



# Proba-V Clouds Detection Round Robin Protocol Discriminant Analysis CNR-UNIBAS

Umberto Amato\*, Guido Masiello^, Carmine Serio^

\*Istituto per le Applicazioni del Calcolo CNR, Napoli (Italy)

^Scuola di Ingegneria, Università della Basilicata, Potenza (Italy)



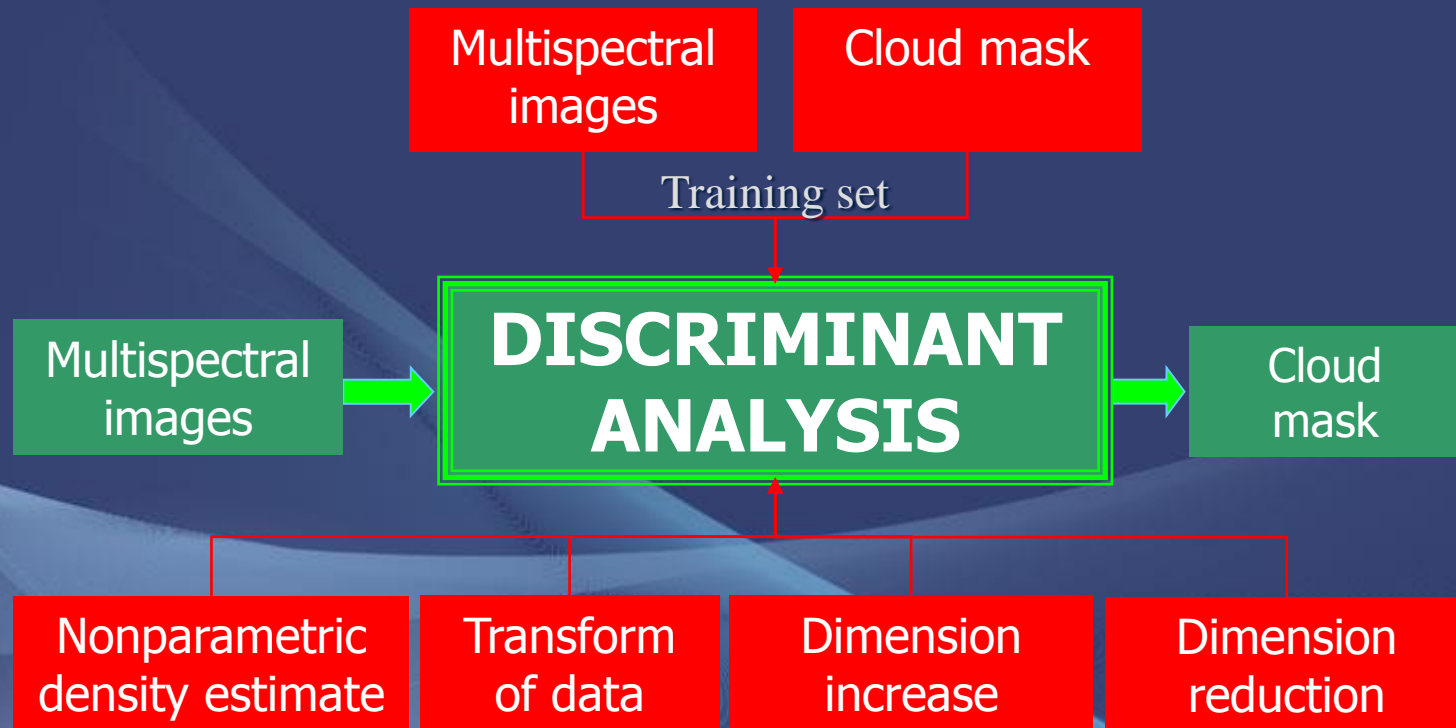
# Proba-V Clouds Detection Round Robin Protocol Discriminant Analysis CNR-UNIBAS

- Methodology
- Results (training data set)
- Results (partial test data set)
- Lessons learnt (to be continued...)

# Proba-V Clouds Detection Round Robin Protocol Discriminant Analysis CNR-UNIBAS

- Methodology
- Results (training data set)
- Results (partial test data set)
- Lessons learnt (to be continued...)

# DISCRIMINANT ANALYSIS



# METHODOLOGY

*Training data set* - Cloud mask SEVIRI - MODIS - SEVIRI&MODIS

*Surface mask* – GlobCover - GLCC - Proba-V

*Classification methodology* – Linear Discriminant Analysis - Quadratic Discriminant Analysis - Principal Component Discriminant Analysis - Cumulative Discriminant Analysis - K-NN - Support Vector Machines - Regression

*Augmenting dimension* (nonlinearity)

*Reducing dimension*

# METHODOLOGY

*Training data set* - Cloud mask SEVIRI - MODIS - SEVIRI&MODIS

*Surface mask* – GlobCover - GLCC - Proba-V

*Classification methodology* – Linear Discriminant Analysis - Quadratic Discriminant Analysis - Principal Component Discriminant Analysis - Cumulative Discriminant Analysis - K-NN - Support Vector Machines - Regression

*Augmenting dimension* (nonlinearity)

*Reducing dimension*

# TRAINING CLOUD MASK – SEVIRI

SEVIRI onboard Meteosat Second Generation (MSG)

Geostationary

15 min repetition time

3Km ssp. spatial resolution

60° around 0° latitude and longitude (one hemisphere)

Grid 3712x3712

11,953,264 useful pixels (86.75%)

Fusion with Proba-V: closest pixel in time and space regardless of the difference

SEVIRI cloud mask: clear, cloudy, uncertain

# TRAINING CLOUD MASK – SEVIRI

SEVIRI onboard Meteosat Second Generation (MSG)

Geostationary

15 min repetition time

3Km ssp. spatial resolution

60° around 0° latitude and longitude (one hemisphere)

