Precision Agriculture (2025) 26:34
https://doi.org/10.1007/s11119-025-10224-6

In season estimation of economic optimum nitrogen rate with remote sensing multispectral indices and historical telematics field-operation data

Morteza Abdipourchenarestansofla[1,2] · **Hans-Peter Piepho**[1]

Accepted: 16 January 2025

Abstract

Accurate estimation and spatial allocation of economic optimum nitrogen (N) rates (EONR) can support sustainable crop production systems by reducing chemical compounds to be applied to the ground while preserving the optimum yield and profitability Smart Farming (SF) techniques such as historical precision agriculture (PA) machinery data, satellite multispectral imagery, and on-machine nitrogen adjustment sensors can bring together state-of-the-art precision in determining EONR. The novelty of this study is in introducing an efficient optimization framework using SF technology to enable real-time and prescription based EONR application execution. An optimization strategy called response surface modelling (RSM) was implemented to support decision making by fusing multiple sources of information while keeping the underlying computation simple and interpretable. Here, a field of winter wheat with an area of 7 ha was used to prove the proposed concept of determining EONR for each location in the field using auxiliary variables called multispectral indices (MSIs) derived from Sentinel 2. Three different image acquisition dates before the actual N application were considered to find the best time combination of MSIs along with the best MSIs to model yield. The best MSIs were filtered out through three phases of feature selection using analysis of variance (ANOVA), Lasso regression, and model reduction of RSM. For the date 2020.03.25, 14 out of 21 MSIs exhibited a significant interaction with the N applied as determined through an on-machine N sensor. For dates 2020.03.30 and 2020.04.04, the numbers of significant indices were identified as 6 and 10, respectively. Some of the MSIs were no longer significant after five days of the growth period (5-day interval between Sentinel 2 revisits). The best model demonstrated an average prediction error of 14.5%. Utilizing the model's coefficients, the EONR was computed to be between 43 kg/ha and 75 kg/ha for the target field. By incorporating MSIs into the fitted model for a given N range, it was demonstrated that the shape of the yield-N relation (RSM) varied due to field heterogeneity. The proposed analytical approach integrates farmer engagement by participatory annual post-mortem analysis. Using the determined RSM approach, retrospective assessment compares economically optimal N input, based on observed MSIs values to each location, with the actual applied rates.

Extended author information available on the last page of the article

Published online: 17 February 2025

Keywords Economic optimum nitrogen rate · Multispectral indices · Optimization · Response surface model · Winter wheat

Introduction

Nitrogen (N) is a vital nutrient and often the most limiting factor in winter wheat production. It supports promoting tillering, enhancing photosynthesis, and building protein in the grain. Hence, the allocation of this resource must be regulated over the growth stages by managing the timing, rate, form, and placement of N fertilization applications. Farmers tend to use more chemical substances in the hope of achieving maximum yield, which raises environmental concerns. A sustainable crop production system requires the regulated allocation of resources for activities like seeding, fertilization, and crop management. Monitoring crop health and crop N content is the basis of efficient fertilization. A study by Smerald et al. (2023) demonstrates that efficient redistribution of N fertilizer could significantly impact the global distribution of cereal production, leading to changes in trade patterns and food self-sufficiency levels. Smart farming technologies such as remote sensing (RS) imagery provide a comprehensive resource for achieving sustainability in crop production systems (Rane, 2023). RS technologies have gradually replaced traditional methods such as field surveys and laboratory testing for diagnosing crop health and N content. Soil and climate data availability is often limited, especially for small or remote farms. RS data enables a direct measurement of crop status and compensate to some extent for the lack of data for nitrogen estimation. There have been studies that prove the applicability of multispectral indices (MSIs) in measuring crop chlorophyll content to monitor crop physiological and phenological status (Zhou et al., 2019). Studies have also shown a great potential of vegetation indices derived from multispectral and hyperspectral imagery in near real-time crop health and status monitoring (Zheng et al., 2022). For instance, Fan et al. (2023) conducted a comprehensive analysis of hyperspectral single-band, two-band, and three-band spectral indices, identifying the optimal three-band index (TBI 5: 530, 734, 514) for accurate and stable estimation of potato plant nitrogen content (PNC) across different years, cultivars, and growth periods using UAV hyperspectral remote sensing. Another study by Li et al. (2023) utilized multispectral features in conjunction with machine learning (ML) algorithms to predict total nitrogen content. Their approach achieved a coefficient determination ranging from 0.37 to 0.70, indicating moderate predictive capability depending on the dataset and conditions. While these methods are promising, ML-based approaches, such as those employed by de Lara et al. (2023) in modeling EONR, can be computationally intensive and require extensive training periods, making them less practical for real-time applications. Our proposed model, in contrast, uses auxiliary variables like MSIs in classical regression models to enhance its flexibility and practicality. By integrating MSIs into a response surface modeling (RSM) framework, this study introduces a more interpretable and computationally efficient alternative for real-time estimation of EONR. Recent advancements in precision agriculture have highlighted the potential of integrating remote sensing, geographic information systems (GIS), and sensor technologies to optimize nitrogen use efficiency while minimizing environmental impacts (Weiss et al., 2020; Agrahari et al., 2021). However, most previous studies have focused on estimating plant N content via spectral indices rather than determining the EONR. Moreover, these studies generally rely

on calibration methods, with limited integration of on-machine sensors and RS data to determine EONR at a fine spatial scale. This paper presents a novel, comprehensive, end-to-end solution for determining EONR by integrating feature selection, crop functioning via RSM, and the determination of EONR into a unified framework. The proposed framework also introduces RSM as a practical optimization scheme, offering better scalability and flexibility compared to traditional quadratic or plateau functions used in many commercial systems. Notably, this approach moves away from small-plot designs typically used in precision agriculture research, instead favoring whole-field trials for in-field EONR estimation. This approach is not only more practical but also facilitates large-scale implementation, reducing the bias and random effects associated with smaller trial plots. The objective of this research is to explore data fusion that integrates on-machine crop sensing technology and satellite-based MSIs from three distinct dates prior to the actual N application. By investigating the influence of image acquisition dates on feature selection and spectral index choice, this study aims to provide a practical solution for setting up real-world on-farm experimentation (OFE) for site-specific farming. This paper is the first to comprehensively provide a framework for EONR estimation using MSIs with RSM, optimizing both computational efficiency and practical scalability for large-scale field operations.

Materials and methods

Experimental site

The experimental farm is located in southwest Germany. This region is characterized by diverse environmental factors that influence agriculture and land use practices. The region typically receives around 700–900 millimeters of precipitation per year, with variations across different parts of the area. The precipitation patterns in the area play a crucial role in determining agricultural practices, water management strategies, and vegetation growth. The soil types in the study area vary but are predominantly characterized by fertile loamy soils with good drainage properties. The climate is classified as a temperate oceanic climate with mild temperatures and relatively high humidity levels.

Experimental design

Phenological changes alter a crop's growth cycle and lead to variations in the assimilation of light, temperature, water, and nutrients required. Apart from nutrition, the other factors—light, temperature, and water—are confounding variables that can influence the resulting yield. Therefore, this study focuses on a controllable factor, namely nutrition application, specifically mineral N. In contrast to other N-treatment experiments that employ strip, plot, or block designs, this study utilizes an on-machine sensor called N-sensor (Fig. 1) to apply N in real-time as the crop-sensing technology assesses crop characteristics for the whole field. The N-sensor technology is provided by a third party (Yara N-sensor). The higher the sensor value, the less fertilizer will be applied. The operator defines minimum and maximum of input N for the whole field. This approach covers the entire field, making it a management zone design to choose the N-level for each location in the field. As a result of applying sensor technology as described, however, there is substantial variation in the actual amounts of

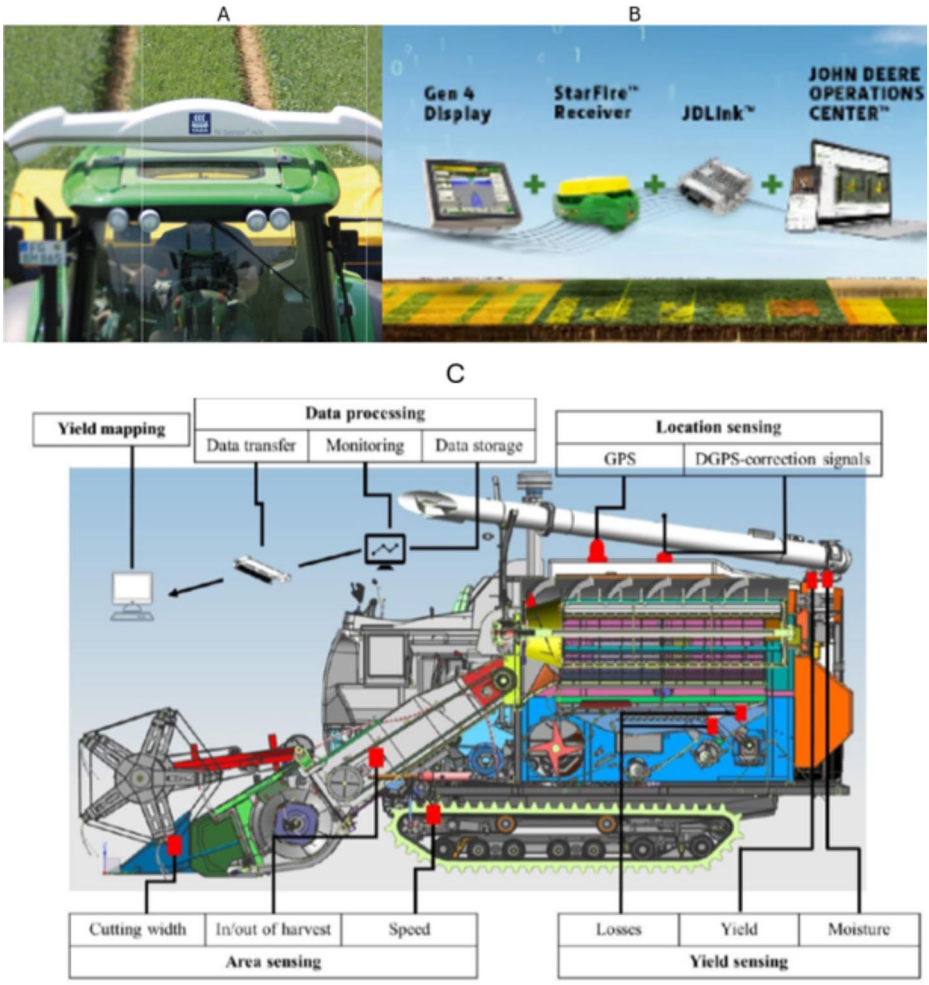

Fig. 1 Illustration of precision agriculture technologies used in nitrogen and yield management. (A) Displays the N-sensor technology (Weyand, 2022) used for real-time N adjustment based on crop reflectance. (B) Shows the PA equipment, including the Gen 4 Display, StarFire Receiver, JDLink, and John Deere Operations Center, which are utilized for data collection, monitoring, and documentation. (C) Presents a detailed schematic of harvester yield monitoring technology (Leroux, 2020), highlighting systems for yield mapping, location sensing, and data processing, integrated with GPS and DGPS correction signals to enhance field operation accuracy

N applied, providing sufficient heterogeneity to be able to fit the RSM. In other words, due to the fact that the YARA sensor probably deviated to some extent from the EONR in many cases, there is enough variation around the EONR to be able to estimate the RSM.

