Mapping Knowledge Domains of Regenerative Agriculture with a Focus on On-Farm Nitrogen Fertilization Experimentation and Response Surface Regression

Dissertation to obtain the doctoral degree of Agricultural Sciences *(Dr. sc. agr.)*

Faculty of Agricultural Sciences

University of Hohenheim

Institute of Crop Science (340)

Submitted by

Morteza Abdipourchenarestansofla

from Boyer Ahmad, Yasuj

2025

Dean: Prof. Dr. Ralf Vögele

Date of oral examination: 29 Sep 2025

Examination Committee

Supervisor and Reviewer	Prof. Dr. Hans-Peter Piepho
Examiner	Prof. Dr.-Ing. Peter Pickel
Examiner	Prof. Dr. Simone Graeff-Hönninger
Head of the examination committee	Prof. Dr. Martin Hasselmann

Contents

Table of Figures

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

Table of Tables

Chapter 2:

Chapter 3:

Chapter 4:

Abbreviations and acronyms

ANOVA - Analysis of Variance

API - Application Programming Interface

BBCH - Biologische Bundesanstalt, Bundessortenamt, and CHemical industry (Phenological growth stages)

CM - Contribution Margin

EONR - Economic Optimum Nitrogen Rate

FMI - Farm Management Information System

GA - Generative Agriculture

GIS - Geographic Information System

IoT - Internet of Things

KDM - Knowledge Domain Mapping

ML - Machine Learning

MSI - Multispectral Index

N - Nitrogen

NUE - Nitrogen Use Efficiency

PA - Precision Agriculture

PNC - Plant Nitrogen Content

PoC - Proof of Concept

RA - Regenerative Agriculture

RMSE - Root Mean Square Error

RS - Remote Sensing

RSM - Response Surface Model

SF - Smart Farming

TC - Total Cost

TR - Total Revenue

Summary

In the face of growing environmental concerns and the global demand for sustainable agriculture, achieving balanced nitrogen (N) management is critical for both maximizing crop productivity and maintaining environmental health. This dissertation proposes an innovative framework to address this challenge within the scope of regenerative agriculture, which emphasizes sustainable farming practices. Regenerative agriculture aims to reduce chemical inputs while maintaining yield levels yet implementing these practices at scale is complex due to the intricate interactions between biological, environmental, and technological factors on farms. This research tackles these challenges by introducing a Knowledge Domain Mapping (KDM)-based framework, integrating advanced technologies—including remote sensing, Internet of Things (IoT) telemetry, geospatial sciences, statistical modeling, machine learning, and cloud computing—to create a holistic and scalable system for optimizing nitrogen applications.

Central to this research is the accurate estimation and spatial allocation of the Economic Optimum Nitrogen Rate (EONR), a crucial element for reducing nitrogen use and enhancing yield. A key contribution of this study is the development of a robust Response Surface Model (RSM) that leverages multispectral indices (MSIs) from Sentinel-2 imagery, historical IoT telemetry data, and on-machine nitrogen sensors. This RSM approach allows for precise EONR predictions tailored to field-specific conditions, reducing the need for traditional plot-based trials and achieving an average prediction error of just 14.5%. Applied to a 7-hectare winter wheat field, the model successfully predicted EONR values ranging from 43 kg/ha to 75 kg/ha across zones, showcasing the adaptability and accuracy of RSM for field-specific nitrogen recommendations. This precision-focused approach exemplifies the study's goal of minimizing environmental impacts while ensuring sustainable yield improvements.

Beyond the initial field-level implementation, this research examines the generalizability of the RSM framework using two modeling strategies: a single RSM across fields and a weighted average model that aggregates individual field-specific RSMs. The weighted model demonstrated superior adaptability and high prediction accuracy, with a root mean square error (RMSE) of 11 kg N/ha for the EONR, highlighting the framework's potential for broader application across different agricultural settings. Such generalizability supports the framework's adoption in diverse farming environments, enabling precise and informed nitrogen management at scale.

To facilitate widespread adoption and practical application, the dissertation also introduces a cloud-based infrastructure that integrates the KDM framework with real-time IoT data and satellite imagery. Leveraging cloud services like Amazon Web Services (AWS) Batch for job orchestration, Amazon S3 for scalable data storage, and RDS Postgres for structured data management, this

infrastructure allows for seamless handling of both real-time and historical data. Spatial interpolation techniques, such as Kriging, enhance the model's capability to generate real-time nitrogen prescription maps, enabling precise nutrient management for large-scale agricultural operations. Automated data quality control and data harmonization embedded within this cloud architecture provide a strong foundation for managing increasing data volumes and diverse field conditions, making the system cost-effective, adaptable, and efficient for modern agriculture.

In summary, this dissertation maps regenerative agriculture via a comprehensive roadmap for translating regenerative agriculture principles into practical, operational nitrogen management practices. Through KDM an interdisciplinary approach is mapped by the integration of advanced modeling, data processing, and cloud technologies. This framework enables sustainable crop management and aligns with global goals for environmentally responsible food production. The innovations introduced here support a scalable, data-driven approach to agricultural sustainability, bridging scientific research with real-world applications to meet the evolving demands of modern agriculture.

Zusammenfassung

Angesichts wachsender Umweltbedenken und der globalen Nachfrage nach nachhaltiger Landwirtschaft ist ein ausgewogenes Stickstoffmanagement (N-Management) entscheidend, um sowohl die Produktivität der Kulturen zu maximieren als auch die Umweltgesundheit zu erhalten. Diese Dissertation schlägt ein innovatives Rahmenkonzept vor, um diese Herausforderung im Kontext der regenerativen Landwirtschaft zu bewältigen, die nachhaltige Anbaumethoden betont. Ziel der regenerativen Landwirtschaft ist es, den Einsatz chemischer Betriebsmittel zu reduzieren und gleichzeitig stabile Erträge zu gewährleisten. Die Umsetzung solcher Praktiken im großen Maßstab ist jedoch aufgrund der komplexen Wechselwirkungen zwischen biologischen, ökologischen und technologischen Faktoren auf landwirtschaftlichen Betrieben äußerst anspruchsvoll. Diese Forschung begegnet diesen Herausforderungen durch die Einführung eines auf Knowledge Domain Mapping (KDM) basierenden Rahmenwerks, das fortschrittliche Technologien wie Fernerkundung, Telemetrie des Internets der Dinge (IoT), Geowissenschaften, statistische Modellierung, maschinelles Lernen und Cloud-Computing integriert, um ein ganzheitliches und skalierbares System zur Optimierung des Stickstoffeinsatzes zu schaffen.

Im Mittelpunkt dieser Forschung steht die präzise Schätzung und räumliche Verteilung der wirtschaftlich optimalen Stickstoffdosis (Economic Optimum Nitrogen Rate, EONR), ein entscheidendes Element zur Reduzierung des Stickstoffeinsatzes und zur Ertragssteigerung. Ein zentraler Beitrag dieser Arbeit ist die Entwicklung eines robusten Response Surface Models (RSM), das multispektrale Indizes (MSIs) aus Sentinel-2-Satellitenbildern, historische IoT-Telemetriedaten und Sensordaten von Stickstoffapplikationsgeräten nutzt. Dieser RSM-Ansatz ermöglicht genaue EONR-Vorhersagen, die auf feldspezifische Bedingungen abgestimmt sind, wodurch der Bedarf an traditionellen, parzellenbasierten Versuchen reduziert wird. Mit einem durchschnittlichen Vorhersagefehler von nur 14,5 % wurde das Modell in einem 7 Hektar großen Winterweizenfeld angewendet und konnte EONR-Werte von 43 kg/ha bis 75 kg/ha in verschiedenen Zonen erfolgreich vorhersagen. Dies verdeutlicht die Anpassungsfähigkeit und Genauigkeit des RSM für feldspezifische Stickstoffempfehlungen. Dieser präzisionsorientierte Ansatz illustriert das Ziel der Studie, Umweltbelastungen zu minimieren und gleichzeitig nachhaltige Ertragssteigerungen zu gewährleisten.

Über die anfängliche Anwendung auf Feldebene hinaus untersucht diese Arbeit die Generalisierbarkeit des RSM-Rahmenwerks mit zwei Modellierungsstrategien: einem einheitlichen RSM über mehrere Felder hinweg und einem gewichteten Durchschnittsmodell, das individuelle, feldspezifische RSMs zusammenfasst. Das gewichtete Modell zeigte eine überlegene

Anpassungsfähigkeit und hohe Vorhersagegenauigkeit mit einem Root Mean Square Error (RMSE) von 11 kg N/ha für die EONR und unterstreicht das Potenzial des Rahmenwerks für eine breitere Anwendung in verschiedenen landwirtschaftlichen Kontexten. Diese Generalisierbarkeit unterstützt die Akzeptanz des Rahmenwerks in unterschiedlichen Anbausystemen und ermöglicht eine präzise und informierte Stickstoffbewirtschaftung im großen Maßstab.

Um eine breite Anwendung und praktische Umsetzung zu fördern, führt die Dissertation zudem eine cloudbasierte Infrastruktur ein, die das KDM-Rahmenwerk mit Echtzeit-IoT-Daten und Satellitenbildern integriert. Durch den Einsatz von Cloud-Diensten wie AWS-Batch für die Job-Orchestrierung, Amazon S3 für skalierbare Datenspeicherung und RDS Postgres für die strukturierte Datenverwaltung ermöglicht diese Infrastruktur eine nahtlose Verarbeitung von Echtzeit- und historischen Daten. Räumliche Interpolationstechniken wie Kriging erweitern die Fähigkeit des Modells, Echtzeit-Stickstoffempfehlungskarten zu erstellen, was eine präzise Nährstoffbewirtschaftung für großflächige landwirtschaftliche Betriebe ermöglicht. Automatisierte Qualitätssicherung und Datenharmonisierung, die in diese Cloud-Architektur eingebettet sind, bieten eine solide Grundlage für die Bewältigung wachsender Datenmengen und vielfältiger Feldbedingungen und machen das System kosteneffizient, anpassungsfähig und leistungsfähig für die moderne Landwirtschaft.

Zusammenfassend bietet diese Dissertation eine umfassende Roadmap, um Prinzipien der regenerativen Landwirtschaft in praktische, operative Stickstoffmanagementpraktiken zu übersetzen. Durch KDM wird ein interdisziplinärer Ansatz abgebildet, der fortschrittliche Modellierung, Datenverarbeitung und Cloud-Technologien integriert. Dieses Rahmenwerk ermöglicht eine nachhaltige Pflanzenbewirtschaftung und unterstützt die globalen Ziele einer umweltbewussten Nahrungsmittelproduktion. Die eingeführten Innovationen fördern einen skalierbaren, datengetriebenen Ansatz für landwirtschaftliche Nachhaltigkeit und schlagen eine Brücke zwischen wissenschaftlicher Forschung und realen Anwendungen, um den sich wandelnden Anforderungen der modernen Landwirtschaft gerecht zu werden.

Chapter 1: General Introduction

1.1. Background and Motivation

The research presented in this thesis builds on an extended effort initially supported by funding from the DEMETER Horizon 2020 project (857202) supported by European Union and John Deere European Technology Innovation Center (ETIC) in Kaiserslautern, Germany. DEMETER H2020 aimed to establish a cross-domain digital agri-food space to improve interoperability across European agri-food sectors. John Deere contributed to this EU initiative by actively participating in research activities and applying project insights on its corporate farms. The findings of this work were communicated and validated with project stakeholders, ensuring practical relevance and alignment with industry needs.

This research identifies key knowledge domains essential for regenerative agriculture, focusing on in-season nitrogen management. By combining scientific insights with scalable solutions, it aims to create a sustainable framework for nitrogen management. To achieve this goal, the study introduces a novel approach for determining the Economic Optimum Nitrogen Rate (EONR) in winter wheat, validated through field trials and practical applications. The framework is then expanded to ensure adaptability across diverse field conditions, supporting more generalized nitrogen optimization. Finally, the study refines these methods into a production-ready solution, underpinned by cloud-based infrastructure, enabling efficient, data-driven decision-making for crop management in regenerative agricultural systems.

1.2. Mapping Knowledge Domains of Regenerative Agriculture

Regenerative Agriculture (RA) is increasingly valued as a sustainable approach to farming, prioritizing soil health, biodiversity, and ecosystem resilience (Sharma et al., 2024; Pontius & McIntosh, 2024; LaCanne & Lundgren, 2018). Unlike conventional agriculture, RA aims to regenerate degraded soils, improve water retention, and foster biodiversity by employing minimal tillage, crop diversification, organic inputs, and carbon sequestration (Gosnell et al., 2020). These principles support long-term soil productivity, aligning with global goals for climate resilience and food security (Schreefel et al., 2020).

Knowledge Domain Mapping (KDM) serves as the framework for this research, integrating multiple scientific and technological domains to optimize RA practices. KDM enables a comprehensive understanding of RA by organizing information and identifying research gaps, which is essential for advancing sustainable and site-specific practices (Chen et al., 2019). The application of KDM in RA encompasses:

1. **Soil Science and Health**: Central to RA, soil health is enhanced through practices like cover cropping, minimal tillage, and crop rotation, which build soil organic matter, support microbial diversity, and boost nutrient cycling (Montgomery et al., 2017; Lal, 2020).

2. **Agroecology and Biodiversity**: RA encourages crop diversification and agroforestry to reduce pest pressure and enhance resilience. These practices improve nitrogen use efficiency, stabilize nutrient levels, and support beneficial soil microbes (Schreefel et al., 2020; Gosnell et al., 2020).

3. **Climate Science**: RA practices such as no-till farming, cover cropping, and organic amendments contribute to carbon sequestration and greenhouse gas reduction, fostering soil and climate resilience (Lal, 2020; Montgomery et al., 2017).

4. **Technological Integration**: Tools like GIS, remote sensing, Internet of Things (IoT), and machine learning facilitate precise monitoring and management in RA. These technologies allow for efficient nutrient application and soil condition assessment, supporting RA's data-driven goals (Kaur et al., 2024; Dutta et al., 2019).

Data-Driven Nitrogen Management is integral to RA, as efficient nitrogen (N) application reduces environmental risks such as soil acidification, water contamination, and greenhouse gas emissions. This research employs advanced nitrogen management techniques, using tools like remote sensing and IoT for precise nitrogen application aligned with ecological goals (Sharma et al., 2024; Dutta et al., 2019). By enhancing nitrogen efficiency through data-driven strategies, this research establishes a foundation for improving soil and ecological health, with findings adaptable to future trials involving organic fertilizers.

Broader Ecological Benefits of RA include improved water quality, reduced emissions, and increased biodiversity. By limiting synthetic inputs, RA protects natural resources and supports stable agricultural ecosystems (Paustian et al., 2016; Kremen & Miles, 2012). Through KDM, this dissertation synthesizes interdisciplinary insights to scale RA practices effectively, demonstrating that RA can address modern agriculture's dual goals of productivity and sustainability (World Economic Forum, 2024; Sustainable Markets Initiative, 2024).

This research centers on optimizing nitrogen management as a critical agronomic challenge in RA. By developing a robust, data-driven framework for sustainable nitrogen application tailored to crop needs, it bridges scientific insights with practical, on-farm applications, providing a foundational approach adaptable to diverse fertilization methods for enhancing both soil and ecological health.

1.2.1. Mobile IoT and Remote Sensing

Recent advances in Mobile IoT technology, coupled with remote sensing (RS), are revolutionizing precision agriculture by enabling real-time, data-driven decisions to optimize nitrogen application and estimate the EONR. At the core of Mobile IoT in precision agriculture is the collection of machine telemetry and agronomic data, which provides insights into field operations, machinery efficiency, soil conditions, and crop health. Machine telemetry data, including GPS-tracked machinery routes, fuel usage, and operation metrics, help track and manage field activity, thus supporting efficient nitrogen application by tailoring practices to specific field zones. In addition, agronomic data collected from field operations offers valuable historical information on crop performance, soil health, and nutrient needs, further enabling site-specific nitrogen optimization (Weiss et al., 2020; Rane, 2023).

Mobile IoT and Precision Agriculture

Mobile IoT systems in precision agriculture integrate sensors and GPS on farm equipment to monitor field operations in real time. This allows for the seamless collection and analysis of data directly from machinery, providing actionable insights that can be adjusted based on crop needs, soil conditions, and environmental factors (Rane et al., 2023). This integration helps optimize nitrogen application by allowing farmers to make informed decisions based on field-level data and machinery performance. Historical telemetry data, for instance, can reveal patterns in crop nutrient requirements, thus allowing for more precise nitrogen application that aligns with EONR (Li et al., 2023).

Remote Sensing for Nitrogen Optimization

Remote sensing, especially through multispectral and hyperspectral imaging, has significantly enhanced nitrogen management in agriculture. Multispectral sensors capture data in discrete spectral bands, which are particularly useful for deriving vegetation indices like the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) that correlate with crop chlorophyll content, canopy structure, and nitrogen status (Zhou et al., 2019; Zheng et al., 2022). Hyperspectral sensors, with their ability to capture hundreds of narrow spectral bands, offer detailed information on plant physiology and biochemistry, making them valuable for detecting early nutrient deficiencies or crop stress (Weiss et al., 2020). Satellite-based multispectral imagery, for example, provides coverage at a broad scale and is suitable for monitoring crop health throughout the season, aiding in nitrogen adjustment based on temporal changes in vegetation indices.

N-Sensors for In-Field Nitrogen Adjustment

Machine-mounted N-sensors, which measure real-time nitrogen status in crops, offer a critical tool for site-specific nitrogen adjustment tailored to crop needs. These sensors, often mounted on tractors or other farm equipment, measure crop reflectance, capturing real-time data on nitrogen concentration in the canopy. By measuring indicators such as chlorophyll content, these sensors help farmers apply nitrogen precisely where it is needed, preventing over-application and reducing environmental impacts (Agrahari et al., 2021). This technology supports data-driven nitrogen management that adapts to variable field conditions, thus optimizing the EONR.

Integrating IoT and Remote Sensing Data for EONR Estimation

The combination of Mobile IoT data and remote sensing imagery forms a comprehensive framework for data-driven nitrogen management. Machine telemetry and agronomic data provide field-specific insights on operational efficiency and soil conditions, while remote sensing and N-sensors supply accurate, timely measurements of crop health and nutrient status (Dash et al., 2024). Fusing these data sources enables the creation of precise nitrogen application maps, facilitating EONR estimation by aligning nitrogen supply with crop demands over time. Historical data from IoT sensors can further enhance EONR calculations by accounting for seasonal patterns, soil types, and field history, making predictions more reliable and tailored to each unique field environment (Kaur et al., 2024).

By utilizing these interconnected technologies, this data-driven approach to nitrogen optimization not only improves crop yield and quality but also supports the sustainable use of resources, reducing the environmental impact of nitrogen over-application in regenerative agriculture.

1.2.2. On Farm Experimentation

Recent advancements in on-farm experimentation (OFE) have enhanced agricultural trials' effectiveness and precision, particularly in managing spatial variability and fostering farmer-researcher collaboration. Traditionally, OFE relied on plot-based or strip-based trials (see Figure 1.1) that, while useful, struggle to capture the inherent spatial heterogeneity within fields.

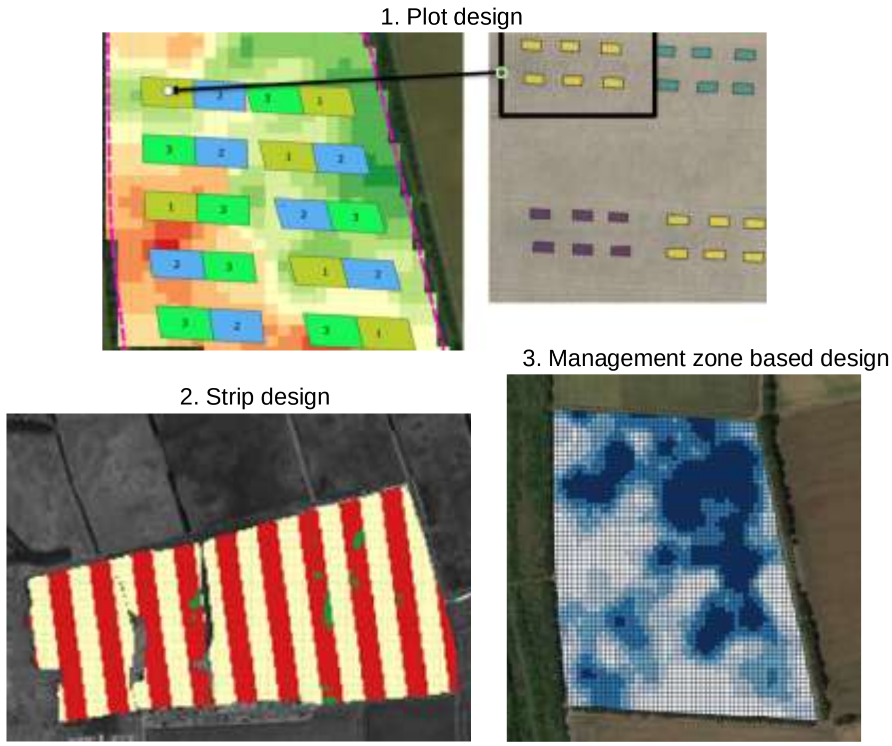

Figure 1.1. Transition from traditional experimental design to site-specific design.

This figure. Demonstrate traditional designs (1 and 2) versus large-scale, site-specific design (3) for nitrogen management trials. Such designs are resource-intensive, require extensive manual setup, and sample only a small portion of the field. Consequently, these traditional methods limit the applicability and scalability of findings, which can hinder precision agriculture's goals of site-specific management and sustainability.

Toffolini and Jeuffroy (2022) emphasized the importance of collaboration in OFE, highlighting how adaptive experimental designs tailored to individual farms can enhance relevance and support adoption. This approach underscores the role of farmer involvement and adaptable design in effectively managing diverse field conditions. A significant contribution from Cao, Grose, Brown, and Rakshit (2022) further demonstrated that systematic experimental designs are often more reliable than randomized layouts for capturing treatment effects across spatially variable fields. By aligning experimental layouts with natural field variation, systematic designs address a critical need in OFE for simplicity and scalability in large fields.

Advanced statistical modeling has also improved frameworks for addressing spatial variability, allowing for more accurate, site-specific recommendations. For instance, Trevisan, Bullock, and Martin (2021) used geographically weighted regression models to capture local crop yield responses, demonstrating the economic benefits of site-specific input management. Similarly, Tanaka (2021) found that spatial linear mixed models—both isotropic and anisotropic—provide greater precision than ordinary least squares regression for on-farm trials. These insights underscore the value of spatially adaptive models in precision agriculture, where field heterogeneity is a key variable.

With the integration of precision agriculture technologies, such as mobile IoT, GPS, and remote sensing, on-farm experimentation has evolved from traditional, isolated plot-based setups to a more practical, scalable approach using management zones. Management zones, characterized by similar soil and environmental conditions, act as experimental units that better represent real-world farming contexts and naturally occurring field variability. This approach allows for trials across broader zones instead of individual plots, providing a more realistic perspective on field conditions and reducing the complexity of trial setup.

By eliminating the need for complex randomization, management zone-based experimentation allows for efficient trial application, where treatments are assigned directly to zones with similar conditions. This method not only saves time and resources but also enhances data accuracy by reflecting true spatial variability within fields. Management zone-based designs facilitate integration with Farm Management Information Systems (FMIS), allowing farmers to conduct retrospective analyses, refine EONR models, and continuously adapt to environmental changes captured through high-resolution remote sensing.

The scalability of management zone-based experimentation encourages wider adoption of precision agriculture practices, supporting a seamless transition from small-scale trials to full-field applications. Integrated with decision-support systems, this method aligns site-specific nitrogen recommendations with daily farming operations, thereby enhancing both economic efficiency and environmental sustainability. By reducing dependency on resource-heavy traditional designs, the zone-based approach offers a scalable, site-specific solution that aligns with regenerative agriculture goals of long-term soil health and ecological stability, supporting sustainable and productive large-scale farming.

Management zones replace traditional plots by focusing on naturally occurring field variability, offering a scalable and practical approach that aligns with production-scale management practices.

1.2.3. Cloud Storge and Computing

Cloud Storage and Computing: Enabling Scalable Regenerative Agriculture

Cloud storage and computing have become essential technologies for modern agriculture, particularly within precision agriculture practices, by enabling scalable data management, analytics, and operational decision-making. Cloud computing refers to the delivery of computing services over the internet, including data storage, processing, and analytics capabilities. Cloud platforms provide these resources on-demand, allowing organizations to scale resources as needed, manage costs effectively, and enhance data accessibility and collaboration (Armbrust et al., 2021).

In the context of regenerative agriculture, cloud computing facilitates the transition of scientific insights and on-farm experimentation into operationally scalable practices by enabling seamless data flow and integration across diverse data sources. Such systems enhance decision-making frameworks by hosting complex data pipelines that merge real-time and historical data from remote sensing, agricultural machinery, and environmental sensors, allowing for comprehensive data-driven insights.

The Role of Cloud Computing in Smart Farming Data Management

Recent research illustrates the transformative role of cloud computing in managing, analyzing, and supporting decision-making processes in agriculture, especially in geospatial data processing. Cloud systems provide the necessary infrastructure to handle diverse data streams—such as telemetry from agricultural machinery, satellite imagery, and field documentation—that drive precision and regenerative agriculture. For instance, Yadav et al. (2024) explore cloud-based frameworks that integrate multi-source agricultural data, including satellite imagery and telemetry, to enhance nitrogen management. By centralizing vast telemetry datasets, the study shows how cloud platforms facilitate real-time decision-making at the field level, enabling farmers to adjust nitrogen application rates dynamically. This underscores cloud computing's capacity to support high-resolution data analytics, which is pivotal for precision agriculture. Similarly, Dash et al. (2024) developed a specialized cloud framework for large-scale geospatial analysis tailored to nitrogen management in winter wheat fields. Their findings highlight how cloud computing enhances real-time monitoring capabilities, providing farmers with immediate field insights that support dynamic nitrogen application decisions. This framework exemplifies how cloud systems enable scalable agricultural applications by managing complex data flows and enhancing the accuracy of resource allocation. Moreover, the GeoEkuiper system by Huang and Deng (2024) offers a cloud-native solution for real-time geospatial data processing, designed to operate effectively even on resource-constrained devices. GeoEkuiper is efficient at handling large-scale data related to soil moisture and nitrogen levels, providing critical support for granular decision-making in crop management. This system is particularly relevant in optimizing nitrogen application, as it enables farmers to utilize data-driven insights tailored to specific crop requirements and environmental conditions. In addition, Morchid et al. (2024) present a smart irrigation system that integrates IoT and cloud platforms to optimize water usage, promoting sustainable resource management. This study demonstrates how cloud-based systems can offer precise control over water resources, aligning with regenerative agriculture's principles by enhancing sustainability in agricultural practices. These studies collectively showcase cloud computing's capabilities in supporting data-driven agriculture through enhanced data integration, governance, and decision-making. Building upon these foundational works, our proposed cloud framework is specifically optimized to support decision-making in nitrogen management by determining the EONR. This framework integrates telemetry data from agricultural machinery with satellite imagery, offering a comprehensive, scalable solution for large-scale nitrogen optimization. Through real-time data processing and analytical functions, our framework aims to enable more sustainable, precise nitrogen management practices, thus operationalizing regenerative agriculture principles on a broader scale.

