Applied Spectroscopy

The novel tool of Cumulative Discriminant Analysis applied to IASI cloud detection

G. Masiello, C. Serio, S. Venafra,

SI/UNIBAS, School of Engineering, University of Basilicata, Potenza, Italy

U. Amato, I. De Feis

IAC/CNR, Ist. Applicazione del Calcolo, Napoli, Italy

L. Lavanant

Centre de Meteorologie Spatiale, Lannion, France

R. Stuhlmann, S. Tjemkes

EUMETSAT, Darmstadt, Germany

Outline

1. Description of Test Statistics
   - Thermal Contrast
   - Shape
2. Methodology to combine statistics in a single cloud mask
   - CDA
   - N-Dimensions
3. Results from the analysis with two globally distributed IASI database
   - D_DIFA1 and D_DIFA2
4. Conclusions and Future development

Cloud detection basics

- In the thermal infrared cloud detection relies on the fact that normally clouds are colder than the underlying surface (thermal contrast)
- Clear sky spectra tend to be different in shape of those in cloudy conditions

Cloud detection scheme: The baseline

Test Statistic	Method
h_s	Based on shape similarity between a couple of spectra (Observed and Reference)
Surface temperature tests ***χ^2***	Based on χ^2 –like variable defined on a couple of skin temperature values (T_s, T_s^R), with T_s, estimated from the spectrum and T_s^R a suitable reference
T_0	Based on the Brightness Temperature, T_0 @ 833 cm^{-1} contrasted against a suitable threshold
Window slope test ***W***	Based on the difference T_2-T_0, with T_2, the Brightness Temperature @ 900 cm^{-1}
ΔT_{CO2}	Split window test based on CO_2 Q-branch @ 791 cm^{-1}, $\Delta T_{CO2}=T_4-T_5$, with T_4 and T_5 the brightness temperature @ 790.5 cm^{-1} and 791.75 cm^{-1}
Spatial Coherence test ***sh***	Based on a cluster of $n \times n$ nearby pixels. Let $\{T_0\}_{i,\ i=1,..,nxn}$, The set of T_0 corresponding to a given $n \times n$ clusters, *sh* is the standard deviation of this set.

1. Description
2. Methodology
3. Results
4. Conclusion

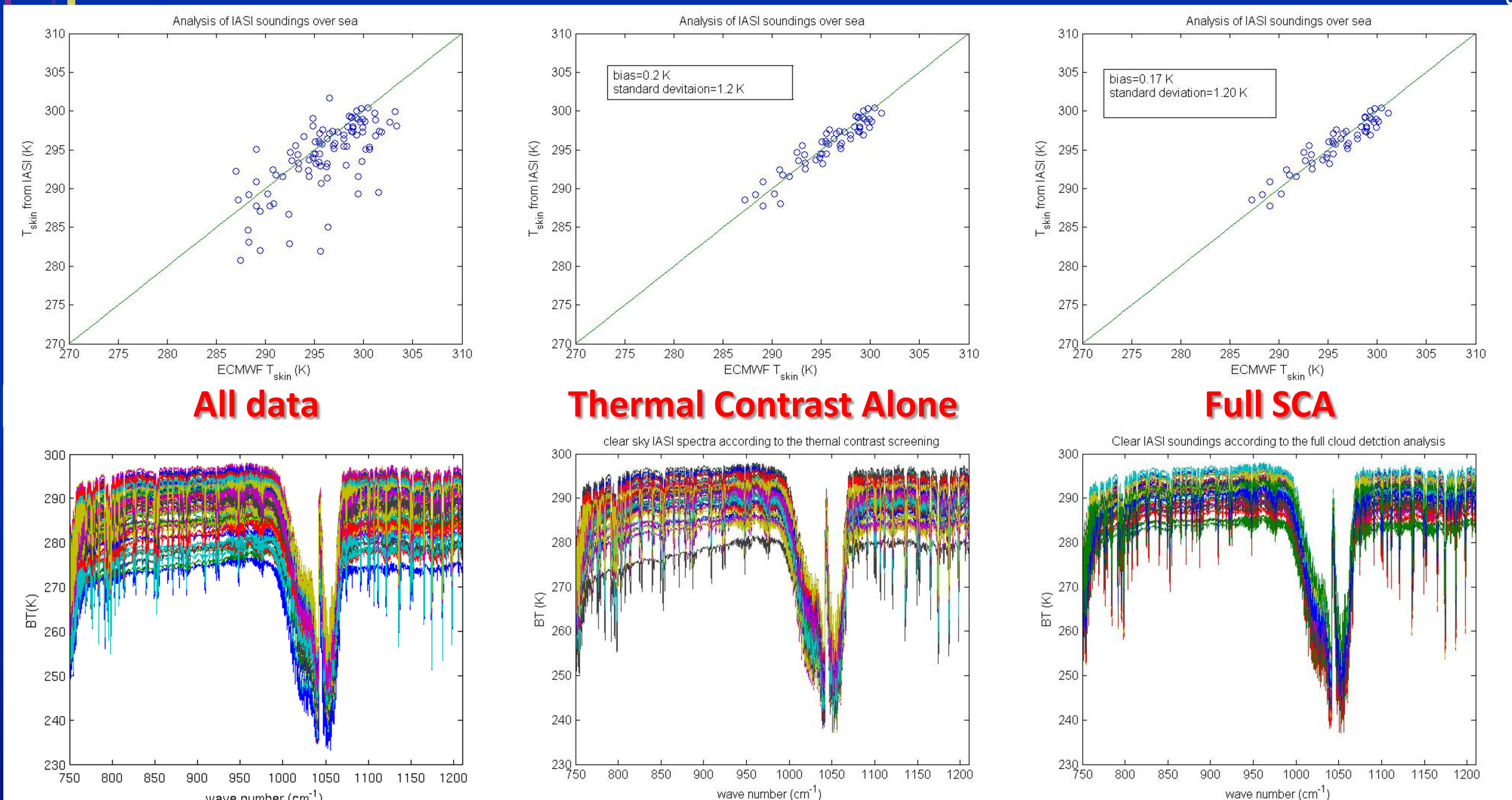
Type of statistics

- T_0: window brightness temperature, BT @ 833 cm^{-1}
- ΔT_{CO2} split test (sensitive to surface pressure as well)
- Normalized thermal contrast
  χ^2 test

