CHARACTERIZING CROPPING SYSTEMS AND THEIR PRODUCTIVITY USING MULTI-SOURCE REMOTE SENSING DATA AND DATA MINING FOR FOOD SECURITY

Elodie VINTROU, Valentine LEBOURGERGOIS, A. Bégué, D. Ienko, M. Teisseire, F.R. Andriandrahona
WHY CHARACTERIZING THE CROPPING SYSTEMS?

- To assess the overall productivity of the system (intensive vs low-input agriculture...)
- To assess environmental risks (over-exploitation of ground water resources, water quality degradation, ...)

To understand the overall sustainability of agriculture on a territory
(facing increasing population, climate change ....)

In a food security context

To assist early warning systems in developing countries
for estimating the food production to compensate for the lack of food
by food aid or imports

AGRHYMET

 SENTINEL-2 for Science Workshop, 20-22 May 2014, Frascati, Italy
EARLY WARNING SYSTEMS FOR FOOD SECURITY: HOW DO THEY WORK?

Different components

- Availability, accessibility, prices stability...

For a given crop:

\[
\text{Yield} \times \text{Crop acreage} = \text{Production}
\]

- National statistics (quant.)
- Crop growth model (quant.)
- Anomalies (qual.)

- National statistics (quant.)
- High or very high spatial resolution images
PROJECT GLOBAL METHODOLOGY

1a. MAPPING THE CULTIVATED DOMAIN (CROPLAND MASK)
   b. MAPPING THE CROPPING SYSTEMS (inside the cropland mask)

2. EXTRACTING PHENOLOGICAL DATES AND ESTIMATING CROP YIELD
STUDY ZONE

- Region of 60*60 km near Antsirabe highlands in Madagascar

SPECIFIC CONSTRAINTS FOR REMOTE SENSING

Cloud cover, fragmented landscapes, small size fields (0.03 ha), associated crops...
SENTINEL-2 MISSION simulation

Two main image sources:

- SPOT acquisition antenna in Reunion (SEAS-OI)
- SPOT4-Take5 project (CNES)
### MULTIPLE SOURCE DATA

<table>
<thead>
<tr>
<th>Data</th>
<th>Crop Mask</th>
<th>Cropping systems mapping</th>
<th>Perspectives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pleides</strong>&lt;br&gt;(mono-date / maximum of growing season)&lt;br&gt;(0.5 m)</td>
<td></td>
<td></td>
<td>Very high spatial resolution</td>
</tr>
<tr>
<td><strong>SPOT date 1</strong>&lt;br&gt;&lt;br&gt;&lt;br&gt;<strong>SPOT date 2</strong>&lt;br&gt;&lt;br&gt;&lt;br&gt;<strong>SPOT date 3</strong>&lt;br&gt;&lt;br&gt;&lt;br&gt;<strong>SPOT date n</strong>&lt;br&gt;(10 – 20 m)</td>
<td></td>
<td></td>
<td>High temporal resolution</td>
</tr>
<tr>
<td><strong>DEM</strong></td>
<td></td>
<td></td>
<td>Toposequence</td>
</tr>
<tr>
<td><strong>Ground databases</strong>&lt;br&gt;(Cropping systems, cropping practices, yield...)</td>
<td></td>
<td></td>
<td>Thematic information</td>
</tr>
</tbody>
</table>

DATA ACQUISITION COMPLETED FOR THE 2012-2013 AGRICULTURAL SEASON AND ACQUISITION ONGOING FOR THE 2013-2014 SEASON
SATELLITE DATA: SPOT 2012-2013 time series (25 images / 8 months)

Context and objectives
Data
Crop Mask
Cropping systems mapping
Perspectives
SATELLITE DATA: SPOT vs. PLEIADES ON OUR STUDY ZONE

Context and objectives

Data

Crop Mask

Cropping systems mapping

Perspectives

SENTINEL-2 for Science Workshop, 20-22 May 2014, Frascati, Italy
GROUND DATA (2013)

- 400 GPS points (2013)

- 1020 GPS points (2014) (thanks to new equipment!)

