Antsirabe (Madagascar)

JECAM/GEOGLAM Science Meeting, Kiev, 11-12 October, 2016

Lebourgeois V., Dupuy S., Gaetano R., Ameline M., Diaz S., Butler S., Bégué A.

CIRAD - TETIS Research Unit

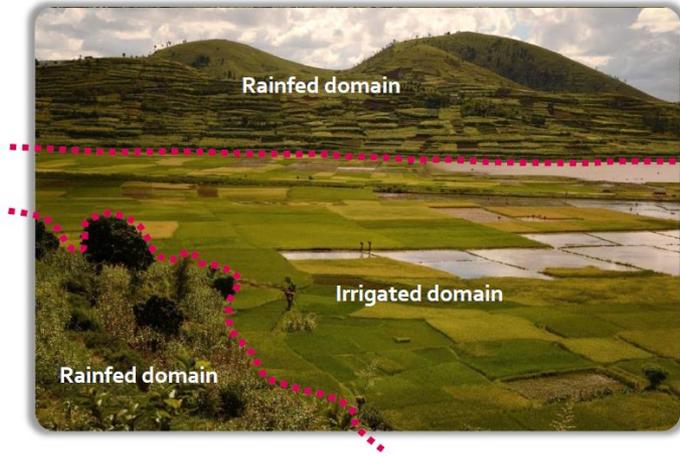# 1. Site Description & Project Objectives

Fig. 1 *Illustration of irrigated and rainfed agricultural landscapes*

STUDY SITE

- **Location:** 60*60 km (3600 km²) in the Central Highlands, Vakinankaratra Region (Lat.: 19.43° S / Long.: 47.04° E)
- **Topography:** hills / plains / shallows (1400-2000m)
- **Soils:** Clayey
- **Drainage class / irrigation:** Moderate / irrigation canals
- **Crop calendar:** October - April
- **Field size:** 0.03 ha
- **Climate and weather:** Subtropical humid (cloudy !)
- **Agricultural methods used:** Manual Tillage / Hoeing / Fertilization with manure more or less mixed with ashes (few NPK inputs due to availability and cost) / Irrigation on terraces or basins, rainfed crops on the hills

OBJECTIVES: Crop identification and crop area estimation + Rice yield estimation

A complex site because of:

- Very (very!) small field size
- Cropping practices (associated crops)
- Natural vegetation around rainfed fields
- Cloudy conditions (rainy season…)

2. Data (2014-2015 growing season)

SATELLITE DATA

- **HSR time series:** 21 images (10 m SPOT 5 & 15 m pansharpened Landsat 8)

- **VHSR coverage:** PLEIADES (0.5m) – 9 tiles acquired around the peak of the growing season

- **Average frequency of the time series:** 14 days (Fig. 2)
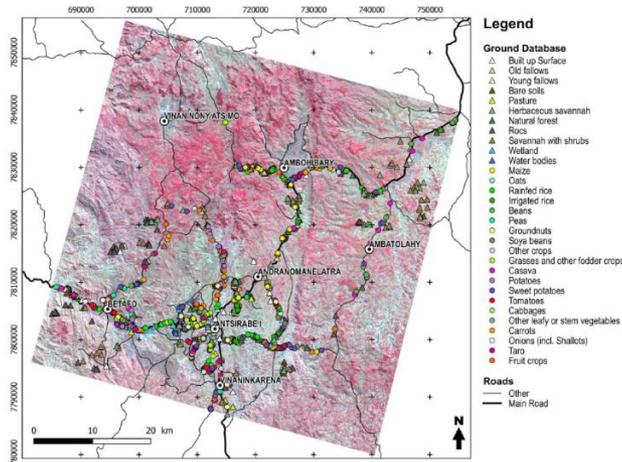
- **Auxilary data:** Digital Elevation Model (SPOT DEM)

Fig. 2 *Chronogram and quicklooks of the 2014-2015 satellite acquisitions. The cloud percentage represents the cloud proportion over the ground database for each acquisition date. For PLEIADES coverage, the 16 March 2015 (D15 for date 15 of the time series) was choosen to represent the median acquisition date of the different tiles.*

Fig. 3 *GPS waypoints on land use for 2014-2015 growing season (background: Landsat 8 satellite image acquired on 15th June 2015, map projection: WGS84 Zone 38 S)*

