UC Davis

UC Davis Electronic Theses and Dissertations

Title

Development of Transferable Grapevine Nitrogen Retrieval Algorithm

Permalink

https://escholarship.org/uc/item/17j6h27t

ISBN

9798290613789

Author

Peanusaha, Sirapoom

Publication Date

2025-06-15

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Development of Transferable Grapevine Nitrogen Retrieval Algorithm

By

SIRAPOOM PEANUSAHA

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biological Systems Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Alireza Pourreza, Chair

_____

Ali Moghimi

_____

Matthew W. Fidelibus

Committee in Charge

2025

ABSTRACT

Nitrogen (N) and nutrient management are central to sustainable grapevine production, supporting optimal yield, maximizing economic return, and reducing environmental pollution. Over the past two decades, remote sensing technologies have emerged as high potential tools for monitoring plant nutritional status. Unmanned aerial systems (UAS) and advances in computing capacity offer significant potential for nutrient management applications by enabling cost-effective acquisition of high-resolution imagery across large areas. Despite their growing popularity, traditional vegetation indices (VIs) and purely data-driven approaches face substantial limitations in generalizability, often failing to transfer across sensors, locations, and environmental conditions due to their empirical nature.

This dissertation aims to develop generalizable, and interpretable methodologies for N and nutrient estimation in grapevines by leveraging physically based radiative transfer models (RTM) across multiple spatial scales. The first part of the study investigates leaf-level spectral dynamics using the PROSPECT model, evaluating the sensitivity of VIs and RTMs to biochemical and biophysical traits across different leaf ages. Comparative analysis reveals that data-driven methods achieved superior predictive performance (e.g., Gaussian Process Regression attained $R^2 \approx 0.78$), whereas physically based models provided mechanistic interpretability grounded in biophysical parameters. A hybrid approach that combined physically based RTM-derived traits with a Random Forest Regressor yielded lower accuracy ($R^2 \approx 0.54$) while reducing training data requirements by approximately 50%. Nevertheless, the accumulation of errors from RTM inversion and subsequent model training suggests that caution is warranted when applying the hybrid strategy.

Building on these mechanistic insights, the second part of the dissertation scales the modeling framework to the canopy level through integration with the HELIOS 3D ray-tracing

framework. A novel pipeline was developed in which synthetic multispectral imagery was generated under varying canopy structures, sun angles, and leaf optical properties to quantify how each factor influences overall canopy reflectance, and this imagery was used to train a deep learning model. A major milestone achieved in this study was the complete replacement of data-intensive imaging with synthetic multispectral data for model training, representing an efficient and scalable alternative to conventional data-driven approaches. The trained models demonstrated the ability to predict nutrient-related traits from real-world UAS imagery, generating spatial variability maps that identify zones requiring nutrient intervention. Although domain gaps between synthetic and real reflectance data persist the findings confirm the feasibility of using physically simulated data to train transferable models for field deployment.

Collectively, this dissertation presents a hybrid modeling strategy that combines physics-based simulation and machine learning to address the fundamental trade-offs between accuracy, generalizability, and interpretability in nutrient monitoring. The resulting framework provides a strong foundation for advancing precision agriculture by enabling N estimation under physically based approach and supporting future standardization of remote sensing workflows using synthetic reference scenes.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Professor Alireza Pourreza for his unwavering support, guidance, patience, and generous investment of time throughout my Ph.D. journey — especially during the toughest times. His mentorship has been a cornerstone of my academic and personal growth. I am also sincerely thankful to Professor Matthew W. Fidelibus and Professor Ali Moghimi for their collaborative spirit and enduring patience in working with me. Their support and insight significantly enriched my research experience. My gratitude goes to Professor Stavros Vougioukas, Professor Andre Daccache, Professor Farzaneh Khorsandi, and Professor Irwin Donis-Gonzalez for their support and guidance as members of my qualifying exam committee. I am incredibly grateful to all members of the Digital Agriculture Lab, especially Prof. Xin Zhang, Dr. Hamid Jafarbiglu, Momtanu Chakraborty, Yuto Kamiya, Mohammadreza Narimani, Parastoo Farajpoor, German Zuniga-Ramirez, and Mario Salinas. Thank you for your friendship, collaboration, and for making the lab such a heartwarming place to work. I also wish to acknowledge all the mentors and teachers who guided me along the way and helped ignite my passion for research — far too many to name but never forgotten. Lastly, I gratefully acknowledge the financial support from the United States Department of Agriculture – National Institute of Food and Agriculture (USDA-NIFA) and the California Table Grape Commission, whose support made this research possible.

DEDICATIONS

To my parents, Pitak and Sataporn, who gave me strength and wisdom. I would not be here without your love, care, and unwavering support.

To my wife, Atinoot, who has stood by my side through every step of this journey.

TABLE OF CONTENTS

Chapter 1: Introduction 1
Importance of Nitrogen Management in Agriculture and Grapevine 1
Importance of Grapevine Industry in California & Nitrogen Pollution 1
Traditional Nitrogen Retrieval Methods 1
Nitrogen Retrieval Using Remote Sensing 2
Mechanistic Modeling (Physically Based Approaches) 3
Goal and Objectives 5
Chapter 2: Literature Review 6
Nitrogen Management in Grapevine Cultivation 6
Importance and Requirements of Nitrogen in Grapevines 6
Relationship of Nitrogen to Yield, Fruit Quality, and Canopy Vigor 6
Environmental Impacts of Nitrogen Fertilization 7
Overview of Remote Sensing for Nitrogen Detection 8
Leaf level and canopy level prediction 8
Area-based traits and mass-based traits 9
Hyperspectral and Multispectral Imaging 9
Approaches for N Retrieval using Remote Sensing Technology 11
Vegetation Indices 11
Data Driven Approaches 13
Physically Based, Radiative Transfer Model and Ray-tracing Model 15
Chapter 3: Nitrogen retrieval in grapevine (Vitis vinifera L.) leaves by hyperspectral sensing 18
Background 18
Materials and Methods 19
Plant Materials and Study Sites 19
Proximal Hyperspectral Scan and Ground Truth Data 22
N retrieval approaches 24
Results 28

VI performance and inconsistent relationship between chlorophyll and N 29
Estimated RTM parameters and their correlation with N_{mass} and N_{area} 32
Performance comparison between single input approach and multiple inputs approach 34
Discussion..... 37
Mass-based Vs. area-based leaf biochemical parameters 37
Trade-off between spectral regions and prediction accuracy 39
Conclusion 40
Chapter 4: Advancing Canopy-Scale Nutrient Estimation in Grapevine (*Vitis. Vinifera. L*) with Synthetic Multispectral Data from 3D RTMs..... 42
Background..... 42
Material and Methods 44
Field Site and Experimental Setup 44
Multispectral image campaigns..... 45
Multispectral image preprocessing 46
HELIOS 3D biophysical modeling framework..... 47
Radiation Transport Plug-in 47
Plant Architecture Plug-in..... 48
Plant Object File Generation and Dimension Reduction..... 48
Spectral PCA Analysis and Synthetic Spectra Generation 49
Scene Setup and Processing 50
Data Preparation and Deep Learning Model 51
Results..... 52
PCA comparison between real and synthetic data generated from HELIOS 52
Model Performance on Synthetic Data (5-Fold Cross-Validation) 53
Boxplot comparing the prediction results in each season 56
Relationship Between Vegetation Indices and Predicted Chlorophyll and Nitrogen 57
Spatial Distribution of Predicted Canopy Traits 59
Discussion..... 62
Trait-Specific Prediction Performance: Biochemical vs. Structural..... 62

Framework Implications for application of digital twin technology in remote sensing of agriculture 65
Conclusion 66
Limitation and Future works 68
Chapter 5: Overall Conclusion and Future Work 69
Overall Conclusion 69
Future Research 70
Reference 72
Appendix 80

TABLE OF FIGURES

Figure 1 The distribution of N_{mass} and Narea ground truth data. 23

Figure 2 The N_{mass} Vs. N_{area} scatter plot 29

Figure 3: Performance of top five vegetation indices on predicting Nmass and Narea 30

Figure 4: Correlation between Chlorophyll estimated by PROSPECT PRO 31

Figure 5: Estimated Chlorophyllarea trend found across potted Flame Seedless and Solbrio 32

Figure 6: Correlation coefficient between direct RTM biochemical parameters 33

Figure 7: Correlation heatmap between mass-based and area-based PROSPECT 33

Figure 8: Performance comparison of major N retrieval approaches from all three collection ... 35

Figure 9: Feature importance of RFR+RTM hybrid models for Nmass and Narea prediction 36

Figure 10: Map of the experimental site at Shafter, CA 45

Figure 11: the PCA of the real and synthetic data selected for the simulation 50

Figure 12: Principal component comparison of synthetic (red) and real (green) canopy 53

Figure 13: Predicted vs. Actual Trait Values from 5-Fold Cross-Validation on Synthetic Data . 55

Figure 14: Predicted Trait Distributions on Real-World Grape Canopy Data Across Seasons.... 57

Figure 15: Relationship between vegetation indices (NDVI and NDRE) and model traits 58

Figure 16: Spatial Distribution of Predicted LAI, Chlorophyll, and Nitrogen 60

Figure 17: Estimated Canopy-Level Nutrient Mass for Nitrogen and Chlorophyll 61

TABLE OF TABLES

Table 1 Derived Ground Truth Information 23

Table 2 VIs used for estimation of N content 24

Table 3 Change in R^2 values for data-driven approaches after spectral resolution reduction 37

Table 4 Data Collection and metadata range of the real MS image 46

Table 5 Parameters Used for Grapevine Canopy Model Generation in HELIOS Simulations 49

TABLE OF ABBREVIATIONS

Abbreviation	Full Term
N	Nitrogen
UAS	Unmanned Aerial System
VIs	Vegetation Indices
RTM	Radiative Transfer Model
GPR	Gaussian Process Regression
RFR	Random Forest Regression
LAI	Leaf Area Index
MS	Multispectral
HS	Hyperspectral
PLSR	Partial Least Squares Regression
NDVI	Normalized Difference Vegetation Index
NDRE	Normalized Difference Red Edge Index
SWIR	Short-Wave Infrared
VNIR	Visible and Near-Infrared
CHL	Chlorophyll
EWT	Equivalent Water Thickness
LMA	Leaf Mass per Area
ANT	Anthocyanin
CAT	Carotenoids
PROT	Protein Content
N_{struct}	Leaf Structure Parameter
MCARI	Modified Chlorophyll Absorption in Reflectance Index
MTCI	MERIS Terrestrial Chlorophyll Index
NDNI	Normalized Difference Nitrogen Index
TCARI	Transformed Chlorophyll Absorption Ratio Index
PRI	Photochemical Reflectance Index
MRENDVI	Modified Red Edge Normalized Difference Vegetation Index
VREI2 (VOG2)	Vogelmann Red Edge Index 2
ARI2	Anthocyanin Reflectance Index 2
SR533	Simple Ratio 533/565
SAVI	Soil-Adjusted Vegetation Index
DLS	Downwelling Light Sensor
PCA	Principal Component Analysis
FNN	Fully Connected Neural Network

Chapter 1: Introduction

Importance of Nitrogen Management in Agriculture and Grapevine

Accurate monitoring and management of nitrogen (N) in plant tissues is essential in agriculture due to its direct impact on crop yield, crop quality, and growers' profitability. Furthermore, overuse of N fertilizer has led to ground water pollution in agricultural intensive regions, especially in California. N is a critical macronutrient for grapevine growth, influencing photosynthetic efficiency, plant vigor, and fruit development. Insufficient N availability limits grape yield and negatively affects fruit quality, while excessive N fertilization may cause excessive canopy growth, reduced fruit quality, and environmental pollution issues such as groundwater contamination and increased greenhouse gas emissions [1], [2].

Importance of Grapevine Industry in California & Nitrogen Pollution

Grapevine cultivation is a major agricultural industry globally, with more than 75 million metric tons of grape produced annually. California is particularly prominent, producing nearly 95% of the U.S. table grape supply, which amounts to approximately 8,600 metric tons each year [3]. The San Joaquin Valley in California, a significant production region, relies on N fertilizers to maintain productivity. Over-application of N fertilizers on farms in this region has caused nitrate pollution of groundwater, prompting regulatory actions by California to require growers to develop N management plans aimed at reducing environmental impacts [4], [5].

Traditional Nitrogen Retrieval Methods

Plant tissue analysis has traditionally been employed to assess N status of grapevines, guiding fertilization decisions. Growers typically collect leaf blades or petioles during specific growth stages, such as bloom or veraison, from vines distributed throughout their vineyards. The leaf normally selected for sampling at bloom is positioned directly opposite the fruit cluster, typically

situated around the 3rd to 5th leaves from the base of a shoot. This leaf has been considered representative of vine nutritional status at bloom, and extension services regularly recommend its use for tissue sampling [6], [7]. At the onset of fruit ripening (veraison), which occurs later in the summer, recently expanded leaves are selected (rather than leaves opposite clusters). Traditional leaf sampling methods have several significant limitations. Leaf sampling is labor-intensive, costly, and provides limited spatial resolution, restricting detailed understanding of within-field variability[8], [9], [10]. This lack of spatial detail constrains the effectiveness of precision nutrient management approaches aimed at addressing spatially variable N demands within vineyards [9], [11].

Nitrogen Retrieval Using Remote Sensing

Remote sensing (RS) technologies, such as multispectral (MS) and hyperspectral (HS) imaging, have shown great potential as efficient tools for monitoring grapevine N status. Cameras are typically mounted on Unmanned Aerial Systems (UAS), enabling rapid and cost-effective acquisition of imagery from altitudes ranging approximately between 20 to 120 meters. This technology has potential to provide growers with high-resolution maps on a per-vine nutrient level showing nutrient status variability across their fields.

Among remote sensing methods, vegetation indices (VIs) have been most frequently utilized due to their simplicity, low computational demand, and effective normalization that reduces environmental noise [12], [13]. Commonly used indices include the Normalized Difference Vegetation Index (NDVI) and Normalized Difference Red Edge Index (NDRE), which primarily correlate with chlorophyll content. However, despite their widespread usage, VIs exhibit limitations including saturation effects at higher vegetation densities, and their sensitivity to N is

indirect since chlorophyll represents only a small fraction of total leaf N content [9], [14], [15], [16].

Linear non-parametric regression methods, such as Partial Least Square Regression (PLSR), are frequently applied to hyperspectral data because of their robustness in handling highly correlated and noisy spectral measurements [17]. Although PLSR can achieve high accuracy, its main limitation is the need for local calibration, restricting its transferability across different vineyards or varying environmental conditions [17], [18], [19].

Non-linear, non-parametric regression methods, including Gaussian Process Regression (GPR), have recently gained prominence for their ability to handle complex spectral relationships and provide uncertainty quantification [17]. However, despite their predictive accuracy, these machine learning approaches often exhibit limitations related to overfitting, particularly when transferring models to new datasets collected under different environmental conditions [17], [20]

Mechanistic Modeling (Physically Based Approaches)

Physically based RTM offer significant promise for N retrieval by simulating how radiation interacts with plant leaves and canopies, guided explicitly by underlying biochemical and biophysical traits.

The PROSPECT model is a widely used RTM that accurately simulates leaf reflectance and transmittance, enabling estimation of biochemical properties, including leaf N content. A recent advancement of PROSPECT, the PROSPECT PRO version, has improved the retrieval of leaf biochemical traits such as protein content—a key biochemical trait directly related to N levels. The improved accuracy in protein estimation significantly enhances N retrieval accuracy and improves

the potential for algorithm transferability across different growing conditions and plant varieties [21].

Extending leaf-level modeling to canopy-level scenarios requires explicit consideration of canopy structure complexities, such as leaf angle distribution, leaf area index (LAI), and illumination geometry These factors are effectively modeled through advanced ray-tracing techniques, such as those implemented in the HELIOS 3D modeling framework. HELIOS is primarily a sophisticated 3D plant modeling tool that incorporates physically based RTM into its detailed scene generation. It also integrates the PROSPECT model to accurately simulate leaf spectral reflectance and transmittance based on leaf biochemical parameters. This model can simulate not only RGB but also multispectral and hyperspectral images under precisely defined environmental and canopy conditions. It enables the systematic variation of key parameters, including leaf biochemical properties, canopy structure, LAI, sun angle (zenith and azimuth angle), and other detailed physical settings [22].

A critical advantage of 3D ray-tracing simulation is its capability to generate synthetic datasets that can serve as an extensive training source for machine learning algorithms. These synthetic datasets, automatically annotated with precise biophysical and biochemical labels, substantially reduce the labor and resources required for data collection and manual annotation. Moreover, synthetic data generated by HELIOS can account for diverse scenarios of environmental conditions and canopy architectures, thus helping to build highly robust, generalizable N retrieval algorithms.

Goal and Objectives

This dissertation aims to advance knowledge in N and nutrient retrieval in grapevines using remote sensing technologies. My first objective focuses on developing mechanistic approaches at the leaf level, employing physically based RTMs to fundamentally understand how variations in leaf biochemical and biophysical parameters influence spectral reflectance. Specifically, this part of the research seeks to (1) compare the performance of five main N retrieval approaches, (2) assess the impact of leaf age on model accuracy, and (3) evaluate models calibrated using both mass-based (N_{mass}) and area-based (N_{area}) N measures. The insights gained from this foundational leaf-level study will establish a critical understanding of the dynamics governing spectral reflectance, essential before scaling up to the canopy level.

Building upon this understanding, my second objective extends physically based modeling from the leaf level to the canopy scale using the HELIOS 3D ray-tracing model. At this stage, I aim to (1) evaluate the feasibility of training deep learning models using synthetic imagery generated through HELIOS simulations, subsequently applying these trained models to real-world multispectral images for accurate canopy-level N prediction. Additionally, I intend to (2) systematically examine how N prediction performance is influenced by critical canopy structural and environmental parameters, specifically azimuth angle and leaf area index (LAI). Lastly, I will (3) investigate the effects of variations in leaf biochemical and biophysical traits on N prediction at the canopy level. This approach directly aligns with growers' needs, providing robust, transferable, and actionable insights for precise N management decisions. By integrating detailed mechanistic insights gained at the leaf level with advanced modeling techniques at the canopy scale, this dissertation seeks to establish comprehensive and reliable methodologies for precision N management in grapevine cultivation.

Chapter 2: Literature Review

Nitrogen Management in Grapevine Cultivation

Importance and Requirements of Nitrogen in Grapevines

N is an essential macronutrient critically important for grapevine growth, fruit development, and productivity. Adequate N availability supports essential plant physiological functions, including photosynthesis, cell division, and the synthesis of vital proteins and enzymes [14]. Grapevines have varying N demands throughout their growth cycle, particularly during key phenological stages including budbreak, bloom, fruit set, and veraison [23], [24].

Growers typically apply N fertilizers based on vine nutrient status assessments determined by plant tissue analysis conducted at bloom and veraison. Traditional methods involve collecting leaf blades or petioles from grapevines, with samples taken from the leaf opposite the cluster at bloom and from the most recently matured leaf at veraison, as these have been shown to best represent the vine's nutritional status [6], [7]. N fertilizers are generally applied to coincide with critical periods of N uptake in grapevines when vines exhibit peak N absorption [25]. .

Relationship of Nitrogen to Yield, Fruit Quality, and Canopy Vigor

N availability directly influences grape yield, fruit quality, and canopy vigor. Properly managed N fertilization enhances yield by ensuring adequate canopy development, fruit set, and berry growth, with optimal N typically resulting in ma ximum grape productivity and high fruit quality [24]. For example, N deficiency reduces vine vigor and fruit mass s [26], smaller berries and delayed ripening due to impaired photosynthesis and carbohydrate accumulation [27], [28], and altered anthocyanin biosynthesis that diminishes color and quality in red cultivars [27], all of which negatively affect yield and vineyard profitability.

Conversely, excess N fertilization promotes excessive vegetative growth that reduces yield and negatively affects fruit quality. High N availability results in dense, overly vigorous canopies, which limit light penetration and airflow, increasing the risk of fungal infections such as powdery mildew and *Botrytis cinerea*-induced berry rot [29]. Shading also reduces fruit exposure to light, and leads to poor fruit composition[30]. Moreover, the N dilution effect, where higher biomass dilutes N concentration per tissue unit, lowers N use efficiency and increases variability in berry quality [30]. Excess N delays ripening and reduces sugar accumulation in the must due to competition between vegetative and reproductive growth [31]. It can also suppress anthocyanin biosynthesis, producing pale berries and diminishing wine color intensity [31].

Environmental Impacts of Nitrogen Fertilization

Improper N management practices, particularly over-application of N fertilizers, pose significant risks to environmental sustainability. Excessive N applications increase nitrate leaching into groundwater, leading to contamination of drinking water supplies, posing substantial public health risks [4], [5]. In California's San Joaquin Valley, extensive agricultural N fertilizer use has notably contributed to widespread nitrate contamination of groundwater, necessitating regulatory interventions[4] .

N overuse is also associated with surface water pollution, causing eutrophication, harmful algal blooms, and degraded aquatic ecosystems, further emphasizing the critical importance of precise N management [5]. To address the environmental impacts of N use, California has implemented regulatory measures that promote more sustainable agricultural practices. Through programs such as the Irrigated Lands Regulatory Program (ILRP), growers are required to develop and implement N Management Plans that track and report N inputs and outputs. These policies aim to reduce nitrate contamination in groundwater by improving transparency and accountability in nutrient

management [32]. Complementing these regulatory efforts, the state also encourages best management practices such as fertigation, cover cropping, and regular soil testing. These strategies, along with grower education initiatives, support the transition toward more efficient and environmentally responsible N use in agriculture [32].

Effective N management thus simultaneously addresses crop productivity, economic viability, and environmental sustainability. Adoption of advanced N monitoring technologies, such as remote sensing, has emerged as a promising approach to optimize nutrient applications, particularly N in this study, while minimizing environmental risks associated with traditional N management practices.

Overview of Remote Sensing for Nitrogen Detection

Leaf level and canopy level prediction

Leaf-level retrieval focuses on the acquisition of biochemical or nutrient traits from single leaves, typically scanned in situ using portable spectrometers. By examining individual leaves in isolation, leaf-level approaches often demonstrate higher model accuracy because they incorporate fewer confounding factors, including soil reflectance and field variability such as sun angle [33], and atmospheric condition [34], and canopy complexity [10]. In contrast, canopy-level retrieval aims to estimate these biochemical traits, particularly N content, at larger spatial scales, such as entire plants, rows, or blocks, by deploying sensors on platforms like UAS. Canopy-level data collection is more directly applicable for large-scale decision-making, such as fertilization [35].

In a comprehensive survey of crop N sensing in agriculture between 1994 and 2019, Berger et al. [17] found that proximal sensing, with instruments such as ASD spectrometers in the 350–2500 nm wavelength range, accounted for 75% of the 167 evaluated studies. Proximal sensing remains central to basic research, as it facilitates detailed examination of spectral–physiological

associations while minimizing the confounding effects of incident solar angle, atmospheric variables, and canopy structures. Meanwhile, regional-scale spectroscopy offers broader spatial coverage and faster repeated observations, albeit necessitating more advanced processing and calibration workflows. With current initiatives devoted to bridging leaf-level precision and canopy-scale implementation, combining both measurement scales have the potential to advance site-specific N management strategies.

Area-based traits and mass-based traits

Area-based N content (N_{area}) is typically defined per unit of leaf or canopy surface, for instance, in g/cm^2 at the leaf scale, g/m^2 at the canopy scale, or kg per hectare in a field. Meanwhile, mass-based N concentration (N_{mass}) is measured per unit of dry mass ([18]. In tissue test lab reports, N is reported in the mass-based form. However, N_{area} can be converted to N_{mass} by dividing by leaf mass per area (LMA), which is the ratio of dry mass to leaf area [36].

