# ESA Sen4Stat
# Sentinels for Agricultural Statistics

## D16.0 v1 – ER
## "Exploitation Report"

Issue/Rev: 1.1

Date Issued: 30-03-24

Milestone: 6

Submitted to European Space Agency

# Consortium Partners

| Participant Organisation Name | Acronym | City, Country |
|---|---|---|
| Université catholique de Louvain | UCLouvain | Louvain-la-Neuve, Belgium |
| CS Romania | CS RO | Craiova, Romania |
| Systèmes d’Information à Référence Spatiale SAS | SIRS | Villeneuve d’Ascq, France |
| Universidad Polytecnica de Madrid | UPM | Madrid, Spain |

## Contact

Université catholique de Louvain – Earth and Life Institute
Place de l’Université, 1 – B-1348 Louvain-la-Neuve – Belgium
Email : Sophie.Bontemps@uclouvain.be
Internet : https://uclouvain.be/en/research-institutes/eli/elie

## Disclaimer

# Document sheet

## Authors and Distribution

| | |
|---|---|
| **Authors** | Sophie Bontemps, Nicolas Deffense, Boris Norgaard, Pierre Defourny |
| **Distribution** | ESA - Zoltan Szantoi |

## Document Status

| Issue/Rev. | Date | Reason |
|---|---|---|
| 0.1 | 01/02/2024 | Initial version of the deliverable |
| 1.0 | 30/06/2024 | First version submitted to ESA |
| 1.1 | 30/03/2025 | Update of the first version, after RIDs received from ESA |

## Detailed record sheet

*From version 1.0 to 1.1*

| Comment | Section | Problem description | Change |
|---|---|---|---|
| ESA RID | 3.2.1.1 | Blurry text in figure | Figure 3-1: quality of the figure improved |
| ESA RID | 3.2.1.1 | RMSE needs to be clarified | Definition of the error metrics are provided on page 23 (MAE, RMSE, RRMSE) and their values are commented in the paragraphs below on pages 23 and 24. Message of figure 3-1 clarified |
| ESA RID | 3.2.1.1 | Blurry text in figure | Figure 3-2: quality of the figure legend improved |
| ESA RID | 3.2.1.2 | Blurry text in figure | Figures 3-5 and 3-6: quality of the figure improved, French removed |
| ESA RID | 3.2.1.2 | Blurry text in figure | Figure 3-8: French removed |
| ESA RID | 3.2.2 | Blurry text in figure | Figure 3-10: quality of the figure improved, French removed |

# Table of contents

## List of figures

# List of tables

# 1 Introduction

## 1.1 Purpose and scope

This document is the Exploitation Report (ER) of the Sentinels for Agricultural Statistics (Sen4Stat) project funded by the European Space Agency (ESA).

The overall objective for the Sen4Stat project is to facilitate the uptake of Earth Observation (EO) information in the National Statistical Offices (NSO) supporting the agricultural statistics. Special attention shall be given to develop and demonstrate EO products and best practices for agriculture monitoring relevant for Sustainable Development Goals (SDG) reporting and monitoring their progress at national scale

The ER is one of the key outputs of the Task 5 (WP 5000) of the Sen4Stat project, named “Full-scale demonstration” (Figure 1-1). It aims at documenting the feedback received from the pilot NSOs about the Sen4Stat EO products and about the system (if tested).

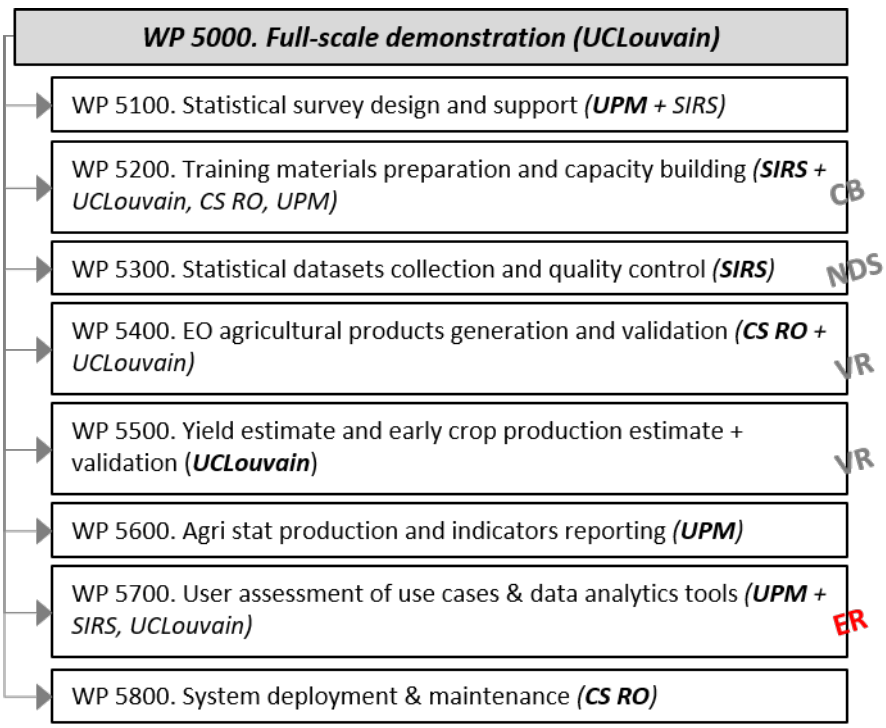


Figure 1-1. Organization of the Task 5 activities (from [AD.2])

## 1.2 Structure of the document

After this introduction, this document contains 2 sections, dedicated to Spain and Senegal, the two pilot countries for which statistical use cases have been successfully developed.

## 1.3 References

### 1.3.1 Applicable documents

| ID | Title | Reference | Issue/Rev. | Date |
|---|---|---|---|---|
| AD.1 | Statement of Work for ESA Sentinels for Agricultural Statistics | EOEP-EOPS-SW-17-015 | 1.0 | 15/03/2017 |
| AD.2 | Sen4Stat Implementation Proposal - Chapter 5 | | 1.0 | 12/05/2017 |

Table 1-1. Applicable documents

### 1.3.2 Reference documents

| ID | Title |
|---|---|
| RD.1 | Sen4CAP Validation Report, v1.2, 21/05/2021 |

Table 1-2. Reference documents

### 1.3.3 Acronyms and abbreviations

| Acronym | Definition |
|---|---|
| AAS | Annual Agricultural Survey |
| AD | Applicable Document |
| EO | Earth Observation |
| ER | Exploitation Report |
| ESA | European Space Agency |
| FAO | Food and Agriculture Organization |
| GPS | Global Positioning System |
| ID | Identifier |
| NSO | National Statistical Office |
| ODK | Open Data Kit Collect |
| RD | Reference Document |
| SDG | Sustainable Development Goal |
| Sen4CAP | Sentinels for Common Agricultural Policy |
| Sen4Stat | Sentinels for Agricultural Statistics |

Table 1-3. List of acronyms and abbreviations

# 2 Spain

## 2.1 NSO's requirements and theoretical background of the use cases

**Design-consistence** is the basic requirement of official statistics, and **design-based inference** is the statistical approach to achieve this requirement. This approach consists in two main tasks: (i) sampling design and (ii) estimation.

In the former, a probabilistic scheme for selecting the sample is defined and the sampling distribution generated by this scheme is identified. In the latter, the required estimates of the population (agriculture) characteristics are computed using sample data and estimators. The estimates uncertainty is assessed using the sampling distribution identified in the former.

In sampling, two kinds of data are clearly differentiated: **survey data** and **auxiliary data**. **Survey data** must be of high quality, i.e. unbiased and reliable. To ensure the required quality, they must be carefully collected on the ground, using well stablished protocols that include the use of unbiased measurement instruments. In addition, their quality must be controlled using the same protocol than for data collection. As a result, survey data are expensives and are thus official statistics. This is why survey data are observed in a sample of the population as small as possible to achieve the required estimates accuracy.

**Ancillary data** are not required to be unbiased, nor that their reliability to be very high: it is enough that they are available for free and that their correlation with the survey data is not null. EO data belongs to the auxiliary category: they are available for free and their correlation with most of the survey variables considered in agricultural statistics is non-null. To study EO data contribution to the agriculture statistics, use cases will consider the two main tasks of the statistical framework: (i) sample design and (ii) estimation.

Spanish NSO's expectations in terms of EO data can be summarized as follows:

1) *Improving design-based* crop acreage estimators by *optimizing* the integration of EO and ground data in the calculation of crop acreage statistics *to reduce the survey costs (without increasing the standard error);*
2) *Improving design-based* crop production estimators by *optimizing* the integration of EO and ground data in the calculation of crop acreage statistics *to reduce the survey costs (without increasing the standard error);*
3) Improving design-based estimators to *disaggregate the estimates* obtained at the national level, at the level of minor administrative units (region/province/canton/county);
4) Improving the timeliness of the estimates.

Improving design-based estimators (points 1 and 2 above) is a comprehensive objective as it implies improving both the estimation procedure and the sample design.

Various statistical methods developed to integrate EO ancillary data with survey data exist in the sampling survey literature. The statistical methods that can answer Spanish NSO's expectations are cost-efficiency, domain and small area estimation, timeliness and optimizing the sample design.

### 2.1.1 Cost-Efficiency

Official statistics is expensive, mainly because they must be based on high quality unbiased and reliable data. As a result, the survey cost is a key criterion for choosing one among a set of sampling techniques. The other key criterion is accuracy, which includes both bias and sampling variance: from now on, we limit ourselves to design-consistent sampling surveys which are unbiased (or approximately unbiased) so that the accuracy measure will be the sampling variance. Cost-efficiency can therefore be calculated as the survey cost multiplied by the sampling variance, so that it integrates the two key criteria in only one.

As already mentioned, in order to improve design-based estimators, it is convenient to differentiate between EO contributions to (i) the sampling design, and to (ii) the estimation. This use case focus on the estimation while the EO contribution for optimizing the sampling design will be treated in a different use case (see section 2.1.4).

In this cost-efficiency use case, what we do is:

a) Integrating EO data with the statistical survey provided by the NSO,
b) Evaluating the effect on the cost-efficiency of the sampling design currently used.

To evaluate this effect, we compare the cost-efficiency of the current sampling design where only ground data (without EO data) are used, with the cost-efficiency of the current sampling design where both ground data and EO data are used.

Spanish NSO expects that EO data can reduce the survey cost, without increasing the standard error.

