D3.4 DATA PRODUCTS VALIDATION REPORT

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	D3.4 Data products validation report		
Work Package	WP3 – Earth Observation Data Products		
Date of Delivery	**Contractual** M18	**Actual**	M18
Nature	Report	**Dissemination Level**	Public
Lead Beneficiary	DRAXIS		
Responsible Author	Stavros Tekes (DRAXIS), Petros Gasteratos (DRAXIS), Giorgos Letsos (DRAXIS)		
Contributions from	AgroApps, NOA, EV ILVO		

Document History

Version	Issue Date	Stage	Description	Contributor
D0.1	3/2/2022	Draft	Draft for review	DRAXIS
D0.2	18/2/2022	Draft	Review feedback	AgroApps
F1.0	28/2/2022	Final	Final version	DRAXIS

Disclaimer

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

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

CONTENT

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

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

1 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.

In this document there is the description of every product and service that was developed in ENVISION. More specifically, for every business case the pilot, the data collection, the validation results, the limitations and the next steps 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. The described results are the initial ones, based on historical data, and it is just the first attempt of developing the ENVISION products/ services. Further development, calibration and validation will be performed during the lifetime and will be presented in the forthcoming deliverables.

An update of this deliverable will be provided in M34 in the D3.6 Data product validation report (final version).

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

2 Validation Planning

The validation planning element will be used to describe the pilot and samples data to be acquired for the data products and services validation. The size, shape and location of the project's study areas will not be further described herein, since all the relevant information is provided in the respective deliverables (D3.2 Catalogue on auxiliary data and available repositories to be incorporated and D3.3 Data products initial 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, and their statistical correlation analysis with satellite data.

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 section will explain how these decisions were derived to meet the specifications of the planned interpretation (e.g. accuracy and precision) or analysis.

Furthermore, this element will be used to describe how the "right" data or imagery (type, quality and resolution) will be acquired for the operational function and validation procedure of the ENVISION project. The element will determine whether different resolutions of data are to be used in different parts of the project.

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

3 ENVISION Data Products description

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

3.1.1 Pilot and samples description

3.1.1.1 Analytics on Vegetation and Soil Index Time-series

Retrieve Analytics and Aggregated Statistics on custom Areas using GIS querying on the top of the DataCube

The methodology behind the analytics retrieval and the statistics was described in depth in D3.3. It's worth mentioning that the procedures under this service generate multiple results either by using DataCube directly or by making spatial queries to the database. The first one includes **plots** and **graphs**, whereas the second one is related to **a series of aggregated values** for each parcel or a set of parcels. This service can be particularly useful for the NPA, **assisting in the evaluation** of the outcomes of the different products and services provided for CAP monitoring purposes, namely Stubble Burning Identification, Grassland Mowing Event Detection, Natura 2000 Hotspot Detection, Minimum Soil Coverage for Soil Erosion, Smart Sampling and Crop Type Mapping, which have been described in detail in D3.3.

Minimum Soil Cover

The methodology for detecting a minimum soil cover over parcels to enhance protection of soil against erosion is described in detail in D3.3. The evaluation of the methodology is based on a validation dataset has been provided by the NPA (D3.2). This dataset includes 293 black fallow parcels, which extend all across the country (Figure 1). In more detail, these are cases that have been declared as "black fallow" but did not comply with the rules of GAEC4, which enforce a sowing to take place before November. Finally, on-the-spot-inspections for these fields were conducted from mid-November to mid-December.

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

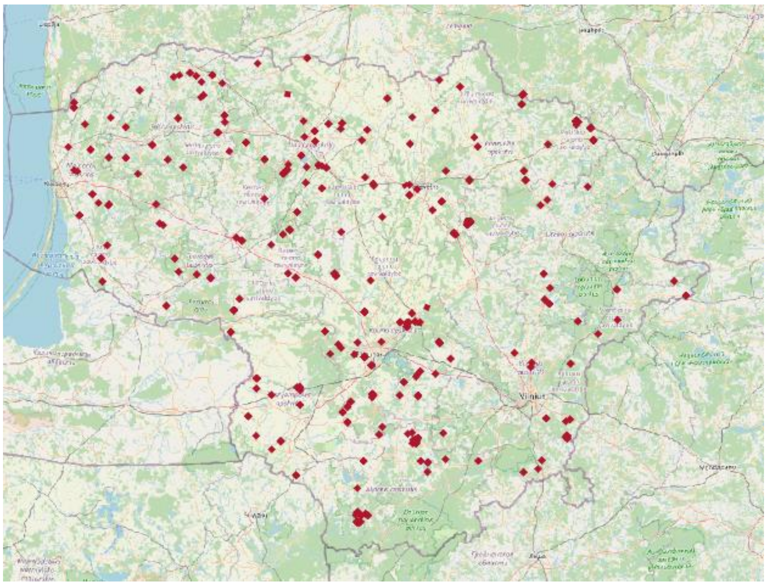


Figure 1: Validation samples of black fallow GAEC 4 incompliances

Runoff Risk Assessment for the Reduction of Water Pollution in Nitrate Vulnerable Areas (GAEC 1/ SMR 1)

As GAEC 1 rules require not to spread manure and/or slurry in the coastal protection zones of water bodies marked in the Surface Water Protection Zone layer, we have developed a run-off risk assessment taking into account the proximity of each parcel to the closest water surfaces. The latter layer is retrieved from the hydrographic network of Lithuania provided by NPA (Figure 2). The proposed methodology for the risk estimation is described in detail in D3.3. No validation data for this task are available.

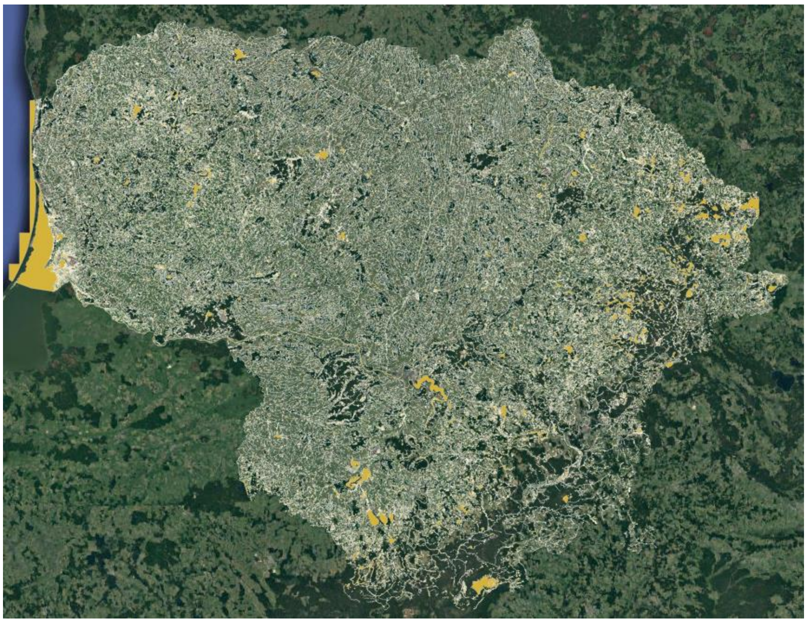


Figure 2: Lithuania water-bodies and hydrographic network

Stubble Burning Identification

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

Stubble Burning Identification product is described in D3.3. This product is developed to answer the GAEC 6 requirement that is related to the burning of agricultural plants and their stubble. Due to lack of validation data, since NPA did not provide sufficient validation data (except 3 cases), the validation of the results in this area is done optically using photo-interpretation. The analytics procedures, however, can assist in an efficiently and timeless evaluation process.

<u>Harvest event detection</u>

The methodology regarding the Harvest event detection algorithm is described in D3.3. The lack of validation data is also an issue in this case. Thus, we collected a sample of 407 random samples from arable cultivations (Winter Cereals, Spring Cereals, Vegetables, Potatoes, etc.) in order to evaluate the precision of identification of harvest events visually, through photo-interpretation. Specifically, we identified abrupt changes in the NDVI and RBG time series, based on the crop type and the expected harvest period of the cultivation year.

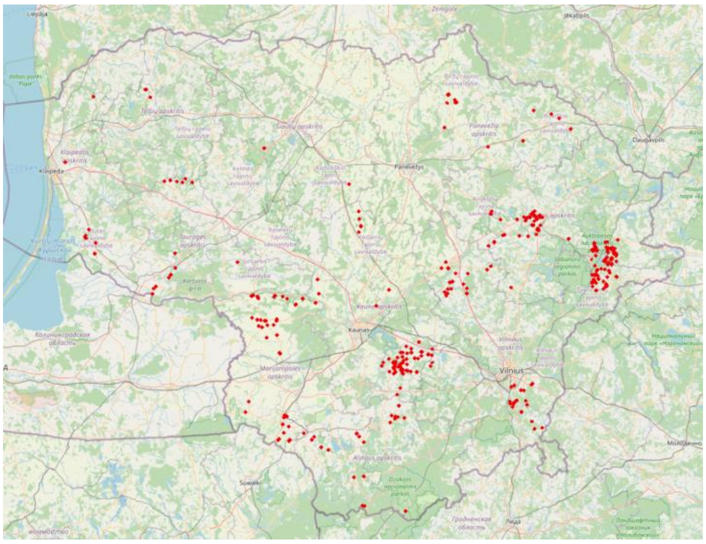


Figure 3: Random Sample picked for the identification of Harvest Events via photo-interpretation

3.1.1.2 Cultivated Crop Type Maps (CCTM)

Methodology of Cultivation Crop Type Map product for Lithuania is described in detail in D3.3. For the validation of the product, Sentinel-1 and Sentinel-2 images from January of 2020 until the September of the same year were used. Specifically, statistical measures of Mean, Median and Standard Deviation calculated from the pixels encompassed in the LPIS buffered geometries for all the s1/s2 bands and indices, referred in D3.3. Moreover, a sample training dataset of about 100,000 instances for 16 crop classes and a respective validation one of more than 20,000 cases were extracted from 23 different areas across Lithuania territory (see Figure 4) and used for the model training and the subsequent evaluation product outputs respectively, as provided from the NPA (mentioned also in D3.2).

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

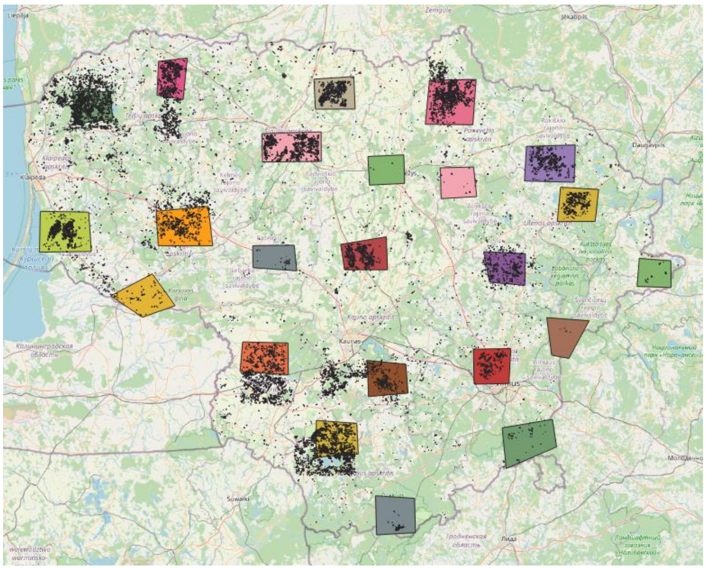


Figure 4: Sampled Areas of Lithuanian Validation Crops

3.1.1.3 Grasslands Mowing Event Detection

The methodology for the Grasslands Mowing Events Detection product is composed of a pipeline of consecutive routines and is described in detail in D3.3.

For the training and evaluation of the Data Fusion product, we collected pixels from several grassland fields, distributed across the entire Lithuanian territory, in order to incorporate different NDVI behaviours. As a result, the dataset includes various profiles of sparsity and local characteristic of the different areas. Apart from NDVI time series, we also utilize all SAR and InSAR time-series of the respective pixels.

Additionally, for the requirements of Mowing Detection model, NPA provided only a few validation cases related to the compliance of the respective grassland fields (if a mowing event occurred inside the mandatory timeframe). For that reason, we adapted two different approaches in order to generate the necessary event timestamps, sufficient to train our models and validate the service's results.

1. *Events Instances based on Photo-Interpretation*

Firstly, a blind photo-interpretation process was carried out by three experts to generate annotated event instances, based on optical Sentinel-2 imagery. Specifically, two of the experts worked independently on two different areas, visualized in Figure 6. For every detection, an approximate timeframe was identified, regarding the starting and finishing date of the event, based on the respective Sentinel image acquisitions, as well as the interpreter's confidence and the estimated percentage of the mowed area (see Figure 5). Consecutively, the third expert evaluated the results of the other two (with 95% agreement) and decided on the final dataset. Finally, mowing compliance was easy to be inferred using Table 5: Lithuania National Mowing Regulations of D3.3. **Overall, we produced a dataset of 2,344 grassland fields instances, including information regarding both the compliance and mowing event timeframe.**

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

	Ori_hold	Ori_id	ri_cro	unique_id	Area	Perimeter	grazing	SEN4CAP_c	NPA_c	mow_n	m1_dstart	m1_dend	m1_conf	1_per	Compliancy	Remarks
1	1002942841	1147703	DGP	1002942841-145...	9.789...	1388.122...	0	1	1	0	-	-	0	0	2	-
2	1002944361	290734	DGP	1002944361-151...	3.100...	769.7450...	0	1	0	1	22-5-2020	26-6-2...	100	90	1	-
3	1002950337	23158	DGP	1002950337-150...	0.230...	259.5199...	0	2	1	1	9-9-2020	19-9-2...	20	50	2	{m1:small}
4	1002952182	175840	DGP	1002952182-151...	0.990...	622.4442...	0	2	0	0	-	-	0	0	2	-
5	1002952671	237706	DGP	1002952671-147...	7.640...	1341.128...	0	1	0	1	15-8-2020	20-8-2...	40	10	2	{m1:partial}
6	1002952671	237698	DGP	1002952671-148...	0.920...	391.2890...	0	2	1	0	-	-	0	0	2	-
7	1002952671	237703	DGP	1002952671-147...	4.820...	1445.764...	0	2	1	2	10-8-2020	15-8-2...	10	50	2	{m1:partial}
8	1002955549	500294	DGP	1002955549-148...	0.530...	379.7484...	0	1	0	1	22-5-2020	26-6-2...	90	100	1	-
9	1002956057	870818	DGP	1002956057-151...	1.720...	822.6799...	0	1	1	1	26-6-2020	26-7-2...	80	50	1	{m1:partial}
10	1002956057	870808	DGP	1002956057-151...	1.220...	552.4066...	0	1	1	1	26-6-2020	26-7-2...	90	70	1	{m1:shadows}

Figure 5: Preview of Photo-Interpretation results shapefile

Despite the small number of grassland fields included in this approxiamation, we should keep in mind that the training of the DL model was performed on the pixel level, which resulted in more than 100,000 cases. Specifically, each pixel is assigned with a positive label in a particular time instance if i) there is an event at this timestamp, for the parcel in which the pixel belongs and ii) there is a noticeable drop of the pixel's NDVI value (more than 0.2). That way, we ensure that we will consider events on a part of a parcel that has not been entirely mowed.

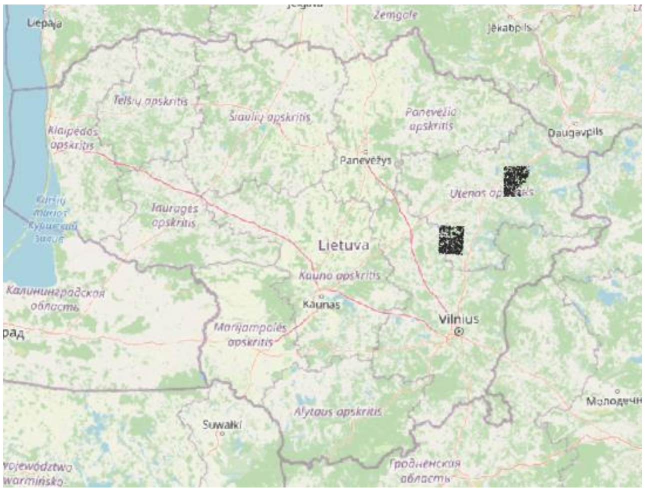


Figure 6: Areas that Mowing Photo-interpretation took place

2. NPA validation samples

In order to exploit efficiently the conformity validation cases provided by the NPA (described in D3.2), a matching of the respective cases with the results provided by the SEN4CAP project took place. SEN4CAP mowing detection service provides information regarding the estimated timestamp that an event took place as well as the level of confidence of the prediction. For that reason, we collect all cases that match with the corresponding NPA validations on the level of conformity (Figure 7) and we keep only the most confident ones. Labels were assigned on the pixels of each parcel, in same fashion as described in the previous paragraph. The final dataset included almost 6,000 grassland fields collected from five different areas (Figure 8).

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

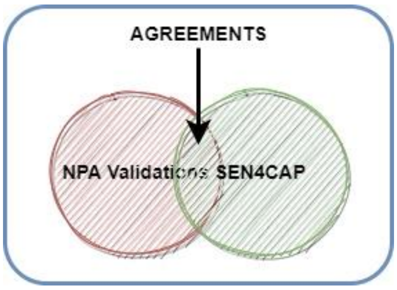


Figure 7: Intersection of NPA validations with SEN4CAP mowing results

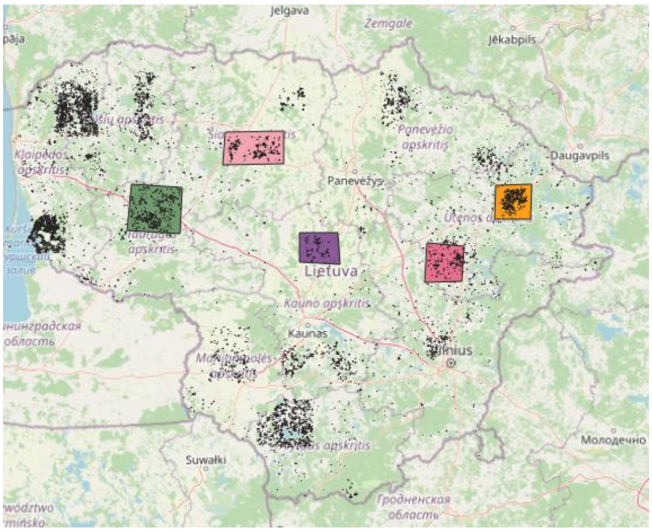


Figure 8: Sampled areas of NPA mowing conformity used

3.1.2 Validation Results

3.1.2.1 Analytics on Vegetation and Soil Index Time-series

Retrieve Analytics and Aggregated Statistics on custom Areas using GIS querying on the top of the DataCube

The methodology behind the analytics retrieval and the statistics was described in depth in D3.3. It's worth mentioning that the procedures under this service generate **multiple and various results** either by using **DataCube** directly or by making **spatial queries to the database**. The first one includes plots and graphs, whereas the second one is related to a series of aggregated values for each parcel or a set of parcels. It is worth mentioning that visualizations can be extracted for **any s1/s2 product, any aggregated statistic measurement, either on the pixel or on the parcel level and for a specific parcel, a crop type or a crop family**. Below are presenting some examples, which highlight the potential of big earth observation data and analytics:

1. **Animation of the temporal evolution of an area.** In Figure 9 and Figure 10, the NDVI index was calculated for the entire year of 2019 for a specific parcel. The calculated values can be used as an input for an animated plot which reveals the evolution of the vegetation throughout the selected time range.
2. **Temporal Statistics over an area.** The temporal statistics exploit the time dimension as in the previous sub-task. The difference here is that instead of creating an animation, we plot the data. The usage of these tools enhances the CAP monitoring by providing useful information about the evolution of the vegetation throughout a user-defined time range. The examples

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

below depict two parcels declared as spring cereals, but predicted as Winter cereal via the crop classification process of ENVISION. It's worth mentioning that this tool plays **a crucial role to validation of results generated by the rest of back- end processes** such as the grassland mowing and the crop classification. Specifically, the most of these services raise alerts. Decision-makers have the potential to examine effortless these alerts by exploring temporal statistics in any dimensions (time, location and band/index), without the need for relying on non-exhaustive, time-consuming and complex methods.

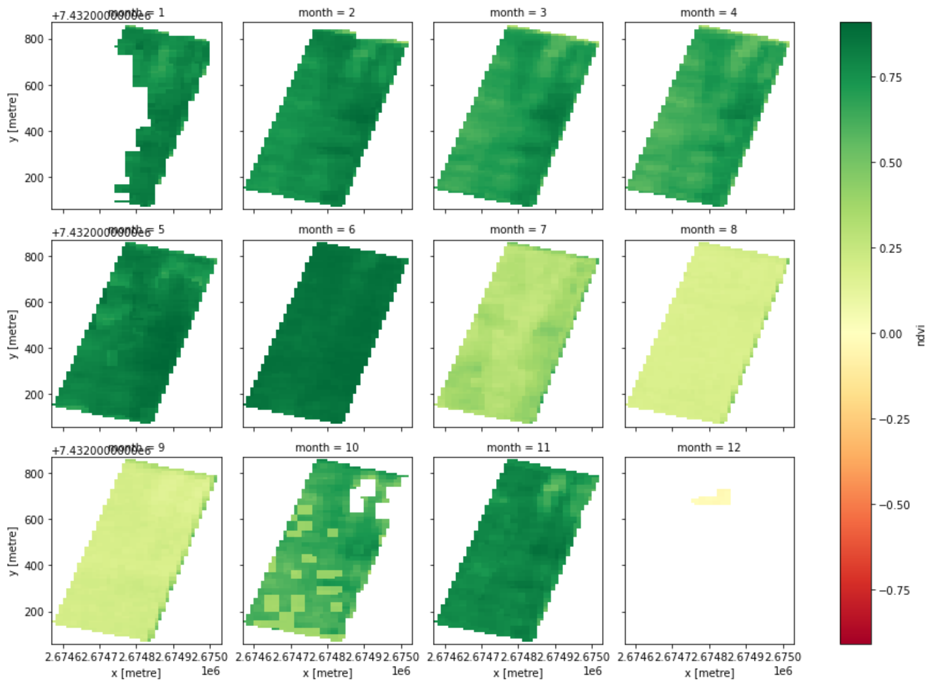


Figure 9: Median monthly NDVI for parcel for which false-declaration-alert has been raised

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

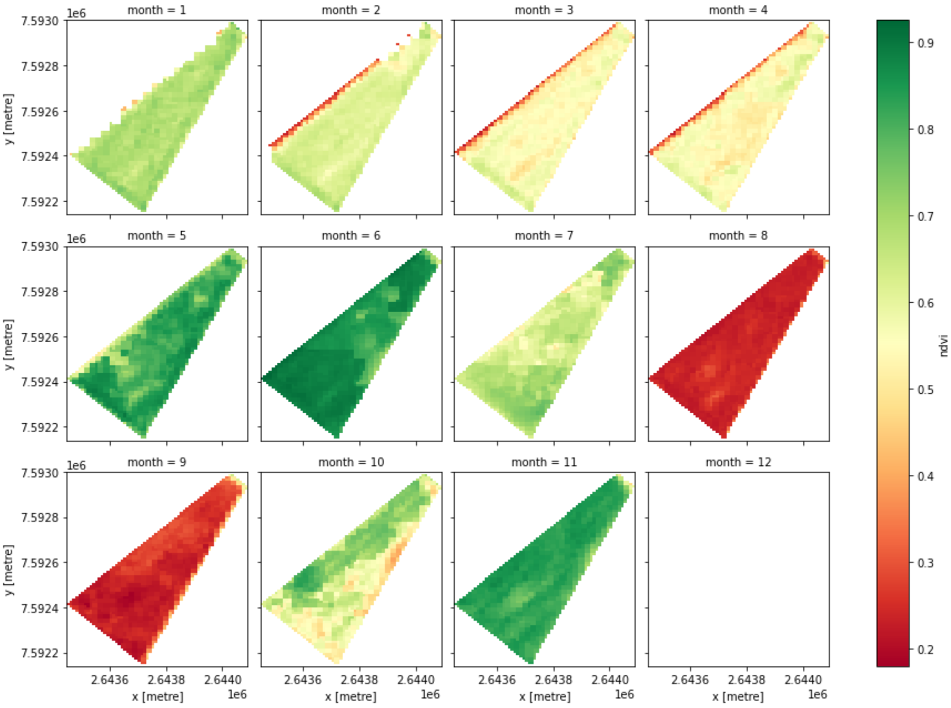


Figure 10: Mean monthly NDVI for parcel for which false-declaration-alert has been raised

3. **Smoothed time-series**. While the previous examples highlight the values of each pixel, these plots take into account the mean value of cloud-free pixels inside an area. As there can be gaps between two or more calculations due to cloud presence, a smoothing process takes place. Thus, patterns can be more noticeable and reveal trends throughout the years.

