YIPEEO: Yield Prediction and Estimation using Earth Observation

[Product Validation Report v2.0]

[ID PVRv2.0]

Version 2.0

[10/05/2024]

Submitted by:

Global Change Research Institute CAS (CzechGlobe)

in cooperation with:

TU Wien and EODC

This document was compiled in response of the ESA/AO/1-11144/22/I-EF: ESA Express Procurement Plus – EXPRO+ Theme 2 - Yield Estimation & Forecasting.

This document provides deliverable D3.2 the Product Validation Report (PVR) describing the results of model performance and validation activities of the project YIPEEO.

Number of pages: 24

Authors:	CzechGlobe: Lucie Homolová (LH), Miroslav Pikl (MP), Marian Švik (MŠ), Lenka Bartošová (LB) TUW: Emanuel Büechi (EB), Wouter Dorigo (WD)		
Circulation (internal):	Project consortium		
External:	ESA		
Issue	Date	Details	Editor
0.1	15.1.2024	First draft	LH
1.0	1.2.2024	Final adjustments	LH
2.0	22.4.2024	Include and address comments from ESA	EB

For any clarifications please contact Lucie Homolová (homolova.l@czechglobe.cz).

List of figures

List of tables

Acronyms

CV	Cross-validation
CR	Cross-polarisation Ratio
CZ	Czechia
LT	Lead time (one, two, three or four months before the harvest)
NDVI	Normalized Difference Vegetation Index
NL	The Netherlands
PVR	Product Validation Report (this document)
RMSE	Root Mean Square Error
ubRMSE	unbiased Root Mean Square Error
RF	Random Forest
S-1/-2	Sentinel-1 / Sentinel-2
UA	Ukraine
XGB	Extreme Gradient Boosting

Executive summary

This document is one of two deliverables within Task 3 – Development and Validation described in the proposal of the ESA YIPEEO project. The scope of this document - D3.2 Product Validation Report (PVR) - is to provide a comprehensive summary of the model performance analysis and validation activities for field level crop yield estimation. This subtask on model validation and performance assessment was led by the CzechGlobe-RS team with the support from the TUW and CzechGlobe-A teams.

Due to unexpected delays in the delivery of field level yield data and the availability of satellite predictors, the model development and validation is currently limited to machine learning methods using Sentinel-1 and Sentinel-2 predictors and yield data from Czechia (Rostěnice farm) only. Independent testing was carried out using field data from Czechia (Polkovice farm), the Netherlands and Ukraine for all years (2017 – 2022). The development and testing of methods using a process-based model (Hermes) is ongoing.

Machine learning models have been developed for winter wheat, spring barley and grain maize. Cross-validation during the model development showed very similar performance for both machine learning methods (Random Forest and Extreme Gradient Boosting). The explained variance increased with time closer to the harvest. Models combining both S-1 and S-2 predictors performed better than models trained separately. The models explained 65% of the variance for winter wheat yields one month before the harvest, 55% for spring barley and over 70% for grain maize.

Testing the machine learning models on the independent datasets, either temporal or spatial split, showed limited performance and transferability of the models in time and space. Therefore, to proceed with Task 4 to scale up the yield models to the regional level, we decided to use the Extreme Gradient Boosting method with data from Czechia and neighbouring countries like Slovakia, where we expect similar conditions to Czechia.

1 Introduction

The aim of Task 3.3 is to assess the performance of crop yield models developed at the field level. The validation activities were divided into two phases. First, a random split between training and validation subsets and a 30-fold cross-validation (CV) was used during the development and optimisation of the machine learning methods. We used 30-folds as a tradeoff between computation time and statistical significance. Larger number of folds did not lead to significantly different results. Second, independent test data were used to validate the applicability of the machine learning models in time and space.

Two machine learning methods were used, Random Forest (RF) and Extreme Gradient Boosting (XGB), both described in detail in the D3.1 ATBD. Due to unexpected delays in the delivery of field-level yield data in this project, the models were developed using the data from Czechia (Rostěnice farm) only.

2 Methods

2.1 Validation set-up

An important part of the model development and validation setup was to split the data into training, validation, and test data. As defined in D1.1 RB, approximately 60% of the data was used for training, 20% for validation (i.e. for the model optimization and cross-validation) and the remaining 20% for independent testing. The results in the D3.2 PVR document are therefore refer to model optimisation and cross-validation using the validation subset and independent model testing using the 20% of data that are not used for model development. The set-up of model validation is divided as follows:

1) Random train-validation split: This was used during model development to ensure that the model received a maximum amount of information for both annual and geographical yield distributions. We used a 30-fold random CV.

2) Temporal split: this was done using a leave-one-year-out validation. Each year from 2016 to 2022 was used once for testing purposes. The remaining years are then used for training and validation. This can be considered as the most realistic validation. An operational crop yield forecast, can as well only be trained with data from the previous years, and does not have any information about the crop yields of the forecasted year. The temporal split was performed for three major crops – winter wheat, spring barley and grain maize.
3) Spatial split: as most crop yield data is available for the Rostěnice farm in Czechia. For most other countries, there is not enough data to train a machine learning model. Therefore, we trained and validated the model using the data from Rostěnice only, and test it on all the other countries and farms. Data from another farm in Czechia (Polkovice), Netherlands and Ukraine were used separately as test sites. The spatial split was only done for winter wheat models, as there are not enough field level yields for other crops to perform this validation.

2.2 Validation statistics

For model validation we used following statistics:

- Explained variance (eq. 1) used during model training and cross-validation only
- Pearson's coefficient of determination – R^2 (eq. 2),
- Bias – B (eq. 3),
- Root mean square error - RMSE (eq. 4),
- Relative root mean square error – rRMSE (eq. 5), and
- Unbiased root mean square error – ubRMSE (eq. 6)

$$explained\ variance = 1 - \frac{var(Oi - Pi)}{var(Oi)} \quad \text{eq.1}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Pi - Oi)^2}{\sum_{i=1}^{n}(Oi - \bar{O})^2} \quad \text{eq. 2}$$

$$B = \frac{1}{n}\sum_{i=1}^{n}(Pi - Oi) \quad \text{eq. 3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Pi - Oi)^2} \qquad \text{eq. 4}$$

$$rRMSE = \frac{RMSE}{\bar{O}} \qquad \text{eq. 5}$$

$$ubRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Pi - Oi)^2 - B^2} \qquad \text{eq. 6}$$

were *Pi* are predicted, *Oi* observed yield values, $\bar{O}$ is mean of observed yields.

