and sub-basin scales, and how this error evolves in time.

Figs. 2 and 3 show the evolution of the SST bias of Mediterranean Sea and Ligurian Sea models respectively. The first thing that may be noticed is that the (lower resolution) Mediterranean Sea model has a smaller bias than the (higher resolution) Ligurian Sea model. The bias in the Mediterranean Sea model (Fig. 2) is always smaller than 0.5 °C. However, the improvement of the *Forecast* over the *Free run* is small. The assimilation of SST seems to have little impact on the Mediterranean Sea model bias. In the Ligurian Sea model (Fig. 3) the variations of bias over time are more evident. The *Forecast* has a smaller bias than the *Free run* over the first seven weeks of the experiment, with improvements as high as 0.5 °C. However, during the two last weeks of the experiment, the *Free run* has a much smaller bias than the *Forecast*. We can better understand what happens at the end of the experiment by looking at the temporal evolution of the mean SST over the Ligurian Sea for each model run and the observations (Fig. 4). In this figure we see that the *Free run* is systematically colder than the observations. The assimilation of the SST brings the model nearer to the actual temperature, so the *Forecast* is nearer to the observations than the *Free run*. On 24 August the *Free run* has a smaller bias than the *Forecast*, due to the combined action of two errors that, by coincidence, cancel each other: a too cool SST on 17 August and an incorrect warming trend contradicting the observations. The *Forecast*, on the other hand, has a correct mean SST on 17 August after the assimilation (the *Analysis*), but the erroneous warming trend results in a too high mean SST on 24 August. An increased assimilation rate (for example, each five days instead of each week), would probably improve the *Forecast* results, avoiding a large divergence between the model and the observations. The *Free run* presents a decreasing trend similar to the *Forecast* from 24 to 31 August, which indicates that the cooling episode at the end of the experiment is also present in the atmospheric forcings. However, there is a week of delay between the beginning of the decreasing trend in the observations and the model results. This delay can be caused by a late response of the model to the atmospheric forcings or by an actual delay between the observed SST and the atmospheric forcings.

Fig. 2: *Bias time evolution for the SST of the Forecast, Free run and Analysis in the Mediterranean Sea model.*
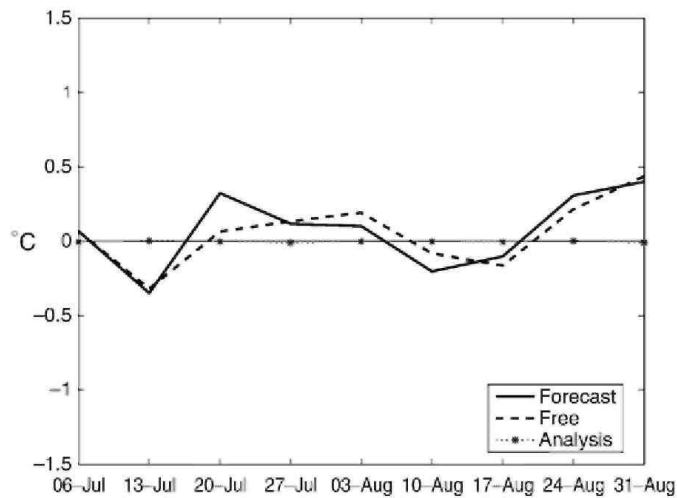
Figs. 5 and 6 show the RMS error of the Mediterranean Sea and the Ligurian Sea models respectively. The Mediterranean Sea model *Forecast* has a smaller error than the *Free run* during the whole experiment, with improvements as high as 0.5 °C. The assimilation has a larger effect on RMS error than on the bias. The RMS error on the Mediterranean Sea model is overall better than the RMS error on the Ligurian Sea model. This is due to an effect called double penalty (Hoffman et al., 1995; Ebert and McBride, 2000). Low-resolution models represent the reality in a smooth way, without much of the observed small-scale variability. High-resolution models represent this small-scale variability more realistically, in the form of eddies, meanders, etc. Even if the later model looks more realistic, it can account for a bigger error than the low-resolution model, because of the incorrect positioning of the small-scale features. The high-resolution model has more degrees of freedom than the low-resolution model. These additional degrees of freedom correspond to small spatial scales and have thus a limited predictability.

Fig. 3: *Bias time evolution of Forecast, Analysis and Free run SST in the Ligurian Sea model.*
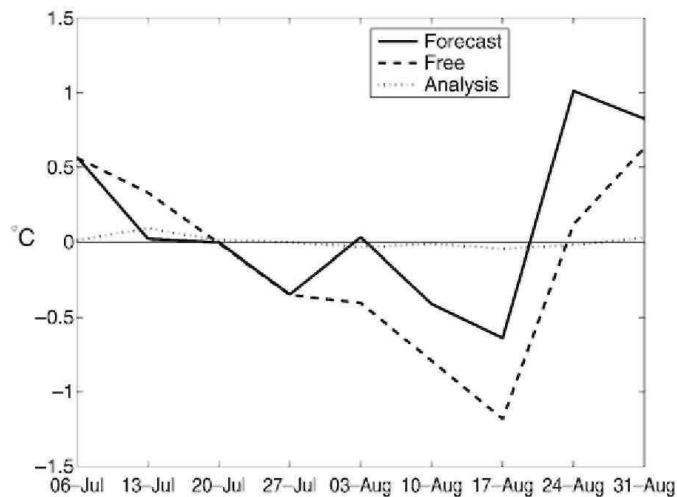
The Ligurian Sea model RMS error (Fig. 6) reflects again the cooling episode error on 24 August, but overall the *Forecast* performs better than the *Free run*. A significant fraction of the RMS error in the Ligurian Sea model is due to the bias. It is therefore interesting to decompose the RMS error into the contribution of the bias and the RMS error based on observations and model SST without their mean. The later RMS is called centred RMS, and one can show that RMS, bias and centred RMS are related by Eq. (24). The centred RMS error (Fig. 7) shows that the bias (as shown in Fig. 3) is the major source of error on 17 and 24 August. At the beginning of the experiment the RMS error is mainly due to the SST distribution (high centred RMS), while at the end of the experiment the SST bias is more important than the error in the temperature distribution of the Ligurian Sea. The origin of the error in the Ligurian Sea (distribution of the SST or mean value) is therefore variable in time. A

bias at the scale of the Ligurian Sea is more likely to be associated to errors in the heat fluxes, and an erroneous SST distribution may result from misrepresented hydrodynamic features such as frontal positions, phase of meanders and position of eddies.

Fig. 4: *Time evolution of the mean SST in the Ligurian Sea model. The mean SST of the Analysis is almost identical to the observed mean SST.*
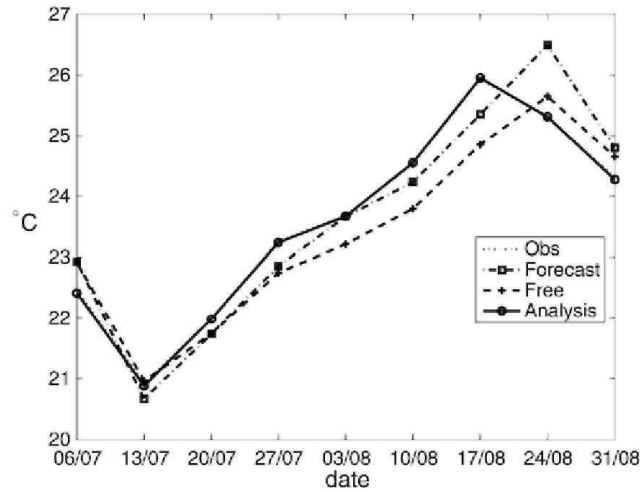

Fig. 5: *RMS error time evolution for the SST of the Forecast, Free run and Analysis in the Mediterranean Sea model.*
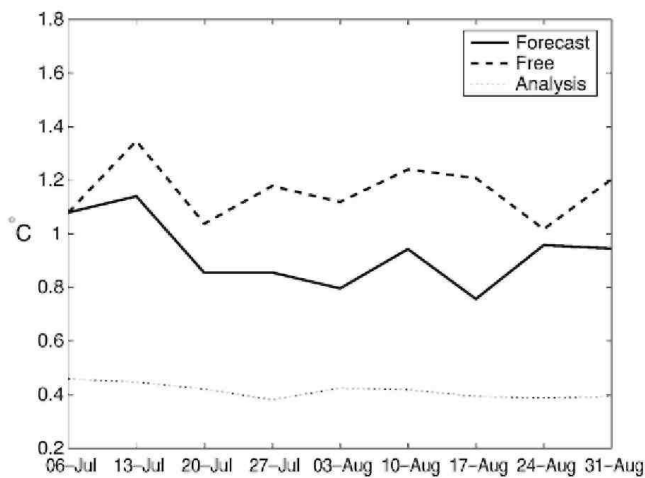
Fig. 6: *RMS error time evolution of Forecast, Analysis and Free run SST in the Ligurian Sea model.*
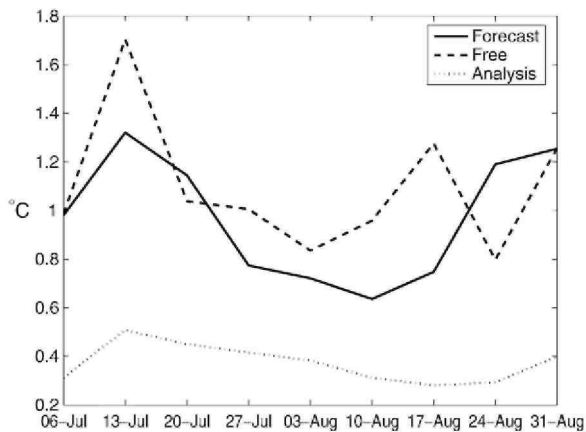
5. Spatio-temporal analysis using Wavelet Transforms

The error analysis of the Ligurian Sea model is pursued with a wavelet transform. This decomposition emphasises the distribution of the error at different spatial scales, related to distinct physical processes. We have decomposed the Ligurian Sea model and DLR SST observations into eight resolution levels with a two-dimensional discrete Haar wavelet. The model has been interpolated to the resolution of the observations (~ 1 km). As the initial model resolution is 1.5 × 1.9 km, we must bear in mind that the model is not able to resolve features smaller than its resolution when interpreting the results of the wavelet decomposition. The resolution levels reached by the wavelet decomposition are 1, 2, 4, 8, 16, 32, 64 and 128 km. The decomposition allows us to analyse the error at each of these spatial scales. The figures presented in this section that do not have information about the time distribution are time averages.

