OCEAN EDR CAL VAL
Estimating Satellite Products Uncertainty Using *In Situ* Data
(Delivery n. 1 for PO # N00173-11-P-0039)

STUDY OF COVARIANCE AS A FUNCTION OF CLOUD COVER

Alberto Alvarez, Giuliana Pennucci and Charles Trees
NURC – NATO Undersea Research Centre

Summary: *This report is the first delivery of the project Calibration/Validation (Cal/Val) Plan for Oceans Environmental Data Records (ERDS) for the Visible Infrared Spectrometer (VIIRS) sensor on the Preparatory Project (NPP) and National Polar Orbiting Operational Environmental Satellite System (NPOESS).*
The focus of the analysis presented here if the evaluation of the covariance field in SST and ocean color images as a function of cloud cover by excluding data from the analyses of clear images, as if clouds were present.
This analysis is required to define the maximum cloud coverage percentage for proper covariance estimation.

Keywords: *sea surface temperature, ocean color, covariance.*

1. Background

One of goals of the to the VIIRS Cal/Val project is to find the adequate number and allocation of *in situ* observational resources to improve satellite products by reducing observational error. This requires the implementation of methodologies to merge remote sensed images with *in situ* data. To accomplish this method was developed to decompose the observed fields into spatially varying mean and Gaussian residual fields.
The approach is based on an isotropic diffusion operator that preserves frontal structures:

$$\psi(\vec{x},t) = \psi(0,t) + \int_0^t \nabla \cdot \left(e^{-\left(\frac{\|\nabla\psi\|}{K}\right)^2} \nabla\psi \right) dt \qquad (1),$$

where $\psi(\vec{x},t)$ represents the oceanographic measured field from the satellite at a given time t and with a known observation error. Using an optimization process that finds the Gaussian residual field ($\varepsilon(\vec{x})$), the covariance of $\varepsilon(\vec{x})$ is assumed to be a function of the relative distance between two points:

$$\left\langle \varepsilon(\vec{x})\varepsilon(\vec{x}') \right\rangle = c_\varepsilon(\|\vec{x} - \vec{x}'\|) \qquad (2).$$

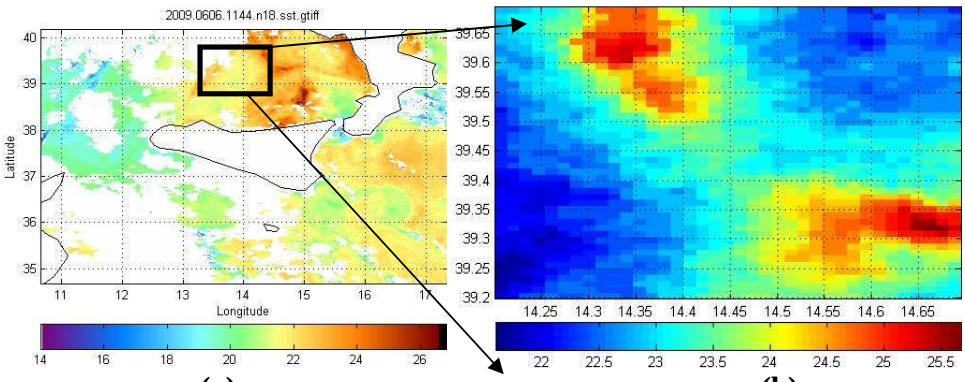
2. The Covariance analysis (spatial domain)

The focus of this report is the evaluation of the covariance field (Equation 2) in satellite imagery (such as sea surface temperature - SST or ocean color data) as a function of cloud cover. In other words, is the covariance of remote sensed data dependent on cloud cover (as assumed) and if so, can it be represented as a function of cloud size and/or position?
For this purpose a few satellite acquisitions (from both AVHRR and MODIS) have been selected and a set of clear-free areas have been identified to evaluate the "clear-covariance" approach. The data will then be corrupted by simulating of the presence of clouds and then the resulted covariance will be compared with the "clear-covariance".
In the next paragraph the results from the iteration of this procedure will be presented.

