Enhancing grassland nitrogen estimation: A multiscale approach through optical reflectance spectroscopy and hybrid modeling techniques

A thesis presented in partial fulfilment of the requirements for the degree of
Doctor of Philosophy
in
Earth Science
at Massey University, Palmerston North,
New Zealand.

Mohammad Hossain Dehghan-Shoar

2025

Contents

List of Tables

List of Figures

Abstract

Optical remote sensing technology has emerged as a powerful tool for assessing vegetation characteristics, particularly nitrogen (N) concentration (N%) in heterogeneous grasslands. Accurate estimation of N% is crucial for farmers, as it directly influences grassland productivity and plays a key role in sustainable land management. Accurate N assessments optimize fertilizer use, boosting productivity, lowering costs, and enhancing environmental modeling to address impacts such as N leaching and greenhouse gas emissions. Despite significant progress, challenges and knowledge gaps remain, highlighting the need for continued research to fully harness remote sensing's potential in agricultural management and its impact on livestock productivity.

This thesis aims to advance the accurate estimation of grassland N% by integrating physically-based, empirical-statistical, and hybrid models using optical reflectance spectroscopy data. The research focuses on three primary objectives:

1. To estimate N% in grasslands using optical reflectance spectroscopy, data will be collected across multiple scales, including ground-, leaf-, canopy-, and satellite-scale observations.
2. To improve the universality and adaptability of grassland N% models through a hybrid approach that combines data from various optical sensors across multiple scales.
3. To account for and quantify uncertainties in grassland N% prediction models.

The thesis addresses the challenge of uncertainty by conducting a comprehensive analysis of its sources and developing methods, such as Physically Informed Neural Networks (PINN), to account for them. Key strategies include data fusion techniques for integrating diverse data sources and improving atmospheric correction methods. A unified methodology combining empirical-statistical and physically-based approaches is proposed to enhance generalization. Machine learning algorithms play a pivotal role in feature selection and optimization, further improving model accuracy and transferability. The developed methods undergo evaluation using independent validation data collected from heterogeneous grasslands across different periods and locations. Results demonstrate that integrating physically-based and empirical-statistical approaches significantly improves model accuracy and transferability, providing a deeper understanding of the factors influencing vegetation traits. This thesis highlights the importance of advanced techniques, including machine learning, deep learning algorithms, Radiative Transfer Models (RTM), and data fusion methods, for precisely characterizing vegetation traits, contributing to more sustainable and efficient grassland management practices.

Acknowledgements

I am profoundly grateful for the guidance and support I have received throughout my research journey. At the forefront of my gratitude are my esteemed supervisors: Prof. Gabor Kerezturi, Prof. Ian Yule, Dr. Reddy Pullangari, Dr. Alvaro Orsi, and Dr. James Hanly. Their dedication, expertise, and unwavering encouragement have been the cornerstone of my progress. Their insightful feedback and willingness to share their vast knowledge have greatly contributed to my personal and professional growth, fostering in me a belief in my potential that has been a driving force behind my success. I am also deeply indebted to Dr. Lorenzo De Simone, who first introduced me to remote sensing during my Master's degree and provided invaluable support. His guidance laid the foundation for my journey in this field, for which I will always be grateful.

My family has been the pillar of my strength throughout this journey. I owe an eternal debt of gratitude to my parents, Majid Dehghan-Shoar and Zahra Behnood, for their unwavering support and encouragement. Their emotional and financial backing have been indispensable. My siblings, Zeinab and Mahdi Dehghan-Shoar, have also been a constant source of belief and motivation, providing steadfast encouragement during challenging times.

In remembrance of my late mother, Tahereh Goodarzi, I hold her memory close to my heart. Her enduring love and support continue to inspire me every day. I am profoundly grateful to my in-laws, Abbas Khodamoradi and Afsaneh Ghodousi, for their kindness, wisdom, and unwavering encouragement, which have greatly enriched my academic pursuits.

Special thanks go to my beloved nieces, Rose and Naaz Khodamoradi, whose unique perspectives and steadfast support have helped me overcome countless challenges. Their presence has been a source of joy and strength throughout this journey.

I am equally indebted to the Nasirzadeh family for their unwavering support, belief in me, and the warmth with which they have embraced me into their lives. Their encouragement and guidance have provided immense comfort and strength.

Finally, I express my deepest gratitude to my wonderful wife, Zahra Sadat Nasirzadeh Oskooei. Her love, patience, and unwavering support have been the foundation of my journey. Zahra's presence in my life is a constant reminder of the strength that comes from love and partnership. For her, I am forever grateful.

Chapter 1

Introduction

The escalating challenge of global hunger underscores a pressing international concern for the coming decades. Forecasts by the Food and Agriculture Organization of the United Nations (FAO) predict that the world's population will surge to 9.9 billion by 2050, significantly increasing the demand for nutrient-rich foods, particularly meat and dairy products (Brook et al., 2021, Bergeret, 2017, Ramankutty et al., 2008). Given the limited availability of arable land, maximizing the efficiency and productivity of existing agricultural areas is essential. This will likely require doubling or even tripling agricultural production from currently cultivated land without expanding the production area.

Grasslands, the primary feed source for livestock, are essential for producing meat and dairy products. Enhancing grassland quality by accurately estimating nitrogen (N) concentration (N%) is key to optimizing production. Effective N management enables farmers to optimize fertilizer use, improve grassland productivity, and mitigate environmental impacts such as eutrophication, soil acidification, and greenhouse gas emissions (GHG) (Wang et al., 2019a, White et al., 2000, Briggs et al., 2005, Xu et al., 2014, Tiscornia et al., 2019, Yan et al., 2019). However, overexploitation of grasslands can lead to ecosystem degradation, raising significant environmental concerns (Briggs et al., 2005, Alexandratos and Bruinsma, 2012, Julian et al., 2017).

Improved estimates of grassland N content significantly enhance the accuracy of models predicting nitrous oxide emissions (N_2O), such as Overseer, which currently relies on long-term average data (McVeagh, 2019, Wheeler et al., 2011a, 2008). More precise, farm-specific estimates are valuable for addressing annual variations in emissions and tracking GHG at farm, regional, and national levels, since livestock N intake directly influences nitrous oxide emissions from urine patches, better grassland N estimates will improve model accuracy (Di and Cameron, 2002, Ledgard et al., 2009).

Understanding grassland N content related to crude protein content over time offers insights into livestock performance (Reid et al., 2015). Monitoring these changes throughout the season helps explain variations in productivity and informs decisions regarding supplementary feed, ultimately optimizing livestock health and productivity (Ju et al., 2009, Good and Beatty, 2011).

Integrating advanced methods for estimating grassland N concentrations, particularly through optical remote sensing combined with empirical-statistical and hybrid models, significantly enhances the accuracy of mechanistic models like Overseer (Wheeler et al., 2011a, Williams et al., 2013, Wheeler et al., 2011b). These methods not only improve nutrient estimations but also provide a more comprehensive view of nutrient management, enabling more reliable simulations of nutrient dynamics and environmental impacts under varying climate, topography, and soil conditions, thereby addressing the spatial and temporal variability in grasslands where traditional monitoring methods may fall short (Rendel et al., 2020).

Remote sensing technologies play a crucial role in the development of advanced farm management strategies by providing precise, high-resolution data at regular intervals, enabling farmers to make effective decisions regarding fertilizer applications, thereby reducing over-application and minimizing environmental impacts (Liaghat et al., 2010). Remote sensing facilitates more informed decision-making by accurately identifying critical areas within grasslands that require targeted interventions, such as regions with higher nutrient leaching risks or areas prioritized for biodiversity conservation. This precision enhances nutrient use efficiency and supports a more sustainable N cycle, ultimately contributing to improved long-term agricultural productivity and the broader objective of sustainable land management (Sishodia et al., 2020).

Both conventional and cutting-edge methods and technologies are employed to address challenges in estimating N levels in grasslands. Optical remote sensing has emerged as a non-destructive, cost-effective alternative to traditional monitoring methods (Darvishzadeh et al., 2011, Pullanagari et al., 2011, 2021). By measuring light reflectance across various segments of the electromagnetic spectrum, optical remote sensing can estimate biochemical and biophysical characteristics of grassland, such as chlorophyll and protein content. Combined with empirical-statistical, physically-based, and hybrid models, these methods provide accurate predictions of N levels. Advanced algorithms, such as One-Dimensional neural networks, further enhance accuracy by leveraging non-linear relationships between spectra and environmental data, despite variations in climate and soil types (Pullanagari et al., 2021, Dehghan-Shoar et al., 2023a,b). These advancements contribute to a balanced N cycle, improved biodiversity conservation, and sustainable ecological resource management, leading to a more resilient and productive agricultural landscape.

1.1 Research objectives and structure

While remote sensing technology offers numerous advantages for vegetation characterization, it faces significant challenges, primarily related to the absence of robust data fusion techniques grounded in physical principles and models resilient to sensor variations (Pullanagari et al., 2012, Dehghan-Shoar et al., 2023b,c, Hossain et al., 2024a). These models are indispensable for ensuring consistent and accurate monitoring of grassland dynamics, especially considering the inherent variability in N% measurements within these ecosystems.

The application of optical remote sensing technology to estimate grassland N% presents several enduring challenges and knowledge gaps. One of the foremost challenges involves the uncertainty and errors arising from various sources, including environmental factors, data acquisition, and processing variables, inaccuracies in atmospheric correction, and errors in ground truth data obtained through laboratory and field methods. This uncertainty extends to the modeling of N content using empirical-statistical, physically-based, or hybrid approaches, partly due to the inherently complex and ill-posed nature of Radiative Transfer Models (RTMs) (Féret et al., 2021, Mutanga, 2004a, Pullanagari et al., 2021, Yang et al., 2020a).

As fundamental tools in remote sensing, RTMs simulate the interaction of light with the Earth's surface and atmosphere. However, they are often considered ill-posed because small variations in input variables can lead to large discrepancies in the output, making it difficult to retrieve accurate information about surface properties like N% (Quan et al., 2015). This complexity arises from the fact that multiple combinations of variables can produce similar spectral signatures. Therefore, it is imperative to identify and address the multitude of sources contributing to uncertainty when modeling grassland characteristics.

Another critical challenge lies in the transferability of models developed for estimating grassland N content using empirical-statistical and physically-based methods. While considerable progress has been made in this domain, many of these models remain site-specific and are tailored to function with particular sensors. This underscores the need to harness advances in optical remote sensing technology, including creating new physically-based models characterized by high accuracy and robustness across diverse regions and temporal scales while also being compatible with multiple optical sensing instruments.

In pursuit of these solutions, this research endeavors to achieve three primary objectives:

1. To estimate N% in grasslands using optical reflectance spectroscopy, data must be collected across multiple scales, including dry/ground-, leaf-, canopy, and satellite-scale observations.

2. To improve grassland N% models' universality and adaptability through a hybrid approach that combines data from various optical sensors across multiple scales.

3. To account for and quantify uncertainties in grassland N% prediction models.

This research focuses on developing reliable and precise methods for estimating N% across grasslands using optical sensing data. A key challenge arises from the spatial variability of N levels, heavily influenced by concentrated urine patches and areas with low background inorganic N. To address this, our approach improves the accuracy of N% estimation, enabling more effective monitoring and prediction of nitrogen responsiveness under diverse conditions. These advancements contribute to targeted and sustainable grassland management practices. The findings of this study have been consolidated into four journal articles and one conference paper,

each detailing specific innovations in these techniques. Notably, all four journal articles have been accepted by peer-reviewed journals, and the conference paper has also been successfully accepted, as outlined below.

- Dehghan-Shoar, M. H., Orsi, A. A., Pullanagari, R. R., Yule, I. J. (2023). A hybrid model to predict nitrogen concentration in heterogeneous grassland using field spectroscopy. Remote Sensing of Environment, 285, 113385.

- Dehghan-Shoar, M. H., Kereszturi, G., Pullanagari, R. R., Orsi, A. A., Yule, I. J., Hanly, J. (2024). A physically informed multi-scale deep neural network for estimating foliar nitrogen concentration in vegetation. International Journal of Applied Earth Observation and Geoinformation, 130, 103917.

- Hossain, D. S. M., Gabor, K., Pullanagari, R. R., Ian, J. Y., Alvaro, O., James, H. (2024, April). Multi-scale estimation of vegetation nitrogen concentration using a physically informed neural network. In 2024 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS) (pp. 1-3). IEEE.

- Dehghan-Shoar, M. H., Pullanagari, R. R., Orsi, A. A., Yule, I. J. (2023). Simulating spaceborne imaging to retrieve grassland nitrogen concentration. Remote Sensing Applications: Society and Environment, 29, 100912.

- Dehghan-Shoar, M. H., Pullanagari, R. R., Kereszturi, G., Orsi, A. A., Yule, I. J., Hanly, J. (2023). A Unified Physically Based Method for Monitoring Grassland Nitrogen Concentration with Landsat 7, Landsat 8, and Sentinel-2 Satellite Data. Remote Sensing, 15(10), 2491.

This Ph.D. thesis is organized into nine chapters.

- **Chapter 1 Introduction** - This chapter examines the current challenges in monitoring grassland characteristics and addresses the need for alternative methods in assessing grassland health and dynamics. It also outlines the main objectives of this thesis, providing a brief description of each subsequent chapter and the expected outcomes and contributions of the research.

- **Chapter 2: Literature Review** - This chapter reviews relevant literature on optical sensing for N estimation in grasslands, providing the theoretical background and context for the research.

- **Chapter 3: A Hybrid Model to Predict N% in Heterogeneous Grassland Using Field Spectroscopy** - This chapter presents a hybrid model that combines empirical and physically-based approaches to estimate grassland N%. It utilizes a large spectral dataset collected over eight years from grassland farms across New Zealand. The model's performance was tested on spatially and temporally independent data, demonstrating superior accuracy compared to other models. **This chapter has been published in the literature** (Dehghan-Shoar et al., 2023a).

- **Chapter 4: A Physically Informed Multi-Scale Deep Neural Network for Estimating Foliar N% in Vegetation** - This chapter introduces a Physically Informed Deep Neural Network that integrates spectral and RTM features to estimate N% across different scales. The model demonstrated enhanced robustness and improved prediction accuracy over existing models, providing new insights into the relationship between spectral data and N%. **This chapter has been published in the literature** (Dehghan-Shoar et al., 2024a).

- **Chapter 5: Enhancing Grassland N Monitoring through Simulated Spaceborne Sensor Data and Hybrid Modeling** - The fifth chapter discusses the use of spaceborne optical imaging for continuous monitoring of grassland N%. It presents a method that combines field spectroscopy with the Soil-Plant-Atmosphere Radiative Transfer model to simulate Top Of Atmosphere data from any spaceborne sensor. The model was independently validated using real Sentinel-2 TOA reflectance, highlighting its potential for monitoring grassland N%. **This chapter has been published in the literature** (Dehghan-Shoar et al., 2023c).

- **Chapter 6: A Unified Physically Based Method for Monitoring Grassland N% with Landsat 7, Landsat 8, and Sentinel-2 Satellite Data** - The sixth chapter presents a unified and physically-based method for estimating N% from Landsat 7, 8, and Sentinel-2 satellite data. It leverages a soil-plant-atmosphere radiative transfer model and a Gaussian process regression to develop a hybrid model for N estimation. This method overcomes the limitations of diverse sensor configurations and complex atmospheric interactions, providing promising results. **This chapter has been published in the literature** (Dehghan-Shoar et al., 2023b).

- **Chapter 7: General Discussion** - This chapter synthesizes the key findings and insights from all previous chapters. The discussion explores the broader implications of the findings and how they contribute to the current state of knowledge in the field.

- **Chapter 8: Conclusions and Future Directions** - This chapter summarizes the main contributions of this research, outlines the conclusions drawn from these findings, and identifies potential areas for further investigation and development in the field of optical sensing for grassland management.

- **Chapter 9: Appendix** - This chapter provides supplementary results and additional details that support the findings presented in the main chapters of this study. The appendix includes expanded data, figures, and methodological clarifications that were referenced throughout the thesis.

Throughout these chapters, this research aims to contribute to the advancement of robust and accurate methods for estimating grassland characteristics, particularly N%, using optical sensing data. These advancements could improve the sustainability and efficiency of grassland management practices while reducing traditional N assessment's negative environmental impacts.

Chapter 2

Literature Review

Here, we will explore the theory and applications of optical sensing for monitoring vegetation characteristics.

2.1 Electromagnetic radiation

The electromagnetic spectrum encompasses an infinite range of frequencies, wavelengths, and photon energies, extending from extremely short wavelengths like gamma rays, to extremely long wavelengths, like radio waves. Within this vast spectrum, the optical domain spans from 400 to 2500 nm and includes both reflective and emissive components, depending on the observed object. This domain is further divided into three spectral sub-domains: the visible region (VIS) from 400 to 700 nm, the near-infrared region (NIR) from 700 to 1400 nm, and the shortwave infrared region (SWIR) from 1400 to 2500 nm. While the entire electromagnetic spectrum covers a much broader range, agricultural applications primarily focus on the 400-2500 nm region, which is critical for monitoring vegetation health, soil properties, and water content (Mulla, 2013, Khanal et al., 2020, Hank et al., 2019).

2.2 Electromagnetic Radiation and Vegetation Monitoring

Electromagnetic radiation, a form of energy that travels through a vacuum or other media, consists of photons, each carrying a specific amount of energy. Photons interact with matter through processes such as absorption, reflection, scattering, and transmission. When absorbed, photons can excite electrons in molecules, raising them to higher energy states. These excited electrons may transfer energy to neighboring molecules, emit light upon returning to lower energy states (fluorescence), or dissipate energy as heat.

Photon absorption is critical in photosynthesis, where chlorophyll molecules within reaction centers capture and utilize light energy. Absorbed photon energy excites electrons, initiating a series of redox reactions that generate Adenosine Triphosphate (ATP) and Nicotinamide Adenine Dinucleotide Phosphate (NADPH) (Smith et al., 2024). These energy carriers drive the Calvin cycle, enabling plants to synthesize glucose and other essential compounds (Raines,

2003).

Understanding the interaction between electromagnetic radiation and vegetation is fundamental to using optical sensing for vegetation monitoring. Optical sensing leverages the unique spectral responses of vegetation, which result from their physical and chemical properties, including pigments, water content, and cellular structure. Accurate estimates of vegetation characteristics, such as chlorophyll concentration, protein content, and leaf area, are achieved by analyzing spectral variations in reflected or transmitted radiation (Féret et al., 2021). This analysis underpins applications in agriculture, forestry, and environmental monitoring.

The energy carried by a photon is directly proportional to its frequency (f), as described by Planck's equation (Equation 2.1):

$$E = h \cdot f \tag{2.1}$$

Here, E is the photon's energy, f its frequency, and h Planck's constant ($6.626 \times 10^{-34}\,\mathrm{J \cdot s}$) (Prakash, 2000, Landgrebe, 2003). Frequency (f) is inversely proportional to wavelength (λ) according to Equation 2.2:

$$f = \frac{c}{\lambda} \tag{2.2}$$

In this equation, c represents the speed of light ($3.00 \times 10^{8}\,\mathrm{m/s}$), and λ is the wavelength. The inverse relationship between frequency and wavelength implies that shorter wavelengths carry higher energy, while longer wavelengths carry lower energy. This principle forms the foundation of optical sensing technologies that rely on detecting specific spectral bands to monitor vegetation attributes (Pisharoty, 1983, Elachi and Van Zyl, 2021, Platnick, 2000).

When electromagnetic radiation interacts with vegetation, absorbed energy excites electrons in molecules such as chlorophyll and proteins (Meroni et al., 2009). This excitation alters the spectral characteristics of reflected or transmitted radiation. Optical sensors analyze these spectral variations to estimate vegetation attributes accurately (Campbell, 1972). For example, specific absorption bands are tuned to detect chlorophyll content, providing critical insights into plant health and growth dynamics. By leveraging the principles of electromagnetic radiation (Equations 2.1 and 2.2), optical sensing instruments can distinguish vegetation properties based on their spectral traits.

The Sun, as a nearly perfect blackbody radiator, emits radiation across a range of wavelengths that interact with vegetation (Abdelbary, 2024). Optical sensing instruments exploit these interactions by capturing variations in vegetation reflectance and absorption, which are essential for monitoring environmental and agricultural processes.

In contrast, the Earth emits radiation that can be modeled as blackbody radiation to a first approximation, but its actual emission spectrum is influenced by atmospheric and surface

conditions, which cause deviations from the ideal blackbody curve. The amount and nature of radiation an object emits are influenced by its temperature, which can be determined using Wien's displacement law. Wien's displacement law, a fundamental principle in the study of thermal radiation, states that the wavelength of maximum emission is inversely proportional to the object's temperature. Therefore, as an object's temperature increases, the wavelength at which it emits most radiation decreases. The equation, $\lambda_{\max} T = \text{constant}$, represents Wien's displacement law, where $\lambda_{\max}$ is the wavelength of maximum emission, and T is the absolute temperature of the object. This equation applies to any object emitting thermal radiation, including stars, planets, and everyday objects such as stoves and light bulbs. By measuring an object's maximum emission wavelength, its temperature can be determined using Wien's displacement law. This principle finds applications in various areas, including the remote sensing of the surface of the Earth temperature using thermal sensors.

Equation 2.3 links the wavelength (λ), expressed in nanometers (nm), with the temperature (T) in Kelvin (K). This relationship, known as Wien's displacement law, allows us to estimate the wavelength at which a surface emits the maximum radiant flux per unit area:

$$\lambda_{\max} = \frac{2898}{T} \tag{2.3}$$

Here, T is the temperature in Kelvin, and $\lambda_{\max}$ is the wavelength of maximum emission in nanometers. For example, the Sun, with an effective surface temperature of approximately 5772 K, emits its peak radiant flux at a wavelength of about 502 nm. The Earth, with a mean surface temperature of approximately 288 K, emits its peak radiant flux at around 10,070 nm, corresponding to the thermal infrared spectrum (Dorigo et al., 2007).

Understanding emission spectra is essential for applications in remote sensing. While the Sun's radiation dominates the visible spectrum, Earth's thermal emissions provide critical information for analyzing surface temperatures, energy balance, and environmental processes. These insights form the basis of many remote sensing techniques, particularly in environmental and agricultural monitoring.

In contrast, the Earth's peak radiant flux occurs at 9660 nm, corresponding to a cooler temperature of about 300 K. According to the findings of Slater (1980), solar irradiation is comparable to self-emitted thermal energy in the 2500 to 6000 nm range. Therefore, the total amount of radiation a surface receives, or irradiance, can be expressed as the flux density (ψ) per unit area (A) using Equation 2.4.

$$E = \frac{d\psi}{dA} \tag{2.4}$$

Radiance is a fundamental concept in optical remote sensing, crucial for quantifying the amount of electromagnetic radiation reflected or emitted from a surface in a specific direction. Radiance is typically represented in optical sensing as the flux density per unit projected area

and unit solid angle, as indicated in Equation 2.5.

$$L = \frac{E}{\pi} \tag{2.5}$$

Radiance is measured in units of watts per square meter per steradian (W m^{-2} sr^{-1}). This unit reflects the fact that radiance accounts for the directionality of radiation, considering how much radiation is emitted or reflected in a particular direction from a given surface area. In contrast, irradiance, measured in watts per square meter (W m^{-2}), represents the total power of electromagnetic radiation incident on a surface from all directions, without considering the angle of incidence. Understanding the difference between radiance and irradiance is essential in remote sensing because radiance helps determine the angular distribution of reflected or emitted radiation, which is critical for accurate surface property measurements.

For a comprehensive understanding of radiance, it is essential to differentiate between two types of irradiance: direct and diffuse. Direct irradiance (E_{Sun}) refers to sunlight reaching a surface directly without scattering by the atmosphere or surrounding objects. This is typically measured using a pyrheliometer, which focuses sunlight onto a detector through a narrow field-of-view (usually 5 degrees). In contrast, diffuse irradiance (E_{Sky}) refers to scattered radiation that reaches a surface from all directions after interacting with atmospheric molecules, aerosols, and other objects. Diffuse irradiance is typically measured using a pyranometer, which has a cosine collector to measure irradiance over a hemisphere. Pyranometers are calibrated in watts per square meter (W m^{-2}) and generally have a spectral range between 300 and 3000 nm (Elminir, 2007, Bernhard et al., 1997).

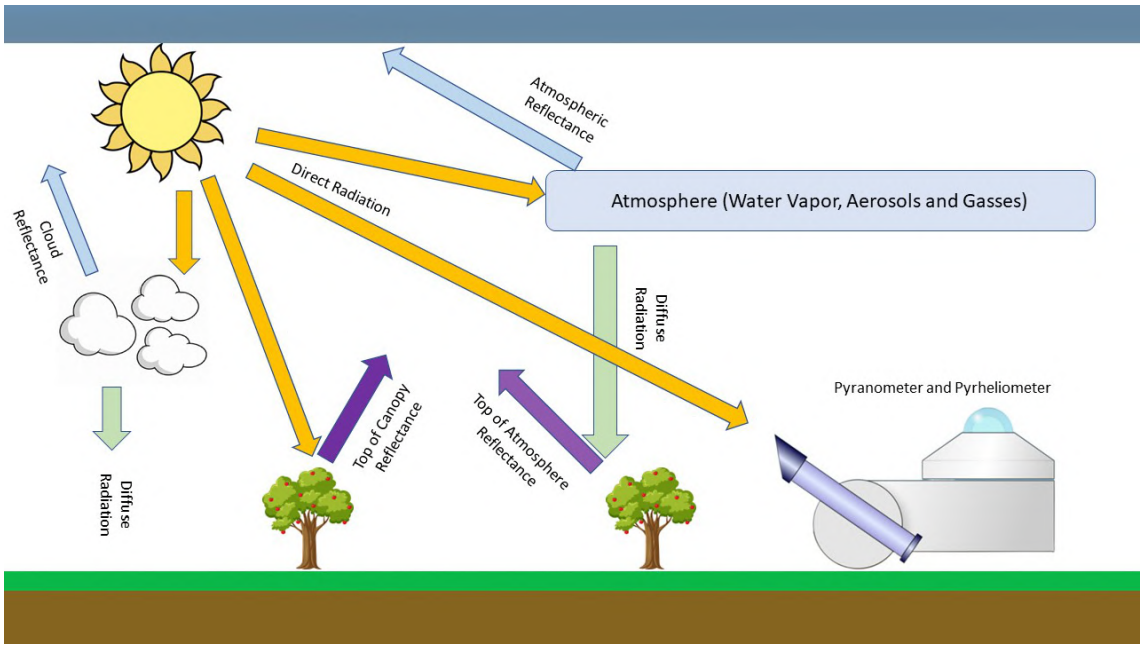


Figure 2.1: Visualizing the impact of Earth's atmosphere, clouds, and vegetation on incoming solar radiation - a simplified illustration.

Top of Canopy (TOC) reflectance represents a surface's spectral properties, such as its biochemical composition and biophysical characteristics (Verhoef, 1998, 1984). It is defined as the ratio of upwelling radiance (L), measured at the TOC, to downwelling irradiance (E), measured at the bottom of the atmosphere, as shown in Equation 2.6:

$$p_\lambda = \pi \cdot \frac{L}{E} \tag{2.6}$$

Here, p_λ is the reflectance at a given wavelength λ, L is the upwelling radiance, and E is the downwelling irradiance. The reflectance is wavelength-specific, meaning it varies depending on the spectral properties of the surface being observed. Reflectance measured at TOC reveals a surface's optical characteristics under specific illumination conditions.

Before optical sensors can analyze surface radiance, Digital Numbers (DN) recorded by the sensor must be converted into radiance values through radiometric calibration. A DN is a numerical value assigned to each pixel in a remotely sensed image, representing the energy detected by the sensor (Wyatt, 2012). These values are converted into radiance using gain (G) and offset (O) parameters, as shown in Equation 2.7:

$$\text{Radiance} = G \cdot \text{DN} - O \tag{2.7}$$

Gain represents the ratio of radiance to the DN value, while the offset accounts for baseline radiance when the DN is zero. Radiometric calibration ensures that DN values are proportional to the incoming radiance, enabling accurate and reliable remote sensing data for scientific and practical applications (Benhadj et al., 2024).

2.3 Atmospheric Correction

The reflectance of a surface or object measured above the atmosphere is called Top of Atmosphere (TOA) reflectance (Mousivand et al., 2015). The impact of atmospheric gases, such as ozone, carbon dioxide, and oxygen, on spectral reflectance is not considered in TOA reflectance when observing an object or surface. Therefore, TOA reflectance can be considered a lower-quality product than TOC reflectance. Atmospheric correction is a critical step in optical remote sensing when monitoring the spectral characteristics of objects or surfaces. This is because atmospheric gases can significantly influence the incoming or reflected radiance and scattering. Accurate atmospheric correction is necessary to ensure that the observed spectral signals are representative of the object or surface being monitored and not distorted by atmospheric effects. Various atmospheric correction algorithms have been developed to correct the impact of atmospheric gases on the spectral reflectance of objects or surfaces (Richter and Schläpfer, 2002, Schläpfer and Richter, 2002, Berk et al., 2006).

Atmospheric gases in Earth's atmosphere can absorb and scatter electromagnetic radiation, altering the surface's spectral reflectance or the observed object. The influence of these gases on spectral reflectance can differ depending on the electromagnetic radiation's wavelength (Ulaby

et al., 1981, Aggarwal, 2004). For instance, oxygen contributes significantly to absorption at 760 nm, while water vapor absorption occurs within the 900 to 2500 nm range. Atmospheric gases can significantly impact the precision of optical sensing data. Thus, atmospheric correction is an essential procedure for eliminating the effects of atmospheric interference on reflectance values (Berk et al., 2006).

This procedure employs mathematical algorithms to compensate for atmospheric effects and estimate the surface reflectance that would have been observed without atmospheric interference. Various methods have been developed for atmospheric correction, including empirical, Radiative Transfer Model (RTM), and hybrid approaches (Kondratyev et al., 2013, Gao et al., 2009, Rahman and Dedieu, 1994). Empirical methods use ground measurements of atmospheric and surface properties to derive atmospheric correction coefficients. In contrast, model-based methods apply atmospheric RTM to simulate atmospheric effects. Hybrid approaches merge elements of both empirical and model-based methods, thereby improving accuracy and efficiency. The choice of atmospheric correction method depends on several factors, such as the instrument used, atmospheric conditions, and the application of the optical sensing data (Gao et al., 2006, 2009). Various methods are used for atmospheric correction in optical sensing, and each is described below:

1. Scene-Based Empirical approaches: These are extensively used for atmospheric correction of optical remote sensing data. Several such approaches have been developed, including Automated Image Registration and Atmospheric Correction (AIRAC) (Conel et al., 1987), and Internal Average Reflectance (IAR) (Kruse, 1988). The IAR approach involves calculating a scene's average spectrum, which is then used to estimate the relative reflectance spectrum for any pixel within the scene. This is achieved by dividing each pixel's spectra by the average spectrum. Scene-based empirical approaches are beneficial when atmospheric conditions across a scene vary widely, and traditional RTM approaches may be inaccurate.

2. Radiative Transfer Model (RTM) methods: This is another approach to atmospheric correction. These approaches allow the simulation and inversion of atmospheric properties such as ozone and oxygen (Gao et al., 1993). Several atmospheric correction algorithms are available for the retrieval of atmospheric properties, including Atmosphere CORrection Now (ACRON), the FastLine-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH), the High-accuracy Atmospheric Correction for Hyperspectral Data (HATCH), and a series of Atmospheric and Topographic Correction (ATCOR) (Qu et al., 2003, Richter and Schläpfer, 2019). These algorithms rely on the principles of RTM and use mathematical and physical models to correct optical data. These approaches are helpful when atmospheric conditions are uniform across the scene and can deliver highly accurate correction results.

3. Hybrid approaches combine RTM and empirical approaches for atmospheric correction. These approaches enhance traditional atmospheric correction by eliminating residual errors often associated with RTM (Goetz et al., 1998, Boardman, 1998). Radiative Transfer

Models (RTM) are used to correct large-scale atmospheric effects in these approaches, while the empirical approach corrects the remaining small-scale effects. This approach has proven to provide highly accurate atmospheric correction results across various applications.

2.4 Different Types of Atmospheric Scattering

Electromagnetic spectrum regions with reasonable transmittance levels are known as "atmospheric windows" and are frequently used for monitoring surface properties through remote sensing techniques. These regions' wavelengths range from 0.3-15 micrometers and encompass VIS, NIR, SWIR, and thermal infrared (TIR) regions (Salisbury and D'Aria, 1992). In addition to atmospheric absorption caused by various gases like carbon dioxide, ozone, and water vapor, solar radiance flux incident on the surface of the Earth and reaching the optical sensor is also impacted by atmospheric scattering (McCartney, 1977). Scattering is the process of light redirection in various directions due to light's interaction with particles or molecules in the atmosphere. Three primary types of spectral scattering exist: Rayleigh, Mie, and non-selective scattering:

1. Rayleigh scattering is a type of spectral scattering that occurs when the size of the intercepting particles is much smaller than the wavelength of the radiation. This scattering is more pronounced at shorter wavelengths, typically less than 500 nanometers, corresponding to the blue region of the visible spectrum (Liang, 2003). The reason shorter wavelengths are more affected lies in their higher energy, as described by Equation 2.1, and their interaction with matter. The degree of scattering is inversely proportional to the fourth power of the wavelength ($1/\lambda^4$), meaning shorter wavelengths scatter much more strongly than longer wavelengths (Seinfeld and Pandis, 2016). Rayleigh scattering results from light interacting with small gas molecules, such as oxygen, in Earth's atmosphere. These molecules, being much smaller than the wavelength of visible light, scatter blue and violet light more strongly than other colors. This phenomenon explains why the sky appears blue during the day; sunlight undergoes significant Rayleigh scattering as it transits through the atmosphere (Young, 1982). However, this scattering effect can influence the accuracy of remote sensing measurements, particularly in the blue and ultraviolet regions of the spectrum.

2. Mie scattering is another spectral scattering type that happens when the intercepting particles' size is comparable to the incoming radiation's wavelength (Liang, 2003). This size-dependent scattering type is more prominent in shorter wavelengths, especially in the visible spectrum's blue region. Unlike Rayleigh scattering, Mie scattering is a wavelength-independent process that occurs when light interacts with particles of a size similar to the light's wavelength. This scattering type commonly results from atmospheric aerosols, such as dust, pollen, smoke, and suspended water particles (Nussenzveig and Wiscombe, 1980). These particles can scatter light in all directions, reducing the light beam's intensity and causing a whitish haze. Mie scattering presence can also impact the accuracy of remote sensing measurements, especially in the visible and NIR spectrum regions.

3. Non-selective scattering is the third type of spectral scattering and occurs when the particle sizes are much larger than the radiation's wavelength (e.g., ice particles) (Tomasi, 1979). This size-independent scattering type can affect the electromagnetic spectrum's entire range. Larger particles in the atmosphere, such as clouds, fog, and smoke, generally cause non-selective scattering. These particles can scatter light equally in all directions, reducing the light beam's intensity and causing a decrease in the observed objects' contrast and color saturation. Approximately 30% of the total radiance in the blue region of the electromagnetic spectrum is influenced by non-selective scattering (Berk et al., 2006). Nonselective scattering can significantly impact the accuracy of remote sensing measurements, particularly in the VIS and NIR spectrum regions.

As solar irradiance navigates through the atmosphere en route to a surface and viewing instrument, it undergoes atmospheric absorption and scattering, influenced by various factors, including solar observation geometry, terrain altitude, and sensor viewing direction. The optical thickness, a measure of the distance solar radiation travels through the atmosphere, is a product of the atmospheric extinction coefficient and path length (Berk et al., 2006, Nakajima and King, 1990).

Four major fluxes associated with atmospheric absorption and scattering include solar irradiance, atmospheric irradiance, surface reflectance, and radiance observed by the sensor. However, these atmospheric effects can compromise the accuracy of remote sensing measurements, particularly in the visible and NIR region. To rectify this, atmospheric correction techniques are employed to eliminate atmospheric effects from remotely sensed data (Yang et al., 2020a). These techniques model the scattering and absorption processes in the atmosphere to adjust the remotely sensed data, enabling accurate interpretation of surface properties. The following are key factors that impact optical remotely sensed data:

1. One flux associated with atmospheric scattering is path radiance. In remote sensing, path radiance signifies the amount of radiation a sensor observes along the path between itself and a target object, influenced by various atmospheric factors (Kaufman, 1993). Path radiance is impacted by molecular and particulate scattering, two types of atmospheric scattering. As previously mentioned, molecular Rayleigh scattering transpires due to interactions between electromagnetic radiation and atmospheric molecules, and its contribution to path radiance depends on the radiation's wavelength. Specifically, Rayleigh scattering decreases proportionally to the fourth power of the wavelength, or λ^{-4}. The general path radiance term amalgamates these different types of scattering and can significantly influence the amount and distribution of electromagnetic radiation a sensor observes along a specific path. Thus, it is crucial to correctly correct path radiance in remote sensing applications to guarantee accurate and reliable results. By accounting for path radiance effects, remote sensing algorithms can isolate the target object's signal more effectively and reduce atmospheric interference's impact on final measurements.

2. Unlike direct solar radiation, which travels unscattered from the Sun to the surface of

the Earth, sunlight scattered by the atmosphere in all directions, including toward the surface of the Earth, is called Diffuse solar radiation (Liu and Jordan, 1960). Diffuse solar radiation is a vital flux that can impact measurement accuracy. It can cause the surface to appear brighter and shadows less distinct, thereby complicating distinguishing between different surface features. Moreover, diffuse radiation can alter the reflectance properties of surfaces by causing multiple reflections and scattering of light within the surface material, which could result in errors in remote sensing measurements. The observed reflectance needs to represent the actual surface properties accurately. Therefore, it is critical to account for diffuse solar radiation's effects in remote sensing applications.

3. Adjacency radiance, another flux associated with atmospheric effects on remote sensing measurements, refers to the reflected radiation from areas neighboring the observed target, scattered by the air volume, and entering the sensor's Instantaneous Field of View (IFOV) (Bulgarelli and Zibordi, 2018). The amount of adjacency radiance entering the sensor's IFOV can significantly affect the accuracy of remote sensing measurements, particularly in the VIS and NIR regions of the spectrum. This flux relies on several factors, including solar observation geometry, the surface of the Earth and atmosphere properties, and the sensor's characteristics.

Taking into account the various atmospheric effects previously discussed, the total radiance measured by an optical sensor can be expressed as follows (Zhao et al., 2024, Slater et al., 1987):

$$L_{\text{Total radiance}}(\lambda) = L_{\text{Path radiance}}(\lambda) + L_{\text{Reflected}}(\lambda) + L_{\text{Adjacency radiance}}(\lambda) \tag{2.8}$$

Here, $L_{\text{Total radiance}}(\lambda)$ is the total radiance measured by the sensor, $L_{\text{Path radiance}}(\lambda)$ is the path radiance caused by scattering and absorption in the atmosphere, $L_{\text{Reflected}}(\lambda)$ is the radiance reflected from the surface of the Earth, and $L_{\text{Adjacency radiance}}(\lambda)$ is the adjacency radiance scattered by air volumes adjacent to the target being observed. Atmospheric correction techniques are imperative to remove the influence of atmospheric scattering and absorption from the remotely sensed data, thereby securing accurate measurements of the reflected radiance from the surface of the Earth.

2.5 Other factors impacting optical remote sensing data

Accurate sensor calibration, image pre-processing, and spectral analysis are fundamental in managing factors that impact optical remote sensed data. They ensure the accuracy and reliability of remote sensing data for various scientific and practical applications, such as environmental monitoring, land use analysis, and disaster management. Geo-rectification is a vital process in optical remote sensing, where geometric distortions in the images obtained are corrected to align them with the surface of the Earth (Schläpfer et al., 2000). This is crucial because remote sensing images, often obtained from a distance, are prone to various factors causing geometric distortions, including sensor orientation, terrain relief, atmospheric effects, and sensor motion. Georectification is accomplished using ground control points (GCPs), reference points on the surface of the Earth with known geographic coordinates (Smhh and Atkinson, 2001). The

GCPs calculate transformation parameters that can warp and stretch the raw image to match ground coordinates accurately. The georectification process's accuracy is critical for various remote sensing applications, including mapping, land use analysis, and environmental monitoring. Proper selection of GCPs, appropriate image pre-processing, and suitable georectification algorithms are essential to achieve high-quality and accurate remote sensing data. Alongside georectification, other factors such as spectral distortion, sensor degradation, and spectral shift can significantly impact the accuracy and reliability of optical remote sensing data.

1. Spectral distortion can be induced by various factors, including sensor motion during data acquisition, atmospheric turbulence, and other environmental factors (Yokoya et al., 2010). This can lead to non-linear distortion in the image, affecting the data's accuracy. It is crucial to manage spectral distortion using suitable correction methods (e.g., smile correction) (Ceamanos and Douté, 2010). Corrections can be achieved using different approaches, such as motion sensors or ground reference points. These methods can improve the accuracy and reliability of the remotely sensed data for various scientific and practical applications like environmental monitoring, land use analysis, and disaster management.

2. Sensor degradation refers to the gradual decline of sensor performance over time, resulting from various factors such as radiation exposure, temperature fluctuations, and mechanical wear and tear (Christophe et al., 2005). This leads to sensor spectral sensitivity changes, affecting the data's accuracy and reliability. Sensor degradation can cause substantial changes in sensor performance by altering its electronic properties, leading to a non-uniform sensor response.

3. Spectral shift is a phenomenon that can occur in remote sensing data when there is a displacement of wavelength positions in the acquired image (Tsai and Philpot, 1998). Several factors, including atmospheric absorption and scattering, sensor calibration errors, and instrument noise, can cause this displacement. Atmospheric gases such as water vapor, ozone, and carbon dioxide can absorb and scatter light in the atmosphere, causing a shift in spectral wavelengths (Guanter et al., 2006). This can result in errors in the interpretation of the remote sensing data, particularly when detecting surface features. Calibration errors can also contribute to spectral shift when the sensor is not calibrated correctly or there is a change in the sensor's performance over time. This can lead to changes in the sensor's spectral sensitivity, causing a shift in spectral wavelengths. Instrument noise can also contribute to spectral shifts in remote sensing data. Various factors, including electronic, thermal, and detector noise, can cause this. These noise sources can introduce random fluctuations in the measured signal, leading to spectral analysis and interpretation errors.

2.6 Topographical correction

In addition to the mentioned factors, topography is crucial in determining the amount of radiation that reaches a surface, especially in optical remote sensing. Variations in terrain altitude

can lead to differences in illumination conditions, which in turn can result in inaccurate quantification of land surface characteristics, such as vegetation indices (Richter et al., 2009) (Figure 2.2). For example, rugged terrain can cause shadows and overexposure in optical data, leading to errors in interpreting optical remotely sensed data. For the accurate capture of spectral measurements in remote sensing, it is essential to account for topographical factors influencing upward radiation.

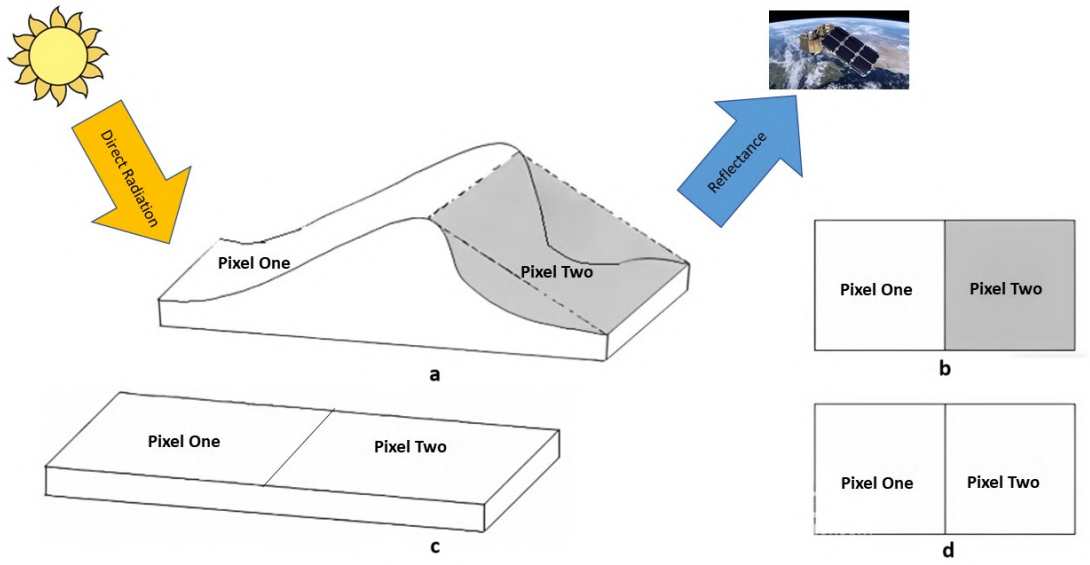


Figure 2.2: Highlighting the influence of rough terrain on incoming solar radiation and its effect on sensor-detected reflectance as seen in Figures a and b.

The surface of the Earth can be smooth, corrugated, or porous. Each surface type has different scattering properties that influence the incoming radiation. The scattering of incoming radiation can be classified as forward scatter, isotropic, or backscatter. Understanding these surface characteristics and the scattering of incoming radiation is crucial for topographic correction. Topographic correction is a necessary pre-processing step in optical remote sensing that removes topographical effects to accurately quantify land surface characteristics and vegetation indices.

Several studies have underscored the importance of topographic correction in accurately monitoring vegetation indices, including N, which is essential for crop health and management (Berger et al., 2020b). For instance, forward scattering occurs when incoming radiation is scattered in the forward direction, often associated with smooth surfaces like water bodies (Hudson and Knopoff, 1967). Isotropic scattering happens when radiation is scattered equally in all directions, typically associated with diffuse surfaces such as forests (Clark, 1983). Backscattering occurs when radiation is scattered back towards the source and is often associated with rugged and porous surfaces like mountainous regions (Sullivan et al., 2013). Understanding the scattering of incoming radiation and surface characteristics is thus critical for accurate topographic

correction in remote sensing. By accounting for these effects, obtaining more accurate and precise measurements of land surface characteristics, including vegetation indices like protein and chlorophyll content, is possible.

1. The three scattering factors caused by a surface interacting with incoming radiation are forward, isotropic, and backscatter (Jin, 1993):

   (a) Forward scattering occurs when the wavelength of the incoming radiation is smaller than the size of the observed surface or object. This type of scattering often occurs on smooth surfaces like water bodies, and it is associated with refraction, focusing effects, and backscatter through internal reflection (Han et al., 2014, Diner and Martonchik, 1985). The reflected radiation from these surfaces forms a plane, defined by the incident radiation ray and the surface normal, making an angle with the local standard equal to the angle of incidence. The result is a diffuse reflection of the incoming radiation scattered in the forward direction, often leading to bright areas in the resulting imagery. Understanding the effects of forward scattering is critical for accurate topographic correction in remote sensing.

   (b) Isotropic scattering is caused by small particles or objects with a rugged surface and interior irregularities (Verhoef, 1998). In this type of scattering, the incoming radiation can penetrate the objects from various angles. Isotropic scattering is often observed in small particles and other dielectrics. Opaque objects with a smooth surface that is randomly oriented also exhibit an average, isotropic micro scattering law (Sundberg et al., 2010). This type of scattering occurs when the radiation is scattered equally in all directions, regardless of the direction of incidence. The result is a diffuse reflection of the incoming radiation, scattered uniformly in all directions. Isotropic scattering is vital in topographic correction in remote sensing, particularly in areas with diffuse surfaces such as forests.

   (c) Backscattering is a type of micro-scattering that occurs in opaque objects with rugged surfaces that are neither metallic (causing strong reflections) nor transparent (causing refraction) (Fung et al., 1992, Santoro et al., 2011). When the light source is directly behind such objects, they appear bright to an observing sensor, hence the term "backscatter." In backscattering, the optical sensor only sees the illuminated face of the object, resulting in a bright appearance. However, the surface appears less bright when viewed from other angles, creating a shadow-like effect. Backscattering is often observed in rugged and porous surfaces such as mountainous regions and can impact the accuracy of remote sensing measurements. Understanding the effects of backscattering is crucial for accurate topographic correction in remote sensing, particularly in areas with rugged and porous surfaces.

2. Surfaces with a high albedo diffusively reflect light, independent of micro-scattering properties (Wang et al., 2014). Albedo is a dimensionless measure that represents the ratio of reflected radiation to incident radiation over an entire surface area, accounting for all wavelengths and viewing angles. In contrast, surface reflectance refers to the proportion

of incident radiation reflected at a specific wavelength and angle of observation. A surface with a high albedo reflects more incident radiation, resulting in a brighter appearance, while a low-albedo surface absorbs more radiation and appears darker. Albedo plays a critical role in topographic correction in remote sensing, as it determines the total radiation reflected onto the sensor, integrating overall angular and spectral contributions.

3. The albedo of a surface plays a critical role in how a remote sensing sensor detects and measures radiation. A surface with a low albedo tends to absorb more incident radiation, but how it reflects the remaining radiation depends on its scattering properties. For instance, a dark object with a smooth surface reflects light into a narrow specular peak, a sharp, mirror-like reflection in a specific direction. Conversely, a surface with isotropic scattering characteristics reflects light in a broad backscatter peak, a diffuse reflection in all directions. The scattering properties of a surface, whether it exhibits forward, isotropic, or backscattering, determine how radiation is detected and measured by the sensor. Considering these properties is crucial for obtaining accurate and precise measurements of land surface characteristics, including vegetation indices, through topographic correction in remote sensing.

Topographic correction is a crucial pre-processing step in optical remote sensing for accurately quantifying land surface characteristics, including biophysical and biochemical properties (de Oliveira et al., 2021). Terrain altitude variations can cause differences in illumination conditions, leading to inaccuracies in quantifying land surface characteristics. In rugged terrain, shadows and overexposure to optical data can lead to errors in interpreting remotely sensed data. To account for these topographical effects, the observed radiance is transformed into a flat terrain using one of several topographic correction methods (Figure 2.3).

The three commonly used topographic correction methods in the literature are empirical normalization, geometrical correction techniques, and more advanced procedures that account for non-Lambertian surfaces (Vázquez-Jiménez et al., 2017). The simplest method is empirical normalization, which uses a simple ratio of bands. The geometrical correction techniques assume that the surface is Lambertian or non-specular and reflects light equally in all directions. However, because most surfaces are non-Lambertian, more advanced procedures that assume both Lambertian and non-Lambertian surfaces have been developed. The Minnaert correction method accounts for the non-Lambertian properties of a surface, such as its anisotropy or directional reflectance. C-correction is another approach that corrects for topographical effects on the radiance signal by accounting for the surface slope and aspect angle, which affect the amount of illumination a surface receives. Sun-canopy-sensor (SCS+C) correction is a third commonly used approach that accounts for the three-dimensional effects of the terrain and vegetation structure on incoming radiation. Some of the most commonly used topographical corrections are as follows:

1. Cosine correction is a widely used topographic correction technique in remote sensing. It involves normalizing the reflected radiance of an observed region based on the assumption that the total irradiance of a unit area is proportional to the cosine of the incidence angle.

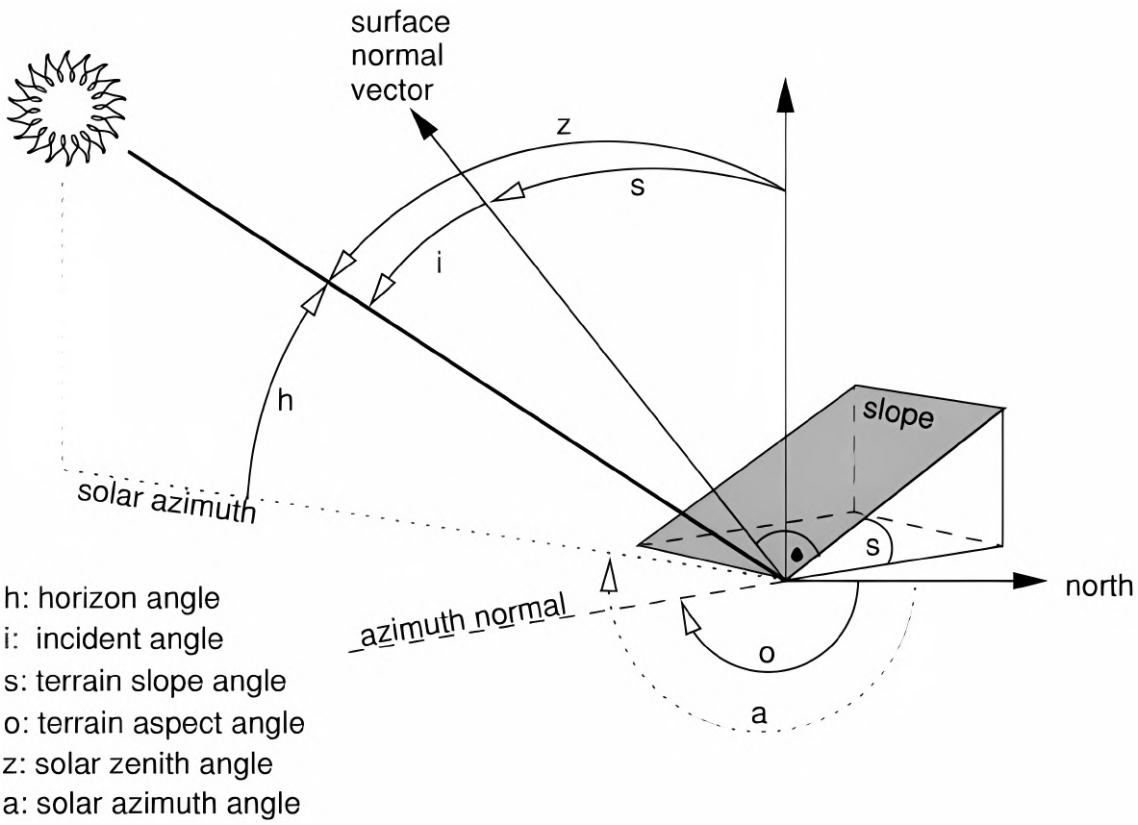


Figure 2.3: Visual representation of the correlation between terrain features and solar angles. Adapted from `https://grass.osgeo.org/grass82/manuals/i.topo.corr.html`.

The cosine correction function is given by:

$$L_n = L\left(\frac{\cos\theta}{\cos i}\right) \tag{2.9}$$

In this equation, L_n represents the normalized reflectance, L represents the uncorrected reflectance, θ represents the solar zenith angle, and i represents the incidence angle. However, a limitation of the cosine model is that it does not account for the diffuse irradiance caused by atmospheric and terrain sources (Vázquez-Jiménez et al., 2017). Regions with low illumination conditions caused by irradiance will receive a considerable proportion of diffuse radiation, resulting in high levels of brightness calculated by the cosine correction. As a result, the cosine correction may not be suitable for correcting topographic effects in areas with significant diffuse radiation. This limitation has led to more advanced topographic correction methods that consider the effects of atmospheric and terrain sources on reflected radiance (Vázquez-Jiménez et al., 2017). Despite this limitation, cosine correction remains a useful and widely used method in remote sensing for simple topographic correction tasks.

2. The Minnaert Correction is a method used for the topographic correction of optical data. One of the critical limitations of the cosine correction model is that it assumes all surfaces are Lambertian (Proy et al., 1989, Kimes and Kirchner, 1981, Conese et al., 1993). The

Minnaert Correction addresses this limitation by introducing a constant value to estimate the non-Lambertian nature of a surface. However, it is not capable of representing the scattering direction. The Minnaert Correction model can calculate the strength of the relationship between radiance and view angle using a constant value (Minnaert, 1941, Justice et al., 1981). This constant value is calculated through a linear regression between the exitance angle, which is the radiant flux leaving a surface per unit surface area, and the cosine of the angle of incidence (Smith et al., 1980, Colby, 1991). The slope of the regression line is then used as the Minnaert constant, which acts as a correction factor to compensate for the over-correction of illumination or brightness due to the incidence angle.

3. The statistical-empirical correction methods are based on the correlation between the surface illumination condition and the terrain characteristics, such as the slope angle (Vázquez-Jiménez et al., 2017). These methods use empirical regression to normalize the reflectance data considering topographical factors and rotate the regression line horizontally, enabling the transformation of an object/surface radiance independent of the incidence angle (Vázquez-Jiménez et al., 2017). However, the effectiveness of this approach is highly dependent on the strength of the correlation between the terrain characteristics and the illumination conditions (Gao and Zhang, 2009).

4. The C-Correction method attempts to address the limitations of the cosine correction model by introducing a moderator parameter to the cosine correction model. This constant, known as C, is calculated based on the linear relationship between the cosine of the incidence angle (i) and the observed reflectance (L) (Teillet et al., 1982). The parameter C is similar to the impact of diffuse Sky irradiance and is introduced to the cosine correction model as an additive component to compensate for over-correction due to incidence angle. The C-Correction method can preserve the spectral information captured by sensors and is commonly used for correcting illumination conditions in rugged terrain (Meyer et al., 1993, Riaño et al., 2003). Equation 2.10 shows the linear relationship between the observed reflectance and the cosine of the incidence angle, where a and b are constant coefficients (Vázquez-Jiménez et al., 2017). Equation 2.11 shows the C-Correction function, where L_n is the normalized reflectance, θ is the solar zenith angle, and C is the correction factor.

$$L = a + b\cos i \quad C = a/b \tag{2.10}$$

$$L_n = L((\cos\theta + C)/(\cos i + C)) \tag{2.11}$$

5. The Sun-Canopy-Sensor Correction (SCS) model is a topographical correction method that improves the cosine correction model by correcting the illuminated vegetation regions (Gu and Gillespie, 1998). The SCS model projects the illuminated vegetation cover from a sloped surface to a horizontal one to undertake normalization. By assuming the reflected radiation from an illuminated surface is independent of topographical factors, the integrated reflectance from the irradiated canopy surface will be proportional to its

area, where the terrain slope is considered (Gu and Gillespie, 1998). The SCS model is represented by equation 2.12, where α represents the terrain slope, i represents the incidence angle, and θ represents the solar zenith angle (Vázquez-Jiménez et al., 2017).

$$L_{\mathrm{n}} = L((\cos\theta \cdot \cos\alpha)/(\cos i)) \tag{2.12}$$

Although the SCS model has achieved better performance than the cosine correction model when undertaking topographical correction, it does not consider the impact of diffuse irradiance. Consequently, the SCS model may over-correct spatial regions where the unit area's slope does not face the illumination source.

6. The SCS model is a widely used topographic correction approach that corrects the illumination conditions of vegetation regions. However, the SCS model is limited by an over-correction issue in regions with slopes not facing the illumination source. To address this limitation, the SCS+C correction method combines the constant component from the C-Correction model with the SCS model, reducing over-correction issues. The SCS+C correction model introduces a moderator parameter, C, to the SCS model, which acts as an additive component to account for the impact of diffuse irradiance. By incorporating the C constant into the SCS model, the SCS+C model can accurately correct the illumination conditions of vegetation regions, reducing the over-correction commonly encountered with the SCS model. The SCS+C correction equation is shown in equation 2.13, where θ is the solar zenith angle, α is the terrain slope, and i is the incidence angle.

$$L_{\mathrm{n}} = L((\cos\theta \cdot \cos\alpha + C)/(\cos i + C)) \tag{2.13}$$

2.7 Optical remote sensing instruments

In remote sensing, various optical sensing instruments are used to monitor objects. Each instrument has unique specifications that determine its performance, such as spectral resolution, radiometric resolution, signal-to-noise ratio (SNR), and spatial resolution (Jeziorska, 2019). Therefore, selecting the appropriate optical sensing instrument for a specific application is critical. Below are vital factors about optical remote sensing instruments:

1. Radiometric resolution is an essential parameter in remote sensing that refers to the sensitivity of a sensor to detect slight differences in the intensity of electromagnetic radiation (Navalgund et al., 2007). It determines the degree of detail a sensor can capture in an image, particularly in quiet or low-contrast areas. A sensor with high radiometric resolution can distinguish between objects with slight differences in brightness values, leading to a more accurate and precise measurement. The radiometric resolution is expressed in the number of bits used to digitize the sensor data, where higher radiometric resolution provides better image quality and more accurate measurements (Verde et al., 2018).

2. FOV (Field of View) in remote sensing refers to the total angular extent of the scene visible to the imaging system. The maximum angle the sensor can capture in a single image is determined by the sensor's optical design (Schläpfer et al., 2014). A larger FOV can capture a larger scene but may also result in lower spatial resolution, while a smaller FOV can provide higher spatial resolution. However, it may require multiple images to cover a larger area. FOV and IFOV (Instantaneous Field of View) are related parameters, but they describe different aspects of the remote sensing imaging system. While IFOV measures the system's spatial resolution and is defined as the angle subtended by a single detector element on the optical system's axis, FOV determines each image's coverage area. FOV and IFOV also impact the spectra that a remote sensing imaging system can capture (Cracknell, 1998). The size of the FOV determines the amount of background radiation that can be captured along with the target signal, which can affect the accuracy of the measurements. In addition, the IFOV determines the level of spectral mixing between different targets within a single pixel. A smaller IFOV can better isolate the signal from a single target and reduce the spectral mixing. A larger IFOV may result in more mixing between different targets within a pixel. Accurate characterization and calibration of both FOV and IFOV are crucial for ensuring that remote sensing data is high quality and can be used effectively for scientific and practical applications, including spectral analysis and interpretation.

3. Spectral resolution refers to an optical sensor's ability to capture information about a scene at specific wavelengths. It is determined by the bandwidth and the number of spectral bands that the sensor can detect, its optical design, detection technology, detector element size, and SNR (Liu et al., 2015). SNR describes the strength of a signal compared to the background noise within a system. The signal represents the amount of electromagnetic radiation detected by the sensor from the scene, while noise includes any unwanted electromagnetic radiation or electronic interference that could degrade the measurement. Typically, a high SNR indicates a high-quality measurement where the signal is much stronger than the noise (Joiner et al., 2013), while a low SNR leads to reduced measurement accuracy and precision. Various factors affect SNR, such as exposure time and ambient temperature. Increasing exposure time can improve the signal strength relative to the noise, thereby increasing SNR. However, increasing sensitivity can amplify both the signal and noise, potentially offsetting the benefits of increased signal detection. Therefore, balancing signal enhancement and noise reduction is vital for obtaining high-quality measurements. A sensor with high spectral resolution can distinguish between different materials based on their unique spectral signatures, enabling accurate object identification and classification.

Based on the spatial coverage, optical sensing instruments can be broadly divided into three groups, as follows (Figure 2.4):

1. Aerial optical sensing refers to sensors mounted on fixed-wing or rotary-wing aircraft, as well as Unmanned Aerial Vehicles (UAVs). Aerial optical sensing has several advantages

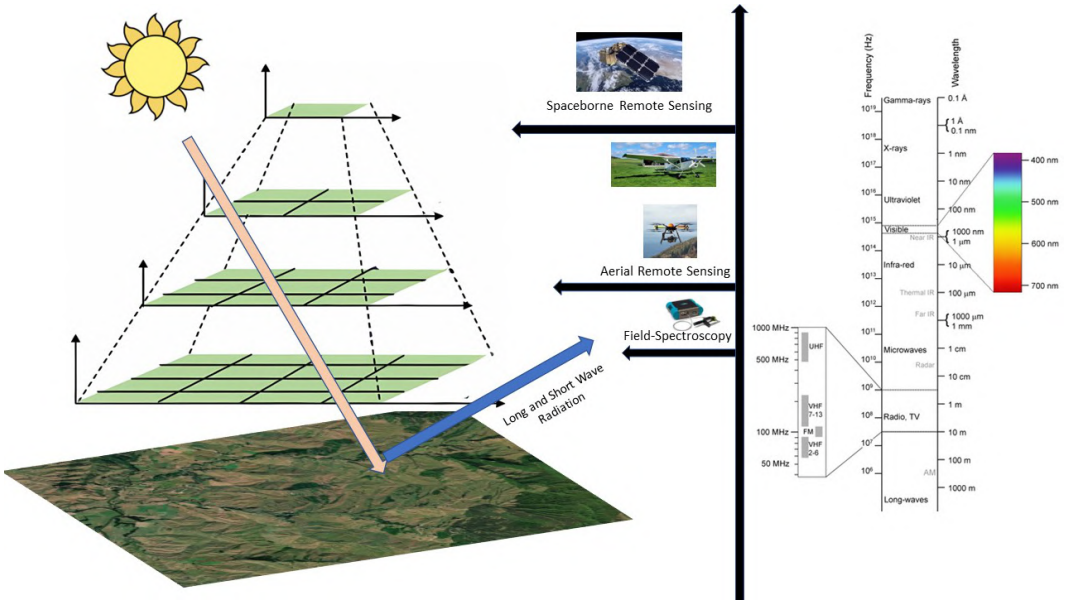


Figure 2.4: Depiction of different types of optical sensing techniques is categorized based on spatial coverage.

over other sensing methods. It provides a more comprehensive coverage area, more systematic data collection, and better spatial resolution than ground-based sensing methods, making it ideal for small-scale and regional commercial applications. Additionally, it is less expensive and time-consuming than traditional methods, making it a more efficient data collection method. One example of aerial optical sensing in action is the study conducted by Pullanagari et al. (2016), where hyperspectral imagery was collected using an aircraft equipped with a hyperspectral sensor. The researchers used hyperspectral imagery to develop maps of nutrient concentration, specifically nitrogen (N), in mixed grasslands. By analyzing the spectral signatures of different vegetation types in the hyperspectral imagery, the researchers could estimate the amount of N in the vegetation, an essential nutrient for plant growth. Land managers can use these maps of nutrient concentration to optimize fertilizer application and improve the health and productivity of grasslands.