Fig. 7: *Centred RMS error time evolution for the SST of the Forecast, Free run and Analysis in the Ligurian Sea model.*
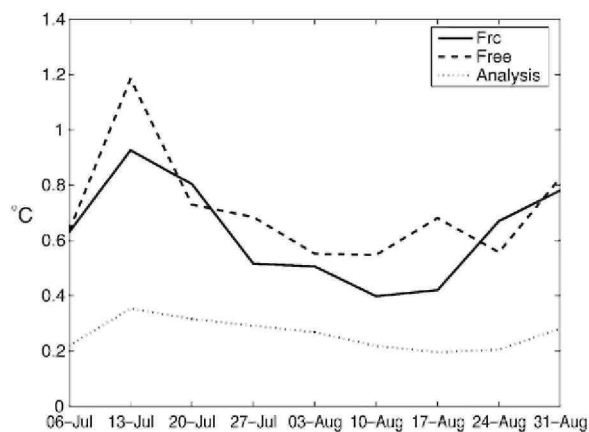
Fig. 8: *Time average RMS error at each spatial scale including the bias term.*
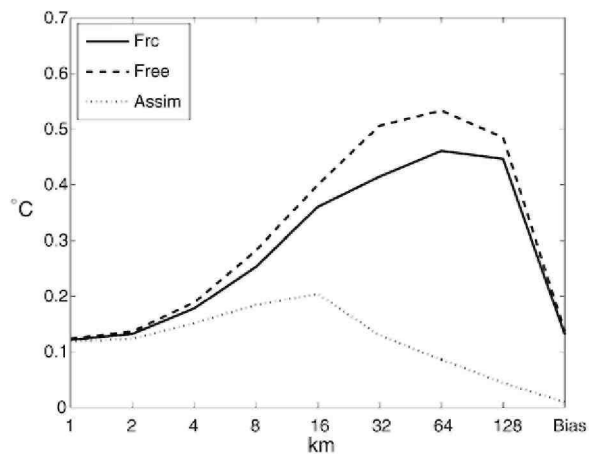
The RMS error for each wavelet scale is depicted in Fig. 8. The highest error reduction from the *Free run* to the *Forecast* occurs from 16 to 128 km. The *Analysis* error shows a maximum at 16 km. The fast decay of the *Analysis* RMS error from this maximum to larger scales suggest that the assimilation is mainly impacting large scales, which we can explain as follows: the structure of the correction from the assimilation is determined by the observations and the model error covariance. The error covariance acts as a filter and controls which scales of the model are corrected by the assimilation. The error covariance matrix is obtained by an ensemble simulation and is therefore neither invariant in time nor isotropic. The model error covariance is limited to a maximum correlation length of 47 km, as explained in Barth et al. (2007-this issue). This maximum correlation length corresponds to scales 6 to 7 (32 to 64 km) in the wavelet decomposition. Corrections are not strictly limited to 47 km, but their amplitude decreases significantly for scales smaller than scale 6. Coherent features

with length scales higher than 47 km are also corrected. This explains the peak observed in the *Analysis* RMS error (Fig. 8) : the correction acts mainly at scales larger than 32 km, so scales 16 and smaller are less affected by the assimilation.

Fig. 9: *Percentage each scale, including the bias term, contributes to the time and space average total RMS error.*
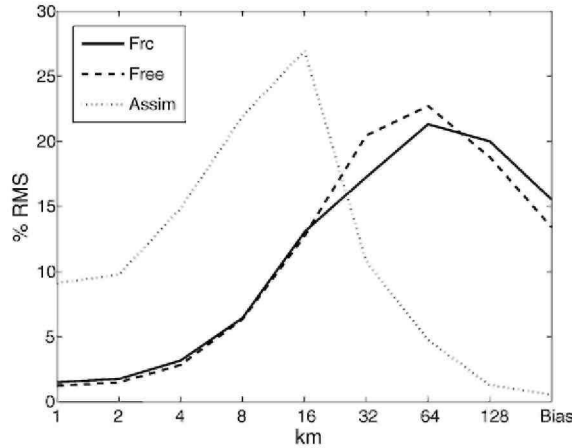

Table 1: *RMS^2 at each scale, including the bias term, for Forecast, Free run and Analysis wavelet coefficients, and the total RMS^2 of the corresponding model run*

Scale (km)	1	2	4	8	16	32	64	128	Bias	RMS^2 model
Free	0.01	0.02	0.04	0.08	0.16	0.26	0.29	0.23	0.17	1.26
Forecast	0.01	0.02	0.03	0.06	0.13	0.17	0.21	0.20	0.15	1.01
Analysis	0.01	0.01	0.02	0.03	0.04	0.02	0.01	0.002	0.001	0.14

Units are °C^2.

Fig. 10: *Standard deviation of the correction, model trend and trend of the error for each scale. Scale 256 stands for the basin-wide contribution calculated from the approximation at the highest scale of the decomposition.*
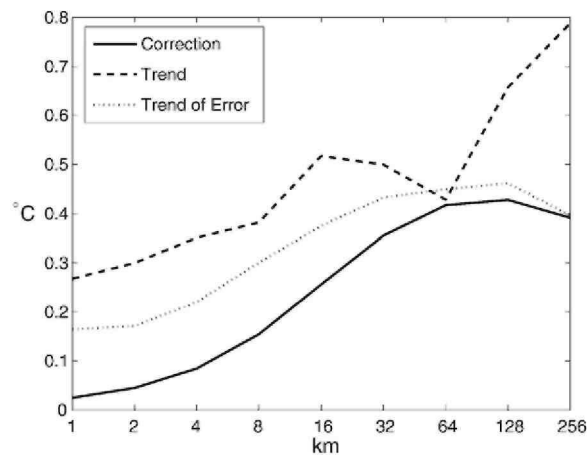
The percentage of RMS^2 (in Fig. 9 and Table 1) shows how the total RMS^2 is divided among scales, or how each scale contributes to the total RMS error, according to Eq. (23). The distribution of RMS is very similar in the *Free run* and in the *Forecast,* with scales 32 and 64 km of the *Forecast* contributing in a lesser amount to the total error than in the *Free run,* due to the correction of the assimilation. Scales smaller than 32 km contribute

little to the total error. Most of the SST model error is located at large scales, and is efficiently corrected by the assimilation. This finding basically confirms the choice of the maximum correlation length, but a small improvement can be expected by choosing a lower value of the maximum correlation length, since scale 16 km still accounts for 13% of the total error of the *Forecast*. This correlation length should not be smaller than 16 km, as scales 1 to 8 km have each a small impact on the total error of the model.

Fig. 11: *Evolution in time of the standard deviation of the Forecast trend of the model error.*
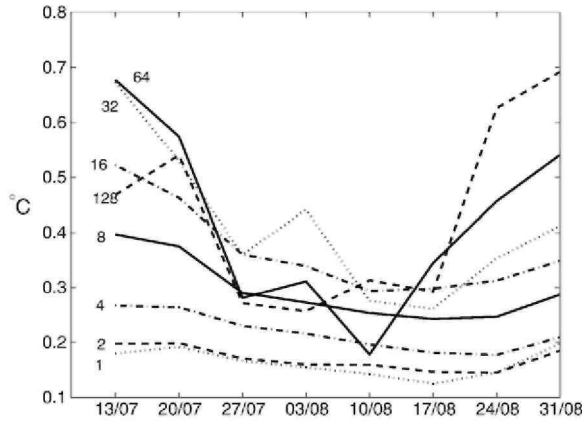
Fig. 12: *SST for (a) observations, (b) Analysis, (c) Free run and (d) Forecast on 13 July 2000.*
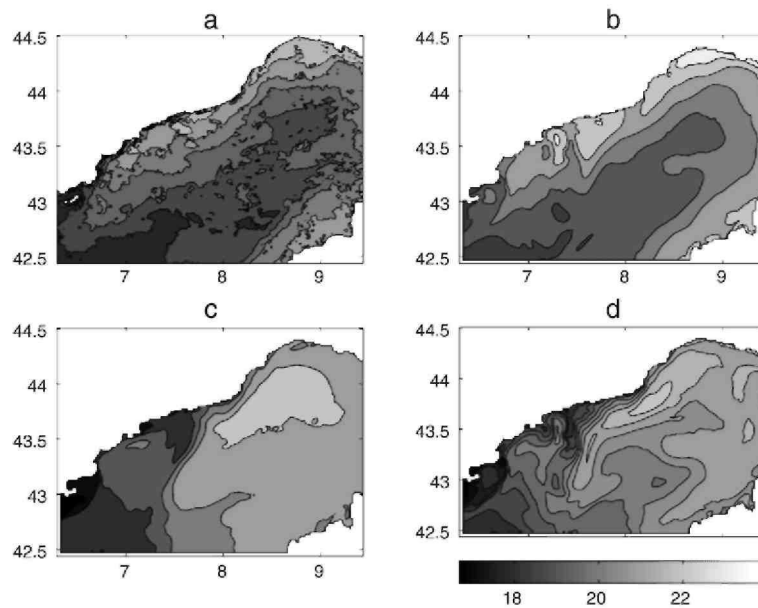
The time evolution of the *Analysis* run is the cumulated effect of the "free" model evolution and the correction applied by the assimilation. Since both effects have very different nature, it makes sense to study them separately. The correction ($CR_t = A_t - F_t$, with A_t and F_t the *Analysis* and the *Forecast* at time step t) and the model trend ($T_t = F_t - A_{t-1}$, with A_{t-1} the *Analysis* at time step t-1) are calculated at each scale for the two-dimensional SST fields:

$$CR_{m,t} = A_{m,t} - F_{m,t} \qquad\qquad (27)$$

$$T_{m,t} = F_{m,t} - A_{m,t-1} \qquad\qquad (28)$$

with $CR_{m,t}$ the correction, $T_{m,t}$ the model trend, $A_{m,t}$ the *Analysis* and $F_{m,t}$ the *Forecast* at time t and scale m. We

will compare these measures to the trend of the model error TE_t, also decomposed at each scale:

$$TE_{m,t} = (F_{m,t} - O_{m,t}) - (A_{m,t-1} - O_{m,t-1}) \qquad (29)$$

with $TE_{m,t}$ the trend of the model error, and $O_{m,t}$ the observations at time t and scale m. TE measures how the error of the model changes with time for each of the resolution scales. The standard deviation of these three measures is shown in Fig. 10.

The assimilation mainly corrects scales larger than 32 km, as seen by the rapid increment of the standard deviation of the correction. This explains the error reduction from the *Free run* to the *Forecast* RMS error at those scales seen in Fig. 8. The correction can be viewed as the part of the trend of the model that is considered error by the assimilation scheme. The trend of the model, on the other hand, is also higher at large scales, because the large scales in the SST have higher variability than the small scales. There is a peak in the trend of the model at 16-32 km, corresponding to the meso-scale, which presents a high activity in the Ligurian Sea related to the formation of small eddies and meanders along the NC. The high values at 128 km and the mean term account for the basin-wide trend of summer heating. The trend of the *Free run* (not showed) has a similar distribution, so the maximums at 16 to 32 km and the basin-wide scale are not introduced by the assimilation. Except for the scale 64 km, the standard deviation of the model trend is higher than the standard deviation of the trend of the model error and of the correction. This is an indication of skill of the model at all scales except 64 km.

It is also interesting to see the evolution in time of the standard deviation of the trend of the model error at each scale (Fig. 11). Scales 1 to 16 km present a steady pattern with higher values of the trend of the model error with the increasing scale. Larger scales (32 to 128 km) have a variable contribution to the total error. We can identify the scales that cause the different errors found in Section 4. For example, scales 32 and 64 km dominate the trend of the model error on 13 July, when the main error found in the model was due to RMS error (Fig. 6). On the other hand, on 24 August scale 128 km is the most important source of error in the model, followed by the 64 km scale. On this day, the bias was the most important source of error on the model (Fig. 3).

We have shown that the impact of the assimilation is stronger at scales larger than the correlation length. Correction at scales of 64 km and larger accounts for basin-wide changes, as the general temperature distribution in the Ligurian Sea. We will look now at how the correction affects the small scales. We take as an example the SST on 13 July, which is a day with a particular high RMS error in both the *Forecast* and *Free run*. In Fig. 12 the SST distribution for the observations, *Free run, Forecast* and *Analysis* is shown. The observations (panel a) show a high variability, that the model *(Free run* in panel c and *Forecast* in panel d) fails to represent. The general circulation of the Ligurian Sea is clearly seen in the observations, with the warm NC describing a cyclonic gyre along the Italian and French coasts, and with colder temperatures in the centre of the basin. The *Free run,* in panel c, fails to represent the described main circulation, which is the cause of the high error obtained this day. The *Free run* SST distribution is very different from the observations, and the variability is smaller. The *Forecast,* in panel d, has a better SST distribution, and the path of the NC can be detected. The *Analysis,* in panel b, is the result of the assimilation of the observations shown in panel a into the *Forecast* shown in panel d. The Ligurian Sea SST in the *Analysis* is close to the observations since the observed SST has been assimilated. The spatial distribution of the correction for this day at scale 4 km is shown in Fig. 13. The 4 km scale has been chosen as representative for the small-scale, as smaller scales may not have been well resolved by the model. The assimilation is mainly correcting the SST along the French coast and the NC, which is clearly mispositioned in the model (too far offshore). The misposition of the NC induces also an error on small scales. The position of this small-scale error is clearly linked to the large-scale current. From a physical point of view, the interpretation is straight forward, since the small-scale features are generated at the large-scale fronts. Such an analysis could not be possible with Fourier decomposition, since the location in space is not retained in the Fourier space.

Fig. 13: *Spatial distribution of the standard deviation of the correction of the Forecast by the assimilation on 13 July 2000 at scale 4 km. The magnitude of the wavelet coefficients ($\sqrt{\sum_{v=1}^{3} D_v^2}$, where D_1, D_2 and D_3 correspond to the horizontal, vertical and diagonal direction of details) is used in this figure to calculate the standard deviation of the correction.*
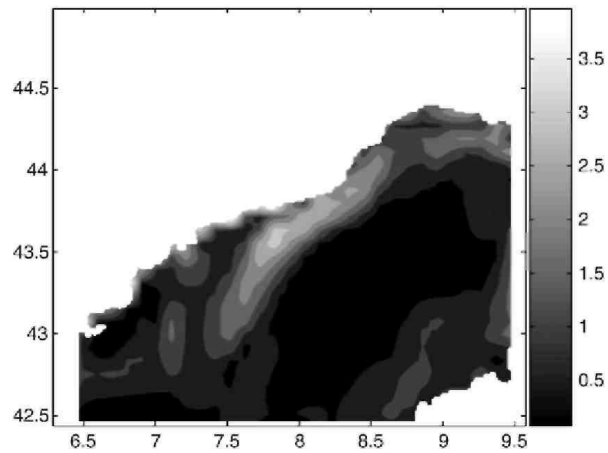

Fig. 14: *Spatial distribution of the standard deviation of the correction induced by the assimilation at scale 4 km for the 9 weeks of the experiment. The standard deviation of the correction is calculated as specified in Fig. 13.*
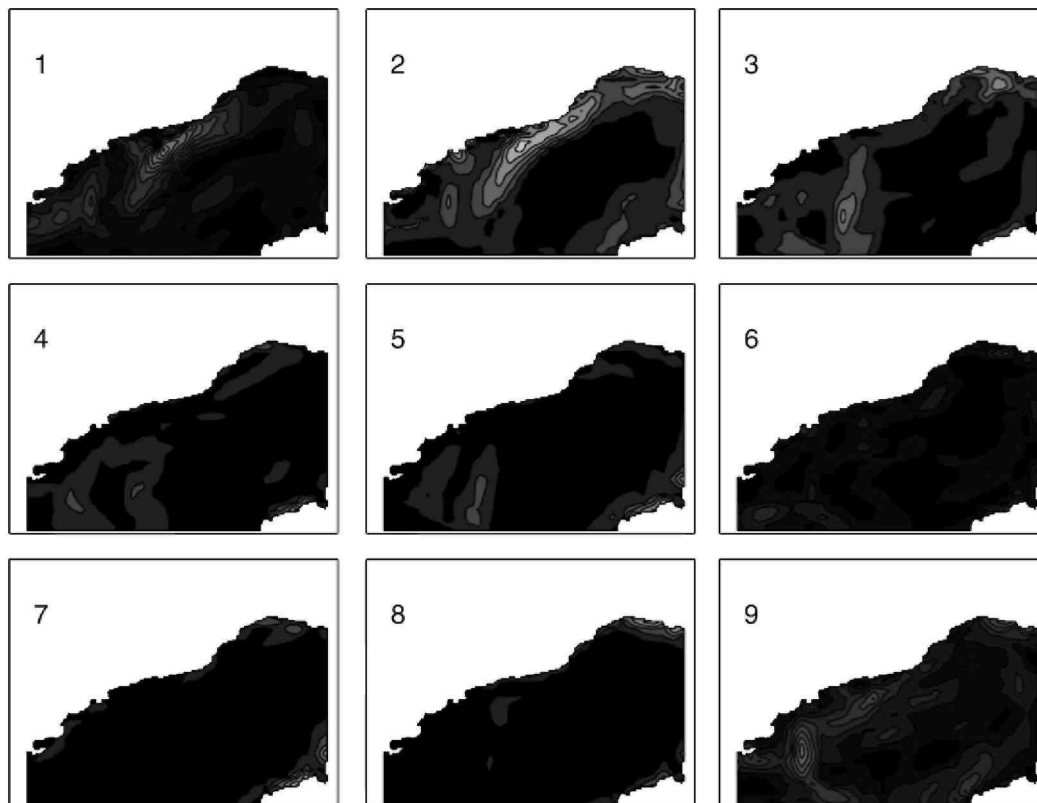
Fig. 15: *Taylor diagram for Free run (in grey), Forecast (in black) and Analysis (in white).*
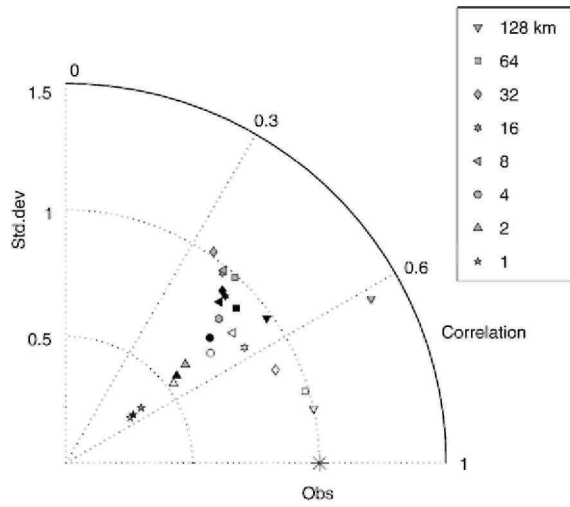
Fig. 16: *Standard deviation distribution for the Forecast, Free run, Analysis and observed SST.*
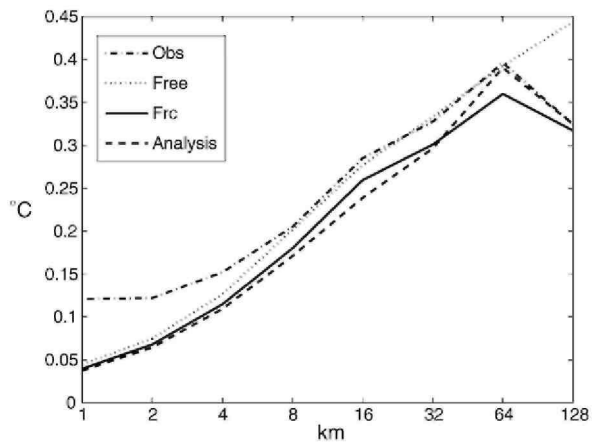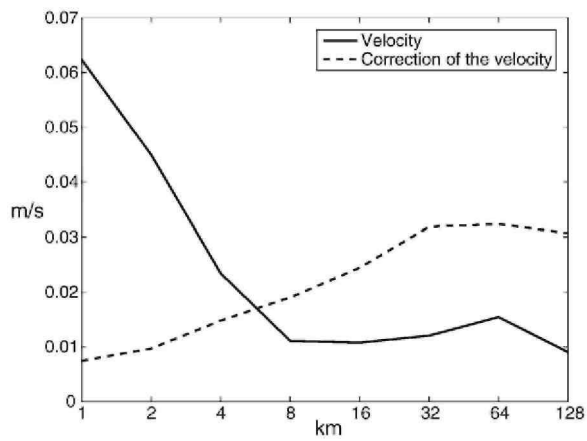