Grid 3712x3712

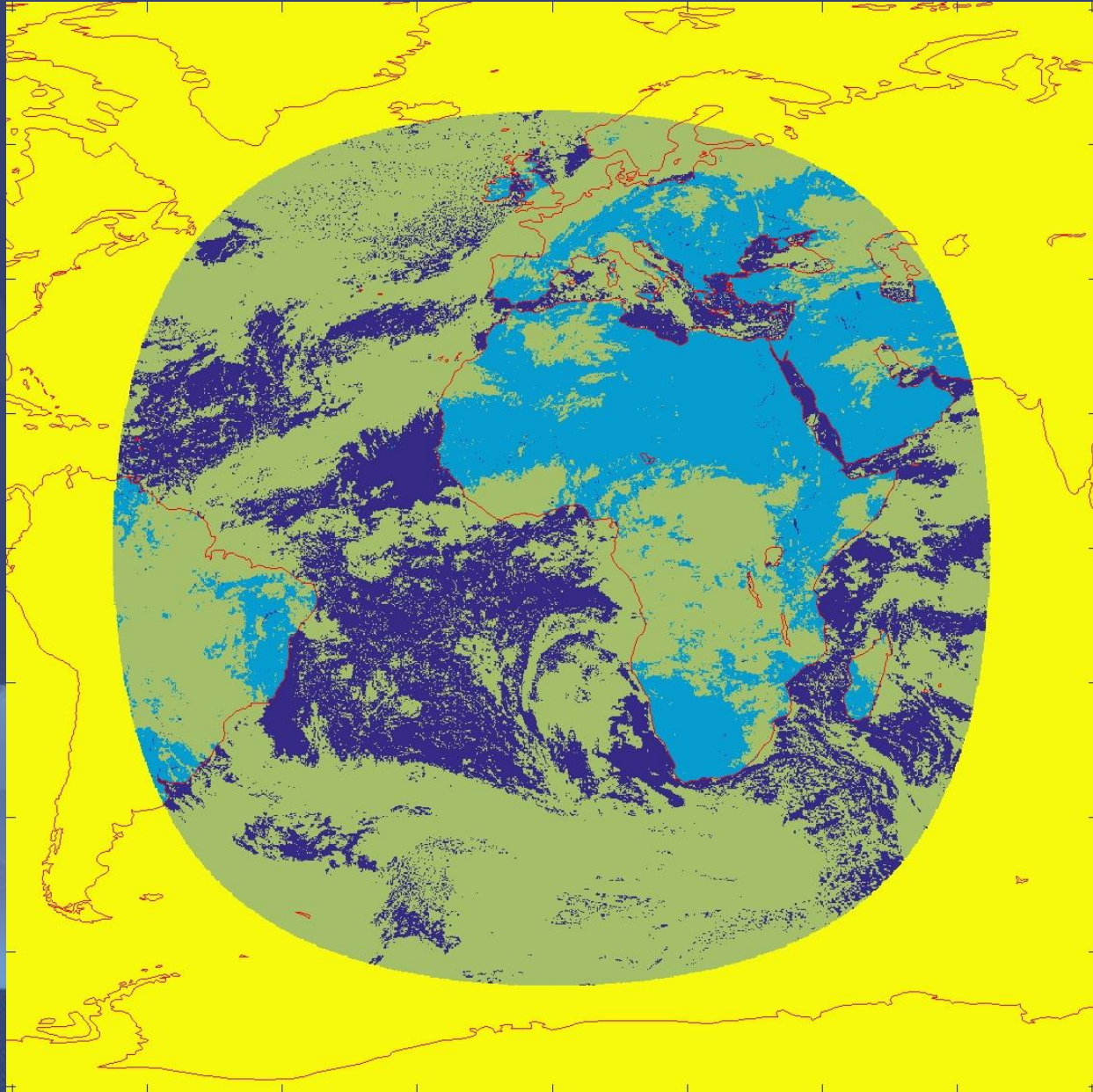
11,953,264 useful pixels (86.75%)

Fusion with Proba-V: closest pixel in time and space  
regardless of the difference

SEVIRI cloud mask: clear, cloudy, uncertain



# TRAINING CLOUD MASK – SEVIRI



# TRAINING CLOUD MASK – MODIS

MODIS sensor onboard EOS Terra and Aqua satellites

Polar

1Km spatial resolution (products MOD35, MYD35)

2276 files co-registered with the full Proba-V data set

Fusion with Proba-V: Each MODIS pixel mapped into one (closest) Proba-V pixel in space. Time difference less than  $\pm 30$  min

