D3.6 DATA PRODUCTS VALIDATION REPORT (FINAL VERSION)

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	D3.6 Data products validation report (final version)			
Work Package	WP3 – Earth Observation Data Products			
Date of Delivery	**Contractual**	M38	**Actual**	M38
Nature	Report	**Dissemination Level**		Public
Lead Beneficiary	DRAXIS			
Responsible Author	Eirini Pantopoulou (DRAXIS)			
Contributions from	Agathoklis Dimitrakos (AgroApps), Maria Banti (AgroApps), Panos Ilias (EV ILVO), Bert Callens (EV ILVO), Iason Tsardanidis (NOA), Thanassis Drivas (NOA), Alexia Tsouni (NOA), Stella Girtsou (NOA)			

Document History

Version	Issue Date	Stage	Description	Contributor
D0.1	16/10/2023	Draft	Draft for review	DRAXIS
D0.2	27/10/2023	Draft	Comments from review	AgroApps
F1.0	31/10/2023	Final	Final version for submission	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

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

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

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

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

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

This deliverable is the validation report of the Earth Observation (EO) data products developed during the ENVISION project. The produced EO services are tailored to monitor agricultural malpractices and the environmental impacts. This document highlights the performance of every product that was developed in ENVISION. More specifically, for every business case the pilot, the data collection, the validation results and limitations are described. This work package aims at designing and developing the EO data products of the ENVISION platform, which will address all the potential customers' specific needs. An initial version of this deliverable was provided in M18 as part of D3.4, the Data Product Validation Report. This version builds upon that initial report, incorporating additional insights and results for further validation based on refinement processes. The results presented herein represent the final outcomes based on historical data, summarizing our efforts throughout the development of ENVISION products and services. For your reference, Table 1 below presents a matrix that aligns the developed data products with the corresponding outputs within the service.

Table 1: ENVISION data products and the final services provided

ID	Related Task	Data Product	Business Case	Services	Service Provider
DP1	Task 3.3	Analytics on Vegetation and Soil Index Time-series	NMA	Harvest events detection	NOA
			NMA & CAPO	Stubble burning identification on arable land	
			CAPO	Detection of illegal land clearing in Natura2000 protection areas	
			NMA & CAPO	Minimum soil cover for soil erosion	
			NMA & CAPO	Runoff risk assessment for the reduction of water pollution in nitrate vulnerable areas	
DP2	Task 3.4	Cultivated crop type maps	NMA & CAPO	Confirmation of GSAA	NOA
				Smart sampling for OTSC inspections	
				Crops diversification compliance	
DP3	Task 3.5	Grassland mowing events detection	NMA	Grassland activity monitoring and management	NOA
DP4	Task 3.6	Soil condition monitoring	LV	Top-soil qualitative soil organic carbon estimations	EV ILVO
DP5	Task 3.7	Crop growth Monitoring and identification of organic farming practices	OCS	Distinction of organic farming practices	AgroApps
				Crop growth monitoring	

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

Introduction

The validation planning section serves to outline the pilot areas and sample data used for validating our data products and services. We won't deep into the detailed specifications of the project's developed methodologies and study areas, as all relative information is available in the respective deliverables (D3.3, "Data product initial report" and D3.7, "Data products final report").

ENVISION project involved a number of discrete pilot areas (business cases), data of differing dates, quality, resolution, or scale that will be used both during the validation procedure and the operational function. The general concept of the validation strategy consists of collecting the in-situ data, provided by the business cases' end-users of ENVISION, as well as remote sensing data collected both from the end-user and service providers during the operational period of project.

This element will describe all the relevant products of the locational data collection and image acquisition design, will define the key attributes to measured and validated, and will indicate the number and type of samples (e.g. geospatial data requirements, samples definition and description, satellite data acquired) expected. It will also describe where, when and how measurements or images were acquired.

Within validation planning, decisions are made on the type and number of samples and locations of observations. This deliverable will explain how these decisions were derived to meet the specifications of the planned interpretation (e.g. accuracy and precision) or analysis.

Useful Metrics:

Several essential metrics will used to evaluate the performance of the results. These metrics are crucial tools for assessing the performance and accuracy of your deliverable's analysis, providing a comprehensive view of how well your models or methods are performing in a wide range of scenarios.

Overall Accuracy:
Overall accuracy is a straightforward metric that measures the proportion of correctly classified instances to the total number of instances in a dataset. It is expressed as:

$$accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions}$$

Cohen's Kappa Coefficient:
Cohen's Kappa assesses the agreement between observed and expected classifications, considering the possibility of random chance. It's particularly useful for problems with imbalanced class distributions. The formula is:

$$k = \frac{P_0 - P_e}{1 - P_e}$$

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

where P_o is the observed agreement, and P_e is the expected agreement.

Precision, Recall, and F1 Score:
These metrics are fundamental in binary classification tasks. Precision (User Accuracy) measures the proportion of true positive predictions among all positive predictions, high precision means that when the model predicts something as positive, it is often correct. Recall (Producer Accuracy) gauges the ability to identify all actual positives, high recall means the model can effectively find most of the positive cases in the dataset. The F1 score combines them to find a balance between precision and recall; this is useful when you want to find a trade-off between them. It is particularly valuable when you need a single metric that considers both false positives and false negatives.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

$$F1\ Score = 2\frac{Precision \cdot Recall}{Precision + Recall}$$

Mean Absolute Error (MAE):
MAE quantifies the average absolute difference between predicted and actual values. It's an excellent measure of prediction accuracy and is calculated as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Mean Squared Error (MSE):
MSE computes the average of squared differences between predicted and actual values, giving more weight to large errors:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

R-squared (R^2) or Coefficient of Determination:
R^2 measures the goodness of fit of a regression model. It indicates the proportion of the variance in the dependent variable that is explained by the independent variables. The formula for R^2 is:

$$R^2 = 1 - \frac{SSR}{SST}$$

where SSR is the sum of squared residuals and SST is the total sum of squares.

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

1. BC1: Monitoring multiple environmental and climate requirements of CAP – Lithuania

This section outlines the validation dataset in order to evaluate the outputs provided by the respective data products (DP1-DP3) applied in case of Lithuania (BC1). Detailed information on the methodology of the respective algorithms developed is provided in D3.7.

1.1. DP1. Analytics on Vegetation and Soil-Index Time-series

This data product is designed to analyze time-series data related to vegetation and soil indices in Lithuania. It offers several algorithmic components:

- Minimum Soil Cover for Soil Erosion: This feature provides data on soil percentage and minimum soil cover, which helps assess the risk of soil erosion in different regions of Lithuania, particularly in agricultural areas. It can assist in making informed land management decisions to prevent soil erosion.
- Runoff Risk Assessment for the Reduction of Water Pollution in Nitrate Vulnerable Areas: This component assesses the risk of runoff and water pollution in nitrate vulnerable areas of Lithuania. It can be valuable in managing and mitigating water pollution, especially in regions with intense agricultural activities.
- Harvest Event Detection: This feature identifies the occurrence of harvest events in agricultural areas. It can be used to track crop harvesting seasons, optimize farm operations, and monitor crop yields.
- Stubble Burning Identification: This component is designed to detect and identify instances of stubble burning, which can be a concern for air quality and environmental impact. It can help monitor and enforce regulations related to stubble burning practices.

Sampling Description

Minimum oil cover for soil erosion

This service is designed to detect and promote the adoption of minimum soil cover practices to safeguard soil against erosion. In order to evaluate the effectiveness of this method, we carefully gathered data through on-site field inspections (OTSC) conducted by the NPA. These inspections took place at the end of the 2022 cultivation period, immediately following the initial operational implementation of ENVISION, and were guided by the alerts generated by the system's algorithm. The validation dataset comprises 82 instances of black fallow parcels, each exceeding a minimum size of 0.5 hectares. These parcels were strategically distributed across the entire country. It's important to note that out of these cases, 42 were found to be non-compliant with the relative regulations.

According to these, black fallow lands must be sown or planted with agricultural crops before the 15th of November each year. Therefore, our validation dataset primarily includes instances where these regulations were not adhered to, highlighting the need for detection through the service. To ensure the precision of our assessments, it's worth mentioning that on-site inspections for these fields occurred within a specific timeframe, starting in mid-November and concluding in mid-December of

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

the same year. This carefully chosen timeframe was intentionally selected to guarantee accurate evaluations of compliance with the minimum soil cover regulations.

Furthermore, to enhance the quality of the assessment, NOA experts conducted photo-interpretations into the total 207 predicted cases for 2022. This involved carefully examining cloud-free Sentinel-2 images to provide additional validation and enrich the assessment with a broader set of samples. More specifically, this process is facilitated utilizing the datacube services hosted in Creodias and developed by NOA. This advanced system significantly streamlines the photo-interpretation process and allows for the efficient and fast analysis of the available Sentinel-2 satellite data. Taking into advantage the geometry of the fields, this scientifically rigorous method ensures precise and reliable results into our evaluation. An analysis on the number of available samples collected is listed in Table 2 below.

Runoff risk assessment for the reduction of water pollution in nitrate vulnerable areas

Under CAP regulations, it is imperative to avoid the application of manure and/or slurry in the coastal protection zones around water bodies as delineated in the Surface Water Protection Zone layer. In response, we have devised a runoff risk assessment procedure that considers the proximity of each agricultural parcel to the nearest water surfaces. Our assessment relies on data sourced from Lithuania's hydrographic network, generously provided by the NPA, as illustrated in Figure 1. It's important to note that this service was finally deployed only into a sample pilot area of Lithuania during the year 2022, as indicated by the red-highlighted region in Figure 1.

Additionally, it's worth mentioning that the output generated for this service is of a qualitative nature. In other words, our validation process primarily focuses on **qualitative assessments** rather than quantitative metrics due to the inherently conductive nature of the service. This is a risk assessment; as such it cannot be measured directly for the respective paying agencies. Overall, this approach ensures that we maintain a high level of compliance with CAP regulations in the pilot area while providing valuable insights into the effectiveness of our runoff risk assessment procedure.

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

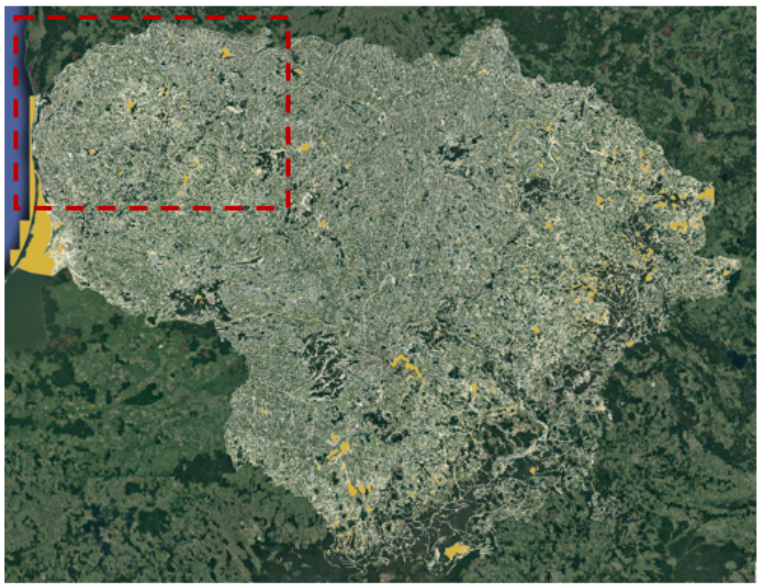


Figure 1: Lithuania's hydrographic network and pilot area for runoff risk assessment depicted with red colour.

Stubble burning identification on arable land

The stubble burning identification, has been specifically provided to address CAP requirements concerning the burning of agricultural plants and stubbles (an example is depicted in Figure 2 below). Unfortunately, for the year 2022, we faced a challenge as the NPA could only provide us with 17 validation burn parcels instances recorded from the local fire departments, which was insufficient for a thorough evaluation. To address this limitation and ensure the accuracy of our results, we took an additional step. Again, we turned into optical methods through photo-interpretation using datacube services. In this process, we closely examined the total amount of 127 burned indicated cases, which was quite manageable. Our team meticulously reviewed these cases to determine if stubble burning had actually occurred, relying on available Sentinel-2 images. This extra step was necessary due to the absence of the usual validation data.

Figure 2: Example case of stubble burning in arable crops in Lithuania.

Harvest event detection

To assess the performance of harvest event detection, we relied on validation data provided by the NPA. Specifically, they supplied us with 197 cases for the year 2022 and 77 cases for the year 2023,

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

each accompanied by the precise date of the harvest event. Additionally, to further enhance the evaluation process, NOA conducted again a photo-interpretation exercise. We randomly selected samples from various arable cultivations, including winter cereals, spring cereals, vegetables, potatoes, and others. This exercise aimed to evaluate how accurately our service identified harvest events through visual inspection within a short time-range before or after the predicted date. An analysis on the number of available samples collected is listed in Table 2 below.

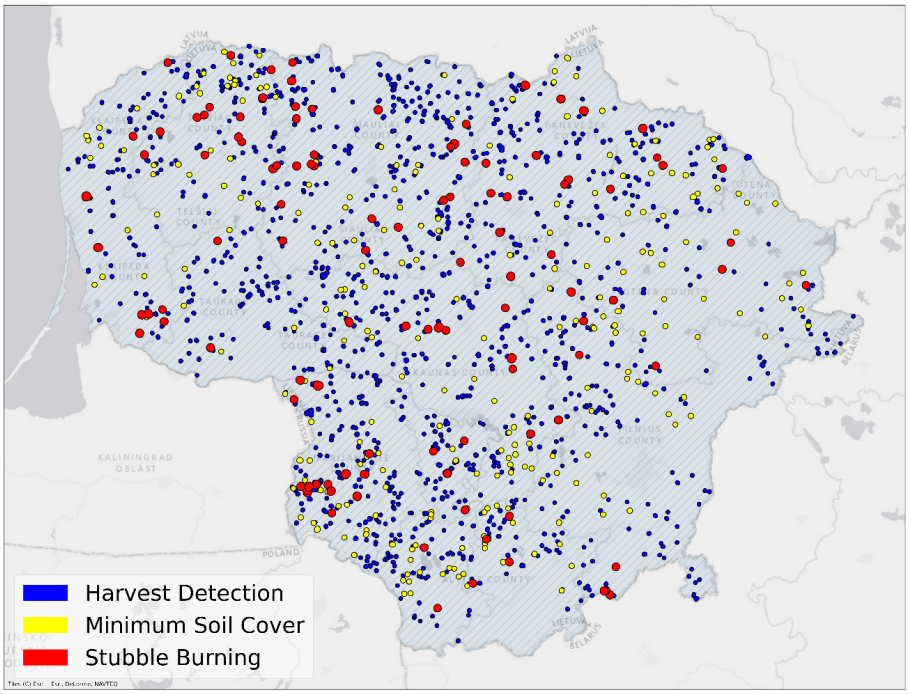


Figure 3: Validation samples geographical distribution across Lithuania territory for DP1.

Table 2: Summary numerical analysis of the samples collected for the validation of DP1 services in Lithuania.

	2022 (NPA)	2022 (NOA)	2022 (Total)	2023 (NPA)	2023 (NOA)	2023 (Total)
Minimum soil cover	42	207	**249**	-	-	-
Stubble Burning	17	127	**144**	-	-	-
Harvest Event Detection	197	1013	**1210**	100	1322	**1422**

Validation Results

Minimum soil cover for soil erosion

The minimum soil cover detection algorithm was exclusively applied for the year 2022, since this CAP requirement should be checked after mid-November of the current year of implementation. Unfortunately, for the year 2023, this assessment couldn't be carried out due to the end of the project and the termination of the Creodias infrastructure contract on August 31, 2023. As described also in D3.7, a binary mask is calculated based on the mutual fulfilment of the following conditions:

i. $0 < NDVI < 0.3$

ii. $0 < SAVI < 0.35$

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

iii. $B2 - B1 > 0$

iv. $B3 - B2 > 0$

v. $NBR2 < 0.35$

Furthermore, to categorize an area as bare ground from satellite observations, a minimum of 20% of clear pixels must show soil presence, even if the entire area isn't entirely devoid of vegetation. These specific thresholds have been selected to optimize the trade-off between recall and precision in the results.

Based on the aforementioned parameters, the algorithm flagged 207 alert cases of black fallow fields, amounting to less than 2% of the total black fallow declarations. Through Sentinel-2 image analysis, the algorithm achieved a precision **accuracy rate of 83%**. This precision is based on no evidence of any sown practice with other agricultural crops in 171 out of the 207 parcels, using at least one cloud-free image available for analysis between late August and late November. Unfortunately, we cannot objectively calculate recall due to the absence of a representative ground truth samples. However, it's noteworthy that 40 of the predicted cases intersected with the 42 non-compliant cases provided by NPA, resulting in a **rate of success of over 95%** on the available minimum soil cover incompliant cases.

Run-off Risk Assessment

The run off risk assessment algorithm takes into account the parcel's proximity to water surfaces. Initially, the algorithm iterates over every vertex of the parcel and calculate the water proximity. The minimum distance to a water surface is assigned as the corresponding value to each parcel. In addition, the Revised Universal Soil Loss Equation (RUSLE) has been calculated as it estimates the annual soil loss that is due to erosion through a factor-based approach using as input variables, described in D3.3.

it's crucial to clarify that the components on which RUSLE depends on (P, R, K, LS, and C) are spatially distributed variables. These factors are not represented by single, uniform values for an entire country but are calculated at a pixel level (10m spatial resolution by resampling process). The spatial variability accounts for differences in land use, terrain, and other localized factors, making them more accurate and context-specific. All the involved parameters are downloaded for the whole Europe from the ESDAC, cropped to each country borders and resampled to the Sentinel 2 spatial resolution (10 m), except from the LS and C factor, which were calculated using LPIS and NDVI again at 10m resolution. The runoff risk assessment model, based on RUSLE and water proximity, leverages these spatially distributed parameters to provide precise and localized insights, which are crucial for understanding the environmental dynamics and potential erosion risk within specific regions. Therefore, while it's not feasible to provide a single value per component for each country, the methodology ensures that the model accounts for the inherent heterogeneity of environmental conditions across the landscape, resulting in more accurate and region-specific assessments. Having calculated RUSLE and the minimum distance from a water surface, every parcel is labelled with a risk category as the following Table 3 indicates:

Table 3: The rules for runoff risk assessment.

RUSLE	Water Proximity (meters)			
	<=10	<=50	>50	>100
<=4	High	Low	Low	Very Low

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

>4 and <=8	High	Moderate	Low	Very Low
>8 and <=15	High	High	Moderate	Very Low
>15	Very High	Very High	Moderate	Very Low

Taking into consideration the values of water proximity and RUSLE, runoff risk has been computed for each parcel. The risk level for many parcels is high due to the fact that they are close to water surfaces. An analysis of the results exported is presented in Table 4. Furthermore, Figure 4 depicts the visualization of the parcels along with their categories and the water surfaces around them. Both layers' data is directly retrieved from the ENVISION database.

Table 4: Run-off Risk Assessment Results.

	Very Low or Low Risk	Moderate Risk	Very High or High Risk
Algorithm	79072	25167	5107

Figure 4: Visualization of the run-off risk for a subset of parcels along with the water surfaces around them.

Stubble Burning Identification

Similar to minimum soil cover, the algorithm's application was limited to the operational year 2022, resulting in the identification of 127 stubble burning cases. After conducting a quality check on the results provided using Sentinel-2 imagery, it was determined that only 31 (25%) of these were indeed genuine burning events. It's worth noting that out of the 17 ground truth cases reported by NPA and the local fire departments, more than half (12 cases) were among those identified by the algorithm.

A notable challenge encountered in Lithuanian cases was the distinction between burning of stubbles and plowing practices taking place in arable lands. These two activities often exhibit similar behavior in the monitoring indices, leading to potential misidentification (see Figure 5). Despite this, NPA expressed satisfaction with the algorithm's performance since it successfully pinpointed the majority of actual burn events and also revealed new instances of non-compliance (see Figure 6).

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

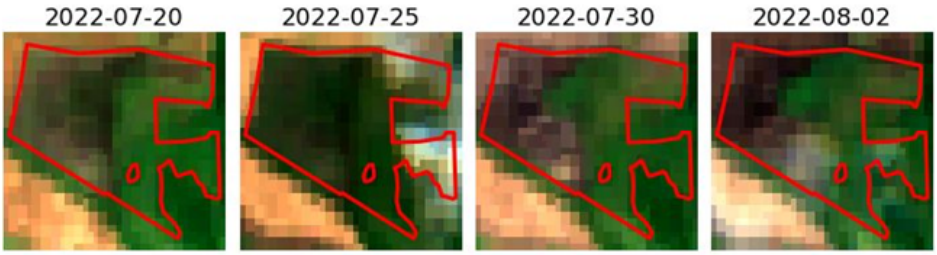


Figure 5: Example of plowing practise in arable land that was wrongly identified as stubble burning.

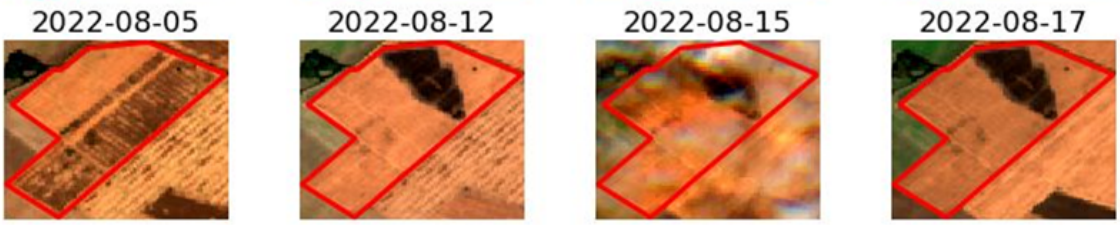


Figure 6: A stubble burning event successfully identified by the algorithm between 5 August and 12 August and was not reported by the local fire departments.

In D3.7, it's noted that stubble burning events typically exhibit significant spectral discrepancies across the affected area and the evaluated pixels, with only a small portion of the parcel actually being burned. To accurately identify burnt regions, a secondary method calculates the mean and standard deviation of post NBR after the event. True stubble burning events in Lithuania are identified when post_NBR mean value is above a specific threshold (threshold_1), and the standard deviation of post_NBR remains considerably high (threshold_2). Below is a summary analysis of the accuracy (Table 5) obtained based on the 17 events instances provided by NPA and local fire stations.

Table 5: Stubble burning detection performance after secondary filtering for some indicative set of parameters.

set	Mean (NBR) (threshold_1)	Std (NBR) (threshold_2)	Total number of burned cases detected	Total number of correct cases detected out of 17 provided by NPA	Precision (%)
1	-0.12	0.02	9902	13	0.13
2	-0.06	0.02	4867	13	0.27
3	-0.03	0.02	755	13	1.72
4	-0.12	0.04	5899	13	0.22
5	-0.06	0.04	3678	13	0.35
6	-0.03	0.04	654	12	1.83
7	-0.12	0.06	1212	12	0.99
8	**-0.06**	**0.06**	**127**	**12**	**9.45**
9	-0.03	0.06	61	5	8.20
10	-0.12	0.08	151	6	3.97
11	-0.06	0.08	27	2	7.41
12	-0.03	0.08	0	0	-

Harvest Event Detection

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

The “harvest event detection” algorithm is a threshold-based routine looking for abrupt drops on VIs values (refer on D3.7). These thresholds have been set up on samples collected from previous cultivation periods (i.e., 2020, 2021 and 2022) in order to maximize the trade-off between recall and precision. In Table 6 below, the top-5 combinations of threshold parameters are presented in descending order based on their F1-score for harvest cases of 2022.

Table 6 Top-5 set of parameters for harvest event detection in Lithuania based on harvested samples provided by NPA for 2022.

	dNDVI	R_NDVI	NDMI	dNDMI	R_NDMI	PSRI	dPSRI	BSI	dBSI	Precision (%)	Recall (%)
1	**0.05**	**0.001**	**0.2**	**0.03**	**0.001**	**0.08**	**0.03**	**-0.12**	**0.03**	**99.4**	**96.6**
2	0.05	0.001	0.15	0.03	0.001	0.08	0.01	-0.1	0.03	99.1	94.9
3	0.06	0.001	0.2	0.02	0.001	0.08	0.01	-0.12	0.03	99.5	93.8
4	0.05	0.001	0.15	0.03	0.001	0.08	0.03	-0.12	0.03	98.6	94.3
5	0.04	0.001	0.2	0.02	0.001	0.08	0.03	-0.12	0.03	98.8	93.8

The algorithm performs quite well, particularly in predicting the day of the harvest event with remarkable accuracy. During the evaluation process for both operational years of implementation, it consistently achieved satisfactory results, whether in cases of actual harvest events or non-harvest scenarios. Specifically, when assessing the samples provided by NPA, the algorithm demonstrated its effectiveness. A prediction has been considered as correct only if this was detected within a range of 6 days before to 18 days after a known cut. In 2022, it correctly identified 172 out of 178 actual harvested cases and 18 out of 19 non-harvested cases. For 2023, it achieved a perfect recall rate on 77 harvested cases and only missed 2 out of 23 non-harvested cases. Table 6 summarize the analysis of these results. Additionally, a distribution analysis of the deviation between the estimated day of the harvest event and the actual date was performed for the recovered harvested cases, providing insights into the algorithm's precision in predicting harvest events for both years (see Figure 7).

In the samples evaluated by NOA through photo-interpretation, the values of recall and precision were consistently high for both harvested and non-harvested occasions in both 2022 and 2023 (see Table 7). These results align perfectly with the findings from NPA's field inspections, demonstrating the algorithm's reliability and corroborating its performance. An example of Sentinel-2 images capturing a harvest event predicted from the algorithm is depicted in Figure 8.

Table 7: Analysis on recall, precision and support of the harvest event detection on samples collected.

Cultivation Period		2022			2023		
Dataset	Condition	Recall	Precision	Support	Recall	Precision	Support
NPA	No Harvest	0.947	0.750	19	0.913	1.0	23
	Harvest	0.966	0.994	178	1.0	0.975	77
NOA	No Harvest	0.853	0.901	197	0.893	0.888	205
	Harvest	0.979	0.977	816	0.979	0.980	1117

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

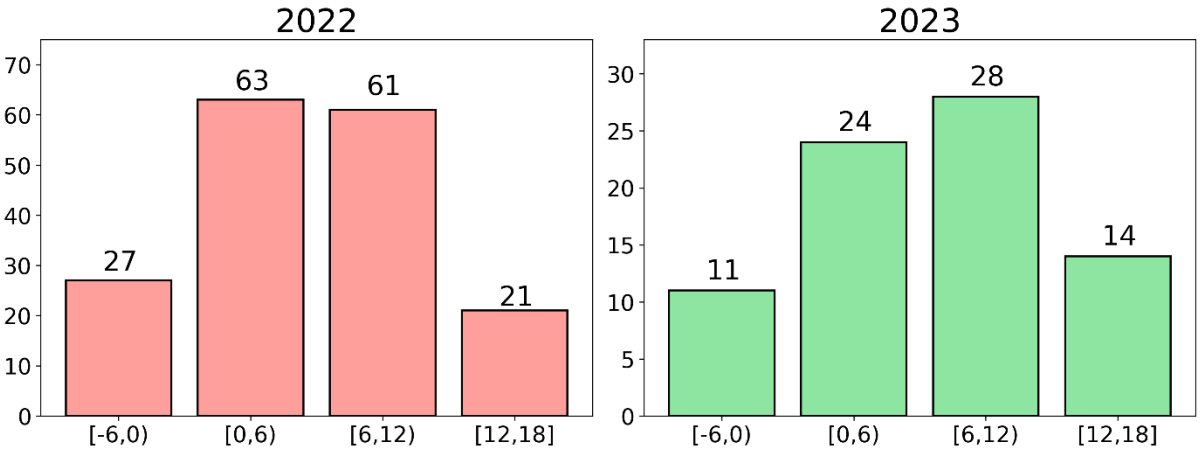


Figure 7: Counts of the deviation in days between estimated day of the parcel's harvest and the actual date reported on NPA's OTSCs validations.

Figure 8: A harvest event example successfully identified by the algorithm between 21 July and 5 August.

Discussion and limitations

The algorithmic components within DP1 exhibit an impressive overall accuracy of 98% according to NPA, rendering them suitable for integration into in-house infrastructures. Notably, the Harvest Events Detection algorithm demonstrates exceptional performance across diverse regions.

Nevertheless, several limitations have surfaced.

First, the Minimum Soil Cover for Soil Erosion algorithm relies on November data, a period with extended cloud coverage. This dependence on clear-sky conditions during a cloudy month poses a risk of reduced accuracy if essential Sentinel-2 images are obscured by clouds. This vulnerability may compromise the algorithm's reliability.

Second, there is a temporal misalignment in the scheduling of the Minimum Soil Cover for Soil Erosion algorithm compared to other components. November scheduling for this algorithm conflicts with the timing requirements of the rest of the algorithms. This misalignment can lead to operational challenges and necessitate extending the Creodias subscription to cover the entire cultivation year, potentially imposing financial burdens.

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

The Run-off Risk Assessment for Water Pollution Reduction in Nitrate Vulnerable Areas evaluation is qualitative as it operates as a risk assessment algorithm. It identifies high-risk areas based on proximity to water bodies and soil characteristics. Water pollution is primarily influenced by farming practices, and quantitative measurements are not feasible within this framework. Nevertheless, it can guide control bodies on the strategic monitoring of farming practices and regulatory compliance.

Finally, the Stubble Burning Identification algorithm encounters difficulties in distinguishing between farming tillage activities and actual stubble burning events. This limitation may lead to false positives or negatives, potentially hindering effective monitoring and regulation of stubble burning practices. Nevertheless, NPA has expressed satisfaction with the algorithm's performance. They justify this by keeping the total number of predictions relatively low and by ensuring that the majority of actual burnt cases are included. This approach suggests that the algorithm, while not perfect, still serves the NPA's needs adequately and aligns with their operational requirements.

All in all, the modules mentioned rely on manually optimized parameters to operate effectively. These parameters, often fine-tuned through human expertise, are crucial in influencing the modules' behavior and performance. Through careful adjustments, operators can enhance the modules' capabilities. While manual optimization can be a time-intensive process, it plays a significant role in achieving desired outcomes for the implementation of the respective services to other regions. This human touch, guided by experience and domain knowledge, can lead to finely tuned systems that operate with precision and efficiency.

1.2. DP2. Cultivated Crop Type Maps (CCTM)

Product Description

To assess the performance of cultivated crop type maps product, we rely on a robust validation dataset provided by NPA. This dataset includes a substantial number of samples, with 51,061 instances for 2022 and 35,077 for 2023, during the nationwide deployment of ENVISION.

We've taken great care to collect these samples to ensure they represent a balanced cross-section of different types of crops (see Figure 9). These samples are sourced through a two-fold approach by NPA: first, from remote sensing images, primarily utilizing the Sentinel-2 satellite data to identify and categorize crop types. Second, a significant portion of the validation dataset is derived from on-site field visits, where experts conduct in-person inspections to validate the accuracy of the crop type maps. The number of the aforementioned available samples is visualized in Figure 10. It's worth noting that we've selected fields of varying sizes to cover different types of farming landscapes (refer to Figure 11). Additionally, as it can be depicted in Figure 12, we've collected samples evenly from various parts of Lithuania to ensure a well-rounded representation.

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

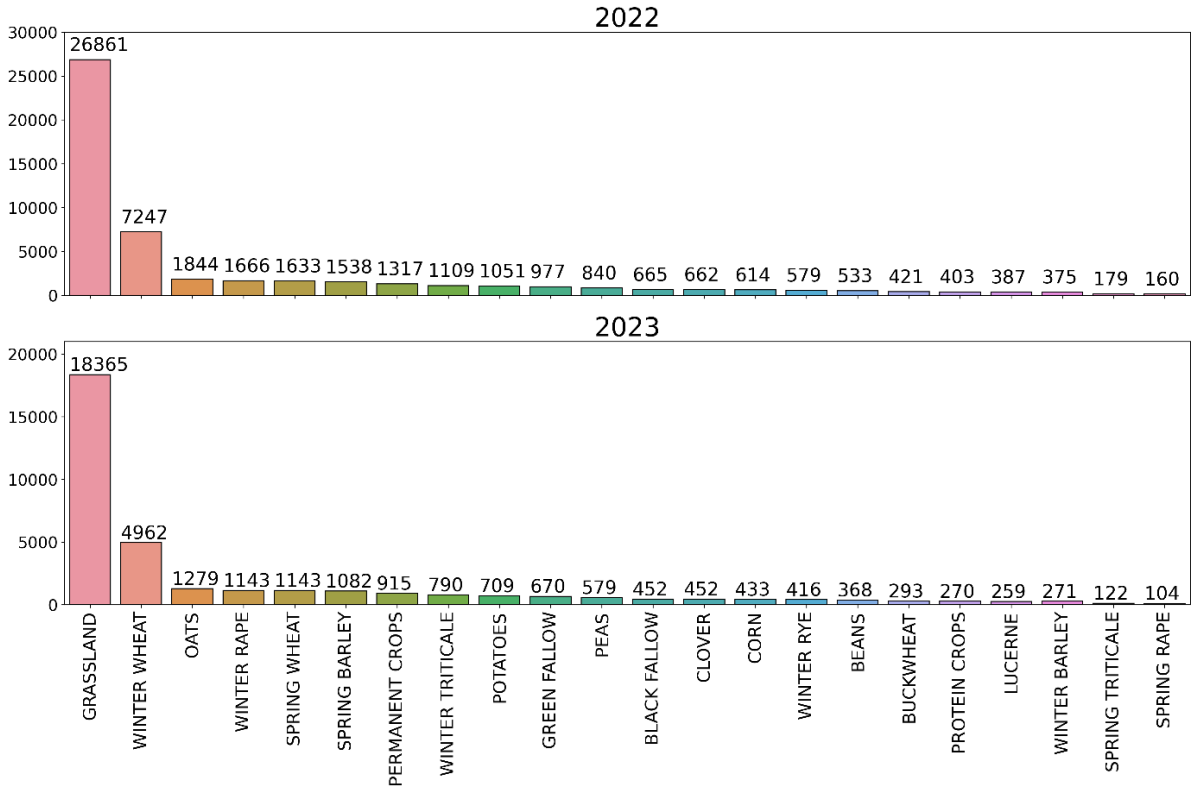


Figure 9: Crops distribution of validation samples for CCTM for 2022 and 2023.

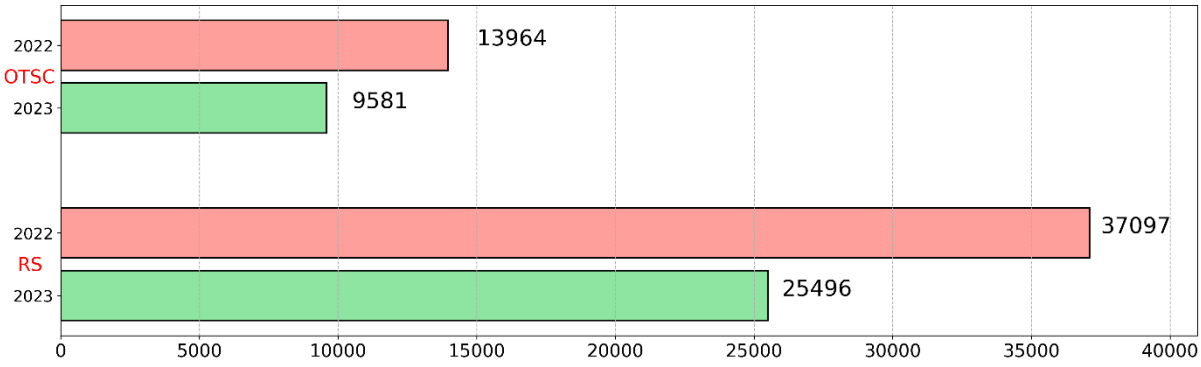


Figure 10: Number of available OTSC and RS validation for both operational year.

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

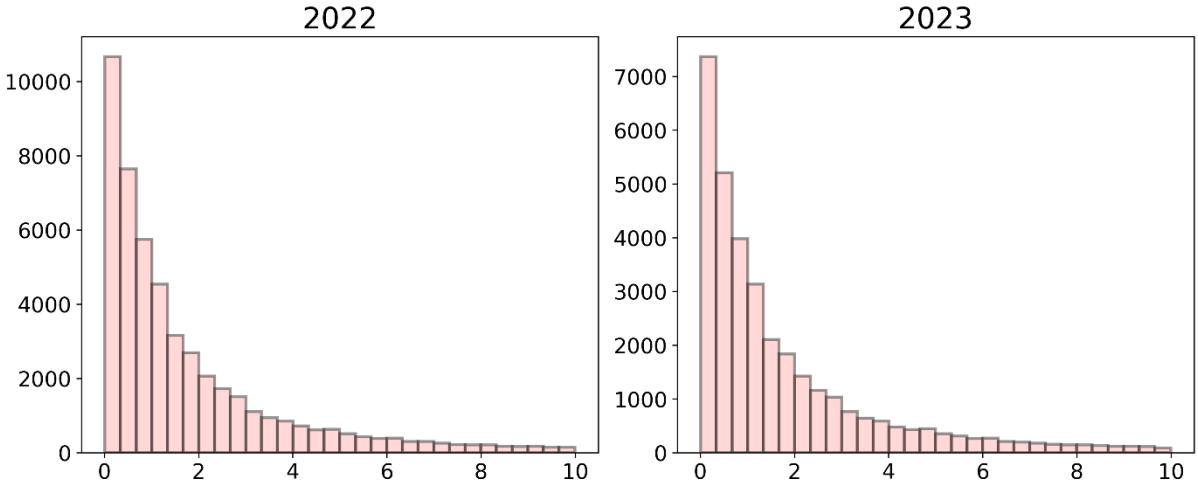


Figure 11: Distribution of parcel size of the CCTM validation. Area is calculated in hectares (ha).

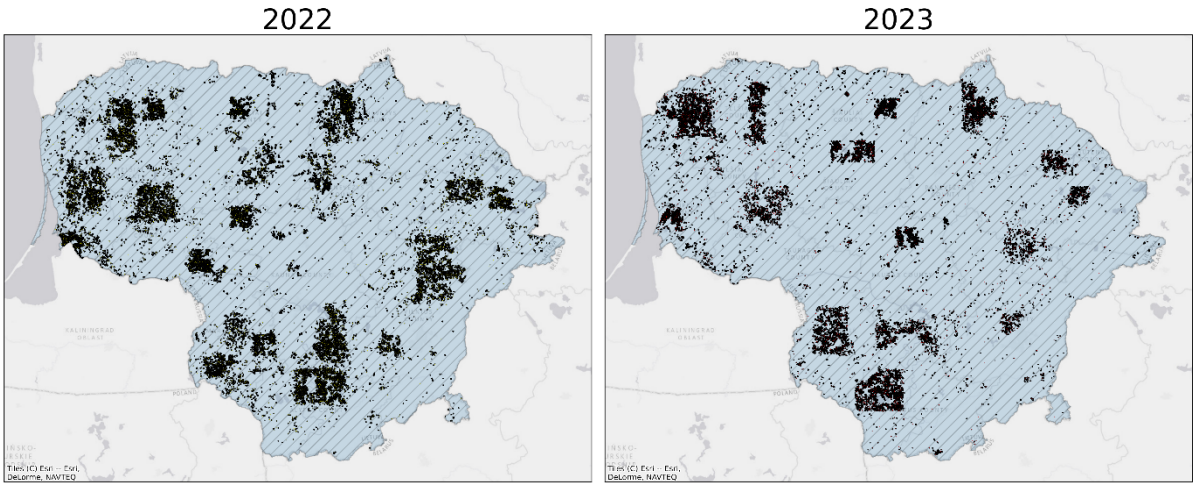


Figure 12: Validation samples geographical distribution across Lithuania territory for CCTM DP.

