Mapping Crops and their Biophysical Characteristics with Polarimetric Synthetic Aperture Radar and Optical Remote Sensing

Part 3: Sen4Stat Open-Source Toolbox for Large Scale Crop Mapping (Theory and Practical)

Pr. Pierre Defourny (UCLouvain)

26 April, 2022
Training Outline

12 April
Part 1: SAR Polarimetry for Agriculture (Theory and Practice)

19 April
Part 2: Practical SAR Polarimetry with Sentinel-1, RCM, & SAOCOM Imagery for Agriculture

April 26
Part 3 - Sen4Stat Open-Source Toolbox (Theory and Practical)

May 3
Part 4: Crop-Specific Time Series Analysis for Growth Monitoring
Table of Contents

- Section 1: Challenges for large scale crop mapping at high spatial resolution
- Section 2: Large scale crop mapping using Sen4Stat toolbox
- Section 3: Practical using Sen4Stat toolbox for crop type mapping
- Section 4: Q&A session
New satellite missions and IT (r)evolutions have changed the game for agriculture.

Server farms with full EO data archives
Exploitation at low/no cost

Free, open and long term data policy (US, EU, …)

Open source processing tools using ML & AI algo.

EO ag. products => (intra-) field level
1. Cropland mapping
2. Crop type mapping
3. Biophys. variables (fCover, LAI, fAPAR, …)
4. Crop condition monitoring
5. Ag. practices monitoring

How to take advantage of Copernicus Sentinel-1 and Sentinel-2 data for crop mapping at national scale?
0 days 00 hours 00 minutes
Sentinel-2 constellation: summer solstice
Section 1: Challenges for large scale crop mapping at high spatial resolution
Challenges for large scale mapping

In addition to large data volume and the IT requirements

• Observing satellite systems at high resolution do not cover large area at once
  – Sentinel-2 instrument swath: 290 km
  – Sentinel-1 interferometric wide swath mode: 250 km
  – Overlap between images acquired from adjacent orbit
  – Increasing overlap due to converging orbit to North Pole

⇒ **Heterogeneity of observation density** across the area

• Cloud coverage frequency quite variable (optical systems) and nonsystematic acquisition plan over space and time (optical and SAR systems)

⇒ **Heterogeneity of valid observation density** across the area

• Significant **agro-climatic gradients** introducing large variability in crop calendars and vegetation development across the area
Challenges for large scale crop mapping – why is it so special?

3 challenges already identified

Temporal resolution critical for crop discrimination

Example of South Sudan (619.745 km²)

Nbr of granules for satellite data:
scenes (Landsat) and tiles (Sentinels)

Nbr of valid images (cloud cover < 10%)
from March 2017 to Feb. 2018
Challenges for large scale crop mapping

Example of South Sudan – **first solution to cope with heterogeneous valid observ. density**

Nbr of granules for satellite data: scene (Landsat) and tile (Sentinels)

Monthly synthesis (June 2017) by averaging all cloud free reflectances

=> **artefacts** due to the variability of input data

=> **reduction of temporal resolution** to monthly
Challenges for large scale crop mapping

Temporal synthesis methods averaging all cloud free reflectances acquired during the period

Sen4Stat: Weighted Average Synthesis Processor
- weight as function of date (central date distance)
- weight as function of Aerosol Optical Thickness
- weight as function of distance to clouds
- weight as function of sensor (S2>L8)

=> reduction of the compositing artefacts

=> monthly temporal resolution not really suitable for crop discrimination

Comparison of temporal synthesis methods

(Vancutsem et al., IJRS 2008)
Western Cape Province monitored by Sentinel 2 monthly composites
2016 winter grain production region (South Africa)
Challenges for large scale crop mapping

SAR temporal synthesis by averaging all SAR observations over the period for noise reduction

Mean of the coherence values as derived from Sentinel-1 interferometric pairs for March 2017 over Netherlands

Sentinel-1 color composite (blue: mean of the July coherence; green: mean of the March coherence; red: seasonal standard deviation)
Challenges for large scale crop mapping

Temporal information very critical for crop discrimination

*Second solution to cope with heterogeneous obs. density over large area*

Production of an **optical or SAR time series** which is:

- **gap-filled** using a linear interpolation method with respect to missing data (clouds, cloud shadows, no data, ...)
- **temporally resampled** at regular interval

⇒ This approach provides the same “observation” dates over the whole area independently from the sensor orbits and cloud cover which is very critical to run most of the classification methods.
Challenges for large scale crop mapping

Example of South Sudan

Major rainfall gradient across the country defining two different crop calendars

Stratification in zones to be classified separately
Challenges for large scale crop mapping

**Example of South Sudan – stratification**

A model to calibrate for each stratum requires collection of enough ground samples for each zone covering all crop types but also all non-cropland classes.

