Yield prediction in winter wheat using machine learning; improving implemented farm management tool

By Mette Kramer Langgaard, Department of Plant and Environment, SEGES Innovation P/S.

Accurate quantification of expected yield before harvest is important to estimate the absolute nitrogen requirement of the crop and consequently adjust the last nitrogen fertilizer application. Since 2020, Danish farmers have had access to a yield forecast model in the web-based management tool, CropManager. The model in production uses Sentinel-2 L1C data to predict winter wheat yield four times during the growth season from April to August.

The objectives of this study were 1) to increase the accuracy and robustness of the yield prediction model in winter wheat by adding more data and new features to the model and 2) to implement the new model in the web-based management tool CropManager used by Danish Farmers.

Originally, the forecast model in CropManager was based on data from 2016 and 2017 alone collected in 2018, which consist of yield data (combine harvesters) from 106 winter wheat fields (1,125 hectares), field polygons, Sentinel-2 L1C data (13 spectral bands and the indices NDVI, NDRE, and MSAVI2), the Danish terrain height model (with a resolution of 0.4 meter), and weather data (including air temperature, evaporation, precipitation, soil temperature, and global radiation).
In 2022 more than 6300 hectare (400 fields) of yield data from combine harvesters were collected from 2016 to 2021. Table 1 shows the distribution of data between years. Combining data collected in 2018 and 2022 gives a total of 747,500 observations (10 x 10 m grids). In addition to adding more data to the original model, new features will be included (wheat variety, soil type, soil water status, evapotranspiration, and crop rotation).

Results from 2020 showed that machine learning (ML) models were able to predict winter wheat yields on field level with a MAE of 622 kg/ha before harvest. With more than 6 times the amount of data in 2022 together with new features, the performance of the models is expected to improve. Several experiments are to be conducted using different ML algorithms and different validation methods (cross-validation and removing entire harvest years as validation). The performance of the models will be evaluated using mean absolute error (MAE), root mean square error (RMSE), R^2 and the standard deviation (SD) of MAE. The results will be ready by the end of 2022.

Table 1. Yield data from combine harvesters from 2016 to 2021 with the distribution of fields and hectares each year and the average yield level.

Year	Yield data		
	Number of fields	Hectare	Avg. yield (ton/ha)
2021	169	3.611	6,6
2020	41	438	7.6
2019	51	619	8.3
2018	51	647	6.6
2017	64	826	8.0
2016	21	190	8.3
Sum	**400**	**6.352**	

STØTTET AF
Promilleafgiftsfonden for landbrug
SEGES INNOVATION