All the above ensures a comprehensive and reliable validation dataset that enables us to thoroughly evaluate the performance of our crop type maps product in real-world national-scale operational scenarios.

Validation Results

Crops Classification Results Performance

As detailed in D3.3 and D3.7, multiple crop type maps are generated during the cultivation period, starting from early April and continuing until the end of August. The accuracy of these models gradually improves as more data is incorporated into the analysis, reaching its peak performance by the end of August.

Table 8 displays the validation results for various machine learning models evaluated in both early September 2022 and 2023. Among these models, some have notably extended inference times.

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

Notably, the Random Forest model stands out by delivering the optimal results in terms of accuracy and elapsed time.

Table 8: Classification performance for different machine learning models based on the predictions provided at the early September for 2022 and 2023.

	2022				2023			
	RF	SVM	XGBoost	MLP	RF	SVM	XGBoost	MLP
Recall (Macro Avg.)	**0.75**	0.68	0.75	0.66	**0.75**	0.68	0.75	0.66
Recall (Weighted Avg.)	**0.90**	0.86	0.92	0.85	**0.90**	0.86	0.92	0.85
Precision (Macro Avg.)	**0.88**	0.90	0.88	0.86	**0.88**	0.90	0.88	0.86
Precision (Weighted Avg.)	**0.90**	0.92	0.91	0.89	**0.90**	0.92	0.91	0.89
F1-Score (Macro Avg.)	**0.78**	0.72	0.79	0.69	**0.78**	0.72	0.79	0.69
F1-Score (Weighted Avg.)	**0.89**	0.85	0.90	0.79	**0.89**	0.85	0.90	0.79
Overall Accuracy	**0.90**	0.87	0.91	0.84	**0.90**	0.87	0.91	0.84
Kappa Coeff.	**0.85**	0.82	0.86	0.79	**0.85**	0.82	0.86	0.79
Elapsed Time (min.)	**5.3**	35.1	19.6	5.9	**4.1**	27.3	12.9	4.0

Table 9 provides the classification report for 22 distinct crop classes, reflecting the validation analysis conducted throughout the cultivation period for 2022 and 2023. This report offers a snapshot of the model's final classification results, which were assessed at the beginning of September.

In terms of accuracy, the classifier performs quite well for most crop classes, achieving high accuracy rates. However, there are exceptions, particularly for classes with very limited support data (e.g., clover, green fallow, lucerne). Additionally, the performance tends to be lower for permanent crops, often due to confusion with grasslands. Overall, the model demonstrates an impressive level of accuracy, coming very close to the 90% mark for both cultivation years, with similar performance observed in both years. It's important to note that certain classes, such as mixed crops (e.g., agricultural mixes, other vegetables, and other crops on arable land), were excluded from this analysis because they lack a distinct spectral profile, and the classification model struggles to perform well on these cases.

Table 9: Classification Report based on the predictions provided at the early September for 2022 and 2023.

	2022				2023			
	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
Beans	0.87	0.96	0.91	533	0.87	0.96	0.91	368
Black Fallow	0.78	0.75	0.77	665	0.79	0.76	0.77	452
Buckwheat	0.78	0.92	0.84	421	0.79	0.91	0.84	293
Clover	0.98	0.20	0.33	662	0.98	0.20	0.33	452
Corn	0.93	0.96	0.95	614	0.93	0.95	0.94	433
Grassland	0.92	0.97	0.95	26861	0.92	0.97	0.95	18365
Green Fallow	0.75	0.64	0.69	977	0.76	0.64	0.70	670
Lucerne	0.93	0.18	0.30	387	0.95	0.16	0.28	259
Oats	0.77	0.88	0.82	1844	0.77	0.89	0.83	1279

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

Peas	0.87	0.89	0.88	840	0.88	0.89	0.88	579
Permanent Crops	0.73	0.45	0.56	1317	0.74	0.47	0.57	915
Potatoes	0.70	0.84	0.76	1051	0.69	0.85	0.76	709
Protein Crops	0.97	0.63	0.76	403	0.95	0.63	0.76	270
Spring Barley	0.89	0.78	0.83	1538	0.88	0.78	0.83	1082
Spring Rape	0.95	0.79	0.87	160	0.96	0.78	0.86	104
Spring Triticale	0.98	0.59	0.73	179	0.97	0.58	0.73	122
Spring Wheat	0.86	0.84	0.85	1633	0.86	0.84	0.85	1143
Winter Barley	0.96	0.77	0.86	375	0.95	0.77	0.85	271
Winter Rape	0.97	0.97	0.97	1666	0.96	0.97	0.97	1143
Winter Rye	0.94	0.85	0.89	579	0.93	0.85	0.89	416
Winter Triticale	0.84	0.73	0.78	1109	0.84	0.72	0.78	790
Winter Wheat	0.92	0.93	0.93	7247	0.92	0.93	0.93	4962
Macro Avg.	**0.88**	**0.75**	**0.78**	**51061**	**0.88**	**0.75**	**0.78**	**35077**
Weighted Avg.	**0.90**	**0.90**	**0.89**		**0.90**	**0.90**	**0.89**	
Overall Accuracy	**0.90**				**0.90**			
Kappa Coeff.	**0.85**				**0.85**			

The results exhibit a noticeable improvement as the cultivation period progresses (see Figure 13). This enhancement can be attributed to the increasing clarity of spectral characteristics, allowing for more effective discrimination between different crop types. Notably, spring crops such as beans, peas, and corn demonstrate a significant boost in accuracy after June of the monitored year. Crucially, the results

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

for black fallow have achieved a satisfactory level of performance at the end of the period, aligning well with the minimum expectations specified by NPA.

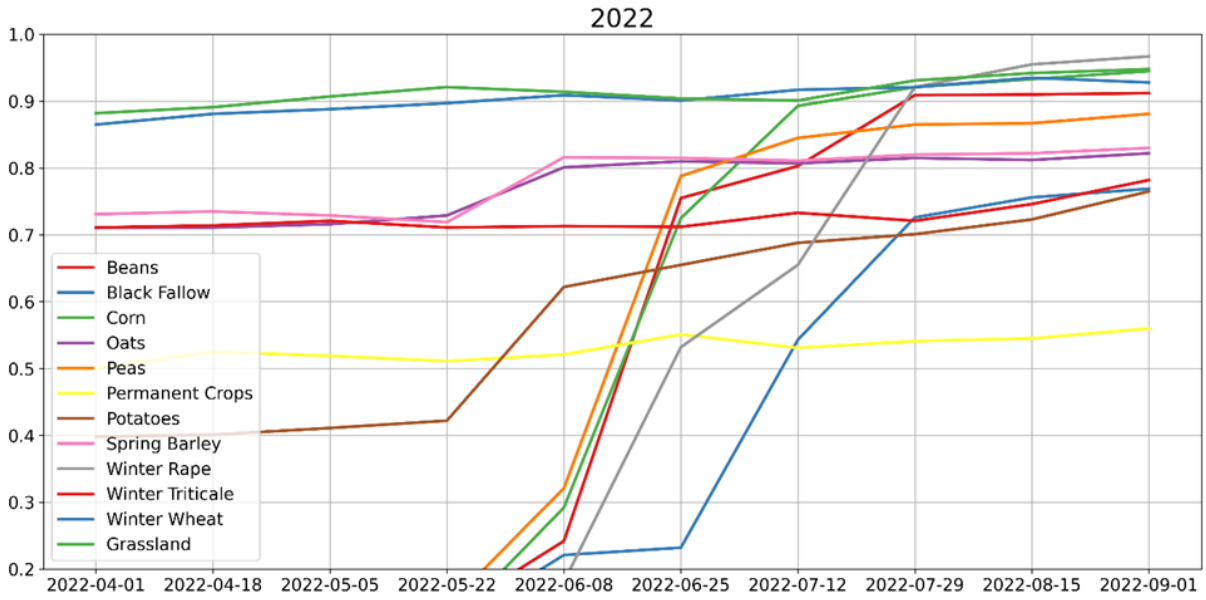


Figure 13: Classifier F1 score Progress over Cultivation Period of 2022. Results are similar for 2023.

In summary, our model provides probabilities for all available crop classes, and these probabilities sum up to 1. It's intuitive that higher probabilities correspond to higher accuracies, as expected. To quantify the difference between the most confident prediction and the second one, we can assess the likelihood of correctly identified cases. As illustrated in Figure 14, as this difference becomes more significant, accuracy gradually improves. However, it's important to note that this approach impacts the total number of cases predicted. Notably, for around 60% of the total cases, we observe a difference higher than 0.6, while only about 20% exhibit a difference higher than 0.9. This trade-off between the algorithm's accuracy level and the number of predicted cases is a critical factor that will guide our approach in interpreting the results. It serves as the foundation for providing the smart sampling service, which aims to pinpoint the most certain instances of incorrect farmer declarations. These

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

identified cases will be invaluable to end-users for strategically planning their field inspections and conducting thorough monitoring activities.

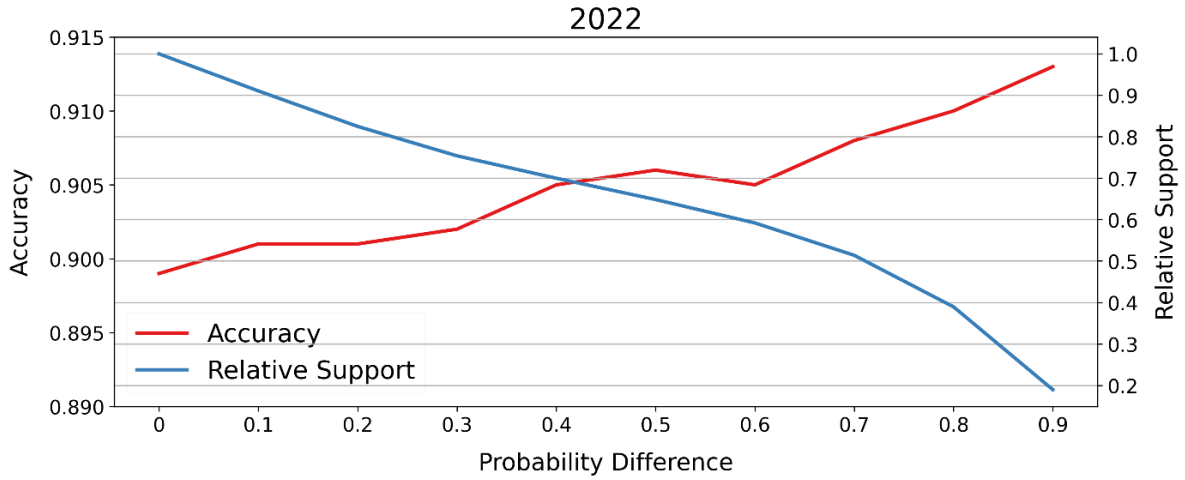


Figure 14: Accuracy and Relative Support (i.e., number of cases above this threshold/ total number of cases) trade-off for different values of probability difference between the 2 first most confident predictions.

Results Interpretability

Tables 10 and 11 below depict the producer and user accuracy of the different cases respectively, as well as the loss of information among classes and how the model confuses them.

The producer accuracy table (i.e., recall table) indicates the crop type distribution of the false negative instances, namely what crop types does the model predict when it makes a mistake, for each one of the different classes. For example, for the case of clover we can see that 73% of actual clover cases has been predicted mistakenly as grass. Clover and grass have very similar spectral signatures, but the latter has almost 30x more samples in the dataset, and thus the model reasonably struggles to identify the clovers. Similarly, all the winter cereals are confused them since they belong in the same group of cereals and present very similar characteristics.

On the other hand, the user accuracy table (i.e., precision table) indicates the crop type distribution of the false positive instances, namely what is the ground truth of the predictions that the model makes a mistake. For instance, we can see that from the total predicted black fallow cases, 78% were indeed black fallows, while 7% were in fact grasslands. Interestingly, in the occasion of clover we see that, even though the algorithm can cannot distinguish clovers from grasslands as stated before, almost everything (98%) that has been predicted as clover, is indeed a clover.

These confusion tables are really useful for results interpretability. It is obvious that the model performs better in terms of User's Accuracy instead of Producer's Accuracy, which means that the model can identify successfully the spectral behavior of almost all crop types. This is significant for the sub-sequent smart sampling algorithm, since it is based on the predictions and their level of confidence in order to highlight the respective alerts of false declarations. Overall, these results were calculated at the end of August of 2022, when NPA was needing to have the first results in order to start strategically planning their OTSC campaigns. Results for 2023 cultivation period were similar.

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

Table 10: Lithuania Producer Accuracy Table for 2022. Results for 2023 are similar.

Crop Name	Declared Parcels	Well Classified	Producer Accuracy	Confusion Class 1	1%	Confusion Class 2	2%	Confusion Class 3	3%	Rest %
Beans	533	513	0,962	Oats	0,01	Peas	0,01	Spring Barley	0,01	0,01
Black Fallow	665	502	0,755	Grassland	0,12	Green Fallow	0,04	Potatoes	0,03	0,06
Buckwheat	421	387	0,919	Grassland	0,04	Potatoes	0,01	Oats	0,01	0,02
Clover	662	133	0,201	Grassland	0,73	Potatoes	0,02	Black Fallow	0,02	0,04
Corn	614	589	0,959	Potatoes	0,01	Buckwheat	0,01	Green Fallow	0,01	0,02
Grassland	26861	26158	0,974	Permanent Crops	0,01	Potatoes	0,01	Winter Wheat	0	0,01
Green Fallow	977	624	0,639	Grassland	0,11	Oats	0,05	Buckwheat	0,04	0,16
Lucerne	387	68	0,176	Grassland	0,8	Potatoes	0,01	Oats	0,01	0,02
Oats	1844	1629	0,883	Grassland	0,02	Spring Wheat	0,02	Spring Barley	0,02	0,05
Peas	840	751	0,894	Potatoes	0,03	Oats	0,02	Green Fallow	0,02	0,04
Permanent Crops	1317	596	0,453	Grassland	0,51	Potatoes	0,02	Black Fallow	0	0,02
Potatoes	1051	885	0,842	Grassland	0,09	Peas	0,02	Winter Wheat	0,01	0,04
Protein Crops	403	253	0,628	Oats	0,14	Grassland	0,1	Green Fallow	0,05	0,09
Spring Barley	1538	1195	0,777	Oats	0,08	Spring Wheat	0,06	Grassland	0,02	0,06
Spring Rape	160	127	0,794	Winter Rape	0,06	Peas	0,06	Buckwheat	0,04	0,04
Spring Triticale	179	105	0,587	Spring Wheat	0,13	Oats	0,12	Spring Barley	0,07	0,09
Spring Wheat	1633	1376	0,843	Oats	0,06	Winter Wheat	0,02	Spring Barley	0,02	0,06
Winter Barley	375	290	0,773	Winter Wheat	0,12	Grassland	0,05	Spring Barley	0,01	0,04
Winter Rape	1666	1610	0,966	Winter Wheat	0,02	Grassland	0,01	Peas	0	0,01
Winter Rye	579	495	0,855	Winter Wheat	0,06	Grassland	0,05	Winter Triticale	0,02	0,02
Winter Triticale	1109	810	0,73	Winter Wheat	0,19	Grassland	0,04	Winter Rye	0,01	0,03
Winter Wheat	7247	6775	0,935	Grassland	0,02	Winter Triticale	0,02	Spring Wheat	0	0,02

Table 11: Lithuania User Accuracy Table for 2022. Results for 2023 are similar.

Crop Name	Classified Parcels	Well Classified	User Accuracy	Confusion Class 1	1%	Confusion Class 2	2%	Confusion Class 3	3%	Rest %
Beans	592	513	0,867	Grassland	0,02	Spring Barley	0,02	Green Fallow	0,02	0,07

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

Black Fallow	641	502	0,783	Grassland	0,07	Green Fallow	0,06	Winter Wheat	0,02	0,07
Buckwheat	496	387	0,78	Green Fallow	0,08	Grassland	0,05	Spring Barley	0,02	0,06
Clover	136	133	0,978	Green Fallow	0,01	Grassland	0,01	Winter Wheat	0	0
Corn	632	589	0,932	Grassland	0,02	Oats	0,01	Winter Wheat	0,01	0,03
Grassland	28348	26158	0,923	Permanent Crops	0,02	Clover	0,02	Lucerne	0,01	0,03
Green Fallow	830	624	0,752	Grassland	0,07	Black Fallow	0,03	Oats	0,03	0,12
Lucerne	73	68	0,932	Grassland	0,05	Clover	0,01	Winter Wheat	0	0
Oats	2120	1629	0,768	Spring Barley	0,06	Spring Wheat	0,04	Grassland	0,03	0,1
Peas	864	751	0,869	Green Fallow	0,03	Potatoes	0,03	Spring Rape	0,01	0,06
Permanent Crops	813	596	0,733	Grassland	0,25	Winter Wheat	0,01	Clover	0,01	0
Potatoes	1264	885	0,7	Grassland	0,12	Oats	0,02	Winter Wheat	0,02	0,14
Protein Crops	262	253	0,966	Grassland	0,02	Green Fallow	0,01	Spring Barley	0	0
Spring Barley	1341	1195	0,891	Oats	0,02	Spring Wheat	0,02	Winter Wheat	0,01	0,05
Spring Rape	133	127	0,955	Green Fallow	0,04	Winter Wheat	0,01	Peas	0	0
Spring Triticale	107	105	0,981	Spring Wheat	0,01	Winter Triticale	0,01	Winter Wheat	0	0
Spring Wheat	1601	1376	0,859	Spring Barley	0,05	Oats	0,03	Winter Wheat	0,02	0,04
Winter Barley	303	290	0,957	Winter Wheat	0,03	Winter Rye	0	Buckwheat	0	0,01
Winter Rape	1665	1610	0,967	Winter Wheat	0,01	Spring Rape	0,01	Grassland	0	0,01
Winter Rye	528	495	0,938	Winter Wheat	0,03	Winter Triticale	0,02	Green Fallow	0,01	0,01
Winter Triticale	963	810	0,841	Winter Wheat	0,13	Winter Rye	0,01	Winter Barley	0,01	0,01
Winter Wheat	7349	6775	0,922	Winter Triticale	0,03	Grassland	0,01	Winter Barley	0,01	0,03

Towards smart sampling

Throughout the entire cultivation period, we rely on prediction confidence levels to identify confidently declared cases that may be incorrect. Specifically, focusing on the last classification run conducted at the end of August, we find that more than 90% of mismatched cases predicted by the model are indeed wrongly declared cases by the farmers. Impressively, in 85% of these cases, the model accurately predicts the correct crop type. Furthermore, our methodology successfully identifies around 85% of all actual wrongly declared cases, demonstrating it's high recall rate. The essence of the

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

smart sampling algorithm is grounded in the belief that the most confident model predictions reflect the truth. Therefore, if a prediction doesn't agree strongly with the true label, we consider it a wrongly declared case.

In addition, alert cases are dynamically determined based on the probabilities associated with the results. We assess the level of alerts by considering the difference in the probabilities between the two most confident predictions (critical parameter a) and the total number of cases flagged as wrongly declared throughout the cultivation year (critical parameter b), also named as persistent misclassifications. This assessment is visualized using a traffic light system. For highrisk alerts (level 2 and 3), these parameters are configured to characterize the estimated percentage of false declaration, which in BC of Lithuania is approximately 3%, based on former applicants declarations.

In Figure 15, precision and recall are displayed for different percentages of disagreements among the total number of declarations. Misclassified (based on their initial declarations) instances are sorted inversely based on their confidence intervals between the two most confident classifier predictions. The optimal trade-off between precision and recall is achieved at roughly 5%, a bit higher to the expected number of false declarations. Parameter a is the confidence interval threshold used to identify the most confident disagreements. It's dynamically set to represent a ratio of 5% of the total instances, slightly exceeding the expected false declaration percentage.

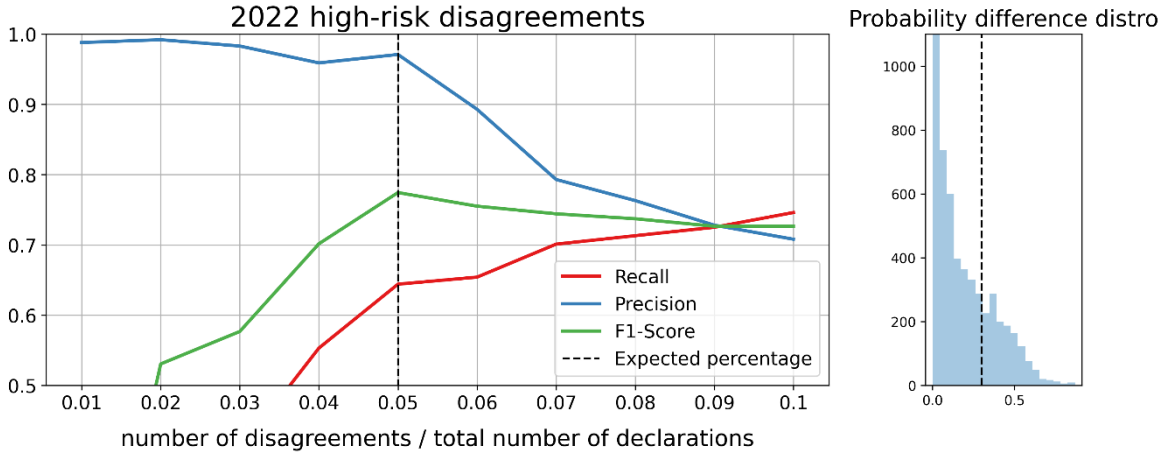


Figure 15: Evaluation of accuracy on high-risk disagreement vs the number of alert cases based on the confidence interval between two major predictions distribution (parameter a).

Parameter b, termed "Persistence," represents the count of times a sample has been consistently misclassified in various classification iterations. This was set at a value of 2.

In Figures 16 and 17 below, we illustrate the progress of precision and recall for high-risk alert cases (level 2 and 3) during the cultivation period for 2022 and 2023, respectively. Precision approaches near-perfect values early in the cultivation period, while the recall of false declarations reaches its peak at the end of the year. The rest of the actual wrongly declared cases have been either allocated in lower-risk alert cases or not detected at all. Figure 20 illustrates the distribution of actual wrongly declared cases categorized by the different alert levels generated by our system.

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

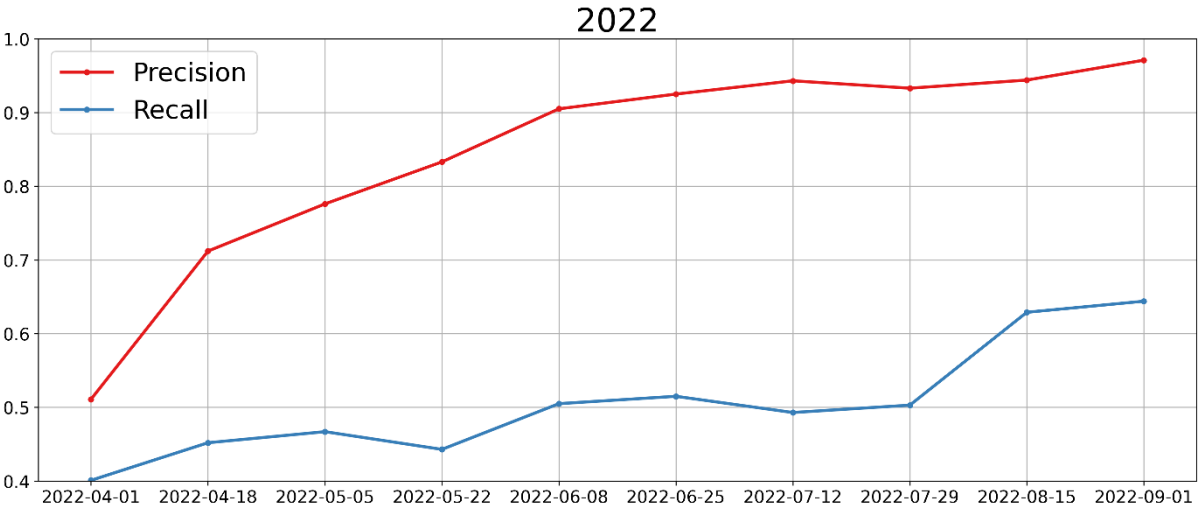


Figure 16: Progress of precision and recall of high-risk alert cases (level 2 and 3) during 2022.

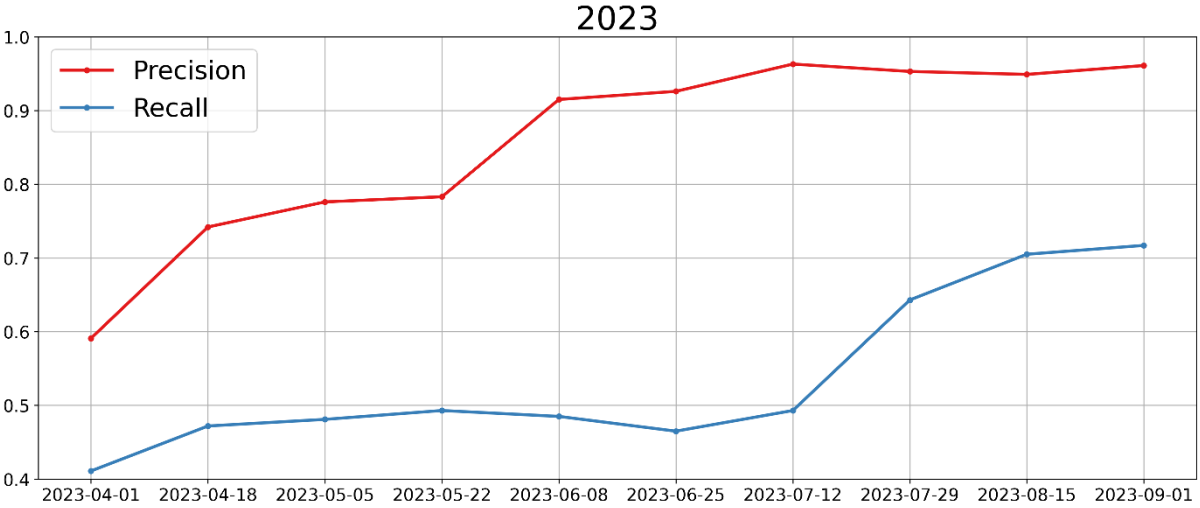


Figure 17: Progress of precision and recall of high-risk alert cases (level 2 and 3) during 2023.

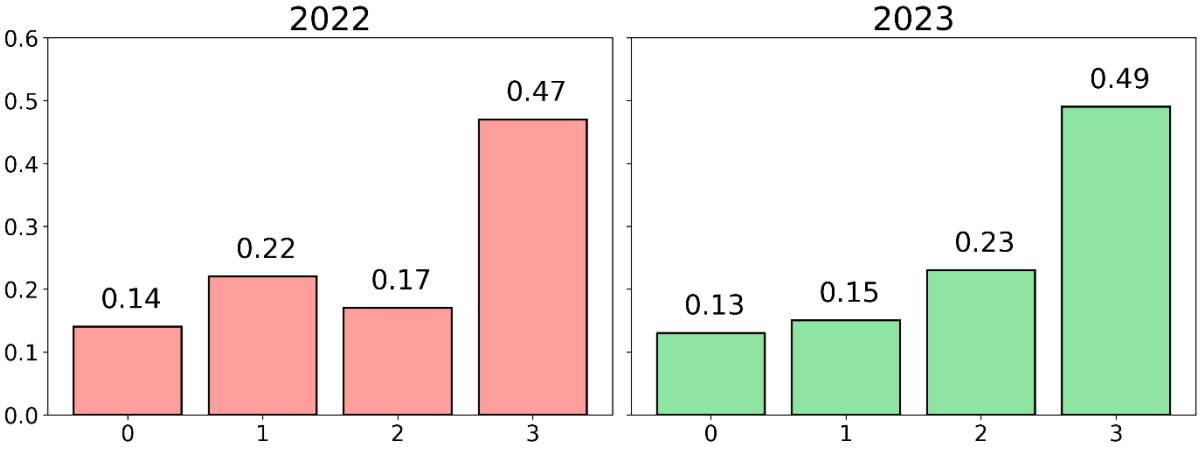


Figure 18: Portion of actual wrongly declared cases distribution among the various risk alerts output.

In Figures 19 and 20, we provide visual representations of two parcels indicated as high-risk alerts on 2022 to highlight the differences between the average NDVI behaviour of the declared crop type

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

(orange colour), the predicted crop type from the classifier (green colour), and the actual NDVI time series of the specific parcel (blue colour). These figures clearly show that in both cases, the curve of the sample and the average curve of the predicted crop type closely resemble each other.

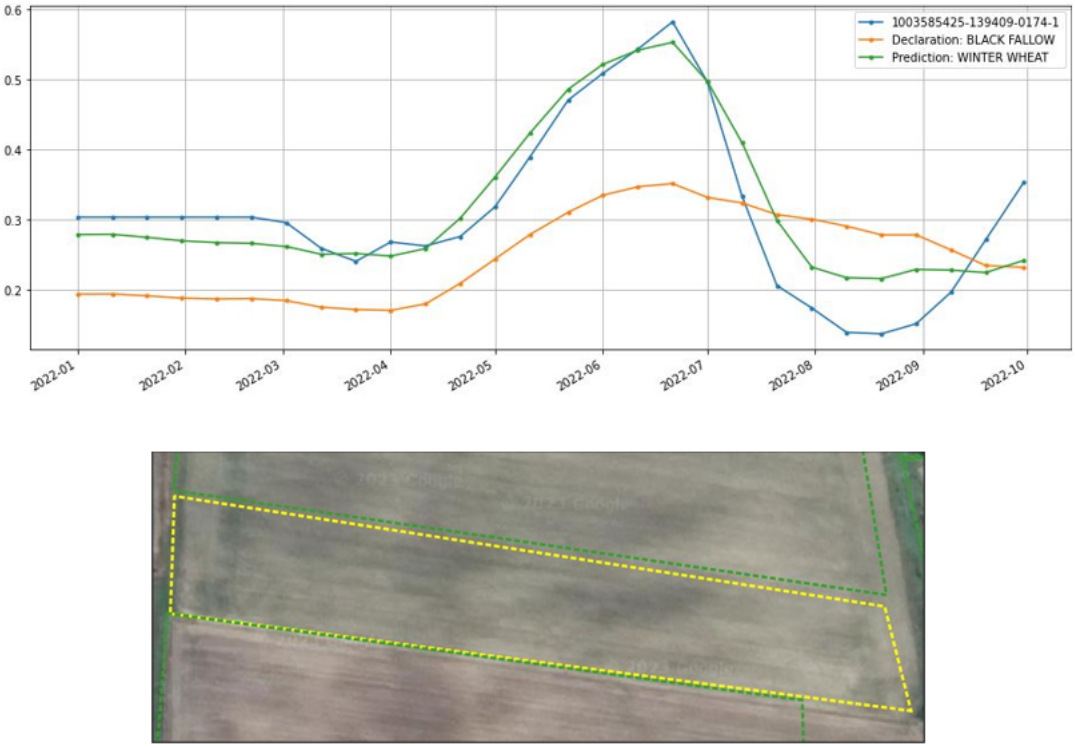


Figure 19: NDVI and S2 image of a case predicted as Winter Wheat and declared as Black Fallow. According to this plot, black fallows present lower values of NDVI during the cultivation period, which is definitely not evident here.

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

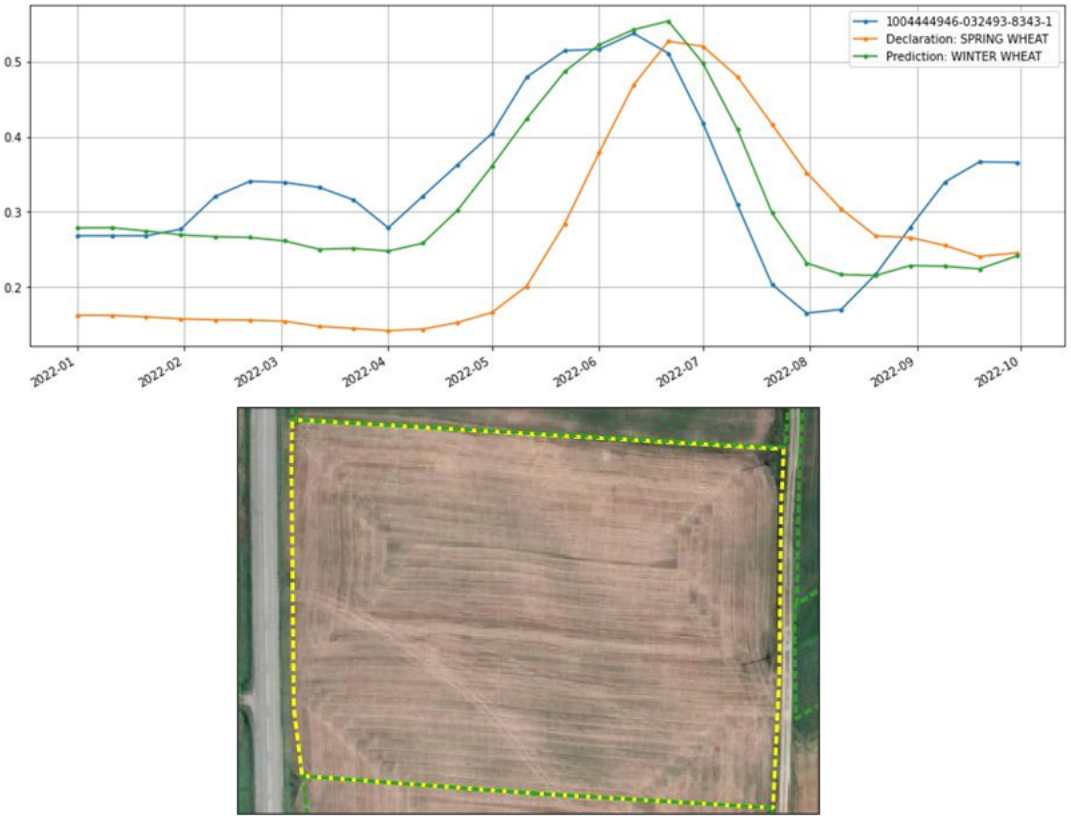


Figure 20: NDVI and S2 image of a case predicted as Winter Wheat and declared as Spring Wheat. According to this plot, spring wheat present raise of NDVI on late May. On the other hand, winter wheat presents a raise earlier during the cultivation period, which is more similar.

Discussion and limitations

The results demonstrate an exceptional level of accuracy, suggesting practical applicability. Through a semi-automatic approach (users can set their own confidence thresholds of acceptance), users can utilize the results and confidence levels provided as guidance for result extraction, facilitating the transition towards more target cases.

In the context of the smart sampling scenario, we initially established strict confidence parameters to enhance precision, yielding only a limited number of cases with exceptionally high precision. However, as we progress towards an exhaustive monitoring scenario, it becomes necessary to relax these parameters. Doing so allows us to capture a greater number of alerts, with a primary objective of maximizing recall.

1.3. DP3. Grassland Mowing Event Detection

Product Description

The Grassland Mowing Event Detection data product involves a two-step process, meticulously detailed in D3.3 and D3.7.

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

In the initial step, which focuses on reconstructing dense NDVI time series, we've undertaken a structured approach for both year of 2022 and 2023. Specifically, we've gathered pixel time series data from grassland parcels distributed within six predefined bounding boxes (i.e., study sites), strategically located across various regions of Lithuania (see Figure 21). This selection accounts for the diverse landscape characteristics present in these regions. The pixel time series have been methodically chosen to maximize the availability of cloud-free Sentinel-2 imagery, meaning they contain a minimal proportion of cloud coverage for better representation. This approach ensures a robust assessment of the capability of our S1/S2 fusion model in reconstructing NDVI over hidden timestamps within these time series. A numerical analysis on these samples is following in Table 12.

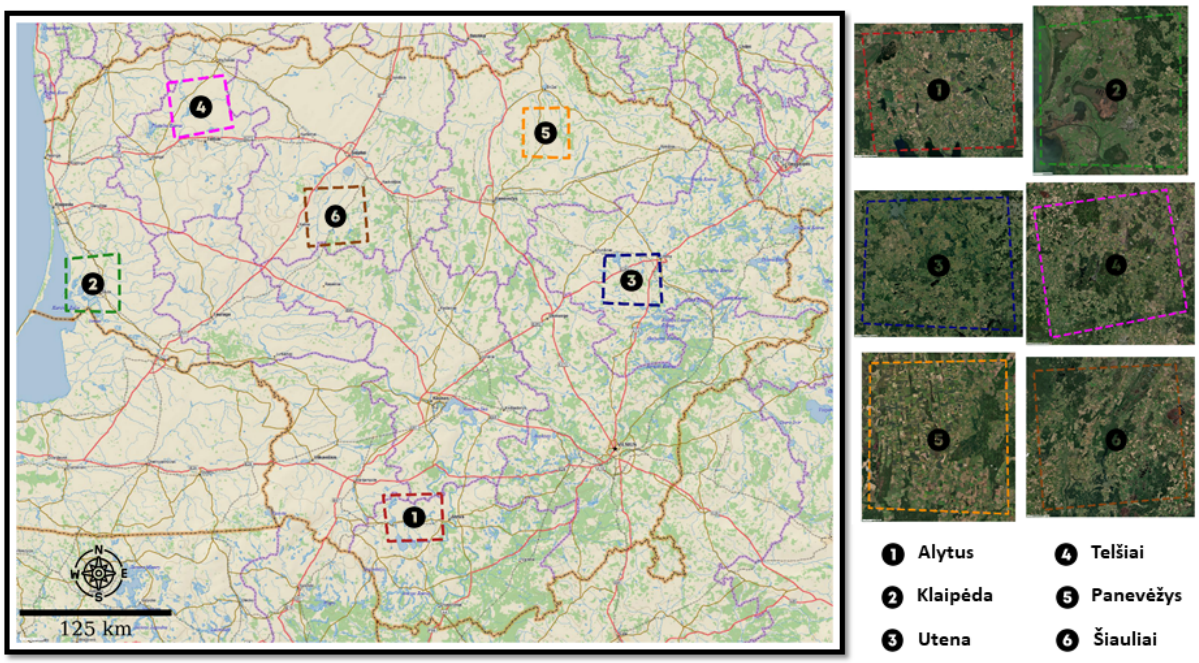


Figure 21: Map of Lithuania. The boxes correspond to the relevant regions for evaluation of S1/S2 fusion model for NDVI reconstruction.

Table 12: Numerical analysis of the samples time series collected for the validation of NDVI reconstruction for grasslands based on S1/S2 fusion model.

	2022			2023		
	# parcels	# pixels	Avg. parcel size (ha)	# parcels	# pixels	Avg. parcel size (ha)
Region 1	1211	116432	1.32	1317	198472	1.24
Region 2	1021	121543	2.01	988	98323	2.26
Region 3	1201	98065	1.68	1119	97156	1.66
Region 4	754	55043	1.77	982	64973	1.69
Region 5	712	58734	1.83	756	59004	1.89
Region 6	543	33212	1.74	621	39821	2.01
Total	5442	483029	1.70	5783	557749	1.74

In addition, to fulfill the prerequisites for the subsequent mowing detection model, NPA provided a limited number of validation cases – specifically, 197 for 2022 and 133 for 2023. These cases include

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

both instances where grassland fields were in compliance with CAP regulations and cases where they were not. Moreover, they include significant information about the number and precise dates of the actual mowing events (if they performed) for both years.
In order to enhance the quality of our assessment, NOA conducted an extensive photo-interpretation exercise involving more than 1000 grassland sample cases for both 2022 and 2023, utilizing NOA's data cube services as previously described. In this meticulous process, a blind photo-interpretation approach was employed, involving three independent experts.