Many studies support N_{area} for remote sensing applications, since leaf-level traits tend to correlate more strongly with surface area rather than with concentration alone. Additionally, N_{area} offers advantages for scaling between leaf and canopy level, because it can be multiplied by the leaf area index [37]. Nonetheless, N_{area} is more challenging to measure since the leaf area can change dramatically as the leaf shrinks and can be altered more easily.

In contrast, N_{mass} draws closer attention to the fraction of plant mass that is N, which indicates general plant health [38]. Moreover, N_{mass} in remote sensing model training is more widely-used, because dried and ground leaf samples with no data about leaf area available [37].

Hyperspectral and Multispectral Imaging

RS technologies, such as multispectral (MS) and hyperspectral (HS) aerial imagery, have the potential to measure plant nutrient status with a relatively high margin of confidence [39], [40].

MS and HS imaging represent two ends of the spectrum in remote sensing technologies used for N detection, each with distinct roles and trade-offs. HS sensors capture reflectance across hundreds of narrow, continuous spectral bands—typically spanning 350 to 2500 nm—enabling the detection of subtle biochemical features such as chlorophyll, water, and especially protein-related absorption peaks in the short-wave infrared region (SWIR), which are critical for accurate N estimation[8], [21]. This spectral richness makes HS imaging a powerful research tool for identifying the most informative spectral bands [34].

However, HS systems also come with notable limitations. The vast number of bands introduces strong multicollinearity, necessitating dimensionality reduction techniques like PCA or LASSO [38], [41] or Ensemble Band Selection [42]. Additionally, the volume of data generated by HS sensors requires high-capacity storage, advanced computing resources, and specialized software for calibration and correction. These factors limit the operational scalability and make HS imaging less suitable for commercial deployment in routine agricultural monitoring [38].

In contrast, multispectral (MS) sensors, typically comprising 5–12 discrete bands in the visible, near-infrared, and sometimes mid-infrared regions, offer a more practical balance between cost, accessibility, and ease of data handling [38], [43], [44]. MS sensors are lightweight, widely compatible with UAS platforms, and well-suited for operational crop monitoring over large areas. However, due to their limited spectral resolution, they may miss critical information if key biochemical features fall outside the selected bands. This constraint can lead to reduced sensitivity and accuracy, particularly for substances with similar spectral characteristics.

Importantly, the strengths of HS imaging can be leveraged to inform the design of optimized multispectral sensors. By analyzing HS data to determine which narrow bands are most strongly associated with N-related traits [34], researchers can guide industry toward building MS cameras

that include only the most relevant bands, thus combining scientific rigor with operational efficiency. This research-to-application pathway underscores the complementary nature of HS and MS technologies in advancing precision agriculture.

Approaches for N Retrieval using Remote Sensing Technology

Vegetation Indices

VIs are mathematical combinations of reflectance values at specific spectral bands, typically designed to highlight vegetation characteristics such as greenness, chlorophyll content, or canopy density. Among the most widely used is the Normalized Difference Vegetation Index (NDVI), calculated as the normalized difference between near-infrared (NIR) and red reflectance. The underlying principle of NDVI and similar indices is based on the fact that green vegetation absorbs red light (due to chlorophyll) and reflects NIR light, enabling a proxy estimation of plant vigor and health [45], [46]. Other indices, such as the Normalized Difference Red Edge (NDRE), Modified Chlorophyll Absorption in Reflectance Index (MCARI), and Normalized Difference Nitrogen Index (NDNI), utilize red-edge and shortwave infrared (SWIR) regions, which are often more sensitive to N content due to their relation with biochemical absorption features [43], [47].

VIs have been extensively applied across various crops and phenotyping contexts to estimate N content and other plant traits. In viticulture, NDVI and NDRE have been used to assess spatial variability in canopy vigor, which is closely related to N content. Hall et al. [48] showed that remote sensing imagery using NDVI provided insights into vine vigor and potential yield quality within vineyards [48]. Moghimi et al. [9] applied UAS-based multispectral imagery combined with machine learning to predict leaf N concentration in grapevines, emphasizing the utility of narrowband VIs centered on red-edge wavelengths. Similarly, Omidi et al. [42] used ensemble feature selection from hyperspectral data to identify the most informative bands for N detection in

grape leaves, which were largely outside the standard commercial multispectral ranges, suggesting the value of customized indices. Rustioni et al. demonstrated that reflectance spectroscopy could differentiate between specific nutrient deficiencies (e.g., N, K, Mg) in grapevine leaves, which supports the development of diagnostic VIs for nutrient management [16].

In cereal crops, VIs have also proven effective. Magney et al. used RapidEye satellite imagery to map N uptake in wheat fields over three growing seasons, finding that NDRE outperformed other VIs with an R^2 of 0.81 [49]. Vizzari et al. demonstrated that NDVI derived from Sentinel-2 imagery could guide variable rate N fertilization in wheat, achieving environmental and economic benefits without compromising yield or protein content [50]. Taskos et al. also evaluated canopy-based sensors such as the ACS-430 and found that red-edge-based VIs correlated well with leaf N, chlorophyll concentration, and yield-related traits in grapevines [51]. These examples underscore the widespread application of VIs in both research and operational settings for non-destructive and high-throughput estimation of N-related traits.

However, despite their popularity and ease of use, vegetation indices have several limitations. One major constraint is that they estimate N indirectly via chlorophyll content, which can be affected by multiple environmental stresses beyond N deficiency, such as drought, pests, or aging [45], [46]. This can introduce uncertainties, especially in mixed-stress field conditions. Another well-known issue is the saturation of indices like NDVI at high LAI or canopy density, which limits their sensitivity to N differences in vigorous or mature crops [38], [49], [52]. More specifically, Vis could also be affected by different trellis architecture. Furthermore, the performance of VIs can vary with crop type, growth stage, and environmental conditions, often necessitating site-specific calibration or integration with additional spectral or structural information [1].

In summary, VIs offer a simple, rapid, and widely adopted approach for N status estimation and general plant health monitoring. While effective in many contexts, their indirect nature, sensitivity to canopy conditions, and varying performance across stages and crops highlight the need for complementary methods or enhancements, such as integration with machine learning models, to overcome inherent limitations and improve robustness.

Data Driven Approaches

Linear non-parametric regression, commonly referred to as chemometrics, utilizes statistical and mathematical transformations to reduce high-dimensional hyperspectral data into simplified features suitable for modeling plant properties such as N content [53]. A foundational chemometric method is Principal Component Analysis (PCA), which projects correlated hyperspectral variables into uncorrelated principal components, thereby compressing the dataset's dimensionality while retaining major sources of variance [54]. However, the most widely adopted chemometric method for N estimation is Partial Least Squares Regression (PLSR) due to its robustness in managing spectral noise and multicollinearity, enabling high predictive accuracy across noisy datasets [33]. For example, Schmidtke et al. applied PLSR on Fourier-transform infrared spectroscopy data from wood and root samples of wine grapes and achieved an R^2 of up to 0.97, illustrating the technique's precision in localized calibration contexts [19]. Similarly, Ye et al. demonstrated that both PLSR and Multiple Linear Regression could effectively predict N content in apple tree leaves and canopies from hyperspectral data, with R^2 values exceeding 0.77 depending on preprocessing strategy [55]. Despite their effectiveness, chemometric models require extensive site- and condition-specific calibration, making them unsuitable for direct transfer across different seasons, geographic locations, plant varieties, or growth stages [56], [57].

To overcome these limitations and improve scalability, non-linear non-parametric regression approaches, particularly machine learning (ML) algorithms, have gained traction with the increasing availability of computational resources [20]. Gaussian Process Regression (GPR), a kernel-based method introduced by Rasmussen (2005), is particularly valuable in environmental modeling due to its probabilistic framework that not only provides a point estimate but also an associated confidence interval, enhancing interpretability and reliability in natural systems [56], [58]. Other commonly applied ML algorithms include Support Vector Regression (SVR), which handles nonlinear relationships through kernel trick applications [59], and Random Forest Regression (RFR), an ensemble-based method robust to overfitting and capable of managing multicollinearity effectively [57], [60]. In their grapevine N studies, Omidi et al. employed six ML algorithms including RFR and SVM to develop an ensemble feature selection pipeline that identified key wavebands from over 800 hyperspectral samples, leading to improved N prediction performance[42]. Chancia et al. applied both the ensemble feature selection and PLSR techniques on aerial hyperspectral imagery to estimate grape leaf N concentration, obtaining R^2 values of 0.44 and 0.45, respectively, demonstrating comparative performance between ML and traditional chemometrics [34]. In tobacco plants, Fu et al. utilized an ensemble of six regression models, including artificial neural networks (ANN), support vector machines (SVM), and Gaussian processes, to predict photosynthetic traits, such as maximum carboxylation rate of Rubisco, from hyperspectral reflectance, where stacking these models outperformed individual regressors in both accuracy and robustness [61].

Nonetheless, the adaptability and high dimensional flexibility of ML approaches come at the cost of increased risk of overfitting, particularly when models are trained on narrowly scoped datasets. Overfitting occurs when algorithms become excessively tuned to the noise or anomalies within the

training set, reducing the generalizability and reproducibility of the model across other datasets or field conditions [56]. While techniques like cross-validation and regularization can mitigate this risk, the lack of a direct mechanistic or physical link between spectral input and biological traits limits the interpretability of most ML models. This limitation is critical in agricultural decision-making, where actionable insight requires models that are not only accurate but also physiologically meaningful and transferable across temporal and spatial domains.

In summary, chemometric and machine learning approaches each present distinct advantages in modeling complex plant traits such as N content. However, both require careful design considerations, chemometrics for calibration transferability and ML for model generalization and interpretability.

Physically Based, Radiative Transfer Model and Ray-tracing Model

Physically-based RTM offer a robust alternative to data-driven approaches by simulating the interaction of light with vegetation based on physical laws [21], [62], [63], [64], [65]. These models are grounded in radiative transfer theory and can be applied at the leaf and canopy scales. Leaf-level models such as the PROSPECT family simulate leaf reflectance and transmittance as functions of biochemical parameters including chlorophyll, water content, and structural traits [63]. The most recent iteration, PROSPECT-PRO, extends this capability by incorporating additional constituents such as proteins and carbon-based compounds. This enhancement is particularly relevant for N estimation, as proteins constitute the major N pool in foliage [21], [66].

RTMs can be implemented in two operational modes. In the forward mode, the model predicts spectral reflectance from known plant traits. Conversely, in the inverse mode, physiological traits are retrieved from measured spectral data through algorithms such as optimization or lookup tables [63], [64], [67]. In various crop systems, PROSPECT has shown promise in estimating N-related

parameters. For instance, Wang et al. used the inverse PROSPECT model to estimate leaf mass per area (LMA), equivalent water thickness (EWT), and chlorophyll content across multiple tree species [68]. They reported a R^2 of 0.71 for leaf N_{area} Such physically-based methods provide enhanced interpretability and generalizability due to their foundation in measurable biophysical properties [18]. PROSPECT-PRO's capability to retrieve protein content, a direct biochemical proxy for N, offers a distinct advantage over traditional spectral indices, which generally infer N indirectly via chlorophyll estimation [21].

At the canopy scale, models such as PROSAIL integrate PROSPECT's leaf-level outputs with canopy structural descriptors, including LAI, solar angle, and observer geometry. However, PROSAIL assumes homogeneous canopies and is thus limited in its ability to represent complex plant architectures found in orchards and vineyard. To address this, three dimensional (3D) RTMs using ray tracing algorithms, such as the LESS model [69], [70] and, HELIOS model [22], simulate light transport through detailed 3D canopy geometries, capturing the anisotropic scattering and shadowing effects essential for accurate reflectance modeling in structurally heterogeneous crops.

Recent studies highlight the potential of 3D ray tracing RTMs for N estimation in orchard systems, demonstrating promising results in almond canopies [71]. Pourreza et al. applied a 3D PROSAIL inversion to retrieve canopy biochemical and structural parameters, which were subsequently used as inputs to machine learning models for predicting N content from hyperspectral canopy reflectance [10]. Cheng et al. coupled the 3D ray tracing model LESS with PROSPECT-D to estimate leaf chlorophyll content in apple orchards through an inversion approach based on a look-up table strategy. Lei et al. paved the way for applying the 3D HELIOS ray tracing model to nutrient retrieval tasks by demonstrating its ability to generate realistic synthetic canopy imagery in the RGB domain. Their work used HELIOS-generated images to train object detection models

for tasks such as bean leaf localization and strawberry counting, illustrating the potential of 3D RTMs to reduce the resource demands of collecting large, high-quality training datasets. However, the HELIOS model has yet to be tested in the MS domain or for biochemical trait retrieval using plant spectroscopy.

While RTMs require precise input parameters and are computationally intensive, particularly in the inverse mode, their theoretical rigor and biophysical interpretability make them a valuable framework for generalized N retrieval. Furthermore, hybrid approaches that integrate physically-derived traits into machine learning workflows have shown promise in enhancing transferability across environmental conditions and agroecosystems [10], [56], [69], [71].

Chapter 3: Nitrogen retrieval in grapevine (Vitis vinifera L.) leaves by hyperspectral sensing

Background

RS has emerged as a powerful tool for estimating plant biochemical traits, including N, offering an alternative to traditional destructive sampling methods that are labor-intensive, costly, and limited in spatial coverage [8], [9], [10], [11], [34]

Despite its potential, RS-based N estimation at the canopy level faces several challenges. Variability introduced by sun-sensor-target geometry, background reflectance, canopy structure, and atmospheric interference can significantly obscure the relationship between reflectance and leaf N content[10], [33], [34]. These confounding factors limit the transferability and reliability of models built solely on canopy-level data. As a result, leaf-level spectral analysis has become a crucial focus area for improving the mechanistic understanding of N retrieval from RS data.

At the leaf scale, reflectance spectra are less affected by structural and external noise, enabling clearer associations between biochemical content and spectral response. However, commonly used vegetation indices (VIs), such as NDVI or NDRE, often saturate in dense vegetation and primarily reflect chlorophyll content, which comprises only about 1.7% of total leaf N [14]. This limitation is particularly evident in younger leaves, where chlorophyll levels may be low despite high N content, leading to underestimation of true N status [9].

Machine learning (ML) approaches, such as partial least squares regression (PLSR), random forest, and Gaussian process regression, have been employed to model N from hyperspectral data with high accuracy but often require local calibration and are sensitive to data domain shifts [56]. In contrast, RTMs, such as PROSPECT and PROSPECT-PRO, offer a physics-based framework for

simulating leaf reflectance and estimating N-related traits like leaf mass per area (LMA), chlorophyll, and protein content [21]. These models allow both forward simulation and inverse retrieval, making them highly suitable for evaluating N content under varying physiological and environmental conditions.

Leaf-level datasets offer a controlled framework to investigate the foundational sources of spectral variability that influence canopy reflectance. Factors such as leaf age, structure, and biochemical composition contribute to the complexity of canopy spectral signals. By isolating and characterizing the spectral influence of these individual factors at the leaf scale, we can deepen our understanding of how N interacts with other leaf biochemical parameters. Ultimately, this knowledge supports the development of more robust, transferable nutrient estimation algorithms that account for natural variability in canopy structure and composition.

Therefore, investigating N estimation using hyperspectral sensing at the leaf level provides a more controlled environment for understanding spectral-biochemical relationships. The insights gained from these controlled leaf-scale studies are foundational to developing robust, generalizable N monitoring models at canopy level.

Materials and Methods

Plant Materials and Study Sites

Potted Vines

On 22 April 2021, dormant bench grafted ‘Flame Seedless’ grapevines (*Vitis vinifera* L.) on Freedom rootstock were planted into 37-liter polyethylene pots filled with a substrate consisting of sphagnum peat, perlite, and coarse sand in a 1:2:0.5 ratio. The peat portion of the substrate was amended with approximately 40 g dolomitic lime per 3.785 l of peat. The exterior of each pot was painted white to minimize substrate temperature, and the pots were placed on an unshaded gravel

bed outdoors at the University of California Agricultural and Natural Resources Kearney Research and Extension Center in Parlier, CA. At planting, the substrate in each pot was irrigated to field capacity with a multi-outlet drip ring (NetBow™, Netafim USA, Fresno, CA). The potted Flame Seedless vines were weighed then and periodically after bud break (3 May 2021) to estimate evapotranspiration. Vines were irrigated as needed to avoid water deficit stress, using estimated daily evapotranspiration as a guide for water replacement.

In late May, vines of uniform appearance were randomly assigned to receive one of three N fertigation regimes, which consisted of different modified Hoagland's nutrient solutions, similar to those described by Schreiner and Scagel (2017), except the N amount applied to pots was either 7.50, 3.75, or 1.12 mM. Each vine received 3 L of solution weekly (1L per day, three days/week) through August. The three N fertigation regimes aimed to induce a range of leaf N contents from insufficient to sufficient. Each vine was trained to a single shoot tied to a bamboo stake. Each stake extended approximately 1.2 m above the substrate, and any additional growth beyond the height of the stake was allowed to become a pendant. Two datasets were obtained from potted vines in June and July 2021. Nine potted vines (three replicates of each treatment) were collected in each harvest.

Field-grown vines

Sobrio table grapevines (*Vitis vinifera L.*) grafted to Freedom rootstock were planted on 3 March 2021 in a vineyard at the Kearney Research and Extension Center. The vines were spaced 2.13 m within rows and 3.66 m between rows at planting. Plots of five consecutive vines in a row were assigned to one of four N fertilizer treatments, replicated ten times. The treatments included split applications of 0, 12.5, 25, or 50 g N per vine annually, half from $CaNO_3$ and the remainder from $(NH_4)_2SO_4$. In 2021, a bud break occurred on 24 March 2021, and then vines were trained up

bamboo stakes and onto open-gable trellises like those described in Williams and Fidelibus (2016). Fertilizers were applied on 26 April and 18 May 2021. In January 2022, each vine was pruned to four canes per vine that were tied onto parallel horizontal wires that bracketed the vine trunks. Budbreak occurred on 14 March 2022, and fertilizers were applied on 11 April and 19 May 2022. In 2022, Solbrio grapevines were irrigated at approximately 70% crop evapotranspiration, using degree days from the Parlier CIMIS station to estimate the crop coefficient and ET_o described by Williams et al.

Leaf sample collection and processing

In the 'Flame Seedless' potted-vine study, nine vines, three from each N fertigation regime, were destructively harvested at each sampling campaign on 29 June 2021 and 27 July 2021. All the primary leaves on each vine's single shoot were collected, clipped from their petioles with shears, placed into individual prelabelled zip-top mylar bags, and stored in a cooler with ice until measurements could be completed. All measurements were completed when the samples were collected at each sampling campaign. Leaves were collected and numbered from the oldest leaf at the base of a shoot, acropetally to the youngest leaf at the tip. Leaves with damaged leaf blade areas wider than 2.54 cm in diameter were ignored because damaged areas, such as holes, will alter the spectral signature. The leaf position (order from the base) was recorded and used to rank leaves by age. Damaged leaves were discarded from the dataset, but their position was recorded to ensure the leaf order number reflected accurate leaf age. A similar approach was taken with 'Solbrio,' where a representative shoot was collected on 16 June 2022 from a vine in two treatment replicates. All the leaves on each shoot were collected like that described for 'Flame Seedless.' Across three leaf sampling campaigns, we collected 664 leaf samples. The breakdown is 182 from Flame Seedless I, 287 from Flame Seedless II, and 195 from Solbrio.

Proximal Hyperspectral Scan and Ground Truth Data

The fresh weight of each leaf was determined by weighing it on a digital scale, tared to the mylar bag weight, with an accuracy of 0.0001 grams. The area of each leaf was then measured with an LI-COR leaf area meter (LI-3100, LI-COR Bioscience, NE, USA) with an accuracy of 0.01 cm^2. The proximal spectral reflectance of leaves was then collected by a field spectrometer, SVC HR-1024i (Spectra Vista Corp, NY, USA), a contact probe spectrometer with a spectra range of 350-2500 nm. The spectrometer is integrated with a leaf scanner with a built-in light source to provide constant light intensity. Undamaged leaf surfaces near the base of each leaf were used for spectral scans. To maintain rigorous accuracy and consistency in reflectance measurements, we employ a white spectralon disk for calibration. The instrument also features an automated dark current correction mechanism. Our measurement protocol mandates intermittent reference reflectance assessments of the spectralon disk, executed after every quintet of target leaf measurements. Initial multi-point measurements on select leaves confirm uniform reflectance across leaf surfaces, thereby justifying subsequent single-point spectral acquisition at the midrib for each leaf sample. The leaves were then rinsed with deionized water, placed in labeled coin envelopes and dried to a constant weight in a forced-air drier set to 55 ºC. Dried samples were ground in a mill and submitted to a commercial laboratory (Valley Tech Ag Laboratory Services, Tulare, CA) to determine total N using the automated combustion method. Table 1 shows the formula for calculating the derived ground truth data label.

For subsequent analysis, hyperspectral scan data is converted to CSV format. Overlapping bands are eliminated in two specific spectral regions—971-1010 nm and 1893-1911 nm. This process addresses the issue of spectral overlap arising from Gaussian-shaped response functions in the detector array.

Table 1 Derived Ground Truth Information

Variable	Formula	References
Leaf Mass Area	$\frac{Leaf\ dry\ mass\ [g]}{Leaf\ area\ [cm^2]}$	[72]
Equivalent Water Thickness (EWT)	$\frac{Leaf\ fresh\ mass\ [g] - Leaf\ dry\ mass\ [g]}{Leaf\ area\ [cm^2]}$	[73]
Leaf N per area (N_{area})	$\frac{N\% \times Leaf\ dry\ mass\ [g]}{Leaf\ area\ [cm^2]}$	[74]
Relative Leaf Age	$1 - \frac{Leaf\ Position\ No.}{Highest\ Leaf\ Position\ No.}$	

Figure 1 shows the range of ground truth N labels generated by the per-leaf ground truth data (lab-based N measurement).

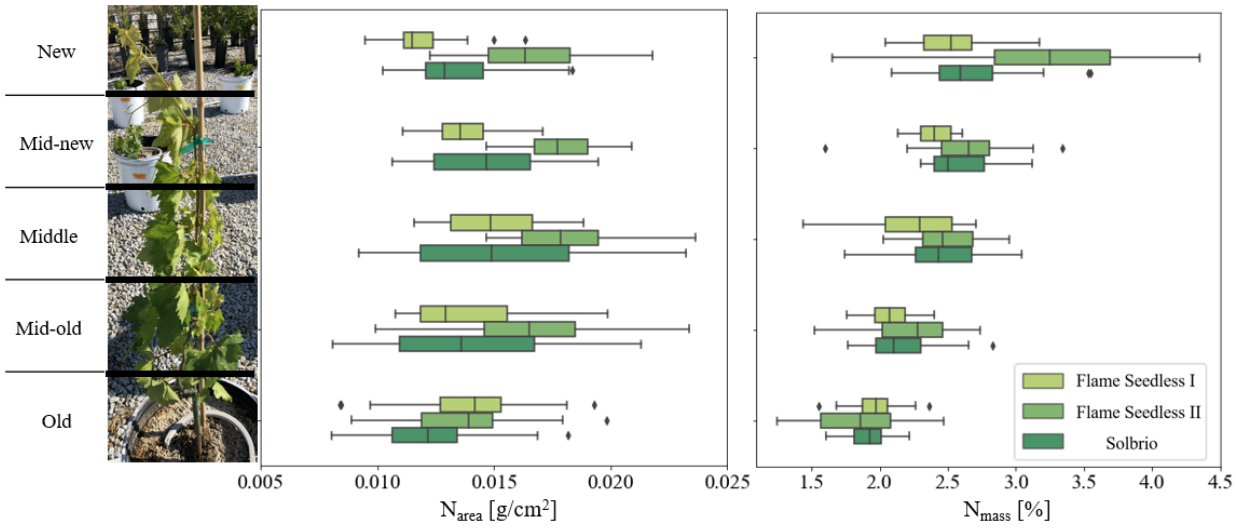


Figure 1 The distribution of N_{mass} and Narea ground truth data. The potted Flame Seedless II dataset has the widest range of N_{mass} and Narea. N_{mass} appeared to be the lowest in the oldest leaves near the base of the shoots.