Fig. 17: *Standard deviation distribution of the surface velocity and the correction of surface velocity induced by the assimilation.*

The correction on the other days of the model run is shown in Fig. 14. We can see that the correction around the NC is done during the two first weeks of the experiment. Other zones that are corrected by the assimilation are the southwest corner of the model domain and the northern coast of Corsica. The correction in the centre of the Ligurian Sea is very small, because the variability of this zone is smaller than along the path of the coastal cyclonic currents (Northern Current, Eastern Corsican Current and West Corsican Current).

Fig. 15 contains the Taylor diagram for *Free run, Forecast* and *Analysis* for the eight scales of the wavelet decomposition. This diagram displays the standard deviation, correlation and centred RMS of the model runs and compares it to observations (see description of the Taylor diagram in Section 3.3.2). The main advantage of this diagram is that it summarises several error measures into one graphic, and we can have a general vision of the model skill. Scales 16 to 128 km are very near to the observed standard deviation, but their correlation and centred RMS is worst than for smaller scales. In addition, in large scales the impact of assimilation is higher: the *Forecast* is systematically nearer to the observations than the *Free run.* At small scales, the assimilation is causing a decrease on the *Forecast* standard deviation (the *Free run* has already a small standard deviation, but this is accentuated by the assimilation). The better correlation and centred RMS observed at small scales indicates that they are able to represent more accurately the observations, although the effect on standard deviation must be further studied.

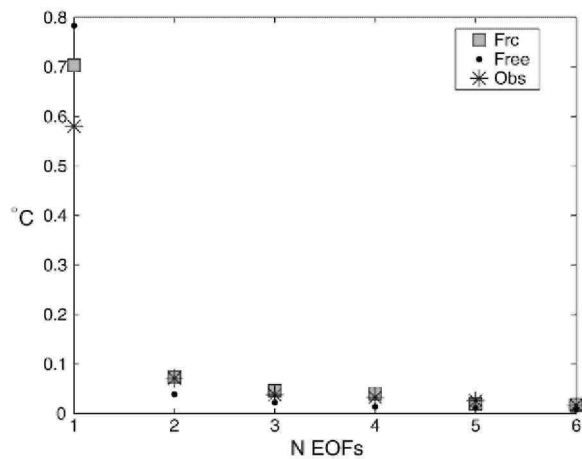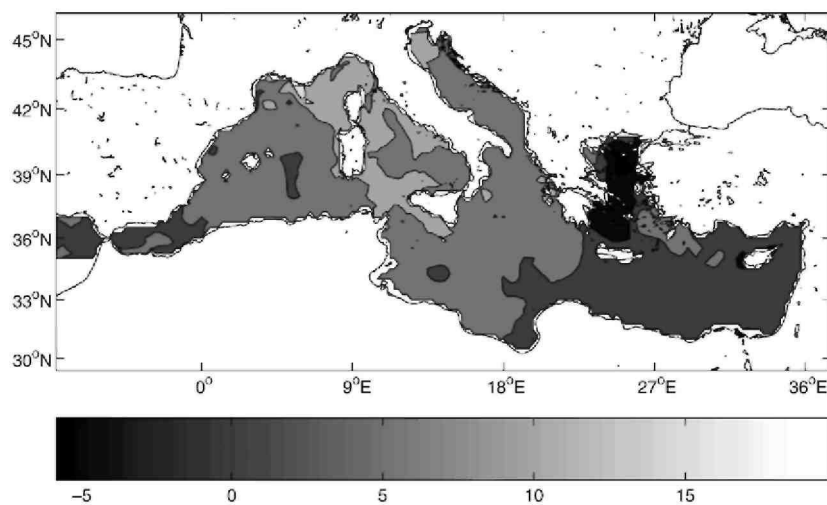**Fig. 18:** *Variance of Free run, Forecast and observations for each EOF mode.*

Fig. 19: *First spatial mode for the Free run.*

The causes for the decrease in the SST standard deviation can be multiple. For example, a strong horizontal diffusion can eliminate the small features from the model results. A strong vertical mixing can also weaken the surface variability by increasing the thermal inertia of the surface water, and thus damp the horizontal signal of small-scale features. Again, the model is not able to represent features smaller than 2 km, and the small scales close to the model resolution are not well resolved by it. In addition, the small-scale information added by the assimilation of high-resolution data is related to the correlation length of the assimilation scheme. Scales smaller than this correlation length are not modified by the assimilation. A certain amount of variance at small scales in the observations is also due to noise, not present in the model.

The variance distribution at different spatial scales of the wavelet decomposition depends strongly on the nature of the physical variable. The study of the standard deviation of two different variables, SST and currents, points out some of these differences. The SST standard deviation distribution among scales for each model run, *Analysis* and observations is shown in Fig. 16. The *Analysis* intersects the *Forecast* at about 32 km, so the assimilation is causing a decrease of variance at scales smaller than about the maximum correlation length. The *Free run* has a higher variance than the *Forecast* at all scales, so the decrease in variance at small scales is transmitted to large scales in the *Forecast*. Both *Free run* and *Forecast* present a high difference with observations at scales 1 to 3 (1 to 4 km), for the reason stated above.

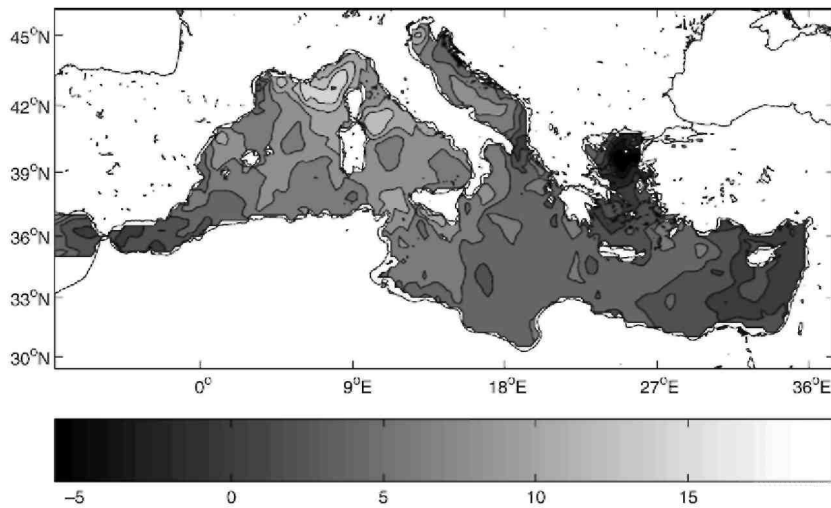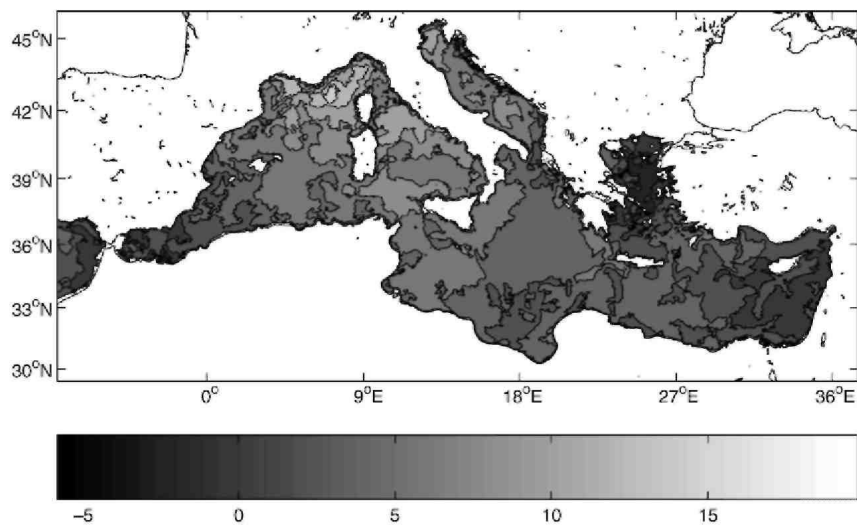**Fig. 20:** *First spatial mode for the Forecast.*