The RSM then takes the applied N and spectral indices to model yield as the response variable. Five fields of winter wheat from a specific farm were selected for the analysis, with each field treated as an individual experiment. Mineral N fertilization is commonly split into three or four applications during the season. To simplify the analysis, the most common N application (ammonium sulfate saltpeter) in winter wheat was considered in the

analyses. Note that the fertilizer product type does not limit the analysis. The experimental design implemented in this trial facilitates an exploration of the selection and applicability of specific spectral indices for N determination. This was achieved through an examination of interactions obtained by cross-products involving input N and selected spectral indices, which were measured before the actual N application. It is noteworthy that the acquisition time of satellite imagery can influence these interactions. This study focuses on a single field to establish the foundational data pipeline for preprocessing and integration, making it a proof of concept for determining in-season EONR. The target experimental field has an area of 7 ha with a comprehensive set of variables covering the entire field, including harvest yield observations, as-applied N sensor readings, and 21 MSIs measured before the N application operation. A second-order RSM was developed for the entire field, necessitating the integration of diverse and heterogeneous map layers, thereby requiring appropriate data preprocessing steps.

Data description

Three main datasets are used for this study: harvester yield data, sensor readings during spraying applications, and multispectral satellite imagery. The field operation data such as harvested yield and N application rates were documented through the John Deere Operation Center for analysis. Data from Sentinel satellite imagery was collected from three different dates before the application. The multispectral imager on the Sentinel-2 satellite offers a diverse collection of 13 spectral bands ranging from visible through near infrared to shortwave infrared. The dates 2020.03.25, 2020.03.30, and 2020.04.04 were selected to download Sentinel satellite data because they represent the most recent acquisition dates prior to the nitrogen application on the winter wheat crop. The goal was to capture up-to-date spectral information to assess the crop's characteristics just before the N application. By using satellite data closest to the application timing, it ensures accurate monitoring of the crop's condition, allowing for precise prescription mapping to optimize nitrogen use and improve yield outcomes. See Table 1 for details about trial data.

The amount of N applied was determined by the N-sensor for each location on the field and the applied rates were documented through sensor readings capability and internet connectivity in John Deere Operations Center (Fig. 1). By the end of the season, the crop was harvested and the amount of harvested yield for each location in the field was measured by combine harvester yield monitoring technology (Fig. 1). The harvest observations were documented in the Operation Center.

Twenty-one different indices were calculated from Sentinel-2 spectral bands for data fusion with field operation data from the target field. The acquisition of the imagery was done on at least three different revisiting dates before the actual N application. Since the assumption is that there exists a significant interaction between MSIs and N-rate, the MSIs must be acquired before the actual N application. This study investigates 21 MSIs derived from multispectral imagery to estimate the economical optimum of N in terms of net-return by proposing a second order RSM regression via factoring in historical telematics field-operations such as harvester observation and N sensor readings. Therefore, the goal is to develop an RSM with the best set of 21 MSIs combined from all three different acquisition dates (Table 1) before the actual N application took place. Several N treatments were mapped out in the trial based on a management zones design and the N-rates were deter-

Table 1 Data sources used for this study are field-operation data and their corresponding satellite imagery. The crop growth stages are also provided as BBCH for better understanding the relation of image acquisition date and crop fertility stage

Operation	Product	Range	Date	BBCH Stage
Nitrogen application	Ammonium sulfate saltpeter	20–75 kg/ha with 26% Ammonium sulfate saltpeter	2020.04.06	-
Harvest	Winter wheat	3.7–14 t/ha	2020.07.13	-
Seeding	Winter wheat	-	2019.11.15	-
Seninel-2 imagery	Sentinel-2 L1C BAND	-	2020.03.25 2020.03.30 2020.04.04	BBCH Stage 30–31: Beginning of stem elongation. BBCH Stage 31–32: Stem elongation continues. BBCH Stage 32–33: Further stem elongation, second node detectable

mined through crop sensing technology, which was mounted on the machine for real-time determination of N needed for the crop (see Fig. 2). To achieve the objective of this study one of the trial fields was utilized.

Data preparation

Given the differing spatial resolutions of the map layers in this study, an automated data preprocessing pipeline was developed to facilitate the final analysis of map layers. Additionally, this pipeline serves as an automated solution for scaling up in a data production environment (see Fig. 3).

Detecting faulty data, also known as outliers, in the documented harvester yield observation data and sensor readings from applicator machines is essential for accurate analysis. Telematics machinery observations from agriculture fields are contaminated with various sources of error that obscure the spatial structure of the observations. These error sources are well-known, mostly inherited from the dynamic nature of operation, speed changes, irregular field topography, non-fully used cutting bar during harvest operation, and start/end delays of headland and filling/emptying times. Cleaning the documented telematics machinery data was achieved through an unsupervised machine learning approach proposed by Abdipourchenarestansofla & Piepho (2022). The proposed method leverages the inter-

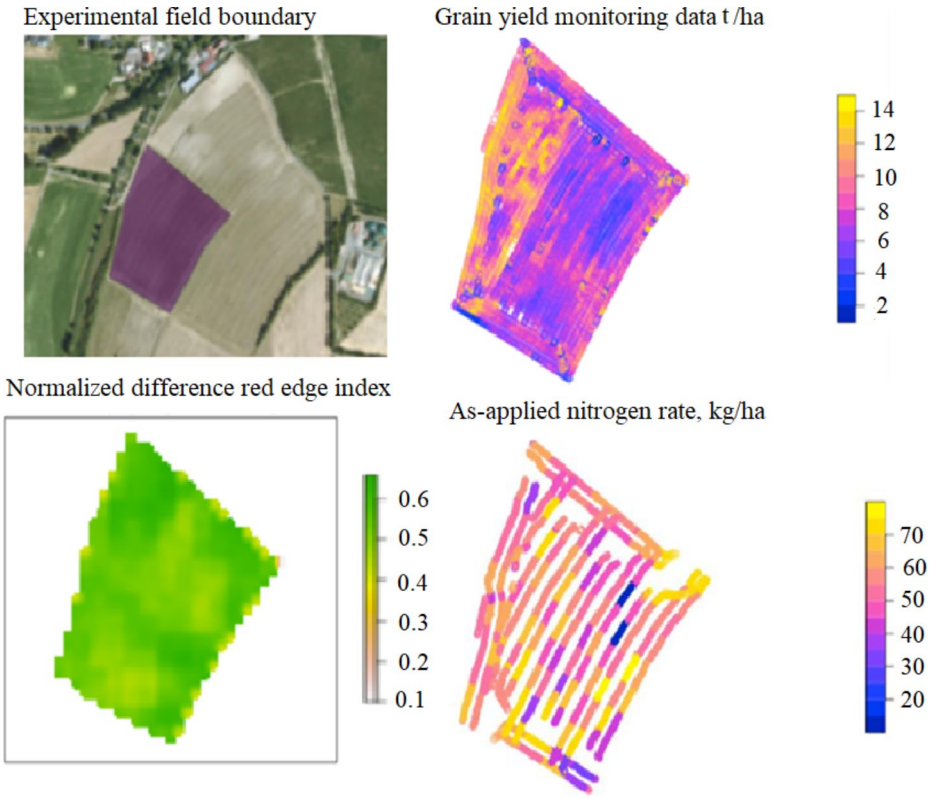

Fig. 2 Overview of the experimental field and associated data layers used in the study. The top-left image shows the experimental field boundary. The bottom-left image illustrates the Normalized Difference Red Edge Index (NDRE), a vegetation index derived from Sentinel-2 multispectral imagery, representing crop health variability. The top-right panel depicts grain yield monitoring data collected through precision agriculture equipment, while the bottom-right panel represents as-applied nitrogen rates (kg/ha) sensor reading data during field operation. These spatial data layers, presented in a combination of raster and discrete formats, are crucial for analyzing relationships between crop performance and nitrogen application. Hence, appropriate data processing and preparation is required (see Fig. 3) to enable map overlay analysis for this study

quartile range and Robust Kernel Outlier Factor as global and spatial outlier detection strategies, respectively. The automation of the developed data integration and map layer analysis pipeline for estimating N at production level requires a data quality component that ensures not only the accuracy of sensor measurements but also the accuracy of the documented telematics data such as GPS location accuracy, metadata accuracy, and user data entry as they enter the field for that corresponding operation (Abdipourchenarestansofla & Schroth, 2022). An example of the effect of outliers on the analysis is demonstrated in Fig. 4.