N retrieval approaches

We compared the performances of five main N retrieval approaches: linear parametric regression, non-linear parametric regression, non-linear non-parametric regression, physically-based, and a hybrid approach that combines RTM and data-driven methods[17]. All three datasets (two potted vines dataset and one 'Solbrio' veraison dataset) were combined and evaluated to test each approach's consistency.

Linear Parametric Regression Approach

Multiple VIs were selected from previous studies that explored the use of VI for canopy N content estimation in various crops. A complete list of VIs used in this study is listed in Table 2. Each VI is then used individually as a primary regressor to predict N_{mass} and N_{area}. The VIs were selected based on a Tian et al. study that assessed multiple hyperspectral VIs most effective for N monitoring in rice [75]. The coefficient of determination (R^2), Root Mean Squared Error (RMSE), and Normalized RMSE (NRMSE) of each index are reported.

Table 2 VIs used for estimation of N content

Index	Formula	Reference
Normalized Difference Red Edge Index (NDRE)	$\frac{R_{790} - R_{720}}{R_{790} + R_{720}}$	[76]
MERIS Terrestrial Chlorophyll Index (MTCI)	$\frac{R_{754} - R_{709}}{R_{709} - R_{681}}$	[77]
Modified DATT index (MDATT)	$\frac{R_{721} - R_{744}}{R_{721} - R_{714}}$	[78]
Normalized Difference Nitrogen Index (NDNI)	$\frac{\log\left(\frac{1}{R_{1510}}\right) - \log\left(\frac{1}{R_{1690}}\right)}{\log\left(\frac{1}{R_{1510}}\right) + \log\left(\frac{1}{R_{1690}}\right)}$	[79]

Red Edge Normalized Difference Vegetation Index (RENDVI)	$\frac{R_{750} - R_{705}}{R_{750} + R_{705}}$	[80]
Transformed Chlorophyll Absorption Ratio (TCARI)	$3\,[(R_{700} - R_{670}) - 0.2(R_{700} - R_{550})\left(\frac{R_{700}}{R_{670}}\right)]$	[81]
Normalized Difference 570/530 Photochemical Reflectance Index (PRI)	$\frac{R_{531} - R_{570}}{R_{531} + R_{570}}$	[82]
Modified Red Edge Normalized Vegetation Index (MRENDVI)	$\frac{R_{750} - R_{705}}{R_{750} + R_{705} - 2R_{445}}$	[83]
Vogelmann Red Edge index 2 (VREI2 or VOG2)	$\frac{R_{734} - R_{747}}{R_{715} + R_{726}}$	[84]
Anthocyanin Reflectance Index 2 (ARI2)	$R_{800}\,(\frac{1}{R_{550}} - \frac{1}{R_{700}})$	[85]
Modified Chlorophyll Absorption in Reflectance Index 2 (MCARI2)	$\frac{1.5[2.5(R_{800} - R_{670}) - 1.3(R_{800} - R_{550})]}{\sqrt{(2R_{800} + 1)^2 - (6R_{800} - 5\sqrt{R_{670}}) - 0.5}}$	[86]
Simple Ratio (533,565) (SR533)	$\frac{R_{533}}{R_{565}}$	[75]
Modified Simple Difference 705 Index (R705)	$\frac{R_{705}}{R_{717} + R_{491}}$	[75]
Modified Simple Difference 434 Index (R434)	$\frac{R_{434}}{R_{496} + R_{401}}$	[75]

Linear Non-parametric Regression Approach

PLSR was deployed on a combined hyperspectral dataset of all three campaigns. The Savitzky-Golay (SavGol) filter was employed with a window size of 17 and a polynomial order of 2 as a

preprocessing step before PLSR analysis. This filter reduced noise and variability in the dataset and helped to enhance the predictive power and accuracy of the PLSR model. The number of PLSR components was optimized, and the performance of PLSR regression was determined similarly to VIs regression performance. For hyperparameter optimization, the combined dataset were partitioned into training and test sets, maintaining a proportion of 60% for training data and 40% for testing data. This division method ensured a substantial training dataset for model development while allowing a rigorous assessment of model performance on unseen data. GridSearchCV from the scikit-learn package was used for the training dataset to determine the number of optimal components. The optimum number of PLSR components was 32 bands for mass-based and 24 bands for area-based Nitrogen estimation.

We employed a 5-fold cross-validation strategy to assess the PLSR model, fine-tuned with optimal hyperparameters on the combined dataset. Evaluation metrics such as R2, RMSE, and NRMSE were used to assess the model's performance. This approach offered predictions for each data point and ensured that the model was trained on diverse data. This methodology is a robust measure of the model's generalization capabilities on new, unseen data and aids in identifying potential overfitting by evaluating the model across multiple data subsets.

Non-linear Non-parametric Regression Approach

In this study, the spectral reflectance profile of each leaf was used to train GPR and RF. In the RF model, the reflectance dataset was preprocessed by merging adjacent wavelengths with a correlation higher than >0.95, as per the adaptive downscaling approach outlined by [42]. This reduced the initial 994 wavebands to 62 spectral windows to construct the RF model.

For the GPR model, hyperparameter tuning was conducted via grid search on the training dataset. This resulted in selecting a composite kernel comprising constant kernel, Radial Basis Function

(RBF), and white noise components. This kernel configuration is commonly employed in GPR to capture systematic patterns (via the constant and RBF components) and random noise (via the white noise component). Similar to the PLSR model, a 5-fold cross-validation strategy was implemented to rigorously train and evaluate the model for both mass-based and area-based N estimation.

Physical-based Model Inversion Approach

In this experiment, we employed PROSPECT PRO in the inverse mode to estimate the leaf primary biochemical and biophysical traits of each leaf. Each leaf's biochemical traits were then used as a regressor for estimating the N content of the leaves. As a pre-processing step, the reflectance values for each sample are resampled to create a uniform wavelength interval of 1 nm. This standardized wavelength interval is a prerequisite for the input data used in PROSPECT PRO. The inverse model can predict the following leaf traits: Carbon-based Constituents (CBC), Protein (PROT), Equivalent Water Thickness (EWT), Chlorophyll A and B (CHL), Carotenoids (CAT), Anthocyanin (ANT), Brown Pigment (BROWN), and Leaf Structure (N_{Struct}). The RTM algorithm estimated these parameters in an area-based unit (g/cm^2) by default. We converted the area-based output to mass-based traits by dividing them by the actual LMA from ground truth data.

Hybrid Approach

A hybrid approach is defined as a combination of RTM and data-driven approaches [87]. For instance, simulated leaf biochemical parameters derived from RTM can be used to train a machine learning model, such as RFR and GPR. The spectra profile of the sample is then fed into the RTM, and leaf biochemical traits are estimated. The estimated traits are then used to train RFR or GPR models for N content prediction [88]. In this study, we coupled GPR and RFR with leaf biochemical traits estimated by PROSPECT PRO to predict N_{mass} and N_{area}. The parameters from PROSPECT PRO used in the hybrid model are PROT, EWT, ANT, CHL, CAT, and N_{struct}. For

RFR, the feature ranking importance was extracted to understand each trait's role in N retrieval. The data splitting and cross-validation techniques used were similar to the PLSR method.

Results

Ground truth data

As demonstrated in Figure 1, younger leaves exhibit higher N_{mass}, even though the estimated Chlorophyll labels are lower in younger leaves than in older ones. Medium-aged leaves tend to have the highest level of N_{area}. In all three datasets, 'Solbrio' has the biggest range of N_{area}. For $N_{mass,}$ the leaves at the middle part of the shoot tend to best represent the shoot's average N content. For N_{area} the trend is unclear.

The relationship between N_{mass} and N_{area} is generally considered to be linear (shown in Figure 2). The potted Flame Seedless II dataset shows a higher number of younger leaves with a non-linear relationship between N_{mass} and N_{area}. Solbrio, grown in the field, shows the most linear relationship between N_{mass} and N_{area}. More information is required to confirm whether the differences observed were due to management practices or varieties. N_{mass} appears to be less variable than N_{area} for leaves in most age classes except the newest leaves of Flame Seedless in the second data collection campaign. It is unclear why the newest leaves in that dataset had particularly high variability in N_{mass}. Still, the youngest leaves grow rapidly, which might lead to rapid changes in N content. Similar variations in NO_3-N content across short timescales were reported in 'Thompson Seedless' petioles collected in early spring [23].

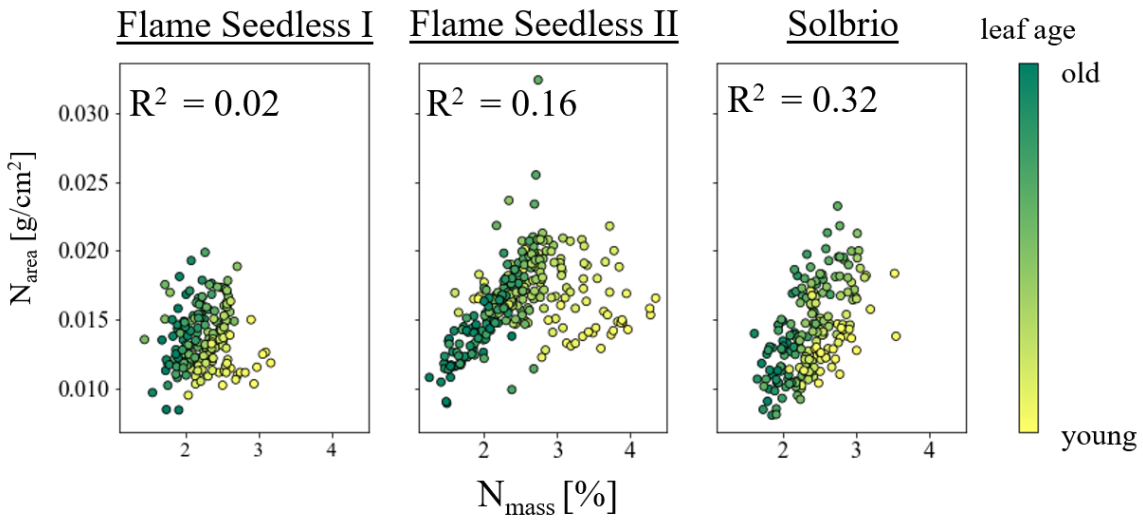
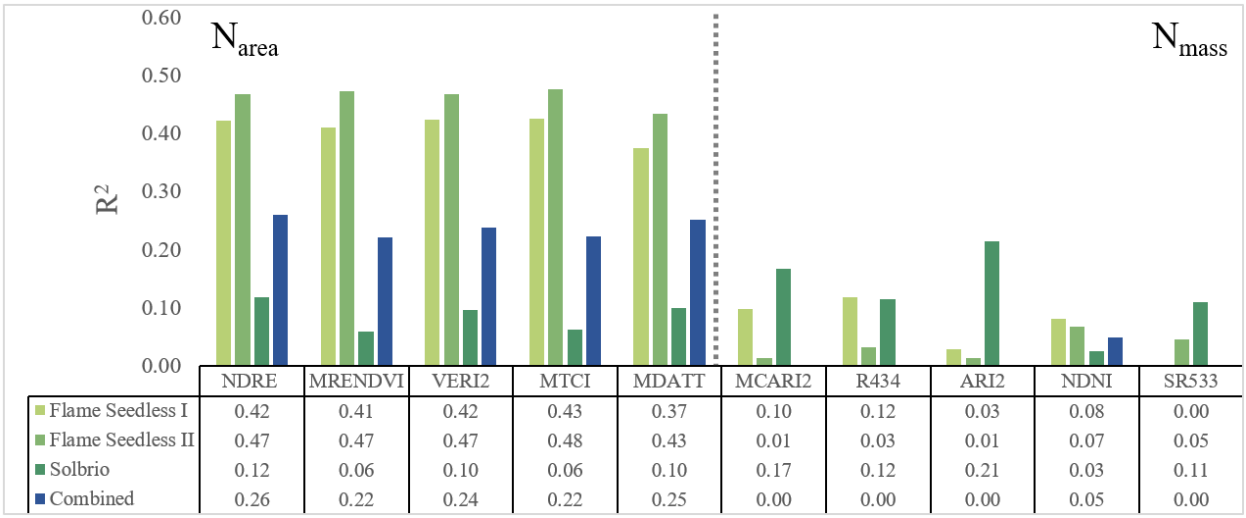


Figure 2 The Nmass Vs. Narea scatter plot showing the general correlation between Nmass and Narea. 'Solbrio' leaves had a linear relationship between Nmass and Narea regardless of leaf age.

VI performance and inconsistent relationship between chlorophyll and N

Figure 3 shows the predictive accuracies of N_{mass} and N_{area} as obtained by the top five VIs. When calibrated by N_{area}, the VIs demonstrated enhanced performance within the potted 'Flame Seedless' dataset, whereas the 'Solbrio' dataset exhibited marginally improved performance when calibrated with N_{mass}. In contrast to the Flame Seedless dataset, the combined dataset experienced a decline in the R^2 value by approximately 20 percent when VIs were employed to estimate N_{area}, and the value was predominantly near zero when VIs were utilized to estimate N_{mass}. This underlines the inherent limitations in the generalizability of VI-based N retrieval across varying temporal, spatial, and cultivar contexts.



	NDRE	MRENDVI	VERI2	MTCI	MDATT	MCARI2	R434	ARI2	NDNI	SR533
Flame Seedless I	0.42	0.41	0.42	0.43	0.37	0.10	0.12	0.03	0.08	0.00
Flame Seedless II	0.47	0.47	0.47	0.48	0.43	0.01	0.03	0.01	0.07	0.05
Solbrio	0.12	0.06	0.10	0.06	0.10	0.17	0.12	0.21	0.03	0.11
Combined	0.26	0.22	0.24	0.22	0.25	0.00	0.00	0.00	0.05	0.00

Figure 3: Performance of top five vegetation indices on predicting Nmass and Narea for each dataset separately as well as a combined dataset. Performances of VIs are high for Narea prediction in 'Flame Seedless' I and II, despite low performance in 'Solbrio'. Combining the datasets lowered the performance of VIs, suggesting challenges in applying VIs universally on different datasets.

Figure 4 shows the high correlation (average Pearson's R 0.98) between Chlorophyll estimated by PROSPECT PRO and NDRE in each dataset. However, the trend between Chlorophyll and N_{mass} is inconsistent through different datasets and leaf age. Older leaves at the bottom of the shoot show a high correlation with Chlorophyll, but younger leaves have high N_{mass} even though they show relatively low NDRE. Figure 6 shows chlorophyll concentration increases along with leaf age before dropping in a downward bell curve shape. The relationship in Figure 5 suggests that $Chlorophyll_{area}$ also shares a similar downward bell curve shape relationship with leaf age, which is more pronounced and evident within the Flame Seedless datasets.

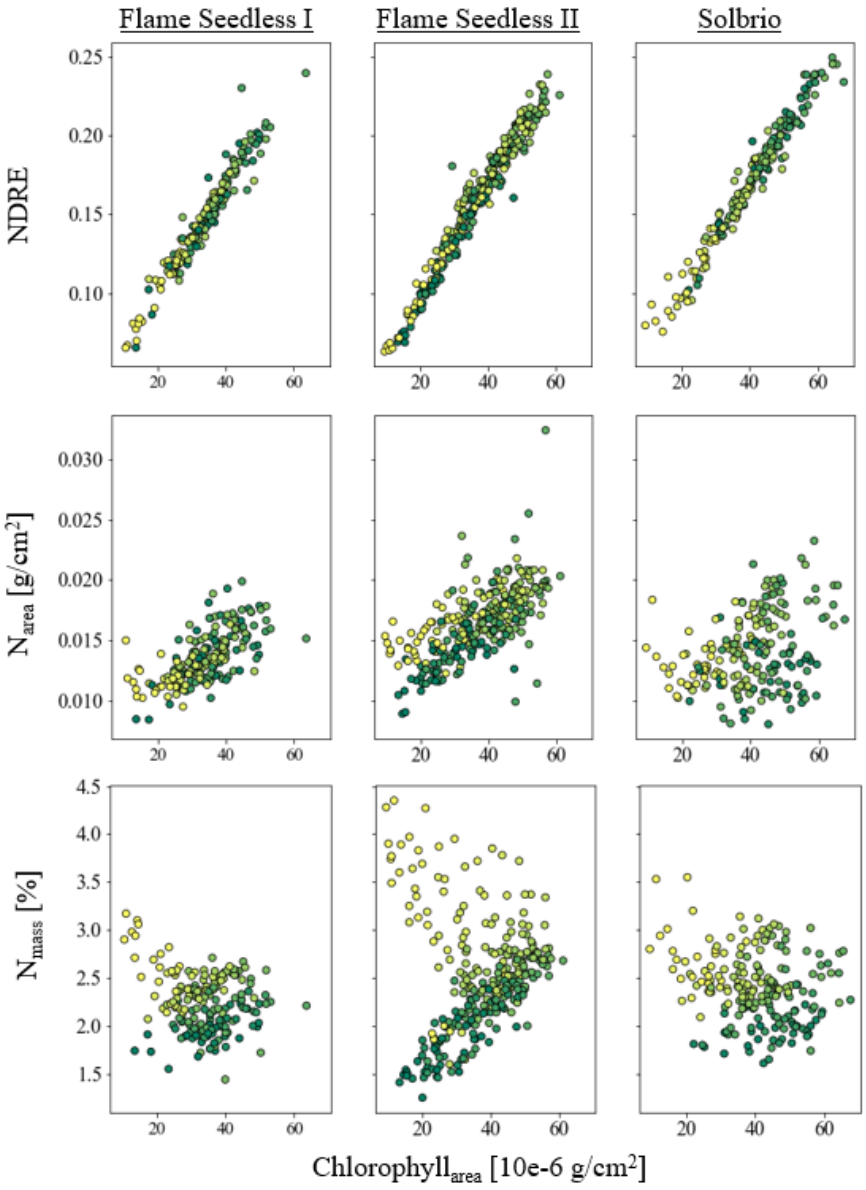


Figure 4: Correlation between Chlorophyll estimated by PROSPECT PRO using reflectance, NDRE index, Nmass, and Narea.Chlorophyll, which is highly correlated with NDRE, showed inconsistent with trends with both Nmass and Narea. NDRE and Chlorophyll. which is highly correlated with NDRE, showed inconsistent whnt trends with both N_{mass} and N_{area}. NDRE and Chlorophyll$_{area}$ show a high correlation, while Chlorophyll$_{area}$ shows no significant relationship with N_{mass}.

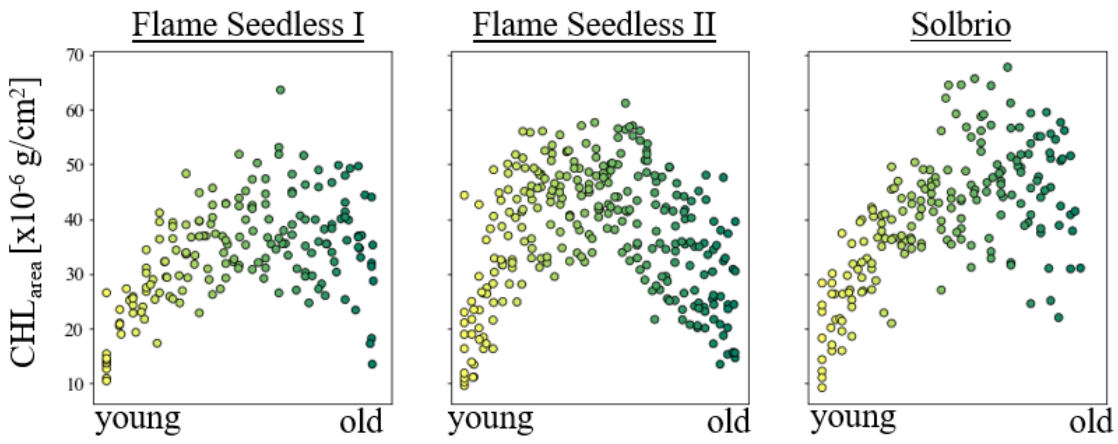


Figure 5: Estimated Chlorophyll$_{area}$ trend found across potted Flame Seedless and Solbrio vines showing a bell shape curve relationship between chlorophyll and leaf age.

Estimated RTM parameters and their correlation with N_{mass} and N_{area}

The accuracy of EWT and LMA prediction by RTM was determined with the ground truth data. The NRMSE of EWT and LMA were 19% and 24 %, respectively. The R^2 value between Protein$_{mass}$ and ground truth protein content obtained by the 6.25 conversion factor of TN is 0.45. The NRMSE level was similar to what [89] where the NRMSE of estimated EWT and LMA using only reflectance with prior N_{struct} estimation was 21.4% and 38.9%, respectively. To visually represent the correlations, we constructed a correlation heatmap. In this heatmap, each cell color corresponds to the magnitude and direction of the correlation between two specific leaf biochemical parameters, N_{mass} and N_{area}. Mass-based leaf biochemical traits showed a higher correlation with $N_{mass,}$ and area-based leaf biochemical parameters showed a higher correlation with N_{area}. Protein$_{mass,}$ and EWT$_{mass}$ show the strongest correlation with N_{mass} (see Figure 7), while Chlorophyll$_{area}$ and Protein$_{area}$ show the highest correlation with N_{area}. Figure 6 shows a consistent relationship of Protein$_{mass}$ for N_{mass} prediction across all three datasets, while Chlorophyll$_{mass}$ - showed inconsistent prediction power for $N_{mass.}$ Interestingly, N_{struct}, which represents fundamental leaf physical parameters, did not appear to be correlated with either N_{mass} or N_{area}.

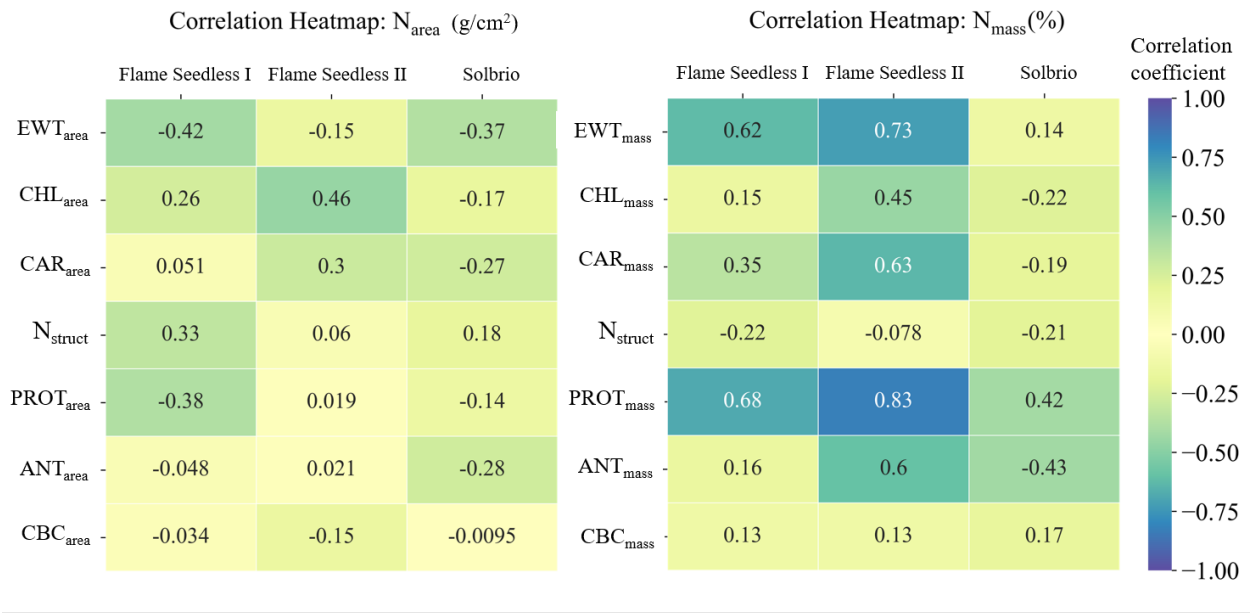


Figure 6: Correlation coefficient between direct RTM biochemical parameters and Nmass and Narea in each dataset.
Chlorphyll, both mass-based and area-based, shows an inconsistent relationship with both N_{mass} and $N_{area,}$ while $Protein_{mass}$ show a consistent positive correlation with N_{mass} across all datasets.

