	<p><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> i</p>
---	--	---

## DIMITRI Algorithm Theoretical Basis Document [01]

# Automated Cloud Screening in DIMITRI 3.0




**ESA contract:** 4000106294

**ARGANS Reference:** 003-013: MO-SCI-ARG-TN-004

**Date:** 28<sup>th</sup> May 2014


**Version:** 1.0

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a  <b>Revision:</b> 1.0  <b>Date:</b> 28/05/2014  <b>Page:</b> ii</p>
---	---	---

## Authors


NAME	AFFILIATION	CONTACT
K. Barker	ARGANS Ltd, UK	<a href="mailto:kbarker@argans.co.uk">kbarker@argans.co.uk</a>
D. Marrable	ARGANS Ltd, UK	<a href="mailto:dmarrable@argans.co.uk">dmarrable@argans.co.uk</a>
C. Mazeran	Solvo, France	<a href="mailto:constant.mazeran@solvo.fr">constant.mazeran@solvo.fr</a>

## Acknowledgements

	<b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening	<b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> iii
---	--	---

## TABLE OF CONTENTS

<b>1</b>	<b>Introduction .....</b>	<b>1</b>
1.1	Scope of this ATBD .....	1
1.2	DIMITRI .....	1
<b>2</b>	<b>Cloud screening in DIMITRI_v2.0.....</b>	<b>4</b>
2.1	Landsat ACCA Algorithm .....	4
2.2	GlobCarbon .....	4
2.3	VGT Operational.....	5
2.4	Cloud screening performance .....	6
2.5	Potential areas for improvement.....	17
<b>3</b>	<b>Improved Automated Cloud Screening Implementation in DIMITRI Version 3.0.....</b>	<b>19</b>
3.1	Overview of approaches .....	19
3.2	Method 1: Spatial Scale Variability (SSV) .....	19
3.3	Method 2: BRDF Variability Threshold BRDFVT .....	20
3.4	Output files generated by the SSV cloud screening.....	21
3.5	Output files generated by the BRDF cloud screening .....	22
3.6	DIMITRI modules.....	22
3.6.1	New RGB generation.....	22
3.6.2	SSV cloud screening modules .....	22
3.6.3	BRDF cloud screening modules.....	23
3.7	HMI updates and user options.....	24
3.7.1	Main window .....	24
3.7.2	SSV cloud screening options .....	24
3.7.3	BRDF cloud screening options .....	26
<b>4</b>	<b>Results of cloud screening method implementations .....</b>	<b>28</b>
4.1	Methodology results .....	28
4.1.1	Training datasets.....	28
4.2	MERIS SSV cloud screening over Libya4.....	29
4.2.1	MERIS SSV cloud screening over SPG .....	35
4.2.2	AATSR SSV cloud screening over Libya4 .....	38
4.2.3	MERIS BRDF cloud screening over Libya4.....	41

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> iv</p>
---	---	--

4.2.4	AATSR BRDF cloud screening over Libya4 .....	44
<b>5</b>	<b>Discussion and Conclusions.....</b>	<b>48</b>
<b>6</b>	<b>Appendix.....</b>	<b>1</b>
6.1	Cloud screening statistics .....	1

## List of Figures

Figure 1:	DIMITRI v2.0 screenshot.....	3
Figure 2	Percentage of images where the auto cloud screening agrees with the manual cloud screening.....	9
Figure 3	Percentage of images where the auto cloud screening agrees with the manual cloud screening.....	9
Figure 4	Percentage of images where the auto cloud screening agrees with the manual cloud screening at DomeC.....	10
Figure 5	Percentage of images where the auto cloud screening agrees with the manual cloud screening at Libya4 .....	10
Figure 6	Percentage of images where the auto cloud screening agrees with the manual cloud screening at SIO .....	11
Figure 7	Percentage of images where the auto cloud screening agrees with the manual cloud screening at SPG .....	11
Figure 8	Percentage of images where the auto cloud screening agrees with the manual cloud screening at TuzGulu.....	12
Figure 9	Percentage of images where the auto cloud screening agrees with the manual cloud screening at Uyuni .....	12
Figure 10	The performance of the auto cloud screening of all of the sensors combined at each site.....	13
Figure 11	The number of manually cloud screened MERIS images at Amazon with respect to date. Each bar represents approximately 1 week. ....	13
Figure 12	The number of manually cloud screened MERIS images at BOUSSOLE with respect to date. Each bar represents approximately 1 week. ....	14
Figure 13	The number of manually cloud screened MERIS images at DomeC with respect to date. Each bar represents approximately 1 week. ....	14
Figure 14	The number of manually cloud screened MERIS images at Libya4 with respect to date. Each bar represents approximately 1 week. ....	15
Figure 15	The number of manually cloud screened MERIS images at SIO with respect to date. Each bar represents approximately 1 week. ....	15
Figure 16	The number of manually cloud screened MERIS images at SPG with respect to date. Each bar represents approximately 1 week. ....	16


	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> v</p>
---	---	---

Figure 17 The number of manually cloud screened MERIS images at TuzGulu with respect to date. Each bar represents approximately 1 week. ....	16
Figure 18 The number of manually cloud screened MERIS images at Uyuni with respect to date. Each bar represents approximately 1 week. ....	17
Figure 19 DIMITRI main window.....	24
Figure 20 DIMITRI SSV cloud screening window .....	25
Figure 21 DIMITRI BRDF cloud screening window.....	27
Figure 22 MERIS Libya4 20020507 RGB (top) and SSV cloud detection at 412 nm (bottom) .....	30
Figure 23 MERIS Libya4 20020612 RGB (top) and SSV cloud detection at 412 nm (bottom) .....	31
Figure 24 Libya4 20020504 RGB (top) and SSV cloud detection at 412 nm (bottom) .....	32
Figure 25 MERIS Libya4 20020729 RGB (top) and SSV cloud detection at 412 nm (middle) and at 865 nm (bottom).....	33
Figure 26 MERIS Libya4 20060129 former RGB (top), new RGB (middle) and SSV cloud detection at 412 nm (bottom).....	34
Figure 27 SPG 20020723 RGB (top) and SSV cloud detection at 412 nm (bottom) .....	35
Figure 28 SPG 20020615 RGB (top) and SSV cloud detection at 412 nm (bottom) .....	36
Figure 29 SPG 200200616 RGB (top) and SSV cloud detection at 412 nm (bottom) .....	37
Figure 30 SPG 20020521 RGB (top) and SSV cloud detection at 412 nm (bottom) .....	38
Figure 31 AATSR 20020723 RGB (top) and SSV cloud detection at 555 nm (bottom) .....	39
Figure 32 AATSR 20021001 RGB (top) and SSV cloud detection at 555 nm (bottom) .....	40
Figure 33 AATSR 20020928 RGB (top) and SSV cloud detection at 555 nm (bottom) .....	41
Figure 34 MERIS BRDF cloud screening analysis over Libya4 .....	42
Figure 35 MERIS cloud screening performance over Libya4 for the DIMITRI nominal implementation (top) and BRDF method (bottom). The nominal screening is considered as clear when AUTO_CS=0 and as cloudy when AUTO_CS>0.....	43
Figure 36 Examples of MERIS wrong manual screening over Libya4 (20060129, 20040203, 20090309 from top to bottom). While the MANUAL_CS is set to 0 in the DB, small clouds are visible in the ROI. ....	44
Figure 37 AATSR BRDF cloud screening analysis over Libya4.....	45
Figure 38 AATSR cloud screening performance over Libya4 for the DIMITRI nominal implementation (top) and BRDF method (bottom). The nominal screening is considered as clear when AUTO_CS=0 and as cloudy when AUTO_CS>0.....	46
Figure 39 Examples of AATSR wrong manual screening over Libya4 (left 20030622 and right 20050607). While the MANUAL_CS is set to 0 in the DB, small clouds are visible in the ROI. ....	47

## List of Tables

Table 1: Sensors and sites included in the DIMITRI v2.0 database .....	3
Table 2 Current database statistics of Amazon .....	6


	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> vi</p>
---	---	--

Table 3 Current database statistics of Boussole ..... 6

Table 4 Current database statistics of DomeC..... 7

Table 5 Current database statistics of Libya4 ..... 7

Table 6 Current database statistics of SIO ..... 7


Table 7 Current database statistics of SPG ..... 8

Table 8 Current database statistics of TuzGulu ..... 8

Table 9 Current database statistics of Uyuni ..... 8

Table 10 Number of observations in Libya4 training datasets used hereafter, for 3 classes ..... 28

Table 11 Number of observations in SPG training datasets used hereafter, for 3 classes ..... 28

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 1</p>
---	---	---

# 1 Introduction

## 1.1 Scope of this ATBD

Under ESA contract 4000106294 ('Earth Observation Multi-Mission Phase-E2 Operational Calibration: assessment of enhanced and new methodologies, technical procedures and system scenarios'; hereafter referred to as 'MOSAEC'), there was a requirement to improve the automated cloud screening procedure that is implemented in DIMITRI.

This document describes briefly the current algorithms implemented and details the methodologies which have augmented and improved these current methodologies.


This ATBD will:

- Describe the principles of this method;
- Describe what is needed for the implementation in DIMITRI
- Describe the updates made to DIMITRI Human Machine Interface (HMI) and how the user can use this methodology.

## 1.2 DIMITRI

The Database for Imaging Multi-Spectral Instruments and Tools for Radiometric Intercomparison (DIMITRI) is an open-source software giving gives users the capability of long term monitoring of instruments for systematic biases and calibration drift, with a database of L1b top of atmosphere radiance and reflectances from a number of optical medium resolution sensors.

DIMITRI comes with a suite of tools for comparison of the L1b radiance and reflectance values originating from various medium resolution sensors over a number of radiometrically homogenous and stable sites (Table 1) at TOA level, within the 400nm – 4µm wavelength range. The date range currently available is 2002 to 2012. DIMITRI's interface enables radiometric intercomparisons based on user-selection of a reference sensor, against which other sensors are compared. DIMITRI contains site reflectance averages and standard deviation (and number of valid pixels in the defined region of interest, or ROI), viewing and solar geometries and auxiliary and meteorology information where available; this allows extractions of windspeed and direction, surface pressure, humidity and ozone concentration from MERIS products, and water vapour and ozone concentration from VGT-2 products. Each observation is automatically assessed for cloud cover using a variety of different automated algorithms depending on the radiometric wavelengths available; manual cloud screening is also visually performed using product quicklooks to flag misclassified observations. DIMITRI also provides a platform for radiometric intercalibration from User defined matching parameters: geometric, temporal, cloud and ROI coverage. Other capabilities and functions include: product reader and data extraction routines, radiometric recalibration & bidirectional reflectance distribution function (BRDF) modelling, quicklook generation with ROI overlays, instrument spectral response comparison tool, VEGETATION simulation.

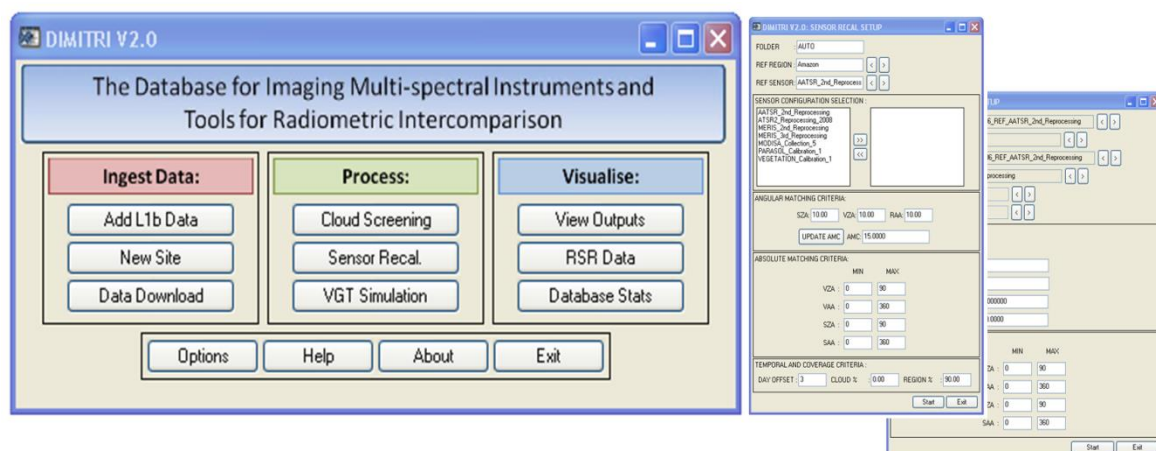
	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 2</p>
---	---	---

DIMITRI v2.0 has these two methodologies:

1. **Radiometric intercomparison based on angular and temporal matching**, based on the methodology of Bouvet (2006) and Bouvet et al (2007): Concomitant observations made under similar geometry and within a defined temporal window are intercompared at similar spectral bands.
2. **Radiometric intercomparison of VEGETATION simulated and actual observations**, making use of the ability to combine timeseries from all sensors into one “super sensor” and fitting a 3-parameter BRDF model to all observations to simulate TOA spectra of VEGETATION-2 (Bouvet, 2011).

DIMITRI v3.0 is evolved from DIMITRI v2.0 and has two additional methodologies and an improved automated cloud screening and cloud screening tool:

1. **Absolute vicarious calibration over Rayleigh Scattering**, based on the methodology of Hagolle et al (1999) and Vermote et al (1992) and utilising open ocean observations, to simulate molecular scattering (Rayleigh) in the visible and comparing against *the observed*  $\rho_{toa}$  to derive a calibration gain coefficient;
2. **Vicarious calibration over sunglint**, based on the methodology of Hagolle et al (2004); similar to Rayleigh scattering approach but accounting for sunglint reflectance contribution;
3. **Improved automated cloud screening**, exploiting the spatial homogeneity (smoothness) of validation sites when cloud free and applying a statistical approach utilising  $\sigma(\rho_{toa})$  over a ROI, and defining variability thresholds, such as dependence on wavelength and surface type.







	<b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening	<b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 3
---	--	---

Figure 1: DIMITRI v2.0 screenshot

Table 1: Sensors and sites included in the DIMITRI v2.0 database

SENSOR	SUPPLIER	SITE	SITE TYPE
AATSR (Envisat)	ESA	Salar de Uyuni, Bolivia	Salt lake
MERIS, 2 <sup>nd</sup> and 3 <sup>rd</sup> reprocessing (Envisat)	ESA	Libya-4, Libyan Desert	Desert
ATSR-2 (ERS-2)	ESA	Dome-Concordia (Dome-C), Antarctica	Snow
MODIS-A (Aqua)	NASA	Tuz Golu, Turkey	Salt Lake
POLDER-3 (Parasol)	CNES	Amazon Forest	Vegetation
VEGETATION-2 (SPOT5)	VITO	BOUSSOLE, Mediterranean Sea	Marine
		South Pacific Gyre (SPG)	Marine
		Southern Indian Ocean (SIO)	Marine

DIMITRI is maintained by ESA and ARGANS, and is freely available at <http://www.argans.co.uk/dimitri/>.

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 4</p>
---	---	---

## 2 Cloud screening in DIMITRI\_v2.0

In DIMITRI\_v2.0 the automated cloud screening cloudiness is expressed as a percentage of cover and stored in the DIMITRI database. The cloudiness can also be determined manually by the user.

As a potentially more sophisticated algorithm Landsat ACCA was applied wherever the bands supported by a particular instrument made its application possible. For instruments which lacked the thermal bands heavily utilised by the Landsat ACCA algorithm (i.e. MERIS & Parasol) the Globcarbon algorithm is used, and in the case of VEGETATION-2 which has very different band characteristics, the operational VGT cloud screening algorithm is applied.

### 2.1 Landsat ACCA Algorithm

The Landsat 7 ACCA algorithm is a two pass algorithm. The first pass consists of the application of a sequence of eight filters to each pixel of the image data. The filters classify the pixels as not cloud, ambiguous, ambiguous possibly desert, and ambiguous possibly snow using a variety of band ratios and composites. All pixels not classified by the time the final filter is reached are cloud. The second pass uses statistical analysis of the pixels Landsat band 6 values to determine whether ambiguous pixels are cloudy or not.

DIMITRI function:


CLOUD\_MODULE\_LCCA.pro

### 2.2 GlobCarbon

The Globcarbon algorithm consists of two filters which use band ratios to respectively discriminate cloud from snow and bright land from cloud. As applied to MERIS the algorithm uses a ratio of band 9/13 for the snow filter and band 2/10 for the bright land filter. All pixels identified as bright in the operational flags are input to these two filters. Bright pixels not classified as snow or bright land are classified as cloud. Ratio thresholds for each of the filters were taken from the Globcarbon paper. Early on in testing it was found that the algorithms effectiveness was substantially improved if operational flags for land and water were also used in the screening.

DIMITRI functions:

CLOUD\_MODULE\_GLOBCARBON.pro (MERIS)

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 5</p>
---	---	---

CLOUD\_MODULE\_GLOBCARBON\_P.pro (POLDER)

Fixed reflectance thresholds:

MERIS: Snow Threshold = 1.0    Cloud Threshold = 0.8

POLDER: Snow Threshold = 0.9386    Cloud Threshold = 1.736

## 2.3 VGT Operational

Reflectance information from all four bands was extracted over the validation sites and automatically cloud screened using the operational VEGETATION cloud screening algorithm (Lissens et al., 2000). SPOT VEGETATION provides TOA reflectance for 4 bands ranging from the visible to SWIR wavelengths; all radiometric data, as well as the auxiliary ozone and water vapour information is extracted from each data product.

DIMITRI function:

CLOUD\_MODULE\_VGT.pro

## 2.4 Cloud screening performance

To assess the cloud screening performance, it is important to assess the current state of the database by the following criteria: how many images have been 'manually cloud screened i.e. assessed by an expert users and how often the expert user's assessment agrees with the current cloud screening methodologies. The following section presents a way to visualise the state of the database in regards to cloudscreening.

Table 2 - Table 9 & Figure 2 - Figure 9 shows the total number of cloud screened images as well as the number of manually cloud screened images and the percentage of manually screened images that agree with the auto-cloud screening methods.