After removing outliers in the row observation from telematics machinery (spreader and harvest observation), the data were interpolated to 5 m spatial resolution. In small fields where heterogeneity can happen at a small scale, aggregation results in a higher degree of homogeneity across the field whereas for a larger field size (e.g., >30 ha) it is possible to preserve spatial variability when the sensor readings are aggregated. The applied rate

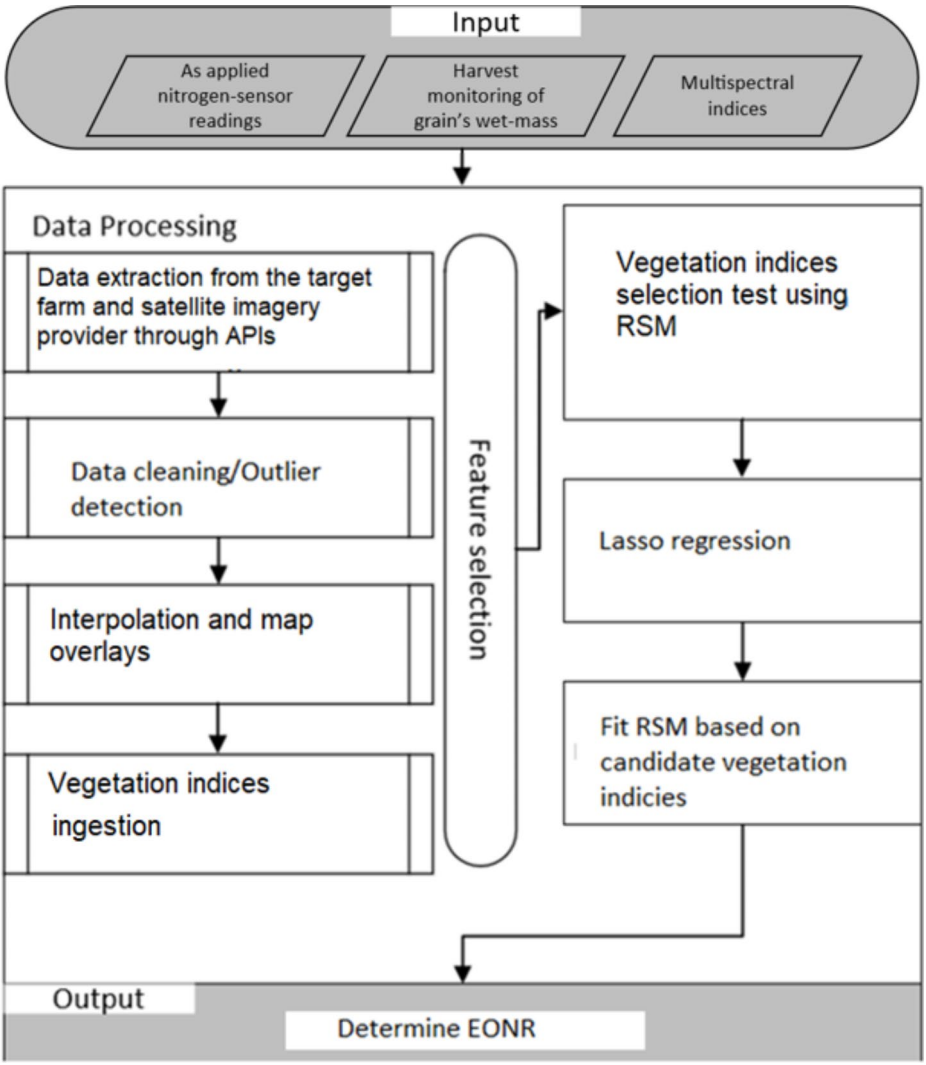

Fig. 3 Pipeline for data preprocessing and integration for final analysis. The figure represents the overall workflow of data preparation and integration automation pipeline. The historical field operation data were automatically extracted through Restful Application Programming Interfaces (APIs) provided by Operation Center, the John Deere cloud solution for Farm Management Information System (FMIS)

was recorded on a spreader machine where individual GPS-data points (applied rate for each GPS location) do not represent the exact region/area where the N was distributed. In other words, the N was distributed at that location plus the neighboring locations with some radius, and to recover that a variogram-based interpolation can be a reasonable choice for recovering the data. With this approach, the statistical dispersion or in general descriptive statistics of the data do not change significantly before and after interpolation. When fitting the RSM, all interactions between N and MSIs were assessed, as well as the expected quadratic relation (rising and falling of the response curve as N increases). Such observa-

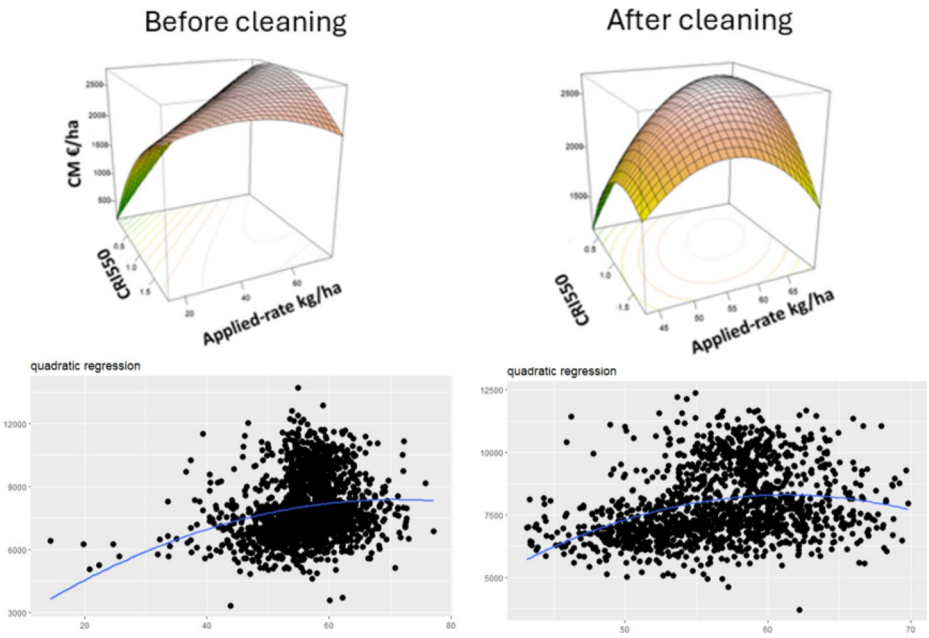

Fig. 4 The figure illustrates the importance of outlier removal. The top panels show the response surface model before and after removing outliers, where the response surface becomes more reliable and reveals a global yield maximum. The bottom panels depict the quadratic regression fitted for yield and applied nitrogen, highlighting the effect of outliers before and after cleaning

tions are consistent with the general principles of spatial data analysis and geostatistics. The approach of comparing distributions and visualizing map layers for different field sizes is a sound methodology for deciding which data handling method (aggregation or interpolation) best suits specific needs and objectives of the underlying study. It is essential to adapt such data analysis techniques to the spatial characteristics and scale of the problem that is being addressed. Such insights can help guide more informed decision-making in PA.

Response surface methodology

Crop characteristics at relevant phenological stages can be indicators of N deficiency. Hence, modeling yield response was performed to input using RSM that also includes the MSIs as proxies for N deficiency at relevant phenological stages. The shape of the response to N for a given value of an index changes spatially only when there is an interaction between N and that index. Additionally, for the RSM to work, apart from the significant interaction, the quadratic term for applied N needs to be negative (Box & Draper, 2007). This ensures that the conditional polynomial at a location in the field, given the values of the covariates which was plugged in, has a maximum. The choice of MSIs and the best timing can be determined by significance test for interaction in a second order RSM with interactions and their quadratic terms. RSM is a collection of statistical and mathematical techniques that are essential for developing, improving, and optimizing processes (Box & Draper, 2007). RSM finds its broadest utility in scenarios where multiple input variables have the potential to impact a

process's performance, or the quality attribute associated with it. In this context, the performance measure or quality characteristic is referred to as the response. These input variables are also referred to as independent variables (N and MSIs) where N is under the control of either the scientist or the farmer, or potentially both. The key hypothesis of this study is inspired by the idea that the MSIs derived from satellite data can support in site-specific modeling of the amount of N needed for the plant. Hence it assumes a significant interaction between as-applied N rate by crop sensor technology and vegetation indices derived from satellite data. If there is an interaction, the EONR at a specific site can be analytically modeled through an RSM. Analyzing two quantities experiments with second-order RSM enables exploitation of the maximum return (Piepho & Edmondson, 2018). Assuming the data are standardized to a 5-meter spatial resolution and are ready for analysis, each individual cell of the grid map represents an area of 25 m², constituting a single observation. Hence given a harvester yield observation layer and its pairs of explanatory variables (i.e., as applied N rate and one of the MSIs) the RSM is given by

$$y_i = \alpha + \beta_N n_i + \beta_{NN} n_i^2 + \beta_X x_i + \beta_{XX} x_i^2 + \beta_{NX} n_i x_i + \epsilon_i \quad (1)$$

where y_i is the harvested yield (kg/ha) at the i-th location, n_i is the rate of applied N recommended by on-machine crop sensing technology for the same location, x_i denotes the value of computed multispectral index for a given time at the very same location before the actual N, α is the intercept, $\beta_{NX} n_i x_i$ is the interaction term, $\beta_{NN} n_i^2$ and $\beta_{XX} x_i^2$ are the quadratic terms for applied N and spectral index, and ϵ is a random term variable having zero expectation and constant variance. Using parameters α, β_N, β_{NN}, β_X, β_{XX}, and β_{NX}, the method of least squares (Box and Draper, 2007) is used to estimate the parameters. The key feature of the RSM is that for fixed x_i there exists a quadratic model in n_i. For this, one can analytically determine the maximum response and EONR. The optimal N-rate will differ between different levels of x_i only if there is a significant interaction between applied rate and given vegetation index. In fact, presence of an interaction is necessary for the central hypothesis of Precision Farming to hold (Piepho et al., 2011). Additionally, for the proposed RSM model to work the β_{NN} is expected to be < 0. The response surface is computed with the 'rsm' package in the R language. This package provides several functions to facilitate classical RSM (Lenth, 2009).

Vegetation indices selection.