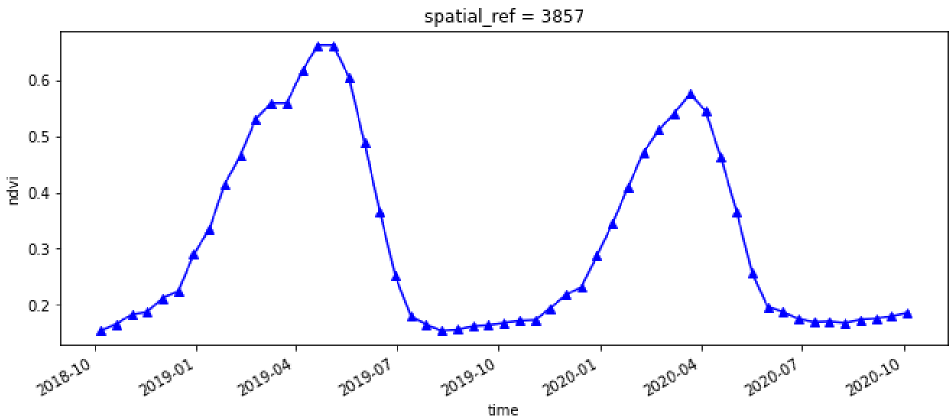


Figure 11: NDVI smoothed mean values for a parcel for two years

4. **HeatMap of Clear Pixels.** The production of a heatmap regarding clear pixels contributes to a better and more effective data analysis. For instance, Figure 12reveals a frequency analysis of the cloud-free pixels included in a parcel over a selected time window. This is very useful in cases of pixel based solutions for other products, since **the optimal selection of the pixels** (the ones with **the highest number of cloud-free instances**) can be easily acquired.

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

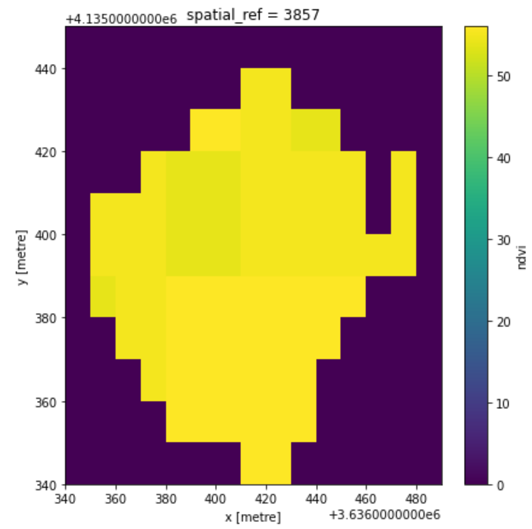


Figure 12: Heatmap of clear pixels included in a parcel

5. **Index Anomalies**. One of the primary goals of this task is to identify possible changes over the time. Thus, this sub-task focus on understanding possible anomalies to any chosen index. For example, NDVI measures the greenness of plant leaves, which indicates an overall vegetative health. As we have dense measurements for NDVI, there is the potential of comparing current NDVI value, either for a day or for a month, to the average computed NDVI over one or more years. From this comparison, often called NDVI Anomaly, we can determine the changes in an area regarding the drought or crop production. Figure 13 illustrates the NDVI standardized anomaly using short and long-time ranges values where short is the mean NDVI and long is the yearly NDVI.

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

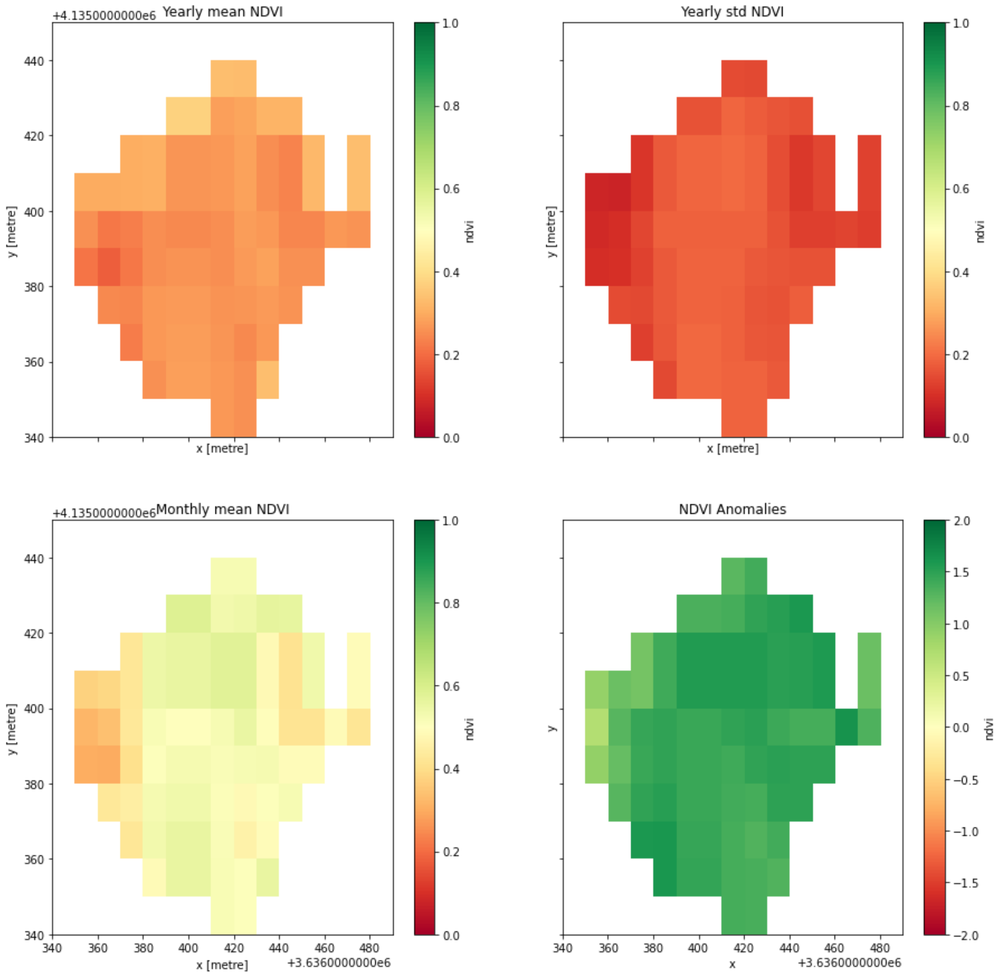


Figure 13: Yearly mean NDVI, yearly standard deviation of NDVI, Monthly mean NDVI and NDVI anomaly for a specific parcel.

6. **Spatial Queries.** As the aggregated results have been calculated, they are stored in the database. Thus, a series of simple and complex queries can take place. These queries can help users to monitor parcels over an area throughout the time. This geospatial database will be exploited by the ENVISION platform, where the latter will be connected in order to retrieve all the required data and visualize them via plots, layers in maps and raw text. At the same time, there is the potential for users to **connect directly to the database** so the data become available for process, analysis and integration to several GIS systems and applications. It becomes apparent that the storage of this information is of a high importance as it **can be exploited by anyone**. A number of queries are listed below in order to showcase the discrete and various information provided by ENVISION services and of which the output has initially visualized in CreoDIAS platform:
    a. **List the mowing events for a specific parcel**

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

```sql
SELECT event_start_date,event_end_date,event_confidence
from parcels,parcels_mowing
where country_id = 1 and parcel_id = '1011586814-017455-7612-2' and parcels.db_id = parcels_mowing.db_id_id
```

Data Output | Explain | Messages | Notifications

	event_start_date date	event_end_date date	event_confidence double precision
1	2019-06-01	2019-06-11	91
2	2019-06-19	2019-06-28	87

Figure 14: Query for collecting back the results from grassland mowing events detection

b. List the average area per crop type

```sql
SELECT crop_code,AVG(ST_AREA(geom))
FROM parcels, crops
where declared_crop_id_id = crops.crop_id
group by crop_code
order by AVG(ST_AREA(geom)),cast(crop_code as integer)
limit 10
```

Data Output | Explain | Messages | Notifications

	crop_code character varying (100)	avg double precision
1	193	403.8530588568311
2	205	919.6634593821516
3	161	973.1831965693099
4	236	1774.286990206512
5	56	1821.4268250483096
6	222	1868.8569339764863
7	233	1969.2431246279064
8	181	2212.3051899156585
9	23	2309.006612438152
10	46	2493.095013344273

Figure 15: Query for calculating the distance between two parcels

c. List the distance between two or more parcels

```sql
SELECT ST_DISTANCE(p1.geom,p2.geom)
FROM parcels p1, parcels p2
where p1.parcel_id = '1011585518-052418-0810-1' and p2.parcel_id = '1011585518-052418-7866-2'
```

Data Output | Explain | Messages | Notifications

	st_distance double precision
1	1838.0739789723586

Figure 16: Query for calculating the distance between two parcels

d. List the false declarations based on predictions over an area or/and per farmer

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

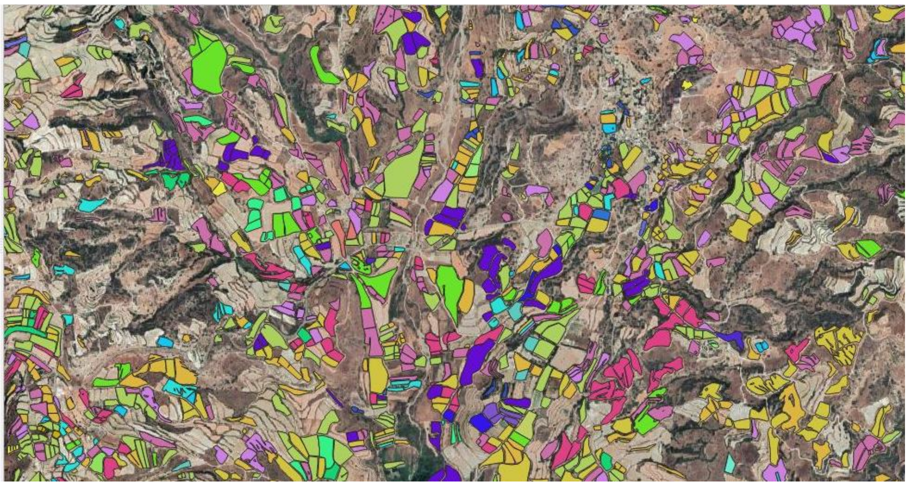


Figure 17: Visualization of the query's result regarding false declarations based on predictions over an area

Minimum soil Cover

The minimum soil cover detection algorithm (described in D3.3) has been applied to a validation set of **281** parcels. The accuracy of the algorithm is 61% as it detects soil presence in 171 of the 281 parcels of which we have at least one cloud-free image to analyse from end of September to December, whereas the most of the on-the-spot checks for the sample data has taken place at the end of November. It is worth mentioning that for a high number of misclassified parcels there was only one cloud-free image or/and a small number of cloud-free pixels, which implies the increase of wrong predictions' percentage.

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. Taking into consideration the values of water proximity and RUSLE, runoff risk has been computed for each parcel. The risk level for the most parcels is high due to the fact that they are close to water surfaces. More results are presented in Table 1. At the same time Figure 18 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 1: Run-off Risk Assessment Results

	Very Low or Low Risk	Moderate Risk	Very High or High Risk
Algorithm	79072	110	251067

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

Figure 18: Visualization of the run-off risk for a subset of parcels along with the water surfaces around them

Stubble Burning Identification

The Stubble Burning Identification algorithm is a threshold-based approach (as described in D3.3) for finding burnt areas. The lack of validation data and the existence of very few burnt parcels in the country, do not allow for an evaluation of the performance, in this case. However, there are several validation data in for the area of Cyprus, where the same algorithm is applied, and the results are presented in Section 3.2.

Harvest Event Detection

The Harvest event detection algorithm, as already mentioned in D3.3, is threshold-based approach looking for abrupt drops in NDVI values. The algorithm performs quite well, predicting with very close proximity the day of the harvest event. Specifically, the evaluation process includes a sample of 407 randomly selected arable cultivations that we have checked using photo-interpretation of the RGB and NDVI components. Given that, for the majority of the cases (334 out of 407) we managed to identify the harvest events on time, within a range of less than 6 days, and even better, 163 from those within a range less than 3 days (Figure 19). Finally, we can see that the maximum error in term of DoYs is 24 days.

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

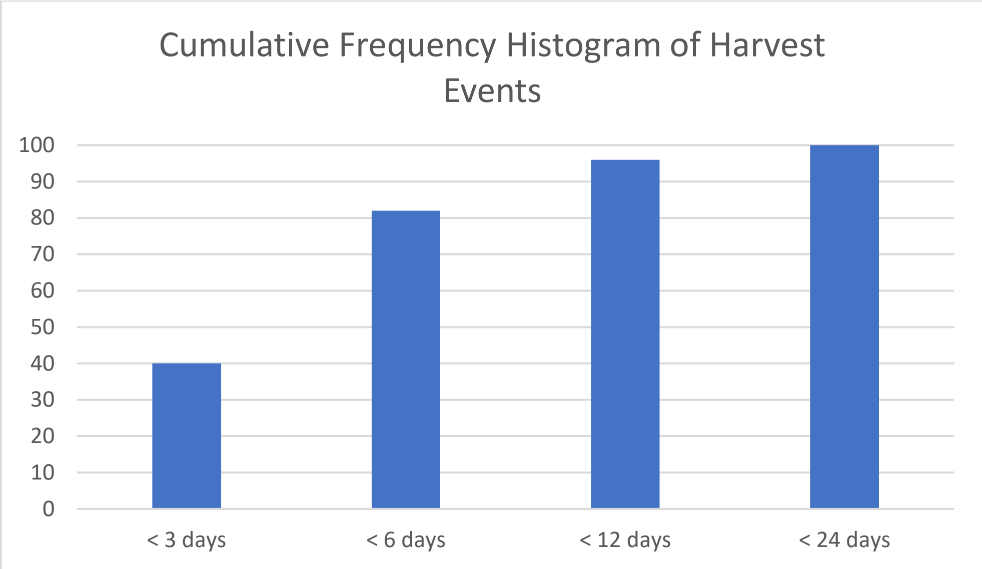


Figure 19: Cumulative frequency (%) of distance in DoYs between the models prediction and validation

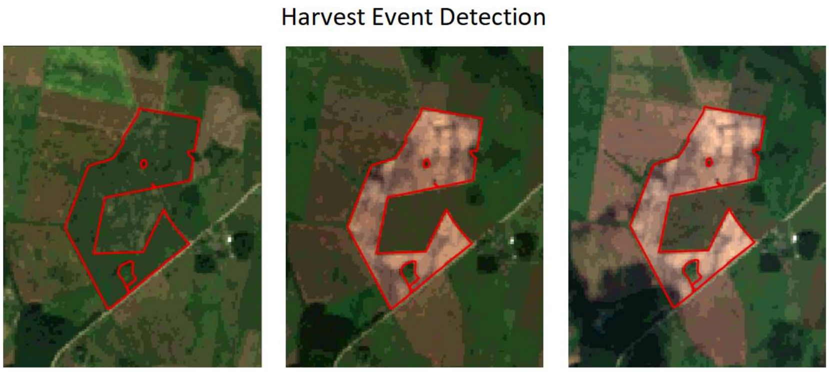


Figure 20: Parcel before and after harvesting, as detected from the Harvest Event Detection algorithm

3.1.2.2 Cultivated Crop Type Maps (CCTM)

As explained in detail in D3.3, multiple crop type maps are produced throughout the cultivation period, starting from early April until the end of August. The accuracy of the different models increases gradually, as more acquisitions are included in the input, resulting to an optimal performance at the end of August. Figure 21 presents the F1 score for **16 different crop classes**, in which the validation analysis during the cultivation period was performed.

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

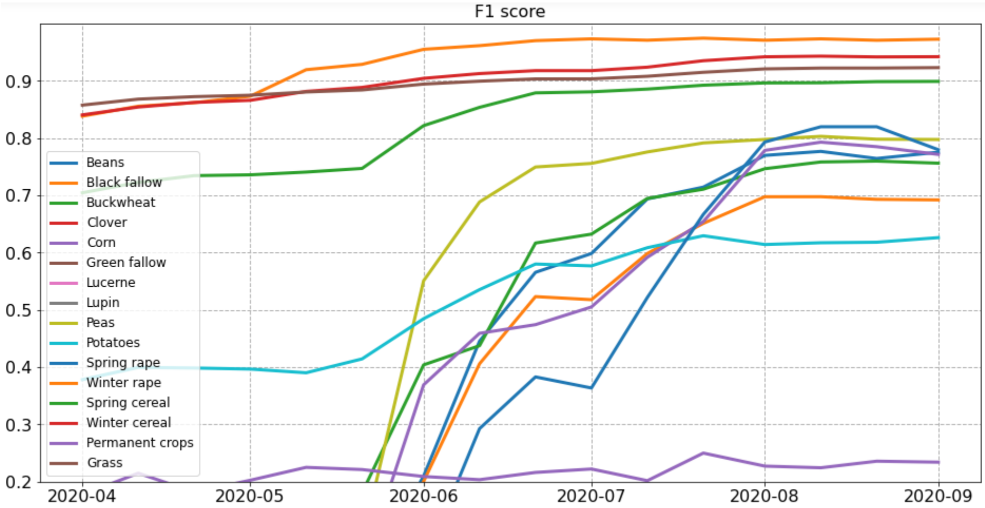


Figure 21: Classifier Performance (F1 score) Progress over Cultivation Period

Table 2 and Figure 22 illustrate the final classification report of the model, performed at the beginning of September. Accuracy of the classifier in relatively high for almost all the crop classes, with the exception of specific cases that have very low support (clover, green fallow, Lucerne, lupin). We should also mention that the classes of mixed crops, like agricultural mixes, other vegetables, protein plant and other crops on arable land, were excluded from the current study analysis, since they do not have a clear spectral profile and the performance of the classification model is very poor on those cases.

Table 2: Classification Report based on the predictions provided at the early September

	UA	PA	F1-Score	Support
Beans	0.78	0.77	0.78	155
Black Fallow	0.77	0.63	0.69	712
Buckwheat	0.73	0.79	0.76	237
Clover	1.0	0.05	0.10	279
Corn	0.73	0.81	0.77	167
Green Fallow	0.00	0.00	0.00	81
Lucerne	0.00	0.00	0.00	126
Lupin	1.00	0.04	0.07	28
Peas	0.89	0.72	0.80	437
Potatoes	0.57	0.69	0.63	544
Spring Rape	0.92	0.68	0.78	34
Winter Rape	0.97	0.98	0.97	818
Spring Cereal	0.86	0.92	0.90	3354
Winter Cereal	0.94	0.95	0.94	4946
Permanent Crops	0.37	0.17	0.23	140
Grass	0.90	0.95	0.92	8052
Macro Avg.	0.72	0.57	0.58	20110
Weighted Avg.	0.88	0.89	0.88	20110
Overall Accuracy			0.89	
Kappa			0.85	

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

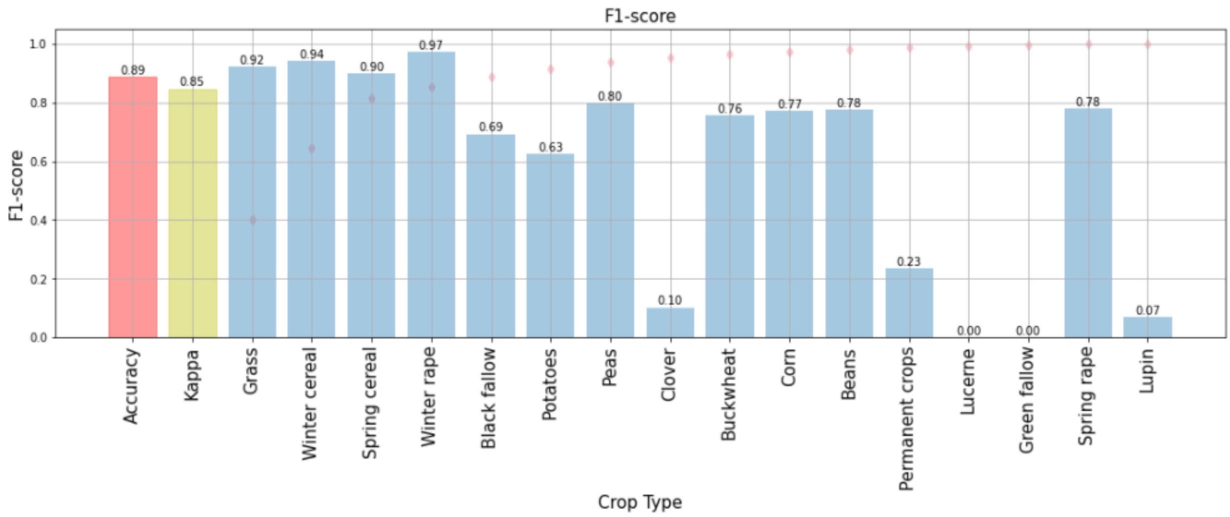


Figure 22: F1 score based on the predictions provided at the early September

Tables 3 and 4 below depict the producer and user accuracy of the different cases, as well as the loss of information among classes and how the model confuses them. Specifically, the producer accuracy table (Table 3) 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 78% 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, 1st confusion class for all the annual crops (e.g potatoes, spring rape, peas, etc.) is the spring cereals since they belong in the same higher taxonomy class.

On the other hand, the user accuracy table (Table 4) 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 beans 78% were indeed beans, while 11% were in fact peas. Interestingly, in the occasion of clover we see that, even though the algorithm can cannot distinguish clovers from grasslands as stated before, everything that has been predicted as clover, is indeed a clover. On the other hand, in the case of green fallow and lucerne the model does not classify any instance as such.

From the results it is obvious that the model performs much 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 2020, against actual validation, when NPA needs to have the first results in order to start planning their OTSC campaigns. Eventually, in case we could accumulate more images of September into the feature space, and of course, more training samples into the training dataset, the model gives even better results.

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

Table 3: Lithuania Producer Accuracy Table

	Crop Code	Crop Name	Declared parcels	Well Classified	Producer Accuracy	Confusion class 1	% 1	Confusion class 2	% 2	Confusion class 3	% 3	Rest %
0	15.000000	Beans	155.000000	119	0.768000	Spring cereal	0.070000	Potatoes	0.050000	Peas	0.050000	0.060000
1	18.000000	Black fallow	712.000000	446	0.626000	Grass	0.120000	Winter cereal	0.090000	Spring cereal	0.090000	0.080000
2	26.000000	Buckwheat	237.000000	186	0.785000	Potatoes	0.110000	Spring cereal	0.050000	Grass	0.040000	0.020000
3	33.000000	Clover	279.000000	15	0.054000	Grass	0.780000	Winter cereal	0.040000	Potatoes	0.040000	0.080000
4	37.000000	Corn	167.000000	135	0.808000	Potatoes	0.100000	Grass	0.030000	Spring cereal	0.020000	0.040000
5	55.000000	Green fallow	81.000000	0	0.000000	Grass	0.310000	Black fallow	0.250000	Spring cereal	0.120000	0.320000
6	75.000000	Lucerne	126.000000	0	0.000000	Grass	0.830000	Potatoes	0.060000	Winter cereal	0.040000	0.080000
7	76.000000	Lupin	28.000000	1	0.036000	Spring cereal	0.250000	Beans	0.210000	Potatoes	0.180000	0.320000
8	104.000000	Peas	437.000000	315	0.721000	Spring cereal	0.120000	Grass	0.040000	Beans	0.040000	0.080000
9	109.000000	Potatoes	544.000000	376	0.691000	Spring cereal	0.120000	Grass	0.080000	Winter cereal	0.040000	0.080000
10	134.000000	Spring rape	34.000000	23	0.676000	Spring cereal	0.120000	Potatoes	0.060000	Buckwheat	0.060000	0.090000
11	157.000000	Winter rape	818.000000	798	0.976000	Winter cereal	0.020000	Grass	0.000000	Spring cereal	0.000000	0.000000
12	180.000000	Spring cereal	3354.000000	3099	0.924000	Grass	0.020000	Winter cereal	0.020000	Potatoes	0.020000	0.010000
13	181.000000	Winter cereal	4946.000000	4694	0.949000	Grass	0.030000	Spring cereal	0.020000	Black fallow	0.000000	0.000000
14	2000.000000	Permanent crops	140.000000	24	0.171000	Grass	0.790000	Black fallow	0.010000	Spring cereal	0.010000	0.010000
15	3000.000000	Grass	8052.000000	7612	0.945000	Winter cereal	0.020000	Spring cereal	0.010000	Potatoes	0.010000	0.010000

Table 4: Lithuania User Accuracy Table

	Crop Code	Crop Name	Classified parcels	Well Classified	User Accuracy	Confusion class 1	% 1	Confusion class 2	% 2	Confusion class 3	% 3	Rest %
0	15.000000	Beans	152.000000	119	0.783000	Peas	0.110000	Lupin	0.040000	Potatoes	0.030000	0.040000
1	18.000000	Black fallow	577.000000	446	0.773000	Grass	0.080000	Green fallow	0.030000	Potatoes	0.030000	0.080000
2	26.000000	Buckwheat	255.000000	186	0.729000	Grass	0.070000	Peas	0.040000	Spring cereal	0.040000	0.130000
3	33.000000	Clover	15.000000	15	1.000000	Grass	0.000000	Permanent crops	0.000000	Winter cereal	0.000000	0.000000
4	37.000000	Corn	183.000000	135	0.738000	Spring cereal	0.050000	Grass	0.050000	Potatoes	0.030000	0.130000
5	55.000000	Green fallow	0.000000	0	nan	nan	nan	nan	nan	nan	nan	nan
6	75.000000	Lucerne	0.000000	0	nan	nan	nan	nan	nan	nan	nan	nan
7	76.000000	Lupin	1.000000	1	1.000000	Grass	0.000000	Permanent crops	0.000000	Winter cereal	0.000000	0.000000
8	104.000000	Peas	353.000000	315	0.892000	Black fallow	0.020000	Beans	0.020000	Potatoes	0.020000	0.040000
9	109.000000	Potatoes	657.000000	376	0.572000	Grass	0.130000	Spring cereal	0.080000	Black fallow	0.050000	0.160000
10	134.000000	Spring rape	25.000000	23	0.920000	Green fallow	0.080000	Grass	0.000000	Permanent crops	0.000000	0.000000
11	157.000000	Winter rape	823.000000	798	0.970000	Winter cereal	0.010000	Grass	0.010000	Black fallow	0.000000	0.010000
12	180.000000	Spring cereal	3541.000000	3099	0.875000	Grass	0.030000	Winter cereal	0.020000	Potatoes	0.020000	0.050000
13	181.000000	Winter cereal	5020.000000	4694	0.935000	Grass	0.030000	Spring cereal	0.020000	Black fallow	0.010000	0.010000
14	2000.000000	Permanent crops	65.000000	24	0.369000	Grass	0.550000	Spring cereal	0.030000	Winter cereal	0.020000	0.030000
15	3000.000000	Grass	8443.000000	7612	0.902000	Clover	0.030000	Winter cereal	0.020000	Permanent crops	0.010000	0.040000

As explained in detail in D3.3, we utilize the level of the predictions' confidences throughout the entire cultivation period to display the most confident wrongly declared cases. Specifically, taking into consideration only the last classification run performed at the end of August, **we identify more than 90% of the wrongly declared cases** from the farmers (high PA) and in the 85% of them with the have predicted the correct crop type also. However, the total amount of wrongly predicted cases is much greater than the actual wrongly declared cases, which results in low user's accuracy. The smart sample algorithm (described in D3.3) presented suboptimal performance of recall (almost 35%) against the

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

actual validations provided by NPA. The nature of the smart sampling algorithm lies upon the fact that the most confident predictions of a model reflect the truth, and thus if a prediction does not match with the true label, we assume this is a wrongly declared case. However, the NPA's validation data contain wrongly declared parcels of very similar crop types (in their spectral signatures), such as clovers declared as grasslands, cases that our model would never be able to identify with high confidence. The smart sampling algorithm has been tested successfully in operational scenarios in the past with very high precision. Therefore, a precision of 50% (compared to the validation data) along with photointerpretation that we executed in the results, strongly indicate that are **more cases that are not included into the actual ground truth misclassifications set and can be characterized as actual wrongly declared.**

The dataset of 2020 validation instances was devised from NPA using only remoting-sensing data, since on-the-spot expeditions were not feasible due to COVID-19 outbreak and the constraint regulations existed that period. For that reason, we consider that many of the actual wrongly declared cases were impossible to be distinguished, even from the best of the remote sensing experts, especially in cases where the crops' signatures have very similar behaviour. In Figure 23 and Figure 24 we present 2 parcels that showcase the difference between the average NDVI behaviour of the crop type of the NPA validation (blue colour), the crop type of the classifier's prediction (orange colour) and NDVI time series of the specific parcel (stripped green line). It is clear that in both cases the curve of the sample and the respective average curve of the predicted crop type are much more alike.