2. Similar to aerial optical sensing, spaceborne optical sensing involves using sensors to collect optical data. However, in this case, the sensors are mounted on a satellite orbiting the Earth, which provides a significantly larger coverage area than aerial optical sensing. This type of optical sensing is commonly used for large-scale scientific applications, owing to its ability to provide data over vast regions of the surface of the Earth. The data collected by spaceborne sensors can be used for various applications, including environmental monitoring, disaster management, and climate studies. Gao et al. (2020b) used spaceborne imagery to monitor nutrient ratios in the Alpine Grassland Ecosystem of the Tibetan Plateau and found that it is possible to use MODIS (Moderate Resolution Imaging Spectroradiometer) to measure the N and phosphorus ratio of the vegetation cover. EnMAP (Environmental Mapping and Analysis Program) is a German hyperspectral satellite mission designed to monitor and analyze the Earth's environment (Guanter et al., 2015). It provides detailed spectral information across VIS, NIR, and SWIR. EnMAP supports

various applications, such as vegetation monitoring, water quality assessment, soil analysis, and urban studies, by offering high spectral and spatial resolution data. PRISMA (PRecursore IperSpettrale della Missione Applicativa) is an Italian hyperspectral imaging satellite mission developed by the Italian Space Agency (ASI) (Acito et al., 2022). It features a hyperspectral sensor that captures data from the visible to the SWIR. PRISMA is primarily used for environmental monitoring, natural resource management, and agricultural applications, providing detailed spectral information to support these activities.

3. Proximal sensing instruments are designed to collect high-quality spectral signatures of objects close to the instrument. These instruments are widely used in scientific experiments to obtain precise and accurate measurements of the spectral characteristics of material. Proximal sensing instruments are commonly used in agriculture, forestry, and ecology research, as well as in industrial applications such as quality control and material testing. However, the main limitation of proximal sensing instruments is their small coverage area, which makes them less useful for monitoring large spatial domains. Researchers often must sample multiple locations to obtain representative spectral signatures of an area of interest, which can be time-consuming and labor-intensive. To address this limitation, Pullanagari et al. (2021) proposed using empirical approaches to estimate grassland N concentration (N%) using proximal sensing instruments. In their study, the researchers collected proximal sensing data from multiple locations in grasslands and combined them with field measurements of N% to develop a predictive model. The model was then used to estimate the N% of grasslands across a large spatial domain. Using empirical approaches to calibrate proximal sensing data, the researchers overcame the limitation of a small coverage area and obtained accurate estimates of N% across a large spatial domain. This approach could have important implications for improving our understanding of nutrient cycling and ecosystem function in grassland ecosystems and other ecosystems with similar limitations. Overall, using proximal sensing instruments, combined with empirical approaches, can provide valuable insights into the spectral characteristics of objects at high accuracy and precision.

This categorization based on spatial coverage is essential for selecting the most appropriate optical sensing instrument to achieve the desired level of detail in the monitoring task. Multispectral (MS) and hyperspectral (HS) sensors are commonly used optical sensing instruments in remote sensing applications. Instruments like HS sensors, which capture narrowband spectral features in the electromagnetic spectrum, can provide detailed information on the chemical composition of objects (Dehghan-Shoar et al., 2023a). In contrast, MS instruments capture broader spectral information and are helpful for various applications, including land use/land cover mapping, change detection, and environmental monitoring (Dehghan-Shoar et al., 2023b,c).

Instruments which capture narrow band spectral information are designed to capture spectral signals at very fine spectral resolutions. They can provide detailed information on the spectral properties of objects, including subtle differences in vegetation characteristics, such as vegetation protein content and stress (Féret et al., 2021, Mohammed et al., 2019). This

makes HS instruments useful for scientific applications, such as mapping vegetation chlorophyll content. However, HS instruments have some disadvantages compared to MS instruments, including their higher cost and the more significant data storage requirements for HS data (Dehghan-Shoar et al., 2023a). HS data can also be more complex to process and analyze than MS data, which can be challenging for some applications. On the other hand, MS instruments can measure general vegetation characteristics using vegetation indices like the Normalized Difference Vegetation Index (NDVI) (Schinasi et al., 2018). MS instruments cannot distinguish subtle differences in vegetation characteristics such as diseases and stress, as they can only measure generic characteristics like high or low vigor (Guerini Filho et al., 2020). In contrast, HS instruments are capable of identifying physiological crop traits. They can accurately estimate vegetation nutrient deficiencies by correlating spectral signature with changes in plant physiology (Pullanagari et al., 2016).

Figure 2.5: Molecular electromagnetics resonance overtones (Féret et al., 2021)

2.8 Methods for Monitoring Vegetation Properties

Measured reflectance is a fundamental aspect of remote sensing as it provides valuable information about the optical properties of objects (Curran, 1989, Mutanga, 2004a) (Table 2.1). Reflectance arises from the combined effects of scattering, absorption, and radiation transmission within the vegetation canopy. The amount of radiation the vegetation canopy reflects is closely related to the vegetation's biochemical and biophysical variables. For instance, leaf chlorophyll primarily absorbs VIS radiation for photosynthesis, making VIS reflectance useful for monitoring leaf chlorophyll content, an essential indicator of plant health (Feret et al., 2008).

Table 2.1: Absorption features related to particular biochemicals, demonstrating the specific wavelengths at which certain compounds exhibit unique absorption characteristics, pivotal for remote sensing and biochemical analysis (Curran, 1989, Mutanga, 2004a, Kumar et al., 2001, Elvidge, 1990)

Wavelength (nm)	**Absorbing Compounds**	**Absorption Mechanism**
430	Chlorophyll a	Electron transition
460	Chlorophyll a	Electron transition
640	Chlorophyll b	Electron transition
660	Chlorophyll a	Electron transition
910	Protein	C-H stretch, 3rd overtone
930	Oil	C-H stretch, 3rd overtone
970	Water, starch	O-H bend, 1st overtone
990	Starch	O-H stretch, 2nd overtone
1020	Protein	N-H stretch
1040	Oil	C-H stretch, C-H deformation
1200	Lignin	C-H stretch, 2nd overtone
1220	Water, cellulose, starch, lignin	O-H bend, 1st overtone
1400	Water	O-H bend, 1st overtone
1420	Lignin	C-H stretch, C-H deformation
1450	Starch, sugar, water, lignin	O-H stretch, 1st overtone
1490	Cellulose, sugar	C-H stretch, C-H deformation
1510	Protein, nitrogen	N-H stretch, 1st overtone
1530	Starch	O-H stretch, 1st overtone
1540	Starch, cellulose	O-H stretch, 1st overtone
1580	Starch, sugar	O-H stretch, 1st overtone
1690	Lignin, starch, protein	C-H stretch, 1st overtone
1730	Protein	C-H stretch
1736	Cellulose	O-H stretch
1780	Cellulose, sugar, starch	C-H stretch, 1st overtone
1820	Cellulose	O-H stretch, C-O stretch
1900	Starch	O-H stretch, H-O-H deformation
1924	Cellulose	O-H stretch, O-H deformation
1940	Water, protein, lignin, cellulose,	O-H stretch, C-H deformation
1960	Starch, nitrogen	O-H stretch, O-H rotation
1980	Starch, sugar	O-H stretch, C-O stretch
2000	Protein	N-H asymmetry
2060	Starch	O-H deformation, C-O deformation
2080	Protein, nitrogen	N-H stretch, N-H rotation
2100	Starch, sugar	O-H stretch, O-H deformation, C-O-C stretch
2130	Protein	N-H stretch
2180	Protein, nitrogen	N-H rotation, C-H stretch, C-O stretch, C=O stretch
2240	Protein	C-H stretch
2250	Starch	O-H stretch, O-H deformation
2270	Cellulose, sugar, starch	C-H stretch, O-H stretch, C-H rotation, CH2 rotation
2280	Starch, cellulose	C-H stretch, CH2 deformation
2300	Protein, nitrogen	CH2 rotation, C=O stretch, N-H stretch
2310	Oil	C-H bend, 2nd overtone
2320	Starch	C-H stretch, C-H deformation
2340	Cellulose	C-H stretch, O-H deformation
2350	Cellulose, nitrogen, protein	CH2 rotation, C-H deformation

The NIR region of the electromagnetic spectrum is most suitable for estimating the canopy's biomass and Leaf Area Index (LAI) (Friedl et al., 1994, Darvishzadeh et al., 2008a, Atzberger et al., 2015). This is because NIR radiation penetrates the vegetation canopy and reflects from the underlying structure, such as branches and stems. This reflected radiation is related to the canopy structure, such as the number and orientation of leaves and provides information about the canopy's biophysical properties. In contrast, the SWIR region includes information on plant water, protein content, and other biochemical properties. The SWIR region is absorbed by water and organic compounds, and the amount of radiation reflected is related to the water and biochemical content of the vegetation canopy (Figure 2.5) (Berger et al., 2018a, Curran, 1989). Using multiple spectral bands in remote sensing can provide a more comprehensive understanding of vegetation's biochemical and biophysical properties, making it a valuable tool for various applications. Observing and analyzing these spectral properties of vegetation provides valuable information about plant growth, health, and ecological functioning. Remote sensing is used in various fields, including agriculture, forestry, ecology, and climatology, to monitor changes in vegetation and the environment (Obermeier et al., 2019, Dorigo et al., 2007). Therefore, understanding reflectance and its relationship with vegetation properties is essential for remote sensing applications, especially those related to managing and monitoring natural resources.

Chlorophyll is a widely used proxy for estimating N in vegetation due to its close relationship with photosynthesis. However, N-containing biochemical constituents, such as proteins, have absorption features in different parts of the electromagnetic spectrum (Curran, 1989). Only a fraction of N is present in chlorophyll pigments (less than 15% wt), which makes it challenging to estimate N using only chlorophyll content, particularly across different seasons and locations (Fourty et al., 1996, Darawsheh and Bouranis, 2006). Vegetation protein content, on the other hand, is the primary repository of N in grasslands and can be detected in the SWIR part of the electromagnetic spectrum (Berger et al., 2018a,b, 2020a). Thus, to accurately estimate vegetation N, it is crucial to identify and use appropriate spectral bands across the electromagnetic spectrum, including the SWIR region where the absorption features of vegetation protein content are present (Curran, 1989). Several studies have emphasized incorporating multiple spectral bands across the electromagnetic spectrum to estimate vegetation N accurately. For example, Dechant et al. (2017) confirmed the importance of SWIR absorption features in accurately estimating vegetation N (Dechant et al., 2017). Similarly, Asner et al. (2011) reported that leaf biochemical traits could not be accurately predicted using information captured only from the VIS absorption features (Asner et al., 2011). Therefore, to obtain more accurate and precise measurements of vegetation N, it is necessary to use multiple spectral bands across the electromagnetic spectrum, including the SWIR region, to accurately capture the absorption features of different biochemical constituents.

Optical sensing technology has numerous applications for estimating vegetation characteristics, including LAI, chlorophyll content, vegetation water content, aboveground biomass, and protein content. LAI is commonly estimated using the NDVI and has been widely used for vegetation monitoring (Friedl et al., 1994, Darvishzadeh et al., 2008a, Atzberger et al., 2015).

Reflectance data in the visible range of the electromagnetic spectrum can be used to estimate chlorophyll content in vegetation (Darvishzadeh et al., 2011, Tong and He, 2017). Meanwhile, reflectance data in the SWIR range of the electromagnetic spectrum is used for estimating vegetation water content (Davidson et al., 2006, He et al., 2017). Aboveground biomass can be estimated using VIS, NIR, and SWIR reflectance data. Recent studies have demonstrated the efficacy of combining data from multiple spectral bands to estimate aboveground biomass, including high-resolution satellite data (Jin et al., 2014, Ali et al., 2017, Knoblauch et al., 2017, Quan et al., 2017, He et al., 2019). Protein content, which is essential for grasslands, can be estimated using reflectance data in the SWIR range of the electromagnetic spectrum (Pullanagari et al., 2021, Clevers and Kooistra, 2012, Pullanagari et al., 2011, Ullah et al., 2012). As protein content can be used as a proxy for estimating N content, critical for crop management and health, accurate estimation of vegetation protein content is important in remote sensing.

Measured reflectance is a fundamental aspect of remote sensing and provides valuable information about the optical properties of objects. Understanding the relationships between reflectance and vegetation characteristics, such as LAI, chlorophyll content, vegetation water content, aboveground biomass, and protein content, is essential in various fields such as agriculture, forestry, ecology, and climatology for monitoring changes in vegetation and the environment. While chlorophyll content is commonly used as a proxy for estimating N in vegetation, using multiple spectral bands across the electromagnetic spectrum, including the SWIR, is necessary for more accurate and precise measurements of vegetation N content. Incorporating multiple spectral bands in remote sensing can provide a more comprehensive understanding of vegetation's biochemical and biophysical properties, making it a valuable tool for various applications.

2.9 Estimation of Vegetation Biochemical and Biophysical Properties

Three primary methods are commonly employed to estimate the biochemical and biophysical properties of grassland vegetation using optical sensing data: (1) Empirical-statistical methods, (2) Physically-based models, and (3) Hybrid approaches.

2.9.1 Empirical-statistical methods

Empirical-statistical methods are a common approach used to estimate vegetation biochemical and biophysical properties and machine learning algorithms are often utilized to create high-performance and robust models (Pullanagari et al., 2021, 2011, Kuo and Sloan, 2005). These models are beneficial for estimating N content, as they can capture complex, nonlinear patterns in high-dimensional spectral data (Atzberger et al., 2015, Pullanagari et al., 2011). However, despite their high accuracy, empirical-statistical models have limitations. They are typically calibrated to a specific set of conditions, such as environmental factors, and their performance may be limited to those conditions (Wang et al., 2020). For example, an empirical-statistical model that performs well in a specific region or vegetation type may not generalize to other regions

or vegetation types, requiring recalibration of the model. Additionally, empirical-statistical methods do not provide a physical understanding of the underlying processes that govern the observed spectral signals, making it difficult to extrapolate results beyond the range of data used for calibration.

In contrast, physically-based models are grounded in the physical principles of light-vegetation interactions and provide a better understanding of the underlying mechanisms that govern the observed spectral signals (Jacquemoud and Baret, 1990, Jacquemoud et al., 2009a,b). Physically-based models have the potential to provide more accurate and detailed information, including canopy structure, leaf biochemistry, and plant water content (Féret et al., 2021, Feret et al., 2008). Physically-based models have been widely used for estimating vegetation biochemical and biophysical properties, including N content (Li et al., 2018b). These models have also been used to simulate vegetation dynamics and carbon sequestration (Li et al., 2019, Berger et al., 2018a). In recent years, a hybrid approach combining the strengths of empirical-statistical and physically-based models has emerged. These data-driven models are based on the physical principles of light-vegetation interactions. However, they also utilize machine learning algorithms to capture complex, nonlinear patterns in the spectral data (Berger et al., 2018a). Data-driven physically-based models can provide more accurate and transferable vegetation biochemical and biophysical property predictions, making them well-suited for large-scale applications.

Empirical-statistical methods face significant challenges when estimating grassland N content due to the collinearity and redundancy of some spectral features. Additionally, many spectral bands captured using HS data can limit the accuracy of N predictions due to the curse of dimensionality (Kuo and Sloan, 2005). Increasing the number of ground truth samples can address this issue, but this approach may only sometimes be feasible due to the limited availability of large samples. Therefore, alternative approaches have been developed to address this issue while decreasing the number of samples. Dimension reduction techniques such as Principle Component Analysis (PCA) and feature selection methods such as Recursive Feature Elimination (RFE) have been used to reduce the dimensionality of the data and improve the accuracy of N estimation models (Sankey et al., 2008, Amirruddin et al., 2020, Abdi and Williams, 2010, Pullanagari et al., 2018). Dimension reduction techniques eliminate redundant spectral bands, thereby reducing data dimensionality, which can improve the accuracy of N estimation models. The PCA, for instance, is a widely used dimension reduction technique that can transform original spectral data into a smaller set of uncorrelated variables (i.e., principal components), reducing data redundancy and improving model accuracy. On the other hand, feature selection techniques select a subset of spectral bands that are most relevant for estimating N. The RFE, for example, selects spectral bands based on their importance in N estimation and iteratively removes the minor significant bands until the optimal set of spectral bands is achieved.

Using dimension reduction and feature selection techniques can significantly improve the accuracy of N estimation models. By eliminating redundant and irrelevant spectral bands, these

techniques enable the development of more efficient and accurate empirical-statistical models, which are crucial for monitoring and managing grasslands sustainably. Friedl et al. (1994) demonstrated the effectiveness of empirical linear regression models for estimating biophysical parameters of the tallgrass prairie using Landsat data. The study found that incorporating terrain variables such as slope angle and elevation significantly improved the accuracy of regression models for estimating biomass. Similarly, Davidson et al. (2006) used the ordinary least squares method to evaluate the effectiveness of various vegetation indices in estimating vegetation water content. The study revealed that SWIR-derived vegetation indices were effective indicators of vegetation water content in grassland and shrubland areas, where the critical water absorption bands are present (Gao, 1996).

Empirical-statistical models have been used in various studies to estimate grassland vegetation properties. Psomas et al. (2011) used proximal HS data and a multiple linear regression algorithm with narrowband NDVI-type vegetation indices to estimate aboveground biomass in grasslands. Similarly, Ullah et al. (2012) predicted grassland biomass and N using five vegetation indices and band-depth analysis based on the multiple linear regression method using the Medium Resolution Imaging Spectrometer (MERIS) data. Jin et al. (2014) developed regression models to estimate biomass in a temperate grassland by combining field sampling with five vegetation indices from MODIS reflectance data. The research indicates a strong correlation between biomass and vegetation indices from optical remote sensing data, particularly those derived from the NIR region. However, empirical-statistical models have limitations regarding generalization to different datasets and environmental conditions over which they were calibrated. Therefore, developing transferable models that can perform consistently over different spatial and temporal scales is a crucial challenge in the field.

Sakowska et al. (2016) investigated the potential use of Sentinel-2 for estimating the seasonal changes in grassland biophysical parameters based on three regression methods (Sakowska et al., 2016). Their study found that using empirical-statistical methods with narrowband indices derived from Sentinel-2 data performed well for estimating canopy chlorophyll content. They also noted the substantial contribution of SWIR to explain the variability of these parameters. In addition to empirical-statistical methods, machine and deep learning methods have gained popularity for estimating vegetation characteristics using remote sensing data. These methods are advantageous over conventional methods as they can establish highly nonlinear relationships between the vegetation variables or reflectance data and response variables. For example, Pullanagari et al. (2016) compared four machine-learning methods to predict the nutrient content of mixed grassland using airborne HS imagery. In their study, the RFR algorithm yielded the highest accuracy for most nutrients, including N. Similarly, Lu and He (2019b,a) evaluated four different methods for estimating grassland chlorophyll content using HS imagery and found that the RFR method performed the best. Using empirical-statistical methods, Tong and He (2017) evaluated many vegetation indices for estimating chlorophyll content in mixed grassland. Their results indicated that narrowband indices from the visible spectrum, particularly

red-edge, performed better than broadband vegetation indices for estimating chlorophyll content. The study highlights the importance of using suitable vegetation indices and the potential of empirical-statistical methods for estimating vegetation properties.

2.9.2 Physically-based and hybrid methods

Physically-based RTM have gained popularity for estimating vegetation biophysical and biochemical parameters (Atzberger et al., 2013, Gastellu-Etchegorry et al., 2004, Estévez et al., 2021, Pu et al., 2020). RTMs simulate the physical interaction of incoming solar radiation with a vegetative canopy and can describe canopy reflectance as a function of viewing and illumination geometry, vegetation biophysical and biochemical parameters, canopy, atmosphere, and soil background characteristics. RTMs can predict leaf and canopy properties by simulating the observed reflectance spectrum as a function of vegetation biophysical and biochemical parameters. Model inversion processes can retrieve vegetation parameters using RTMs (Yang et al., 2020a, Darvishzadeh et al., 2011, Atzberger et al., 2015, Verhoef, 1984). Physically-based models can provide accurate estimates of vegetation biophysical and biochemical parameters, including N content, by considering the physical properties and processes of the vegetation. Two key RTMs are used to estimate vegetation characteristics: leaf-based and canopy-based models. These models depend on accurate input data, such as spectral information, atmospheric properties, and knowledge of vegetation structure and composition.

Despite their accuracy, physically-based models can be computationally intensive and require extensive parameterization, which can be challenging to obtain for field applications. Additionally, physically-based models require detailed information about vegetation and the surrounding environment to generate accurate results. Field measurements of canopy structure, composition, and other parameters are necessary for accurate modeling results. Nevertheless, physically-based models are increasingly used to estimate vegetation properties and are expected to play a more significant role in future research. Additionally, RTMs provide a means of investigating and understanding the cause-and-effect relationships between biophysical and biochemical characteristics and response variables such as vegetation N. However, the accuracy achieved using RTM is limited by our current understanding of the biophysical and biochemical processes that affect radiation's scattering, absorption, and transmission when interacting with a vegetative canopy. Therefore, ongoing research is needed to improve the accuracy of RTM and to develop a better understanding of these processes. Two key RTMs are used for estimating vegetation characteristics: leaf and canopy scale RTMs.

Leaf Radiative Transfer Models (RTM)

Our understanding of how photons interact with leaf structure and biochemical parameters is still limited, but several methods are commonly used to describe the scattering of radiation in leaves (Jacquemoud and Baret, 1990). These methods are based on five critical models used for simulating the interaction between leaves and incoming radiation:

1. The N-flux model is a leaf optical model that utilizes the Kubelka-Munk theory to simulate the interaction between light and leaves, considering them as slabs of diffusing and absorbing material. However, a critical limitation of the N-flux model is its inability to accurately identify absorption coefficients and link them to leaf biochemical parameters, as highlighted by previous studies (Richter and Fukshansky, 1996, Fukshansky et al., 1991).

2. The plate model is a commonly used architecture for retrieving leaf biochemical parameters. This model considers a leaf as a stream of plates with empty spaces between them and is based on the Kubelka-Munk theory and Lambertian scattering across the leaf. This model is suitable for retrieving leaf biochemical parameters similar to the N-flux model. However, it is unable to retrieve small absorption features as reported by Fourty et al. (1996) and Jacquemoud and Baret (1990).

3. Ray tracing models utilize a sampling simulation strategy to accurately represent the internal structure of leaves (Govaerts and Verstraete, 1998). These models can simulate the interaction of radiation with a surface but require detailed information about the interacting surface, such as its morphology. While ray tracing models can provide a realistic overview of leaf structure, they are computationally expensive and may require significant computational resources to run simulations accurately.

4. Radiosity models use a method of calculation that describes the amount of energy reflected from a surface, given a specific surface area (Liang, 2003). The radiosity model represents a leaf as a set of discrete reflecting and transmitting components with unique parameters, accurately representing the leaf's interaction with photons. However, this model is computationally expensive, requiring substantial computing power and time to execute (Baranoski and Rokne, 1997).

5. Stochastic models rely on statistical Markov Chain simulations to estimate the transition probabilities between different radiation states, which can provide valuable information about how radiation interacts with a surface and the likelihood of transitions between different states. An example of a stochastic model that can simulate radiation interactions with vegetation is the Stochastic Light Optimization Procedure (SLOP) model introduced by Maier et al. (1999).

Canopy Radiative Transfer Models (RTM)

Canopy-level RTMs are commonly used with leaf-level models to estimate vegetation characteristics such as biomass, chlorophyll, and N. Different canopy RTMs exist but most common models assume the vegetation canopy as a turbid medium. In a turbid medium, the canopy

leaves are considered minor and randomly distributed, acting as absorbing and scattering elements with no physical size. This simplification of the canopy allows RTMs to be computationally efficient and capable of handling large-scale datasets. Verhoef (1984) proposed a one-dimensional model for approximating canopy characteristics, which assumes a horizontally homogeneous and infinite canopy with vertical variability and finiteness. With this model, it is possible to describe the radiance transfer through a vegetation canopy (Verhoef, 1984). This model has been widely used to estimate vegetation biophysical parameters, such as LAI, through inversion of the model. However, this approach has limitations when addressing heterogeneous canopies, such as those found in orchards or forests, where the canopy is vertically stratified and non-uniform. In such cases, more complex RTMs considering the canopy structure and spatial heterogeneity are necessary to accurately estimate vegetation parameters (Gastellu-Etchegorry et al., 2004, Estévez et al., 2021).

Despite developing more complex canopy RTMs, current models still have limitations in accurately representing the vegetation canopy, particularly assuming that the canopy is a turbid medium. This is due to the vegetation canopy's complex physical structure and biochemical properties, which can result in non-Lambertian surfaces, multiple scattering, and significant energy interactions between the leaves and the surrounding environment. These limitations can result in inaccuracies in the estimated vegetation parameters, particularly for densely vegetated areas, and further research is necessary to address these issues. The following are the key models for modeling canopy RTM:

1. Geometrical item models are another type of RTM that assumes that the surface of a vegetation canopy consists of opaque geometrical objects with predetermined spatial distributions and optical characteristics. In these models, the vegetation canopy is represented by discrete elements, such as needles, branches, or leaves, arranged in an organized pattern to form the canopy. These elements' geometrical shapes and optical properties are defined beforehand, and their interactions with incoming radiation are simulated. Geometrical item models are often combined with the turbid medium assumption, making them well-suited for representing forests and orchards. One commonly used geometrical item model is the ray tracing model, which traces individual light rays through a canopy and calculates their interactions with each canopy element. This approach provides a more detailed description of radiation transfer within a canopy than the turbid medium assumption, but it can be computationally expensive. In addition to the ray tracing model, several 3D hybrid RTMs, combining elements of leaf and canopy-level RTMs, are used in research. The Discrete Anisotropic Radiative Transfer (DART) model is an example of a 3D hybrid RTM that addresses the limitations of the turbid medium assumption. The DART model simulates the radiation transfer through a canopy by tracing individual rays of light through the canopy, considering the geometry of the individual canopy elements and the leaves' radiation scattering and absorption properties. The DART model effectively simulates the light environment within a forest canopy and estimates the amount of radiation absorbed by individual leaves (Gastellu-Etchegorry et al., 2004). Overall, geometric item models and 3D hybrid RTMs provide a more detailed description of radiation transfer

through a vegetation canopy than traditional canopy-level RTMs. However, these models can be computationally expensive and require detailed information about the vegetation canopy and its optical properties.

2. Monte Carlo Ray Tracing (MCRT) models simulate the interactions between photons and matter and are widely used to study light transport through vegetation canopies. MCRT models are based on the principle of probability and statistical sampling and can provide a realistic representation of how radiation interacts with heterogeneous 3D canopies. These models can also account for the statistical nature of radiative transfer behavior. MCRT models are particularly effective in modeling complex canopy structures, including forests and urban vegetation (Govaerts and Verstraete, 1998). Despite their advantages, MCRT models require a high level of detail to describe radiation transfer, which makes them computationally intensive. MCRT models must simulate many photons to obtain a statistically meaningful result, which can be computationally expensive. Moreover, they require a detailed description of the vegetation structure, including leaf size, orientation, and distribution, which can be challenging to obtain from field measurements. Despite the computational challenges, MCRT models provide a highly accurate representation of the interaction of radiation with complex canopies. They have been widely used in various applications, such as forest remote sensing (Govaerts and Verstraete, 1998, Gastellu-Etchegorry et al., 2016).

Radiative Transfer Model (RTM) Inversion

Inverting RTM consists of finding a set of input variables that best match the simulated and actual reflectance observed by a sensor. There are four standard methods of inverting RTM parameters:

1. Iterative optimization techniques are widely used in remote sensing applications for estimating vegetation biophysical and biochemical properties (Vohland and Mader, 2007, de Wit, 1999, Lavergne et al., 2007). This method determines the best spectral fit between simulated and actual reflectance a sensor captures. This process involves an iterative procedure that minimizes a loss function such as a Mean Squared Error (MSE). The iterative procedure continues until the desired accuracy is achieved or until a stopping criterion is met. The goal is to find the best possible solution that minimizes the difference between the observed reflectance and the reflectance calculated by the model. One commonly used approach is based on the maximum likelihood estimator, which determines the likelihood that a particular set of parameters is responsible for the observed reflectance. To maximize this likelihood, the iterative procedure continues to adjust the parameters, often the vegetation's biophysical and biochemical properties. This method can also provide an uncertainty estimation of the solution, which is valuable information for decision-making. Various mathematical and statistical approaches can be used to minimize the loss function in the iterative optimization process. For example, the Levenberg-Marquardt algorithm is a popular optimization method that can handle non-linear problems (Ranganathan,

2004). The Gauss-Newton algorithm is another commonly used optimization method that is faster than the Levenberg-Marquardt algorithm but can be more sensitive to initial conditions (Wang, 2012). Other optimization methods include the conjugate gradient, quasi-Newton, and genetic algorithms. Iterative optimization techniques have been used in various applications. For instance, Vohland and Mader (2007) used iterative optimization techniques to retrieve vegetation water content from hyperspectral data. Dechant et al. (2017) used these methods to estimate vegetation biomass from Landsat data. Using MODIS data, Lavergne et al. (2007) applied iterative optimization techniques to estimate chlorophyll content in a boreal forest.

2. Look Up Table (LUT) techniques involve generating an LUT using an RTM with different combinations of input parameter values. A loss function searches the LUT for spectra that closely match the observed spectra. The method often uses the average or median of several best-fitting spectra to find a solution (Weiss et al., 2000). One of the key advantages of this method is its computational efficiency in retrieving vegetation biochemical and biophysical parameters. However, it has some limitations, such as the need to generate new LUTs based on new situations and its underperformance when the number of input variables becomes very high. LUT techniques have been widely used in various remote sensing applications, such as estimating chlorophyll content and LAI. Additionally, LUT-based methods have been applied to estimate vegetation biophysical and biochemical properties from different types of sensors, including hyperspectral and multispectral sensors. Despite their limitations, LUT techniques remain a valuable tool for retrieving vegetation properties, especially in scenarios where computational resources are limited.

3. Hybrid techniques combine RTMs and machine learning (ML) algorithms. Machine learning is a subset of artificial intelligence that uses data to train models for various applications (Lary et al., 2016). The use of machine learning algorithms to invert RTM parameters has become popular due to its computational efficiency (Liang, 2003). Hybrid techniques involve using an RTM to generate a large dataset representing different vegetation types, which is then used alongside a machine learning algorithm to create a model for predicting vegetation biochemical and biophysical parameters. One of the significant advantages of hybrid techniques is their ability to handle large datasets and retrieve vegetation properties efficiently. Using machine learning algorithms also allows for integrating different data types, such as remote sensing and ground data. This method can also handle non-linear relationships between variables, a significant limitation of other modeling approaches. However, hybrid techniques also have some limitations. For example, the model's accuracy depends on the training dataset's quality and representativeness. Furthermore, the method requires a large amount of input data, which may be challenging in some situations. Despite these limitations, hybrid techniques have shown promise in estimating vegetation characteristics accurately. For example, recent studies have shown that hybrid models can accurately estimate vegetation biophysical and biochemical parameters, such as LAI and chlorophyll content. These studies demonstrate the potential of hybrid models for improving the accuracy of vegetation characterization using RTM

parameters.

4. Emulator techniques have become popular in recent years as they address some of the traditional limitations of RTM inversion processes, such as high computational complexity. RTM emulators are surrogate models that approximate the RTM with high accuracy, which enables computationally efficient and accurate simulations of vegetation biochemical and biophysical parameters. A training set of RTM input-output pairs is used to create an emulator to train a surrogate model that can approximate the RTM. Once the emulator is trained, it can be used to predict the output values for new sets of input parameters. One of the main advantages of using an emulator is that it can drastically reduce the computational time required to perform RTM simulations. This approach is beneficial when the RTM is computationally expensive, or the input parameter space is high-dimensional. For instance, Gómez-Dans et al. (2016) used an emulator technique to model the variation of simulated reflectance spectra due to changing leaf water content in a corn crop. The results showed that the emulator produced accurate predictions while reducing the computational time by several orders of magnitude compared to a full RTM simulation. Emulators have been used in various applications, including crop yield forecasting, biomass estimation, and vegetation classification. The emulator approach also enables the propagation of uncertainty estimates in the predicted values, making it a valuable tool for decision-making. However, one of the main challenges with this approach is that the emulator's accuracy could be improved by the quality and quantity of the training data used to train the model.

2.9.3 Limitations of Radiative Transfer Models (RTM)

RTM is a powerful technique for estimating vegetation biophysical and biochemical parameters, but it is known to be a problem that could be solved. The ill-posed nature of RTM arises from the numerous structural and biophysical parameters that need to be accurately determined in the inversion process. This presents a challenge as uncertainties in the input data can propagate errors and result in incorrect estimation of the vegetation parameters (Atzberger and Richter, 2012, Li et al., 2018a). Several methods have been developed to address these uncertainties. One such method is the combination of single variables to form synthetic variables. In this approach, the ill-posed problem is converted into a well-posed one by combining multiple variables that affect the same biophysical or biochemical parameter. This reduces the number of unknowns in the inversion process, leading to more accurate estimations of vegetation parameters. For instance, Darvishzadeh et al. (2008b) used a synthetic variable approach to estimate the LAI from hyperspectral data. Another approach is to incorporate prior knowledge about model parameters, which can be used to constrain the inversion process. Prior knowledge can come from various sources, such as field measurements or literature values, and can be incorporated into the inversion process as a constraint. For example, Dorigo et al. (2007) used prior knowledge about soil moisture to improve the accuracy of soil moisture retrieval from passive microwave remote sensing data.

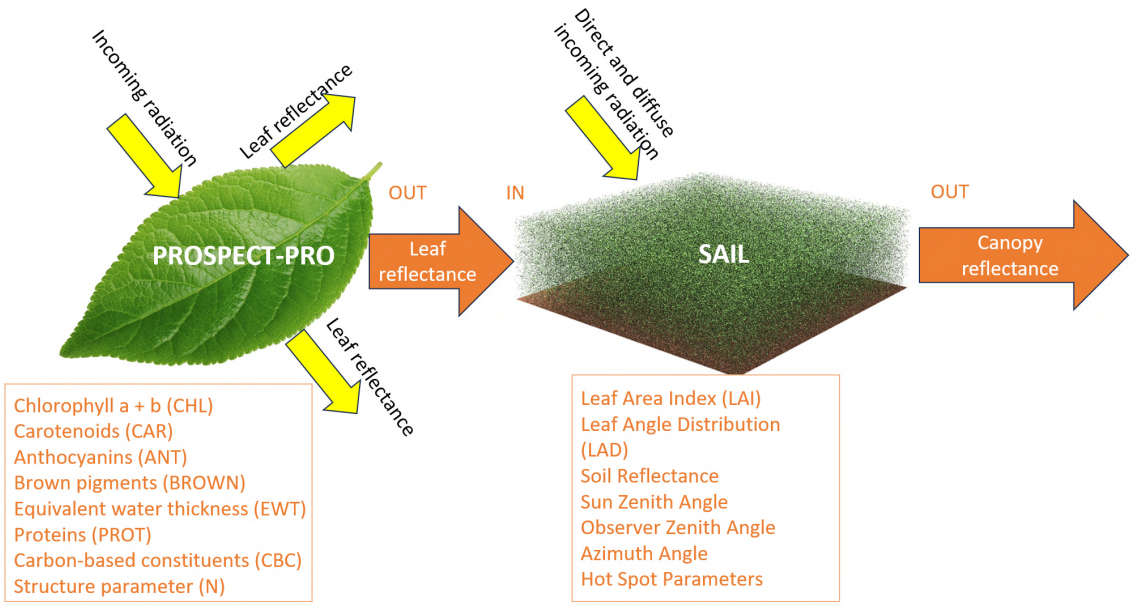


Figure 2.6: Figure depicting the architecture of PROSAIL, which couples the leaf RTM PROSPECT-PRO and the canopy RTM SAIL.

2.9.4 Radiative Transfer Models (RTM) Used in Literature

Physically-based algorithms have been developed to retrieve vegetation characteristics using RTM that directly link vegetation properties such as chlorophyll content and LAI with the canopy reflectance measured by optical sensors. These algorithms are based on radiative transfer theory; therefore, retrieving vegetation characteristics can be performed using spectral reflectance data. For instance, Darvishzadeh et al. (2008a) used the PROSAIL (PROSPECT+SAIL) leaf-canopy RTM in a LUT-based method to estimate LAI and chlorophyll content of mixed grassland (Figure 2.6). They developed a LUT based on a PROSAIL simulation with various combinations of LAI, chlorophyll content, and soil reflectance. The LUT was used to search for the best match between the observed reflectance and the simulated reflectance of the PROSAIL model. Their study found that the physically-based method accurately predicted canopy chlorophyll content. Similarly, other physically-based algorithms, such as the Soil-Canopy Observation of Photosynthesis and Energy (SCOPE) model, have been used to retrieve vegetation characteristics such as LAI, chlorophyll, and vegetation water. These models simulate the physical interaction of light with vegetation and account for the absorption and scattering of light by vegetation components, atmospheric effects, and soil reflectance (Yang et al., 2017, Verrelst et al., 2015b). The retrieval process is based on optimizing the difference between observed and modeled reflectance spectra, with model parameters such as LAI and chlorophyll content used as the input variables.

RTM is a powerful technique for estimating vegetation biophysical and biochemical variables. However, the ill-posed nature of RTM arises from the numerous structural and biophysical parameters that need to be accurately determined in the inversion process. This presents

a challenge as uncertainties in the input data can propagate errors and result in incorrect estimation of the vegetation parameters (Atzberger and Richter, 2012, Li et al., 2018a). Several methods have been developed to address these uncertainties. One such method is the combination of single variables to form synthetic variables. In this approach, the ill-posed problem is converted into a well-posed one by combining multiple variables that affect the same biophysical or biochemical parameter. This reduces the number of unknowns in the inversion process, leading to more accurate estimations of vegetation parameters.

Physically-based algorithms have been developed to retrieve vegetation characteristics using RTM that directly link vegetation properties such as chlorophyll content and LAI with the canopy reflectance measured by optical sensors. These algorithms are based on radiative transfer theory; therefore, retrieving vegetation characteristics can be performed using spectral reflectance data. For instance, Si et al. (2012) demonstrated the successful estimation of grassland LAI and chlorophyll content using the inversion of the PROSAIL RTM with MERIS reflectance data (Si et al., 2012). Darvishzadeh et al. (2011) compared the Partial Least Squares Regression (PLSR) and LUT approaches for estimating LAI in grassland using airborne hyperspectral imagery and found that LAI could be estimated using PROSAIL with accuracy comparable to the empirical-statistical approaches (Darvishzadeh et al., 2011). Atzberger et al. (2015) assessed two RTM inversion methods and two empirical-statistical modeling approaches for estimating grassland LAI using similar data as Darvishzadeh et al. (2011). They found that estimating LAI using LUT-based RTM yields similar results to the empirical-statistical methods (Atzberger et al., 2015). However, they also discovered that the accuracy and robustness of empirical-statistical models decrease when the calibration size of the model decreases.

Similarly, other physically-based algorithms, such as the Soil-Canopy Observation of Photosynthesis and Energy (SCOPE) model, have been used to retrieve vegetation characteristics such as LAI, chlorophyll, and vegetation water. These models simulate the physical interaction of light with vegetation and account for the absorption and scattering of light by vegetation components, atmospheric effects, and soil reflectance. The retrieval process is based on optimizing the difference between observed and modeled reflectance spectra, with model parameters such as LAI and chlorophyll content used as the input variables. Quan et al. (2017) used the PROSAIL RTM to estimate grassland biomass, demonstrating higher accuracy compared to empirical methods. The study found that PLSR performed similarly to the physical method (Quan et al., 2017). Lu and He et al. (2019) explored three types of methods (empirical and physically-based) for estimating grassland chlorophyll content using hyperspectral imagery. They concluded that PLSR and RFR algorithms performed better than linear regression and RTM methods for estimating chlorophyll content (Lu and He, 2019a,b).

There have been efforts to use RTM for retrieving vegetation N, such as the PROSPECT-PRO model, which allows for the estimation of leaf-level protein content (Jacquemoud and Baret, 1990, Féret et al., 2021, Li et al., 2018b). PROSPECT-PRO is a leaf-level RTM based

on the PROSPECT model developed by Jacquemoud and Baret (1990). This model can simulate the directional-hemispherical reflectance and transmittance of green monocotyledon and dicotyledon species and senescent leaves over the solar spectrum from 400 to 2500 nm. The leaf structure parameter N, which represents the average number of air/cell wall interfaces within the mesophyll, and the leaf biochemical variables are the two types of input variables used by this model, with the latter being modified from the original formulation of the model (Fourty et al., 1996, Jacquemoud et al., 2009a). While it was previously believed that chlorophylls were solely responsible for light absorption in the VIS spectrum, it is now known that carotenoids (including xanthophyll pigments) and anthocyanins may also play a significant role in green leaves.

Berger et al. (2020b) proposed a novel method to estimate crop leaf-level protein content using the 4SAIL canopy-level RTM, calibrated with ground-measured spectral data. Their study revealed that the method could accurately identify protein-rich areas in agricultural fields. Furthermore, recent efforts have been made to develop more sophisticated 3D RTMs that consider the impact of biophysical parameters on reflectance, absorption, and scattering. For instance, the DART model considers the effects of vegetation clumping, which can improve the accuracy of biophysical and biochemical retrievals from remote sensing data (Gastellu-Etchegorry et al., 2017).

Despite advances in RTMs, accurately retrieving vegetation N% remains challenging due to the ill-posed nature of RTMs and the influence of factors such as viewing geometry, sensor geometry, illumination, atmospheric conditions, and terrain factors. Several studies have addressed these limitations. For example, Darvishzadeh et al. (2008a,b) used an iterative optimization approach based on a maximum likelihood estimator to estimate vegetation parameters from hyperspectral data. Additionally, Verrelst et al. (2016a) developed an active learning approach combining an RTM and an iterative optimization algorithm to retrieve N content, showing promising results. Furthermore, Zhao et al. (2012, 2013) investigated the relationships between vegetation N and hyperspectral vegetation indices, which could improve N retrieval accuracy. However, Berger et al. (2020b) found that most RTMs are generated using a limited number of vegetation samples, leading to uncertainties with new datasets and sources. Thus, increasing the sample size and variety of vegetation types used to generate RTMs is essential for improving accuracy.

To enhance the robustness and applicability of RTMs like RAdiation transfer Model Intercomparison (RAMI) for vegetation N% monitoring, several future directions are crucial (Zhou et al., 2023). Firstly, incorporating more diverse datasets will better represent different vegetation types and environmental conditions. Secondly, refining algorithms and integrating machine learning techniques can improve RTMs' accuracy and computational efficiency. Furthermore, addressing current limitations in modeling leaf optical properties, canopy structure, and soil background effects will enhance the precision of vegetation N% monitoring across varied ecosystems. Exploring new methodologies, such as 3D modeling and advanced remote sensing techniques, can further advance the reliability of these models in different environmental conditions.

Physics-Informed Neural Networks (PINNs) extend hybrid approaches by embedding physical laws, such as radiative transfer equations, directly into neural network architectures (Cao and Weng, 2024). By integrating these laws as constraints within the loss function, PINNs enhance predictive accuracy and interpretability by ensuring alignment with established physical principles (Raissi et al., 2019). This capability is especially advantageous when training data are sparse or noisy, as the embedded physical constraints guide the model toward physically plausible solutions, mitigating overfitting and enhancing robustness (Zhang et al., 2024).

In vegetation monitoring, PINNs have emerged as a powerful tool for estimating biophysical and biochemical properties by leveraging radiative transfer models and hyperspectral data. For example, Dehghan-Shoar et al. (2024a) demonstrated how PINNs can model canopy reflectance, effectively capturing complex interactions between light and vegetation. Additionally, PINNs can generate synthetic datasets that adhere to physical constraints, further enriching the training data for hybrid machine learning models (Dehghan-Shoar et al., 2023a). This dual role as both predictive and data-generation tools highlights their versatility within hybrid frameworks.

Despite their promise, PINNs present unique challenges. The computational intensity of solving differential equations during training is a significant barrier, particularly for large-scale applications. However, recent advancements in optimization techniques, such as adaptive learning rates and enhanced numerical solvers, are alleviating these challenges and making PINNs increasingly feasible for operational use (Zhou et al., 2023). Their integration of data-driven methods with mechanistic understanding makes PINNs a promising avenue for addressing the complexities of vegetation monitoring across diverse environmental conditions.

Hybrid approaches, including innovations like PINNs, exemplify the transformative potential of combining empirical-statistical and physically-based methods. These methods enable scalable and accurate estimation of vegetation properties while addressing challenges such as dimensionality and collinearity. The success of hybrid frameworks hinges on the careful calibration and validation of RTM-simulated datasets and the representativeness of training data. While computational demands remain a constraint, the proliferation of cloud computing resources and the increasing availability of hyperspectral and multispectral datasets are rapidly mitigating these barriers (Weiss et al., 2000).

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Mohammad Hossain Dehghan Shoar
Name/title of Primary Supervisor:	Associate Professor Gabor Kereszturi
In which chapter is the manuscript /published work:	3

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

Dehghan-Shoar, M. H., Orsi, A. A., Pullanagari, R. R., Yule, I. J. (2023). A hybrid model to predict nitrogen concentration in heterogeneous grassland using field spectroscopy. Remote Sensing of Environment, 285, 113385.

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:

- The percentage of the manuscript/published work that was contributed by the candidate:

- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	
Date:	03-Sep-2024
Primary Supervisor's Signature:	Gabor Kereszturi Digitally signed by Gabor Kereszturi Date: 2024.09.03 22:05:52 +12'00'
Date:	3-Sep-2024

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

Chapter 3

A hybrid model to predict nitrogen concentration in heterogeneous grassland using field spectroscopy

3.1 Abstract

Field spectroscopy is a rapid and non-destructive tool that estimates vegetation's nitrogen (N) concentration (N%). Empirical and physically-based models are widely used for retrieving N%. However, model transferability to different times and locations and feature redundancy remain the two key challenges of field spectroscopy analysis. Here, we addressed these problems by developing a hybrid method (i.e., a combination of physically-based (PROSAIL) and empirical models) to retrieve N% in grasslands. We used a large spectral dataset with more than 6000 samples collected over eight years (2009-2016) for grassland farms across New Zealand. The hybrid model combines the features derived from PROSAIL inversion and an empirical model and develops a predictive model using a Gaussian Process Regression (GPR) algorithm. The model performance is tested on spatially and temporally independent data and compared with PROSAIL and empirical models. The hybrid model achieves higher performance with an RMSE (%N), R^2 and Mean Prediction Interval Width (MPIW) of 0.27, 0.78 and 0.26 as compared to empirical (RMSE=0.28, R^2=0.77 and MPIW=0.32) and physically-based models (RMSE=0.33, R^2=0.65 and MPIW=0.56). In addition, the hybrid model significantly outperforms the physically-based and empirical models during autumn (RMSE=0.32, R^2=0.78 and MPIW=0.11) and summer (RMSE=0.27, R^2=0.80 and MPIW=0.16) seasons.

3.2 Introduction

Addressing world hunger is a pressing issue that will be faced globally in the coming decades. According to the Food and Agriculture Organization of the United Nations (FAO), the world's population will grow from 7.6 to 10 billion between 2017 and 2067, which creates unprecedented pressure on food production (Bergeret, 2017). Meat and dairy production varies significantly

globally, but the use of pastures or grasslands to produce meat and dairy products is very common (Ramankutty et al., 2008, FAO, 2011).

To maintain high-quality grasslands, synthetic fertilizers are commonly applied at regular intervals. Nitrogen (N) is considered a key macro-nutrient that supports grassland growth and development, and timely application of N fertilizers optimizes grassland productivity (Tilman et al., 1996, Fay et al., 2015, Novoa and Loomis, 1981). However, overuse of N fertilizer is costly and can cause serious environmental harm. Therefore, regular monitoring of N status in grasslands is important in order to optimize fertilizer application while reducing negative environmental effects (Julian et al., 2017).

Conventional methods of monitoring the N status of grasslands are time-consuming and costly and thus limit the scope of the analysis to small area experiments (Nocita et al., 2015). Therefore, the application of novel methodologies enables a better way of monitoring grassland nitrogen concentration (N%). Field spectroscopy or hyperspectral remote sensing allows for a non-destructive, rapid, and cost-effective alternative to conventional methods for monitoring grassland N% (see, e.g. Milton, 1987, Milton et al., 2009, for a review of applications of field spectroscopy). Compared to multispectral systems, spectroscopy data captures detailed spectral measurements relevant to biochemical constituents such as protein, cellulose, lignin, and chlorophyll (Novoa and Loomis, 1981). In order to obtain accurate and robust canopy N% estimations from Spectroscopic data, several modeling approaches have been proposed (Verrelst et al., 2015a, 2019a, Pullanagari et al., 2011). These are based on empirical modeling, mechanistic physically-based simulations, or a combination of both (hereafter called a hybrid approach).

Empirical approaches currently leverage the power of machine learning and deep learning algorithms to deliver high prediction performance and robustness on different datasets. These models are widely applied to retrieve N% due to their ability to recognize complex, nonlinear patterns in high dimensional data (e.g. Pullanagari et al., 2021). However, despite the high accuracy, model performance is limited to the specific set of conditions over which they are trained. In particular, the lack of transferability to different datasets impacts the generalization of the insights obtained by the model to different but similar datasets (Wang et al., 2020). Furthermore, many spectral bands available from field spectroscopy can limit the accuracy of N predictions due to the so-called *curse of dimensionality* (Kuo and Sloan, 2005). One method of addressing this issue is by increasing the number of ground reference samples to ensure statistical confidence (e.g. Thenkabail et al., 2018, Weiss et al., 2000), which ultimately limits their usefulness to large samples. Different approaches have been utilized to address this issue without increasing the number of samples, with varying degrees of success (Wang et al., 2022a, Bossung et al., 2022).

In contrast to empirical approaches, mechanistic methods (e.g. Radiative Transfer Models (RTMs)) are often used for monitoring vegetation characteristics by simulating the physical

interaction of incoming solar radiation with vegetative canopy (Atzberger et al., 2013, Gastellu-Etchegorry et al., 2004, Estévez et al., 2021, Pu et al., 2020, Verhoef, 1984). RTMs can be used to monitor leaf and canopy properties by simulating the observed reflectance spectrum as a function of several biophysical and biochemical properties (Yang et al., 2020a). A process of model inversion enables the retrieval of the combination of variables that result in the best match between the measured and predicted spectrum (Vohland and Mader, 2007, de Wit, 1999, Lavergne et al., 2007).

Since RTMs consider a wide range of variables, they can be transferable to different spatial and temporal scales (Pellissier et al., 2015, Wang et al., 2019b). In addition, RTMs enable us to understand the cause-and-effect relationship between biophysical and biochemical properties and response variables such as N%. However, the accuracy delivered with this approach depends on our current understanding of the biophysical processes that determine the scattering, absorption, and transmission of photons when interacting with a vegetative canopy (Liang, 2003, Berk et al., 2006). This is further limited by model simplifications (e.g., radiation transfer is typically solved in 1D) (Jacquemoud et al., 2009a, Verhoef, 1984). The large number of structural and biophysical variables that need to be constrained lead to uncertainties in the model inversion (the so-called ill-posed nature of mechanistic models) (Atzberger and Richter, 2012, Sun et al., 2022, Li et al., 2018a). Furthermore, the uncertainty of each inverted RTM variable causes propagated errors in the analysis workflow. As a result, identifying and addressing different sources of uncertainty in RTM remains a critical issue. Several methods are developed to address the uncertainty associated with this issue (Rivera et al., 2013). Commonly used methods to reduce the uncertainty associated with the RTM inversion process are by combining single variables into synthetic variables and use of prior knowledge about model variables (Darvishzadeh et al., 2008a, Dorigo et al., 2007).

Hybrid approaches provide an alternative method of monitoring grassland N% by combining the characteristics of empirical and physically-based methods. These approaches have become popular and have produced promising results (Berger et al., 2020a). Key benefits of hybrid approaches include faster computation times for the inversion of RTM variables and the ability to mitigate the natural collinearity and band redundancy associated with Spectroscopic data. A common way of using hybrid methods for monitoring vegetation variables is by generating a large training dataset of simulated biophysical and biochemical properties and their associated spectral signatures (Darvishzadeh et al., 2011, He et al., 2019). These are used alongside machine learning algorithms to find an empirical relationship between the simulated spectra and biophysical and biochemical variables. The generated models are then used on real-world datasets to infer the biophysical and biochemical variables, which are then used to predict grassland N%. Hybrid approaches can improve the performance of their predictions, but they also inherit many of the problems associated with fully empirical or mechanistic models, such as model transferability (Wang et al., 2020).

In this work, we address the various challenges described above by developing empirical,

physically-based, and hybrid models to predict grassland N%. We measure the performance of our models by using independent data collected during different seasons and locations, which encompass heterogeneous land over wide spatial and temporal scales in New Zealand (see Pullanagari et al., 2021, for details of the dataset). Furthermore, we investigated the uncertainty associated with the inverted RTM variables and conducted feature importance analysis to identify the most important RTM variables and spectral bands for monitoring grassland N%.

3.3 Material and Methods

3.3.1 Environmental conditions

Sampling was conducted in 17 dairy and hill country farms across the south and north islands of New Zealand from 2009 to 2016 in different seasons (autumn, winter, spring, and summer). The sampled farms consist of different grazing patterns based on the type of livestock (e.g., sheep, deer, and cattle). Due to optimal temporal conditions, grasslands in New Zealand grow continuously during the year. New Zealand has temperate weather conditions, with the mean annual precipitation between 600 mm and 1600 mm and a mean annual temperature of $10°C$ to $16°C$ *. In addition, due to the diversity of soil types in New Zealand, field campaigns were conducted on farms where there is significant variability in the soil pH (between 4.5 and 7.5), Olsen P (between 2 and 217), and Potassium (between 1 and 74) levels. Detailed descriptions regarding the field campaigns and dataset used in this work are stated by Pullanagari et al. (2021).

3.3.2 Spectral measurements

The canopy spectral measurements were collected on areas covered by grass with an estimated Leaf Area Index (LAI) of one to three to minimize the spectral contribution of the soil background. The samples were collected on heterogeneous grassland and the main grass species are illustrated in Table 3.1.

Grassland canopy spectral measurements were captured using two types of spectro-radiometers with a spectral range of 350 to 2500 nm :

1. ASD Field-Spec 4 Hi-Res (Malvern Panalytical Ltd., Malvern, United Kingdom)
2. ASD FieldSpecPro (Analytical Spectra Devices, Inc., Boulder, CO, USA)

Both sensors have a 512-element silicon photodiode, which provides a resolution of 3 nm and a sampling interval of 1.4 nm in the Visible Near Infra Red Region (VNIR) (<1000 nm). In addition, the sensors also consist of two sets of thermoelectric (TE)-cooled indium gallium arsenide (InGaAs) photodiode arrays, which help provide a spectral resolution of 10–12 nm and 8 nm with a sampling interval of 1 nm in the Short Wave Infra-Red Region (SWIR) (>1000 nm). The canopy spectral measurements were captured using an artificial probe, a Canopy Pasture Probe (CAPP), which consists of an artificial light source (50-W tungsten-quartz-halogen bulb) (Kawamura et al., 2009) (Figure 3.1).

*https://niwa.co.nz/education-and-training/schools/resources/climate/overview

Table 3.1: Table containing the list of primary and secondary grass species in the dataset. While the most dominant species are browntop (*Agrostis capillaris*) and clover (*Trifolium repens*) several secondary grass species are present in less intensive farming areas.

Grass type	Latin name	English name	Description
Primary	*Agrostis capillaris*	Browntop	Dairy and hill country
Primary	*Capillaris*	Clover	Dairy and hill country
Secondary	*Lolium perenne*	Ryegrass	Hill country
Secondary	*Pennisetum clandestinum*	Kikuyu grass	Hill country
Secondary	*Cynosurus cristatus*	Dogstail	Hill country
Secondary	*Dactylis glomerata*	Cocksfoot	Hill country
Secondary	*Phleum pratense*	Timothy	Hill country
Secondary	*Anthoxanthum odoratum*	Xweet vernal	Hill country
Secondary	*Holcus lanatus*	Yorkshire fog	Hill country

Figure 3.1: Collection of grassland spectra using the Canopy Pasture Probe (CAPP) during a field campaign (Kawamura et al., 2009).

Due to variable weather conditions, the light source was placed in an inverted dark plastic bin to maintain consistent illumination throughout the scanning period. The spectral readings were optimized using a white reference panel (Sanches et al., 2009). The sampling sites were located in different geographic locations and selected based on stratified random sampling. Environmental variables such as slope, aspect, and elevation were considered for stratification as they influence grassland growth and performance. The sampling locations were equally distributed over the selected farms to cover various environmental conditions. Slope angle, slope aspect, and soil types were used for our stratification, and we categorized our sampling based on the slope categories of 0–8, 8–16, 16–25, and 25. We used the Digital Elevation Model (DEM) available from Land Information in New Zealand (LINZ) for our sampling design. In addition, the spectral measurements for each farm were collected within two days to minimize uncertainty.

At each sampling site, a quadrant with a size of $0.5\,\mathrm{m}^2$ was placed on the grassland to take

measurements (Hutchinson et al., 2016, Cosgrove et al., 1998, Pullanagari et al., 2012). A total of 20 spectral measurements were captured by rotating the bin and changing the location of the bin within the quadrant. The quadrant size and number of spectral measurements were selected to capture and represent grass heterogeneity within the quadrant. The bin was connected to the spectro-radiometer using a fiber optic cable. The distance between the optic cable and grass samples was 40 cm. Following this, the average spectrum was calculated and pre-processed to be used for the analysis.

3.3.3 Chemical analysis

After collecting each sample's spectral measurements, grass samples were cut using an electric shearing machine and stored in a cooling storage box. All sampling locations were geo-referenced using an RTK-GPS unit, and after collecting the spectral measurements, photos of the samples were collected for a visual assessment. The collected grass samples were sent to Analytical Research Laboratories (ARL), Napier, New Zealand, to determine the N% of each sample using the Kjeldahl method (Lynch and Barbano, 1999). This process involved drying, grinding, and homogenizing each sample using a mixer mill. Finally, the N% of each sample was estimated using a QuikChem® 8500 flow injection analyzer.

3.3.4 Data pre-processing and dataset creation

The spectral regions of 350–450 nm and 2450–2500 nm of the collected spectral data were removed from the analysis due to significant noise in the spectra. The noise in the remaining spectra was minimized by employing a Savitzky–Golay filter with a smoothing window width of 13 and a polynomial order of 3 without calculating the derivatives (Press and Teukolsky, 1990). Finally, the analysis-ready datasets were created using the N% results acquired from the chemical analysis and collected spectral measurements. The construction of the empirical, hybrid, and physically-based methods are discussed in the following subsections.

3.3.5 Modelling approaches

Physically-based model

The physically-based model relies on the inversion of biochemical and biophysical variables (e.g., LAI and Cab) of the grass samples using an iterative inversion procedure. Following this, the inverted physically-based model variables were used alongside an empirical model to predict grassland N%. In this work, we used the well-known physically-based model known as PROSAIL. PROSAIL is composed of two physically-based models known as SAIL, a canopy bidirectional reflectance model developed by Verhoef (1984) and PROSPECT, a leaf optical properties model first proposed by Jacquemoud and Baret (1990). We implemented the latest PROSPECT variant, PROSPECT-PRO (Féret et al., 2021), due to its ability to invert protein content in different vegetation species using the measured spectra. Our method of retrieving

variables using PROSAIL consists of running many forward and inverse simulations while attempting to minimize the difference between the simulated and measured spectra using the sum of squared differences as a loss function. The inversion process involved running 50,000 simulations in parallel using Markov-Chain Monte Carlo (MCMC) with 1000 walkers. We used the Python ensemble sampling toolkit called EMCEE, which implements an Affine Invariant MCMC (Foreman-Mackey et al., 2019, Goodman and Weare, 2010). This particular variant of MCMC was adopted due to its higher performance and efficiency. The variables of the search space used can be seen in Table 4.2.

Table 3.2: The PROSAIL variable search space, which was used alongside Markov-Chain Monte Carlo (MCMC) to undertake the inversion process.

Variable name	Max value	Max value	Abbreviation	Units
Structure variable	1	3	N	
Chlorophyll content	1	100	Cab	g/cm^{-2}
Carotenoid content,	1	30	Car	g/cm^{-2}
Brown pigment content	0	0.2	Cbrown	
Equivalent water thickness	0	0.02	Cw	cm
Dry matter content	0	0.02	Cm	g/cm^{-2}
Protein content	0	0.02	Cp	g/cm^{-2}
Carbon-based constituents	0	0.02	CBC	g/cm^{-2}
Anthocyanins content	1	10	Anth	g/cm^{-2}
Leaf Area Index	1	3	LAI	
Leaf angle distribution a	10	90	LIDFa	Degrees
Leaf angle distribution b	10	90	LIDFb	Degrees
Background reflectance	-	-	RSOIL	
Dry/Wet soil factor	-	-	PSOIL	
Hotspot parameter	0	1	HOTSPOT	
Solar zenith angle	0	0	SZA	Degrees
Observer zenith angle	20	20	VZA	Degrees
Relative azimuth angle	160	160	RAA	Degrees

Following the inversion of the PROSAIL variables, the mean and one-sigma standard deviation of each inverted variable were calculated. We used the mean values of the inverted area-based PROSAIL variables alongside a Gaussian Process Regression (GPR) algorithm with a Matérn covariance function to infer grassland N% (GPy, 2012). The covariance function of the GPR algorithm was selected using a random search optimization algorithm (Bergstra and Bengio, 2012). We validated our model by splitting our large spectral dataset into training and validation sets using a ratio of 80% (n=4800) and 20% (n=1200). As the development of transferable models is very challenging, we ensured the validation set is temporally and spatially independent to measure our model performance accurately (Wang et al., 2020).

In addition, we further validated the model performance using seasonal cross-validation to measure how the models performed during different seasons. This process consisted of creating a model using data captured during three seasons (e.g., autumn, winter, and spring) and validating the generated model using data captured during the remaining independent season

(e.g., summer). This procedure was repeated for the four seasons of spring (n=2940), autumn (n=2005), summer (n=854), and winter (n=208), where data were available. We validated the performance of our models by using statistical metrics including R^2, Root Mean Squared Error (RMSE) (%N), and Mean Prediction Interval Width (MPIW) (Khosravi et al., 2011).

Empirical model

The empirical model takes advantage of the relationship between the spectral reflection captured by the spectro-radiometer to monitor N%. A challenge in this method for monitoring grassland N% is the collinearity and redundancy of some spectral features. To address this issue, we excluded highly correlated spectral bands, bands which have a linear correlation of greater than 0.99%, and undertook feature selection by implementing a genetic algorithm (GA) to identify the most important spectral bands for monitoring N% (Calzolari, 2021, Leardi et al., 1992).

GA is an iterative procedure that selects spectral bands by encoding the inclusion and exclusion of each spectral band. This procedure associates a bit (included and excluded band) with each spectral band and creates chromosomes using different spectral bands. We set the crossover and independent crossover values for each chromosome to set the probability for the spectral bands, which are passed to the next iteration. In addition, the tournament selection parameter was set to determine which bands will continue to the next iteration.

A key limitation of GA is the possibility of finding the local optimal solution rather than identifying the global optimal solution. One way of overcoming this issue is by introducing the possibility of mutations (random addition of variables in each iteration). To do this, we used a set of values for mutation and independent mutation probabilities for each iteration, which introduces new bands in each iteration.

GA requires the use of a surrogate algorithm to undertake feature selection. In this case, a GPR with a Matérn kernel was used. At the end of each generation of the GA, a model was calibrated using the selected spectral bands and the surrogate algorithm. The performance or fitness of the model was then evaluated using an independent test set and 10-fold cross-validation. The GA parameters were chosen based on the overall performance of the model, which is illustrated in Table 3.3.

