D3.2 CATALOGUE ON AUXILIARY DATA AND AVAILABLE REPOSITORIES TO BE INCORPORATED

Project: Monitoring of Environmental Practices for Sustainable Agriculture Supported by Earth Observation

Acronym: ENVISION

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 869366.

Document Information

Grant Agreement Number	869366	**Acronym**	ENVISION
Full Title	Monitoring of Environmental Practices for Sustainable Agriculture Supported by Earth Observation		
Start Date	1st September 2020	**Duration**	36 months
Project URL	https://envision-h2020.eu/		
Deliverable	D3.2 Catalogue on auxiliary data and available repositories to be incorporated		
Work Package	WP3 – Earth Observation data products		
Date of Delivery	**Contractual** M9	**Actual**	M9
Nature	Report	**Dissemination Level**	Public
Lead Beneficiary	DRAXIS		
Responsible Author	Panagiota Syropoulou (DRAXIS), George Efthimiou (DRAXIS)		
Contributions from	NOA, EV ILVO, AgroApps		

Document History

Version	Issue Date	Stage	Description	Contributor
D0.1	21/5/2021	Draft	Draft for review	DRAXIS
D0.2	26/5/2021	Draft	Review feedback	NOA
F1.0	31/5/2021	Final	Final version	DRAXIS

Disclaimer

This document and its content reflect only the author's view, therefore the EASME is not responsible for any use that may be made of the information it contains!

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

CONTENT

LIST OF FIGURES

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

LIST OF TABLES

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Executive Summary

The aim of this deliverable is to provide a catalogue on the available auxiliary data and repositories that will be incorporated (data format and kind) as part of the project's products and services.

This deliverable is based on the Deliverable 4.1 Architecture and Services Specifications report which defines the system architecture, the generic integration framework as well as provides the specifications (data, visualisation) of the services that will be developed during the project. Furthermore, this deliverable will be the basis for future deliverables, such as D3.3 Data products initial report, D3.5 Report on collected auxiliary data and D3.7 Data products final report.

Specifically, the sections of this deliverable are:

Section 1 – ENVISION Data Requirements: Contains the overview of the data that are needed to be collected in order to start developing the ENVISION services.

Section 2 – ENVISION Data Summary – 1st round: Contains an analysis of the data that have been collected from the service provides up-to-now (1st round) per business case.

Section 3 – Conclusions: Contains the main conclusions of this deliverable.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

1 ENVISION Data Requirements

ENVISION contributes to the achievement of Common Agricultural Policy's (CAP's) environmental objectives, offering the tools for the continuous, large-scale and uninterrupted monitoring of farm management activities with regards to sustainability. These tools reinforce the monitoring of environmental- and climate-friendly agricultural practices stemming from European Union (EU) policy, ensuring that the agricultural activities do not severely impact the climate and nature.

ENVISION fully exploits the wealth of data made available through GEOSS and Copernicus and its synergetic use with other data to develop data products, such as:

- Cultivated crop type maps
- Soil Organic Carbon
- Distinction of organic – conventional farming
- Grassland mowing/ ploughing
- Soil erosion

ENVISION makes use of heterogeneous types of available data (EO-based, in situ, open data, and historical on-field check data), as well as state-of-the-art technologies and methodologies (automatic pixel/ texture/ object-oriented change detection and classification methods, machine learning, data fusion, multi-source and multi-temporal data management) for providing a fully-automated and scalable toolbox of services, built in close collaboration with its future customers.

ENVISION develops services that best fit the needs of Paying Agencies (PAs) and Organic Certification Bodies (OCBs), helping them to master the complex processes of monitoring farmers' performance in relation to the environmental rules stemming from EU policy. These services will be tested and validated in an operational environment not only by the project business partners, but also by a group of Lighthouse Customers (LHCs) that will actively involve into the project.

The diagram below depicts the correlations among the different components/ aspects of the ENVISION project and how the various data sources will feed the ENVISION services.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

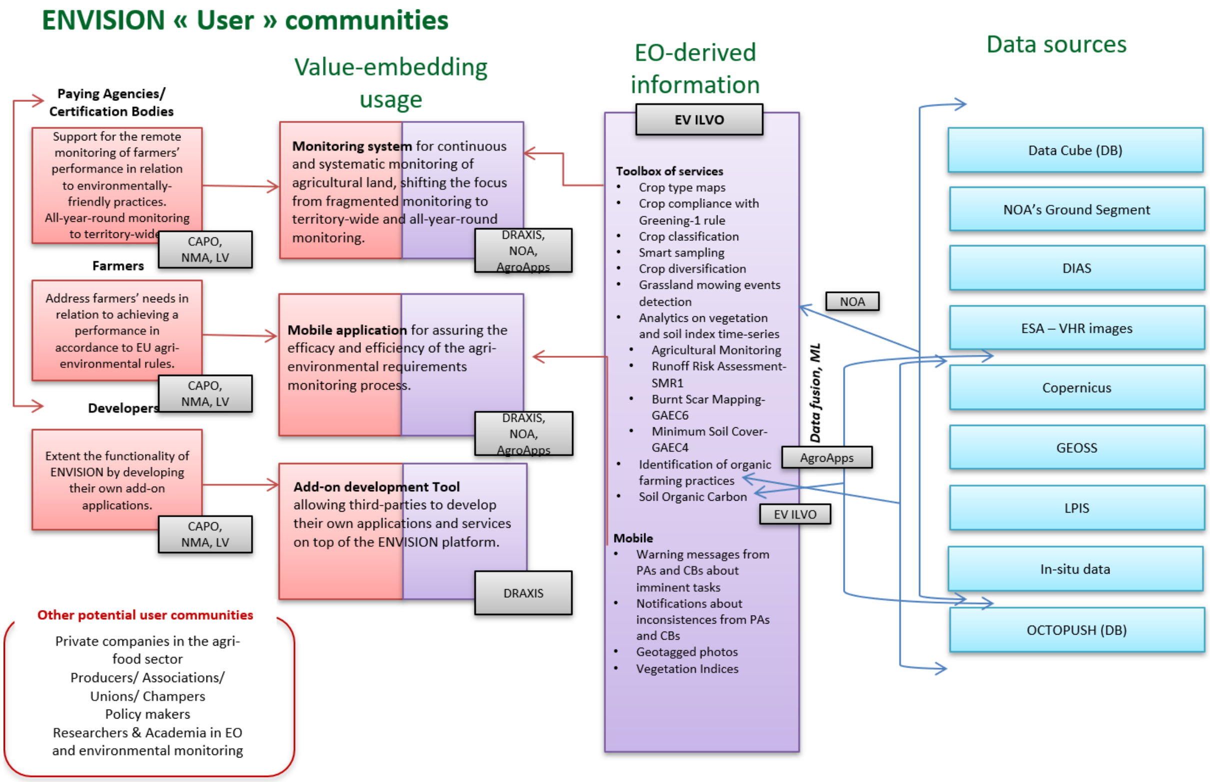


Figure 1: Diagram of ENVISION concept

Based on a preliminary analysis of the ENVISION services' development needs and requirements, the following lists were structured and circulated to the business cases partners. Specifically, these lists present the requirements from the service providers' side for what data should be collected in order to start developing the services.

The list for the conventional identification information is the following:

- Parcels: Vector polygons of parcels (.shp), at least 30,000 to 50,000 per paying agency, at least 3 years (i.e. 2017-2020), source = LPIS
- Parcels of agricultural grassland, at least 20,000 per paying agency, at least 3 years (i.e. 2017-2020), source=LPIS. These parcels will be used for grassland mowing detection.
- Declarations: The declarations for the abovementioned "parcels" dataset, at least 3 years of declarations. The declarations should be at the lowest level of crop description, i.e. the declared labels used for Greening 1 - wheat, barley, oats and not cereals, which is a higher-level description.
- The phenology of the principal crops to be investigated in each area of interest.
- Results of the spot inspections (at least the last year). OTSCs verifying the cultivated crop type are necessitated. Any other compliance decision on GAECs, SMRs and Greenings would be very helpful, as well.
- Actual yield of the provided fields
- Area of Interest (AOI)

The list for the distinction between organic vs convention farming practices is the following:

- Parcels vector data, polygon data, parcel geometry, parcel crop type, at least 30,000 to 50,000 per paying agency, at least 3 years (i.e. 2017-2020), source = LPIS

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

- For the same AOI we need this information for conventional and organic farming practices
- Declarations: The declarations for the abovementioned “parcels” dataset, at least 3 years of declarations
- Actual yield of the provided fields

Furthermore, an additional list was sent to the business cases partners, consisting of a number of layers. These layers will be presented on top of the ENVISION maps and the users will be able to manipulate the spatial objects in order to perform geographic operations and spatial analysis by using GIS tools. However, these data layers will be defined based on the need of the end-users and the availability of data.

The list of the data layers is the following:

- Base layer
- Very High-Resolution Images: 2 VHR Spot images, as received from JRC, for the years of inspection as mentioned above
- Orthophotos: Any orthophotos available
- Hydrographic networks - as detailed as possible (e.g. watershed delineation)
- RIvers- Hydro Note
- River flow rates, levels through time
- Abstraction locations and rates (spatial and temporal resolutions)
- Irrigation rates
- Rainfall data (station location, temporally resolved)
- Roads
- Watercourse maps
- Water surface
- Water flows
- 50m contours- lines/ points
- Reclassified map based on MDT 25m
- Nitrate Vulnerable Zones
- Botanical Heritage Sites
- Natura 2000
- Land use/ Land cover Maps
- ILOT and Sub-ILOT

2 ENVISION Data Summary – 1st round

This section provides an analysis of the data that have been collected from the service provides up-to-now (1st round) per business case.

The ENVISION project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 869366

2.1 The Business Case of Cyprus (Cyprus Agricultural Payments Organisation – CAPO)

This Business case will focus on employing ENVISION's services to monitor Cross-Compliance, Greening, and Rural Development Programmes' (RDP's) climate-environmental-requirements. Through the period of the business case implementation, the PA will evaluate which requirements can be effectively monitored using Earth Observation (EO) technology using the principle of lowest cost with the maximum effort.