MODIS cloud mask: cloudy, uncertain clear, probably clear, confident clear

# TRAINING CLOUD MASK – MODIS

MODIS sensor onboard EOS Terra and Aqua satellites

Polar

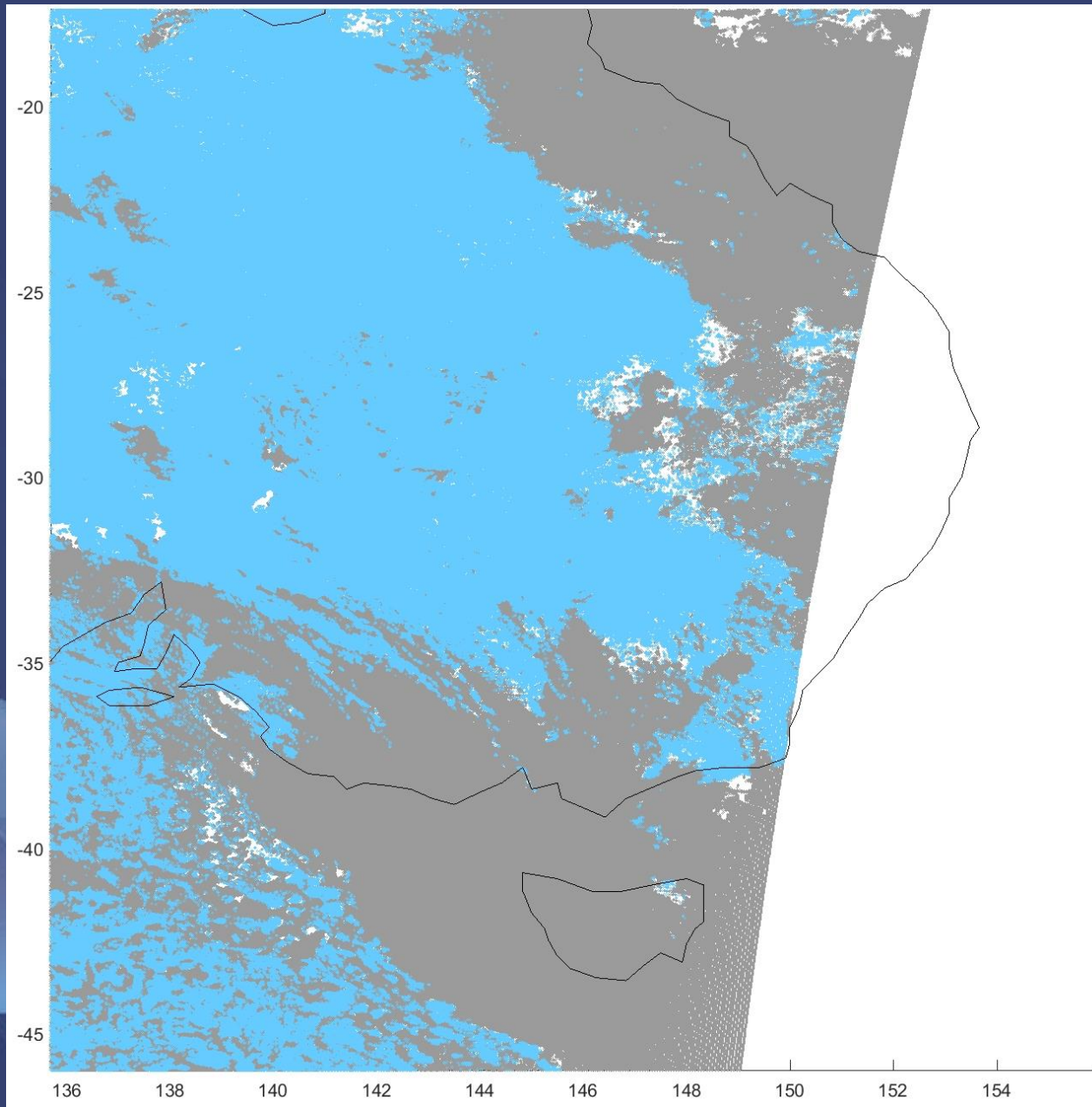
1Km spatial resolution (products MOD35, MYD35)

2276 files co-registered with the full Proba-V data set

Fusion with Proba-V: Each MODIS pixel mapped into one (closest) Proba-V pixel in space. Time difference less than  $\pm 30$  min

MODIS cloud mask: cloudy, uncertain clear, probably clear, confident clear

# TRAINING CLOUD MASK – MODIS



# TRAINING CLOUD MASK – SEVIRI&MODIS

Retains only pixels where cloudy or clear labels of SEVIRI and MODIS agree

# SURFACE MASK – GlobCover

2005 ESA initiative jointly with JRC, EEA, FAO, UNEP, GOFC-GOLD and IGBP

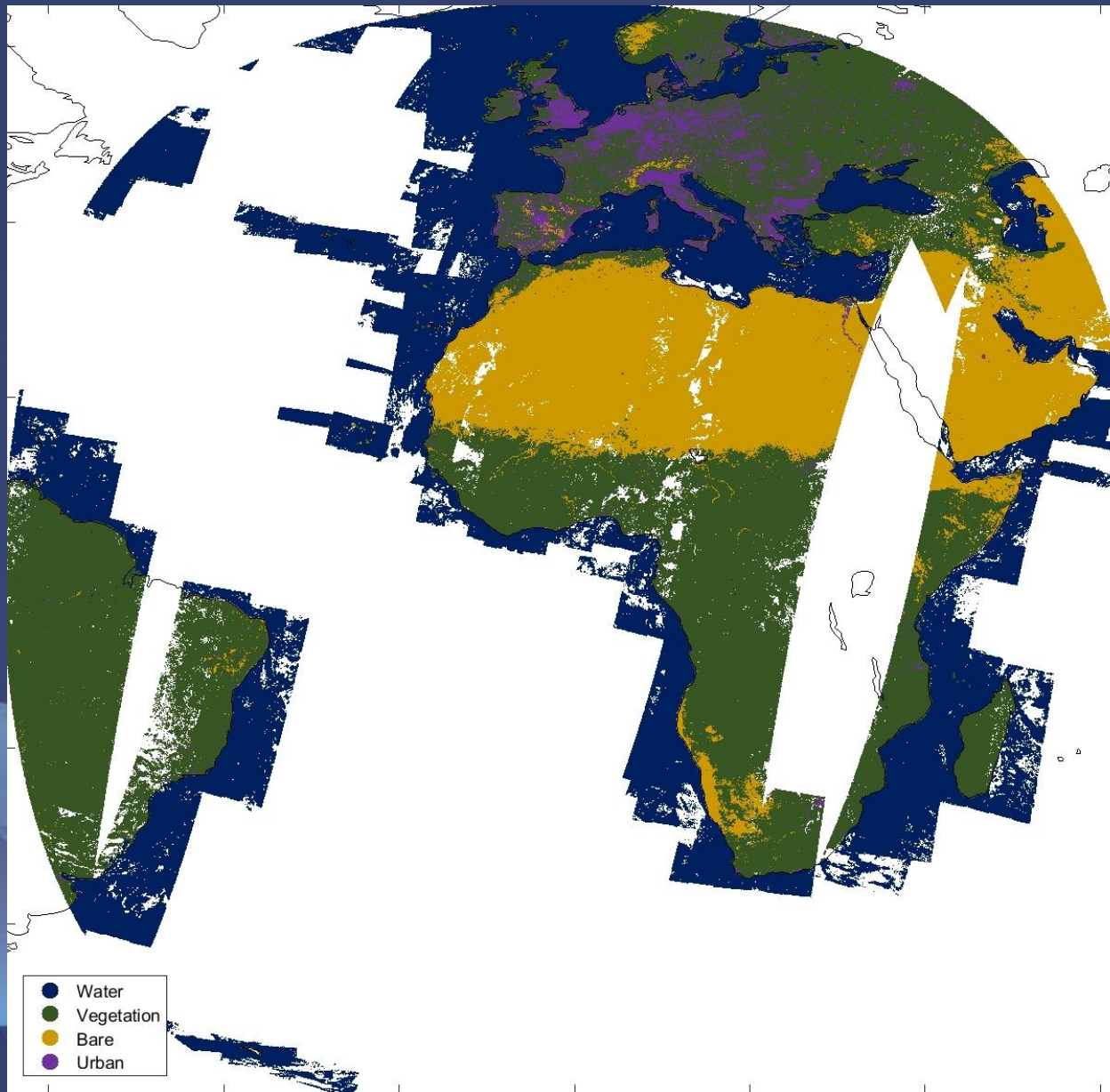
Data from MERIS sensor onboard ENVISAT  
Coverage January - December 2009  
300 m spatial resolution