Table 2 Current database statistics of Amazon

Sensor	AutoCS	ManualCS	Total Agree	Total Agree (%)
AATSR	935	36	5	13.89%
ATSR2	172	23	19	82.61%
MERIS	3205	390	101	25.90%
MODISA	3047	1184	351	29.65%
PARASOL	1239	199	60	30.15%
VEGETATION	3907	170	60	35.29%
Total	12505	2002	596	29.77%

Table 3 Current database statistics of Boussole

Sensor	AutoCS	ManualCS	Total Agree	Total Agree (%)
AATSR	442	252	10	3.97%
ATSR2	243	67	3	4.48%
MERIS	4123	1538	57	3.71%
MODISA	4425	1980	238	12.02%
PARASOL	1737	957	153	15.99%
VEGETATION	6365	1384	78	5.64%
Total	17335	6178	539	8.72%

Table 4 Current database statistics of DomeC

Sensor	AutoCS	ManualCS	Total Agree	Total Agree (%)
AATSR	1545	814	552	67.81%
ATSR2	1104	259	188	72.59%
MERIS	6757	4392	175	3.98%
MODISA	9073	8033	5959	74.18%
PARASOL	4301	3754	1311	34.92%
VEGETATION	1273	0	0	0.00%
Total	24053	17252	8185	47.44%

Table 5 Current database statistics of Libya4

Sensor	AutoCS	ManualCS	Total Agree	Total Agree (%)
AATSR	1293	314	4	1.27%
ATSR2	210	49	1	2.04%
MERIS	3085	2003	155	7.74%
MODISA	4090	2094	247	11.80%
PARASOL	813	26	1	3.85%
VEGETATION	5734	1514	916	60.50%
Total	15225	6000	1324	22.07%

Table 6 Current database statistics of SIO

Sensor	AutoCS	ManualCS	Total Agree	Total Agree (%)
AATSR	734	43	2	4.65%
ATSR2	69	67	56	83.58%
MERIS	5007	531	299	56.31%
MODISA	3570	1215	327	26.91%
PARASOL	313	10	1	10.00%
VEGETATION	0	0	0	0.00%
Total	9693	1866	685	36.71%


	<b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening	<b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 8
---	--	---

Table 7 Current database statistics of SPG

Sensor	AutoCS	ManualCS	Total Agree	Total Agree (%)
AATSR	891	28	1	3.57%
ATSR2	68	29	23	79.31%
MERIS	4642	191	64	33.51%
MODISA	5242	1802	903	50.11%
PARASOL	502	37	2	5.41%
VEGETATION	0	0	0	0.00%
<b>Total</b>	<b>11345</b>	<b>2087</b>	<b>993</b>	<b>47.58%</b>

Table 8 Current database statistics of TuzGulu

Sensor	AutoCS	ManualCS	Total Agree	Total Agree (%)
AATSR	967	101	18	17.82%
ATSR2	225	36	10	27.78%
MERIS	3821	1235	223	18.06%
MODISA	4195	1256	137	10.91%
PARASOL	1090	824	57	6.92%
VEGETATION	5853	738	86	11.65%
<b>Total</b>	<b>16151</b>	<b>4190</b>	<b>531</b>	<b>12.67%</b>

Table 9 Current database statistics of Uyuni

Sensor	AutoCS	ManualCS	Total Agree	Total Agree (%)
AATSR	570	465	404	86.88%
ATSR2	223	115	107	93.04%
MERIS	2804	2069	1975	95.46%
MODISA	3607	2848	2163	75.95%
PARASOL	461	446	126	28.25%
VEGETATION	4240	312	254	81.41%
<b>Total</b>	<b>11905</b>	<b>6255</b>	<b>5029</b>	<b>80.40%</b>

The values in Tables 2 - 9 were calculated by comparing the automated cloud-screening data with the manually screened data. The forth column shows the number of data points in the database where the manual cloud-screening values agree with the automated values. Lastly the performance score shows the percentage of data points where the manual and automatic screening agree as a percentage of the total number of points that have been manually screened.

It should be noted that on re-inspection of the database for the discovery of potential training sets, there were many incorrectly manually classified images. Comparing whether the auto cloud screening agrees with manual cloud screening implies that the manually cloud screen images are in fact true. Therefore the numbers quoted in the previous tables should not be taken as a perfect metric of the performance of the current methodology.

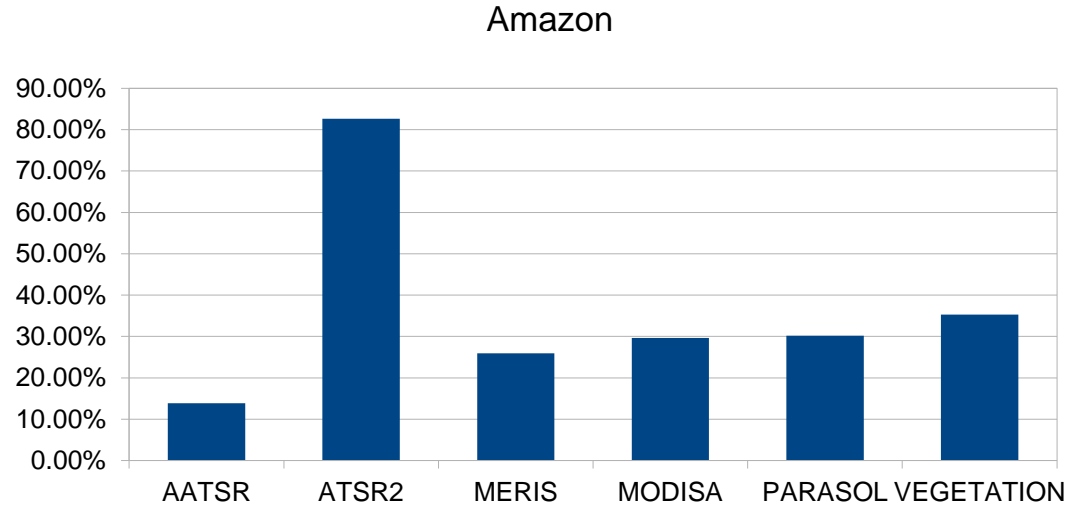


Figure 2 Percentage of images where the auto cloud screening agrees with the manual cloud screening

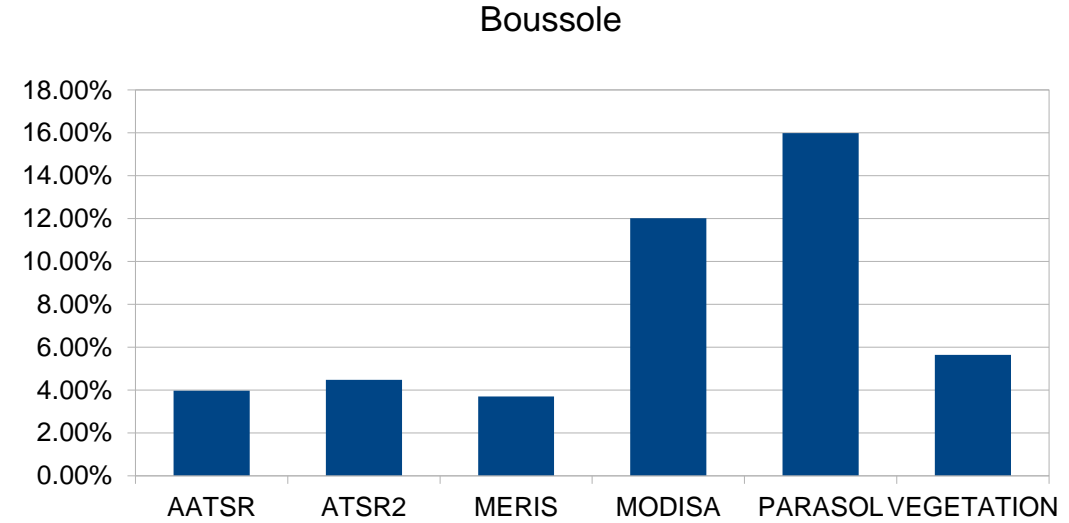


Figure 3 Percentage of images where the auto cloud screening agrees with the manual cloud screening

### DomeC

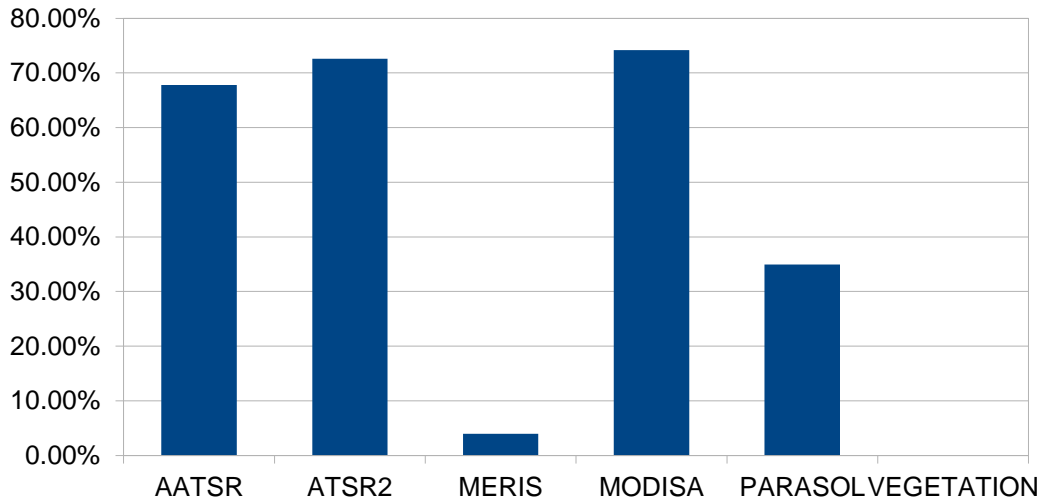


Figure 4 Percentage of images where the auto cloud screening agrees with the manual cloud screening at DomeC

### Libya4

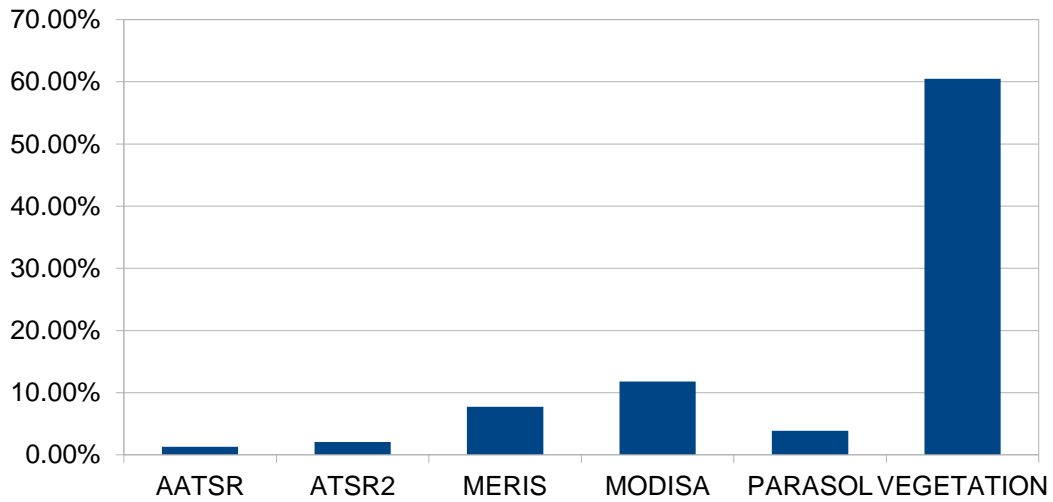


Figure 5 Percentage of images where the auto cloud screening agrees with the manual cloud screening at Libya4



### SIO

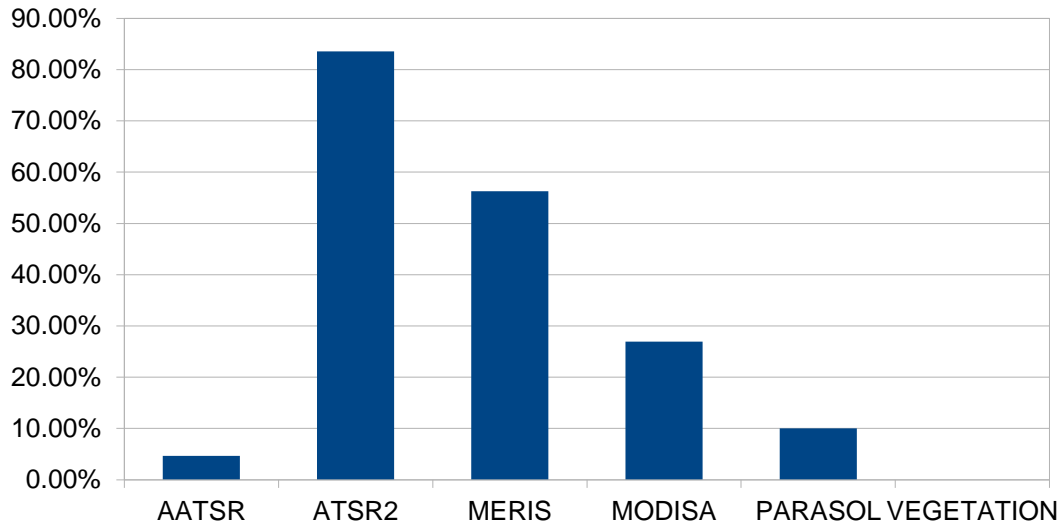


Figure 6 Percentage of images where the auto cloud screening agrees with the manual cloud screening at SIO

### SPG

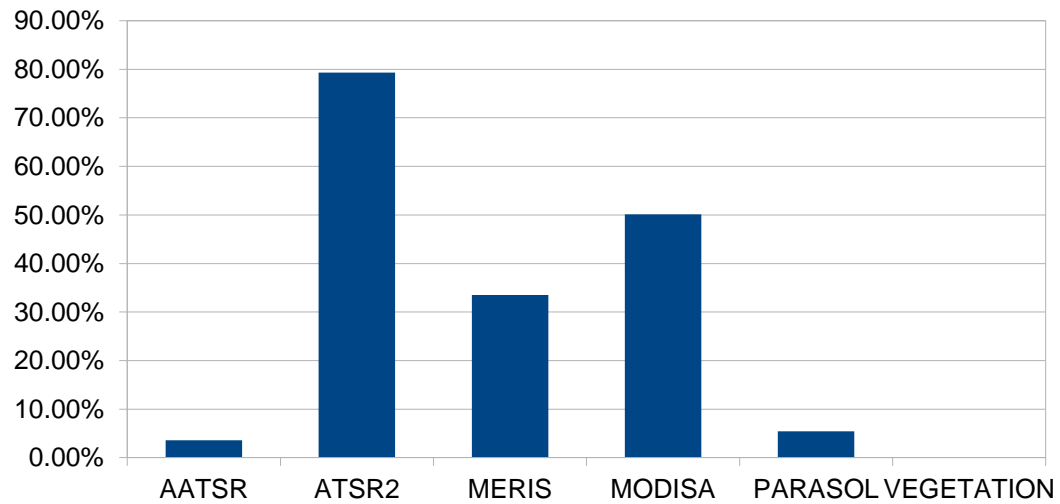


Figure 7 Percentage of images where the auto cloud screening agrees with the manual cloud screening at SPG

### TuzGulu

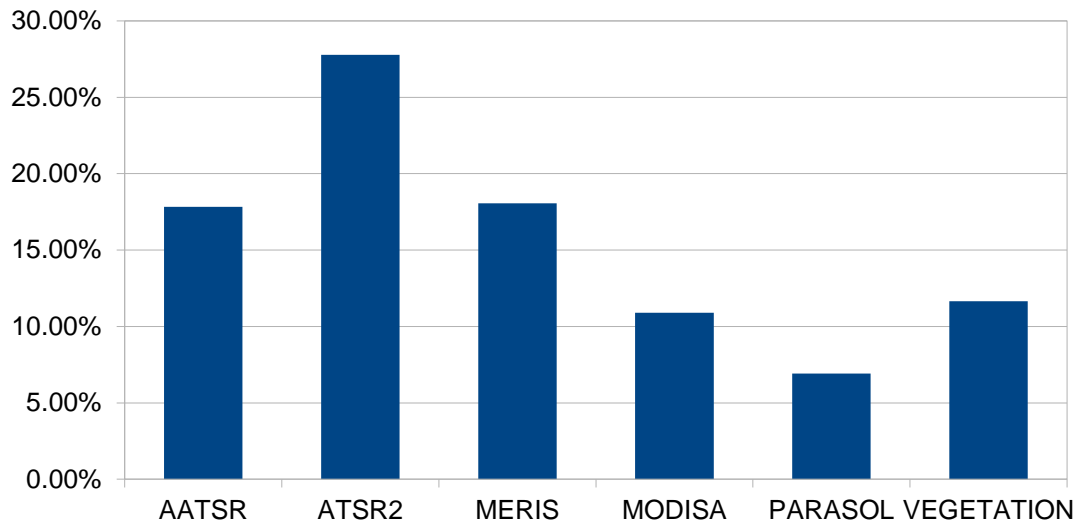


Figure 8 Percentage of images where the auto cloud screening agrees with the manual cloud screening at TuzGulu

### Uyuni

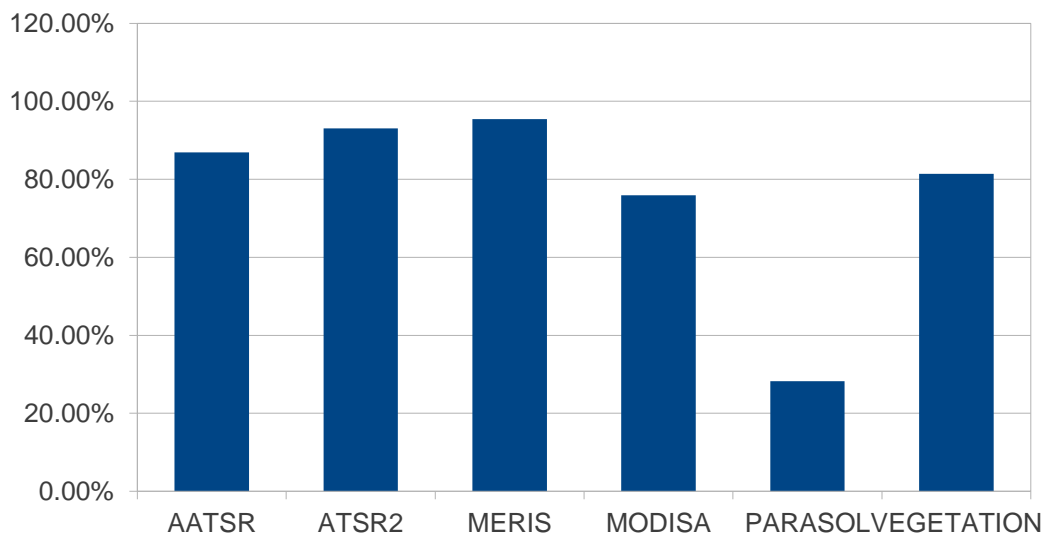


Figure 9 Percentage of images where the auto cloud screening agrees with the manual cloud screening at Uyuni

### All Sensors

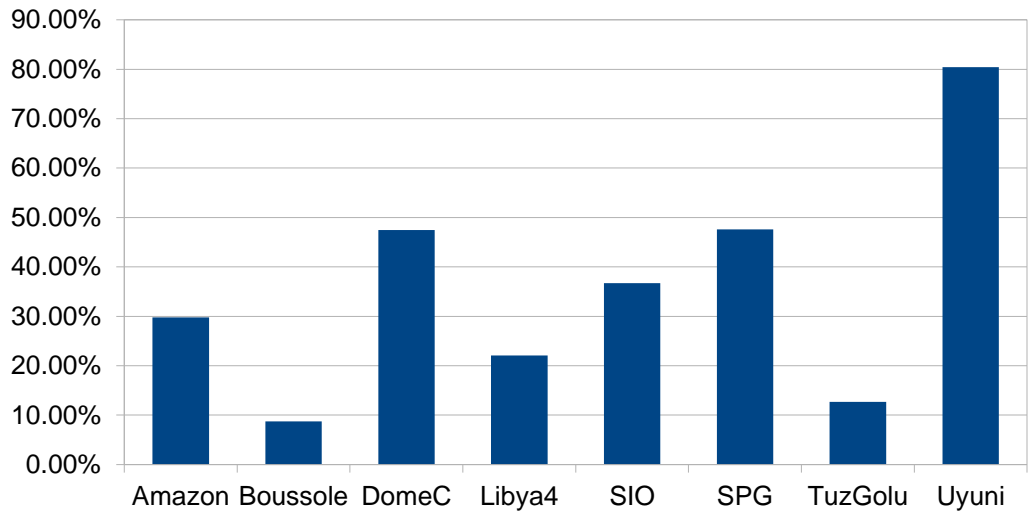


Figure 10 The performance of the auto cloud screening of all of the sensors combined at each site.