Cloud Computing: A Cornerstone of Regenerative Agriculture realization Practices

The scalability, data integration, and analytical capacity provided by cloud computing are fundamental to operationalizing regenerative agriculture at scale. By enabling real-time insights into crop health, soil conditions, and nutrient requirements, cloud platforms empower farmers to implement data-driven, sustainable practices, particularly in nitrogen management. This aligns with the knowledge domains mapped in regenerative agriculture by facilitating efficient, technology-driven farming practices. The integration of cloud computing within regenerative agriculture practices thus forms a critical technological foundation for advancing sustainable, large-scale farming solutions.

1.3. Economic Optimum Nitrogen Rate

EONR represents the nitrogen application rate that maximizes economic returns by balancing yield benefits against fertilizer costs. Determining EONR is critical for achieving optimal productivity without excess nitrogen, which can cause environmental harm through runoff and greenhouse gas emissions (Arnall et al., 2014). EONR estimation aligns with sustainable agricultural practices by addressing both economic and ecological goals. Achieving an accurate EONR ensures that N fertilization supports key processes like tillering, photosynthesis, and protein synthesis in crops, especially winter wheat, while minimizing environmental impacts (Cui et al., 2009; Wang et al., 2023).

1.3.1. In Season Mineral Fertilization

In-season nitrogen (N) fertilization for winter wheat refers to the application of N during key growth stages, particularly when the crop's nutrient demands are at their peak. This targeted approach allows farmers to supply only as much nitrogen as needed by the crop, promoting both agronomic and environmental benefits. A study by Cui et al. (2009) demonstrated that in-season nitrogen management (INM) strategies, which determine optimal nitrogen rates (ONR) based on soil nitrate-N tests and specific crop growth phases, can significantly enhance yield while reducing environmental impacts. Conducted across nine sites in the North China Plain, the study found that ONR ranged from 71 to 170 kg N ha^{-1}, with yields at ONR averaging 6.5 t ha^{-1}, matching the maximum yield predicted by yield response curves. Importantly, the study highlighted that applying more nitrogen beyond the ONR did not improve yield but increased residual soil nitrate, leading to higher nitrogen losses and potential environmental harm.

Efficient nitrogen application across a crop's growth stages—through regulated timing, rate, form, and placement—has become a priority to avoid the excessive use of chemical fertilizers, which have raised environmental concerns due to their potential for runoff and greenhouse gas emissions (Rane, 2023). Thus, a sustainable crop production system necessitates regulated resource allocation, particularly for fertilization and other crop management activities, to meet both productivity and ecological goals. To achieve this, it is crucial to monitor crop health and nitrogen content, which serve as foundations for efficient and sustainable fertilization strategies.

On-farm experiments provide valuable insights into how optimal nitrogen management can balance agronomic and environmental goals. For instance, Cui et al. (2009) demonstrated through in-season nitrogen management (INM) that targeting nitrogen applications at critical growth stages allows farmers to balance productivity with environmental responsibility, aligning nitrogen application rates with crop requirements and indigenous nitrogen sources.

Recently, the use of active canopy sensors and multispectral indices has improved in-season nitrogen management by enabling real-time monitoring of crop health. Meloni et al. (2024) demonstrated that the Normalized Difference Red Edge (NDRE) index could effectively guide N fertilization for winter wheat, achieving an R^2 value of 0.479 in grain yield prediction. While this approach improves efficiency, it also highlights a limitation: using RS indices alone may lack the precision needed for spatially variable nitrogen requirements across different fields. Furthermore, advances in remote sensing, especially through the integration of on-machine sensors like the RapidSCAN, have facilitated in-season nitrogen status diagnosis and top-dress recommendations in crops such as rice (Lu et al., 2024). Their machine learning-based approach using NDVI and NDRE indices showed an accuracy rate of 89% and improved nitrogen use efficiency (NUE) and yield by significant margins compared to farmer practices. These technologies illustrate the growing role of in-season monitoring for real-time, site-specific nitrogen adjustments, supporting sustainable production systems in modern agriculture.

1.3.2. Modeling Economic Optimum Nitrogen Rate

Early studies on EONR estimation primarily used traditional models with yield response curves from field trials. Arnall et al. (2014) evaluated nitrogen responsiveness across 261 years of data from cereal production trials in the U.S., finding that yield potential and nitrogen responsiveness were independent variables. This indicated that both factors must be considered independently for more accurate nitrogen recommendations. Such foundational approaches provided essential insights but did not account for complex interactions between environmental factors and crop health, which limits EONR accuracy for varied field conditions.

Recent advancements in remote sensing (RS) and machine learning (ML) have enabled more sophisticated EONR modeling. For example, studies using RS-based vegetation indices like NDVI and NDRE allow for near-real-time crop health assessment. However, RS-based approaches alone, while useful for detecting nitrogen status, may not capture spatial variability in nitrogen requirements effectively (Meloni et al., 2024). Combining RS data with ML algorithms has shown promise in refining EONR predictions. Lu et al. (2024) integrated NDVI and NDRE data with environmental factors in an ML model, achieving R^2 values of 0.94-0.96 for critical nitrogen concentration (Nc) prediction in rice. While their approach improved diagnostic accuracy and NUE, it also underscores the high computational demands and complexity of ML models, which may limit their scalability for real-time nitrogen management.

Comparative studies reveal that ML-based EONR models, despite their complexity, may not always offer substantial accuracy improvements over simpler methods. A recent ML-based study achieved an RMSE of 13.7% across multiple fields, which is only slightly better than our RSM-based approach with an RMSE of 14.3%. This marginal improvement raises questions about the practicality of using complex and less interpretable ML algorithms. Additionally, Li et al. (2023) pointed out that ML models require extensive training datasets and can vary significantly in EONR accuracy depending on the algorithm and data inputs. Scaling these models also poses challenges: potentially leading to unreliable site-specific EONR predictions for new fields (Niemeyer et al., 2021).

Our Approach: Integrating RSM with On-Machine N Sensors

In this study, we propose an EONR estimation model that combines Response Surface Modeling (RSM) (Box & Draper, 2007) with multispectral indices (MSIs) and on-machine nitrogen sensors. By using RSM, which efficiently captures interactions between variables in a second-order polynomial, we address the computational limitations associated with complex ML algorithms while maintaining accuracy. Our approach integrates satellite-based MSIs from Sentinel-2 imagery with on-machine sensors, providing real-time, field-specific nitrogen recommendations. This model achieved competitive results, with R^2 values comparable to ML-based studies, and an RMSE of 14.3%, demonstrating effective accuracy without the high computational burden of ML methods.

In conclusion, our RSM-based model stands as a practical and scalable solution for nitrogen management, aligning with sustainable agriculture practices while offering a reliable alternative to ML algorithms. Integrating RSM with on-machine sensors enables a feasible approach for precision nitrogen management, balancing accuracy, simplicity, and field applicability. This approach places our model among the recent advancements in EONR estimation, emphasizing the value of combining simplicity with robust, real-world performance.

1.4. Objectives

- This research centers on developing a comprehensive and scalable framework for nitrogen management in agriculture experimentation that paves the way towards regenerative agriculture. The objectives of this study are:
- Knowledge Domain Mapping in Regenerative Agriculture: To systematically map out the interdisciplinary knowledge domains ranging from geospatial sciences to machine learning that converge in regenerative agriculture, particularly within nitrogen management.
- Optimizing Nitrogen Application Through Advanced Data Integration: To optimize EONR for winter wheat by integrating mobile IoT data, multispectral satellite imagery, and geospatial analysis with statistical methods.
- Data Automation and Cloud Deployment: To develop automated data preprocessing pipelines and cloud deployment solutions, supporting scalable and real-time field management practices.
- Practical Application and Model Generalization: To test the scalability and generalization of the nitrogen optimization model across diverse field conditions, providing actionable insights for large-scale agricultural operations and supporting on-farm experimentation as a practical alternative to traditional plot-based designs.

1.5. Significance and Novelty

The novelty of this dissertation lies in its practical approach to regenerative agriculture, achieved through a diverse integration of knowledge domains including remote sensing, IoT, geospatial sciences, statistical optimization, and cloud storge and computing technologies. By bridging data-driven insights with real-world applications, this research advances KDM's role in agriculture, paving the way for effective resource allocation and ecosystem-friendly farming. This framework offers a viable model for farmers, researchers, and policymakers, highlighting the transformative potential of technology towards regenerative agriculture. The findings are expected to contribute to the ongoing discourse on sustainable agriculture by demonstrating the role of interdisciplinary knowledge and advanced data technologies in promoting responsible, high-yield farming practices.

This dissertation provides a robust, scalable framework for the sustainable application of nitrogen, reinforcing the concept that each technological advancement—when strategically integrated contributes to the broader goal of regenerative, environmentally sustainable agriculture. Through this research, the agricultural community gains a comprehensive, technology-enabled approach for nutrient management, offering a practical pathway to regenerative practices that meet today's food production needs without compromising future environmental health.

1.6. Structure

Chapter two introduces a second-order Response Surface Model (RSM) to estimate EONR by integrating on-machine nitrogen sensors and multispectral indices (MSIs) from Sentinel-2 satellite imagery. Using 21 MSIs from three different acquisition dates, the model examined the interaction between MSIs and as applied nitrogen rate recommendations with N-sensor. Through feature selection methods like ANOVA, it identified key indices that effectively estimate yield response to nitrogen at the field level, addressing spatial variability with higher precision. By employing whole-field trials, this approach avoids the limitations of small-plot designs, ensuring that nitrogen recommendations adapt to specific crop needs across different sub-field locations. The findings highlight that field heterogeneity can influence EONR, with optimal nitrogen levels varying spatially within the field. Practical applications of this approach include real-time nitrogen adjustments and prescription mapping, which can be transmitted to machines for precise field applications. While the model's flexibility and simplicity support real-time decisions, the research also identifies areas for further improvement, such as data preprocessing, scalability across environmental conditions, and inclusion of protein content in yield modeling for a more comprehensive assessment of EONR. Future studies are encouraged to validate the approach across diverse crops and regions, considering seasonal and environmental variations to further refine the model's applicability in PA.

Chapter three Illustrates the scalability of the developed EONR estimation framework across divers fields. This study builds on previous research using response surface modeling (RSM) to estimate EONR and examines the framework's adaptability across five diverse winter wheat fields. Two modeling approaches were tested: a pooled model, where four fields were used to train a universal RSM, and an average model, which created individual RSMs for each field and combined them via weighted averaging. Results show that the average model achieved better accuracy with an RMSE of 11 kg/ha, compared to 19 kg/ha in the pooled model, and maintained EONR predictions within a consistent range of 59.5 to 85.7 kg/ha. This method demonstrated better generalization and adaptability, especially under varying field conditions, confirming the model's potential for scaling up nitrogen management across different agricultural settings in PA. The average model provides a more robust and reliable estimate for EONR, minimizing risks of under-fertilization and aligning closely with typical nitrogen application practices reported by farmers. Through K-fold cross-validation, the study showed that the average model mitigates the impact of uncertainties, such as weather and market conditions, which are critical for precise nitrogen recommendations. Future research is encouraged to integrate real-time IoT and satellite data, improving model responsiveness and environmental sustainability. These findings underscore the framework's potential for enhancing nitrogen application strategies in PA, offering a scalable, data-driven tool that supports sustainable crop production across diverse agricultural landscapes.

Chapter four introduces a scalable cloud-based framework designed to optimize nitrogen application by estimating the EONR for PA. The system processes vast amounts of agronomic telemetry data from farm machinery and satellite multispectral imagery, combining these inputs to deliver real-time nitrogen prescription maps tailored to diverse field conditions. Using AWS services such as Batch for job orchestration, Amazon S3 for scalable data storage, and RDS Postgres for structured data management, the architecture dynamically harmonizes spatially heterogeneous data, applying techniques like Kriging interpolation for accurate nitrogen recommendations. The framework efficiently integrates both real-time and historical data streams, enabling farmers to make data-driven decisions that boost crop productivity while reducing environmental impact. The system’s flexibility and scalability are validated through a case study, where it successfully manages large-scale regional data to generate actionable insights for nitrogen management across various farming environments. With automated data archiving, quality control, and regional data management, this cloud-native solution adapts to growing data volumes while reducing costs. It not only enhances real-time decision-making but also sets a foundation for future advancements by supporting additional data sources such as soil sensors and weather data. This framework represents a robust solution for sustainable PA, enabling farmers to maximize nitrogen use efficiency and enhance resource management in modern farming.

Chapter five underscores the role of this research in advancing regenerative agriculture through a comprehensive framework that integrates remote sensing, machine learning, and cloud computing to enhance nitrogen management, using an interdisciplinary approach facilitated by Knowledge Domain Mapping (KDM). Each chapter contributes to the realization of regenerative agricultural goals by aligning nutrient application with crop demands, thereby reducing environmental impact and promoting soil health. The KDM approach maps and integrates diverse knowledge fields—spanning geospatial sciences, agronomic telemetry, and data analytics—ensuring that each domain complements the others to support precise, adaptable, and environmentally conscious farming practices. This interdisciplinary approach not only enhances EONR estimation across varied field conditions but also aligns with regenerative principles by improving resource use efficiency, supporting soil regeneration, and minimizing chemical dependency. Through this framework, the dissertation offers a pathway toward scalable, data-driven solutions that make regenerative agriculture feasible in modern, large-scale farming systems.

Chapter 2: In Season Estimation of Economic Optimum Nitrogen Rate with Remote Sensing and Historical Telematics Field-Operation Data

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 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 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.

Keywords: Economic optimum nitrogen rate, multispectral indices, optimization, response surface model, winter wheat

2.1. 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. The availability of soil and climate data is limited for every farm. RS data enable a direct measurement of crop status and compensates 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) employed spectral indices to measure plant nitrogen content (PNC) as a vital indicator in assessing N status in the potato crop; however, their approach was independent of year, cultivar, and growing period. Another study focuses on multispectral features along with machine learning (ML) to predict total N content, with a coefficient of determination ranging from 0.37-0.70 (Li et al., 2023). While these methods are promising, ML-based approaches, such as those employed by de Lara et al. (2023) in modeling Economic Optimum Nitrogen Rate (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 (PA) 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 schema compared to traditional quadratic or plateau functions used in many commercial systems. Notably, this approach moves away from small-plot designs typically used in PA 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.

2.2. Material and Methods

2.2.1. 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.

2.2.2. 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 (Figure 2.1) 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 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, there is substantial variation in the actual amounts of N applied, providing sufficient heterogeneity to be able to fit the RSM. In other words, since 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 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 analysis. 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.

2.2.3. 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 2.1 for details about trial data.

Table 2.1. Data sources used in this study.

Operation	Product	Range	Data	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	-
Sentinel-2 imagery	Sentinel-2 L1C BAND	-	2020.03.25	BBCH Stage 30-31: Beginning of stem elongation.
			2020.03.30	BBCH Stage 31-32: Stem elongation continues.
			2020.04.04	BBCH Stage 32-33: Further stem elongation, second node detectable

The crop growth stages are also provided in table 2.1 as BBCH for better understanding the relation of image acquisition date and crop fertility stage.

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 (Figure 2.1). 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 (Figure 2.1). The harvest observations were documented in the Operation Center.

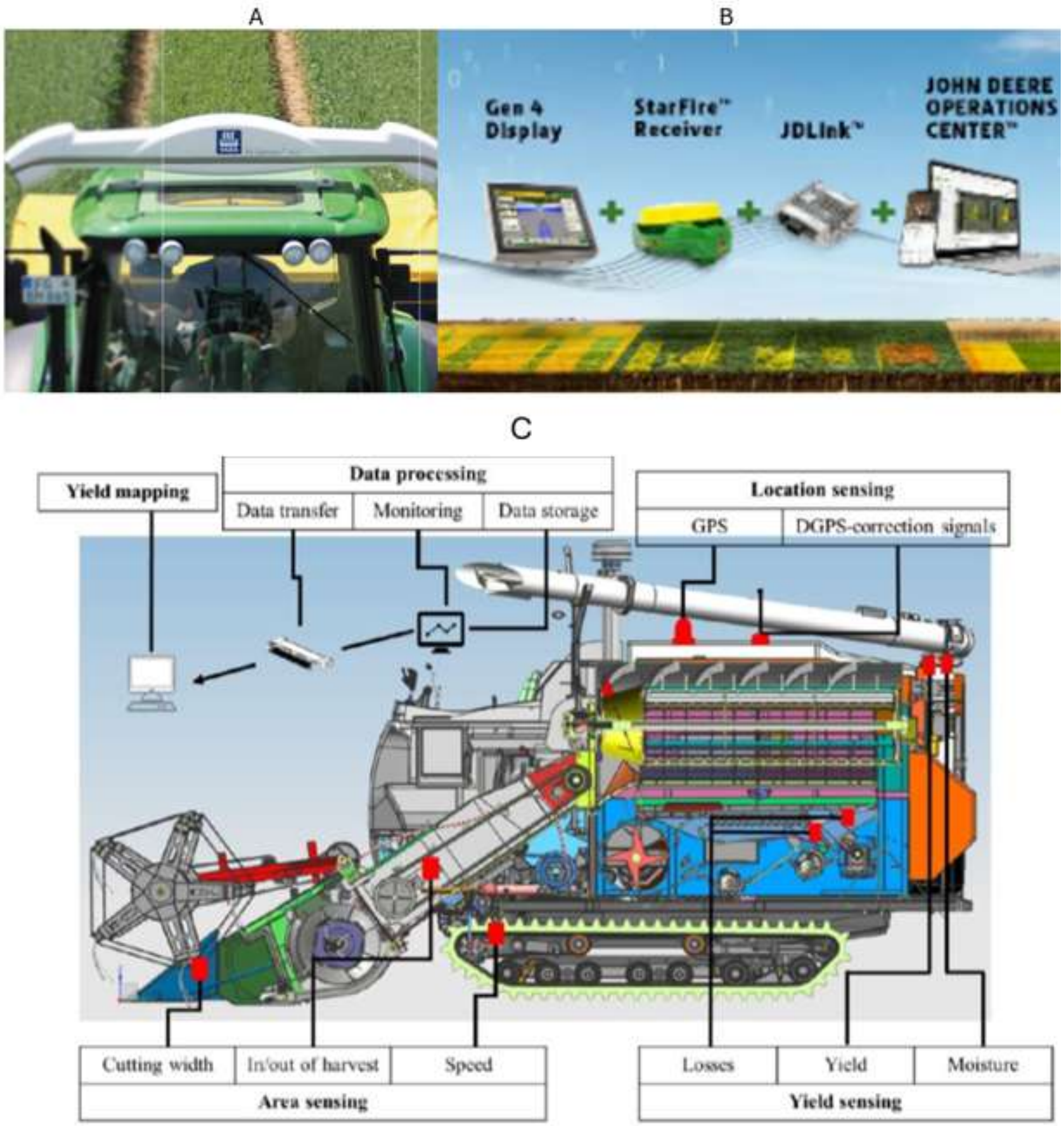

Figure 2.1. Technology used for On-Farm Experimentation.

On the left, image A illustrates the N-sensor technology (Weyand, 2022) and on the top right image B, the PA equipment demonstrated that is used for data collection and documentation. On the bottom right image C, a detailed picture (Leroux, 2020) of harvester yield monitoring technology is demonstrated. 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 2.1) 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 determined through crop sensing technology, which was mounted on the machine for real-time determination of N needed for the crop (see Figure 2.1). To achieve the objective of this study one of the trial fields was utilized.

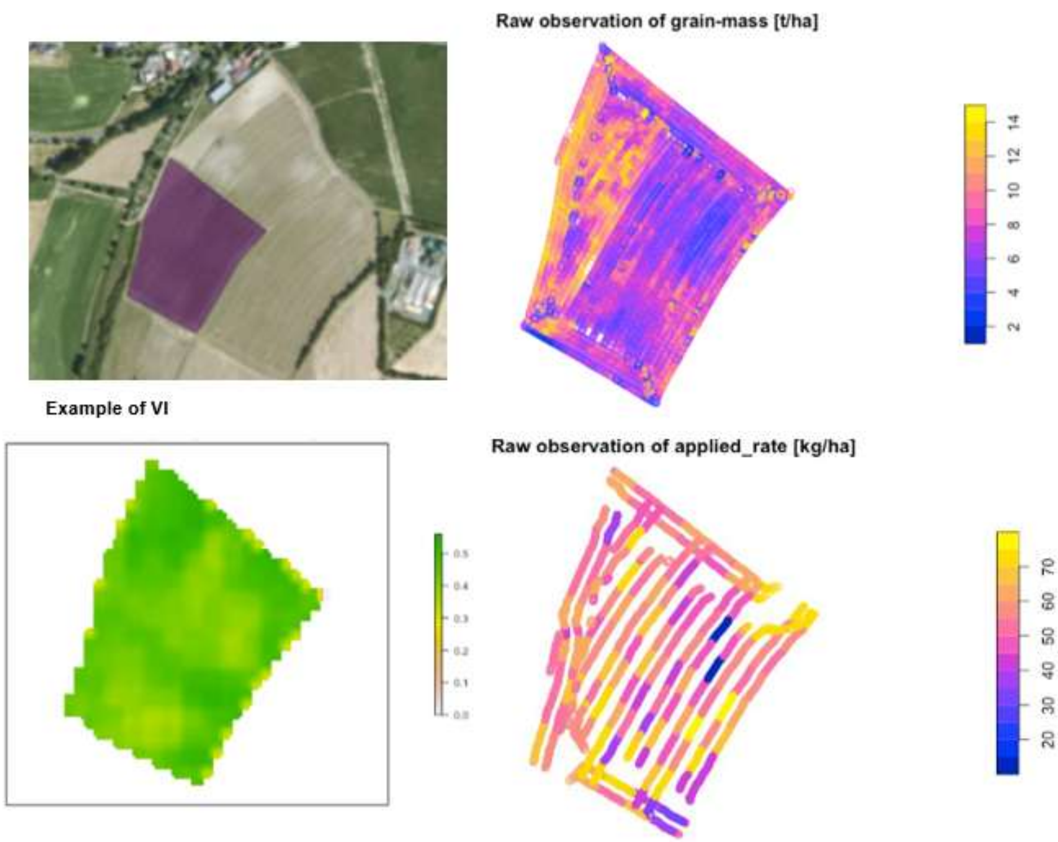

Figure 2.2. Illustrates the target field for this study.

The row observations were documented via PA equipment, and the VI is a vegetation index Normalized Difference Red Edge Index (NDRE) derived from sentinel-2 multispectral imagery. This figure represents the spatial representation of the input data for this study. As can be seen, data are in different formats. The field-operation data (Applied-rate kg/ha is the applied N, and the grain mass t/ha is harvesting yield observation) are presented in a discrete manner while the VI example is presented in Raster form. Hence, appropriate data processing and preparation is required (see Figure 2.3) to enable map overlay analysis for this study.

2.2.4. 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 Figure 2.3).

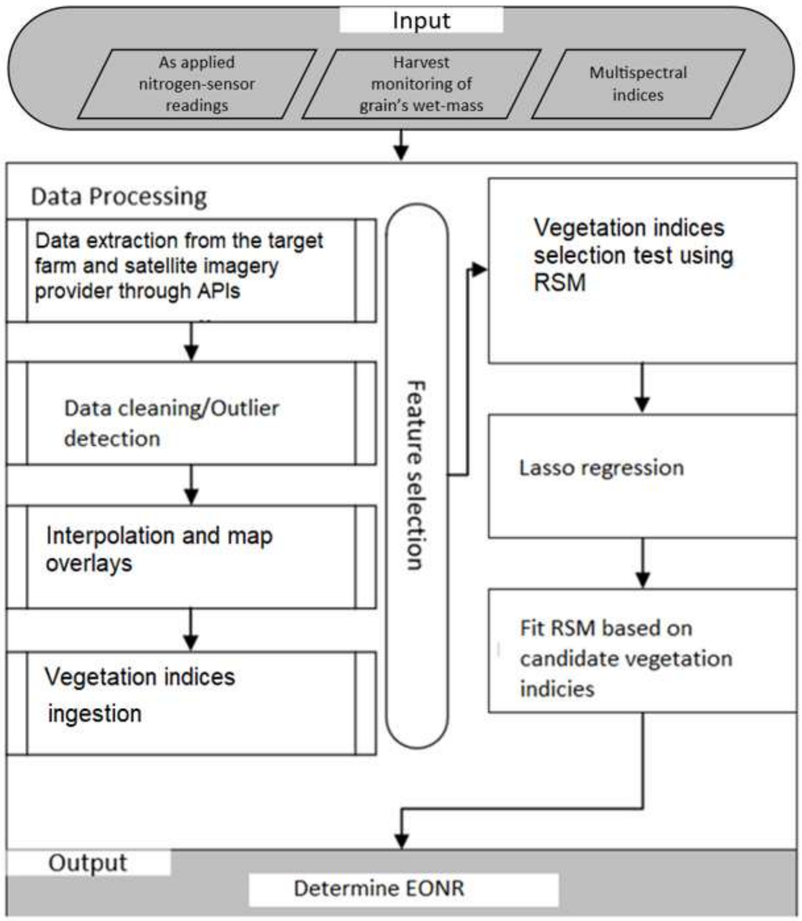

Figure 2.3. Pipeline for data preprocessing and integration for final analysis.

Figure 2.3 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).

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, nonfully 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 interquartile 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 Figure 2.4.

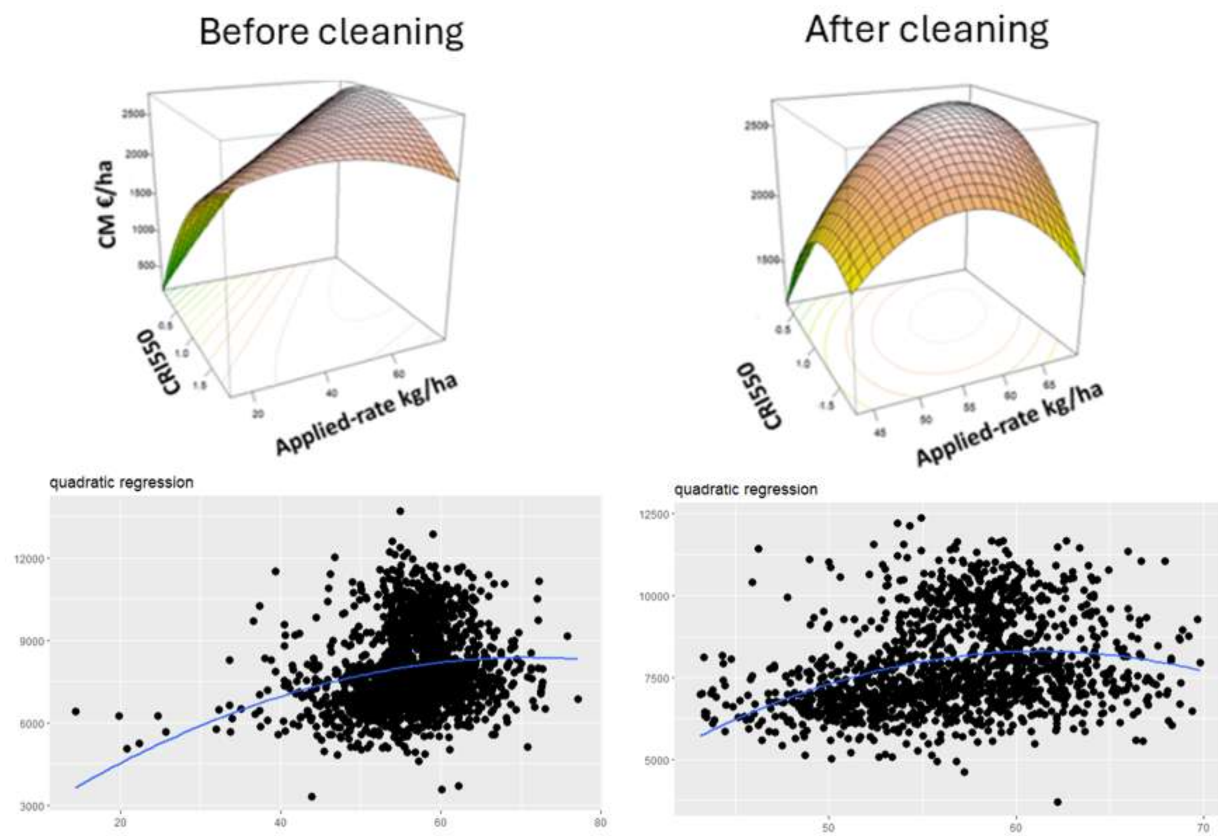

Figure 2.4. Illustrates the importance of outlier removal.