Grid 55,800x129,600 pixels:  $[-65^{\circ}, 80^{\circ}]$  latitude;  $[-180^{\circ}, 180^{\circ}]$  longitude

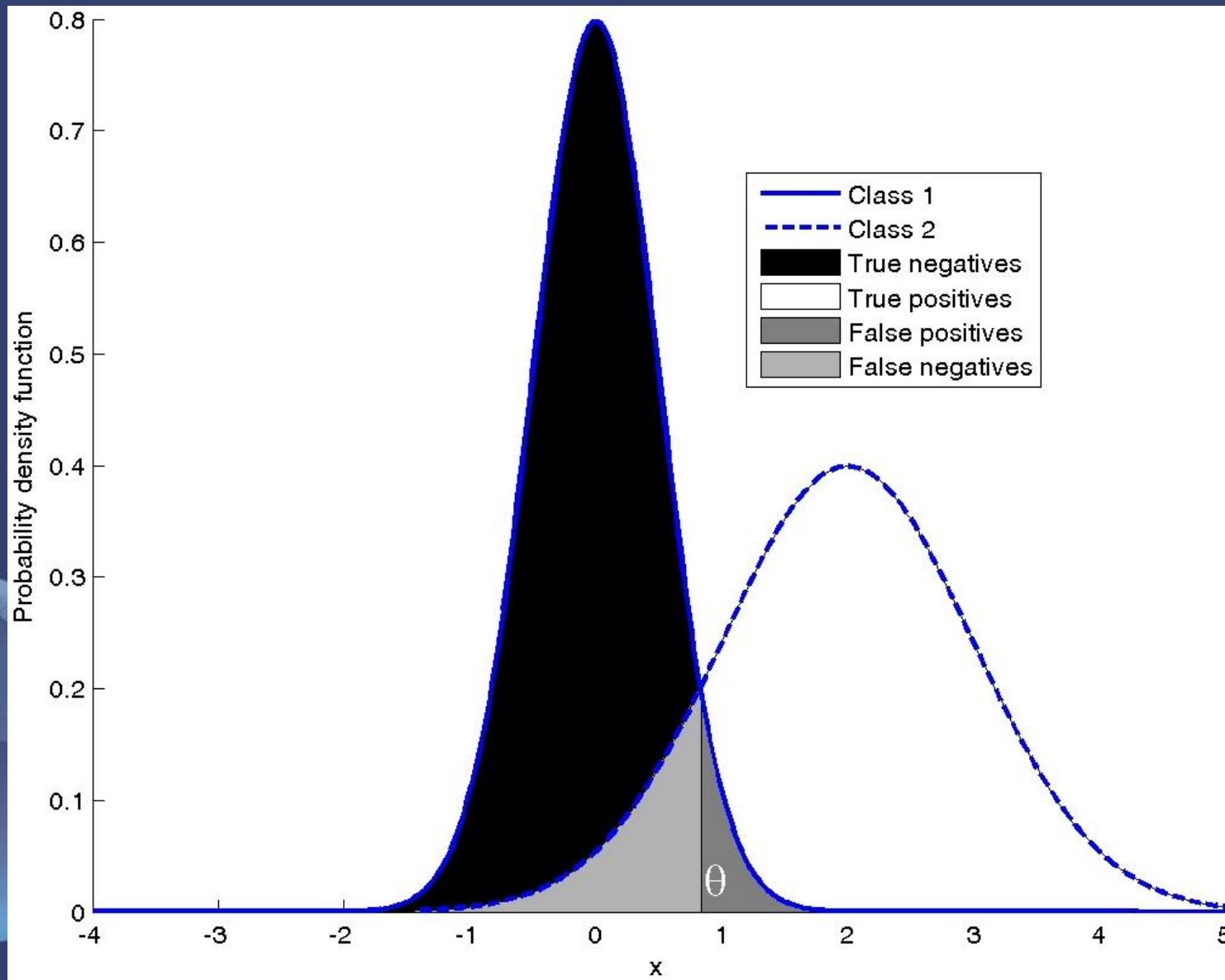
22 different surface classes for GlobCover

Reduced to 5 classes (Water, Vegetation, Bare land, Urban, Snow/Ice)