In the following lines, let's call the ground data "G" and the EO data "EO"

It is assumed that unitary cost (cost per sampling unit) is constant $C_0$ (equal for every sampling unit), so that the total cost, $C=C_0n_G$, is proportional to the ground sample size, $n_G$. Let $V_G$ be the sampling variance using the current sampling design where only ground data G are observed (without EO data). Then, the cost-efficiency of the current sampling design is $C_0n_G*V_G$.

Let's be $n_{G+EO}$ the sample size when EO data is integrated in the sampling subject to the constraint that $V_{G+EO}= V_G$. The total cost using the current sampling design plus EO data is $C_0n_{G+EO}$, and the cost-efficiency is $C_0n_{G+EO}*V_G$.

The effect of the integration of EO data in the current sampling survey subject to the constraint $V_{G+EO}= V_G$, is $EO_{effect}= (C_0n_G*V_G- C_0n_{G+EO}*V_G)/ C_0n_G*V_G=1-n_{G+EO}/n_G$. The expected effect is a reduction, $(n_G-n_{G+EO}) =EO_{effect}*n_G$, of the sample size without increasing the sampling error (since the constraint $V_{G+EO}= V_G$ holds). The new sample size would be $n_{G+EO}= (1-EO_{effect})*n_G$, and the new survey cost would be $C_{G+EO}=C_0n_{G+EO}= (1-EO_{effect})C_0n_G$; the reduction factor, $(1-EO_{effect})=1/RE$ , is the inverse of the relative efficiency.

This use case will be applied to the estimators of the crop acreage for the main crops in the areas of interest.

### 2.1.2 Domain and small area estimation

In Spain, the current sample was designed to achieve the required estimates accuracy at the national level. However, reliable estimates over minor administrative areas, such as region/province and county, are also required without increasing the sample size – this has been stated by the NSO.

In a minor administrative area, the sample size will be always lower than at the national level and, as a result, the estimators' accuracy will be also lower. In the literature on sampling survey:

- a domain is a part of the population (say a region/province) where the sample size is big enough for the design-based estimator to be sufficiently precise for most uses;
- a small area is a part of the population (say a county) where, due to the small sample size, the design-based estimator is not sufficiently precise for most uses.

In this use case, we will consider an alternative GREG estimator, which is design-based and more accurate for domain estimation than the projective estimator used at national-level. For small area estimation, we will use a model-based estimator to "borrow strength" from related small areas in order to obtain precise estimates for a given small area. This estimator makes optimal use of the available data, according to statistical criteria, and allows for providing estimates even in counties where the sample size is null. The use of EO data is key for this application.

This use case will be applied applied to the estimators of the crop acreage for the main crops in the areas of interest.

### 2.1.3 Timeliness

Official statistics are published a long time after the end of the campaign, thus being not available at the right time to take decisions. The Spanish NSO expects that EO data will contribute improving this timeliness. Three main applications were initially considered here:

- Getting crop acreage forecasts at the mid-season and crop yield forecast one month before the harvest;
- Supporting a more rapid publication of consolidated statistics;
- Providing multi-seasonal estimates.

The demonstration of the second application is unfortunately not in our hands. It will be discussed with the NSO during our iterations but providing a clear demonstration in the framework of the project is not feasible.

As for the latest application, it is very challenging because it requires a two-phase sampling strategy and regression estimators for integrating ground and EO data. The first phase sample would be the current sample selected by the country. For each season, a second phase sample would need to be selected among the sampling units included in the first phase. The simplest version of this design would be a sample with two independent components: one component as a pure panel sample common to every season, and the other component as a specific sample for each season (the supplementary sample). We would aim at showing how to estimate (i) the total of the survey variables in each season and its annual aggregated, as well as (ii) the change of these totals between two consecutive seasons and its annual aggregate. While looking attractive, this multi-seasonal estimation might be difficult to achieve as it requires that one pilot country implements this two-phase sampling strategy. After discussion with the NSO, it was not considered for the project.

### 2.1.4 Optimizing the sample design

The procedure for elaborating agricultural statistics begins with the design of the sample for ground data collection, and it finishes with the computation of the required estimates. In all the other use cases, the starting point is the sample already existing in the pilot country, and we focus on how the integration of ground and EO data improves the estimators.

However, *optimizing* the sample design is also key, and has been mentioned by the Spanish NSO. The specific expectation here is to generate a map of irrigated areas that would be used to update the stratification that serves as a basis for them to define their samples allocation.

## 2.2 Cost-efficiency use case

Figure 2-1 presents the results for the cost-efficiency use case in Castilla y Leon, for the three main crops which are wheat, maize and sunflower. For each crop, the figure shows the acreage estimate based on ground data only (ESYRCE being the name of the agricultural survey) and on the coupling of ESYRCE with EO data. More interestingly, the figure also provides the confidence intervals around these estimates. The EO impact is a systematic reduction of the interval, and thus of the sampling error (highlighted in Figure 2-2). As a result, the relative efficiency of the coupling between EO and ESYRCE datasets is high.

| | Data | Acreage (hectares) | Uncertainty | | | Relative efficiency |
|---|---|---|---|---|---|---|
| | | | 95% Confidence Interval (hectares) | | Sampling Error (CV%) | |
| | | | Limits | Amplitude | | |
| Barley (F-Score: 0,875) | Ground (ESYRCE) | 980 081 | Lw: 924 173<br>Up: 1 037 461 | 113 288 | 2,95 | ------ |
| | Gound+RS | 923 026 | Lw: 899 364<br>Up: 946 689 | 47 325 | 1,31 | 5,73 |
| Maize (F-score: 0,970) | Ground (ESYRCE) | 210 558 | Lw: 176 727<br>Up: 244 389 | 67 662 | 3,18 | ------ |
| | Ground+RS | 132 901 | Lw: 124 606<br>Up: 141 195 | 16 589 | 8,2 | 16,64 |
| Wheat (F-score: 0,880) | Ground (ESYRCE) | 877 717 | Lw: 827 030<br>Up: 928 404 | 101 374 | 2,94 | ------ |
| | Ground+RS | 812 088 | Lw: 787 877<br>Up: 836 299 | 48 422 | 1,52 | 4,38 |

Figure 2-1. Cost-efficiency use case in Castilla y Leon (Spain, 2020)

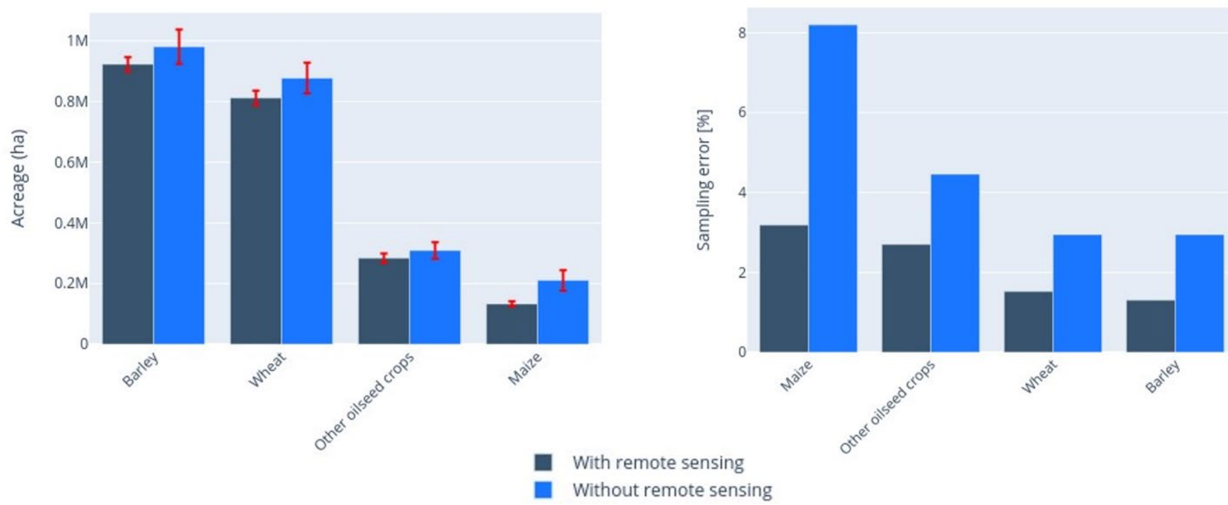


Figure 2-2. Acreage estimation of the predominant classes in Castilla y Leon (2020) presented with their confidence interval (left) and sampling error (right), with (grey) and without (blue) EO data

Similarly, Figure 2-3 presents the results for the cost-efficiency use case in Andalusia, for the three main crops which are olive groves, wheat and sunflower. Here also, the EO impact is a systematic reduction of the interval, and thus of the sampling error, as shown in Figure 2-4, and it can be concluded that the relative efficiency of the coupling between EO and ESYRCE datasets is high.

| | Data | Acreage (hectares) | Uncertainty | | | Relative efficiency |
|---|---|---|---|---|---|---|
| | | | 95% Confidence Interval (hectares) | | Sampling Error (CV%) | |
| | | | Limits | Amplitude | | |
| Olive groves (F-Score: 0,897) | **Ground (ESYRCE)** | 1 624 187 | Lw: 1 529 848<br>Up: 1 718 525 | 188 677 | 2,96 | ------ |
| | **Ground+RS** | 1 740 863 | Lw: 1 693 061<br>Up: 1 788 664 | 95063 | 1,40 | 3,896 |
| Wheat (F-score: 0,834) | **Ground (ESYRCE)** | 398 357 | Lw: 358 596<br>Up: 438 119 | 79523 | 5,09 | ------ |
| | **Ground+RS** | 391 056 | Lw: 369 952<br>Up: 412 161 | 42209 | 2,75 | 3,55 |
| Oilseed crops (F-score: 0,931) | **Ground (ESYRCE)** | 229 049 | Lw: 197 700<br>Up: 260 399 | 62699 | 6,98 | ------ |
| | **Ground+RS** | 210 232 | Lw: 196 088<br>Up: 224 377 | 28289 | 3,43 | 4,91 |

Figure 2-3. Cost-efficiency use case in Andalusia (Spain, 2020)

