



Estimating winter wheat yield and grain quality using Sentinel-2 imagery and meteorological data

Sarah Carola Vecchietti

Independent project • 30 credits

Swedish University of Agricultural Sciences, SLU

Department of Crop Production Ecology

Agriculture Programme – Soil and Plant Sciences (Agronomprogrammet – mark/växt)

Examensarbeten / Institutionen för mark & miljö, SLU

Partnumber: 2026:16

Uppsala 2026



Estimating winter wheat yield and grain quality using Sentinel-2 imagery and meteorological data

Sarah Carola Vecchietti

Supervisor:	Julianne de Castro Oliveira, SLU, Department of Crop Production Ecology
Assistant supervisor:	Björn Ringselle, RISE Research Institutes of Sweden
Examiner:	Kristin Persson, SLU, Department of Soil and Environment; Soil nutrient cycling

Credits:	30 credits
Level:	Advanced level, A2E
Course title:	Independent project in biology
Course code:	EX1026
Programme/education:	Agronomprogrammet- mark växt
Course coordinating dept:	Department of Crop Production Ecology
Place of publication:	Uppsala
Year of publication:	2026
Cover picture:	Ultuna testbed, photo by the author, 2026
Title of series:	Examensarbeten/Institutionen för mark och miljö
Part number:	2026:16

Keywords:	winter wheat, grain yield, grain quality, Sentinel-2, remote sensing, machine learning, meteorological data
------------------	--

Swedish University of Agricultural Sciences
Faculty of Natural Resources and Agricultural Sciences (NJ)
Department of Crop Production Ecology

Abstract

Accurate estimation of winter wheat (*Triticum aestivum* L.) yield and grain quality can support crop monitoring and agricultural management. This study evaluated the potential of Sentinel-2 imagery and meteorological data to estimate winter wheat grain yield, crude protein concentration and starch concentration at an experimental testbed site in Ultuna, Uppsala, Sweden. Data from four harvest years (2020, 2022, 2023 and 2025) were analysed, comprising 54 observations for grain yield and 48 observations for crude protein and starch concentrations. The experimental plots represented contrasting fertiliser-application treatments, although fertiliser products, nutrient composition and total nutrient inputs were not consistently documented or evaluated across years. Complementary analyses then investigated the effects of fertiliser treatment and year on grain yield and crude protein concentration, as well as associations between selected soil properties using mixed-effects models.

Sentinel-2 spectral bands and vegetation indices were combined with meteorological variables aggregated according to crop development stages. Two predictor-selection approaches were compared: an initial screening based on statistical and agronomic criteria and recursive feature elimination (RFE). Multiple linear regression (MLR), random forest (RF) and support vector machine regression (SVM) were evaluated using leave-one-out cross-validation. Additionally, unmanned aerial vehicle (UAV) imagery collected in 2025 was compared at the plot level with Sentinel-2 imagery using the vegetation index Visible Atmospherically Resistant Index (VARI) to investigate the spatial patterns captured by the two platforms.

Predictors from the vegetative growth stage were particularly important for all three response variables. RFE improved the best predictive performance, although the magnitude of improvement differed among traits. For grain yield, SVM achieved the best performance after RFE (RMSE = 957 kg ha⁻¹; R² = 0.87). For crude protein concentration, RF performed best (RMSE = 0.45 percentage points (pp); R² = 0.84), while SVM showed the best performance for starch concentration (RMSE = 0.67 pp; R² = 0.77). The final SVM grain-yield model was externally validated using 18 independent plot-year observations from the Swedish Infrastructure for Ecosystem Science (SITES) Agroecological Field Experiment (SAFE) at Lönnstorp Research Station in Skåne, Sweden. External performance was substantially lower (RMSE = 1,199 kg ha⁻¹; R² = 0.21), indicating limited transferability beyond the conditions represented at Ultuna. Comparable external grain-quality data were not available.

Mixed-effects analyses showed significant effects of fertiliser treatment and year on grain yield, whereas the Treatment × Year interaction was not significant. In contrast, crude protein concentration showed a significant Treatment × Year interaction, indicating that the response to fertiliser treatment varied among years. UAV imagery collected in 2025 revealed spatial differences in crop development among experimental blocks, highlighting the potential of high-resolution imagery for characterising fine-scale variability within the experimental site.

Overall, Sentinel-2 imagery combined with meteorological data showed potential for estimating winter wheat yield and grain quality within the experimental conditions represented at Ultuna. However, the external validation demonstrated that strong internal predictive performance did not fully transfer to an independent site, highlighting the need for broader calibration and validation across multiple sites and growing seasons.

Keywords: winter wheat, grain yield, grain quality, Sentinel-2, remote sensing, machine learning, weather

Table of contents

List of tables	7
List of figures	9
Abbreviation	12
1. Introduction	13
1.1 Environmental and management drivers of winter wheat yield and grain quality	13
1.2 Remote sensing for agricultural monitoring and precision agriculture	14
1.3 Modelling and feature selection for crop production	15
1.4 Research gap and study objectives	16
2. Materials and methods	19
2.1 Study area and experimental design	19
2.2 Spatial data collection	21
2.3 Sentinel-2 data acquisition, preprocessing and feature extraction	22
2.3.1 Image acquisition and cloud-shadow masking	22
2.3.2 Spectral bands and vegetation indices	23
2.3.3 Plot-level feature extraction and data availability	25
2.3.4 Sentinel-2 data processing and feature aggregation	26
2.3.5 Soil data	26
2.3.6 Agronomic data	27
2.4 Meteorological data	28
2.5 UAV data	29
2.6 Statistical and predictive analyses	30
2.6.1 Statistical analysis of agronomic response variables	30
2.6.2 Modelling datasets and preprocessing	31
2.6.3 Predictor selection and predictive modelling	31
2.6.4 Cross-validation and evaluation metrics	32
2.6.5 UAV–Sentinel-2 comparison	33
2.7 External validation	34
3. Results	36
3.1 Variability of the response variables	36
3.1.1 Grain yield	36
3.1.2 Crude protein concentration	38
3.1.3 Relationship between crude protein and starch concentration	40
3.2 Soil properties of the experimental site	40
3.3 Soil spatial variability and association with grain yield under different fertiliser treatments	41
3.4 Environmental and spectral variability	43
3.4.1 Weather conditions during the growing seasons	43

3.5	Sentinel-2 index variability across crop stages and years.....	46
3.6	Grain yield prediction	47
3.6.1	Predictors selected for grain yield	47
3.6.2	Initial predictor screening	48
3.6.3	Recursive feature elimination	49
3.6.4	Joint RFE of Sentinel-2 and meteorological predictors	51
3.7	Crude protein prediction results.....	51
3.7.1	Predictors selected for crude protein concentration.....	51
3.7.2	Initial predictor screening	52
3.7.3	Recursive feature elimination	53
3.8	Starch prediction results	55
3.8.1	Predictors selected for starch concentration.....	55
3.8.2	Initial predictor screening	56
3.8.3	Recursive feature elimination	58
3.9	External validation.....	59
3.10	UAV.....	61
3.11	Comparison of spatial detail in Sentinel-2 and UAV imagery	61
4.	<i>Discussion</i>	63
4.1	Main findings.....	63
4.2	Agronomic interpretation	63
4.2.1	Grain yield	63
4.2.2	Crude protein	67
4.2.3	Starch	69
4.3	Effect of predictor-selection strategy.....	71
4.4	Comparison of modelling algorithms	73
	External validation	75
4.5.....		75
4.6	UAV-derived spatial variability in crop development	76
4.7	Implications for testbed development and future research	77
4.8	Limitations	79
5.	<i>Conclusion</i>.....	82
	<i>References</i>.....	84
	<i>AI declaration</i>	90
	<i>Popular science summary</i>	91
	<i>Acknowledgements</i>	92
	<i>Appendix</i>	94

List of tables

Table 1. Crop rotation in experimental blocks A, B and C from 2020 to 2025 at the RISE testbed site in Ultuna, Uppsala.	21
Table 2. Spectral bands of the Sentinel-2 MSI sensor extracted from Google Earth Engine and used as candidate predictor variables.(ESA/Copernicus 2026)....	23
Table 3. Vegetation indices calculated from Sentinel-2 surface reflectance bands and used as remote-sensing predictors in the modelling workflow.	24
Table 4. Availability of Sentinel-2 acquisition dates for winter wheat plots after cloud and shadow masking.	25
Table 5. Summary of the independent Lönnstorp dataset used for external validation of the yield prediction model.	35
Table 6. Type III ANOVA results from the linear mixed-effects model for winter wheat grain yield. Num df = numerator degrees of freedom; Den df = denominator degrees of freedom using the Kenward-Roger method; F = F-statistic used to test the significance of each fixed effect.	37
Table 7. Type III ANOVA results from the linear mixed-effects model for crude protein concentration. Num df = numerator degrees of freedom; Den df = denominator degrees of freedom using the Kenward-Roger method. F = F-statistic used to test the significance of each fixed effect.	39
Table 8. Summary of soil texture, organic matter and pH across the 27 experimental plots based on the 2024 soil survey.	41
Table 9. Type III ANOVA results from linear mixed-effects models evaluating clay content and soil organic matter as covariates of winter wheat grain yield. F = F-statistic used to test the significance of each fixed effect.	42
Table 10. Predictors selected for winter wheat grain yield using the initial screening and recursive feature elimination strategies. Predictor names indicate the data source, summary statistic and crop stage. For Sentinel-2 predictors, the first statistic refers to the plot-level zonal statistic and the second to the temporal aggregation within the crop stage; weather predictor names indicate the meteorological summary and crop stage.	47
Table 11. Predictive performance of the models for winter wheat yield using leave-one-out cross-validation (LOOCV).	48
Table 12. Predictive performance of multiple linear regression, support vector machine and random forest models for winter wheat yield evaluated using leave-one-out	

cross-validation. RMSE and bias are expressed in kg ha ⁻¹ . Bias was calculated as predicted minus observed yield.	50
Table 13. Leave-one-out cross-validated performance of models using predictors selected through joint RFE of Sentinel-2 and meteorological variables.	51
Table 14. Predictors selected for crude protein concentration. Predictor names indicate the data source, summary statistic and crop stage. For Sentinel-2 predictors, the first statistic refers to the plot-level zonal statistic and the second to the temporal aggregation within the crop stage; weather predictor names indicate the meteorological summary and crop stage.	51
Table 15. Predictive performance of the models for crude protein concentration using leave-one-out cross-validation (LOOCV).	52
Table 16. Leave-one-out cross-validated performance of random forest, support vector machine and multiple linear regression models for predicting crude protein concentration using RFE-selected predictors.	54
Table 17. Predictors selected for starch concentration using the initial screening and recursive feature elimination strategies. Predictor names indicate the data source, summary statistic and crop stage. For Sentinel-2 predictors, the first statistic refers to the plot-level zonal statistic and the second to the temporal aggregation within the crop stage; weather predictor names indicate the meteorological summary and crop stage.	56
Table 18. Predictive performance of the models for starch concentration using leave-one-out cross-validation (LOOCV).	57
Table 19. Leave-one-out cross-validated performance of support vector machine, random forest and multiple linear regression models for predicting starch concentration.	58
Table 20. Comparison of predictor ranges between the Ultuna training dataset and the independent SITE Lönnstorp external validation dataset for the predictors retained in the final SVM yield model.	60

List of figures

Figure 1. Geographic location of the testbed, the experimental site at RISE in Uppsala, Sweden.	20
Figure 2. Spatial layout of the experimental blocks A, B and C at the RISE testbed site in Ultuna, Uppsala.	21
Figure 3. Location and layout of the external validation plots at the Lönnstorp experimental field. Background: RGB orthophoto ©Lantmäteriet (2025).	35
Figure 4. Plot-level yield maps for the available harvest years. Each polygon represents one experimental plot and is coloured according to measured grain yield (kg ha ⁻¹). The maps show spatial variation in grain yield among plots and differences in grain yield levels among years.	37
Figure 5. Grain yield under three application treatments in 2020, 2022 and 2023. Coloured points represent individual plot observations. Black points represent estimated marginal means from the linear mixed-effects model, averaged across years, and error bars indicate 95% confidence intervals. Different letters indicate significant differences among treatments according to Tukey-adjusted pairwise comparisons ($p < 0.05$).	38
Figure 6. Estimated marginal means of winter wheat crude protein concentration (% dry matter) according to the number of fertiliser applications in 2020, 2022 and 2023. Error bars represent 95% confidence intervals. Different letters indicate significant differences among fertiliser treatments within each year based on Tukey-adjusted pairwise comparison ($p < 0.05$).	39
Figure 7. Relationship between crude protein and starch concentration in winter wheat grain. Points represent individual plot observations, colours indicate harvest year, and the black line represents the fitted linear regression with 95% confidence interval. β represents the estimated regression slope.	40
Figure 8. Spatial distribution of clay content (%) across the experimental plots based on the 2024 soil survey.	42
Figure 9. Spatial distribution of soil organic matter (%) across the experimental plots based on the 2024 soil survey.	42
Figure 10. Seasonal variation in daily mean vapour pressure deficit (VPD) across selected winter wheat crop years. The black line represents daily mean VPD, while the shaded backgrounds indicate the crop-development stages: establishment, overwintering, vegetative growth and grain filling. Each panel represents one harvest year. The dashed horizontal line at 1 kPa is included as	

a visual reference to facilitate comparison among years. Data prior to January 2020 were unavailable.	44
Figure 11. Daily mean air temperature during the winter wheat crop year for 2020, 2022, 2023 and 2025. Shaded areas indicate establishment, overwintering, vegetative growth and grain filling. Data prior to January 2020 were unavailable.	45
Figure 12. Cumulative climatic water balance across the winter wheat crop year for the harvest years 2020, 2022, 2023 and 2025. Background shading indicates establishment, overwintering, vegetative growth and grain filling. The dashed horizontal line represents a neutral water balance. Data prior to January 2020 were unavailable.	46
Figure 13. Sentinel-2 index distributions by crop stage and harvest year. Boxplots show the distribution of plot-level mean for NDRE, NDVI and NDWI during the vegetative and grain filling stages. Individual points represent the mean index value within one plot for one Sentinel-2 acquisition date.	47
Figure 14. Relationship between observed and leave-one-out cross-validation (LOOCV)-predicted winter wheat yield obtained with the support vector machine model. The dashed line represents the 1:1 relationship between observed and predicted values. Symbols indicate the year of observation.	49
Figure 15. Observed versus leave-one-out cross-validated predicted winter wheat yield for the support vector machine model using satellite and meteorological predictors selected through RFE. Each point represents one plot-year observation, with symbols indicating harvest year. The dashed 1:1 line represents perfect agreement between observed and predicted yield.....	50
Figure 16. Relationship between observed and leave-one-out cross-validation (LOOCV) -predicted crude protein concentration obtained using the random forest model. The dashed line represents the 1:1 relationship between observed and predicted values. Different symbols indicate the year of observation.	53
Figure 17. Observed versus leave-one-out cross-validated predicted crude protein concentration for the random forest model. Different symbols indicate harvest years, while the dashed 1:1 line represents perfect agreement between observed and predicted values.....	55
Figure 18. Relationship between observed and leave-one-out cross-validation (LOOCV)-predicted starch concentration for the Random Forest model. The dashed line represents the 1:1 relationship between observed and predicted values. Different symbols indicate the year of observation.....	57
Figure 19. Observed versus leave-one-out cross-validated predicted starch concentration for the support vector machine model. Different symbols indicate harvest years.	

The dashed 1:1 line represents perfect agreement between observed and predicted values.....	59
Figure 20. Observed and predicted winter wheat grain yield for the external validation dataset from the SITES SAFE Lönstorp experimental site (2018-2023). Predictions were obtained using the SVM yield model developed from the Ultuna dataset. The dashed line represents the 1:1 relationship between observed and predicted yield.....	60
Figure 21. Relationship between Sentinel-2- derived VARI on 18 May 2025 and UAV-derived VARI on 8 May 2025 across experimental plots. Each points represents one experimental plot, with the mean VARI calculated from the pixels within that plot for each platform.....	61
Figure 22. Comparison of spatial resolution between Sentinel-2 and UAV vegetation-index imagery across the experimental plots in May 2025. (a) Sentinel-2-derived VARI on 18 May 2025; (b) UAV-derived VARI on 8 May 2025.	62
Figure 23. Examples of weed presence in UAV RGB orthomosaics acquired at the RISE testbed during the 2025 growing season. The images represent different experimental plots and acquisition dates.....	81

Abbreviation

Abbreviation	Description
ANOVA	Analysis of variance
ET _o ,	Reference evapotranspiration
GAI	Green Area Index
GDD	Growing degree days
GEE	Google Earth Engine
GNSS	Global Navigation Satellite System
LAI	Leaf Area Index
LOOCV	Leave-one-out cross-validation
MLR	Multiple linear regression
MSI	Multispectral Instrument
NDRE	Normalised Difference Vegetation Index
NDWI	Normalised Difference Water Index
NIR	Near -infrared
NIT	Near-infrared transmittance
pp	Percentage points
RF	Random forest
RFE	Recursive feature elimination
RGB	Red-green-blue
RMSE	Root mean square error
RTK	Real-time kinematic
SAFE	SITES agroecological Field Experiment
SAR	Synthetic Aperture Radar
SCL	Scene Classification Layer
SITES	Swedish Infrastructures for Ecosystem Science
SPAD	Soil Plant Analysis Development
SVM	Support vector machine
SWIR	Short-wave infrared
UAV	Unmanned aerial vehicle
VARI	Visible Atmospherically Resistant Index
VPD	Vapour pressure deficit
XGBoost	eXtreme Gradient Boosting

1. Introduction

Cereal crops play a fundamental role in global food security, accounting for approximately 42% of global dietary energy supply and 36% of global protein supply in 2023 (FAO 2023). Among cereals, wheat (*Triticum aestivum* L.) is one of the most widely cultivated and economically important crops and plays a central role in global food production (Shiferaw et al. 2013). Consequently, accurate monitoring and prediction of wheat productivity are increasingly important for supporting global food security and sustainable agricultural systems (Papadopoulos et al. 2024).

Wheat is cultivated in both winter and spring forms, which differ in their seasonal development and vernalisation requirements. Winter wheat is typically sown in autumn and requires exposure to prolonged cold before flowering, whereas spring wheat has little or no vernalization requirement and can develop more rapidly after spring sowing (Hyles et al. 2020).

In Sweden, wheat is the largest cereal crop by cultivated area. In 2025, wheat was grown on 525.600 ha, representing more than half of the total cereal area, with winter wheat corresponding to approximately 90% of the total wheat area (Jordbruksverket 2025). The predominance of winter wheat highlights its agronomic importance under Swedish growing conditions.

Winter wheat generally has a longer developmental cycle than spring wheat, which may provide greater opportunities for resource capture, biomass accumulation, and grain formation, potentially resulting in higher yields. However, a longer growing season does not necessarily result in higher grain yield, as yield formation also depends on the availability and utilisation of resources, particularly radiation, water, and nutrients, as well as environmental conditions during critical developmental stages (Slafer et al. 2023). Therefore, monitoring crop growth and nitrogen accumulation across successive phenological stages can provide valuable information for understanding grain yield formation (Meng et al. 2013).

1.1 Environmental and management drivers of winter wheat yield and grain quality

Winter wheat yield is influenced by meteorological conditions, soil properties and agronomic management. Temperature, precipitation, and solar radiation affect crop growth, yield formation and grain composition throughout the phenological stages, with their effects depending on the timing and duration of exposure during crop development (Srivastava et al. 2022).

Grain protein concentration is particularly influenced by nitrogen availability, uptake, and remobilisation, which depend on soil conditions, water availability, and fertiliser management (Khalid et al. 2023). In contrast, starch accumulation is closely associated with photosynthetic activity and the duration and environmental conditions of grain filling (Filip et al., 2023). Grain yield and protein concentration often exhibit an inverse relationship, partly

because greater carbohydrate accumulation can dilute protein concentration in the grain (Cormier et al. 2016)

Temperature plays an important role in regulating crop development and phenology. High temperatures can accelerate crop development, shorten grain filling, and impact grain yield, while frost events during sensitive development stages can reduce grain number and yield (Barlow et al. 2015). In Sweden, spring frosts are not uncommon, although frost events during the reproductive stages of winter wheat are relatively rare (Strandberg et al. 2026).

Water availability also influences wheat growth and grain development. Water stress can reduce photosynthetic activity and affect both yield and grain quality, with particularly detrimental effects when occurring around flowering and during grain filling (Filip et al. 2023). Vapour pressure deficit (VPD) reflects atmospheric evaporative demand, and high VPD can increase transpiration demand and reduce stomatal conductance, ultimately limiting photosynthesis and plant growth (Grossiord et al. 2020).

In addition to meteorological conditions, soil properties and agronomic management can contribute to spatial and temporal variation in wheat productivity. In particular, soil texture and climatic conditions can influence the response to nitrogen fertilisation, affecting grain yield. Soil characteristics influence water and nutrient availability, while nitrogen fertilisation also influences grain protein concentration (Wang et al. 2023). Interannual differences in weather conditions can also contribute to variation in crop responses to nitrogen fertilisation and grain yield among growing seasons (Addy et al. 2020). Therefore, considering soil properties, fertiliser management and harvest year alongside meteorological conditions can consequently provide a more comprehensive understanding of the factors underlying observed yield and quality variability.

1.2 Remote sensing for agricultural monitoring and precision agriculture

Conventional crop monitoring commonly relies on field observations and in situ sampling conducted by farmers, advisors, authorities and researchers to assess crop development, identify potential stresses, and estimate yield and grain quality. Although these methods provide essential ground-truth information, field data acquisition, particularly over large areas, can be time-consuming, costly, labour-intensive and difficult to apply repeatedly (Wu et al. 2023). Moreover, crop monitoring information is needed at different spatial scales, from field and within-field assessments to support farmers' management decisions, as well as regional and national evaluations that can support agricultural planning, production forecasting, and decision-making.

Remote sensing provides a complementary, non-destructive approach for monitoring spatial and temporal variation in crop conditions. It enables the acquisition of information about objects or surfaces without direct physical contact, primarily using sensors that detect the reflected or emitted electromagnetic radiation (NASA 1999). These sensors operate in different regions of the electromagnetic spectrum, including visible, infrared,

microwave, and ultraviolet wavelengths. Depending on the application, remote sensing systems may differ in their spatial, spectral, radiometric, and temporal resolutions (Mulla 2013).

Sensors can also be deployed on several platforms, including satellites, unmanned aerial vehicles (UAVs), and proximal systems (Weiss et al., 2019), providing different levels of detail for the target objects. The spectral information obtained can be analysed directly or used to calculate vegetation indices, which provide indicators of crop vigour, vegetation cover, and physiological status.

Satellite and UAV platforms may provide complementary information at different spatial and temporal scales. For example, the Sentinel-2 multispectral Earth observation mission, managed by the European Space Agency (ESA), provides freely available imagery with systematic acquisitions and spatial resolutions ranging from 10 to 60 m, which are generally suitable for characterising crop conditions at the field and broader spatial scales (ESA 2015). However, these spatial resolutions may be insufficient for detecting fine-scale variability, particularly in small fields or experimental plots (Bukowiecki et al. 2021). In addition, users cannot control satellite acquisition timing, and cloud cover may reduce the availability of usable optical observations. In Sweden, Alshihabi et al. (2024) investigated the use of Sentinel-2 vegetation indices for mapping within-field variation in winter wheat yield. Their study demonstrated the application of satellite-derived spectral information for yield mapping in precision agriculture, while also highlighting the importance of spatial scale when interpreting yield variability.

In contrast, UAV imagery can provide centimetre-scale spatial resolution, making it particularly useful for monitoring small experimental plots and within-plot variability (Haghighattalab et al. 2016). However, UAV data acquisition generally requires more operational effort and processing than accessing existing satellite imagery. UAV missions must be planned and conducted individually, and the resulting large image datasets may require substantial processing (Yang et al. 2017). Limited flight endurance, adverse weather conditions (e.g., clouds, rain, and wind), safety requirements, regulations, and operator expertise may also constrain their use. Previous comparisons have shown that UAV imagery can capture fine-scale crop variation, while Sentinel-2 remains useful for representing broader spatial patterns (Li et al. 2022).

Therefore, combining remote sensing observations with environmental data can further support crop monitoring and assessment by providing complementary information on vegetation conditions and the factors influencing crop development.

1.3 Modelling and feature selection for crop production

Modelling approaches for crop prediction range from conventional linear regression to non-linear algorithms such as random forest (RF) and support vector machine (SVM). In wheat, Sentinel-2 observations have been successfully combined with machine learning methods to estimate grain yield,

with model performance depending on both the algorithm and the temporal information provided to the model (Segarra et al. 2022).

Developing predictive models requires reference measurements of the target variables, such as grain yield and grain quality parameters. Grain yield can be measured by harvesting field plots or using yield monitoring systems installed on combine harvesters, while grain quality parameters, such as protein and starch concentrations can be determined through laboratory analysis of harvested grain samples (Pancorbo et al. 2023). These reference measurements are needed both to train predictive models and to evaluate their performance. One validation approach is cross-validation, in which different subsets of the available observations are held out from model training and used for validation. Independent datasets from other sites or growing seasons can also be used for external validation to assess model transferability under different conditions (Wolters et al. 2022).

Because spectral bands, vegetation indices, temporal summaries and environmental data can create highly redundant and potentially correlated predictors, feature selection can be used to reduce model complexity and improve interpretability. The recursive feature elimination (RFE) method, for example, progressively removes less informative predictors based on their contribution to model performance (Ramezan 2022). Appropriate predictor selection and validation strategies are therefore particularly important when the number of observations with measured values of the target variables available for model development and evaluation is limited.

1.4 Research gap and study objectives

Winter wheat grain and quality are influenced by numerous factors, including meteorological conditions such as temperature, precipitation, solar radiation, and VPD, as well as soil characteristics and agronomic practices. Although remote sensing and meteorological data have been applied to crop monitoring and yield and grain quality predictions, challenges remain when these approaches are applied at the scale of small experimental plots. Satellite observations provide indirect information on crop conditions and may be constrained by spatial resolution, while the influence of meteorological conditions varies across phenological stages. Therefore, further investigation is needed to assess whether integrating satellite-derived crop information with stage-specific meteorological variables can improve the prediction of wheat yield and grain quality traits at plot scale.

Estimating grain quality before harvest is also important, as crude protein concentration is a key indicator of wheat suitability as, for example, for breadmaking, while starch concentration contributes to its nutritional characteristics. Predicting these parameters can therefore provide useful information for assessing grain quality and planning its potential end uses (Khalid et al. 2023; Filip et al. 2023).

The historical dataset available from the RISE experimental testbed in Ultuna, Uppsala, Sweden, also provided an opportunity to investigate agronomic variability across fertiliser treatments and growing seasons.

Although the general effects of fertilisation and environmental conditions on wheat yield and grain quality are well established, the magnitude and consistency of these effects within this specific testbed had not previously been evaluated. Therefore, an exploratory analysis was included to assess the effects of fertiliser treatment and harvest year on grain yield and quality, as well as potential associations between selected soil properties and grain yield in the fertilisation treatments. This analysis was included to characterise the variability represented in the historical dataset, provide context for interpreting the predictive models, and support future applications of the testbed data.

Additionally, UAV imagery collected at the experimental site during the 2025 growing season provided an opportunity to investigate crop variability at a finer resolution than that available from Sentinel-2. Comparing vegetation indices from UAV and Sentinel-2 imagery could help assess how well satellite observations capture plot-level differences and spatial patterns identified with higher-resolution imagery.

Therefore, the main objective of this study was to evaluate the potential of integrating Sentinel-2 multispectral imagery and meteorological data to explain and predict winter wheat grain yield and grain quality, specifically crude protein and starch concentrations, at plot scale under experimental field conditions in a testbed in Uppsala, central Sweden.

To address this overall objective, the study had the following specific objectives:

1. To characterise agronomic variability within the historical testbed dataset by assessing the effects of fertiliser treatment and year on winter wheat grain yield and crude protein concentration, and by evaluating whether selected soil properties were associated with variation in grain yield.
2. To identify the Sentinel-2 and meteorological variables most informative for estimating winter wheat grain yield, crude protein concentration, and starch concentration.
3. To evaluate whether the feature selection method RFE improves predictive performance compared with predictor selection based on statistical and agronomic criteria.
4. To compare the performance of multiple linear regression (MLR), RF and SVM regression for estimating grain yield and grain quality traits.
5. To assess the transferability of the selected yield model using an independent external dataset from the Swedish Infrastructure for Ecosystem Science (SITES) Agroecological Field Experiment (SAFE) at Lönnstorp Research Station in Skåne, Sweden.
6. To assess the relationship between UAV- and Sentinel-2-derived Visible Atmospherically Resistant Index (VARI) values at the plot level and compare the spatial patterns captured by the two platforms.

In the present study, prediction refers to a late-season, pre-harvest estimation of final grain yield and grain quality. Predictor variables were derived from remotely sensed and meteorological information available during the growing season.

Based on these objectives, the following research questions were formulated:

1. What are the main sources of agronomic variability in the historical testbed dataset, and how do fertiliser treatment, harvest year and selected soil properties relate to winter wheat grain yield and crude protein concentration?
2. Which Sentinel-2 and meteorological variables are most informative for predicting winter wheat grain yield and grain quality?
3. Does RFE improve predictive performance compared with predictor selection based on agronomic criteria and correlation analysis?
4. Do machine-learning models outperform multiple linear regression in predicting yield, crude protein content, and starch content?
5. How well does the selected yield model perform when applied to an independent experimental dataset from another site?
6. What spatial differences in crop growth and vegetation patterns can be identified from UAV-derived vegetation information, and to what extent do UAV- and Sentinel-2-derived VARI show similar plot-level patterns?

2. Materials and methods

2.1 Study area and experimental design

The study was conducted at the RISE experimental testbed in Ultuna, Uppsala, Sweden (59.811° N, 17.653°E) (Figure 1). The region has a cold-temperate climate, characterized by cold winters and relatively mild summers. According to the Swedish climate normal period (1991–2020), July is the warmest month with mean temperatures of approximately 17 °C. Climate projections for Uppsala County indicate warming trends, a longer growing season, and increased precipitation, particularly during winter (SMHI 2026). The topography of the study is predominantly flat.