3 Performance of machine learning methods at the field level

3.1 Model validation during the development and training – random split

For model validation during the development and training, we used a random split to a training and validation dataset with a 30-fold CV. Following graphs show the mean and standard deviation obtained from the 30-fold CV. We evaluated three optimisation techniques and Figure 1 shows that hyperparameter tuning increases the explained variance. We also evaluated the importance of predictors derived from S-1 and S-2 data for winter wheat (Figure 2), spring barley (Figure 3) and grain maize (Figure 4) models. Especially for early forecasts, four months before the harvest, models using S-1 predictors performed better than models using predictors from S-2. Conversely, forecasts one to two months before the harvest performed better with S-2 predictors. Nevertheless, the combination of both S-1 and S-2 predictors always resulted in a higher explained variance of the models.

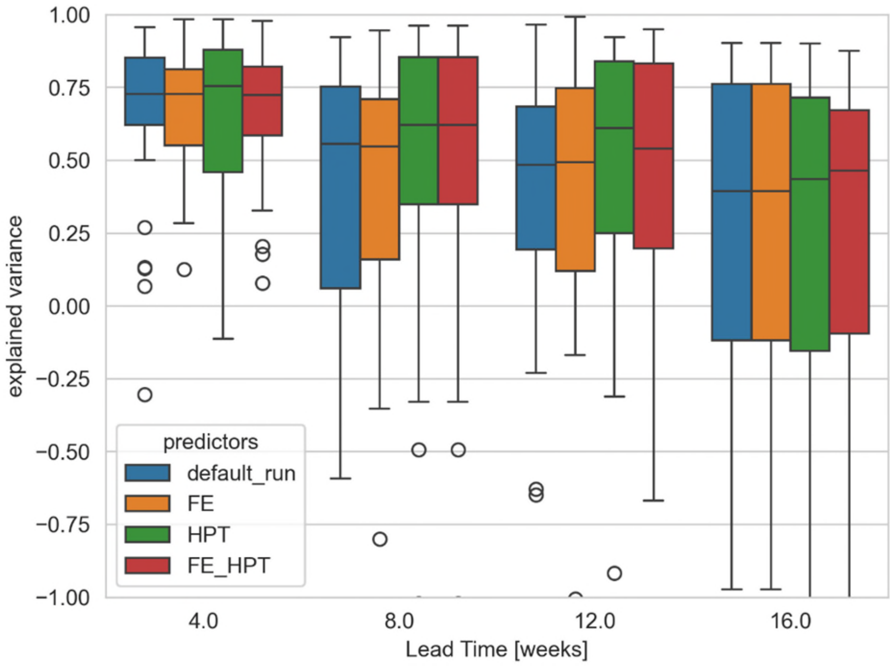

Figure 1. Influence of different optimization techniques (FE – feature elimination, HPT – hyperparameter tuning, FE_HPT - combination of both) on model performance. Cross-validation (random 30-fold CV) of winter wheat model using the XGB method.

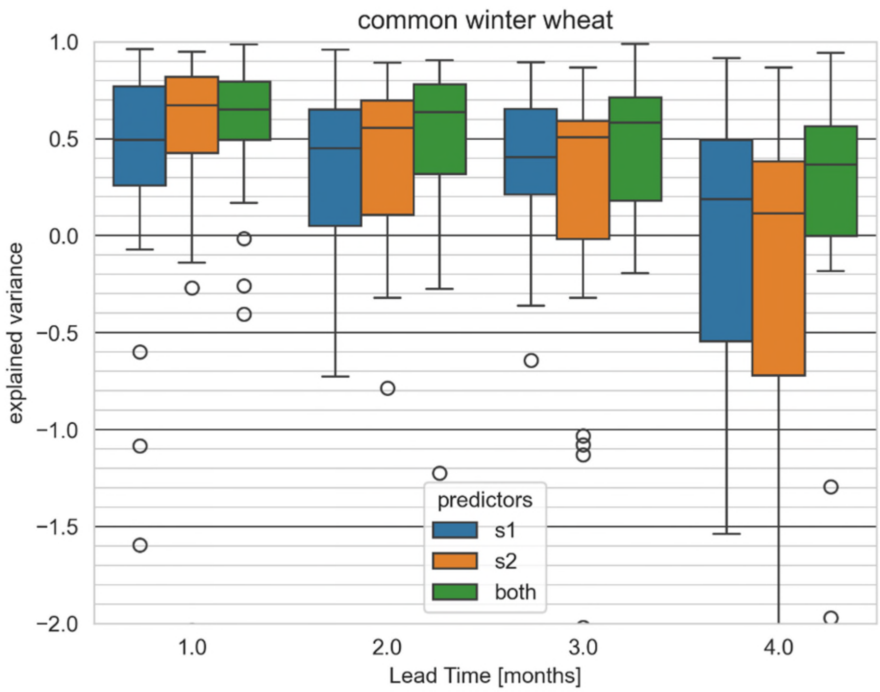

Figure 2. Winter wheat forecast using XGB method. The three barplots per lead-time show the model performance using only Sentinel-1 data as predictors, only Sentinel-2 data, and combination of both. The barplots describe a random 30-fold cross-validation.

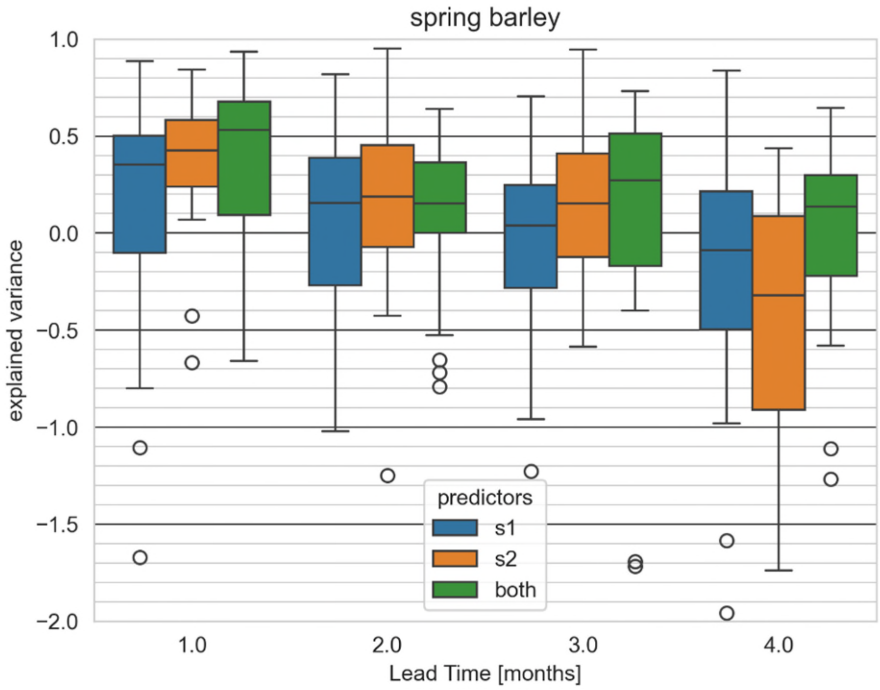

Figure 3. Same as Figure 2 for spring barley.

