POLITIQUE SCIENTIFIQUE FEDERALE

FEDERAAL WETENSCHAPSBELEID

Research programme for Earth Observation Stereo II

Contract n° SR/12/140

HiSea
High resolution merged satellite Sea surface temperature fields

GeoHydrodynamics and Environmental Research (GHER)
University of Liège

Final report

21 Feb 2013

For the PARTNERSHIP: the coordinator

Table of Contents

1.Executive summary....................2
2.Report....................2
2.1.Comparison of in situ and satellite SST data in the western Mediterranean Sea....................3
2.2.Detection of outliers in satellite data using spatial coherence....................3
2.3.Development of a technology for merging several datasets using DINEOF....................3
2.4.Application of the DINEOF-merging technology to various SST datasets....................4
2.5.Application of DINEOF-OI to other datasets....................5
2.6.DINEOF forecasts....................5
2.7.Dissemination....................5
a)Papers published in peer-reviewed journals....................5
b)Oral and poster presentations at international conferences and meetings....................6
c)Organisation of the 44th International Liège Colloquium....................6
3.Main conclusions of HiSea....................7
List of Annexes....................8

1. Executive summary

Several satellites measure Sea Surface Temperature (SST), each with different technical specificities and error sources. Together with in situ data, they form a highly complementary data set. The creation of merged SST products, integrating the strengths of each of its components and minimising their weaknesses, is however not an easy task, but it is certainly a desirable goal that has generated a large amount of research over the last years. The work developed in the frame of HiSea (High resolution merged satellite Sea surface temperature fields) aimed to provide a technique to merge several data sources using DINEOF (Data Interpolating Empirical Orthogonal Functions), by addressing the following tasks:

- Development of a technology that allows to merge different data sets at very different sampling intervals (in space and time) and create an integrated product at the highest sampling frequency and with the highest quality possible.

- Provide improved, merged analyses of variables such as SST and Total Suspended Matter.

- Improve our understanding of the relation between variables (and take advantage of this improved knowledge to ameliorate the analyses).

- Using the above-mentioned developments, explore the capability of DINEOF to produce forecasts based on multi-variate EOFs and model forecasts.

DINEOF is a technique to infer missing data in satellite data sets. In this project DINEOF was further developed so that it can merge different data sets. First, an initial DINEOF reconstruction of a data set with a high spatial resolution is made, and the EOF basis obtained is used as the covariance matrix needed to subsequently include into the analysis other data sources (satellite and in situ). Error estimations for each data set are used to weight their influence in the final product.

Merged high-resolution (in space and time) SST data sets and error statistics were obtained for the Mediterranean Sea for 2009. Analyses with total suspended matter were also performed in the North Sea for the year 2008. The improvements made to the base technique used throughout the project (DINEOF) will be made available freely and openly to the scientific community (source code and documentation). Results have been published in international peer-reviewed journals (an additional manuscript is under preparation).

2. Report

The main results and conclusions obtained during HiSea are summarised here. The full reports and publications can be found in the Annexes.

2.1. Comparison of in situ and satellite SST data in the western Mediterranean Sea

A comparison between in situ and satellite sea surface temperature (SST) was realised in the western Mediterranean Sea for the year 1999. Five international databases were used to extract in situ data for the desired period and zone: World Ocean Database (WOD), MEDAR/Medatlas, Coriolis, International Council for the Exploration of the Sea (ICES) and International Comprehensive Ocean-Atmosphere Data Set (ICOADS). A careful error estimation was performed, classified by type of platform and database. Given to the heterogeneity of the data used, very different conclusions were reached when comparing them individually to satellite data. Among the major conclusions it was seen that ship data, from the ICOADS database, were the most numerous, and as such they are a valuable source of data. However, the error assessment between these data and satellite data showed a large bias and RMS error, due to the heterogeneity of the measuring approaches and sensors used from ships. Data from other platforms compared generally well with satellite data, with RMS errors from 0.6ºC to 0.9ºC and small biases.

The study was published in Ocean Dynamics, and can be found in Annex I

A. Alvera-Azcárate, C. Troupin, A. Barth, and J.-M. Beckers. Comparison between satellite and in situ sea surface temperature data in the Western Mediterranean Sea. Ocean Dynamics, 61(6):767-778, 2011.

2.2. Detection of outliers in satellite data using spatial coherence

Satellite data sets often contain outliers (i.e., anomalous values with respect to the surrounding pixels), mostly due to undetected clouds and rain or to atmospheric and land contamination. A methodology to detect outliers in satellite data sets was developed. The approach was based on a truncated Empirical Orthogonal Function (EOF) basis obtained by analysing the data with DINEOF. The information rejected by this EOF basis was used to identify suspect data. A proximity test and a local median test were also performed, and a weighted sum of these three helped to accurately detect outliers in a data set. Most satellite data undergo automated quality-check analyses. The approach developed here exploited the spatial coherence of the geophysical fields, therefore detecting outliers that would otherwise pass such checks. The methodology was applied to infrared sea surface temperature (SST), microwave SST and chlorophyll-a concentration data over different domains, showing the applicability of the technique to a range of variables and temporal and spatial scales. A series of sensitivity tests and validation with independent data were also conducted.

This study was published in Remote Sensing of Environment, and can be found in Annex II.

A. Alvera-Azcárate, D. Sirjacobs, A. Barth, and J.-M. Beckers. Outlier detection in satellite data using spatial coherence. Remote Sensing of Environment, 119:84-91, 2012.

2.3. Development of a technology for merging several datasets using DINEOF

This was the core activity of the project, and consisted on developing an extension to

DINEOF to merge several datasets. The basic DINEOF technique is described at depth at the following papers: Beckers and Rixen(2003), Alvera-Azcárate et al (2005), Beckers el at (2006), Alvera-Azcárate et al (2007).The methodological approach used in HiSea is described in Annex III in detail. First, a two-step approach was developed, by performing first a DINEOF analysis of one dataset (from a polar-orbiting satellite) and using the obtained EOF basis in a second step to merge several datasets. An embedded approach was then developed, by basically performing step 2 of the previous approach within DINEOF, each time that the EOF basis is updated in the iterative process. The aim of testing these two approaches was to see the impact of the iterative process in the quality of the final fields of SST. Finally, an methodology to use a correlated error covariance matrix in DINEOF-OI was developed.

2.4. *Application of the DINEOF-merging technology to various SST datasets*

Several applications of the DINEOF-OI technique have been performed in order to analyse the obtained results. A first test was realised with AVHRR SST data and the in situ data described in section 2.1. The aim was to develop the technique with a small dataset. It was seen that using the DINEOF-OI approach improved the results over a DINEOF step alone (using only satellite data). We therefore moved one step further and apply DINEOF-OI to two satellite SST datasets.

A polar-orbiting dataset (AVHRR NAR-18) and a geostationary dataset (SEVIRI) were used for the year 2009. The merging results retain the high spatial correlation of the AVHRR NAR-18 dataset (~2 km) and the high temporal resolution of the SEVIRI dataset (3 hours). The results were compared with OSTIA (Operational Sea Surface Temperature and Sea Ice Analysis), a daily global analysis of all available satellite and in situ data at a resolution of ~6 km. OSTIA data are produced in the frame of the MyOcean project. The root mean square error between the DINEOF-OI analysis and the OSTIA analysis is 0.39ºC, with a bias of -0.09ºC (DINEOF-OI colder than OSTIA). This results demonstrate that the technique developed in the project gives accurate results. It must be noted that the DINEOF-OI uses only 2 satellite datasets and only a rough correction for the existing bias between these datasets is performed. It is therefore expected that the quality of the DINEOF-OI data increases with the use of more data and a better bias removal technique.

Several tests were also realised to establish the impact of using a correlated covariance matrix in the quality of the final products. The results show that the error can be indeed reduced when using a correlated covariance matrix. This study suggested that errors are correlated at different scales (we found local minima at 20 km, 300 km and 700 km), which would depend on the source of the error (e.g. the longest lengths scales can be due to atmospheric processes).

The detailed results are presented in Annex IV.

2.5. *Application of DINEOF-OI to other datasets*

In order to demonstrate the usefulness of the technique outside the main variable analysed in this project, SST, and in other domain than the Mediterranean Sea, a merging analysis has been performed using Total Suspended Matter (TSM) data over the North Sea, using again a polar-orbiting dataset (MODIS) and a geostationary dataset (SEVIRI). These data were provided by K. Ruddick (MUMM, Brussels). Annex V presents the results, that, although preliminary (an improved error variance estimation should be used) show the potential of DINEOF-OI to be also used with variables like ocean colour. The results are compared to in situ data to assess their quality, showing that the merged data are able to reproduce them accurately.

2.6. *DINEOF forecasts*

We have applied DINEOF to address the problem of forecasting a variable measured by satellite (TSM) using (i) previous satellite measurements of this variable, and (ii) a numerical model run in the same region. The aim was to provide forecasts of physical variables (such as currents, wind or bottom stress) at several days lead. Several combinations of variables were tried, but the correlation between the model variables and the satellite TSM were not high enough to provide meaningful results. Recommendations for future developments are to use better-correlated variables (maybe model SST), or using as a parameter the number of EOFs retained by DINEOF (choosing the one that provides better forecasts). A detailed report of the forecasting activities can be found in Annex VI.

2.7. *Dissemination*

The activities performed during the duration of HiSea have been presented at several international conferences and published in peer-reviewed papers. A dedicated webpage (http://www.gher.ulg.ac.be/HiSea/) contains the main results and conclusions obtained through the project.

a) Papers published in peer-reviewed journals

[1] A. Alvera-Azcárate, D. Sirjacobs, A. Barth, and J.-M. Beckers. Outlier detection in satellite data using spatial coherence. Remote Sensing of Environment, 119:84-91, 2012.

[2] A. Alvera-Azcárate, C. Troupin, A. Barth, and J.-M. Beckers. Comparison between satellite and in situ sea surface temperature data in the Western Mediterranean Sea. Ocean Dynamics, 61(6):767-778, 2011.

In addition, a manuscript is in preparation describing the DINEOF-OI technique.

b) Oral and poster presentations at international conferences and meetings

[1] A. Alvera-Azcárate, A. Barth, M. E. Toussaint and J.-M. Beckers. HiSea: High resolution merged satellite sea surface temperature fields. Belgian Earth Observation Day (BELSPO). Oudenburg, Belgium, 25 May 2011. Oral presentation + poster.

[2] Aida Alvera-Azcárate, Alexander Barth, and Jean-Marie Beckers. Merging satellite and in situ sea surface temperature data using DINEOF. Geophysical Research Abstracts, Vol. 13, EGU2011-5939. EGU General Assembly 2011, Oral presentation.

[3] Aida Alvera-Azcárate, Alexander Barth, and Jean-Marie Beckers. Satellite and in situ sea surface temperature comparison and merging in the Mediterranean Sea. Third International Workshop on Advances in the Use of Historical Marine Climate Data (MARCDAT-III). 2-6 May 2011, Frascati (Italy). Oral presentation.

[4] A. Alvera-Azcárate, C. Troupin, A. Barth, and J.-M. Beckers. An EOF-based technique to compute merged high resolution sea surface temperature fields. 44th International Liège Colloquium on Ocean Dynamics, 7-11 May 2012. Liège (Belgium).

c) Organisation of the 44th International Liège Colloquium

The 44th edition of the International Liège Colloquium was devoted to the topic “Remote sensing of colour, temperature and salinity – new challenges and opportunities”. The colloquium was co-organized by K. Ruddick (MUMM) and A. Alvera-Azcárate (Ulg). The official webpage is http://modb.oce.ulg.ac.be/?page=colloquium&year=2012

A total of 120 abstracts were submitted and keynote speakers presented the latest developments on the fields of remote sensing of colour, temperature and salinity.

A special issue is under preparation for publication at Remote Sensing of Environment. Twenty six manuscripts were submitted and are now in various stages of the revision process.

3. Main conclusions of HiSea

Given the results presented in this report and its annexes, the following conclusions can be drawn:

– A technique for merging several data sources (satellite and in situ) based on the methodology DINEOF has been developed (called DINEOF-OI). The correlation of the error covariance matrix can be taken into account, and tests showed the influence of atmospheric process on the error variance. The results obtained with DINEOF-OI are robust and can have a high temporal and spatial resolution.

– Two approaches were used in this project (two-step approach and embedded approach). The error assessment gives similar results for both approaches, although the embedded one seems to have more small scale variability. The embedded approach is, as of now, slower than the two step approach.

– DINEOF-OI can be easily applied to different variables. During HiSea, sea surface temperature and total suspended matter were used and both variables were successfully analysed.

– We developed a technique to remove outliers from the satellite data before the analyses. This outlier detection exploits spatial coherence of the data, and it helps to improve the quality of the final fields. Tests were made on various domains and variables.

– TSM forecasts using hydrodynamical model variables like currents and bottom stress do not provide meaningful results because of the lack of correlation between the model variables and TSM. Other variables should be tried (like model SST) to obtain improved forecasts of TSM.

List of Annexes

Annex I: Comparison of in situ and satellite SST data in the western Mediterranean Sea

Annex II: Detection of outliers in satellite data using spatial coherence

Annex III: Methodology

Annex IV: Application of DINEOF-OI to various SST datasets

Annex V: Application of DINEOF-OI to Total Suspended Matter data

Annex VI: Forecast of Total Suspended Matter in the southern North Sea using DINEOF

Annexx VII: Project file

Annex I

Comparison of in situ and satellite SST data in the western Mediterranean Sea

Ocean Dynamics (2011) 61:767–778
DOI 10.1007/s10236-011-0403-x

Comparison between satellite and in situ sea surface temperature data in the Western Mediterranean Sea

Aida Alvera-Azcárate · Charles Troupin · Alexander Barth · Jean-Marie Beckers

Received: 8 November 2010 / Accepted: 7 March 2011 / Published online: 13 April 2011

Abstract A comparison between in situ and satellite sea surface temperature (SST) is presented for the Western Mediterranean Sea during 1999. Several international databases are used to extract in situ data (World Ocean Database, MEDAR/Medatlas, Coriolis Data Center, International Council for the Exploration of the Sea and International Comprehensive Ocean-Atmosphere Data Set). The in situ data are classified into different platforms or sensors (conductivity–temperature–depth, expendable bathythermographs, drifters, bottles, and ships), in order to assess the relative accuracy of these type of data with respect to Advanced Very High Resolution Radiometer SST satellite data. It is shown that the results of the error assessment vary with the sensor type, the depth of the in situ measurements, and the database used. Ship data are the most heterogeneous data set, and therefore present the largest differences with respect to in situ data. A cold bias is detected in drifter data. The differences between satellite and in situ data are not normally distributed. However, several analysis techniques, as merging and data assimilation, usually require Gaussian-distributed errors. The statistics obtained during this study will be used in future work to merge the in situ and satellite data sets into one unique estimation of the SST.

Keywords Sea surface temperature · Data comparison · Satellite · In situ · Mediterranean Sea

Responsible Editor: Pierre-Marie Poulain

This article is part of the Topical Collection on *Multiparametric observation and analysis of the Sea*

A. Alvera-Azcárate (✉) · C. Troupin · A. Barth · J.-M. Beckers
AGO-GHER-MARE, University of Liège,
Allée du Six Aout, 17, Sart Tilman, Liège, 4000, Belgium
e-mail: a.alvera@ulg.ac.be

A. Alvera-Azcárate · A. Barth
F.R.S.-FNRS (National Fund for Scientific Research),
Brussels, Belgium

1 Introduction

Sea surface temperature (SST) is one of the key variables for the estimation of the state of the world climate (Donlon et al. 2009) and is considered one of the essential climate variables by the World Meteorological Organization. High-quality SST data sets are needed for various applications, including numerical weather prediction, ocean forecasting, and climate research. SST can be measured with different sensors and platforms, but these do not provide a homogeneous estimation of the SST because of the specificity of each sensor and platform type. Two major groups of platforms can be described, satellite and in situ SST. These methods of measurement are very different, hence the SST measured by each of them can differ in terms of spatial and temporal coverage, bias, effective depth of measurement, etc. In order to understand the differences and similarities between the SST measurements made from in situ platforms and satellite platforms, it is necessary to carefully assess the error and biases between them (Castro et al. 2008).

In this work, a comparison between in situ and satellite SST data is undertaken in the Western Mediterranean Sea. The ultimate objective in mind, which is not part of this work, is to use this error assessment for an optimal merging of these two data sources. Developments of new approaches for merging satellite and in situ SST are needed (Donlon et al. 2009), for which we need to take into account the specificities of each type of measure and the differences and biases between them. The results obtained through this exercise can be beneficial for other applications, as the improvement of platform and sensor design, the detection of drifts and biases in particular platforms, etc.

Several databases that compile large amounts of in situ data are available today. Each of these databases provides interesting data sets at the global and local scale. For some regional applications, the use of global databases (such as World Ocean Database, (WOD) or the International Comprehensive Ocean-Atmosphere Data Set, (ICOADS)) can turn out to be incomplete, so the use of more local or specialized databases becomes necessary. It is not yet well established how each of these databases are constructed, and quality control is certainly not homogenized among them. Therefore, in any work aiming to use data combined from different databases, a comparison between them becomes necessary. Moreover, each type of in situ sensor has specific errors (e.g., Emery et al. 2001a; Kent et al. 2010), therefore in the error assessment of this work, the different in situ sensors used will be compared with the satellite data separately.

There are several works which have undertaken the task of comparing SST data sources in order to obtain a better understanding of the differences and biases between them. For example, Kent and Challenor (2006) performed a global assessment of ship data errors from 1970 to 1997. Emery et al. (2001a) and Xu and Ignatov (2010) made a validation of global in situ buoys and ship data with the purpose of satellite calibration. A local error assessment between in situ and satellite data was realized by Barton (2007). Castro et al. (2008) show an error assessment of global infrared and microwave satellite data for merging purposes. Also, a large amount of research over the last years has been devoted to the merging of different SST sources (e.g., Guan and Kawamura 2004; Reynolds et al. 2007; Gentemann et al. 2009). All these works emphasize the difficulty in comparing SST from different platforms given their specific characteristics. These works also show, indirectly, that it is not easy to generalize the results of each of these works to a given zone in the world ocean, a given set of data, and a given application.

This work is organized as follows: section 2 describes the different satellite and in situ data sets used in this work. A description of the basic statistics of these data is included in section 3. Then an error assessment between the different data is performed, first by data type and then by database in section 4. Conclusions are presented in section 5.

2 Description of data

2.1 In situ data

The domain chosen for this study is the Western Mediterranean Sea (35.1°N-44.3°N; 6°W-15.6°W, see Fig. 1). Temperature data for this domain were downloaded from various databases:

- World Ocean Database 2005 (WOD05, Locarnini et al. (2006), http://www.nodc.noaa.gov/).
- MEDAR/MedAtlas (MEDAR-Group (2002), http://www.ifremer.fr/medar/).
- Coriolis Data Center (http://www.coriolis.eu.org/).
- International Council for the Exploration of the Sea (ICES, http://www.ices.dk/).
- International Comprehensive Ocean-Atmosphere Data Set (ICOADS, http://icoads.noaa.gov/, release 2.5).

The year 1999 was chosen for this study because of the large number of in situ data available. In more recent years less in situ data are available, as these still have to make their way to the international databases mentioned above.

Only the measurements taken at a maximum of 5 m depth are retained. For profile data, only the shallowest data point above 5 m depth is retained, and only those data for which the depth is known are kept (some of the data only indicate they are surface data, with no mention on the measurement depth). From these data, duplicates were automatically detected by identifying data located within a radius of 1/100° in both latitude

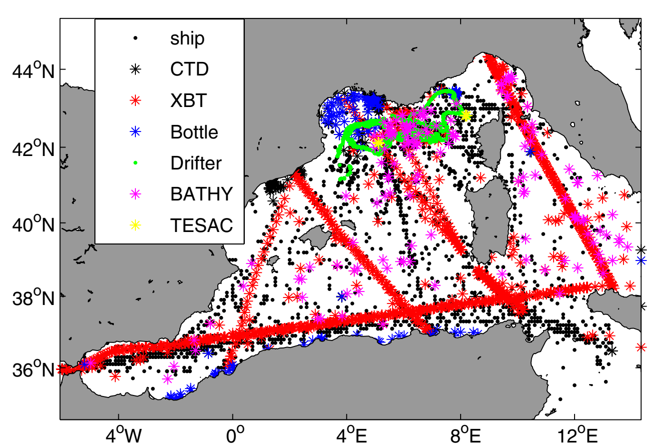


Fig. 1 Spatial distribution of the in situ data divided by sensor/platform type

and longitude, and that were taken within the same hour. If these temperature measurements differed by less than 1/100°C, they were considered as duplicates, and the one with the highest precision was retained. From the 7,731 initial data, a total of 6,636 data are kept after a basic quality check, that included the check for duplicates and the elimination of data that deviate more that ±3 standard deviations from the mean of the in situ data. Some of the data sets provide quality flags of their data, however to avoid discrepancies between quality flags used by the different data sets, none of these were used. Therefore, all data presented in this work have undergone the same quality checks, described above. The average depth of the retained data is 2 m, although a large part of the data (2,343 measurements, or 35% of the data) indicate the shallowest measurement as 0 m, with no indication of how close to the surface the measurement was taken in reality. From these, 1,951 are drifting buoys (i.e., 85% of the 2,343 data), and the rest are expendable bathythermographs (XBT) (6.6%), BATHY data (5.9%), and conductivity–temperature–depth (CTD) (3.4%). The surface drifter data used in this work are from the Surface Velocity Program, and have a sensor depth of about 0.2 m (P.-M. Poulain, personal communication), which can explain that their depth is referred as 0 m.

Information about the sensor/platform for each measurement is also recorded, with seven categories in total: (1) CTD data, (2) XBT, (3) floats/drifters, (4) low-resolution CTD or bottles, and (5) ship data. Also, (6) BATHY and (7) TESAC sensors are among the types of sensors found, but as no information about in which platform these are loaded, we kept them as separate categories. Finally, a number of them did not identify the sensor/platform. These unknown data were not included in this work. The resulting data distribution is shown in Fig. 1, in which the sensor/platform type is also specified. Ship data are the most numerous set, followed by drifters, XBTs and CTDs. The total number for each sensor is detailed in Table 1, along with the average depth for each of them. Although in general the study zone is well sampled when taking into account all types of measurements, individual sensor distribution is very heterogeneous, specially CTDs, bottles and drifters, which may introduce spatial biases.

Table 1 Number of observations available for this work, average depth for each of them, and coincident satellite observations

	In situ	Depth (m)	Satellite (day time)	Satellite (night time)
CTD	320	0.86	124	95
XBT	1,043	1.41	287	325
Bottle	260	1.96	73	64
Float/drifter	1,994	0.06	737	729
Bathy	141	0.02	46	60
Tesac	13	0	5	0
Ship	2,865	4	969	1,008
Total	6,636		2,241	2,281

All databases used in this work contain data from several sensors and platforms. ICOADS is the only one that provides ship data. Other type of data are also available at ICOADS, although we decided to keep only the ship data as it appeared that the sensor type of a large amount of these additional data did not have been correctly identified. For example, there were data classified as moored buoys that, given their spatial and temporal distribution, were more likely XBT data or ship data. From the ship data, 78% of the SST has been recorded at the engine room intake, 11% at the hull contact sensor, 10% of measurements were taking using a bucket, and there was also a small percentage (1%) with an unknown method of measurement.

2.2 Satellite data

Advanced Very High Resolution Radiometer (AVHRR) SST data on board the NOAA Polar Orbiting Environmental Satellite series for 1999 were downloaded from the NASA Jet Propulsion Laboratory Physical Oceanography Distributed Active Archive Center (PODAAC, http://podaac.jpl.nasa.gov). The horizontal resolution is about 5 km and both daytime and night-time passages were obtained. Infrared radiometers measure the top 10 μm of the sea surface (Robinson 2004), although they are calibrated using bulk temperature from buoys in oceans around the world (Emery et al. 2001b; Robinson 2004). However, it is unclear if the Mediterranean Sea is well covered by these calibration measures (Emery et al. 2001a) for a given year, so the actual depth represented by the satellite SST in the Mediterranean Sea is not well known.

A Data Interpolation Empirical Orthogonal Functions (DINEOF, Beckers and Rixen (2003); Alvera-Azcárate et al. (2005)) analysis of the data was realized to the AVHRR data to identify and remove outliers from the original data set, following Alvera-Azcárate et al. (2011), which proposes an improvement from the methodology used in Sirjacobs et al. (2011). Outliers are defined as data that present anomalous values with respect to the surrounding pixels. Examples of phenomena giving rise to the presence of outliers are cloud edges, haze areas, contrails, or cloud shadows. Pixels for which the analysis–observation difference (the residuals) are larger than the statistically expected misfit calculated during the analysis are identified as suspect. Two additional tests are realized, one checking the deviation of each pixel from a local median, and another verifying the proximity of each pixel to a cloud. The outliers are identified by combining the results from these three tests. A total of 0.12% and 0.15% of the data were removed from the day-time and night-time data sets, respectively. DINEOF is a technique to reconstruct missing data using an EOF basis, although at this stage only the outlier removal was performed. In the rest of the paper, the original (cloudy but outlier-free) AVHRR data are used.

To extract the satellite data at the in situ positions, a linear interpolation has been realized. If the satellite data were missing at the in situ location, then the nearest satellite pixel was chosen. In the presence of clouds it is common that the nearest pixel is also cloudy, so at the end, from the 6,636 in situ data available, a total of 2,241 day-time satellite data and 2,281 night-time data were present. The hour of the day at which the in situ data were taken was not considered for the interpolation, so both day-time and night-time satellite data are compared with the same set of in situ data, with no consideration of the hour on which the data were taken. The rationale behind this is that normally all available in situ data are used to create daily merged satellite/in situ maps, otherwise the number of in situ data would be too small. Also, these merged fields aim to represent the daily average SST.

3 Data statistics

The temporal distribution of the in situ and satellite data was examined. Figure 2 shows the number of data taken in each month of 1999. There are about 300–400 in situ observations per month, except for September, October and November, during which more than 800 measurements were made per month. The number of satellite data interpolated at the in situ locations is much smaller, due principally to the presence of clouds in this data set. As a result, there are about 34% of in situ data that can be compared with day-time or night-time satellite data. The distribution through the hours of the day (Fig. 3) of the in situ data is homogeneous, with only a slightly higher percentage of data taken during day-time hours (54.56%) compared with night-time hours (45.44%). The comparison with satellite data will therefore not be biased by inhomogeneous data distribution. There are specific times of the day with a very high number of data. These have been identified as ship data. Presumably, a high number of ships have an automated procedure for the recording of surface temperature data, and it is possible that this procedure is more often established at precise hours four times a day.

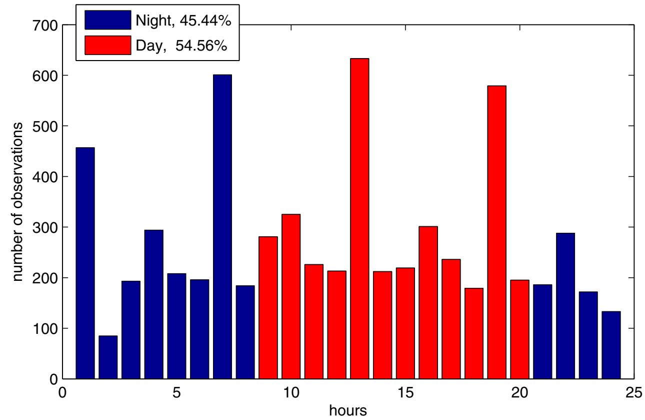


Fig. 3 Number of in situ observations for each hour of the day

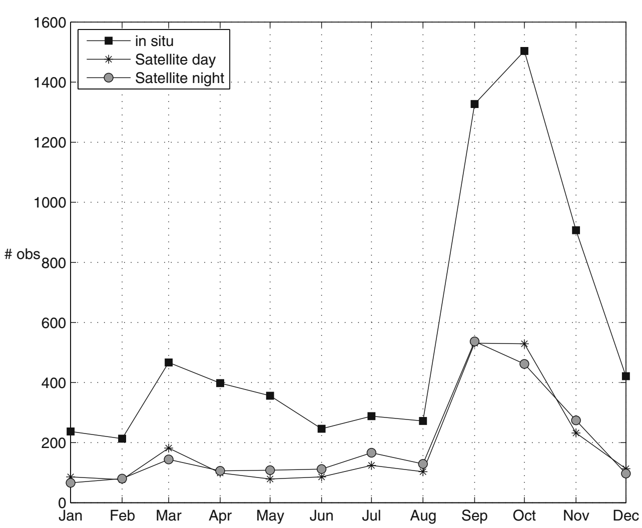


Fig. 2 Number of in situ observations, satellite day-time data and satellite night-time data for each month of the year 1999

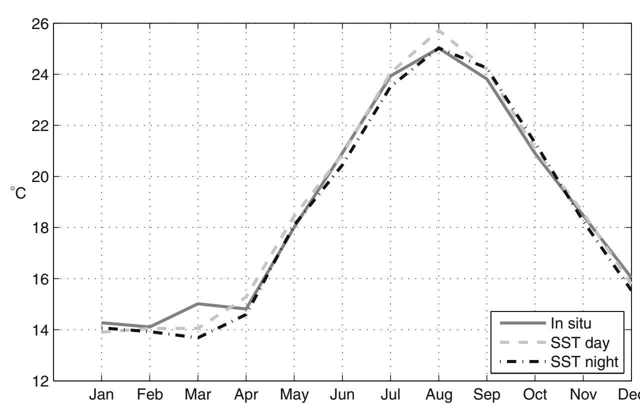


Fig. 4 Monthly average temperature for in situ, day-time and night-time satellite data

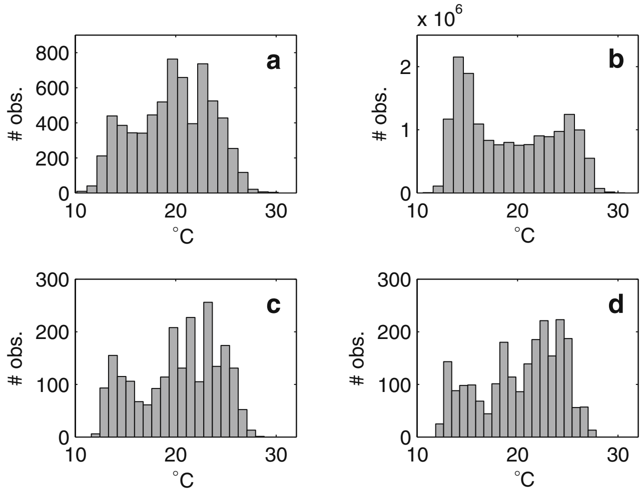


Fig. 5 Temperature distribution in 1999 of **a** all in situ data, **b** all day-time satellite data, **c** day-time satellite data interpolated at the in situ positions, and **d** night-time satellite data interpolated at the in situ positions. Note that the vertical scales are different in each subplot

The monthly average temperature for the in situ and satellite data (interpolated at the in situ data locations) is shown in Fig. 4. Both day-time and night-time satellite data reproduce closely the annual temperature cycle as described by the in situ data, although day-time satellite temperatures present an anomalously high temperature in August. Both day-time and night-time satellite data sets are about 1°C colder than in situ data in March.

Figure 5 shows the temperature distribution of the in situ and satellite data. The distribution for both types of data presents two peaks, a small one at about 14°C and a larger one at about 20°C to 25°C. In order to assess the effect on the heterogeneous spatio-temporal distribution of the in situ measurements, a histogram of the full day-time satellite SST data set is included as well (the distribution of night-time satellite SST is similar to day-time satellite SST). The full satellite data set distribution presents two peaks as well, although in this case the cold peak is much larger than the warm peak. The warm peak in the in situ data might be due to the higher number of these data being collected from September to November, which as seen on Fig. 4 have an average value of about 20°C to 22°C. Fewer in situ data are collected during the winter months, which is probably the reason why the cold peak is less pronounced than the warm peak.

In order to establish the effect of the presence of clouds in the satellite data distribution, we have calculated the distribution for the cloud-free SST data obtained by applying DINEOF (Beckers and Rixen 2003; Alvera-Azcárate et al. 2005) to the data set

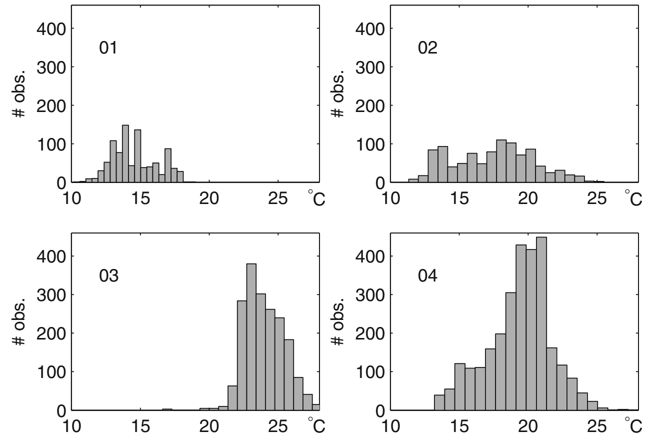


Fig. 6 Temperature distribution of in situ data for each season (*1*, January–March; *2*, April–June, *3*, July–September; *4*, October–December)

used in this work. The temperature distribution of a pentad AVHRR climatology (available at http://data.nodc.noaa.gov/pathfinder/CoralAtlas/PathfinderSST_Climatologies/5day/) has been calculated as well. Both of these data sets present a distribution similar to the cloudy data set in Fig. 5 (not shown), which gives us confidence in that it represents correctly the Western Mediterranean SST distribution. Given these results, we can confirm that the different spatial and temporal distributions of in situ and satellite data are causing the differences observed in the temperature distributions of Fig. 5.