- h_s test
- window slope test, W
  0 K < BT(900)-BT(833) <1.5 K
- Spatial Coherence *sh*

Thermal Contrast **Shape & Morphology**

Udine exercise

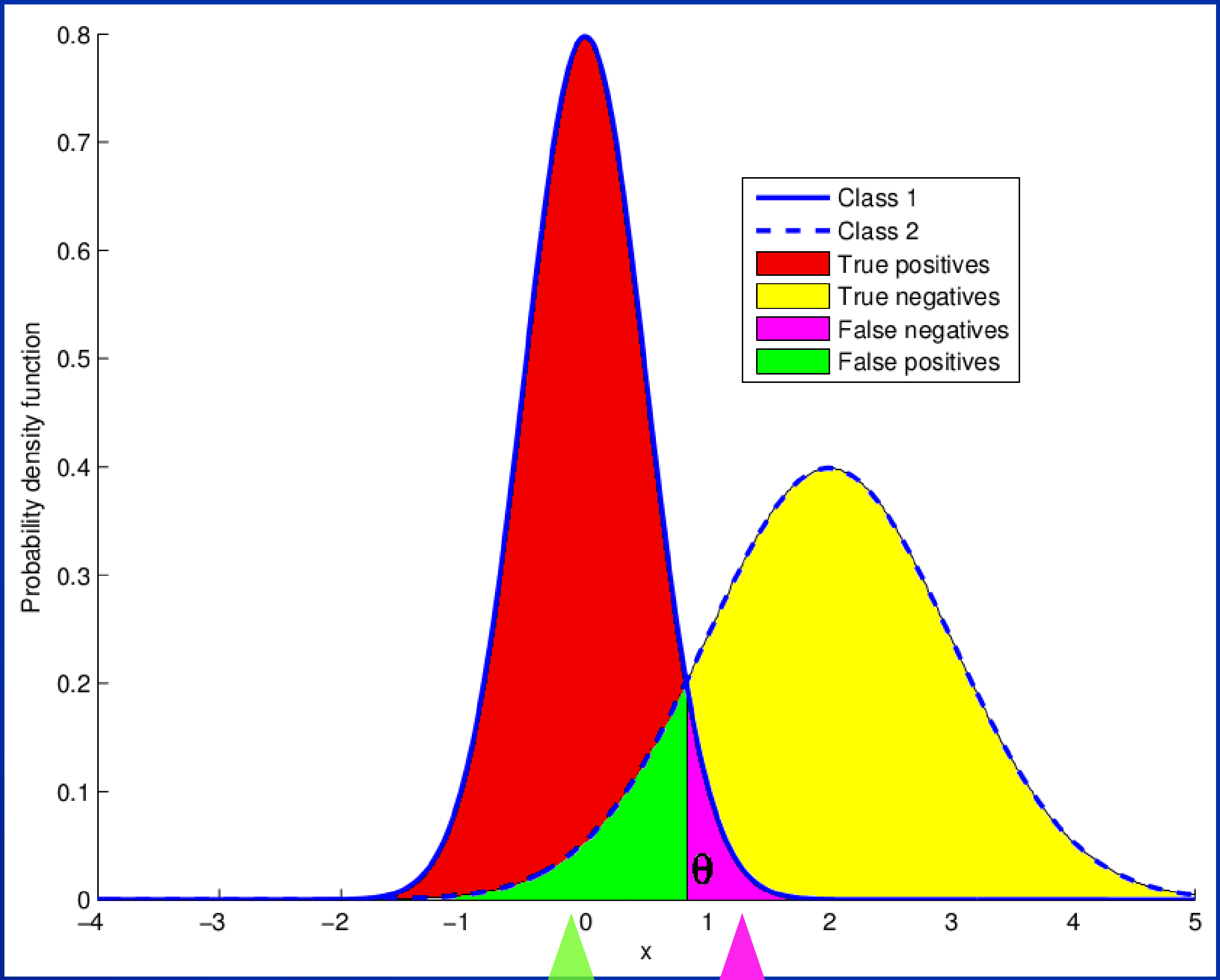


Upper panels show the scatter plot of ECMWF Skin Temperature against the one derived from the spectra.

Morphology Tests detect cirrus cloud contaminated spectra classified clear by Thermal Contrast tests alone.