Feature selection involves the process of choosing a subset of pertinent features from a larger dataset. One method to gauge the significance of each feature is by fitting a model multiple times, each time utilizing a distinct set of variables or features, and subsequently assessing the model's performance. Considering the existence of three image acquisition dates and 21 MSIs (see Table 2) for each given acquisition date, complete scrutinizing of all the index-date combinations available as x_i was considered. The calculation of MSIs is given by Sentinel Hub Custom Scripts (2024). The screening of features for MSIs was meticulously conducted through three feature selection phases to identify the optimal set for N-modeling. The acquisition dates of satellite imagery were determined to span three consecutive days of Sentinel-2 revisiting time (every five days) from March 25th, 2020, to April 4th, 2020. Each of the three dates for MSIs undergoes three distinct phases independently of feature selection. The process begins with Phase One leveraging analysis of variance (ANOVA) tests using RSM output, followed by Phase Two using Lasso regression,

Table 2 21 MSIs derived from sentinel-2 multispectral imagery for this study that are used for ingestion with yield map and as-applied N map, see Fig. 3

MSIs	Description	MSIs	Description
ARI2	Anthocyanin reflectance index. Plant stress responses, understanding plant pigmentation.	EVI	Enhanced Vegetation Index. Assesses the health and vigor of vegetation.
ARVI	Atmospherically Resistant Vegetation Index. Measure of vegetation health.	EVI2	Enhanced Vegetation Index. Assesses the health and vigor of vegetation.
ATSAVI	Adjusted transformed soil– adjusted VI Reliable measure of vegetation vigor and health.	GLI	Green Leaf Index. Monitoring plant growth, photosynthetic activity, and stress responses in vegetation.
CARI	Chlorophyll Absorption Ratio Index. Monitoring plant health, photosynthetic activity, and stress responses.	GOSAVI	Green Optimized Soil Adjusted Vegetation Index. Assesses vegetation health while accounting for soil background influences.
CARI2	Chlorophyll Absorption Ratio Index 2. Enhancement of the original Chlorophyll Absorption Ratio Index.	LCI	Leaf Chlorophyll Index. Estimates the chlorophyll content of plant leaves.
CCCI	Canopy Chlorophyll Content Index. Assesses chlorophyll content in the canopy of vegetation.	MCARI	Modified Chlorophyll Absorption in Reflectance Index. Minimizes the influence of soil background and atmospheric interferences on chlorophyll estimates.
CHLGREEN	Green-Leaf Chlorophyll Index. Provides a quantitative measure of chlorophyll concentration in plant foliage.	MCARI1	Modified Chlorophyll Absorption in Reflectance Index 1.
CRI550	Carotenoid Reflectance Index. Provides information on the relative abundance of carotenoids in plant tissues.	MCARI2	Modified Chlorophyll Absorption in Reflectance Index 2.
CRI700	Carotenoid Reflectance Index.	NDRE	Normalized Difference Read Edge Index. Quantifying chlorophyll concentration, detecting early signs of stress, and evaluating plant vigor.
CVI	Chlorophyll vegetation index	NDVI	Normalized Difference Vegetation Index. Monitors changes in vegetation over time, assess plant productivity.
TCI	Triangular chlorophyll index. Provides information on chlorophyll concentration, photosynthetic activity, and plant stress responses.		

and concludes with Phase Three through a model reduction approach using RSM. All three phases are described in the following sections. In the end, the candidate MSIs (Montero et., 2023) from each date were combined to create the final second-order RSM.

Phase one Among all 21 MSIs for each day, 63 were the total MSIs to be filtered for final downstream modeling. This number of features makes modeling complex and can lead to problems posed by collinearity. Therefore, the feature selection process started with selecting the MSIs that have significant interaction with as-applied N. Given a dataset with MSIs,

where represents a single spectral index (i.e., Chlorophyll Absorption Ratio Index) and is the number of observations for a given date before actual N. For each variable we fit the proposed RSM (see Eq. 1) and perform analysis of variance (ANOVA) tests on the fitted model. The significance of the interaction of each MSI at the time of fitting with N-applied is checked. If the interaction is significant, we keep that MSI, if not that MSI is eliminated from further processing. A significance level of 0.01 for coefficient is considered. The mentioned process in Phase One was automated for all the individual MSIs separately across all three different dates. Fitting the above mentioned second-order response surface model for individual MSIs enables us to perform an ANOVA for each set of MSIs. After running the ANOVA test for all the individual MSIs acquired from all three dates, all the MSIs that have significant interaction with N applied through on-machine N sensor from all the dates were collected. All candidate MSIs were taken to the next phase for all three dates as the next step of feature selection.

Phase two To enhance model simplicity and avoid collinearity in fitting, a Lasso regression was employed as the second step in feature selection. The approach involves systematically applying Lasso to select candidate variables of interest (MSIs) from the initial phase. Initially, the significant MSIs were categorized into three classes (see Table 3) based on multispectral imagery, covering diverse crop characteristics (Zhao et al., 2022; Prasad et al., 2000). The categorization of MSIs is based on the parameters that were derived as indices from multispectral imagery. According to the literature each index represents a particular characteristic of the crop. Therefore, the MSIs are categorized into three groups labeled Biochemical, Chlorophyll, and Vegetation to distinguish the crop characteristics associated with the MSIs. For instance, the Anthocyanin and Carotenoid are identified as Biochemical properties of the plant (Zhao et al., 2022). Subsequently, the Lasso was applied to each class at specific dates, utilizing L1-norm regularization to identify the most relevant features within each category.

During feature selection, to uphold the concept of second-order RSM, quadratic and interaction terms were added along with linear terms for all combinations, and Lasso regression was applied for each category at each date. In other words, when a single index and N were used as predictors for the response, the quadratic and interaction terms for the given index,

Table 3 Three different categories were identified during the first phase of feature selection. After first phase of feature selection the MSIs are categorized in the three categories for the second phase of feature selection by Lasso regression

MSIs Category	Spectral Indices
Biochemical	ARI2 - Anthocyanin Reflectance Index CRI550 - Carotenoid Reflectance Index CRI700 - Carotenoid Reflectance Index
Chlorophyll	CARI - Chlorophyll Absorption Ratio Index CARI2 - Chlorophyll Absorption Ratio Index CCCI - Canopy Chlorophyll Content Index CHLGREEN - Chlorophyll Green CVI - Chlorophyll Vegetation Index LCI - Leaf Chlorophyll Index
Vegetation	EVI - Enhanced Vegetation Index NDVI - Normalized Difference Vegetation Index NDRE - Normalized Difference Red Edge Index GOSAVI - Green Optimized Soil Adjusted Vegetation Index ARVI - Atmospherically Resistant Vegetation Index

as well as a quadratic term for N, were generated and fed into the Lasso model. The results are presented in the "Results" section.

Phase three The second-order RSM was fitted to the combination of all the MSI categories and dates selected in the second phase and was considered as the full model to start the third iteration of feature selection. The original Eq. (1) was expanded to all the combination of MSIs that were filtered by previous phases. The criterion for finding the best combination is the significance of coefficients in the ANOVA table. Additionally, it is required that $\beta_{NN} < 0$, as a quadratic relation between the response and input nitrogen is expected.

In RSM, the concept of functional marginality, as advocated by (Nelder, 2000), underscores the importance of constructing "well-formed" polynomials. This entails including all marginal terms associated with each term in the polynomial to ensure the response surface maintains its properties under arbitrary linear transformations of predictor variables. For instance, given x and n as one the selected index and applied N, respectively, for that index if a quadratic term like x^2 is included, its corresponding linear term x must also be incorporated. Similarly, when incorporating interaction terms, such as xn, both x and n should be included. This approach ensures the stability and meaningfulness of the response surface model, aligning with the broader concept of well-posed mathematical problems.

Site-specific determination of economic optimum N-rate

At each location in the field, the EONR was computed via a set of candidate MSIs derived from sentinel-2 satellite imagery. To make an optimal decision for N-rate it is important to ask, "what are the extra (marginal) costs and what are the extra (marginal) benefits associated with the decision." The EONR is the point where an incremental change in N input costs equals an incremental change in the value of product produced. Let's assume the harvested winter wheat grain for a given location is 5000 kg/ha and the return on each kg is 0.4 €/kg denoted by P_W, then the value per ha is called total revenue (TR) given by

$$TR = P_W \times kg\ per\ ha = 2000\ \text{Euro}/ha$$

The cost of input N was calculated by multiplying 0.147€ /kg cost of product (P_N) purchased times the amount of products used per ha (60 kg/ha) and that is called total cost (TC) given by

$$TC = P_N \times kg\ N\ applied\ per\ ha = 25.08\ \text{Euro}/ha$$

Consequently, the return is given by

$$Return = P_W \times Y(n) - P_N \times n \tag{2}$$

The winter wheat yield per ha, denoted as Y, is determined by N inputs, and expressed as a function $Y(n)$. It is important to observe that while an optimal scenario suggests that increasing nitrogen can enhance yield, there exists a diminishing returns phenomenon. In the best-case scenario, higher nitrogen application may result in increased yield, but as the

nitrogen input continues to rise, the incremental gain in yield diminishes. This phenomenon is commonly referred to as the "Law of Diminishing Returns" (Fausti et al., 2017). The Return function is also called Contribution Margin (CM). CM is sales revenue minus variable costs (Athearn et al., 2021). In order to determine EONR a second order RSM is required. Taking the model with the estimated parameters, the N that maximizes the contribution margin is the EONR. The key idea is to plug in the local values of the index x (or indices) into the regression model, determine the coefficients of the quadratic polynomial for the local response to N, given these local values of *x*, and then determine the optimal N rate at the location. Assuming fixing a single covariate x at the value observed for the given location in the field, collecting polynomial terms for regression on n, $Y(n)$ is given by

$$Y(n) = c_0 + c_1 n + c_2 n^2 \tag{3}$$

where $c_0 = \alpha + \beta_X x$ is the intercept, $c_1 = \beta_N + \beta_{NX} x$ and $c_2 = \beta_{NN}$ are linear and quadratic term, respectively, which were acquired through RSM model fitted by fusing MSIs, on-machine N adjustment sensor, and harvested yield observation as the response. The first derivative of the yield with respect to the inputs (n) in the second-order polynomial equation corresponds to the slope of (n). In this context, the first derivative $\frac{\partial Y}{\partial n}$ signifies the incremental change in yield due to a marginal change in the nitrogen rate. The first derivative of Eq. (3) is defined as

$$\frac{\partial Y}{\partial n} = Y'(n) = c_1 + 2c_2 n \tag{4}$$

The derivate $\frac{\partial Y}{\partial n}$ is also called the marginal product of the input N. The economic significance of this equation lies in its ability to indicate the rate at which N per ha was transformed into yield per ha. Next by substituting $Y(n)$ in Eq. (2) with Eq. (3), and taking the derivative of the return with respect to n, we find

$$\frac{\partial\, Return}{\partial n} = P_W (c_1 + 2c_2 n) - P_N \tag{5}$$

where P_W represents the marginal revenue, which is the selling price of winter wheat, while P_N stands for the cost per unit of nitrogen fertilizer. $P_W (c_1 + 2c_2 n)$ represents the marginal product of nitrogen. By setting the derivative in (6) equal to zero, the EONR for nitrogen using the following equation can be defined by

$$(c_1 + 2c_2 n) = \frac{P_N}{P_W} \tag{6}$$

From this, the solution for EONR can be obtained by solving for *n* in Eq. (6), yielding

$$EONR = \left(\frac{P_N}{P_W}\right)\left(\frac{1}{2c_2}\right) - \left(\frac{c_1}{2c_2}\right) \tag{7}$$

The illustration here was for a single covariate *x*, but the method is readily extended to cover any number of additional covariates.

