

**DEVELOPMENT OF SENSING SYSTEM BASED ON FOLIAR SPECTRAL
REFLECTANCE TO ESTIMATE NUTRIENTS IN POTATO PLANTS**

by

Reem Abukmeil

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University

Halifax, Nova Scotia

April 2024

Dalhousie University is located in Mi'kma'ki, the
ancestral and unceded territory of the Mi'kmaq.

We are all Treaty people.

DEDICATION

This PhD thesis dissertation is dedicated to the memory of my father, and my brother, whose presence I deeply miss and who unfortunately could not witness the completion of my doctoral journey. I believe they would have experienced immense pride in my accomplishments if they had been alive today.

Author,

Reem Abukmeil, Eng.

TABLE OF CONTENTS

List of Tables	vi
List of Figures	viii
Abstract	xi
List of Abbreviations and symbols used	xii
Acknowledgments	xiv
1. Chapter 1: Introduction	1
1.1. Background	1
1.2. Application of spectroscopy for nutrients assessment in agriculture	4
1.2.1. Remote sensors - advantages, and challenges	7
1.2.2. Proximal sensors - advantages and challenges	10
1.3. The application of Vis-NIRS in estimating nutrients in potato plants	14
1.4. Spectroscopy data analysis	15
1.5. Research goal and objectives	17
1.6. Research Hypothesis	17
2. Chapter 2: assessment of the correlation between petiole chemical content and leaf reflectance spectra of potato plants	19
2.1. Background	19
2.2. Data collection of petioles and leaves	19
2.2.1. Data sampling of outdoor fields	19
2.2.2. Data sampling of indoor growing area	21
2.3. Analysis of samples (chemical testing and spectral analysis)	24
2.4. Dataset construction using linear relationship	26
2.5. Rationale for model selection	27
2.6. Estimation of macro and micronutrients in 2020	29
2.6.1. Spatial and temporal distribution of potato petioles chemical content	29

2.6.2. Determination of the significant wavebands.....	34
2.6.3. Model performance measures in estimating nutrients	40
2.7. Estimation of macro and micronutrients in 2021	46
2.7.1. Spatial and temporal distribution of petiole chemical content.....	46
2.7.2. Determination of the significant wavebands.....	47
2.8. Estimation of P at multiple seasons: a case study	50
2.9. Conclusion	56
3. Chapter 3: Calibrating lab and field spectral instruments for nutrient when reading reflectance of potato plants	58
3.1. Background	58
3.2. Data collection of lab and field reflectance spectra	58
3.2.1. Samples acquisition of laboratory and field-based spectrophotometers	58
3.2.2. Raw spectra from lab and field spectral instruments	61
3.3. Calibration of lab and field spectra at each waveband as a single entity	63
3.3.1. Raw spectra of in-field datasets in comparison to lab datasets.....	66
3.3.2. Calibrated of in field datasets in comparison to lab datasets	68
3.3.3. Verifying the calibration at each waveband in estimating NPK.....	71
3.4. Conclusion	75
4. Chapter 4: comprehensive methodology to estimate nutrients using combination of machine learning techniques.....	76
4.1. Background	76
4.2. Demonstration of multivariate relationships based on ground truth data	76
4.3. Model selection and concept.....	78
4.4. Improvement in estimating nutrients	80
4.4.1. Model performance measures	80
4.4.2. Model performance measures at feature selection method.....	85

4.4.3. Assessing the performance gains from adding nonlinear models	88
4.5. Two vs. three-season model estimation	91
4.5.1. Positions of the significant wavebands	95
4.5.2. Performance of estimation models using the stacked multivariate regression pipeline.....	95
4.6. Conclusion	98
5. Chapter 5: Rapid estimation of nutrients in potato plants using Leaf reflectance spectra	99
5.1. Background	99
5.2. Optimizing number of leaves for the in-field spectral analysis	100
5.3. Steps of the developed sensing system	104
5.3.1. Acquiring spectra measurements	104
5.3.2. Storing the leaf spectral data	104
5.3.3. Sending the stored spectra to Microsoft Azure cloud	105
5.3.4. Computing and delivering the nutrient values of the petiole based on the stored leaf spectral data	107
5.4. General overview of the developed sensing system components	108
5.5. Conclusion	110
6. Chapter 6: Conclusion and future plans	111
6.1. Importance of incorporating various fertilization schemes.....	112
6.2. Influence outliers at the ground truth level	112
6.3. Influence of bouncing reflectance spectra on wavelength range reduction	115
6.4. Cost comparison between chemical testing versus sensing	116
6.5. Potential for enhancements at proposed sensing system front-end.....	117
References.....	119

LIST OF TABLES

Table 1.1: List of studies used different machine learning algorithms to extract information from the NIR data to estimate nutrients	16
Table 2.1: Operating specifications of NIRS DS2500 Analyzer.	24
Table 2.2: Number, range, and first four significant wavebands resulting from Lasso MLR modeling at the dried testing mode in Season 2020	35
Table 2.3: Number, range, and first four significant wavebands resulting from Lasso MLR modeling at the fresh testing mode in Season 2020.....	35
Table 2.4: The validation results of Lasso MLR models estimating elements at the best testing mode.....	43
Table 2.5: Descriptive content of the petiole nutritional concentrations during the entire growth season in 2021.	47
Table 2.6: Number, range, and first four significant wavebands resulting from Lasso MLR modeling at the dried testing mode in Season 2021.	48
Table 2.7: Number, range, and first four significant wavebands resulting from Lasso MLR modeling at the fresh testing mode in Season 2021.....	48
Table 2.8: The validation results of Lasso MLR models estimating in Season 2021 at the best testing mode.....	49
Table 2.9: Performance of regression models for dried leaves.....	54
Table 2.10: Performance of regression models for fresh leaves.....	55
Table 3.1: Operating specifications of the two spectrophotometers.....	60
Table 3.2: Prediction performance in estimating NPK using the lab vs. field spectra, and lab vs field calibrated spectra of both modes.....	75
Table 4.1: Prediction performance values in estimating macro and micronutrients in single and across seasons of the best testing mode.....	81
Table 4.2: Prediction performance of the stacked multivariate regression at the base and meta layers of the macro-nutrients at the dried mode.....	83
Table 4.3: Prediction performance of the stacked multivariate regression at the base and meta layers of the micro-nutrients at the dried mode.	83
Table 4.4: Prediction performance of the stacked multivariate regression at the base and meta layers of macro-nutrients at the fresh mode.....	84

Table 4.5: Prediction performance of the stacked multivariate regression at the base and meta layers of micro-nutrients at the fresh mode.....	84
Table 4.6: Prediction performance of the stacked multivariate regression at the base and meta layers after the feature selection using MRMR of the macro-nutrient at the dried mode.....	86
Table 4.7: Prediction performance of the stacked multivariate regression at the base and meta layers after the feature selection using MRMR of the micro-nutrient at the dried mode.....	87
Table 4.8: Prediction performance of the stacked multivariate regression at the base and meta layers after the MRMR method of the macro-nutrients at the fresh mode.	87
Table 4.9: Prediction performance of the stacked multivariate regression at the base and meta layers after the MRMR method of the micro-nutrients at the fresh mode.	88
Table 4.10: Prediction performance of using nonlinear models at the meta layer of the cross-validation dataset at dried mode.....	89
Table 4.11: Prediction performance of using nonlinear models at the meta layer of the cross-validation dataset at fresh mode.	90
Table 4.12: Prediction performance of using nonlinear models at the meta layer of the testing dataset at dried mode.....	90
Table 4.13: Prediction performance of using nonlinear models at the meta layer of the testing dataset at fresh mode.	91
Table 4.14: Descriptive content of petiole nutritional concentrations during the entire growth season in 2022.	92
Table 4.15: Positions of the first four significant wavebands resulting from multivariate stacked regression of three seasons (2020, 2021, 2022).....	95
Table 4.16: Prediction performance of the stacked multivariate regression at the meta layer of the testing dataset at the dried mode of two vs three-season model.....	97
Table 4.17: Prediction performance of the stacked multivariate regression at the meta layer of the testing dataset at the fresh mode of two vs three-season model	97
Table 5.1: Results of the statistical analyses of one data point.....	103
Table 6.1: Preliminary expected costs in a year to operate the proposed sensing system by a consultant vs the current sampling cost based on 360 samples in season.....	117

LIST OF FIGURES

Figure 1.1: Current approach of petioles samples collection for lab chemical testing.	4
Figure 1.2: The spectrum of sunlight at sea level on a bright day. The red circles refer to the water absorption wavebands by atmospheric water vapor.	6
Figure 1.3: Comparison between: (a) non-imaging, and (b) imaging spectroscopy.....	6
Figure 2.1: Location of the open-field farms where samples were collected from NB. Green circles refer to the areas where samples collected in 2020, 2021, and 2022. Blue circle shows an additional area for data collection in 2022.	20
Figure 2.2: (a) Steps of sample collection. (b) Chemical testing of petioles. (c) Spectral analysis of dried and fresh leaves performed over the two modes of dried and fresh leaves.	21
Figure 2.3: Layout and internal view of the indoor growing area during potato planting (a) location of sensors inside, and the location of LED lights using vertical lines, (b) early growing stage, (c) vegetation growing stage.	23
Figure 2.4: Difference between reflectance spectra of dried and fresh leaves.	25
Figure 2.5: Steps taken to handle the raw data as inputs into the dataset (a) raw format of the chemical content results, (b) raw format of receiving the spectra data, and (c) development of a dataset based on a linear relationship.	26
Figure 2.6: Temporal distribution for each nutrient during the growing season. Horizontal lines represent the limits of the maximum (red) and minimum (yellow) range of nutrients in potato petioles as recommended by A&L Canada Laboratories Inc in Ontario. Blue dots refer to data points of Season 2020, and black dots refer to the data points taken from indoor growing area.....	34
Figure 2.7: Pearson's correlation (r) for dried (brown) and fresh leaves (green) across the spectrum. Absolute peaks represent the highest r . The solid line outlines the testing mode of the highest r^2 using Lasso Regression, while dashed line represents the testing mode with less r^2 . The bars present the region of the four most significant wavebands by Lasso Regression. The intensity of the grey scale of the bar intensity gives the sequence of important wavebands from darker to brighter.	40
Figure 2.8: The validation results of the actual versus estimated concentrations of the testing mode (green for fresh, brown for dried) of the highest r^2 . Blue box shows the normal range of nutrients in potato petioles as recommended by A&L Canada Laboratories.	45
Figure 2.9: A descriptive diagram showing the positioning of the fundamental absorption waveband in the mid infrared region of the spectrum (mid-IR), and the positions of the bounced overtones within the NIR range.....	50

Figure 2.10: Position of the wavebands selected by PLSR and Lasso to estimate P across seasons: (a) dried mode, (b) fresh mode. The shaded bars refer to the water absorbance wavelengths..... 56

Figure 3.1: Lab, and field (raw), intensities for a representative data point of the test data. in the (a) dried and (b) fresh modes. The blue regions highlight water absorbance wavelengths..... 62

Figure 3.2: General structure of the proposed calibration modeling structure. (A) Matched spectra obtained from lab and field measurements. (B) Extracting a target variable (intensity at a specific waveband) and associated predictors (intensities at the local neighborhood of the target waveband). (C) Building of predictive models using the target and predictor variables values across all training data points. (D) Resulting models. 65

Figure 3.3: Flow of the calibration method towards developing calibrated field spectra involving the neighbor-based selection of variables followed by SVR..... 66

Figure 3.4: Lab, field (raw), and field (calibrated) intensities for a representative data point of the test data. (a) dried and (b) fresh modes. The blue regions highlight water absorbance wavelengths..... 67

Figure 3.5: Absolute error at each waveband between lab and field calibrated spectra, vs. lab and field spectra at the dried mode of the (a) cross-validation, and (b) holdout split. The thick middle line represents the mean absolute error..... 70

Figure 3.6: Absolute error at each waveband between lab and field calibrated spectra, vs. lab and field spectra at the fresh mode of the (a) cross-validation, and (b) holdout split. The thick middle line represents the mean absolute error..... 71

Figure 3.7: Estimated error of (a) N, (b), P, and (c) K estimation values of the five cross-validation folds and the holdout split of the dried mode by field and field calibrated spectra. 73

Figure 3.8: Estimated error of (a) N, (b), P, and (c) K estimation values of the five cross-validation folds and the holdout split of the fresh mode by field and field calibrated spectra. 74

Figure 4.1: Spearman correlation values among nutrient concentrations. The numbers are the sample estimates based on the data collected in this research, while the colors indicate the expected direction of each association based on literature. 78

Figure 4.2: Concept of stacked multivariate regression at base and meta layers. Y represents the predicted target feature at the base layer and \hat{Y} presents the predicted target feature at the meta layer. 79

Figure 4.3: Stacked multivariate model development and model performance assessment. 85

Figure 4.4: Trends of the temporal distribution for each nutrient at each farm compared during three growing season. Black, blue, and green lines refer to data points of Seasons 2020, 2021, and 2022, respectively. 94

Figure 5.1: Steps of sample collection, and spectral analysis for one data point. 101

Figure 5.2: Spectral measurements of one data point. (a) fresh leaves, (b) dried leaves. Shaded boxes highlight the water absorbance ranges. 103

Figure 5.3: Collection of spectral measurement in the field of both dried and fresh leaves. 104

Figure 5.4: Flowchart of the spectral files pre-processing steps performed. 106

Figure 5.5: Flowchart of the developed script to be run at the remote processing cloud. 108

Figure 5.6: Schematic of the proposed sensing system for non-destructively determining petiole nutrient values in near real-time from plant leaves in field, (a) acquiring spectral measurements of a plant leaf (tip leaf of 4th petiole), (b) storing the leaf spectral data in a computer memory, (c) sending the stored spectra to a remote storage device (d) computing the nutrient values of the petiole based on the stored leaf spectral data, and (e) delivering near-time results indicating the nutrient values of the petiole via a remote processing platform. 109

Figure 6.1: Differences in chemical concentrations between two labs of same samples of petioles at week (w). Blue refers to the results given by A&L Canada Laboratories, and the orange concentrations given by AgSource Laboratories. Sampling period of each field started from 45 days after planting till 135 days. 114

ABSTRACT

In this research project, a sensing system based on spectroscopy was developed to offer a solution to the problem of slow and tedious method of understanding the status of nutrients in potato plants using chemical tissue analysis. In the past several researchers have explored the use of ground-based sensors to estimate nutrient content in different crops based on leaf chemical content. However, in potatoes, petiole rather than leaf is the main organ required for tissue testing. Hence, the step in this research investigated the relationship between the chemical composition of petioles and the reflectance spectra of leaves. Initially, a linear relationship was assumed between the chemical content of petioles and leaf spectra within the Vis to NIR spectral range. Two modes of lab spectral analysis, dried and fresh, were considered. Univariate linear regression models were constructed to estimate each nutrient individually. The results showed that a correlation between the chemical contents of potato petioles and the leaf spectrum was confirmed for 12 elements. The results further showed that the predictive power was higher in the dried mode, but also N, P, K showed a high potential to be estimated based on fresh spectra.

Next, interlinkages between the nutrients were further assessed to enhance the predictive performance. Once the interlinkage was confirmed, a machine learning pipeline based on multivariate stacked regression was developed and implemented as it outperformed the univariate regression in estimating all nutrients. Finally, a full solution was developed based on portable spectrophotometer which can be deployed in the field. The spectrophotometer connects to a cloud computing, where the pipeline is stored, once internet is available to provide the farmer with nutrient status of plants in a matter of minutes in the case of fresh leaves and within 24 hours in the case of dried leaves. The rapid estimation of the nutrient status will help in better management of the amount of fertilizer applied during the potato growing season once the system is deployed in field.

LIST OF ABBREVIATIONS AND SYMBOLS USED

Abbreviation*	Meaning
Al	Aluminum
A_n	Actual concentrations at n -th data point from 1 to C
AOAC	Association of Official Analytical Chemists
B	Boron
B_i	Regression coefficient of the i -th waveband
β^0	Intercept
C	Number of data points
Ca	Calcium
CAD	Canadian dollar
Cu	Copper
$^{\circ}\text{C}$	Degree Celsius
E_n	Estimated concentrations at n -th data point from 1 to C
Fe	Iron
IR	Infrared
K	Potassium
Lasso	Least Absolute Shrinkage and Selection Operator
Max.	Maximum
Mg	Magnesium
MLR	Multiple linear regression
Min.	Minimum
MRMR	Minimum redundancy - maximum relevance
Mn	Manganese
N	Nitrogen
N	Index of data point 1, ..., C-1, C
NIR	Near-infrared
NIRS	Near-infrared spectroscopy
nRMSE	Normalized root mean square error
No.	Sample size

Abbreviation*	Meaning
P	Phosphorus
PLSR	Partial Least Squares Regression
r^2	Coefficient of determination
R	Pearson correlation
r_s	Spearman correlation
RF	Random Forest
RMSE	Root mean square error
RPD	The ratio of (standard error of) Prediction to (standard) Deviation
RPIQ	The ratio of performance to interquartile distance
S	Sulfur
SD	Standard deviation
SEP	Standard Error of Prediction
SWIR	Short wave infrared
SVR	Support Vector Regression
UV	Ultraviolet
Vis	Visible range
VNIR	Very near-infrared
XGB	Extreme Gradient Boosting
x_i	Spectral results of the leaves of the i -th waveband
Y	Chemical results of petioles
Zn	Zinc
Z	Vector of spectrum inputs
λ	Complexity parameter (Lambda)

* They appear in the list in alphabetical order rather than in the order in which they appear in the text.

ACKNOWLEDGMENTS

I am deeply grateful to the numerous individuals who have provided invaluable support throughout my graduate studies at Dalhousie University. First and foremost, I extend my utmost gratitude to Dr. Ahmad Al-Mallahi for his exceptional guidance, and profound knowledge, which have played a pivotal role in the successful completion of my doctoral project.

I am sincerely thankful to Dr. Felipe Campelo, a member of my advisory committee, for generously devoting his time and sharing his valuable expertise in data mining during the various stages of my project. Additionally, I would like to express my appreciation to Dr. Balakrishnan Prithiviraj for facilitating my access to his laboratory.

I would also like to thank New Brunswick Enabling Agricultural Research and Innovation program of the Canadian Agricultural Partnership (NBCAP), McCain Foods, Potatoes New Brunswick, and the Natural Science and Engineering Research Council of Canada (NSERC) for funding my project over the course of four years. I would like to acknowledge McCain Foods Company for granting me access to the fields at the Farm of the Future in New Brunswick, where I collected milestone samples. Additionally, I extend my thanks to Dave Bell, Bell Crop Services Inc., for his cooperation in facilitating my access to various farms in Lakeville and Waterville, New Brunswick. I would also like to acknowledge Picketa Systems Inc. for their collaboration and assistance in granting me access to different farms in Grand Falls, New Brunswick. Their support has been invaluable in the successful collection of my samples.

Furthermore, I want to thank Atlantic Farm Mechanization for their additional and generous scholarship. Special recognition goes to the Mitacs Globalink Research Award for generously funding my internship at Aston University, United Kingdom, which contributed

significantly to the successful completion of major project deliverables. I am also grateful to BMO Financial Group for their generous graduate scholarship in agriculture and to the Faculty of Agriculture of Dalhousie University for selecting me as a recipient of in-course scholarships.

I extend my gratitude to the research group, whose friendly demeanor and engaging personalities have created a pleasant research environment. I would like to express my appreciation to my dearest friends for their unwavering motivation, which has propelled me towards achieving my academic goals.

I extend my deepest appreciation to my mother (Refqa Abukmeil), my sisters (Rania, Suhad, Heba, Abeer and Lubna Abukmeil), my brothers (Alaa and Mohamed Abukmeil), my in-laws (Jeffrey, Misti, and Cody Campbell), and all other family members for their support and care during this transformative journey. Their love and encouragement have been indispensable in shaping my success today.

Lastly, I reserve a special dedication for my husband, (Colton Campbell) for his unwavering support, and constant encouragement. I am profoundly grateful for his presence in my life.

CHAPTER 1: INTRODUCTION

1.1. BACKGROUND

Potato (*Solanum tuberosum* L) is the premier vegetable crop across Canada, whose production encompasses regions from coast to coast, registering \$1.5 billion in potato receipts in 2021 (Agriculture and Agri-Food Canada, 2021-2022). Insights from the 2022 Canadian Potato Acreage Report underscore certain developments in the country potato crop, particularly in Eastern provinces where a reduction in seeded acreage has been observed (Potato growers of Canada, 2022). The report attributes this phenomenon primarily to elevated input costs for fuel and fertilizer, acting as pivotal determinants in shaping acreage trends. The implications of these cost increases have steered attention toward refining nutrient management practices to optimize crop quality and yields.

Thus, nutrient management holds paramount importance among potato growers. The nutritional composition of the potato plant is a response to the availability of macro and micronutrients in soil. Macronutrients are taken in large quantities in regard to their physiological functions in plant metabolism and for the tuber yield formation, including nitrogen (N), phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), and sulfur (S), while the micronutrients are taken in small quantities such as manganese (Mn), zinc (Zn), iron (Fe), sodium (Na), copper (Cu), aluminum (Al), and boron (B), whose inclusion in the fertilizer schedule is very necessary to sustain production and quality (Koch et al., 2020).

Application of fertilizers commonly includes pre-planting, banding before or at planting, side-dressing after planting, spraying the foliage during crop growth, and injection into irrigation water (Rowe, 1993). Each method has its own significance, for

instance, pre-plant application is important to eliminate the need for additional fertilizer during planting, and the bands are necessary to provide benefits for early growth, side-dressing is necessary during the growth stage, for nutrients that will not stimulate early growth such as N. Spraying onto foliage is rather effective for treating some nutrient deficiencies and for overcoming the nutrients that are probably less available for plant uptake due to soil fixation. Finally injecting fertilizers into irrigation water has the advantage of providing nutrients according to the needs of the crop (Rowe, 1993).

Many factors affect the response of a potato plant to fertilizer applications, including soil microorganisms that play roles in nutrient cycling, tillage practices that affect soil compaction, and rooting depth, overirrigation that may cause leaching of nutrients, large weed populations that compete for nutrient uptake, and pest existence that cause damage to healthy plants and thus prevent nutrients uptake (Rowe, 1993). Thus, a careful management of fertilizer application is required within the season. An additional factor for the careful management of in-season application is that potato plants are late-maturing cultivars which means the need for N nutrients might be greater within the growth stage (Rowe, 1993). P fertilization is commonly applied to the soil in the fall or in the spring before planting, and it may also be banded at planting (Rowe, 1993). However, P is also applied in foliar application once P deficiency develops during the season (Hopkins et al., 2014). Similarly, K, S, and micronutrients are available for foliar application only once the deficiency is likely to occur. Ca and Mg are needed for potato production in acid soils, and thus commonly applied before and at planting (Rowe, 1993).

The amount of recommended fertilizer application is commonly based on yield potentials, field history, and soil tests (Bohl and Johnson, 2010). Chemical testing of plant

tissues is also widely used as a diagnostic tool to determine nutrient status during crop growth (Rowe, 1993). Soil tests are commonly conducted before planting to determine the levels of nutrients existing in farms, however, the challenges of this method encounter shallow sampling depth, mis-selection of sampling zones, and immobile nutrients in soils that are unnecessarily taken by plants (Jafarbiglu, and Pourreza, 2022). As a result, the focus on tissue-based nutrient analysis increased to be an indicative tool in determining the levels of nutrients in plants (Motsara and Roy, 2008).

In particular, the growers in New Brunswick (NB) care about the application not only for N, but also for P, Mg, Mn, B, Zn, and S, as the soil in NB is acidic and may contain high levels of soluble Al and Mn (Fahmy et al., 2010). This led the growers to apply different types of limes (such as dolomitic limestone) to change those soluble micronutrients into less soluble chemical forms and increase the availability of Ca and Mg (Gransee and Führs, 2013). On the other hand, monitoring B is necessary as it is susceptible to causing hollow heart disease (Isaacs, 2020). Zn is also a key micronutrient for potato plants for its role in photosynthesis and cell growth (Isaacs, 2020). Therefore, tissue testing is followed as a decision tool for growers to manage the quantity and time of nutrient application.

Nevertheless, the complexity of collecting tissue samples is burdening the process. For instance, tissue sampling of potato plants has a particular protocol to be followed, otherwise analysis of samples gives significantly different results that do not represent the actual nutritional status of the plant. The protocol states that the petiole¹ of the fourth leaf from the top of the potato plant must be collected for chemical testing because their nutrient

¹ The slender stalk that connects the leaf blade (lamina) to the stem.

concentrations reflect well the nutritional status of the plant (Rowe, 1993). The common practice is to strip the leaflets off the petiole and collect 40-50 petioles from different neighbouring potato plants from a specific sampled area (Zebarth et al., 2007) (Figure 1.1). As the nutrient concentrations in tissues vary from one growth stage to another due to the differences in the mobility of nutrients within the plant, it is recommended to start collecting the tissue samples every 7-10 days starting 40-45 days after planting (Rowe, 1993). However, as the challenges encountered in selecting the correct petiole and the time gap between sample collection and testing, using spectral techniques are under study to identify nutritional status by using the reflectance spectra of leaves.

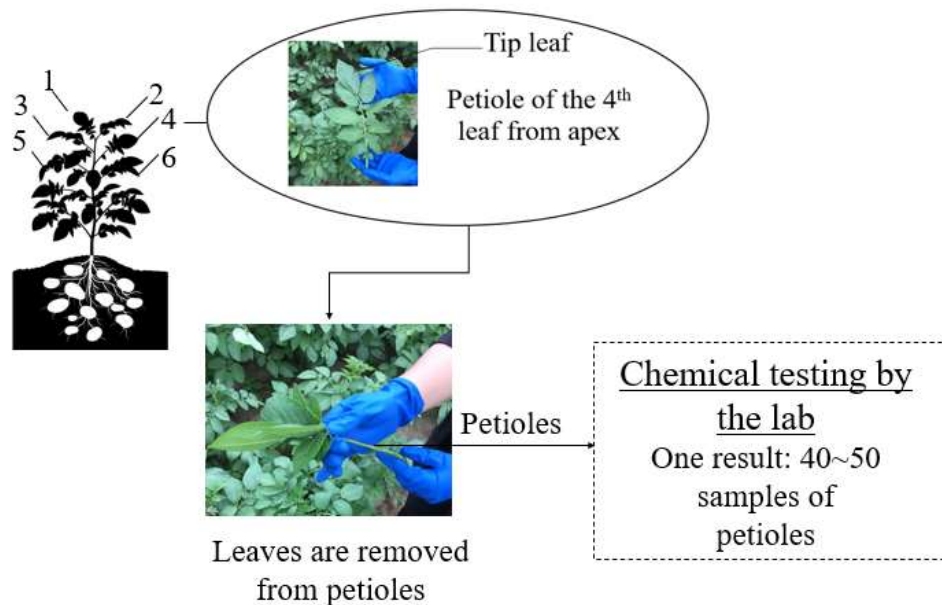


Figure 1.1: Current approach of petioles samples collection for lab chemical testing

1.2. APPLICATION OF SPECTROSCOPY FOR NUTRIENTS ASSESSMENT IN AGRICULTURE

Herschel's discovery of infrared (IR) energy back in 1800 marked the beginning era of understanding the existing phenomenon of energy absorption that occurs in the IR region (Williams et al., 2019). The discovery was about energy absorbers that

fundamentally occur within the mid-IR, and their overtones occurring in the NIR region (Williams et al., 2019). This discovery later translated into industrial applications that were first initiated by Karl Norris for agricultural purposes (Williams et al., 2019). Their application was the first to successfully use NIR spectroscopy (NIRS) to determine the moisture content and protein in soybeans (Williams et al., 2019).

The work concept of NIRS is to measure the intensity of interaction between light and matter via a detector that receives the reflected or transmitted light (Worsfold and Zagatto, 2017). When incident light hits a plant leaf, some of it gets reflected directly from a sample holding no information, some other light energy will get completely scattered within the sample, and other energy would be diffusely reflected or transmitted reaching a spectroscopic detector (Williams et al., 2019). In plants, diffuse reflectance within the Vis range primarily occurs due to photosynthetic pigments, whereas water molecules in the leaf absorb the light in the short-wave infrared (SWIR) region (Figure 1.2). Although the reflectance spectrum of leaves is consistent, they differ between plant types which give different intensities of reflected light. Previous studies have identified significant spectral wavebands within the Vis and very near-infrared range (VNIR) (400 - 1100 nm) and the SWIR (1000 - 3000 nm) in applications related to forestry and crop analysis (Saari et al., 2011; Jenal et al., 2019a; Jenal et al., 2019b). The NIRS has been introduced to agriculture either by using non-imaging or imaging spectral instruments such as remote sensors and ground-based sensors (Figure 1.3) - as explained in more detail in the following section.

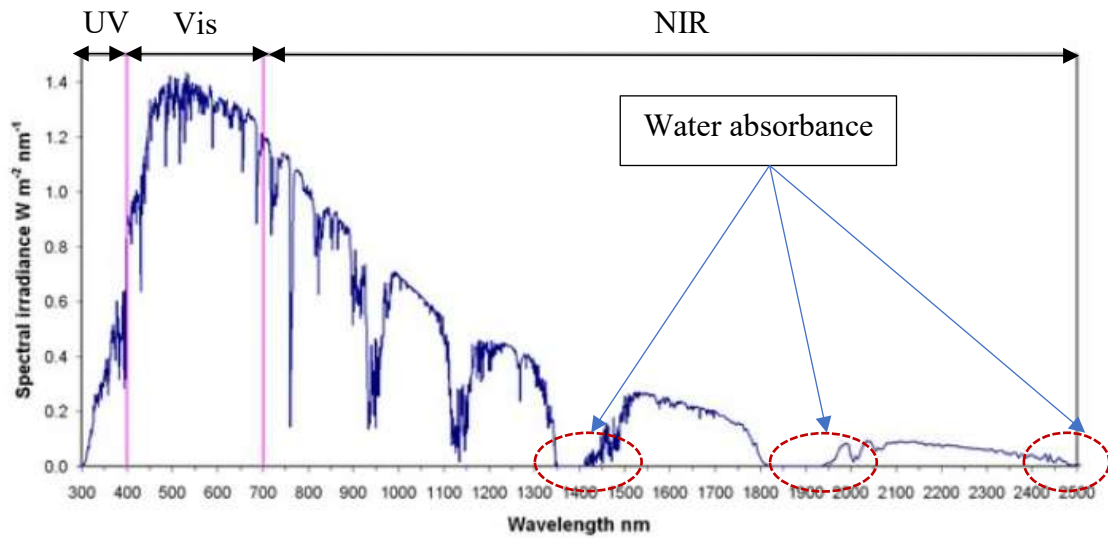
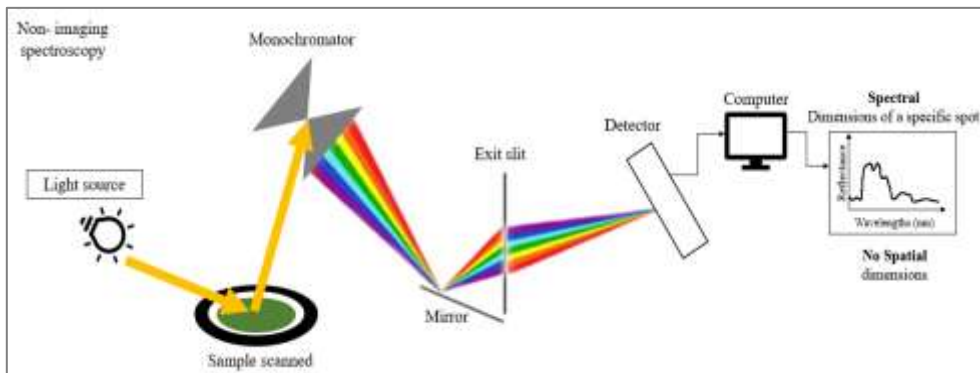
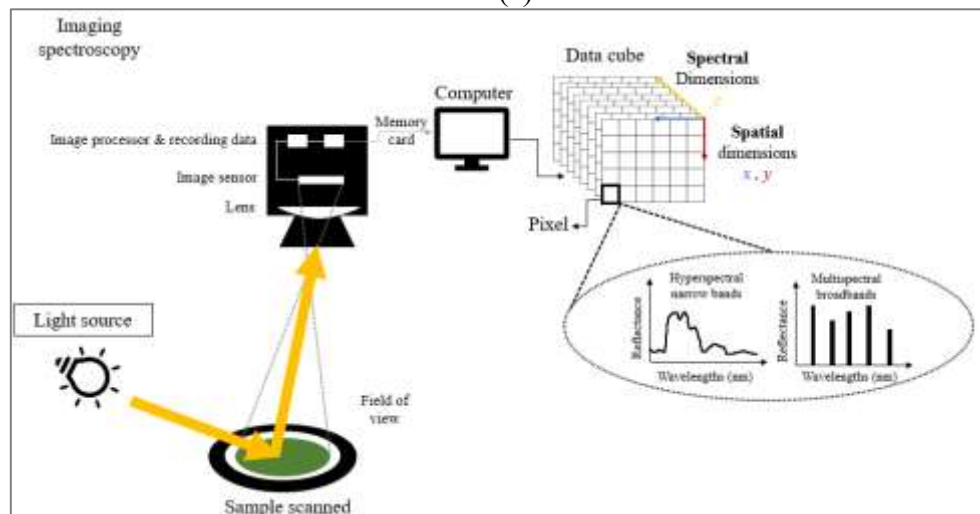


Figure 1.2: The spectrum of sunlight at sea level on a bright day. The red circles refer to the water absorption wavebands by atmospheric water vapor



(a)



(b)

Figure 1.3: Comparison between: (a) non-imaging, and (b) imaging spectroscopy

1.2.1. Remote sensors - advantages, and challenges

Applications of remote sensor platforms in agriculture such as satellites and aircraft are widely used to detect stressed plants by obtaining the reflectance information using electromagnetic range of Vis and NIR regions. Choosing among these platforms depends on which information is required (Jafarbiglu, and Pourreza, 2022). For instance, satellite maps cover wide areas, but at a lower spatial resolution, while the unmanned aerial vehicle (UAV) can provide high spatial resolutions, but at a small-scale area (Jafarbiglu, and Pourreza, 2022).

The use of the Vis and NIR regions in remote sensors is based on selective wavebands called vegetation indices (VIs) (Morcillo-Pallarés et al., 2019). The rationale behind the use of VIs is their simplicity to reduce the undesired effects related to variations of other canopy properties, background soil reflectance, solar illumination, and atmospheric composition. Those VIs are combinations of two or more wavelengths designed to highlight essential information about vegetation based on the concept that the chlorophyll pigment of the healthy plant reflects green wavebands and absorbs the red ones (Solano-Alvarez et al., 2022). In 1969, a simple algebraic relation between NIR and red wavebands called “Normalized Difference Vegetation Index (NDVI)” was proposed as an indicator to simplify the complexity of collected satellite images by quickly delineating vegetation and vegetative stress (Kriegler et al., 1969). The NDVI was approved for its capability to assess plant health for its sufficient correlation with wheat cover and its effects on grain yield using red (670 nm) and NIR (780 nm) wavebands (Phillips et al., 2004). NDVI at the red and NIR wavebands (670 nm and 780 nm) were also found to be

highly correlated with rice N content as a potential tool to improve N fertilization (Rehman et al., 2019).

The common use of remote sensing imagery in agriculture came from its free access using multi-spectral earth-observing satellites including Landsat 8, Sentinel-2 (S2), and Sentinel-3 (S3), where such imagery applications are composed of several wavebands. For instance, the Landsat 8 satellite produces 11 band images, with the wavelength of the bands being between 435 and 12,510 nm, and S2 produces 13 bands within the range of 433 and 2280 nm (Morcillo-Pallarés et al., 2019). However, the use of satellite-based remote sensing is still limited due to their low spatial resolution and high altitude (Jafarbiglu, and Pourreza, 2022). Thus, the application of UAV for field phenotyping was emerged.

An earlier study assessed the ability of UAV equipped with a multispectral camera at wavebands of 520–600, 630–690, and 760–900 nm to highlight the differences in geometrical and spectral canopy characteristics between eight olive cultivars (Caruso et al., 2021). In their study, they concluded that there was no relationship between NDVI and fruit yield when abiotic and biotic factors affected final fruit yield. A further study assessed the potential of a multispectral imaging camera to study canopies in the olive orchard (Caruso et al., 2022). Their results confirm the reliability of UAV imagery in estimating the olive geometrical canopy characteristics. A study by Pádua et al. (2020) proposed the use of different imagery sensors such as RGB, multispectral and thermal imagery combined to acquire high spatiotemporal resolution of images for the aim to extract individual grapevine geometrical and biophysical parameters within vineyard plots. Their study confirmed the effectiveness of their method to assess different estimated geometrical and biophysical parameters for the vineyard.

However, the use of specific wavebands within the multispectral imagery and Vis might drop some important information in other parts of the spectrum (Zhang et al., 2022). For instance, monitoring leaf K content in rice leaves was improved by developing three-band spectral indices involving the red edge wavebands instead of two-band spectral indices (Lu et al., 2020). Another study concluded that the complex structure of plant-environment interactions can be assessed using the information provided by different VIs (Bojinov et al., 2022). Other studies evaluated the potential of hyperspectral camera using UAV platforms to detect plant status. A study explored the feasibility of estimating crop traits in maize using hyperspectral camera of a spectral range within 450-950 nm (Shu et al., 2021). Their study successfully identified the sensitive wavebands for the optimal estimation of maize traits. Another study used hyperspectral camera within the range of 400 to 1000 nm and between 953 and 2518 nm to monitor the crop nutrient status of potatoes over two growing seasons (Liu et al., 2021a). Their study demonstrated the capacity of imaging spectroscopy to quantify the nutrient status and to predict tuber yield in potatoes across cultivars, growth stages, and growing seasons.

The large size of data generated from hyperspectral camera still complicates the computing process. In addition, the use of remote sensors of both multi and hyperspectral camera also requires the exclusion of the disturbances caused by wind and varying sunlight (Svensgaard et al., 2014; Rasti et al., 2018). Such disturbances could alter the robustness of VIs by environmental and climatic conditions (Bajocco et al., 2022). Moreover, the electromagnetic signals must pass through the atmosphere to reach the sensor on the remote sensing platform (either satellite or UAV), and based on the atmospheric condition, some

noise will be intermingled with signals, e.g., water vapor strongly absorbs wavebands of 1350 - 1480 nm and 1780 - 2032 nm (Asner and Martin, 2015).