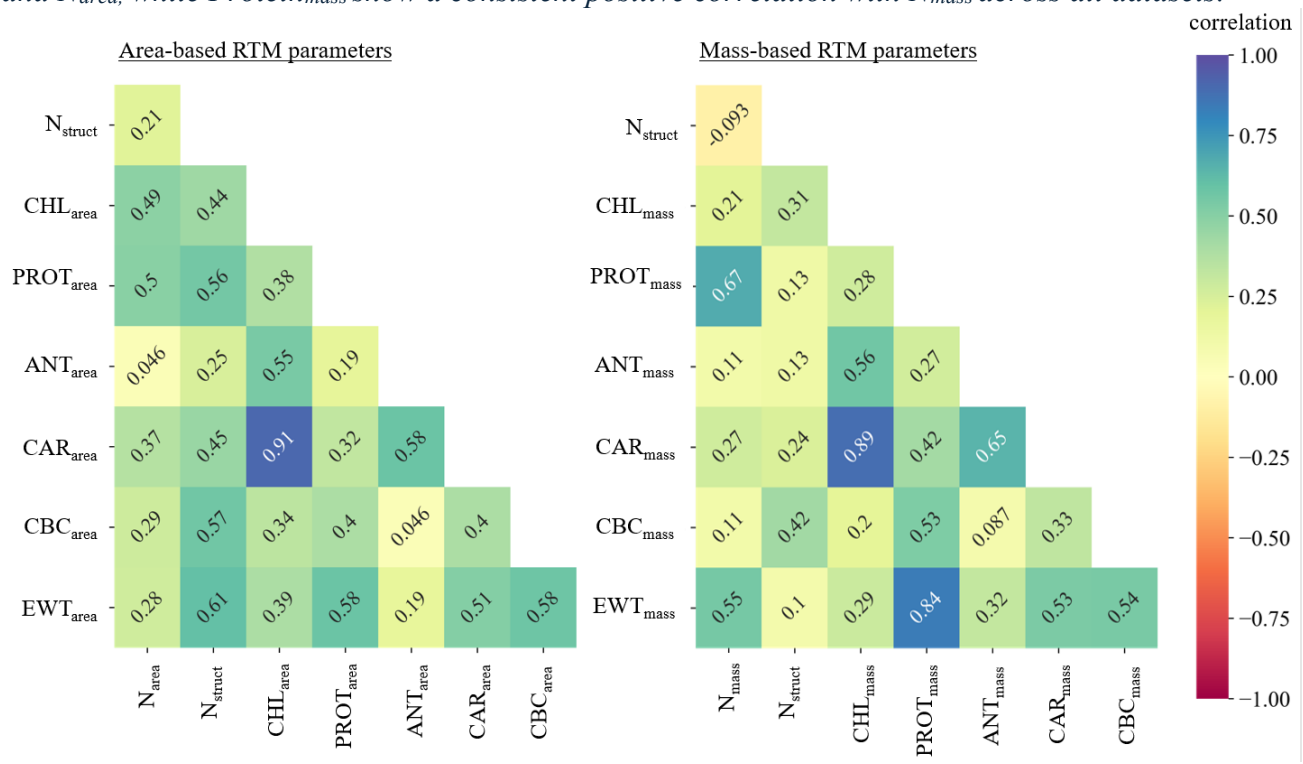


Figure 7: Correlation heatmap between mass-based and area-based PROSPECT-estimated leaf biochemical traits to Nmass, Narea, and leaf age.

Performance comparison between single input approach and multiple inputs approach

Figure 8 shows the performance of various N retrieval methods. Chemometric (PLSR) and ML (GPR) approaches, which require the entire 350-2500 nm reflectance spectrum, showed the best R^2 value and NRMSE performance. While calibration by N_{area} performed better in the VIs approach, N_{mass} was better predicted by PLSR, GPR, Physical modeling, and hybrid models using RFR. VIs and Physical-based approach R^2 values are based on simple linear regression using only one parameter at a time to make predictions. In contrast, GPR, PLS, and hybrid approaches use multiple inputs to make a prediction. It is important to note that the VI approach fits well with older leaves. Still, the R^2 value dropped significantly because of the irregular behavior of younger leaves shown in light green in N_{mass} prediction.

GPR, PLSR, and hybrid approaches show the top three highest R^2 values using a 5-fold cross-validation technique, which is expected since they take more inputs than single input approaches, namely VIs and a Physical-based model. The hybrid method that trains RF with leaf biochemical parameters estimated by PROSPECT showed good performance, although not comparable to PLSR and GPR. The hybrid model offered about 20% less accuracy in the cross-validated result and 5% higher NRMSE compared to chemometrics and ML approaches. Still, it uses a subset of the hyperspectral data instead of the full spectrum. The physical-based approach using $Protein_{mass}$ as a predictor of N_{mass} showed high potential as it is capable of predicting N with an R^2 value up to 0.45 using only one input. The hybrid approach trained on mass-based RTM parameters only increases the N_{mass} prediction accuracy by 9% but uses six more variables. The scatter plot of GPR, PLSR, and hybrid approaches for N_{mass} prediction show higher variation in younger leaves because there are fewer data points to train the model. In N_{mass} values beyond 3.5%, there is more non-

linear spectral signature behavior similar to the downward bell-shaped curve relationship found between chlorophyll and leaf age.

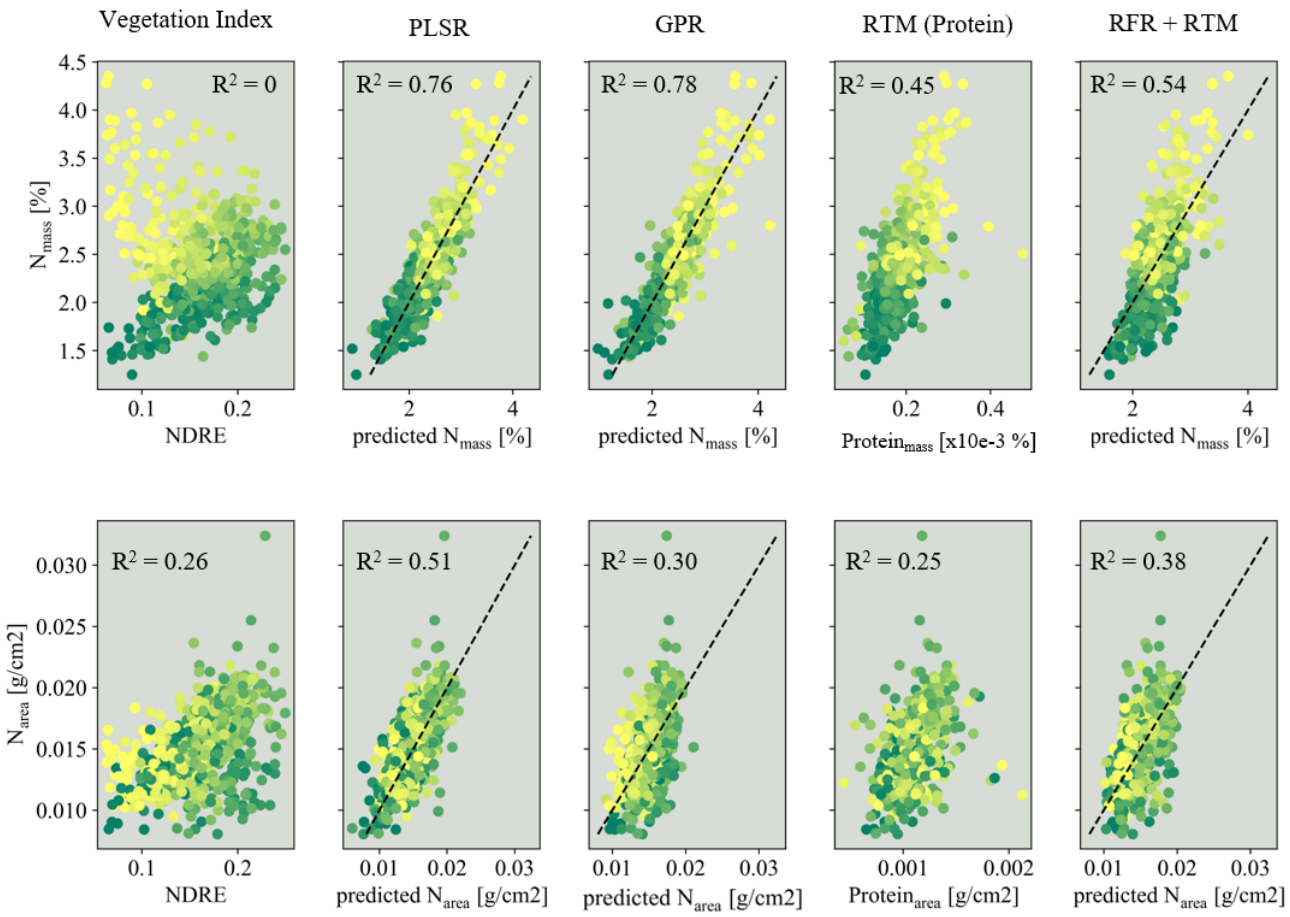


Figure 8: Performance comparison of major N retrieval approaches from all three collection campaigns. Lighter green reflects younger leaf age, while darker green reflects older age.

Figure 9 shows the importance of $Protein_{area}$ and $Protein_{mass}$ in the decision-making process of the RF model among all eight leaf traits estimated by PROSPECT inversion. All biochemical traits from RTM are normalized by area by default. $Protein_{mass}$ was calculated using the ground truth leaf LMA. N_{mass} clearly showed a more consistent relationship with $Protein_{mass}$ compared to $Chlorophyll_{mass,}$ which showed a similar irregularity that also appeared in traditional VI, such as

NDRE. Nonetheless, $Chlorophyll_{mass}$ showed the second highest feature importance after $Protein_{mass}$ which meant that Chlorophyll content still plays a significant role in N_{mass} prediction. For N_{area} prediction, $Protein_{area}$, $Chlorophyll_{area}$, and $Anthocyanin_{area}$ are the top three relevant features.

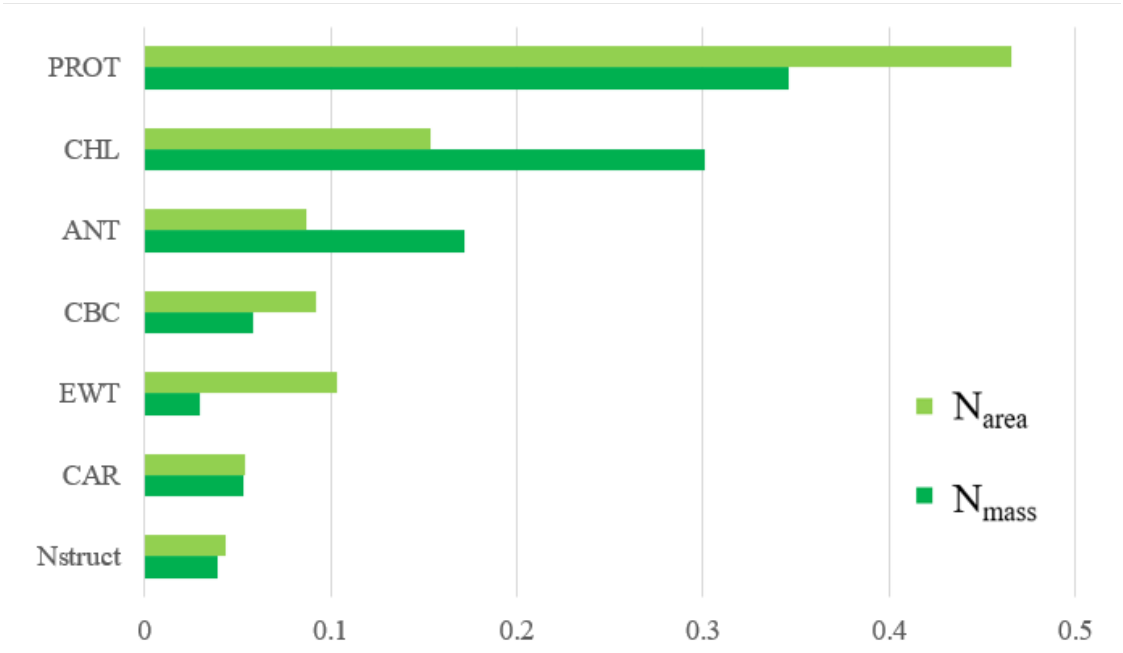


Figure 9: Feature importance of RFR+RTM hybrid models for Nmass and Narea prediction. Protein is highly important in both Nmass and Narea prediction in hybrid models.

Table 3 shows a reduction in N prediction accuracy once the waveband was reduced from full spectrum (350-2500 nm) to only visible-near infrared range (VNIR), commonly used for aerial spectral sensing at the canopy level. GPR and PLS suffer a reduction of 31.5% drop in performance once shortwave-infrared (SWIR) is dropped.

Table 3 Change in R2 values for data-driven approaches when the available spectral data is reduced from VNIR+SWIR to only VNIR

Method	R^2 350 – 2500 nm (VNIR+SWIR)	R^2 350-900 nm (VNIR)	ΔR^2
GPR on N_{mass}	0.78	0.38	-0.40
GPR on N_{area}	0.30	0.23	-0.07
PLS on N_{mass}	0.76	0.41	-0.35
PLS on N_{area}	0.51	0.36	-0.15

Discussion

Mass-based Vs. area-based leaf biochemical parameters

The potted 'Flame Seedless' II dataset exhibited the widest range of the N_{mass} and N_{area} labels, which could be attributed to the prolonged differential N treatment applied compared to Flame Seedless I. The bell curve shape relationship between leaf age and chlorophyll content (Fig. 5) resonates with findings by Poni et al [90]. The scatter plot depicting the relationship between mass-based PROSPECT estimated protein content and N_{mass} highlights the potential of using protein content as a complementary factor alongside chlorophyll content and traditional VIs for N retrieval. Notably, $protein_{mass}$ displays a consistent positive correlation with $N_{mass,}$ whereas $chlorophyll_{mass}$ still exhibits an irregular trend in younger leaves. In the PROSPECRT-PRO model, protein content estimation becomes feasible using optimal waveband between 2100-2140 nm and 2160-2180 nm, a considerably smaller bandwidth (60 nm) than the entire hyperspectral spectrum (350-2500 nm)

employed in PLSR or GPR. The relationship between N_{mass} and leaf age shows a dilution effect similar to observations made by others (Williams, 1987).

Regarding N_{area} performance, the VIs approach stands out due to their alignment with strong correlations between N and visible range VIs (400-700 nm), where chlorophyll content exerts a pronounced spectral signature. The apparent decline in VIs' efficacy in N_{mass}, prediction might be attributed to the non-linear association between chlorophyll and leaf age, leading to corresponding non-linearity between chlorophyll-based VIs (e.g., NDRE) and N_{mass}. Consequently, the normalization of high N_{mass} in younger leaves by smaller mass and leaf area culminates in reduced N_{area} , fostering a generally linear relationship between N and Chlorophyll, as evidenced in Figure 5. This characteristic enhances the performance of single VIs in predicting N through simple linear regression.

Figure 7 shows a high correlation consistency of $Protein_{mass}$ for N_{mass} prediction compared to inconsistency in $Chlorophyll_{mass}$. This supports Berger's suggestion that protein might be a good candidate for N_{mass} retrieval. The significant level of correlation between EWT and N_{mass} could be because EWT_{mass} is highly correlated with $Protein_{mass}$ (Figure 8). The absorption bands used to retrieve Protein with PROSPECT PRO are highly affected by EWT [21] which could be the reason for the high correlation between EWT and Protein. The R^2 value between $Protein_{mass}$ and EWT_{mass} is 0.71. Therefore, future studies should aim to understand the relationship between EWT and N because EWT predicted by PROSPECT can be affected by multiple other factors, such as irrigation and whether the leaf spectral profile was collected from fresh or dry leaf.

The parameter "N_{struct}" characterizes the internal complexity of a leaf's structure. A lower value (1–1.5) suggests a simpler compact mesophyll tissue, often observed in monocots. On the other hand, a high value (1.5–2) signifies a more intricate mesophyll structure with greater intercellular

air, typically found in eudicots [21]. Initially, it was anticipated that N_{struct} would exhibit significant importance due to its reflection of leaf structure, which varying N levels could influence. Surprisingly, this study found that N_{struct} exhibited minimal feature importance and correlation with N. This discrepancy might stem from the intricate interplay between N and leaf structure, potentially influenced by additional factors. For instance, in some monocot plants such as tall fescue, increased N application has been linked to an enhanced leaf mesophyll area [91] while in rice, it led to thinner leaves and reduced chlorophyll content [92]. Additionally, other biochemical parameters, such as Protein, may undergo more substantial changes compared to leaf structure in response to N treatment. The intricate relationship between N and leaf structure warrants further investigation to fully comprehend the underlying dynamics.

The N_{mass} prediction is about 15-40% more accurate than N_{area} prediction, except when using the VIs approach, where N_{mass} performs poorly across different indices and datasets. Berger et al. and Kattenborn et al. suggested that using N_{area} could be more suitable for upscaling from leaf to canopy level because N_{area} does not require dry matter content. However, from the result shown in this study, using N_{area} to monitor plant N status may reduce retrieval accuracy [37], [56]. Leaf area measurement might introduce a higher error margin than leaf dry mass because a leaf can be deformed during data collection, impacting the measurement, especially in large data collection campaigns. In contrast, the dry mass will remain unchanged regardless of the leaf shape. Secondly, using N_{area} to represent leaf N status assumes that N content is uniformly distributed across the leaf surface, which may not be the case.

Trade-off between spectral regions and prediction accuracy

Figure 9 shows that data-driven methods, namely PLSR and GPR, consistently provide higher prediction accuracy than VIs and RTM parameters. However, RTM parameters require

significantly less spectral data while retaining an acceptable and consistent accuracy. In N_{mass} retrieval, $Protein_{mass}$ showed high potential compared to all other RTM traits. On the other hand, in N_{area} prediction, $Chlorophyll_{area}$, $Protein_{area}$, and $Anthocyanin_{area}$ showed high feature importance. N_{area} prediction (R^2) using RTM traits was, on average, 20% less accurate than N_{mass}.

Contrary to the result shown in N_{mass}, $Anthocyanin_{area,}$ which is highly correlated to the $Chlorophyll_{area,}$ showed relatively high feature importance compared to other leaf biochemical parameters when predicting N_{area}. The higher feature importance exhibited by $Anthocyanin_{area}$ could be because of the difference in variety. Solbrio is a black variety, whereas Flame Seedless is a red variety, which might affect anthocyanin levels in leaves. The Pearson's correlation between $Anyhocyanin_{area}$ and $Chlorophyll_{area}$ of Flame Seedless I, Flame Seedless II, and Solbrio was shown to be 0.51, 0.30, and 0.75, respectively. Therefore, $Anthocyanin_{area}$ shows higher feature importance because its value intrinsically indicates the variety of each data point.

Conclusion

The performance of machine learning (GPR) and chemometrics approach (PLS) using the entire 350-2500 nm spectrum outperformed linear regression by VIs, physical-based modeling, and hybrid approaches. However, the hybrid model combining physical-based modeling and machine learning showed high potential in reducing about 50% of the hyperspectral bands required to retrieve N levels while maintaining an acceptable accuracy level. Protein, as one of the RTM traits, shows high potential to be used as a predictor of N_{mass} because protein is consistently correlated with N regardless of leaf age, unlike chlorophyll. In young leaves, chlorophyll level is low while N is high, whereas in old leaves, chlorophyll levels are low and N is low. Thus, chlorophyll has an age-dependent relationship with N, which may reduce the accuracy of VIs as indicators of N. Although leaf age affects the accuracy of N retrieval by VIs, it does not impact the accuracy of the

PLS, GPR, and hybrid approach when hyperspectral data from the entire spectrum of visible, near-infrared, and short-wave infrared is available. Leaf age is highly correlated with N_{mass}. However, leaf age does not have a significant relationship with N_{area}. Furthermore, N_{mass} is consistently proven to be a better indicator of N regarding prediction accuracy. N_{mass} has already been widely used in viticulture because measuring dry mass is less time-consuming than measuring the leaf area.

Based on the results from this study, suggestions for future studies are the following: (1) the effect of EWT on retrieving protein estimation should be studied and understood because EWT_{mass} and $Protein_{mass}$ are found to be significantly correlated. Moreover, the waveband between 2100-2180 nm, which was used for protein estimation by PROSPECT, is highly affected by EWT.

Chapter 4: Advancing Canopy-Scale Nutrient Estimation in Grapevine (*Vitis. Vinifera. L*) with Synthetic Multispectral Data from 3D RTMs

Background

Accurate estimation of canopy-level nutrient status remains a critical challenge in precision viticulture. While remote sensing approaches have increasingly been used to monitor nutrient variability across vineyards, most current methods remain constrained either by empirical calibration requirements or by oversimplified assumptions regarding canopy structure and light interactions. To address these limitations, physically based radiative transfer modeling offers a mechanistic approach for simulating spectral reflectance from vegetation by explicitly modeling the interaction of light with leaf and canopy structures based on fundamental biophysical principles.

Among 3D RTMs, PROSAIL is widely used to estimate leaf biochemical traits such as chlorophyll, water content, and N through simulated reflectance at the leaf and canopy scales [21], [63]. However, the simplified canopy structure assumed in these models, such as uniform leaf distributions or flat homogeneous layers, does not reflect the heterogeneous, clumped geometries commonly observed in real vineyards. As a result, these models are less effective at reproducing realistic canopy-level reflectance in perennial fruit crops such as grapevines and almonds, especially under complex illumination scenarios.

To overcome these limitations, 3D ray-tracing models have emerged as a powerful class of physically based models capable of simulating light transport within fully 3D plant architectures. The HELIOS framework provides an extensible C++ platform for modeling 3D radiative transfer using a GPU-accelerated Monte Carlo ray-tracing engine [93]. HELIOS includes modular architecture with dedicated plug-ins for realistic plant architecture generation and optical modeling,

making it suitable for simulating heterogeneous, dynamic vineyard canopies under varying illumination.

A key capability enabled by HELIOS is the generation of synthetic multispectral imagery under fully controlled structural, biochemical, and environmental conditions. This capability supports the training of deep learning models on large, labeled datasets where the ground truth is precisely known. Recent studies have demonstrated the promise of synthetic reflectance data for training machine learning models to estimate N and other canopy traits, particularly in scenarios where empirical data is limited or challenging to acquire [8], [22], [69], [71]. However, the application of such synthetic-to-real domain transfer in viticulture remains underexplored.