Here's how the process unfolded: Two experts independently analyzed random grassland samples throughout Lithuanian territory, with a particular focus on instances featuring minimal cloud coverage in the available Sentinel-2 imagery for more objective assessment. For each mowing event detection, these experts identified an approximate timeframe, taking into account the estimated starting and finishing dates of the event based on the corresponding Sentinel-2 image acquisitions (see example Figure 22). They also assessed their confidence in the detection and estimated the percentage of the mowed area. Subsequently, a third expert evaluated the results provided by the initial two experts, and decisions were made based on the most confident of their findings. Remarkably, a high level of agreement (95%) was achieved among the experts, reinforcing the reliability of the dataset. Overall, we concluded into a dataset comprising 2,308 grassland field instances for 2022 and 1,954 for 2023, homogeneously distributed across Lithuania (see Figure 23). Tables 13 and 14 contain a numerical analysis of the validation dataset samples based on the validation method collected and the number of annotated mowing events respectively.

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

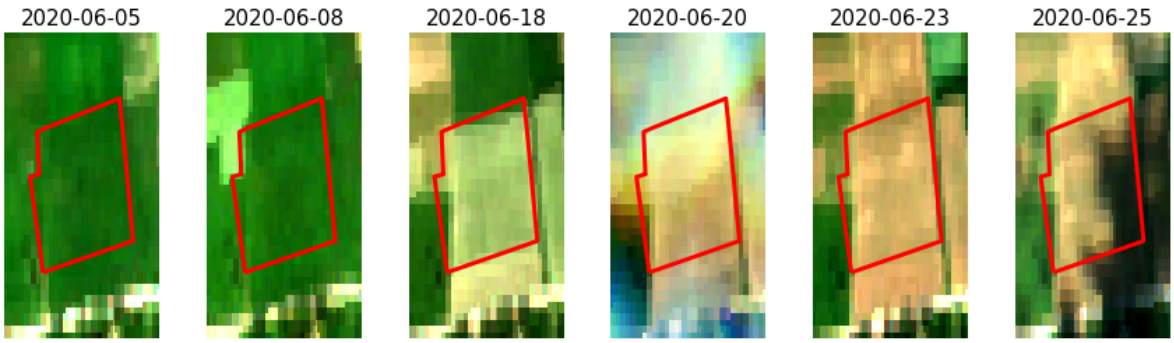


Figure 22: Example of Sentinel-2 time series images for grassland mowing events annotation. An event has been performed between 8 and 18 of June.

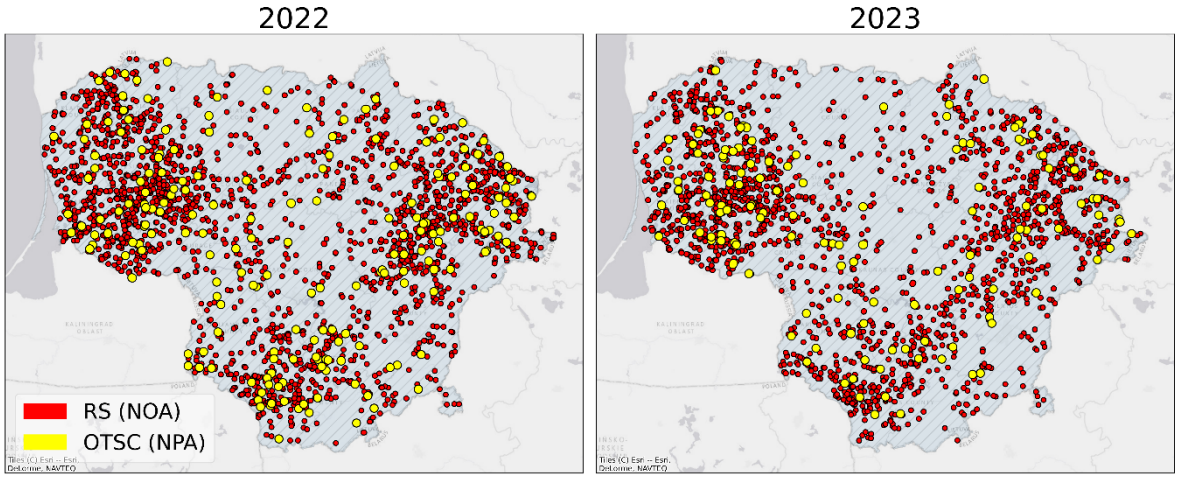


Figure 23: Grassland mowing samples geographical distribution across Lithuania territory.

Table 13: Numerical analysis of the samples collected for the validation of DP3 services in Lithuania.

	2022 (NPA)	2022 (NOA)	2022 (Total)	2023 (NPA)	2023 (NOA)	2023 (Total)
Mowing Event Detection	197	2113	**2310**	133	1822	**1955**

Table 14: Number of mowing events performed analysis.

	2022	2023
No evidence of mowing event	321	326
1 mowing event	1825	1553
2 mowing events	101	55
More than 2 mowing events	63	21

Validation Results

Data Fusion

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

To address the challenge of abrupt change detection in Lithuanian grasslands, where extensive cloud coverage frequently disrupts the continuity of Sentinel-2 optical imagery, we have employed a Deep Learning Architecture based on Recurrent Neural Networks (RNN). This architecture exploits Sentinel-1 Synthetic Aperture Radar (SAR) data, which is independent of weather conditions, in combination with cloud-free Sentinel-2 data. Our objective is to take advantage of the consistent temporal information from Sentinel-1 at the pixel level and the temporal pattern-tracking capability of RNNs to generate continuous and dense NDVI time series.

We evaluated the performance of our approach using random time steps for multiple pixel time series extracted from study sites (Region 1 to Region 6) during the years 2022 and 2023. In these evaluations, we deliberately concealed the actual NDVI values. The results, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Coefficient of Determination (R^2), are visualized in Figure 24, demonstrating the effectiveness of our S1/S2 fusion method. The methodology exhibits robust performance, with a mean MAE of 0.0258, mean MSE of 0.0015, and an R^2 value of 0.914.

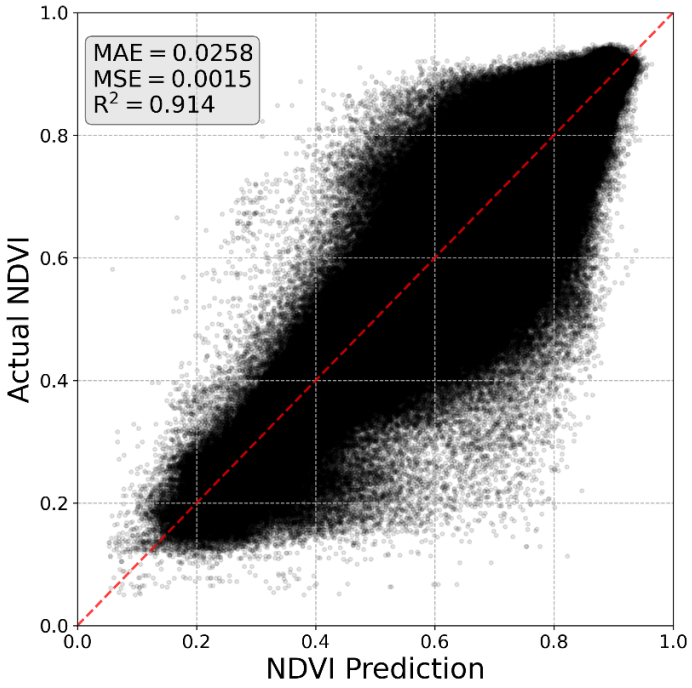


Figure 24: Scatter plot between the actual NDVI values and the SF predictions for all samples collected.

Figure 25 displays scatter plots for the six study regions. Overall, a strong correlation between the ground truth and the SF prediction is evident. Region 5 and 6 stands out with the least favorable performance, marked by an average error of 0.029 and 0.031, respectively. Nevertheless, there are no notable performance discrepancies among the other regions.

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

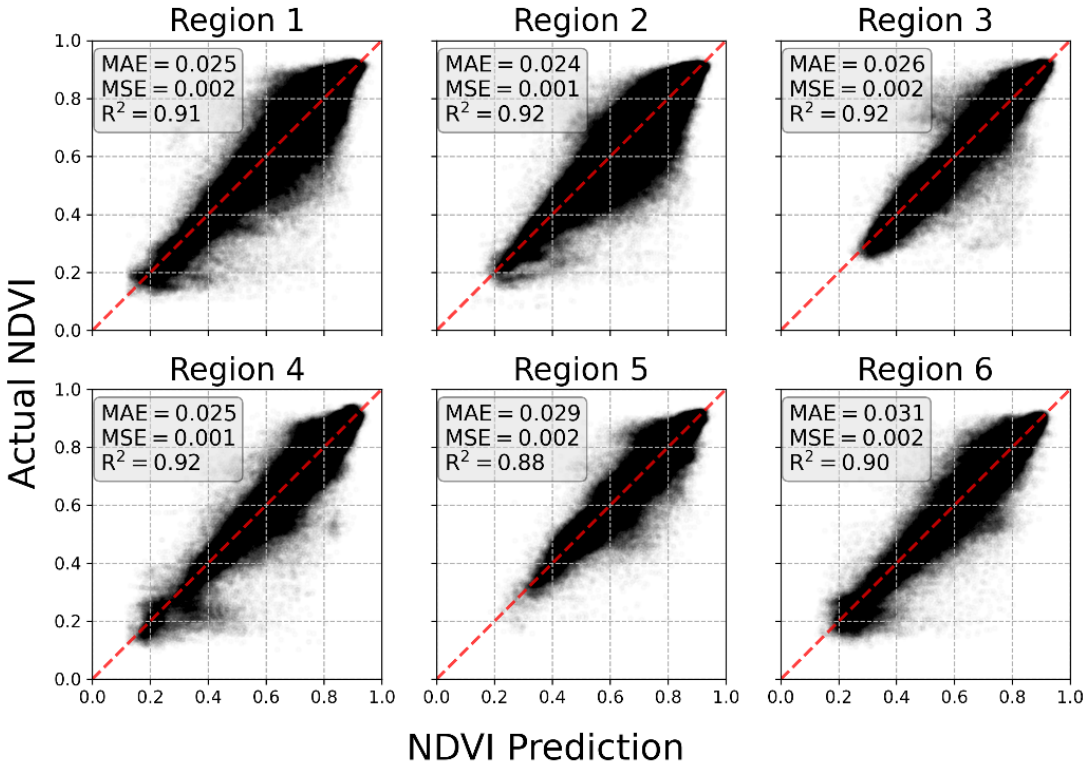


Figure 25: Scatter plots between the actual NDVI values and the SF predictions for each study region.

Figure 26 provides insight into the average and standard deviation of MAE and the average magnitude of NDVI drops on each inference date. Notably, abrupt NDVI reductions are a common occurrence from early June to mid-August, suggesting the possibility of mowing or grazing activities in Lithuanian grasslands during the summer season. Predicting these sharp NDVI declines poses a greater challenge, resulting in a relatively lower performance of the SF model in such instances.

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

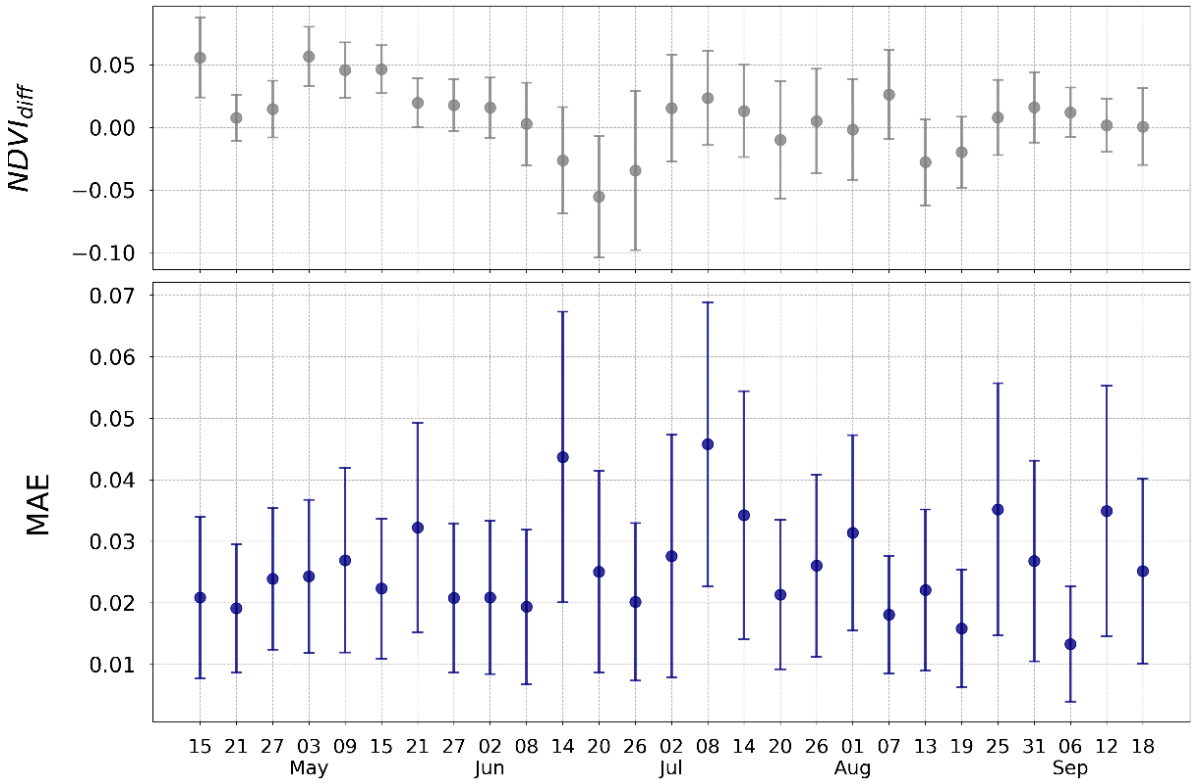


Figure 26: The upper bar plot displays the distro of NDVI drops, while the lower box plot shows the relevant values of MAE distribution on each inference date.

In Figure 27, we evaluate the model's performance under varying cloud coverage conditions (i.e., number of cloudy observations out of to the total number of available observaitons). The analysis of the results across different cloud coverage scenarios reveals that when cloud coverage is less than 50%, the SF model achieves a MAE of 0.025. However, as the number of cloudy timestamps in the time series increases, the MAE also shows an upward trend. More specifically, when comparing errors between the lowest and highest cloud coverage in the time series, there is an increase of 0.02 (rising from 0.02 to 0.04) in the mean error, while the standard deviation remains consistently around 0.015 for the SF model. This demonstrates the SF model's robustness.

In regions with substantial cloud coverage, extended gaps in Sentinel-2 data acquisitions can occur, sometimes lasting for months. While most interpolation methods are effective for short-term data gaps, which are more common, their reliability diminishes as the gaps become larger.

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

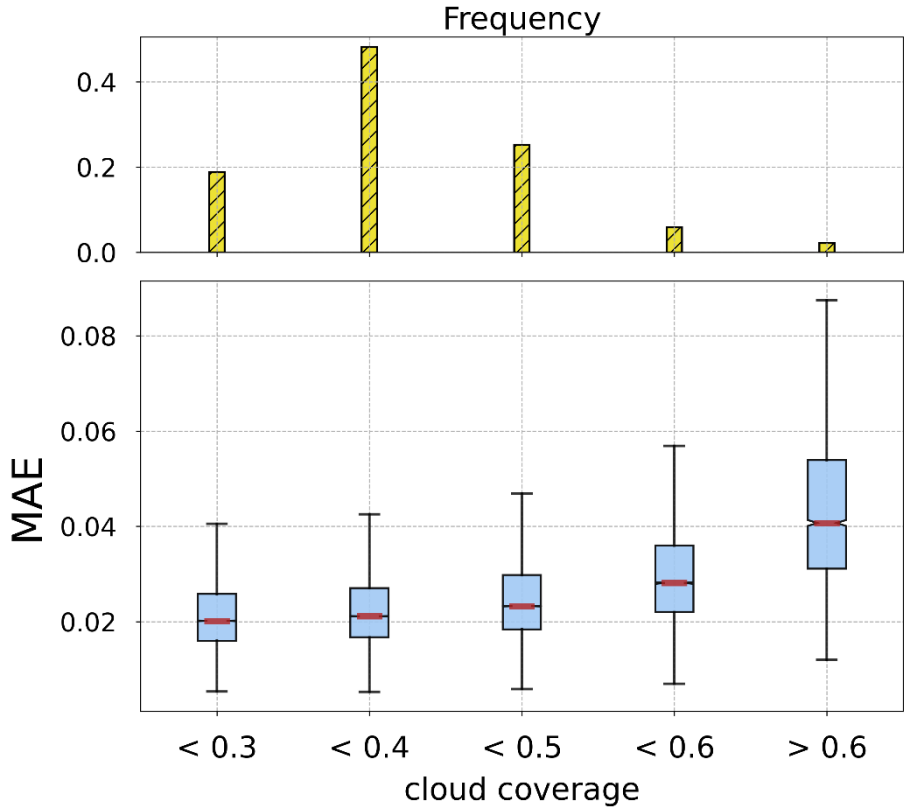


Figure 27: The upper histogram shows the frequency of each cloud coverage scenario while the low box plot shows a comparison of the MAE for the different cloud coverage scenarios.

Furthermore, when examining the distribution of MAE for varying lengths of consecutive missing values in Figure 28, it becomes evident that the SF model consistently produces stable results, even in cases where there are ten consecutive missing NDVI images. In the most challenging situations, the SF model maintains an average MAE of around 0.025, with a minimal standard deviation. As expected, less favorable results are observed in cases with gaps exceeding 7-8 timestamps (more than a month), often associated with significant NDVI reductions related to activities like mowing or grazing. Nevertheless, our method effectively reconstructs NDVI curves in many instances, as in example Figure 29. This figure illustrates a mean parcel time series constructed from initial NDVI inputs, where timestamps associated to the mowing event are hidden. The SF model effectively captures and highlights events that may not be apparent when using baseline interpolation methods.

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

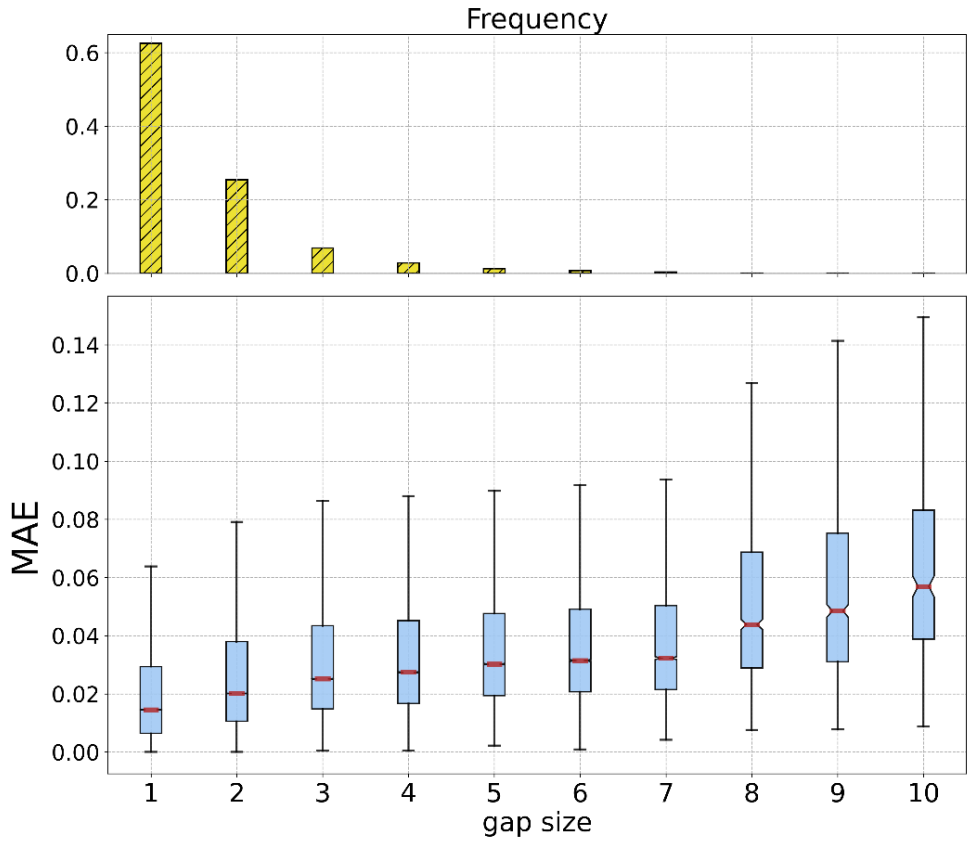


Figure 28: The upper histogram shows the frequency of the number of consecutive missing values in the grasslands' NDVI time series while the low box plot shows a comparison of the MAE for the different number of consecutive missing values (gap size).

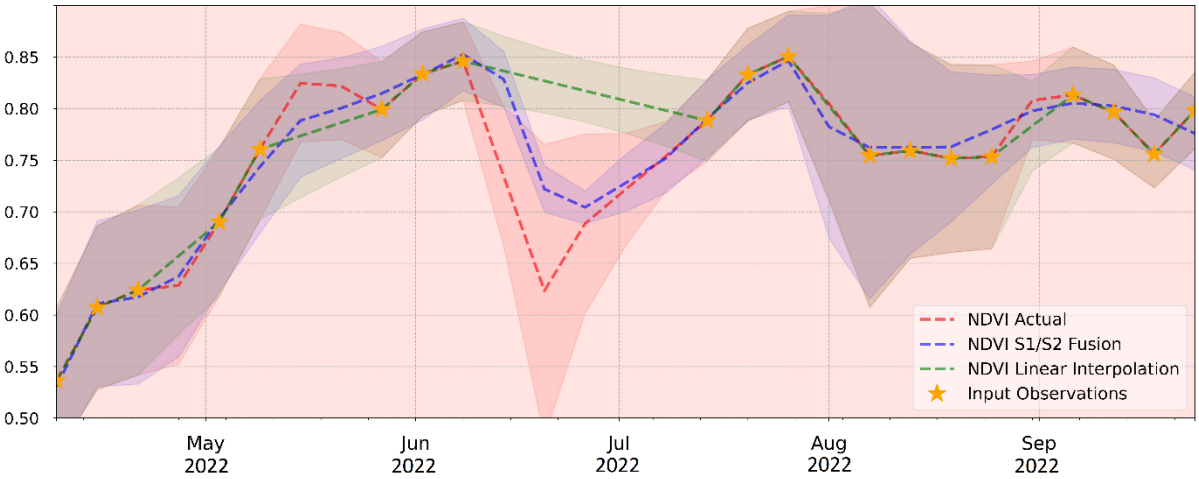


Figure 29: NDVI reconstruction example related to a hidden mowing event. Stars show input NDVI values, the blue line represents the SF predictions, and the red line shows the actual. Green line shows the results predicted used linear interpolation based on star inputs.

Grassland Mowing Event Detection

We have developed a deep learning architecture for detecting mowing events. This model uses an RNN to analyze time series data comprising newly generated NDVI values and corresponding S1 backscatter coefficients. The aim is to pinpoint the 6-day window (see example Figure 30 below) during which a

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

mowing event occurred. By applying relevant mowing regulations based on the grassland type, we can deduce whether a mowing event took place, typically before the end of August. This analysis can help assess farmers' compliance.

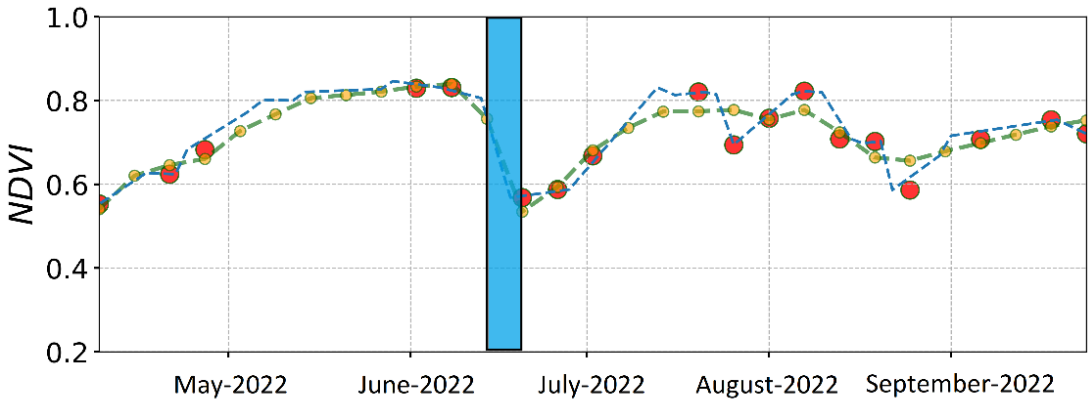


Figure 30: Mowing Event Detected as result of sudden NDVI drop.

The model's performance with a total accuracy approximately of 98% for both years, as demonstrated in Table 15, which showcases its recall and precision for both mowed and non-mowed cases. The results indicate that the model effectively addresses scenarios involving compliant and non-compliant farmers.

Furthermore, Table 16 offers a more detailed analysis by categorizing the total number of events. This fine-grained examination reveals the model's adaptability across various event scenarios. It's worth noting a slight decline in performance when dealing with more than two total events. Nevertheless, this minor decrease doesn't impact the ultimate determination of a farmer's compliance. The model only needs to identify at least one event to make this assessment.

Table 15: Analysis on recall, precision and support of the mowing event detection analyzed for mowed or not mowed cases.

Cultivation Period	2022			2023		
Condition	Precision	Recall	Support	Precision	Recall	Support
Non-mowed	0.987	0.941	321	0.985	0.988	326
Mowed	0.976	0.983	1989	0.986	0.999	1629
Total Accuracy	**0.977**		**2310**	**0.986**		**1955**

Table 16: Analysis on recall, precision and support of the mowing event detection analyzed for the different number of total mowing events performed.

Cultivation Period		2022			2023		
Dataset	Condition	Precision	Recall	Support	Precision	Recall	Support
NPA	No Evidence of Mowing Events	**0.941**	**0.842**	19	0.800	0.727	11

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

	1 Mowing Event	**0.925**	**0.980**	152	0.953	0.990	102
	2 Mowing Events	**0.875**	**0.636**	22	0.706	0.600	20
	More than 2 Mowing Events	**1.000**	**0.750**	4	-	-	0
NOA	No Evidence of Mowing Events	**0.990**	**0.947**	302	0.991	0.997	315
	1 Mowing Event	**0.989**	**0.998**	1673	0.995	0.999	1451
	2 Mowing Events	**0.877**	**0.899**	79	0.969	0.886	35
	More than 2 Mowing Events	**0.907**	**0.831**	59	0.813	0.619	21

Furthermore, in Figure 31, the scatter plot reveals a strong correlation between predicted and reference dates, which are expressed in Days of the Year. The high coefficient of determination (R^2 = 0.95) and low Mean Absolute Error (MAE of 2.58 days) indicate the model's ability to accurately identify mowing events. Specifically, the majority of mowing events fall within a 12-day range, equivalent to the time gap between two consecutive Sentinel acquisitions. This proximity is primarily attributed to the subsequent time shifting of the new exported NDVI measurements, as previously described in the data fusion process.

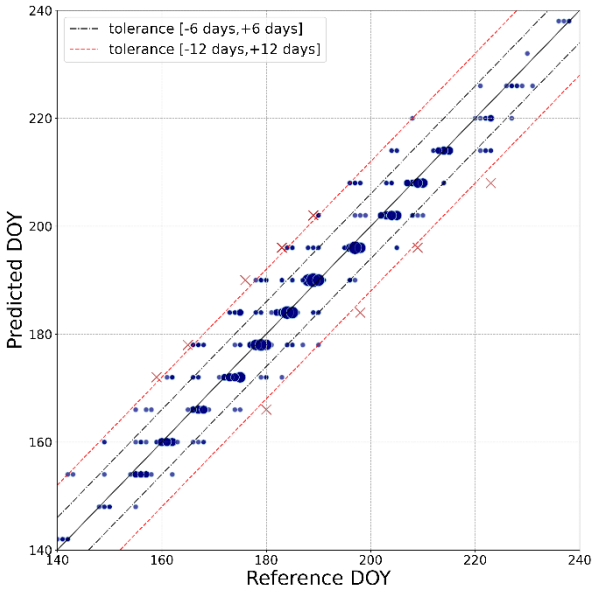


Figure 31: Reference day of the year (DOY) for the mowing events and the DOY predicted by the model.

Evaluating the impact of cloud coverage on the model's performance is essential. As highlighted in the data fusion section, mowing events can be obscured by extended cloud cover, potentially leading to inaccurate notifications of farmer non-compliance. In Figure 32, we examine the model's robustness

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

across varying levels of cloud coverage. The results demonstrate that our model's performance remains consistent, even in the presence of extensive cloud coverage. The proportion of accurately identified mowing cases remains stable, even in the most extreme scenarios. This robustness signifies that our model can effectively handle cloud-related challenges, ensuring reliable results for users, regardless of cloud conditions.

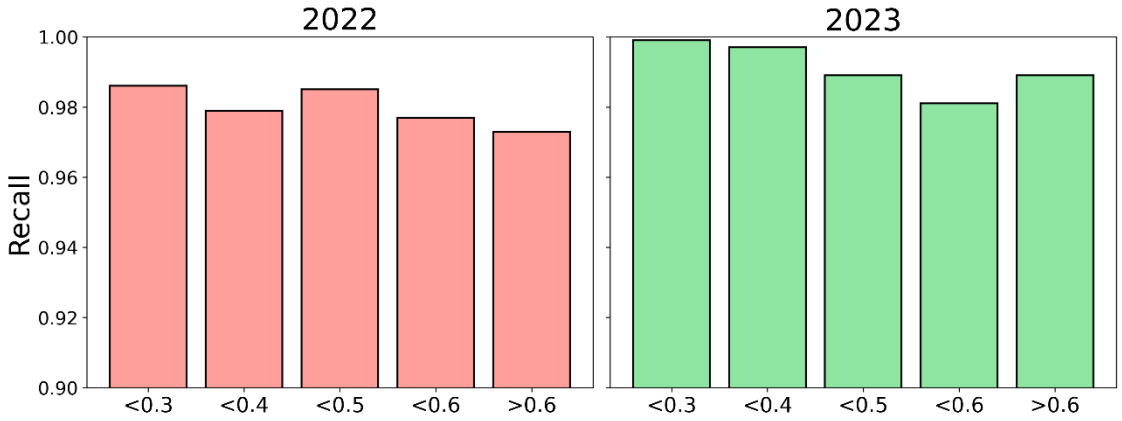


Figure 32: Recall performance on different cloud coverage scenarios. Cloud coverage is calculated as the ratio of total cloudy timestamps to the total number of timestamps available.

Additionally, parcel size is a potentially crucial factor to consider in our analysis As described in D3.7, results are provided initially, on each pixel individually and aggregated statistics is used to provide us with a representative level of confidence regarding the extent and the exact time instance that a mowing event took place for each parcel. Figure 33 illustrates the model's capability to correctly identify actual events based on different parcel sizes. Notably, the model's performance remains consistently stable across all scenarios, yielding optimal results, especially in cases with larger parcels. This stable performance across varying parcel sizes aligns with our expectations, as the total number of the available pixels becomes higher.

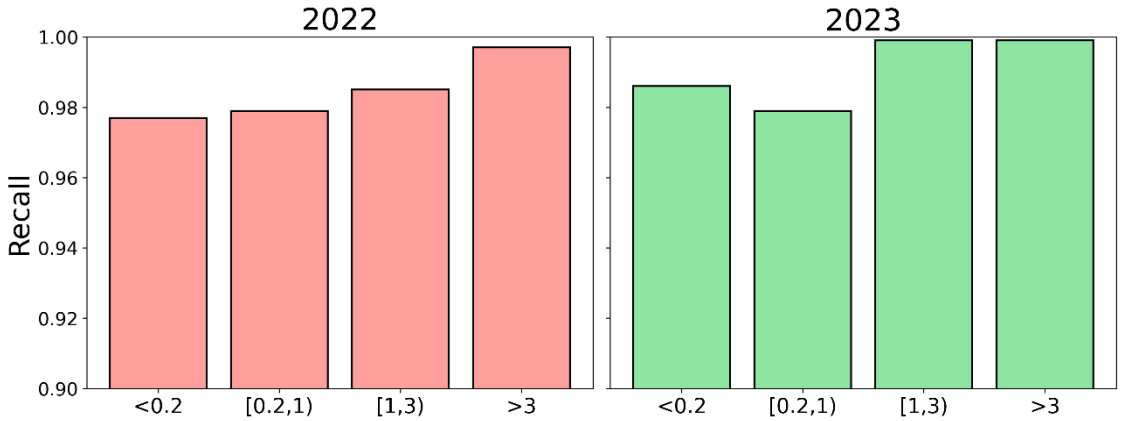


Figure 33: Recall performance on different parcel size scenarios. Area is calculated in herctares.

Lighthouse Customers: Case of Flanders

The grassland mowing event detection algorithm was developed and applied in a pilot sub-area of Flanders for the 2022 growing season for 107,726 total cases, using GSAA farmer's declarations provided by LV. Since there were no training labels available for mowing events, the methodology

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

involved a data fusion step for NDVI reconstruction and a subsequent threshold-based event detection approach, as detailed in BC.1. of D3.7.

In the initial phase, where the primary focus was on reconstructing dense NDVI time series, we followed a structured approach similar to the one used in Lithuania. This involved transfer learning, where we fine-tuned a pre-existing model on the NDVI time series data from two evaluation regions within Flanders (see Figure 34).

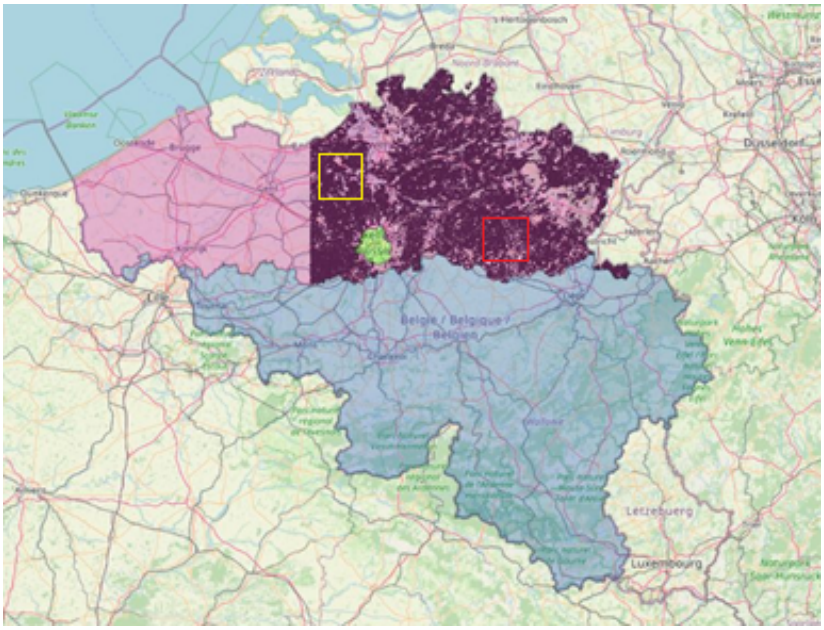


Figure 34: "Grassland Mowing Event Detection" product applied for the area of Flanders (pink layer) in Belgium for 2022. Test areas are extracted from two evaluation sites, Region 1 (yellow colour) and Region 2 (red colour).

Figure 35 displays scatter plots for the two study regions. Overall, a strong correlation between the ground truth and the SF prediction is evident, similar to Lithuanian standards.

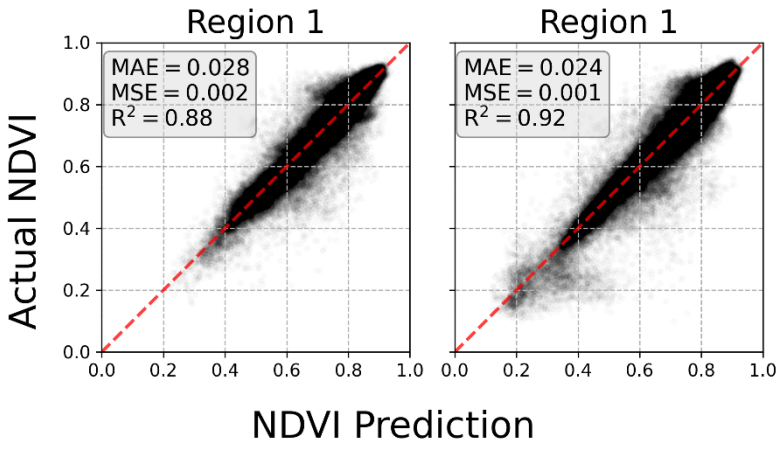


Figure 35: Scatter plots between the actual NDVI values and the SF predictions for the two study regions in Flanders.

To assess mowing event detection in Flanders, NOA conducted a photo-interpretation evaluation since LV did not provide their own validations. NOA's evaluation included 1448 sample grassland cases from

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

the two regions, utilizing again NOA's data cube services. This meticulous process employed a blind photo-interpretation approach, involving three independent experts, similar to the approach used in Lithuania. The results of this evaluation are summarized in Table 17 for the two regions, respectively.

Table 17: Analysis on recall, precision and support of the mowing event detection analyzed for the two study regions.

Region	**Region 1**			**Region 2**		
Condition	Precision	Recall	Support	Precision	Recall	Support
No Evidence of Mowing Events	0.922	0.884	121	0.937	0.908	98
1 Mowing Event	0.959	0.972	386	0.921	0.930	302
2 Mowing Events	0.936	0.940	248	0.910	0.922	231
More than 2 Mowing Events	0.818	0.783	23	0.778	0.718	39
Total Accuracy	**0.942**		**778**	**0.912**		**670**

The results in Flanders are comparable to those in Lithuania, although slightly less favorable. It's worth noting that the accuracy of the results could potentially be improved by providing training labels for mowing events. However, generating a critical sample of over 10,000 cases (as expected) for such training is a time-consuming process. The threshold-based routine relies on predefined NDVI thresholds, and Table 18 showcases the top five accuracy scores achieved for the two regions using these specific threshold values.

Table 18: Top-5 set of parameters for threshold-based mowing event detection in Flanders.

set	NDVI_th	NDVI_r	Accuracy – Region 1 (%)	Accuracy – Region 2 (%)
1	**-0.08**	**0.001**	**94.2**	**91.2**
2	-0.10	0.001	94.4	89.9
3	-0.12	0.001	91.9	90.8
4	-0.08	0.005	88.7	89.1
5	-0.06	0.001	88.7	86.3

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

Discussion and limitations

The Grassland Mowing Detection product in Lithuania has demonstrated remarkable performance and robustness in monitoring grasslands. The algorithm displays notable sensitivity to critical bottleneck factors like parcel size and cloud cover, showcasing a full automisation and adaptability across diverse parcel sizes and resilience in dealing with varying cloud coverage conditions.

The anomaly related to Sentinel-1B, which led to reduced SAR temporal resolution for 6 to 12 days, is expected to have some impact. Nevertheless, the algorithm's outstanding performance suggests that the effect may be less severe than initially anticipated (also discussed in D3.7).

However, certain limitations should be considered. Extensive training data is essential (especially for the part of NDVI reconstruction using the S1/S2 Fusion model), and implementing the algorithm at a national scale demands substantial computational resources, especially with a pixel-based approach. It's vital to be cautious though, when working with exceptionally large training datasets, as they can increase the risk of overfitting. To mitigate the risk of absence of training data, we propose employing threshold-based approaches, similar to SEN4CAP project, sufficiently implemented for the lighthouse customer case of Flanders. To address the computational challenges, we recommend adopting a pixel-based approach only for small parcels (e.g., less than 1 hectare), which is critical for algorithm precision. For larger parcels, an approach utilizing average time-series data can be employed, optimizing computational efficiency while maintaining accuracy.