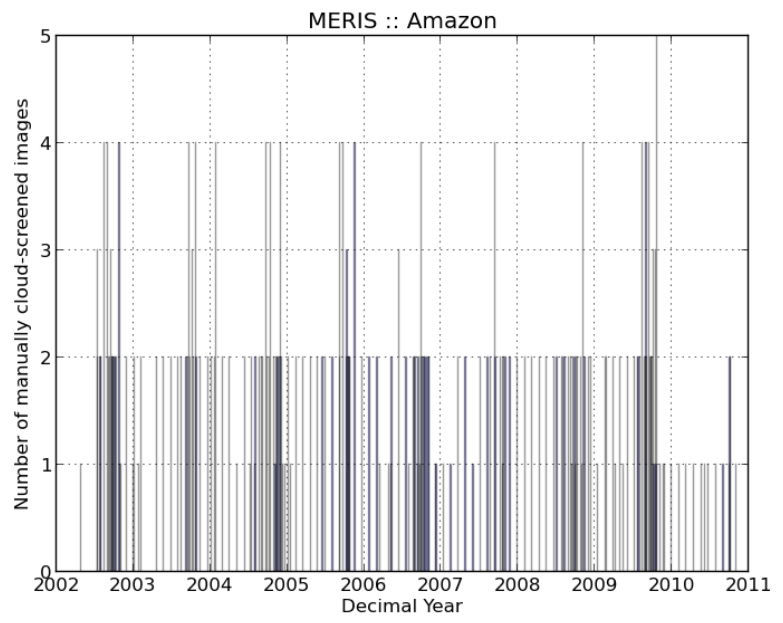


Figure 11 The number of manually cloud screened MERIS images at Amazon with respect to date. Each bar represents approximately 1 week.

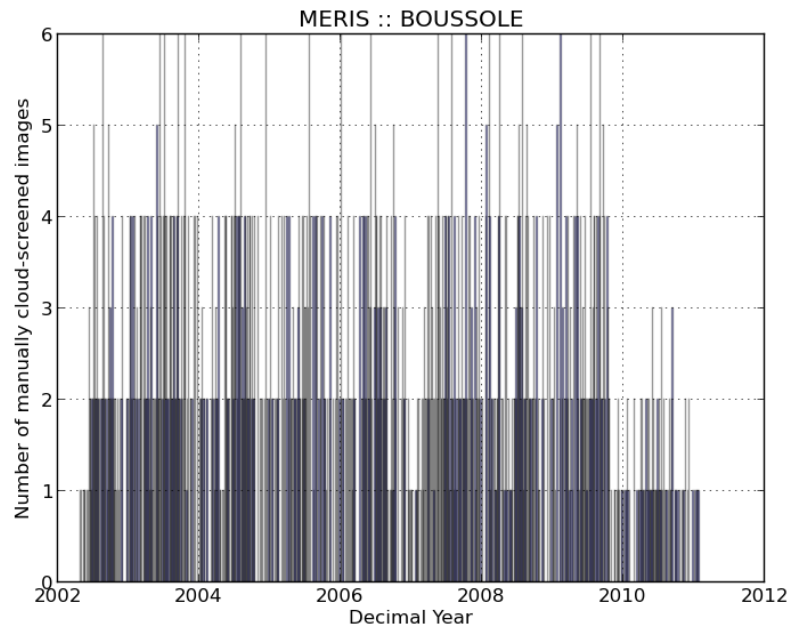


Figure 12 The number of manually cloud screened MERIS images at BOUSSOLE with respect to date. Each bar represents approximately 1 week.

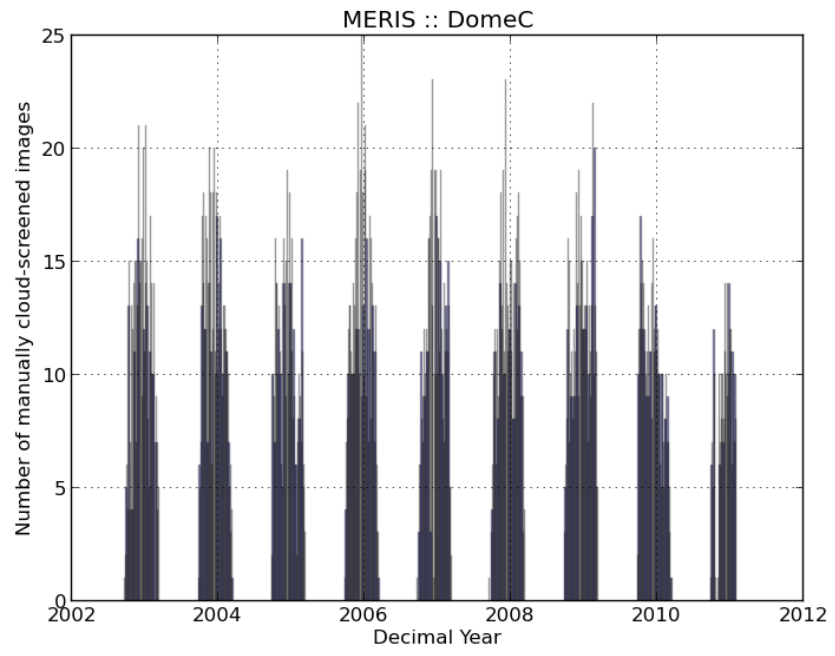


Figure 13 The number of manually cloud screened MERIS images at DomeC with respect to date. Each bar represents approximately 1 week.

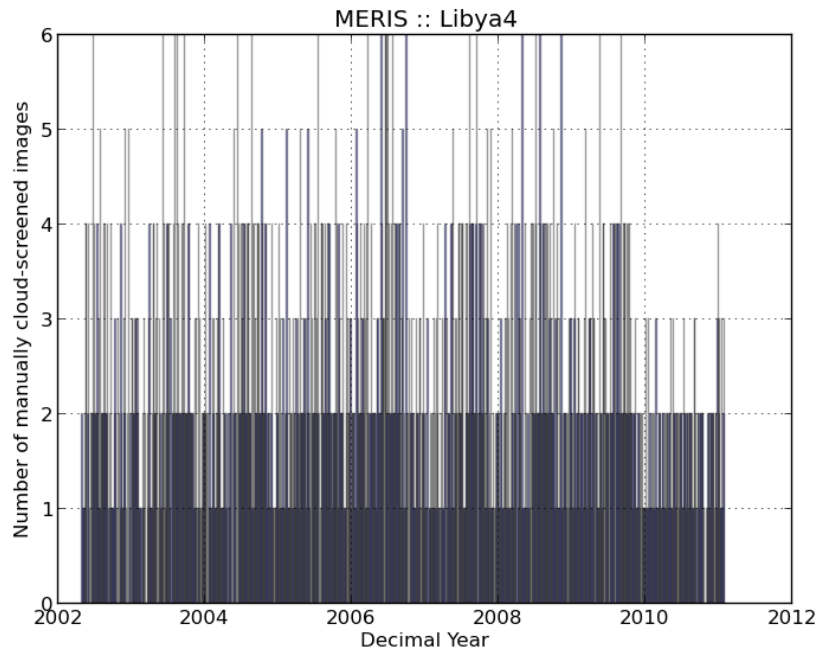


Figure 14 The number of manually cloud screened MERIS images at Libya4 with respect to date. Each bar represents approximately 1 week.

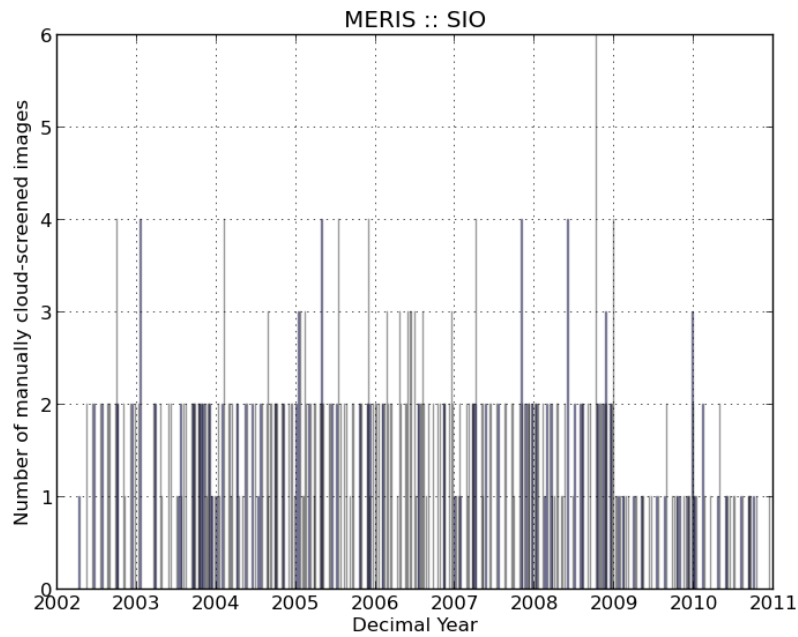


Figure 15 The number of manually cloud screened MERIS images at SIO with respect to date. Each bar represents approximately 1 week.

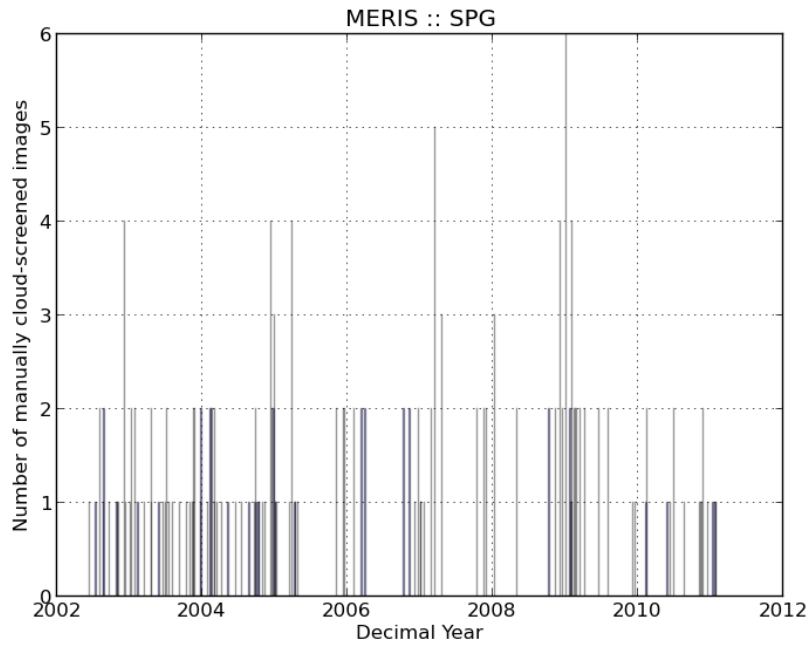


Figure 16 The number of manually cloud screened MERIS images at SPG with respect to date. Each bar represents approximately 1 week.

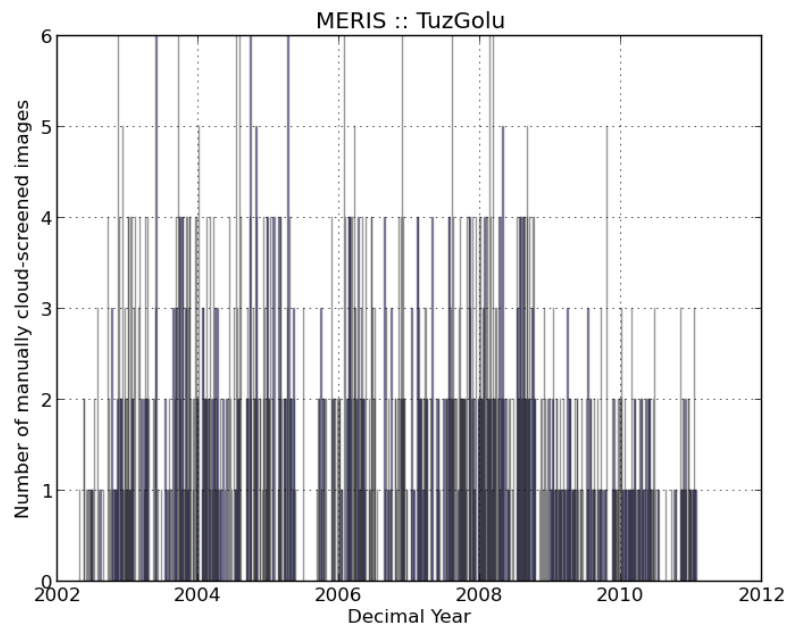


Figure 17 The number of manually cloud screened MERIS images at TuzGulu with respect to date. Each bar represents approximately 1 week.

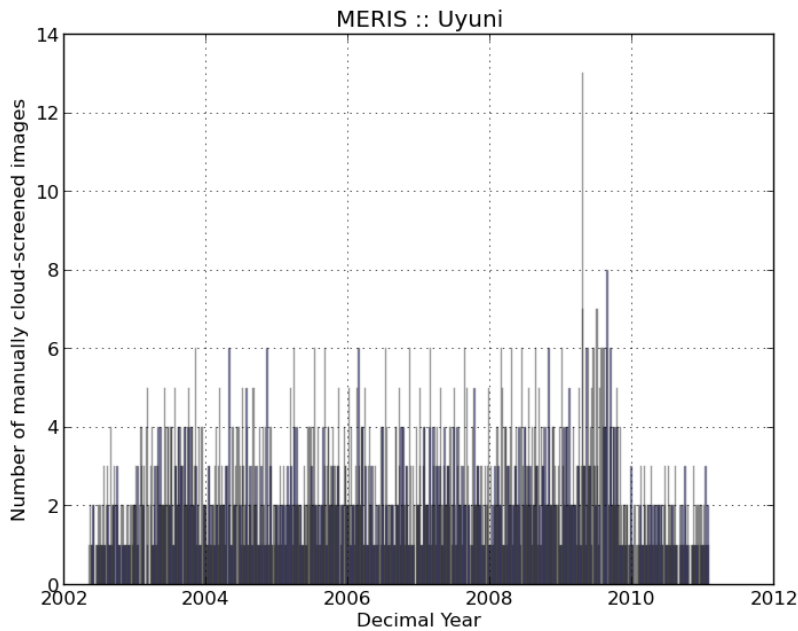



Figure 18 The number of manually cloud screened MERIS images at Uyuni with respect to date. Each bar represents approximately 1 week.

[SEE APPENDIX : CLOUD SCREENING STATISTICS FOR STATISTICS FOR OTHER SENSORS](#)

## 2.5 Potential areas for improvement

Previous assessments of the existing cloud screening algorithms show that all are potentially adequate for semi-automated cloud screening purposes. However, there are currently a number of issues with both arising from the application of algorithms to sensors they were not originally designed for.

Tests on MERIS data initially showed Globcarbon significantly underestimating cloud area. This was because a substantial amount of the cloud pixels over land were being classified incorrectly by the algorithm as snow. Therefore, these did not appear in the final cloud mask. Comparison of a range of images containing no snow (Amazon), mixed cloud & snow (Boussole, Tuz Golu) and only snow (Dome-C) suggested that the threshold value for the MERIS band 9/13 ratio test that determines the cloud/snow distinction derived from the Globcarbon paper was set too low. On the basis of the images tested it was established that a threshold value of approximately 1.35 to 1.4 would significantly improve the accuracy of the cloud areas calculated. This corresponds to the adjustment to the thresholds that should be made to allow for the fact that the values described in the source paper are based on reflectances and the values in the product assessed through the JAVA test platform are radiances. A proportional increase for the bright land mask from 0.8 to 1.1 was implemented but not found to be as critical to the performance of the

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 18</p>
---	---	--


algorithm in the tests. Any increase beyond 1.1 resulted in too much misclassification of dark land areas as bright land in areas such as the Amazon.

Extension of the algorithm to PARASOL products required selection of the PARASOL bands closest in frequency to the MERIS bands. The effectiveness of these band ratios and the new thresholds required by the filters was determined empirically.

The test results for the Landsat ACCA algorithm currently show poor accuracy. This is because cloud coverage in any product is currently being associated only with the amount of cold cloud which the algorithm finds in the data on the first pass. Pixels which are potentially ambiguous either as bright land or snow are currently being completely excluded from the cloud area estimation. This sometimes results in underestimation of cloud area and sometimes in over estimation depending on the degree of ambiguity that the algorithm finds in the data. There are two underlying reasons for this behaviour. The first is that the use of bands from instruments which are equivalent but not identical to the Landsat bands requires that new thresholds be found and set. Improved thresholds would mean improved classification of pixels as cloud or as ambiguous and would result in more effective screening by the filters of the first pass of the algorithm.