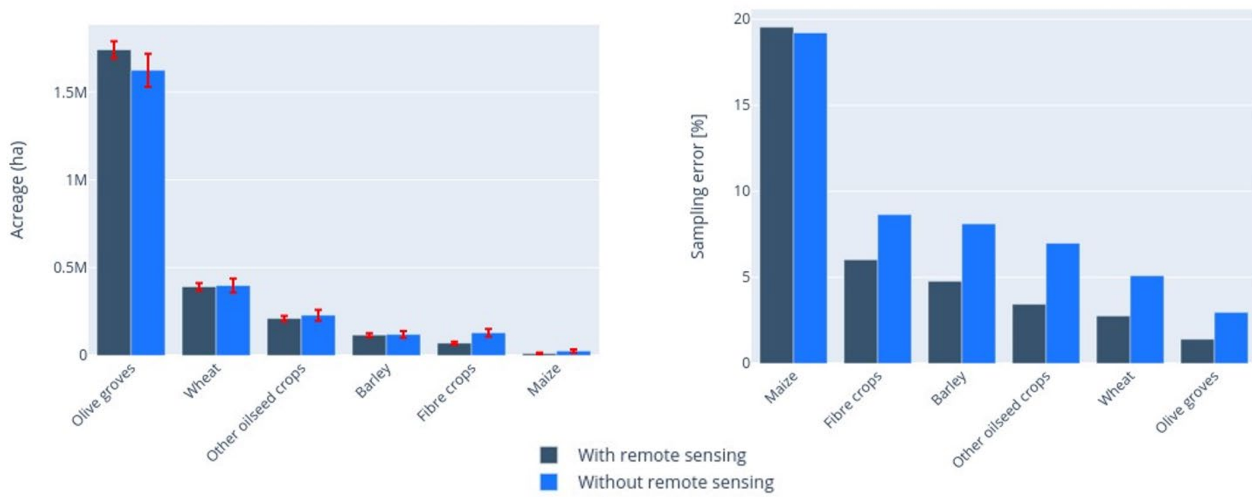


Figure 2-4. Acreage estimation of the predominant classes in Andalusia (2020) presented with their confidence interval (left) and sampling error (right), with (grey) and without (blue) EO data

It is planned to do the same analysis with the national-scale land cover map but the quality at the current moment is not good enough, mainly due to artefacts. This exercise will be carried out during the project extension.

The cost-efficiency use case was also considered for the yield estimation, but in a slightly different way. Indeed, the estimation of the yield was set up to show that estimating yield on a larger sample of data (i.e. with EO data) can improve confidence in aggregate statistics by virtually increasing the number of data points collected in the survey. The cost-efficiency use case was therefore presented in the Validation Report, intrinsically linked with the yield estimation in itself. The main outcome is reminded here in Table 2-1.

Table 2-1. Yield estimation of the provinces of Castilla-y-Lèon (kg/ha) given by ESYRCE and both models (Null and RS). The average yield, the standard deviation, and the mean absolute error computed on the ten repetitions of estimation are presented.

| | ESYRCE | | Null Model (10x) | | | | S4S RS Model | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | Yield | N | Mean | Sd | MAE | N | Mean | Sd | MAE |
| Àvila | 151 | 4250.2 | 107 | 4241.5 | 83.0 | 84.7 | 150 | 4232.4 | 34.9 | 37.9 |
| Burgos | 446 | 4852.4 | 315 | 4826.8 | 64.9 | 69.6 | 446 | 4764.3 | 38.2 | 88.1 |
| Leòn | 52 | 3792.7 | 37 | 3822.0 | 103.8 | 109.7 | 52 | 3817.5 | 57.0 | 59.2 |
| Palencia | 304 | 4585.6 | 211 | 4602.1 | 32.3 | 39.2 | 302 | 4557.5 | 17.0 | 29.9 |
| Salamanca | 122 | 4204.3 | 87 | 4155.8 | 63.1 | 81.5 | 122 | 4155.8 | 57.9 | 72.3 |
| Segovia | 294 | 4169.5 | 206 | 4168.0 | 52.5 | 52.8 | 294 | 4134.1 | 35.4 | 50.1 |
| Soria | 275 | 3617.5 | 192 | 3640.1 | 35.2 | 40.3 | 275 | 3542.6 | 26.8 | 74.9 |
| Valladolid | 460 | 4588.2 | 320 | 4574.6 | 37.8 | 41.4 | 459 | 4531.1 | 26.5 | 57.1 |
| Zamora | 206 | 4600.0 | 142 | 4586.8 | 65.0 | 67.2 | 204 | 4569.1 | 54.7 | 60.4 |
| Castilla Y Leòn | 2310 | 4437.2 | 1617 | 4426.5 | 16.5 | 20.8 | 2304 | 4391.9 | 14.0 | 45.3 |

We showed that the model incorporating the remote sensing variables, although not capable of accurately estimating yields at the plot scale, can be used to synthetically augment data in poorly represented statistical units and thus improve the robustness of estimates at this scale. Since the use of the RS model greatly reduces the standard deviation of the estimates, it is likely that improving the performance of the estimation model at the field level would allow the number of samples to be measured in the field to be reduced while maintaining the same confidence in the estimates.

## 2.3 Domain and small area estimation use case

This use case aims at demonstrating how EO data contribute to estimate agricultural statistics over small administrative units for which there is not enough sample to ensure a good accuracy of the estimators. The targeted administrative units in Spain are illustrated in Figure 2-5.

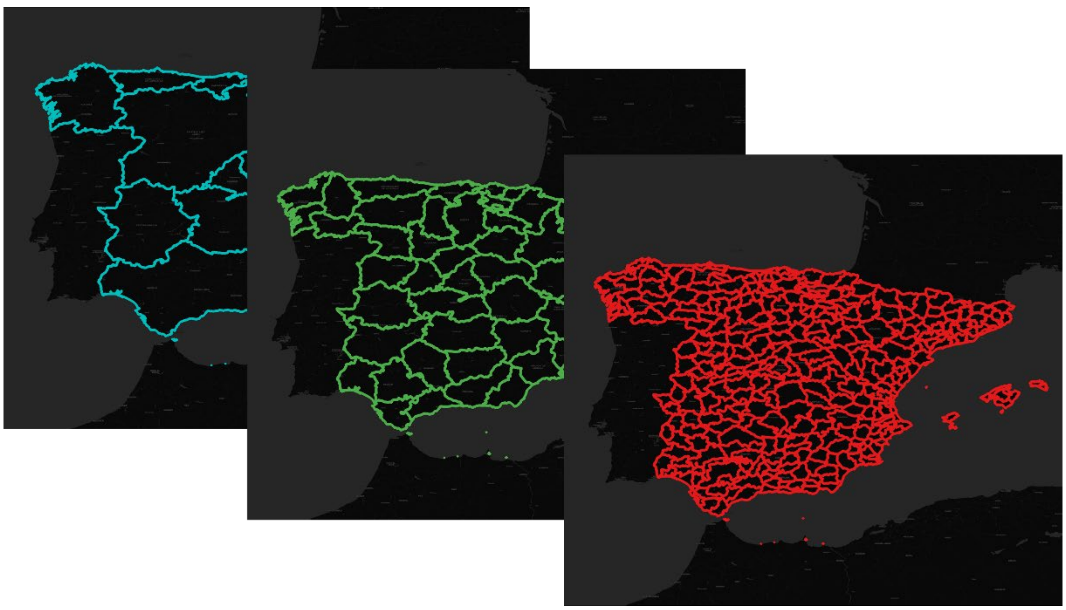

Figure 2-5. Illustration of the administrative units targeted by the use case: autonomous community, provinces and municipalities

The test case was demonstrated in Castilla y Leon. Figure 2-6 shows the estimates of barley acreage for four provinces in Castilla y Leon. It can be seen that the estimates are quite similar without and with EO data, but that the sampling error is significantly reduced when EO data is integrated. This is illustrated in a different way in , emphasizing the decrease of the sampling error.

| REGION | ESYRCE | | ESYRCE+EO | | Relative Efficiency |
|---|---|---|---|---|---|
| | Acreage (has.) | Error (CV%) | Acreage (has.) | Error (CV%) | |
| León | 6853.2 | 24.15 | 6834.5 | 16.20 | 2.3 |
| Palencia | 88602.0 | 7.36 | 90535.3 | 3.33 | 4.7 |
| Valladolid | 128209.5 | 5.57 | 119707.4 | 2.66 | 5.1 |
| Zamora | 12324.2 | 17.71 | 10948.2 | 8.16 | 6.4 |
| TOTAL AREA | 235989.1 | 4.37 | 228028.5 | 1.98 | 5.2 |

Figure 2-6. Acreage estimates for barley in 4 provinces in the region of Castilla y Leon (2020), obtained without (ESYRCE columns) and with (ESYRCE+EO columns) EO data

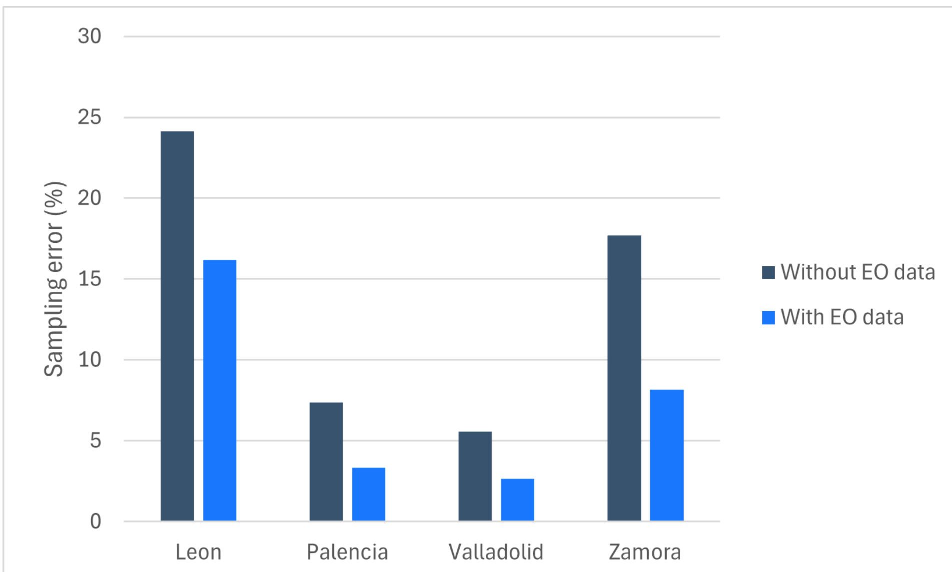


Figure 2-7. Reduction of the sampling error thanks to EO data for acreage estimates at provincial level – the case of barley in Castilla y Leon (2020)

The positive impact of EO data has also been proven in the estimation of barley acreage statistics at the level of the municipalities. For these administrative units, it is not possible to obtain statistics using only ground data: the very low number of samples would induce very low accuracy of the statistics. But using EO data, acreage estimates can be obtained with a sampling error remaining reasonable (i.e. less than 20%). This is shown in Figure 2-8.

| Municipality | | Acreage | |
|---|---|---|---|
| | | Has. | Error (CV%) |
| 49020 | Belver de los Montes | 212.96 | 29.1 |
| 49043 | Castroverde | 2914.22 | 8.0 |
| 49156 | Pinilla de Toro | 963.30 | 10.0 |
| 49168 | Quintanilla del Monte | 466.65 | 20.3 |
| 49219 | Toro | 615.91 | 14.0 |
| 49235 | Vezdemarbán | 1358.22 | 12.6 |
| 49250 | Villalpando | 560.05 | 39.1 |
| 49252 | Villamayor de Campos | 1056.23 | 11.1 |
| 49260 | Villanueva del Campo | 784.03 | 13.2 |
| 49263 | Villar de Fallaves | 844.16 | 11.0 |
| 49267 | Villardondiego | 516.40 | 11.5 |
| 49270 | Villavendimio | 656.07 | 10.4 |
| Total Zamora | | 10948.2 | 8.16 |

Figure 2-8. Acreage estimates for barley in the municipalities of the Zamora province (2020), obtained thanks to the integration of EO data

As for the yield, the yield estimation was also applied at provincial level, with conclusive results as it can be shown in Table 2-2 (presented and discussed in more details in the Validation Report).