The in situ data distribution has been also represented separately for each season (Fig. 6): the

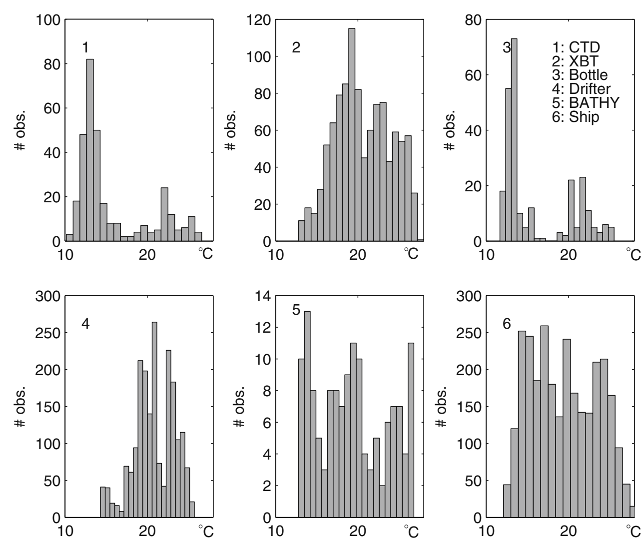


Fig. 7 Temperature distribution of in situ data for each sensor. The *x*-axis denotes the temperature in degrees Celsius, and the *y*-axis indicates the number of observations. Note the different scales in the vertical axes

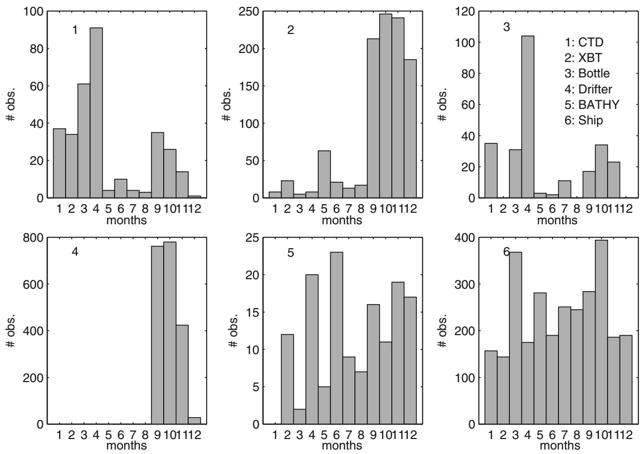


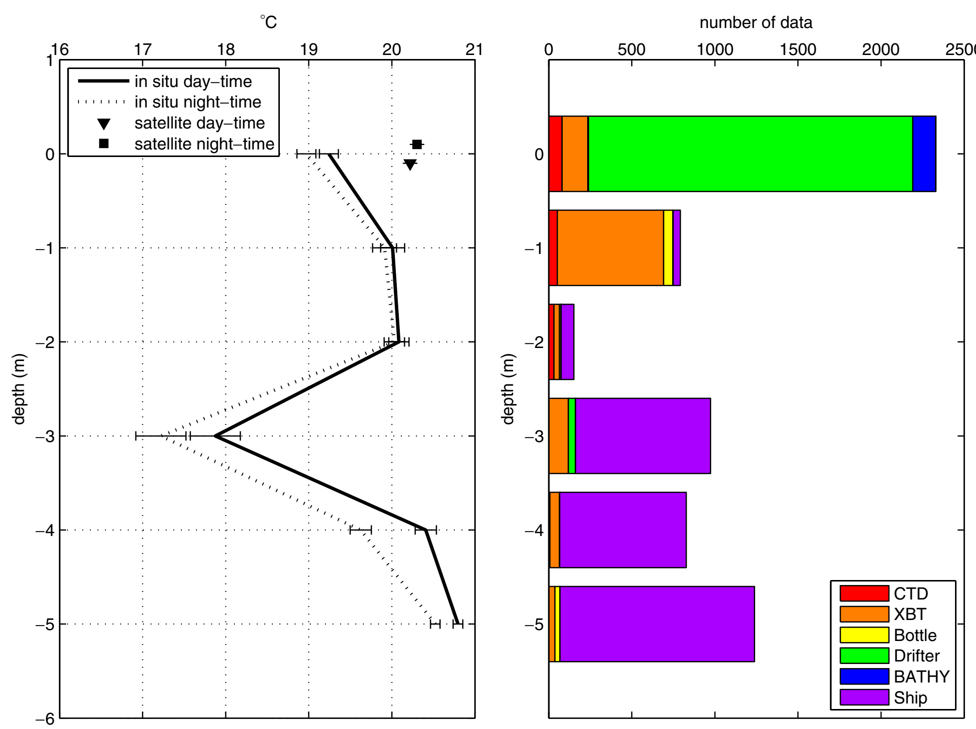
Fig. 8 Temporal distribution of in situ data. The *x*-axis denotes the months of the year, and the *y*-axis indicates number of observations. Note the different scales in the vertical axes

contribution of winter data (season 1, January to March) and spring data (season 2, April to June) to the cold peak becomes apparent in this figure, while the warm peak in the distribution of in situ data is due, as thought, to the high number of data taken in late summer and fall. The distribution of the satellite data interpolated to the in situ positions is very similar to the distribution of in situ data for each season, in terms of position and relative size of peaks and minimum and maximum values.

The temperature distribution for each sensor type, and their distribution in time (Figs. 7 and 8, respectively) add some information on the in situ data distribution: the warm peak is mainly due to drifter and XBT data, both data sets mostly taken from September to December. The cold winter peak consists mainly of ship, CTD and to a lesser extent, bottle data. Ship data distribution is quite homogeneous during the year (Fig. 8), therefore the histogram for this type of data (Fig. 7) is the most similar to the complete distribution of satellite data.

The average temperature for each depth considered (from 0 to 5-m depth) was calculated using all in situ data. In order to verify our approach of using all in situ data when comparing with satellite data, we divided the in situ data into day-time and night-time data, using the day-night distribution used in Fig. 3. The results are presented in Fig. 9, where the average temperature for day and night satellite data (interpolated to the in situ positions) is included as well. It can be seen that, for data up to 3 m depth, the difference between in situ data and satellite data is larger that any differences induced by the time of the day at which the data are taken. This validates the approach of using all in situ data (day and night time) in our subsequent comparisons. For data at 4 and 5 m depth, it is less straightforward to make this assertion. Also, it appears that in situ data at 1–2 m depth are the closest to the satellite estimate, which indicates that the satellite data are effectively

Fig. 9 *Left panel*, average temperature with depth for the in situ data, divided in night-time and day-time data. The *horizontal bars* represent the standard error of the mean. The average temperature for day and night satellite data (interpolated to the in situ positions) is also included. Note that the depth of the night and day satellite data is different from 0 m only to improve their readability. *Right panel*, number of data for each sensor category at each depth



representing the bulk temperature of the Mediterranean Sea for this particular data set. The temperature minimum observed at 3 m depth is mostly caused by ship data. These ship data are taken during all months in 1999 and homogeneously through all the Western Mediterranean Sea, so it is unlikely that these data come from a single mis-calibrated sensor. It may be possible that there is a bias inherent to a specific ship temperature sensor mounted at 3 m depth, or induced by the architecture of a specific type of ship. The identification of the particular type of ships carrying a temperature sensor at 3 m depth is however beyond the scope of this work. These data should probably be removed from the merging analysis.

4 Comparison between in situ and satellite data

As stated previously, the ultimate objective in mind for this work, although not part of it, is to use the error assessment for an optimal merging of the satellite and in situ data. Given the small number of in situ observations compared with satellite observations, the impact of the in situ data can be very limited in the final merged product. Because of that, in this section we compare all available in situ data to either the night-time or day-time satellite data. The error statistics obtained in this section will determine if using all in situ data (regardless of the hour of the day at which they are taken) to be merged with either night-time or day-time satellite data is a valid approach. A summary of the error statistics between in situ and satellite data can be found in Table 2. In general, both day-time and night-time satellite data present very similar RMS error and correlation, so a more detailed study is needed, dividing the in situ data by sensor type and by data set.

4.1 Error by sensor type

Several error measures (bias, root-mean squared (RMS) error, correlation and standard deviation of the different data sets) are used to assess the differences between in situ and satellite data. The last three measures are nicely condensed into the Taylor diagram (Taylor 2001), which is presented in Fig. 10 for the comparison between day-time satellite and in situ data, and in Fig. 11 for the comparison between night-time satellite and in situ data. For each of these comparisons, the data have been grouped into months (Figs. 10a and 11a), and into sensor-type (Figs. 10b and 11b). In this last case, the average value for each month is previously subtracted from the data to remove the annual cycle. Note that the satellite data are normalized by the standard deviation of the in situ data. In situ data are positioned in the x-axis with a standard deviation of 1 and the error of the other data being compared is established as the linear distance to this point (centered RMS), the angle to the x-axis (correlation) and distance to the origin (standard deviation, where data

Table 2 Errors between in situ data and AVHRR satellite data

	Bias (°C)	RMS (°C)	Correlation	Anomaly correlation
Day time	0.16	1.1	0.96	0.7
Night time	−0.12	1.2	0.95	0.7

Day time and night time refer to satellite time passage, but all in situ data are compared with each of these data sets

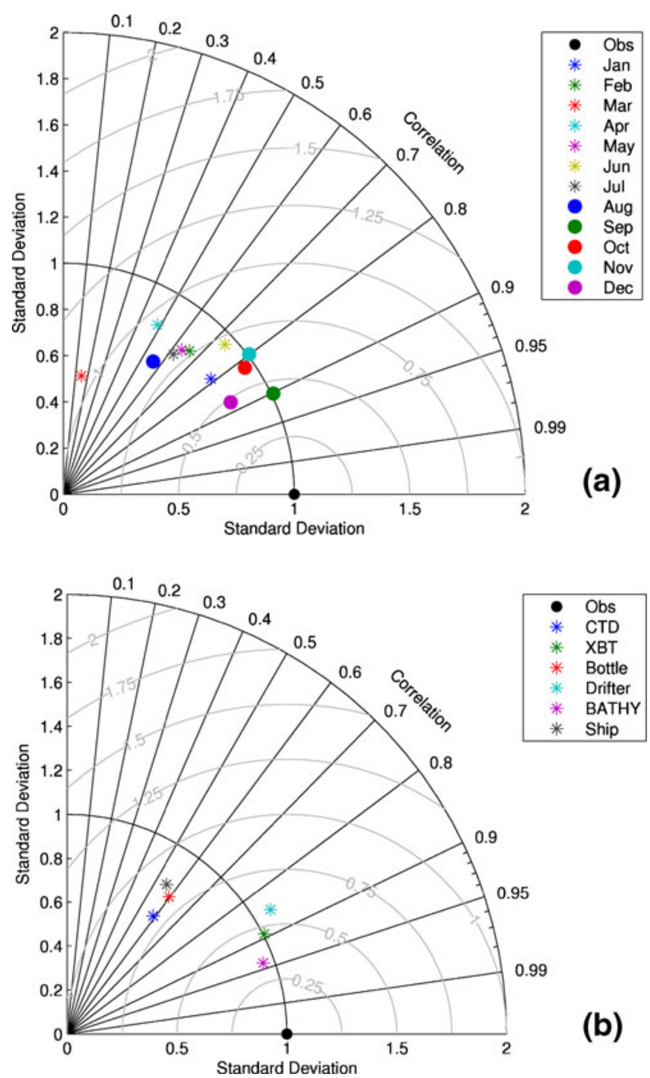


Fig. 10 Taylor diagram of the comparison between day-time satellite with in situ data for **a** each month and **b** each sensor. *Gray isolines* represent the normalized and centered RMS error

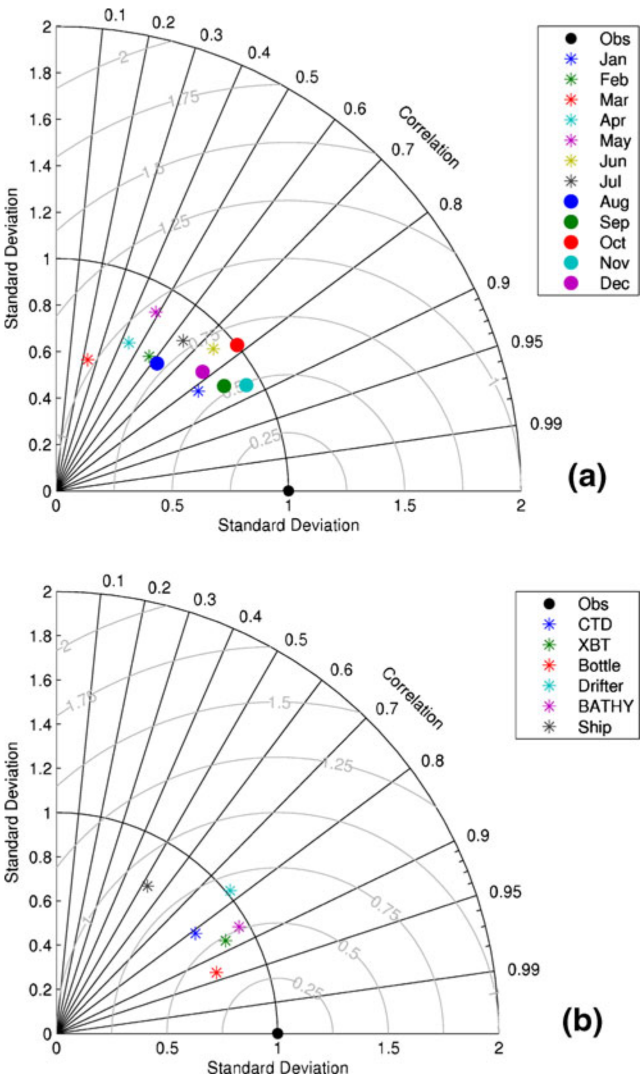


Fig. 11 Taylor diagram of the comparison between night-time satellite with in situ data for **a** each month and **b** each sensor. *Gray isolines* represent the normalized and centered RMS error

with a standard deviation lower/higher than one have a standard deviation lower/higher than the reference data set).

For both night and day satellite data, spring and summer months present the highest RMS errors and lowest correlations. There is no apparent difference between night-time and day-time satellite data, which would be indicated by a higher clustering of one of them around the reference point (the in situ data). If the Taylor diagrams are repeated without the ship data (figures not shown), then it becomes clear that night-time data are closer to the in situ data than day-time satellite data. It appears that the ship data introduce a higher variability in the comparisons. When looking at the error for each type of sensor, the comparison with night-time satellite data yields the best results, with ship data the platform with the worst performance, as already mentioned above. One of the best comparisons is obtained between bottle data and night-time satellite data. In the comparison with day-time satellite data, bottle, CTD and ship data present the highest errors. Note that the bias is not included in Taylor diagrams. The fact that night-time satellite data compares better to in situ data than day-time satellite data reinforces the practice of relating the in situ SST measurements to night-time satellite SST estimates, to avoid the problem of diurnal warming.

In order to complete the information included in Figs. 10 and 11, Table 3 contains the bias and centered RMS error (i.e., the RMS error without the contribution of the bias) of the satellite data respect to in situ data, and the standard deviation of both satellite and in situ data, divided again by sensor type. The highest bias is found in the comparison with ship data, with night-time satellite data half a degree colder than in situ data. The comparison with drifters also yields a high bias, for both night-time and day-time satellite data, and in this case the satellite data are warmer than the drifter data. The sign of the bias cannot be explained in either case by the average depth at which these measurements are taken (Table 1), because ship data are the deepest (4 m depth in average) and drifter data are one of the shallowest (0.06 m depth in average). Our results agree with other works that account for a cold bias in drifter data respect to ship data (e.g., Emery et al. 2001a; Ingleby 2010), and a warm bias in ship data (e.g., Kent et al. 1993, 2010), more specifically in engine room intake measurements. Given that the ship data set we are using consists mostly of engine room intake measurements (78.5%), the observed warm bias must be mostly due to this effect. The centered RMS error is the highest when comparing satellite data with ship data, CTD data, and BATHY data.

The error assessment mentioned above compares satellite and in situ data only at those positions were satellite data is available. In other words, when clouds are present the error between in situ and satellite data can obviously not be calculated. Any error measure concerning infrared satellite data will result in a "clear-sky" estimation. This explains the apparent contradiction between, for example, the warm bias of ship data presented in Table 3 and the average temperature (colder than satellite data) of ship data in Fig. 9. In Fig. 9, all ship data are used to calculate the average value, whereas in Table 3 only ship data for which there are matching satellite observations are used. Considering that the surface of the sea might be warmer under clear sky conditions (because of the increased solar radiation reaching the surface of the sea), it is

Table 3 Errors between in situ data and AVHRR satellite data, for each sensor type

		Bias(°C)	Centered RMS(°C)	σ satellite (°C)	σ in situ(°C)
CTD	Day time	0.42	0.75	4.33	4.33
	Night time	−0.1	0.61	4.6	
XBT	Day time	−0.0	0.69	3.43	3.37
	Night time	−0.15	0.66	3.63	
Bottle	Day time	0.27	0.82	4.47	4.18
	Night time	−0.05	0.51	4.54	
Drifter	Day time	0.48	0.52	2.19	2.43
	Night time	0.42	0.68	2.47	
Bathy	Day time	−0.18	0.68	4.0	4.25
	Night time	−0.33	0.89	3.85	
Tesac	Day time	−0.27	0.12	0.13	0.47
	Night time	–	–	–	
Ship	Day time	−0.1	1.4	4.5	4
	Night time	−0.5	1.5	4.4	

Day time and night time refer to satellite time passage, but all in situ data are compared with each of these data sets

therefore expected that the in situ data average is warmer when we consider only those points where there are no clouds.

The particular contribution of each sensor/platform type to the overall bias is presented as an histogram in Fig. 12, again divided by sensor type. A peak is found at about −4°C caused by ship data, both for the day-time and night-time satellite data sets. This appears to indicate an erroneous set of ship data with values much warmer than satellite data, rather than an wrongly detected cold zone in the satellite data set (for example an undetected cloud), which will likely appear at more than one sensor type. Further investigation on the origin of this peak in the histogram reveals that these are data taken consecutively from the 10th to the 30th March 1999 in the Gulf of Lions area, which indicates that these data may come from a single ship. The difference of about 1°C between satellite and in situ data during March (Fig. 4) is explained by the presence of these ship data. The histogram in Fig. 12 also shows the positive bias of satellite data respect to drifter data, and the warm bias of night-time satellite data respect to CTDs.

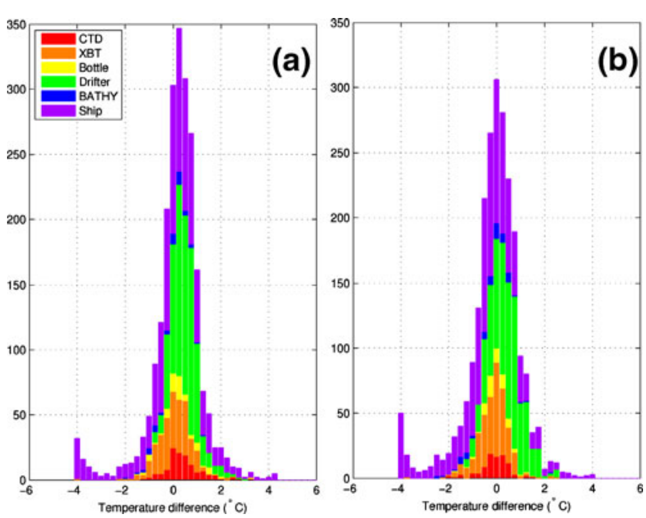


Fig. 12 Histogram of the difference between satellite and in situ data, for **a** day-time satellite data and **b** night-time satellite data. Negative (positive) values represent a cold (warm) bias in the satellite data with respect to the in situ data

Visually, some of the distributions appear to be skewed. In order to test if the distributions can be described as Gaussian, an Anderson–Darling test (Anderson and Darling 1952) has been applied. For all data types, together the test concluded that with a confidence of 99%, the difference distribution is not normal for both the day-time and night-time cases. Applying the test sensor by sensor, all sensors present a non-normal distribution at the 99% level of confidence, except BATHY for day-time data and bottles and drifters for night-time data, for which the hypothesis of non-normality cannot be rejected. The non-normality of the data difference distributions has important consequences for several analysis techniques, as in data assimilation and the merging of the satellite and in situ data sets.

4.2 Error by database

A last test is performed to assess the quality of the in situ data organized by database. The results of this assessment are presented in Fig. 13 and in Table 4. ICOADS data are composed solely of ship data, and the rest of the databases provided data from all the other platforms except ship data. This makes difficult the comparison between ICOADS and the rest of the databases. As already discussed in the previous subsection, ship data from ICOADS presents the highest RMS error and biases, which is reflected in Fig. 13

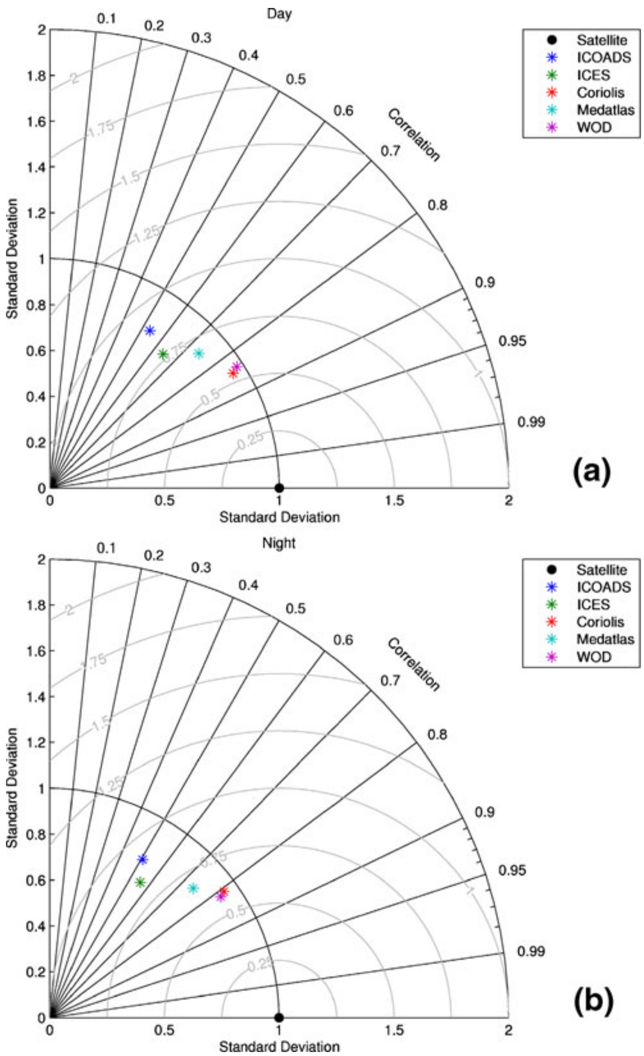


Fig. 13 Taylor diagram presenting the error between in situ data (organized by database) and satellite data (used as reference). *Top panel*, day-time satellite data are used. *Bottom panel*, night-time satellite data are used. The annual cycle has been subtracted before the calculation of the errors. *Gray isolines* represent the normalized and centered RMS error

and Table 4. ICES data follow in the Taylor diagram. There is little difference in the quality of each database when comparing with day-time or night-time satellite data. Apart from ICOADS, the databases WOD and Coriolis present the highest bias (Table 4), with satellite colder than in situ data for WOD and warmer than in situ data for Coriolis. In general, all presented databases have very similar correlation with satellite data, although when the annual cycle is removed, ICOADS data and ICES data perform poorly in terms of correlation.

5 Conclusions

A comparison between in situ and satellite sea surface temperature (SST) has been realized in the Western Mediterranean Sea for 1999. Five international databases have been used to extract in situ data for the desired period and zone: World Ocean Database (WOD), MEDAR/Medatlas, Coriolis, International Council for the Exploration of the Sea (ICES) and International Comprehensive Ocean-Atmosphere Data Set (ICOADS). The in situ data have been classified into different platforms or sensors, in order to compare the relative accuracy of these type of data respect to the satellite data. The statistics obtained during this study will be used in future work for merging purposes.

Ship data, from the ICOADS database, are the most numerous, and as such they are a valuable source of data. However, the error assessment between these data and satellite data shows a large bias and RMS error. A series of suspect data were identified with a large bias (more than 3°C warmer than satellite data) and that presumably came from a single ship as they were localized in time and space. One must bear in mind that the ships collecting surface temperature data are very heterogeneous in size, and therefore the resulting measurements are very heterogeneous as well. In addition, measurements from ships (and other platforms in general) are not made with the purpose of complementing satellite data, therefore they do not necessarily represent the same temperature.

Table 4 Errors between in situ data and AVHRR satellite data, for each data base

		Bias(°C)	Centered RMS (°C)	Correlation	Anomaly corr.
ICOADS	Day time	0.05	1.43	0.95	0.53
	Night time	0.5	1.5	0.94	0.51
ICES	Day time	−0.36	0.72	0.99	0.64
	Night time	0	0.48	0.99	0.56
Coriolis	Day time	0.11	0.64	0.98	0.86
	Night time	0.24	0.66	0.98	0.81
Medatlas	Day time	−0.01	0.68	0.98	0.81
	Night time	0.15	0.56	0.99	0.74
WOD	Day time	−0.39	0.62	0.98	0.84
	Night time	−0.25	0.73	0.97	0.82

Day time and night time refer to satellite time passage, but all in situ data are compared with each of these data sets

Other types of data performed more homogeneously, with RMS errors of 0.6°C to 0.9°C and small biases. The largest bias was detected for drifter data, which was in average 0.48°C and 0.42°C colder than day-time and night-time satellite data, respectively. This cold bias in drifter data is maybe unexpected given that the drifters average depth was of 0.06 m, one of the shallowest among the platforms considered in this work. The fact that satellite data are calibrated using in situ buoys measuring bulk temperatures can be the reason for this bias. If a given sensor is directly exposed to the air, for example during calm sea conditions, this might induce as well a cold bias in the drifters measurements. The fact that most of the drifters were released during fall in 1999, a period during which there can have been a cooling of the air in the Gulf of Lions region, reinforces this possibility. If the exposure to the air is responsible for the cold bias during fall, then the opposite should be verified too, i.e., that drifters deployed during summer and under calm sea conditions will present a warm bias. This cannot be verified in this work, as there were no drifter data deployed during summer. The smallest bias between satellite data and in situ data was observed at 1–2 m depth, which confirms that the satellite are representing the bulk temperature of the Mediterranean Sea, at least for this particular data set.

The satellite-in situ SST difference distribution is generally not normal, as shown by an Anderson–Darling test at the 99% level of confidence. This result is obtained when using all in situ data types and for individual sensor types. Only the difference distribution between satellite and BATHY, bottle and drifter data cannot be considered non-normal using the mentioned test. The non-normality of the data difference distributions has important consequences for several analysis techniques, as in data assimilation and the merging of the satellite and in situ data sets. This factor therefore needs to be taken into account in future work.

Apart from the error assessment by data type, the average error for each database was as well calculated. ICOADS, containing only ship data, presents the highest errors. The rest of the databases have similar RMS errors among them, but in terms of bias, WOD and ICES presented the highest deviations respect to satellite data.

The comparison of satellite infrared data with in situ data is limited by the presence of clouds in the atmosphere, which prevents the infrared radiation from the sea surface to reach the satellite. The error measures presented in this work reflect therefore clear-sky conditions. The presence or absence of clouds influences the sea surface temperature. In the absence of clouds, the solar radiation reaching the surface of the sea during the day increases. The absence of clouds affects also the net long-wave radiation budget at the ocean surface (the amount of long-wave radiation reflected back to the ocean is reduced). This may specially affect the bias between in situ and satellite data, as well as other error measures.

The results obtained in this work emphasize that the differences between in situ and satellite SST data can be affected by various factors (database, sensor or platform type, specific bias at a particular platform, etc). A careful study of these differences is needed prior to any work aiming to use these sources of data in a joint manner.

Acknowledgements This work was realized in the context of the HiSea—SR/12/140 project funded by the Belgian Science Policy (BELSPO) in the frame of the Research Program For Earth Observation "STEREO II". The AVHRR Oceans Pathfinder SST data were obtained through the online PO.DAAC Ocean ESIP Tool (POET) at the Physical Oceanography Distributed Active Archive Center (PO.DAAC), NASA Jet Propulsion Laboratory, Pasadena, CA. The different providers for the databases used (World Ocean Database, MEDAR/MedAtlas, Coriolis Data Center, the International Council for the Exploration of the Sea and the International Comprehensive Ocean-Atmosphere Data Set) are also acknowledged for their important work of compiling extensive amounts of data and making them publicly available. The National Fund for Scientific Research, Belgium, is acknowledged for funding the post-doctoral positions of A. Alvera-Azcárate and A. Barth and for funding C. Troupin's thesis through a FRIA grant. This is a MARE publication.

References

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 surface temperature. Ocean Model 9:325–346. doi:10.1016/j.ocemod.2004.08.001

Alvera-Azcárate A, Sirjacobs D, Barth A, Beckers J-M (2011) Outlier detection in satellite data using spatial coherence. Rem Sens Environ (submitted)

Anderson TW, Darling DA (1952) Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Ann Math Stat 23:193–212

Barton I (2007) Comparison of In Situ and Satellite-Derived Sea Surface Temperatures in the Gulf of Carpentaria. J Atmos Oceanic Technol 24:1773–1784

Beckers J-M, Rixen M (2003) EOF calculations and data filling from incomplete oceanographic data sets. J Atmos Oceanic Technol 20(12):1839–1856

Castro SL, Wick GA, Jackson DL, Emery WJ (2008) Error characterization of infrared and microwave satellite sea surface temperature products for merging and analysis. J Geophys Res 113(C03010). doi:10.1029/2006JC003829

Donlon C, Casey KS, Robinson IS, Gentemann CL, Reynolds RW, Barton I, Arino O, Stark J, Rayner N, Le Borgne P, Poulter D, Vazquez-Cuervo J, Armstrong E, Beggs H, Llewellyn-Jones D, Minnett PJ, Merchant CJ, Evans R

(2009) The GODAE high-resolution sea surface temperature pilot project. Oceanography 22(3):34–45

Emery W, Baldwin DJ, Schlüssel P, Reynolds R (2001a) Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. J Geophys Res 106(C2): 2387–2405

Emery W, Castro S, Wick G, Schluessel P, Donlon C, (2001b). Estimating sea surface temperature from infrared satellite and in situ temperature data. Bull Am Meteorol Soc 82(12):2773–2785

Gentemann C, Minnett P, Sienkiewicz J, DeMaria M, Cummings J, Jin Y, Doyle J, Gramer L, Barron C, Casey K, Donlon C (2009) MISST: the multi-sensor improved sea surface temperature project. J Oceanogr 22(2):78–89

Guan L, Kawamura H (2004) Merging satellite infrared and microwave SSTs: methodology and evaluation of the New SST. J Oceanogr 60:905–912

Ingleby B (2010) Factors affecting ship and buoy data quality: a data assimilation perspective. J Atmos Ocean Technol 27:1476–1489

Kent C, Kennedy J, Berry D, Smith R (2010) Effects of instrumentation changes on sea surface temperature measured in situ. WIREs Climate Change 1(5):718–728. doi: 10.1002/wcc.55

Kent C, Taylor P, Truscott B, Hopkins J (1993) The accuracy of voluntary observing ships' meteorological observations - results of the VSOP-NA. J Atmos Ocean Technol 10:591–608

Kent E, Challenor P (2006) Toward estimating climatic trends in SST. Part II: random errors. J Atmos Ocean Technol 23:476–486

Locarnini RA, Mishonov AV, Antonov JI, Boyer TP, Garcia HE (2006) World ocean atlas 2005, volume 1: temperature. tech. rep., NOAA. Washington. 182 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

Reynolds R, Smith T, Liu C, Chelton D, Casey K, Schlax M (2007) Daily high-resolution-blended analyses for sea surface temperature. J Climate 20:5473–5496

Robinson IS (2004) Measuring the oceans from space. The principles and methods of satellite oceanography. Springer, London. 669 pp

Sirjacobs D, Alvera-Azcárate A, Barth A, Lacroix G, Park Y, Nechad B, Ruddick K, Beckers J-M (2011) Cloud filling of ocean color and sea surface temperature remote sensing products over the southern north sea by the data interpolating empirical orthogonal functions methodology. J Sea Res 65:114–130

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res 106(D7):7183–7192

Xu F, Ignatov A (2010) Evaluation of in situ sea surface temperatures for use in the calibration and validation of satellite retrievals. J Geophys Res 115:C09022. doi:10.1029/2010JC006129

Annex II

Detection of outliers in satellite data using spatial coherence

Remote Sensing of Environment 119 (2012) 84–91

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Outlier detection in satellite data using spatial coherence

A. Alvera-Azcárate [a,b,*], D. Sirjacobs [c], A. Barth [a,b], J.-M. Beckers [a]

[a] AGO-GHER-MARE, University of Liège, Allée du 6 Août 17, B5, Sart Tilman, 4000 Liège, Belgium
[b] National Fund for Scientific Research, FNRS-FRS, Belgium
[c] Department of Life Sciences, Boulevard du Rectorat 27, B22, Sart Tilman, 4000 Liège, Belgium

ARTICLE INFO

Article history:
Received 15 September 2010
Received in revised form 8 November 2011
Accepted 3 December 2011
Available online xxxx

Keywords:
Outlier detection
Satellite data
Empirical Orthogonal Functions
Sea surface temperature
Chlorophyll-a concentration

ABSTRACT

Satellite data sets often contain outliers (*i.e.*, anomalous values with respect to the surrounding pixels), mostly due to undetected clouds and rain or to atmospheric and land contamination. A methodology to detect outliers in satellite data sets is presented. The approach uses a truncated Empirical Orthogonal Function (EOF) basis. The information rejected by this EOF basis is used to identify suspect data. A proximity test and a local median test are also performed, and a weighted sum of these three tests is used to accurately detect outliers in a data set. Most satellite data undergo automated quality-check analyses. The approach presented exploits the spatial coherence of the geophysical fields, therefore detecting outliers that would otherwise pass such checks. The methodology is applied to infrared sea surface temperature (SST), microwave SST and chlorophyll-a concentration data over different domains, to show the applicability of the technique to a range of variables and temporal and spatial scales. A series of sensitivity tests and validation with independent data are also conducted.