One additional challenge to the remote sensing platforms is related to the soil-adjustment factor which is used as a canopy structure parameter while extracting the information from the images. This factor is quite critical in minimizing soil background noise when the canopy is heterogeneous with low vegetation cover (Zhen et al., 2020). Remote VIs data also need to be calibrated through ground-truth data to minimize errors associated with spectra capturing and variation in the plant, background effects, and climatic variability (Yang et al., 2017). Thus, proximal ground-based sensors can be a suitable alternative to remote sensing platforms.

1.2.2. Proximal sensors - advantages and challenges

Proximal sensing has been assessed for its ability to determine phenotyping changes in both indoor and field studies using imaging and non-imaging spectroscopy. A study assessed the use of hyperspectral camera of a spectral range from 380 to 1012 nm to determine phenotypic changes on lettuce plants grown in a controlled environment growth chamber for lettuce (Simko et al., 2016). Their study showed that hyperspectral camera was able to detect changes in the total chlorophyll and anthocyanin contents in lettuce. Another study concluded the possibility of using hyperspectral camera to determine the shikimic acid concentration in glyphosate-tolerant transgenic maize (Feng et al., 2018). Their study was implemented in a greenhouse environment and the hyperspectral images were acquired over the Vis and NIR range from 380 to 1030 nm. Another study incorporated the results of using the same camera to assess the disease severity of barley cultivars in a greenhouse (Thomas et al., 2018). Hyperspectral camera was also assessed

for its ability to detect mold disease in tomato leaves grown in a controlled room using the spectral range of 380 to 1030 nm (Kong et al., 2014). A hyperspectral camera within the spectral range from 550 to 1700 nm was successfully able to quantify chemical properties in maize and soybean plants (Pandey et al., 2017). Another study used a laboratory hyperspectral camera over a range of 380 nm to 1030 nm to estimate the seed yield of oilseed rape leaves taken from field trials (Zhang and He, 2013). Undoubtedly, those indoor experiments provided a more controlled phenotyping environment and higher spatial resolution compared to field setups, as the effect of plant geometry on illumination and reflectance variation is more distinct. However, such an environment of non-biological reflectance variation can mask actual biological effects which might misguide the interpretation of hyperspectral data (Mertens et al., 2021).

Other studies explored the application of hyperspectral camera in outdoor field conditions. A study assessed various optical sensing including hyperspectral camera between 970 nm and 1670 nm for non-destructive field-based phenotyping in grain cereals (Busemeyer et al., 2013). Their study concluded the need to develop mechanical interfaces of machines to enhance the economic value of using standard machines. Another study used on-the-go proximal hyperspectral camera (400 nm to 1000 nm) using natural illumination to classify a large number of grapevines under field conditions (Gutiérrez et al., 2018). Their study showed the potentials of deploying on-the-go hyperspectral camera to classify many grapevine varieties based on real outdoor conditions.

Nevertheless, the challenge for image processing is that the reference values should be known. Determination of these values at each pixel position in an image is not feasible, and thus an average reference value is commonly used. The drawback in selecting an

average reference pixel is that it is simple in clear images but can be less obvious when images have a poor spatial structure (Juan, 2019). Another study mentioned that if the reference pixel can not be selected randomly among the pixels with a high average spatial, it could be specified using prior knowledge of the study area (Yunjun et al., 2019). However, selecting a prior known reference pixel will consume a long time of processing. Thus, the non-imaging spectroscopy using hyperspectral and multispectral sensors has been intensively studied to estimate crop traits (Table 1.1).

Few studies assessed the performance of hyperspectral cameras in comparison to non-imaging spectrophotometer. A study evaluated the Vis and NIR taken by a camera in comparison with a dedicated NDVI sensor to calculate NDVI and to identify water stress in apples and citrus (Glenn and Tabb, 2019). Their study showed that the NDVI measurement by the camera images was correlated with measurements from a chlorophyll meter. A study used a non-imaging spectrophotometer to verify the accuracy and feasibility of camera images to retrieve maize leaf chlorophyll content (Zhang et al., 2012). It indicated that estimating chlorophyll content was feasible using hyperspectral camera.

Yet, studying hyperspectral sensing is needed to identify the significant wavebands to develop models that could be transformed into multispectral devices in later stages. For instance, some commercial sensors are already available in the market using specific wavebands of multispectral sensors. The most well-known are the N-sensor (Yara Canada, Regina, Saskatchewan), Crop Circle (Holland Scientific, Nebraska, USA), and GreenSeeker handheld crop sensor (Trimble Inc., Westminster, Colorado, USA) which are capable of real-time assessment of plant N status with the simultaneous application of N fertilizers (Mezera et al., 2021). Such sensors correlate multispectral reflectance

wavebands to capture the red and NIR data for subsequent measurements of NDVI. A study found that both Yara-N and GreenSeeker were able to describe N variation in the wheat and corn canopies (Tremblay et al., 2009). Another study concluded that CropCircle and GreenSeeker effectively measure the NDVI for maize fields (Shaver et al., 2011). Another study assessed the application of the same sensors of CropCircle and GreenSeeker in detecting P deficiency in potato plants (Jasim et al., 2020). The study showed that both sensors had a low predictive value for P uptake. The study revealed that such sensors were exclusively used to detect N stress rather than P stress.

This exclusive use of sensors in the market for N detection came from its importance as a key protein component in chlorophyll molecules and enzymes crucial for photosynthetic processes (Koch et al., 2020). However, the shortage of other nutrients, particularly K, and P can result in different stress-induced responses such as restricted shoots and root growth, early defoliation of older leaves, and decreased biomass yield (Mee et al., 2016). On the other hand, excessive plant nutrition, and particularly an excess of N, can lead to lush growth which leads to dense plant canopies and consequently traps humidity within the canopy, creating conditions for fungal diseases, in addition to the risk of significant portions of N (Koch et al., 2020). Moreover, other nutrients may require exploiting the range of spectrum from Vis to NIR rather than the analysis over specific spectral regions (Peng et al., 2020). The appearance of advanced analytical methods allows producing high multidimensional data of Vis and NIR spectra to extract relevant information encrypted in the NIR range.

1.3. THE APPLICATION OF VIS-NIRS IN ESTIMATING NUTRIENTS IN POTATO PLANTS

In the previous section, an overview has been discussed on the deployment of sensing techniques such as non-imaging and imaging spectrophotometers in agriculture to detect the nutritional status of the plants. Ground-based sensors are already delivered to markets to estimate plant properties (Gabriel et al., 2017), and remote sensors are also widely used to detect stressed plants by obtaining electromagnetic wave reflectance information from the canopy (Xue and Su, 2017). However, both sensors have drawbacks related to canopy reflectance including atmospheric and soil interference as discussed earlier (Muñoz-Huerta et al., 2013), and therefore, several studies analyzed the reflectance at the leaf level to eliminate noise coming from atmospheric and soil interference such as Mahajan et al. (2021), Peng et al. (2020), and Liao et al. (2012). These studies are based on the chemical testing of the leaves as a reference point. Unlike potato plants, the petiole is main organ for chemical testing and not the leaf. One study used chlorophyll meter to assess N in potato petiole for the varieties of Ranger Russet and Russet Burbank early in the growing season (Davenport et al., 2005). Another study assessed potato leaf N content using leaf reflectance based on leaf chlorophyll and protein content (Botha et al., 2006). Finally, Liu et al. (2021a) were able to predict foliar N and petiole nitrate at different wavelength regions of different potato cultivars and planting seasons using aerial hyperspectral platform. No studies found in literature that have assessed the petiole chemical content using Vis-NIR data of leaves. In addition, no studies compared the estimation of petiole chemical contents using fresh and dried spectral scan of leaves.

1.4. SPECTROSCOPY DATA ANALYSIS

Part of the complexity of spectroscopy data analysis comes from overlapping effects of multiple nutrient concentrations on the spectral signature, which needs to be considered when estimating those concentrations using the intensity at different wavebands. In addition, the little spectral differences between different samples make it tough to interpret the data. As such, different machine learning algorithms have been employed to address such complexities (Sampaio and Brites, 2022). Such algorithms are commonly used for modeling spectral data and there are two types: (a) classification, which predicts qualitative results, and (b) regression, which predicts quantitative results. The concept of work of such algorithms is to learn a model structure from labeled data, i.e., data points for which the value of the outcome variable is known. Once trained, these models are deployed to estimate the outcome variable for new data points (Sampaio and Brites, 2022). Several studies assessed the robustness of using classification and regression algorithms for agricultural purposes (Table 1.1). One study compared the efficacy of combining linear and nonlinear algorithms to estimate the foliar macro and micronutrient status of mango (Mahajan et al., 2021). Their study concluded that PLSR-combined with SVR, RF, XGB, KNN machine learning models were found to be the best to predict most of the nutrients. All the studies mentioned in Table 1.1 involved the use of machine learning algorithms using leaf chemical content as the reference values.

Table 1.1: List of studies used different machine learning algorithms to extract information from the NIR data to estimate nutrients

Nutrient under study/ Crop	Wavelength region (nm)	Used method of data analysis	Reference
N, P, K, Fe, Mn, Zn, Cu/ Fingered citron leaf	1100 – 2498	PLSR, PCR	Liao et al. (2012)
N, P, K, Ca, Mg, Fe, Zn, Mn/ Orange leaf	400 – 1000	PLSR	Menesatti et al. (2010)
N, K, Ca, Mg, B, Fe, Cu, Mn, Zn/ Citrus leaf	830 – 2600	PLSR	Galvez-Sola et al. (2015)
K/ Ping'ou hybrid hazelnut	310 - 1130	Linear and nonlinear models	Zhao, and Pan (2020)
Fe, Zn/ Rice leaf	325 - 1075	PCA, combined least squares- support vector machine	Shao and He (2013)
N, P, K, Mg, Fe/ Bean leaf	Multispectral sensor at 640, 550, 475 , 510 – 590, 730, 769, 710	LDA	Lazarević et al. (2022)
N/ Potato foliage	Multispectral sensor at 670, 730, 780	RF	Peng et al. (2021)
N, P, K, Mg, S, Cu, Fe, Mn, Zn/ Valencia-orange leaf	380 – 1020	RF, and KNN, Lasso, ridge regression, SVM, ANN	Oscos et al. (2020)
N, P, K/ Aquaponics- grown plants	350 – 2500	PCA, PLSR, RF	Taha et al. (2022)
N, P, K/ Different plant species including rice, and corn	350 – 2500	PLSR, SVR	Zhai et al. (2013)

The studies, which tried to estimate nutrients based on petiole chemical content, used linear regression model in the study by Davenport et al. (2005), PLSR and ordinary least squares regression by Liu et al. (2021a). Assessing the performance of other nonlinear machine learning algorithms was not covered although they showed high capacity in extracting information from NIR as the study of Mahajan et al. (2021). In addition, the multicollinearity between nutrients in relation to the spectral data might be more biologically relevant. For instance, K and Mg are known for their antagonisms as

competitive cations for plant uptake, and Ca and Mg are also known for their synergic contribution to maintaining a stable pH in soil (Koch et al., 2020). Therefore, assessing the interlinkages among the nutrients when extracting the information from the spectral data is required to convey crucial information about the sample.

1.5. RESEARCH GOAL AND OBJECTIVES

Based on the research gaps identified earlier, the goal of this project is to develop a non-destructive sensing system which estimates nutrients in potato plants. This system shall improve the current sampling method, which is going to help in reducing the analysis time. The objectives of this research are to:

- understand the relationship between nutrient elements in petioles and spectral reflectance of leaves by comparing fresh and dried leaves and finding significant wavelengths.
- find interlinkages between nutrients and their effects on the predictive performance using reflectance spectra.
- develop a machine learning pipeline based on stacked multivariate models for less-destructive and non-destructive sensing system for dried and fresh leaves, respectively based on the interlinkage
- develop a comprehensive solution for rapid delivery of nutrient status in potato plants that outperforms the current practice of chemical tissue analysis.

1.6. RESEARCH HYPOTHESIS

The nutritional requirements of the potato plant change at various stages of crop development. The stages are divided into five: sprout development, vegetative growth, tuber initiation, tuber bulking, and maturation. Photosynthesis begins during the second

stage to provide energy for the growth and development of vegetative parts of the plant. Roots actively absorb plant nutrients from the soil during this stage to support the following stages of tuber initiation and bulking. Thus, the portion of nutrients absorbed by the plant depends upon the mobility of each nutrient within the plant (Rowe, 1993).

Nutritional disorders can result from deficiencies, and excesses between nutrients in the soil or as they are taken up by the plant. Nutrient deficiencies in potato plants can sometimes be identified from visible symptoms on the tip leaves. For example, P deficiency will cause dark green leaves with margins rolling upward, Fe deficiency will turn leaves yellow to nearly white with leaf tips and edges remaining green, and Cu deficiency will wilt the tip leaves causing them to die (Rowe, 1993).

Even though nutrient deficiencies could appear on leaves, visual diagnoses can still be misleading as other diseases resemble the same symptoms. Moreover, by the time symptoms are visible, it is often too late to make corrections and thus more precise and timely identification of nutrient deficiencies in potato plants is commonly obtained from a chemical testing of plant tissues (Rowe, 1993). As mentioned earlier, the deployment of spectral instruments in evaluating the nutritional composition of plants showed immense success within the Vis and NIR ranges. However, as the petiole is main organ of potato plants for estimating nutrients, the application of spectral instruments over a petiole is impractical.

Therefore, hypothesis 1 of this research is that there exists an indirect correlation between the spectral data obtained from the leaves of potato plants, particularly focusing on the tip leaves, and the chemical concentration in petioles. Hypothesis 2 is that the interlinkages between nutrients may influence the estimation performance on one another.

CHAPTER 2: ASSESSMENT OF THE CORRELATION BETWEEN PETIOLE CHEMICAL CONTENT AND LEAF REFLECTANCE SPECTRA OF POTATO PLANTS

2.1. BACKGROUND

Utilizing spectral instruments in agriculture has demonstrated significant promise for evaluating the nutritional composition of plants. This can be achieved through either remote or proximal platforms (See Section 1.2). Deploying spectral instruments over a petiole mentioned to be impractical due to their slender shape. Thus, one of the research gaps identified is that the correlation between the petiole chemical content and leaf spectra needs to be analyzed. In addition, spectral analysis over the range of 400 nm to 2500 nm has not been widely analyzed to predict potato nutrient status. Therefore, the aim of this chapter is to analyze the relationship between petiole chemical testing and leaf reflectance spectra within this range of spectra.

2.2. DATA COLLECTION OF PETIOLES AND LEAVES

2.2.1. Data sampling of outdoor fields

The experimental procedure of this research adhered to the existing protocol used by potato growers in NB for sample collection, preparation, and chemical testing. As a first step, a total of 40 data points were collected from subplots at two potato farms located in Lakeville, NB, focusing on the Russet Burbank variety (Figure 2.1). This dataset is referred to as "open-field data". Each data point comprised of 40 petioles and 40 leaves, which were subjected to lab chemical testing and spectral analysis, respectively. This quantity of petioles was determined based on the requirement of DairyLand Lab Inc. (Arcadia, Wisconsin, USA), where the analysis took place. They specified a dry weight of 3 grams (fresh weight > 32 grams) for accurate chemical testing. Sampling took place between late

June and late September 2020 every other week at 40-45 days after planting. This time was chosen based on the suggestion given by Zebarth et al. (2007), as this period provides the most accurate assessment of crop nutrient status. Random sampling was conducted within the same subplots throughout the season at each location. After peeling the leaves off the petioles, they were immediately vacuum packed into sampling bags and refrigerated to maintain their freshness before shipment. The samples were packed with ice bags and delivered to the lab, with a two-day time lag until reception. Figure 2.2 shows the steps taken for sampling and analysis. Both petiole and leaf samples from the farm were collected from the fourth leaf from the tip of the plants.



Figure 2.1: Location of the open-field farms where samples were collected from NB. Green circles refer to the areas where samples collected in 2020, 2021, and 2022. Blue circle shows an additional area for data collection in 2022

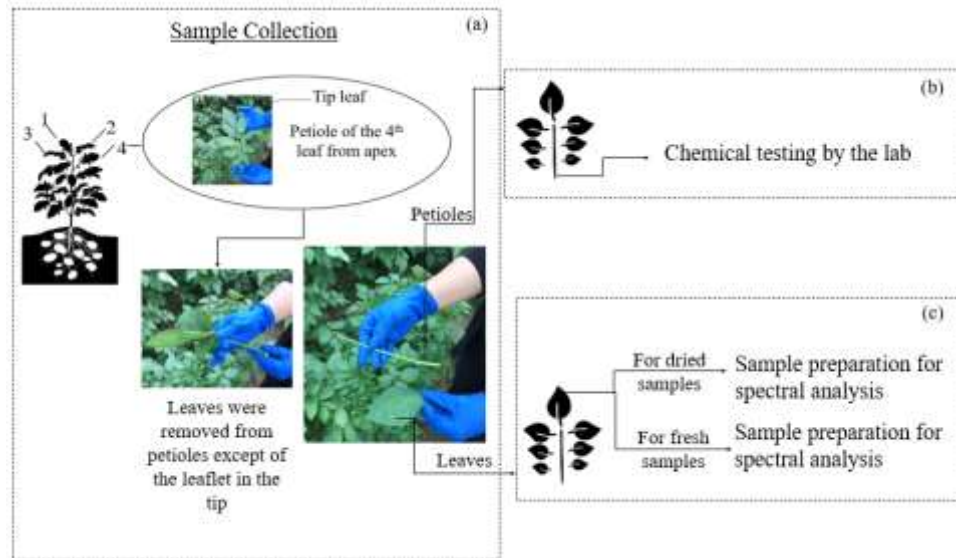


Figure 2.2: (a) Steps of sample collection. (b) Chemical testing of petioles. (c) Spectral analysis of dried and fresh leaves performed over the two modes of dried and fresh leaves

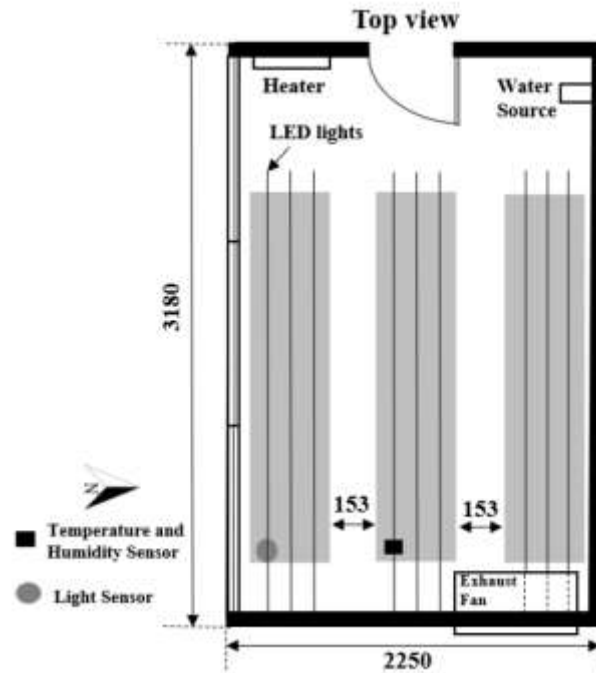
2.2.2. Data sampling of indoor growing area

Crop growth is affected by several factors from the time of planting until harvesting, including fertilization, irrigation, weed, and pest control. Later fertilizer application is up to the growers experience to try maximizing the yield. Due to the complexities occurring at the open-field farming, there is a likelihood of interdependence among the nutrients, impacting the overall nutrient status of the plants. Conducting independent experiments to test each nutrient separately would indeed be a favorable approach to understand the individual effects of each nutrient. Thus, indoor growing area gives me the opportunity to apply different fertilization schemes.

The designated area utilized for this purpose was initially a sunroom located at the Department of Engineering in the Agriculture Campus of Dalhousie University in Truro, Nova Scotia (45°22'30.1"N 63°15'50.6" W). In addition, various measures were implemented to regulate lighting and temperature levels to maintain a controlled

environment. This was achieved through the integration of sensors, microcontrollers, and actuators (Figure 2.3). The experiment ran over a time span from August till December 2020.

By conducting the experiment in a controlled environment, it was possible to vary the fertilization practices systematically. Additionally, conducting the experiment in an indoor setting provided the advantage of enabling sample collection off-season. Two fertilization schemes were followed for the indoor growing, an over application for NPK in one group (20-20-20 NPK application) and a cut in P content for the second group (22-0-22 NPK weekly) from the fourth week until the end of the experiment. Twenty data points were collected following the same sampling methodology presented for the open-field sample collection discussed in Section (2.2.1).



(a)



(b)



(c)

Figure 2.3: Layout and internal view of the indoor growing area during potato planting (a) location of sensors inside, and the location of LED lights using vertical lines, (b) early growing stage, (c) vegetation growing stage

During the practical implementation of the experiment, certain challenges were encountered related to sample size and environmental conditions. One of the challenges arose from the fact that the petioles of the potato plants grown indoors had a lower weight compared to those grown in open-farming conditions. As a result, the sample size obtained from the indoor growing was insufficient to conduct the necessary chemical testing that

required a minimum weight of 3 g. To address this issue, it was necessary to collect 70 petioles instead of 40 per data point.

2.3. ANALYSIS OF SAMPLES (CHEMICAL TESTING AND SPECTRAL ANALYSIS)

The 60 data points collected from open-farming, and indoor growing were considered sufficient to proceed with the initial analysis to find the correlation between petiole chemical content and leaf reflectance spectra. The chemical testing of petioles of each data point included a comprehensive analysis of all macro and micronutrients that are commonly examined by Canadian farmers. The petioles were subjected to a drying process at a consistent temperature of 55-60 degrees Celsius (°C) for a duration of 16-24 hours, ensuring that a constant weight was achieved. Chemical testing for all macro and micro-nutrients was conducted using the official methods prescribed by the Association of Official Analytical Chemists (AOAC).

The leaves at each data point were divided into two distinct testing modes: fresh and dried, with each group consisting of 20 leaves, henceforth called *fresh* and *dried*. The dried group of leaves underwent the same drying procedure as the petioles. The leaves were analyzed for their spectral reflectance at each mode using NIRS Analyzer (DS2500, Metrohm USA Inc.). The operating specifications of the spectrophotometer are illustrated in Table 2.1, including the range of wavelengths and spectral resolution.

Table 2.1: Operating specifications of NIRS DS2500 Analyzer

Item	Specification
Measurement Mode	Reflectance
Wavelength Range	400 - 2500 nm
Detectors	Silicon (400 - 1100 nm) and Lead Sulfide (1100 - 2500 nm)
Optical Bandwidth	8.75 ±0.1 nm
Spectral resolution	0.5 nm
Number of wavelengths	4,200

The NIRS Analyzer utilized in this research measures the reflectance of leaves across a range of 400-2500 nm. However, the data generated by the NIRS Analyzer WinISI software were presented in terms of absorbance ($\log(1/reflectance)$) after conversion. The spectral measurements were obtained at an interval of 0.5 nm, resulting in a total of 4,200 readings. The absorbance values were converted back to reflectance values using the relationship ($10^{-Absorbance}$). Instead of utilizing the entire set of 4,200 readings, a single reading was selected at intervals of 8 nm (equivalent to every 16 readings considering the spectral resolution of 0.5 nm) starting from 404 nm. These representative spectral signatures were used for data analysis, resulting in a total of 262 readings. Figure 2.4 shows the raw reflectance spectra of both testing modes highlighting the water absorption wavebands at 1350 to 1480 nm, 1780 nm to 2032 and after 2400 nm as mentioned in Chapter 1 (Asner and Martin, 2015).

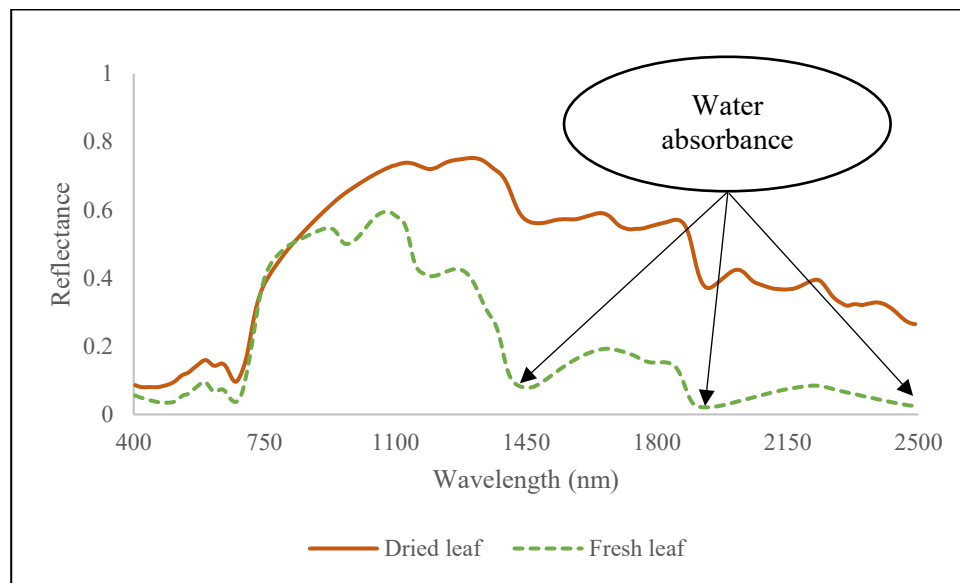


Figure 2.4: Difference between reflectance spectra of dried and fresh leaves

2.4. DATASET CONSTRUCTION USING LINEAR RELATIONSHIP

In this initial analysis, an assumption was made that a linear relationship exists between the chemical content of petioles and the spectral reflectance of leaves. Figure 2.5 (a) shows the format in which the lab analysis of petioles is received, where macronutrients and Na are listed in percentages (%), whereas the rest of micronutrients are in particles per million (ppm). Figure 2.5 (b) also shows the format of receiving the raw spectra in an absorbance format.

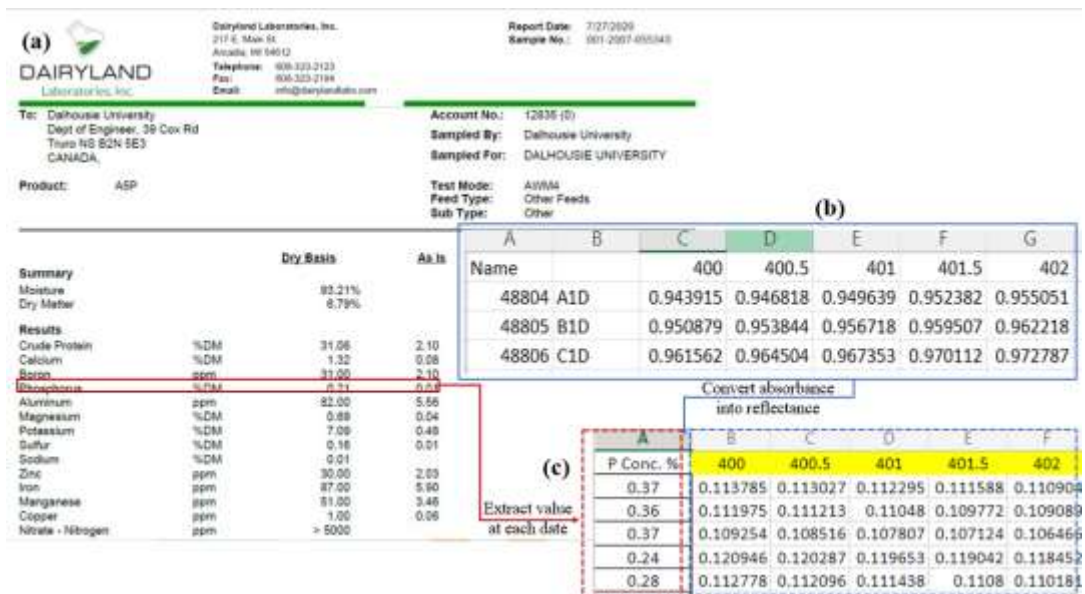


Figure 2.5: Steps taken to handle the raw data as inputs into the dataset (a) raw format of the chemical content results, (b) raw format of receiving the spectra data, and (c) development of a dataset based on a linear relationship

Based on the aforementioned assumption, multiple univariate linear regression (MLR) models were employed, and as a first step, datasets were built by the spectra acted as the predictor variables (x), and the measured nutrient content acted as the response variable (y). The datasets were then formulated by using the standard MLR equation (2.1):

$$y = \beta_0 + \sum_{i=1}^Z x_i \beta_i + \epsilon, \quad (2.1)$$

where y represents the chemical content of the petiole, x_i is the spectral intensity of the i -th waveband, β_0 is the intercept, Z is the total number of wavebands, β_i is the regression coefficient for the i -th waveband, and ϵ is the model residual. The fresh dataset derived from the spectral analysis of fresh leaves, and the dried dataset obtained from the spectral analysis of dried leaves were prepared for data analysis and processing using the R statistical language (R Version 4.0.2; R Core Team, 2021).

2.5. RATIONALE FOR MODEL SELECTION

For the models described above, the number of predictors (wavelengths) was larger than the number of responses (chemical results). This is a common problem in dealing with NIRS data analysis, which results in a risk of overfitting, where an over specified model fails to filter out the noise from the training data, resulting in a poor generalization ability to new data. In such cases, subset selection methods, such as Lasso MLR, can be valuable. Lasso performs regularization by shrinking or setting some coefficients to zero, effectively selecting a subset of predictors that have the most explanatory power in relation to the target variable (Hastie et al., 2008). By reducing the number of predictors, Lasso helps to mitigate the risk of overfitting and improve prediction accuracy. The selection of the optimal complexity parameter, λ , is crucial in Lasso regularization. Larger values of λ impose greater penalization on non-zero coefficients, resulting in increased shrinkage and sparsity in the model. The process of selecting the appropriate λ involves finding the value that minimizes the root mean squared error (RMSE) (Hastie et al., 2008).

Efficient algorithms, as described by Hastie et al. (2008), are available to compute the entire path of solutions as the value of λ gets varied. This allows for a systematic evaluation of different models with varying levels of shrinkage. By performing successive

model training and performance assessment using iterative hold-out techniques like cross-validation, the optimal value of λ can be determined based on the smallest average RMSE, indicating the best trade-off between model complexity and predictive accuracy. Lasso regularization was used in this work as an effective technique for handling the challenge of the large number of predictors (262 wavelengths) compared to the 60 data points by effectively shrinking uninformative coefficients to zero and retaining only the most informative wavelengths in the model.

To implement Lasso regularization, the *glmnet* and *caret* packages in R were utilized. The training and validation of Lasso MLR models of each nutrient were carried out using a 5-fold cross-validation. To perform the cross-validation, the dataset was divided into five folds. Lasso MLR model was trained on four folds while utilizing the remaining fold as the validation set. This process was repeated five times, with each fold serving as the validation set once. By averaging the performance across the five folds, a robust evaluation of the model performance was obtained.

Initially, a Pearson's correlation (r) analysis was conducted to examine the relationship between the wavelengths and the content of each element. This analysis aimed to identify the wavelengths that exhibited the highest absolute correlation values, which could be considered as key wavelengths for the subsequent models. Later, the evaluation of the model performance was conducted by calculating the coefficient of determination (r^2) between the actual and estimated concentrations. To further assess the model performance, a Ratio of (standard error of) Prediction to (standard) Deviation (RPD) was calculated. This ratio, described by Williams (2019), is calculated using Equation (2.2).

$$\text{RPD} = \frac{\sqrt{\{(\sum A_n^2 - [\sum A_n]^2 / C) / (C-1)\}}}{\sqrt{\{(\sum (A_n - E_n)^2 - [\sum (A_n - E_n)]^2 / C) / (C-1)\}}}, \quad (2.2)$$

where C is the number of data points, A_n is the actual concentrations at n -th data point from 1 to C , E_n is the estimated concentrations at n -th data point from 1 to C , and n is the index of data point 1, ..., $C-1$, C . Following the classification suggested by Williams (2019), the RPD values for predicting functionality factors, such as grain texture, were categorized as excellent (> 4.1), very good ($\geq 3.5 - 4.0$), good ($\geq 3.0 - 3.4$), fair ($\geq 2.5 - 2.9$), and poor (< 2.0). In another study related to the monitoring of foliar nutrient status in mango using spectral indices, Mahajan et al. (2021) proposed a different classification for RPD, categorizing it as excellent (> 2), acceptable ($\geq 1.4 - 2.0$), and non-reliable (< 1.40). Considering that mango, like potato, is a horticultural crop (Saúco, 1997) and that the latter classification scheme was applicable, the classification suggested by Mahajan et al. (2021) was followed.

2.6. ESTIMATION OF MACRO AND MICRONUTRIENTS IN 2020

2.6.1. Spatial and temporal distribution of potato petioles chemical content

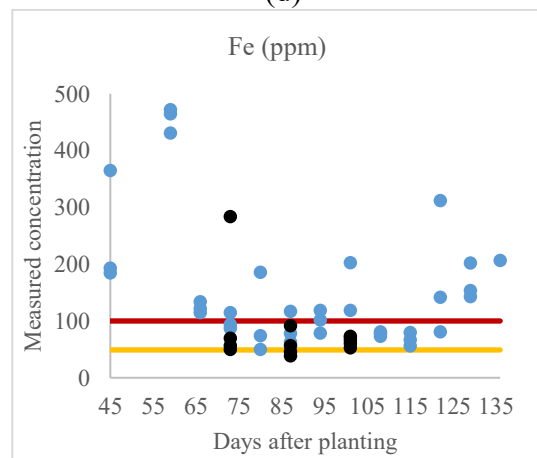
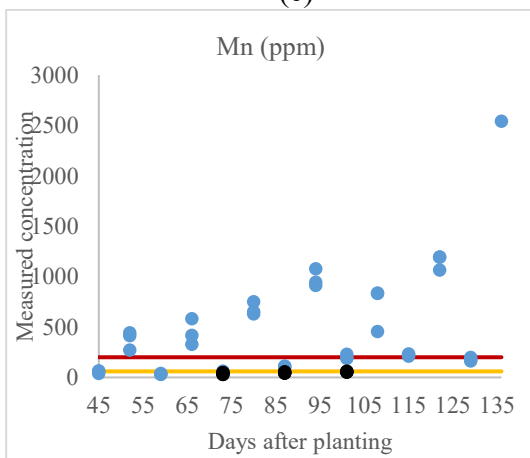
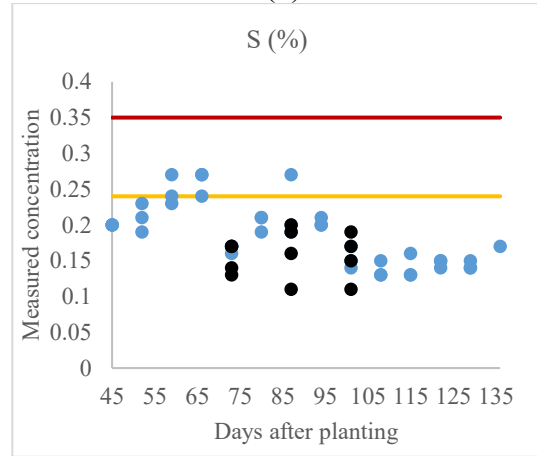
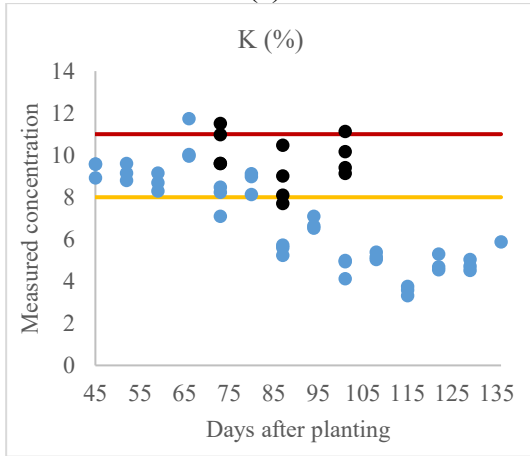
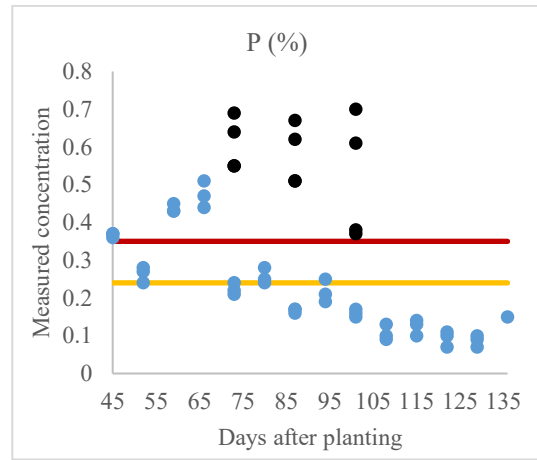
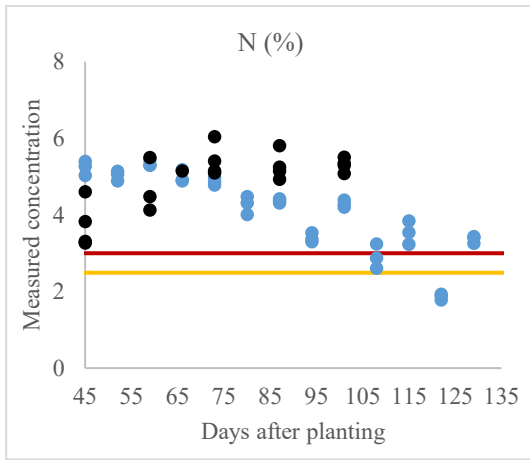
Figure 2.6 presents the results of the chemical testing conducted throughout the growing season, indicating the range of minimum and maximum values for each element compared to the recommended normal range of nutrients in potato petioles, as advised by A & L Canada Laboratories Inc in Ontario and the University of Minnesota (Kaiser and Rosen, 2018). The figure demonstrates the temporal concentration of each nutrient and its distribution within the normal range. The figure also depicts the spatial distribution of the measured concentrations, distinguishing between samples taken from the farm and from the indoor growing areas.