Below is the analysis of the provided data:

Parcel records received refer to parcels of different crops, years and farming practices as follows:

- 2017: 322.335 (318.634 conventional, 3.701 organic)

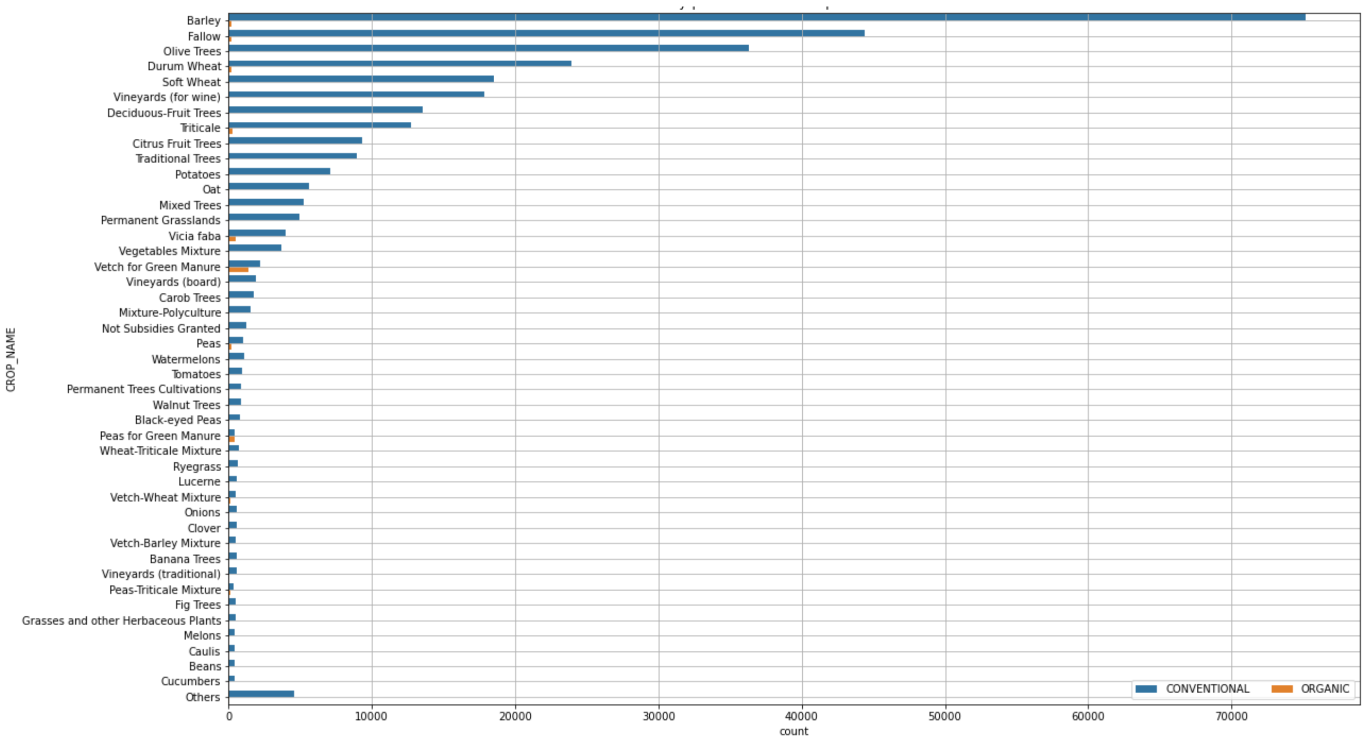


Figure 2: Cyprus Declarations Crops Count 2017 per crop type

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

- 2018: 326.789 (323.139 conventional, 3.650 organic)

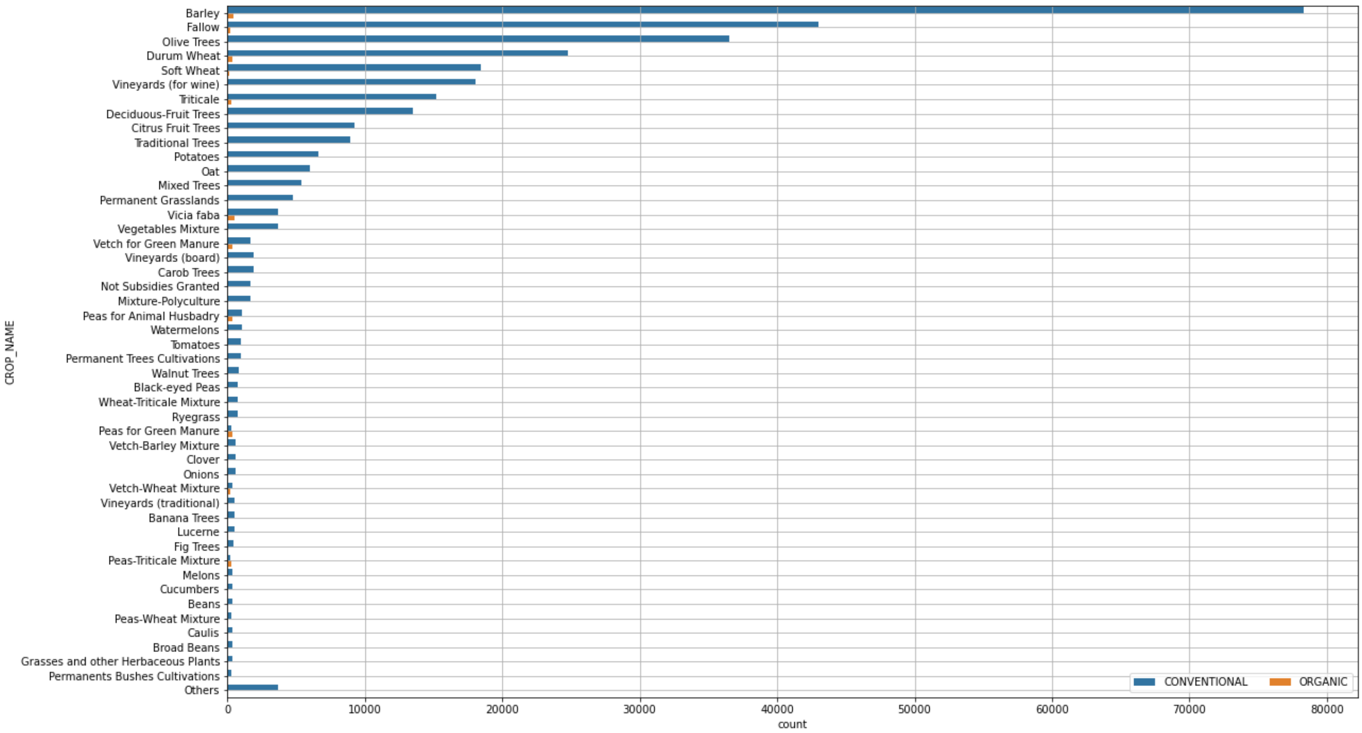


Figure 3: Cyprus Declarations Crops Count 2018 per crop type

- 2019: 327.543 (324.086 conventional, 3.457 organic)

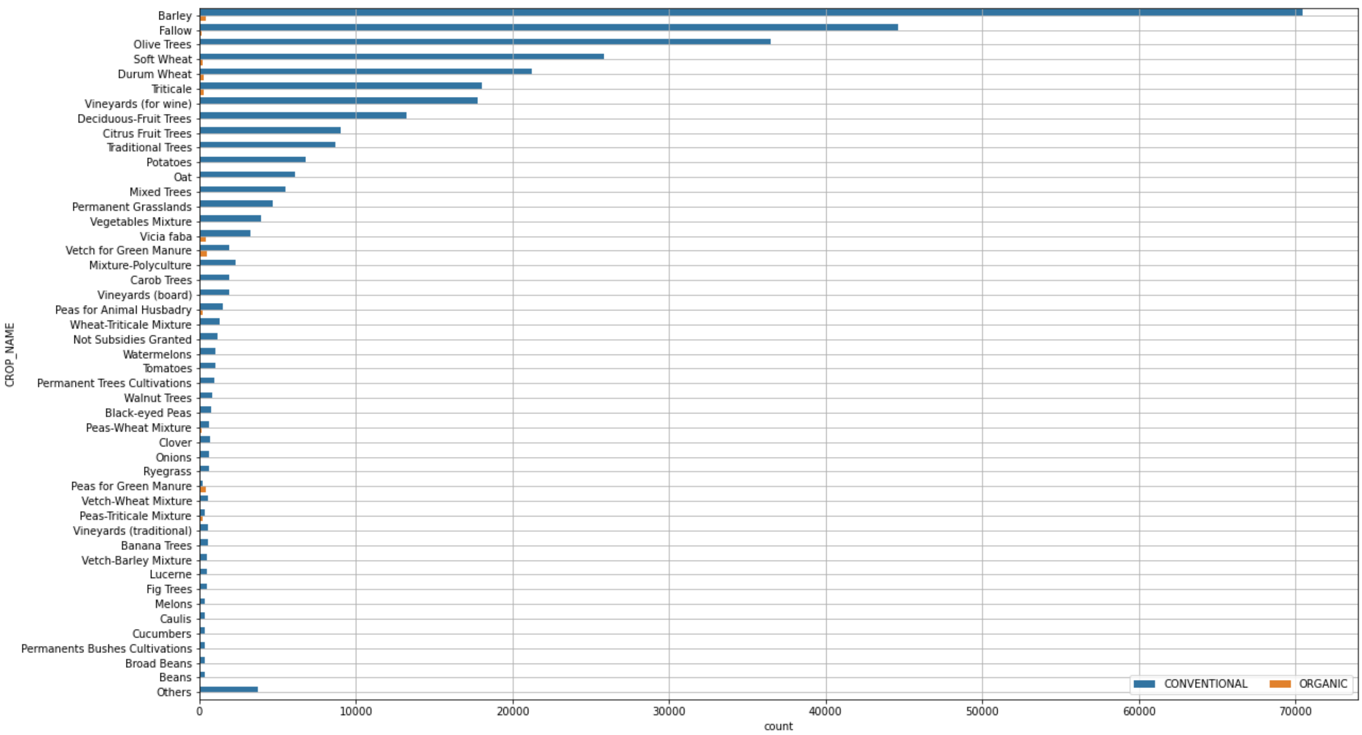


Figure 4: Cyprus Declarations Crops Count 2019 per crop type

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

- 2020: 330.659 (info available only for conventional)

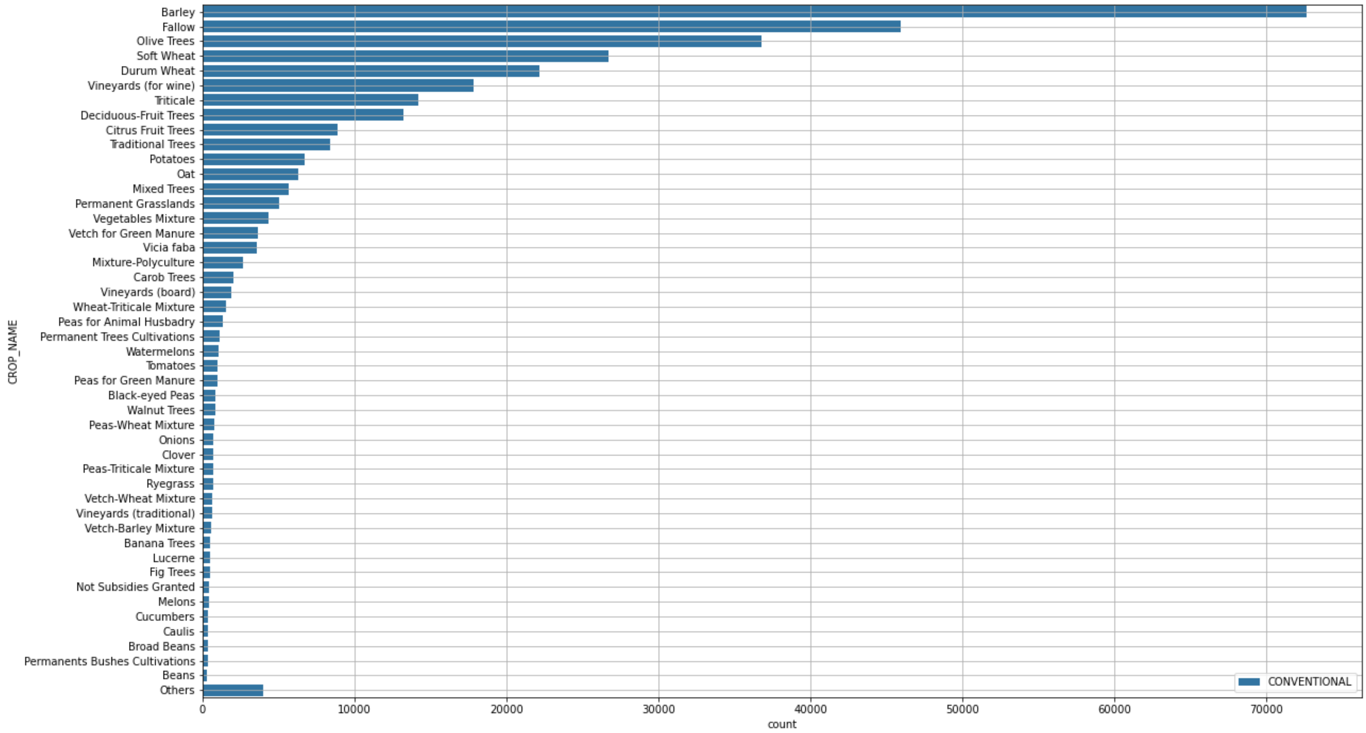


Figure 5: Cyprus Declarations Crops Count 2020 per crop type

Additionally, for the year 2019, we have several validated cases (through OTSC or Remote Sensing Assessment) that accounts for almost ~ 10% of the total dataset:

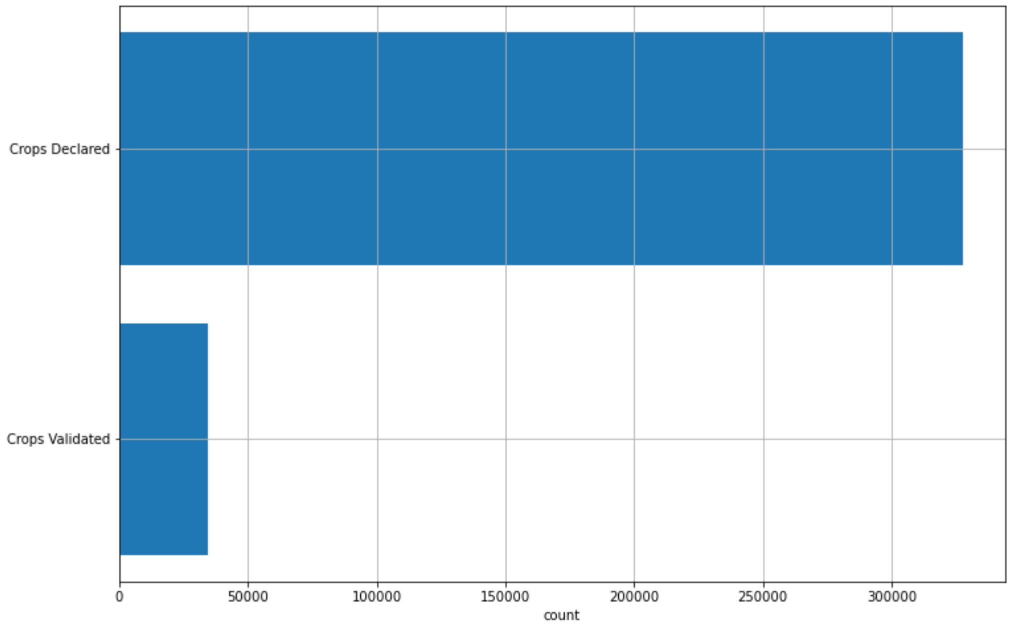


Figure 6: Cyprus Data for 2019

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Moreover, some auxiliary data in order to check GAEC 4, GAEC 6, SMR 1 and other cross-compliance requirements have been provided. The following plot depicts the number of respective cases:

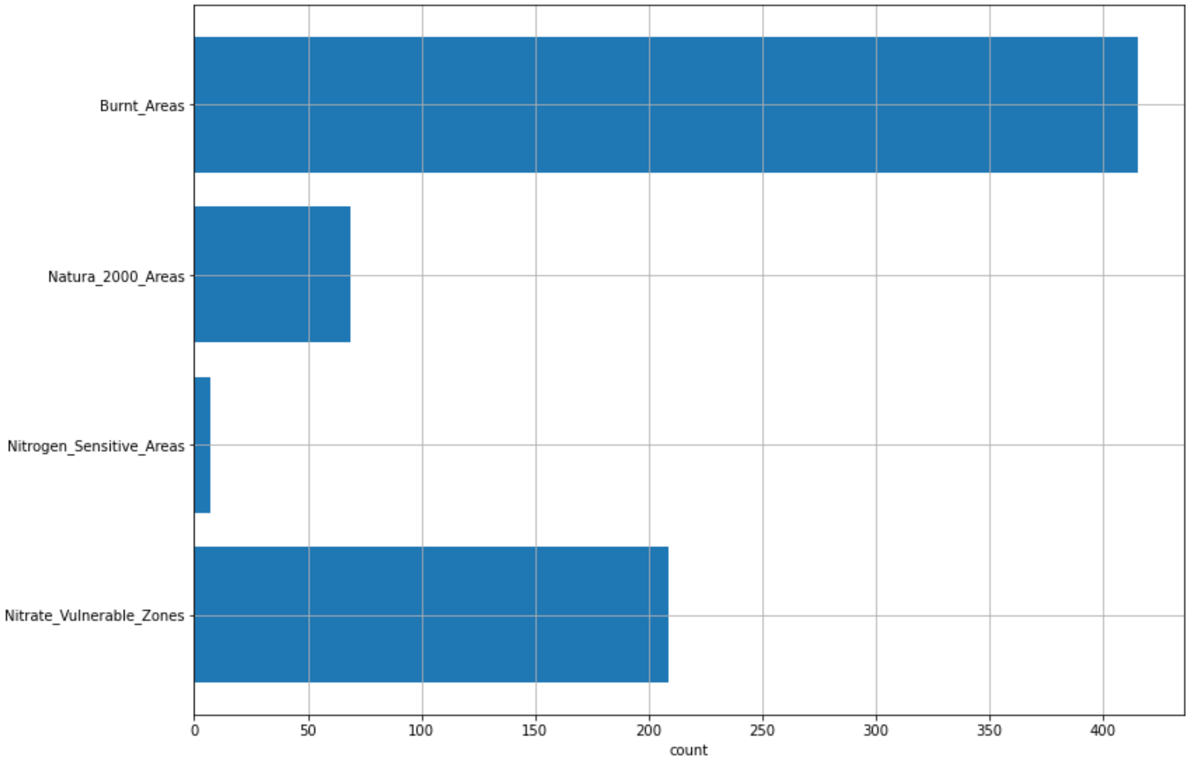


Figure 7: Cyprus Auxiliary Data

CreoDIAS will be used as the primary resource of retrieving Sentinel data, as according the Deliverable 3.1 Cost-benefit analysis, seems to be currently the best-fit solution for ENVISION in terms of budget and the offered services. Specifically, Sentinel-1 GRD, Sentinel-1 SLC and Sentinel-2 Level-2A data are going to be exploited. The Sentinel-2 Level-2A products are offered in the most of the cases as Bottom of Atmosphere (BOA) reflectance images derived from the associated Level-1C products. However, a subset of the Level-1C products has not been transformed to Level-2A. Thus, Level-1C products will be also used as an input for the generation of the missing Level-2A products.
Some of the issues that have been faced with the provided data are the following:

- Two or more crop types have been associated with the same parcel number
- Small size of a series of parcels
- Wrongly declared parcels cultivations in order to comply with local subsidy regulations
- Intense natural vegetation may confuse algorithms
- Several cultivations in the same parcel declared as one

Very High Resolution (VHR) data may be needed so to map or/and validate the existence of more complex structures inside the parcels.

2.2 The Business Case of Lithuania (National Paying Agency – NPA)

This Business case will focus on employing ENVISION's services to monitor Cross-Compliance, Greening, and Rural Development Programmes' (RDP's) climate-environmental-requirements. Through the period of the business case implementation the PA will evaluate which requirements can

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

be effectively monitored using Earth Observation (EO) technology using the principle of lowest cost with the maximum effort.

Below is the analysis of the provided data:

Parcel records received refer to parcels of different crops, years and farming practices as follows:

- 2018: 1.153.795 (1.129.566 conventional, 24.229 organic)

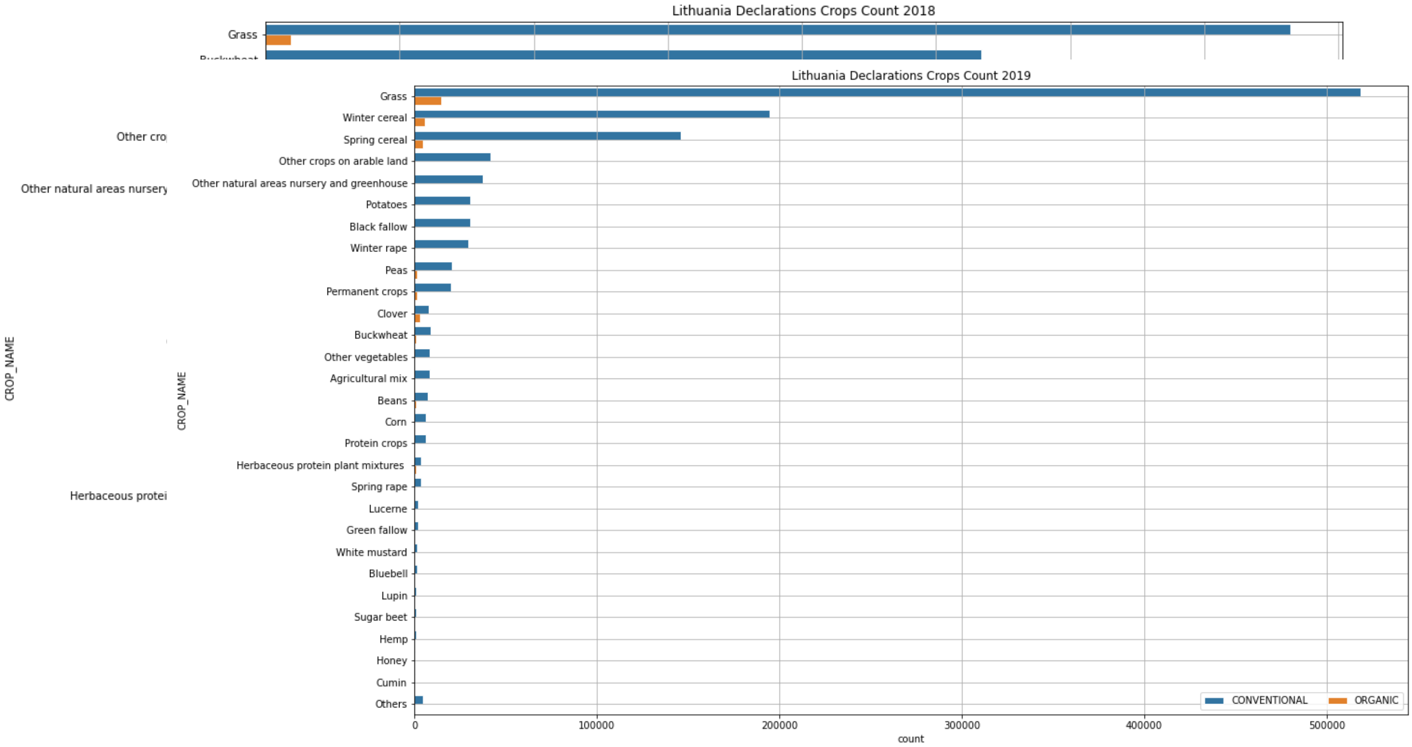
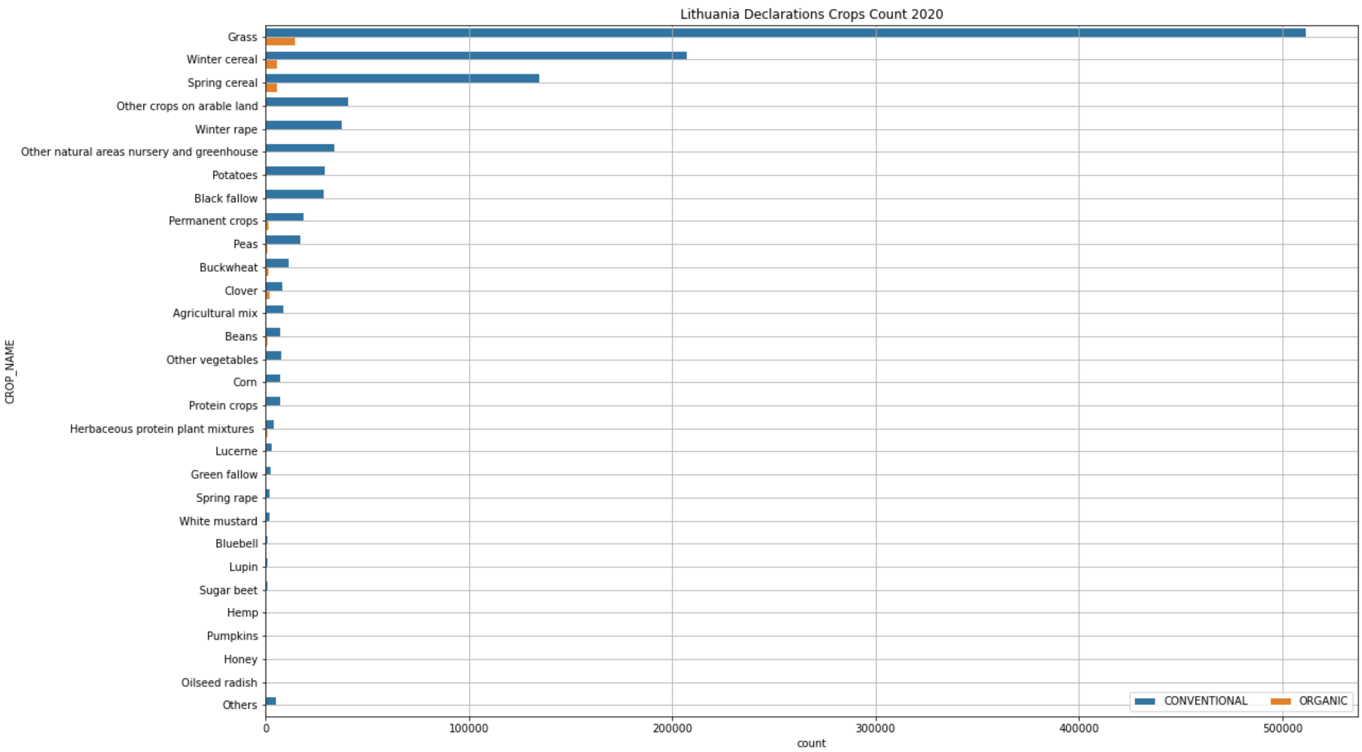


