NutriBudget

Optimisation of nutrient budget in agriculture

D3.3 Report and github-repo with algoritms for derivation of KPI from sensing data

Funded by the European Union

Cover Delivery Report

Project Information	
Acronym	NutriBudget
Title	Optimisation of nutrient budget in agriculture
Project no.	101060455
Type of Action	RIA
Website	*https://www.nutribudget.eu/*
Deliverable Information	
Title	Report and github-repo with algorithms for derivation of KPI from sensing data
WP number and title	WP3 – Developing the NutriKPI monitoring system
Lead Beneficiary	NMI
Authors	Kees van den Dool (NMI), Sven Verweij (NMI), Wim Bussink (NMI), Francesca Degan (Arvalis), and Gerard H. Ros (WU)
Reviewers	Ylivainio, K, Degan, F
Description	This report describes the development of an IT-service to monitor the growth of catch crops from space
Type	R – Document, report
Dissemination Level	PU - Public
Status	draft
Submission due date	31-01-2025
History of Changes	
Version 0.1	Draft created by NMI (17-12-2024)
Version 1.0	Final version by NMI after revisions of ARVALIS and LUKE sent to UGent (03-02-2025)
Version 1.0	Final version approved by UGent and uploaded (04-02-2025)

Preface

The NutriBudget project aims to develop and implement a prototype of an integrated nutrient management platform (NutriPlatform) in various regions across Europe, as a decision support tool (DST) for farmers, advisors, European policy makers and regional authorities. The development of the NutriPlatform will be based on knowledge from existing and new field-tested agronomic mitigation measures linked to advanced NutriModels, which integrate various nutrient models, common data standards and relevant monitoring indicators. Thereby, Nutribudget will contribute to systemically optimizing nutrient flows and budgets across different agricultural production systems and regions in the EU to reduce nutrient losses to the environment and related impacts. The NutriModels will be able to operate at different scales: for farmers at the farm level and for regional authorities and policy makers at the regional to EU level, taking into account a holistic, sustainable and data-driven perspective on agriculture, linking the flow of nutrients between soil, water, air, plants, animals, feed and food with specific validated technological or nature-based mitigation measures within a financially viable transition route towards the desired nutrient status, as described in the Zero Pollution Action Plan and the Farm to Fork Strategy.

To assess the actual farm performance in view of agronomic and environmental targets, an integrative key performance indicator (KPI) framework has been designed to monitor the transition from the current to the desired status to have optimised farming systems (conventional, agro-ecological and organic in animal and crop production) in equilibrium with maximum agricultural performance and minimal environmental pressure. As such, this framework will guide the actual decision support as well the identification of appropriate roadmaps to reach the desired status for soil surpluses of carbon and nutrients in view of targets for soil quality, water quality, climate, biodiversity and crop production.

Assessing and monitoring the sustainability of farms and farming systems requires farm specific data regarding soil properties, and soil, crop and nutrient management practices. One of the main properties affecting the impact of farming systems on soil fertility, biodiversity, water quality, and climate is the presence of cover crops during the winter period. This study evaluates whether the presence of cover crops can be detected from space, allowing a tailor-made evaluation of farming systems and minimizing farmers input in the decision support tool. This activity is part of Task 3.3, focused on the development of simple IT-services supporting the temporal and spatial monitoring of farm performance for decision support on farm level. The developed model and supporting algorithms will be made available for use in the actual DST development in WP5.

We are very grateful to all partners who contributed to the development of the NutriKPI framework: Ghent University (Belgium), Luke (Finland), Yara International (Norway), PWC (France), Arvalis (France), Beta Technology Center (Spain), Wageningen University & Research (the Netherlands), the Rural Investment Support for Europe Foundation (RICE), the Universita Degli Studi di Milano (Italy), Proman Management (Austria), Sveriges Landbruksunversitet (Sweden), Nutrient Management Institute (the Netherlands), Acqua & Sole (Italy), Impact (Belgium), Stockholms Universitet (Sweden) and the Forschungsinstitut fur Biologischen Landbau Stiftung (Switzerland). Lastly we thank Francesca Degan and Ylivainio Kari for reviewing this report.

Executive Summary

This report (deliverable D3.3 of the NutriBudget project, Task 3.3) describes the development of a spatial explicit machine learning model that can predict the occurrence of cover crops over time and space across all farming systems in the EU. Given the potential benefits of cover crops to soil fertility, water quality, biodiversity and climate (known from scientific studies), this illustrates the potential of remote sensing derived indices to give farmers insights in some basic farm properties affecting its contribution to a sustainable agriculture.

After the introduction, we describe the methodology to train a machine learning model using open source satellite data and ground-truth data being collected in France. The best model was able to predict the occurrence of cover crops with a precision (proportion of positive predictions) and accuracy of about 70%. This implies that the developed model is applicable for first identification of the occurrence of cover crops. Additional guidelines are given for further improvement and upscaling.

The derivation of these remote sensing indicator (the presence of a cover crop) shows that basic crop management can be monitored from space, thereby facilitating the input generation of the NutriPlatform (making its user more friendly for the user) and providing concrete, tailor-made information on a relevant mitigation measure that can be applied to improve soil fertility, biodiversity, water quality and carbon storage for climate mitigation. The developed model is open-source and may be implemented into the IT architecture of the decision support tool (developed in WP5). It also facilitates the model input generation for the impact assessment of measures tested with the NutriModels in WP2.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ALE	Accumulated Local Effect
DST	Decision Support Tool
EC	European Commission
ERA5	ECMWF Reanalysis v5
GEE	Google Earth Engine
LIPS	Land Information Parcel Systems
NMI	Nutriënten Management Instituut
RS	Remote Sensing
RVI	Radar Vegetation Index
SAR	Synthetic Aperture Radar
S1	Sentinel 1
S2	Sentinel 2
FIP	Feature Importance Plot
VV	Sentinel 1 single co-polarization, vertical transmit/vertical receive
VH	Sentinel 1 single co-polarization, vertical transmit/horizontal receive
WP	Work Package (of the Nutribudget project)

1. Introduction

1.1. The NutriBudget project

Food systems face multiple challenges globally, including the need to produce sufficient crops of high nutritional quality while reducing the environmental impacts of this production, especially in the face of climate change and biodiversity loss (Silva and Giller, 2021). This is of particular importance due to concerns that yield increases are insufficient to meet future food demands (Zhang et al., 2021). Agronomic measures optimising the nutrient use efficiency of fertilizers and crop yield on one hand and soil health on the other are promising solutions to mitigate environmental impacts while maximising crop production. The number of studies examining impacts of agronomic measures is rapidly increasing, many examining the impact of crop diversification, cover crops, reduced tillage, and precision fertilization practices (Bolinder et al., 2020; Eagle et al., 2017; Young et al., 2021). Both the applicability and impact of these measures are highly dependent on the local agro-ecological properties of the food system. Though this dependency of measures on site properties is known from ample evidence from field experiments and modelling studies, their impact on the effectiveness of agronomic measures is often ignored, thereby hampering the guidance and actual implementation on farm and field level. One of the main objectives of the Nutribudget project is therefore to develop and implement a prototype of an integrated nutrient management platform, called NutriPlatform, in various regions across Europe, as a decision-support tool for farmers, advisors, European policy makers and regional authorities.

A wide range of relevant DSTs have recently been developed, providing decision options for policymakers and farmers (Ros et al., 2023). However, in contrast to the integrated NutriModels and NutriPlatform that will be developed in NutriBudget, most of these are only strongly focused on one topic, such as soil protection (Oleson et al., 2016; Sarangi et al., 2004), precision agriculture (Venkatalakshmi & Devi, 2013), fertigation management (Elia & Conversa, 2015) or specific nutrient measures (Hewett et al., 2004). Others have been developed for a specific geographic context (Manos, et al., 2007) or for land evaluation and spatial planning of sustainable management operations (De la Rosa et al., 2004). Very few tools evaluate agronomic measures quantitatively using impact assessment on targeted environmental as well as agronomic outcomes.