This testbed has been operational since 2018 and was established as an experimental infrastructure to support the digitalisation of agriculture and the evaluation of emerging technologies under field conditions. It was designed to represent a practical crop rotation system under realistic agricultural conditions rather than a strictly controlled experimental design. Its main purpose was to provide infrastructure to integrate and test technologies for crop monitoring and precision management, including applications related to fertilisation efficiency, crop status assessment, and tools for more precise targeting of plant protection products. Since 2023, the testbed has been part of AgrifoodTEF, a European initiative for the testing and validation of new technologies in agriculture (AgrifoodTEF 2026).

The historical data collected at the testbed had not previously been consolidated into a structured and documented dataset. Instead, the available information consisted of separate files from different experimental years, with variations in field configurations, measurement procedures and data formats. In addition, responsibility for the testbed and its associated datasets had shifted between various operators over time, and no continuous standardised system for data management, documentation, and archiving had been maintained. As a result, some information was fragmented or difficult to retrieve. Consequently, a substantial preliminary part of this study consisted of identifying, retrieving, reconstructing and reorganizing the historical data to determine their suitability for comparison across years and integration into subsequent analyses. Reconstructing the historical dataset also provided a clearer overview of limitations in the experimental design, field documentation, and data continuity, helping improve future use of the testbed.

The testbed consists of three experimental blocks (A, B and C) (Figure 2). Each block measures approximately 47m in width and 172 m in length, resulting in an area of about 8084 m², and is divided into nine plots numbered from 1 to 9. Fertilisation treatments were applied across the plots to introduce spatial heterogeneity, creating varied field conditions suitable for testing sensors, remote sensing approaches, and agricultural technologies. To prevent confounding variables, no differences were introduced in herbicides and pesticide practices. The crop rotation in blocks A, B, and C from 2020 to 2025 is summarized in Table 1.

The dataset utilized in this study spans the period from 2020 to 2025, specifically focusing on harvest years 2020, 2022, 2023 and 2025. The year 2021 was excluded due to discrepancies in the experimental design, and 2024 was excluded because winter wheat was not included in the crop rotation.

Since 2025, flower strips containing species mixtures characteristic of Swedish agricultural landscapes have been established within the experimental area to enhance biodiversity and support ecological functions. In addition, several weed species occur within the experimental blocks, frequently at higher densities than those typically observed in commercial agricultural fields. These features contribute to increased habitat heterogeneity and should be considered when interpreting results and extrapolating findings to commercial farming systems.

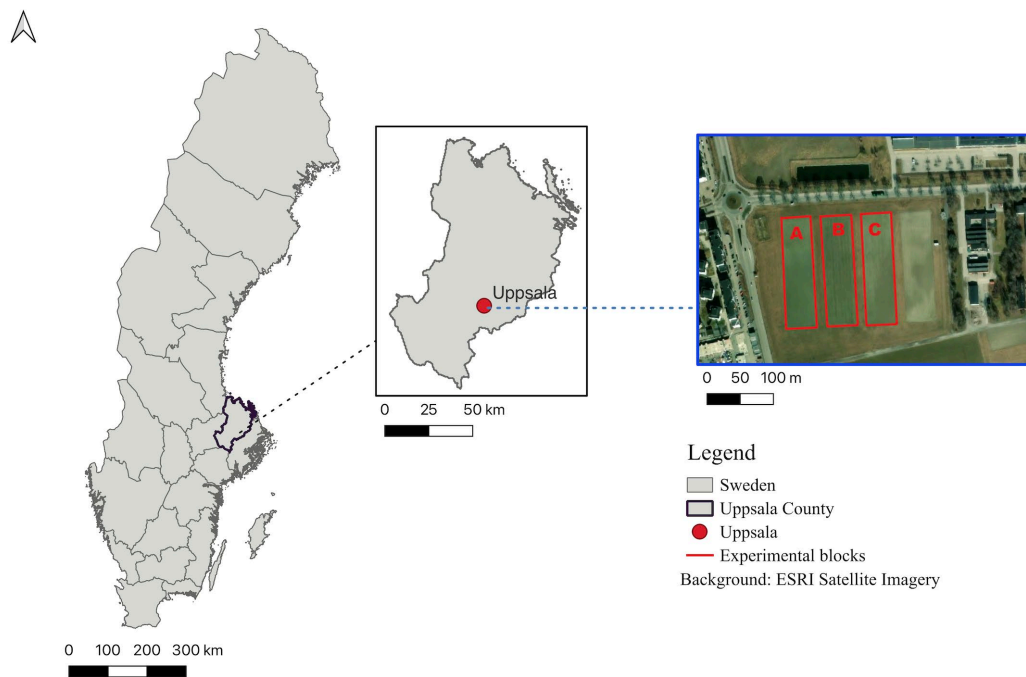


Figure 1. Geographic location of the testbed, the experimental site at RISE in Uppsala, Sweden.



Figure 2. Spatial layout of the experimental blocks A, B and C at the RISE testbed site in Ultuna, Uppsala.

Table 1. Crop rotation in experimental blocks A, B and C from 2020 to 2025 at the RISE testbed site in Ultuna, Uppsala.

Year	A	B	C
2020	Peas	Spring barley	Winter wheat
2021	Winter wheat	Spring barley	Winter oilseed rape
2022	Winter oilseed rape	Peas	Winter wheat
2023	Rye	Winter wheat	Spring barley
2024	Peas	Winter oilseed rape	Peas
2025	Winter wheat	Winter wheat	Winter wheat

2.2 Spatial data collection

As no shapefile describing the experimental layout was available, a new spatial dataset of plot boundaries was developed for this study. Plot limits were reconstructed from reference points collected in the field using the SparkFUN RTK Express Plus receiver (SparkFun Electronics, Niwot, CO, USA) connected to a smartphone via the SW Maps application. The receiver was configured to receive real-time corrections via the NTRIP service of the Swedish national SWEPOS network (Lantmäteriet, Gävle, Sweden), providing centimetre-level positioning accuracy under RTK fixed-solution conditions.

During the measurements, the receiver was kept in a vertical position with an unobstructed view of the sky to minimise signal interference caused by vegetation, buildings or other obstacles. As no survey pole was available, the receiver was held manually at a constant height, while maintaining a stable position during each measurement. Each reference point was collected by keeping the receiver stationary for a few seconds and verifying the presence of an RTK fixed solution before saving the measurement. The collected coordinates were exported as shapefiles and processed in QGIS 4.0.3 “Norrköping” (QGIS Development Team 2026) to delineate the experimental plots.

Boundary points collected around the block were first used to define the block outlines. These outlines were then converted into polygons and used as the basis for generating a regularly aligned plot grid. The polygon orientation angle was calculated and normalised to zero before plot subdivision. After the plot grid was created, it was rotated back to its original orientation and geographic position. The workflow made it possible to generate plot-level shapefiles suitable for the subsequent spatial analyses. An internal buffer of 0.5 m was applied to the plot boundaries within each plot to reduce edge effects, minimise the risk of overlap between adjacent plots and improve the reliability of the remote sensing data extraction.

2.3 Sentinel-2 data acquisition, preprocessing and feature extraction

2.3.1 Image acquisition and cloud-shadow masking

Sentinel-2 imagery was acquired and processed using Google Earth Engine (GEE), a cloud-based platform designed for large-scale geospatial analysis (Gorelick, Noel et al. 2017). Level-2A surface reflectance images were obtained from the COPENICUS/S2_SR_HARMONIZED collection and combined with cloud-probability data from the COPENICUS/S2_CLOUD_PROBABILITY collection. Images with more than 50% cloud cover ($CLOUDY_PIXEL_PERCENTAGE > 50\%$) were excluded.

Cloudy pixels were identified in the whole image using the s2cloudless cloud-probability product. Pixels with a cloud probability above 60% were classified as clouds. Potential cloud shadows were identified as dark pixels with B8 reflectance below 0.15 (scaled value below 1500). To avoid misclassifying water as cloud shadow, pixels classified as water (Scene Classification Layer class 6) were excluded from the dark-pixel mask. Cloud shadows were projected in the direction opposite to the mean solar azimuth, up to a maximum distance of 10 pixels, corresponding to approximately 100 m at 10-m resolution. The cloud-and-shadow mask was then refined using focal minimum and focal maximum operations with a radius of two pixels. This also created an approximately 20-m buffer around detected cloud and shadow areas.

Images acquired from 1 April until the harvest date of each study year were used to cover the main winter wheat growing season. Only plots cultivated

with winter wheat in the corresponding year were included, according to the crop-rotation scheme (Table 1). Images acquired on the same date were combined into daily mosaics to avoid duplicate observations caused by overlapping Sentinel-2 acquisitions. Finally, the cloud-and-shadow mask was applied to the surface reflectance bands before calculating vegetation indices and extracting spectral information for each plot.

2.3.2 Spectral bands and vegetation indices

The Sentinel-2 Multispectral Instrument (MSI) provides observations in the visible, red-edge, near-infrared (NIR), and short-wave infrared (SWIR) regions of the electromagnetic spectrum. The spectral bands have spatial resolutions of 10, 20, or 60 m, depending on the wavelength region and the intended application. The visible bands provide information on canopy colour and pigment absorption, while the red-edge and near-infrared bands are particularly sensitive to vegetation condition, canopy structure, chlorophyll content, and biomass. The SWIR bands provide complementary information related to vegetation and soil water content.

The spectral variables extracted from the Sentinel-2 Level-2A surface reflectance collection included bands B1–B9, B11, and B12. Bands B1 and B9, which are primarily intended for aerosol and water-vapour applications and have a resolution of 60 m, were retained as candidate predictors as part of the broad exploratory screening of Sentinel-2 spectral information. Although these bands are less commonly used for vegetation monitoring, they were included in the modelling workflow together with other available spectral bands. The characteristics of the Sentinel-2 bands used as candidate predictors are presented in Table 2.

Table 2. Spectral bands of the Sentinel-2 MSI sensor extracted from Google Earth Engine and used as candidate predictor variables. (ESA/Copernicus 2026).

Band	Spectral region	Central wavelength (nm)	Spatial resolution (m)
B1	Coastal aerosol	443	60
B2	Blue	490	10
B3	Green	560	10
B4	Red	665	10
B5	Red-edge 1	705	20
B6	Red-edge 2	740	20
B7	Red-edge 3	783	20
B8	Near infrared (NIR)	842	10
B8A	Narrow near infrared	865	20
B9	Water vapour	945	60
B11	Short-wave infrared 1 (SWIR 1)	1610	20

In addition to the individual spectral bands, three normalized-difference indices were calculated: the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Red-Edge Index (NDRE), and the Normalized Difference Water Index (NDWI). These indices were selected to represent complementary aspects of crop condition, including canopy greenness and biomass, chlorophyll status, and vegetation water content. In addition, the Visible Atmospherically Resistant Index (VARI) was calculated from the visible bands to enable comparison with UAV-derived VARI.

NDVI was calculated using the red and near-infrared bands and was used as a general indicator of vegetation greenness and canopy development. NDRE combined the narrow near-infrared and red-edge bands and was included as an indicator of chlorophyll content and crop physiological status. NDWI was calculated using the near-infrared and SWIR bands and was used to capture variations associated with vegetation water content. The equations and Sentinel-2 band combinations used to calculate the indices are reported in Table 3.

Table 3. Vegetation indices calculated from Sentinel-2 surface reflectance bands and used as remote-sensing predictors in the modelling workflow.

Name	Abbreviation	General equation	Sentinel-2 implementation	Reference
Normalized Difference Vegetation Index	NDVI	$\frac{NIR - Red}{NIR + Red}$	$\frac{B8 - B4}{B8 + B4}$	(Rouse et al. 1974)
Normalized Difference Water Index	NDWI	$\frac{NIR - SWIR}{NIR + SWIR}$	$\frac{B8 - B11}{B8 + B11}$	(Gao 1996)
Normalized Difference Red Edge index	NDRE	$\frac{NIR - Red\ Edge}{NIR + Red\ Edge}$	$\frac{B8A - B5}{B8A + B5}$	(Barnes et al. 2000)
Visible Atmospherically Resistant Index	VARI	$\frac{Green - Red}{Green + Red - Blue}$	$\frac{B3 - B4}{B3 + B4 - B2}$	2026-10-15 13:01:00

The inclusion of the Sentinel-2 red-edge bands was particularly relevant for winter wheat monitoring because reflectance in this spectral region is sensitive to changes in leaf chlorophyll concentration, canopy structure, and plant physiological condition. Red-edge-based variables may therefore provide information that is complementary to conventional indices such as NDVI. Moreover, NDVI can become less sensitive under dense-canopy conditions, whereas red-edge indices may retain a greater capacity to distinguish variations in crop vigour and productivity (Lu et al. 2025).

2.3.3 Plot-level feature extraction and data availability

Sentinel-2 spectral information was extracted using the EPSG:32633 coordinate reference system, corresponding to WGS 84/UTM zone 33N. Spectral-band and vegetation-index values were extracted at a nominal spatial scale of 10 m.

Because Sentinel-2 bands have different native spatial resolutions, the bands originally available at 20 or 60 m were evaluated at the common 10-m extraction scale. No explicit resampling method was specified in the Google Earth Engine workflow. Consequently, Google Earth Engine's default nearest-neighbour resampling method was applied when values from the lower-resolution bands were requested at a 10-m scale. This procedure aligned the different spectral bands within a common extraction grid but did not increase the amount of spatial information contained in the original 20- or 60-m bands.

The buffered polygons were used as the spatial units for spectral extraction, and no additional geometric buffer was applied within Google Earth Engine. The inward buffer reduced the potential influence of neighbouring plots, plot boundaries, tramlines, and mixed pixels on the extracted spectral values.

For each Sentinel-2 acquisition date, spectral-band and vegetation-index values were spatially summarised within each buffered plot. The calculated statistics included the mean, median, maximum, minimum, standard deviation, and the 25th and 75th percentiles. These statistics were used to characterise both the central tendency and the within-plot spatial variability of the spectral response.

Plot-level mean values were calculated using the default weighted reducer in Google Earth Engine. Pixels partially intersecting the buffered plot boundaries contributed to the calculated mean according to the proportion of their area falling within the polygon. Because the experimental plots were relatively small compared with the 10 m Sentinel-2 grid, restricting the extraction to pixels completely contained within the polygons would have reduced the number of available pixels per plot. The internal buffer was therefore used to reduce edge effects, while partially intersecting pixels were retained to preserve sufficient spatial information. The number of valid NDVI pixels remaining after cloud and shadow masking was also recorded for each plot and acquisition date. This variable was retained as an indicator of image availability and the amount of valid spectral information contributing to each plot-level observation.

The number of usable Sentinel-2 acquisition dates varied among study years because of differences in satellite-image availability, cloud cover, atmospheric conditions, and the number of observations retained after cloud and shadow masking. The final number of acquisition dates available for each harvest year and experimental block is reported in Table 4.

Table 4. Availability of Sentinel-2 acquisition dates for winter wheat plots after cloud and shadow masking.

Year	Block	No. of plots	Available acquisition dates
------	-------	--------------	-----------------------------

2020	C	9	32
2022	C	9	28
2023	B	9	23
2025	A, B, C	27	26

2.3.4 Sentinel-2 data processing and feature aggregation

The CSV files exported from GEE were imported into R (R Core Team 2026). Each Sentinel-2 observation was associated with the corresponding harvest year and plot identifier. A common plot-year identifier was therefore used to link the satellite observations with the plot-level agronomic and yield datasets. Spatial statistics for each plot and acquisition date had already been calculated in GEE. Temporal aggregation in R involved summarising these plot-level observations across acquisition dates within predefined periods of the growing season. Before temporal aggregation, data availability was assessed by summarising the acquisition dates, the number of observations available per year, and the number of plots represented in each study year. The number of available images and the mean number of valid pixels were also calculated for each plot-year combination and retained as data-availability and quality-control indicators. These variables were not used as predictors.

Sentinel-2 spectral bands and vegetation indices were retained as candidate numerical predictors. Two groups of Sentinel-2 features were subsequently generated: seasonal predictors and crop-stage predictors. Seasonal predictors were calculated for each plot-year using Sentinel-2 observations acquired from 1 April (included) up to, but excluding, the harvest date of the corresponding year. For each spectral band and vegetation index, the temporal mean, maximum, minimum, and standard deviation were calculated across the available acquisition dates within this period for each plot and year.

Crop-stage predictors were generated separately. The plot-level spatial. Statistics calculated in GEE, including the mean, maximum, minimum, median standard deviation, 25th percentile and the 75th percentile, were aggregated in R across acquisition dates within the predefined crop development stages: vegetative growth (April–May) and grain filling (June–August, up to the harvest date). For each spatial statistic, temporal statistics, including the mean, maximum, minimum, median and standard deviation were calculated across the acquisition dates belonging to each stage. The resulting seasonal and crop-stage variables constituted the Sentinel-2 candidate predictor set used in the subsequent predictor selection procedures described in Section 2.6.4.

2.3.5 Soil data

Soil data from the 2024 survey were provided by RISE. According to the available laboratory report “Markkartering 2024 varje småruta”, one laboratory sample was reported for each experimental plot and analysed by Agrilab AB. The reported analyses included pH, AL-extractable phosphorus (P-AL),

potassium (K-AL), magnesium (Mg-AL), calcium (Ca-AL), aluminium (Al-AL) and iron (Fe-AL); HCL-extractable potassium (K-HCL) and copper (Cu-HCL); as well as soil organic matter, clay, silt and sand contents, and soil texture class. The available report did not provide further information on the field sampling procedure, including sampling depth or spatial distribution of subsamples used for soil texture and soil organic matter determination, or the solution used for pH measurement. No separate sampling protocol or reference containing these methodological details was available, and they could therefore not be documented further.

The soil data were used to characterise the experimental site and to describe spatial variability within and among the experimental blocks. Clay content and soil organic matter were additionally evaluated in separate linear mixed-effects models, as covariates of grain yield to assess whether variation in these soil properties was associated with yield after accounting for fertiliser treatment and year.

Soil variables were not included as predictors with the Sentinel-2 and meteorological yield-prediction models.

In addition to the soil properties obtained from the 2024 soil survey, data from soil sensors installed within the testbed were also evaluated as a potential source of plot-level information. However, inspection of the sensor records revealed limitations, as they have not been installed at consistent locations across years, preventing the construction of a continuous and spatially comparable time series. The soil sensor data were therefore excluded from subsequent analyses.

2.3.6 Agronomic data

Plot-level agronomic data were provided by RISE for the harvest years 2020, 2021, 2022, 2023 and 2025. Data from 2021 were excluded from subsequent statistical and predictive analyses because the experimental design differed from that used in the other study years. The dataset used for predictive modelling therefore comprised the harvest years 2020, 2022, 2023 and 2025. The field experiments analysed in this study were conducted in Block C in 2020, Block C in 2022, Block B in 2023, and Blocks A, B and C in 2025.

The agronomic dataset included winter wheat grain yield, grain quality, fertilisation records and information on field management operations for each experimental plot. Grain yield was reported in kg ha⁻¹ and adjusted to a standard grain moisture content of 15%. Grain crude protein and starch concentrations were expressed as percentages of dry matter. The available historical records did not document who was responsible for harvesting the grain samples prior to 2025. According to information provided by Lövsta farm, grain quality analyses in 2025 were performed using near-infrared transmittance (NIT). A nitrogen-to-protein conversion factor of 5.7 was used for crude protein. Further information on the analytical instrument, calibration procedure, sample preparation and laboratory protocol was not available. The analytical methods used for grain quality measurements in the earlier study years were also not documented in the available records.

Agronomic records were used to characterise differences in fertiliser management among plots and to support the interpretation of spatial and temporal variability in grain yield and crude protein. The statistical analysis of fertiliser treatment was restricted to the comparable years 2020, 2022 and 2023. The 2025 data were not included in this analysis because the fertilisation scheme was not directly comparable in the three treatment levels used in the earlier experimental years. Treatment was defined according to the number of fertiliser applications, comprising three levels: no fertiliser application (0), two fertiliser applications (2), and three fertiliser applications (3). These treatment levels were used to evaluate the effects of fertiliser management and its interaction with year on grain yield and crude protein concentration.

Detailed fertilisation records were available for the study years, although the level of information differed among years. For 2020–2023, the available records reported application dates and fertiliser application rates, but the fertiliser product and its nutrient composition were not specified.

Consequently, the corresponding N, P and K application rates could not be reliably calculated for these years. In 2020 plots 1–3 had no recorded fertiliser applications, plots 4–6 received two applications, and plots 7–9 received three applications. In 2020, 450 kg ha⁻¹ of fertiliser product was applied on 8 April, followed by 220 kg ha⁻¹ on 12 May. Plots assigned to the three-application treatment received an additional 390 kg ha⁻¹ on 11 June. In 2021, the corresponding application rates were 450 kg ha⁻¹ on 7 April, 220 kg ha⁻¹ on 31 May and, for plots 7–9, an additional 220 kg ha⁻¹ on 27 June. In 2022, 450 kg ha⁻¹ was applied on 21 April and 220 kg ha⁻¹ on 12 May, followed by an additional application of 220 kg ha⁻¹ on 17 June in plots 7–9.

In 2025, the fertiliser product was identified as Axan NS 2. Applications of 444 kg ha⁻¹ on the 3 April and 167 kg ha⁻¹ on 4 June were recorded. Based on the reported nitrogen concentration of 27% N, these corresponded to approximately 120 and 45 kg N ha⁻¹ respectively. No phosphorus or potassium was supplied by this product.

2.4 Meteorological data

Meteorological stations installed within the testbed were initially evaluated as potential sources of weather data. Their records were inspected for temporal coverage, continuity and data quality. However, inconsistencies were identified that prevented their use as reliable predictors in the subsequent analyses.

Consequently, meteorological data were obtained from the SLU Ultuna meteorological station (59°49'03" N, 17°39'14" E), located approximately 500 m from the experimental area. The station is operated by SLU and is part of a long-term meteorological monitoring programme at Ultuna. Its proximity to the experimental area provided a source of local meteorological data for the study.

Daily meteorological records were collected for the study years and included mean, minimum and maximum air temperature (°C), precipitation (mm), relative humidity (%), solar radiation (MJ m⁻² day⁻¹), wind speed (m s⁻¹), reference evapotranspiration (ET₀; mm day⁻¹), and VPD (kPa).

Weather variables were summarized for the period from April to harvest and for calendar-based periods corresponding approximately to major phases of the winter wheat growing cycle. These periods were defined as establishment (September-November), overwintering (December-February), vegetative (March-May) and grain filling (June-August). These periods were based on calendar months rather than field observations of phenological stage and should therefore be regarded as approximations of the corresponding crop phases. To associate autumn and early winter meteorological conditions with the corresponding crop, observations from September to December were assigned to the following harvest year. Because meteorological records for 2019 were unavailable, the establishment period could not be represented for the 2020 crop year. The overwintering period for the 2020 crop was represented only by January and February 2020, because December 2019 was unavailable.

Additional agroclimatic variables were derived from the daily observations, including growing degree days (GDD), climatic water balance, frost days, severe-frost days, and absolute minimum and maximum air temperatures. GDD were calculated from daily air temperature using a base temperature of 5 °C (GDD_base5), and daily values were summed within each analysis period to obtain cumulative GDD (GDD_sum_base5). GDD is commonly used to represent thermal accumulation in relation to crop development (McMaster & Wilhelm 1997). Daily climatic water balance was calculated as the difference between daily precipitation and reference evapotranspiration ($P - ET_0$, mm) (Allen et al. 1998). Daily water balance was subsequently summed within each analysis period to obtain the cumulative water balance. Frost days were defined as days with a minimum air temperature below 0 °C, while severe frost days were defined as days with minimum temperature below -10 °C (Zhang et al. 2011).

To characterise both the magnitude and temporal variability of weather conditions, the meteorological variables were summarised using means, sums, absolute extremes and standard deviations, depending on the type of variable. Mean values were calculated for air temperature, VPD, relative humidity, solar radiation and wind speed, whereas cumulative values were calculated for precipitation, ET_0 , GDD and water balance. Minimum and maximum air temperatures were also retained to describe extreme temperatures.

2.5 UAV data

RGB and multispectral UAV imagery were acquired over the testbed on 7 April, 8 May, 12 June and 21 July 2025. These image collections had not previously been processed or evaluated as orthomosaics and were therefore inspected and processed as part of the present study. Image coverage and processing quality varied among acquisition dates. Some acquisitions did not provide complete coverage of the experimental testbed, resulting in gaps in the reconstructed imagery, while for other acquisitions a complete orthomosaic could not be generated. Consequently, only the RGB imagery acquired on 8

May provided sufficiently complete coverage for the plot-level comparison performed in this study.

The RGB images acquired on 8 May were uploaded to the Solvi platform (Solvi AB, Gothenburg, Sweden), where an orthomosaic was generated and the VARI index was calculated. The UAV images contained positional information from the onboard GNSS, which allowed them to be approximately georeferenced during orthomosaic generation. This georeferencing allowed the resulting orthomosaic to be spatially aligned with the experimental plot shapefile in QGIS and used for plot-level extraction. However, RTK correction or ground control points were not available during the flight campaign, so positional accuracy was lower than expected for an RTK- or ground-control-based workflow. A reflectance calibration panel was also not available during image acquisition.

The resulting VARI orthomosaic was then imported into QGIS and overlaid with the experimental plot boundaries. Zonal statistics were calculated for each plot to summarize the VARI values within the plot area, including the mean, median, minimum, maximum and standard deviation. Mean UAV-derived VARI values were subsequently used as the plot-level value for comparison with Sentinel-2 derived VARI, and the resulting dataset was analysed in R.

2.6 Statistical and predictive analyses

2.6.1 Statistical analysis of agronomic response variables

Linear mixed-effects models were used to assess the effects of fertiliser treatment and harvest year on winter wheat grain yield and crude protein concentration. Treatment and year were included as fixed factors, together with their interaction, while plot was included as a random effect to account for repeated observations from the same experimental plots. The models were fitted using the 'lmer' function in R. These analyses were based on the comparable treatment data available for 2020, 2022 and 2023.

Fertiliser treatments were not randomly assigned to plots within the experimental blocks but followed a fixed spatial arrangement, with plots 1–3 receiving no fertilisation application, plots 4–6 receiving two applications, and 7–9 receiving three applications. Consequently, some of the observed differences among treatments could also reflect spatial differences among plots, such as variation in soil conditions, drainage or compaction. Furthermore, if plots repeatedly receiving no fertilisation, continued nutrient removal through crop harvest without the replacement of nutrients through fertilisation could have progressively reduced soil nutrient availability over time. Consequently, differences in grain yield and crude protein concentration may reflect not only fertilisation applied during the current growing season but also cumulative effect of previous nutrient management. Therefore, the observed differences cannot be attributed only to fertilisation, and the analysis was used mainly to describe treatment-related variation within the historical dataset rather than to establish a strictly causal fertiliser effect.

The significance of treatment, year and the Treatment $\times$ Year interaction was assessed using Type III analysis of variance (ANOVA), with denominator

degrees of freedom estimated using the Kenward–Roger method. Statistical significance was evaluated at $\alpha = 0.05$. Estimated marginal means (EMMs) were subsequently calculated using the `emmeans` package. Pairwise comparisons were adjusted for multiple testing using the Tukey method. When the Treatment $\times$ Year interaction was not significant, treatment and year effects were evaluated using marginal means averaged over the levels of the other factor. When the interaction was significant, treatment effects were examined separately within each year.

2.6.2 Modelling datasets and preprocessing

Candidate meteorological variables for predictive modelling were derived from the April to harvest period and from the crop development periods described in Section 2.4. Variables with insufficient observations or no variability were removed before predictor selection as a quality control step.

Separate predictive models were developed for three response variables: winter wheat grain yield, crude protein concentration and starch concentration. The modelling dataset included winter wheat observations from the 2020, 2022, 2023 and 2025 harvest years. The 2021 season was excluded from the mixed-effects analyses and the predictive modelling because its experimental design differed from that of the comparable study years. Year 2024 was not included because winter wheat was not cultivated in the experimental blocks during that year.

The relationship between crude protein and starch concentration was additionally examined using simple linear regression. Only observations for which both crude protein and starch measurements were available were included. Data from 2023 were excluded because starch concentration data were not available in the RISE project files reviewed for this study. The available documentation did not indicate whether these measurements had not been performed or had simply not been recorded in the available files. The regression slope (β) indicates the average increase in the response variable associated with a one-unit increase in the predictor.

Each observation represented one plot-year combination and was identified by harvest year and plot ID. The final datasets contained 54 observations for grain yield and 48 observations for both crude protein and starch concentration.

Missing data were assessed separately for each response variable using complete-case analysis. Complete-case analysis was specified as the procedure for handling missing values. However, no observations were removed from the final modelling datasets because of missing values.

2.6.3 Predictor selection and predictive modelling

Two alternative predictor-selection approaches were evaluated independently: statistical and agronomic screening approach and recursive feature elimination (RFE). Candidate predictors comprised Sentinel-2 variables and meteorological variables derived from the temporal aggregation periods described above.

In the statistical and agronomic screening approach, pairwise Pearson correlation matrices were examined separately for the Sentinel-2 and meteorological candidate predictors. The Sentinel-2 candidate set comprised 164 predictors, corresponding to 13,366 pairwise correlations, whereas the meteorological candidate set comprised 20 predictors and 190 pairwise correlations. The meteorological predictor set showed greater intercorrelation and redundancy than the Sentinel-2 predictor set. The median absolute pairwise correlation was $|r| = 0.81$ for meteorological predictors, compared with $|r| = 0.55$ for Sentinel-2 predictors. Based on this greater observed redundancy, a more stringent correlation threshold was applied to meteorological predictors ($|r| \geq 0.85$), whereas a threshold of $|r| \geq 0.90$ was used for Sentinel-2 predictors to reduce multicollinearity while retaining potentially complementary spectral information. When pairs of predictors exceeded the specified correlation threshold, variables were further evaluated based on their relationship with the response variable and their agronomic relevance before the final predictor set was retained.

RFE was applied independently of the statistical and agronomic screening approach. It was first performed separately for Sentinel-2 and meteorological predictors to identify reduced predictor subsets within each data source. Random forest was used within the RFE procedure to rank predictor importance, and candidate predictor subsets were evaluated using five-fold cross-validation repeated five times. For grain yield, an additional RFE analysis was performed on the combined set of Sentinel-2 and meteorological candidate predictors to assess whether joint selection across the two data sources resulted in a different predictor subset.