Discriminant analysis

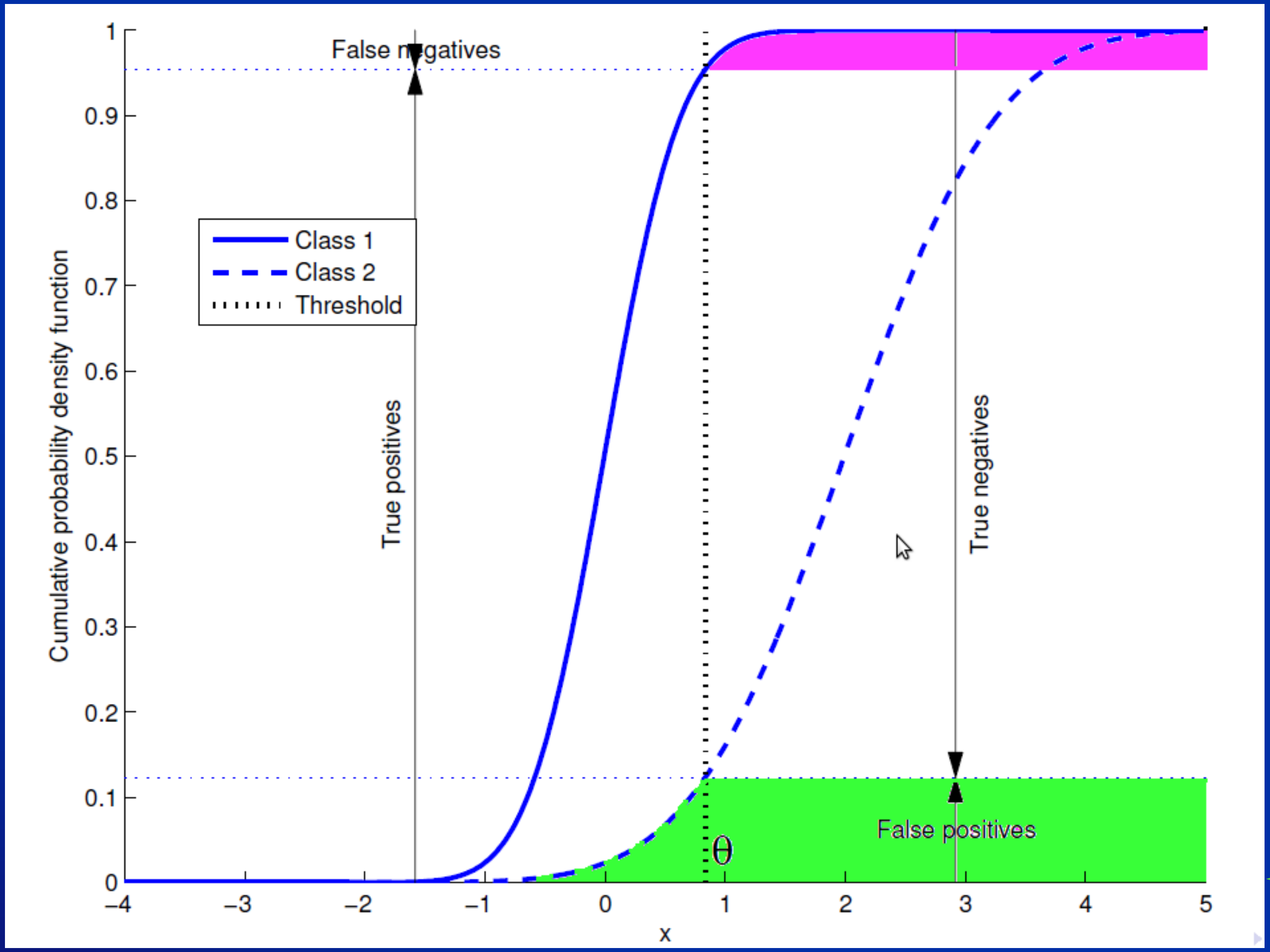
Discriminant Analysis and more generally classification methods formally rely on the minimization of an objective (cost) function that penalizes misclassifications of the training dataset.



I-type error is defined as the fraction of pixels being Clear and retrieved as Cloudy (False negatives).

II-type error the fraction of pixels being Cloudy and classified as Clear (False positives).

I and II type errors in terms of the cumulative density function



I-type error is defined as the fraction of pixels being Clear and retrieved as Cloudy (False negatives).

II-type error the fraction of pixels being Cloudy and classified as Clear (False positives).

A novel approach: Cumulative Discriminant Analysis (CDA)

- CDA is based on the cumulative distribution function
- CDA is based on the following requirements:
  - Minimization of both errors of the I and II type
  - Unbiased nonparametric estimate
- Advantages
  - Fully non parametric estimates of the cdf
    - ECDF, Empirical Cumulative Density Function
  - Consistent estimator (surely converges to the true cumulative density function)
  - Uniform convergence over x

Basic of the methodology

$$\Gamma(x, \vartheta) = \begin{cases} 1 \text{ (Clear)} & \text{if } x \leq \vartheta \\ 2 \text{ (Cloudy)} & \text{if } > \vartheta, \end{cases}$$

$$E^{\mathrm{I}} = 1 - F^{\mathrm{Clear}}(\vartheta)$$

$$E^{\mathrm{II}} = F^{\mathrm{Cloudy}}(\vartheta)$$

Optimal threshold ϑ:

$$\hat{\vartheta} = \underset{\vartheta}{\operatorname{argmin}}\, \mathcal{C}(\mathbf{x}^{\mathrm{training}}, \vartheta)$$

- Worst classification among Clear and Cloudy conditions

$$\mathcal{C}(\mathbf{x}^{\mathrm{training}}, \vartheta) = \max(E^{\mathrm{I}}, E^{\mathrm{II}})$$