1. Introduction

Sea surface temperature (SST) data sets undergo a series of quality-check analyses aimed at removing data contaminated with aerosols, clouds, rain, dust, etc. Most quality-check procedures act on a pixel-by-pixel basis (*e.g.*, Esaias et al., 1998; Kilpatrick et al., 2001), so the spatial coherence of the data is not adequately exploited. As a consequence, outliers may remain in the products made available for research and monitoring purposes (Donlon et al., 2002; Lazarus et al., 2007; Merchant et al., 2008b). Some tests do exploit the spatial coherence of the data (*e.g.*, Coakley & Bretherton, 1982), as applied by May et al. (1998), although not all suspect data may be flagged as bad or removed from the final data set. Stringent tests need to be applied in order to remove these outliers from the data set, although there is no consensus on how best to do this. The presence of these outliers in satellite SST data introduces biases that make the comparison between different satellite products difficult. The validation of different satellite data sets is consequently also made difficult, as is the correction of possible differences between them. Some applications, such as data assimilation, can also be affected by the presence of these outliers, which must be removed before assimilation (*e.g.*, Okamoto & Derber, 2006).

In this study we propose a methodology to detect outliers in satellite data sets that uses Empirical Orthogonal Functions (EOFs) to highlight suspect data. This EOF basis is calculated by means of DINEOF (Data Interpolating EOFs, Beckers and Rixen (2003), Alvera-Azcárate et al. (2005)), a technique to reconstruct missing values in satellite data sets. Along with the EOF basis test, two additional tests are used to increase the robustness of the detection of outliers: a proximity test (as most outliers occur at the edges of clouds, land, rain or the satellite swath) and a local median test.

The study is organized as follows. In Section 2, the data sets used for testing and validating the outlier detection technique are presented. Section 3 describes the methodology used in this study and this technique is then applied to an infrared SST data set and validated in Section 4. In order to demonstrate the applicability of the technique to other variables, an application using microwave SST and one using chlorophyll-a concentration are presented respectively in Sections 5 and 6. Conclusions are presented in Section 7.

2. Data sets

Advanced Very High Resolution Radiometer (AVHRR) SST fields produced by the Ocean & Sea Ice Satellite Application Facility (O&SI SAF) were obtained through the Medspiration website (ftp://ftp.ifremer.fr/ifremer/medspiration/data/). The domain of study is the western Mediterranean Sea (see Fig. 1), and the data have been interpolated on a ~2 × 2 km grid. There are two SST estimates per day, but only daytime data are used in this study, in order to test the robustness of the outlier detection technique in the presence of diurnal warming

* Corresponding author at: AGO-GHER-MARE, University of Liège, Allée du 6 Août 17, B5, Sart Tilman, 4000 Liège, Belgium. Tel.: +32 43663664.
E-mail address: a.alvera@ulg.ac.be (A. Alvera-Azcárate).
URL: http://modb.oce.ulg.ac.be/aida (A. Alvera-Azcárate).

0034-4257/$ – see front matter
doi:10.1016/j.rse.2011.12.009

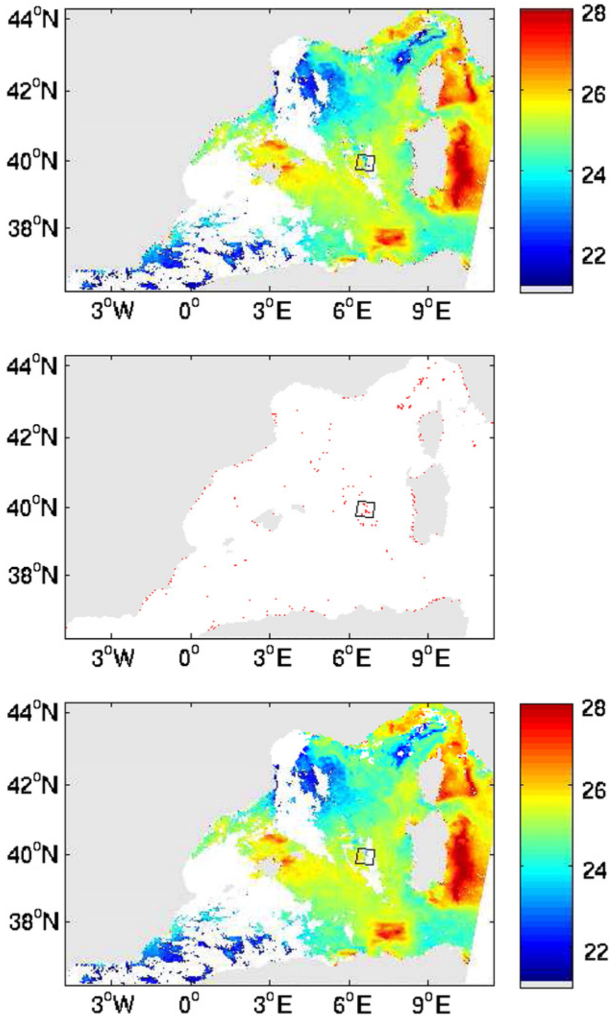


Fig. 1. Study domain: the western Mediterranean Sea. Top panel: SST (°C) on 4 July 2010. Middle panel: outliers detected. Bottom panel: SST data with outliers removed. The black square indicates the zone where a detail is shown in Fig. 2.

events (which should not be detected as outliers). Six months of data are used, from 7 January 2010 to 7 July 2010.

Daily global SST data from the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) from 1 August 2010 to 31 December 2010 are also used, in order to verify the applicability of the methodology to outliers due to the presence of rain and to test the methodology on a global scale. These data have been interpolated on a 0.25×0.25 degree grid, and daytime passes are used, as for the AVHRR data set. The data were downloaded from ftp://ftp.ssmi.com/.

Another test was carried out using chlorophyll-a concentration from the Sea-viewing Wide Field-of-view Sensor (Sea-WiFS), on board the SeaStar spacecraft (http://oceancolor.gsfc.nasa.gov/). These data are 8-day composites covering the Caribbean Sea from 1 January 2004 to 31 December 2004, and have a spatial resolution of 0.1 degrees.

For the verification of the results obtained, level 3 MODIS Aqua SST data at a spatial resolution of 4 km were used in the domain and time frame of the AVHRR data set. These data were downloaded from http://oceancolor.gsfc.nasa.gov/. In addition, Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) data were also used for the validation of the western Mediterranean Sea data set. These data are an analysis incorporating SST information from various satellites, and are available daily with a resolution of about 5 km (Stark et al., 2007). These data were downloaded through the MyOcean portal (http://operation.myocean.eu/).

3. Method

3.1. DINEOF description

DINEOF (Data INterpolating Empirical Orthogonal Functions) is a parameter-free technique based on an iterative EOF decomposition to calculate missing data in satellite data sets. A temporal and spatial average is removed from the data, and the missing data are initialised to zero (*i.e.*, to an unbiased first guess). The first EOF mode is then calculated from this data set, which is used to infer a new estimate for the missing data. This procedure is repeated until convergence is obtained for the values given to the missing data with the first EOF mode. Subsequently, the two leading EOFs are taken and the process is repeated until convergence; then three EOF modes are used, and so on. The optimal number of EOFs needed to calculate the values at the missing locations is determined by cross-validation: a small percentage of valid data (typically 1% of the total data) are initially set apart and flagged as missing. Once convergence is reached for a given number of EOF modes, a root mean square error is calculated between the newly obtained estimate and the initial data set. The number of modes that minimises this error is considered optimal. Note that not all modes need to be calculated, as one can consider that if the error increases steadily for 3 consecutive modes, a minimum has been reached. Error maps can be calculated for the reconstructed data using an Optimal Interpolation approach (e.g., Daley, 1991) in which the DINEOF EOF basis is used to construct a covariance field (Beckers et al., 2006). DINEOF was first described in Beckers and Rixen (2003), and an adaptation to handle the large data sets typical of satellite imagery can be found in Alvera-Azcárate et al. (2005).

The EOF basis calculated within DINEOF is based on the mean and covariance of the original data. The probability distribution of the data can be completely defined by the mean value of the data and the EOF basis if the original data are normally distributed. While this is the case for SST, other variables, however, do not present a Gaussian distribution (*e.g.*, biological variables such as chlorophyll-a concentration, or total suspended matter). In these cases, a transformation of the original data needs to be performed prior to the DINEOF analysis. A logarithmic transformation can be used, for example, although other transformations may also be used.

3.2. Outlier detection

Using the truncated EOF basis retained by DINEOF as optimal to reconstruct the missing data, and interpreting the information contained in the higher-order EOFs that have been discarded as noise, outliers can be detected within DINEOF as those pixels for which the analysis-observation difference (the residuals) is larger than the statistically expected misfit calculated during the analysis. Sirjacobs et al. (2011) used the ratio between the analysis residuals and the expected standard deviation of the residuals to identify outliers in a given data set. For a given time, the normalised residual is computed by:

$$O_i = \frac{X_i^a - X_i^o}{\Delta_i}, \quad \text{for } X_i^o \text{ not missing} \tag{1}$$

where $i = 1, \ldots, m$ is the spatial index, X_i^a is the analysed value, reconstructed by DINEOF, X_i^o is the original data set (*i.e.*, before applying DINEOF, and where some indexes are undefined due to missing data) and Δ_i is the expected misfit calculated as:

$$\Delta_i = \sqrt{\mu_{\text{eff}}^2 - \sum_{k=1,N} E_{i,k}^2} \tag{2}$$

where N is the number of EOFs retained by DINEOF for the reconstruction and $k = 1, \ldots N$. μ_{eff}^2 is an estimation of the average noise in the original field, calculated as the cross-validation error obtained with DINEOF (normalised by the spatial correlation length of the data error to account for the correlation of the error in the data set; see Beckers et al. (2006) for more details). $\sum_{k=1,N} E_{i,k}^2$ is the expected error for each pixel i, calculated as:

$$\mathbf{E} = \mathbf{L}_p \mathbf{S}_C \tag{3}$$

where the columns of $\mathbf{L}_p$ (size $m \times N$) are the spatial EOFs multiplied by the corresponding singular values, $\mathbf{S}_C$ (size $N \times N$) is a square root factorisation (Cholesky factorisation) of $\mathbf{C}$, which is given by:

$$\mathbf{C} = \mathbf{S}_C \mathbf{S}_C^T = \mu_{\mathrm{eff}}^2 \left(\mathbf{L}_p^T \mathbf{L}_p + \mu_{\mathrm{eff}}^2 \ \mathbf{I}_N \right)^{-1} \tag{4}$$

with $\mathbf{I}_N$ the identity matrix of size $N \times N$. The threshold value to classify a given pixel as an outlier following Eq. (1) is proposed to be 3 in Sirjacobs et al. (2011) meaning that for a Gaussian-distributed misfit, 0.3% of the data would fall into this category. However, if the expected misfit Δ is not accurately estimated, the detection of outliers using this approach is not robust. This shortcoming is even more important as we are interested in the extreme values of O.

In order to mitigate this problem, we propose in this study an improvement to the outlier classification proposed by Sirjacobs et al. (2011). The median and the Median Absolute Deviation (MAD) are used instead of the standard deviation, as statistics based on the median are more robust to the presence of outliers in the data than statistics based on the standard deviation (Wilks, 1995). The median of the outlier index O (obtained with Eq. 1) is calculated for each image:

$$O_m = \mathrm{median}(O) \tag{5}$$

and the MAD between these two quantities is performed:

$$\delta = \mathrm{mad}(O) = 1.4826 \ \ \mathrm{median}|O - O_m| \tag{6}$$

The factor 1.4826 is introduced such that, for a normal distribution, the MAD is equal to the standard deviation (*e.g.*, Reimann et al., 2008). An index of the likelihood that a given pixel is an outlier is therefore given by:

$$O_{\mathrm{eof}} = \left| \frac{O - O_m}{\delta} \right| \tag{7}$$

This test examines the spatio-temporal coherence of the data, penalising those pixels that are inconsistent with the EOF basis. To strengthen the outlier classification, two additional tests are performed. A proximity test is performed in order to penalise the proximity to cloud, rain or land pixels, as many outliers in satellite data are contaminated by these. For a given image, if a pixel is originally classified as cloud, rain or land, all pixels in its vicinity (typically one pixel, but this can be modified) are penalised as potential outliers, so that $O_{\mathrm{prox}} = 3$ for those pixels and $O_{\mathrm{prox}} = 0$ for the rest.

Finally, as a third test, a local median is calculated for each image over a given window size. The MAD is again calculated for the data inside a window of a given size:

$$\delta_{\mathrm{median}} = \mathrm{mad}(X^o) \tag{8}$$

and an estimation of outlier pixels can be therefore obtained:

$$O_{\mathrm{median}} = \left| \frac{X^o - \mathrm{median}(X^o)}{\delta_{\mathrm{median}}} \right| \tag{9}$$

The final classification of a pixel as an outlier is made by a weighted sum of the three tests described above:

$$O_{\mathrm{final}} = w_{\mathrm{eof}} O_{\mathrm{eof}} + w_{\mathrm{prox}} O_{\mathrm{prox}} + w_{\mathrm{median}} O_{\mathrm{median}} \tag{10}$$

where:

$$w_{\mathrm{mad}} + w_{\mathrm{prox}} + w_{\mathrm{median}} = 1 \tag{11}$$

The weights give the possibility of penalising more heavily the aspects considered to be more problematic in a given data set.

4. Application to AVHRR data

4.1. EOF basis determination using DINEOF

By applying DINEOF to the six months of AVHRR data described in Section 2, a total of 15 EOFs were found to be optimal using the cross-validation technique. These EOFs explain 99.21% of the total variability. The remaining 0.79% of variability, filtered out by the EOF basis, consists mainly of noise, although it may also contain small scale and transient features that have too weak a signal to be retained in the first 15 EOFs. Although very limited, some small scale information might therefore be lost from the initial data set.

4.2. Outlier detection

The outlier detection method was then applied to the AVHRR data set. Clouds were used as the factor to classify outliers in the proximity test. Fig. 1 shows an example of the original data on 4 July 2010, along with the detected outliers (red dots in the middle panel) and the original data with the outliers removed. An equal value of 1 / 3 is given to all three weights described in Section 3, and the size of the window over which the median is calculated is 20×20 pixels. The threshold for O_{final}, above which a pixel is considered an outlier, has been set to 3. Several types of outliers are present in the original data: near clouds, scattered in cloudless zones and along the coastline. The technique is able to detect most of these, as can be seen in the middle panel of Fig. 1. The quality of the SST once the outliers have been removed is improved (bottom panel of Fig. 1).

Note that large zones of very high SST located east of Corsica and Sardinia islands are not classified as outliers. These zones are affected by diurnal warming because of the blocking of westerly winds by the islands mountains (Merchant et al., 2008a). In order to test the impact of the domain size in the classification of these warming events, a larger domain of the AVHRR data set was used. This larger domain contains the western Mediterranean Sea and the north-east Atlantic Ocean, extending over ~3000 km in latitude and ~5200 km in longitude (compared to the ~1200 km by ~2100 km of the domain presented in Fig. 1), and over the same time period. The EOF basis retained by DINEOF consists of 11 EOFs, which account for 99.83% of the total variability. Applying the same criteria to the detection of outliers as in the example described above, the diurnal warming events seen in Fig. 1 are not classified as outliers by our technique (data not shown). Therefore, the outlier detection technique does not penalise large zones of anomalous data resulting, for example, from an atmospheric event, and this is true for different domain sizes.

A detail of the western Mediterranean domain is shown in Fig. 2 (see Fig. 1 for location). Together with the original data, the result of each of the three tests is shown. Also included is the SST without outliers for two thresholds: 2.5 and 3, as well as the pixels classified as outliers for these two cases. Pixels that appear in dark red (values larger than 3) for each test are classified as outliers in this example. In this example, the O_{prox} test is the strictest (classifying more pixels as outliers) so the addition of the O_{eof} and O_{med} tests modulates this classification.

4.3. Validation of the outlier detection technique

In order to verify the accuracy of the method in detecting outlier data, the Root Mean Square (RMS) difference was calculated between

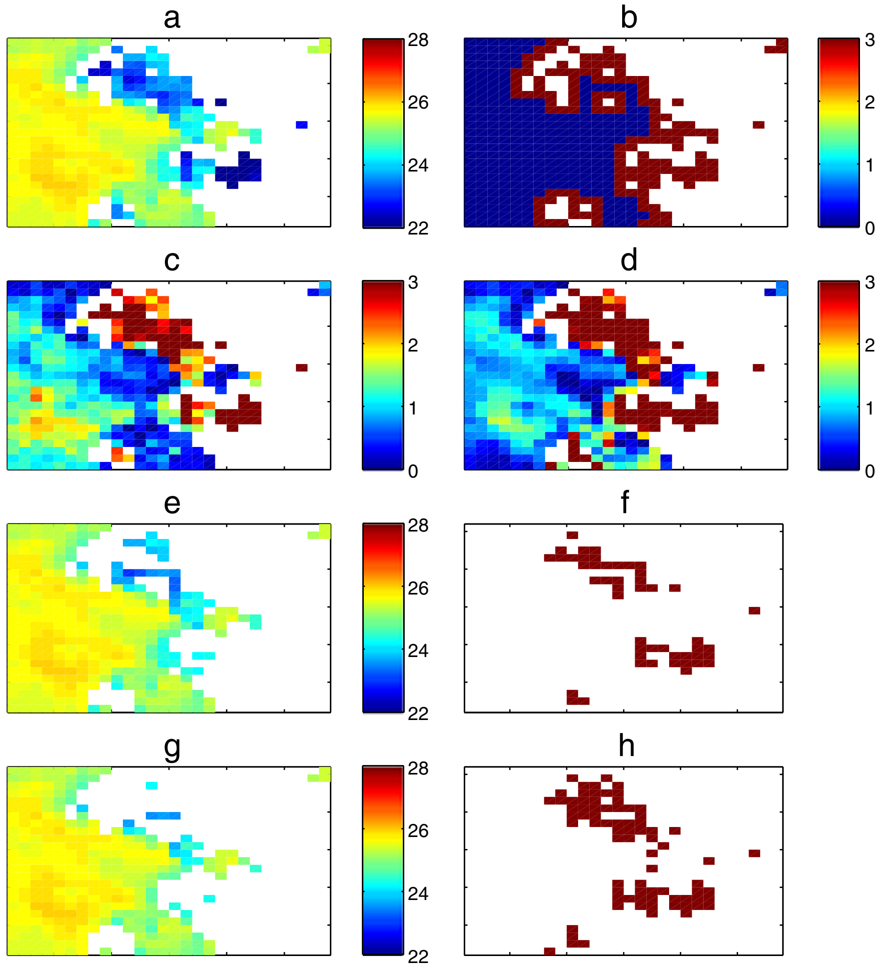


Fig. 2. SST detail (°C) and outlier tests in the small domain of the western Mediterranean Sea shown in Fig. 1. Panel a: original data; panel b: proximity test; panel c: EOF test; panel d: local median test; panel e: SST withoutliers removed, applying a threshold of 3; panel f: outliers detected with a threshold of 3; panel g: SST with outliers removed, applying a threshold of 2.5; panel h: outliers detected with a threshold of 2.5. See text for a detailed description of each sub-test.

the AVHRR SST field and two reference fields: MODIS Aqua SST and OSTIA SST fields, both for 4 July 2010. Table 1 shows the RMS difference for various combinations of weights and threshold values, along with the number of outliers detected with each case. Only the RMS difference for pixels classified as outliers is presented. For pixels not classified as outliers, the RMS difference is of about 0.9 °C for the comparison between AVHRR and MODIS and 1.4 °C for the comparison between AVHRR and OSTIA, regardless of the weights and threshold used. This is due to the higher number of data that enter this computation, making this statistic more robust.

Table 1
RMS difference between AVHRR and two reference data sets (MODIS SST and OSTIA SST) on 4 July 2010. Different weights and thresholds are applied. The number of data classified as an outlier for each case is included (as the total number and as the percentage with respect to the whole data set). Only the RMS difference for pixels classified as outliers is presented.

w_{eof}	w_{prox}	w_{med}	Threshold	RMS_{MODIS} (°C)	RMS_{OSTIA} (°C)	number of outliers (%)
1/3	1/3	1/3	3	2.6	3.13	1335 (0.7%)
1/2	1/4	1/4	3	1.35	3.10	1442 (0.75%)
1/4	1/2	1/4	3	1.29	3.12	1259 (0.65%)
1/4	1/4	1/2	3	1.36	3.06	1416 (0.74%)
1	0	0	3	1.32	2.72	3657 (1.9%)
0	0	1	3	1.16	2.72	2063 (1.1%)
0	1	0	–	1.06	1.33	17533 (9%)
1/3	1/3	1/3	2	1.53	2.20	4841 (2.5%)
1/2	1/4	1/4	2	1.31	2.34	5665 (3%)
1/4	1/2	1/4	2	1.23	1.72	7631 (4%)
1/4	1/4	1/2	2	1.28	2.29	4266 (2.2%)
1	0	0	2	1.3	2.11	11571 (6%)
0	0	1	2	1.13	2.13	4791 (2.5%)

For data classified as outliers, the RMS difference is larger than the values obtained for non-outlier data in all cases. This shows that the pixels detected in the AVHRR data are in fact different from the values in the MODIS and OSTIA SST, and are therefore very likely to be outliers. The smallest RMS difference is obtained when only the O_{prox} test is applied. This result is to be expected because this test bases its classification only on the proximity of each pixel to a cloud, regardless of its value. This test should always be applied as a complement to the other two.

The classification using an equal weight of 1 / 3 for each test and a threshold of 3 detects the most suspect data (the RMS difference is the largest of the table, for both comparisons), with a minimal data loss (0.7% of the complete data set). Other combinations result in more data being classified as outliers. The determination of the weights and threshold to be used is dependent on the data set and the degree to which one wants to remove outliers. Note that the outlier population detected for each combination of weights is different, which results in very different RMS errors.

An additional test was carried out to test the influence of the median window size in the detection of outliers. Fig. 3 shows the results of the median test for median window sizes of 2×2, 10×10, 20×20,

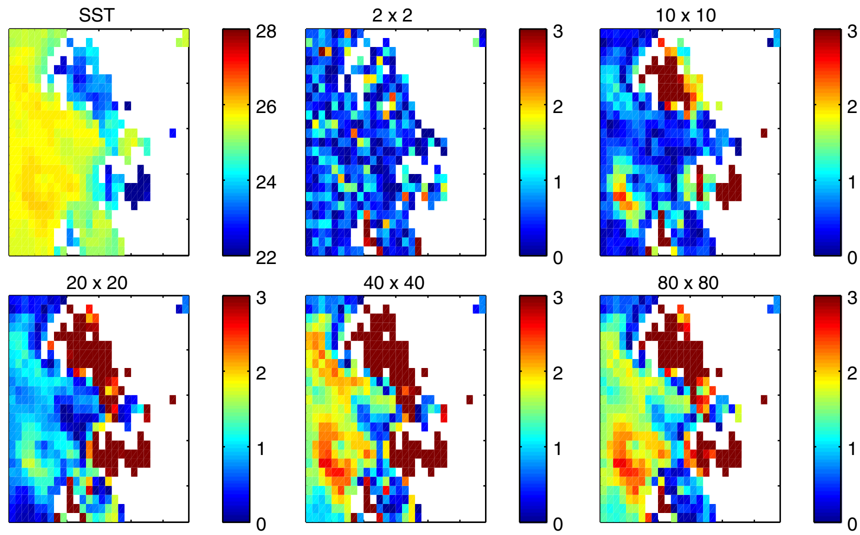


Fig. 3. SST detail (°C) and effect of window size in the median test. Windows of 2 × 2, 10 × 10, 20 × 20, 40 × 40 and 80 × 80 pixels were used.

40 × 40 and 80 × 80 pixels, together with the original SST to assess the impact of this parameter. It can be seen that a too small window size (*i.e.*, 2 × 2 pixels) does not give robust results, as very few of the suspect cold pixels in the SST are given high values with the median test. For large median window sizes (40 × 40 pixels and 80 × 80 pixels), some good-quality pixels obtain high scores and are therefore subject to being classified as outliers when they are not. Moderate median window sizes of 10 × 10 and 20 × 20 pixels give the best results, and this validates our choice of using a 20 × 20 pixel window size for the median test. Moreover, the larger the window size, the longer the computational time needed to the compute the median test; therefore moderate window sizes are preferred.

4.4. Comparison with the Sirjacobs et al. (2011) method

As mentioned in Section 3.2, the EOF test is based on the one presented in Sirjacobs et al. (2011), although some improvements have been made to make the approach more restrictive in the classification of outliers. In order to show the effect of using Eq. (7) instead of Eq. (1), we carried out an outlier classification test using these two approaches. The left panel of Fig. 4 shows again the SST detail in the western Mediterranean Sea as shown by a black square in Fig. 1. The centre panel of Fig. 4 shows the points classified using the approach of Sirjacobs et al. (2011), using a threshold of 3 as in previous tests. The right panel of Fig. 4 shows the result of the EOF test suggested in this study (*i.e.*, without applying the median or the proximity test), again classifying as outliers those pixels that exceed the threshold of 3. As can be seen, applying the method proposed in Sirjacobs et al. (2011) leads to the detection of fewer outliers than with the approach

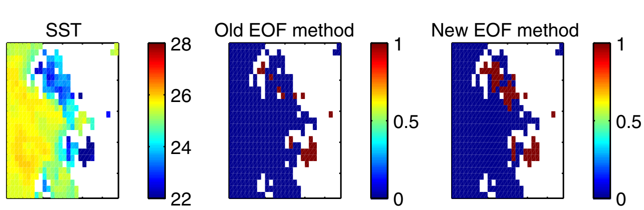


Fig. 4. Outliers detected using the EOF test as described in Eq. (1) (centre panel) and outliers detected using the new approach suggested in this study (Eq. 7, right panel). Note that in the new approach onlythe EOF test is applied, and not the proximity and median tests. The original SST (°C) is shown in the left panel for reference.

suggested in this study. Over the whole domain of study, on 4 July 2010, the approach by Sirjacobs et al. (2011) detected a total of 1931 outliers, and the EOF test suggested here detected 3657 outliers. This example shows that the new approach is better able to detect outliers.

5. Application to microwave SST data

The outlier detection methodology was also applied to the TMI data described in Section 2. Microwave sensors are able to "see" through clouds, but not through rain, so the proximity test used the presence of rain in a given pixel as the condition to flag the pixels in its vicinity. First, DINEOF was applied in order to compute the truncated EOF basis. Twelve EOFs were retained as optimal by DINEOF, which explain a total of 99.67% of the initial variance. An example of the missing data reconstruction obtained with DINEOF can be seen in Fig. 5. Note that most of the missing data in the TMI data set are due to the gaps between the satellite swaths, which are more pronounced near the equator. As these are not static, DINEOF can provide an estimation of the SST under these gaps as well as under other gaps caused mainly by the presence of rain. By using the spatio-temporal information contained in the truncated EOF basis, the reconstruction retains the meso-scale information observed in the original data, in the form of eddies and meanders.

An example of the detection of outliers is shown in Fig. 6 for the domain delimited by a black rectangle in Fig. 5. The original data are shown, together with the results of the individual outlier tests, and the final SST once the outliers have been removed (two thresholds for the outlier test, 2.5 and 3, are shown). The size of the window over which the median test is calculated is 20 × 20 pixels, as in the previous example. Note however that given the coarser resolution of the TMI SST product, the median test acts on a larger scale than for the AVHRR example. It can be seen that the technique is also capable of detecting outliers due to rain-contaminated pixels, as these are also flagged by both the median and the EOF tests. The proximity test also helps in the classification of the final outliers. The quality of the resulting SST once the outliers are removed is good for both thresholds used.

6. Application to chlorophyll-a concentration data

In order to test the outlier detection technique with a variable other than SST, the Sea-WiFS chlorophyll-a concentration data mentioned in Section 2 were used. These data are 8-day composites, and

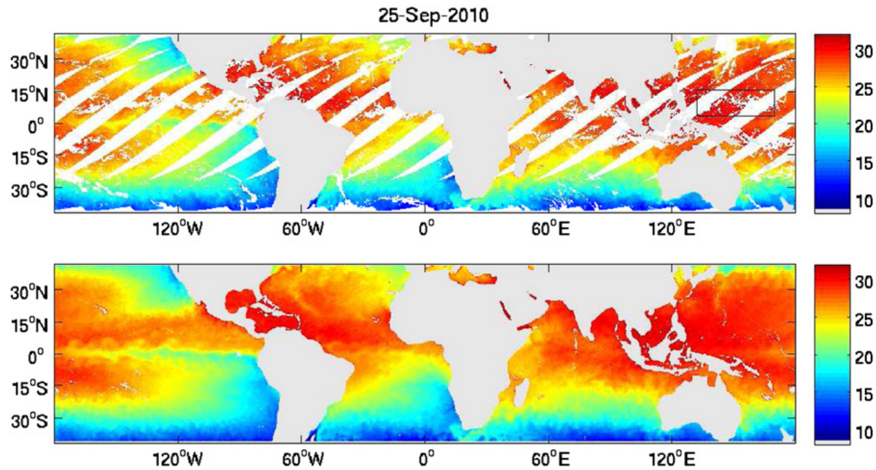


Fig. 5. TMI SST (°C) missing data reconstruction for 25 September 2010. Top panel: original data. Bottom panel: DINEOF reconstruction.

as such they contain less missing data than daily fields (about 22% of clouds and other missing data), and the possibility of outliers is diminished because of the averaging used in the compositing. The EOF basis calculated by DINEOF consists of the 9 most dominant EOFs, which explain 92.5% of the total variance. A logarithmic transformation of the chlorophyll-a data was performed before using DINEOF, because chlorophyll-a concentration data do not have a Gaussian distribution, as discussed in Section 3.1. This transformation was kept for the detection of outliers. The size of the window over which the median test is calculated is 20×20 pixels.

Although the presence of outliers is less evident in this data set, there are instances when some pixels have values very different from their surroundings, probably due to a persistent cloudy situation which decreases the amount of data available for compositing. Fig. 7 shows the chlorophyll-a concentration data for the period 9 to 16 January 2004, and, as can be appreciated in the image showing the southeastern part of the domain (top-right panel of Fig. 7),some data might be considered as outliers. In particular, some high values are observed within the high chlorophyll-a concentration plume located in the middle of the image, and along the coast. Given the composite nature of the data, a threshold of 3 might be too restrictive and thus classify as outliers some data that might be correct (see Fig. 7, bottom left panel). A less restrictive threshold of 6 is in this case appropriate to detect those suspect data mentioned above. This example shows that the outlier detection technique is also able to detect outliers that arise as a result of a compositing strategy.

7. Conclusions

We have presented a technique to detect outliers in satellite data sets, based on the combination of three tests that exploit the spatial coherence of the data. The proposed technique was tested using sea surface temperature (SST) data from an infrared sensor and a microwave sensor, as well as chlorophyll-a concentration data. These examples showed that the technique can be applied to variables with very different characteristics.

The combined use of the three proposed tests (EOF basis test, proximity test and local median test), is able to accurately detect outliers in the data set, as demonstrated by comparing the results with independent satellite SST products. Using only one or two of the three tests might give sub-optimal results. For example, when very few data are present in the window used for the local median test (because data are missing due to the presence of clouds, rain, etc.), non-optimal results might be obtained. Also, using the median test alone might

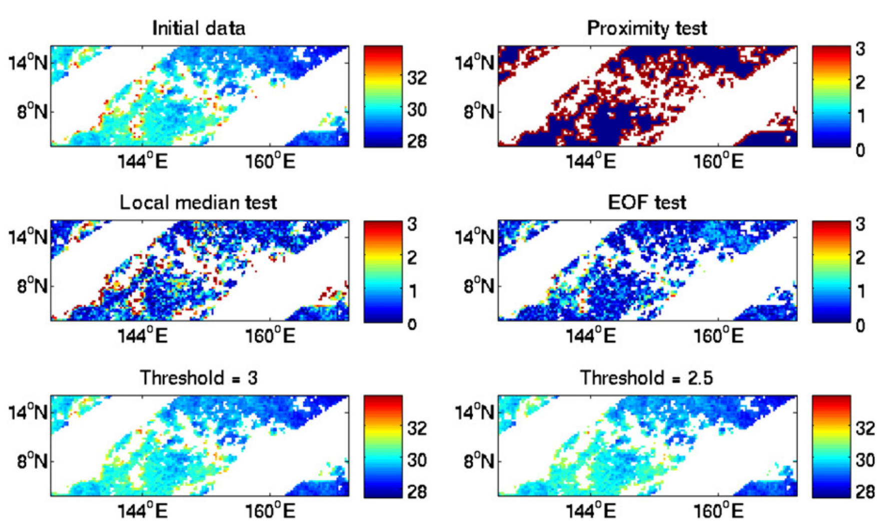


Fig. 6. SST detail (°C) and outlier tests in the domain shown by a black rectangle in Fig. 5. The original data are shown in the top left panel, and the data without outliers are shown in the two bottom panels, using two thresholds (2.5 and 3) when combining the individual outlier tests. The individual outlier classifications for each of the tests performed are also shown.