Table 2-2. Comparison of ESYRCE Yield estimation given by Province with the Sen4Stat RS yield estimation model applied on all the barley fields of the survey.

| | ESYRCE | | Sen4Stat RS Model | |
|---|---|---|---|---|
| | N | Yield [kg/ha] | N | Yield [kg/ha] |
| Àvila | 151 | 4250.2 | 330 | 4297.8 |
| Burgos | 446 | 4852.4 | 2530 | 4678.7 |
| Leòn | 52 | 3792.7 | 276 | 4077.1 |
| Palencia | 304 | 4585.6 | 1068 | 4541.2 |
| Salamanca | 122 | 4204.3 | 279 | 4193.0 |
| Segovia | 294 | 4169.5 | 775 | 4327.8 |
| Soria | 275 | 3617.5 | 662 | 3611.5 |
| Valladolid | 460 | 4588.2 | 1556 | 4676.4 |
| Zamora | 206 | 4600.0 | 624 | 4462.6 |
| Castilla Y Leòn | 2310 | 4437.2 | 8100 | 4483.0 |

## 2.4 Timeliness

Three main applications were considered in this use case. They are reviewed one by one in the below paragraphs.

The first one aimed at getting crop acreage forecasts at the mid-season and crop yield forecast one month before the harvest.

Getting crop acreage forecasts necessitates the availability of the survey data (ESYRCE) and a crop type map. The project can only address the availability of the crop type map and we can say that it is possible to have this map at the mid-season (if ESYRCE is available) or from the mid-season as soon as ESYRCE is available. Indeed, in a separate project, it was already demonstrated that the accuracy of the mid-season crop type map was quite good and reached the plateau for most of the crops (except the summer crops like sunflower). This is shown in Figure 2-9, coming from the Sentinels for Common Agriculture (Sen4CAP) project.

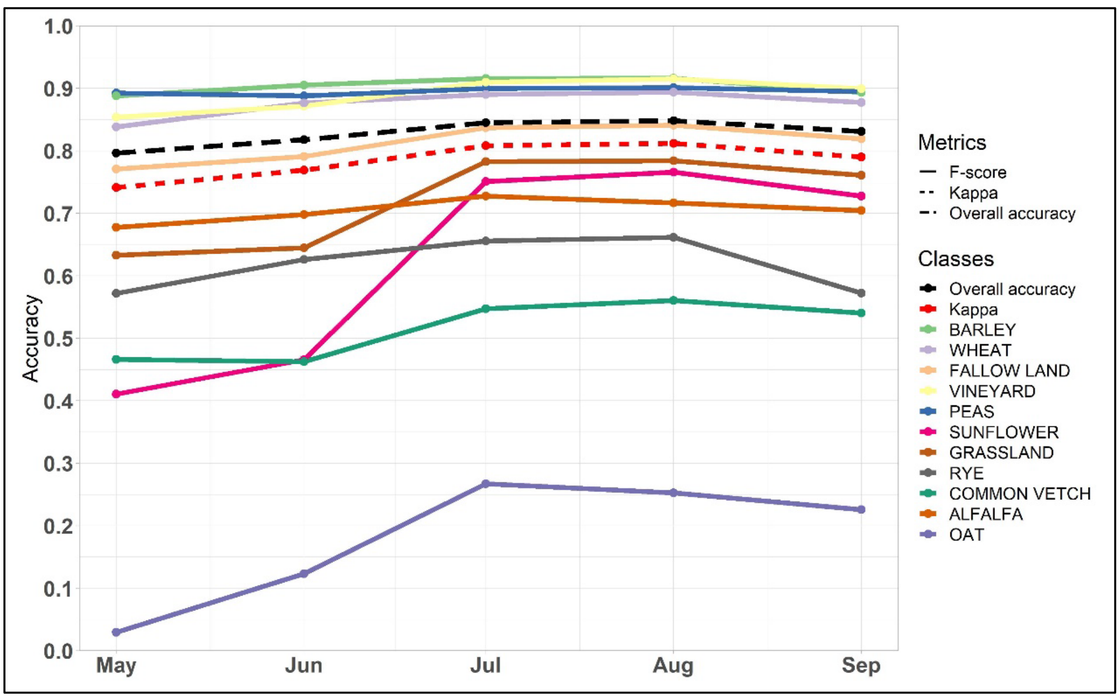


Figure 2-9. Evolution of the accuracy of the crop type classes (with the F-Score metrics) along the season in Castilla y Leon (2018 – from the Sen4CAP project)

As for the yield (and thus production) estimates, it is based on metrics at 3 key phenological stages, which are the emergence, the maximum of vegetation and the senescence. As soon as these moments are over, the estimates can be produced. Because senescence often takes place before the harvest, it means that estimates are possible to obtain before that moment.

## 2.5 Sampling design use case

A map distinguishing between the irrigated and non-irrigated crops was produced at national scale. It has been delivered to the Spanish NSO and the analysis is under way. Their feedback would be reported in an updated version of this report during the project extension.

## 2.6 Feedback from NSO

During the project, we had three official workshops with the Spanish NSO (February 2021, March 2021, May 2022, October 2023) and multiple iterations by email. Each of these iterations were the occasion to refine our mutual understanding: UCLouvain to better understand the NSO's expectations and the NSO to better understand what is feasible or not with EO.

The feedback from the different use cases were quite positive. The interest was very high for the cost-efficiency use case. Additional questions came about the possibility to reduce even more the survey cost by using EO data to interpret samples. Indeed, not all the ESYRCE samples are visited on the field each year because of the economic cost. Samples of specific land cover which are not expected to experience change between years are either photo-interpreted or re-used from the year before.

High interest was also expressed about the irrigation map. The map was shared with the NSO and as already stated, their analysis is in progress.

During the extension, the cost-efficiency use case will demonstrated at national scale. We might also consider the additional request to test the use of EO data to support the samples interpretation. Finally, based on the feedback about the irrigation map, additional activities might also come up for this use case.

# 3 Senegal

## 3.1 NSO's requirements

Senegalese NSO's expectations in terms of EO data can be summarized as follows:

1) *Improving design-based* crop acreage estimators by *optimizing* the integration of EO and ground data in the calculation of crop acreage statistics *to reduce the standard error while providing unbiased estimates* (therefore reducing the coefficient of variation without increasing the number of samples);
2) *Improving design-based* crop production estimators by *optimizing* the integration of EO and ground data in the calculation of crop production statistics *to reduce the standard error of the estimate while not increasing the number of samples;*
3) *Improving design-based* crop acreage and crop production estimators by *optimizing* the integration of EO and ground data in the calculation of crop acreage statistics *to reduce the survey costs (without increasing the standard error);*
4) Improving the sampling design, with the aim of *improving/optimizing the spatial allocation of the samples.*

Improving design-based estimators (points 1 and 2 above) is a comprehensive objective as it implies improving both the estimation procedure and the sample design.

Following the use cases introduced in the previous section for Spain, the statistical methods that can answer Senegalese NSO's expectations are cost-efficiency and optimization of the sample design.

The NSO also expressed their interest to use EO data to check the reliability of the survey data and reduce the uncertainty associated with this data. The statistical framework presented in section 2.1 assumes that these survey data are unbiased and reliable but in practice, this might not be totally the case. An additional use case will therefore concern the data collection protocol.

## 3.2 Data collection protocol use case

The Annual Agricultural Survey (AAS) in senegal is a list frame survey over all country, with a master frame that is the same since 2013. In 2013, an agriculture census took place, which allowed listing all active farmers (identifying new ones and removing the ones who stopped since the previous census). In parallel, a mapping exercise aimed at listing the active farmers by village, resulting in a map of agricultural households by village.

2000 holdings are selected thorough a stratified sampling from the 526.000 holdings in Senegal, which corresponds to ~ 0.4%. These 2000 holdings are spread in all the Senegalese departments, in direct ratio to the size of these departments. The same holdings are visited during 2 consecutive years and then, a new sample of holdings is drawn (2015-2016, 2017-2018, 2019-2020, 2021-2022-2023, 2024-205).

In each holding, farmers are interviewed, and GPS measurements are done in all fields belonging to the household. GPS coordinates are recorded at the parcel-level in the form of **points** and the **parcel area is measured (but outlines are not recorded)**. For the main season crops (not off-season crops), information is collected about crop type, crop area and production. Surveys are conducted annually, during the second half of the season (i.e. starting in August) and in any case, before the harvest. The surveys are carried out using a decentralized approach, through regional offices.

The use case about data collection protocol is an iterative use case, where we learn from one year, propose improvements and then test these improvements, learn again and make new propositions.

### 3.2.1 Iteration 1: 2020-2022

The first iteration took place in 2020, based on the AAS 2018. Using the AAS 2018, it was not possible to map crop types accurately (only cropland). The main issues identified were:

- The use of points to record field localizations: the Garmin GPS is used to calculate the area of the fields, but the GPS tracks are not recorded. We could therefore rely only on points and not on polygons, which decrease the performance of remote sensing images classification;
- Following the protocol, a GPS point is taken in each field. Unfortunately, it is often not located inside it but instead at the plot boundaries. It is therefore difficult to make a clear link between the point coordinates and a specific crop type;
- Minor crops were not enough represented in the survey to allow a performant discrimination.

The 2020 activities concluded that there was a clear potential for Earth Observation data at the condition that it can build on suitable ground data for algorithms calibration. The discussions showed that the AAS protocol requires only little adjustments to make the collected data suitable for this EO data integration. The main aspect to improve was the use of polygons instead of points.