<table>
<thead>
<tr>
<th>Crop / Non crop</th>
<th>Land cover class</th>
<th>Nr of samples</th>
<th>Proportion in crop / no crop categories</th>
<th>Proportion in total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non crop</td>
<td>Bush</td>
<td>1504</td>
<td>44%</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>Forests</td>
<td>566</td>
<td>17%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>Grasslands and meadows</td>
<td>565</td>
<td>16%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>Shrubland</td>
<td>408</td>
<td>12%</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>Wetland</td>
<td>93</td>
<td>3%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>Water bodies</td>
<td>78</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>Fallow</td>
<td>76</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>Built-up surface</td>
<td>71</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>Bare soil</td>
<td>64</td>
<td>2%</td>
<td>1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Crop type</th>
<th>Nr of samples</th>
<th>Proportion in crop / no crop categories</th>
<th>Proportion in total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Sorghum</td>
<td>369</td>
<td>31%</td>
<td>8%</td>
</tr>
<tr>
<td>Late Sorghum</td>
<td>244</td>
<td>21%</td>
<td>5%</td>
</tr>
<tr>
<td>Maize</td>
<td>122</td>
<td>10%</td>
<td>3%</td>
</tr>
<tr>
<td>Groundnut</td>
<td>110</td>
<td>10%</td>
<td>3%</td>
</tr>
<tr>
<td>Early Sorghum Mixed</td>
<td>105</td>
<td>9%</td>
<td>2%</td>
</tr>
<tr>
<td>Cassava</td>
<td>79</td>
<td>7%</td>
<td>2%</td>
</tr>
<tr>
<td>Late Sorghum Mixed</td>
<td>55</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>Maize Mixed</td>
<td>48</td>
<td>4%</td>
<td>1%</td>
</tr>
<tr>
<td>Sesame</td>
<td>13</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>Rice</td>
<td>10</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>Bulrush Millet</td>
<td>9</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>Beans</td>
<td>7</td>
<td>1%</td>
<td>0%</td>
</tr>
</tbody>
</table>

=> 4602 samples well distributed
Section 2: Large Scale Crop Mapping Using Sen4Stat Toolbox
ESA open-source SAR and optical toolbox for operational crop type mapping and monitoring over very large areas

- Process Sentinel-1, Sentinel-2 and Landsat-8&9 time series according to the state-of-the-art including advanced SAR products (coherence, gamma naught, …)
- Deliver automatically or on request **5 types of products** (processors) in near real-time along the satellite data acquisition or off-line:
  1. **10m optical cloud free temporal synthesis and SAR temporal synthesis**
  2. **Time series of spectral indices** (NDVI, coherence, …) and **biophysical variables** (LAI, fCover, fAPAR)
  3. **10m crop type maps** along the season based on in situ dataset and stratification
  4. **Large set of crop growth conditions metrics** (including meteorological data)
  5. **Crop yield estimation** at various aggregation levels (national, regional, …)
– ESA open-source SAR and optical toolbox for operational crop mapping and monitoring over very large areas
ESA open-source SAR and optical toolbox with a graphical user interface to configure the system to launch the production and monitor the processes

**System configuration**

### Start of the season

**Sen4Stat: parameters settings**

- **Area of Interest**: Shapefile to be uploaded
- **Monitoring period**: Start and end dates to be defined
- **S1 - S2 - L8&9**: To be selected
- **Data sources**: ESA&USGS – AWS – DIAS – local storage...

### End of the season...

**Sen4Stat: field campaign**

- **Sampling design**: Stratification and sampling
- **Field visit**: In situ data collection – early survey
- **In situ data collection – mid-season survey**
- **Data upload**: Field data quality control and formatting

**Mali stratification (PIRT)**
system operation for crop growth monitoring:

Leaf Area Index (LAI) production in near-real time along the season
System operation for crop growth monitoring: **Leaf Area Index (LAI)** production in near real-time along the season.

Sen4Stat: 2016 – 2017 Leaf Area Index at 10 m

**Leaf Area Index**

*Sen4Stat* profile for a maize field near Potchefstroom - South Africa.
System operations running according to different modes:
NRT with orchestrator (fully automated) or on request

**Automated mode** through the web graphical user interface (GUI)

a) based on the Orchestrator with by-default parameterization, automatic data download and processing until the end of the season, on-time delivery

=> operational production in near real time (NRT)

b) Processor execution on user request, with by-default parameterization

**Manual mode**: to run processor independently, with custom parameters

a) through the GUI, with the Custom job approach

b) through SNAP software (only processor of Level 3 and 4)

c) in command line through a Linux terminal
Available reference dataset collected according to an area or list sampling frame (national statistical surveys, Ministry of Agriculture, ...)