Figure 8: Lithuania Declarations Crops Count 2018 per crop type

- 2019: 1.185.411 (1.148.133 conventional, 37.278 organic)

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Figure 9: Lithuania Declarations Crops Count 2019 per crop type



- 2020: 1.178.899 (1.141.583 conventional, 37.316 organic)

Additionally, for the years of **2019** and **2020**, we have several validated cases (through OTSC or Remote Sensing Assessment) that accounts for almost ~ 10% of the total dataset.

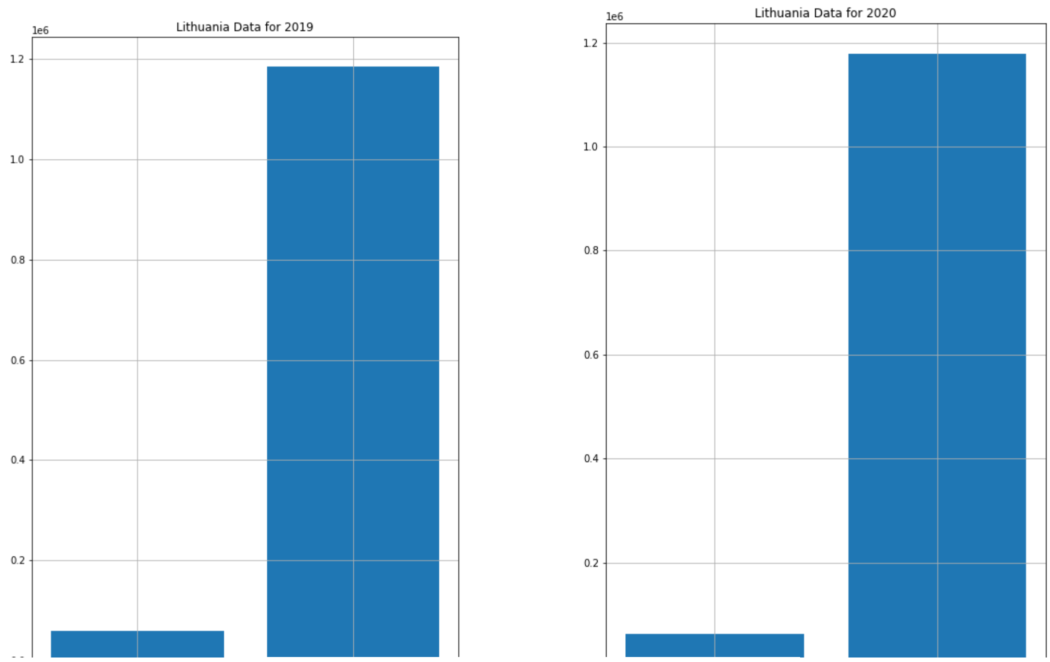


Figure 11: Lithuania Data for 2019 - 2020

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Moreover, data according mowing compliance is at our disposal for all the respective years mentioned: In addition, some auxiliary data in order to check GAEC 4, GAEC 6, SMR 1 and other cross-compliance

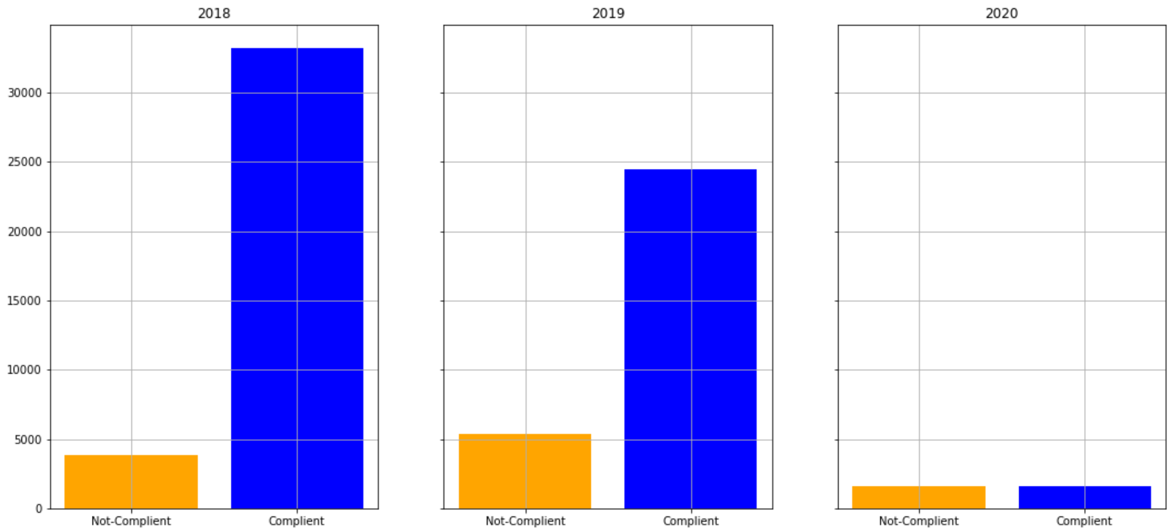


Figure 12: Data for mowing compliance for the years 2018 - 2020

requirements have been provided. The following plot depicts the number of respective cases:

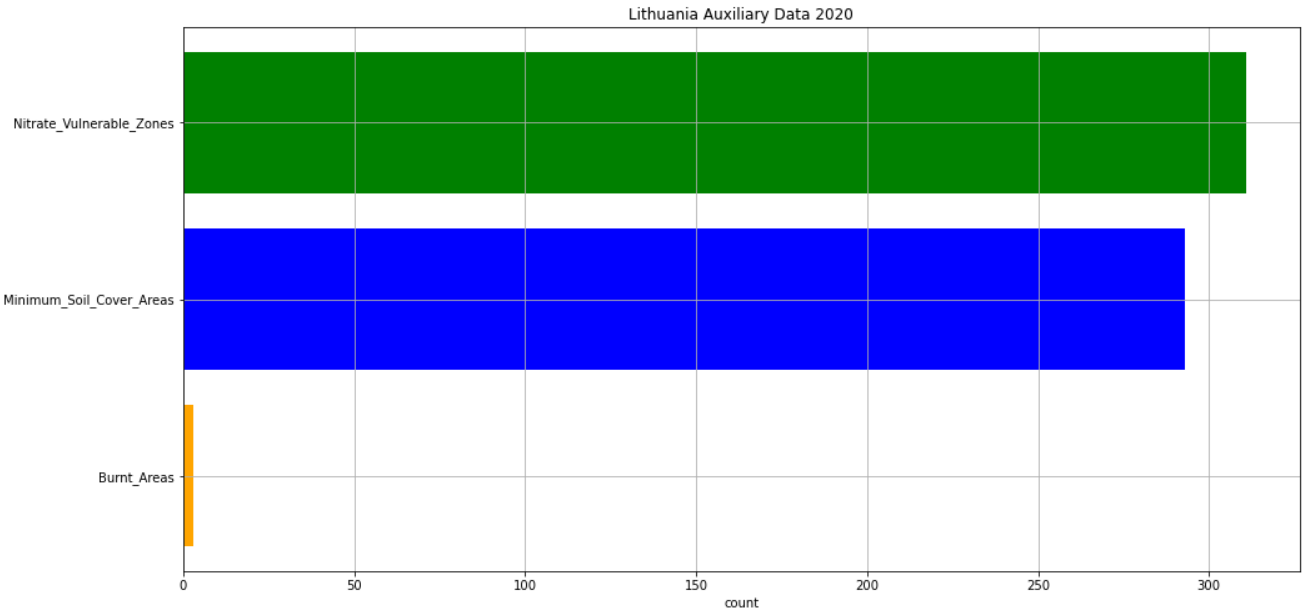


Figure 13: Lithuania Auxiliary Data 2020

NPA has also provided the hydrographic vectors including ten types of water bodies.
CreoDIAS will be used as the primary resource of retrieving Sentinel data. Specifically, Sentinel-1 GRD, Sentinel-1 SLC and Sentinel-2 Level-2A data is going to be exploited. The Sentinel-2 Level-2A products are offered in the most of the cases as Bottom of Atmosphere (BOA) reflectance images derived from the associated Level-1C products. However, a subset of the Level-1C products has not been

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

transformed to Level-2A. Thus, Level-1C products will be also used as an input for the generation of the missing Level-2A products.
The main issues that have been faced with the provided data are:

- High Sentinel data sparsity due to frequent inclement weather phenomena and clouds cover
- Slow processing of SLC Sentinel-1 data needed for the coherence generation

Very High Resolution (VHR) data may be needed so to map or/and validate the existence of more complex structures inside the parcels

2.3 The Business Case of Serbia (Doo Organic Control System Subotica – OCS)

This case will demonstrate how the uptake of EO technology can improve the overall monitoring of organic certification requirements such as farmland expansion, biodiversity, GHG emissions, water and soil.

