92. A new method for satellite-based derivation of site-specific yield potentials of winter wheat for precision farming applications

L. Hagn[1,*], M. Mittermayer[1], J. Schuster[1], F. Leßke[2] and K.-J. Hülsbergen[1]
[1]*Technical University of Munich, Liesel-Beckmann-Straße 2, 85354 Freising, Germany;*
** Ludwig.hagn@tum.de*
[2]*University of Applied Sciences Weihenstephan Triesdorf, Vöttingerstraße 27, 85354 Freising, Germany*

Abstract

Accurate site-specific yield potential determination is essential for sustainable agricultural practices, enabling efficient resource use and reducing environmental impacts. This study presents a satellite-based method to predict winter wheat yield using a six-year dataset from conventional and organic farming systems. Two models were evaluated: a non-linear support vector machine (SVM) model and a multilinear model with stepwise backward regression. The SVM model demonstrated higher predictive accuracy, with an R^2 of 0.80, a MAE of 0.58 t/ha, and an RMSE of 0.79 t/ha, compared to the multilinear model (R^2=0.75, MAE=0.69 t/ha). Expanding the dataset to include diverse soils, climates, and growth stages could further enhance model robustness and applicability.

Keywords: multilinear regression, remote sensing, satellite data, support vector machine, yield estimation

Introduction

Determining the site-specific yield potential of winter wheat (*tricitum aestivum*) is crucial for accurate management of agricultural production (Gebbers und Adamchuk, 2010). It is the basis of variable rate applications (VRA) such as targeted nitrogen (N) applications or site-specific seeding rates. In heterogeneous fields, VRA based on site-specific yield potentials enable the efficient use of resources while reducing inputs and maintaining high yields. This contributes to sustainable agriculture by avoiding over-fertilization and minimizing the impact of nitrogen emissions (Mittermayer *et al.*, 2021; Schuster *et al.*, 2022).

One significant challenge is that many legally prescribed fertilization systems rely on the target yield at the field level as an input factor to calculate the appropriate amount of fertilizer (GFO, 2020; Mittermayer *et al.*, 2024). If the target yield is overestimated, this can result in over-fertilization, which has negative environmental and economic consequences. Even more challenging is the application of site-specific fertilization systems in heterogeneous fields, as yield potentials are required on site-specific level, regardless of the chosen site-specific fertilization system.

A key technology for determining the site-specific yield potential has become the use of satellite data (Hagn *et al.*, 2024). This offers a consistent and scalable method to determine yield variability across fields, providing a precise foundation for agricultural management decisions. Many studies have been conducted on the topic of satellite-based yield prediction. A systematic literature review by van Klompenburg *et al.* (2020) revealed that temperature, rainfall and soil type are among the most commonly used variables for yield prediction. However, to enhance the accuracy of the models, it is essential to integrate additional agronomic factors, such as crop variety, critical growth stages (Wang *et al.*, 2024), and the impact of variable weather conditions on crop development (Zare *et al.*, 2024). A significant challenge in yield prediction is to enhance model robustness and adaptability across diverse farming conditions (e.g. conventional and organic farming systems).

Data quality further complicates yield modelling. Many studies use combine harvester yield data as a reference or training data set (Ji *et al.*, 2023). However, not all combine harvester datasets are entirely

reliable as they are often generated under time constraints and without the necessary but critical calibrations (Fulton *et al.*, 2018) reliable due to a lack of proper calibration. The most accurate data comes from plot combine harvesters equipped with integrated weighing scales. However, creating high-quality datasets using plot combine harvesters is a labour-intensive and costly process, limiting the number of studies that utilize this method (Spicker, 2017).
Currently, there is no reliable system for determining site-specific yields available in Germany. This study aimed to address this gap by developing a reliable method for satellite-based derivation of site-specific yield potentials of winter wheat. Using a six-year dataset from organic and conventional farming systems, this study evaluated two approaches for yield prediction: a support vector machine (SVM) model and a multilinear model with stepwise backward regression. This is a critical first step in identifying key parameters that influence grain yield. It forms the basis for the development of valuable applications in precision agriculture.

Material and methods

The study was conducted at three research stations of the Technical University of Munich – Thalhausen (A) and Roggenstein (B) and Viehhausen (C), located 30 km west and north of Munich. While Thalhausen and Roggenstein are managed under conventional farming practices, Viehhausen (C) is operated under organic farming conditions. The inclusion of data from fields managed under organic conditions aimed to include lower yield ranges as well, to ensure that the development of the algorithm accounts for a wider range of yield variability.