JECAM Guidelines – ad hoc “windshield survey” by motorized vehicles (car, motorbike) selecting a set of appropriate roads (set 1) and complemented by regular additional transect (set 2) using secondary roads and tracks to try to reduce the spatial bias brought about by roadside sampling.

In situ data collection required for annual crop type mapping
- Requirements for the in situ data collection

- In situ data for **calibration (training)**: sampling to cover the diversity of situations existing in the study area (possibly a national territory) in order to **cover the range of possible signatures for the different elements of interest** (i.e. most crop types and the main non cropland classes).

- In situ data for **validation** to estimate the products accuracy (with a confidence interval) using a **statistically-sound sampling to be objective and independent**; for logistic reasons, sampling not strictly random but 2-stage sampling (with PSU and ESU) to assess the crop types (one field campaign)

⇒ **Villages partition**

\[\text{Primary Sampling Unit (PSU):} \quad \text{villages randomly selected}\]

\[\text{Elementary Sampling Unit (ESU):} \quad \text{fields randomly selected along roads}\]

⇒ **calibration and validation field campaigns for crop type mapping can be combined but the sampling design should be statistically sound to obtain fully independent dataset.**
1) High quality reference data collection:
   - Crop type acquired from field campaigns
   - Non-crop data obtained by visual interpretation of very high spatial resolution imagery available online

2) Quality control: geometry consistency, typology standardisation, homogeneity check...

3) Splitting polygons for calibration (25%) and validation (75%) for each class (alternative strategies might be required in case on poorly balanced in situ dataset)

4) Enhancing the samples number for marginal crop type using Synthetic Minority Over-sampling Technique (SMOTE)
• The Overall Accuracy (OA) is calculated as the total number of correctly classified pixels \( (n_{ii} \sim \) diagonal elements of the confusion matrix) divided by the total number of test pixels:

\[
OA = \frac{\sum_{i=1}^{r} n_{ii}}{\sum_{i=1}^{r} \sum_{j=1}^{r} n_{ij}}
\]

• **Precision** or **User's Accuracy** (UA) for the class \( i \) is the fraction of correctly classified pixels with regard to all pixels classified as this class \( i \) in the classified image:

\[
UA_{i} = \frac{\sum_{j=1}^{r} n_{ij}}{\sum_{j=1}^{r} n_{ij}}
\]

• **Recall** or **Producer's Accuracy** (PA) for the class \( i \) is the fraction of correctly classified pixels with regard to all pixels of that ground truth class \( i \):

\[
PA_{i} = \frac{\sum_{j=1}^{r} n_{ij}}{\sum_{j=1}^{r} n_{ij}}
\]

• **F-Score** (also known as F-1 score or F-measure) is the harmonic mean of the Precision and Recall and reaches its best value at 1 and worst score at 0:

\[
F_{\text{Score}} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]
– System operation in automated mode to deliver in NRT 10-m crop type maps at national scale

Before the start of the monitoring period

**System initialization**

- Cloud computing or automatic EO data download
- Manual in situ data upload

EO data providers

[Defourny et al., RSE2019]
- System operation in automated mode to deliver in NRT
10-m crop type maps at national scale

Before the start of the monitoring period

System initialization

- Cloud computing or automatic EO data download
- Manual in situ data upload

Monitoring period

EO data providers

Operators

Mid-season crop type map
238 397 km²
System operation in automated mode to deliver in NRT 10-m crop type maps at national scale

Before the start of the monitoring period

System initialization

Monitoring period

- Cloud computing or automatic EO data download
- Manual in situ data upload

EO data providers

Operators

Mid-season crop type

238 397 km²

EO data providers

- Sentinel-2
- Sentinel-1
- Landsat

EO data providers

- Sentinel-2
- Sentinel-1
- Landsat
– System operation in automated mode to deliver in NRT 10-m crop type maps at national scale

Before the start of the monitoring period

System initialization

Monitoring period

- Cloud computing or automatic EO data download
- Manual in situ data upload

EO data providers

- Sentinel-2
- Sentinel-1
- Landsat

Cloud computing or automatic EO data download

Manual in situ data upload

Mid-season crop type map

238,397 km²

End of season crop type map & accuracy assessment

Operators

NASA’s Applied Remote Sensing Training Program
– Very first 10-m national cropland and crop type maps based on Sentinel-2a only (Ukraine, July 2016)

