



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.

© 2011 Elsevier Inc. All rights reserved.

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 $\sim 2 \times 2$ 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).

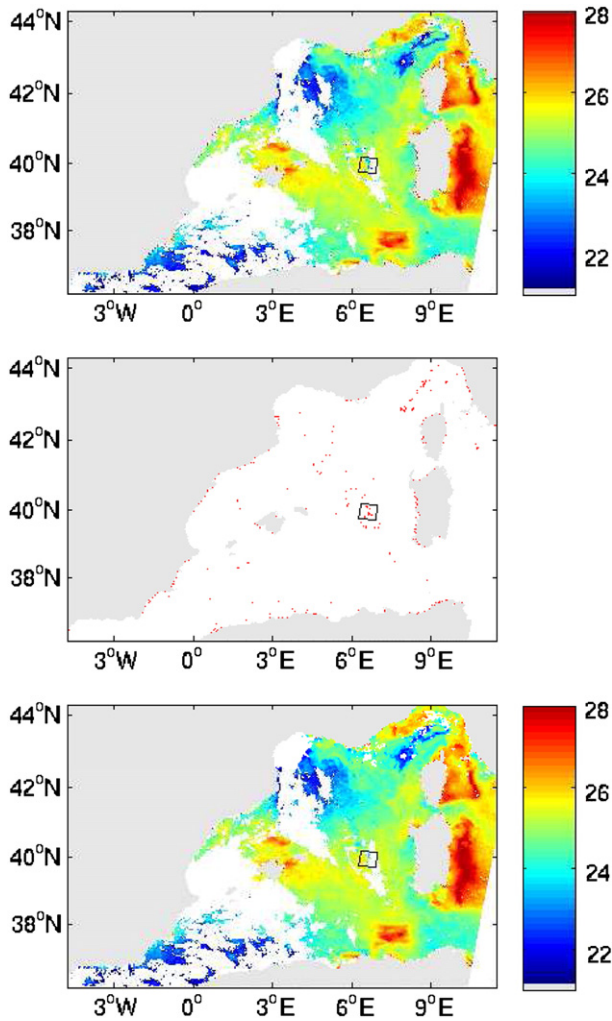


Fig. 1. Study domain: the western Mediterranean Sea. Top panel: SST (°C) on 4 July 2010. Middle panel: outliers detected. Bottom panel: SST data with outliers removed. The black square indicates the zone where a detail is shown in Fig. 2.

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 (SeaWiFS), 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:

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

where $i = 1, \dots, m$ 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:

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

where N is the number of EOFs retained by DINEOF for the reconstruction and $k = 1, \dots, N$. 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:

$$\mathbf{E} = \mathbf{L}_p \mathbf{S}_C \quad (3)$$

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:

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

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:

$$O_m = \text{median}(O) \quad (5)$$

and the MAD between these two quantities is performed:

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

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:

$$O_{\text{eof}} = \left| \frac{O - O_m}{\delta} \right| \quad (7)$$

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 $O_{\text{prox}} = 3$ for those pixels and $O_{\text{prox}} = 0$ 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:

$$\delta_{\text{median}} = \text{mad}(X^o) \quad (8)$$

and an estimation of outlier pixels can be therefore obtained:

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

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

$$O_{\text{final}} = w_{\text{eof}} O_{\text{eof}} + w_{\text{prox}} O_{\text{prox}} + w_{\text{median}} O_{\text{median}} \quad (10)$$

where:

$$w_{\text{mad}} + w_{\text{prox}} + w_{\text{median}} = 1 \quad (11)$$

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

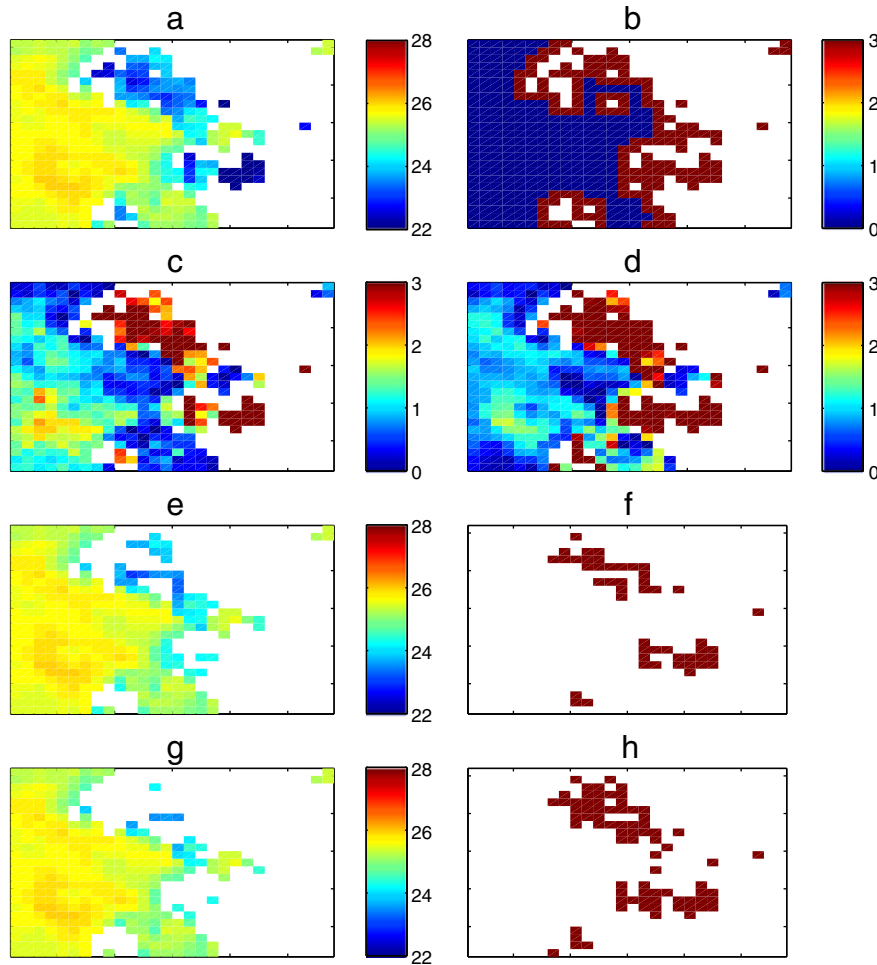


Fig. 2. SST detail (°C) and outlier tests in the small domain of the western Mediterranean Sea shown in Fig. 1. Panel a: original data; panel b: proximity test; panel c: EOF test; panel d: local median test; panel e: SST withoutliers removed, applying a threshold of 3; panel f: outliers detected with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

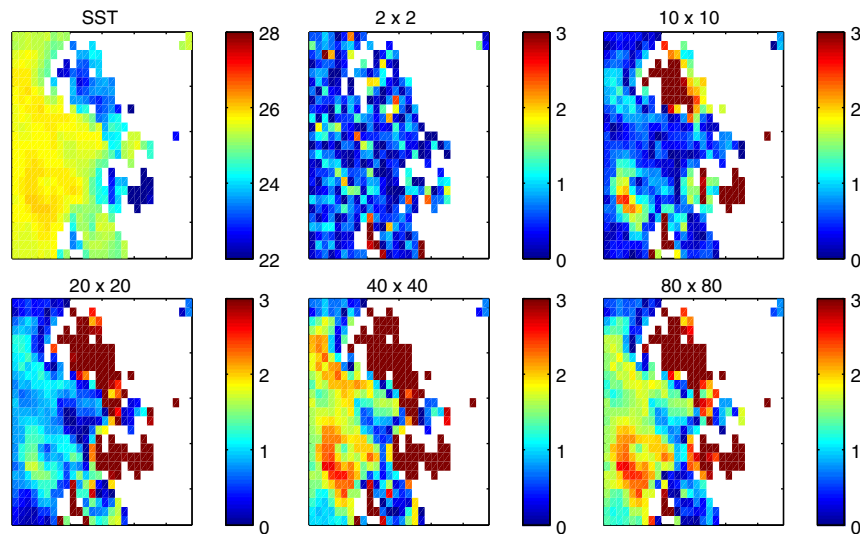


Fig. 3. SST detail ($^{\circ}\text{C}$) and effect of window size in the median test. Windows of 2×2 , 10×10 , 20×20 , 40×40 and 80×80 pixels were used.

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 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

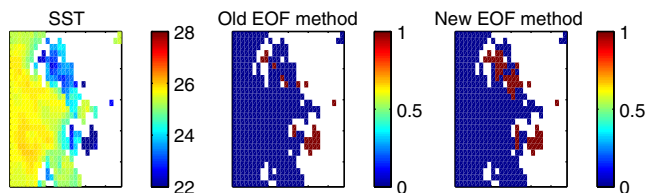


Fig. 4. Outliers detected using the EOF test as described in Eq. (1) (centre panel) and outliers detected using the new approach suggested in this study (Eq. 7, right panel). Note that in the new approach only the EOF test is applied, and not the proximity and median tests. The original SST ($^{\circ}\text{C}$) 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 SeaWiFS chlorophyll-a concentration data mentioned in Section 2 were used. These data are 8-day composites, and