# SURFACE MASK – GlobCover

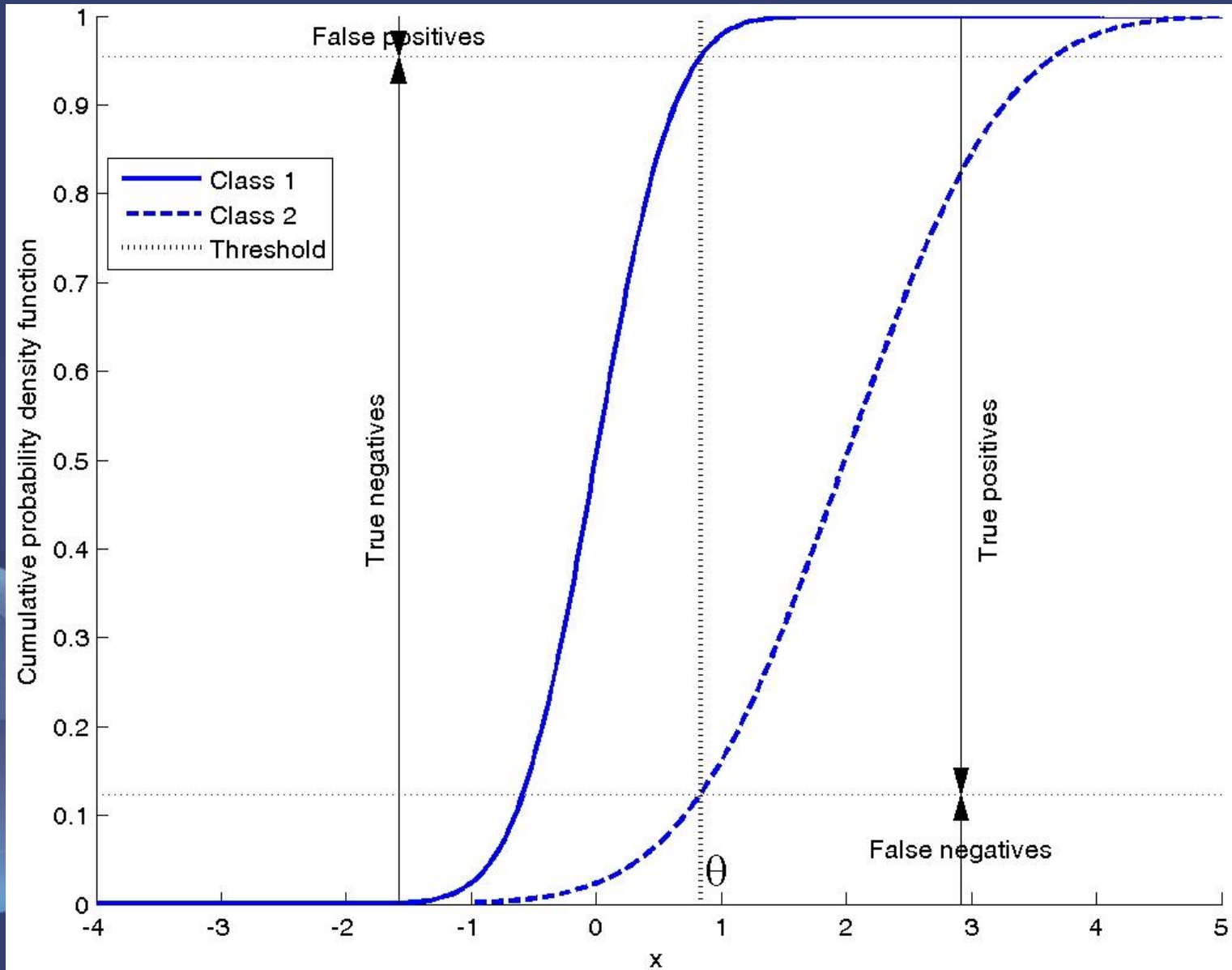


# CUMULATIVE DISCRIMINANT ANALYSIS CDA





# CUMULATIVE DISCRIMINANT ANALYSIS CDA



# CUMULATIVE DISCRIMINANT ANALYSIS CDA

One variable:

$$\hat{F}_{Clear}(x) \quad \hat{F}_{Cloudy}(x)$$

Empirical cumulative functions of reflectance in clear and cloudy conditions, respectively

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} C(\mathbf{X}; \theta)$$

$$C(\mathbf{X}; \theta) = \min \left( \max \left( 1 - \hat{F}_{Clear}(\theta), \hat{F}_{Cloudy}(\theta) \right), \max \left( \hat{F}_{Clear}(\theta), 1 - \hat{F}_{Cloudy}(\theta) \right) \right)$$

direction



$\mathbf{X}$  training reflectance and corresponding clear/cloudy labels

Multivariate: Hypothesis of independence

# Proba-V Clouds Detection Round

## Robin Protocol

### • Methodology Discriminant Analysis CNR-UNIBAS

- Results (training data set)
- Results (partial test data set)
- Lessons learnt (to be continued...)

# Proba-V Clouds Detection Round Robin Protocol Discriminant Analysis CNR-UNIBAS

- Methodology
- Results (training data set)
- Results (partial test data set)
- Lessons learnt (to be continued...)

# RESULTS – Training data set

	<b>SEVIRI Cloud mask</b>	<b>MODIS Cloud mask</b>
<b>Clear</b>	46,432,741 (50.6%)	120,640,955 (40.1%)
<b>Cloudy</b>	45,311,833 (49.4%)	187,331,229 (59.9%)
<b>Total</b>	91,744,574	300,972,184

# RESULTS – Training data set (silver standard)

CDA SEVIRI	Water			Vegetation			Bare land			Urban			Snow/Ice		
	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy
Training size	3.000.000	1.012.707	1.987.293	3.000.000	1.149.659	1.850.341	3.000.000	2.624.951	375.049	300.645	83.887	216.758	60.291	24.038	36.253
Test size	3.000.000	1.012.707	1.987.293	3.000.000	1.149.659	1.850.341	3.000.000	2.624.951	375.049	300.645	83.887	216.758	60.291	24.038	36.253
BLUE	80,0	79,7	80,2	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,4	81,7	79,0	78,2	79,6
RED	80,8	80,4	81,0	75,8	75,7	75,8	59,3	59,2	59,8	78,2	78,0	78,3	62,7	62,1	63,1
NIR	81,1	81,0	81,2	76,8	76,6	76,9	55,9	55,9	56,3	77,5	77,4	77,6	80,1	80,0	80,2
SWIR	80,9	80,7	81,0	62,1	61,9	62,2	62,4	62,5	62,1	67,2	67,1	67,3	62,2	61,5	62,7
BLUE-RED	80,7	80,7	80,8	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	79,0	78,8	79,2
BLUE-NIR	81,1	81,0	81,2	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	80,1	80,0	80,2
BLUE-SWIR	80,9	80,7	81,0	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	79,0	78,8	79,2
RED-NIR	81,1	81,0	81,2	76,8	76,8	76,8	59,5	59,5	59,6	78,2	78,1	78,2	80,1	80,0	80,2
RED-SWIR	80,9	80,7	81,0	75,8	75,7	75,8	62,3	62,3	62,3	78,2	78,1	78,2	62,7	62,1	63,1
NIR-SWIR	81,1	81,0	81,2	76,8	76,8	76,8	62,3	62,3	62,3	77,5	77,4	77,6	80,1	80,0	80,2
BLUE-RED-NIR	81,1	81,0	81,2	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	80,1	80,0	80,2
BLUE-RED-SWIR	80,9	80,7	81,0	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	79,0	78,8	79,2
BLUE-NIR-SWIR	81,1	81,0	81,2	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	80,1	80,0	80,2
RED-NIR-SWIR	81,1	81,0	81,2	76,8	76,8	76,8	62,3	62,3	62,3	78,2	78,1	78,2	80,1	80,0	80,2
BLUE-RED-NIR-SWIR	81,1	81,0	81,2	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	80,1	80,0	80,2
P.C. 1	80,9	80,9	80,9	76,9	76,9	76,9	58,7	58,7	58,7	78,9	78,9	78,9	77,6	77,6	77,6
P.C. 1-2	80,9	80,9	80,9	76,9	76,9	76,9	70,7	71,4	65,8	79,7	81,9	78,8	77,6	77,6	77,6
P.C. 1-2-3	80,9	80,9	80,9	76,9	76,9	76,9	70,7	71,2	67,1	78,9	78,9	78,9	77,6	77,6	77,6
P.C. 1-2-3-4	80,9	80,9	80,9	76,9	76,9	76,9	74,1	75,4	64,4	78,9	78,9	78,9	77,6	77,6	77,6
MAX-MIN	81,0	NIR		81,3	BLUE		67,1	P.C. 1-2-3		81,6	BLUE-RED		80,0	NIR	