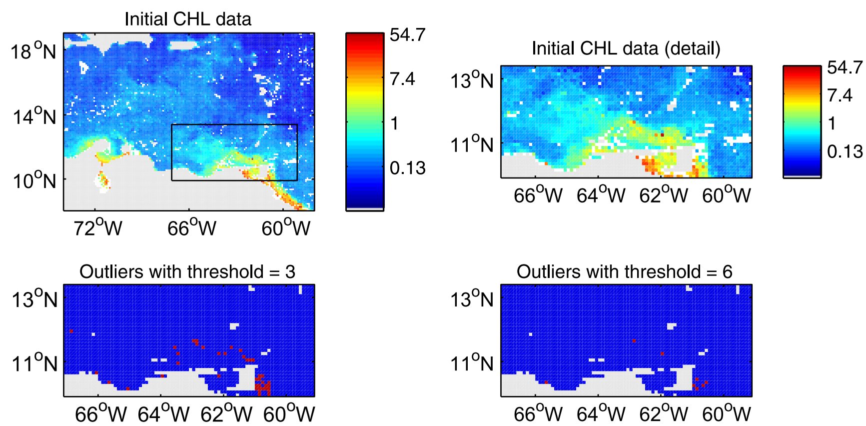


Fig. 7. Top left panel: 8-day composite of chlorophyll-a concentration (in mg/m^3) in the Caribbean Sea, corresponding to the period 9 to 16 January 2004. Top right panel: detail of the same data in the southeast of the domain. Bottom left: outliers detected using a threshold of 3, and (bottom right panel) a threshold of 6.

penalise too strongly data situated along a strong front. The EOF test alone might classify as an outlier an anomalous event happening only once in the analysed time series. The combination of the three tests can help mitigate these effects.

It has been shown that the methodology presented in this study, as it is based on a series of spatial coherence tests, only flags data that stand out as anomalous in relation to their surroundings. Therefore, large zones of anomalous data resulting from an atmospheric event, such as diurnal warming events, for example, are generally not penalised by this technique.

The weight given to each of the tests, as well as the threshold over which a pixel is classified as an outlier, can be adapted to each data set and to the future applications of the data. This gives the capability to adjust the sensitivity to different factors. As explained in this study, additional tests can be implemented, such as a land-proximity test, which may penalise pixels near to the coastline. Nevertheless, in the examples shown in this study some coastal outliers had already been detected by our approach.

The difficulty in detecting outliers in a data set lies in the fact that there is no unique definition of an outlier. This definition might vary depending on the specific application, the quality of the data, and even the expectations of the person using these data. The methodology presented in this study allows the adjustment of the degree to which one classifies a given pixel as an outlier by varying the weights of the different tests and the final threshold, and the approach can therefore be adapted to each specific case. For example, the size of the window used in the median test can be used to influence the size of the structures being classified as outliers. In general, single pixels or small zones (a few pixels) are penalized by the proposed tests, and not coherent structures. One could increase the size of the window in the median test so that larger structures are more penalized. However, if these structures are recurrent, like the warming event observed in the example given for the Mediterranean Sea, the EOF test will not penalize them, so they will not be classified as outliers.

The methodology can be applied on a global scale, although the computing resources needed would be high (particularly if working with high-resolution data sets). However, an alternative approach for global applications could be to calculate the EOF basis on a sub-basin scale, which would allow for a longer time-series for less computational cost. Such an EOF basis is capable of representing better the meso-scale processes of the sub-basins, and therefore the detection of outliers may be more robust using this approach.

The source code of DINEOF, the technique on which the EOF basis test is based, along with the outlier test detection technique described in this study, are freely available at http://modb.oce.ulg.ac.be/wiki.

Acknowledgements

This study was carried out within the context of the HiSea (SR/12/140) and Geocolour (SR/00/139) projects funded by the Belgian Science Policy (BELSPO) within the framework of the Research Program for Earth Observation “STEREO II”. MODIS Aqua l3m SSTdata and SeaWiFS chlorophyll-a concentration data were obtained from the Ocean Color website http://oceancolor.gsfc.nasa.gov/. North Atlantic Region AVHRR SST level 3 data were produced by the Ocean & Sea Ice Satellite Application Facility (O&SI SAF) and obtained through the Medspiration website (ftp://ftp.ifremer.fr/ifremer/medspiration/data/. OSTIA data were obtained through the MyOcean portal (http://operation.myocean.eu/). Microwave TMI data were downloaded from ftp://ftp.ssmi.com/. Two anonymous reviewers are acknowledged for their useful and constructive comments. The National Fund for Scientific Research, Belgium, is acknowledged for funding the positions of A. Alvera-Azcárate and A. Barth. This is a MARE publication.

References

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 surface temperature. *Ocean Modelling*, *9*, 325–346, doi: 10.1016/j.ocemod.2004.08.001.

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., Barth, A., & Alvera-Azcárate, A. (2006). DINEOF reconstruction of clouded images including error maps. Application to the sea surface temperature around Corsican Island. *Ocean Science*, *2*(2), 183–199.

Coakley, J., & Bretherton, F. (1982). Cloud cover from high-resolution scanner data: Detecting and allowing for partially filled fields of view. *Journal of Geophysical Research*, *87*(C7), 4917–4932.

Daley, R. (1991). *Atmospheric data analysis.* : Cambridge University Press 457 pp.

Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J. B. W., & Murray, M. J. (2002). Toward improved validation of satellite sea surface skin temperature measurements for climate research. *Journal of Climate*, *15*(4), 353–369.

Esaias, W. E., Abbott, M. R., Barton, I., Brown, O. B., Campbell, J. W., Carder, K. L., et al. (1998). An overview of MODIS capabilities for ocean science observations. *IEEE Transactions on Geoscience and Remote Sensing*, *36*(4), 1250–1265.

Kilpatrick, K. A., Podestá, G. P., & Evans, R. (2001). Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database. *Journal of Geophysical Research*, *106* (C5), 9179–9197.

Lazarus, S. M., Calvert, C. G., Splitt, M. E., Santos, P., Sharp, D. W., Blottman, P. F., et al. (2007). Real-time, high-resolution, spaceâtime analysis of sea surface temperatures from multiple platforms. *Monthly Weather Review*, *135*, 3158–3173.

May, D., Parmeter, M., Olszewski, D., & McKenzie, B. (1998). Operational processing of satellite sea surface temperature retrievals at the naval oceanographic office. *Bulletin of the American Meteorological Society*, *79*(3), 397–407.

Merchant, C. J., Filipiak, M. J., Le Borgne, P., Roquet, H., Autret, E., Piolle, J. F., et al. (2008). Diurnal warm-layer events in the western Mediterranean and European shelf seas. *Geophysical Research Letters*, *35*, L04601.

Merchant, C. J., Le Borgne, P., Marsouin, A., & Roquet, H. (2008). Optimal estimation of sea surface temperature from split-window observations. *Remote Sensing of Environment*, *112*, 2469–2484.
Okamoto, K., & Derber, J. C. (2006). Assimilation of SSM/I radiances in the NCEP global data assimilation system. *Monthly Weather Review*, *134*, 2612–2631.
Reimann, C., Filzmoser, P., Garrett, R., & Dutter, R. (2008). *Statistical data analysis explained: Applied environmental statistics with R*: Wiley 362 pp.
Sirjacobs, D., Alvera-Azcárate, A., Barth, A., Lacroix, G., Park, Y., Nechad, B., et al. (2011). Cloud filling of ocean color and sea surface temperature remote sensing products over the southern north sea by the data interpolating empirical orthogonal functions methodology. *Journal of Sea Research*, *65*(1), 114–130.
Stark, J. D., Donlon, C. J., Martin, M. J., & McCulloch, M. E. (2007). *Ostia: An operational, high resolution, real time, global sea surface temperature analysis system.* Oceans '07 IEEE*Marine challenges: coastline to deep sea., Aberdeen, Scotland (UK)*.
Wilks, D. S. (1995). *Statistical methods in the atmospheric sciences*: Academic Press.

Annex III

Methodology

1 DINEOF-OI

A technique to merge data sets has been developed during this project. The analysis is based on the formalism of optimal interpolation (OI) and the error covariance is expressed using a set of spatial EOFs obtained from DINEOF, instead of using an analytical expression, which is the most common way of describing the covariance matrix in OI. An EOF-based error covariance represents therefore more realistically the complex spatial variability of the data sets. This code has been written as a post-processing step for DINEOF: first a data set is analysed using DINEOF and the truncated EOF basis retained by it is used in the second step to merge the data set used in DINEOF with another data set.

The spatial EOFs are scaled:

$$\boldsymbol{U}_s = \frac{1}{\sqrt{n}} \boldsymbol{U} \boldsymbol{\Sigma}$$

where n is the number of time steps, **U** are the spatial EOFs and **Σ** are the singular values.

The analysis step can be described as follows:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{P}\,\mathbf{H}^T\,(\mathbf{H}\,\mathbf{P}\,\mathbf{H}^T + \mathbf{R})^{-1}\,(\mathbf{y}_o - \mathbf{H}\,\mathbf{x}_b)$$

where $\mathbf{x}_a$ is the analysis, $\mathbf{x}_b$ is the background field, **P** is the covariance matrix (based on the scaled truncated spatial EOF basis), **R** is the error covariance of the data, $\mathbf{y}_o$ are the data being analysed and **H** is an operator to extract the data at the observations location. The error variance of the observations can be specified as a single number or as pixel-by-pixel values. Single sensor error estimates (SSES) provided with the satellite data sets can be therefore used. The error covariance matrix **R** can be a diagonal matrix (meaning that the error covariance of the observations are de-correlated) or it can have non-null off-diagonal terms, to take into account the correlation of the error covariance of the observations. If no estimation of SSES is available, a first guess for the error variance can be estimated using the information rejected by the truncated EOF basis, considering this as a measure of noise in the dataset:

$$\mu^2 = \frac{1}{mn - M} \sum_{x_{ij}\, not\ missing} \left(x_{ij}^2 - x_{r,ij}^2 \right)$$

with m the number of spatial data, M the number of missing data, *x* the original data and x_r the data reconstructed with DINEOF. This is a first guess and can be adjusted. Because of the high spatial redundancy typical of satellite data, a strong spatial correlation within these data is normally found. The error variance of the satellite data has to take into account this redundancy:

$$\mu_e^2 = r\,\mu^2$$

where:

$$r = \frac{L^2}{\Delta x \Delta y}$$

where L is the correlation length of the observational error and Δx, Δy are the zonal and meridional resolution respectively.

2 Embedded DINEOF-OI

The technique described above was used either as a two-step process (EOFs calculated in DINEOF and the OI step preformed afterwards) or as an embedded process: within DINEOF, and at each time the EOF is updated in the iterative process, the merging with the geostationary data was realised. The merged estimation is then used in the initial dataset to perform a new EOF decomposition. The methodological approach is the same. Updating the initial data with the merged estimate can allow to obtain a more consistent merged dataset, as the merging is done iteratively and not only in one step.

3 Use of a correlated error covariance matrix in DINEOF-OI

To derive an “optimal” analysis it is necessary to know the error covariance of the background field (**P**) and the error covariance of the observations (**R**). Sophisticated methods are used to estimated the error covariance of the background field, but often simple diagonal form is used for **R**. Ignoring the correlation of observations results in giving too much weight to the observations. Several approaches avoiding the issue of correlated observations errors have been used, such as sub-sampling of the data (not all the data is used), binning i.e. degrading the spatial distribution of the data (some spatial information is lost) or inflation of **R** (pretend that the data are less precise because there is some redundancy).

The theoretical framework for dealing with spatially corrected errors is however well understood. Given the background estimate ($\mathbf{x}^f$), the observations ($\mathbf{y}^o$) and their relationship as expressed in the observation operator (**H**)

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{P}\mathbf{H}\left(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right)$$

The error covariance (**P**) is expressed in terms of EOFs (columns of the matrix **S**):

$$\mathbf{P} = \mathbf{S}\mathbf{S}^T$$

One can avoid the formulation of the error covariance by performing the following eigenvector decomposition:

$$(\mathbf{HS})^T \mathbf{R}^{-1} (\mathbf{HS})^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$$

Using the Sherman–Morrison–Woodbury formula, the analyses can be efficiently computed by

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{S}\,\mathbf{U}\,(\mathbf{I}+\mathbf{\Lambda})^{-1}\,\mathbf{U}^T\,(\mathbf{H}\mathbf{S})^T\,\mathbf{R}^{-1}\,(\mathbf{y}^o - \mathbf{H}\,\mathbf{x}^f)$$

Even for a relatively small domain such as the Western Mediterranean Sea and relatively low resolution sensor, the size of the matrix **R** can be quite large and the computation of its inverse can be challenging. For the present case, the matrix **R** has the size approximately 10 000 x 10 000 for Seviri data.

The error covariance of the observation **R** is decomposed in variance (**D**) and correlation (**C**):

$$\mathbf{R} = \mathbf{D}\,\mathbf{C}\,\mathbf{D}$$

The single sensor error statistics is used to specify the error variance of the SST. Error is assumed to come from a spatially correlated and non-spatially correlated component:

$$\mathbf{C} = (1-\alpha)\,\mathbf{I} + \alpha\,\mathbf{C}'$$

The spatially correlated error has the form:

$$\mathbf{C}_{ij} = \exp(-|\mathbf{r}_i - \mathbf{r}_j|^2 / L^2)$$

where $|\mathbf{r}_i - \mathbf{r}_j|$ is the distance between points *i* and *j.* The spatial structure of the observation error covariance is thus parameterized by the fraction of correlated error (α) and correlation length (*L*).

Annex IV

Application of DINEOF-OI to various SST datasets

1 Application of the merging technique to satellite + in situ data sets in the western Mediterranean Sea

The in situ and satellite data described in Appendix I were used for a first test of the merging technique. Only night-time satellite data were used as following the satellite - in situ comparison, this data set was nearer to in situ observations.

One year of AVHRR data was used (1999), which contained 64 % of missing data due mainly to the presence of clouds in the atmosphere. A first DINEOF analysis was realised on the SST anomalies (the seasonal cycle removed is presented in figure 3), which found by cross-validation 14 EOF modes as the optimal set for the reconstruction of missing data. These EOFs explained 95.62% of the total variance and a cross-validation error of 0.4 °C was obtained for the reconstruction.

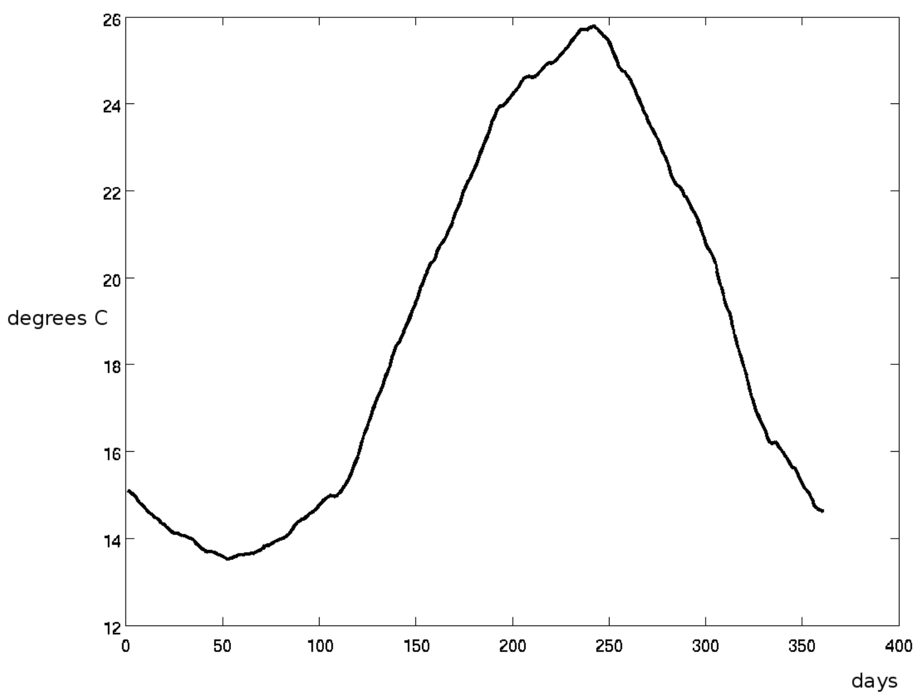


Figure 1: seasonal cycle of SST removed prior to the DINEOF analysis.

The first three temporal and spatial EOFs obtained from the DINEOF analysis are shown in figures 4, 5 and 6. These EOFs represent the variability of the western Mediterranean SST respect to the seasonal cycle of figure 3, and together they explain about 80% of the total variability. There are clear zones of high variability, like the Alboran Sea and the Gulf of Lions. The temporal modes are highly variable and it is difficult to infer a temporal scale in these EOFs, although one can see a 4 to 6 month variability on the first and second modes.

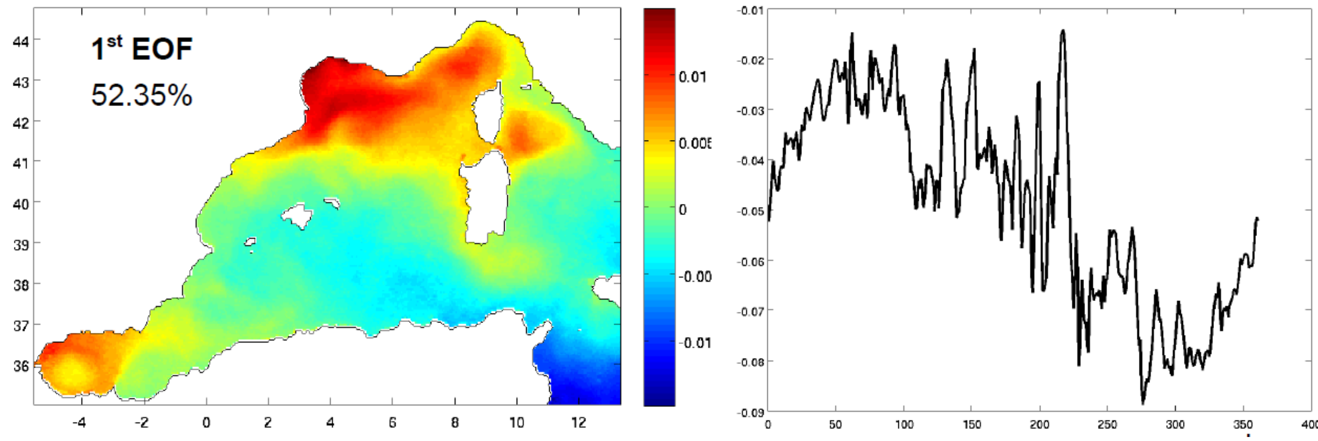


Figure 2: First spatial (left) and temporal (right) EOF mode obtained with DINEOF for the AVHRR data set. The percentage of variability explained by this mode is shown in the left panel. The units of the temporal mode x-axis are days since 1st January 1999.

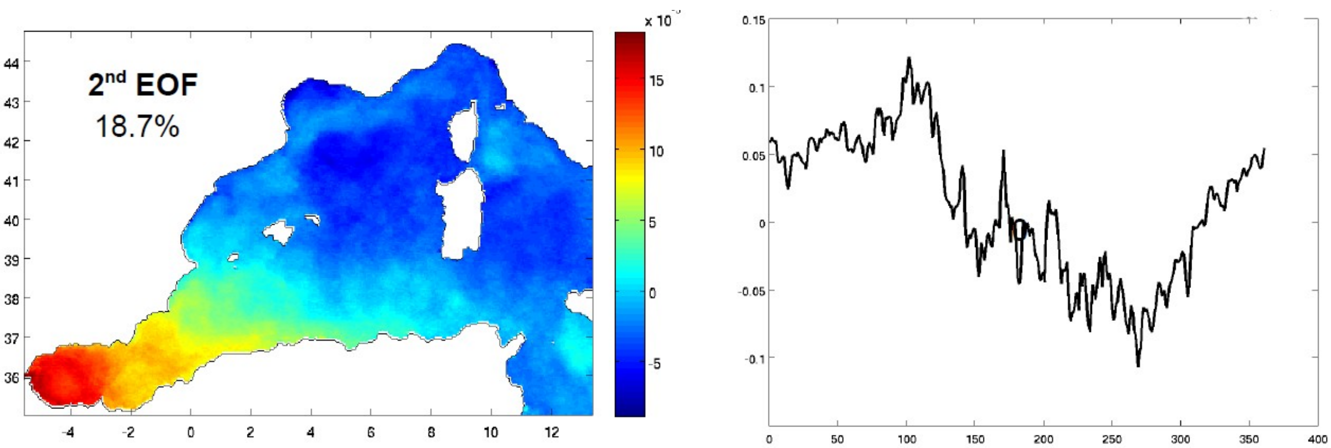


Figure 3: Second spatial (left) and temporal (right) EOF mode obtained with DINEOF for the AVHRR data set. The percentage of variability explained by this mode is shown in the left panel. The units of the temporal mode x-axis are days since 1st January 1999.

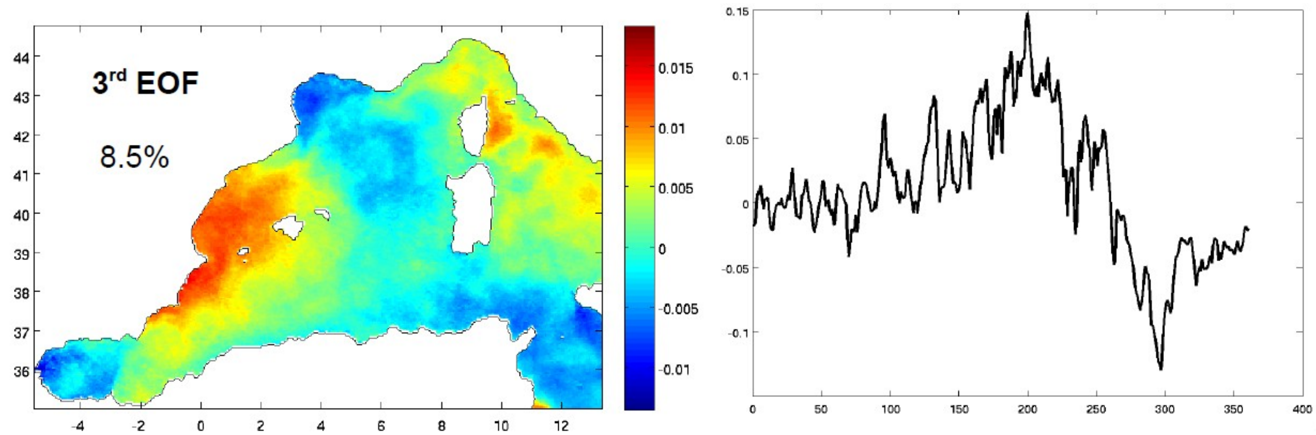


Figure 4: Third spatial (left) and temporal (right) EOF mode obtained with DINEOF for the AVHRR data set. The percentage of variability explained by this mode is shown in the left panel. The units of the temporal mode x-axis are days since 1st January 1999.

The covariance function used in the OI step is based on these EOFs (the three shown in figures 4 to 6 and 11 additional EOFs not shown here). Figures 7 and 8 show two examples of the spatial covariance of a given point (in particular, one in the Balearic Sea and another in the Gulf of Lions) with respect to the rest of sea points in the western Mediterranean Sea. The variance of the first EOF was reduced to produce a more localized response.

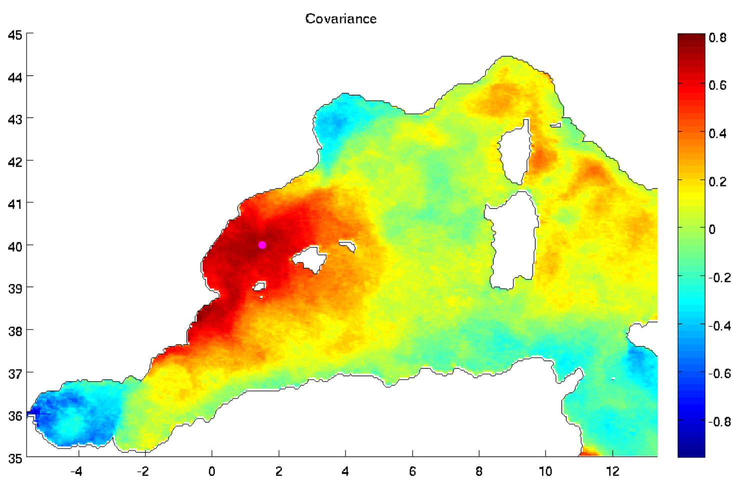


Figure 5: covariance between a point in the Balearic Sea (magenta dot) and the rest of the domain.

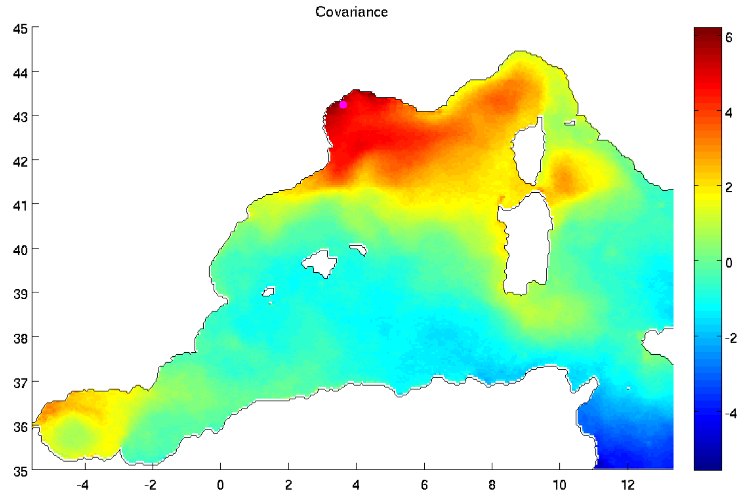


Figure 6: covariance between a point in the Gulf of Lions (magenta dot) and the rest of the domain.

We must stress the fact that the covariance explained with the EOF basis is non-parametric, i.e. it is based on the data alone. This explains the richness of the structures that can be seen in the covariance fields. There are however also some shortcomings to the use of a truncated EOF basis: spurious long-distance correlations are observed, specially in figure 7, and this must be probably due to the limited number of EOFs used in the analysis (we recall we used 14 EOFs).

An OI step is therefore done using the satellite data and the in situ data, with the EOF basis as covariance function. Figure 9 shows an example of the initial AVHRR data and the result after merging them with available in situ data for the 16 October 1999. The error for the different in situ data are those given in table 1.

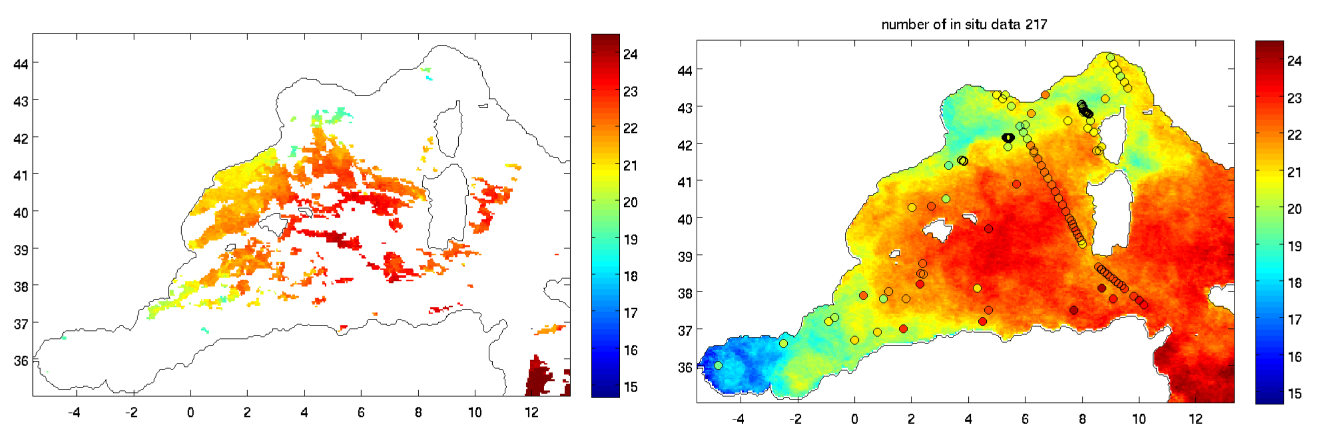


Figure 7: (left) Initial SST from AVHRR; (right) DINEOF-OI merged SST using satellite and in situ

In order to validate the results obtained with the DINEOF-OI approach, 10% of the in situ data, at random locations, was taken aside for cross-validation. It was seen that using the DINEOF-OI approach improved the results over a DINEOF step alone (using only satellite data). The errors obtained through the cross-validation are summarised in table 2.

Table 1: RMS error between the analysis field obtained by DINEOF or DINEOF-OI and all in situ data and a cross validation subset of in situ data. The highlighted fields show the errors obtained with the DINEOF approach (only satellite data) and the DINEOF-OI approach taking aside the cross-validation set.

	All in situ data (°C)	10% of cross-validation data (°C)
DINEOF	1.12	1.07
DINEOF-OI with all in situ data	1.08	1.04
DINEOF-OI without CV data	1.08	1.06

2 Application of DINEOF-OI to polar-orbiting + geostationary satellite SST data: two-step approach

The next step was to apply DINEOF-OI to a combination of polar-orbiting and geostationary data, with the aim of calculating a merged data set with the spatial resolution of the polar-orbiting satellite data and the temporal resolution of the geostationary satellite data.

For the polar-orbiting satellite, the AVHRR NAR18 data were used, with a spatial resolution of~2km and two daily passages (at 2:00 and 12:00). For the geostationary satellite, SEVIRI data were used, with a spatial resolution of ~10km spatial resolution and 8 daily fields (3-hourly). Different periods were used, and here the results for 1 January to 31 Mars 2009 are presented. In order to correct for the bias between the AVHRR NAR18 data and the SEVIRI data, 0.32ºC were removed from the latter.

To compare the results, OSTIA SST data were downloaded for the same period. As it is shown in figures 8 and 9, the correspondence between the OSTIA data and the DINEOF-OI merged data is very good, with the OSTIA data smoother than the DINEOF-OI data. In figure 9, for example, the warm signal of the Northern Current along the French coast is better represented in the DINEOF-OI field.

The root mean square error between OSTIA and DINEOF-OI is 0.39ºC, with a bias of -0.09ºC (DINEOF-OI colder than OSTIA). For the comparison, DINEOF-OI data at 7:30 were used, as OSTIA represents foundation temperature free of diurnal warming. DINEOF-OI data are however, 3-hourly, so the complete variation of the western Mediterranean SST diurnal cycle can be assessed. As an example, the complete series of SST for the 20 January 2009 is shown in Figure 10. A smooth transition between images can be observed, with the warmest part in the south of the domain becoming cooler through the day.