It was therefore decided to conduct a specific exercise in 2021, in the department of Nioro du Rip. Between August and November 2021, a dedicated field data campaign was implemented to complement the official AAS, with the funding of FAO. Based on the Garmin GPS and a tablet, the pilot survey aimed at ensuring that field polygons are systematically recorded to obtain data compatible with remote sensing. The use of CSEntry as main software for AAS is subject to an agreement between the NSO and the FAO and was not easily modifiable for field data collection (e.g. addition of polygon recording and new questions for integration into remote sensing). The Open Data Kit Collect (ODK) tool installed on an Android tablet was therefore chosen as additional tool for collecting information about the heterogeneity of plots, the presence or not of mixed crops, the crop cutting for yield information and for ensuring the recording of parcels delineation.

The defined pilot protocol discussed with the NSO consisted in tracing the plots on one hand with the Garmin GPS and on the other hand with the tablet via the ODK Collect application. The data were sent daily to a server accessible by both protagonists.

Five teams worked on the field and simultaneously, the NSO and UCLouvain did a quick quality control of the data each day in order to adjust the survey in real time if necessary. For example, it was difficult for some investigators to properly delimit the fields at the beginning of the campaign. It was necessary to redefine the way points were taken. The data were to be used to produce a crop type map and a yield prediction map, both disaggregated by municipality.

Two comparative analyses were conducted to determine the quality of the area measurement data: one based on the entire dataset in the Nioro du Rip department during the pilot project, and a second based on discussions with the NSO.

#### 3.2.1.1 Comparative analysis on all the surfaces collected in the Nioro department

The data collected in the field by the tablet in manual mode are compared to the Garmin GPS data on the same surfaces. In general, the use of two separate devices during the agricultural survey was complicated for the investigators who stopped the protocol quickly. Some 231 polygons have been collected by ODK collect, and after cleaning, 199 polygons remain to be compared.

To statistically quantify the deviations of the tablet dataset from the GPS reference, the following indicators are used:

- The average absolute error, representing the absolute difference between the predicted value ($x_i$) and the actual values ($y_i$), and averaging it across the dataset (n values):

$$MAE = \frac{1}{n} + \sum_{i=1}^{n} |x_i - y_i|$$

- The Root Mean Squared Error/Deviation (RMSE) measures the average magnitude of the errors and is concerned with the deviations from the actual value:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}$$

- The Relative Root Mean Square Error, the dimensionless form of the RMSE:

$$RRMSE = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(y_i)^2}}$$

The MAE gives an idea of the quality of the tablet dataset. By removing aberrant polygons, the mean absolute error on the entire consistent dataset is 0.3 ha. More than half of the plots studied are less than one hectare in size. This error of 0.3 ha represents almost 1/3 of the area error for half of the dataset. The accuracy of the model seems therefore quite low.

The root mean square error is 0.43. This index provides an indication of the variability of the prediction quality. The larger the difference between the MAE and the RMSE, the larger the variance of the individual errors in the dataset. The lower is the RMSE value, the better is the accuracy of the dataset. Since the RMSE gives a relatively high weight to large errors, a value of 0.43 means that the dataset has large errors. In particular, the value indicates that the variance of the model is relatively high, reaching nearly 45% of the mean of the observations. The RMSE depends on the size of the dataset. To compensate for this effect, it is interesting to normalize it. The RRMSE, dimensionless form of the RMSE, expresses the relative average error in percentage, and its value is 2.5%.

Figure 3-1 relates the regression line of the tablet dataset to the reference line (GPS Garmin). The value of the slope is 0.99, close to the value of the reference slope. The position of the regression line and the bias of 0.27 indicates that the tablet data generally underestimates the reference data. This bias is confirmed by a visual analysis of the data.

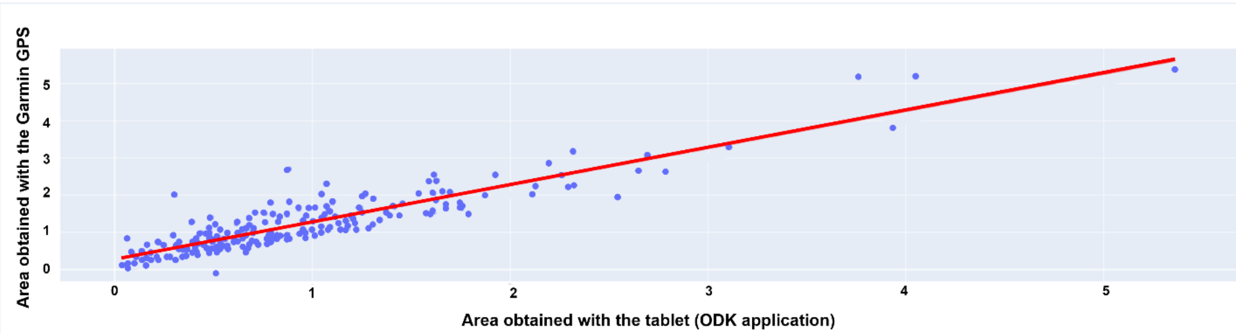


Figure 3-1. Predicted areas by the ODK app on an Android tab (blue points), the red line is the regression line related to the predicted dataset and the green line, the reference related to the GPS Garmin dataset

Several measurements deviate significantly from the reference, which can be explained visually. Indeed, ODK Collect's measurements in manual mode shows big weaknesses (Figure 3-2). The erroneous surfaces are caused by:

- error in encoding the points by the enumerator;
- waiting time: the tablet's GPS takes time to stabilize its position with acceptable accuracy. A shortened waiting time can then lead to an error of up to 30 meters from the position of the investigator;
- recording error by the tablet: it is possible that the device or application does not record a point correctly or does not record it at all.

Statistically and visually, the tablet dataset is less accurate than the one recorded with the GPS and has a bias that underestimates the baseline value.

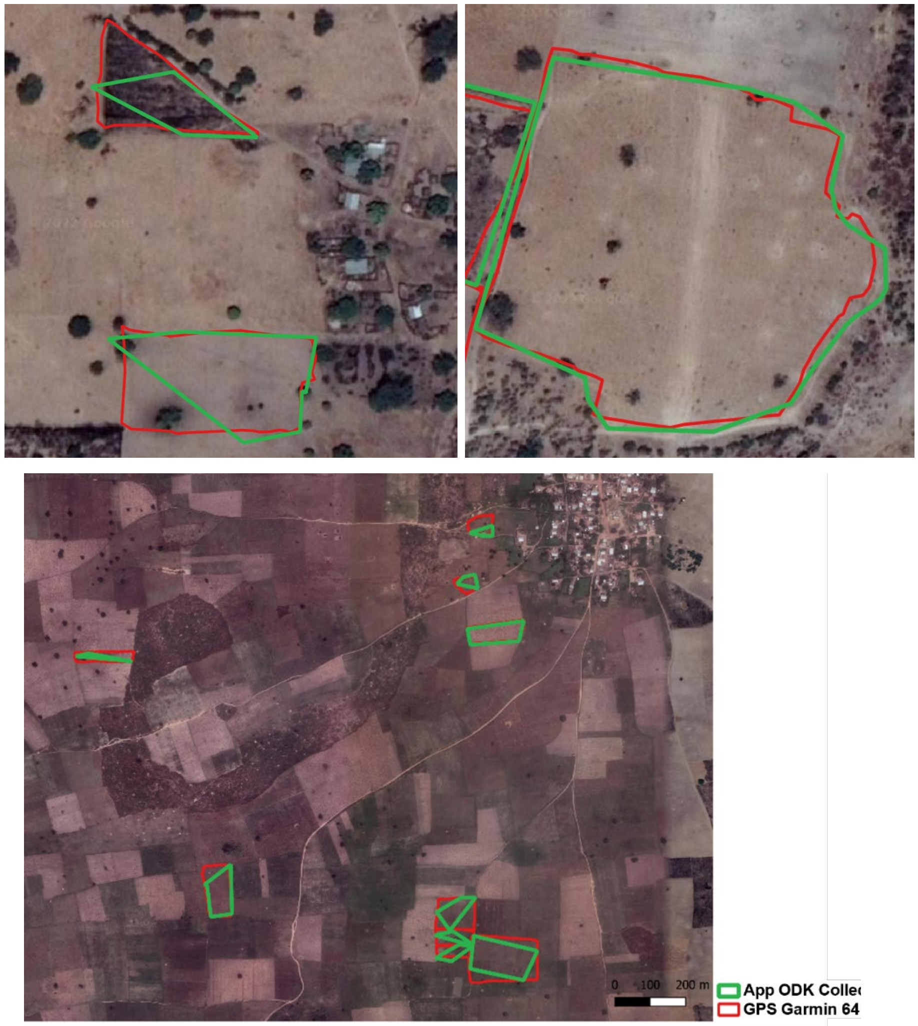


Figure 3-2. Parcel's polygons from Garmin GPS in red and from the tablet's GPS in green.

### 3.2.1.2 Comparative analysis on small surfaces collected in Dakar

The outcomes of the analysis presented in the previous section were presented during a workshop with the NSO in Dakar (April 2022). During the discussions, it was reported that the accuracy of the areas collected by the ODK Collect application through an Android tablet lost its accuracy in the case of small estimated areas.

To provide clear answers to these statements, it was proposed at the Dakar workshop to conduct tests of small area surveys with both types of tools (Garmin 64 GPS and ODK Collect through an Android tablet or a smartphone). The accuracy of the surface data collected by the tablet would then be compared to the reference data taken with a Garmin 64 GPS.

The tests carried out by DAPSA in an open space away from the offices are supported in the field by the laboratory of applied remote sensing of the Cheikh Anta Diop University of Dakar. The small area selected for this study of 2,500 m2 is equivalent to ¼ ha and is often the smallest area of rainy season crops except for marginal speculations (cowpea, bissap, etc.).

The protocol provides for a comparison of ODK and Garmin measurements with actual perimeter and area measurements taken in the field. Three main geometric shapes should be traced using quilts and a 200-meter tape: a square, a rectangle of any shape and an irregular shape with at least 5 points of inflection.

For each geometric shape, three measurements are taken simultaneously: one with manual ODK (i.e. each time the enumerator changes the direction of walking, it has to record a GPS point), a second with automatic ODK (i.e. no manual encoding of changes in direction) and a third with Garmin GPS on each of the plot shapes. In total, nine measurements are made by each device. Table 3-1shows the areas measured in m² by the different tools.