GROUND DATA

- **Land use characterization:** 1020 GPS waypoints + 199 additionally points added to the field database for non-crop class by photo interpretation of PLEAIDES VHSR imagery, making a total of 1219 points (860 cropped and 359 non-cropped) (Fig. 3). Plot boundaries were digitized on VHSR imagery to obtain a polygon (object) database. Attributes on land use were recorded according to the JECAM nomenclature (Fig. 4)

- **Rice yield estimation:** 124 GPS waypoints associated with attributes (seeding and transplant dates, variety, fertilization, toposequence, straw biomass, full grain weight, empty grain weight, photo) and plot boundaries digitized on PLEIADES images
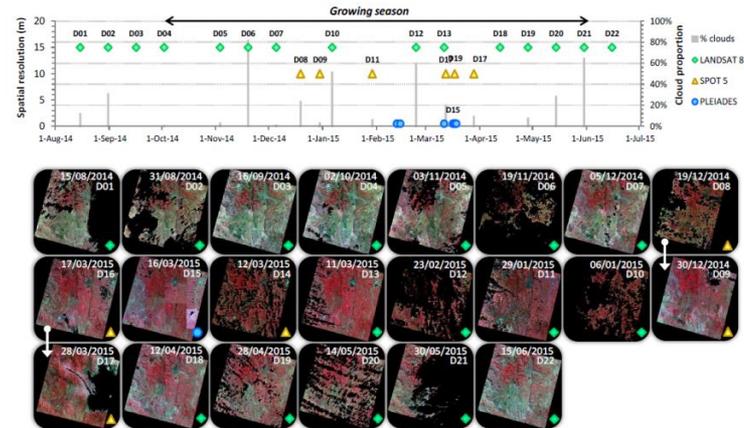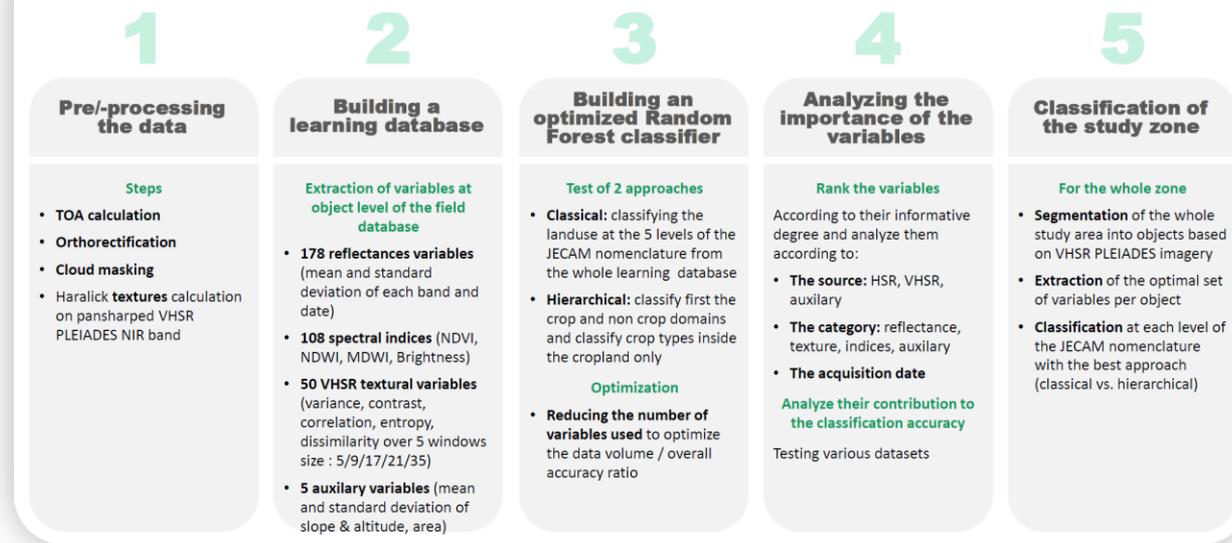
Level 1 Cropland	Level 2 Land Cover	Level 3 Crop Group	Level 4 Crop Class	Level 5 Sub Class
Crop	Annual cropland	Cereals	Maize	Maize
			Oats	Oats
			Rice	Rainfed rice
				Irrigated rice
		Leguminous	Beans	Beans
			Peas	Peas
		Oilseed crops	Groundnuts	Groundnuts
			Soya beans	Soya beans
		Other crops	Other crops	Other crops
			Grasses and other fodder crops	Grasses and other fodder crops
		Root or tuber crops with high starch or inulin content	Casava	Casava
			Potatoes	Potatoes
			Sweet potatoes	Sweet potatoes
		Vegetables and melons	Fruit-bearing vegetables	Tomatoes
			Leafy or stem vegetables	Cabbages
				Other leafy or stem vegetables
			Root bulb or tuberous vegetables	Carrots
				Onions (incl. Shallots)
				Taro
	Lignous crop	Fruit crops	Fruit crops	Fruit crops
Non Crop	Built-up surface	Built-up Surface	Built-up Surface	Built-up Surface
	Fallows	Old fallows	Old fallows	Old fallows
		Young fallows	Young fallows	Young fallows
	Natural spaces	Bare soils	Bare soils	Bare soils
		Grassland	Pasture	Pasture
			Herbaceous savannah	Herbaceous savannah
		Forest	Forest	Forest
		Rocs	Rocs	Rocs
		Shrub land	Savannah with shrubs	Savannah with shrubs
	Water bodies	Water bodies	Water bodies	Water bodies
	Wetland	Wetland	Wetland	Wetland