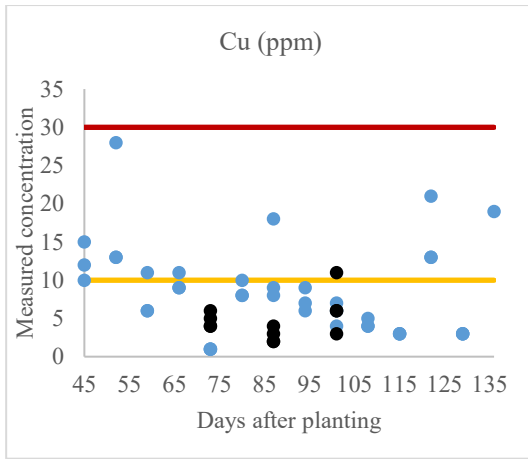
The high concentrations of NPK at the beginning of the season reflect the practice of applying fertilizers extensively during early stages to meet the plant nutrient requirements throughout the vegetative and reproductive phases, as well as to prevent deficiencies later in the season (Figure 2.6 (a, b, c)). The decrease in the NPK concentrations during the growing season can be attributed to dilution as plant biomass increases (Du et al., 2020; Gómez et al., 2020). The increased uptake of N and P in later stages has also been observed in previous studies (Liu et al., 2021a; Rosen et al., 2014), respectively. The slight increase in NPK concentrations towards the end of the season may indicate that the potato plant uptake of NPK has reached the maturity phase of the tubers, and the movement of NPK elements from foliage to underground tubers has ceased.

A similar trend is observed for S throughout the season, as its uptake is related to translocation within the plant, contributing to both yield and quality (Koch et al., 2020). The elevated S concentration can be attributed to the initial high K application in the form of K_2SO_4 , as shown in Figure 2.6 (d), in response to the concentrations in Figure 2.6 (c) (Koch et al., 2020). Mg is recognized as a competitive cation to K in terms of plant uptake (Koch et al., 2020). This implies that a low concentration of K in plant samples would lead to an increase in Mg concentrations, as observed towards the end of the season in Figure 2.6 (h). Similar findings were reported in a previous study that highlighted a cation antagonism between K and Mg, resulting in a significant decrease in Mg concentrations with high K supply in potato plants (Koch et al., 2020). The high values of Ca concentrations observed during the growing season in Figure 2.6 (i) can be attributed to the probable transport of Ca via the xylem rather than its transportation for tuber formation (Koch et al., 2020). Ca and Mg are well-known for their role in maintaining a stable pH in

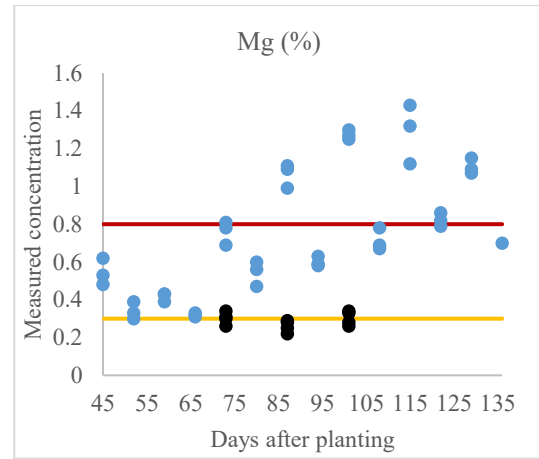
soil through the application of lime in the form of $\text{CaMg}(\text{CO}_3)_2$ (Government of New Brunswick, Department of Agriculture, Aquaculture and Fisheries, 2011).

The supply of Ca is commonly applied not only to neutralize soil pH but also to inhibit the uptake of Al and Mn, which can cause toxicity in potato plants. This could explain why Al concentrations in Figure 2.6 (j) fall within the maximum normal range for both farm and indoor data. Additionally, there are three instances where Al concentrations exceeded the normal range, which may indicate the unreliability of chemical testing during those specific sampling times. The application of Ca can also influence the uptake of Zn due to the decrease in soil acidity. A previous study stated that when lime is added to raise the pH, Zn becomes less available to potato plants (Koch et al., 2020). This can justify the decrease in Zn concentrations after 70 days from planting, as shown in Figure 2.6 (k), which coincides with the increase in Ca illustrated in Figure 2.6 (i). Moreover, low B concentrations are typically found in acidic soils (Ahmad et al., 2012). Therefore, the addition of lime in the form of Ca can enhance B concentration uptake, consistent with the chemical results shown in Figure 2.6 (i, l). The random distribution of Mn and Fe concentrations thereafter raises concerns about the reliability of the chemical testing (Figure 2.6 (e, f)).

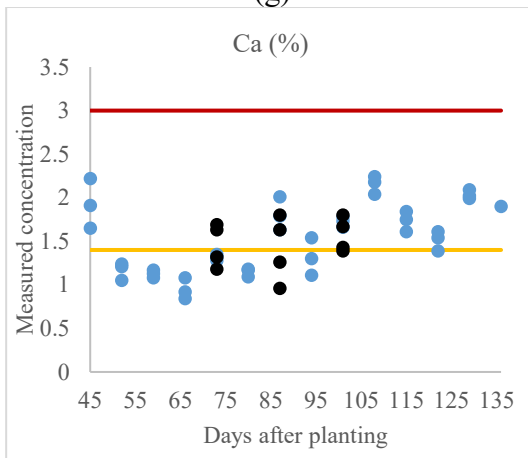




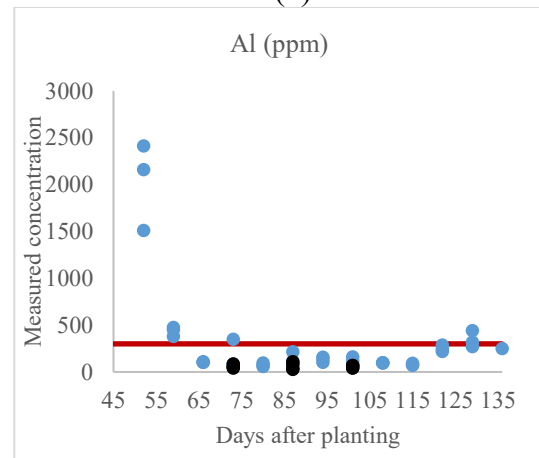
(g)



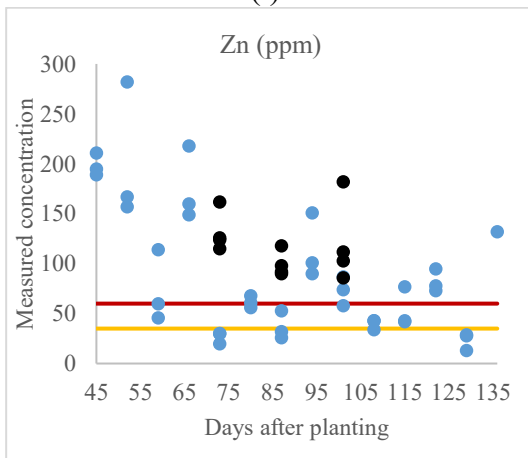
(h)



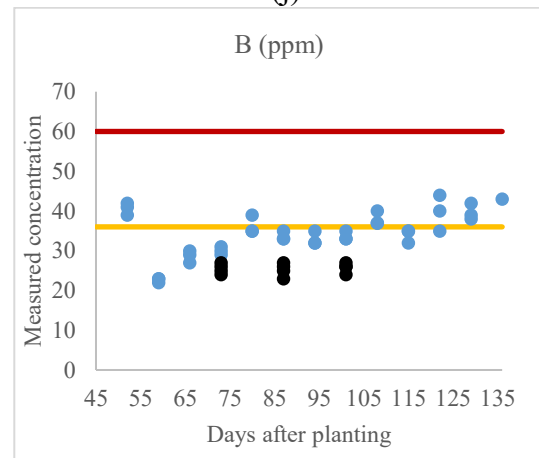
(i)



(j)



(k)



(l)

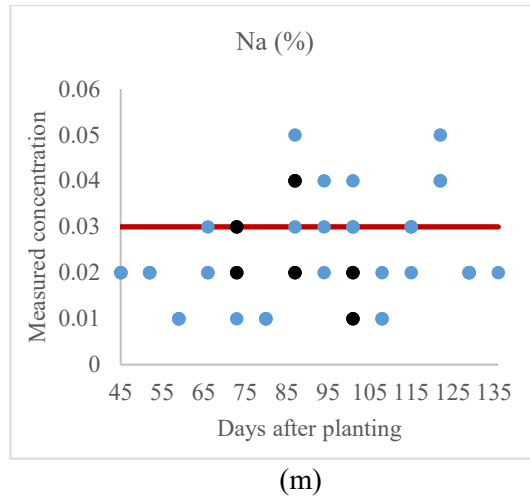


Figure 2.6: Temporal distribution for each nutrient during the growing season. Horizontal lines represent the limits of the maximum (red) and minimum (yellow) range of nutrients in potato petioles as recommended by A&L Canada Laboratories Inc in Ontario. Blue dots refer to data points of Season 2020, and black dots refer to the data points taken from indoor growing area

2.6.2. Determination of the significant wavebands

The absorption wavebands observed in the Vis-NIR range are commonly referred to as overtones, as they correspond to specific functional groups. In some cases, two absorbers may coincide to such an extent that an absorption waveband appears near the sum of the frequencies of the two fundamental wavebands. To address this, the recommendation of Williams (2019) was followed, and the first four significant wavebands of both modes are presented in Tables 2.2 and 2.3, and the corresponding Figure 2.7 represented by the vertical bars. The presented significant wavebands are ordered based on their significance score identified by the model regardless of their sequencing order.

Table 2.2: Number, range, and first four significant wavebands resulting from Lasso MLR modeling at the dried testing mode in Season 2020

Element	Number of wavebands	Range of wavebands (nm)	First four significant wavebands (nm)
N	10	524 - 1460	652, 708, 524, 1452
P	10	404 - 1924	708, 404, 540, 700
K	11	404 -1932	524, 700, 708, 652
Ca	20	404 - 2100	404, 444, 588, 540
Mg	14	404 - 1940	700, 532, 1716, 524
S	18	404 - 1916	404, 588, 516, 1452
Mn	22	428 - 2492	660, 628, 428, 492
Zn	12	468 - 2124	1932, 524, 1852, 532
Fe	19	404 - 2316	1932, 636, 524, 2308
B	11	412 -1932	684, 1932, 412, 460
Cu	23	428 - 2484	1940, 676, 428, 1716
Al	17	404 - 2316	1932, 2308, 524, 652
Na	2	724 - 732	724, 732, 980, 692

Table 2.3: Number, range, and first four significant wavebands resulting from Lasso MLR modeling at the fresh testing mode in Season 2020

Element	Number of wavebands	Range of wavebands (nm)	First four significant wavebands (nm)
N	13	404 - 1828	660, 684, 404, 484
P	10	404 - 948	404, 436, 588, 708
K	17	404 - 2300	404, 428, 588, 948
Ca	11	404 – 1924	876, 764, 548, 756
Mg	20	404 - 1932	668, 508, 404, 428
S	11	548 - 1204	588, 668, 732, 884
Mn	11	404 - 1700	404, 556, 740, 732
Zn	12	404 - 1100	596, 668, 540, 1100
Fe	24	404 - 1932	684, 660, 572, 1068
B	4	404 - 708	404, 708, 548, 588
Cu	17	428 - 1924	436, 596, 540, 700
Al	19	404 - 1708	684, 572, 916, 548
Na	20	548 - 2148	548, 972, 700, 1028

Figure 2.7 illustrates the outcomes derived from the highest positive r observed between the chemical contents of petioles and the reflectance measurements across a wavelength range of 404 nm to 2492 nm for both modes. The regions corresponding to the four most significant wavebands, as determined by Lasso MLR, are represented by vertical

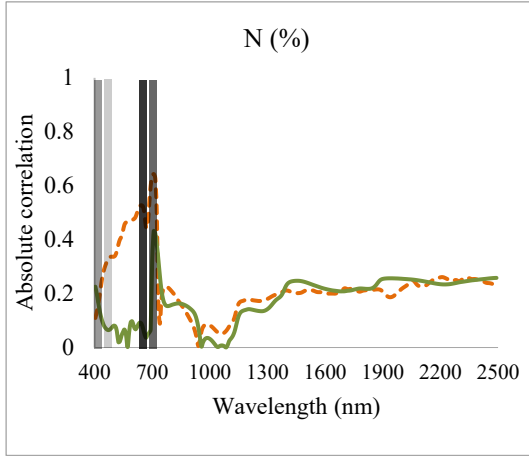
bars. These bars have a standardized width of 20 nm, which has been arbitrarily selected to evaluate the potential overlap of the most influential wavebands within the same range.

Figure 2.7 illustrates that among all macronutrients, S exhibited the highest r value in the fresh testing mode (Figure 2.7 (a-f)). A similar outcome was observed using Lasso MLR model, which yielded a closely matched r^2 value in both testing modes (Table 2.4). Additionally, P and K showed the highest r values in the dried testing mode (Figure 2.7 (b, c)), with Lasso MLR indicating similar r^2 values in both modes. In contrast to Pearson's correlation, Lasso MLR training model revealed that N had the highest r^2 values for the fresh testing mode (Figure 2.7 (a)).

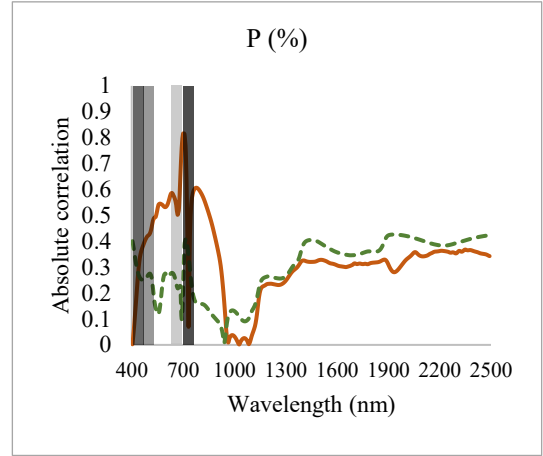
The most significant wavebands for N, P, and Ca were found in the Vis range, as depicted in Figure 2.7 (a, b, d), except for one waveband in the NIR range for K, Mg, and S (Figure 2.7 (c, e, f)). This observation may be attributed to the synergistic effect among these nutrients, as explained earlier in Section 2.6.1. The significant wavebands in the Vis range likely to contribute to the highest correlation observed in the fresh testing mode for N, K, P, and S, as they are less likely to interfere with water absorbance spectra in the NIR range (Prananto et al., 2020). Similar findings regarding significant wavebands were reported for P prediction in the VNIR region of the spectrum in corn canopy (Siedliska et al., 2021), as well as for S and K prediction in the Vis-VNIR range of the spectrum in mango leaves (Mahajan et al., 2021).

On the other hand, the dried testing mode revealed the highest r values for all micronutrients, consistent with the results obtained from Lasso MLR modeling, except for Fe, which exhibited a comparable r^2 value in both testing modes (Figure 2.7 (i) (Table 2.4). Among the micronutrients, Mn displayed four significant wavebands exclusively in the Vis

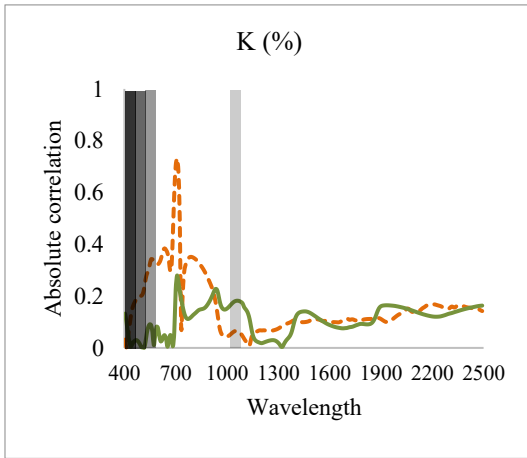
range (Figure 2.7 (g)), while B had a single significant waveband in the NIR region (Figure 2.7 (j)), despite its potential interference with Ca (Section 2.6.1). This unique waveband for B could potentially serve as a distinctive feature. Zn, Fe, Cu, and Al shared a similar pattern of having two significant adjacent wavelengths in the Vis and NIR ranges (Figure 2.7 (h, i, k, l)). Importantly, these NIR wavebands did not interfere with the significant wavebands associated with the macronutrients affecting them, such as K and Ca (Section 2.6.1). The correlation analysis of the spectral data highlighted the prominence of significant wavebands in the Vis-NIR region. This finding aligns with previous studies conducted by Osco et al. (2020) for predicting macro and micronutrient content in oranges and Ling et al. (2019) for detecting concentrations of leaf nutritional elements.



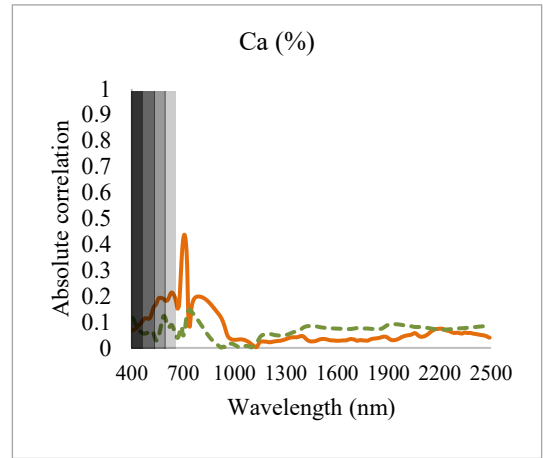
(a)



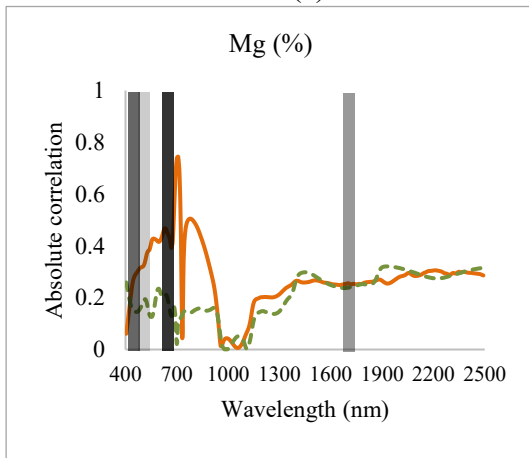
(b)



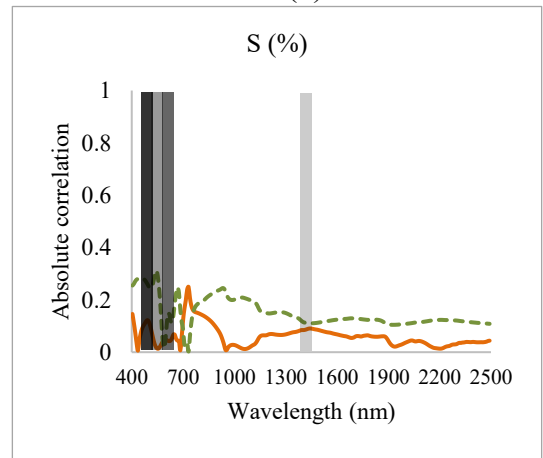
(c)



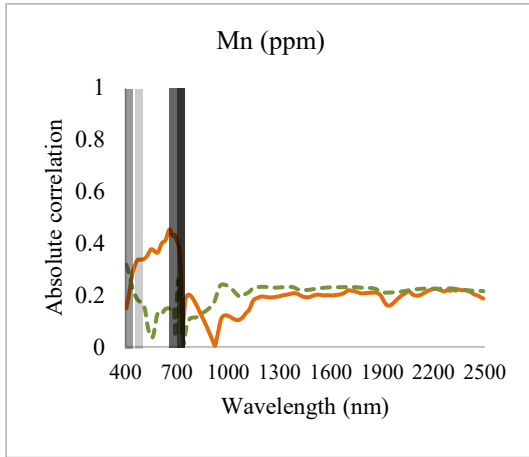
(d)



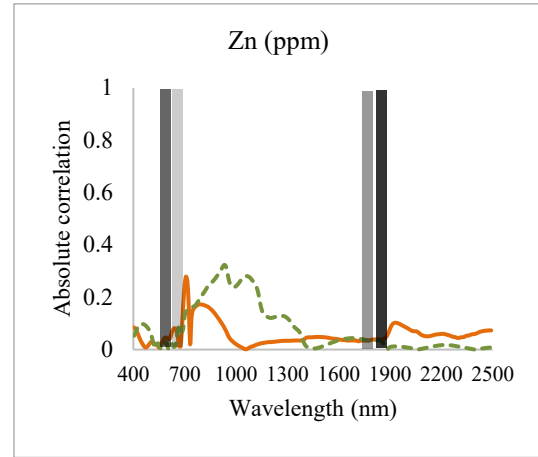
(e)



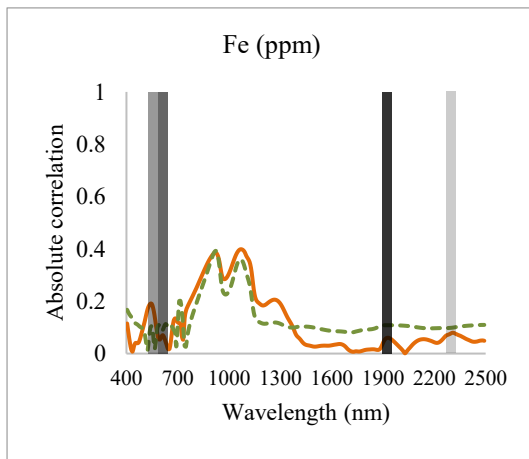
(f)



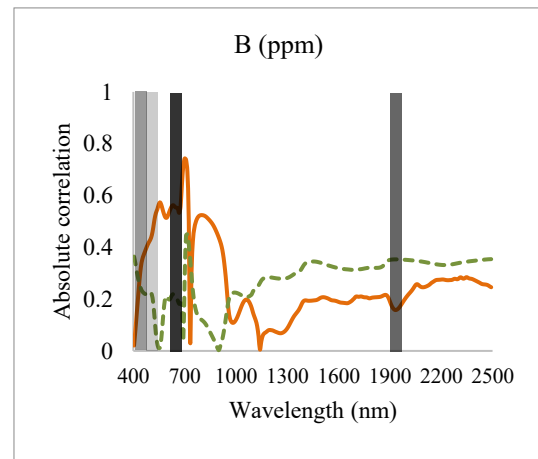
(g)



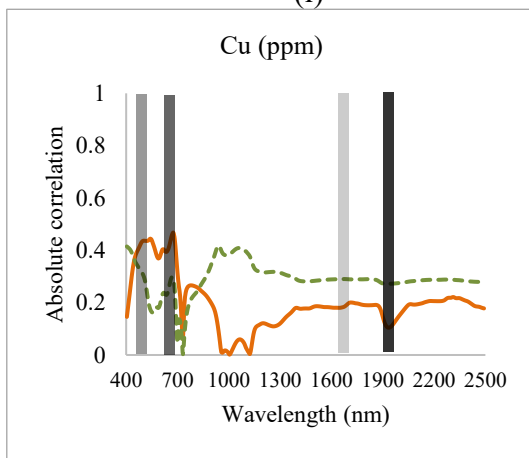
(h)



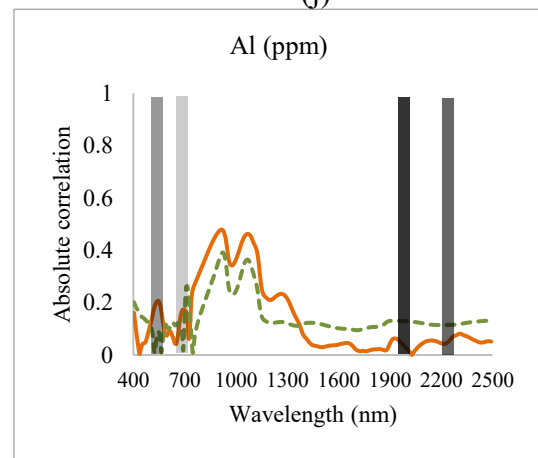
(i)



(j)



(k)



(l)

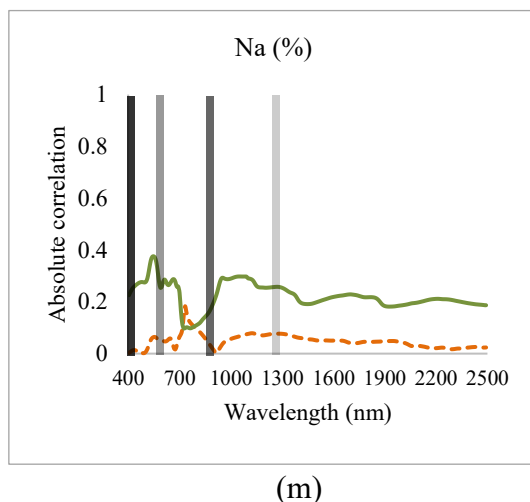


Figure 2.7: Pearson's correlation (r) for dried (brown) and fresh leaves (green) across the spectrum. Absolute peaks represent the highest r . The solid line outlines the testing mode of the highest r^2 using Lasso Regression, while dashed line represents the testing mode with less r^2 . The bars present the region of the four most significant wavebands by Lasso Regression. The intensity of the grey scale of the bar intensity gives the sequence of important wavebands from darker to brighter

2.6.3. Model performance measures in estimating nutrients

The evaluation of Lasso MLR models indicated a strong performance in estimating most macronutrients, as evidenced by the high RPD values reported in Table 2.4. The estimation of S exhibited a slightly lower value, yet still acceptable RPD value. These favorable RPD values indicate that the developed models captured a substantial portion of the variance observed in the chemical results.

For instance, even though most N concentrations (93%) exceeded the normal range of 2.49-3% as depicted in Figure 2.8 (a), the model still provided reasonable estimations. A similar pattern was observed in P estimation model shown in Figure 2.8 (b). Additionally, K and Mg estimation models displayed a satisfactory distribution of estimated concentrations around the fitting line, despite the concentrations falling outside the normal range (Figure 2.8 (c, e)).

Ca estimation model demonstrated a low correlation but a high RPD value of 2.55 (Table 2.4), suggesting a reasonable variance in the actual measurements shown in Figure 2.8 (d). On the other hand, S was the only macronutrient with concentrations consistently below the normal range. However, its model performance yielded reasonable results in terms of r^2 and RPD values compared to other macronutrients (Table 2.4). It is worth considering that Ca and S models might benefit from the inclusion of more diverse chemical concentrations in the dataset as shown in Figure 2.6 (i, d).

For the estimation of micronutrients, the RPD evaluation indicated acceptable to excellent performances, as illustrated in Table 2.4. Lasso MLR models demonstrated significantly high correlation values, except for Mn. However, Mn estimation model exhibited an excellent RPD value in Table 2.4, suggesting a good estimation performance despite the variability in chemical concentrations depicted in Figure 2.6 (e).

It is important to note that the estimation of Al was less reliable at the beginning of the season, as explained in Section 2.6.2 and depicted in Figure 2.6 (j). The presence of unreliable Al concentrations during training may have introduced some bias to the model. In Figure 2.8 (l), Al model displayed favorable estimation results for concentrations below 300 ppm, while three data points beyond 1000 ppm were underestimated. To address this, an additional analysis was conducted on Al dataset after shuffling the data points to eliminate the order of temporal effects of concentrations. Following the shuffling as a pre-processing step, Lasso testing model exhibited an improvement in the r^2 value from 0.67 to 0.71, while the RPD value slightly decreased from 2.61 to 2.39. Furthermore, considering the three data points shown at day 52 in Figure 2.6 (j) as concentrations outside the accepted range, the r^2 value of the testing model marginally decreased from 0.67 to

0.62, while the RPD value increased from 2.61 to 3.78. Thus, further investigation is required for AI modeling, with larger datasets, to enhance the reliability and performance of the model.

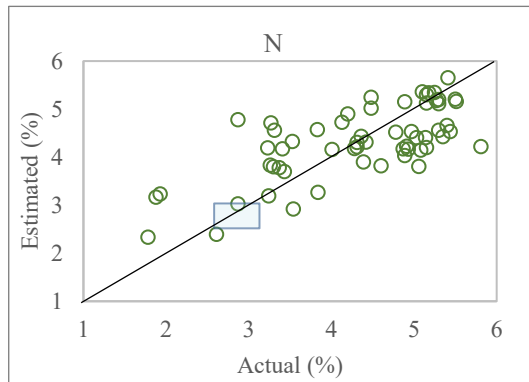
Similarly, Figure 2.6 (f) and Figure 2.8 (i) indicated three inconsistent Fe concentration values compared to the others throughout the growing season. However, the cross-validation results of Lasso model still yielded a high r^2 value of 0.65 and an excellent RPD value of 3.66 (Table 2.4). This high RPD value suggests that the model estimation error is relatively small when compared to the actual values. Nonetheless, due to the lower confidence in these three Fe concentrations compared to the rest of the dataset, re-evaluating the model performance by excluding these values was decided. The results showed a decrease in the r^2 value from 0.65 to 0.40, while the RPD value remained relatively unchanged at 4.24 compared to 4.23. Although the r^2 value decreased, the range of actual Fe concentrations used to train the model became more reliable without the inclusion of those inconsistent values. On the other hand, the estimation of other micronutrients showed good alignment with the fitting line (Figure 2.8 (g, h, j, k)), and the model performance, as measured by RPD values, demonstrated acceptable to excellent accuracy for Zn, Cu, Mn, and B, respectively (Table 2.4).

Table 2.4: The validation results of Lasso MLR models estimating elements at the best testing mode

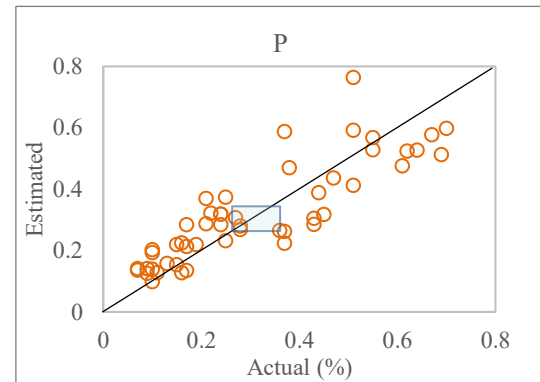
Element	Unit	Testing mode used for modeling	Validation results	
			r^2	RPD
N	%	Fresh	0.59	3.06
P	%	Dried	0.74*	2.26
K	%	Fresh	0.75**	2.44
Ca	%	Dried	0.32	2.55
Mg	%	Dried	0.77	2.85
S	%	Dried	0.50*	1.82
Mn	ppm	Dried	0.24	2.30
Zn	ppm	Dried	0.54	2.26
Fe	ppm	Dried	0.65*	3.66
B	ppm	Dried	0.62	2.08
Cu	ppm	Dried	0.58	2.09
Al	ppm	Dried	0.67	2.61
Na	%	Fresh	0.19	4.36

* r^2 values are less than 0.04 compared to values of testing mode for fresh removed leaves.

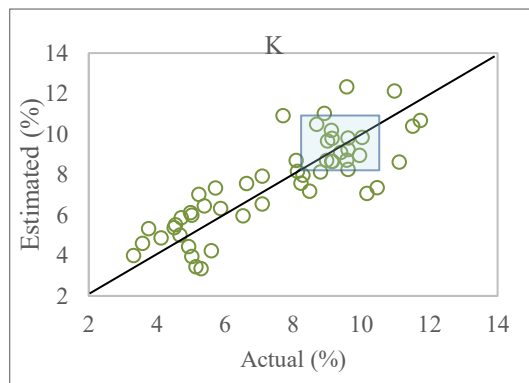
** r^2 values are less than 0.02 compared to values of testing mode for dried ground leaves.



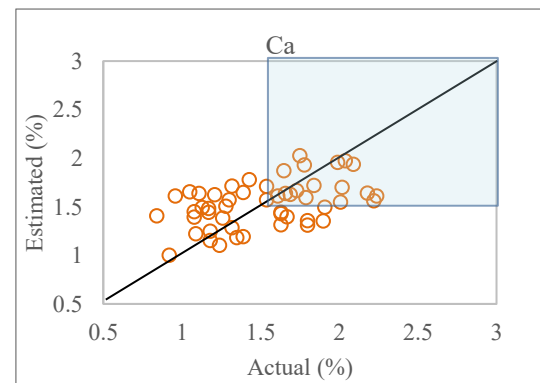
(a)



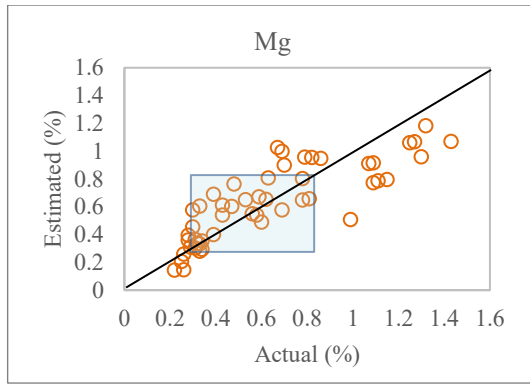
(b)



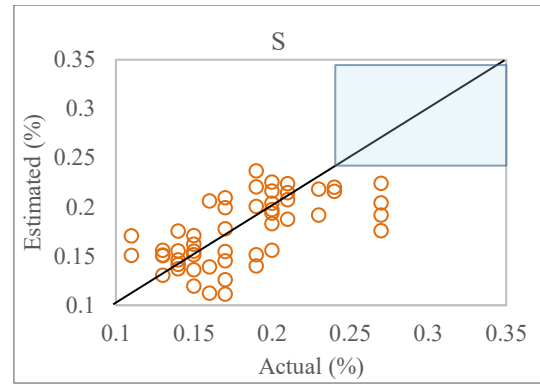
(c)



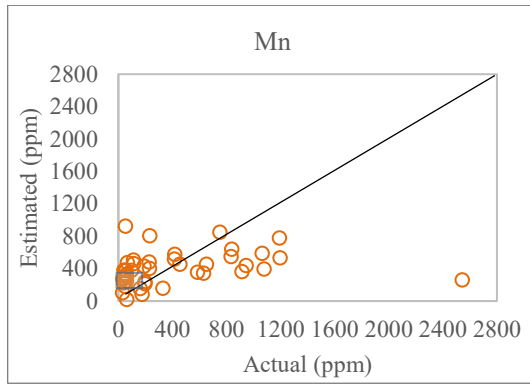
(d)



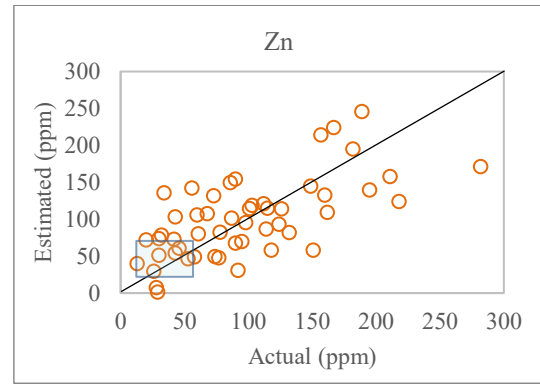
(e)



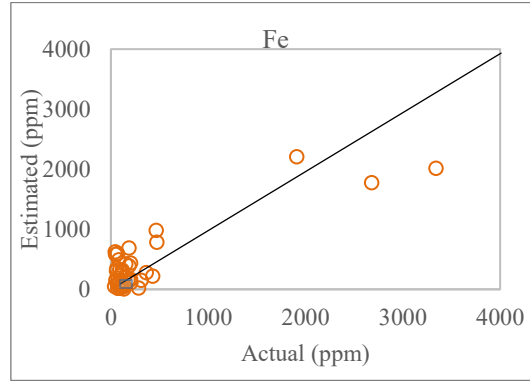
(f)



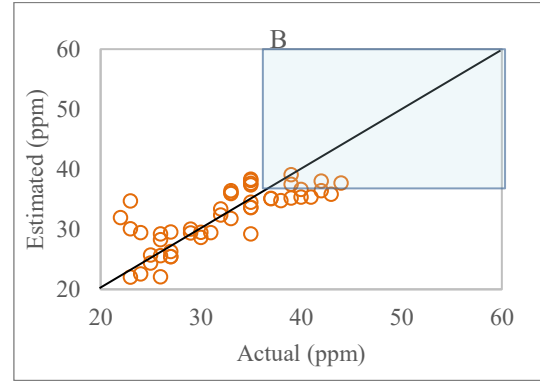
(g)



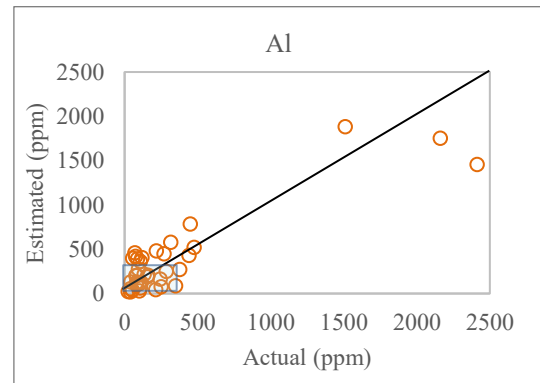
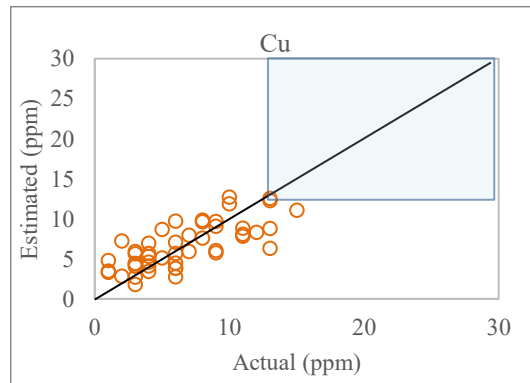
(h)



(i)



(j)



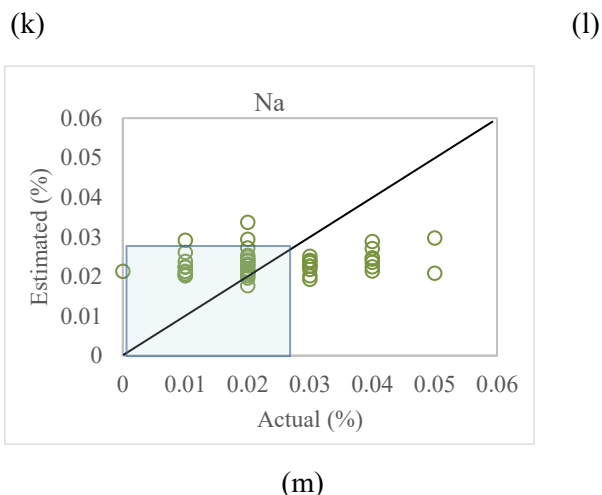


Figure 2.8: The validation results of the actual versus estimated concentrations of the testing mode (green for fresh, brown for dried) of the highest r^2 . Blue box shows the normal range of nutrients in potato petioles as recommended by A&L Canada Laboratories

Na exhibited a distinct pattern compared to other micronutrients. Throughout the growing season, Na concentrations displayed a discrete distribution with a dominant concentration observed at 0.02% (Figure 2.6 (m)). It is worth noting that Na analysis is conducted in percentage units rather than ppm, unlike the other micronutrients. This choice of units may not effectively capture changes in Na concentrations occurring within the season. Despite this discrete distribution, Pearson's correlation analysis did not reveal any unusual patterns when compared to other elements (Figure 2.7 (m)). However, Lasso MLR model consistently provided an estimation value of 0.02% for Na (Figure 2.8 (m)), disregarding the actual values which reached up to 0.05% in Figure 2.6 (m). Consequently, this estimation model yielded a low r^2 value of 0.19 (Table 2.4). It is important to note that the RPD value of 4.36 (Table 2.4) suggests a low variance in the actual data, making it difficult to accurately evaluate the efficiency of the estimation capacity (Parrini et al., 2021). It remains uncertain whether the poor model performance is due to unreliable Na

chemical concentrations or if there is indeed no correlation between petiole Na content and leaf spectrum. Therefore, Na was excluded from further analysis in this work.

