



Near-infrared reflectance spectroscopy for nursery *Eucalyptus* seedlings and cuttings production.

By

Nkosikhona Trywell Mkhize

BSc. (UKZN), BSc. Hons. (UKZN)

Submitted in partial fulfilment of the requirements for the degree of

Master of Science in Agriculture: Plant Pathology

In the

Discipline of Plant Pathology

School of Agricultural, Earth and Environmental Sciences

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

February 2025

Dissertation Summary

A critical function and purpose of a nursery is for seedlings and cuttings production for farming systems. Nurseries provide farms with known pure varieties that are healthy and pathogen-free. Therefore, seed and subsequent seedling quality assessments are of great importance. Traditional methods for assessing seed (i.e. germination, and vigour) and hedge (i.e. clonal purity and nutrient content) either require long and laborious procedures or expensive chemical/molecular analysis, making them less accessible to nursery managers. Near-infrared reflectance spectroscopy (NIRS) is a rapid low-cost analysis tool easily integrated into various processes and industries, making it a viable evaluation technique. Hence, this study developed prediction models for seed germination, foliar nutrient content and hedge/clonal purity and tested whether NIRS can improve specific nursery quality control tasks, namely, seed germination and vigour testing, foliar nutrient content analysis and hedge clonal purity assessments via clonal identification.

Seed germination rate and vigour are important characteristics that impact the success rate of plantation establishment. Hence, a study aimed to evaluate the feasibility of NIRS in predicting *Eucalyptus* seed germination and seed vigour was initiated. Species included in the study were *E. nitens*, *E. dunnii*, *E. benthamii* and *E. macarthurii*, which were accessed from 191 seed lots. Seed germination and vigour were determined through seed germination assays in 96-well culture plates for 10 days and used as the reference data for model development and validation. The results indicated that NIRS can predict seed germination with high accuracy in most of the species tested (RMSEP: 0.357 – 73.5%, R^2 : 58.46 – 95.42%) and collectively as a global species model (RMSEP: 11.4 – 18.8%, R^2 : 73.68 – 84.99%). The results for seed vigour also indicated that NIRS can predict with accuracy for the species tested (RMSEP: 5.68 – 126%, R^2 : 78.91 - 83.57%) and collectively as a global species model (RMSEP: 18.8%, R^2 : 71.37%). These NIR models showed promise for enhancing *Eucalyptus* seed germination and vigour testing. This study was the first to report the use of NIR as a viable tool for seed germination and vigour in nursery operations prioritising *Eucalyptus* plantations.

Eucalyptus clones are intensively propagated in nurseries and planted to establish commercial plantations. The nutritional status of *Eucalyptus* cuttings impacts their

ability to establish viable hedges and subsequently impacts the success of plantation establishment. Therefore, a study was conducted to determine whether NIRS can be used to measure and assess *Eucalyptus* foliar nutrient levels. Foliar material was sampled from the field and pot trials that were part of long-term monitoring nutrient depletion trials, sampled from 1 year-old coppice trees and pot trial samples, with a total of 382 samples with a total of 1146 spectra excluding independent test data set. The spectra were captured using a Bruker MPA NIRS. Foliar nutrient quantification was done on an Agilent MP-AES 4200 and Leco CNS and was used as the reference data to create NIR nutrient calibration models. Global models were developed for key nutrients (N, P and K) and correlated well with the reference nutrient data, having $R^2 > 89\%$ for *E. dunnii*, G×N and G×U. This indicate that NIRS was able to model certain elements well, whilst others to a lesser extent. The significance of this study was that spectroscopic techniques, like NIRS, have promise to measure plant nutrient levels. However, this requires more investigation and development.

The accurate identification of clonal purity is a critical step in forestry industries to ensure uniformity, optimize productivity, and maintain genetic integrity. Therefore, the ability of NIRS to determine hedge clonal purity was evaluated. The foliar material of 4 *Eucalyptus* varieties, G×U400, G×NPP, GL222 and G×U488, were sampled from commercial hedges at Top Crop Nursery (Cramond, Pietermaritzburg). The samples were dried, milled and presented to the NIRS for spectra acquisition, and the reference data used for model calibration relied on the nursery label of the hedges used. Principal component analysis (PCA) of the spectral signatures was able to cluster the samples to their respective groups using 2 and 3-dimensional PCA plots. Categorical models were created using the Ident module of the Bruker OPUS software. Upon validation of the clonal purity model, a confusion matrix indicated that all varieties except G×U400 could be correctly identified, possibly due to contamination with GL222 at this nursery, which is the problem this approach aims to resolve. The model accuracy for G×NPP, G×U400, G×U488 and GL222 were 100.00, 60.00, 92.00 and 68.00%, respectively. Considering the potential contamination of the reference material, the models performed reasonably well and warrant further development. Further development would include the use of DNA validation of calibration and validation samples. This study highlights the potential of NIRS as a transformative tool

for clonal verification in large-scale forestry operations, with implications for improving efficiency and reducing operational costs.