3. Results using AVHRR data (sea surface temperature, SST)

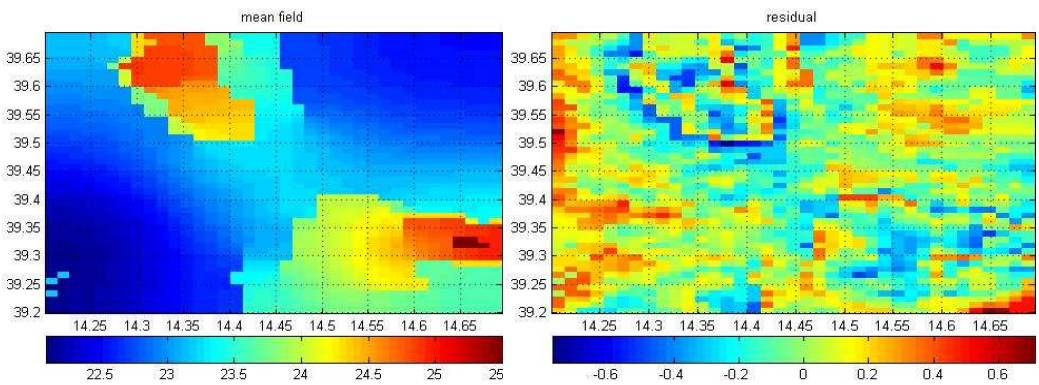
As discussed above how does the covariance analysis change with respect to cloud cover. For this purpose a simulated set of clouds with random position and varying pixel size have been added to the clear image and then the covariance results with the simulated clouds was compared with the results obtained in the cloud free image, assuming the following steps.

- **STEP 1.** A clear area (cloud free) have been identified in a satellite image (SST from AVHRR), as presented in Figure 1a and b.
- **STEP 2.** The selected area was decomposed (using equation 1) into a spatially varying mean (Figure 2a) and Gaussian Noise fields (Figure 2b).



(a) **(b)**

Figure 1: Sea surface temperature image acquired on the 6th of June 2009 during an AVHRR overpass of Sicily, Italy (a). The black box overlays the clear study area of ~54 by 53 kms of cloud-free data (b).



(a) **(b)**

Figure 2: Mean field (a) and Gaussian Residual field $\varepsilon(\vec{x})$ (b) of the selected area. $\varepsilon(\vec{x})$ is a Gaussian realization with a mean of $\mu_\varepsilon = 0.0047$ (units?) and a standard deviation of $\sigma_\varepsilon = 0.2047$ (units?).

- **STEP 3.** Using Equation 2, the covariance matrix was computed and fitted to a covariance model (Figure 3), where the value at zero spatial lag is obtained by an exponential fitting of the obtained curve. In this case, the fitting law is given by:

$$C_\varepsilon(\|\vec{x}-\vec{x}'\|) = 0.0238 \cdot e^{\frac{-\|\vec{x}-\vec{x}'\|}{3476}} \qquad (3).$$

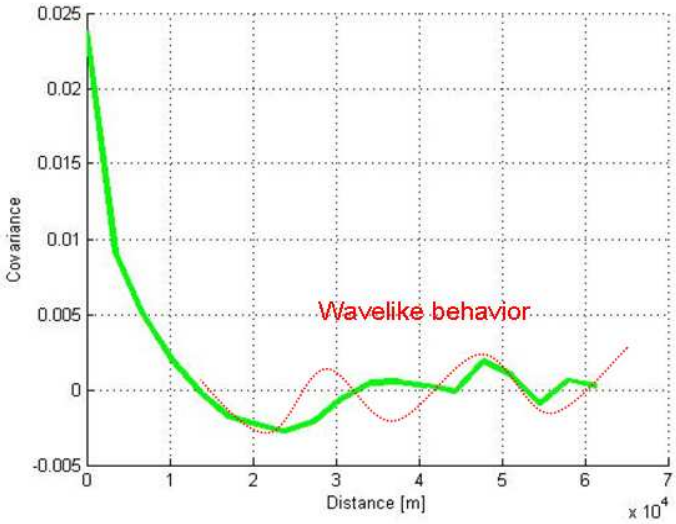


Figure 3: Covariance function in terms of the spatial lag obtained on the clear SST area of Figure 2. The wave like behavior starts around $1.5x10^4$ meters (~15 kms of distance).

It's important to note that the observed "wavelike behavior" (for distances bigger than $1.5x10^4$ kms) is an artifact of the longitudinally banded structure in the original image (that depends of the sensor error).