2.7. ESTIMATION OF MACRO AND MICRONUTRIENTS IN 2021

2.7.1. Spatial and temporal distribution of petiole chemical content

An additional 144 data points was collected in Season 2021 following the same sample collection, handling, and analysis methods presented in Sections 2.2 and 2.3, focusing on the Russet Burbank variety. Sample collection was expanded to cover seven open farms in Florenceville, Lakeville in NB (Figure 2.1). The petiole chemical content is presented in Table 2.5. The noticeable variations in the SD values of Mn, Fe, and Al concentrations may refer to the reliability confidence in the ground truth values of the chemical testing. Another reason of such variations could refer to the heterogeneous distribution that indeed existed amongst the samples taken from the same subplot. The latter reason is less likely to be the reason, as the sampling was consistent within the same subplot during the entire growing season. Therefore, those variations might later impact the confidence level of the estimation models, as it is hard to judge on the efficiency of the low estimation capacity whether it is a reason for non-reliable chemical concentrations or there is in fact no correlation between petiole content and leaf spectrum.

Table 2.5: Descriptive content of the petiole nutritional concentrations during the entire growth season in 2021

Macro-nutrients	N	P	K	Ca	Mg	S
Min. measured	1.59	0.07	2.05	0.93	0.22	0.12
Max. measured	5.5	0.47	12.26	3.59	2.08	0.29
Min. recommended*	2.49	0.24	8	1.4	0.3	0.24
Max. recommended*	3	0.35	11	3	0.8	0.35
Mean	3.59	0.19	6.63	1.87	0.90	0.20
SD	0.90	0.08	2.39	0.61	0.42	0.04
Micro-nutrients	Mn	Zn	Fe	Cu	Al	B
Min. measured	102.00	11.00	39.00	1.00	1.00	13.00
Max. measured	1400.00	105.00	4307.00	52.00	1241.00	228.00
Min. recommended*	60	35	50	10	ND	36
Max. recommended*	200	60	100	30	300	60
Mean	492.32	43.03	246.43	7.50	200.40	50.65
SD	262.06	20.94	401.20	6.08	225.92	39.80

* Normal range in nutrient concentrations stated by A&L Canada Laboratories Inc. ND: not defined.

2.7.2. Determination of the significant wavebands

The initial analysis by Lasso MLR highlighted the importance of around 100 significant wavebands distributed within the Vis and NIR ranges for all nutrients. An assumption was made that 300 data points for a linear regression model with at most 100 features plus the intercept term shall provide a reasonable number of residual degrees-of-freedom in the model helping in mitigating the risk of overfitting. Lasso models were trained and tested using the data points collected from Season 2021 and Tables (2.6, 2.7) show the range of the significant wavebands to estimate each nutrient along with the four significant wavebands of each mode presented based on their significance score. Table 2.8 shows the validation results in estimating nutrients of the mode that had the highest correlation value.

Table 2.6: Number, range, and first four significant wavebands resulting from Lasso MLR modeling at the dried testing mode in Season 2021

Element	Number of wavebands	Range of wavebands (nm)	First four significant wavebands (nm)	
			Season 2020 ²	Season 2021
N	17	492-2348	652, 708, 524, 1452	2052, 1444, 2308, 2332
P	22	492-2308	708, 404, 540, 700	2052, 1524, 2300, 1444
K	18	516-1828	524, 700, 708, 652	1684, 1828, 1820, 1428
Ca	25	404-2484	404, 444, 588, 540	1820, 1724, 1892, 1684
Mg	19	508-2212	700, 532, 1716, 524	2212, 1676, 1324, 1684
S	16	404-2484	404, 588, 516, 1452	2052, 2316, 700, 1444
Mn	33	420-2292	660, 628, 428, 492	596, 1884, 492, 732
Zn	26	404-2220	1932, 524, 1852, 532	1420, 1516, 540, 404
Fe	5	404-1196	1932, 636, 524, 2308	804, 404, 812, 1196
B	14	508-2172	684, 1932, 412, 460	1652, 2172, 700, 724
Cu	25	404-2300	1940, 676, 428, 1716	2212, 2300, 652, 484
Al	13	412-2188	1932, 2308, 524, 652	692, 716, 668, 1380

Table 2.7: Number, range, and first four significant wavebands resulting from Lasso MLR modeling at the fresh testing mode in Season 2021

Element	Number of wavebands	Range of wavebands (nm)	First four significant wavebands (nm)	
			Season 2020 ³	Season 2021
N	23	404 - 2492	660, 684, 404, 484	1444, 532, 436, 572
P	17	404 - 2492	404, 436, 588, 708	404, 652, 708, 620
K	24	404 - 1972	404, 428, 588, 948	620, 404, 932, 780
Ca	18	404 - 2276	876, 764, 548, 756	940, 612, 772, 1668
Mg	25	404 - 2484	668, 508, 404, 428	404, 764, 668, 916
S	17	452 - 1932	588, 668, 732, 884	700, 684, 1436, 724
Mn	30	404 - 2460	404, 556, 740, 732	612, 2452, 1908, 2460
Zn	23	404 - 2492	596, 668, 540, 1100	572, 404, 1884, 2492
Fe	6	476 - 1892	684, 660, 572, 1068	700, 1092, 1108, 476
B	27	444 - 2484	404, 708, 548, 588	668, 684, 1476, 644
Cu	34	404 - 2492	436, 596, 540, 700	2284, 1956, 2492, 1092
Al	29	404 - 2492	684, 572, 916, 548	1444, 2180, 468, 1700

² Correspond to Table 2.2.

³ Correspond to Table 2.3

Table 2.8: The validation results of Lasso MLR models estimating in Season 2021 at the best testing mode

Element	Unit	Testing mode used for modeling	Validation results	
			r^2	RPD
N	%	Dried	0.75*	1.96
P	%	Dried	0.75*	2.05
K	%	Dried	0.63*	1.68
Ca	%	Dried	0.49	1.4
Mg	%	Dried	0.74	1.96
S	%	Dried	0.64	1.69
Mn	ppm	Dried	0.42	1.29
Zn	ppm	Dried	0.37*	1.3
Fe	ppm	Dried	0.24	1.14
B	ppm	Dried	0.67	1.81
Cu	ppm	Dried	0.24	1.13
Al	ppm	Dried	0.54	1.5

* r^2 values are less than 0.05 compared to values of another testing mode.

One notable observation from training models on the data from two separate seasons is the shift in the range of most significant wavelengths from the Vis range to the NIR range. This phenomenon can be explained by considering both chemical and modeling viewpoints. A study, which assessed the nutrient content of cotton leaves using wavelength range between 1300 nm – 2500 nm (Prananto et al., 2021), concluded that predicting N corresponded not only to the presence of chlorophyll compounds that fall within the Vis spectrum, but also to protein compounds where their significant wavebands hover around 2054 nm and 2712 nm. Their findings correspond to the results of this research in estimating N in 2020, and 2021. This chemical perspective offers a rationale for explaining this phenomenon. From the modeling perspective, those results suggest that the model functionality extends across a broad range of wavebands which could be linked to how the chosen model (Lasso) works. For instance, Lasso model can identify the most significant waveband from a pair of highly correlated wavebands, selects it and removes the others. Another example is given where the significant wavebands to assess P were found within

the range of 2390 nm - 2400 nm as those spectral ranges are absorbed by phytates, nucleic acids, phosphoproteins, phospholipids where P mostly exists (Prananto et al., 2021). This may explain the results of P in Season 2021 (Table 2.6). However, the significant wavebands in Season 2020 were basically locating within the Vis range, which could be attributed to the phenomenon of absorptive wavebands and their overtones bouncing over a combination of wavebands (Williams, 2019) (Figure 2.9).

Those overtones are broadened as compared to their counterparts found in the mid-IR region and each bounce produces some absorption of the beam at several bounces before reaching the detector. Those bouncing wavebands at various positions of the spectrum may have the same significance to estimate a specific nutrient at a narrower or a boarder range, which may justify the reason of shifting the significant wavebands from Vis to NIR. However, under this level of research, no intensive analysis was done to identify the major absorption wavebands and their bounces of each nutrient.

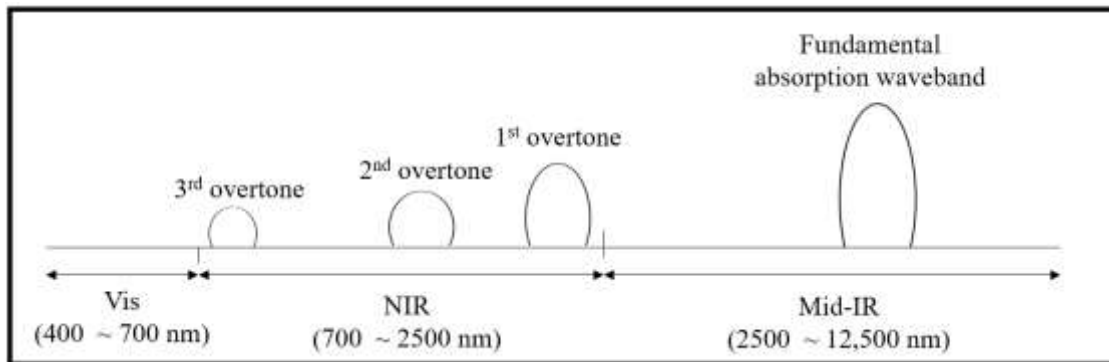


Figure 2.9: A descriptive diagram showing the positioning of the fundamental absorption waveband in the mid infrared region of the spectrum (mid-IR), and the positions of the bounced overtones within the NIR range

2.8. ESTIMATION OF P AT MULTIPLE SEASONS: A CASE STUDY

A more detailed study was performed on estimating P across seasons, as the performance of a generalizable model may be influenced by samples variability across

different seasons. For instance, Botha et al. (2006) found the accuracy of a model of one season to estimate petiole nitrate concentration using dried leaf chlorophyll content in different potato cultivars drops when used for samples of another season. The study referred to the reason for sample variability induced by prolonged drought conditions in that specific year of the study. Similarly, Liu et al. (2021a) were able to predict foliar N by correlating data of petiole nitrate and hyperspectral images of different potato cultivars and planting seasons. Their study executed over the dried plant tissue concluded that generalizable models are possible if training measurements from all developmental stages of interest and for multiple growing seasons include variable weather conditions within model development. Another reason for choosing P for a deeper analysis is that P showed a high consistency in the chemical results from the same subplot over the two seasons. For that purpose, a detailed study was implemented to compare the performance of three datasets in estimating P: (i) Dataset 1: develop a model based solely on data from Season 2020, (ii) Dataset 2: train model using Season 2020 data and validate its performance on Season 2021, and (iii) Dataset 1+2: create a combined model using data from both seasons. In addition to Lasso, PLSR was also used to predict the leaf P content using the spectral data in single and across seasons. The generated dataset between P content and leaf spectra of both fresh and dried leaves was split into 80% training and 20% testing.

The models were trained using 5-fold cross-validation, with the RMSE, and normalized RMSE (nRMSE)⁴ used as the error statistic. Another parameter used for assessing the model performance is the Ratio of Performance to Interquartile Distance

⁴ Calculated in reference to RMSE divided by the maximum value of observations minus the minimum value of observations.

(RPIQ) which is defined as the interquartile range of the reference values divided by the root mean square error (Bellon-Maurel et al., 2010). The RPIQ takes both the prediction error and the variation of reference values into account, in contrast to the residual prediction deviation (RPD). The RPD assumes a normal distribution of the reference values which may lead to erroneous conclusions, in contrast to RPIQ where no distributional assumption is made (Bellon-Maurel et al., 2010). In their study, they concluded the greater the RPIQ, the better the model predictive capacity is to be. A previous study classified the regression model performance based on the RPIQ values as very poor (<1.5), poor ($1.5-2.0$), good ($>2.0-2.5$), and very good (>2.5) (Mahajan et al., 2021).

Tables 2.10 and 2.11 show the performance parameters of two linear regression models, PLSR and Lasso of both dried and fresh modes. The cross-validation performance of Datasets 1 and 2 exhibited good r^2 values between 0.532 and 0.780 (Tables 2.9, 2.10). The difference in the model performance of each season may be attributed to the number of data points being used for model training. In Dataset 1, 80% of data from Season 2020 were taken for training the model, however, Dataset 2 trained the model using all data points of Season 2020. With this larger number of data points available for training, the model may be better able to capture the underlying patterns in the data. Generally, Lasso outperformed the PLSR substantially (r^2 : 0.695 vs. 0.588 of Dataset 1 of dried mode, r^2 : 0.404 vs. 0.171 of Dataset 2 of fresh mode) (Tables 2.9 and 2.10), which could be due to the way Lasso models select the relevant wavebands. The PLSR builds a reduced set of latent variables to avoid the multicollinearity through a principal component transformation. In contrast, Lasso shrinks the 262 correlation coefficients to a smaller number (~ 20 of each dataset) to determine the concentration of P.

The selection of specific wavebands by Lasso, particularly under the fresh mode, seemed to be significant as the full range of spectrum could trigger the non-linear responses of a leaf spectrum due to the water absorption signals in the NIR range (Sheng et al., 2022). Similar results are shown by Dyar et al. (2012) who concluded that Lasso directly determined which wavelength channels can best represent the elemental concentrations, better than PLSR models. RMSE (0.056 for Lasso vs. 0.070 for PLSR on Dataset 2 – see Table 2.9) and the estimated RPIQ (>2 in all cases) of single-season models indicate good to very good predictions for dried and fresh leaves. Moreover, the r^2 obtained by Lasso models for the testing datasets showed performance comparable with the cross-validation estimation, suggesting that the models did not overfit the training data.

When the trained model used to test data points from the same season in 2020, both PLSR, and Lasso exhibited a good performance of both modes (r^2 : 0.695 dried mode, r^2 : 0.684 of fresh mode) (Tables 2.9, 2.10). When the model of Season 2020 was used to validate data points taken from Season 2021, the model performance exhibited poorly at r^2 0.437 in the dried mode, and 0.404 in the fresh mode. This poor estimation could refer to the fact that the differences in spectral measurements may occur across seasons due to different factors associated with plant diseases, deficiency of other nutrients, and physiological stresses (Muñoz-Huerta et al., 2013; Liu et al., 2021a). If these differences were not included within training, the model may not recognise some of the patterns when tasked with generating predictions for new data.

For across seasons analysis, both linear models produced a slightly lower performance when data from the two seasons were pooled together. For instance, the r^2 of Lasso model was 0.695 for Season 2020 in comparison to 0.668 for the combined seasons

of dried mode (Table 2.9, 2.10). However, models using the combined season data presented very good performance measures of RPIQ. The outperformance of Lasso is likely referred to the same reason suggested in the single-season models, namely the ability of Lasso to select wavebands that are specifically relevant to estimate the output of P.

Although, the cross-season models performed less than the single-season models the performance measures of r^2 , RMSE, nRMSE, and RPIQ indicate the possibility to generate a model that would work at different growing seasons. That means by combining data points across seasons, the model might be provided with a broader range of possible relationships between waveband intensity and nutrient concentration, allowing it to learn these patterns and potentially enhancing its generalisation ability.

Table 2.9: Performance of regression models for dried leaves

	Dataset 1		Dataset 2		Dataset 1+2	
	PLSR	Lasso	PLSR	Lasso	PLSR	Lasso
Cross-validation						
RMSE (*)	0.081	0.114	0.087	0.097	0.066	0.053
nRMSE	0.184	0.259	0.138	0.153	0.060	0.034
r^2 (*)	0.649	0.592	0.780	0.735	0.530	0.696
Testing						
RMSE	0.057	0.054	0.170	0.109	0.069	0.058
nRMSE	0.204	0.191	0.270	0.172	0.178	0.149
r^2	0.588	0.695⁵	0.508	0.437	0.543	0.668
RPIQ	2.844	3.048	1.960	3.280	2.854	3.415

* Average values of the 5-fold cross-validation

- Bold font refers to the model that has the highest r^2 value

⁵ Correspond to Table 2.4 where model gave an r^2 value of 0.745 in estimating P. The difference between these two values (0.695 instead of 0.745) referred to that this analysis of Season 2020 excluded the data points of the indoor growing alike to what was reported in Table 2.4. Though, the indoor data points of P expanded the range of P concentrations which gave a better distribution of model estimations at 60 data points in total.

Table 2.10: Performance of regression models for fresh leaves

	Dataset 1 (Season 2020)		Dataset 2 (Season 2021)		Dataset 1+2 (Seasons combined)	
	PLSR	Lasso	PLSR	Lasso	PLSR	Lasso
Cross-validation						
RMSE (*)	0.109	0.066	0.131	0.112	0.072	0.058
nRMSE	0.248	0.151	0.209	0.177	0.164	0.131
r^2 (*)	0.532	0.758	0.665	0.684	0.486	0.646
Testing						
RMSE	0.077	0.065	0.192	0.180	0.065	0.060
nRMSE	0.274	0.232	0.304	0.234	0.167	0.153
r^2	0.569	0.684	0.171	0.404	0.528	0.608
RPIQ	2.124	2.507	1.444	1.582	3.248	3.560

* Average values of the 5-fold cross-validation

- Bold font refers to the model that has the highest r^2 value

Regarding the position of the significant wavebands, both models imposed similar positions of the most important wavelengths within the Vis range of the spectrum (400 - 700 nm) (Figure 2.10 (a, b)). Additional significant wavebands were selected within the SWIR by Lasso, which may highlight the importance of this range of spectrum to estimate P. The dried mode in particular exhibited relevant wavelength regions within 1350 - 1480, and 1780 - 2032 nm that were not masked by water features, which are known as water absorbance ranges. Wavelengths in the SWIR region are predominantly associated with light absorption by proteins, N, cellulose, starch, and sugar, where P plays an important role in protein synthesis (Naumann et al., 2020). Thus, neglecting the water content in the leaf by taking the spectrum over the dried mode can be important to diagnose P status in the plant. Findings from earlier also suggested that the SWIR region is useful for distinguishing levels of P in plant tissues studies (Siedliska et al., 2021; Knox et al., 2012; Prananto et al., 2021).

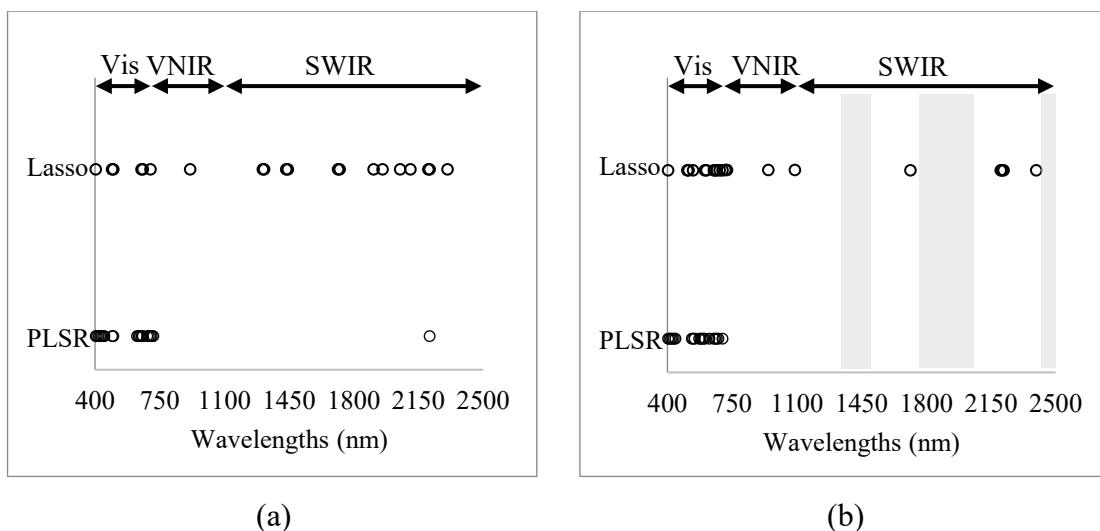


Figure 2.10: Position of the wavebands selected by PLSR and Lasso to estimate P across seasons: (a) dried mode, (b) fresh mode. The shaded bars refer to the water absorbance wavelengths

2.9. CONCLUSION

The results presented under this chapter showed that there is indeed a correlation between the chemical contents of potato petioles and the leaf spectrum for 12 out of the 13 nutrients, with Na being the only exception. The results further showed that the correlation is higher in the dried mode of leaves for most of the nutrients. Still, there is a comparable correlation to estimate N, P, K, S, and Fe in the fresh mode. The results of regression models over a different season in 2021 under different growing and climatic conditions showed a high potential to estimate all macro and micronutrients based on the RPD values, except Mn, Zn, Fe, and Cu.

Furthermore, the enhanced predictive performance observed in Season 2021 compared to 2020 for most nutrients signifies that the models learned more about data patterns using a large data size. In Season 2021, the results also showed that the predictive power was higher in the dried mode, but also N, P, K showed a high potential to be estimated based on fresh spectra matching the results in 2020. Moreover, the results of P

after compelling the data from two consecutive seasons showed that the estimation of nutrients is still well performed based on the RPIQ values using two linear models, PLSR, and Lasso.

CHAPTER 3: CALIBRATING LAB AND FIELD SPECTRAL INSTRUMENTS FOR NUTRIENT WHEN READING REFLECTANCE OF POTATO PLANTS

3.1. BACKGROUND

The spectral analysis shown in Section 2.3 underwent through laboratory-based spectrophotometer. Since, one of the objectives of this research is to enable the estimation of nutrients in field, field-based spectral scanning was necessary. Although laboratory and field spectral instruments are based on the same technology, calibration needs to be developed to minimize biases and noise that can emerge from using different spectral instruments. The importance of evaluating spectral data obtained from various instruments before assessing the capabilities of data modeling was emphasized in a study conducted by Krinitskiy et al. (2021). The ability to calibrate the data of different instruments allowed to maximize the dataset by combining data points created by two different instruments.

3.2. DATA COLLECTION OF LAB AND FIELD REFLECTANCE SPECTRA

3.2.1. Samples acquisition of laboratory and field-based spectrophotometers

The reference spectral measurements in this research were obtained using an NIRS analyzer (DS2500, Metrohm USA Inc.) for both fresh and dried leaves in the laboratory setting. In contrast, the field spectral measurements were collected using a portable spectrophotometer (FieldSpec4 Standard-Res., Malvern Panalytical Technologies, Malvern, Worcestershire, UK). For fresh leaves, a leaf clip was used to capture the reflectance spectrum, while dried leaves were scanned on a turntable after grinding. The operating specifications of the two spectrophotometers are presented in Table 3.1.

In Season 2021, a total of 45 data points were collected from the Russet Burbank potato variety, obtained from six subplots across four potato farms in Florenceville-Bristol,

NB, Canada. Each subplot contributed to six or nine data points, depending on the initial data collection date. Within each data point, observations were taken from 60 randomly selected leaves. These leaves were divided into two groups: Group 1 for laboratory testing, consisting of 40 leaves (20 fresh and 20 dried), and Group 2, which included 20 fresh leaves initially scanned in the field and then dried for subsequent dried spectral scan.

Group 1 leaves were vacuum-packed immediately after removing the petioles, placed in sampling bags, and refrigerated before shipment following the same protocol discussed in Section 2.2.1. Each sample was packed with an ice bag and shipped to DairyLand Lab Inc. (Arcadia, Wisconsin, USA) for fresh and dried spectral analysis using the NIRS analyzer. The dried leaves were placed in a paper envelope and then in an oven at 55-60 °C for 16-24 hours. The spectral observations of both dried and fresh leaves were taken in a black cup to minimize the impact of stray light. The leaves were symmetrically trimmed to fit the cup size of the spectrophotometer for consistent measurements.

Group 2 leaves were detached from the petioles for the in-field spectral analysis using the leaf clip of the Vis-NIR spectrophotometer. A black background was chosen for the leaf clip as it provided better modeling accuracy compared to a white background, as supported by Wang et al. (2017). A white reference measurement was taken every ten readings to ensure normalized in-field spectral measurements. After measuring the spectrum of fresh leaves, they were dried, ground into a 1 mm powder, and then scanned on the turntable. The ground powder was placed in a glass Petri dish with a white reference panel (Spectralon panel, LabSphere Inc., North Sutton, USA) on the rotating turntable. The thickness of the ground powder in the petri dish was approximately 3 cm to ensure

complete diffuse reflectance, following the recommendation by Williams (2019) for wheat and similar-sized seeds.

Table 3.1: Operating specifications of the two spectrophotometers

Item	Specifications	
	NIRS analyzer ⁶	Field spectrophotometer
Brand	FOSS DS2500	ASD FieldSpec4
Measurement Mode	Reflectance	Reflectance
Wavelength Range	400 - 2500 nm	350 – 2500 nm
Spectral resolution	0.5 nm	1 nm
Dried leaves presentation to instrument	Samples cup	Sample turntable
Fresh leaves presentation to instrument	Samples cup	Leaf clip
Weight	27 kg	5.44 kg
Mechanical environment	Stationary	Portable (leaf clip), Stationary (turntable)
Voltage supply	100-240 V AC	90-240 VAC or external battery
Format of outputs	Nir format	Export to ASCII text files is possible

To create the laboratory datasets, the laboratory-based spectra were developed using the same methodology discussed in Section 2.3. Two datasets were formed, one for fresh leaves and another for dried leaves, representing the laboratory spectral measurements in their respective testing modes, and hereafter called “*lab datasets*.” The data generated by the ASD spectrophotometer Indico Pro software provided reflectance values at every 1 nm from 350 nm to 2500 nm. To reduce the impact of noise, the spectral range from 350 nm to 399 nm, which had a comparatively lower signal-to-noise ratio, was excluded from the analysis. For the field datasets, the in-field spectral measurements of fresh leaves were averaged across 20 leaves for each data point. In contrast, the in-field spectra of dried leaves were averaged every three readings. This averaging process resulted in two datasets representing the in-field spectral measurements for fresh and dried leaves, hereafter called “*field datasets*.” Both the lab and field datasets were down sampled to an

⁶ Correspond to Table 2.1.

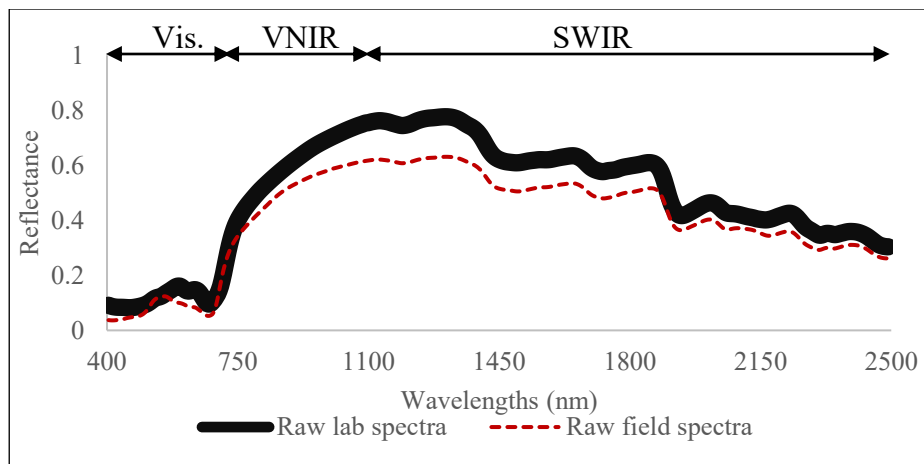
8 nm resolution, resulting in a total of 262 wavebands with centers ranging from 404 nm to 2492 nm. This down sampling process ensured consistent data representation across the different datasets.

3.2.2. Raw spectra from lab and field spectral instruments

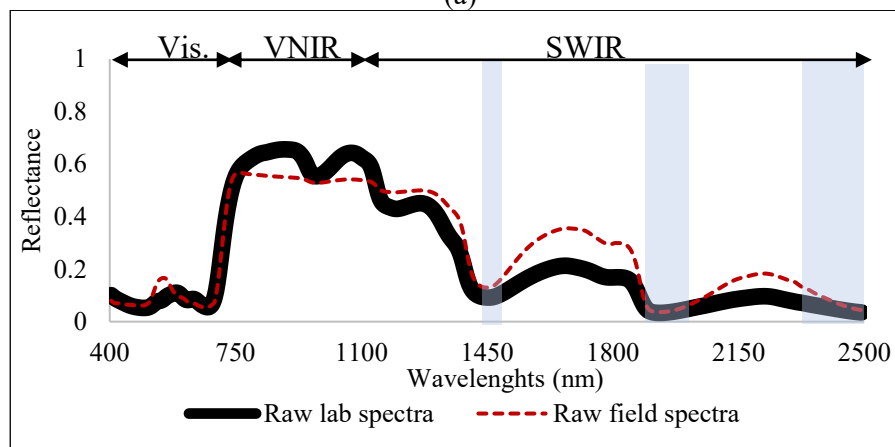
The comparison of spectral measurements between dried leaves conducted in the field and in the lab revealed some interesting findings. In the VNIR-SWIR range, both measurements showed similar peak signals of reflectance values (Figure 3.1 (a)). However, there was a noticeable difference in the peak signals observed in the red and green regions of the Vis range. The lab measurements had a peak signal in the red region (~660 nm), while the in-field measurements had a peak signal in the green region (~540 nm). This prominent spike in the green waveband range for the in-field measurements is believed to be caused by chlorophyll reflectance, as suggested by previous studies (Peng et al., 2020; Botha et al., 2006).

On the other hand, the shift of the reflectance peak to the red region for the lab measurements could potentially be attributed to vegetation stress. This stress may have resulted from the time gap between leaf collection and lab analysis, leading to a decrease in chlorophyll content. This decrease could alter the proportion of light-absorbing pigments and result in reduced overall absorption with chlorophyll (Sonobe and Wang, 2017; Carter and Miller, 1994). It is known that exposure to stress can induce the production of flavonoids as a secondary antioxidant defense system in plant tissues (Kumar and Pandey, 2013). Studies have shown that spectra from deteriorated leaves tend to have higher values around ~685 nm compared to green leaves (Mattila et al., 2021). Moving on to the comparison of fresh leaves, both lab and in-field measurements of the Vis-VNIR-SWIR

reflectance spectra exhibited similar spectral features (Figure 3.1 (b)). These features included a green peak around 560-580 nm, a red edge transition around 740 nm, a plateau from 800 to 1300 nm, and water absorption features at 1400, 1900, and 2500 nm. Overall, the lab measurements showed lower reflectance values in the Vis range compared to the in-field measurements, and this pattern was also observed in the VNIR-SWIR range. Within the wavelength range of 700 nm to 1500 nm, both datasets displayed similar spectral patterns, with high reflectance peaks observed for the in-field measurements at around 940 nm and 1200 nm.



(a)



(b)

Figure 3.1: Lab, and field (raw), intensities for a representative data point of the test data. in the (a) dried and (b) fresh modes. The blue regions highlight water absorbance wavelengths

3.3. CALIBRATION OF LAB AND FIELD SPECTRA AT EACH WAVEBAND AS A SINGLE ENTITY

Arai (2013) and Coliban et al. (2020) conducted evaluations on the calibration between remote sensing data and surface reflectance by employing a nonlinear mixing model and artificial neural networks. Silva et al. (2018) utilized an SVR model with a Gaussian radial basis kernel to predict sugar content in grape berries using spectral imaging measurements, demonstrating a satisfactory performance on unseen data from different grape varieties. Ren et al. (2023) highlighted the suitability of SVR for calibrating remote sensing data with excellent generalization capabilities. Their study also emphasized the significance of parameter selection in SVR and its impact on estimating chlorophyll concentration and suspended sediment concentration in water bodies. Similarly, Guo et al. (2021) emphasized the importance of variable selection prior to implementing SVR for accurate calibration in predicting soil nutrients using Vis-NIR spectra. To enhance protein prediction accuracy in wheat, Xu and Wang (2014) employed a neighbor-based weighted regression and SVR, incorporating local information among samples to achieve improved results. The application of SVR models was then extended to calibrate every waveband separately, using the local neighborhood of the target waveband. This method suggests avoiding the assumption of uniformity in the calibration function throughout the entire spectrum, and instead, it learns distinct transformation functions that consider potential nonlinearities in the mapping functions.

The assumption that the reference data is the lab spectra (target variables acting as y), and the field spectra acting as response variables (x) was maintained. The spectral data were first split into six subsets based on the source plot of the leaves to prevent data leakage

due to within-plot dependencies between observations. One of the subsets was reserved as a holdout split, to enable evaluation of the final performance of the calibration model on data that was not at all used for model development. The remaining five were used for model development, with 5-fold cross validation used to obtain performance estimates as a part of the model development process. The methodology developed for mapping field measurements into values compatible with lab spectra involved employing nonlinear regression models to fit the reflectance values for each target waveband. To begin, the set of response variables was prepared as described in Equation (3.1):

$$y = \{y_{ki}; i = 404, 412, \dots, 2492, k = 1, \dots, n_{samples}\}, \quad (3.1)$$

where y_{ki} is a target waveband of a specific sample. For each response y_{ki} , a corresponding set of predictors x_{ki} was calculated based on a subset of the reflectance values of the field measurement using Equation (3.2):

$$x_{ki} = \{x_{k(i+j)}, x_{k(i+j)}^2; j \in \{0, \pm 8, \pm 16\}; k = 1, \dots, n_{samples}\}, \quad (3.2)$$

where x_{ki} is subject to $400 \leq (i + j) \leq 2500$. This equation resulted in a maximum of 10 predictive features - the matching waveband (x_{ki}) plus up to two lower and two higher wavebands and their squared values, truncated at both ends of the measurement range. The consideration of the neighbor wavebands in addition to the central one aimed at capturing the local influence of the adjacent variables for possible relevant valuable information, and the addition of squared values allows an additional degree of nonlinearity between the target and response wavebands. The model development and calculation were performed using R. Figure 3.2 illustrates the developed calibration methodology on how to extract the target and predictor variables for a given sample k .

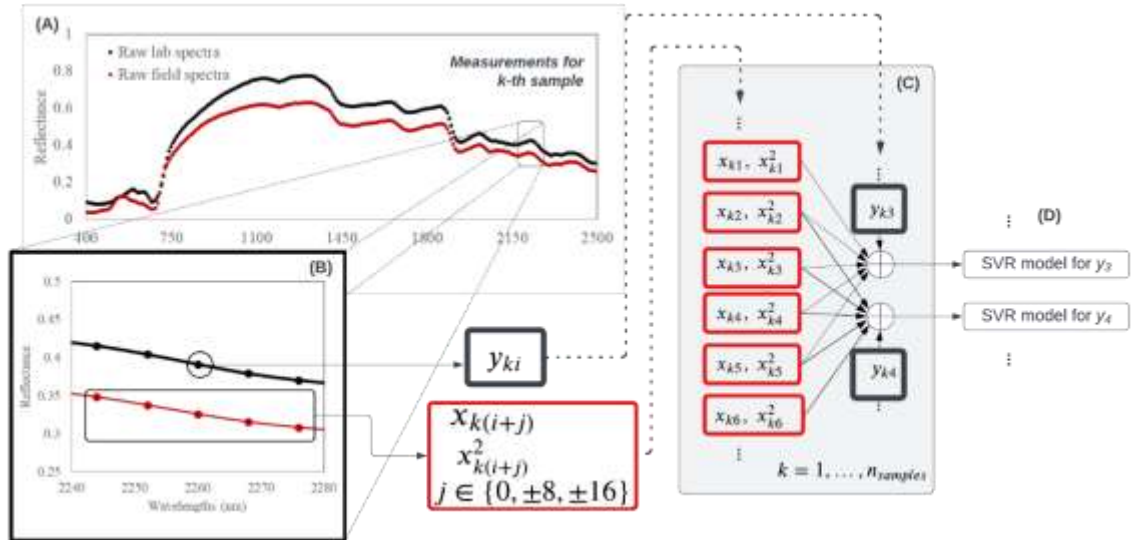


Figure 3.2: General structure of the proposed calibration modeling structure. (A) Matched spectra obtained from lab and field measurements. (B) Extracting a target variable (intensity at a specific waveband) and associated predictors (intensities at the local neighborhood of the target waveband). (C) Building of predictive models using the target and predictor variables values across all training data points. (D) Resulting models

The next step was to implement SVR models at each target waveband to give the estimated field calibrated spectra. The 262 resulting models were applied to the testing data points, resulting in the estimation of the field calibrated spectra. Figure 3.3 shows the steps of applying the calibration method using lab and in-field spectra.