Results

Feature selection

Figures 5 and 6, and 7 demonstrate only the significant MSIs and for each date if we count each bar in the plot, we come to the number of significant MSIs. Notably, for the date 2020.03.25, 14 out of 21 MSIs exhibited a significant interaction with the N applied as determined through an on-machine N sensor. For dates 2020.03.30 and 2020.04.04, the numbers of significant indices were identified as 6 and 10, respectively. Some of the MSIs were no longer significant after five days of the growth period (5-day interval between Sentinel 2 revisits). For instance, NDVI is known as a limited index for higher canopy cover due to saturation issues (Wang et al., 2016), as observed in the analysis. NDVI was significant on the first date of 2020.03.25 and therefore is presented in the plot of coefficients (Fig. 5). However, as time passed with every five days of Sentinel revisit, NDVI was no longer significant (not in the plot of coefficients in Fig. 6 for dates 2020.03.30 and Fig. 7 for 2020.04.04), due to rapid growth of the crop around tillering and stem elongation stages resulting in a higher canopy cover, which, in turn, affected the NDVI readings.

Lasso regression

The generation of all terms (linear, quadratic, and interaction terms) and fitting Lasso regression for all the MSIs in each class at each date was done through an automated script. For simplicity the result of Lasso approach only for one date and a particular category is provided here. The procedure is the same for the other categories and dates.

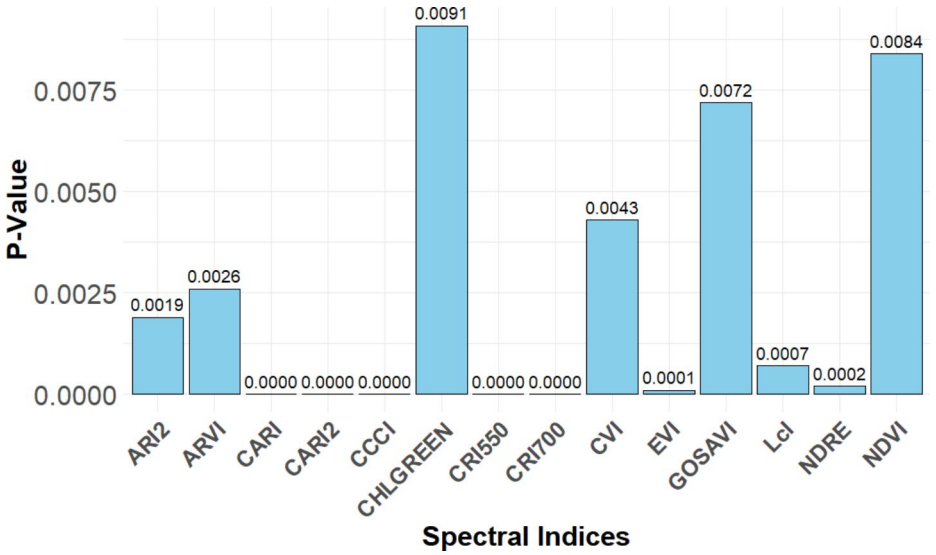

Fig. 5 MSIs with significant p-values (on top of each bar) extracted from ANOVA table for 2020.03.25, computed via a single RSM fit. For the mentioned date out of 21 indices only 14 indices are significant that are shown in this plot

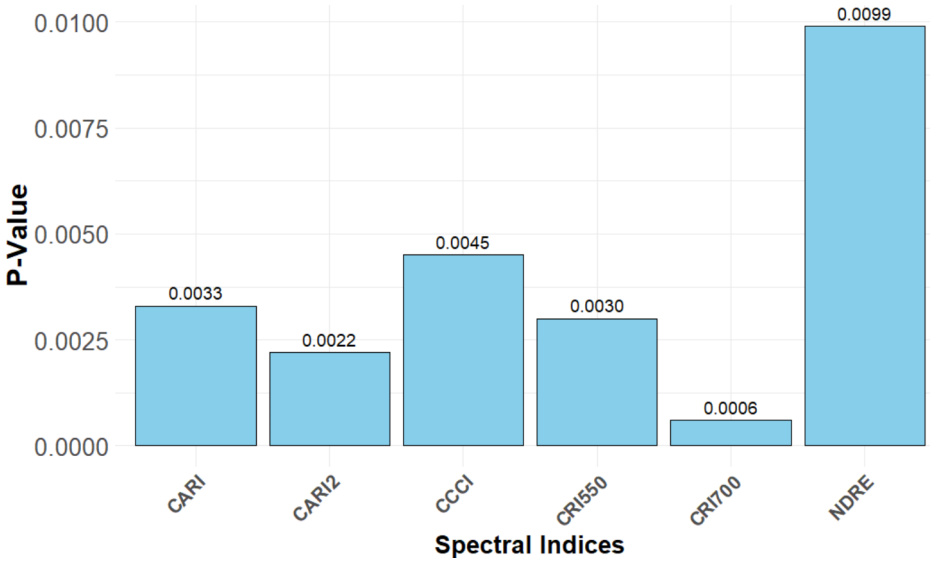

Fig. 6 MSIs with significant p-values (on top of each bar) extracted from ANOVA table for 2020.03.30, computed via a single RSM fit. For the mentioned date out of 21 indices only 6 indices are significant that are shown in this plot

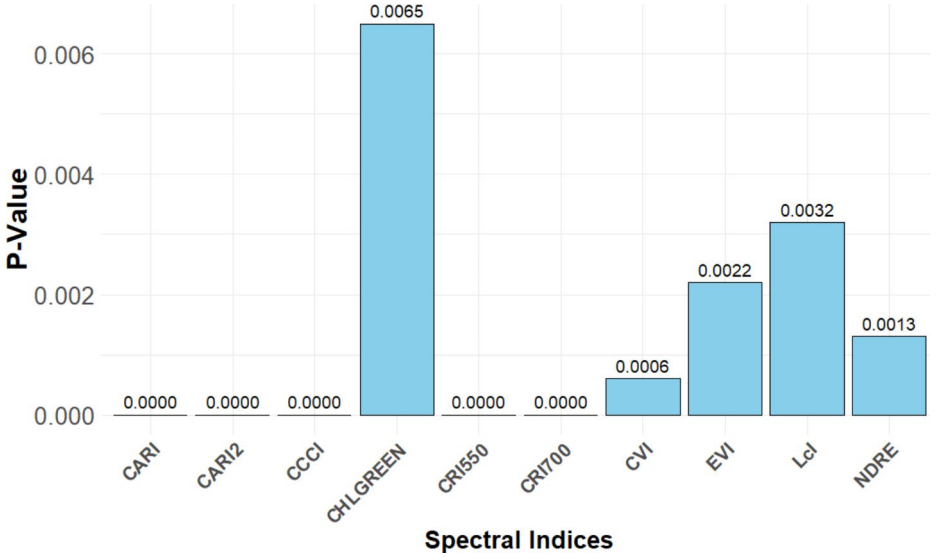

Fig. 7 MSIs with significant p-values (on top of each bar) extracted from ANOVA table for 2020.04.04, computed via a single RSM fit. For the mentioned date out of 21 indices only 10 indices are significant that are shown in this plot

Table 4 Example of linear, quadratic, and linear terms of MSIs from Chlorophyll category for date "2020.03.30" introduced in Lasso regression

Linear term	Quadratic	Interaction
$CARI$	CARI^2	$CARI : CCCI$, $CARI : CARI2$
$CCCI$	$CCCI^2$	$CCCI : CARI$, $CCCI : CARI2$
$CARI2$	CARI2^2	$CARI2 : CARI$, $CARI2 : CCCI$
$Nitrogen$	$Nitrogen^2$	$CCCI : Nitrogen$, $CARI : Nitrogen$, $CARI2 : Nitrogen$

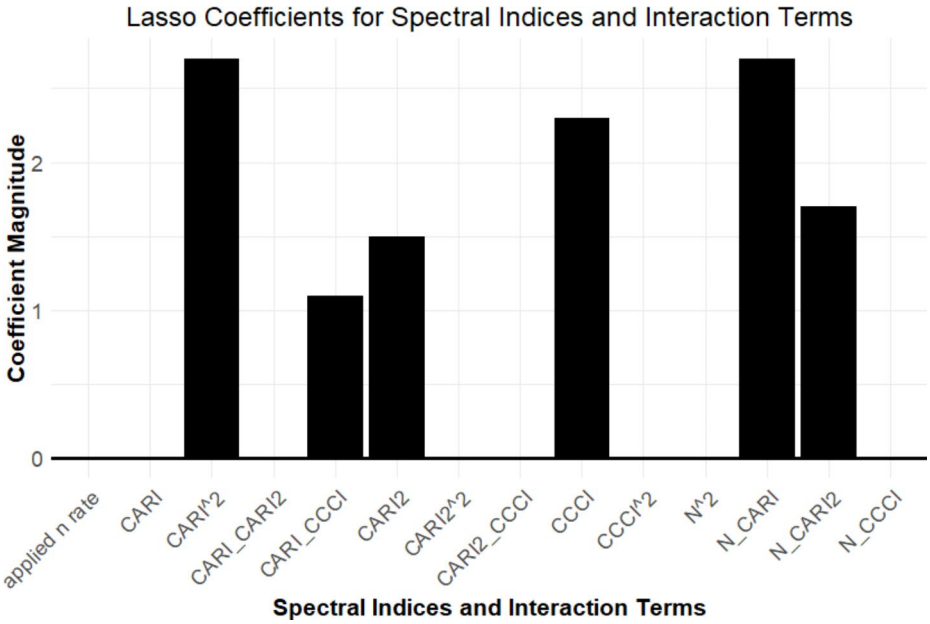

Fig. 8 Lasso coefficients for MSIs from chlorophyll category in 2020.03.30. As can be seen only CARI and CARI2 were selected as the candidate MSIs for the next step. Lasso coefficients for MSIs from the chlorophyll category on 2020-03-30 show that only CARI and CARI2 were selected as candidate MSIs for the next step. Among all interaction terms (N: CARI, N:CARI2, and N: CCCI), only those with non-negative coefficients were considered. Therefore, CARI and CARI2 were selected because their interaction term in the Lasso regression were not equal to zero, unlike the CCCI index

Hence in the above example in Table 4 as input dataset to Lasso, the goal was to select a set of MSIs whose coefficients only for the interaction terms are bigger than zero and take them to the modeling phase. In the example, the coefficients of $CARI : Nitrogen$ and $CARI2 : Nitrogen$ are not equal to zero and the Lasso coefficient for $CCCI : Nitrogen$ is zero (see Fig. 8). This indicates that CCCI is no longer a candidate for the third phase of feature selection.