Table 3-1. Areas measured by tools used in AAS for the small plot measurement test project

| Parcel's number | Area measured by the Garmin GPS (= reference) | Area measured by ODK in automated mode (m²) | Area measured by ODK in manual mode (m²) |
|---|---|---|---|
| 1 | 2272.7 | 2364.0 | 2370.0 |
| 2 | 4458.7 | 4359.9 | 2294.0 |
| 3 | 2411.2 | 2407.4 | 4160.7 |
| 4 | 3600.7 | 3499.9 | 3453.0 |
| 5 | 2894.4 | 2939.8 | 3491.0 |
| 6 | 4152.4 | 4322.5 | 3701.5 |
| 7 | 7877.3 | 8206.3 | 6199.8 |
| 8 | 14025.5 | 13510.5 | 8508.6 |
| 9 | 19687.4 | 19156.13 | 13946.54 |
| **Total area** | **61380.4** | **48125.1** | **48125.1** |

Figure 3-3 shows the different plots measured, in red the delimitation of the plots by the Garmin 64 GPS, in green, those measured with the ODK application on the tablet in an automatic way and finally in orange, those measured with ODK in a manual way.

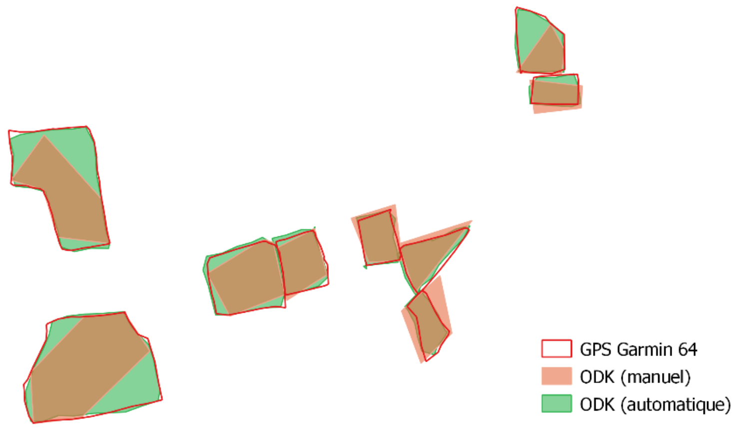


Figure 3-3. Parcel's delineations by the Garmin GPS in red, by the tablet with ODK Collect in manual mode in orange and in automated mode in green.

As there were no ground-truth measurement provided (perimeter and area), the GPS Garmin 64 is taken as reference. A visual analysis shows that the surfaces taken by ODK Collect automatically seem to match the reference, unlike the surfaces taken by ODK Collect in manual mode.

Several hypotheses, as mentioned previously, can justify the notable differences between the Garmin accuracy and the "manual" ODK accuracy (Figure 3-4).

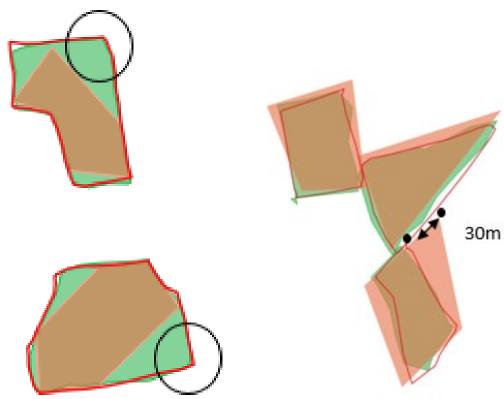


Figure 3-4. Noticeable differences between Garmin GPS data and tablet data via manual ODK Collect

The total area measurement in automatic mode on the tablet underestimates the reference by 614 m², giving an accuracy of 98.99%, while the manual mode underestimates it by 13255 m², giving an accuracy of 78.4%. As a reminder, accuracy is the closeness distance of a set of data to the real value.

The measurements are plotted on two graphs in Figure 3-5 and Figure 3-6. The trendline of the automatic ODK measurements (Figure 3-5, orange) shows a small deviation from the line of the reference measurements (green) and thus a good overall accuracy, while for the ODK measurements in manual mode, the trendline deviates strongly from the reference (Figure 3-6).

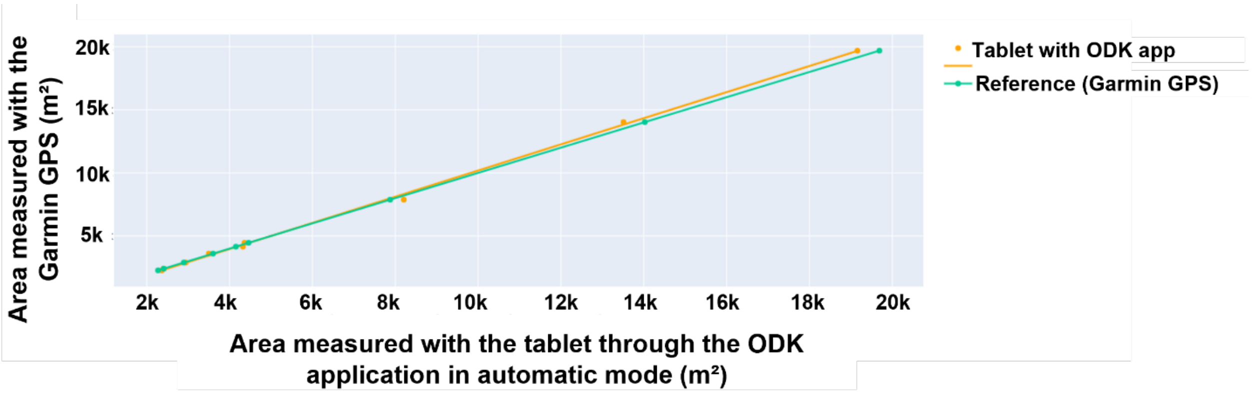


Figure 3-5. Comparison of the measurements made between the tablet via ODK Collect in automatic mode and the reference (Garmin 64 GPS)

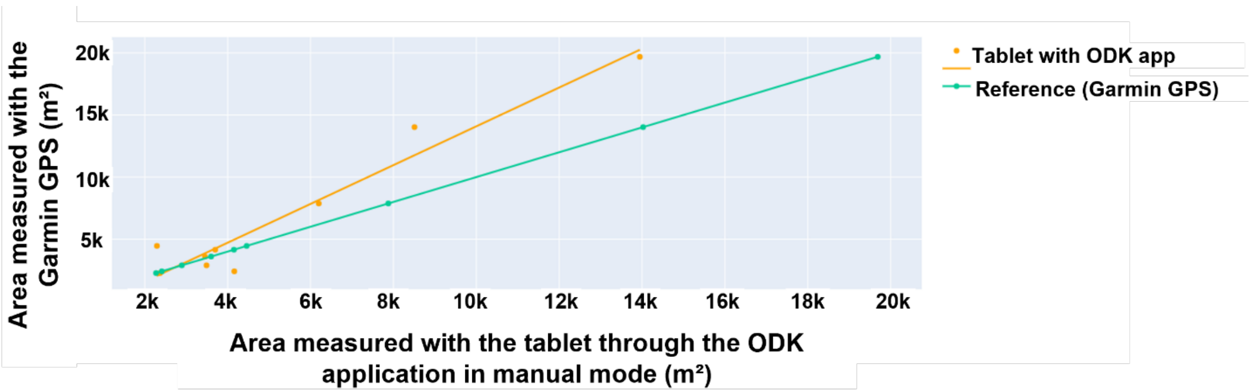


Figure 3-6. Comparison of the measurements made between the tablet via ODK Collect in manual mode and the reference (Garmin 64 GPS)

Regarding the two measurements taken by the tablet, it is clear that the automatic point measurement provides a result that is much more accurate and closer to the reference result than the manual measurement. It avoids recurrent errors due to bad encoding of points, time to fix the GPS position of the tablet or problems in recording the points by the tool. It is therefore not advisable to use the tablet for surface acquisition in manual mode with the ODK Collect application.

In order to fully determine the effectiveness of the tablet, the accuracy of the position of the points taken by the tablet's automated mode have also be analyzed.

The location accuracy is obtained by calculating the minimum distance between 5350 points taken randomly along the lines of the Garmin surfaces and the edges of the tablet polygons. This results in an average distance of 1.96m with a standard deviation of 1.52m. Over 96% of the dataset has an accuracy error of less than 5m.

To get rid of the shape of the plot (e.g., an elongated plot could have a higher bias than a compact plot), a bias calculation is performed based on the area of the data in excess of the reference and in deficit to the reference (Figure 3-7) divided by the GPS lengths appended to the corresponding areas. It results in an average bias of 2.75m per perimeter unit for the excess areas and 2.54m of missing areas.

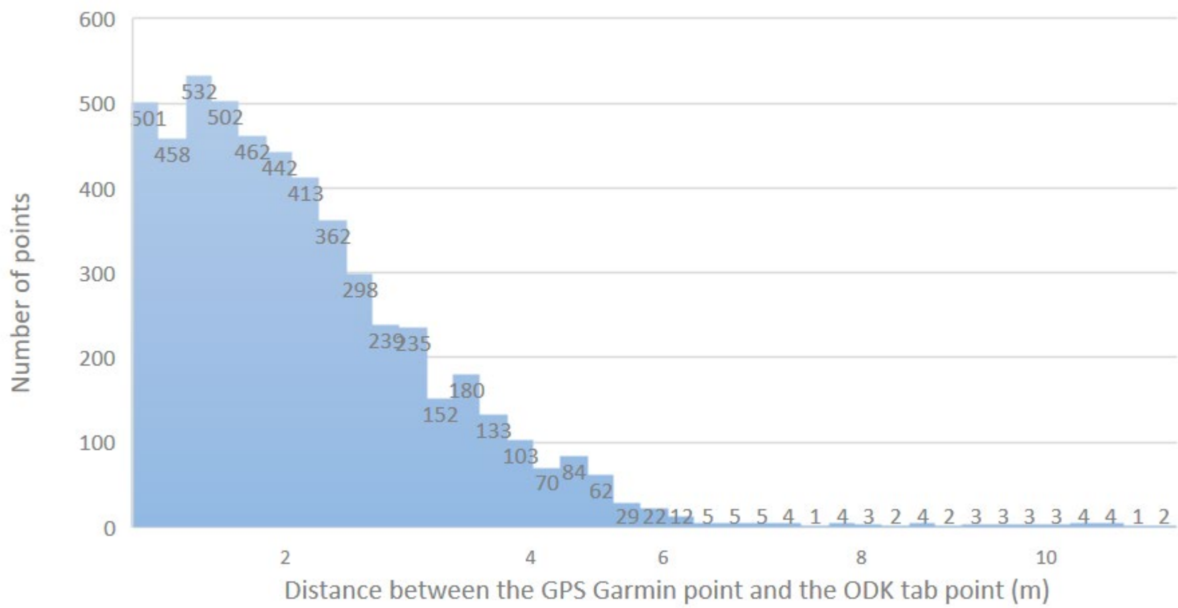


Figure 3-7. Distance in meters between 5350 points taken randomly along the Garmin GPS delineations and the polygon delineations drawn by the ODK application on a tablet

➢ **Yield data**

A key feature of the Nioro pilot survey was the measurement of yields of selected crops through a crop cut square to produce a yield prediction map. The results of the crop cut and the yield per crop arrived a few months after the field survey (Figure 3-8). Some problems detected in the data made it difficult to use the data:

- the data were aggregated by household without plot information. It was therefore impossible to relate yield predictions to measured yields;
- the names used to link the different data were sometimes incorrect (naming and reporting problems).