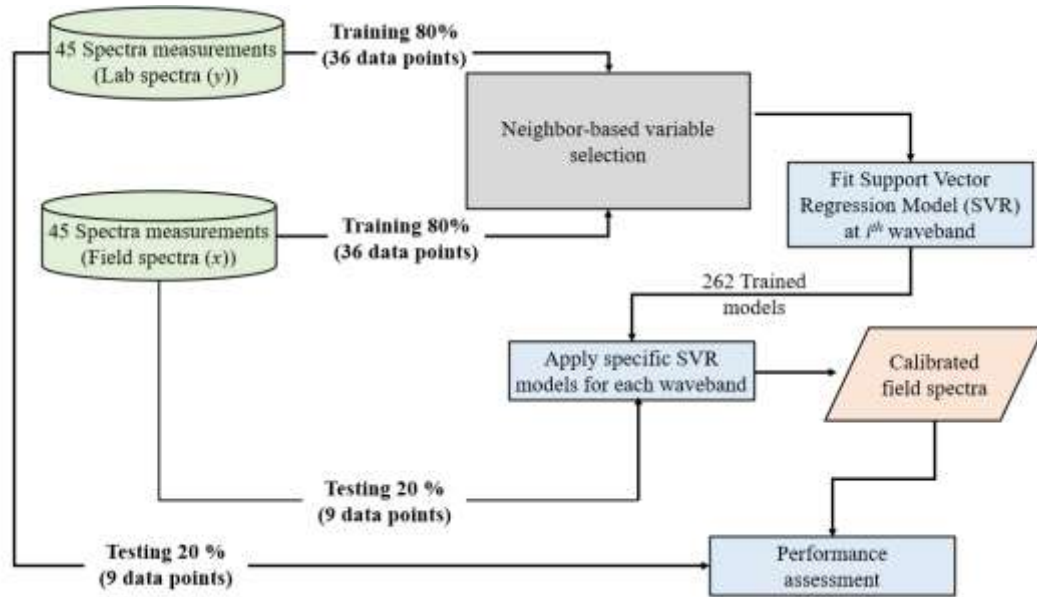


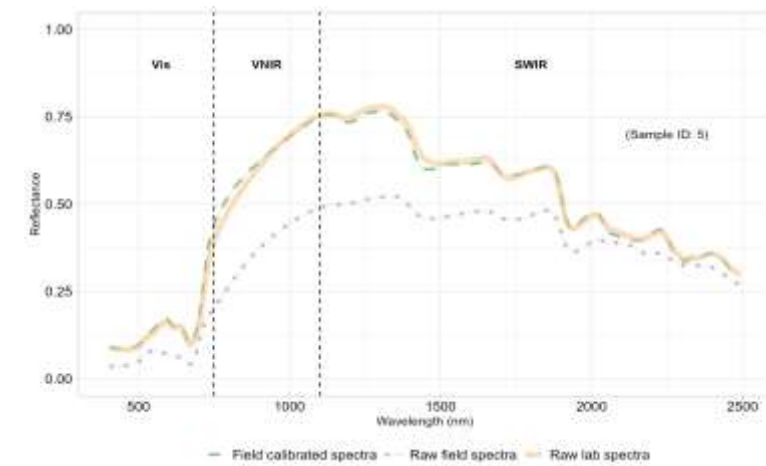
Figure 3.3: Flow of the calibration method towards developing calibrated field spectra involving the neighbor-based selection of variables followed by SVR

3.3.1. Raw spectra of in-field datasets in comparison to lab datasets

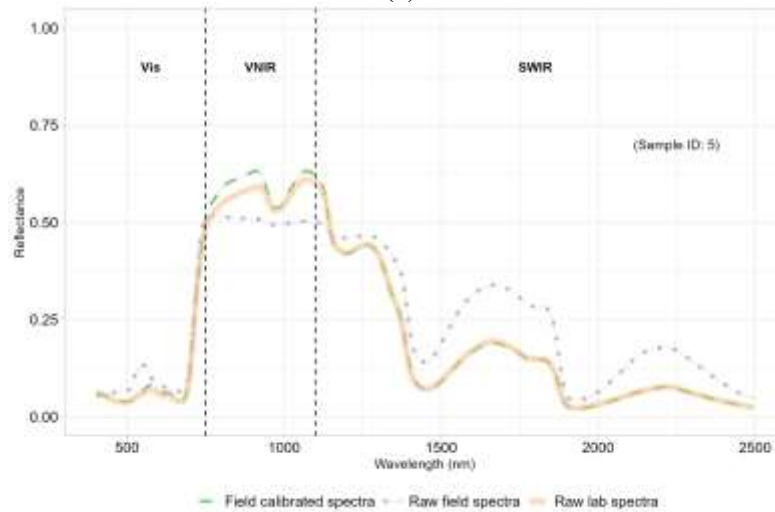
Figure 3.4 presents the resulting field calibrated spectra of both modes in reference to lab and field spectra. The calibration could successfully shift the spiked reflectance values from green (540 nm) back to red waveband region at around 660 nm (Figure 3.4 (a)), which may refer to the application of SVR as a nonlinear model and the consideration of the neighboring wavebands influence. The influence of a waveband width up to 32 nm (Equation 3.2) was enough to modify the signal peak from green (500 nm – 565 nm) to red (625 nm – 750 nm) as both regions are confined. Moreover, the plateau of the spectral reflectance values within the NIR-SWIR range of the field calibrated spectra matched the similar signals at the lab-based spectra (Figure 3.4 (a)), which shows the successfulness of applying the nonlinear SVR model at each waveband between lab and field spectra.

Likewise, the calibration process effectively achieved alignment between the plateau of the spectrum within the SWIR range in the fresh mode, while also adjusting the

green peak around the 560-580 nm range (Figure 3.4 (b)). Additionally, the calibration method addressed discrepancies between the lab and field spectra in the waveband region from 760 nm to 1140 nm. This was accomplished by leveraging SVR to capture and replicate the nonlinear patterns between the two sets of spectra. The differences between the field spectra and the field calibrated spectra, in comparison to the lab spectra, were depicted in (Figure 3.5) through the visualization of the average absolute errors across each waveband for the nine testing data points.



(a)



(b)

Figure 3.4: Lab, field (raw), and field (calibrated) intensities for a representative data point of the test data. (a) dried and (b) fresh modes. The blue regions highlight water absorbance wavelengths

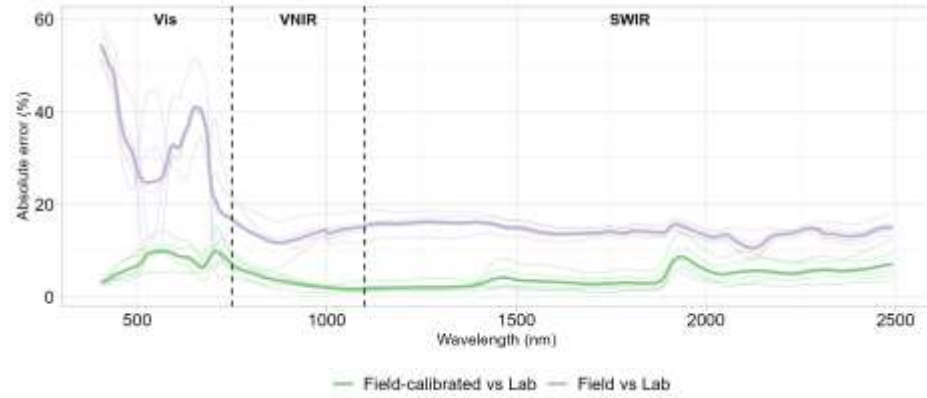
3.3.2. Calibrated of in field datasets in comparison to lab datasets

In Figure 3.5, a significant reduction in the absolute error values between the lab and field calibrated spectra observed in the dried mode. This reduction is particularly notable within the Vis range, where the absolute error decreased from 60% to 10% (Figure 3.5). This indicates that by utilizing the SVR model at each waveband, along with the neighboring variables and their squared terms, it was possible to align the nonlinear patterns between the lab and field spectra. The reduction in absolute error at 540 nm and 708 nm noticed to be not substantial, as these regions of the spectrum initially exhibited minimal differences between the raw field and lab spectra, as shown in Figure 3.4 (a). Additionally, the calibration demonstrated success within the SWIR range, with a consistent reduction of approximately 13% in absolute error, since the raw lab and field spectra initially had similar plateau reflectance values (Figure 3.4 (a)).

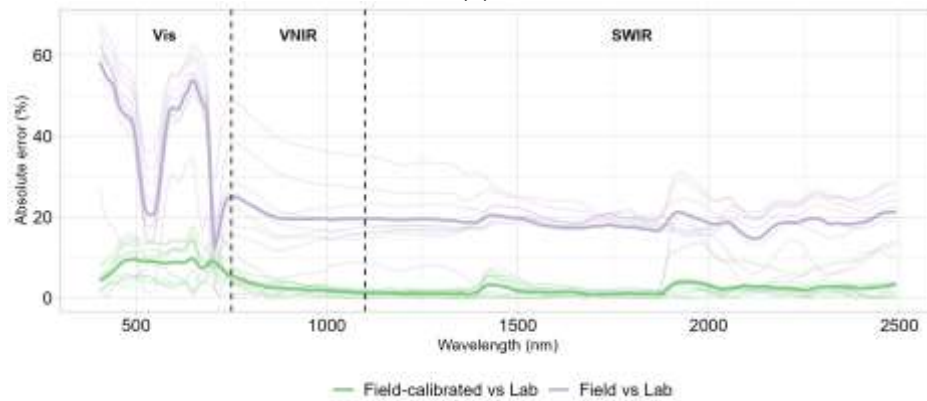
Figure 3.6 reveals a remarkable reduction in error values not only within the Vis range but also the SWIR when utilizing field calibrated spectra in the fresh mode. The calibration method significantly reduced discrepancies between the lab and field spectra by approximately 80% within the green waveband range, where the peak signals between the lab and field were noticeably different. This significant reduction was achieved by adjusting the spike of spectral features from the green to red waveband, aligning the values with those of the lab spectra. However, the reduction in absolute error did not exceed 15% at 560 nm, 770 nm, and 1160 nm – 1300 nm, which corresponded to the low discrepancies that was already present between the lab and field raw spectra in those specific ranges (Figure 3.4 (b)). Notably, the developed calibration method further contributed to minimizing discrepancies within the SWIR range in the fresh mode, with an error reduction

of approximately 100% (Figure 3.6). The figure shows that the high discrepancies between the raw lab and field spectra within the SWIR region were not solely limited to the water absorbance wavebands but were present throughout the entire region.

This can be attributed to sample handling more than variations in environmental conditions between field and lab measurements, as both spectrophotometers measured the reflectance spectra of fresh leaves using artificial light, without being influenced by ambient environmental conditions in the field. However, sample handling might introduce differences in reflectance values when using both spectrophotometers. In a lab setting, fresh leaf samples were prepared and handled more carefully, ensuring consistent measurement conditions. On the other hand, in the field setting, obtaining consistent measurements might be challenging due to sample variability caused by variable leaf coverage within the leaf clip, resulting in inconsistent measurements. The reason reflectance spectra within the SWIR range are more sensitive to such sample variability is that this region exhibits lower reflectance values compared to the Vis and VNIR ranges, meaning that the signal-to-noise ratio within this region is lower (Williams, 2019). Therefore, inconsistent sample handling could introduce more variability to the measurements within the SWIR range compared to the Vis and VNIR ranges, as shown in Figure 3.6.

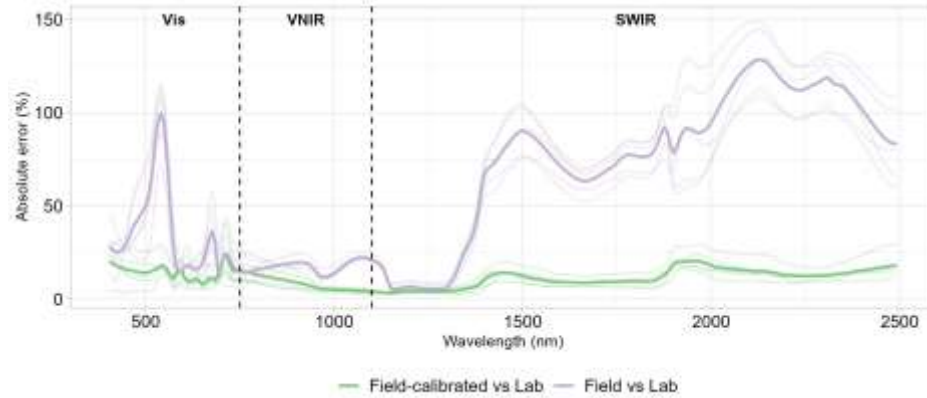


(a)

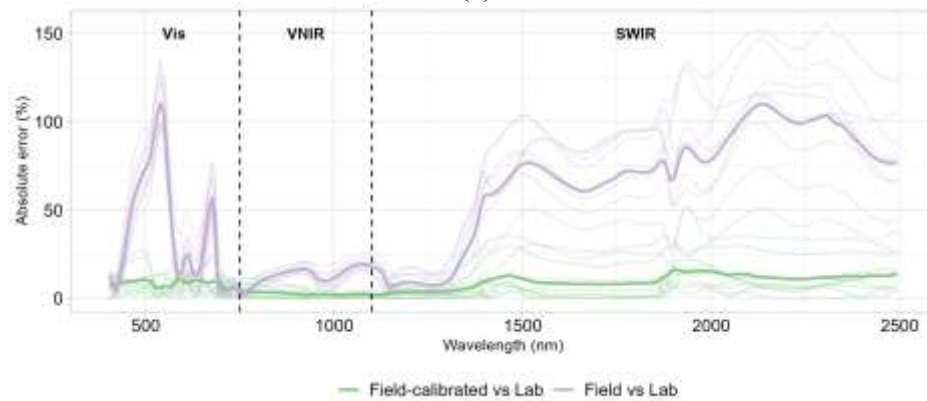


(b)

Figure 3.5: Absolute error at each waveband between lab and field calibrated spectra, vs. lab and field spectra at the dried mode of the (a) cross-validation, and (b) holdout split. The thick middle line represents the mean absolute error



(a)



(b)

Figure 3.6: Absolute error at each waveband between lab and field calibrated spectra, vs. lab and field spectra at the fresh mode of the (a) cross-validation, and (b) holdout split. The thick middle line represents the mean absolute error

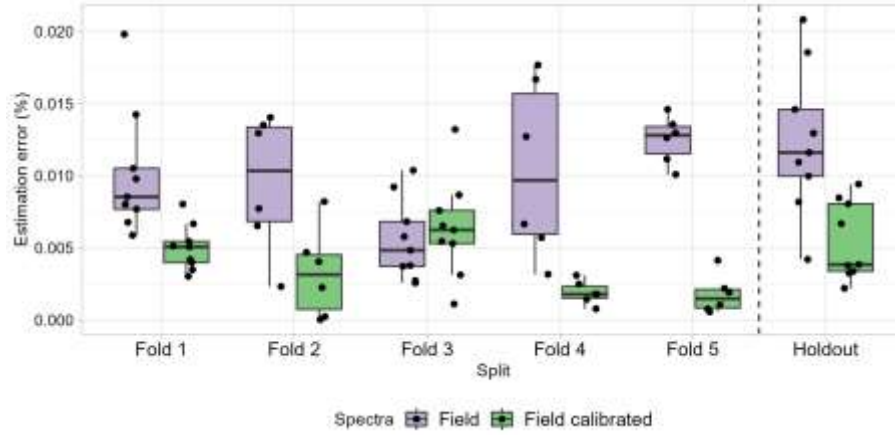
3.3.3. Verifying the calibration at each waveband in estimating NPK

To verify the performance of the proposed calibration method, NPK content in potato petioles of the 45 data points was estimated using the lab and field measurements before and after calibration (See section 3.2.1). For each type, a dataset was built where the chemical composition of NPK acted as responses and the spectra acted as predictors following same protocol discussed in Section 2.4. Then, NPK models generated in Section 2.6 were used to test the 9 testing data points as those models showed a reasonable trend to estimate these three elements on data collected over an entire season. The RMSE and

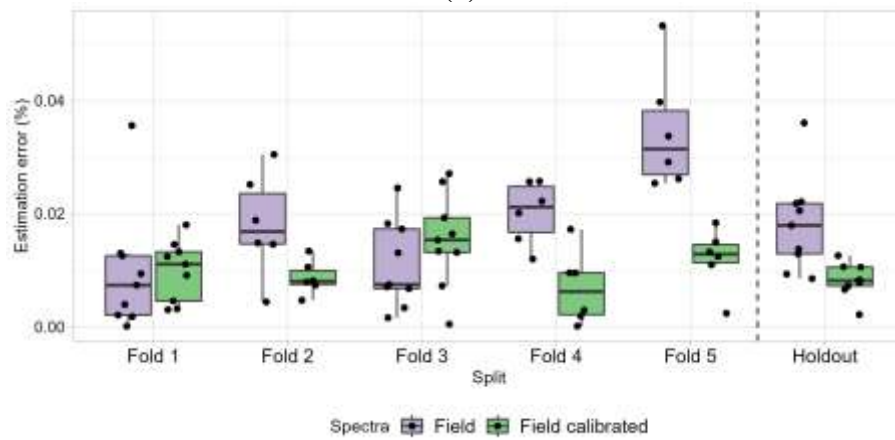
nRMSE between the actual concentrations and the estimated ones given by lab, field, and field calibrated spectra was used to assess the calibration performance.

The results showed an improved alignment between the lab and field calibrated spectra, as shown in Figures 3.7 and 3.8. In the dried mode, the mean differences in estimated nutrient levels between models fed the lab readings and the field-calibrated values for the holdout set were 0.55%, 0.82%, and 0.76% for N, P, and K, respectively in comparison with mean discrepancies of 1.24 %, 1.81%, and 2.35% when using the raw (uncalibrated) field values. In the fresh mode, a greater improvement was observed for the three nutrients: 0.57% vs. 1.67% for N, 1.12% vs. 3.15% for P, and 1.03 vs. 4.45% for K. It is noteworthy that the deviation in the estimation of K dropped to around 1% after calibration, in comparison to approximately 4.5% when using the uncalibrated field data.

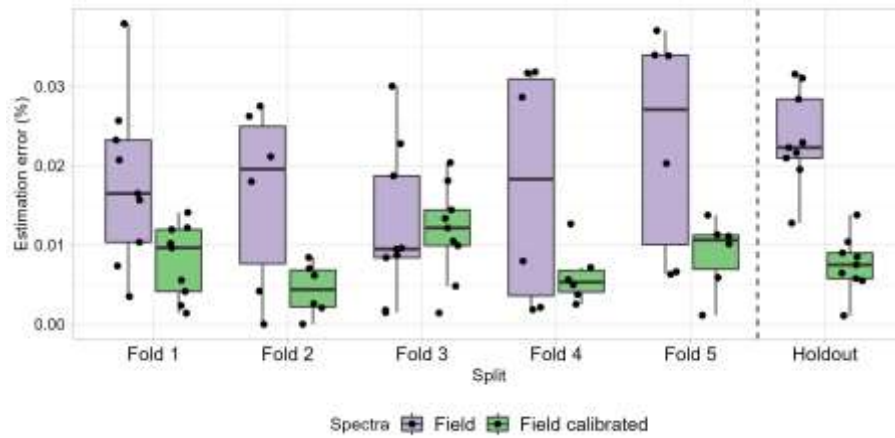
The results consistently demonstrate improvements in nutrient estimation following the calibration process. The modest improvement observed in the dried mode can be attributed to specific spectral features present in the raw field spectra, where significant wavebands are positioned around the wavelength regions (690 nm - 720 nm) that already closely resemble the corresponding features observed in the lab spectra (Figure 3.5 (a)). On the other hand, the substantial improvement observed in the fresh mode can be attributed to the significant wavebands located within the SWIR range, where the calibration process succeeded in reducing discrepancies by up to 100% (Figure 3.5 (b)). These results suggest that the calibration model effectively mitigates differences between the two spectral measurement modes across the spectrum. This can be particularly beneficial when estimating nutrient concentrations that rely on SWIR spectra, such as Zn, Fe, B, Cu, and Al.



(a)

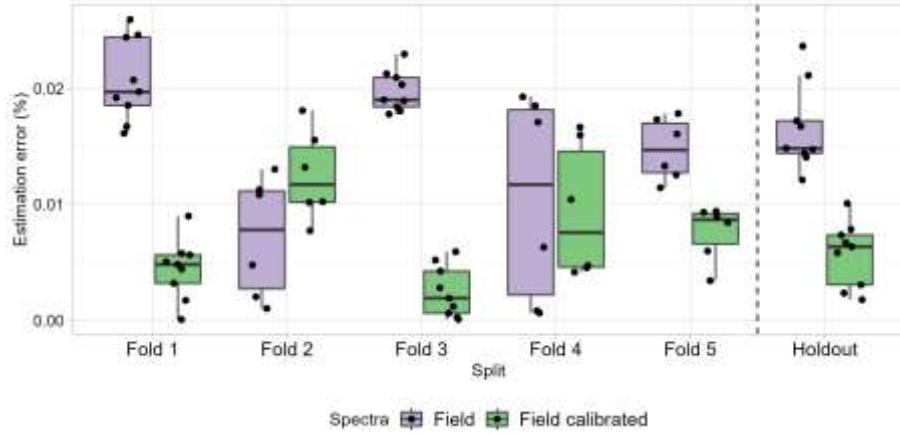


(b)

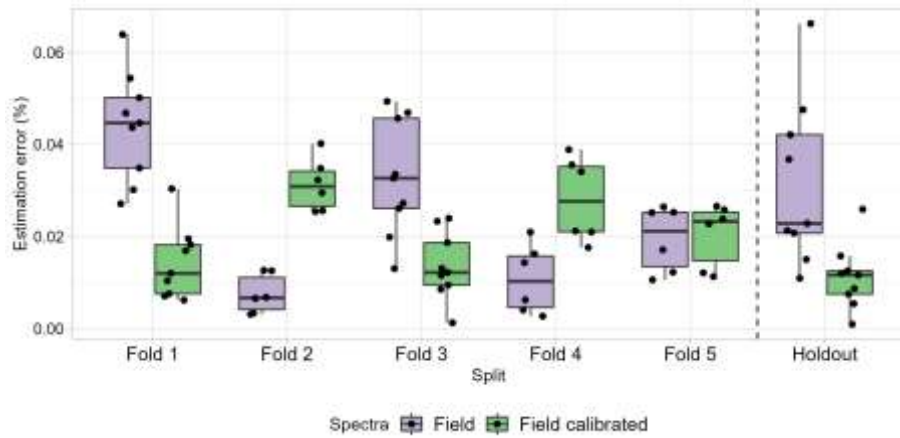


(c)

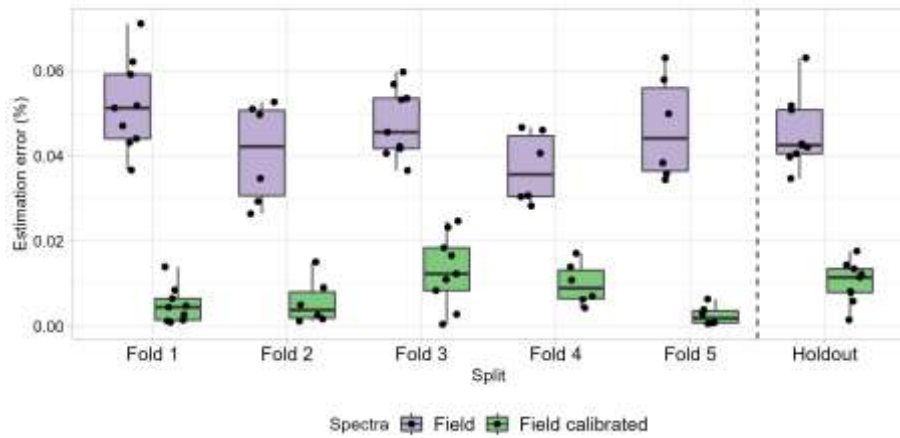
Figure 3.7: Estimated error of (a) N, (b) P, and (c) K estimation values of the five cross-validation folds and the holdout split of the dried mode by field and field calibrated spectra



(a)



(b)



(c)

Figure 3.8: Estimated error of (a) N, (b) P, and (c) K estimation values of the five cross-validation folds and the holdout split of the fresh mode by field and field calibrated spectra

Table 3.2: Prediction performance in estimating NPK using the lab vs. field spectra, and lab vs field calibrated spectra of both modes

Estimated values using spectra		Dried mode			Fresh mode		
		N	P	K	N	P	K
Lab vs. field	RMSE	0.057	0.005	0.172	0.069	0.009	0.308
	nRMSE	0.888	0.907	1.210	0.841	0.510	1.259
Lab vs. field calibrated	RMSE	0.026	0.002	0.059	0.026	0.003	0.076
	nRMSE	0.250	0.129	0.434	0.653	0.469	0.591

3.4. CONCLUSION

The developed calibration method using a neighbor selection of variables with their squared terms prior to the implementation of SVR at each waveband could successfully align the discrepancies between lab and field spectra. The calibration further contributed to minimizing the discrepancies between lab and field spectra within the SWIR range of the fresh mode up to 100%. The developed calibration method was verified by estimating NPK in potato petioles using the three datasets of spectra: lab, field, and field calibrated spectra. The general consistent improvement in estimating NPK using field calibrated spectra in reference to lab estimations validates the ability of the developed calibration method in eliminating the discrepancies between the two spectral measurements. Though, the results of NPK estimation in the fresh mode highlight the importance of having consistent spectral measurements for nutrients that have significant wavebands positioned within the SWIR.

CHAPTER 4: COMPREHENSIVE METHODOLOGY TO ESTIMATE NUTRIENTS USING COMBINATION OF MACHINE LEARNING TECHNIQUES

4.1. BACKGROUND

As discussed in Chapter 1, the crop reflectance was evaluated for its capacity in evaluating the nutritional content in a plant by using the Vis-NIR region of the spectrum. Chapter 2 further approved the initial hypothesis of the indirect relationship existing between potato petioles and leaf reflectance. Chapter 2 also discussed the successful use of Lasso models to address the high dimensionality of the spectral data by shrinking the irrelevant wavebands to zero. However, Lasso models did not account for the interlinkages between nutrient concentrations. This chapter addresses the interlinkages between nutrients and their influence on the performance prediction. To evaluate multivariate relationships, datasets from the 2020, 2021 and 2022 seasons were employed.

4.2. DEMONSTRATION OF MULTIVARIATE RELATIONSHIPS BASED ON GROUND TRUTH DATA

To establish a baseline, the existing relationships among nutrients were investigated and confirmed to identify parameters influencing other elements in a multivariate modeling. The interlinkages among the nutrients were found by the literature to be arranged around four groups, as depicted in Figure 4.1, where black and red lines indicate positive and negative correlations: (1) N is influenced by other macronutrients - P, K, Mg, and S (Schleuss et al., 2020; Koch et al., 2020; Muttucumaru et al., 2013; Naumann et al., 2020); (2) P is influenced by Ca and S, as well as N, and two micronutrients, Al and B (Abbasian et al., 2018; Naumann et al., 2020; Steiner et al., 2012; Rad et al., 2022); (3) K, while sharing N, S, and Mn with P, is also influenced by Mg and Fe (Heenan and Campbell,

1981; Koch et al., 2020; Schmidt et al., 2020), and (4) Ca is influenced by P and B, as well as Mg, zinc Zn, and Cu (Alaoui-Sossé et al., 2004; Koch et al., 2020).

To validate these interlinkages, Spearman correlation (r_s) was calculated for pairs of petiole nutrient contents expected to exhibit dependencies. Spearman correlation assesses the non-linear relationship between two variables without assuming any specific distribution for the variables (Schober et al., 2018). The *cor* function in R software was employed to evaluate nutrient relationships based on 179 data points.

Figure 4.1 illustrates nutrient interlinkages based on literature (represented by the line color) and the sign of r_s values calculated from our data. The figure shows a comprehensive alignment between r_s values and the existing literature. As an example, the positive r_s value of 0.832 observed between N and P concurs with the findings reported by Schleuss et al. (2020), indicating a positive correlation between these elements. Similarly, the counteracting effect, as denoted by r_s of -0.637 between Ca and P, aligns with the conclusions drawn by Abbasian et al. (2018). Furthermore, the calculated r_s value of -0.825 between K and Mg provides an explanation for the antagonistic relationship between these elements, in contrast to the synergic effect between Ca and Mg with r_s value of 0.704.

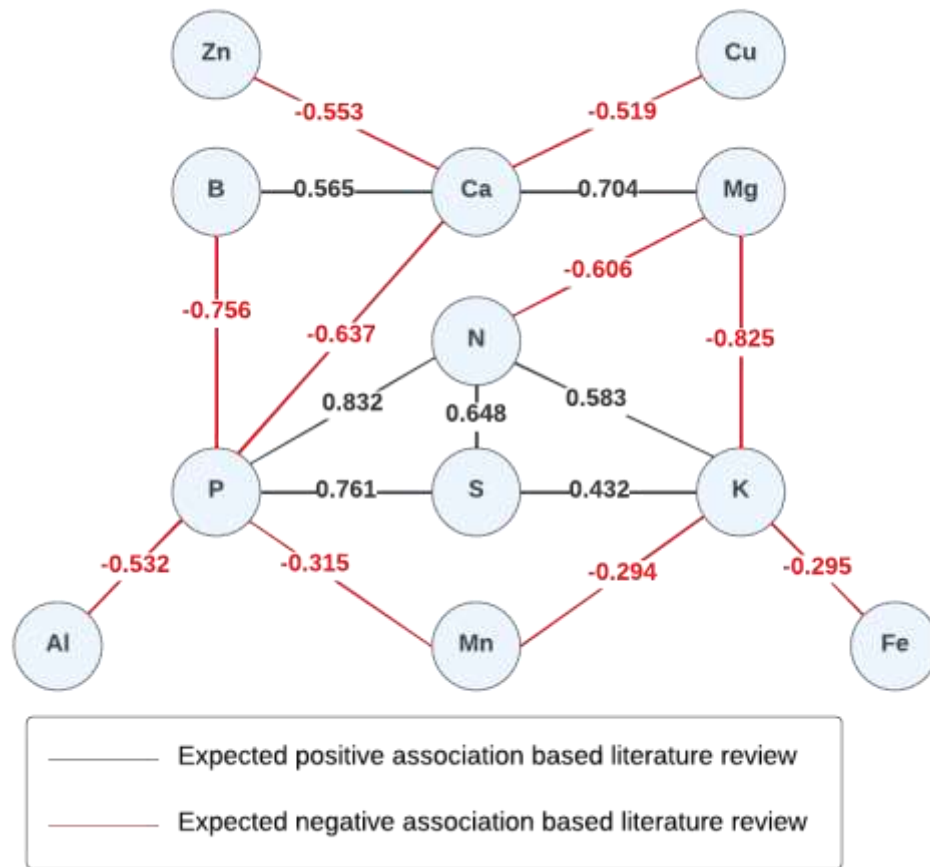


Figure 4.1: Spearman correlation values among nutrient concentrations. The numbers are the sample estimates based on the data collected in this research, while the colors indicate the expected direction of each association based on literature

4.3. MODEL SELECTION AND CONCEPT

While the interdependence of nutrient concentrations is well-known in agriculture, there is a lack of previous research addressing this specific issue. However, in clinical research, the analysis of correlated outcomes is frequently conducted. – e.g., when studying how different symptoms may result from the same underlying causes. In particular, a technique known as “*stacked multivariate regression*” can be used to predict potentially correlated outputs by fitting multiple univariate regressions organised in two layers (Rauschenberger, and Glaab, 2021).

In the first layer (“*base layer*”), a regression model is developed to independently estimate each target variable using the original predictive features (like the models described in Chapter 2, where chemical concentrations were predicted separately based on reflectance values). The outputs from the base layer are then used as inputs in a second layer (“*meta layer*”), where they are treated as predictors. The base layer focuses on learning the relationship between features and individual target variables, while the second layer considers potential dependencies among the different target outcomes. Figure 4.2 illustrates the concept of the stacked multivariate regression.

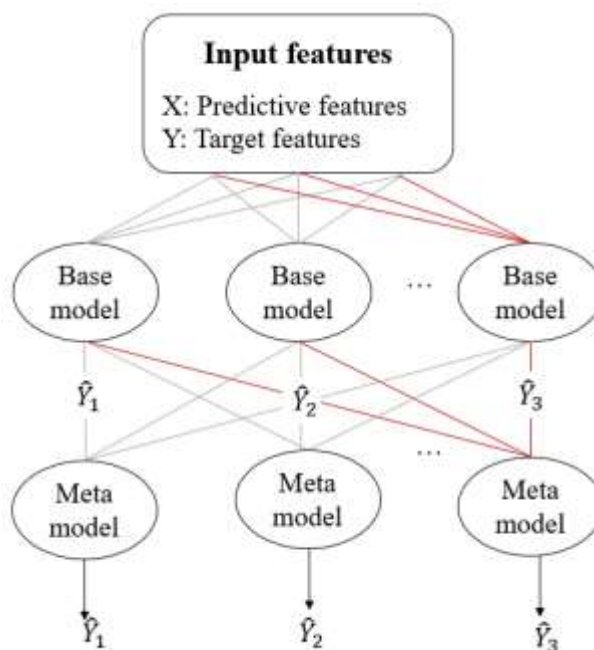


Figure 4.2: Concept of stacked multivariate regression at base and meta layers. \hat{Y} represents the predicted target feature at the base layer and \hat{Y} presents the predicted target feature at the meta layer

As a step to develop the stacked multivariate regression, two matrices were developed, the inputs of X matrix (spectral data), and outputs of Y matrix (nutrients). The

*joinet*⁷ package in R software was used to implement linear models at both layers (Rauschenberger and Glaab, 2021), and a fixed random seed was set for reproducibility.

4.4. IMPROVEMENT IN ESTIMATING NUTRIENTS

4.4.1. Model performance measures

Considering the remarkably high number of features (262 wavebands) when compared to the number of observations (179 data points of Seasons 2020 and 2021)⁸, Lasso was selected to be implemented in both layers using the *joinet* package. The training was implemented using 80% of the data points, with the remaining 20% reserved to assess the model performance on unseen data. The performance of the estimation at both layers was assessed using r^2 , RMSE, and nRMSE to estimate the effect of adding the meta modeling on the estimation performance of each nutrient. The model fitting and assessment were separated for the dried and fresh leaves data.

A comparison was first done on the estimated values between singles and cross seasons at the base layer (Table 4.1). The difference in the model performance at each season may be attributed to differences in spectral measurements occurring due to several factors associated with plant diseases, deficiency of other nutrients, and physiological stresses. The enhancement in the estimation performance of N and P of the base layer can be attributed to the increase in sample size. A larger volume of data can lead to improved r^2 values by reducing the uncertainty in model coefficient estimates, especially if the additional data contains relevant information about the relationship between chemical

⁷ The "*joinet*" package in R software offers the Lasso and Ridge regression methods only, as of April 2023.

⁸ Number of data points taken from open farms in Season 2020 and Season 2021 is 40, and 144 respectively. Among these data points, a total of 179 observations included results for all 12 nutrients, with no missing values in the nutrient matrix.

content and reflectance spectrum variables. However, it should be noted that simply increasing the data size does not necessarily guarantee improvements in generalization performance, particularly if the new data comes from a different data-generating distribution. Factors such as variations in outdoor conditions or sampling methodologies across seasons can influence the relationship between reflectance spectra and nutrient concentration, potentially leading to a decreased model performance in a combined-season scenario, which might explain the noticeable decrease in the estimation performance of B, Zn, and Cu, in particular (Table 4.1). Additionally, discrepancies in spectral measurements resulting from factors like plant diseases, deficiencies in other nutrients, and physiological stresses can also contribute to a reduced model performance when combining data from different seasons (Muñoz-Huerta et al., 2013; Liu et al., 2021a). Finally, the change in the testing mode of the highest estimation performance might also refer to the change in spectral characteristics of a sample from one season to another.

Table 4.1: Prediction performance values in estimating macro and micronutrients in single and across seasons of the best testing mode

Element	Season 2020		Season 2021		Season 2020 + 2021	
N	Fresh	0.593	Dried	0.75*	Dried	0.804
P	Dried	0.745*	Dried	0.75*	Dried	0.668
K	Fresh	0.748 *	Dried	0.63*	Fresh	0.328
Ca	Dried	0.321	Dried	0.49	Dried	0.451
Mg	Dried	0.765	Dried	0.74	Dried	0.721
S	Dried	0.504 *	Dried	0.64	Fresh	0.610
Mn	Dried	0.241	Dried	0.42	Dried	0.407
B	Dried	0.615	Dried	0.37*	Fresh	0.696
Zn	Dried	0.535	Dried	0.24	Dried	0.254
Fe	Dried	0.647 *	Dried	0.67	Dried	0.368
Cu	Dried	0.578	Dried	0.24	Dried	0.218
Al	Dried	0.668 *	Dried	0.54	Dried	0.542

* r^2 values are less than 0.04 compared to values of another testing mode.

Tables 4.2, 4.3, 4.4, and 4.5 present the performance of the multivariate regression at the meta layer in comparison to the base layer of both modes: dried and fresh. As illustrated, the estimation performance did not improve at the meta layer of all nutrients. For example, N and Mg are the only macronutrients where the r^2 values improved at the meta layer from 0.804 to 0.818, and 0.721 to 0.740 in the dried mode (Table 4.2). In addition, K, Ca, S, and Zn are the only nutrients that witnessed an improvement in the meta layer in comparison to the base layer in the fresh mode (Tables 4.4, 4.5). This would happen due to different factors including that the interlinkages between nutrients did not necessarily improve the estimation of a specific nutrient at the meta layer, the high dimensionality of the input features (wavebands) may cause noise in underlying the true patterns, and lastly, the assumption of linearity relationship at the meta layer does not necessarily reflect the existing relationships between nutrients interlinkages.

The former factor is not likely to be the reason for the unimproved estimation performance at the meta layer, as evidenced by Figure 4.1 which visibly shows intricate biological relationships between nutrients. To address the second factor, a feature reduction method called minimum redundancy - maximum relevance (MRMR) feature selection for each nutrient was used to extract the most explanatory features prior to the base layer models (Ding and Peng, 2005). In addition, different nonlinear models (RF, SVR, and XGB) were assessed at the meta layer models to enable the model to capture potential nonlinear relationships between the concentrations of different nutrients. Figure 4.3 illustrates the modified approach that was followed for the estimation pipeline at two layers.

Table 4.2: Prediction performance of the stacked multivariate regression at the base and meta layers of the macro-nutrients at the dried mode

			N	P	K	Ca	Mg	S
Base layer	Training	r^2	0.798	0.845	0.763	0.541	0.795	0.696
		RMSE	0.435	0.049	1.215	0.402	0.196	0.022
		nRMSE	0.111	0.112	0.119	0.146	0.111	0.129
	Testing	r^2	0.804	0.774 ⁹	0.296	0.451	0.721	0.528
		RMSE	0.420	0.030	1.660	0.430	0.200	0.030
		nRMSE	0.123	0.143	0.226	0.203	0.154	0.231
Meta layer	Training	r^2	0.778	0.848	0.738	0.498	0.796	0.682
		RMSE	0.456	0.050	1.305	0.439	0.200	0.023
		nRMSE	0.117	0.114	0.128	0.160	0.113	0.133
	Testing	r^2	0.818	0.774	0.264	0.350	0.740	0.501
		RMSE	0.440	0.030	1.670	0.450	0.200	0.030
		nRMSE	0.129	0.143	0.227	0.212	0.154	0.231

* Bold font for the r^2 that is higher in the meta layer than the base layer of a nutrient.