Data acquisition and data processing
Quantitative yield data was obtained from grain yield measurements using well-calibrated and validated New Holland combine harvesters (models CX 7.80, CX 7.90 and CX 8050). Qualitative yield data (core-threshing) was gathered from approximately 743 plots with a Wintersteiger Delta plot combine harvester, equipped weight scales, generating highly accurate, georeferenced core threshing data.
The core-threshing layout was aligned with the satellite grid, ensuring that each satellite grid cell (20×20 m) contains a georeferenced core threshing plot (15×1.8 m). To account for yield variations within fields, a relative biomass potential map (rel. BMP) according to Hagn *et al.* (2024) was integrated into the plot layout, ensuring that plots covered both high- and low-yield zones, effectively recording diverse yield variations (Figure 1).
Data were collected between 2018 and 2024, encompassing dry, wet, and typical years to account for interannual climate variability and improve yield prediction across diverse conditions.

The satellite imagery from Sentinel-2 satellites, covering the period from 2018 to 2024 was analyzed. Images were selected at key growth stages (GS) of winter wheat (GS 32, 55 and 65).
The red edge inflection point (REIP), which has proven to correlate well with plant nitrogen content and grain yield, has been incorporated into the model as an indicator of yield potential.

$$\mathrm{REIP} = 700 + 40\left(\frac{\frac{(670\mathrm{nm}+780\mathrm{nm})}{2} - 700\mathrm{nm}}{740\mathrm{nm} - 700\mathrm{nm}}\right) \qquad (1)$$

Furthermore, meteorological data (monthly average temperatures: AVG Temp. Jan.–Jul.), precipitation levels from January to July (AVG Prec. Jan.–Jul.), georeferenced soil information from the German soil survey (soil productivity index 1 and 2) alongside with relative heights of the fields were integrated with yield data for model developing. The German soil survey provides a general overview of the spatial distribution of soil types and their productivity (Blume, 2002), which served as a key data source for the specific soil conditions at each site.

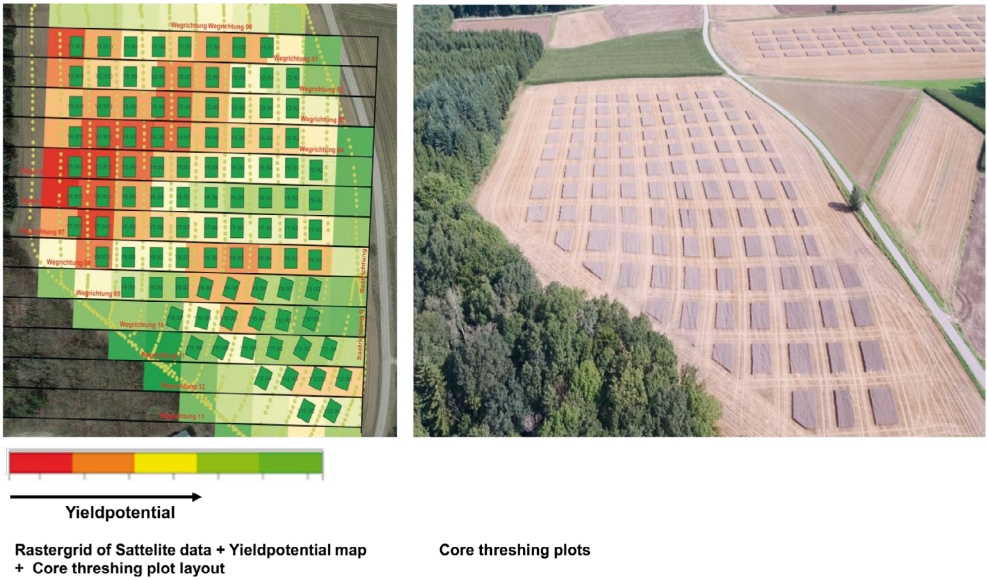

Figure 1. Plot layout of core threshing data. (a) Yield potential map, satellite raster grid and core threshing plots; (b) core threshing plots.

Yield prediction models

Two approaches were evaluated for modelling and predicting grain yield. The first approach involved a multilinear model with stepwise backward regression, where less important variables were removed step by step based on the Akaike information criterion (AIC) to identify the key parameters for yield prediction. The second approach used a non-linear support vector machine (SVM) model with a radial basis kernel, which incorporated all available data (predictors) to predict yield. Before training the models, the dataset was cleaned by removing outliers identified using Cook's Distance to prevent distortion in the results. The cleaned data was then randomly divided so that 80% was used for training and 20% for testing and validation. The performance of the models was evaluated using three key metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2.