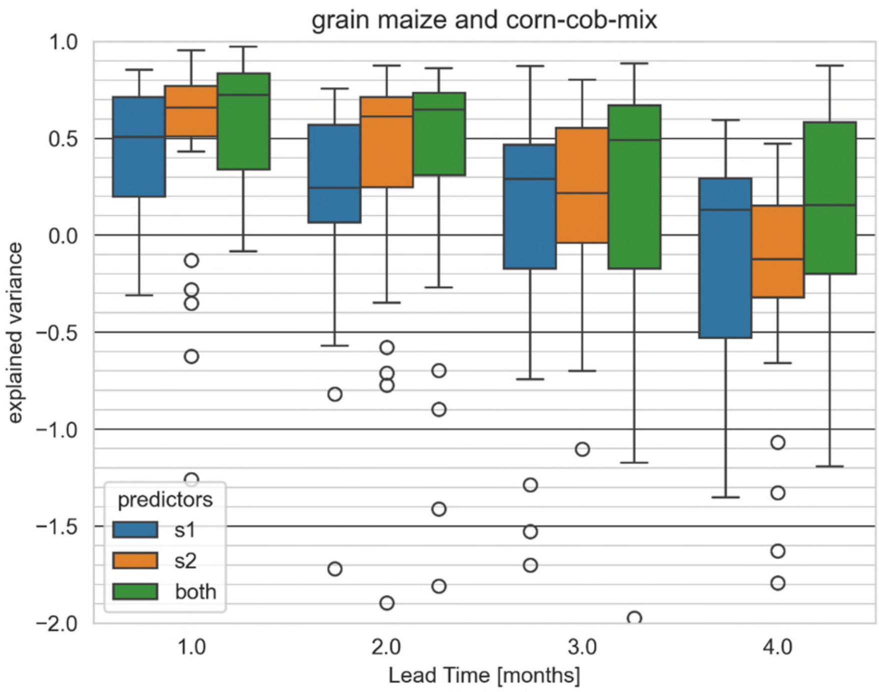

Figure 4. Same as Figure 2 for grain maize

3.2 Model validation using independent test data – temporal split

Temporal split validation was done for each crop and for RF and XGB models trained on CZ data with leave-one-year-out validation. Here, each year was sued once as test data and the model was trained on the remaining years. The results are summarised in Tables 1 and 2. Figure 5 shows R^2 statistics and Figure 6 shows RMSE. The temporal split showed different performance between years for different crops. In general, R^2 did not exceed 0.5 and relative RMSE was on average around 29%. Figure 7 shows the scatter plot between observed and estimated yields for grain maize using the XGB model one month before the harvest.

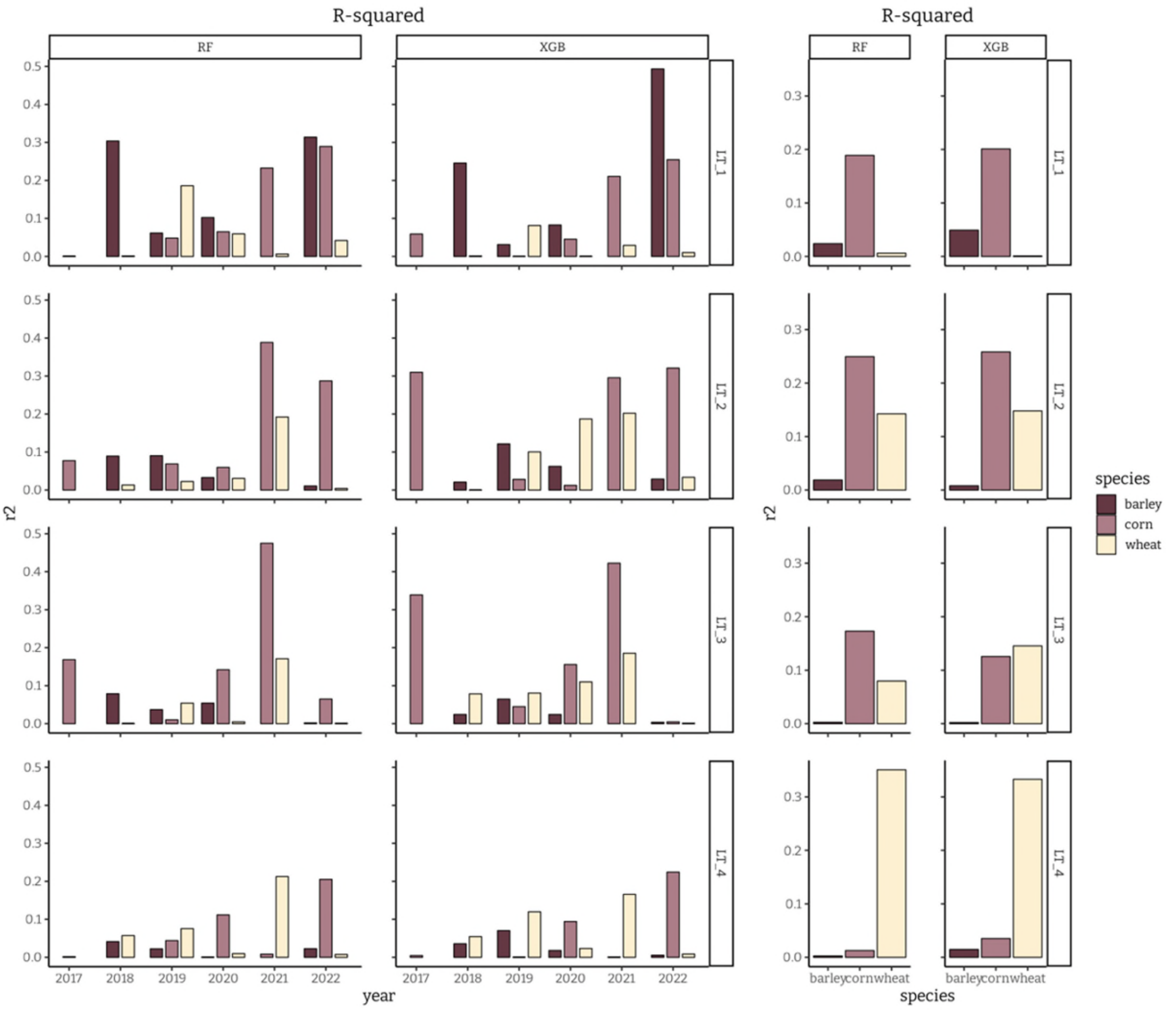

Figure 5. Model validation using independent test data with leave-one-year-out validation – Pearson's coefficient of determination (R^2) for crop yield models trained using CZ field level data (Rostěnice farm) only. The left panels show statistics for individual years used as the test data, the right panel shows statistics for all years combined.