Since the results were not usable for remote sensing, one solution could have been to use the exact location of the crop cutting square requested in the ODK Collect survey. However, it appears that the location point was not taken on the crop cut but at the edges of the plots. This observation was confirmed during the workshop in Dakar in April 2022, where it was agreed that measurement squares were not properly georeferenced.

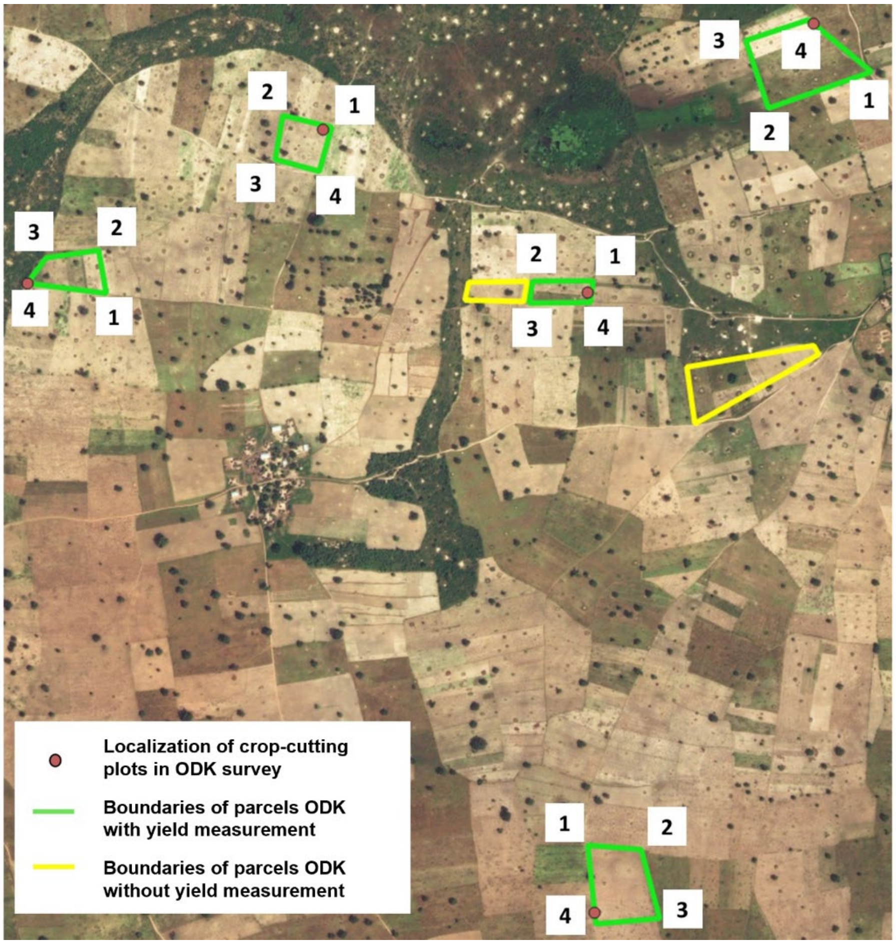


Figure 3-8. Location of crop cutting squares

- **Recommendations for next iteration on AAS protocol improvement**

In the case of the pilot test in Nioro, a few new aspects in the protocol were aimed at improving the field campaign and making it compatible with remote sensing, such as adding information to the main survey (crop heterogeneity, presence or absence of mixed crops, taking a yield square and associated position). After discussion, it was found that qualitative heterogeneity was not very helpful in determining yield, but that it would be interesting to investigate further. Crop mixing does not seem to be important in Senegal, except at the edges of plots, on areas too small to be detected by satellite images (bissap or cowpea). As for the yield squares, the test in the Nioro was not conclusive because the data did not allow for a link between the field information and the satellite image. However, it is important to continue to investigate and improve the use of yield squares in order to produce a useful yield map for the country.

Studies on the accuracy of the tablet in terms of points and areas calculated do not support the claim that the tablet can replace the Garmin GPS at this time. The GPS was taken as a reference, which means that all the claims about the errors of the tablet are based on the claim that the GPS is accurate. However, it is not simple to qualify the exact error on the positioning of the Garmin GPS. Since this error is not known, it is easy to believe that the Garmin GPS is better than the tablet. In fact, the tablet error may result from a combination of errors from both devices. A more thorough and detailed protocol could quantify the actual errors for each device. However, there are no plans to include one in this project.

### 3.2.2 Iteration 2: 2023-2024

The 2023-2024 activities aimed at extending the previous activities to 6 departments (Nioro du Rip, Koungheul, Dagana, Mbacke, Kolda and Tambacounda).

A Letter of Agreement was signed between the NSO and the UCLouvain to define the updates in the AAS protocol to be implemented over these 6 districts. The most important points of this LoA are:

- The new protocol is planned to be implemented in two phases: a first phase beginning in August when EAA interviewers collect data in the field, and a second phase when they return to the field in October and November.
- ***Phase 1:***
    - In accordance with the EAA's initial protocol, the surveyors circle each plot with a Garmin 64 GPS to obtain the surface area. **The outline of each plot is recorded on a server;**
    - In addition to the initial protocol, a GPS point is taken with the table and the SurveySolutions software, in the middle of each plot in order to avoid naming problems between the EAA data on the tablet and the plot outlines taken with the Garmin 64 GPS. In cases where it would be difficult to reach the middle of the plot, the point is taken at least 20 m from the edge of the plot. For small plots less than 40 m wide, the point is taken in the middle.
    - Still following the initial protocol, a point is taken in the middle of each crop cutting plot with the Garmin 64 GPS ;
    - During this phase, the NSO ensures that the protocol is respected and makes the enumerators aware of the importance of collecting GPS coordinate;
    - NSO also ensures quality control of the geometry of the plots taken by GPS and supplied by the enumerators. UCLouvain supports NSO in its quality control approach.
- ***Phase 2:***
    - In each crop cutting plot, the required number of plants/panicles is taken at random and the harvest of the square is assessed on the basis of the weight of the grains in dry matter.

This report focuses on the **crop type data** because the **crop yield data are not yet available** at the moment of writing this report.

Collected polygons for which data could be retrieved (not all geometries could be linked to collected data) are displayed in Figure 3-9.

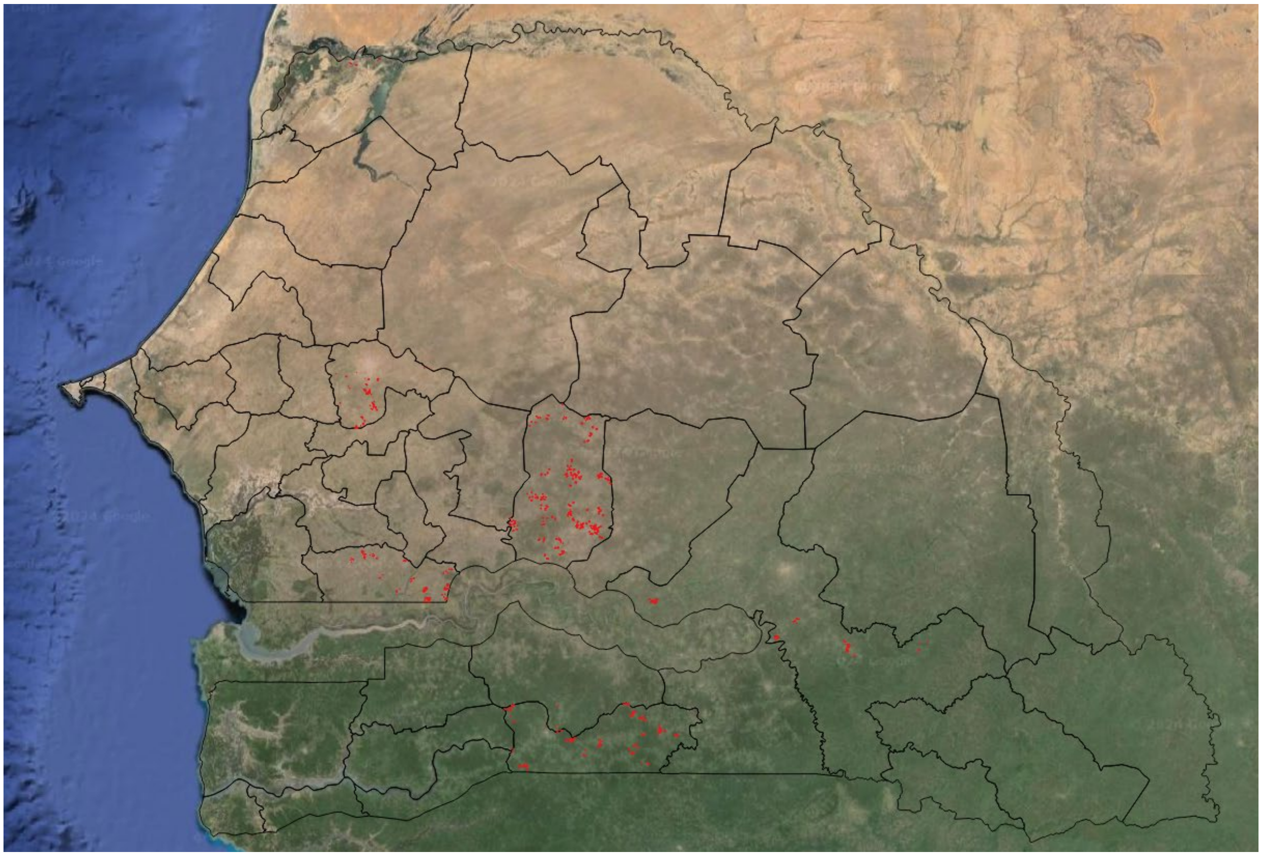

Figure 3-9 Polygons (highlighted in red) collected during the 2023 field campaign.