- **STEP 4.** Using the retrieved Covariance (C_{CLEAR}) in Figure 3 as the best reference for comparison, several clouds (block of NaNs) have been simulated in different positions (as schematized in Figure 4) and the correspondent covariances have been evaluated and compared with C_{CLEAR} (Figure 5).

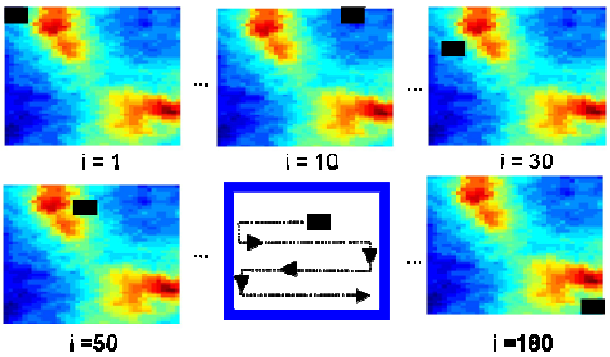


Figure 4: Simulation of a cloud of 3x3 pixels (~1% cloud coverage) in different positions (a shift of 3 pixels from upper left to lower right) on the original clear image (47 x 37 pixel). This produces a total of 180 images.

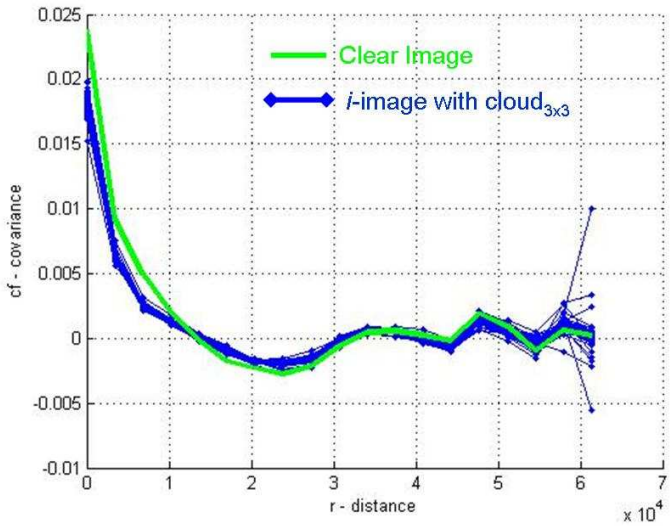


Figure 5: Covariances of the 180 images (with cloud coverage ~=1%). I would plot the average covariance of the 180 images as one line with the Std error/deviation at selected distances. Also what is the error in the clear image at selected distances? Does it change as a function of distance? This should also be plotted, if possible.

- **STEP 5.** Figure 5 shows that, as expected, the covariance is cloud-dependent in the sense that is dependent on the amount of information (image/structures). To generalize this result, the procedure in STEP 4 has been iterated increasing the box-cloud size from 3x3 pixels (~1% cloud coverage) to 35x35 pixels (~70% cloud coverage) that corresponds to 539 images. For each image the covariance has been computed and the results displayed in Figure 6.

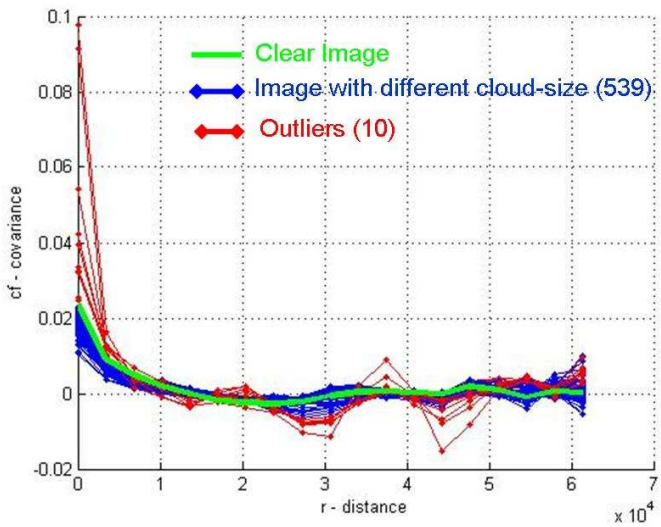


Figure 6: Covariances of the 539 images (with cloud coverage from 1% to 70%). The 98% of the realization remain under the green line (clear case), as expected.