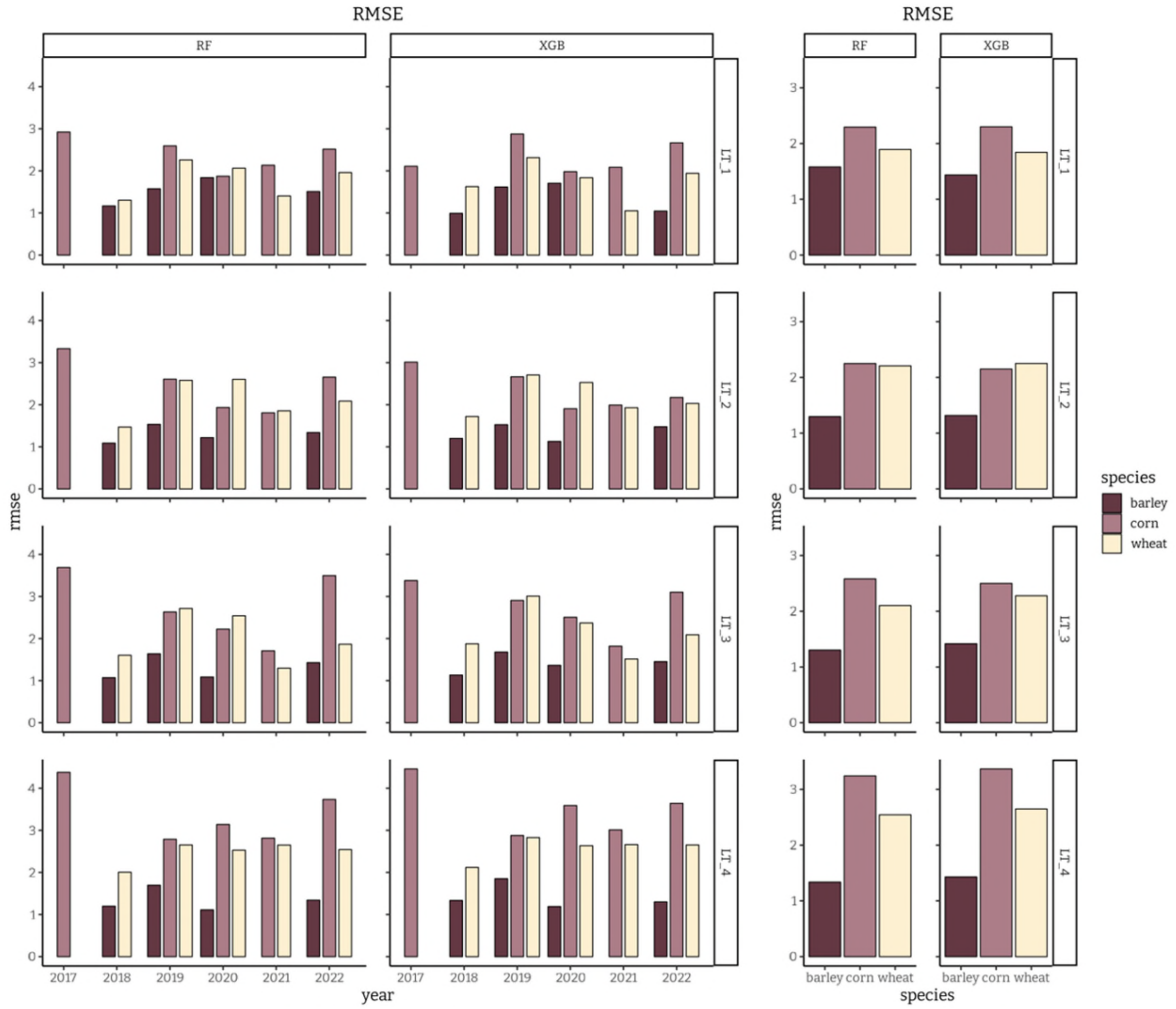

Figure 6. Model validation using independent test data with leave-one-year-out validation – root mean square error (RMSE) for crop yield models trained using CZ field level data (Rostěnice farm) only. The left panels show statistics for individual years used as the test data, the right panel shows statistics for all years combined.

Table 1. Model validation statistics using independent test data for leave-one-year-out validation - individual years.