Legend
- Crop
- Non crop

Context and objectives Data Crop Mask Cropping systems mapping Perspectives

Background: SPOT 5 image, 3 March 2013
© CNES 2013, Distribution Astrium Services / Spot Image S.A., France, All rights reserved

SENTINEL-2 for Science Workshop, 20-22 May 2014, Frascati, Italy
APPROACHES

2 steps

- Crop / non crop mask
- Cropping system mapping

2 approaches

- object-based image analysis (RS method)
- multisource data mining techniques (DM method)
CROP / NON CROP MASK: Object-based analysis

Concept

- Segmentation in homogeneous objects (groups of pixels) and classification

Pixel approach → Object approach
CROP / NON CROP MASK : Object-based analysis

Data used

- 2 SPOT images (maximum of the growing season and dry season)
CROP / NON CROP MASK : Object-based analysis

Data used

- 2 SPOT images (maximum of the growing season and dry season)
- NDVI
## CROP / NON CROP MASK : Object-based analysis

### Data used

- 2 SPOT images (maximum of the growing season and dry season)
- NDVI
- DEM (from 1400 m to 2000 m) + Slope

### Data

<table>
<thead>
<tr>
<th>Context and objectives</th>
<th>Data</th>
<th>Crop Mask</th>
<th>Cropping systems mapping</th>
<th>Perspectives</th>
</tr>
</thead>
</table>

### Images

- **High altitude**
- **Low altitude**
- **High slope**
- **Low slope**

*SENTINEL-2 for Science Workshop, 20-22 May 2014, Frascati, Italy*
CROP / NON CROP MASK : Object-based analysis

Data used

- 2 SPOT images (maximum of the growing season and dry season)
- NDVI
- DEM + Slope
- Water system (extracted from DEM)
CROP / NON CROP MASK : Object-based analysis

Data used

- 2 SPOT images (maximum of the growing season and dry season)
- NDVI
- DEM + Slope
- Water system (extracted from DEM)
- Ground database about crop and no crop (400 GPS waypoints)
CROP / NON CROP MASK: Object-based analysis

Segmentation results
CROP / NON CROP MASK: Object-based analysis

1. Removing urban objects
   - Urban
   - Other
   Expert rules
   Ex: 
   - Urban = Brightness < 700 AND Euclidian texture > 4 AND NDVI < 0.48

2. Classification toposequence
   - Urban
   - Other
   - Shallows
   - Downslopes
   - Middle or high slopes
   - Uplands
   Expert rules
   Ex: 
   - Shallows / basins = Cross the river system

3. Classification crop / non crop
   - Urban
   - Other
   - Shallows
   - Crop
   - Non crop
   - Downslopes
   - Crop
   - Non crop
   - Middle or high slopes
   - Crop
   - Non crop
   - Uplands
   - Crop
   - Non crop

   Supervised classification using our ground DB.
CROP / NON CROP MASK: Object-based analysis

Context and objectives

Data

Crop Mask

Cropping systems mapping

Perspectives

Shallows: crop
Shallows: non crop
Downslope: crop
Downslope: non crop
Slope: crop
Slope: non crop
Upland: crop
Upland: non crop

% crop 58%
% non crop 42%
CROP / NON CROP MASK : Object-based analysis

Confusion matrix (5-fold cross validation)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Classification</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>crop</td>
<td>202</td>
<td>26</td>
<td>228</td>
<td></td>
</tr>
<tr>
<td>non-crop</td>
<td>33</td>
<td>63</td>
<td>96</td>
<td></td>
</tr>
</tbody>
</table>

Commission error 14% 29%
Omission error 11% 34%

OVERALL ACCURACY = 82%

Dissymmetry between crop and non-crop classes (dissymmetry in the reference data set? ).
Discovering new and useful knowledge from huge data

Data mining is the application of specific algorithms for extracting patterns from data.

Challenge: Spatio-temporal and heterogeneous data

Several steps: Data preprocessing (integration and cleaning), Data Mining (pattern extraction), Pattern validation

For what? Clustering, classification, summarization ...
<table>
<thead>
<tr>
<th>Context and objectives</th>
<th>Data</th>
<th>Crop Mask</th>
<th>Cropping systems mapping</th>
<th>Perspectives</th>
</tr>
</thead>
</table>

**CROP / NON CROP MASK : Data mining techniques**

1. **Attributes extraction at the plot scale**
   Radiometric, textural, temporal, and static information + cropping system type

2. **Extraction of frequent patterns for crop and non crop using different data mining algorithms**
   Learning on 80% of the database

3. **Patterns analysis and interpretation**: the classifier is built

4. **Crop / non crop classification on the entire zone**:
   The classifier or prediction model is able to classify new unseen objects only from their remotely-sensed attributes.