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

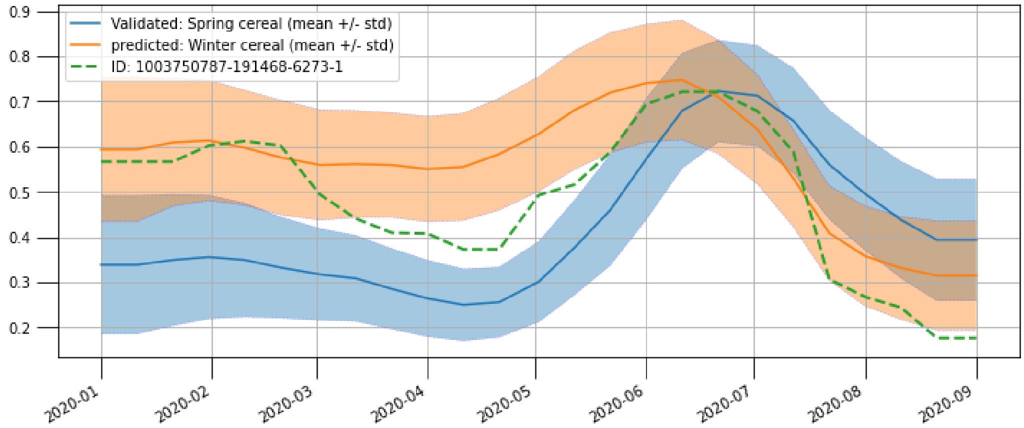


Figure 23: NDVI of a case predicted as Winter Cereal and the label given is Spring Cereal

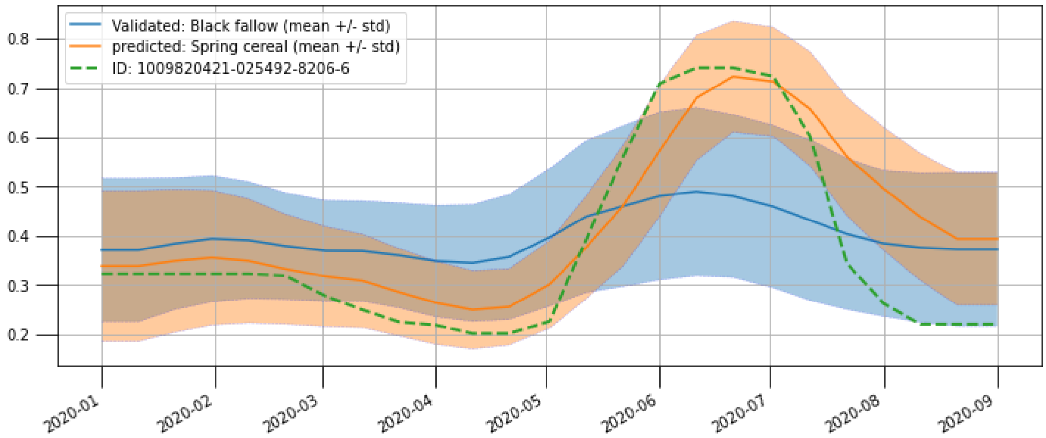


Figure 24: NDVI of case predicted as Spring Cereal and the label given is Black Fallow

Therefore, in order to highlight the efficiency of the smart sampling algorithm, we created a new set of misclassifications based on the predictions of all the RF models trained during cultivation period. This dataset includes all cases that have been identified as alerts (cases that have been predicted differently with high confidence, more information in D3.3) from all different RF models (persistent misclassifications). The threshold of confidence is set accordingly, so as the total number of the persistent misclassifications correspond to the excepted wrong declarations indicated by the NPA.
Using the smart sampling routine, in Figure 25 we can see the final total of candidate wrong declarations as this formed at the end of August where it approximates the total of the persistent misclassifications.

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

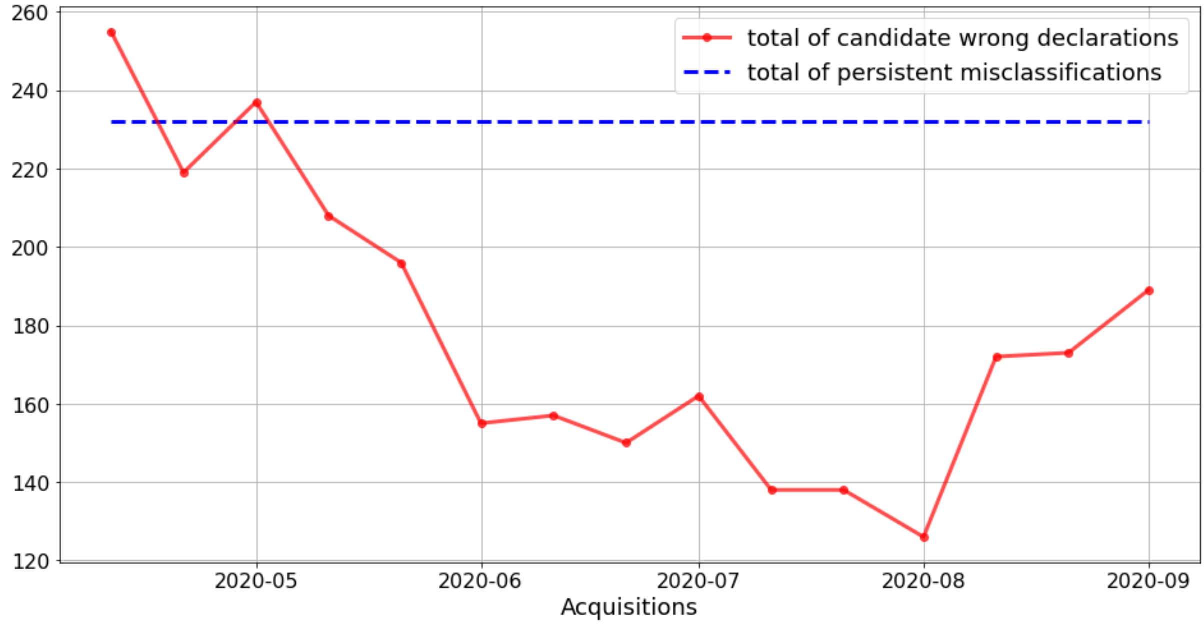


Figure 25: Number of indicated wrongly declared cases modified throughout cultivation period

Furthermore, in Figure 26 below we visualize the progress of Precision and Recall of the routine against persistent misclassifications. These cases were indicated as alerts in more than 8 different models. In order to evaluate even more the results, we conducted photointerpretation in all of the alerts of the last run. The outcomes of this indicated that almost all of these alerts were indeed wrongly declared cases, with an **overall accuracy of more than 90%,** which strengthens even more the assumption that there should be more wrongly declared parcels in the validation dataset.

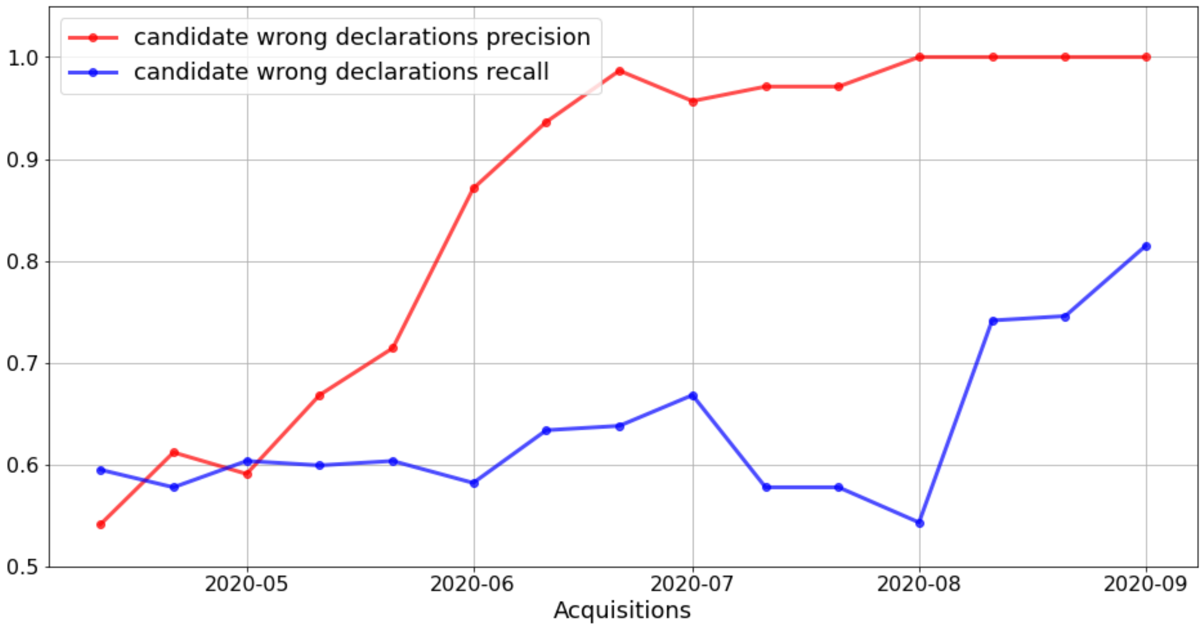


Figure 26: Precision and Recall of smart sampling algorithm over cultivation period

Overall, we have to mention that on the described scenario, we initialize the confidence parameters on a very strict manner in order to optimize the degree of the precision of the indicated cases, displaying only few of them but with very high precision. However, moving towards the scenario of the

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

exhaustive monitoring, we should relax the respective parameters and as result acquire more alerts, which aims at maximizing the recall.

Last but not least, regarding the traffic light output of the product we can visualize the different indicated predictions returned from smart sampling routine based on the taxonomy differentiation between the predicted category and the declared one (as mentioned in D3.3). For instance, cases that have been indicated as alerts from the smart-sampling algorithm and predicted from the classification model as winter crops but have been declared as fallows can be characterized as ***high-risk alerts***, while other cases that the higher level of prediction and declaration is the same, as ***medium-risk alerts***.

3.1.2.3 Grasslands Mowing Events Detection

Several of the following results presented in 2021 AGU Fall Meeting (Kontoes, Charalampos; Tsardanidis, Iasonas; et al. "Deep Learning for Event Detection on Grasslands", B42C-07 presented at 2021 AGU Fall Meeting, 13-17 Dec. https://doi.org/10.5281/zenodo.5995583)

Data Fusion

In order to detect abrupt changes on grasslands, we need uninterrupted optical imagery time-series. Nevertheless, the continuity of the Sentinel-2 time-series is often hindered by intense cloud coverage, which is even more of an issue for Northern European countries, such as Lithuania. In order to tackle this problem, we applied a Deep Learning Architecture based on Recurrent Neural Networks that uses as inputs Sentinel-1 Synthetic Aperture (SAR) data as presented in D3.3, which are weather independent, and the available Sentinel-2 data that are characterized as cloud free observations. Our goal is to **exploit the available Sentinel 1 information,** on the pixel level, which are provided with a constant time step, and the ability of RNN architectures to track temporal patterns, **in order to export continuous and dense NDVI time series**. The performance evaluation was conducted on random time steps, in which we have artificially hidden the actual NDVI values. Results on the Mean Absolute Error (MAE) and Mean Squared Error (MSE) are visualized in Figure 27 below, displaying the difference between using our fusion method and a baseline linear interpolation. The proposed methodology presents **significantly better** performance, with a **mean MAE 0.0279 and mean MSE 0.0017** compared to 0.025 and 0.0071 of the interpolation method.

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

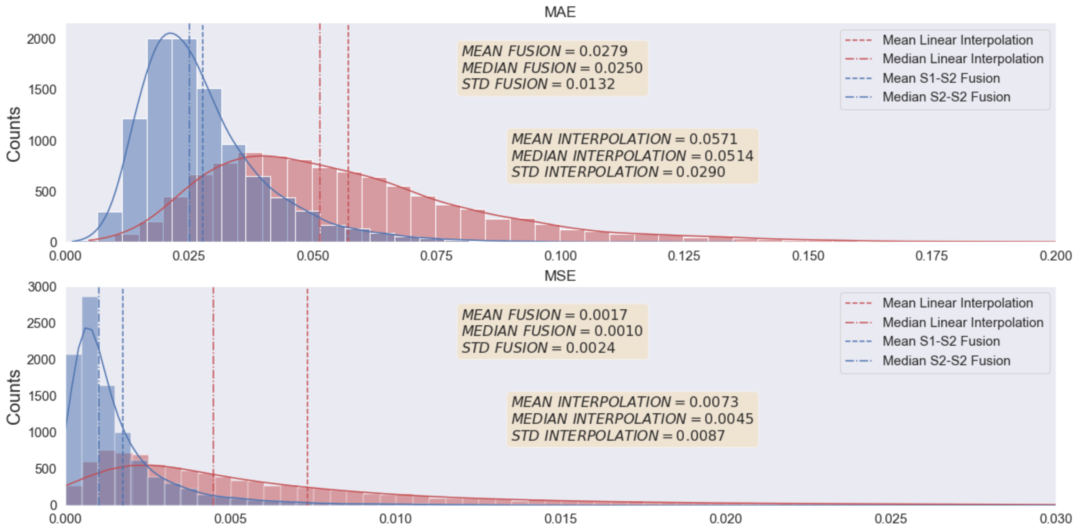


Figure 27: MAE and MSE acquired on randomly hidden NDVI values

Furthermore, preliminary results of our data fusion methodology showed us that this ancillary routine is at place to provide smooth time series with small errors and significantly high correlation compared to the actual measurements, independently of the area (Figure 28).

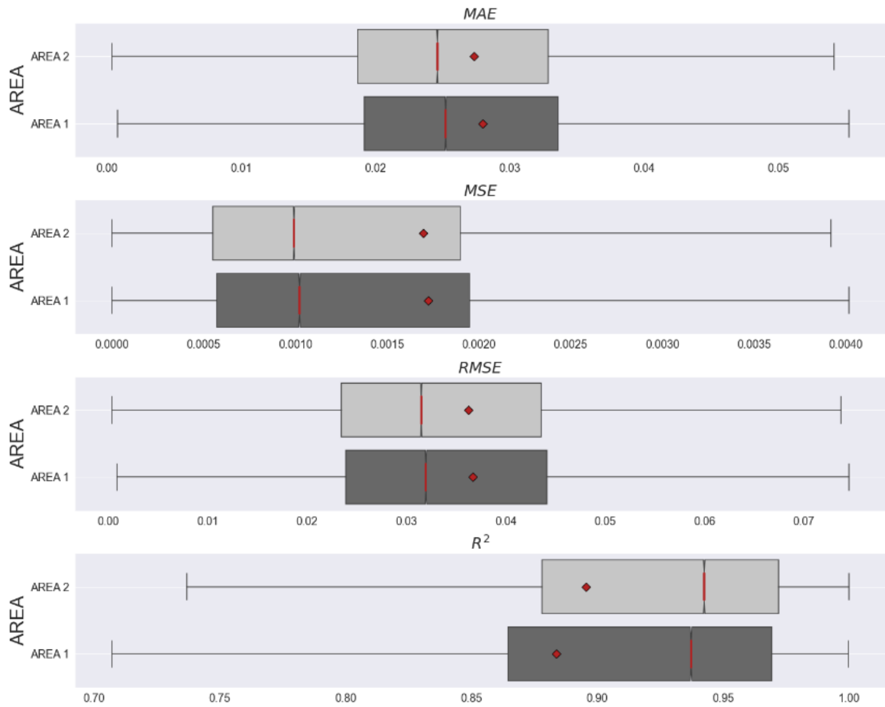


Figure 28: Data Fusion Performance on Areas I and II

Moreover, the S1/S2 data fusion pre-processing step, is able to eliminate the noise coming from individual cloudy cases that preprocessing cloud masks were not able to detect, something that especially for the mowing detection task consist a major problem since these abrupt changes can be erroneous identified as potential mowing events (see Figure 29).

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

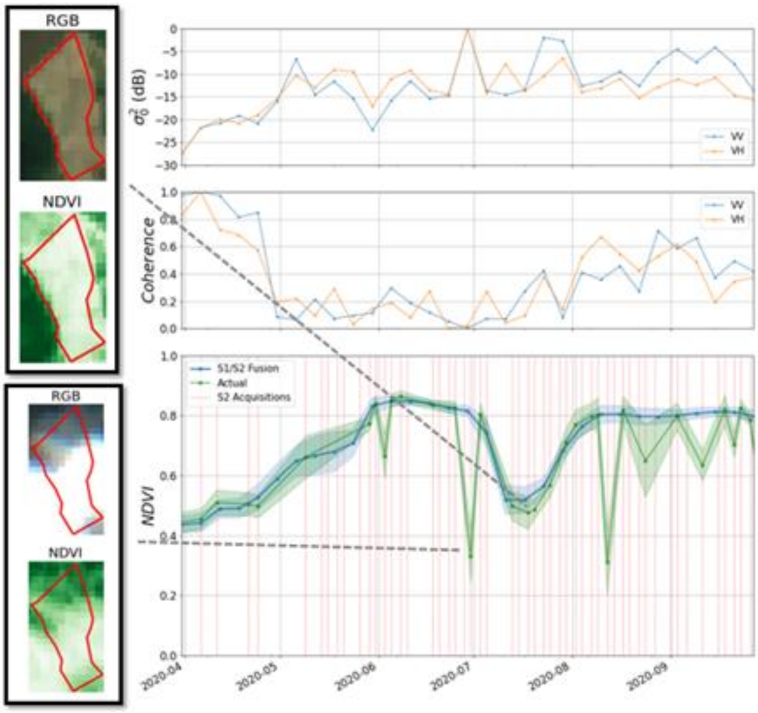


Figure 29: Data Fusion Routine is able to eliminate potential noise coming from non-detected cloud measurements

<u>*Grasslands Mowing Events Detection*</u>

A mowing event usually can be identified from a sudden drop of the NDVI values and a simultaneous increase in the respective SAR bands. For this task, we implement a similar DL architecture with the previous fusion case described above. An RNN based model that uses as inputs the new artificially created NDVI time series of a fixed time step, along with the S1 data (backscatter and coherence) tries to identify the 6-day timeframe (**WHEN**) of an event happened. Subsequently, from the latter and with the usage of the relative mowing regulations depending the type of the grasslands, we can infer the occurrence of an event (**IF**) took place and as a result, the farmers conformity.

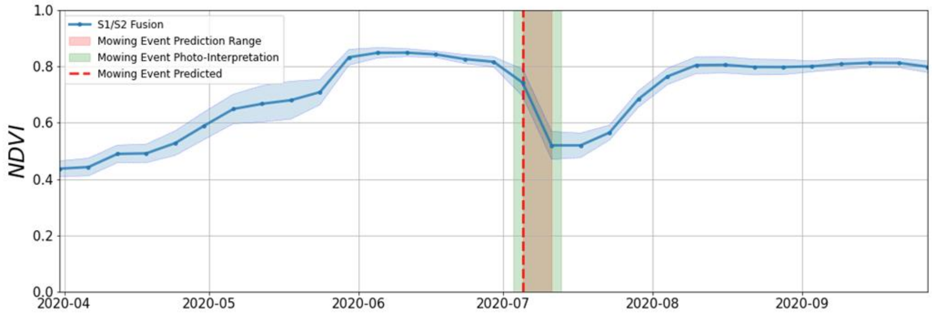


Figure 30: Mowing Event Detected as result of sudden NDVI drop

Results are evaluated parcel wise (since NPA evaluations are addressing to parcel level too) by applying majority voting on initial results extracted on the pixel level. From the scatter plot in Figure 31 we can see that there is a high correlation between predicted and reference date (expressed in Days Of Year), pointing out that in general we are able to identify the respective mowing events precisely enough. Specifically, we can see that in the majority of the cases, the events are detected within a range of less than 12 days, which is a gap of two successive Sentinel acquisitions, and we assume this is due to the

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

subsequent time shifting of the new exported NDVI measurements, after data fusion step described in D3.3.

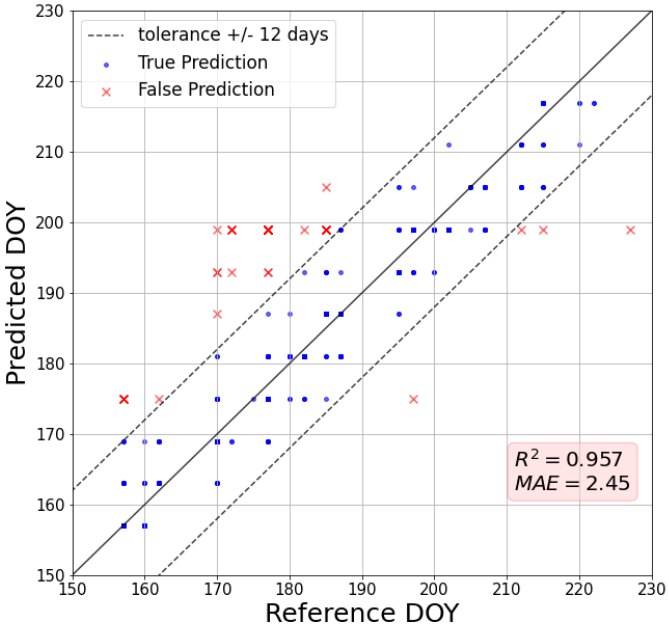


Figure 31: Reference day of the year (DOY) for the mowing events and the DOY predicted by the model

Despite that defining of the exact date of an event can be crucial, it is much more important to monitor if the events took place within a predefined time period, to answer the compliance requirements of the farmers. For that reason, using the respective outputs of the estimated timeframes of mowing occurrences, we can infer the results in the level of conformity by simply applying the respective regulations regarding the grassland type and the period that an event has been identified. Figure 32 below presents the model's performance for the two areas. It should be noted here that for the Overall Accuracy metric for 'when model', we considered correct any prediction that was identified one acquisition before or after the actual event.