Method		RF					XGB				
Year	Lead time	R^2	Bias [t/ha]	RMSE [t/ha]	rRMSE [%]	ubRMSE [t/ha]	R^2	Bias [t/ha]	RMSE [t/ha]	rRMSE [%]	ubRMSE [t/ha]
Winter wheat											
2018	1	0.00	-0.77	1.30	0.28	1.05	0.00	-1.14	1.63	0.35	1.16
	2	0.01	-1.05	1.47	0.31	1.02	0.00	-1.23	1.72	0.37	1.21
	3	0.00	-1.17	1.60	0.34	1.09	0.08	-1.24	1.88	0.40	1.41
	4	0.06	-1.75	2.01	0.43	0.98	0.05	-1.87	2.12	0.45	0.99
2019	1	0.19	-1.87	2.26	0.43	1.27	0.08	-1.86	2.31	0.44	1.38
	2	0.02	-1.87	2.58	0.49	1.78	0.10	-1.93	2.71	0.51	1.89
	3	0.05	-1.96	2.72	0.51	1.88	0.08	-2.17	3.01	0.57	2.09
	4	0.08	-1.88	2.65	0.50	1.87	0.12	-1.92	2.83	0.54	2.08
2020	1	0.06	1.85	2.06	0.26	0.91	0.00	1.64	1.84	0.23	0.82
	2	0.03	2.44	2.60	0.33	0.91	0.19	2.22	2.53	0.32	1.21
	3	0.00	2.42	2.54	0.32	0.79	0.11	2.11	2.37	0.30	1.08
	4	0.01	2.33	2.53	0.32	0.99	0.02	2.31	2.64	0.33	1.27
2021	1	0.01	1.00	1.40	0.19	0.98	0.03	0.07	1.05	0.14	1.05
	2	0.19	1.40	1.85	0.25	1.21	0.20	1.32	1.93	0.26	1.41
	3	0.17	0.64	1.30	0.18	1.13	0.19	0.89	1.52	0.21	1.23
	4	0.21	2.23	2.65	0.36	1.43	0.17	2.21	2.66	0.36	1.48
2022	1	0.04	1.52	1.96	0.27	1.24	0.01	1.40	1.94	0.27	1.35
	2	0.00	1.58	2.09	0.29	1.36	0.03	1.53	2.03	0.28	1.34
	3	0.00	1.08	1.87	0.26	1.52	0.00	1.40	2.09	0.29	1.55
	4	0.01	2.31	2.54	0.35	1.06	0.01	2.37	2.65	0.36	1.19
Spring barley											
2018	1	0.30	0.70	1.17	0.21	0.93	0.25	0.29	0.99	0.18	0.95
	2	0.09	-0.28	1.09	0.19	1.05	0.02	-0.35	1.20	0.21	1.15
	3	0.08	-0.15	1.07	0.19	1.06	0.02	0.05	1.13	0.20	1.13
	4	0.04	-0.26	1.20	0.21	1.17	0.04	-0.37	1.33	0.24	1.28
2019	1	0.06	-0.52	1.57	0.29	1.49	0.03	-0.50	1.62	0.30	1.54
	2	0.09	-0.43	1.53	0.28	1.47	0.12	-0.51	1.53	0.28	1.44
	3	0.04	-0.63	1.64	0.30	1.51	0.06	-0.79	1.68	0.31	1.49
	4	0.02	-0.52	1.70	0.31	1.61	0.07	-0.64	1.85	0.34	1.74
2020	1	0.10	-1.57	1.84	0.32	0.96	0.08	-1.28	1.71	0.29	1.13
	2	0.03	-0.63	1.22	0.21	1.04	0.06	-0.45	1.13	0.19	1.03
	3	0.05	-0.47	1.09	0.19	0.98	0.02	-0.76	1.37	0.24	1.14
	4	0.00	-0.16	1.11	0.19	1.10	0.02	-0.03	1.19	0.21	1.19
2022	1	0.31	1.28	1.51	0.24	0.80	0.49	0.80	1.05	0.17	0.67
	2	0.01	0.81	1.34	0.21	1.06	0.03	0.98	1.48	0.24	1.11
	3	0.00	1.05	1.43	0.23	0.97	0.00	1.00	1.45	0.23	1.05
	4	0.02	0.86	1.34	0.22	1.03	0.01	0.78	1.30	0.21	1.04
Grain maize											
2017	1	0.00	-2.25	2.92	0.38	1.87	0.06	-1.47	2.11	0.27	1.52
	2	0.08	-3.00	3.33	0.43	1.44	0.31	-2.73	3.01	0.39	1.27
	3	0.17	-3.46	3.69	0.48	1.29	0.34	-3.17	3.38	0.44	1.16
	4	0.00	-3.96	4.38	0.57	1.86	0.00	-4.10	4.46	0.58	1.74
2019	1	0.05	0.23	2.59	0.20	2.58	0.00	-0.53	2.87	0.23	2.83
	2	0.07	0.43	2.61	0.20	2.57	0.03	0.16	2.66	0.21	2.66
	3	0.01	0.34	2.63	0.21	2.61	0.04	0.13	2.91	0.23	2.90
	4	0.04	-0.19	2.79	0.22	2.78	0.00	0.60	2.88	0.23	2.81

2020	1	0.07	0.80	1.87	0.15	1.69	0.05	1.03	1.98	0.15	1.69
	2	0.06	0.64	1.93	0.15	1.82	0.01	0.74	1.91	0.15	1.76
	3	0.14	1.14	2.22	0.17	1.91	0.16	1.68	2.51	0.20	1.86
	4	0.11	2.69	3.14	0.25	1.62	0.09	2.99	3.59	0.28	1.99
2021	1	0.23	-0.66	2.13	0.18	2.03	0.21	-0.38	2.08	0.18	2.05
	2	0.39	-0.02	1.81	0.16	1.81	0.30	0.36	1.99	0.17	1.96
	3	0.48	0.33	1.71	0.15	1.68	0.42	0.44	1.82	0.16	1.77
	4	0.01	1.36	2.81	0.24	2.46	0.00	1.37	3.01	0.26	2.68
2022	1	0.29	1.05	2.51	0.22	2.28	0.25	0.56	2.66	0.24	2.61
	2	0.29	0.66	2.66	0.24	2.57	0.32	0.42	2.17	0.19	2.13
	3	0.06	2.83	3.50	0.31	2.05	0.00	2.11	3.10	0.28	2.27
	4	0.21	1.45	3.73	0.33	3.44	0.22	0.74	3.64	0.32	3.57

Table 2. Model validation statistics using independent test data for leave-one-year-out validation - all years combined.

Method		**RF**					**XGB**				
Crop type	**Lead time**	**R^2**	**Bias [t/ha]**	**RMSE [t/ha]**	**rRMSE [%]**	**ubRMSE [t/ha]**	**R^2**	**Bias [t/ha]**	**RMSE [t/ha]**	**rRMSE [%]**	**ubRMSE [t/ha]**
winter wheat	1	0.01	0.37	1.89	0.29	1.86	0.00	0.06	1.84	0.28	1.84
	2	0.14	0.53	2.21	0.33	2.14	0.15	0.43	2.25	0.34	2.21
	3	0.08	0.19	2.10	0.32	2.09	0.15	0.23	2.28	0.34	2.27
	4	0.35	0.82	2.54	0.38	2.41	0.33	0.81	2.65	0.40	2.52
spirng barley	1	0.02	-0.27	1.58	0.28	1.56	0.05	-0.35	1.44	0.25	1.39
	2	0.02	-0.23	1.30	0.23	1.28	0.01	-0.18	1.32	0.23	1.31
	3	0.00	-0.16	1.31	0.23	1.30	0.00	-0.26	1.42	0.25	1.40
	4	0.00	-0.08	1.34	0.23	1.33	0.01	-0.11	1.43	0.25	1.43
grain maize	1	0.19	-0.03	2.30	0.20	2.30	0.20	0.00	2.30	0.20	2.30
	2	0.25	0.11	2.25	0.20	2.25	0.26	0.23	2.15	0.19	2.14
	3	0.17	0.91	2.58	0.23	2.42	0.13	0.83	2.50	0.22	2.36
	4	0.01	1.15	3.24	0.28	3.03	0.04	1.04	3.37	0.29	3.20