This study builds upon leaf-level RTM by extending it to the canopy scale through HELIOS. The primary aim is to evaluate whether synthetic images produced from HELIOS simulations can be used to train deep learning models that accurately predict canopy-level N content when applied to real-world multispectral data. A critical aspect of this investigation is determining the feasibility and accuracy of domain transfer across synthetic and empirical domains, where differences in reflectance due to canopy structure, background, and illumination may degrade predictive performance.

By combining physically based modeling with deep learning, this work provides a novel framework for simulating, training, and evaluating canopy N estimation under a range of conditions relevant to operational viticulture. The approach is aligned with growers' needs for robust, transferable, and interpretable N monitoring solutions that support sustainable and precise N management across heterogeneous vineyard landscapes.

Material and Methods

Field Site and Experimental Setup

The research was conducted in a commercial table grape vineyard located in Shafter, Kern County, California, planted with *Vitis vinifera* L. Valley Pearl grafted onto Freedom rootstock. Established in 2016, the vineyard employs a quadrilateral cordon training system supported by a Y-shaped open gable trellis. Vines are spaced 1.80 meters apart within rows and 3.65 meters between rows, which are oriented east–west. Each row contains approximately 102 vines, and the vineyard consists of a total of 104 rows.

The experimental plots were situated centrally within the vineyard, specifically in rows 64 and 65. Standard viticultural practices were applied throughout the vineyard, with the exception of N fertilization in the designated experimental area. Instead, a controlled N treatment regime was established, applying three distinct fertilization rates: no N, 42 lbs. actual N per acre (1x rate) and 105 lbs. actual N per acre (2.5x rate). This design was applied to a total of 150 vines distributed across 30 experimental plots, creating a broad range of foliar N concentrations for assessing vine responses under variable N availability.

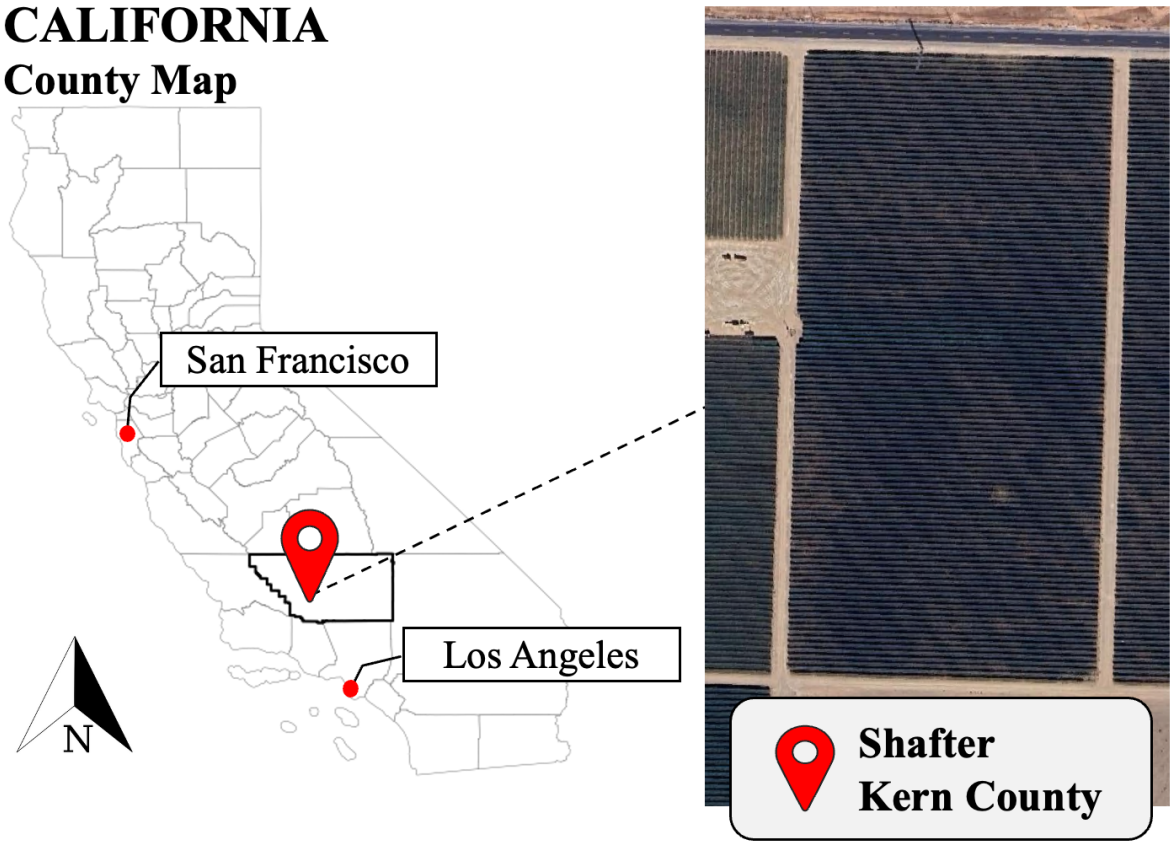


Figure 10: Map of the experimental site at Shafter, CA

Multispectral image campaigns

MS imagery was collected across the growing seasons of 2020 and 2021, encompassing key phenological stages including prebloom, bloom, and veraison. Image acquisition was conducted using a DJI Phantom 4 MS system mounted on a DJI Phantom 4 PRO drone (FC6310, DJI, Shenzhen, China). The MS sensor captures five distinct spectral bands: blue (450 nm ± 16 nm), green (560 nm ± 16 nm), red (650 nm ± 16 nm), red edge (730 nm ± 16 nm), and near-infrared (NIR, 840 nm ± 26 nm). Flights were performed at altitudes ranging from 80 to 95 meters above ground level. Each image has a spatial resolution of 1600 × 1300 pixels.

The sun zenith angle at the time of image capture ranged from 12 to 26 degrees, as detailed in Table 4. The drone was equipped with a downwelling light sensor (DLS) to measure incident light

for radiometric calibration; however, these measurements are known to be highly sensitive to sensor tilt during flight [33]. To address this issue, a reflectance reference tarp was imaged before, during, and after each flight for radiometric correction. A calibration tarp with a known and stable reflectance of 26% was used as a ground reference to support consistent reflectance-based post-processing.

Table 4 Data Collection and metadata range of the real MS image

Date	Phenological State	Accumulated GDD	Zenith angle range (o)	Flight Height (m)	Pixel size (cm/px)
4-23-2020	Prebloom	556	23 - 26	80	4.7
7-1-2020	Veraison	2517	14 - 16	85	5.2
4-29-2021	Bloom	720	21 - 22	95	5.6
6-23-2021	Veraison	2332	11 - 12	95	5.7

Multispectral image preprocessing

A custom-developed data processing pipeline by the Digital Agriculture Lab (DAL) at the University of California, Davis, was used to generate a spatial grid for identifying and locating the center of each individual vine. Raw MS images were calibrated, radiometrically corrected, and adjusted for fisheye lens distortion using DAL's processing framework, along with images of a reference panel captured before, during, and after each flight mission. The calibrated images were then imported into Pix4D to produce an orthomosaic image of the entire vineyard. Using the orthomosaic as a spatial template, DAL's customized pipeline was employed to identify the raw image in which each vine appeared closest to the image center[33], [94], thereby minimizing the hotspot effect. Each vine image was then individually clipped from its corresponding calibrated

raw image. These clipped images were subsequently processed through a canopy segmentation algorithm, which applied SAVI [95] index thresholding to remove soil and mixed pixels, resulting in a segmented canopy image for each vine.

HELIOS 3D biophysical modeling framework

The HELIOS framework (version 1.3.28, released February 25, 2025) was used for simulating radiative transfer within 3D plant canopies. HELIOS is an open-source C++ modeling architecture designed to integrate modular plug-ins for various environmental and biophysical processes [93]. The framework supports GPU acceleration. Core functionality is managed through the Context class, which governs geometric primitives, data structures, and interactions among coupled plug-ins.

Two key HELIOS modules were employed in this study: the radiation transport plug-in and the plant architecture plug-in. These were selected based on their suitability for simulating spectral radiative transfer and generating physiologically realistic canopy geometries, respectively.

Radiation Transport Plug-in

The HELIOS radiation plug-in implements a reverse Monte Carlo ray-tracing algorithm optimized for GPU computation. This module allows detailed simulation of radiative energy exchange, including both direct and diffuse solar radiation, as well as thermal emissions, under thermodynamically consistent conditions. User-defined parameters include camera geometry (e.g., position, orientation, image resolution, and horizontal field of view) and waveband configuration (e.g., RGB, multispectral, or user-specified spectral bands). Optical properties such as reflectivity and transmissivity can be assigned at the level of individual primitives, enabling fine-grained control over scene interactions with light [22].

Plant Architecture Plug-in

The HELIOS plant architecture (PA) plug-in was utilized to procedurally generate grapevine canopy models. The plug-in supports more than 19 crop types and simulates growth through recursive branching and organ placement. For grapevine, the initial structure at simulation time zero corresponds to a 3-year-old vine. Canopy growth can be simulated over up to 365-time steps, with daily resolution, allowing for realistic temporal development of shoots and leaves.

Key structural parameters, including phytomer number, leaf distribution, and branching architecture, were defined to match typical vineyard configurations. The resulting canopy geometries were then used as inputs for the radiation simulation module.

Plant Object File Generation and Dimension Reduction

Plant object files were generated using varying simulation time steps, leaf sizes, leaf angles, and leaf decay rates, as summarized in Table 5. As the time step increased, the number of developed leaves also increased, leading to significantly larger file sizes. For instance, rendering a 3×3 grid of tree models with high time step object files exceeding 25 minutes of processing time due to the complexity and volume of geometry. To address this computational limitation, a dimension reduction strategy was implemented. Specifically, for object files exceeding 50 MB in size, a quadratic edge collapse decimation algorithm was applied to reduce geometry while preserving essential structural features. This process ensured that file sizes were reduced to 50% of their original size or maintained below 50 MB, thereby enhancing computational efficiency during simulation and rendering. The generated canopy models exhibited a LAI range between 1 and 8, which aligns with typical values observed in vineyard-grown grapevines. 150 grapevine object files were created to use for simulating a diverse range of grapevine canopy.

Table 5 Parameters Used for Grapevine Canopy Model Generation in HELIOS Simulations

Parameter (HELIOS variable name)	Range
Time step (age)	190, 200, 210, 230, 231, 233, 234, 235, 236, 237, 239, 240, 245, 250, 260
Leaf scale (phytomer_parameters.leaf.pitch)	0.1, 0.12, 0.13, 0.15, 0.155, 0.16, 0.165, 0.17, 0.175, 0.18, 0.19, 0.2, 0.21, 0.22, 0.23, 0.25
Leaf pitches (phytomer_parameters.leaf.prototype_scale)	Uniform distribution of (-80, 70), (-74, -64), (-60, -40), (-25, -15), (-24, -14), (-20, -10) degrees
Decay Rate (vegetative_bud_break_probability_decay_rate)	0.0, 0.5, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95
Minimum bud break probability (vegetative_bud_break_probability_min)	0.08, 0.085, 0.09, 0.095, 0.1, 0.11, 0.12

Spectral PCA Analysis and Synthetic Spectra Generation

A principal component analysis (PCA) of grapevine leaf spectral properties was conducted using data adapted from internal grapevine RTM project, which included measurements across different canopy location (top, middle, and bottom) and phenological stages. The dataset consisted of 127 leaf-level spectra, each with corresponding reflectance and transmittance values. A two-dimensional PCA was performed on the full spectral range, accounting for approximately 95% of the total variation (see Figure 11). A convex hull was constructed around the PCA distribution to

define the boundary of the empirical spectral domain. To ensure adequate coverage of spectral variability, a 10% buffer was applied to expand the modeled domain beyond the convex hull.

Within this extended PCA space, 300 synthetic spectra were generated by evenly distributing sample points throughout the domain. These PCA-derived points were then projected back to the original spectral space to retrieve corresponding reflectance and transmittance profiles. The resulting synthetic spectra were subsequently used to define leaf optical properties for each simulation scene, enabling a representative sampling of realistic leaf-level spectral variability across the canopy.

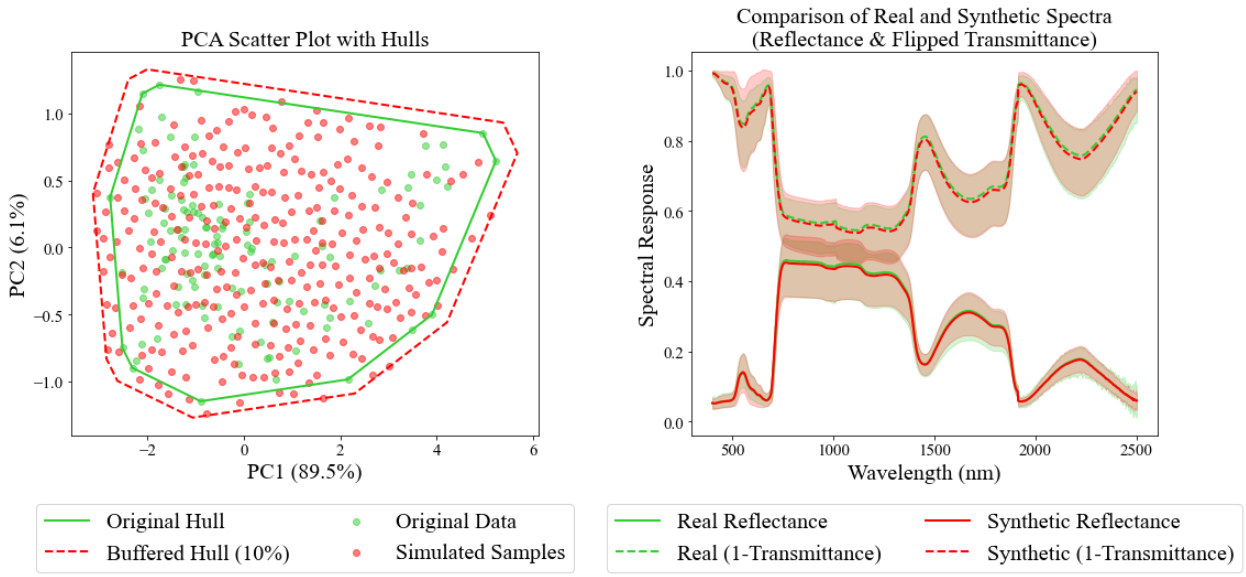


Figure 11: the PCA of the real and synthetic data selected for the simulation

Scene Setup and Processing

Each simulation scene was configured using a 3×3 grid of duplicated tree models to improve computational efficiency. Prior testing confirmed that the reflectance profile of a tree positioned at the center of the grid did not differ significantly when compared to that in more computationally intensive, heterogeneous tree arrangements. This validated the use of homogeneous tree replication for scene construction without compromising spectral accuracy.

Varying solar zenith angles were incorporated to match real-world field conditions. These angles were computed using the HELIOS API based on geospatial metadata, including latitude, longitude, time zone, and the exact timestamp of UAS image acquisition for each tree. Default atmospheric parameters, such as temperature, air pressure, and relative humidity, were applied uniformly across all scenes to standardize radiative transfer conditions.

A 50% reflectance reference panel was placed in each scene, positioned deliberately outside of canopy shadow regions. This reference panel served as a calibration target and was used during post-processing to calculate pixel-level reflectance values across the rendered image.

Data Preparation and Deep Learning Model

Data Preparation

Reflectance data from both real and synthetic sources were preprocessed using min-max scaling before training. Specifically, reflectance histograms for each spectral band were normalized to a 0–1 range to ensure consistent feature scaling across datasets.

For the output variables (y values), robust scaling was employed to reduce the impact of potential outliers. The Leaf Area Index (LAI) values used for training were uniformly distributed from approximately 1 to 8 to realistically represent the variability observed in real-world LAI measurements. Conversely, traits derived from principal component analysis (PCA) of spectral data were normally distributed, aligning with empirical observations of nutrient variability.

Deep Learning Model Architecture

The deep learning model employed was a multi-input fully connected neural network (FNN), structured to handle multiple data sources simultaneously. It featured two parallel input pathways: one dedicated to reflectance histograms and the other designed specifically for solar zenith angle

data. The zenith angle data were represented as the cosine of the zenith angle, a practice similar to [96].

Each pathway contained multiple dense (fully connected) layers paired with rectified linear unit (ReLU) activation functions to introduce non-linearity and enhance feature extraction capabilities. Dropout layers and L2 regularization were strategically integrated into each branch to prevent overfitting.

Following individual processing, the reflectance histogram features, and zenith angle information were merged to jointly inform the final predictions. The model output consisted of multiple simultaneous trait predictions, employing a multi-trait output layer which provide more computational efficiency. This design choice capitalized on correlations among nutrient traits, resulting in a reduction in computational demand and improved predictive performance through better utilization of trait interdependencies.

Results

PCA comparison between real and synthetic data generated from HELIOS

Principal component analysis (PCA) indicated areas of overlap and divergence between the real and synthetic spectral datasets (Figure 12). The spectral PCA scores formed clear phenological groupings in the same area, with bloom and prebloom data clustering closely due to their similarity in accumulated growing degree days (GDD), and veraison spectra similarly forming a distinct cluster reflecting consistent phenological characteristics on the positive PC1 side of the plot.

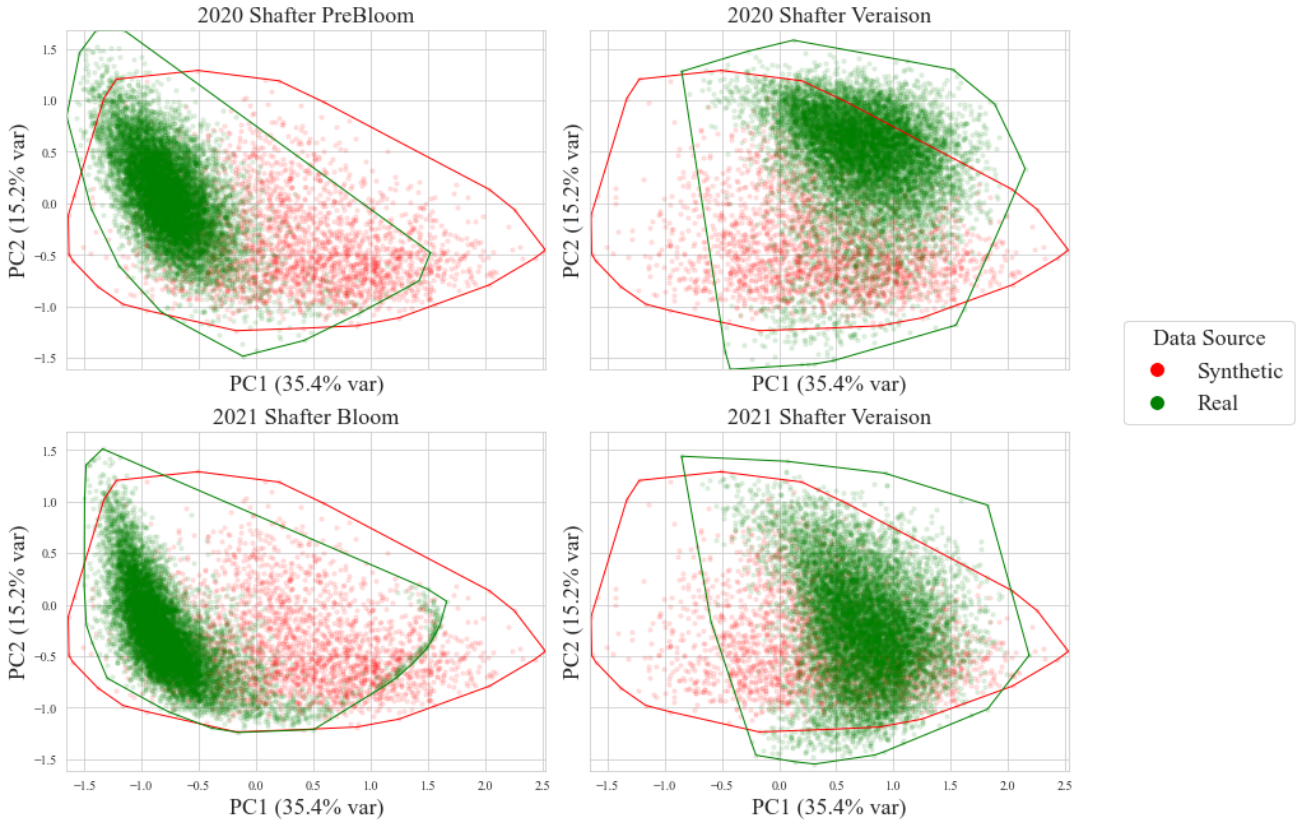


Figure 12: Principal component comparison of synthetic (red) and real (green) canopy Reflectance histograms show PC1 (35.4 % variance) versus PC2 (15.2 % variance) for all 4 imaging campaigns. The separation and overlap between red and green point clouds visualize the spectral domain gap at each phenological stage, with the largest divergence appearing during veraison.

Model Performance on Synthetic Data (5-Fold Cross-Validation)

Figure 13 shows the actual vs predicted plot between of all leaf level traits using 5-fold cross validation in 2000 synthetic images and their ground truth label based on Farajpoor et al. [97] model inversion. Five-fold cross-validation on the synthetic training set produced trait-specific coefficients of determination (R^2) ranging from 0.24 for copper (Cu) to 0.96 for chlorophyll (CHL). These values are consistent with the inversion-derived labels reported in Farajpoor et al. [97], where band-importance peaks sometimes the five DJI P4M wavebands. Leaf area index (LAI) exhibited moderate accuracy ($R^2 \approx 0.57$; RMSE ≈ 1.1), with systematic overestimation below LAI

≈ 3 and underestimation above LAI ≈ 5, indicating regression toward the population mean (≈ 4). A comparable bias was observed for N per unit area (Narea). Conversely, biochemical pigments were retrieved with high accuracy: CHL, ANT, and CAR each achieved $R^2 > 0.90$ and RMSE < 3 % of their respective mean values, demonstrating stable performance across the full concentration range. Nitrogen in structural compounds (N_{struct}) also showed strong agreement with ground truth ($R^2 = 0.88$; RMSE = 0.06), confirming reliable estimation of leaf structural variability despite the sensor's limited spectral resolution.