# RESULTS – Training data set (silver standard)

	Water			Vegetation			Bare land			Urban			Snow/Ice		
	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy
<b>Alg 2 SEVIRI</b>	81,1	81,0	81,2	81,4	81,3	81,4	64,3	64,2	64,9	81,6	81,6	81,6	80,1	80,0	80,2
<b>Alg 3 MODIS</b>	80,1	80,1	80,1	78,3	78,3	78,4	83,8	83,8	83,9	82,2	82,2	82,2	79,5	79,6	79,4
<b>Alg 1 SEVIRI&amp;MODIS</b>	89,0	89,0	89,0	88,8	88,8	88,8	82,7	82,6	82,8	87,7	87,5	87,7	86,5	86,4	86,5

BLUE-RED-NIR-SWIR spectral bands

Clear and Cloudy success percentages coincide (by construction)

SEVIRI&MODIS cloud mask best performance (NOT conclusive!)

# Proba-V Clouds Detection Round

## Robin Protocol

### • Methodology Discriminant Analysis CNR-UNIBAS

- Results (training data set)
- Results (partial test data set)
- Lessons learnt (to be continued...)



# Proba-V Clouds Detection Round Robin Protocol Discriminant Analysis CNR-UNIBAS

- Methodology
- Results (training data set)
- Results (partial test data set)
- Lessons learnt (to be continued...)

# RESULTS – Test dataset (gold standard)

Surface typology	Alg 2 - SEVIRI			Alg 3 - MODIS			Alg 1 - SEVIRI&MODIS		
	Global	Clear	Cloudy	Global	Clear	Cloudy	Global	Clear	Cloudy
Water (326 scenes)	71.8	61.0	88.9	81.6	82.5	80.2	78.2	73.0	86.5
Vegetation (921 scenes)	83.3	62.8	91.4	78.3	72.4	80.6	82.7	63.6	90.3
Bare land (83 scenes)	73.5	78.6	68.3	73.5	81.0	65.9	73.5	81.0	65.9
Urban (9 scenes)	100	100	-	100	100	-	100	100	-
Ice/snow (10 scenes)	100	-	100	90	-	90	100	-	100
<b>Global (1350 scenes)</b>	<b>80.1</b>	<b>64.1</b>	<b>89.9</b>	<b>79.0</b>	<b>77.5</b>	<b>79.8</b>	<b>81.3</b>	<b>69.3</b>	<b>88.5</b>

BLUE-RED-NIR-SWIR spectral bands

Clear and Cloudy success percentages do not coincide

MODIS cloud mask best balanced performance

# Proba-V Clouds Detection Round

## Robin Protocol

### • Methodology Discriminant Analysis CNR-UNIBAS

- Results (training data set)
- Results (partial test data set)
- Lessons learnt (to be continued...)

# Proba-V Clouds Detection Round Robin Protocol Discriminant Analysis CNR-UNIBAS

- Methodology
- Results (training data set)
- Results (partial test data set)
- Lessons learnt (to be continued...)

# LESSONS LEARNT – Cloud mask

- pixels with MODIS cloud mask: 281,842,754
- pixels also having SEVIRI cloud mask: 73,617,537
- **pixels with coinciding cloud mask: 65,238,704 (88.6%)**
  
- pixels with coinciding clear cloud mask: 257,341,98 (39,4%)
- pixels with coinciding cloudy cloud mask: 39,504,506 (60.6%)
- pixels having MODIS and SEVIRI cloud mask that are clear according to SEVIRI cloud mask: 29,435,399 (40.0%)
- pixels having MODIS and SEVIRI cloud mask that are cloudy according to SEVIRI cloud mask: 44,182,138 (60.0%)
- pixels having MODIS and SEVIRI cloud mask that are clear according to MODIS cloud mask: 30,411,830 (41.3%)
- pixels having MODIS and SEVIRI cloud mask that are cloudy according to MODIS cloud mask: 43,205,707 (58.7%)