Finally, the study demonstrated the capability of NIRS to expedite the analysis of various nursery processes. Once this research is refined and optimized, it can save on nursery analysis costs and improve the monitoring of seed lots and hedges. Therefore, the current study has contributed to advancing forestry nursery operations.

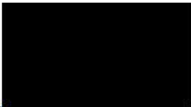
Declaration

I, **Nkosikhona Trywell Mkhize** declare that:

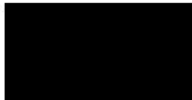
1. The research reported in this thesis, except where otherwise indicated or acknowledged, is my original work.
2. This thesis has not been submitted in full or in part for any degree or examination to any other university.
3. This thesis does not contain other persons' data, pictures, graphs, or other information unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, their words have been re-written, but the general information attributed to them has been referenced; where their exact words have been used, their writing has been placed inside quotation marks and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the Internet.



Nkosikhona Trywell Mkhize (Candidate)

Dr Nokwazi Carol Mbili (Supervisor):  _____

Dr Richard Burgdorf (Co-Supervisor):  _____

Dr Danvir Ramesar (Co-supervisor):  _____

Acknowledgments

I would like to express my sincere gratitude to the following:

My supervisor, Dr Richard Burgdorf, for initiating this project and selecting me to work on it; my co-supervisor, Dr Danvir Ramesar at the ICFR, who has provided me with the technical and scientific support I needed during this work; and Dr Nokwazi Mbili, my academic supervisor for her assistance, guidance, and constructive criticism throughout the study. This has been a transformative experience for me.

Furthermore, I wish to thank the Top Crop Nursery team, Mr Mike Kruger and Ms Mpumi Hadebe, for their support and assistance during field samplings. I also wish to express my gratitude to the technical staff at the Institute for Commercial Forestry Research (ICFR) for their support, guidance and technical assistance throughout my research. Thanks to my fellow students at the ICFR and the University of KwaZulu-Natal for moral support and assistance throughout this work. I extend huge gratitude and appreciation to NTC for funding and technical support, which made this research possible. I would like to express my great appreciation to the ICFR for accommodating this research and supporting me throughout the project.

Finally, I wish to thank my family and friends for their unending faith in me as a student at UKZN. My wish is to make them proud and for them to know that I am eternally grateful for all they have done.

Table of Contents

Dissertation Summary	ii
Declaration	v
Acknowledgments	vi
Table of Contents	vii
List of Figures	x
List of Tables	xi
List of Abbreviations	xii
General Introduction	xiv
Chapter 1	1
The applications of Near infrared spectroscopy in forestry seedling production: A review	1
Abstract	1
1.1 Introduction	1
1.2 Background of forestry and its importance	3
1.3 Seedling production for forestry	4
1.4 Seedling germination	5
1.5 Seedling quality assessment procedures	5
1.6 Morden technology	6
1.7 Near-infrared spectroscopy (NIRS)	7
1.8 The use of NIR	9
1.9 Clone identification	10
1.10 Impact and future work	11
1.11 References	12
Chapter 2	16

Screening of <i>Eucalyptus</i> species for seed germination and vigour using near-infrared reflectance spectroscopy.....	16
Abstract.....	16
2.1 Introduction	17
2.2 Materials and methods.....	18
2.2.1 Sample identification and sub-sampling	18
2.2.2 Seed preparation and spectral acquisition	18
2.2.3 Assessment of seed germination and vigour	19
2.2.4 NIR model development and validation	20
2.3 Results	20
2.3.1 NIR model performance for predicting <i>Eucalyptus</i> seed germination	20
2.3.2 NIR model performance for predicting seed germination vigour for <i>Eucalyptus</i> varieties	29
2.4 Discussion.....	30
2.5 Conclusion	33
2.6 References.....	33
Chapter 3	37
Evaluation of near-infrared reflectance spectroscopy for low-cost foliar nutrient analysis of <i>Eucalyptus</i> seedlings.....	37
Abstract.....	37
3.1 Introduction	37
3.2 Materials and methods.....	39
3.2.1 Sample collection and preparation.....	39
3.2.2 Assessment of foliar nutrient analysis.....	39
3.2.3 NIR spectra acquisition and model development.....	40
3.3 Results	40
3.3.1 NIR model performance for predicting foliar nutrient levels in <i>Eucalyptus</i> varieties	40