Cost Function

Basic of the methodology
Two Dimensions

I-type Error

$$E^{\mathrm{I}} = \underbrace{\int_{-\infty}^{\vartheta_1}\int_{-\infty}^{\infty} f^{\mathrm{Clear}}(x_1,x_2)\,dx_1\,dx_2}_{a+b} + \underbrace{\int_{\vartheta_1}^{\infty}\int_{-\infty}^{\vartheta_2} f^{\mathrm{Clear}}(x_1,x_2)\,dx_1\,dx_2}_{c}$$

$$= F_1^{\mathrm{Clear}}(\vartheta_1) + F_2^{\mathrm{Clear}}(\vartheta_2) - F_1^{\mathrm{Clear}}(\vartheta_1)F_2^{\mathrm{Clear}}(\vartheta_2)$$

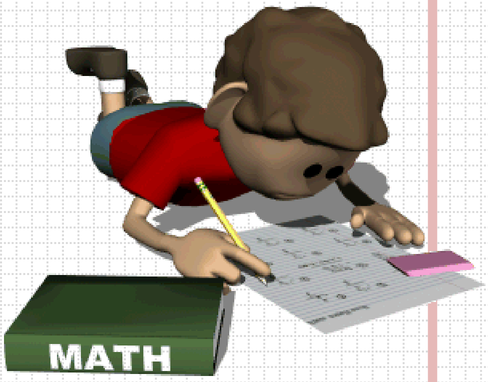
II-type Error

$$E^{\mathrm{II}} = 1 - \left(F_1^{\mathrm{Cloudy}}(\vartheta_1) + F_2^{\mathrm{Cloudy}}(\vartheta_2) - F_1^{\mathrm{Cloudy}}(\vartheta_1)F_2^{\mathrm{Cloudy}}(\vartheta_2)\right)$$

Then

$$E^{\mathrm{I}} = 1 - F_1^{\mathrm{Clear}}(\vartheta_1)F_2^{\mathrm{Clear}}(\vartheta_2)$$

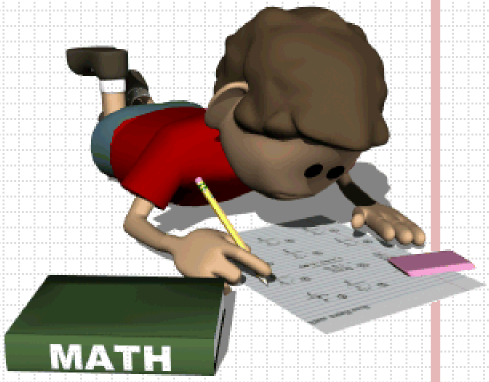
$$E^{\mathrm{II}} = F_1^{\mathrm{Cloudy}}(\vartheta_1)F_2^{\mathrm{Cloudy}}(\vartheta_2).$$



Basic of the methodology: Three Dimensions

- $D = 3$

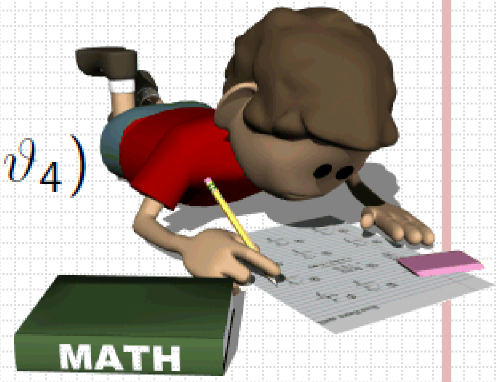
$$
\begin{aligned}
E^{\mathrm{I}} = & F_1^{\mathrm{Clear}}(\vartheta_1) + F_2^{\mathrm{Clear}}(\vartheta_2) + F_3^{\mathrm{Clear}}(\vartheta_3) \\
& - F_1^{\mathrm{Clear}}(\vartheta_1)F_2^{\mathrm{Clear}}(\vartheta_2) - F_1^{\mathrm{Clear}}(\vartheta_1)F_3^{\mathrm{Clear}}(\vartheta_3) \\
& - F_2^{\mathrm{Clear}}(\vartheta_2)F_3^{\mathrm{Clear}}(\vartheta_3) \\
& + F_1^{\mathrm{Clear}}(\vartheta_1)F_2^{\mathrm{Clear}}(\vartheta_2)F_3^{\mathrm{Clear}}(\vartheta_3)
\end{aligned}
$$



Basic of the methodology: Four and more Dimensions

- $D = 4$

$$
\begin{aligned}
E^{\mathrm{I}} = & F_1^{\text{Clear}}(\vartheta_1) + F_2^{\text{Clear}}(\vartheta_2) + F_3^{\text{Clear}}(\vartheta_3) + F_4^{\text{Clear}}(\vartheta_4) \\
& - F_1^{\text{Clear}}(\vartheta_1)F_2^{\text{Clear}}(\vartheta_2) - F_1^{\text{Clear}}(\vartheta_1)F_3^{\text{Clear}}(\vartheta_3) \\
& - F_1^{\text{Clear}}(\vartheta_1)F_4^{\text{Clear}}(\vartheta_4) - F_2^{\text{Clear}}(\vartheta_2)F_3^{\text{Clear}}(\vartheta_3) \\
& - F_2^{\text{Clear}}(\vartheta_2)F_4^{\text{Clear}}(\vartheta_4) - F_3^{\text{Clear}}(\vartheta_3)F_4^{\text{Clear}}(\vartheta_4) \\
& + F_1^{\text{Clear}}(\vartheta_1)F_2^{\text{Clear}}(\vartheta_2)F_3^{\text{Clear}}(\vartheta_3) \\
& + F_1^{\text{Clear}}(\vartheta_1)F_2^{\text{Clear}}(\vartheta_2)F_4^{\text{Clear}}(\vartheta_4) \\
& + F_2^{\text{Clear}}(\vartheta_2)F_3^{\text{Clear}}(\vartheta_3)F_4^{\text{Clear}}(\vartheta_4) \\
& - F_1^{\text{Clear}}(\vartheta_1)F_2^{\text{Clear}}(\vartheta_2)F_3^{\text{Clear}}(\vartheta_3)F_4^{\text{Clear}}(\vartheta_4)
\end{aligned}
$$



- and so on for higher D.