Results

The SVM model demonstrates greater alignment with the observed data, resulting in a MAE of 0.58 t/ha, an RMSE of 0.79 t/ha, and an R^2 value of 0.80. In comparison, the multilinear model with stepwise backward regression demonstrated slightly lower predictive accuracy, with a MAE of 0.69 t/ha, a RMSE of 0.89 t/ha, and an R^2 value of 0.75 (Table 1). This is further illustrated by the scatter plots (Figure 2), where the SVM model model demonstrates a tighter clustering of points around the diagonal reference graph, indicating lower variability and stronger predictive alignment. In contrast, the multilinear model shows a wider spread of points, reflecting greater variability in predictions.

Table 1. Model evaluation: yield prediction accuracy.

Model	MAE (t/ha)	RMSE (t/ha)	R^2
SVM model	0.58	0.79	0.80
Backward-regression model	0.69	0.75	0.75

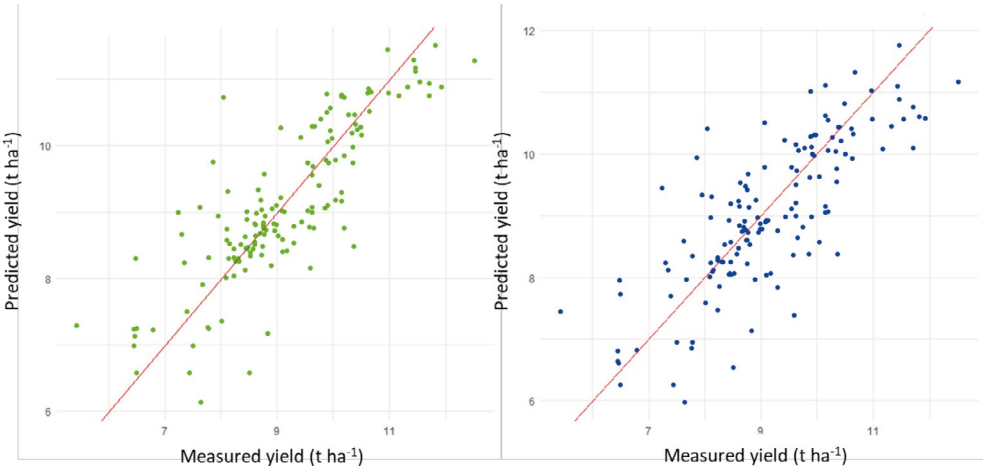

Figure 2. Predicted vs measured yield: SVM model (left), linear model (right).

The backward regression analysis initially started with 20 predictors, as shown in Table 2. The regression algorithm reduced the number of predictors to 11, simplifying the model while retaining the most influential explanatory variables. The final model included REIP values from different growth stages, soil productivity indices and average temperatures from January to May. This stepwise reduction process improved the interpretability of the model and highlighted the critical factors influencing yield variability.

Figure 3 illustrates the progressive increase in AIC during the backward regression steps as key parameters were systematically removed. This analysis provides insight into the predictors that have a substantial impact on model performance and those that can be excluded without significantly compromising model quality. The dataset demonstrates that the REIP values of growth stages GS32 and GS55 have the most significant influence, followed by REIP GS 65 and one of the soil productivity indices. These results underline the high suitability of satellite-derived data for winter wheat yield prediction.

Table 2. Input parameters of models.

Integrated predictors of the backward-regression model	Significance	AIC
Full model		4480.6
Rel. height	**	4485.6
Soil productivity index 1	***	4518.3
Soil productivity index 2	***	4529.8
REIP GS 65	***	4560.5
REIP GS 32	***	4666.2
AVG Temp. Apr.	***	4679.8
AVG Temp. Mar.	***	4684.1
AVG Temp. Feb.	***	4684.9
AVG Temp. Jan.	***	4689.9
AVG Temp. May	***	4691.9
REIP GS 55	***	4820.4