To move agricultural systems towards a more efficient and sustainable future, it is important to understand existing opportunities for operational and strategic management and to have simple and robust indicators to monitor agricultural and environmental performance. Where Ros et al. (2023) developed an integrative framework to evaluate and monitor the overall farm performance in view of agronomic and environmental objectives, there is additional need for simple agri-environmental indicators that help farmers to increase their sustainability. For agronomic, environmental, and economic reasons, the need for spatialized information about agricultural practices is therefore expected to rapidly increase. Remote sensing has been proven to be an effective tool for monitoring cropping practices (Bégué et al., 2018). Due to a large variety of on-board sensors on an increasing number of satellites, the spectral and temporal properties of the land surface resulting from crop and soil management practices can be captured and monitored at different spatial and temporal scales. The practices involved are highly variable across climatic zones and regions across Europe, and most of the scientific studies evaluating the relevance of remote sensing (RS) are limited to case studies. This shortcoming strongly depends on the wide variety and variability of agricultural practices, requiring data intensive (over time and space) monitoring programs of these practices and the associated remote sensing data. The increasing availability of remote sensing data across Europe at high spatial and temporal resolution, along with the emergence of new data processing techniques such as data mining and deep learning, suggests that remote sensing data could boost the use of cheap and reliable farm indicators and help farmers improve farm performance in view of various agronomic and environmental targets.

1.2. Identification of cover cropping

The NutriBudget decision support tool provides guidance to farmers, advisors, and policy makers on the implementation of mitigation measures to reduce the environmental impact of fertilizer application and enhance nutrient efficiency. To ensure the applicability and efficacy of the advice provided, it is necessary to collate data from a range of sources. This information encompasses data pertaining to soil characteristics, cultivation practices, fertilizer application, meteorological conditions, and farm management procedures. A significant drawback of requiring users to input a substantial amount of information is that it may result in users abandoning the application before receiving advice, thereby preventing them from utilizing the knowledge gained in the NutriBudget project. It is therefore crucial to minimise the amount of information that users have to enter to ensure that the knowledge from the NutriBudget project is successfully transferred to farmers, advisors and policy makers. This will also help to reduce the barriers to using the application.

One potential solution is to provide the user with default values, thereby eliminating the need for the user to enter the data manually while directly reflecting on the farms' contribution to various ecosystem services. The efficacy of defaults is contingent upon their accuracy; otherwise, the advice may prove ineffective or even inapplicable, potentially eroding the user's trust in the application.

As a testcase this deliverable aims to assess the feasibility of developing an algorithm to determine whether a field has a cover crop during the winter. In accordance with the European Commission, the term "cover crop" is defined in EC No. 1200/2009 as follows:

> *"It is the arable land on which plants are sown specifically to reduce the loss of soil, nutrients and plant protection products during the winter or other periods when the land would otherwise be bare and susceptible to losses. The economic interest of these crops is low, and the main goal is soil and nutrient protection. Normally they are ploughed in during spring before sowing another crop and are not harvested or used for grazing."*

Several researchers (Ahmed et al., 2023; KC et al., 2021; Najem et al., 2023; Shang et al., 2020) have argued that the use of remote sensing techniques can facilitate the detection of the presence of green manures. With RS it is possible to detect changes in soil texture due to, for example, tillage or crop growth. It is also possible to classify crop types and quantify aspects like biomass production. The advantage of RS compared to field observations is that it allows for the monitoring of large areas over time. Additionally, the freely available data make it a cost-effective method.

1.3. Objective

The goal of this deliverable is to develop an algorithm that can give a reliable prediction on the use of cover crops across the EU. In this way we can ensure that the algorithm will be applicable within the NutriBudget application that will be implemented throughout Europe. Since data sources on cover crop use are lacking in nearly all European countries, using satellite data offers a solution. The fact that data from the Sentinel program are publicly available, provides the advantage of easy data access.

2. Methodology

For the development of the cover crop detection algorithm, we decided to use machine learning. The reason for this is that machine learning is a highly effective method for retrieving patterns from (large) datasets and is applicable for retrieving these patterns from satellite data. A dataset of fields from the past for which the crop and use of a cover crop is known can be combined with satellite observations for these fields at the corresponding times. A machine learning model can be trained to recognize the presence of green manures based on satellite images.

2.1. Datasets used

Several datasets are retrieved and used to develop a machine learning model that can predict the presence of a cover crop.

Satellite data

For the satellite data we will use the observations from the European Sentinel-1 (S1) constellation. S1 uses a RS methodology called synthetic aperture radar (SAR), which scans the Earths' surface using radio waves. The morphology of the vegetation is one of the factors that influences the way the waves are reflected and recorded by the satellite. The variation in backscatter between different crops is significant enough to be able to distinguish between crop types. The advantage of SAR over optical remote sensing, used for example by Sentinel-2, is that it is not dependent on cloud-free days as the radio waves let S1 'see through' the clouds. Therefore, it is possible to monitor seasonal dynamics more accurately (Najem et al., 2023).

Cover crop datasets

To train a robust machine learning model it is necessary to have a dataset with sufficient variation in space and time. In this way, the model can learn to distinguish more accurately between different crops and green manures and the chance of bias is reduced. However, there is no European dataset in which the use of cover crops is registered. There are Land Information Parcel Systems (LIPS) used by member states, but the data structure and availability of the data differs per country and in most systems the presence of cover crops is not registered. There is one exception for the LIPS in France where not only the presence but also the cover crop type has been recorded. Since France is a large country with a large variety in crops and agricultural regions, this dataset is very suitable to evaluate the development of the cover crop detection algorithm.

For the training model we took a random weighted subset of all agricultural fields in France for the years 2019, 2021 and 2021. Different fields were selected for the various years. The year 2020 was excluded due to some issues in the LIPS of that year. This resulted in a set of 17,118 fields with observations. In the table below we show the share of the different crops in the dataset as well as per crop the percentage of the fields that have a cover crop. In Figure 1 we show a map with the location of all the fields. One can see that the selected fields are spread all over France.

Table 1. *Number of fields (n) per crop type in the dataset used to derive a machine learning model predicting the occurrence of cover crops.*

Crop name	n	Crop name	n
Alfalfa	199	**Olive plantations**	94
Beans	199	**Onions**	96
Beetroot	100	**Onobrychis**	99
Cabbage	87	**Pasture meadow grassland**	294
Buckwheat	98	**Peas**	295
Carrots	90	**Plums**	100
Chickpeas	99	**Potatoes**	197
Chicory	99	**Soy soybeans**	100
Clover	97	**Spelt**	95
Cocksfoot catgrass	95	**Spring barley**	783
Dry pules	438	**Spring common soft wheat**	96
Fallow land	293	**Spring oats**	146
Festuca	97	**Sunflower**	193
Flax linen	294	**Sweet chestnuts**	198
Fodder roots	97	**Temporary grass**	98
Fresh vegetables	91	**Tree wood forest**	293
Grain maize	492	**Truffle**	98
Green silo maize	288	**Unmaintained**	91
Hazelnuts hazel	98	**Unspecified orchard**	100
Hemp	100	**Unspecified permanent crops**	97
Industrial crops	95	**Unspecified cereals**	247
Lavender	98	**Vineyards**	291
Lentils	98	**Walnuts**	97
Lolium	98	**Winter barley**	2049
Melon	97	**Winter common soft wheat**	5128
Millet	196	**Winter hard wheat**	98
Miscanthus	96	**Winter oats**	95
Not known	96	**Winter rapeseed**	199
Nurseries	98	**Winter rye**	93
Oak	98	**Winter triticale**	738