This process was applied to each individual category at a given date for all the datasets. The Lasso regression shrinks the small coefficients to zero and hence retains the most relevant features in the data set for final modeling. The MSIs candidates are presented in Table 5.

RSM model output

Given the selected candidate MSIs from Lasso regression in Table 4 several RSM models were fitted with different combinations of MSIs candidates. Following feature selection in the third phase through model reduction of RSM, candidate MSIs for fitting harvest yield were determined, including $CARI2_{2020.03.25}$, $ARI2_{2020.03.25}$, $CARI_{2020.03.30}$, $\mathrm{CRI550}_{2020.03.30}$, $CARI2_{2020.04.04}$ and $CRI700_{2020.04.04}$. The ANOVA table indicated the significance of all terms in the model summary. The β_{NN} coefficient for n^2 was found to be negative, aligning with the criteria of a second-order model, where an increase in N led to a diminishing return in yield. The best model demonstrated an average accuracy error of 14.5% using root mean square error (RMSE). Utilizing the model's coefficients, the EONR range was computed to be between 43 kg/ha and 71 kg/ha. Eventually it turned out that the contribution of NDER and LCI in the model along with the rest of selected indices leads to bias in determination of EONR with large negative and large positive EONR value. However, with model reduction by eliminating the LCI and NDRE, the final model determined EONR in a reasonable range as recommended by the literature and also by the farmer's input. The final model for determination of EONR is represented as follows:

$$
\begin{aligned}
y = {} & \alpha + \beta_N N + \beta_{NN} N^2 + \beta_{CARI2_{2020.03.25}} CARI2_{2020.03.25} \\
& + \beta_{CARI2_{2020.03.25.2}} CARI2_{2020.03.25}{}^2 + \beta_{ARI2_{2020.03.25}} AR12_{2020.03.25} \\
& + \beta_{ARI2_{2020.03.25.2}} ARI2_{2020.03.25}{}^2 + \beta_{CARI\,2020.03.30} CARI_{2020.03.30} \\
& + \beta_{CARI\,2020.03.30.2} CARI_{2020.03.30}{}^2 + \beta_{\mathrm{CRI550}2020.03.30} \mathrm{CRI550}_{2020.03.30} \\
& + \beta_{\mathrm{CRI550}2020.03.20.2} \mathrm{CRI1550}_{2020.03.20}{}^2 + \beta_{CARI2\,2020.04.04} CARI2_{2020.04.04} \\
& + \beta_{\mathrm{CARI2}2020.04.04.2} CARI2_{2020.04.04}{}^2 + \beta_{\mathrm{CRI700}2020.04.04} CRI700_{2020.04.04} \\
& + \beta_{CRI700\,2020.04.04.2} CRI700_{2020.04.04}{}^2 + \beta_{\mathrm{N.CARI2}2020.03.25} \mathrm{N{*}CARI2}_{2020.03.25} \\
& + \beta_{N.ARI2\,2020.03.25} N * ARI2_{2020.03.25} + \beta_{N.CARI\,2020.03.30} N * \mathrm{CARI}_{2020.03.30} \\
& + \beta_{N.CRI550\,2020.03.30} N * \mathrm{CRI550}_{2020.03.30} + \beta_{N.CARI2\,2020.04.04} N * CARI2_{2020.04.04} \\
& + \beta_{N.CRI700\,2020.04.04} N * CRI700_{2020.04.04}
\end{aligned}
$$

EONR estimation validation

Plot a bundle of response curves Plotting the proposed model at different locations, it may be demonstrated that the shape of the response varies across different geographical locations (latitude, longitude) in the field. As shown above, the values for,, and were computed through plugging the MSIs for each corresponding geographical location in the field (see Table 6).

Table 5 Final candidate MSIs for the third phase of feature selection by RSM model reduction. After applying Lasso regression to all categories and dates, the following MSIs are identified and selected from the following categories presented in this table

MSIs Category	Spectral Index	Date
Biochemical	ARI2	2020.03.25
Chlorophyll	CARI2	2020.03.25
Chlorophyll	CARI	2020.03.30
Biochemical	CRI550	2020.03.30
Chlorophyll	CARI2	2020.04.04
Biochemical	CRI700	2020.04.04
Chlorophyll	LCI	2020.04.04
Vegetation	NDRE	2020.04.04

One can also visualize the response curve at each geographical location. Starting with one value of intercept c_0, and one value of coefficient c_1, and constant c_2, the shape of yield-N response can be drawn at each location in the field. The result can be seen in Fig. 9. A bundle of curves for several points in the field is also plotted (see Fig. 10) to reflect the spatial heterogeneity of EONR across the field.

Discussion

In this investigation, a second-order RSM was developed as an optimization strategy for determining the EONR. The methodology involved integrating N recommendations provided by on-machine N-sensor technology with MSIs derived from Sentinel-2 L1C. A total of 21 MSIs were extracted from three image acquisition dates, and these indices, along with crop sensing technology, were employed to ascertain the EONR across the field by fitting a second-order RSM. To explore the interaction between MSIs from multispectral imagery and N-rate recommendations obtained through crop sensing technology, ANOVA was conducted using a second-order RSM fitted to each individual index. A significance level of 0.01 was employed for a selective choice of spectral indices. The determination of the EONR involves modeling yield with N alongside cofactors such as soil, topographical, and other environmental variables that influence yield outcomes. The analytical algorithm employed for this purpose can range from statistical methods to machine learning (ML). For instance, a study by de Lara et al. (2023) utilized environmental data to model EONR using ML, highlighting the challenge of generalizing findings across different fields, while the proposed model incorporates auxiliary variables (MSIs) rather than the aforementioned factors. This underscores the flexibility of our proposed model, which has the potential for integration with such factors to enhance its capabilities. It is important to note that ML-based approaches can be computationally intensive and less practical for real-time applications unless they undergo extensive training over a prolonged period. The determination of the EONR through RSM, which fused spectral vegetation indices with N-sensor recommendations, showcased a more robust estimation of N rate in modeling harvested yield. From a statistical point of view the true EONR remains uncertain (Nigon et al., 2019), and it is not fully clear what are the underlying limiting factor(s) in modeling response yield that reflect the sensitivity of choice of $Y(N)$ (Meyer-Aurich et al., 2022). It is noteworthy that the second-order RSM plays a crucial role in the accurate estimation of input N compared to plateau functions. Research conducted by Li et al. (2023) regarding the profitability of EONR in various trial setups revealed that a Latin square design or a specific pattern strip design tends to yield higher average profits compared to designs involving gradual N rate changes across the spatial layout (such as whole-field trials based on management zones), and therefore should be avoided if directly relevant information at field or farm level is required. Modeling yield via RSM and utilizing MSIs as auxiliary variables can automatically account for spatial autocorrelation, as the satellite image covers the entire field with a fine grid-size that captures granular spatial variability across the field. Such predictive variables reduce random effects during modeling with RSM, as opposed to using sparse environmental samples from the field. Our approach underscores the benefits of steering clear of small-plot designs and instead favoring whole-field trial designs, while also incorporating MSIs, for the specific purpose of determining in-field EONR. This approach allows for the

Table 6 The fitted model incorporates a set of acquired variable inputs (MSIs) to derive the coefficients c_0, c_1, and c_2, ultimately determining the maximum N-optimum. The highlighted rows in this table serve as visual representations of the corresponding responses. Note that c_2 is the coefficient of N^2 which is a scalar that is equal to −2.56

CARI2 2020.03.25	ARI2 2020.03.25	CARI2 2020.03.30	CRI550 2020.03.30	CARI2 2020.04.04	CRI700 2020.04.04	c_0	c_1	Max *N*-optimum
0.170876	0.5582213	0.2843609	0.3698111	0.1824791	1.510801	−1789.041	278.7468	54
0.1561802	0.4564984	0.267444	−1.123592	0.1636587	1.353497	−1248.175	285.8728	55
0.16303	0.5685227	0.2890172	−0.6702452	0.1695313	1.867213	−1404.962	324.8939	63
0.1615508	0.5573206	0.2813489	−1.102917	0.1664962	1.251833	−5393.32	313.3343	61
0.1682062	**0.6064325**	**0.3026892**	**−0.8783947**	**0.1705964**	**1.797223**	**−2100.952**	**357.4955**	**70**

adjustment of N based on the actual crop requirements at any given time, leading to a better precision for N allocation while considering the economic optimum point for each sub-region as N levels vary across the field. It is crucial to acknowledge that alterations in the spatial resolution of the input dataset and the data cleaning strategy may yield varied results in feature selection. Therefore, the fundamental significance of this research lies within the proposed methodological and analytical framework for modeling EONR, with the understanding that data pre-processing methods are subject to change based on expert knowledge and available data. By incorporating new MSIs into the fitted model for a given N range, it was demonstrated that the shape of the yield-N relation (RSM) varied spatially due to field heterogeneity. For example, two different locations in the field yielded distinct economic optimum N rates, approximately 70 kg at one geographical position and around 53 kg at another. To assess the effectiveness of the applied N compared to the estimated EONR, three main scenarios are anticipated in the dataset (Fig. 11). The first scenario occurs when the applied nitrogen quantity closely matches the EONR recommendation. The second scenario arises when the applied nitrogen falls below the optimal level, indicating under-fertilization for that specific location. Lastly, the third scenario occurs when the field is overfertilized for that location, signifying that the applied nitrogen exceeds the optimal point.