DIFA Database for training, cross-check and validation

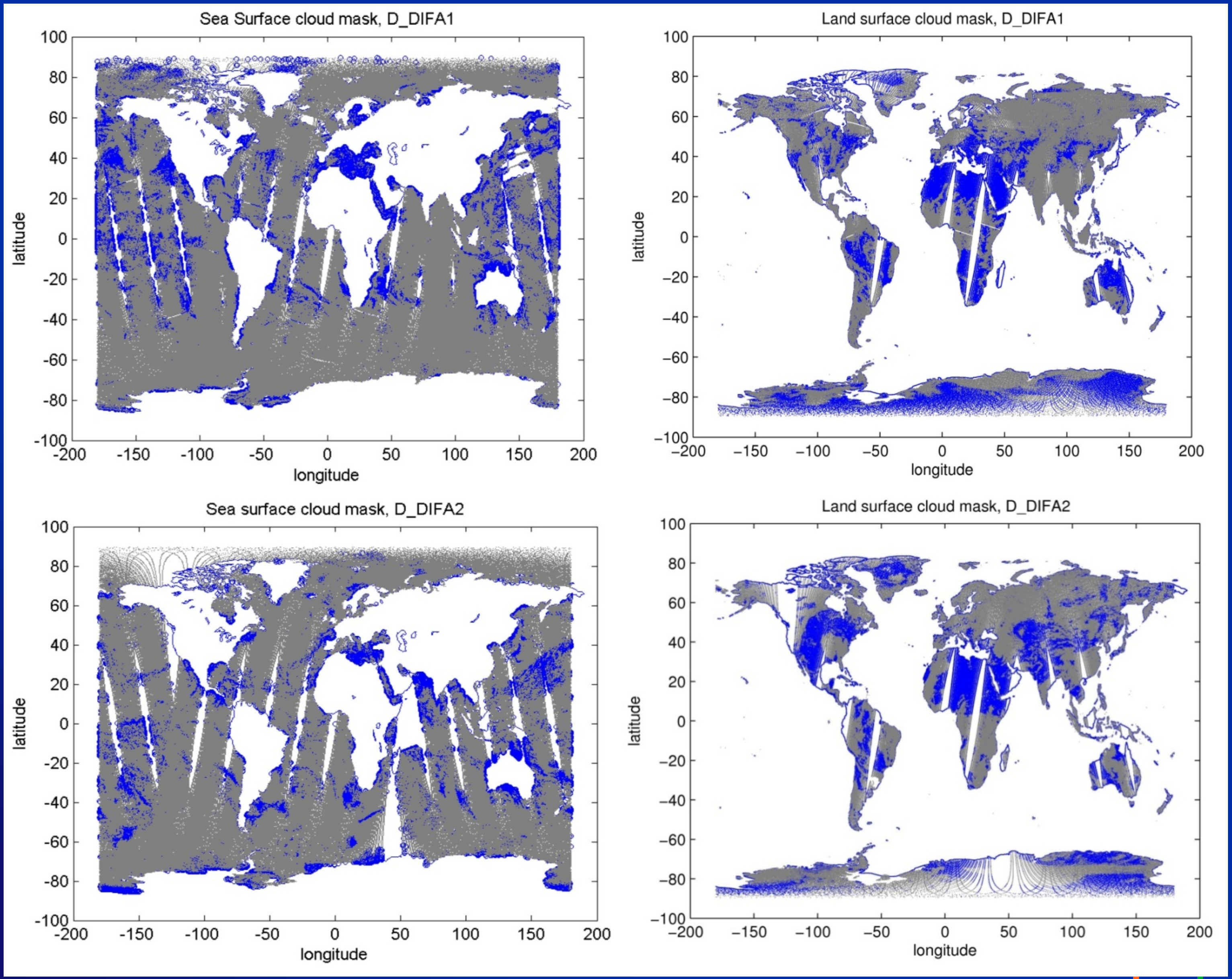
- We identified and developed two data sets of IASI spectra that are:
  - Qualified for sky-type (clear/cloudy) through the CMS cloud mask, which is based on the co-location of the IASI footprint with AVHRR imagery.
  - Complemented with skin temperature fields from the ECMWF analysis, co-located in space and in time with IASI footprints.
- D_DIFA1:
  - 640000 IASI spectra (8 orbits)
  - 12 h global acquisition on 18 November 2009
- D_DIFA2:
  - 800000 IASI spectra (9 orbits)
  - 15 h global acquisition on 22 July 2007

Cloud mask D_DIFA1 & D_DIFA2 (reference AVHRR cloud mask, Grey Cloudy, Blue Clear)