The selected spectral bands were used to create calibration and validation datasets. Finally, a model was calibrated and validated using a GPR algorithm with a Matérn covariance function to monitor grassland N%, similar to the physically-based model described previously.

Table 3.3: The genetic algorithm (GA) variables, which were used for feature selection.

Genetic algorithm	Variable value
Fitness function	Mean Squared Error (MSE)
Population size	50
Crossover probability	0.1
Mutation probability	0.05
Number of generations	10000
Independent crossover probability	0.5
Independent mutation probability	50
Tournament size	0.5

Hybrid model

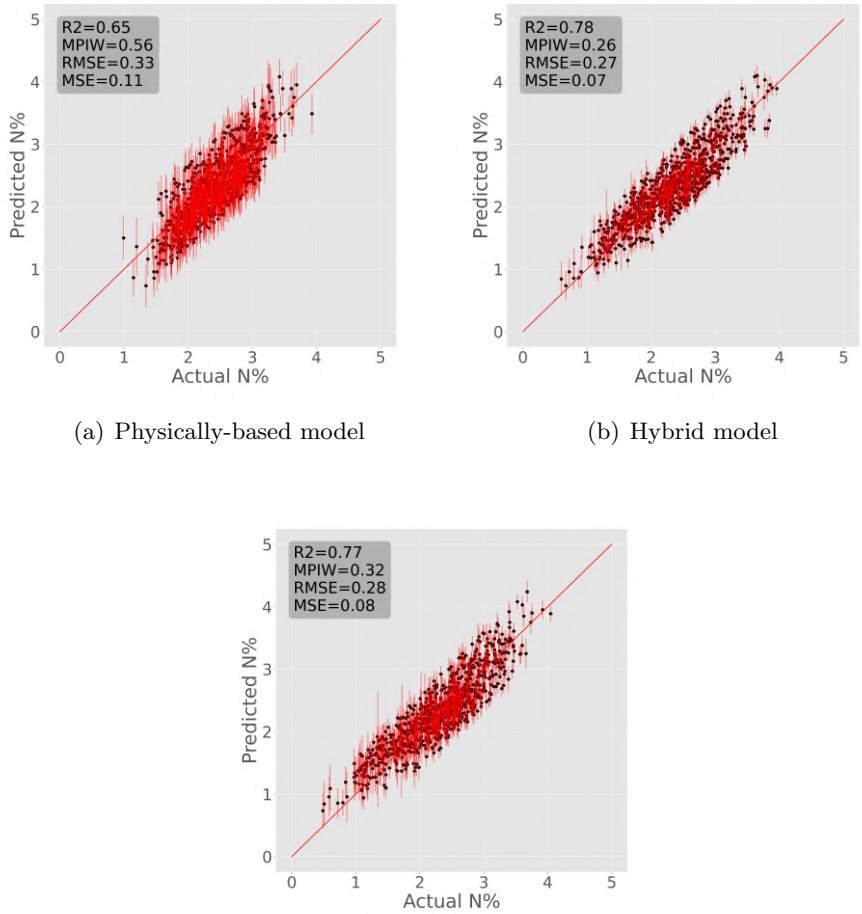
The hybrid model combines the spectral bands and the inverted PROSAIL variables used in the physically-based model to monitor grassland N%. The inverted grassland biochemical and biophysical variables derived from the physically-based model were added as further features to the empirical dataset to combine the empirical and physically-based models. Following this, a GA, as described previously, was implemented and used to undertake feature selection using the hybrid dataset. Hence, the hybrid model receives as input features the list of spectral bands and inverted PROSAIL biophysical and biochemical variables.

Similar to the physically-based and empirical models, a GPR with a Matérn covariance function was built to predict grassland N%. Our hybrid model was validated and evaluated using independent data collected from different locations and times (similar to the other models). In addition, we conducted feature importance analysis by computing Shaply values to investigate the key features that contributed most to the prediction of N%. By using the library, SHAP (SHapley Additive exPlanations) enables the retrieval of Shapley values to help explain the outputs of different machine learning models (Lundberg et al., 2020).

3.4 Results

Here, we present the results for predicting grassland N% using hybrid, empirical, and physically-based models. Figure 3.2 shows the performance of the different grassland N% models, which are validated using data collected from spatio-temporally independent farms during different seasons.

The results demonstrated that the hybrid model outperforms both the physically-based and empirical models. We used a t-test to measure if the difference between the results were significantly different. The difference between the accuracy of the empirical and hybrid models with the physically-based model was significant ($p < 0.05$). However, the difference in accuracy between the empirical and hybrid models was insignificant ($p > 0.05$). We further investigated the performance of the different models using seasonal cross-validation. We found that the



(a) Physically-based model

(b) Hybrid model

(c) Empirical model

Figure 3.2: Predicted nitrogen (N) concentration (N%) was obtained using the different modeling approaches against the actual values. Each panel corresponds to a different modeling approach. Black circles denote the model predictions. The red bars correspond to the standard deviation obtained with the Gaussian Process Regressor (GPR). The legend on each panel shows the value of R^2, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Prediction Interval Width (MPIW) for each model.

hybrid model outperforms the empirical and physically-based model in the autumn and summer seasons with RMSE, R^2 and MPIW (Figure 3.3). Overall, the highest performance for all seasons was achieved by the hybrid model.

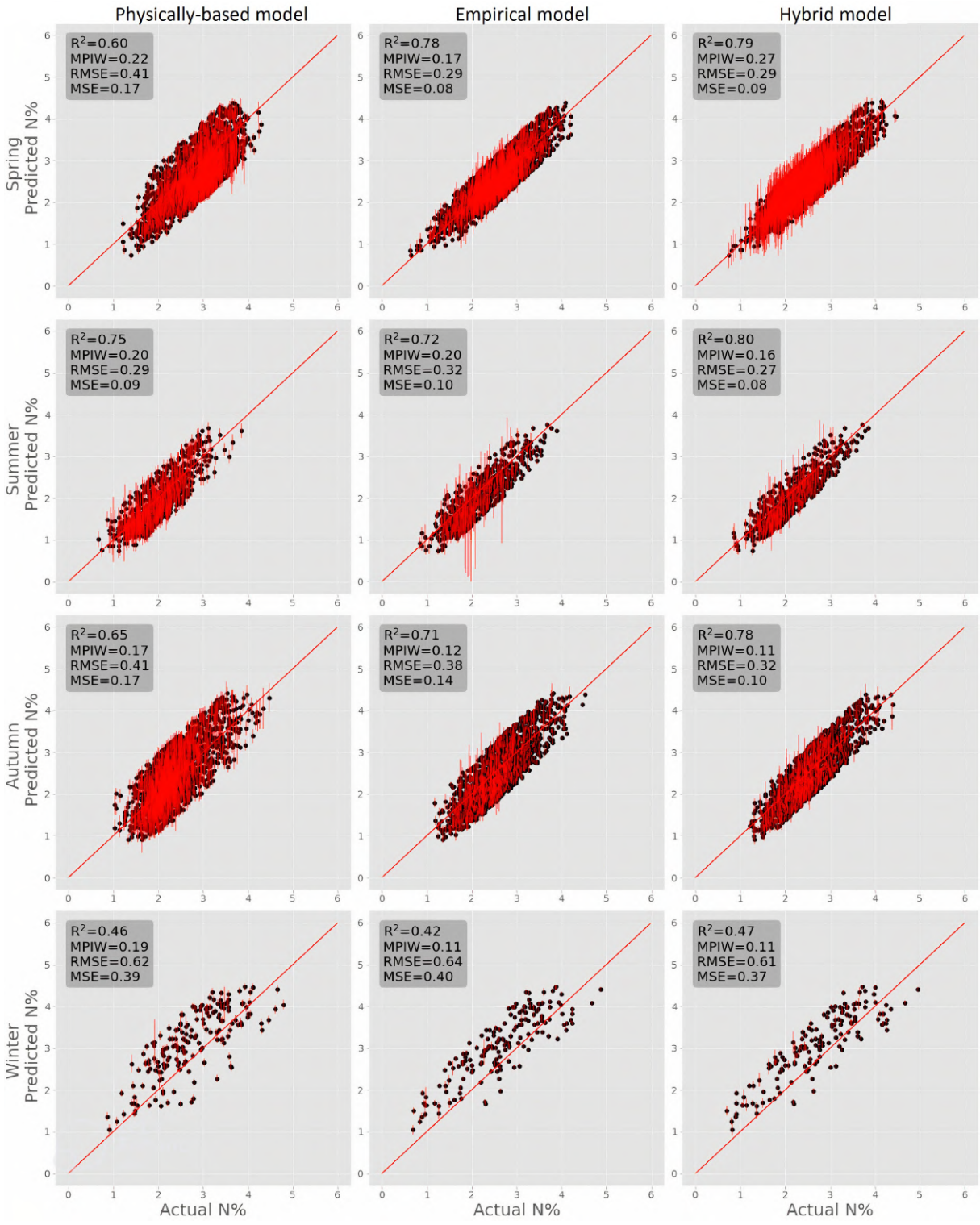


Figure 3.3: The change in performance of the hybrid, physically-based, and empirical models in different seasons. The scatter plots in the first column were generated using the physically-based model, while the plots depicted in the second and third columns correspond to the empirical and hybrid models. In addition, each row corresponds to different seasons. The first row was generated using samples collected during spring, while the plots in the second, third, and fourth rows were generated using samples collected during summer, autumn, and winter.

3.4.1 Spectral features selected

In the empirical model, a total of 65 out of 2101 spectral bands were selected across the spectrum, while in the hybrid model, 52 out of 2101 spectral features were selected. In addition to the spectral features, chlorophyll content (Cab), dry matter content (Cm), water thickness (Cw), LAI, and protein content (Prot) derived from PROSAIL were selected in the hybrid approach. The majority of the spectral bands selected were located in the SWIR of the spectrum, in which the wavelength intervals [2050-2080], [830-906], [1610, 1850] nm were repetitively selected (Figure 3.4). Other key bands were in the NIR region of the spectrum, which are associated with canopy structure characteristics (Mousivand et al., 2014, Verrelst et al., 2019b).

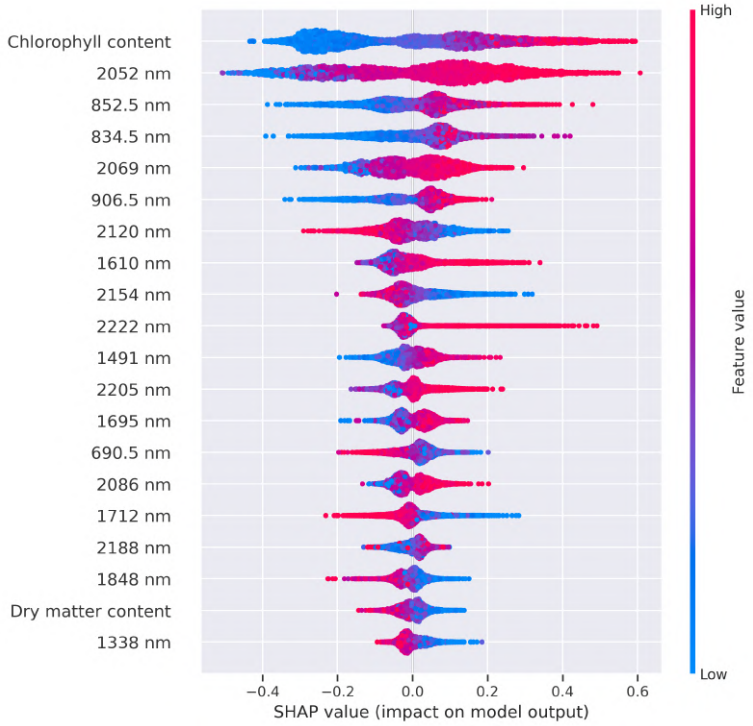


Figure 3.4: Illustration of the most important features of the hybrid dataset. The most important features are presented on the left-hand side of the figure, while the color bar on the right-hand side illustrates the value of each feature. The plot illustrates that a higher chlorophyll content (Cab) corresponds with a higher nitrogen (N) concentration (N%).

3.4.2 Relationship between PROSAIL derived variables and N%

In the following, we characterized the inverted biochemical variables derived from PROSAIL. We computed the pairwise cross-correlation coefficient to quantify the intercorrelation between the inverted variables and N%, as shown in Figure 3.5.

We found a strong linear correlation between grassland equivalent water thickness, chlorophyll, and carotenoid content. There is also a moderate relationship between chlorophyll and protein content, with grassland N%. The most important PROSAIL-derived features that were used for the prediction of grassland N% are found to be chlorophyll, protein and dry matter content, and equivalent water thickness.

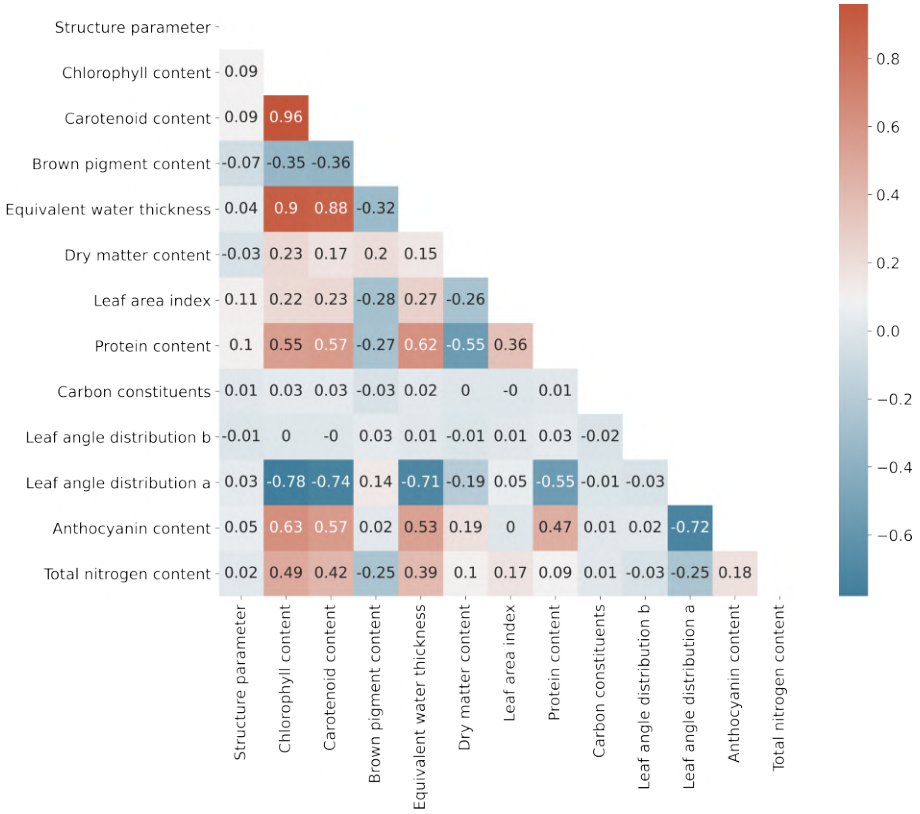


Figure 3.5: Figure illustrating the correlation plot for the inverted PROSAIL variables and nitrogen (N) concentration (N%).

3.4.3 Uncertainty of the inverted PROSAIL variables

An MCMC analysis was undertaken to investigate the posterior distribution of the inverted biophysical and biochemical values of the PROSAIL model. In Figure 3.6, we show the resulting (68-95%) confidence intervals around the best fit. The uncertainty associated with some variables was reasonable (e.g., chlorophyll content) while for others (e.g. carbon constituents) the uncertainty was significantly larger. Figure 3.6 (a) is generated to investigate the impact of the physical variables of the RTM on the retrieval accuracy of the biochemical variables.

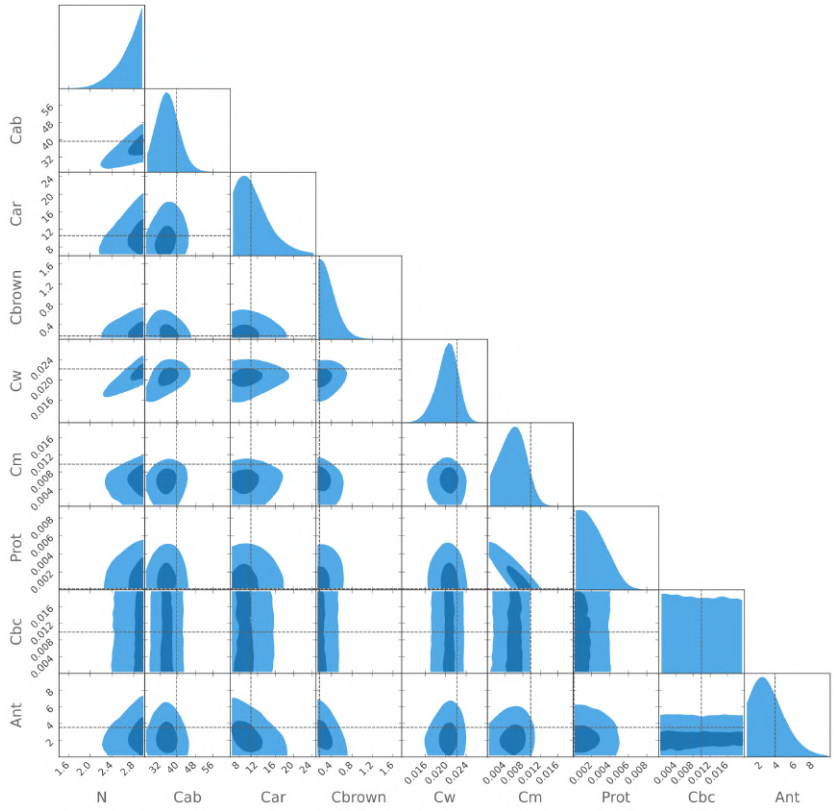
As seen in Figure 3.6 (b), it is very challenging to retrieve PROSAIL variables accurately when insufficient information is known about the collected samples, and it is impossible to constrain some RTM variables. Sample spectra were generated using PROSAIL while fixing the values for LAI and Leaf angle distribution a and b to investigate this. The LAI value was fixed to 3, while the leaf angle distributions a and b were fixed to 45 and 60, respectively. Followed by this, other variables were inverted using MCMC as previously (Figure 3.6 (a)).

Constraining the physical variables of PROSAIL enables the retrieval of the biochemical variables to improve the inversion process. To further validate this, we conducted PROSAIL simulations using MCMC, running 50,000 simulations, with and without fixing LIDFa, LIDFb, and LAI (the fixed values for LIDFa, LIDFb, and LAI are 20, 20, and 2 accordingly). This resulted in the creation of two distributions for Cp and Cab. Using a z-test, we compared the distribution of the inverted Cp and Cab variables with and without fixing the biophysical variables of PROSAIL. The results illustrated that the distribution of the inverted PROSAIL variables for Cp and Cab are significantly different ($p < 0.05$).

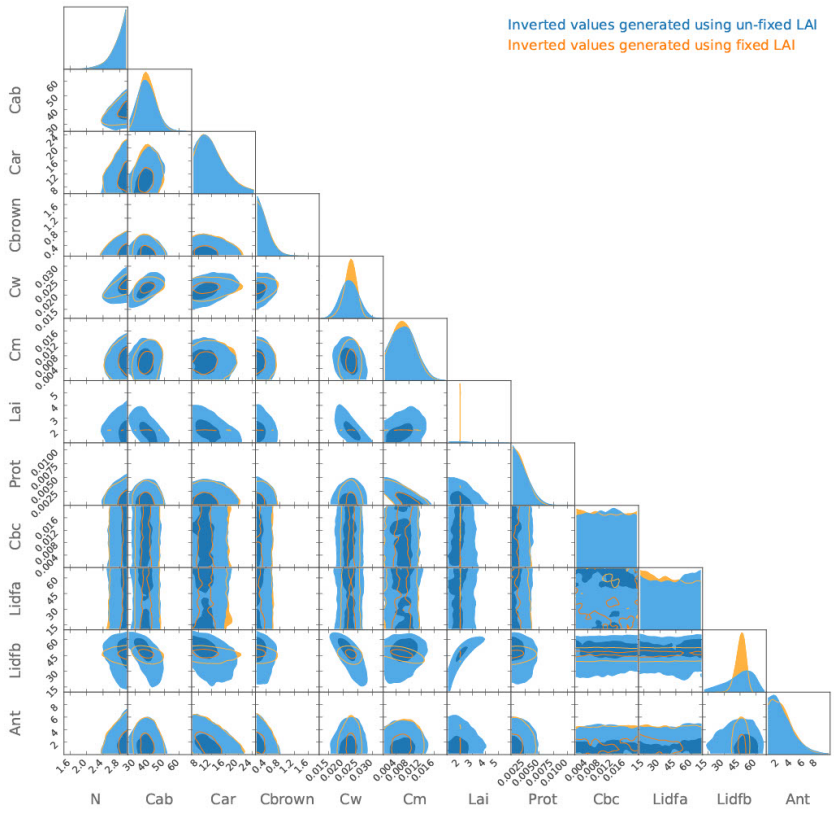
3.4.4 Performance of the different models

Figure 3.4 shows that the most informative features of the hybrid dataset were chlorophyll content and the spectral bands at 2060, 2052, 852.5, and 834.5 nm. Summarizing the selected bands highlights the importance of the SWIR region of the electromagnetic spectrum for predicting grassland N%.

The uncertainty associated with the physically-based model is significantly higher than the hybrid and empirical models (Figure 3.2). We investigated the impact of the training set's size on the different models' performance. The best-performing model is the hybrid model, while the physically-based model yielded the least performance irrespective of the calibration size (Figure 3.7). In addition, the uncertainty of all the models decreased when using a larger calibration dataset. The performance of the models started saturating when the calibration size approached 3500 samples (Figure 3.7). This can be because this calibration size ($n = 3500$) contains the highest spectral variability required for our models to perform optimally. We can also see that the hybrid model performs better with smaller datasets ($n < 500$); however, the performance of the hybrid and empirical models is similar with a larger calibration dataset ($n > 800$). The physically-based model has a poorer performance than the hybrid and empirical models, but it has much lower overall uncertainty.



(a) Constrained model



(b) Unconstrained model

Figure 3.6: Corner plots demonstrating the marginalized results of the inverted PROSAIL variables. The different shades of blue correspond to each inverted variable's uncertainty (sigma) levels. Plot (b) illustrates the margins associated with retrieved PROSAIL variables with and without fixing the Leaf Area Index (LAI). We investigated the impact of the biophysical variables of the Radiative Transfer Model (RTM) using simulated data (plot (a)). As seen in plot (a) by constraining the physical variables of the RTM the uncertainty associated with inverted biochemical variables decreases significantly.

3.5 Discussion

3.5.1 The performance of empirical, physically-based and hybrid models

In this study, we used a hybrid approach to improve the retrieval accuracy and transferability of grassland N% from field spectroscopy data and compared the performance with empirical and physically-based methods. Overall, the hybrid model outperformed the empirical and physically-based models with improved accuracy on a spatially and temporally independent dataset (Figure 3.2). Féret et al. (2021) indicated scaling from leaf to canopy scale, reflectance anisotropy, and canopy structure influence the retrieval accuracy. Despite canopy scale effects, this study has improved the retrieval accuracy. However, detailed understanding and modeling are required to measure the influence of canopy scale factors on N% estimation. In addition to the accuracy, the hybrid approach reduced the uncertainty levels associated with each predicted sample, which indicates consistent performance under different environmental conditions. The hybrid model combined the contribution from multiple spectral regions, which are actively linked to fundamental N absorptions and RTM-derived variables (Curran, 1989, Fourty et al., 1996).

Like other studies, RTM inversion makes it challenging to accurately infer grassland's biophysical and biochemical variables when many unknown variables exist in the RTM search space. It is very challenging to accurately invert protein content without constraining physical variables of the RTM, which was also highlighted by Féret et al. (2021) (Figure 3.6). This is because the physical variables of the RTM significantly impact the spectral regions where the protein absorption bands are present. Although we constrained the search space for sensor, solar geometry, and LAI, the results were not as accurate as the hybrid approach.

Another biochemical component derived from RTM, commonly used for monitoring grassland N%, is chlorophyll a+b or Cab. The spectra associated with Cab are found in the visible region of the electromagnetic spectrum. As seen in Figure 3.8, the absorption of Cab is very strong in the visible region of the spectrum, which will enable its accurate retrieval using RTM.

However, despite the common use of Cab for monitoring N%, fundamentally N containing biochemical constituents are located in different parts of the electromagnetic spectrum, and only a small portion of N is present in chlorophyll pigments (Curran, 1989, Fourty et al., 1996). Moreover, grassland physiology and the natural change in chlorophyll content in different seasons may not be a suitable proxy for monitoring N% across different seasons and locations (Darawsheh and Bouranis, 2006). This study indicates that multiple absorption features, including chlorophyll absorption, must accurately estimate N%. As a result, the hybrid model performs better during the summer and autumn seasons, with better RMSE and uncertainty values (Figure 3.3). Therefore, by combining the spectral features and PROSAIL-derived variables, the hybrid model can accurately incorporate the absorption features required for monitoring N% during different seasons.

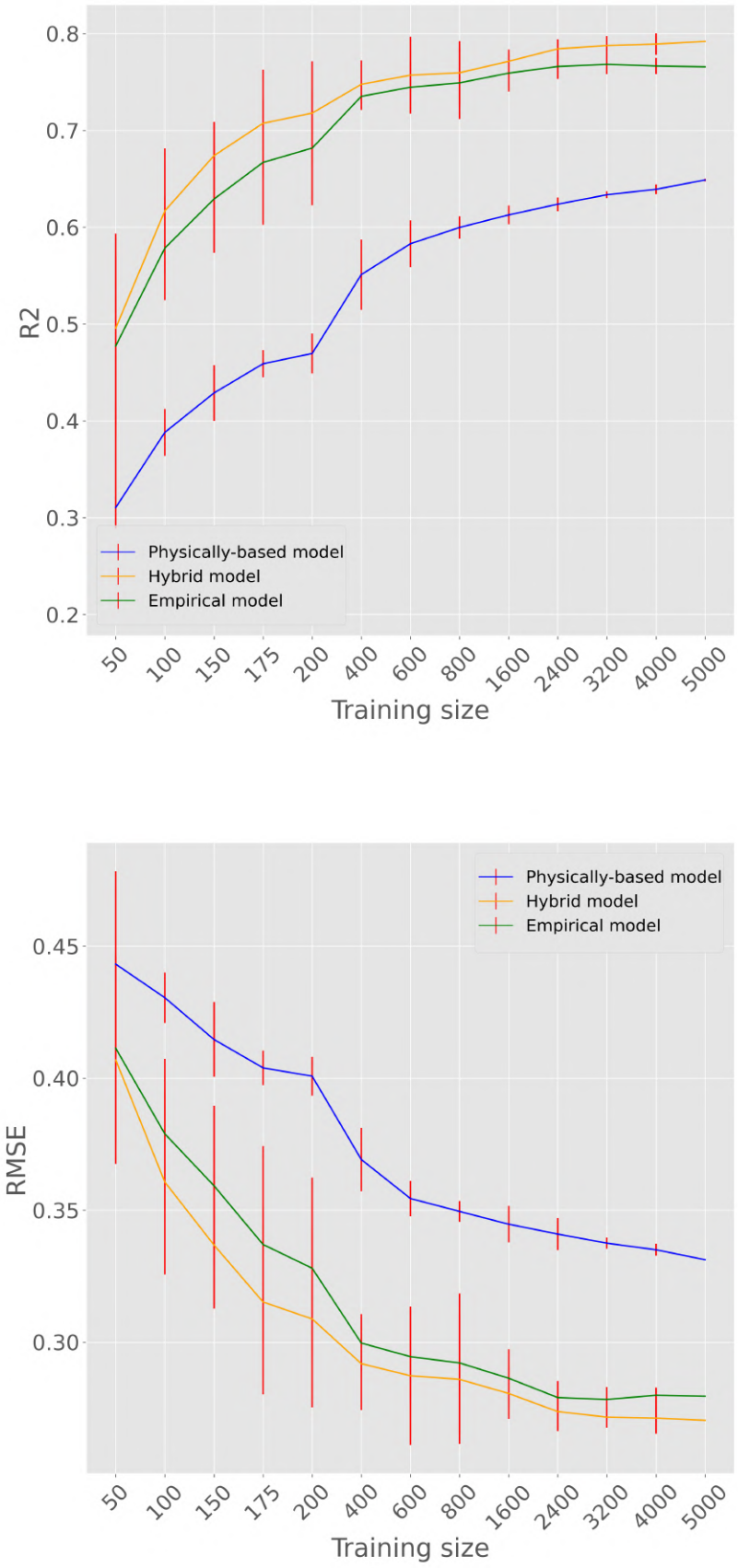


Figure 3.7: Figure illustrating the change in the performance of the hybrid, empirical, and physically-based models based on the size of the training dataset. The vertical red lines correspond to the uncertainty associated with predicted values.

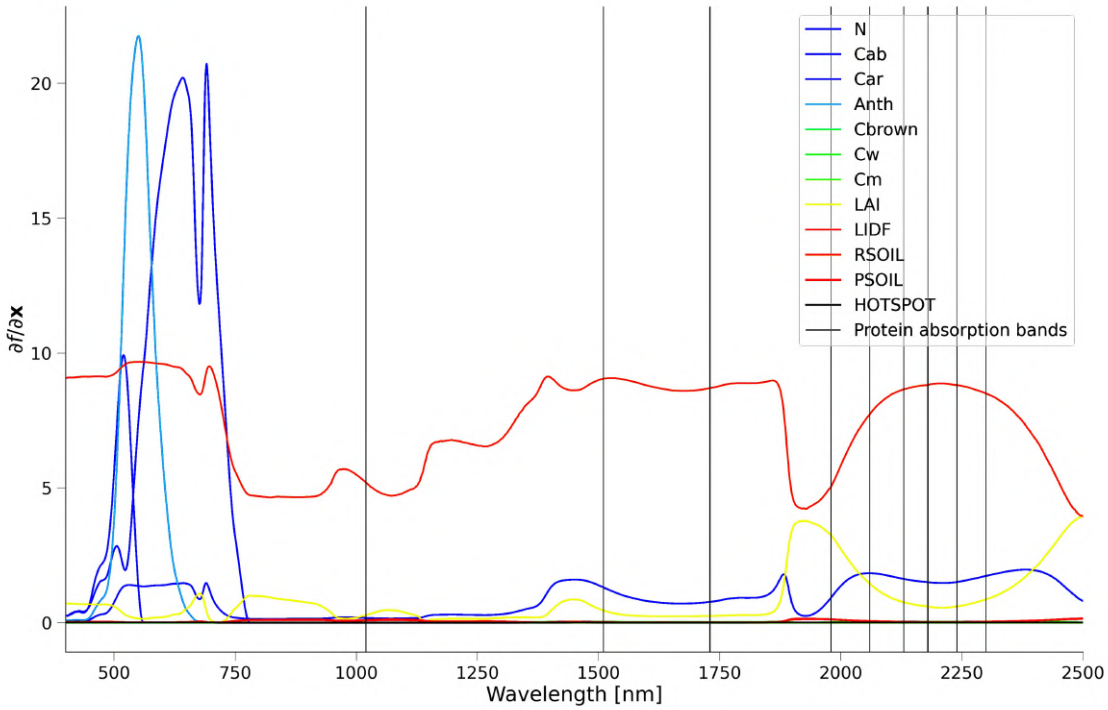


Figure 3.8: Illustration of protein absorption bands and sensitivity analysis of PROSAIL variables. It appears that the physical variables of the Radiative Transfer Model (RTM) provide a significant impact on the entire spectra, while some variables, such as chlorophyll content (Cab) and Anthocyanins (Anth), only impact the visible range of the spectra. This indicates the importance of constraining the physical variables of the RTM to retrieve the biochemical variables accurately.

Furthermore, significant challenges are associated with creating models using large training datasets, such as increased computation time (Verrelst et al., 2016a). Therefore, it is important to identify the optimal number of training samples needed for creating high-performing models. As a result, we investigated the impact of training dataset size on the performance of our hybrid, empirical, and physically-based models (Figure 3.7).

We found that the performance of our models starts saturating when using a training dataset of 3500 samples. This is because the 3500 samples used contain the spectral variability required for monitoring N% accurately. Verrelst et al. (2016a) investigated machine learning algorithms with RTM simulated datasets and highlighted the importance of using high-quality and informative samples rather than very large training datasets. Identifying and creating optimal and informative training datasets is critical, as it will reduce costs, decrease computation requirements, and increase model performance.

Figure 3.7 shows that the hybrid model performs better using a smaller training dataset. This is because the hybrid model takes advantage of the spectral and PROSAIL-derived features for monitoring N%. In addition, the physically-based model resulted in less uncertainty when using smaller training datasets, illustrating that the physically-based model is more robust when using small datasets as it relies on biophysical and biochemical variables for monitoring N%.

3.5.2 Identification of the optimal spectral features

A common method of monitoring heterogeneous grassland characteristics is using captured spectral measurements alongside empirical-statistical algorithms (Darvishzadeh et al., 2014, Quan et al., 2017, Proctor et al., 2017). For example, The use of Partial Least Squares Regression (PLSR) alongside grassland hyperspectral data is very popular, achieving different levels of accuracy (Foster et al., 2017, Pullanagari et al., 2012, Obermeier et al., 2019). PLSR transforms the collected spectral information into new predictor components with a lower feature space to explain target variables (N%) (Wold et al., 2001, Berger et al., 2020b,a). However, the use of such methods often produces poorer results with more uncertainty in comparison with more sophisticated machine learning approaches (Pullanagari et al., 2012).

Pullanagari et al. (2021) used a large spectral dataset with GPR, PLSR, and One Dimensional Convolutional Neural Network (1D-CNN) algorithms for predicting grassland N%. In addition, a GA was used to identify key spectral bands for monitoring grassland N% in this work. Here, it was found that the 1D-CNN and GPR significantly outperform PLSR. This study suggests that because PLSR transforms spectra into lower feature space in a linear orthogonal fashion, it is less suitable for identifying complex non-linear relationships between the spectra and target variables. As a result, it is important to identify spectral bands that are highly sensitive to grassland protein and chlorophyll content to develop high-performing machine learning and deep learning-based models for monitoring N%.

In addition, since the canopy spectral signal is influenced by many confounding factors (e.g., canopy structure and water content, bidirectional effects), selecting the optimal spectral domain is critical to retrieve accurate estimates of N%. This study used the GA method to identify the N% sensitive bands. As a result of this, the accuracy of the hybrid and empirical models increased significantly while maintaining stable performance on the independent dataset. Similarly, Berger et al. (2020a) used the GP-BAT algorithm to select optimal bands for the estimation of canopy N% from PROSAIL-PRO simulated spectra (Verrelst et al., 2016b).

Pullanagari et al. (2018) selected optimal bands for estimating protein in heterogeneous grasslands using recursive feature elimination with random forest (RFE-RF) and concluded that band selection could improve the model robustness by excluding the noisy bands. Despite the low signal-to-noise ratio (SNR) in SWIR, as illustrated in Figure 3.4, the majority of the bands were selected from the SWIR of the spectrum, which is related to fundamental protein absorption bands (Curran, 1989). In addition to the protein absorption bands, chlorophyll, and dry matter content were also selected as good predictors of N% for the hybrid method.

As seen in Figure 3.8, the LAI and other structural variables provided a strong impact on canopy reflectance, Knyazikhin et al. (2013), also stressed the role of canopy structure on N% retrieval, but the relationship was not clear. However, in this study, the structural information

was indirectly addressed by selecting bands in the NIR of the spectrum. We believe that the selected bands are used to compensate for the impact of LAI and Leaf Inclination Distribution Function (LIDF) on the spectra, particularly the SWIR of the spectrum.

3.5.3 Impact of season on the retrieved RTM variables

Figure 3.3 shows that the performance of the physically-based model is much higher during summer, autumn, and spring seasons and drops significantly in winter. The hybrid model provided higher performance than the physically-based model for all seasons, but the model's performance also decreased significantly during winter. This is because low temperature and limited solar radiation restrict photosynthesis, in which chlorophyll is the main component (Wingler and Hennessy, 2016, Hurtado-Uria et al., 2013). Lower temperature in winter slows ryegrass growth significantly, which will cause the Cab content present in grassland to remain low (Nagelmüller et al., 2016).

In addition, low temperature and incoming solar radiation will impact the growth and development of grass varieties differently, causing grass varieties to have different biochemical and biophysical characteristics under the same seasonal conditions (Lonati et al., 2009). As a result, chlorophyll content is not a robust predictor for heterogeneous grassland N% in winter. One reason that can cause the poorer performance of the physically-based model may be because PROSAIL is only capable of deriving the biochemical and biophysical proxies used in a regression model to predict N%. Therefore, using a physically-based model that can predict grassland N% directly may enable the retrieval of N% across different seasons more accurately (Li et al., 2018b).

3.5.4 Interpretation of the inverted biophysical and biochemical variables

Correlation analysis was undertaken by using PROSAIL-derived biochemical and biophysical variables and N%. As seen in Figure 3.5, chlorophyll, carotenoid, and canopy water content provided a very high correlation. This could be because chlorophyll and water content are closely linked to the transport of nutrients in plants, while carotenoid and chlorophyll content both contribute to the absorption of light energy (Bielinis et al., 2015, Gitelson et al., 2006).

Figure 3.5 also illustrates that protein content provides a moderate relationship with N%. The correlation between protein and N content is well known, as proteins consist mainly of N (Berger et al., 2020b). However, one cause of this moderate relationship may be because N% or N concentration provides a lower sensitivity to protein content. Another factor that may influence the correlation between these PROSAIL-derived variables may be due to uncertainty associated with the inverted variables.

3.5.5 Future recommendation of using RTM to predict N% in heterogeneous grasslands

One of the key advantages of using hybrid or empirical methods is that complex and highly non-linear information is embedded in the spectra, which may not be implemented within the RTM (such as the 3D structure of the canopy). Therefore, significant research is required to develop a more sophisticated RTM that incorporates more detailed information regarding the biochemical and biophysical components of grassland.

Due to the increasing number of new spaceborne hyperspectral sensors such as EnMAP and PRISMA, the continuous improvement of RTM is very important as it will enable monitoring of the biophysical and biochemical characteristics of grassland at a global scale (Labate et al., 2009, Guanter et al., 2015). However, due to the coarse spatial resolution of these sensors, the performance of hybrid approaches needs to be validated. This is not only of significant scientific importance but will allow the continuous and timely monitoring of grassland N%, which is required to optimize synthetic fertilizer application to heterogeneous grasslands.

3.6 Conclusion

Monitoring and managing grassland quality is key to efficient feed and protein conversion by grazing animals. Grassland quality is highly dependent on timely monitoring and application of N, a key macro-nutrient supporting grassland productivity. This work has developed three methods for monitoring N% of heterogeneous grassland using field spectroscopy, including empirical, physically-based, and hybrid models. The following are the key findings of this study:

1. Due to the ill-posed nature of RTM, the presence of multiple grass species, and the impact of the structural variables on the spectra, the accurate inversion of PROSAIL variables is very challenging.

2. Season has a significant impact on the model performance, winter models are less accurate than other seasons which could be due to low temperature, decrease in incoming solar radiation and grassland physiology. In addition, the hybrid model performs significantly better in the summer and autumn seasons.

3. Although Cab is commonly used as a proxy for determining N%, due to the limited amount of N present in Cab, it cannot be used robustly for approximating grassland N%. The results suggest that by successfully identifying canopy structural variables, it could be possible to estimate grassland N% accurately.

4. The hybrid model is a robust alternative method for estimating heterogeneous grassland N% as it helps address the key limitations of RTM and empirical methods while maintaining its key advantages.

Upcoming spaceborne hyperspectral sensors such as EnMAP will deliver a unique probe of environmental processes at a global scale. To ensure optimal exploitation of such a rich dataset,

we anticipate that hybrid models that leverage mechanistic and machine learning empirical modeling, such as the one presented here, will play a key role. This is particularly important for the analysis of heterogeneous grassland. Continuous monitoring of grassland N% on a global scale is essential for optimizing chemical fertilizer use, promoting environmentally sustainable practices, and achieving broader climate goals. This data will enable on-farm optimization, while also enhancing regional, national, and international efforts in climate adaptation and the accurate estimation of greenhouse gas emissions.

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Mohammad Hossain Dehghan Shoar
Name/title of Primary Supervisor:	Associate Professor Gabor Kereszturi
In which chapter is the manuscript /published work:	4

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

Dehghan-Shoar, M. H., Kereszturi, G., Pullanagari, R. R., Orsi, A. A., Yule, I. J., Hanly, J. (2024). A physically informed multi-scale deep neural network for estimating foliar nitrogen concentration in vegetation. International Journal of Applied Earth Observation and Geoinformation, 130, 103917.

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:

- The percentage of the manuscript/published work that was contributed by the candidate:

- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	
Date:	03-Sep-2024
Primary Supervisor's Signature:	Gabor Kereszturi Digitally signed by Gabor Kereszturi Date: 2024.09.03 22:06:21 +12'00'
Date:	3-Sep-2024

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

Chapter 4

A Physically Informed Multi-Scale Deep Neural Network for Estimating Foliar Nitrogen Concentration in Vegetation

4.1 Abstract

This study introduces a Physically Informed Deep Neural Network (PINN) that leverages spectral data and Radiative Transfer Model insights to improve nitrogen (N) concentration (N%) estimation in vegetation, addressing the complexities of physical processes. Utilizing a comprehensive spectroscopy dataset from various species across dry/ground (n=2010), leaf (n=1512), and canopy (n=6007) scales, the study identifies 13 spectral bands key for chlorophyll and protein quantification. Key bands at 2276 nm, 755 nm, 1526 nm, 2243 nm, and 734 nm emerged vital for accurate N% prediction. The PINN outperforms partial least squares regression and standard deep neural networks, achieving an R^2 of 0.71 and an RMSE of 0.42 (%N) on an independent validation set. Results indicate dry/ground data performed best (R^2 = 0.9, RMSE = 0.24 %N), with leaf and canopy data showing lower efficacy (R^2 = 0.67, RMSE = 0.45 %N; R^2 = 0.65, RMSE = 0.46 %N, respectively). This multi-scale approach provides insights into spectral and N% relationships, enabling precise estimation across vegetation types and facilitating the development of transferable models. The PINN offers a new avenue for analyzing remote sensing data, demonstrating the significant potential for accurate, scale-spanning N% estimation in vegetation. Further enriching our analysis, the inclusion of seasonal data significantly enhanced our model's performance in field spectroscopy, with notable improvements observed across summer, spring, autumn, and winter. This adjustment underlines the model's increased accuracy and predictive capability at the field spectroscopy scale, emphasizing the vital role of integrating environmental factors, including climatic and physiological aspects, in future research.

4.2 Introduction

Grasslands are indispensable for the global economy and ecological sustainability, serving as a crucial source of essential resources like food, income, and biofuels (Taubert et al., 2012). They play a pivotal role in mitigating climate change and providing critical ecosystem services such as soil conservation, water regulation, and biodiversity conservation, all of which are vital for human well-being (FAO, 2020, Taubert et al., 2012, Dehghan-Shoar, 2018, Petermann and Buzhdygan, 2021). Accurately estimating grassland nitrogen (N) levels is paramount for effective management (Goetz and Whitehead, 1971). Notably, grasslands are a significant source of food for livestock, highlighting the importance of adopting sustainable management practices, including precise N level monitoring, to ensure a stable and sustainable global food supply (Ramankutty et al., 2008).

Assessing N concentration (N%) in grasslands faces various challenges, including labor-intensive processes, limited spatial coverage, intrusive sampling methods, accuracy concerns, and high costs. This has prompted interest in alternative monitoring methods (Pullanagari et al., 2021, Dehghan-Shoar et al., 2023a,b). Spectroscopy, applied at different scales like dry/ground vegetation, leaf, and canopy, has emerged as a valuable technique for estimating N% in vegetation (Mutanga, 2004b). While this method can enhance performance, it often leads to empirical and data-driven models, which may lack generalizability. Despite achieving reasonable accuracy at individual scales, accurately estimating N% across multiple scales remains challenging due to diverse factors affecting spectral reflectance at each scale. For instance, measurements at the canopy scale, sensitive to energy-matter interactions, are crucial for determining vegetation biochemical and biophysical variables (Kokaly, 2001). However, accuracy at this scale can be influenced by variables like signal-to-noise ratio (SNR), sensor specifications, and environmental conditions such as soil exposure and water on leaves (Queally et al., 2022, Pullanagari et al., 2021).

To reliably monitor N% across different scales, it is essential to consider factors such as Leaf Area Index (LAI), canopy-scale leaf angle, internal leaf structure, and water content at both leaf and ground levels (Féret et al., 2021, Verhoef, 1984). This multi-scale approach represents a significant advancement over traditional methods reliant on soil or plant tissue sampling and subsequent laboratory analysis. While it offers improved performance and generalizability, capturing nuanced variations in N levels across diverse grasslands, its potential for facilitating timely decision-making in environmental management and agricultural practices requires further empirical validation. Nevertheless, its scalability and cost-effectiveness present a promising alternative for comprehensive assessments of grassland N distribution (Dehghan-Shoar et al., 2023a).

Models estimating N% at multiple scales, including leaf, canopy, and ground, are crucial for effective grassland N monitoring. This approach provides a nuanced understanding of N distribution and addresses the complexities of diverse grassland ecosystems. By integrating these scales, this strategy fills a critical research gap, offering a comprehensive solution for N

monitoring. Each scale presents unique challenges. Canopy scale data, offering extensive area coverage, are influenced by factors like atmospheric conditions and canopy structure (Verhoef, 1984, Queally et al., 2022). Although insightful for understanding plant responses to varying N%, leaf scale studies are labor-intensive and less suited for whole ecosystems (Féret et al., 2021). Dry/ground scale data, known for accurate N% measurements, face limitations due to potential sample preparation errors, the need for specialized equipment, and cost (Féret et al., 2021). Balancing cost and accuracy is the key challenge in remote sensing agriculture and environmental monitoring applications. Addressing these complexities necessitates the development of robust models that integrate physical principles, aiming to improve vegetation mapping for agricultural and environmental monitoring.

To enhance the accuracy of N% estimation in vegetation, we have devised a method that melds Physics-Informed Neural Networks (PINN) with Radiative Transfer Modeling (RTM) (Raissi et al., 2019). This innovative approach not only adheres to physical laws but also integrates vital biochemical and biophysical factors influencing plant growth. Our methodology underscores the significance of data directly linked to vegetation development, elucidating its pivotal role in precision agriculture. This is particularly evident in discerning key spectral features across various vegetation scales—such as at the dry/ground level, leaf, and canopy (Chen et al., 2022, Pullanagari et al., 2021). Further research bolsters our approach, highlighting the substantial benefits of merging spectral, structural, and textural data with traditional spectral reflectance methods. This amalgamation proves particularly effective in surmounting challenges commonly encountered with RTM, thereby enhancing the estimation of vegetation parameters (Ishaq et al., 2023, Dhakal et al., 2023, Dehghan-Shoar et al., 2023a,c, Pullanagari et al., 2021).

The adoption of a multi-dimensional strategy enhances the analysis of plant nutrients. For instance, studies involving unmanned aerial vehicles (UAVs) demonstrate that incorporating multispectral imaging with machine learning algorithms—factoring in canopy features (e.g., LAI)—results in more accurate biomass estimation (Dhakal et al., 2023). This underscores the advantages of employing diverse data types for precise plant analysis (Camps-Valls et al., 2023, Ishaq et al., 2023). Furthermore, the fusion of spectral data with other physical information proves advantageous in monitoring vegetation and estimating forage biomass, thus accentuating the value of integrating varied data types in plant studies (Pullanagari et al., 2018). Emphasizing the importance of incorporating a wide array of information further bolsters our comprehensive approach to understanding plant behavior (Pullanagari et al., 2018, Ishaq et al., 2023). Through the application of PINN alongside RTM, our technique augments N% estimation accuracy in vegetation while mitigating the ill-posedness typically associated with RTM. This method addresses the complexities associated with N% estimation through a holistic and integrated strategy, meeting the demands of contemporary agricultural and environmental research (Camps-Valls et al., 2023). It underscores the essential role of incorporating a broad spectrum of physically-based data, thereby enhancing the precision and depth of vegetation nutrient monitoring and affirming the significance of PINN in advancing precision agriculture.

4.3 Material and Methods

4.3.1 Canopy scale measurements

Between 2009 and 2016, we surveyed 17 dairy and hill country farms across New Zealand's South and North Islands. These farms were selected for their exposure to New Zealand's temperate climate, which allows for year-round grass growth moderated by annual precipitation ranging from 600 mm to 1600 mm. For detailed field campaign information and dataset specifics, refer to Pullanagari et al. (2021) and Dehghan-Shoar et al. (2023a). Spectral data were collected from canopies of seven principal grass species during spring (n=2940), autumn (n=2005), summer (n=854), and winter (n=208). We used ASD FieldSpec 4 Hi-Res and ASD FieldSpec PRO spectroradiometers with 512-element silicon photodiodes, offering high-resolution Visible Near Infrared (VNIR) (350-1000 nm) and Shortwave Infrared (SWIR) (1000-2500 nm) to collect spectral measurements. The Canopy Pasture Probe (CAPP) facilitated grassland spectral data capture, paired with a 50-W tungsten-quartz-halogen bulb for consistent illumination across varying weather conditions, calibrated against a white reference for optimal accuracy (Sanches et al., 2009). Stratified random sampling accounted for environmental factors like slope and elevation, with 20 spectral readings per 0.5 m^2 area to account for grassland diversity and reduce noise. Grass samples were collected, preserved, and geolocated using an RTK GPS for N% analysis by the Kjeldahl method and processed at the Analytical Research laboratory, Napier, New Zealand, involving sample drying, grinding, and homogenization for analysis with a QuikChem® 8500 analyzer.

4.3.2 Dry/ground scaled measurements

In 2014, a study on six New Zealand hill country and dairy farms assessed the spectral reflectance of grass samples, collecting 2010 samples for in-situ and laboratory analysis. The samples were dried, ground, and homogenized to ensure consistency. Spectral reflectance measurements were conducted using an ASD FieldSpec PRO spectroradiometer with a contact probe, taking 20 readings per sample to reduce noise and enhance data reliability. The instrument was calibrated with Spectralon™ standards to ensure accuracy. This approach aimed to estimate N% in the grass, following methodologies from previous research (Dehghan-Shoar et al., 2023c,a), capturing variability within each sample through comprehensive spectral analysis.

4.3.3 Leaf scale measurements

In our study, we leveraged two paramount leaf scale datasets: the LOPEX93 dataset (n=304 spectral reflectance samples) from the Leaf Optical Properties Experiment Database (Hosgood et al., 1995) and leaf spectra of N stressed maize (n=1210) from the Ecological Spectral Information System (EcoSIS) (Grzybowski et al., 2021, Ge et al., 2019), which span a wide array of plant species and conditions. The LOPEX93 dataset, developed at the Joint Research Centre in Ispra, Italy, encompasses a comprehensive collection of biophysical, biochemical, and spectral data for over 50 plant species. This collection includes data on fresh and dried leaves and

conifer needles. It stems, from spectral data acquisition conducted via a Perkin Elmer Lambda 19 spectrophotometer, ensuring high spectral resolution across the 400-2500 nm wavelengths. The resolution ranged from 1-2 nm in the VNIR (Visible and Near Infrared) spectrum to 4-5 nm in the SWIR (Short Wave Infrared) spectrum. This dataset notably includes reflectance and transmittance measurements, along with actual N percentages determined through the Kjeldahl method. Conversely, the maize dataset (Grzybowski et al., 2021, Ge et al., 2019) focuses on spectra from maize leaves subjected to N stress, recorded using a FieldSpec 4 spectroradiometer across the 350–2500 nm wavelengths. In this dataset, N levels were quantified employing the Dumas method, facilitated by a LECO FP428 analyzer. Many samples in the leaf dataset are accompanied by leaf N content per leaf dry mass (%DW) values. To ensure measurements were consistent and comparable across different environments or conditions, we selectively harmonized the data from the dry/ground, canopy, and leaf scales, aligning them based on the samples' total mass or volume as necessary. This harmonization was particularly important in scenarios where the N measurement units varied, ensuring our analysis maintained a uniform approach. In cases where the data was already consistent or the measurement discrepancies did not affect the analysis, we opted not to harmonize, preserving the original data integrity. This selective process of alignment allowed us to integrate biomass and dry weight information effectively, ensuring leaf scale datasets corresponded accurately with those from dry/ground and canopy levels. Our harmonization also involved processing reflectance data from each dataset, where we conducted pre-processing to exclude outliers in spectral signatures or samples with missing N values. This careful pre-processing safeguarded the data's reliability, essential for incorporating a wide array of spectral and biochemical data in a thorough examination of our leaf scale dataset. After this meticulous preparation, we had a total of 1,512 samples ready for our study.

4.3.4 Inversion of biophysical and biochemical variables

We employed the SAIL and PROSPECT-PRO RTM models, integrating dry/ground, leaf, and canopy scale data, to accurately estimate key biophysical and biochemical variables such as LAI, Protein Content (Cp), and Chlorophyll Content (Cab). These variables are pivotal for assessing vegetation health and growth dynamics. PROSAIL, a merger of the SAIL and PROSPECT-PRO models, facilitates the precise inversion of these variables by modeling canopy reflectance and vegetation biochemical properties (Féret et al., 2021). SAIL, introduced by Verhoef (1984), simulates canopy reflectance, accounting for light-plant interactions, while PROSPECT-PRO, initially conceptualized by Jacquemoud and Baret (1990) and later refined by Féret et al. (2021), is renowned for its effectiveness in biochemical estimation at leaf scale. The inversion process involved extensive forward and inverse RTM simulations in minimizing discrepancies between observed and modeled spectra. This optimization utilized the Affine Invariant Markov Chain Monte Carlo (MCMC) method, supported by 1000 walkers and 50,000 simulations, leveraging the Python EMCEE ensemble toolkit for its robustness and precision in error minimization (Foreman-Mackey et al., 2019). We refined our inversion methodology for enhanced accuracy by fine-tuning key variables within the PROSAIL model, such as viewing and illumination

angles, as detailed in Table 4.2. We calculated the mean and standard deviation for variables of each sample across three scales: dry/ground (n=2010), leaf (n=1512), and canopy (n=6007), thus compiling a comprehensive dataset comprising n=9529 samples (Table 4.1).

Table 4.1: Sample size and summary statistics for nitrogen (N) concentration (N%) across different scales and combined

Dataset	**Max**	**Mean**	**Min**	**Samples**	**Std Dev**
Dry/ground scale	5.11	2.85	0.97	2010	0.75
Leaf scale	6.00	2.61	0.68	1512	0.81
Canopy scale	5.32	2.54	0.73	6007	0.79
Combined	6.00	2.62	0.68	9529	0.80

For dry/ground and leaf scale data, we utilized the PROSPECT-PRO model, renowned for its accurate simulation of optical properties including the structure variable, chlorophyll content, carotenoid content, brown pigment content, equivalent water thickness, dry matter content, protein content, carbon-based constituents, and anthocyanins content. For canopy scale data, we employed the PROSAIL model, which integrates both leaf and canopy structural effects alongside leaf scale properties such as LAI, leaf angle distribution parameters (a and b), and the hotspot parameter (Figure 4.1). To guarantee data independence and minimize sample overlap, we meticulously divided the dataset into independent calibration (n=7623) and validation (n=1906) subsets, ensuring that samples for canopy and dry/ground scales were collected from distinct spatial and temporal locations. Specifically, the calibration subset included dry/ground (1608 samples), LOPEX (244 samples), Maize stress (966 samples), and canopy data (4805 samples). The validation subset consisted of dry/ground (402 samples), LOPEX (60 samples), Maize stress (242 samples), and canopy data (1202 samples). This strategic division, underlining spatial and temporal separation, facilitates a detailed and unbiased evaluation of our models across various vegetative states and scales.

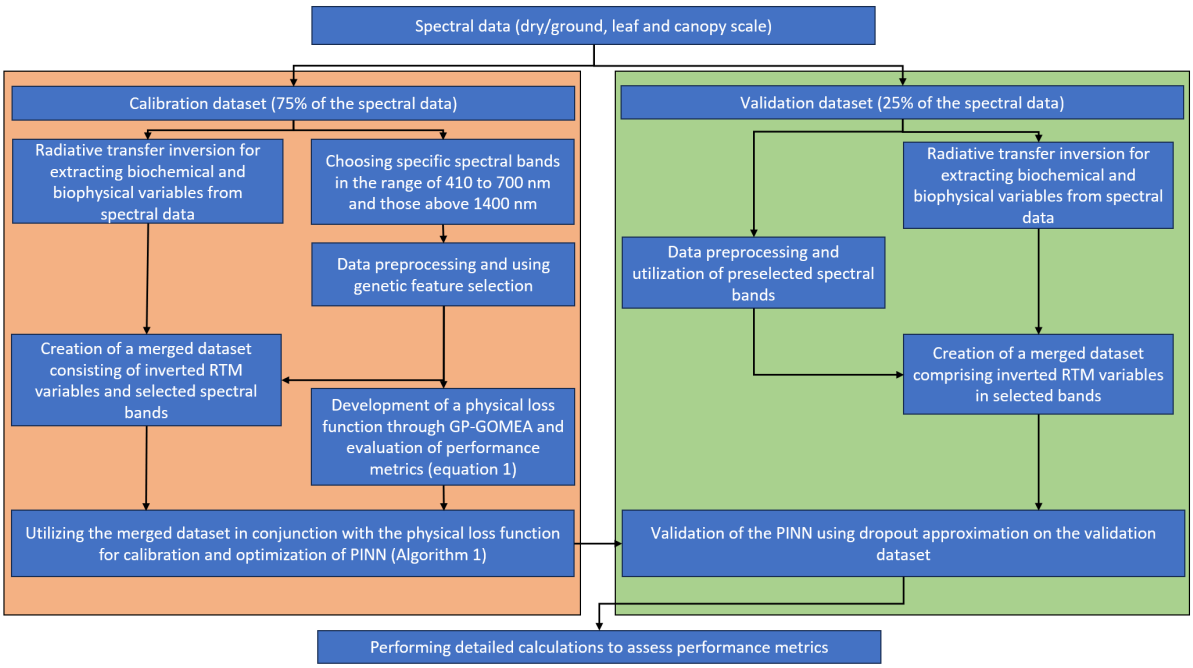


Figure 4.1: Flowchart depicting the methodology used in this study to predict nitrogen (N) concentration (N%) employing Physics-Informed Neural Networks (PINN). Spectral bands chosen through Genetic Feature Selection (GA) and inverted variables from leaf levels at dry/ground, leaf, and canopy scales served as input for the PINN model.

Table 4.2: The search space for PROSAIL variables was optimized using Markov Chain Monte Carlo (MCMC) for inversion, setting the Equivalent Water Thickness (Cw) to zero for dry/ground samples to enhance performance while leaving Cw unconstrained for leaf and canopy data.

Variable name	Min value	Max value	Abbreviation	Units
Structure variable	1	3	N	
Chlorophyll content	1	100	Cab	cm2.microg-1
Carotenoid content	1	30	Car	cm2.microg-1
Brown pigment content	0	0.2	Cbrown	
Equivalent water thickness	0	0.02	Cw	cm-1
Dry matter content	0	0.02	Cm	cm2.g-1
Protein content	0	0.02	Cp	cm2.g-1
Carbon based constituents	0	0.02	CBC	cm2.g-1
Anthocyanins content	1	10	Anth	cm2.nmol-1
Leaf Area Index	1	3	LAI	
Leaf angle distribution a	10	90	LIDFa	
Leaf angle distribution b	10	90	LIDFb	
Background reflectance	-	-	RSOIL	
Dry/Wet soil factor	-	-	PSOIL	
Hotspot parameter	0	1	HOTSPOT	
Solar zenith angle	0	0	SZA	Degrees
Observer zenith angle	20	20	VZA	Degrees
Relative azimuth angle	160	160	RAA	Degrees

4.3.5 Physically-based equation for estimation of N%

In our study, we formulated a physically-based equation for accurately estimating N% by focusing on spectral regions sensitive to Cab within the 410-780 nm range and Cp beyond 1400 nm, identifying these regions as critical for precise N% predictions (Féret et al., 2021). To optimize our model, we implemented Genetic Feature Selection (GA), a process that successfully pinpointed 13 vital spectral bands intimately linked with the presence of Cab and Cp (Jarvis and Goodacre, 2005). This selection was crucial for refining our approach, as GA methodically sifted through a vast dataset to identify the most informative spectral bands, significantly enhancing the predictive model's accuracy, especially in applications like remote sensing or spectroscopy (Dehghan-Shoar et al., 2023a). GA refined spectral band selection for N% estimation in vegetation by simulating natural evolutionary processes. Initially, GA generates a diverse array of potential solutions, each representing a different combination of spectral bands. Solutions are assessed for their predictive accuracy of N%, with the most effective ones selected for reproduction. During the crossover phase, these selected solutions exchange portions of their genetic code, creating new combinations that potentially enhance predictive performance. Mutation introduces random variations, ensuring diversity and aiding in the discovery of novel, effective band combinations. This cycle of selection, crossover, and mutation repeats, gradually honing in on a highly efficient set of spectral bands. The iterative, evolutionary nature of GA allows

it to dynamically explore and optimize the selection space, effectively narrowing down the vast array of possible combinations to those most critical for accurate N% estimation. This process showcases GA's ability to focus on the most relevant spectral information, markedly improving the model's efficiency and accuracy in complex predictive modeling tasks.

In our work, we leveraged the Gene-pool Optimal Mixing Evolutionary Algorithm for Genetic Programming (GP-GOMEA) to enhance N% estimation across varied vegetative conditions such as dry/ground, leaf, and canopy levels (Virgolin et al., 2021). GP-GOMEA excels by efficiently merging segments of pre-existing solutions, significantly elevating the fitness of successive generations. This is achieved through a comprehensive analysis of interactions between solution components, directing the algorithm towards the most viable solutions. Importantly, GP-GOMEA is adept at estimating interdependencies among model components, employing this knowledge to execute crossover operations of interdependent components en bloc (Virgolin et al., 2020). This approach ensures the preservation of their collective efficacy, a critical feature for solving complex problems that demand integrated solutions. GP-GOMEA is notably skilled at deriving models in the form of compact symbolic expressions, which are often interpretable, providing an added layer of understanding. By applying GP-GOMEA, we developed a precise model that selectively utilized 10 of the initially identified 13 key spectral bands associated with crucial plant elements—protein and chlorophyll—essential for accurate N% estimation. The algorithm's decision to select 10 features reflects a strategic focus on maximizing model efficiency and information utility, by sidelining features with limited additional value or those contributing to multicollinearity, thereby improving the model's predictive accuracy and generalizability. This process not only illustrates GP-GOMEA's capacity to pinpoint and prioritize the most impactful variables for complex estimations but also highlights its significant methodological contribution to precision agriculture, emphasizing its ability to find models that balance simplicity with interpretability and effectiveness.

Our methodology intricately weaves the selected spectral bands into a PINN, leveraging the absorption characteristics of Cab and Cp to enhance N% predictions. This integration serves as a crucial regularization factor within the network. Further elevating our model's clarity and analytical depth, we incorporate SHapley Additive exPlanations (SHAP) for a comprehensive feature importance analysis of the 13 GA selected bands (Lundberg and Lee, 2017). This inclusion of SHAP values illuminates the specific contributions of various spectral inputs toward the model's predictions, significantly boosting both the precision and the interpretability of our results. By meticulously setting hyperparameters for both GA and GP-GOMEA, as detailed in Table 4.3, we ensure a refined and detailed process for N% estimation, reflecting our dedication to achieving a model that is both accurate and transparent in its predictive capabilities.

4.3.6 Creating the PINN

To estimate N%, our study utilized a PINN incorporating a physically-based equation derived through the GP-GOMEA and optimized via TensorFlow across 10,000 iterations using Bayesian

Table 4.3: Genetic algorithm (GA) and symbolic regression hyperparameters used.