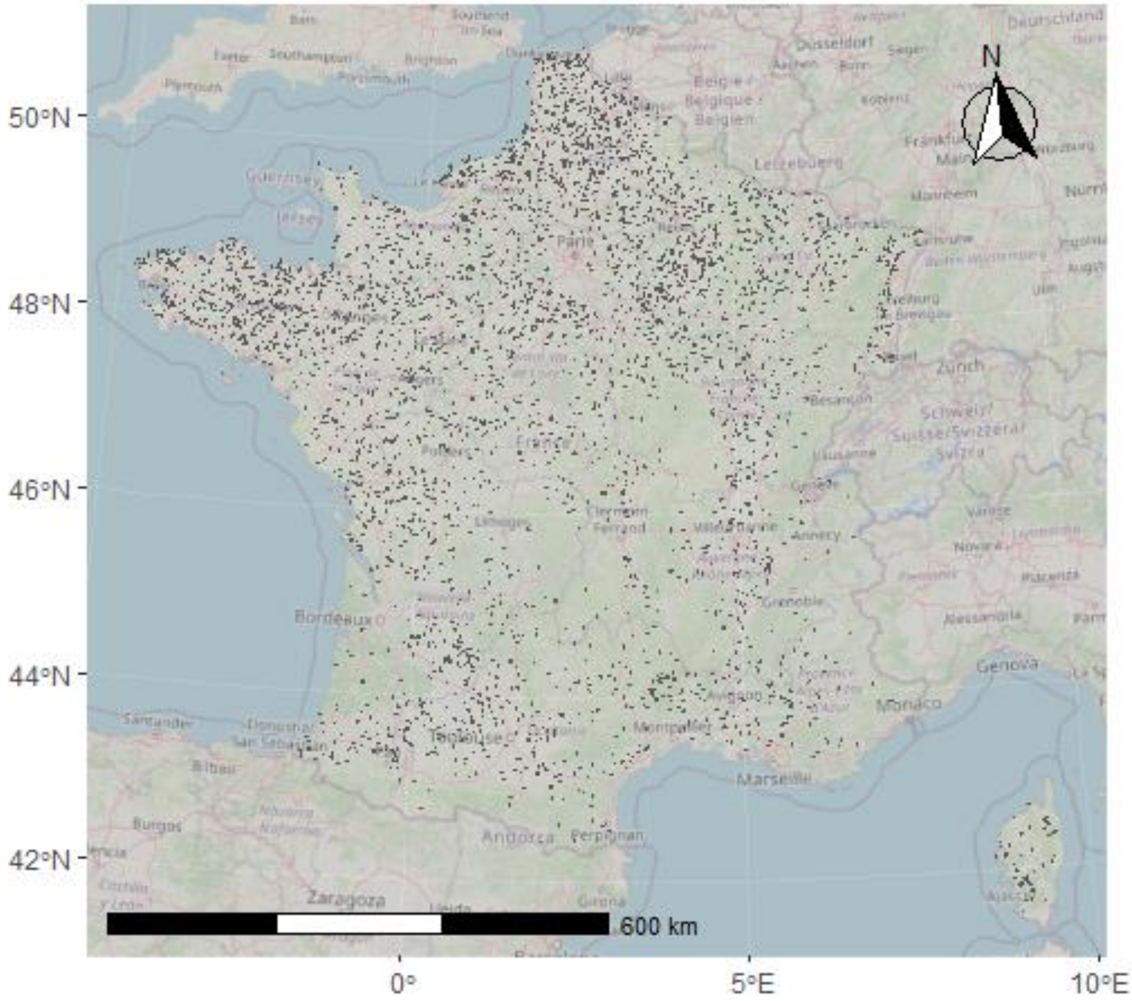


Figure 1 Location of the selected fields in France for model training

2.2. Data processing

Defining the targets to be predicted

When training models, we trained two types of models targeting diverse types of targets. The first model was trained to distinguish whether a field was fallow or not during the winter period. This was a simple approach to see if the model could detect whether a crop was growing during winter. The second model was trained to classify the land use based on four categories in the winter months: winter crop, perennial, cover crop or fallow. This is a more complex approach that requires the model to differentiate between crop types. Winter crops and perennials already grow during the winter months, so there will be no cover crop. Therefore, the model should know that in these cases non-fallow does not mean that there is a cover crop.

Selecting the relevant features, being the explanatory variables

The S1 data for the selected fields were collected using Google Earth Engine (GEE), which is a geospatial analytics platform that contains multiple geospatial datasets. In GEE the S1 data are already partly pre-processed, such as thermal noise removal and terrain correction. In machine learning the variables used to explain the observed variation in the target variable over time and space are called model features. The features used from the S1 data were the VV (single co-polarization, vertical transmit / vertical receive) and VH (single co-polarization, vertical transmit / horizontal receive) bands and the viewing angle of the satellite. These are the data that the model uses to 'read' the surface. In the table below we present the descriptive statistics of the S1 data for the selected fields for the three years.

Table 2. *Observed variation in the VV (single co-polarization, vertical transmit / vertical receive) and VH (single co-polarization, vertical transmit / horizontal receive) and angle features of the satellite data for the selected fields for the three years.*

Parameter	VH	VV	angle
Min	-46.37	-59.41	30.06
Q25	-20.82	-13.28	34.34
Median	-18.74	-11.48	38.40
Mean	-19.00	-11.61	38.18
Q75	-16.94	-9.82	42.12
Max	-2.73	13.90	46.26

Besides the crop and soil morphology, weather affects the backscatter of the radio waves of S1, for example due to soil moisture or snow cover. Therefore, we also collected snow, evaporation, and precipitation data for the selected fields from the ERA5-Land (ECMWF Reanalysis v5) dataset, which is also available in GEE. ERA5-Land is a reanalysis of the global climate and weather for the past 8 decades. Based on model data and observations, a global and consistent dataset is produced that contains hourly weather data. This results in fairly accurate weather data for any location across Europe. The dataset is updated daily and has a latency of 5 days, which also makes the dataset useful in the future. The descriptive statistics of the ERA5-Land data are given in the table below. The data are for the selected fields for the three years.

Table 3. *Observed variation in the weather related covariates for the selected fields for the three years..*

	Snow cover (%)	Snow depth water equivalent (mm)	Evaporation sum (mm)	Precipitation sum (mm)
Min	0.0	0.0	-10.00	0.00
Q25	0.0	0.0	-2.93	0.0035
Median	0.0	0.0	-2.03	0.26
Mean	0.55	0.27	-2.11	2.47
Q75	0.0	0.0	-1.20	2.94
Max	100	820.0	0.00	120.00

In more detail, **snow cover** represents the fraction (0-1) of the cell / grid-box occupied by snow (similar to the cloud cover fields of ERA5-Land). **Snow depth water equivalent** is the depth of snow from the snow-covered area of a grid box. Its units are meters of water equivalent, so it is the depth the water would have if the snow melted and was spread evenly over the whole grid box. The ECMWF Integrated Forecast System represents snow as a single additional layer over the uppermost soil level. The snow may cover all or part of the grid box. The **evaporation sum** is the accumulated amount of water that has evaporated from the Earth's surface, including a simplified representation of transpiration (from vegetation), into vapor in the air above. This variable is accumulated from the beginning of the forecast to the end of the forecast step. The ECMWF Integrated Forecasting System convention is that downward fluxes are positive. Therefore, negative values indicate evaporation and positive values indicate condensation. **The precipitation sum** the collected liquid and frozen water, including rain and snow, that falls on the Earth's surface. It is the sum of large-scale precipitation (that precipitation which is generated by large-scale weather patterns, such as droughts and cold fronts) and convective precipitation (generated by convection which occurs when air at lower levels in the atmosphere is warmer and less dense than the air above, so it rises). Precipitation variables do not include fog, dew or the precipitation that evaporates in the atmosphere before it lands at the surface of the Earth. This variable is accumulated from the beginning of the forecast time to the end of the forecast step. The units of precipitation are depth in meters. It is the depth the water would have if it were spread evenly over the grid box. Care should be taken when comparing model variables with observations, because observations are often local to a particular point in space and time, rather than representing averages over a model grid box and model time step.

Data transformations and standardizations

All variables were collected on a daily scale. With evaporation and precipitation, we calculated the precipitation surplus which in turn was used to calculate a cumulative precipitation deficit throughout the year.

After collecting the data, a final processing step was performed. The S1 data only had observations once every few days and required normalization. First the S1 observations were normalised per field per year, for which we used two different methods.

- The first method (Najem et al., 2023) was VV/VH scaling using the formula: (VV - min(VV)) / (max(VV) - min (VV)).
- The second method (Kaplan et al., 2021) was using the radar vegetation index (RVI) using the formula: (4 * VH) / (VV + VH).

To join the S1 and weather data for each field, the observations were aggregated on a weekly and monthly basis, using the average values of the features. We also calculated these values with VV and VH values corrected for the incidence angle. For that we used the following formula: $\beta / (\sin((90 - \theta) * \pi / 180))^3$, where β is the backscattered signal and θ the incidence angle in degrees (Kaplan et al., 2021).