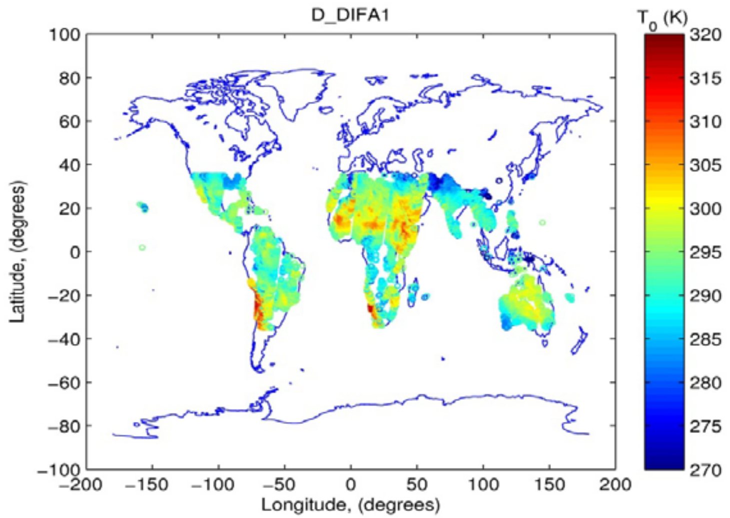
Sea | **Land**

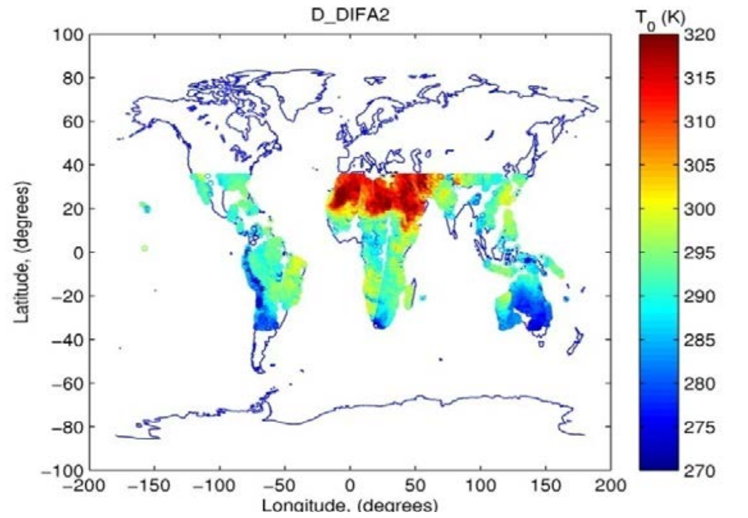
D_DIFA1, 18/11/09

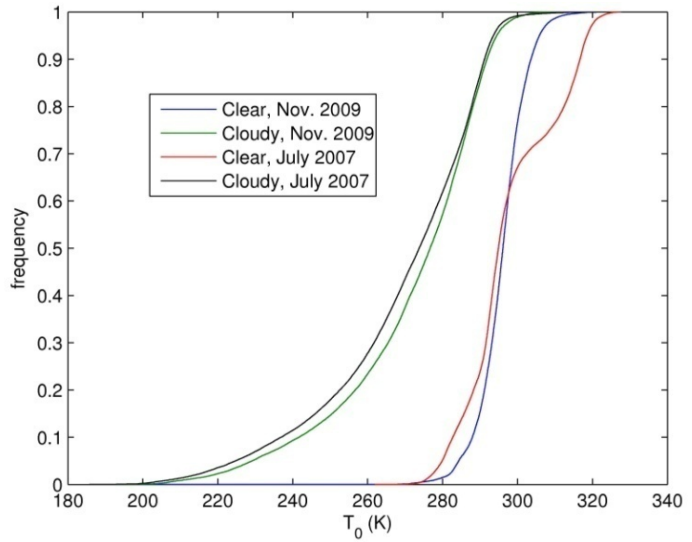
D_DIFA2, 22/07/07

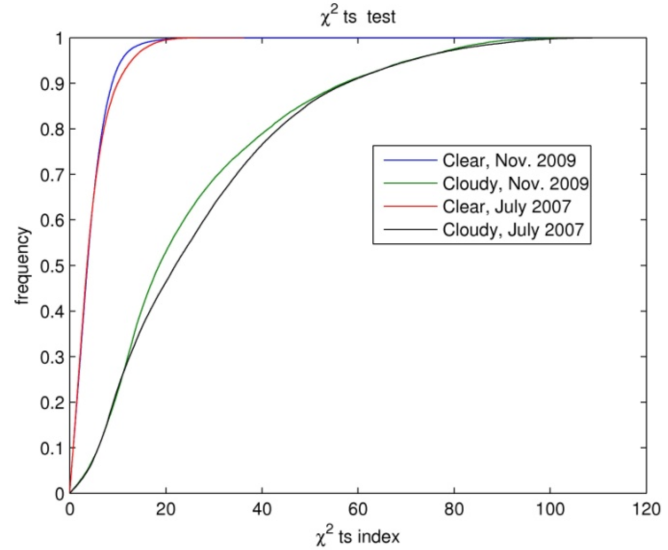


Tropical, land surface









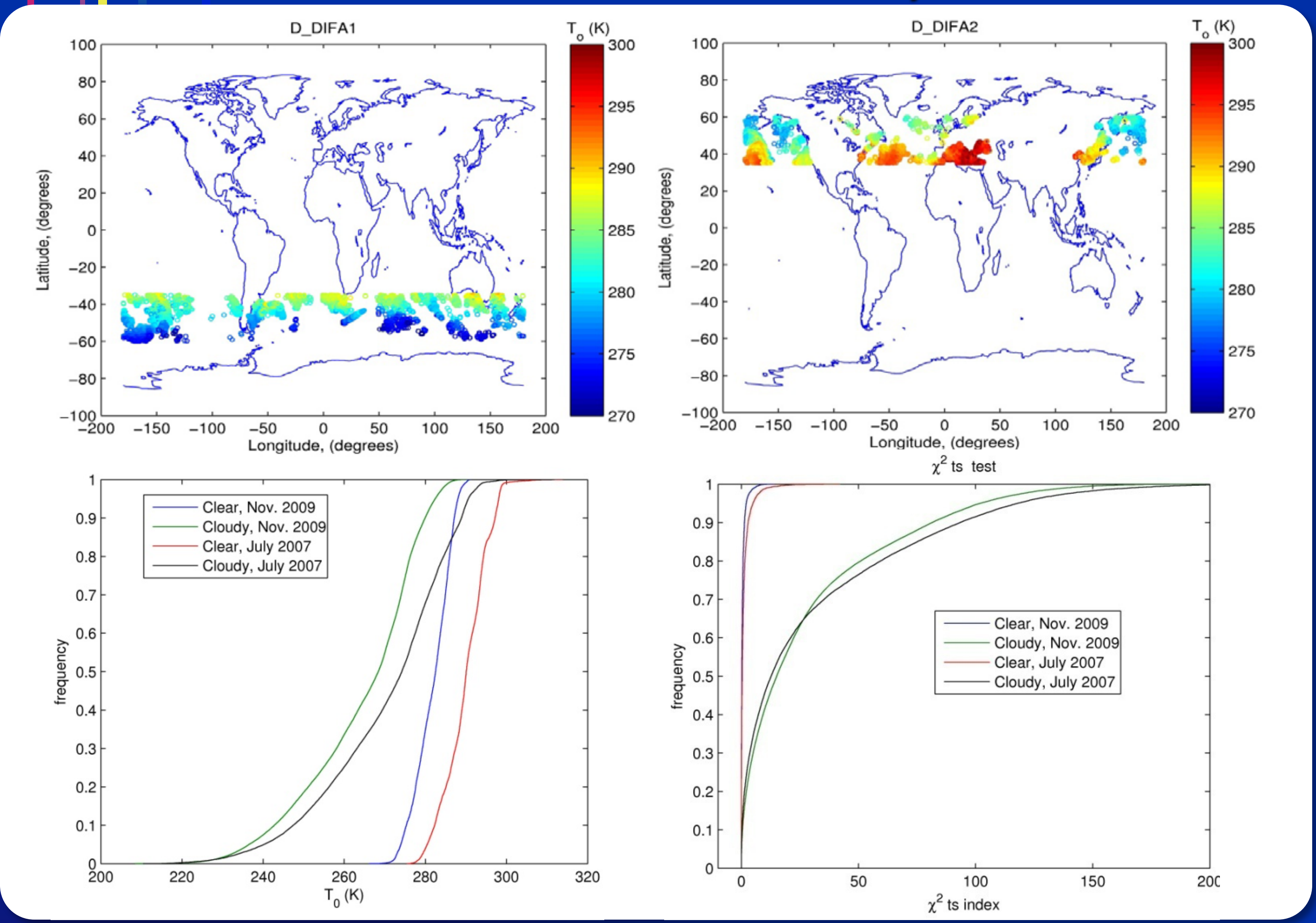
Total Score

$$\frac{n_{11}+n_{22}}{n_{12}+n_{21}} =$$

84.4 %

CMS	DIFA-SCA (D_DIFA1 + D_DIFA2)		
	Clear	Cloudy	Total
Clear	25168+ 30152 (**83.9 %**)	3465+7182 (**16.1 %**)	28633+37334 (**100 %**)