2. BC2: Monitoring multiple environmental and climate requirements of CAP – Cyprus

This section outlines the validation dataset in order to evaluate the outputs provided by the respective data products (DP1 and DP2) applied in case of Cyprus (BC2). Detailed information on the methodology of the respective algorithms developed is provided in D3.7.

2.1. DP1. Analytics on Vegetation and Soil-Index Time-series

This data product is designed to analyze time-series data related to vegetation and soil indices in Lithuania. It offers several algorithmic components:

- Minimum Soil Cover for Soil Erosion: This feature provides data on soil percentage and minimum soil cover, which helps assess the risk of soil erosion in different regions of Cyprus, particularly in agricultural areas. It can assist in making informed land management decisions to prevent soil erosion.
- Runoff Risk Assessment for the Reduction of Water Pollution in Nitrate Vulnerable Areas: This component assesses the risk of runoff and water pollution in nitrate vulnerable areas of Cyprus. It can be valuable in managing and mitigating water pollution, especially in regions with intense agricultural activities.

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

- Detection of illegal land clearing in Natura2000 protection areas: This feature is designed to identify unauthorized agricultural activities within Natura 2000 regions, ensuring the preservation of their ecological integrity and facilitating the monitoring of potential violations.
- Stubble Burning Identification: This component is designed to detect and identify instances of stubble burning, which can be a concern for air quality and environmental impact. It can help monitor and enforce regulations related to stubble burning practices.

Sampling Description

Minimum oil cover for soil erosion

This service promotes the adoption of minimum soil cover practices to prevent erosion. The algorithm assesses soil percentages on areas with slopes exceeding 10%, similar to the approach used in Lithuania. To evaluate its effectiveness, NOA conducted photo-interpretations for the years 2022 and 2023, during which the module was applied. These assessments involved a thorough examination of cloud-free Sentinel-2 images using Creodias' datacube services, prior to the main cultivation season (until March). These instances were evenly distributed across Cyprus, guided by the system's alerts, and covered 1479 out of 21761 cases with slopes greater than 10% in 2022. In 2023, a total of 1346 alerts for minimum soil cover violations were identified out of 21853 cases with slopes exceeding 10% (see Figure 39).

Runoff risk assessment for the reduction of water pollution in nitrate vulnerable areas

Under CAP regulations, it is imperative to avoid the application of manure and/or slurry in the coastal protection zones around water bodies as delineated in the Surface Water Protection Zone layer. In response, we have devised a runoff risk assessment procedure that considers the proximity of each agricultural parcel to the nearest water surfaces inside Nitrate Vulnerable zones. Our assessment relies on data sourced from Cyprus hydrographic network and nitrate vulnerable areas, generously provided by the CAPO, as illustrated in Figure 36. The service was deployed for the whole Cyprus for the year 2022.

Additionally, it's worth mentioning that the output generated for this service is of a qualitative nature. In other words, our validation process primarily focuses on **qualitative assessments** rather than quantitative metrics due to the inherently conductive nature of the service. This is a risk assessment; as such it cannot be measured directly for the respective paying agencies. Overall, this approach ensures that we maintain a high level of compliance with CAP regulations in the pilot area while providing valuable insights into the effectiveness of our runoff risk assessment procedure.

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

Figure 36: Hydrographic network of Cyprus (yellow colour) and Nitrate Vulnerable Areas (purple colour).

Stubble burning identification on arable land

The identification of stubble burning serves the specific purpose of meeting CAP requirements related to the burning of agricultural residues, as illustrated in example Figure 37 below. However, CAPO did not provide us with validation data for burn parcels in recent years. To overcome this limitation, we resorted to optical methods, employing photo-interpretation with datacube services. During this process, we closely examined a manageable total of 220 cases of indicated burning in 2022 and 295 cases in 2023 (see Figure 39). Our team meticulously reviewed these instances, relying on available Sentinel-2 images to determine whether stubble burning had indeed taken place. This additional step became necessary due to the absence of customary validation data.

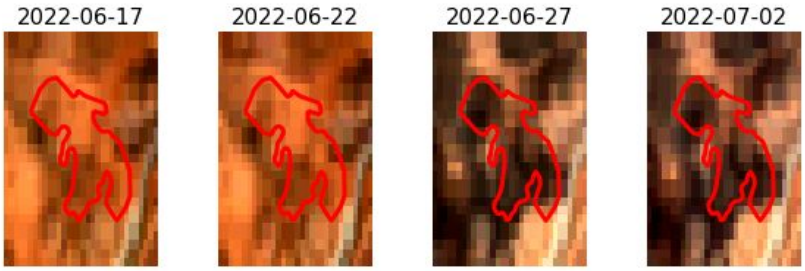


Figure 37: Example case of stubble burning in arable crops in Cyprus.

Detection of illegal land clearing in Natura2000 protection areas

Natura 2000 is a protected area network in the European Union aimed at preserving Europe's endangered species and habitats. In Cyprus, agricultural activity is generally prohibited within these areas (see Figure 38), except with special permission. To detect intense activity within Natura 2000 regions, a pixel-based routine has been developed that evaluates several vegetation and soil indices. To ensure policy compliance, Eligible Agricultural Areas from the LPIS are excluded from the analysis within Natura 2000 sites. This helps distinguish authorized interventions from potentially unauthorized ones.

It's worth mentioning that the output generated for this service is of a qualitative nature. In other words, our validation process primarily focuses on **qualitative assessments** rather than quantitative metrics due to the inherently conductive nature of the service.

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

Figure 38: Natura 2000 network sites in Cyprus.

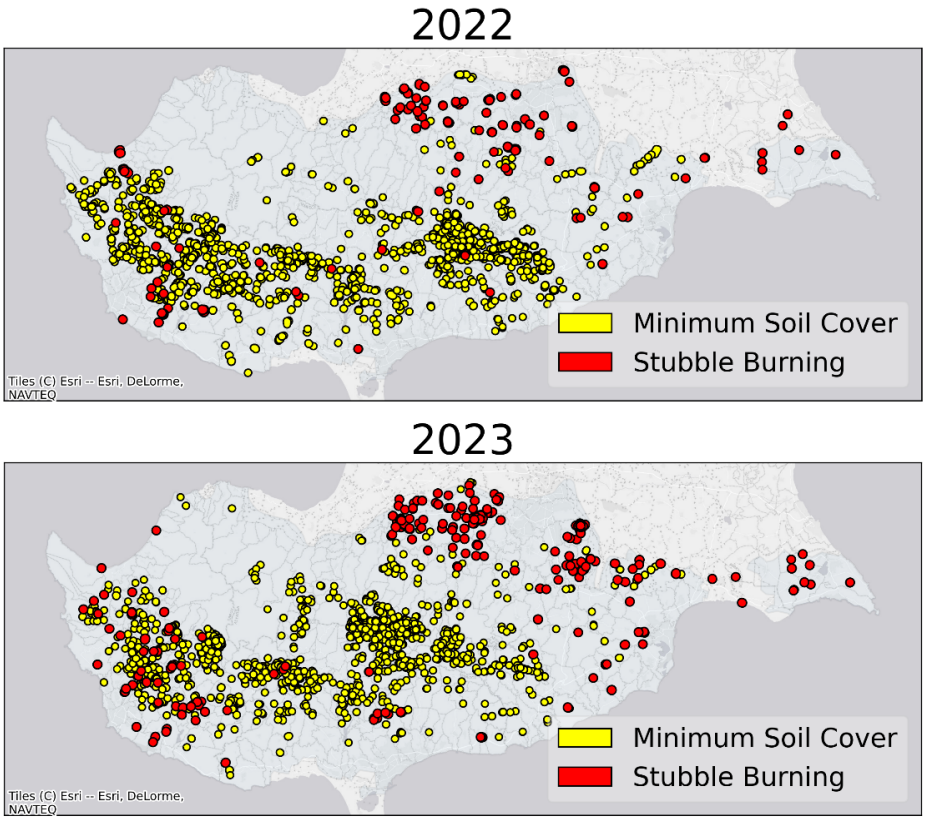


Figure 39: Alert cases for Minimum Soil Cover and Stubble Burning across Cyprus for DP1.

Validation Results

Minimum soil cover for soil erosion

The minimum soil cover detection algorithm was applied in both the 2022 and 2023 cultivation periods for parcels with slopes over 10%. A binary mask is created based on the following conditions, as outlined in D3.7:

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

i. $ndvi\ lower < ndvi < ndvi\ upper$

ii. $savi\ lower < savi < savi\ upper$

iii. $B2 - B1 > (B2 - B1)\ lower$

iv. $B3 - B2 > (B3 - B2)\ lower$

v. $NBR2 < NBR2\ upper$

To classify an area as bare ground from satellite observations, a minimum of percentage of clear pixels must indicate the presence of soil, even if vegetation is still present. Table 19 displays the top-5 threshold parameter combinations in descending order based on precision accuracy. By using the best combination, we achieved an accuracy of 88.9% for 2022 and 83.2% for 2023.

Table 19: Top-5 set of parameters for minimum soil cover in for 2022 and 2023.

	NDVI LOWER	NDVI UPPER	SAVI LOWER	SAVI UPPER	B2-B1	B3-B2	NBR2	CLEAR PIXELS (%)	PRECISION 2022	PRECISION 2023
1	**0**	**0.25**	**0**	**0.35**	**0**	**0**	**0.4**	**20**	**88.9**	**83.2**
2	0	0.3	0	0.35	0	0	0.4	20	86.3	82.1
3	0	0.25	0	0.4	0	0	0.35	20	86.1	82.0
4	0	0.25	0	0.3	0	0	0.35	20	82.0	79.9
5	0	0.15	0	0.3	0	0	0.35	20	77.4	78.9

Run-off Risk Assessment

The run off risk assessment algorithm takes into account the parcel's proximity to water surfaces. The methodology here is the exact same with BC of Lithuania. By taking into consideration the values of water proximity and RUSLE, runoff risk has been computed for each parcel. A numerical description of the results is presented in Table 20. In addition, Figure 40 depicts the visualization of the parcels along with their categories and the water surfaces around them. Both layers' data is directly retrieved from the ENVISION database.

Table 20: Run-off Risk Assessment Results.

	Very Low or Low Risk	Moderate Risk	Very High or High Risk
Algorithm	43562	40834	1035

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

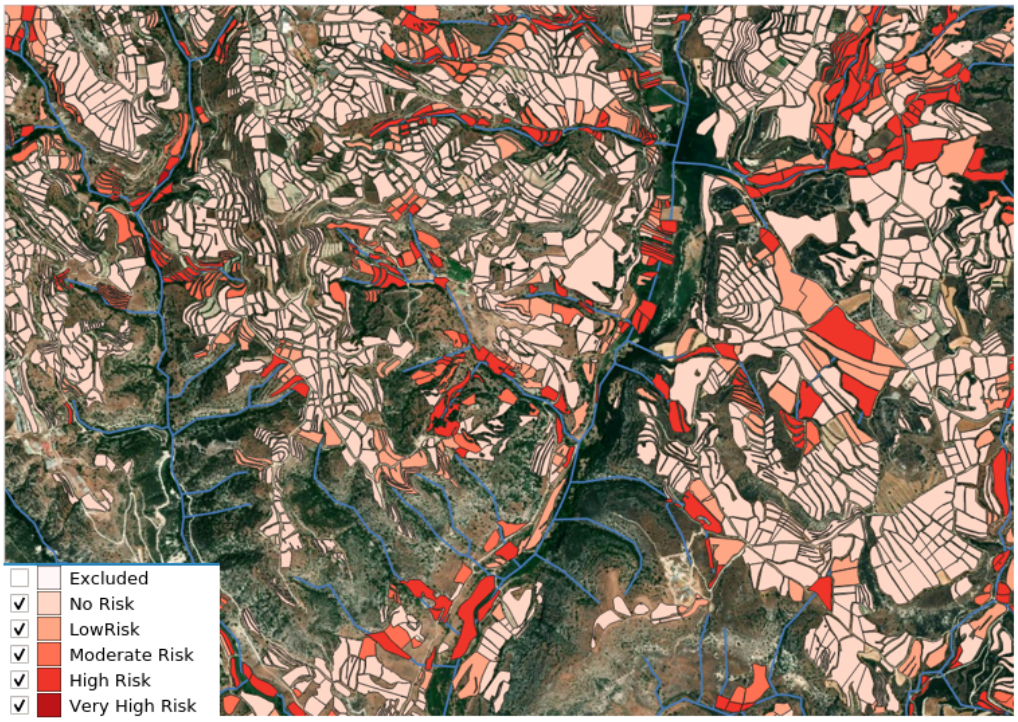


Figure 40: Visualization of the run-off risk for a subset of parcels along with the water surfaces around them.

Stubble Burning Identification

The algorithm was used for both the 2022 and 2023 cultivation periods. Following a quality assessment using Sentinel-2 imagery, an exceptional level of accuracy was achieved: 91.9% for 2022 and 94.2% for 2023. Notably, a significant portion of the detected cases can be attributed to wildfires, that were relatively easy to be identified.

Table 21 lists the top-5 threshold parameter combinations derived from various vegetation and soil indices, ordered by precision accuracy. Furthermore, an illustrative example of stubble burning detected by the algorithm is provided in Figure 41.

Table 21: Top-5 set of parameters for stubble burning for 2022 and 2023.

	NBR UPPER	SAVI UPPER	NDMI UPPER	PSRI UPPER	NDWI LOWER	BSI LOWER	PRECISION 2022	PRECISION 2023
1	**-0.05**	**0.15**	**-0.05**	**0.25**	**-0.25**	**0.1**	**91.9**	**94.2**
2	-0.05	0.15	-0.1	0.25	-0.25	0.05	85.5	90.7
3	-0.1	0.15	-0.05	0.25	-0.2	0.05	84.7	88.9
4	-0.1	0.2	-0.05	0.25	-0.2	0.05	80.6	85.1
5	-0.05	0.15	-0.1	0.25	-0.2	0.1	80.6	83.8

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

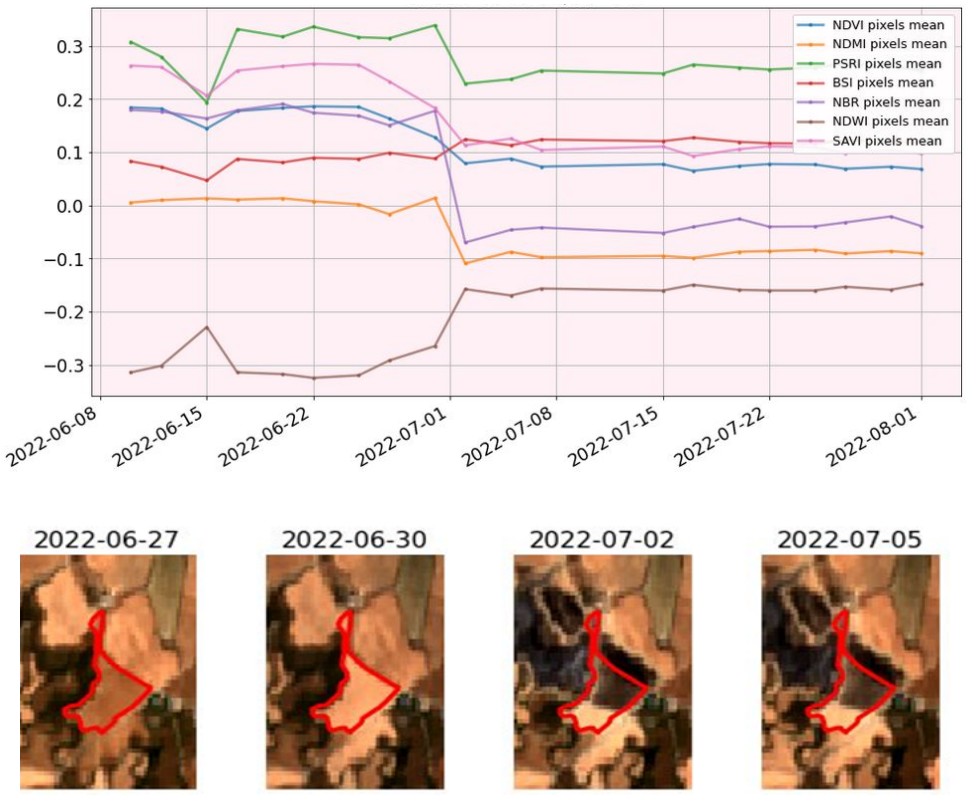


Figure 41: Visual representation of stubble burning event successfully identified by the algorithm between 30 June and 2 July.

Detection of illegal land clearing in Natura2000 protection areas

The Natura 2000 Hotspot Detection algorithm employs a threshold-based approach, as described in D3.7. While the absence of validation data limits precise evaluation, our manual inspection of predictions via photointerpretation reveals strong performance in identifying relevant cases. However, the majority of the detected events require further evaluation through CAPO's verification mechanisms, including both authorized and unauthorized activities within Natura 2000 regions.

Figure 42: An example of an area where illegal land clearing occurred in 2022 and correctly identified from the algorithm, depicting the situation before and after the illegal clearing.

It's worth noting that we utilize the same methodology as the Lithuanian BC (BC1) for detecting harvest events at the level of pixel, integrating the analysis of various vegetation and soil indices (e.g., NDVI,

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

NDMI, PSRI, and BSI) over time. In Table 22, we present the selected parameters aimed at achieving results of high quality, minimizing the inclusion of noisy false alerts (primarily caused by seasonal vegetation changes) while capturing a substantial number of genuine alert cases. The final output (see Figure 43) serves as an advisory tool to assist control authorities in identifying potential illegal activities.

Table 22: Critical parameters used for Detection of illegal land clearing in Natura2000 protection areas in 2022 and 2023.

	NDVI	NDMI	PSRI	BSI
1	0.15	0.2	0.1	-0.1

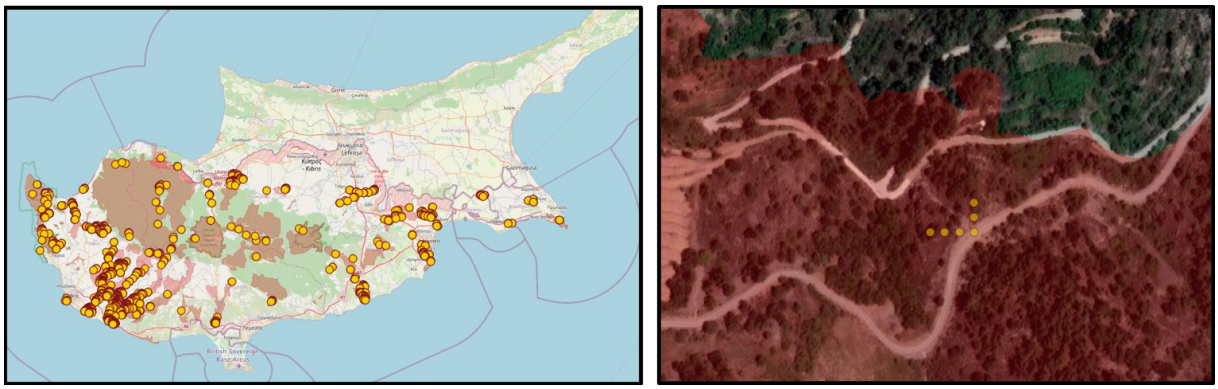


Figure 43: Cyprus Natura2000 Alert Pixels Detected example for 2022. The identification of alert pixels, signals potential instances of unauthorized clearing activities.

Discussion and limitations

The algorithmic components within DP1 exhibit satisfactory overall accuracy, making them suitable for integration into in-house infrastructures. Notably, the Stubble Burning and Minimum Soil Cover detection algorithms display the best performance.

However, it is crucial to acknowledge certain limitations.

First, the evaluation of Stubble Burning and Minimum Soil Cover relied on qualitative assessment through NOA's photo-interpretation. CAPO experts provided only a qualitative picture of the outcomes as they are using them for their current operation, already from 2022. The guidance by the algorithms outcomes was satisfactory according to their declaration

The Run-off Risk Assessment for Water Pollution Reduction in Nitrate Vulnerable Areas evaluation is qualitative as it operates as a risk assessment algorithm. It identifies high-risk areas based on proximity to water bodies and soil characteristics. Water pollution is primarily influenced by farming practices, and quantitative measurements are not feasible within this framework. Nevertheless, it can guide control bodies on the strategic monitoring of farming practices and regulatory compliance.

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

Finally, the evaluation of the Detection of illegal land clearing in Natura 2000 protection areas algorithm relied solely on qualitative photo interpretations by NOA. This pixel-based algorithm was exhaustively applied to all Natura2000 zones and fine-tuned accordingly to reduce false-positive indications through parameter configuration. Additionally, to mitigate further noise from the intense variation of the vegetation in large forestry regions, the analysis concentrated on assessing only the boundaries of these areas, reducing the impact on intensive activity detection. The final assessment of detected cases requires further verification by CAPO to distinguish authorized from unauthorized activities within Natura 2000 regions. Nonetheless, the algorithm successfully identified multiple instances of illegal activity.

All in all, the modules mentioned rely on manually optimized parameters to operate effectively. These parameters, often fine-tuned through human expertise, are crucial in influencing the modules' behavior and performance. Through careful adjustments, operators can enhance the modules' capabilities. While manual optimization can be a time-intensive process, it plays a significant role in achieving desired outcomes for the implementation of the respective services to other regions. This human touch, guided by experience and domain knowledge, can lead to finely tuned systems that operate with precision and efficiency.

2.2. DP2. Cultivated Crop Type Maps (CCTM)

Product Description

To evaluate the cultivated crop type maps produced, we employ a rigorous validation dataset supplied by CAPO. This dataset comprises 12,681 instances for 2022 and 4,563 for 2023, gathered nationwide during ENVISION deployment. The dataset is meticulously curated to ensure a diverse representation of crop types (see Figure 44). It combines two methods: utilizing Sentinel-2 satellite data for remote sensing and expert on-site field visits for validation. These samples encompass varying field sizes, covering diverse farming landscapes and uniformly distributed samples across Cyprus (see Figure 47).

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

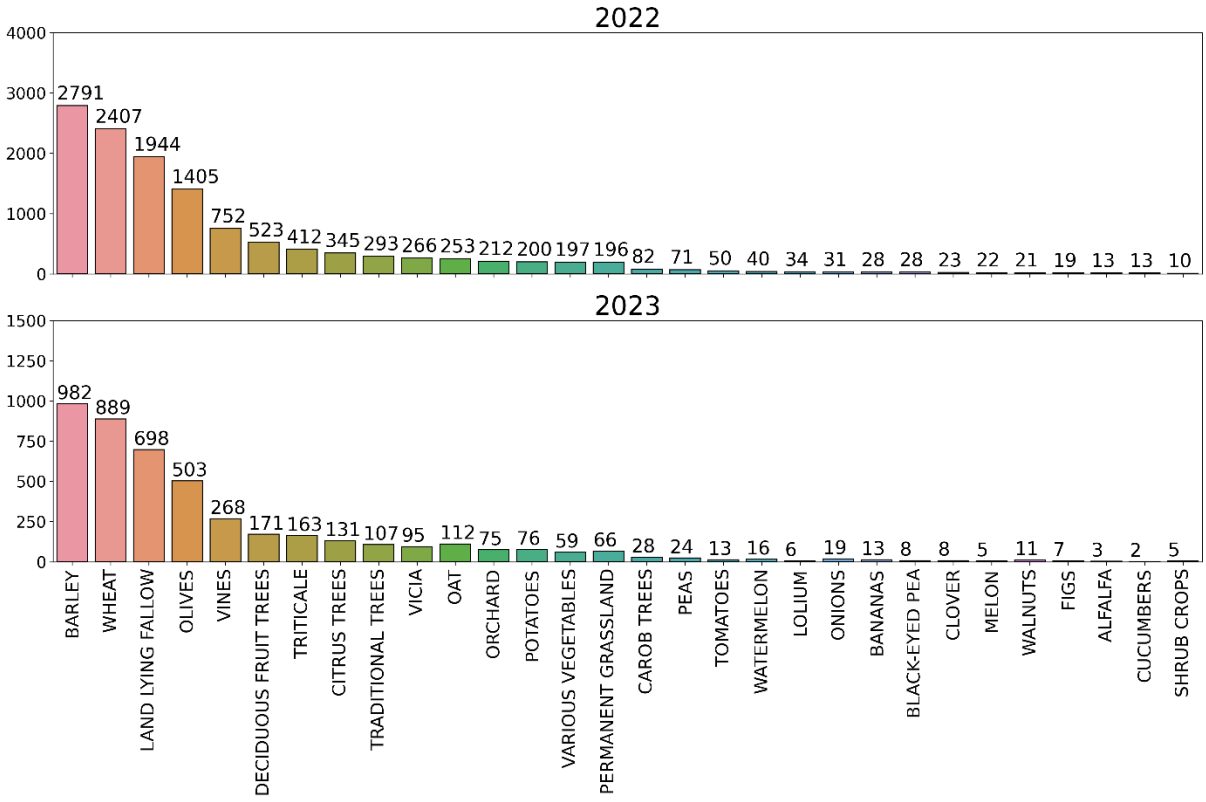


Figure 44: Crops distribution of validation samples for CCTM for 2022 and 2023.

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

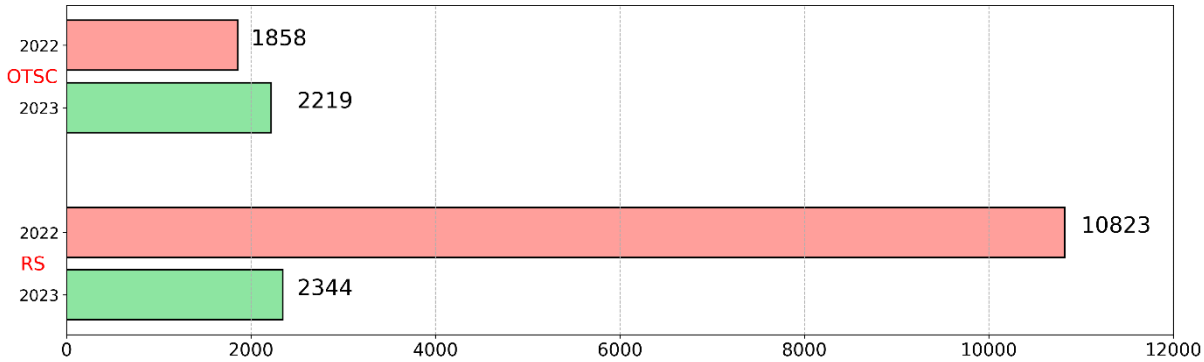


Figure 45: Number of available OTSC and RS validation for both operational year.

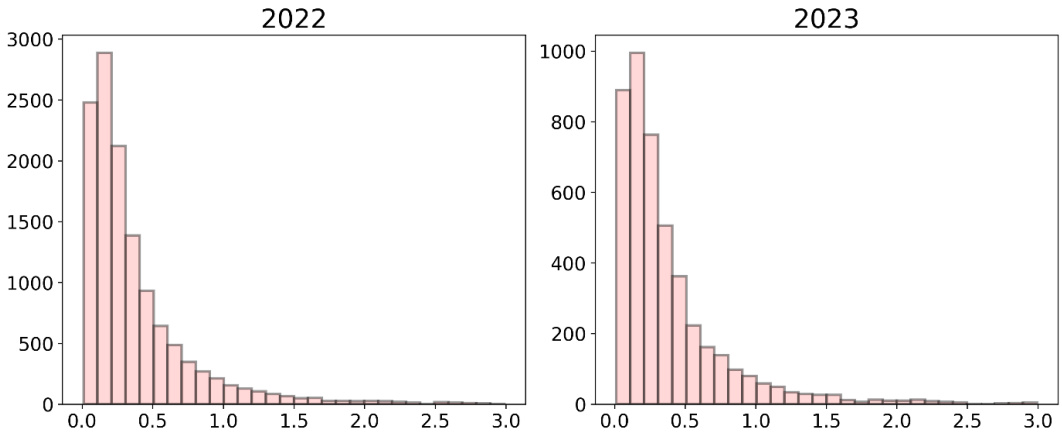


Figure 46: Distribution of parcel size of the CCTM validation. Area is calculated in hectares (ha).

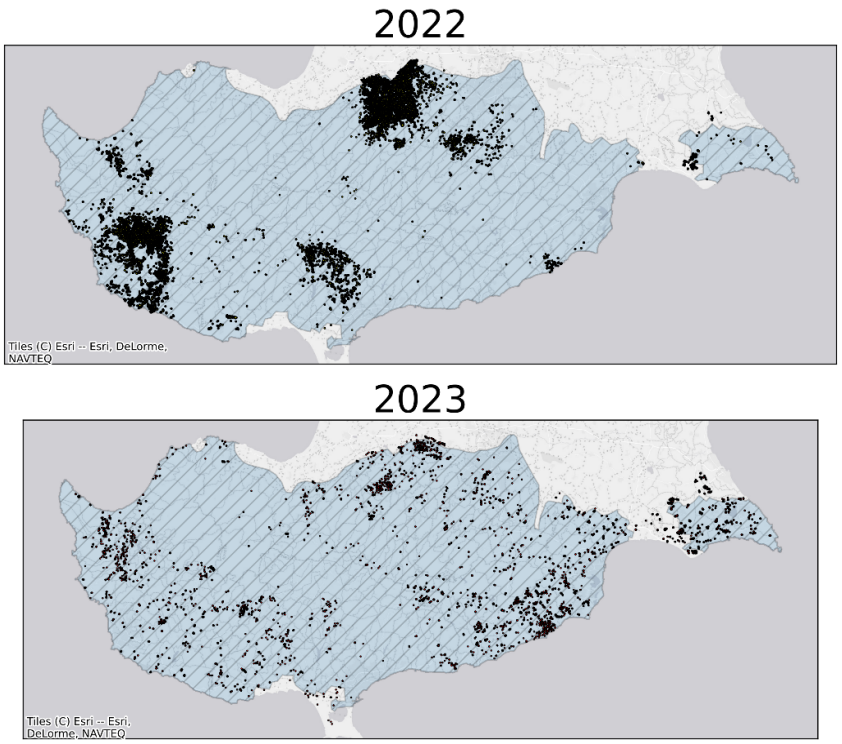


Figure 47: Validation samples geographical distribution across Cyprus territory for CCTM DP.

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

Validation Results

Crops Classification Results Performance

As outlined in D3.3 and D3.7, we generate multiple crop type maps from mid October of the previous year, until May of the current year. Model accuracy improves as more data accumulates, peaking it's optimal performance by the end of May.

Table 23 presents validation results for different machine learning models in early June for both 2022 and 2023. Notably, the Random Forest model outperforms others in accuracy and processing time.

Table 23: Classification performance for different machine learning models based on the predictions provided at the early June for 2022 and 2023.

	2022				2023			
	RF	SVM	XGBoost	MLP	RF	SVM	XGBoost	MLP
Recall (Macro Avg.)	**0.83**	0.71	0.81	0.73	**0.83**	0.71	0.82	0.72
Recall (Weighted Avg.)	**0.83**	0.70	0.81	0.74	**0.84**	0.71	0.82	0.74
Precision (Macro Avg.)	**0.90**	0.82	0.87	0.86	**0.90**	0.83	0.88	0.86
Precision (Weighted Avg.)	**0.83**	0.74	0.80	0.78	**0.84**	0.74	0.81	0.79
F1-Score (Macro Avg.)	**0.86**	0.77	0.84	0.80	**0.85**	0.77	0.85	0.81
F1-Score (Weighted Avg.)	**0.83**	0.74	0.82	0.77	**0.84**	0.74	0.83	0.77
Overall Accuracy	**0.83**	0.74	0.81	0.78	**0.84**	0.73	0.82	0.79
Kappa Coeff.	**0.80**	0.71	0.77	0.73	**0.81**	0.71	0.78	0.74
Elapsed Time (min.)	**2.1**	10.1	6.6	2.3	**0.9**	6.9	3.1	0.9

Table 24 presents a classification report for 30 crop classes, representing a snapshot of the model's performance in early June during the cultivation periods of 2022 and 2023.

The classifier exhibits strong accuracy across most crop classes, though challenges arise for mixed vegetable classes, characterized by their blended features. Overall, the model consistently achieves an accuracy above 80% in both years. Notably, the model tends to prioritize precision over recall, a factor that can aid regulatory bodies in ensuring accurate declarations and inspect potential cases of false declarations.

Table 24: Classification Report based on the predictions provided at the early June for 2022 and 2023.

	2022				2023			
	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
ALFALFA	0.83	0.77	0.80	13	0,50	0,67	0,57	3
BANANAS	1.00	0.93	0.96	28	1,00	1,00	1,00	13
BARLEY	0.79	0.85	0.82	2791	0,79	0,86	0,82	982
BLACK-EYED PEA	0.96	0.86	0.91	28	1,00	0,63	0,77	8
CAROB TREES	0.99	0.83	0.90	82	1,00	0,82	0,90	28
CITRUS TREES	0.93	0.85	0.89	345	0,98	0,85	0,91	131
CLOVER	0.63	0.96	0.76	23	0,62	1,00	0,76	8
CUCUMBERS	0.93	1.00	0.96	13	1,00	1,00	1,00	2

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

DECIDUOUS FRUIT TREES	0.90	0.82	0.86	523	0,86	0,87	0,86	171
FIGS	1.00	0.74	0.85	19	1,00	0,71	0,83	7
LAND LYING FALLOW	0.82	0.74	0.78	1944	0,82	0,74	0,78	698
LOLIUM	1.00	0.82	0.90	34	0,83	0,83	0,83	6
MELON	0.94	0.77	0.85	22	1,00	0,40	0,57	5
OAT	0.94	0.83	0.88	253	0,94	0,89	0,92	112
OLIVES	0.81	0.86	0.84	1405	0,86	0,85	0,85	503
ONIONS	0.96	0.81	0.88	31	1,00	0,84	0,91	19
ORCHARD	1.00	0.77	0.87	212	1,00	0,77	0,87	75
PEAS	1.00	0.86	0.92	71	1,00	0,67	0,80	24
PERMANENT GRASSLAND	0.87	0.94	0.90	196	0,87	0,89	0,88	66
POTATOES	0.73	0.87	0.79	200	0,81	0,95	0,87	76
SHRUB TREES	1.00	0.80	0.88	10	1,00	1,00	1,00	5
TOMATOES	0.89	0.82	0.85	50	0,79	0,85	0,81	13
TRADITIONAL TREES	0.91	0.80	0.85	293	0,98	0,86	0,92	107
TRITICALE	0.72	0.81	0.76	412	0,78	0,87	0,82	163
VARIOUS VEGETABLES	0.96	0.52	0.68	197	0,98	0,66	0,79	59
VICIA	0.95	0.83	0.89	266	0,99	0,80	0,88	95
VINES	0.89	0.93	0.91	752	0,92	0,92	0,92	268
WALNUTS	1.00	0.81	0.89	21	1,00	0,91	0,95	11
WATERMELON	1.00	0.85	0.92	40	0,93	0,81	0,87	16
WHEAT	0.79	0.83	0.81	2407	0,80	0,86	0,83	889
Macro Avg.	**0.90**	**0.83**	**0.86**	**12681**	**0.90**	**0.83**	**0.85**	**4563**
Weighted Avg.	**0.83**	**0.83**	**0.83**		**0.84**	**0.84**	**0.84**	
Overall Accuracy	**0.83**				**0.84**			
Kappa Coeff.	**0.80**				**0.81**			

The results show a clear improvement as the cultivation period advances (refer to Figure 48). This enhancement is primarily attributed to the growing clarity of spectral characteristics, facilitating more precise discrimination among various crop types. Remarkably, cereals (i.e., Barley, Wheat etc.) and the rest of arable cultivations (e.g., vegetables, vicia crops, etc.) notably enhance their accuracy after mid-January in the observed year. Of particular importance is the satisfactory performance achieved for

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

miscellaneous case of land lying fallow, aligning with CAPO's specified minimum expectations.

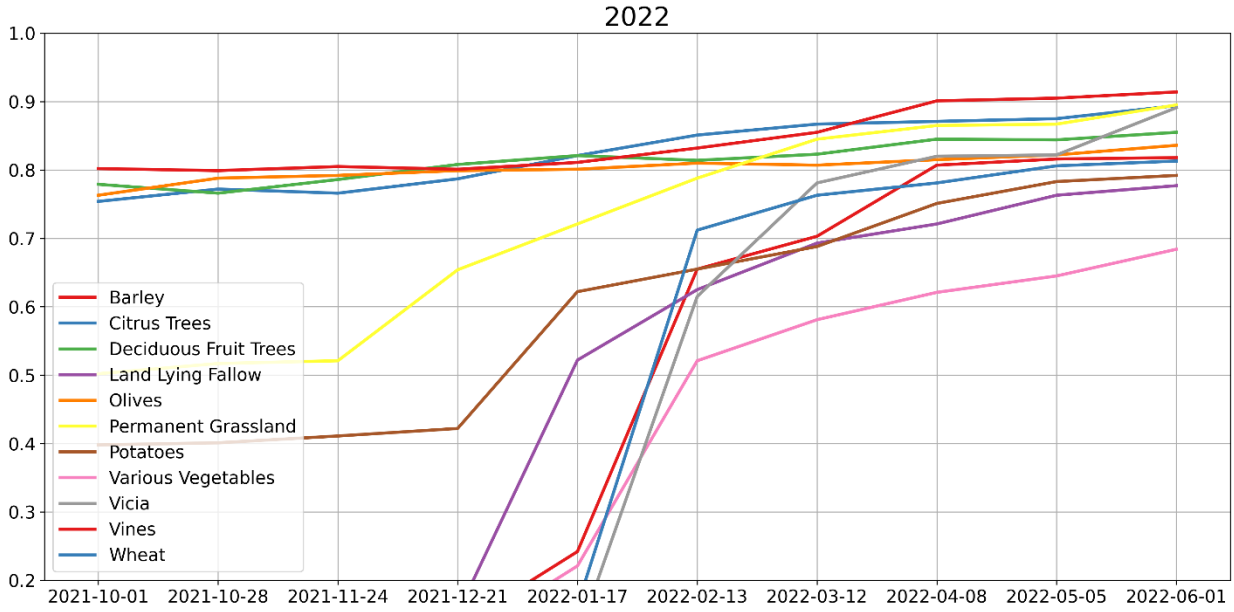


Figure 48: Classifier F1 score Progress over Cultivation Period of 2022. Results are similar for 2023.

Furthermore, our results reveal that the model's performance is minimally impacted by the size of the parcels (see Figure 49). As detailed in D3.7, the methodology operates at the pixel level for cases with less than 10 clear pixels (where results are subsequently aggregated at the parcel level through majority voting). These cases are excluded from the model's training. Consequently, as long as even a single pixel (equivalent to approximately 0.01 hectares) remains within a 5-meter inward buffer, results can be derived. As anticipated, performance tends to improve for larger parcels. This suggests that parcel size has limited influence on the model's accuracy, reinforcing the methodology's robustness.