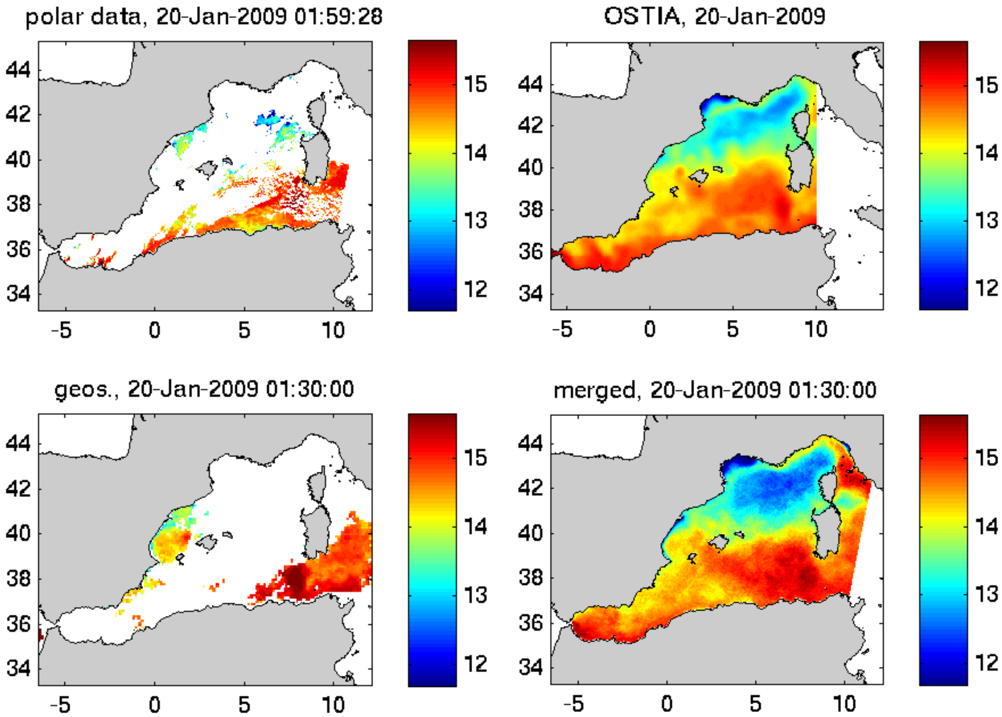


Figure 8: Example of merged SST on 20 January 2009. Top left: initial polar data. Bottom left: initial geostationary data. Top right: OSTIA data; Bottom left: DINEOF-OI data (two step approach)

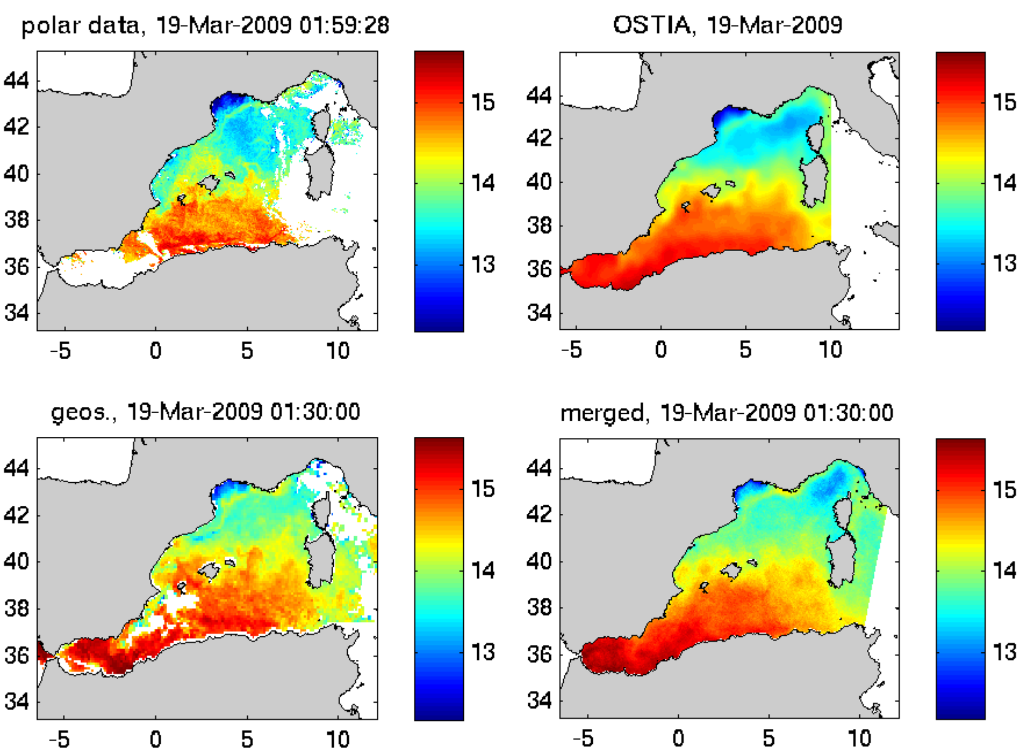


Figure 9: Example of merged SST on 19 Mars 2009. Top left: initial polar data. Bottom left: initial geostationary data. Top right: OSTIA data; Bottom left: DINEOF-OI data (two-step approach)

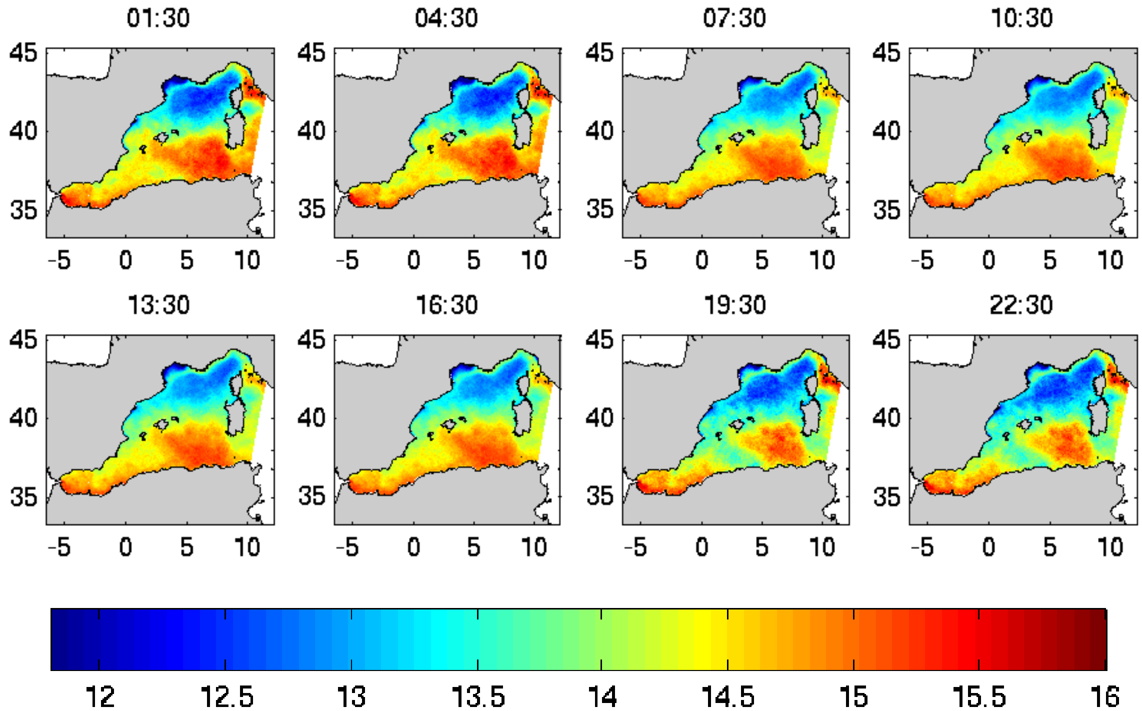


Figure 10: DINEOF-OI 3-hourly reconstruction of western Mediterranean SST for 20 January 2009.

The merged data were also compared to independent in situ data. The Coriolis database (http://www.coriolis.eu.org/) was used. All available data for 2009 were downloaded, consisting of drifting buoys, Argo profilers, moorings, XBTs, CTDs and gliders). We kept for the comparison the shallowest measure in case of profiles, and in any case no observations below 5 m depth were retained. The RMS between the merged DINEOF-OI data and the in situ data is 1.3ºC with a bias of -0.13ºC.

3 Application of DINEOF-OI to polar-orbiting + geostationary satellite SST data: embedded approach

The same analyses preformed in section 2 were realised using the embedded DINEOF-OI approach, in which the merging step takes place each time the EOF basis is updated within DINEOF. The results are very similar, although more small-scale variability can be seen in the embedded approach. To illustrate this we show in figure 11 the DINEOF-OI embedded estimation for 20 January 2009, to be compared with figure 8. The large scale features are similar, and the difference is as said, in the smaller features that are more abundant in the embedded case.

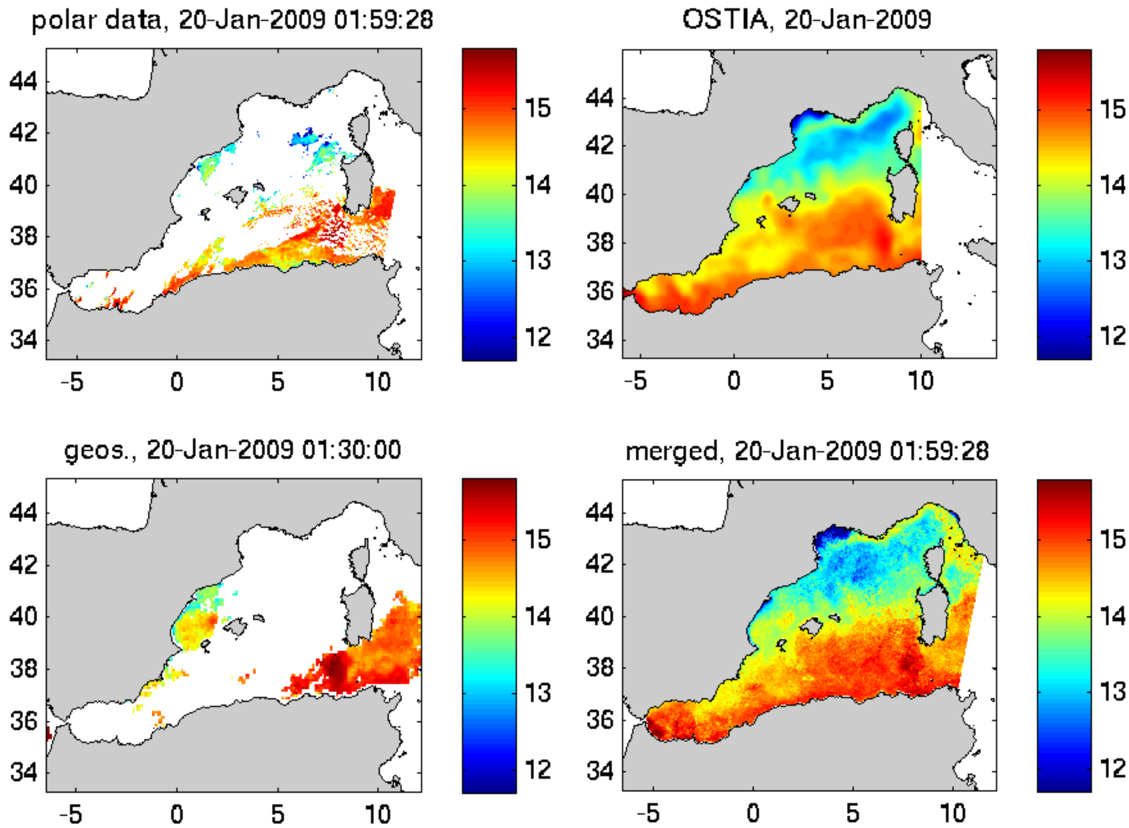


Figure 11: Example of merged SST on 20 January 2009, using the embedded DINEOF-OI. Top left: initial polar data. Bottom left: initial geostationary data. Top right: OSTIA data; Bottom left: DINEOF-OI data (embedded approach)

Comparisons with OSTIA and with the Coriolis in situ data were also performed, and all values are summarised in table 2. As can be seen, the two-step approach has a lower RMS when comparing to OSTIA and the embedded approach has a lower RMS when comparing with the in situ data.

Table 2. Error assessment of the merged dataset, using DINEOF-OI in two steps and the embedded DINEOF-OI.

	OSTIA		Coriolis in situ data	
	RMS (ºC)	Bias (ºC)	RMS (ºC)	Bias (ºC)
DINEOF-OI (two steps)	0.39	-0.093	1.3	-0.13
DINEOF-OI (embedded)	0.6	0.21	1.12	0.55

A difference between the two approaches is, as of now, the CPU time needed to perform the calculations. The embedded approach performs the analysis step iteratively, therefore the total CPU time is more elevated than the two-step approach, where the update is only done once.

4 Using a spatially correlated error covariance matrix.

Using the same combination of polar-orbiting and geostationary datasets of the previous section, the use of a spatially correlated error covariance matrix was tested. The EOFs were computed based on observations from NAR 18 from 1 January 2009 to 30 June 2009. Data used for the reconstruction are from 17 May 2009, which has the lowest cloud cover in data set (figure 12).

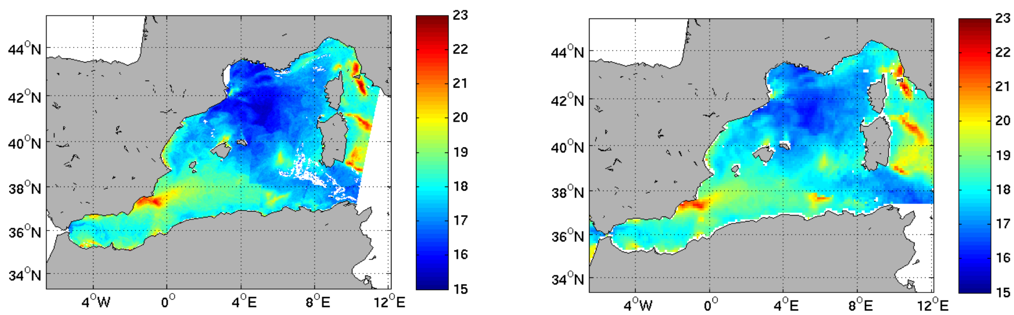


Figure 12: SST from NAT 18 and SERVIRI for 17 May 2009.

1% (or 40 %) of observations from AVHRR NAR-18 were removed from the dataset for cross-validation. The objective of the reconstruction is to determine if assuming spatial correlated observations for SEVIRI is beneficial by varying the fraction of correlated error (α) and correlation length L.

The benefit of taking explicitly the spatial correlation of the observation error into account is tested using cross-validation: 1% of the NAR 18 data (chosen randomly) are taken out of the data set and not used in the analysis. The analysis is first performed with a diagonal **R** matrix. The RMS error compared to the cross-validation data without taking the correlation into account is 0.587 °C.

A series of tests were performed with different correlation lengths *L* and different values for the fraction of correlated error *a*. The cross-validation error is given in table 3. The minimal cross-validation error was found for a correlation length of 20 km and a fraction of correlated error of 0.6. The cross-validation error is reduced by a relatively modest 0.01 °C. Visually the results (figure 13) are quite similar. The difference of the reconstruction reveals that the differences reach about 0.3° in the Northern Tyrrhenian Sea and in the Western Provençal Basin.

Table 3: Cross-validation error taking aside 1% of the data for varying values of the correlation length L (in km) and fraction of fraction of correlated error a.

		correlation length			
		5	**10**	**20**	**30**
fraction of correlated error	**0.2**	0.584	0.582	0.581	0.581
	0.4	0.584	0.579	0.585	0.695
	0.6	0.580	0.583	**0.576**	0.580
	0.8	0.674	0.584	0.642	0.655

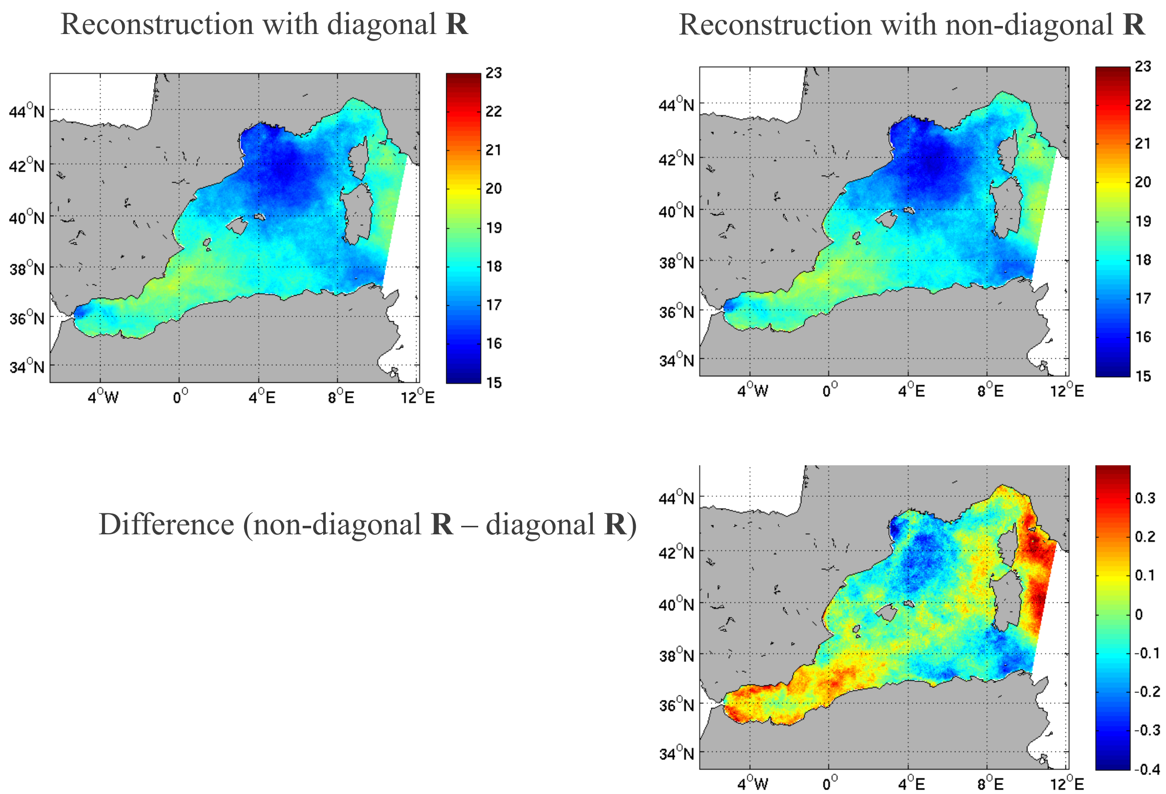


Figure 13: Reconstructed SST using NAR 18 and SERVIRI of the 11 May 2009 using diagonal observation error covariance R, a non-diagonal R and the difference.

The percentage of cross-validation points was increased to 40% to study how the reconstruction error behaves under a more typical cloud cover. In fact, with a low percentage of cross-validation points, the removed data are isolated pixels whose information can be relatively easily recovered using the neighbouring pixels. Using 40% of cross-validation points, the RMS error without taking the correlation into account is 0.603°C. It is interesting to note that the optimal cross-validation error is obtained for the same values of the correlation length and fraction of correlated error as before. The reduction of the RMS error, 0.02°C, is larger than the case the 1% cross-validation points.

Table 4: Cross-validation error taking aside 40% of the data for varying values of the correlation length L (in km) and fraction of correlated error a.

		correlation length			
		5	**10**	**20**	**30**
fraction of correlated error	**0.2**	0.598	0.594	0.592	0.592
	0.4	0.599	0.589	0.597	1.403
	0.6	0.589	0.591	**0.581**	0.588
	0.8	0.688	0.606	1.142	0.676

Finally, we performed a series of analyses using larger correlation lengths.

Table 4: Cross-validation error taking aside 40% of the data for varying values of the correlation length L (in km) and fraction of correlated error, for longer correlation lengths.

		Correlation length (km)						
		50	**100**	**200**	**300**	**500**	**700**	**1000**
Fraction of correlated error	**0.2**	1.558	1.1459	0.5986	**0.5827**	0.6014	**0.5812**	0.6154
	0.4	0.6128	0.6065	0.6419	1.0649	0.7915	0.599	0.5925
	0.6	1.1249	0.7125	0.9742	0.8011	0.6967	0.6492	0.6432
	0.8	0.648	0.614	3.2523	0.8927	3.9199	0.9894	3.5744

While the results obtained with a correlation length of 20 km are still minimizing the cross-validation error, we found two additional length scales that are very near the value obtained for 20 km: 300 km and 700 km. This may reflect the multiscale nature of the SST errors, and suggests that errors are correlated at different scales, which would depend on the source of the error (e.g. the longest lengths scales can be due to atmospheric processes).

Annex V

Application of DINEOF-OI to Total Suspended Matter data

In order to test if DINEOF-OI can also work with other variables and in other domains, ocean colour data over the North Sea have been used. A polar-orbiting dataset (from MODIS), with a spatial resolution of 2 km, and a geostationary dataset)SEVIRI) with a spatial resolution of about 6 km and a temporal resolution of 15 min. The approach used is the same as used throughout the project: an EOF basis is calculated for the higher spatial resolution data and used as covariance matrix to merge the two data sources.

We worked directly with marine reflectance data, for the period June-July 2008. Sixteen EOFs were retained by DINEOF. An error variance of 1 was used for both MODIS and SEVIRI data.

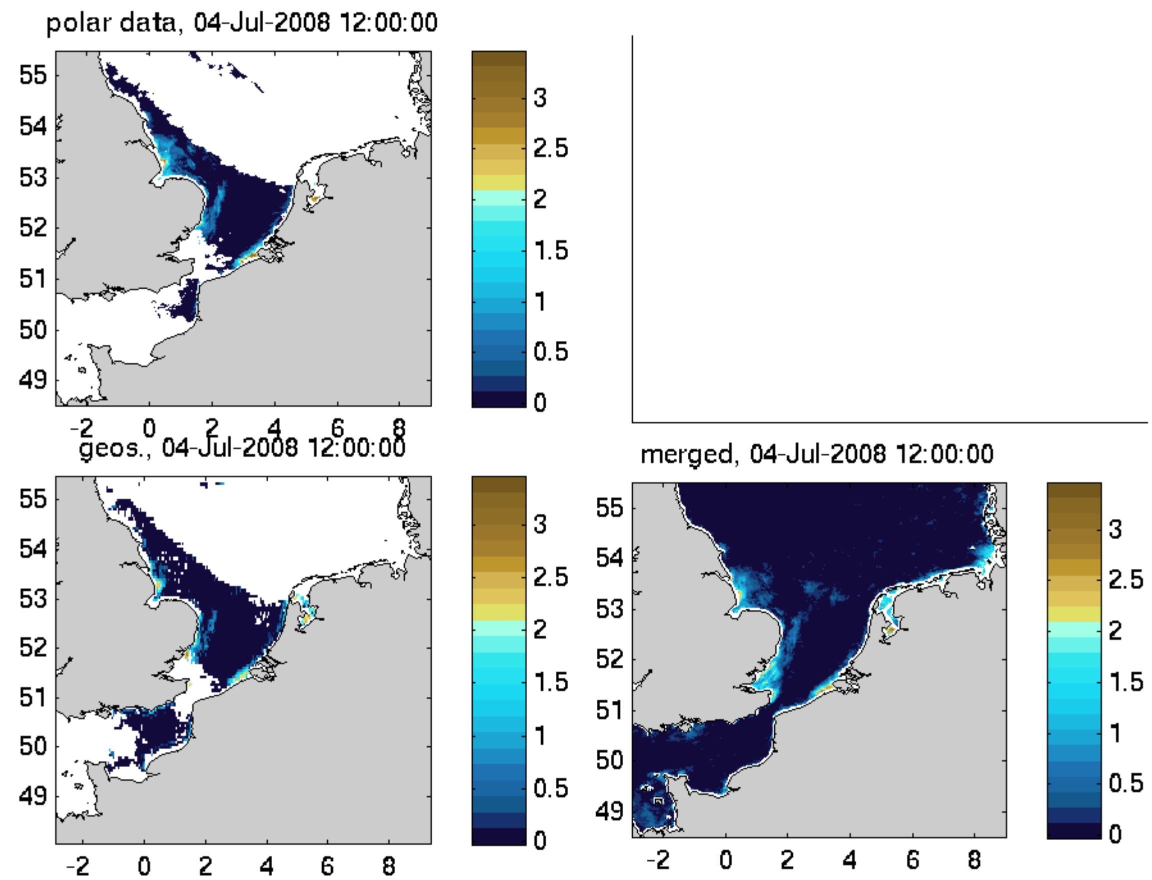


Figure 1: Top panel: initial MODIS marine reflectance data; bottom left: initial SEVIRI marine reflectance data; bottom right: reconstruction by DINEOF-OI. Date: 04 July 2008. Data have been transformed to Total Suspended Matter (g/m3) and are displayed in logarithmic scale.

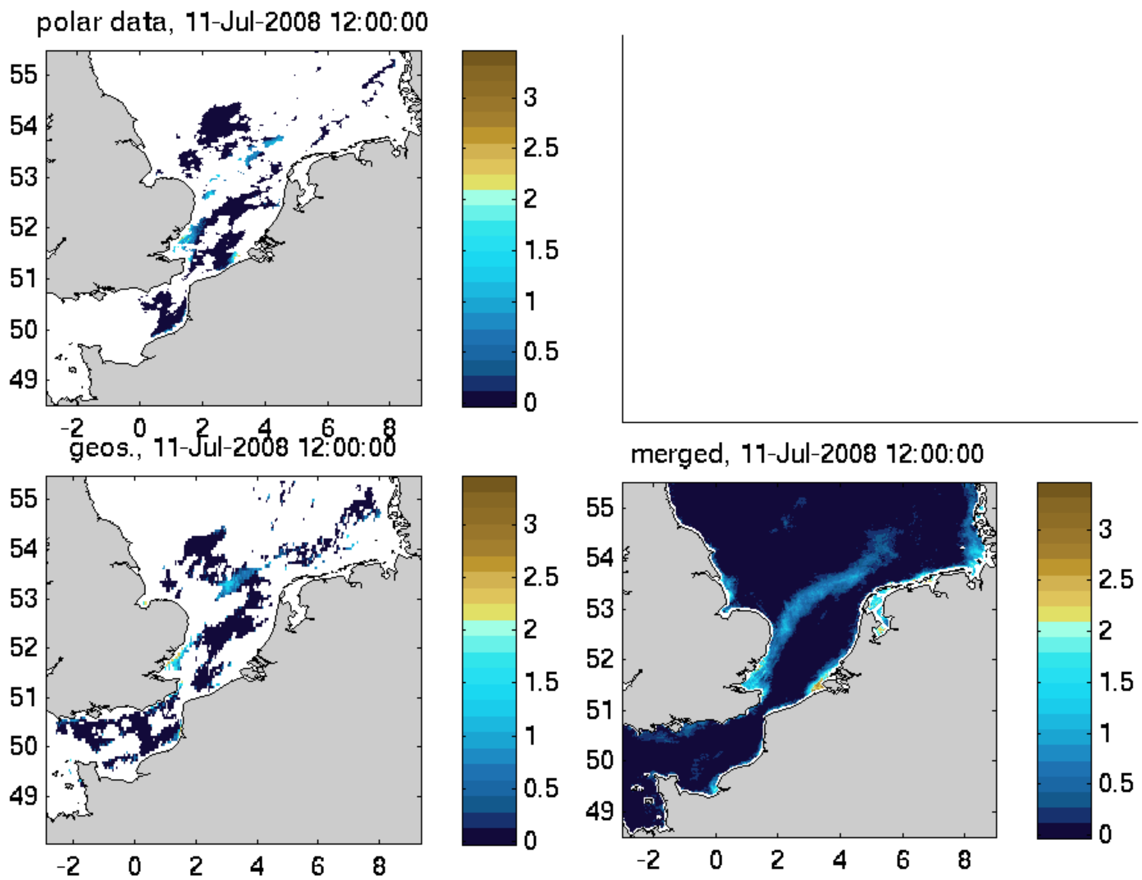


Figure 2: Top panel: initial MODIS marine reflectance data; bottom left: initial SEVIRI marine reflectance data; bottom right: reconstruction by DINEOF-OI. Date: 11 July 2008. Data have been transformed to Total Suspended Matter (g/m3) and are displayed in logarithmic scale.

The merged dataset has the temporal resolution of the SEVIRI dataset (*i.e.* 15 min). The quality of the results was assessed using in situ data from the West Gabbard and Warp measuring buoys, provided by Cefas (see their position in figure 3). This comparison showed that the technique is able to provide accurate reconstructions (see table 1). In situ data are in Formazin Turbidity Units (FTU). In order to compare to the in situ data, satellite reflectance was transformed to Formazin Nephelometric Units (FNU).

We expect that using a more realistic error variance will help to further improve the quality of the merged dataset.

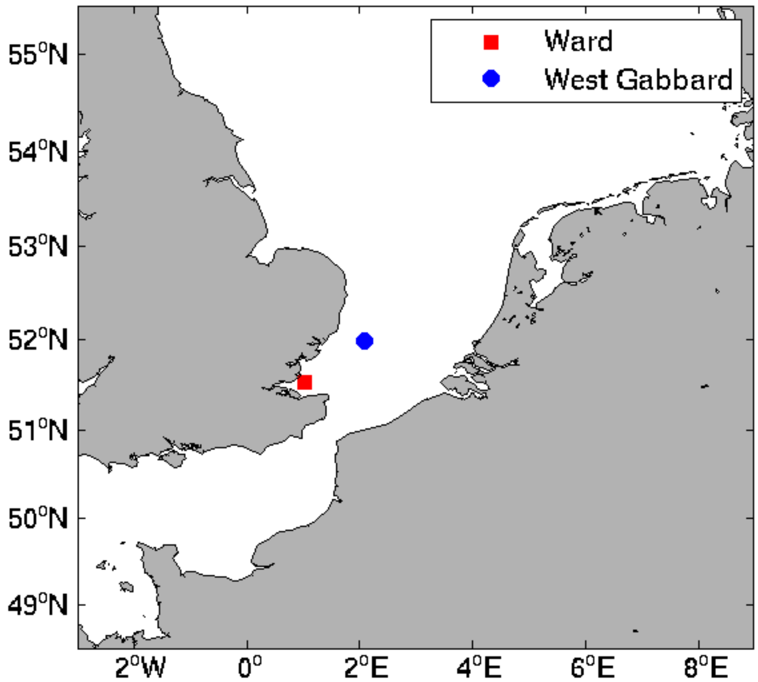


Figure 3: position of the in situ buoys used for comparison with our data.

Table 1: error assessment of the initial datasets and the merged product provided by DINEOF-OI. Units are FNU/FTU.

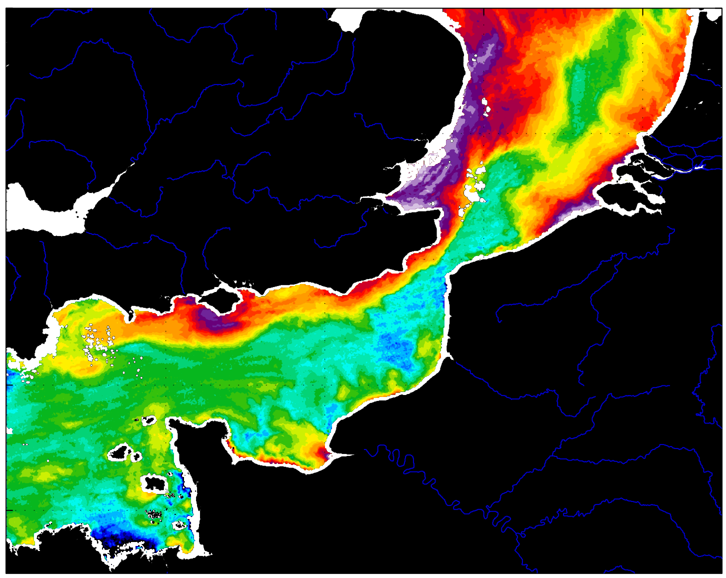
	Ward		West Gabbard	
	RMS	Bias	RMS	Bias
MODIS	6.99	-0.97	2.19	0.45
SEVIRI	2.39	1.1	0.69	0.08
DINEOF-OI	5.89	-2.12	0.88	-0.49

Annex VI

Forecast of Total Suspended Matter in the southern North Sea using DINEOF

Forecast of Total Suspended Matter in the southern North Sea using DINEOF

C. Troupin[1] , A. Alvera-Azcárate[1], G. Lacroix[2]



[1] Université de Liège, Department of Astrophysics, Geophysics and Oceanography (AGO), GeoHydrodynamics and Environment Research (GHER), Allée du 6-Août, 17, Sart-Tilman B5a, 4000 Liège, Belgium
[2] Remote Sensing and Ecosystem Modelling (REMSEM) team,
Management Unit of the North Sea Mathematical Models (MUMM)

✉ctroupin@ulg.ac.be

Contents

1 Numerical model **3**

1.1 Spatial domain 3

1.2 Time period 3

1.3 Model variables 5

1.4 Temporal variations 5

1.5 River run-offs 7

2 TSM satellite data **8**

2.1 Temporal variation 8

2.2 Seasonal cycle 8

2.3 Cloud coverage 11

3 Model variable analysis **12**

3.1 Correlations 12

3.2 Fourier analysis 12

4 DINEOF analysis of TSM **15**

4.1 Validation points 15

4.2 Transformation on data 15

4.3 Number and importance of the modes 16

4.4 Reconstructed fields 17

4.5 Interpretation of the modes 18

5 DINEOF forecast **21**

5.1 Univariate reconstructions 21

5.1.1 Expected error and number of modes 22

5.1.2 Correlation and RMS 24

5.1.3 Results 25

5.2 Multivariate reconstruction 27

5.3 TSM and Bottom stress 27

5.3.1 Expected error and number of modes 27
5.3.2 Correlation and RMS 28
5.3.3 Results 28
5.4 TSM and sea surface elevation 30
5.4.1 Correlation and RMS 30
5.4.2 Results 30
5.5 Summary 30

6 Conclusions **32**

Introduction

The objective of this work is to provide forecast of a variable measured by satellite using a time series of images of this variable combined with the outputs of a numerical model. In the present work the tested variable will be the total suspended matter (TSM) for which forecast will be constructed in the southern North Sea.

DINEOF:	remote-sensing measurements	+	model variables
	$\downarrow$		$\downarrow$
	no explicit parametrization		forecasts

In Section 1, the numerical model are presented. Then we describe the TSM data (Section 2). Section 4 is focused on the DINEOF reconstructions using only TSM, while in Section 5 we combine TSM data and model outputs to generate TSM forecasts.

1 Numerical model

The model outputs are obtained from the Remote Sensing and Ecosystem Modelling (REMSEM) team, Management Unit of the North Sea Mathematical Models (MUMM) (Lacroix *et al.*, 2004).

1.1 Spatial domain

The domain extends from 48°30'N to 57°N and from 4°W to 9°E. The zonal resolution is 1/12° and the meridional resolution is 1/24°. The grid is made up of 157×205 grid cells.