Variable	**GA variables**	**GP-GOMEA variables**
Fitness function	MSE	MSE
Population size	100	64
Crossover probability	0.1	0.5
Mutation probability	0.05	0.5
Number of generations	10000	No limit
Independent crossover probability	0.5	0.5
Independent mutation probability	50	0.5
Tournament size	0.5	—
Max tree height	—	6
Linear scaling	—	True
Function set	—	[+, −, _p, /_$sqrt$, _$p$log]

optimization (Schell, 2022, Grantbuster et al., 2021). We have integrated the physical loss function into the PINN by implementing Equation 1 within a custom loss function built using TensorFlow. Additionally, we have introduced weighting factors for both the standard Mean Squared Error (MSE) and the physical loss, as proposed by Grantbuster et al. (2021). The PINN model inputs include inverted RTM variables (N, Cab, Car, Cbrown, Cw, Cm, Cp, CBC, and Anth) and 13 spectral bands identified by GA. Bayesian optimization refined the PINN's architecture, improving performance through a Gaussian process model on hyperparameter efficiency (Table 4.4 and Figure 9.1) (Bergstra et al., 2015). Additionally, dropout-based Bayesian approximation quantified the uncertainty in PINN predictions, creating a predictive distribution that assesses model uncertainty, thus offering insights into the model's prediction confidence levels (Junhwan et al., 2022) (Algorithm 1). The PINN's efficacy was verified using an independent dataset, evaluating its accuracy with metrics like Root Mean Squared Error (RMSE), MSE, and coefficient of determination (R^2), detailed in our methodology section. We further benchmarked the PINN by comparing its performance with several established regression methods, each with unique characteristics and applications. Partial Least Squares Regression (PLSR) is a technique that reduces the predictors to a smaller set of uncorrelated components and performs linear regression on these components, making it effective in handling situations where predictors are highly collinear. Support Vector Regression (SVR) works by mapping input features into high-dimensional space and finding the hyperplane that best fits the data with a margin of tolerance, ideal for complex, non-linear relationships (Kecman, 2005). Random Forest Regression (RF) utilizes an ensemble of decision trees to improve prediction accuracy and control over-fitting, making it versatile for various data types without the need for extensive parameter tuning (Liang et al., 2018). Gaussian Process Regression (GPR), on the other hand, is a probabilistic model that provides not only predictions but also the uncertainty of those predictions,

making it suitable for cases where estimation of prediction uncertainty is crucial (Pullanagari et al., 2021). In addition, to evaluate the robustness of each model against bias and noise, we introduced heteroscedastic noise into calibration datasets spanning different scales. Random values (-0.45 to 0.45) were added to the original values, creating variable noise variance.

Table 4.4: Summary of key variables utilized in the multi-scale Physics-Informed Neural Networks (PINN) model for nitrogen (N) concentration (N%) in vegetation.

Parameter	**Selected value**	**Search space**
Model Architecture	Sequential	
Hidden Layers	Relu	[Linear, ReLU, and Sigmoid]
Dropout Layer	0.2	[0.1-0.5]
Output Layer	Linear	
Optimizer	Adam	
Number of layers	3	[2-5]
Number of neurons	100	[50-300]
Loss Functions	MSE and physical loss	
Loss Weights	[0.2, 0.8]	
Batch Size	10000	
Epochs	20000	

Algorithm 1 Algorithm depicting Physics-Informed Neural Networks (PINN) with Bayesian approximation using dropout.

Initialize: the model variables (weights and biases) and set the dropout rate
Load: the calibration data, physical values (N%), and RTM variables and spectral bands
Define: the physical loss function, which calculates the difference between the network's outputs and physical values
for each calibration sample **do**
    Input: the current data (spectral bands and RTM variables) into the PINN
    Apply Dropout : randomly drop out units during forward Propagation with the specified rate
    Forward Propagation : propagate the inputs through the network to calculate the outputs
    Calculate Loss : calculate the physical loss and mean squared error
    Backpropagation: use the physical loss and mean squared error to update the model variables
end for
Make Predictions: use the trained model to make predictions on new data by turning off the Dropout for prediction
Bayesian Approximation : use Monte Carlo sampling to obtain posterior distributions of the model variables, given the calibration data and physical loss function

4.4 Results

4.4.1 Variables inversion

We applied MCMC numerical optimization to invert the RTM variables, enabling the assessment of uncertainties in dry/ground, leaf, and canopy data (Figure 4.2). Dry/ground data inversion showed superior performance and reduced uncertainty compared to leaf and canopy data. Canopy scale data exhibited lower uncertainty in the visible spectrum than NIR and SWIR, which is attributed to the stronger influence of biophysical factors. However, canopy inversion demonstrated reduced performance compared to leaf data, highlighting the critical role of interleaf interactions.

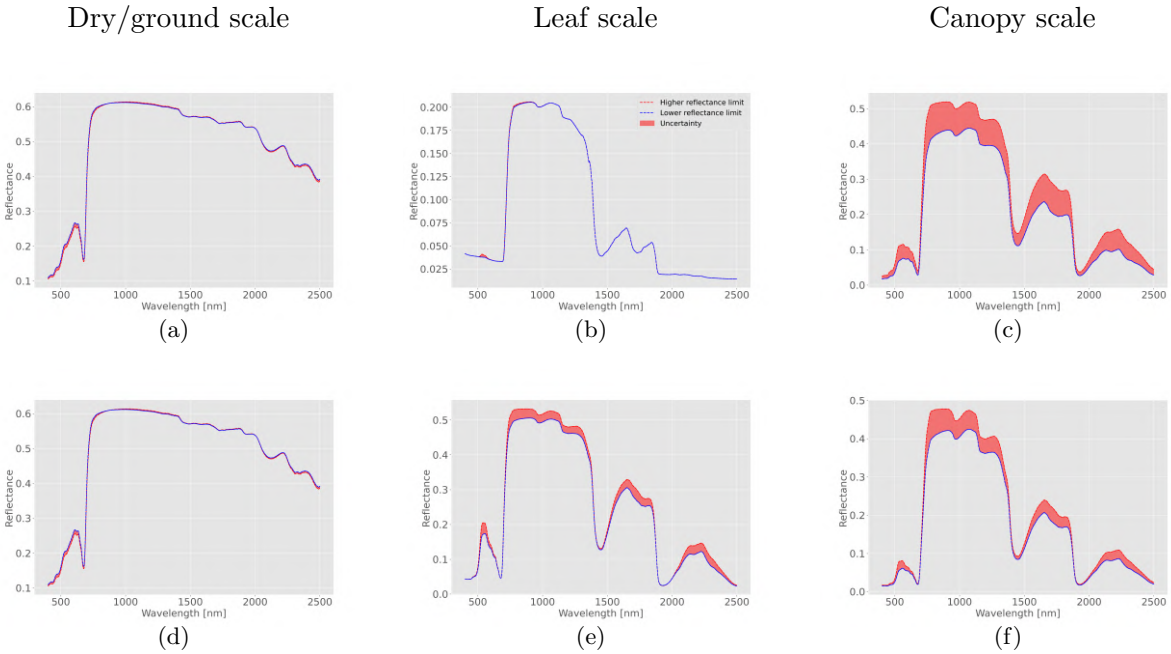


Figure 4.2: Spectral band results from the inversion of PROSPECT-PRO and PROSAIL variables across dry/ground (a and d), leaf (b and e), and canopy (c and f) scales exhibit a trend of increasing uncertainty. This escalation is particularly significant from dry/ground to leaf, and even more so at the canopy scale, signaling issues related to the complexity of canopy structure and the inherent ill-posed nature of Radiative Transfer Models (RTM). These uncertainties underscore the challenges in accurately capturing the intricate interactions within canopy systems through spectral data.

4.4.2 Physically-based equation

We utilized GA for feature selection, a crucial step that helped us identify the fundamental bands correlating with Cab and Cp. We then applied GP-GOMEA to derive equation 1, allowing us to integrate a physical loss function into the PINN. This function, shown below in Figure 4.3, is designed to mitigate the adverse effects of noise and bias and is expressed as follows:

$$N\% = A - \frac{B \cdot (C \cdot \log(\frac{D}{E}))}{\left(\frac{E}{D}\right)^{\frac{1}{4}} \cdot F \cdot G \cdot \sqrt{\frac{E \cdot \log(D)}{J}} \cdot \log\left(\frac{\sqrt{H}}{\log(I)}\right)} \tag{4.1}$$

$A = 1.27$

$B = 17.95$

$C = \log(569\,\text{nm})$

$D = \sqrt{734\,\text{nm}}$

$E = \sqrt{755\,\text{nm}}$

$F = \log(2077\,\text{nm})$

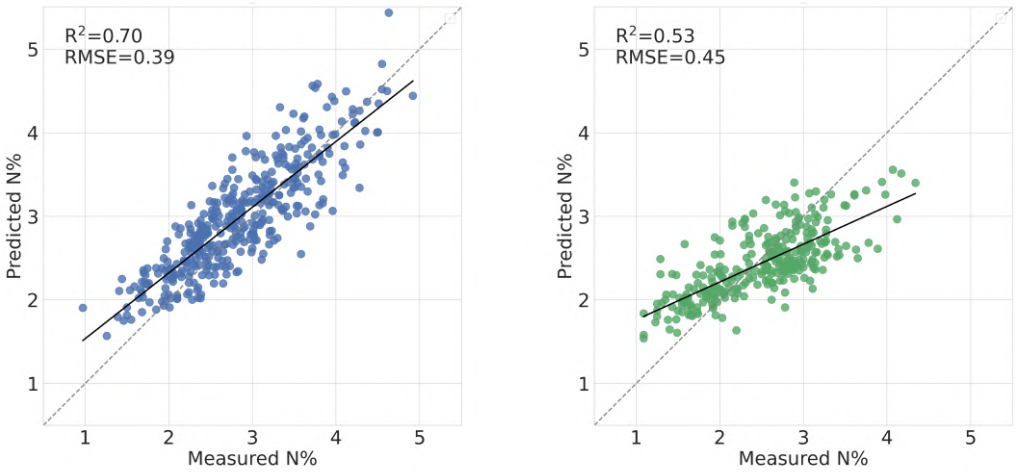
$G = \log\left(\frac{2276\,\text{nm}}{2177\,\text{nm}}\right)$

$H = \log(1559\,\text{nm})$

$I = \frac{2276\,\text{nm}}{2077\,\text{nm}}$

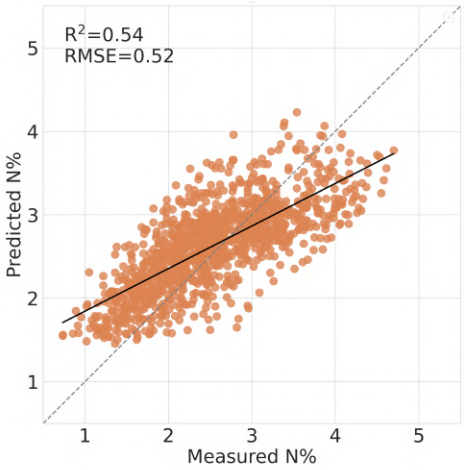
$J = 467\,\text{nm}$

Our research identified key spectral bands at 2276 nm, 755 nm, 1526 nm, 2243 nm, 734 nm, 2177 nm, 2375 nm, 1559 nm, 569 nm, 1760 nm, 2077 nm, 2473 nm, and 467 nm as crucial for estimating N% where most key spectral bands were identified in the SWIR (Figure 4.4 b). Using equation 4.1 and independent validation data (Figure 4.3), we achieved 0.70 R^2 (RMSE = 0.39 %N) for dry/ground, 0.53 R^2 (RMSE = 0.45 %N) for leaf, and 0.54 R^2 (RMSE = 0.52 %N) for canopy scale data. While these bands consistently influenced N% predictions across scales, their overall impact varied (Figure 4.4 a). Notably, bands at 2276 nm, 755 nm, 1526 nm, 2243 nm, and 734 nm were most influential for N% estimation across all scales (Figure 4.4 a). We observed an inverse correlation between reflectance at 1526 nm, 2243 nm, and 734 nm and increased N%, as detailed in Figure 4.5. Significant non-linear relationships between N% and these spectral bands were evident (Figure 4.5 e, f, g, and j).



(a) Dry/ground scale

(b) Leaf scale



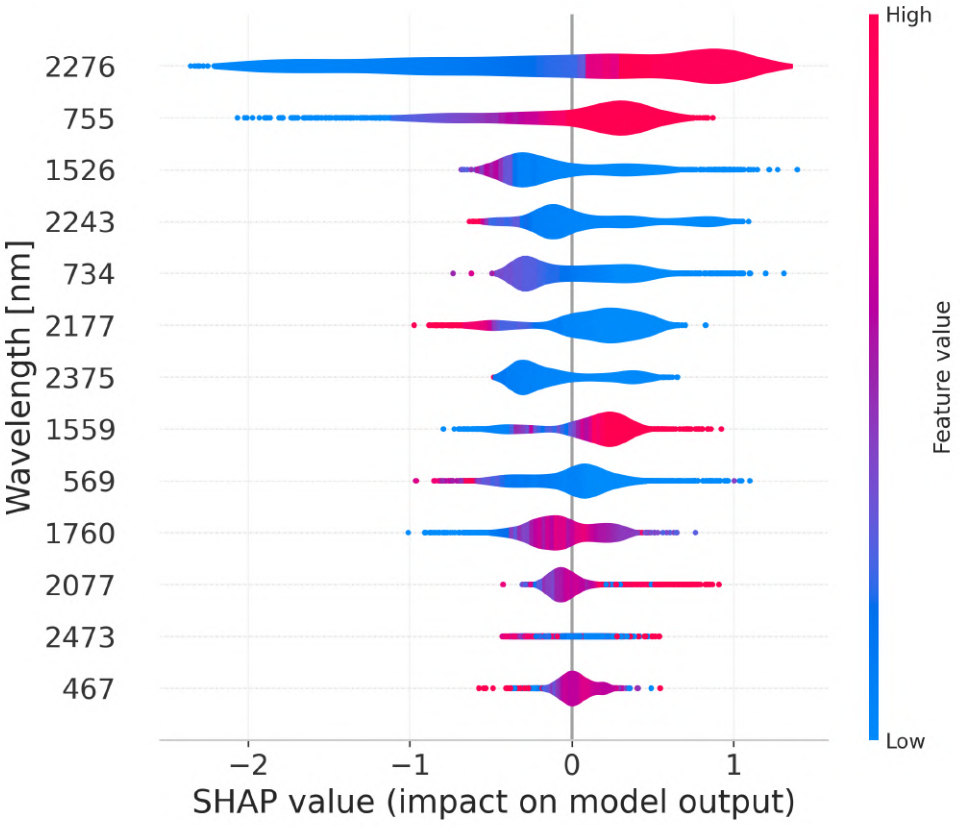
(c) Canopy scale

Figure 4.3: Equation 1, derived from GP-GOMEA with a physical loss function, effectively estimates nitrogen (N) concentration (N%) across scales: dry/ground (panel a), leaf (panel b), and canopy (panel c). Its performance, influenced by key spectral bands from Genetic Feature Selection (GA), demonstrates precision and adaptability in various grassland environments.

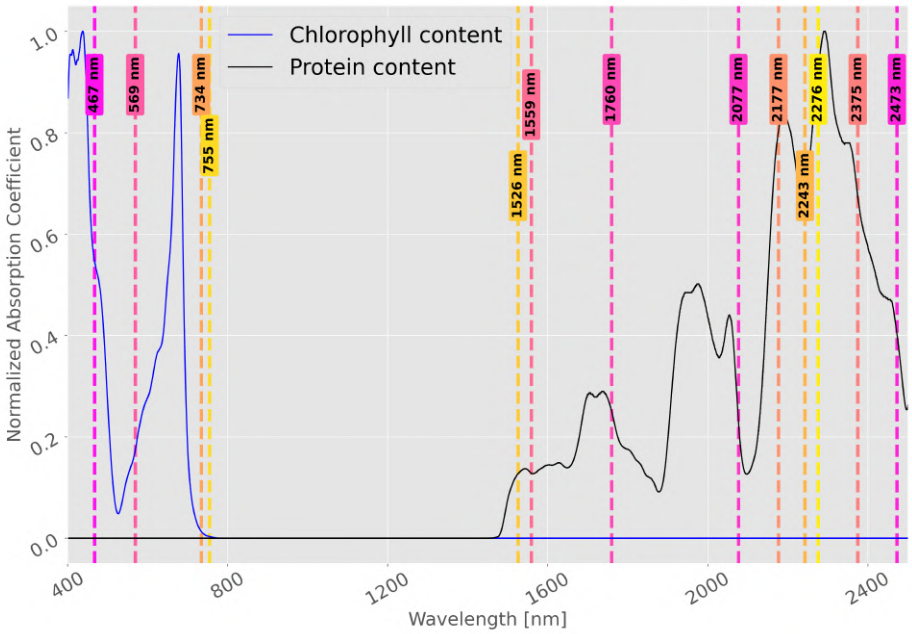
We created 13 dependence plots to scrutinize the interplay between pivotal spectral bands, as pinpointed by GA, and their correlation with N% content. This suite of plots vividly demonstrates how each chosen spectral band interacts with N levels within plant tissues, offering a deep dive into the predictive power of specific wavelengths (Figure 4.5). Subplots (a) to (m) systematically explore these relationships, with particular emphasis on bands at 2276 nm and 755 nm (subplots c and d), marked as primary predictors of N%. The 755 nm wavelength, linked to Sun Induced Fluorescence (SIF), emerges as a significant factor in N% estimation, spotlighting the utility of SIF in agricultural monitoring (Mohammed et al., 2019).

Moreover, the 2276 nm band is distinguished by its high SHAP value, reinforcing its indispensable role in accurate N% prediction. Contrastingly, the 2375 nm and 2473 nm bands within the SWIR range, alongside the 467 nm band in the visible spectrum (subplots l, m, and a, respectively), exhibit lower SHAP values. This suggests a relatively lesser impact on N% estimation, though their inclusion provides a fuller spectral analysis narrative. Notably, some plots reveal U-shaped or parabolic trends, underscoring the non-linear dynamics between certain spectral bands and N%. This complexity is critical for understanding the multifaceted nature of spectral data analysis in agricultural contexts. Subplot (a) sets the stage by comparing 467 nm with 1760 nm, revealing distinct clusters that imply a strong dependency of spectral reflectance on N levels. Moving through the series, subplot (b) analyzes the 569 nm and 1760 nm wavelengths, suggesting an inverse relationship reflective of how vegetation absorption properties vary with N content across these frequencies.

Subplot (c) focuses on 734 nm and 755 nm, showcasing a dispersed data pattern that highlights the intricate relationship these wavelengths share with vegetation N%, pivotal for refining spectral analysis techniques. As we progress to subplots (d) through (m), each plot unfolds further layers of insight. For instance, subplots (d) and (e), which contrast 755 nm with 467 nm and 1526 nm with 2243 nm respectively, delve into the diverse ways N% influences plant reflectance, showcasing a range of correlation degrees and data dispersion patterns.

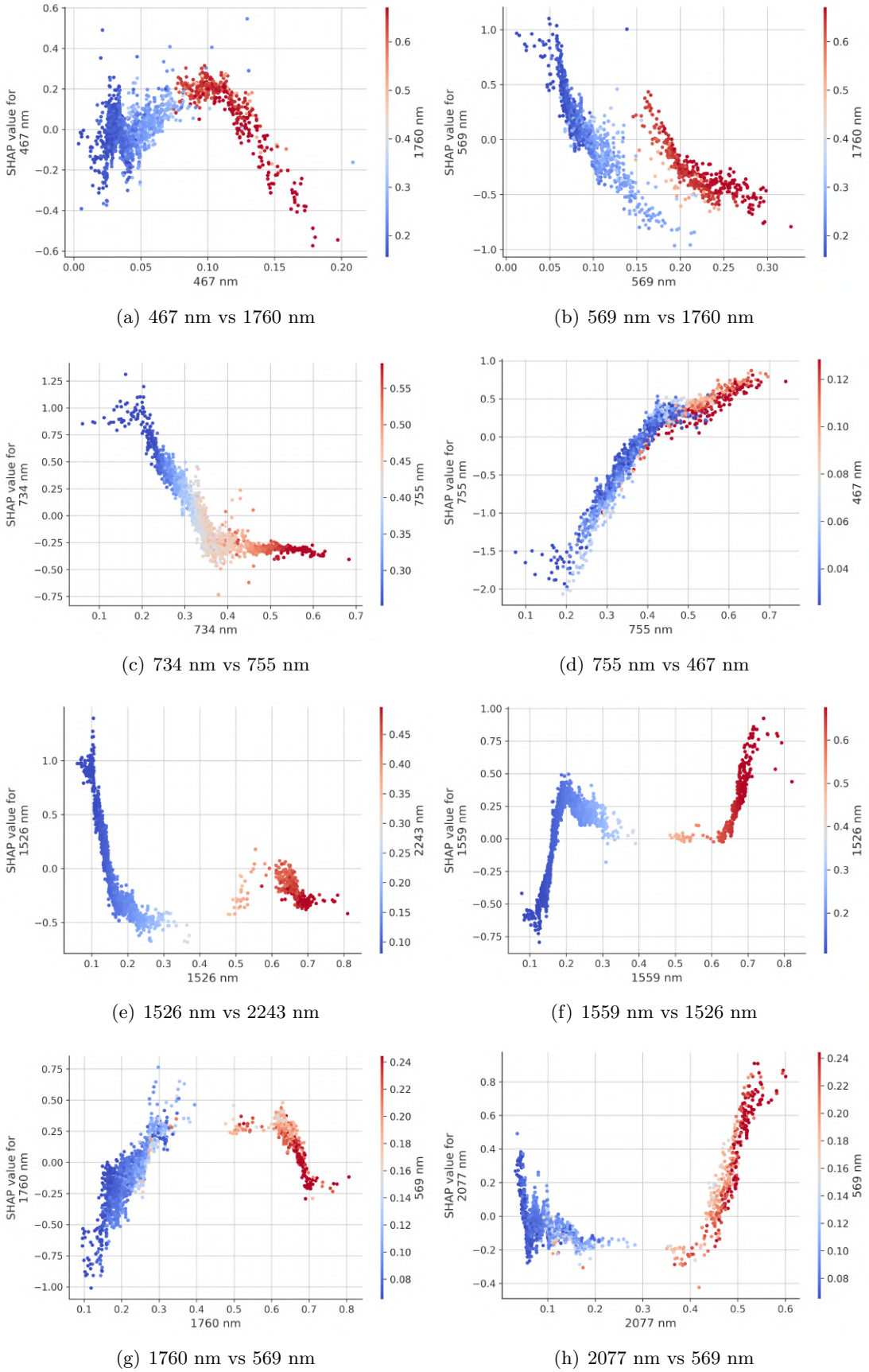


(a) Feature importance



(b) Spectral absorption

Figure 4.4: Plot (a) shows a SHAP (SHapley Additive exPlanations) summary highlighting key spectral bands for nitrogen (N) concentration (N%) estimation. Plot (b) reveals the sensitivity of these bands to protein and chlorophyll, crucial for the physical loss function and Genetic Feature Selection (GA), with chlorophyll and protein absorption marked in blue and black, respectively. The absorption coefficients are generated using PROSPECT-PRO.



(a) 467 nm vs 1760 nm

(b) 569 nm vs 1760 nm

(c) 734 nm vs 755 nm

(d) 755 nm vs 467 nm

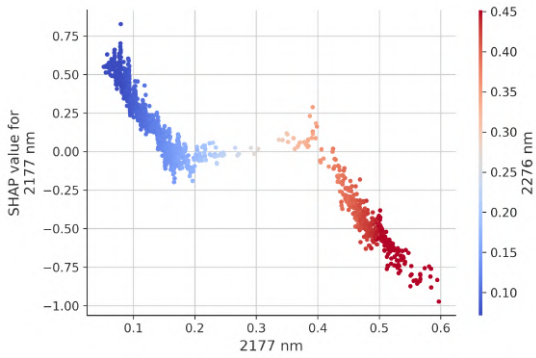
(e) 1526 nm vs 2243 nm

(f) 1559 nm vs 1526 nm

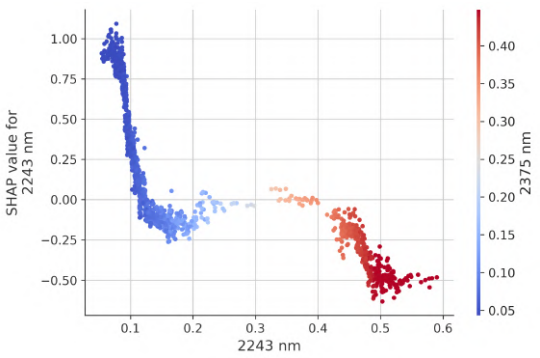
(g) 1760 nm vs 569 nm

(h) 2077 nm vs 569 nm

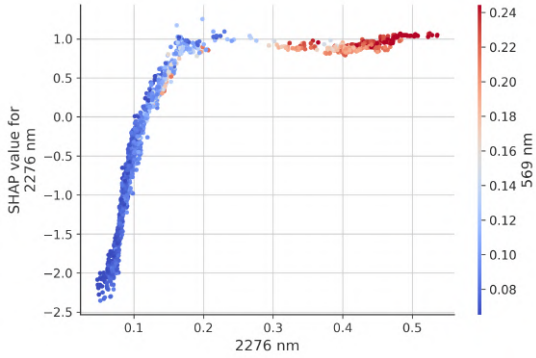
Figure 4.5: Integrated reflectance interaction plots across vegetation scales for nitrogen (N) concentration (N%) estimation. The plots depict mixed data from dry/ground, leaf, and canopy scales, plotting the reflectance of one spectral band on the x-axis against another band on the y-axis. The color intensity represents the SHAP (SHapley Additive exPlanations) values for the band on the y-axis, offering insights into its predictive relevance. To interpret the figure, observe how the data points' position and color variation signify the inter-band reflectance relationships and their combined predictive influence on N% across the mixed scales.



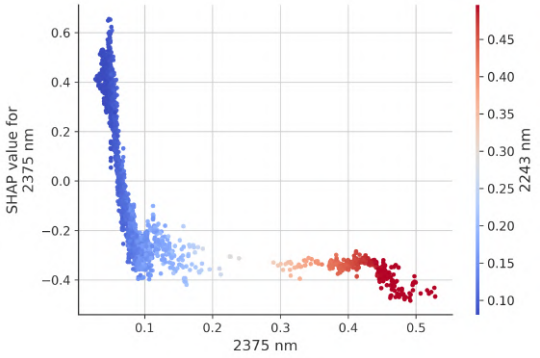
(a) 2177 nm vs 2276 nm



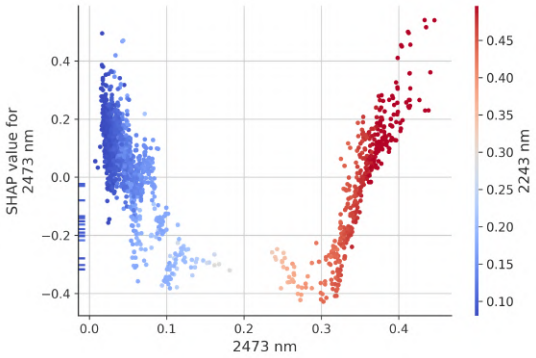
(b) 2243 nm vs 2375 nm



(c) 2276 nm vs 569 nm



(d) 2375 nm vs 2243 nm



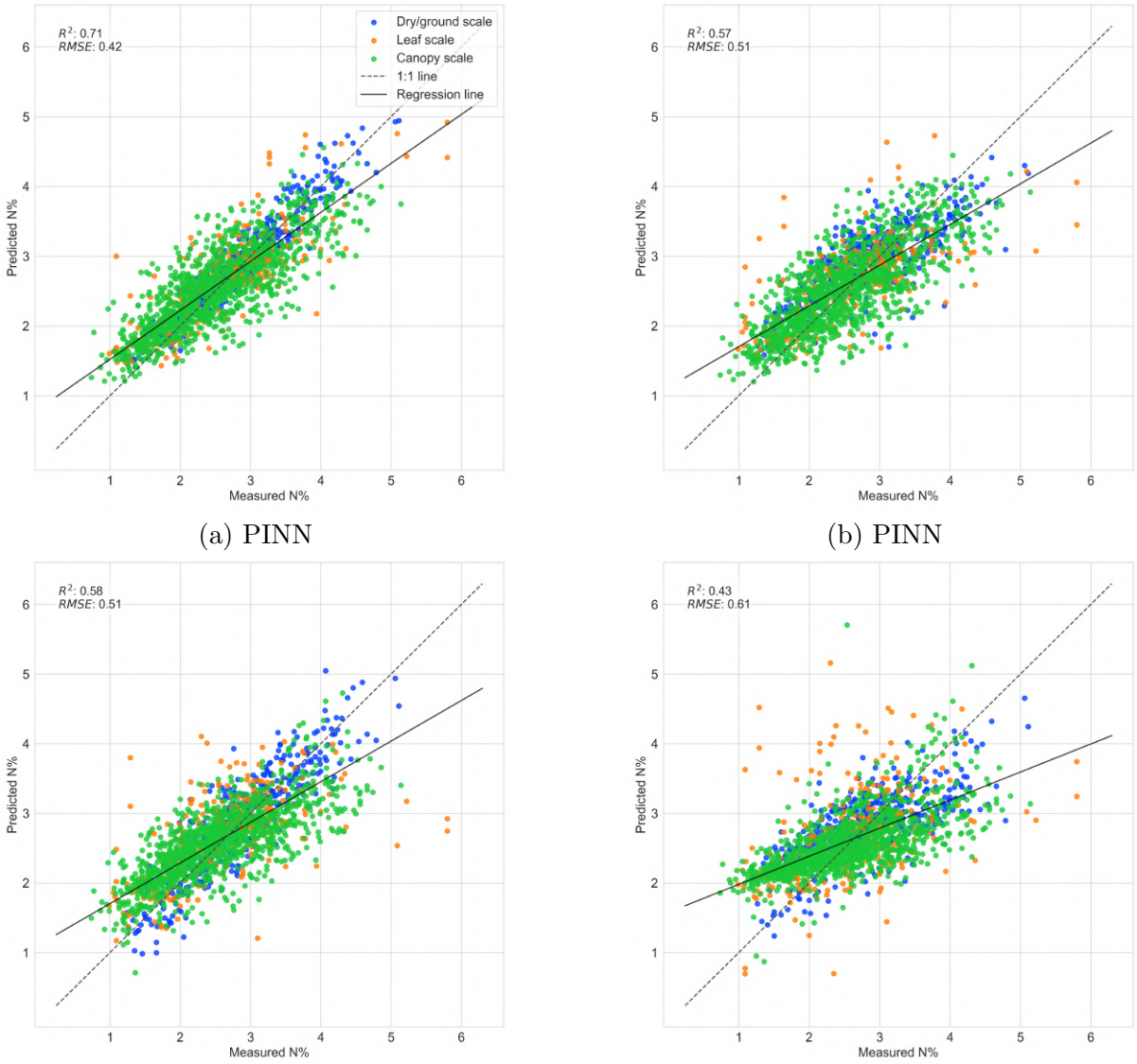
(e) 2473 nm vs 2253 nm

Figure 4.5: Continued.

4.4.3 Performance of PINN

The PINN's performance was assessed with RTM-derived variables exclusively and in combination with important spectral bands. Critical RTM-derived variables for N% estimation were Cab, Cw, Cbc, and Cp. Integrating spectral bands with RTM features improved model validation (R^2 = 0.71, RMSE = 0.42 N%) compared to using RTM variables alone (R^2 = 0.57, RMSE = 0.51 N%) (Figure 4.6). Including spectral bands reduced error and outliers, evidenced by a reduced Interquartile Range (IQR) of 1.5. A z-test ($p<0.05$) confirmed significantly lower uncertainty in models with spectral bands (Figure 9.3).

Spectral bands and RTM variables **Only Radiative Transfer Model (RTM) variables**

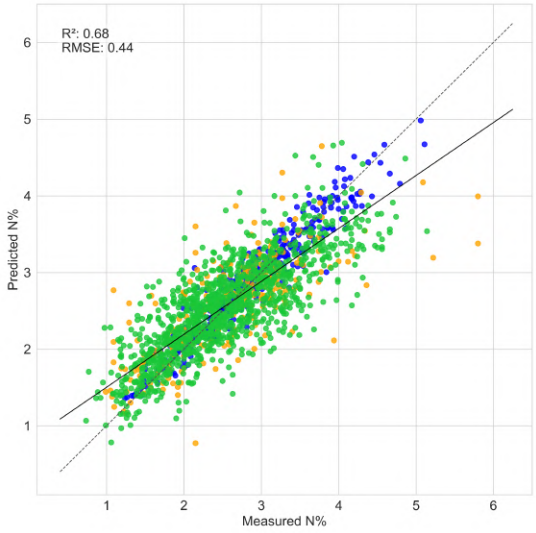


(a) PINN (b) PINN

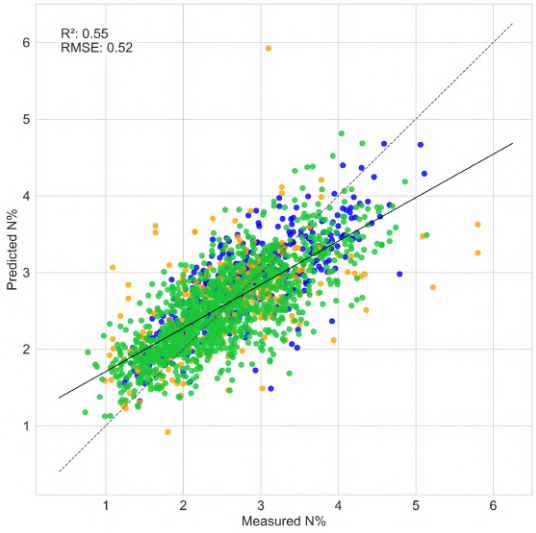
(c) PLSR (d) PLSR

Figure 4.6: The comparative analysis of Physics Informed Neural Network (PINN), Partial Least Squares Regression (PLSR), Gaussian Process Regression (GPR), Random Forest Regression (RF), and Support Vector Regression (SVR) in estimating nitrogen (N) concentration (N%) based on spectral bands related to chlorophyll and protein content demonstrates the superior performance of PINN. These results illustrate that PINN outperforms traditional models in accurately estimating N%.

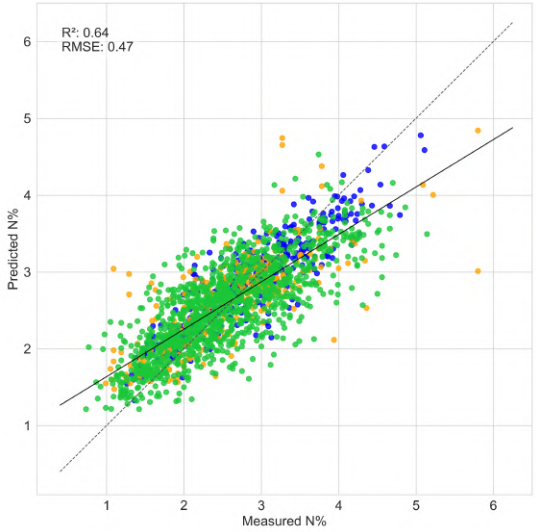
Spectral bands and RTM variables **Only Radiative Transfer Model (RTM) variables**



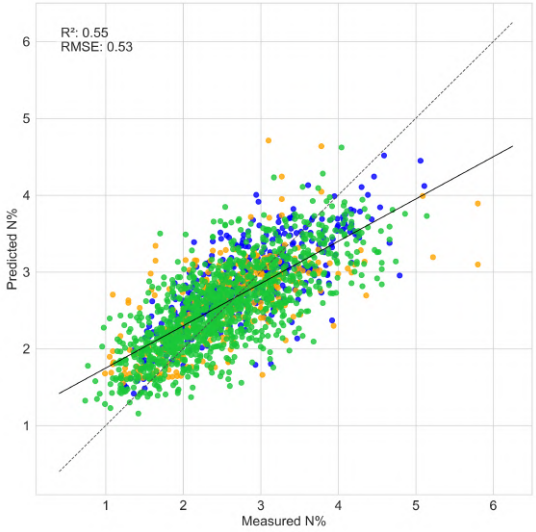
(e) GPR



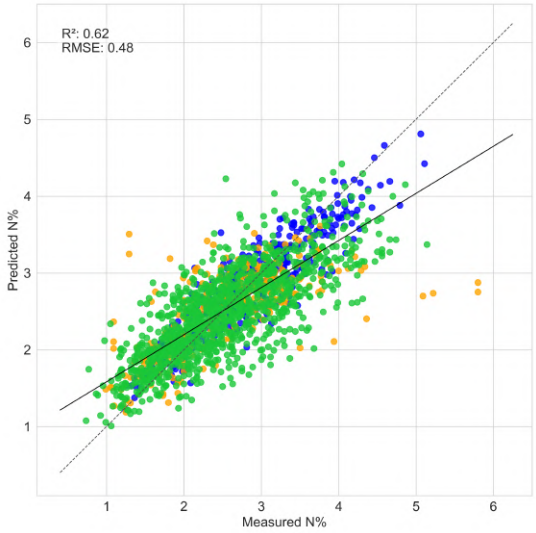
(f) GPR



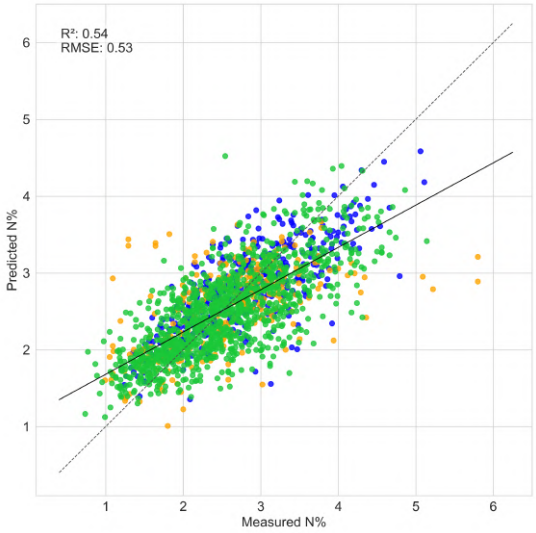
(g) RF



(h) RF



(i) SVR



(j) SVR

Figure 4.6: Continued.

Table 4.5: Comparison of R^2 and Root Mean Square Error (RMSE) values across various methods: Physics Informed Neural Network (PINN), Partial Least Squares Regression (PLSR), Gaussian Process Regression (GPR), Random Forest Regression (RF), and Support Vector Regression (SVR).

Category	**Method**	R^2	**RMSE (N%)**
Dry/ground scale	PINN (RTM variables)	0.61	0.47
	PINN (RTM variables and spectral bands)	0.90	0.24
	PLSR (RTM variables)	0.55	0.50
	PLSR (RTM variables and spectral bands)	0.78	0.35
	GPR (RTM variables)	0.61	0.46
	GPR (RTM variables and spectral bands)	0.89	0.23
	RF (RTM variables)	0.58	0.48
	RF (RTM variables and spectral bands)	0.79	0.34
	SVR (RTM variables)	0.60	0.47
	SVR (RTM variables and spectral bands)	0.83	0.30
Leaf scale	PINN (RTM variables)	0.44	0.59
	PINN (RTM variables and spectral bands)	0.67	0.45
	PLSR (RTM variables)	0.40	0.74
	PLSR (RTM variables and spectral bands)	0.50	0.60
	GPR (RTM variables)	0.4	0.60
	GPR (RTM variables and spectral bands)	0.51	0.55
	RF (RTM variables)	0.43	0.59
	RF (RTM variables and spectral bands)	0.57	0.50
	SVR (RTM variables)	0.37	0.64
	SVR (RTM variables and spectral bands)	0.44	0.57
Canopy scale	PINN (RTM variables)	0.57	0.51
	PINN (RTM variables and spectral bands)	0.65	0.46
	PLSR (RTM variables)	0.50	0.58
	PLSR (RTM variables and spectral bands)	0.55	0.52
	GPR (RTM variables)	0.57	0.50
	GPR (RTM variables and spectral bands)	0.63	0.47
	RF (RTM variables)	0.54	0.52
	RF (RTM variables and spectral bands)	0.59	0.49
	SVR (RTM variables)	0.55	0.51
	SVR (RTM variables and spectral bands)	0.58	0.50
Combined	PINN (RTM variables)	0.57	0.51
	PINN (RTM variables and spectral bands)	0.71	0.42
	PLSR (RTM variables)	0.43	0.61
	PLSR (RTM variables and spectral bands)	0.58	0.51
	GPR (RTM variables)	0.55	0.52
	GPR (RTM variables and spectral bands)	0.68	0.44
	RF (RTM variables)	0.55	0.53
	RF (RTM variables and spectral bands)	0.64	0.47
	SVR (RTM variables)	0.54	0.53
	SVR (RTM variables and spectral bands)	0.62	0.48

In our study, we conducted a comprehensive analysis of PINN, PLSR, GPR, RF, and SVR performances, utilizing RTM variables and spectral bands across various scales (Figure 4.6 Table 4.5. At the dry/ground scale, the PINN method utilizing solely RTM variables demonstrated a moderate correlation, with an R^2 of 0.61 and RMSE of 0.47 N%. This performance was significantly enhanced upon the inclusion of spectral bands, achieving an R^2 of 0.90 and RMSE of 0.24 N%. Notably, GPR closely matched PINN's performance with spectral bands, achieving an R^2 of 0.89 and RMSE of 0.23 N%, indicating its strong predictive capability in this context. Other methods like RF and SVR also showed improvements with spectral integration, with RF achieving an R^2 of 0.79 and RMSE of 0.34 N%, and SVR reaching an R^2 of 0.83 and RMSE of 0.30 N%.

At the leaf scale, PINN models using RTM variables alone indicated moderate efficacy (R^2 = 0.44, RMSE = 0.59 N%), which improved with the addition of spectral bands (R^2= 0.67, RMSE = 0.45 N%). The performance of PLSR, GPR, RF, and SVR was generally lower in this scale compared to PINN, with PLSR showing the weakest performance with RTM variables alone (R^2 = 0.40, RMSE = 0.74 N%). The canopy scale analysis revealed moderate correlations for PINN models both with RTM alone (R^2 = 0.57, RMSE = 0.51 N%) and when combined with spectral bands (R^2 = 0.65, RMSE = 0.46 N%). Similar to the leaf scale, GPR, RF, and SVR showed moderate performances, with GPR reaching an R^2 of 0.63 and RMSE of 0.47 N% when combining RTM variables with spectral bands, highlighting the competitive nature of these methods at larger scales. Across all scales combined, the PINN model integrating RTM variables with spectral bands outperformed its single-variable counterpart, achieving an R^2 of 0.71 and RMSE of 0.42 N%. This comparison underscores the enhanced predictive capability of PINN with spectral integration. This pattern was consistent but varied in magnitude among the other methods (PLSR, GPR, RF, and SVR), reflecting the distinct advantages and limitations of each method in leveraging RTM variables and spectral bands for regression analysis.

4.4.4 Performance of the PINN in the presence of noise

This assessed PINN performance and the equation's behavior under different noise conditions, offering comprehensive applicability insights. Leveraging equation 1 as a physical loss function improved PINN accuracy (R^2 = 0.66, RMSE = 0.46 N%), compared to lower performance without it (R^2 = 0.40, RMSE = 0.61 N%) (Table 4.6 and Figure 4.7). Utilization of the physical loss function in the PINN reduced uncertainty, minimized bias, and enhanced the depiction of input spectral data relationships with N%. Testing across scales—dry/ground, leaf, and canopy—revealed consistent improvements with the physical loss function. Canopy scale model performance notably increased (R^2 = 0.62, RMSE = 0.48 N% with physical loss; R^2 = 0.41, RMSE = 0.59 N% without). Similar enhancements were observed at leaf (R^2 = 0.50, RMSE = 0.56 N% with; R^2 = 0.30, RMSE = 0.65 N% without) and dry/ground scales (R^2 = 0.85, RMSE = 0.29 N% with; R^2 = 0.48, RMSE = 0.54 N% without). To enhance the validity of our results amidst the presence of noise and uncertainty in the data, we undertook a comparative analysis. This analysis juxtaposed the performance of PINN, characterized by a physical loss function, with that of traditional machine learning algorithms, namely PLSR, RF, SVR, and GPR. PINN,

distinguished by its explicit incorporation of underlying physical principles, exhibited superior R^2 values, underscoring a more robust model fit than its conventional counterparts. PLSR and SVR have an R^2 of 0.57 and an RMSE of 0.51, implying moderate performance. Lastly, RF and GPR have slightly lower R^2 values of 0.52 and 0.53, respectively, with GPR having an RMSE of 0.53, indicating a compromise between fit quality and predictive error. This finding underscores the potential of integrating physical knowledge to markedly enhance predictive accuracy (Table 4.6 and Figure 9.2).

Table 4.6: The comparison of the coefficient of determination (R^2) and Root Mean Square Error (RMSE) values for the Physics Informed Neural Network (PINN) utilizing RTM variables alongside critical spectral bands, both with and without the inclusion of physical loss, is examined alongside the performance of Gaussian Process Regression (GPR), Random Forest Regression (RF), and Partial Least Squares Regression (PLSR) under noisy conditions at various scales (dry/ground, leaf, canopy, and a combined).

Category	**Method**	R^2	**RMSE (N%)**
Dry/ground scale	PINN (with physical loss)	0.85	0.29
	PINN (without physical loss)	0.48	0.54
	PLSR	0.71	0.41
	RF	0.43	0.56
	SVR	0.61	0.47
	GPR	0.69	0.42
Leaf scale	PINN (with physical loss)	0.50	0.56
	PINN (without physical loss)	0.30	0.65
	PLSR	0.31	0.66
	RF	0.38	0.62
	SVR	0.40	0.61
	GPR	0.41	0.65
Canopy scale	PINN (with physical loss)	0.62	0.48
	PINN (without physical loss)	0.41	0.59
	PLSR	0.55	0.52
	RF	0.47	0.57
	SVR	0.57	0.51
	GPR	0.58	0.50
Combined	PINN (with physical loss)	0.66	0.46
	PINN (without physical loss)	0.40	0.61
	PLSR	0.57	0.51
	RF	0.52	0.54
	SVR	0.57	0.51
	GPR	0.53	0.53

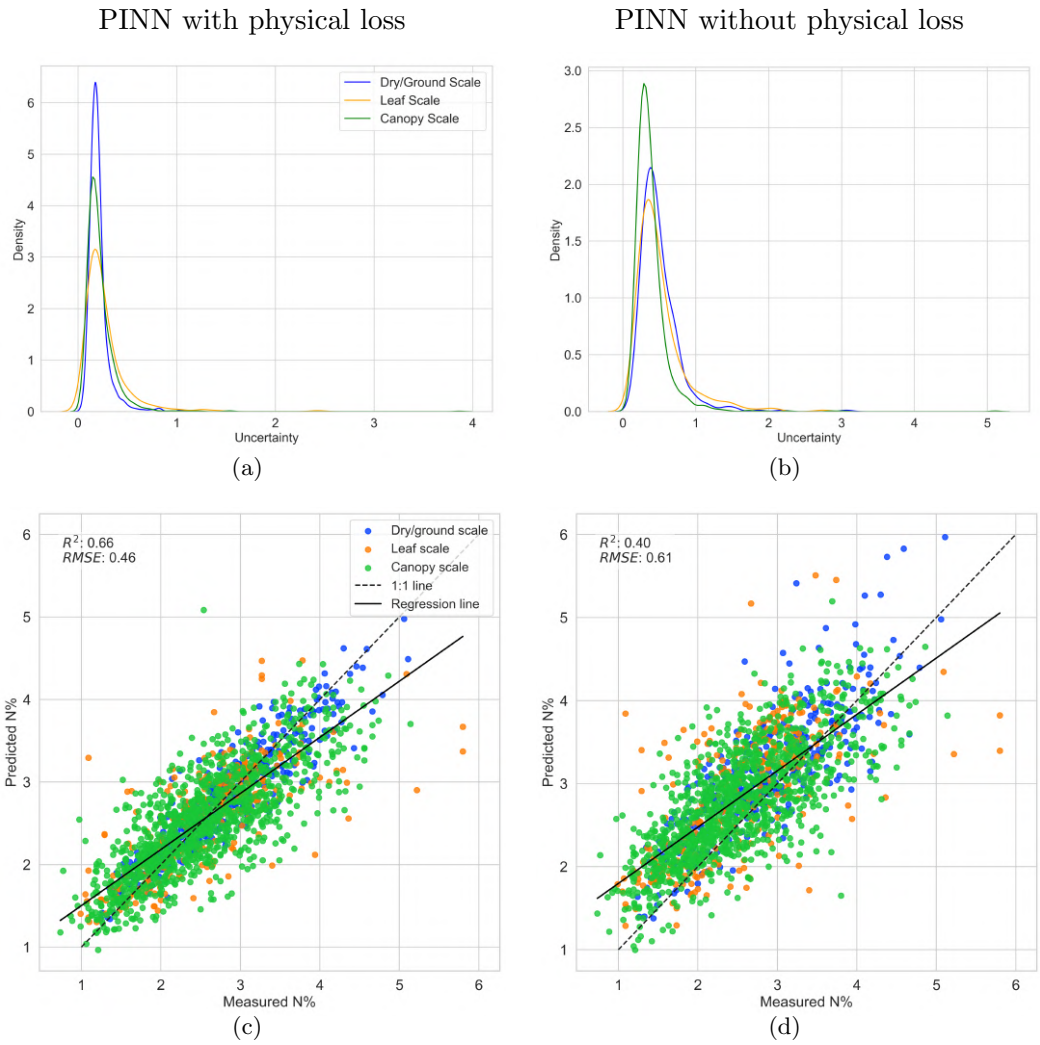


Figure 4.7: The efficacy of a physical loss function in improving the performance of the Physics-Informed Neural Network (PINN) model in the presence of noisy data is demonstrated in this study. Plots a and b illustrate the uncertainty of the estimations, with axis labels providing full definitions, such as nitrogen (N) concentration (%N). Plots c and d present a comparative analysis of the model's accuracy with and without the inclusion of heteroscedastic noise in the dataset, indicating that incorporating physical knowledge significantly improves accuracy and reduces bias. All abbreviations and units are defined to ensure the figure is self-contained and comprehensible.

4.5 Discussion

4.5.1 Inversion of RTM variables

The reliability of MCMC methods for inverting RTM variables from spectral data can be undermined by insufficient preliminary observations, which create data gaps, and by uninformative priors that are essential to the Bayesian framework but may lead to unstable posterior distributions. These issues are evident in our graphical uncertainty analysis across various scales—dry/ground, leaf, and canopy—presented in Figure 4.2. At the dry/ground scale (subplots a and d), uncertainty decreases with increasing wavelength, hinting at more consistent model predictions in the infrared spectrum due to less absorption by soil and ground components. Conversely, at the leaf scale (subplots b and e), uncertainty peaks at key photosynthetic absorption wavelengths, reflecting complex light interactions with leaf biochemistry. The canopy scale (subplots c and f) shows even greater uncertainty due to multiple light scattering and absorption, influenced by factors such as leaf orientation, canopy structure, and background reflectance, highlighting the intricacies of canopy modeling. These complexities illustrate the need for tailored strategies in remote sensing to reduce uncertainty at each scale. Additionally, MCMC's inherent uncertainties—attributable to its stochastic nature and potential Monte Carlo errors, especially in complex or limited models—can affect its utility for N management decisions. Alternative approaches like variational data assimilation and particle filters offer some resolution but face their challenges, such as handling non-Gaussian errors and the 'curse of dimensionality', which requires more computational resources with increasing model dimensions, as discussed by (Moral et al., 2006).

Our analysis further explores challenges in spectral data inversion, such as noise and complex biophysical structures, which increase uncertainty. We leverage prior knowledge to minimize these and focus on spectral bands that yield accurate RTM estimates (Dehghan-Shoar et al., 2023a). We find the highest accuracy in dry/ground scale reflectance data, attributing it to reduced complexity in vegetation's biophysical and interleaf characteristics. The dry/ground scale model outperforms canopy scale models, highlighting the significant impact of canopy structure on N% estimation and the effectiveness of controlled environments in mitigating canopy complexity's effect on N% estimation (Liu et al., 2023). Imposing constraints on biophysical and interleaf variables enhances inversion accuracy and reduces uncertainty (Jacquemoud and Baret, 1990). Including additional spectral bands has decreased uncertainty and improved performance, attributed to the unique absorption features across the spectral curve. However, an excessive number of spectral bands may lead to overfitting, suggesting future research should examine the impact of spectral band selection on model performance.

4.5.2 Selection of important spectral bands for N% estimation

Our study employed a GA for feature selection to identify key spectral bands crucial for N% estimation (Leardi et al., 1992). The bands predominantly identified fall within the 2000-2500 nm, 1400-2000 nm, and 410-780 nm ranges, consistent with previous research findings (Féret

et al., 2021, Pullanagari et al., 2012, Dehghan-Shoar et al., 2023b,a, Curran, 1989). However, we have noted minor discrepancies between the selected bands and the absorption peaks for fundamental N containing compounds as delineated in the PROSPECT-PRO model (Figure 4.4 b). Specifically, while the band at 2276 nm directly aligns with the N compound absorption peaks, reinforcing its selection (Féret et al., 2021), discrepancies in other bands, such as the 755 nm band's close association with the O2A absorption band, prompt further reflection (Mohammed et al., 2019). These discrepancies might stem from several factors. For instance, certain selected bands, though not directly overlaying the N absorption peaks, may exhibit the highest correlation with N% due to their indirect influence on photosynthetic efficiency and plant health, as evidenced by their sensitivity to chlorophyll content and photosynthetic activity. This is particularly true for the 734 nm band within the red-edge region, which, despite its theoretical detachment from direct N% estimation, provides significant insights due to N's impact on photosynthesis. It's also conceivable that the optimal selection of bands results from a combination of factors, including their direct and indirect relationships with N content, the complex interplay of spectral signatures, and plant physiological responses to varying N levels. These elements, coupled with the GA's capability to discern patterns within multi-dimensional datasets, might explain the preference for bands that do not strictly align with known N absorption peaks. Given these considerations, our findings underscore the need for further research to elucidate the underlying mechanisms governing the correlation between these spectral bands and N% estimation. This exploration could reveal novel insights into the spectral characteristics of vegetation and their intricate relationship with N availability, potentially leading to more refined approaches in remote sensing applications for nutrient management.

Our investigation into spectral characteristics, particularly within the critical ranges of 2000-2500 nm, 1400-2000 nm, and 410-780 nm, has yielded valuable insights for estimating N%. However, this study also highlights the complex challenges associated with varying scales of spectral data collection. The complexity of harmonizing spectral data is underscored by the differences in measurement approaches, such as the bidirectional reflectance from dry/ground samples versus directional-hemispherical reflectance in fresh leaf samples, as seen in the LOPEX93 dataset. Bridging the gap between the hemispherical-directional reflectance factor (HDRF) and hemispherical reflectance (HRF) facilitates accurate data integration and analysis. We developed Equation 1, focusing on selecting spectral bands after thoroughly examining HDRF effects across the spectrum. This process ensures these bands exhibit minimal HDRF influence, enhancing the equation's reliability amidst diverse measurement techniques.

4.5.3 Performance of the physical loss function

Our study implements the GP-GOMEA algorithm to develop a physical loss function (equation 1) for calibrating the PINN, focusing on spectral bands at 569 nm, 734 nm, 755 nm, 2077 nm, 2177 nm, 1559 nm, and 467 nm due to their direct correlation with N% estimation (Figure 4.4 a). This approach necessitates meticulous spectral band selection to balance the risks of overfitting and underfitting, given the symbolic expression's complexity generated by the algorithm,

which might pose interpretation challenges.

We highlight the complexity of non-linear interactions between these spectral bands in achieving accurate N% predictions. This perspective challenges the traditional reliance on linear models, such as PLSR, GPR, RF and SVR, underscoring the need for models that capture complex dynamics, as linear models often oversimplify the nuanced realities of spectral data analysis (Figure 4.5). Notably, bands at 1526 nm and 2243 nm show lower reflectance correlating with higher N%, emphasizing the importance of a diverse spectral range to reduce prediction noise and uncertainty.

In the context of our study, these high SHAP values underscore the particularly influential nature of these pairs in estimating N%. This observation illuminates the interplay between different spectral bands and their cumulative impact on the model's predictive capability, adding another layer of complexity and understanding to our predictive framework.

The complexity of interpreting the algorithm's symbolic expression indicates the need for advancements in transparent and effective feature selection and regression techniques. Scientific computing involves using advanced computing capabilities to solve complex scientific problems. It integrates mathematics, physics, and computer science to simulate and analyze natural and engineered systems. This field relies on numerical methods, algorithms, and computational tools to perform experiments in a virtual environment, allowing scientists to understand phenomena that are difficult, dangerous, or impossible to study physically. Scientific computing enables precise predictions, optimizations, and the exploration of scenarios across various disciplines, from climate modeling to bioinformatics, enhancing our understanding of the world.

Scientific Machine Learning (SciML) represents a cutting-edge synthesis of machine learning and traditional scientific disciplines, aimed at improving predictive models by incorporating physical laws into algorithms such as Physics-Informed Neural Networks (PINNs). This inclusion of physical loss functions, as noted by (Camps-Valls et al., 2023), enhances the accuracy and interpretability of the models. It reflects a paradigm shift towards models that are not just data-driven but also rooted in the principles governing natural phenomena. The differential performance of PINNs with and without physical constraints emphasizes the significant benefits of domain-specific knowledge in boosting prediction accuracy for N estimation—a critical factor in resource management for sectors like agriculture and forestry. Furthermore, the principles underpinning SciML have broader implications for spectral data analysis, offering potential advancements in the detection and quantification of elements such as arsenic and antimony, which are correlated with N levels, as reported by (Chakraborty et al., 2022, Rodriguez-Gomez et al., 2023).

4.5.4 Role of PINN in an improved N%

We introduce a novel PINN that enables the integration of domain knowledge in deep neural networks for monitoring N% in vegetation. This approach, grounded in scientific principles (e.g., protein absorption), can significantly improve model performance and mark a new era in vegetation spectroscopy research (Ballschmiter and Katz, 1969). Our study demonstrates that PINNs can outperform traditional models like PLSR, GPR, RF, and SVR by incorporating system physics knowledge, a first in this field (Figure 4.6 and Table 4.5), echoing the success of the PINN with multiscale spectral data (Wan et al., 2022).

Integrating a physical loss function into our model leads to notable performance improvements, even with noisy data, by reducing prediction bias and uncertainty (Figure 4.7 and Table 4.6). These results are consistent with the findings of Buster et al. (2022), which demonstrate that integrating physical principles with deep learning provides effective regularization and improves model generalizability. However, the applicability of PINNs is contingent upon the availability of a relevant physical loss function. Hence, their estimation relies on indirect measures through other biochemical and biophysical proxies, like the association between phosphorus content and SWIR spectra, mediated by protein absorption bands (Chakraborty et al., 2022, Rodriguez-Gomez et al., 2023). This highlights the potential limitations of PINNs in contexts where direct physical relationships are not well-defined.

4.6 Conclusion

This study leveraged a comprehensive multiscale spectral dataset to develop a novel PINN model for predicting N%, integrating spectral and radiative transfer model (RTM) attributes with a custom physical loss function. This model identified critical spectral bands (2276 nm, 755 nm, 1526 nm, 2243 nm, and 734 nm) strongly correlated with N%, essential for accurate estimation. It exhibited outstanding performance across different scales, with a notable performance decrease from dry/ground to canopy scales due to canopy complexity and variable water content. Specifically, at the dry/ground scale, the model achieved an R^2 of 0.90 and an RMSE of 0.24 %N, while at canopy and leaf scales, it recorded R^2 values of 0.65 and 0.67 and RMSEs of 0.46 %N and 0.45 %N, respectively, resulting in an overall R^2 of 0.71 and an RMSE of 0.42 %N, outperforming models using fewer spectral bands.

Future research aims to enhance model accuracy by incorporating aerial and spaceborne remote sensing data, examining the impacts of topography and atmospheric conditions on N% predictions. This approach could improve robustness and real-world applicability. Our model's success over the PLSR model, by capturing non-linear relationships between spectral bands and RTM variables, offers insights for refining N% estimation. Accurate N% estimation has potential implications for advancing sustainable land-use and vegetation management practices.

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Mohammad Hossain Dehghan Shoar
Name/title of Primary Supervisor:	Associate Professor Gabor Kereszturi
In which chapter is the manuscript /published work:	5

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

Dehghan-Shoar, M. H., Pullanagari, R. R., Orsi, A. A., Yule, I. J. (2023). Simulating spaceborne imaging to retrieve grassland nitrogen concentration. Remote Sensing Applications: Society and Environment, 29, 100912.

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:
- The percentage of the manuscript/published work that was contributed by the candidate:
- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	
Date:	03-Sep-2024
Primary Supervisor's Signature:	Gabor Kereszturi Digitally signed by Gabor Kereszturi Date: 2024.09.03 22:06:52 +12'00'
Date:	3-Sep-2024

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

Chapter 5

Enhancing Grassland Nitrogen Monitoring through Simulated Spaceborne Sensor Data and Hybrid Modeling

5.1 Abstract

Spaceborne optical imaging enables continuous nitrogen (N) concentration (N%) monitoring in grasslands. However, the differences in instrumental setup, image pre-processing, wavelength coverage, and sampling rate pose a challenge when leveraging data from multiple spaceborne instruments. We develop a method that uses field spectroscopy and the Soil-Plant-Atmosphere Radiative Transfer (SPART) model to simulate any Top Of Atmosphere (TOA) spaceborne sensor. The purpose of this method is to allow better integration of field spectroscopy data with spaceborne optical imagery. We develop a hybrid model using simulated data and a Random Forest Regressor, which is then independently validated using real Sentinel-2 TOA reflectance collected during 2016 and 2020. Our model achieves an independent validation accuracy of 0.44 (RMSE), MPIW of 0.11, and R^2 of 0.55, demonstrating the potential of this methodology for monitoring grassland N% using field spectroscopy.

5.2 Introduction

Grasslands are the backbone of the global meat and dairy sectors and thus play a critical role in assuring food security. Globally, grasslands cover approximately 40% of the land surface (Wang et al., 2019a), and contribute positively to human well-being in many ways, including as a source of carbon sequestration (Tiscornia et al., 2019, O'Mara, 2012, Yan et al., 2019). Monitoring grasslands helps meat and dairy farmers increase productivity and react to climate change. Nitrogen (N) is a fundamental nutrient and driver of grassland productivity (Guo et al., 2016). Farmers and managers apply N in the form of chemical fertilizer to increase productivity.

However, its overuse is costly and can hurt the environment, causing ecological and environmental damage (Ahmed et al., 2017). Therefore, it is critical to monitor the N concentration (N%) of grasslands to optimize the application of chemical N fertilizers.

Traditionally, monitoring grassland N% is undertaken using field campaigns, collecting grass samples, and measuring the N% of each sample in a laboratory setting in the form of wet chemistry. Although this approach provides reliable and accurate information about grassland characteristics, this method is time-consuming and costly. By reducing costs associated with labor and laboratory material, field spectroscopy provides a more time and cost-efficient alternative for monitoring grassland N% (Pullanagari et al., 2021, Dehghan-Shoar et al., 2023a). However, both methods are inadequate for continuous monitoring at large spatial scales. Spaceborne Multi-spectral Imagery (MSI) has become a convenient alternative for monitoring fundamental vegetation characteristics globally (Estévez et al., 2022). Although MSI's spectral resolution and quality are lower than proximal spectroradiometers, MSI has been shown to successfully monitor grassland N% dynamics (Fernández-Habas et al., 2021, Gao et al., 2020b).

Two well-known spaceborne MSI instruments are the twin constellation of Sentinel-2A and 2B (S2). These instruments have a revisit time of up to 2–3 days in mid-latitudes and have a Ground Sampling Distance (GSD) of 10, 20, and 60 m. In addition, they cover a large spectral region between 400 and 2500 nm with 13 spectral bands, which are suitable for monitoring N% (Frampton et al., 2013, Szantoi and Strobl, 2019, Berger et al., 2020a). However, there are challenges associated with using spaceborne instruments for monitoring N%. In addition to differences in sensor characteristics such as spectral bandpasses and revisit dates, the requirement for atmospheric correction remains a key challenge. Spaceborne instruments can measure the ratio of electromagnetic radiation flux reflected to the incident solar radiation of a surface, known as Top of Atmosphere (TOA) reflectance. Although it is possible to use TOA reflectance for monitoring vegetation characteristics, atmospheric variables can introduce significant uncertainties in TOA reflectance. Therefore, it is important to develop reliable and accurate methods that can influence vegetation variable retrieval that is less sensitive to instrumental setup, wavelength coverage, and sampling rate and has less uncertainty.

5.3 Material and Methods

5.3.1 Environmental conditions

We conducted field campaigns during 2016-2020 on multiple dairy and hill country farms across New Zealand. The farms had different grazing patterns, with grass growing continuously during the year due to optimal weather conditions. The weather conditions included a mean annual precipitation of 600 mm and 1600 mm and a mean annual temperature of 10 ° C to 16 ° C. In addition, the farms had high levels of variability in soil characteristics such as pH, phosphorus, and potassium levels. Detailed descriptions regarding the field campaigns and dataset used in this work are stated by Pullanagari et al. (2021) and Dehghan-Shoar et al. (2023a).