Fig. 4 *JECAM nomenclature for Antsirabe site*

3. Methods

1 Pre/-processing the data

Steps
- TOA calculation
- Orthorectification
- Cloud masking
- Haralick **textures** calculation on pansharped VHSR PLEIADES NIR band

2 Building a learning database

Extraction of variables at object level of the field database
- **178 reflectances variables** (mean and standard deviation of each band and date)
- **108 spectral indices** (NDVI, NDWI, MDWI, Brightness)
- **50 VHSR textural variables** (variance, contrast, correlation, entropy, dissimilarity over 5 windows size : 5/9/17/21/35)
- **5 auxilary variables** (mean and standard deviation of slope & altitude, area)

3 Building an optimized Random Forest classifier

Test of 2 approaches
- **Classical:** classifying the landuse at the 5 levels of the JECAM nomenclature from the whole learning database
- **Hierarchical:** classify first the crop and non crop domains and classify crop types inside the cropland only

Optimization
- **Reducing the number of variables used** to optimize the data volume / overall accuracy ratio

4 Analyzing the importance of the variables

Rank the variables
According to their informative degree and analyze them according to:
- **The source:** HSR, VHSR, auxilary
- **The category:** reflectance, texture, indices, auxilary
- **The acquisition date**

Analyze their contribution to the classification accuracy

Testing various datasets

5 Classification of the study zone

For the whole zone
- **Segmentation** of the whole study area into objects based on VHSR PLEIADES imagery
- **Extraction** of the optimal set of variables per object
- **Classification** at each level of the JECAM nomenclature with the best approach (classical vs. hierarchical)

Main steps:

- Build a **learning database** (more than 300 variables per polygon of the ground database)

- Build a **Random Forest classifier** (Classical vs. Hierarchical) and **optimize** it (reducing the number of variables to obtain an optimal overall accuracy + reduce time computation)

- **Analyze the contribution of each source or category of variables** on overall accuracy + strategic acquisition dates

4. Results (focus on cropland and crop type classification)

4.1 CLASSICAL VS. HIERARCHICAL APPROACH

- At level 1 (Cropland – 2 classes) "Cropland" and "Non-Cropland" were classified with an overall accuracy of 91.7% (Kappa 0.82). A f-score of 0.93 was observed for "Cropland" class showing a very good classification accuracy for this specific class.
- Results showed an improvement of f-score for more than 70% of the classes using the hierarchical approach at the more detailed level of the JECAM nomenclature (level 5 – Sub Class having 25 classes) (Fig. 5).
- With hierarchical approach, "Non Crop" classes were well classified whatever the level of the nomenclature (Overall accuracy > 80%, Kappa > 0.75) and the classification accuracy of "Crop" classes decreased from level 3 (Crop Group) to level 5 (Sub Class) (Fig. 6).
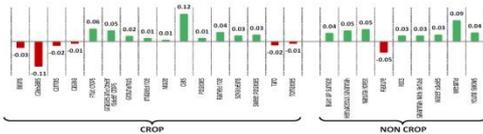
Fig. 5 Difference in class f-score between hierarchic and classic approaches at Sub Class level.