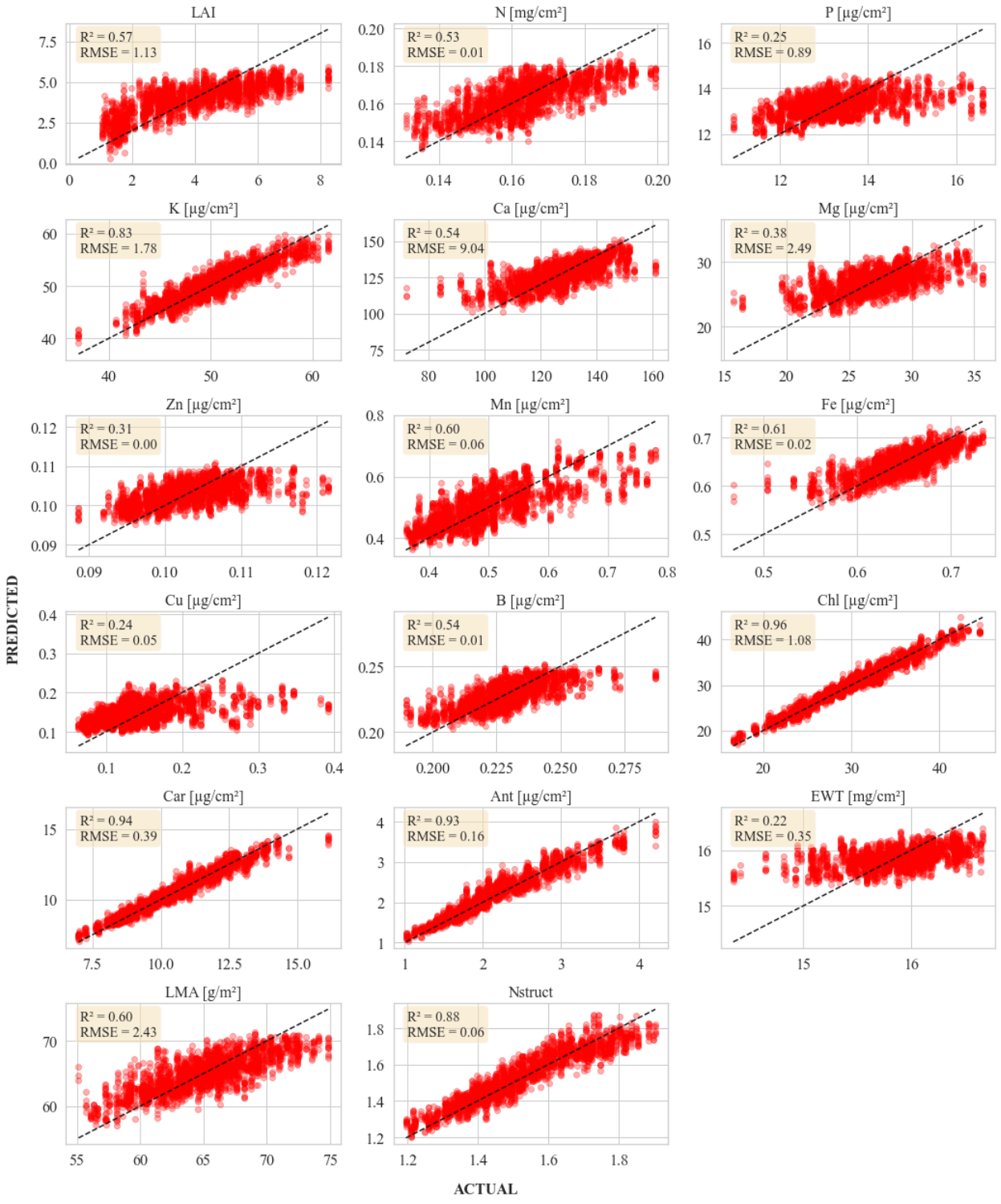


Figure 13: Predicted vs. Actual Trait Values from 5-Fold Cross-Validation on Synthetic Data Stratified by LAI and Solar Zenith Angle.

Boxplot comparing the prediction results in each season

Figure 14 shows the boxplot comparisons between traits used for training the model and the predicted output for real images from 4 imaging campaigns. CHL, LMA, and Narea all increased from prebloom and bloom to veraison, reflecting the physiological maturation of leaves as the leaves thicken and chlorophyll accumulates, giving a darker green appearance. An exception to this general progression is the LAI: Pre-bloom canopies displayed the highest density (averaged ≈ 5), Bloom 2021 averaged ≈ 4, and the two veraison campaigns registered lower values (≈ 3.0 in 2020 and ≈ 3.5 in 2021). veraison samples also exhibited markedly higher LMA than earlier stages; in both 2020 and 2021, predicted LMA values extended beyond the inter-quartile range of the synthetic training data, indicating structural thickening not fully represented in the synthetic domain. Across all traits, predicted distributions for field datasets were more compressed than those of the synthetic priors, yet biochemical variables—CHL, ANT , CAR, and LMA—were still recovered across their expected dynamic ranges. Predicted Narea rose monotonically with accumulated growing degree days, diverging from the decline reported Farajpoor et al. points from prebloom through fruitset, followed by a slight rebound at veraison [97].

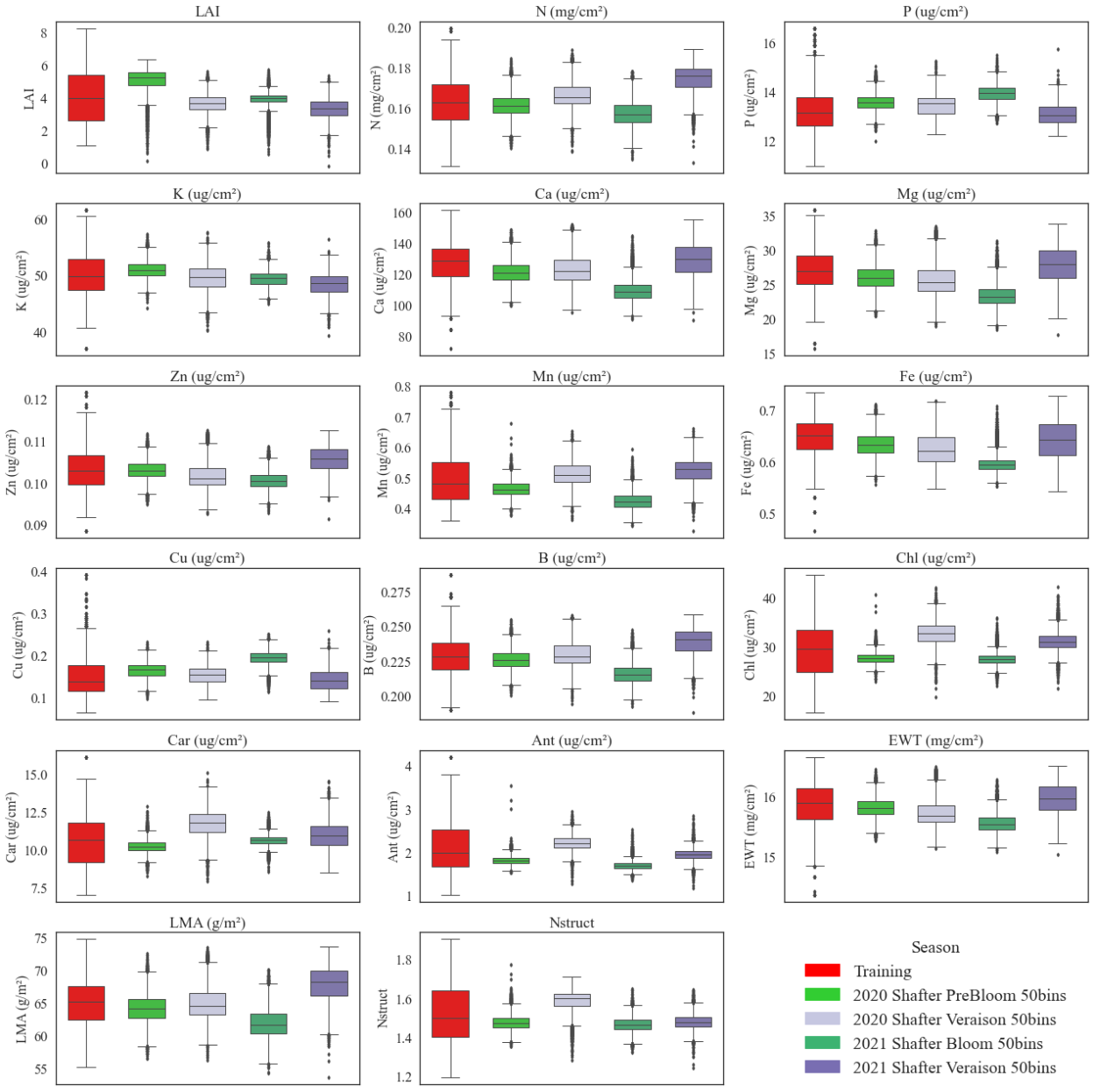


Figure 14: Predicted Trait Distributions on Real-World Grape Canopy Data Across Seasons Using a Deep Learning Model Trained on all Synthetic Data available

Relationship Between Vegetation Indices and Predicted Chlorophyll and Nitrogen

NDVI and NDRE were evaluated as indicators of CHL and N content using the average reflectance values from segmented real canopy images (Figure 16). Vegetation indices are often used as proxies for N because they are highly correlated with chlorophyll, which in turn correlates with N.

Although leaf-level studies have shown limited performance of vegetation indices for predicting N directly, both NDVI and NDRE showed strong correlations with the model's predicted chlorophyll values. Using ordinary least squares regression, NDVI achieved an R^2 of 0.51 and NDRE achieved an R^2 of 0.73 for predicted chlorophyll. These results suggest that the model is able to capture key spectral trends and that chlorophyll, whose signal is clearly expressed in the five-band multispectral sensor, can be predicted with good accuracy using this approach.

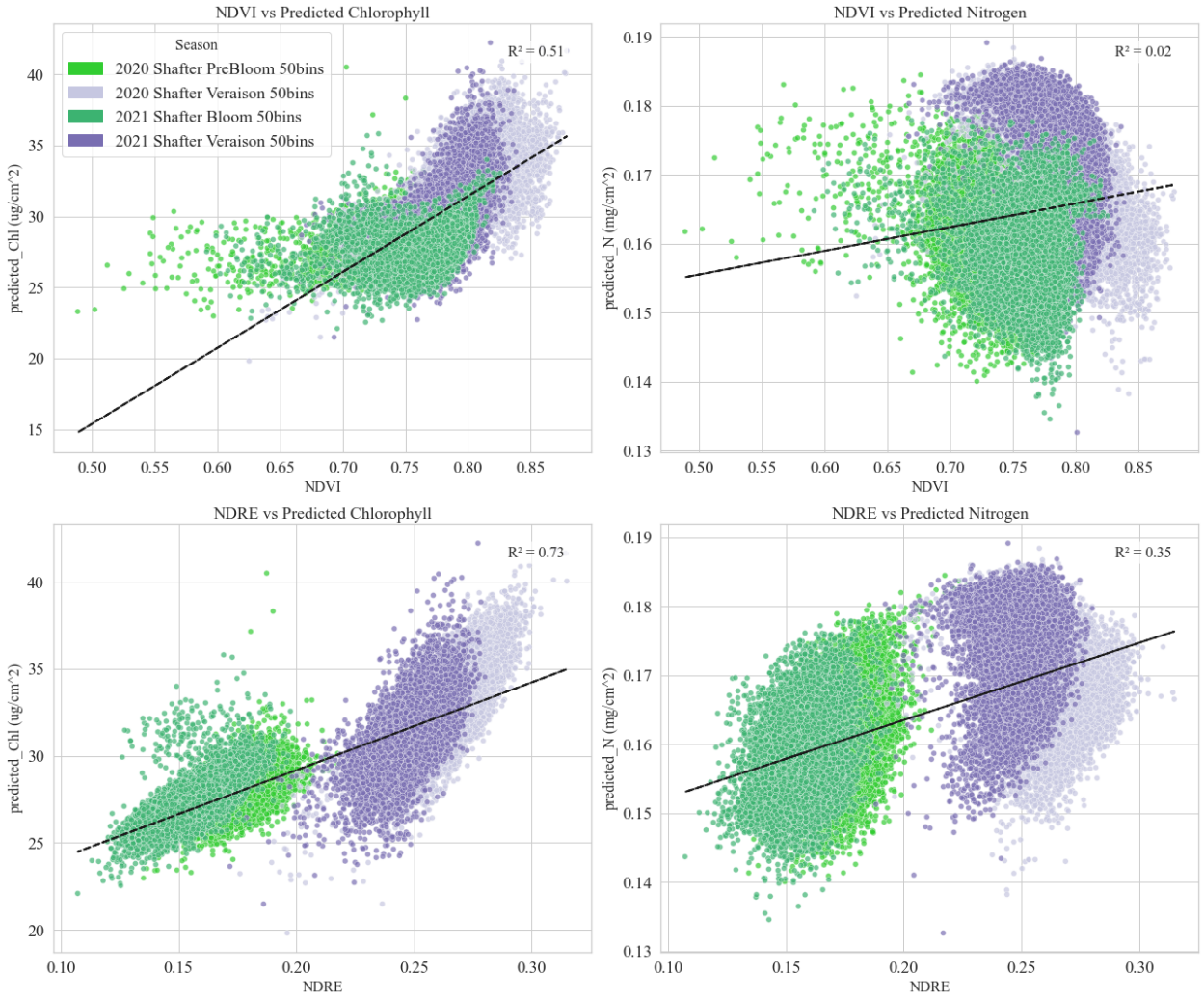


Figure 15: Relationship between vegetation indices (NDVI and NDRE) and model predicted chlorophyll and N across real canopy images. Each subplot shows the ordinary least squares (OLS) regression between NDVI or NDRE and predicted CHL or N values. NDVI and NDRE were computed as the average index values from segmented real canopy images across four field

seasons. R^2 values indicate the strength of correlation between vegetation indices and model predictions.

Spatial Distribution of Predicted Canopy Traits

Figure 16 and 17 show spatial maps of predicted canopy traits across four image acquisition campaigns. Each rectangle corresponds to an individual vine; blank areas represent missing vines or locations where no canopy pixels were detected after segmentation.

Predicted CHL maps showed consistent spatial patterns, with lower CHL values repeatedly observed on the east side of the vineyard across all campaigns. A localized increase in CHL prediction was observed in the upper central region of the Bloom 2021 dataset, likely due to changes in illumination caused by transient cloud cover during image acquisition.

N predictions exhibited greater temporal variability, and possibly an inverted trend of N between bloom and veraison. In bloom 2021, higher N predictions were concentrated in the center-East region of the field, whereas this same area showed increased N predictions in both veraison campaigns.

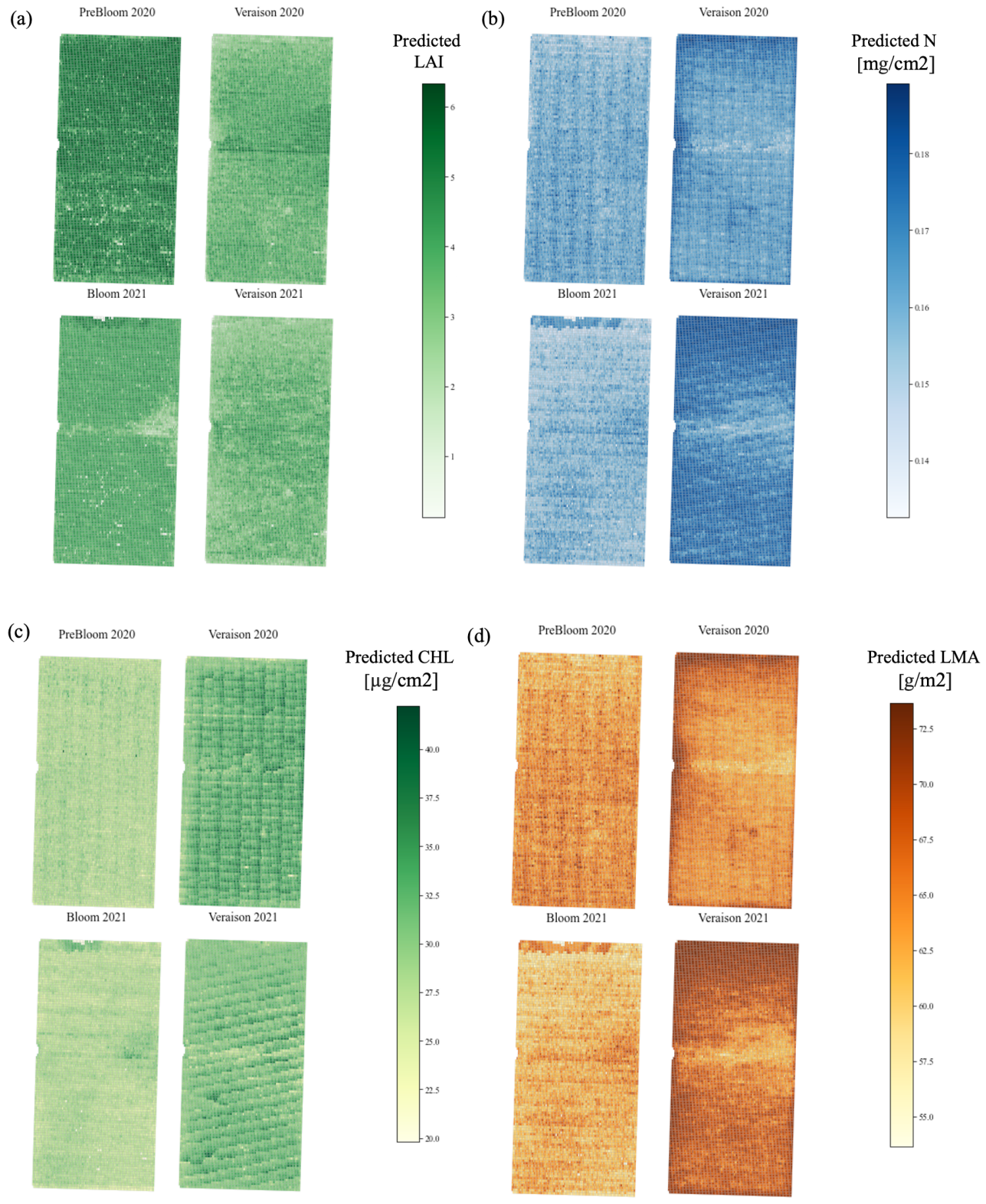


Figure 16: Spatial Distribution of Predicted LAI, Chlorophyll, and Nitrogen Across Four Phenological Stages in a Table Grape Vineyard.

Each subplot represents per-vine predictions of key canopy the model trained on synthetic data. Rows correspond to predicted leaf area index (a), N content per unit leaf area (b), chlorophyll content per unit leaf area (c) and leaf mass area (d), while columns represent different field campaigns. Each rectangle denotes an individual vine; blank areas indicate missing vines.

Predicted LAI was multiplied by leaf-level trait estimates, such as N_{area}, CHL, to approximate total canopy-level nutrient mass per vine, providing a management-relevant indicator of total nutrient availability across different canopy sizes. This design allows management interpretation where, for example, two canopies with similar leaf-level N but differing canopy sizes would show higher total N mass in the larger canopy due to increased leaf area. In Figure 17, part a and b, it is interesting to observe the region in the center right that exhibited high LAI x N in veraison 2020, transitioned to become the region with lowest LAI x N value in bloom 2021.

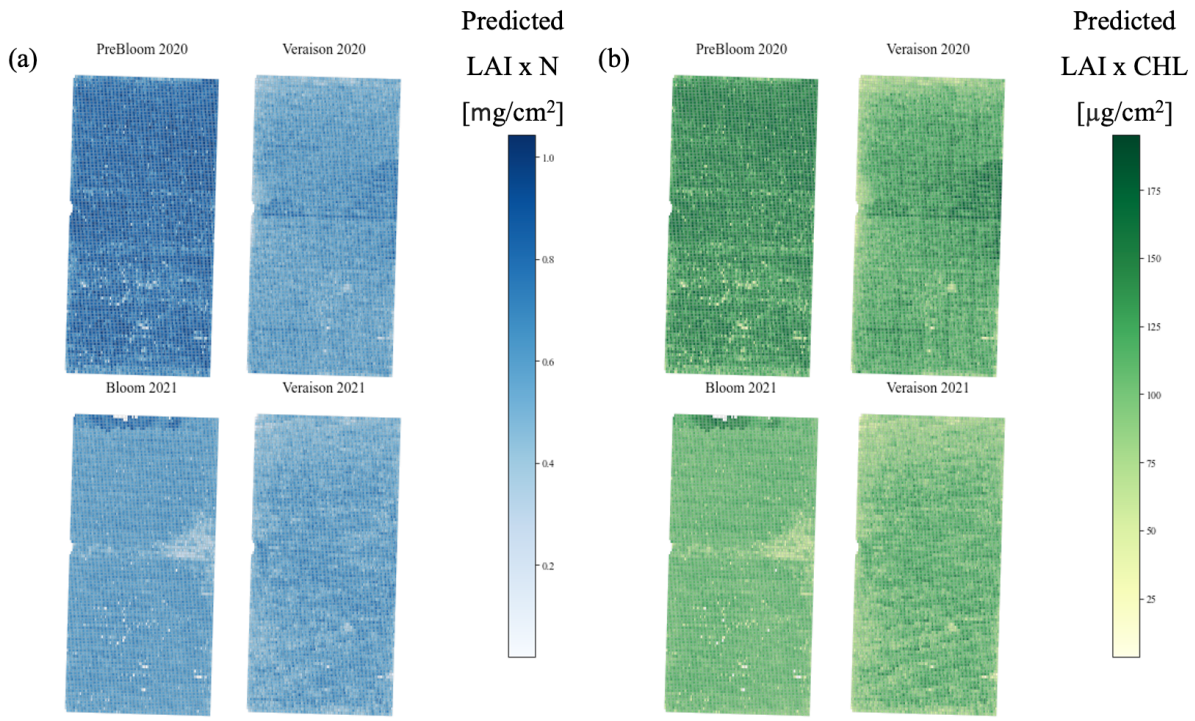


Figure 17: Estimated Canopy-Level Nutrient Mass for Nitrogen and Chlorophyll Across Four Phenological Stages.
(a) Spatial distribution of predicted canopy N_{mass} per unit ground area, calculated as the product of predicted LAI and leaf-level nitrogen content (mg per cm²). (b) Corresponding spatial distribution of canopy-level chlorophyll mass per unit ground area, computed as LAI multiplied by leaf-level chlorophyll concentration (µg per cm²).

Discussion

Trait-Specific Prediction Performance: Biochemical vs. Structural

The model exhibited lower predictive performance for LAI compared to biochemical traits such as CHL, CAR, and ANT. This discrepancy can be attributed in part to the fundamental differences in how these traits are expressed spectrally. LAI, as a structural canopy attribute, is indirectly inferred through spatial reflectance patterns, which may be altered during image preprocessing. Specifically, the canopy segmentation process using the SAVI was designed to maximize the removal of soil and mixed pixels. While effective in isolating canopy pixels, this process may also inadvertently discard important spatial cues necessary for accurately estimating LAI, particularly at the canopy margins. Furthermore, biochemical traits benefit from their inherent interdependencies. In the multi-trait modeling framework adapted from Farajpoor et al. [97], the spectral expression of CHL, CAR, and ANT is jointly constrained by biologically plausible relationships, enabling the model to leverage shared variance and improve trait inference. In contrast, LAI values in the HELIOS simulations were independently varied across scenes through random selection of canopy models. This randomization limits the degree to which LAI co-varies with leaf biochemical composition, thereby reducing the model's ability to learn stable, multivariate associations for LAI prediction. It is also worth noting that LAI values in the training dataset followed a uniform distribution across the range of 1 to 8, whereas most biochemical traits exhibited approximately normal distributions. While this difference in distribution does not directly explain the lower performance for LAI, it provides additional context for how the data may have influenced model learning behavior.

The prediction distributions for real data appeared notably less diverse than those of the synthetic training data, which can be attributed to several contributing factors. First, the synthetic dataset was intentionally designed to maximize diversity in both canopy structure and leaf biochemical

traits. Reflectance and transmittance inputs were sampled across a wide spectral principal component space derived from multiple grapevine varieties, canopy heights, and phenological stages. To further enhance model generalizability, a 10% buffer was added around the empirical spectral domain to expand coverage (see Figure 11). In contrast, the real-world dataset was sourced from a single vineyard and a single grapevine variety, albeit across multiple field seasons. This inherent disparity in input variability likely contributed to the broader range of predictions in synthetic data relative to the more constrained output distributions observed in real-world applications.