Fig. 21: *First spatial mode for the observations.*

We have compared the SST standard deviation of Fig. 16 to the surface velocity standard deviation, shown in Fig. 17. The surface velocity is directly related to the kinetic energy. The correction applied to the currents after the assimilation of SST (Eq. (27)) is also represented in this figure. The distribution of the velocity standard deviation (higher values for small scales) is very different to the SST distribution. The SST is mainly influenced by the atmospheric forcings, which have a typical variability scale larger than the variability scales found in the ocean. The standard deviation distribution of the surface velocity reflects the variability scales typical of the ocean, so small scales of the velocity have higher variability than the SST small scales. The energy distribution for the currents correction is however similar to the distribution of the correction of the temperature, because the structure of the velocity covariance is directly related to the structure of the temperature covariance.

6. Spatio-temporal analysis using EOFs

Wavelet transforms are very useful to analyse a model at different resolution levels. However, to analyse the whole Mediterranean Sea model we prefer to use EOFs. The irregular land-sea boundaries of the Mediterranean Sea model will affect the wavelet coefficients, making very difficult the analysis. EOFs are not affected by irregular boundaries, because only the sea points are retained for their calculation. Recall that for the computation of EOFs no missing data can be present in the analysed matrix. In this section, the reconstructed Pathfinder SST data introduced in Section 3.2 are used in the comparison with model results.

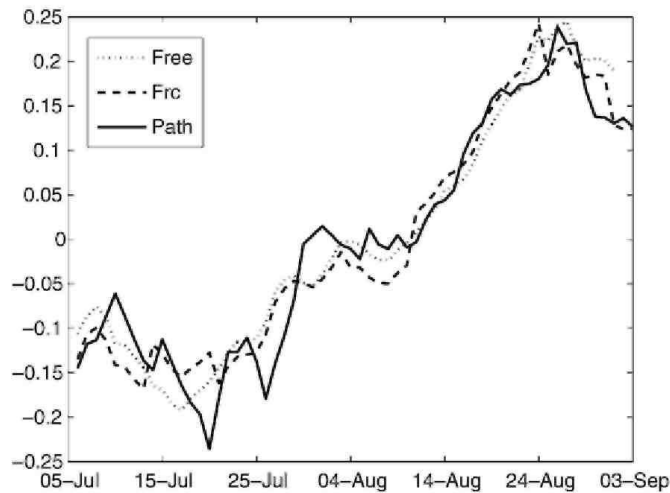**Fig. 22:** *First temporal mode for Free run, Forecast and observations.*

Table 2: *Correlation between the model and the observations for each EOF mode*

	Spatial mode		Temporal mode	
Mode	Forecast	Free run	Forecast	Free run
1st	0.89	0.78	0.96	0.96
2nd	0.48	0.56	0.69	0.77
3rd	0.22	0.2	0.16	0.37

Model and Pathfinder SST data have different resolutions, so the number of points representing the Mediterranean Sea is different for each of these matrices. Pathfinder SST has 139188 sea points and the Mediterranean Sea model grid has 4155 sea points. Both matrices have 60 time instants. For the direct comparison between the EOFs obtained with Pathfinder SST and the EOFs obtained with the model results, the ratio between the number of points must be taken into consideration. The relationship between the number of points and the singular values (S) of the EOFs based on the observations and model for this case is:

$$\frac{\sum\limits_{i} \mathbf{S}^2_{\text{model},i}}{\sum\limits_{i} \mathbf{S}^2_{\text{obs},i}} \sim \frac{N_{\text{model}}}{N_{\text{obs}}} \sim 0.03 \qquad\qquad (30)$$

For the direct comparison between model and observations singular values, the scaling factor $\sqrt{N}$ must be used. The right EOFs U are then multiplied by the singular values S when displayed to take into consideration the difference in the number of points. The spatial EOFs depicted in this section represent °C, and the spatial EOFs are adimensional. For the computation of the EOFs, the two-month time and space average from 6 July to 1 September is subtracted from the matrices.

The first six EOFs have been calculated for each of the studied matrices (*Free run, Forecast* and observations). The number of EOFs chosen corresponds to the optimal number of EOFs retained by DINEOF to reconstruct the Pathfinder SST data set (see Section 3.2). The variance explained by each mode is shown in Fig. 18, calculated as $V_i = S^2_{\text{data set},i}/P_{\text{data set}}$, with $P_{\text{data set}}$ the total number of points (in space and time) of the data set, *Forecast, Free run* or observations. The three most important modes account for a big part of the total variance explained by the six first EOFs. Thus, only these three modes will be kept for the study.

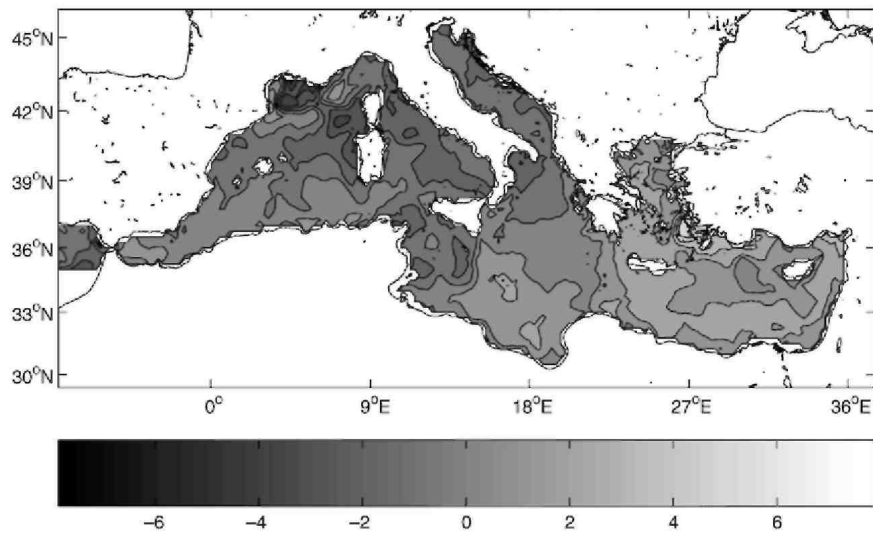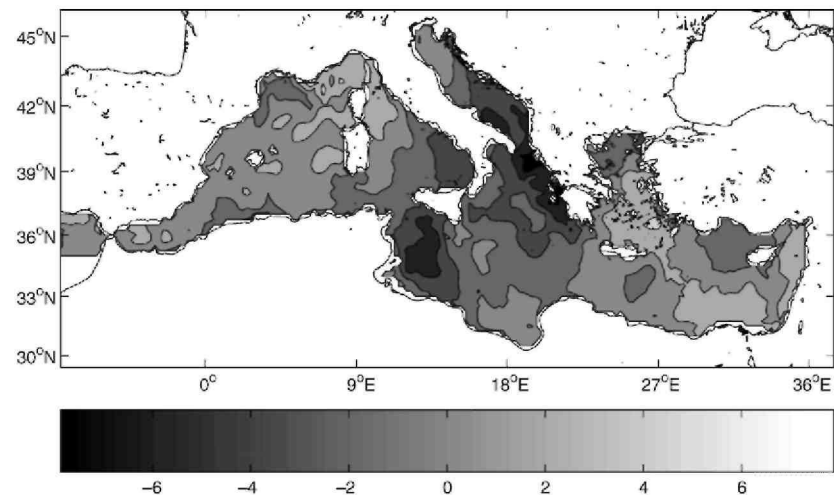**Fig. 23:** *Second spatial mode for the Free run.*
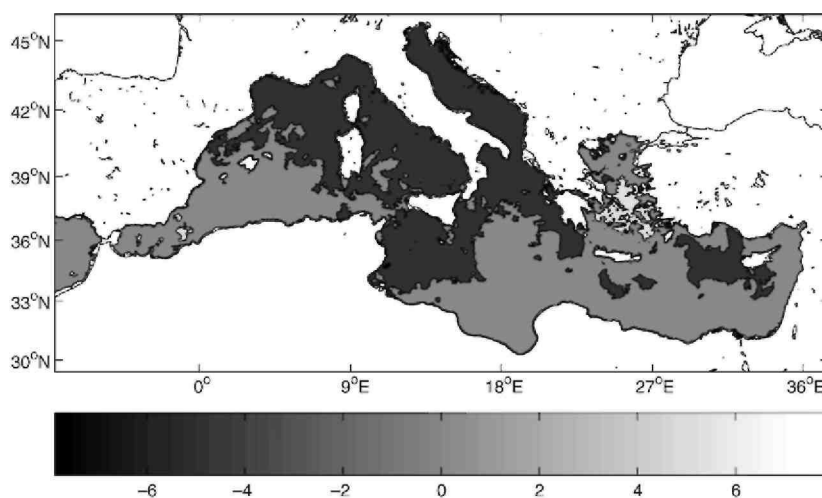