	Cropland (level 1)	Land Cover (level 2)		Crop Group (level 3)		Crop Class (level 4)		Sub Class (level 5)	
		Crop	Non Crop	Crop	Non Crop	Crop	Non Crop	Crop	Non Crop
Overall Accuracy	91.7%	96.6%	90.7%	70.2%	83.2%	64.1%	81.3%	64.4%	80.2%
Kappa	0.82	0.69	0.75	0.61	0.79	0.59	0.78	0.61	0.76

Fig. 6 Overall accuracy and Cohen's Kappa obtained with the hierarchical approach for each level of the JECAM nomenclature

4.2 OPTIMIZATION

- Many variables of our dataset are correlated or redundant. Forcing the Random Forest algorithm to use the more important variables (ranked thanks to the mean decrease accuracy measure) with an optimal reduced set of chosen variables improved slightly the overall accuracy of the classifications at each level of the nomenclature (Fig. 7).
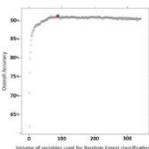 - In general, results showed that the more the level of nomenclature was complex (with a lot of classes), the more the optimal volume of variables was high (Fig. 8).

Fig. 8 Optimal volume of variables obtained for each level of the nomenclature.

Fig. 7 Evolution of cross-validated (5 folds) overall accuracy according to the set of optimal variables used for classification at Cropland level
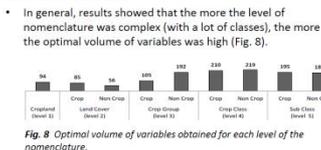
4.3 ANALYSIS OF THE VARIABLES OF IMPORTANCE

- **Category (Fig. 9 left):** Indices and reflectances are always present in a significant volume (except for crop classes at level 2 – Land Cover). Textures were predominant to discriminate crop classes ("Lignous Crop" from "Annual Crop") at level 2 – Land Cover but are not useful to classify cropland (level 1), non crop classes at level 2 (Land Cover) and crop classes at level 3 (Crop Group). Despite their small volume (only 5 variables over 341), auxiliary variables (mainly slope & altitude) were always present inside the 30 more important variables, whatever the level of classification.
- **Source (Fig. 9 right):** HRS variables are predominant whatever the level of classification, except for discriminating crop classes ("Lignous Crop" from "Annual Crop") at level 2 (Land Cover).
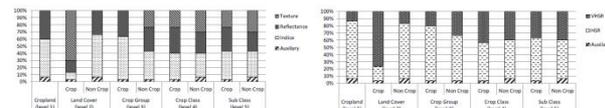
Fig. 9 Proportion of the different categories (left) and source (right) of variables inside the 30 more important variables (ranked according to the mean decrease accuracy measure) for each level of the JECAM nomenclature.

- **Acquisition date (Fig. 10):** Acquisition date 15 (D15 – 16/03/15), which corresponds to PLEAIDES coverage, was the most used because of the initial volume of variables (63, which includes 50 textures) calculated from this date. SPOT 5 images from D09 (30/12/14) and D17 (17/03/15) was also important to discriminate crop classes (in green color gradient) whatever the level of classification.
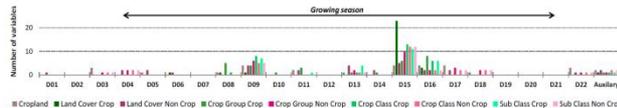
Fig. 10 Volume of variables per date inside the 30 more important variables (ranked according to the mean decrease accuracy measure) for each level of the JECAM nomenclature.

4.4 CONTRIBUTION OF THE VARIABLES TO THE CLASSIFICATION ACCURACY

- Results (Fig. 11) showed that VHSR used alone or along with auxiliary data (datasets 6 and 7) was not sufficient to obtain interesting accuracy results. The best accuracies were obtained using the whole data set (All). Eliminating the VHSR textures from the whole data set (Dataset 1) did not significantly impact the results.