Three regression methods were compared for predicting winter wheat grain yield, crude protein concentration and starch concentration: multiple linear regression (MLR), random forest (RF) and support vector machine regression (SVM). These methods were selected to compare a simple linear approach with two machine-learning methods capable of modelling more complex relationships between predictors and crop responses. Models were fitted separately using the predictor sets obtained from alternative predictor-selection approaches.

2.6.4 Cross-validation and evaluation metrics

Model performance was evaluated using leave-one-out cross-validation (LOOCV) at the plot-year level. In each iteration, one observation was excluded, the model was fitted using all remaining observations, and a prediction was generated for the excluded observation. This procedure was repeated until every observation had been used once for validation.

The predictor sets obtained from the alternative feature-selection approaches were evaluated separately using the same model-validation framework. RFE was performed before the final leave-one-out cross-validation procedure and was not repeated within each validation iteration.

For the MLR, RF and SVM predictive models, continuous predictors were centred by subtracting their training mean and scaled by dividing by their standard deviation. Within each LOOCV iteration, the preprocessing

parameters were estimated using only the training and subsequently applied to the corresponding held-out observation. This prevented information from the validation observation from influencing predictor preprocessing.

The validation was primarily intended to evaluate the ability of the models to reproduce plot-level variability across the years and environmental conditions represented in the dataset.

Model performance was assessed using root mean square error (RMSE), predictive coefficient of determination (R^2) and bias, as defined in Equations 1–3. RMSE was expressed in kg ha^{-1} for grain yield and in percentage points of dry matter for crude protein and starch concentrations.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2}$$

Equation 1

$$R^2 = 1 - \frac{\sum (\gamma_i - \hat{\gamma}_i)^2}{\sum (\gamma_i - \bar{\gamma})^2}$$

Equation 2

$$Bias = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)$$

Equation 3

2.6.5 UAV–Sentinel-2 comparison

UAV-derived and Sentinel-2-derived VARI were compared at plot level to assess whether the two platforms captured similar spatial patterns in crop greenness. Mean UAV-derived VARI values from 8 May 2025, obtained from the zonal statistics were matched with the corresponding mean Sentinel-2-derived VARI values from 18 May 2025 using the unique plot identifier (id_cell). Observations with missing values in either dataset was excluded from the comparison.

The relationship between UAV- and Sentinel-2-derived VARI was evaluated descriptively using scatterplots, with experimental blocks A, B and C distinguished by colour. A linear trend was fitted across all available plots to describe the overall direction of the relationship, and the regression slope (β) was used as a descriptive measure of the change in UAV-derived VARI associated with Sentinel-2-derived VARI. Because the two images were acquired 10 days apart and differed substantially in spatial resolution and acquisition characteristics, the analysis was intended as a descriptive comparison of plot-level spatial patterns rather than a direct validation of one platform against the other.

2.7 External validation

External validation data were obtained from the SITES Agroecological field Experiment (SAFE) at Lönnerstorp Research Station, Skåne, Sweden (Barreiro & Albertsson 2022). The Lönnerstorp Research Station is operated by the Swedish University of Agricultural Sciences (SLU) and SITES is a national research infrastructure supporting long-term field-based ecosystem research and environmental monitoring across nine research stations in Sweden (Barreiro & Albertsson 2022). Established in 2015, SAFE is a long-term agricultural field experiment designed to investigate different cropping systems and their environmental and agronomic performance.

The experiment comprises four cropping systems: conventional reference, organic, agroecological intensification and perennial cereal production, with replicates across four experimental blocks. The conventional reference system follows a conventional four-year crop rotation consisting of spring barley, winter oilseed rape, winter wheat and sugar beet, representative of crops presents in southern Sweden. The dataset evaluated in this study comprised the conventional reference cropping system. Available records were screened to identify plots cultivated with winter wheat between 2018 and 2023 (Table 5). Three plots were selected for each year, resulting in an independent dataset of 18 observations (Figure 3). Grain yields and harvest dates were obtained from SAFE records. Because crops are rotated among plots within the experimental system, the plots differed among years. The spatial boundaries of the selected plots were used to extract Sentinel-2 data for each growing season. Satellite data were processed using the same cloud and shadow masking, spatial statistics and crop-stage aggregation procedure applied to the testbed dataset in Ultuna. Meteorological variables were similarly derived using the same definitions and crop-stage periods used for the model development. The resulting dataset was used as an independent external dataset to evaluate the previously developed winter wheat grain yield prediction model. This validation allowed the assessment of model transferability in different environmental and management conditions. As grain-quality measurements were not available for the selected SAFE observations, external validation was restricted to the grain yield model.

Table 5. Summary of the independent Lönnstorp dataset used for external validation of the yield prediction model.

Year	Plot IDs	n	Harvest date	Yield range (kg ha ⁻¹)
2018	95, 72, 28	3	2018-07-25	4,680–7,240
2019	98, 75, 31	3	2019-08-15	8,831–8,826
2020	97, 74, 30	3	2020-08-06	7,757–9,659
2021	96, 73, 29	3	2021-08-16	8,060–9,121
2022	95, 72, 28	3	2022-08-11	8,482–9,642
2023	31, 75, 98	3	2023-08-05	6,175–7,942

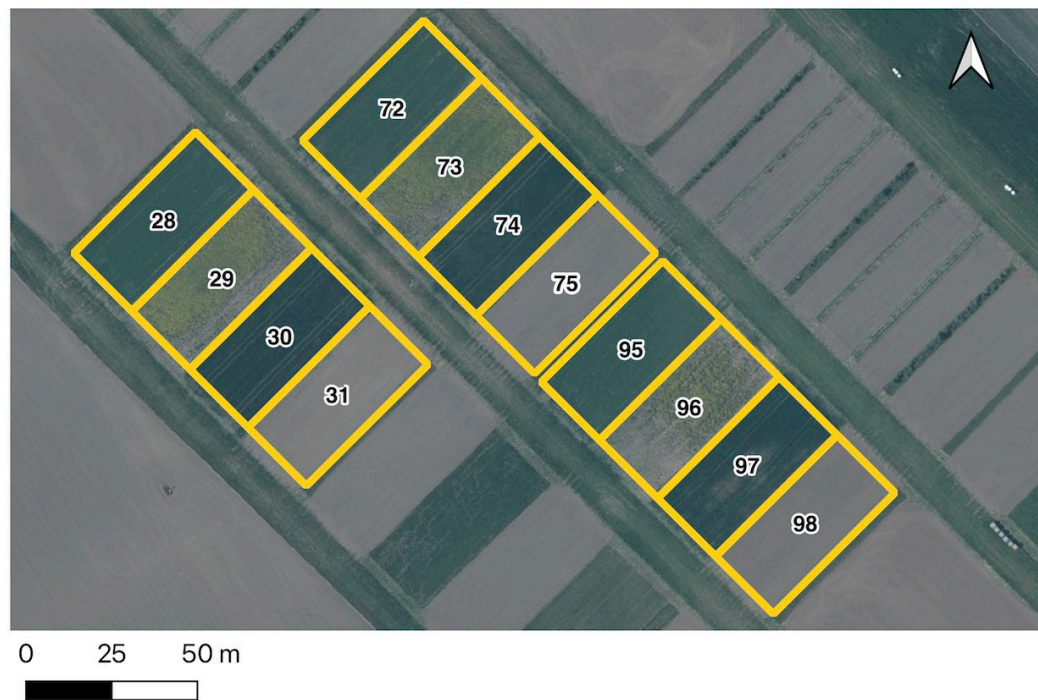


Figure 3. Location and layout of the external validation plots at the Lönnstorp experimental field. Background: RGBI orthophoto ©Lantmäteriet (2025).

3. Results

3.1 Variability of the response variables

3.1.1 Grain yield

Grain yield varied among years and experimental plots (Figure 4). The linear mixed-effects model showed significant effects on both fertiliser treatment ($F = 16.43$, $p = 0.0037$) and year ($F = 71.45$, $p < 0.001$), whereas the Treatment $\times$ Year interaction was not significant ($F = 0.95$, $p = 0.467$; Table 6) indicating that the effect of fertilisation on grain yield did not differ significantly among the analysed years.

Estimated marginal mean grain yield increased from 3,893 kg ha⁻¹ in the untreated plots to 6,372 kg ha⁻¹ and 7,055 kg ha⁻¹ in plots receiving two and three fertiliser applications, respectively.

Tukey-adjusted pairwise comparisons showed that the untreated treatment had significantly lower grain yield than both the two-application treatment ($p = 0.0125$) and the three-application treatment ($p = 0.0038$), whereas the two- and three-application treatments did not differ significantly ($p = 0.508$; Figure 5).

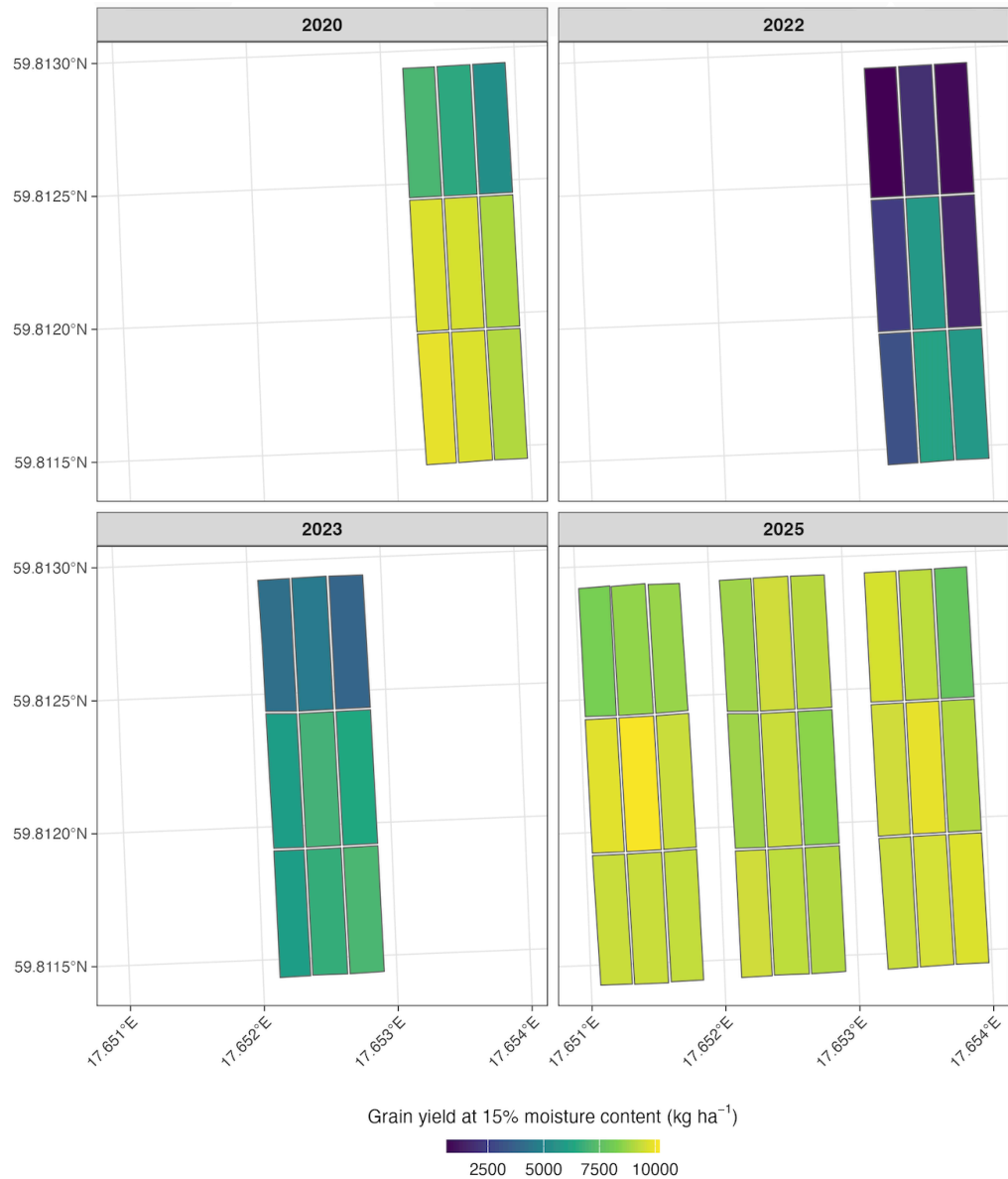


Figure 4. Plot-level yield maps for the available harvest years. Each polygon represents one experimental plot and is coloured according to measured grain yield (kg ha⁻¹). The maps show spatial variation in grain yield among plots and differences in grain yield levels among years.

Table 6. Type III ANOVA results from the linear mixed-effects model for winter wheat grain yield. Num df = numerator degrees of freedom; Den df = denominator degrees of freedom using the Kenward-Roger method; F = F-statistic used to test the significance of each fixed effect.

Effect	Num df	Den df	F	p-value
Treatment	2	6	16.43	0.0037
Year	2	12	71.45	<0.001
Treatment × Year	4	12	0.95	0.467

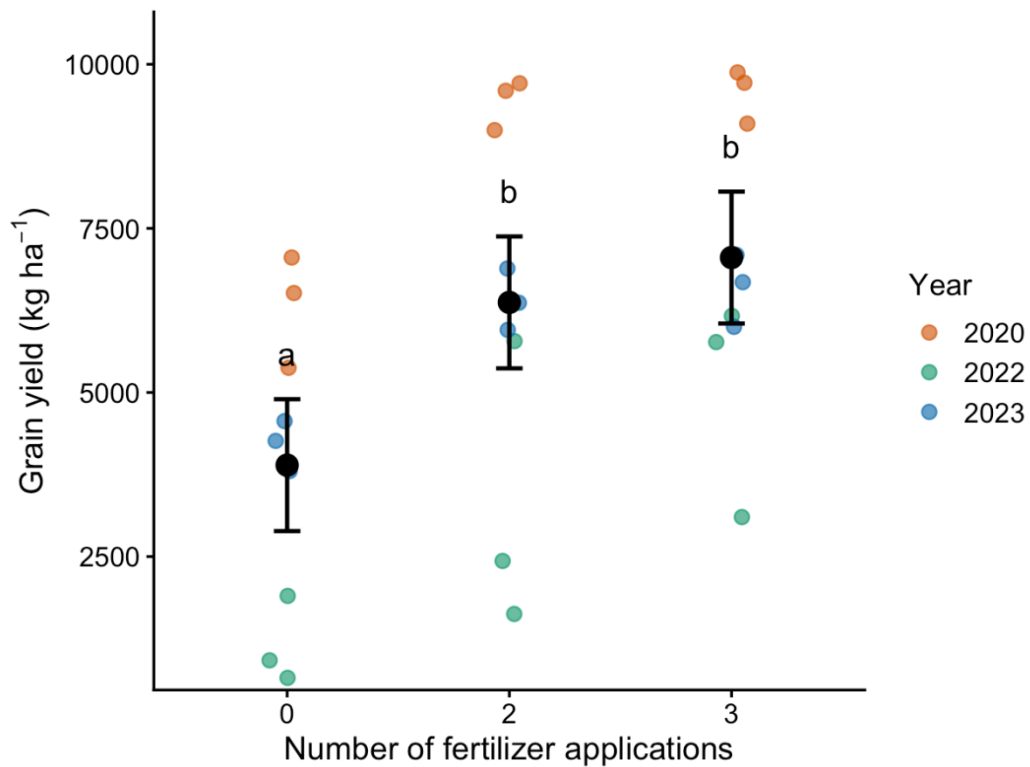


Figure 5. Grain yield under three application treatments in 2020, 2022 and 2023. Coloured points represent individual plot observations. Black points represent estimated marginal means from the linear mixed-effects model, averaged across years, and error bars indicate 95% confidence intervals. Different letters indicate significant differences among treatments according to Tukey-adjusted pairwise comparisons ($p < 0.05$).

3.1.2 Crude protein concentration

Crude protein concentration was significantly affected by the Treatment $\times$ Year interaction ($F = 17.54$, $p < 0.001$; Table 7), indicating that the effect of fertiliser treatment differed among years. Consequently, treatment effects were examined separately within each year using estimated marginal means and Tukey-adjusted pairwise comparisons. The estimated marginal means for each treatment and year are shown in Figure 6, with different letters indicating significant differences among fertiliser treatments within each year ($p < 0.05$).

In 2020, crude protein concentration increased progressively with the number of fertiliser applications, with significant differences among all three treatment levels. In 2022, crude protein concentration was significantly lower in the untreated treatment than in both fertilised treatments, whereas the two- and three-application treatments did not differ. A similar pattern was observed in 2023, where the untreated treatment differed significantly from both fertilised treatments, while no significant difference was detected between two and three applications.

Table 7. Type III ANOVA results from the linear mixed-effects model for crude protein concentration. Num df = numerator degrees of freedom; Den df = denominator degrees of freedom using the Kenward-Roger method. F = F-statistic used to test the significance of each fixed effect.

Effect	Num df	Den df	F-value	p-value
Treatment	2	6	219.08	<0.001
Year	2	12	102.24	<0.001
Treatment $\times$ Year	4	12	17.54	<0.001

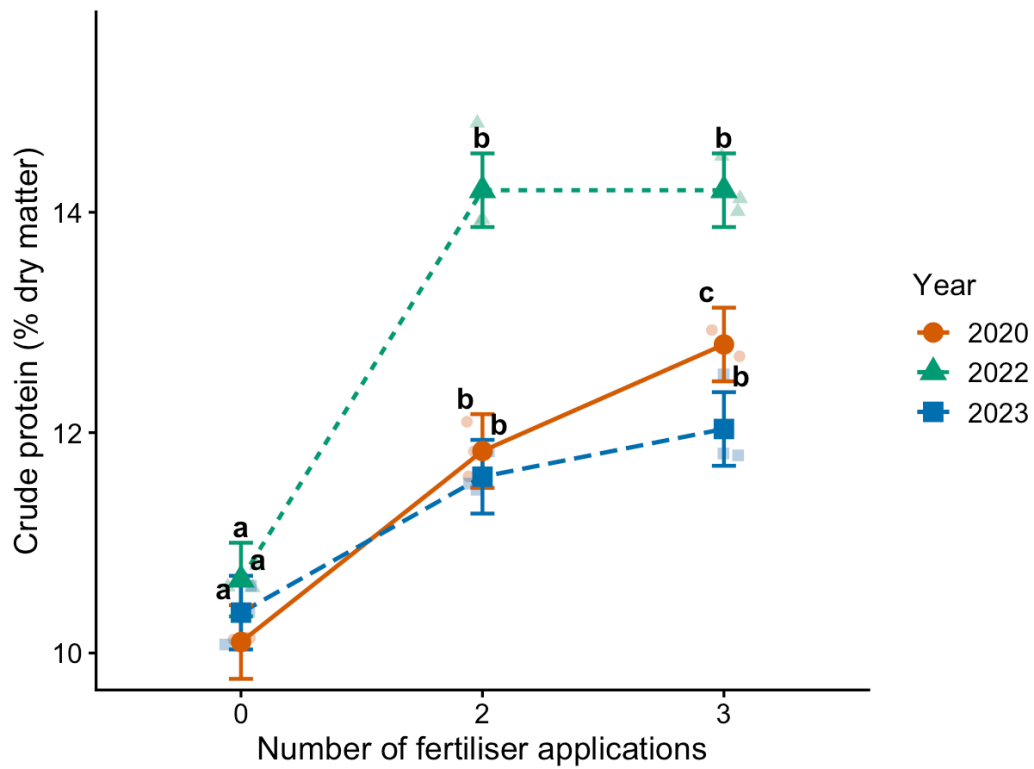


Figure 6. Estimated marginal means of winter wheat crude protein concentration (% dry matter) according to the number of fertiliser applications in 2020, 2022 and 2023. Error bars represent 95% confidence intervals. Different letters indicate significant differences among fertiliser treatments within each year based on Tukey-adjusted pairwise comparison ($p < 0.05$).

3.1.3 Relationship between crude protein and starch concentration

A negative relationship was observed between crude protein and starch concentration (Figure 7). Starch concentration decreased as crude protein concentration increased, with an estimated regression slope of $\beta = -1.05$.

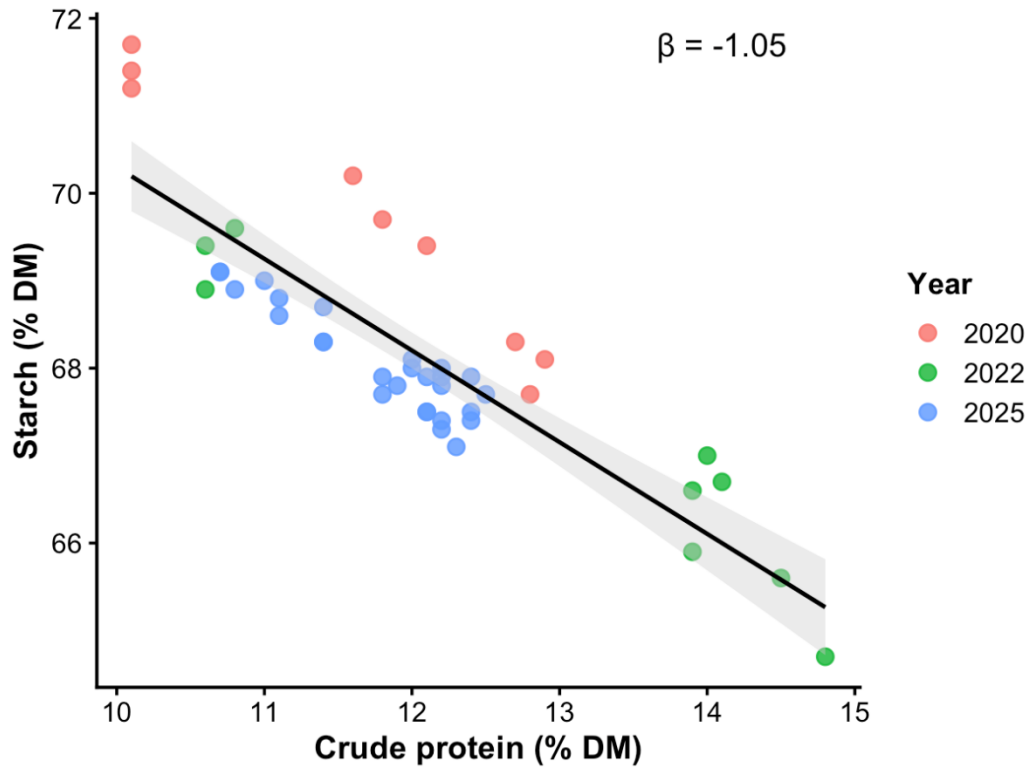


Figure 7. Relationship between crude protein and starch concentration in winter wheat grain. Points represent individual plot observations, colours indicate harvest year, and the black line represents the fitted linear regression with 95% confidence interval. β represents the estimated regression slope.

3.2 Soil properties of the experimental site

Soil properties measured in the 2024 survey are summarised in Table 8. Median values were used to summarise soil properties because some variables, particularly pH, showed asymmetric distributions and relatively extreme values. The median was therefore considered a more robust measure than the mean for describing soil conditions across the plots. The experimental site was characterised by a relatively high clay and silt content, with median values of 37% and 45%, respectively, while median sand content was 16%. Soil organic matter ranged from 1.0 to 2.8%, with a median of 2.0%, while soil pH ranged

from 6.3 to 8.0, with a median of 6.5. These ranges indicate spatial heterogeneity in soil conditions within the experimental site.

Historical documentation from 2015 indicated that part of the experimental site had previously been used for a liming experiment. However, detailed information on the treatment layout and application rates was unavailable. Therefore, it was not possible to determine whether previous liming treatments contributed to the observed spatial variation in soil pH.

Table 8. Summary of soil texture, organic matter and pH across the 27 experimental plots based on the 2024 soil survey.

Soil property	Median	Range	Unit
Clay content	37	33–45	%
Silt content	45	38–49	%
Sand content	16	11–24	%
Soil organic matter	2.0	1.0–2.8	%
Soil pH	6.5	6.3–8.0	—

3.3 Soil spatial variability and association with grain yield under different fertiliser treatments

Spatial variation in soil properties was observed across the experimental site (Figure 8 and Figure 9). Clay content ranged from 33 to 45% across the plots, with a median of 37%, and showed differences both within and among experimental blocks. Soil organic matter ranged from 1.0 to 2.8%, with a median of 2.0%, and displayed a heterogeneous spatial distribution across the site.

To assess whether the observed spatial variation in clay content and soil organic matter contributed to grain yield variability, each soil property was included separately as a covariate in the linear mixed-effects model. Neither clay content nor soil organic matter had a significant effect on grain yield ($p = 0.636$ and $p = 0.850$, respectively; Table 9). In both models, fertiliser treatment and year remained significant, whereas their interaction was not significant.

Figure 8. Spatial distribution of clay content (%) across the experimental plots based on the 2024 soil survey.



Figure 9. Spatial distribution of soil organic matter (%) across the experimental plots based on the 2024 soil survey.

Table 9. Type III ANOVA results from linear mixed-effects models evaluating clay content and soil organic matter as covariates of winter wheat grain yield. F = F -statistic used to test the significance of each fixed effect.

Model	Effect	F-value	p-value
Clay model	Treatment	11.86	0.009

Model	Effect	F-value	p-value
Soil organic matter model	Year	66.59	<0.001
	Clay	0.23	0.636
	Treatment × Year	0.73	0.586
	Treatment	12.76	0.008
	Year	68.68	<0.001
	Organic matter	0.04	0.850
	Treatment × Year	0.86	0.513

3.4 Environmental and spectral variability

3.4.1 Weather conditions during the growing seasons

Weather conditions were examined across all four crop development stages: establishment, overwintering, vegetative and grain filling. VPD varied among harvest years and crop stages (Figure 10). During the vegetative stage, VPD generally increased from early spring towards late May, although the magnitude of this increase differed among years. The most pronounced increase was observed in 2023, while VPD in 2025 remained relatively low during March and April before increasing in May. During grain filling, VPD was generally higher and more variable, with short-term peaks occurring at different times among the years. These patterns indicate interannual variation in the timing and intensity of atmospheric evaporative demand.

Daily air temperature across the full winter wheat crop year showed the expected seasonal progression from cooler conditions during establishment and overwintering to warmer conditions during vegetative growth and grain filling (Figure 11). The timing and magnitude of these temperature changes differed among harvest years, indicating contrasting thermal conditions during crop development. For the 2020 crop year, meteorological data prior to January 2020 were unavailable, preventing comparison of the establishment and early overwintering periods with the other years.

Cumulative climatic water balance also differed markedly among crop years (Figure 12). The timing and persistence of periods with positive or negative water balance varied among years, indicating differences in the balance between precipitation and atmospheric water demand throughout crop development.

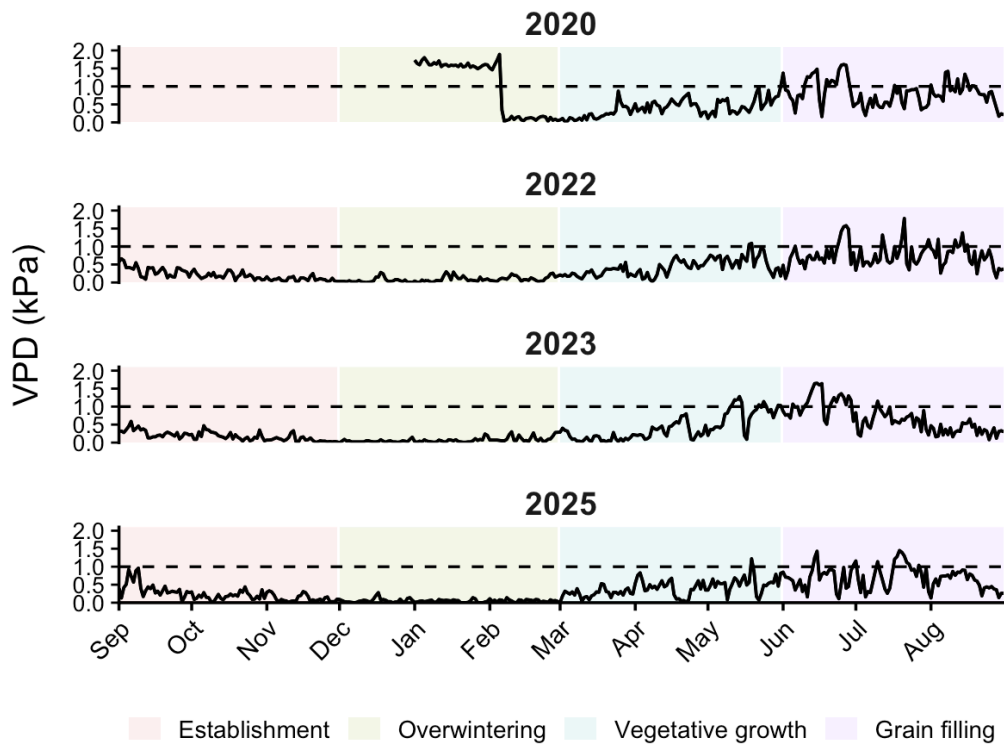


Figure 10. Seasonal variation in daily mean vapour pressure deficit (VPD) across selected winter wheat crop years. The black line represents daily mean VPD, while the shaded backgrounds indicate the crop-development stages: establishment, overwintering, vegetative growth and grain filling. Each panel represents one harvest year. The dashed horizontal line at 1 kPa is included as a visual reference to facilitate comparison among years. Data prior to January 2020 were unavailable.