We also performed a normalisation of the S1 data using a correction for the viewing angle. However, the models that used these normalised values did not perform better than the one without the correction for angle and are therefore not discussed or presented further in this report.

2.3. Developing ML model

Benchmarking and model evaluation

Before starting the training of a machine learning model, a benchmark is created for independent model evaluation. Therefore, the dataset is divided into splits for training, testing and validation. Random subsets were taken from the dataset for testing (15% of the data) and validation (10% of the data).

The metrics used to assess the model's performance were precision, accuracy, recall, F1 score and Cohen's kapa.

- Precision (also called positive predictive value) represents the proportion of positive predictions that are correct.
- The (overall) accuracy is computed by the ratio between the number of the correctly classified test samples and the total test samples. Mean Accuracy of Class N is computed by the ratio between the number of the correctly classified test samples that are labelled as N and the total test samples that are labelled as N.
- Recall (also known as sensitivity) measures the proportion of actual positive cases correctly identified. Also known as sensitivity or true positive rate.

- F1 score (also called F-score or F-measure) is a measure of a test's accuracy. It considers both the precision and the recall of the test to compute the score. The F1 score is the harmonic average of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.
- Cohen's kappa is a statistic that measures inter-rater agreement for qualitative (categorical) items. It is a score that expresses the level of agreement between two annotators on a classification problem. It is generally thought to be a more robust measure than simple percent agreement calculation, as κ takes into account the possibility of the agreement occurring by chance.

These metrics were calculated per target group and crop type. Since some crop types, such as winter wheat and winter barley, have a large share in the dataset, we decided to use an equal amount of fields in the validation set per crop category. This benchmark was finally used for evaluating the model performance for different targets and features.

Model training

The machine learning method we used for developing the cover crop detection algorithm was XGBoost (Extreme Gradient Boosting) (Chen and Guestrin, 2016). XGBoost is a powerful machine learning algorithm that uses an ensemble of decision trees to improve prediction accuracy. It is highly efficient, scalable, and capable of handling both regression and classification tasks. XGBoost works by boosting weak models iteratively to correct errors from previous trees, making it very useful for large datasets with complex patterns.

Parameters that influence the design of the ML can highly influence the performance. Therefore, multiple models were trained with different configurations for the hyperparameters. To efficiently find the optimal set of hyperparameters with the best performance, a technique called Bayesian optimization is used. With Bayesian Optimization a probabilistic model is built to predict the performance of the model using configuration (Chen et al., 2019). By evaluating many possible configurations, the most promising configuration is selected to be next model to train. Hyperparameters were tuned using Bayesian optimization by maximizing the accuracy for the samples in the validation set.

During the model training process, we tried multiple combinations of targets, features, corrections and aggregations in order to find the best approach to detect the presence of cover crops. For these combination, multiple models were trained, from which we selected the ones that had the best performance. Only the latter are described in this report (see section 3).

Model Explanation

When one has trained a model, there are a few methods to interpret how the model arrives at its prediction, thereby opening the black box of the develop machine learning. As a means of model sensitivity analysis we used the following two methods: feature importance and Accumulated Local Effects (ALE).

- Permutation feature importance measures the increase in the prediction error of the model after we permuted the feature's values, which breaks the relationship between the feature and the true outcome. The importance of a feature is calculated as the increase of the models' prediction after permuting the feature. A feature is "important" if shuffling its values increases the model error, because in this case the model relied on the feature for the prediction. A feature is "unimportant" if shuffling its values leaves the model error unchanged, because in this case the model ignored the feature for the prediction (Breiman, 2001; Fisher, Rudin and Doinici, 2018; Wei, Lu and Song, 2015). More details and background information can be found in the book Interpretable Machine Learning from Christoph Molnar (2020).
- Accumulated local effects describe how features influence the prediction of a machine learning model on average. ALE plots are a faster and unbiased alternative to partial dependence plots. If features of a machine learning model are correlated, the partial dependence plot cannot be trusted. The computation of a partial dependence plot for a feature that is strongly correlated

with other features involves averaging predictions of artificial data instances that are unlikely in reality. This can greatly bias the estimated feature effect. ALE plots solve this problem by calculating – also based on the conditional distribution of the features – differences in predictions instead of averages. More details can be found here.

In summary, the feature importance gives an overview of how much the selected features contribute the model prediction, i.e. predicting the variation in space and time for the occurrence of cover crops and the crop type. Variables with less importance can be skipped as explanatory variable. The ALE-plot is used to explain the effect of a feature on the prediction, by showing how changes in a feature impact the prediction.

3. Results Model Performance

3.1. Model accuracy and precision

In table 4 we present the metrics of the best preforming models. The best models are found using a combination of preprocessing steps, including the scaling method, angle correction, and ERA5-Land included as covariate. The overall accuracy varied between 67 and 71% as did the precision. The recall varied from 0.74 to 0.77 and the F1 score varied from 0.64 to 0.70. In table 5 we present the hyperparameters for the best performing binary model (E) and multiclass model (F).

Table 4. *Model performance of the top-5 best scoring ML models to predict cover crop presence.*

Name *	Scaling	Angle correction	ERA5-Land included	Accuracy	Precision	Recall	F1
A – binary	VV/VH	No	No	0.665	0.670	0.737	0.641
B – binary	VV/VH	No	No	0.690	0.690	0.739	0.673
C – binary	RVI	No	No	0.690	0.690	0.741	0.673
D – binary	RVI	No	Yes	0.694	0.694	0.748	0.676
E – binary	RVI	Yes	Yes	0.712	0.712	0.765	0.697
F – multiclass	RVI	Yes	Yes	0.689	0.689	0.732	0.686

* model A makes use of weekly aggregated data where all other models make use of monthly averaged data.

Table 5. *Hyper parameters for the best performing binary and multiclass model*

Variable	E - binary	F - multiclass
booster	gbtree	gbtree
eta	0.235	0.172
gamma	0	0
max_depth	9	9
min_child_weight	1	1
max_delta_step	0	0
subsample	1	1
sampling_method	uniform	uniform
colsample_bytree	1	1
colsample_bylevel	1	1
colsample_bynode	1	1
lambda	9.632	7.159
alpha	0.572	0.136
tree_method	hist	hist
grow_policy	depthwise	depthwise
sample_type	uniform	uniform
normalize_type	tree	tree

The results for the first two models, A and B, show that aggregation of the observations on the monthly level results in a slightly better model performance compared to a situation where this aggregation has not been done. The comparison of model B and C in turn tells us that there is hardly any difference between VV/VH scaling and RVI. However, the RVI models performed better than the VV/VH scaling when the models were expanded with more features and a correction for the incidence angle. Adding the ERA-5 data resulted in a very small improvement in the performance (model D compared to model C). When comparing model D to model E, we see that correcting the backscatter for incidence angle resulted in a bigger improvement and there we saw that adding the ERA-5 data did have an impact on the model performance. The models we discussed up to now only included the binary class models.

Before evaluating the multiclass model, we will discuss the confusion matrixes of the best performing binary models (see Figure 2). Analysing the model performance, we conclude that:

- for 65% of the cases the model correctly predicts whether a field is non fallow, which means there was a cover crop, winter crop or a perennial crop. The precision is much higher, for 94% of the non-fallow fields the prediction is correct.
- for 88% of the cases the model correctly predicted when a field was fallow. However, the precision was fairly low, only 49%, meaning that less than half of real-world fallow fields were identified by the model. The dataset contains crops that never have a cover crop, such as potatoes and soybeans. Note that for these crops it is not necessary to use the model, since one knows beforehand that there will be no cover crop in winter with current agricultural practices in France. Nevertheless, when we remove these crops from the matrix, the overall model performance decreases.

In table 6 below we present the performance per crop, based on the predictions of model E, for those crops that have at least 10 occurrences in the validation set. The prediction of crop cover in winter varies strongly per crop type. Some are always predicted correctly, such as winter wheat, winter barley and winter triticale, while others have a very low accuracy, such as other dry pulses and unspecified cereals. The crops presented in the table below comprise 45% of the observations in the validation set.