5.3.2 Spectral measurements

The Top of Canopy (TOC) reflectance of grassland was captured using the field spectroradiometers ASD Field-Spec 4 Hi-Res (Malvern Panalytical Ltd., Malvern, United Kingdom) and ASD FieldSpecPro (Analytical Spectra Devices, Inc., Boulder, CO, USA). The TOC reflectance measures the fraction of incoming solar radiation that is reflected by the surface of the Earth, in this case, grassland. These instruments have a spectral range of 350 to 2500 nm and a 512-element silicon photodiode providing a spectral resolution of 3 nm and sampling interval of 1.4 nm in the Visible Near Infrared (VNIR) region (<1000 nm).

Furthermore, the sensors consisted of two sets of thermoelectric (TE)-cooled indium gallium arsenide (InGaAs) photodiode arrays which help provide a spectral resolution of 10–12 nm and 8 nm with a sampling interval of 1 nm in the Shortwave Infrared (SWIR) region (>1000 nm)(Pullanagari et al., 2021, Danner et al., 2015). We captured the TOC reflectance of different grass species on areas with an estimated Leaf Area Index (LAI) of 1-3. This was undertaken to minimize the impact of the soil background on the collected spectra, as illustrated by Pullanagari et al. (2021).

We collected samples from different geographic locations using a random stratified sampling technique. Slope, aspect, and soil types were used for stratification. Categorization was undertaken using slope values of 0–8, 8–16, 16–25, and 25 degrees calculated from the open-source Digital Elevation Model (DEM) accessible from Land Information in New Zealand (LINZ). A quadrant with a size of $0.5\,m^2$ was located on the grassland at each sampling location. The grass within the quadrant was harvested using an electric shearing machine and stored in a cooling storage box (Hutchinson et al., 2016, Cosgrove et al., 1998, Pullanagari et al., 2012).

All sampling locations were geo-referenced using an RTK-GPS system. Finally, the collected grass samples were sent to Analytical Research Laboratories (ARL), Napier, New Zealand, to determine the N% of each sample using the Kjeldahl method (Kirk, 1950, Lynch and Barbano, 1999). This process consisted of drying, grounding, and homogenizing each sample using a mixer mill (Frommer et al., 2011).

Furthermore, we collected S2 TOA reflectance data from the Google Earth Engine (GEE) platform for five farms where field campaigns were conducted (Gorelick et al., 2017). We used imagery captured within ten days, during which imagery was unavailable on the same sampling day. This was undertaken as grassland N% does not change significantly within this period. Furthermore, we spatially re-sampled the S2 TOA data to 10-meter (m) spatial resolution for this analysis. Finally, datasets were created by extracting the spectra for regions of interest using the geo-referenced locations collected during the field campaigns (Table 6.1). The areas contaminated with shadows and other surfaces (e.g., artificial surfaces) were excluded from the analysis.

Table 5.1: Table containing the dates and locations where the field campaigns were undertaken, and Sentinel-2 (S2) Top of Atmosphere (TOA) imagery is acquired.

Latitude	Longitude	Field sampling date	Image acquisition date
-40.7	175.8	18/04/2016	26/04/2016
-39.3	174.3	22/10/2018	21/09/2018
-43.9	171.5	4/04/2019	30/03/2019
-40.1	175.2	28/11/2019	5/12/2019
-40.1	175.2	18/11/2020	11/11/2020

5.3.3 Modelling grassland N%

The collected spectral regions between 350–450 nm and 2450–2500 nm of the spectroradiometer data were removed from the analysis due to the significant amount of noise in the spectra. We minimized the noise in the remaining spectra by using a Savitzky–Golay filter with a smoothing window width of 13 (13 nm) and polynomial order of 3 (Schafer, 2011, Press and Teukolsky, 1990). We used the Spectral Response Function (SRF) of S2 made available by the European Space Agency (ESA) to simulate S2 TOC reflectance. We completed this process by calculating the sum of the products of SRF and field spectroradiometer data divided by the sum of SRF to measure the true bandpass. Figure 5.1 demonstrates the SRF of S2 spectral bands alongside a sample collected using the field spectroradiometer.

Following the simulation of S2 TOC reflectance using the field spectroradiometer, we inverted the biochemical and biophysical variables using the Soil-Plant-Atmosphere Radiative Transfer (SPART) model (Yang et al., 2020a). The SPART model incorporates three computationally efficient RTMs, including:

- The Brightness-Shape-Moisture (BSM) soil reflectance RTM (Verhoef et al., 2018)
- The PROSAIL canopy RTM (Jacquemoud and Baret, 1990, Verhoef, 1984)
- The Simplified Method for Atmospheric Correction (SMAC) RTM (Rahman and Dedieu, 1994)

We implemented a Markov Chain Monte Carlo (MCMC) numerical optimization algorithm alongside a loss function (total sum of error) to search for the best SPART variables (Foreman-Mackey et al., 2019). This process consisted of searching for SPART variables that minimize the difference between the simulated and actual S2 TOC reflectance within a given range of variable values (Table 6.2) (Simon Blanke, 2020). We used MCMC because it can retrieve the uncertainty associated with each inverted SPART variable. We set the soil variables, geographical location, viewing, illumination geometry, and Day Of Year (DOY) required by SPART to constant values. This enables the MCMC only to invert the atmospheric and vegetation variables. The MCMC algorithm used an ensemble sampler with 100 walkers, which ran for 10000 iterations for each sample. We performed thinning and calculated each inverted variable's mean and uncertainty (standard deviation).

Table 5.2: Search window used with the Markov Chain Monte Carlo (MCMC) and SPART for simulating Sentinel-2 (S2) Top of Canopy (TOC) reflectance from S2 Top of Atmosphere (TOA) reflectance.

SPART variable	variable ID	Search space	RTM	Units
Air pressure	Pa	900-1100	SMAC	hPa
Aerosol optical thickness	aot550	0-1	SMAC	-
Water vapour	uh2o	0-2.5	SMAC	g/cm^{-2}
Ozone content	uo3	0-0.4	SMAC	cm-atm
Structure variable	N	1.5-2.5	PROSAIL	-
Chlorophyll content	Cab	10-90	PROSAIL	micro g/cm^{-2}
Carotenoid content	Car	2-9	PROSAIL	micro g/cm^{-2}
Brown pigment content	Cs	0-0.1	PROSAIL	-
Equivalent water thickness	Cw	0-0.2	PROSAIL	cm
Dry matter content	Cdm	0-0.1	PROSAIL	g/cm^{-2}
Protein content	Cp	0-0.02	PROSAIL	g/cm^{-2}
Carbon constituents	CBS	0-0.02	PROSAIL	g/cm^{-2}
Anthocyanin content	ant	0-7	PROSAIL	$\mu g/cm^{-2}$
Leaf Area Index	LAI	1-4	PROSAIL	-
Leaf angle distribution a	LIDFa	-0.5-0.5	PROSAIL	degree
Leaf angle distribution b	LIDFb	-0.5-0.5	PROSAIL	degree
Soil brightness	B	0.5	BSM	-
Soil moisture percentage	SMp	50	BSM	percentage
Soil moisture carrying capacity of the soil	SMC	0.25	BSM	-
Single water film optical thickness	film	0.0150	BSM	cm

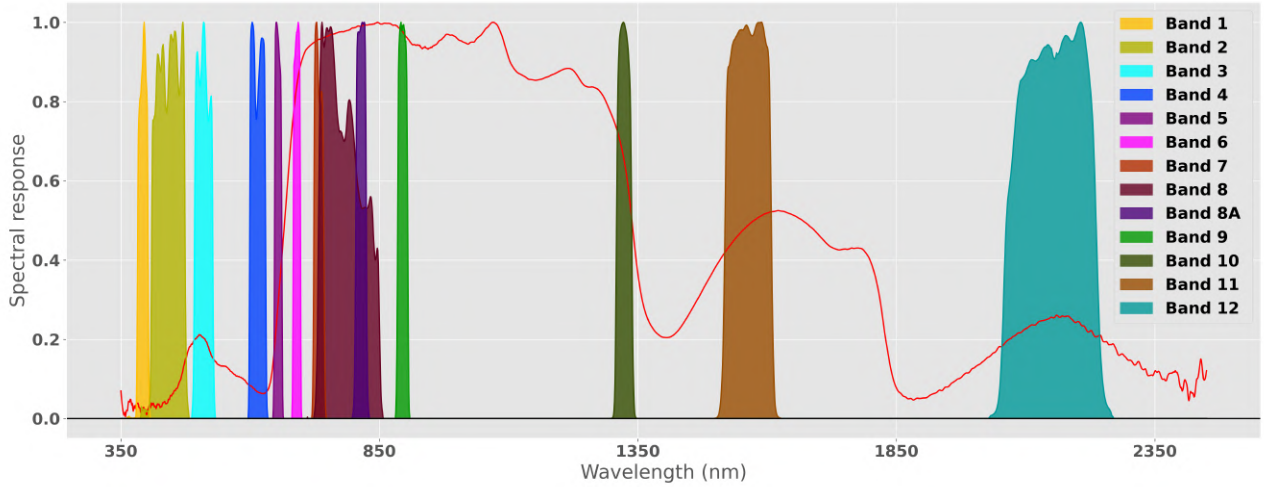


Figure 5.1: Sample spectroradiometer Top of Canopy (TOC) spectra collected during field campaigns alongside the Spectral Response Function (SRF) for Sentinel-2 (S2) spaceborne sensor.

We used the mean values for each inverted SPART variable to simulate S2 TOA spectra. We removed the coastal aerosol, water vapor, and SWIR-cirrus bands from the spectra. Furthermore, due to the high uncertainty associated with the SWIR region of the simulated S2 TOA spectra, only spectral bands between 500 and 1000 nm were used in the modeling process. We used the simulated S2 TOA spectra alongside a Random Forest Regressor (RFR) algorithm to develop a model for predicting grassland N%. We optimized the RFR variables using a random search algorithm, which identified 2550 as the optimal number of estimators and a value of 5 for the max depth of the RFR (Bergstra and Bengio, 2012). We calculated the uncertainty of the model by using the confidence intervals of the predicted values (Polimis et al., 2017).

We used Leave-One-Out (LOO) cross-validation using the calibration dataset to assess the performance of the RFR model and undertake optimization. In addition, we independently validated our model performance using actual S2 TOA data (Table 6.1) downloaded from GEE. The calibration dataset consisted of 1604 samples collected by the spectroradiometer, while the validation dataset consisted of 477 samples captured during different periods in different locations. The model performance was evaluated by using statistical metrics such as R^2, Root Mean Squared Error (RMSE), and Mean Prediction Interval Width (MPIW). The workflow of the methodology used in this work is illustrated in Figure 5.2.

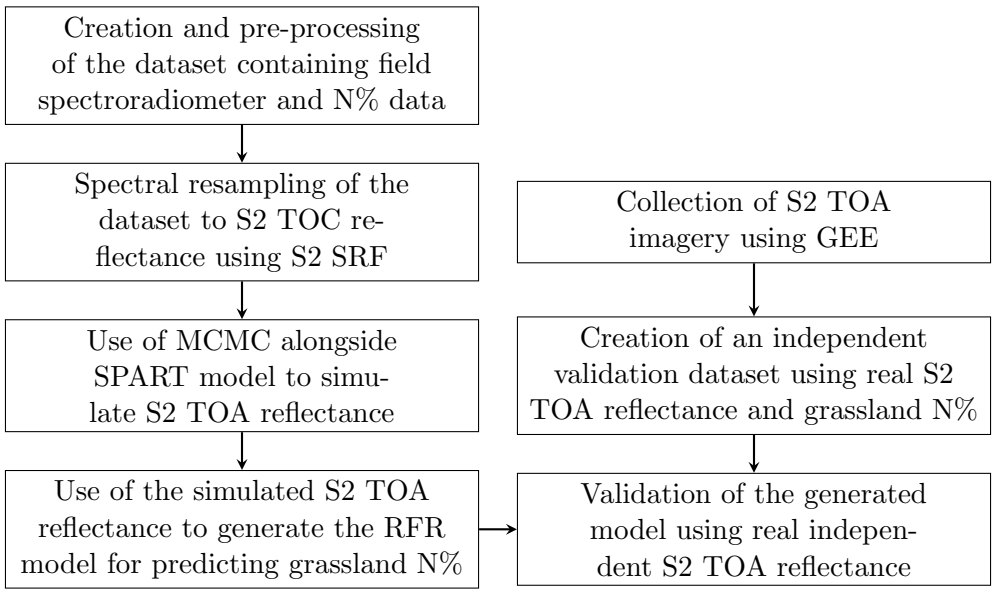


Figure 5.2: Figure containing the methodology to develop the Random Forest Regressor (RFR) model using simulated data.

5.4 Results

The methodology outlined in the previous section allows us to use field spectroscopy to constrain the biochemical and biophysical variables in the SPART model to simulate S2 reflectance. The confidence in the constraints of these variables, obtained from the model inversion, also propagates uncertainties in the simulated spectra. In the following, we analyze the impact of

uncertainties in the model variable inversion and how these affect the predicted N% obtained from simulated S2 reflectance.

5.4.1 The impact of variable uncertainties on the simulated S2 reflectance

Here, we investigate the impact of different SPART variables on simulated S2 TOA reflectance. Figure 5.3 illustrates that the atmospheric variables impact the NIR of the spectrum. Canopy biophysical variables (LAI and LIDFa) significantly impact the entire spectra, particularly in the NIR and SWIR.

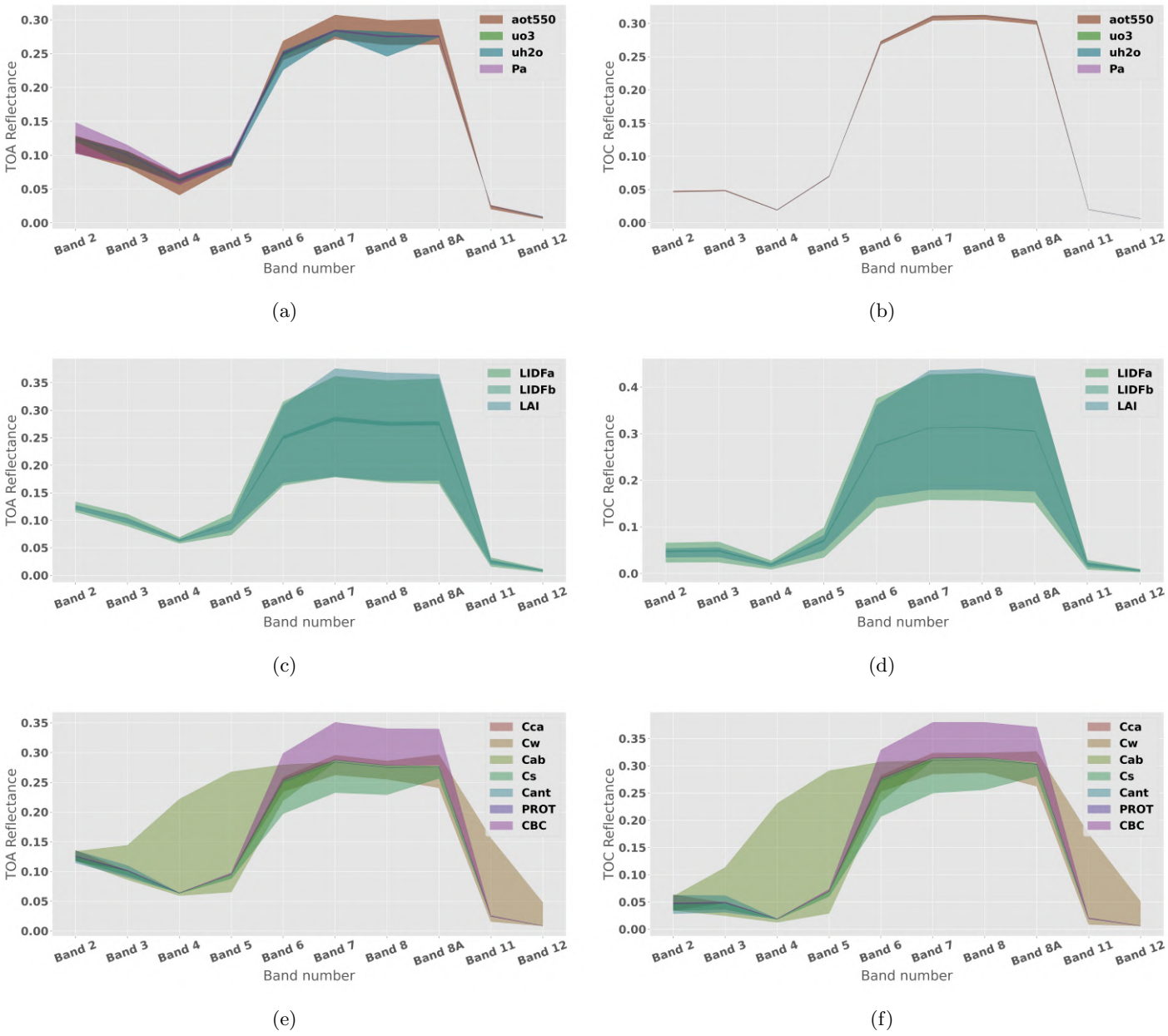


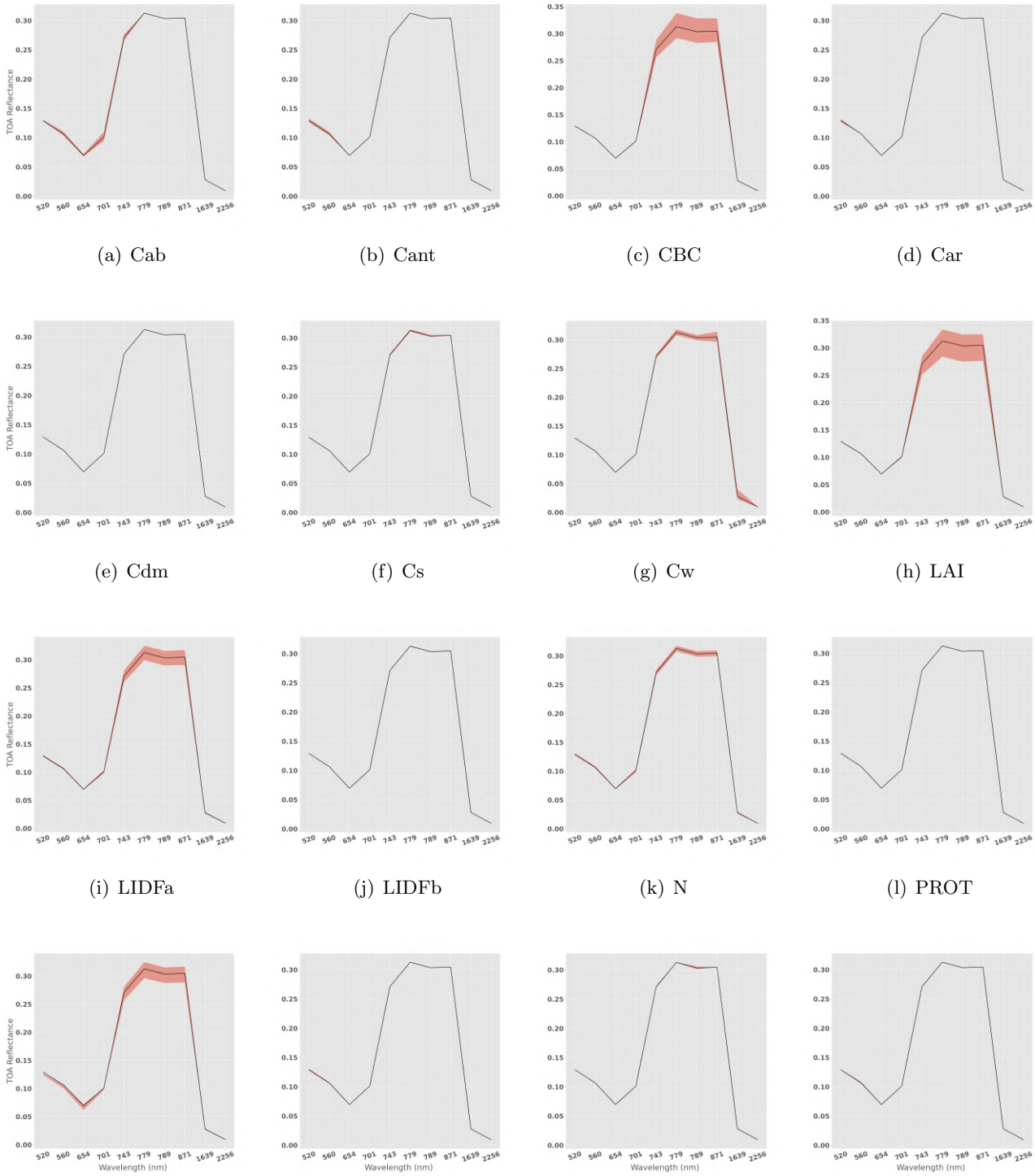
Figure 5.3: The impact of the atmospheric, biophysical, and biochemical SPART variables on Sentinel-2 (S2) Top of Atmosphere (TOA) and Top of Canopy (TOC) reflectance.

Among biochemical variables, Cab, CBC, Cs, and Cw strongly impact the spectra. Figure 5.3 shows that in biophysical and biochemical variables atmospheric correction, the influence from the atmosphere was significantly reduced (Figure 3b).

We have investigated the impact of the uncertainty associated with each inverted SPART variable on the simulated S2 TOA spectra (Figure 5.4). As illustrated in Figure 5.4, the spectral region between 700 and 1000 nm contains significant uncertainty, which is due to some different variables, including LAI, aot550, LIDFa, and CBC. The uncertainty associated with this region can cause a significant decrease in model performance and stability, as Band 8A is the most critical spectral feature for monitoring grassland N% in our model.

Model performance can also be affected by the uncertainty caused by the different SPART variables. The uncertainty associated with the NIR of S2 TOA reflectance appears to come from aot550. Cw has a small impact on the SWIR region of the spectrum, which is consistent with Figure 5.3, where it is shown that there is no significant uncertainty associated with other inverted SPART variables. This can be because we are accurately inverting these variables or the overall impact of these variables on the spectra is limited (e.g., structure variable).

Figure 5.5 illustrates how the simulated S2 reflectance compares against real S2 reflectance. Uncertainties in the simulated spectra are computed from propagating the uncertainties in the RTM variables to the reflectance of each channel. The simulated spectra show limited agreement with the real S2 TOA reflectance. The visible range displays smaller uncertainties compared to the NIR region. The simulated reflectance in the NIR region is typically higher than that in the real spectra. However, this is compensated by larger model uncertainty.



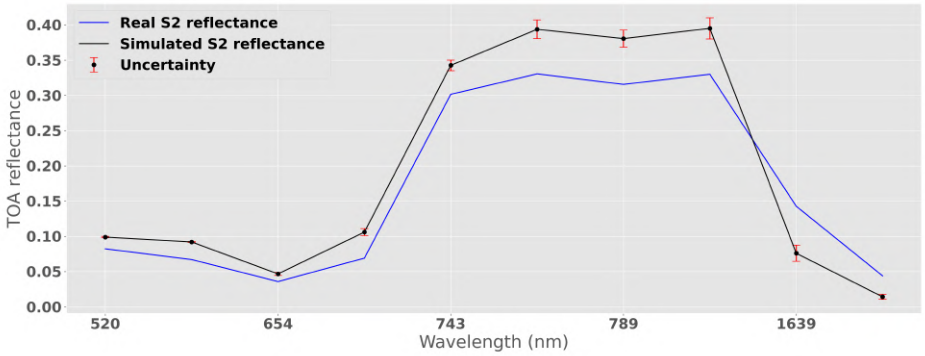
(a) Cab (b) Cant (c) CBC (d) Car

(e) Cdm (f) Cs (g) Cw (h) LAI

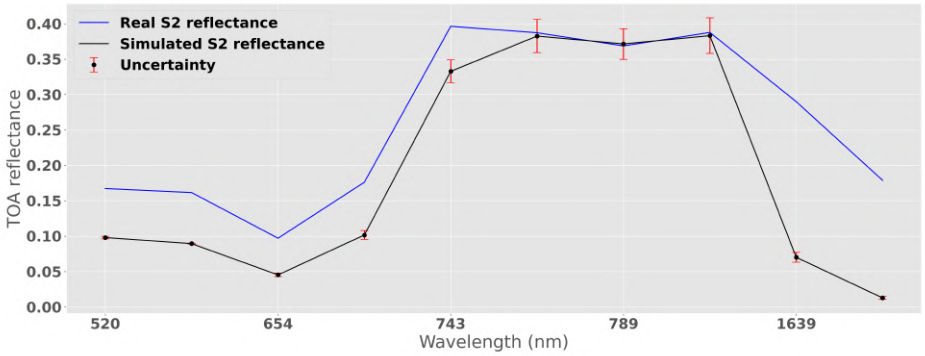
(i) LIDFa (j) LIDFb (k) N (l) PROT

(m) aot550 (n) Pa (o) uh2o (p) uo3

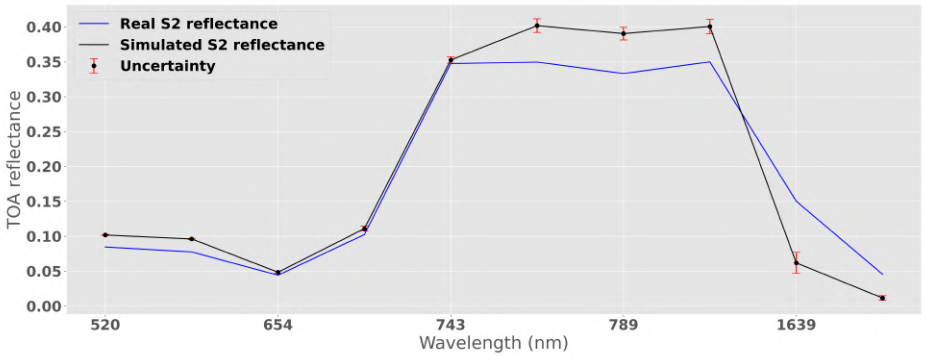
Figure 5.4: The impact of uncertainty in SPART variables on simulated Sentinel-2 (S2) Top of Atmosphere (TOA) reflectance. Subplots (a–p) show the influence of individual variables: (a) Cab (chlorophyll content), (b) ant (anthocyanin content), (c) Car (carotenoid content), (d) Cs (brown pigment content), (e) Cw (equivalent water thickness), (f) Cdm (dry matter content), (g) Cp (protein content), (h) LAI (Leaf Area Index), (i) LIDFa (leaf angle distribution parameter a), (j) LIDFb (leaf angle distribution parameter b), (k) Nm (structure variable), (l) CBS (carbon based constituents), (m) aot550 (aerosol optical thickness), (n) Pa (air pressure), (o) uh2o (water vapor), and (p) uo3 (ozone content). The red-shaded regions represent the uncertainty range, indicating the sensitivity of TOA reflectance to variability in these parameters.



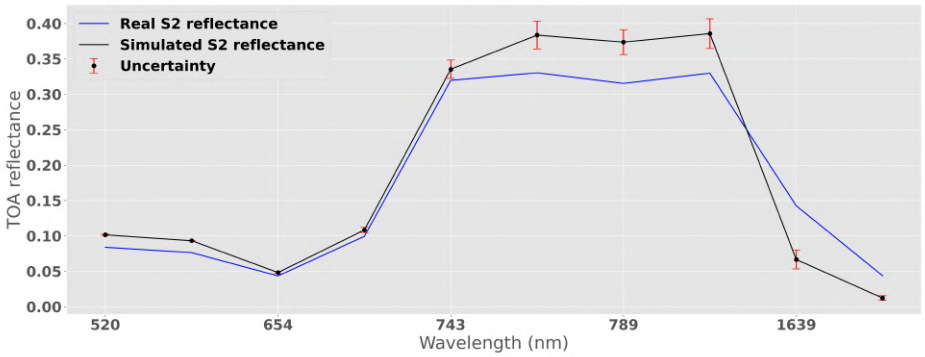
(a)



(b)

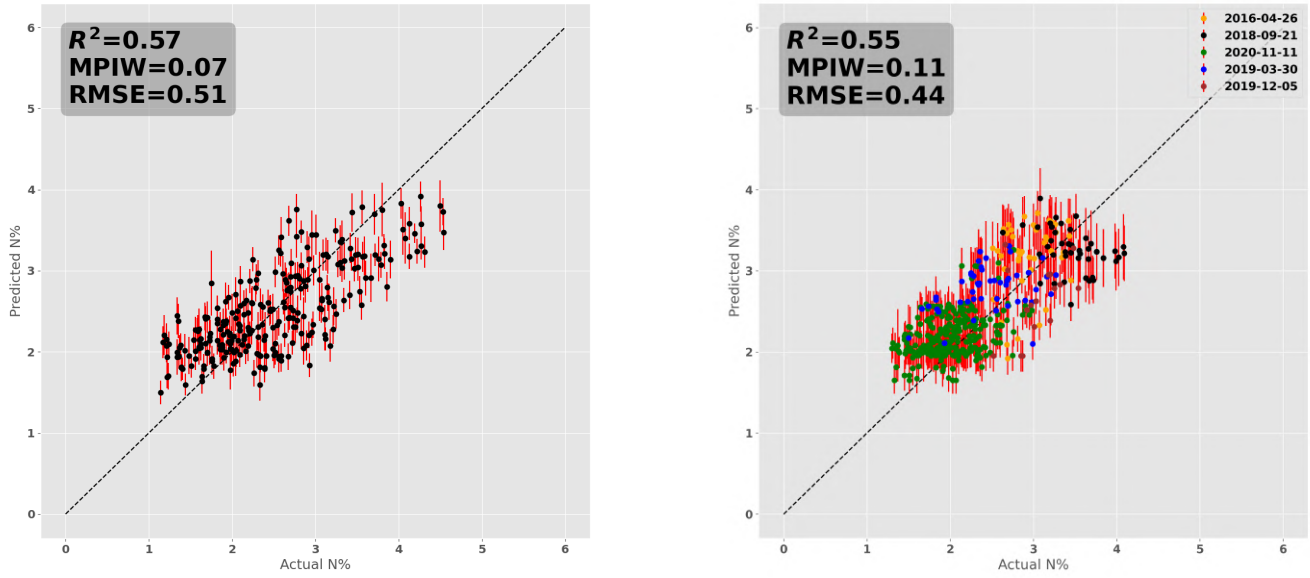


(c)



(d)

Figure 5.5: Comparison of real (blue) and simulated (black) Sentinel-2 (S2) Top of Atmosphere (TOA) reflectance for four samples, with red vertical lines indicating uncertainty in the simulated spectra. Subplot (a) shows the first sample with moderate uncertainty across wavelengths; (b) highlights slightly higher uncertainty near the peaks; (c) depicts minor deviations between real and simulated reflectance; and (d) demonstrates strong agreement with minimal uncertainty.



(a) (b)

Figure 5.6: (a) Predicted versus actual grassland N% using simulated Sentinel-2 (S2) Top of Atmosphere (TOA) reflectance, with $R^2 = 0.57$, Mean Prediction Index Width (MPIW) = 0.07, and Root Mean Squared Error (RMSE) = 0.51. (b) Predicted versus actual grassland nitrogen (N) concentration (N%) using a model calibrated on simulated S2 TOA reflectance and validated with real S2 TOA reflectance, achieving $R^2 = 0.55$, MPIW = 0.11, and RMSE = 0.44. In (b), colors correspond to the acquisition dates of the real S2 TOA images. Vertical red lines indicate the uncertainty associated with each predicted sample, and the dashed diagonal line represents the 1:1 relationship between predicted and actual values.

5.4.2 Model performance for predicting grassland N%

We use independent validation data to measure the performance of our model predictions for grassland N% using simulated S2 TOA reflectance. Figure 5.6 shows that our model can predict N% with an average RMSE of 0.51.

We use real independent S2 TOA reflectance data to benchmark our model's performance to test our grassland N% model. As illustrated in Figure 5.6 the RMSE obtained with real data is 0.44, thus only slightly better than what is obtained with our simulated spectra. The uncertainties and slight systematic biases illustrated in Figure 5.5 have thus a small impact on the overall estimate of N%.

5.4.3 The relative importance of S2 bands

To investigate the impact of different spectral bands for predicting N%, we compute Shapley values using the python library SHAP (SHapley Additive exPlanations) (Lundberg et al., 2020). We generated Figure 5.7 to investigate the impact of each band on the model prediction. Band

8A and Bands 5, 2, 4, and 3 are the most important spectral bands. These bands likely contribute to predicting N% because they are also known to predict grassland Cab, which is the strongest indicator of N% other than Cp (Delloye et al., 2018).

We can see that higher Band 8a and Band 2 values contribute to an N% prediction, which is above the average value. However, higher values of Band 5 contribute to an N% prediction less than the average N% value. As this band is associated with Cab absorption, this behavior can be due to the uncertainty associated with the simulation of Band 5.

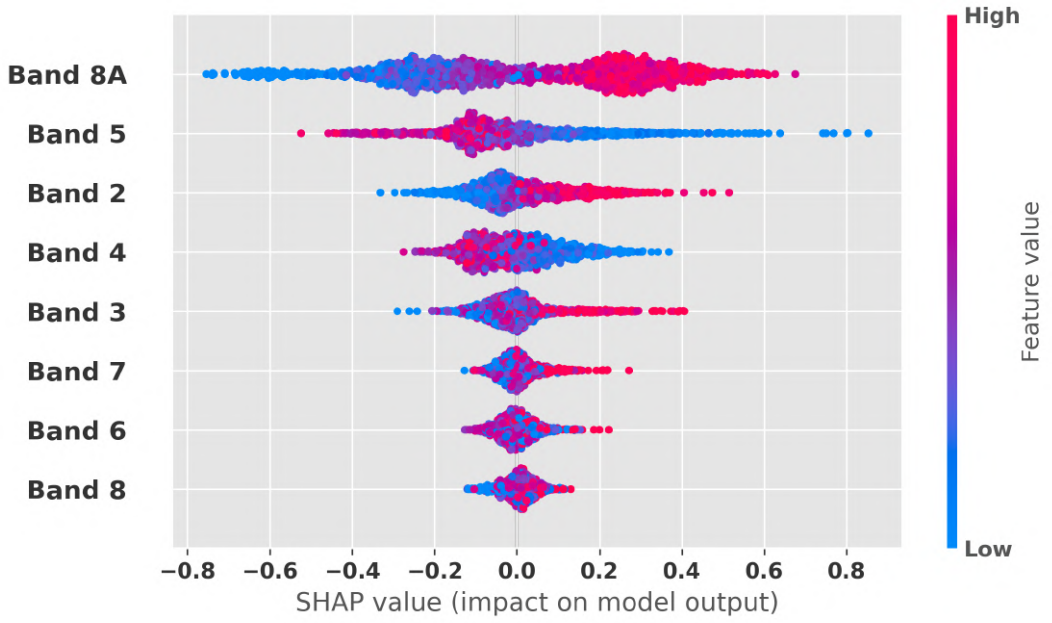


Figure 5.7: The relationship between Sentinel-2 (S2) Top of Atmosphere (TOA) spectral bands and grassland nitrogen (N) concentration (N%) using SHAP (SHapley Additive exPlanations) values. The x-axis shows SHAP values (impact on model output), with positive values increasing predicted N% and negative values decreasing it. The y-axis lists the spectral bands, while the color gradient (blue to red) represents low to high feature values. This visualization highlights the contribution and importance of each spectral band to the predicted N%.

5.5 Discussion

5.5.1 Model performance using real S2 TOA reflectance

In this work, we have developed a methodology for predicting grassland N% using a model calibrated on simulated S2 TOA reflectance from field spectroscopy data. Our model showed that it is possible to predict grassland N% during different seasons and locations (Figure 5.6). The results of our methodology demonstrated that it is possible to simulate S2 TOA reflectance using field spectroradiometer data and develop a model for predicting grassland N% using simulated S2 TOA reflectance alongside the SPART RTM (Figure 5.5). This work enables the retrieval of grassland N% without undertaking atmospheric correction.

Since the SPART model uses three different RTMs, it is critical to constrain as many variables as possible to improve the RTM inversion process (Yang et al., 2020a). Failure to constrain these variables will significantly impact the accuracy of the simulated spectra. Looking at Figures 5.3 and 5.4, we can see that the uncertainty associated with the different SPART variables can significantly impact the S2 TOA reflectance. Verrelst et al. (2019b) investigated the potential for retrieving the biophysical and biochemical variables from S2 TOA reflectance. In their work, it is suggested that the retrieval of Cab, Cw, and LAI from S2 TOA reflectance data is feasible.

Furthermore, Verrelst et al. (2019b) shows that the atmospheric variables have a limited impact on the SWIR region of the spectrum, which agrees with our finding (Figure 5.3). However, because of the impact of atmospheric variables on NIR, it is difficult to invert variables influenced by this region accurately (e.g., Cdm). As atmospheric conditions can vary significantly based on geographical location, a detailed investigation should be conducted to understand better the impact of different atmospheric variables on S2 TOA reflectance for monitoring N%.

As illustrated in Figure 5.3 we can see that the physical variables significantly impact the entire S2 reflectance. Therefore, a prerequisite of a high-performance RTM is the proper parametrization of the physical variables of an RTM. However, correct parametrization of the physical variables of RTMs remains a challenging issue, and further research should be undertaken to develop methods to improve RTM parametrization. One method of parameterizing RTM is by using tools including Light Detection and Ranging (LiDAR). LiDAR can be used to construct the structure of the canopy, thereby measuring the biophysical variables required to constrain RTM. Another way of parametrizing an RTM can be done by using Look Up Tables (LUT) alongside statistical machine learning algorithms such as Gaussian Process Regressors (GPR) to retrieve variable values and the uncertainty associated with these values (Sinha et al., 2020, Estévez et al., 2021, Weiss et al., 2000). We can use these approximate values to reduce the search space required for the inversion process. This will not only help address the ill-posedness of RTM but will also help decrease the uncertainty associated with the estimated values.

5.5.2 Impact of S2 TOA spectral bands on N% predictions

The most important band for predicting grassland N% is Band 8A (865 nm), which is a function of Cab absorption (Figure 5.7). Nitrogen is a key element for developing vegetation chlorophyll (Bassi et al., 2018, Clevers and Gitelson, 2013). Conversely, Homolova et al. (2013) showed a moderate relationship between Cab and N. This is because only a small proportion of total N% is present in Cab.

A key advantage of using Cab for monitoring grassland N% is that this biochemical variable can easily be detected using the visible region of the spectrum (Figure 5.3). However, due to the small amount of N% present in Cab, it is important to use other biochemical variables such as Cp for monitoring N%. As proteins are a key N containing biochemical in vegetation, Berger

et al. (2020b) suggests that Cp can be used as an alternative biochemical vegetation component for retrieving N content.

There are challenges in retrieving Cp due to the uncertainties associated with variable inversion. Most protein absorption bands are present in the spectrum's SWIR (Bands 11 and 12), highly influenced by vegetation structural variables. Thus, it is very challenging to retrieve Cp without constraining other RTM variables (e.g., LAI and Cw).

As illustrated in 5.4, the model has a significant risk of aggregated uncertainty. This will have a significant effect on model stability and performance. It is critical to monitor, measure, and minimize the uncertainty added to the model at different levels to find effective ways of managing errors.

5.5.3 Recommendation of using RTM for predicting N% in grasslands

Our methodology enables field spectroscopy data to simulate different satellite data, including the highly anticipated EnMAP spaceborne instrument. These simulated datasets can then be used to create models for estimation of grassland N% alongside real spaceborne data (Guanter et al., 2015).

One key challenge in using spaceborne MSI for monitoring grassland N% in rough terrains is the impact of topographic effects on the spectra. Rugged terrain strongly affects spectral reflectance, creating variable illumination and shadow effects from slopy areas (Yin et al., 2022). Therefore, we believe topographic correction should be routinely undertaken when using MSI to monitor grasslands in rugged terrain (Yin et al., 2018, 2022, Vanonckelen et al., 2013).

Cloud coverage is another key factor that causes challenges for timely monitoring of grassland N%. To address this challenge, it is possible to use multiple satellite imagery to achieve a high temporal resolution. However, because of spaceborne instruments' different viewing and illumination geometry, sensor harmonization is required to create a high temporal resolution of a scene (Franch et al., 2019). Different sensors can be harmonized using Bidirectional Reflectance Distribution Function (BRDF) correction. BRDF correction is a function that considers sensor viewing and illumination angles to normalize an image to reduce the directional effects of different sensors (Wu et al., 2019, Nguyen et al., 2020).

However, BRDF correction is challenging, and there are different methods to undertake BRDF correction (Li et al., 2012, 2010). Franch et al. 2018 suggested that BRDF correction has a limited impact on the Normalized Difference Vegetation Index (NDVI) (Franch et al., 2018). However, BRDF correction can significantly affect other spectral regions, such as NIR and SWIR. Therefore, It is important to undertake an in-depth study to understand the impact of BRDF correction on grassland N% prediction using different spaceborne instruments.

By successfully addressing these challenges, it is possible to harmonize different spaceborne MSI and develop sensor invariant models, allowing for continuous monitoring of grassland N%. In addition, by using geospatial platforms such as GEE, it is possible to process numerous imagery and develop software applications that farmers and farm managers can readily use to enable the optimization of farms, increase grassland productivity, and have a significant positive impact on the environment (Bonfil et al., 2021).

5.6 Conclusion

S2 TOA reflectance allows for continuous monitoring of grassland N%, enabling the optimization of N fertilizer application to grasslands. In this work, we proposed a methodology using RTM and RFR to create a hybrid model using simulated S2 TOA reflectance from field spectroradiometer data. We validated our model using real S2 TOA reflectance for the different hill country and dairy farms across New Zealand during different periods where data was available. The following are the key findings of our study:

1. It is possible to create a model using simulated S2 TOA reflectance for monitoring grassland N% with the SPART model alongside the RFR algorithm.

2. Propagation of errors (uncertainty) occurs at different stages of the RTM inversion process. This makes retrieving some RTM variables (e.g., Cp) difficult.

3. The proposed model can monitor grassland N% during different periods and locations.

4. It is important to constrain as many RTM variables as possible to reduce the propagation of uncertainty in the model.

5. Due to the challenges associated with retrieving grassland Cp using simulated S2 TOA reflectance, Cab is a key variable for predicting grassland N%.

Our work should be expanded using different sensors, including the newly launched Landsat-9 and the highly anticipated EnMAP spaceborne hyperspectral instrument. Using upcoming spaceborne hyperspectral instruments enables the accurate estimation of large-scale grassland biochemical and biophysical variables. Due to the high level of spectral information provided by hyperspectral data, it could be possible to invert grassland Cp, enabling accurate estimation of N%. Furthermore, by harmonizing different spaceborne instruments, temporal information can constrain RTM and reduce uncertainty.

It is important to investigate BRDF correction methods and their impact on predicting grassland N%. This will enable us to develop transferable models that can be used alongside different spaceborne sensors and cloud computing platforms for monitoring grassland N% at large scales continuously. Our findings can help farmers increase farm productivity, significantly reduce costs associated with traditional N% assessment, and reduce the negative impact of chemical N fertilizers on the environment.

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Mohammad Hossain Dehghan Shoar
Name/title of Primary Supervisor:	Associate Professor Gabor Kereszturi
In which chapter is the manuscript /published work:	6

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

Dehghan-Shoar, M. H., Pullanagari, R. R., Kereszturi, G., Orsi, A. A., Yule, I. J., Hanly, J. (2023). A Unified Physically Based Method for Monitoring Grassland Nitrogen Concentration with Landsat 7, Landsat 8, and Sentinel-2 Satellite Data. Remote Sensing, 15(10), 2491.

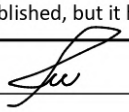
( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:

- The percentage of the manuscript/published work that was contributed by the candidate:

- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	
Date:	03-Sep-2024
Primary Supervisor's Signature:	Gabor Kereszturi Digitally signed by Gabor Kereszturi Date: 2024.09.03 22:07:24 +12'00'
Date:	3-Sep-2024

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

Chapter 6

A Unified Physically-Based Method for Monitoring Grassland Nitrogen Concentration with Landsat 7, Landsat 8, and Sentinel-2 Satellite Data

6.1 Abstract

The increasing number of satellite missions provides vast opportunities for continuous vegetation monitoring, which is crucial for precision agriculture and environmental sustainability. However, accurately estimating vegetation traits, such as nitrogen (N) concentration (N%), from Landsat 7 (L7), Landsat 8 (L8), and Sentinel-2 (S2) satellite data is challenging due to the diverse sensor configurations and complex atmospheric interactions. To address these limitations, we developed a unified and physically-based method that combines a soil–plant–atmosphere radiative transfer (SPART) model with the Bottom of Atmosphere (BOA) spectral bidirectional reflectance distribution function. This approach enables us to assess the effect of rugged terrain, viewing angles, and illumination geometry on the spectral reflectance of multiple sensors. Our methodology involves inverting radiative transfer model variables using numerical optimization to estimate N% and creating a hybrid model. We used Gaussian process regression (GPR) to incorporate the inverted variables into the hybrid model for N% prediction, resulting in a unified approach for N% estimation across different sensors. Our model shows a validation accuracy of 0.35 (RMSE %N), a mean prediction interval width (MPIW) of 0.35, and an R^2 of 0.50, using independent data from multiple sensors collected between 2016 and 2019. Our unified method provides a promising solution for estimating N% in vegetation from L7, L8, and S2 satellite data, overcoming the limitations posed by diverse sensor configurations and complex atmospheric interactions.

6.2 Introduction

Grassland productivity is critical for the meat and dairy sectors due to the ever-increasing demand for high-quality meat and dairy products (Steinfeld et al., 2007). As a significant component of the vegetation photosynthetic apparatus, nitrogen (N) concentration (N%) plays a crucial role in grassland productivity. N based chemical fertilizers are commonly applied to grasslands to increase productivity, which can harm the environment (Rouse et al., 1999, Bassi et al., 2018, Howarth, 2008). Therefore, optimizing the application of N chemical fertilizers to grasslands is essential. Traditionally, grassland N% is monitored by collecting grass samples and sending them to a chemical lab for wet chemistry analysis. An alternative method of monitoring grassland N% is using spaceborne optical imagery (Pullanagari et al., 2016, Frampton et al., 2013, Szantoi and Strobl, 2019, Berger et al., 2020a). Spaceborne optical imagery measures the electromagnetic radiation flux reflected by the incident solar radiation of a surface, commonly referred to as reflectance.

Reflectance is the ratio of the amount of light reflected from a surface to the amount of incident light. Vegetation has a unique reflectance profile across different wavelengths of the electromagnetic spectrum, allowing for the identification of specific chemical properties, including N%, which could be detected using the visible (VIS), near infrared (NIR), and short wave infrared (SWIR) regions of the spectrum (Dehghan-Shoar et al., 2023a). Reflectance can be categorized into Top of Atmosphere (TOA) and Top of Canopy (TOC) reflectance (Verhoef, 1984, 1998). TOA reflectance represents the measurements of the electromagnetic radiation flux reflected by the earth's surface and atmosphere. Although it includes atmospheric effects that may distort the actual reflectance profile of the vegetation, TOA reflectance is widely used in remote sensing applications due to its accessibility and global coverage (Estévez et al., 2020). By applying atmospheric correction techniques, TOA reflectance data can be converted to TOC reflectance, which refers to the measurements of the radiation flux reflected only by the earth's surface, excluding atmospheric effects (Richter and Schläpfer, 2011). The TOC reflectance is considered more relevant for studying vegetation properties, such as N%, because it provides a more accurate representation of the actual reflectance profile of the vegetation by minimizing atmospheric interference (Richter and Schläpfer, 2019). However, the use of TOA reflectance in combination with appropriate atmospheric correction methods can still provide valuable information for monitoring grassland productivity and optimizing N fertilizer application.

The presence of chlorophyll and protein content, which are closely associated with N%, results in distinct spectral reflection that can be detected by spaceborne optical instruments (Curran, 1989, Féret et al., 2021, Dehghan-Shoar et al., 2023c,a, Pullanagari et al., 2021). By analyzing these spectral signatures and the associated reflectance values, it is possible to estimate the N% of vegetation. Advanced algorithms and models are developed to interpret these spectral data and derive accurate N% estimations. This method allows for the non-destructive and continuous monitoring of N% at large spatial scales, making it more efficient than traditional sampling methods (Fernández-Habas et al., 2021). By leveraging the unique reflectance properties of vegetation and advanced data processing techniques, optical remote sensing provides

an effective alternative for monitoring grassland N%.

Monitoring N% in grasslands using spaceborne optical imagery faces significant challenges, including cloud coverage and image acquisition frequency. One solution to address these issues is to utilize a combination of multiple optical sensors, such as Landsat 7 (L7), Landsat 8 (L8), and Sentinel-2 (S2), for continuous and comprehensive monitoring. These sensors provide a robust monitoring solution by increasing the data sample size and offering more frequent images. For example, L7 features a 16-day revisit time, L8 has a 16-day revisit time on the same path and a 5-day revisit time on adjacent paths, while S2 has a 5-day revisit time (Main-Knorn et al., 2017, Goward et al., 2001, Roy et al., 2014). Leveraging multiple sensors with varying revisit times minimizes the impact of cloud coverage or poor image quality on any individual image. However, effectively using the interchangeability of these sensors still needs to be improved. First, each sensor has unique spectral and spatial resolutions and radiometric calibration, leading to inconsistencies when merging data from different sources (Useya and Chen, 2018). Pre-processing and normalization are required to ensure accurate and comparable results. Second, the varying revisit times of these sensors can result in temporal mismatches, making it challenging to create a seamless time series of observations (Shao et al., 2019). Innovative methodologies and tools to fully exploit the potential of using multiple optical sensors for N% monitoring in grasslands must be developed to overcome these challenges.

A common approach to tackling this issue is through a sensor harmonization process called the bi-directional reflectance distribution function (BRDF) (Nguyen et al., 2020, Claverie et al., 2018). The BRDF accounts for the directional dependence of a target's reflected energy as a function of illumination and viewing geometries, producing cross-sensor consistent surface reflectance products (Lucht et al., 2000, Roy et al., 2016b). By converting the imagery to the earth's surface point directly below the satellite and considering factors such as incidence and reflected zenith and azimuth angles, wavelength, and spectral irradiance, BRDF enables the continuous monitoring of grasslands' biochemical and biophysical properties, including N%. Although BRDF is a powerful solution, it can be computationally demanding due to its wavelength dependence and may pose challenges when used with sensors featuring numerous spectral bands (Queally et al., 2022). Despite these hurdles, integrating multiple optical sensors with BRDF presents a promising strategy for overcoming cloud coverage and image acquisition time challenges in grassland N% monitoring (Lucht et al., 2000).

In grassland monitoring, topography presents a significant challenge (Liang, 2003). Its influence on spectral imagery is often underestimated, yet topography plays a crucial role in the spectral analysis of vegetation in an uneven terrain (Dymond et al., 2001, Gu and Gillespie, 1998, Chi et al., 2021). This is because areas with slopes facing the light source receive more significant amounts of down-welling radiation, resulting in increased illumination (Iqbal, 2012). Consequently, this hinders the accurate retrieval of biochemical and biophysical properties of vegetation using empirical–statistical- or physically-based methods. Moreover, the anisotropic nature of non-Lambertian surfaces adds complexity, as the radiance detected by the observing

sensor is further influenced by surface characteristics (Ono et al., 2010, Colby, 1991).

Many topographical correction methods have been developed for Lambertian and non-Lambertian surfaces (Dozier and Frew, 1981). A common method of undertaking topographical correction is by using the cosine method. This method uses the linear relationship between the radiance measured and the viewing incident angle to undertake topographical correction (Richter, 1997, Teillet et al., 1982). However, this method has certain limitations, such as the fact that it does not consider the impact of diffuse irradiance. To address these limitations of the cosine method, Gu and Gillespie (1998) proposed using the sun–canopy–sensor (SCS) topographic correction method, which accounts for some BRDF effects over rugged terrain. This method improves upon the cosine correction method by normalizing the illuminated canopy area of a scene. However, as the SCS method does not consider the impact of diffuse irradiance, this topographic correction technique may result in over-correction for slopes facing away from the sun. To address this limitation of the SCS method, Soenen et al. (2005) proposed a moderator variable (C) that simulates the effect of diffuse sky illumination. This variable is derived by calculating the linear relationship between the radiance and cosine of the incidence angle Soenen et al. (2005), Gu and Gillespie (1998).

Estimating the biochemical and biophysical properties of vegetation can be achieved using physically-based RTMs (Jacquemoud and Baret, 1990, Verhoef, 1998). These models monitor the vegetation properties by simulating the interaction between incoming solar radiation and vegetative canopies (Atzberger et al., 2013, Gastellu-Etchegorry et al., 2004, Estévez et al., 2021, Pu et al., 2020, Verhoef, 1984). One approach involves variable inversion, where RTMs simulate the observed reflectance spectrum as a function of several biophysical and biochemical variables, such as leaf Area Index (LAI), protein content (Cp), and chlorophyll content (Cab) (Vohland and Mader, 2007, de Wit, 1999, Lavergne et al., 2007). Inverting these variables allows RTMs to find the best match between measured and predicted spectra.

Accurately and reliably inverting the biochemical and biophysical variables of RTMs is challenging due to the ill-posed nature of RTM problems and our limited understanding of the biophysical processes governing photon scattering, absorption, and transmission when interacting with vegetative canopies (Liang, 2003, Berk et al., 2006, Atzberger and Richter, 2012, Sun et al., 2022, Li et al., 2018a). Additionally, while the impact of topographical features on spectral reflectance is well-established, the influence of topography on the retrieval of N% using RTM needs further exploration Li et al. (2012). Constraining all sources of uncertainty and potential error is crucial to improve N% retrieval using optical data. The soil–plant–atmosphere radiative transfer (SPART) is a specialized RTM that simulates the interactions between incoming solar radiation, soil, plants, and the atmosphere (Yang et al., 2020b,a). It integrates atmospheric, canopy, vegetation, and soil RTMs, inverting the vegetation's biochemical and biophysical variables using TOA reflectance. By accounting for complex processes, such as light absorption, reflection, and transmission through different components of the terrestrial

environment, the SPART model offers valuable insights into vegetation characteristics, enhancing remote sensing-based estimations of biophysical and biochemical properties for precision agriculture and environmental monitoring applications. Utilizing TOA reflectance enables the retrieval of grassland characteristics without requiring atmospheric correction, increasing computational efficiency and reducing potential aggregated errors often encountered due to incorrect atmospheric correction procedures.

To estimate grassland N%, we propose using BRDF and topographically corrected (using the SCS+C method) L7, L8, and S2 TOA data in conjunction with the SPART model (Yang et al., 2020b,a). Given the varying specifications of L7, L8, and S2 sensors, such as their number of bands, unique machine learning models are often developed for each sensor to estimate N% (Roy et al., 2016a, D'Odorico et al., 2013). However, this approach imposes considerable computational demands when assessing grassland N% on a large scale. By employing SPART to convert spectra from different sensors into a unified feature space, it becomes feasible to construct a single model that leverages RTM-derived biochemical and biophysical variables for N% prediction. The primary advantage of RTM is its foundation in a physical model that simulates the interaction between electromagnetic radiation and vegetation canopies. This enables precise N% predictions under varying environmental conditions, which can be challenging using machine learning techniques relying solely on sensor-specific data. It is essential to combine RTM and machine learning methods to harness the strengths of both. RTM models the physical processes governing N%, while machine learning integrates supplementary environmental information, resulting in more accurate and efficient models for N% prediction. Machine learning techniques offer greater speed and computational efficiency than RTM, making them valuable for large-scale applications. These models also boast increased flexibility, incorporating diverse input data such as weather, soil moisture, and other environmental factors, which enhances prediction accuracy while maintaining efficiency for large-scale use. However, it is crucial to recognize that machine learning models may yield different scientific understanding than RTM. Combining both methods can result in improved predictions and a deeper comprehension of the underlying physical processes governing N% (Dehghan-Shoar et al., 2023c).

In this study, we aimed to develop a unified approach for estimating grassland N% using data from multiple sensors. To harmonize data from different sensors (L7, L8, and S2), we employed the SPART model to invert the grass's biochemical and biophysical properties through numerical optimization. This process utilized pre-processed, topographically corrected TOA reflectance data, subsequently calibrating the machine learning model. We assessed the model's performance by validating it with independent data from the same satellite sources. A comprehensive statistical analysis was conducted to evaluate the model's predictive accuracy and investigate the influence of sensor viewing angles and scene topography variations on the acquired spectra.

6.3 Material and Methods

6.3.1 Field Campaigns and Chemical Analysis

We collected 700 grass samples between 2016 and 2019 from five different hill country and dairy farms across the North and South Islands of New Zealand, ensuring significant variability in our study (Table 6.1). We employed random stratified sampling based on slope angle, aspect, and soil types. Slope angles were categorized into groups 0–8, 8–16, 16–25, and 25+ degrees. The open-source digital elevation model (DEM) from Land Information in New Zealand (LINZ) was used for our sampling design. At each location, a 0.5 square meter quadrant was placed on the grassland, and the grass was harvested using an electric shearing machine before being stored in a cooling box (Hutchinson et al., 2016, Cosgrove et al., 1998, Pullanagari et al., 2012). Each site was geo-referenced with an RTK-GPS system, and photographs were taken for visual assessment. Grass samples were sent to Analytical Research Laboratories (ARL) in Napier, New Zealand, for N% measurement using the Kjeldahl method (Lynch and Barbano, 1999). Samples were dried, ground, and homogenized with a mixer mill (Frommer et al., 2011). To ensure accurate and reliable measurements, we collected data from grass-covered areas with a LAI range of one to four. The LAI is a dimensionless indicator that quantifies the leaf area per unit of ground area, providing essential information on the vegetation's structure and density. This LAI range allows for the precise capture of the canopy's spectral characteristics while minimizing the influence of the surrounding soil background, ensuring proper sensor calibration for vegetation canopy detection and reducing the contribution of soil background to the spectral measurements. By adhering to these sampling guidelines, our data collection was robust and representative of the grassland conditions in the study area. Detailed accounts of our field campaigns and dataset can be found in Pullanagari et al. (2021), Dehghan-Shoar et al. (2023a).

6.3.2 Spaceborne Optical Imagery Harmonization and Topographical Correction

We identified and downloaded L7, L8, and S2 TOA imagery for the dates on which our samples were collected (Table 6.1). This process was undertaken using the Google Earth Engine (GEE) cloud computing platform and the geo-referenced sampling locations (Gorelick et al., 2017). We used satellite imagery captured within seven days with less than 15 percent cloud coverage. We harmonized the imagery captured by different sensors using BRDF, hence re-projecting the L7, L8, and S2 TOA imagery to Nadir (Schaaf et al., 2002). We utilized the spectrally equivalent MODIS (moderate resolution imaging spectroradiometer) bands L7, L8, and S2, along with the corresponding spectral BRDF model parameters suggested by Roy et al. (2016b, 2017). This method was selected based on its demonstrated high performance and ease of calculation use when harmonizing both the S2 and L8 TOA data (Zhang et al., 2018).

Table 6.1: Table containing the list of dates and locations where the field campaigns were undertaken, and Landsat 7 (L7), Landsat 8 (L8), and Sentinel-2 (S2) Top of Atmosphere (TOA) imagery were acquired.

Latitude	Longitude	Field Sampling Date	Image Acquisition Date	Instrument
−40.7	175.8	18 April 2016	26 April 2016	Sentinel-2
−43.9	171.5	04 April 2019	30 March 2019	Sentinel-2
−39.3	174.3	22 October 2018	18 October 2018	Landsat 8
−40.7	175.8	18 April 2016	20 April 2016	Landsat 7
−40.1	175.2	28 November 2019	23 November 2019	Landsat 7

To ensure that the data from different data sources with varying spectral and radiometric characteristics are comparable and consistent, we implemented the modified sun-canopy-sensor (SCS+C) model for the topographic correction of each image. The SCS+C method enhances the standard sun–canopy–sensor topographic correction by incorporating diffuse atmospheric irradiance (Chi et al., 2021, Soenen et al., 2005). To perform topographic correction, we used metadata-based observation and solar angle information, as well as topographic information calculated from the shuttle radar topography mission (SRTM) Digital Elevation Model (DEM), including slope and slope aspect (Van Zyl, 2001). Additionally, we co-registered the DEM layer and the optical imagery to improve alignment and reduce misregistration errors. Finally, the L7, L8, and S2 TOA reflectance and DEM imagery were spatially resampled to 10-meter (m) spatial resolution using the nearest neighbor algorithm for further analysis.

6.3.3 RTM Variable Inversion

The spectral signature for each sample was extracted from the pre-processed L7, L8, and S2 imagery. The extracted spectra correspond to values of a single pixel overlapping with the captured GPS locations. Due to the impact of shadow and cloud coverage, some samples were manually removed from the dataset. Using numerical optimization, we used the SPART model with the extracted L7, L8, and S2 TOA data to invert the biochemical and biophysical variables. This model uses a vegetation canopy RTM, commonly referred to as PROSAIL. PROSAIL was developed using the SAIL and PROSPECT models first proposed by Verhoef (1984) and Jacquemoud and Baret (1990). In this work, we implemented the latest variation of PROSPECT called PROSPECT-PRO, as it enables the forward and inverse simulation of vegetation protein content (Féret et al., 2021). The SPART model also couples the brightness–shape–moisture (BSM) and simplified method for atmospheric correction (SMAC) RTM models; hence, it is capable of simulating the biochemical and biophysical variables of vegetation from TOA data (Yang et al., 2020b,a, Verhoef et al., 2018, Rahman and Dedieu, 1994).

We used a Markov chain Monte Carlo (MCMC) numerical optimization algorithm with a total sum of error as the loss function to invert each sample's biophysical and biochemical variables using SPART (Foreman-Mackey et al., 2019). We used MCMC, as it can sample from high-dimensional probability distributions associated with RTM problems. In addition, MCMC can calculate the uncertainty associated with each inverted variable. We parametrized the

MCMC sampling algorithm using an ensemble sampler with 100 walkers operating for 10,000 simulations per sample. To address the ill-posed nature of RTM, we constrained some of the SPART variables based on prior knowledge and available data, including soil parameters, geographical location, atmospheric parameters, viewing and illumination geometry, and day of the year (DOY) required by SPART (Table 6.2) (Atzberger and Richter, 2012). We estimated the SPART ozone, aerosol, and water vapor values using the Total Ozone Mapping Spectrometer (TOMS), MODIS/Terra Aerosol Cloud Water Vapor, and the NCEP/NCAR 40-Year Reanalysis Project datasets available from GEE (Kalnay et al., 2018, Hubanks et al., 2008). After undertaking the simulations using MCMC, we performed thinning to calculate the mean and one-sigma standard deviation for each inverted variable.

Table 6.2: The SPART parameter search space used for inversion of the biophysical and biochemical parameters from Landsat 7 (L7), Landsat 8 (L8), and Sentinel-2 (S2).

SPART Parameter	**Parameter ID**	**Search Space**	**RTM**	**Units**
Air pressure	Pa	900–1100	SMAC	hPa
Aerosol optical thickness	aot550	0–1	SMAC	-
Water vapor	uh2o	0–2.5	SMAC	g/cm^{-2}
Ozone content	uo3	0–0.4	SMAC	cm-atm
Structure parameter	N	1.5–2.5	PROSPECT-PRO	-
Chlorophyll content	Cab	10–90	PROSPECT-PRO	micro g/cm^{-2}
Carotenoid content	Car	2–9	PROSPECT-PRO	micro g/cm^{-2}
Brown pigment content	Cs	0–0.1	PROSPECT-PRO	-
Equivalent water thickness	Cw	0–0.2	PROSPECT-PRO	cm
Dry matter content	Cdm	0–0.1	PROSPECT-PRO	g/cm^{-2}
Protein content	Cp	0-0.02	PROSPECT-PRO	g/cm^{-2}
Carbon constituents	CBS	0–0.02	PROSPECT-PRO	g/cm^{-2}
Anthocyanin content	ant	0–7	PROSPECT-PRO	g/cm^{-2}
Leaf Area Index	LAI	1–4	4SAIL	-
Leaf angle distribution a	LIDFa	−0.5–0.5	4SAIL	degree
Leaf angle distribution b	LIDFb	−0.5–0.5	4SAIL	degree
Soil brightness	B	0.5	BSM	-
Soil moisture percentage	SMp	50	BSM	percentage
Soil moisture carrying capacity of the soil	SMC	0.25	BSM	-
Single water film optical thickness	film	0.0150	BSM	cm

6.3.4 Retrieval of Grassland N%

After retrieving each sample's biochemical and biophysical variables using the L7, L8, and S2 reflectance, we focused on developing a model to predict grassland N%. We employed a Gaussian process regressor (GPR) algorithm that leverages the inverted biochemical and biophysical variables from TOA imagery for model calibration (Vovk, 2013). The GPR is a robust non-parametric method modeling the relationship between input and output variables. It assumes that the function values at any set of input points are jointly Gaussian-distributed, enabling

uncertainty estimates for predictions at new, unseen input points (GPy, 2012). A kernel function, chosen to measure the similarity between input points, is vital for performing GPR. The kernel function influences the model's behavior and helps learn complex patterns in the data. Various kernel functions exist, including the squared exponential kernel, the Matern kernel, and the radial basis function (RBF) kernel. We optimized our model using the random search optimization (RSO) algorithm, which identified the RBF kernel as the optimal parameter for the GPR (Bergstra and Bengio, 2012).