Therefore, implementing informative resource management is essential to allocate the optimum N rate at the sub-field level. Studies show that a concise timing of N application can be effective in increasing Nitrogen Use Efficiency (NUE), which can lead to reducing the number of split N applications. A study by Schulz et al. (2015) demonstrated that reducing the number of split N applications can lead to higher yield and at the same time increase in the NUE if the late first N application can take place between advanced tillering and beginning of stem elongation either broadcast or placed. Therefore, achieving the appropriate total distribution of N throughout the season and combining accurate placement or spatial distribution of N with proper timing can improve production systems and sustainability by optimizing the utilization of chemical compounds. The proposed approach can be implemented in practical applications through real-time adjustment of N or prescription mapping. In a real-time application scenario, integrating on-machine N sensors with multispectral imagery can generate a robust and accurate EONR, while a prescription map can be generated just before entering the field and transmitted to a sprayer for execution. The computational time for RSM methodology enables a viable real-time application. Highlighting the limitation in estimating EONR, a study by Baum et al. (2024) with maize ranks the influence of genotype, environment, and management (G × E × M) factors on EONR. Environment (41%) has the most impact, followed by management (31%) and genetics (27%). Key drivers include weather variability and soil nitrogen carryover, with management and genetic adaptations helping buffer climate change effects on EONR and maize yield. Another important limitation is the great temporal variation of EONR that can affect the estimation of EONR. A study attempted to calibrate the APSIM model to assess nitrogen needs in different field zones, finding it effectively predicts corn yield but underestimates the EONR. Temporal variability in EONR was greater than spatial, and better site-specific data could improve accuracy (Thompson et al., 2024). The limitations of the proposed framework in general can be described in three main aspects: Data Preparation: this research acknowledges that changes in spatial resolution, interpolation, and extrapolation strategies may impact feature selection results. Variations in data cleaning strategies could also influence model performance. Model Scalability: The scalability of the proposed

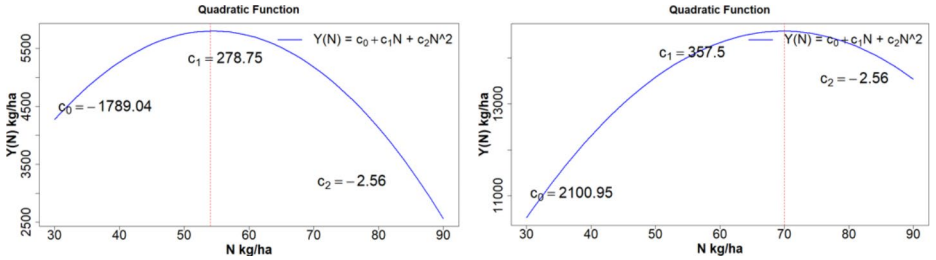

Fig. 9 Representation of quadratic functions for two different geographical positions in the field, showing variation in yield response to nitrogen application. The left plot demonstrates a maximum yield at 54 kg/ha of nitrogen with an average yield of ~4000 kg/ha, while the right plot shows a maximum yield at 70 kg/ha of nitrogen with an average yield of ~6500 kg/ha. These differences highlight the variability in yield potential and nitrogen requirements across the field. The spatial distribution of the Economic Optimum Nitrogen Rate (EONR) is depicted as a prescription map (see Fig. 10) and can be transmitted to machinery for execution

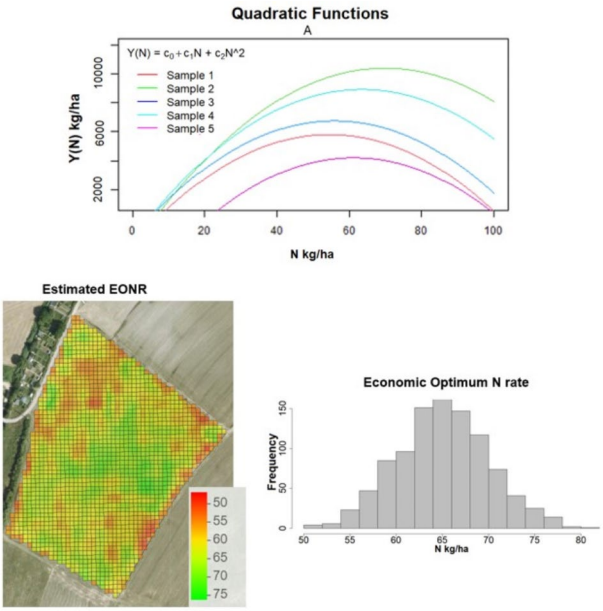

Fig. 10 Plot A is an example of a bundle of data points (samples) at different geographical positions which represent different yield-N response shapes (different intercept and maximum point) across the field. The map indicated by "A" demonstrates the spatial distribution of the estimated optimum N-rate derived from the prediction set given the estimated parameters. The spatial distribution of the EONR is illustrated through a prescription map with 5-meter spatial resolution, which can be transmitted to the machine for accurate N application. This map is the output of EONR calculations which reveal spatial clusters of similar EONR, aligning with the anticipated natural behavior where nearby areas exhibit similar characteristics compared to distant areas. In essence, this reflects the principle that proximity implies similarity, as observed in nature. C is the histogram distribution of estimated EONR with min and max values of EONR ranging from 45 kg to 75 kg, respectively

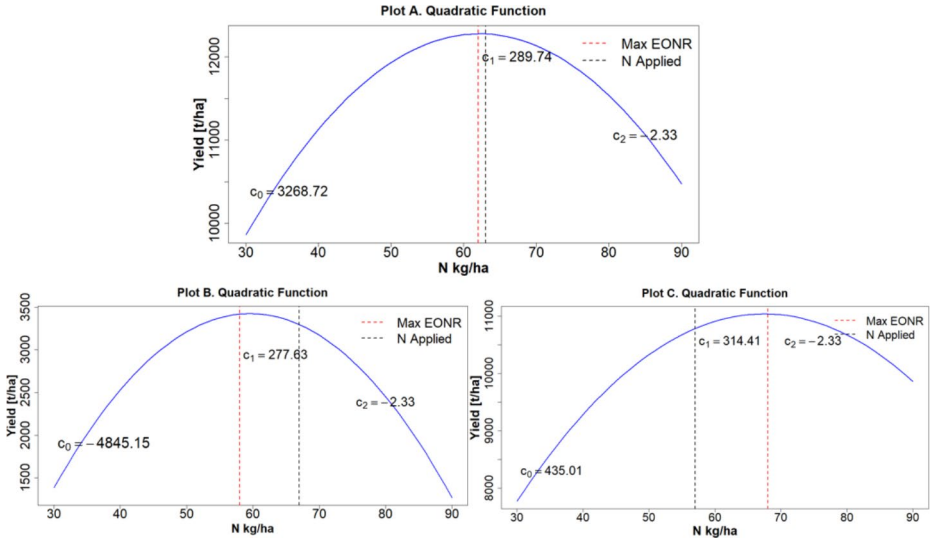

Fig. 11 Illustration of three scenarios: A shows a location where the applied nitrogen aligns closely with the estimated EONR, B displays an overfertilization scenario in a different location within the field, and C depicts an under-fertilization scenario at another location in the same field

model across different environmental conditions remains a challenge. New studies should validate this method under varying soil, climate, and crop conditions. Future Studies: Future research should incorporate protein content as a response variable to better represent both yield quantity and quality in the EONR model.

Conclusion

The advancement in on-farm experimentation using PA equipment has empowered the collection of detailed data from individual fields, facilitating a robust analysis. The wealth of observations across a field aids in minimizing biases and random effects during analysis, enabling informed decision-making on a field-by-field basis without the necessity of integrating data from other fields for generalization purposes. The limited availability of soil and climate data on individual farms is mitigated to some extent by indirect measurement of crop status enabled by RS data, thereby addressing the challenges associated with N estimation in agriculture. The proposed method involves the conversion of reflectance values into MSIs that accurately reflect crop nutritional status. Continuous refinement of the model through the integration of several years of observations enhances the accuracy of determining the EONR. This study utilizes observational data from a single field as a proof of concept for determining in-season EONR. While cross-validation was not within the scope of this paper, incorporating observations from other fields within the same farm and crop can facilitate a meta-analysis and lead to the development of a more generalizable model across the farm. This, in turn, enables robust cross-validation through the leave-one-field-out method. Exploring the generalizability of RSM models using meta-analytic approaches to integrate results from multiple fields is an interesting subject for future research. It is crucial

to consider the crop growth stage and spatial resolution of the indices for improved accuracy in yield monitoring and N estimation. Additionally, integrating various forms of information further enhances the precision of estimation. The RSM provides an interpretable and less complex way of integrating MSIs and on-machine N sensor technology, supporting the determination of the N optimum rate. Careful consideration of optimal dates and MSIs is vital for accurate modeling of N response. RSM optimization approaches prove more robust in estimating the EONR compared to simple quadratic or plateau approaches. Data quality plays a pivotal role in understanding and modeling the true relationship between input and output. The integration of diverse data sources can enhance model accuracy, resulting in a precise estimation of the EONR. Continuous refinement and adaptation of these methodologies are essential for advancing the field of PA. Future studies can bring in protein content as a second response along with the mass of the harvested grain to come up with an EONR that represents not only the optimum quantity of the yield but also takes the quality indicated by protein into the account. In the proposed analytical approach, the engagement of farmers through a participatory approach, involving annual *post mortem* analysis at the end of each season, can be easily integrated. The retrospective assessment using the developed RSM approach of the N input that would have been economically optimal, given the observed MSIs values for each location, allows farmers a systematic comparison with the actual rates applied. The integration of on-machine sensor technology and satellite imagery enables real-time N application execution, hence resulting in better accuracy.

Acknowledgements The authors acknowledge the support of the John Deere Intelligent Solution Group and a partner farm in facilitating this research. Appreciation is extended to Tobias Füge, the corresponding grower, for supporting the trial and granting permission for data sharing and open access publication for the research community.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The datasets analyzed during the current study are available along with the source code on GitHub in "In-season-EONR-estimation-with-smart-farming-data" repository, which can be found under this link.

Declarations

Ethical Statement This manuscript adheres to the ethical standards and guidelines set forth by the Precision Agriculture journal at Springer. Any potential conflicts of interest have been disclosed, and the research has been carried out with the highest degree of integrity and transparency. The authors affirm that any previously published work that has informed this research has been properly cited. The team is committed to upholding the principles of ethical conduct in research and scholarly publication.