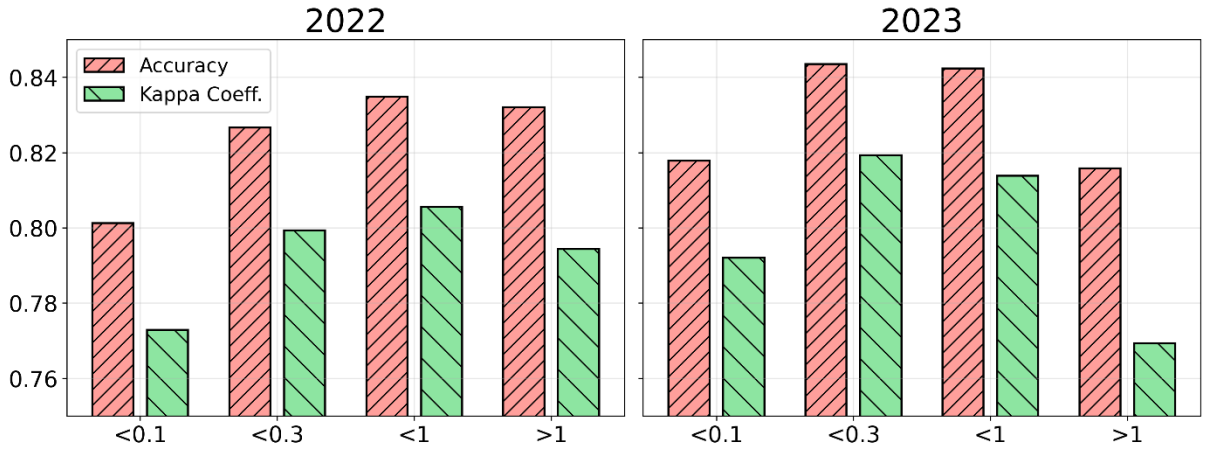


Figure 49: Classifier overall Accuracy and Kappa score for different parcel size (in hectares).

In summary, our model assigns probabilities to all available crop classes, with these probabilities summing up to 1. It's logical that higher probabilities correspond to higher accuracies, as expected. To gauge the accuracy difference between the most confident prediction and the second one, we can

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

evaluate the likelihood of correctly identified cases, as shown in Figure 50. As this difference becomes more substantial, accuracy gradually improves.

However, it's crucial to recognize that this approach affects the total number of predicted cases. Notably, about 60% of the total cases exhibit a difference higher than 0.2, while only approximately 20% show a difference higher than 0.5. This trade-off between algorithm accuracy and the number of predicted cases is pivotal in shaping our approach to interpreting results. It forms the basis for our smart sampling service, which aims to pinpoint the most certain instances of incorrect farmer declarations. These identified cases will be invaluable to end-users for strategic field inspections and comprehensive monitoring activities.

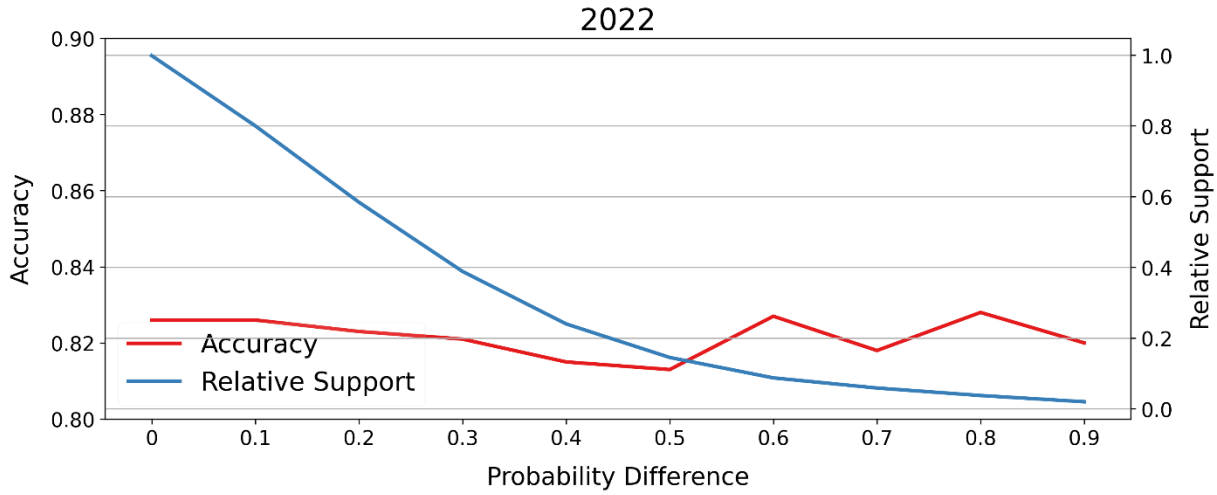


Figure 50: Accuracy and Relative Support (i.e., number of cases above this threshold/ total number of cases) trade-off for different values of probability difference between the 2 first most confident predictions.

Results Interpretability

Tables 25 and 26, displayed below, present producer and user accuracy data, respectively, along with information loss and confusion matrices that shed light on the model's misclassifications. The producer accuracy table, unveils the crop types mistakenly predicted by the model for false negatives, illuminating the taxonomy relationships. For instance, it reveals that 15% of actual vegetable instances were erroneously predicted as potatoes, given their similar taxonomy. Likewise, cereals are often confused due to shared characteristics. Conversely, the user accuracy table, discloses the actual ground truth for false positive predictions made by the model. In many cases, the model correctly predicts nearly all instances.

These confusion matrices offer invaluable insights for result interpretation. Notably, the model excels in User's Accuracy, suggesting its ability to successfully discern the spectral behaviors of various crop types. This holds significant implications for the subsequent smart sampling algorithm, which relies on predictions and their confidence levels to pinpoint false declarations. These results were computed at the start of June 2022, enabling CAPO to strategically plan their OTSC campaigns. The 2023 results exhibited a similar pattern, reinforcing the model's robust performance.

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

Table 25: Cyprus Producer Accuracy Table for 2022. Results for 2023 are similar.

Crop Name	Declared Parcels	Well Classified	Producer Accuracy	Confusion Class 1	1%	Confusion Class 2	2%	Confusion Class 3	3%	Rest %
Alfalfa	13	10	0,77	Wheat	0,15	Barley	0,08	Olives	0	0
Bananas	28	26	0,93	Barley	0,07	Wheat	0	Olives	0	0
Barley	2791	2372	0,85	Wheat	0,06	Land Lying Fallow	0,04	Triticale	0,02	0,03
Black-Eyed Pea	28	24	0,86	Wheat	0,11	Vines	0,04	Olives	0	0
Carob Trees	82	68	0,83	Land Lying Fallow	0,06	Barley	0,05	Wheat	0,04	0,02
Citrus Trees	345	293	0,85	Barley	0,05	Olives	0,04	Wheat	0,03	0,02
Clover	23	22	0,96	Wheat	0,04	Olives	0	Bananas	0	0
Cucumbers	13	13	1,00	Wheat	0,00	Olives	0	Bananas	0	0
Deciduous Fruit Trees	523	430	0,82	Olives	0,05	Vines	0,04	Barley	0,03	0,05
Figs	19	14	0,74	Olives	0,16	Wheat	0,05	Deciduous Fruit Trees	0,05	0
Land Lying Fallow	1944	1433	0,74	Barley	0,09	Olives	0,07	Wheat	0,06	0,04
Lolium	34	28	0,82	Barley	0,12	Wheat	0,06	Olives	0	0
Melon	22	17	0,77	Land Lying Fallow	0,14	Wheat	0,05	Potatoes	0,05	0
Oat	253	210	0,83	Barley	0,09	Wheat	0,05	Land Lying Fallow	0,02	0,02
Olives	1405	1211	0,86	Wheat	0,04	Land Lying Fallow	0,04	Barley	0,03	0,03
Onions	31	25	0,81	Wheat	0,06	Barley	0,06	Potatoes	0,06	0
Orchard	212	164	0,77	Olives	0,07	Wheat	0,05	Barley	0,03	0,08
Peas	71	61	0,86	Wheat	0,04	Barley	0,03	Vicia	0,03	0,04
Permanent Grassland	196	184	0,94	Wheat	0,03	Barley	0,02	Olives	0,01	0,01
Potatoes	200	174	0,87	Wheat	0,05	Barley	0,04	Land Lying Fallow	0,02	0,01
Shrub Trees	10	8	0,80	Wheat	0,10	Barley	0,1	Oat	0	0
Tomatoes	50	41	0,82	Land Lying Fallow	0,04	Barley	0,04	Vines	0,04	0,06
Traditional Trees	293	233	0,80	Barley	0,06	Olives	0,04	Wheat	0,03	0,08
Triticale	412	335	0,81	Wheat	0,09	Barley	0,08	Land Lying Fallow	0,01	0,01
Various Vegetables	197	103	0,52	Potatoes	0,15	Wheat	0,07	Barley	0,06	0,19
Vicia	266	222	0,84	Barley	0,06	Wheat	0,05	Land Lying Fallow	0,03	0,03

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

Vines	752	702	0,93	Wheat	0,03	Barley	0,03	Olives	0	0
Walnuts	21	17	0,81	Deciduous Fruit Trees	0,10	Olives	0,05	Citrus Trees	0,05	0
Watermelon	40	34	0,85	Barley	0,10	Wheat	0,03	Various Vegetables	0,03	0
Wheat	2407	1993	0,83	Barley	0,08	Land Lying Fallow	0,04	Triticale	0,02	0,02

Table 26: Cyprus User Accuracy Table for 2022. Results for 2023 are similar.

Crop Name	Classified Parcels	Well Classified	User Accuracy	Confusion Class 1	1%	Confusion Class 2	2%	Confusion Class 3	3%	Rest %
Alfalfa	12	10	0,83	Wheat	0,08	Barley	0,08	Olives	0,00	0,00
Bananas	26	26	1,00	Wheat	0,00	Olives	0,00	Barley	0,00	0,00
Barley	2998	2372	0,79	Wheat	0,07	Land Lying Fallow	0,06	Olives	0,02	0,07
Black-Eyed Pea	25	24	0,96	Potatoes	0,04	Wheat	0,00	Olives	0,00	0,00
Carob Trees	69	68	0,99	Land Lying Fallow	0,01	Wheat	0,00	Olives	0,00	0,00
Citrus Trees	314	293	0,93	Olives	0,03	Orchard	0,02	Land Lying Fallow	0,01	0,02
Clover	35	22	0,63	Barley	0,11	Wheat	0,09	Oat	0,06	0,11
Cucumbers	14	13	0,93	Tomatoes	0,07	Wheat	0,00	Olives	0,00	0,00
Deciduous Fruit Trees	476	430	0,90	Olives	0,03	Land Lying Fallow	0,02	Various Vegetables	0,01	0,04
Figs	14	14	1,00	Wheat	0,00	Olives	0,00	Bananas	0,00	0,00
Land Lying Fallow	1754	1433	0,82	Wheat	0,06	Barley	0,06	Olives	0,03	0,04
Lolium	28	28	1,00	Wheat	0,00	Olives	0,00	Bananas	0,00	0,00
Melon	18	17	0,94	Barley	0,06	Wheat	0,00	Olives	0,00	0,00
Oat	223	210	0,94	Various Vegetables	0,02	Wheat	0,01	Barley	0,01	0,01
Olives	1490	1211	0,81	Land Lying Fallow	0,09	Barley	0,02	Deciduous Fruit Trees	0,02	0,06
Onions	26	25	0,96	Barley	0,04	Wheat	0,00	Oat	0,00	0,00
Orchard	164	164	1,00	Wheat	0,00	Oat	0,00	Bananas	0,00	0,00
Peas	61	61	1,00	Wheat	0,00	Oat	0,00	Bananas	0,00	0,00
Permanent Grassland	212	184	0,87	Barley	0,05	Land Lying Fallow	0,04	Wheat	0,01	0,03
Potatoes	239	174	0,73	Various Vegetables	0,13	Barley	0,06	Wheat	0,04	0,05
Shrub Trees	8	8	1,00	Wheat	0,00	Oat	0,00	Bananas	0,00	0,00

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

Tomatoes	46	41	0,89	Various Vegetables	0,07	Vicia	0,02	Land Lying Fallow	0,02	0,00
Traditional Trees	257	233	0,91	Land Lying Fallow	0,04	Olives	0,03	Barley	0,01	0,02
Triticale	467	335	0,72	Barley	0,14	Wheat	0,12	Land Lying Fallow	0,01	0,01
Various Vegetables	107	103	0,96	Barley	0,01	Vicia	0,01	Land Lying Fallow	0,01	0,01
Vicia	234	222	0,95	Land Lying Fallow	0,02	Wheat	0,02	Peas	0,01	0,00
Vines	785	702	0,89	Deciduous Fruit Trees	0,03	Land Lying Fallow	0,03	Olives	0,02	0,03
Walnuts	17	17	1,00	Wheat	0,00	Oat	0,00	Bananas	0,00	0,00
Watermelon	34	34	1,00	Wheat	0,00	Oat	0,00	Bananas	0,00	0,00
Wheat	2528	1993	0,79	Barley	0,07	Land Lying Fallow	0,05	Olives	0,02	0,07

Towards smart sampling

As previously explained for the BC of Lithuania, alert cases are adaptively determined based on associated result probabilities. We evaluate alert levels by considering two critical parameters:

- a, which denotes the difference in probabilities between the top two confident predictions.
- b, referred to as persistent misclassifications, representing the total number of cases flagged as wrongly declared throughout the cultivation year.

This evaluation is visually represented through a traffic light system. For high-risk alerts (level 2 and 3), these parameters are configured to reflect an estimated percentage of false declarations, approximately 8-9% in the case of Cyprus's BC, based on historical data. In Figure 51, precision and recall are graphed across different disagreement percentages within the total declarations. Instances misclassified based on their initial declarations are sorted inversely based on their confidence intervals between the two most confident classifier predictions. The optimal balance between precision and recall is achieved at approximately 10%, slightly exceeding the expected false declaration percentage. Parameter a serves as the confidence interval threshold, dynamically set to represent around 10% of total instances. This approach helps ensure effective alert identification for a more robust assessment of false declarations.

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

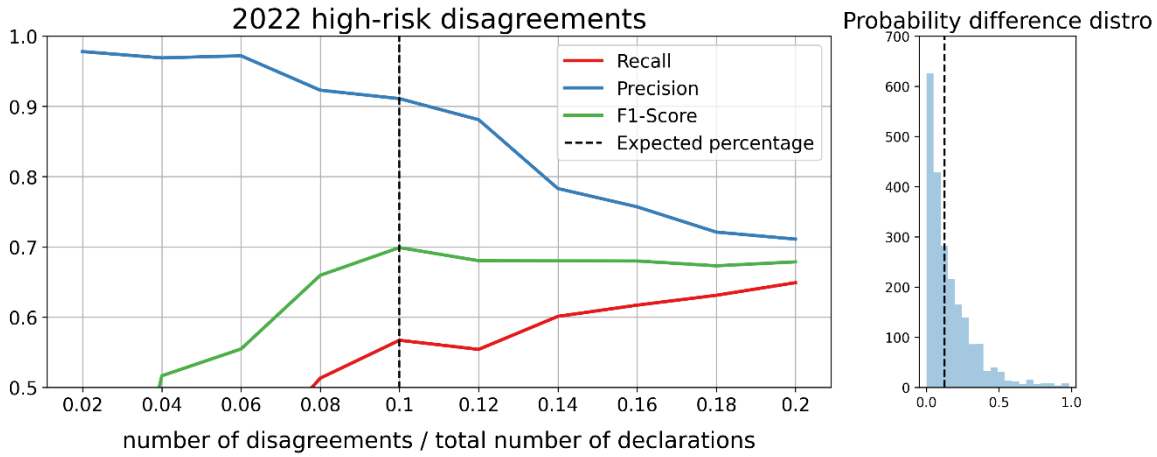


Figure 51: Evaluation of accuracy on high-risk disagreement vs the number of alert cases based on the confidence interval between two major predictions distribution (parameter a).

Parameter b, which signifies the count of times a sample is consistently misclassified in multiple classification iterations, is consistently set at a value of 2.

In Figures 52 and 53 presented below, we visualize the progression of precision and recall concerning high-risk alert cases (level 2 and 3) during the cultivation periods of 2022 and 2023, respectively. Precision approaches near-perfect values early in the cultivation period, while the recall of false declarations peaks towards the end of the year. Other actual wrongly declared cases are either assigned to lower-risk alert categories or remain undetected. Figure 54 portrays the distribution of actual wrongly declared cases categorized by the different alert levels generated by our system.

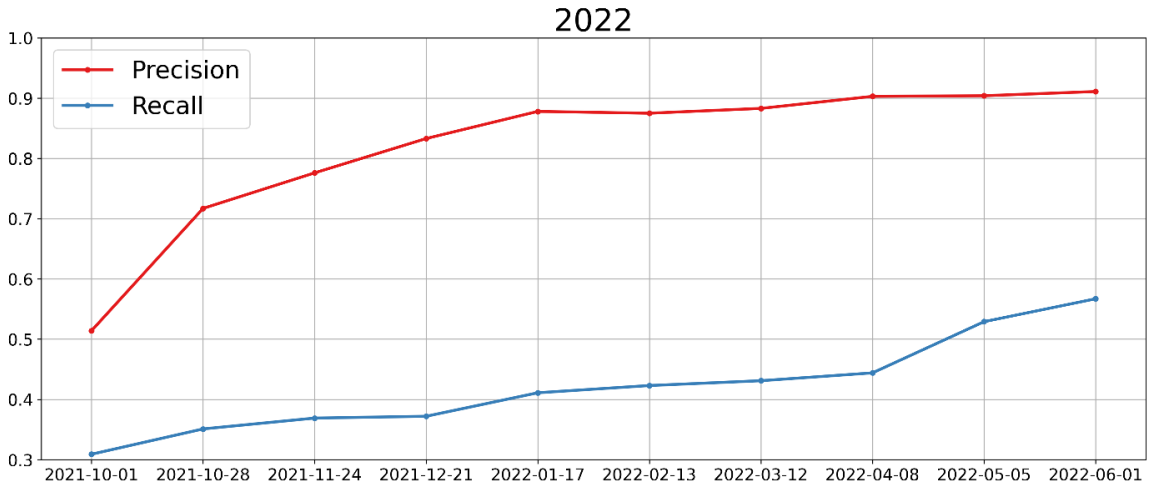


Figure 52: Progress of precision and recall of high-risk alert cases (level 2 and 3) during 2022.

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

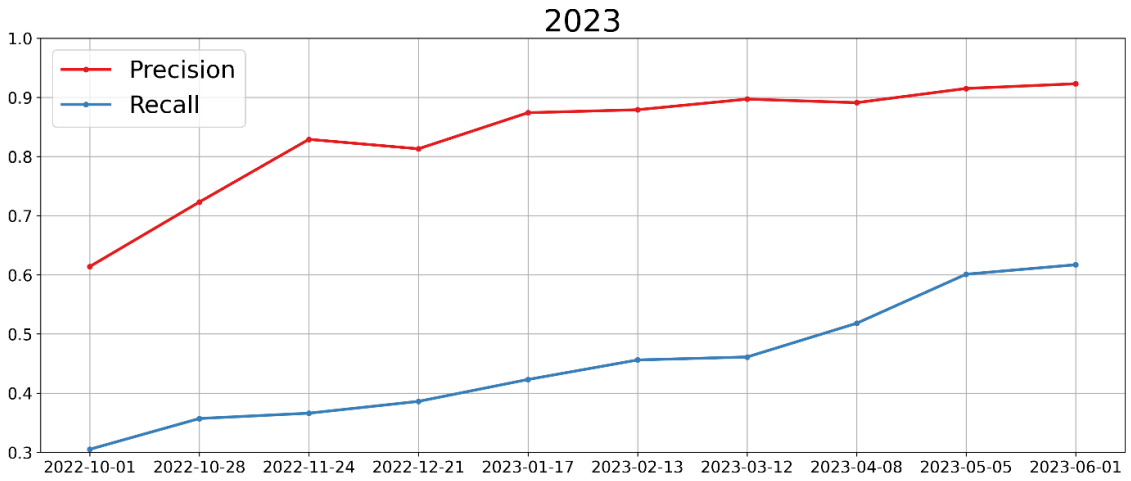


Figure 53: Progress of precision and recall of high-risk alert cases (level 2 and 3) during 2023.

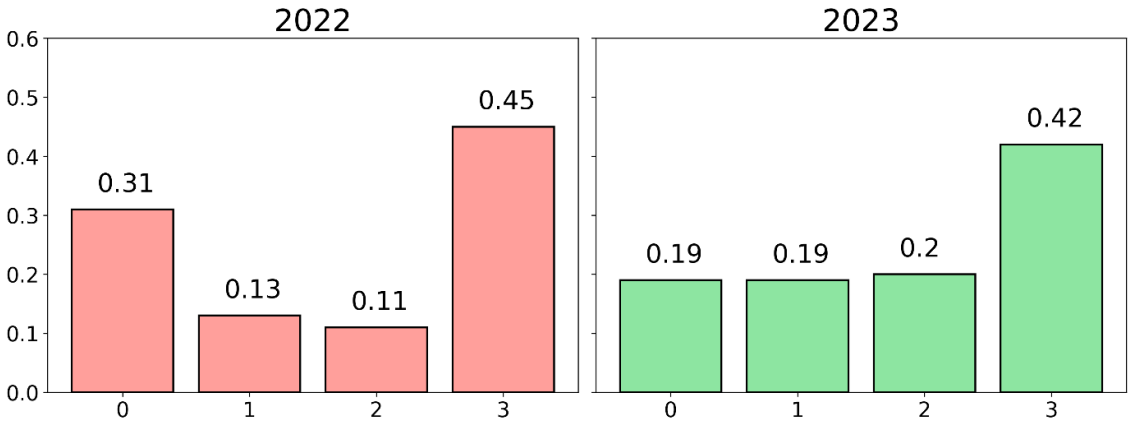


Figure 54: Portion of actual wrongly declared cases distribution among the various risk alerts output.

In Figures 55 and 56, we provide visual representations of two parcels indicated as high-risk alerts on 2022 to highlight the differences between the average NDVI behaviour of the declared crop type (orange colour), the predicted crop type from the classifier (green colour), and the actual NDVI time series of the specific parcel (blue colour). These figures clearly show that in both cases, the curve of the sample and the average curve of the predicted crop type closely resemble each other.

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

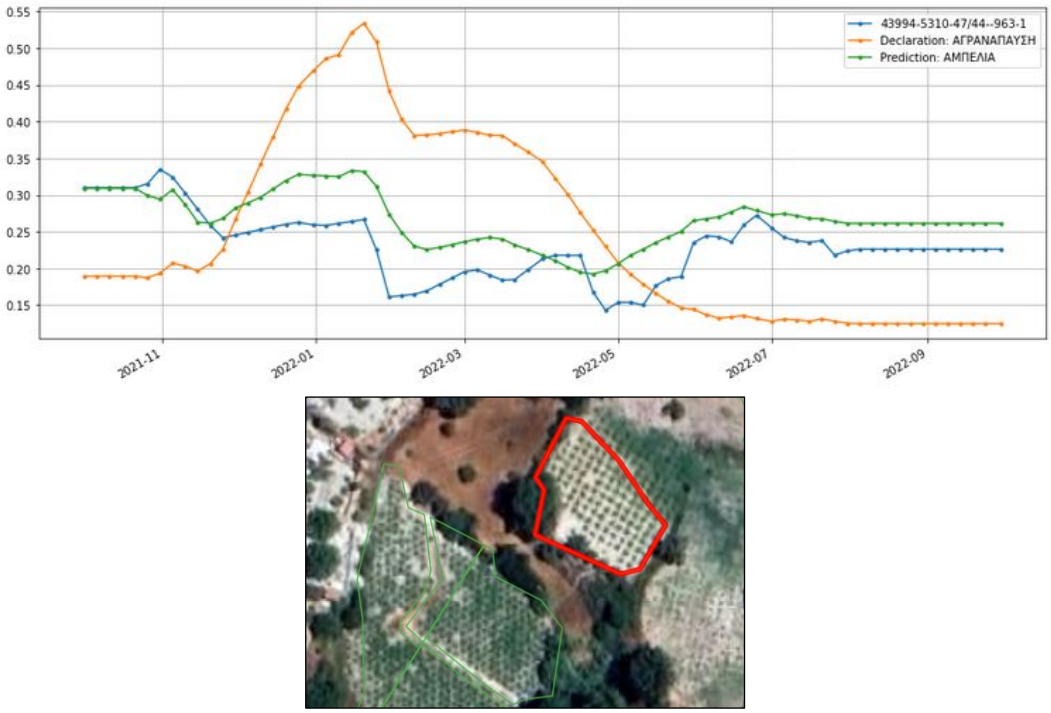


Figure 55: NDVI and S2-image of a case predicted as Vines and declared as Land Lying Fallow. According to this plot, fallows present much different NDVI signal during the cultivation period, which is definitely not evident here.

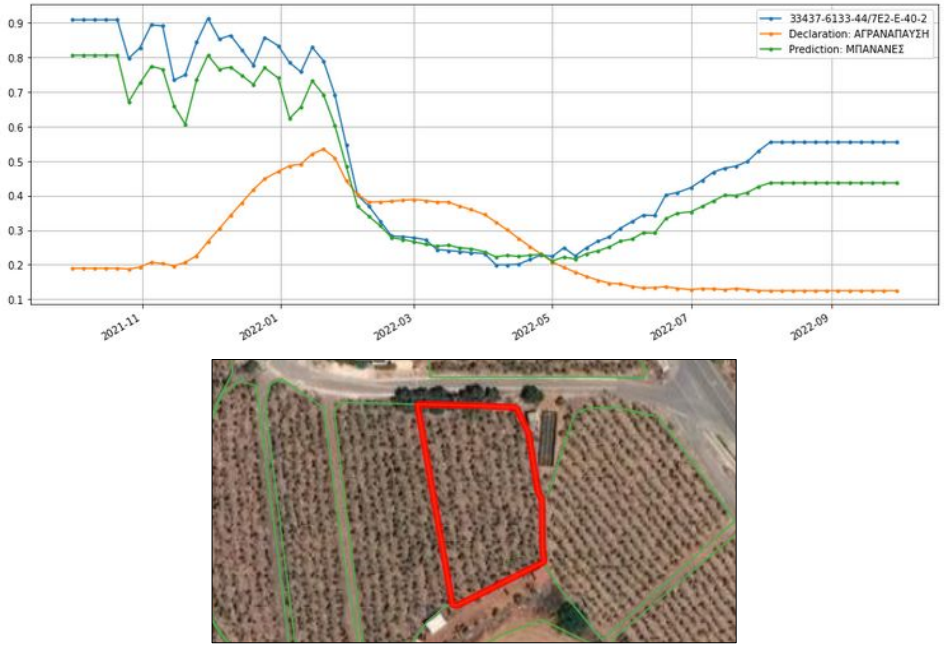


Figure 56: NDVI of a case predicted as Banana Trees and declared as Land Lying Fallow. According to this plot, fallow should present different characteristics. On the other hand, the NDVI signal is almost identical with the average NDVI of the Banana trees cases.

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

Discussion and limitations

In summary, our results demonstrate an exceptional level of accuracy, indicating practical applicability. Through a semi-automatic approach (users can set their own confidence thresholds of acceptance), users can leverage these results and their associated confidence levels to guide their focus towards more specific cases. Initially, in the context of the smart sampling scenario, we set stringent confidence parameters to enhance precision, resulting in a limited number of cases with exceptionally high precision. However, as we transition to a more exhaustive monitoring scenario, we relax these parameters to capture a greater number of alerts, with the primary goal of maximizing recall.

To address the challenge of training with inaccurately labelled data effectively since there is a significant high portion of false applicants' declarations (8-9%) in GSAA, we implement stacking ensembles, elaborated in detail in D3.7. This methodology involves training multiple base hierarchical models, each on an individual subset of the original dataset. The core aim is to harness the diversity of these models, each excelling at capturing different aspects of the data. During training, each base hierarchical model operates on its own dataset subset and utilizes the hierarchical structure of crop classification to enhance predictions at lower levels. Once trained, the stacking ensemble combines their predictions using majority voting, with the final decision based on the majority vote. This ensemble approach harnesses the collective knowledge of the models, resulting in a more comprehensive data representation and improved classification accuracy, consistently above 80%. Importantly, when dealing with false or inaccurately labelled data, this methodology offers enhanced resilience. The base models, trained independently on different data subsets, are less susceptible to the influence of individual instances of incorrect labelling. Additionally, the majority voting mechanism helps filter out erroneous predictions, as incorrect predictions are unlikely to achieve a majority consensus among the models.

Furthermore, our results reveal that one of Cyprus's primary challenges, the relatively small average parcel size, minimally impacts the model's performance. The methodology, detailed in D3.7, operates at the pixel level for cases with less than 10 clear pixels, with results aggregated at the parcel level through majority voting. These smaller cases are excluded from the model's training. Cases with not any clear S2 pixel after buffering (approximately 0.01 hectares), are excluded from the model's estimation. Last but not least, performance tends to improve for larger parcels, affirming the methodology's robustness and reliability.

3. BC3: Monitoring the condition of soil – Belgium

EV ILVO acts as a data provider in the Envision projects, as it is described in D4.1 and due to this, it is needed to satisfy two primary requirements: i) To develop soil quality data products that adequately support the needs for CAP monitoring at E.U. level. Ii) To deliver those products to the Envision platform to allow the easy and effective deployment of the products and to support the provision of the Envision services, considering also the Envision technological framework. To support both primary requirements within the current reporting period EV ILVO:

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

- Further, automated and completed the topsoil organic carbon prediction process, as that is described in D3.3, D3.4 and D3.7, in a way to be able to deliver soil quality data products to the Envision platform but to allow the use of Spatial-Temporal Asset Catalog (STAC) services as described in D3.7.
- Improved the way we provided the data products to the Envision platform, as described in D3.7, conforming with the provided directions coming from the contracting authority in a way to support further interoperability.
- Adopted an API-based architectural approach to allow either the direct use of the topsoil ML models (see more at D1.9 2nd Progress Report) or the development of applications that make on-demand requests for a data product for specific AOI (polygon), similar to the one demonstrated at the AgriTEF Dag in Flanders on Jun 6 2022.
- Improved the accuracy of the topsoil organic carbon prediction models and finalised the modelling process, and covered all mentioned D1.9 scenarios.
- Developed and presented the Flemish Soil Quality Maps that use as an indicator the predictions of the topsoil organic carbon model, also considering the pedological conditions, as described in D3.7. The data products deliver information at pixel and parcel levels.
- Adapted the developed models to allow the application to other E.U. regions by using INSPIRED harmonised data sets, as described in D3.7.

3.1. Methodology

EV ILVO defined a methodology (Figure 8) that enables current scientific research outcomes and delivers on a large scale soil quality data products using indicators that rely on topsoil organic carbon predictions. As we have presented in other deliverables like D3.4, D3.3 and D3.7, the methodology allows the continued development of data products at regional (for example, in Flanders, Figure 9), National and E.U. levels. The data products provide the information at pixel and parcel levels aiming to cover the CAP needs for soil organic carbon monitoring in cropland, supporting P.A.s applying their strategic plans.

The significant methodological phases have remained the same within the current reporting period. However, we took action, and we performed adjustments to:

- Further, automated and completed the topsoil organic carbon prediction process, as that is described in D3.3, D3.4 and D3.7 in a way to be able to deliver soil quality data products to the Envision platform and to allow the use of Spatial-Temporal Asset Catalog (STAC) services as described in D3.7 (Figure 10).
- It is now possible to deploy the topSOC ML prediction model at each time stamp (satellite image) of the cloudless bare soil collection and compare the results using only the synthetic layer (Figure 11).
- Improved the way we provided the data products to the Envision platform, as described in D1.6, conforming with the provided directions coming from the contracting authority in a way to support further interoperability.
- Additionally, EV ILVO will deliver both Data products A and B at ZENODO and the modelling metadata using the OGC GeoPackage 1.3.1 to support interoperability. However, we still need to provide LV with detailed information on the Flemish model accuracy, the information related to

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

the sampling campaign and the lab measurements using a report that will follow the provided data products.

- Adopted an API-based architectural approach to allow either the direct use of the topsoil ML models (see more at D1.9 2nd Progress Report) or the development of applications that make on-demand requests for a data product for specific AOI (polygon), similar to the one demonstrated at the AgriTEF Dag in Flanders on Jun 6 2022 (Figure 12).
- The use of STAC services and the provision of the data products or the TopSOC ML predictions by using APIs allows not only the effective collaboration with other partners, for example, NOA, but also the future provision of the data as a service to the Envision platform and the collaboration with Flemish Governmental projects like the Soil Passport.
- Improved the accuracy of the topsoil organic carbon prediction models and finalised the modelling process, and covered all mentioned D1.9 scenarios.
- Developed and presented the Flemish Soil Quality Maps that use as an indicator of the predictions of the topsoil organic carbon model, also considering the pedological conditions, as described in D3.7. The data products deliver information at pixel and parcel levels (Figure 14).
- Adapted the developed methodology to allow the application to other E.U. regions by using INSPIRED harmonised data sets, as described in D3.7 (Figure 13).

3.2. Product description

For a more text like description and more details, see deliverable 3.7 and 3.5. For source data and data formats see table 3.5.

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

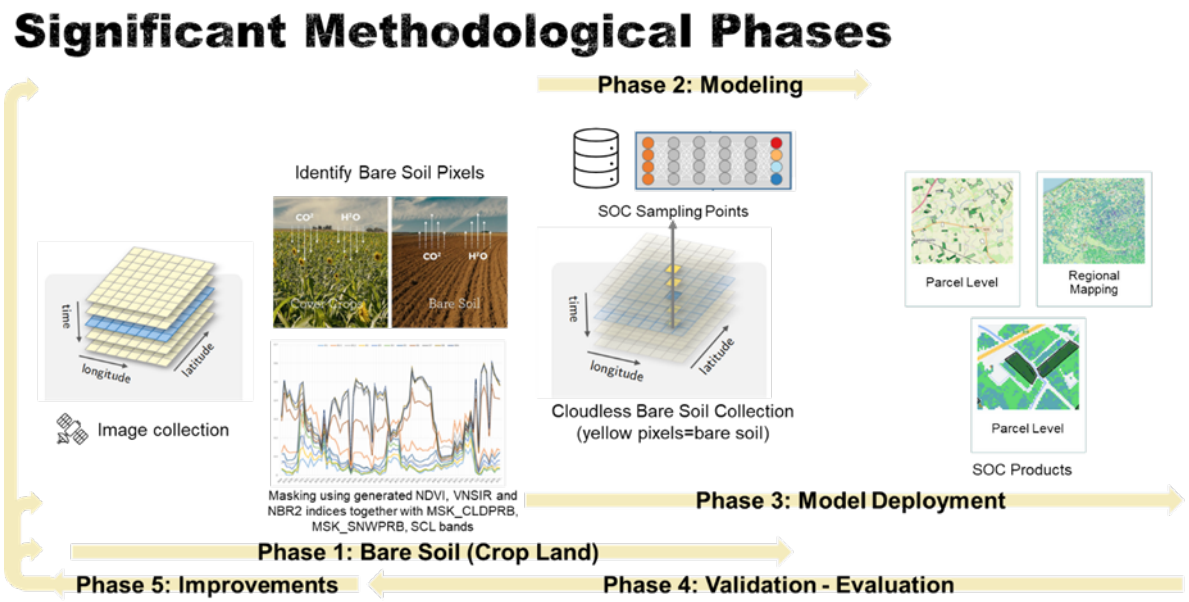


Figure 57: Significant methodological phases supporting large-scale SOC mapping and development of soil quality indicators at pixel (intra-field) and parcel level (aggregation).

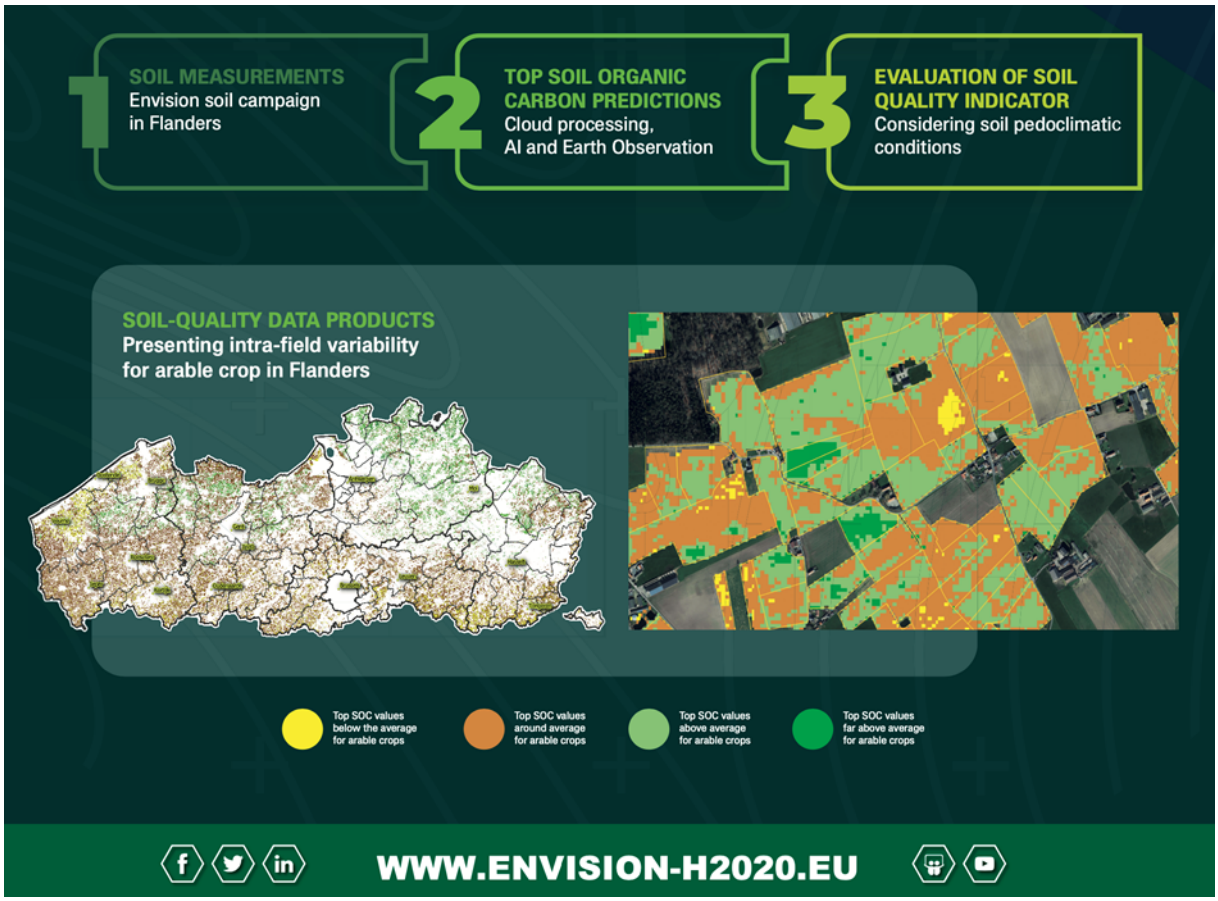


Figure 58: Soil Quality data product presented at the AgriTEF Day on the 6th o June 2023.

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

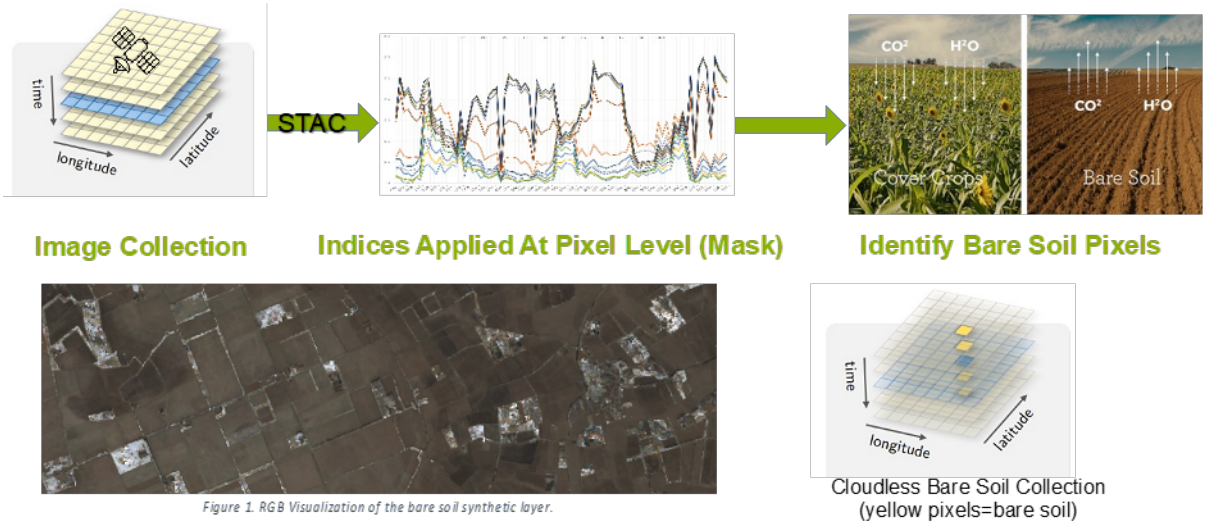


Figure 59: For the development of the data products, access to the satellite image collections or to other data products is being done using the Spatial Temporal Asset Catalogs service (STAC).