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

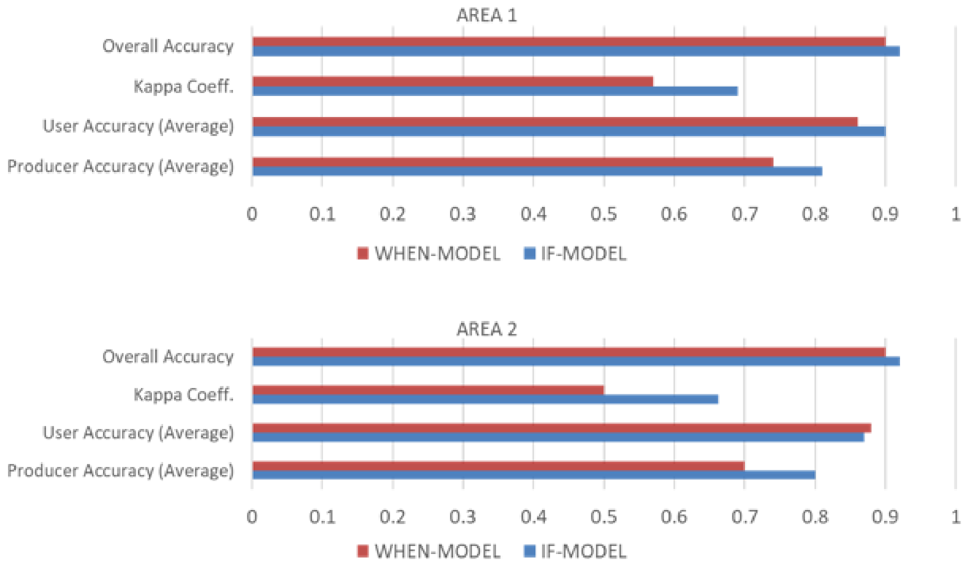


Figure 32: Accuracy assessment of event occurrence for the evaluated parcels for both cases

The performance is satisfactory for both cases and areas. For the case of conformity, the problem is more straightforward and, thus, we achieve a higher Overall Accuracy compared to the detection of the exact time of events. However, the performance in terms of Cohen Kappa Coefficient is not optimal, since compliant some cases are neglected due to the extensive cloud coverage throughout the year. As we mentioned also in D3.3, the actual percentage of non-compliant cases is approximately 5% of the total declarations. On cases of sparse NDVI time-series, the model presents low sensitivity, since it is not able to pinpoint any explicit drop of the NDVI, which characterizes an event. As a result, many compliant cases are characterized wrongly as non-compliant. These are the cases we are trying to eliminate with the addition and the optimization of the aforementioned data fusion routine in the entire pipeline.

Finally, yet importantly, by using the samples exported using the intersection of SE4CAP results and the respective validations we have from the NPA, we assess our methodology, on compliance level, for 5 different validation areas all around Lithuania (Figure 33). Similarly, we notice that the metrics are quite high for almost all areas. We observe, though, a less than optimal performance for *area 3* and *area 1*, where we had a rather sparse time-series due to extended cloud coverage.

Moreover, Table 5 presents precision, recall and F1 score for both classes. The relatively low precision (UA) of the non-compliant cases is due to the cloud coverage issue mentioned as explained earlier.

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

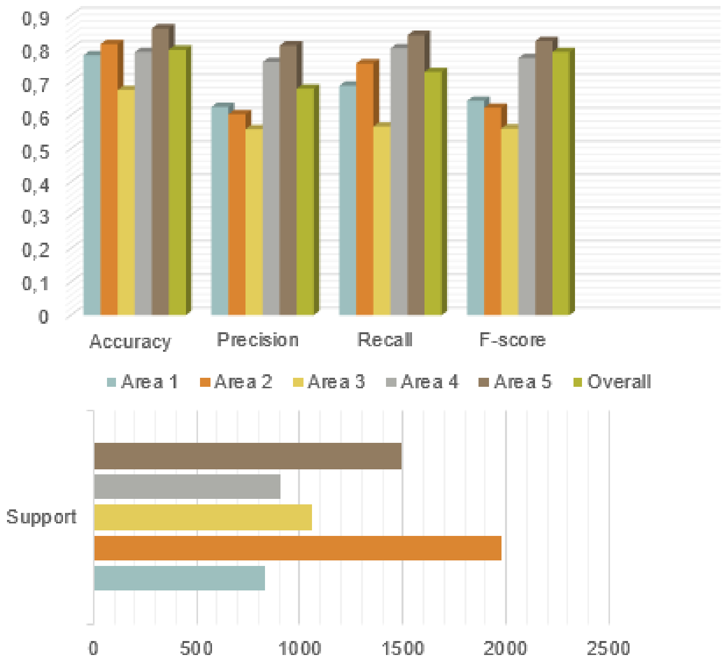


Figure 33: Accuracy Assessment for every sampled area

Table 5: Classification Report for all five sampled areas

	UA	PA	F1-Score	Support
Non-Compliant	0.45	0.64	0.53	1106
Compliant	0.92	0.84	0.87	5164
Macro Avg.	0.68	0.74	0.70	6270
Weighted Avg.	0.84	0.80	0.81	6270
Overall Accuracy			0.80	

3.1.3 Limitations and Next Steps

3.1.3.1 Analytics on Vegetation and Soil Index Time-series

Minimum soil erosion

The main limitation of the methodology, as it has been also highlighted in the results, is the **frequent presence of cloudy pixels**. Thus, the number of cloud-free images is dramatically reduced resulting a dwindling number of data to be analysed. The preliminary version of the algorithm works well on identifying soil cover. However, it does need improvement to increase the overall accuracy. Next steps include the exploration of the effectiveness of the Data Fusion methodology presented in D3.3 in gap filling of SAVI time series. The usage of this product will create **denser time-series**, giving the potential for a more accurate analysis and identification of bare soil. The future work will also focus on testing more complex threshold methodologies, **using multiple vegetation** indices as well as **S1 data to alleviate the cloud coverage** issue. Finally, given that we will acquire annotated data (either from the NPA or by generating them, through blind photointerpretation) we are planning to experiment with

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

machine learning algorithms, to address the issue of thresholding techniques and further enhance the algorithm's performance.

Runoff Risk Assessment for the Reduction of Water Pollution in Nitrate Vulnerable Areas (GAEC 1/ SMR 1)

The runoff risk assessment algorithm takes into consideration the calculation of the water proximity for each vertex of the polygon. It becomes apparent that this computation increases the time complexity, especially for the current business case, as the number of parcels is very high. We are planning to optimize the time complexity by exploiting the parallel processing procedures. A future work will focus on generating useful information and improving runoff risk estimation by introducing more sophisticate modules to the algorithm, such as the aspect of the parcel, to improve the runoff risk estimation. Towards that direction, the calculation of RUSLE may also be revised.

Stubble Burning Identification

As already mentioned in D3.3 the main limitation in the BC of Lithuania is the **extensive cloud coverage**, which results in **sparse NBR time-series**. To restrict the impact of this phenomenon, we are planning to explore the effectiveness of the Data Fusion methodology presented in D3.3 in gap filling of NBR time series. Another issue was **the lack of validation data**. As it has already been mentioned, we are going to use this baseline methodology as a tool to manually **generate an annotated dataset**, through blind photointerpretation of three experts. Finally, we are planning to enhance the existing methodology (with more rules in multiple s1/s2 products) and implement alternative ones as well as apply them on the pixel level and not on parcel level as of now, producing results with higher detail (e.g. fully damaged, partially damaged and no damage).

Harvest Event Detection

Similarly, limitations and future work as in case of Stubble Burning Identification product, are applied also here. The cloud coverage issue will be addressed as well with the incorporation of supplementary S1 data and the threshold-based methodology will be enhanced with more input features, such as other vegetation indices.

3.1.3.2 Cultivated Crop Type Maps (CCTM)

In this deliverable, we provided a preliminary version of this product and the results are very promising. However, difficulties regarding the accuracy of the classification model and efficiency of smart-sampling algorithm emerged. The RF model struggles to identify some specific crop type. A major issue in the Lithuania BC is the extensive **cloud coverage** which does not allow for clean and smooth spectral signatures, which is of a high importance on the current approach. As a first step, we are going to explore **the Data Fusion product** (presented in D3.3) for **gap filling in other crop types and Sentinel-2 indices,** as well as bands, and also generate a plethora of new indices to enhance the feature space. Moreover, **implementation of DL architect**ures that are more robust and resilient to cloud coverage, and do not usually require any pre-processing too, **will be considered**.

Another limitation here is that, since we are using aggregated statistics on the parcel-level, **we do not exploit the spatial context during the modelling process**. Therefore, in the future, we will focus on ML and/or DL approaches that exploit both temporal and spatial characteristics (e.g. TempCNNs) and **we may work on the pixel level**, in order to improve the classification performance and also expand in even more crop types.

The aforementioned actions are expected to improve the classification performance, and consequently, this will also have an impact on the accuracy of the smart-sampling methodology.

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

Finally, based on the user's feedback and moving towards exhaustive monitoring, multiple variations of the smart sampling algorthim may be evaluated.

3.1.3.3 Grasslands Mowing Events Detection

Taking into account the results depicted in the previous sub-section, the main problem we have to deal with again, is that of extensive cloud coverage that characterizes Lithuania and in general Northern-European countries. Frequent cloud appearance throughout the year results in sparse NDVI time-series and obvious consequences in the performance of the mowing event monitoring algorithm. The development and the optimization of an efficient data fusion method is of high significance in order to alleviate the problem described and increase the accuracy of the successive mowing detection model. For that reason, continuous update on the current state-of-the-art regarding S1/S2 fusion methodologies is necessary in order to exploit S1 (coherence and backscatter) data the best way possible. Hence, it will be able to track as many as possible of the mowing events, and subsequently fine-tune our models towards the improvement of the precision of the characterized non-compliant cases. Moreover, by getting rid of cloudy measurements, which is translated also in a NDVI drop, we are about to reduce the percentage of false positive mowing events detected due to sudden drop of the NDVI.

Moreover, **validation data regarding the timestamps of the mowing events is one of the main restrictions** we have faced until now. NPA has provided only validation data with regards to compliance or not, and not the timeframes that the mowing events took place. The acquisition of more validation mowing event timestamps is of utmost importance for both training and validating our models. It is in our current plans to design a more integrated and complete photo-interpretation action in order to generate more such instances, with supplementary incorporation of rapid revisit imagery and VHR data.

Finally, **limitations also appeared in the discrimination of potential grazing events**. These are cases of scattered subtle changes in grasslands areas, mainly for animal husbandry reasons. With Sentinel-2 resolution (~10 m) is impossible to identify changes of this magnitude. The possibility of an integration of VHR imagery on operation level in the future will allow us to reconsider the enhancement of a product answering also this very specific user requirement.

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

3.2.1 Pilot and samples description

3.2.1.1 Analytics on Vegetation and Soil Index Time-series

Retrieve Analytics and Aggregated Statistics on custom Areas using GIS querying on the top of the DataCube

The methodology behind the analytics retrieval and the statistics was described in depth in D3.3. As in the previous business case, this service can be particularly useful for the paying agency of Cyprus, namely CAPO, assisting in the evaluation of the outcomes of the different products and services provided for CAP monitoring purposes (e.g. Stubble Burning Identification, Minimum Soil Coverage, Smart Sampling, etc.), which have been described in detail in D3.3.

Minimum Soil Cover

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

The methodology for detecting a minimum soil cover over parcels to enhance protection of soil against erosion is described in detail in D3.3. In order to validate the results of the methodology, CAPO provided a sample of 2866 parcels, which have been cross-checked via Remote Sensing. Specifically, only 18 of them found to be non-compliant. Figure 34 presents the spatial distribution of the aforementioned fields.

Figure 34: Soil Erosion Validated Samples: Compliant (green) Non-Compliant (red)

Runoff Risk Assessment for the Reduction of Water Pollution in Nitrate Vulnerable Areas

To answer the GAEC 1 and SMR 1 requirement, a runoff risk assessment for the reduction of water pollution in nitrate vulnerable areas has been developed, which takes into account the proximity to the closest surface waters. The related methodology is analysed in D3.3. The validation set provided by CAPO includes 8427 parcels that were checked for compliance of GAEC 1 rule. However, the current methodology cannot exploit this validation dataset, since it requires the hydrographic network as input, which was of extremely poor quality in case of Cyprus and could not be considered as valid.

Figure 35: GAEC 1 non-compliant cases (red) and Nitrate Sensitive Areas of Cyprus (yellow)

Natura 2000 Hotspot Detection

The methodology regarding the Natura 2000 Hotspot Detection algorithm is described in depth in D3.3. As for the validation procedure of this data product, CAPO did not provide any labelled data, regarding illicit activity inside Natura 2000 regions. Due to the absence of any insight about the activity in Natura regions, the results of the algorithm were validated intuitively using photo-interpretation in different areas throughout the country.

Stubble Burning Identification

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

The methodology for the Stubble Burning Identification is a threshold-based technique and it is fully described in D3.3. For the evaluation of the algorithm's performance, we used a sample of 415 observations, located into two small regions (Figure 36), where many neighbouring parcels got burnt simultaneously, probably during small wildfires. The wildfires took place at 20/05/2020 (the right one on the map) and 20/06/2020 (the left one on the map).

Figure 36: Stubble Burning Validation Samples

3.2.1.2 Cultivated Crop Type Maps (CCTM)

he methodology of the Cultivation Crop Type Map product for Cyprus is described in detail in D3.3. For the validation of the product, Sentinel-1 and Sentinel-2 images from the October of 2018 until the July of 2019 were used as inputs. Specifically, for every evaluated field, a sample of 10 random representative pixels extracted using the LPIS buffered geometries for all the s1/s2 band and indices referred in D3.3. A training dataset of more than 100,000 pixels referring to 13 crop classes and a validation one of a similar size were exported from 5 different sampled area across Cyprus territory (see Figure 37) and used for training the ML models and the evaluating their outputs.

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

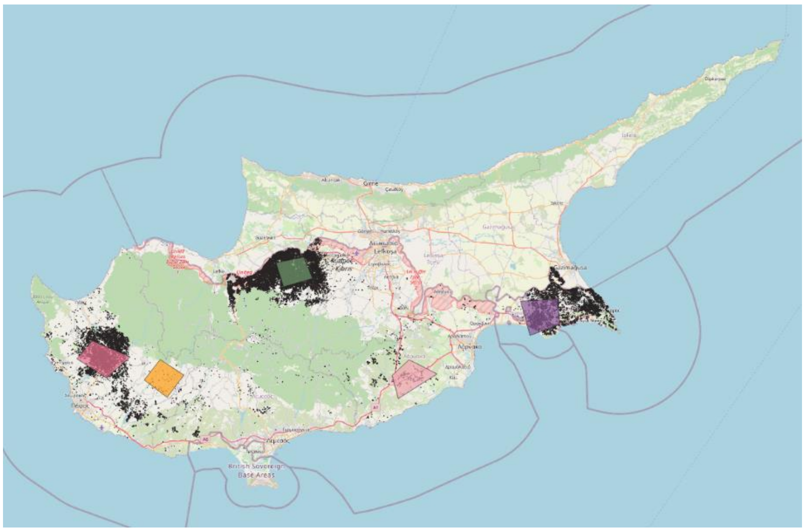


Figure 37: Sampled Areas of Cyprus Validation Crops

3.2.2 Validation results

3.2.2.1 Analytics on Vegetation and Soil Index Time-series

Retrieve Analytics and Aggregated Statistics on custom Areas using GIS querying on the top of the DataCube

1. **Temporal Statistics over an area.** The temporal statistics exploit the time dimension as in the previous sub-task. As already mentioned before for the BC1, the usage of this tool enhances the CAP monitoring by providing useful information about the evolution of the vegetation throughout a user-defined time range. The example below (Figure 38) showcases the validation process for an alarm generated from a back-end process. Specifically, the alarm is related to non-compliance to GAEC 4 rules. Thus, users can effortlessly explore the SAVI and any other index in order to decide whether the alarm is false or not. Complexity of decision is reduced as the tool provides critical information in many dimensions.

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

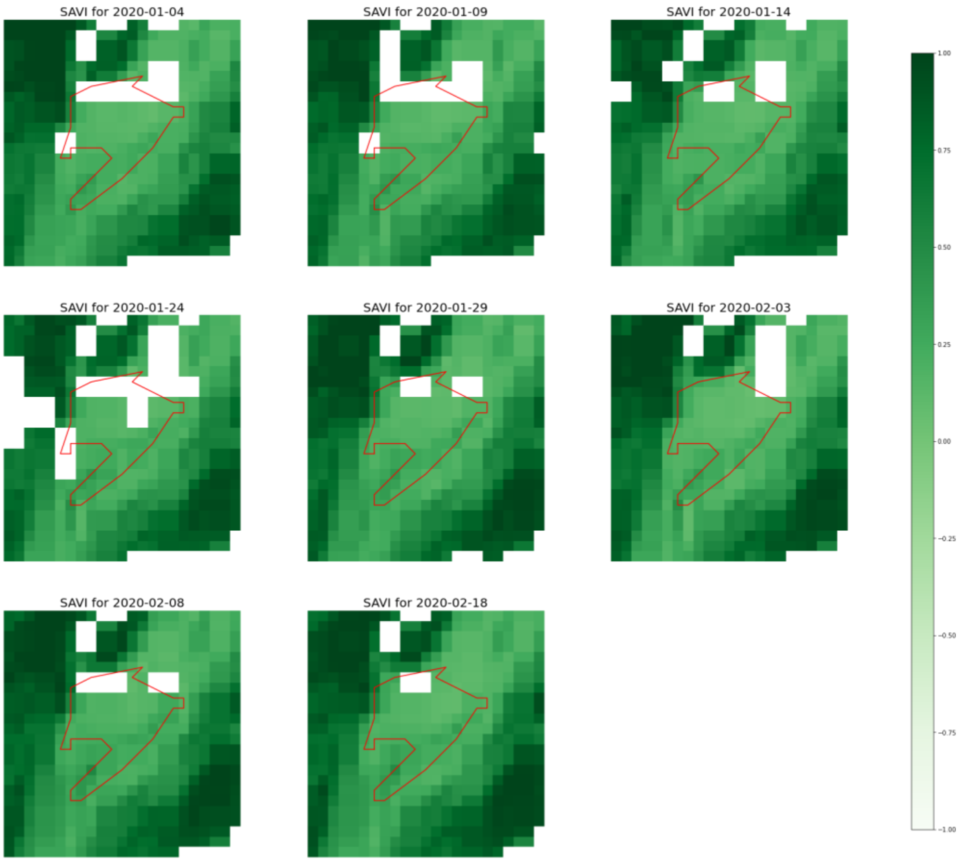


Figure 38: SAVI for parcel flagged as non-compliant to GAEC 4

2. **Smoothed time-series**. Except from investigating each parcel individually, smoothed time series can provide aggregated results for either a parcel or a group of them. Therefore, the mean or any other metric can be calculated and smoothed, aiming at providing noticeable and reveal patterns and trends throughout the years. Figure 39 illustrates the resamples biweekly NDVI for all parcels declared as barley, in Cyprus.

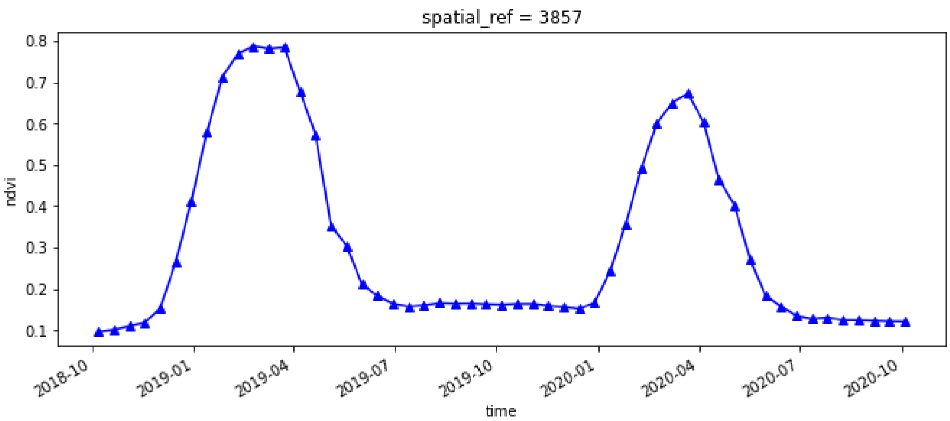


Figure 39: Biweekly NDVI during two years for parcels declared as barley.

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

Minimum Soil Cover

The algorithm for checking GAEC 4 rule performs quite well in detecting soil erosion in Cyprus having an overall accuracy of 98.71%. However, we identify correctly **2819** out of **2843 compliant parcels**, but there are **only 5 correct predictions out of 18 non-compliant parcels** (Table 6). This results to an average recall of 63.5 % and a precision of 58.4%. Moreover, 5 parcels have been totally excluded from the evaluation process due to either extensive cloud coverage or their very small size. The overall performance of the algorithm can be considered **quite satisfying, given the simplicity of this baseline threshold-based approach and the extremely high imbalance between the two classes**.

Table 6: Soil Cover for Soil Erosion Results

		Predicted		
		Compliant	**Non Compliant**	
Ground Truth	**Compliant**	**2819**	24	2843
	Not Compliant	13	**5**	18
		2832	29	

Runoff Risk Assessment for the Reduction of Water Pollution in Nitrate Vulnerable Areas (GAEC 1/ SMR 1)

The estimation and risk analysis for GAEC 1 and SMR 1 respectively requires a full hydrographic layer for the area of interest. Currently, there is a lack of such a layer in Cyprus as mentioned in 3.2.1.1, and thus, water proximity cannot be calculated. However, RUSLE estimation has been already completed and is depicted in Figure 40.

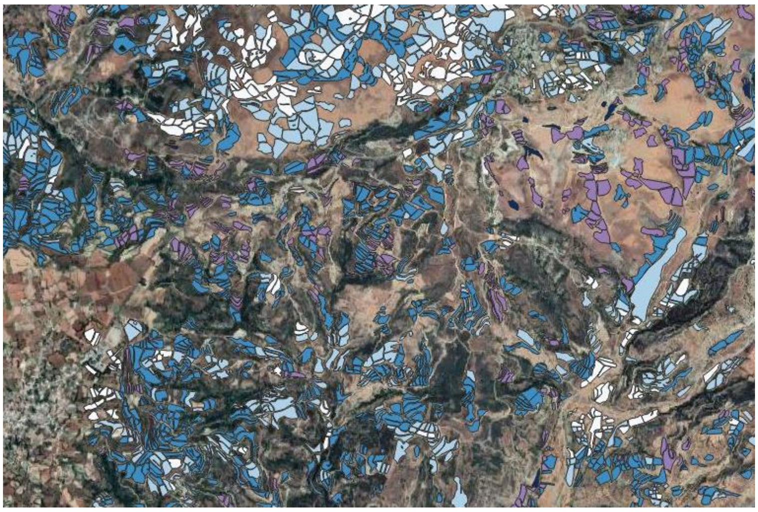


Figure 40: RUSLE estimation

Natura 2000 Hotspot Detection

Natura 2000 Hotspot Detection algorithm is a threshold-based approach, as already mentioned in D3.3. The lack of a validation data does not allow for accurate evaluation, however, by manually

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

inspecting the predictions through photointerpretation, we observed a high performance in terms of recall. The vast majority of the identified events are correct, for both permitted and illegal actions inside Natura 2000 regions. Besides, it is worth mentioning that we apply here the same methodology, as the one described in the Lithuanian BC (BC1) for the harvest event detection. The results of the relative product (Figure 19) also indicate a quite good performance of the respective methodology. Figure 41 illustrates an example, where on the left side an illegal land clearing event is identified, and on the right side it is optically validated.

Figure 41: Land clearing event detected outside arable land parcel considered illegal practice

Figure 42 presents all the land clearing events that were detected from the algorithm (black colour), inside the 'Agia Aikaterini' Natura site in Cyprus (brown colour). This dataset has been provided to CAPO to assist them in a photointerpretation procedure for generating validation data. Finally, by extracting the arable land parcels where activity is permitted, we can provide a final map containing the illegal activities inside Natura2000 Areas.

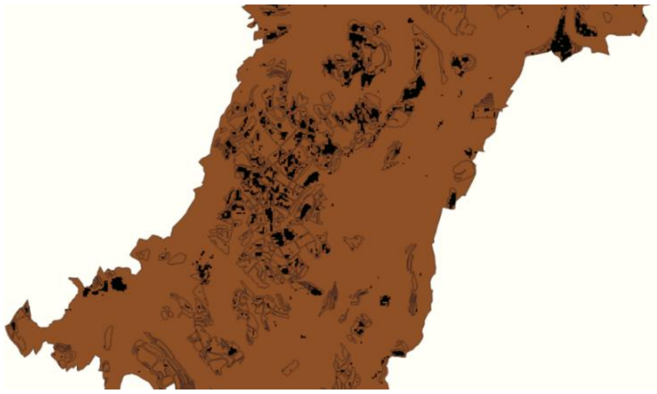


Figure 42: Land clearing events detected in Agia Aikaterini Natura Site

<u>*Stubble Burning Identification*</u>

The methodology of the Stubble Burning Identification is based on two thresholds of NBR and dNBR time-series. The accuracy depends highly on the values of these thresholds. For the proposed values (see D3.3), we classify correctly 264 out of the 439 burnt parcels, which is translated to a recall 60.13%. It is worth mentioning that for each of these cases, we identify the exact timeframe of the event. Moreover, out of the 226,735 examined fields for the cultivation period of 2019, we identified 3,978 burnt areas (approximately 1.8%). We expect that this percentage corresponds to a realistic scenario,

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

however we do not have any relative feedback yet. Therefore, the thresholds can be modified in order to be in accordance with the expected ratios of burnt fields.