5. **Validation on 20% of the database**
<table>
<thead>
<tr>
<th>Plot ID</th>
<th>Crop</th>
<th>Cropping systems</th>
<th>NDVI MAX1</th>
<th>NDVI MEAN1</th>
<th>NDVI MAX2</th>
<th>NDVI MEAN2</th>
<th>NDVI MAX3</th>
<th>NDVI MEAN3</th>
<th>Texture variance MAX1</th>
<th>Texture contrast MAX1</th>
<th>Distance to river</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Crop</td>
<td>Maize</td>
<td>2800</td>
<td>2296</td>
<td>3721</td>
<td>2952</td>
<td>2736</td>
<td>2516</td>
<td>...</td>
<td>8</td>
<td>...</td>
<td>315</td>
</tr>
<tr>
<td>2</td>
<td>Crop</td>
<td>Rainfed Rice</td>
<td>2880</td>
<td>2238</td>
<td>4036</td>
<td>2874</td>
<td>3693</td>
<td>3262</td>
<td>...</td>
<td>32</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Crop</td>
<td>Maize</td>
<td>2236</td>
<td>1658</td>
<td>3602</td>
<td>1999</td>
<td>3195</td>
<td>2452</td>
<td>...</td>
<td>30</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Crop</td>
<td>Soybean</td>
<td>2032</td>
<td>1859</td>
<td>3167</td>
<td>2704</td>
<td>2500</td>
<td>2470</td>
<td>...</td>
<td>43</td>
<td>...</td>
<td>55</td>
</tr>
<tr>
<td>5</td>
<td>Crop</td>
<td>Maize</td>
<td>2296</td>
<td>1773</td>
<td>2848</td>
<td>2183</td>
<td>2769</td>
<td>2103</td>
<td>...</td>
<td>25</td>
<td>...</td>
<td>218</td>
</tr>
<tr>
<td>6</td>
<td>Non crop</td>
<td>Natural Vegetation</td>
<td>4074</td>
<td>3470</td>
<td>4740</td>
<td>3853</td>
<td>4632</td>
<td>3963</td>
<td>...</td>
<td>23</td>
<td>...</td>
<td>253</td>
</tr>
<tr>
<td>7</td>
<td>Crop</td>
<td>Cassava</td>
<td>3148</td>
<td>2168</td>
<td>4363</td>
<td>2976</td>
<td>4415</td>
<td>3248</td>
<td>...</td>
<td>58</td>
<td>...</td>
<td>292</td>
</tr>
<tr>
<td>8</td>
<td>Non crop</td>
<td>Natural Vegetation</td>
<td>2602</td>
<td>2233</td>
<td>3445</td>
<td>2690</td>
<td>3718</td>
<td>3056</td>
<td>...</td>
<td>152</td>
<td>...</td>
<td>308</td>
</tr>
<tr>
<td>9</td>
<td>Crop</td>
<td>Soybean</td>
<td>1903</td>
<td>1522</td>
<td>1816</td>
<td>1609</td>
<td>2926</td>
<td>2339</td>
<td>...</td>
<td>45</td>
<td>...</td>
<td>345</td>
</tr>
<tr>
<td>10</td>
<td>Non crop</td>
<td>Natural Vegetation</td>
<td>2841</td>
<td>2356</td>
<td>3791</td>
<td>3151</td>
<td>3710</td>
<td>3137</td>
<td>...</td>
<td>32</td>
<td>...</td>
<td>164</td>
</tr>
<tr>
<td>11</td>
<td>Non crop</td>
<td>Natural Vegetation</td>
<td>2817</td>
<td>2500</td>
<td>5185</td>
<td>4370</td>
<td>5797</td>
<td>5235</td>
<td>...</td>
<td>90</td>
<td>...</td>
<td>185</td>
</tr>
</tbody>
</table>

About 200 variables (qualitative and quantitative).
CROP / NON CROP MASK : Data mining techniques

Results for the different algorithms used:

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Overall accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaiveBayes</td>
<td>78.40%</td>
</tr>
<tr>
<td>J48</td>
<td>78.70%</td>
</tr>
<tr>
<td>SMO (SVM)</td>
<td>84.25%</td>
</tr>
<tr>
<td>RandomForest</td>
<td>82.10%</td>
</tr>
</tbody>
</table>

Confusion matrix (5-fold cross validation) for the SVM classification:

<table>
<thead>
<tr>
<th></th>
<th>crop</th>
<th>non-crop</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>crop</td>
<td>214</td>
<td>15</td>
<td>229</td>
</tr>
<tr>
<td>non-crop</td>
<td>36</td>
<td>59</td>
<td>95</td>
</tr>
</tbody>
</table>

Commission error 16% 20%
Omission error 7% 38%

OVERALL ACCURACY = 84%
### CROP / NON CROP MASK: methods comparison

<table>
<thead>
<tr>
<th></th>
<th>CROP</th>
<th>NON CROP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Omission error</td>
<td>Omission error</td>
</tr>
<tr>
<td></td>
<td>RS_method 11%</td>
<td>DM_method 7%</td>
</tr>
<tr>
<td>Producer accuracy</td>
<td>RS_method 89%</td>
<td>DM_method 93%</td>
</tr>
<tr>
<td>Commission error</td>
<td>RS_method 14%</td>
<td>DM_method 16%</td>
</tr>
<tr>
<td>User accuracy</td>
<td>RS_method 86%</td>
<td>DM_method 84%</td>
</tr>
<tr>
<td></td>
<td>RS_method 34%</td>
<td>DM_method 38%</td>
</tr>
<tr>
<td>Producer accuracy</td>
<td>RS_method 66%</td>
<td>DM_method 62%</td>
</tr>
<tr>
<td>Commission error</td>
<td>RS_method 29%</td>
<td>DM_method 20%</td>
</tr>
<tr>
<td>User accuracy</td>
<td>RS_method 71%</td>
<td>DM_method 80%</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td><strong>82%</strong></td>
<td><strong>84%</strong></td>
</tr>
</tbody>
</table>

Both methods provided stable results for the crop class
CHARACTERIZING CROPPING SYSTEMS: object-based image analysis

4 classes (cropping systems) inside the crop mask

- Rice (rainfed)
- Maïs (rainfed)
- Rice (irrigated)
- Other crops (soybean, carrot, cassava... (rainfed)

Supervised classification with 80% of the ground DB - 5 repetitions -
CHARACTERIZING CROPPING SYSTEMS: object-based image analysis

Confusion matrix (5-fold cross validation):

<table>
<thead>
<tr>
<th>CLASSIFICATION</th>
<th>Rainfed rice</th>
<th>Irrigated rice</th>
<th>Rainfed maize</th>
<th>Other crops</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainfed rice</td>
<td>18</td>
<td>6</td>
<td>12</td>
<td>9</td>
<td>45</td>
</tr>
<tr>
<td>Irrigated rice</td>
<td>2</td>
<td>30</td>
<td>1</td>
<td>4</td>
<td>37</td>
</tr>
<tr>
<td>Rainfed maize</td>
<td>13</td>
<td>6</td>
<td>45</td>
<td>24</td>
<td>88</td>
</tr>
<tr>
<td>Other crops</td>
<td>10</td>
<td>5</td>
<td>19</td>
<td>18</td>
<td>52</td>
</tr>
<tr>
<td><strong>Commission error</strong></td>
<td><strong>63%</strong></td>
<td><strong>15%</strong></td>
<td><strong>56%</strong></td>
<td><strong>62%</strong></td>
<td>222</td>
</tr>
<tr>
<td><strong>Omission error</strong></td>
<td><strong>58%</strong></td>
<td><strong>36%</strong></td>
<td><strong>42%</strong></td>
<td><strong>67%</strong></td>
<td></td>
</tr>
</tbody>
</table>

**OVERALL ACCURACY = 50%**

The errors are related to the cropping system class.
- Smaller errors for « irrigated rice » class (as expected)
- Larger errors for « other crops » class (very heterogeneous)
- Confusion between rainfed maize and rice (same seasonality)
CHARACTERIZING CROPPING SYSTEMS : using Data Mining

Results for the different algorithms :