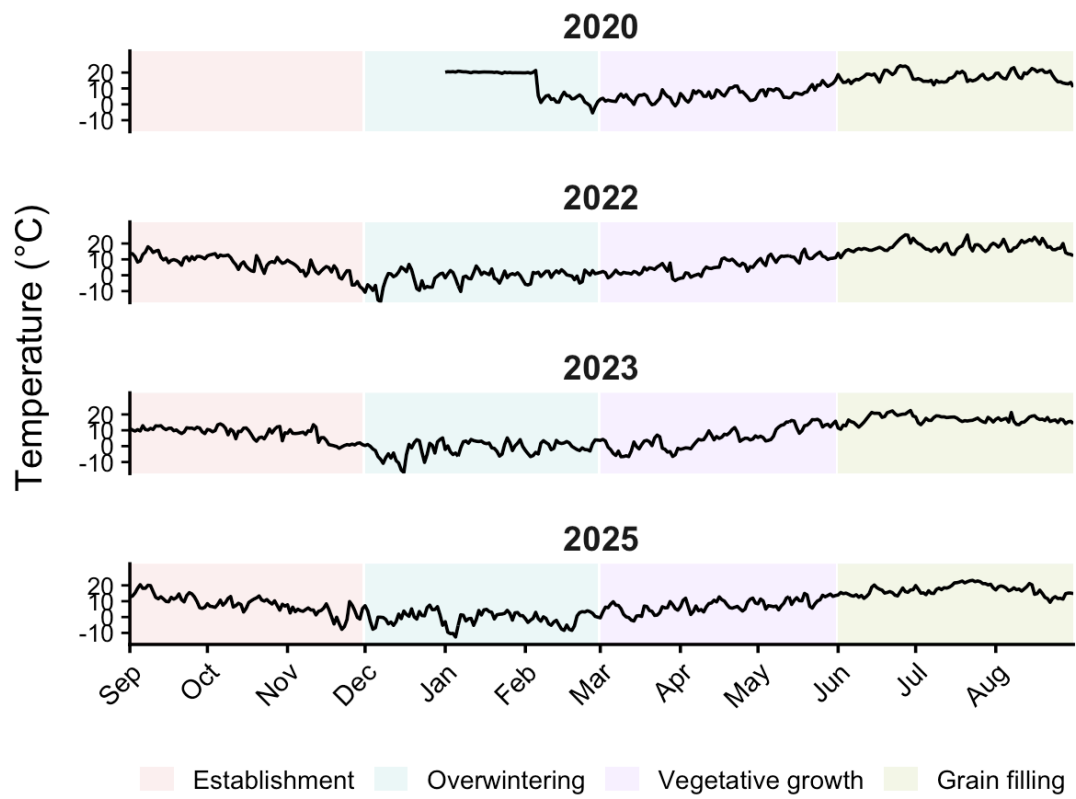


Figure 11. Daily mean air temperature during the winter wheat crop year for 2020, 2022, 2023 and 2025. Shaded areas indicate establishment, overwintering, vegetative growth and grain filling. Data prior to January 2020 were unavailable.

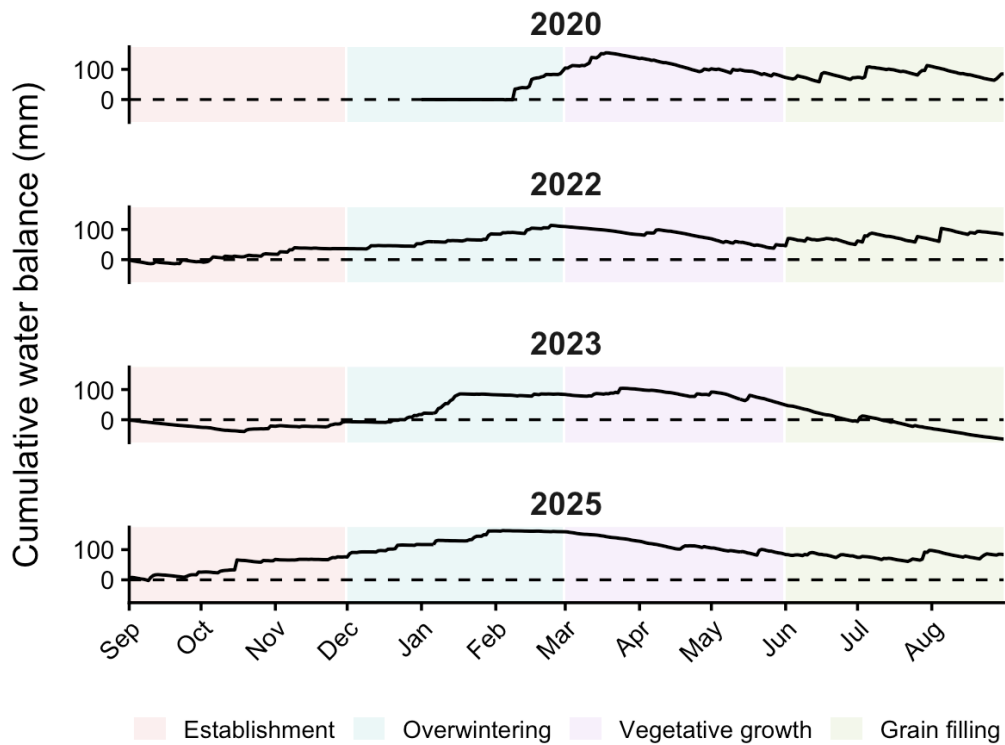


Figure 12. Cumulative climatic water balance across the winter wheat crop year for the harvest years 2020, 2022, 2023 and 2025. Background shading indicates establishment, overwintering, vegetative growth and grain filling. The dashed horizontal line represents a neutral water balance. Data prior to January 2020 were unavailable.

3.5 Sentinel-2 index variability across crop stages and years

Sentinel-2 index values varied among harvest years and crop stages (Figure 13). During the vegetative stage, NDRE and NDVI showed the lowest median values in 2022, while higher values were generally observed in 2021 and 2025. The 2021 Sentinel-2 observations were included in this descriptive analysis to provide temporal continuity in the assessment of spectral variability, although this year was not included in the predictive modelling because yield and grain-quality measurements were unavailable. During the vegetative stage, 2025 showed a markedly wider distribution of spectral index values than other harvest years, particularly for NDRE and NDVI, indicating greater variability in canopy conditions during this period.

During grain filling, NDRE and NDVI values were generally higher than during the vegetative stage, although differences among years remained evident. NDWI also varied among years, with relatively low values in 2022 and higher median values in 2020 and 2025.

Overall, the distributions indicate interannual and stage-dependent variability in the spectral response of winter wheat.

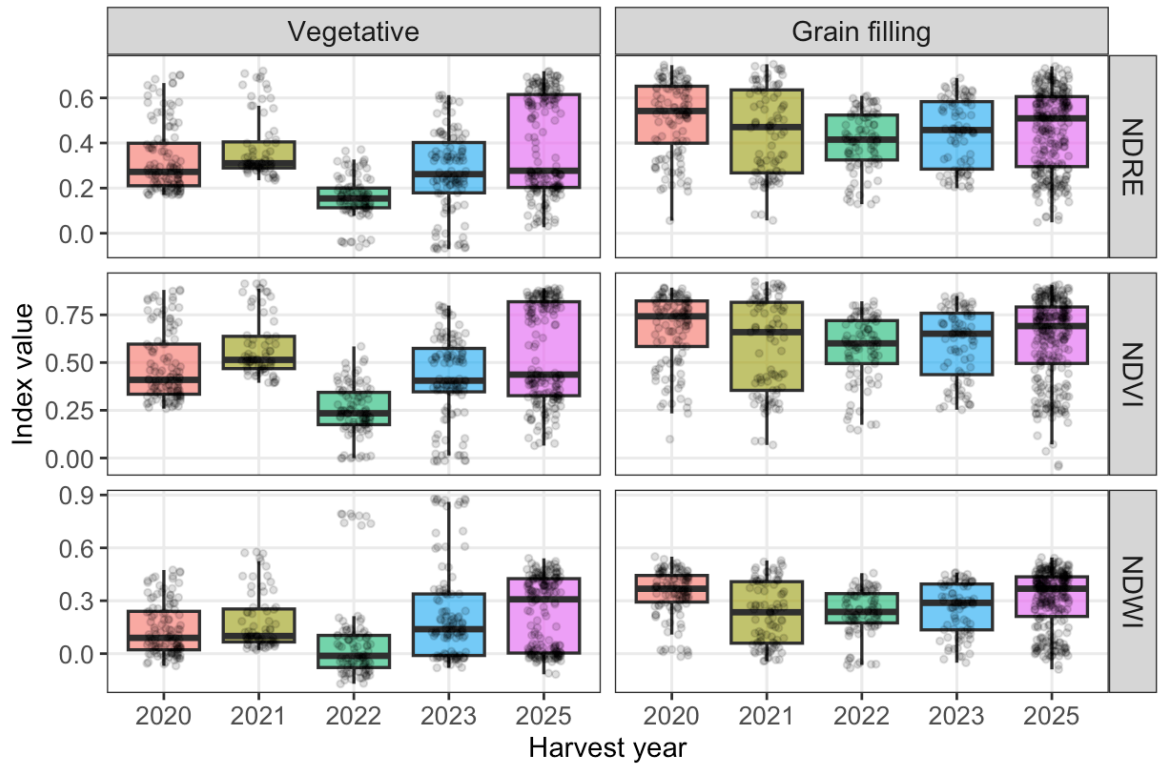


Figure 13. Sentinel-2 index distributions by crop stage and harvest year. Boxplots show the distribution of plot-level mean for NDRE, NDVI and NDWI during the vegetative and grain filling stages. Individual points represent the mean index value within one plot for one Sentinel-2 acquisition date.

3.6 Grain yield prediction

3.6.1 Predictors selected for grain yield

The initial screening procedure, based on correlation and collinearity screening combined with agronomic interpretation, retained four predictors for grain yield: two Sentinel-2 variables and two meteorological variables (Table 10). All selected predictors were derived from the vegetative stage. RFE retained a smaller set consisting of one Sentinel-2 predictor and one meteorological predictor, both also derived from the vegetative stage.

Table 10. Predictors selected for winter wheat grain yield using the initial screening and recursive feature elimination (RFE) strategies. Predictor names indicate the data source, summary statistic and crop stage. For Sentinel-2 predictors, the first statistic refers to the plot-level zonal statistic and the second to the temporal aggregation within the crop stage; weather predictor names indicate the meteorological summary and crop stage.

Selection strategy	Data source	Selected predictor
Initial screening	Sentinel-2	b4_min_min_vegetative
Initial screening	Sentinel-2	ndre_median_max_vegetative
Initial screening	Weather	frost_days_vegetative_
Initial screening	Weather	vpd_mean_vegetative
RFE	Sentinel-2	ndvi_max_max_vegetative
RFE	Weather	air temp mean vegetative

3.6.2 Initial predictor screening

Overall, the predictors retained through the initial correlation- and collinearity-based screening procedure indicated that conditions during the vegetative growth stage were particularly relevant for yield prediction. The selected predictor set included two Sentinel-2 variables, together with the number of frost days and mean vapour pressure deficit during the vegetative stage, highlighting the importance of both crop spectral response and weather conditions.

The three modelling approaches showed similar predictive performance (Table 11). SVM achieved the lowest root mean square error (RMSE = 1,105 kg ha⁻¹) and the highest R² (0.82), followed closely by MLR (RMSE = 1,114 kg ha⁻¹, R² = 0.82) and RF (RMSE = 1,133 kg ha⁻¹, R² = 0.81). Prediction bias ranged from -46 kg ha⁻¹ for SVM to +27 kg ha⁻¹ for RF, with MLR showing a bias of -6 kg ha⁻¹.

All three modelling approaches provided comparable yield predictions, although SVM showed marginally better predictive performance. The relationship between observed and LOOCV-predicted yield for the SVM model is shown in Figure 14. Observed-versus-predicted plots for MLR and RF are provided in Appendix Figures A7 and A8. Predicted values generally followed the 1:1 line. Larger deviations were observed for some observations in 2022, for which the model tended to overpredict low yields and underpredict intermediate yields. Overall, the predicted values captured the broad variation in yield, although prediction accuracy varied among individual observations.

Table 11. Predictive performance of multiple linear regression (MLR), support vector machine (SVM) and random forest (RF) models for winter wheat yield using leave-one-out cross-validation (LOOCV). RMSE. and bias are expressed in kg ha⁻¹. Bias was calculated as predicted minus observed yield

Model	RMSE (kg ha ⁻¹)	R ²	Bias (kg ha ⁻¹)
SVM	1,105	0.82	-46
MLR	1,114	0.82	-6
RF	1,133	0.81	+27

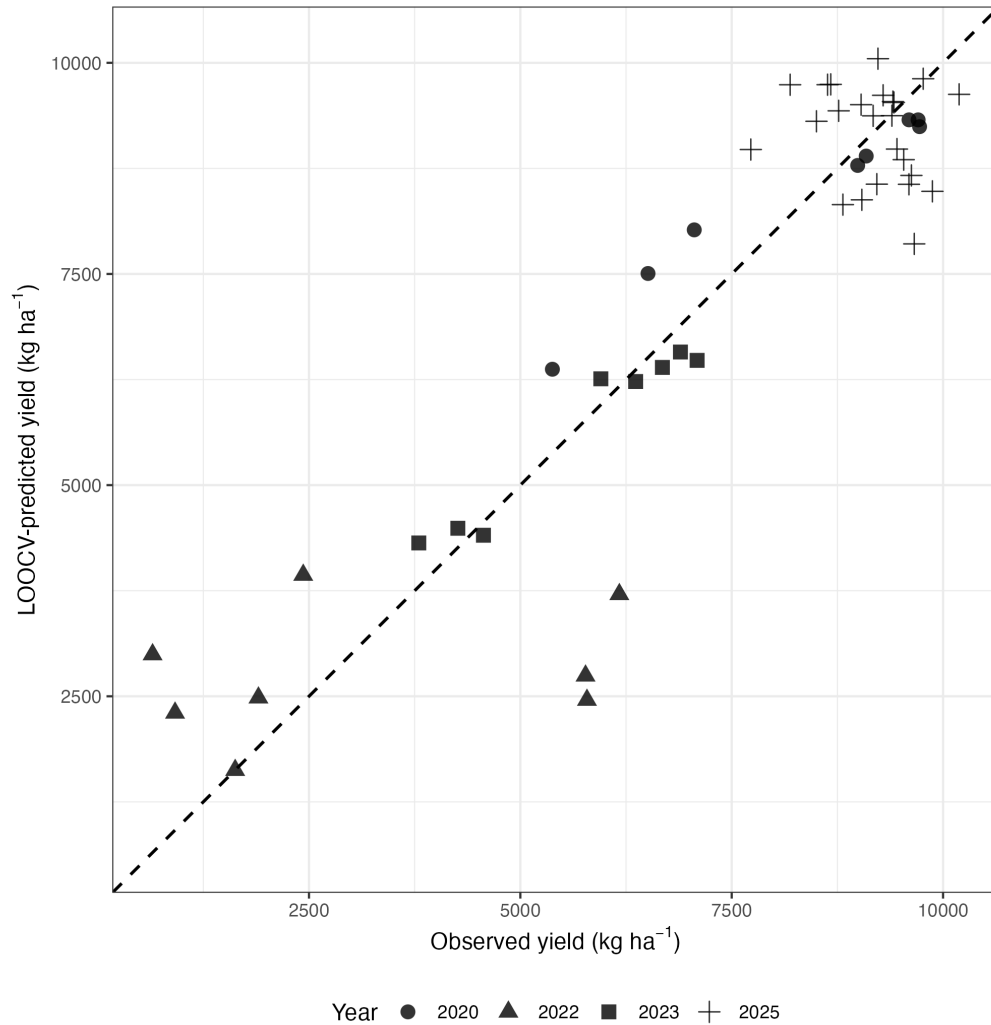


Figure 14. Relationship between observed and leave-one-out cross-validation (LOOCV)-predicted winter wheat yield obtained with the support vector machine (SVM) model. Each point represents one plot-year observation, with symbols indicating harvest year. The dashed 1:1 line represents perfect agreement between observed and predicted yield.

3.6.3 Recursive feature elimination

The SVM achieved the best predictive performance, with the lowest RMSE (957.0 kg ha⁻¹). The model showed a positive bias of 100 kg ha⁻¹, indicating a slight tendency to overestimate yield. RF showed intermediate predictive performance, with an RMSE of 1069.6 kg ha⁻¹ and an R² of 0.83. Multiple linear regression produced the least accurate predictions, with an RMSE of 1561.0 kg ha⁻¹ and an R² of 0.65. The LOOCV-predicted yields generally followed the observed values, although prediction accuracy differed among years. The largest deviations from the 1:1 line occurred in 2022, whereas the

2020 and 2023 observations were generally predicted more accurately. For the high-yielding 2025 observations, predictions were concentrated within a relatively narrow range, indicating reduced discrimination among plots at high yield levels. (Table 12; Figure 15). The corresponding observed versus predicted plots for the RF and MLR models are provided in Appendix Figures A1 and A2.

Table 12. Predictive performance of multiple linear regression, support vector machine and random forest models for winter wheat yield evaluated using leave-one-out cross-validation. RMSE and bias are expressed in kg ha^{-1} . Bias was calculated as predicted minus observed yield.

Model	RMSE (kg ha^{-1})	R ²	Bias (kg ha^{-1})
SVM	957.0	0.87	+100
RF	1069.6	0.83	+14
MLR	1561.0	0.64	-27

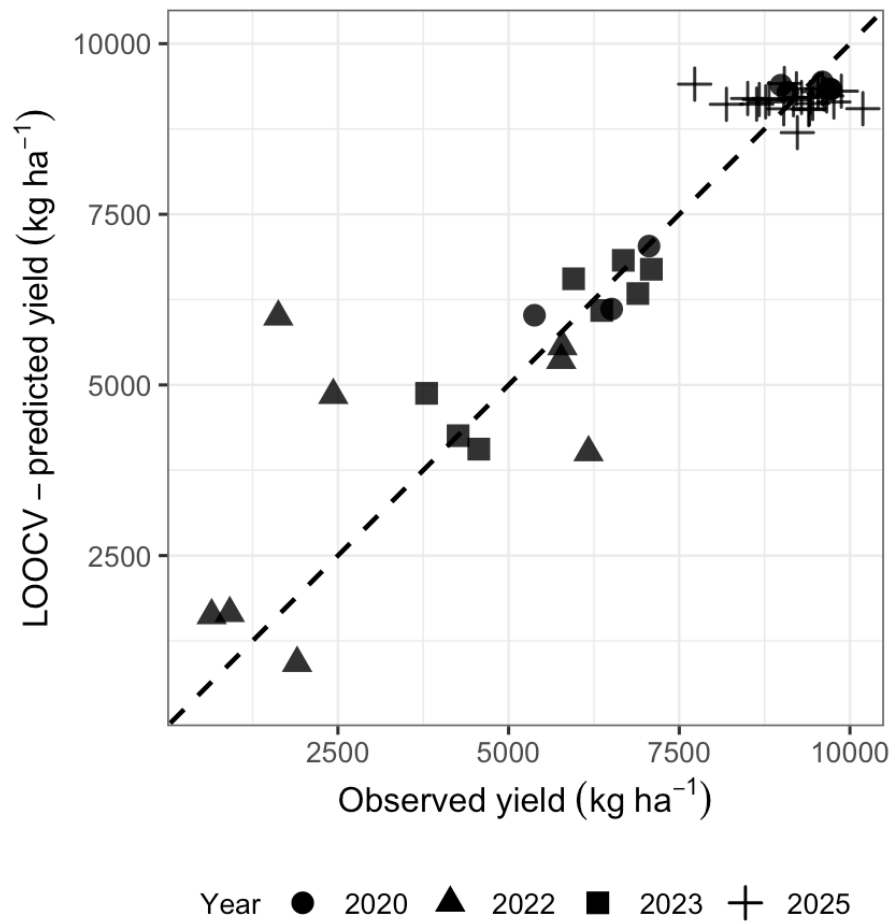


Figure 15. Relationship between observed and leave-one-out cross-validation (LOOCV)-predicted winter wheat yield for the support vector machine (SVM) model using satellite and meteorological predictors selected through recursive feature elimination (RFE).

Each point represents one plot-year observation, with symbols indicating harvest year. The dashed 1:1 line represents perfect agreement between observed and predicted yield.

3.6.4 Joint RFE of Sentinel-2 and meteorological predictors

As an additional analysis, recursive feature elimination was performed jointly on the Sentinel-2 and meteorological candidate predictors. The joint RFE retained two predictors, `ndvi_max_max_vegetative` and `ndvi_p75_max_grain_filling`, both derived from Sentinel-2 imagery, while no meteorological variables were selected. Using these predictors, RF achieved the best LOOCV performance, with an RMSE of 1,025 kg ha⁻¹ and an R² of 0.85, followed by MLR (RMSE = 1,108 kg ha⁻¹, R² = 0.82) and SVM (RMSE = 1,179 kg ha⁻¹, R² = 0.81; Table 13). Compared with the separate RFE approach, for which SVM achieved an RMSE of 957 kg ha⁻¹ and an R² of 0.87, joint predictor selection did not improve predictive performance.

Table 13. Leave-one-out cross-validated performance of models using predictors selected through joint RFE of Sentinel-2 and meteorological variables.

Model	RMSE (kg ha ⁻¹)	R ²	Bias (kg ha ⁻¹)
RF	1024.5	0.85	+65
MLR	1108.2	0.82	-11
SVM	1179.4	0.81	+171

3.7 Crude protein prediction results

3.7.1 Predictors selected for crude protein concentration

RFE selected seven predictors for crude protein concentration: six Sentinel-2 variables and one meteorological variable (Table 14). Five of the six satellite predictors were derived from the vegetative stage, while one NDWI predictor represented the grain-filling stage. The meteorological RFE retained the variability in relative humidity during vegetative growth.

Table 14. Predictors selected for crude protein concentration using the initial screening and recursive feature elimination (RFE). Predictor names indicate the data source, summary statistic and crop stage. For Sentinel-2 predictors, the first statistic refers to the plot-level zonal statistic and the second to the temporal aggregation within the crop stage; weather predictor names indicate the meteorological summary and crop stage.

Selection strategy	Data source	Selected predictor
Initial screening	Sentinel-2	B11_p75_mean vegetative
Initial screening	Sentinel-2	B12_max_max vegetative

Selection strategy	Data source	Selected predictor
Initial screening	Weather	Maximum air temperature
RFE	Sentinel-2	b12_p75_mean_vegetative
RFE	Sentinel-2	b9_p75_min_vegetative
RFE	Sentinel-2	b8a_p75_mean_vegetative
RFE	Sentinel-2	b11_p75_mean_vegetative
RFE	Sentinel-2	ndwi_max_max_grain_filling
RFE	Sentinel-2	b11_p75_max_vegetative
RFE	Weather	rh_sd_vegetative

3.7.2 Initial predictor screening

The three modelling approaches showed greater differences in predictive performance for crude protein concentration than for yield (Table 15). RF achieved the lowest RMSE (0.72 percentage points (pp)) and the highest R^2 (0.57), outperforming MLR (RMSE = 0.87 pp, R^2 = 0.38) and SVM (RMSE = 0.93 pp, R^2 = 0.29). Prediction bias was low for all three models, with RF showing a bias of 0.035 pp. Overall, RF provided the most accurate predictions of crude protein concentration among the evaluated models. The relationship between observed and LOOCV-predicted crude protein concentration for the best performing RF model is shown in Figure 16. Observed-versus-predicted plots for the other two models, SVM and MLR, are provided in the Appendix Figures A9 and A10. Predicted values generally increased with observed crude protein concentration, although prediction accuracy varied across the observed range. Most observations with crude protein concentrations around 14% were predicted comparatively close to the 1:1 line, whereas several observations within the intermediate range were underestimated. Predictions for 2025 showed relatively large deviations from the 1:1 line for several observations, suggesting lower predictive accuracy for this harvest year, while a particularly large underprediction was also observed for one observation from 2022. Overall, the predicted values covered a narrower range than the observed measurements, indicating a compression of prediction towards the central range and a reduced ability to reproduce extreme values.

Table 15. Predictive performance of the multiple linear regression (MLR), support vector machine (SVM) and random forest (RF) models for crude protein concentration using leave-one-out cross-validation (LOOCV). RMSE and bias are expressed in percentage points (pp). Bias was calculated as predicted minus observed crude protein concentration

Model	RMSE pp	R^2	Bias pp
MLR	0.87	0.38	0.012
SVM	0.93	0.29	0.230

Model	RMSE pp	R ²	Bias pp
RF	0.72	0.57	0.035

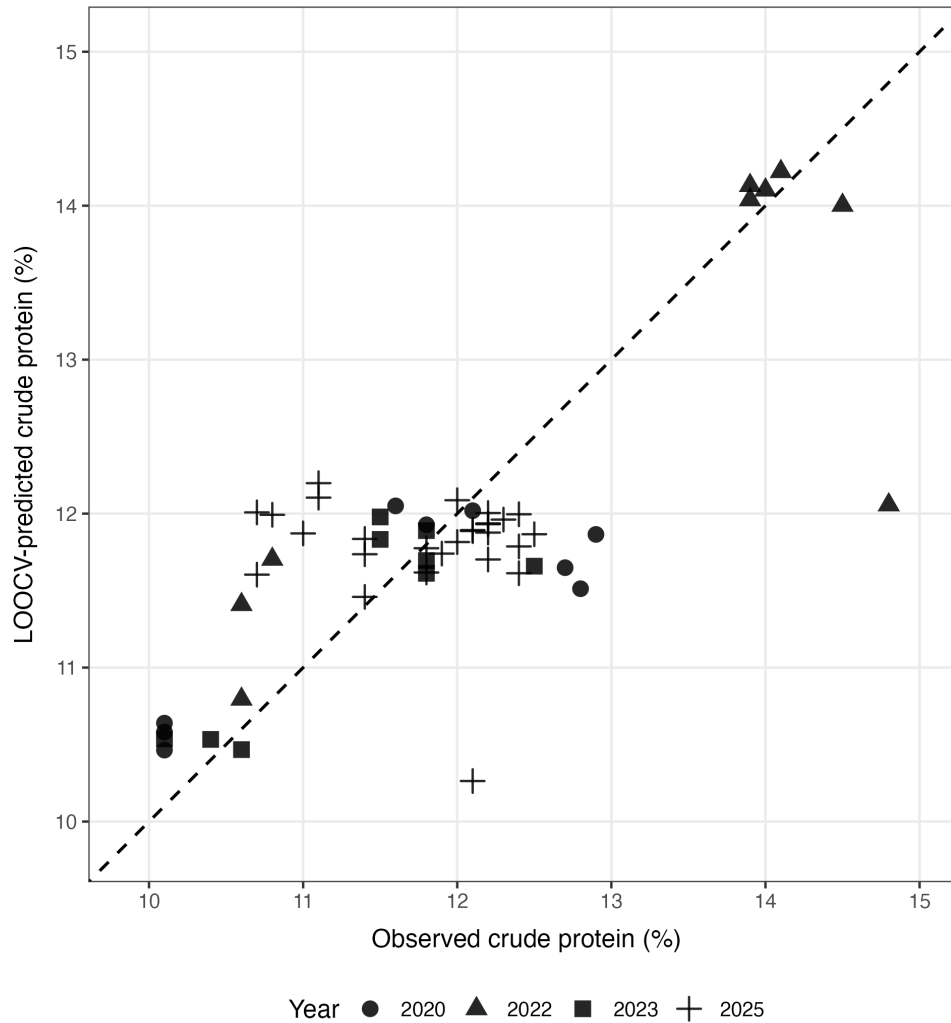


Figure 16. Relationship between observed and leave-one-out cross-validation (LOOCV) -predicted crude protein concentration obtained using the random forest (RF) model. Each point represents one plot-year observation, with symbols indicating harvest year. The dashed 1:1 line represents perfect agreement between observed and predicted values.

3.7.3 Recursive feature elimination

For crude protein concentration, RFE retained six Sentinel-2 predictors and one meteorological predictor. The selected satellite variables represented SWIR reflectance from bands B11 and B12 during the vegetative stage, red-edge reflectance from B8a during the vegetative stage, narrow near-infrared reflectance B8A during the vegetative stage, reflectance from the water-vapour band B9 during the vegetative stage, and maximum NDWI during the grain-

filling stage. More specifically, the retained variables were:

b12_p75_mean_vegetative, b9_p75_min_vegetative,
b8a_p75_mean_vegetative, b11_p75_mean_vegetative,
ndwi_max_max_grain_filling, and b11_p75_max_vegetative.

The selected meteorological predictor was the standard deviation of relative humidity during the vegetative stage (rh_sd_vegetative).

RF achieved the best predictive performance for crude protein concentration, with the lowest LOOCV RMSE (0.45 pp) and the highest R^2 (0.84; Table 16). SVM showed similar but slightly lower performance (RMSE = 0.47 pp, R^2 = 0.83), whereas MLR produced the largest prediction error (RMSE = 0.54 pp, R^2 = 0.76). Observed and RF-predicted values generally followed the 1:1 line, although predictions were compressed towards the central range, with lower concentrations tending to be overestimated and the highest concentrations underestimated (Figure 17). The corresponding plots for the MLR and SVM models are provided in the Appendix Figures A3 and A4. The small negative bias (−0.054 pp) indicated a slight overall tendency towards underestimation.

Table 16. Predictive performance of multiple linear regression (MLR), support vector machine (SVM) and random forest (RF) for crude protein concentration evaluated using leave-one-out cross-validated (LOOCV) with predictors selected through recursive feature elimination (RFE). RMSE and bias are expressed in percentage points (pp). Bias was calculated as predicted minus observed crude protein concentration.