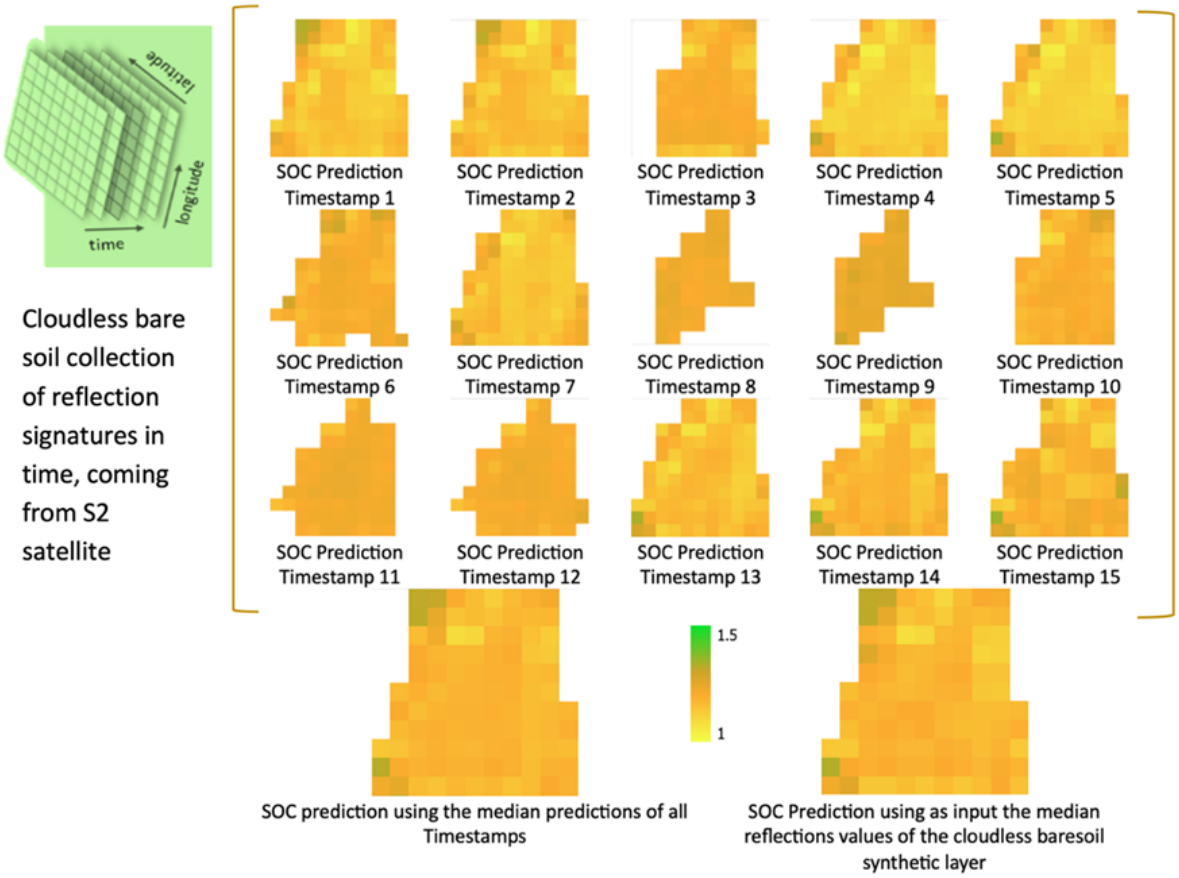


Figure 60: Within the current reporting period, EV ILVO automated further the data development process, adding the ability to deploy an ML on the synthetic layer or for each bare soil cloud collection layer. Each layer represents different timestamps within the collection. The ability to deploy the ML for each timestamp enables the topsoil organic carbon prediction separately or the statistical process of the predictions. Both support the

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

monitoring with the further assessment of the accuracy of the model and the dynamic visualisation of the results.

Figure 61: Demonstration of the possibility of providing the data products by using an API. This way, a farmer can request to see the intra-field soil quality conditions only for his parcels. The demonstration took place on the Flemish AgriTEF Day, collaborating with the Flemish Department of Agriculture (LV), allowing EV ILVO to consume an API that delivers per Farm the agricultural parcels. We used DjustConnect authorisation and data consent services to overcome GDPR issues successfully.

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

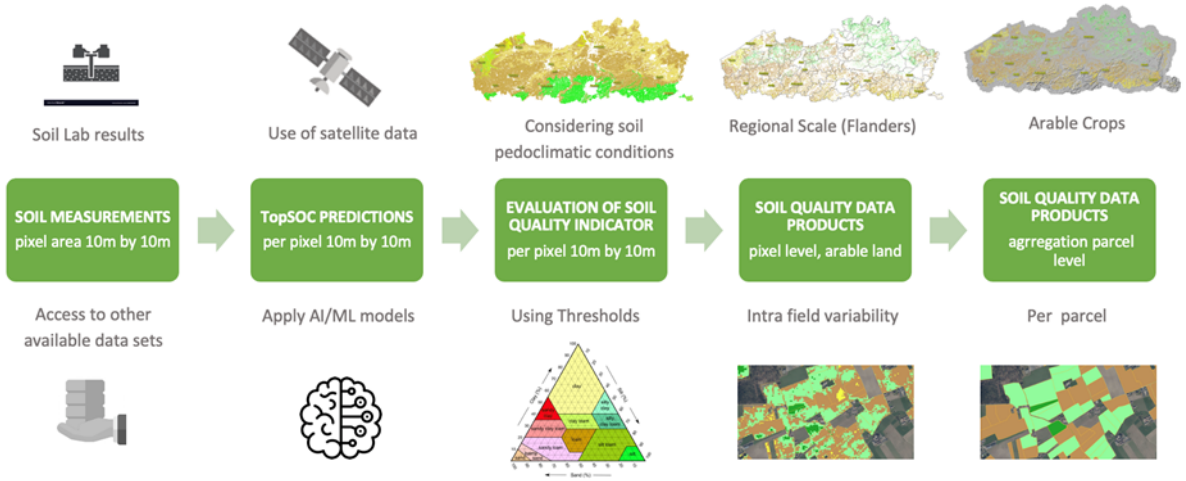


Figure 62: Passing from topSOC prediction to the Development of Soil Quality data products at pixel and parcel level, considering pedoclimatic conditions. By using INSPIRED harmonised data, applying the same steps to other E.U. regions is possible.

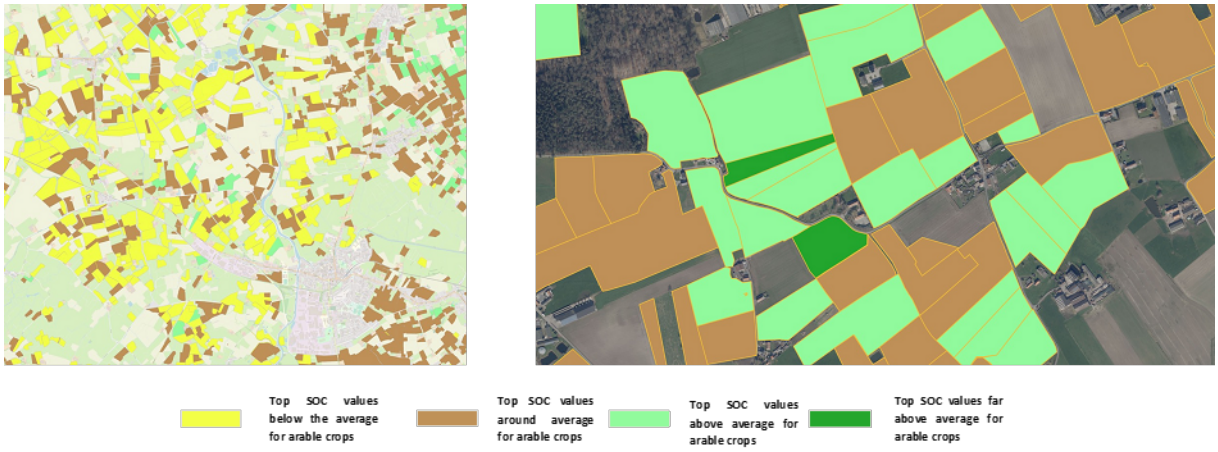


Figure 63: Envision Soil Quality products at pixel and parcel level, covering the Flemish region.

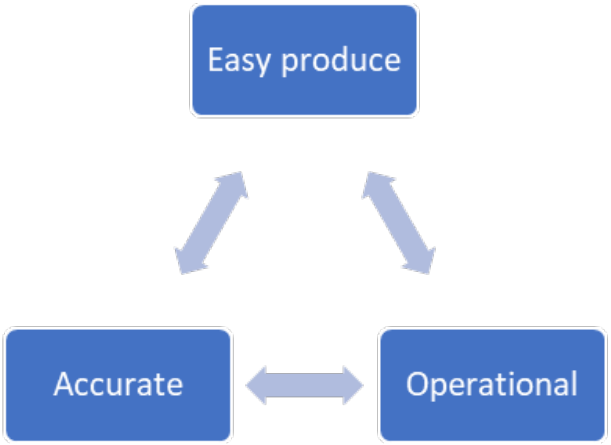


Figure 64: EV ILVO SOC methodology tries to balance three goals to achieve large-scale applicability. First, easy to produce, second to be operational and third, to achieve the needed accuracy levels to support the CAP needs for SOC monitoring.

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

3.3. Validation criteria and results

Soil Organic Carbon monitoring

Within the current reporting period, we continue the work on the modelling, aiming to improve the accuracy. We developed and tested several regression models within the **third iteration of our product developments**. In this deliverable, we will present the results of the three basic model scenarios:

- **Scenario A:** Input parameters are only the reflection values. This scenario is the continuation of the basic scenario of the first iteration. Because the models have the same input parameters, we compared the modelling results with those from the first iteration.
- **Scenario B:** We have included an extra input parameter, the soil association type.
- **Scenario C:** We have included an extra input parameter, the soil association type and the period (Month).

Table 27: Executed SOC modelling scenarios.

	Modelling scenarios for the SOC Models						
Iterations period	1st iteration Oct 20-Feb 22	**2nd iteration** **Marc – August 2022**			**3nd iteration** **Sept 2022 – June 2023**		
Scenario Name	**Scenario A**	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario A**	**Scenario B**	**Scenario C**
Input parameters	Reflections Values of all S2 bands	Reflections Values of all S2 bands	Reflections Values of all S2 bands soil association type	Reflections Values of all S2 bands soil association type period (Month).	Reflections Values of all S2 bands	Reflections Values of all S2 bands soil association type	Reflections Values of all S2 bands soil association type period (Month).
Parameters Value	Median value per band for the period May 2018	All reflection values per band for the period May 2018 until August 2022			All reflection values per band for the period May 2018 until April 2023		

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

	until Dec 2021						
Reflection values extracted	Cloudless Bare Soil Collection Layer	Cloudless Bare Soil Collection Layer			Cloudless Bare Soil Collection Layer		
SOC measurements region	Flanders	Flanders			Flanders		
Training sampling / cross validation strategy	80%, 20%, 10% random sampling	80%, 20%,20% random sampling /Fold Group using point I.D.	80%, 20%, 20% random sampling /Fold Group using point I.D.	80%, 20%, 20% random sampling /Fold Group using point I.D.	80%, 20%,20% random sampling /Fold Group using point I.D.	80%, 20%, 20% random sampling /Fold Group using point I.D.	80%, 20%, 20% random sampling /Fold Group using point I.D.
Model Code	**01MedBands**	**02TSeBands**	**02TSeBandsSoil**	**02TSeBandsSoilMon**	**03TSeBands**	**03TSeBandsSoil**	**03TSeBandsSoilMon**

In the third iteration, we followed an approach similar to the second approach, which means:

- We train the models using the reflection values coming from all timestamps per pixel.
- The model sampling strategy is 80%, 20%, 20%, which means 20% of the sampling points consist of the unseen data set, and from the 80% of the seen data set, the 80% consist of the training set and 20% the test set.
- In scenarios B and C, we use point I.D. to define a cross-validation strategy that considers that records belong to specific measurement points and makes sure that records belonging to the same point cannot be in different folds for Cross-Validation or both in calibration and test set for prediction, thus preventing overfitting. The unseen data is also selected so that there the point id's of the unseen data is not present in the seen dataset.
- After testing, we decided not to proceed with Scenario D, because we didn't show any improvement in accuracy.
- After testing, we decided not to proceed with using markers that point out the phenological states due to the difficulty on the method's applicability. The provided data set do not allow the specific definition of the start and end of a crop growing cycle.
- The TopSOC prediction supported the development of Soil Quality indicators. There as explained in D3.7 we use the distribution method to identify the zone in the Flemish region where the

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

expected TopSOC is below, close, or above the average, considering also the pedoclimate conditions. This approach allows the identification of zones within the parcel where the conditions are not favourable.

In the following pages, we will present the **modelling validation results** with some figures and tables that describe the model's performance, how the parameters contribute to the results and other information that can better support the end users of this data product to understand the model's accuracy.

Table 28: 3rd iteration model validation results for different scenarios.

Scenarios		**Scenario B**	**Scenario A**	**Scenario C**
Model Code		03TSeBandsSoil	03TSeBands	03TSeBands
Models with the best performance	E.T.	R^2:0.70	R^2:0.68	R^2:0.68
		RPD:1.84	RPD:1.78	RPD:1.78
	MLP	R^2:0.65	R^2:0.37	R^2:0.37
		RPD:1.68	RPD:1.26	RPD:1.26
	catboost	R^2:0.58	R^2:0.62	R^2:0.62
		RPD:1.55	RPD:1.62	RPD:1.62
	Basian Ridge	R^2:0.44	R^2:0.42	R^2:0.42
		RPD:1.33	RPD:1.34	RPD:1.34

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

3.4. Discussion

The results are on independent, not before seen data by the models. But the results can be inconsistent/ unstable with different divisions of training and validation/ test data.
This mainly signifies certain combinations of soil associations with certain spectral signatures are then not present in the training set, because spectral models do not work well outside or their calibration range.
More top soil OC samples, especially for underrepresented bare soil spectral signatures/ soil associations combinations could help remedy this in the future.

Technological Roadmap see deliverable 3.7, also 3.7 for limits algorithm, requirements.

Deliverable 3.5 for input/output table and data format, schemas/ workflows in 3.5 for soil campaign design, 3.7 for schemas/ workflows top soil OC modelling, OC regression map generation, and Soil Quality data product.

4. BC4: Monitoring of organic farming requirements – Serbia

4.1. Product description

The present section of the deliverable deals with the validation process of the D5 product, developed in the context of the Envision project by AgroApps, and in particular in the sub-service "Distinction of Organic Farming Practices". It follows a “Data product as service” business model aiming to provide an Agriculture Monitoring System for Certification Bodies that seek a surveillance tool for the assessment of the validity of farming practice declarations.

The data product being validated is a vector geospatial feature which is served through the ENVISION platform, or alternatively could be served via WFS to the user, in a shapefile format. It contains the parcel polygon boundary geometries, and as far as attributes, the evaluation of the Classification as Organic/Conventional farming practice, in the representation of a traffic light system. Its values are given in a standardised confusion matrix terminology, and depict the result of the prediction in regards with what was initially declared. The main details of this data product, regarding the input data, the coverage area, the methods incorporated as well as the output, are summarised in the following table.

Table 29: D5 Product - Distinction of Organic Farming Practices service outline.

Product	Service	Data Input	Data Format	Thematic Content	Spatial Distribution	Data Processing	Applied Method	Data Output	Data Format
DP5	**Distinction of Organic Farming Practices**	**LPIS (GeoSerbia)**	SQL table, Shapefile,Geojson	Land Parcel Identification System	User Defined, (across Serbian Administration Units)	*Spatial Aggregation*	*Dissolve*		

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

		GSA (OCS)	SQL table, csv	Farming Practice Declarations		*Spatial Proximity*	*Inner Buffering*	
							Compute Parcel GeometryArea	
							Compute Parcel Geometry Elongation	
						Data Descriptive Statistics	*Frequency Distribution*	
						Spatial Sampling	*Random Points*	
		SoilGridsTM	Raster Grids (.tif)	Soil Organic Carbon	User Defined Areas of Interest	*Raster Calculation*	*Compute Soil Organic Matter*	
				Sand/Silt/Clay content		*Raster Reclassification, Conditional*	*Compute USDA Soil Texture*	
		Sentinel 2 MSI	Raster Grids (.SAFE)	VIS-NIR-SWIR reflectance	User Defined Areas of Interest			
		L1C				*Atmospheric Correction*	*Sen2Cor*	
		L2A				*Feature Extraction*	*Vegetation Indices*	
							Quality Masking	
						Temporal Gap Fill	*Spline Interpolation*	
						Smoothing – Temporal Derivatives	*Savitzky – Golay Filters*	
						Image Texture	*GLCM Metrics*	

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

				Vegetation Phenology	*Double Sigmoid Curve Fitting*		
				Dimensionality Reduction	*PCA*		
				Outlier Detection	*Isolation Forest*		
				Novelty Detection	*One-Class SVM*		
				ML Classification Algorithm	*XG-Boost (XGB)*		
				Data Augmentation			
				Zonal Functions	*Zonal Parcel Area Tabulate*		
					Zonal Parcel Statistics (Mean, Standard Deviation)		
						TrafficLight – Confusion Metrics	Shapefile

The relevant metadata provided to the user concerning the product are the accuracy evaluation metrics of the crop specific classification models, the spatial and temporal extent of the service job,

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

the EO features used, and of course, the descriptive statistics of the input data concerning the farming practice declarations.

The objectives of the product are, to implement the training of classification algorithms with EO data, to predict farming practice at plot level for a specified crop type, and to report to the user information on the validity of the declaration in the form of a traffic light system that examines the following possibilities:

- a parcel was predicted conventional while being organic
- a parcel was predicted organic while being conventional
- a parcel was predicted correctly as conventional
- a parcel was predicted correctly as organic

It allows the user to reach conclusions on the compliance of the crop declaration. For example, the case where a parcel is predicted as organic while being conventional, is an indication of non-compliance of the declaration with the prediction of the classification models, which would need to be further checked.

The input data, for the 2022 and 2023 pilot business cases, included spatial information from the Serbian LPIS and their associated attributes from the GSA statements. The table schema of the GSA was proposed to include the following fields:

- Parcel ID (unique id - Primary Key)
- Applicant ID
- Year of Declaration
- Farming Practices: The declared type of farming practices (organic or conventional)
- Declared Crop Code
- Crop Variety
- Crop Type Category

The geographical distribution of the parcel data in the Serbian territory is depicted on the maps that follow, while statistical data concerning their statistics are given in the subsequent tables.

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

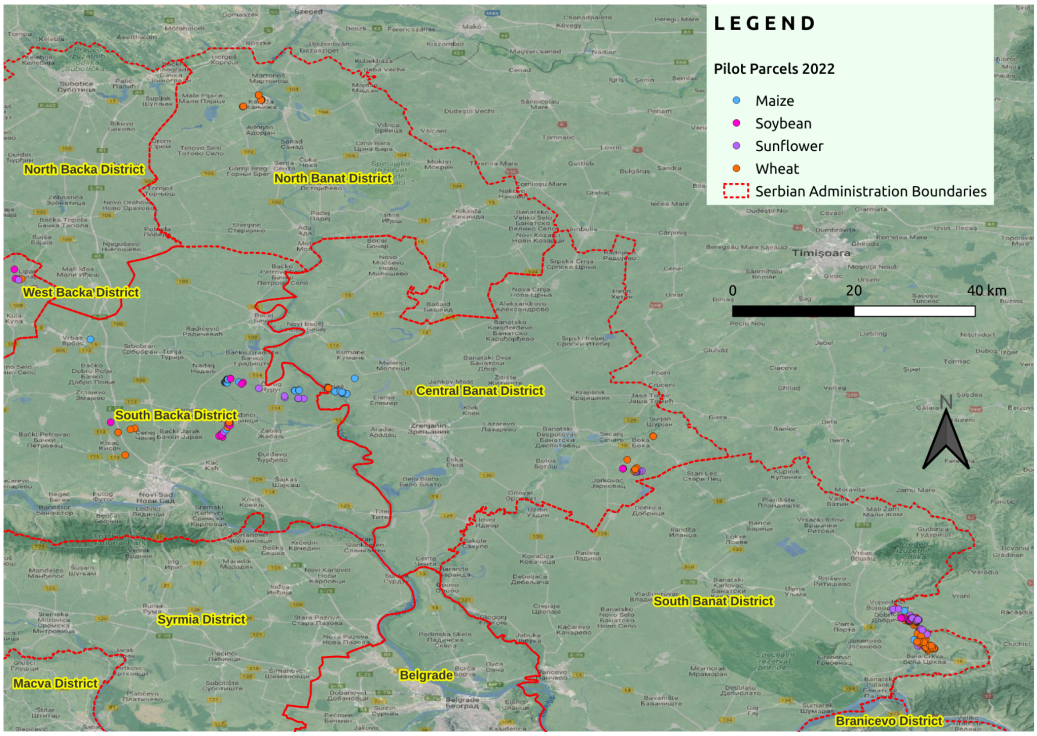


Figure 65: Geographic Distribution of 2022 pilot parcels.

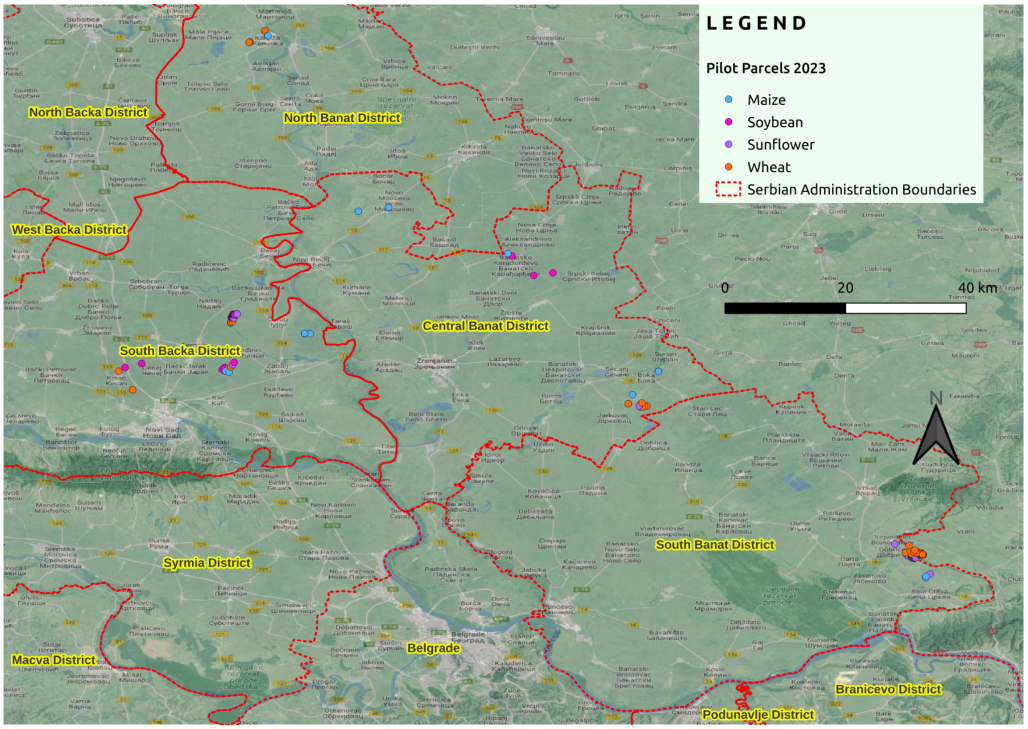


Figure 66: Geographic Distribution of 2023 pilot parcels.

Table 30: Data Descriptive Statistics Distribution of 2022 pilot parcels.

Distribution of parcel count and area per crop						
Crop Type	***Conventional***		***Organic***		***Total Area***	***Total Count***
	Area (Ha)	***Count***	***Area (Ha)***	***Count***		

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

Maize	*0*	*0*	*402.6802*	*54*	***402.6802***	***54***
Soybean	*2.56*	*4*	*330.0166*	*81*	***332.5766***	***85***
Sunflower	*5.8197*	*9*	*405.8672*	*87*	***411.6869***	***96***
Wheat	*9.0466*	*18*	*1069.8639*	*149*	***1078.9105***	***167***
Total Result	***17.4263***	***31***	***2208.4279***	***371***	***2225.8542***	***402***

Table 31: Data Descriptive Statistics Distribution of 2023 pilot parcels.

Distribution of parcel count and area per crop						
Crop Type	***Conventional***		***Organic***		***Total Area***	***Total Count***
	Area (Ha)	***Count***	***Area (Ha)***	***Count***		
Maize	*1.383*	*4*	*32.644*	*15*	***34.027***	***19***
Soybean	*0*	*0*	*165.738*	*26*	***165.738***	***26***
Sunflower	*9.9289*	*18*	*177.4786*	*45*	***187.4075***	***63***
Wheat	*6.525*	*10*	*374.8306*	*139*	***381.3556***	***149***
Total Result	***17.8369***	***32***	***750.6912***	***225***	***768.5281***	***257***

The analysis of the shape and geometry of the polygons provided by the Serbian LPIS was carried out by calculating a shape index that quantifies the elongation and is indicative of the number of "useful and representative" pixels that can be aggregated on the parcel when calculating the zonal statistics. The aim was to establish the adequacy of the dataset in providing a sufficient set for training machine learning algorithms. The following relationship was thereafter implemented:

$$SHAPE\ ELONGATION = 4 \cdot \pi \cdot \frac{SHAPE\ AREA}{(SHAPE\ PERIMETER)^2}$$

The index values range in the interval [0,1], to represent polygon shapes that span from elongated to circular geometries. The distribution of Shape Elongation Index values in the pilot plot data for the years 2022 and 2023 is shown in the histogram charts below, while a table of statistics for the geometry parameters is given below. The plot of typical plots representing the average Elongation and Area values is further given in the figures below.

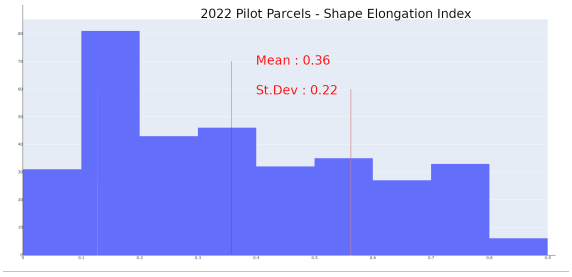


Figure 67: Shape Elongation Histogram of 2022 pilot parcels.

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

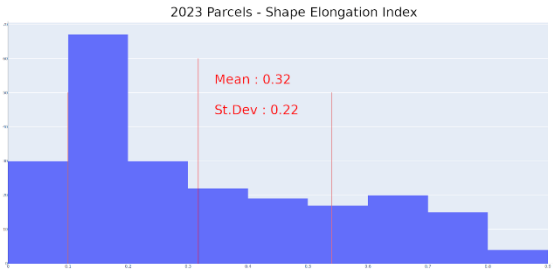


Figure 68: Shape Elongation Histogram of 2023 pilot parcels.

Table 32Distribution of Shape Statistics of 2022 pilot parcels.

2022 Shape Geometry Statistics		
	Elongation	**Area (Ha)**
Mean	0.36	5.94
Median	0.31	0.72
St.Dev	0.22	15.7

Table 33: Distribution of Shape Statistics of 2023 pilot parcels.

2023 Shape Geometry Statistics		
	Elongation	**Area (Ha)**
Mean	0.32	3.3
Median	0.24	0.7
St.Dev	0.22	9

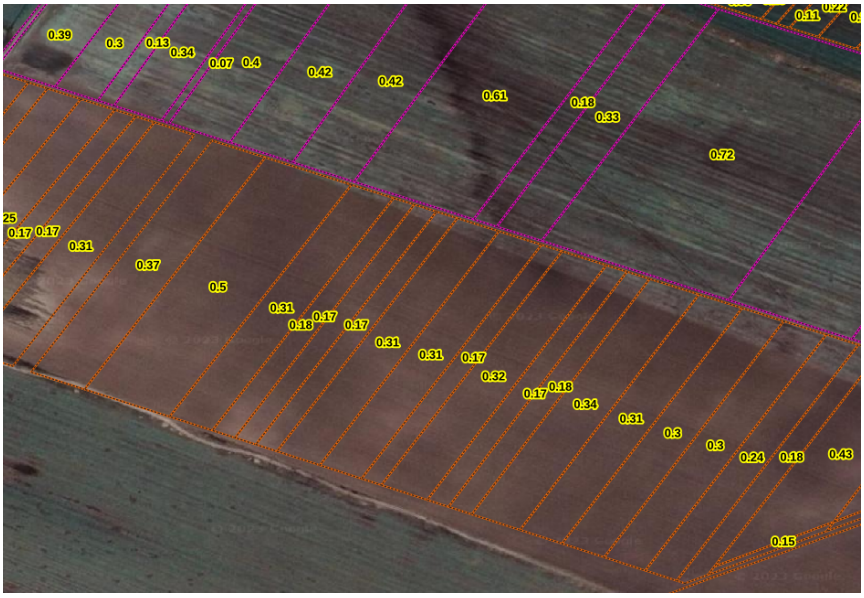


Figure 69: Typical shape representations of 2022 and 2023 pilot parcels. The shape elongation index value of each parcel is valued.

The key conclusions that emerged from the preliminary evaluation of the input data were as follows:

- The data show a wide geographical dispersion in the regions of Central Banat, North Banat, South Banat, South Backa and West Backa, for which variable climatic conditions are assumed

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

- The geometry of the Serbian LPIS polygons provided for the 2022 and 2023 pilots showed particularly small and elongated parcels
- The count and the monitored area of data per crop, and even more so when analysed within different varieties, was found to be in small numbers.
- The distribution between Conventional and Organic plots was highly heterogeneous, with the latter strongly dominating the dataset

It should be noted that the ground truth data were obtained in 3 batches following requests to increase the sample size. However, the Outlier Analysis via visual inspection was performed on each take yielded a significant proportion of the data as not eligible. Visual inspection is a relatively simple process that requires the user to have a fairly basic knowledge, which could possibly be trained, of what a timeseries NDVI profile displays in a crop, as well as the sowing/cutting dates in the area of interest. The user observes the profile of each field-sample in the training set, and scores it as to the correctness of its statement. More specifically, a visual inspection of NDVI phenology curves was performed to identify instances that did not capture the typical profile of the crops under consideration, and were communicated to OCS for further investigation. In the following figure, a typical example of a non - eligible parcel that was addressed to OCS for validation is presented. The observed NDVI curve clearly displays a parcel that couldn't be classified as a wheat crop.

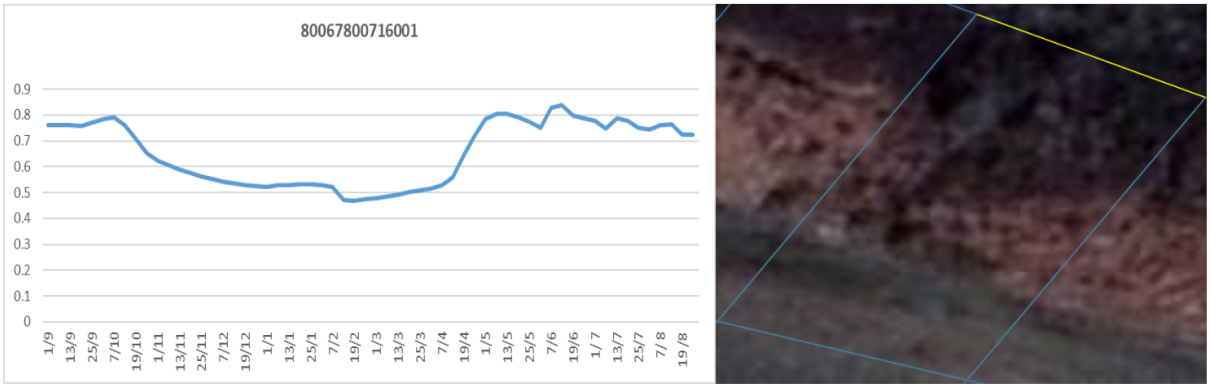


Figure 70: Detected data outlier. Non eligible crop type declaration.

These data quality control procedures resulted in the exception of several parcels from the sample, which can be clearly indicated in the following tables for the pilot cases for the years 2022 and 2023.

Table 34: Data Outliers Descriptive Statistics Distribution of 2022 pilot parcels - Non Eligible data.

Data Outliers - Distribution of parcel count and area per crop - 2022						
	Conventional		Organic		Total Area	Total Count
Crop Type	Area (Ha)	Count	Area (Ha)	Count		
Maize	0	0	99.6302	10	**99.6302**	**10**
Soybean	0	0	73.7966	23	**73.7966**	**23**
Sunflower	1.2597	2	13.8872	2	**15.1469**	**4**
Wheat	0.5466	1	50.4439	30	**50.9905**	**31**
Total Result	**1.8063**	**3**	**237.7579**	**65**	**239.5642**	**68**

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

Table 35: Data Outliers Descriptive Statistics Distribution of 2023 pilot parcels - Non Eligible data.

Distribution of parcel count and area per crop - OUTLIERS						
	Conventional		Organic		Total Area	Total Count
Crop Type	Area (Ha)	Count	Area (Ha)	Count		
Maize	0	0	0	0	**0**	**0**
Soybean	0	0	0	0	**0**	**0**
Sunflower	1.1969	2	2.4806	7	**3.6775**	**9**
Wheat	1.267	2	24.3806	22	**25.6476**	**24**
Total Result	**2.4639**	**4**	**26.8612**	**29**	**29.3251**	**33**

The final distribution of data from the 2022/2023 pilots as formulated by the verification procedures described previously, is recorded in the tables that follow. The number and area of pilot plots by farming practice, in the different crops and geographical areas are listed.

Table 36: Descriptive Statistics Distribution of 2022 pilot parcels - Cleaned dataset.

Distribution of parcel count and area per crop - CLEANED DATA						
	Conventional		Organic		Total Area	Total Count
Crop Type	Area (Ha)	Count	Area (Ha)	Count		
Maize	0	0	303.05	44	**303.05**	**44**
Soybean	2.56	4	256.22	58	**258.78**	**62**
Sunflower	4.56	7	391.98	85	**396.54**	**92**
Wheat	8.5	17	1019.42	119	**1027.92**	**136**
Total Result	**15.62**	**28**	**1970.67**	**306**	**1986.29**	**334**

Table 37: Descriptive Statistics Distribution of 2023 pilot parcels - Cleaned dataset.

Distribution of parcel count and area per crop - CLEANED DATA						
	Conventional		Organic		Total Area	Total Count
Crop Type	Area (Ha)	Count	Area (Ha)	Count		
Maize	1.383	4	32.644	15	**34.027**	**19**
Soybean	0	0	165.738	26	**165.738**	**26**
Sunflower	8.732	16	174.998	38	**183.73**	**54**
Wheat	5.258	8	350.45	117	**355.708**	**125**
Total Result	**15.373**	**28**	**723.83**	**196**	**739.203**	**224**

Table 38: Geographical distribution of parcel area per crop of 2022 pilot parcels - Cleaned dataset.

Geographical distribution of parcel area per crop -2022						
Geographic Districts	Farming Practice	Area (Ha)				Total Area (Ha)
		Maize	Soybean	Sunflower	Wheat	
Central Banat	Conventional	0	2.56	4.56	8.5	**15.62**
Central Banat	Organic	3.14	0	0	0.75	**3.89**
North Banat	Organic	1.3	0	0	3.07	**4.37**

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

South Banat	Organic	35.55	93.6	272.67	941.86	**1343.68**
South Backa	Organic	263.06	160.28	115.22	73.18	**611.74**
West Backa	Organic	0	2.34	4.09	0.56	**6.99**
Total Result		**303.05**	**258.78**	**396.54**	**1027.92**	**1986.29**

Table 39: Geographical distribution of parcel count per crop of 2022 pilot parcels - Cleaned dataset.

Geographical distribution of parcel count per crop -2022						
Geographic Districts	**Farming Practice**	**Area (Ha)**				**Total Area (Ha)**
		Maize	**Soybean**	**Sunflower**	**Wheat**	
Central Banat	Conventional	0	4	7	17	**28**
Central Banat	Organic	12	0	0	4	**16**
North Banat	Organic	3	0	0	9	**12**
South Banat	Organic	2	45	75	95	**217**
South Backa	Organic	27	11	8	10	**56**
West Backa	Organic	0	2	2	1	**5**
Total Result		**44**	**62**	**92**	**136**	**334**

Table 40: Geographical distribution of parcel area per crop of 2022 pilot parcels - Cleaned dataset.

Geographical distribution of parcel area per crop - 2023						
		Area (Ha)				
Geographic Districts	**Farming Practice**	**Maize**	**Soybean**	**Sunflower**	**Wheat**	**Total Area (Ha)**
Central Banat	Conventional	1.383	0	8.732	5.258	**15.373**
Central Banat	Organic	23.461	6.899	0	0	**30.36**
North Banat	Organic	0.969	0	0	2.234	**3.203**
South Banat	Organic	4.201	0	76.519	255.996	**336.716**
South Backa	Organic	4.013	158.839	98.479	92.22	**353.551**
Total Result		**34.027**	**165.738**	**183.73**	**355.708**	**739.203**

Table 41: Geographical distribution of parcel count per crop of 2022 pilot parcels - Cleaned dataset.

Geographical distribution of parcel count per crop - 2023						
Geographic Districts	**Farming Practice**	**Area (Ha)**				**Total Area (Ha)**
		Maize	**Soybean**	**Sunflower**	**Wheat**	
Central Banat	Conventional	0	4	7	17	**28**
Central Banat	Organic	12	0	0	4	**16**
North Banat	Organic	3	0	0	9	**12**
South Banat	Organic	2	45	75	95	**217**
South Backa	Organic	27	11	8	10	**56**
Total Result		**44**	**60**	**90**	**135**	**329**

It became apparent from the above evidence that the pilot data did not meet the data recommendations set for ML model training and for achieving a desired level of accuracy. Although there are no minimum requirements of the service about the quantity and quality of the input data,

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

these issues definitely have a big effect on the algorithm performance. Therefore, there are data size recommendations that relate with the spatial extent of the dataset (how localised is the dataset in regards with soil & climatic conditions and their effect on crop growth) and its sub-stratification within crop varieties, with a "lose" rule of thumb of at least "50 times the number of involved crop varieties" parcel samples. The data quality is influenced by the balance between the number of samples within organic/conventional classes, and the total samples for each variety, aiming for a uniform data distribution. For this important reason, the decision was made regarding Data Product Validation as follows:

- to not carry out, as is normally done, training with the data of pilot plots, but direct inference on them using the most accurate models trained with historical data of previous years, extraction of validation metrics on the unseen data of 2022/2023. In this way to test the work hypothesis of whether an accurate model can have good generalisation qualities not only on unseen data of the same year, but also on subsequent growing seasons.
- to validate the traffic light system for the Identification of Organic Farming Practices at the farm level.