The results in Figure 6 confirm the covariance dependence on the cloud-size (loss of information) in the 98% of the realization simulated. The remaining 2% is represented of 10 realizations that can be considered as outliers, as described in the following paragraph.

3.1 Outlier Analysis

As written before, our approach is based on the assumption that the measured oceanographic filed (the image) can be decomposed into a spatially varying mean and a Gaussian residual fields, ($\varepsilon(\vec{x})$)..

For these reasons, the optimization process that looks for a Gaussian residual field $\varepsilon(\vec{x})$ has been checked by applying some standard goodness-of-fit tests. Figure 7 shows the Kurtosis, Skewness and Mean values of the 539 realizations, the red points represent the outliers that we want to analyze.

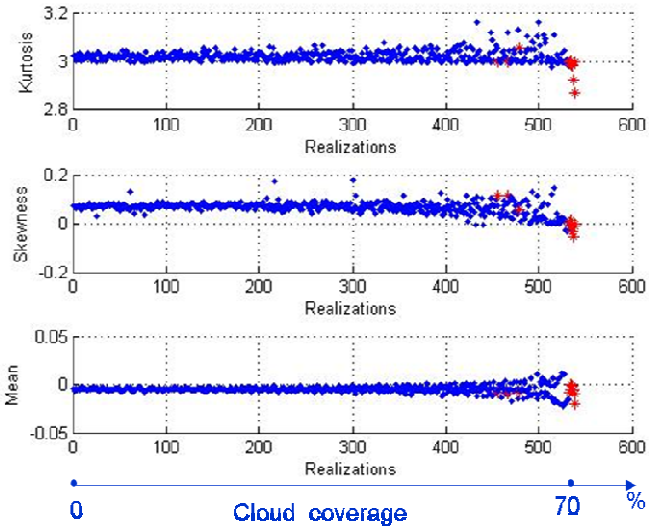


Figure 7: Kurtosis, Skewness and Mean analysis.

The analysis of Figure 7 shows that the outliers turn up when the cloud coverage is bigger than 60-65% of the image. In these cases, the decomposition of the observed field is not well represented as a spatially varying mean and a Gaussian Noise. Figure 8 confirmed this in the histograms that compared a case of good realization (cloud coverage lesser than 10%) with one outlier.

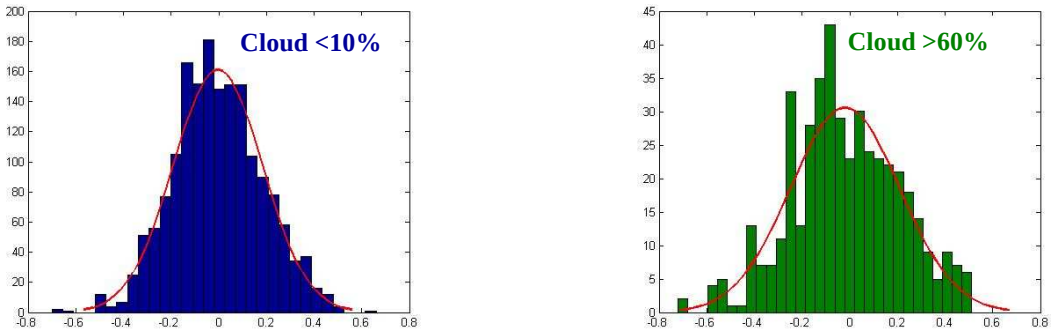


Figure 8: Histogram of $\varepsilon(\vec{x})$ for two realizations of the SST analysis. In the left side a

realization with cloud percentage lesser than 10%, while the right side shows a realization with coverage bigger than 60%.