A further confounding factor is that Landsat 7 ACCA is a two pass algorithm and only the first pass is currently implemented. As described above, the first pass runs images through a series of filters which perform a preliminary classification of pixels leaving some flagged as ambiguous. The second pass uses an analysis of the distribution of values in band 6 (thermal band) to arrive at a final classification of these ambiguous pixels. The algorithm is therefore currently only reporting pixels as cloud when no first pass filter is able to classify the pixel as ambiguous. It is estimated that a further 3-5 days work would be required to implement the second pass of the algorithm and rerun the tests and perform the analysis necessary to determine the appropriate thresholds for each of the filters on the different instruments.



	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 19</p>
---	---	--

## 3 Improved Automated Cloud Screening Implementation in DIMITRI Version 3.0

### 3.1 Overview of approaches

In addition to the current cloud screening methodologies, two new approaches are described here. The first method, Spatial Scale Variability (SSV), measures the statistical variance of the per-pixel values from the mean top of atmosphere value over the area of interest. If the variance is above an empirically derived threshold then the scene is considered cloudy. The second method, BRDF Variability Threshold (BRDFVT), measures how closely the BRDF of a satellite measurement over a site complies with an empirically derived model of the BRDF for that site. If the BRDF deviates beyond a certain threshold from the modelled BRDF the scene is considered cloudy.

### 3.2 Method 1: Spatial Scale Variability (SSV)

Firstly a training set of images are selected such that images can be classified into three groups and defined as: clear-sky, part-cloud or full-cloud. A curve of how mean and standard deviation change with increasing resolution is produced for all of the training images. This procedure is defined below.


A low resolution is defined that is 2 x 2 pixels. The scene is interpolated to the new resolution and the standard deviation of the TOA value over those pixels is calculated. The resolution size is increased until the averaging window is the same resolution as the scene. The mean and standard deviation are calculated for each resolution as shown below.

$$\mu(m, n, \lambda) = \frac{1}{(m \times n)} \sum_{j=1}^n \cdot \sum_{i=1}^m \cdot \rho_{i,j}(\lambda)$$

$$\sigma(m, n, \lambda) = \sqrt{\frac{1}{(n \times m)} \sum_{j=1}^n \cdot \sum_{i=1}^m \cdot (\rho_{i,j}(\lambda) - \mu(m, n, \lambda))^2}$$

$$\{m \in \mathbb{Z} \mid 1 < m \leq \text{window length}\}$$

$$\{n \in \mathbb{Z} \mid 1 < n \leq \text{window width}\}$$

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 20</p>
---	---	--

A power law approximation for the changing variability is computed and the gradient of the slope compared against a set of defined thresholds.

$$\sigma(m, n, \lambda) \approx Ax(\lambda)^B$$

Where A is the linear scaling factor and B is the slope factor. Both A and B are found using least-squares regression.

The thresholds are computed by analysing the change in reflectance variability with pixel resolution on a number of data products within DIMITRI. The magnitude of the standard deviation of the reflectance are considered within the algorithm to allow less strict thresholds over low reflectance targets such as the oceans and stricter thresholds over bright targets where the difference between the surface and cloud is small.


This method of subsampling allows classification of observations as either clear sky, fully cloud, and possibly part cloud; it cannot however determine the estimated cloud percentage of the observation. It is likely that there will not be clearly defined categories for the gradient of the power law approximation; the subsampling test is therefore performed in addition to the currently implemented cloud screening algorithms. Should the subsampling results fall within a defined ‘ambiguous’ region of gradients (at the crossover from clear sky to cloudy) then the radiometric ratio threshold test should be used to determine the cloud coverage classification.

### 3.3 Method 2: BRDF Variability Threshold BRDFVT

A time series of observations are acquired at different viewing and illumination angles in a given spectral interval is acquired with enough clear sky days, verified for instance with visual inspection. An empirical BRDF model is fit to these TOA BRDF data. Such model and associated uncertainties are subsequently used to predict the estimated magnitude of the observed signal corresponding to a given geometry. Any departures from this model prediction including the uncertainties are used to interpret the scene as having the presence of clouds.

The cloud screening approach described above is not impacted by temporal variability because it is applied pixel by pixel. Additionally, because DIMITRI employs a matchups approach, timescales are kept similar by definition and temporal variability is not a significant factor.

Firstly a set of clear sky images are selected over the site of interest. A temporal BRDF model is simulated by fitting a weighted curve through multiple measurements over a site at different times and sun-sensor viewing geometries.

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 21</p>
---	---	--

Images are cloud screened by computing the BRDF and calculating the relative bias between the calculated BRDF and the model. If the residual is greater than an empirically derived threshold, then the scene is considered cloudy.

$$\beta(t, d) = \frac{B_f^*(t, d) - \widehat{B}_f(t, d)}{\widehat{B}_f(t, d)}$$

Where t is time and d is solar direction.

The uncertainty is estimated by weighting the curve fit procedure with the variance of the BRDF observation.

$$\delta\beta(t, d) = \beta(t, d) \left( \left| \frac{\delta\widehat{B}_f^*(t, d)}{\widehat{B}_f^*(t, d)} \right| + \left| \frac{\delta\widehat{B}_f(t, d)}{\widehat{B}_f(t, d)} \right| \right)$$

Where  $\delta\widehat{B}_f^*(t, d)$  is the standard deviation of the pixels covering the area of interest.


### 3.4 Output files generated by the SSV cloud screening

The output directory is named SITE\_DATE\_SSV\_CS\_SENSOR\_PROC where DATE is the date of run. Two types of files are systematically generated for each SSV cloud screening run:

- **SSV\_CS\_LOG.txt:** log file summarising all options of the run (parameters).
- **SCENE\_SSV.JPG:** plots of the standard-deviation as function of the subsampling window size, for both the processed scene and the training classes (see example in ). SCENE is the filename of the scene, similar to the RGB filename except the extension.

Furthermore, when the SSV is run for the first time, the fitting coefficients of the training stage are stored in an IDL SAV file placed in the Input directory containing input training dataset files (see SSV CS options hereafter), i.e. Site\_SITE/SENSOR/Proc\_PROC:

- **CLOUD\_TRAINING\_SSV\_FIT\_B.dat:** IDL save file relative to training of band B containing the following variables:
  - N\_CLASS: number of training classes
  - CLASS: array containing each class names

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 22</p>
---	---	--

- COLOUR\_CLASS array containing each class colour code
- SSV\_BAND: the training band (must correspond to field B of the filename)
- A\_SSV: array of fitting coefficient A for each class
- B\_SSV: array of fitting coefficient B for each class

### 3.5 Output files generated by the BRDF cloud screening

Output directory is named SITE\_DATE\_BRDF\_CS\_SENSOR\_PROC where DATE is the date of run. Six types of files are systematically generated for each SSV cloud screening run:

- **BRDF\_CS\_LOG.txt:** log file summarising all options of the run (parameters).
- **BRDF\_CS\_ANALYSIS\_SITE\_SENSOR\_PROC.JPG:** plot of the simulated over observed TOA signal ratio as function of time, for all observations providing a manual classification as either cloud or clear (see example on Figure 34)
- **AUTO\_CS\_PERF\_SITE\_SENSOR\_PROC.JPG:** histogram plot performance of the nominal DIMITRI cloud screening (AUTO\_CS field in DB), for clear and cloudy conditions as referred by the manual classification (result of the nominal screening is considered as cloudy if AUTO\_CS>0)
- **BRDF\_CS\_PERF\_SITE\_SENSOR\_PROC.JPG:** histogram plot performance of the BRDF cloud screening, for clear and cloudy conditions as referred by the manual classification
- **BRDF\_CS\_PERF\_SITE\_SENSOR\_PROC.CSV:** text file containing the performance number used in the histogram plots (in %).
- **DIMITRI\_DATABASE\_BRDF\_CS.CSV:** subset of the DIMITRI database file corresponding to user options (site, sensor, etc.) with BRDF cloud screening output instead of AUTO\_CS field.


## 3.6 DIMITRI modules

### 3.6.1 New RGB generation

All quicklook modules have been updated in order to draw a red box around the ROI and not mask the ROI itself (see example in Figure 26). This is important as the previous red overlay on full ROI often hid some thin patterns and scenes were erroneously manually screened as clear; examples of such misclassification are given in the results hereafter.

### 3.6.2 SSV cloud screening modules

The SSV method is implemented in the single cloudscreening/ssv\_cloud\_screening.pro routine.

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 23</p>
---	---	--

This uses the IDL CONGRID function to subsample the window and LINFIT to compute the fitting coefficients (with formerly log-transformed inputs). Note that for allowing 2x2, 3x3, ...NxN pixels subsampling, the largest square windows inside the ROI is chosen; this may lost pixels at last corner acquisition when the ROI is not square.

The module mainly consists of two stages:

- The training stage (optional, see option below): this stage computes the fitting coefficients ( $A_c, B_c$ ) for each class  $c$  detected in the Input directory. The final fitting coefficients of a class correspond to the median of all fitting coefficients of individual observations.
- The screening stage: it first computes the standard-deviation  $\sigma(i)$  of each subsample window  $i$  of the current scene and then measures its distance to each class by the Euclidian norm:

$$\chi(c) = \sum_{i=0}^N (\sigma(i) - A_c i^{B_c})^2$$

The class with lowest  $\chi$  is considered as the most representative of the scene.

### 3.6.3 BRDF cloud screening modules


The BRDF method is implemented in the single cloudscreening/brdf\_cloud\_screening.pro routine and makes use of already existing BRDF modules. Similarly to VEGETATION TOA simulations, the TOA signal is first corrected for gaseous absorption (with exact integration on the sensor RSR) but not corrected for Rayleigh and aerosol scattering. The module mainly consists in two stages:

The training stage: this stage creates a fake “super sensor” file based on the clear sky training dataset and then calls the Roujean BRDF module to compute the K1, K2, K3 coefficients over a given binning sample.

The screening stage: the signal of current scene is simulated at TOA with respect to the Roujean BRDF and actual geometry; because the training dataset generally does not cover continuously the full sensor time-series, the closest BRDF bin to actual acquisition time is chosen. The simulated signal is then compared to the actual measurement by the ratio:

$$Ratio(\lambda) = \left| \frac{\rho_{TOA}^{sim}(\lambda) - \rho_{TOA}^{obs}(\lambda)}{\rho_{TOA}^{obs}(\lambda)} \right|$$

The scene is classified as cloudy when this ratio is above a threshold given by the user.

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 24</p>
---	---	--

NOTE: BECAUSE THIS MODULE CALLS BRDF ROUTINES COMPILED AFTER THE CLOUD MODULES, A COMPILATION ERROR APPEAR WHEN COMPILING ONCE. HENCE DIMITRI MUST BE COMPILED TWICE, BY TYPING TWICE @COMPILE DIMITRI

## 3.7 HMI updates and user options

### 3.7.1 Main window

The main window has been updated with a new column dedicated to cloud screening.

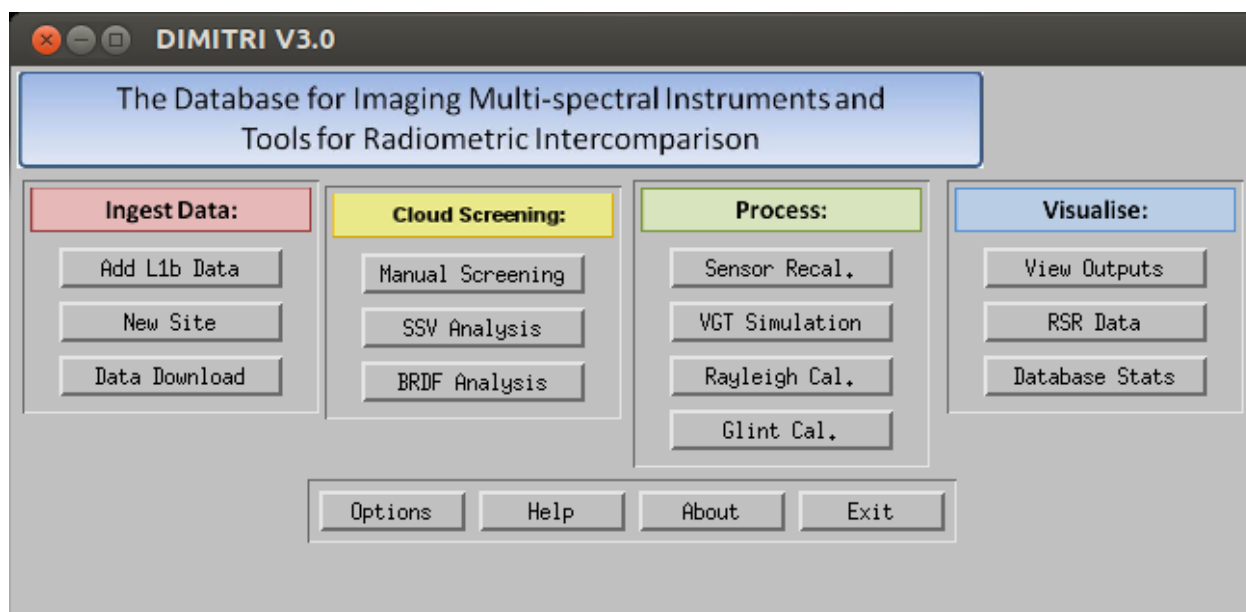


Figure 19 DIMITRI main window

### 3.7.2 SSV cloud screening options

The DIMITRI SSV cloud screening window is shown in Figure 20. Options are:

- Case study (region, sensor, processing version, time range)
- Coverage criteria. This applies both for the training stage (if activated) and the screening stage.
- Cloud screening option: skip the training stage or not and select the band used for standard-deviation computation. Skipping the training stage can be of interest when the fitting coefficients already exists from a previous run because computation may take a relatively long time, depending on number of classes and number of scenes per classes.

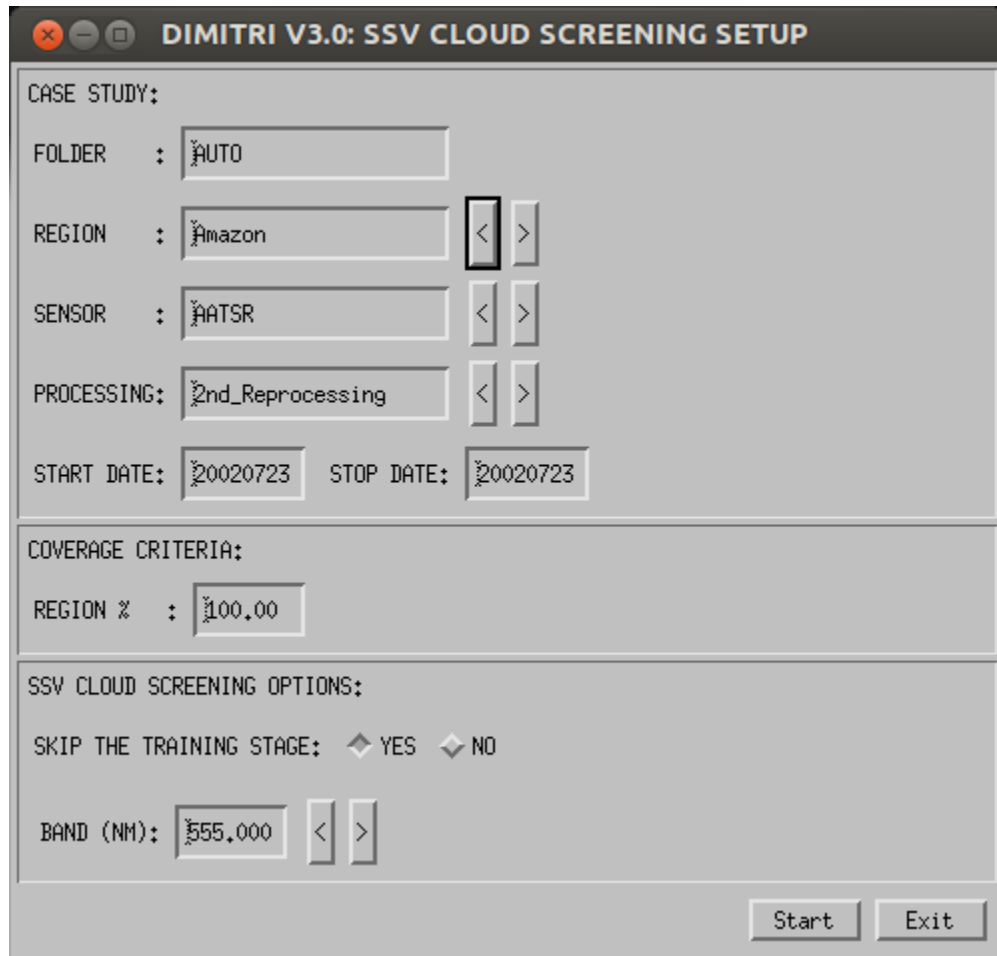


Figure 20 DIMITRI SSV cloud screening window

The training stage, when not skipped, automatically searches for all training datasets matching the pattern


#### **DIMITRI\_DATABASE\_CLOUD\_TRAINING\_CLASS.CSV**

in the input directory corresponding to the run, i.e.

#### **Input/Site\_SITE/SENSOR/Proc\_PROC/**

Where CLASS can be any string identifying a class, e.g. CLEAR, PARTIALLY\_CLEAR, etc. The format of these training files is exactly similar as the standard DIMITRI\_DATABASE.CSV file; a convenient way to generate a training file is thus to take a subset of the DB file that corresponds to manual inspection or any other choice.

The training is then conducted on all CSV datasets identified and the suffixes are used in the output plots to name each class.

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 26</p>
---	---	--

NOTE: THERE ARE AS MANY SCENE\_SSV.JPG OUTPUTS AS SCENES CORRESPONDING TO THE CHOSEN TIME RANGE. HENCE IT IS WORTH LIMITING EACH SSV CLOUD SCREENING RUN TO FEW DAYS OR FEW MONTHS.

### 3.7.3 BRDF cloud screening options

- The DIMITRI BRDF cloud screening window is shown on **Error! Reference source not found..** Options are:
- Case study (region, sensor, processing version, time range)
- Coverage criteria. This applies both for the BRDF computation stage and the screening stage
- BRDF training parameters, including:
  - Clear sky database for BRDF computation
  - Size of bins in days
  - Minimal number of observation per bin
  - Viewing and solar angles range
  - Option to clean the output directory for all temporary BRDF outputs
- BRDF cloud screening parameters, i.e. band of the TOA ratio and threshold to detect clouds

The clear sky database is exactly similar to training datasets of the SSV cloud screening and must correspond to good condition for BRDF computation (typically user can select the clear sky class of the SSV cloud screening).

NOTE: WHEN THE CLEAR SKY DATABASE IS SCARCE, THE BIN PERIOD MUST BE LARGE (E.G. 100 DAYS OR MORE), OTHERWISE THE ROUJEAN BRDF CANNOT FIND ENOUGH OBSERVATIONS INSIDE THE BINS.