1.2 Time period

Two periods are considered:

January 2003: outputs are available every 30 minutes, yielding 1489 time steps. This period is used to

- compute the correlations between the different model variables and discard some of them.
- determine the temporal resolution necessary to capture the variability of the TSM.

1 January to 31 December 2003 , with a smaller set of variables and a longer time step (Section 3).

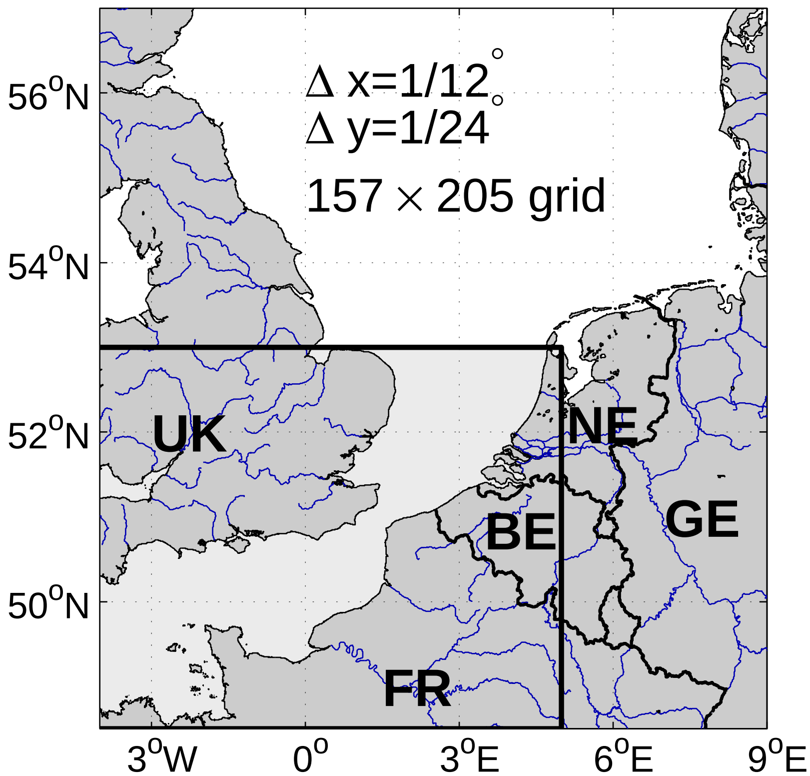


Figure 1: Spatial domain where the outputs of the numerical model are available. The shaded rectangle delimits the area where TSM is available. Main rivers are indicated in blue.

1.3 Model variables

For a first analysis, ten variables were extracted, as described in Tab. 1. Vertical velocity was not chosen because the North Sea is not stratified in the region of interest.

Table 1: Variables extracted from the model. The abbreviations (first column) will be used in the following sections.

Abbreviations	Name	Units	
WINDU	U-Wind velocity	[m/s]	
WINDV	V-Wind velocity	[m/s]	
UM2ATC	Mean depth U-velocity	[m/s]	
VM2ATC	Mean depth V-velocity	[m/s]	
U2ATCSurf	Surface U-velocity	[m/s]	(σ-layer 20/20)
V2ATCSurf	Surface V-velocity	[m/s]	(σ-layer 20/20)
U2ATCBot	Bottom U-velocity	[m/s]	(σ-layer 1/20)
V2ATCBot	Bottom V-velocity	[m/s]	(σ-layer 1/20)
ZETA2	Surface elevation	[m]	($\pm$ mean depth)
BSTOT	Bottom stress	[m^2/s^2]	

1.4 Temporal variations

Each variable is spatially averaged in order to provide a time series (Fig. 2, 1489 values per variable). All the ocean variables display tidal oscillations. The zonal wind is westward most of the time, while the meridional component often switches from north to south. At surface, the velocity roughly varies between -0.3 and 0.3 m/s. At the bottom, the maximal values are lower than 0.15 m/s. The bottom stress exhibits two periods of larger amplitudes, probably related to the wind intensity.

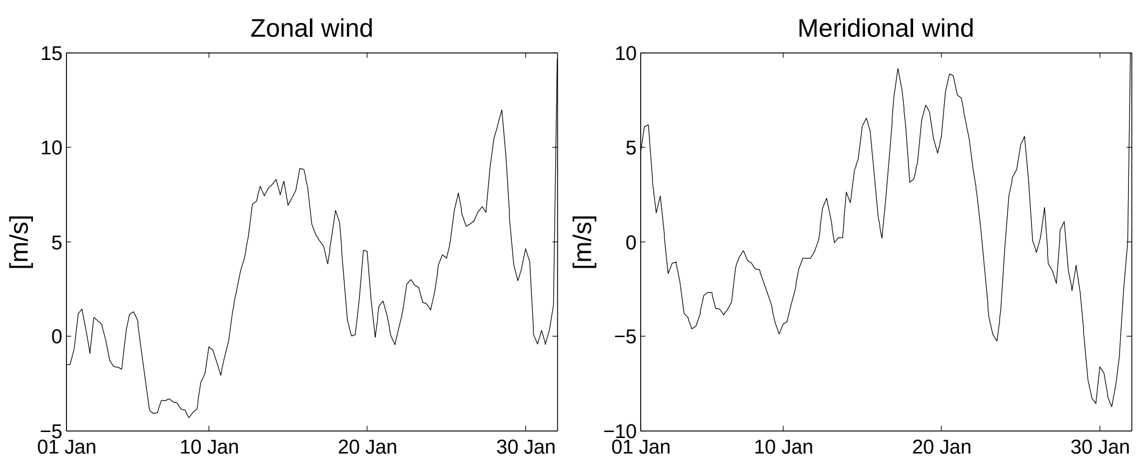


Figure 2: Time evolution of the variables averaged on the domain.

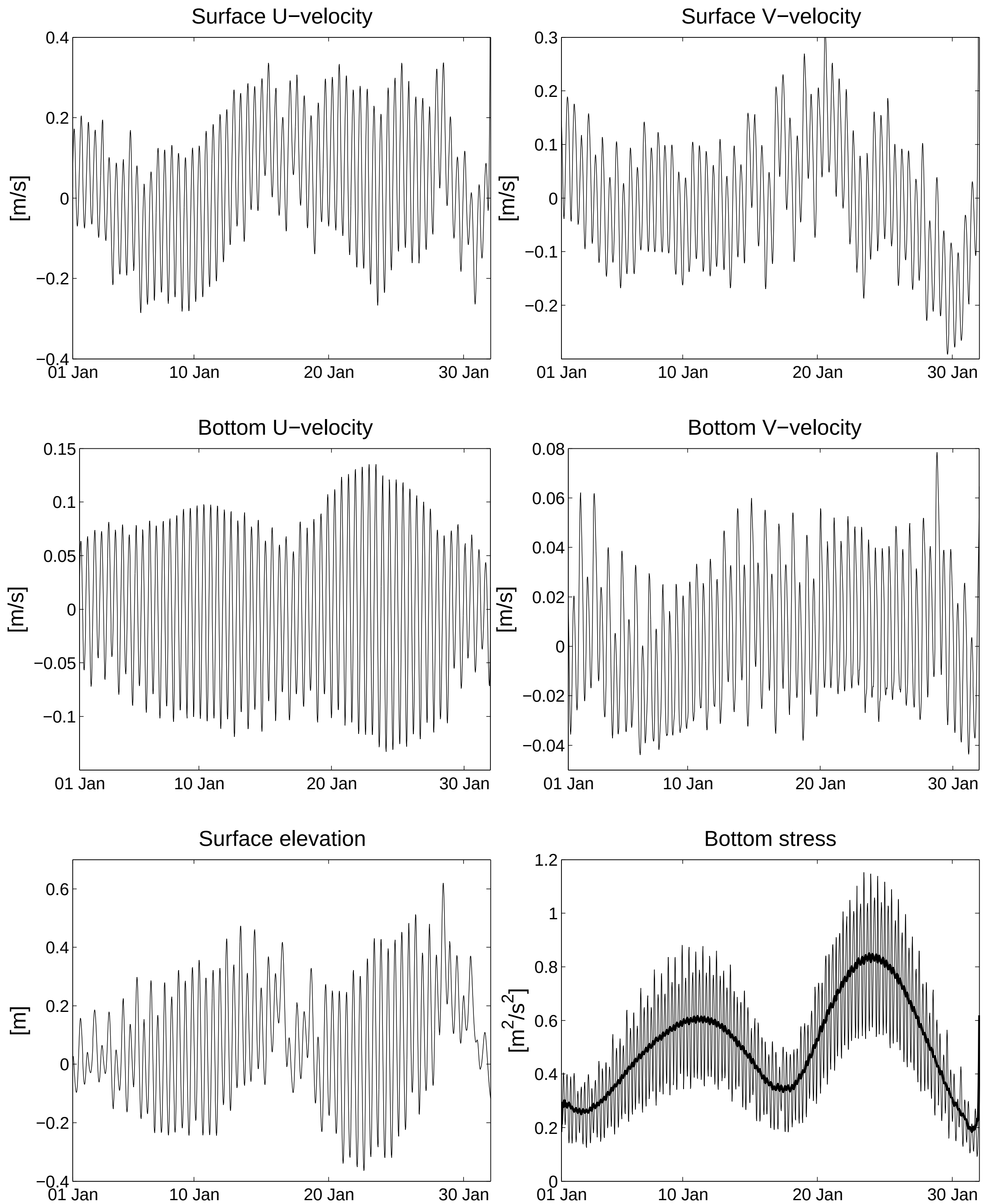


Figure 2: Continued.

1.5 River run-offs

Rivers discharges are one of the main source of TSM, along with coastal erosion (Fettweis *et al.*, 2007). The main rivers influencing the studied region are: the rivers Meuse, Rhine, Scheldt and Thames. Their rates of flow are represented in Fig. 3. Though these four rivers have different locations and the measurement have different time resolution, their signal are similar, if the amplitudes are disregarded. Measurements were obtained from Savina *et al.* (2010), who used river discharges to study transport of common sole larvae with a numerical model.

The main features we observe are:

- The maximal flows in early January, with values roughly one order of magnitude higher than the flows in spring and summer.
- A secondary maximum taking place in February.
- A period of weaker flow from April to November.

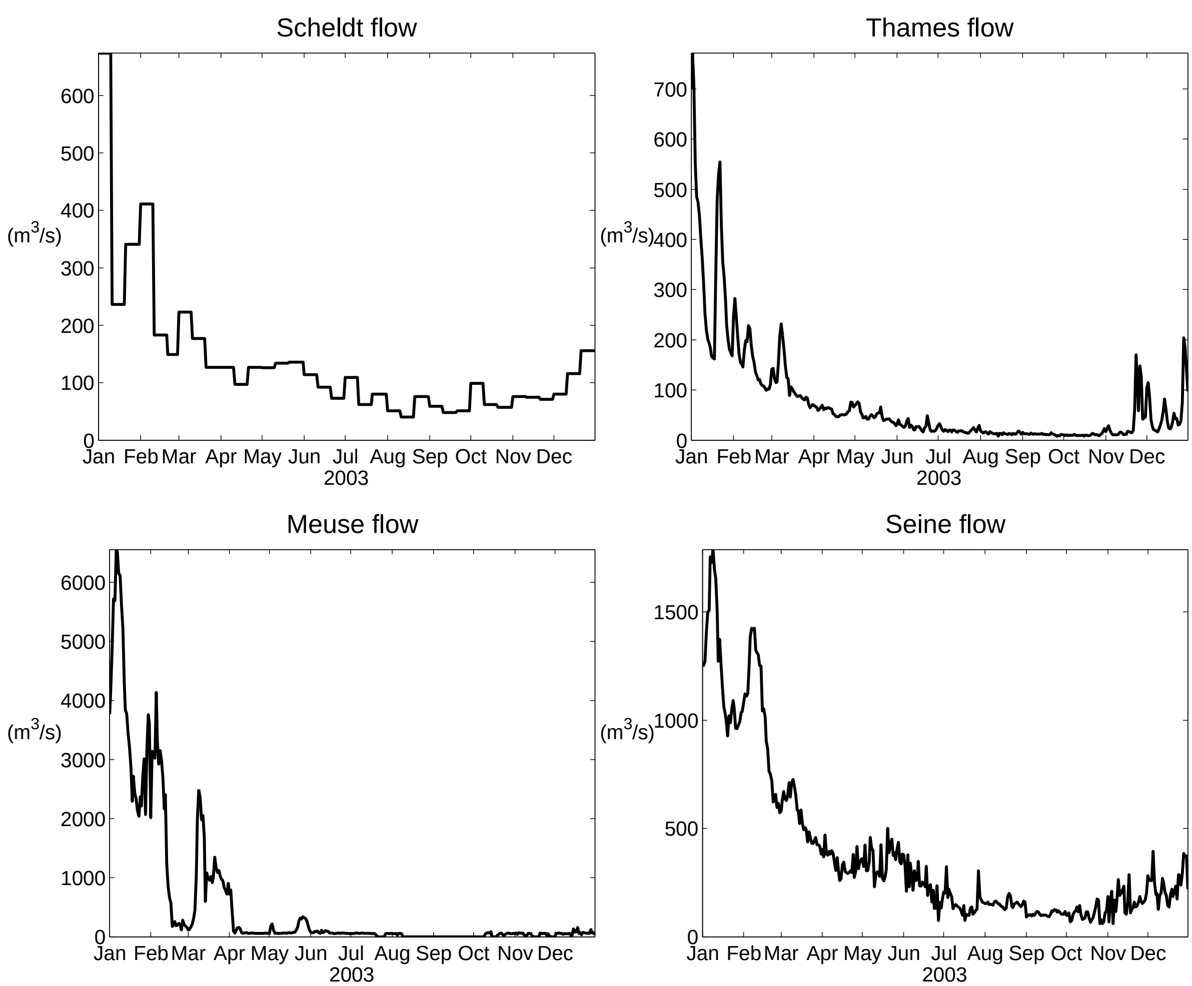


Figure 3: Rates of flow of the main rivers of the domain.

2 TSM satellite data

The total suspended matted (TSM) is measured by satellite MODIS and is obtained in the box 48°30'N to 53°N and from 4°W to 5°E. The processing of the data is described in details in Nechad *et al.* (2011). The zonal resolution is 14/100° (and the meridional resolution is 9/100°, leading to a 501×645 grid. Figure 4 shows the mean TSM field for 2003.

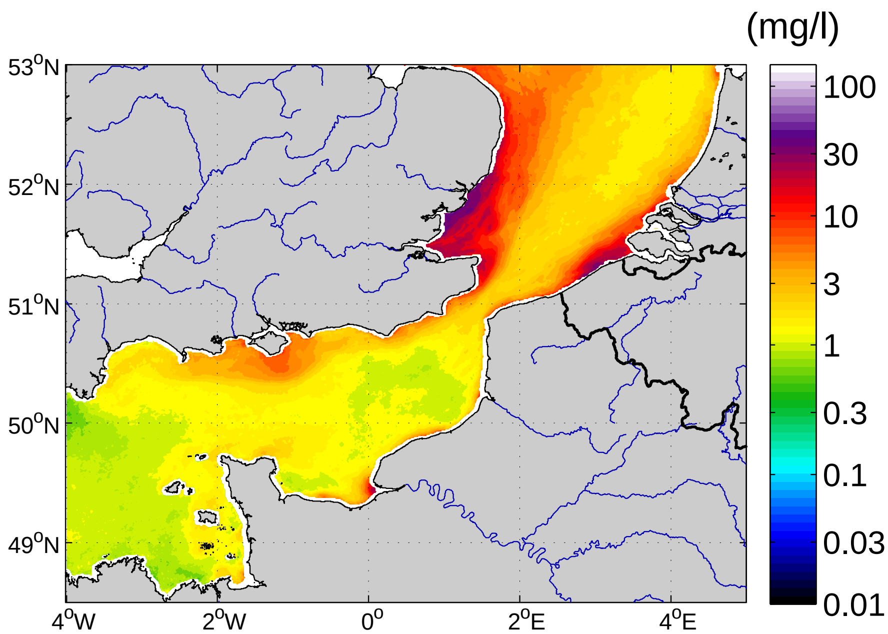


Figure 4: Mean TSM concentration for 2003.

2.1 Temporal variation

Figure 5 shows the domain-averaged TSM for the year 2003. From this time series, it is not easy to identify a particular cycle, though the winter values appear higher than the summer values. Strong peaks with value larger than 20 mg/l occur several times between January and August. Note that this time series has to be considered with precaution, as it is constructed on incomplete images, of which the spatial coverage may strongly vary from one day to another.

2.2 Seasonal cycle

The monthly-averaged fields (Fig. 6) display the maximal concentrations (30 mg/l) north of the Channel (≈ 51°N), especially along the coasts and between December and March. The lowest concentrations are observed in the southern part of the studied region.

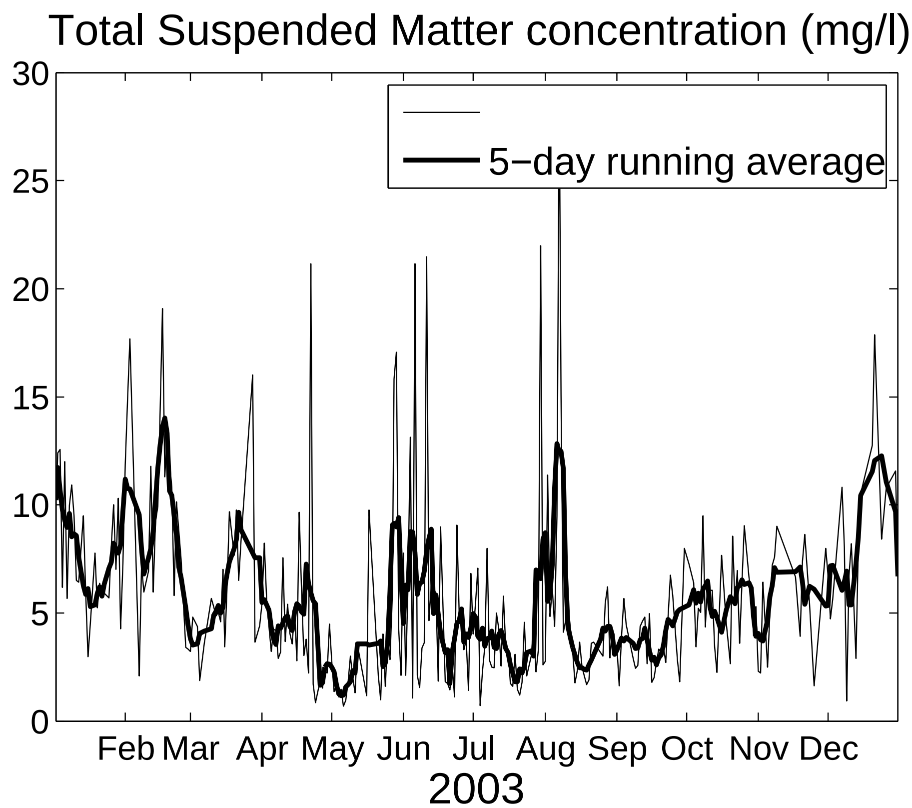


Figure 5: Time-averaged TSM concentration for 2003.

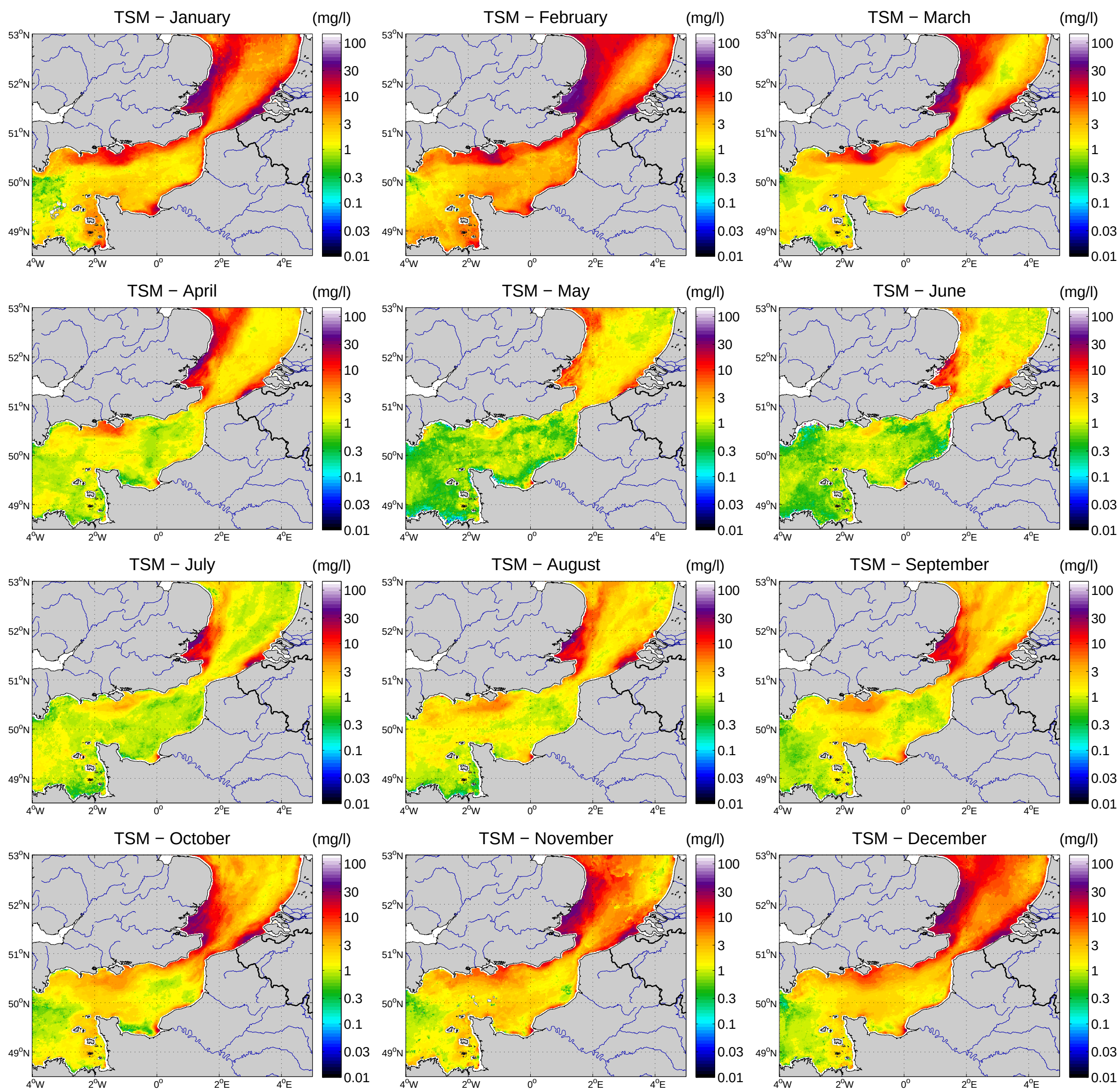


Figure 6: Monthly-averaged fields of TSM. Note the logarithmic scale

2.3 Cloud coverage

The data coverage indicates, for each pixel of the image, the ratio between the number of images where a TSM value is attributed to this pixel over the total number of images (Fig. 7a). Note that the initial time series we worked on was already processed, and all the images with less than 2% of valid pixels were removed.

Concerning the seasonal cycle of the spatial coverage, Figure 7b demonstrates that the quality of the images is highly variable, ranging from slightly more than 2% for the worst images to more than 97% for the best images.

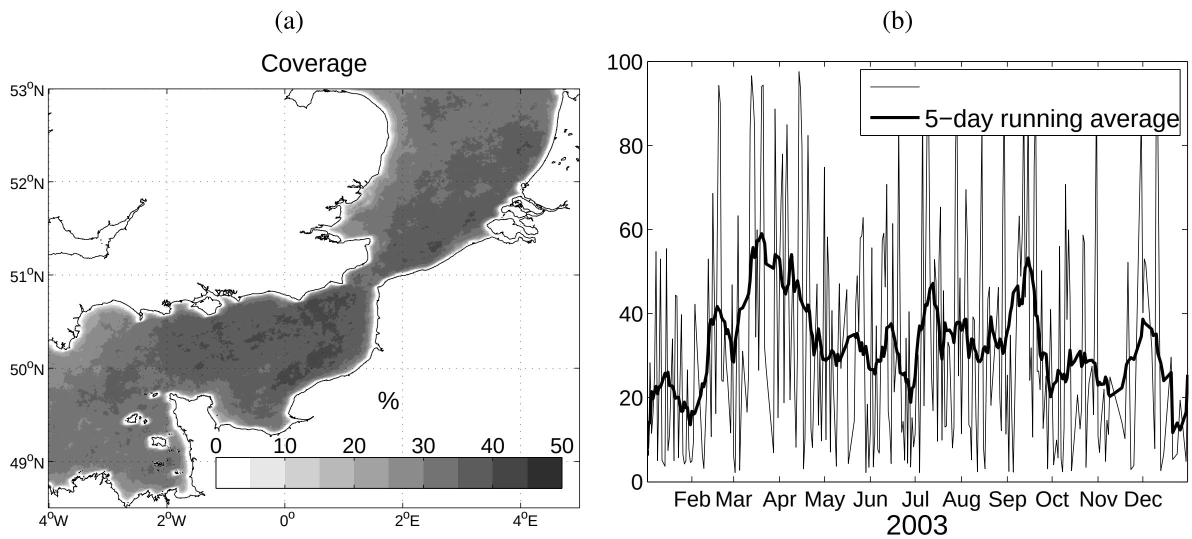


Figure 7: Spatial distribution (a) and time evolution (b) of valid pixels for the considered series of images.

3 Model variable analysis

In this section, the model variables are examined in order to determine an adequate time step and to limit the number of variables.

3.1 Correlations

The correlations between the different variables are computed as follows: let us assume that $\mathbf{A}_1$ and $\mathbf{A}_j$ are column-vectors containing the data after removal of the missing values. After the removal of the mean value:

$$\begin{aligned}\mathbf{A}'_i &= \mathbf{A}_i - \mathrm{mean}(\mathbf{A}_i)\\ \mathbf{A}'_j &= \mathbf{A}_j - \mathrm{mean}(\mathbf{A}_j)\end{aligned}$$

for $i, j = 1, 2, \ldots 10$ (the number of considered variables, see Tab. 1), the correlation is computed as:

$$r(i,j) = \frac{\mathbf{A}_i'^T \cdot \mathbf{A}'_j}{\sqrt{\left(\mathbf{A}_i'^T \cdot \mathbf{A}'_i\right)\left(\mathbf{A}_j'^T \cdot \mathbf{A}'_i\right)}}$$

We observe large correlation ($\geq 90\%$) between the individual velocity components at the surface, the bottom, or averaged over the water column ([`u2atcsurf`,`u2atcbot`, `um2atc`] and [`v2atcsurf`,`v2atcbot`, `vm2atc`])..

Moderate correlation ($40\% \leq r \leq 50\%$) between surface elevation and variables related to zonal velocity (`u2atcsurf`,`u2atcbot` and `um2atc`).

For the next analysis, the 6 model variables that will be conserved are: the bottom stress, the two components of the depth-averaged velocity, the two wind components, and the sea surface elevation.

3.2 Fourier analysis

In order to determine the main frequencies of the variables, the time series of Section 1.4 are applied a Fourier transform . For all the variable related to velocity, the Fourier spectrum evidences energy peaks at

- $\omega = 0.2224$Hz ($T = 12h29'$) for the bottom velocity components and the surface elevation (Fig. 10);
- $\omega = 0.2252$Hz ($T = 12h20'$) for the surface and averaged velocity components (not shown).

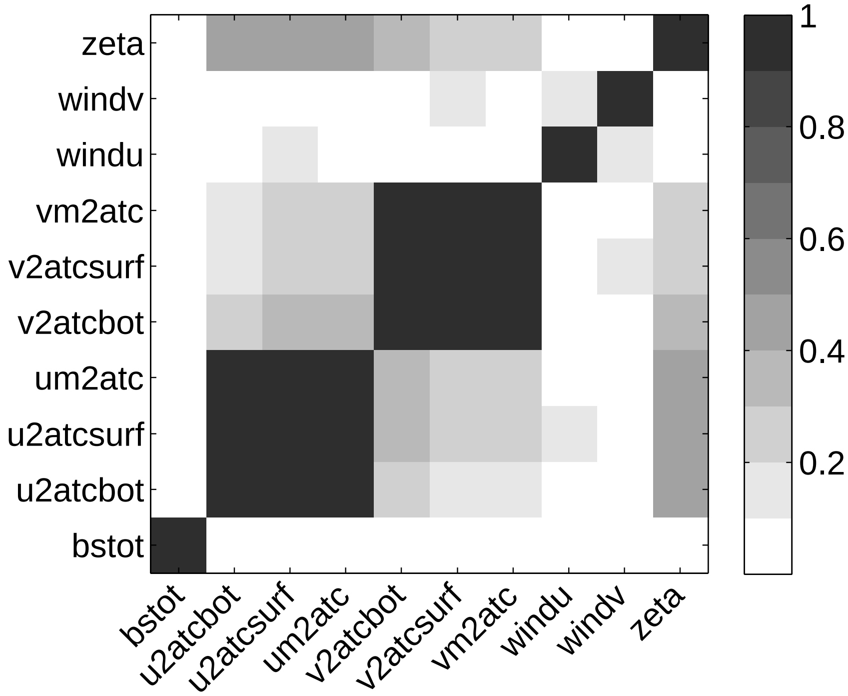


Figure 8: Correlation between the 10 initial model variables for January 2003. Darker colors indicate stronger correlations. Variable abbreviations are presented in Tab. 1.

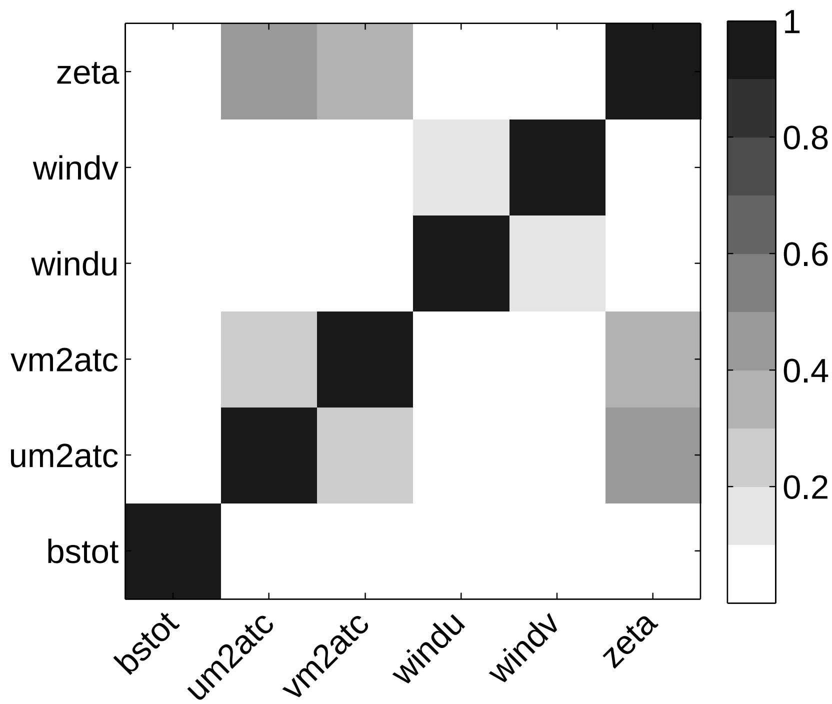


Figure 9: Correlation between the 6 selected model variables for the year 2003.

These values correspond to the M_2 tidal frequency, which is dominant in the region of study. The dominant frequencies regarding the wind components are not easily identifiable (Fig. 11).

Variable	bstot	u2atcbot	u2atcsurf	um2atc	v2atcbot
Frequency (10^{-4}Hz)	0.4476	0.2224	0.2252	0.2252	0.2224
Variable	v2atcsurf	vm2atc	windu	windv	zeta
Frequency (10^{-4}Hz)	0.2252	0.2252	0.0461	0.0353	0.2224

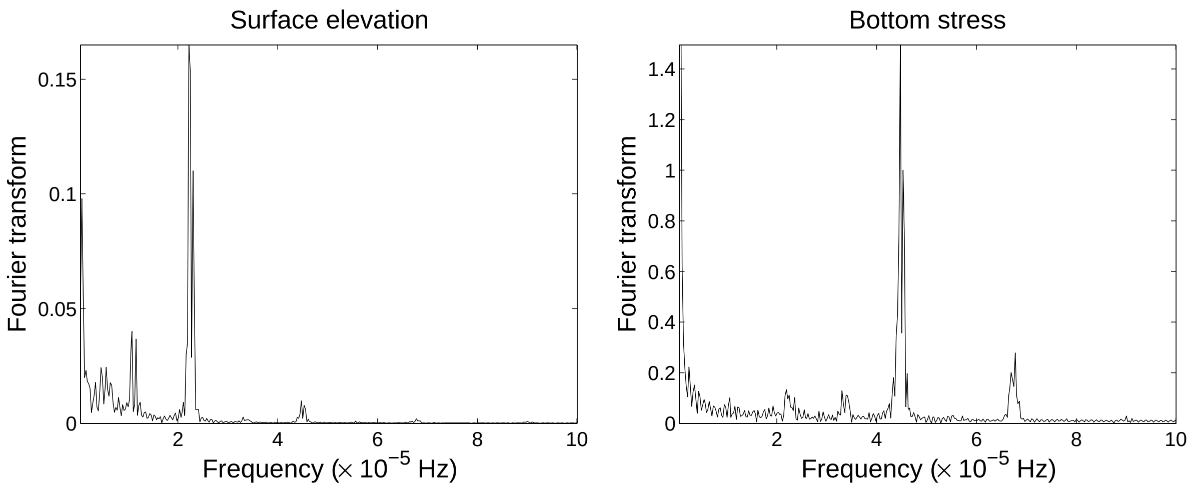


Figure 10: Energy spectra of the sea surface elevation and the bottom stress.