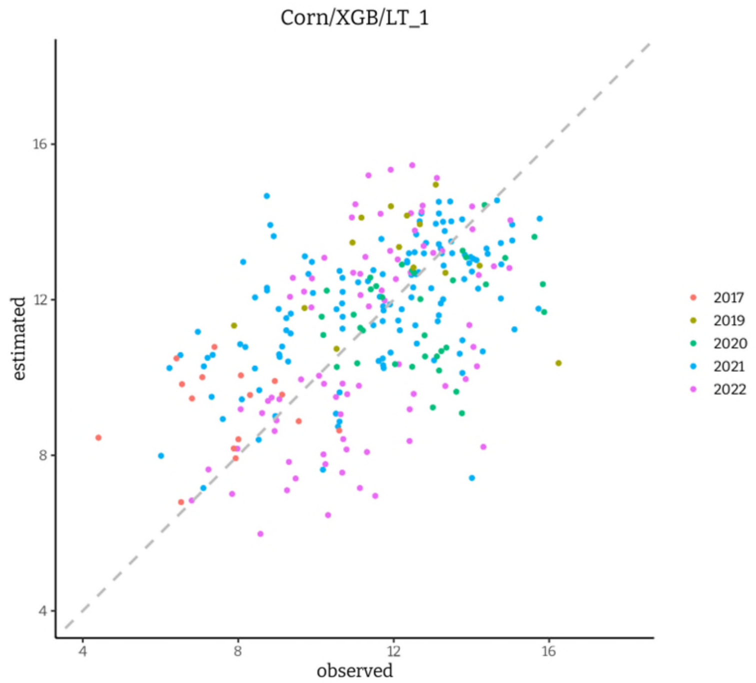

Figure 7. Scatter plot between observed and estimated field-level yields for grain maize using XGB model one month before the harvest with leave-one-year-out validation.

3.3 Model validation using independent test data – spatial split

The spatial split validation was done for winter wheat only. The RF and XGB models were trained on CZ data from all years and applied to another farm in Czechia (Polkovice) and to data from the Netherlands and Ukraine. The results are summarised in Table 3 and Figure 9. When the models were applied on data from another farm in Czechia, the R^2 was between 0.27 and 0.66 depending on the lead time and the relative RMSE was around 30% for the RF model, which slightly outperformed the XGB models (Table 3). Application to Dutch and Ukrainian data failed. In the science cases of Task 5, we will further explore the potential of the transfer-learning approach with more regional scale data to obtain better forecasts for regions outside of Czechia.

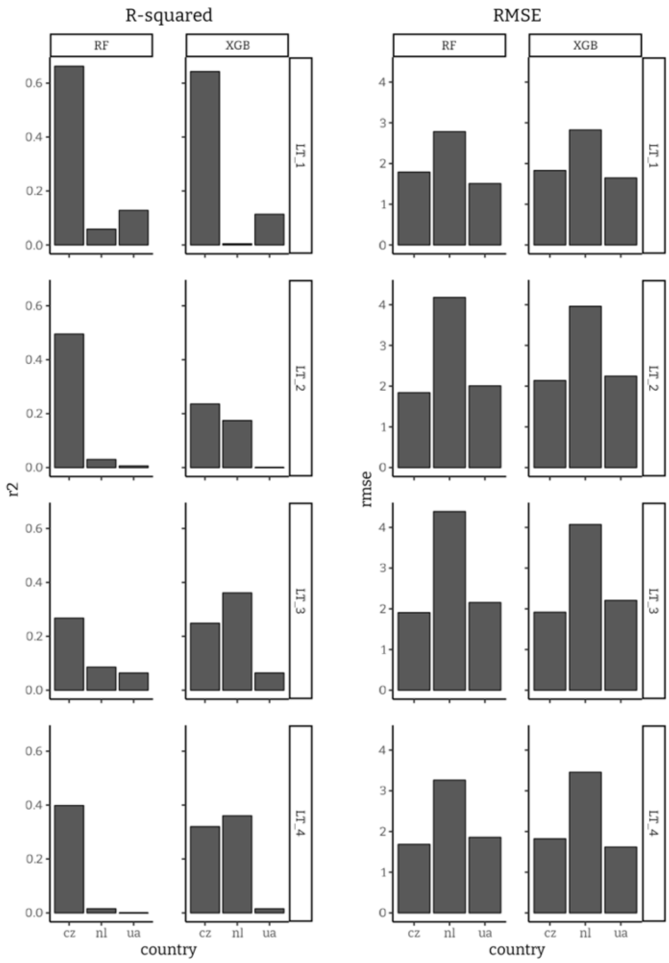

Figure 8. Model validation using independent test data from other countries – Pearson's coefficient of determination R^2 (left panel) and root mean square error RMSE (right panel). The model for winter wheat yield estimation was trained using CZ field level data (Rostěnice farm) and tested on data from CZ (Polkovice farm), NL and UA.

Table 3. Model validation statistics using independent test data from other countries. The model for winter wheat yield estimation was trained using CZ field level data (Rostěnice farm) and tested on data from CZ (Polkovice farm), NL and UA.

Method		RF					XGB				
Country	Lead time	R^2	Bias [t/ha]	RMSE [t/ha]	rRMSE [%]	ubRMSE [t/ha]	R^2	Bias [t/ha]	RMSE [t/ha]	rRMSE [%]	ubRMSE [t/ha]
CZ	1	0.66	-1.26	1.79	0.31	1.27	0.64	-1.33	1.83	0.32	1.26
	2	0.50	-1.04	1.84	0.32	1.52	0.24	-1.26	2.14	0.37	1.73
	3	0.27	-0.89	1.91	0.33	1.69	0.25	-0.85	1.92	0.34	1.72
	4	0.40	-0.63	1.68	0.30	1.56	0.32	-0.81	1.82	0.32	1.63
NL	1	0.06	2.60	2.78	0.27	1.01	0.00	2.62	2.83	0.28	1.06
	2	0.03	4.00	4.18	0.41	1.20	0.17	3.53	3.96	0.39	1.79
	3	0.09	4.05	4.39	0.43	1.69	0.36	3.65	4.07	0.40	1.80
	4	0.02	3.06	3.26	0.32	1.14	0.36	2.90	3.46	0.34	1.87
UA	1	0.13	-0.59	1.51	0.28	1.39	0.11	-0.71	1.65	0.31	1.49
	2	0.01	-1.34	2.01	0.37	1.49	0.00	-1.56	2.25	0.42	1.62
	3	0.06	-1.28	2.16	0.40	1.74	0.06	-1.32	2.21	0.41	1.77
	4	0.00	-1.15	1.86	0.35	1.46	0.02	-0.73	1.62	0.30	1.44