Fig. 11 Classification accuracies (Cohen's Kappa) obtained using the 8 experimental datasets at each level of the nomenclature

4.5 CLASSIFICATION OF THE STUDY AREA

- The whole study area was classified using the optimized classifiers obtained with the hierarchic approach at each level of the nomenclature (Fig. 12)
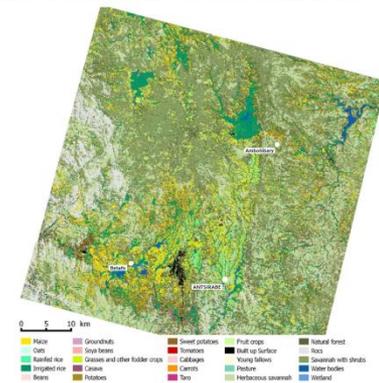- For each object, an information about the class membership probability is given

Fig. 12 Classification of the study area at the more detailed level (level 5 - Sub Class) of the JECAM nomenclature

Main results:

- **Improvement** of f-score for more than 70% of the crop classes **using the hierarchical approach**
- **Predominance of VHSR variables** (textures!) when ranking the variables by importance
- BUT…. Eliminating the VHSR data from the whole data set did not significantly impact the accuracy results !

5. Ongoing work: Integration of Sentinel-2 data

DATA acquired over 2015-2016 growing season

In 2015-2016, the study area was expanded to cover 2/3 of the Vakinankaratra Region (7470 km²)

- **1401 GPS Waypoints** (1125 Crop, 276 Non Crop)
- **20 HSR images:** 10 Landsat 8 (over 24 available acquisitions) and 10 Sentinel-2 (over 21 available images)
- **1 VHSR coverage:** 2 SPOT 6 images

METHOD

- The same method as the one established in 2015 was used, excepted that only the average values per objetc were extracted (as the results over the 2014-2015 growing season showed that standard deviations were the less discriminant variables).
- For Sentinel-2 data, 60 m resolution bands (B1- aerosols, B8 – water vapour, B10 – cirrus) were not considered

PRELIMINARY RESULTS

- Such as for the previous growing season, results showed an **improvement of f-score for more than 70% of the classes using the hierarchical approach** at the more detailed level of the JECAM nomenclature (level 5 – Sub Class having 25 classes)
- Accuracy results (Fig. 14) observed at each level of the nomenclature were less satisfying than for the previous growing season, particularly for Non Crop classes (10 - 5% decrease for levels 3 to 5). For Crop classes, this accuracy loss was less important (1-3 %)
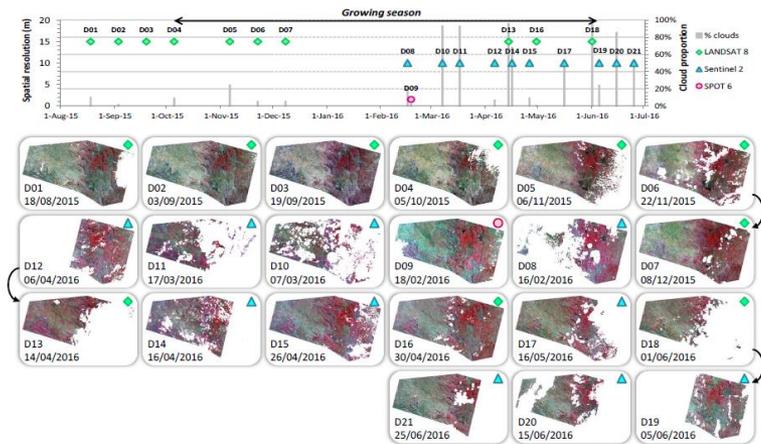

Fig. 13 *Chronogram and quicklooks of the 2015-2016 satellite acquistions. The cloud percentage represents the cloud proportion over the ground database for each acquisition date..*

	Cropland (level 1)	Land Cover (level 2)		Crop Group (level 3)		Crop Class (level 4)		Sub Class (level 5)	
		Crop	Non Crop	Crop	Non Crop	Crop	Non Crop	Crop	Non Crop
Overall Accuracy	90.9%	95.7%	80.5%	68.9%	72.8%	62.4%	64.2%	61.3%	65.9%
Kappa	0.68	0.67	0.73	0.57	0.41	0.53	0.62	0.58	0.64

Fig. 14 *Overall accuracy and Cohen's Kappa obtained in 2015-2016 with the hierarchical approach for each level of the JECAM nomenclature*

What about Sentinel-2 ?

- **Same approach** used for the 2015-2016 growing season **with integration of Sentinel-2 data** (+Landsat 8)

- Preliminary results showed **equivalent classification accuracies for crop type classification** (but with a not optimal S2 dataset because of clouds ! => waiting for Sentinel-2b !)

- During 2016-2017 growing season, the **Sen2Agri Toolbox** will be tested in real time for cropland and crop type mapping over a 90 000 km² area to serve as a use case.

Thank you! 🇲🇬 *Misaotra !*