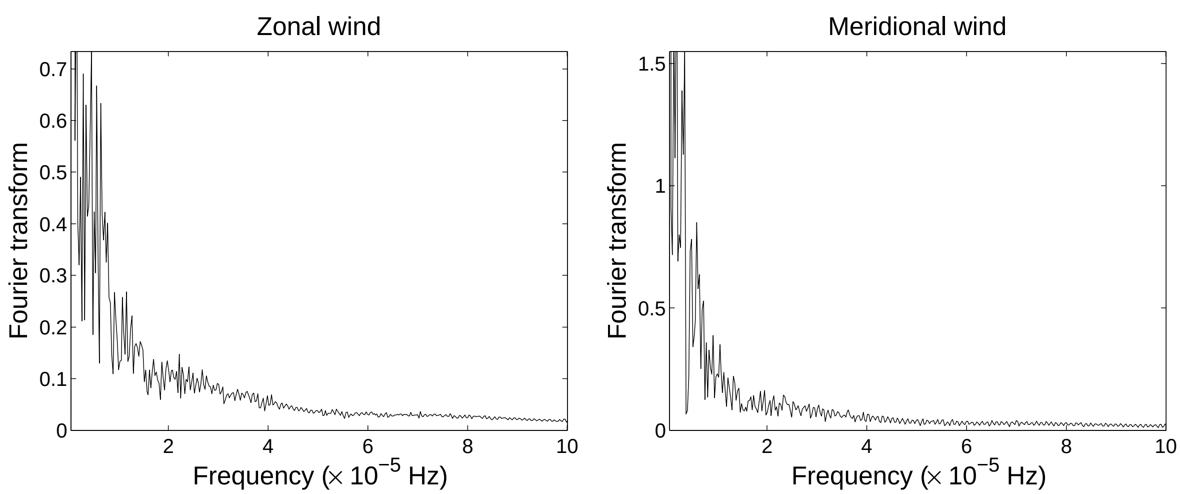


Figure 11: Energy spectra of the wind components

4 DINEOF analysis of TSM

In this section, mono-variate DINEOF analysis of TSM is carried out using different options:

1. For the validation: we use either random data points, or artificial clouds for the cross-validation. The validation points are used to determine the optimal number of modes.
2. For the variable: we process the original TSM values or apply a transformation on them. (logarithm or *anamorphosis*).

The combinations of these options lead to 6 cases presented hereinafter.

4.1 Validation points

Two methods are tested:

1. A set of 500 cross-validation points are randomly selected.
2. The validations points have the shapes of real clouds and are extracted from random images. Here we add artificial clouds on the 4 most covered images, leading to 3.04 % of cloud cover added.

4.2 Transformation on data

DINEOF is a mathematical method, hence it does not know about physics. It means that unrealistic values (e.g., negative concentrations, very high/low value of a field) may occur in the reconstructed fields, especially when the distribution of data is very different from a normal one (Fig. 12a). A simple bypass consists in working with the logarithm of the variable, perform the DINEOF analysis, and then transform back the reconstructed field. The corresponding histogram is shown in Fig. 12b.

The second transformation implemented is the anamorphosis: all the data are considered and we try to find an empirical transformation that will make the distribution close to a normal distribution (Fig. 12c). After the analysis, it is easy to go back to the original variable, since the empirical transformation is known (Fig. 12d).

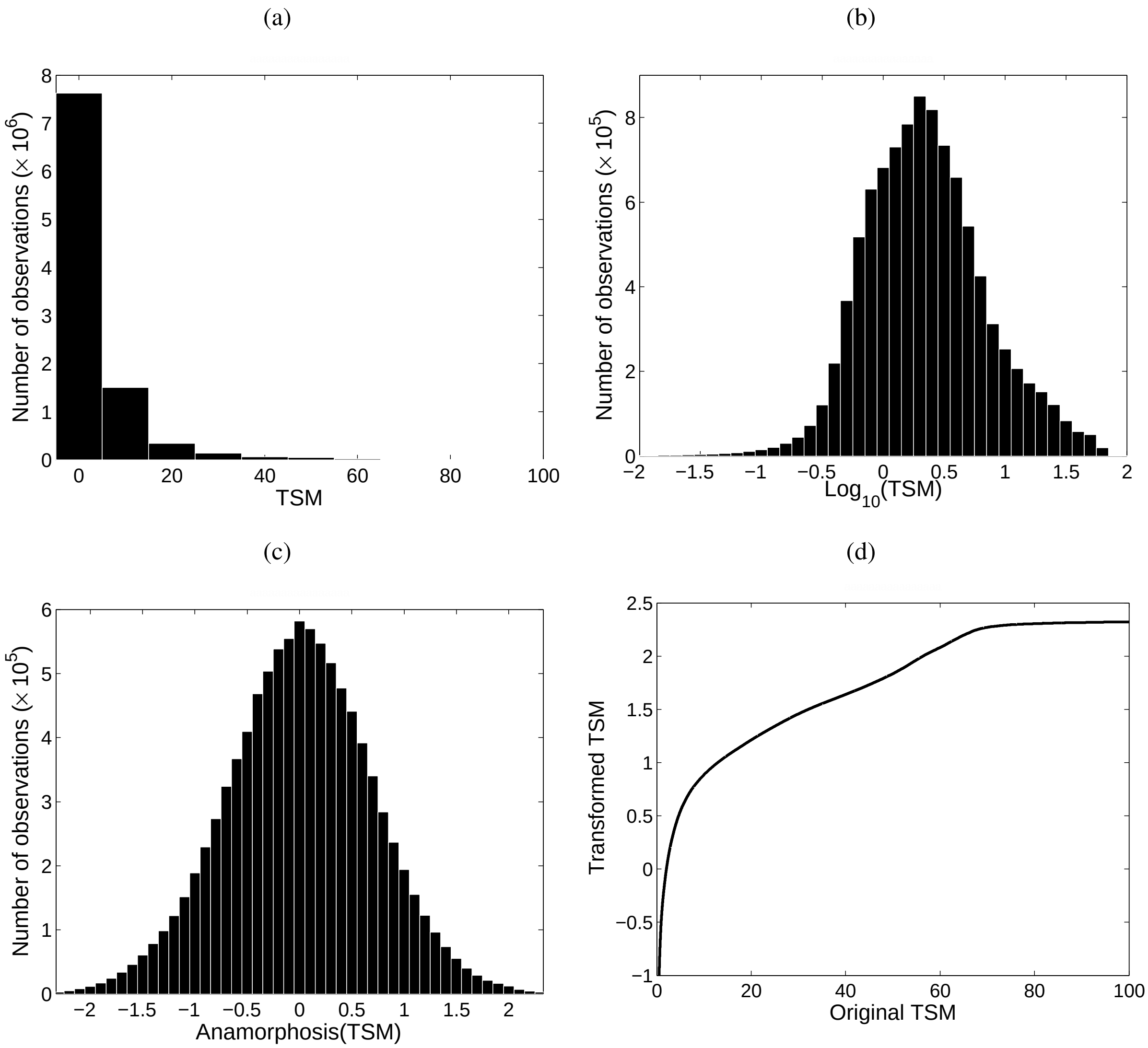


Figure 12: Histogram of the original data (a) and of the data after transformation: logarithm (b) and anamorphosis (c). Empirical transformation function (d).

4.3 Number and importance of the modes

The six configurations lead to different numbers of modes, as summarized in Tab. 2. When working with the original data (i.e., no transformation applied), the number of modes is 4, whatever the method used for the validation. When working with the logarithm, the number of modes goes from 15 (random points) to only 4 when the validation is made through artificial clouds. Finally, the number of modes relative to the anamorphosis case doesn't change much with the two validations techniques.

Also note that the data transformation improves the convergence during the iterative process in DINEOF: without any transformation, the mean number of iterations per mode is larger than 183 (the maximum number of iterations, 300, being frequently reached), whereas when we use transformation, this number decreases to about 90. As our goal is to set an implementation of DINEOF producing forecasts, any reduction of the computational time is welcome.

Table 2: Optimal number of modes for the reconstruction for different options of validation and data transformation.

Validation method \ Transformation	Original	Log	Anamorphosis
Random points	4	15	15
Clouds	4	4	13

Table 3: Importance of the first six modes (in %) for the different methods.

	Random points			Clouds		
Mode	Original	Logarithm	Anamorphosis	Original	Logarithm	Anamorphosis
1	70.67	44.44	42.58	70.67	52.05	43.07
2	10.35	27.84	29.48	10.35	31.88	30.58
3	7.52	5.20	5.25	7.52	5.49	4.97
4	5.96	3.13	3.24	5.96	3.00	3.17
5	–	2.97	2.99	–	–	2.89
6	–	2.09	1.79	–	–	1.83

4.4 Reconstructed fields

The different options were tested for the year 2003 and are presented in Fig. 13. To illustrate the different reconstructions, a day with a low cloud coverage is selected (15 February).

Using the anamorphosis can lead to problems when the analysis of the considered variable provides values out of the bounds where the transformation function (Fig. 12d) is defined. To solve this problem, it is necessary to extend the function out of the initial bounds.

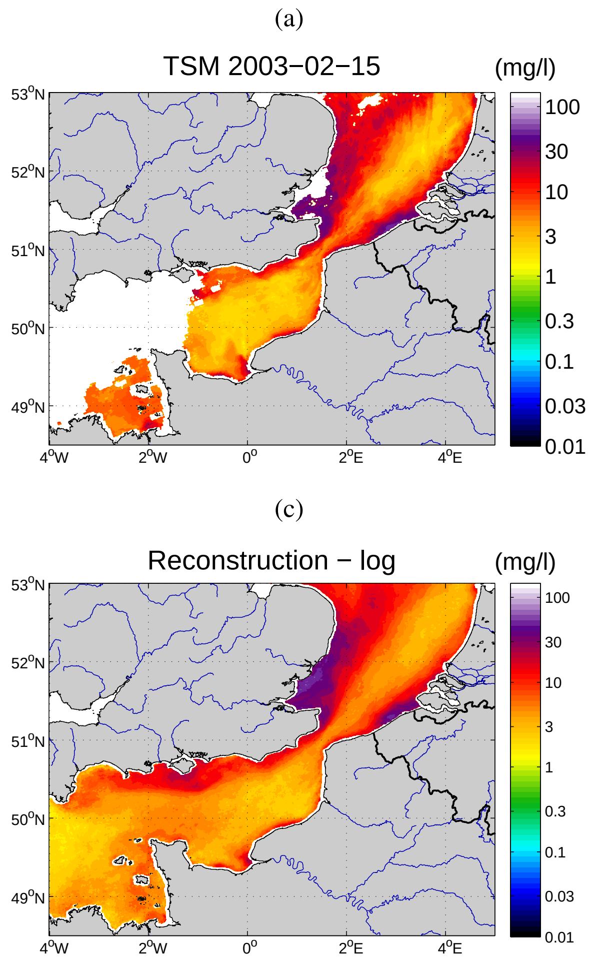


Figure 13: Original data (a) and reconstructions for February 15, 2003: no transformation (b), logarithm (c) and anamorphosis (d). The cross-validation is performed using artificial clouds.

4.5 Interpretation of the modes

The modes represented in Fig. 14 concerns the case where the logarithm of the concentration is processed and where artificial clouds are used for the cross validation.

The first spatial mode has the largest values near the river mouths, especially in the northern part of the domain (Rhine, Scheldt, Thames). The souther part of the studied region, south of the Strait of Dover (51°N), is characterized by negative values for the spatial model. The corresponding temporal mode exhibits minimal values during summer. It is positive all year long, except for a few values in early August (the minimum is reached on 07-Aug-2003). The data corresponding to this day exhibits very large values of TSM ($>$ 30 mg/l) in the south-western part of the domain, and not along the coastline. These high concentrations may be explained by external forcing (e.g., wind), or may be outliers, since the values are more often on the order of 1 mg/l in this area. The first mode accounts for about 52% of the total variability.

The second mode (31.8774% of the total variance) has a pattern less obvious to interpret. The values of the spatial modes are negative almost everywhere, the only exceptions being the Tames and the Scheldt mouths. The field reaches its most negative values in the coastal areas of France and England. The temporal mode displays a decreasing amplitude from May to mid-February, with a strong increase in spring.

The third mode (5.4905 % of the total variance) has a spatial structure similar to the first mode in the northern part of the domain, except that the signs are different. South of the Strait of Dover, there is a dipole structure, with the positive values approximatively west of 1°W. The temporal modes oscillates between negative and positive values, The extremal values observed for the first mode (early August) also appears here. The fourth mode (not shown) account for 3% of the total variance.

The evolutions of the river flows (Section 1.5) are not rejected in any of the principal modes shown in Fig. 14. The reason may be that a large part of the domain is not affected by rivers, so that the river signal is not visible in the principal components.

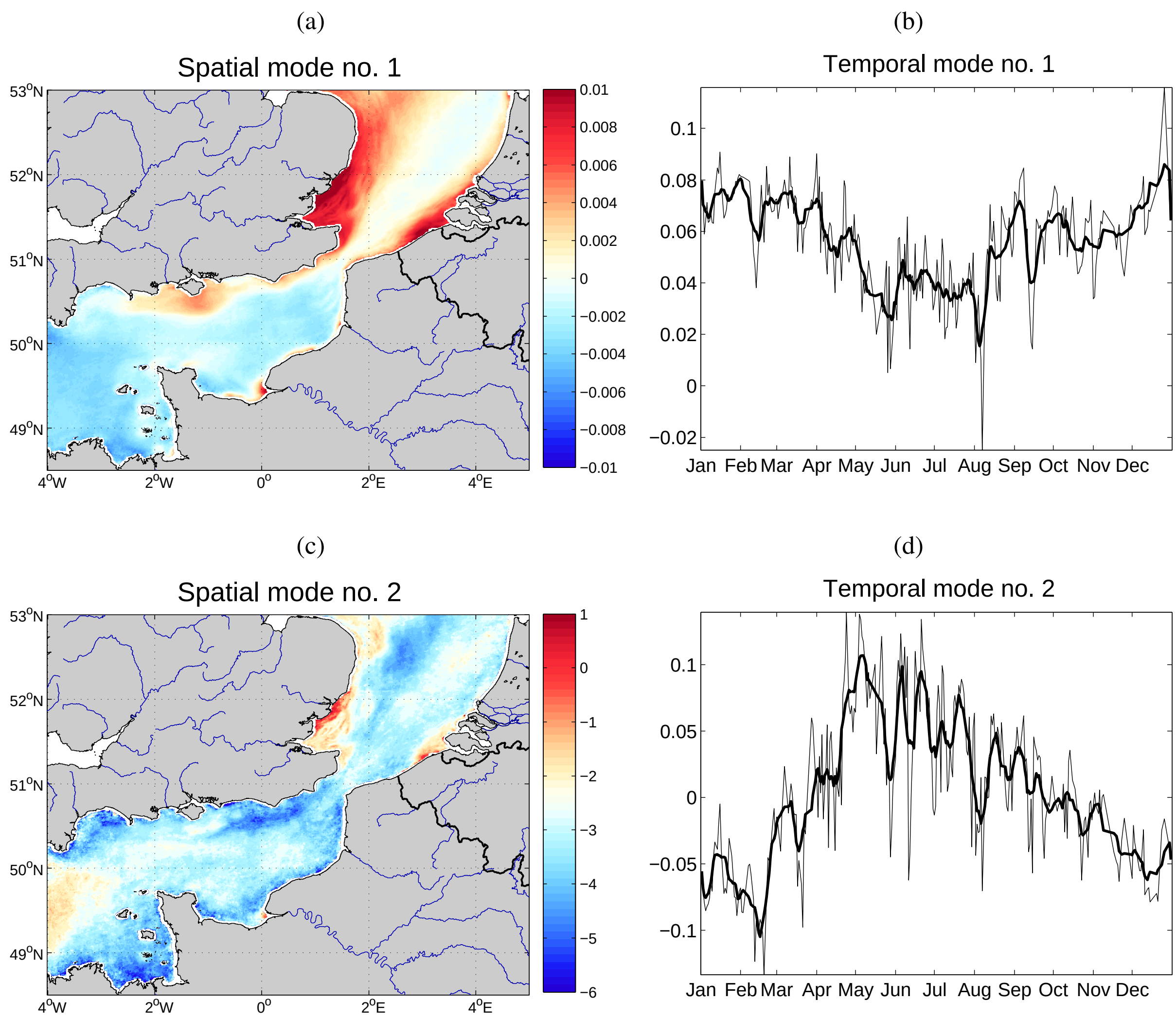


Figure 14: First three spatial and temporal modes. The thick black curves are 10-day running means.

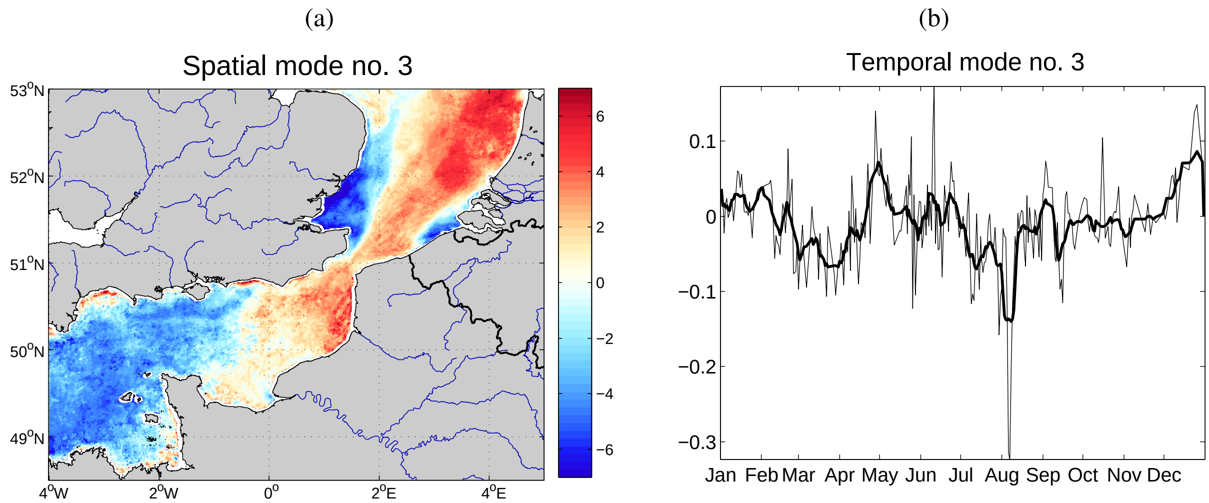


Figure 14: Continued.

5 DINEOF forecast

This section is focused on the development of DINEOF forecasts using a time series of TSM images (Section 2) along with a numerical model run in the same region (Section 1).

The method to test the forecast is the following:

1. From the TSM time series, select N_1 days of data that will serve for the reconstruction.
2. Add N_2 days of empty data, that will be reconstructed as pseudo-forecasts.
3. From the model results, extract the $N_1 + N_2$ model outputs that are the closest in time from the TSM images (approximatively taken at 1.30 PM).
4. Perform the reconstruction with different combinations of model variables.

As a first test, we word with

- $N_1 = 100$ TSM images,
- $N_2 = 10$ days of forecast.

Taking into account the available images, this means that the forecast will be made on the period 1–10 May 2003.

To assess the quality of a forecast, a comparison is made with the original TSM images for days $N_1 + 1$, $N_1 + 2$, ... The first day of forecast (1 May 2003) has a good spatial coverage, allowing for more reliable statistics on the misfits.

5.1 Univariate reconstructions

DINEOF offers the filtering of the temporal covariance matrix as a way to ensure coherence between images that are close to each other (Alvera-Azcárate *et al.*, 2009). Two parameters controls the filtering of the temporal covariance matrix:

- α (unit: squared days) specifies the strength of the filter,
- p (no unit) the reach of the filter.

Prior to the DINEOF forecast using the model outputs, a series of test with different combinations of the filtering parameters are run. There is a condition to guarantee the filter stability is related to the minimal time interval between two consecutive images:

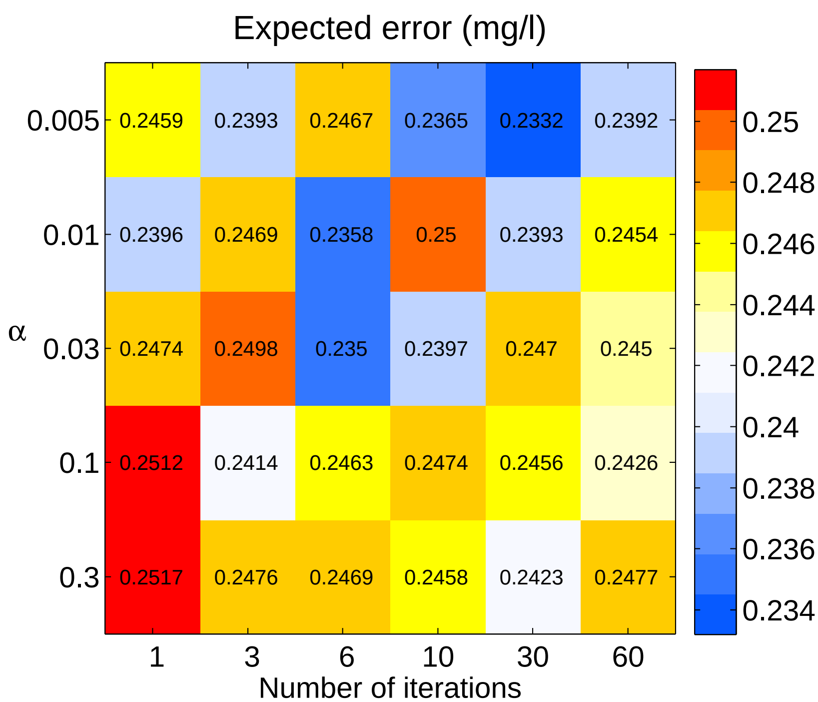
$$\alpha \leq \frac{\min(\Delta t)^2}{2}.$$

For the validation, we used artificial clouds added to the 2 clearest images, leading to a total of 4.4282 % of cloud cover. The results of the tests are presented in Figs. 15 and 16.

5.1.1 Expected error and number of modes

The lowest expected error was obtained for $\alpha = 0.005$ d^2 and 25 iterations of the filter. With these parameters, 9 modes were necessary to make the reconstruction. From Fig. 15, it is not easy to infer a tendency for the influence of the parameters on the expected error. However, it appears that large (low) values of α combined with low (high) values of p lead to large (small) errors. Also, for large values of p, the convergence of DINEOF tends to be slower.

(a)



(b)

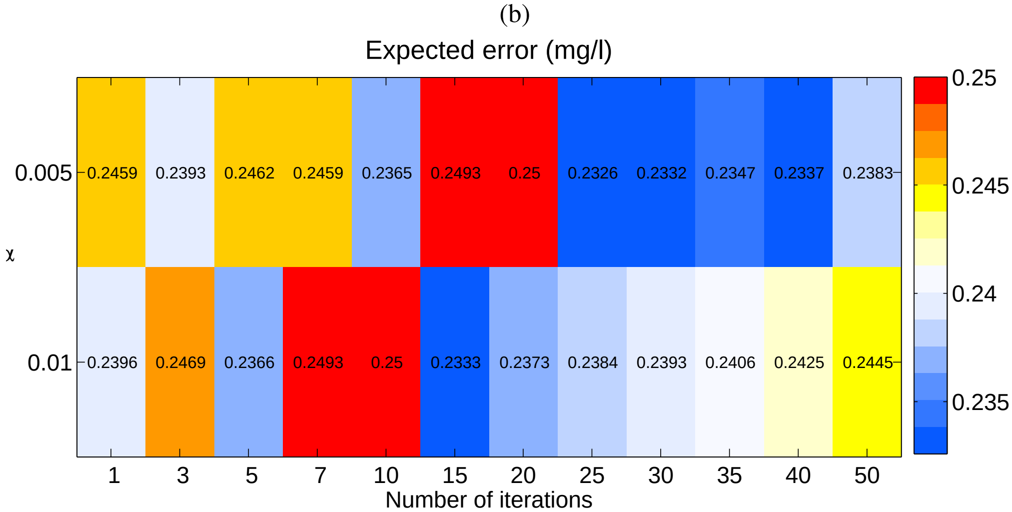
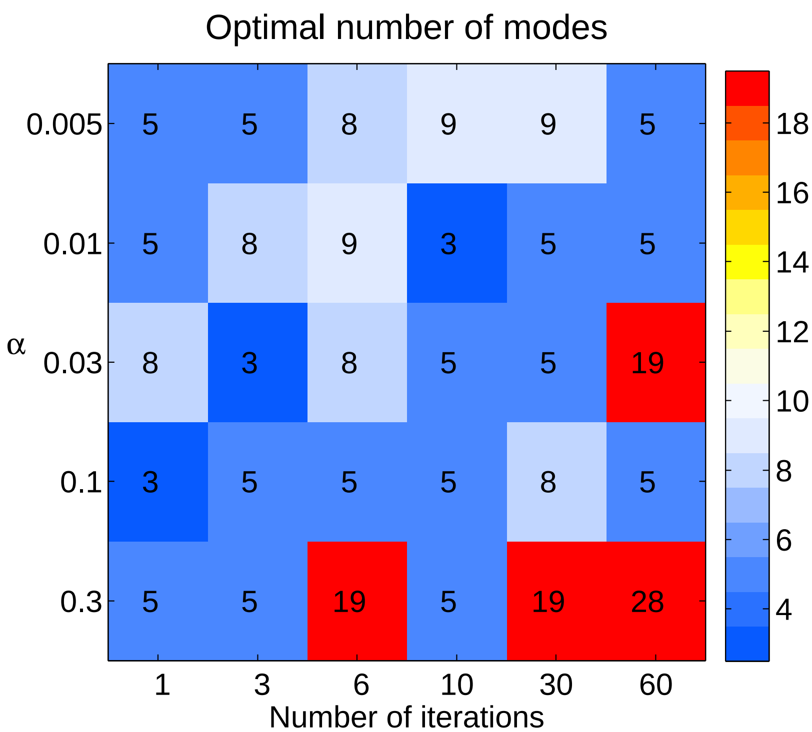


Figure 15: Expected error of the reconstruction for different couples (α, p).

The number of EOF's (Fig. 16) is stable, for the most part between 5 and 9. In some occasions it takes larger values (19, 28), probably because of the particular combinations of parameters for the filtering.

(a)



(b)

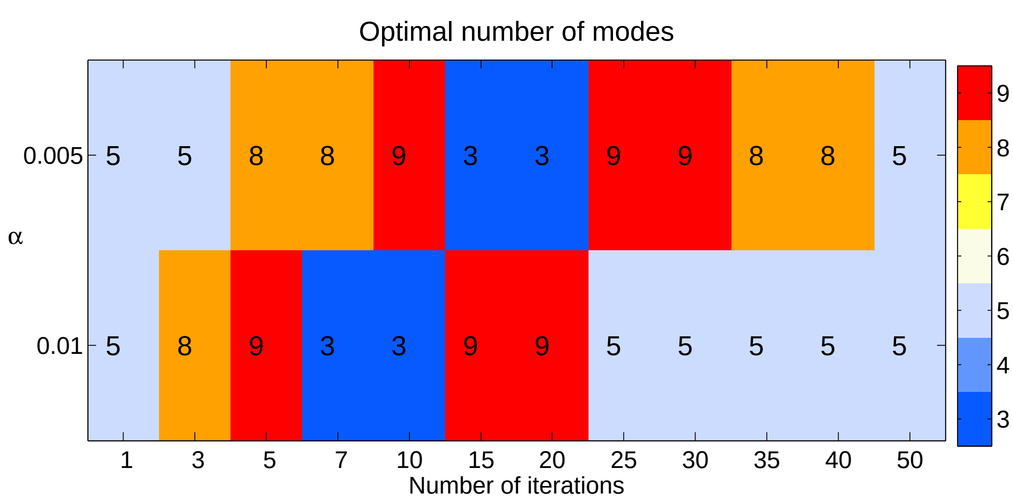


Figure 16: Optimal number of modes for different couples (α, p).

5.1.2 Correlation and RMS

Additional statistics are calculated by comparing the pseudo-forecasts for the period 1-10 May, 2003 with the original satellite images available for this period (Fig. 19): the correlation, and the RMS of the differences between the fields.Here we concentrate on the first day of forecast (1 May 2003).

The largest values of correlation (around 80%; Fig. 17) are obtained for large values of α and p. The values leading to the minimal expected error (Section 5.1.1) lead to a correlation of 74.35%. The smallest values of RMS (around 0.26; Fig. 24) are also obtained for large values of α and p.

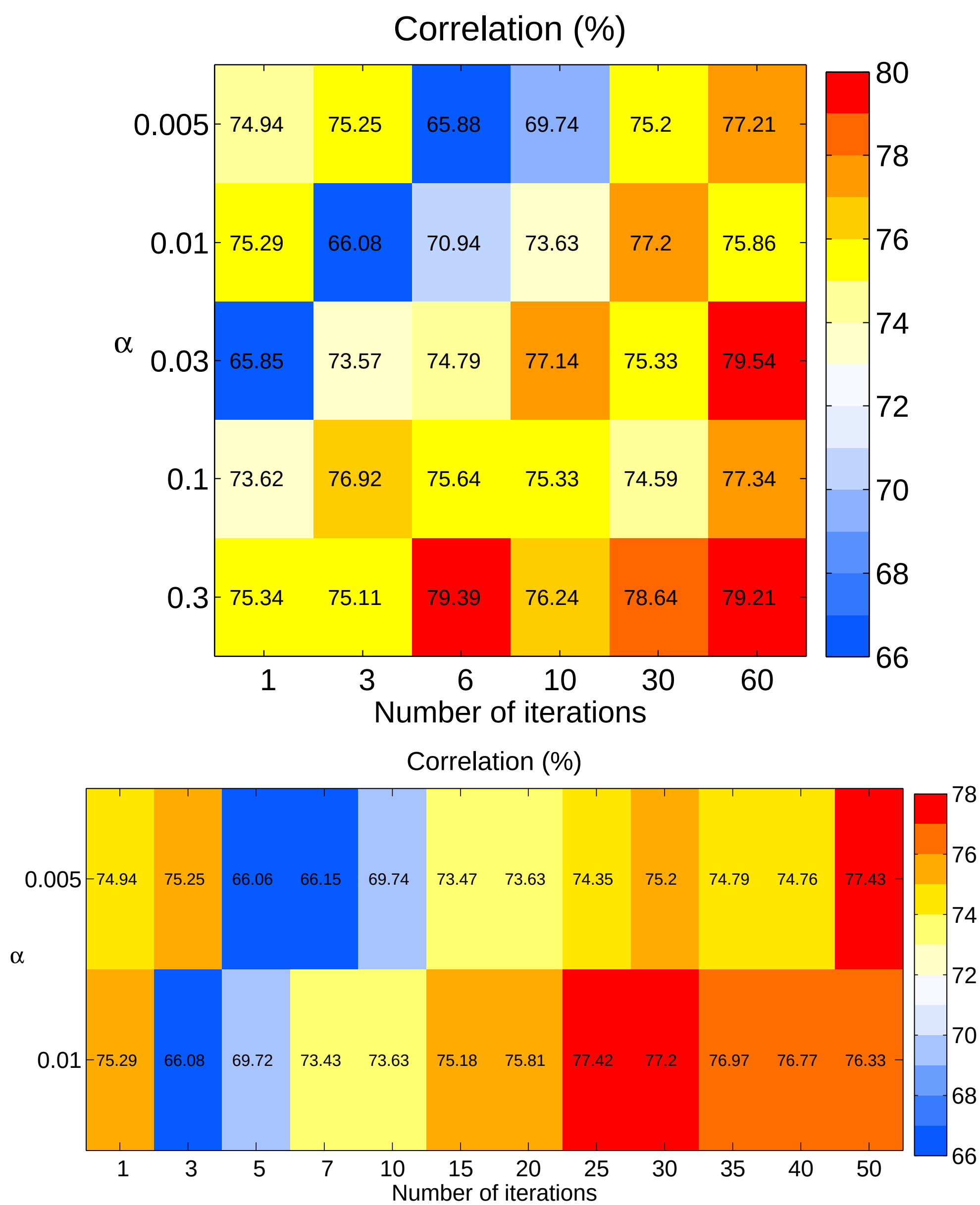


Figure 17: Correlation between the forecast and the original field for different couples (α, p).