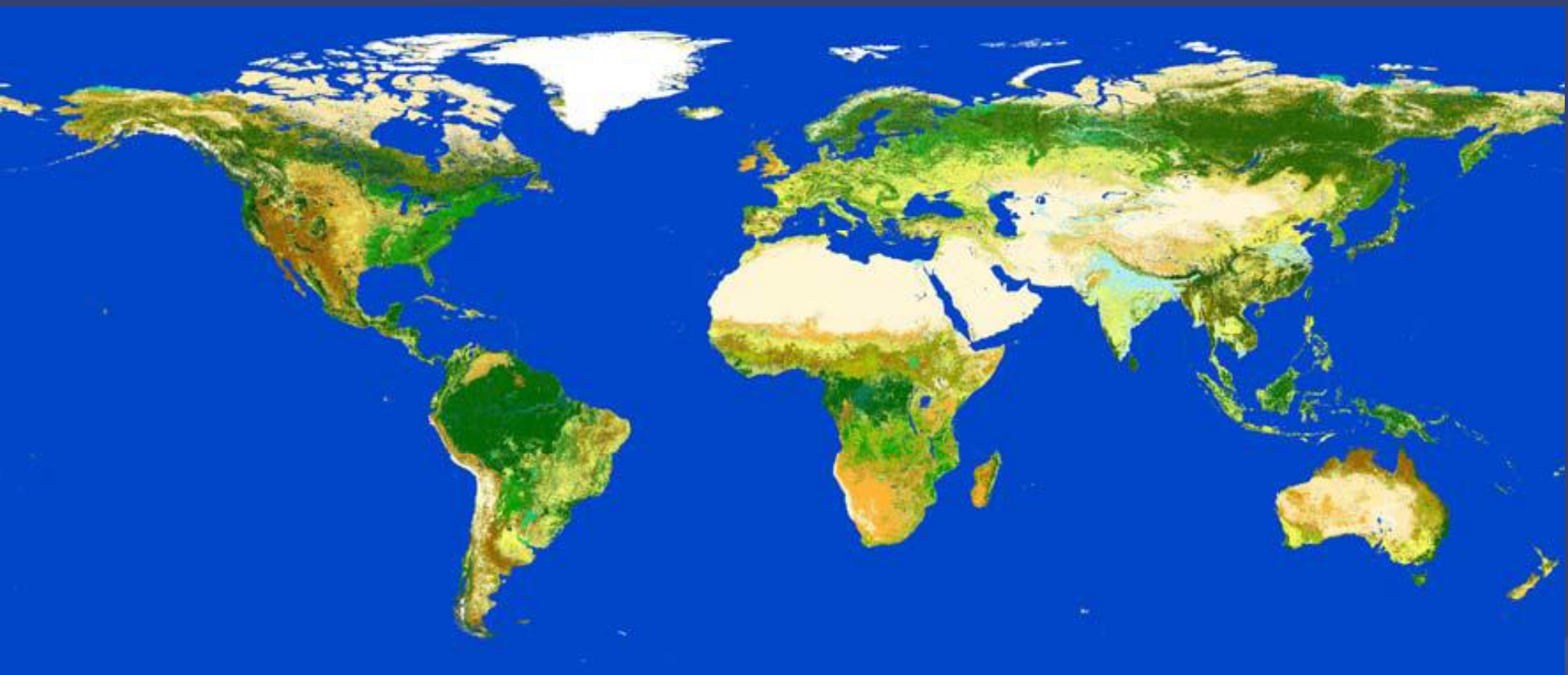
# LESSONS LEARNT – Cloud mask

- Proba-V test pixels with MODIS cloud mask: 43 out of 1,350
- Proba-V test pixels with SEVIRI cloud mask: 6 out of 1,350
- **pixels with coinciding Proba-V and MODIS cloud mask: 39 out of 43 (90.7%)**
- **pixels with coinciding Proba-V and SEVIRI cloud mask: 5 out of 6 (83.3%)**

# LESSONS LEARNT – Surface mask

- pixels having GlobCover and Proba cloud mask: 296,917,269
- Water pixels according to GlobCover: 114,756,349 (38.6%)
- Land pixels according to GlobCover: 182,160,920 (61.4%)
- Water pixels according to Proba: 103,764,133 (35.0%)
- Land pixels according to Proba: 193,153,139 (65.0%)
  