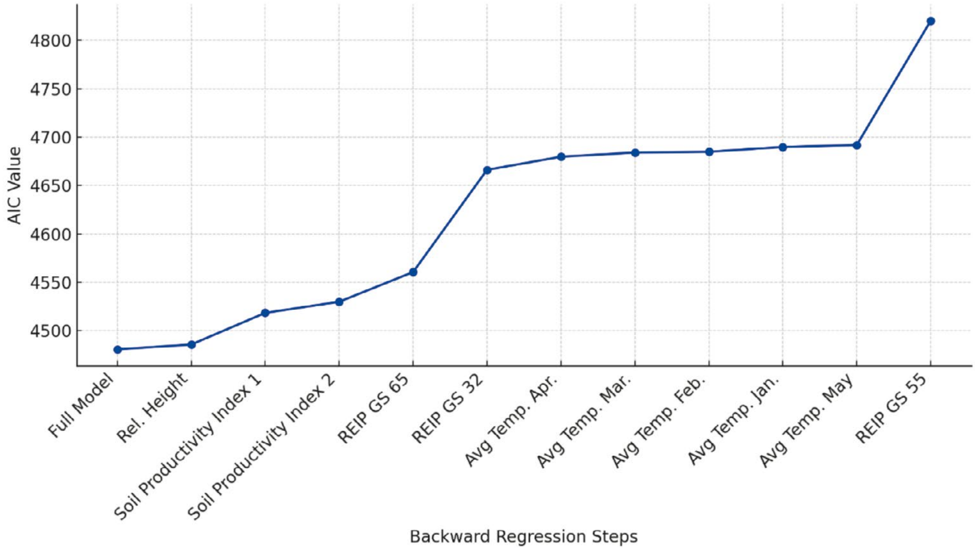

Figure 3. AIC development during backward regression steps.

Discussion

The results of this study demonstrate that both the SVM model and stepwise backward-regression-model can effectively predict grain yield, with the SVM model showing higher predictive accuracy. The stepwise backward-regression-model demonstrated the most influential predictors. These models have the potential to contribute to several practical applications in agriculture, such as site-specific nitrogen management and variable rate seeding based on predicted yield potential or the derivation of management zones. Nevertheless, further precision in yield prediction should be accomplished, as demonstrated by Cheng *et al.* (2022).

One limitation of the study is the selection of GS analyzed. While key stages such as GS 32, 55 and 65 were included, the critical period between GS 65 and GS 92, when crops absorb nitrogen crucial for final yield (Kichey *et al.*, 2007), was not covered. Additionally, the REIP, which is well-suited for green vegetation, is less effective for ripening crops due to declining chlorophyll levels. Incorporating complementary indices, such as the NDSI, (Rodrigues *et al.* (2018)), and extending the analysis to later growth stages could improve the model's ability to capture nutrient uptake and canopy maturity, thereby enhancing yield predictions.

Another limitation is the dataset's limited representation of diverse soil and climate conditions. While the study covered multiple years and climatic conditions, it predominantly included fields with soils with high water-holding capacity, which explains why precipitation was not a significant predictor, even in dry years. Expanding the dataset to include fields with sandy or less fertile soils and regions with more variable precipitation patterns could improve the model's applicability to water-limited environments and heterogeneous farming systems. To further enhance the model robustness and applicability, future work should focus on expanding the dataset to include more diverse soil-climate regions (quantitative data) and growth stages, while integrating additional spectral indices. These steps are crucial for the development of a robust and adaptable prediction system.

Conclusions

This study demonstrated the potential of satellite data combined with advanced modelling approaches to accurately predict winter wheat yield.

The SVM model outperformed the multilinear regression model ($R^2=0.80$, MAE = 0.58 t/ha), demonstrating its utility for precision agriculture. Key factors influencing yield variability included REIP values, soil productivity indices, and temperatures. While the method showed promising results, further improvements could be realized through the incorporation of later growth stages and the expansion of datasets to cover a more diverse range of soil and climate conditions. These enhancements will contribute to the support of sustainable and efficient farming practices.

Acknowledgment

This study was funded by the Federal Office for Agriculture and Food.

References

Blume, H.-P. (2002). Some aspects of the history of German soil science. Journal of Plant Nutrition and Soil Science, 165(4), 377. https://doi.org/10.1002/1522-2624(200208)165:4<377::AID-JPLN377>3.0.CO;2-K

Cheng, E., Zhang, B., Peng, D., Zhong, L., Le Yu, Liu, Y., …, & Yang, S. (2022). Wheat yield estimation using remote sensing data based on machine learning approaches. Frontiers in Plant Science 13, 1090970. https://doi.org/10.3389/fpls.2022.1090970

Fulton, J., Hawkins, E., Taylor, R., & Franzen, A. (2018). Yield monitoring and mapping. in: precision agriculture basics. Wiley, Hoboken, NJ, pp 63–77.