Below is the analysis of the provided data:

Out of 4316 parcel records for which crop information has been received, 3703 were successfully imported to the database having all related information including the field of Geometry. Those 3703 parcel records refer to parcels of different crops, years and farming practices as follows:

- 2220 conventional parcels
- 1483 organic parcels

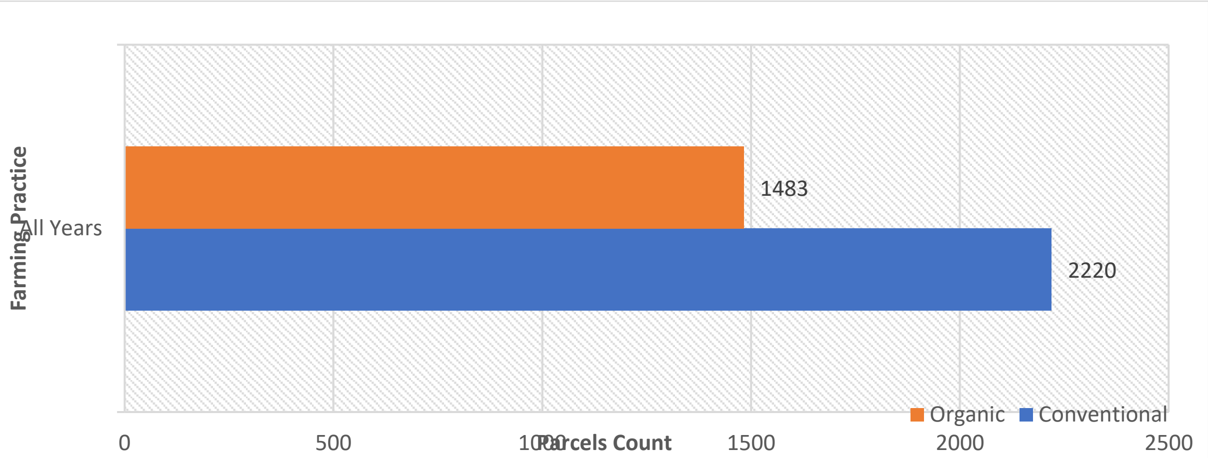


Figure 14: Total Organic and Conventional Parcels Count

- 2016: 566 (550 organic, 16 conventional)
- 2017: 280 (254 organic, 26 conventional)
- 2018: 857 (289 organic, 568 conventional)
- 2019: 1008 (229 organic, 779 conventional)
- 2020: 992 (161 organic, 831 conventional)

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

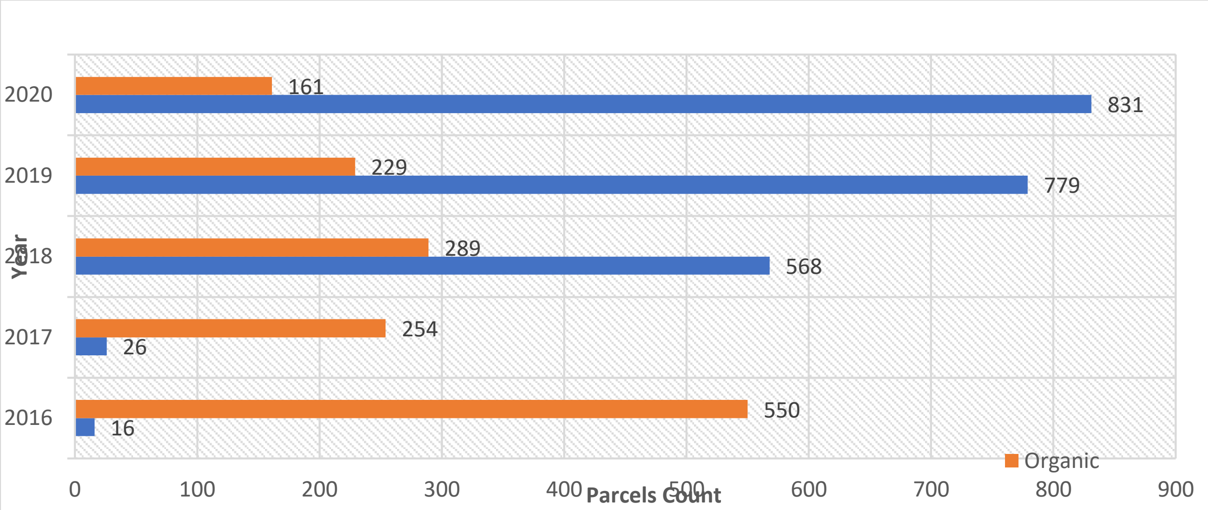


Figure 15: Organic and Conventional Parcels Count (2016-2020)

- 1164 wheat (538 organic, 626 conventional)
- 1130 corn (148 organic, 982 conventional)
- 423 Soybean (211 organic, 212 conventional)
- 940 Sunflower (541 organic, 399 conventional)
- 46 Barley (45 organic, 1 conventional)

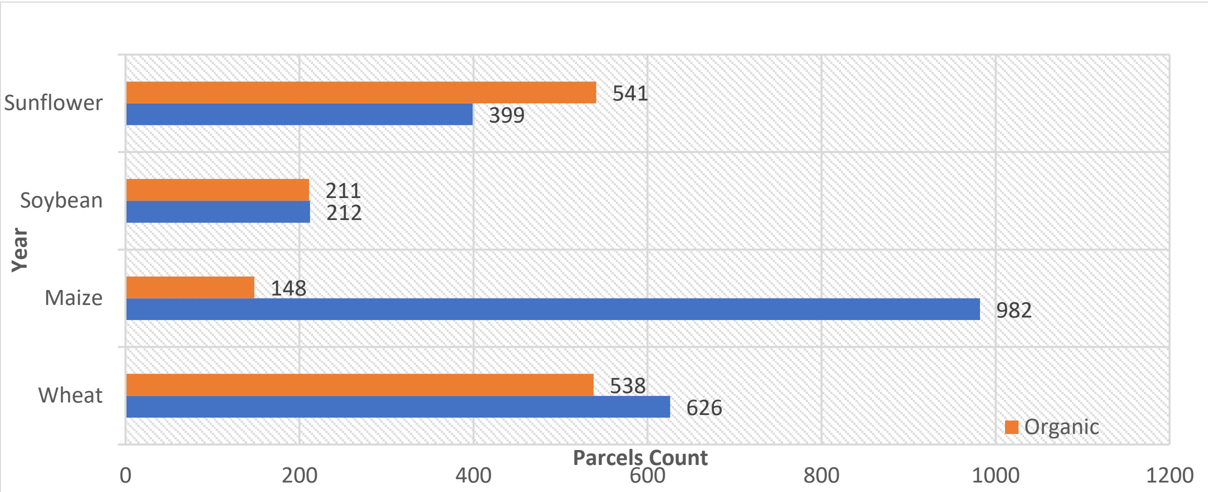


Figure 16: Serbia Declaration Crops Count 2016-2020

The data repository that is going to be used is the CreoDIAS platform. Atmospherically corrected Sentinel-2 Level-2A images are going to be retrieved and in cases where Level-2A images are not available, Level-1C would be retrieved and processed with sen2cor algorithm in order to become Level-2A products.

In order to achieve a fairly successful discrimination between Organic and Conventional crops, a sufficient number of representative pixels is required. Those pixels can be identified since they are totally located inside parcels of known crop characteristics. Since the pixel size is given (10m*10m), the size and the shape of the parcels should be sufficiently large, so that it totally contains pixels and consequently those pixels are representative of the crop type and practice. Consequently, there are two key-factors regarding the usefulness of the parcel data stemming both from the need to have

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

sufficient number of representative pixels; the size & shape of each parcel, the number of parcels available.

A. Parcel Geometry Characteristics
B. Number of parcels per classification category
C. Parcel dispersion & Relevance

A - Parcel Geometry Characteristics

The geometry characteristics analysis of the received parcels showed that

- In general, the average parcel size is small, meaning that despite the number of parcels might be sufficient (which is not), the number of contained useful pixels per parcel is small and so is the total number of pixels.
- 275 / 3703 are very small to have any chance to include an entire pixel
  Parcel_area < 0.2ha, given the pixel size (10m*10m = 100 m2 = 0.01ha)
- At least 1200 / 3703 have elongated shape (ratio: perimeter / area > very high values)

However, in many cases the long parcels are located next to each other. Therefore, a further step was carried out in order to unify (dissolve) neighbouring parcels of the same category. The following example demonstrates how the unification worked; parcels of the same category and season (in this case Wheat Organic 2016) that have common boarders (direct neighbouring) are unified to form one large parcel.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

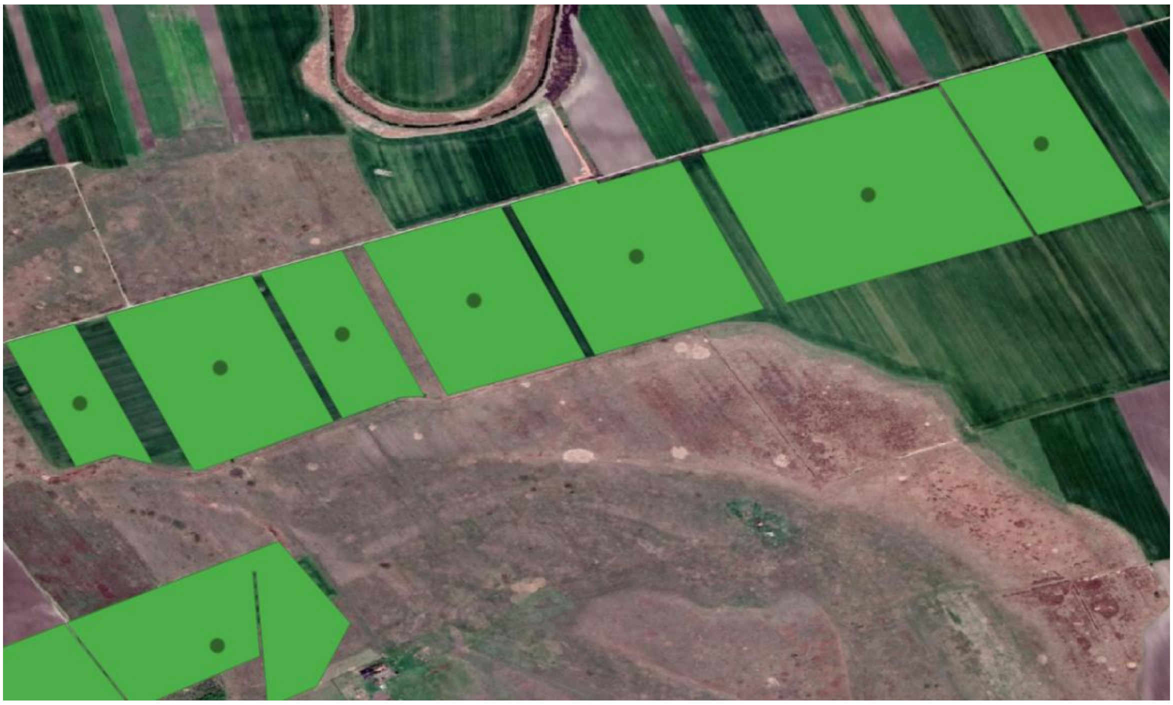


Figure 17: Unification Example

After this unification the total number of parcels was eventually reduced from 3703 to 1661 but the average parcel size increased. In the following tables the number of parcels per farming practice is presented in detail for all year and all crops before and after the parcel unification. In the blue part of the tables, the crucial Area analysis is given in the form of **Mean, Max, Min and Standard Deviation of the area** for each category.