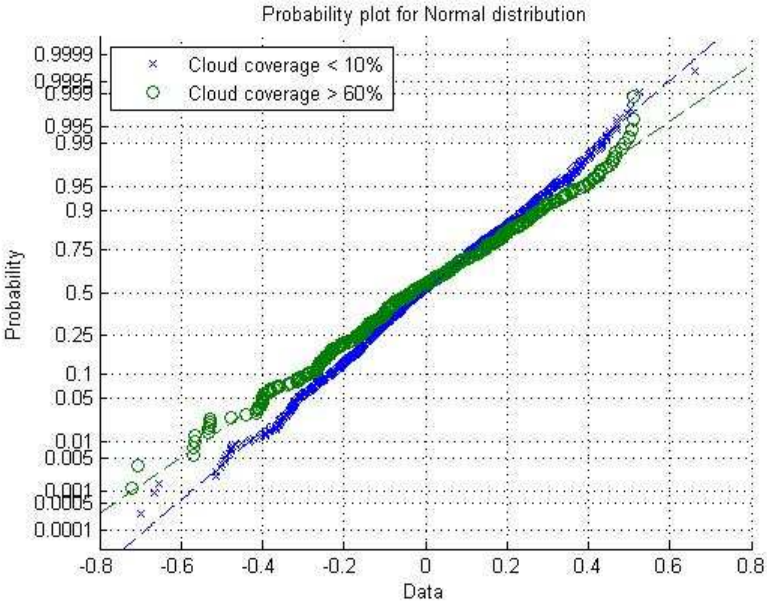


Figure 9: Normal probability plot of the distribution showed in Figure 8.

The results above suggest that a first check can be made performing a goodness-of-fit test of the $\varepsilon(\vec{x})$ distribution in the sense that the covariance analysis have to be implemented just on the realization that satisfy the normal distribution condition. To have a confirmation of that, we have applied the covariance analysis on other satellite data (as presented in the following paragraph).

4. Results using MODIS data (water leaving radiances, nLw)

Figure 10 shows a clear area identified from a MODIS image that was acquired on the 27th of May 2005. The covariance analysis was implemented with the methodology described in paragraph 3 on several MODIS images; results presented here are from the water-leaving radiance at 532 nm.

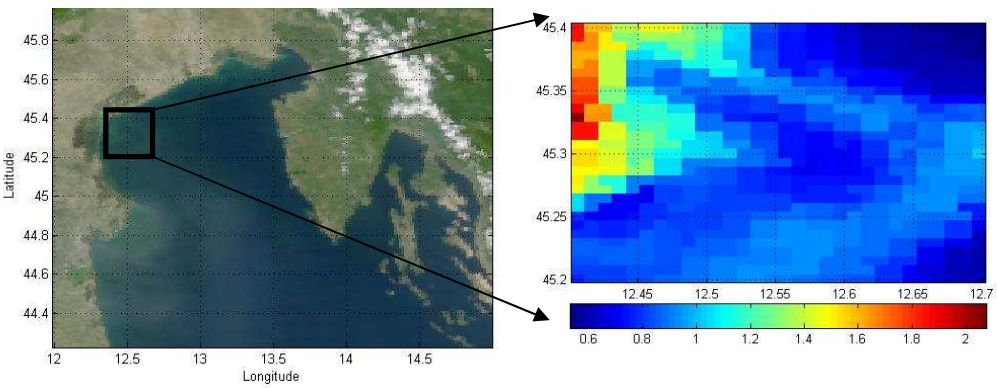


Figure 10: Normalized water leaving radiance (nLw) at 532 nm from a MODIS image that was acquired on the 27th of May 2005. The black box overlays the cloud-free study area of ~29 by 30 kilometers.

As done for the SST data, the Monte Carlo simulation has been implemented on the MODIS area and the covariances have been computed (as showed in Figure 11).

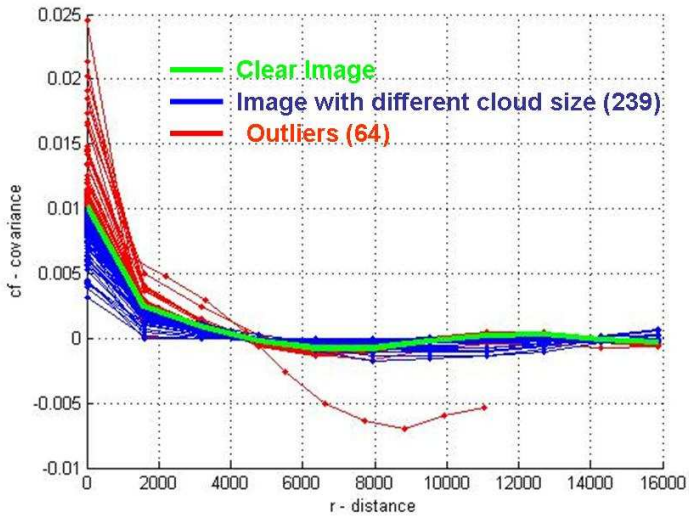


Figure 11: Covariances of the 539 images (with cloud coverage from 1% to 70%). The 74% of the realization remain under the green line (cloud-free case).