Table 4.3: Prediction performance of the stacked multivariate regression at the base and meta layers of the micro-nutrients at the dried mode

			Mn	B	Zn	Fe	Cu	Al
Base layer	Training	r^2	0.573	0.650	0.236	0.336	0.334	0.446
		RMSE	206.893	21.405	32.454	407.734	5.041	180.129
		nRMSE	0.082	0.100	0.121	0.096	0.099	0.119
	Testing	r^2	0.407	0.585	0.254	0.368	0.218	0.542
		RMSE	255.390	23.330	29.280	647.230	5.320	452.400
		nRMSE	0.189	0.212	0.188	0.196	0.197	0.187
Meta layer	Training	r^2	0.573	0.721	0.289	0.343	0.366	0.452
		RMSE	224.569	22.558	31.471	407.391	5.136	189.451
		nRMSE	0.089	0.105	0.117	0.095	0.101	0.125
	Testing	r^2	0.451	0.547	0.314	0.408	0.208	0.579
		RMSE	252.210	21.240	26.970	651.140	5.430	488.100
		nRMSE	0.187	0.193	0.173	0.197	0.201	0.202

* Bold font for the r^2 that is higher in the meta layer than the base layer of a nutrient.

⁹ Correspond to Table 2.9 where model gave an r^2 value of 0.668 in estimating P in combined seasons. The difference between these two values (0.774 instead of 0.668) is referred to the implementation of the stacked regression over two seasons, and the indoor concentrations of P alike to what was reported in Table 2.9.

Table 4.4: Prediction performance of the stacked multivariate regression at the base and meta layers of macro-nutrients at the fresh mode

			N	P	K	Ca	Mg	S
Base layer	Training	r^2	0.763	0.757	0.702	0.610	0.737	0.597
		RMSE	0.476	0.052	1.262	0.399	0.231	0.026
		nRMSE	0.122	0.119	0.130	0.145	0.124	0.154
	Testing	r^2	0.554	0.691	0.328	0.113	0.630	0.610
		RMSE	0.625	0.050	1.765	0.533	0.235	0.028
		nRMSE	0.178	0.208	0.183	0.212	0.142	0.178
Meta layer	Training	r^2	0.749	0.773	0.692	0.530	0.725	0.587
		RMSE	0.492	0.054	1.293	0.437	0.241	0.028
		nRMSE	0.126	0.124	0.133	0.159	0.129	0.165
	Testing	r^2	0.586	0.731	0.307	0.094	0.661	0.562
		RMSE	0.630	0.045	1.757	0.532	0.235	0.028
		nRMSE	0.179	0.189	0.183	0.211	0.141	0.177

* Bold font for the r^2 that is higher in the meta layer than the base layer of a nutrient.

Table 4.5: Prediction performance of the stacked multivariate regression at the base and meta layers of micro-nutrients at the fresh mode

			Mn	B	Zn	Fe	Cu	Al
Base layer	Training	r^2	0.292	0.625	0.267	0.196	0.106	0.390
		RMSE	306.031	22.056	32.628	379.252	5.939	166.915
		nRMSE	0.122	0.103	0.120	0.089	0.116	0.111
	Testing	r^2	0.082	0.696	0.236	0.282	0.018	0.318
		RMSE	232.345	28.268	28.828	588.926	4.981	420.149
		nRMSE	0.229	0.141	0.188	0.178	0.217	0.174
Meta layer	Training	r^2	0.263	0.661	0.322	0.253	0.265	0.435
		RMSE	313.205	23.500	32.468	394.691	5.798	171.932
		nRMSE	0.125	0.110	0.120	0.092	0.114	0.114
	Testing	r^2	0.105	0.735	0.094	0.288	0.065	0.438
		RMSE	220.865	27.824	30.402	625.054	5.015	443.689
		nRMSE	0.218	0.139	0.199	0.189	0.218	0.184

* Bold font for the r^2 that is higher in the meta layer than the base layer of a nutrient.

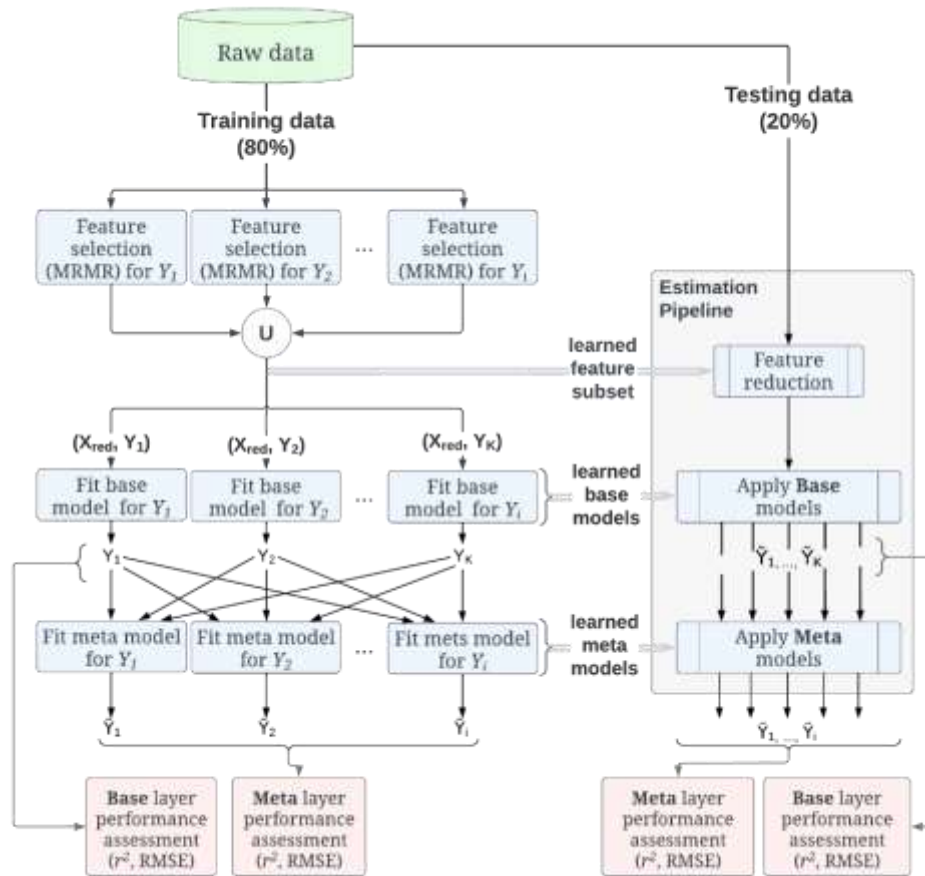


Figure 4.3: Stacked multivariate model development and model performance assessment

4.4.2. Model performance measures at feature selection method

Tables 4.6, 4.7, 4.8, and 4.9 show the prediction performance indicators of r^2 , RMSE, and nRMSE of using the MRMR feature selection step prior to fitting the stacked multivariate regression for both modes. The results of the testing dataset show that the *a priori* feature reduction has indeed improved the estimation performance of most nutrients. For example, the r^2 values of P and Mg improved from 0.774 to 0.818, and from 0.740 to 0.802 in the dried mode, and the r^2 value of estimating N has been improved from 0.586 to 0.653. The estimation performance at the meta layer of some nutrients remained stable around the same value as the base layer models. For example, the r^2 value of estimating P

was 0.819 at both layers in the dried mode (Table 4.6), and S estimation performance was 0.513 in the fresh mode (Table 4.8). This could be explained, for instance, if the full information about the concentration of some nutrients assumed to be already captured by the selected wavebands, in which case information about the (predicted) concentrations of other nutrients would not add further predictive power. The explanation might be valid, which means that if the waveband features are insufficient to estimate the concentration of a particular nutrient in the base layer, that nutrient contribution to the prediction of other nutrients in the meta layer is also likely to be weak. Another consideration arises from the assumption that the interdependent nutrients exhibit a linear relationship, which may not accurately reflect the actual scenario. Thus, three different nonlinear models were applied at the meta layer to assess whether the lack of improvement in the performance at the meta layer is indeed attributable to the linearity assumption.

Table 4.6: Prediction performance of the stacked multivariate regression at the base and meta layers after the feature selection using MRMR of the macro-nutrient at the dried mode

			N	P	K	Ca	Mg	S
Base layer	Training	r^2	0.837	0.802	0.852	0.747	0.871	0.728
		RMSE	0.381	0.044	0.941	0.309	0.150	0.021
		nRMSE	0.099	0.101	0.092	0.113	0.085	0.123
	Testing	r^2	0.895	0.819	0.485	0.463	0.802	0.558
		RMSE	0.356	0.036	1.459	0.391	0.157	0.027
		nRMSE	0.108	0.169	0.198	0.184	0.120	0.206
Meta layer	Training	r^2	0.841	0.803	0.856	0.754	0.874	0.728
		RMSE	0.375	0.044	0.926	0.303	0.148	0.021
		nRMSE	0.098	0.101	0.091	0.111	0.084	0.126
	Testing	r^2	0.895	0.819	0.505	0.470	0.817	0.560
		RMSE	0.350	0.037	1.441	0.399	0.150	0.027
		nRMSE	0.108	0.172	0.196	0.189	0.115	0.206

* Bold font for the r^2 that is higher in the meta layer than the base layer of a nutrient.

Table 4.7: Prediction performance of the stacked multivariate regression at the base and meta layers after the feature selection using MRMR of the micro-nutrient at the dried mode

			Mn	B	Zn	Fe	Cu	Al
Base layer	Training	r^2	0.753	0.738	0.248	0.134	0.296	0.472
		RMSE	163.547	19.697	32.199	394.035	5.135	173.474
		nRMSE	0.065	0.097	0.120	0.092	0.103	0.113
	Testing	r^2	0.696	0.401	0.296	0.249	0.205	0.379
		RMSE	182.213	24.116	28.757	635.056	5.372	455.018
		nRMSE	0.135	0.215	0.184	0.192	0.202	0.192
Meta layer	Training	r^2	0.738	0.418	0.148	0.306	0.472	0.738
		RMSE	19.692	28.215	389.940	5.112	176.252	19.692
		nRMSE	0.064	0.097	0.105	0.091	0.098	0.121
	Testing	r^2	0.696	0.401	0.523	0.300	0.197	0.379
		RMSE	181.188	24.125	24.293	623.223	5.382	463.651
		nRMSE	0.134	0.196	0.155	0.188	0.203	0.195

* Bold font for the r^2 that is higher in the meta layer than the base layer of a nutrient.

Table 4.8: Prediction performance of the stacked multivariate regression at the base and meta layers after the MRMR method of the macro-nutrients at the fresh mode

			N	P	K	Ca	Mg	S
Base layer	Training	r^2	0.775	0.751	0.782	0.599	0.789	0.657
		RMSE	0.446	0.050	1.140	0.387	0.192	0.024
		nRMSE	0.120	0.108	0.114	0.135	0.101	0.133
	Testing	r^2	0.706	0.754	0.571	0.395	0.789	0.513
		RMSE	0.496	0.038	1.376	0.419	0.167	0.026
		nRMSE	0.181	0.210	0.165	0.199	0.117	0.180
Meta layer	Training	r^2	0.775	0.752	0.784	0.599	0.790	0.657
		RMSE	0.445	0.050	1.133	0.398	0.191	0.023
		nRMSE	0.117	0.107	0.114	0.134	0.101	0.138
	Testing	r^2	0.706	0.759	0.559	0.395	0.789	0.513
		RMSE	0.496	0.038	1.421	0.404	0.172	0.026
		nRMSE	0.168	0.211	0.165	0.201	0.118	0.176

* Bold font for the r^2 that is higher in the meta layer than the base layer of a nutrient.

Table 4.9: Prediction performance of the stacked multivariate regression at the base and meta layers after the MRMR method of the micro-nutrients at the fresh mode

		Mn	B	Zn	Fe	Cu	Al	
Base layer	Training	r^2	0.345	0.681	0.217	0.098	0.107	0.526
		RMSE	265.918	21.760	32.856	401.245	5.769	164.476
		nRMSE	0.100	0.106	0.122	0.093	0.105	0.109
	Testing	r^2	0.177	0.425	0.207	0.304	0.076	0.598
		RMSE	286.933	23.256	30.271	626.997	5.726	410.855
		nRMSE	0.247	0.142	0.191	0.193	0.201	0.172
Meta layer	Training	r^2	0.345	0.681	0.312	0.161	0.213	0.526
		RMSE	266.281	22.628	30.752	387.482	5.418	167.658
		nRMSE	0.101	0.103	0.114	0.090	0.105	0.111
	Testing	r^2	0.177	0.425	0.440	0.596	0.186	0.598
		RMSE	286.880	21.351	25.984	578.183	5.373	425.277
		nRMSE	0.242	0.142	0.169	0.177	0.202	0.178

* Bold font for the r^2 that is higher in the meta layer than the base layer of a nutrient.

4.4.3. Assessing the performance gains from adding nonlinear models

Three different nonlinear models (RF, SVR, and XGB) were assessed at the meta layer to estimate nutrients in correlation to other nutrients of both modes. MRMR was also implemented before applying Lasso at the base layer. As the *joinet*¹⁰ package only provides implementations of Lasso and ridge regression models, a script was built using R packages *randomForest*, *e1071* and *xgboost* to incorporate the nonlinear models into the stacked regression pipeline (Liaw and Wiener, 2002; Meyer et al., 2014; Chen and Guestrin, 2016).

Tables 4.10 and 4.11 present the performance assessment measures after implementing those nonlinear models on the cross-validation datasets. The results show that the pipeline composed of MRMR feature selection, Lasso model at the base layer, and

¹⁰ The "joinet" package in R software offers Lasso and Ridge regression methods only, as of April 2023. <https://cran.r-project.org/web/packages/joinet/index.html>

XGB model at the meta layer resulted in the highest performance for all nutrients, when compared with similar pipelines using meta layer models based on RF and SVR. Tables 4.12 and 4.13 show that applying those pipelines improved the estimation performance of all nutrients of the testing dataset except for S. For example, the results show that the estimation improved from 0.817 for Mg to 0.855 by XGB, 0.836, and 0.871 by SVR in the dried mode. Likewise, the estimation performance for Cu improved from 0.197 to 0.557 by XGB, 0.444 by RF, and 0.532 by SVR. This combination of modeling had also a substantial impact on the estimation performance of the nutrients in the fresh mode. For example, the estimation of Zn is improved by XGB from 0.440 to 0.729, Fe (from 0.596 to 0.905), and Al (from 0.598 to 0.870).

Table 4.10: Prediction performance of using nonlinear models at the meta layer of the cross-validation dataset at dried mode

	RF			SVR			XGB		
	r^2	RMSE	nRMSE	r^2	RMSE	nRMSE	r^2	RMSE	nRMSE
N	0.976	0.158	0.041	0.925	0.262	0.070	0.997	0.049	0.013
P	0.969	0.019	0.041	0.857	0.039	0.086	0.999	0.004	0.009
K	0.977	0.397	0.039	0.922	0.694	0.069	0.997	0.141	0.014
Ca	0.963	0.133	0.049	0.854	0.242	0.088	0.995	0.044	0.016
Mg	0.978	0.067	0.038	0.936	0.109	0.062	0.999	0.014	0.008
S	0.959	0.009	0.053	0.824	0.017	0.100	0.992	0.004	0.024
Mn	0.959	87.542	0.035	0.723	180.19	0.071	0.994	25.3	0.010
B	0.95	9.194	0.045	0.794	18.233	0.088	0.998	1.767	0.008
Zn	0.926	12.81	0.048	0.599	26.646	0.100	0.993	3.263	0.012
Fe	0.83	217.296	0.048	0.318	360.457	0.084	0.997	22.496	0.005
Cu	0.894	2.432	0.047	0.542	4.394	0.086	0.993	0.546	0.011
Al	0.934	72.385	0.050	0.65	146.456	0.097	0.996	15.37	0.010

* Bold font for the r^2 that is higher in the meta layer between the three nonlinear models of each mode.

Table 4.11: Prediction performance of using nonlinear models at the meta layer of the cross-validation dataset at fresh mode

	RF			SVR			XGB		
	r^2	RMSE	nRMSE	r^2	RMSE	nRMSE	r^2	RMSE	nRMSE
N	0.964	0.193	0.051	0.87	0.343	0.087	0.993	0.079	0.020
P	0.969	0.019	0.039	0.88	0.036	0.080	0.998	0.004	0.009
K	0.967	0.472	0.048	0.865	0.898	0.089	0.997	0.144	0.015
Ca	0.933	0.18	0.062	0.703	0.34	0.114	0.994	0.049	0.018
Mg	0.97	0.077	0.042	0.873	0.153	0.080	0.997	0.024	0.013
S	0.95	0.01	0.059	0.809	0.018	0.106	0.99	0.004	0.024
Mn	0.934	110.802	0.047	0.606	219.789	0.091	0.992	32.558	0.013
B	0.945	9.827	0.049	0.775	18.821	0.090	0.996	2.282	0.011
Zn	0.931	13.039	0.051	0.516	27.569	0.103	0.989	4.213	0.016
Fe	0.898	192.705	0.054	0.219	380.468	0.088	0.997	25	0.006
Cu	0.86	2.791	0.055	0.381	4.989	0.096	0.990	0.704	0.014
Al	0.925	76.392	0.055	0.606	155.835	0.104	0.995	17.516	0.012

* Bold font for the r^2 that is higher in the meta layer between the three nonlinear models of each mode.

Table 4.12: Prediction performance of using nonlinear models at the meta layer of the testing dataset at dried mode

	RF			SVR			XGB		
	r^2	RMSE	nRMSE	r^2	RMSE	nRMSE	r^2	RMSE	nRMSE
N	0.903	0.332	0.101	0.894	0.338	0.098	0.897	0.331	0.101
P	0.875	0.033	0.157	0.865	0.032	0.148	0.834	0.034	0.176
K	0.525	1.434	0.197	0.609	1.31	0.174	0.519	1.465	0.209
Ca	0.541	0.341	0.163	0.61	0.312	0.148	0.477	0.366	0.167
Mg	0.836	0.143	0.109	0.871	0.127	0.095	0.855	0.139	0.114
S	0.592	0.025	0.192	0.592	0.025	0.185	0.527	0.028	0.200
Mn	0.769	170.728	0.127	0.735	168.919	0.126	0.760	160.295	0.116
B	0.358	21.038	0.196	0.444	19.833	0.176	0.406	20.988	0.207
Zn	0.733	19.583	0.130	0.421	26.524	0.167	0.776	28.149	0.168
Fe	0.827	516.212	0.155	0.542	612.554	0.187	0.831	452.584	0.115
Cu	0.444	4.566	0.169	0.532	4.636	0.173	0.577	4.109	0.167
Al	0.833	373.664	0.161	0.414	464.676	0.194	0.926	298.149	0.116

Table 4.13: Prediction performance of using nonlinear models at the meta layer of the testing dataset at fresh mode

	RF			SVR			XGB		
	r^2	RMSE	nRMSE	r^2	RMSE	nRMSE	r^2	RMSE	nRMSE
N	0.771	0.438	0.161	0.817	0.391	0.125	0.765	0.456	0.176
P	0.883	0.028	0.192	0.829	0.031	0.175	0.716	0.049	0.213
K	0.542	1.49	0.157	0.577	1.45	0.172	0.56	1.562	0.162
Ca	0.505	0.38	0.196	0.473	0.383	0.189	0.568	0.417	0.209
Mg	0.755	0.175	0.127	0.736	0.18	0.120	0.755	0.173	0.143
S	0.607	0.023	0.169	0.635	0.022	0.156	0.499	0.028	0.194
Mn	0.24	277.039	0.230	0.274	268.807	0.216	0.234	278.864	0.266
B	0.481	20.236	0.139	0.515	21.254	0.122	0.416	21.922	0.158
Zn	0.77	20.374	0.120	0.362	26.772	0.167	0.729	31.303	0.170
Fe	0.746	495.101	0.159	0.629	590.885	0.184	0.905	368.131	0.138
Cu	0.318	4.91	0.206	0.424	4.825	0.187	0.369	4.981	0.214
Al	0.704	374.995	0.152	0.444	452.055	0.192	0.870	275.778	0.148

4.5. TWO VS. THREE-SEASON MODEL ESTIMATION

Additional 98 data points in Season 2022 were collected (See Figure 2.1). The chemical content of each of the nutrients is presented in Table 4.14. Also, Figure 4.4 shows the nutrient concentrations of the data points taken over the three growing seasons in 2020, 2021, and 2022. Most nutrient concentrations exhibited similar trends of either degradation or increase over the growing season across the three years, with the exception of Fe, Al, and B. Such exception could refer to different fertilization schemes at different seasons especially that data collection in 2022 was expanded covering different farms at different areas. Furthermore, these trends align with the interdependence observed among nutrient concentrations, such as the negative relationship between Mg and K (Figure 4.4 (h, c)), as well as the positive relationship between Ca and Mg (Figure 4.4 (i, h)) (refer to Section 4.1 for more details).

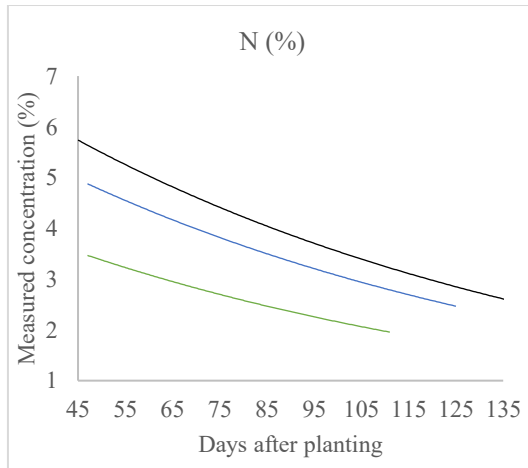
The variations in Fe, Al, and B concentrations during Season 2020 compared to the other seasons (Figure 4.4 (f, j, l)) may be attributed to differences in the fertilization

schemes implemented at various farms between the seasons. These findings are consistent with specific fertilization practices influencing the levels of these nutrients in the soil and subsequently impacting plant nutrient uptake.

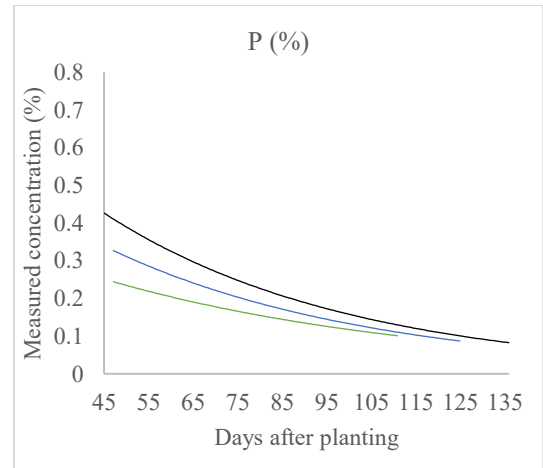
Table 4.14: Descriptive content of petiole nutritional concentrations during the entire growth season in 2022

Macro-nutrients	N	P	K	Ca	Mg	S
Min. measured	1.83	0.08	3.63	0.96	0.35	0.12
Max. measured	4.38	0.36	13.25	3.02	1.58	0.27
Min. recommended*	2.49	0.24	8	1.4	0.3	0.24
Max. recommended*	3	0.35	11	3	0.8	0.35
Mean	2.88	0.19	7.37	1.50	0.81	0.19
SD	0.59	0.07	2.11	0.42	0.27	0.03
Micro-nutrients	Mn	Zn	Fe	Cu	Al	B
Min. measured	56.10	13.30	56.00	3.70	11.10	22.80
Max. measured	516.80	131.40	264.20	61.70	156.70	61.60
Min. recommended*	60	35	50	10	ND	36
Max. recommended*	200	60	100	30	300	60
Mean	253.37	41.30	96.39	10.19	38.07	32.02
SD	124.95	26.83	44.08	10.74	31.12	6.54

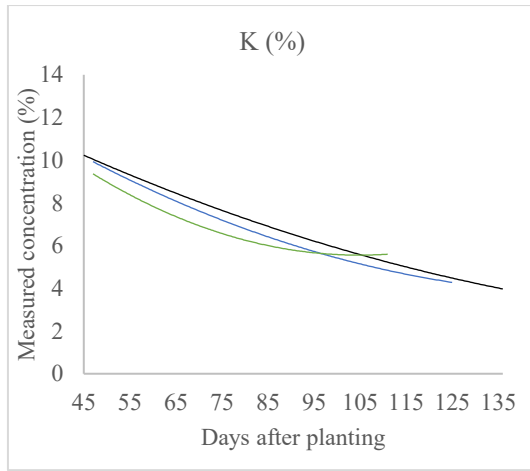
* Normal range in nutrient concentrations stated by A&L Canada Laboratories Inc. ND: not defined.



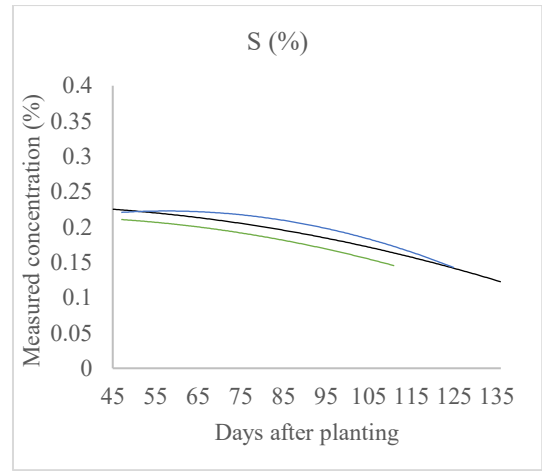
(a)



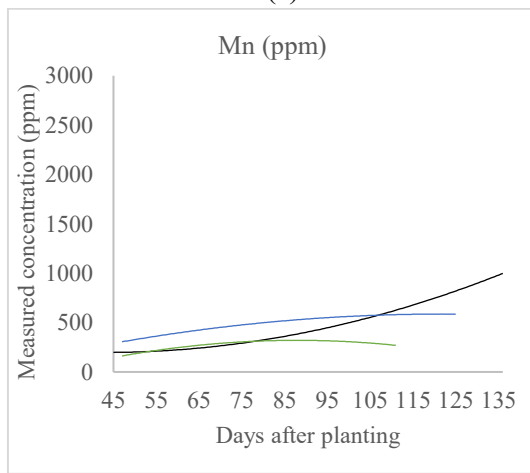
(b)



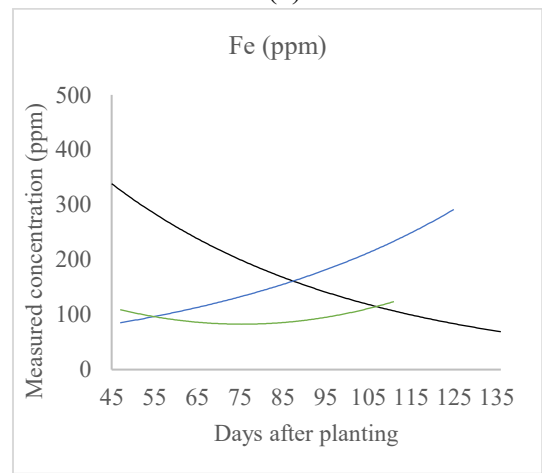
(c)



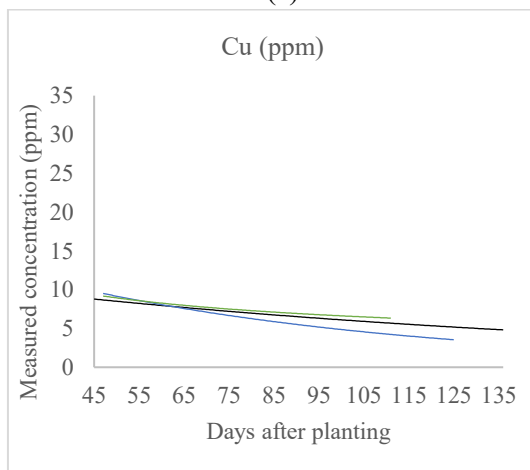
(d)



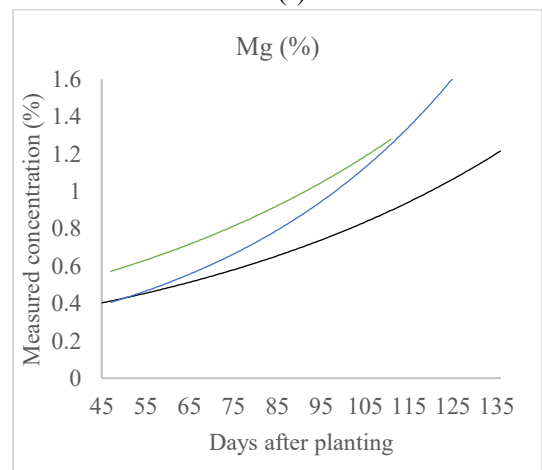
(e)



(f)



(g)



(h)

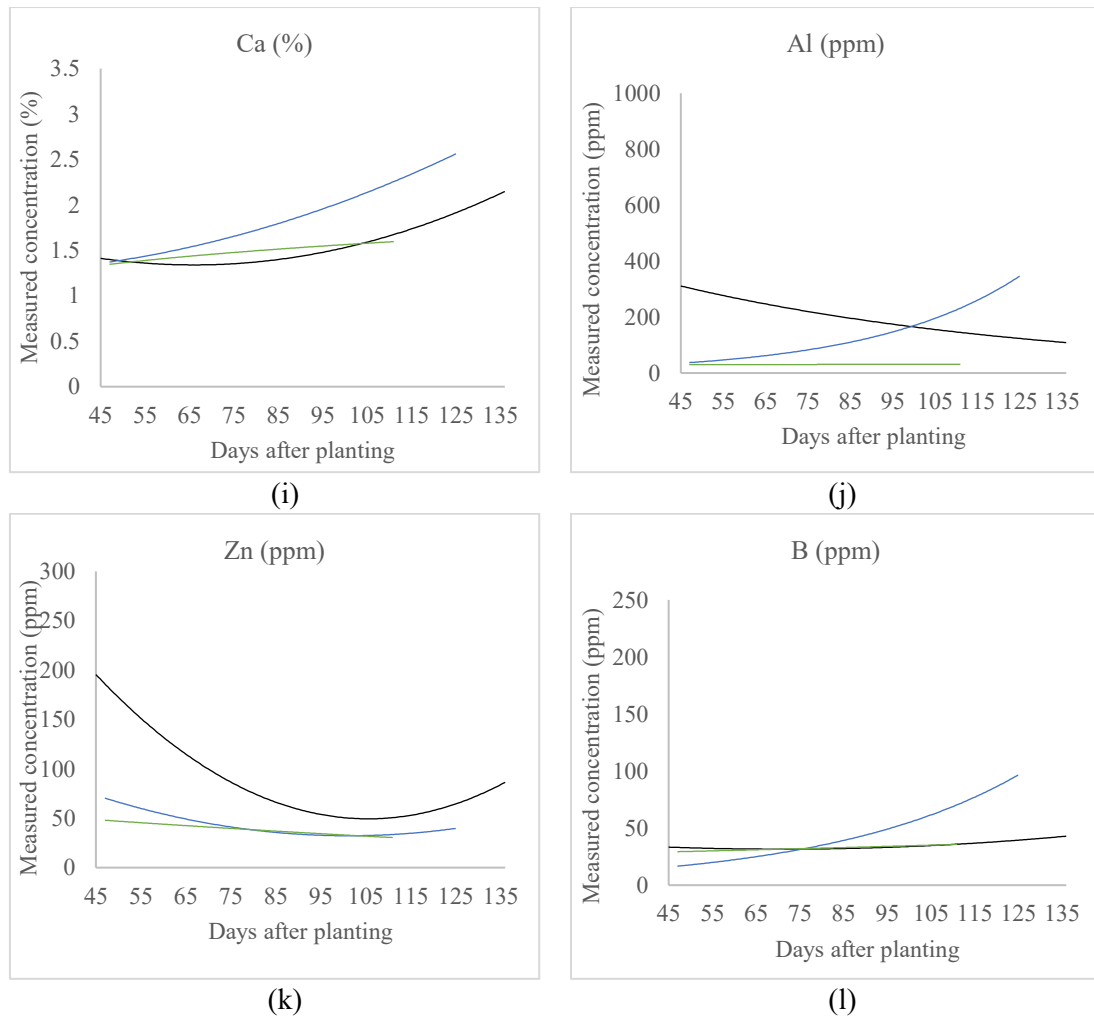


Figure 4.4: Trends of the temporal distribution for each nutrient at each farm compared during three growing season. Black, blue, and green lines refer to data points of Seasons 2020, 2021, and 2022, respectively

In conducting spectral analysis, the field spectrophotometer was employed to gather spectra from all the samples during Season 2022. The decision to collect field spectra rather than lab spectra was decided as the objective is to enable the estimation of nutrients on-the-go. Subsequently, the calibration method developed in Chapter 3 was successful to eliminate the discrepancies between lab and field spectral instruments, and thus the calibration method deployed on spectral data of 2022 before pooling them with the lab spectra of 2020 and 2021.

4.5.1. Positions of the significant wavebands

Table 4.15 presents the first four significant wavebands of each nutrient of both modes. Pooling all data points of three seasons highlighted significant wavebands not only within the Vis range, but also within the NIR as was previously noticed in single season modeling (more details in Section 2.6.4). As discussed earlier, two absorbers may coincide to such an extent that an absorption waveband appears near the sum of the frequencies of the two fundamental wavebands, and this might explain the appearance of wavebands swinging between Vis and NIR of the same nutrient at different seasons. Studying these absorption characteristics and identifying the swing bands was not further investigated under the scope of this research.

Table 4.15: Positions of the first four significant wavebands resulting from multivariate stacked regression of three seasons (2020, 2021, 2022)

	Dried mode	Fresh mode
N	428, 2204, 1308, 1940	916, 1908, 636, 708
P	788, 700, 404, 1452	404, 588, 1532, 820
K	772, 700, 1452, 404	404, 588, 1484, 540
Ca	764, 700, 404, 1500	404, 636, 1452, 804
Mg	780, 700, 404, 1420	404, 580, 1548, 796
S	700, 980, 1924, 476	940, 588, 1924, 668
Mn	764, 700, 1924, 1332	404, 636, 1924, 700
B	780, 700, 428, 1308	1044, 644, 1932, 772
Zn	404, 2204, 1308, 1932	404, 636, 1860, 708
Fe	836, 700, 1468, 404	1116, 588, 708, 1932
Cu	668, 1932, 1308, 700	1972, 652, 740, 700
Al	1100, 692, 748, 1932	1100, 596, 1932, 780

4.5.2. Performance of estimation models using the stacked multivariate regression pipeline

As per the discussion of Section (4.3.3.), the utilization of a hybrid approach involving MRMR, Lasso, and XGB resulted in the best estimation of all correlated nutrients of the cross-validation, and thus it was chosen to be the main selected pipeline for

the testing dataset over the three seasons pooled together. The data were then randomly split into an 80:20 ratio for training and testing. The performance indicators are presented in Tables (4.16, 4.17) for both modes.

The results obtained from the dried mode analysis revealed a notable improvement in nutrient estimation across three seasons as compared to using data from only two seasons. Specifically, N, K, Ca, and B exhibited enhanced model performance when trained on a larger dataset (Table 4.16). A larger dataset provides a more representation of the variations and patterns present in the spectral data, enabling the model to learn more effectively from the underlying trends. However, this was not the case for K, Mg, Zn, and Fe where the r^2 value reduced from 0.825, 0.827, 0.768, 0.941 to 0.628, 0.625, 0.616, and 0.335, respectively (Table 4.16). The reason could be attributed to the presence of different conditions across the individual seasons, which may have introduced additional variance into the data causing the model to be less capable of accommodating the diverse range present in the combined dataset. The estimation of Ca, S, Mn, and Al across seasons gave comparable performance levels to the two-season modeling.

In the fresh mode analysis, the estimation of most nutrients demonstrated a noteworthy improvement when considering data from all three seasons, which can be attributed to the larger dataset leading to enhanced model performance (Table 4.17). However, for Mg, S, Zn, and Cu, the model performance was comparatively lower when using data from all three seasons compared to two-season modeling which could possibly refer to the same reason discussed above of varying conditions. Specifically, the collection of samples from diverse farms at different locations during the third season introduced different environmental factors and conditions that might not adequately be represented in

the training dataset. Despite the reductions in estimation performance of some nutrients, these models continue to exhibit impressive accuracy for the majority, underscoring their potential for nutrient estimation in field conditions.

Table 4.16: Prediction performance of the stacked multivariate regression at the meta layer of the testing dataset at the dried mode of two vs three-season model

	Two-season model ¹¹			Three-season model		
	r^2	RMSE	nRMSE	r^2	RMSE	nRMSE
N	0.886	0.346	0.101	0.895	0.294	0.090
P	0.825	0.037	0.176	0.628	0.052	0.147
K	0.480	1.539	0.209	0.616	1.447	0.175
Ca	0.500	0.355	0.167	0.584	0.357	0.139
Mg	0.827	0.148	0.114	0.625	0.231	0.139
S	0.608	0.026	0.200	0.559	0.025	0.157
Mn	0.770	156.054	0.116	0.761	142.361	0.126
B	0.314	22.769	0.207	0.505	16.759	0.129
Zn	0.768	26.178	0.168	0.616	23.544	0.153
Fe	0.941	380.194	0.115	0.335	498.144	0.151
Cu	0.465	4.498	0.167	0.510	6.207	0.102
Al	0.932	279.960	0.116	0.867	237.821	0.099

Table 4.17: Prediction performance of the stacked multivariate regression at the meta layer of the testing dataset at the fresh mode of two vs three-season model

	Two-season model ¹²			Three-season model		
	r^2	RMSE	nRMSE	r^2	RMSE	nRMSE
N	0.601	0.620	0.176	0.667	0.501	0.155
P	0.685	0.051	0.213	0.644	0.046	0.190
K	0.533	1.554	0.162	0.677	1.372	0.129
Ca	0.208	0.527	0.209	0.418	0.465	0.182
Mg	0.680	0.238	0.143	0.573	0.247	0.149
S	0.498	0.031	0.194	0.317	0.031	0.204
Mn	0.104	269.864	0.266	0.261	328.406	0.132
B	0.311	31.644	0.158	0.541	19.564	0.150
Zn	0.675	26.027	0.170	0.338	30.635	0.199
Fe	0.730	454.277	0.138	0.868	259.802	0.079
Cu	0.256	4.919	0.214	0.055	8.466	0.297
Al	0.677	356.807	0.148	0.808	208.525	0.087

¹¹ Correspond to Table 4.12.

¹² Correspond to Table 4.13.

4.6. CONCLUSION

The results show that the implementation of the multivariate stacked regression outperformed univariate regression in estimating all nutrients in potato plants based on leaf reflectance, indicating that considering the correlated outcomes during modeling is relevant for the ability to adequately predict nutrient concentrations. The results also suggest that reducing the impacts of the high dimensionality of the feature space by combining feature reduction and model regularization is significant in improving the estimation capacity for all nutrients. The approach of using two-level modeling with Lasso at the base layer and a nonlinear model at the meta layer showed an improvement in estimating correlated outcomes. The best combination - consisting of MRMR feature selection followed by Lasso at the base layer and XGB at the meta layer - was found to provide the best predictive performance for most of the nutrients in both modes: dried and fresh. Those results show the possibility of simultaneously and effectively estimating nutrients in potato plants based on foliar spectral reflectance.