Cloudy	5551+7693 (**15.4 %**)	30543+42224 (**84.6%**)	36094+49917 (**100 %**)

Mid-Lat Summer, sea surface



Total Score

$$\frac{n_{11}+n_{22}}{n_{12}+n_{21}} =$$

86.1 %

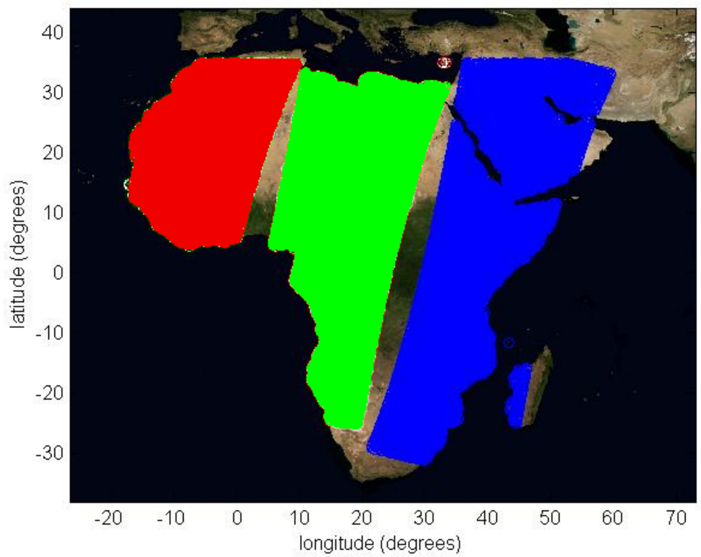
CMS	DIFA-SCA (D_DIFA1 + D_DIFA2)		
	Clear	Cloudy	Total
Clear	6690+ 5695 (**84.2 %**)	1060+1276 (**15.9 %**)	7750+6971 (**100 %**)
Cloudy	8684+5915 (**13.6 %**)	65295+27300 (**86.4%**)	73979+33215 (**100 %**)