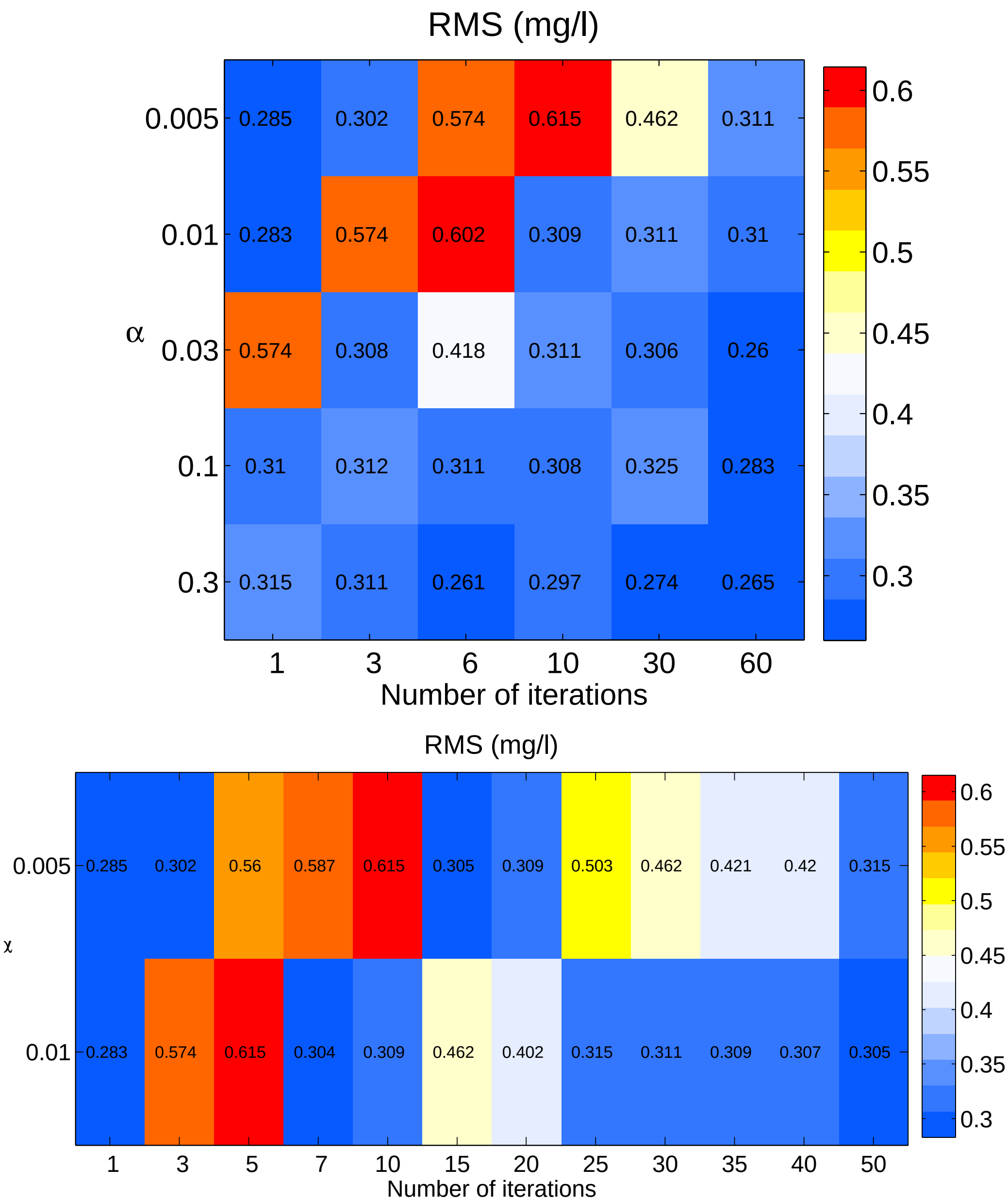


Figure 18: RMS of the difference between the forecast and the original field for different couples (α, p).

5.1.3 Results

The TSM forecasts for the period 1–3 May 2003 are performed using different combinations of parameters obtained as described in the previous section. For the sake of conciseness, only the following cases are presented:

- $\alpha = 0.005$ d^2 and $p = 30$
- $\alpha = 0.03$ d^2 and $p = 60$

- $\alpha = 0.3$ d^2 and $p = 6$

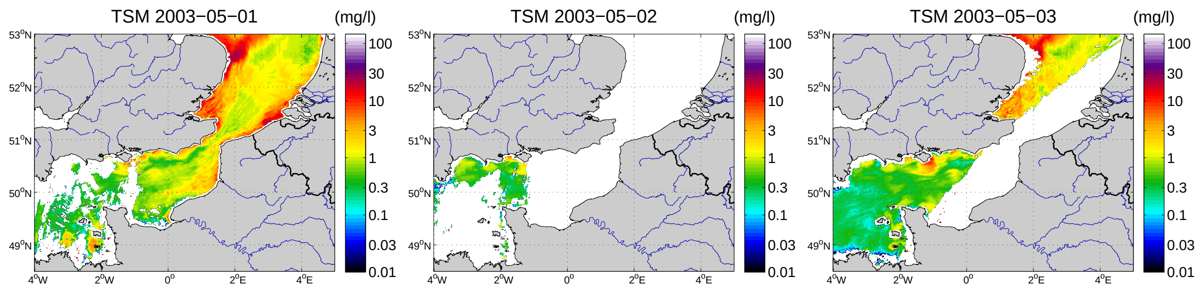


Figure 19: Original measurements of TSM corresponding to the period 1-3 May 2003.

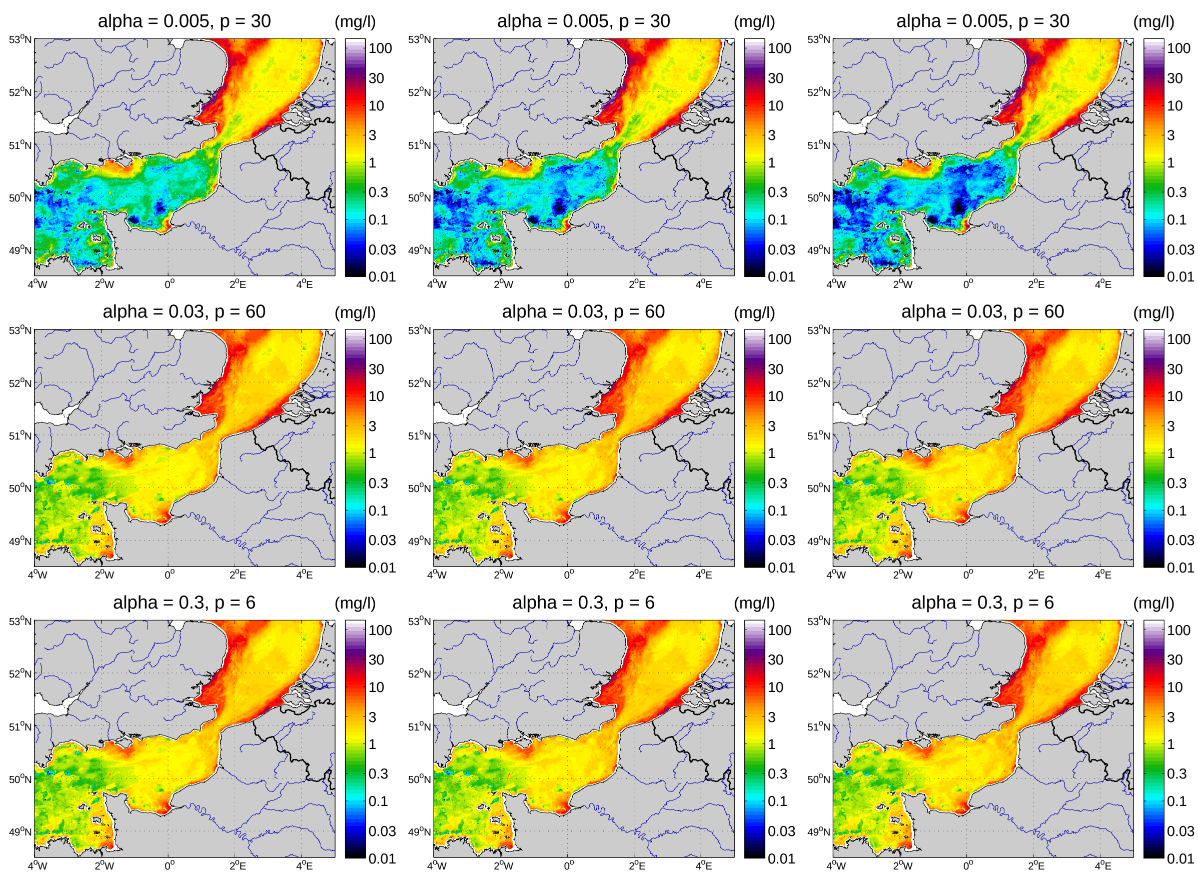


Figure 20: DINEOF forecasts the period 1-3 May 2003. The filter parameters are indicated on the top of each sub-figure.

For all the combinations of parameters, the high concentration along the coast north of 51°N are well reproduced. The forecast with $\alpha = 0.005$ d^2 and $p = 30$ underestimates the TSM in the south-western part of the domain.

The fields with ($\alpha = 0.03$ d^2, $p = 60$) and ($\alpha = 0.3$ d^2, $p = 6$) are very similar. Also, the three forecasts don't exhibit significant changes from one day to another. This is due to the choice of the parameters: the filtering of the time-covariance matrix is so that the forecast is almost a copy of the last day of data.

5.2 Multivariate reconstruction

For the multivariate reconstructions, the variables (TSM and select model variable(s)) have to be normalized:

- The vectors are normalized in order to have a unit variance.
- A second normalization is performed in order to take into account the different spatial resolutions and coverages for the considered variables (North *et al.*, 1982). Without this second normalization, the variable with the best spatial resolution would have more weight in the reconstruction.

In the present case, the model results located outside the TSM region (Fig. 1) are discarded, in order to have a common region. The ratio of the spatial resolution products yields:

$$r = \frac{\Delta \mathrm{lon}_{\mathrm{Model}} \times \Delta \mathrm{lat}_{\mathrm{Model}}}{\Delta \mathrm{lon}_{\mathrm{TSM}} \times \Delta \mathrm{lat}_{\mathrm{TSM}}} \tag{1}$$

$$= \frac{1/12^\circ \times 1/24^\circ}{14/100^\circ \times 9/100^\circ} \tag{2}$$

$$= 27.5573. \tag{3}$$

One land-sea mask was also created for each model variable, because the masks are not always the same (e.g., the wind is computed all over the domain).

5.3 TSM and Bottom stress

In this first reconstruction, the model bottom stress is used in order to improve the TSM forecast.

5.3.1 Expected error and number of modes

The next figures present again the expect error and the optimal number of modes for different combinations of parameters α and p. Figure 21a roughly shows that increasing α and/or p increases the expected error.

Concerning the number of modes, low values of α and p lead to a larger number of modes (around 15), but similarly to the univariate analysis, particular combinations lead to large number of EOF's, for instance $(0.3, 6)$ and $(0.03, 60)$.

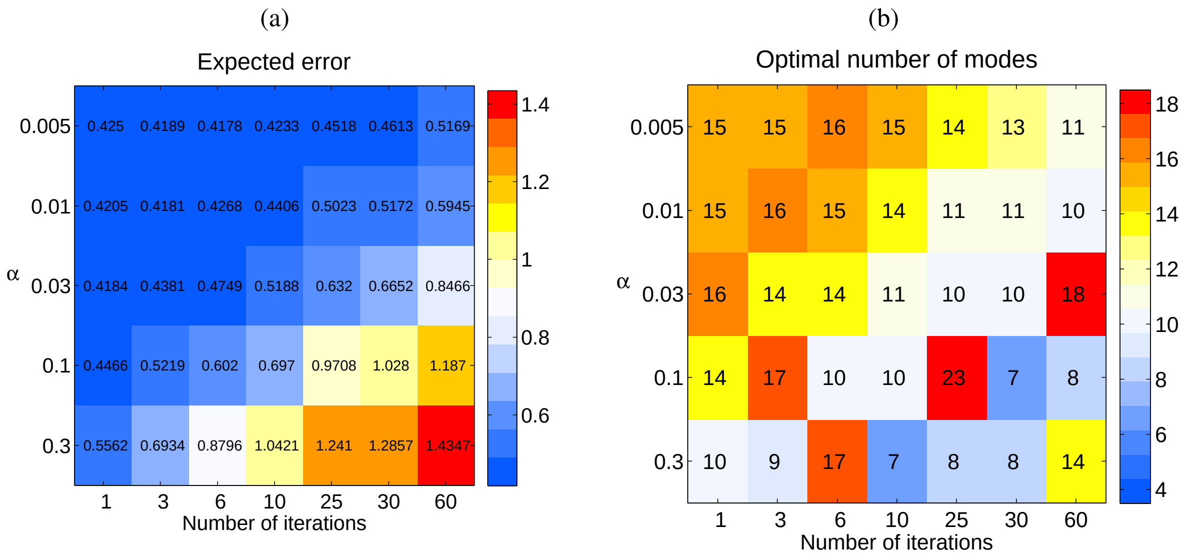


Figure 21: Expected error (a) and optimal number of modes (b) for different couples (α, p).

5.3.2 Correlation and RMS

Three combinations of α and p lead to correlations close to 80% (red tiles in Fig. 24). Two of them have a large value for the number of iterations: $p = 60$. For $\alpha = 0.03$ d^2 and $p = 60$, this means that periods shorter than

$$T = 2\pi\sqrt{\alpha p} = 8.42 \text{ d}$$

are filtered out from the data set. The comparison with Figs. 17 and 24 reveals that in general, better correlations are obtained for the first day of forecast when using TSM alone, though for some combinations of parameters, including the bottom stress in the reconstruction slightly improves the correlations. The best correlation ($\approx 77\%$) is obtained for $\alpha = 0.01$ d^2 and $p = 25$. The corresponding RMS is 0.283 mg/l.

5.3.3 Results

The corresponding reconstructions is presented along with the original data (Fig. 23). The zones of high TSM concentrations along the coasts in the northern part of the domain are correctly reproduced, while the lower concentration south of the Channel are almost not present in the reconstruction.

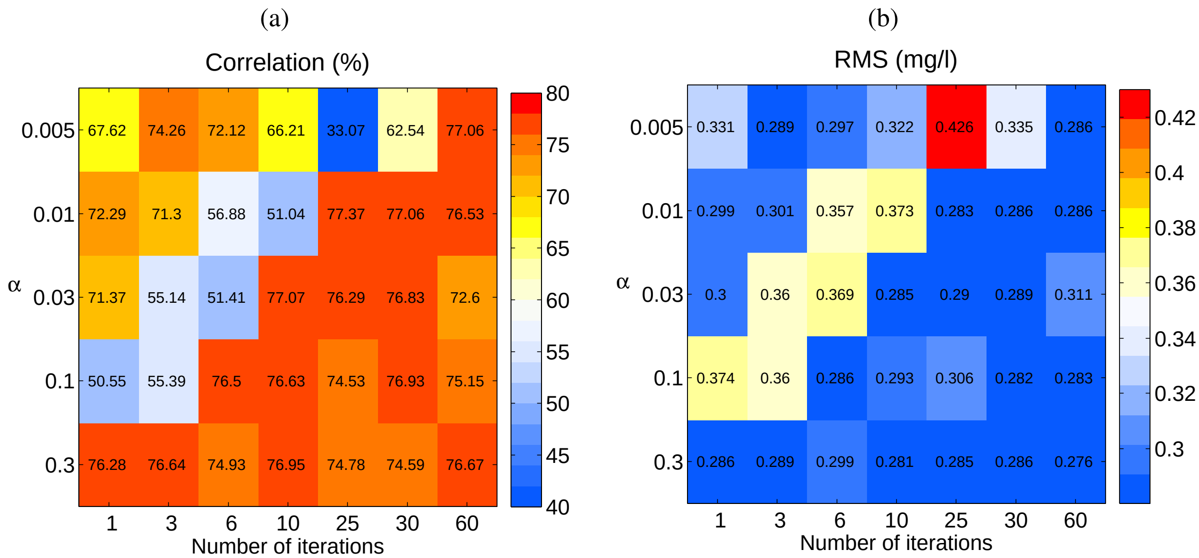


Figure 22: Correlation (a) and RMS (b) of the difference between the forecast and the original field for different couples (α, p).

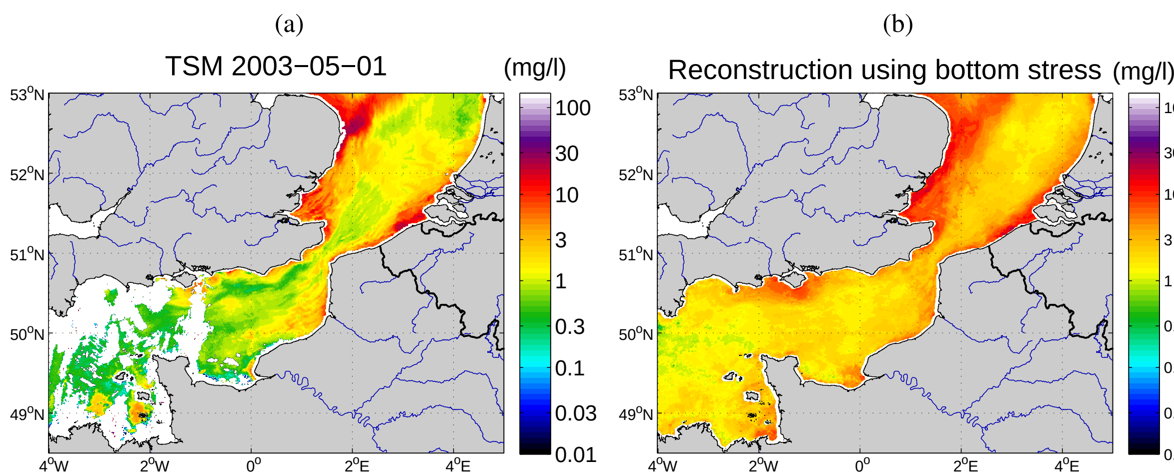


Figure 23: Original TSM and reconstruction on 1 May 2003 using the model bottom stress and $\alpha = 0.01\ d^2$, $p = 25$.

5.4 TSM and sea surface elevation

Other model variables were checked in other to see if they can provide better results and different optimal combinations of parameters.

5.4.1 Correlation and RMS

In this case, the sea surface elevation was selected. The best combination (α, p) is not the same as in the previous cases: $\alpha = 0.005$ d^2 and $p = 10$, leading to a correlation close to 78% and a RMS of 0.27 mg/l. The results are slightly improved with respect to the reconstruction using TSM and the bottom stress.

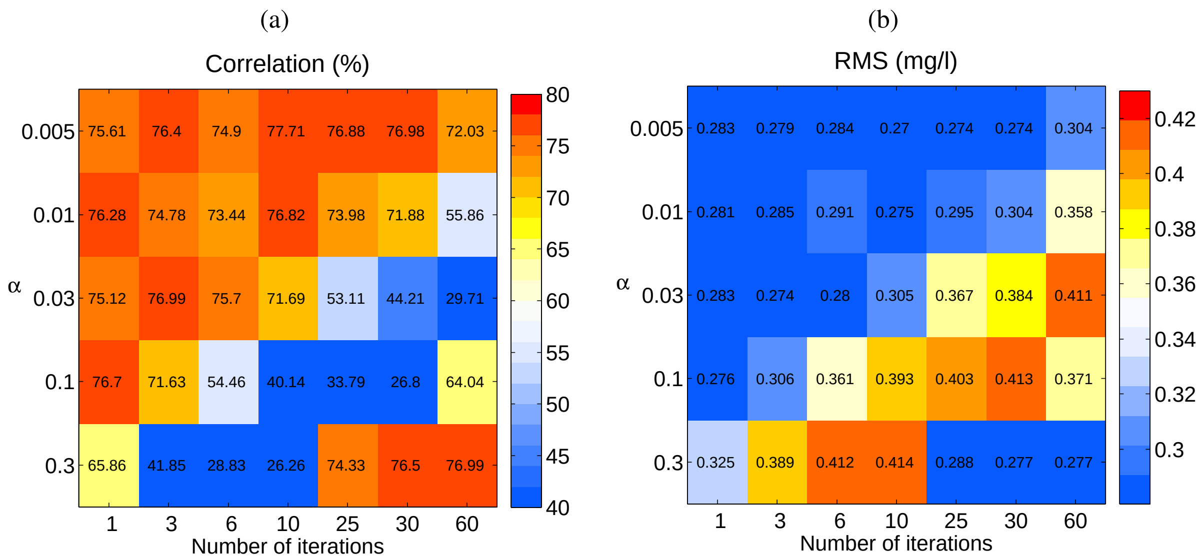


Figure 24: Correlation (a) and RMS (b) of the difference between the forecast and the original field for different couples (α, p).

5.4.2 Results

The lower values of the southern part of the domain are better reproduced than in the previous case with the bottom stress.

5.5 Summary

The different test carried out to examine the possible improvements of the reconstruction using a model variable are summarised in Table 4. The main information is that none of the analysis combining the TSM and a model variable provided better results (as measured by the correlation) than the case when TSM was used alone. Several reasons may explain this observation:

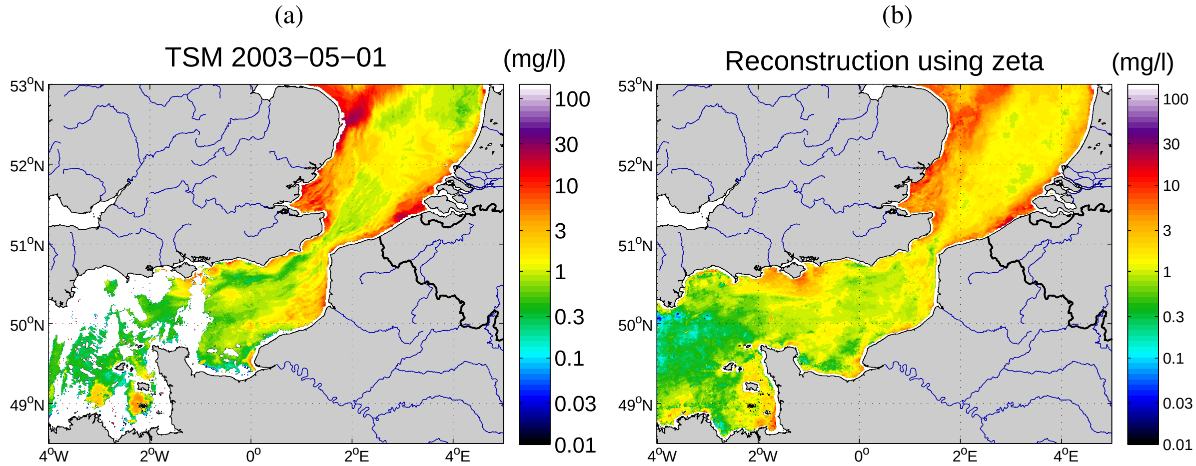


Figure 25: Original TSM and reconstruction on 1 May 2003 using the model sea surface elevation and $\alpha = 0.005\ d^2$*,* $p = 10$*.*

- The measures of the quality of the forecast (correlation or RMS) don't constitute the best option.
- The quality of the forecast depends on the selected number of modes for the reconstruction. The optimal number is the one which minimised the expected error, but not necessarily the number that provides the best forecast.
- The correlations between the TSM and the model variables are not strong enough.

The values for α and p that lead to the strongest correlations are different in each case.

Table 4: Correlation, RMS and best combination of parameters for different reconstructions (univariate and multivariate). Note that the case (TSM + velocity) was not presented in the previous sections.

	Correlation	RMS (mg/l)	α (d^2)	p
TSM only	79.54	0.260	0.03	60
	79.39	0.261	0.3	6
TSM + bstot	77.37	0.283	0.01	25
TSM + zeta	77.71	0.270	0.005	10
TSM + velocity	76.08	0.335	0.3	30

6 Conclusions

In this report, we address the problem of forecasting a variable measured by satellite (TSM) using

- previous satellite measurements of this variable,
- a numerical model run in the same region and able to provide forecast of physical variables (velocity components, wind, bottom stress, ...).

In DINEOF, two parameters determine the filtering of the covariance matrix:

1. α, which specifies the strength of the filter,
2. p, which controls the reach of the filter.

Both in univariate and multivariate analysis, the choice of these parameters may improve the quality of the forecast, as measured by the RMS and the correlation.

In the southern part of the North Sea, the various tests performed did not show any improvement in the forecast, meaning that the information contained in the numerical model does not help for the reconstruction.

Future work will consist in running similar analysis, but with two variables already well-correlated (e.g., forecast the chlorophyll concentration using sea-surface temperature obtained with the model).

In the particular case of TSM, other sources of information could be used to better constrain the reconstruction, for instance the river flows, which influence the TSM concentrations near to the coast, or the tidal coefficients (e.g., Rivier *et al.*, 2012).

Acknowledgments

J. Van Campenhout and F. Petit (ULg) are acknowledged for their help with the river data.

References

Alvera-Azcárate, A., Barth, A., Sirjacobs, D., & Beckers, J.-M. (2009). Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF. *Ocean Science*, **5(4)**: 475–485. doi:10.5194/os-5-475-2009.
URL www.ocean-sci.net/5/475/2009/

Fettweis, M., Nechad, B. & den Eynde, D. V. (2007). An estimate of the suspended particulate matter (SPM) transport in the southern North Sea using SeaWiFS images, in situ measurements and numerical model results. *Continental Shelf Research*, **27(10-11)**: 1568–1583. doi:10.1016/j.csr.2007.01.017.
URL http://www.sciencedirect.com/science/article/pii/S0278434307000477

Lacroix, G., Ruddick, K., Ozer, J. & Lancelot, C. (2004). Modelling the impact of the Scheldt and Rhine/Meuse plumes on the salinity distribution in Belgian waters (southern North Sea). *Journal of Sea Research*, **52**: 149–163. doi:10.1016/j.seares.2004.01.003.

Nechad, B., Alvera-Azcárate, A., Ruddick, K. & Greenwood, N. (2011). Reconstruction of MODIS total suspended matter time series maps by DINEOF and validation with autonomous platform data. *Ocean Dynamics*, **61**: 1205–1214. doi:10.1007/s10236-011-0425-4.
URL http://www.springerlink.com/content/yj51321705624143/fulltext.pdf

North, G. R., Bell, T. L. & Cahalan, R. F. (1982). Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review*, **110(7)**: 699–706. doi:10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2.
URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%281982%29110%3C0699%3ASEITEO%3E2.0.CO%3B2

Rivier, A., Gohin, F., P., Bryère, Petus, C., Guillou, N. & Chapalain, G. (2012). Observed vs. predicted variability in non-algal suspended particulate matter concentration in the English Channel in relation to tides and waves. *Geo-Marine Letters*, **32**: 139–151. doi:10.1007/s00367-011-0271-x.

Savina, M., Lacroix, G. & Ruddick, K. (2010). Modelling the transport of common sole larvae in the southern North Sea: Influence of hydrodynamics and larval vertical movements. *Journal of Marine Systems*, **81(1-2)**: 86–98. doi:10.1016/j.jmarsys.2009.12.008.
URL http://www.sciencedirect.com/science/article/pii/S0924796309003467

List of Figures

1 Spatial domain where the outputs of the numerical model are available. 4

2 Time evolution of the variables averaged on the domain. 5

3 Rates of flow of the main rivers of the domain. 7

4 Mean TSM concentration for 2003. 8

5 Time-averaged TSM concentration for 2003. 9

6 Monthly-averaged fields of TSM. 10

7 Spatial distribution and time evolution of valid pixels for the considered series of images. 11

8 Correlation between the 10 initial model variables for January 2003. 13

9 Correlation between the 6 selected model variables for the year 2003. 13

10 Energy spectra of the sea surface elevation and the bottom stress. 14

11 Energy spectra of the wind components 14

12 Histogram of the original data and of the data after transformation. 16

13 Original data and reconstructions for February 15, 2003. 18

14 First three spatial and temporal modes. 19

15 Expected error of the reconstruction for different couples (α, p). 22

16 Optimal number of modes for different couples (α, p). 23

17 Correlation between the forecast and the original field for different couples (α, p). 24

18 RMS of the difference between the forecast and the original field for different couples (α, p). 25

19 Original measurements of TSM corresponding to the period 1-3 May 2003. . . 26

20 DINEOF forecasts the period 1-3 May 2003. 26

21 Expected error (a) and optimal number of modes (b) for different couples (α, p). 28

22 Correlation (a) and RMS (b) of the difference between the forecast and the original field for different couples (α, p). 29

23 Original TSM and reconstruction on 1 May 2003 using the model bottom stress and $\alpha = 0.01$ d^2, $p = 25$. 29

24 Correlation (a) and RMS (b) of the difference between the forecast and the original field for different couples (α, p). 30

25 Original TSM and reconstruction on 1 May 2003 using the model sea surface elevation and $\alpha = 0.005$ d^2, $p = 10$. 31

List of Tables

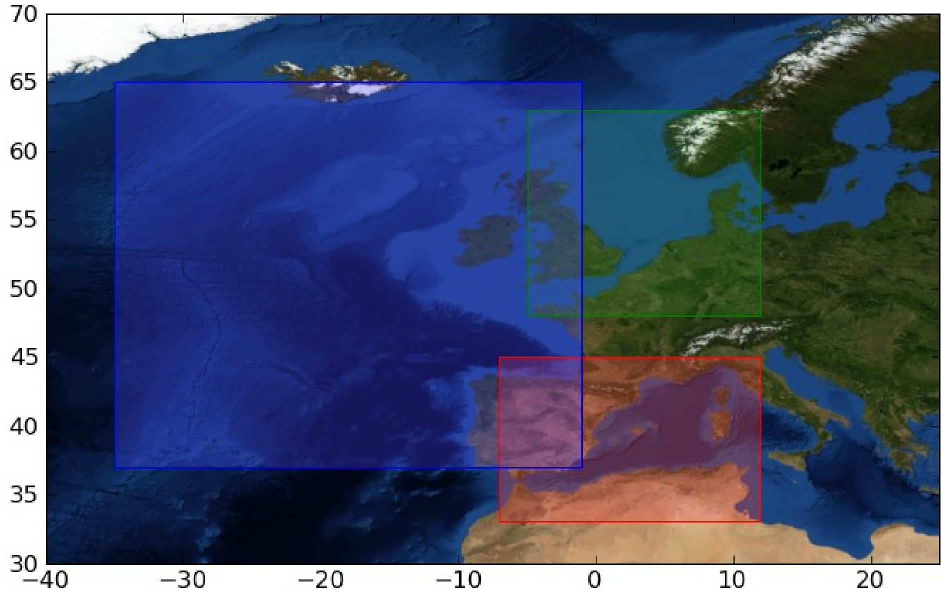
1 Variables extracted from the model. 5
2 Optimal number of modes for the reconstruction for different options of validation and data transformation. 17
3 Importance of the first six modes for the different methods. 17
4 Correlation, RMS and best combination of parameters for different reconstructions. 31

Annex VII

Project file

SR/12/140 - HISEA

(Geographic) study area : The North Sea will be the main region of study. Additional tests will be carried out in the western Mediterranean Sea, and the north-east Atlantic Ocean, if needed for specific developments.



Context and objectives (max 14 lines)

Several satellites measure Sea Surface Temperature (SST), each of these with different technical specificities and error sources. Together with in situ data, they form a highly complementary data set. The creation of merged SST products, integrating the strengths of each of its components and minimising their weaknesses, is however not an easy task, but it is certainly a desirable goal that has generated a large amount of research over the last years. The objectives of this project are: (i) To develop a technology that allows to merge different data sets at very different sampling intervals (in space and time) and create an integrated product at the highest sampling frequency and with the highest quality possible. (ii) To provide improved, merged analyses of variables such as SST and chlorophyll. (iii) Obtain a better understanding of inter-sensor differences, and of the diurnal cycle of the studied variables. (iv) To better understand the relation between variables (and take advantage of this improved knowledge to ameliorate the analyses). (v) Using the above-mentioned developments, explore the capability of DINEOF to produce SST forecasts based on multi-variate EOFs and model forecasts. (vi) Finally, to improve DINEOF to meet user needs and required precision.

Methodology (max 11 lines)

DINEOF is a technique to infer missing data is satellite data sets. In this project we will further develop DINEOF so that it can merge different data sets. First, an initial DINEOF reconstruction of a data set with a high spatial resolution will be made, and the EOF basis obtained will be used as the covariance matrix needed to subsequently include into the analysis other data sources (satellite and in situ). Error estimations for each data set will be used to weight their influence in the final product.

Special attention will be given to the diurnal cycle and the effect of diurnal warming in the quality of the measurements, and multivariate DINEOF analyses will be performed to investigate the influence of variables like wind and turbidity in these warming events. Finally, by combining satellite SST fields and model forecasts using a multivariate DINEOF, we will investigate the capability of DINEOF to produce SST forecasts, which will be compared to the forecast provided by numerical analyses.

Results expected (max 16 lines)

Merged high-resolution (in space and time) SST data sets and error statistics will be obtained for the study zones, for variable time frames within 2008 and 2010. The improvements made to the base technique used throughout the project (DINEOF) will be made available freely and openly the the scientific community (source code and documentation). Also, statistical parameters characterising the diurnal cycle and the difference between skin and bulk temperature will be obtained through this project. Finally, the technology to forecast SST based on statistical information and model data will be obtained as well. The portability of the developed techniques to other variables and domains will be reported. All results will be published in international peer-reviewed journals.

Products and services (if applicable)

Execution

Period: December 2010 - November 2012

Laboratory: AGO-GHER-MARE, University of Liège

Discipline (select one or more appropriate disciplines)

Oceans & coasts

Environment

General Earth observation