- **pixels agreeing for Water and Land: 279,345,259 (94.1%)**

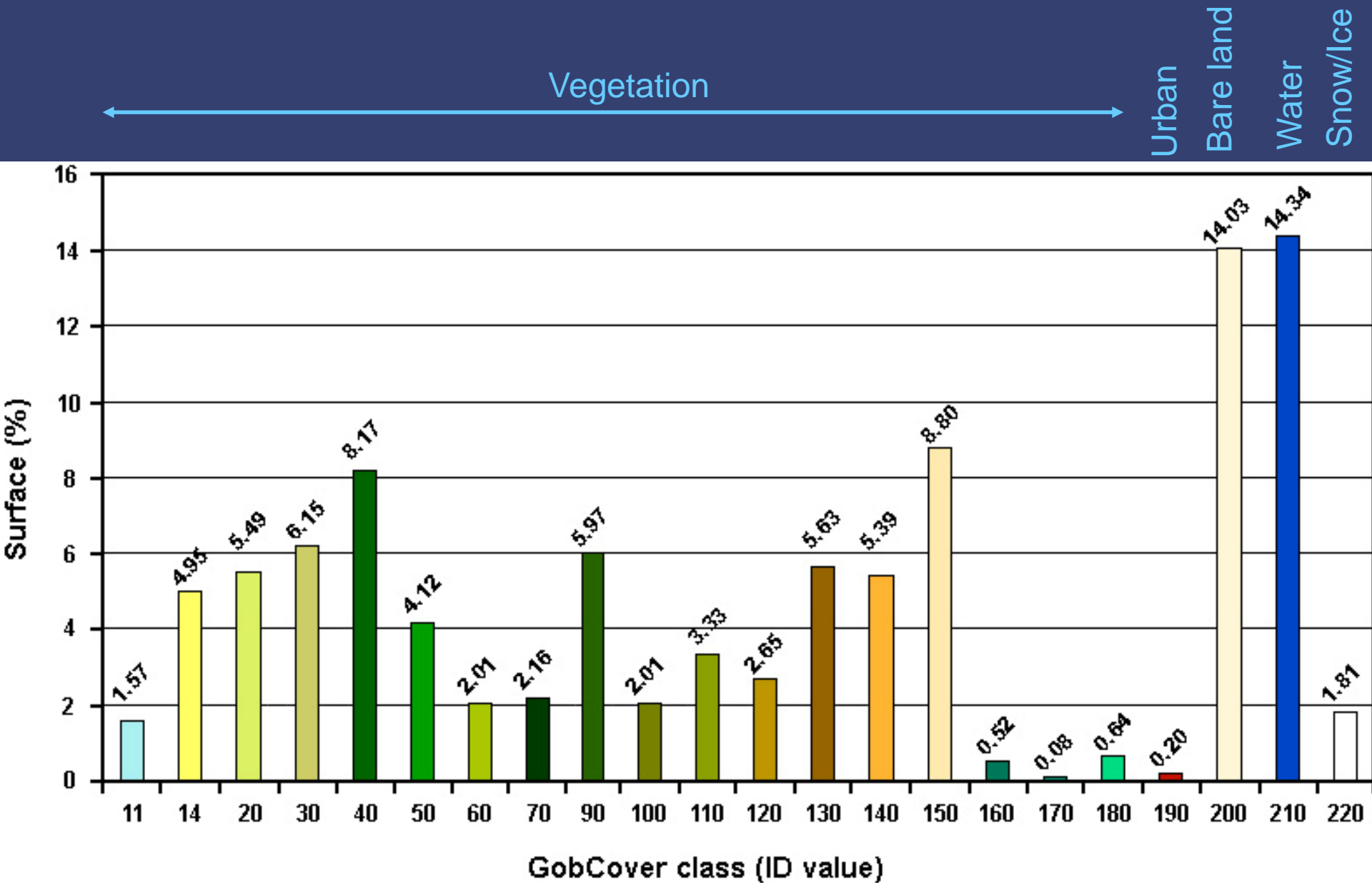
# LESSONS LEARNT – Surface mask



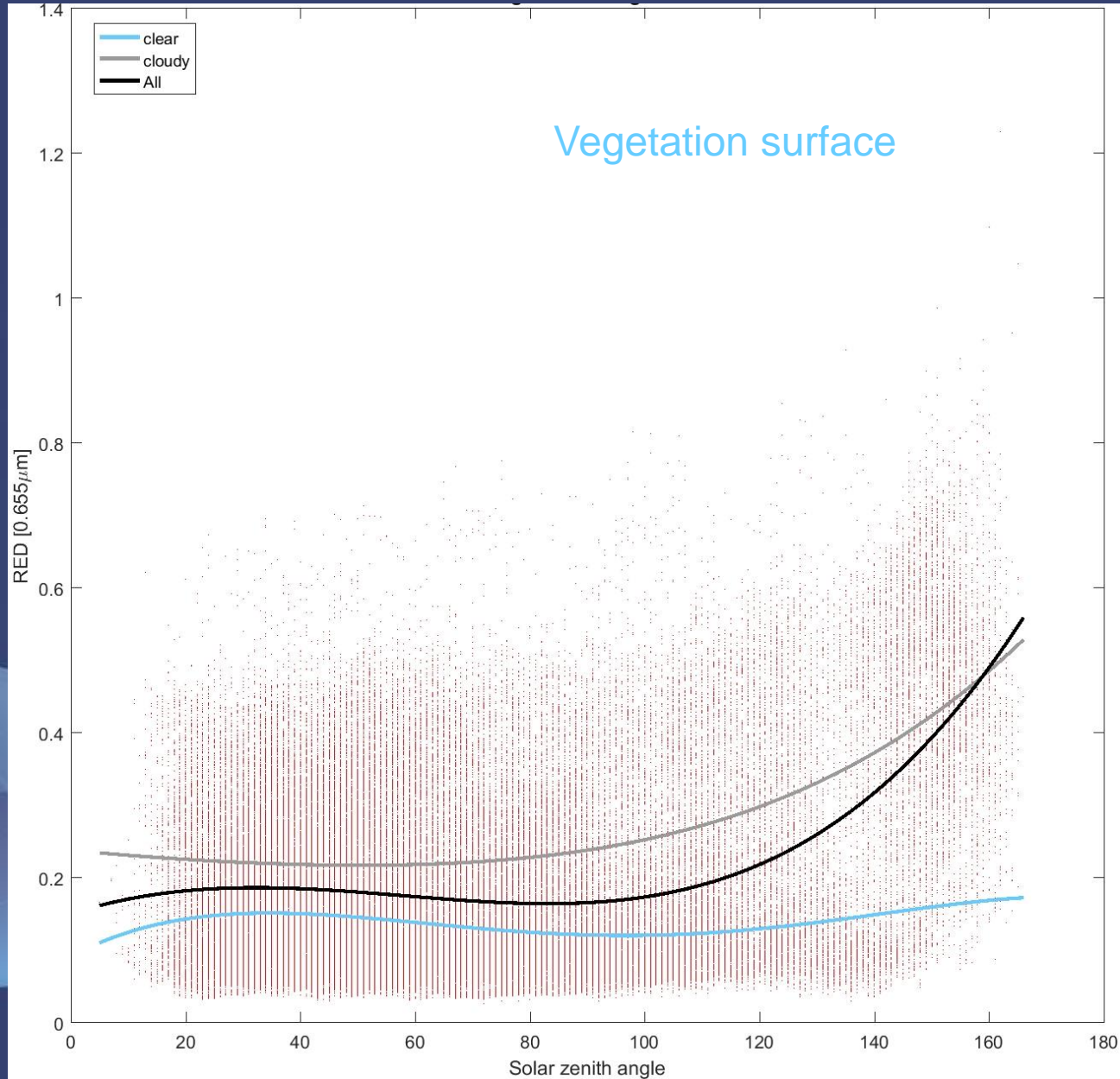
22 GlobCover land classes



# LESSONS LEARNT – Surface mask



# LESSONS LEARNT – Solar zenith angle



# LESSONS LEARNT – Classification method

- There is still room for improved classification (see the agreement rate between MODIS and SEVIRI silver standard, 88.6%, even though obtained with infrared channels)
- Different methodologies (e.g., regression)
- Cost function; Control of the Type I and Type II errors
- Nonlinearities (native nonlinear methodologies; augmentation of variables)
- Spatial features (e.g., correlation)

# LESSONS LEARNT – Proba-V Gold standard

Proba-V is a unique data set accurately annotated that meets the «Gold» label for supervised classification

Pls make data available, with all due credits

**Thank you!**