Overall accuracy: 96% (Defourny et al., 2019)

Overall accuracy: 81%
– Cropland map at mid and end of the season for the two strata of South Sudan

F1-Score of cropland increases with the amount of EO data available along the season
- Crop mapping in Mali (2017)

- Successful crop mapping at 10m resolution in collaboration with ICRISAT
- ~4.5 TB of Sentinel-2 data (2152 S2 and 542 L8 images)

Cropland mask for the 2017 growing season (end-of-season)
OA = 96%

Crop type map for the 2017 growing season (end-of-season)
OA = 66%
Sen2Agri system: Open source system precursor of Sen4Stat running in NRT or off-line, running locally or in the cloud

Version 2.0 released

Documented and downloadable at www.esa-sen2agri.org/resources/software/

Monthly cloud free surface reflectance composite at 10-20 m

Cropland mask at 10m updated every month
Crop type map at 10 m for the main regional crops
Vegetation status map at 10 m delivered every week (NDVI, LAI, ICover, fAPAR)

Sen2Agri as a service on the EODC cloud (e.g. for WFP)
Open-source operational solution for national processing in near real time and for local applications

- Under GNU-GPL License
- Open-source software based on Orfeo ToolBox
- Cluster-ready architecture for distributed processing
- Integration with Sentinel-2 ToolBox
- Operational system required: CentOS 7 (GNU/Linux)

State-of-the-art SAR processing for high level products

5 main processors to be possibly combined
State-of-the-art SAR processing for high level products for amplitude and coherence

\( \beta_0 \) (slant range geometry)

\( \sigma_0 \) (corrected for local inc. angle)

\( \gamma_0 \) (flattened corrected using DEM)
Crop type mapping module (S4S_Crop mapping processor)

- Random Forest as classifier by default (deep learning alternative: Neural Net [Transformer])
- Specific quality control of in situ dataset to balance the calibration dataset
- Large set of features used by the Random Forest model including:
  - S2 & L8 surface reflectance time series and S1 amplitude and coherence time series
  - NDVI, NDWI, Red Edge Indices, brightness and S1 polarization ratio time series
  - Temporal metrics from above time series and S1 temporal metrics from SAR time series
– Random Forest (RF)

- Decision Trees ~ recursive partitions calibrated on training set to get regions (nodes) increasingly homogeneous with respect to their class variable

- Random Forest ~ ensemble-learning algorithm that combines multiple classifications of the same data to produce higher classification accuracies

- Sequence:
  - **Bagging**: individual trees are grown from different subsets of training data (random subsampling of the original data for growing each tree)
  - **Single majority vote** to identify the most frequent class decided from all trees
Very similar accuracies for Random Forest and Neural Net (Transformer)

Overall Accuracy (OA): 80% (crop type) & 88% crop group
NASAs Applied Remote Sensing Training Program

Random Forest classification based on S2 & L8 time series only

Overall Accuracy: 96%
F-Score cropland: 0.97
F-Score non-cropland: 0.88

2000 GPS traces for holdings selected through a stratified sampling from the 526,000 holdings in Senegal (~ 0.4%)
- Quality of in situ data as major bottleneck today

Poor accuracy of crop type map because of the unbalanced ground observation dataset

<table>
<thead>
<tr>
<th></th>
<th>Arachide</th>
<th>Mais</th>
<th>Millet</th>
<th>Nièbé</th>
<th>Sorgho</th>
<th>Autres</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arachide</td>
<td>23295</td>
<td>0</td>
<td>6</td>
<td>1101</td>
<td>158</td>
<td>1184</td>
</tr>
<tr>
<td>Mais</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Millet</td>
<td>178</td>
<td>0</td>
<td>918</td>
<td>93</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Nièbé</td>
<td>3445</td>
<td>0</td>
<td>0</td>
<td>11318</td>
<td>158</td>
<td>1068</td>
</tr>
<tr>
<td>Sorgho</td>
<td>1996</td>
<td>0</td>
<td>0</td>
<td>1200</td>
<td>443</td>
<td>270</td>
</tr>
<tr>
<td>Autres</td>
<td>5037</td>
<td>0</td>
<td>84</td>
<td>2610</td>
<td>191</td>
<td>4256</td>
</tr>
</tbody>
</table>

| Total Areas (km²) | groundnut | goundnut | berisssap | bissap | coton | cotton | diakhatou | fonio | fonio | gombo | gombo | maize | maize | millet | millet | nièbé | nièbé | nièbé | nièbé | piment | piment | rice | rice | rice | rice | rainned | rainned | rice | rice | rice | rice | soeum | soeum | sorghum | sorghum | watermelon | watermelon | other | other | other | other | void | void | void | void | void |
|-------------------|-----------|---------|-----------|--------|-------|--------|-----------|-------|-------|-------|-------|-------|-------|---------|---------|-------|-------|-------|-------|--------|--------|-------|-------|-------|-------|--------|--------|---------|---------|-----------|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
Remote Sensing Training Program