We validated our model using spatially and temporally independent data from L7, L8, and S2 TOA reflectance data to predict grassland N%. The dataset we used in this study consisted of 377 for S2, 277 for L7, and 46 for L8 samples. We used 30% ($n = 210$) of our total dataset for independent validation, while the remaining samples were used for calibration and cross-validating purposes. Due to the imbalanced nature of our dataset, we used the synthetic minority over-sampling technique for regression with Gaussian noise (SMOGN) on the calibration data. This method performs the synthetic minority over-sampling technique for regression (SMOTER) with traditional interpolation and Gaussian noise (SMOTER-GN) (Branco et al., 2017). To achieve the best possible results, we optimized the default parameters for the method, setting the number of neighbors for over-sampling to 6, perturbation or noise percentage to 0.02, and a threshold of 0.5. Furthermore, we identified the most informative biochemical and biophysical variables for predicting grassland N% using the recursive feature elimination algorithm (RFE) and GPR (Pullanagari et al., 2018, Guyon et al., 2002). The RFE is a commonly used feature selection technique that fits a model and removes the least important variables from a dataset until the existing requirements have been met (specified number of features, in our case, five).

Following this procedure, we calculated the mean and standard deviation of the predicted values. These values are then used to measure the performance of our methodology by calculating statistical metrics, including R^2, root mean squared error (RMSE), and mean prediction interval width (MPIW) (Khosravi et al., 2011). In addition, we investigated the impact of slope angle, slope aspect, viewing, and illumination geometry on the spectra captured by each sensor, by comparing the impact of these variables on the simulated and real harmonized and topographically corrected reflectance. In addition, we investigated the impact of the topographical correction on our model performance using the validation dataset. The flowchart containing the methodology used in this study is illustrated in Figure 6.1.

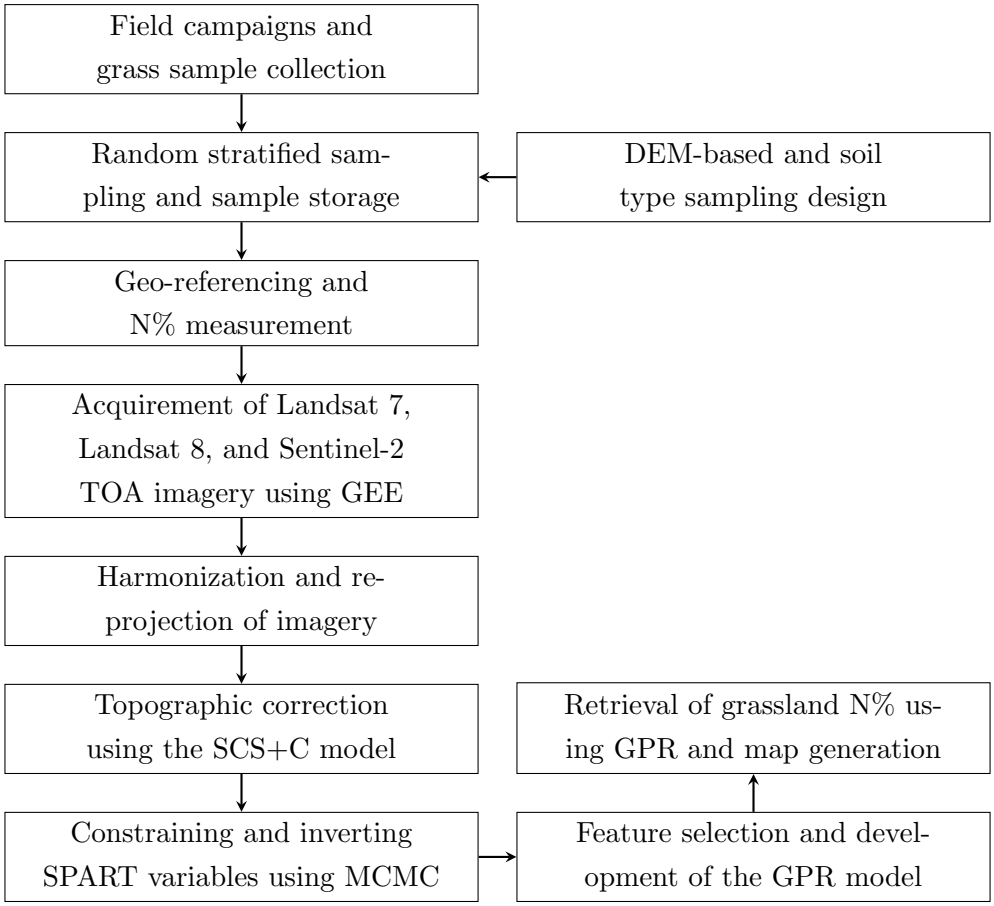


Figure 6.1: Flowchart illustrating the methodology for grassland nitrogen (N) concentration (N%) retrieval using SPART variables and Gaussian Process Regressor (GPR). The process includes field campaigns for grass sample collection, Digital Elevation Model (DEM) based and stratified sampling design, geo-referencing and N% measurement, acquisition and harmonization of Landsat and Sentinel-2 imagery, topographic correction using the SCS+C model, and inversion of SPART variables via Markov Chain Monte Carlo (MCMC). The GPR model is developed through feature selection and used for N% retrieval and map generation.

6.4 Results

6.4.1 Impact of Viewing and Illumination Geometry on Simulated TOA Reflectance

By fixing all SPART parameters while changing the viewing and illumination angles of SPART (shaded green area of Figure 6.2), we estimated the sensitivity of TOA reflectance to viewing geometry. This indicates that the viewing zenith angle impacts the VIS and NIR of TOA reflectance, while the illumination zenith angle has a limited impact. A higher viewing zenith angles result in lower reflectance in the NIR, with a decrease of greater than 10%. Lower illumination zenith angle values increase the TOA reflectance. Similarly, variation of the illumination zenith angle has a limited influence, with a decrease of less than 3%, on other parts of the spectrum, such as the SWIR.

The impact of viewing and illumination geometry on TOA reflectance is noticeably pronounced in rugged terrain, whereas this effect is relatively minimal in flat areas. As illustrated in Figure 6.3, a considerable visual difference between rugged and flat terrains regarding TOA reflectance highlights the sensitivity of rugged terrain to illumination and viewing geometry. Moreover, the red circles in the same figure further emphasize the limited impact observed in flat regions, which is consistent with the notion that the effect of viewing and illumination geometry on TOA reflectance is generally more pronounced in rugged terrain.

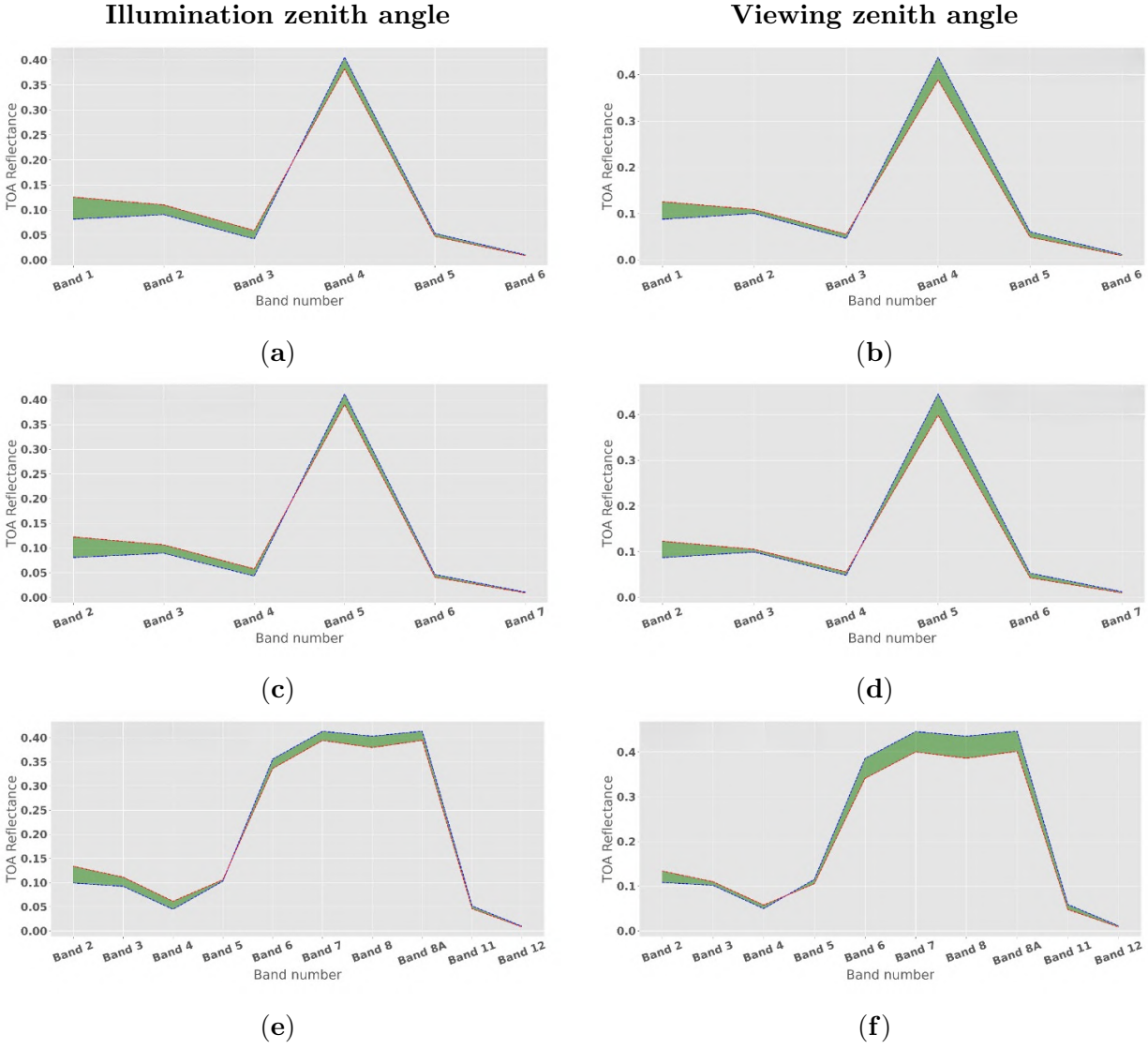
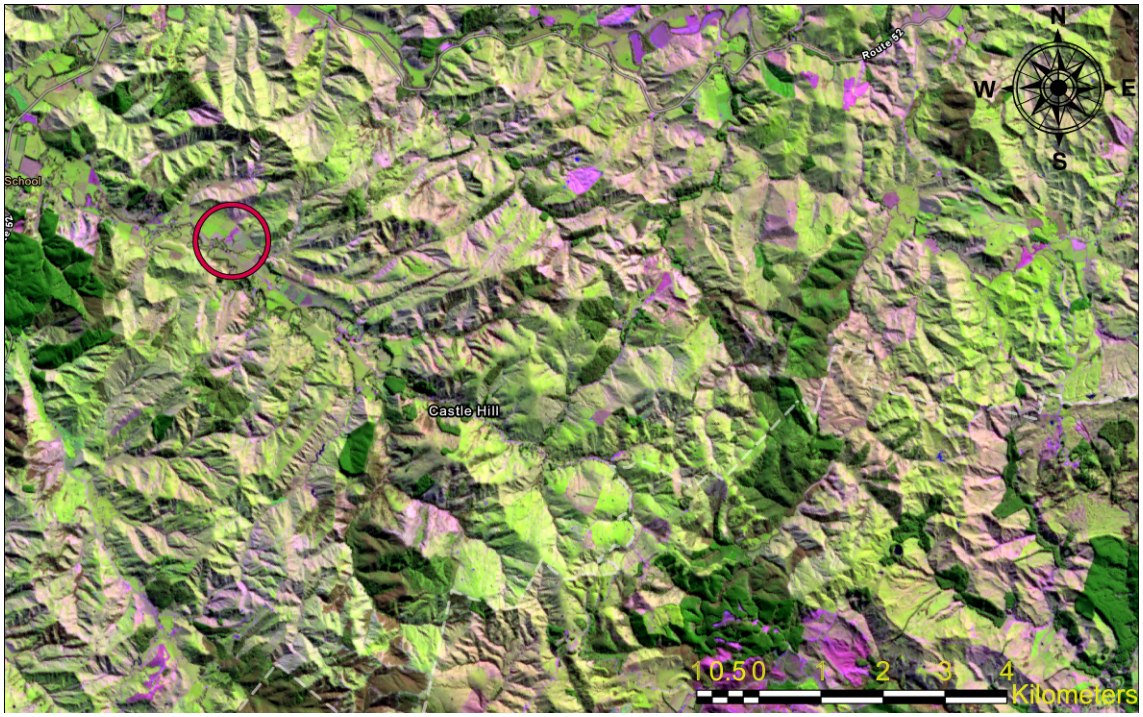
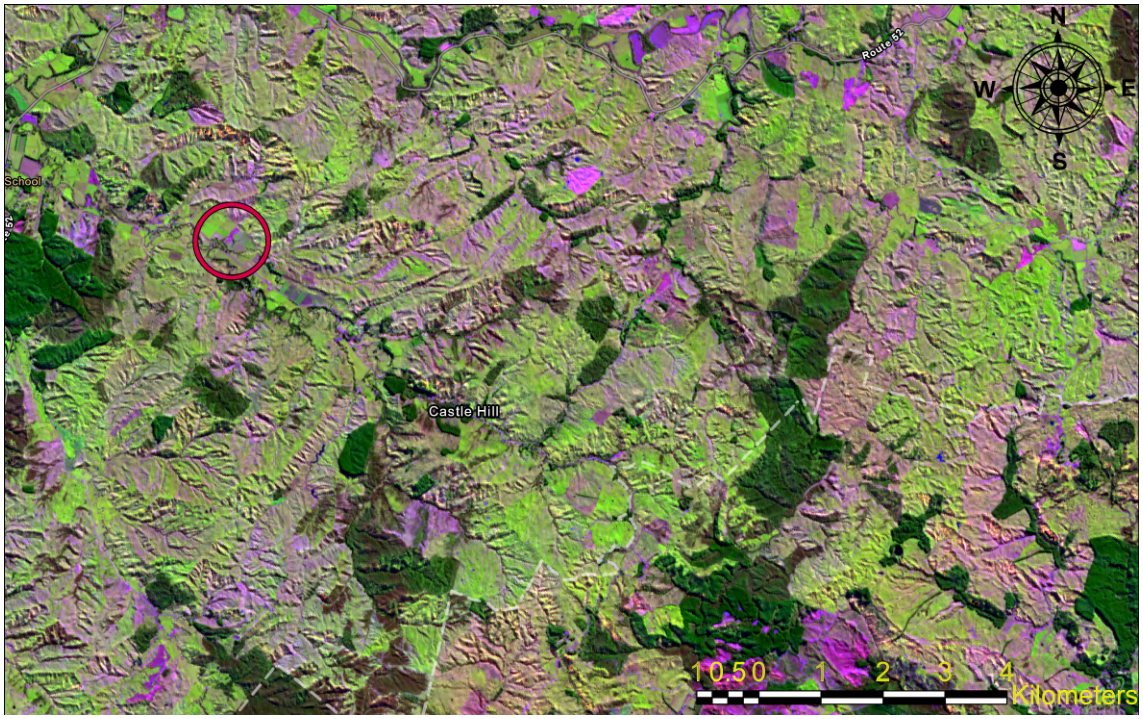


Figure 6.2: The impact of viewing and illumination geometry on simulated Top of Atmosphere (TOA) reflectance for Landsat 7 (L7; a, b), Landsat 8 (L8; c, d), and Sentinel-2 (S2; e, f) across different bands. Subplots (a), (c), and (e) represent the influence of illumination zenith angles, while (b), (d), and (f) show the effect of viewing zenith angles. The dashed red lines correspond to higher zenith angles (greater deviation from nadir), and the dashed blue lines correspond to lower zenith angles (closer to nadir). The plots illustrate variations in TOA reflectance caused by changes in sensor viewing and solar illumination geometry across spectral bands, highlighting sensitivity to angular configurations.



(a) Sentinel-2 TOA image without topographical correction



(b) Topographically corrected Sentinel-2 TOA image

Figure 6.3: Impact of topographic correction on Sentinel-2 (S2) Top of Atmosphere (TOA) imagery in rugged terrain for grassland present in rugged terrain in the north island, New Zealand. The figure is a composite of Band 11, Band 8A, and B4 of an S2 TOA image. The red circles highlight the limited impact observed in flat regions in subfigures **a** and **b**.

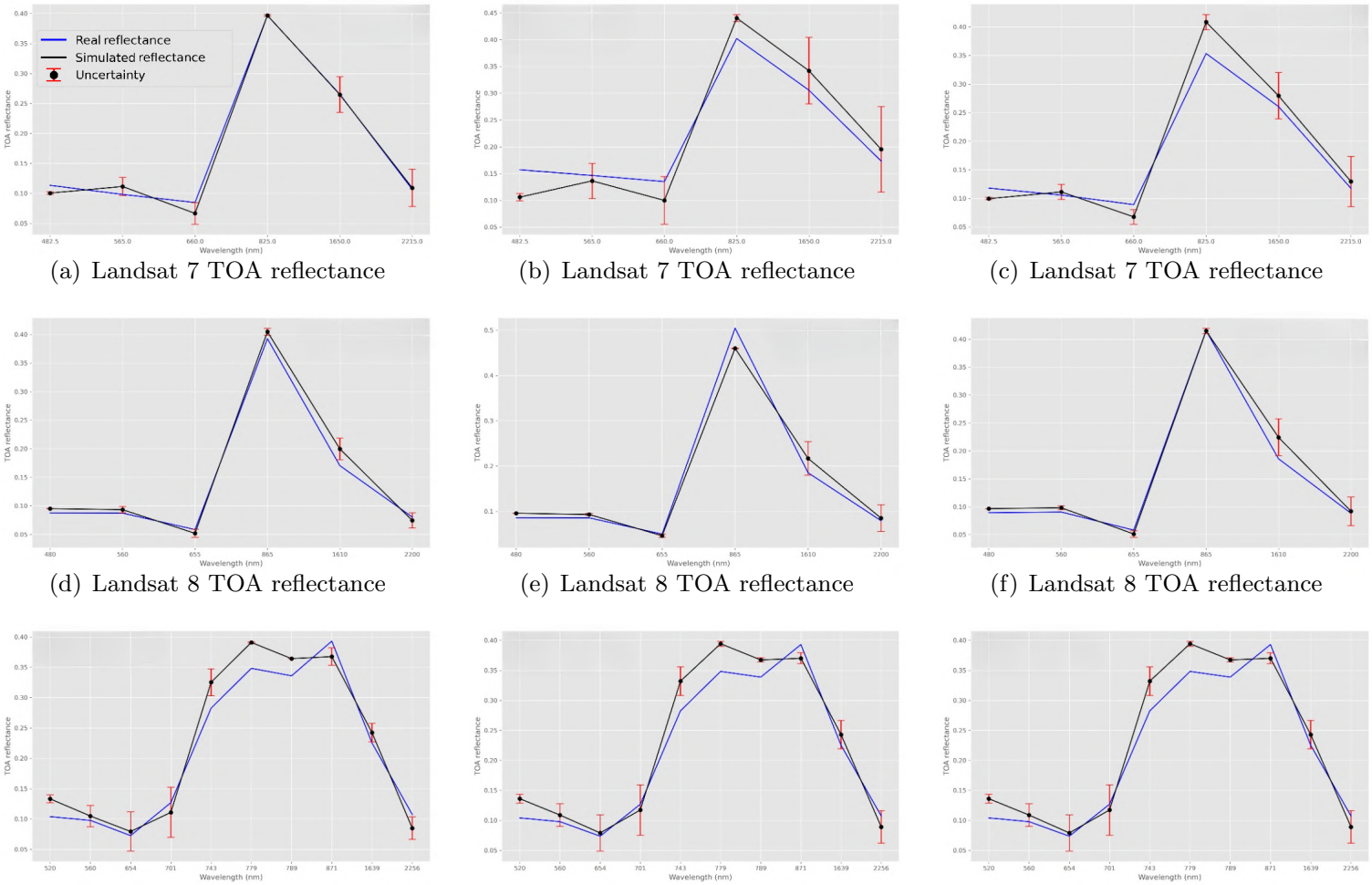
6.4.2 Inversion of SPART Variables

We used SPART and an MCMC numerical inversion method to simulate spectral data and retrieve the biochemical and biophysical variables. The simulation aimed to minimize the difference between the actual and simulated spectral data as described in (Dehghan-Shoar et al., 2023c). This simulation showed higher uncertainty in the NIR region of the L7, L8, and S2 data sets, as demonstrated in Figure 6.4.

6.4.3 Impact of Topographic Correction on TOA Reflectance

We applied the SCS+C method for topographic correction after implementing BRDF to harmonize sensor data. Our findings reveal a notable visual contrast in rugged terrain before and after topographic correction. At the same time, minimal differences are observed in flat areas (as emphasized in red in Figure 6.3). Following topographic correction, the NIR spectral reflectance decreases in rugged terrain, aligning with our observations in Figures 6.3 and 6.5. This reduction in NIR spectral reflectance can be due to the variations in surface slope and orientation associated with rugged terrain. Such variations can influence the amount of sunlight reflected by the surface, and topographic correction compensates for these disparities, yielding a more precise estimate of the surface reflectance.

We investigated the impact of slope angle and slope aspect on each spectral band of S2 TOA reflectance, as illustrated in Figure 6.6. Our results show that a higher slope value is usually associated with a greater difference (on average, better than 10% in the NIR). In particular, the difference between the reflectance values is typically higher when the aspect is northeast and southwest-facing. This is because the terrain is facing the sun during image acquisition. Furthermore, the NIR (B6, B7, B8, and B8A) is highly impacted by steep terrain in comparison with the visible region (B2, B3, and B4) and SWIR (B11 and B12).



(a) Landsat 7 TOA reflectance (b) Landsat 7 TOA reflectance (c) Landsat 7 TOA reflectance

(d) Landsat 8 TOA reflectance (e) Landsat 8 TOA reflectance (f) Landsat 8 TOA reflectance

(g) Sentinel-2 TOA reflectance (h) Sentinel-2 TOA reflectance (i) Sentinel-2 TOA reflectance

Figure 6.4: Plots depicting the uncertainty associated with the simulated and actual Top of Atmosphere (TOA) reflectance for Landsat 7 (L7; a, d, g), Landsat 8 (L8; b, e, h), and Sentinel-2 (S2; c, f, i). Columns one, two, and three correspond to L7, L8, and S2 TOA reflectance, respectively. Each row represents a different dataset: the first row (a–c) shows the highest uncertainty across bands, the second row (d–f) displays moderate uncertainty, and the third row (g–i) illustrates the lowest uncertainty. The blue line represents actual reflectance, the black line represents simulated reflectance, and red vertical lines denote the uncertainty for each band. These plots highlight variations in reflectance and associated uncertainties for each satellite and dataset configuration.

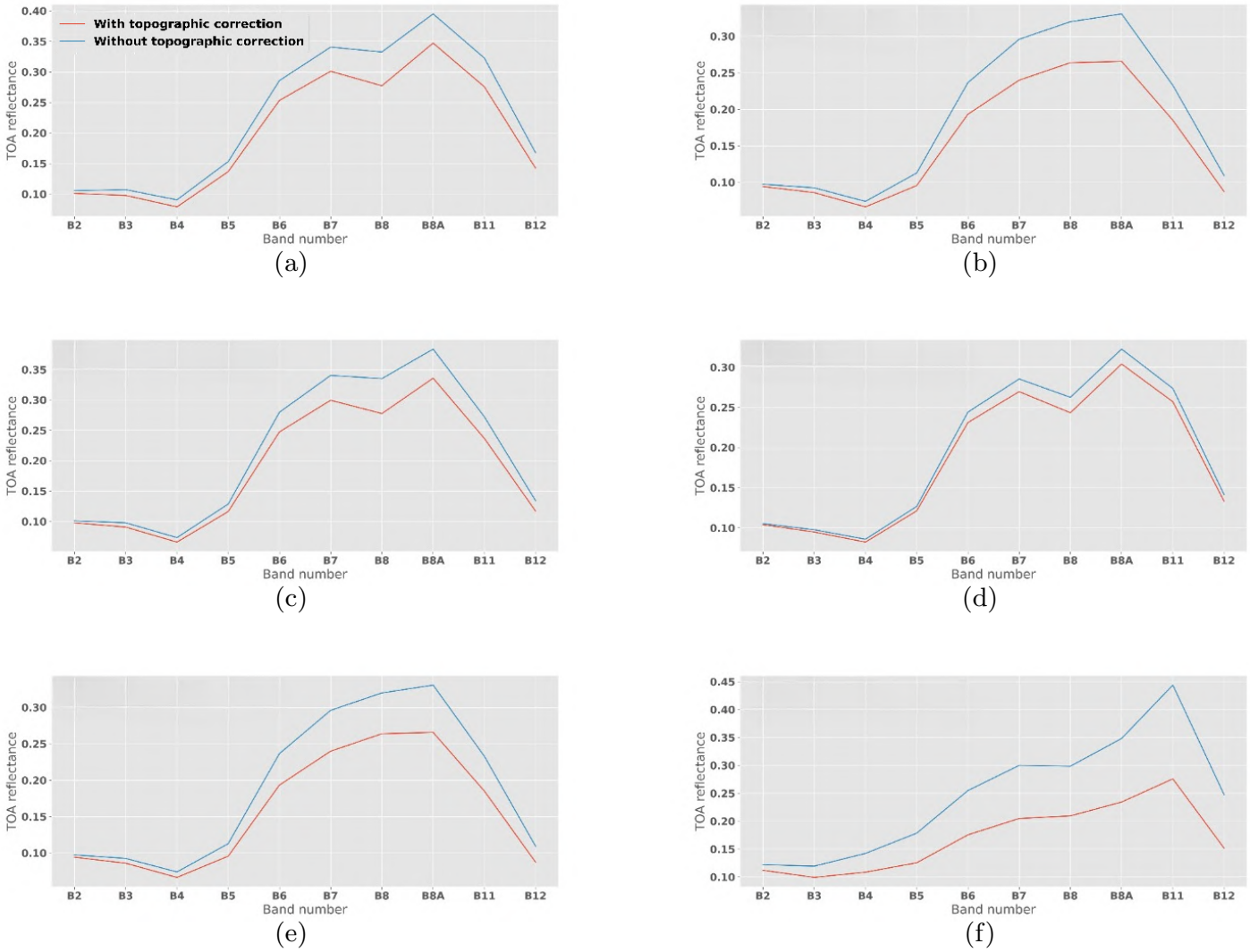
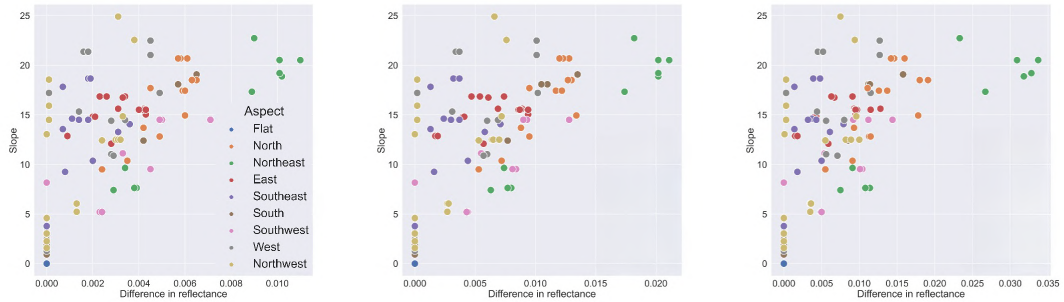


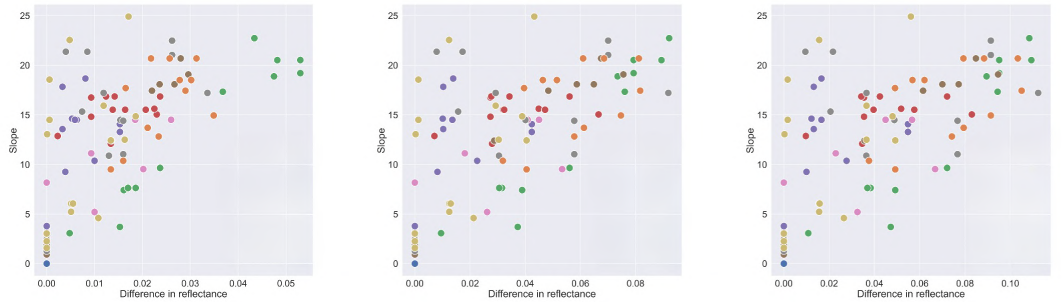
Figure 6.5: Assessment of the influence of topographic correction on Sentinel-2 (S2) Top of Atmosphere (TOA) reflectance. Subfigures (a–f) illustrate the impact of topographic correction on different samples by comparing reflectance values obtained with (blue line) and without (red line) topographic correction. The x-axis represents the spectral bands (B2 to B12), and the y-axis shows the TOA reflectance values. These plots highlight the variations caused by topographic correction across spectral bands, demonstrating its effect in reducing biases and improving consistency in reflectance retrieval. Each subfigure represents a unique sample analyzed in this study.

6.4.4 Validation Results of the Model

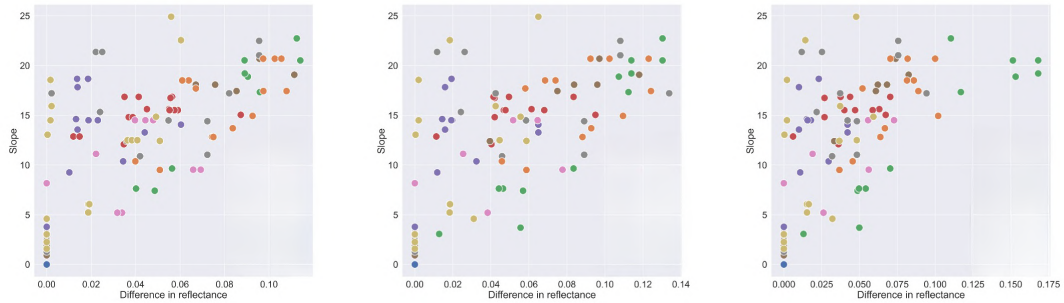
The Cab, Cw, Cbc, LIDFa, and LAI were identified as the most informative features for predicting N% in vegetation based on the RFE technique. Among these variables, Cab, representing chlorophyll-a concentration, is the most important due to its strong correlation with N%, as both are essential for photosynthesis (Dehghan-Shoar et al., 2023c, Wang et al., 2022b). Cw is related to N% through its role in cellular hydration, while Cbc is indirectly associated with N% through the carbon-N balance in plants (Gao et al., 2020a). LIDFa provides insights into canopy structure influenced by N availability, and LAI indicates vegetation density, which is also affected by N availability (Pullanagari et al., 2021). Using these biochemical and biophysical variables derived from TOA reflectance, we calibrated our model with an optimized GPR algorithm for predicting N% in vegetation.



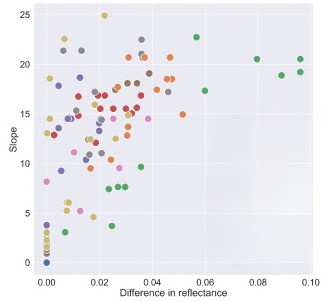
(a) B2 (b) B3 (c) B4



(d) B5 (e) B6 (f) B7



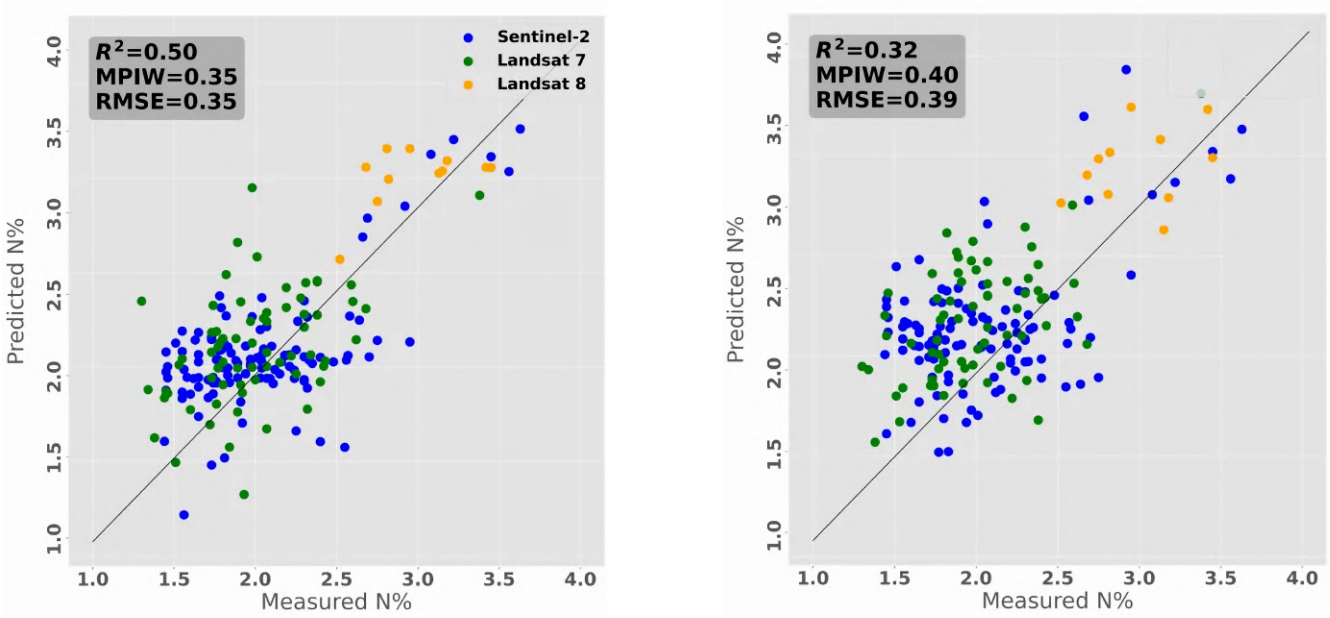
(g) B8 (h) B8A (i) B11



(j) B12

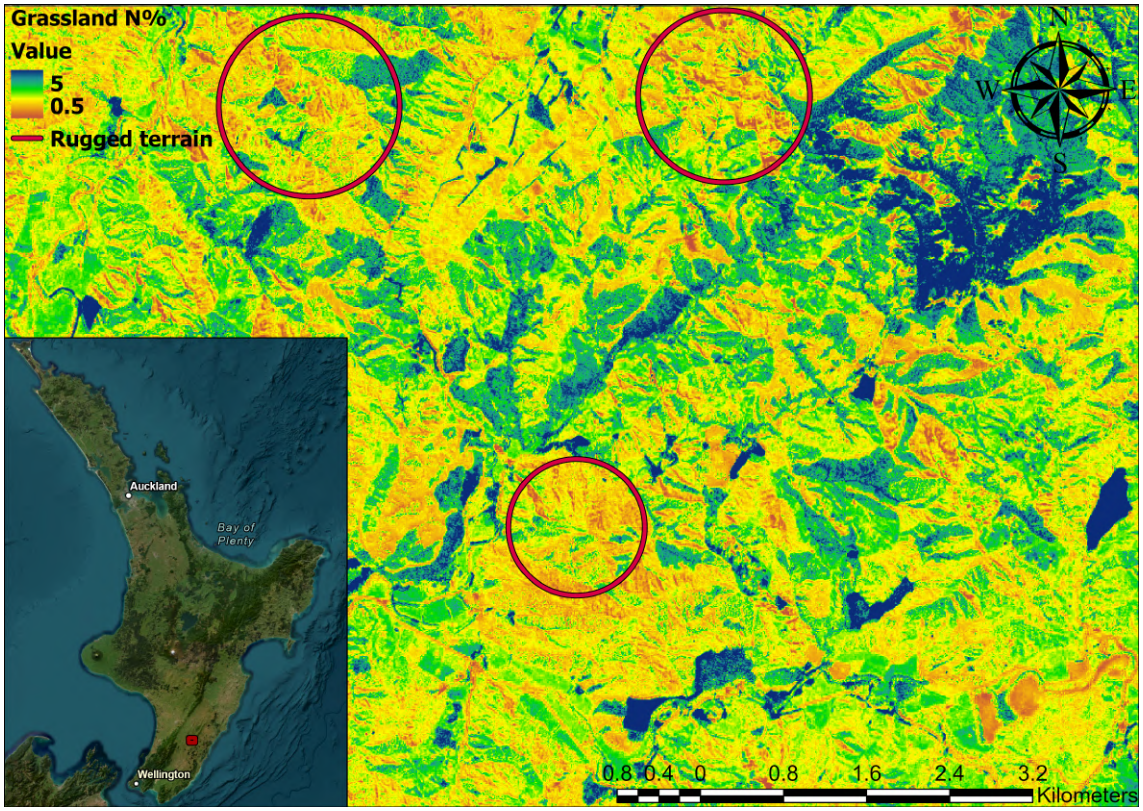
Figure 6.6: The impact of slope angle and slope aspect on Sentinel-2 (S2) Top of Atmosphere (TOA) reflectance before and after topographic correction across bands B2 to B12. Subfigures (a–j) show scatterplots with the x-axis representing the difference in reflectance between corrected and non-corrected data, and the y-axis showing slope angle. Points are color-coded by slope aspect (flat, north, northeast, east, southeast, south, southwest, west, northwest). These plots highlight the influence of slope angle and aspect on reflectance correction, with notable variation across spectral bands.

We evaluated our GPR model's performance using spatially and temporally independent biochemical and biophysical data from L7, L8, and S2 satellite imagery. Our model demonstrated moderate performance using topographically corrected TOA reflectance data, achieving R^2 values of 0.5, RMSE values of 0.35, and MPIW values of 0.35, as shown in Figure 6.7 a. However, when not using topographically corrected data, the model's performance declines, resulting in R^2 values of 0.32, RMSE values of 0.40, and MPIW values of 0.39, as depicted in Figure 6.7 b. The difference in prediction accuracy can be attributed to the importance of accounting for topographic effects on reflectance data, which helps improve the model's ability to accurately estimate biochemical and biophysical properties. Additionally, combined uncertainties from model inversion, BRDF effects, and topography contribute to the observed discrepancies in model performance, which are also present in N% maps generated (Figure 6.8). Addressing these sources of uncertainty is essential for enhancing the reliability and accuracy of remote sensing-based models when estimating vegetation properties.

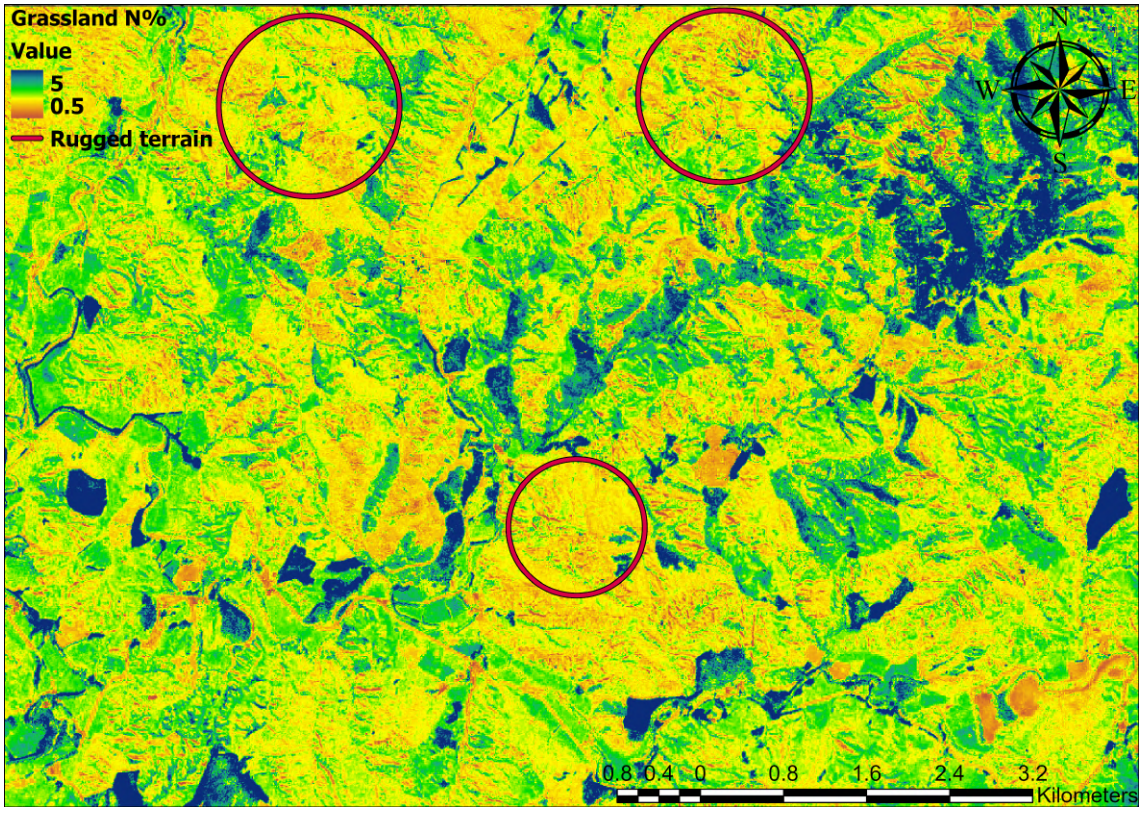


(a) (b)

Figure 6.7: Performance of the model for monitoring N% using harmonized and topographically corrected Landsat 7 (L7), Landsat 8 (L8), and Sentinel-2 (S2) imagery. Subfigure (a) shows independent validation results with corrections, achieving $R^2 = 0.50$, Mean Prediction Interval Width (MPIW) = 0.35, and Root Mean Squared Error (RMSE) = 0.35. Subfigure (b) presents independent validation results without corrections, with $R^2 = 0.32$, MPIW = 0.40, and RMSE = 0.39. The diagonal line represents the 1:1 relationship between predicted and measured nitrogen (N) concentration (N%), while points are color-coded by imagery source: S2 (blue), L7 (green), and L8 (orange). These results highlight the improvement in model performance with topographic corrections.



(a) Sentinel-2 TOA image N% map without topographical correction



(b) Sentinel-2 TOA N% map with topographical correction

Figure 6.8: The impact of the topographic correction on Sentinel-2 (S2) Top of Atmosphere (TOA) derived nitrogen (N) concentration (N%) in grasslands in rugged terrain in the North Island of New Zealand is shown, with the inset map in figure (**a**) highlighting the study area. Subfigures (**a**,**b**) illustrate the discrepancies in grassland N% estimation due to the effects of rugged terrain. The red circles illustrate the impact of topography on the predicted N% values.

6.5 Discussion

6.5.1 Impact of Viewing and Illumination Geometry

Vegetation Cab and Cp are often used as proxies for predicting grassland N% as these two variables are closely linked to vegetation N% (Homolova et al., 2013, Berger et al., 2020b,a). Since the spectral absorption associated with Cab is present in the VIS of the spectrum, Cab can be used as a robust proxy for monitoring grassland N% without applying BRDF correction. However, despite the promising results achieved for monitoring N using Cab, Cp is the main source of N in vegetation, and the accurate retrieval of Cp is critical for monitoring N% (Féret et al., 2021). Key bands for estimating protein are scattered across the electromagnetic spectrum (particularly in the SWIR); as a result, BRDF remains an important step for inverting Cp using RTM. It is important to identify and constrain all sources of uncertainty and error to accurately invert RTM variables using numerical inversion (MCMC) (Figure 6.4). However, there is uncertainty with inverted variables derived using MCMC uncertainties associated with the NIR and SWIR caused by several factors, including the inversion process of SPART variables. As many factors impact the SWIR (e.g., atmospheric variables and viewing geometry), it is important to identify and isolate different sources of error and uncertainty (including viewing and illumination geometry) for inverting Cp, thus predicting accurate N%. This becomes more critical when developing invariant sensor models due to the differences in the sensor specifications.

Our study aimed to evaluate the impact of viewing and illumination geometry on spectra, particularly in the VIS and NIR regions. The results revealed that geometry significantly affects both regions, consistent with previous research by Mao et al. (2020) that reported the sensitivity of the spectral regions between 760 nm to 890 nm and 450 nm to 680 nm with geometric factors. This sensitivity can be attributed to the anisotropy of the vegetative canopy and its interaction with radiation as explained by (Proy et al., 1989). Moreover, our study showed that predicted biomass values are more sensitive to geometric factors than biochemical and biophysical variables impacted by the VIS region, consistent with the findings of Buchhorn et al. (2016), who investigated the impact of geometry on various variables, including biomass. Ignoring these factors can lead to misleading conclusions, as demonstrated by Mao et al. (2020). Therefore, our results highlight the importance of considering viewing and illumination geometry, particularly in the NIR region, when predicting biomass values. Additionally, neglecting BRDF effects, as highlighted by Buchhorn et al. (2016), can result in inaccurate estimations of biochemical and biophysical variables of vegetation.

6.5.2 Impact of Topography

Topographical correction can significantly impact spectra when the terrain is steep and facing the sun as shown in Figure 6.6. This is because steep terrain receives more solar irradiance, which affects the measured reflectance (Iqbal, 2012, Bishop et al., 2019). Topographical factors significantly impact longer wavelengths (greater than 750 nm), making it difficult to retrieve

certain biochemical and biophysical variables, such as LAI, Cw, and Cbc, which have characteristic absorption in these regions. This can lead to inaccurate estimation of N% using SPART. To investigate this issue, we compared the performance of our model with and without topographical correction. Our findings show that topographical correction improves model performance and transferability, particularly in rugged shadow-dominated terrain. de Oliveira et al. (2021) investigated the interaction between topographic factors on vegetation indices derived from hyperspectral imagery. They found that most vegetation indices related to biochemical and biophysical variables and plant physiology have some degree of anisotropy with terrain illumination effects, consistent with our findings (Figures 6.3 and 6.5). There are significant visual differences before and after topographic correction. This change in reflectance could have a significant impact on the inversion of SPART variables, adding error and uncertainty to our model.

In addition, we observed the "over-correction" of spectra, where the terrain is not facing the sun. The SCS+C proposed by Soenen et al. (2005) reduces the over-correction of spectra by introducing the C parameter as an empirical coefficient to calculate diffuse skylight irradiance. Given the simplicity of the SCS+C model and lack of consideration of the anisotropic irradiance and reflectance, this model cannot generalize well to multiscale topographic regions Bishop et al. (2019). Bishop et al. (2019) has discovered that it is very challenging to accurately retrieve biophysical variables of a scene by using empirical topographical correction methods. A detailed investigation is required to develop transferable and physically-based topographical correction techniques for retrieving grassland spectra in rugged terrain.

A limitation of this work is that the topographical correction procedure used is a low spatial resolution DEM layer. This limitation can lead to errors and uncertainties in the topographically corrected image, impacting the estimation of biophysical and biochemical variables, including LAI and Cab. The implications of inaccurate estimation of these variables can have far-reaching consequences for vegetation monitoring and management (e.g., estimation of N%). Higher-quality data sources, such as light detection and ranging (LiDAR) products, can provide highly accurate and detailed elevation data for topographical correction. This could reduce errors and uncertainties in the corrected image, improving biophysical and biochemical variable estimation accuracy and, therefore, N% retrieval. It is essential to note that using higher-quality data sources, such as LiDAR, can come at a higher cost regarding financial resources and time. Therefore, researchers and practitioners must carefully consider the trade-offs between data quality and cost-effectiveness when selecting data sources for topographical correction.

6.5.3 The Performance of the Model

The proposed model can predict grassland N% using the biochemical and biophysical variables derived from multiple satellite imagery (Figure 6.7). The independent validation results of our model are promising and show good performance, consistent with results achieved in other work (Figure 6.4) (Dusseux et al., 2022, Dehghan-Shoar et al., 2023c). Furthermore, our results demonstrate that the topographically corrected data perform better in terms of RMSE, MPIW,

and R^2 compared to data retrieved without undertaking topographical correction.

Using RFE, we identified Cab, LAI, LIDFa, Cbc, and Cw as the critical variables for predicting N%. The importance of these variables highlights the need for topographical correction and BRDF correction, which can improve the accuracy of the predictions. By inverting the variables and using physical characteristics from the RTM, we can develop transferable models that consider several physical parameters impacting the spectra. This approach allows us to combine physical and empirical information, taking advantage of both methods and improving the robustness of our predictions. The selection of these biophysical and biochemical variables by RFE depicts the importance of above ground biomass (AGB) and chlorophyll for estimating N%. Although there is no direct relationship between N% and AGB, there is an indirect relationship between the two because higher N% positively correlates with AGB for grasslands (Chen et al., 2018). Our research found that LAI, Cw, and Cbc are essential for estimating N% proxies for fresh AGB. This aligns with the findings of Wocher et al. (2022), which showed that LAI, Cw, Cp, and Cbc could be used to estimate AGB. It is important to note that while models trained using spectra and similar empirical and statistical methods may still provide helpful information, they may not be as transferable as models that combine physical and empirical information. By considering both methods, we aim to improve the accuracy and reliability of our predictions of N%.

Even though Cp is the main N-containing biochemical in grassland, the RFE algorithm has yet to identify this variable as necessary. This is because of the uncertainty and error in the inversion process and limitations of the sensors (e.g., low spectral and spatial resolution and poor signal-to-noise ratio of the SWIR). Alternatively, Cab is selected as an essential variable for estimation N%, as Cab is the second-most N-containing biochemical variable in vegetation (Berger et al., 2020b, Homolova et al., 2013). The lack of selection of Cp for predicting N% highlights the importance of identifying and addressing different sources of error and uncertainty, including topography and geometry when monitoring grassland N% in rugged terrain.

6.5.4 Enhancing RTM for Monitoring Grassland N% in Rugged Terrain

Enhancing the accuracy of the proposed model for grassland N% prediction requires addressing factors that affect input variable quality and inversion process precision. This can be achieved using higher spectral resolution instruments, such as hyperspectral imagery with a superior signal-to-noise ratio (SNR), to collect high-quality spectral data and enable a more precise inversion process (Dehghan-Shoar et al., 2023a). Improvements to the model's accuracy can also result from refining the statistical and physically-based models used in the inversion process, such as algorithm advancements, data incorporation, or parameter adjustments (Rivera et al., 2013, Xu et al., 2022). Moreover, exploring new methods for obtaining RTM variables and enhancing the inversion process may be beneficial. Addressing these factors can improve model accuracy and provide reliable grassland N% monitoring results.

This study demonstrated promising results in monitoring grassland N% using BRDF and topographical correction techniques. Although our approach offers higher temporal resolution through various optical spaceborne instruments, periods with limited optical imagery remain. Therefore, exploring alternative methods for improving temporal resolution in N% monitoring is essential. Further enhancements are required to monitor grassland biophysical and biochemical variables. Our results suggest that integrating BRDF and topographical correction techniques has the potential for grassland N% monitoring. However, limitations exist, including limited temporal resolution and the need to refine statistical and physically-based models within the inversion process. Future research could explore physics-informed machine learning algorithms for vegetation N% monitoring, given the strong correlation between N% Cab and Cp (Karniadakis et al., 2021, Raissi et al., 2017). Implementing such strategies may enable more accurate and reliable monitoring across different vegetation species.

6.6 Conclusion

Our study presents a novel physically-based sensor workflow for monitoring grassland N% using spaceborne optical imaging. Our approach integrates multiple RTMs with L7, L8, and S2 TOA reflectance data, BRDF for sensor harmonization, and SCS+C for topographical correction. The key findings of our work are as follows:

1. Our approach achieves an independent validation accuracy of 0.35 (RMSE %N), a mean prediction interval width value of 0.35, and an R^2 of 0.50 using independent validation data from multiple sensors between 2016 and 2019, demonstrating the potential for the cost-efficient monitoring of grassland N% using various spaceborne optical instruments in rugged terrain.

2. We investigated the impact of rugged terrain, viewing, and illumination on different sensors' spectral reflectance for estimating grassland N%, showing that viewing and illumination geometry can significantly impact spectra, particularly in longer wavelengths. Moreover, topographic correction is essential for monitoring grassland characteristics in rugged terrain.

3. To address the ill-posed nature of RTM, it is essential to identify and address sources of uncertainty, including topography and viewing and illumination geometry. Future research should investigate the impact of different BRDF and topographical correction methods on retrieving grassland N%.

4. Although our proposed methodology provides higher temporal resolution for monitoring grassland N%, there are still periods where acquiring optical imagery is challenging. Therefore, it is crucial to investigate alternative methods of continuously monitoring grassland characteristics.

5. Further investigation should be undertaken using physically guided machine learning algorithms to monitor N%. This will enable the development of high-performance sensor models for monitoring vegetation characteristics across different species.

Overall, our study represents a significant step forward in developing an efficient approach for monitoring grassland N%, with important implications for improving the sustainability of agriculture and the environment. Our approach offers a low-cost and continuous monitoring solution for grassland N% that reduces the negative impact of chemical N fertilizers on the environment. Additionally, our approach is available on cloud services, providing increased accessibility for users.

Chapter 7

General Discussion

Optical sensors are emerging as valuable tools in precision agriculture, particularly for enhancing the productivity and sustainability of grasslands. Unlike traditional methods, such as wet chemistry analysis, these sensors offer a non-destructive and near real-time means of assessing grass quality (Berger et al., 2020b, Estévez et al., 2020). Despite their potential, the application of optical remote sensing for monitoring grassland dynamics presents several challenges. For instance, cloud cover can hinder the collection of reliable measurements, complicating the extraction of accurate nutrient information from spectral data. Additionally, various confounding factors can influence the accuracy of these measurements (Camps-Valls et al., 2023). This thesis explores methods to estimate nitrogen (N) concentration (N%) in grasslands, leveraging optical sensing technology to support sustainable agricultural practices. The goal is to enhance grassland productivity for livestock feeding while maintaining ecological integrity.

A thorough evaluation of various models is conducted, including physically-based, empirical-statistical, and hybrid models developed for characterizing vegetation traits using different optical reflectance spectroscopy data. This evaluation includes assessing the advantages and limitations of each model type. Physically-based models offer valuable insights such as key information regarding the biophysical and biochemical status of vegetation and have high generalizability but face uncertainties and challenges in data integration due to their inherently ill-posed nature (Sun et al., 2022). Empirical-statistical models are noted for their predictive accuracy, while hybrid models combine the strengths of both to improve generalizability, accuracy, and ease of use. To address prediction models using optical remote sensing, this research has three objectives:

1. To estimate N% in grasslands using optical reflectance spectroscopy, data must be collected across multiple scales, including dry/ground-, leaf-, canopy, and satellite-scale observations.

2. To improve grassland N% models' universality and adaptability through a hybrid approach that combines data from various optical sensors across multiple scales.

3. To account for and quantify uncertainties in grassland N% prediction models.

This study crafted four cornerstone Chapters, focusing on grassland N estimation and aligning with the research objectives (Figure 7.1). The approach encompasses an expansive spatial analysis, from dry/ground scale spectral data to the utilization of multisensor spaceborne data, highlighting the capability for hybrid model development for each spatial scale. These models integrate with various optical sensing technologies, ensuring frequent monitoring opportunities.

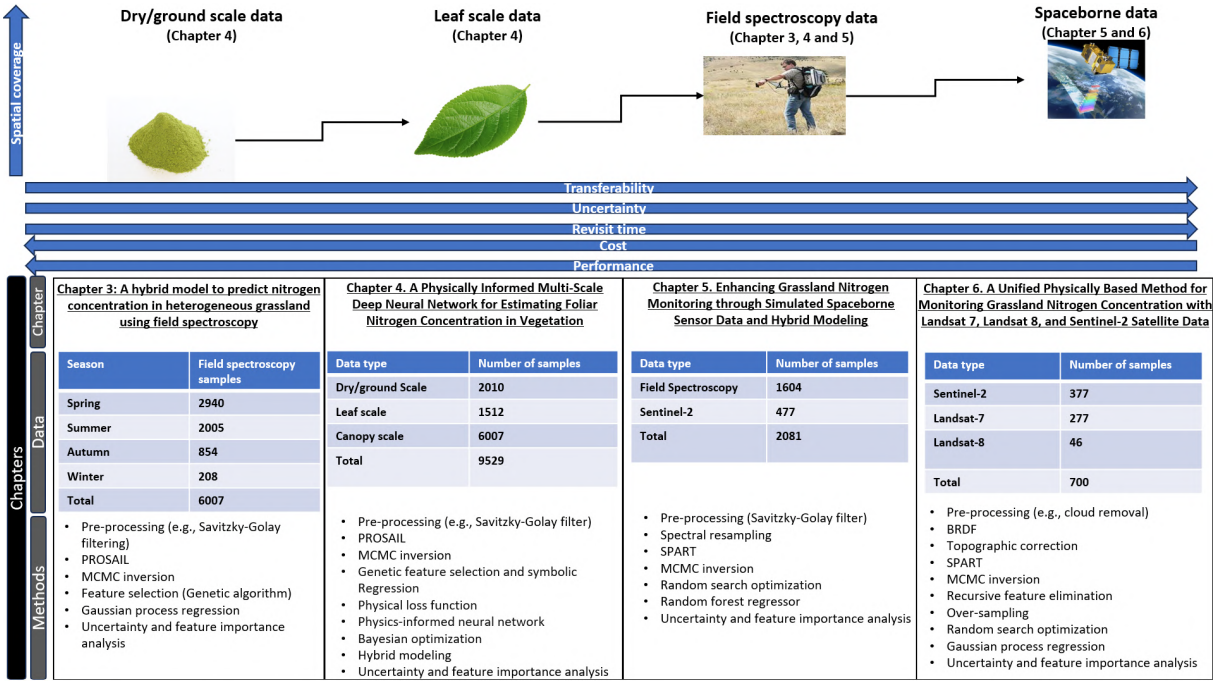


Figure 7.1: Graphical abstract of the research Chapter in this thesis, illustrating the dry/ground-, leaf-, canopy, and satellite-scale observations.

Our analysis reveals that with broader spatial coverage comes increased uncertainty attributed to atmospheric conditions, sensor limitations, and temporal changes in the scenes. In contrast, superior spectral data (e.g., narrower bandwidth) can improve performance. The research methodically outlines specific pre-processing techniques, data inversion methods, feature selection strategies, and regression analyses to address and mitigate these challenges, thereby enhancing data integrity. The research balances spatial coverage and revisits frequency, cost-efficiency, and performance quality, emphasizing the transferability of methods and the quantification of uncertainties. Integrating various data sources and methodologies provides a comprehensive framework, significantly advancing the monitoring of grassland N levels. In the subsequent sections, we will detail the key findings related to our objectives.

7.1 Model accuracy

To enhance N% estimation accuracy in grasslands, our research leverages the comprehensive capabilities of multiscale spectral data, covering the spectrum from 350 to 2500 nm. This approach amalgamates the strengths of various spectroscopic measurements—ranging from proximity to satellite imagery—creating a framework for the consistent and adaptive monitoring of grassland ecosystems. Central to our methodology is the calibration of models to align spectral and

spatial characteristics across various platforms, such as the signal-to-noise ratio (SNR). This calibration process is pivotal, ensuring the fidelity of multiscale spectral data fusion, for precise N% estimations.

Chapter 3 presents a detailed exploration of a substantial dataset compiled from various seasons and locations throughout New Zealand, leading to the formulation of a hybrid methodology (Pullanagari et al., 2021). This methodology marries inverted PROSAIL variables with the information provided by key spectral bands associated with N%, thus normalizing the dataset for uniform application across different temporal and spatial contexts (Jacquemoud et al., 2009b). Our findings show a strong correlation between sample size and model performance, stabilizing at a sample size of n=3500 (Figure 3.7). This plateau elucidates the necessity of curating a representative dataset for the reliable estimation of N% in grasslands, highlighting the efficiency of a thoughtfully designed sampling strategy. Such strategies facilitate the development of more effective models with optimized calibration datasets and that can reduce sampling cost in future monitoring.

Furthermore, our investigations reveal that seasonal variations impact model performance due to physiological alterations within the grassland biome with the best results achieved in the summer and spring. This observation suggests that integrating seasonal data or incorporating physiological insights (e.g., vegetation growth stage) into the modeling framework could enhance model accuracy (Kokaly et al., 2009). Additionally, the inherent heterogeneity of grassland ecosystems, particularly variables such as the Leaf Area Index (LAI), profoundly influences spectral signatures, especially within the Short-Wave Infrared (SWIR) and Near-Infrared (NIR) regions (Figure 3.8). Consequently, enriching our models with biodiversity and species distribution data might refine model performance further (Lausch et al., 2016).

The work in Chapter 3 highlights the economic challenges of agricultural modeling (Figure 7.1). High costs of extensive sampling create a major obstacle for accurate grassland N% estimation. This pushes us to find new ways to reduce sample sizes or use existing datasets to reduce these costs. Chapter 5 explores using field spectroscopy data to calibrate models with Sentinel-2 data (Perich et al., 2021). This method proved useful for creating high-quality datasets for models designed for spaceborne instruments with lower spectral and spatial resolutions (Figure 5.6). However, there is a noticeable decrease in performance compared to direct field spectroscopy, leading to important discussions on the trade-offs in remote sensing methods.

Chapter 5 explores the pivotal role of Radiative Transfer Models (RTM) in bridging the variability in spectral data, such as differences in bandwidth or sampling intervals, which result from various sensor characteristics and environmental conditions like viewing and illumination geometry (Figure 6.2). The RTM's utility in variable inversion and its ability to synthesize high spectral resolution datasets from lower-quality sources, such as Sentinel-2, highlights a potential pathway to ensure uniform spectral data generation across multiple instruments. This approach emphasizes the indispensable accuracy afforded by high-resolution field spectroscopy

data in aligning Sentinel-2 Top Of Atmosphere (TOA) data, particularly when accounting for atmospheric variables influencing spectra. Challenges associated with SWIR inversion and the critical role of chlorophyll in N% estimation are underscored by Sakowska et al. (2016), who highlight the intricate relationship between technological capability and ecological insight in advancing grassland monitoring methodologies.

Chapter 4 explores model performance for N% estimation across diverse observational scales—from the precision of lab and leaf spectroscopy to the scene scales, and scope of field spectroscopy. The consistent performance of these models across varying scales (dry/ground, leaf, canopy, and satellite scales), with a noted enhancement in performance at higher spatial resolutions (Figures 4.6 and 9.4). This phenomenon, attributed not only to the controlled environments and superior spectral data available at these scales but also to the advanced methodologies employed, advocates for integrating varied data sources to refine N% estimation accuracy (Dehghan-Shoar et al., 2024a, Hossain et al., 2024a). By leveraging hybrid methods that combine different data sets and analytical techniques, the overall accuracy and reliability of N% estimation can be significantly improved. Central to this advancement is the application of Physically Informed Neural Networks (PINN), which, by anchoring model performance in physically-based equations using key biochemical variables essential for N% estimation (Figure 4.1). The necessity of incorporating spectral bands with non-linear relationships with N% into the models seems to be important; particularly as Partial Least Squares Regression (PLSR) showcased limited effectiveness compared to PINN. This limitation primarily stems from the requirements for capturing the numerous protein content absorption bands essential for precise N% estimation (Table 2.1). A finding from this inquiry is the potential of leveraging physical principles and functions to enhance model performance, particularly in datasets characterized by noise, by systematically reducing bias (Figure 4.6).

Progressing to the concluding segments of our research, detailed in Chapter 6, we delineate the formulation of a model dedicated to N% estimation through a harmonized methodology integrating data from Landsat 7, 8, and Sentinel-2. This approach, crafted around physical processes and optimized for spaceborne and multispectral imagery challenges, validates its efficacy and feasibility. The analysis reiterates the role of the visible spectrum in N% estimation (Figure 6.7), resonating with observations from Chapter 5. Furthermore, our findings show the substantial impact of incorporating the Bidirectional Reflectance Distribution Function (BRDF) and topographical corrections in elevating model precision, particularly in landscapes where terrain complexity presents significant analytical challenges (Figure 6.6). These revelations underscore the need to address and mitigate the limitations and variabilities inherent in data acquired through different sensors.