Table 6. *Model performance to predict crop cover in winter for the main crop with best binary model E.*

Crop	Accuracy	Precision	Recall	F1	Kappa	n
Winter common soft wheat	1.00	1.00	1.00	1.00	1.00	77
Other dry pulses	0.30	0.30	0.50	0.38	0.30	30
Winter barley	1.00	1.00	1.00	1.00	1.00	29
Fallow land	0.45	0.45	0.50	0.47	0.45	20
Peas	0.90	0.90	0.50	0.64	0.90	20
Spring barley	0.85	0.63	0.92	0.74	0.85	20
Winter triticale	1.00	1.00	1.00	1.00	1.00	18
Grain maize	0.38	0.38	0.40	0.39	0.37	16
Unspecified cereals	0.19	0.57	0.54	0.55	0.19	16
Alfalfa	0.33	0.33	0.50	0.40	0.33	15
Millet	0.69	0.69	0.50	0.58	0.69	13
Beans	0.42	0.42	0.50	0.45	0.42	12
Green silo maize	0.75	0.76	0.75	0.75	0.75	12
Potatoes	0.90	0.90	0.50	0.64	0.90	10

The confusion matrix of the multiclass model is also presented in Figure 2. Here we see that it is challenging for the model to differentiate between the winter crop and the cover crop. Overall, we can say that this model is not good enough for predicting the presence of a crop during winter. The prediction of cover crops is very accurate for all those cases where a cover crop existed: in 95% of the cases the prediction is correct. That can make this model suitable for certain implementations in the NutriBudget application, even though the precision for classifying the type of cover crops is only 51%.

Model F is up to now the best performing model for the multiclass classification, which made use of the RVI and ERA5-Land data, including a correction for the incidence angle. The accuracy and precision of this model is, however, two percent lower than for the best binary model. That is logical, since there are more classes the model should predict. We further discuss the performance of the model at the end of this paragraph, using the confusion matrix.

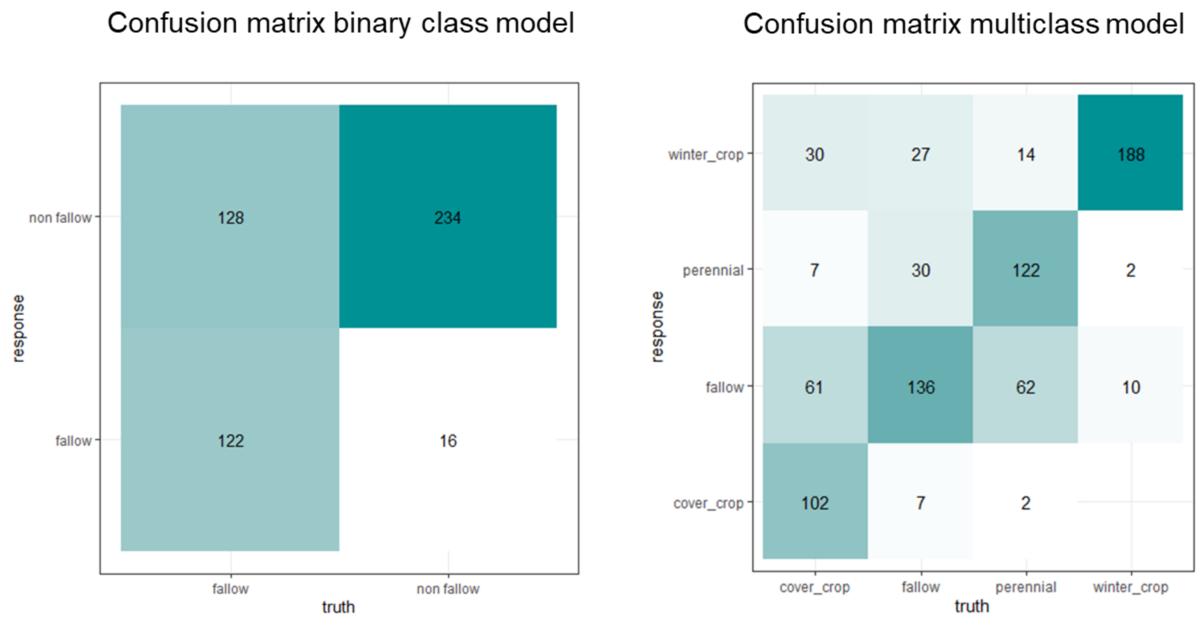


***Figure 2**. Confusion matrix for the best binary class model (left) and best multiclass model (right) predicting the occurrence of a cover crop during winter.*

3.2. Exploring the model features

Here we summarize the main features allowing the machine learning model to predict the occurrence of the cover crop. From the variance importance analysis, it became clear that the RVI features, as well the precipitation surpluses or deficits, are among the top features affecting the binary classification models (see figures in Annex 1). The variables with limited contribution include the snow depth and snow cover. Same findings were found for the multi-class models. For both the binary and the multiclass models the best performing features are RVI features in April and May (multiclass model) or June (binary model). For the binary model this is not very logical, since crop cover in winter should be predicted. These results suggest that the model uses observations earlier in the year to do a crop classification which is then used to predict crop cover in winter. In the binary model there are also precipitation surplus and deficit features with a negative impact on the model, which suggests that some of these features are not relevant and should be left out in the future.

When zooming into the ALE plots of the two best performing features per model we see that they show logical trends (see figures in Annex 2). In all the plots we see that the accumulated dependence decreases with higher RVI values. That is logical since a low RVI values, will most likely mean that there is no crop cover, while high RVI values mean crop cover. The observations on the crop cover are used by the model to determine the crop category.

4. Discussion

For the assessment of remote sensing data to measure relevant farm properties that control the fate of carbon and nutrients, we selected the presence of cover crops as an example and developed a modelling procedure to identify their presence. We illustrated this using a national dataset from France, with 17,118 fields across a large region with variable agroecological properties. Overall, we developed a machine learning allowing us to predict the occurrence of cover crops with 70% precision and accuracy.

One of the Good Agricultural and Environmental Conditions under the new Common Agricultural Policy in Europe stipulates that member states must require farmers to apply crop rotation. This practice is also included in the list of activities supported by eco-schemes. A summary of the first strategic plans shows that most member states incorporate crop rotations, either with a secondary or cover crops during the winter season to protect soils (EC, 2022). The use of techniques to develop spatial explicit datasets from remote sensing data as done in this study is highly valuable to monitor farmers' activities as well as the influence of market developments and policies on farmers' decisions and the subsequent impact on the climate and the environment. Since the new CAP emphasizes performance and results, assessments with detailed data like the one produced in this study will become more important.

With the best-performing models for the binary and multiclass classification it is possible to develop two different implementations for use in the NutriBudget project. The binary model is most effective in recognizing whether a field had crop cover during winter, while the multiclass model can specifically recognize cover crops. The first is essentially what the European Commission is aiming for: to be able to assess whether fields are covered in winter. On the other hand, the multiclass model is more of use for farmers to show that cover crops (as input for the NutriModel) will boost the agricultural sustainability of arable farming systems. If the model predicts that a cover crop has been used, the farmer can be confident that that prediction is correct, reducing the time spent filling in the data.

Given the development, we learned several additional insights being discussed below.

First, the results for the binary model have shown that aggregating the satellite data on a monthly basis is better than the aggregation on a weekly basis. The use of RVI showed to be slightly better compared to the VV/VH scaling. Also, the addition of the ERA5-Land data resulted in an improvement of the model, as did the correction for the incidence angle. This means that it is relevant to add the addition layers of context in the form of weather data and that correcting for the incidence angle helps to improve the accuracy and precision of these machine learning models. From the feature importance plots, we learned however, that the precipitation and deficit data from not all the months contribute to the model prediction. We can also conclude that snow cover and depth can be ignored since they did not contribute to explaining the spatial and temporal variation in the occurrence of cover crops. In other climate zones, where snow occurs more often in winter, these variables could be of more importance. To make the model also applicable for these regions, it could be valuable to collect data from these regions or from winters with significant amounts of snow in France. These are insights to keep in mind for the development of improved models in the future.