Table 1: Number & Area Stats of parcels (pre- & post- Unification) for Wheat Organic and Wheat Conventional

Wheat	Before unification	After unification	Before unification	After unification	Before unification	After unification
	Organic	Organic Dissolved	Conventional	Conventional Dissolved	Total	Total Dissolved
2016	87	35	2	2	89	37
2017	66	25	4	4	70	29
2018	200	41	187	109	387	150
2019	77	20	213	128	290	148
2020	107	41	220	137	327	178
Total	**538**	**162**	**626**	**380**	**1163**	**542**
Mean	**3.51**	**11.63**	**1.19**	**1.97**		
St Dev	9.07	16.94	2.11	2.92		
Min	0.1	0.01	0.07	0.15		
Max	84.36	90.02	32.74	32.74		

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Table 2: Number & Area Stats of parcels (pre- & post- Unification) for Maize Organic and Maize Conventional

Maize	Before unification	After unification	Before unification	After unification	Before unification	After unification
	Organic	Organic Dissolved	Conventional	Conventional Dissolved	Total	Total Dissolved
2016	81	22	4	4	85	26
2017	15	10	10	9	25	19
2018	8	8	242	118	250	126
2019	35	10	357	164	392	174
2020	8	6	369	188	377	194
Total	**148**	**56**	**982**	**483**	**1129**	**539**
Mean	**4.72**	**14.23**	**1.52**	**3.09**		
St Dev	2.67	20.81	3.67	5.49		
Min	0.5	0.04	0.01	0.07		
Max	21	80.41	57.51	57.51		

Table 3: Number & Area Stats of parcels (pre- & post- Unification) for Soybean Organic and Soybean Conventional

Soybean	Before unification	After unification	Before unification	After unification	Before unification	After unification
	Organic	Organic Dissolved	Conventional	Conventional Dissolved	Total	Total Dissolved
2016	89	19	6	6	95	25
2017	73	20	7	7	80	27
2018	13	10	51	33	64	43
2019	17	9	79	38	96	47
2020	5	5	69	40	74	45
Total	**197**	**63**	**212**	**124**	**409**	**187**
Mean	**2.89**	**15.23**	**3.49**	**5.19**		
St Dev	1.08	21.09	0.53	12.36		
Min	0.09	0.09	0.3	0.37		
Max	10	88.6	4.9	118.61		

Table 4: Number & Area Stats of parcels (pre- & post- Unification) for Sunflower Organic and Sunflower Conventional

Sunflower	Before unification	After unification	Before unification	After unification	Before unification	After unification
			Conventional			
2016	288	68	4	3	292	71
2017	89	21	4	4	93	25
2018	58	17	88	51	146	68
2019	97	34	130	78	227	112
2020	8	7	173	110	181	117
Total	**541**	**147**	**399**	**246**	**939**	**393**
Mean	**2.08**	**12.35**	3.37	**2.22**		

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

St Dev	0.96	18.00	0.60	3.00		
Min	0.01	0.01	2.2	0.26		
Max	6.2	135.95	9.5	29.46		

B - Number of parcels per classification category

Having in mind the target of 600 useful parcels per for each category and looking at the current available parcels it is apparent that we are standing too far from the target even without unifying the parcels.

Table 5: Number of parcels per Category in comparison to the target

Category	Initially available	After unification (useful parcels)	Target
Wheat Organic	537	**162**	600
Wheat Conventional	626	**380**	600
Maize Organic	147	**56**	600
Maize Conventional	982	**483**	600
Sunflower Organic	540	**147**	600
Sunflower Conventional	399	**246**	600
Soybean Organic	197	**63**	600
Soybean Conventional	212	**124**	600

C – Parcel Dispersion & Relevance

Another issue that should be noted here is that in many cases, **elongated single parcels are located scattered** an area making it impossible to unify them with neighbouring ones, making uncertain any possibility of usefulness.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

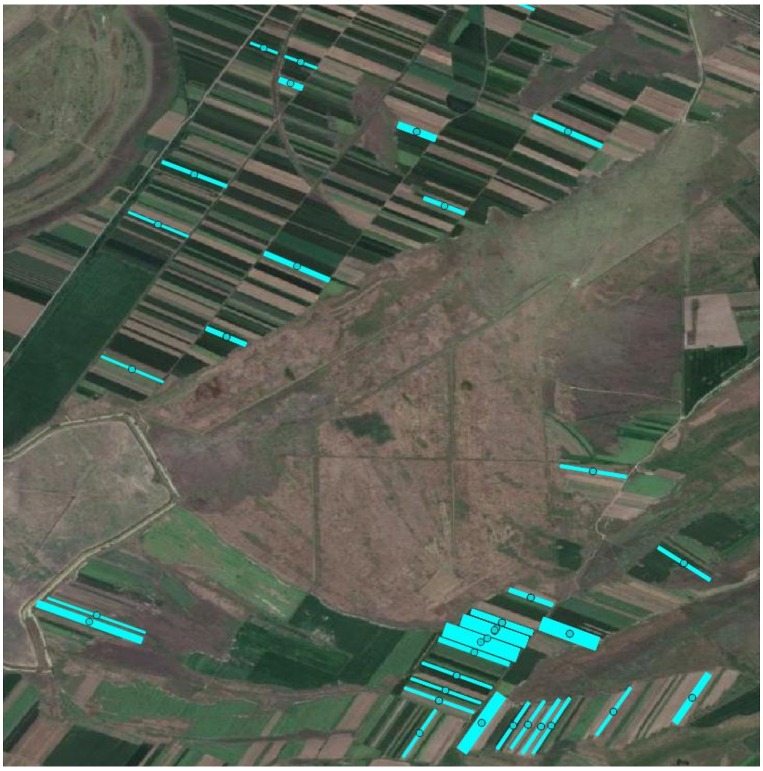


Figure 18: Elongated parcels scattered around an area

Finally, there are parcels that contain land cover not relevant with the crops, like bush/tree boundaries or roads.

Figure 19: Parcels that contain land cover

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

2.4 The Business Case of Belgium (Vlaamse Gewest – LV)

This business case will focus on a particular data product of ENVISION for monitoring CAP's soil requirements (Soil Organic Carbon) and the maintenance of soil organic matter level relative to the current and future CAP requirements.

Below is the analysis of the provided data:

The soil sampling campaign aimed to collect samples that can cover most of the SOC variability in croplands of the Flanders region. Therefore, the soil samples have been collected within the different soil regions insisting on agricultural parcels to ensure a large variability in soil types and SOC content. The SOC variability is necessary to build an effective prediction model to map SOC at a regional scale. For this purpose, we exploited the link between SOC content and spectral behaviour in the optical domain: the Sentinel-2 bands were used as feature space to determine where to collect samples by the Kennard – Stone algorithm. To ensure the proper quantity of soil samples for each soil type, we carried out a stratified feature-based approach for the sampling selection. The strata are 11 soil association regions based on the Soil Association Map of Flanders.

2.4.1 Data resources

Land Parcel Identification System (LPIS) data

LV provided the agricultural parcels[1] (shapefile format Lbgbrprc20.shp and Meta_Landbouwgebruikspercelen_LV_2020.pdf) on the latest records of the Land Parcel Identification System (LPIS) system of Flanders. The files consisted of 202,450 agricultural parcels, and one of the data fields described the primary crop type. We split this dataset into the 11-soil association region (see Table 6) we selected, and we removed all the parcels where the land use was not annual crops, thus all the parcels with low or null possibility to observe the Soil at bare conditions.

[1] https://www.geopunt.be/catalogus/datasetfolder/13cc8a4e-7292-4cb8-afc7-6e07d052b01a

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Table 6: Area info per soil association in the Flemish Region

Association code	Description	Translation	ha
15	natte zand- en lemig-zandgronden met humus of/en ijzer B horizont	wet sandy and loamy sandy soils with humus or / and iron B horizon	164971.4
60	natte alluviale gronden zonder profielontwikkeling	wet alluvial soils without profile development	108012.9
19	complex van de associaties 15 + 17	complex of associations 15 + 17	96310.5
38	niet gedifferentieerde zandlemige of lemige substraatgronden op klei-zandcomplex	undifferentiated sandy or loamy substrate soils on a clay-sand complex	85603.35
29	natte zandleemgronden met textuur B horizont of met verbrokkelde textuur B horizont	wet sandy loam soils with texture B horizont or with crumbled texture B horizont	85123.55
17	natte zand- tot licht-zandleemgronden met kleur B horizont of met textuur B horizont	wet sandy to light sandy loam soils with color B horizont or with texture B horizont	73545.93
32	leemgronden met textuur B horizont: matig droge associatie	loamy soils with texture B horizont: moderately dry association	66666.85
27	natte licht-zandleem- en zandleemgronden met verbrokkelde textuur B horizont	wet light sandy loam and sandy loam soils with crumbled texture B horizont	66129.74
14	droge zand- en lemig-zandgronden met humus of/en ijzer B horizont	dry sandy and loamy soils with humus or / and iron B horizon	61962.18
31	leemgronden met textuur B horizont: normale associatie	loamy soils with texture B horizont: normal association	59817.75
2, 3, 4, 5, 6, 7, 8, 9, 10	Polders	Polders	84003
		ALL soil associations selected	952147.2
		Flemish territory	1368207

Satellite data

We used all the level 2A Sentinel-2 images acquired between 2018 and 2021 (about 3500 images, covering the Flemish region) to select only pixels in the LPIS parcels having spectral characteristics referable to bare soil conditions. We made use of the GEE resources and S2 repositories.

```
var collection = ee.ImageCollection('COPERNICUS/S2_SR')
.filterDate('2018-05-25', '2021-04-30')
.filterBounds(geometry)
```

https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR

Soil and Lucas data and SOC measurements

We made use of the Bodemassociaties Atlas van België (see Meta_Bodemassociaties.pdf). The Atlas Digital vectorial dataset with an overview of the occurrence and classification of soil associations in Flanders, as published in the Atlas of Belgium (edition 1972), on map sheets 11A and 11B (on map scale 1: 500,000). A soil association is a substantive and spatial grouping of soil series (legend units of the Belgian soil map 1: 20,000). An association is characterized by a free constant ratio between the area occupied by a number of typical soil series. Often the nature of the association is indicated in terms of a dominant, associate and included series. For areas for which no soil map was available, extrapolations were made using geological information and expert knowledge.

Regarding the use of Lucas data, we decided not to exploit the European LUCAS topsoil database due to the good first results of the local SOC models based only on the Sentinel-2 spectra. However, if there is a future modelling need, we can integrate the collected scanned results and adjust our methodology.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

2.4.2 Sampling selection algorithm

The availability of a consistent number of ancillary data (or covariates) allows applying sampling strategies according to the feature space and not only based on the geographical space. Provided that the covariates are strongly related to the target variable, i.e., to the soil organic carbon (SOC), the sampling strategy based on feature space can ensure to collect soil samples representative of the whole range of SOC values within the investigated area. For these purposes, remote sensing data cheaply provides covariates over large areas. The physical link between spectral data in the optical domain and SOC exists and is widely exploited in remote sensing context, thus the absorbance/reflectance values at a given wavelength can be considered as covariates related to the target variable and consequently, the spectral variability can be exploited for sampling strategies based on feature space. However, some absorption features are quite broad and they can partly overlap with spectral region related to a different soil property. For this reason, it is desirable using the whole spectrum as covariates instead of a single band or a narrow region. SOC prediction models exploit most of the spectral regions across the electromagnetic spectrum between 400 and 2500 nm and this is due to the large heterogeneity of the components of the organic matter.