In addition to this range mismatch, model underfitting may also have played a role, particularly for traits such as LAI. The model occasionally defaulted to predicting values clustered around the global average, suggesting insufficient learning of the underlying structural-spectral relationships. For LAI, prediction is further complicated by factors such as leaf angle and canopy light scattering, which influence reflectance without altering LAI itself. Since the same LAI value can yield different spectral profiles depending on leaf orientation, the model may have struggled to disentangle these effects without additional structural cues. In this study, LAI prediction was included in an effort to extend trait estimation from the leaf level to the canopy level via composite traits (e.g., LAI × biochemical concentration), but the spectral signal alone may not have been sufficient for robust inference in the presence of these confounding variables.

Phenological Dynamics in Predicted Biochemical and Structural Traits

The differences in model performance across phenological stages reflect underlying physiological changes in grapevine canopy development and demonstrate the model's adaptability to these evolving traits. Key canopy characteristics such as CHL, LMA, and LAI all increased from prebloom and bloom to veraison stages, consistent with the expected phenological progression.

These transitions are also visually apparent in the RGB imagery, where leaf color shifts from lighter to darker green as chlorophyll accumulates. LMA increases as leaves mature and thicken, and LAI expands due to canopy growth throughout the season. The model successfully captured these seasonal dynamics, with predicted trait values reflecting this temporal evolution in both synthetic and real datasets.

The HELIOS model played a central role in enabling the simulation of such trait-relevant variability by offering extensive control over structural parameters. Canopy properties such as leaf size, leaf angle, LAI, and even canopy leaf density profile was varied across simulation scenes, facilitating the creation of highly diverse and biologically plausible canopy configurations. Furthermore, HELIOS supported the generation of multispectral datasets under varying solar zenith angles, allowing the model to learn directional reflectance effects, which are known to influence canopy spectra as shown in Figure A2 and supported by Jafarbiglu et al [33]. This structural and angular variability was essential in allowing the deep learning model to infer phenological differences. Indeed, principal component analysis of reflectance spectra from real data showed clear clustering between bloom/prebloom and veraison stages, and the model's trait predictions followed a similar trend. This alignment indicates that HELIOS simulations successfully preserved meaningful phenological relationships, enhancing the model's ability to generalize across stages of canopy development.

Vegetation Indices as Validation Tools

NDVI and NDRE are commonly used spectral indices in crop monitoring due to their established correlations with chlorophyll content. Although these indices were not very effective for N content prediction in grapevines (See Figure 4 in Chapter 3), their sensitivity to chlorophyll makes them useful proxies for evaluating the plausibility of model-predicted chlorophyll distributions. In this

study, both average NDVI and NDRE were computed from real imagery and regressed against chlorophyll values predicted by the deep learning model trained exclusively on HELIOS-generated synthetic data. The results showed strong linear correlations, with NDRE achieving an R^2 of 0.73 and NDVI an R^2 of 0.51. These findings support the credibility of the predicted chlorophyll maps and, more broadly, validate the ability of HELIOS-simulated data to train models capable of producing biochemically meaningful outputs when applied to real-world imagery. This highlights the value of vegetation indices not as replacements for model-based biochemical prediction but as accessible benchmarks for spectral trend consistency and domain transfer reliability.

Beyond serving as a validation step, these results demonstrate the practical potential of the HELIOS framework for remote sensing applications. While HELIOS has previously been applied to generate synthetic RGB images for object detection tasks such as fruit counting [98], this study is among the first to demonstrate its viability for spectrally intensive use cases, such as nutrient prediction from multispectral reflectance. This expands HELIOS's utility from structural realism into biochemical modeling.

Framework Implications for application of digital twin technology in remote sensing of agriculture

The HELIOS framework offers several advantages that make it particularly well suited for simulating remote sensing data on a scale. First, it provides modular plug-ins for plant architecture, enabling users to parameterize diverse canopy features such as leaf angle, leaf size, shoot orientation, and canopy density. This allows researchers to create structurally realistic digital twins of vineyard or orchard systems, tailored to match specific field configurations. Second, HELIOS supports fine-grained environmental control, including the manipulation of solar zenith and azimuth angles, facilitating batch simulation under varied illumination conditions. These

capabilities are essential for training robust deep learning models that generalize across spatial and temporal variability in remote sensing data.

However, despite its strengths, the current limitations in computational efficiency present challenges for broader adoption. For example, simulating a 3×3 vineyard scene using detailed canopy models (~1 million primitives per vine) required approximately 20 minutes per image on a high-end local workstation (Intel Core i9-7980XE, 128 GB RAM, NVIDIA GTX 1080Ti). This computational cost limits scalability, particularly for users with constrained resources. To overcome this, the current study leveraged distributed computing on an AWS EC2 Fleet, using 32 g4dn.2xlarge instances equipped with NVIDIA T4 GPUs, to accelerate scene generation in distributed computing manner. While this solution enabled the production of a sufficiently large training set, it underscores the need for ongoing optimization of simulation workflows to reduce run time and increase accessibility for researchers across institutions.

In summary, this work establishes HELIOS not only as a flexible and biophysically grounded modeling environment but also as a viable engine for producing synthetic datasets that support trait estimation in real remote sensing imagery. With further improvements in simulation speed and structural realism, HELIOS could serve as a foundational tool for the development of transferable, data-driven models in vineyard management and beyond.

Conclusion

This study demonstrates the feasibility of using synthetic multispectral imagery generated by the HELIOS 3D RTM to train deep learning models for predicting canopy biochemical traits in grapevine. The trained models showed successful domain transfer when applied to real UAS imagery, particularly for traits with strong spectral expression such as CHL, CAR, and ANT. In

contrast, structural traits like LAI were predicted with lower accuracy, likely due to limitations in spectral expressivity and potential information loss during canopy segmentation.

Principal component analysis and Euclidean distance metrics confirmed a non-trivial domain gap between synthetic and real spectral data; however, high correlations between predicted chlorophyll and vegetation indices (see Figure 15) provided independent validation of the model's spectral similarity with real image and generalizability. While trait predictions across phenological stages partially reflected expected physiological trends—such as increasing CHL and LMA from prebloom to veraison—the consistency varied across traits. CHL predictions aligned reasonably well with seasonal canopy development; however, N predictions showed spatial inconsistencies, with some regions exhibiting reversed patterns between bloom and veraison, indicating limited reliability in capturing true temporal dynamics for certain traits.

The HELIOS framework proved effective for simulating realistic canopy structures and environmental variability through its modular plug-ins and support for scene-level parameterization, enabling the generation of structurally and spectrally diverse digital twins. However, the computational intensity of high-resolution scene rendering remains a barrier to scalability. This study addressed this limitation by leveraging distributed simulation across multiple GPU-enabled cloud instances, but further improvements in simulation efficiency are necessary to broaden accessibility.

Overall, this work establishes a biophysically grounded and scalable pipeline for trait estimation in precision agriculture, and positions HELIOS as a viable engine for digital twin development for nutrient retrieval algorithm applications

Limitation and Future works

Homogeneous Neighboring Trees

All scenes in this study duplicated the center vine's architecture and biochemical traits across its eight neighbors to minimize the combinatorial space of simulations and time required to run the simulations. While computationally less expensive, this assumption ignores the vine-to-vine variability present in real vineyards and may encourage the model to overfit to perfectly symmetrical canopies. Introducing variation, either by randomly sampling from the existing OBJ library or by diversifying spectra profile used for each neighboring tree, would produce more realistic spectral backgrounds and reduce the risk of systematic bias. Overall, the effects of spectra from neighboring trees must be further understood for a digital twin framework can be considered robust.

Single-Trait Leaf Spectra per Tree

Each simulated vine was assigned a single, uniform set of leaf biochemical properties, neglecting the within-canopy gradients documented in Chapter 4. A more realistic approach would allocate multiple leaf-trait clusters (e.g., top, middle, bottom spectral profile) within a single canopy, weighted to reflect observed proportions. This stratified representation would yield composite canopy spectra that better match field measurements and could enhance the fidelity of N and LAI predictions.

Persistent Domain Gap in Dense Canopies

Synthetic canopies exhibited a pronounced central reflectance peak at the center of the canopy with rapid lateral fall-off towards to edge of the canopy, whereas real images showed uniformly high reflectance across the canopy. Closing this gap will require denser foliage toward canopy edges or trellis geometries that mimic the Y-shaped open-gable system in the study vineyard. Equally important is first identifying why the current simulation pipeline fails to reproduce the

uniformly high internal reflectance observed in real imagery—whether this shortfall stems from oversimplified leaf orientation, or insufficient multiple-scattering events. A clear mechanistic understanding will ensure that subsequent architectural changes address the true source of the discrepancy rather than merely increasing scene complexity by simply increasing the number of leaves.

Future research should therefore couple architectural refinements with performance-oriented strategies, such as mesh decimation, or algorithm optimization, to preserve realism without rendering the simulation workflow impractical for large-scale simulation.

Chapter 5: Overall Conclusion and Future Work

Overall Conclusion

This dissertation establishes a comprehensive framework for N estimation in grapevines by integrating physically based modeling approaches at both the leaf and canopy scales. The first component of the research focused on evaluating mechanistic RTMs at the leaf level to understand how biochemical and biophysical traits influence spectral reflectance. Through a comparative analysis of N retrieval methods, the study revealed that while purely data-driven approaches often yield high prediction accuracy, their reliance on large training datasets and limited generalizability constrain their practical application. Physically based models, in contrast, offer interpretability and require fewer data but are challenged by narrow predictive capacity and ill-posed inversion problems due to spectral ambiguities. The integration of machine learning with radiative transfer inversion, however, introduced new sources of error propagation, particularly in mass-based N estimates, underscoring the importance of trait-specific calibration and error sensitivity analysis.

Building upon the mechanistic understanding developed at the leaf scale, the second part of the study extended these insights to the canopy level using HELIOS, a 3D ray-tracing RTM. HELIOS enabled the generation of synthetic multispectral imagery across varying canopy architectures, illumination geometries, and biochemical trait distributions. A key milestone achieved in this dissertation was the replacement of data-intensive hyperspectral imaging with synthetic multispectral data for model training, representing an almost opposite philosophy to conventional data-driven methods. This synthetic-data-driven approach provides a practical alternative in scenarios where field data acquisition is costly, sparse, or logistically constrained. Deep learning models trained entirely on synthetic datasets were able to estimate real-world canopy N content with moderate to high accuracy, particularly for traits such as chlorophyll, carotenoids, and structural N. The domain gap between synthetic and real reflectance data remained a limiting factor, especially for traits sensitive to canopy segmentation and LAI variability. Nonetheless, this hybrid modeling approach demonstrated that synthetic data, when derived from mechanistic simulations, can serve as a viable surrogate for empirical data in scenarios where ground truth is sparse or difficult to collect.

The combined outcomes of both studies underscore the value of hybrid frameworks that couple physically based RTM with machine learning to enhance the generalizability, interpretability, and scalability of nutrient retrieval systems in grapevine. By grounding predictions in biophysical principles while leveraging data-driven flexibility, the approach contributes a robust methodology that supports both fundamental understanding and practical deployment for precision agriculture.

Future Research

Advancing this hybrid modeling framework presents multiple opportunities. First, improvements in computational efficiency are necessary to scale 3D radiative transfer simulations for large-scale

or near-real-time applications. Reducing GPU rendering time and optimizing canopy structure parameterization will enable broader adoption of synthetic training data pipelines. Second, further refinement of canopy architecture realism is needed to reduce domain gaps, particularly for trellis-trained grapevines where structure is highly management-dependent. Expanding the diversity of leaf spectral inputs and incorporating intra-canopy variation will enhance transferability across phenological stages, cultivars, and trellis systems.

Reference

[1] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Comput. Electron. Agric.*, vol. 151, no. November 2017, pp. 61–69, 2018, doi: 10.1016/j.compag.2018.05.012.

[2] Y. Fu *et al.*, "An overview of crop nitrogen status assessment using hyperspectral remote sensing: Current status and perspectives," *Eur. J. Agron.*, vol. 124, no. December 2020, p. 126241, 2021, doi: 10.1016/j.eja.2021.126241.

[3] CDFA, "California Agricultural Statistics Review 2018-19," *Calif. Agric. Stat. Rev.*, pp. 1–121, 2019.

[4] T. Harter *et al.*, "Addressing Nitrate in California's Drinking Water with a focus on Tulare Lake Basin and Salinas Valley Groundwater," *Rep. State Water Resour. Control Board Rep. Legis.*, pp. 1–78, 2012.

[5] Z. Jiang, Y. Zhong, J. Yang, Y. Wu, H. Li, and L. Zheng, "Effect of nitrogen fertilizer rates on carbon footprint and ecosystem service of carbon sequestration in rice production," *Sci. Total Environ.*, vol. 670, pp. 210–217, 2019, doi: 10.1016/j.scitotenv.2019.03.188.

[6] H. A. J Benton Mills and J. Jr, *Plant analysis handbook II: a practical sampling, preparation, analysis, and interpretation guide.* Micro-Macro Pub., 1996.

[7] J. Robinson, "Critical plant tissue values and application of nutritional standards for practical use in vineyards.," in *Proceedings of the soil environment and vine mineral nutrition symposium, San Diego, California, USA, 29-30 June, 2004*, D. R. S. L. P. Christensen, Ed., American Society for Enology and Viticulture, 2005, pp. 61–68.

[8] M. Chakraborty, A. Pourreza, S. Peanusaha, P. Farajpoor, S. D. S. Khalsa, and P. H. Brown, "Integrating hyperspectral radiative transfer modeling and Machine learning for enhanced nitrogen sensing in almond leaves," *Comput. Electron. Agric.*, vol. 234, p. 110195, Jul. 2025, doi: 10.1016/j.compag.2025.110195.

[9] A. P. Moghimi Alireza; Zuniga-Ramirez, German; Williams, Larry E. ;. Fidelibus, Matthew W., "A Novel Machine Learning Approach to Estimate Grapevine Leaf Nitrogen Concentration Using Aerial Multispectral Imagery," *Remote Sens.*, vol. 12, no. 21, pp. 3515-NA, 2020, doi: 10.3390/rs12213515.

[10] A. Pourreza, Y. Kamiya, S. Peanusaha, H. Jafarbiglu, A. Moghimi, and M. W. Fidelibus, "Nitrogen retrieval in grapevine (Vitis vinifera L.) canopy by hyperspectral imaging," *Comput. Electron. Agric.*, vol. 229, p. 109717, Feb. 2025, doi: 10.1016/j.compag.2024.109717.

[11] G. Anderson, J. van Aardt, P. Bajorski, and J. V. Heuvel, "Detection of wine grape nutrient levels using visible and near infrared 1nm spectral resolution remote sensing," *Auton. Air Ground Sens. Syst. Agric. Optim. Phenotyping*, vol. 9866, no. May 2016, p. 98660H, 2016, doi: 10.1117/12.2227720.

[12] J. G. P. W. Clevers, "Beyond NDVI: Extraction of Biophysical Variables From Remote Sensing Imagery," in *Remote Sensing and Digital Image Processing*, vol. NA, NA vols., Springer Netherlands, 2014, pp. 363–381. doi: 10.1007/978-94-007-7969-3_22.

[13] J. C.-V. Verrelst Gustau; Muñoz-Marí, Jordi; Rivera, Juan Pablo; Veroustraete, Frank; Clevers, Jan G. P. W. ;. Moreno, Jose, "Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, no. NA, pp. 273–290, 2015, doi: 10.1016/j.isprsjprs.2015.05.005.

[14] F. S. B. Chapin Arnold J. ;. Field, Christopher B. ;. Waring, Richard H., “Plant Responses to Multiple Environmental FactorsPhysiological ecology provides tools for studying how interacting environmental resources control plant growth,” *BioScience*, vol. 37, no. 1, pp. 49–57, 1987, doi: 10.2307/1310177.
[15] R. Giovos, D. Tassopoulos, D. Kalivas, N. Lougkos, and A. Priovolou, “Remote sensing vegetation indices in viticulture: A critical review,” *Agric. Switz.*, vol. 11, no. 5, 2021, doi: 10.3390/agriculture11050457.
[16] L. Rustioni, D. Grossi, L. Brancadoro, and O. Failla, “Iron, magnesium, nitrogen and potassium deficiency symptom discrimination by reflectance spectroscopy in grapevine leaves,” *Sci. Hortic.*, vol. 241, no. July, pp. 152–159, 2018, doi: 10.1016/j.scienta.2018.06.097.
[17] K. Berger *et al.*, “Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions,” *Remote Sens. Environ.*, vol. 242, no. March, p. 111758, 2020, doi: 10.1016/j.rse.2020.111758.
[18] K. Berger *et al.*, “Retrieval of aboveground crop nitrogen content with a hybrid machine learning method,” *Int. J. Appl. Earth Obs. Geoinformation*, vol. 92, no. April, p. 102174, 2020, doi: 10.1016/j.jag.2020.102174.
[19] L. M. Schmidtke, J. P. Smith, M. C. Müller, and B. P. Holzapfel, “Rapid monitoring of grapevine reserves using ATR-FT-IR and chemometrics,” *Anal. Chim. Acta*, vol. 732, pp. 16–25, 2012, doi: 10.1016/j.aca.2011.10.055.
[20] S. C. Nawar Ronald; Halcro, Graham; Mulla, David J. ;. Mouazen, Abdul Mounem, “Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review,” *Adv. Agron.*, vol. 143, no. NA, pp. 175–245, 2017, doi: 10.1016/bs.agron.2017.01.003.
[21] J. B. Féret, K. Berger, F. de Boissieu, and Z. Malenovský, “PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents,” *Remote Sens. Environ.*, vol. 252, no. October 2020, 2021, doi: 10.1016/j.rse.2020.112173.
[22] T. Lei, J. Graefe, I. K. Mayanja, M. Earles, and B. N. Bailey, “Simulation of Automatically Annotated Visible and Multi-/Hyperspectral Images Using the Helios 3D Plant and Radiative Transfer Modeling Framework,” *Plant Phenomics*, vol. 6, Apr. 2024, doi: 10.34133/plantphenomics.0189.
[23] P. Christensen, “Seasonal Changes and Distribution of Nutritional Elements in Thompson Seedless Grapevines,” *Am. J. Enol. Vitic.*, vol. 20, no. 3, pp. 176–190, 1969, doi: 10.5344/ajev.1969.20.3.176.
[24] P. Christensen, “Nutrient Level Comparisons of Leaf Petioles and Blades in Twenty-Six Grape Cultivars Over Three Years (1979 through 1981),” *Am. J. Enol. Vitic.*, vol. 35, no. 3, pp. 124–133, 1984, doi: 10.5344/ajev.1984.35.3.124.
[25] T. Verdenal *et al.*, “Understanding and managing nitrogen nutrition in grapevine: A review,” *Oeno One*, vol. 55, no. 1, pp. 1–43, 2021, doi: 10.20870/OENO-ONE.2021.55.1.3866.
[26] A. M. Metay Jessica; Guilpart, Nicolas; Christophe, Angélique, “Nitrogen supply controls vegetative growth, biomass and nitrogen allocation for grapevine (cv. Shiraz) grown in pots,” *Funct. Plant Biol. FPB*, vol. 42, no. 1, pp. 105–114, 2014, doi: 10.1071/fp14062.
[27] E. B. Soubeyrand Cyril; Hilbert, Ghislaine; van Leeuwen, Cornelis; Delrot, Serge; Gomès, Eric, “Nitrogen supply affects anthocyanin biosynthetic and regulatory genes in grapevine

cv. Cabernet-Sauvignon berries.," *Phytochemistry*, vol. 103, no. NA, pp. 38–49, 2014, doi: 10.1016/j.phytochem.2014.03.024.
[28] E. C. Soubeyrand Sophie; Beauvoit, Bertrand; Dai, Zhanwu; Cluzet, Stéphanie; Hilbert, Ghislaine; Renaud, Christel; Maneta-Peyret, Lilly; Dieuaide-Noubhani, Martine; Mérillon, Jean-Michel; Gibon, Yves; Delrot, Serge; Gomès, Eric, "Constraint-Based Modeling Highlights Cell Energy, Redox Status and α-Ketoglutarate Availability as Metabolic Drivers for Anthocyanin Accumulation in Grape Cells Under Nitrogen Limitation," *Front. Plant Sci.*, vol. 9, no. NA, pp. 421–421, 2018, doi: 10.3389/fpls.2018.00421.
[29] T. Z. Thomidis Eleftheria; Koundouras, Stefanos; Karagiannidis, Constantinos; Navrozidis, Ioannis; Nikolaou, Nikolaos, "Effects of nitrogen and irrigation on the quality of grapes and the susceptibility to Botrytis bunch rot," *Sci. Hortic.*, vol. 212, no. NA, pp. 60–68, 2016, doi: 10.1016/j.scienta.2016.09.036.
[30] J.-L. V. Spring Thibaut; Zufferey, Vivian; Viret, Olivier, "Nitrogen dilution in excessive canopies of Chasselas and Pinot noir cvs," *OENO One*, vol. 46, no. 3, pp. 233–240, 2012, doi: 10.20870/oeno-one.2012.46.3.1520.
[31] G. S. Hilbert J. P. ;. Molot, C. ;. Giraudon, J. ;. Milin, M. ;. Gaudillère, Jean-Pierre, "Effects of nitrogen supply on must quality and anthocyanin accumulation in berries of cv. Merlot," *Vitis J. Grapevine Res.*, vol. 42, no. 2, pp. 69–76, 2015, doi: 10.5073/vitis.2003.42.69-76.
[32] T. S. Rosenstock, D. Liptzin, J. Six, and T. P. Tomich, "Nitrogen fertilizer use in California: Assessing the data, trends and a way forward," *Calif. Agric.*, vol. 67, no. 1, pp. 68–79, Jan. 2013, doi: 10.3733/ca.E.v067n01p68.
[33] H. P. Jafarbiglu Alireza, "Impact of sun-view geometry on canopy spectral reflectance variability," *ISPRS J. Photogramm. Remote Sens.*, vol. 196, no. NA, pp. 270–286, 2023, doi: 10.1016/j.isprsjprs.2022.12.002.
[34] R. B. Chancia Terry; Heuvel, Justine E. Vanden; van Aardt, Jan, "Assessing Grapevine Nutrient Status from Unmanned Aerial System (UAS) Hyperspectral Imagery," *Remote Sens.*, vol. 13, no. 21, pp. 4489-NA, 2021, doi: 10.3390/rs13214489.
[35] M. P. Rast Thomas H., "Earth Observation Imaging Spectroscopy for Terrestrial Systems: An Overview of Its History, Techniques, and Applications of Its Missions," *Surv. Geophys.*, vol. 40, no. 3, pp. 303–331, 2019, doi: 10.1007/s10712-019-09517-z.
[36] F. F. Baret Th., "Radiometric Estimates of Nitrogen Status of Leaves and Canopies," in *Diagnosis of the Nitrogen Status in Crops*, vol. NA, NA vols., Springer Berlin Heidelberg, 1997, pp. 201–227. doi: 10.1007/978-3-642-60684-7_12.
[37] T. Kattenborn, F. Schiefer, P. Zarco-Tejada, and S. Schmidtlein, "Advantages of retrieving pigment content [µg/cm2] versus concentration [%] from canopy reflectance," *Remote Sens. Environ.*, vol. 230, no. September 2018, p. 111195, 2019, doi: 10.1016/j.rse.2019.05.014.
[38] D. Li *et al.*, "Application of unmanned aerial vehicle optical remote sensing in crop nitrogen diagnosis: A systematic literature review," *Comput. Electron. Agric.*, vol. 227, p. 109565, Dec. 2024, doi: 10.1016/j.compag.2024.109565.
[39] L. Fan *et al.*, "Hyperspectral-based estimation of leaf nitrogen content in corn using optimal selection of multiple spectral variables," *Sens. Switz.*, vol. 19, no. 13, 2019, doi: 10.3390/s19132898.
[40] M. M. Khakbazan Alan P. ;. Huang, Jianzhong, "Economic evaluation of variable rate nitrogen management of canola for zones based on historical yield maps and soil test