✕ ◀ ▶
**DIMITRI V3.0: BRDF CLOUD SCREENING SETUP**

**CASE STUDY:**

FOLDER :

REGION :  < >

SENSOR :  < >

PROCESSING:  < >

START DATE:  STOP DATE:

---

**COVERAGE CRITERIA:**

REGION % :

---

**BRDF TRAINING OPTIONS:**

CLEAR SKY DATABASE :  Browse

BRDF BIN PERIOD (DAYS) :

BRDF OBS. THRESHOLD :

ANGLES	MIN	MAX
VZA :	<input type="text" value="0"/>	<input type="text" value="90"/>
VAA :	<input type="text" value="0"/>	<input type="text" value="360"/>
SZA :	<input type="text" value="0"/>	<input type="text" value="90"/>
SAA :	<input type="text" value="0"/>	<input type="text" value="360"/>

CLEAN TEMPORARY BRDF OUTPUTS: ◆ YES ◆ NO

---

**BRDF CLOUD SCREENING PARAMETERS:**

BAND (NM) :  < >

TOA THRESHOLD (%):

Start
Exit

Figure 21 DIMITRI BRDF cloud screening window

## 4 Results of cloud screening method implementations

### 4.1 Methodology results

#### 4.1.1 Training datasets

For the purpose of illustrating and validating the methods, training datasets were generated based on the cloud screening information already existing in the database. Scenes were first selected randomly from the full database for the following three classes:

- Clear: MANUAL\_CS=0
- Partially clear: MANUAL\_CS=1 and AUTO\_CS<0.3
- Cloudy: MANUAL\_CS=1 and AUTO\_CS>5

The number of scenes is about 20% on the total manually screened data, but varies upon sensor to have a decent size in the training dataset.


Then a visual inspection was briefly made to check selected scenes are suitable for each class, and add or remove some.

Table 10 Number of observations in Libya4 training datasets used hereafter, for 3 classes

	Class Clear	Class Partially clear	Class Cloudy
<b>AATSR</b>	15	16	0
<b>ATSR2</b>	18	4	0
<b>MERIS 3<sup>rd</sup> reproc.</b>	77	28	15
<b>MODISA</b>	20	15	14
<b>PARASOL</b>	0	16	0
<b>VEGETATION</b>			

Table 11 Number of observations in SPG training datasets used hereafter, for 3 classes

	Clear	Partially clear	Cloudy
<b>MERIS 3<sup>rd</sup> reproc.</b>	21	14	18

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 29</p>
---	---	--

## 4.2 MERIS SSV cloud screening over Libya4

There is a noticeable distinction between the three classes, particularly between clear and cloudy (Figure 22). Evolution of the standard-deviation with the subsampling size is very weak for each class, yet noisier when cloudy for small sizes.

Figure 22, Figure 23 and Figure 24 show respectively a correct cloud, clear sky and partially cloudy detection at band 412 nm.

For very small clouds, the distinction between clear and partially clear becomes difficult, and possibly a unique threshold on standard deviation would be more robust than the Euclidian norm to distinguish both classes; alternatively, the use of a band in the NIR, like 865 nm, introduces more dynamics in the standard deviation and allow to retrieve the partial cloudiness (case of 20020729, see Figure 25).

Interestingly, the SSV method allows detecting clouds which were discarded by manual inspection, see for instance case of 20060129 on Figure 26.

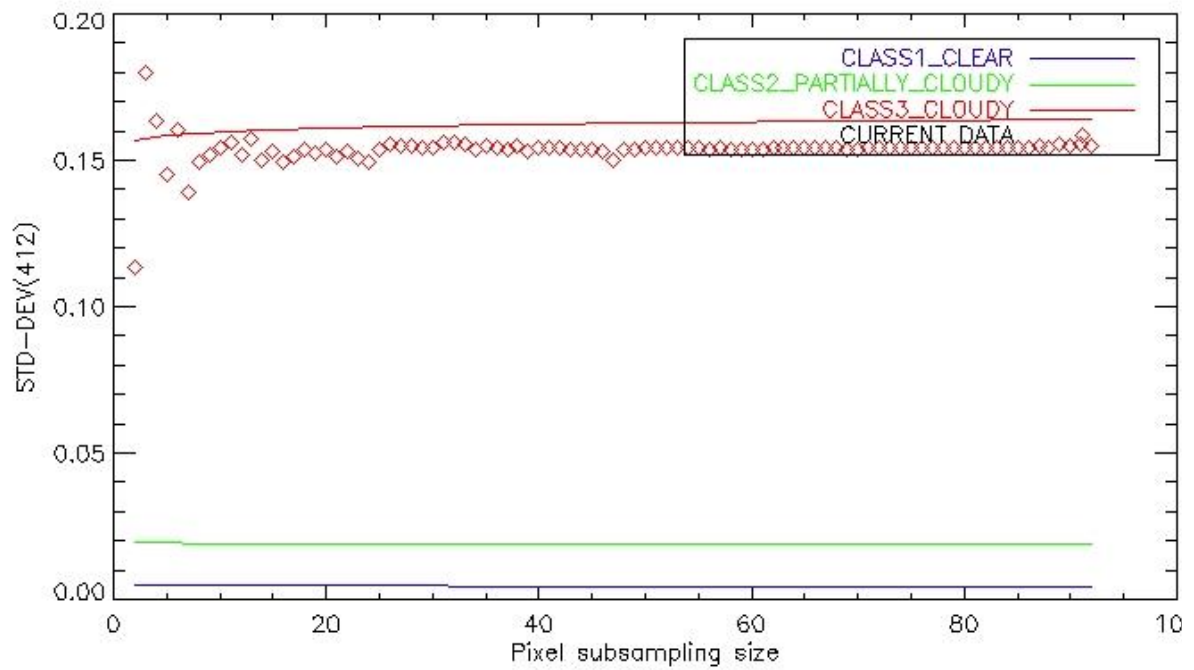


Figure 22 MERIS Libya4 20020507 RGB (top) and SSV cloud detection at 412 nm (bottom)

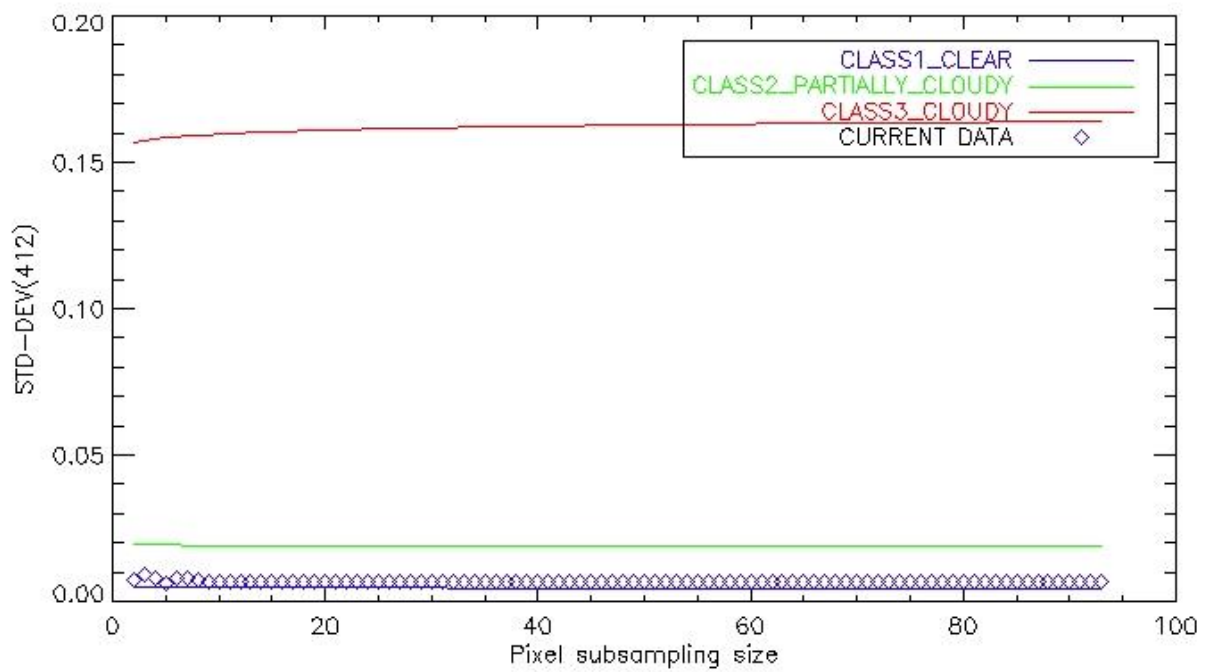
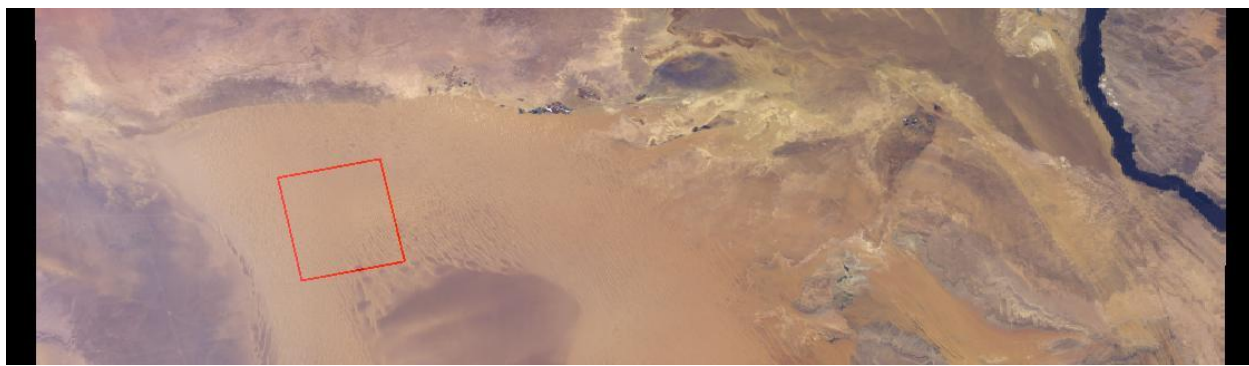


Figure 23 MERIS Libya4 20020612 RGB (top) and SSV cloud detection at 412 nm (bottom)

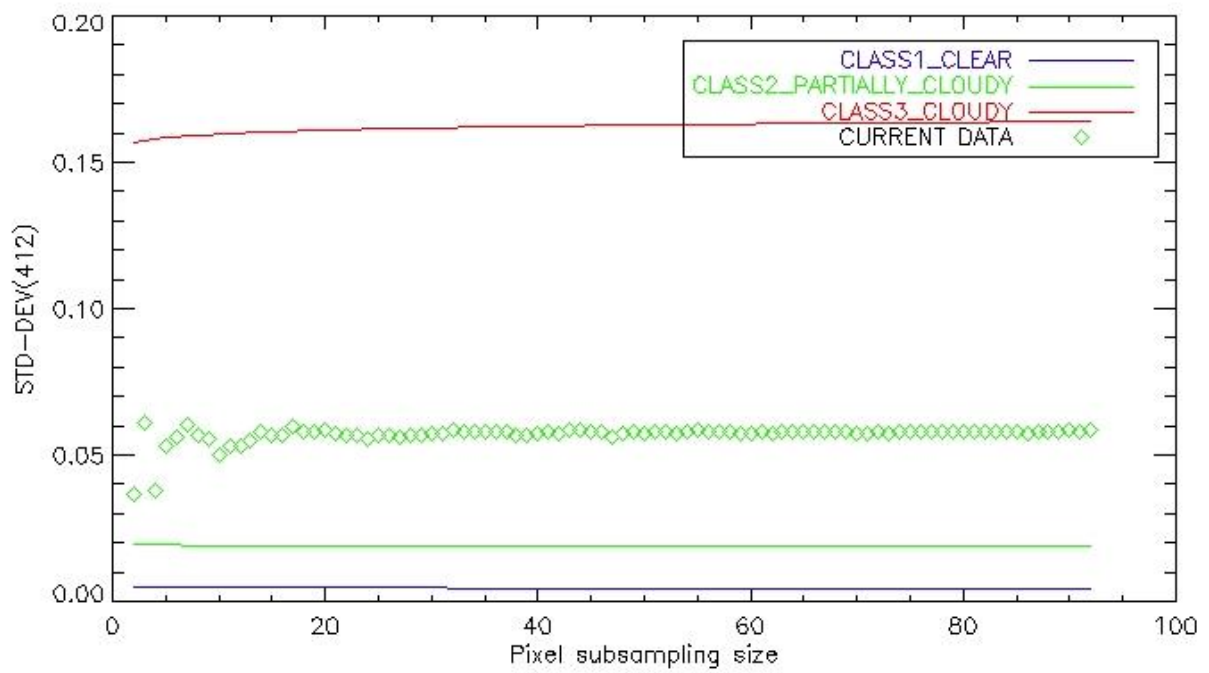
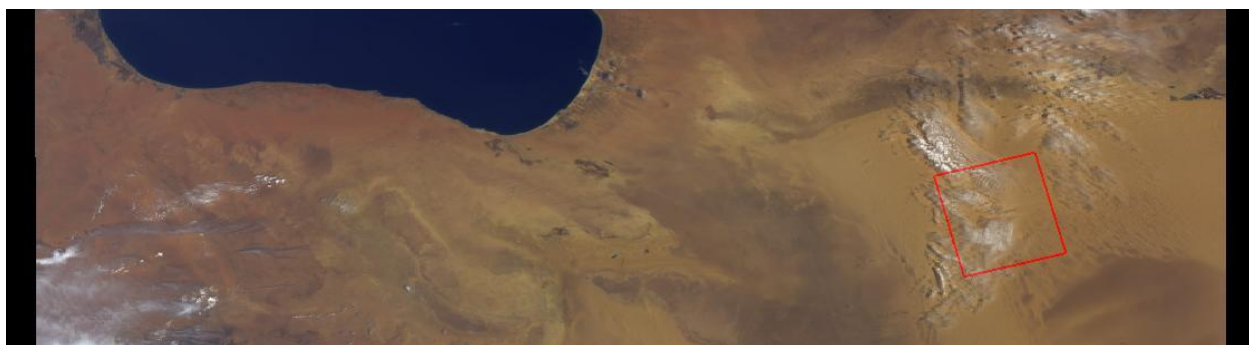


Figure 24 Libya4 20020504 RGB (top) and SSV cloud detection at 412 nm (bottom)

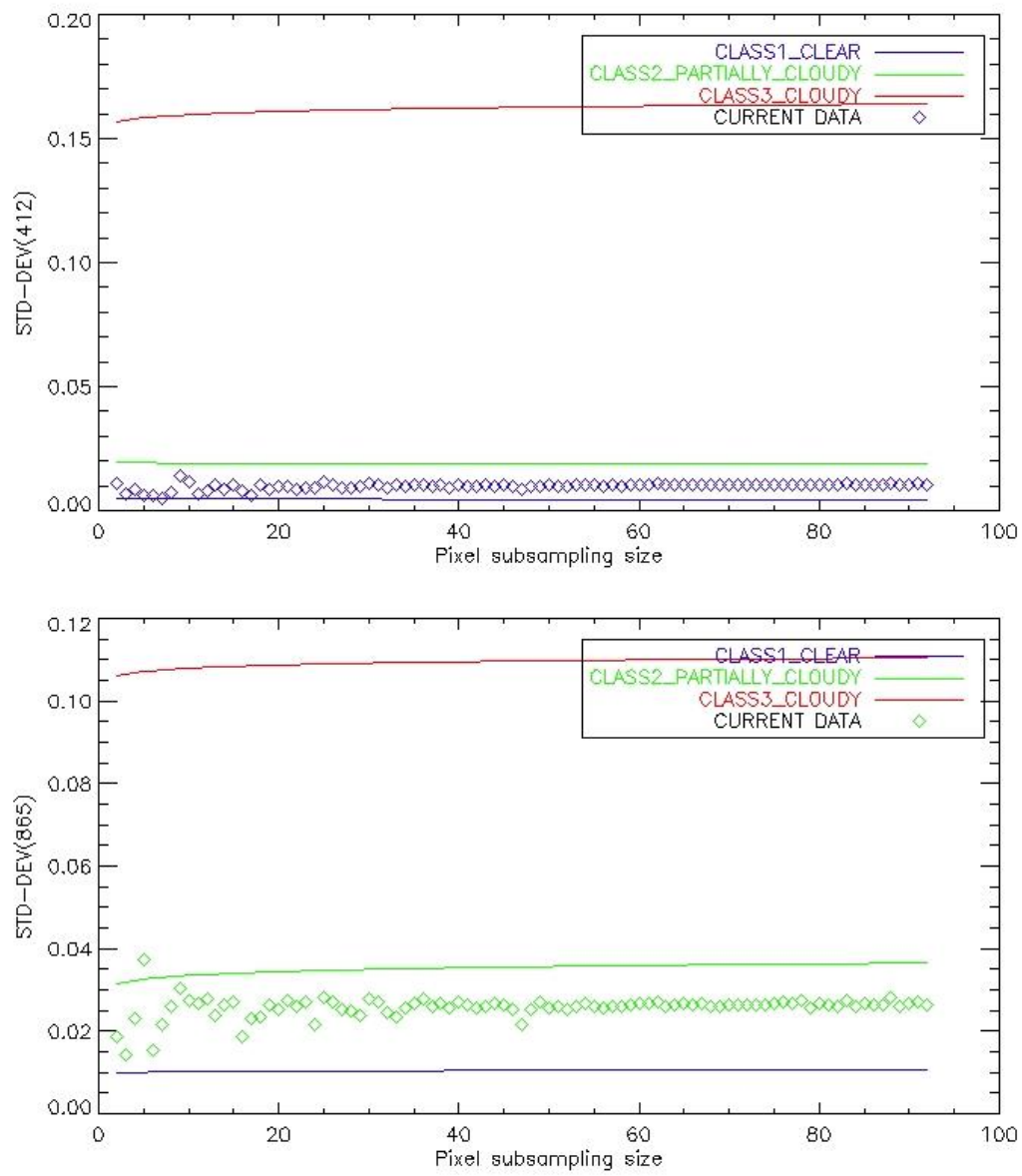


Figure 25 MERIS Libya4 20020729 RGB (top) and SSV cloud detection at 412 nm (middle) and at 865 nm (bottom)