3.4 Comparison validation strategies

The last three chapters showed significant differences between the validation strategies. The random cross-validation during the model development showed promising results for all three crops (R^2 between 0.5 and 0.75 for LT1). The temporal validation showed poor results for all years ($R^2 < 0.3$). The spatial validation for the Netherlands and Ukraine also showed poor results, while the spatial validation for Polkovice (CZ) showed acceptable results. Therefore, in this chapter we will analyse these results in more detail and try to find explanations for these differences.

3.4.1 Temporal split

Here we used each year once as test data and trained the model on the remaining years. Looking at the distributions of the crop yields within the different years (Figure 9), we can see that this may be problematic for some years. The maize and winter wheat yields show a clear gap in the distributions. In the first years (2017 and 2018 for maize and 2017-2019 for wheat) the yields are much lower than afterwards. For maize even the first two years are clearly different from each other and are therefore difficult to predict for a model that is trained for a different range of yields. The last four years have at least a similar maize yield value range. However, having only four years for maize, and three years for winter wheat, may be too few to train the model properly. For spring barley, this tendency is less clear. Only 2021 stands out with significantly higher yields. However, here we can see an increasing trend over the years, which can be related to improvements in agricultural practices than to changes in the predictors. A last point is related to these factors: the range of crop yields within a year is much smaller than between the years. This means that the range of crop yields in the test year is often small and therefore more difficult to predict. This is also reflected in the low correlations between the predictors and the crop yield within individual years (see D3.1 ATBD Chapter 2).

The impact of these factors is much less for a random test-train split. There the ranges of values between the training and testing are much closer and thus easier to predict.

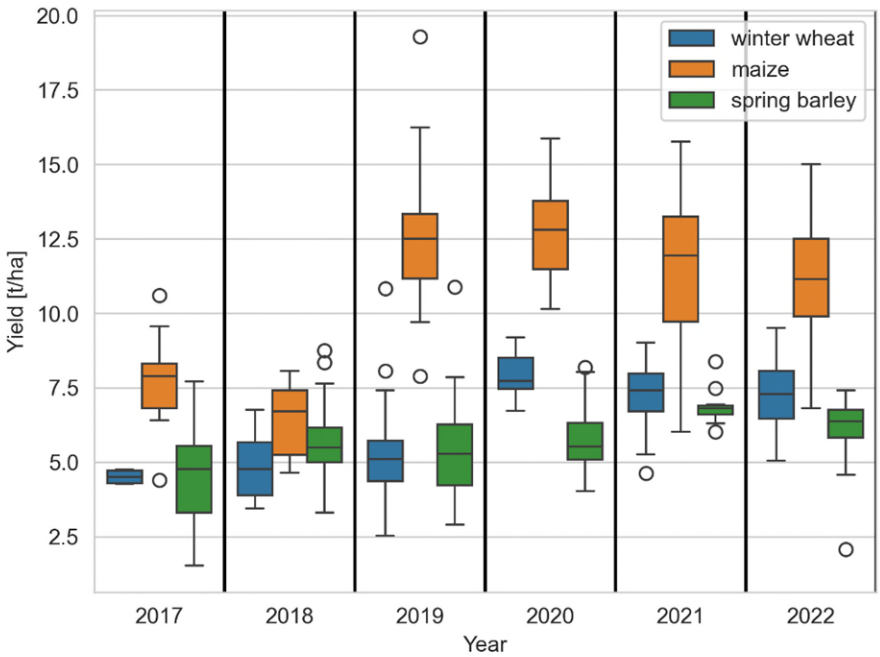

Figure 9. Comparison of the crop yield distributions per crop and year.

3.4.2 Spatial split

The spatial test-train split was used to evaluate how well, a model trained for Rostěnice could be applied in different regions. An obvious limitation of this approach is that crop growing conditions can be very different between countries, i.e., crop yields in arid regions are mainly limited by water availability, whereas in colder countries, temperature is the limiting factor. Therefore, such cross-applicability of the model dependents on the climate. While the Netherlands and Czechia are in the same Koeppen-Geiger-climate class (Cfb: warm temperate, fully humid, warm summers), Ukraine is in a different climate class (Dfb: snow, fully humid, warm summers) (Kottek et al., 2006). In addition, different agricultural practices (irrigation, fertilisation, tillage…) make such comparisons difficult.

When we look at the distribution of the yield data, we can see that the winter wheat yields in Ukraine and Czechia (Rostěnice and Polkovice) are quite similar, while yields in the Netherlands are higher (Figure 10).

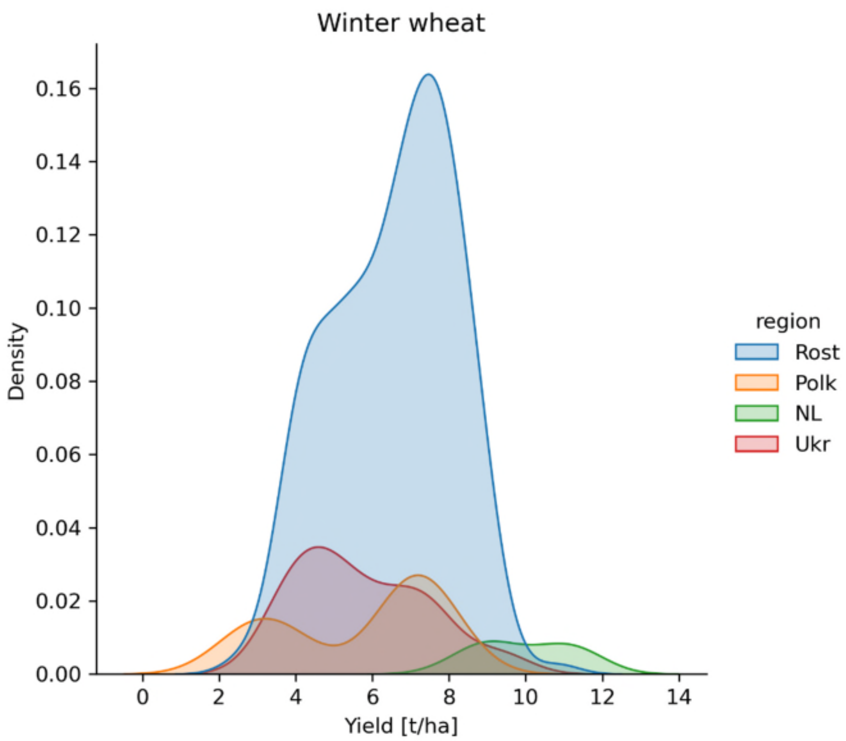

Figure 10. Comparison of the winter wheat yields per region (Rost - Rostěnice farm in Czechia, Polk – Polkovice farm in Czechia, NL – the Netherlands and Ukr – Ukraine).