This Figure represents the response surface model before and after removing outliers on the top. The shape of response surface becomes more reliable, exposing a maximum global yield. On the bottom, the quadratic regression fitted for yield and applied N depicts the effect of outliers before and after cleaning.

After removing outliers in the row observation from telematics machinery (spreader and harvest observation), the data were interpolated to 5m 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 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 observations 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.

2.3. 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 amount of N needed for the plant. Hence it assumes a significant interaction between applied rate with 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 is ready to be analyzed, each individual cell of the grid map represents an area of 25 m^2, which constitutes a single observation (i.e., for a 7-ha field, there are approximately 1500 observations). 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 parameter β_{NN} is expected to be < 0, implying a diminishing return repose curve at each location. The response surface is computed with the 'rsm' package in the R language. This package provides several functions to facilitate classical RSM (Lenth, 2009).

2.4. 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 2.2) 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, 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.

Table 2.2. MSIs derived from Sentinel-2.

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.		

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 make 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 $x_1, x_2, x_3, \dots, x_n$, where $x_i \in \mathbb{R}^m$ represents a single spectral index (i.e., Chlorophyll Absorption Ratio Index) and m is the number of observations for a given date before actual N. For each variable X we fit the proposed RSM (see equation 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 MSI is eliminated from further processing. A significance level α of 0.01 for coefficient β_{NX} 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 2.3) 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, Lasso was applied to each class at specific dates, utilizing L1-norm regularization to identify the most relevant features within each category.

Table 2.3. Identified MSIs categories.

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

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

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, as well as a quadratic term for N, were generated and fed into the Lasso model. The results are presented in the "result" 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 equation (1) was expanded to all the combinations 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.

2.5. 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

$$TR = P_W \times \ kg\ per\ ha = 2000\ €/ha$$

The cost of input N was calculated by multiplying 0.417€/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

$$TC = P_N \times \ kg\ N\ applied\ per\ ha = 25.08\ €/ha$$

Consequently, the return is given by

$$Return = P_W \times Y(n) - P_N \times n \qquad (3)$$

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., 2018). 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

$$Y(n) = c_0 + c_1 n + c_2 n^2 \qquad (4)$$

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 equation (4) is defined as

$$\frac{\partial Y}{\partial n} = Y'(n) = c_1 + 2c_2 n \qquad (5)$$

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 equation (3) with equation (4), and taking the derivative of the return with respect to n, we find

$$\frac{\partial Return}{\partial n} = P_W(c_1 + 2c_2 n) - P_N \qquad (6)$$

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 ca be defined by

$$(c_1 + 2c_2 n) = \frac{P_N}{P_W} \qquad (7)$$

From this, the solution for EONR can be obtained by solving for *n* in equation (7), yielding

$$\text{EONR} = \left(\frac{P_N}{P_W}\right)\left(\frac{1}{2c_2}\right) - \left(\frac{c_1}{2c_2}\right) \qquad (8)$$

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

2.6. Results

2.6.1. Feature selection

Figures 5, 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., 2020), 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 (Figure 2.5). However, as time passed with every five days of Sentinel revisit, NDVI was no longer significant (not in the plot of coefficients in Figure 2.6 for dates 2020.03.30 and Figure 2.7 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.

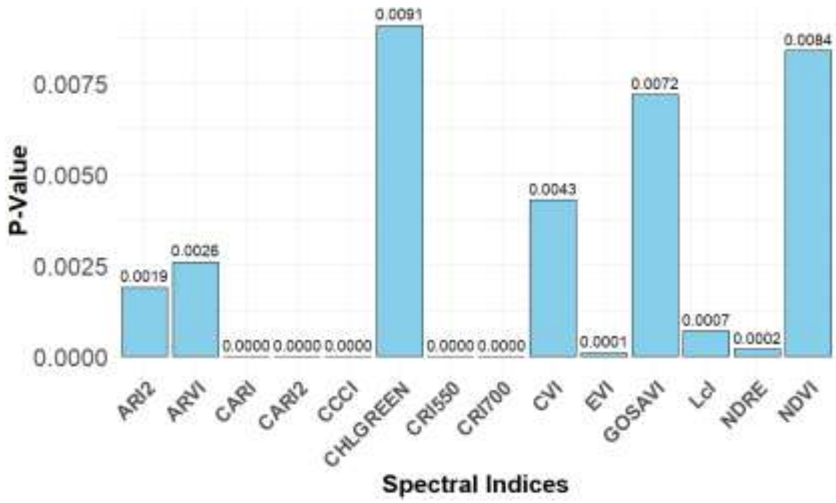

Figure 2.5. Significant MSIs for the first data of image acquisition.

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.

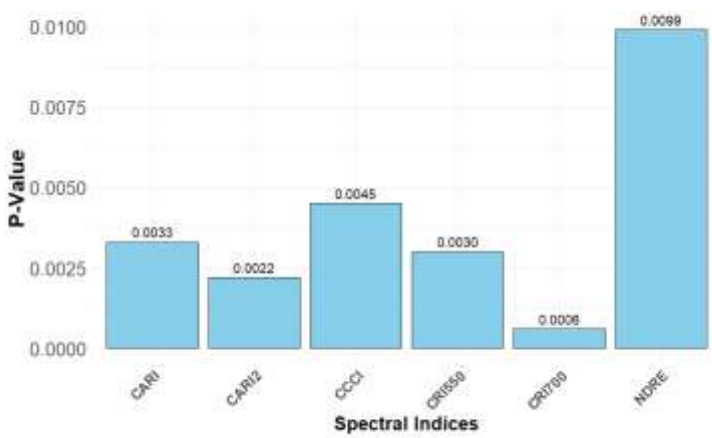

Figure 2.6. Significant MSIs for the second data of image acquisition

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.

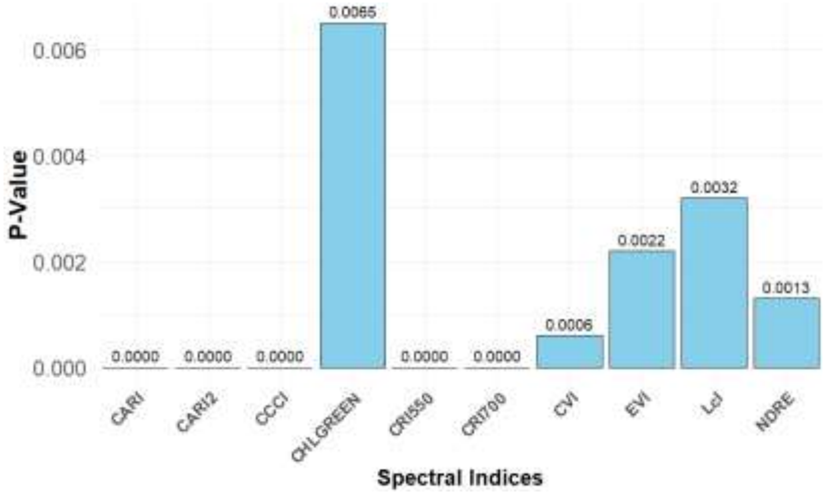

Figure 2.7. Significant MSIs for the third data of image acquisition

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.

2.6.2. 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.

Table 2. 4. Example of linear, quadratic, and interaction terms of MSIs MSIs from Chlorophyll category for date "2020.03.30" introduced in Lasso regression.

Linear term	Quadratic	Interaction
$CARI$	CARI^2	$CARI{:}CCCI$, $CARI{:}CARI2$
$CCCI$	$CCCI^2$	$CCCI{:}CARI$, $CCCI{:}CARI2$
$CARI2$	$\mathrm{CARI2}^2$	$CARI2{:}CARI$, $CARI2{:}CCCI$
$Nitrogen$	$Nitrogen^2$	$CCCI{:}Nitrogen$, $CARI{:}Nitrogen$, $CARI2{:}Nitrogen$

Hence in the above example in Table 2.4 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 Figure 2.8). This indicates that CCCI is no longer a candidate for the third phase of feature selection.

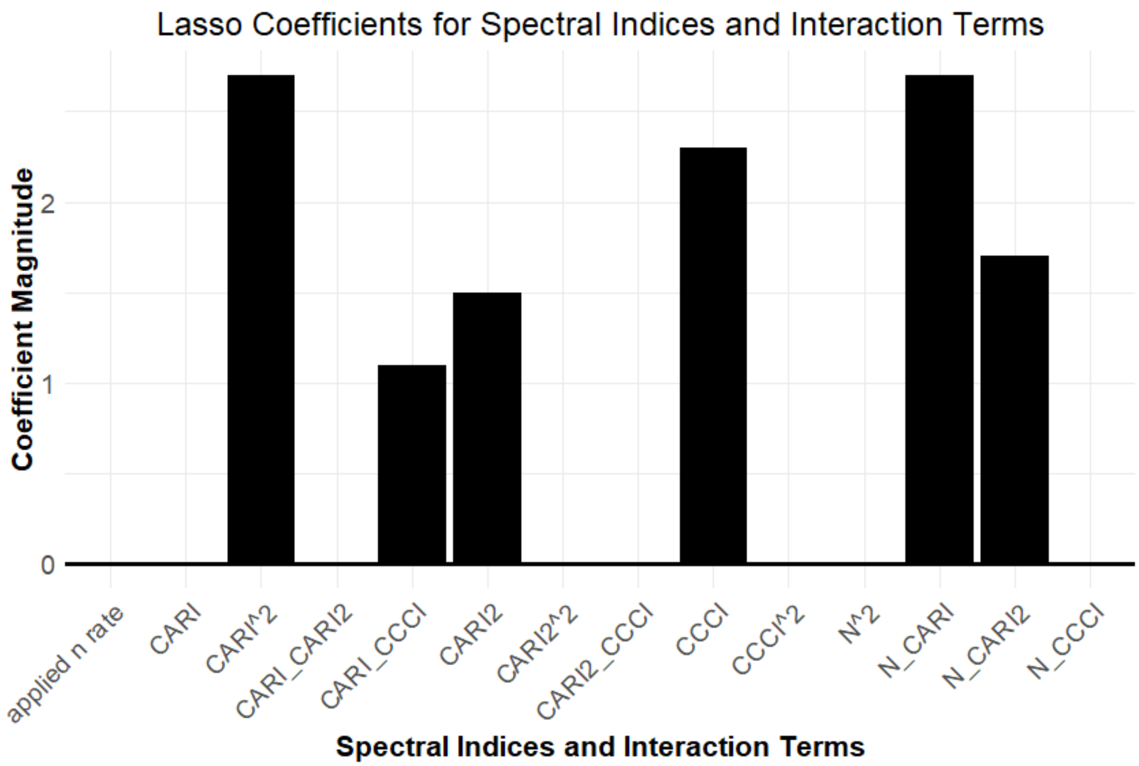

Figure 2.8. Lasso coefficients for MSIs from chlorophyll category in the second date of acquisition.

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.

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 2.5.

Table 2.5. Final candidate MSIs for the third phase of feature selection by RSM model reduction.

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

After applying Lasso regression to all categories and dates, the above MSIs are identified and selected from the following categories presented in this table.

2.6.3. RSM model output

Given the selected candidate MSIs from Lasso regression in table 2.4 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 leads 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:

$$y = \alpha + \beta_N N + \beta_{NN} N^2 + \beta_{CARI2_{2020.03.25}} CARI2_{2020.03.25} + \beta_{CARI2_{2020.03.25}} {CARI2_{2020.03.25}}^2 + \beta_{ARI2_{2020.03.25}} ARI2_{2020.03.25} + \beta_{ARI2_{2020.03.25}} {ARI2_{2020.03.25}}^2 + \beta_{\mathrm{CARI}_{2020.03.30}} CARI_{2020.03.30} + \beta_{CARI_{2020.03.30}} {CARI_{2020.03.30}}^2 + \beta_{\mathrm{CRI550}_{2020.03.30}} \mathrm{CRI550}_{2020.03.30} + \beta_{\mathrm{CRI550}_{2020.03.30}} {\mathrm{CRI550}_{2020.03.30}}^2 + \beta_{CARI2_{2020.04.04}} CARI2_{2020.04.04} + \beta_{CARI2_{2020.04.04}} {CARI2_{2020.04.04}}^2 + \beta_{CRI700_{2020.04.04}} CRI700_{2020.04.04} + \beta_{CRI700_{2020.04.04}} {CRI700_{2020.04.04}}^2 + \beta_{N.CARI2_{2020.03.25}} N * CARI2_{2020.03.25} + \beta_{N.ARI2_{2020.03.25}} N * ARI2_{2020.03.25} + \beta_{N.\mathrm{CARI}_{2020.03.30}} N * \mathrm{CARI}_{2020.03.30} + \beta_{N.\mathrm{CRI550}_{2020.03.30}} N * \mathrm{CRI550}_{2020.03.30} + \beta_{N.CARI2_{2020.04.04}} N * CARI2_{2020.04.04} + \beta_{N.CRI700_{2020.04.04}} N * CRI700_{2020.04.04} \quad (2)$$

2.6.4. EONR estimation validation

Plot a bundle of response curves: Plotting the proposed model $Y(n)$ 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 c_0, c_1, and c_2 were computed through plugging the MSIs for each corresponding geographical location in the field (see Table 2.6).

Table 2.6. A bundle of response curves.

$CARI2$ $2020.03.25$	$ARI2$ $2020.03.25$	$CARI$ $2020.03.30$	$CRI550$ $2020.03.30$	$CARI2$ $2020.04.04$	$CRI700$ $2020.04.04$	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**

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.

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 Figure 2.9. A bundle of curves for several points in the field is also plotted (see Figure 2.10) to reflect the spatial heterogeneity of EONR across the field.

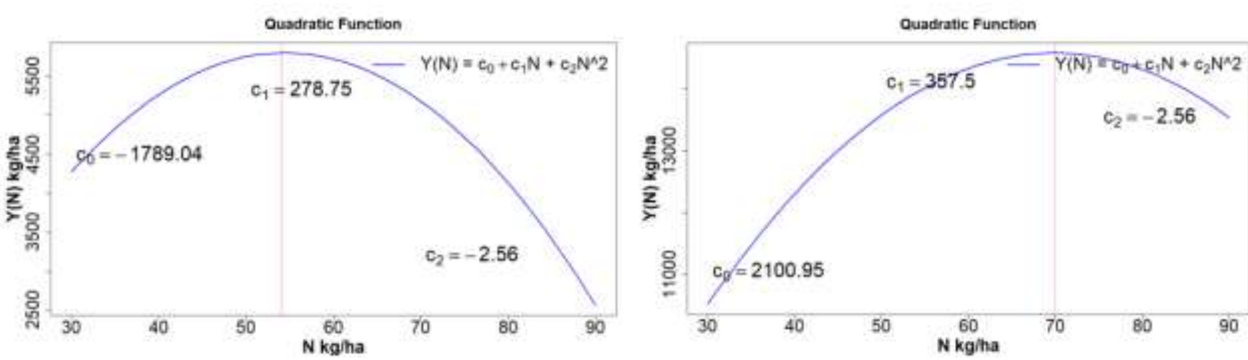

Figure 2.9. The plot of a bundle response curves.

Two different geographical position which have two different intercepts c_0, and values of coefficient c_1. As can be seen the maximum varies at each geographical position in the field. For the left-side max is at ~ 54kg and for the right-side max is at ~ 70kg. It is noteworthy that the average yield differs for each location, indicating that one area in the field can have an average yield of ~ 4000kg while in the other area the average yield is ~ 6500kg as demonstrated from left to the right plot, respectively. The spatial distribution of EONR is depicted as a prescription map (see Figure 2.10) and can be sent over the network connection to the machine for execution.

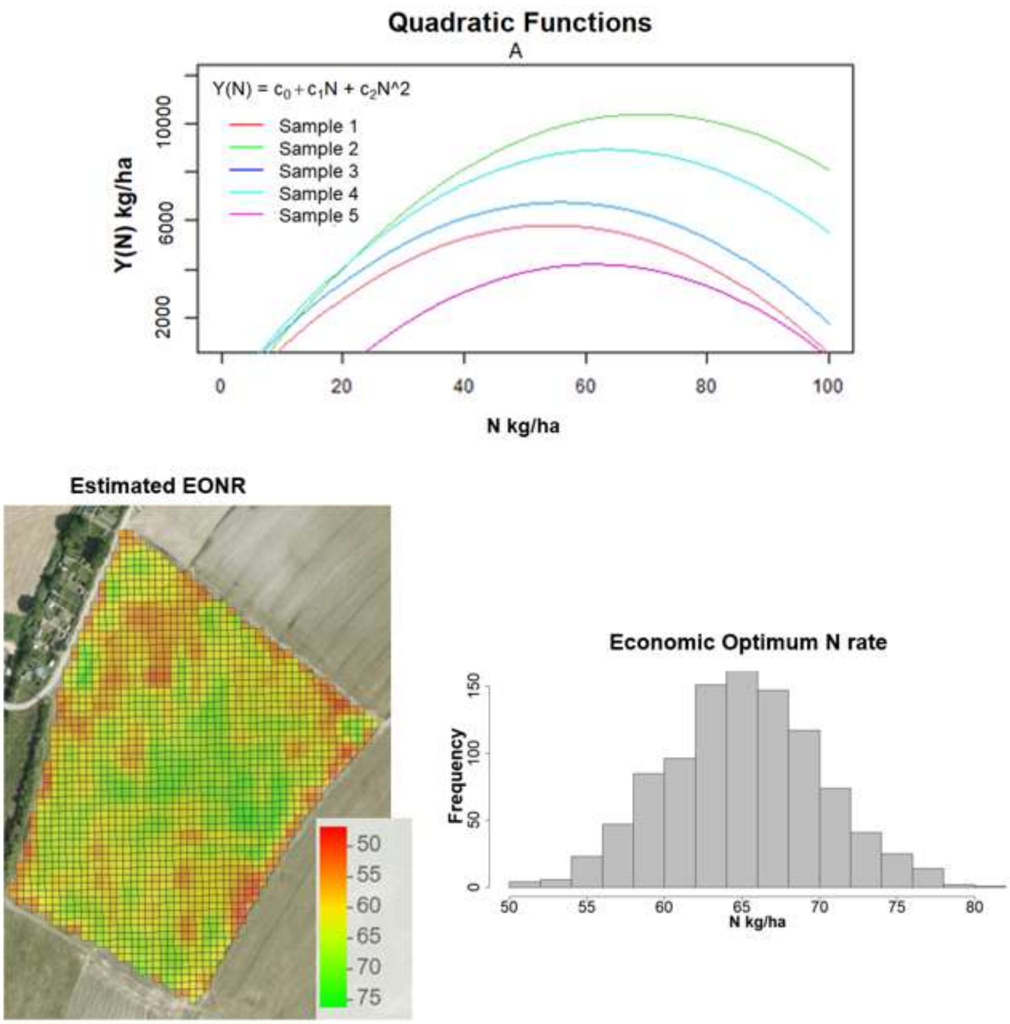

Figure 2.10. Spatial variability of EONR across different parts of the field.

In Figure 2.11, 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.

2.7. 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 significant 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 factors 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 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 (Figure 1.11). 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.

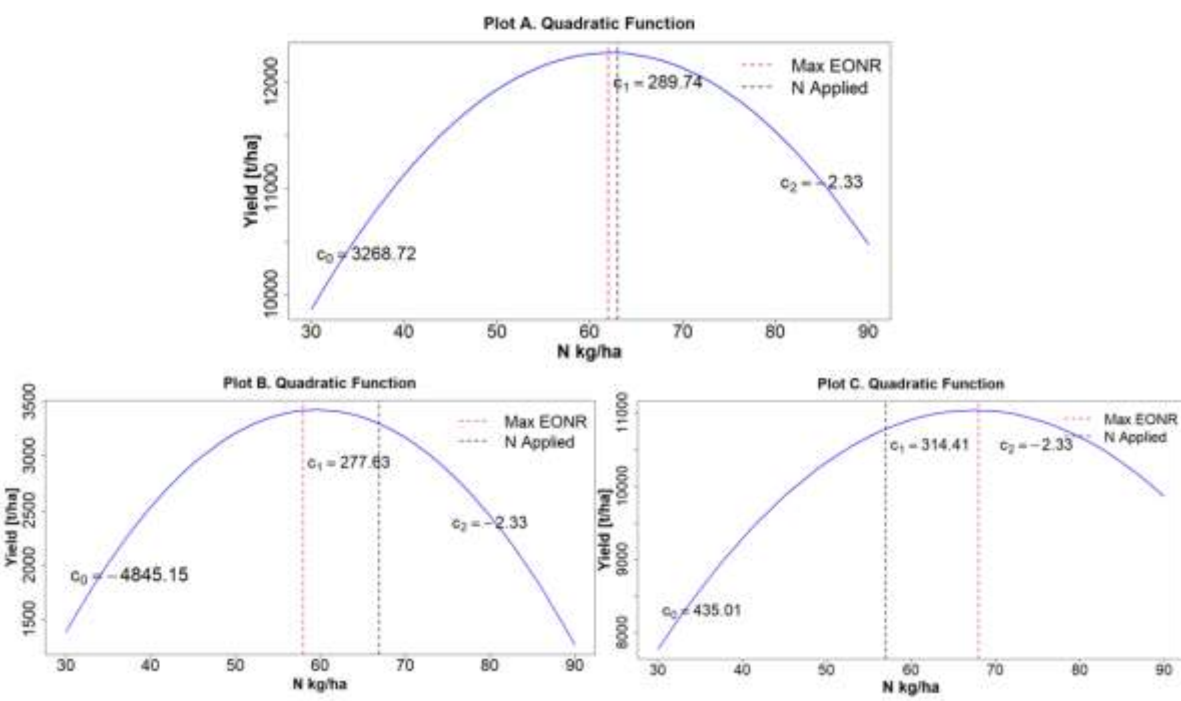

Figure 2.11. Illustration of three scenarios of N management outcomes.

In Figure 2.12, 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.

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 the machine. The computational time for RSM methodology enables a viable real-time application. Highlighting the limiation 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 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.

2.8. 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.

Declarations

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

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.

Acknowledgement

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.

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.

Chapter 3: Model Generalization for in-Season Estimation of Economic Optimum Nitrogen Rate Across Diverse Winter Wheat Fields.

Abstract

The accurate estimation and spatial allocation of economic optimum nitrogen (EONR) rates are crucial for sustainable crop production, reducing chemical inputs while maintaining optimal yield and profit. Building on our previous work, which demonstrated the efficacy of using response surface modeling (RSM) in conjunction with historical and real-time mobile Internet of Things (IoT) data to determine EONR. This follow-up study aims to evaluate the generalization capabilities of the proposed framework across diverse agricultural datasets in five different fields. Leveraging additional data sources, we tested the robustness and scalability of an RSM strategy to support decision-making in a variety of field conditions. Two modeling approaches were developed to find the best generalization approach for EONR estimation. The first model is a “pooled model” where four experimental fields are used to fit a single RSM, then cross-validated with the test field. In contrast, the second approach employed a weighted averaging methodology named the “average model”. Instead of fitting a pooled model across all four fields, each field received its own RSM fit, and a weighted average framework was introduced to create a representative model that integrates the characteristics of all individual fields into a unified predictive model. A comparison of a pooled model and an average model is made to showcase the performance of different modeling approaches in estimating EONR. The results indicate that the framework with model weighted averaging maintains high prediction accuracy and reliability, with an average root mean square error (RMSE) of 11kg N/ha for the EONR and a range of the predicted EONR from 59.5 to 85.7 kg/ha, while the pooled model RMSE was 19kg N/ha and a range of predicted EONR from -0.52 to 91.4 kg/ha. The average model demonstrates the framework's adaptability to different field conditions. This follow-up research confirms the framework's potential for widespread adoption in precision agriculture, offering a scalable and robust solution for optimizing nitrogen application across diverse agricultural settings. The insights gained from this study contribute to the development of more sustainable and efficient farming practices, leveraging advanced geospatial data, statistical, and machine learning techniques.

Keywords: Precision Agriculture, Economic Optimum Nitrogen Rate (EONR), Average Model, model scalability, Response Surface Modeling (RSM)

3.1. Introduction

Effective nitrogen (N) management is critical in agricultural practices, particularly for winter wheat production, where nitrogen is often the most limiting nutrient. An efficient and reliable optimization model that can capture the complexity of historical and real-time mobile Internet of Things (IoT) data, such as response surface model (RSM) (Abdipourchenarestansofla & Piepho, 2024), is essential to support making informative nutrient management decisions. This framework aids in optimizing nitrogen applications based on the economic optimum nitrogen rate (EONR). The ability to generalize and scale recommendations across different field conditions was tested in this paper. By effectively monitoring crop health and nitrogen content, farmers can make informed decisions that lead to efficient fertilization practices, ultimately contributing to regenerative agriculture (Sharma et al., 2024, Smerald et al., 2023). Implementing an accurate N estimation model empowers farmers to balance crop needs with environmental stewardship, facilitating a shift towards more responsible and efficient nitrogen management practices. Mobile IoT Smart farming technologies and multispectral imagery provide a comprehensive resource for achieving sustainability in crop production systems (Rane, 2023). Another important component to achieving sustainability is the identification of a model that minimizes error and avoids biased fertilizer rates when it comes to scalability. Modeling can be performed using machine learning and statistical approaches. In practical applications, the utilization of data assimilation systems remains limited due to their rigidity and complexity. As a result, many systems depend on simplistic relationships between indices and the target variable, such as nitrogen (Clevers & Gitelson, 2013; Diacono et al., 2013; Jin et al., 2018). This study was based on a framework proposed by Abdipourchenarestansofla and Piepho (2024) to evaluate the scalability of the framework under various field conditions. The proposed framework facilitates the recommendation of variable rate nitrogen application by utilizing historical nitrogen application maps, harvest yield monitoring data, and spectral indices derived from satellite data. The framework was tested on a single field, and the findings demonstrated its effectiveness in determining the EONR at each point in the field. This follow-up study aims to address the challenges of generalizing and scaling the estimation of EONR across multiple fields with heterogeneous characteristics. While our original framework demonstrated promising results for a single field, the ability to apply this framework to diverse agricultural settings remains untested. To address this, we will employ an average model through leave-one-field-out cross-validation strategies to fit a model that can scale up to more diverse environmental conditions. This approach is validated using unseen data to ensure robustness and reliability in various field conditions. Our primary objective is to demonstrate that the framework developed in our original paper can be effectively generalized and scaled, providing a robust solution for optimizing nitrogen application across different agricultural environments.

3.2. Material and Method

3.2.1. Experimental Site

The experimental farm is situated in southwest Germany, a region known for its diverse environmental factors affecting agriculture and land use. The area typically receives 700-900 millimeters of annual precipitation, with variations across different parts. These precipitation patterns are crucial for agricultural practices, water management, and vegetation growth. The region's soil types are predominantly fertile loamy soils with good drainage properties. The climate is temperate oceanic, characterized by mild temperatures and relatively high humidity levels.