Synthetic layer

In the ENVISION project we used soil sampling strategy based on a stratified feature-based approach, in which the feature space consists of the satellite spectral data retrieved by a multi-temporal analysis of the Sentine-2 (S2) data and the strata are the soil associations of the Soil map of the Flemish region that assure an adequate geographical distribution and a proportioned representation of the all-soil types insisting in the Flemish region.

First, we made a bare soil composite image using a pixel-based multi-temporal analysis using S2 images acquired from 2018 to 2021. The selection of the bare soil pixels was carried out according to the computation of indices and a cloud mask that can detect green and dry vegetation and high soil moisture content that can affect the soil spectrum shape. The output of this analysis is a synthetic bare soil layer ‘(SBSL) for croplands in each soil association region. Croplands were detected using the parcels of the Land Parcel Identification System (LPIS) provided by LV.

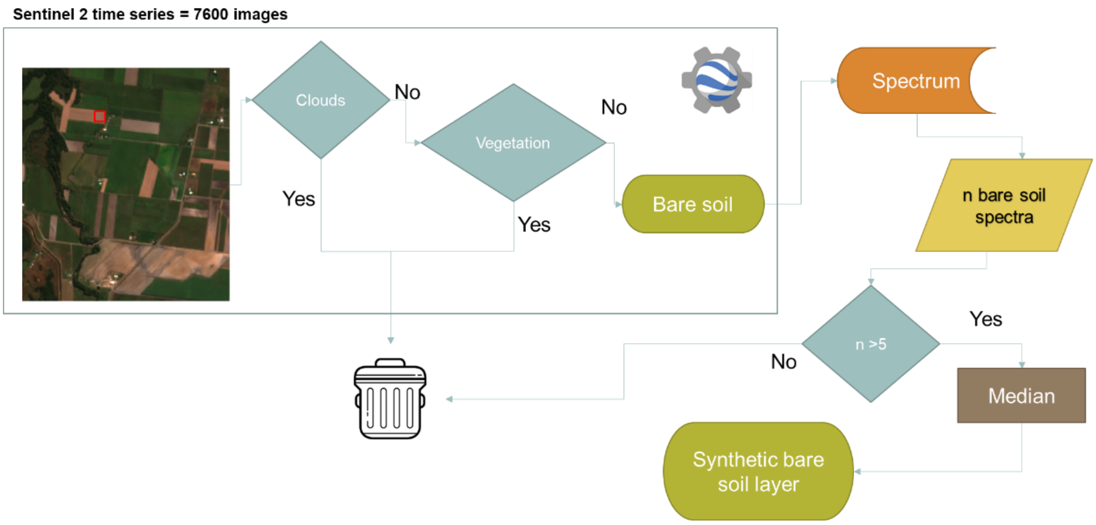


Figure 20: A flow presenting the production of a synthetic bare soil layer.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Algorithm

For each soil association region, we extracted the S2 spectra from each pixel of the SBSL that will form the feature space to assess the geographical position of the soil samples by the Kennard-Stone algorithm. This algorithm allowed to select n samples uniformly distributed over the feature space from all the spectra of the SBSL within each soil association region, thus optimizing the coverage of the spectral variability. First, the algorithm finds the two spectra that are furthest apart based on Euclidean distance assigning them to the calibration dataset and removing them from the input matrix. Then, the procedure is repeated until the number of the samples within the calibration dataset is equal to n.

Sample suggested points (outputs)

We collected samples from 171 locations, 21 more than the original estimation overcoming budget restrictions. To increase farmers' awareness and respect their rights to decide if they want to participate in the Soil Campaign, EV ILVO worked together with the LV, and LV sent a formal request to the Farmers. We identify which availability from the selected points. This way, the geographical distribution of the collected samples ensures the right proportion among soil associations, which in turn allows to properly cover the SOC variability in the Flanders region.

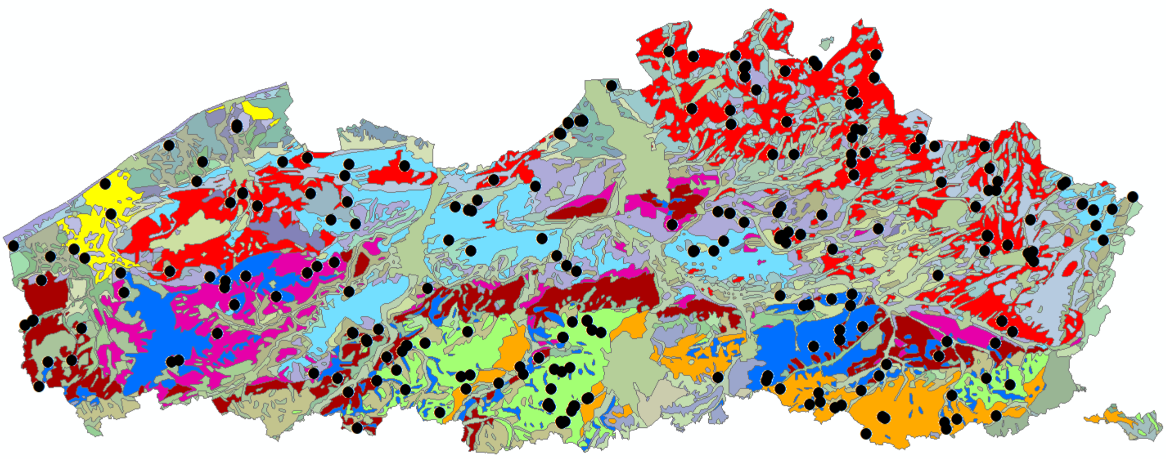


Figure 21: Kennard-Stone algorithm selected points (total 245 to deliver the needed flexibility on final selection

2.4.3 Sample collection and lab measurements

Sampling points

From the 245 sampling points that were selected, 5 points were not made available by the farmers (refuse to participate) while 13 points could not be sampled due to the presence of temporary grassland (I.e., parcels with low or null possibility to observe the Soil at bare conditions) and farm buildings. From the remaining 227 points, 171 were sampled within the project, thereby ensuring a good distribution across the different soil association classes that were identified for Flanders.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

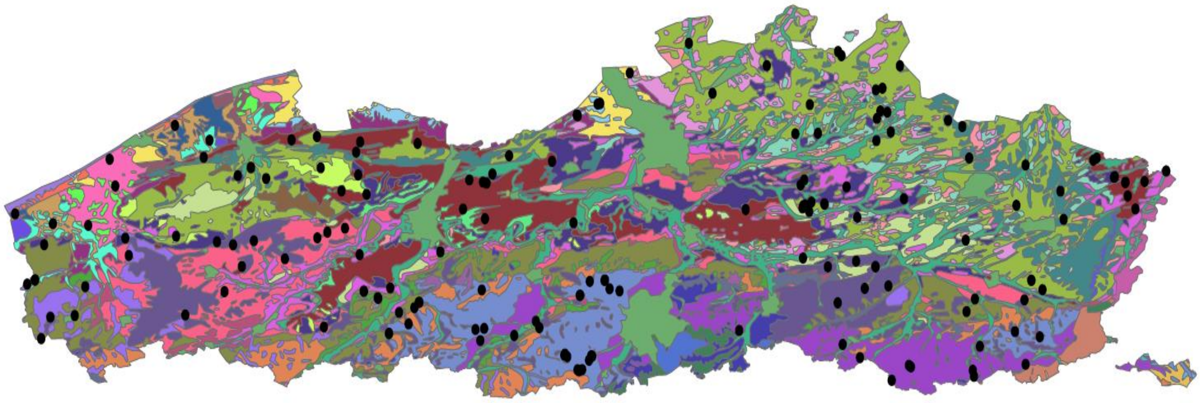
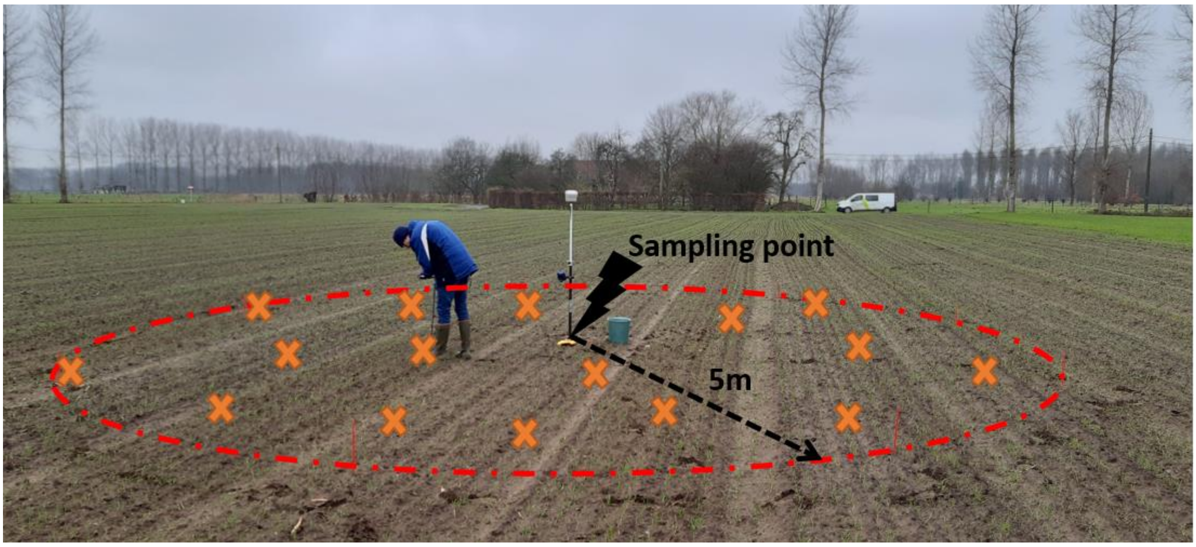


Figure 22: Final sampled locations (171)

2.4.4 Sampling protocol

In the field, the location of the sampling point was identified by means of a Stonex S10 RTK GPS (centimetre accurate). Further, an area with a radius of 5m around the sampling point was marked by red sticks after which 16 subsamples from the topsoil (0-10cm) were collected randomly within the sampling area by means of an auger with a diameter of 2.5cm. The subsamples were thoroughly mixed and stored in a labelled plastic bag for transportation to the laboratory. A picture was taken at each sampling site and some general field characteristics were monitored (e.g., land cover, soil conditions, tillage).