4.1 Outlier Analysis

Expecting a behavior similar to the previous SST case, a first check was performed evaluating the Kurtosis, Skewness and Mean values of the 539 realizations. However, as seen from the following figure, it seems that in this case outliers aren't related to the cloud coverage.

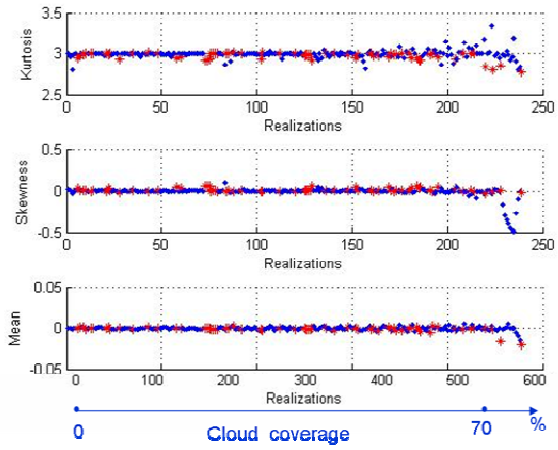
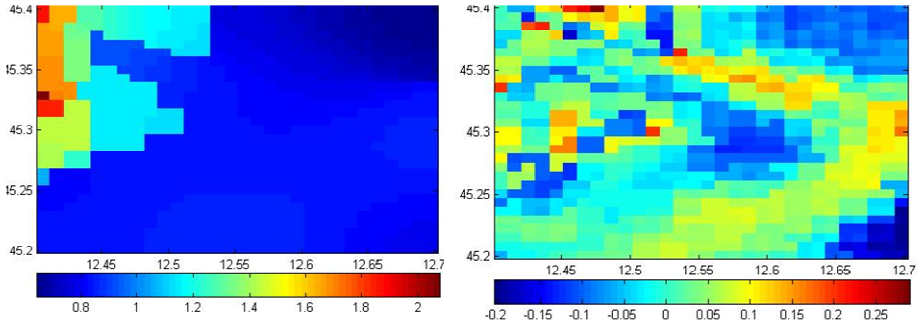


Figure 12: Kurtosis, Skewness and Mean analysis.

This situation can be explained if we take into account the Bayesian concept[1]. Our covariance analysis depends on the results retrieved applying the anisotropic diffuse operator (Equation 1) that should provide the "best" estimate of our oceanographic field into a spatially varying mean and a Gaussian residual field.
It's straightforward that the solution depends on the starting data set (an image as a random variable), in particular it depends on the size, amplitude and number of structures involved in the analysis. In particular for the MODIS case presented, the number of structures is less with respect to the SST case discussed in Paragraph 3. It is possible to compare the SST mean field and residual field of Figure 2 with those retrieved from the MODIS case (see Figure 13).



(a) **(b)**

Figure 13: Mean field (a) and Gaussian Residual field $\varepsilon(\vec{x})$ (b) of the MODIS cloud-free area.

The results showed above together with those retrieved in Paragraph 3 suggest that there is a need to define an additional "quality index" that takes into account of this limitation, as described in detail in the next paragraph.

5. Definition of "quality indices" to classify the covariance analysis to respect clouds

As suggested from the results showed in Paragraphs 3 and 4, to better qualify the covariance field in SST and ocean color images as a function of cloud cover, there is a requirement to define a few more indices that relate the quality of the covariance analysis with the cloud cover percentage.
For this purpose three indices have been defined:

- Number of structures within the oceanographic field (amount of information);
- Goodness of the mean filed estimation;
- Goodness of the residual field realization (the $\varepsilon(\vec{x})$ distribution is Gaussian?);

The first takes into account the particular statistical realization considered and gives an idea of the limit of the covariance evaluation, while the last two are needed to check the goodness of the statistical model retrieved using the anisotropic diffusion operator in Equation 1 (a sort of goodness-of-fit test).