3.2.2. Experimental Design

Phenological changes significantly affect a crop's growth cycle, influencing the assimilation of light, temperature, water, and nutrients. This study focuses on the controllable factor of nutrition application, specifically mineral nitrogen (N). Unlike traditional N-treatment experiments that use strip, plot, or block designs, this study employs an on-machine sensor called the N-sensor provided by Yara International ASA, Oslo, Norway, to apply N in real-time across the whole field. The N-sensor assesses crop characteristics across the entire field, allowing the operator to define minimum and maximum input nitrogen levels. Higher sensor values result in less fertilizer application. This management zone design enables the selection of N levels for each field location, creating substantial variation in the actual amounts of N applied and providing sufficient heterogeneity for fitting the RSM. The RSM then uses the applied N and spectral indices to model yield, the response variable. Before any N application, the Nmin values are calculated from the soil sampling to demonstrate the variability of available N across the fields. See Figure 3.1.

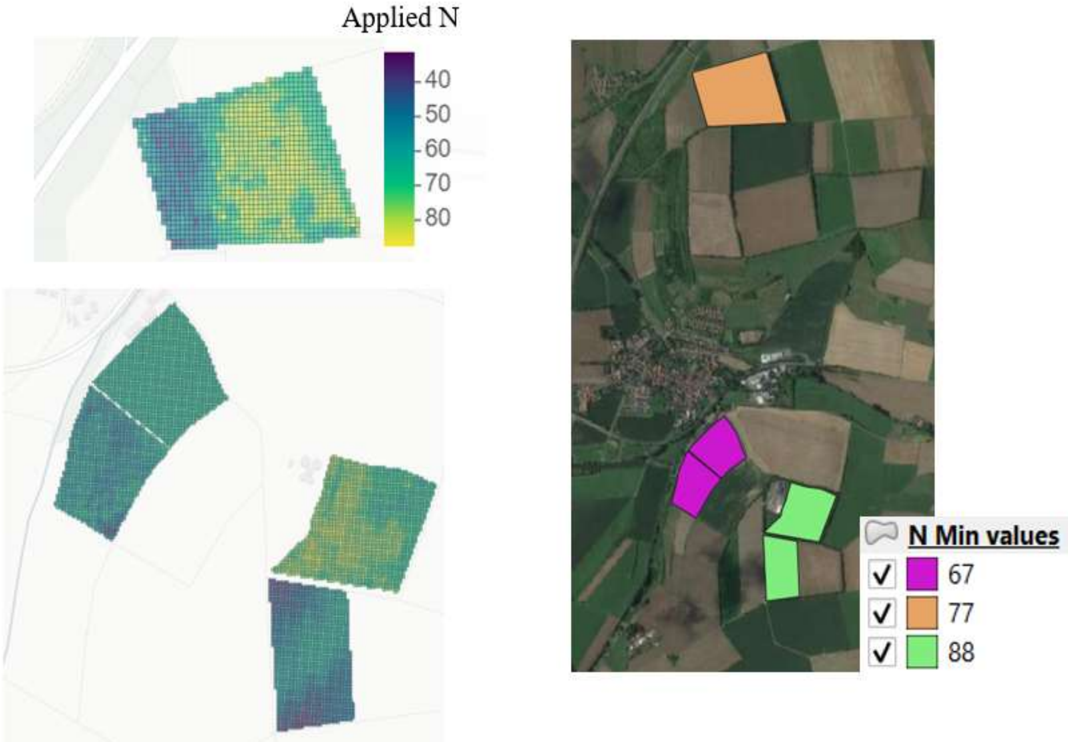

Figure 3.1. Experimental units on the target farm with Winter Wheat crop

In Figure 3.1. Each polygon represents a field boundary which is considered as an experimental unit. The Nmin values represent the amount of nitrogen in the soil for each field.

The Nmin values were subtracted from the final N application map to avoid unintended overfertilization. In other words, if the N application rate for the map (EONR) is 50 kg/ha on average, the final application rate would be $50 - Nmin$ for the respective field since there exists N in the soil and hence the determined value of EONR can be reduced by the amount of Nmin in the soil. Mineral N fertilization is typically split into three or four applications during the season. For simplicity, the analysis focuses on the most common N application in winter wheat, ammonium sulfate saltpeter. The fertilizer product type does not limit the analysis. The first N application was applied late in the season. The experimental design facilitates exploring the selection and applicability of specific spectral indices for N determination. This is done by examining interactions obtained from cross-products involving input N and selected spectral indices measured before the actual N application. The acquisition time of satellite imagery can influence these interactions. This study uses five fields to assess the scalability of the proposed framework through a leave-one-out cross-validation approach and average model strategy. Each of the five fields, with varying environmental conditions, is treated as an individual experiment. The fields include comprehensive variables such as harvest yield observations, as-applied N sensor readings, and

21 MSIs measured before the N application operation. A second-order RSM is developed for the entire field, necessitating the integration of diverse and heterogeneous map layers and appropriate data preprocessing steps. By applying this approach, we aim to demonstrate the framework's robustness and scalability in optimizing nitrogen application across diverse agricultural settings.

3.2.3. Model Training and Evaluation

Two modeling strategies were developed to find the best generalization approach for EONR estimation. The first model is a “pooled model”. A pooled model is a statistical approach that combines data from multiple groups or units into a single dataset to estimate a common set of parameters, assuming that the relationships between the variables are consistent across all units. For the pooled model four experimental fields were used to fit a single RSM, which is then tested on the test set. In contrast, the second approach employed a weighted averaging methodology denoted here as the “average model”. Instead of fitting a pooled model across all four fields, each field received its own RSM fit, and a weighted average strategy was introduced to create a representative model that integrates the characteristics of all individual fields into a unified predictive model (refer to section “Model Averaging and Generalization”). To assess the performance of the modeling strategies a leave-one-out cross-validation strategy was employed. This involved partitioning the five fields into training and test set. Training the model on four subsets and testing it on the test set was done on each possible partition in turn. Integrating cross-validation into the modeling process ensured that both modeling strategies generalize effectively to unseen data. By iterative training and testing the model across multiple folds, the Root Mean Square Error (RMSE) was calculated to evaluate predictive accuracy. This comprehensive approach enhanced the stability and reliability of the model's performance, minimizing overfitting while ensuring generalizability across different datasets. The RMSE computation as

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(\widehat{EONR}_{model\ averging} - \widehat{EONR}_{test\_set}\right)^2}{n}} \qquad (1)$$

where $\widehat{EONR}_{model\ averging}$ is the estimated EONR from the average model and $\widehat{EONR}_{test\_set}$ is the estimated EONR from the test set. The goal is to be to minimize the error to $\widehat{EONR}_{test\_set}$ when a generalization approach for modeling is introduced. In this case $\widehat{EONR}_{model\ averging}$ and $\widehat{EONR}_{singel\ model}$ were used to cross-validate with the $\widehat{EONR}_{test\_set}$. The one that has the minimum error to $\widehat{EONR}_{test\_set}$ is represented as the candidate modeling strategy for the generalization of the EONR framework.

The average RMSE across 5 folds was computed for the "pooled model" and "average model". The lower the average of RMSE, the closer to the estimated EONR from the test field. This indicates a difference between the EONR estimates from the two models.

3.2.4. Data Description

This study utilizes three primary datasets: harvester yield data, sensor readings during spraying applications, and multispectral satellite imagery. The data were collected from five experimental units (fields), each with the same set of variables and crop type. The N applications were carried out on different dates based on crop needs and farmer decisions, with harvest dates also varying among the five fields. However, the dates of N application and harvest operations did not differ by more than five days between fields. The field operation data, such as harvested yield and N application rates, were documented through the John Deere Operation Center. Data from Sentinel-2 satellite imagery were collected from three different dates before the N 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. Detailed information about the trial data is provided in Table 3.1.

Table 3.1. Data source for this study.

Operation	Product	Range	Data
Nitrogen application	Ammonium sulfate saltpeter	20-75 kg/ha 26%N	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

The quantity of nitrogen (N) applied at each location in the field was determined by the N-sensor, with the application rates recorded through the sensor's readings and internet connectivity in the John Deere Operations Center. At the end of the season, the crop was harvested, and the yield for each location was measured using yield monitoring technology in the combine harvester. These harvest observations were then documented in the Operations Center. The details of the measurement technology of the above-mentioned data are described in the previous study by Abdipourchenarestansofla and Piepho (2024).

Twenty-one different indices were calculated from the spectral bands of Sentinel-2 for data fusion with field operation data from the target area. The imagery was acquired on at least three different revisit dates before the actual nitrogen (N) application. This approach assumes that there is a significant interaction between MSIs and N rates, necessitating the acquisition of MSIs before the actual N application. In this study, only 6 indices are used from 21 MSIs derived from multispectral imagery to estimate the EONR. The 6 indices are selected from the findings out of the first research by Abdipourchenarestansofla and Piepho (2024). The objective is to leverage the developed framework from the reference study mentioned earlier to develop a more robust RSM that can be generalized across different fields with different Nmin values. Various N treatments were mapped in the trial based on a management zones design, with N rates determined through crop sensing technology mounted on the machinery for real-time assessment of the N required for the crop. To address the varying spatial resolutions of map layers in this study, an automated data preprocessing pipeline was established to streamline the final analysis and support scalability in data production. Identifying and eliminating outliers in harvester yield observations and sensor readings from applicator machines is crucial for accurate analysis. Telematics data from agricultural machinery often contains errors due to operational dynamics, such as speed fluctuations, irregular field topography, and incomplete use of cutting bars during harvesting. An unsupervised machine learning approach, as proposed by Abdipourchenarestansofla & Piepho (2023), was employed to clean the telematics data, utilizing interquartile range and Robust Kernel Outlier Factor for outlier detection. For effective data integration and analysis of nitrogen (N) application at production levels, it is vital to ensure the quality of sensor measurements and the accuracy of documented telematics data, including GPS accuracy and metadata (Abdipourchenarestansofla & Schroth, 2022). After outlier removal, the data were interpolated to a 5m spatial resolution. In smaller fields, this aggregation resulted in greater homogeneity, while larger fields (e.g., >30 ha) maintained spatial variability. The applied N rates collected from the spreader do not accurately represent the exact area where N was applied, as distribution occurs in the vicinity of GPS points. Variogram-based interpolation was used to recover this data effectively, ensuring the descriptive statistics remained consistent pre- and post-interpolation. When fitting the RSM, interactions between N and MSIs were evaluated, revealing a quadratic relationship indicative of optimal N application. This aligns with established spatial data analysis principles. Comparing distributions and visualizing map layers for various field sizes is a robust methodology for determining the most suitable data handling techniques (aggregation or interpolation) tailored to the study's objectives. Adapting these analysis methods to the spatial characteristics of the issue at hand can enhance decision-making in precision agriculture (PA).

3.3. Model Averaging and Generalization

Crop characteristics at relevant phenological stages serve as indicators of nitrogen (N) deficiency. In this study, yield response modeling was conducted using RSM, incorporating MSIs as proxies for N deficiency at critical growth stages. The response shape to N for a given index value varies spatially when there is an interaction between N and that index. For RSM to function effectively, it is essential that the interaction is significant and that the quadratic term for applied N is negative (Box & Draper, 2007). This condition ensures that the conditional polynomial at a specific location in the field, based on the covariate values, achieves a maximum yield response (Abdipourchenarestansofla & Piepho, 2024). RSM encompasses a range of statistical and mathematical techniques crucial for developing, improving, and optimizing agricultural processes (Box & Draper, 2007). The methodology includes an average model approach utilizing leave-one-out cross-validation across five distinct fields. Each field is iteratively excluded as a test set while the RSM is employed to fit the data from the remaining four fields individually. The average coefficients obtained from the individual RSM models fitted to the training set (the four fields) are calculated to create a final averaged model used to estimate the EONR. RMSE was used as a tool for accessing model accuracy. This averaging technique supports compensation for heterogeneity within the data and enhances the scalability and generalizability of the model across varying crop conditions. By ensuring that the RSM can adapt to different agronomic practices and environmental factors, this approach facilitates more informed nitrogen management strategies in PA. Modeling RSM can be described as follows, given a harvester yield observation and its pairs of explanatory variables (i.e., as applied N rate and one of the MSIs) the RSM is given by

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

where y_i is the harvested yield (kg/ha) at the i-th location in a given field, 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, is the model 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, ϵ is a random term variable having zero expectation and constant variance (Abdipourchenarestansofla, & Piepho, 2024). The least squares method (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 x_i only if there is a significant interaction between the applied rate and the given vegetation index. The presence of interaction is necessary for the central hypothesis of Precision Farming to hold, additionally, for the proposed RSM model to work the β_{NN} is expected to be < 0 that can indicate the quadratic shape of the repose curve at each location (Abdipourchenarestansofla, & Piepho, 2024). The response surface is computed with the ‘RSM’ package in the R language. This package provides several functions to facilitate classical RSM (Lenth, 2009). The widely used approach for selecting the best model based on information criteria may not always be straightforward to apply (Piepho & Williams, 2021). The average model combines predictions from multiple models to enhance predictive performance, mitigate overfitting, and improve robustness (Lyons et al., 2019) which has a large impact on the estimated EONR. The term "average model" refers to a model that combines the results of multiple individual models to produce a single predictive model. Averaging the parameters from the five fitted models essentially creates a new model that represents the general trend or behavior observed across all five crop fields. The computation of the weighted average of the coefficients from each model is calculated as

$$\theta_{weighted} = \sum_{m=1}^{K} w_m \theta_m, \qquad (2)$$

In this formulation, $\theta_{weighted}$ represents the weighted average of the coefficients form multiple models. Each model, indexed by m, had an associated coefficient vector θ_m and a corresponding weight w_m. The summation over $m = 1$ to K (where K denotes the number of models) aggregates the weighted contributions of each model’s coefficient. Thus, $\theta_{weighted}$ reflects a combined estimation that accounts for the relative importance (weight) of each model’s contribution.

Equation (2) describes how the average parameter estimate is calculated across multiple models. Since all individual models share the same set of parameter estimates—specifically, the six candidate MSIs, as-applied N, their quadratic terms, and interactions—but differ only in the datasets, formula (2) demonstrates the computation of the average model for a single parameter estimate for simplicity. This approach applies uniformly across all model parameter estimates. A step-by-step computation of the weighted model is described as follows.

1. Fit Individual Models:

Fit a separate model to each dataset (four different crop fields) and obtain the coefficients and performance metrics.

2. Calculate RMSE:

For each model, calculate RMSE. RMSE measures the average deviation of predicted values from the actual values. See equation 3 for RMS calculation.

3. Determine weights:

Normalize the performance metrics to calculate weights for each model. For R-squared, higher values indicate better model performance, while for RMSE, lower values indicate better performance. In this study, the weight is calculated only based on RMSE.

Using the following formulas to derive weights based on RMSE:

$$w_m = \frac{\frac{1}{RMSE_m}}{\sum_{k=1}^{n} \frac{1}{RMSE_k}} \tag{3}$$

Here, w_m is the weight for the $m\text{-}th$ model, and $RMSE_m$ is the root mean square error of the $m\text{-}th$ model, and K is the total number of the models.

This formulation ensures that the weights w_m will sum to 1 across all models. Once these weights are computed, the weighted average of the coefficients from each model can be calculated using equation 2.

A demonstration of the computation of the proposed weighted average modeling formulation is given in the following. For simplicity let's consider only one parameter estimate β_{NN} for the quadric term N^2;

Let's say we have fitted four models to different datasets and calculated their R-squared values and RMSE as follows:

Table 3.2. Output of the models for individual fit of RSM on each experimental field.

Model	Adjusted R-squared	RMSE kg/ha	Coefficient (θ) N^2
1	0.31	644	-121.319
2	0.35	1517	-9.4944
3	0.45	1116	-0.17877
4	0.70	1097	-6.65809

Using R-squared for weights:

1. Calculate weights:

$$w_1 = \frac{0.31}{0.31 + 0.35 + 0.45 + 0.70} = \approx 0.1713$$

$$w_2 = \frac{0.35}{0.31 + 0.35 + 0.45 + 0.70} \approx 0.1934$$

$$w_3 = \frac{0.45}{0.31 + 0.35 + 0.45 + 0.70} \approx 0.2486$$

$$w_4 = \frac{0.70}{0.31 + 0.35 + 0.45 + 0.70} \approx 0.3867$$

2. Compute the weighted average of the coefficients using equation (2):

$$\theta_{weighted} = (0.1713 \times -121.319) + (0.1934 \times -9.4944) + (0.2486 \times -0.17877) + (0.3867 \times -6.65809) \approx -25.241$$

Using RMSE for weights:

1. Calculate weights:

$$w_1 = (1/644) / (1/644 + 1/1517 + 1/1116 + 1/1097) \approx 0.3866 w_2$$
$$= (1/1517) / (1/644 + 1/1517 + 1/1116 + 1/1097) \approx 0.1641 w_3$$
$$= (1/1116) / (1/644 + 1/1517 + 1/1116 + 1/1097) \approx 0.2230 w_4$$
$$= (1/1097) / (1/644 + 1/1517 + 1/1116 + 1/1097) \approx 0.2263$$

2. Compute the weighted average of the coefficients:

$\theta_{weighted}$ = (0.3866 × -121.319) + (0.1641 × -9.4944) + (0.2230 × -0.17877) + (0.2263 × -6.65809) ≈ -50.011

Weighted average coefficient using Adjusted R-squared: $\theta_{weighted} \approx -25.241$
Weighted average coefficient using RMSE: $\theta_{weighted} \approx -50.011$

In summary, weighted averaging using R-squared or RMSE allows for a more refined approach to combine model predictions by accounting for their relative performance. The weights derived from these metrics help ensure that better-performing models have a greater influence on the final averaged coefficients, providing a robust estimation that can enhance predictions in the nitrogen management framework.

3.4. Vegetation indices selection

Vegetation indices derived from multispectral imagery can be susceptible to multicollinearity, necessitating a robust feature selection procedure before modeling. Initially, 21 indices were obtained for three separate dates to develop the EONR framework. During the framework's development in the first study (Abdipourchenarestansofla & Piepho, 2024), 6 indices (Table 3.3) were ultimately selected. Considering three image acquisition dates and 21 MSIs for each date, a thorough examination of all index-date combinations designated as x_i was performed. The feature screening for MSIs was methodically executed through three distinct phases of feature selection to identify the optimal set for nitrogen (N) modeling. The satellite imagery acquisition dates were chosen to cover three consecutive days within the Sentinel-2 revisit cycle (every five days), from March 25, 2020, to April 6, 2020. Each of the three MSI dates underwent an independent feature selection across the three phases. The process commenced with Phase One, which employed analysis of variance (ANOVA) tests based on RSM output, followed by Phase Two utilizing lasso regression, and concluded with Phase Three, which involved a model reduction approach using RSM. For details of the selection procedure during all three phases see "vegetation indices selection" in Abdipourchenarestansofla & Piepho (2024). In the end, the candidate MSIs (Montero et al., 2023) from each date were combined to create the final second-order RSM.

Table 3.3. Candidate MSIs derived from sentinel-2 multispectral imagery for this study.

MSIs	Description	MSIs	Description
$ARI2_{2020.03.25}$	Anthocyanin reflectance index. Plant stress responses, understanding plant pigmentation	$CARI2_{2020.03.25}$	Chlorophyll Absorption Ratio Index 2 Enhancement of the original Chlorophyll Absorption Ratio Index
$CARI_{2020.03.30}$	Chlorophyll Absorption Ratio Index Monitoring plant health, photosynthetic activity, and stress responses	$CARI2_{2020.04.04}$	Chlorophyll Absorption Ratio Index 2 Enhancement of the original Chlorophyll Absorption Ratio Index

$CRI700_{2020.04.04}$	Carotenoid Reflectance Index	$CRI550_{2020.03.30}$	Carotenoid Reflectance Index Provides information on the relative abundance of carotenoids in plant tissues

3.5. Site-Specific Determination of Economic Optimum N-Rate

The EONR represents the nitrogen application rate that maximizes net returns. It is a key principle in agricultural economics that seeks to balance the potential crop yield with the costs associated with nitrogen fertilizer application. This approach is crucial for optimizing agricultural practices to attain maximum yield while ensuring economic efficiency. The calculation of the EONR is described in Abdipourchenarestansofla & Piepho (2024). The economic return is given by the following equation,

$$Return = P_W \times Y(n) - P_N \times n \qquad (3)$$

where the winter wheat yield per ha is denoted as Y, and is determined by nitrogen inputs, and is therefore expressed as a function $Y(n)$. The $Y(n)$ modeling is the functional equation given in formula (1). Indeed, the yield function $Y(n)$ relies on the RSM, which provides a framework for understanding how various factors, including nitrogen inputs, influence crop yield. The RSM incorporates multiple independent variables, including nitrogen application rates (N) and multispectral indices (X) derived from remote sensing data. The yield response $Y(n)$ varies spatially across the field due to several factors: 1. Variability in MSIs: The MSIs, which represent crop characteristics, vary across the field because of differing environmental conditions. These conditions can include soil type, moisture levels, and nutrient availability, all of which influence crop characteristics and health. 2. Interaction between nitrogen and MSIs: The interaction between nitrogen application rates and the spatially varying MSIs is a crucial aspect of the model. As the N input changes, its effect on yield is influenced by the local environmental conditions represented by the MSIs. For instance, areas with higher vegetation indices may respond differently to nitrogen application than areas with lower indices due to their inherent growth potential and nutrient status. 3. Field Heterogeneity: Agricultural fields are inherently heterogeneous, meaning that the response to nitrogen applications will not be uniform. The RSM captures these spatial interactions, allowing for tailored nitrogen management strategies that optimize yield based on localized conditions. 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., 2018). 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.

3.6. Results

3.6.1. Cross-validation results

The model performance evaluation is done utilizing RMSE. As the estimation of EONR is done through yield modeling, the stability of model performance during the test phase is essential to estimate a reliable range for EONR. Hence the RMSE values are demonstrated in Figure 3.2.

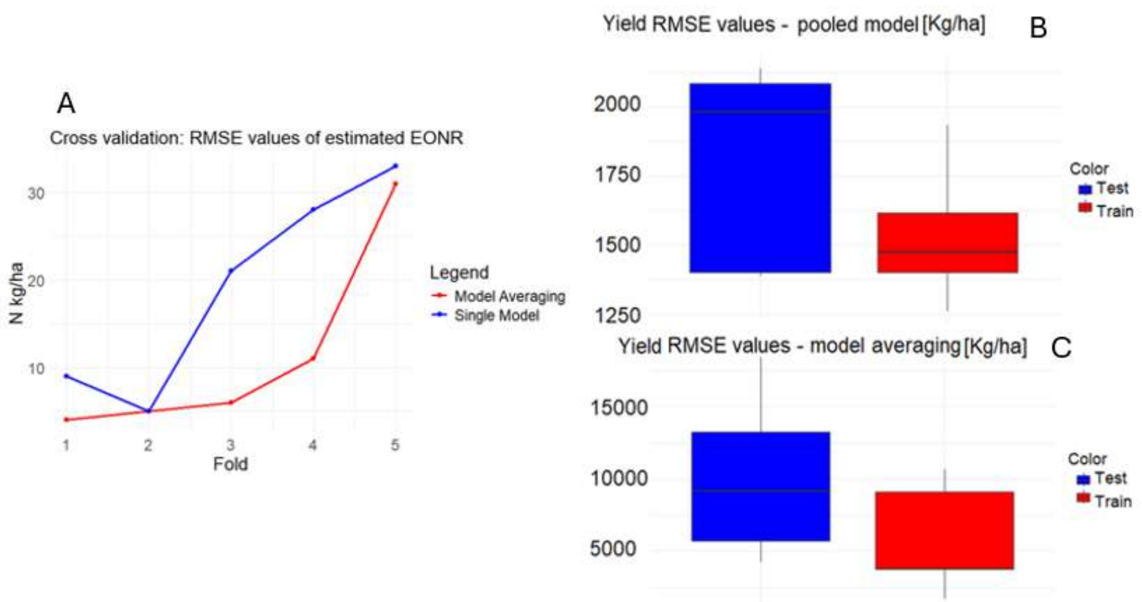

Figure 3.2. Cross-validation results for the EONR modeling approaches. **(A)** The RMSE values of the estimated EONR for each fold, comparing the Single Model (blue) against Model Averaging (red). **(B)** Box plots showing the distribution of Yield RMSE values for the **pooled model** during training (red) and testing (blue). **(C)** Box plots showing the distribution of Yield RMSE values for the **model averaging** approach during training (red) and testing (blue).

Figure 3.2 illustrates the comparison between two modeling approaches for estimating the economically optimal nitrogen rate (EONR) and the associated yield predictions across multiple folds of cross-validation. The graph in figure 3.2 labeled A displays the RMSE values for EONR estimates for each fold using both the single model and the model averaging approach. The single model (blue line) exhibits consistently higher RMSE values across folds compared to the model averaging approach (red line), suggesting that model averaging achieves better predictive stability and accuracy for EONR estimation.

Graphs B and C provide insights into the RMSE values for yield predictions. In B, the pooled model's RMSE values are significantly higher for the test dataset (blue box) compared to the train dataset (red box), reflecting the expected generalization gap. In C, the model averaging approach demonstrates a similar pattern, but the training RMSE is higher relative to the pooled model, indicating a trade-off between model complexity and generalization. Overall, the figure highlights the superiority of model averaging in reducing errors in EONR predictions while maintaining reasonable yield prediction accuracy.

3.6.2. A comparison of Pooled Model vs. average model

Estimating a reliable and robust model is vital for determining EONR. A comparison of the Pooled Model vs the average model is demonstrated through prescription maps (EONR maps) for individual fields (Figure 3.3).

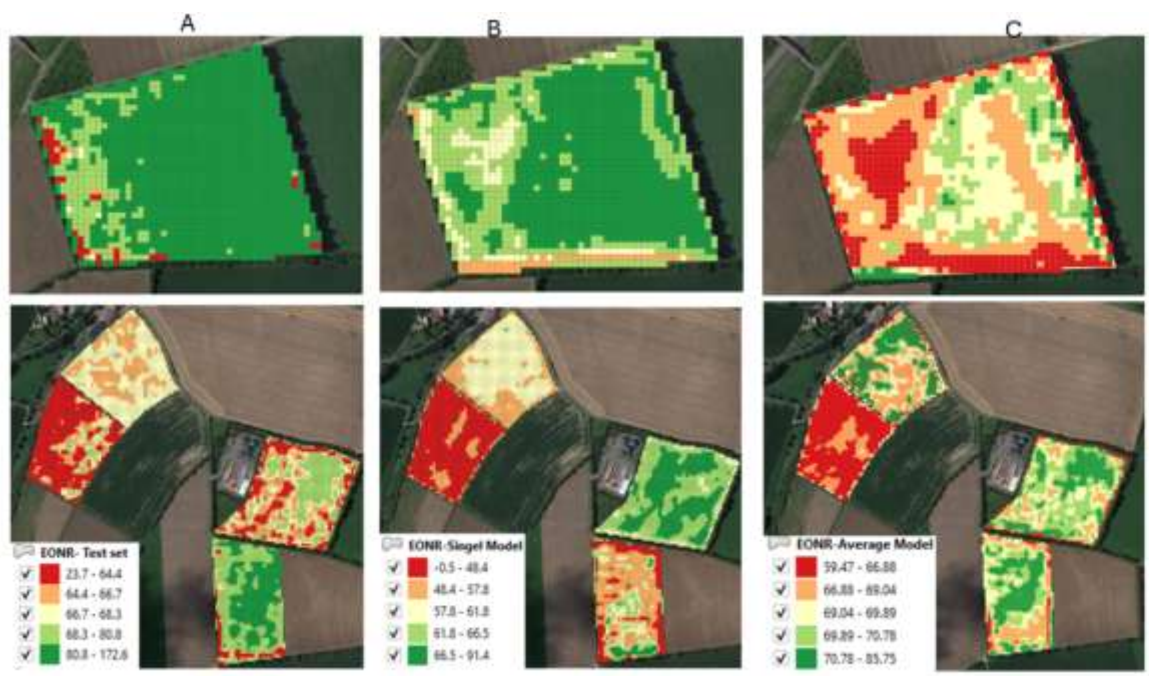

Figure 3.3. Comparison of the spatial distribution of output EONR on three models of EONR distribution. A represents the actual applied rate acquired from the YARA sensor, B is the EONR estimated from the pooled model, and C is the EONR estimated with the average model.