Model	RMSE pp	R^2	Bias pp
RF	0.45	0.84	−0.054
SVM	0.47	0.83	+0.042
MLR	0.54	0.76	−0.001

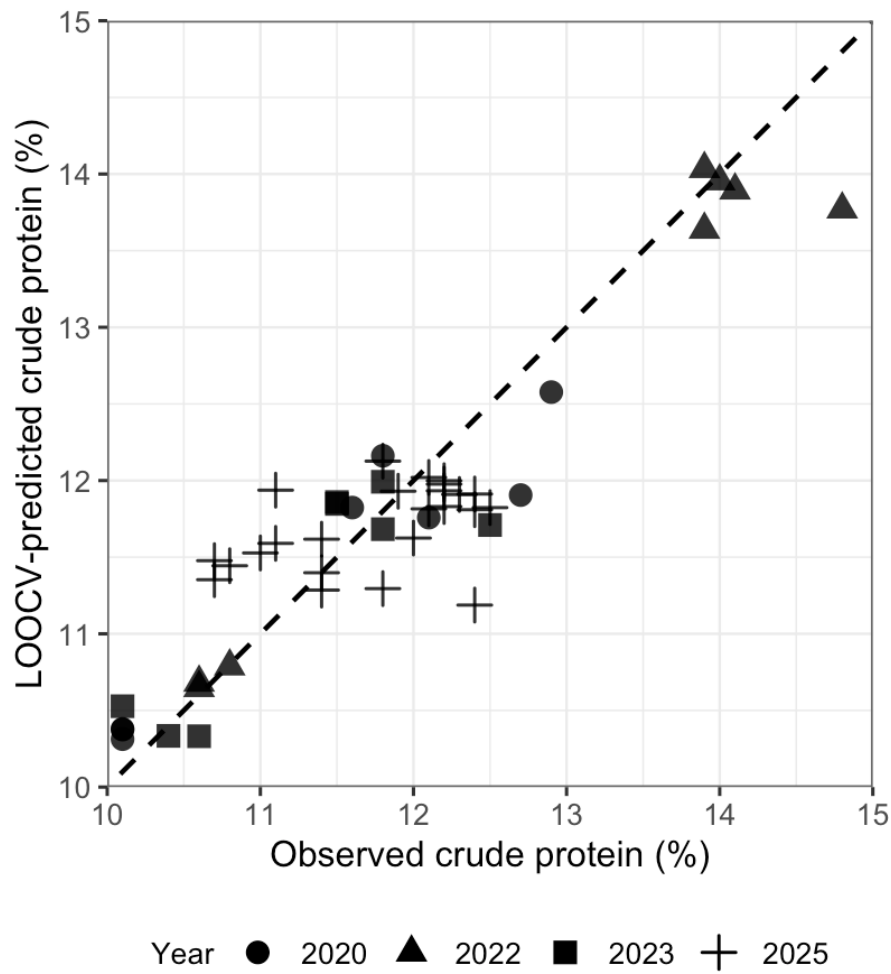


Figure 17. Relationship between observed and leave-one-out cross-validation (LOOCV) -predicted crude protein concentration obtained using random forest (RF) model. Each point represents one plot-year observation, with symbols indicating harvest year. The dashed 1:1 line represents perfect agreement between observed and predicted values.

3.8 Starch prediction results

3.8.1 Predictors selected for starch concentration

The initial screening procedure retained three predictors for starch concentration: two Sentinel-2 variables and one meteorological variable (Table 17). The satellite predictors represented mean B8A and B9 reflectance during the vegetative stage, while cumulative precipitation during vegetative growth was retained as the meteorological predictor.

Recursive feature elimination also retained three predictors: two Sentinel-2 variables and one meteorological variable. The selected satellite predictors represented the temporal mean of plot-level median B1 reflectance during the vegetative stage (b1_median_mean_vegetative) and the temporal minimum of

the plot-level 25th percentile of NDRE during grain filling (ndre_p25_min_grain_filling). Climatic water balance during the vegetative stage (water_balance_vegetative) was retained as the meteorological predictor.

Table 17. Predictors selected for starch concentration using the initial screening and recursive feature elimination (RFE) strategies. Predictor names indicate the data source, summary statistic and crop stage. For Sentinel-2 predictors, the first statistic refers to the plot-level zonal statistic and the second to the temporal aggregation within the crop stage; weather predictor names indicate the meteorological summary and crop stage.

Selection strategy	Data source	Selected predictor
Initial screening	Sentinel-2	B8A_mean_mean_vegetative stage
Initial screening	Sentinel-2	B9_mean_mean vegetative stage
Initial screening	Weather	Cumulative precipitation during the vegetative stage
RFE	Sentinel-2	b1_median_mean_vegetative
RFE	Sentinel-2	ndre_p25_min_grain_filling
RFE	Weather	water_balance_vegetative

3.8.2 Initial predictor screening

For starch concentration, the RF model showed the best predictive performance among the three evaluated approaches (Table 18), achieving the lowest RMSE (0.71 pp) and the highest R^2 (0.73). MLR showed lower predictive accuracy (RMSE = 1.06 pp; R^2 = 0.39), while SVM produced the least accurate predictions (RMSE = 1.12 pp; R^2 = 0.32). Bias was close to zero for RF, with the lowest absolute bias (0.014 pp), and MLR (+0.025 pp), whereas SVM showed a larger negative bias (−0.141 pp). Overall, RF provided the most accurate predictions of starch concentration. The relationship between observed and LOOCV-predicted starch concentration for the RF model is shown in Figure 17. The corresponding plots for the MLR and SVM models are provided in Appendix Figures A11 and A12. Predicted values generally increased with observed starch concentration, indicating that the model captured a substantial proportion of the variability among samples. However, the predicted values covered a narrower range than the observed measurements. Lower starch concentrations, particularly among observations from 2022, tended to be overestimated, whereas the highest concentrations, mainly observed in 2020, were underestimated. Predictions for intermediate starch concentrations were generally closer to the 1:1 line, although variation remained among individual observations. Overall, this pattern indicates a compression of predictions towards the central range, with larger prediction errors occurring at the extremes of the observed starch distribution.

Table 18. Predictive performance of multiple linear regression (MLR), support vector machine (SVM) and random forest (RF) models for starch concentration evaluated using leave-one-out cross-validation (LOOCV). RMSE and bias are expressed in percentage points (pp). Bias was calculated as predicted minus observed starch concentration.

Model	RMSE pp	R ²	Bias pp
MLR	1.06	0.39	0.025
SVM	1.12	0.32	−0.141
RF	0.71	0.73	0.014

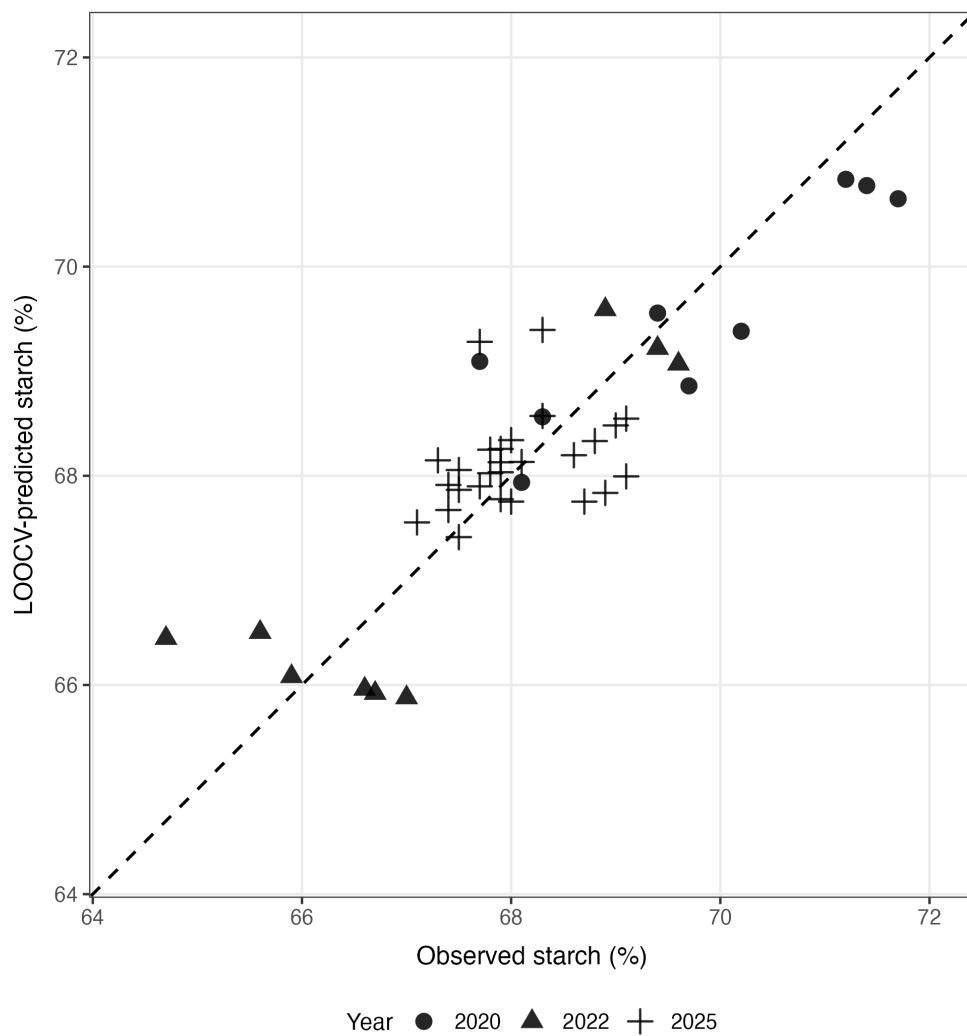


Figure 18. Relationship between observed and leave-one-out cross-validation (LOOCV)-predicted starch concentration for the random forest (RF) mode using satellite and meteorological predictors selected through the initial screening procedure. Each point represents one plot-year observation, with symbols indicating

harvest year. The dashed line represents the 1:1 relationship between observed and predicted values. Different symbols indicate the year of observation.

3.8.3 Recursive feature elimination

SVM achieved the best predictive performance for starch concentration, with the lowest LOOCV RMSE (0.67 pp) and the highest R^2 (0.77; Table 19). RF showed similar performance (RMSE = 0.70 pp; $R^2 = 0.77$), whereas MLR produced the largest prediction error (RMSE = 0.91 pp; $R^2 = 0.62$). Observed and SVM-predicted concentrations generally followed the 1:1 line, although the predicted range was narrower than the observed range (Figure 19). The corresponding plots for the MLR and RF models are provided in the Appendix Figures A5 and A6. The lowest observed value was overestimated, while some of the highest concentrations were underestimated. The negative bias (−0.011 pp) indicated no relevant systematic tendency towards over- or underestimation.

Table 19. Predictive performance of multiple linear regression (MLR), support vector machine (SVM) and random forest (RF) models for starch concentration evaluated using leave-one-out cross-validated (LOOCV) with predictors selected through recursive feature elimination (RFE). RMSE and bias are expressed in percentage points (pp). Bias was calculated as predicted minus observed starch concentration.

Model	RMSE (% dry matter)	R^2	Bias
SVM	0.67	0.77	−0.011
RF	0.70	0.77	+0.013
MLR	0.91	0.62	−0.034

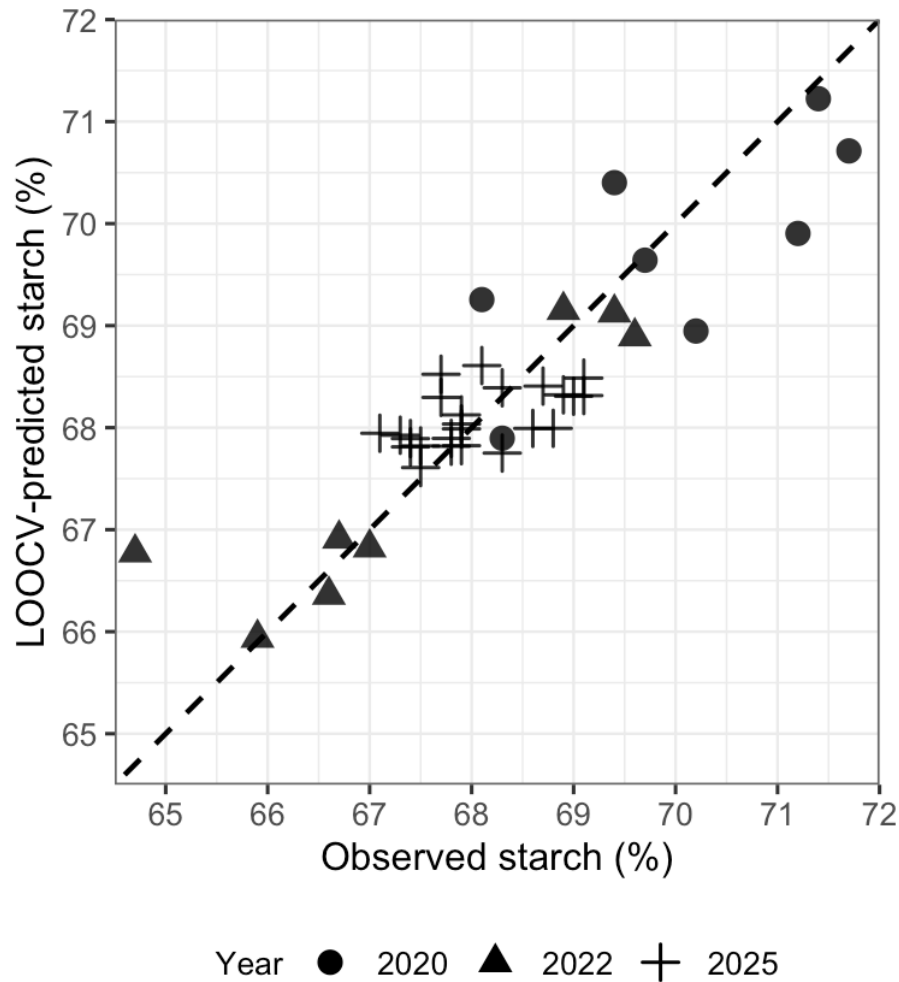


Figure 19. Relationship between observed and leave-one-out cross-validation (LOOCV)-predicted starch concentration for the support vector machine (SVM) model using satellite and meteorological predictors selected through recursive feature elimination (RFE). Each point represents one plot-year observation, with symbols indicating harvest year. The dashed 1:1 line represents perfect agreement between observed and predicted values.

3.9 External validation

External validation of the SVM yield model was performed using the independent winter wheat observations from the SITES SAFE Lönnstorp experimental site, covering the growing season 2018-2023. The model showed lower predictive performance at the external site than during the internal validation at Ultuna (Figure 20). The relationship between observed and predicted yield resulted in an R^2 of 0.21, with an RMSE of 1198.7 kg ha⁻¹ and a bias of -532 kg ha⁻¹. Predicted yields varied over a narrower range than the observed values, with an overall tendency towards underestimation at higher observed yields. Individual observed and predicted values and prediction errors are reported in Appendix Table A13.

A comparison of the predictor ranges used by the SVM model showed that some Lönnstorp observations extended beyond the range represented in the Ultuna training dataset (Table 20). Maximum vegetative stage NDVI ranges from 0.50 to 0.91 in the Ultuna training data and from 0.80 to 0.94 at Lönnstorp. Mean vegetative-stage air temperature ranged from 5.1 to 7.1°C at Ultuna and from 6.8 to 8.1°C at Lönnstorp. The external dataset included conditions that were only partly represented in the trained data.

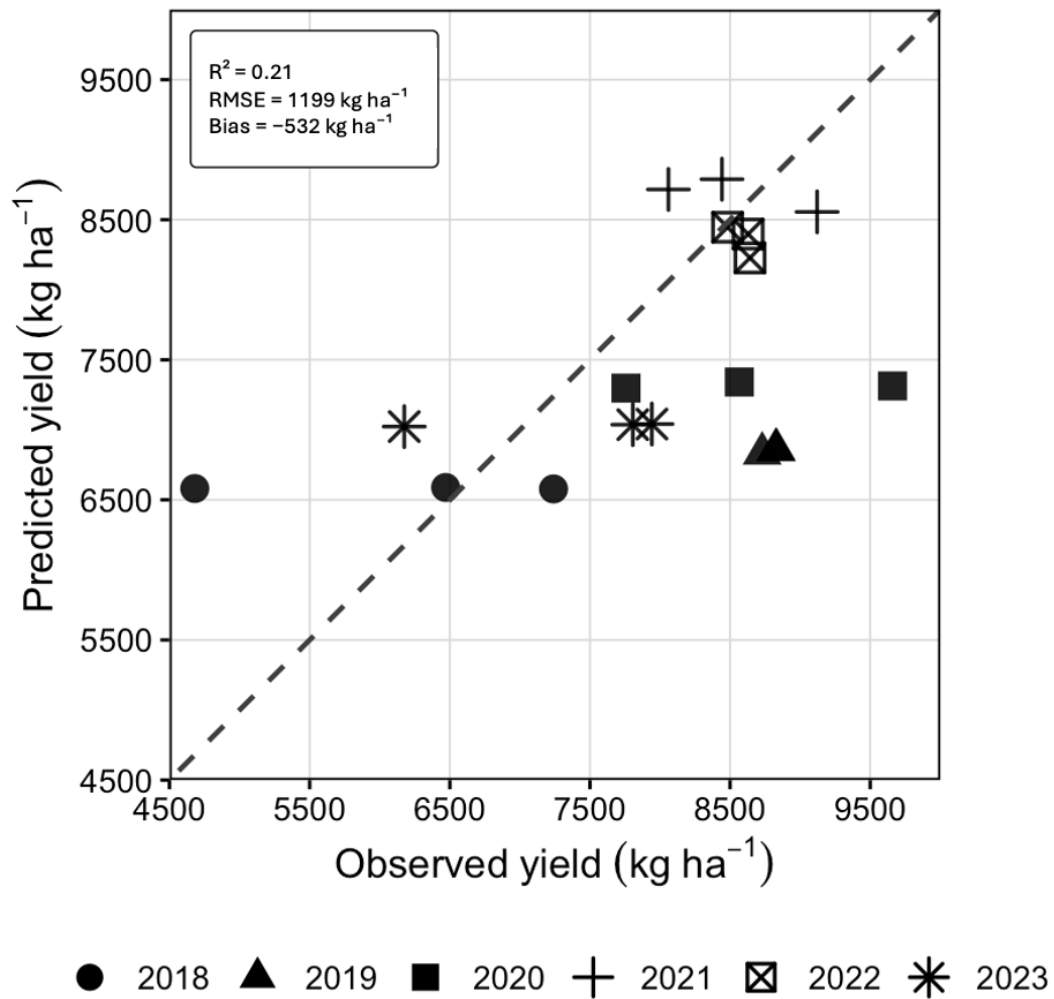


Figure 20. Observed and predicted winter wheat grain yield for the external validation dataset from the SITES SAFE Lönnstorp experimental site (2018-2023). Predictions were obtained using the SVM yield model developed from the Ultuna dataset. The dashed line represents the 1:1 relationship between observed and predicted yield.

Table 20. Comparison of predictor ranges between the Ultuna training dataset and the independent SITE Lönnstorp external validation dataset for the predictors retained in the final SVM yield model.

Predictor	Ultuna training range	Lönnstorp range
-----------	-----------------------	-----------------

Maximum vegetative stage NDVI	0.50–0.91	0.80–0.94
Mean vegetative stage air temperature (°C)	5.1–7.1	6.8–8.1

3.10 UAV

A positive relationship was observed between Sentinel 2-derived VARI and UAV-derived VARI across the experimental plots (Figure 21). Despite differences in spatial resolution between the two platforms, the regression slope was positive ($\beta = 0.43$), indicating that higher Sentinel-2 VARI values were generally associated with higher UAV-derived VARI values.

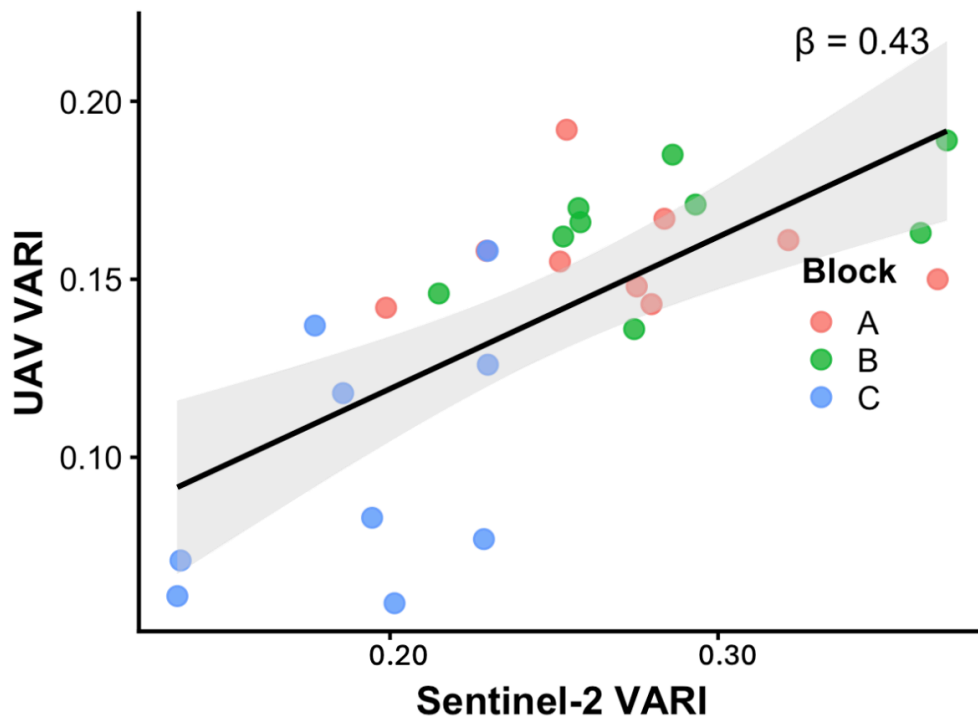


Figure 21. Relationship between Sentinel-2- derived VARI on 18 May 2025 and UAV-derived VARI on 8 May 2025 across experimental plots. Each points represents one experimental plot, with the mean VARI calculated from the pixels within that plot for each platform.

3.11 Comparison of spatial detail in Sentinel-2 and UAV imagery

A clear difference in the level of spatial detail was observed between the Sentinel-2 and UAV vegetation-index maps (Figure 22). The Sentinel-2-derived VARI image showed relatively coarse and spatially homogeneous

areas within the experimental plots, reflecting the larger satellite pixel size. In contrast, the UAV-derived VARI image revealed substantially finer within-plot spatial variability and small-scale patterns that were not distinguishable in the Sentinel-2 image. Despite these differences, both platforms captured similar broad spatial patterns across the experimental area. Plots with higher Sentinel-2 VARI generally also showed higher UAV-derived VARI, with Block B tending towards higher values in both datasets and Block C more frequently occurring at the lower end of the distributions.

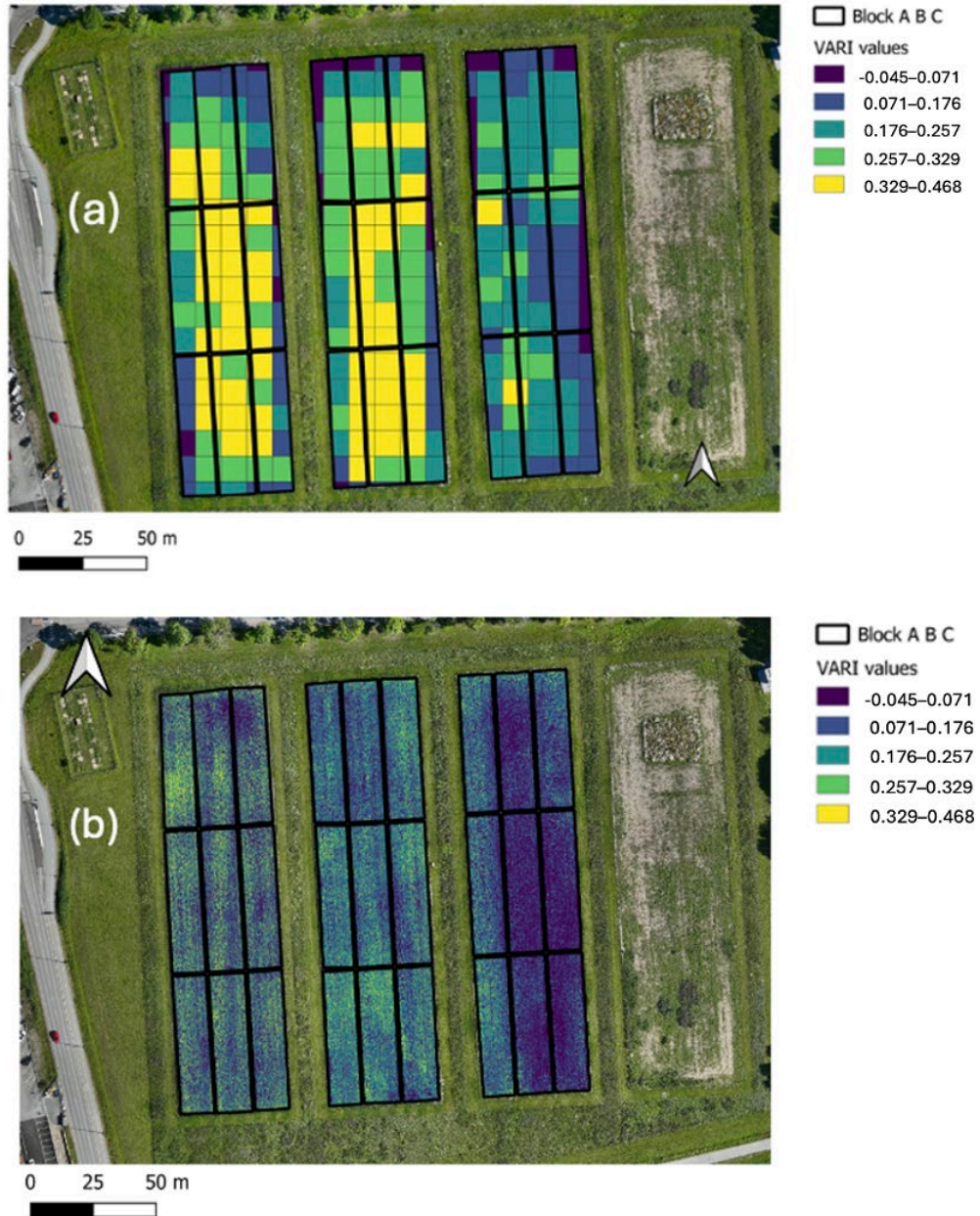


Figure 22. Comparison of spatial resolution between Sentinel-2 and UAV vegetation-index imagery across the experimental plots in May 2025. (a) Sentinel-2-derived VARI on 18 May 2025; (b) UAV-derived VARI on 8 May 2025.

4. Discussion

4.1 Main findings

The main finding of this study was that Sentinel-2 and meteorological variables representing the vegetative stage contained substantial information for explaining plot-level variation in winter wheat yield and grain quality. RFE improved the best predictive performance, although the magnitude of improvement differed among traits. Non-linear or more flexible modelling approaches provided advantages for crude protein and starch concentration, whereas multiple linear regression remained competitive for grain yield when the initially screened predictor set was used.

In addition to the predictive modelling results, the analysis of the agronomic response variables showed substantial effects of both management and year. Grain yield differed significantly among fertiliser treatments and harvest years, while the Treatment $\times$ Year interaction was not significant (Table 6, Figure 5), suggesting that the overall treatment response was relatively consistent across the years included in this analysis. In contrast, crude protein concentration showed a significant Treatment $\times$ Year interaction (Figure 6, Table 7) indicating that the response of grain protein concentration to fertiliser treatment depended more strongly on the growing season. These results highlight the importance of considering both management and interannual environmental variability when interpreting the relationships between remotely sensed meteorological and grain response variables.

4.2 Agronomic interpretation

4.2.1 Grain yield

The mixed-model analysis provided further agronomic context for the yield variability observed across the study. Fertiliser treatment had a significant effect on grain yield, with the untreated treatment producing lower yields than both fertilised treatments (Table 6, Figure 4). This response is consistent with the well-established role of nitrogen fertilisation in increasing wheat productivity, as demonstrated in long-term field experiments such as the Broadbalk Wheat Experiment at Rothamsted (Lawes & Gilbert 1864). N fertilisation has consistently been associated with increased wheat yield, although the magnitude of the response depends strongly on the amount of N supplied, soil conditions, management and climate (Yokamo et al. 2023). Similar patterns have been reported in Swedish field experiments. For example, Delin et al. (2005) evaluated several N application rates in winter wheat in south-western Sweden and showed that yield responses to fertiliser N varied with plant-available soil N and local field conditions. Hamnér et al. (2017) further showed, across seven field trials in southern and central Sweden with N rates ranging from 0 to 240 kg N ha⁻¹, that nitrogen supply strongly influenced nutrient accumulation during winter wheat development. However,

the absence of a significant difference between the two fertilised treatments in the present study suggests that the yield response may have started to plateau under the conditions of this study.

Harvest year had an even stronger effect on yield, indicating that fertiliser treatment alone cannot explain the observed differences in crop performance. The meteorological profiles also showed clear interannual differences in VPD, air temperature and water balance (Figure 10–12). Long-term experiments in winter wheat have similarly demonstrated year-to-year variation in both grain yield and the response to N fertilisation, with temperature and rainfall during specific periods of crop development contributing to this variation (Addy et al. 2020).

The significant year effect should not be interpreted as a purely meteorological effect. In the present dataset, harvest year represents differences in seasonal growing conditions that were not fully characterised and may have captured variation in crop management, soil conditions, previous crop effects and other unmeasured factors. Therefore, the mixed model supports the interpretation that winter wheat yield at the testbed resulted from the combined influence of nutrient availability and interannual environmental variability rather than from any single factor.

Soil data indicated spatial heterogeneity across the experimental testbed, particularly in soil texture and organic matter (Table 8, Figures 8–9). The site was predominantly characterised by heavy clay and clay loam soils, with some differences among blocks. However, comparisons between the 2018 and 2024 soil surveys should be interpreted cautiously because the sampling and spatial-processing methods were not fully comparable.

The 2024 soil survey showed spatial variability in clay content and soil organic matter across the experimental plots. When these variables were included separately as covariates in linear mixed-effects models for grain yield, neither clay content nor soil organic matter showed a significant association with yield (Table 8). In contrast, fertiliser treatment and year remained significant in both models, whereas the Treatment $\times$ Year interaction remained non-significant. These results suggest that, within the range of soil conditions represented at the site, variation in clay content and soil organic matter did not explain additional detectable variation in grain yield after accounting for treatment and year.

This finding does not imply that soil properties were unimportant for crop development. Only clay content and soil organic matter were formally evaluated as covariates, whereas other soil characteristics, including nutrient availability, drainage, soil structure and interactions among soil properties, may also influence crop growth and yield. Furthermore, the limited and unbalanced dataset made it difficult to separate soil-related spatial effects from differences associated with year, block and management. The available documentation did not indicate that the experimental blocks were deliberately selected to represent contrasting soil conditions. The observed soil variability was therefore treated as an existing characteristic of the site rather than as an intentionally designed experimental factor.

The Sentinel-2 index distributions also showed notable interannual differences during vegetative growth. In 2025, NDVI and NDRE displayed a

markedly wider range than in the other harvest years (Figure 13). This may partly reflect the broader spatial representation of the 2025 dataset, since winter wheat was cultivated in all three experimental blocks, whereas the other analysed harvest years were represented by a single block. The 2025 distribution may therefore have captured greater variability in crop development in addition to temporal variation within the vegetative period.