Fig. 24: *Second spatial mode for the Forecast.*

6.1. First EOF mode

The first spatial mode (Figs. 19-21 for *Free run, Forecast* and observations respectively) represents the summer SST distribution in the Mediterranean Sea. The spatial distribution of the SST is very similar for the three data sets, or in other words, the model (both the *Free run* and the *Forecast)* is able to accurately represent the SST distribution of the first EOF in the Mediterranean Sea. In the already warm waters of the summer Mediterranean Sea, two zones are contributing to the surface warming: the north-western Mediterranean (Gulf of Lions and Ligurian Sea) and the Northern Adriatic Sea. The Gulf of Lions and the Northern Adriatic are influenced by river runoff of the Rhone River and the Po River respectively. Po River water is warmer than the surrounding open seawater in summer *(e. g.* Gacić et al., 1997) so it influences the warming of the Northern Adriatic. The Rhone River runoff has been also found to be warmer than the surrounding waters of the Gulf of Lions, at least in year 2000 *(e. g.* Younes et al., 2003). On the other hand, the water entering the Mediterranean Sea by the Gibraltar Strait and the Bosphore Strait are the only two sources of cold water. It must be mentioned that the Black Sea is not explicitly included in the model. The signal observed at the Bosphore Strait is the result of the relaxation of the model to the climatology.

The first temporal mode (Fig. 22) represents the summer heating trend from July to August. The correlation between the observations and the *Forecast* is better than for the Free run (see Table 2). The amplitude minimum found in July is slightly shifted in the model respect to the observations. The RMS error between the model and the observations first temporal mode is 0.04 for both *Free run* and *Forecast.* If the first temporal mode of the model is forwarded 2 days in time, the RMS error is minimised to 0.035, indicating a slight shift between model and observations. However, the maximum at the end of the experiment is reached at the same time by the model and the observations. The temporal shift is only present the first half of the experiment, which explains that only a small improvement has been reached by shifting the model.

Fig. 25: *Second spatial mode for the observations.*


6.2. Second EOF mode

The second EOF spatial mode in all three fields (Figs. 23-25) presents a zonal crest-trough-crest structure. However, the crest of the *Forecast* in the Western Mediterranean is displaced to the North compared to the observations. The distribution of the *Free run* is closer to the observations than the *Forecast,* as indicated by the correlation in Table 2. On the other hand, the amplitude is better reproduced by the *Forecast* than by the *Free run.* The three data sets show a similar time evolution of the second EOF mode (Fig. 26).

While the first mode represents the seasonal heating forced by the atmosphere, the second EOF mode accounts for the intrinsic ocean variability, mostly at the meso-scale and sub-basin scale. The second EOF mode of the observations shows the variability of the Rhone River Plume, an upwelling west of Corsica and Sardinia, cold waters off the Sicilian south coast, the Ierapetra Gyre (southeast of Crete, Ayoub et al. (1998) and Rhodes Gyre

(west of Cyprus, Milliff and Robinson (1992). In the second EOF mode of the observations and *Free run,* these features are correlated in the same way, except the Ierapetra Gyre in the Free run, which presents the opposite correlation pattern than the observations. This means that the SST of the Ierapetra Gyre decreases with time in the observations, and increases in the *Free run.* In the *Forecast,* the Ierapetra Gyre is in phase with the observations, as well as the cold waters off the Sicilian south coast and the Rhodes Gyre. The Rhone River Plume is not present in the *Forecast.* The double anticyclonic gyre system in the Alboran Sea is clearly visible in the second EOF of the *Forecast,* but is not captured by the second EOF of the observations and *Free run.*

The assimilation has a positive effect on the first EOF mode. However, the second EOF mode, representing the dynamical features not directly forced by the heat fluxes, is deteriorated by the assimilation. This indicates that the assimilation adversely affects the equilibrium of these dynamical structures during the analysis.

Fig. 26: *Second temporal mode for Free run, Forecast and observations.*
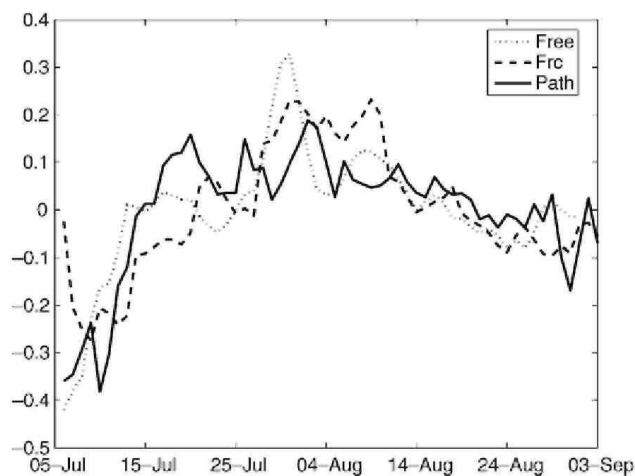
6.3. Third EOF mode

The third EOF of *Free run* and observations (Figs. 27 and 28 respectively) are in good agreement, especially in the Provençal Basin and the Levantine Basin. The Rhodes Gyre is again visible in the third EOF of both data sets. The Algerian current is also present in both, but it is stronger and better defined in the observations than in the *Free run.* However, the third EOF of these two data sets has the opposite sign on the Tunisian Shelf. The third EOF of the *Forecast* (Fig. 29) is quite different from the observations and *Free run.* In particular, the *Forecast* shows an East-West modulation of the Algerian Current, instead of the North-South modulation visible in the Free run and observations. The structures in the Levantine Basin and Ionian Sea are also different from the observations. Since the third and fourth EOFs of the Forecast have a similar variance (see Fig. 18), only a small change can alter the order of these EOFs. Indeed, we found that the fourth EOF of the Forecast (not shown) is very similar to the third EOF of the observations and *Free run.*

7. Conclusions

In this work we have analysed the results of a two-way nested model in the Mediterranean Sea. The overall behaviour of the model is good, and the assimilation of SST and SLA improves its quality in terms of RMS error and bias when compared to a *Free run.* However, some aspects of the model can be ameliorated. For example, when the atmospheric heating/cooling trend is in disagreement with the observed SST (as seen at the end of this experiment), an assimilation frequency shorter than one week is necessary to prevent the model to diverge from the observed SST. We have seen that the low-resolution model has a smaller RMS error than the high- resolution model, due to the double penalty effect. This effect is related to the increased small-scale variability in the high-resolution model, which is less predictable. This effect has to be considered when interpreting model results at different resolutions.
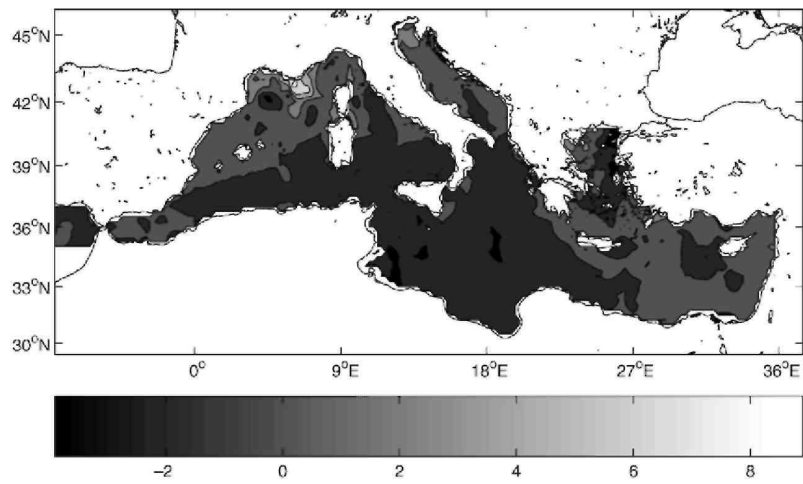
Fig. 27: *Third spatial mode for the Free run.*