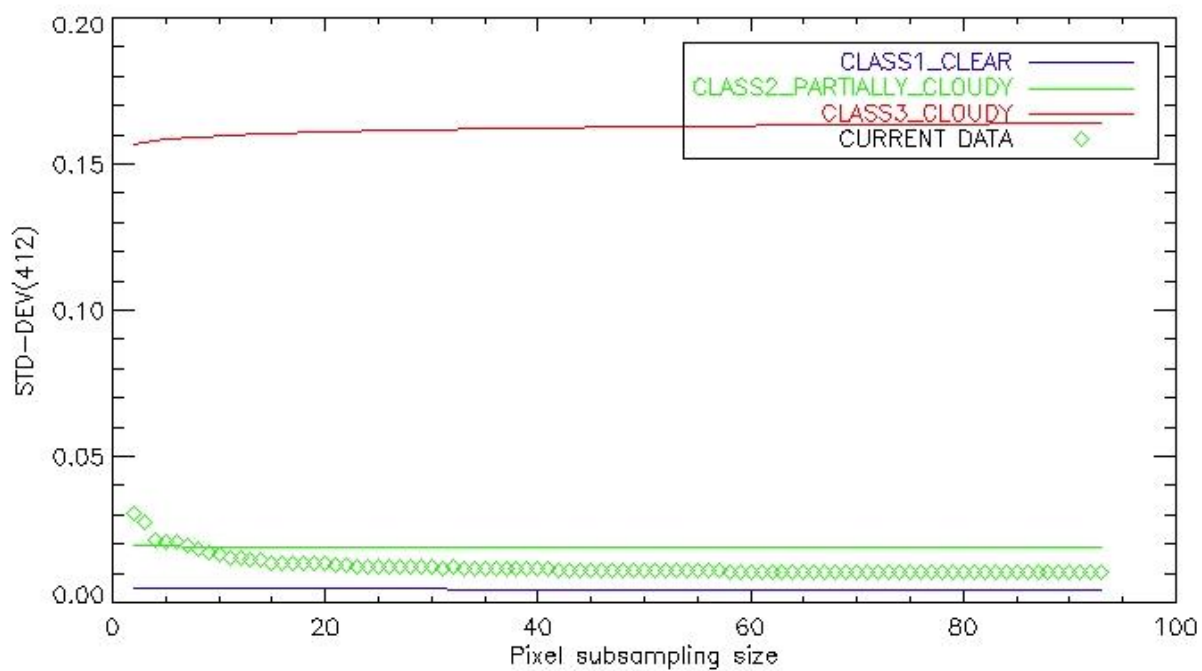
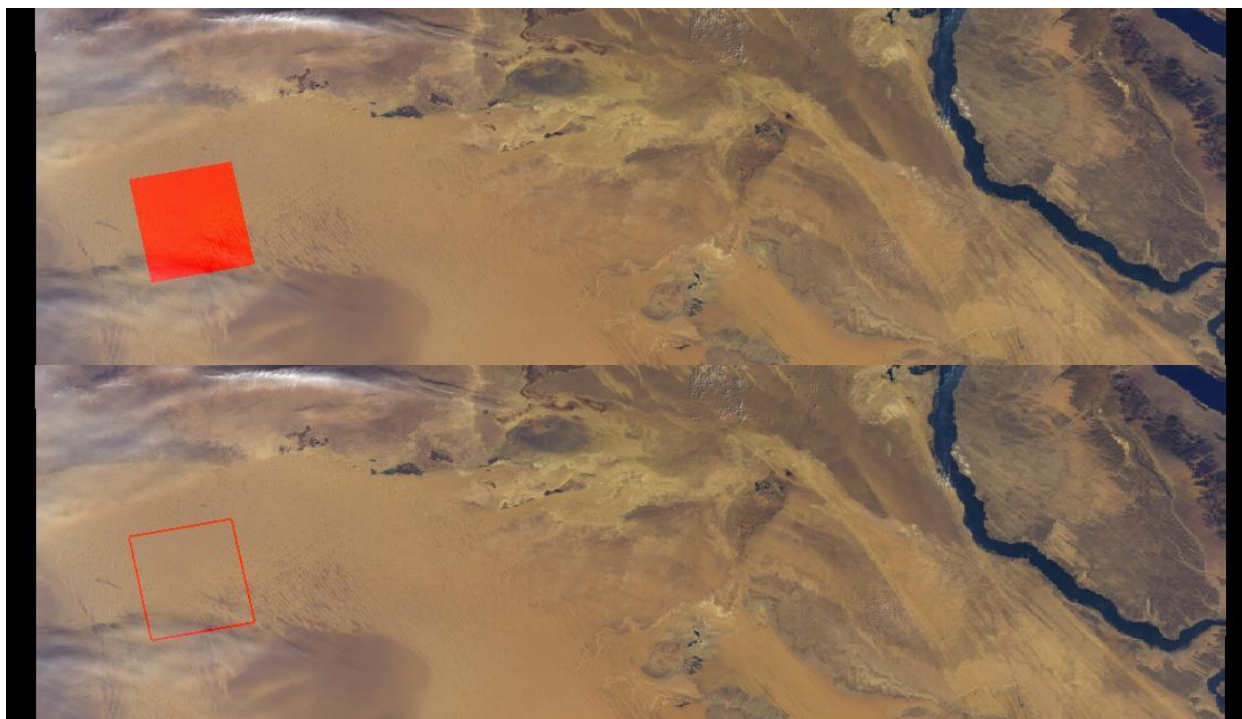


Figure 26 MERIS Libya4 20060129 former RGB (top), new RGB (middle) and SSV cloud detection at 412 nm (bottom)



#### 4.2.1 MERIS SSV cloud screening over SPG

The distinction between the cloudy and partially cloudy classes is now hardly visible over SPG, possibly because the corresponding training datasets are too similar (see e.g. Figure 27). The SSV method still allows distinguishing clear and cloudy scenes, as shown on Figure 27 to Figure 29 for respectively cloudy, clear and partially cloudy conditions. When the cloud coverage becomes too homogeneous, the standard-deviation decreases and the algorithm may incorrectly select the clear class (see example on Figure 30); in such a case, the standard-deviation is however large enough for distinguishing the clouds and a threshold could provide the correct classification.

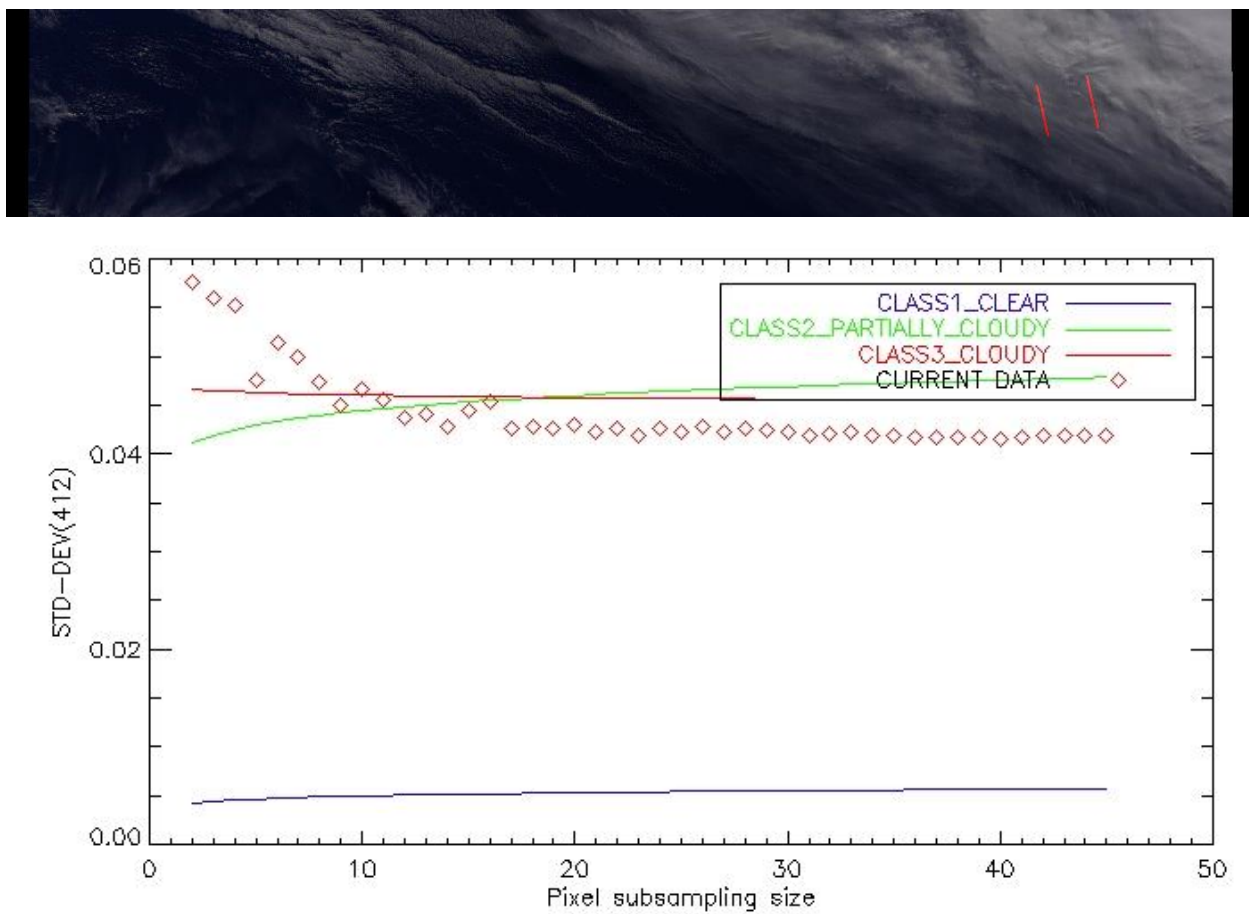


Figure 27 SPG 20020723 RGB (top) and SSV cloud detection at 412 nm (bottom)

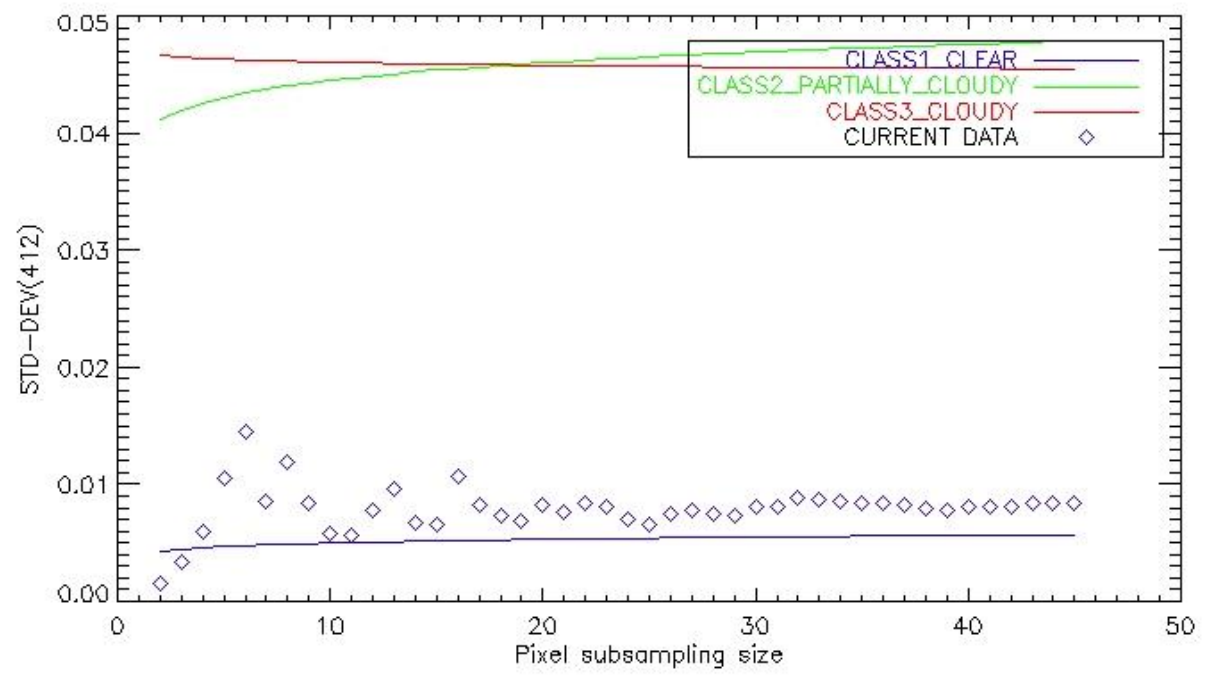


Figure 28 SPG 20020615 RGB (top) and SSV cloud detection at 412 nm (bottom)

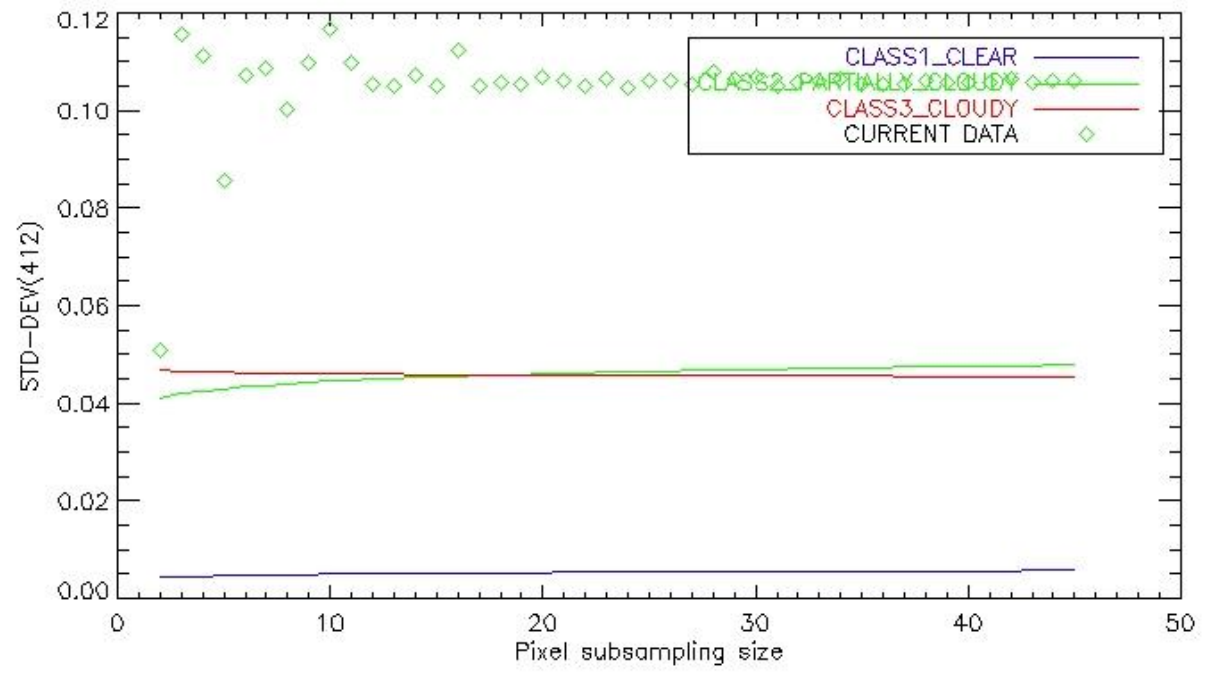
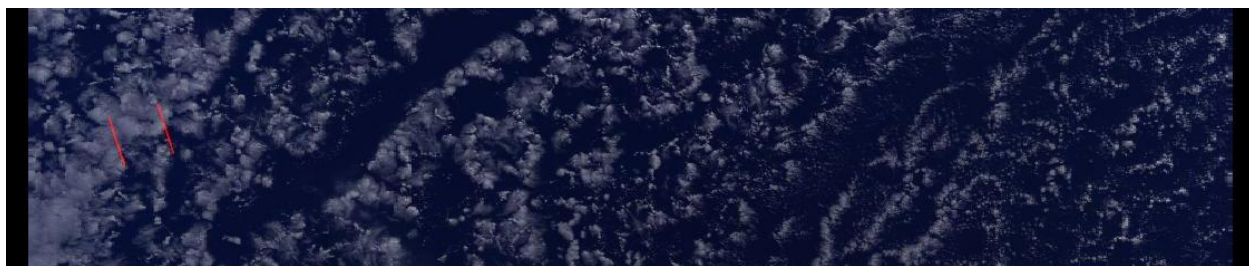


Figure 29 SPG 200200616 RGB (top) and SSV cloud detection at 412 nm (bottom)

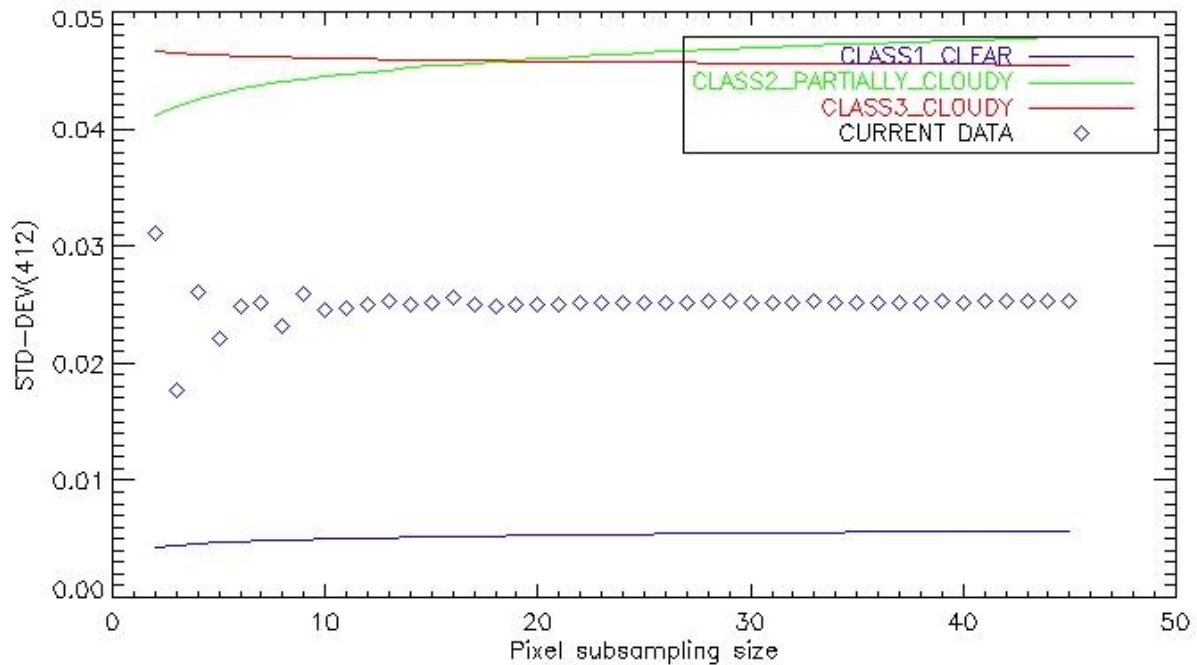


Figure 30 SPG 20020521 RGB (top) and SSV cloud detection at 412 nm (bottom)

#### 4.2.2 AATSR SSV cloud screening over Libya4

Examples of AATSR SSV correct detection for clear and cloudy sky are given respectively on Figure 31 and Figure 32. When cloud level becomes too small, the SSV method is unable to detect them (Figure 33).