Among the Sentinel-2 predictors, the vegetative-stage minimum of Sentinel-2 Band 4 (`'b4_min_min_vegetative'`), representing the lowest within plot minimum B4 reflectance recorded across satellite acquisition dates during the vegetative period, was negatively associated with grain yield. Sentinel-2 B4 represents reflectance in the red region of the spectrum, where chlorophyll absorption is strong. Lower red reflectance is generally associated with a greater contribution of green vegetation and stronger chlorophyll absorption (Frampton et al. 2013). Agronomically, this may be because the minimum red reflectance captures the period of the strongest red absorption by the canopy, potentially corresponding to maximum canopy greenness or chlorophyll activity. A median, in contrast, represents the more typical red reflectance condition across the vegetative stage and is less influenced by individual acquisition dates with unusually low reflectance values. Minimum-based metrics are also more sensitive to extreme observations than median-based metrics, which are more robust to outliers (White et al. 2014).

In contrast, the maximum across acquisition dates of the plot level median NDRE values during the vegetative stage (`'ndre_median_max_vegetative'`) showed a positive association with grain yield. NDRE is sensitive to canopy chlorophyll status and vegetation vigour and may retain greater sensitivity than NDVI under relatively dense canopy conditions (Delegido et al. 2011). In this variable, the use of the median reduces the influence of extreme pixel values, whereas retaining the maximum values across the vegetative period may capture the stage at which the canopy expressed its strongest red-edge response. Because red-edge reflectance is closely related to chlorophyll content and green leaf area, this seasonal maximum may have represented a period of particularly vigorous crop development. In contrast, a temporal median would describe more typical conditions across the vegetative period and could reduce the influence of short periods of maximum canopy development. Higher NDRE values might have reflected better crop nutritional status, canopy development and biomass accumulation.

A similar pattern was observed for `'ndvi_max_max_vegetative'`, representing the maximum across acquisition dates of the within plot maximum NDVI values during the vegetative period, which was also positively associated with grain yield. Higher NDVI values generally indicate greater green vegetation cover and canopy development, suggesting that plots with more vigorous vegetative growth tended to produce higher yields. However, NDVI becomes progressively less sensitive as canopy cover and biomass increase because of saturation effects (Tenreiro et al. 2021). The simultaneous selection of NDVI- and NDRE-based predictors therefore suggests that both general canopy development and chlorophyll-related variation contained information relevant to final yield.

Meteorological predictors provided complementary information on the environmental conditions experienced by the crop. The number of frost days during vegetative growth was negatively associated with grain yield. This relationship is consistent with the adverse effects that frost exposure can have on wheat development, although the magnitude of frost damage depends strongly on crop developmental stage and temperature severity (Leske & Biddulph 2022). In the present dataset, a greater frequency of frost events during the vegetative development was therefore associated with lower final yield.

VPD was also negatively associated with yield. Higher VPD represents greater atmospheric evaporative demand and can increase crop water loss. Under sufficiently high evaporative demand, plants may reduce stomatal conductance to limit transpiration, which can also restrict carbon assimilation and biomass accumulation (Joshi et al. 2022). The observed negative association may therefore indicate that seasons characterised by greater atmospheric water demand during vegetative growth were less favourable for canopy development and subsequent yield formation.

Mean air temperature during the vegetative period showed the opposite pattern and was positively associated with grain yield. Under the relatively cool growing conditions of Ultuna, warmer spring temperatures within the range observed in this study may have favoured vegetative growth, particularly leaf expansion and biomass accumulation (Yu et al. 2023). This relationship should nevertheless be interpreted within the climatic conditions represented by the dataset and should not be taken to imply that increasing temperature is always beneficial. Excessively high temperatures, particularly during sensitive developmental stages, can accelerate crop development and reduce wheat yield (Fu et al. 2023).

Overall, the selected predictors consistently point to the importance of crop conditions during the vegetative period for final grain yield. Higher yield was associated with higher NDRE and NDVI values and lower red-band (B4) reflectance during the vegetative stage. Higher NDVI values are generally consistent with greater green vegetation cover and canopy development, while higher NDRE values may indicate differences in chlorophyll status. Lower B4 reflectance indicates that smaller proportion of red light was reflected by the canopy, which may be associated with greater chlorophyll absorption. However, these spectral measurements provide indirect indicators of canopy conditions rather than direct measurements of chlorophyll concentration or biomass.

Higher yield was also associated with fewer frost events, lower atmospheric evaporative demand and relatively warmer conditions. Taken together, these results suggest that both crop vigour and the environmental conditions associated with vegetative development contributed to the yield differences observed among plots and years. The relationships should, however, be interpreted as associations within the conditions represented in the dataset rather than as evidence of direct causal effects.

4.2.2 Crude protein

The agronomic analysis showed that crude protein concentration increased with fertilisation, but unlike grain yield, the magnitude of the response varied among harvest years, as indicated by the significant Treatment $\times$ Year interactions (Table 6, Figure 5). This pattern is consistent with previous studies showing that nitrogen fertilisation generally increases grain protein concentration in winter wheat, but the strength of this response depends on environmental and growing conditions (Ghimire et al. 2021). Similar variability has been reported under Swedish conditions. Delin et al. (2005), using several N application rates in winter wheat in south-western Sweden, found that crude protein responses to fertiliser N varied among years and local field conditions. Long et al. (2024) likewise reported a significant Year $\times$ N interaction for grain protein concentration, indicating that the response of grain protein concentration to increasing N supply differed among years. These findings support the results of the present study, in which the overall increase was greatest in 2022, more gradual in 2020 and smaller in 2023. Such interannual differences may reflect differences in crop status, nitrogen uptake, remobilisation, soil nitrogen availability and weather conditions during the grain filling period (Barbottin et al. 2005).

Environmental conditions can also modify the balance between grain yield and grain protein concentration, since high biomass and grain production may dilute grain nitrogen concentration when nitrogen supply does not increase proportionally (Bogard et al. 2010). Therefore, the significant Treatment $\times$ Year interaction observed here suggests that the effect of fertilisation on grain protein concentration was not constant but depended on the conditions prevailing during the growing season. However, the treatment effect should be interpreted cautiously in this study. Treatment represented the number of fertiliser applications rather than a consistently quantified nitrogen rate, and complete information on total nitrogen input was not available for all the years. The result therefore demonstrates differences among fertiliser treatments and years, but it does not allow the response of grain protein concentration to nitrogen application rate to be quantified directly.

The spectral predictors selected for crude protein concentration (Table 14) were predominantly associated with the SWIR, near-infrared and red-edge regions. B11 and B12 were repeatedly retained among the selected predictors. Both bands are in the SWIR region, which is sensitive to vegetation water content and changes in canopy condition. SWIR wavelengths have also been used in vegetation nitrogen assessment because spectral absorption features related to proteins and other nitrogen-containing compounds occur within this region (Curran 1989). However, the broad Sentinel-2 SWIR bands should not be interpreted as direct measurements of plant nitrogen or grain protein concentration. In wheat, the inclusion of SWIR wavelengths has been shown to improve the estimation of flag-leaf nitrogen concentration compared with models based on visible and infra-red wavelengths (Bruning et al. 2019). Their association with final grain protein concentration may therefore be indirect, as SWIR reflectance is also strongly influenced by vegetation water and dry matter content, as well as by canopy structural characteristics (Liu et al. 2015). These characteristics change during crop development and senescence and may

reflect physiological conditions associated with nitrogen uptake and remobilisation to the grain. Consequently, SWIR predictors may capture crop conditions associated with subsequent grain protein formation rather than a direct spectral response to grain protein itself.

B8A, centred at approximately 865 nm, is located at the long-wavelength end of the Sentinel-2 red-edge region, at the transition toward the narrow infrared, providing complementary information from the narrow near-infrared region. Reflectance in the red-edge and adjacent near-infrared wavelengths is sensitive to vegetation biophysical and biochemical characteristics, including chlorophyll content, leaf area and canopy structure (Clevers et al. 2017). These characteristics are closely related to nitrogen status, and Sentinel-2 has previously been used to estimate canopy nitrogen content in winter wheat (Bossung et al. 2022). The selection of B8A may therefore indicate that differences in canopy development and nitrogen-related physiological status during the vegetative growth were associated with variation in grain protein content.

Previous research in southern Sweden has also investigated the potential of multispectral remote sensing for predicting grain crude protein concentration in winter wheat. Wolters et al. (2022) developed and evaluated prediction models based on UAV multispectral imagery, identifying the chlorophyll index derived from near-infrared and red-edge reflectance as a useful predictor. The models were evaluated using leave-one-trial-out cross-validation to assess their performance across independent experimental trials.

The contribution of B9 is more difficult to interpret physiologically. Sentinel-2 B9 is located within a spectral region strongly affected by atmospheric water-vapour absorption and is primarily intended to support atmospheric correction (Martins et al. 2017). Its selection should therefore not be interpreted as evidence of a direct physiological relationship with grain protein concentration. Given its relatively coarse spatial resolution and its sensitivity to atmospheric water vapour, B9 may instead have captured broader environmental or spectral differences associated with the growing conditions represented in the dataset. Alternatively, its selection may have resulted from a chance association, particularly given the limited number of observations and the large number of candidate predictors.

The selection of an NDWI-related predictor further suggests that crop water status contributed to crude protein prediction. As discussed, NIR–SWIR water indices are sensitive to vegetation water content (Gao 1996) and the relationship with crude protein is therefore likely to reflect differences in canopy water status and crop development rather than a direct measurement of protein concentration. This interpretation is consistent with evidence that wheat grain protein concentration can respond to water availability and environmental conditions (Wan et al. 2022).

Meteorological predictors also contributed to crude protein prediction. Air temperature is an important driver of metabolic rates of crop development and can influence the duration of grain filling, nitrogen uptake and assimilation and the subsequent remobilisation of nitrogen from vegetative tissues to the grain. In particular, temperature during and after anthesis can alter grain development and final protein concentration (Savill et al. 2018). Relative-humidity

variability may instead reflect differences in atmospheric moisture conditions and evaporative demand, which can influence crop water status and biomass production (Wan et al. 2022). These environmental conditions may therefore affect the balance between carbohydrate accumulation and nitrogen concentration in the developing grain.

The interpretation of these predictors should also consider the relationship between grain yield and protein concentration. Grain protein concentration is often negatively associated with yield because greater carbohydrate accumulation can increase grain mass more rapidly than nitrogen accumulation, resulting in a dilution of grain nitrogen and therefore protein concentration (Maeoka et al. 2020). Consequently, spectral or meteorological variables associated with vigorous crop growth and high biomass production may influence protein concentration indirectly through their effects on both yield formation and nitrogen partitioning.

Overall, the selected predictors suggest that final crude protein concentration reflected the combined influence of nitrogen management, canopy development, crop water status and atmospheric conditions. However, Sentinel-2 spectral bands and meteorological variables should not be interpreted as direct measurements of grain protein. Rather, they appear to capture crop physiological and environmental conditions associated with nitrogen uptake, biomass accumulation and nitrogen remobilisation. Because fertilisation information was incomplete and could not consistently be expressed as kg N ha^{-1} , some of the variation attributed to spectral and meteorological predictors may also reflect undocumented differences in nitrogen management. The identified relationships should therefore be interpreted primarily as associations within the conditions represented by the dataset.

4.2.3 Starch

The grain-quality analysis showed a negative relationship between crude protein and starch concentration (Figure 6). The simple linear regression produced a slope of $\beta = -1.05$, indicating that a one-percentage-point increase in crude protein concentration was associated with an estimated decrease of approximately 1.05 pp in starch concentration. This pattern suggests that variation in these two major components of grain composition was not independent and is consistent with a trade-off between protein and carbohydrate accumulation in the grain.

This relationship should nevertheless be interpreted cautiously. The analysis included only observations for which both crude protein and starch measurements were available, and data from 2023 were excluded because starch measurements were incomplete. Moreover, the regression describes an association between the two grain quality traits and does not demonstrate that variation in crude protein directly caused changes in starch concentration. Instead, both traits may have responded to common environmental and agronomic conditions during grain development. These conditions can alter the relative accumulation of nitrogen compounds and carbohydrates in the grain,

thereby affecting the final concentrations of protein and starch (Cormier et al. 2016).

The spectral predictors selected for starch concentration indicate that crop condition during the vegetative stage contained information associated with final grain composition. During tillering and stem elongation, canopy structure and biomass develop rapidly (Xie et al. 2016). Remote-sensing observations during this period may therefore capture differences in crop vigour and physiological status that influence subsequent biomass production and grain development.

Several of the selected spectral variables were related to canopy development and vegetation status. B8A, located in the red-edge region, is sensitive to vegetation biophysical properties and may reflect differences in canopy structure and biomass development (Clevers et al. 2017). B9 is more difficult to interpret physiologically because it lies within a spectral region strongly influenced by atmospheric water-vapour absorption and is primarily used for atmospheric applications (Martins et al. 2017). Its association with starch concentration should therefore be interpreted cautiously and not as evidence of a direct relationship with grain composition.

Water-related conditions were also represented among the selected predictors. Cumulative precipitation during vegetative growth was retained by the initial screening, while vegetative-stage climatic water balance was selected through RFE. These variables suggest that differences in water availability during crop development were associated with final starch concentration. Adequate water supply supports canopy development and photosynthesis, whereas water limitation can restrict biomass accumulation and alter subsequent grain filling. However, the observed relationships should be interpreted as associations rather than evidence that precipitation or water balance directly determined starch accumulation (Sheng et al. 2022).

NDRE during grain filling provides a potentially more direct indication of crop condition during the period in which grain composition is established. NDRE is sensitive to canopy chlorophyll status and may reflect the persistence of green, photosynthetically active tissue. Higher chlorophyll retention during grain filling may support continued assimilate production and carbohydrate supply to the developing grain (Hassan et al. 2021). The minimum across Sentinel-2 acquisition dates of the within plot 25th percentile of NDRE values during grain filling (ndre_p25_min_grain_filling) may indicate that the lower end of the within-plot NDRE distribution during grain filling contained relevant information for predicting final starch concentration. The 25th percentile represents the NDRE value below which 25% of the pixel values within a plot fall, reflecting the lower part of NDRE distribution while being less sensitive to extreme values than the absolute minimum. This predictor may therefore capture differences in spatial heterogeneity of canopy senescence or stress among plots. Such differences could influence photosynthetic activity and carbon assimilation during grain filling, when starch is accumulated in the developing grain. Nevertheless, this interpretation should be considered indicative, as NDRE is an indirect measure of canopy condition rather than a direct measure of starch accumulation.

B1 requires caution. Sentinel-2 B1 is centred at approximately 443 nm and is primarily designed for aerosol-related applications, with a native spatial resolution of 60 m (Gascon et al. 2017). Given the relatively small experimental plots used in this study, plot-level B1 values may have been influenced by surrounding surfaces, mixed pixels and the resampling required during extraction. Its selection should therefore not be interpreted as evidence of a specific physiological relationship with starch concentration. Rather, B1 may have captured broader spectral or environmental differences within the dataset.

Overall, the predictors selected for starch concentration suggest that both early crop development and environmental conditions related to water availability were associated with final grain composition, while crop status during grain filling provided additional information. The negative association between crude protein and starch further indicates that starch concentration should be interpreted within the broader context of grain composition rather than as an isolated quality trait. As with the other response variables, the Sentinel-2 and meteorological predictors should therefore be regarded primarily as indirect indicators of crop physiological and environmental conditions associated with final starch concentration.

4.3 Effect of predictor-selection strategy

The third research question examined whether recursive feature elimination (RFE) improved predictive performance compared with the initial predictor-screening approach. Overall, the best-performing model improved after RFE for all three response variables, although the magnitude of improvement differed substantially among traits.

For grain yield, the lowest RMSE decreased from 1,105 kg ha⁻¹ with the initially screened predictors to 957 kg ha⁻¹ after RFE, a reduction of 148 kg ha⁻¹, or approximately 13% (Tables 10–11). Nevertheless, the resulting RMSE was relatively high compared with values reported in previous satellite-based winter wheat studies. Bouras et al. (2023) reported RMSE values of 600–680 kg ha⁻¹ after assimilating Sentinel-2 derived LAI into a crop growth model in southern Sweden. Similarly, Marszalek et al. (2022) using Sentinel-2 and climatological data for multi-year winter wheat yield prediction in southern Germany, reported an RMSE of 560 kg ha⁻¹. These comparisons indicate that the prediction error obtained in the present study was comparatively high, although direct comparison should be interpreted cautiously because of differences in dataset size, spatial scale, predictor sets and validation procedures. Predictive R² simultaneously increased from 0.82 to 0.87. In this case, RFE also reduced the predictor set from four variables to two, indicating that a smaller set of Sentinel-2 and meteorological predictors was retained and slightly improved the information relevant to yield prediction. The joint RFE retained only Sentinel-2 predictors (Table 12), indicating that within the Ultuna dataset, the spectral variables contained most of the information selected for grain yield prediction. This result should not be interpreted as evidence that meteorological data are generally unnecessary for yield prediction. However,

the relative importance of meteorological predictors may differ among sites and growing seasons. The present study did not establish whether including meteorological data would improve model performance when applied to independent sites or years. This possibility therefore requires further investigation. The strong contribution of the spectral predictors may partly reflect their ability to integrate the effects of environmental conditions on crop development through the spectral response of the canopy.

However, the satellite only solution required information from both the vegetative and grain-filling stages, whereas the separate RFE approach combining Sentinel-2 and meteorological data retained only predictors from the vegetative stage. From a practical perspective, this represents a trade-off between data requirements and prediction timing. A satellite-only model may be easier to apply where reliable meteorological data are unavailable, but it would require observations later in the growing season. In contrast, a model based on Sentinel-2 and meteorological information from the vegetative stage could potentially provide an earlier yield estimate, potentially providing more time for harvest planning and other management decisions. However, this potential advantage has not been evaluated under operational forecasting conditions and would require further validation across independent years and sites.

The largest improvement was observed for crude protein concentration (Tables 14–15, Figures 15–16). The RMSE of the best-performing model decreased from 0.72 to 0.45 pp, representing a reduction of approximately 37%, while predictive R^2 increased from 0.56 to 0.83. Unlike for grain yield, however, this improvement was accompanied by an increase in the number of selected predictors, from three in the initial screening to seven after RFE. The improved performance may therefore reflect the contribution of several complementary spectral and meteorological variables rather than greater model parsimony. At the same time, retaining seven predictors from only 48 observations increases the possibility that some of the selected relationships are specific to the available dataset.

For starch concentration, the improvement associated with RFE was comparatively small (Tables 17–18, Figures 17–18). The lowest RMSE decreased from 0.71 to 0.67 pp, corresponding to an improvement of approximately 5%, while predictive R^2 increased from 0.73 to 0.77. Both selection approaches retained three predictors. This suggests that much of the information relevant to starch concentration had already been captured by the initially screened predictor set, and that changing the selected variables provided only a modest additional gain in predictive performance.

Taken together, these results indicate that the use of RFE was generally beneficial, although the magnitude of the improvement differed among response variables. It provided a clear improvement for crude protein concentration, a moderate improvement for grain yield, and only a limited improvement for starch concentration. These findings suggest that RFE may provide a useful and more systematic alternative to the relatively laborious initial predictor-screening procedure used in this study. However, its effectiveness should still be evaluated using larger datasets and when models are tested across independent sites and years.

The apparent improvement after RFE should be interpreted cautiously because RFE was performed once using the full dataset before the final leave-one-out cross-validation. Although repeated five-fold cross-validation was used within the RFE procedure to evaluate predictor subsets, observations later used as validation cases in the final LOOCV had already contributed to the feature selection process. Therefore, the final LOOCV did not provide a completely independent evaluation of the full modelling pipeline, and predictive performance may be optimistic compared with a nested approach in which RFE is repeated using only the training data within each validation iteration. In nested cross-validation, predictor selection is repeated separately within each training set, so that observations used for validation do not influence which predictors are selected. However, even this approach would only provide a more rigorous estimate of internal predictive performance within the available datasets and would not guarantee transferability to new sites or years. Nevertheless, the validation performed in this study provides useful information on how consistently the identified relationships can predict observations within the range of conditions represented at Ultuna. It also allows a direct comparison of modelling approaches and predictor-selection strategies under the same validation framework, showing that RFE improved predictive performance and that Sentinel-2 and meteorological variables contained useful information for explaining variation in yield and grain quality.

Despite these limitations, the analyses provide an exploratory assessment of which Sentinel-2 and meteorological predictors are associated with variation in winter wheat yield and grain quality at Ultuna, and how different modelling approaches perform when applied to the same dataset. The lower LOOCV errors obtained after RFE suggest that systematic feature selection may be beneficial. The results are therefore valuable for identifying promising predictors and modelling approaches for further investigation, rather than demonstrating their general applicability. Furthermore, the reduced predictive capacity observed during external validation at Lönnstorp highlights the importance of distinguishing between apparent within-site predictive accuracy and model transferability. These results provide a basis for improving future modelling strategies, particularly through the use of larger datasets (multi-year and multi-site) and independent validation.

4.4 Comparison of modelling algorithms

The fourth research question examined whether the non-linear modelling approaches, random forest (RF) and support vector machine regression (SVM), outperformed multiple linear regression (MLR). Model comparison was based not only on cross-validation metrics, but also on the observed-versus-predicted relationships, which helped identify systematic deviations from the 1:1 line and differences in model behaviour across the observed range. The best-performing models are presented in the Results section, while the corresponding plots for the remaining models are included in the Appendix.

For grain yield, the three algorithms performed very similarly when the initially screened predictors were used. The differences in RMSE and

predictive R^2 were small, and the observed-versus-predicted plots showed broadly comparable agreement with the 1:1 line. This suggests that the additional flexibility of SVM and RF provided little advantage over the simpler linear model under this predictor set. However, the relatively small dataset should be considered when interpreting these differences, as the limited number of observations may have restricted the ability to distinguish clearly between the predictive performance of the three modelling approaches.

This finding is relevant to previous Swedish research on satellite-based yield estimation. Alshihabi et al. (2024) demonstrated that a relatively simple empirical relationship between Sentinel-2 vegetation indices and yield measurements could be used to map within-field winter wheat yield variation. Although their study addressed a different prediction task and spatial scale, it illustrates that model complexity is not necessarily a prerequisite for useful yield estimates. In the present study, the relatively competitive performance of MLR with the initially screened predictors similarly suggests that more complex algorithms do not always provide improvements.

The comparison changed after RFE. SVM showed the best predictive performance, followed by RF, whereas MLR performed substantially worse. The observed-versus-predicted relationships were consistent with this pattern, with the non-linear models showing better agreement with observed yield than MLR. Under the reduced predictor set, the greater flexibility of SVM and RF therefore appeared more advantageous. Nevertheless, given the relatively small dataset, these differences should be interpreted cautiously. A larger dataset covering a wider range of growing seasons and environmental conditions would be needed to determine whether the apparent advantage of the non-linear models over MLR is robust and generalisable to independent data. The dataset represented only a limited number of years and environmental conditions, and model ranking may change when relationships between predictors and yield differ from those represented during model development. A simpler linear model may, for example, be more stable in new conditions if the underlying relationships remain approximately linear, whereas more flexible models may be more sensitive to patterns specific to the available observations.

For crude protein concentration, RF provided the strongest predictive performance under both predictor-selection strategies. After RFE, SVM performed similarly, while MLR remained less accurate. This suggests that crude protein prediction may have benefited from models capable of representing more complex relationships among the selected spectral and meteorological predictors. Nevertheless, the observed-versus-predicted plots also showed that prediction quality varied across the observed concentration range, highlighting the importance of considering model behaviour in addition to overall performance metrics.

For starch concentration, model performance depended more strongly on predictor selection. RF clearly performed best with the initially screened predictors, whereas SVM and RF showed very similar performance after RFE. MLR was consistently less accurate. The observed-versus-predicted plots were therefore particularly useful for assessing whether models with similar overall

metrics also showed similar agreement with the 1:1 line and comparable behaviour at low and high starch concentrations.

Taken together, these results show that greater model complexity did not consistently lead to better predictive performance. For grain yield, MLR was competitive when the initially screened predictors were used, whereas the non-linear models showed a clearer advantage after RFE. RF was generally the strongest model for crude protein concentration, while the best-performing approach for starch depended on the predictor-selection strategy. Thus, no single algorithm was consistently superior across all response variables and predictor sets. Model comparison should therefore consider not only RMSE and R^2 , but also systematic deviations from the 1:1 line, performance across the observed range, model complexity, interpretability and the intended application.

4.5 External validation

The external validation provided a more demanding test of the transferability of the final SVM grain-yield model beyond the conditions represented at Ultuna. When applied to the independent Lönnstorp dataset, comprising 18 plot-year observations from 2018–2023, predictive performance was substantially lower than in the internal LOOCV, with R^2 decreasing from 0.87 to 0.21 and RMSE increasing from 957 to approximately 1,200 kg ha⁻¹. The negative bias (−532 kg ha⁻¹) also indicated an overall tendency to underestimate yield at the external site, particularly at higher observed yields. This reduction in performance suggest that relationships identified within the Ultuna dataset did not fully transfer to a different site and set of growing seasons. Differences in environmental conditions and predictor ranges may have contributed to this result, as some Lönnstorp observations extended beyond the conditions represented in the Ultuna training data. Nevertheless, the external validation is useful because it provides an independent assessment of model generalisability and demonstrates that strong internal cross-validation performance should not be interpreted as evidence of equivalent performance at new sites.

The importance of independent validation has also been highlighted in Swedish research on remote sensing based on grain quality prediction. Wolters et al (2022) developed models for winter wheat grain protein concentration using UAV multispectral reflectance data and evaluated their performance across independent experimental trials. Although the predicted variables and sensing platform differed from those used for yield prediction in the present study, both studies address the challenge of applying relationships derived from remote sensing observations beyond the original calibration data. The lower predictive performance observed at Lönnstorp therefore highlights the importance of evaluating model transferability across independent sites and growing seasons.

The external validation results also highlight the importance of expanding the temporal and environmental range represented in the training dataset. Including observations from the 2026 growing season and subsequent years would increase temporal replication and reduce the current imbalance among

years. However, additional years from Ultuna alone may not be sufficient to ensure model transferability. Future model development should therefore also include observations from independent sites and environmental conditions that extend the range represented in the training data. Such datasets would support year and site validation and provide a stronger basis for assessing model robustness.

4.6 UAV-derived spatial variability in crop development

The UAV imagery revealed spatial variability in crop greenness across the experimental site (Figure 21). Differences were visible both among and within the experimental blocks, with Block B generally showing higher VARI values, while Block C tended to show lower and more heterogeneous values. These spatial patterns indicate that crop development was not uniform across the experimental area and that plot-level differences in canopy condition were already detectable during the early growing season.

The comparison between UAV and Sentinel-2 imagery highlighted a clear difference in the level of spatial detail captured by the two platforms. Sentinel-2-derived VARI represented broader spatial patterns across the experimental plots, whereas the UAV imagery revealed considerably finer within-plot variation. This difference is particularly relevant in the present experimental setting, where the plots are relatively small in relation to the spatial resolution of Sentinel-2. Individual satellite pixels may therefore integrate reflectance from different parts of a plot or from neighbouring surfaces, reducing the ability to distinguish small-scale spatial variation. In contrast, the centimetre-scale UAV imagery allowed within-plot heterogeneity to be visualised more clearly. This finer spatial detail could be particularly useful during field monitoring or when planning field measurements, as localised differences in VARI may help identify areas within individual plots that require closer investigation. This approach could improve the selection of sampling locations and provide information on spatial variation. However, these potential applications were not directly evaluated in the present study.

Despite these differences in spatial resolution, the two platforms showed a similar overall spatial pattern. In the May comparison, a positive fitted relationship was observed between Sentinel-2-derived and UAV-derived VARI, with a regression slope of $\beta = 0.43$. Plots with higher Sentinel-2 VARI therefore generally also showed higher UAV-derived VARI (Figure 21). Block B tended to occupy the higher range of both UAV and Sentinel-2 VARI, whereas several observations from Block C occurred at the lower end of the distributions. This suggests that both platforms captured a common broad spatial gradient in canopy greenness, although the UAV imagery provided substantially greater spatial detail.

The UAV–Sentinel-2 comparison should nevertheless be interpreted cautiously. The May UAV and Sentinel-2 observations were separated by ten days. Part of the difference between the two datasets may therefore reflect crop development between acquisition dates rather than differences between the sensing platforms alone. In addition, the UAV imagery was acquired without

RTK positioning or radiometric calibration, which may have introduced additional spatial or radiometric uncertainty.

VARI primarily reflects variation in visible canopy greenness and should therefore be interpreted as an indirect indicator of crop condition. Higher VARI values may indicate a greener or more developed canopy, but they do not represent a direct measurement of biomass, grain yield or grain quality.

4.7 Implications for testbed development and future research

The results indicate that Sentinel-2 imagery combined with local meteorological data can capture variability associated with winter wheat grain yield and quality within the Ultuna experimental site. Several predictors retained in the final models originated from the calendar defined as vegetative period, corresponding to March–May for meteorological data and April–May for Sentinel-2 observations. This suggests that information collected during spring growth may already contain useful indications of final crop performance. This is agronomically plausible because canopy development, tillering and environmental conditions during this period can influence subsequent biomass production and yield formation. However, the potential for reliable early season yield forecasting requires further evaluation.

These findings are consistent with previous studies showing that satellite observations can provide useful information for winter wheat yield estimation. Under Swedish conditions, Bouras et al. (2023) showed that assimilating Sentinel-2 derived LAI improved winter wheat yield estimation in southern Sweden. Similar findings have been reported in Europe. Marszalek et al. (2022), for example, demonstrated that Sentinel-2 observations combined with climatological variables could be used for multi-year winter wheat yield prediction in southern Germany. Although these studies used different modelling approaches and do not directly establish the predictive value of the specific spring variables selected here, they provide a broader context for the potential of integrating spectral and environmental information.

Future modelling studies could also investigate the integration of Sentinel-1 synthetic aperture radar (SAR) observations with Sentinel-2 imagery and meteorological data. Sentinel-1 could provide complementary information on crop structure and moisture related conditions, with the additional advantage of acquiring observations under cloudy conditions. Furthermore, alternative machine learning, such as eXtreme Gradient Boosting (XGBoost), could be evaluated alongside MLR, RF and SVM to assess whether they improve predictions of grain yield and quality.