In addition to the similarity of the crop yields, the similarity of the predictors between test and train data is also crucial (Meyer and Pebesma, 2021). The distribution of the predictors shows a different pattern: only the Ukraine has significantly different NDVI and Sig40 CR values from Rostěnice (Figure 11). The NDVI is lower in the Ukrainian data than in the others. This is related to the slight shifts in harvest dates. In the Ukraine winter wheat is often harvested in the beginning of August, whereas in Czechia the harvest date is end of July. Therefore, the crop cycle is also shifted and NDVI in June is expected to be lower in Ukraine compared to Czechia. This will be addressed in the upcoming tasks by shifting the observation dates of the predictors for the Ukraine.

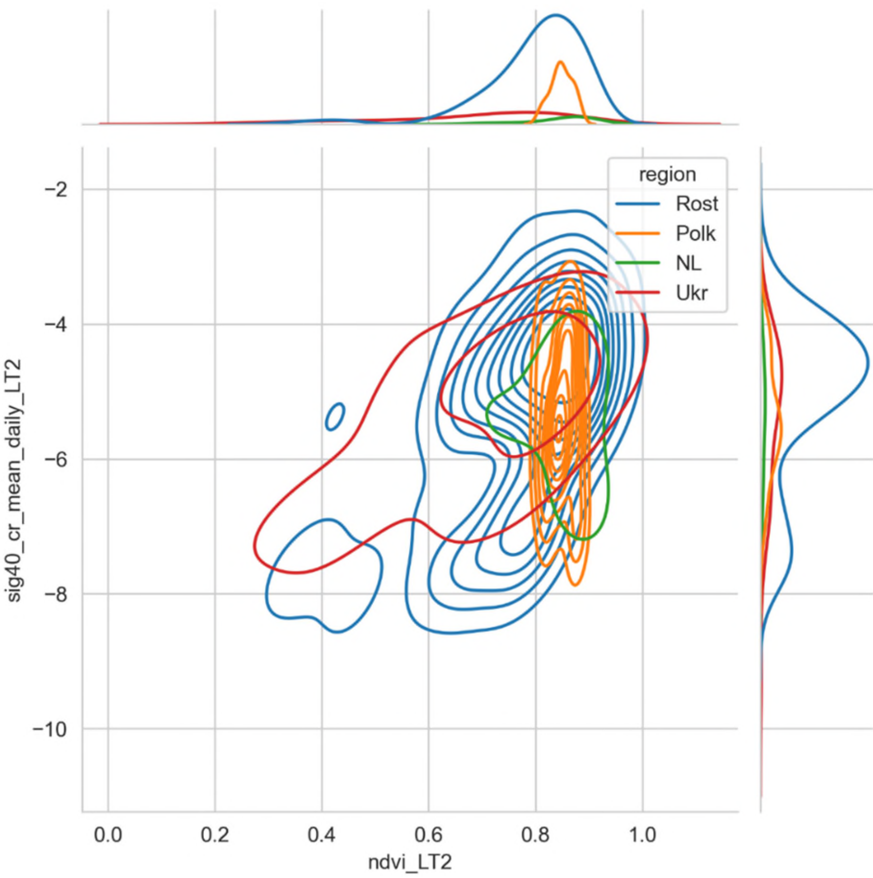

Figure 11. Distribution of two predictors (Sentinel-1 Sig40 CR and Sentinel-2 NDVI for lead-time 2) for the different regions.

When we look at the performance of the models, we can see, that they work for Polkovice only. The Netherlands and Ukraine show a very poor performance (Figures 5 and 6). This can be expected, as the Polkovice farm is quite close to the Rostěnice farm and thus the most similar regarding to relation of yields to predictors. For the Netherlands, the much higher yields made it impossible for the model to predict them correctly. For Ukraine, the differences in the predictors and perhaps even the different climate zone led to the poor performance.

A final factor that needs to be mentioned is the high variability of the model performance in the random cross-validation (Figures 2-4). Some of the 30 folds performed really well, while others showed an explained variance below 0. This means, that even by just predicting the average crop yields for all years and fields better forecasts would have been obtained. If we look at the two validations done here (spatial and temporal splits), we only have one and seven folds for the validation. A final validation is therefore difficult as a different model run could give very different results.

4 Performance of process-based crop yield models at the field level

Methods based on process-based models (Hermes model) are being parameterised and tested with field data from the Polkovice farm (Czechia), so there are no results to present yet. The validation of process-based models will be delivered in the revised version of PVR.

5 Conclusions and method selection

Machine learning models were developed for winter wheat, spring barley and grain maize. Two machine learning methods were tested – Random Forest and Extreme Gradient Boosting. Cross-validation during the model development showed very similar performance for both machine learning methods. Explained variance increased with time closer to the harvest. Models combining both S-1 and S-2 predictors outperformed models that were trained separately. The XGB models explained 65% of the variance for winter wheat yields one month before the harvest, 55% for spring barley and about 75% for grain maize.

Testing the machine learning models on the independent datasets, either temporal or spatial split, showed limited performance and transferability of the models in time and space. We observed that applying models trained on data from Czechia to other countries, such as the

Netherlands and Ukraine, was problematic. Therefore, to proceed with Task 4 to scale up the yield models to the regional level, we decided to proceed with the Extreme Gradient Boosting method, which will be tested on data from Czechia and neighbouring countries like Slovakia, where we expect more similar conditions to Czechia.

6 Reference

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel F. (2006): World map of the Köppen-Geiger climate classification updated, Meteorol. Z., 15 (3), pp. 259-263, 10.1127/0941-2948/2006/0130

Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods in Ecology and Evolution, 12(9), 1620–1633. https://doi.org/10.1111/2041-210X.13650