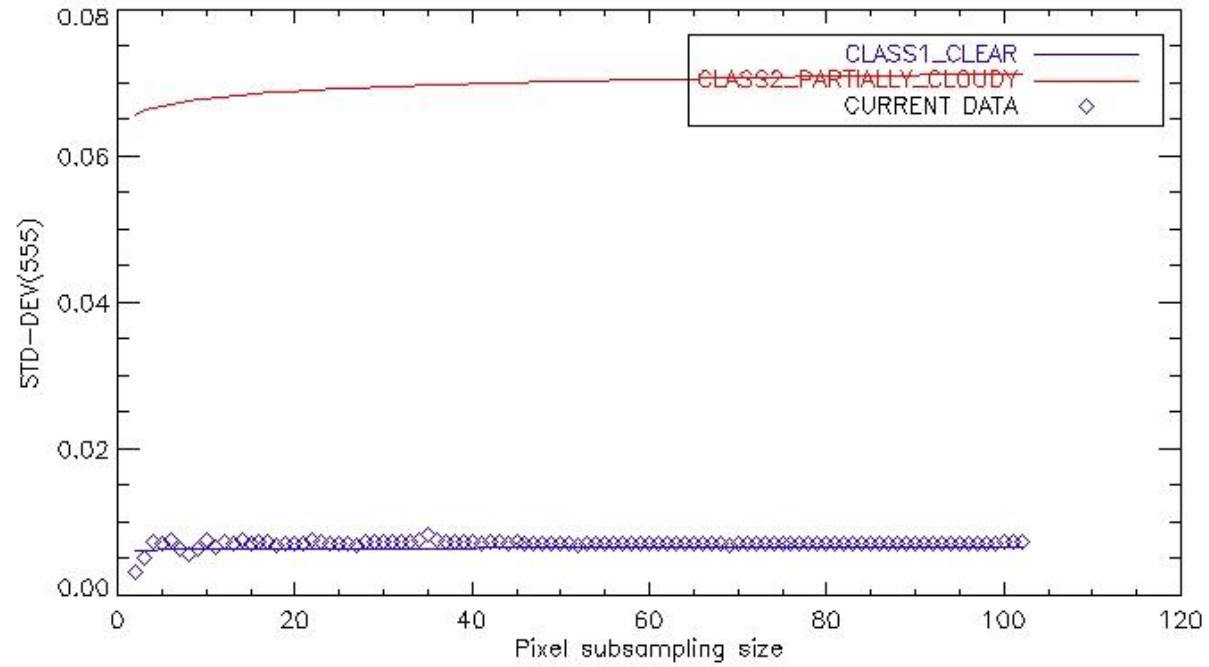


Figure 31 AATSR 20020723 RGB (top) and SSV cloud detection at 555 nm (bottom)

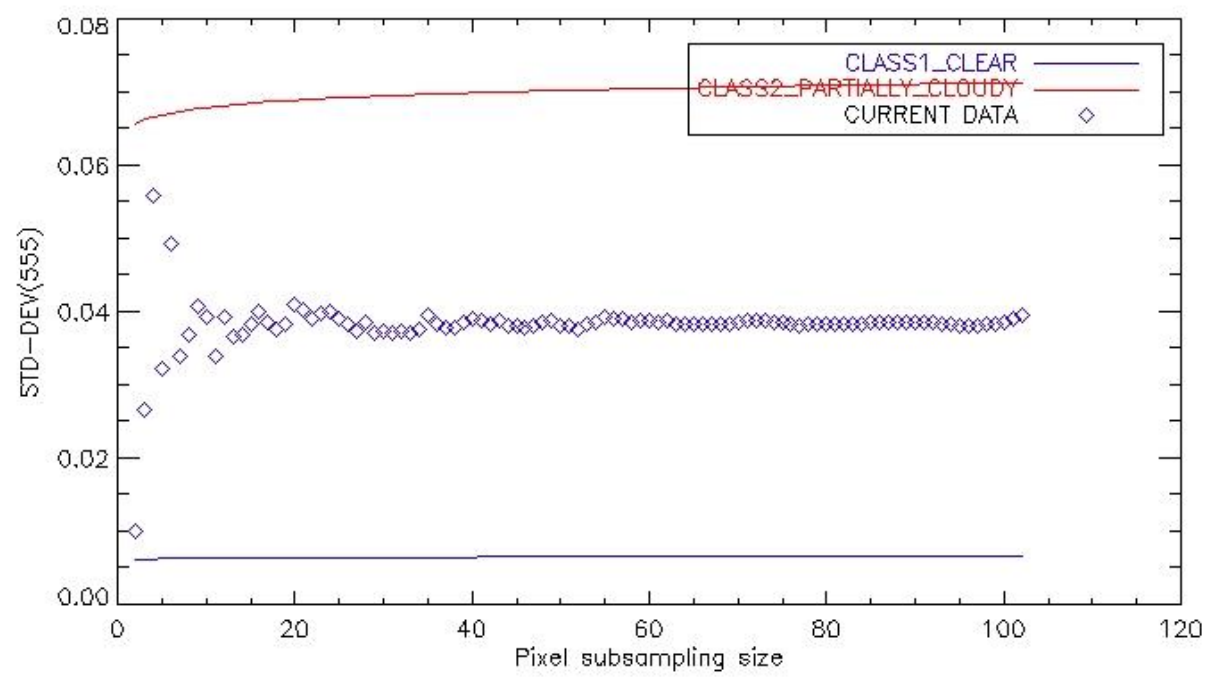
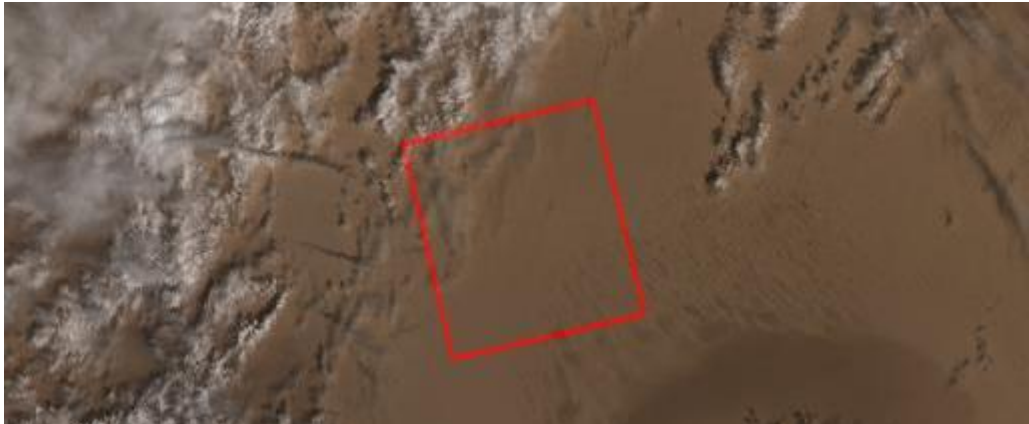


Figure 32 AATSR 20021001 RGB (top) and SSV cloud detection at 555 nm (bottom)



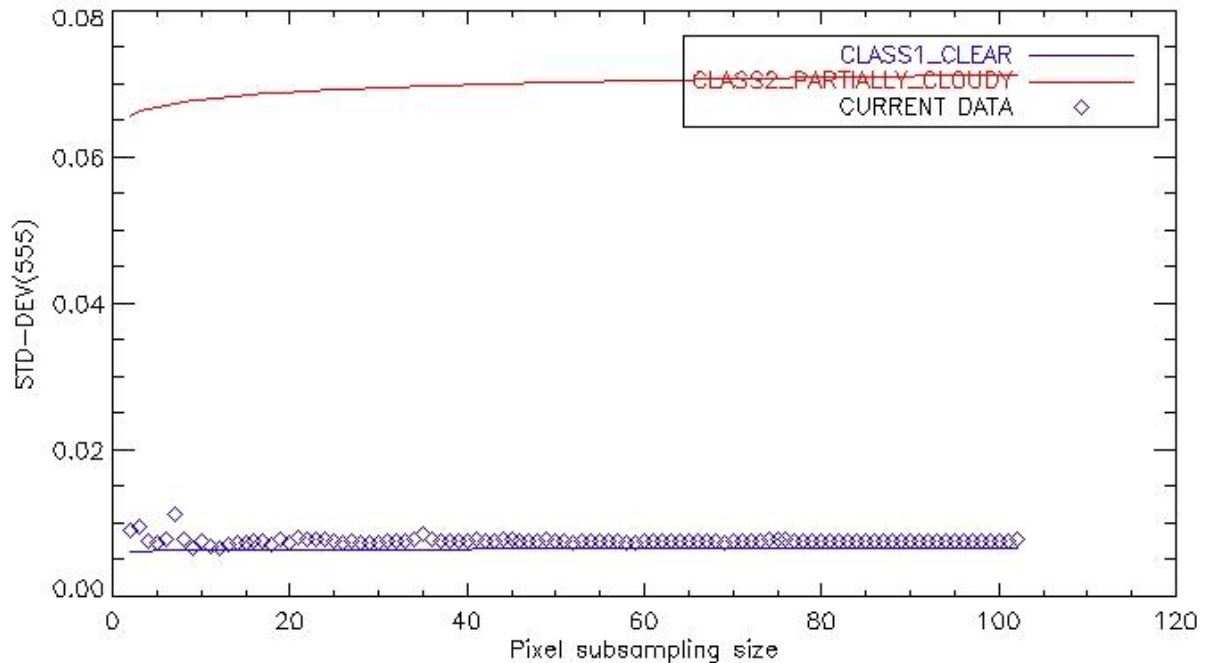


Figure 33 AATSR 20020928 RGB (top) and SSV cloud detection at 555 nm (bottom)

#### 4.2.3 MERIS BRDF cloud screening over Libya4

The method discussed here uses the same clear class training data set previously defined for the SSV method. A large binning period has been chosen (200 days) because the training dataset contains only few scenes. The TOA test is performed at 412 nm with a 2.5% threshold.

TOA signal ratio is plotted on Figure 34 where the colours refer to the cloud identification. The overall performance is plotted on Figure 35 (bottom), with the performance of nominal DIMITRI cloud screening method (AUTO\_CS) reminded on the top figure. The BRDF cloud detection performs very well, with 75% success for cloud detection and 73% success for clear sky.

Furthermore it is worth noting that most of the 25% clear sky scenes apparently wrongly detected (yellow dots on Figure 34) correspond to an erroneous manual screening and are actually cloudy. Examples are provided on Figure 36 and shows that the BRDF method is able to detect very small cloud patterns. This means also that the displayed 100% performance of the nominal algorithm is artificially high.

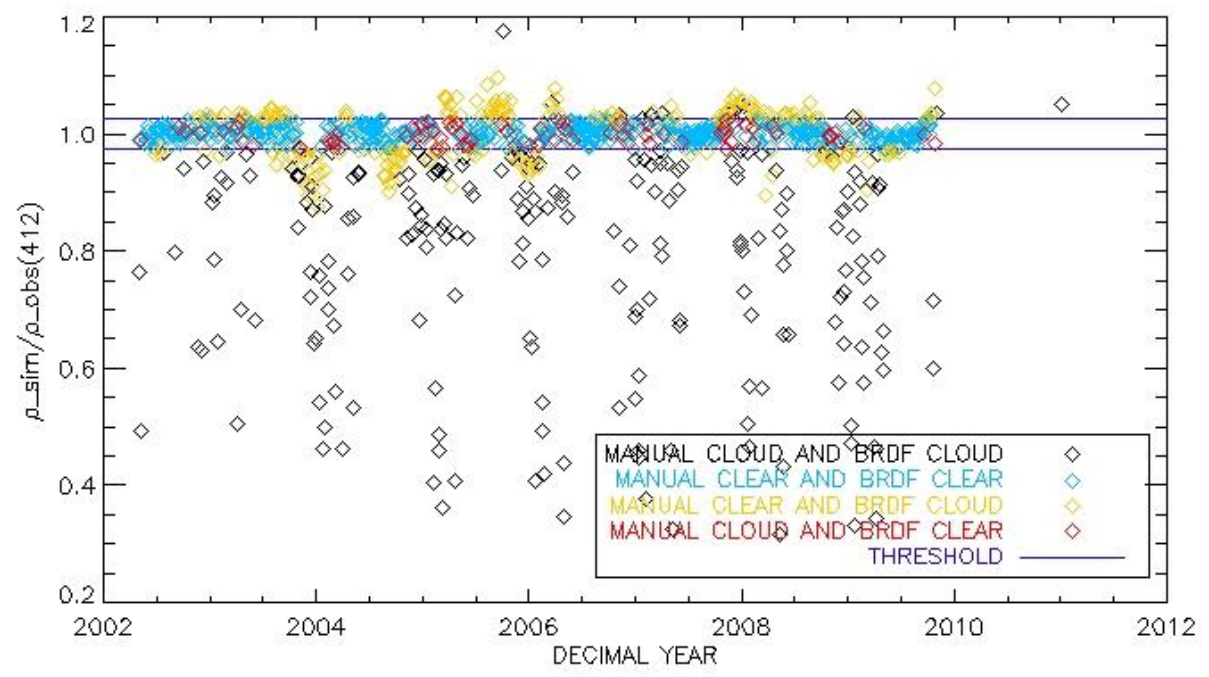


Figure 34 MERIS BRDF cloud screening analysis over Libya4



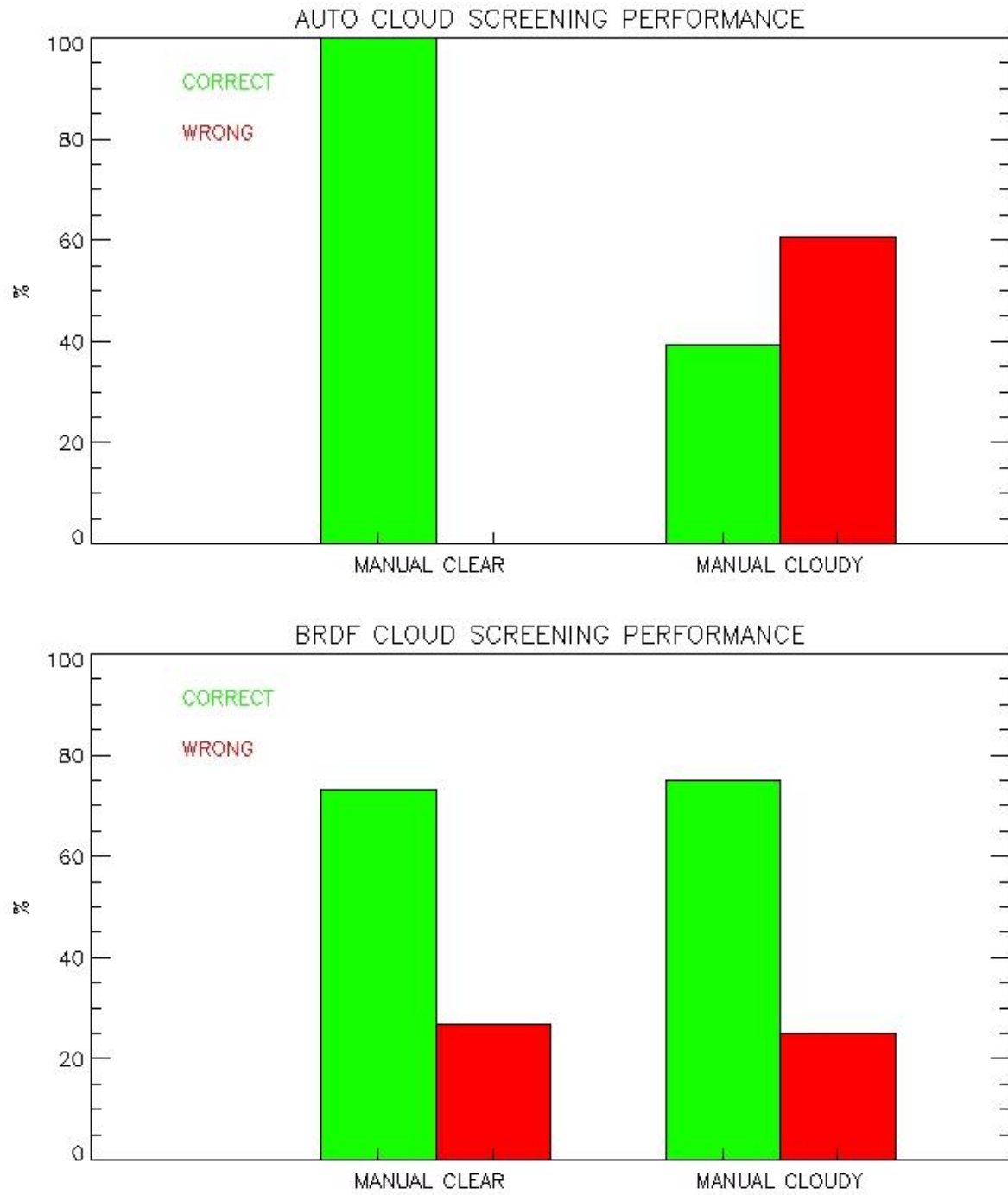



Figure 35 MERIS cloud screening performance over Libya4 for the DIMITRI nominal implementation (top) and BRDF method (bottom). The nominal screening is considered as clear when AUTO\_CS=0 and as cloudy when AUTO\_CS>0

	<p><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p>Reference: MO-SCI-ARG-TN-004a Revision: 1.0 Date: 28/05/2014 Page: 44</p>
---	--	--