Beyond the predictive modelling, an important contribution of the thesis was the reconstruction and systematic organisation of the historical information available for the RISE testbed. The available datasets had not previously been analysed, and the integration of yield, grain-quality, meteorological, soil and remote sensing information revealed several limitations related to experimental design, field documentation, sensor deployment and data management.

Identifying these limitations provides a basis for improving both the scientific value and the practical operation of the testbed.

Recommended improvements include introducing randomisation when experimental treatments are established, documenting field operations, more consistently, improving the positioning and recording of soil sensors, and using RTK positioning and calibration panels during UAV data acquisition.

Some improvements have started to be implemented during the 2026 growing season. A GPS-based harvest mapping system from Lövsta farm was used during the 2026 harvest to improve the spatial documentation of yield data. The improved organisation and documentation of the testbed also supported activities with three external clients within the European AgrifoodTEF project, illustrating how this work can contribute to future experimental and demonstration activities.

Future studies should also investigate whether high-resolution UAV imagery provides additional information beyond Sentinel-2, particularly in small plots where satellite observations may be affected by mixed pixels and non-crop vegetation. One promising approach is to link field measurements, UAV observations and Sentinel-2 imagery through crop biophysical parameters such as Green Area Index (GAI). Direct field measurements could subsequently support the calibration or validation of Sentinel-derived estimates. Bukowiecki et al. (2021) demonstrated the potential of using UAV-derived GAI information to calibrate Sentinel-2 observations in winter wheat. This approach could provide an intermediate spatial scale between field measurements and satellite imagery.

Ongoing development also includes the cleaning and restructuring of the historical meteorological dataset. These data are intended to be integrated with observations from three fixed field camera systems, including RGB, vegetation-index and night-time imagery, together with UAV imagery, Sentinel-2 observations and soil measurements. One potential application is to investigate whether combinations of these data sources can support earlier detection of crop stress or disease occurrence.

To support such applications, more systematic ground-truth observations will be required. These should include crop phenology, visual disease scoring, dates of disease occurrence and progression, as well as detailed records of management operations and their timing. Chlorophyll measurements using a SPAD chlorophyll meter (Konica Minolta, Japan) are also planned to provide field-based information on relative leaf chlorophyll content and crop nitrogen status. Additional measurements such as soil infiltration tests and soil penetration resistance could improve the interpretation of both crop observations and soil sensor data.

Together, these developments could strengthen the role of the testbed as a platform for integrating agronomic measurements, environmental monitoring and remote sensing.

4.8 Limitations

The modelling dataset covered four harvest years (2020, 2022, 2023 and 2025) and was strongly unbalanced among years, with 2025 contributing substantially more observations than the other seasons. This limited temporal replication makes it difficult to separate year effects from differences associated with block, management and environmental conditions.

Model performance was evaluated using LOOCV at the individual plot-year level. Consequently, observations from the same harvest year could occur in both the training and validation datasets. This is particularly relevant for the meteorological predictors, because all plots within the same year shared the same weather values.

Therefore, the reported performance primarily represents the ability of the models to predict plot-level variability within the years, and environmental conditions already represented in the dataset, rather than their ability to predict an entirely unseen growing season.

Leave-one-year-out validation would have provided a more direct assessment of temporal transferability. However, with only four harvest years and markedly different sample sizes among years, such validation would have relied on a very small number of heterogeneous validation groups. Future studies including additional and more evenly represented growing seasons would allow a more robust evaluation of model performance across years.

A further limitation concerns the relationship between RFE and model validation. RFE was performed using repeated five-fold cross-validation before the final LOOCV of the predictive models. Feature selection was therefore not nested within the final validation procedure, meaning that observations later used for validation indirectly contributed to the selection of the predictor set. This may result in somewhat optimistic estimates of predictive performance for the RFE-based models. The RFE results should therefore be interpreted as estimates of internal predictive performance within the available dataset rather than fully independent estimates of model performance in new growing seasons.

The interpretation of yield and grain-quality variability was also limited by incomplete agronomic metadata. Information on cultivar, sowing date, tillage, crop protection and some harvest procedures was not consistently available. Fertilisation records were also incomplete as the nitrogen application rates and composition of fertilisers used were not documented for year 2020–2023. Consequently, some variation attributed to spectral or meteorological predictors may also reflect unrecorded differences in crop management, particularly for crude protein concentration.

The spatial resolution of Sentinel-2 represents an additional source of uncertainty. The bands used in the analysis had native spatial resolutions between 10 and 60 m, although the imagery was exported at a spatial scale of 10 m. A 0.5 m inward buffer was applied before spectral extraction to reduce boundary effects. However, this distance was small relative to the Sentinel-2-pixel dimensions and therefore could not ensure that all contributing pixels represented only winter wheat. A substantially wider buffer would also have reduced the already limited area available for spectral extraction within each

plot. Plot-level values may therefore still have been influenced by mixed pixels containing neighbouring vegetation, plot margins, tramlines or other surfaces.

A further limitation concerns the definition of crop development stages. The stages used for temporal aggregation were based on fixed calendar periods rather than field-observed phenology. This represents a pragmatic approach that is commonly used in remote sensing studies when detailed phenological observations are unavailable. In the present study, most of the historical data had been collected before the start of this work and were not originally collected for the specific purpose of linking remote sensing observations with crop phenology. Because the timing of wheat development can vary among years depending on temperature, sowing date and other growing conditions, the same calendar period may therefore have represented slightly different developmental stages in different years. Consequently, observations assigned to the same calendar period may not have represented equivalent physiological stages across years, potentially introducing additional variability into the spectral and meteorological predictors and affecting model performance.

An additional potential limitation concerns weed presence within the experimental area. Visual inspection of the UAV RGB orthomosaics acquired during the 2025 growing season revealed the presence of weeds in different experimental plots (Figure 23). However, weed abundance and spatial distribution were not systematically quantified, and this information was not included in the dataset used for predictive modelling. Weed vegetation may have contributed to the spectral signal recorded by Sentinel-2, particularly where crop and weed vegetation occurred within the same pixel. Liu et al. (2024) reported similarities in the spectral characteristics of wheat and grass weeds, highlighting the difficulty of distinguishing them using vegetation indices. Furthermore, competition for light, water and nutrients may have influenced winter wheat growth and grain yield, as Lui et al. (2024) demonstrated that increasing weed infestation could negatively affect wheat biomass and grain yield. Weed presence therefore represents a potential source of agronomic and spectral variability that was not explicitly accounted for in the models. This variability may have contributed to the prediction errors and the relatively high RMSE observed for grain yield, as the spectral information may not always have accurately represented winter wheat growth alone. However, because weed abundance was not quantified and observations were limited to the UAV imagery, its contribution to model prediction errors could not be directly assessed.

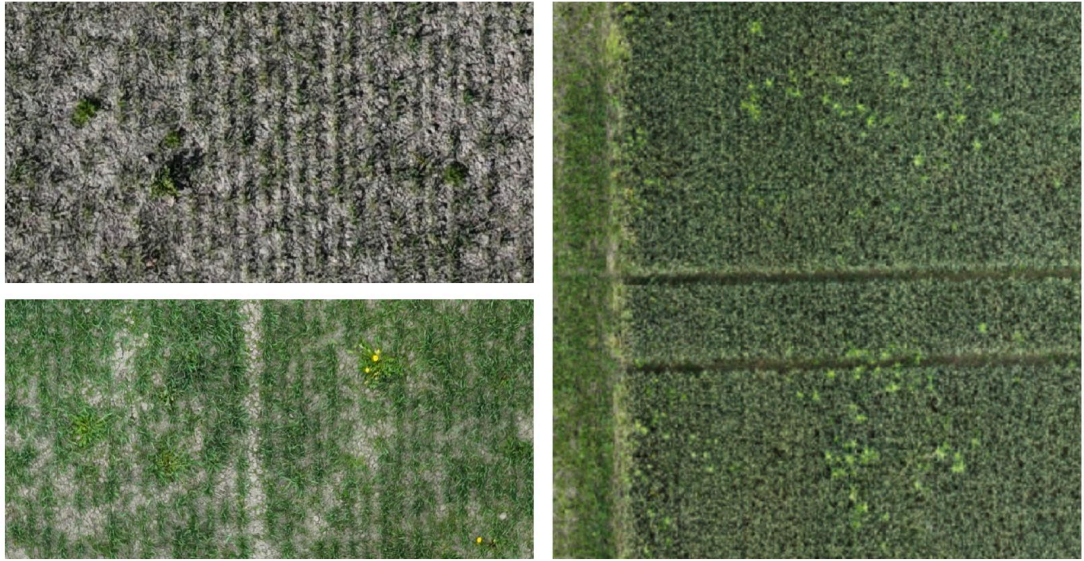


Figure 23. Examples of weed presence in UAV RGB orthomosaics acquired at the RISE testbed during the 2025 growing season. The images represent different experimental plots and acquisition dates.

Finally, the comparison between UAV and Sentinel-2 imagery was subject to additional methodological limitations. The acquisitions were not simultaneous, with a ten-day difference in May. Consequently, part of the observed difference between platforms may reflect crop development between acquisition dates rather than sensor or spatial-resolution effects alone. In addition, the UAV imagery was acquired without RTK positioning or radiometric calibration.

5. Conclusion

This study evaluated the potential of combining Sentinel-2 imagery and meteorological data aggregated by over calendar-defined periods to estimate winter wheat grain yield, crude protein concentration and starch concentration at an experimental site in central Sweden. Overall, the results showed that spectral and meteorological information collected during crop development contained substantial information related to plot-level variation in both grain yield and grain quality. Predictors representing the vegetative growth stage were particularly important across the three response variables, while information from the grain-filling stage also contributed to some grain-quality models.

RFE improved the best internal predictive performance for all three response variables, although the magnitude of improvement differed among traits. Using the RFE-selected predictors, SVM provided the best performance for grain yield ($\text{RMSE} = 957 \text{ kg ha}^{-1}$; $R^2 = 0.87$), RF performed best for crude protein concentration ($\text{RMSE} = 0.45 \text{ pp}$; $R^2 = 0.84$), and SVM achieved the best performance for starch concentration ($\text{RMSE} = 0.67 \text{ pp}$; $R^2 = 0.77$). However, no modelling algorithm was consistently superior across all response variables and predictor-selection strategies. MLR remained competitive for grain yield when the initially screened predictors were used, indicating that, within the predictor sets evaluated in this study, greater model complexity did not necessarily lead to improved predictive performance. However, this finding does not exclude the possibility that more complex models could perform better with a different set of predictors.

As expected from established agronomic knowledge, the complementary analyses further demonstrated that grain yield was strongly associated with both fertiliser treatment and harvest year, while the response of crude protein concentration to fertiliser treatment differed among years. UAV imagery revealed fine-scale spatial differences in vegetation pattern that were less clearly represented by Sentinel-2, illustrating the complementary value of high-resolution UAV observations for small experimental plots.

The internal predictive results should be interpreted within the environmental conditions represented in the Ultuna dataset. External validation at SITES SAFE Lönnstorp showed substantially lower predictive performance, demonstrating that the strong internal performance did not fully transfer to an independent site and set of growing seasons. The limited and unbalanced number of harvest years, plot-level cross-validation, non-nested feature selection, and incomplete agronomic metadata further constrain the generalisability of the models. The spatial resolution of Sentinel-2 relative to the small experimental plots may also affect transferability because plot-level spectral values could partly reflect mixed pixels and surrounding surfaces. Relationships learned under the specific spatial configuration of the Ultuna testbed may therefore not be reproduced in sites with different plot sizes or surrounding vegetation. Further validation across additional years and environments, using year-blocked or external validation and more complete information on crop management, is therefore required.

References

- Addy, J.W.G., Ellis, R.H., Macdonald, A.J., Semenov, M.A. & Mead, A. (2020). Investigating the effects of inter-annual weather variation (1968–2016) on the functional response of cereal grain yield to applied nitrogen, using data from the Rothamsted Long-Term Experiments. *Agricultural and Forest Meteorology*, 284, 107898.
<https://doi.org/10.1016/j.agrformet.2019.107898>
- AgrifoodTEF (2026). *About agrifoodTEF | agrifoodTEF*. AgrifoodTEF.
<https://agrifoodtef.eu/about-agrifoodtef> [2026-06-29]
- Allen, Raes D., Pereira L. S., & Smith M. (1998). *Crop evapotranspiration-Guidelines for computing crop water requirements-FAO I56*. FAO.
https://www.researchgate.net/publication/235704197_Crop_evapotranspiration-Guidelines_for_computing_crop_water_requirements-FAO_Irrigation_and_drainage_paper_56 [2026-09-12]
- Alshihabi, O., Persson, K. & Söderström, M. (2024). Easy yield mapping for precision agriculture. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, 74 (1), 2411950.
<https://doi.org/10.1080/09064710.2024.2411950>
- Barbottin, Lecomte C., Bouchard C., & Jeuffroy M .H. (2005). Nitrogen Remobilization during Grain Filling in Wheat. *Crop Science - CROP SCI*, 45.
https://www.researchgate.net/publication/238080894_Nitrogen_Remobilization_during_Grain_Filling_in_Wheat
- Barlow, K.M., Christy, B.P., O’Leary, G.J., Riffkin, P.A. & Nuttall, J.G. (2015). Simulating the impact of extreme heat and frost events on wheat crop production: A review. *Field Crops Research*, 171, 109–119.
<https://doi.org/10.1016/j.fcr.2014.11.010>
- Barnes, E., Clarke, T.R., Richards, S.E., Colaizzi, P., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E. & Thompson, T.L. (2000). Coincident detection of crop water stress, nitrogen status, and canopy density using ground based multispectral data. *Proceedings of Conference Name: Fifth International Conference on Precision Agriculture*, Bloomington, Minnesota, USA, 2000.
<https://www.tucson.ars.ag.gov/unit/publications/PDFfiles/1356.pdf?>
- Barreiro, A., & Albertsson, J. (2022). *SITES Agroecological Field Experiment (SAFE)*. Department of Biosystems and Technology, Swedish University of Agricultural Sciences. <https://doi.org/10.23700/4ASX-8870>
- Bogard, M., Allard, V., Brancourt-Hulmel, M., Heumez, E., Machet, J.-M., Jeuffroy, M.-H., Gate, P., Martre, P. & Le Gouis, J. (2010). Deviation from the grain protein concentration–grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. *Journal of Experimental Botany*, 61 (15), 4303–4312.
<https://doi.org/10.1093/jxb/erq238>
- Bossung, C., Schlerf, M. & Machwitz, M. (2022). Estimation of canopy nitrogen content in winter wheat from Sentinel-2 images for operational agricultural monitoring. *Precision Agriculture*, 23 (6), 2229–2252.
<https://doi.org/10.1007/s11119-022-09918-y>
- Bouras, E. houssaine, Olsson, P.-O., Thapa, S., Díaz, J.M., Albertsson, J. & Eklundh, L. (2023). Wheat Yield Estimation at High Spatial Resolution through the Assimilation of Sentinel-2 Data into a Crop Growth Model. *Remote Sensing*, 15 (18), 4425. <https://doi.org/10.3390/rs15184425>
- Bruning, B., Liu, H., Brien, C., Berger, B., Lewis, M. & Garnett, T. (2019). The Development of Hyperspectral Distribution Maps to Predict the Content

- and Distribution of Nitrogen and Water in Wheat (*Triticum aestivum*). *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.01380>
- Bukowiecki, J., Rose, T. & Kage, H. (2021). Sentinel-2 Data for Precision Agriculture?—A UAV-Based Assessment. *Sensors (Basel, Switzerland)*, 21 (8), 2861. <https://doi.org/10.3390/s21082861>
- Clevers, J.G.P.W., Kooistra, L. & Van den Brande, M.M.M. (2017). Using Sentinel-2 Data for Retrieving LAI and Leaf and Canopy Chlorophyll Content of a Potato Crop. *Remote Sensing*, 9 (5), 405. <https://doi.org/10.3390/rs9050405>
- Cormier, F., Foulkes, J., Hirel, B., Gouache, D., Moënné-Loccoz, Y. & Le Gouis, J. (2016). Breeding for increased nitrogen-use efficiency: a review for wheat (*T. aestivum* L.). *Plant Breeding*, 135 (3), 255–278. <https://doi.org/10.1111/pbr.12371>
- Curran, P.J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30 (3), 271–278. [https://doi.org/10.1016/0034-4257\(89\)90069-2](https://doi.org/10.1016/0034-4257(89)90069-2)
- Delegido, J., Verrelst, J., Alonso, L. & Moreno, J. (2011a). Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content. *Sensors*, 11 (7), 7063–7081. <https://doi.org/10.3390/s110707063>
- Delegido, J., Verrelst, J., Alonso, L. & Moreno, J. (2011b). Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content. *Sensors*, 11 (7), 7063–7081. <https://doi.org/10.3390/s110707063>
- Delin, S., Lindén, B. & Berglund, K. (2005). Yield and protein response to fertilizer nitrogen in different parts of a cereal field: potential of site-specific fertilization. *European Journal of Agronomy*, 22 (3), 325–336. <https://doi.org/10.1016/j.eja.2004.05.001>
- ESA (2015). Sentinel-2 User Handbook. Document. https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook
- ESA/Copernicus (2026). *Sentinel-2. documentation*. <https://documentation.dataspace.copernicus.eu/Data/SentinelMissions/Sentinel2.html#sentinel-2-level-2a-top-of-canopy-toc> [2026-07-01]
- FAO (2023). Food balance sheets 2010–2023. <https://doi.org/10.4060/cd7162en>
- Filip, E., Woronko, K., Stepień, E. & Czarniecka, N. (2023). An Overview of Factors Affecting the Functional Quality of Common Wheat (*Triticum aestivum* L.). *International Journal of Molecular Sciences*, 24 (8), 7524. <https://doi.org/10.3390/ijms24087524>
- Frampton, W.J., Dash, J., Watmough, G. & Milton, E.J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 83–92. <https://doi.org/10.1016/j.isprsjprs.2013.04.007>
- Fu, J., Bowden, R.L., Jagadish, S.V.K. & Prasad, P.V.V. (2023). Genetic variation for terminal heat stress tolerance in winter wheat. *Frontiers in Plant Science*, 14. <https://doi.org/10.3389/fpls.2023.1132108>
- Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58 (3), 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, F., Pflug, B., Bieniarz, J., Clerc, S., Pessiot, L., Trémas, T., Cadau, E., De Bonis, R., Isola, C., Martimort, P. & Fernandez, V. (2017). Copernicus Sentinel-2A Calibration and Products

- Validation Status. *Remote Sensing*, 9 (6), 584.
<https://doi.org/10.3390/rs9060584>
- Ghimire, D., Das, S., Mueller, N.D., Creech, C.F., Santra, D., Baenziger, P.S., Easterly, A.C., Maust, B. & Maharjan, B. (2021). Effects of cultivars and nitrogen management on wheat grain yield and protein. *Agronomy Journal*, 113 (5), 4348–4368. <https://doi.org/10.1002/agj2.20836>
- Gitelson, A.A., Kaufman, Y.J., Stark, R., & Rundquist, D. (2002). Novel Algorithms for Remote Estimation of Vegetation Fraction. 80, 76–87
- Gorelick, Noel, Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27.
<https://doi.org/10.1016/j.rse.2017.06.031>
- Grossiord, C., Buckley, T.N., Černusak, L.A., Novick, K.A., Poulter, B., Siegwolf, R.T.W., Sperry, J.S. & McDowell, N.G. (2020). Plant responses to rising vapor pressure deficit. *New Phytologist*, 226 (6), 1550–1566.
<https://doi.org/10.1111/nph.16485>
- Haghighattalab, A., González Pérez, L., Mondal, S., Singh, D., Schinstock, D., Rutkoski, J., Ortiz-Monasterio, I., Singh, R.P., Goodin, D. & Poland, J. (2016). Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. *Plant Methods*, 12 (1), 35.
<https://doi.org/10.1186/s13007-016-0134-6>
- Hamnér, K., Weih, M., Eriksson, J. & Kirchmann, H. (2017). Influence of nitrogen supply on macro- and micronutrient accumulation during growth of winter wheat. *Field Crops Research*, 213, 118–129.
<https://doi.org/10.1016/j.fcr.2017.08.002>
- Hassan, M.A., Yang, M., Rasheed, A., Tian, X., Reynolds, M., Xia, X., Xiao, Y. & He, Z. (2021). Quantifying senescence in bread wheat using multispectral imaging from an unmanned aerial vehicle and QTL mapping. *Plant Physiology*, 187 (4), 2623–2636.
<https://doi.org/10.1093/plphys/kiab431>
- Hyles, J., Bloomfield, M.T., Hunt, J.R., Trethowan, R.M. & Trevaskis, B. (2020). Phenology and related traits for wheat adaptation. *Heredity*, 125 (6), 417–430. <https://doi.org/10.1038/s41437-020-0320-1>
- Jordbruksverket (2025). *Jordbruksmarkens användning 2025. Slutlig statistik*. [text]. <https://jordbruksverket.se/om-jordbruksverket/jordbruksverkets-officiella-statistik/jordbruksverkets-statistikrapporter/statistik/2025-10-22-jordbruksmarkens-anvandning-2025.-slutlig-statistik> [2026-09-10]
- Joshi, J., Stocker, B.D., Hofhansl, F., Zhou, S., Dieckmann, U. & Prentice, I.C. (2022). Towards a unified theory of plant photosynthesis and hydraulics. *Nature Plants*, 8 (11), 1304–1316. <https://doi.org/10.1038/s41477-022-01244-5>
- Khalid, A., Hameed, A. & Tahir, M.F. (2023). Wheat quality: A review on chemical composition, nutritional attributes, grain anatomy, types, classification, and function of seed storage proteins in bread making quality. *Frontiers in Nutrition*, 10.
<https://doi.org/10.3389/fnut.2023.1053196>
- Lawes, J.B. & Gilbert, J.H. (1864). *Report of Experiments on the Growth of Wheat for Twenty Years in Succession on the Same Land*. Clowes.
- Leske, B.A. & Biddulph, T.B. (2022). Estimating Effects of Radiation Frost on Wheat Using a Field-Based Frost Control Treatment to Stop Freezing Damage. *Genes*, 13 (4), 578. <https://doi.org/10.3390/genes13040578>
- Li, M., Shamshiri, R.R., Weltzien, C. & Schirrmann, M. (2022). Crop Monitoring Using Sentinel-2 and UAV Multispectral Imagery: A Comparison Case Study in Northeastern Germany. *Remote Sensing*, 14 (17), 4426.
<https://doi.org/10.3390/rs14174426>

- Liu, S., Peng, Y., Du, W., Le, Y. & Li, L. (2015). Remote Estimation of Leaf and Canopy Water Content in Winter Wheat with Different Vertical Distribution of Water-Related Properties. *Remote Sensing*, 7 (4), 4626–4650. <https://doi.org/10.3390/rs70404626>
- Liu, T., Zhao, Y., Wang, H., Wu, W., Yang, T., Zhang, W., Zhu, S., Sun, C. & Yao, Z. (2024). Harnessing UAVs and deep learning for accurate grass weed detection in wheat fields: a study on biomass and yield implications. *Plant Methods*, 20 (1), 144. <https://doi.org/10.1186/s13007-024-01272-6>
- Long, D.S., Reardon, C.L. & Adams, C.B. (2024). Critical levels and nitrogen fertilizer equivalencies for grain protein in winter wheat. *Agronomy Journal*, 116 (1), 339–348. <https://doi.org/10.1002/agj2.21499>
- Lu, C., Yang, M., Dong, S., Liu, Y., Li, Y. & Pan, Y. (2025). Improving Winter Wheat Yield Estimation Under Saline Stress by Integrating Sentinel-2 and Soil Salt Content Using Random Forest. *Agriculture*, 15 (14), 1544. <https://doi.org/10.3390/agriculture15141544>
- Maeoka, R.E., Sadras, V.O., Ciampitti, I.A., Diaz, D.R., Fritz, A.K. & Lollato, R.P. (2020). Changes in the Phenotype of Winter Wheat Varieties Released Between 1920 and 2016 in Response to In-Furrow Fertilizer: Biomass Allocation, Yield, and Grain Protein Concentration. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.01786>
- Marszalek, M., Körner, M. & Schmidhalter, U. (2022). Prediction of multi-year winter wheat yields at the field level with satellite and climatological data. *Computers and Electronics in Agriculture*, 194, 106777. <https://doi.org/10.1016/j.compag.2022.106777>
- Martins, V.S., Barbosa, C.C.F., De Carvalho, L.A.S., Jorge, D.S.F., Lobo, F.D.L. & Novo, E.M.L. de M. (2017). Assessment of Atmospheric Correction Methods for Sentinel-2 MSI Images Applied to Amazon Floodplain Lakes. *Remote Sensing*, 9 (4), 322. <https://doi.org/10.3390/rs9040322>
- McMaster & Wilhelm (1997). Growing degree-days: one equation, two interpretations. *Agricultural and Forest Meteorology*, 87, 291–300
- Meng, Q., Yue, S., Chen, X., Cui, Z., Ye, Y., Ma, W., Tong, Y. & Zhang, F. (2013). Understanding Dry Matter and Nitrogen Accumulation with Time-Course for High-Yielding Wheat Production in China. *PLOS ONE*, 8 (7), e68783. <https://doi.org/10.1371/journal.pone.0068783>
- Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114, 358–371
- NASA (1999). Remote Sensing - NASA Science. <https://science.nasa.gov/earth/earth-observatory/remote-sensing/> [2026-08-22]
- Pancorbo, J.L., Alonso-Ayuso, M., Camino, C., Raya-Sereno, M.D., Zarco-Tejada, P.J., Molina, I., Gabriel, J.L. & Quemada, M. (2023). Airborne hyperspectral and Sentinel imagery to quantify winter wheat traits through ensemble modeling approaches. *Precision Agriculture*, 24 (4), 1288–1311. <https://doi.org/10.1007/s11119-023-09990-y>
- Papadopoulos, G., Arduini, S., Uyar, H., Psiroukis, V., Kasimati, A. & Fountas, S. (2024). Economic and environmental benefits of digital agricultural technologies in crop production: A review. *Smart Agricultural Technology*, 8, 100441. <https://doi.org/10.1016/j.atech.2024.100441>
- QGIS Development Team (2026). *QGIS Geographic Information System* (4.0.3 “Norrköping”). QGIS Association. <https://www.qgis.org>
- R Core Team (2026). *R: A Language and Environment for Statistical Computing*.
- Ramezan, C.A. (2022). Transferability of Recursive Feature Elimination (RFE)-Derived Feature Sets for Support Vector Machine Land Cover

- Classification. *Remote Sensing*, 14 (24), 6218.
<https://doi.org/10.3390/rs14246218>
- Rouse, J.W., Jr., Haas, R.H., Schell, J.A. & Deering, D.W. (1974). Monitoring Vegetation Systems in the Great Plains with ERTS. In: *NASA Special Publication*. 309.
<https://ui.adsabs.harvard.edu/abs/1974NASSP.351..309R> [2026-07-01]
- Savill, G.P., Michalski, A., Powers, S.J., Wan, Y., Tosi, P., Buchner, P. & Hawkesford, M.J. (2018). Temperature and nitrogen supply interact to determine protein distribution gradients in the wheat grain endosperm. *Journal of Experimental Botany*, 69 (12), 3117–3126.
<https://doi.org/10.1093/jxb/ery127>
- Segarra, J., Araus, J.L. & Kefauver, S.C. (2022). Farming and Earth Observation: Sentinel-2 data to estimate within-field wheat grain yield. *International Journal of Applied Earth Observation and Geoinformation*, 107, 102697.
<https://doi.org/10.1016/j.jag.2022.102697>
- Sheng, K., Xu, L., Wang, M., Lei, H. & Duan, A. (2022). The end-use quality of wheat can be enhanced by optimal water management without incurring yield loss. *Frontiers in Plant Science*, 13.
<https://doi.org/10.3389/fpls.2022.1030763>
- Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M. & Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security*, 5 (3), 291–317. <https://doi.org/10.1007/s12571-013-0263-y>
- Slafer, G.A., Savin, R. & Sadras, V.O. (2023). Wheat yield is not causally related to the duration of the growing season. *European Journal of Agronomy*, 148, 126885. <https://doi.org/10.1016/j.eja.2023.126885>
- SMHI (2026). *Upplands klimat. SMHI*. [text].
<https://www.smhi.se/kunskapsbanken/klimat/klimatet-i-sveriges-landskap/upplands-klimat> [2026-06-29]
- Srivastava, A.K., Safaei, N., Khaki, S., Lopez, G., Zeng, W., Ewert, F., Gaiser, T. & Rahimi, J. (2022). Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. *Scientific Reports*, 12 (1), 3215. <https://doi.org/10.1038/s41598-022-06249-w>
- Strandberg, G., Nilsson, C., Wilcke, R.A.I. & Wiréhn, L. (2026). Late spring frost may hinder Scandinavian farmers from making advantage of the longer vegetation period. *Communications Sustainability*, 1 (1), 133.
<https://doi.org/10.1038/s44458-026-00141-8>
- Swedish University of Agricultural Sciences (SLU) (2025). *SITES Agroekologiska Fältexperiment (SAFE) | slu.se. SLU*.
<https://www.slu.se/om-slu/organisation/institutioner/biosystem-teknologi/lonnstorp/sites-agroekologiska-faltexperiment-safe/> [2026-10-05]
- Tenreiro, T.R., García-Vila, M., Gómez, J.A., Jiménez-Berni, J.A. & Fereres, E. (2021). Using NDVI for the assessment of canopy cover in agricultural crops within modelling research. *Computers and Electronics in Agriculture*, 182, 106038. <https://doi.org/10.1016/j.compag.2021.106038>
- Wan, C., Dang, P., Gao, L., Wang, J., Tao, J., Qin, X., Feng, B. & Gao, J. (2022). How Does the Environment Affect Wheat Yield and Protein Content Response to Drought? A Meta-Analysis. *Frontiers in Plant Science*, 13.
<https://doi.org/10.3389/fpls.2022.896985>
- Wang, Y., Peng, Y., Lin, J., Wang, L., Jia, Z. & Zhang, R. (2023). Optimal nitrogen management to achieve high wheat grain yield, grain protein content, and water productivity: A meta-analysis. *Agricultural Water Management*, 290, 108587. <https://doi.org/10.1016/j.agwat.2023.108587>

- White, J.C., Wulder, M.A., Hobart, G.W., Luther, J.E., Hermosilla, T., Griffiths, P., Coops, N.C., Hall, R.J., Hostert, P., Dyk, A. & Guindon, L. (2014). Pixel-Based Image Compositing for Large-Area Dense Time Series Applications and Science. *Canadian Journal of Remote Sensing*, 40 (3), 192–212. <https://doi.org/10.1080/07038992.2014.945827>
- Wolters, S., Söderström, M., Piikki, K., Börjesson, T. & Pettersson, C.-G. (2022). Predicting grain protein concentration in winter wheat (*Triticum aestivum* L.) based on unpiloted aerial vehicle multispectral optical remote sensing. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, 72 (1), 788–802. <https://doi.org/10.1080/09064710.2022.2085165>
- Wu, B., Zhang, M., Zeng, H., Tian, F., Potgieter, A.B., Qin, X., Yan, N., Chang, S., Zhao, Y., Dong, Q., Boken, V., Plotnikov, D., Guo, H., Wu, F., Zhao, H., Deronde, B., Tits, L. & Loupian, E. (2023). Challenges and opportunities in remote sensing-based crop monitoring: a review. *National Science Review*, 10 (4), nwac290. <https://doi.org/10.1093/nsr/nwac290>
- Xie, Q., Mayes, S. & Sparkes, D.L. (2016). Optimizing tiller production and survival for grain yield improvement in a bread wheat × spelt mapping population. *Annals of Botany*, 117 (1), 51–66. <https://doi.org/10.1093/aob/mcv147>
- Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R., Feng, H., Zhao, X., Li, Z., Li, H. & Yang, H. (2017). Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives. *Frontiers in Plant Science*, 8. <https://doi.org/10.3389/fpls.2017.01111>
- Yokamo, S., Irfan, M., Huan, W., Wang, B., Wang, Y., Ishfaq, M., Lu, D., Chen, X., Cai, Q. & Wang, H. (2023). Global evaluation of key factors influencing nitrogen fertilization efficiency in wheat: a recent meta-analysis (2000-2022). *Frontiers in Plant Science*, 14. <https://doi.org/10.3389/fpls.2023.1272098>
- Yu, H., Gao, Z., Zhao, J., Wang, Z., Li, X., Xu, X., Jian, H., Bian, D., Cui, Y. & Du, X. (2023). The Effects of Phased Warming during Late Winter and Early Spring on Grain Yield and Quality of Winter Wheat (*Triticum aestivum* L.). *Agronomy*, 13 (7), 1909. <https://doi.org/10.3390/agronomy13071909>
- Zhang, X., Alexander, L., Hegerl, G.C., Jones, P., Klein Tank A.M.C, Peterson, T.C., & Zwiers, F.W. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, 2 (6). <https://doi.org/10.1002/wcc.147>

AI declaration

Generative artificial intelligence, including Copilot (Microsoft) and Gemini (Google) was used during the preparation of this thesis to support translation, language editing, troubleshooting and refinement of R and GEE JavaScript code. In particular, Gemini was used during the development of the R scripts for debugging, syntax correction and code refinement.