The efficacy of physically-based methods as discussed by Berger et al. (2020a) highlights the transformative potential of such hybrid approaches in environmental science. Similarly, the emphasis on the visible spectrum and the critical evaluation of sensor capabilities align with the findings of Perich et al. (2021), who underscored the evolving challenges in remote sensing data

interpretation amidst technological advancements and environmental variability. For instance, Perich et al. (2021) discusses how new sensor technologies can enhance the accuracy of vegetation monitoring but also introduce complexities in data calibration due to varying environmental conditions. By refining and calibrating models to effectively interpret spectral datasets of varied quality, we establish a pathway for advanced agricultural monitoring. Future studies will advance data harmonization techniques and incorporate physically-based deep learning methods to bolster data fusion and model flexibility (Queally et al., 2022, Perich et al., 2021). This will substantially enhance our capacity for agricultural monitoring and management.

Our methodologies introduce several benefits and advantages for monitoring grasslands and enhancing their management, including more accurate vegetation assessment, efficient data collection, and informed decision-making.

1. The rigorous calibration and validation of high-quality data ensure these methodologies are not only transferable across different grassland ecosystems but also adaptable for monitoring grassland N cycles and management in New Zealand. This broad applicability is essential for comprehensive monitoring efforts, offering insights into grassland dynamics, such as growth cycles and N fluxes, worldwide. By linking our methods with established tools and studies (e.g., The Agricultural Production Systems sIMulator (APSIM)) on N management, we enhance our understanding of these ecosystems and contribute to better-informed decisions for sustainable grassland management (Hoffmann et al., 2018).

2. By enabling the use of multispectral data (e.g., Sentinel-2), our models open avenues for more comprehensive agricultural monitoring strategies. This approach allows for utilizing existing datasets and applying multispectral techniques without substantially compromising the accuracy of N% estimations. Our models complement existing environmental monitoring methods by providing a cost-effective and scalable solution for large-scale assessments. They integrate seamlessly with traditional ground-based measurements, offering a holistic view of grassland health and N dynamics. By combining remote sensing data with other methods, we can enhance the precision and efficiency of monitoring efforts, leading to more informed and sustainable grassland management practices.

3. The versatility and accessibility of these advanced modeling tools enhance agricultural monitoring by providing actionable insights. By translating our research into user-friendly applications, farmers can readily utilize these tools for grassland management and conservation efforts. This approach ensures that the benefits of advanced methodologies are accessible to farmers, supporting a collaborative and multi-faceted approach to ecosystem sustainability. By integrating these tools into existing agricultural practices, we can provide farmers with precise, data-driven insights that facilitate informed decision-making, ultimately contributing to more sustainable and resilient grassland ecosystems.

4. As we refine our models and integrate emerging data sources, we anticipate significant

enhancements in the accuracy, efficiency, and applicability of grassland monitoring techniques. By incorporating additional data such as soil moisture, thermal imagery, and hyperspectral data, we aim to develop a more comprehensive understanding of grassland ecosystems. These advancements will improve the precision and timeliness of assessments, aiding in better management and conservation practices. Our objectives include advancing real-time data collection, enhancing sensor technologies, and developing sophisticated machine learning algorithms to further refine our monitoring capabilities and support sustainable grassland management.

5. Our approach facilitates data integration across observational scales, from detailed leaf-level measurements to broader canopy and landscape-level observations. This holistic perspective is indispensable for understanding grassland health and function, enabling the development of more effective management and conservation strategies.

Our research addresses the challenges in grassland N% monitoring and paves the way for future advancements. The harmonized methodologies we propose offer practical and scalable solutions for sustainable ecosystem management and conservation initiatives. As we progress, refining data harmonization techniques and adopting innovative approaches such as physically-based deep learning will enhance the adaptability and effectiveness of our models (Sun et al., 2019). This evolution ensures our methodologies remain at the forefront of agricultural modeling and grassland management, ready to tackle current and future challenges.

While our research demonstrates the effectiveness of integrating multiscale spectral data for N% estimation in grasslands, it is crucial to critically assess the influencing factors and potential limitations. Environmental and physiological variability, such as soil composition, moisture levels, and regional climatic differences, may affect the generalizability of our models across diverse ecosystems. These factors can influence spectral signatures, potentially leading to discrepancies. The reliance on high-quality spectral data necessitates robust calibration protocols to address sensor degradation, atmospheric interference, and viewing geometry variations. Additionally, the economic feasibility of large-scale implementation is a concern, especially in resource-limited regions where costs associated with field sampling, sensor deployment, and data processing can be prohibitive. Future research should focus on validating these models across various geographical and temporal contexts, integrating additional environmental variables, and developing cost-effective, scalable solutions using existing datasets and low-cost sensors.

7.2 Universality and adaptability

Our research aims to advance modeling, specifically for estimating grassland N%. This work represents a significant step forward by combining the flexibility of physically-based models with the performance of empirical-statistical methods. This integration creates a more cohesive and

comprehensive methodological framework. This combination addresses and bridges a critical gap in traditional N% modeling practices, improving our models' accuracy and applicability across diverse grassland ecosystems such as dairy farms.

At the core of our methodology is the application of RTM for the detailed inversion of environmental variables, which are then integrated with spectral bands. This approach ensures both theoretical robustness and empirical validation, enabling our models to perform consistently across heterogeneous grasslands in different locations. Our models achieve contextual sensitivity, accurately identifying essential attributes, such as protein and chlorophyll content, with reasonable precision.

Model transferability is important; however, developing transferable models faces significant challenges due to several factors such as sensor variability and environmental differences (Pullanagari et al., 2021). We demonstrate a significant enhancement in model transferability through extensive validations across various geographical locations, supplemented by seasonal cross-validation (Figure 3.3). This improvement, achieved by integrating RTM-derived variables with specific spectral bands, emphasizes the role of key absorption bands in creating adaptable models suited for diverse environmental conditions. However, our exploration also reveals challenges, particularly during winter, where model transferability is notably limited, likely reflecting physiological changes (e.g., growth stage) within the vegetation (Berger et al., 2020b, Mutanga, 2004a, Mutanga and Skidmore, 2004). This phenomenon, compounded by the variability arising from the mixing of diverse grass species compositions, highlights the need for further research to elucidate the impact of species-specific structural traits on model accuracy.

Our hypothesis posits that precisely identifying fundamental spectral signatures indicative of vital vegetation attributes, such as chlorophyll and protein content, is essential for developing models based on optical reflectance spectroscopy (Table 2.1). However, the widespread applicability of these models across varied environmental conditions is influenced by a complex array of factors, including differences in soil properties, plant species, moisture levels, and atmospheric conditions. Additionally, the quality of spectral and spatial data (e.g., spatial and spectral resolution) and the presence of significant absorption bands play crucial roles. Our preliminary work with field spectroscopy, which involves control over sensor-to-vegetation proximity and consistent lighting conditions, suggests that mitigating the impacts of atmospheric and environmental variables, such as aerosol concentration and cloud cover, could significantly improve model transferability and overall performance (Figure 9.5).

The use of high-quality field spectroscopy data, validated against Sentinel-2 imagery, underscores the crucial role of data fidelity and atmospheric conditions in the efficacy of models. Rooted in the four-stream theory and operationalized through the SPART RTM, this approach enables the inversion of variables across various environmental layers (Yang et al., 2020a). Challenges encountered within the SWIR region, primarily due to atmospheric distortions such as water vapor, which overlaps with important protein bands, highlight a common obstacle in

remote sensing—the difficulty of accurately discerning ground truth amidst fluctuating atmospheric conditions (Figure 5.3).

Despite these challenges, our identification of the red-edge and visible spectral regions as important for accurate N% estimation aligns with the findings of Zarco-Tejada et al. (2021), who emphasized the sensitivity of these spectral regions to chlorophyll content, making them reliable indicators of plant health and nutritional status. This insight underscores the importance of understanding spectral responses to various physiological conditions, such as chlorophyll content and plant water status, as well as environmental conditions, including soil moisture and atmospheric composition. This underscores a common theme in the literature on the use of optical sensing for agricultural modeling: the necessity for comprehensive datasets and robust modeling techniques to accurately capture and predict vegetation dynamics (Berger et al., 2018a, Pullanagari et al., 2021, Berger et al., 2020b).

Our investigation into integrating SPART-based methods for data fusion from Landsat 7, Landsat 8, and Sentinel-2, particularly focusing on BRDF and topographical corrections, addresses an important aspect of remote sensing when trying to develop transferable models (Richter and Schläpfer, 2019). These corrections are essential for enhancing model transferability and accuracy across diverse landscapes, especially in rugged terrains where the visible spectrum is key for grassland N% estimation (Figure 6.7). This methodological approach aligns with the findings of Gupta and Shukla (2020), who highlighted the importance of BRDF corrections in reducing variability introduced by different observation geometries, thus enhancing the consistency of satellite-derived datasets. Our findings illustrate a trade-off between achieving model accuracy and ensuring broad model transferability, encapsulating the inherent challenge in N% modeling: optimizing models for specific contexts while ensuring their applicability across grasslands. This balance is both necessary and desirable, emphasizing the importance of models that can adapt to variability in data without compromising their predictive accuracy.

Incorporating PINN into our modeling framework, as detailed in Chapter 4, represents a significant advancement in grassland N% estimation. This novel integration combines the structural rigor and domain expertise of physically-based models with the versatile accuracy of empirical-statistical models. PINNs notably reduce prediction noise and bias, enhancing the models' adaptability and precision under diverse environmental conditions by effectively capturing non-linear relationships and complexities in the underlying data (Figure 4.6). This approach underscores the critical contribution of physically-based models to developing models that are consistent and universally applicable across grasslands (Dehghan-Shoar et al., 2024b).

The application of PINNs exemplifies the trend of using advanced computational techniques, such as neural networks and physics-based models, to more accurately simulate and predict complex environmental phenomena, thereby addressing challenges like climate change, ecosystem dynamics, and resource management. Similar to the efforts described by Raissi et al. (2019) in non-linear partial differential equations, our work adapts this innovative approach to the

agricultural domain, showcasing its potential to improve predictions in highly variable natural systems. This aligns with the growing recognition within the scientific community of the need for models that accurately represent the dynamic interactions within ecosystems (Camps-Valls et al., 2023).

Our methodology acknowledges the dynamic and complex nature of grassland ecosystems, particularly emphasizing the significant impact of seasonal variations on N% estimation. The physiological changes experienced by grasslands throughout the year, as reflected in their spectral signatures, necessitate the development of adaptable and accurate models across seasons. This requirement aligns with the findings of Thieme et al. (2024), who highlighted the critical role of intercomparison of same-day remote sensing data for measuring winter cover crop biophysical traits. Such studies underscore the importance of accounting for seasonal and sensor variability to enhance predictive accuracy and better understand ecosystem dynamics.

The exploration of data source selection—from lab and field to spaceborne observations—underscores the challenges of achieving high model accuracy and transferability. Our findings, which reveal the effectiveness of models calibrated on high-quality spectral data in improving transferability and reducing costs, contribute to an ongoing dialogue on the trade-offs between data quality and model performance.

7.3 Accounting and quantifying uncertainties in grassland N% prediction models

Accounting and quantifying uncertainties and errors in optical sensing data for grassland N% estimation is one of the objectives of our research. This is crucial from a management perspective because understanding and addressing these uncertainties ensures more accurate and reliable data, leading to better decision-making in grassland management. Accurate estimation of N% is vital for optimizing fertilization strategies, improving crop yields, and minimizing environmental impacts. We sought to find and constrain 'known unknowns' within optical sensing, leading to methodological advancements that enhance N% estimation precision (Povey and Grainger, 2015a). Our strategy prominently involved utilizing Markov Chain Monte Carlo (MCMC) to invert biochemical and biophysical variables through RTM such as SPART and (PROSPECT+SAIL) PROSAIL, quantifying the uncertainties inherent to RTM inversion, notably its ill-posed nature (Chapters 3, 4, 5, and 6) (Figure 5.4) (Dong et al., 2019). We acknowledged and addressed challenges in MCMC, including slow convergence, high computational demand, and sensitivity to initial values and tuning parameters. Strategies such as enhancing convergence diagnostics, employing adaptive MCMC, and applying robust uncertainty quantification were instrumental in mitigating these issues (Cotter et al., 2013). One strategy we utilized was optimizing the initial starting values and upper and lower limits of the search space, enabling faster and more accurate inversions (Chernozhukov and Hong, 2003).

Our approach also refined model performance by constraining highly uncertain MCMC inverted RTM variables and introducing pre-processing and feature selection workflows (Figure 3.6). In addition, methods like spectral averaging and genetic feature selection were key in identifying spectral bands closely linked to N%, particularly those indicative of protein and chlorophyll content (Chapters 3 and 4). These techniques improved model precision by emphasizing relevant spectral features and streamlined the feature selection process, reducing data dimensionality and bolstering computational efficiency. These processing steps identified key spectral bands at 2276 nm, 755 nm, 1526 nm, 2243 nm, and 734 nm for precise N% estimation (Chapter 4) (Cui et al., 2011).

Furthermore, we addressed the influence of terrain, sensor, and illumination geometry on optical remote sensing data, particularly in the NIR spectrum (Chapter 6). Applying topographical and Bidirectional Reflectance Distribution Function (BRDF), corrections were key in mitigating errors from terrain-induced anomalies and the diverse geometries of solar illumination and sensor viewing angles when using spaceborne imagery (Chapter 5) (Figure 6.5) (Lee et al., 2021). However, the overestimation of spectral reflectance in certain terrains led us to advocate for advanced BRDF and topographical correction techniques to ensure consistent performance across diverse instruments and complex terrain-illumination interplays (Chapter 6).

Advanced modeling techniques like Gaussian Process Regression, confidence intervals, and dropout approximation significantly contributed to our model uncertainty quantification, enhancing the robustness and reliability of our analysis framework (Chapters 3, 4, 5, and 6). A notable method adopted was the formulation of a physical loss function using critical spectral bands associated with N% and Genetic Symbolic Regression (GSR) alongside PINN, addressing RTM's ill-posedness challenges and augmenting model resistance to ground truth data noise and bias (Figure 4.6).

Our exploration revealed PINN's comparative robustness and predictive accuracy over classical methodologies like Partial Least Squares Regression (PLSR) (Table 4.5). PINN's integration of physical principles within the learning paradigm empowers the model to assimilate data while considering biophysical variables. We enhanced PINN's capabilities by leveraging GSR to derive the loss function, ensuring it aligns with physical laws and adapts to intricate data patterns. However, we were mindful of GSR's limitations, including sensitivity to data noise and the propensity for model overfitting, especially with high-dimensional datasets (Virgolin et al., 2020, 2021). Addressing these challenges, we implemented noise filtering, complexity penalization, and cross-validation, ensuring the robustness and generalizability of the model developed through this integrated approach.

PINN represents a significant advancement in optical sensing for complex, non-linear systems. Our research also highlighted the importance of seasonal data in reducing model uncertainty due to its impact on grassland physiology, advocating for adaptive, seasonally-aware modeling techniques in heterogeneous grasslands (Chapter 3).

We also addressed 'unknown unknowns'—non-linear systematic errors inherent in remote sensing's approximations and assumptions (Povey and Grainger, 2015a). Our approach involved implementing correction algorithms for systematic errors and integrating advanced algorithms, such as BRDF and topographical correction, to mitigate random spectral noise. This enhanced data reliability and reduced model uncertainty (Chapters 3, 4) (Vögtli et al., 2024, Povey and Grainger, 2015b). Signal processing techniques, such as the Savitzky–Golay filter, were crucial in reducing the impact of random noise. Care was taken to preserve critical data features by optimizing its parameters, ensuring the retention of subtle protein absorption bands in SWIR (Féret et al., 2021).

In this work, we identified different sources of errors and uncertainties in optical remote sensing for modeling grassland N% from various chapters. The main categories of errors include sensor-related, radiometric, geometric, atmospheric, environmental, spectral dataset, model, and ground truth and validation errors (Table 7.1). For example, sensor-related issues such as instrument aging and sensor noise degrade data quality, addressable through regular maintenance and advanced noise reduction techniques. Radiometric errors like calibration inaccuracies require regular calibration, while geometric errors such as image distortions need correction algorithms. Atmospheric conditions like aerosols and cloud cover impact readings and necessitate atmospheric correction techniques. Environmental factors, including canopy heterogeneity and terrain variations, benefit from time-series analysis and topographical corrections. Integrating diverse spectral datasets and minimizing human errors demand robust data integration frameworks and rigorous training. Model errors from assumptions and parameter uncertainty call for continuous validation and robust statistical methods. Finally, ground truth and validation errors, including GPS inaccuracies, highlight the need for enhanced data collection and quality control. These findings emphasize the importance of advanced techniques and attention to improving the accuracy and reliability of optical remote sensing in agricultural modeling.

Remote sensing data encompass numerous errors and uncertainties from various sources that impact spectral data. For example, atmospheric attributes can significantly affect the SWIR due to the presence of water absorption bands, while topography can have a substantial impact on the NIR (Figure 6.6). One of the most critical issues in this domain is the propagation of errors at different stages throughout the remote sensing system. This propagation can skew the estimation of N%, leading to inaccuracies with broad implications for environmental assessment and management. To address these challenges, it is essential to identify, quantify, and constrain these error sources. A key aspect of this approach involves employing atmospheric correction techniques, such as RTM-based inversion of atmospheric variables, which can constrain the impact of atmospheric effects on spectra. Additionally, advanced noise reduction strategies, like spectral averaging, increase overall SNR while reducing high-dimensional spectral redundancy and the impact of noise. These methods are instrumental in refining the quality of collected data, thereby enhancing the reliability of N% estimation (Richter and Fukshansky, 1996, Dehghan-Shoar et al., 2023b).

Table 7.1: Categorization of sources of uncertainty and error in optical remote sensing for modeling grassland nitrogen (N) concentration (N%).

Category	**Source of Uncertainty/Error**	**Description**	**Potential Solutions**
Sensor-Related Errors	Instrument Aging	Degradation of sensor performance over time.	Regular maintenance and sensor replacement.
	Sensor Noise	Electronic noise affecting data quality.	Advanced noise reduction techniques.
	Signal-to-Noise Ratio (SNR)	Low SNR leading to poor data quality.	Enhancing sensor technology and signal processing.
	Spectral Response Variability	Variability in sensor's spectral sensitivity.	Using sensors with higher spectral sensitivity.
Radiometric error	Calibration Errors	Inaccuracies in sensor calibration affecting data validity.	Regular calibration against known standards.
	Metadata Quality	Incomplete or inconsistent metadata affects data usability.	Comprehensive metadata documentation and correction.
Geometric error	Geometric Distortions	Distortions in images due to sensor optics and platform instability.	Use of geometric correction algorithms.
	Viewing Angle Errors	Errors due to the angle at which the sensor views the target.	Angle correction strategies such as BRDF.
Atmospheric error	Aerosols, Scattering, Glint and Gases	Impact of atmospheric particles and gases on readings.	Use of atmospheric correction techniques.
	Cloud Cover and Haze	Cloud interference altering data.	Cloud detection and masking algorithms.
	Sun Angle Variations	Variability in illumination affecting data quality.	Timing data collection to optimize sun angle.
Environmental error	Canopy heterogeneity	Variations in land cover over time, influenced by factors such as sward maturity, grazing regime, grazing load/history, type of animal grazing, and seasonal changes in species composition (e.g., different grasses or clover dominating at different stages).	Time-series analysis, dynamic land cover modeling, and incorporating detailed grazing and land use history.
	Terrain	Influence of topography composition and terrain on spectral data, including variations due to slope, aspect, and altitude.	Undertaking topographical correction and incorporating terrain-related variables in models.
	Moisture Variability	Fluctuations in moisture levels affecting grassland productivity and spectral signatures.	Integrating soil moisture data and monitoring seasonal moisture changes.
Spectral Dataset error	Integration of Diverse Data Sources	Challenges in harmonizing datasets.	Comprehensive data integration frameworks.
	Data Fusion Challenges	Difficulty integrating diverse datasets.	Using Advanced data fusion techniques.
	Human Error	Mistakes in data handling, analysis, or interpretation.	Training programs, automated systems, and double-check protocols.
Model error	Model Assumptions	Assumptions in models leading to errors.	Regular model validation and refinement.
	Parameter Uncertainty	Inaccuracies in model parameter selection.	Robust statistical methods for parameter estimation.
	Model Calibration	Misalignment between model outputs and observations.	Continuous model calibration with ground data.
	Model Complexity	Complex behaviors leading to errors.	Simplifying models using ensemble techniques.
	Seasonal Variations	Seasonal changes affecting features.	Multi-temporal analysis; season-specific modeling.
Ground Truth dataset and Validation error	Ground Truth Accuracy	Discrepancies between remote sensing and ground data.	Enhanced ground truth data collection and validation methods.
	Validation Methodologies	Limitations in existing validation techniques.	Developing robust validation frameworks.
	GPS Inaccuracy	Errors in the geolocation data due to GPS inaccuracies.	Utilizing high-precision GPS systems and integrating correction methods like differential GPS.
	Data Processing Errors	Errors during data processing stages.	Rigorous quality control in data processing.

Employing different correction techniques (e.g., atmospheric correction) alone is insufficient to tackle the multifaceted nature of uncertainties in remote sensing, especially given the various sources of error in optical sensing data. Effective management of these uncertainties necessitates implementing structured frameworks like the Guide to the Expression of Uncertainty in Measurement (GUM) (Sayer et al., 2020). The GUM framework provides a standardized approach to evaluate and express uncertainty, laying a robust foundation for decision-making in environmental monitoring. By embracing the principles of GUM, it becomes possible to systematically assess uncertainties associated with each component of the remote sensing process (Qiu et al., 2023), including atmospheric interference, sensor inaccuracies, and the inherent variabilities within grassland ecosystems. This comprehensive evaluation guides the development of more accurate and reliable remote sensing methodologies.

Throughout our research, we encountered numerous practical examples that highlighted the complexities and challenges in optical sensing for grassland N% estimation. The major sources of error we encountered were atmospheric, geometric, topographic, and modeling errors, which we attempted to quantify and constrain using various techniques. For instance, we observed significant variations in spectral data due to atmospheric conditions, particularly in the presence of aerosols and varying sun angles, necessitating advanced atmospheric correction techniques. Additionally, the heterogeneity of grassland canopies and varying terrain introduced errors that were mitigated through topographical corrections and BRDF adjustments. These real-world examples underscored the importance of robust pre-processing steps, such as spectral averaging and genetic feature selection, crucial for identifying key spectral bands related to protein and chlorophyll content. Moreover, our exploration of MCMC methodologies revealed the necessity of optimizing initial values and tuning parameters to enhance the convergence and accuracy of RTM inversions. While some sources of uncertainty, such as sensor aging, were more challenging to identify and address, these instances illustrate the practical challenges and solutions that informed our approach to improving N% estimation precision in grasslands. As a result, more sophisticated methods need to be adopted in the future to understand how each source of error interacts and to effectively use the GUM framework.

7.4 Implications

Using optical sensing technologies for precise N% estimation in grasslands signifies a substantial leap toward sustainable agriculture (Tilman et al., 1996, Fay et al., 2015, Novoa and Loomis, 1981). By optimizing N inputs, this research advocates for a less intrusive, more sustainable approach to livestock farming, emphasizing the importance of environmental stewardship. This strategy aims to boost productivity in response to growing global food demands while minimizing the adverse environmental impacts traditionally associated with agricultural practices. The findings suggest that the future of agriculture hinges on integrating technological innovation with ecological conservation, creating a model that enhances productivity and food security

while protecting the planet (Novoa and Loomis, 1981, Goetz and Whitehead, 1971, Fay et al., 2015).

This research makes a significant contribution to the understanding of grassland ecosystems and their management by demonstrating how precision technologies, particularly the monitoring of N% using optical sensing, can be employed. As a result, these methods can be used to optimize nutrient content and management, enabling better decision-making in areas such as feed formulation. This not only improves animal health and productivity but also promotes environmental sustainability by reducing methane emissions through more efficient nutrient conversion in livestock. The implications of this research extend beyond agriculture, highlighting the role of precision technologies in monitoring and maintaining biodiversity, soil health, and ecosystem services (FAO, 2020, 2021, Taubert et al., 2012, Tilman et al., 2006a, O'Mara, 2012). This approach aligns with global efforts to combat climate change and biodiversity loss, emphasizing the need for a symbiotic relationship between technological advancement and environmental conservation (Tilman et al., 2006b, Petermann and Buzhdygan, 2021).

Our research has the potential to significantly enhance pasture growth models, which traditionally rely on limited samples with little species diversity (McVeagh, 2019). Due to a lack of better information, these models have become accepted standards but often fall short in day-to-day management, leading farmers to revert to visual assessments. This underscores the importance of Section 7.3 on uncertainty, as farmers need information with reduced uncertainty and greater accuracy. Integrating extensive and diverse spectral data with pasture growth models can address these limitations, making the models more accurate and adaptable to various environmental conditions and management practices. For instance, McVeagh (2019) highlights the shortcomings of historical models and the need for models that account for species diversity across different conditions such as slope, aspect, and altitude. Our models incorporate detailed factors like biochemical and biophysical variables, offering a holistic understanding of grassland dynamics and providing precise, actionable insights for effective management. To further improve predictions, combining spectral data-driven models with traditional pasture growth models can leverage the strengths of both approaches.

The findings from this thesis highlight the versatility and potential of optical sensing and advanced modeling techniques in various scientific applications. For instance, these methods can be adapted to quantify N cycles, providing valuable insights into nutrient dynamics in agricultural and natural ecosystems. The ability to accurately measure and monitor N cycles can lead to a better understanding and management of nutrient use efficiency, minimizing environmental impacts such as water pollution and greenhouse gas emissions.

Furthermore, the integration of these advanced techniques (e.g., PINN and Scientific Machine Learning (SciML)) can enhance precision agriculture by optimizing fertilizer application, thereby improving crop yields and sustainability (Wijmer et al., 2024). This approach also has potential applications in environmental monitoring, such as tracking changes in vegetation

health and detecting early signs of ecosystem stress (Hossain et al., 2024b). By refining these methods and expanding their applications, this research contributes to the development of more accurate and reliable tools for environmental assessment and management. The advancements in optical sensing and modeling not only improve data quality but also provide a foundation for future studies aimed at addressing critical environmental challenges.

Chapter 8

Conclusions and Future Directions

Our research significantly contributes to estimating grassland nitrogen (N) concentration (N%), offering insights and solutions to advance ecological research and sustainable grassland management. This thesis delves into integrating physically-based and empirical-statistical models, aiming to achieve consistent and reliable performance across diverse temporal periods, geographical locations, and sensors with varying specifications. While the proposed models demonstrated strong reliability, uncertainties related to environmental factors and sensor-specific variations were identified and systematically addressed. By acknowledging challenges such as spatial resolution, signal-to-noise ratio (SNR), spectral resolution, and various sources of error, we emphasize the feasibility and necessity of developing unified, sensor-invariant workflows for continuous grassland monitoring.

These workflows can provide a comprehensive understanding of global grassland characteristics, which is especially valuable during seasons with limited data availability due to cloud cover. Table 7.1 categorizes the different sources of uncertainty and error in optical remote sensing for modeling grassland N%, including sensor-related errors, radiometric errors, geometric errors, atmospheric errors, environmental errors, spectral dataset errors, model errors, and ground truth and validation errors. To ensure consistent and accurate model performance across various scenarios, it is crucial to adopt holistic approaches that systematically address and quantify these uncertainties and errors. Future work should utilize well-established uncertainty frameworks, such as the Guide to the Expression of Uncertainty in Measurement (GUM). By incorporating GUM, we can comprehensively understand the impact of different uncertainties and errors on our models, thereby enhancing their reliability and robustness and ultimately advancing grassland monitoring.

Furthermore, integrating our models with physically-based models opens exciting avenues for research and application. This integration enables the simulation of complex grassland interactions under specific climatic conditions and various management strategies. However, a significant challenge in this endeavor is creating high-quality training datasets, which are crucial for accurate vegetation trait estimation from canopy reflectance. To overcome this challenge, we propose harnessing Radiative Transfer Models (RTMs) to generate synthetic yet realistic training datasets. The strength of RTMs lies in their ability to capture our best understanding

of the biochemical and biophysical processes that govern the interaction of light with vegetation. This approach allows for the calibration of high-performance models using both synthetic and real-world datasets. These datasets should mirror high-quality field data and encompass diverse vegetation types, growth stages, and observational configurations. By combining RTMs with crop models, we can potentially generate extensive and highly realistic datasets. These datasets are instrumental in developing sensor-invariant workflows applicable to a wide range of optical sensors, covering the 400- to 2500-nm spectral range.

The importance of such workflows cannot be overstated, as they ensure consistent and accurate monitoring across various sensors and environmental conditions, significantly enhancing the reliability and applicability of our models. Looking ahead, the future holds even more exciting possibilities, particularly in the realm of digital twins for grassland ecosystems. Digital twins—virtual replicas of physical entities—can revolutionize our ability to simulate, understand, and manage grasslands effectively. By harnessing the synergy of different information sources and leveraging extensive data, we aim to create highly realistic digital twins that reflect the intricate complexities of grassland ecosystems. Such systems will have numerous applications, including:

- **What-If Analysis:**
  - Enables the simulation of various scenarios to predict outcomes before implementing changes in the real-world system.
  - Allows testing of different operational strategies to determine the best course of action.
  - Reduces the costs associated with trial-and-error in physical processes.
- **Process Optimization:**
  - Identifies inefficiencies and bottlenecks in current operations to improve productivity.
  - Enhances performance by simulating adjustments in real-time, leading to optimized workflows.
  - Facilitates continuous improvement by providing insights derived from data analytics.
- **Uncertainty Analysis:**
  - Assesses the impact of uncertain variables on system performance to better prepare for variable outcomes.
  - Improves system understanding by simulating various conditions and their effects.
  - Supports robust design and operation of systems by considering a range of possible future states.
- **Risk Assessment:**
  - Predicts potential failures and their consequences, enabling proactive maintenance and risk mitigation.

  - Evaluates the severity and frequency of potential risks, contributing to a safer operational environment.
  - Enhances resilience by modeling the system's behavior under extreme or unexpected conditions.

- **Decision Support:**
  - Provides a data-driven foundation for making informed decisions.
  - Combines historical data with predictive analytics to guide strategic planning.
  - Increases confidence in decision-making by visualizing the outcomes of different choices.

Our upcoming research initiatives will significantly advance grassland ecology by leveraging digital twins and hybrid modeling methods. These innovative approaches will enable us to simulate various N interaction scenarios in grasslands more effectively. Building on our current work, which centers on consistent monitoring through sensor-invariant workflows and RTMs, our future efforts will enhance correction methodologies and integrate PINN algorithms. These developments aim to boost our models' precision, adaptability, and responsiveness, contributing to a deeper ecological understanding and supporting sustainable management of grasslands.

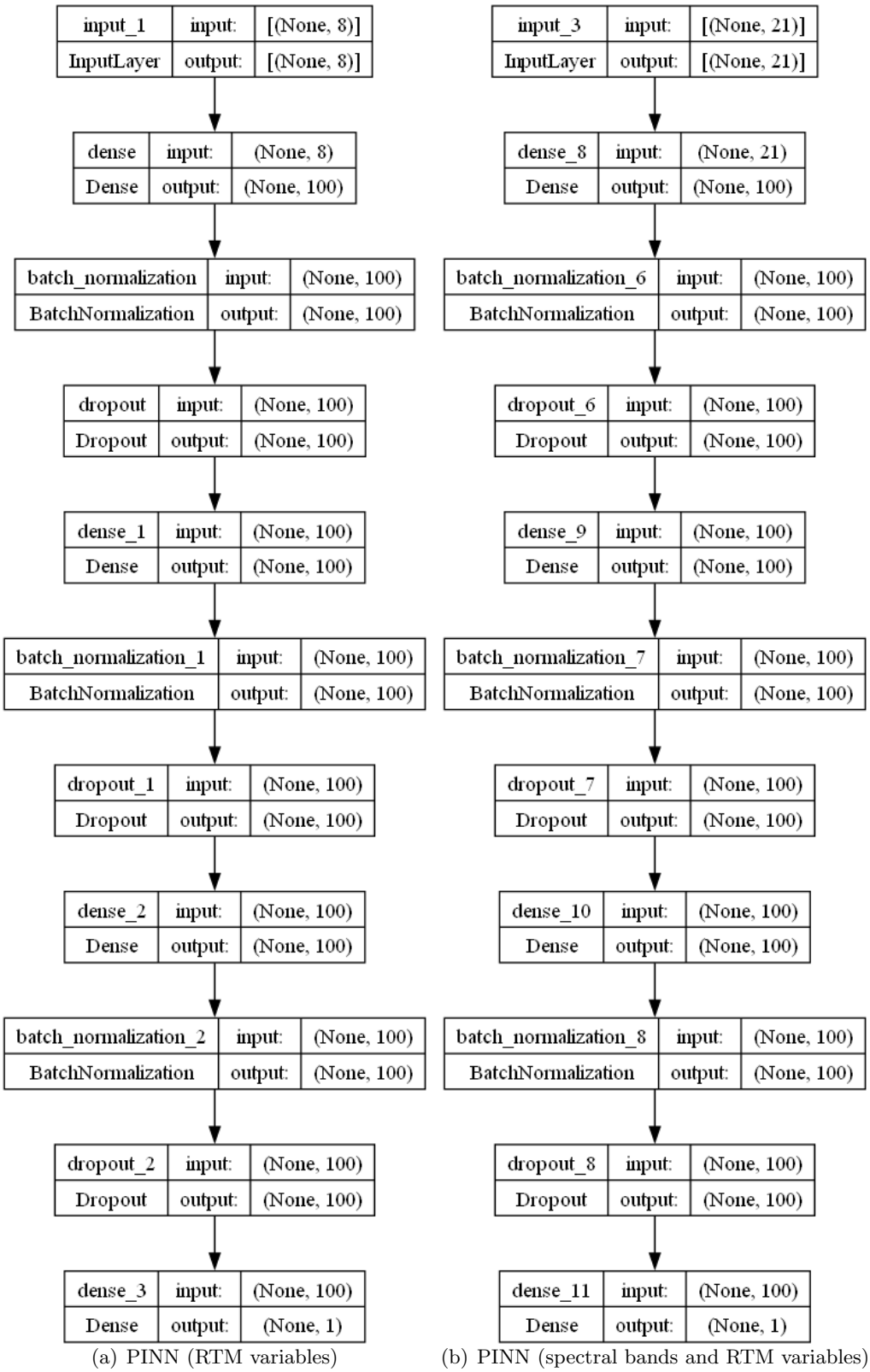
Central to our approach is the development of comprehensive, sensor-invariant workflows. We will create highly realistic digital twins of grassland ecosystems by merging established uncertainty frameworks, physical crop models, and synthetic training datasets from RTMs. These digital twins will facilitate detailed simulations of N interactions, offering a dynamic and comprehensive method for ecological research and monitoring. Our research is poised to make a significant impact in the field of grassland ecology. Combining hybrid modeling techniques, sensor-invariant workflows, and digital twins, we aim to transform grassland ecosystems' understanding and sustainable management, setting a solid foundation for future research and breakthroughs in this vital field.

Chapter 9

Appendix

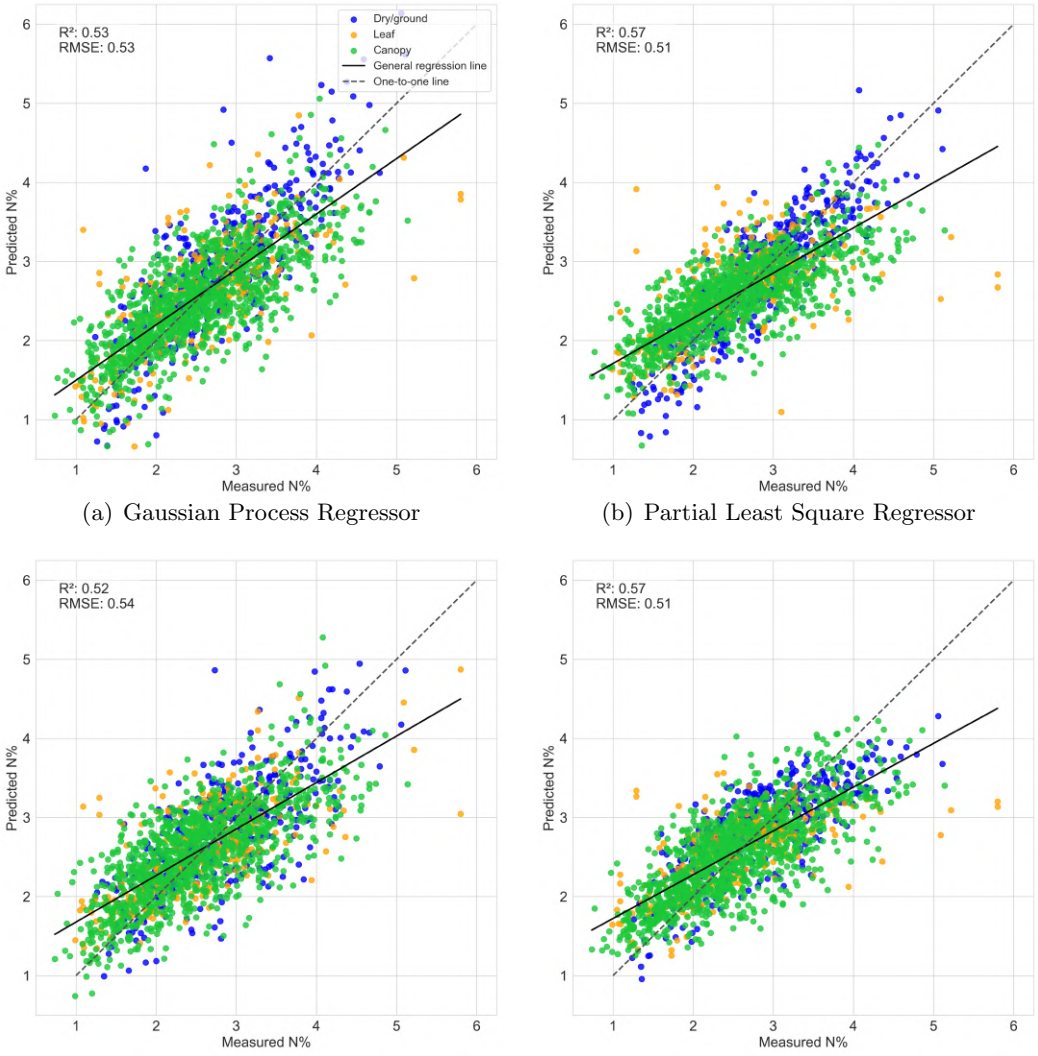
Table 9.1: Table containing the ID list of the acquired Landsat 7 (L7), Landsat 8 (L8), and Sentinel-2 (S2) Top of Atmosphere (TOA) imagery. The imagery was acquired when the cloud coverage was less than 15 percent.

Instrument	**Image ID**
Sentinel-2	COPERNICUS/S2/20160426T221552_20160426T221553_T60GUA
Sentinel-2	COPERNICUS/S2/20190330T222539_20190330T222539_T59GNM
Landsat 8	LANDSAT/LC08/C01/T1_TOA/LC08_073087_20181018
Landsat 7	LANDSAT/LE07/C01/T1_TOA/LE07_07208_20160420
Landsat 7	LANDSAT/LE07/C01/T1_TOA/LE07_072088_20191123



(a) PINN (RTM variables) (b) PINN (spectral bands and RTM variables)

Figure 9.1: Diagram depicting the architecture of a dual-pathway neural network, with each path incorporating batch normalization and dropout layers that converge into dense layers. Image (a) represents the architecture tailored for a Physics-Informed Neural Network (PINN) utilizing solely Radiative Transfer Model (RTM) variables. Image (b) corresponds to the architecture adapted for a PINN that integrates both RTM variables and key spectral bands, indicating a more complex input feature set.



(a) Gaussian Process Regressor

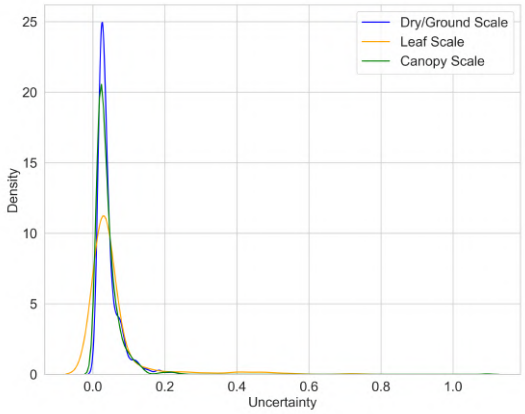
(b) Partial Least Square Regressor

(c) Random Forest Regressor

(d) Support Vector Regressor

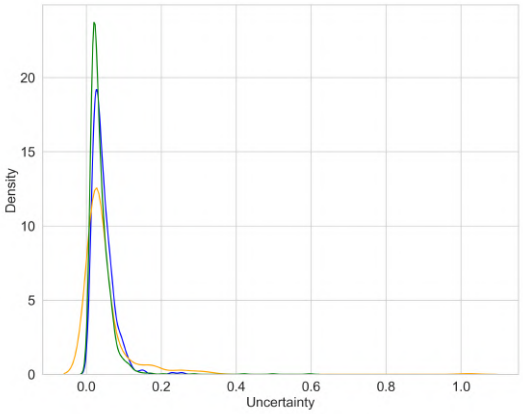
Figure 9.2: Comparative performance of regression models. Subfigures (a), (b), (c), and (d) illustrate the predictive accuracy of four different regression models. (a) Gaussian Process Regression, highlighting its capability to model complex, non-linear relationships with an inherent measure of uncertainty. (b) Partial Least Squares Regression, emphasizing its effectiveness in dealing with highly multivariate and collinear data. (c) Random Forest Regressor, showcases its ensemble learning approach for capturing non-linear interactions and providing robustness against overfitting. (d) Support Vector Regression, demonstrating its application of kernel functions to capture non-linear patterns in data with high dimensionality. Each subfigure includes a scatter plot of the measured vs. predicted values, along with a one-to-one reference line, regression fit, and corresponding R^2 and Root Mean Squared Error (RMSE) values to assess model performance.

Using important spectral bands and Radiative Transfer Model (RTM) variables



(a)

Using only RTM variables



(b)

Figure 9.3: Graphs (a) and (b) detail Physics Informed Neural Network (PINN) model uncertainties highlighting the impact of including or excluding significant spectral bands in PINN.

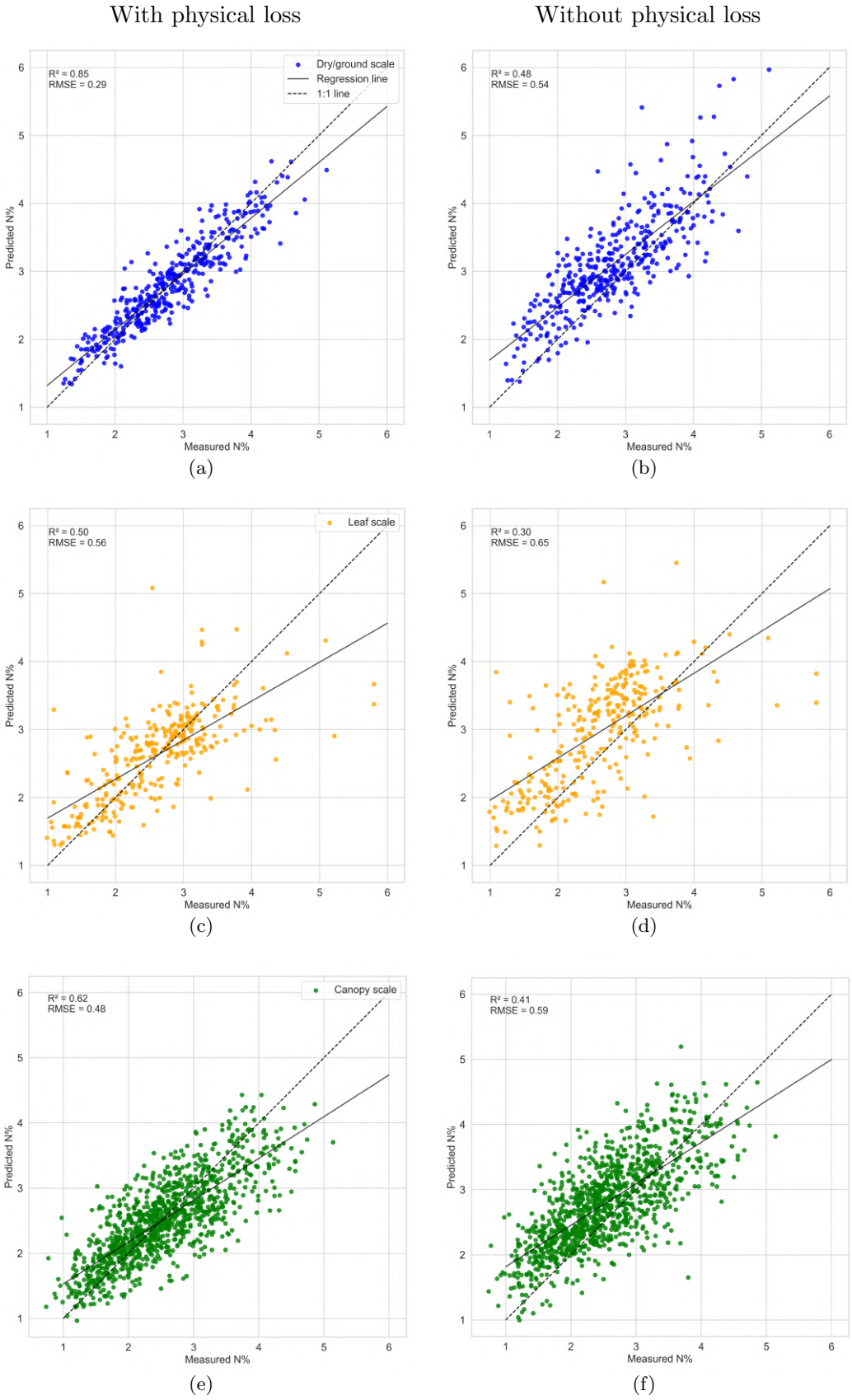


Figure 9.4: Comparison of results obtained by the Physics-Informed Neural Network (PINN) for estimating nitrogen (N) concentration (N%) in the presence of noise, using Radiative Transfer Model (RTM) variables and key spectral bands associated with chlorophyll and protein content across different scales. Subfigures (a) and (b) present results for dry/ground data, while (c–f) show results for leaf and canopy scale data.

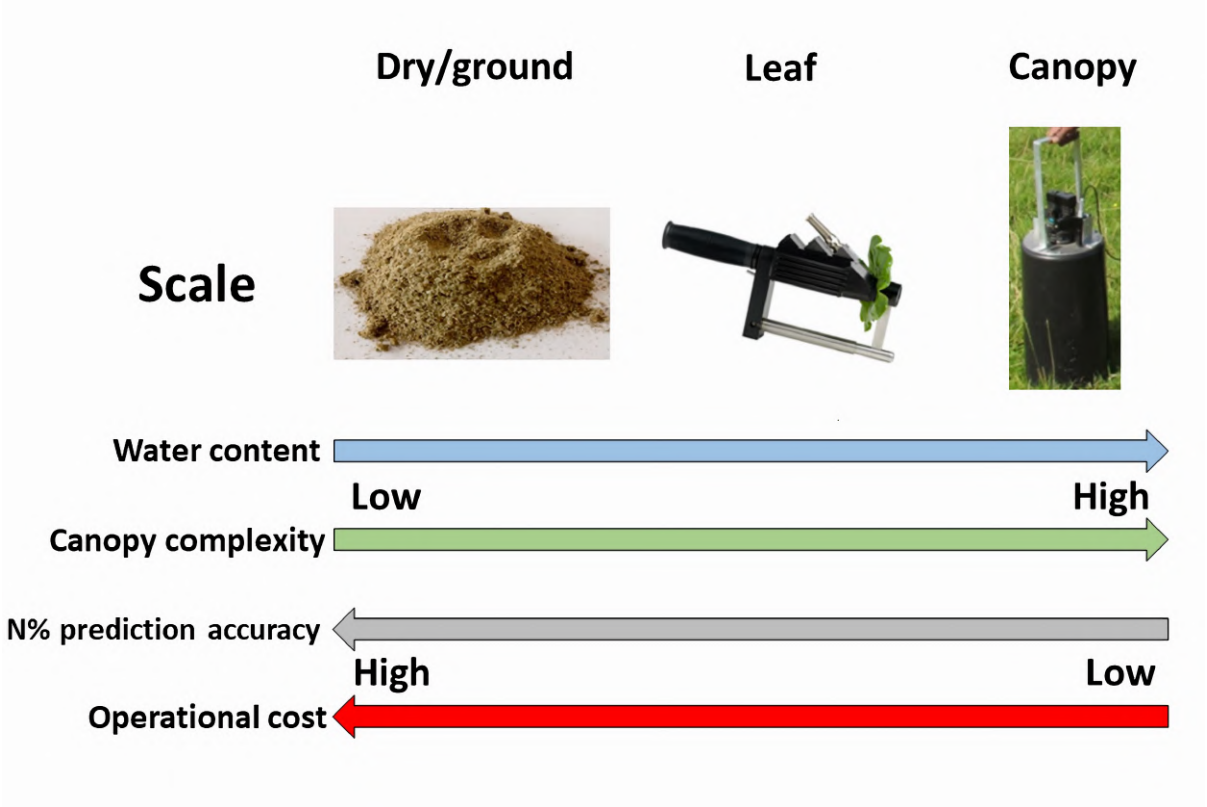


Figure 9.5: The complexity of the canopy structure and its water content may notably affect the accuracy of nitrogen (N) concentration (N%) estimations. However, the increased accuracy linked with dry/ground scale data is associated with heightened operational expenses.

Bibliography

Abdelbary, A. F. (2024). Optimal thermal design parameters of educational remote sensing satellite. *Assiut University Bulletin for Environmental Researches*, 27(2):41–59.

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, 2(4):433–459.

Acito, N., Diani, M., and Corsini, G. (2022). PRISMA Spatial Resolution Enhancement by Fusion With Sentinel-2 Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:62–79.

Aggarwal, S. (2004). Principles of Remote Sensing. *Satellite remote sensing and GIS applications in agricultural meteorology*, 23(2):23–28.

Ahmed, M., Rauf, M., Mukhtar, Z., and Saeed, N. A. (2017). Excessive use of nitrogenous fertilizers: an unawareness causing serious threats to environment and human health. *Environmental Science and Pollution Research*, 24(35):26983–26987.

Alexandratos, N. and Bruinsma, J. (2012). World agriculture towards 2030/2050: the 2012 revision.

Ali, I., Cawkwell, F., Dwyer, E., and Green, S. (2017). Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data—A Machine Learning Approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(7):3254–3264.

Amirruddin, A. D., Muharam, F. M., Ismail, M. H., Tan, N. P., and Ismail, M. F. (2020). Hyperspectral spectroscopy and imbalance data approaches for classification of oil palm's macronutrients observed from frond 9 and 17. *Computers and Electronics in Agriculture*, 178:105768.

Asner, G. P., Martin, R. E., Knapp, D. E., Tupayachi, R., Anderson, C., Carranza, L., Martinez, P., Houcheime, M., Sinca, F., and Weiss, P. (2011). Spectroscopy of canopy chemicals in humid tropical forests. *Remote Sensing of Environment*, 115(12):3587–3598.

Atzberger, C., Darvishzadeh, R., Immitzer, M., Schlerf, M., Skidmore, A., and Le Maire, G. (2015). Comparative analysis of different retrieval methods for mapping grassland leaf area index using airborne imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, 43:19–31.

Atzberger, C., Darvishzadeh, R., Schlerf, M., and Le Maire, G. (2013). Suitability and adaptation of PROSAIL radiative transfer model for hyperspectral grassland studies. *Remote Sensing Letters*, 4(1):55–64.

Atzberger, C. and Richter, K. (2012). Spatially constrained inversion of radiative transfer models for improved LAI mapping from future Sentinel-2 imagery. *Remote Sensing of Environment*, 120:208–218.

Ballschmiter, K. and Katz, J. J. (1969). Infrared study of chlorophyll-chlorophyll and chlorophyll-water interactions. *Journal of the American Chemical Society*, 91(10):2661–2677.

Baranoski, G. V. and Rokne, J. G. (1997). An Algorithmic Reflectance and Transmittance Model for Plant Tissue. In *Computer Graphics Forum*, volume 16, pages C141–C150. Wiley.

Bassi, D., Menossi, M., and Mattiello, L. (2018). Nitrogen supply influences photosynthesis establishment along the sugarcane leaf. *Scientific Reports*, 8(1).

Benhadj, I., Livens, S., Esposito, M., Vercruyssen, N., Van Dijk, C., Soukup, M., Marchi, A. Z., and Maresi, L. (2024). Hyperscout-1 inflight calibration and product validation. *International Journal of Remote Sensing*, 45(7):2486–2517.

Berger, K., Atzberger, C., Danner, M., D'Urso, G., Mauser, W., Vuolo, F., and Hank, T. (2018a). Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: A review study. *Remote Sensing*, 10(1):85.

Berger, K., Atzberger, C., Danner, M., Wocher, M., Mauser, W., and Hank, T. (2018b). Model-based optimization of spectral sampling for the retrieval of crop variables with the PROSAIL model. *Remote Sensing*, 10(12):2063.

Berger, K., Verrelst, J., Féret, J.-B., Hank, T., Wocher, M., Mauser, W., and Camps-Valls, G. (2020a). Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. *International Journal of Applied Earth Observation and Geoinformation*, 92:102174.

Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Wocher, M., Strathmann, M., Danner, M., Mauser, W., and Hank, T. (2020b). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment*, 242:111758.

Bergeret, P. (2017). The future of food and agriculture: Trends and challenges [Note de lecture]. *Bibliographie prospective, n. 29/09*, page 3.

Bergstra, J. and Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13(10):281–305.

Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., and Cox, D. D. (2015). Hyperopt: a Python library for model selection and hyperparameter optimization. *Computational Science amp; Discovery*, 8(1):014008.

Berk, A., Anderson, G. P., Acharya, P. K., Bernstein, L. S., Muratov, L., Lee, J., Fox, M., Adler-Golden, S. M., Chetwynd Jr, J. H., Hoke, M. L., et al. (2006). MODTRAN5: 2006 update. In *Algorithms and technologies for multispectral, hyperspectral, and ultraspectral imagery xii*. SPIE.

Bernhard, G., Mayer, B., Seckmeyer, G., and Moise, A. (1997). Measurements of spectral solar UV irradiance in tropical-Australia. *Journal of Geophysical Research: Atmospheres*, 102(D7):8719–8730.

Bielinis, E., Jozwiak, W., and Robakowski, P. (2015). Modelling of the relationship between the SPAD values and photosynthetic pigments content in Quercus petraea and Prunus serotina leaves. *Dendrobiology*, 73:125–134.

Bishop, M. P., Young, B. W., Colby, J. D., Furfaro, R., Schiassi, E., and Chi, Z. (2019). Theoretical Evaluation of Anisotropic Reflectance Correction Approaches for Addressing Multi-Scale Topographic Effects on the Radiation-Transfer Cascade in Mountain Environments. *Remote Sensing*, 11(23):2728.

Boardman, J. W. (1998). Post-ATREM polishing of AVIRIS apparent reflectance data using EFFORT: a lesson in accuracy versus precision. In *Summaries of the seventh JPL airborne earth science workshop*, volume 1, page 53. JPL Publication Pasadena, CA.

Bonfil, D. J., Michael, Y., Shiff, S., and Lensky, I. M. (2021). Optimizing Top Dressing Nitrogen Fertilization Using VENμS and Sentinel-2 L1 Data. *Remote Sensing*, 13(19):3934.

Bossung, C., Schlerf, M., and Machwitz, M. (2022). Estimation of canopy nitrogen content in winter wheat from Sentinel-2 images for operational agricultural monitoring. *Precision Agriculture*, 23(6):2229–2252.

Branco, P., Torgo, L., and Ribeiro, R. P. (2017). SMOGN: a pre-processing approach for imbalanced regression. In Luís Torgo, P. B. and Moniz, N., editors, *Proceedings of the First International Workshop on Learning with Imbalanced Domains: Theory and Applications*, volume 74 of *Proceedings of Machine Learning Research*, pages 36–50. PMLR.

Briggs, J. M., Knapp, A. K., Blair, J. M., Heisler, J. L., Hoch, G. A., Lett, M. S., and McCARRON, J. K. (2005). An Ecosystem in Transition: Causes and Consequences of the Conversion of Mesic Grassland to Shrubland. *BioScience*, 55(3):243.

Brook, B. W., Buettel, J. C., and Hong, S. (2021). Constrained scenarios for twenty-first century human population size based on the empirical coupling to economic growth.

Buchhorn, M., Raynolds, M. K., and Walker, D. A. (2016). Influence of BRDF on NDVI and biomass estimations of Alaska Arctic tundra. *Environmental Research Letters*, 11(12):125002.

Bulgarelli, B. and Zibordi, G. (2018). On the detectability of adjacency effects in ocean color remote sensing of mid-latitude coastal environments by SeaWiFS, MODIS-A, MERIS, OLCI, OLI and MSI. *Remote Sensing of Environment*, 209:423–438.

Buster, G., Bannister, M., Habte, A., Hettinger, D., Maclaurin, G., Rossol, M., Sengupta, M., and Xie, Y. (2022). Physics-guided machine learning for improved accuracy of the National Solar Radiation Database. *Solar Energy*, 232:483–492.

Calzolari, M. (2021). manuel-calzolari/sklearn-genetic: sklearn-genetic 0.4.0.

Campbell, C. E. (1972). Use of the planck law in remote sensing. In *Remote Sensing of Earth Resources and the Environment*, volume 27, pages 49–56. SPIE.

Camps-Valls, G., Gerhardus, A., Ninad, U., Varando, G., Martius, G., Balaguer-Ballester, E., Vinuesa, R., Diaz, E., Zanna, L., and Runge, J. (2023). Discovering causal relations and equations from data. *Physics Reports*, 1044:1–68.

Cao, M. and Weng, Q. (2024). Embedded physical constraints in machine learning to enhance vegetation phenology prediction. *GIScience & Remote Sensing*, 61(1):2426598.

Ceamanos, X. and Douté, S. (2010). Spectral Smile Correction of CRISM/MRO Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*.

Chakraborty, R., Kereszturi, G., Pullanagari, R., Durance, P., Ashraf, S., and Anderson, C. (2022). Mineral prospecting from biogeochemical and geological information using hyperspectral remote sensing - Feasibility and challenges. *Journal of Geochemical Exploration*, 232:106900.

Chen, J.-B., Dong, C.-C., Yao, X.-D., and Wang, W. (2018). Effects of nitrogen addition on plant biomass and tissue elemental content in different degradation stages of temperate steppe in northern China. *Journal of Plant Ecology*, 11(5):730–739.

Chen, Q., Zheng, B., Chen, T., and Chapman, S. C. (2022). Integrating a crop growth model and radiative transfer model to improve estimation of crop traits based on deep learning. *Journal of Experimental Botany*, 73(19):6558–6574.

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation. *Journal of econometrics*, 115(2):293–346.

Chi, H., Yan, K., Yang, K., Du, S., Li, H., Qi, J., and Zhou, W. (2021). Evaluation of topographic correction models based on 3-D radiative transfer simulation. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5.

Christophe, E., Léger, D., and Mailhes, C. (2005). Quality criteria benchmark for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 43(9):2103–2114.

Clark, R. N. (1983). Spectral properties of mixtures of montmorillonite and dark carbon grains: Implications for remote sensing minerals containing chemically and physically adsorbed water. *Journal of Geophysical Research: Solid Earth*, 88(B12):10635–10644.

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., and Justice, C. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219:145–161.

Clevers, J. G. and Gitelson, A. A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation*, 23:344–351.

Clevers, J. G. and Kooistra, L. (2012). Using Hyperspectral Remote Sensing Data for Retrieving Canopy Chlorophyll and Nitrogen Content. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):574–583.

Colby, J. D. (1991). Topographic normalization in rugged terrain. *Photogrammetric Engineering and Remote Sensing*, 57(5):531–537.

Conel, J. E., Green, R. O., Vane, G., Bruegge, C. J., Alley, R. E., and Curtiss, B. J. (1987). AIS-2 radiometry and a comparison of methods for the recovery of ground reflectance. In *Proceedings of the 3rd airborne imaging spectrometer data analysis workshop*.

Conese, C., Gilabert, M., Maselli, F., and Bottai, L. (1993). Topographic normalization of TM scenes through the use of an atmospheric correction method and digital terrain models. *Photogrammetric Engineering and Remote Sensing*, 59(12):1745–1753.

Cosgrove, G., Betteridge, K., Thomas, V., and Corson, D. (1998). A sampling strategy for estimating dairy pasture quality. In *PROCEEDINGS-NEW ZEALAND SOCIETY OF ANIMAL PRODUCTION*, volume 58, pages 25–28. New Zealand Society of Animal Prod Publ.

Cotter, S. L., Roberts, G. O., Stuart, A. M., and White, D. (2013). MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. *Statistical Science*, 28(3).

Cracknell, A. P. (1998). Review article Synergy in remote sensing-what's in a pixel? *International Journal of Remote Sensing*, 19(11):2025–2047.

Cui, M., Prasad, S., Mahrooghy, M., Bruce, L. M., and Aanstoos, J. (2011). Genetic algorithms and linear discriminant analysis based dimensionality reduction for remotely sensed image analysis. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 2373–2376. IEEE.

Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30(3):271–278.

Danner, M., Locherer, M., Hank, T., Richter, K., et al. (2015). Spectral Sampling with the ASD FIELDSPEC 4. *Remote Sensing*, 10(12):2063.

Darawsheh, M. and Bouranis, D. (2006). Season vs. nutrition-dependent fruit loading: Effects on pigment dynamics of tomato leaves. *Journal of Plant Nutrition*, 29(4):699–715.

Darvishzadeh, R., Atzberger, C., Skidmore, A., and Schlerf, M. (2011). Mapping grassland leaf area index with airborne hyperspectral imagery: A comparison study of statistical approaches and inversion of radiative transfer models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6):894–906.

Darvishzadeh, R., Skidmore, A., Schlerf, M., and Atzberger, C. (2008a). Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sensing of Environment*, 112(5):2592–2604.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., and Cho, M. (2008b). LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(4):409–426.

Darvishzadeh, R., Skidmore, A. K., Mirzaie, M., Atzberger, C., and Schlerf, M. (2014). Fresh biomass estimation in heterogeneous grassland using hyperspectral measurements and multivariate statistical analysis. In *InAGU Fall Meeting Abstracts*, volume 1.

Davidson, A., Wang, S., and Wilmshurst, J. (2006). Remote sensing of grassland–shrubland vegetation water content in the shortwave domain. *International Journal of Applied Earth Observation and Geoinformation*, 8(4):225–236.

de Oliveira, L. M., Galvão, L. S., and Ponzoni, F. J. (2021). Topographic effects on the determination of hyperspectral vegetation indices: a case study in southeastern Brazil. *Geocarto International*, 36(19):2186–2203.

de Wit, A. J. (1999). Application of a genetic algorithm for crop model steering using NOAA-AVHRR data. In *Remote sensing for earth science, ocean, and sea ice applications*, volume 3868, pages 167–181. SPIE.

Dechant, B., Cuntz, M., Vohland, M., Schulz, E., and Doktor, D. (2017). Estimation of photosynthesis traits from leaf reflectance spectra: Correlation to nitrogen content as the dominant mechanism. *Remote Sensing of Environment*, 196:279–292.

Dehghan-Shoar, M. H. (2018). *Use of new generation geospatial data and technology for low cost drought monitoring and SDG reporting solution: a thesis presented in partial fulfillment of the requirement for the degree of Master of Science in Computer Science at Massey University, Manawatū, New Zealand*. PhD thesis, Massey University.

Dehghan-Shoar, M. H., Kereszturi, G., Pullanagari, R. R., Orsi, A. A., Yule, I. J., and Hanly, J. (2024a). A physically informed multi-scale deep neural network for estimating foliar nitrogen concentration in vegetation. *International Journal of Applied Earth Observation and Geoinformation*, 130:103917.

Dehghan-Shoar, M. H., Kereszturi, G., Pullanagari, R. R., Orsi, A. A., Yule, I. J., and Hanly, J. (2024b). A physically informed multi-scale deep neural network for estimating foliar nitrogen concentration in vegetation. *International Journal of Applied Earth Observation and Geoinformation*, 130:103917.