Figure 43 visualizes one of the validation areas (red-coloured outline) with burnt parcels (left side). On the right side, we highlight the output of the Stubble Burning Identification algorithm compared to the validation data (red-coloured outline). In blue colour are represented the parcels declared as arable land and in green those declared as fallow land.

Figure 43: Parcels detected to be burnt compared to the provided validated data (validations are with red, arable lands are with blue and fallow lands with green)

As observed above, the algorithm performs relatively well on detecting burnt areas. Moreover, the algorithm succeeds to provide accurate predictions on detecting the date when each event happened. In Figure 44, multiple neighbouring parcels are found to be burnt at a different time, on consecutive acquisitions. These cases are detected from the algorithm together with the corresponding date of event, both of which were validated by experts through photointerpretation.

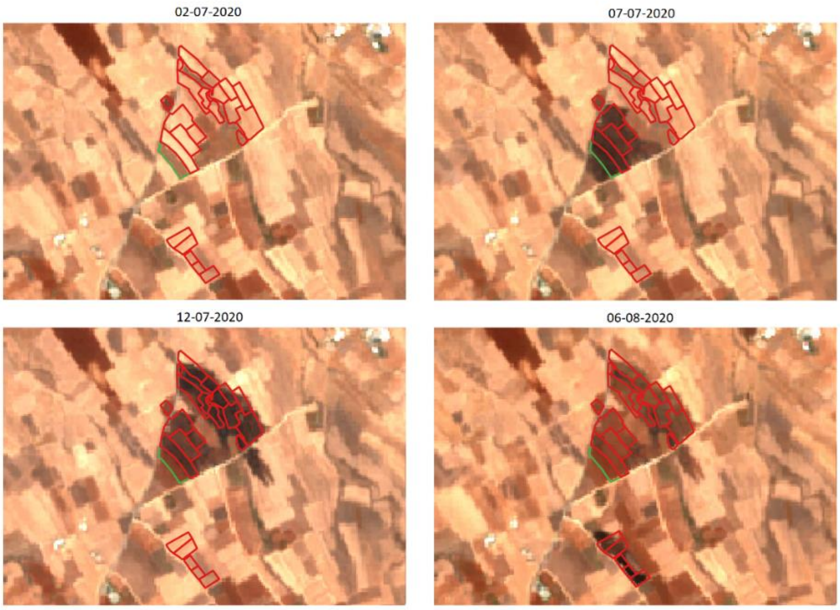


Figure 44: Neighbouring parcels found burnt by the algorithm on consecutive acquisitions (arable lands are with red colour and fallow with green)

One additional service that we provided is the exploitation of the current algorithm, with a view to track the burnt areas from the devastating wildfire took place last summer (July 2021) near Arakapas village in the Limassol. That was a first crash-test of the robustness of our methodology in the general context of Burnt Scar Mapping, where CAPO asked us to apply our model to assist them in the

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

detection of damaged cultivations in the area of interest (AOI). Specifically, initially we exported a raster layer of the estimated pixels detected as burnt, and subsequently based on the percentage of the burnt pixels inside each parcel's geometry, we generate a map layer with the estimated damage. The latter was categorized as "Destroyed","Damaged", "Partially Damaged" or "No Damaged" based on pre-defined thresholds related to the percentage of burnt area, as well as a "Not Assessed" label in cases where clouds were detected. In Figure 45 we present this output map where the deeper the red, the greater the damage was. Overall, user's testimony about this product was very satisfying, claiming that we were able to provide them with quite good results even in the level of discrimination between "Destroyed" and "Damaged", with high accuracy, more precise in comparison to the respective ones from Copernicus Emergency Management Service[1] (EMS) activation and European Forest Fire Information System[2] (EFFIS) (Figure 45). However, few were the cases that missed, especially related with the existence of *greenhouses* in the AOI, which probably should have been marked and excluded from the evaluation in the first place.

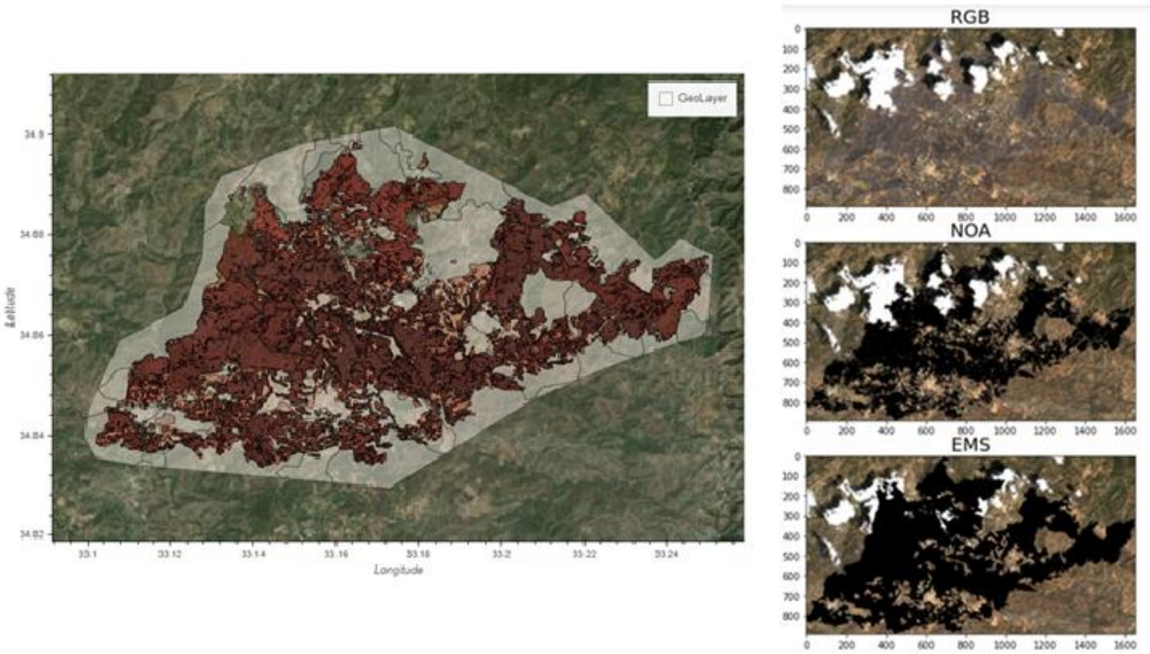


Figure 45: 2021 Cyprus Wildfire damage layer (left), comparison of NOA results with EMS (right)

3.2.2.2 Cultivated Crop Type Maps (CCTM)

In order to evaluate the performance of the SVM hierarchical classification model we trained multiple models throughout the cultivation period, from early March until the end of June of 2019. Similar with the Lithuanian case, the accuracy is increasing gradually as we include more and more acquisitions in the feature space, reaching an optimal performance at the end of June when the cultivation period has been completed. Figure 46 indicates the increase of the Overall Accuracy and the Cohen Kappa coefficient for all the 13 crop classes at the lowest level of taxonomy, in which we performed the

[1] https://emergency.copernicus.eu/
[2] https://effis.jrc.ec.europa.eu/

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

validation analysis. We notice that although both metrics do not reach significantly high levels, the Kappa coefficient metric presents a quite satisfactory performance, based on limitations and particularities of this case (explained in D3.3).

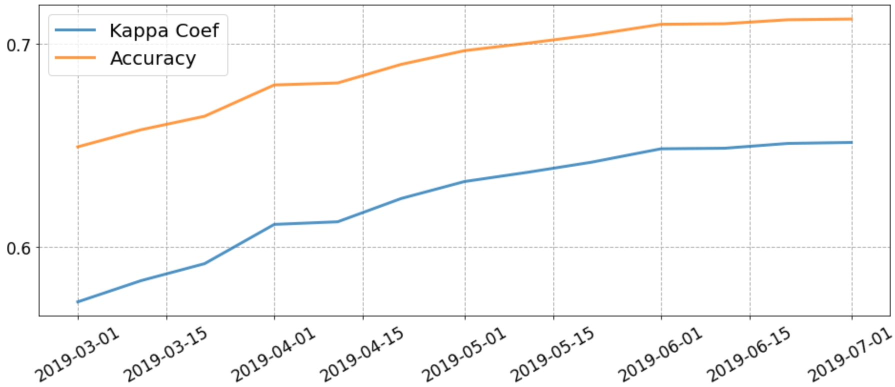


Figure 46: Cyprus Classification Performance over Cultivation Period

As mentioned in the methodology section of D3.3, the fallows have been assessed independently. Specifically, for the rest of the crop classes we have applied a hierarchical ML model approach (based on SVM) in order to exploit information issued from the different layers of crop taxonomy in Cyprus, and map the crops accordingly. Conversely, for the very specific case of fallows, we applied an individual binary SVM classifier on the highest level of the taxonomy to detect only a very small portion of wrong declared fallow cases that the model is very confident.

In Table 7 below, we can see the results, in the pixel-level, acquired at the end of the cultivation period on the lowest level of crop taxonomy, using the hierarchical ML model, which presents a rather mediocre performance. The SVM classifier cannot easily distinguish the different crop types resulting in an accuracy of only 66%. However, for the case of potatoes and vineyards, we observe the highest accuracy, in term of all the metrics, since they have more explicit and clear spectral signatures. As it has been mentioned (section 4.3.1.2 in D3.3), in Cyprus most of the examined crops share very similar spectral characteristics, making the crop identification a very challenging task. Moreover, the small parcel size of the area, which results in a high number of mixels, complicate even more the problem. Thus, we aggregate the predictions in the parcel level, which results in a significant increase in all the metrics (more than 5%), as we can see from Table 11.

Table 7: Classification Report of the main Hierarchical model for the rest classes based on the predictions provided at the late June

	UA	PA	F1-Score	Support
Durum Wheat	0.65	0.55	0.60	21430
Barley	0.65	0.86	0.74	36020
Potatoes	0.85	0.82	0.83	12630
Olive Trees	0.58	0.64	0.61	7980
Citrus Fruit Trees	0.73	0.65	0.69	2110
Vineyards (for wine)	0.71	0.79	0.75	9400
Permanent Grasslands	0.78	0.56	0.65	2050

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

Vegetables Mixture	0.34	0.19	0.24	2050
Traditional Trees	0.55	0.55	0.55	3140
Soft Wheat	0.43	0.15	0.22	9550
Triticale	0.82	0.33	0.48	2220
Deciduous-Fruit Trees	0.31	0.22	0.26	1730
Macro Avg.	0.62	0.53	0.55	110310
Weighted Avg.	0.65	0.66	0.64	110310
Overall Accuracy			0.66	
Kappa			0.57	

Tables 8 and 9 depict the Producer and User Accuracy respectively (on the parcel level), as well as the most frequent cases of misclassification between the actual truth and the predictions. Specifically, the producer accuracy table (

Table 8) 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 some specific crop types, such as Traditional Trees and Olive Trees, a confusion is expected since they belong to the same crop family and share similar characteristics. However, this problem will not have any impact in the subsequent monitoring of

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

crop diversification requirements, since these do not apply on permanent cultivations. Furthermore, by quickly inspecting

Table 8, one can notice that barley is the class that is predicted systematically, when it fails to predict correctly. Barley has the highest samples in the dataset, and even more, the spectral signatures of most crops are very close with each other (see section 4.3.1.2 in D3.3). Therefore, the ML model is not very confident and tends to allocate most of the confusing cases in the barley class.

The user accuracy table (Table 9) 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. Opposite to the

Table 8, we notice here several crop with relatively high accuracies. However, we can notice here the same pattern with regards to what confuses the model. Most of the errors concern the barley crop type. For example, from all the predicted Durum Wheat instances, 33% were in fact Barley. This can be attributed also to the high number of barley samples in the training dataset.

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

Table 8: Cyprus Producer Accuracy Table

	Crop Code	Crop Name	Declared parcels	Well Classified	Producer Accuracy	Confusion class 1	% 1	Confusion class 2	% 2	Confusion class 3	% 3	Rest %
0	1	Durum Wheat	2143.000000	1261	0.588000	Barley	0.330000	Soft Wheat	0.020000	Vineyards (for wine)	0.010000	0.050000
1	3	Barley	3602.000000	3087	0.857000	Durum Wheat	0.060000	Fallow	0.030000	Potatoes	0.010000	0.030000
2	40	Potatoes	1263.000000	1056	0.836000	Barley	0.100000	Fallow	0.020000	Durum Wheat	0.020000	0.020000
3	42	Olive Trees	798.000000	556	0.697000	Vineyards (for wine)	0.160000	Barley	0.060000	Traditional Trees	0.040000	0.060000
4	68	Citrus Fruit Trees	211.000000	145	0.687000	Olive Trees	0.200000	Barley	0.040000	Deciduous-Fruit Trees	0.020000	0.060000
5	70	Vineyards (for wine)	940.000000	807	0.859000	Olive Trees	0.080000	Deciduous-Fruit Trees	0.020000	Durum Wheat	0.020000	0.030000
6	75	Fallow	3325.000000	2715	0.817000	Barley	0.100000	Durum Wheat	0.030000	Potatoes	0.020000	0.030000
7	76	Permanent Grasslands	205.000000	124	0.605000	Barley	0.190000	Olive Trees	0.120000	Traditional Trees	0.040000	0.050000
8	111	Vegetables Mixture	205.000000	39	0.190000	Barley	0.300000	Potatoes	0.270000	Fallow	0.080000	0.150000
9	184	Traditional Trees	314.000000	189	0.602000	Olive Trees	0.200000	Vineyards (for wine)	0.130000	Deciduous-Fruit Trees	0.040000	0.030000
10	226	Soft Wheat	955.000000	131	0.137000	Barley	0.560000	Durum Wheat	0.240000	Potatoes	0.030000	0.040000
11	230	Triticale	222.000000	80	0.360000	Barley	0.350000	Soft Wheat	0.220000	Durum Wheat	0.050000	0.020000
12	233	Deciduous-Fruit Trees	173.000000	35	0.202000	Vineyards (for wine)	0.430000	Olive Trees	0.180000	Traditional Trees	0.090000	0.090000

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

Table 9: Cyprus User Accuracy Table

	Crop Code	Crop Name	Classified parcels	Well Classified	User Accuracy	Confusion class 1	% 1	Confusion class 2	% 2	Confusion class 3	% 3	Rest %
0	1	Durum Wheat	1908.000000	1261	0.661000	Soft Wheat	0.120000	Barley	0.120000	Fallow	0.060000	0.050000
1	3	Barley	5050.000000	3087	0.611000	Durum Wheat	0.140000	Soft Wheat	0.110000	Fallow	0.070000	0.070000
2	40	Potatoes	1272.000000	1056	0.830000	Vegetables Mixture	0.040000	Barley	0.040000	Fallow	0.040000	0.040000
3	42	Olive Trees	897.000000	556	0.620000	Vineyards (for wine)	0.080000	Traditional Trees	0.070000	Fallow	0.050000	0.180000
4	68	Citrus Fruit Trees	180.000000	145	0.806000	Olive Trees	0.090000	Deciduous-Fruit Trees	0.060000	Potatoes	0.020000	0.020000
5	70	Vineyards (for wine)	1135.000000	807	0.711000	Olive Trees	0.110000	Deciduous-Fruit Trees	0.070000	Traditional Trees	0.040000	0.080000
6	75	Fallow	2930.000000	2715	0.927000	Barley	0.040000	Durum Wheat	0.010000	Potatoes	0.010000	0.020000
7	76	Permanent Grasslands	140.000000	124	0.886000	Barley	0.060000	Durum Wheat	0.030000	Vineyards (for wine)	0.010000	0.010000
8	111	Vegetables Mixture	102.000000	39	0.382000	Barley	0.230000	Potatoes	0.170000	Soft Wheat	0.080000	0.150000
9	184	Traditional Trees	298.000000	189	0.634000	Olive Trees	0.090000	Fallow	0.070000	Deciduous-Fruit Trees	0.050000	0.150000
10	226	Soft Wheat	271.000000	131	0.483000	Triticale	0.180000	Barley	0.170000	Durum Wheat	0.130000	0.030000
11	230	Triticale	88.000000	80	0.909000	Soft Wheat	0.030000	Barley	0.030000	Durum Wheat	0.020000	0.000000
12	233	Deciduous-Fruit Trees	85.000000	35	0.412000	Vineyards (for wine)	0.200000	Traditional Trees	0.140000	Olive Trees	0.140000	0.110000

Below is presented the binary SVM's performance in fallow discriminator of the validated cases. Figure 47 displays the F1 score metric only for the "Not fallow" class, for the multiple models throughout the time, which presents an incremental behaviour. The highly imbalanced dataset together with the spectral similarities that fallows share with multiple other crop families makes also this task quite difficult. However, after the last training, the model is able to achieve 35% precision out of the 248 actual wrongly declared cases existed in the validation dataset (see Table 10). For us, this is a very good and auspicious preliminary result, since fallow cases are a very special case, with not at all very clear spectral signature, of very high discrepancy, very similar to the rest of the crop types.

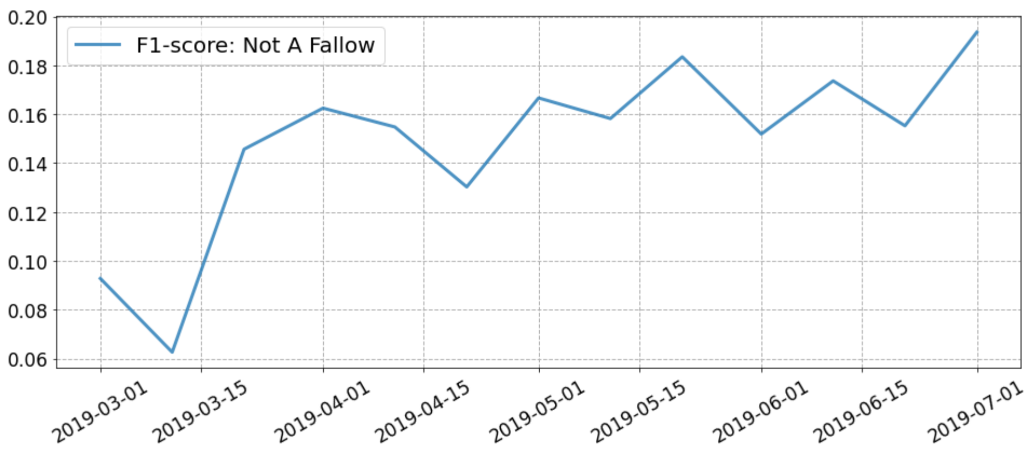


Figure 47: Fallows model Classification Performance over Cultivation Period

Table 10: Classification Report of the Fallows binary model based on the predictions provided at the late June

	UA	PA	F1-Score	Support
Not-Fallow	0.35	0.14	0.2	248

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

Fallow	0.93	0.98	0.95	2775
Macro Avg.	0.64	0.56	0.57	3023
Weighted Avg.	0.88	0.91	0.89	3023
Overall Accuracy			0.91	
Kappa			0.16	

By taking into account the aforementioned, for the declared fallows that have been identified as "not fallow", we predict their crop type using the main Hierarchical ML routine. Overall, in Table 11 and Figure 48, the final classification results are presented for the finest level of the crop taxonomy. These are aggregated results into the parcel level, which results in significant improvement compared to pixel-based classification (Table 7).

Table 11: Classification Report on the finest level of crops taxonomy at late June

	UA	PA	F1-Score	Support
Durum Wheat	0.66	0.59	0.62	2143
Barley	0.61	0.86	0.71	3602
Potatoes	0.83	0.84	0.83	1263
Olive Trees	0.62	0.70	0.66	798
Citrus Fruit Trees	0.81	0.69	0.74	211
Vineyards (for wine)	0.71	0.86	0.78	940
Fallow	0.93	0.82	0.87	3325
Permanent Grasslands	0.89	0.61	0.72	205
Vegetables Mixture	0.38	0.19	0.25	205
Traditional Trees	0.63	0.61	0.62	314
Soft Wheat	0.48	0.14	0.21	955
Triticale	0.91	0.36	0.52	222
Deciduous-Fruit Trees	0.41	0.20	0.27	173
Macro Avg.	0.68	0.57	0.60	14356
Weighted Avg.	0.71	0.71	0.70	14356
Overall Accuracy			0.71	
Kappa			0.65	

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

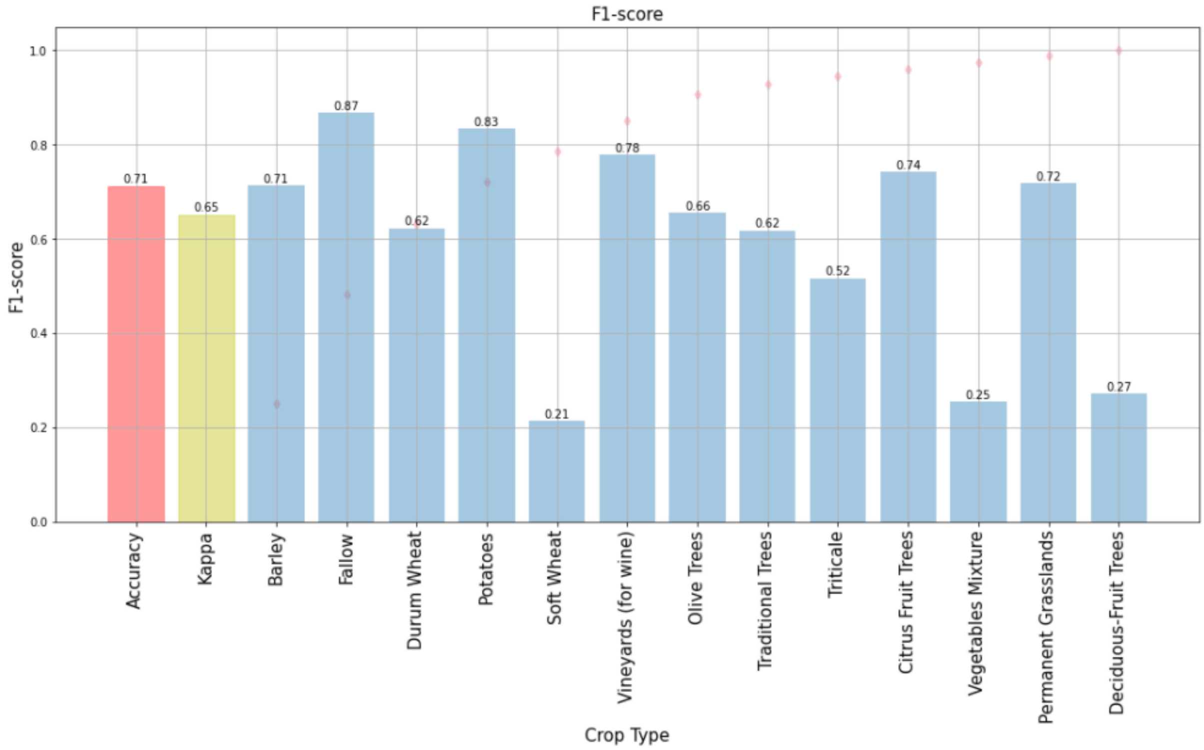


Figure 48: F1 score based on the predictions provided at the late June on the finest level of crops taxonomy

Finally, classification in higher level of taxonomy can be sufficient to address some of the requirements (see D3.3). Thus, using the respective taxonomical categorizations (see Figure 30 "Crop Taxonomy for Cyprus" in D3.3) we can extract the results for the most general crop families (see Table 12 and Figure 49). The performance there, it is very satisfactory, reaching an Overall Accuracy of 86% and Kappa Cohen Coefficient of 0.8, respectively.