Second, the binary model performs well in classifying fields with a crop cover during winter. However, it struggles to accurately identify fallow fields, resulting in a non-reliable classification for non-fallow fields. Despite this limitation, the model could still be practically implemented by focusing on predicting fallow fields, as the accuracy for this classification is very high. When the model predicts the absence of crop cover, that prediction is likely to be reliable. Third, there are various explanations possible for the poor performance of the binary models. The two most important ones are the high occurrence of certain crop categories and the high feature importance of RVI in April and June. The crops presented in the Table 6 comprised 45% of the observations in the validation set, which shows that the other crops have even a lower occurrence. To improve the model prediction for the low performing crop types, the model will also need more observations for these crops, to better recognize their characteristics. Further improvement of the model can also be achieved by only training the model on the autumn and winter months from the dataset. Now the model mainly seems to determine the presence of crop cover in winter based on the observations. By excluding the spring and summer observations from the dataset, the model will be forced to focus on the months in which cover crops are implemented.

Fourth, the dataset seems to show inconsistencies with agricultural practices and regulations in relation to the Nitrate Directive. For example, that cover crops are commonly grown after potatoes, while they do not occur after potatoes in the dataset. Also, in France, cover crops can serve various purposes, such as energy production or animal feed. During the winter, there is no clear distinction between cover crops used as catch crops and those intended for other purposes, as the difference mainly depends on the harvest date (Constantin et al., 2024; Launay et al., 2022). Therefore, in future development of the model it will be relevant to take into account these practices and regulations in the setup and implementation of the dataset.

Fifth, based on results from previous studies, it should be possible to reach higher accuracies for crop cover predictions in winter. Najem et al (2023) have reached recall values up to 95%, for a study in France using Sentinel-1, for example. Their dataset was smaller and contained less crops, which could be a reason for the higher performance. Ahmed et al (2023) also had high model performance, but they made use of Landsat 8 data, which uses optical observations. Adding the visual observations, for example using S2 data, could also improve the model performance as it gives the model more opportunities for characterizing the crop category.

Lastly, for the multiclass model the ERA5-Land data included also relevant features and the angle correction resulted in the model improvements as well. It is logical that the multiclass model is not performing as well as the binary model due to the higher number of categories to be predicted. In fact, the multiclass model as a whole is still too inaccurate for implementation. But since the recall for the crop class ‘cover crop’ is high, it makes the model useful for specifically classifying cover crops, despite the low precision. As mentioned earlier, it is important for a user of the NutriPlatform, that the prediction of the model is accurate. Since that is the case for the detection of cover crops, the farmer can be confident that if the model predicts a cover crop, a cover crop was indeed used.

Note that due to the current lack of information, the model evaluation was conducted only on a dataset from France, assuming that the results are applicable to other regions in Europe. Though France is a particularly good choice due to data availability, the current status of cover crop implementation, and its high diversity of climate regions, the fact that the performance of the results in other regions is still unknown must be addressed in further work. Therefore, we highlight the need for increased data availability and an accompanying definition of what does or does not belong to the cover crop group among EU countries as highlighted by Fendrich et al. (2023). To be able to improve both model types in the future, bigger datasets are needed to distinguish between fallow and field cover and between different crops growing during winter. Ideally, the different crop categories should also be more evenly distributed. Because of the big share of certain crop categories in the current dataset, it can be hard for the model to correctly learn to distinguish between different crop categories.

5. Conclusions and recommendations

Overall, we developed a machine learning allowing us to predict the occurrence of cover crops with 70% precision and accuracy. With the best-performing models for binary and multiclass classification, it is possible to identify whether a field had crop cover during the winter, as well detect a cover crop. Both are useful for use in the actual development of the decision support tool being developed in WP5 of the Nutribudget project. Note that we were only able to train and validate this approach on cropping systems across France given data limitations regarding the presence of cover crops.

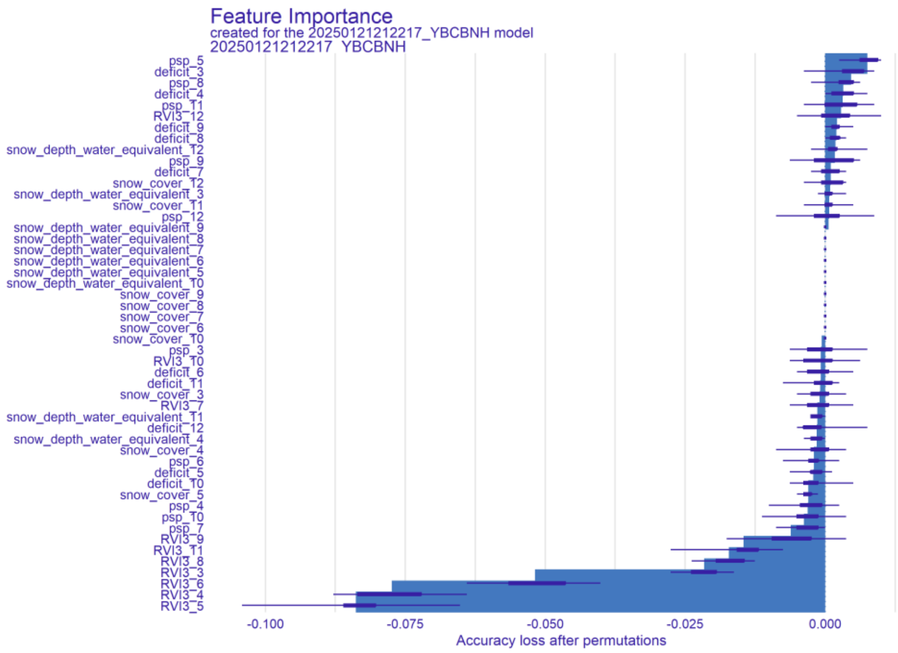
For following work, we recommend:

- Expand the database with country specific datasets regarding the cultivation of cover crops during winter.
- Further model improvement with more crops or additional features like soil and geohydrological properties controlling the presence of crops during winter period.
- Add the developed model as IT service within the IT architecture of the NutriPlatform, allowing others to make use of the prediction / classification model in their own applications.