A comparison between the pooled model and the average model is conducted to assess the effectiveness of different approaches in estimating the EONR. The results indicate that the average model framework, which uses weighted averaging, achieves higher prediction accuracy and reliability, with an average RMSE of 11 kg N/ha and a mean EONR of 69.04 kg N/ha. In contrast, the pooled model exhibited a higher RMSE of 19 kg N/ha and a lower mean EONR of 57.4 kg N/ha. The EONR range predicted by the Pooled Model was from -0.52 to 91.44 kg N/ha, compared to a more reliable range of 59.47 to 85.75 kg N/ha from the average model. This suggests that the average model provides a more stable and realistic EONR estimation, as the pooled model produced a negative minimum value, which is impractical for nitrogen application.

3.7. Discussion

The comparison between the EONR estimated through a pooled model and the average model provides valuable insights into nitrogen management strategies in PA. The Pooled Model, based on a Pooled Model fitted across observations of four different fields and tested on the test set, offers a mean nitrogen of 57.4 kg/ha application rate that reflects the unique agronomic circumstances of each site. Conversely, the average model, which aggregates individual models across multiple fields, yields a higher estimated nitrogen application rate of 69.04 kg/ha, indicating its role in providing a broader perspective on nitrogen needs.

The cross-validation analysis further highlights key differences in the two modeling approaches. RMSE values for EONR predictions across folds show that the pooled model exhibits higher prediction errors compared to the average model, indicating that the average model achieves better predictive stability and accuracy. Moreover, yield RMSE values reveal a generalization gap for the pooled model, with test errors being significantly higher than training errors. While the average model also displays this gap, its RMSE values indicate a trade-off between model complexity and generalization that better aligns with practical nitrogen management needs.

The scalability of these models was effectively demonstrated through K-fold cross-validation across five distinct fields, showcasing their adaptability to various agricultural datasets. This approach not only reinforced the robustness of the models but also highlighted their generalization capabilities. For winter wheat crops in the study region, the average rate of the first nitrogen (N) application is typically reported by farmers to range from 60 kg/ha to 90 kg/ha, based on years of experience, and is also supported by findings from Laidig et al. (2024). The results suggest that the average model demonstrates greater robustness and reliability in estimating the EONR range of (59-85 kg/ha), while the Pooled Model tends to produce a lower N rate (-0.52 – 91 kg/ha), which can lead to under-fertilization and subsequent loss of returns. The main challenge in developing such tools lies in the uncertainties inherent in weather forecasts and crop price projections, which are necessary for estimating yields and returns at various nitrogen rates. Consequently, it is essential to characterize the risks associated with these uncertainties to accurately determine optimal nitrogen rates during the growing season (Chen et al, 2024). A study by Matavel et al. (2024) highlights the complexities in estimating the EONR due to uncertainties in input-output relationships and site-specific factors, demonstrating that the average model can effectively mitigate biases and provide reliable estimates, even in the absence of a known true model.

The integration of advanced modeling techniques, combined with continuous data collection, is essential for enhancing nitrogen management strategies. Future efforts should focus on refining these models by incorporating real-time data from IoT sensors and satellite imagery, thus further optimizing nitrogen application rates while minimizing environmental impact. The insights gained from this study underscore the framework's potential for widespread adoption in PA, providing a scalable and robust solution for optimizing nitrogen application across various agricultural settings. The findings emphasize the importance of both individual model accuracy and the broader applicability of average models in decision-making processes related to sustainable crop production.

Chapter 4: A Scalable Cloud Computing Infrastructure for Decision Making in Precision Agriculture: The Use Case of Estimation of Economic Optimum Nitrogen Rate Optimum Nitrogen Rate

Abstract

This paper presents a comprehensive and scalable cloud computing framework tailored to estimate the Economic Optimum Nitrogen Rate (EONR) for precision agriculture, addressing the challenge of optimizing nitrogen application across diverse farming environments. The framework integrates agronomic telemetry data from machinery, such as harvest monitoring and nitrogen application rates, with multispectral satellite imagery to process large datasets efficiently. Key components of the architecture include Amazon Web Services (AWS) Batch for orchestrating containerized jobs, Amazon S3 for scalable, region-specific data storage, and RDS Postgres for structured data management, ensuring the system can handle both real-time and historical data streams seamlessly. The architecture's core innovation lies in its ability to dynamically process, harmonize, and interpolate spatially heterogeneous datasets using advanced geospatial techniques like Kriging interpolation, producing highly accurate nitrogen prescription maps. These maps are crucial for guiding real-time, data-driven nitrogen application decisions at the field level, enhancing both crop productivity and environmental sustainability. A case study validates the framework’s capability to scale across large regions, efficiently manage telemetry and satellite data, and generate actionable nitrogen management insights. By leveraging cloud-native technologies, this system offers an adaptable, cost-effective solution that enhances the precision and sustainability of nitrogen management practices in modern agriculture. Furthermore, the use of automated data quality control, data harmonization, and scalable cloud storage ensures that the framework is well-equipped to handle increasing data volumes and varying field conditions, providing a robust foundation for future advancements in precision farming.

Keywords: Batch Processing, Agronomic Telematic Data, Cloud Computing, Economic Optimum Nitrogen, AWS, Geospatial Data Processing, Precision Agriculture

4.1. Introduction

The increasing availability of large-scale geospatial data is reshaping precision agriculture, with nitrogen (N) management being a key area of focus. Precision agriculture relies heavily on data-driven decisions to optimize crop yields, enhance resource efficiency, and reduce environmental impact. With the advent of technologies like the Internet of Things (IoT) and cloud computing, the integration of telemetry data from agricultural machinery and satellite multispectral imagery has opened new possibilities for real-time crop monitoring. These technologies provide farmers with detailed insights into field conditions, helping to determine the Economic Optimum Nitrogen Rate (EONR), which is crucial for sustainable and profitable nitrogen management. Despite these advancements, there is still limited work on cloud-based processing and analytics for agronomic telemetry data. Most studies have focused on basic machinery status monitoring, leaving a gap in leveraging historical telemetry data in conjunction with satellite imagery to optimize nitrogen application decisions. Telemetry data, often amounting to terabytes per season, combined with high-resolution satellite imagery, creates challenges in data management and analysis. Traditional data infrastructure methods struggle to keep up with the volume and complexity of these datasets, making it difficult to derive actionable insights in a timely manner. Cloud computing has emerged as a viable solution to these challenges by providing scalable infrastructure for data integration, storage, and real-time analytics. By leveraging cloud services, large datasets can be processed efficiently, enabling timely, data-driven decisions in nitrogen management. Recent studies have demonstrated the effectiveness of cloud-based infrastructures in handling large volumes of satellite and telemetry data. For instance, Yadav et al. (2024) and Dash et al. (2024) highlight how cloud platforms can integrate diverse data sources and provide scalable computing power to support geospatial analysis. This enables the optimization of agricultural inputs like nitrogen at a more granular level, improving decision-making accuracy. The use of cloud computing in precision agriculture has already shown promising results. Jacobsen et al. (2017) demonstrated how elastic cloud computing could enhance the performance of geospatial data analytics on satellite data, leading to improved agricultural outcomes. Similarly, Morchid et al. (2024) presented a smart irrigation system that uses IoT, telemetry, and cloud platforms to optimize water use and improve sustainability. These studies underscore the potential of cloud computing to revolutionize data processing and decision-making in agriculture. Moreover, innovative systems like "GeoEkuiper" by Huang and Deng (2024) offer scalable solutions for real-time spatial data processing, even on resource-constrained devices. This enables efficient handling of large geospatial datasets, such as those used in nitrogen management, through cloud-native processing engines. Other studies have explored plant disease detection frameworks using IoT, embedded systems, and machine learning, demonstrating how these technologies can further enhance the precision and effectiveness of smart agriculture. Specific to agricultural machinery, most work to date has centered on basic data collection and diagnostics. For instance, Hao et al. (2018) developed an ARM-based, embedded Linux system to monitor agricultural machinery, combining real-time diagnostics with cloud integration. However, more advanced cloud-based analytics for optimizing nitrogen management, particularly using telemetry and satellite data remain underexplored.

This paper addresses that gap by proposing a cloud computing framework specifically designed to manage and process large-scale geospatial data for nitrogen management in precision agriculture. The framework integrates both telemetry data from agricultural machinery and satellite imagery, supporting a wide range of data processing and analytical functions. Through a case study, the paper demonstrates the framework's effectiveness in optimizing nitrogen application, increasing productivity, and reducing environmental impacts. By providing a scalable and efficient infrastructure, this framework enables real-time decision-making and supports more sustainable nitrogen management practices in modern agriculture.

4.2. Conceptual Architecture Framework

For better performance, a shell script is the quickest way to get data from a database since it runs close to the database. However, running a dedicated server for this is not cost-effective or sustainable. That is where a batch-based job can help. It provides on-demand processing power without needing to manage servers when running on the cloud environment. We used the AWS Batch to run a containerized app, which is triggered by an AWS lambda function, providing a fast and simple solution. The architecture for computing the EONR leverages various AWS cloud services to handle data ingestion, storage, processing, modeling, and visualization. Designed to process large volumes of agronomic telemetry data and satellite imagery, this system provides flexibility, scalability, and efficiency through containerized batch jobs and automated workflows. The EONR computation integrates historical telemetry data with spectral analysis of satellite imagery, combining nitrogen application data, harvest monitoring data, and vegetation indices (VIs) to model the optimal nitrogen application for winter wheat fields. The following outlines the detailed design and logic of the development of our proposed automated, cloud-based pipeline for computing EONR in precision agriculture using various AWS services. See Figure 4.1.

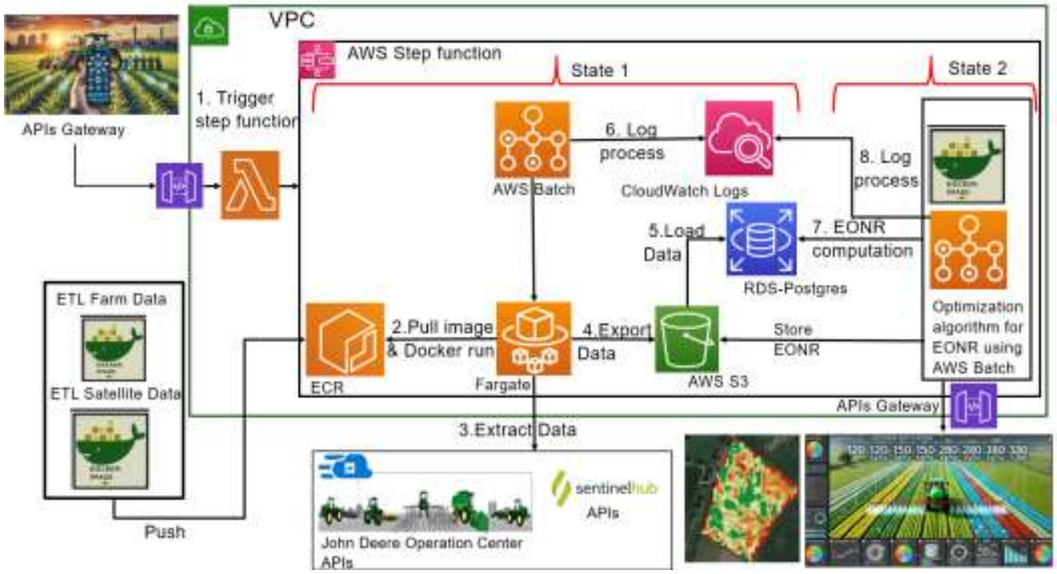

Figure 4.1. Cloud architecture for scaling EONR into production.

The infrastructure presented in Figure 4.1 works as follows: Users initiate the process through an Application Programming Interface (API) Gateway, providing key inputs such as field boundary, crop type, and operation date. The workflow is orchestrated by AWS Step Functions, which triggers two main ETL tasks managed by AWS Batch. The ETL jobs, packaged as Docker images in Amazon ECR, extract telemetry and satellite data from sources such as the John Deere Operation Center and Sentinel Hub. The data is processed on AWS Fargate, stored in Amazon S3, and loaded into an RDS Postgres database. Subsequently, an optimization algorithm for EONR is executed via AWS Batch, utilizing the data in RDS to generate a prescription map. This output is then stored in S3 and made accessible to clients through APIs for real-time use in agricultural equipment, providing an efficient, scalable solution for data-driven nitrogen management. The Virtual Private Cloud (VPC) in this architecture represents an isolated network environment that provides secure access and communication between various AWS services. It ensures that data flows, computations, and storage within the architecture remain secure and isolated from the public internet, adding an extra layer of security to the data and computations related to EONR management.

4.2.1. Data Storage: Amazon S3 and RDS Postgres

The architecture designed for estimating the EONR (Figure 4.1) leverages AWS to address the complex data management requirements of precision agriculture. A crucial component of this system is the data storage strategy, which not only efficiently handles large volumes of telemetry data from machinery and satellite imagery for each farm, but also ensures scalability, compliance, security, and cost efficiency.

Challenges of Data Management in EONR Estimation:

Estimating EONR is a complex process that relies on integrating diverse data sources, such as telemetry data from machinery (e.g., harvest monitoring, as-applied nitrogen application rates) and high-resolution satellite imagery. The primary challenge lies in managing these large, heterogeneous datasets, which continue to grow due to the temporal variability of crop needs and the need for continuous measurement of crop characteristics. Additionally, as farms expand and data collection frequency increases, traditional Farm Management Information Systems (FMIS) struggle to efficiently store, retrieve, and analyze these vast data volumes. Another layer of complexity is added by the need to comply with data protection regulations (see Table 4.1) which requires data to be stored and processed within its geographical boundaries (Bakare et al., 2024). Therefore, a robust, scalable, and regulation-compliant data management solution is essential to support decision-making for on-farm nitrogen application experimentation.

Table 4.1. Data privacy laws around the globe.

Region/country	Privacy Law	Description
European Union	GDPR	General Data Protection Regulation.
California	CCPA	California Consumer Privacy Act.
Canada	PIPEDA	Personal Information Protection & Electronic Documents Act.
Brazil	LGPD	Lei Geral de Proteção de Dados.

Proposed Storage Solution: A Regional and Farm-Specific Approach

To address these challenges, the proposed architecture adopts a regionalized data storage strategy using Amazon S3 and RDS Postgres, allowing both horizontal and vertical scaling. This regional approach ensures compliance with data protection regulations and facilitates cost efficiency. Amazon S3 is employed for unstructured data storage, where each farm within a specific geographical region is allocated a dedicated folder within a region-specific S3 bucket. This setup isolates machine telemetry data and satellite imagery into individual storage locations. The regionalization of S3 storage minimizes data overlap, streamlines data processing, and allows for elastic growth as more farms are onboarded. Additionally, the lifecycle management policies of S3 enable automatic archiving of older data to more cost-effective storage classes, addressing long-term data storage needs.

For structured data, this architecture employs a single RDS Postgres instance per region. Conceptual database schemas (Figure 2) were created to store data such for individual farms such as fields, boundaries, historical nitrogen application and harvest monitoring data with their metadata, along with vegetation indices derived from sentinel-2 satellite imagery. This design consolidates data within a centralized database, allowing for efficient partitioning by farm ID, field ID, or specific field operations. Such segmentation facilitates fast and secure data retrieval. By regionalizing data storage in both S3 and RDS Postgres, the system not only complies with data protection regulations but also minimizes data transfer costs and latency. By leveraging these scalable storage practices, the system ensures it can handle the increasing influx of both telemetry and satellite data, enabling seamless integration with analytics workflows and long-term storage efficiency.

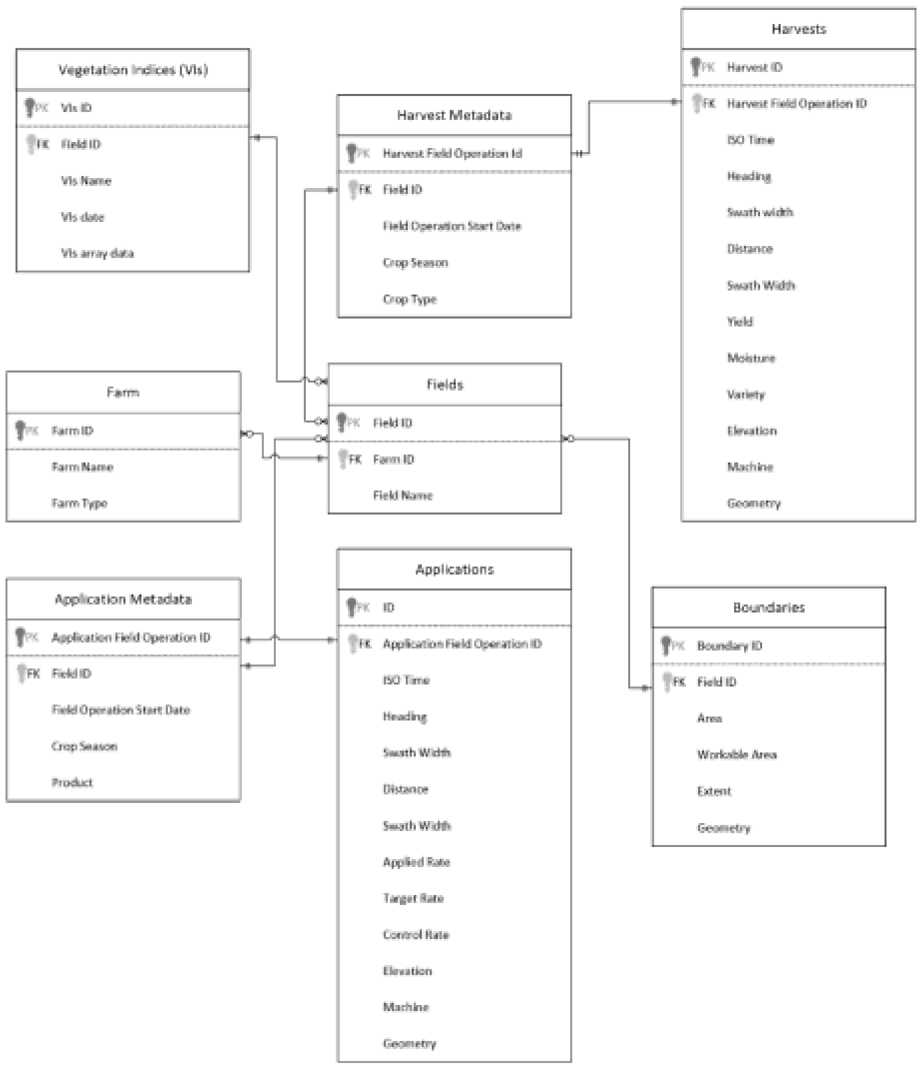

Figure 4.2. Relational database schema for data modeling.

The relational database schema models the relations between key farming entities such as Farms, Fields, Harvests, Applications, and Vegetation Indices (VIs).

The database schema presented in Figure 4.2 enables analytics (computation of EONR) through a fast and robust data retrieval. Farms are linked to fields, which in turn are connected to both harvest and application operations. Metadata tables store detailed information about each operation which enhances the query computation when searching for a particular field operation. Boundaries define the physical limits of fields. The schema centralizes data management by using Field ID and Field Operation ID to link related tables, ensuring efficient tracking of farming activities across various processes.

4.3.2. Data Ingestion and Preparation: AWS Batch with Dockerized Applications

Data Ingestion for Telemetry Data

Telemetry data, which includes machinery operation metrics, as-applied nitrogen (N) application rates, and harvest monitoring data, is collected from agricultural machinery equipped with various sensors. This data is often available through platforms such as the John Deere Operation Center APIs, which provides data in vector formats such as ESRI standard shapefiles. In the architecture outlined (Figure 4.1), telemetry data ingestion begins with the user-triggered AWS Step Functions workflow. The Step Functions orchestrate an ETL (Extract, Transform, Load) pipeline managed by AWS Batch and executed using Docker containers registered in Amazon Elastic Container Registry (ECR). The overall ETL pipeline pulls telemetry data from the John Deere Operation Center using the provided Application Programming Interfaces (APIs), processes the data on AWS Faregate, and stores the cleaned and transformed data in region-specific Amazon S3 buckets. This data is then further loaded into a region-specific Amazon RDS Postgres instance, where it is structured into a unified schema.

Data Ingestion for Satellite Imagery

Satellite imagery is another crucial data source for EONR estimation. In this architecture, satellite data is ingested from sources such as Sentinel Hub APIs, which provide high-resolution imagery with a revisit time of approximately five days. Like telemetry data, satellite imagery is ingested using a separate ETL pipeline managed by AWS Batch and executed in Docker containers on AWS Fargate. The ETL pipeline extracts relevant satellite scenes, computes various vegetation indices (e.g., NDVI, NDRE), and stores the raster images in GeoTIFF format within dedicated folders in region-specific S3 buckets. This strategy ensures that each farm’s satellite imagery is isolated, improving data retrieval efficiency for analysis. The processed vegetation indices are then loaded into RDS Postgres, where they are linked with the corresponding field (see Figure

4.2), creating a comprehensive dataset that includes both satellite imagery and field operation machinery aspects of crop management. The extract satellite data are used to compute relevant vegetation indices for final modeling (see Table 4.2).

Table 4.2. Spectral Indices derived from Sentinel-2 imagery for modeling EONR

MSIs	Description	MSIs	Description
$ARI2_{2020.03.25}$	Anthocyanin reflectance index. Plant stress responses, understanding plant pigmentation	$CARI2_{2020.03.25}$	Chlorophyll Absorption Ratio Index 2 Enhancement of the original Chlorophyll Absorption Ratio Index
$CARI_{2020.03.30}$	Chlorophyll Absorption Ratio Index Monitoring plant health, photosynthetic activity, and stress responses	$CARI2_{2020.04.04}$	Chlorophyll Absorption Ratio Index 2 Enhancement of the original Chlorophyll Absorption Ratio Index
$CRI700_{2020.04.04}$	Carotenoid Reflectance Index	$CRI550_{2020.03.30}$	Carotenoid Reflectance Index Provides information on the relative abundance of carotenoids in plant tissues

Combining Telemetry Data and Satellite Imagery

The integration of telemetry data and satellite imagery occurs within the region-specific RDS Postgres database. During the data loading phase, telemetry data and satellite-derived vegetation indices are linked based on field boundaries, operation dates, and specific crop seasons. This relational setup enables the system to correlate the timing and rate of N application with observed crop health indicators from satellite imagery. The combined dataset provides a holistic view of the field's conditions, supporting the modeling of EONR. When the user initiates the EONR computation through an API Gateway (Figure 4.1), the system retrieves this integrated dataset from RDS Postgres and feeds it into an optimization algorithm managed by AWS Batch. This process allows the algorithm to analyze both historical telemetry data and current crop status (from satellite imagery) to generate a precise nitrogen prescription map. The integration facilitated by the combined data ingestion pipeline ensures that the system provides accurate, real-time insights into nitrogen application decisions. An overview workflow for combining telemetry data and satellite imagery is demonstrated in Figure 4.3.

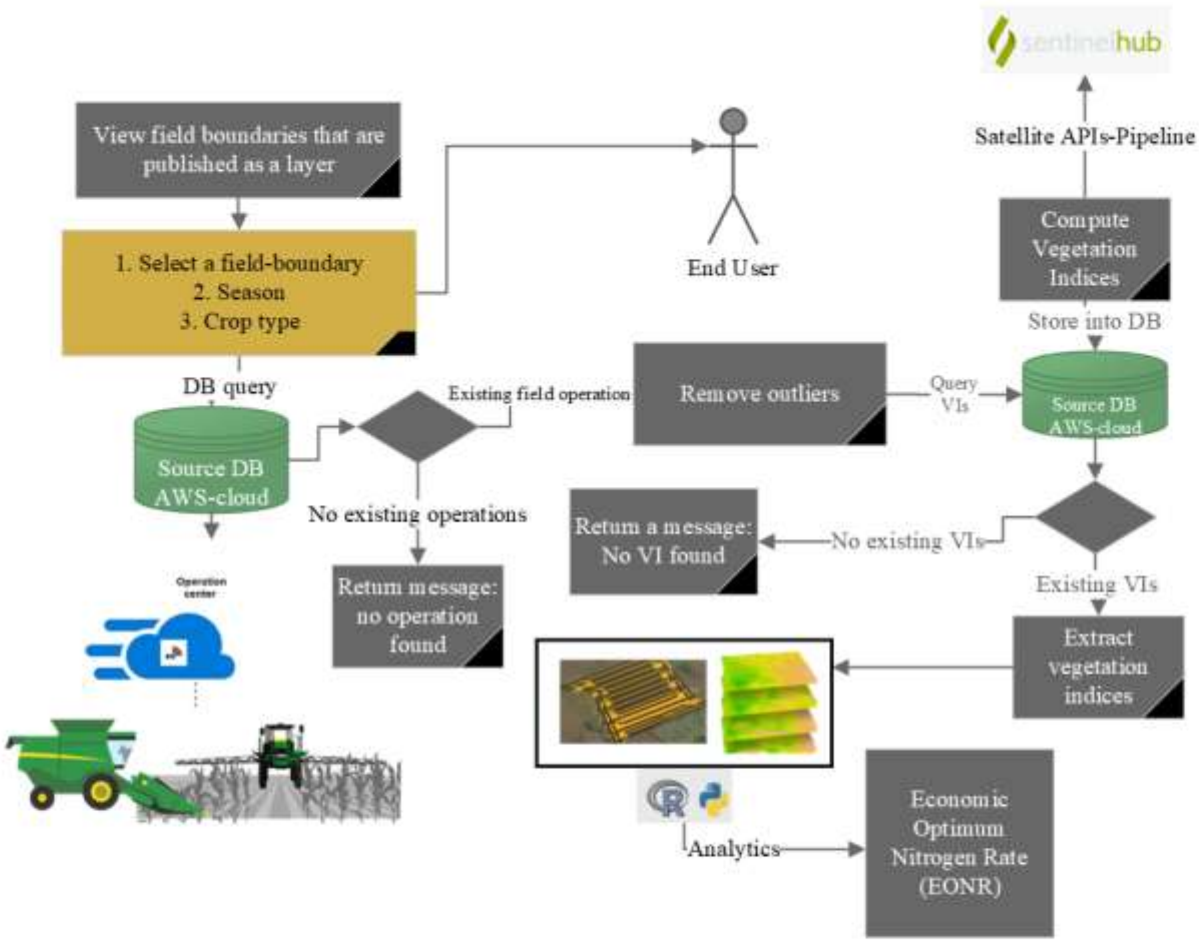

Figure 4.3. Telemetry data and satellite data ingestion.

This workflow seamlessly combines field-level telemetry data with remote sensing imagery to provide comprehensive agricultural solutions, in this case for estimation of EONR. Starting with user input on field boundaries, the system checks the database for existing harvest and nitrogen application operation data and computes vegetation indices using satellite imagery. The data undergoes multiple processing steps, such as removing outliers and storing them in an accessible format for analysis. The result is an informed application solution of EONR, that helps farmers make data-driven decisions to optimize nitrogen use, ultimately supporting sustainable and profitable farming practices.