Table 12: Classification Report on the middle level of crops taxonomy at late June

	UA	PA	F1-Score	Support
Tree Crops	0.80	0.78	0.79	1496
Vines	0.71	0.86	0.78	940
Cereals	0.89	0.94	0.91	6922
Broadleaf Crops	0.83	0.84	0.83	1263
Vegetables	0.38	0.19	0.25	205
Grasslands	0.89	0.61	0.72	205
Fallow	0.93	0.82	0.87	3325
Macro Avg.	0.77	0.72	0.74	14356
Weighted Avg.	0.86	0.86	0.86	14356
Overall Accuracy			0.86	
Kappa			0.80	

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

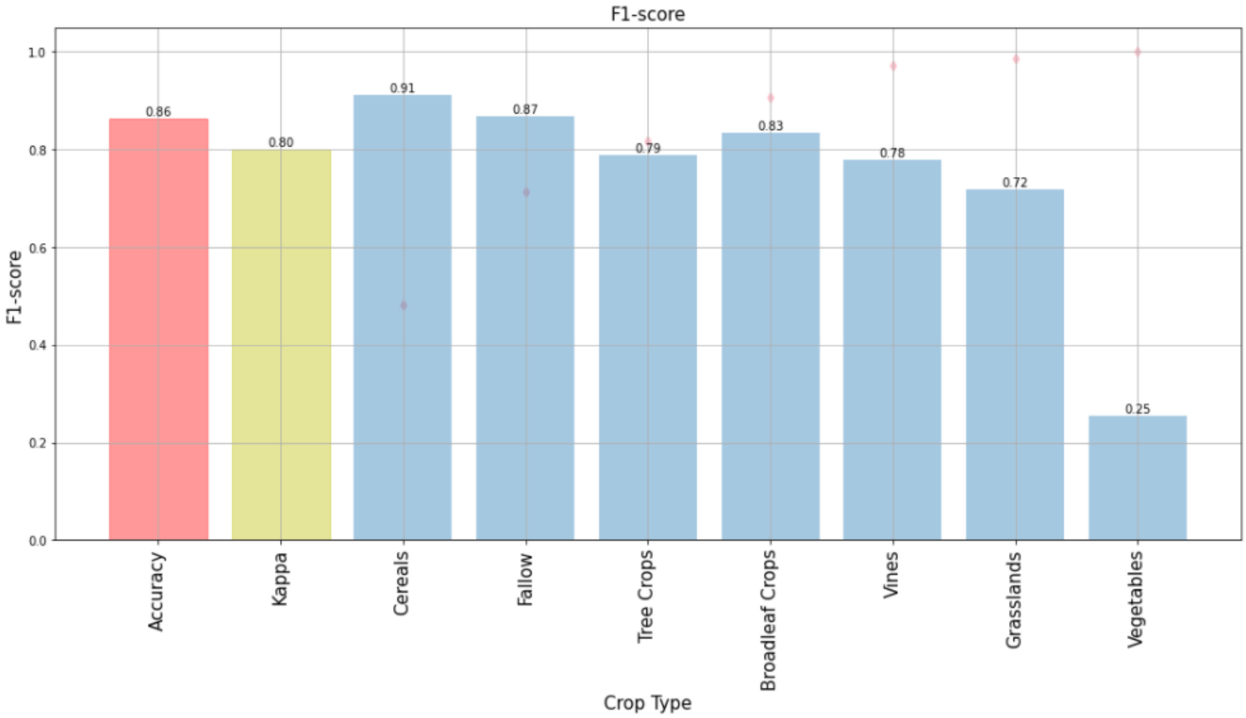


Figure 49: F1 score based on the predictions provided at the late June on the middle level of crops taxonomy

3.2.3 Limitations and Next Steps

3.2.3.1 Analytics on Vegetation and Soil Index Time-series

Minimum Soil Cover

The main limitation in the Business Case of Cyprus is related to **the small size of parcels**. The pixels included in the parcels may be not enough for the algorithm to decide whether there is or not a minimum soil cover. Another issue here is that despite the satisfying amount of validation data, its **high imbalance does not allow for an accurate evaluation of the algorithm's performance**. Therefore, in the near future we will try to collect more non-complaint cases, either with photointerpretation or from OTSCs performed by CAPO. Since the main methodology is the same here as in the Business Case of Lithuania, next steps include also development of **more sophisticated threshold-based algorithms, using multiple vegetation indices as well as S1 data** to alleviate potential cloud coverage, or even **machine learning techniques** given that the **required annotated data** can be collected.

Runoff Risk Assessment for the Reduction of Water Pollution in Nitrate Vulnerable Areas (GAEC 1/ SMR 1)

On the preliminary version of run-off risk assessment for the BC2, RUSLE estimation has been calculated. **Nevertheless, there was no potential for a risk analysis as the hydrography layer provided to us has insufficient data**. The latter considered to be so far the main limitation. Therefore, next steps include the computation of the risk analysis along with an evaluation of the results based on the dataset provided by the CAPO.

Natura 2000 Hotspot Detection

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

As already mentioned in D3.3, the main limitation in the data product of Natura 2000 Hotspot Detection for Cyprus is the **very small parcels and the lack of validation data.** This means that a lot of pixels, in many occasions the majority of the pixels included inside parcels' boundaries, are half inside the parcel and half outside (mixels). For this reason, the algorithm was developed on the pixel level to produce as accurate results as possible. The Natura 2000 Hotspot Detection algorithm, which detects land clearing events using NDVI from Sentinel-2 satellite images, yields satisfactory results so far. In the forthcoming months, Sentinel-1 weather independent images (coherence and backscatter) will be introduced to this data product, in an attempt to fill in gaps in the NDVI time-series caused by cloud coverage. Moreover, we are planning to develop **more sophisticated rule-based approaches that use more vegetation indices**. Finally, a manually **annotated validation dataset will be produced** in order to extract statistics and allow an accurate evaluation of the results.

Stubble Burning Identification

As already mentioned in D3.3 and above, the main limitation in the data product of Stubble Burning Identification for Cyprus **is the very small size of the parcels**. As a result, the algorithm, which is based on calculating the average values of pixels for each parcel, is prone to errors when performed on very small parcels. For this reason, the algorithm will be applied on the **pixel level** both to augment the data from each parcel and to **produce results with higher detail (e.g. fully damaged, partially damaged and no damage)**. Next steps will also focus on improving the performance of the algorithm by testing different thresholds or different approaches and by including more Sentinel-2 indices.

3.2.3.2 Cultivated Crop Type Maps (CCTM)

On this preliminary version of Crop Type Mapping for the BC2, we provided some initial results for the particular case of Cyprus. As we saw, it is a very challenging case and the discrimination on the lowest level of crop taxonomy, required for the monitoring of Greening regulations, is not optimal yet. **The main problem of Cyprus is the small size and the long and narrow shape of the parcels** that do not assist in the extraction of representative spectral characteristics for the various crop types. Work should be done in order to improve the level of performance and be at place to provide acceptable results that will allow us to assimilate them in the smart-sampling algorithm. **The adaptation of a pixel-wise approach is essential** and we are also planning to assess other algorithms (e.g **ML/DL architectures**) in order to **fully exploit the geo-spatial information** of the area, as well as to decide on the integration of additional visual indices into the model's input feature space. Moreover, **the smart sampling algorithm will be modified accordingly in order to satisfy the specific BC requirements**. Moreover, the possibility of an integration of VHR imagery on operation level in the future, will allow us to consider the fusion of supplementary image sources with a view to improve the final results. Last but not least, further **thorough evaluation and improvements will be done for the very special case of fallows** and the respective methodology. In the current scenario, we designed an approach that is able to identify other crop cases that have been declared mistakenly as fallows. Finally, we have to examine also the opposite case as well since it is equally frequent to happen according to our validation data, where we need to identify the actual fallows that have been declared wrongly from the applicants as another crop type.

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

3.3 BC3: Monitoring the condition of soil – Belgium

3.3.1 Pilot and samples description

The Flemish business case focuses on deploying ENVISION service for topsoil Soil Organic Carbon Monitoring in Flanders, Belgium. Currently, the state of agricultural soils is checked by performing soil samplings and conducting laboratory examinations. However, these methods do not provide a continuous overview of soils' state and require significant effort, time, and resources to be committed. Consequently, these types of controls have to be significantly reduced and replaced with a more automated process. The business case is implemented in Belgium, within the Flemish region, involving LV, the Flemish Department of Agriculture and Fisheries and Paying Agency, which is in Flanders' the official PA in charge of the financial support for agriculture and the implementation of CAP. EV ILVO will assist LV, and is responsible for developing the data products that can support the provision of the services. Additionally, in close collaboration with EV ILVO to perform the soil campaign and assess the soil organic carbon model performance.

3.3.1.1 Methodology

To achieve user requirements and other non-functional requirements related to service scalability, we define a methodology that can enable current scientific research outcomes and deliver soil organic carbon products on a large scale. More specific:

- In **Phase One**, the main goal is to develop a <u>Cloudless Collection of Bare Soil Pixels</u>. We filter the original collection by applying cloud masking. After the bare soil pixels were selected by masking using computated indices that can detect green and dry vegetation and high soil moisture content that can affect the soil spectrum shape. We also filter the collection layer by using the ESA Worldwide land cover mapping. Alternatively, we can use the Land Parcel Identification System (LPIS) parcels provided by LV in the masking process. However, that requires more computational power and the ESA Worldwide land cover mapping covers equally the existing croplands and the grasslands.
- In **Phase Two**, we perform the modelling and there, the goal is to create a mapping between the extracted reflection signatures coming from the Bare Soil Collection and the SOC measurements.
- In **Phase Three**, we apply the deployed model to all pixels belonging to larger areas, at regional (Envision, BC3) or even national scale. At this phase, we aggregate the SOC from pixel level to parcel level, and critical decisions are made on presenting the information to the service consumers or the end-users.
- In **Phase Four**, we perform the technical validation and service evaluation, which means validating a complete solution or a segment of a solution that is about to be or has already been implemented to determine how well a solution meets the business needs and delivers value to the organization.
- In **Phase Five**, you need to evaluate and make improvements, considering how the changes at each phase affect the product in other phases and the service itself. It's a critical phase because it supports traceability and monitoring, which means approving and assessing changes to product information to manage it throughout the business analysis effort.

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

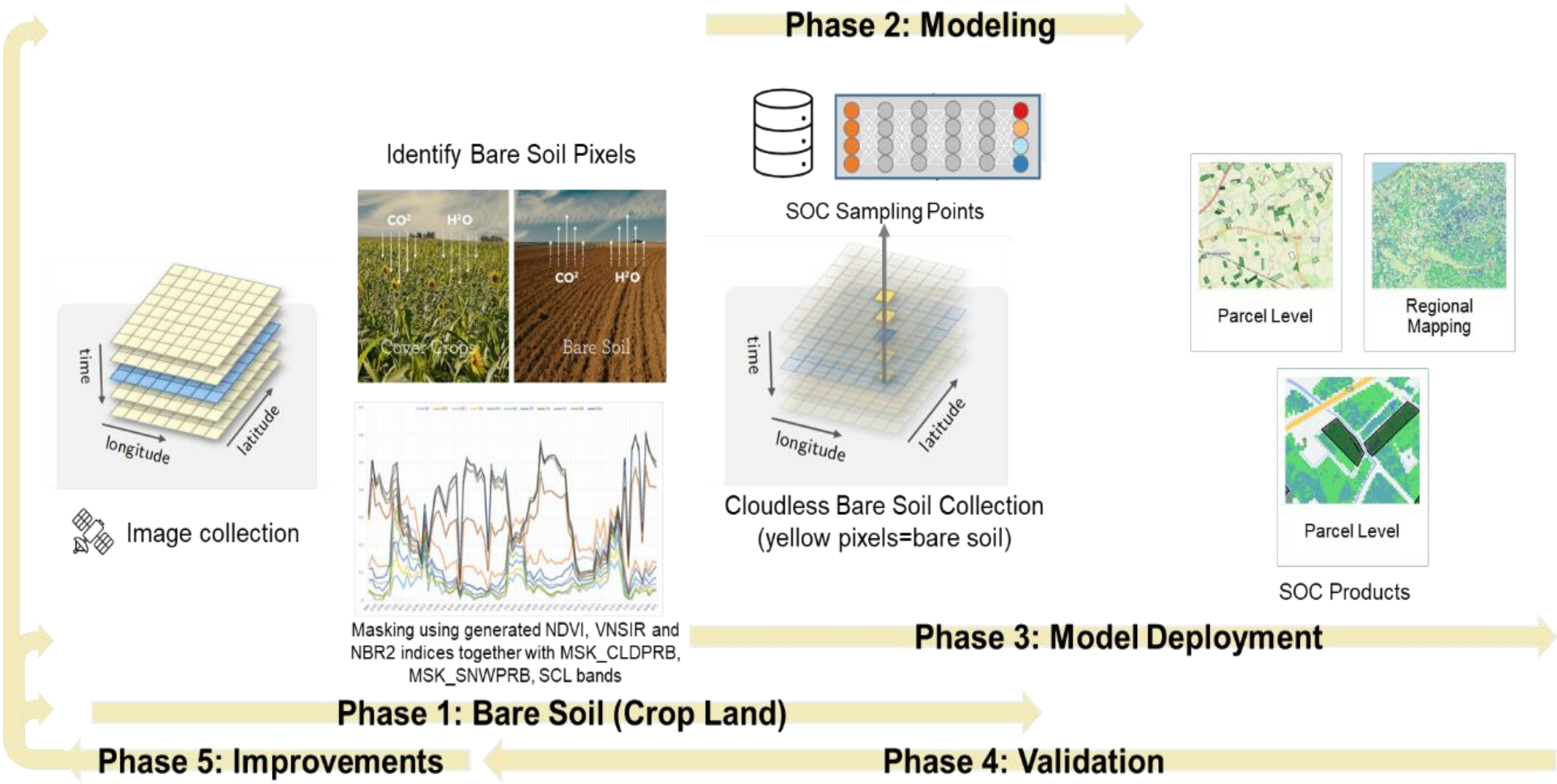


Figure 50: Significant methodological phases

The developed technological environment supports the implementation of the aforementioned methodological phases and provides the needed scalability to leverage Copernicus data resources at the Regional, National and European level. It relies on the Google Earth Engine system, the Colab notebook collaboration environment and PYCARET. PYCARET is an open-source, low-code machine learning library in Python that automates machine learning workflows and supports data preparation, model training, hyperparameter tuning, analysis and interpretability, and model selection.

3.3.1.1.1 Satellite image acquisition and pre-processing

The collection consists of 4598 images of L2A that covered the Flemish region from 2018 until the end of April 2021. The first step was to create a collection of S2 images using the GEE Python APIs to access Data Catalog products (Sentinel 2 MSI, Level 2A) following the latest Legal notice on the use of Copernicus Sentinel Data and Service Information.

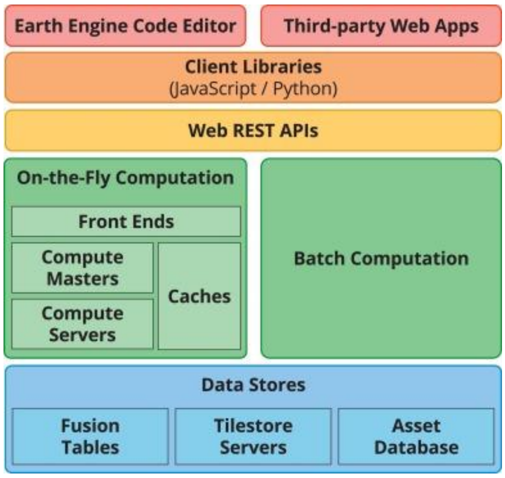


Figure 51. Google Earth Engine (GEE) system architecture diagram. GEE relies on a Client/server programming model.

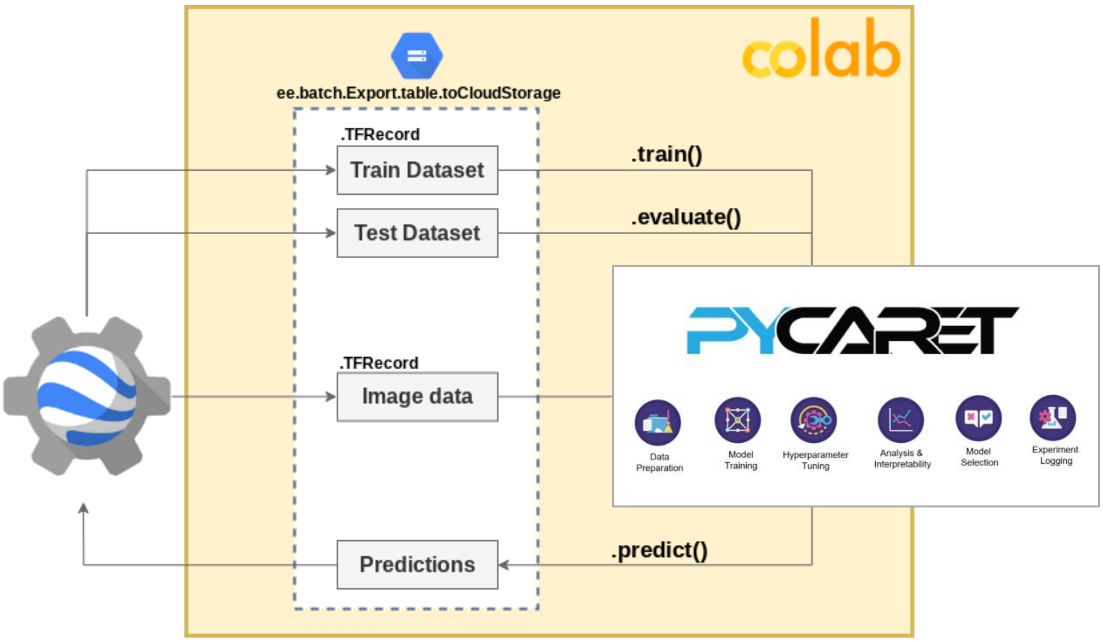


Figure 52. Together with Colab and Pycaret, GEE creates a robust technological framework that allows collaboration, ensures scalability, and supports productivity.

To apply cloud masking, we make use of the S2-L2 bands:

- MSK_CLDPRB 20 meters Cloud Probability Map (missing in some products)
- MSK_SNWPRB 10 meters Snow Probability Map (missing in some products)
- QA60 60m Cloud mask.

And to identify the bare soil pixels, we created and applied a set of extra masks using the NDVI, VNSIR and NBR2 indices.

Indices	Formulas	Upper and Down thresholds suitable to identify for Bare Soil
NDVI	(B08-B04)/(B08+B04)	>-0.25 and <0.35
NBR2	(B11-B12)/(B11+B12)	>0 and <0.1
VNSIR	(2 * RED) - GREEN - BLUE) + (3 *(SWIR2 - NIR)	>0.1

Table 13: To identify the bare soil layer, we created and applied a set of extra masks using the NDVI, VNSIR and NBR2 indices.

Figure 53. RGB Visualization of the bare soil synthetic layer.

```
var cloudless_bare_soil_collection = ee.ImageCollection('COPERNICUS/S2_SR')
                                     .filterDate(startDate, endDate)
                                     .map(maskS2clouds)
                                     .map(maskvlaams)
                                     .map(maskCropland)
                                     .map(addNDVI)
                                     .map(addNBR2)
                                     .map(addVNSIR)
                                     .map(addBSI)
                                     .map(addcount)
                                     .map(maskNDVI)
                                     .map(maskNBR2)
                                     .map(maskVNSIR
```

Figure 54. Part of the code in Java Script, describes the processing and masking steps applied to develop the Cloudless Bare Soil Collection.

3.3.1.1.2 EO data processing and analysis workflow/ algorithms in an automated manner

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

Using the GEE Reducers[3] and the Exporting Data[4] ability, we generate raw data per sampling point (Figure 55). It's possible to automatically extract the completed value set without applying bare soil masking (by activating only the cloud mask function), supporting the data analysis without using the median values per band and sampling point.

B1	B11	B12	B2	B3	B4	B5	B6	B7	B8	B8A	BSI	NBR2	NDVI	VNSIR	NSMI	date	point_id
0.0559	0.1545	0.086	0.0332	0.0564	0.0352	0.0939	0.3243	0.484	0.4823	0.5052	-0.4619966	0.2848233	0.8639614	2.0713	0.2848233	6/3/2018	2
0.068	0.1937	0.1149	0.0415	0.0786	0.066	0.1345	0.3208	0.4136	0.444	0.4646	-0.3030059	0.2553467	0.7411765	1.8008	0.2553467	6/13/2018	2
0.3544	0.3254	0.3128	0.1318	0.1755	0.2134	0.2833	0.4155	0.4855	0.3145	0.5097	0.0938991	0.019743	0.1915135	1.4334	0.019743	6/15/2018	2
0.0324	0.2453	0.1612	0.0421	0.0876	0.137	0.1974	0.2613	0.3115	0.3475	0.3806	-0.0094572	0.2068881	0.4344685	1.2616	0.2068881	6/25/2018	2
0.0387	0.2502	0.1683	0.05	0.0812	0.1334	0.1704	0.2159	0.2655	0.2901	0.3209	0.0601078	0.1956989	0.3700118	1.0765	0.1956989	6/28/2018	2
0.0393	0.2821	0.1793	0.0579	0.1051	0.1799	0.2219	0.25	0.2807	0.3267	0.3514	0.0914245	0.2228002	0.289775	1.0111	0.2228002	6/30/2018	2
0.0592	0.3165	0.2073	0.0809	0.1196	0.1885	0.2316	0.2753	0.3176	0.3492	0.39	0.0800984	0.2084765	0.2988655	1.044	0.2084765	7/5/2018	2
0.0139	0.2903	0.1857	0.0478	0.0868	0.1611	0.1985	0.2297	0.271	0.2984	0.3367	0.1318957	0.2197479	0.2988031	0.9516	0.2197479	7/8/2018	2
0.1183	0.4243	0.2908	0.1529	0.1976	0.2471	0.3069	0.3975	0.4556	0.48	0.51	0.0295177	0.1866872	0.3203136	1.1134001	0.1866872	7/13/2018	2
0.063	0.483	0.3331	0.1111	0.1617	0.2522	0.3054	0.3491	0.3872	0.4246	0.4528	0.1569754	0.1836785	0.2547281	0.6778	0.1836785	7/15/2018	2
0.0499	0.4403	0.303	0.0938	0.1429	0.2225	0.2654	0.2973	0.3315	0.3535	0.3925	0.1941266	0.1847168	0.2274306	0.6483	0.1847168	7/23/2018	2
0.0537	0.3653	0.2827	0.1007	0.1433	0.1985	0.2338	0.2563	0.2854	0.3005	0.32	0.1684974	0.1274691	0.2044088	0.7111	0.1274691	8/2/2018	2
0.0301	0.314	0.2794	0.0866	0.1255	0.1656	0.1866	0.2324	0.2602	0.2631	0.2828	0.1566381	0.0583081	0.2274318	0.7873	0.0583081	8/7/2018	2
0.0365	0.2294	0.1532	0.0812	0.1135	0.1537	0.1745	0.1895	0.2087	0.2214	0.2354	0.1173983	0.1991636	0.1804852	0.9053	0.1991636	8/12/2018	2
0.0576	0.2751	0.197	0.0952	0.1211	0.1571	0.1836	0.2282	0.2551	0.2545	0.2878	0.1055122	0.165431	0.2366375	0.9402	0.165431	8/14/2018	2
0.0584	0.3733	0.3031	0.1182	0.1628	0.1962	0.234	0.3032	0.3479	0.3562	0.3751	0.0911007	0.1037847	0.2896452	0.894	0.1037847	8/17/2018	2
0.0356	0.1978	0.1045	0.0422	0.0841	0.0539	0.1381	0.312	0.3486	0.3603	0.3606	-0.2305106	0.3086338	0.7397393	1.5069	0.3086338	9/1/2018	2
0.0321	0.2189	0.1101	0.0436	0.0907	0.0432	0.1437	0.465	0.5403	0.5271	0.5601	-0.3705572	0.3306991	0.8485008	2.0715	0.3306991	9/11/2018	2
0.1216	0.2965	0.1815	0.1419	0.1716	0.1283	0.209	0.5144	0.6229	0.6279	0.6357	-0.2887996	0.2405858	0.6606718	2.0745	0.2405858	9/13/2018	2
0.0172	0.1988	0.0981	0.0262	0.0675	0.0261	0.1155	0.4809	0.5822	0.6048	0.5856	-0.4744713	0.3391714	0.9172611	2.2019	0.3391714	9/21/2018	2
0.0116	0.1792	0.0915	0.0271	0.0646	0.0245	0.1144	0.4085	0.5076	0.534	0.515	-0.4673117	0.3239749	0.912265	2.0501	0.3239749	9/26/2018	2
0.0123	0.0489	0.0242	0.0153	0.0235	0.0103	0.0362	0.144	0.1785	0.1772	0.18	-0.5295987	0.3378933	0.8901333	1.4115	0.3378933	10/1/2018	2
0.0154	0.2185	0.1236	0.0326	0.0718	0.0358	0.1225	0.4366	0.5286	0.5423	0.5451	-0.3866377	0.2774043	0.876146	2.0126	0.2774043	10/3/2018	2
0.0226	0.1866	0.1254	0.0338	0.0604	0.0396	0.1033	0.3372	0.4225	0.4477	0.4428	-0.3607461	0.1961538	0.8374718	1.7836	0.1961538	10/16/2018	2
0.0142	0.1756	0.0964	0.0185	0.0483	0.026	0.0998	0.386	0.4785	0.4832	0.4986	-0.4267027	0.2911765	0.897879	1.9838	0.2911765	10/21/2018	2

Figure 55. Reflection values per Sentinel 2 band, together with the computed indices and the image data. Sampling point 2.