recommendations.," *Sci. Rep.*, vol. 11, no. 1, pp. 4439-NA, 2021, doi: 10.1038/s41598-021-83917-3.
[41] C. W. Cao Tianli; Gao, Maofang; Li, Yang; Li, Dandan; Zhang, Huijie, "Hyperspectral inversion of nitrogen content in maize leaves based on different dimensionality reduction algorithms," *Comput. Electron. Agric.*, vol. 190, no. NA, pp. 106461-NA, 2021, doi: 10.1016/j.compag.2021.106461.
[42] R. Omidi, A. Moghimi, A. Pourreza, M. E.-H. Aly, and A. S. Eddin, "Ensemble Hyperspectral Band Selection for Detecting Nitrogen Status in Grape Leaves," *ArXiv201004225 CsCV*, 2020, [Online]. Available: http://arxiv.org/abs/2010.04225
[43] L. M. Deng Zhihui; Li, Xiaojuan; Hu, Zhuowei; Duan, Fuzhou; Yan, Yanan, "UAV-based multispectral remote sensing for precision agriculture: a comparison between different cameras.," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, no. NA, pp. 124–136, 2018, doi: 10.1016/j.isprsjprs.2018.09.008.
[44] Y. C. Ma ZhiChao; Fan, YiGuang; Bian, MingBo; Yang, GuiJun; Chen, RiQiang; Feng, HaiKuan, "Estimating potassium in potato plants based on multispectral images acquired from unmanned aerial vehicles.," *Front. Plant Sci.*, vol. 14, no. NA, pp. 1265132-NA, 2023, doi: 10.3389/fpls.2023.1265132.
[45] G. A. Carter and A. K. Knapp, "Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration," *Am. J. Bot.*, vol. 88, no. 4, pp. 677–684, 2001, doi: 10.2307/2657068.
[46] J. L. Hatfield, A. A. Gitelson, J. S. Schepers, and C. L. Walthall, "Application of spectral remote sensing for agronomic decisions," *Agron. J.*, vol. 100, no. 3 SUPPL., pp. 117–131, 2008, doi: 10.2134/agronj2006.0370c.
[47] S. Debnath *et al.*, "Hyperspectral Imaging to detect Age, Defects and Individual Nutrient Deficiency in Grapevine Leaves," *arXiv*, 2020.
[48] A. Hall, D. W. Lamb, B. Holzapfel, and J. Louis, "Optical remote sensing applications in viticulture - A review," *Aust. J. Grape Wine Res.*, vol. 8, no. 1, pp. 36–47, 2002, doi: 10.1111/j.1755-0238.2002.tb00209.x.
[49] T. S. Magney, J. U. H. Eitel, and L. A. Vierling, "Mapping wheat nitrogen uptake from RapidEye vegetation indices," *Precis. Agric.*, vol. 18, no. 4, pp. 429–451, 2017, doi: 10.1007/s11119-016-9463-8.
[50] M. Vizzari, F. Santaga, and P. Benincasa, "Sentinel 2-based nitrogen VRT fertilization in wheat: Comparison between traditional and simple precision practices," *Agronomy*, vol. 9, no. 6, pp. 1–12, 2019, doi: 10.3390/agronomy9060278.
[51] D. K. Taskos Stefanos; Stamatiadis, Stamatis; Zioziou, Eleftheria; Nikolaou, Nikolaos; Karakioulakis, K. ;. Theodorou, Nikolaos, "Using active canopy sensors and chlorophyll meters to estimate grapevine nitrogen status and productivity," *Precis. Agric.*, vol. 16, no. 1, pp. 77–98, 2014, doi: 10.1007/s11119-014-9363-8.
[52] C. F. Jordan, "Derivation of Leaf-Area Index from Quality of Light on the Forest Floor," *Ecology*, vol. 50, no. 4, pp. 663–666, 1969, doi: 10.2307/1936256.
[53] A. M. C. Davies, "What is and what is not chemometrics," *Spectrosc. Eur.*, vol. 24, no. 4, p. 33+36, 2012.
[54] J. P. V. Rivera-Caicedo Jochem; Muñoz-Marí, Jordi; Camps-Valls, Gustau; Moreno, Jose, "Hyperspectral dimensionality reduction for biophysical variable statistical retrieval," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, no. NA, pp. 88–101, 2017, doi: 10.1016/j.isprsjprs.2017.08.012.

[55] X. Ye, S. Abe, and S. Zhang, “Estimation and mapping of nitrogen content in apple trees at leaf and canopy levels using hyperspectral imaging,” *Precis. Agric.*, vol. 21, no. 1, pp. 198–225, Feb. 2020, doi: 10.1007/s11119-019-09661-x.
[56] K. Berger *et al.*, “Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions,” *Remote Sens. Environ.*, vol. 242, p. 111758, Jun. 2020, doi: 10.1016/j.rse.2020.111758.
[57] J. M. Verrelst Jordi; Alonso, Luis; Delegido, Jesús; Rivera, Juan Pablo; Camps-Valls, Gustavo; Moreno, Jose, “Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3,” *Remote Sens. Environ.*, vol. 118, no. NA, pp. 127–139, 2012, doi: 10.1016/j.rse.2011.11.002.
[58] C. E. W. Rasmussen Christopher, *Gaussian Processes for Machine Learning*, vol. NA, no. NA. NA, 2005, p. NA.
[59] F. Michaud, M. Lamas, U. Lugrís, and J. Cuadrado, “A tutorial on support vector regression,” *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, pp. 199–222, 2021, doi: 10.1186/s12984-021-00806-6.
[60] M. R. Segal, “Machine Learning Benchmarks and Random Forest Regression Publication Date Machine Learning Benchmarks and Random Forest Regression,” *Cent. Bioinforma. Mol. Biostat.*, p. 15, 2004.
[61] P. Fu, K. Meacham-Hensold, K. Guan, and C. J. Bernacchi, “Hyperspectral Leaf Reflectance as Proxy for Photosynthetic Capacities: An Ensemble Approach Based on Multiple Machine Learning Algorithms,” *Front. Plant Sci.*, vol. 10, p. 730, Jun. 2019, doi: 10.3389/fpls.2019.00730.
[62] W. A. ; G. Allen H. W. ;. Richardson, Arthur J. ;. Thomas, James R., “Interaction of Isotropic Light with a Compact Plant Leaf,” *J. Opt. Soc. Am.*, vol. 59, no. 10, pp. 1376–1379, 1969, doi: 10.1364/josa.59.001376.
[63] S. B. Jacquemoud Frédéric, “PROSPECT: A model of leaf optical properties spectra,” *Remote Sens. Environ.*, vol. 34, no. 2, pp. 75–91, 1990, doi: 10.1016/0034-4257(90)90100-z.
[64] S. Jean-Baptiste; François Féret C.; Asner Gregory P. ;. Gitelson, Anatoly A. ;. Martin, Roberta E. ;. Bidel, Luc P. R. ;. Ustin, Susan L. ;. Le Maire, Guerric; Jacquemoud, “PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments,” *Remote Sens. Environ.*, vol. 112, no. 6, pp. 3030–3043, 2008, doi: 10.1016/j.rse.2008.02.012.
[65] S. Jean-Baptiste; Gitelson Féret Anatoly A.; Noble Scott D. ;. Jacquemoud, “PROSPECT-D: towards modeling leaf optical properties through a complete lifecycle,” *Remote Sens. Environ.*, vol. 193, no. NA, pp. 204–215, 2017, doi: 10.1016/j.rse.2017.03.004.
[66] R. F. ; A. Kokaly Gregory P. ;. Ollinger, Scott V. ;. Martin, Mary E. ;. Wessman, Carol A., “Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies,” *Remote Sens. Environ.*, vol. 113, no. NA, pp. S78–S91, 2009, doi: 10.1016/j.rse.2008.10.018.
[67] K. Jean-Baptiste; Le Maire Féret G.; Jay Sylvain; Berveiller, Daniel; Bendoula, Ryad; Hmimina, Gabriel; Cheraiet, A. ;. Oliveira, J. C. ;. Ponzoni, Flávio Jorge; Solanki, Twinkle; de Boissieu, F. ;. Chave, Jérôme; Nouvellon, Yann; Porcar-Castell, Albert; Proisy, Christophe; Soudani, “Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine

learning," *Remote Sens. Environ.*, vol. 231, no. 110959, pp. 110959-NA, 2019, doi: 10.1016/j.rse.2018.11.002.
[68] Z. S. Wang Andrew K. ;. Wang, Tiejun; Darvishzadeh, Roshanak; Hearne, John W., "Applicability of the PROSPECT model for estimating protein and cellulose + lignin in fresh leaves," *Remote Sens. Environ.*, vol. 168, no. 168, pp. 205–218, 2015, doi: 10.1016/j.rse.2015.07.007.
[69] J. Cheng *et al.*, "Estimating canopy-scale chlorophyll content in apple orchards using a 3D radiative transfer model and UAV multispectral imagery," *Comput. Electron. Agric.*, vol. 202, p. 107401, Nov. 2022, doi: 10.1016/j.compag.2022.107401.
[70] J. X. Qi Donghui; Guo, Dashuai; Yan, Guangjian, "A Large-Scale Emulation System for Realistic Three-Dimensional (3-D) Forest Simulation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 11, pp. 4834–4843, 2017, doi: 10.1109/jstars.2017.2714423.
[71] D. Oswald, A. Pourreza, M. Chakraborty, S. D. S. Khalsa, and P. H. Brown, "3D radiative transfer modeling of almond canopy for nitrogen estimation by hyperspectral imaging," *Precis. Agric.*, vol. 26, no. 1, p. 12, Feb. 2025, doi: 10.1007/s11119-024-10207-z.
[72] T. Cheng, B. Rivard, A. G. Sánchez-Azofeifa, J. B. Féret, S. Jacquemoud, and S. L. Ustin, "Deriving leaf mass per area (LMA) from foliar reflectance across a variety of plant species using continuous wavelet analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 28–38, 2014, doi: 10.1016/j.isprsjprs.2013.10.009.
[73] J. B. Féret *et al.*, "Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine learning," *Remote Sens. Environ.*, vol. 231, no. October 2018, p. 110959, 2019, doi: 10.1016/j.rse.2018.11.002.
[74] D. Li *et al.*, "Estimation of area- and mass-based leaf nitrogen contents of wheat and rice crops from water-removed spectra using continuous wavelet analysis," *Plant Methods*, vol. 14, no. 1, pp. 1–20, 2018, doi: 10.1186/s13007-018-0344-1.
[75] Y. C. Tian, X. Yao, J. Yang, W. X. Cao, D. B. Hannaway, and Y. Zhu, "Assessing newly developed and published vegetation indices for estimating rice leaf nitrogen concentration with ground- and space-based hyperspectral reflectance," *Field Crops Res.*, vol. 120, no. 2, pp. 299–310, Jan. 2011, doi: 10.1016/j.fcr.2010.11.002.
[76] D. F. Rodriguez Glenn J.; Belford, R.; Christensen, L. K., "Detection of nitrogen deficiency in wheat from spectral reflectance indices and basic crop eco-physiological concepts," *Aust. J. Agric. Res.*, vol. 57, no. 7, pp. 781–789, 2006, doi: 10.1071/ar05361.
[77] J. C. Dash Paul J., "The MERIS terrestrial chlorophyll index," *Int. J. Remote Sens.*, vol. 25, no. 23, pp. 5403–5413, 2004, doi: 10.1080/0143116042000274015.
[78] B. Datt, "Remote Sensing of Chlorophyll a, Chlorophyll b, Chlorophyll a+b, and Total Carotenoid Content in Eucalyptus Leaves," *Remote Sens. Environ.*, vol. 66, no. 2, pp. 111–121, 1998, doi: 10.1016/s0034-4257(98)00046-7.
[79] L. P. Serrano Josep; Ustin, Susan L., "Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals," *Remote Sens. Environ.*, vol. 81, no. 2, pp. 355–364, 2002, doi: 10.1016/s0034-4257(02)00011-1.
[80] A. Gitelson and M. N. Merzlyak, "Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves," *J. Photochem. Photobiol. B*, vol. 22, no. 3, pp. 247–252, 1994, doi: 10.1016/1011-1344(93)06963-4.

[81] D. Haboudane, J. R. Miller, N. Tremblay, P. J. Zarco-tejada, and L. Dextraze, “Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture,” vol. 81, pp. 416–426, 2002.
[82] J. A. ; P. Gamon Josep; Field, Christopher B., “A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency,” *Remote Sens. Environ.*, vol. 41, no. 1, pp. 35–44, 1992, doi: 10.1016/0034-4257(92)90059-s.
[83] B. Datt, “A new reflectance index for remote sensing of chlorophyll content in higher plants: Tests using Eucalyptus leaves,” *J. Plant Physiol.*, vol. 154, no. 1, pp. 30–36, 1999, doi: 10.1016/S0176-1617(99)80314-9.
[84] J. E. Vogelmann, B. N. Rock, and D. M. Moss, “Red edge spectral measurements from sugar maple leaves,” *Int. J. Remote Sens.*, vol. 14, no. 8, pp. 1563–1575, 1993, doi: 10.1080/01431169308953986.
[85] A. A. Gitelson, M. N. Merzlyak, and O. B. Chivkunova, “Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves¶,” *Photochem. Photobiol.*, vol. 74, no. 1, p. 38, 2001, doi: 10.1562/0031-8655(2001)074<0038:opaneo>2.0.co;2.
[86] R. Main, M. A. Cho, R. Mathieu, M. M. O’Kennedy, A. Ramoelo, and S. Koch, “An investigation into robust spectral indices for leaf chlorophyll estimation,” *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 6, pp. 751–761, 2011, doi: 10.1016/j.isprsjprs.2011.08.001.
[87] K. V. Berger Jochem; Féret, Jean-Baptiste; Hank, Tobias; Wocher, Matthias; Mauser, Wolfram; Camps-Valls, Gustau, “Retrieval of aboveground crop nitrogen content with a hybrid machine learning method,” *Int. J. Appl. Earth Obs. Geoinformation ITC J.*, vol. 92, no. NA, pp. 102174-NA, 2020, doi: 10.1016/j.jag.2020.102174.
[88] J. V. Verrelst Jorge; Rivera-Caicedo, Juan Pablo; Lumbierres, Maria; Morcillo-Pallarés, Pablo; Moreno, Jose, “Global Sensitivity Analysis of Leaf-Canopy-Atmosphere RTMs: Implications for Biophysical Variables Retrieval from Top-of-Atmosphere Radiance Data,” *Remote Sens.*, vol. 11, no. 16, pp. 1923-NA, 2019, doi: 10.3390/rs11161923.
[89] L. Spafford, G. le Maire, A. MacDougall, F. de Boissieu, and J. B. Féret, “Spectral subdomains and prior estimation of leaf structure improves PROSPECT inversion on reflectance or transmittance alone,” *Remote Sens. Environ.*, vol. 252, no. March 2020, 2021, doi: 10.1016/j.rse.2020.112176.
[90] S. ; I. Poni C. ;. Silvestroni, O., “Interactions of Leaf Age, Fruiting, and Exogenous Cytokinins in Sangiovese Grapevines Under Non-Irrigated Conditions. II. Chlorophyll and Nitrogen Content,” *Am. J. Enol. Vitic.*, vol. 45, no. 3, pp. 278–284, 1994, doi: 10.5344/ajev.1994.45.3.278.
[91] I. F. Rademacher and C. J. Nelson, “Nitrogen effects on leaf anatomy within the intercalary meristems of tall fescue leaf blades,” *Ann. Bot.*, vol. 88, no. 5, pp. 893–903, 2001, doi: 10.1006/anbo.2001.1527.
[92] L. Jinwen *et al.*, “Responses of rice leaf thickness, SPAD readings and chlorophyll a/b ratios to different nitrogen supply rates in paddy field,” *Field Crops Res.*, vol. 114, no. 3, pp. 426–432, 2009, doi: 10.1016/j.fcr.2009.09.009.
[93] B. N. Bailey, “Helios: A Scalable 3D Plant and Environmental Biophysical Modeling Framework,” *Front. Plant Sci.*, vol. 10, p. 1185, Oct. 2019, doi: 10.3389/fpls.2019.01185.
[94] M. Chakraborty, A. Pourreza, X. Zhang, H. Jafarbiglu, K. A. Shackel, and T. DeJong, “Early almond yield forecasting by bloom mapping using aerial imagery and deep

learning," *Comput. Electron. Agric.*, vol. 212, p. 108063, Sep. 2023, doi: 10.1016/j.compag.2023.108063.
[95] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sens. Environ.*, vol. 25, no. 3, pp. 295–309, Aug. 1988, doi: 10.1016/0034-4257(88)90106-X.
[96] M. Weiss, F. Baret, and S. Jay, "S2ToolBox Level 2 products: LAI, FAPAR, FCOVER," 2016.
[97] P. Farajpoor, A. Pourreza, M. Narimani, A. El-kereamy, and M. Fidelibus, "Multi-Trait Spectral Modeling for Estimating Grapevine Leaf Traits and Nutrients," *Plant Phenomics*, vol. *Under Review*.
[98] T. Choi *et al.*, "DAVIS-Ag: A Synthetic Plant Dataset for Prototyping Domain-Inspired Active Vision in Agricultural Robots," in *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)*, Bari, Italy: IEEE, Aug. 2024, pp. 2524–2530. doi: 10.1109/CASE59546.2024.10711368.

Appendix

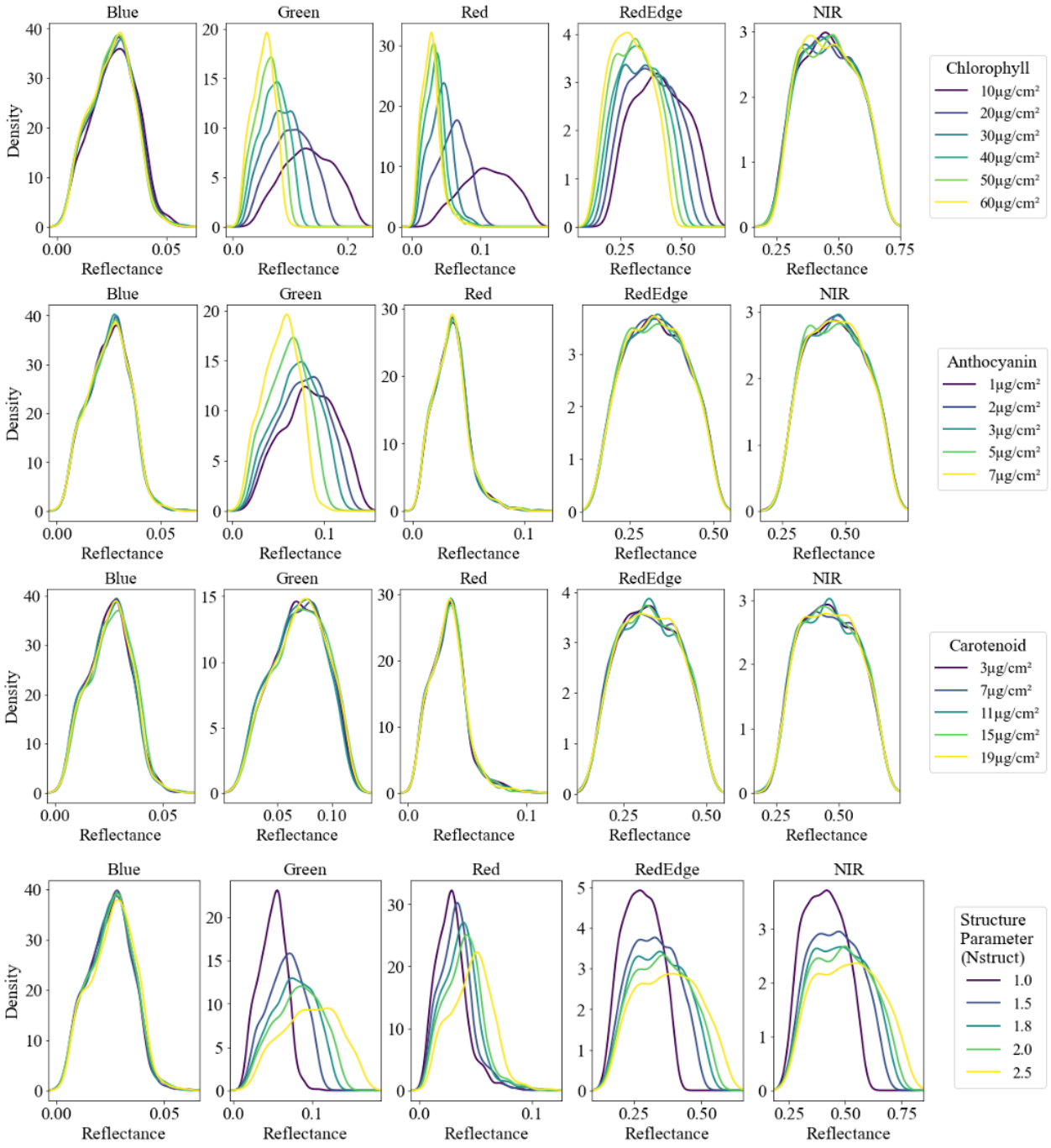


Figure A1: Effects of varying biochemical parameters on reflectance distributions across Phantom 4 multispectral bands. Each subplot displays KDE of reflectance in individual bands as a function of varying biochemical traits, illustrating spectral sensitivity to changes in parameters.

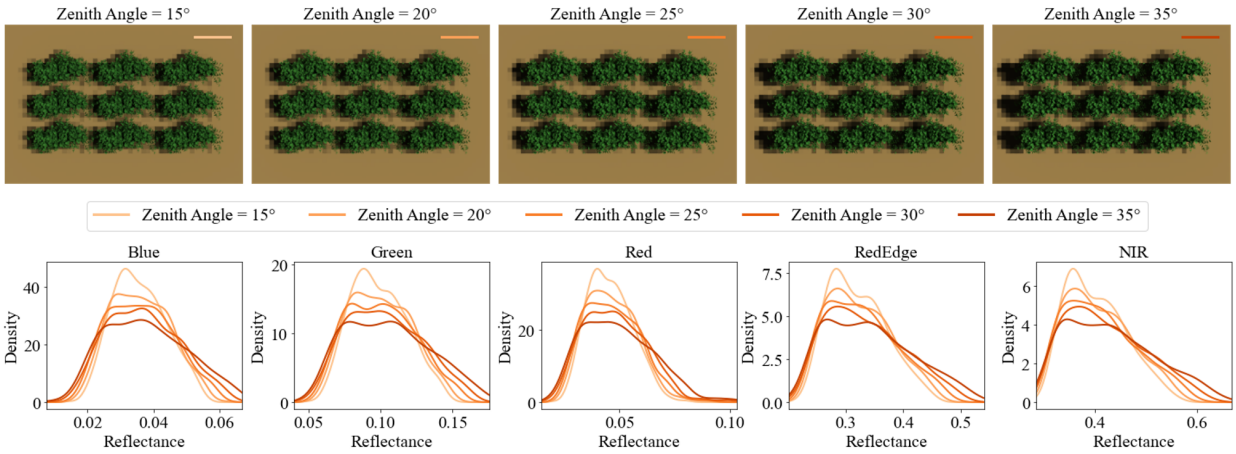


Figure A2: Effects of varying sun zenith angle on the same canopy on 5 multispectral bands captured by DJI P4M camera in HELIOS 3D ray tracing model.