5.1 Index of Structures

As written in Paragraph 4, the covariance solution depends on the starting data set (an image as a random variable), in particular it depends on the size, amplitude and number of structures involved in the analysis.
To quantify this concept an index of structure can be defined as the ratio between the amount of information and the de-correlation length (that represents the space distance over which neighboring data points are correlated at some specified confidence level):

$$I_S = \frac{\sqrt{\text{Size of the clear area}}}{\text{decorrelation distance}} \qquad (4)$$

The figure below shows the values of Is that have been evaluated for each realization of the MODIS Monte-Carlo simulation described above. As expected, the Is value decreases when the cloud size increases (lost of information); excluding the two outliers (red circles) that have to be investigated.

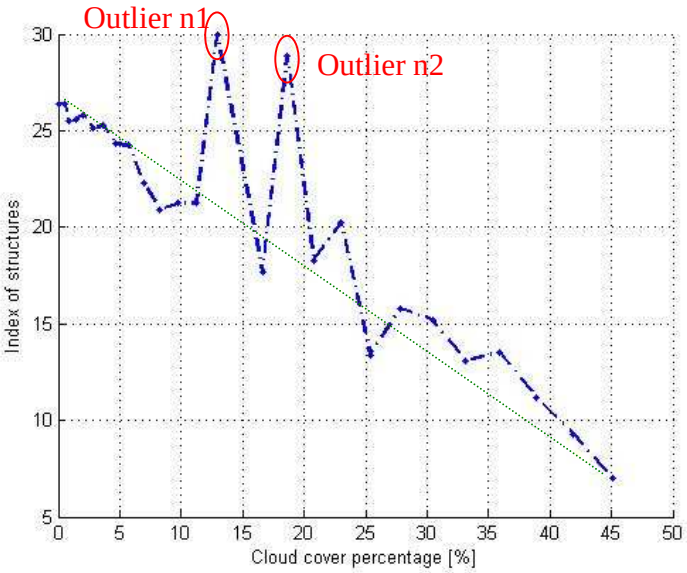
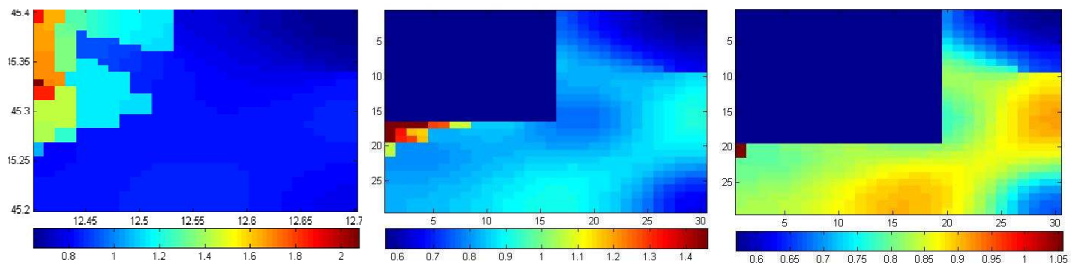


Figure 14: plot of the index of structure of the MODIS simulation. The value 0% represents the number of structure of the starting image (cloud-free).

The reason of these outliers can see as the limit of the model implementation, indeed both realizations represent a singular case of cloud position, as is possible to verify from Figure 15b and 15c where the simulated cloud covers all the upper left part of the image. If the main front/structure area is deleted the analysis focus on the remaining background (on which can be founded more structure) that can be considered as a totally different realization.



(a) **(b)** **(c)**

Figure 15: mean field of the cloud-free area (a) and on the two outliers (b and c) cases that have been identified using the Is analysis.

It's important to note that the Is index analysis presented here, is independent on the goodness of the residual field realization, indeed, for both the outliers, the distribution of the retrieved residual field results a Gaussian.

5.2 Index for the goodness of the residual filed estimation

As introduced above, another possible reason of outliers due to a bad statistical model realization, id the failure of the anisotropic diffuse operator (equation 1) that results in non-Gaussian of the residual field.

To avoid the consideration of these cases we have decide to check mean, Kurtosis and Skewness of each realizations and to apply on them a goodness fit of test to reject the bad case. In particular we apply the Lilliefors test that represents an adaptation of the Kolmogorov–Smirnov test [2].