Picture 1: Sampling points and the subsamples area

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Table 7: Document template used for the soil campaign

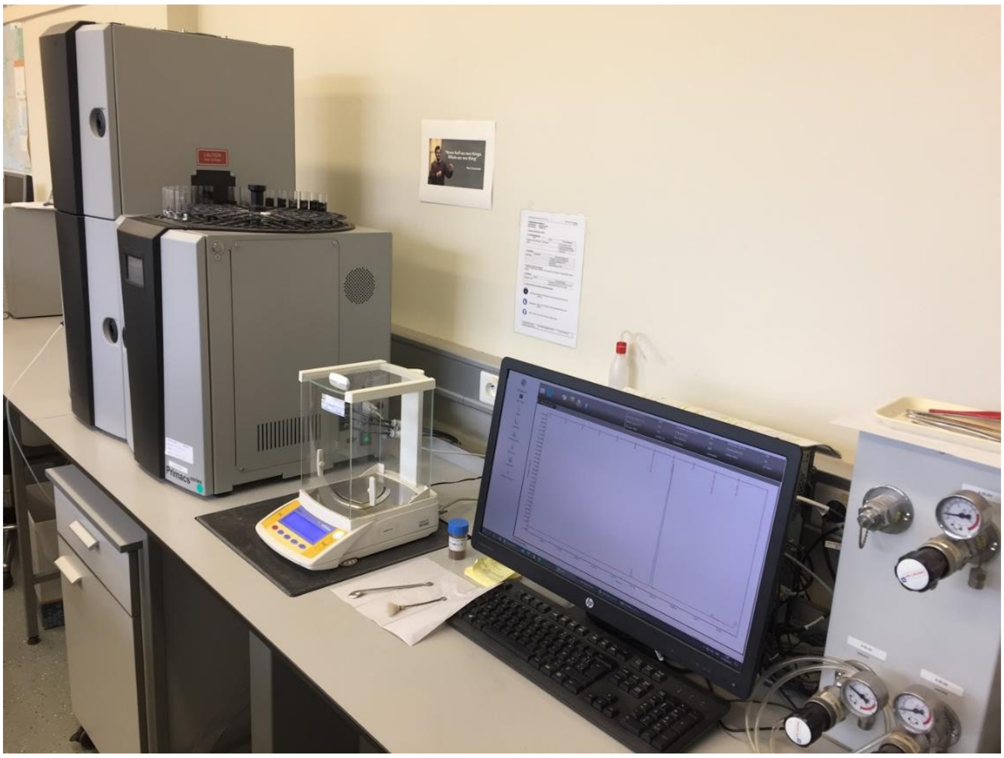
Staalnameformulier ENVISION 2021	
Algemene info	
PLOT ID: …………………………………………………	**Plaats (hoofdgemeente)** ………………………………………………
Datum:………………………………………………	
Staalnemer(s): …………………………………………………………	
Beschrijving bereikbaarheid en ligging proefvlak:	
Terreinkenmerken *(Omcirkel of vul aan)*	
Landgebruik zoals verwacht?	ja/neen, specifeer………………………………………………………
Landbedekking	Braak - Stoppel - Gewas, specifeer………………………………
Bodemtoestand	Normaal - Waterverzadigd - Bevroren - Overstroomd - Uitgedroogd
Macroreliëf:	Vlak - Depressie (droog/nat) - Helling - Top - Plateau
Bodemverstoring	Geen
	Berijding door landbouwmachines
	Erosie
	Andere, specifieer: ………………………………………………
Landbouwactiviteit	Recente bewerking: nee/ja, specifeer………………………………
	Recent bemest: nee/ja, specifeer…………………………………
	Andere, specifieer: ………………………………………………

2.4.5 Lab measurements

Prior to analysis, the soil samples were oven-dried at 40°C for 7 days, ground in a mortar and passed through a 250 µm sieve. Soil organic carbon (SOC) and Total N were measured by dry combustion using a Skalar Primacs SNC-100[2].

[2] https://www.skalar.com/analyzers/carbon-and-nitrogen-analyzers/

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366



Picture 2: Carbon and Nitrogen analysis in solid samples by using the PrimacsSNC-100 which is a modern and flexible sample analyser with integrated 100-position autosampler for determination of Nitrogen (N), Protein, Total Carbon (TC), Total Elemental Carbon (TEC), Total Inorganic Carbon (TIC) and Total Organic Carbon (TOC) all in one unit.

2.4.6 Results

The soil organic carbon content (SOC) of the soil samples is displayed in the scatter plot below. The SOC in the dataset ranged from 0.29% till 12.40% (not shown in the figure). 88% of the samples contained a SOC content between 0,5 and 2,0 %.

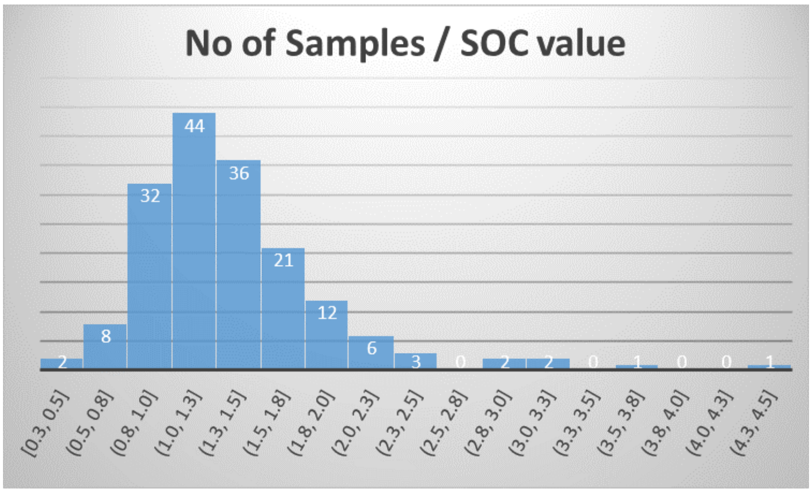


Figure 23: No of samples (y axis) and the estimated SOC value (x axis). From the 171 samples the majority takes SOC values between 0.8 - 1.8 (%/dry soil).

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Table 8:SOC Average valueper soil association

Description	AVR	StdDev
natte zand- en lemig-zandgronden met humus of/en ijzer B horizont	1.49	0.60
natte zand- en lemig-zandgronden met humus of/en ijzer B horizont	1.11	0.20
complex van de associaties 15 + 17	1.40	0.29
droge zand- en lemig-zandgronden met humus of/en ijzer B horizont	1.70	0.64
polders oudland: oude kleiplaatgronden	1.56	0.52
natte alluviale gronden zonder profielontwikkeling	2.35	2.77
natte zand- tot licht-zandleemgronden met kleur B horizont of met textuur B horizont	1.45	0.55
niet gedifferentieerde zandlemige of lemige substraatgronden op klei-zandcomplex	1.13	0.26
leemgronden met textuur B horizont: normale associatie	1.28	0.39
leemgronden met textuur B horizont: matig droge associatie	1.15	0.29
natte licht-zandleem- en zandleemgronden met verbrokkelde textuur B horizont	1.27	0.34

2.4.7 Data Catalogue

Table 9: Data catalogue of the soil campaign

<table>
<tr><th>File name</th><th>Metadata</th></tr>
<tr><td>Lbgbrprc20.shp</td><td>LPIS Argicultural Parcels – Flemish Region</td></tr>
<tr><td>ENVISION_2021_Final dataset soil sampling campaign_27042021.xlsx</td><td>
<table>
<tr><th>Column</th><th>Description</th><th>Unit</th><th>Method</th></tr>
<tr><td>FID</td><td>Identification code</td><td>-</td><td>-</td></tr>
<tr><td>soil_ID</td><td>Soil association code</td><td>-</td><td>-</td></tr>
<tr><td>X (L72)</td><td>x-coordinate</td><td>-</td><td>Lambert 72</td></tr>
<tr><td>Y (L72)</td><td>y-coordinate</td><td>-</td><td>Lambert 72</td></tr>
<tr><td>X (WGS 84)</td><td>x-coordinate</td><td>-</td><td>WGS 84</td></tr>
<tr><td>Y (WGS 84)</td><td>y-coordinate</td><td>-</td><td>WGS 84</td></tr>
<tr><td>OC</td><td>Organic carbon</td><td>%/dry soil</td><td>ISO 10694; BOC; BAM</td></tr>
<tr><td>Ntot</td><td>Total nitrogen</td><td>%/dry soil</td><td>Volgens Dumas via I</td></tr>
<tr><td>TC</td><td>Total carbon</td><td>%/dry soil</td><td>ISO 10694; BOC; BAM</td></tr>
<tr><td>IC</td><td>Inorganic carbon</td><td>%/dry soil</td><td>ISO 10694; BOC; BAM</td></tr>
</table>
Table 10 Field name, description, units and method of the lab measurements results for the 171 soil samples.

The IPR rights of the soil campaign results belong to EV ILVO.</td></tr>
<tr><td>Bodemass.shp</td><td>http://www.geopunt.be/download?container=bodemassociatie&title=Bodemassociatiekaart
see also Meta_Bodemassociaties.pdf</td></tr>
</table>

2.4.8 Annexes

Annex A: ENVISION 2021– STAALNAMEPROTOCOL / SAMPLING PROTOCOL

Version 05/01/2021

SUPPLIES

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

Documents

- map with plot ID and address (Geert folder)
- Labels
- Notice of summons Flemish government
- Staalnameformulier

Materials

- GPS (white or blue depending on availability)
- Tape measure (minimum 5 meters)
- red sticks (minimum 5)
- Stainless steel stingguts with a useful length of 30 cm and an inner diameter of 25 mm, with marking on the gouge of the bottom 10 cm (masking with tape)
- spatula
- 1 mengemmer
- Plastic bag to collect soil steel in the field and transport it to the lab
- Alcoholstift
- Cooler
- Camera/GSM
- Water bottle
- Rag to clean material

Sampling

1. The coordinate of the sampling point (= Plot ID) issearched on the site using the map and the hand GPS and is marked with a red stick. The coordinates of all sampling points can be found in the GPS under 'TOMMY -> ENVISION'. See additional manual for using the GPS.
2. With the help of a tape measure, a circle with a radius of 5m and the gps coordinate is then plotted as the centrepiece. Place at least 4 sticks to indicate the edge of the circle.
3. Then a subs steel is taken at 16 locations, randomly chosen within the surface of the circle to a depth of 10cm
    a. The bottom is first slightly kicked in around the place where drilling will take place. Before the gouge drill is pushed into the ground, the surface is cleared of organic residues such as plant residues (above-ground plant parts such as blades of grass, grass, harvest residues or remnants of green cover) and any residues of organic fertilization (e.g. compost, barn manure,...).
    b. Push the drill perpendicular to ground level in the bottom to the required depth. The drill is turned 1 revolution in a clockwise sense and then pulled up with the opening of the person's gouge away. It is important not to have a ground loss.
    c. Soil located outside the drill body (gouge) is scraped off with the spatula. The upper cm of steel must NOT be removed from the gouge. Clearly visible above-ground plant parts (e.g. blades of grass) may be removed.
    d. Release the full drill by pushing the sample out of the gouge drill with the spatula.
4. Collect all the soil material from the 16 sub-samples in the mixer and mix thoroughly. Then transfer ALL the material to a plastic bag with label and label (PLOT ID).
5. Store the bag in the cooler for transport
6. Take at least one overview photo of the terrain and sampling surface
7. Fill in the sampling form

Storage and steel preparation

8. The samples are stored in the cold store of the lab at <4°C in the box provided for this purpose. Samples may be stored in the cooling room at <4°C for a maximum of 5 days before they can move to the drying oven.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

9. The fresh Soil shall be transferred in a red bowl and placed in the drying oven at 40°C. The soil sample shall not exceed the thickness of 5 cm. If this were the case, a representative partial language would be taken. To ensure a sufficiently dry soil steel, the samples should remain in the drying oven for 7 days. Clearly state the plot ID.

Notes

1. If the plot cannot be sampled at the time of sampling due to the presence of an extremity (e.g. beet heap, manure storage, plot under water, shed, conservatory,...) the sampling is not carried out. The plot is deleted and a next plot is selected (indicate on sampling form!).
2. Within a plot, zones with standing water, avoided when taking the sub-samples.
3. If the plot spreads over 2 adjacent plots, the sub-samples are taken in both lots.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

3 Conclusions

The information presented in this deliverable is the basis of the ENVISION services development and it will be the basis for the following deliverables of WP3. An updated version of these data will be provided in the D3.3 Data products initial report on M18.

The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

End of Document

This project has received funding from the European Union`s Horizon 2020 research and innovation programme under grant agreement No. 869366.