4.2. Criteria

The data product was validated in terms of its accuracy relative to ground truth data from the LPIS and the GSA that are considered the standard, against which the product is compared. Ground truth data kept out of the algorithm training and are considered as "unseen" by the model. Normally they are partitioned out of the input dataset, as a test set, in order to give an unbiased evaluation of the model predictive qualities. For reasons already mentioned in the previous chapter, that relate to the quantity of the pilot dataset, data from 2022 and 2023 seasons were used to validate those historical data models which showed superior accuracy. Accuracy evaluation was conducted with these datasets as "unseen", firstly on a pixel level, and subsequently the traffic light system was evaluated on a parcel level. Confusion matrix notation and terminology was used for this aim. In the problem of statistical classification, a confusion matrix, also known as error matrix is a specific table layout that allows visualisation of the performance of an algorithm, typically a supervised learning one. Each row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class, or vice versa – both variants are found in the literature. The name stems from the fact that it makes it easy to see whether the system is confusing two classes.

To assess the accuracy of our classification scheme, the following metrics were acquired:

- Overall Accuracy
- Precision: Precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances
- Recall: (true positive rate) is the probability of a positive test result, conditioned on the individual truly being positive.
- Specificity: (true negative rate) is the probability of a negative test result, conditioned on the individual truly being negative
- Balanced Accuracy: The balanced accuracy in binary and multiclass classification problems to deal with imbalanced datasets. It is defined as the average of recall obtained in each class.
- F1 Score: also known as balanced F-score or F-measure. The F1 score can be interpreted as a harmonic mean of the precision and recall

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

4.3. Validation Methodology and results

In this chapter the methodology applied for the validation of the product D5: Distinction of Organic Farming Practices in the pilot business cases of the Envision project is developed and subsequently the results of the data validation process are presented. This process aims to test the accuracy with which ML classification models trained from input data can predict farming practice in cases of parcel data that have not participated in the training.

The elaboration of the algorithm training/evaluation methodology followed and implemented in the historical data models, is explained in detail in D3.5, D3.7 and the most recent D1.9 Reporting of 2nd Reporting Period. The validation of the historical data models is thoroughly elaborated on the recent D1.9.

For the inference task, all the steps of the successive components of the data processing flow of the service, described in the above deliverables, were followed, concerning:

- Field Data Import: LPIS+GSA data subset import to the database
- Spatial Data POSTGIS processing + Descriptive Stats
- EO Data Import: Data import from CreoDIAS and Copernicus Dataspace APIs
- SoilGrids Import
- EO Feature Engineering: Vegetation and Image Texture indice timeseries and derivatives
- Data Outlier/Anomaly Detection

For prediction of the classification layer, it was considered to use the most accurate of the models trained in the past, with the use of historical datasets. While the evaluation of these Classification models is presented in detail in D1.9 Reporting of 2nd Reporting Period, a summary of the error metrics is given in the following table.

Table 42: Test Set Validation Results for all crop/year scenarios. Confusion Matrix Error Metrics from Historical data classification models.

Model Training Scenarios	Test Validation Confusion Matrix Metrics					
	Overall Accuracy	Recall	Specificity	Balanced Accuracy	Precision	F1
Maize Samples 2016 Early Season Cycle	0.9141	0.9856	0.6766	0.8311	0.9101	0.9464
Maize Samples 2016 Full Season Cycle	0.9474	0.9622	0.8982	0.9302	0.9622	0.9622
Soybean Samples 2016 Early Season Cycle	0.9624	0.9735	0.8632	0.9183	0.9845	0.9790
Soybean Samples 2016 Full Season Cycle	0.9746	0.9774	0.9438	0.9606	0.9947	0.9860
Sunflower Samples 2019 Early Season Cycle	0.8779	0.9907	0.5333	0.7620	0.8665	0.9244
Sunflower Samples 2019 Full Season Cycle	0.9554	0.9782	0.8857	0.9320	0.9782	0.9782
Sunflower Samples 2020 Early Season Cycle	0.9677	0.9894	0.5217	0.7556	0.9770	0.9832
Sunflower Samples 2020 Full Season Cycle	0.9798	0.9894	0.7826	0.8860	0.9894	0.9894
Sunflower Samples Merged Early Season Cycle	0.9374	0.9576	0.5217	0.7397	0.9762	0.9668
Sunflower Samples Merged Full Season Cycle	0.9818	0.9873	0.8696	0.9284	0.9936	0.9904
Wheat Samples 2018 Early Season Cycle	0.9874	0.9917	0.5714	0.7816	0.9955	0.9936
Wheat Samples 2018 Full Season Cycle	0.9874	0.9910	0.6429	0.8169	0.9962	0.9936
Wheat Samples 2020 Early Season Cycle	0.9635	0.9908	0.4846	0.7377	0.9712	0.9809
Wheat Samples 2020 Full Season Cycle	0.9668	0.9886	0.5846	0.7866	0.9766	0.9826
Wheat Samples Merged Early Season Cycle	0.9863	0.9932	0.3448	0.6690	0.9929	0.9931
Wheat Samples Merged Full Season Cycle	0.9665	0.9842	0.4151	0.6997	0.9812	0.9827

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

What is evident from the historical data classification models Error Metrics is that the most evaluative cases were:

- Maize Crop (2016) - Early/Full Season Models
- Soybean Crop (2016) - Early/Full Season Models
- Sunflower Crop (2019) - Early/Full Season Models
- Wheat (2018) - Early/Full Season Models

EO Features were processed for each Geographic District – Area of Interest. Regarding the temporal coverage of the EO Features produced, the dates defining the Early/Full Prediction seasons are presented on the following table.

Table 43: Temporal coverage of Crop Seasons in ML models.

Crop	Season Start	Early Prediction	Full Season End
Maize	01 February	25 July	01 October
Soybean	01 February	25 July	01 October
Sunflower	01 January	25 July	01 November
Wheat	01 September	01 June	01 August

Model inference was predicted for each crop, at a spatial resolution of 10m, mapping the classification probability (p-values) for discriminating organic from conventional farming practice. At this point, a threshold p-value was determined, for the discretization of the classification result on boolean decision values. The classification probability threshold was decided by optimization after visual inspection and analysis of the ROC curves of the Internal Cross Validation of the models. It was set to a value that minimised False Positive Rate and maximised True Positive Rate, that approximated 0.55.

Evaluation was initially performed at a pixel level. All pixels inside a parcel were labelled according to the farming practice reference declaration, given the assumption that every parcel involves one crop, which was assessed during outlier detection that it holds true. The binary classification inference results were extracted for each parcel with the use of zonal area tabulation. The evaluation of predicted vs reference was based on the confusion matrix created for each classification model, and validation metrics were calculated.

Pilot cases 2022 - Confusion Matrices

Table 44: Pixel based Confusion Matrix of Maize 2016 Early Season Model.

Maize 2016 Early Season Model		
	Predicted Class	
Declared Class	Organic	Conventional
Organic	10569	15910
Conventional	0	0
Total Pixel Result	10569	15910

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

Table 45: Pixel based Confusion Matrix of Maize 2016 Full Season Model.

Maize 2016 Full Season Model		
	Predicted Class	
Declared Class	Organic	Conventional
Organic	21914	4564
Conventional	0	0
Total Pixel Result	21914	4564

Table 46: Pixel based Confusion Matrix of Soybean 2016 Early Season Model.

Soybean 2016 Early Season Model		
	Predicted Class	
Declared Class	Organic	Conventional
Organic	22615	984
Conventional	321	10
Total Pixel Result	22936	994

Table 47: Pixel based Confusion Matrix of Soybean 2016 Full Season Model.

Soybean 2016 Full Season Model		
	Predicted Class	
Declared Class	Organic	Conventional
Organic	18938	4660
Conventional	287	44
Total Pixel Result	19225	4704

Table 48: Pixel based Confusion Matrix of Sunflower 2019 Early Season Model.

Sunflower 2019 Early Season Model		
	Predicted Class	
Declared Class	Organic	Conventional
Organic	14270	34679
Conventional	283	144
Total Pixel Result	14553	34823

Table 49: Pixel based Confusion Matrix of Sunflower 2019 Full Season Model.

Sunflower 2019 Full Season Model		
	Predicted Class	
Declared Class	Organic	Conventional
Organic	14804	34142
Conventional	113	314

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

Total Pixel Result	**14917**	**34456**

Table 50: Pixel based Confusion Matrix of Wheat 2018 Early Season Model.

Wheat 2018 Early Season Model		
	Predicted Class	
Declared Class	**Organic**	**Conventional**
Organic	124268	661
Conventional	842	196
Total Pixel Result	**125110**	**857**

Table 51: Pixel based Confusion Matrix of Wheat 2018 Full Season Model.

Wheat 2018 Full Season Model		
	Predicted Class	
Declared Class	**Organic**	**Conventional**
Organic	124375	552
Conventional	1037	0
Total Pixel Result	**125412**	**552**

Table 52: Evaluation of Historical Data Classification Models with 2022 unseen pilot data.

Prediction Model	**Percentage of pixels that are declared organic and are classified as organic (%) {Recall}**	**Percentage of pixels that are declared conventional and are classified as conventional (%) {Specificity}**	**Percentage of pixels that are classified as organic and are declared organic (%) {Precision}**
Maize			
Maize 2016 Full Season	82.7	N/A	100
Maize 2016 Early Season	39.9	N/A	100
Soybean			
Soybean 2016 Full Season	80.2	13.2	98.5
Soybean 2016 Early Season	95.8	3.02	98.6
Sunflower			
Sunflower 2019 Full Season	30.2	73.5	99.2
Sunflower 2019 Early Season	29.1	33.7	98.05

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

Wheat			
Wheat 2018 Full Season	99.5	0	99.1
Wheat 2018 Early Season	99.4	18.8	99.3

Pilot cases 2023 - Confusion Matrices

Table 53: Pixel based Confusion Matrix of Maize 2016 Early Season Model.

Maize 2016 Early Season Model		
	Predicted Class	
Declared Class	Organic	Conventional
Organic	3628	652
Conventional	122	57
Total Pixel Result	3750	709

Table 54: Pixel based Confusion Matrix of Soybean 2016 Early Season Model.

Soybean 2016 Early Season Model		
	Predicted Class	
Declared Class	Organic	Conventional
Organic	20416	1163
Conventional	0	0
Total Pixel Result	20416	1163

Table 55: Pixel based Confusion Matrix of Sunflower 2019 Early Season Model.

Sunflower 2019 Early Season Model		
	Predicted Class	
Declared Class	Organic	Conventional
Organic	8735	13958
Conventional	167	976
Total Pixel Result	8902	14934

Table 56: Pixel based Confusion Matrix of Wheat 2018 Early Season Model.

Wheat 2018 Early Season Model		
	Predicted Class	
Declared Class	Organic	Conventional
Organic	36340	9027
Conventional	181	491
Total Pixel Result	36521	9518

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

Table 57: Pixel based Confusion Matrix of Wheat 2018 Full Season Model.

Wheat 2018 Full Season Model		
	Predicted Class	
Declared Class	Organic	Conventional
Organic	44923	447
Conventional	672	0
Total Pixel Result	45595	447

Table 58: Evaluation of Historical Data Classification Models with 2023 unseen pilot data.

Prediction Model	Percentage of pixels that are declared organic and are classified as organic (%) {Recall}	Percentage of pixels that are declared conventional and are classified as conventional (%) {Specificity}	Percentage of pixels that are classified as organic and are declared organic (%) {Precision}
Maize			
Maize 2016 Early Season	84.76	31.84	96.74
Soybean			
Soybean 2016 Early Season	94.61		100
Sunflower			
Sunflower 2019 Early Season	38.49	85.38	98.12
Wheat			
Wheat 2018 Full Season	99.01	0	98.52
Wheat 2018 Early Season	80.1	73.06	99.5

Traffic Light System Evaluation

The results of classification are provided in the data product as a vector geospatial feature which is served through the ENVISION platform, or alternatively could be served via WFS to the user, in a shapefile format. It contains the parcel polygon boundary geometries, and as far as attributes, the evaluation of the Classification as Organic/Conventional farming practice, in the representation of a traffic light system. Its values are given in a standardised confusion matrix terminology, and depict the result of the prediction in regards with what was initially declared. The prediction is decided by a configured threshold value on the classification probability, which is the actual output of the algorithm. If the spatial average within the parcel bounds is higher than 0.5 the parcel is inferred as organic. The traffic light values are the following:

- False Negative (FN) also known as type II underestimation error if a parcel was predicted conventional while being organic

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

- False Positive (FP) also known as type I overestimation error if a parcel was predicted organic while being conventional
- True Negative (TN) if a parcel was predicted correctly as conventional
- True Positive (TP) if a parcel was predicted correctly as organic

The outcome of the Traffic Light System for the 2022 and the 2023 pilots is given in a series of maps, on the annex of this deliverable. The symbology representation of it can be observed on the following image.

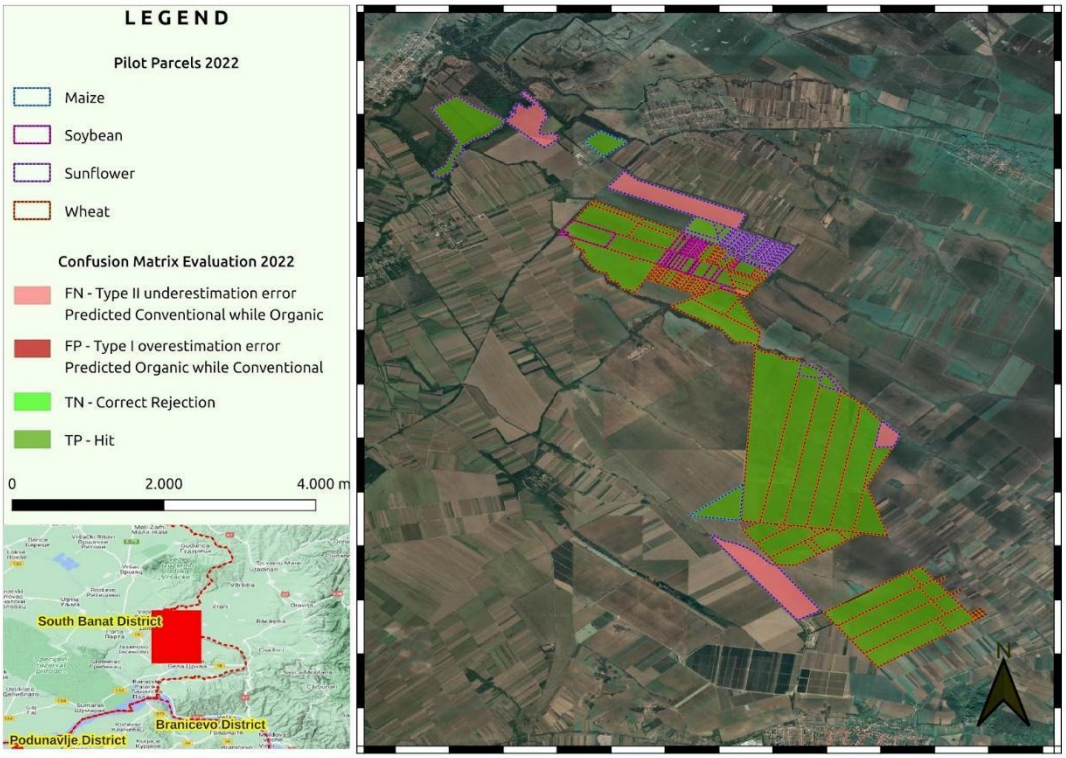


Figure 71: Symbology representation of the D5 data product traffic light system.

A second level of evaluation regarded the traffic light system, and validated the performance among the different geographic regions of Serbia, in terms of parcel based accuracy metrics. On the tables that follow, a parcel based evaluation of the D5 product service "Distinction of Organic Farming Practices" is presented for the 2022 and 2023 pilot parcels.

Pilot cases 2022 - Traffic Light System Evaluation - Predictions vs. Declarations

Table 59: Evaluation of Data Product Traffic Light System - Validation of 2016 Maize Classification Model with 2022 pilot data.

Maize - 2022 pilot parcels											
Evaluation / Geographic Districts	TP	TN	FP	FN	Total Result	Precision	Recall	F1	Specificity	Balanced ACC	ACC
Central Banat	11	0	0	1	**12**	1	0.916	0.956	N/A	N/A	0.916
North Banat	3	0	0	0	**3**	1	1	1	N/A	N/A	1

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

South Banat	2	0	0	0	**2**	1	1	1	N/A	N/A	1
South Backa	23	0	0	4	**27**	1	0.851	0.92	N/A	N/A	0.85
West Backa	0	0	0	0	**0**	N/A	N/A	N/A	N/A	N/A	N/A
Total Result	**39**	**0**	**0**	**5**	**44**	**1**	**0.89**	**0.94**	**N/A**	**N/A**	**0.89**

Table 60: Evaluation of Data Product Traffic Light System - Validation of 2016 Soybean Classification Model with 2022 pilot data.

Soybean - 2022 pilot parcels											
Evaluation											
Geographic Districts	**TP**	**TN**	**FP**	**FN**	**Total Result**	**Precision**	**Recall**	**F1**	**Specificity**	**Balanced ACC**	**ACC**
Central Banat	0	0	4	0	**4**	0	N/A	N/A	0	N/A	0
North Banat	0	0	0	0	**0**	N/A	N/A	N/A	N/A	N/A	N/A
South Banat	45	0	0	0	**45**	1	1	1	N/A	N/A	1
South Backa	4	0	0	7	**11**	1	0.36	0.53	N/A	N/A	0.363
West Backa	2	0	0	0	**2**	1	1	1	N/A	N/A	1
Total Result	**51**	**0**	**4**	**7**	**62**	**0.93**	**0.88**	**0.9**	**0**	**0.44**	**0.82**

Table 61: Evaluation of Data Product Traffic Light System - Validation of 2019 Sunflower Classification Model with 2022 pilot data.

Sunflower - 2022 pilot parcels											
Evaluation											
Geographic Districts	**TP**	**TN**	**FP**	**FN**	**Total Result**	**Precision**	**Recall**	**F1**	**Specificity**	**Balanced ACC**	**ACC**
Central Banat	0	7	0	0	**7**	N/A	N/A	N/A	1	N/A	1
North Banat	0	0	0	0	**0**	N/A	N/A	N/A	N/A	N/A	N/A
South Banat	47	0	0	28	**75**	1	0.626	0.77	N/A	N/A	0.62
South Backa	0	0	0	8	**8**	N/A	0	N/A	N/A	N/A	0
West Backa	2	0	0	0	**2**	1	1	1	N/A	N/A	1
Total Result	**49**	**7**	**0**	**36**	**92**	**1**	**0.58**	**0.73**	**1**	**0.79**	**0.61**

Table 62: Evaluation of Data Product Traffic Light System - Validation of 2018 Wheat Classification Model with 2022 pilot data.

Wheat - 2022 pilot parcels											
Evaluation											
Geographic Districts	**TP**	**TN**	**FP**	**FN**	**Total Result**	**Precision**	**Recall**	**F1**	**Specificity**	**Balanced ACC**	**ACC**
Central Banat	4	0	17	0	**21**	0.19	1	0.32	0	N/A	0.19
North Banat	9	0	0	0	**9**	1	1	1	N/A	N/A	1
South Banat	94	0	0	1	**95**	1	0.98	0.99	N/A	N/A	0.98
South Backa	10	0	0	0	**10**	1	1	1	N/A	N/A	1
West Backa	1	0	0	0	**1**	1	1	1	N/A	N/A	1
Total Result	**118**	**0**	**17**	**1**	**136**	**0.87**	**0.99**	**0.93**	**0**	**0.5**	**0.87**

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

Pilot cases 2023 - Traffic Light System Evaluation - Predictions vs. Declarations

Table 63: Evaluation of Data Product Traffic Light System - Validation of 2016 Maize Classification Model with 2023 pilot data.

Maize - 2023 pilot parcels											
Evaluation / Geographic Districts	TP	TN	FP	FN	Total Result	Precision	Recall	F1	Specificity	Balanced ACC	ACC
Central Banat	4	0	0	0	**4**	1	1	1	N/A	N/A	1
North Banat	2	0	0	0	**2**	1	1	1	N/A	N/A	1
South Banat	2	0	0	0	**2**	1	1	1	N/A	N/A	1
South Backa	7	0	0	0	**7**	1	1	1	N/A	N/A	1
Total Result	**15**	**0**	**0**	**0**	15	**1**	**1**	**1**	**N/A**	**N/A**	**1**

Table 64: Evaluation of Data Product Traffic Light System - Validation of 2016 Soybean Classification Model with 2023 pilot data.

Soybean - 2023 pilot parcels											
Evaluation / Geographic Districts	TP	TN	FP	FN	Total Result	Precision	Recall	F1	Specificity	Balanced ACC	ACC
Central Banat	3	0	0	0	**3**	1	1	1	N/A	N/A	1
North Banat	0	0	0	0	**0**	N/A	N/A	N/A	N/A	N/A	N/A
South Banat	0	0	0	0	**0**	N/A	N/A	N/A	N/A	N/A	N/A
South Backa	22	0	0	1	**23**	1	0.956	0.977	N/A	N/A	0.956
Total Result	**25**	**0**	**0**	**1**	26	**1**	**0.96**	**0.98**	**N/A**	**N/A**	**0.96**

Table 65: Evaluation of Data Product Traffic Light System - Validation of 2019 Sunflower Classification Model with 2023 pilot data.

Sunflower - 2023 pilot parcels											
Evaluation / Geographic Districts	TP	TN	FP	FN	Total Result	Precision	Recall	F1	Specificity	Balanced ACC	ACC
Central Banat	0	15	1	0	**16**	0	N/A	N/A	0.93	N/A	0.93
North Banat	0	0	0	0	**0**	N/A	N/A	N/A	N/A	N/A	N/A
South Banat	19	0	0	8	**27**	1	0.703	0.826	N/A	N/A	0.703
South Backa	0	0	0	11	**11**	N/A	0	N/A	N/A	N/A	0
Total Result	**19**	**15**	**1**	**19**	54	**0.95**	**0.5**	**0.66**	**0.94**	**0.72**	**0.63**

Table 66: Evaluation of Data Product Traffic Light System - Validation of 2018 Wheat Classification Model with 2023 pilot data.

Wheat - 2023 pilot parcels											
Evaluation	TP	TN	FP	FN		Precision	Recall	F1	Specificity		ACC

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

Geographic Districts					Total Result					Balanced ACC	
Central Banat	0	0	8	0	**8**	0	N/A	N/A	0	N/A	0
North Banat	6	0	0	0	**6**	1	1	1	N/A	N/A	1
South Banat	100	0	0	0	**100**	1	1	1	N/A	N/A	1
South Backa	11	0	0	0	**11**	1	1	1	N/A	N/A	1
Total Result	**117**	**0**	**8**	**0**	**125**	**0.94**	**1**	**0.97**	**0**	**0.5**	**0.94**

4.4. Discussion

In the following discussion section, a review of the conclusions drawn in the successive stages of the product's development and evolution as well as in the final evaluation of the traffic light system. A technological roadmap, how the five developed products could be implemented in the area of another member state: standardisation of input data, limits of the algorithm, minimum requirements of training data, schemas how the achieved accuracy and the decision thresholds can be combined.

The task of distinguishing organic from conventional farming practices with the use of EO data was indeed very challenging. Initially, regarding the strategy, decisions had to be made about what EO derived classification features to use for the discrimination. Data dimensionality and more practical reasons regarding the spatial data extent and the available data space posed a certain limit as to how many features to use. The discussion was about whether to focus more on the spectral or the spatiotemporal content of the EO data. Clear scientific evidence about “a defined spectral signature” of an organic farming practice wasn’t found in the literature, rather than some few experimental cases that focused solely on crop/leaf canopy nutrient content. These studies used very high resolution multispectral and hyperspectral data, questioned the problem on specific crop varieties and on a highly local scale experimental plots, relying on abundant ground truth data about nutrient NPK inputs. On Envision, it was known from the start, that such in situ data were not available at a national scale. It was finally resolved, to focus more on the spatiotemporal aspects of vegetation phenology in the EO signal.

In the successive benchmarks and iterations carried out to train the ML algorithms, the central issue was the "Bias-Variance Trade-off" that was noted. During the 1st iteration of ML model training, error values greater than the acceptable threshold were observed. In other words, the model used was not strong enough to produce an accurate prediction. Therefore, this was a case of high bias, which was addressed by increasing the complexity of the models. The number of features was significantly increased by including NDVI Derivatives and Image Texture metrics. A dimensionality reduction was performed and a Boosting Trees classification algorithm was chosen instead of the original SVM, which is suitable for cases of high bias.

At the 2nd iteration, the behaviour of the models was indicative of a case of high variance and a high tendency for over-fitting. The regularisation parameters were chosen in a range of values that made the modelling less "aggressive" in memorising the dataset. Unfortunately, the lack of samples with uniformity of distribution by crop variety and also by geographical distribution became apparent.

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

However, the performance of the models on unseen datasets was not the same across all crops. In the case of Sunflower for 2019 and 2020, the results showed quite good performance and qualities in relation to their over/under estimation (type I & II errors). Their efficiency was quite stable, regardless of the year, both in early and full season prediction.

This is an important finding because it shows that when the algorithms are trained in a year specific manner, with enough data, diverse and balanced in terms of the underlying crop variety, and localised in a limited geographical space, they can give promising perspectives and good results, contributing with their predictions to the specific business case, be it mid or full season planning. This seems to indicate that the resulting product could be viable under certain assumptions regarding the input data it receives, and the accepted error thresholds set by its design.

The generalisation properties of the models, regarding prediction on other years/seasons, was showcased on the pilot business cases validation for the 2022/2023 seasons, and elaborated on the current deliverable. The inference of models trained on data of other years yielded unsatisfactory results in the aspect of underestimating. Unfortunately. ground truth data on pilot years were scarce, unequally distributed, and with many outliers regarding the crop type. Thus, they were not enough to train year specific models over the pilot seasons. A few points to mention:

- Class imbalance and overall scarcity of data among crop varieties of the training datasets affect the predictive qualities of the models, which is evident in the moderate performance of Recall and Specificity metrics. It is worth mentioning that the crap variety is a very important attribute in the discrimination process and thus the dataset should be representative of this aspect.
- Model underestimation, the condition where a parcel was predicted conventional while being organic, gets even worse in the case when models trained with historical data aim to predict future instances of unseen data.
- Regionality of the training dataset also seems to play an important role due to the implications of local soil and climate conditions on vegetation phenology

Data Cleaning and quality control was a very important aspect of the service development due to findings of severe false crop type declaration. What was seen in many conventional/organic declarations was that there were incorrect entries in the crop type attribute. Initial efforts of Outlier detection with unsupervised methods did not yield good results. The hybrid approach was far more precise, but had the disadvantage of requiring the involvement of expert knowledge. Visual inspection is a relatively simple process that requires the user to have a fairly basic knowledge, which could possibly be trained, of what a timeseries NDVI profile displays in a crop, as well as the sowing/cutting dates in the area of interest. The user observes the profile of each field-sample in the training set, and scores it as to the correctness of its statement.
A critical issue is scaling up, and the impact on human resources needed for the hybrid data cleaning process. Ideally, to avoid any bias that may arise from continuous photo-interpretation, it would be legitimate to involve more than one user, in overlapping subsets. Considering that expert estimation could be provided by a limited number of users, it is implied that scale up is not linear, but on the

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

contrary, a large increase in the training set disproportionately increases the human resources required. An alternative could be a concurrent crop classification system, such as the one that was also developed in the project, or similar ones (Sen4Cap etc), and the selection of a training set with high confidence in crop type estimation (high p-value). But also in this case the presence of outliers in LPIS is significant and this has its impact on the quality of the estimation. In conclusion, an important data requirement that would enhance the business case would be the possibility for the user to upload a second auxiliary set of large-scale field visits data, to train the novelty detection algorithm and improve the training dataset without the need for visual inspection of the NDVI profile. This option wasn't developed during the project due to the absence of "real in situ" Field Visits data, but could definitely be a major improvement to the product.

Up to this point, many conclusions underline the difficulty of the task, the particularities of the input dataset and its quality control, the fine tuning of the models, etc. They regard issues that arise from the ML model training and propose recommendations for achieving an acceptable level of accuracy.

However, the concept behind the product is an Agriculture Monitoring Data Service. Respectfully each end user could provide regional-or national parcel data and farming practice declarations, employ EO features, Train/Tune-Validate/Evaluate a ML model and finally predict on a parcel level assessing a value from the traffic-light system that relates with its farming practice conformance. All the issues that came up in the Serbian pilot, and the experience gained through the project, led to the design and implementation of processing components and tools that were incorporated to the service in order to improve its functions. Any user could deploy a new project, train a new algorithm for a crop of interest, and predict on different regions, preferably geographically localised. Towards the end of the Envision project, evidence exists that a user following the recommended guidelines regarding data quality/quantity, feature engineering, and outlier detection could achieve an acceptable level of prediction accuracy. After all, the minimum acceptable error is user subjective and relates with the user's risk analysis of its business case.

A technological roadmap, how the five developed products could be implemented on the area of another member state includes:

Standardisation of input data: LPIS postgres tables/shapefiles/geojson files and GSA tables should be provided by the user with a harmonised content, which means that they should both share a primary key field of the same data type configuration. This would assure that the SQL join of the spatial and attribute tables would be successful. The table schema of the GSA should be proposed to the user to include the following fields

- Parcel ID (unique id - Primary Key)
- Applicant ID
- Year of Declaration
- Farming Practices: The declared type of farming practices (organic or conventional)
- Declared Crop Code
- Crop Variety

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

- Crop Type Category

Limitations of the algorithm: Regard the issues related with,

- The limited sample data support. The particularity of small and elongated parcels, and the boundary buffering that occurred in order to enhance the reliability of the sampled pixels, further decreased the total number of parcels.
- Uncertainty with regards the validity of the organic parcel declarations. It is believed that there were cases of wrongly declared parcels cultivation in order to comply with local subsidy regulations. The case of instances where several cultivations in the same parcel were declared as one, furtherly made things more complicated. Outlier analysis helped in order to filter out non reliable data, but due to its unsupervised nature, it performed well in the cases where the feature space data clouds had good separability properties.
- Spatial distribution of organic/conventional sample data of the same crop, shows that these may be located far apart from each other, and this fact may introduce small shifts on the timeseries temporal scale due to slightly different sowing/harvest dates related to local climatic conditions.

Minimum requirements of training data: There are no minimum requirements about the quantity and quality of the input data, but definitely these issues have a big effect on the algorithm performance. Therefore, there are data size recommendations that relate with the spatial extent of the dataset (how localised is the dataset in regards with soil & climatic conditions and their effect on crop growth) and its sub-stratification within crop varieties, with a “lose” rule of thumb of at least “50 times the number of involved crop varieties” parcel samples. The data quality is influenced by the balance between the number of samples within organic/conventional classes, and the total samples for each variety, aiming for a uniform data distribution.

Schemas how the achieved accuracy and the decision thresholds can be combined: the modification of the classification decision threshold, is done heuristically from the ROC curve of ML tuning cross validation. The outermost upper left point of the curve represents the p-threshold with the lowest error rates and the higher accuracy.

5. Conclusions

The D3.6 Data Products Validation Report provides a comprehensive overview of the data processing, model training, and evaluation methodologies employed for monitoring environmental practices in sustainable agriculture. The report outlines the various steps involved in the data processing flow, from data import to feature engineering and outlier detection. The evaluation of historical data classification models demonstrates the effectiveness of the methodologies in predicting agricultural practices, with specific crops and years highlighted as the most evaluative cases. Challenges, such as discrepancies in reported parcel data and spatial distribution of sample data, are discussed, emphasizing their impact on model performance. Recommendations concerning the quantity and quality of training data are also provided, suggesting a balanced and uniform data distribution for optimal performance. The report references multiple deliverables, ensuring a thorough understanding of the methods and evaluations used.

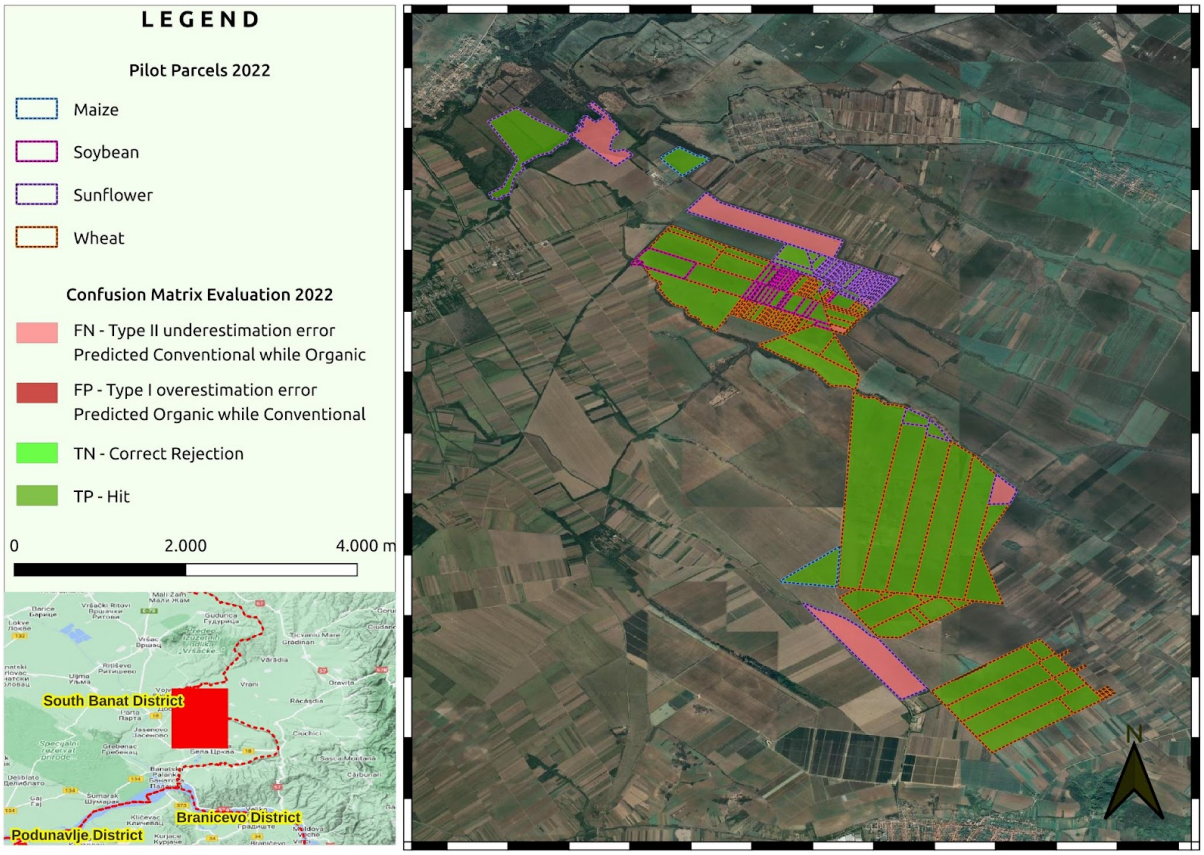
The ENVISION project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 869366

Moving forward, and as a roadmap for prospective projects or service providers, several recommendations emerge. Data augmentation techniques can be employed to support data diversity, effectively addressing discrepancies and spatial distribution challenges. There's potential in exploring new machine learning architectures to further improve accuracy and generalisability. A deeper temporal analysis, especially considering the impacts of climate change on crop yields, could offer valuable insights. Broadening the data sources, perhaps by including meteorological or socio-economic indicators, can provide holistic view. Consistent data quality can be maintained by establishing rigorous protocols for data collection, verification, and preprocessing. Furthermore, establishing a feedback mechanism will allow real-time feedback from farmers and stakeholders, refining model predictions and engaging agricultural experts, environmental scientists, and policymakers will ensure that the methodologies remain grounded in reality and can drive sustainable agricultural policies.