Gebbers, R., & Adamchuk, V.I. (2010). Precision agriculture and food security. Science, 327(5967), 828–831. https://doi.org/10.1126/science.1183899

GFO (2020). DüV. Verordnung, über die Anwendung von Düngemitteln, Bodenhilfsstoffen, Kultursubstraten und Pflanzenhilfsmitteln nach den Grundsätzen der guten fachlichen Praxis beim Düngen (DüV. Ordinance on the use of fertilizers, soil additives, growing media and plant additives in accordance with the principles of good professional practice in fertilization) (Düngeverordnung-DüV). BGBl. S.846, 2020 (German Fertilizer Application Ordinance). Federal Office for Agriculture and Food, Berlin.

Hagn, L., Schuster, J., Mittermayer, M., & Hülsbergen, K.-J. (2024). A new method for satellite-based remote sensing analysis of plant-specific biomass yield patterns for precision farming applications. Precision Agriculture, 25(6), 2801–2830.

Ji, H., He, X., Wang, W., & Zhang, H. (2023). Prediction of winter wheat harvest based on back propagation neural network algorithm and multiple remote sensing indices. Processes, 11(1), 293. https://doi.org/10.3390/pr11010293

Kichey, T., Hirel, B., Heumez, E., Dubois, F., & Le Gouis, J. (2007). In winter wheat (*Triticum aestivum* L.), post-anthesis nitrogen uptake and remobilisation to the grain correlates with agronomic traits and nitrogen physiological markers. Field Crops Research, 102(1), 22–32. https://doi.org/10.1016/j.fcr.2007.01.002

Mittermayer, M., Donauer, J., Kimmelmann, S., Maidl, F.-X., & Hülsbergen, K.-J. (2024). Effects of different nitrogen fertilization systems on crop yield and nitrogen use efficiency - Results of a field experiment in southern Germany. Heliyon, 10(7), e28065. https://doi.org/10.1016/j.heliyon.2024.e28065

Mittermayer, M., Gilg, A., Maidl, F.-X., Nätscher, L., & Hülsbergen, K.-J. (2021). Site-specific nitrogen balances based on spatially variable soil and plant properties. Precision Agriculture, 22(5), 1416–1436. https://doi.org/10.1007/s11119-021-09789-9

Rodrigues, F., Blasch, G., Defourny, P., Ortiz-Monasterio, J. I., Schulthess, U., Zarco-Tejada, P.J., Taylor, J.A., & Gérard, B. (2018). Multi-temporal and spectral analysis of high-resolution hyperspectral airborne imagery for precision agriculture: assessment of wheat grain yield and grain protein content. Remote Sensing, 10(6), 930. https://doi.org/10.3390/rs10060930

Schuster, J., Mittermayer, M., Maidl, F.-X., Nätscher, L., & Hülsbergen, K.-J. (2022). Spatial variability of soil properties, nitrogen balance and nitrate leaching using digital methods on heterogeneous arable fields in southern Germany. Precision Agriculture, 24, 647–676. https://doi.org/10.1007/s11119-022-09967-3

Spicker, A.B. (2017). Entwicklung von Verfahren der teilflächenspezifischen Stickstoffdüngung zu Wintergerste (*Hordeum vulgare* L.) und Winterraps (*Brassica napus* L.) auf Grundlage reflexionsoptischer Messungen. (Development of sensorbased nitrogen fertilization systems for oilseed rape (*Brassica napus* L.) and winter wheat (*Hordeum vulgare* L.)). Dissertation. Technische Universität München, Freising-Weihenstephan.

van Klompenburg, T., Kassahun, A., Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture 177, 105709. DOI: 10.1016/j.compag.2020.105709.

Wang, Z., Zhang, C., Gao, L., Fan, C., Xu, X., Zhang, F. et al. (2024). Time Phase Selection and Accuracy Analysis for Predicting Winter Wheat Yield Based on Time Series Vegetation Index. Remote Sensing 16 (11), 1995. DOI: 10.3390/rs16111995.

Zare, H., Viswanathan, M., Weber, T. K. D., Ingwersen, J., Nowak, W., Gayler, S., Streck, T. (2024). Improving winter wheat yield prediction by accounting for weather and model parameter uncertainty while assimilating LAI and updating weather data within a crop model. European Journal of Agronomy 156, 127149. DOI: 10.1016/j.eja.2024.127149.