CHAPTER 5: RAPID ESTIMATION OF NUTRIENTS IN POTATO PLANTS USING LEAF REFLECTANCE SPECTRA

5.1. BACKGROUND

It is true that the deployment of spectroscopy showed a high potential to overcome the destructiveness of the chemical testing using the spectrum of the tip leaves. Nevertheless, their collection is still undergoing a destructive approach. The laboratory spectrophotometer requires around 20 tip leaves to fill the scanning cup. Further analysis applied to optimise the number of tip leaves required for fresh and dried spectral analysis using the in-field spectrophotometer. Then, based on this analysis and on the findings of the Chapters 2, 3, and 4, a sensing system was developed for non-destructive estimation of petiole nutrient in the field for the fresh scan and less destructive estimation in the dried scan.

The system works as a near real-time estimation in less a minute of N, K, Fe, and Al, and the system estimates Mn, and Al within 24 hrs. The other nutrients showed low r^2 (<0.66) in either mode (Section 4.5.2). I used a threshold value of $r^2 > 0.66$ to consider the model performance is fairly acceptable based on (Williams et al., 2019). The developed system consists of the following steps: (a) acquiring spectral measurements of a plant leaf in the field using the field spectrophotometer at either mode (handheld clip for the fresh mode, and the turntable for the dried mode, for more details see Section 3.2.1), (b) converting the leaf spectral data to txt files using computer software provided by the field spectrophotometer and storing the files in the computer memory, (c) processing the leaf spectral data every 8 nm and sending the stored spectra to Microsoft Azure as a remote storage device for subsequent computation (d) computing the nutrient values of the petiole

based on the stored leaf spectral data, and (e) delivering either near real-time results of the fresh scan, or 24hrs results of the dried scan, indicating the nutrient values of the petiole using a remote processing platform.

5.2. OPTIMIZING NUMBER OF LEAVES FOR THE IN-FIELD SPECTRAL ANALYSIS

In order to optimise the number of tip leaves required, the spectral analysis (fresh and dried) was taken at a smaller number of leaves for 45 data points from open field farms in NB using the in-field spectrophotometer vs. the laboratory one. Each data point contained 20 tip leaves. Figure 5.1 shows the steps taken for the sample collection and spectral analysis. The fresh spectral scan was initially performed leaf by leaf, and finally by averaging the spectral measurements of the 20 tip leaves to be compared with the lab spectral measurements. After the fresh spectral scan, 20 tip leaves was split in half, and then dried at 55 - 60 °C for 16 - 24 hours using a turntable of the ASD spectrophotometer. Each half was scanned and then combined and scanned as a whole. The thickness of ground powder in the petri dish was around 3 cm to ensure a complete diffuse reflectance. This follows the recommendation given by Williams et al. (2019) for wheat and seeds of comparable size.

F-test and *t*-test analyses were applied to evaluate the spectra of fresh leaves, comparing individual leaves against average groups of 5, 10, and 20 leaves. These statistical tests were also conducted on the spectra of dried leaves for each half and then for the combined of 20 dried leaves. The *F*-test was used to test the homogeneity in variances between spectra, whereas the *t*-test was performed to approve the null hypothesis (H_0) of the equal means between the averaged spectra. A *P*-value of 0.05 was used to determine the statistical significance. For the fresh leaves, both analyses were implemented

over the spectra of individual tip leaves. For the dried leaves, the analyses were done over the averaged spectra of ground 20 leaves as well as half of their quantity.

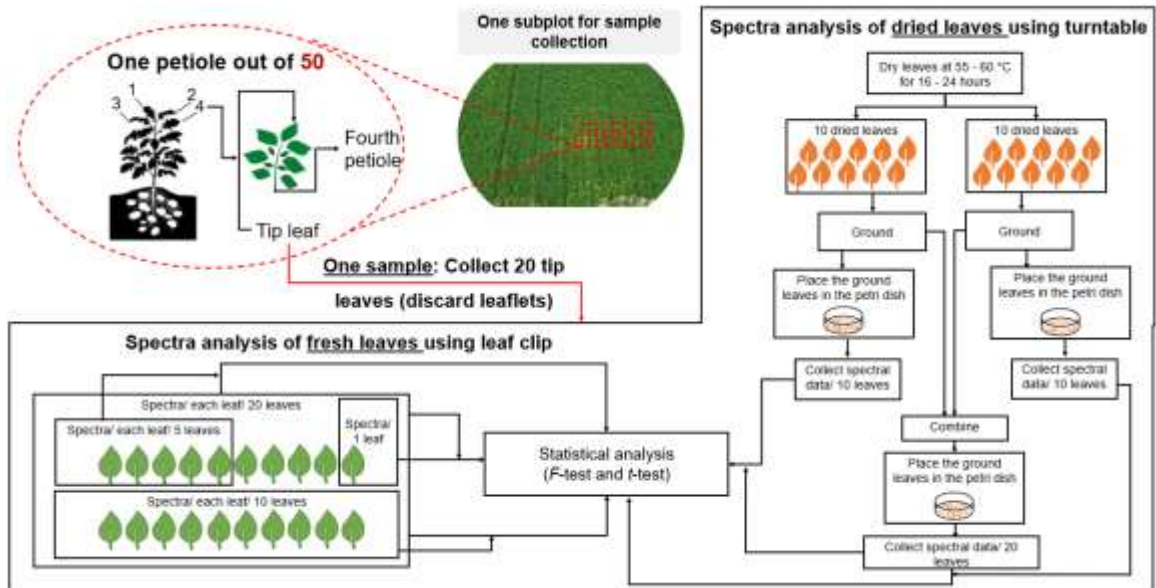
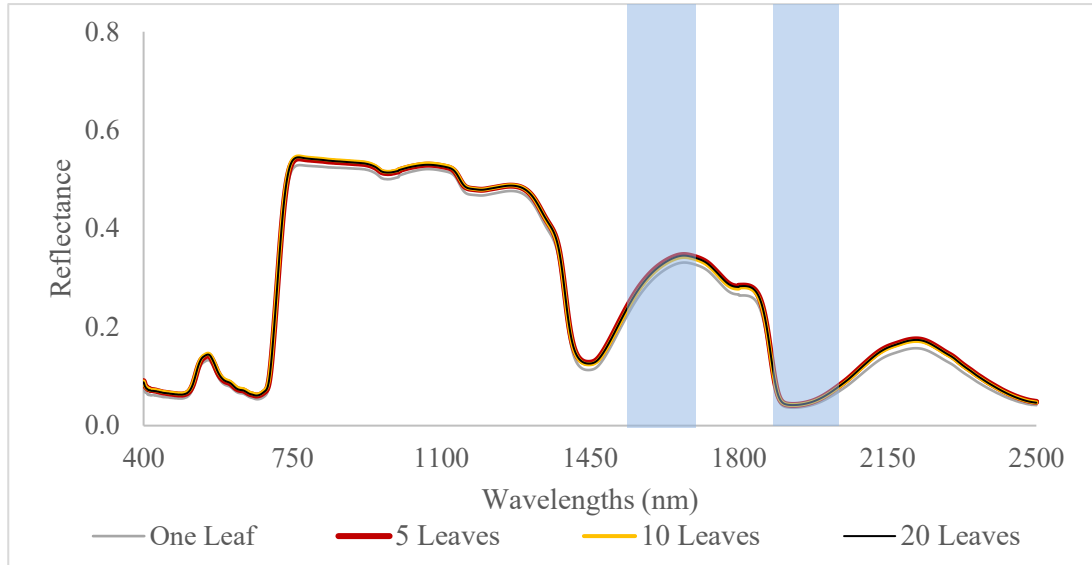


Figure 5.1: Steps of sample collection, and spectral analysis for one data point

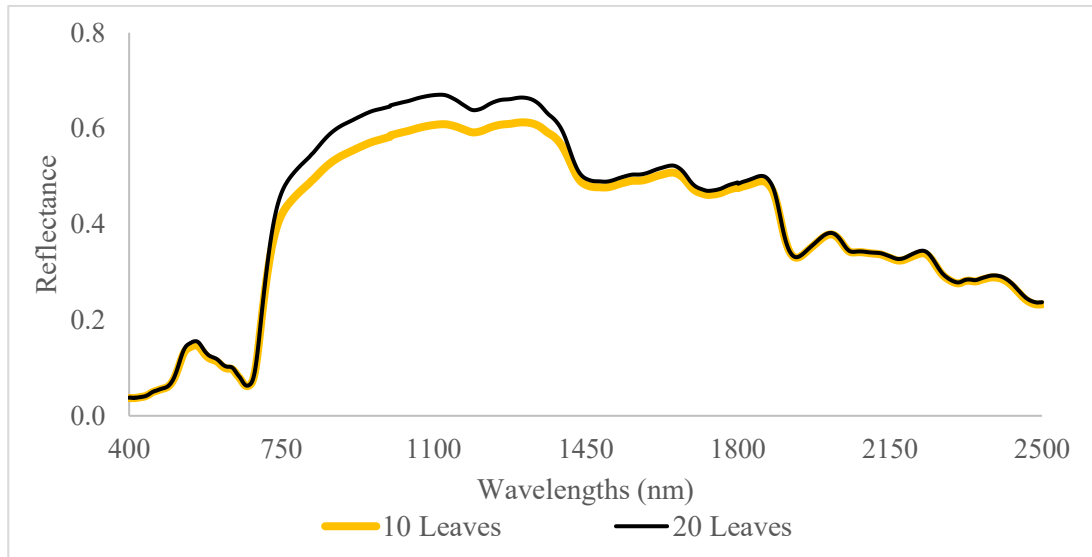
The spectral measurements in Figure 5.2 show that each spectrum of the 45 data points exhibited similar patterns for both fresh and dried leaves, hence, the patterns described in the figure are for one data point. Generally, the patterns visually correspond to no significant variance in reflectance taken within the 20 tip leaves in fresh and dried leaves. The peaks within the Vis and NIR of each fresh leaf were similar, and the low reflectance within the wavelength regions (1350–1480 and 1780–2032 nm) refers to the water absorbance (Figure 5.2a). Likewise, the reflectance peaks of the dried leaves were at the regions shown in Figure (5.2b), although the plateau was higher between 750 - 1400 nm when the spectra were taken over the 20 leaves than the reflectance measurements taken from the 10 ground leaves. This would refer to the thickness of the 10 ground dried leaves below 3 cm, which may have caused incomplete diffuse reflectance.

The F -test of fresh leaves taken individually, every 5, 10, and 20 leaves exposed a P -value > 0.05 (min. 0.53 of Table 5., Figure 5.2a), indicating no significant difference between the variances of the spectral sets of the leaves. The P -value at t -test of >0.05 (min 0.68 of Table 5.1, Figure 5.2a) also concludes that the H_0 will not be rejected. Such analysis of the spectral measurements of the fresh leaves showed equality of means. Thus, one data point could be given by 1 tip leaf only.

For the dried leaves, the P -value at F -test ranged between 0.32 and 0.40 which is greater than the significance level of 0.05, meaning no significant difference between the variances of the two sets of 10 ground leaves as show in Figure 5.2b. Consequently, the t -test was calculated based on the equality of the two variances. The P -value at t -test was 0.59 and 0.68 for each half of the 20 leaves which is greater than 0.05 and thus, the H_0 of the equality of mean for the two halves will not be rejected. Thus, one data point is possible to be obtained by as minimum as 10 dried tip leaves.



(a)



(b)

Figure 5.2: Spectral measurements of one data point. (a) fresh leaves, (b) dried leaves. Shaded boxes highlight the water absorbance ranges

Table 5.1: Results of the statistical analyses of one data point

Testing mode	F -test (-)	P -value at F -test (-)	t -statistic (-)	t -critical (-)	P -value at t -test (-)
One fresh leaf	0.82	0.53	-0.41	1.66	0.68
Five fresh leaves	1.04	0.90	0.15	1.66	0.88
Ten fresh leaves	0.96	0.89	0.18	1.66	0.86
Ten dried leaves	1.35	0.32	-0.54	1.66	0.59

5.3. STEPS OF THE DEVELOPED SENSING SYSTEM

5.3.1. Acquiring spectra measurements

The developed sensing system uses the field spectrophotometer to take the spectral measurements of potato tip leaves at either mode: fresh by using leaf clip (Figure 5.3) or dried by using Turntable (Figure 5.3). The implementation of the developed sensing system necessitates predefined settings. These settings encompass the labelling of each spectral file as either "fresh" or "dried," determined by the testing mode, and subsequently linked with a serial number. The format of each spectral file looks like: "Spectrum xxx_Dried.txt" or "Spectrum xxx_Fresh.txt.", where xxx refers to the serial number of the spectrum file.



Dried scan using a turntable

Fresh scan using a leaf clip

Figure 5.3: Collection of spectral measurement in the field of both dried and fresh leaves

5.3.2. Storing the leaf spectral data

Following the creation of each spectral file, the format provided by the field spectrophotometer is in the ".asd" format. Every file was stored in a predefined directory located on the D drive, utilizing the name "Estimation of concentrations". The Analytical Spectral Devices Inc. also provides an impeded script where the spectral data can be

converted from “.asd” to “.txt” format. The script was used to generate the “.txt” files for all the generated spectral files to ease in reading data for subsequent processing. This process is the first interface by the user to manually convert the format of the generated spectral files.

5.3.3. Sending the stored spectra to Microsoft Azure cloud

Each of the spectral files at the “.txt” format was automatically processed in an embedded script developed under this research to re-arrange the spectral data every 16nm starting from 404 nm. This script was installed at the same computer where the spectral files are stored. The functionality of the script operation is depicted in Figure 5.4. The same script changes the format of each file from the “.txt” to “.xlsx” format and sends each “.xlsx” to Microsoft Azure cloud, a remote cloud storage/ computing platform used for this research to store spectral data and process the computation.

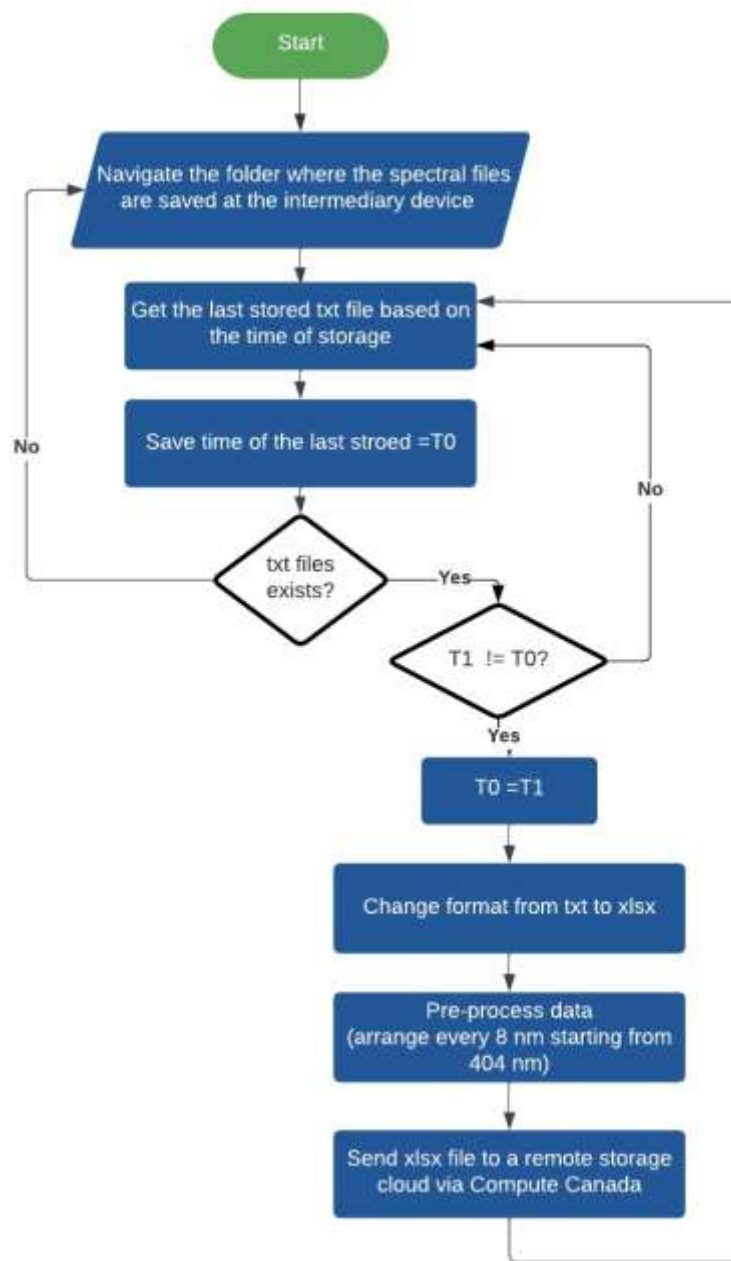


Figure 5.4: Flowchart of the spectral files pre-processing steps performed

5.3.4. Computing and delivering the nutrient values of the petiole based on the stored leaf spectral data

After the successful storage of the spectral files as “.xlsx” format, another script was developed under this research to run the script based on which mode. The script runs by a developed Azure Function App triggered by time (every 1 minute). The concept of the script is presented in Figure 5.5, and the computation is triggered if a new time stamped spectral data is uploaded to the Cloud. The script reads the file name either “Spectrum xxx_Dried.xlsx” or “Spectrum xxx_Fresh.xlsx” and then continues the computation accordingly. Based on which mode, the stored spectral file enters the calibration process discussed in Section 3.4. After the calibration of the spectra, the script takes the calibrated spectra to run the machine learning pipeline discussed in Section 4.4.3. The final predicted concentrations of petiole nutrient values are delivered to the customer via an embedded communication tool in the script that generates SMS with the results. If the script could not find the correct file name by either “Spectrum xxx_Dried.xlsx” or “Spectrum xxx_Fresh.xlsx”, an SMS is sent saying “File name is not correct, please rename”. The script will run upon on the last stored file by checking the storage date/time, otherwise, the script stays tuned till a new spectral file comes up at the remote storage of Microsoft Azure.

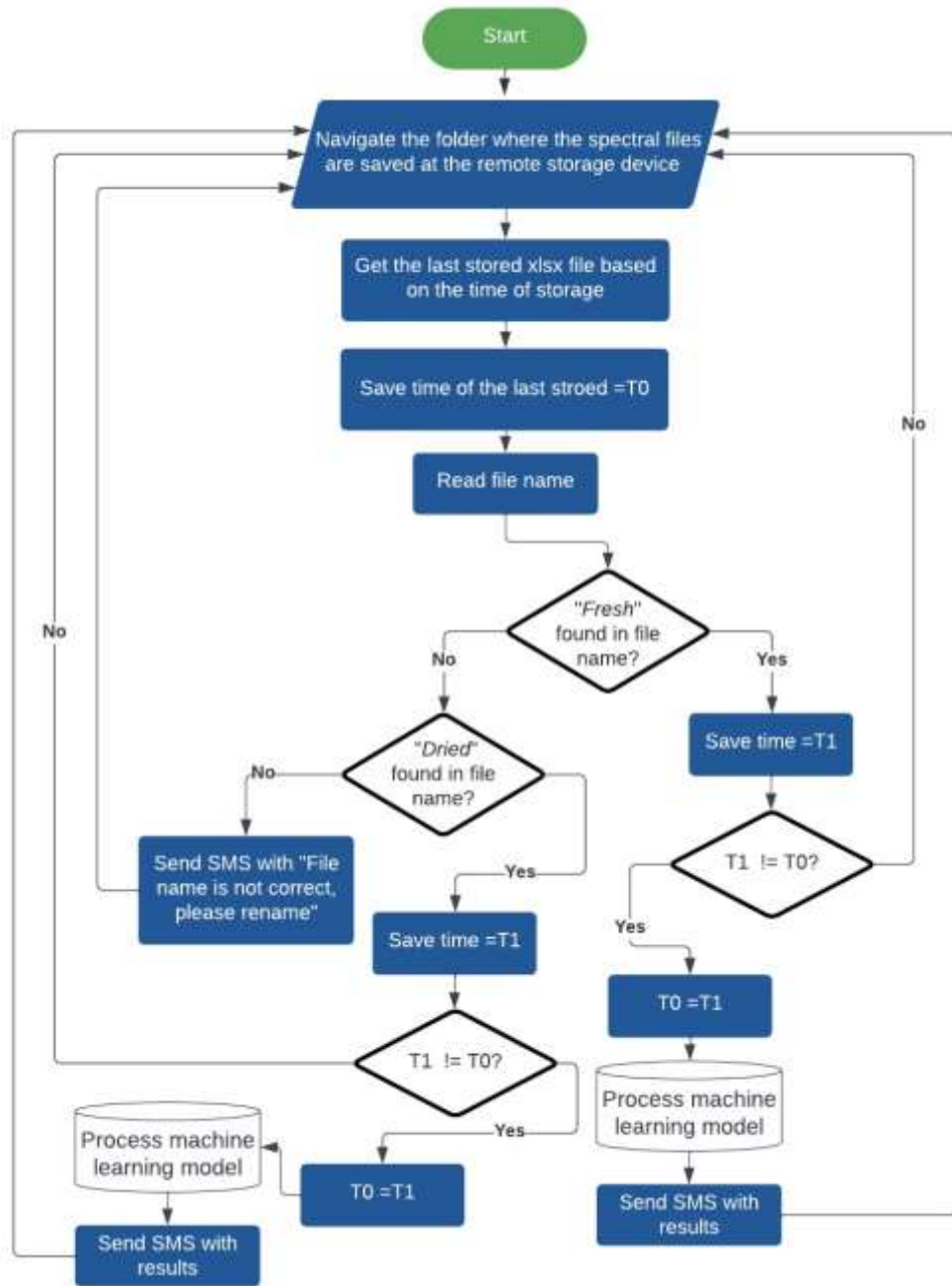


Figure 5.5: Flowchart of the developed script to be run at the remote processing cloud

5.4. GENERAL OVERVIEW OF THE DEVELOPED SENSING SYSTEM COMPONENTS

The preceding sections of this research provide detailed explanations of the various phases encompassed within the developed sensing system. These sections outline the

systematic progression from initial concept to the final implementation of a sensing system for rapid estimation of petiole chemical content in potato plants. Figure 5.6 illustrates the processes undertaken in utilizing the field spectrophotometer for the fresh scan. This schematic presentation provides a holistic view of the systematic methodology, demonstrating how spectral data of fresh potato leaves converged to predict near real-time petiole chemical content in less a minute.

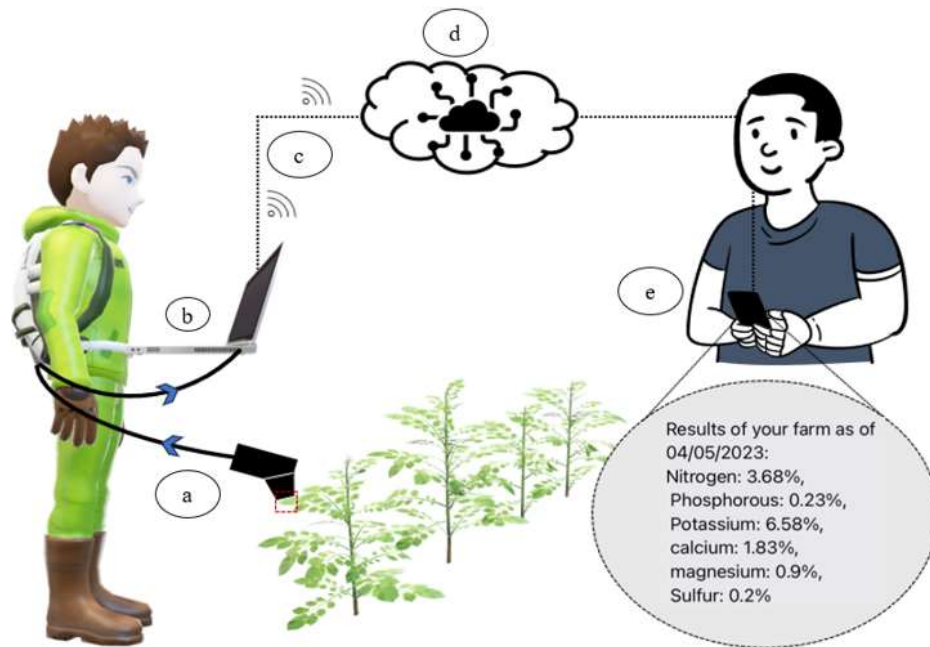


Figure 5.6: Schematic of the proposed sensing system for non-destructively determining petiole nutrient values in near real-time from plant leaves in field, (a) acquiring spectral measurements of a plant leaf (tip leaf of 4th petiole), (b) storing the leaf spectral data in a computer memory, (c) sending the stored spectra to a remote storage device (d) computing the nutrient values of the petiole based on the stored leaf spectral data, and (e) delivering near-time results indicating the nutrient values of the petiole via a remote processing platform

5.5. CONCLUSION

A non-destructive system has been successfully developed for the near real-time determination of petiole nutrient concentrations in less than a minute by analyzing the reflectance spectra of fresh tip leaf, employing a portable spectrophotometer. Additionally, a less destructive system has been developed for the dried leaves to give results in 24 hrs, by also employing a field spectrophotometer. Both systems comprised of acquiring spectral measurements of a plant leaf in the field, pre-processing spectral data, computing the nutrient values, and delivering results using a remote processing platform. This system has fulfilled the research goal, which aimed to enhance the existing sampling method and to expedite the analysis time.

CHAPTER 6: CONCLUSION AND FUTURE PLANS

The findings from this research demonstrate a significant potential for utilizing leaf reflectance spectra to assess the nutrient levels in potato plants. The proposed sensing system put forth in this research has the capability to speed up acquiring the nutritional status of potato plants. For nutrients for which models showed a predictive ability ($r^2 > 0.66$) in the fresh mode of three seasons such as for N, K, Fe, and Al, the proposed sensing system can provide near real-time results in less than a minute instead of the current waiting period of 10-12 days. Similarly, nutrients for which models displayed a predictive performance with $r^2 > 0.66$ in the dried mode (such as Mn, and Al) can be reliably assessed within 24 hours, and the remaining nutrients showed low prediction with $r^2 < 0.66$ in either mode. This suggests that the proposed sensing system can introduce a shift in the current sampling methodology, transitioning from a destructive sampling protocol to a non-destructive one when focusing on fresh leaves. Alternatively, for dried leaves, the proposed method involves a less destructive sampling procedure, reducing the number of required petioles from 40-50 to only 10 leaves. This innovative approach promises to be a valuable tool for farmers, enabling them to efficiently apply the necessary fertilizers and maintain their crops in a timely manner. While the current research showcases promising results, it also acknowledges the potential for continuous improvements in the developed sensing solution, as discussed in the subsequent sections. By addressing these aspects, the proposed system be optimized even further, maximizing its utility and benefits for agricultural applications.

6.1. IMPORTANCE OF INCORPORATING VARIOUS FERTILIZATION SCHEMES

The goal of collecting data across seasons was to ensure that the developed datasets represented real scenarios encompassing a wide range of concentrations, enabling the developed models to learn patterns and relationships between petiole chemical contents and leaf spectra that are applicable across a diverse set of concentrations. This wide sampling range of concentrations enhances the robustness of the developed models.

However, open-field farming presents complex situations where nutrients can interact with each other, influencing the overall nutrient status of plants. Conducting independent experiments to test each nutrient individually would be an ideal method to understand the specific effects of each nutrient. Consequently, it becomes necessary to expand the scope of indoor growing to implement various fertilization schemes.

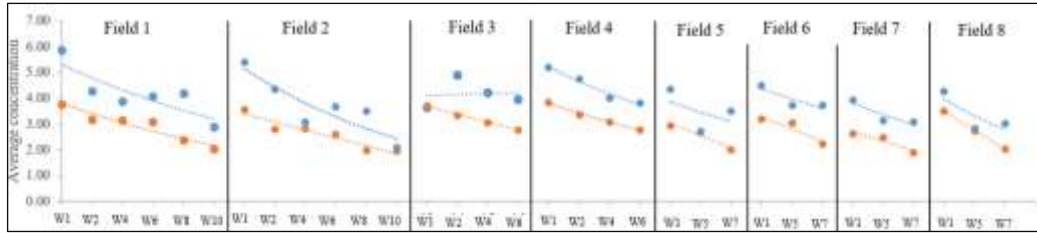
6.2. INFLUENCE OUTLIERS AT THE GROUND TRUTH LEVEL

The reliability of the models used in this research, which involved the prediction of nutrient levels using ground-based spectrophotometers, heavily relies on accurate ground truth values. Standard protocols were followed for sample collection, handling, and shipping, as per the practices employed by potato growers. The laboratory analysis of the samples adhered to reliable and approved methods outlined by AOAC.

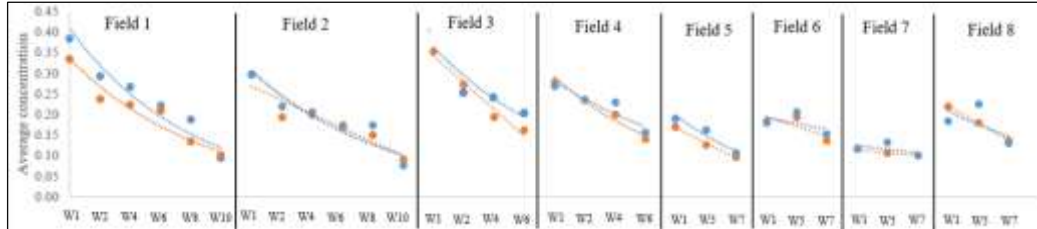
The accuracy of ground truth is generally assumed to be precise. Knowing that, the ground truth is practically impossible to be completely accurate due to the differences between the time the sample was collected and analyzed. The unreliable ground truth is not limited to the time gap between sample collection and analysis, but also it may occur during the analysis of the sample chemical content. The effects of errors in ground truth values on modeling accuracy have been barely assessed by other researchers. A study by Carlotto

M.J. (2009) assessed the errors in ground truth on the thematic accuracy to classify land cover diversity. Their study indicated that the evaluation of a larger scale can be more representative of the true relative performance of the classifier for the diversity of land cover even using many inaccurate ground truth points. However, the ground truth accuracy is more sensitive and required when the scale of the sample that is intended to be modeled is small, such as estimating nutrient levels in plants.

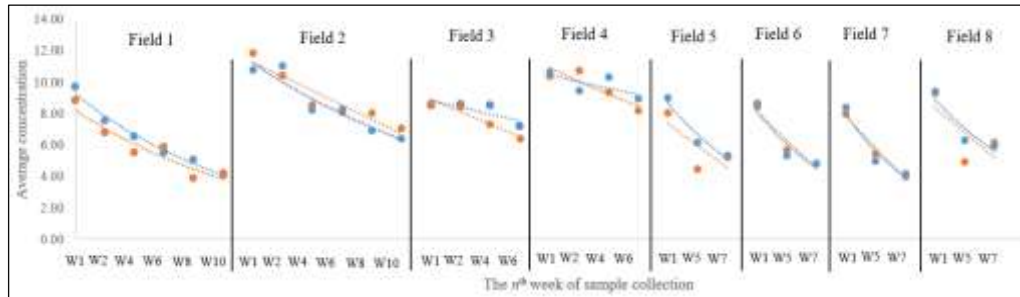
Accuracy in these situations is required because of the later fertilization decisions that will be taken. Under this research, a preliminary evaluation was performed on the ground truth accuracy by sending the samples taken from same spot and time to two different laboratories for chemical testing. Amongst 13 macro and micronutrients, only NPK concentrations gave consistent trends between the two labs (Figure 6.1 a, b, c). However, the differences between the concentrations of other nutrients such as in Mn and B (Figure 6.1 d, e) raise a red flag of the modeling accuracy that may fail to validate the model performance. This could refer to that data might have outliers. However, a methodology to find and verify such outliers was not implemented under this research which might lead to potential distortions the learned relationship between target variables and predictive features. Thus, increasing the dataset might give a space to detect the outliers and remove them.



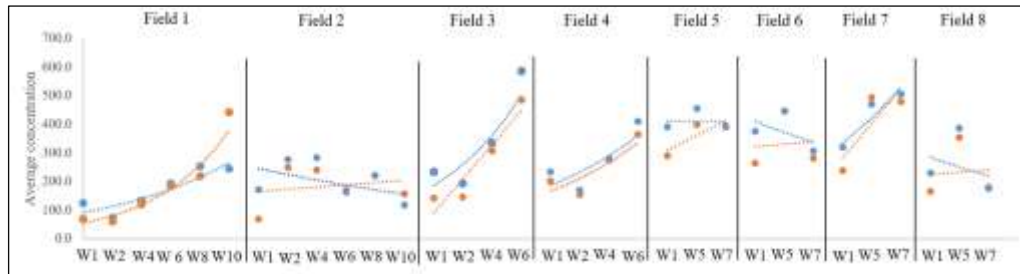
a. N



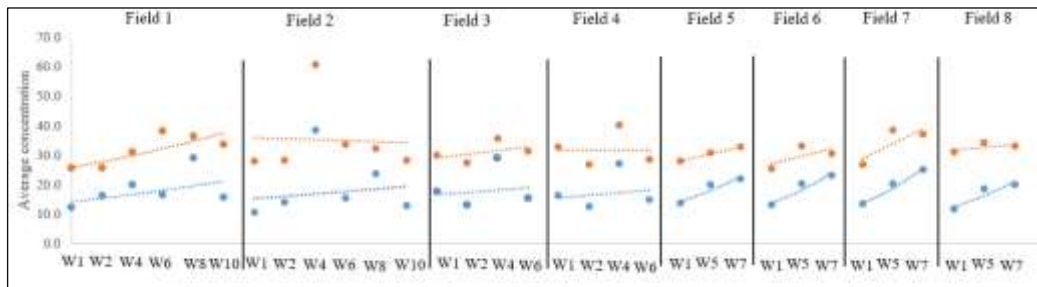
b. P



c. K



d. Mn



e. B

Figure 6.1: Differences in chemical concentrations between two labs of same samples of petioles at week (w). Blue refers to the results given by A&L Canada Laboratories, and the orange concentrations given by AgSource Laboratories. Sampling period of each field started from 45 days after planting till 135 days

6.3. INFLUENCE OF BOUNCING REFLECTANCE SPECTRA ON WAVELENGTH RANGE REDUCTION

The findings of this research emphasized the significance of SWIR for accurately estimating various nutrients, including N and P (Section 2.7.2). The scientific rationale behind this is attributed to the presence of functional groups in the SWIR region, which can correspond to specific nutrients. For example, numerous studies have demonstrated that chlorophyll, predominantly found in the Vis spectrum, is an important indicator of N content in leaf tissues. Additionally, protein content serves as another crucial indicator for estimating the total N content in leaves. These bouncing wavebands from Vis to SWIR, as explained earlier, could be attributed to the phenomenon of absorptive bands and their overtones bouncing over a combination of wavebands (Figure 2.9).

To leverage this phenomenon, the employed models could select significant wavebands based on a scoring applied to pairs of wavebands. The highest scoring pair is chosen, even if both wavebands individually exhibit significance. By employing these models, the required range of wavebands can be narrowed down, thereby reducing the cost associated with the spectrophotometer. However, it should be noted that this research did not conduct an in-depth analysis to determine at which range of wavebands shorter than 2500 nm can be optimal for estimating a specific nutrient.

In summary, this research underscores the importance of SWIR in estimating various nutrients, offering valuable information for specific nutrient detection. By selecting significant wavebands through scoring, the range of required wavebands can potentially be reduced, leading to a cost-effective selection of the required spectrophotometry.

Nevertheless, further investigations are needed to determine the wavelength range for estimating individual nutrients within a shorter range.

6.4. COST COMPARISON BETWEEN CHEMICAL TESTING VERSUS SENSING

It is important to acknowledge that the following cost estimation is preliminary and subject to various factors, such as the consultant pricing strategy, and the number of farmers served. A comprehensive business plan and analysis would be necessary to further validate and refine these projections. Additionally, potential operational costs, maintenance/ depreciation expenses, and other relevant factors should be considered in the overall financial assessment of the proposed sensing system.

Comparatively, the purchase cost of the spectrophotometer (around 100,000 CAD) seems significantly more expensive than the current sampling method. However, it is important to note that the proposed sensing system does not entail individual growers purchasing a spectrophotometer. Instead, the proposal suggests hiring a consultant who can carry out the measurements and provide results to the farmers. Considering the consultant is to operate during the summer season of potato growing, the following parameters are assumed to be the target number of customer in on season:

1. Three months of collecting samples per one year (12 weeks/ year).
2. 5 working days per week (60 days/ year).
3. 6 farms target (sample) per day (360 samples/ year).

Table 6.1 presents a rough estimate of the preliminary costs by following the current petiole chemical testing versus the proposed sensing system. The labor cost does not include only sample collection, but also transportation to the sample collection field. The preliminary analysis shows that if the consultant offers their service at a rate of 50 CAD per sample,

the revenue generated during the three-month operation would amount to 18,000 CAD in the first year not including the spectral instrument price. This means that the consultant could potentially recover the initial investment in the spectrophotometer within four years of operation from the revenue minus the cloud use and the depreciation cost.

Table 6.1: Preliminary expected costs in a year to operate the proposed sensing system by a consultant vs the current sampling cost based on 360 samples in season

Item	Unit cost (CAD)	Total cost (CAD)
Petiole chemical testing costs		
Labor cost for sampling (~ 60 min/ sample)	100/sample	36,000
Chemical testing	45/sample	16,200
Shipment (every week)	23/week	276
Total cost in one season		52,476
Proposed sensing system		
A portable spectrophotometer (400 – 2500 nm)	100,000	100,000
Depreciation of the spectral instrument (7% per year, 14 yrs. lifetime)	7,000	7,000
Labor cost for sampling (~ 30 min/ sample)	50/sample	18,000
Cloud use	200/year	200
Total cost in one season not incl. the spectral instrument		25,200

6.5. POTENTIAL FOR ENHANCEMENTS AT PROPOSED SENSING SYSTEM FRONT-END

The proposed solution involves the utilization of a portable spectrophotometer, with the spectra data being transmitted to a remote server for processing. There are several potential enhancements that could be considered to improve the front-end of the solution.

These include:

1. Data visualization: another enhancement could involve incorporating a data visualization component into the front-end. By visualizing the output of the chemical results on a monitor or display interface connected to user database containing information such as farm locations and sample collection times. Users

could be provided with a graphical representation of each nutrient in reference to the normal range that aids in understanding the results in a more comprehensive manner.