4.3. Telemetry and satellite Data quality challenges and solutions

The estimation of the EONR requires the integration of three primary datasets: telemetry data from agricultural machinery, including as-applied nitrogen application data, harvest monitoring data, and multispectral satellite imagery. These datasets provide a comprehensive view of the field operations and crop health, which are critical for generating accurate EONR estimates. The combination of these datasets allows the system to assess crop responses to nitrogen application and other management practices over time, aiding in more precise nitrogen recommendations (see chapter 2).

Table 4.3. Represents information about data used for model generalization.

Data source	Data type	Acquisition dates	# fields	# datasets
Harvest	ESRI shapefile format	2020.07.13 - 2020.07.24	14	14
Fertilization	ESRI shape file format	2020.04.01 - 2020.04.06	14	14
Satellite (three dates before actual N application)	Geotiff raster layer	2020.03.22 2020.03.27 2020.03.27 2020.04.01 2020.04.01 2020.04.06	14	882 21 VIs for each date $21 \times 3 = 63$ 63 14 = 882

4.3.1. Telemetry Data from Agricultural Machinery

Telemetry data is acquired from the John Deere Operation Center and represents machinery operation metrics, such as harvest monitoring data and as-applied nitrogen application rates. The data is captured during field operations and is provided in ESRI Shapefile format along with a corresponding JSON metadata file. Each shapefile contains spatial point layers that represent the location of each sensor reading during field operations, providing details such as: Harvested grain weight (t/ha), which is a measure of crop yield during harvest operations, elevation, representing the topography of the field, vehicle speed, providing context for the operation's efficiency and uniformity, nitrogen application rate (kg/ha), which indicates the amount of nitrogen applied at each location in the field. These shapefiles provide detailed insights into spatial variability across the field (Figure 4.4).

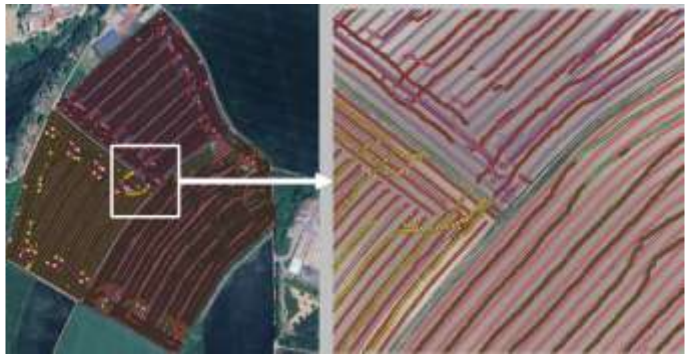

Figure 4.4. Illustration of field operation data, with multiple passes of equipment across the season.

Each pass in during fields operations corresponds to a specific nitrogen application, which is typically split into two applications during the growing season, as per common agricultural practice in the study region. The data also includes passes of the combine harvester, shown as separate layers. These detailed, spatially explicit datasets are critical for analyzing the impact of nitrogen application on crop yield and further optimizing N application for next season.

4.3.2. Satellite Imagery and Spectral Indices

The second major dataset comes from multispectral satellite imagery, which is retrieved from Sentinel-2 satellites through the Sentinel Hub platform. Sentinel-2 provides open access with 10 m spatial resolution and temporal resolution of 5 days revisit in the form of GeoTIFF raster layers, which can be used to compute various vegetation indices (VIs), such as the Chlorophyll Absorption Ratio Index (CARI) and Carotenoid Reflectance Index (CRI700). These indices are essential for monitoring crop health and growth stages (see chapter 2). After computing VIs, the relevant indices are down-sampled to 5 m (Figure 4.5) to better describe the spatial variability of the fields.

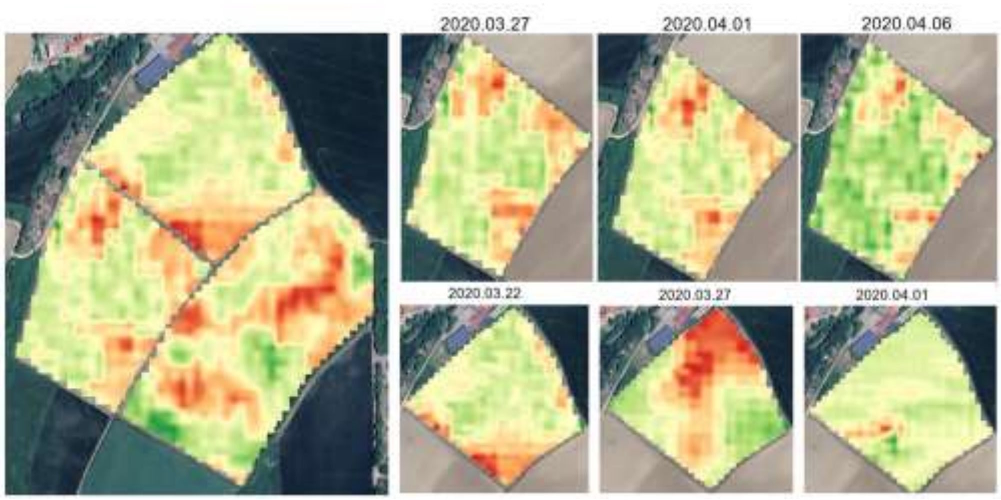

Figure 4.5. Representation of spatial variability in chlorophyll levels across March and April 2020.

This Figure shows CARI maps derived from Sentinel-2 imagery, capturing 5m spatial variability in chlorophyll levels across March and April 2020, with color gradients indicating shifts in crop chlorophyll content over time; these indices inform a model for estimating optimal nitrogen application rates to enhance yield.

The maps in Figure 4.5 show visualize the chlorophyll content of crops changing over time before the actual N application, with areas of higher chlorophyll absorption indicated by greener shades and lower absorption by redder shades. Each map corresponds to a specific date, demonstrating how crop health and photosynthetic activity fluctuate during this period, providing valuable insight for monitoring crop growth and identifying stress zones. These vegetation indices are then used as inputs to the RSM model to estimate the optimal nitrogen rate for improving crop yield.

4.3.3. Data Cleaning and Quality Control

Ensuring the quality of telemetry data and satellite imagery is critical for accurate EONR modeling. Telemetry data from agricultural machinery, such as harvest monitoring and nitrogen application sensor readings, are often contaminated by errors from the dynamic nature of field operations. Common sources of error include speed variations, irregular field topography, incomplete use of cutting bars, and delays in machine operations. These inconsistencies, along with sensor malfunctions (e.g., drift, noise, and dead sensors), create outliers that obscure the spatial structure of the observations, making it difficult to analyze and derive meaningful insights. A robust preprocessing strategy is required to clean the raw telemetry data. To address this, an unsupervised machine learning approach proposed by Abdipourchenarestansofla & Piepho (2022) was employed. The method detects and removes outliers by leveraging both the interquartile range (IQR) for global outlier detection and the Robust Kernel Outlier Factor (RKOF) for spatial outliers. For example, in yield data, where some measurements are recorded due to incomplete machine passes or sensor inaccuracies, the model filters out extreme values and flags suspicious data points. Additionally, metadata validation is incorporated to ensure that field operation documentation (e.g., GPS locations and user inputs) is accurate. Sensor deficiencies and inaccuracies in the metadata (such as incorrect field boundaries or crop season details) are also flagged for correction (Figure 4.6).

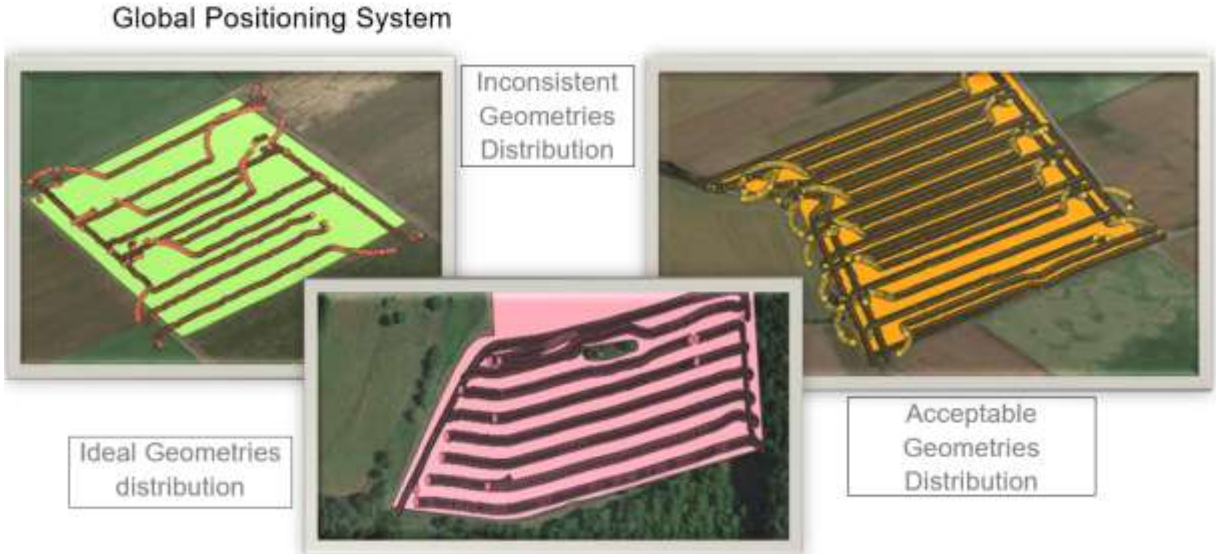

Figure 4.6. Variability in Global Positioning System (GPS) Distribution during Field Operations.

Figure 4.6 illustrates three examples of GPS geometries distribution during field operations. The left image shows an ideal geometries distribution where GPS data points are consistently aligned with the field boundaries and operation path. The right image displays inconsistent geometries distribution, where GPS data points show misalignment and irregular coverage due to machine drift or GPS signal issues. The bottom image presents an acceptable geometries distribution, where the GPS data is mostly aligned but contains some minor deviations, representing an operationally acceptable scenario.

In addition to sensor readings that capture operation attributes at each location, accurate documentation of field operations is essential for ensuring data quality in automated EONR modeling. Inconsistent or incomplete documentation, often caused by a lack of operator knowledge, can lead to errors in field operation records, as demonstrated in Figure 4.7. These inaccuracies compromise the reliability of nitrogen application recommendations and overall farm management decisions if not identified out during data cleaning and processing steps.

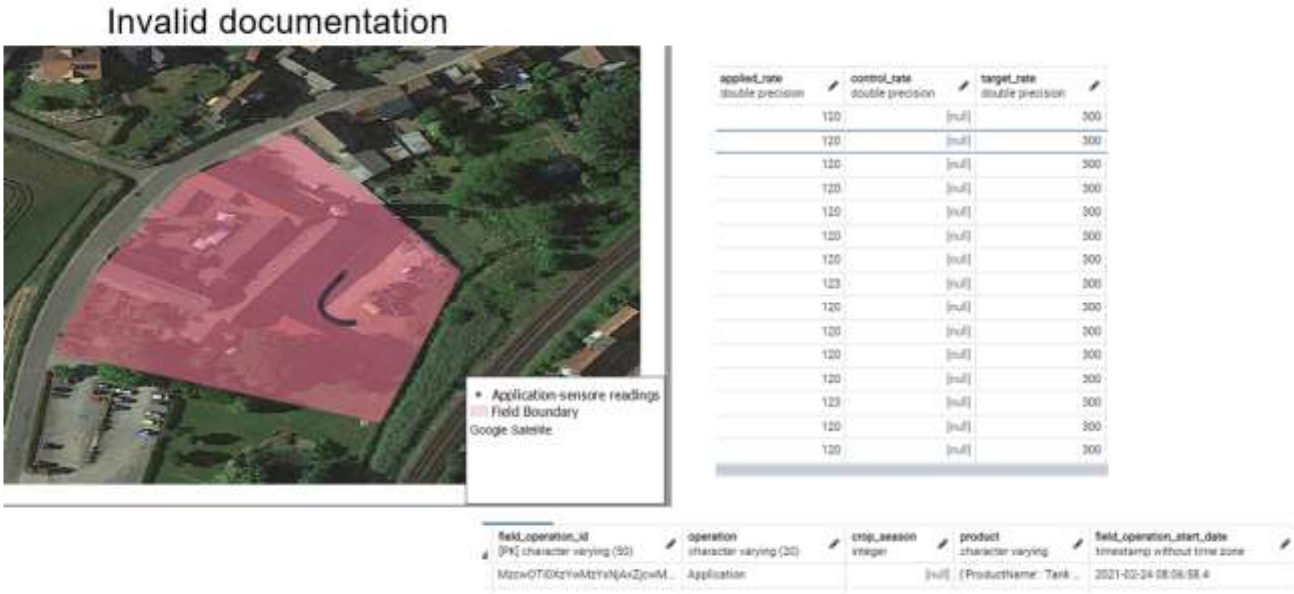

Figure 4.7. Example of Invalid Documentation in Field Operations.

Figure 4.7 shows an instance of invalid documentation during field operations. On the left, the application-sensor readings are incorrectly mapped to a field boundary that does not match the actual operational area, leading to errors in data accuracy. On the right, the data table reveals missing or null values in key operational fields, such as control rate and target rate, as well as incomplete metadata regarding the operation type, crop season, and field operation timestamps. These inconsistencies highlight the challenges posed by poor documentation, which can compromise the accuracy of nitrogen application and field management records.

To address the challenges of data quality of telemetry data the well-established ISO 25012 and 25024 standards were implemented. These standards define several data quality dimensions such as accuracy, completeness, consistency, and credibility. Based on our observation, it was determined that the most important dimensions are consistency, accuracy, and completeness. The results guided the design of a flexible and automated Data Quality Assessment (DQA) solution tailored to the quality changes of telemetry data. The solution ensures that each measurement is scored on a scale from 0% (worst quality) to 100% (best quality) and outputs the results in machine-readable formats defined as JSON format. The DQA Python library (as illustrated in Figure 4.8) implements the procedure for data quality analysis, taking measurement specifications (also provided in JSON) as input. The DQA library processes the incoming data (e.g., ESRI Shapefiles or JSON-based telemetry data) and outputs quality scores to determine whether a particular dataset can pass the DA gate or not. This solution was integrated into ETL-farm data, allowing for automated and scalable data quality control.

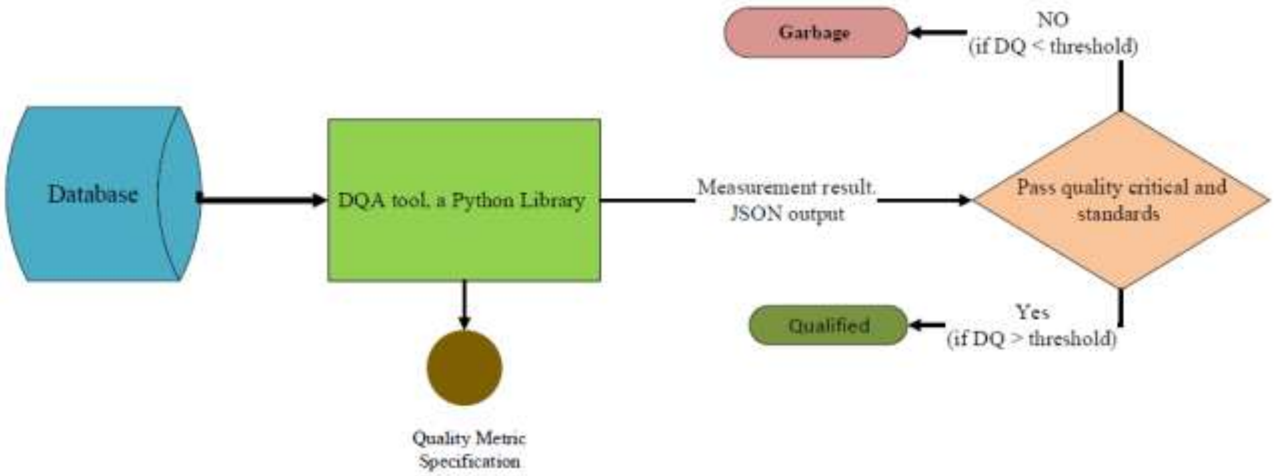

Figure 4.8. Workflow of the Data Quality Assessment (DQA) Tool for Structured Agricultural Data. The diagram illustrates the data quality assessment process using the DQA Python library. Input data from the database is evaluated against a set of predefined quality metric specifications such as accuracy, completeness, consistency, and credibility. The tool produces a JSON output containing data quality score, which is compared against a threshold to determine whether the data is qualified for further use or marked as garbage due to insufficient quality.

4.3.4. Data Processing and Harmonization

One challenge in combining these datasets is the difference in spatial and temporal sampling frequencies between telemetry data and satellite imagery. For example, the sampling frequency of harvest monitoring data is typically higher than that of the nitrogen application data, as shown in Figure 4.9. This discrepancy requires additional geoprocessing to harmonize the datasets into a unified spatial resolution suitable for analysis. Additionally, artifacts in the harvest data (e.g., outlier values or missing data points) must be addressed through data cleaning and preprocessing techniques.

Figure 4.9. Visualization of Field operations sampling frequency.

Figure 4.9 contrasts high-frequency yield data with lower-resolution nitrogen application data, illustrating the need for geoprocessing to harmonize datasets for optimal nitrogen use in crop yield analysis. These are two vector map layers illustrating field operations with differing sampling frequencies. On the left, the harvest monitoring data displays harvested grain yield (ton/ha) at a high sampling frequency, resulting in a detailed and granular map. On the right, the nitrogen (N) application data shows as-applied nitrogen rates (kg/ha), but with a lower spatial resolution due to less frequent sampling compared to the harvesting operation. This difference in sampling frequency demonstrates the need for geoprocessing to harmonize datasets when analyzing yield and input application rates to determine optimal nitrogen use for improving crop yield.

The harmonization process involves smoothing the data, removing outliers, and interpolating missing values to align telemetry and satellite data both spatially and temporally. By converting these datasets into a unified format, the system supports robust map-layer analysis for EONR computation, enabling precise nitrogen recommendations based on the relationship between applied nitrogen and observed crop performance.

After the outlier detection and cleaning process, data normalization is performed to harmonize spatial resolutions across datasets. Telemetry data, such as N application rates and harvest observations, are often captured at different frequencies and spatial resolutions. The telemetry data is then interpolated to a 5-meter spatial resolution using Kriging interpolation which uses variogram to model spatial correlation and predict unknown values. In small fields where heterogeneity is minimal, this interpolation method produces a relatively homogeneous dataset. In larger fields, interpolation helps preserve spatial variability by distributing sensor readings more accurately across neighboring areas. The average field size used in this study was 10 which can be considered for interpolation. For nitrogen application data, the N rate is recorded at individual GPS locations but often reflects application over a broader area, influenced by machine spread patterns. To recover the true distribution of nitrogen, a variogram-based interpolation technique estimates the spread of nitrogen over the surrounding locations. This ensures that the final dataset accurately represents the application areas and avoids overestimating or underestimating EONR. The interpolation process maintains the consistency of descriptive statistics, ensuring that the variability in the data is preserved while adapting it for further spatial analysis and modeling.

Data harmonization is essential for integrating cleaned telemetry data with satellite imagery. Satellite-derived vegetation indices are typically acquired at 10 m resolution which then down sample to higher resolution (5 m) provide insights into crop health across different points in time. The cleaned telemetry data is harmonized with these indices by aligning spatial resolutions and time intervals, allowing for consistent analysis across multiple datasets. The harmonized data enables robust map-layer analysis, ensuring that the nitrogen application rates and observed crop health indices are directly comparable.

4.4. Modeling, EONR Computation and Workflow Orchestration

A RSM framework was developed in chapter 2 as the core optimization algorithm to estimate the EONR. As shown in Figure 4.1, the RSM computation is handled via AWS Batch, which utilizes data from Postgres database and Amazon S3 to generate precise nitrogen prescription maps tailored to individual fields. The model incorporates a comprehensive dataset consisting of historical telemetry data and vegetation indices derived from satellite imagery retrieved from RDS Postgres, allowing it to account for both temporal and spatial variability in crop response since 3 different dates are used before the actual N application. See Figure 4.10.

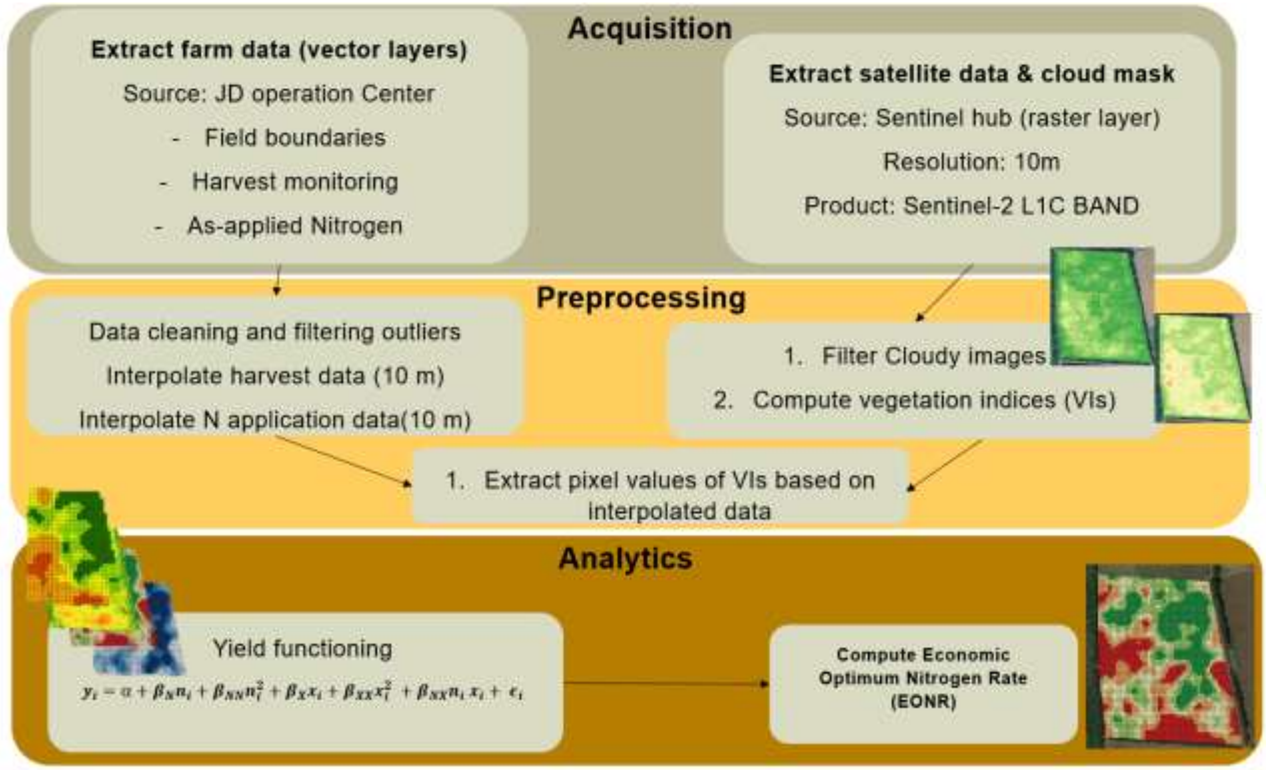

Figure 4.10. Process workflow for computing EONR

The workflow computes EONR by integrating and preprocessing farm and satellite data, then using analytics (RSM) to model yield and determine EONR. The steps for computing EONR through data acquisition, preprocessing, and analytics for optimizing nitrogen applications. It integrates farm data from John Deere Operation Center and satellite data from Sentinel hub, which undergo preprocessing to clean and filter data, interpolate key metrics, and compute vegetation indices. Finally, the analytics stage uses these data to model yield functioning and compute the EONR.

The entire workflow, from data extraction to EONR computation, is orchestrated by AWS Step Functions. This service coordinates the sequential execution of tasks, including initiating ETL processes for farm data and satellite imagery, storing the data, and running the RSM in AWS Batch. When an EONR computation is initiated either through user input via a RESTful API or as part of a scheduled process the Step Functions workflow progresses through its defined states.

It first runs the batch jobs to extract and process telemetry and satellite data, ensuring that this data is properly integrated and loaded into the RDS Postgres database. The final state of the workflow (State 2) then triggers the RSM execution in a separate AWS Batch job, using the integrated datasets to estimate the optimal nitrogen rate.

RESTful APIs plays a critical role in user interaction, enabling users to input key parameters such as field boundary, crop type, and season through a simple web interface. These APIs pass the user inputs to the AWS Step Functions workflow, which then orchestrates the entire analytical pipeline, including data extraction, processing, and eventually, the RSM-based EONR computation. By managing this input and triggering the appropriate processes, the APIs and Step functions ensure that the system dynamically adapts to varying user requirements and data conditions.

AWS Batch dynamically allocates computing resources to run the RSM efficiently, regardless of the size and complexity of the dataset. The containerized environment ensures consistent model execution across different farms and regions, enabling reliable results. Upon completion, the model outputs a nitrogen prescription map stored in S3, which is then made accessible through APIs for real-time use in agricultural equipment. This orchestration, facilitated by Step Functions and RESTful APIs, provides a streamlined, scalable solution for end-to-end EONR estimation in precision agriculture. To enable the computation of EONR the core system architecture is presented in Figure 4.11.

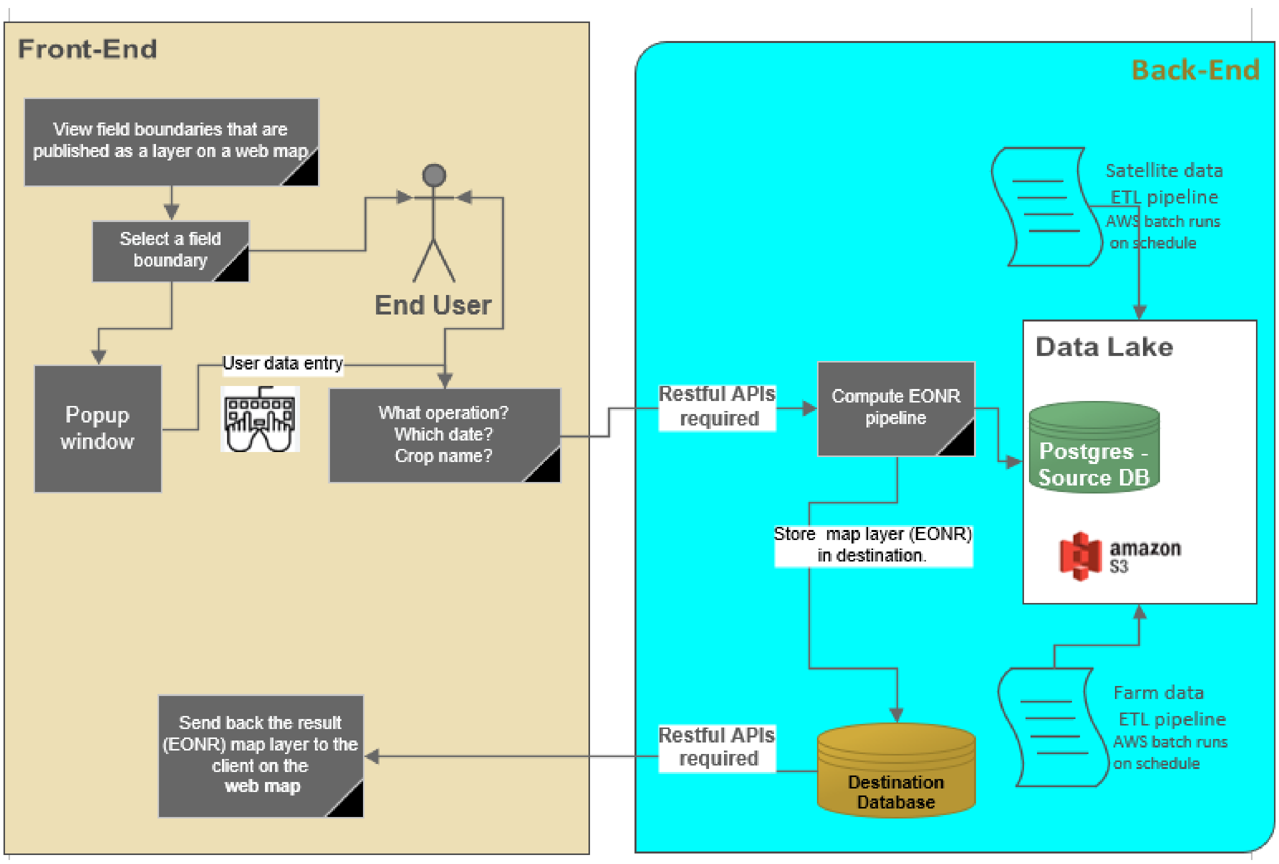

Figure 4.11. lustration of the system architecture for computing EONR.