Dehghan-Shoar, M. H., Orsi, A. A., Pullanagari, R. R., and Yule, I. J. (2023a). A hybrid model to predict nitrogen concentration in heterogeneous grassland using field spectroscopy. *Remote Sensing of Environment*, 285:113385.

Dehghan-Shoar, M. H., Pullanagari, R. R., Kereszturi, G., Orsi, A. A., Yule, I. J., and Hanly, J. (2023b). A Unified Physically Based Method for Monitoring Grassland Nitrogen Concentration with Landsat 7, Landsat 8, and Sentinel-2 Satellite Data. *Remote Sensing*, 15(10):2491.

Dehghan-Shoar, M. H., Pullanagari, R. R., Orsi, A. A., and Yule, I. J. (2023c). Simulating spaceborne imaging to retrieve grassland nitrogen concentration. *Remote Sensing Applications: Society and Environment*, 29:100912.

Delloye, C., Weiss, M., and Defourny, P. (2018). Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. *Remote Sensing of Environment*, 216:245–261.

Dhakal, R., Maimaitijiang, M., Chang, J., and Caffe, M. (2023). Utilizing Spectral, Structural and Textural Features for Estimating Oat Above-Ground Biomass Using UAV-Based Multispectral Data and Machine Learning. *Sensors*, 23(24):9708.

Di, H. and Cameron, K. (2002). Nitrate leaching in temperate agroecosystems: sources, factors and mitigating strategies. *Nutrient cycling in agroecosystems*, 64:237–256.

Diner, D. J. and Martonchik, J. V. (1985). Influence of Aerosol Scattering on Atmospheric Blurring of Surface Features. *IEEE Transactions on Geoscience and Remote Sensing*, GE-23(5):618–624.

D'Odorico, P., Gonsamo, A., Damm, A., and Schaepman, M. E. (2013). Experimental Evaluation of Sentinel-2 Spectral Response Functions for NDVI Time-Series Continuity. *IEEE Transactions on Geoscience and Remote Sensing*, 51(3):1336–1348.

Dong, T., Liu, J., Shang, J., Qian, B., Ma, B., Kovacs, J. M., Walters, D., Jiao, X., Geng, X., and Shi, Y. (2019). Assessment of red-edge vegetation indices for crop leaf area index estimation. *Remote Sensing of Environment*, 222:133–143.

Dorigo, W. A., Zurita-Milla, R., de Wit, A. J., Brazile, J., Singh, R., and Schaepman, M. E. (2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation*, 9(2):165–193.

Dozier, J. and Frew, J. (1981). Atmospheric corrections to satellite radiometric data over rugged terrain. *Remote Sensing of Environment*, 11:191–205.

Dusseux, P., Guyet, T., Pattier, P., Barbier, V., and Nicolas, H. (2022). Monitoring of grassland productivity using Sentinel-2 remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 111:102843.

Dymond, J., Shepherd, J., and Qi, J. (2001). A Simple Physical Model of Vegetation Reflectance for Standardising Optical Satellite Imagery. *Remote Sensing of Environment*, 77(2):230–239.

Elachi, C. and Van Zyl, J. J. (2021). *Introduction to the physics and techniques of remote sensing*. John Wiley & Sons.

Elminir, H. K. (2007). Relative influence of weather conditions and air pollutants on solar radiation–Part 2: Modification of solar radiation over urban and rural sites. *Meteorology and Atmospheric Physics*, 96(3-4):257–264.

Elvidge, C. D. (1990). Visible and near infrared reflectance characteristics of dry plant materials. *International Journal of Remote Sensing*, 11(10):1775–1795.

Estévez, J., Berger, K., Vicent, J., Rivera-Caicedo, J. P., Wocher, M., and Verrelst, J. (2021). Top-of-Atmosphere Retrieval of Multiple Crop Traits Using Variational Heteroscedastic Gaussian Processes within a Hybrid Workflow. *Remote Sensing*, 13(8):1589.

Estévez, J., Salinero-Delgado, M., Berger, K., Pipia, L., Rivera-Caicedo, J. P., Wocher, M., Reyes-Muñoz, P., Tagliabue, G., Boschetti, M., and Verrelst, J. (2022). Gaussian processes retrieval of crop traits in Google Earth Engine based on Sentinel-2 top-of-atmosphere data. *Remote Sensing of Environment*, 273:112958.

Estévez, J., Vicent, J., Rivera-Caicedo, J. P., Morcillo-Pallarés, P., Vuolo, F., Sabater, N., Camps-Valls, G., Moreno, J., and Verrelst, J. (2020). Gaussian processes retrieval of LAI from Sentinel-2 top-of-atmosphere radiance data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167:289–304.

FAO (2011). FAOSTAT. Accessed: 2025-01-01.

FAO (2020). Global livestock production. Accessed: 2025-01-01.

FAO (2021). Animal production. Accessed: 2025-01-01.

Fay, P. A., Prober, S. M., Harpole, W. S., Knops, J. M. H., Bakker, J. D., Borer, E. T., Lind, E. M., MacDougall, A. S., Seabloom, E. W., Wragg, P. D., et al. (2015). Grassland productivity limited by multiple nutrients. *Nature Plants*, 1(7):15080.

Féret, J.-B., Berger, K., de Boissieu, F., and Malenovskỳ, Z. (2021). PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents. *Remote Sensing of Environment*, 252:112173.

Feret, J.-B., François, C., Asner, G. P., Gitelson, A. A., Martin, R. E., Bidel, L. P., Ustin, S. L., le Maire, G., and Jacquemoud, S. (2008). PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment*, 112(6):3030–3043.

Fernández-Habas, J., Moreno, A. M. G., Hidalgo-Fernández, M. T., Leal-Murillo, J. R., Oar, B. A., Gómez-Giráldez, P. J., González-Dugo, M. P., and Fernández-Rebollo, P. (2021). Investigating the potential of Sentinel-2 configuration to predict the quality of Mediterranean permanent grasslands in open woodlands. *Science of The Total Environment*, 791:148101.

Foreman-Mackey, D., Farr, W. M., Sinha, M., Archibald, A. M., Hogg, D. W., Sanders, J. S., Zuntz, J., Williams, P. K., Nelson, A. R., de Val-Borro, M., et al. (2019). emcee v3: A Python ensemble sampling toolkit for affine-invariant MCMC. *Journal of Open Source Software*, 4(43):1864.

Foster, A., Kakani, V., and Mosali, J. (2017). Estimation of bioenergy crop yield and N status by hyperspectral canopy reflectance and partial least square regression. *Precision Agriculture*, 18(2):192–209.

Fourty, T., Baret, F., Jacquemoud, S., Schmuck, G., and Verdebout, J. (1996). Leaf optical properties with explicit description of its biochemical composition: Direct and inverse problems. *Remote Sensing of Environment*, 56(2):104–117.

Frampton, W. J., Dash, J., Watmough, G., and Milton, E. J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82:83–92.

Franch, B., Vermote, E., Skakun, S., Roger, J.-C., Masek, J., Ju, J., Villaescusa-Nadal, J. L., and Santamaria-Artigas, A. (2019). A method for Landsat and Sentinel 2 (HLS) BRDF normalization. *Remote Sensing*, 11(6):632.

Franch, B., Vermote, E., Skakun, S., Roger, J.-C., Santamaria-Artigas, A., Villaescusa-Nadal, J. L., and Masek, J. (2018). Toward Landsat and Sentinel-2 BRDF Normalization and Albedo Estimation: A Case Study in the Peruvian Amazon Forest. *Frontiers in Earth Science*, 6.

Friedl, M., Schimel, D., Michaelsen, J., Davis, F., and Walker, H. (1994). Estimating grassland biomass and leaf area index using ground and satellite data. *International Journal of Remote Sensing*, 15(7):1401–1420.

Frommer, J., Voegelin, A., Dittmar, J., Marcus, M. A., and Kretzschmar, R. (2011). Biogeochemical processes and arsenic enrichment around rice roots in paddy soil: results from micro-focused X-ray spectroscopy. *European Journal of Soil Science*, 62(2):305–317.

Fukshansky, L., Fukshansky-Kazarinova, N., and Remisowsky, A. M. v. (1991). Estimation of optical parameters in a living tissue by solving the inverse problem of the multiflux radiative transfer. *Applied Optics*, 30(22):3145.

Fung, A. K., Li, Z., and Chen, K.-S. (1992). Backscattering from a randomly rough dielectric surface. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2):356–369.

Gao, B.-C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3):257–266.

Gao, B.-C., Davis, C., and Goetz, A. (2006). A review of atmospheric correction techniques for hyperspectral remote sensing of land surfaces and ocean color. In *2006 IEEE International Symposium on Geoscience and Remote Sensing*, pages 1979–1981. IEEE.

Gao, B.-C., Heidebrecht, K. B., and Goetz, A. F. (1993). Derivation of scaled surface reflectances from AVIRIS data. *Remote Sensing of Environment*, 44(2-3):165–178.

Gao, B.-C., Montes, M. J., Davis, C. O., and Goetz, A. F. (2009). Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. *Remote Sensing of Environment*, 113:S17–S24.

Gao, J., Liang, T., Liu, J., Yin, J., Ge, J., Hou, M., Feng, Q., Wu, C., and Xie, H. (2020a). Potential of hyperspectral data and machine learning algorithms to estimate the forage carbon-nitrogen ratio in an alpine grassland ecosystem of the Tibetan Plateau. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163:362–374.

Gao, J., Liu, J., Liang, T., Hou, M., Ge, J., Feng, Q., Wu, C., and Li, W. (2020b). Mapping the Forage Nitrogen-Phosphorus Ratio Based on Sentinel-2 MSI Data and a Random Forest Algorithm in an Alpine Grassland Ecosystem of the Tibetan Plateau. *Remote Sensing*, 12(18):2929.

Gao, Y. and Zhang, W. (2009). A simple empirical topographic correction method for ETM+ imagery. *International Journal of Remote Sensing*, 30(9):2259–2275.

Gastellu-Etchegorry, J., Martin, E., and Gascon, F. (2004). DART: a 3D model for simulating satellite images and studying surface radiation budget. *International Journal of Remote Sensing*, 25(1):73–96.

Gastellu-Etchegorry, J.-P., Lauret, N., Yin, T., Landier, L., Kallel, A., Malenovskỳ, Z., Al Bitar, A., Aval, J., Benhmida, S., Qi, J., et al. (2017). DART: recent advances in remote sensing data modeling with atmosphere, polarization, and chlorophyll fluorescence. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(6):2640–2649.

Gastellu-Etchegorry, J.-P., Yin, T., Lauret, N., Grau, E., Rubio, J., Cook, B. D., Morton, D. C., and Sun, G. (2016). Simulation of satellite, airborne and terrestrial LiDAR with DART (I): Waveform simulation with quasi-Monte Carlo ray tracing. *Remote Sensing of Environment*, 184:418–435.

Ge, Y., Atefi, A., Zhang, H., Miao, C., Ramamurthy, R. K., Sigmon, B., Yang, J., and Schnable, J. C. (2019). High-throughput analysis of leaf physiological and chemical traits with VIS–NIR–SWIR spectroscopy: a case study with a maize diversity panel. *Plant Methods*, 15(1).

Gitelson, A. A., Keydan, G. P., and Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, 33(11).

Goetz, A. F., Heidebrecht, K. B., Kindel, B., and Boardman, J. W. (1998). Using ground spectral irradiance for model correction of AVIRIS data. In *Summaries of the Seventh JPL Airborne Earth Science Workshop January 12-16, 1998*, volume 1.

Goetz, H. and Whitehead, D. C. (1971). The Role of Nitrogen in Grassland Productivity. *Ecology*, 52(4):730.

Gómez-Dans, J. L., Lewis, P. E., and Disney, M. (2016). Efficient Emulation of Radiative Transfer Codes Using Gaussian Processes and Application to Land Surface Parameter Inferences. *Remote Sensing*, 8(2):119.

Good, A. G. and Beatty, P. H. (2011). Fertilizing nature: a tragedy of excess in the commons. *PLoS biology*, 9(8):e1001124.

Goodman, J. and Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27.

Govaerts, Y. M. and Verstraete, M. M. (1998). Raytran: A Monte Carlo ray-tracing model to compute light scattering in three-dimensional heterogeneous media. *IEEE Transactions on Geoscience and Remote Sensing*, 36(2):493–505.

Goward, S. N., Masek, J. G., Williams, D. L., Irons, J. R., and Thompson, R. (2001). The Landsat 7 mission: Terrestrial research and applications for the 21st century. *Remote Sensing of Environment*, 78(1-2):3–12.

GPy (since 2012). GPy: A gaussian process framework in python.

Grantbuster, Rossol, M., Bannister, M., and Hettinger, D. (2021). NREL/phygnn: Release for DOI.

Grzybowski, M., Wijewardane, N. K., Atefi, A., Ge, Y., and Schnable, J. C. (2021). Hyperspectral reflectance-based phenotyping for quantitative genetics in crops: progress and challenges. *Plant Communications*, 2(4):100209.

Gu, D. and Gillespie, A. (1998). Topographic Normalization of Landsat TM Images of Forest Based on Subpixel Sun–Canopy–Sensor Geometry. *Remote Sensing of Environment*, 64(2):166–175.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., et al. (2015). The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sensing*, 7(7):8830–8857.

Guanter, L., Richter, R., and Moreno, J. (2006). Spectral calibration of hyperspectral imagery using atmospheric absorption features. *Applied Optics*, 45(10):2360.

Guerini Filho, M., Kuplich, T. M., and Quadros, F. L. D. (2020). Estimating natural grassland biomass by vegetation indices using Sentinel 2 remote sensing data. *International Journal of Remote Sensing*, 41(8):2861–2876.

Guo, Q., Hu, Z.-m., Li, S.-g., Yu, G.-r., Sun, X.-m., Li, L.-h., Liang, N.-s., and Bai, W.-m. (2016). Exogenous N addition enhances the responses of gross primary productivity to individual precipitation events in a temperate grassland. *Scientific Reports*, 6(1).

Gupta, S. K. and Shukla, D. P. (2020). Evaluation of topographic correction methods for LULC preparation based on multi-source DEMs and Landsat-8 imagery. *Spatial Information Research*, 28(1):113–127.

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, 46:389–422.

Han, X., Shen, J., Yin, P., Hu, S., and Bi, D. (2014). Influences of refractive index on forward light scattering. *Optics Communications*, 316:198–205.

Hank, T. B., Berger, K., Bach, H., Clevers, J. G., Gitelson, A., Zarco-Tejada, P., and Mauser, W. (2019). Spaceborne imaging spectroscopy for sustainable agriculture: Contributions and challenges. *Surveys in Geophysics*, 40:515–551.

He, B., Quan, X., Xu, D., Yin, C., Liao, Z., Qiu, S., Ge, J., and Zhang, Z. (2017). Retrieving Grassland Canopy Water Content by Considering the Information from Neighboring Pixels. *Photogrammetric Engineering amp; Remote Sensing*, 83(8):553–565.

He, L., Li, A., Yin, G., Nan, X., and Bian, J. (2019). Retrieval of grassland aboveground biomass through inversion of the PROSAIL model with MODIS imagery. *Remote Sensing*, 11(13):1597.

Hoffmann, M. P., Isselstein, J., Rötter, R. P., and Kayser, M. (2018). Nitrogen management in crop rotations after the break-up of grassland: Insights from modelling. *Agriculture, ecosystems & environment*, 259:28–44.

Homolova, L., Malenovskỳ, Z., Clevers, J. G., García-Santos, G., and Schaepman, M. E. (2013). Review of optical-based remote sensing for plant trait mapping. *Ecological Complexity*, 15:1–16.

Hosgood, B., Jacquemoud, S., Andreoli, G., Verdebout, J., Pedrini, G., and Schmuck, G. (1995). Leaf optical properties experiment 93 (LOPEX93). *Report EUR*, 16095:1–46.

Hossain, D.-S. M., Gabor, K., Pullanagari, R., Ian, J. Y., Alvaro, O., and James, H. (2024a). Multi-scale estimation of vegetation nitrogen concentration using a physically informed neural network. In *2024 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS)*, pages 1–3. IEEE.

Hossain, D.-S. M., Pullanagari, R., Alvaro, O., and Ian, J. Y. (2024b). Extraction of solar-induced fluorescence (sif) from airborne hyperspectral data. In *2024 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS)*, pages 1–3. IEEE.

Howarth, R. W. (2008). Coastal nitrogen pollution: a review of sources and trends globally and regionally. *Harmful Algae*, 8(1):14–20.

Hubanks, P. A., King, M. D., Platnick, S., and Pincus, R. (2008). MODIS atmosphere L3 gridded product algorithm theoretical basis document. *ATBD Reference Number: ATBD-MOD-30*, 30:96.

Hudson, J. and Knopoff, L. (1967). Statistical properties of Rayleigh waves due to scattering by topography. *Bulletin of the Seismological Society of America*, 57(1):83–90.

Hurtado-Uria, C., Hennessy, D., Shalloo, L., Schulte, R., Delaby, L., and O'CONNOR, D. (2013). Evaluation of three grass growth models to predict grass growth in Ireland. *The Journal of Agricultural Science*, 151(1):91–104.

Hutchinson, K., Scobie, D., Beautrais, J., Mackay, A., Rennie, G., Moss, R., and Dynes, R. (2016). A protocol for sampling pastures in hill country. *Journal of New Zealand Grasslands*, 78:203–209.

Iqbal, M. (2012). *An Introduction to Solar Radiation*. Elsevier.

Ishaq, R. A. F., Zhou, G., Tian, C., Tan, Y., Jing, G., and Jiang, H. (2023). A Systematic Review of Radiative Transfer Models for Crop Yield Prediction and Crop Traits Retrieval. *Remote Sensing*, 16(1):121.

Jacquemoud, S. and Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, 34(2):75–91.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., François, C., and Ustin, S. L. (2009a). PROSPECT+ SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment*, 113:S56–S66.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., François, C., and Ustin, S. L. (2009b). PROSPECT+ SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment*, 113:S56–S66.

Jarvis, R. M. and Goodacre, R. (2005). Genetic algorithm optimization for pre-processing and variable selection of spectroscopic data. *Bioinformatics*, 21(7):860–868.

Jeziorska, J. (2019). UAS for wetland mapping and hydrological modeling. *Remote Sensing*, 11(17):1997.

Jin, X., Li, Z., Feng, H., Xu, X., and Yang, G. (2014). Newly Combined Spectral Indices to Improve Estimation of Total Leaf Chlorophyll Content in Cotton. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(11):4589–4600.

Jin, Y.-Q. (1993). *Electromagnetic scattering modelling for quantitative remote sensing*. World Scientific.

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A., Middleton, E., Huemmrich, K., Yoshida, Y., and Frankenberg, C. (2013). Global monitoring of terrestrial chlorophyll fluorescence from moderate spectral resolution near-infrared satellite measurements: Methodology, simulations, and application to GOME-2. *Atmospheric Measurement Techniques Discussions*, 6(2):3883–3930.

Ju, X.-T., Xing, G.-X., Chen, X.-P., Zhang, S.-L., Zhang, L.-J., Liu, X.-J., Cui, Z.-L., Yin, B., Christie, P., Zhu, Z.-L., et al. (2009). Reducing environmental risk by improving N management in intensive Chinese agricultural systems. *Proceedings of the National Academy of Sciences*, 106(9):3041–3046.

Julian, J. P., De Beurs, K. M., Owsley, B., Davies-Colley, R. J., and Ausseil, A.-G. E. (2017). River water quality changes in New Zealand over 26 years: response to land use intensity. *Hydrology and Earth System Sciences*, 21(2):1149–1171.

Junhwan, C., Seokmin, O., and Joongmoo, B. (2022). Uncertainty estimation in AVO inversion using Bayesian dropout based deep learning. *Journal of Petroleum Science and Engineering*, 208:109288.

Justice, C. O., Wharton, S. W., and Holben, B. (1981). Application of digital terrain data to quantify and reduce the topographic effect on Landsat data. *International Journal of Remote Sensing*, 2(3):213–230.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al. (2018). The NCEP/NCAR 40-year reanalysis project. In *Renewable energy*, pages 146–194. Routledge.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440.

Kaufman, Y. J. (1993). Aerosol optical thickness and atmospheric path radiance. *Journal of Geophysical Research: Atmospheres*, 98(D2):2677–2692.

Kawamura, K., Betteridge, K., Sanches, I. D., Tuohy, M. P., Costall, D., and Inoue, Y. (2009). Field radiometer with canopy pasture probe as a potential tool to estimate and map pasture biomass and mineral components: A case study in the Lake Taupo catchment, New Zealand. *New Zealand Journal of Agricultural Research*, 52(4):417–434.

Kecman, V. (2005). Support vector machines–an introduction. In *Support vector machines: theory and applications*, pages 1–47. Springer.

Khanal, S., Kc, K., Fulton, J. P., Shearer, S., and Ozkan, E. (2020). Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote Sensing*, 12(22):3783.

Khosravi, A., Nahavandi, S., Creighton, D., and Atiya, A. F. (2011). Lower Upper Bound Estimation Method for Construction of Neural Network-Based Prediction Intervals. *IEEE Transactions on Neural Networks*, 22(3):337–346.

Kimes, D. S. and Kirchner, J. A. (1981). Modeling the Effects of Various Radiant Transfers in Mountainous Terrain on Sensor Response. *IEEE Transactions on Geoscience and Remote Sensing*, GE-19(2):100–108.

Kirk, P. L. (1950). Kjeldahl method for total nitrogen. *Analytical Chemistry*, 22(2):354–358.

Knoblauch, C., Watson, C., Berendonk, C., Becker, R., Wrage-Mönnig, N., and Wichern, F. (2017). Relationship between Remote Sensing Data, Plant Biomass and Soil Nitrogen Dynamics in Intensively Managed Grasslands under Controlled Conditions. *Sensors*, 17(7):1483.

Knyazikhin, Y., Schull, M. A., Stenberg, P., Mõttus, M., Rautiainen, M., Yang, Y., Marshak, A., Carmona, P. L., Kaufmann, R. K., Lewis, P., et al. (2013). Hyperspectral remote sensing of foliar nitrogen content. *Proceedings of the National Academy of Sciences*, 110(3).

Kokaly, R. F. (2001). Investigating a Physical Basis for Spectroscopic Estimates of Leaf Nitrogen Concentration. *Remote Sensing of Environment*, 75(2):153–161.

Kokaly, R. F., Asner, G. P., Ollinger, S. V., Martin, M. E., and Wessman, C. A. (2009). Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote sensing of environment*, 113:S78–S91.

Kondratyev, K. Y., Kozoderov, V. V., and Smokty, O. I. (2013). *Remote Sensing of the Earth from Space: Atmospheric Correction*. Springer Science & Business Media.

Kruse, F. A. (1988). Use of airborne imaging spectrometer data to map minerals associated with hydrothermally altered rocks in the northern grapevine mountains, Nevada, and California. *Remote Sensing of Environment*, 24(1):31–51.

Kumar, L., Schmidt, K. S., Dury, S., and Skidmore, A. K. (2001). Imaging Spectrometry and Vegetation Science.

Kuo, F. Y. and Sloan, I. H. (2005). Lifting the curse of dimensionality. *Notices of the AMS*, 52(11):1320–1328.

Labate, D., Ceccherini, M., Cisbani, A., De Cosmo, V., Galeazzi, C., Giunti, L., Melozzi, M., Pieraccini, S., and Stagi, M. (2009). The PRISMA payload optomechanical design, a high performance instrument for a new hyperspectral mission. *Acta Astronautica*, 65(9-10):1429–1436.

Landgrebe, D. A. (2003). *Signal Theory Methods in Multispectral Remote Sensing*. Wiley.

Lary, D. J., Alavi, A. H., Gandomi, A. H., and Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3–10.

Lausch, A., Erasmi, S., King, D. J., Magdon, P., and Heurich, M. (2016). Understanding forest health with remote sensing-part I—a review of spectral traits, processes and remote-sensing characteristics. *Remote Sensing*, 8(12):1029.

Lavergne, T., Kaminski, T., Pinty, B., Taberner, M., Gobron, N., Verstraete, M. M., Vossbeck, M., Widlowski, J.-L., and Giering, R. (2007). Application to MISR land products of an RPV model inversion package using adjoint and Hessian codes. *Remote Sensing of Environment*, 107(1-2):362–375.

Leardi, R., Boggia, R., and Terrile, M. (1992). Genetic algorithms as a strategy for feature selection. *Journal of Chemometrics*, 6(5):267–281.

Ledgard, S., Schils, R., Eriksen, J., and Luo, J. (2009). Environmental impacts of grazed clover/grass pastures. *Irish journal of Agricultural and food research*, pages 209–226.

Lee, K.-S., Seo, M., Choi, S., Jin, D., Jung, D., Sim, S., and Han, K.-S. (2021). Uncertainty analysis of BRDF Modeling Using 6S Simulations and Monte-Carlo Method. *Korean Journal of Remote Sensing*, 37(1):161–167.

Li, F., Jupp, D. L., Reddy, S., Lymburner, L., Mueller, N., Tan, P., and Islam, A. (2010). An Evaluation of the Use of Atmospheric and BRDF Correction to Standardize Landsat Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(3):257–270.

Li, F., Jupp, D. L., Thankappan, M., Lymburner, L., Mueller, N., Lewis, A., and Held, A. (2012). A physics-based atmospheric and BRDF correction for Landsat data over mountainous terrain. *Remote Sensing of Environment*, 124:756–770.

Li, H., Liu, G., Liu, Q., Chen, Z., and Huang, C. (2018a). Retrieval of Winter Wheat Leaf Area Index from Chinese GF-1 Satellite Data Using the PROSAIL Model. *Sensors*, 18(4):1120.

Li, X., Du, H., Mao, F., Zhou, G., Han, N., Xu, X., Liu, Y., Zheng, J., Dong, L., Zhang, M., et al. (2019). Assimilating spatiotemporal MODIS LAI data with a particle filter algorithm for improving carbon cycle simulations for bamboo forest ecosystems. *Science of The Total Environment*, 694:133803.

Li, Z., Jin, X., Yang, G., Drummond, J., Yang, H., Clark, B., Li, Z., and Zhao, C. (2018b). Remote Sensing of Leaf and Canopy Nitrogen Status in Winter Wheat (Triticum aestivum L.) Based on N-PROSAIL Model. *Remote Sensing*, 10(9):1463.

Liaghat, S., Balasundram, S. K., et al. (2010). A review: The role of remote sensing in precision agriculture. *American journal of agricultural and biological sciences*, 5(1):50–55.

Liang, L., Di, L., Huang, T., Wang, J., Lin, L., Wang, L., and Yang, M. (2018). Estimation of leaf nitrogen content in wheat using new hyperspectral indices and a random forest regression algorithm. *Remote Sensing*, 10(12):1940.

Liang, S. (2003). *Quantitative remote sensing of land surfaces*. Wiley.

Liu, B. Y. and Jordan, R. C. (1960). The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Solar Energy*, 4(3):1–19.

Liu, L., Liu, X., and Hu, J. (2015). Effects of spectral resolution and SNR on the vegetation solar-induced fluorescence retrieval using FLD-based methods at canopy level. *European Journal of Remote Sensing*, 48(1):743–762.

Liu, S., Yan, Z., Wang, Z., Serbin, S., Visser, M., Zeng, Y., Ryu, Y., Su, Y., Guo, Z., Song, G., Wu, Q., Zhang, H., Cheng, K., Dong, J., Hau, B. C. H., Zhao, P., Yang, X., Liu, L., Rogers, A., and Wu, J. (2023). Mapping foliar photosynthetic capacity in sub-tropical and tropical forests with UAS-based imaging spectroscopy: Scaling from leaf to canopy. *Remote Sensing of Environment*, 293:113612.

Lonati, M., Moot, D. J., Aceto, P., Cavallero, A., and Lucas, R. J. (2009). Thermal time requirements for germination, emergence and seedling development of adventive legume and grass species. *New Zealand Journal of Agricultural Research*, 52(1):17–29.

Lu, B. and He, Y. (2019a). Evaluating empirical regression, machine learning, and radiative transfer modelling for estimating vegetation chlorophyll content using bi-seasonal hyperspectral images. *Remote Sensing*, 11(17):1979.

Lu, B. and He, Y. (2019b). Leaf Area Index Estimation in a Heterogeneous Grassland Using Optical, SAR, and DEM Data. *Canadian Journal of Remote Sensing*, 45(5):618–633.

Lucht, W., Schaaf, C. B., and Strahler, A. H. (2000). An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Transactions on Geoscience and Remote Sensing*, 38(2):977–998.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):56–67.

Lundberg, S. M. and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in neural information processing systems*, 30.

Lynch, J. M. and Barbano, D. M. (1999). Kjeldahl Nitrogen Analysis as a Reference Method for Protein Determination in Dairy Products. *Journal of AOAC INTERNATIONAL*, 82(6):1389–1398.

Maier, S., Lüdeker, W., and Günther, K. (1999). SLOP: A revised version of the stochastic model for leaf optical properties. *Remote Sensing of Environment*, 68(3):273–280.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., and Gascon, F. (2017). Sen2Cor for Sentinel-2. In *Image and Signal Processing for Remote Sensing XXIII*, volume 10427, pages 37–48. SPIE.

Mao, Z.-H., Deng, L., Duan, F.-Z., Li, X.-J., and Qiao, D.-Y. (2020). Angle effects of vegetation indices and the influence on prediction of SPAD values in soybean and maize. *International Journal of Applied Earth Observation and Geoinformation*, 93:102198.

McCartney, E. J. (1977). Optics of the Atmosphere: Scattering by Molecules and Particles. *Physics Today*, 30(5):76–77.

McVeagh, P. (2019). *Prediction of soil Olsen P through mixed pasture leaf tissue biochemical and biophysical properties, topography, and farm management in New Zealand hill country.* PhD thesis, Massey University.

Meroni, M., Rossini, M., Guanter, L., Alonso, L., Rascher, U., Colombo, R., and Moreno, J. (2009). Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications. *Remote sensing of environment*, 113(10):2037–2051.

Meyer, P., Itten, K. I., Kellenberger, T., Sandmeier, S., and Sandmeier, R. (1993). Radiometric corrections of topographically induced effects on Landsat TM data in an alpine environment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 48(4):17–28.

Milton, E. (1987). Review Article Principles of Field Spectroscopy. *Remote Sensing*, 8(12):1807–1827.

Milton, E. J., Schaepman, M. E., Anderson, K., Kneubühler, M., and Fox, N. (2009). Progress in field spectroscopy. *Remote Sensing of Environment*, 113:S92–S109.

Minnaert, M. (1941). The reciprocity principle in lunar photometry. *The Astrophysical Journal*, 93:403.

Mohammed, G. H., Colombo, R., Middleton, E. M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., et al. (2019). Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. *Remote Sensing of Environment*, 231:111177.

Moral, P. D., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo Samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(3):411–436.

Mousivand, A., Menenti, M., Gorte, B., and Verhoef, W. (2014). Global sensitivity analysis of the spectral radiance of a soil–vegetation system. *Remote Sensing of Environment*, 145:131–144.

Mousivand, A., Verhoef, W., Menenti, M., and Gorte, B. (2015). Modeling Top of Atmosphere Radiance over Heterogeneous Non-Lambertian Rugged Terrain. *Remote Sensing*, 7(6):8019–8044.

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4):358–371.

Mutanga, O. (2004a). *Hyperspectral Remote Sensing of Tropical Grass Quality and Quantity*. Wageningen University and Research.

Mutanga, O. (2004b). *Hyperspectral remote sensing of tropical grass quality and quantity*. Phd thesis - research external, graduation external, Netherlands. ITC Dissertation; 111.

Mutanga, O. and Skidmore, A. (2004). Integrating imaging spectroscopy and neural networks to map grass quality in the Kruger National Park, South Africa. *Remote Sensing of Environment*, 90(1):104–115.

Nagelmüller, S., Kirchgessner, N., Yates, S., Hiltpold, M., and Walter, A. (2016). Leaf Length Tracker: a novel approach to analyse leaf elongation close to the thermal limit of growth in the field. *Journal of Experimental Botany*, 67(6):1897–1906.

Nakajima, T. and King, M. D. (1990). Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. Part I: Theory. *Journal of the Atmospheric Sciences*, 47(15):1878–1893.

Navalgund, R. R., Jayaraman, V., and Roy, P. S. (2007). Remote sensing applications: An overview. *Current Science*, 93(12):1747–1766.

Nguyen, M. D., Baez-Villanueva, O. M., Bui, D. D., Nguyen, P. T., and Ribbe, L. (2020). Harmonization of landsat and sentinel 2 for crop monitoring in drought prone areas: Case studies of Ninh Thuan (Vietnam) and Bekaa (Lebanon). *Remote Sensing*, 12(2):281.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Dor, E. B., Brown, D. J., Clairotte, M., Csorba, A., et al. (2015). Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. *Advances in Agronomy*, 132:139–159.

Novoa, R. and Loomis, R. (1981). Nitrogen and plant production. *Soil Water and Nitrogen in Mediterranean-type Environments*, pages 177–204.

Nussenzveig, H. and Wiscombe, W. (1980). Efficiency factors in Mie scattering. *Physical Review Letters*, 45(18):1490–1494.

Obermeier, W. A., Lehnert, L. W., Pohl, M., Gianonni, S. M., Silva, B., Seibert, R., Laser, H., Moser, G., Müller, C., Luterbacher, J., et al. (2019). Grassland ecosystem services in a changing environment: The potential of hyperspectral monitoring. *Remote Sensing of Environment*, 232:111273.

O'Mara, F. P. (2012). The role of grasslands in food security and climate change. *Annals of botany*, 110(6):1263–1270.

Ono, A., Kajiwara, K., Honda, Y., and Ono, A. (2010). Development of vegetation index using radiant spectra normalized by their arithmetic mean. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, Volume XXXVIII, Part 8*, pages 710–714, Kyoto, Japan.

Pellissier, P. A., Ollinger, S. V., Lepine, L. C., Palace, M. W., and McDowell, W. H. (2015). Remote sensing of foliar nitrogen in cultivated grasslands of human dominated landscapes. *Remote Sensing of Environment*, 167:88–97.

Perich, G., Aasen, H., Verrelst, J., Argento, F., Walter, A., and Liebisch, F. (2021). Crop Nitrogen Retrieval Methods for Simulated Sentinel-2 Data Using In-Field Spectrometer Data. *Remote Sensing*, 13(12):2404.

Petermann, J. S. and Buzhdygan, O. Y. (2021). Grassland biodiversity. *Current Biology*, 31(19):R1195–R1201.

Pisharoty, P. (1983). Introduction to remote sensing. *Proceedings of the Indian Academy of Sciences Section C: Engineering Sciences*, 6:97–107.

Platnick, S. (2000). Vertical photon transport in cloud remote sensing problems. *Journal of Geophysical Research: Atmospheres*, 105(D18):22919–22935.

Polimis, K., Rokem, A., and Hazelton, B. (2017). Confidence Intervals for Random Forests in Python. *The Journal of Open Source Software*, 2(19):124.

Povey, A. and Grainger, R. (2015a). Known and unknown unknowns: uncertainty estimation in satellite remote sensing. *Atmospheric Measurement Techniques*, 8(11):4699–4718.

Povey, A. and Grainger, R. (2015b). Known and unknown unknowns: uncertainty estimation in satellite remote sensing. *Atmospheric Measurement Techniques*, 8(11):4699–4718.

Prakash, A. (2000). Thermal remote sensing: concepts, issues and applications. *International Archives of Photogrammetry and Remote Sensing*, 33(B1; PART 1):239–243.

Press, W. H. and Teukolsky, S. A. (1990). Savitzky-Golay smoothing filters. *Computers in Physics*, 4(6):669–672.

Proctor, C., Lu, B., and He, Y. (2017). Determining the absorption coefficients of decay pigments in decomposing monocots. *Remote Sensing of Environment*, 199:137–153.

Proy, C., Tanre, D., and Deschamps, P. (1989). Evaluation of topographic effects in remotely sensed data. *Remote Sensing of Environment*, 30(1):21–32.

Psomas, A., Kneubühler, M., Huber, S., Itten, K., and Zimmermann, N. (2011). Hyperspectral remote sensing for estimating aboveground biomass and for exploring species richness patterns of grassland habitats. *International Journal of Remote Sensing*, 32(24):9007–9031.

Pu, J., Yan, K., Zhou, G., Lei, Y., Zhu, Y., Guo, D., Li, H., Xu, L., Knyazikhin, Y., and Myneni, R. B. (2020). Evaluation of the MODIS LAI/FPAR algorithm based on 3D-RTM simulations: A case study of grassland. *Remote Sensing*, 12(20):3391.

Pullanagari, R., Dehghan-Shoar, M., Yule, I. J., and Bhatia, N. (2021). Field spectroscopy of canopy nitrogen concentration in temperate grasslands using a convolutional neural network. *Remote Sensing of Environment*, 257:112353.

Pullanagari, R., Kereszturi, G., and Yule, I. (2016). Mapping of macro and micro nutrients of mixed pastures using airborne AisaFENIX hyperspectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117:1–10.

Pullanagari, R., Yule, I., King, W., Dalley, D., and Dynes, R. (2011). The Use Of Optical Sensors To Estimate Pasture Quality. *International Journal on Smart Sensing and Intelligent Systems*, 4(1):125–137.

Pullanagari, R., Yule, I., Tuohy, M., Hedley, M., Dynes, R., and King, W. (2012). In-field hyperspectral proximal sensing for estimating quality parameters of mixed pasture. *Precision Agriculture*, 13(3):351–369.

Pullanagari, R. R., Kereszturi, G., and Yule, I. (2018). Integrating airborne hyperspectral, topographic, and soil data for estimating pasture quality using recursive feature elimination with random forest regression. *Remote Sensing*, 10(7):1117.

Qiu, X., Li, Z., Liu, S., Liu, T., and Jia, G. (2023). Uncertainty modeling and analysis of spaceborne infrared hyperspectral images over rugged land surface. *The International Archives*

of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-1/W2-2023:1523–1529.

Qu, Z., Kindel, B. C., and Goetz, A. F. (2003). The High Accuracy Atmospheric Correction for Hyperspectral Data (HATCH) model. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1223–1231.

Quan, X., He, B., and Li, X. (2015). A Bayesian network-based method to alleviate the ill-posed inverse problem: A case study on leaf area index and canopy water content retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, 53(12):6507–6517.

Quan, X., He, B., Yebra, M., Yin, C., Liao, Z., Zhang, X., and Li, X. (2017). A radiative transfer model-based method for the estimation of grassland aboveground biomass. *International Journal of Applied Earth Observation and Geoinformation*, 54:159–168.

Queally, N., Ye, Z., Zheng, T., Chlus, A., Schneider, F., Pavlick, R. P., and Townsend, P. A. (2022). FlexBRDF: A Flexible BRDF Correction for Grouped Processing of Airborne Imaging Spectroscopy Flightlines. *Journal of Geophysical Research: Biogeosciences*, 127(1):e2021JG006622.

Rahman, H. and Dedieu, G. (1994). SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum. *International Journal of Remote Sensing*, 15(1):123–143.

Raines, C. A. (2003). The Calvin cycle revisited. *Photosynthesis research*, 75:1–10.

Raissi, M., Perdikaris, P., and Karniadakis, G. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2017). Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations.

Ramankutty, N., Evan, A. T., Monfreda, C., and Foley, J. A. (2008). Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, 22(1).

Ranganathan, A. (2004). The Levenberg-Marquardt algorithm. *Tutorial on LM Algorithm*, 11(1):101–110.

Reid, M., O'Donovan, M., Elliott, C., Bailey, J., Watson, C., Lalor, S., Corrigan, B., Fenelon, M., and Lewis, E. (2015). The effect of dietary crude protein and phosphorus on grass-fed dairy cow production, nutrient status, and milk heat stability. *Journal of dairy science*, 98(1):517–531.

Rendel, J., Mackay, A., Smale, P., Manderson, A., and Scobie, D. (2020). Optimisation of the resource of land-based livestock systems to advance sustainable agriculture: A farm-level analysis. *Agriculture*, 10(8):331.

Riaño, D., Chuvieco, E., Salas, J., and Aguado, I. (2003). Assessment of different topographic corrections in Landsat-TM data for mapping vegetation types (2003). *IEEE Transactions on Geoscience and Remote Sensing*, 41(5):1056–1061.

Richter, R. (1997). Correction of atmospheric and topographic effects for high spatial resolution satellite imagery. *Algorithms for Multispectral and Hyperspectral Imagery III*.

Richter, R., Kellenberger, T., and Kaufmann, H. (2009). Comparison of Topographic Correction Methods. *Remote Sensing*, 1(3):184–196.

Richter, R. and Schläpfer, D. (2002). Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: Atmospheric/topographic correction. *International Journal of Remote Sensing*, 23(13):2631–2649.

Richter, R. and Schläpfer, D. (2011). Atmospheric/topographic correction for airborne imagery. *ATCOR-4 user guide*, pages 565–02.

Richter, R. and Schläpfer, D. (2019). Atmospheric and topographic correction (ATCOR theoretical background document). *DLR IB*, 1:0564–03.

Richter, R. and Schläpfer, D. (2019). Atmospheric/topographic correction for satellite imagery (ATCOR-2/3 Tutorial, Version 1.4, April 2019). *DLR Report DLR-IB 564-04/2019*.

Richter, T. and Fukshansky, L. (1996). Optics of a Bifacial Leaf: 1. A Novel Combined Procedure for Deriving the Optical Parameters. *Photochemistry and Photobiology*, 63(4):507–516.

Rivera, J. P., Verrelst, J., Leonenko, G., and Moreno, J. (2013). Multiple cost functions and regularization options for improved retrieval of leaf chlorophyll content and LAI through inversion of the PROSAIL model. *Remote Sensing*, 5(7):3280–3304.

Rodriguez-Gomez, C., Kereszturi, G., Jeyakumar, P., Pullanagari, R., Reeves, R., Rae, A., and Procter, J. N. (2023). Remote exploration and monitoring of geothermal sources: A novel method for foliar element mapping using hyperspectral (VNIR-SWIR) remote sensing. *Geothermics*, 111:102716.

Rouse, J. D., Bishop, C. A., and Struger, J. (1999). Nitrogen pollution: an assessment of its threat to amphibian survival. *Environmental Health Perspectives*, 107(10):799–803.

Roy, D. P., Kovalskyy, V., Zhang, H., Vermote, E. F., Yan, L., Kumar, S., and Egorov, A. (2016a). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment*, 185:57–70.

Roy, D. P., Li, Z., and Zhang, H. K. (2017). Adjustment of Sentinel-2 multi-spectral instrument (MSI) Red-Edge band reflectance to Nadir BRDF adjusted reflectance (NBAR) and quantification of red-edge band BRDF effects. *Remote Sensing*, 9(12):1325.

Roy, D. P., Wulder, M. A., Loveland, T. R., Woodcock, C. E., Allen, R. G., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D. M., Kennedy, R., et al. (2014). Landsat-8: Science

and product vision for terrestrial global change research. *Remote sensing of Environment*, 145:154–172.

Roy, D. P., Zhang, H., Ju, J., Gomez-Dans, J. L., Lewis, P. E., Schaaf, C., Sun, Q., Li, J., Huang, H., and Kovalskyy, V. (2016b). A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, 176:255–271.

Sakowska, K., Juszczak, R., and Gianelle, D. (2016). Remote Sensing of Grassland Biophysical Parameters in the Context of the Sentinel-2 Satellite Mission. *Journal of Sensors*, 2016:1–16.

Salisbury, J. W. and D'Aria, D. M. (1992). Emissivity of terrestrial materials in the 8–14 μm atmospheric window. *Remote Sensing of Environment*, 42(2):83–106.

Sanches, I., Tuohy, M., Hedley, M., and Bretherton, M. (2009). Large, durable and low-cost reflectance standard for field remote sensing applications. *International Journal of Remote Sensing*, 30(9):2309–2319.

Sankey, J. B., Brown, D. J., Bernard, M. L., and Lawrence, R. L. (2008). Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. *Geoderma*, 148(2):149–158.

Santoro, M., Beer, C., Cartus, O., Schmullius, C., Shvidenko, A., McCallum, I., Wegmüller, U., and Wiesmann, A. (2011). Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements. *Remote Sensing of Environment*, 115(2):490–507.

Sayer, A. M., Govaerts, Y., Kolmonen, P., Lipponen, A., Luffarelli, M., Mielonen, T., Patadia, F., Popp, T., Povey, A. C., Stebel, K., et al. (2020). A review and framework for the evaluation of pixel-level uncertainty estimates in satellite aerosol remote sensing. *Atmospheric Measurement Techniques*, 13(2):373–404.

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., Muller, J.-P., et al. (2002). First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment*, 83(1-2):135–148.

Schafer, R. W. (2011). What is a Savitzky-Golay filter?[lecture notes]. *IEEE Signal Processing Magazine*, 28(4):111–117.

Schell, U. (2022). *Keras/Tensorflow*. Carl Hanser Verlag GmbH Co. KG.

Schinasi, L. H., Benmarhnia, T., and De Roos, A. J. (2018). Modification of the association between high ambient temperature and health by urban microclimate indicators: A systematic review and meta-analysis. *Environmental Research*, 161:168–180.

Schläpfer, D., Hausold, A., and Richter, R. (2000). A unified approach to parametric geocoding and atmospheric/topographic correction for wide FOV airborne imagery. Part 1: parametric ortho-rectification process. In *Proc. 2nd EARSeL Workshop on Imaging Spectroscopy, EARSeL, Enschede*, volume 9.

Schläpfer, D. and Richter, R. (2002). Geo-atmospheric processing of airborne imaging spectrometry data. Part 1: Parametric orthorectification. *International Journal of Remote Sensing*, 23(13):2609–2630.

Schläpfer, D., Richter, R., and Feingersh, T. (2014). Operational BRDF effects correction for wide-field-of-view optical scanners (BREFCOR). *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):1855–1864.

Seinfeld, J. H. and Pandis, S. N. (2016). *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons.

Shao, Z., Cai, J., Fu, P., Hu, L., and Liu, T. (2019). Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. *Remote Sensing of Environment*, 235:111425.

Si, Y., Schlerf, M., Zurita-Milla, R., Skidmore, A., and Wang, T. (2012). Mapping spatio-temporal variation of grassland quantity and quality using MERIS data and the PROSAIL model. *Remote Sensing of Environment*, 121:415–425.

Simon Blanke (since 2020). Gradient-Free-Optimizers: Simple and reliable optimization with local, global, population-based and sequential techniques in numerical search spaces.

Sinha, S. K., Padalia, H., Dasgupta, A., Verrelst, J., and Rivera, J. P. (2020). Estimation of leaf area index using PROSAIL based LUT inversion, MLRA-GPR and empirical models: Case study of tropical deciduous forest plantation, North India. *International Journal of Applied Earth Observation and Geoinformation*, 86:102027.

Sishodia, R. P., Ray, R. L., and Singh, S. K. (2020). Applications of remote sensing in precision agriculture: A review. *Remote sensing*, 12(19):3136.

Slater, P., Biggar, S., Holm, R., Jackson, R., Mao, Y., Moran, M., Palmer, J., and Yuan, B. (1987). Reflectance-and radiance-based methods for the in-flight absolute calibration of multispectral sensors. *Remote sensing of environment*, 22(1):11–37.

Slater, P. N. (1980). Remote sensing: Optics and optical systems. *Reading*.

Smhh, D. P. and Atkinson, S. F. (2001). Accuracy of rectification using topographic map versus GPS ground control points. *Photogrammetric Engineering & Remote Sensing*, 67(5):565–570.

Smith, J., Lin, T. L., Ranson, K., et al. (1980). The Lambertian assumption and Landsat data. *Photogrammetric Engineering and Remote Sensing*, 46(9):1183–1189.

Smith, K., Strand, D. D., and Walker, B. J. (2024). Evaluating the contribution of plant metabolic pathways in the light to the atp: Nadph demand using a meta-analysis of isotopically non-stationary metabolic flux analyses. *Photosynthesis Research*, pages 1–13.

Soenen, S. A., Peddle, D. R., and Coburn, C. A. (2005). SCS+ C: A modified sun-canopy-sensor topographic correction in forested terrain. *IEEE Transactions on Geoscience and Remote Sensing*, 43(9):2148–2159.

Steinfeld, H., Gerber, P., Wassenaar, T. D., Castel, V., Rosales, M., Rosales, M., and de Haan, C. (2007). *Livestock's long shadow: environmental issues and options*, volume 32. Annual Reviews.

Sullivan, J. M., Twardowski, M. S., Ronald, J., Zaneveld, V., and Moore, C. C. (2013). Measuring optical backscattering in water. *Light Scattering Reviews 7*, pages 189–224.

Sun, A. Y., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A., and Zhong, Z. (2019). Combining physically based modeling and deep learning for fusing GRACE satellite data: can we learn from mismatch? *Water Resources Research*, 55(2):1179–1195.

Sun, J., Wang, L., Shi, S., Li, Z., Yang, J., Gong, W., Wang, S., and Tagesson, T. (2022). Leaf pigment retrieval using the PROSAIL model: influence of uncertainty in prior canopy-structure information. *The Crop Journal*, 10(5):1251–1263.

Sundberg, R., Richtsmeier, S., and Haren, R. (2010). Monte Carlo based hyperspectral scene simulation. In *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*. IEEE.

Szantoi, Z. and Strobl, P. (2019). Copernicus Sentinel-2 calibration and validation.

Taubert, F., Frank, K., and Huth, A. (2012). A review of grassland models in the biofuel context. *Ecological Modelling*, 245:84–93.

Teillet, P., Guindon, B., and Goodenough, D. (1982). On the Slope-Aspect Correction of Multispectral Scanner Data. *Canadian Journal of Remote Sensing*, 8(2):84–106.

Thenkabail, P. S., Lyon, J. G., and Huete, A. (2018). Advances in hyperspectral remote sensing of vegetation and agricultural crops. In *Fundamentals, Sensor Systems, Spectral Libraries, and Data Mining for Vegetation*, pages 3–37. CRC press.

Thieme, A., Prabhakara, K., Jennewein, J., Lamb, B. T., McCarty, G. W., and Hively, W. D. (2024). Intercomparison of Same-Day Remote Sensing Data for Measuring Winter Cover Crop Biophysical Traits. *Sensors*, 24(7).

Tilman, D., Hill, J., and Lehman, C. (2006a). Carbon-negative biofuels from low-input high-diversity grassland biomass. *Science*, 314(5805):1598–1600.

Tilman, D., Reich, P. B., and Knops, J. M. H. (2006b). Biodiversity and ecosystem stability in a decade-long grassland experiment. *Nature*, 441(7093):629–632.

Tilman, D., Wedin, D., and Knops, J. (1996). Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature*, 379(6567):718–720.

Tiscornia, G., Jaurena, M., and Baethgen, W. (2019). Drivers, process, and consequences of native grassland degradation: Insights from a literature review and a survey in Río de la plata grasslands. *Agronomy*, 9(5):239.

Tomasi, C. (1979). Non-selective absorption by atmospheric water vapour at visible and near infrared wavelengths. *Quarterly Journal of the Royal Meteorological Society*, 105(446):1027–1040.

Tong, A. and He, Y. (2017). Estimating and mapping chlorophyll content for a heterogeneous grassland: Comparing prediction power of a suite of vegetation indices across scales between years. *ISPRS Journal of Photogrammetry and Remote Sensing*, 126:146–167.

Tsai, F. and Philpot, W. (1998). Derivative analysis of hyperspectral data. *Remote Sensing of Environment*, 66(1):41–51.

Ulaby, F. T., Moore, R. K., and Fung, A. K. (1981). *Microwave remote sensing: Active and passive. volume 1-microwave remote sensing fundamentals and radiometry.*

Ullah, S., Si, Y., Schlerf, M., Skidmore, A. K., Shafique, M., and Iqbal, I. A. (2012). Estimation of grassland biomass and nitrogen using MERIS data. *International Journal of Applied Earth Observation and Geoinformation*, 19:196–204.

Useya, J. and Chen, S. (2018). Comparative Performance Evaluation of Pixel-Level and Decision-Level Data Fusion of Landsat 8 OLI, Landsat 7 ETM+ and Sentinel-2 MSI for Crop Ensemble Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(11):4441–4451.

Van Zyl, J. J. (2001). The Shuttle Radar Topography Mission (SRTM): a breakthrough in remote sensing of topography. *Acta Astronautica*, 48(5-12):559–565.

Vanonckelen, S., Lhermitte, S., and Van Rompaey, A. (2013). The effect of atmospheric and topographic correction methods on land cover classification accuracy. *International Journal of Applied Earth Observation and Geoinformation*, 24:9–21.

Vázquez-Jiménez, R., Romero-Calcerrada, R., Ramos-Bernal, R. N., Arrogante-Funes, P., and Novillo, C. J. (2017). Topographic Correction to Landsat Imagery through Slope Classification by Applying the SCS + C Method in Mountainous Forest Areas. *ISPRS International Journal of Geo-Information*, 6(9):287.

Verde, N., Mallinis, G., Tsakiri-Strati, M., Georgiadis, C., and Patias, P. (2018). Assessment of Radiometric Resolution Impact on Remote Sensing Data Classification Accuracy. *Remote Sensing*, 10(8):1267.

Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment*, 16(2):125–141.

Verhoef, W. (1998). *Theory of radiative transfer models applied in optical remote sensing of vegetation canopies.* Wageningen University and Research.

Verhoef, W., Van Der Tol, C., and Middleton, E. M. (2018). Hyperspectral radiative transfer modeling to explore the combined retrieval of biophysical parameters and canopy fluorescence from FLEX–Sentinel-3 tandem mission multi-sensor data. *Remote Sensing of Environment*, 204:942–963.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G., and Moreno, J. (2015a). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties–A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108:273–290.

Verrelst, J., Dethier, S., Rivera, J. P., Munoz-Mari, J., Camps-Valls, G., and Moreno, J. (2016a). Active Learning Methods for Efficient Hybrid Biophysical Variable Retrieval. *IEEE Geoscience and Remote Sensing Letters*, 13(7):1012–1016.

Verrelst, J., Malenovskỳ, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P., and Moreno, J. (2019a). Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. *Surveys in Geophysics*, 40(3):589–629.

Verrelst, J., Rivera, J. P., Gitelson, A., Delegido, J., Moreno, J., and Camps-Valls, G. (2016b). Spectral band selection for vegetation properties retrieval using Gaussian processes regression. *International Journal of Applied Earth Observation and Geoinformation*, 52:554–567.

Verrelst, J., Rivera, J. P., van der Tol, C., Magnani, F., Mohammed, G., and Moreno, J. (2015b). Global sensitivity analysis of the SCOPE model: What drives simulated canopy-leaving sun-induced fluorescence? *Remote Sensing of Environment*, 166:8–21.

Verrelst, J., Vicent, J., Rivera-Caicedo, J. P., Lumbierres, M., Morcillo-Pallarés, P., and Moreno, J. (2019b). Global Sensitivity Analysis of Leaf-Canopy-Atmosphere RTMs: Implications for Biophysical Variables Retrieval from Top-of-Atmosphere Radiance Data. *Remote Sensing*, 11(16):1923.

Virgolin, M., Alderliesten, T., Witteveen, C., and Bosman, P. A. N. (2021). Improving Model-Based Genetic Programming for Symbolic Regression of Small Expressions. *Evolutionary Computation*, 29(2):211–237.

Virgolin, M., De Lorenzo, A., Medvet, E., and Randone, F. (2020). Learning a formula of interpretability to learn interpretable formulas. In *Parallel Problem Solving from Nature–PPSN XVI: 16th International Conference, PPSN 2020, Leiden, The Netherlands, September 5-9, 2020, Proceedings, Part II 16*, pages 79–93. Springer.

Vögtli, M., Schläpfer, D., Schuman, M. C., Schaepman, M. E., Kneubühler, M., and Damm, A. (2024). Effects of atmospheric, topographic, and BRDF correction on imaging spectroscopy-derived data products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:109–126.

Vohland, M. and Mader, S. (2007). Numerical minimisation and artificial neural networks: Two different approaches to retrieve parameters from a canopy reflectance model. In *Proceedings 5th EARSeL Workshop on Imaging Spectroscopy. Bruges*, volume 1.

Vovk, V. (2013). Kernel ridge regression. In *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, pages 105–116. Springer.

Wan, L., Zhou, W., He, Y., Wanger, T. C., and Cen, H. (2022). Combining transfer learning and hyperspectral reflectance analysis to assess leaf nitrogen concentration across different plant species datasets. *Remote Sensing of Environment*, 269:112826.

Wang, J., Lopez-Lozano, R., Weiss, M., Buis, S., Li, W., Liu, S., Baret, F., and Zhang, J. (2022a). Crop specific inversion of PROSAIL to retrieve green area index (GAI) from several decametric satellites using a Bayesian framework. *Remote Sensing of Environment*, 278:113085.

Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R. B., and Chang, Q. (2019a). Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154:189–201.

Wang, Y. (2012). Gauss–newton method. *WIREs Computational Statistics*, 4(4):415–420.

Wang, Y., Suarez, L., Poblete, T., Gonzalez-Dugo, V., Ryu, D., and Zarco-Tejada, P. (2022b). Evaluating the role of solar-induced fluorescence (SIF) and plant physiological traits for leaf nitrogen assessment in almond using airborne hyperspectral imagery. *Remote Sensing of Environment*, 279:113141.

Wang, Z., Chlus, A., Geygan, R., Ye, Z., Zheng, T., Singh, A., Couture, J. J., Cavender-Bares, J., Kruger, E. L., and Townsend, P. A. (2020). Foliar functional traits from imaging spectroscopy across biomes in eastern North America. *New Phytologist*, 228(2):494–511.

Wang, Z., Schaaf, C. B., Strahler, A. H., Chopping, M. J., Román, M. O., Shuai, Y., Woodcock, C. E., Hollinger, D. Y., and Fitzjarrald, D. R. (2014). Evaluation of MODIS albedo product (MCD43A) over grassland, agriculture and forest surface types during dormant and snow-covered periods. *Remote Sensing of Environment*, 140:60–77.

Wang, Z., Townsend, P. A., Schweiger, A. K., Couture, J. J., Singh, A., Hobbie, S. E., and Cavender-Bares, J. (2019b). Mapping foliar functional traits and their uncertainties across three years in a grassland experiment. *Remote Sensing of Environment*, 221:405–416.

Weiss, M., Baret, F., Myneni, R., Pragnère, A., and Knyazikhin, Y. (2000). Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie*, 20(1):3–22.

Wheeler, D., Cichota, R., Snow, V., and Shepherd, M. (2011a). A revised leaching model for OVERSEER® Nutrient Budgets. *Adding to the knowledge base for the nutrient manager.*

Wheeler, D., Cichota, R., Snow, V., and Shepherd, M. (2011b). A revised leaching model for OVERSEER® Nutrient Budgets. *Adding to the knowledge base for the nutrient manager.*

Wheeler, D., Ledgard, S., and DeKlein, C. (2008). Using the OVERSEER nutrient budget model to estimate on-farm greenhouse gas emissions. *Australian Journal of Experimental Agriculture*, 48(2):99–103.

White, T. A., Campbell, B. D., Kemp, P. D., and Hunt, C. L. (2000). Sensitivity of three grassland communities to simulated extreme temperature and rainfall events. *Global Change Biology*, 6(6):671–684.

Wijmer, T., Al Bitar, A., Arnaud, L., Fieuzal, R., and Ceschia, E. (2024). AgriCarbon-EO v1.0.1: large-scale and high-resolution simulation of carbon fluxes by assimilation of Sentinel-2 and Landsat-8 reflectances using a Bayesian approach. *Geoscientific Model Development*, 17(3):997–1021.

Williams, R., Brown, H., Dunbier, M., Edmeades, D., Hill, R., Metherell, A., Rahn, C., and Thorburn, P. (2013). A critical examination of the role of OVERSEER® in modelling nitrate losses from arable crops. *Occasional Report*, 26(8).

Wingler, A. and Hennessy, D. (2016). Limitation of grassland productivity by low temperature and seasonality of growth. *Frontiers in Plant Science*, 7.

Wocher, M., Berger, K., Verrelst, J., and Hank, T. (2022). Retrieval of carbon content and biomass from hyperspectral imagery over cultivated areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 193:104–114.

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130.

Wu, S., Wen, J., Gastellu-Etchegorry, J.-P., Liu, Q., You, D., Xiao, Q., Hao, D., Lin, X., and Yin, T. (2019). The definition of remotely sensed reflectance quantities suitable for rugged terrain. *Remote Sensing of Environment*, 225:403–415.

Wyatt, C. (2012). *Radiometric calibration: theory and methods*. Elsevier.

Xu, D., Guo, X., Li, Z., Yang, X., and Yin, H. (2014). Measuring the dead component of mixed grassland with Landsat imagery. *Remote Sensing of Environment*, 142:33–43.

Xu, L., Shi, S., Gong, W., Shi, Z., Qu, F., Tang, X., Chen, B., and Sun, J. (2022). Improving leaf chlorophyll content estimation through constrained PROSAIL model from airborne hyperspectral and LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, 115:103128.

Yan, J., Zhang, G., Deng, X., Ling, H., Xu, H., and Guo, B. (2019). Does climate change or human activity lead to the degradation in the grassland ecosystem in a mountain-basin system in an arid region of China? *Sustainability*, 11(9):2618.

Yang, P., van der Tol, C., Yin, T., and Verhoef, W. (2020a). The SPART model: A soil-plant-atmosphere radiative transfer model for satellite measurements in the solar spectrum. *Remote Sensing of Environment*, 247:111870.

Yang, P., Verhoef, W., and van der Tol, C. (2017). The mSCOPE model: A simple adaptation to the SCOPE model to describe reflectance, fluorescence and photosynthesis of vertically heterogeneous canopies. *Remote Sensing of Environment*, 201:1–11.

Yang, P., Verhoef, W., and Van der Tol, C. (2020b). Unified Four-Stream Radiative Transfer Theory in the Optical-Thermal Domain with Consideration of Fluorescence for Multi-Layer Vegetation Canopies. *Remote Sensing*, 12(23):3914.

Yin, H., Tan, B., Frantz, D., and Radeloff, V. (2018). Evaluation of topographic correction on forest mapping using Landsat imagery. In *AGU Fall Meeting Abstracts*, volume 2018, pages B31I–2598.

Yin, H., Tan, B., Frantz, D., and Radeloff, V. C. (2022). Integrated topographic corrections improve forest mapping using Landsat imagery. *International Journal of Applied Earth Observation and Geoinformation*, 108:102716.

Yokoya, N., Miyamura, N., and Iwasaki, A. (2010). Preprocessing of hyperspectral imagery with consideration of smile and keystone properties. In *Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques, and Applications III*, volume 7857, pages 73–81. SPIE.

Young, A. T. (1982). Rayleigh scattering. *Physics Today*, 35(1):42–48.

Zarco-Tejada, P. J., Poblete, T., Camino, C., Gonzalez-Dugo, V., Calderon, R., Hornero, A., Hernandez-Clemente, R., Román-Écija, M., Velasco-Amo, M. P., Landa, B. B., Beck, P. S. A., Saponari, M., Boscia, D., and Navas-Cortes, J. A. (2021). Divergent abiotic spectral pathways unravel pathogen stress signals across species. *Nature Communications*, 12(1).

Zhang, C., Yi, Y., Wang, L., Chen, S., Li, P., Zhang, S., and Xue, Y. (2024). Efficient physics-informed transfer learning to quantify biochemical traits of winter wheat from uav multispectral imagery. *Smart Agricultural Technology*, 9:100581.

Zhang, H. K., Roy, D. P., Yan, L., Li, Z., Huang, H., Vermote, E., Skakun, S., and Roger, J.-C. (2018). Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences. *Remote Sensing of Environment*, 215:482–494.

Zhao, C., Wang, Z., Wang, J., and Huang, W. (2012). Relationships of leaf nitrogen concentration and canopy nitrogen density with spectral features parameters and narrow-band spectral indices calculated from field winter wheat (Triticum aestivum L.) spectra. *International Journal of Remote Sensing*, 33(11):3472–3491.

Zhao, K., Valle, D., Popescu, S., Zhang, X., and Mallick, B. (2013). Hyperspectral remote sensing of plant biochemistry using Bayesian model averaging with variable and band selection. *Remote Sensing of Environment*, 132:102–119.

Zhao, Y., He, X., Bai, Y., Xu, F., Jin, X., Li, T., Wang, D., and Gong, F. (2024). Adjacency effect on Rayleigh scattering radiance for satellite remote sensing of river waters. *IEEE Transactions on Geoscience and Remote Sensing*.

Zhou, K., Xie, D., Qi, J., Zhang, Z., Bo, X., Yan, G., and Mu, X. (2023). Explicitly reconstructing RAMI-V scenes for accurate 3-dimensional radiative transfer simulation using the LESS model. *Journal of Remote Sensing*, 3:0033.