2. Integration of GPS information: a further enhancement opportunity lies in connecting the chemical results with the corresponding GPS information collected during the spectral data acquisition process. This integration would facilitate the identification of spatial patterns or trends within the farm, enabling users to gain insights into the geospatial distribution of the analyzed concentrations, and potentially assisting in decision-making processes related to farm fertilization management.
3. Contactless sensor: a further enhancement in the future to develop a rapid estimation of nutrients using a contactless sensor that could be mounted on a tractor.

References

- Abbasian, A., Ahmadi, A., Abbasi, A.R., and Darvishi, B. (2018). Effect of various phosphorus and calcium concentrations on potato seed tuber production. *Journal of Plant Nutrition*, 41 (14): 1765–1777. <https://doi.org/10.1080/01904167.2018.1454955>.
- Abukmeil, R., Al-Mallahi, A.A., and Campelo, F. (2022). New approach to estimate macro and micronutrients in potato plants based on foliar spectral reflectance. *Computers and Electronics in Agriculture*, 198: 107074. <https://doi.org/10.1016/j.compag.2022.107074>.
- Abukmeil, R., Al-Mallahi, A. (2022a) In-field scanning of the spectrum potato plant leaves for a less destructive technique to estimate nutrients. CSBE/SCGAB 2022 Annual General Meeting and Technical Conference in Charlottetown PEI, 24-27 July 2022, Submission ID: 150.
- Agriculture and Agri-Food Canada (2022). Potato Market Information Review 2021-22. <https://agriculture.canada.ca/en/sector/horticulture/reports/potato-market-information-review-2021-22#a6.1> (Accessed on July 06, 2023).
- Ahmad, W., Zia, M.H., Malhi, S.S., Niaz, A., and Saifullah, H. (2012). Boron Deficiency in Soils and Crops: A Review. Editor(s): Goyal, A. In *Crop Plant*. InTech. ISBN: 978-953-51-0527-5. <http://www.intechopen.com/books/crop-plant/boron-deficiency-in-soils-and-crops-a-review>
- Alaoui-Sossé, B., Genet, P., Vinit-Dunand, F., Toussaint, M.L., Epron, D., and Badot, P.M. (2004). Effect of copper on growth in cucumber plants (*Cucumis Sativus*) and its relationships with carbohydrate accumulation and changes in ion contents. *Plant Science*, 166 (5): 1213-1218. <https://doi.org/10.1016/j.plantsci.2003.12.032>.
- Arai, K. (2013). Nonlinear mixing model of mixed pixels in remote sensing satellite images taking into account landscape. *International Journal of Advanced Computer Science and Applications*, 4 (1). <https://doi.org/10.14569/IJACSA.2013.040115>.
- Asner, G.A., and Martin, R. (2015). Spectroscopic remote sensing of non-structural carbohydrates in forest canopies. *Remote Sensing*, 7 (4): 3526–3547. <https://doi.org/10.3390/rs70403526>.
- Bajocco, S., Ginaldi, F., Savian, F., Morelli, D., Scaglione, M., Fanchini, D., Raparelli, E., and Bregaglio, S.U.M. (2022). On the use of NDVI to estimate LAI in field crops: implementing a conversion equation library. *Remote Sensing*, 14 (15): 3554. <https://doi.org/10.3390/rs14153554>.

- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., and McBratney, A. (2010). Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends in Analytical Chemistry*, 29 (9): 1073-1081.
- Bojinov, B., Ivanov, B., and Vasileva, S. (2022). Current state and usage limitations of vegetation indices in precision agriculture. *Bulgarian Journal of Agricultural Science*, 28 (3): 387–394.
- Bohl, W.H., and Johnson, S.B. (2010). Commercial potato production in North America: the potato association of America handbook, 2nd ed. Orono, United States. https://potatoassociation.org/wp458content/uploads/2014/04/A_ProductionHandbook_Final_000.pdf (accessed on September 25, 2023)
- Botha, E. J., Zebarth, B. J., and Leblon, B. (2006). Non-destructive estimation of potato leaf chlorophyll and protein contents from hyperspectral measurements using the PROSPECT radiative transfer model. *Canadian Journal of Plant Science*, 86 (1): 279-291. <https://doi.org/10.4141/P05-017>.
- Busemeyer, L., Mentrup, D., Möller, K., Wunder, E., Alheit, K., Hahn, V., Maurer, H.P., Reif, J.C., Würschum, T., Müller, J., Rahe, F., and Ruckelshausen, A. (2013). BreedVision-A multi-sensor platform for non-destructive field-based phenotyping in plant breeding. *Sensors*, 13 (3): 2830–2847. <https://doi.org/10.3390/s130302830>.
- Carlotto, M.J. (2009). Effect of errors in ground truth on classification accuracy. *International Journal of Remote Sensing*, 30 (18): 4831-4849. <https://doi.org/10.1080/01431160802672864>.
- Carter, G.A., and Miller, R.L. (1994). Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands. *Remote Sensing of Environment*, 50 (3): 295–302. [https://doi.org/10.1016/0034-4257\(94\)90079-5](https://doi.org/10.1016/0034-4257(94)90079-5).
- Caruso, G., Palai, G., Gucci, R., and Priori, S. (2022). Remote and proximal sensing techniques for site-specific irrigation management in the olive orchard. *Applied Sciences*, 12 (3): 1309. <https://doi.org/10.3390/app12031309>.
- Caruso, G., Palai, G., Marra, F.P., and Caruso, T. (2021). High-resolution UAV imagery for field olive (*Olea Europaea* L.) phenotyping. *Horticulturae*, 7: 258. <https://doi.org/10.3390/horticulturae7080258>.
- Coliban, R.M., Marinceş, M., Hatfaludi, C., and Ivanovici, M. (2020). Linear and non-linear models for remotely-sensed hyperspectral image visualization. *Remote Sensing*, 12 (15): 2479. <https://doi.org/10.3390/rs12152479>.
- Davenport, J.R., Perry, E.M., Lang, N.S., and Stevens, R.G. (2005). Leaf spectral reflectance for nondestructive measurement of plant nutrient status. *HortTechnology*, 15 (1):31-35. <https://doi.org/10.21273/HORTTECH.15.1.0031>.

- Ding, C., and Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, 3 (2):185-205.
- Du, L., Li, Q., Li, L., Wu, Y., Zhou, F., Liu, B., Zhao, B., Li, X., Liu, Q., Kong, F., and Yuan, J. (2020). Construction of a critical nitrogen dilution curve for maize in southwest China. *Scientific Reports*, 10: 13084. <https://doi.org/10.1038/s41598-020-70065-3>.
- Dyar, M.D., Carmosino, M.L., Breves, E.A., Ozanne, M.V., Clegg, S.M., and Wiens, R.C. (2012). Comparison of partial least squares and lasso regression techniques as applied to laser-induced breakdown spectroscopy of geological samples. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 70: 51-67. <https://doi.org/10.1016/j.sab.2012.04.011>.
- Environment and Climate Change Canada, Government of Canada New Brunswick (2020, September). Water Resource Report. https://www2.gnb.ca/content/gnb/en/departments/elg/environment/content/water/content/water_quantity.html (Accessed on February 08, 2023).
- Fahmy, S.H., Hann, S.W.R., and Jiao, Y. (2010). Soils of New Brunswick the second approximation. Eastern Canada Soil and Water Conservation Centre. Technical Publication Number: NBSWCC-PRC 2010-01. https://sis.agr.gc.ca/cansis/publications/surveys/nb/nbsa/nbsa_report.pdf (Accessed on February 06, 2023)
- Feng, X., Yu, C., Chen, Y., Peng, J., Ye, L., Shen, T., Wen, H., and He, Y. (2018). Non-destructive determination of shikimic acid concentration in transgenic maize exhibiting glyphosate tolerance using chlorophyll fluorescence and hyperspectral imaging. *Frontiers in Plant Science*, 9: 468. <https://doi.org/10.3389/fpls.2018.00468>.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33 (1): 1-22. <https://doi.org/10.18637/jss.v033.i01>
- Gabriel, J.L., Zarco-Tejada, P.J., López-Herrera, P. J., Pérez-Martín, E., Alonso-Ayuso, M., and Quemada, M. (2017). Airborne and ground level sensors for monitoring nitrogen status in a maize crop. *Biosystems Engineering*, 160:124-33. <https://doi.org/10.1016/j.biosystemseng.2017.06.003>.
- Galvez-Sola, L., García-Sánchez, F., Pérez-Pérez, J.G., Gimeno, V., Navarro, J.M., Moral, R., Martínez-Nicolás, J.J., and Nieves, M. (2015). Rapid estimation of nutritional elements on citrus leaves by near infrared reflectance spectroscopy. *Frontiers in Plant Science*, 6:571. <https://doi.org/10.3389/fpls.2015.00571>.
- Glenn, D. M., and Tabb, A. (2019). Evaluation of five methods to measure normalized difference vegetation index (NDVI) in apple and citrus. *International Journal of Fruit Science*, 19 (2): 191–210. <https://doi.org/10.1080/15538362.2018.1502720>.

- Gómez, M.I., Magnitskiy, S., and Rodríguez, L.E. (2019). Critical dilution curves for nitrogen, phosphorus, and potassium in potato group Andigenum. *Agronomy Journal*, 111 (1): 419–427. <https://doi.org/10.2134/agronj2018.05.0357>.
- Gransee, A. and Führs, H. (2013). Magnesium mobility in soils as a challenge for soil and plant analysis, magnesium fertilization and root uptake under adverse growth conditions. *Plant Soil* 368: 5–21. <https://doi.org/10.1007/s11104-012-1567-y>
- Guo, P., Li, T., Gao, H., Chen, X., Cui, Y., and Huang, Y. (2021). Evaluating calibration and spectral variable selection methods for predicting three soil nutrients using Vis-NIR spectroscopy. *Remote Sensing*, 13 (19): 4000. <https://doi.org/10.3390/rs13194000>.
- Gutiérrez, S., Fernández-Navales, J., Diago, M.P., and Tardaguila, J. (2018). On-the-go hyperspectral imaging under field conditions and machine learning for the classification of grapevine varieties. *Frontiers in Plant Science*, 9: 1102. <https://doi.org/10.3389/fpls.2018.01102>.
- Hastie, T., Tibshirani, R., and Friedman, J. (2008). *The elements of statistical learning: data mining, inference, and prediction*. Second edition. New York: Springer, 2008.
- Heenan, D.P., and Campbell, L.C. (1981). Influence of potassium and manganese on growth and uptake of magnesium by soybeans (*Glycine max* (L.) Merr. cv. Bragg). *Plant Soil*, 61: 447-456. <https://doi.org/10.1007/BF02182025>
- Hopkins, B.G., Horneck, D.A., and MacGuidwin, A. (2014). Improving phosphorus use efficiency through potato rhizosphere modification and extension. *American Journal of Potato Research*, 91: 161–74. <https://doi.org/10.1007/s12230-014-9370-3>.
- Hochmuth, G. J., Maynard, D., Vavrina, C., Hanlon, E., and Simonne, E. (2018). Plant tissue analysis and interpretation for vegetable crops in Florida. Horticultural Sciences Department, UF/IFAS Extension, HS964 series. <https://edis.ifas.ufl.edu/publication/ep081> (Accessed on June 01, 2021).
- Isaacs, J. (2020). Small and mighty: micronutrients in potatoes. Micronutrients are key for potato. *Potatoes in Canada*. <https://www.potatoesincanada.com/small-and-mighty-micronutrients-in-potatoes/> (Accessed on February 06, 2023).
- Jafarbiglu, H., and Pourreza, A. (2022). A comprehensive review of remote sensing platforms, sensors, and applications in nut crops. *Computers and Electronics in Agriculture*, 197: 106844. <https://doi.org/10.1016/j.compag.2022.106844>.
- Jasim, A., Zaeen, A., Sharma, L.K., Bali, S.K., Wang, C., Buzza, A., and Alyokhin, A. (2020). Predicting phosphorus and potato yield using active and passive sensors. *Agriculture*, 10 (11): 564. <https://doi.org/10.3390/agriculture10110564>.

- Jenal, A., Bareth, G., Bolten, A., Kneer, C., Weber, I., and Bongartz, J. (2019a). Development of a VNIR/SWIR Multispectral Imaging System for Vegetation Monitoring with Unmanned Aerial Vehicles. *Sensors*, 19 (24): 5507. <https://doi.org/10.3390/s19245507>.
- Jenal, A., Hüging, H., Ahrends, H.E., Bolten, A., Bongartz, J., and Bareth, G. (2021b). Investigating the Potential of a Newly Developed UAV-Mounted VNIR/SWIR Imaging System for Monitoring Crop Traits—A Case Study for Winter Wheat. *Remote Sensing*, 13 (9): 1697. <https://doi.org/10.3390/rs13091697>.
- Juan, A.d. (2019). Chapter 2.5 - Multivariate Curve Resolution for Hyperspectral Image Analysis. Editor(s): Amigo, J.M. In *Data Handling in Science and Technology*, 32:115–150. Elsevier Ltd. ISBN: 9780444639776. <https://doi.org/10.1016/B978-0-444-63977-6.00007-9>.
- Kaiser, D.E., and Rosen, C.J. (2018). Understanding plant analysis for crops. University of Minnesota Extension. <https://extension.umn.edu/testing-and-analysis/understanding-plant-analysis-crops> (Accessed on June 22, 2021).
- Koch, M., Naumann, M., Pawelzik, E., Gransee, A., and Thiel, H. (2020). The importance of nutrient management for potato production part I: plant nutrition and yield. *Potato Research*, 63:97-119. <https://doi.org/10.1007/s11540-019-09431-2>.
- Kong, W., Liu, F., Zhang, C., Bao, Y., Yu, J., and He, Y. (2014). Fast detection of peroxidase (POD) activity in tomato leaves which infected with botrytis cinerea using hyperspectral imaging. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 118: 498–502. <https://doi.org/10.1016/j.saa.2013.09.009>.
- Knox, N.M., Skidmore, A.K., Prins, H.H.T., Heitkönig, I.M.A., Slotow, R., Van der Waal, C., and de Boer, W.F. (2012). Remote sensing of forage nutrients: combining ecological and spectral absorption feature data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 72: 27-35. <https://doi.org/10.1016/j.isprsjprs.2012.05.013>.
- Kriegler, F.J., Malila, W.A., Nalepka, R.F., Richardson, W. (1969). Preprocessing transformations and their effect on multispectral recognition. *Proceedings of the 6th International Symposium on Remote Sensing of Environment*. Ann Arbor, MI: University of Michigan, 97-131.
- Krinitzkiy, M., Aleksandrova, M., Verezemskaya, P., Gulev, S., Sinitsyn, A., Kovaleva, N., and Gavrikov, Al. (2021). On the generalization ability of data-driven models in the problem of total cloud cover retrieval. *Remote Sensing*, 13 (2): 326. <https://doi.org/10.3390/rs13020326>.
- Kumar, S., and Pandey, A.K. (2013). Chemistry and biological activities of flavonoids: an overview. *The Scientific World Journal*, 2013, Article ID 162750: 16. <https://doi.org/10.1155/2013/162750>.

- Lazarević, B., Carović-Stanko, K., Živčak, M., Vodnik, D., Javornik, T., and Safner, T. (2022). Classification of high-throughput phenotyping data for differentiation among nutrient deficiency in common bean. *Frontiers in Plant Science*, 13: 931877. <https://doi.org/10.3389/fpls.2022.931877>.
- Liao, H., Jianguo, W., Wenrong, C., Weidong, G., and Chunhai, S. (2012). Rapid diagnosis of nutrient elements in fingered citron leaf using near infrared reflectance spectroscopy. *Journal of Plant Nutrition*, 35 (11): 1725–1734. <https://doi.org/10.1080/01904167.2012.698352>.
- Ling, B., Goodin, D.G., Raynor, E.J., and Joern, A. (2019). Hyperspectral analysis of leaf pigments and nutritional elements in tallgrass prairie vegetation. *Frontiers in Plant Science*, 10: 142. <https://doi.org/10.3389/fpls.2019.00142>.
- Liu, N., Townsend, P.A., Naber, M.R., Bethke, P.C., Hills, W.B., and Wang, Y. (2021a). Hyperspectral imagery to monitor crop nutrient status within and across growing seasons. *Remote Sensing of Environment*, 255: 112303. <https://doi.org/10.1016/j.rse.2021.112303>.
- Liu, E., Zhao, H., Zhang, S., He, J., Yang, X., and Xiao, X. (2021b). Identification of plant species in an alpine steppe of northern Tibet using close-range hyperspectral imagery. *Ecological Informatics*, 61: 101213. <https://doi.org/10.1016/j.ecoinf.2021.101213>.
- Liu, J., Sun, Y., Liu, W., Tan, Z., Jiang, J., and Li, Y. (2013). Association of spectroscopically determined leaf nutrition related traits and breeding selection in sassafras tzumu. *Plant Methods*, 17: 33. <https://doi.org/10.1186/s13007-021-00734-5>.
- Li, Y., Al-Sarayreh, M., Irie, K., Hackell, D., Bourdot, G., Reis, M.M., and Ghamkhar, K. (2021). Identification of weeds based on hyperspectral imaging and machine learning. *Frontiers in Plant Science*, 11: 611622. <https://doi.org/10.3389/fpls.2020.611622>.
- Lu, J., Yang, T., Su, X., Qi, H., Yao, X., Cheng, T., Zhu, Y., Cao, W., and Tian, Y. (2020). Monitoring leaf potassium content using hyperspectral vegetation indices in rice leaves. *Precision Agriculture*, 21: 324–348. <https://doi.org/10.1007/s11119-019-09670-w>.
- Mattila, H., Sotoudehnia, P., Kuuslampi, T., Stracke, R., Mishra, K.B., and Tyystjärvi, E. (2021). Singlet oxygen, flavonols and photoinhibition in green and senescing silver birch leaves. *Trees - Structure and Function*, 35 (4): 1267–1282. <https://doi.org/10.1007/s00468-021-02114-x>.
- Mahajan, G. R., Das, B., Murgaokar, D., Herrmann, I., Berger, K., Sahoo, R.N., Patel, K., Desai, A., Morajkar, S., and Kulkarni, R.M. (2021). Monitoring the foliar nutrients status of mango using spectroscopy-based spectral indices and PLSR-combined machine learning models. *Remote Sensing*, 13 (4): 641. <https://doi.org/10.3390/rs13040641>.

- Mee, C.H., Chong, Balasundram, S.K., and Hanif, A.H.M. (2016). Detecting and monitoring plant nutrient stress using remote sensing approaches: a review. *Asian Journal of Plant Sciences*, 16 (1): 1–8. <https://doi.org/10.3923/ajps.2017.1.8>.
- Mertens, S., Verbraeken, L., Sprenger, H., Demuynck, K., Maleux, K., Cannoot, B., De Block, J., Maere, S., Nelissen, H., Bonaventure, G., Crafts-Brandner, S.J., Vogel, J.T., Bruce, W., Inzé, D., and Wuyts, N. (2021). Proximal hyperspectral imaging detects diurnal and drought-induced changes in maize physiology. *Frontiers in Plant Science*, 12: 640914. <https://doi.org/10.3389/fpls.2021.640914>.
- Mezera, J., Lukas, V., Horniaček, I., Smutný, V., and Elbl, J. (2021). Comparison of proximal and remote sensing for the diagnosis of crop status in site-specific crop management. *Sensors*, 22 (1): 19. <https://doi.org/10.3390/s22010019>.
- Menesatti, P., Antonucci, F., Pallottino, F., Rocuzzo, G., Allegra, M., Stagno, F., and Intrigliolo, F. (2010). Estimation of plant nutritional status by Vis–NIR spectrophotometric analysis on orange leaves [*Citrus Sinensis* (L) Osbeck Cv Tarocco]. *Biosystems Engineering*, 105 (4): 448–454. <https://doi.org/10.1016/j.biosystemseng.2010.01.003>.
- Morcillo-Pallarés, P., Rivera-Caicedo, J.P., Belda, S., De Grave, C., Burriel, H., Moreno, J., and Verrelst, J. (2019). Quantifying the robustness of vegetation indices through global sensitivity analysis of homogeneous and forest leaf-canopy radiative transfer models. *Remote Sensing*, 11 (20): 2418. <https://doi.org/10.3390/rs11202418>.
- Motsara, M., and Roy, R.N. (2008). Guide to laboratory establishment for plant nutrient analysis. Food and agriculture organization of the United Nations, Rome. ISBN: 978-92-5-105981-4
- Muñoz-Huerta, R.F., Guevara-Gonzalez, R.G., Contreras-Medina, L.M., Torres-Pacheco, I., Prado-Olivarez, J., and Ocampo-Velazquez, R. V. (2013). A review of methods for sensing the nitrogen status in plants: advantages, disadvantages, and recent advances. *Sensors*, 13: 10823-10843. <https://doi.org/10.3390/s130810823>.
- Muttucumaru, N., Powers, S.J., Elmore, J.S., Mottram, D.S., and Halford, N.G. (2013). Effects of nitrogen and sulfur fertilization on free amino acids, sugars, and acrylamide-forming potential in potato. *Journal of Agricultural and Food Chemistry*, 61 (27): 6734–6742. <https://doi.org/10.1021/jf401570x>.
- Naumann, M., Koch, M., Thiel, H., Gransee, A., and Pawelzik, E. (2020). The importance of nutrient management for potato production part II: plant nutrition and tuber quality. *Potato Research*, 63: 121-137. <https://doi.org/10.1007/s11540-019-09430-3>.
- Ortiz, J.D., Avouris, D., Schiller, S., Luvall, J.C., Lekki, J.D., Tokars, R.P., Anderson, R.C., Shuchman, R., Sayers, M., and Becker, R. (2017). Intercomparison of approaches to the empirical line method for vicarious hyperspectral reflectance calibration. *Frontiers in Marine Science*, 4: 296. <https://doi.org/10.3389/fmars.2017.00296>.

- Osco, L. P., Ramos, A.P.M., Pinheiro, M.M.F., Moriya, É.A.S.M., Imai, N.N., Estrabis, N., Ianczyk, F., de Araújo, F.F., Liesenberg, V., de Castro Jorge, L.A., Li, J., Ma, L., Gonçalves, W.N., Junior, J.M., 1 and Creste, J.E. (2020). A machine learning framework to predict nutrient content in Valencia-orange leaf hyperspectral measurements. *Remote Sensing*, 12 (6): 906. <https://doi.org/10.3390/rs12060906>.
- Parrini, S., Staglianò, N., Bozzi, R., and Argenti, G. (2021). Can grassland chemical quality be quantified using transform near-infrared spectroscopy?. *Animals*, 12 (1): 86. <https://doi.org/10.3390/ani12010086>.
- Pandey, P., Ge, Y., Stoerger, V., and Schnable, J.C. (2017). High throughput in vivo analysis of plant leaf chemical properties using hyperspectral imaging. *Frontiers in Plant Science*, 8: 1348. <https://doi.org/10.3389/fpls.2017.01348>.
- Pádua, L., Adão, T., Sousa, A., Peres, E., and Sousa, J.J. (2020). Individual grapevine analysis in a multi-temporal context using UAV-based multi-sensor imagery. *Remote Sensing*, 12 (1): 139. <https://doi.org/10.3390/rs12010139>.
- Peng, Y., Zhang, M., Xu, Z., Yang, T., Su, Y., Zhou, T., Wang, H., Wang, Y., and Lin, Y. (2020). Estimation of leaf nutrition status in degraded vegetation based on field survey and hyperspectral data. *Scientific Reports*, 10 (1): 4361. <https://doi.org/10.1038/s41598-020-61294-7>.
- Peng, J., Manevski, K., Kørup, K., Larsen, R., and Andersen, M.N. (2021). Random forest regression results in accurate assessment of potato nitrogen status based on multispectral data from different platforms and the critical concentration approach. *Field Crops Research*, 268: 108158. <https://doi.org/10.1016/j.fcr.2021.108158>.
- Phillips, S. B., Keahey, D. A., Warren, J. G., and Mullins, G. L. (2004). Estimating winter wheat tiller density using spectral reflectance sensors for early-spring, variable-rate nitrogen applications. *Agronomy Journal*, 96 (3): 591–600. <https://doi.org/10.2134/agronj2004.0591>.
- Pompilio, L., Marinangeli, L., Amitrano, L., Pacci, G., D'andrea, S., Iacullo, S., and Monaco, E. (2018). Application of the empirical line method (ELM) to calibrate the airborne daedalus-CZCS scanner. *European Journal of Remote Sensing*, 51 (1): 33–46. <https://doi.org/10.1080/22797254.2017.1399087>.
- Potato Growers of Canada (2022). 2022 Estimated Canadian Potato Plantings (Acres). <https://unitedpotatocanada.com/wp-content/uploads/2022/07/Canadian-Potato-Acreage-2022-Estimate-07-15-2022.pdf> (Accessed on June 22, 2023)
- Prananto, A.J., Minasny, B., and Weaver, T. (2020). Chapter one - Near infrared (NIR) spectroscopy as a rapid and cost-effective method for nutrient analysis of plant leaf tissues. *Advances in Agronomy*, 164:1-49. <https://doi.org/10.1016/bs.agron.2020.06.001>.

- Prananto, J. A., Minasny, B., and Weaver, T. (2021). Rapid and cost-effective nutrient content analysis of cotton leaves using near-infrared spectroscopy (NIRS). *PeerJ*, 9: e11042. <https://doi.org/10.7717/peerj.11042>.
- Rad, S.A., Dehghanian, Z., Lajayer, B.A., Gürgan, M., and Astatkie, T. (2022). Chapter 7 - Role of physical and chemical agents in plants for tolerance to boron nutrition. Editor(s): Aftab, T., Landi, M., Papadakis, I.E., Araniti, F., Brown, P.H. In *Boron in Plants and Agriculture*: 147-168. Academic Press. ISBN: 9780323908573. <https://doi.org/10.1016/B978-0-323-90857-3.00013-8>.
- Rasti, B., Scheunders, P., Ghamisi, P., Licciardi, G. and Chanussot, J. (2018). Noise reduction in hyperspectral imagery: overview and application. *Remote Sensing*, 10 (3): 482. <https://doi.org/10.3390/rs10030482>.
- Rauschenberger, A., and Glaab, E. (2021). Predicting correlated outcomes from molecular data. Edited by Jonathan Wren. *Bioinformatics*, 37 (21): 3889-95. <https://doi.org/10.1093/bioinformatics/btab576>.
- Rehman, T. H., Reis, A.F.B., Akbar, N., and Linquist, B.A. (2019). Use of normalized difference vegetation index to assess N status and predict grain yield in rice. *Agronomy Journal*, 111 (6): 2889–2898. <https://doi.org/10.2134/agronj2019.03.0217>.
- Ren, J., Cui, J., Dong, W., Xiao, Y., Xu, M., Liu, S., Wan, J., Li, Z., and Zhang, J. (2023). Remote sensing inversion of typical offshore water quality parameter concentration based on improved SVR algorithm. *Remote Sensing*, 15 (8): 2104. <https://doi.org/10.3390/rs15082104>.
- Rosen, C.J., Kelling, K.A., Stark, J.C., and Porter, G.A. (2014). Optimizing phosphorus fertilizer management in potato production. *American Journal of Potato Research*, 91:145-160. <https://doi.org/10.1007/s12230-014-9371-2>.
- Rowe, R.C. & American Phytopathological Society & Potato Association of America (1993). *Potato health management*. APS Press, St. Paul, Minn.
- Saari, H., Pellikka, I., Pesonen, L., Tuominen, S., Heikkilä, J., Holmlund, C., Mäkyänen, J., Ojala, K., and Antila, T. (2011). Unmanned aerial vehicle (UAV) operated spectral camera system for forest and agriculture applications. *Proceedings Volume 8174, Remote Sensing for Agriculture, Ecosystems, and Hydrology XIII*; 81740H. Event: SPIE Remote Sensing, Prague, Czech Republic. <https://doi.org/10.1117/12.897585>
- Sampaio, S.P., and Brites, C.M. (2022). Near-Infrared Spectroscopy and Machine Learning: Analysis and Classification Methods of Rice. In *Integrative Advances in Rice Research*, edited by Huang, M. IntechOpen. <https://doi.org/10.5772/intechopen.99017>.
- Saúco, V. (1997). Horticultural practices of mango. *Acta Horticulturae*, 455:391-400. <https://doi.org/10.17660/ActaHortic.1997.455.50>.

- Schleuss, P. M., Widdig, M., Heintz-Buschart, A., Kirkman, K., and Spohn, M. (2020). Interactions of nitrogen and phosphorus cycling promote P acquisition and explain synergistic plant-growth responses. *Ecology*, 101: e03003. <https://doi.org/10.1002/ecy.3003>.
- Shaver, T.M., Khosla, R., and Westfall, D.G. (2011). Evaluation of two crop canopy sensors for nitrogen variability determination in irrigated maize. *Precision Agriculture*, 12: 892–904. <https://doi.org/10.1007/s11119-011-9229-2>.
- Shao, Y., and He. Y. (2013). Visible/near infrared spectroscopy and chemometrics for the prediction of trace element (Fe and Zn) levels in rice leaf. *Sensors (Basel)*, 13 (2): 1872-1883. doi:10.3390/s130201872.
- Sheng, Q., Santos-Rivera, M., Ouyang, X., Kouba, A.J., and Vance, C.K. (2022). Near-infrared spectroscopy and mode cloning (NIR-MC) for in-situ analysis of crude protein in bamboo. *Remote Sensing*, 14 (6): 1302. <https://doi.org/10.3390/rs14061302>.
- Schmidt, W., Thomine, S., and Buckhout, T.J. (2020). Editorial: iron nutrition and interactions in plants. *Frontiers in Plant Science*, 10: 1670. <https://doi.org/10.3389/fpls.2019.01670>.
- Shu, M., Shen, M., Zuo, J., Yin, P., Wang, M., Xie, Z., Tang, J., Wang, R., Li, B., Yang, X., and Ma, Y. (2021). The application of UAV-based hyperspectral imaging to estimate crop traits in maize inbred lines. *Plant Phenomics*, 2021. <https://doi.org/10.34133/2021/9890745>.
- Siedliska, A., Baranowski, P., Pastuszka-Woźniak, J., Zubik, M., and Krzyszczak, J. (2021). Identification of plant leaf phosphorus content at different growth stages based on hyperspectral reflectance. *BMC Plant Biology*, 21: 28. <https://doi.org/10.1186/s12870-020-02807-4>.
- Silva, R., Gomes, V., Mendes-Faia, A., and Melo-Pinto, P. (2018). Using support vector regression and hyperspectral imaging for the prediction of oenological parameters on different vintages and varieties of wine grape berries. *Remote Sensing*, 10 (2): 312. <https://doi.org/10.3390/rs10020312>.
- Simko, I., Hayes, R.J., and Furbank, R.T. (2016). Non-destructive phenotyping of lettuce plants in early stages of development with optical sensors. *Frontiers in Plant Science*, 7:1985. <https://doi.org/10.3389/fpls.2016.01985>.
- Singh, L., Mutanga, O., Mafongoya, P., Peerbhay, K., and Crous, J. (2022). Hyperspectral remote sensing for foliar nutrient detection in forestry: A near-infrared perspective. *Remote Sensing Applications: Society and Environment*, 25: 100676. <https://doi.org/10.1016/j.rsase.2021.100676>.

- Solano-Alvarez, N., Valencia-Hernández, J. A., Vergara-Pineda, S., Millán-Almaraz, J.R., Torres-Pacheco, I., and Guevara-González, R.G. (2022). Comparative analysis of the NDVI and NGBVI as indicators of the protective effect of beneficial bacteria in conditions of biotic stress. *Plants*, 11 (7): 932. <https://doi.org/10.3390/plants11070932>.
- Sonobe, S., and Wang, Q. (2017). Towards a universal hyperspectral index to assess chlorophyll content in deciduous forests. *Remote Sensing*, 9 (3): 191. <https://doi.org/10.3390/rs9030191>.
- Statistics Canada (2022). Farm management survey, 2021. <https://www150.statcan.gc.ca/n1/daily-quotidien/221215/dq221215b-eng.htm> (Accessed on February 06, 2023).
- Steiner, F., Zoz, T., Pinto Junior, A.S., Castagnara, D.D., and Dranski, J.A.L. (2012). Effects of aluminum on plant growth and nutrient uptake in young physic nut plants. *Semina: Ciências Agrárias*, 33 (5): 1779-1788. <https://doi.org/10.5433/1679-0359.2012v33n5p1779>.
- Svensgaard, J., Roitsch, T., and Christensen, S. (2014). Development of a mobile multispectral imaging platform for precise field phenotyping. *Agronomy*, 4 (3): 322–336. <https://doi.org/10.3390/agronomy4030322>.
- Taha, M. F., ElManawy, A.I., Alshallash, K.S., ElMasry, G., Alharbi, K., Zhou, L., Liang, N., and Qiu, Z. (2022). Using machine learning for nutrient content detection of aquaponics-grown plants based on spectral data. *Sustainability*, 14 (19): 12318. <https://doi.org/10.3390/su141912318>
- Thomas, S., Behmann, J., Steier, A., Kraska, T., Muller, O., Rascher, U., and Mahlein, A.K. (2018). Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform. *Plant Methods*, 14:45. <https://doi.org/10.1186/s13007-018-0313-8>
- Tremblay, N., Wang, Z., Ma, B.L., Belec, C., and Vigneault, P. (2009). A comparison of crop data measured by two commercial sensors for variable-rate nitrogen application. *Precision Agriculture*, 10: 145-161. <https://doi.org/10.1007/s11119-008-9080-2>.
- Wang, J., Shen, C., Liu, N., Jin, X., Fan, X., Dong, C., and Xu, Y. (2017). Non-destructive evaluation of the leaf nitrogen concentration by in-field visible/near-infrared spectroscopy in pear orchards. *Sensors*, 17 (3): 538. <https://doi.org/10.3390/s17030538>.
- Wang, C., and Myint, S.W. (2015). A simplified empirical line method of radiometric calibration for small unmanned aircraft systems-based remote sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8 (5): 1876–1885. <https://doi.org/10.1109/JSTARS.2015.2422716>.
- Williams, P., Antoniszyn, J., and Manley, M. (2019). Near infrared technology: getting the best out of Light. *AFRICAN SUN MeDIA*. <http://doi.org/10.18820/9781928480310>.

- Worsfold, P. J., and Zagatto, E.A.G. (2017). Spectrophotometry: overview. Editor(s): Worsfold, P., Poole, C., Townshend, A., Miró, M. In Reference Module in Chemistry, Molecular Sciences and Chemical Engineering. Elsevier. ISBN: 9780081019849. <https://doi.org/10.1016/B978-0-12-409547-2.14265-9>.
- Xue, J., and Su, B. (2017). Significant remote sensing vegetation indices: a review of developments and applications. *Journal of Sensors*, 2017: 1353691. <https://doi.org/10.1155/2017/1353691>.
- Xu, K., Gong, Y., Fang, S., Wang, K., Lin, Z., and Wang, F. (2019). Radiometric calibration of UAV remote sensing image with spectral angle constraint. *Remote Sensing*, 11 (11): 1291. <https://doi.org/10.3390/rs11111291>.
- Xu, Y., and Wang, L. (2014). K-Nearest neighbor-based weighted twin support vector regression. *Applied Intelligence*, 41: 299–309. <https://doi.org/10.1007/s10489-014-0518-0>.
- Yang, H., Yang, X., Heskell, M., Sun, S., and Tang, J., (2017). Seasonal variations of leaf and canopy properties tracked by ground based NDVI imagery in a temperate forest. *Scientific Reports*, 7: 1267. <https://doi.org/10.1038/s41598-017-01260-y>.
- Yunjun, Z., Fattahi, H., and Amelung, F. (2019). Small Baseline InSAR Time Series Analysis: Unwrapping Error Correction and Noise Reduction. *Computers & Geosciences*, 133: 104331. <https://doi.org/10.1016/j.cageo.2019.104331>.
- Zebarth, B. J., Leclerc, Y., Moreau, G., and Botha, E. (2004). Rate and timing of nitrogen fertilization of Russet Burbank potato: yield and processing quality. *Canadian journal of plant science*: 855-863.
- Zebarth, B., Moreau, G., and Karemangingo, C. (2007). Nitrogen management for potatoes: petiole nitrate testing. GHG Taking Charge Team Factsheet. https://www2.gnb.ca/content/dam/gnb/Departments/10/pdf/Agriculture/potato_pnit.pdf (Accessed on Feb. 10, 2023)
- Zhai, Y., Cui, L., Zhou, X., Gao, Y., Fei, T., and Gao, W. (2013). Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: comparison of partial least-square regression and support vector machine regression methods. *International Journal of Remote Sensing*, 34 (7): 2502–2518. <https://doi.org/10.1080/01431161.2012.746484>.
- Zhao, S., Pan, C. (2020). Spectral parameter-based models for leaf potassium concentration estimation in Ping'ou hybrid hazelnut. *Journal of Arid Land*, 12: 1083–1092. <https://doi.org/10.1007/s40333-020-0081-y>

- Zhao, X., Zhang, J., Huang, Y., Tian, Y., and Yuan, L. (2022). Detection and discrimination of disease and insect stress of tea plants using hyperspectral imaging combined with wavelet analysis. *Computers and Electronics in Agriculture*, 193: 106717. <https://doi.org/10.1016/j.compag.2022.106717>.
- Zhang, X., and He, Y. (2013). Rapid estimation of seed yield using hyperspectral images of oilseed rape leaves. *Industrial Crops and Products*, 42: 416–420. <https://doi.org/10.1016/j.indcrop.2012.06.021>.
- Zhang, Y., Wu, J., and Wang, A. (2022). Comparison of various approaches for estimating leaf water content and stomatal conductance in different plant species using hyperspectral data. *Ecological Indicators*, 142: 109278. <https://doi.org/10.1016/j.ecolind.2022.109278>.
- Zhang, D., Zhao, J., Huang, W., Zhu, D., Huang, L., Xu, X., Ma, Z., and Wang, J. (2012). Retrieval of maize leaf chlorophyll content using peak-valley characteristic parameters from imaging hyperspectra. *Australian Journal of Crop Science*, 6: 109-115.
- Zhen, Z., Chen, S., Qin, W., Yan, G., Gastellu-Etchegorry, J.P., Cao, L., Murefu, M., Li, J., and Han, B. (2020). Potentials and limits of vegetation indices with BRDF signatures for soil-noise resistance and estimation of leaf area index. *IEEE Transactions on Geoscience and Remote Sensing*, 58 (7): 5092–5108. <https://doi.org/10.1109/TGRS.2020.2972297>.