The implementation of MLR, RF and SVM was also supported by available coding examples and repositories on GitHub. These resources were used as references for programming approaches and model implementation.

All code was reviewed, adapted and executed by the author. Model selection, predictor selection, statistical analyses, interpretation of results and conclusions were carried out and evaluated by the author. A statistician provided guidance on how to implement specific parts of the statistical analyses.

Popular science summary

Estimating crop yield before harvest can help improve agricultural planning and crop management. However, wheat production depends on many factors, including weather conditions and crop development.

This study investigated whether freely available Sentinel-2 satellite images and local weather data could be used to estimate winter wheat grain yield and grain quality at an experimental field in Ultuna, Uppsala. The analysis focused on grain yield, crude protein concentration and starch concentration.

Satellite data provided information about crop greenness, canopy development, chlorophyll and water status, while weather data included variables such as temperature, precipitation, frost and atmospheric dryness. Different statistical and machine-learning methods were compared.

The results showed that information from the vegetative growth stage was particularly useful. Satellite indicators related to crop vigour and chlorophyll, together with weather conditions such as frost, temperature and atmospheric water demand, were associated with differences in grain yield. Satellite and weather variables also provided useful information for estimating crude protein and starch concentration.

Drone images collected in 2025 were analysed separately and showed differences in crop development among experimental plots and blocks.

Overall, the study demonstrates the potential of combining satellite imagery and weather observations for monitoring winter wheat. However, the dataset included only four harvest years and some information on crop management was incomplete. More years of data and validation in other fields are therefore needed before the models can be used more widely.

Acknowledgements

I would like to express my deepest gratitude to Julianne Oliveira, who has accompanied me progressively from my bachelor's thesis to this stage of my academic and professional journey. Over the years, she has been a great source of inspiration and has played an important role in strengthening my passion for remote sensing. Her guidance, encouragement and trust have meant a great deal to me.

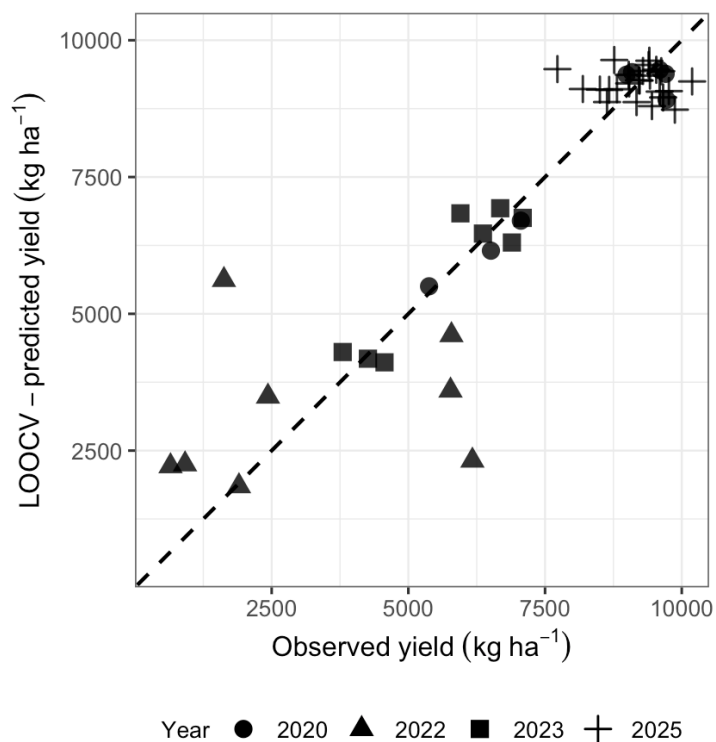
I am equally grateful to Björn Ringselle, whose supervision taught me far beyond the scope of this thesis. Through the many practical activities, I had the opportunity to take part in, he gave me the chance to learn not only about research itself, but also about how knowledge is created in practice, in the field and through collaboration.

It has been a true privilege to have two supervisors whom I have deeply admired and respected, both professionally and personally.

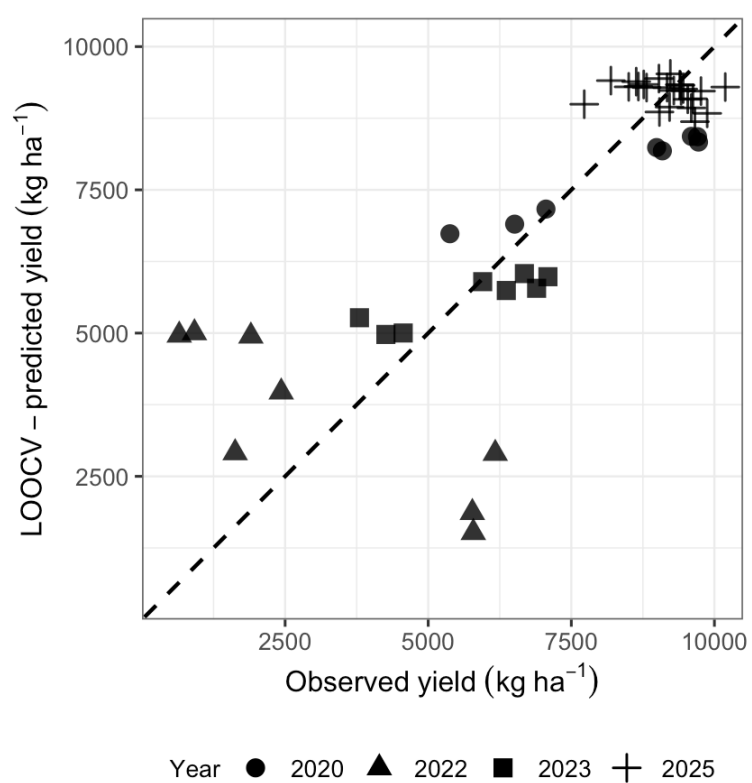
I am especially grateful to Marcos Lana for his invaluable help with the meteorological data and their interpretation, and to Johannes Forkman for his statistical guidance and support. I deeply appreciate the many hours they generously spent discussing the analyses and results with me, sometimes even giving up their breaks to help me understand the data, solve problems and develop my ideas.

I would also like to extend my warmest thanks to everyone I met at RISE. I found there a remarkable combination of brilliant minds, great competence, humanity and a great sense of humour. The people I met made this period much more than simply a thesis project. They filled these months with learning, laughter and memorable experiences, and they are the reason why I will look back on this time with so much joy and affection.

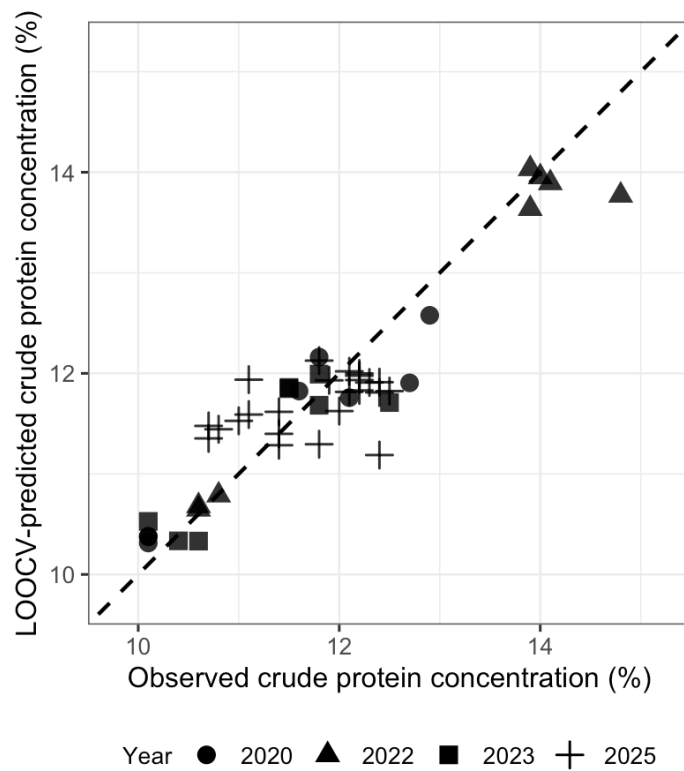
Appendix



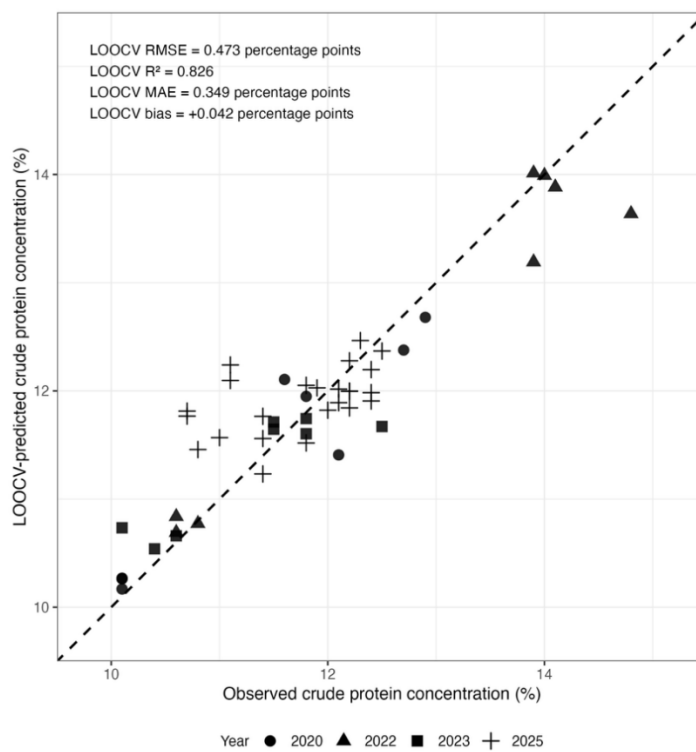
A1. Observed versus leave-one-out cross-validation (LOOCV)-predicted winter wheat grain yield for the random forest model using RFE-selected predictors.



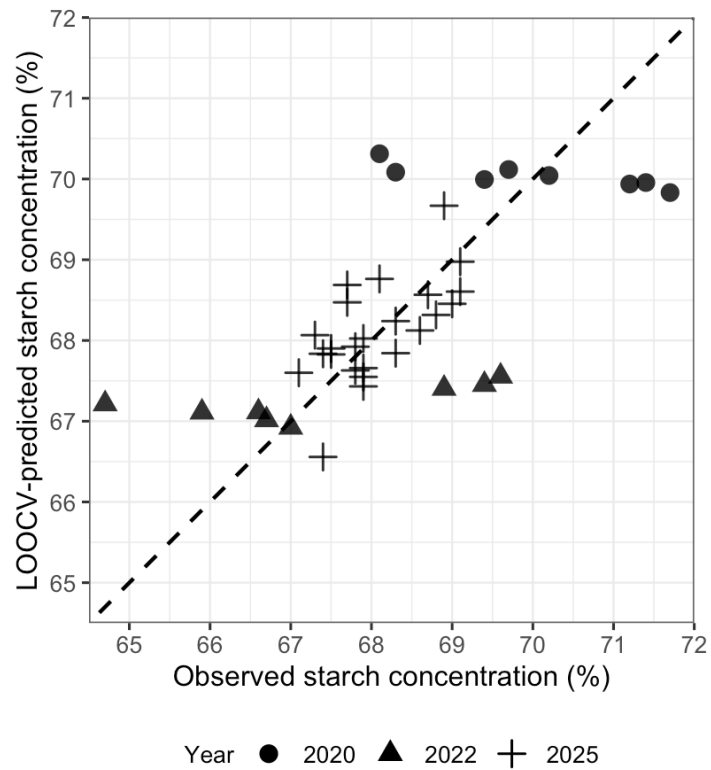
A2. Observed versus leave-one-out cross-validation (LOOCV)-predicted winter wheat grain yield for the MLR model using RFE-selected predictors.



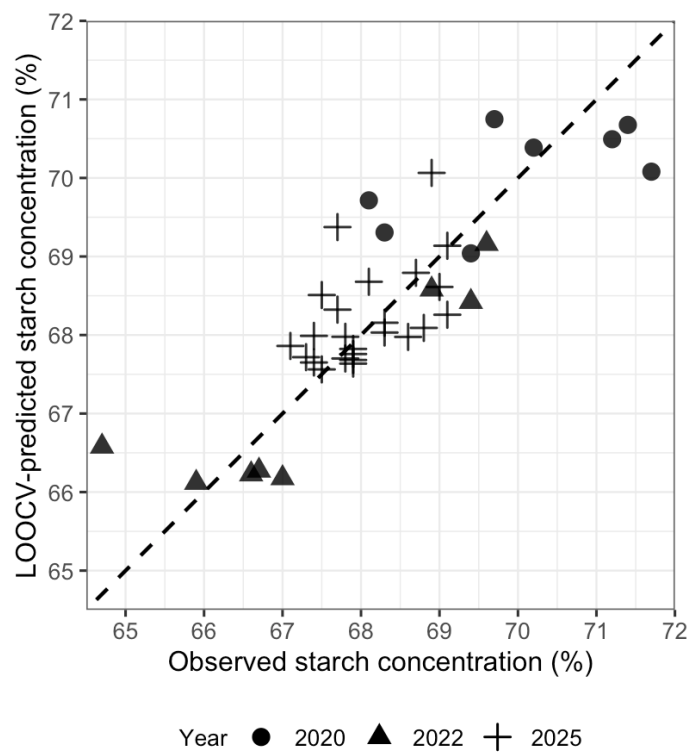
A3. Observed versus leave-one-out cross-validation (LOOCV)-predicted winter wheat grain crude protein concentration for the MLR model using RFE-selected predictors.



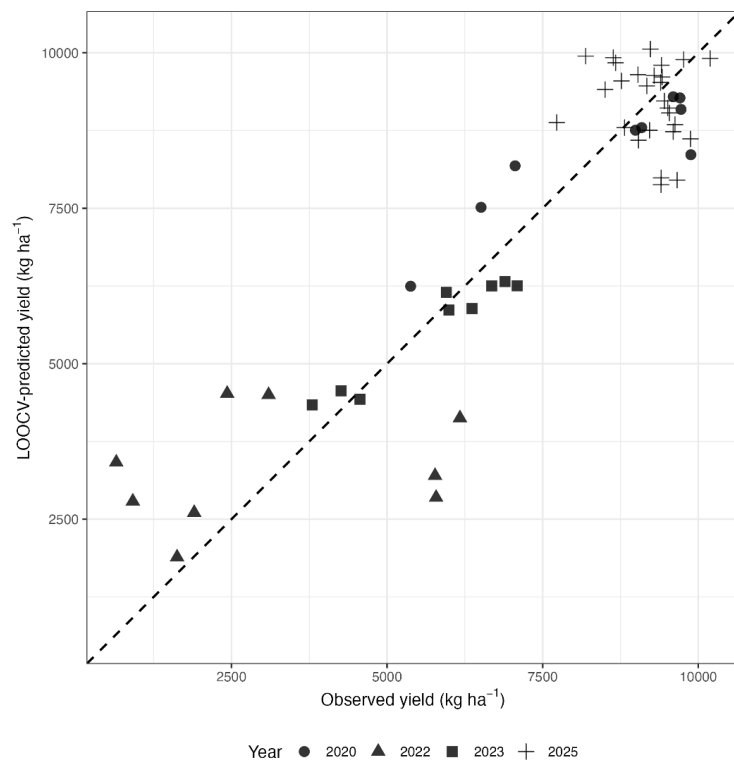
A4. Observed versus leave-one-out cross-validation (LOOCV)-predicted winter wheat grain crude protein concentration for the SVM model using RFE-selected predictors.



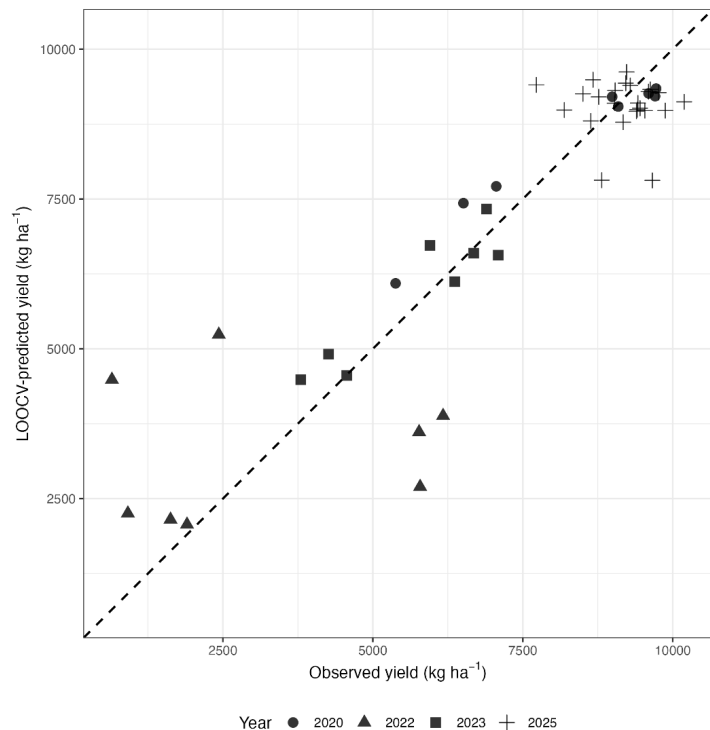
A5. Observed versus leave-one-out cross-validation (LOOCV)-predicted winter wheat grain starch concentration for the MLR model using RFE-selected predictors



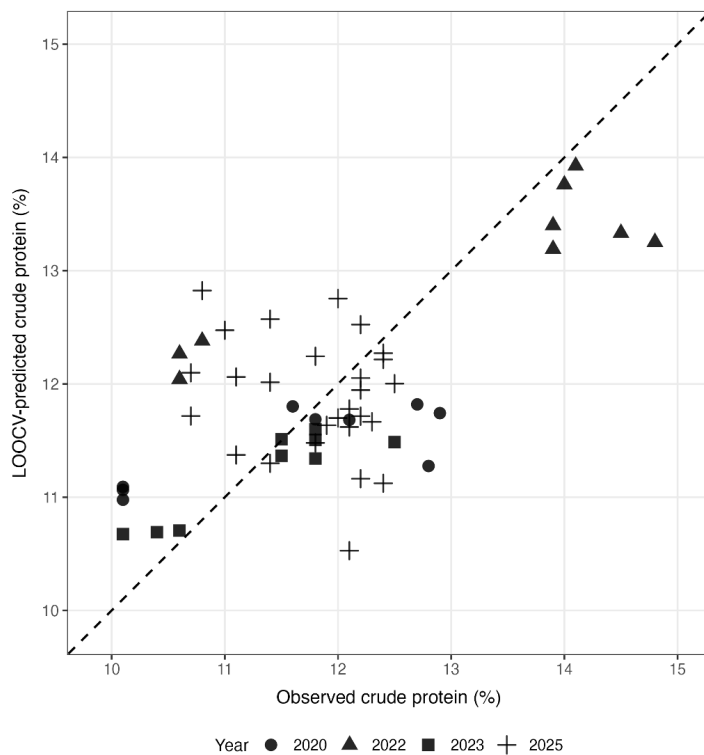
A6. Observed versus leave-one-out cross-validation (LOOCV)-predicted winter wheat grain starch concentration for the RF model using RFE-selected predictors



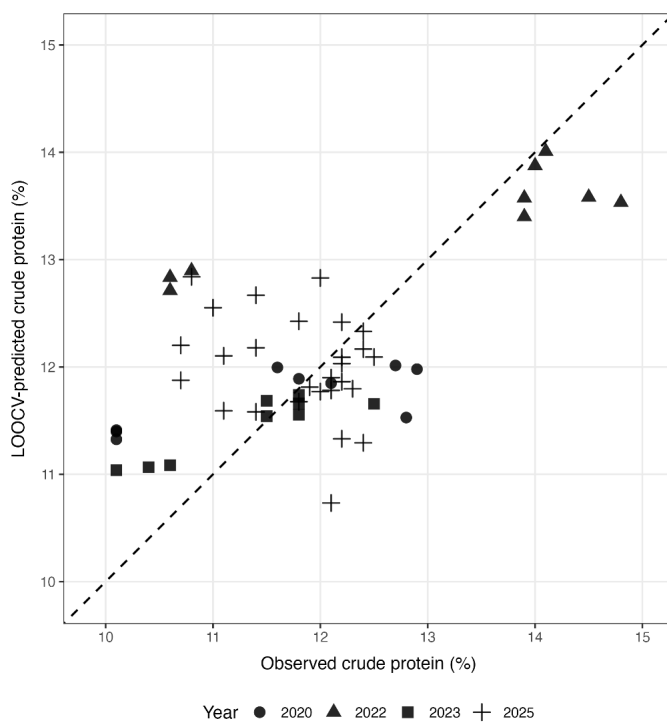
A7. Observed versus leave-one-out cross-validation (LOOCV)-predicted winter wheat grain yield for the MLR model using predictors selected through the initial screening procedure.



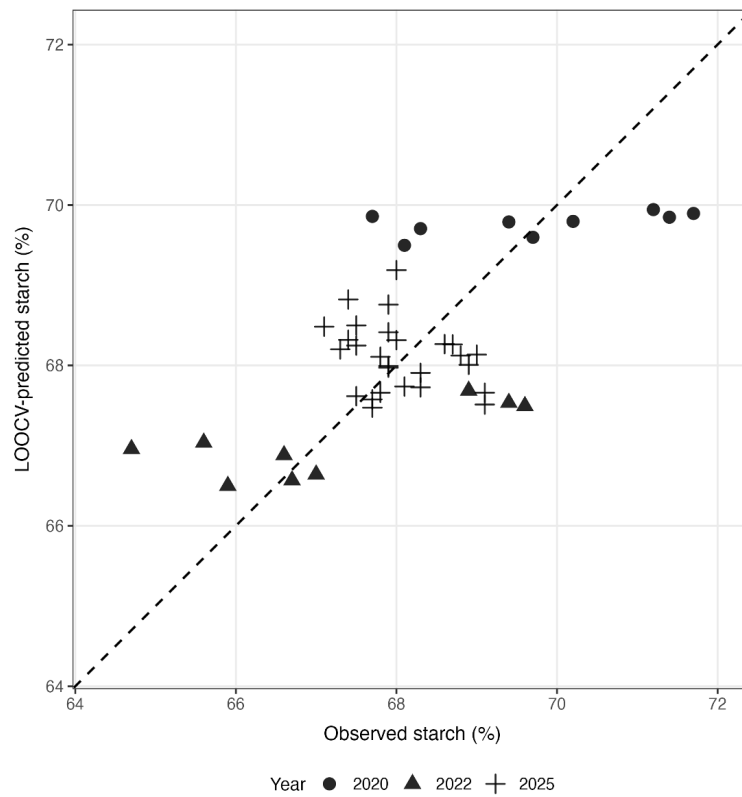
A8. Observed versus leave-one-out cross-validation (LOOCV)-predicted winter wheat grain yield for the RF model using predictors selected through the initial screening procedure.



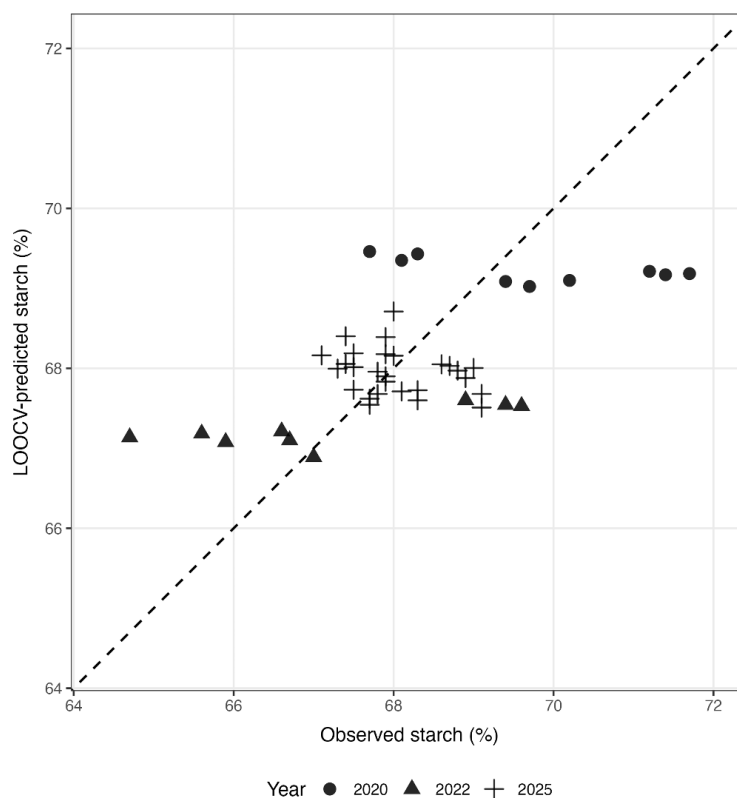
A9. Observed versus leave-one-out cross-validation (LOOCV)-predicted winter wheat grain crude protein concentration for the MLR model using predictors selected through the initial screening procedure.



A10. Observed versus leave-one-out cross-validation (LOOCV)-predicted winter wheat grain crude protein concentration for the SVM model using predictors selected through the initial screening procedure.



A11. Observed versus leave-one-out cross-validation (LOOCV)-predicted winter wheat grain starch concentration for the MLR model using predictors selected through the initial screening procedure.



A12. Observed versus leave-one-out cross-validation (LOOCV)-predicted winter wheat grain starch concentration for the SVM model using predictors selected through the initial screening procedure.

Appendix Table 1. Observed and externally predicted winter wheat grain yield at Lönnstorp for 2018-2023.

Year	Plot	Observed yield (kg ha ⁻¹)	Predicted yield (kg ha ⁻¹)	Difference (kg ha ⁻¹)	Absolute error (kg ha ⁻¹)	Difference (%)
2018	28	6,470	6,588	118	118	+1.8
2018	72	4,680	6,581	1,901	1,901	+40.6
2018	95	7,240	6,578	-662	662	-9.1
2019	31	8,826	6,857	-1,969	1,969	-22.3
2019	75	8,831	6,858	-1,973	1,973	-22.3
2019	98	8,729	6,832	-1,897	1,897	-21.7
2020	30	8,566	7,342	-1,224	1,224	-14.3
2020	74	7,757	7,298	-459	459	-5.9
2020	97	9,659	7,313	-2,346	2346	-24.3
2021	29	8,442	8,789	347	347	+4.1
2021	73	8,060	8,716	656	656	+8.1
2021	96	9,121	8,556	-565	565	-6.2
2022	28	8,629	8,397	-232	232	-2.7

2022	72	8,482	8,446	−36	36	−0.4
2022	95	8,642	8,228	−414	414	−4.8
2023	31	6,175	7,023	848	848	+13.7
2023	75	7,942	7,041	−901	901	−11.3
2023	98	7,806	7,037	−769	769	−9.9

Publishing and archiving

Approved students' theses at SLU can be published online. As a student you own the copyright to your work and in such cases, you need to approve the publication. In connection with your approval of publication, SLU will process your personal data (name) to make the work searchable on the internet. You can revoke your consent at any time by contacting the library.

Even if you choose not to publish the work or if you revoke your approval, the thesis will be archived digitally according to archive legislation.

You will find links to SLU's publication agreement and SLU's processing of personal data and your rights on this page:

☒ YES, I, Sarah Carola Vecchietti, have read and agree to the agreement for publication and the personal data processing that takes place in connection with this

☐ NO, I/we do not give my/our permission to publish the full text of this work. However, the work will be uploaded for archiving and the metadata and summary will be visible and searchable.