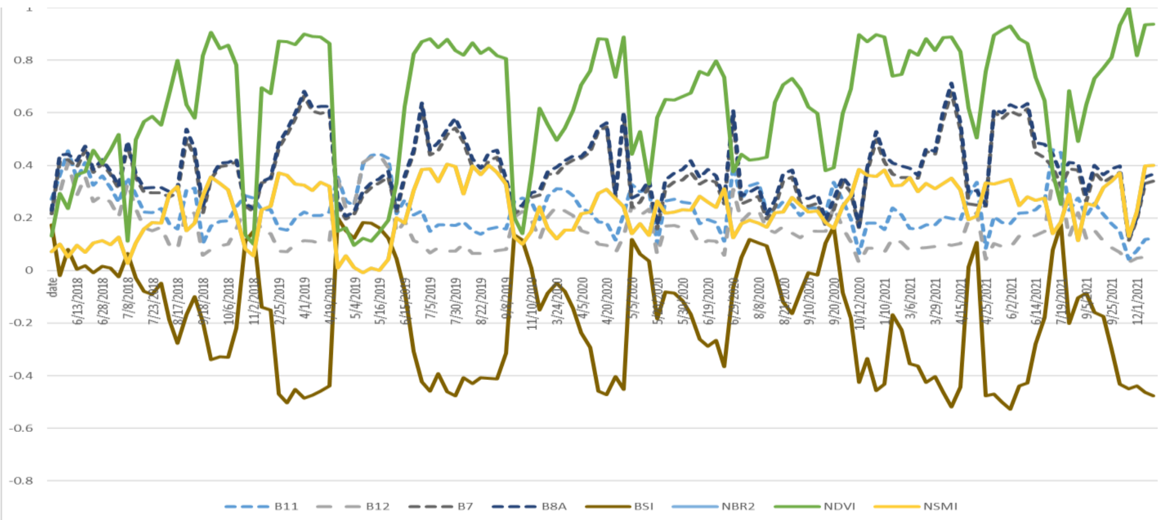


Figure 56. Visualization of reflection bands and indices for the sampling point 33. In total, we have 131 reflection signatures for the period of May- 2018 until the end of 2021. Only 13 reflection signatures correspond to bare soil (10%).

Within COLAB and PYCARET, we perform data preparation, model training, hyperparameter tuning, analysis and interpretability, and model selection.

We tested regression models within the first iteration of our product developments. As presented in Table 14, we create training, test and validation sets within the first iteration that use the median value

[3] Reducers are the way to aggregate data over time, space, bands, arrays and other data structures in Earth Engine. The ee.Reducer class specifies how data is aggregated. The reducers in this class can specify a simple statistic to use for the aggregation (e.g. minimum, maximum, mean, median, standard deviation, etc.), or a more complex summary of the input data (e.g. histogram, linear regression, list)

[4] You can export images, map tiles, tables and video from Earth Engine. The exports can be sent to your Google Drive account, to Google Cloud Storage or to a new Earth Engine asset.

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

per band per sampling point. We follow a strategy of 80%, 20%, 10% and random sampling, which means 10% of the sampling points consist of the unseen data set, and from the 90% of the seen data set, the 80% consist of the training set and 20% the test set. As presented in Figure 58, we compare modelling results from a set of ML algorithms, and we proceed further to model tuning for algorithms that deliver the best results.

Setup() function automatically pre-processing and sampling in the background. It operates on default parameters, but these parameters can be changed according to one's requirement. For more, see https://towardsdatascience.com/machine-learning-made-easier-with-pycaret-907e7124efe6

	Data Type
B1	Numeric
B2	Numeric
B3	Numeric
B4	Numeric
B5	Numeric
B6	Numeric
B7	Numeric
B8	Numeric
B8A	Numeric
B11	Numeric
B12	Numeric
OC	Label

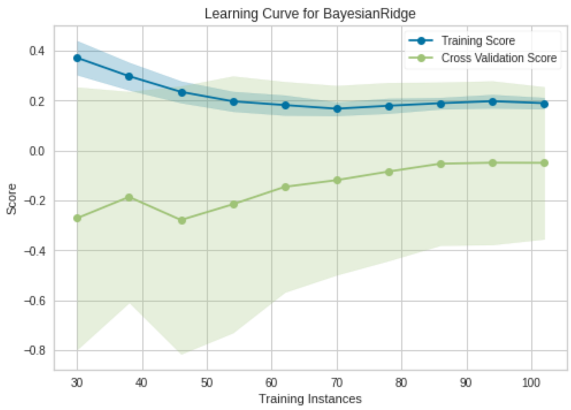
*from pycaret.regression import **
envision =
setup (data = data, target = 'OC',
normalize= True,
transform_target = True,
feature_selection = True,
remove_outliers = True,
combine_rare_levels = True,
rare_level_threshold = 0.05,
remove_multicollinearity = True,
multicollinearity_threshold = 0.95)

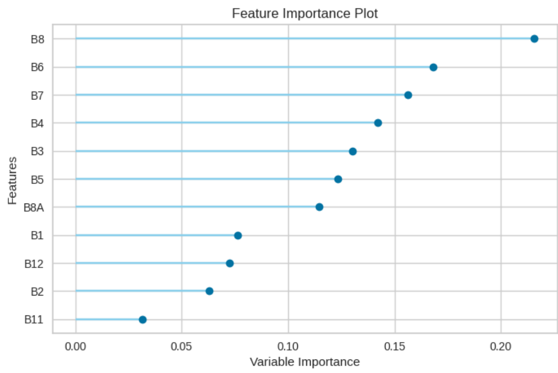
Figure 57. There are around 50 parameters that are to be fed into the setup() function.

lr	Linear Regression	sklearn.linear_model._base.LinearRegression
lasso	Lasso Regression	sklearn.linear_model._coordinate_descent.Lasso
ridge	Ridge Regression	sklearn.linear_model._ridge.Ridge
en	Elastic Net	sklearn.linear_model._coordinate_descent.Elast...
lar	Least Angle Regression	sklearn.linear_model._least_angle.Lars
llar	Lasso Least Angle Regression	sklearn.linear_model._least_angle.LassoLars
omp	Orthogonal Matching Pursuit	sklearn.linear_model._omp.OrthogonalMatchingPu...
br	Bayesian Ridge	sklearn.linear_model._bayes.BayesianRidge
ard	Automatic Relevance Determination	sklearn.linear_model._bayes.ARDRegression
par	Passive Aggressive Regressor	sklearn.linear_model._passive_aggressive.Passi...
ransac	Random Sample Consensus	sklearn.linear_model._ransac.RANSACRegressor
tr	TheilSen Regressor	sklearn.linear_model._theil_sen.TheilSenRegressor
huber	Huber Regressor	sklearn.linear_model._huber.HuberRegressor
kr	Kernel Ridge	sklearn.kernel_ridge.KernelRidge
svm	Support Vector Regression	sklearn.svm._classes.SVR
knn	K Neighbors Regressor	sklearn.neighbors._regression.KNeighborsRegressor
dt	Decision Tree Regressor	sklearn.tree._classes.DecisionTreeRegressor
rf	Random Forest Regressor	sklearn.ensemble._forest.RandomForestRegressor
et	Extra Trees Regressor	sklearn.ensemble._forest.ExtraTreesRegressor
ada	AdaBoost Regressor	sklearn.ensemble._weight_boosting.AdaBoostRegr...
gbr	Gradient Boosting Regressor	sklearn.ensemble._gb.GradientBoostingRegressor
mlp	MLP Regressor	sklearn.neural_network._multilayer_perceptron....
lightgbm	Light Gradient Boosting Machine	lightgbm.sklearn.LGBMRegressor
dummy	Dummy Regressor	sklearn.dummy.DummyRegressor

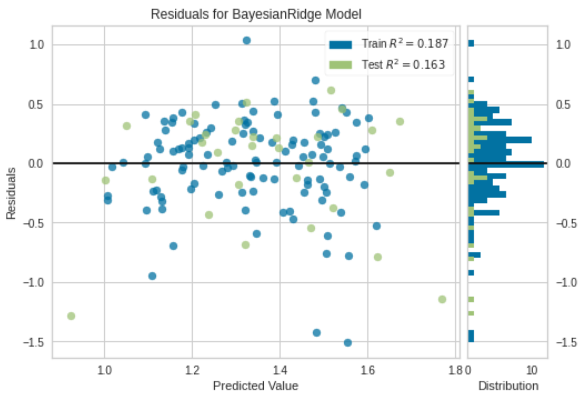
The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366

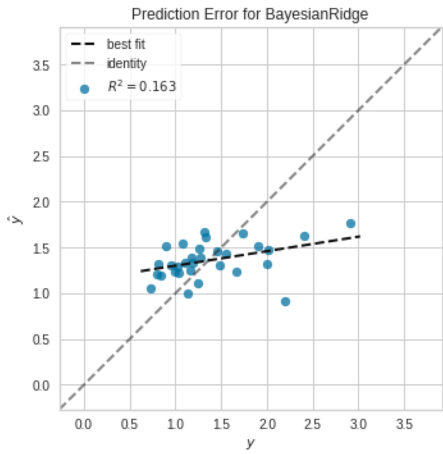
Figure 58. List of available models





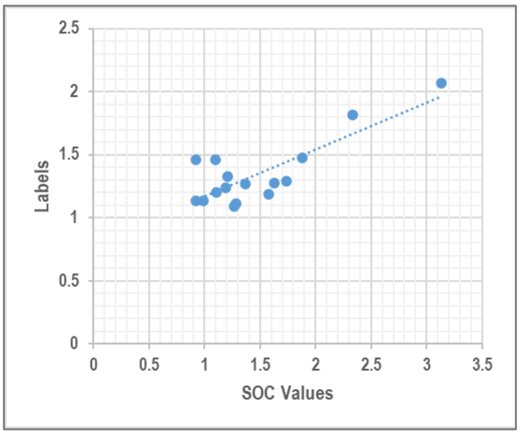
Learning curve *Feature Importance.*





Residuals *Prediction errors*

Figure 59. Evaluation graphs for Bayesian Ridge Model



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

Figure 60. RF model has R2 of 0.26 on the validation set

	Model	MAE	MSE	RMSE
ridge	Ridge Regression	0.2918	0.1539	0.3664
br	Bayesian Ridge	0.2944	0.1601	0.3703
huber	Huber Regressor	0.2954	0.1812	0.3809
omp	Orthogonal Matching Pursuit	0.2946	0.1559	0.3723
lr	Linear Regression	0.3178	0.1907	0.4012
ada	AdaBoost Regressor	0.3164	0.1843	0.4045
rf	Random Forest Regressor	0.3231	0.1875	0.4074
en	Elastic Net	0.3305	0.1930	0.4154
lasso	Lasso Regression	0.3305	0.1930	0.4154
llar	Lasso Least Angle Regression	0.3305	0.1930	0.4154
dummy	Dummy Regressor	0.3305	0.1930	0.4154
knn	K Neighbors Regressor	0.3328	0.1897	0.4175
et	Extra Trees Regressor	0.3454	0.2075	0.4361
lightgbm	Light Gradient Boosting Machine	0.3544	0.2064	0.4382
gbr	Gradient Boosting Regressor	0.3498	0.2299	0.4602
dt	Decision Tree Regressor	0.4153	0.2904	0.5244
par	Passive Aggressive Regressor	0.5957	0.6242	0.6771
lar	Least Angle Regression	0.6943	1.4847	0.8618

Figure 61. Comparing results for a set of models

Models	Seen data set
Bayesian Ridge	MAE: 0.2910 MSE: 0.1548
RF	MAE: 0.3230 MSE: 0.1822
SVM	MAE: 0.3014 MSE: 0.1669
MLP Regressor	MAE: 0.27 MSE: 0.13
MAE: Mean absolute error *MSE: mean squared error*	

Table 14. The models' results use the median values per band for the Total Period May 18 – May 21.

During the first iteration, all the models face difficulties handling SOC's mappings higher than 2.5. That is related to the fact that most of the collected sampling points (171 samples) take SOC values between 0.8 - 1.8 (%/dry soil). Only a few samples (6) take values above 2.5%, which means we don't have a representative train sample to support a mapping within this region. Treating values higher than 2.5% as outliers is an option; however, that means models are not suitable for deployment on regions with different SOC values.

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

3.3.1.1.3 Analytical Methods – EO data products

The deployment of machine learning models is the process of making models available in production where web applications, enterprise software, and APIs can consume the trained model by providing new data points and generating predictions. Normally machine learning models are built so that they can be used to predict an outcome (binary value i.e. 1 or 0 for Classification, continuous values for Regression, labels for Clustering, etc. There are two broad ways of generating predictions (i) predict by batch; and (ii) predict in real-time. In our case we predict by patch and we deploy the model to a synthetic layer of cloudless bare soil collection with median values per band and per pixel. We apply the selected deployed model to all pixels belonging to larger areas, at regional (Envision, BC3) or even national scale to develop the SOC products.

var syn_layer_median = **cloudless_bare_soil_collection**.median()

Developing a synthetic layer is a common practice in many top SOC mapping activities. It delivers benefits related to the simplicity of the process and the fact that it returns the max available coverage (see Figure 63). However, analysis results (see D.3.3 section 4.4.2.1.1 Phase One: Bare Soil Identification) have shown that the long term median reflection values are not always representative.

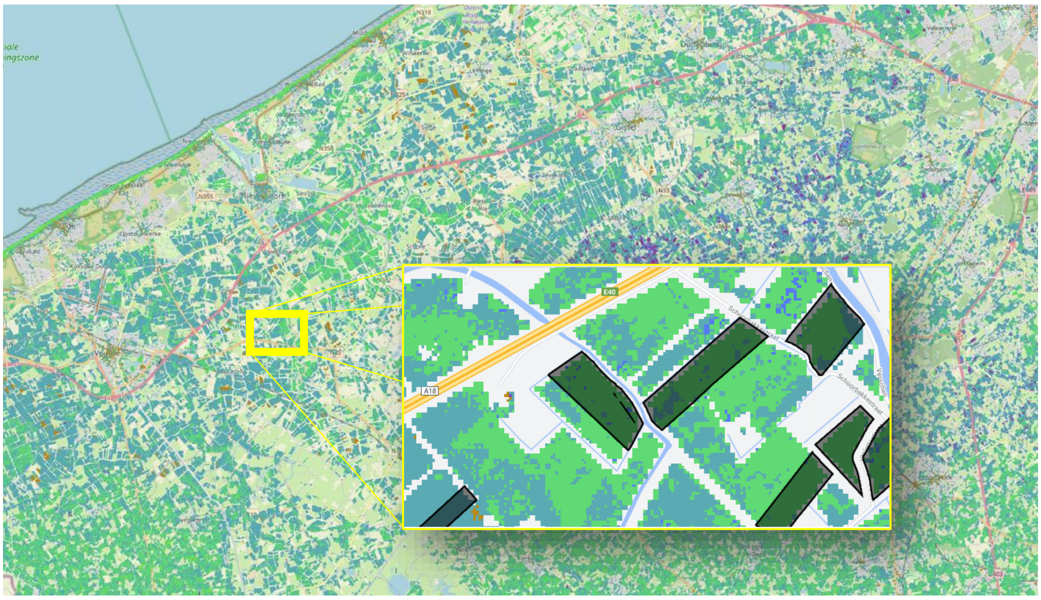
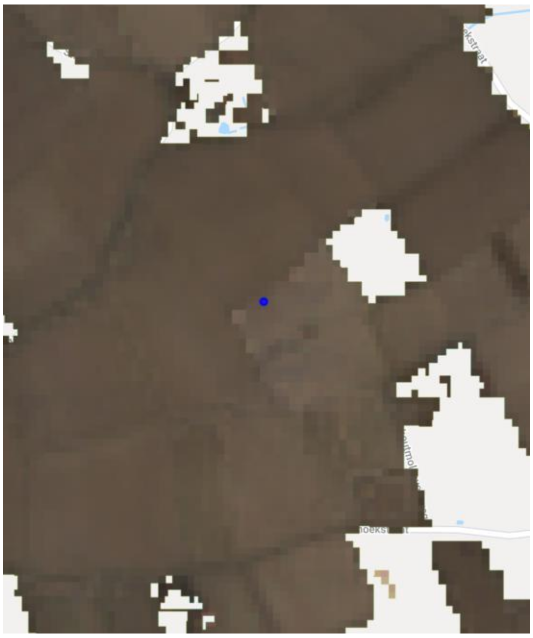


Figure 62. SOC map covering West Flanders. Zoom window overlays a sample of agricultural parcels.

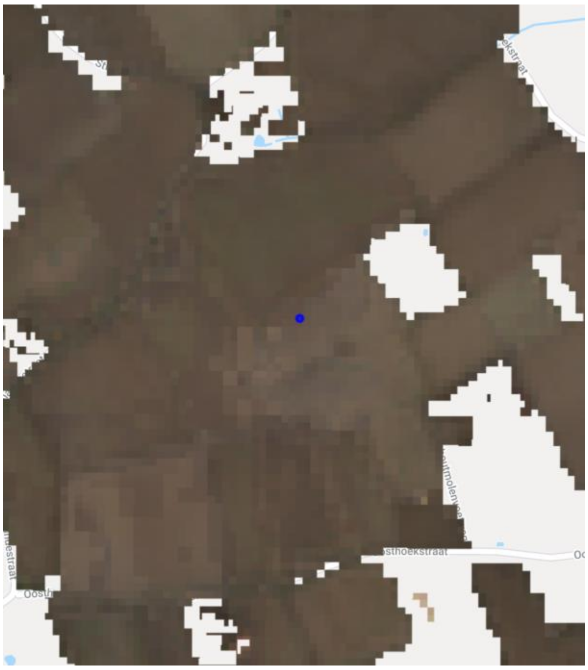
In *Figure 62. SOC map covering West Flanders. Zoom window overlays a sample of agricultural parcels.*

, results are presented in classes (Low, Medium, High) using crisp limits to support decision-making. However, the RMSE of all models receives values close to 0.5, which means the classification results face significant accuracy issues when the labelling receives values close to the crisp limits. Fuzzy logic and the uncertainty transfer to the membership functions can be an option, especially if we manage further to reduce the RMSE to values close to 0.25.

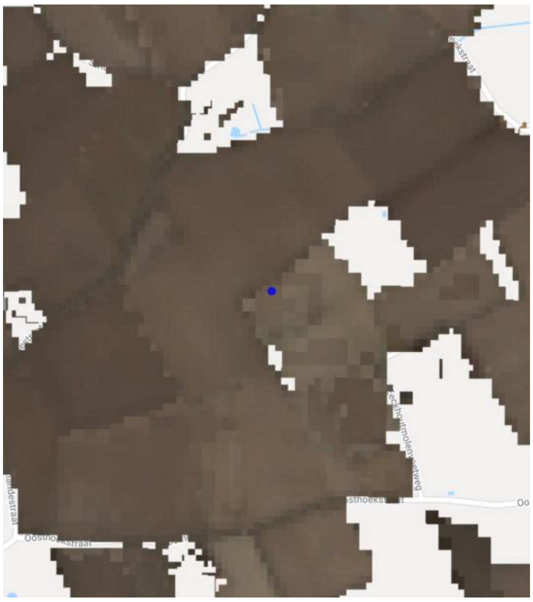
The ENVISION project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869366



RGB visualization of the cloudless bare soil collection for May-2018 until May-2021 using the median values per band.



RGB visualization of a cloudless bare soil collection from May-2018 until May-2019.



RGB visualization of a cloudless bare soil collection from May-2019 until May-2020.

RGB visualization of a cloudless bare soil collection from May-2020 until May-2021. The cloudless bare soil collection mainly does not cover the sampling point area due to clouds.

Figure 63. RGB visualization of continuous-time period stacks of the cloudless bare soil collection area around a soil sampling collection point (point ID 33).

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

3.3.2 Validation results

To perform the validation, we apply the models to a validation set consisting of 10% of the sampling points (unseen). In Table 15. Validation results we provide the validation results for each model

Models	Unseen
Bayesian Ridge	MAE: 0.26 MSE: 0.09
RF	MAE: 0.32 MSE: 0.16
SVM	MAE: 0.32 MSE: 0.25
MLP Regressor	MAE: 0.26 MSE: 0.09
MAE: Mean absolute error *MSE: mean squared error*	

Table 15. Validation results

SOC	Label	Error	Abs Error
0.99	1.188	-0.198	0.198
0.92	1.308	-0.388	0.388
1.28	1.238	0.042	0.042
1.63	1.392	0.238	0.238
1.11	1.144	-0.034	0.034
1.74	1.142	0.598	0.598
1.19	1.058	0.132	0.132
1.88	1.356	0.524	0.524
0.92	1.136	-0.216	0.216
1.21	1.392	-0.182	0.182
1.37	1.258	0.112	0.112
1.58	1.258	0.322	0.322
1.1	1.382	-0.282	0.282
3.13	1.694	1.436	1.436
2.33	1.432	0.898	0.898
1.27	1.066	0.204	0.204
Mean SOC	**Sum of Label**	**Sum of Error**	**MAE**
1.478125	20.444	3.206	0.362875

Table 16. Validation results with errors and calculated MAE, for the RF model.

As presented in Table 13,Table 16. Validation results with errors and calculated MAE, for the RF model. and similar to other models, it rarely fits SOC higher than 2.2. That is a clear message, pointing to where we need to focus our improvement efforts.

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

3.3.3 Limitations

This section will summarize the identified limitations of the first iteration, making references to the previous sections and in Deliverable 3.3 Data products initial report. We will try to classify them at the different phases.

- Phase One: The most crucial product is the cloudless bare soil collection. We use vegetation and moisture indices; however, uncertainty exists on selecting the indices (which indices is suitable? Which combination of indices?) and the definition of the threshold limits (which range?). Even if many scientific papers suggest the optimal combinations of indices and thresholds, it remains a trial–error approach, which is not yet standardized by the Soil community. Quality controls are needed to ensure that the assessed pixels cover bare soil areas on a specific day. Smart and low budget ways are needed to ensure and verify the quality, maintaining the need for in situ data collection at a low level.
- Phase Two and Three: We link the pixel reflections with the measured top Soil Organic carbon as explained in phase two. That means: we need to apply a specific soil sampling collection protocol to perform this link.
    - We reduce the available data sets because it's impossible to use existing collection samples and measurements that do not follow the same protocol.
    - We need to manage the uncertainty, errors etc., coming from the lab measurements, and we need to find ways to transfer that uncertainty to the modelling results.
    - We need to invest resources and increase the cost of new soil campaigns covering the agricultural areas.
    - We need to perform the soil campaigns on periods without cloud coverage.
    - We need to select soil samples on areas covered by bare soil and not vegetation.
    - We need to apply the model on bare soil reflections from different periods than the soil sampling collection. That means different lighting soil moisture, and vegetation conditions.
    - We need to consider how the farm management practices affect the top Soil Organic Carbon but mainly bare soil, moisture levels, etc.
    - We need to consider and deal with the different spatial resolutions of S2 bands and evaluate the influence of adjustment pixels on the pixels with direct spatial links with the soil sampling locations.
    - Soil campaign needs to provide results that generate a representative train and validation set, covering all expected SOC values.

  Even if there is a relation between S2 reflections and topsoil organic carbon, the mentioned points influence the mapping abilities.
- Phase 4: Due to results with lower than the expected accuracy until now, our efforts were focusing mainly on how to improve the model, how to understand better the available data sets, how to improve our methodological steps and select the right technological tools to increase the ability for more iterations, more testing on different scenarios. That limited the time spent on further developing the service's business logic. Even if we have collected the end users' requirements, we have gained more experience in large-scale SOC modelling now and after the first iteration. That experience can support us design better ways to deliver the service and support the SOC monitoring for the CAP needs.

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

3.3.4 Next steps

The following steps aim to improve the modelling results and will be finalized within Phase Five.

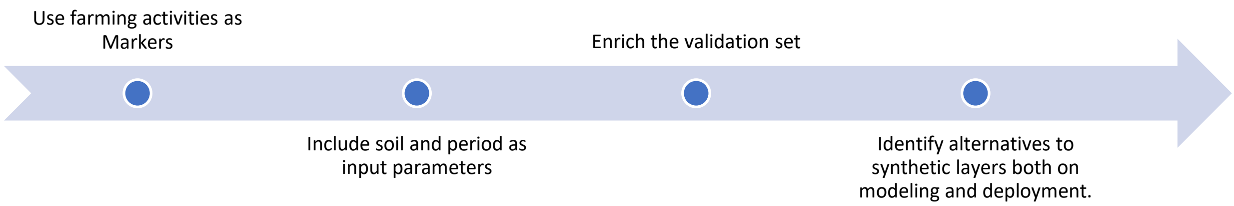


Figure 64. Second iteration - Major improvements

More specific:

- For Phase One: We have already improved the ability to analyse the reflection values for each sampling point and link it with SOC measurements of the same pixel area. We also have tested after the first iteration new indices like the BSI. What is needed is to identify in a better reflection record (which S2 passing day) links with the measure SOC values. The next step is to use data describing farming activities as markers in this process. We will collaborate with LV on this. The use of the median values has limitations, and we are trying to apply alternatives.
- Phase Two: Using more advanced algorithms is an option; however, we need first to generate data sets that describe the mapping better. So we will also use the soil association as input parameters and the period of the collection, for example, the month (Figure 65).
- Phase Three: Decisions on phase One and Two effects significantly affect how we deploy the models. It's impossible to deploy the model on the synthetic layer consisting of the median reflecting values. We need to identify different approaches that are compatible with the structure of the model but also deliver the needed coverage and transfer the accuracy of the model to the SOC products. That is the challenge we need to deal with, and we need to deliver, for example, answers to the following questions:
    - How is the accuracy of the model inherent to the SOC products?
    - Is it possible to apply a model that relies only on reflections from the sampling collecting period (Q1) to a collection consisting of reflections from a non-similar period (Q1 – Q4) as a way to increase the coverage?
    - Is it enough to deliver an accurate SOC map that covers only 20% of the agricultural area uniformly?
    - We can see that we significantly populate the training and validation data set; however, we need to validate this approach. Can we identify reflections signatures by applying functions (for example, min) to specific indices (NDVI)? Is it possible to link SOC results to many reflection signatures from the same pixel area? Or do we need to select reflection signatures that satisfy specific criteria, for example, to exist in the same period (Quarters or months)?
- We need to increase the validation set further. To achieve this, we plan to use the data set of the Flemish soil monitoring framework. The data are expected to be available, and at ENVISION, we have foreseen to follow the same collection protocol.