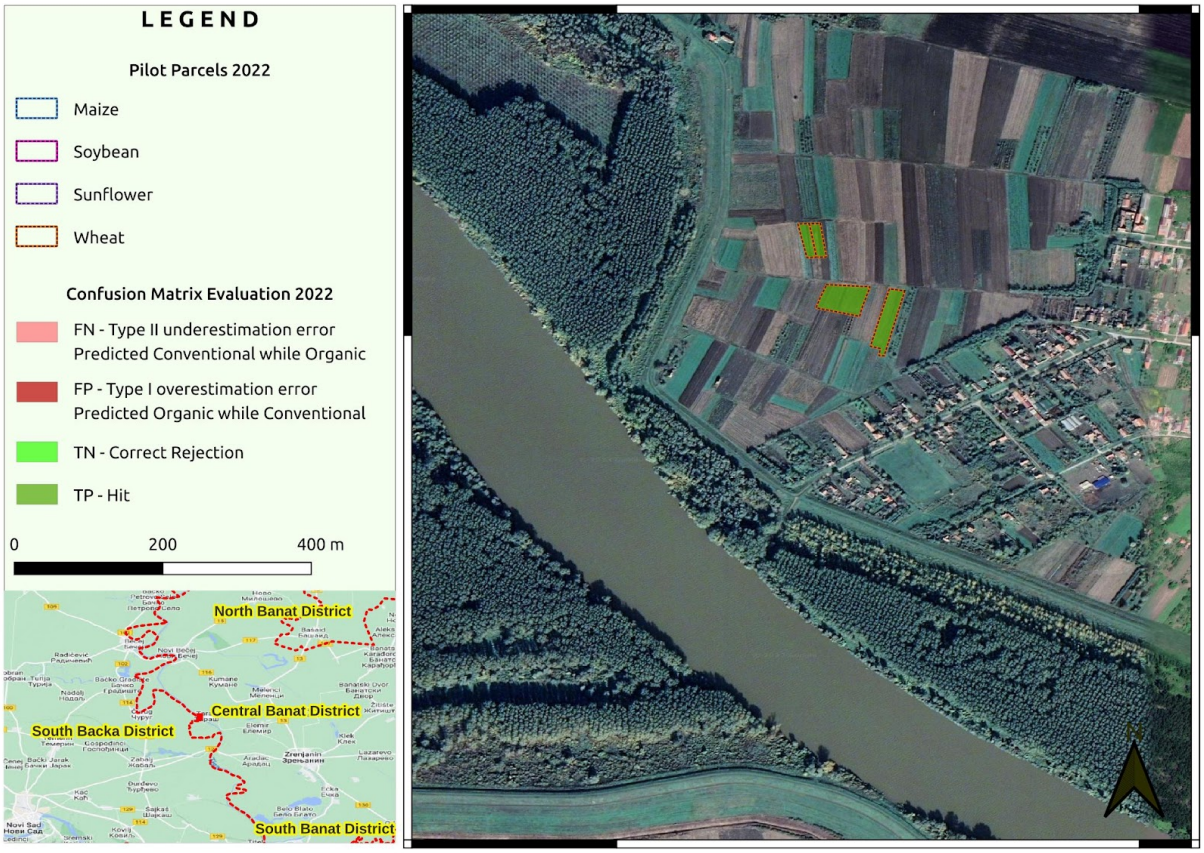
The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

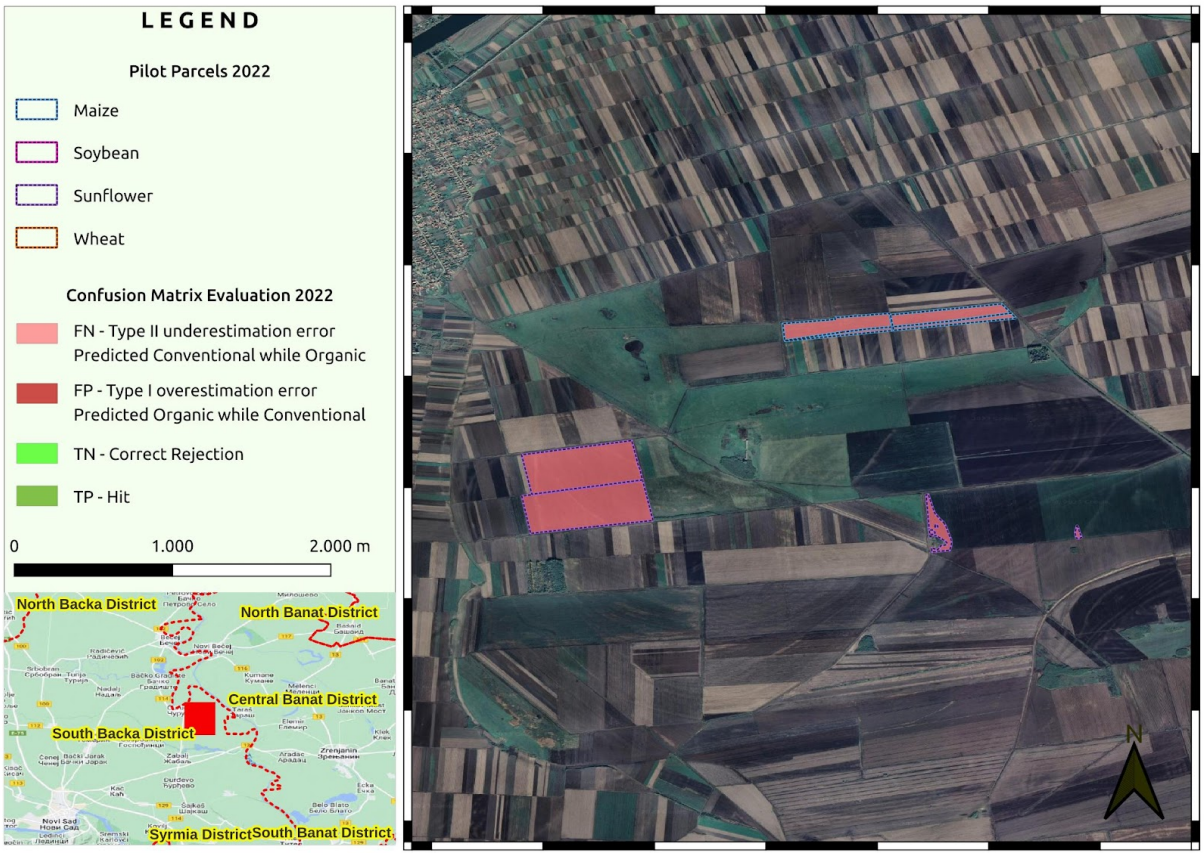
6. ANNEX:

BC4: Monitoring of organic farming requirements – Serbia: Evaluation Maps of the traffic Light System for the 2022 parcel data

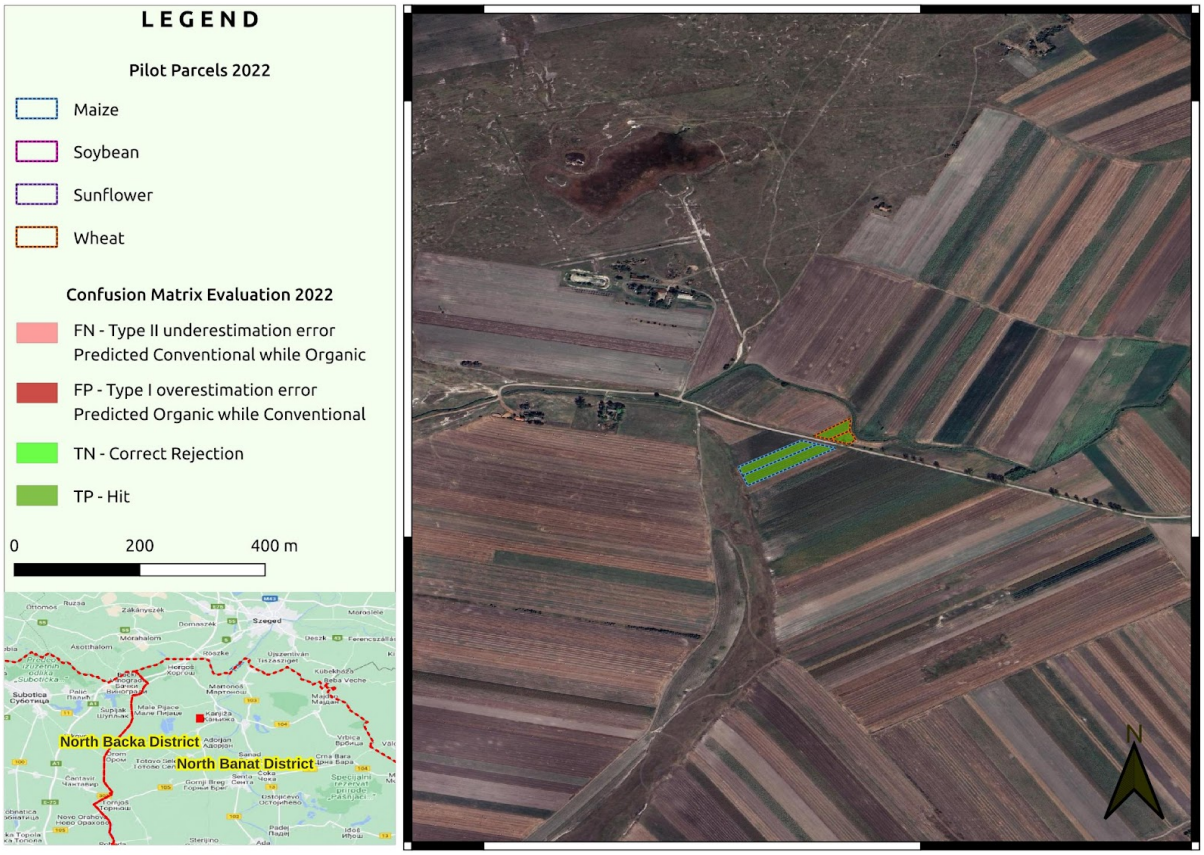


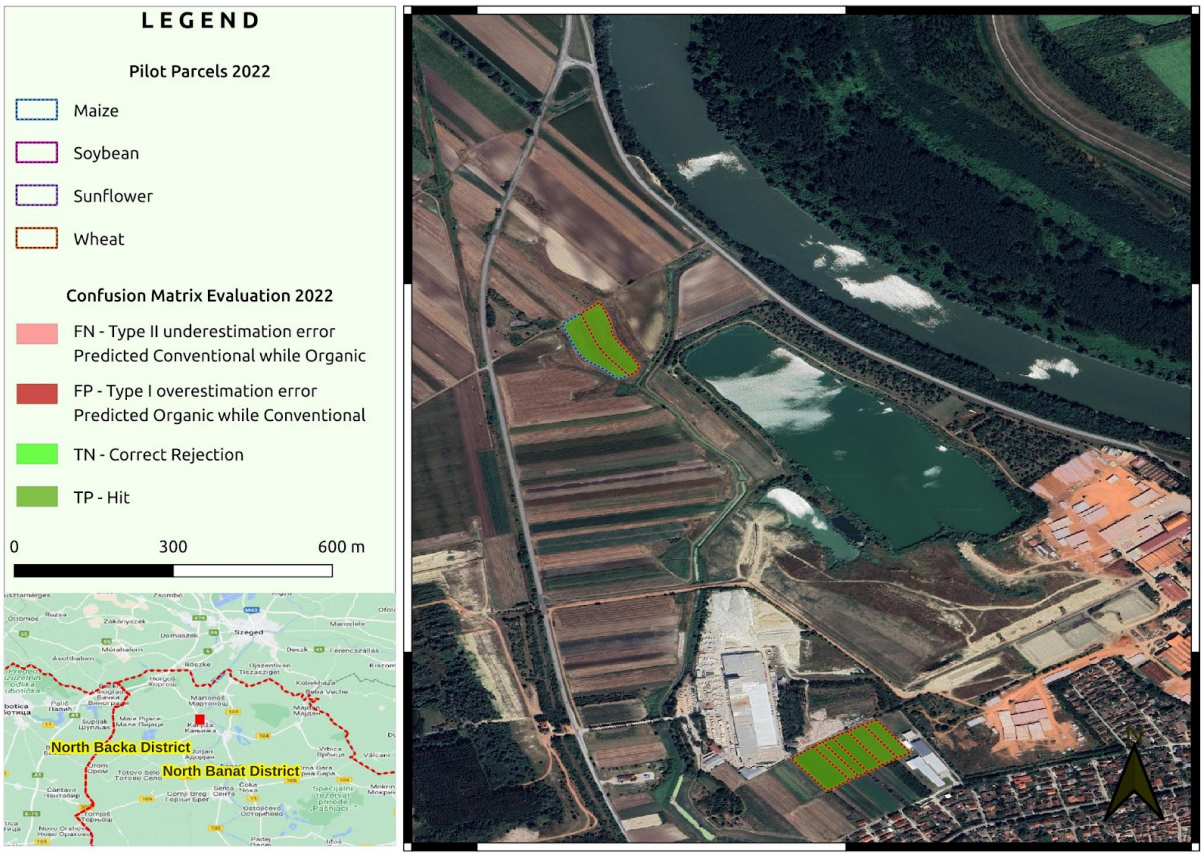
The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366



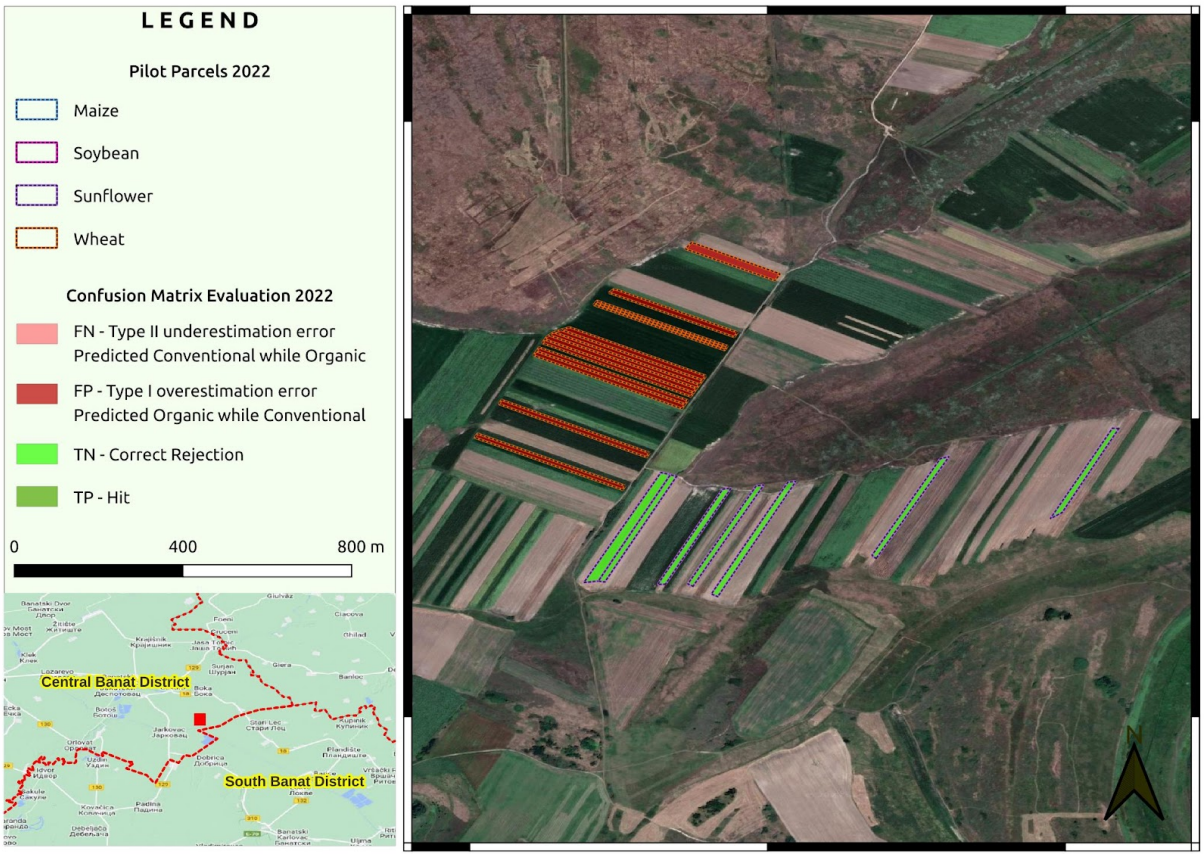


The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366





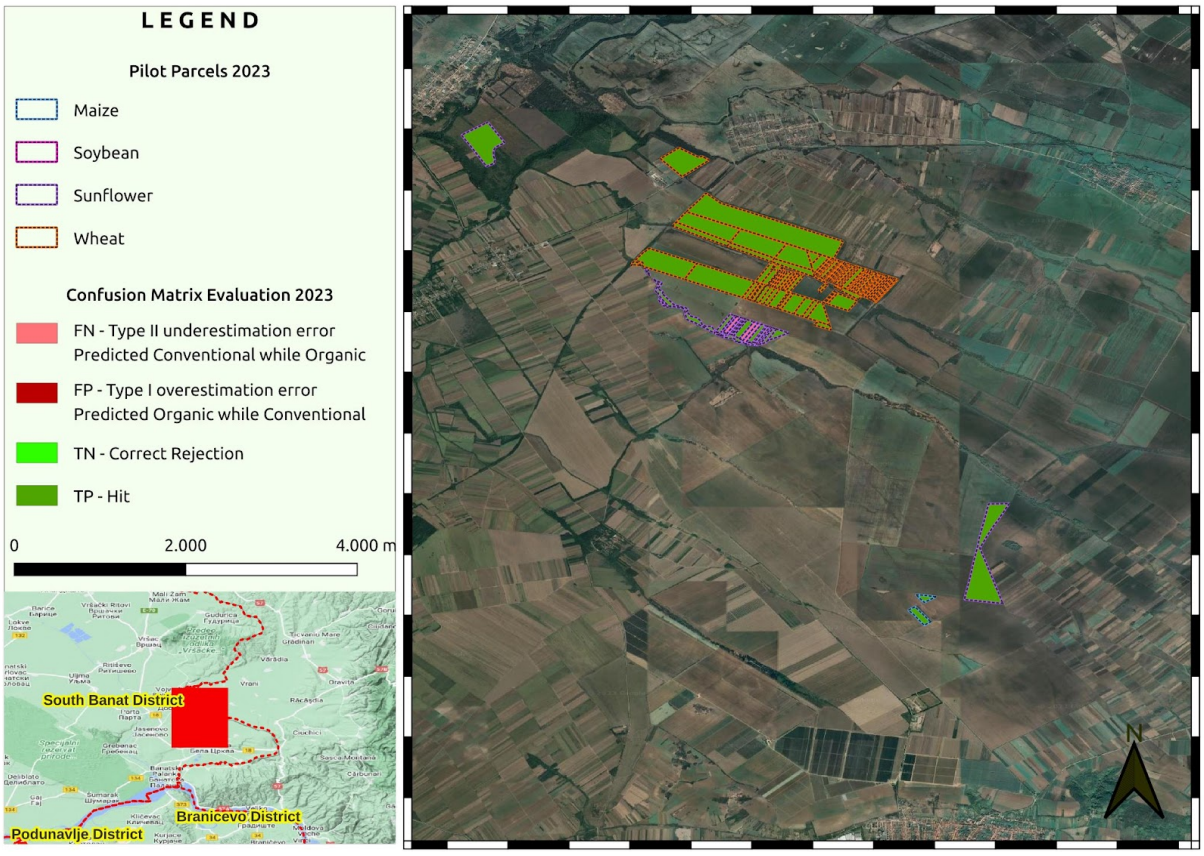
The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366



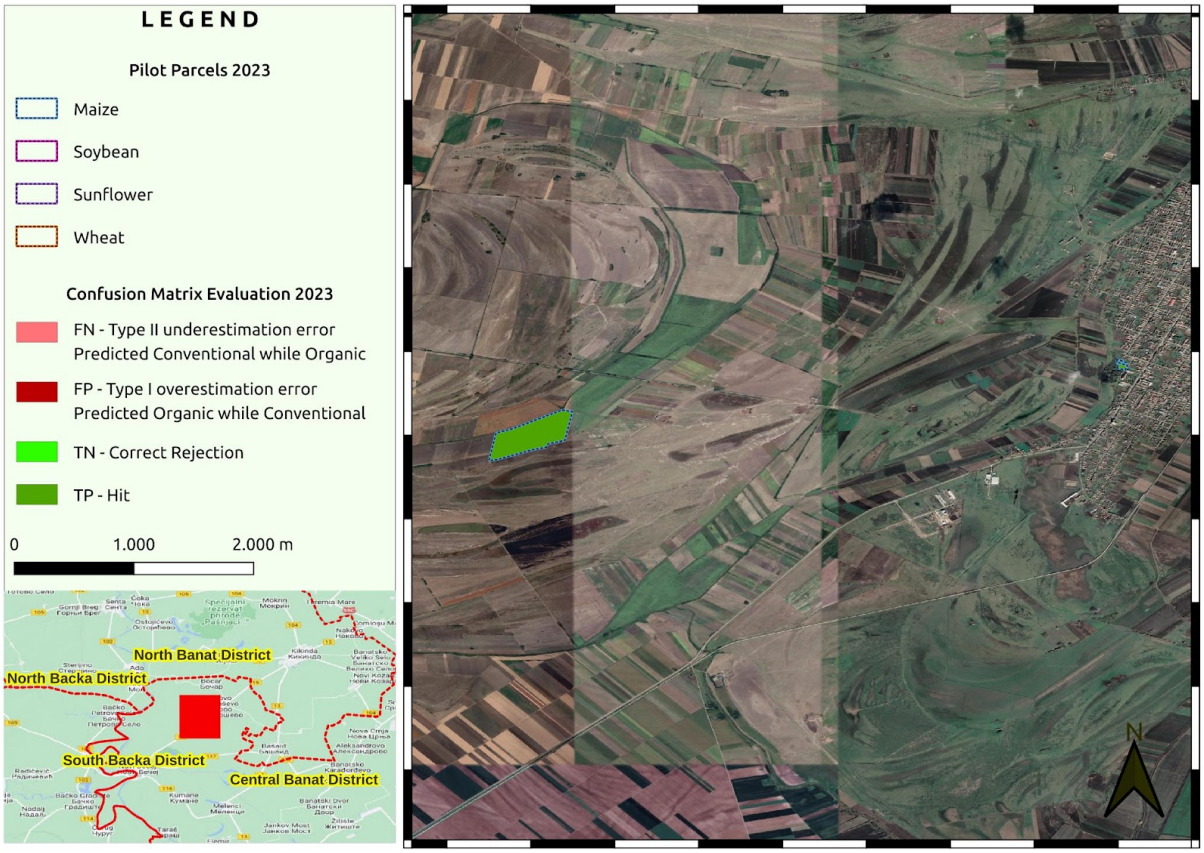
The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

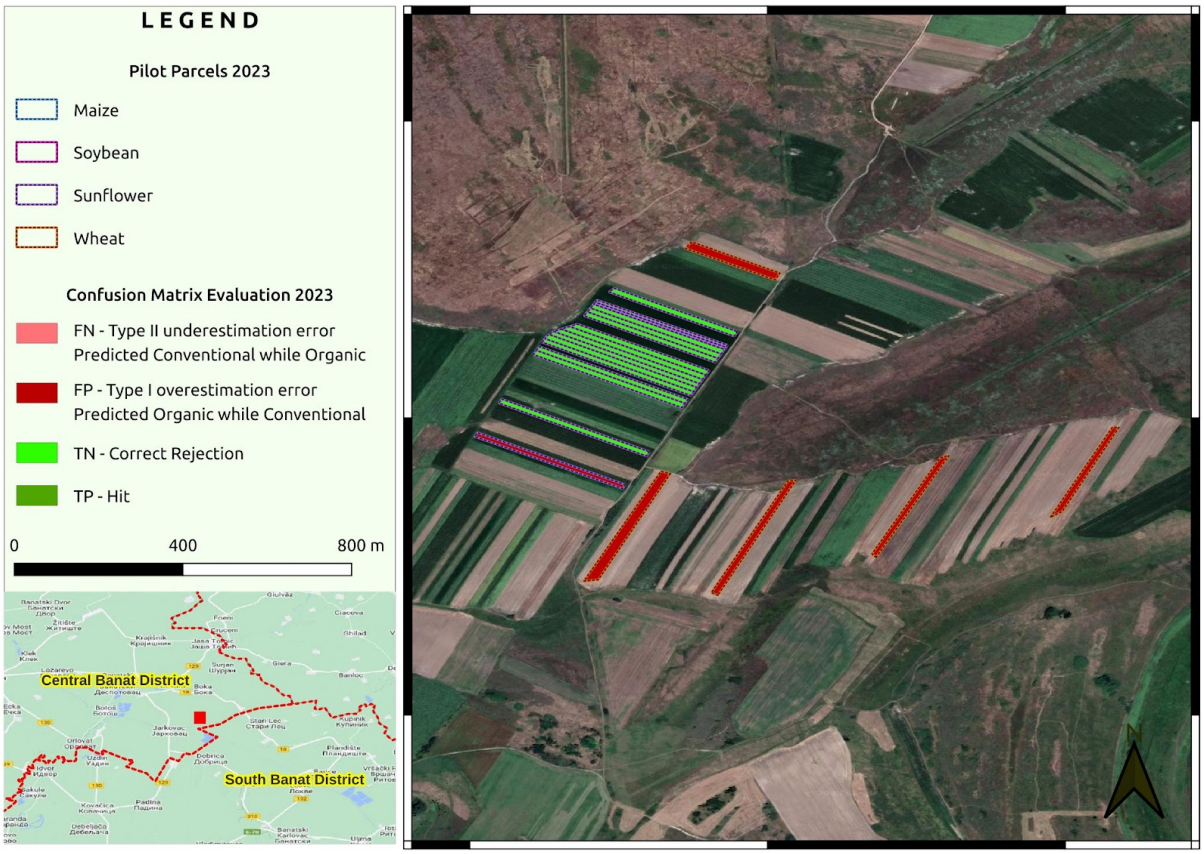
BC4: Monitoring of organic farming requirements – Serbia: Evaluation Maps of the traffic Light System for the 2023 parcel data

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

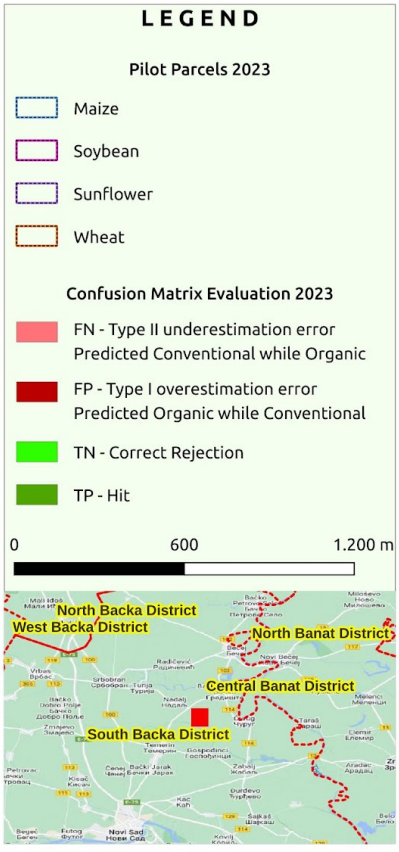


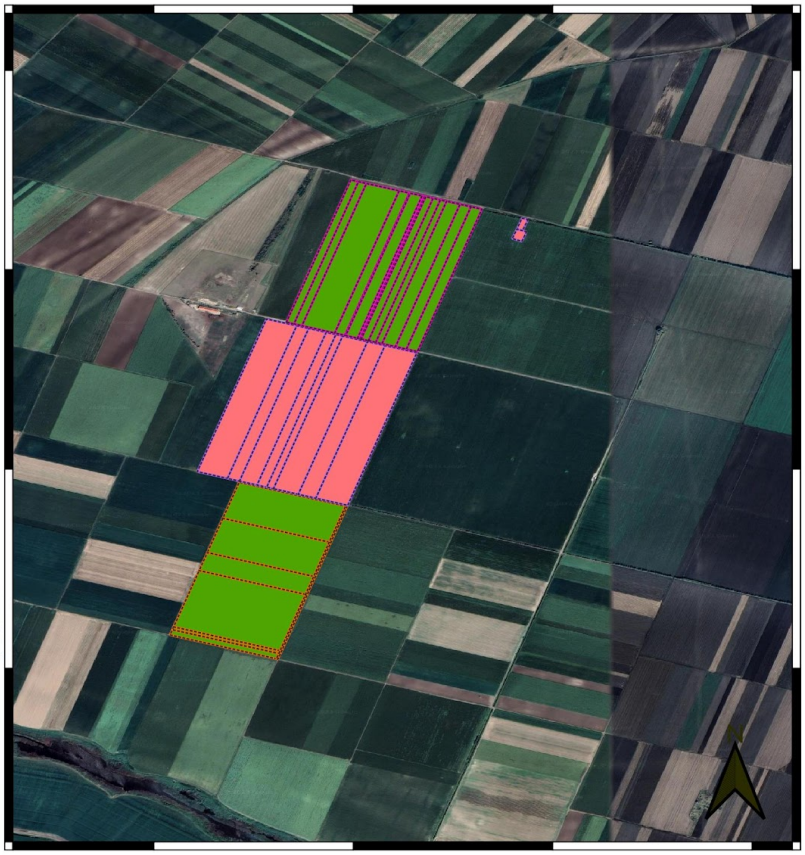
The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

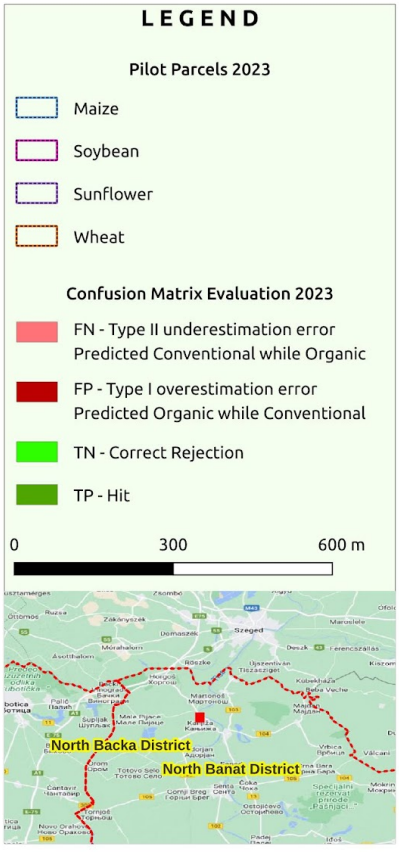




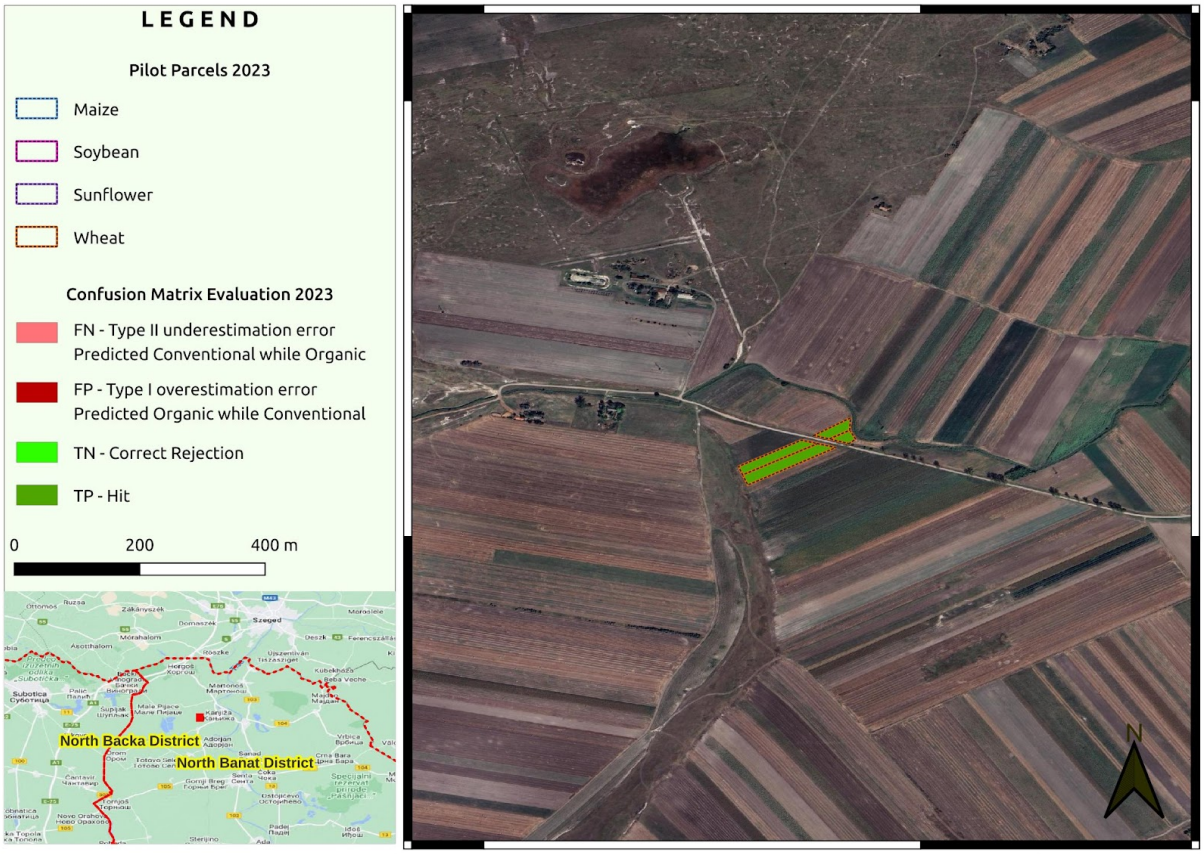
The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

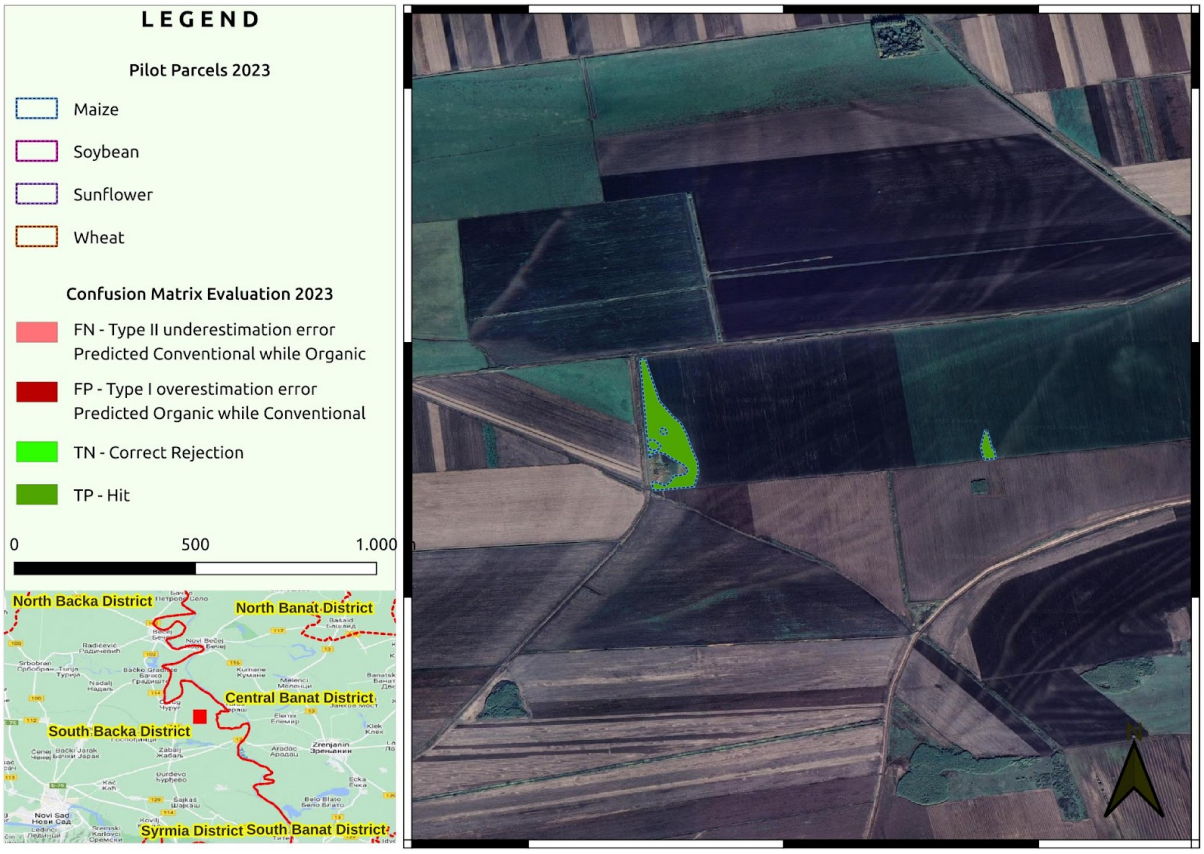




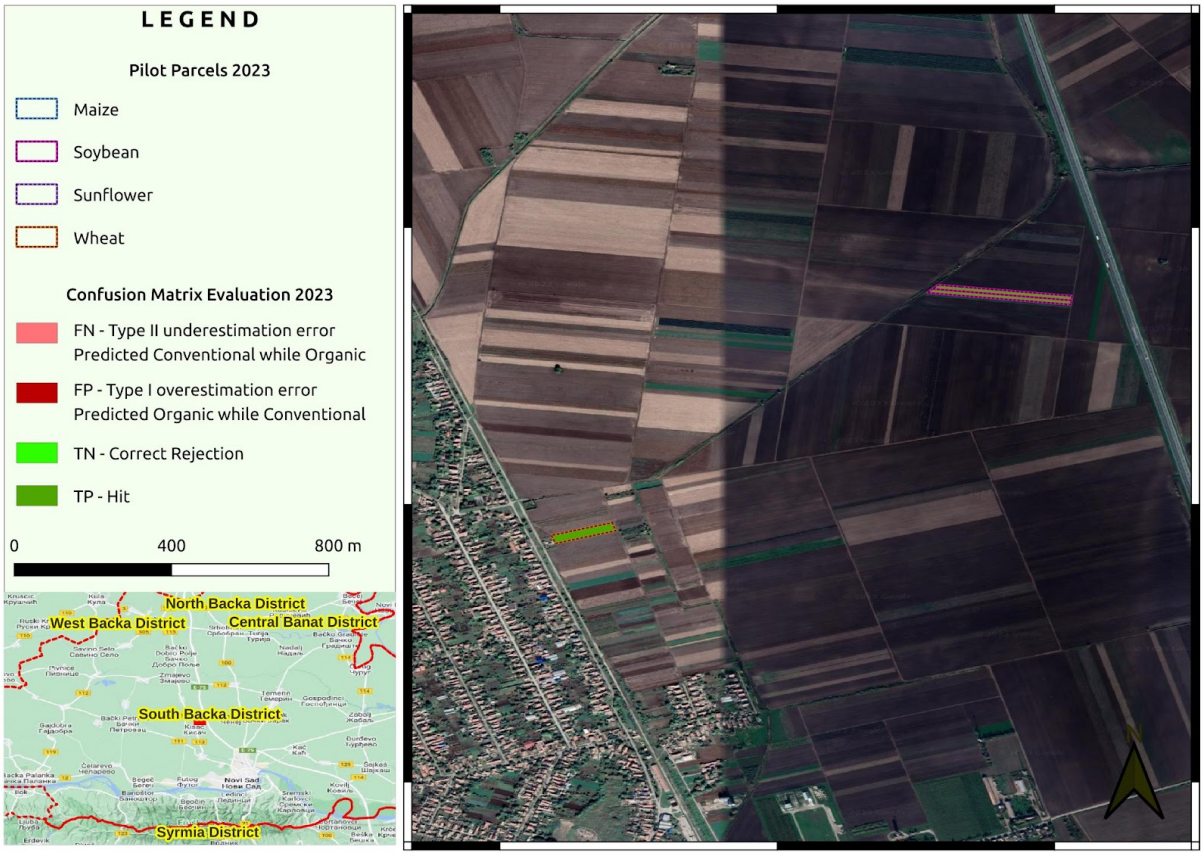


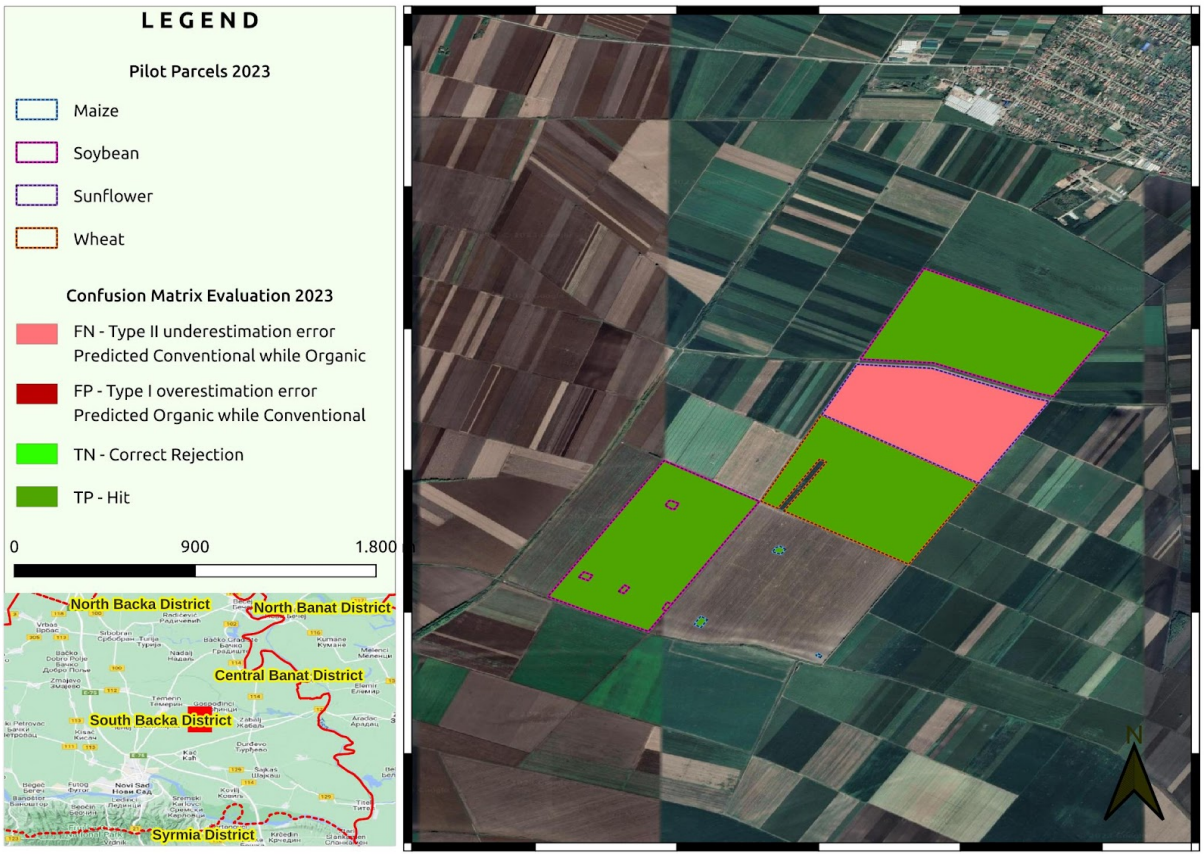
The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366





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





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

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

envision

End of Document

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