Specific in situ data collection for department of Nioro
(Senegal, 2021)

- Calibration dataset
- Validation dataset

Calibration polygons
Validation polygons

Maize
Millet
Groundnut

Maize
Millet
Groundnut

Calibration polygons
Validation polygons

Maize
Millet
Groundnut

0 10 km

Maize
Millet
Groundnut

EOStat

NASA’s Applied Remote Sensing Training Program
Each S2 image covering only a part of the department

- 10-m cropland map using balanced in situ data collection for the Department of Nioro (Senegal)

Overall Accuracy: 97.4%
F-Score cropland: 0.98
F-Score non-cropland: 0.95
10-m crop type map from Sentinel-1 and Sentinel-2 time series acquired over the season and using balanced in situ dataset

Season from 01-05-2021 to 31-12-2021
Overall Accuracy: 84.8 %
- Operational exploitation of satellite missions to improve timely information about crop areas, yields and production

Area sampling frame

EO Multi-sensor dataset

Production of crop type maps

Field surveys

Agricultural statistics improvement

Quality control of survey data

Key information for decision making processes
Section 3: Practical using Sen4Stat toolbox for crop type mapping
Contacts

• Trainers:
  – Pr. Pierre Defourny pierre.defourny@uclouvain.be

• Training Webpage:

• ESA’s EO4Society Website:
  – https://eo4society.esa.int/training-education/

• Twitter: @EOOpenScience
Useful links

- **Sen4Stat**
  https://esa-sen4stat.org/

- **Sen2Agri**
  https://github.com/Sen2Agri/Sen2Agri-System/

- **SNAP download**
  http://step.esa.int/main/download/snap-download/

- **Copernicus Open Access Hub (download Sentinel data)**
  https://scihub.copernicus.eu/

- **NASA ARSET** Agricultural Crop Classification with Synthetic Aperture Radar and Optical Remote Sensing Part 5: Biophysical Variable Retrieval using Optical Imagery to Support Agricultural Monitoring Practices
  https://appliedsciences.nasa.gov/join-mission/training/english/arset-agricultural-crop-classification-synthetic-aperture-radar-and

- **ESA ‘Advanced Training Course on Land Remote Sensing with the Focus on Agriculture’ held in Louvain-la-Neuve, Belgium, on 16-20 September 2019**
  https://eo4society.esa.int/resources/advanced-training-course-on-land-remote-sensing-with-the-focus-on-agriculture/
Sen4Stat Installation Package  (available in June 2022)

- Single installation archive containing all files needed
  - Documentation (docs)
  - Installation scripts
  - Sen4Stat rpm files
  - Sen4Stat Services
  - Other tools and configuration files
- CentOS 7 is required for system installation
- Installation performed through a single script
  - `sudo ./install.sh`
- Two directories with write permissions for all users are needed:
  - `/mnt/archive`
  - `/mnt/upload`
- During installation, all needed third party packages are installed via `yum`. Needed docker images are installed
- PostgreSQL database is used for storing system configuration, product information etc.
- The regular user `sen2agri-service` is created during installation
- System services are created and launched upon installation:
  - `sen4stat-executor` `sen4stat-orchestrator` `sen4stat-http-listener` `sen4stat-demmacs`
  - `sen4stat-demmacs.timer` `sen4stat-monitor-agent` `sen4stat-scheduler` `sen4stat-services`
Homework and Certificate

• Homework Assignment:
  – Answers must be submitted via Google Form
  – Due Date: May 17, 2022

• A certificate of completion will be awarded to those who:
  – Attend all live webinars
  – Complete the homework assignment by the deadline (access from website)
  – You will receive a certificate approximately two months after the completion of the course from: marines.martins@ssaihq.com
Questions?

- Please enter your questions in the Q&A box. We will answer them in the order they were received.
- We will post the Q&A to the training website following the conclusion of the webinar.

https://earthobservatory.nasa.gov/images/6034/pothole-lakes-in-siberia
Thank You!