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Overall accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaiveBayes</td>
<td>44.98%</td>
</tr>
<tr>
<td>J48</td>
<td>52.00%</td>
</tr>
<tr>
<td>SMO (SVM)</td>
<td>55.90%</td>
</tr>
<tr>
<td>RandomForest</td>
<td>50.22%</td>
</tr>
</tbody>
</table>

Confusion matrix (5-fold cross validation) :

<table>
<thead>
<tr>
<th>CLASSIFICATION</th>
<th>Rainfed rice</th>
<th>Irrigated rice</th>
<th>Rainfed maize</th>
<th>Other crops</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainfed rice</td>
<td>21</td>
<td>4</td>
<td>11</td>
<td>7</td>
<td>43</td>
</tr>
<tr>
<td>Irrigated rice</td>
<td>5</td>
<td>32</td>
<td>6</td>
<td>7</td>
<td>50</td>
</tr>
<tr>
<td>Rainfed maize</td>
<td>8</td>
<td>6</td>
<td>49</td>
<td>16</td>
<td>79</td>
</tr>
<tr>
<td>Other crops</td>
<td>11</td>
<td>4</td>
<td>16</td>
<td>26</td>
<td>57</td>
</tr>
</tbody>
</table>

Commission error  | 56%         | 28%           | 42%           | 53%         | 458   |
Omission error    | 51%         | 36%           | 38%           | 54%         |       |

OVERALL ACCURACY  =  56%
# CHARACTERIZING CROPPING SYSTEMS: methods comparison

<table>
<thead>
<tr>
<th>Crop Mask</th>
<th>Omission error</th>
<th>Commission error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RAINFED RICE</strong></td>
<td>58%</td>
<td>51%</td>
</tr>
<tr>
<td><strong>IRRIGATED RICE</strong></td>
<td>36%</td>
<td>36%</td>
</tr>
<tr>
<td><strong>MAIZE</strong></td>
<td>42%</td>
<td>38%</td>
</tr>
<tr>
<td><strong>OTHER CROPS</strong></td>
<td>67%</td>
<td>54%</td>
</tr>
<tr>
<td><strong>OVERALL ACCURACY</strong></td>
<td>50%</td>
<td>56%</td>
</tr>
</tbody>
</table>

Only one class with RS better results: Irrigated rice
CHARACTERIZING CROPPING SYSTEMS: the lessons of the case study

Globally DM is slightly better than OBIA method for cropping system classification

- OBIA classification is very time-consuming
- e-cognition proposed few choices in terms of classification algorithms

The + of Data-mining:
- You can put all the data you have (if the data are not relevant, the algorithm will skip them),
- Once the data base is ready, it is almost instantaneous

Importance of the toposequence in the image processing:
- -> need to incorporate in the image classification, the main driving factors of the cropping systems (when possible), especially for traditionnal agricultural systems where the environment is barely « corrected »
CHARACTERIZING CROPPING SYSTEMS: how to improve the results?

- Larger field campaign: **1000 GPS waypoints for 2013-2014 growing season**
- Contribution of Pleiades images (contribution of texture)
- Contribution of RADARSAT-2 images (free from clouds)

-> Capacities of Data Mining to extract knowledge from **very large amount of heterogeneous data** (satellite images from various sensors, DEM, soil type...) should be further investigated for Sentinel-2 images processing
PRELIMINARY RESULTS : Rice biomass assessment

A survey on 100 (2013) and 130 (2014) fields cultivated with rice (irrigated and rainfed): dry total biomass and grain biomass, sowing and harvest dates.

« On the fly » field selection (not planned).
PRELIMINARY RESULTS: Rice biomass assessment

Field selection using Duveiller’s method (based on time series):

Heterogeneous TS

Homogeneous TS
CHARACTERIZING CROPPING SYSTEMS: how to improve the results?

- Larger field campaign: **1000 GPS waypoints for 2013-2014 growing season**
- Contribution of Pleiades images (contribution of texture)
- Contribution of RADARSAT-2 images (free from clouds)

Next step: RICE YIELD ESTIMATION

- About **130 rice fields** sampled for yield estimation during the 2013-2014 growing season
- Improve field selection based on NDVI profiles
- Identification of phenological transition dates and rice yield estimations through analysis of SPOT **satellite time series metrics**.

*Project goes on thanks to JECAM, SIGMA, Sentinel-2 projects*
Thank you for your attention...