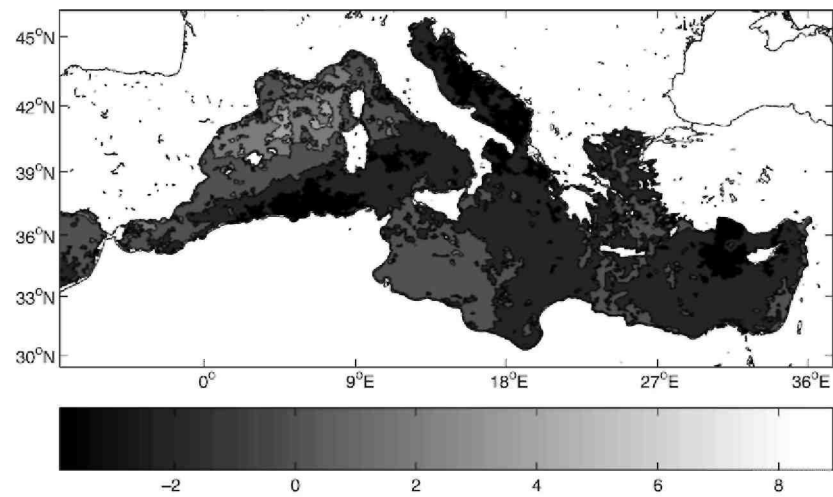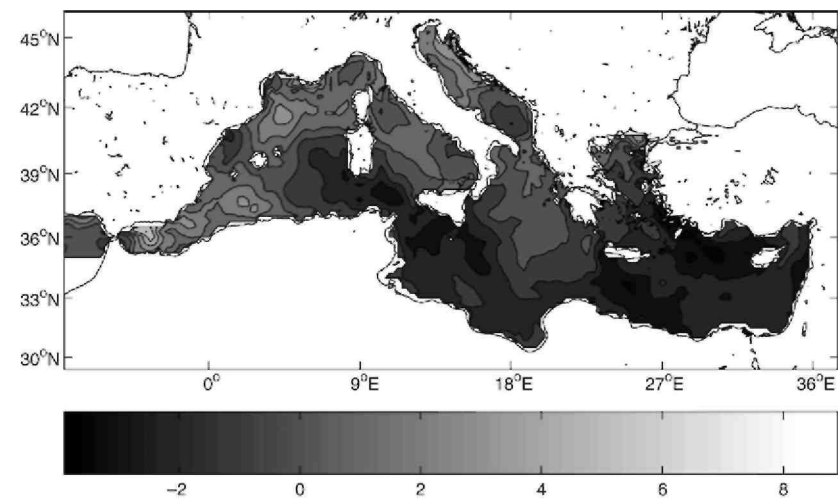**Fig. 28:** *Third spatial mode for the observations.*

Fig. 29: *Third spatial mode for the Forecast.*

A detailed study of the model spatial behaviour, using two-dimensional wavelets, has allowed to point out some weaknesses or aspects that could be improved in the model. First, the wavelet analysis has revealed different behaviours for the small and large-scales in the model.

Large scales are significantly improved by the assimilation, as noticed by the decrease in RMS error and correlation from the *Free run* to the *Forecast.* Also, the standard deviation of the model large scales is very close to the observations, even in the *Free run.* The RMS distribution among scales shows that most of the error in the model is due to the large scales, so the reduction of the total error by the assimilation is due to the improvement of the model accuracy at large scales.

We have detected a problem in the position of the Northern Current (NC) in the Ligurian Sea model. Although only the SST has been studied, we have been able to detect the characteristics of the NC, as it has a very clear signal in the SST. The NC is a feature at large-scale, and its mis-positioning contributes significantly to the total error. However, the wavelet analysis has revealed the link between the position of this current and an error in the small scales. Small-scale errors are high near the NC, especially where the path of the current is mis-represented. The error at the small scales is therefore partially induced by large scales error. Such a finding would not be possible with a classical Fourier analysis, where the spatial information is lost.

The contribution of the small scales to the total error is masked by the large-scale error. By isolating the small-scale with the wavelet analysis, we have revealed model weaknesses at these scales. The small-scale variability of the model is systematically lesser than the variability of the observed SST. The assimilation further reduces this variability, and in addition it does not improve significantly the correlation or the RMS error of the small scales.

Large-scale SST errors in the model can be attributed to poorly known atmospheric heat fluxes, and the position of the NC. However, the detection of the cause of the error in the small scales is more complex to determine. The initial condition does not contain already variability at the small scales. The inability of the model to later generate this variability at small scales can be due to various reasons. For example, a too high horizontal diffusion (explicit and implicit due to the finite resolution) can be responsible for the reduction of the variability at small scales.

The scale distribution of variance depends on the nature of the variable. The SST variance distribution is very different from the velocity variance distribution. The SST variance is dominated by large-scale features, while the surface velocity variance is highest at small scales. We attribute this different behaviour to the fact that the SST distribution in summer is largely influenced by atmospheric forcing, and the typical atmospheric variability scales are much larger than the variability scales in the ocean. The velocity scale distribution reflects the typical size of ocean flow features in the Ligurian Sea.

The multiresolution analysis has also shown that the choice of the maximum correlation length during the assimilation influences directly the error distribution of the model. The maximum correlation length of the assimilation scheme is 47 km, and this value is not in disagreement with the finding that the model has the highest error at scales 32 km and larger. However, a fine-tuning of this parameter could improve the model results, since scale 16 km in the model presents only a slight decrease of error with the assimilation. This scale could benefit from a smaller value of the maximum correlation length. This new value should not be smaller than 16 km, as scales smaller than that contribute little to the total error.

The EOF analysis of the Mediterranean Sea has revealed that the assimilation mainly improves the seasonal heating represented by the first EOF mode. The second EOF mode is degraded by the assimilation. This mode includes the dynamical features not directly forced by the heat fluxes, such as gyres, the Rhone River plume, and coastal upwellings. This indicates that the assimilation does not account for the dynamical balance of these features, and therefore the signal of these structures is weakened by the assimilation.

The EOF modes of the model runs and the observations present overall the same structures for the first and second modes. However, when comparing EOFs, we must take into account that the ranking of the EOFs mode is not necessarily the same for different data sets. If eigenvalues are close (which often happens at higher order modes), a small perturbation can change the order of these EOFs. This happened in the *Forecast* fourth EOF, which is very similar to the third EOF of *Free run* and observations.

It is worthwhile to notice that the wavelet analysis and the EOF decomposition lead to similar conclusions, viewed from a different perspective. The fact that the first EOF is improved by the assimilation is consistent with

the error decrease of large scales in the Ligurian Sea model after the assimilation noticed with the wavelet analysis. However, the assimilation degrades the second EOF mode, which is directly linked with the lack of improvement of the small-scale features in the Ligurian Sea model.

Acknowledgements

References

Albérola, C, Millot, C, 2003. Circulation in the French mediterranean coastal zone near Marseilles: the influence of wind and the Northern Current. Continental Shelf Research 23, 587-610.

Albérola, C, Millot, C, Font, J., 1995. On the seasonal and mesoscale variabilities of the Northern Current during the 'PRIMO-0 experiment in the western Mediterranean Sea. Oceanologica Acta 18, 163-192.

Alvera-Azcárate, A., Barth, A., Rixen, M., Beckers, J.-M., 2005. Reconstruction of incomplete oceanographic data sets using Empirical Orthogonal Functions. Application to the Adriatic Sea. Ocean Modelling 9, 325-346.

Astraldi, M., Gasparini, G.R, 1992. The seasonal characteristics of the circulation in the north Mediterranean basin and their relationship with the atmospheric-climatic conditions. Journal of Geophysical Research 97, 9531-9540.

Astraldi, M., Balopoulos, S., Candela, J., Font, J., Gacic, M., Gasparini, G.R, Manca, B., Theocharis, A., Tintoré, J., 1999. The role of straits and channels in understanding the characteristics of Mediterranean circulation. Progress in Oceanography 44, 65-108.

Ayoub, N., Le Traon, P.Y., De Mey, P., 1998. A description of the Mediterranean surface variable circulation from combined ERS-1 and TOPEX/POSEIDON altimetric data. Journal of Marine Systems 18, 3-40.

Barth, A., Alvera-Azcárate, A, Rixen, M., Beckers, J.-M., 2005. Two-way nested model of mesoscale circulation features in the Ligurian Sea. Progress in Oceanography 66, 171-189.

Barth, A., Alvera-Azcárate, A., Beckers, J.-M., Rixen, M., Vanden-bulcke, L., 2007. Multigrid state vector for data assimilation in a two-way nested model of the Ligurian Sea. Journal of Marine Systems 65, 41-59 (this issue). doi:10.1016/j.jmarsys.2005.07.006.

Beckers, J.-M., 1991. Application of a 3D model to the western Mediterranean. Journal of Marine Science 1, 315-332.

Beckers, J.-M., Rixen, M., 2003. EOF calculations and data filling from incomplete oceanographic data sets. Journal of Atmospheric and Oceanic Technology 20 (12), 1839-1856.

Beckers, J.-M., Rixen, M., Brasseur, P., Brankart, J.-M., Elmoussaoui, A., Crépon, M., Herbaut, C, Martel, F., Van den Berghe, F, Mortier, L., Lascaratos, A., Drakopoulos, P., Korres, P., Pinardi, N., Masetti, E., Castellari, S., Carini, P., Tintore, J., Alvarez, A, Monserrat, S., Parrilla, D., Vautard, R., Speich, S., 2002. Model intercomparison in the Mediterranean. The MedMEx simulations of the seasonal cycle. Journal of Marine Systems 33-34, 215-251.

Boer, G.J., Lambert, S.J., 2001. Second-order space-time climate difference statistics. Climate Dynamics 17, 213-218.

Brankart, J.M., Brasseur, P., 1998. The general circulation in the Mediterranean Sea: a climatological approach. Journal of Marine Systems 18,41-70.

Brankart, J.M., Testut, C.E., Brasseur, P., Verron, J., 2003. Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean modes: application to a 1993-96 hindcast of the North Atlantic Ocean circulation. Journal of Geophysical Research 3074. doi:10.1029/2001JC001198.

Brasseur, P., Beckers, J.-M., Brankart, J.M., Schoenauen, R., 1996. Seasonal temperature and salinity fields in the Mediterranean Sea: climatological analyses of a historical data set. Deep-Sea Research. Part 1. Oceanographic Research Papers 43, 159-192.

Brasseur, P., Ballabera, J., Verron, J., 1999. Assimilation of altimetric observations in a primitive equationmodel of the Gulf stream using a Singular Evolutive Extended Kalman filter. Journal of Marine Systems 22 (4), 269-294.