The system architecture demonstrates where user access a web map to view farm field boundaries, input data, and receive EONR map layers; Amazon S3 stores raw data, while Postgres (RDS) holds processed, structured data for quick retrieval and analysis, ultimately displaying results on the front-end.

The system architecture diagram presented in Figure 4.11 is for computing EONR. This diagram outlines a system architecture where the front-end allows users to view field boundaries for a specific farm on a web map, input relevant details, and receive calculated results (EONR map layers). The back-end consists of two key storage components: Amazon S3 acts as the Data Lake, responsible for storing large volumes of raw, unstructured, and semi-structured data (e.g., satellite imagery, farm data). It serves as the primary storage for incoming datasets that require further processing. Postgres (RDS) is a relational database used for storing structured and processed data, such as processed farm and satellite data and associated metadata. Once data has been cleaned, processed, and structured, it is stored here for fast retrieval and analysis. The result of EONR stored in the destination database and sent back to the front-end for user visualization.

4.4.1. Analytics and algorithms

The foundation of the proposed framework was proposed in our original study about EONR estimation in second chapter. This study introduced an efficient optimization framework for determining EONR. The framework utilized RSM to combine multiple data sources for real-time and prescription-based EONR applications. To conceptualize a scalable cloud computing solution, this integrates a larger number of crop fields (Figure 4.12) to enable a robust and accurate generalization of RSM framework for estimating EONR across many different agricultural fields.

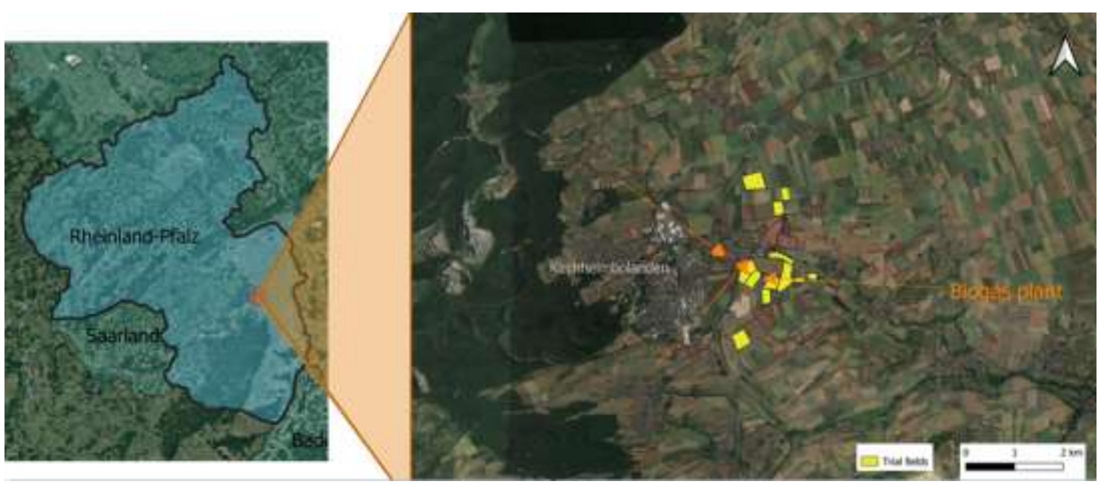

Figure 4.12. Illustration of the experimental fields for this study.

A total of 14 fields were selected for the experimentation, and all the fields were cultivated with the same variety of winter wheat crops. The farm is in Central Europe, southwest of Germany.

The RSM provides a flexible yet powerful optimization strategy for determining EONR by modeling the interaction between nitrogen application rates and crop performance at each location.

RSM for EONR Estimation

The RSM framework can be described by the following model, which relates harvester yield observations to applied nitrogen rates and VIs data:

$$y_i = \beta_0 + \beta_N n_i + \beta_{NN} n_i^2 + \beta_X x_i + \beta_{XX} x_i^2 + \beta_{NX} n_i x_i + \epsilon_i \quad (1),$$

where:

- y_i represents the harvested yield (kg/ha) at location i,
- n_i is the applied nitrogen rate (kg/ha) as recommended by on-machine crop sensors at location i,
- x_i is the value of the selected multispectral index (MSI) derived from satellite imagery for location i,
- β_0 is the intercept, and the remaining β terms represent the linear, quadratic, and interaction coefficients for nitrogen and the MSI,
- ϵ_i is the error term, assumed to be normally distributed with zero mean and constant variance.

This RSM captures the nonlinear relationship between nitrogen applications, crop yield, and spectral indices. The quadratic terms allow the model to account for diminishing returns at high nitrogen application rates, while the interaction term β_{NX} captures how the nitrogen response varies with crop health as indicated by the MSI.

To generalize the RSM framework across multiple fields and improve its scalability, we developed a weighted average model that adapts the base RSM to different agricultural fields. Instead of fitting a pooled model across all fields, individual RSMs were fit for each field, and a weighted average approach was introduced to integrate these models into a unified predictive framework. This method enables more accurate EONR estimates across diverse fields with varying conditions. Hence the computation of weighted average model is given as follows:

$$\theta_{weighted} = \sum_{m=1}^{K} w_m \theta_m, \quad (2),$$

where, $\theta_{weighted}$ represents the weighted average of the coefficients form multiple models. Each model, indexed by m, had an associated coefficient vector θ_m and a corresponding weight w_m. The summation over $m = 1$ to K (where K denotes the number of models) aggregates the weighted contributions of each model's coefficient. Thus, $\theta_{weighted}$ reflects a combined estimation that accounts for the relative importance (weight) of each model's contribution

This weighted average approach ensures that the model captures field-specific variations while maintaining generalizability across multiple fields. Each field contributes to the final model in proportion to its relevance, resulting in a more adaptable and accurate EONR estimation framework.

Equation (2) describes how the average parameter estimate is calculated across multiple models. Since all individual models share the same set of parameter estimates—specifically, the six candidate MSIs (Table 4.2), as-applied N, their quadratic terms, and interactions—but differ only in the datasets, formula (2) demonstrates the computation of the average model for a single parameter estimate for simplicity. This approach applies uniformly across all model parameter estimates.

The model was deployed as a was the batch processing module that was integrated into the overall application architecture (see Figure 4.1). This enabled a real time computation of EONR whenever a user triggers the application to generate an EONR map

4.4.2. Integration into Scalable Cloud Computing

RSM-based EONR computation is deployed in a scalable AWS Batch processing module, which integrates seamlessly into the overall cloud architecture (Figure 4.1). The model processes the telemetry and spectral data stored in Amazon S3 and RDS Postgres databases, allowing for real-time EONR calculation whenever a user triggers the application. Users can input specific field boundary, crop types, and other parameters through an API, and the AWS Batch job computes the optimal nitrogen prescription based on the weighted average RSM. This cloud-native architecture ensures that the computation scales efficiently in response to increased data volumes, whether from new farms or additional telemetry and satellite imagery. By leveraging containerized processing in AWS Batch, the system can handle large datasets and perform EONR calculations in real-time, offering precision farming insights to farmers across diverse geographical regions. The integration of a weighted average model further enhances the system's ability to generalize across multiple fields, ensuring that the solution remains adaptable and scalable for various regions.

4.5. Discussion

This study underscores the transformative potential of integrating cloud computing into precision agriculture, specifically focusing on the estimation of the EONR. By designing a scalable architecture leveraging cloud-native technologies, the framework bridges the gap between large-scale datasets and actionable insights. The system integrates telemetry data from agricultural machinery with satellite-derived multispectral indices (MSIs) to generate nitrogen prescription maps that enhance crop productivity and environmental sustainability.

Cloud computing and IoT have emerged as pivotal technologies in advancing precision agriculture. Existing studies have demonstrated their potential in improving scalability, resource efficiency, and decision-making. For instance, Bouali Et-Taibi et al. (2024) presented a cloud-based IoT-enabled smart irrigation system that optimizes water usage through real-time sensor data and weather predictions, addressing challenges such as water scarcity and climate variability. Similarly, Shafiya Banu et al. (2024) introduced a cloud-based nutrient management system integrating IoT, CNNs, and real-time soil sensors to deliver site-specific fertilizer recommendations. These studies highlight the critical role of cloud-enabled systems in addressing environmental challenges and optimizing resource allocation.

The modular and scalable architecture proposed in this study advances these existing frameworks. Utilizing AWS Batch for dynamic scaling, Amazon S3 for robust storage, and RDS Postgres for structured data management, the system ensures seamless automation from data ingestion to EONR computation. These components align with the scalable infrastructures discussed by Kalyani et al. (2024), who emphasized cloud–fog–edge architectures for handling data heterogeneity and supporting real-time decision-making. Unlike the irrigation-focused system by Bouali Et-Taibi et al. (2024), the framework in this study targets nitrogen management, addressing challenges related to spatial variability and field-level nutrient heterogeneity.

Data quality and preprocessing are foundational to precision agriculture systems, as noted in previous literature. Bouali Et-Taibi et al. (2024) stressed the importance of big data preprocessing, integrating sensor data with weather forecasts for accurate irrigation management. Similarly, Shafiya Banu et al. (2024) employed techniques such as data augmentation and noise filtering to improve CNN model accuracy in nutrient management. This study builds on these principles by incorporating advanced techniques like Kriging interpolation and outlier detection, ensuring high-quality nitrogen recommendations that capture the interplay between crop health, environmental factors, and nitrogen application.

The integration of MSIs with nitrogen application data marks a significant methodological advancement. By addressing sub-field-level variability, the system provides granular nitrogen recommendations, complementing the findings of Shafiya Banu et al. (2024), who utilized IoT-enabled soil sensors for site-specific nutrient optimization. Furthermore, this study employs weighted average RSMs enhance adaptability across diverse farming environments, addressing the scalability concerns highlighted by Bouali Et-Taibi et al. (2024) and aligning with Kalyani et al. (2024), who advocated for advanced modeling techniques in digital twin systems.

Scalability and interoperability remain persistent challenges in precision agriculture. Bouali Et-Taibi et al. (2024) noted the need for standardized communication protocols to aggregate diverse data sources effectively. This study addresses similar concerns by leveraging predictive auto-scaling and proposing future integration of additional data sources, such as soil sensors and weather data, to enhance system precision and adaptability. These efforts align with Kalyani et al. (2024), who emphasized the importance of integrating multiple data streams for real-time agricultural decision-making.

The practical implications of this framework extend to delivering actionable, real-time nitrogen recommendations. By enabling data-driven decisions, the system optimizes resource use and enhances sustainability, paralleling the goals of Bouali Et-Taibi et al. (2024) and Shafiya Banu et al. (2024). User-friendly interfaces bridge the gap between complex data analytics and farmer decision-making, empowering stakeholders to implement site-specific recommendations effectively.

In conclusion, this study contributes to the growing body of literature on cloud-enabled precision agriculture by integrating advanced preprocessing, scalable architectures, and innovative modeling techniques. By addressing challenges such as data heterogeneity, spatial variability, and scalability, it builds upon foundational work by Bouali Et-Taibi et al. (2024), Shafiya Banu et al. (2024), and Kalyani et al. (2024). Future research should validate this framework across diverse environmental and agricultural conditions and explore its integration with broader precision agriculture practices, such as irrigation and nutrient deficiency risk management, to support the transition toward sustainable, resilient, and data-driven food systems.

4.6. Conclusion

This paper presents a scalable cloud computing framework designed to optimize nitrogen management through the estimation of EONR in precision agriculture. By integrating agronomic telemetry data from machinery and satellite-derived multispectral imagery, the framework addresses key challenges of processing large geospatial datasets, enabling real-time, data-driven decision-making to improve nitrogen use efficiency and sustainability.

The proposed architecture leverages AWS services such as Batch for scalable job orchestration, S3 for efficient data storage, and RDS Postgres for structured data management and analytics. This modular, cloud-native design ensures flexibility, scalability, and cost-effectiveness while maintaining compliance with regional data protection regulations. The integration of historical telemetry data with real-time satellite imagery enables the generation of accurate, site-specific nitrogen prescription maps, which enhance both crop productivity and environmental sustainability.

Key contributions of the framework include its capacity to dynamically scale with increasing data volumes and its region-specific data storage strategy, which ensures efficient data retrieval and processing. Automated data archiving further enhances long-term storage efficiency while minimizing costs, making the framework suitable for large-scale precision farming operations. The case study results validate its effectiveness in generating timely and accurate nitrogen recommendations, resulting in improved crop yields and reduced environmental impact.

This framework advances the state of precision agriculture by addressing challenges related to data heterogeneity, spatial variability, and real-time analytics. While it currently demonstrates significant improvements in scalability and resource optimization, future work could focus on integrating additional data sources such as soil sensors and weather data to further enhance precision. Extending the framework to support a broader range of crops and farming practices could also unlock new opportunities for optimizing decision-making across diverse agricultural systems.

In summary, this study contributes a robust, cloud-enabled solution for modern farming that supports sustainable nitrogen management. By promoting resource efficiency and environmental stewardship, the framework aligns with the goals of sustainable and resilient food systems, offering a pathway for advancing precision agriculture on a global scale.

Chapter 5: General Discussion

This chapter synthesizes the multidisciplinary approaches undertaken in this dissertation to realize scalable and sustainable agricultural practices, particularly in nitrogen management. Leveraging Knowledge Domains Mapping (KDM), the work integrates insights from agronomy, data science, remote sensing, and cloud-based computing to address the critical challenges of Economic Optimum Nitrogen Rate (EONR) estimation. The discussion spans the modeling of EONR through on-farm experimentation, evaluating statistical and machine learning (ML) methodologies, and proposing a robust, data-driven cloud architecture to enable real-time, on-demand experimentation. By bridging the gap between science and technology, this dissertation aims to advance regenerative agricultural practices, offering solutions that optimize nitrogen use efficiency (NUE), reduce environmental impact, and enhance ecosystem sustainability.

The study aligns with recent literature emphasizing the transformative potential of advanced AI and real-time data integration in agriculture. For instance, Ravikumar et al. (2024) demonstrated that real-time nitrogen monitoring using IoT devices and machine learning algorithms significantly reduces nitrogen losses while maintaining or increasing crop yields, a finding echoed in this work. Furthermore, adaptive AI applications in precision agriculture (Akintuyi, 2024) highlight the importance of integrating AI with IoT for resource optimization, cost savings, and sustainability—concepts central to the cloud architecture proposed in this dissertation. These advancements underscore the need for accessible, scalable AI solutions and collaborative frameworks to fully realize the benefits of technology in promoting regenerative agriculture.

5.1 Integrative Framework for Optimized Nitrogen Management in Regenerative Agriculture

This research presents an integrated, data-driven framework for nitrogen management designed to align with the principles of regenerative agriculture (RA). RA emphasizes restoring ecosystem health, improving soil and nutrient cycles, and enhancing climate resilience through sustainable farming practices (Lal, 2020; Giller et al., 2021). One of the most critical components of RA is nitrogen optimization, which directly impacts soil health, crop productivity, and environmental sustainability. By targeting the EONR with field-specific precision, the proposed framework contributes to advancing RA by enhancing nitrogen use efficiency (NUE) while minimizing the ecological footprint.

Effective nitrogen management is essential to address the overuse of synthetic fertilizers, a significant contributor to environmental degradation. Excessive nitrogen application leads to nitrate leaching, greenhouse gas emissions, and soil acidification, all of which undermine RA's goals (Lyu et al., 2024; Manzeke et al., 2023). Addressing these issues requires innovative, site-specific strategies for nitrogen optimization. For instance, Ravikumar et al. (2024) demonstrated that in-season nitrogen optimization using real-time sensing technologies significantly reduces nitrogen losses while maintaining crop yields. Building on such approaches, this framework integrates telemetry data from agricultural machinery with multispectral satellite imagery, enabling dynamic, site-specific EONR estimations. This combination facilitates precision agriculture practices that enhance NUE while supporting ecological balance. Robertson and Vitousek (2009) highlighted the importance of site-specific nitrogen management for optimizing resource use and maintaining soil fertility, objectives directly aligned with this study.

The framework employs Response Surface Modeling (RSM) to calculate optimal nitrogen rates based on real-time field data and spectral indices, such as NDVI and CARI. This approach enables farmers to tailor nitrogen applications to their crops' specific needs and environmental conditions, reducing waste and promoting sustainable farming practices. By aligning nitrogen application with the nuanced requirements of individual fields, the system ensures that RA’s principles of resource efficiency and environmental stewardship are met.

The integration of cloud computing technologies enhances the scalability, adaptability, and regulatory compliance of this framework, further supporting RA principles. Modern precision agriculture generates massive volumes of data from diverse sources, including sensors, remote sensing, and farm management systems. Traditional infrastructures struggle to process such data efficiently. Cloud computing addresses these challenges by providing scalable, high-throughput platforms for data integration and analysis. The framework uses Amazon S3 for unstructured telemetry and satellite data storage, ensuring efficient data retrieval and harmonization, while RDS Postgres manages structured data like geospatial information and metadata, supporting regulatory compliance with standards such as GDPR (Bakare et al., 2024). This seamless processing of telemetry and satellite data enables real-time decision-making.

Cloud computing's potential for sustainable agriculture is well-documented. Saravanan et al. (2023) and Wei et al. (2023) highlighted how cloud platforms empower farmers by enabling resource optimization and operational efficiency. Saravanan et al. (2023) also explored integrating technologies such as the Internet of Things (IoT), Ensemble Learning, and Explainable AI (XAI) within a Reinforcement Learning framework, demonstrating how these tools can enhance precision agriculture by increasing output and reducing resource consumption. By leveraging Amazon Web Services (AWS) Batch for containerized processing, the proposed framework dynamically allocates computational resources based on workload, ensuring cost efficiency and responsiveness. This scalability is critical for supporting the global adoption of RA practices across diverse environmental and regulatory contexts.

Adaptive AI technologies further complement the framework by introducing self-learning algorithms that adapt over time, improving accuracy and responsiveness in EONR predictions. These systems adjust to changing environmental conditions and crop responses, enhancing the precision of nitrogen management strategies. Akintuyi (2024) reviewed the transformative role of adaptive AI in precision agriculture, underscoring its ability to optimize farm operations through real-time data analysis and cloud computing.

By combining advanced analytics, cloud computing, and adaptive AI, the proposed framework demonstrates how cutting-edge technologies can operationalize RA principles. Nitrogen optimization, a cornerstone of regenerative practices, is achieved through field-specific precision and real-time responsiveness, reducing environmental harm and enhancing resource efficiency. The scalability, regulatory compliance, and adaptability of the framework position it as a robust tool for modern agricultural transformation.

This integrative approach not only addresses the challenges of nitrogen management but also aligns with RA's broader goals by promoting sustainable and climate-resilient farming systems. Future enhancements could include integrating additional data sources, such as soil sensors and weather forecasts, and adopting blockchain for data traceability. These advancements would further enhance the framework's utility and scalability, driving a global transition toward regenerative agriculture.

5.2 Methodological Approach: Response Surface Modeling vs. Machine Learning in EONR Estimation

Conventional statistical approaches, such as quadratic, linear-plateau, and quadratic-plateau models, have been extensively applied to estimate the EONR by modeling the relationship between nitrogen input and crop yield. These models are celebrated for their simplicity and ease of interpretation, providing actionable insights for farmers and decision-makers. However, they often falter in scenarios involving spatial variability and heterogeneous field conditions due to their rigid functional assumptions and limited capacity to incorporate multiple environmental and management factors. Addressing these limitations, recent studies (Zhang and Brorsen, 2024; Tanaka et al., 2024) have explored more advanced statistical approaches, such as multi-degree splines, and machine learning (ML) models.

Statistical Approaches in Nitrogen Modeling

Zhang and Brorsen (2024) proposed a multi-degree spline approach for EONR estimation, achieving an RMSE of 12.5% in yield modeling, highlighting its flexibility and scalability compared to traditional functions like quadratic models. This approach strikes a balance between accuracy and interpretability, addressing the complexity of site-specific nitrogen management.

Similarly, Colaço et al. (2024) assessed digital strategies for nitrogen management using multi-method evaluations on on-farm experimentation data. They emphasized the value of combining spatial statistical modeling techniques with site-specific data, reporting high accuracy in identifying nitrogen response zones while ensuring practical nitrogen rate recommendations.

de Lara et al. (2023) explored the application of ML methods in predicting site-specific EONR using on-farm precision experimentation data. Their results demonstrated the potential of ML in handling large datasets and spatial variability but highlighted inconsistencies across models and datasets, with RMSE values ranging between 15% and 28% depending on the algorithm.

Machine Learning Approaches

Machine learning methods, such as Random Forest (RF), XGBoost, Support Vector Regression (SVR), and Artificial Neural Networks (ANN), have been widely investigated for nitrogen modeling. Tanaka et al. (2024) demonstrated that while ML models can achieve high accuracy in yield prediction, with RMSE values ranging between 14.2% and 31.5%, their sensitivity to covariate selection and algorithm choice introduces significant uncertainty in EONR recommendations. Similarly, Kakimoto et al. (2022) explored causal forest models in nitrogen management, noting their ability to uncover spatially variable relationships but also highlighting challenges in interpretability and the reliance on large datasets.

Ennaji et al. (2023) provided a comprehensive review of ML applications in nutrient management, emphasizing their advantages in capturing complex interactions between inputs and environmental factors. However, they also highlighted ML's limitations in interpretability and the risk of overfitting, which can result in unreliable fertilizer recommendations.

Comparative Insights and Advancing RSM

While ML models provide a data-driven pathway for handling large, complex datasets, their limitations in interpretability, consistency, and scalability often hinder their practical application for EONR estimation. By comparison, our Response Surface Model (RSM) achieved a comparable RMSE of 14.7% in modeling yield responses while maintaining a reliable range for EONR estimation. Unlike ML methods, our approach provided consistent nitrogen rate recommendations with reduced variability and practical interpretability.

Moreover, RSM demonstrated scalability across five winter wheat fields, confirming its robustness under diverse field conditions. A key advantage of the RSM approach lies in its capacity for retrospective analysis, enabling nitrogen management strategies to be refined for subsequent growing seasons. This capability not only enhances decision-making but also supports the integration of historical and current field data, further strengthening its utility in sustainable and scalable nitrogen management systems.

By combining the strengths of traditional statistical modeling with the adaptability required for precision agriculture, RSM emerges as a reliable, interpretable, and scalable tool for EONR estimation, outperforming more complex and less reliable ML approaches in practical agricultural applications.

5.3 Model Validation: Average Model vs. Pooled Model for EONR Estimation

Effective nitrogen (N) management practices are pivotal for optimizing the EONR, as they must weigh heavily on key factors such as genotype (G), environment (E), and management (M). Recent studies indicate that the environment accounts for 41% of EONR variability, making it the main drive, while management and genetics contribute 31% and 27%, respectively. This underscores the importance of adaptable N management strategies that consider these factors to improve nitrogen use efficiency and mitigate the impacts of climate change on crop production.

The generalizability and scalability of models for estimating EONR are critical to their successful application across diverse environmental and field conditions. This section explores the effectiveness of data-driven strategies using RSMs, introduced in Chapter 2, for in-season EONR estimation in winter wheat fields. A detailed comparative analysis highlights the strengths and limitations of a single-model approach versus a model-averaging technique (our data driven modeling approach) under varied winter wheat field, showcasing the benefits of leveraging weighted averaging for enhanced prediction accuracy and reliability.

Key Differences Between Pooled Model and Average Model

1. Model Structure

- **Pooled Model**: This approach fits a single RSM across four fields and validates it on a fifth, assuming field uniformity. While efficient, this assumption restricts the model's ability to capture the inherent variability in spatial and environmental conditions.
- **Average Model**: By contrast, the average model independently fits RSMs to each field and integrates these using a weighted averaging technique. This approach accounts for field-specific characteristics, creating a unified framework that reflects the diversity of agricultural settings.

2. Accuracy

- The **average model** significantly outperformed the Pooled Model, achieving a root mean square error (RMSE) of 11 kg N/ha compared to 19 kg N/ha for the Pooled Model.
- The **average model** provided reliable EONR estimates ranging from 59–85 kg N/ha, closely aligning with practical field management practices. Conversely, the Pooled Model's estimates included implausible values, such as negative EONR rates (-0.52 to 91 kg N/ha), reflecting its inability to generalize effectively.

3. Performance and Practicality

- The **average model** exhibited greater robustness and reliability, capturing field-level variability and delivering EONR recommendations consistent with farmer practices. Its predictions better accounted for the heterogeneity in soil, crop, and environmental interactions.
- The **pooled model**, with a broader prediction range and lower accuracy, struggled to generalize across diverse conditions, leading to increased risks of under- or over-fertilization.

The findings align with previous research emphasizing the importance of model adaptability and localized calibration. For instance, Matavel et al. (2024) demonstrated the efficacy of model-averaging in accurately estimating ex-post EONR, showcasing its superiority for managing field variability. Similarly, Mousavi (2024) proposed a localized model selection framework to refine input rate predictions, and Hegedus (2022) highlighted the necessity of site-specific nitrogen recommendations to balance profitability and environmental impact.

In addition to methodological advancements, the significant contributions of genotype (G) × management (M) factors to EONR variability underscore opportunities for targeted interventions. Factors such as interannual weather variability, crop radiation use efficiency, and soil inorganic nitrogen carryover from previous seasons play pivotal roles. Management strategies, including optimized planting dates and densities, can influence yield across all nitrogen rates. Importantly, with 59% of EONR variability being manageable through genetic and management strategies, the potential to enhance nitrogen use efficiency and buffer climate change impacts is evident.

This analysis underscores the superiority of model-averaging techniques for providing scalable, accurate, and field-relevant nitrogen management recommendations. By accounting for field-specific variability and integrating key influencing factors, the average model ensures practical applicability and aligns with sustainable nitrogen management goals. Implementing G × M strategies offers a pathway to optimize productivity while mitigating environmental impacts, supporting targeted N fertilizer recommendations and informed EONR decision-making in heterogeneous agricultural systems.