Summary of results: D_DIFA1

Set: D_DIFA1, Year 2009

Climate Zone	Clear	Cloudy	Best Performance Physical Space Total score (%)
Land			
1: Tropics	28693	36108	86.0
2: Mid-Lat North	9093	12931	84.5
3: Mid-Lat South	187	2210	89.0
4: High-lat North	1	17	
5: High-lat South	47	1166	87.3
7: Ice-snow <1km	13839	49761	71.0
8: Ice-snow >1km	20844	26040	59.3
Sea			
1: Tropics	35095	151891	81.3
2: Mid-lat North	7750	73979	88.1
3: Mid-lat South	3293	37648	89.6
4: High-lat North	56	2288	92.0
5: High-lat South	1018	15233	93.4
6: Sea Ice	6835	106118	60.9

D_DIFA2 data set for AFRICA
Validation with SEVIRI Cloud Mask



SEVIRI Cloud mask	IASI-DIFA cloud mask		
	Clear	Cloudy	Total
Clear	7310 (87%)	1064 (13%)	8374 (100%)
Cloudy	194 (13%)	1288 (87%)	1482 (100%)

Total score = 87%

SEVIRI Cloud mask	IASI-DIFA cloud mask		
	Clear	Cloudy	Total
Clear	3576 (99%)	34 (1%)	3610 (100%)
Cloudy	168 (8%)	1893 (92%)	2061 (100%)

Total score = 96.6%

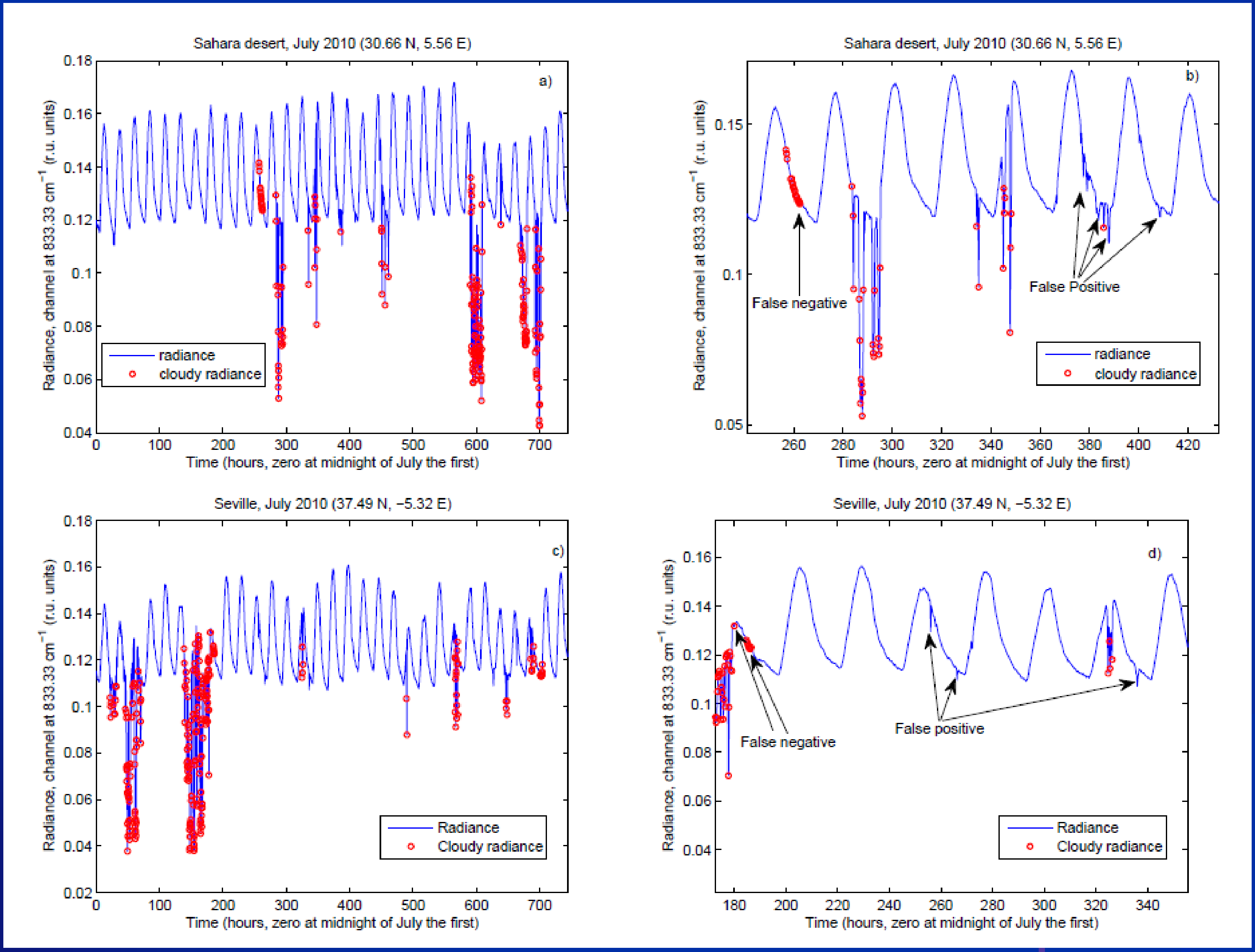
SEVIRI Cloud mask	IASI-DIFA cloud mask		
	Clear	Cloudy	Total
Clear	5188 (95%)	289 (5%)	5477 (100%)
Cloudy	22 (0.7%)	3127 (99.3%)	3149 (100%)

Total score = 96.4%

Concluding remarks

- We have invented a new tool CDF and developed the methodology to combine more statistics in a single cloud mask.
- Static Application Data has been generated for land and sea surface
- Once compared with suitably filtered AVHRR cloud masks we have a total agreement, which is better than 85% for any kind of surface

Future works for the Geostationary Platform



- Time Continuity