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Abdipourchenarestansofla, M., & Piepho, H.P. (2022). Yield observation outlier detection with unsupervised machine learning in harvest machines. *VDI-Berichte, No. 2395, 2022.* Online: 79th International Conference on Agricultural. https://doi.org/10.51202/9783181023952-343

Abdipourchenarestansofla, M., & Schroth, C. (2022). The importance of data quality assessment for machinery data in the field of agriculture. *VDI-Berichte, No. 2395. 2395.* Online: 9th International Conference on Agricultural. https://doi.org/10.51202/9783181023952-495

Agrahari, R. K., Kobayashi, Y., Tanaka, T. S. T., Panda, S. K., & Koyama, H. (2021). Smart fertilizer management: the progress of imaging technologies and possible implementation of plant biomarkers in agriculture. *Soil Science and Plant Nutrition, 67(3)*, 248–258. https://doi.org/10.1080/00380768.2021.1897479

Athearn, K., Yarick, M., & Parks, N. (2021). Knowing Your Product Costs: A Primer for Farmers and Food Entrepreneurs. *FE1103, 11/2021", EDIS*. doi: 10.32473/edis-FE1103-2021.

Baum, M. E., Sawyer, J. E., Castellano, M. J., & Archontoulis, S. V. (2024). Ranking genotype, environment, management effects on the optimum nitrogen rate for maize: A cropping system modeling analysis. *Agronomy Journal, 116*(4), 1775–1791. Retrieved from https://doi.org/10.1002/agj2.21596

Box, G.E.P., & Draper, N.R. (2007). *Response Surfaces, Mixtures, and Ridge Analyses.* New York: 2nd Edition, John Wiley & Sons. Retrieved from https://onlinelibrary.wiley.com/doi/book/10.1002/0470072768

Clevers, J.G.P.W., & Gitelson, A.A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *International Journal of Applied Earth Observation and Geoinformatics, 23*, 344-351. doi:https://doi.org/10.1016/j.jag.2012.10.008

de Lara, A., & Mieno, T., & Luck, J.D., & Puntel, L.A. (2023). Predicting site-specific economic optimal nitrogen rate using machine learning methods and on-farm precision experimentation. *Precision Agric, 24*, 1792–1812. doi:https://doi.org/10.1007/s11119-023-10018-8

Diacono, M., Rubino, P., & Montemurro, F. (2013). Precision nitrogen management of wheat. A review. *Agronomy for Sustainable Development, 33*, 219-241. https://doi.org/10.1007/s13593-012-0111-z

Fan, Y., Feng, H., & Yue, J. (2023). Comparison of different dimensional spectral indices for estimating nitrogen content of potato plants over multiple growth periods. *Remote Sens, 15*, 1–17. doi:https://doi.org/10.3390/rs15030602

Fausti, S., Erickson, B.J., Clay, D.E., & Carlson, C.G. (2017). Deriving and using an equation to calculate economic optimum fertilizer and seeding rates. In S. A. David, & E. Clay, *Practical Mathematics for Precision Farming* (pp. 181–189). American Society of Agronomy. doi:https://doi.org/10.2134/practicalmath2016.0027

Guérif, M., Houlès, V., & Baret, F. (2007). Remote sensing and detection of nitrogen status in crops. Application to precise nitrogen fertilization. *In: 4th International Symposium on Intelligent Information Technology in Agriculture*, (pp. 26–29). Beijing. Retrieved from https://hal.inrae.fr/hal-02824189

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., & Wang, J. (2018). A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy, 92*, 141–152. doi:https://doi.org/10.1016/j.eja.2017.11.002

Lenth, R. V. (2009). Response-surface methods in R, using rsm. *In Journal of Statistical Software, 32(7)*, 1–17. doi:10.18637/jss.v032.i07

Leroux, C. (2020, January 3). *Aaspexit*. Retrieved from Aaspexit- Precision Agriculture: https://www.aspexit.com/yield-maps-in-precision-agriculture/

Li, X., Mieno, T. & Bullock, D.S. (2023). The economic performances of different trial designs in on-farm precision experimentation: a Monte Carlo evaluation. *Precision Agric, 24*, 2500–2521. doi:https://doi.org/10.1007/s11119-023-10050-8

Li, Z., Zhou, X., Cheng, Q., Fei, S., & Chen, Z. (2023). A machine-learning model based on the fusion of spectral and textural features from UAV multi-sensors to analyse the total nitrogen content in winter wheat. *Remote Sens, 15*(8), 2152. doi:https://doi.org/10.3390/rs15082152

Meyer-Aurich, A., & Karatay, Y. N. (2022). Greenhouse gas mitigation costs of reduced nitrogen fertilizer. *Agriculture, 12*(9), 1438. doi:https://doi.org/10.3390/agriculture12091438

Montero, D., Mahecha, M. D., Martinuzzi, F., Söchting, M., & Wieneke, S. (2023). A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research. *Scientific Data, 10*, 197. doi:https://doi.org/10.1038/s41597-023-02096-0

Nelder, J. A. (2000). Functional marginality and response-surface fitting. *Journal of Applied Statistics, 27(1)*, 109–112. doi:https://doi.org/10.1080/02664760021862

Nigon, T. J., Yang, C., Mulla, D. J., & Kaiser, D. E. (2019). Computing uncertainty in the optimum nitrogen rate using a generalized cost function. *Computers and Electronics in Agriculture, 167*, 105030. doi:https://doi.org/10.1016/j.compag.2019.105030

Piepho, H. P., Richter, C., Spilke, J., Hartung, K., Kunick, A., & Thöle, H. (2011). Statistical aspects of on-farm experimentation. *Crop and Pasture Science, 62(9)*, 721–735. Retrieved from https://doi.org/10.1071/CP11175

Piepho, H.-P., & Edmondson, R. N. (2018). A tutorial on the statistical analysis of factorial experiments with qualitative and quantitative treatment factor levels. *J Agron Crop Sci, 204(5)*, 429–455. doi: https://doi.org/10.1111/jac.12267

Prasad S., &. T. (2000). Hyperspectral vegetation indices and their relationships with agriculture crop characteristics. *Remote Sensing of Environment, 71(2)*, 158–182. doi:https://doi.org/10.1016/S0034-4257(99)00067-X

Rane, N. L., Giduturi, M., Choudhary, S. P., & Pande, C. B. (2023). Remote sensing (RS) and geographical information system (GIS) as a powerful tool for agriculture applications: efficiency and capability in agricultural crop management. *International Journal of Innovative Science and Research Technology, 8(4)*, 264–274. doi:https://doi.org/10.5281/zenodo.7845276

Schulz, R., Makary, T., Hubert, S., Hartung, K., Gruber, S., Donath, S., Döhler, J., Wei, K., Ehrhart, E., Claupein, W., & Piepho, H. P. (2015). Is it necessary to split nitrogen fertilization for winter wheat? On-farm research on Luvisols in South-West Germany. *The Journal of Agricultural Science Cambridge, 153(4)*, 575–587. https://doi.org/10.1017/S0021859614000288

Sentinel Hub Custom Scripts (2024). Index Database (INDEX DB) for Sentinel-2. Available online: https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/indexdb/.

Smerald, A., Kraus, D., Rahimi, J., Fuchs, K., Kiese, R., Butterbach-Bahl, K., & Scheer, C. (2023). A redistribution of nitrogen fertiliser across global croplands can help achieve food security within environmental boundaries. *Commun Earth Environ, 315(4)*, 1–11. Retrieved from https://doi.org/10.1038/s43247-023-00970-8

Thompson, L.J., Archontoulis, S.V. & Puntel, L.A. (2024). Simulating within-field spatial and temporal corn yield response to nitrogen with APSIM model. *Precision Agriculutre*. Retrieved from https://doi.org/10.1007/s11119-024-10178-1

Wang, C., Feng, M. C., Yang, W. D., Ding, G. W., Sun, H., Liang, Z. Y., & Qiao, X. X. (2016). Impact of spectral saturation on leaf area index and aboveground biomass estimation of winter wheat. *Spectroscopy Letters, 49(4)*, 241–248. https://doi.org/10.1080/00387010.2015.1133652.

Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A metareview. *Remote Sensing of Environment, 236*, 1–39. https://doi.org/10.1016/j.rse.2019.111402

Weyand, B. (2022). *Reduce your carbon footprint with green fertilizer*. Retrieved from www.effizientduengen.de: https://www.effizientduengen.de/2022/co2-fussabdruck-mit-gruenem-duenger-reduzieren/.

Yousfi, S., Marin Peira, J. F., Rincón De La Horra, G., & Mauri Ablanque, P. V. (2019). remote sensing: useful approach for crop nitrogen management and sustainable agriculture. In M. C. M. Hasanuzzaman, *Sustainable Crop Production.* Rijeka: IntechOpen. doi:10.5772/intechopen.89422.

Zera, J., Lukas, V., Horniacek, I., Smutný, V., & Elbl, J. (2021). Comparison of proximal and remote sensing for the diagnosis of crop status in site-specific crop management. *Sensors, 22(1)*, 19. doi:https://doi.org/10.3390/s22010019

Zhao, X., Zhang, Y., Long, T., Wang, S., Yang, J. (2022). Regulation mechanism of plant pigments biosynthesis: Anthocyanins, Carotenoids, and Betalains. *Metabolites, 12*(9), 871 Retrieved from https://doi.org/10.3390/metabo12090871

Zheng, J., Song, X., Yang, G., Du, X., Mei, X., Yang, X. (2022). Remote sensing monitoring of rice and wheat canopy nitrogen: A review. *Remote Sensing, 14*(22), 5712. https://doi.org/10.3390/rs14225712

Zhou I., Huang W., Zhang J., Kong W., Casa R., Huang Y. (2019). A novel combined spectral index for estimating the ratio of carotenoid to chlorophyll content to monitor crop physiological and phenological status. *International Journal of Applied Earth Observation and Geoinformation, 76*, 128–142. https://doi.org/10.1016/j.jag.2018.10.012.

Authors and Affiliations

Morteza Abdipourchenarestansofla[1,2] · Hans-Peter Piepho[1]

✉ Morteza Abdipourchenarestansofla
morteza.abdipour92@gmail.com

[1] Biostatistics Unit, Institute of Crop Science, University of Hohenheim, Stuttgart, Germany

[2] Intelligence Solution Group, John Deere GmbH & Co KG, Kaiserslautern, Germany