Briggs, W.M., Levine, R.A., 1997. Wavelets and field forecast verification. Monthly Weather Review 125, 1329-1341.

Casati, B., Ross, G, Stephenson, D.B., 2004. A new intensity-scale approach for the verification of spatial precipitation forecasts. Meteorological Applications (RMS) 11, 141-154.

Chapa, S.R., Rao, B.V., Durga-Prasad, G.S.S., 1998. Application of wavelet transform to Meteosat-derived cold cloud index data over South America. Monthly Weather Review 126, 2466-2481.

Daubechies, I., 1992. Ten Lectures on Wavelets. SIAM. 357 pp.

Daubechies, I., Guskov, 1., Schröder, P., Swelden, W., 1999. Wavelets on irregular point sets. Philosophical Transactions of the Royal Society of London. A 357 (1760), 2397-2413.

Denis, B., Laprise, R., Caya, D., 2003. Sensitivity of a regional climate model to the resolution of the lateral boundary conditions. Climate Dynamics 20, 107-126.

Ebert, E.E., McBride, J.L., 2000. Verification of precipitation in weather systems: determination of systematic errors. Journal of Hydrology 239, 179-202.

Font, J., Garcia-Ladona, E., Gorriz, E., 1995. The seasonality of mesoscale motion in the Northern Current of the western Mediterranean: several years of evidence. Oceanologica Acta 18, 207-219.

Fournier, A., 2002. Atmospheric energetics in the wavelet domain. Part I: governing equations and interpretation for idealized flows. Journal of the Atmospheric Sciences 59, 1182-1197.

Fournier, A., 2003. Atmospheric energetics in the wavelet domain. Part II: time-averaged observed atmospheric blocking. Journal of the Atmospheric Sciences 60, 319-338.

Gacic, M., Marullo, S., Santoleri, R., Bergamasco, A., 1997. Analysis of the seasonal and inetrannual variability of the sea surface temperature field in the Adriatic Sea from AVHRR data (1984-1992). Journal of Geophysical Research 102, 22937-22946.

Gates, W.L., Boyle, J.S., Covey, C, Dease, C, Doutriaux, CM., Drach, R.S., Fiorino, M., Glecker, P.J., Hnilo, J.J., Mariais, S.M., Phillips, T.J., Potter, G.L., Santer, B.D., Sperber, K.R., Taylor, K.E., Williams, D.N, 1999. An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). Bulletin of the American Meteorological Society 80 (1), 29-55.

Hoffman, R.N., Liu, Z., Louis, J.-F, Grasotti, C, 1995. Distortion representation of forecast errors. Monthly Weather Review 123, 2758-2770.

Jolliffe, I.T., Stephenson, D.B., 2003. Forecast verification. A Practitioner's Guide in Atmospheric Science. Wiley.

Kilpatrick, K.A., Podestá, G.P, Evans, R., 2001. Overview of the NOAANASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated match up database. Journal of Geophysical Research 106 (C5), 9179-9197.

Kumar, P., Foufoula-Georgiou, E., 1993. A multicomponent decomposition of spatial rainfall fields. 1. Segregation of large-and small-scale features using wavelet transforms. Water Resources Research 29(8), 2515-2532.

Kumar, P., Foufoula-Georgiou, E., 1997. Wavelet analysis for geophysical applications. Reviews of Geophysics 35 (4), 385-412.

Larnicol, G, Le Traon, P.Y., Ayoub, N, De Mey, P., 1995. Mean sea level and surface circulation variability of the Mediterranean Sea from 2 years of topex/poseidon altimetry. Journal of Geophysical Research 100 (C12), 25163-25177.

Lermusiaux, P.F.J., 2001. Evolving the subspace of the three-dimensional multiscale ocean variability: Massachusetts bay. Journal of Marine Systems 29, 385-422.

Lermusiaux, P.F.J., 2002. On the mapping of multivariate geophysical fields: sensitivities to size, scales, and dynamics. Journal of Atmospheric and Oceanic Technology 19, 1602-1637.

Liang, X.S., Robinson, A.R., 2004. A study of the Iceland-Faroe frontal variability using the multiscale energy and vorticity analysis. Journal of Physical Oceanography 34, 2571-2591.

Liang, X.S., Robinson, A.R., 2005. Localized multiscale energy and vorticity analysis. I. Fundamentals. Dynamics of Atmospheres and Oceans 38 (3-4), 195-230.

Lindsay, R.W., Percival, D.B., Rothrock, D.A., 1996. The discrete wavelet transform and the scale analysis of the surface properties of sea ice. IEEE Transactions on Geoscience and Remote Sensing 34 (3), 771-787.

Livezey, R.E., Chen, W.Y., 1983. Statistical field significance and its determination by Monte Carlo techniques. Monthly Weather Review 111,46-59.

Malanote-Rizzoli, P., Manca, B.B., Ribera d'Alcalà, M., Theocharis, A., Brenner, S., Budillon, G, Ozsoy, E., 1999. The Eastern Mediterranean in the 80s and in the 90s: the big transition in the intermediate and deep circulations. Dynamics of Atmospheres and Oceans 29, 365-395.

Mallat, S., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (n7), 674-693.

Mallat, S., 1998. A Wavelet Tour of Signal Processing. Academic Press. 637 pp.

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.

Milliff, R.F., Robinson, A.R., 1992. Structure and dynamics of the Rhodes Gyre system and dynamical interpolation for estimates of the mesoscale variability. Journal of Physical Oceanography 22,317-337.

Millot, C, 1999. Circulation in the Western Mediterranean Sea. Journal of Marine Systems 20, 423-442.

Molcard, A, Pinardi, N, Iskandarani, M., Haidvogel, D.B., 2002. Wind driven general circulation of the Mediterranean Sea simulated with a Spectral Element Ocean Model. Dynamics of Atmospheres and Oceans 35, 97-130.

Murphy, AH., Epstein, E.S., 1989. Skill Scores and Correlation Coefficients in Model Verification. Monthly Weather Review 117,572-581.

Naveau, P., Oh, H.-S., 2004. Polynomial wavelet regression for images with irregular boundaries. IEEE Transactions on Image Processing 13 (6), 773-781.

Oh, H.-S., Naveau, P., Lee, G, 2001. Polynomial boundary treatment for wavelet regression. Biometrika 88 (1), 291-298.

Pham, D.T., Verron, J.C.R.M., 1998. A singular evolutive extended kalman filter for data assimilation in oceanography. Journal of Marine Systems 16, 323-340.

Pinardi, N, Masetti, E., 2000. Variability of the large scale general circulation of the Mediterranean Sea from observing and modelling: a review. Palaeogeography, Palaeoclimatology, Palae-oecology 158, 153-173.

Raick, C, Alvera-Azcárate, A., Barth, A., Brankart, J.-M., Soetaert, K, Grégoire, M., 2007. Application of a SEEK filter to a 1d biogeochemical model of the Ligurian sea: twin experiments and real in situ data assimilation. Journal of Marine Science 65, 561-583 (this issue). doi:10.1016/j.jmarsys.2005.06.006.

Rao, Y.R., Murthy, C.R., 2001. Comparative Analysis of Spectral, Empirical Orthogonal Function, and Wavelet Decompositions for an Upwelling Event. Journal of Atmospheric and Oceanic Technology 18, 704-710.

Reynolds, R.W., Smith, T, 1994. Improved global sea surface temperature analyses using optimum interpolation. Journal of Climate 7, 929-948.

Sammari, C, Millot, C, Prieur, L., 1995. Aspects of the seasonal and mesoscale variabilities of the Northern Current in the Western Mediterranean Sea inferred from the prolig-2 and pros-6 experiments. Deep-Sea Research. Part 1. Oceanographic Research Papers 42(6), 893-917.

Smith, W.H.F., Sandwell, D.T., 1997. Global sea floor topography from satellite altimetry and ship depth soundings. Science 277, 1956-1962.

Spedding, G.R., Browand, F.K, Huang,N.E., Long, S.R., 1993. A 2-D complex wavelet analysis of an unsteady wind-generated surface wave field. Dynamics of Atmospheres and Oceans 20, 55-77.

Sweldens, W., 1997. The lifting scheme: a construction of second generation wavelets. SIAM Journal on Mathematical Analysis 29 (2), 511-546.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research 106 (D7), 7183-7192.

Torrence, C, Compo, G.P, 1998. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society 79 (1), 61-78.

Tusseau, M.H., Mouchel, J.M., 1994. Nitrogen inputs to the Gulf of Lions via the Rhône river. In: Martin, J., Barth, H. (Eds.), Water Pollution Research Reports. Proceedings of the EROS 2000 Workshop, Hamburg. Commission of the European Communities. Mar.

Viudez, A., Tintoré, J., Haney, R., 1996. Circulation in the Alboran Sea as determined by quasi-synoptic hydrographic observations. Part I: three dimensional structure of the two Anticyclonic Gyres. Journal of Physical Oceanography 26 (5), 684-705.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences. Academic Press. 467 pp.

Yano, J.I., Moncrieff, M.W., Wu, X., Yamada, M., 2001. Wavelet analysis of simulated tropical convective cloud systes; Part I: basic analysis. Journal of the Atmospheric Sciences 58, 850-867.

Younes, W.A.N., Bensoussan, N, Romano, J.-C, Arlac, D., Lafont, M.-G, 2003. Seasonal and interannual variations (1996-2000) of the coastal waters east of the Rhone River mouth as indicated by the SORCOM series. Oceanologica Acta 26 (4), 311-321.