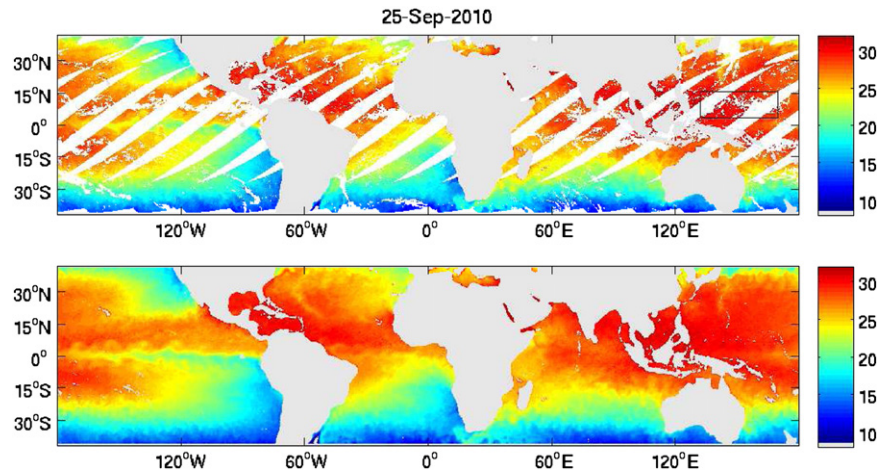


Fig. 5. TMI SST (°C) missing data reconstruction for 25 September 2010. Top panel: original data. Bottom panel: DINEOF reconstruction.

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

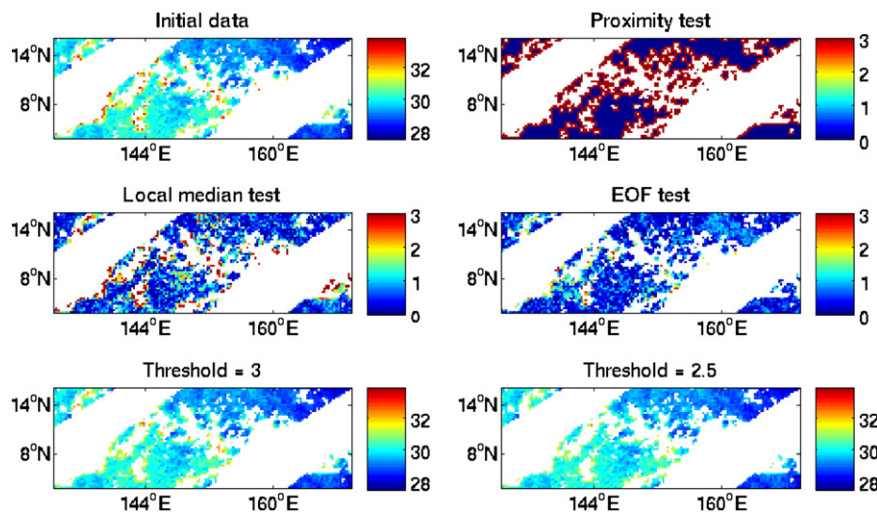


Fig. 6. SST detail (°C) and outlier tests in the domain shown by a black rectangle in Fig. 5. The original data are shown in the top left panel, and the data without outliers are shown in the two bottom panels, using two thresholds (2.5 and 3) when combining the individual outlier tests. The individual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

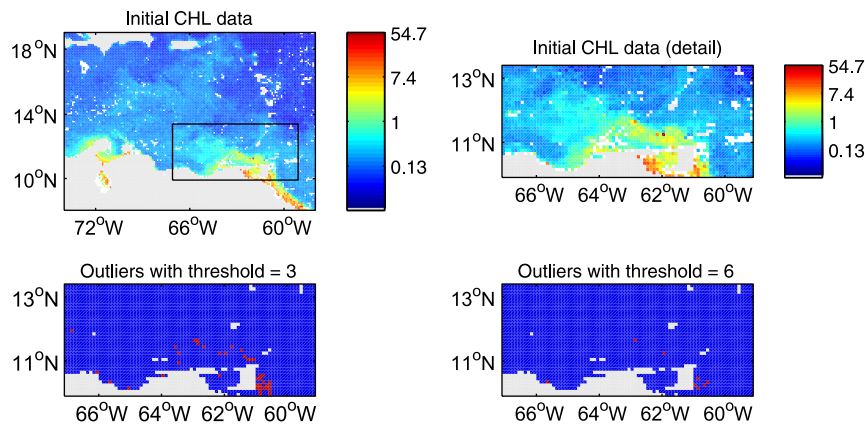


Fig. 7. Top left panel: 8-day composite of chlorophyll-a concentration (in mg/m^3) in the Caribbean Sea, corresponding to the period 9 to 16 January 2004. Top right panel: detail of the same data in the southeast of the domain. Bottom left: outliers detected using a threshold of 3, and (bottom right 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 13m SST data 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.
- Esaías, 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, space-time 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.