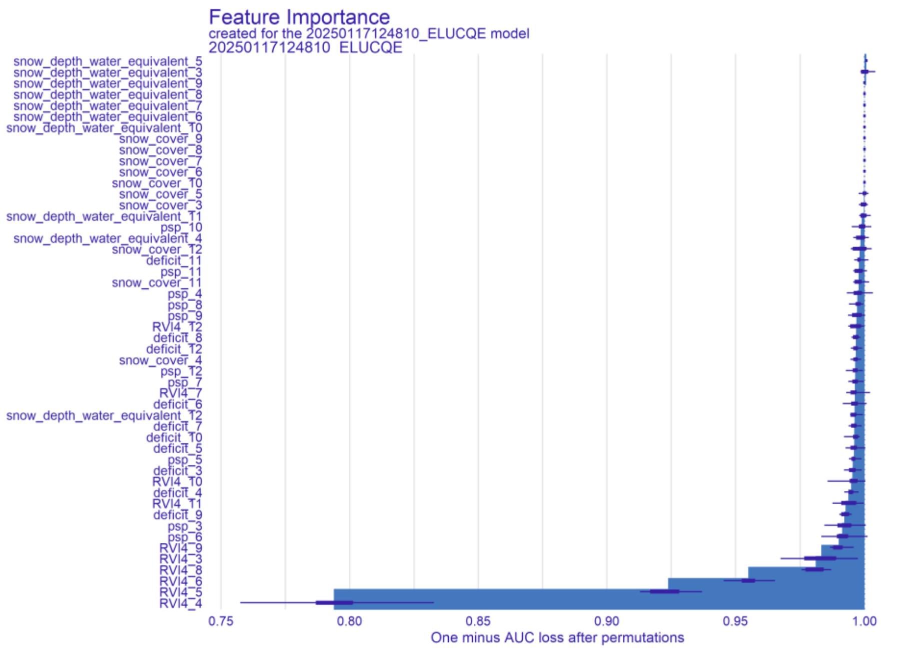
Annexes

Annex 1 FIP plots

Example binary classification model

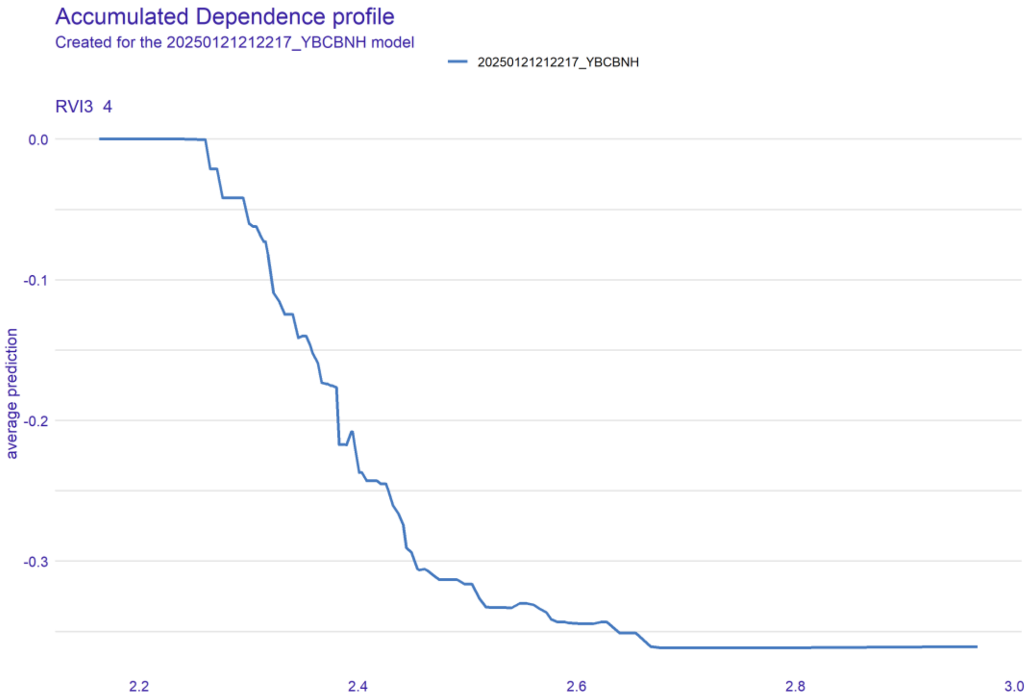


Example multi-class classification model

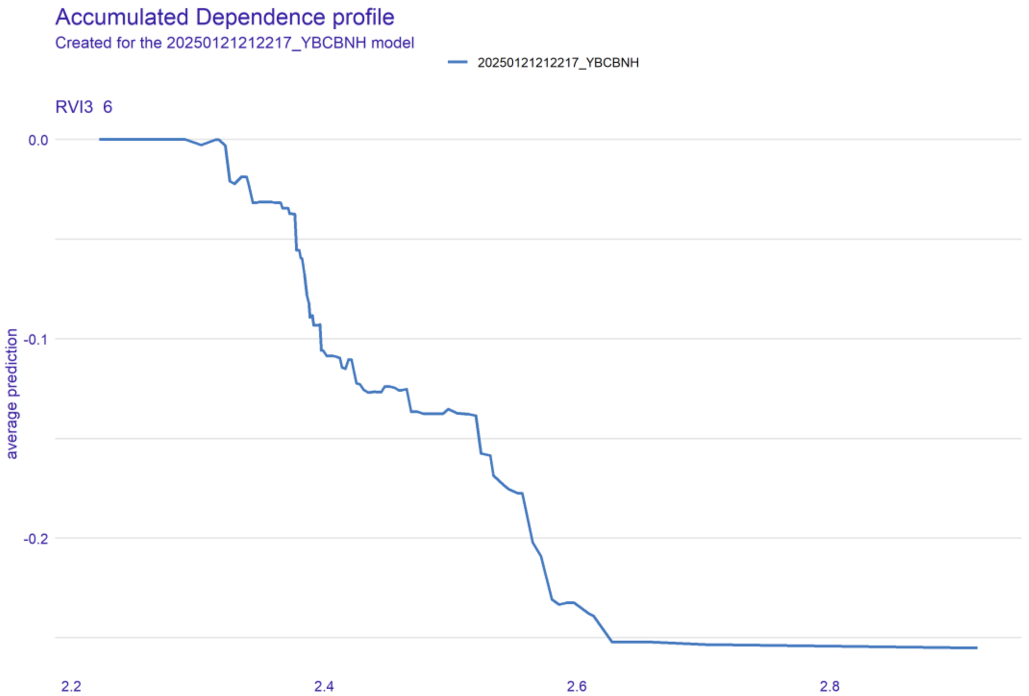


Annex 2 ALE plots

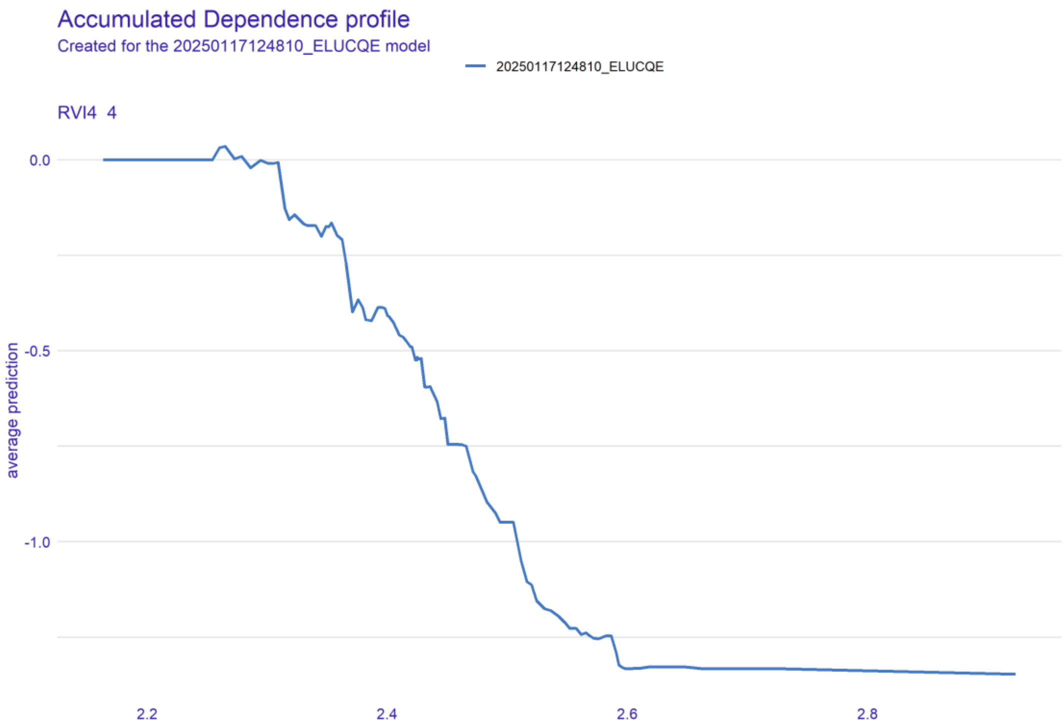
ALE plot for RVI in April for the binary model