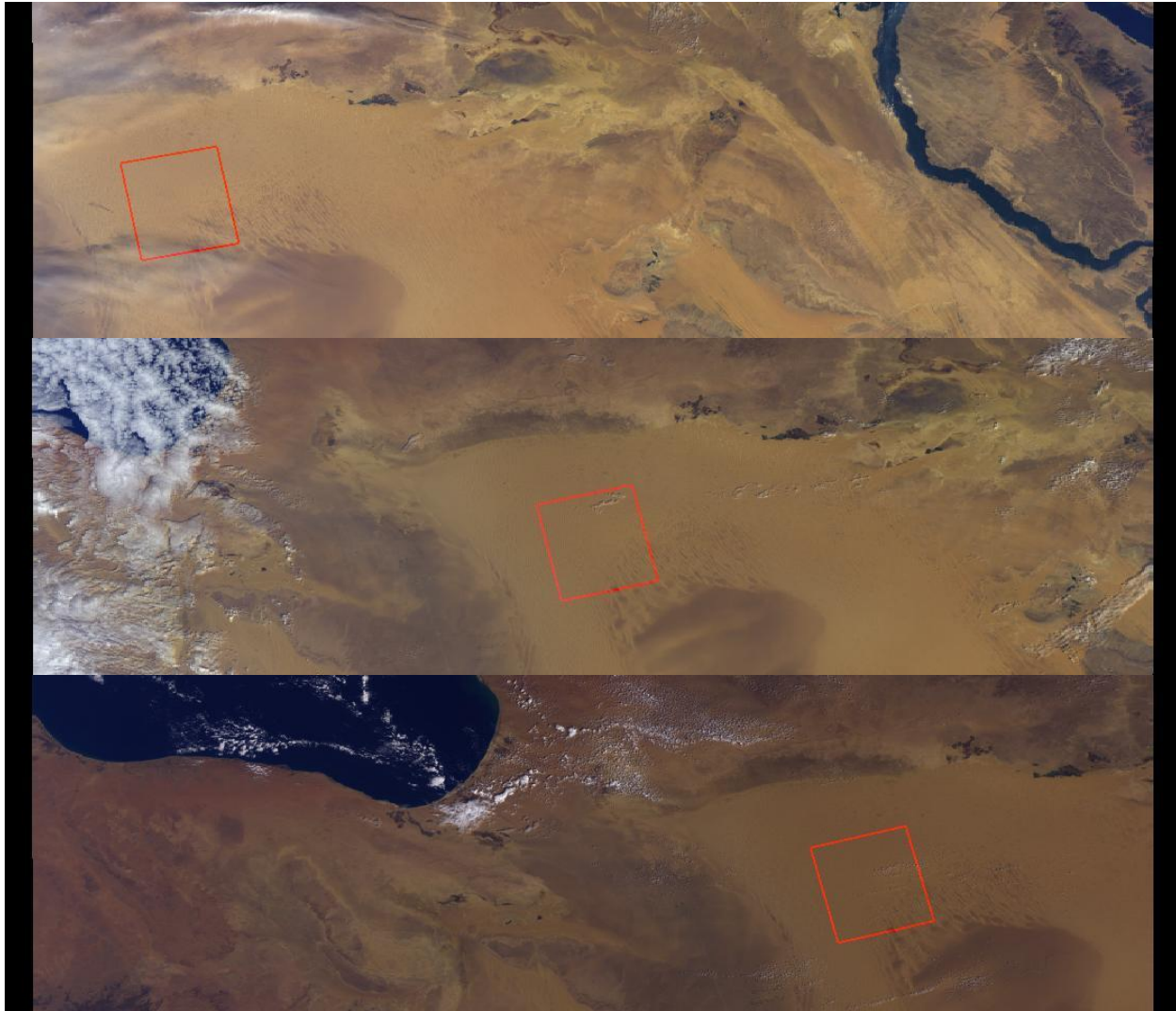


Figure 36 Examples of MERIS wrong manual screening over Libya4 (20060129, 20040203, 20090309 from top to bottom). While the MANUAL\_CS is set to 0 in the DB, small clouds are visible in the ROI.

#### 4.2.4 AATSR BRDF cloud screening over Libya4

The same analysis and conclusion applies to the BRDF cloud screening of AATSR data, computed at band 555nm (see Figure 37 and Figure 38). Again, many scenes are erroneously manually classified as clear in the database but in fact they are cloudy (see two examples on Figure 39) this error artificially reduces the reported performance of the BRDF algorithm over clear site (about 75%, see also yellow points in Figure 37)

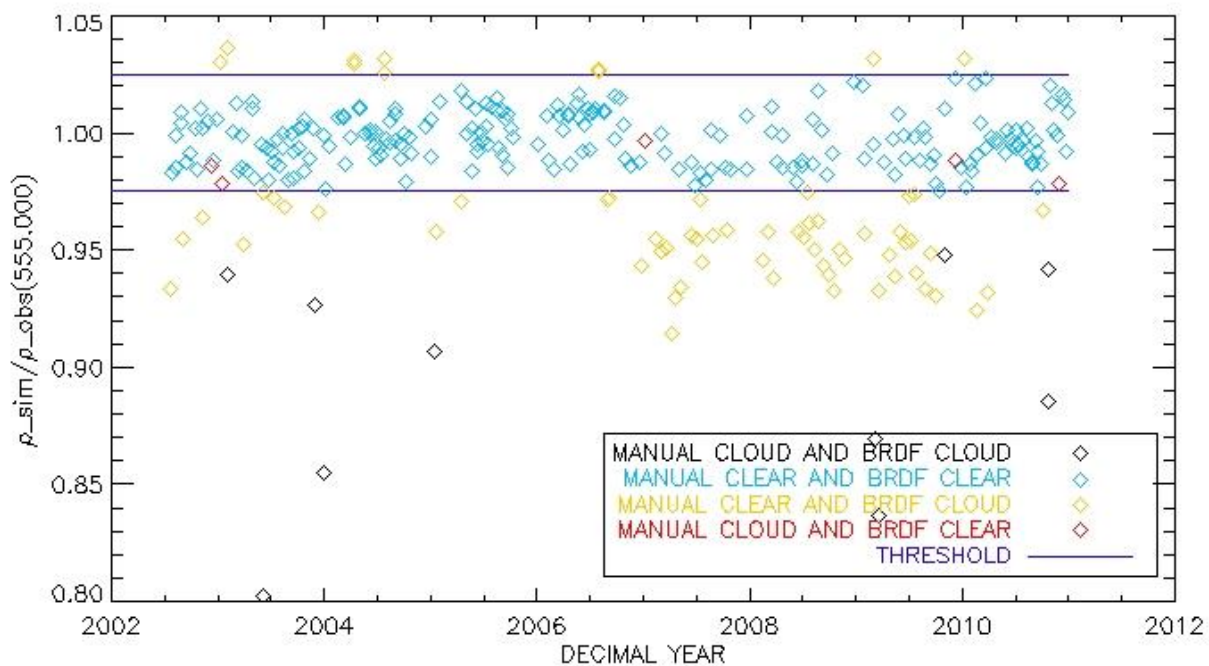
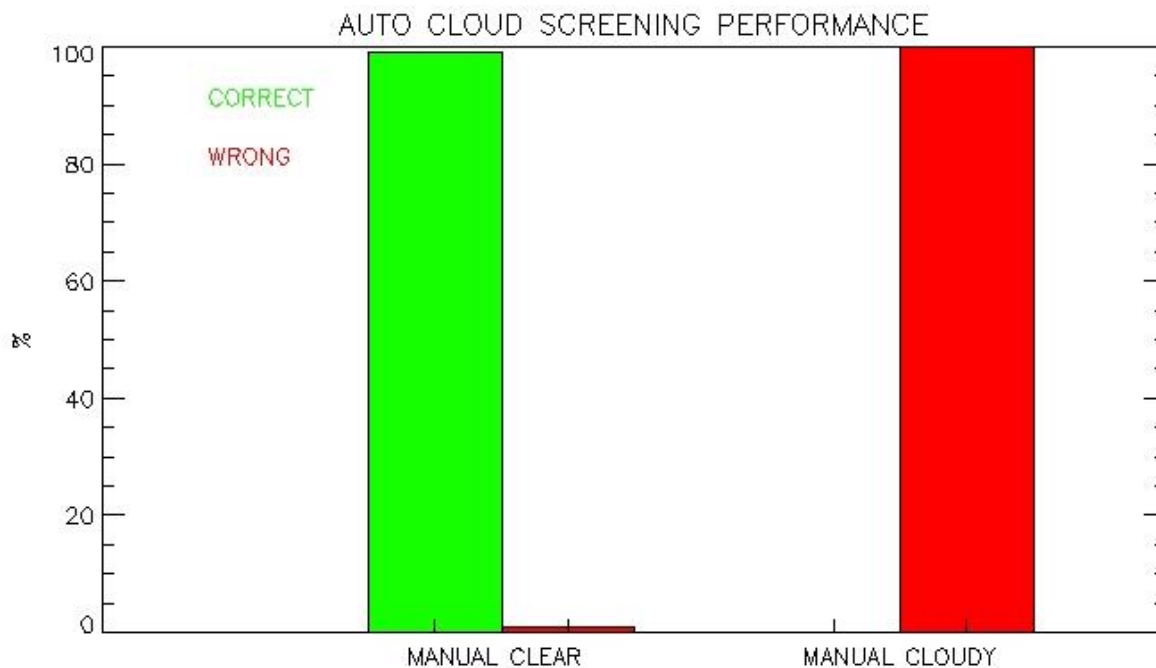


Figure 37 AATSR BRDF cloud screening analysis over Libya4



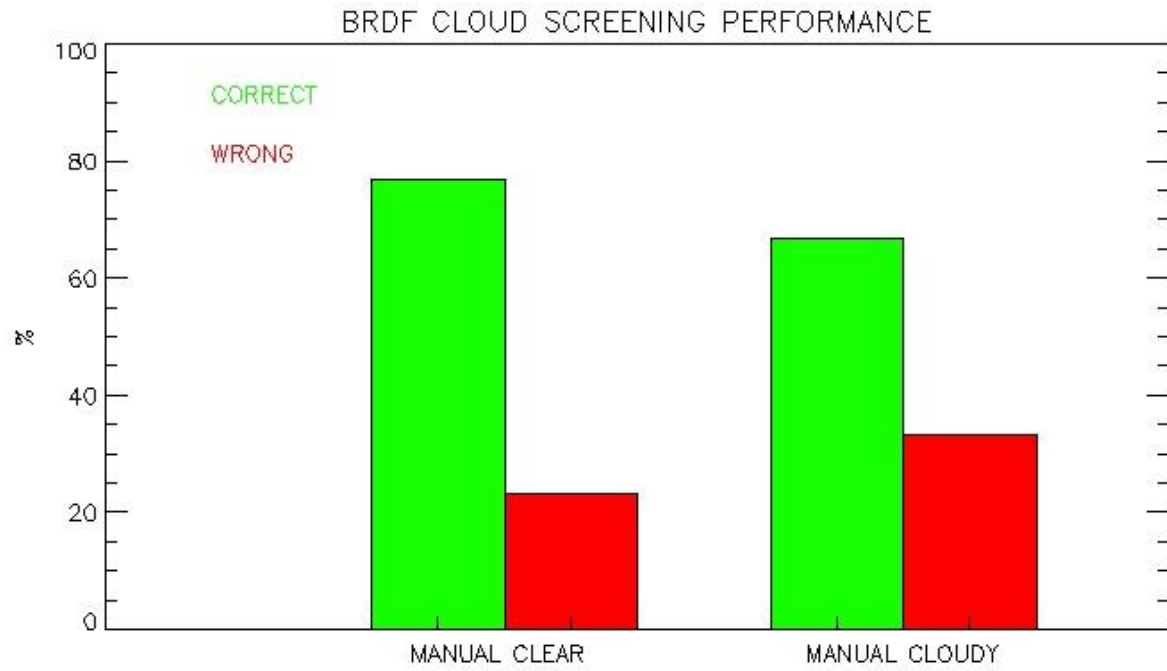



Figure 38 AATSR cloud screening performance over Libya4 for the DIMITRI nominal implementation (top) and BRDF method (bottom). The nominal screening is considered as clear when AUTO\_CS=0 and as cloudy when AUTO\_CS>0

	<p><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 47</p>
---	--	--

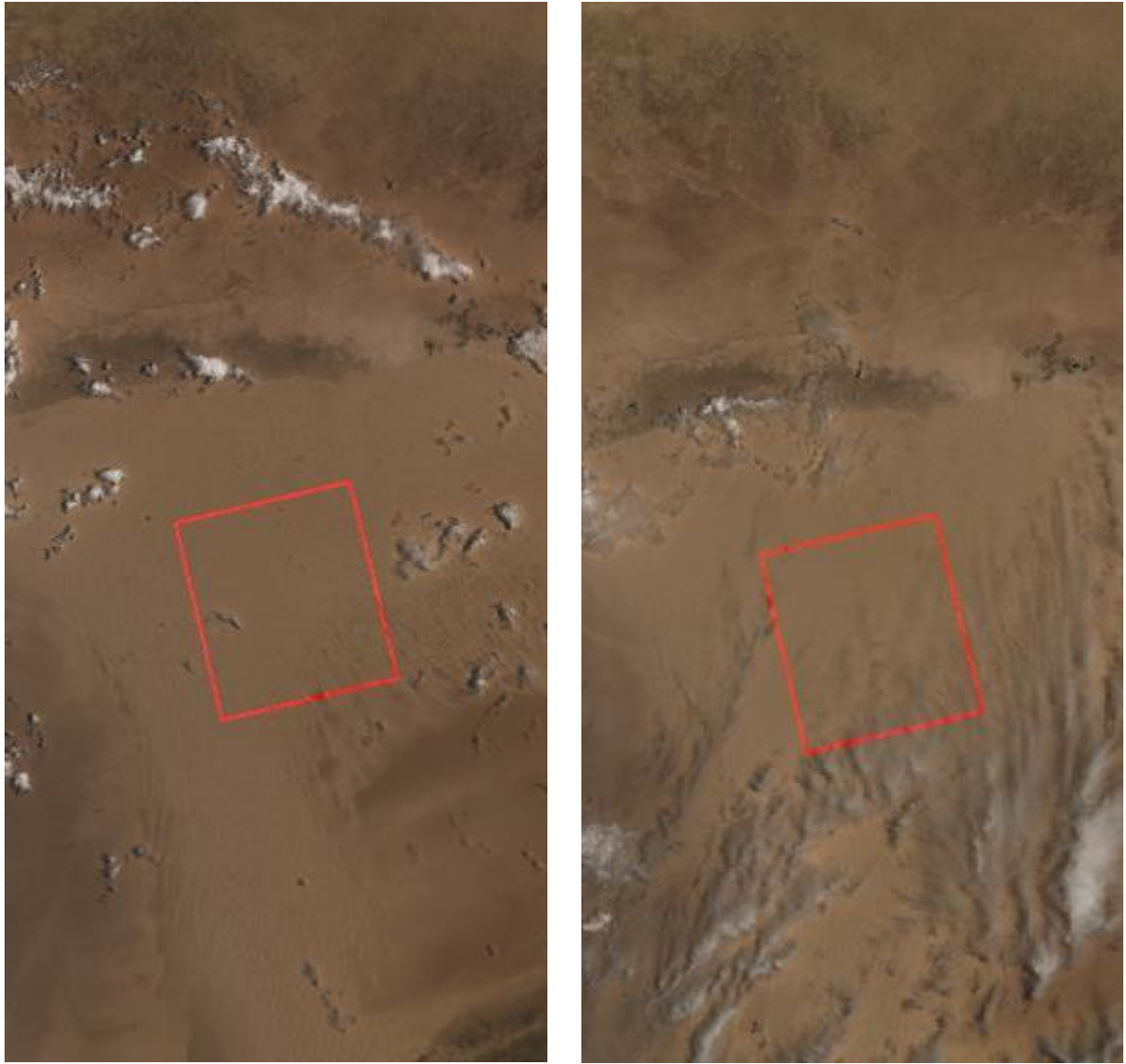



Figure 39 Examples of AATSR wrong manual screening over Libya4 (left 20030622 and right 20050607). While the MANUAL\_CS is set to 0 in the DB, small clouds are visible in the ROI.

	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 48</p>
---	---	--

## 5 Discussion and Conclusions


Further analysis of the current DIMITRI database is required before a robust investigation of the cloud screening methodologies can be completed. During the development of the new methodologies, it was found that the current version of the database has many misclassified manually screened scenes. This makes it difficult to ‘truth’ the models using a large series of images. It is believed that one possible reason for the misclassifications of the manually screened scenes is that DIMITRI V2 drew a red shaded box over the ROI. This made it difficult to see by eye, especially at DomeC where there was the largest found number of misclassified images. DIMITRI V3 subverts this problem with the new ingestion routine. The new routine simply draws an un-shaded box around the ROI making it possible to see into the ROI. It is therefore recommended that the manual cloud screening be redone. As the database is getting quite large, it is recommended that a standard subset be defined and the manual cloud screening be redone. The subset should include a more even number of images across the different sites and sensors than the current manually screened set. It is believed that the new ingestion routine will help fix currently incorrectly screened images.

There is currently no analysis done over the DomeC site. This is due to the very high misclassification rate stated previously. This makes it to impossible automatically generate a training set based on the current DIMITRI database as there is no confidence that the subset will be valid. Further to this, even with the new ingestion routine improving the clarity over the ROI, DomeC still remains a very difficult site to manually cloud screen, especially for partly cloudy scenes.

It was discovered in the SSV method that the power law that was fitted to the standard deviation as a function of window size may not be the best function to fit to. In fact, a linear fit may be more appropriate. It was found that that for the very low resolutions that the variance is mostly noise and this can be a challenge for the fitting routine to fit to accurately. The curve very quickly converges to a straight line. For this reason it may be better to simply take the absolute standard deviation over the entire scene because the changing window size doesn’t appear to give any more information about the scene. It is recommended that this be investigated further in future evolutions of DIMITRI.


The SSV method has very high computational requirements. Training the model can take quite some time and so does classifying the images. The method often has trouble discriminating between cloudy and partly cloudy, particularly over marine targets. The method can however, provide information on whether or not the image has some cloud in it. As a binary classifier it performs quite well on the test dataset. However, the advantage that was foreseen for this



	<p align="center"><b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening</p>	<p><b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 49</p>
---	---	--

method was that it was a ternary classifier. The BRDF method works better and faster if a binary classifier is sufficient.

The BRDF cloud screening method is very useful at identifying scenes with cloud. The BRDF method is able to detect very small amounts of cloud in an ROI even when the training dataset for BRDF computation is scarce (e.g. 15 scenes for AATSR used here, with a binning period of 200 days). Contrary to the SSV cloud screening, it has even helped us to identify scenes erroneously labelled as manually clear in the DB. Hence this approach could be a good candidate to complete the DIMITRI nominal cloud screening run at ingestion time.

	<b>DIMITRI_v3.0 ATBD [01]</b> Automated Cloud Screening	<b>Reference:</b> MO-SCI-ARG-TN-004a <b>Revision:</b> 1.0 <b>Date:</b> 28/05/2014 <b>Page:</b> 1
---	--	---

## 6 Appendix

### 6.1 Cloud screening statistics