5.4 Field Trial Design: Site-Specific vs. Randomized Block/Plot Design

On-Farm Experimentation (OFE) is revolutionizing agricultural research by embedding scientific trials directly within farmers' fields. This participatory approach integrates farmers into the experimentation process, enabling real-world, field-scale studies that address the challenges of modern agriculture. By aligning scientific methodologies with farmers' local knowledge and management practices, OFE fosters a collaborative framework for co-learning and innovation. As Lacoste et al. (2021) emphasize, OFE departs from conventional small-plot trials to focus on demand-driven research that delivers actionable, context-specific insights. Although classical experimental designs such as RCBD, Latin squares, and plot-based trials provide strong statistical control, their effectiveness decreases in real-world agricultural settings where spatial heterogeneity dominates. As Piepho et al. (2011) emphasize, failing to account for spatial correlation in field trials leads to inflated residual variance and reduced power to detect treatment effects. Their work highlights that agricultural experiments, especially those conducted at field scale require designs and analytical methods that explicitly model spatial structure. This principle is particularly relevant for OFE, where operational constraints often limit randomization and replication, making spatial models essential for extracting reliable treatment estimates from heterogeneous fields.

Digital technologies, such as sensors, yield mapping, and satellite imagery, empower OFE by enabling precise data collection and analysis, which supports decision-making tailored to the spatial and temporal variability of fields.

In advancing this paradigm, site-specific field trials have emerged as a key methodology, where treatments are applied across entire fields divided into management zones or fine spatial grids. This design captures high-resolution data using IoT sensors, satellite-based multispectral indices (MSIs), and precision machinery, effectively mapping the spatial variability within fields. Studies like those by Li et al. (2023) and Schulz et al. (2015) demonstrate the value of MSIs in reducing random effects and enhancing model stability, making site-specific trials reliable tools for estimating EONR. These trials deliver granular data, enabling tailored recommendations that align treatment strategies with specific field conditions.

In contrast, traditional designs such as Randomized Complete Block Design (RCBD), stripe, and plot trials, while scientifically rigorous, often fail to capture the heterogeneity of real-world fields. Their resource-intensive setup and reliance on controlled environments limit scalability and applicability for on-farm decision-making (Kakimoto et al., 2022). Small-plot trials, in particular, struggle to account for spatial variability, thereby diminishing their relevance in contemporary agricultural research, which demands adaptability to variable conditions.

To address these limitations, OFE and site-specific trials have gained prominence. To address these limitations, Modern OFE and site-specific trial frameworks inherently align with the design philosophy advocated by Piepho et al. (2011), who showed that spatially informed regression and mixed-effects models can compensate for design limitations and enhance inference quality. By integrating spatial modeling with large numbers of geo-referenced observations generated through yield monitors, MSIs, and machine telemetry, OFE trials achieve higher statistical efficiency than traditional designs. This approach ensures that treatment effects are estimated with improved precision, even under restricted randomization or unreplicated treatment layouts, reinforcing the suitability of OFE methodologies for contemporary precision agriculture.

For instance, Cho et al. (2021) proposed single-strip treatment trials as a statistically robust and practical alternative to block or plot designs. This approach simplifies implementation while effectively capturing field variability. It aligns with the evolution toward whole-field On-Farm Precision Experimentation (OFPE), where precision tools assess treatment effects across heterogeneous field conditions. OFPE replaces traditional stripe trials by providing more detailed spatial data and enhancing the statistical detection of heterogeneous effects, thus supporting more accurate modeling of crop responses.

The advantages of site-specific trials extend far beyond statistical robustness. By integrating tools like MSIs and IoT sensors, these trials illuminate spatial variability and enable real-time monitoring of key variables such as soil moisture, crop health, and yield. Research by Mousavi (2024) and Hegedus (2022) highlights the actionable insights gained from these trials, which improve farm profitability and sustainability, bridging the gap between experimental research and practical application.

This transition from traditional designs to site-specific trials signifies a paradigm shift in agricultural research. Precision agriculture technologies enable these trials to overcome the limitations of traditional approaches, offering richer datasets, scalable designs, and actionable insights applicable to real-world farming practices. Innovations like single-strip treatment trials (Cho et al., 2021), and Liu and Vanhatalo (2020) proposed a Bayesian rejection sampling design to improve spatiotemporal survey efficiency for log-Gaussian Cox processes. Such studies ensure both immediate and long-term value from on-farm experiments. By leveraging these advancements, site-specific trials align with the broader goals of precision agriculture—driving innovation, profitability, and sustainability. They empower researchers and farmers to navigate the complexities of modern agriculture, ultimately shaping the future of agricultural research and practice.

5.5 Practical Implementation and Real-World Applications of Scalable EONR Solutions

The integration of cloud computing and precision agriculture technologies offers a scalable solution for optimizing EONR, addressing both practical implementation and real-world nitrogen management strategies. By deploying the RSM-based EONR model within a cloud-native framework high-volume data processing is enabled, supporting real-time calculations on demand. Rajasekaran et al. (2024) highlight how cloud computing has transformed smart farming by providing real-time data collection, analysis, and storage, enabling remote monitoring, crop management, and resource optimization while addressing challenges such as data security. Saravanan et al. (2023) emphasize how cloud computing, coupled with data mining algorithms, enhances agricultural advancements by enabling efficient processing of large datasets and improving resource management, thus ensuring the scalability required for diverse field conditions. Similarly, Imalka et al. (2022) demonstrate the practical application of technology-driven solutions through the Ceylon E-Agro mobile application, which leverages IoT and cloud-based tools to provide real-time, scalable support for agricultural decision-making. Together, these studies illustrate how cloud computing and IoT technologies bridge the gap between data-driven insights and field-level applications, enabling adaptive, scalable solutions in varying agricultural contexts. In chapter four we proposed a scalable cloud computing framework tailored to EONR estimation, integrating agronomic telemetry data and multispectral satellite imagery for efficient processing of large datasets. Leveraging AWS services like Batch, S3, and RDS Postgres, the system harmonizes telemetry and satellite inputs using geospatial techniques such as Kriging interpolation, generating accurate nitrogen prescription maps that improve crop productivity and environmental sustainability. RSM forms the core of the framework, generalizing EONR computation across diverse fields using a weighted average approach. Users can input field-specific parameters through an API, orchestrating real-time calculations based on integrated datasets. The framework's scalability supports diverse geographies and additional data sources like soil sensors and weather data, offering robust adaptability and precision. Predictive auto-scaling during peak server loads ensures operational efficiency, while features like regionalized data storage comply with global data security regulations. By bridging technological advancements with practical applications, the proposed framework enables real-world implementation of sustainable nitrogen management practices in precision agriculture.

5.6. Challenges and Future Directions

The adoption of RSM with integration of historical machine agronomic telemetry data and Multispectral Indices (MSIs) in nitrogen management has demonstrated significant potential for in season estimation of EONR. However, several challenges remain, particularly related to data quality and the inherent variability of environmental conditions. Uncertainties in crop response to nitrogen under diverse weather scenarios persist as a critical limitation. Enhancing the model's accuracy by integrating IoT sensor data and high-frequency satellite imagery could mitigate some of these uncertainties, offering finer temporal and spatial resolutions. Additionally, expanding the model to include factors such as protein content could align nitrogen management practices with economic and quality-based objectives, providing a more comprehensive approach. Practical implementation also faces computational challenges. Gradient-based optimization techniques for large-scale datasets and the inclusion of auto-scaling mechanisms during peak server loads could enhance processing efficiency and scalability.

Exploring blockchain-based traceability system could further improve data reliability and transparency across the supply chain. Moreover, incorporating self-learning algorithms offers an opportunity to develop dynamic, adaptive models that can evolve with changing field conditions.

Engaging farmers in a participatory approach could further refine nitrogen recommendations. Post-season analyses and direct feedback loops would enable continuous model improvement while fostering trust and adoption among stakeholders. This collaborative strategy aligns with sustainable agricultural practices, ensuring that nitrogen management not only maximizes yield but also minimizes environmental impact.

Going forward, these advancements coupled with interdisciplinary integration of agronomic, technological, and socio-economic factors will be critical in addressing current limitations and advancing the scalability and sustainability of nitrogen management practices in modern agriculture.

References

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., & Zaharia, M. (2021). *A View of Cloud Computing*. Communications of the ACM, 53(4), 50–58.

Abdipourchenarestansofla, M., & Piepho, H. P. (2022). Yield observation outlier detection with unsupervised machine learning in harvest machines. *VDI-Berichte, No. 2395.* Online: 79th International Conference on Agricultural. https://doi.org/10.51202/9783181023952-34

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

Agapiou, A., & Lysandrou, V. (2021). Remote sensing archaeology: Tracking and mapping evolution in European journal publications. *Remote Sensing, 13*(2), 224. https://doi.org/10.3390/rs13020224

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

Akintuyi, O. B. (2024). Adaptive AI in precision agriculture: A review: Investigating the use of self-learning algorithms in optimizing farm operations based on real-time data. *Open Access Research Journal of Multidisciplinary Studies, 7*(2), 16–30. https://doi.org/10.53022/oarjms.2024.7.2.0023

Andrade-Sanchez, P., Gorelick, S. M., Zhang, X., & Wang, W. (2019). Enhancing site-specific crop management through integration of high-resolution environmental data and precision agricultural technology. *Precision Agriculture, 20*(3), 601–615. https://doi.org/10.1007/s11119-018-9592-4

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., & Zaharia, M. (2021). A view of cloud computing. *Communications of the ACM, 53*(4), 50–58.

Athearn, K., Yarick, M., & Parks, N. (2021). Knowing your product costs: A primer for farmers and food entrepreneurs. *FE1103, 11/2021," EDIS.* https://doi.org/10.32473/edis-FE1103-2021

Arnall, D. B., Mallarino, A. P., Ruark, M. D., Varvel, G. E., Solie, J. B., Stone, M. L., Mullock, J. L., Taylor, R. K., & Raun, W. R. (2014). Relationship between grain crop yield potential and nitrogen response. *Agronomy Journal 106*(6), 2335. https://doi.org/10.2134/agronj2013.0034

Bakare, S. S., Adeniyi, A. O., Akpuokwe, C. U., & Eneh, N. E. (2024). Data privacy laws and compliance: A comparative review of the EU GDPR and USA regulations. *Computer Science & IT Research, 5*(3), 45–67. https://doi.org/10.51594/csitrj.v5i3.859

Banu, M. S., Abuthahir, A. A. M., Mohandas, R., Rajan, D. A. J., Meenakshi, B., & Mohankumar, N. (2024). Cloud-based intelligent nutrient management system for precision agriculture using CNN. *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS),* 971–976. https://doi.org/10.1109/ICSCSS60660.2024.10624911

Berghuijs, H. N. C., Silva, J. V., Reidsma, P., & de Wit, A. J. W. (2024). Expanding the WOFOST crop model to explore options for sustainable nitrogen management: A study for winter wheat in the Netherlands. *European Journal of Agronomy, 154,* 127099. https://doi.org/10.1016/j.eja.2024.127099

Bouali Et-Taibi, M., Abid, M. R., Boufounas, E.-M., Morchid, A., Bourhnane, S., Abu Hamed, T., & Benhaddou, D. (2024). Enhancing water management in smart agriculture: A cloud and IoT-based smart irrigation system. *Results in Engineering, 22,* 102283. https://doi.org/10.1016/j.rineng.2024.102283

Box, G. E. P., & Draper, N. R. (2007). Response surfaces, mixtures, and ridge analyses. *John Wiley & Sons.* https://doi.org/10.1002/0470072768

Chen, X., Chambers, R. G., Bandaru, V., Jones, C. D., Ochsner, T. E., Nandan, R., Irigireddy, B. C., Lollato, R. P., Witt, T. W., & Rice, C. W. (2024). Mid-season nitrogen management for winter wheat under price and weather uncertainty. *Field Crops Research, 316,* 109509. https://doi.org/10.1016/j.fcr.2024.109509

Cho, J. B., Guinness, J., Kharel, T., Maresma, Á., Czymmek, K. J., van Aardt, J., & Ketterings, Q. M. (2021). Proposed method for statistical analysis of on-farm single strip treatment trials. *Agronomy, 11*(10), 2042. https://doi.org/10.3390/agronomy11102042

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 Geoinformation, 23,* 344–351. https://doi.org/10.1016/j.jag.2012.10.008

Colaço, A. F., et al. (2024). Digital strategies for nitrogen management in grain production systems: Lessons from multi-method assessment using on-farm experimentation. *Precision Agriculture.* Published online January 9, 2024. https://doi.org/10.1007/s11119-024-10128-x

Conant, R. T., Cerri, C. E. P., Osborne, B. B., & Paustian, K. (2017). Grassland management impacts on soil carbon stocks: A new synthesis. *Ecological Applications, 27*(2), 662–668. https://doi.org/10.1002/eap.1473

Cui, Z., Zhang, F., Chen, X., Dou, Z., & Li, J. (2009). In-season nitrogen management strategy for winter wheat: Maximizing yields, minimizing environmental impact in an over-fertilization context. *Field Crops Research, 113*(3), 143–150. https://doi.org/10.1016/j.fcr.2009.12.004

Dash, S., Mohanty, J. R., & Barik, R. K. (2024). Geospatial web services for geospatial cloud computing paradigm: Architectures and selective applications case studies. *2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU),* Bhubaneswar, India, 01–06. https://doi.org/10.1109/IC-CGU58078.2024.10530754

Dash, S., Shukla, A., & Singh, R. (2024). Cloud-based framework for geospatial data analysis in precision agriculture. *Journal of Agricultural Informatics, 15*(2), 45–58.

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 Agriculture, 24,* 1792–1812. 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

Dutta, R., Zhang, Y., & Pathak, A. (2019). Machine learning for precision agriculture: Challenges and future directions. *Computers and Electronics in Agriculture, 165,* 104021. https://doi.org/10.1016/j.compag.2019.104021

Ennaji, O., et al. (2023). Machine learning in nutrient management: A review. *Artificial Intelligence in Agriculture, 9,* 1–11. https://doi.org/10.1016/j.aiia.2023.06.001

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 Sensing, 15,* 1–17. 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 (Eds.), *Practical Mathematics for Precision Farming* (pp. 181–189). American Society of Agronomy. https://doi.org/10.2134/practicalmath2016.0027

Food and Agriculture Organization of the United Nations. (2022). Agricultural production statistics 2000–2021. *FAOSTAT Analytical Brief 60.* Rome. https://doi.org/10.4060/cc3751en

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

Giller, K. E., Hijbeek, R., Andersson, J. A., & Sumberg, J. (2021). Regenerative agriculture: An agronomic perspective. *Outlook on Agriculture.* https://doi.org/10.1177/0030727021998063

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202,* 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Hegedus, P. B. (2022). Optimizing site-specific nitrogen fertilizer management based on maximized profit and minimized pollution (Doctoral dissertation). *Montana State University, College of Agriculture.* Available at: https://scholarworks.montana.edu/handle/1/16982

Huang, Y., & Deng, L. (2024). GeoEkuiper: A cloud-native system for real-time geospatial data processing. *IEEE Transactions on Cloud Computing, 12*(3), 123–135.

Imalka, L. A., Gunawardana, K. G. A., Kodithuwakku, K. M. S. K., Arachchi, H. K. E., Harshanath, S. M. B., & Rajapaksha, S. (2022). Farming through technology-driven solutions for agriculture industry: Ceylon E-Agro mobile application. *2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC),* Hyderabad, India, 306–310. https://doi.org/10.1109/R10-HTC54060.2022.9929335

Inoue, Y., Kira, O., & Soga, Y. (2021). Leveraging machine learning for precision nitrogen management in regenerative farming systems. *Precision Agriculture, 22*(5), 913–931. https://doi.org/10.1007/s11119-021-09792-3

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. https://doi.org/10.1016/j.eja.2017.11.002

Kakimoto, S., et al. (2022). Causal forest approach for site-specific input management via on-farm precision experimentation. *Computers and Electronics in Agriculture, 199,* 107164. https://doi.org/10.1016/j.compag.2022.107164

Kaur, T., Himshikha, Singh, A., Brar, S. K., Kaur, S., & Kaur, J. (2024). Enhancing nutrient recycling through regenerative practices under different agroecosystems. In Kumar, S., Meena, R. S., Sheoran, P., & Jhariya, M. K. (Eds.), *Regenerative Agriculture for Sustainable Food Systems.* Springer, Singapore. https://doi.org/10.1007/978-981-97-6691-8_9

Kremen, C., & Miles, A. (2012). Ecosystem services in biologically diversified versus conventional farming systems: Benefits, externalities, and trade-offs. *Ecology and Society, 17*(4), 40. https://doi.org/10.5751/ES-05035-170440

LaCanne, C. E., & Lundgren, J. G. (2018). Regenerative agriculture: Merging farming and natural resource conservation profitably. *PeerJ, 6,* e4428. https://doi.org/10.7717/peerj.4428

Laidig, F., Feike, T., Lichthardt, C., Schierholt, A., Piepho, H. P. (2024). Correction to: Breeding progress of nitrogen use efficiency of cereal crops, winter oilseed rape and peas in long-term variety trials. *Theoretical and Applied Genetics, 137*(153). https://doi.org/10.1007/s00122-024-04630-z

Lal, R. (2020). Regenerative agriculture for food and climate. *Journal of Soil and Water Conservation, 75*(5), 123A–124A.

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

Leroux, C. (2020). Yield maps in precision agriculture. *Aspexit - Precision Agriculture.* Retrieved from 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 Agriculture, 24,* 2500–2521. 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 analyze the total nitrogen content in winter wheat. *Remote Sensing, 15*(8). https://doi.org/10.3390/rs15082152

Liu, J., & Vanhatalo, J. (2020). Bayesian model-based spatiotemporal survey designs and partially observed log Gaussian Cox process. *Spatial Statistics, 35,* 100392. https://doi.org/10.1016/j.spasta.2019.100392

Lu, J., Dai, E., Miao, Y., & Kusnierek, K. (2024). Developing a new active canopy sensor- and machine learning-based in-season rice nitrogen status diagnosis and recommendation strategy. *Field Crops Research, 109540.* https://doi.org/10.1016/j.fcr.2024.109540

Lyu, H., Li, Y., Wang, Y., Wang, P., Shang, Y., Yang, X., Wang, F., & Yu, A. (2024). Drive soil nitrogen transformation and improve crop nitrogen absorption and utilization: A review of green manure applications. *Frontiers in Plant Science, 14,* 1305600. https://doi.org/10.3389/fpls.2023.1305600

Lyons, S. E., Tang, Z., Booth, J., & Ketterings, Q. M. (2019). Nitrogen response models for winter cereals grown for forage. *Journal of Agronomy and Crop Science, 205*(2), 48–261.

Matavel, C. E., Meyer-Aurich, A., & Piepho, H. P. (2024). Model-averaging as an accurate approach for ex-post economic optimum nitrogen rate estimation. *Precision Agriculture, 25,* 1324–1339. https://doi.org/10.1007/s11119-024-10113-4

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

Meloni, R., Cordero, E., Capo, L., Reyneri, A., Sacco, D., & Blandino, M. (2024). Optimizing nitrogen rates for winter wheat using in-season crop N status indicators. *Field Crops Research,* 109545. https://doi.org/10.1016/j.fcr.2024.109545

Manzeke-Kangara, M. G., Joy, E. J. M., Lark, R. M., Redfern, S., Eilander, A., & Broadley, M. R. (2023). Do agronomic approaches aligned to regenerative agriculture improve the micronutrient concentrations of edible portions of crops? A scoping review of evidence. *Frontiers in Nutrition, 10,* 1078667. https://doi.org/10.3389/fnut.2023.1078667

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). https://doi.org/10.1038/s41597-023-02096-0

Morchid, M., Boulila, W., & Farah, I. R. (2024). IoT-based smart irrigation system using cloud computing for sustainable agriculture. *Sustainable Computing: Informatics and Systems, 32,* 100678.

Mousavi, M., Mieno, T., & Bullock, D. S. (2024). A new model selection approach based on local economically optimal input rate. *2024 Annual Meeting, July 28–30, New Orleans, LA,* 343892, Agricultural and Applied Economics Association. https://doi.org/10.22004/ag.econ.343892

Niemeyer, C., Nasielski, J., Janovicek, K., & Van Eerd, L. (2021). Yield can explain interannual variation in optimum nitrogen rates in continuous corn. *Nutrient Cycling in Agroecosystems, 121,* 115–128. https://doi.org/10.1007/s10705-021-10168-z

Nelder, J. A. (2000). Functional marginality and response-surface fitting. *Journal of Applied Statistics, 27*(1), 109–112. 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.* https://doi.org/10.1016/j.compag.2019.105030

Nath, S. (2024). A vision of precision agriculture: Balance between agricultural sustainability and environmental stewardship. *Agronomy Journal, 116,* 1126–1143. https://doi.org/10.1002/agj2.21405

Ogungbuyi, M. G., Guerschman, J. P., Fischer, A. M., Crabbe, R. A., Mohammed, C., Scarth, P., Tickle, P., Whitehead, J., & Harrison, M. T. (2023). Enabling regenerative agriculture using remote sensing and machine learning. *Land, 12*(6), 1142. https://doi.org/10.3390/land12061142

O'Donoghue, T., Minasny, B., & McBratney, A. (2024). Digital regenerative agriculture. *npj Sustainable Agriculture, 2,* 5. https://doi.org/10.1038/s44264-024-00012-6

Piepho, H. P., & Williams, E. R. (2021). Regression models for order-of-addition experiments. *Biometrical, 36*(8), 1673–1687.

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. 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. *Journal of Agronomy and Crop Science, 204*(5), 429–455. https://doi.org/10.1111/jac.12267

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

Pretty, J., Benton, T. G., Bharucha, Z. P., Dicks, L. V., Flora, C. B., Godfray, H. C. J., Goulson, D., Hartley, S., Lampkin, N., Morris, C., Pierzynski, G., Springmann, M., & Wratten, S. (2018). Global assessment of agricultural system redesign for sustainable intensification. *Nature Sustainability, 1*(8), 441–446. https://doi.org/10.1038/s41893-018-0114-0

Pontius, J., & McIntosh, A. (2024). Regenerative agriculture. In *Environmental Problem Solving in an Age of Climate Change.* Springer Textbooks in Earth Sciences, Geography, and Environment. https://doi.org/10.1007/978-3-031-48762-0_11

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G. P., & Smith, P. (2016). Climate-smart soils. *Nature, 532*(7597), 49–57. https://doi.org/10.1038/nature17174

Primi, R., Grossi, G., Danieli, P. P., Vitali, A., Lacetera, N., & Ronchi, B. (2024). State of the art and challenges in the environmental labelling for animal food products. *Italian Journal of Animal Science, 23*(1), 1104–1123.

Qin, Z., Wang, T., & Li, X. (2023). Cloud-based data management and analytics for precision agriculture: A review. *Computers and Electronics in Agriculture, 198,* 107006.

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

Ravikumar, S., Vellingiri, G., Sellaperumal, P., Pandian, K., Sivasankar, A., & Sangchul, H. (2024). Real-time nitrogen monitoring and management to augment N use efficiency and ecosystem sustainability – A review. *Hazards Advances, 7*(2), 100466. https://doi.org/10.1016/j.hazadv.2024.100466

Rajasekaran, J., Samad, S. R. A., & Ganesan, P. (2024). Cloud computing for smart farming: Applications, challenges, and solutions. In *Intelligent Robots and Drones for Precision Agriculture.* Springer. https://doi.org/10.1007/978-3-031-51195-0_20

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/.

Sharma, C., Pathak, P., Kumar, A., & Gautam, S. (2024). Sustainable regenerative agriculture allied with digital agri-technologies and future perspectives for transforming Indian agriculture. *Environment, Development and Sustainability.* https://doi.org/10.1007/s10668-024-05231-y

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

Saravanan, S. K., Nisha, F., Rohit, V. R., Lenin, J., Selvam, P. D., & Rajmohan, M. (2023). Impact of cloud computing on agricultural advancement using data mining algorithms. *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS),* 1570–1575. https://doi.org/10.1109/ICACRS58579.2023.10404669

Shafiya Banu, M., Antony Joseph Rajan, D., Ariff Mohamed Abuthahir Riazulhameed, A., Meenakshi, B., Mohandas, R., & Mohankumar, N. (2024). Cloud-based intelligent nutrient management system for precision agriculture using CNN. *Proceedings of the 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS), IEEE.* https://doi.org/10.1109/ICSCSS60660.2024.10624911

Sustainable Markets Initiative. (2024). World's leading food & farming businesses launch action plan to scale regenerative farming, warning speed of progress must triple to tackle the impacts of climate change. Retrieved from https://www.sustainable-markets.org/news/world-s-leading-food-farming-businesses-launch-action-plan-to-scale-regenerative-farming-warning-speed-of-progress-must-triple-to-tackle-the-impacts-of-climate-change/

Sela, S., van Es, H. M., Moebius-Clune, B. N., Marjerison, R. D., Gomes, S., & Melkonian, J. (2022). Adapt-N recommendations reduce environmental losses. *What's Cropping Up?, 25*(5). Retrieved from https://blogs.cornell.edu/whatscroppingup/2015/11/06/adapt-n-recommendations-reduce-environmental-losses/

Tanaka, T. S. T., et al. (2024). Can machine learning models provide accurate fertilizer recommendations? *Precision Agriculture, 25,* 1839–1856. https://doi.org/10.1007/s11119-024-10136-x

Toffolini, Q., & Jeuffroy, M. H. (2022). On-farm experimentation practices and associated farmer-researcher relationships: A systematic literature review. *Agronomy for Sustainable Development, 42,* 114. https://doi.org/10.1007/s13593-022-00845-w

Vitousek, P. M., Naylor, R., Crews, T., David, M. B., Drinkwater, L. E., Holland, E., Johnes, P. J., Katzenberger, J., Martinelli, L. A., Matson, P. A., Nziguheba, G., Ojima, D., Palm, C. A., Robertson, G. P., Sanchez, P. A., Townsend, A. R., & Zhang, F. S. (2009). Nutrient imbalances in agricultural systems and their environmental impacts. *Science, 324*(5934), 1022–1026. https://doi.org/10.1126/science.1170261

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

Wang, Y., Peng, Y., Lin, J., Wang, L., Jia, Z., & Zhang, R. (2023). Optimal nitrogen management to achieve high wheat grain yield, grain protein content, and water productivity: A meta-analysis. *Agricultural Water Management,* 108587. https://doi.org/10.1016/j.agwat.2023.108587

Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A metareview. *Remote Sensing of Environment, 236,* 111402. https://doi.org/10.1016/j.rse.2019.111402

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

Xu, T., Yu, L., He, W., & Yuan, Y. (2020). A cloud-based platform for large-scale remote sensing data processing and analysis. *ISPRS International Journal of Geo-Information, 9*(4), 236. https://doi.org/10.3390/ijgi9040236

Yara International ASA. (n.d.). *Yara N-Sensor: Variable Rate Nitrogen Management*. Retrieved from https://www.yara.us/crop-nutrition/tools-and-services/n-sensor

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

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

Acknowledgements

The author extends sincere gratitude to the John Deere Intelligent Solutions Group and a partner farm for their invaluable support in facilitating this research. Special appreciation is directed to Tobias Füge, the corresponding grower, for his essential support in trial implementation and his permission to share data openly for the benefit of the research community.

Heartfelt thanks to my primary supervisor, Professor Hance-Peter Piepho, for his unwavering guidance, expertise, and encouragement throughout this journey. I am grateful to my co-supervisor, Prof. Dr. Peter Pickle, whose support during my 5 years of work at the John Deere European Technology Innovation Center has been indispensable.

This work, from ideation and research to implementation, was completed independently by me. However, I would like to acknowledge the efforts of my master's students, who, under my guidance during their internships at John Deere, Kaiserslautern, contributed to specific coding tasks. Their diligent work on developing automated tools, such as data pipelines and cloud infrastructure for data management and storage, was conducted under my direct instructions and played a valuable role in supporting this research.