Table 3-2 provides a quantitative insight about this field campaign, providing the collected data (in SurveySolution and in the form of GPX) and the remaining data after the quality control. The quality control includes the following steps: keeping only IDs existing both in SurveySolution and in GPX, keeping only valid crop class, removing duplicated geometries.

Table 3-2. Number of data collected and after quality control

| | Number of samples in AAS | Number of GPX | Number of samples after quality control |
|---|---|---|---|
| **Total** | 12827 | 3925 | 2215 |
| **Dagana** | | 10 | 8 |
| **Kolda** | | 785 | 430 |
| **Kongheul** | | 1231 | 598 |
| **Mbacke** | | 1264 | 678 |
| **Nioro** | | 462 | 334 |
| **Tambacounda** | | 173 | 167 |

A significant loss of data comes from the absence of link between the SurveySolution database and the GPX data. The shared dataset contained detailed information on each field visited during their in-situ campaign, alongside corresponding GPX tracks delineating the boundaries of crop polygons (i.e., parcels). The GPX files represent the precise geospatial extents of these fields, while the attribute data contains identifiers and other descriptive information about the crops within each field.

Initial analysis of this dataset revealed inconsistencies between the crop identifiers (IDs) listed in the attribute file and those associated with the GPX tracks. These discrepancies included instances of duplicate entries and mismatches between the crop IDs in the attribute table and those embedded in the GPX metadata. Such inconsistencies posed challenges for data integration, potentially leading to data loss or misinterpretation of field boundaries, ultimately reducing the reliability and usability of the dataset.

To resolve this, first, duplicate crop identifiers in both the attribute data and the GPX metadata were identified and systematically removed. Once the dataset was cleaned of duplicates, a spatial join method was applied, wherein the crop center point coordinates, provided in the attribute data, were matched to the GPX polygons based on spatial proximity. This process used the longitude and latitude of each crop's center point to determine the nearest polygon representing the field boundary. The join operation ensured that each crop record was spatially linked to its correct field.

Although this distance-based matching method introduced some uncertainty, particularly in areas where fields were densely packed or irregular in shape, it was the most effective solution available. By ensuring that each crop identifier was paired with its corresponding field boundary, the dataset was rendered more reliable.

The distribution of the crops within the collected data is shown in Figure 3-10.

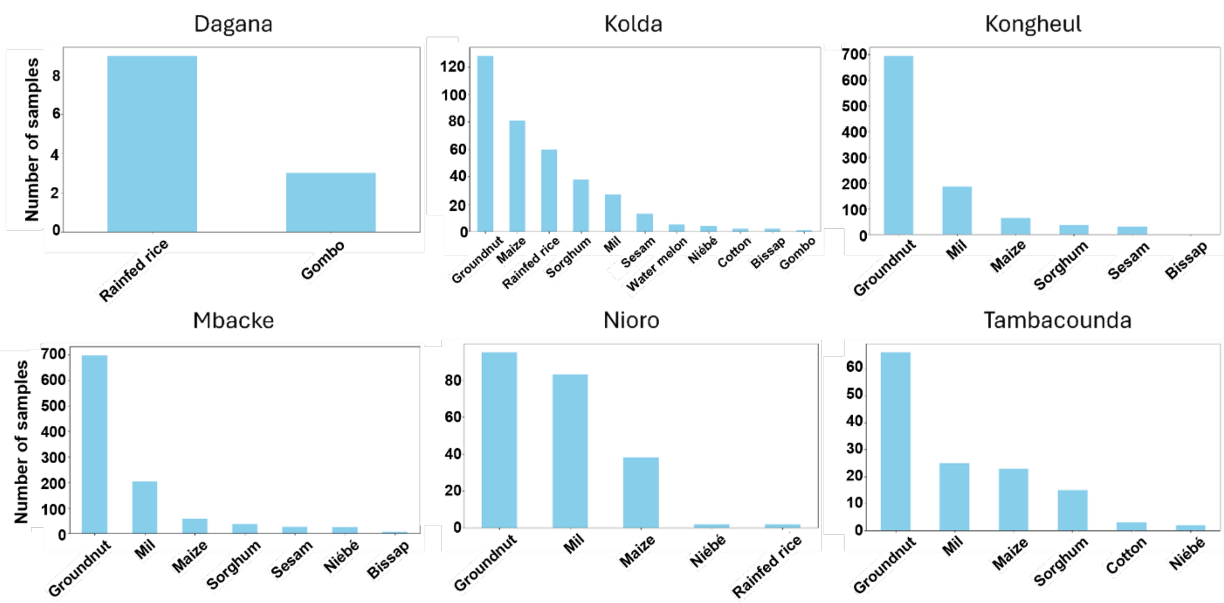


Figure 3-10. Crop distribution in the dataset collected in each department during the 2023 field campaign

## 3.3 Cost-efficiency use case

The implementation of the cost-efficiency use case in Senegal is more complex than in Spain, because the sampling design is not the same. While it is an area frame in Spain, this is a complex (4 stages) list frame of households in Senegal (Figure 3-11).

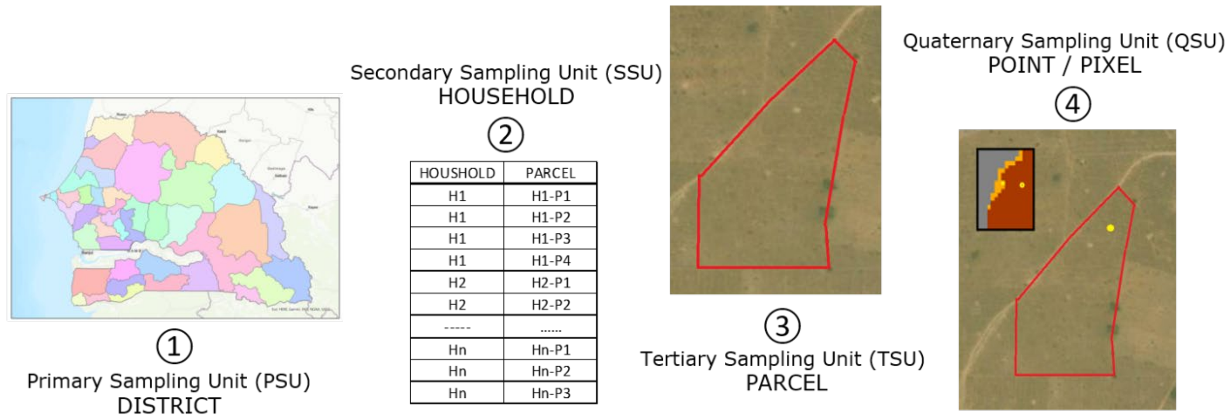


Figure 3-11. 4-stage list frame sample design in Senegal.

As a result, the integration of remote sensing and ground data cannot be done using linear models, but multinomial logit models are needed. These multinomial models deal with the uncertainties and generate probabilities that a pixel of a given class in the map is actually this given crop on the ground.

The cost-efficiency use case has been demonstrated in terms of crop acreage estimates using the crop type map generated over the department of Nioro. No demonstration was done for the crop production estimates since no yield estimation has been achieved so far.

Figure 3-12 presents the crop acreage estimates in the department of Nioro, for the two main crops which are millet and groundnut while Figure 3-13 shows the efficiency of using the crop type map to support this estimation of crop acreages.

| Crop type | Acreage (hectare) | Uncertainty | | | | |
|---|---|---|---|---|---|---|
| | | Standard error | Coefficient of variation (%) | Limits of 95% confidence interval | | |
| | | | | Lower | Upper | Amplitude |
| Millet | 89215 | 3661.103 | 4.11 | 81978.88 | 96330.4 | 14351.52 |
| Groundnut | 78815 | 2923.94 | 3.71 | 73089.15 | 84550.98 | 11461.82 |

Figure 3-12. Crop acreage estimates using EO and ground data in Nioro (Senegal, 2021)

| Crop type | Standard errors of proportion estimators | | Relative efficiency of Classified RS data |
|---|---|---|---|
| | Using only ground data | Using ground & RS data | |
| Millet | 3.37 | 1.90 | 3.13 |
| Groundnut | 3.34 | 1.52 | 4.80 |

Figure 3-13. Efficiency of using the crop type map for crop acreage estimation in Nioro (Senegal, 2021)

While it was not requested, the spatial disaggregation use case was also tested and successfully demonstrated: as shown in Figure 3-14, acreage estimates are available at the "arrondissement" levels with a reasonable error (expressed as the coefficient of variation).

| Arrondissement | Millet | | Groundnut | |
|---|---|---|---|---|
| | Acreage (has.) | Error (CV%) | Acreage (has.) | Error (CV%) |
| Medina Sabakh | 20067.21 | 8.6 | 19765.36 | 7.3 |
| Paoskoto | 38316.02 | 5.3 | 35018.93 | 4.0 |
| Wack Ngouna | 30831.77 | 11.9 | 24030.71 | 10.7 |
| Total Nioro | 89215,00 | 4.11 | 78815,00 | 3.71 |

Figure 3-14. Crop acreage estimates at the district (arrondissement) level in Nioro (Senegal, 2021)

## 3.4 Sampling design use case

The sampling design use case expected from the NSO requires an active participation of the NSO. At the time of writing this report, the collaboration between the project and the NSO has not been possible on that topic.

## 3.5 Feedback from NSO

During the project, we had two official workshops with the Senegalese NSO (March 2021, April 2022) and multiple iterations by email. Each of these iterations were the occasion to refine our mutual understanding: UCLouvain to better understand the NSO's expectations and the NSO to better understand what is feasible or not with EO.

The feedback from the different use cases were quite positive but the main focus was clearly the update of the AAS protocol. The updated aimed at increasing the reliability of the collected data and their compatibility with EO data. The process of changes implementation is a process that takes time. It requires testing a modification, demonstrating its added-value, discussing the feasibility of its implementation and monitoring its implementation. Many interactions between the project and the NSO concerned this use case, including technical sessions on quality control procedures. The biggest success is clearly the implementation of the proposed adjustments in the protocol itself.

The next step for this use case is clearly the evaluation of the protocol in terms of yield data and the proposition of adjustments if needed. This will be started with the evaluation of the 2023 data, which will be documented in the report of the project extension.

Interest was also high for the cost-efficiency use case (also waiting for crop production demonstration). Much less for the sampling design application, which will probably be cancelled for the next steps.