6. Results introducing the quality indices to avoid outliers

As presented in the previous pages, the performance of our analysis is closely related to the goodness of the statistical model realization that has to verify some essential condition. For the intrinsic characteristic of this statistical procedure, it is no possible to provide a standard procedure that allows to always retrieving a good realization and consequently good estimation of the covariance. However the introduction of the quality indices presented above can be considered as a good start to identify some bad realization and to decrease the number of outliers.

In Figure 16a and 16b are displayed the same results of the SST (a) and MODIS (b) simulations presented in Figure 6 and Figure 11 (respectively) after a preliminary check that consider only the realizations that respect the quality indices criteria described above. As expected the outliers number is considerably reduced ~22% (a: from 10 to 3, b: from 64 to 10).

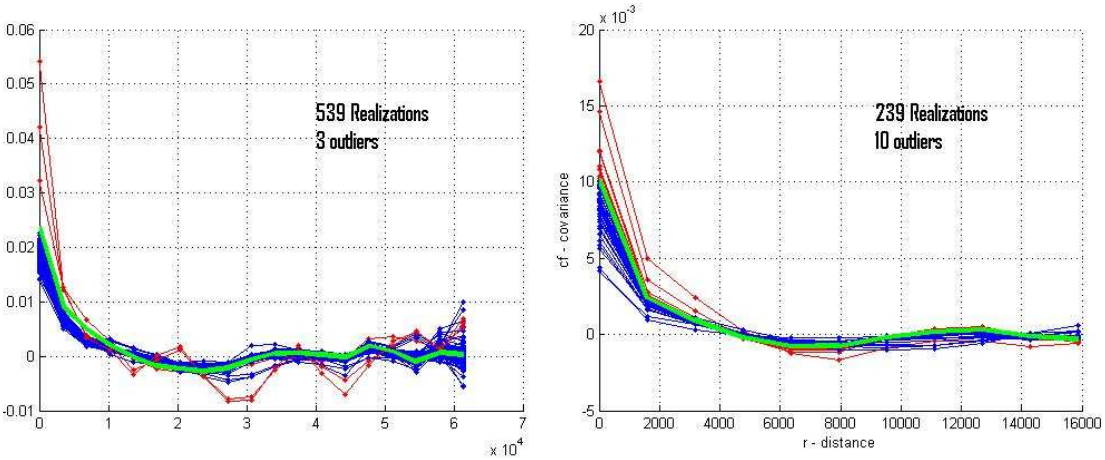


Figure 16: Covariances of the two SST (a) and MODIS (b) simulations after the preliminary check to consider only the realizations with good quality indices. As expected the outliers number is considerably reduced ~22%.

7. Conclusion

This report resumes the first delivery of the project Calibration/Validation (cal/val) Plan for Oceans Environmental Data Records (ERDS) for the Visible Infrared Spectrometer (VIIRS) sensor on the Preparatory Project (NPP) and National Polar Orbiting Operational Environmental Satellite System (NPOESS).

In particular the focus of the analysis presented here if the evaluation of the covariance field in SST and ocean color images as a function of cloud cover.

The results show that the covariance analysis is directly dependent on clouds size/position, in the sense that the introduction of clouds causes a loss of information and decrease the correlation length (as expected). However, the validity of this consideration is directly related to the goodness of the realization of interest. To check that we have defined two quality indices those relate the quality of the covariance analysis with the cloud coverage percentage. The first one is an index that takes into account of the number of structures in the oceanographic field (~amount of information); while the second take into account of the goodness of the residual field realization (the evaluation of $\varepsilon(\vec{x})$ using equation 1).

8. References

[1] Lewis, J.M, Lakshmivarahan, S. and S. K. Dhall, Dynamic data assimilation: a least squares approach, Cambridge University Press, 2006.

[2] Gary S. E. Lagerloef and R. Bernstein, “Empirical Orthogonal function Analysis of advanced very high resolution Radiometer Surface Temperature Patterns in Santa Barbara Channel”, Journal of geophysical Research, Vol. 93, no. C6, pages 68636873, June 15,1988.

[2] Lilliefors, H.W., "On the Komogorov-Smirnov test for normality with mean and variance unknown," Journal of the American Statistical Association, vol. 62, 1967, pp. 399-402.