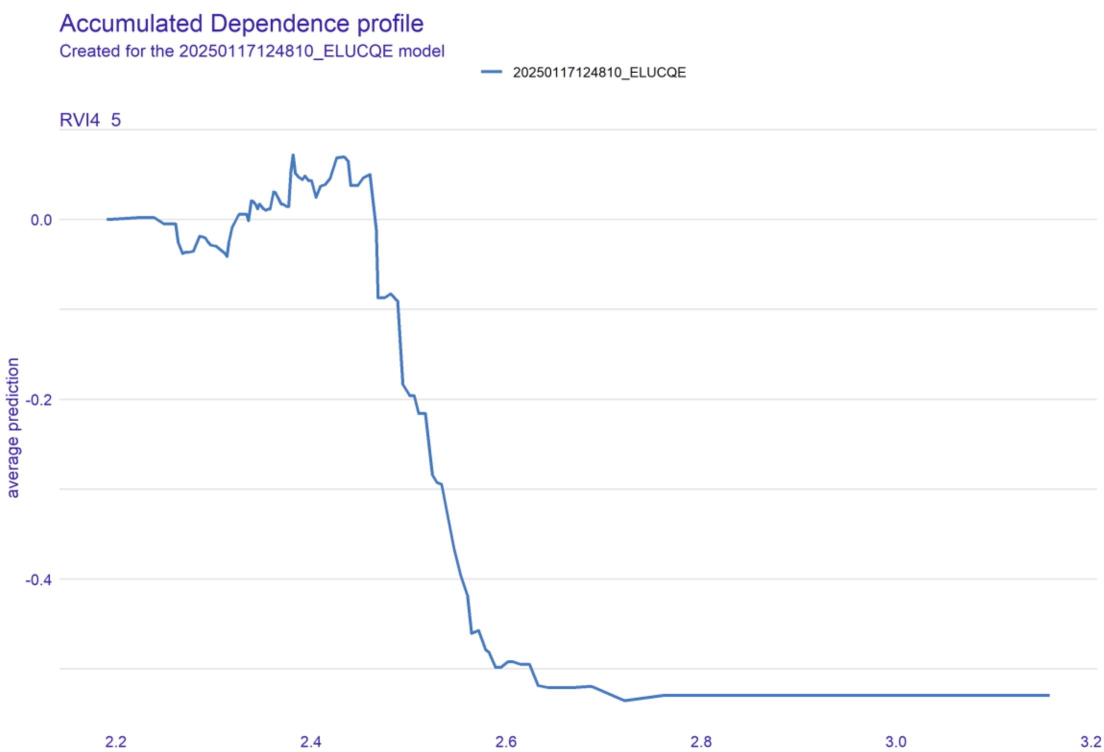
ALE plot for RVI in June for the binary model



ALE plot for April for the multiclass model



ALE plot for May for the multiclass model



List of References

1. Ahmed, Z., Nalley, L., Brye, K., Green, V. S., Popp, M., Shew, A. M. & Connor, L. (2023). Winter-time cover crop identification: A remote sensing-based methodological framework for new and rapid data generation. *International Journal of Applied Earth Observation and Geoinformation*, *125*, 103564.
2. Begue A, Arvor, D., Bellon, B., Betbeder, J., De Abelleyra, D., Ferraz, R.P.D., Lebourgeois, V., Lelong, C., Simoes, M. & S.R. Veron (2018). Remote Sensing and Cropping Practices: A review. *Remote Sensing* 10, 10,99, rs10010099.
3. Bolinder, M.A., Crotty, F., Elsen, A., Frac, M., Kismányoky, T., Lipiec, J., Tits, M., Tóth, Z., Kätterer, T. (2020). The effect of crop residues, cover crops, manures and nitrogen fertilization on soil organic carbon changes in agroecosystems: a synthesis of reviews. Mitig. Adapt. Strateg. Glob. Chang. 25, 929–952. https://doi.org/10.1007/s11027-020-09916-3
4. Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.
5. Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
6. Constantin, J., Minette, S., Vericel, G., Jordan-Meille, L., & Justes, E. (2024). MERCI: a simple method and decision-support tool to estimate availability of nitrogen from a wide range of cover crops to the next cash crop. Plant and Soil, 494(1), 333-351.
7. De la Rosa, D., Mayol, F., Diaz-Pereira, E., Fernandez, M. & D. De la Rosa Jr. (2004). A land evaluation decision support system (MicroLEIS DSS) for agricultural soil protection with special reference to the Mediterranean region. Environmental Modelling & Software, 19, 929–942.
8. Eagle, A.J., Christianson, L.E., Cook, R.L., Harmel, R.D., Miguez, F.E., Qian, S.S., Diaz, D.A.R. (2017). Meta-Analysis Constrained by Data: Recommendations to Improve Relevance of Nutrient Management Research. Agron. J. 109, 1–9.
9. Elia A & G Conversa (2015). A decision support system (GesCoN) for managing fertigation in open field vegetable crops. Part I-methodological approach and description of the software. Frontiers in Plant Science, 6, 1–18. https://doi.org/10.3389/fpls.2015.00319
10. Fendrich, A.N., Matthews, F., Van Eynde, E., Carozzi, M., Li, Z., d'Andrimont, R., Lugao, E., Marin, P., Ciais, P. & P. Panagos (2023). From regional to parcel scale: A high-resolution map of cover crops across Europe combining satellite data with statistical surveys. Science of Total Environment, 873, 162300.
11. Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, *20*(177), 1-81.
12. Hewett CJM, Quinn PF, Whitehead PG, Heathwaite AL & NJ Flynn (2004). Towards a nutrient export risk matrix approach to managing agricultural pollution at source. Hydrology and Earth System Sciences, 8, 834–845. https://doi.org/10.5194/hess-8-834-2004

13. Kaplan, G., Fine, L., Lukyanov, V., Manivasagam, V. S., Tanny, J., & Rozenstein, O. (2021). Normalizing the local incidence angle in sentinel-1 imagery to improve leaf area index, vegetation height, and crop coefficient estimations. *Land*, *10*(7), 680.
14. KC, K., Zhao, K., Romanko, M., & Khanal, S. (2021). Assessment of the spatial and temporal patterns of cover crops using remote sensing. *Remote Sensing*, *13*(14), 2689.
15. Molnar, C. (2020). *Interpretable machine learning: A guide for making black box models explainable*. Leanpub.
16. Launay, C., Houot, S., Frédéric, S., Girault, R., Levavasseur, F., Marsac, S., & Constantin, J. (2022). Incorporating energy cover crops for biogas production into agricultural systems: benefits and environmental impacts. A review. *Agronomy for Sustainable Development*, *42*(4), 57.
17. Najem, S., Baghdadi, N., Bazzi, H., Lalande, N., & Bouchet, L. (2023). Detection and mapping of cover crops using Sentinel-1 SAR remote sensing data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
18. Olesen, J.E., Smith, P., Ingram, J., McVittie, A., Naumann, S., Frelih-Larsen, A., Dooley, E., (2016). SmartSOIL Key Findings, Sustainable farm Management Aimed at Reducing Threats to SOILs under climate change.
19. Ros, G.H., van den Dool, K., van Dam, N., & S le Clech (2023) D3.1 Overview of existing indicators used in national and European policies and market initiatives in relation to agronomic and environmental aims. Nutribudget report, 51 pp.
20. Sarangi A, Madramootoo CA & C Cox (2004). A decision support system for soil and water conservation measures on agricultural watersheds. Land Degradation and Development, 15, 49–63. https://doi.org/10.1002/ldr.589
21. Shang, J., Liu, J., Poncos, V., Geng, X., Qian, B., Chen, Q., ... & Walters, D. (2020). Detection of crop seeding and harvest through analysis of time-series Sentinel-1 interferometric SAR data. *Remote Sensing*, *12*(10), 1551.
22. Silva, J.V., Giller, K.E. (2021). Grand challenges for the 21st century: What crop models can and can't (yet) do. J. Agric. Sci. https://doi.org/10.1017/S0021859621000150
23. Venkatalakshmi B & P Devi (2013). Decision support system for precision farming, 4, 76–81.
24. Wei, P., Lu, Z., & Song, J. (2015). Variable importance analysis: A comprehensive review. *Reliability Engineering & System Safety*, *142*, 399-432.
25. Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., & Deng, S. H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. Journal of Electronic Science and Technology, 17(1), 26-40.
26. Young, M.D., Ros, G.H. & W. de Vries (2021) Impacts of agronomic measures on crop, soil, and environmental indicators: a review and synthesis of meta-analysis. Agriculture, Ecosystems & Environment, 319, 107551.
27. Zhang, X., Yao, G., Vishwakarma, S., Dalin, C., Komarek, A.M., Kanter, D.R., Davis, K.F., Pfeifer, K., Zhao, J., Zou, T., D'Odorico, P., Folberth, C., Rodriguez, F.G., Fanzo, J., Rosa, L., Dennison, W., Musumba, M., Heyman, A., Davidson, E.A. (2021). Quantitative assessment of agricultural

sustainability reveals divergent priorities among nations. One Earth 4, 1262–1277. https://doi.org/10.1016/j.oneear.2021.08.015

Optimisation of nutrient budget in agriculture

Project Coordinators:

Prof. Erik Meers, Erik.Meers@UGent.be
Dr. Ivona Sigurnjak, Ivona.Sigurnjak@UGent.be
Ghent University, Sint Pietersnieuwstraat 25, Ghent 9000, Belgium.

The Consortium:

Funded by the European Union