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

Interactions

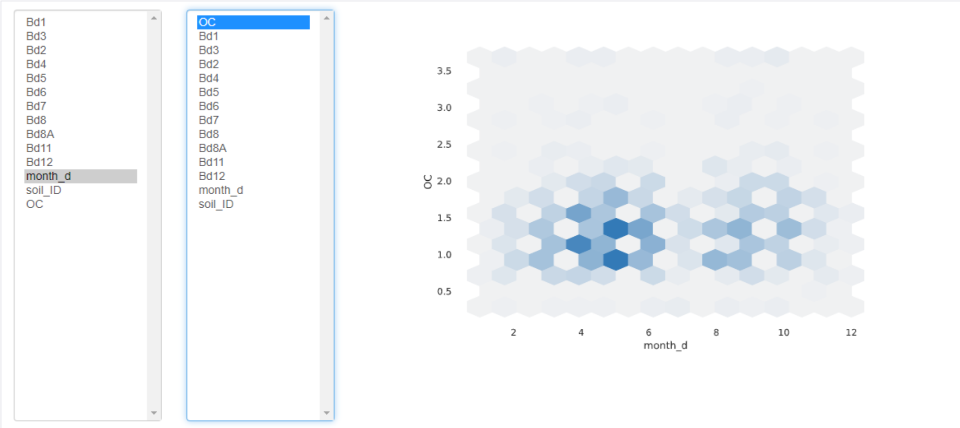


Figure 65. Visualization of the interaction between the collection period and the SOC values.

3.4 BC4: Monitoring of organic farming requirements – Serbia

3.4.1 Pilot and samples description

The methodological framework for the Organic Crop Practice Identification product for the Serbian business case is described in detail in D3.3. and is composed of the model training methods and model deployment methods in the traffic light system, at the operational mode of the service. The validation of the product consisted of the successive preliminary steps of Vegetation Feature extraction from Sentinel 2 images and Ground truth data sampling of the EO derived products, which resulted to the creation of the training -validation dataset, and the resultant application of a machine learning framework for the creation of crop specific models. The service addresses to 2 Certification Bodies of Serbia and thus the pilot area support for the Lighthouse customers of the project is the whole Serbia country region. However, the product validation was conducted a small data subset of the Serbia LPIS, provided along with the Organic/Conventional farming practice declaration, the Certification bodies. The spatial data include parcels from 4 crops of interest, namely Maize, Soybean, Sunflower and Wheat, and a general preview of the geographic distribution of the data is presented on the following map.

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

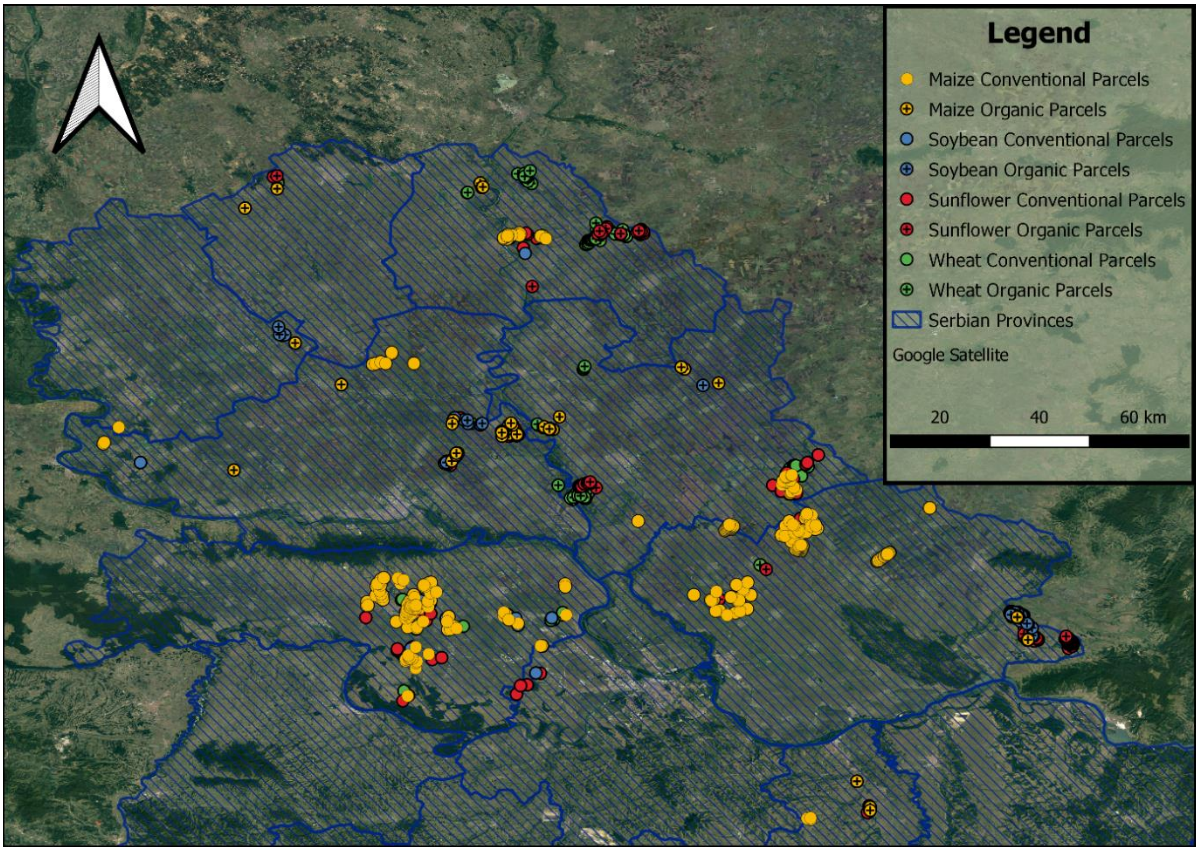


Figure 66. Geographic distribution of organic/conventional parcels for the 4 crops of interest

Data collection for validation purposes included crop declarations from the years 2016-2021, and the count statistics for each crop is presented on the following tables, as well as a summary table for total dataset area available for training/ validation.

Table 17. Count Stats of parcels for Wheat Organic and Wheat Conventional

Wheat	Organic	Organic Dissolved	Conventional	Conventional Dissolved	Total	Total Dissolved
2016	87	35	2	2	89	37
2017	65	25	4	4	69	29
2018	198	41	187	109	385	150
2019	77	20	213	128	290	148
2020	217	68	219	137	436	205
2021	132	31	28	15	160	46
Total	**776**	**220**	**653**	**395**	**1429**	**615**
Mean	**3.51**	**13.76**	**1.19**	**1.97**		
St Dev	9.07	18.86	2.11	2.88		
Min	0.1	0.01	0.07	0.15		
Max	84.36	90.5	32.74	32.74		

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

Table 18: Count Stats of parcels for Maize Organic and Maize Conventional

Maize	Organic	Organic Dissolved	Conventional	Conventional Dissolved	Total	Total Dissolved
2016	79	22	4	4	83	26
2017	15	10	10	9	25	19
2018	8	8	242	118	250	126
2019	35	10	355	164	390	174
2020	8	6	365	188	373	194
2021	27	17	77	34	104	51
Total	**172**	**73**	**1053**	**517**	**1225**	**590**
Mean	**4.72**	**12.54**	**1.52**	**3.05**		
St Dev	2.67	19.91	3.67	5.36		
Min	0.5	0.04	0.01	0.07		
Max	21	80.41	57.51	57.51		

Table 19: Count Stats of parcels for Soybean Organic and Soybean Conventional

Soybean	Organic	Organic Dissolved	Conventional	Conventional Dissolved	Total	Total Dissolved
2016	89	19	6	6	95	25
2017	71	20	7	7	78	27
2018	13	10	51	33	64	43
2019	17	9	78	38	95	47
2020	5	5	68	40	73	45
2021	18	6	6	6	24	12
Total	**213**	**69**	**216**	**130**	**429**	**199**
Mean	**2.89**	**15.86**	**3.49**	**5.00**		
St Dev	1.08	21.40	0.53	12.11		
Min	0.09	0.09	0.3	0.26		
Max	10	88.6	4.9	118.61		

Table 20: Count Stats of parcels for Sunflower Organic and Sunflower Conventional

Sunflower	Organic	Organic Dissolved	Conventional	Conventional Dissolved	Total	Total Dissolved
2016	288	68	4	3	292	71
2017	89	21	4	4	93	25
2018	58	17	88	51	146	68
2019	96	34	130	78	226	112
2020	101	19	173	110	274	129
2021	11	9	13	12	24	21
Total	**643**	**168**	**412**	**258**	**1055**	**426**
Mean	**2.08**	**12.42**	3.37	**2.24**		
St Dev	0.96	17.90	0.60	3.02		
Min	0.01	0.01	2.2	0.26		
Max	6.2	135.95	9.5	29.46		

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

Table 21: Total Area Stats of parcels for all crops, Organic and Conventional

	Total Area (ha)	
	Conventional	Organic
maize	1244.36	159.91
soybean	492.82	544.34
sunflower	501.22	936.91
wheat	692.93	2058.98

The main particularities faced on the validation of the Organic Crop Practice Identification with EO data, service product, were:

- The very small and elongated parcel geometries, found on the Serbian LPIS. This local spatial characteristic, keeping in mind an additional boundary buffer clipping, limits significantly the data sampling availability.
- The cloudy weather conditions. Clouds and shadows are masked from the NDVI images, which results on many gaps on the timeseries. This particularity is partially handled by temporal interpolation pre – processing of the NDVI image stack.
- The unbalanced dataset between organic and conventional parcels. As it can be noticed from the tables above, some crops have quite unbalanced distribution. Trying to prevent biasing the ML models, a lot of data pixels remained finally unused in the training/validation process.

3.4.2 Validation results

The steps that followed for the training of ML models for organic practice identification is presented on the following flowchart.

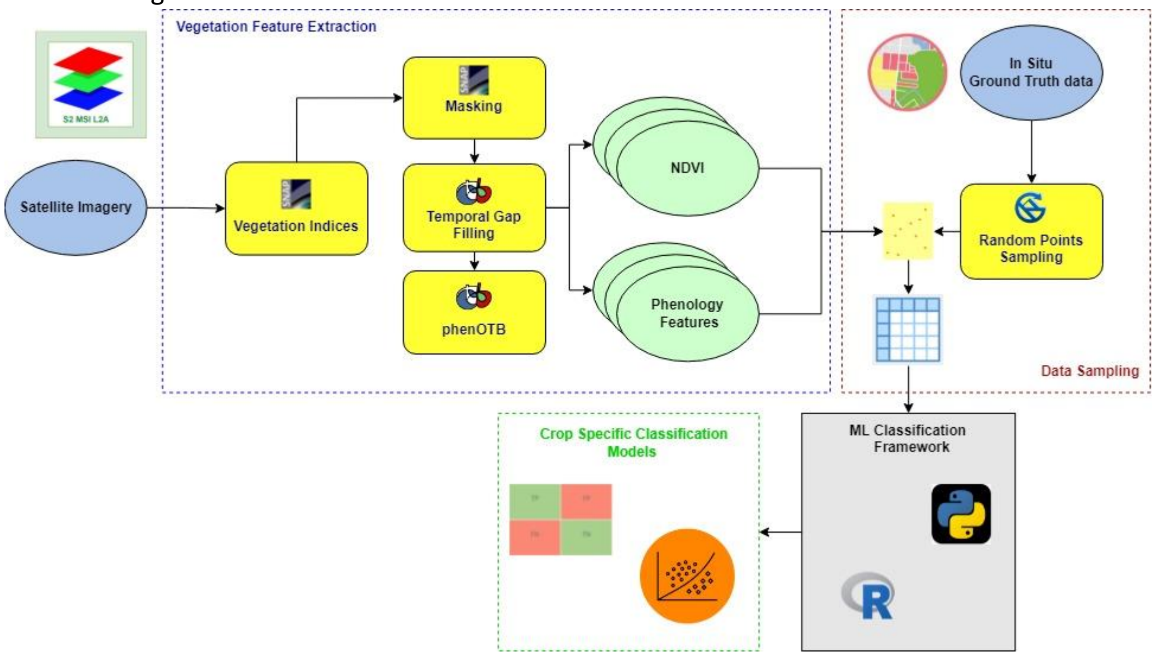


Figure 66: Methodological framework for the training of ML models for organic practice identification

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

Regarding the Vegetation Feature Extraction from Sentinel 2 images, a stick masking approach was followed, allowing only vegetation and barren land pixels to pass, and filtering out the effects of cloudy atmospheric conditions. This resulted on the rejection of 48% of the available data, and as a counter action for the recovery of the NDVI profiles, an interpolation method was applied. The following graph shows the results of gap – filling techniques to the restoration of the timeseries, that are considered reasonably satisfactory. Despite the fact that the decision rule for the masking was strict, not allowing ambiguous land cover pixels (dark objects, unclassified), the spline interpolation provided better results that it would if the abovementioned classes were considered valid.

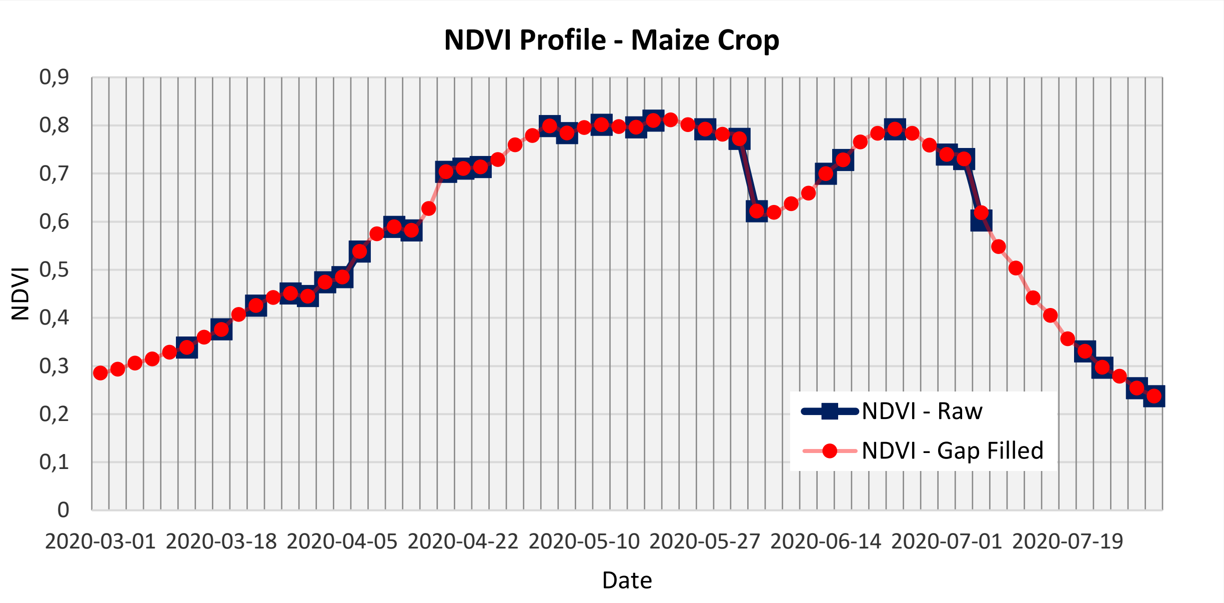


Figure 67: Gap Filling of NDVI profiles on a Maize crop parcel

Phenology feature extraction was approached through the calculation of metrics such as the Crop Growth slope, the Length of the NDVI plateau and Senescence slope, with an assumption that these features could act as discriminant information in order to identify organic farming. On the following graphs, two typical cases of conventional and organic maize crop NDVI profiles are presented along with the values of the calculated phenology metrics. Indeed, an indication exists that the phenology assumption actually holds true and that the extracted features could be of significant importance in the classification process. The NDVI plateau length seems to be larger on conventional crops, while the emergence/senescence slopes seem to be steeper.

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

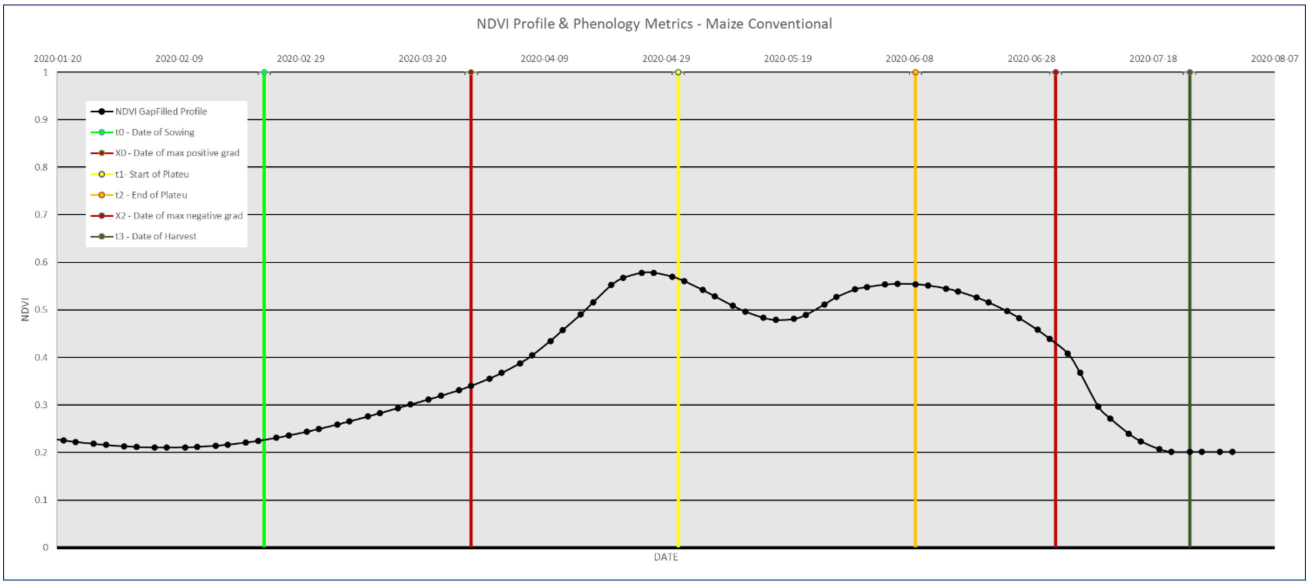


Figure 68: NDVI profile and Phenology metrics for Conventional Maize Crop

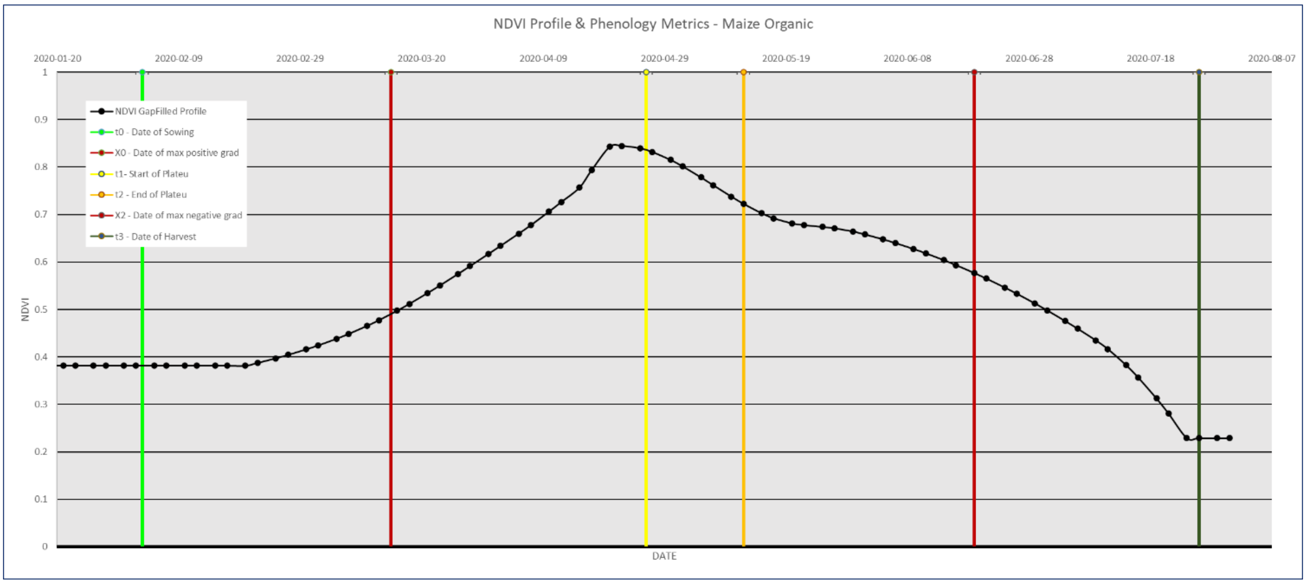


Figure 69: NDVI profile and Phenology metrics for Organic Maize Crop

The classification task regarded the binary discrimination of organic/conventional farming practice on 4 main crops of interest. For winter crops, emerging from mid-October to early August, the models regard wheat, and sunflower, while for the summer crops, emerging from April to September, they predict for maize and soybean. Modeling considers EO timeseries profiles from the whole growing cycles, therefore prediction occurs at the end of the growing season, as defined by the harvesting date of the crop.

The main properties of the Machine Learning framework that was followed for the validation of the product is summarized on the following table. As a result, 4 crop specific models were trained.

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

Table 22: Main ML methodological framework

ML Framework	
Data Sampling	Pixel based - Random Points
Outlier Analysis	PCA - Residuals & Influence
Classifier	C-SVM
Preprocess	Center & Scale at [-1,1]
Kernel	Radial Basis Function
Hyperparameters Grid Search	Cost (C)
	Epsilon (ε)
	Gamma (γ)
Model Validation Method	Nested Resampling
Model Evaluation Metrics	Overall Accuracy
	Precision
	Recall
	F1 Score
	Kappa

The following table summarizes the classification results, evaluated through the nested cross validation resampling. Overall prediction accuracies showcase relatively medium performance remaining in the range of 65 – 73,4 %. Especially the Maize model has a value of Khat of 59% which is low performance, obviously related to the small sample support for the organic class. Some derived conclusions from this first iteration of model training concern issues such sample data support, or the need of further introduction of more image features as predictor variables, and they are referenced on the Limitations and Next Steps subchapter that follows.

Table 23: Classification Results from Crop Specific models evaluation

		F1 Score	**Overall Accuracy**	**Kappa**
Winter Crops	**Wheat**	0.72	0.73	0.67
	Sunflower	0.70	0.71	0.65
Summer Crops	**Maize**	0.64	0.65	0.59
	Soybean	0.65	0.67	0.61

3.4.3 Limitations and Next steps

In this deliverable, a preliminary version of the Organic Crop Practice Identification product service was provided, bringing on a first iteration of results that showcase medium prediction performances, as concluded from the models' evaluation. The limitations present on this BC are related to

- 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 cultivations in order to comply with local subsidy

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

regulations. The case of instances where several cultivations in the same parcel 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.
- The issue of Spectro-temporal separability is considered key, on the task of classifying organic from conventional farming practice. NDVI profiles of these 2 classes seem to be overlapping on the created feature space.
- The phenology features that were incorporated in the models seem to help discrimination in a significant manner, but we simply may need to add more image features in order to enhance the classifier.

The next steps over which, model training and product validation will focus on future iterations, will consider the following actions and improvements:

- Sample dataset enhancement through the request from the Serbia CBs of more validation data that concern the same spatiotemporal extent (2016-2021).
- Incorporation of EO Image Texture features in the classification predictors. An assumption has been made that Organic vs Conventional farming practice may imprint significant spatial patterns and context of NDVI values regarding the homogeneity of radiometric values across different spatial lags. We believe that the assimilation of GLCM image texture features, such as Contrast, Entropy and Correlation, derived from the NDVI layers, will improve the classification results.
- The further use of more phenology features is as well considered for the improvement of the classifiers' performance. Specifically, the extraction of NDVI 1st and 2nd Derivative layers with the use of Savitzky- Golay moving window filtering algorithm, will accentuate the rates of change throughout the profile, and may help on the classifier improvement, in a significant manner.
- The use of H.H Resolution imagery will also be considered, with the suggestion of the Planetscope product, as a solution in order to deal with the small /elongated parcel geometries and validate the existence of more complex structures inside the parcels.

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

4 Conclusions

This deliverable is the first version of the EO data product validation report, providing the initial results that derived from the services development/ training based on historical data that were provided from the business cases' users. Further development/ training/ analysis will be performed and the final and updated results of the data product validation process will be presented in the D3.6 Data product validation report (final version) at M34.

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

End of Document

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