3.3.2 NIR model performance of global <i>Eucalyptus</i> nutrient models	44
3.4 Discussion.....	47
3.5 Conclusion	49
3.6 References.....	50
Chapter 4	52
Evaluation of the ability of near-infrared spectroscopy to assess varietal purity of <i>Eucalyptus</i> hedges	52
Abstract.....	52
4.1 Introduction	53
4.2 Materials and methods.....	54
4.2.1 Sample collection.....	54
4.2.2 Sample processing	54
4.2.3 NIR model development	54
4.3 Results	55
4.3.1 Cluster analysis	55
4.3.2 Independent evaluation of model reliability	56
4.4 Discussion.....	58
4.5 Conclusion	59
4.6 References.....	60
Chapter 5	62
Thesis overview	62
5.1 Introduction	62
5.2 Research objectives and major findings.....	62
5.3 Recommendations and future research	64
Appendices	65
Appendix A: Near-infrared reflectance spectroscopy for low-cost foliar nutrient analysis of forestry seedlings	65

List of Figures

Figure 2.1: Representative plots comparing the NIR seed germination values between calibration and validation datasets, analysed using the global models for day 3, 7 and 10 intervals.

Figure 2.2: Level of accuracy of the day 10 global seed germination model assessed against independent samples of different *Eucalyptus* varieties.

Figure 3.1: Representative plots of foliar nutrient data set, showing test validation data set A- Calibration and validation test for total (N) nitrogen, B - calibration and validation test for total phosphorus (P) and C - calibration and validation test for total potassium (K).

Figure 3.2: The correlation graphs set for true vs predicted values between nutrient models for (A) nitrogen, (B) phosphorus and (C) potassium on independent test for *E. dunnii*, G×N and G×U.

Figure 4.1: The 2-d PCA cluster presentation of data points are shown to be grouped to distinctive clusters.

Figure 4.2: The 3-D PCA presentation of the sphere of clustered data set. The coloured coded as green for G×U400, red for G×NPP, yellow for GL222 and blue for G×U488.

List of Tables

Table 1.1: An Infrared fraction, wavelengths between 780 and 2500 nm (12,500–4000 cm^{-1} , a wavenumber) (Francisco *et al.*, 2017).

Table 2.1: Description of *Eucalyptus* seed lots used.

Table 2.2: NIR model performance of seed germinated in *Eucalyptus* species after 3 days of incubation.

Table 2.3: NIR model performance of seed germinated in *Eucalyptus* species after 7 days of incubation.

Table 2.4: NIR model performance of seed germinated in *Eucalyptus* species after 10 days of incubation.

Table 2.5: Sample report for selected values produced by NIRS prediction of total germination percentage of *E. dunnii*, *E. nitens*, *E. macarthurii* and *E. benthamii* compared to the true values.

Table 2.6: NIR model performance of seed germinated in *Eucalyptus* species cumulative sum (vigour) after incubation.

Table 3.1: NIR model for nutrient analysis using *Eucalyptus* varieties (*E. dunnii*, hybrid G×N, and G×U, comparing R^2 of global model and single species analysis.

Table 3.2: NIR *Eucalyptus* global model performance metrics for foliar nutrients.

Table 3.3: Independent test dataset for N, P and K using the global model for samples of *E. dunnii*, hybrid G×N, and G×U.

Table 4.1: The presentation of confusion matrix of independent data testing.

Table 4.2: The presentation of how the method performed with the actual data to predict the unknown samples and model robustness.

Table 4.3: *Eucalyptus* varietal-wise prediction statistics.

Table A1: Independent test dataset for NPK on Global model and single species (*E. dunnii*, hybrid G×N, and G×U).

Table A2: NIR report of N, P and K values predicted for different *Eucalyptus* hybrids.

List of Abbreviations

FRAL - Forestry Research and Analytical Laboratory
FT-NIR - Fourier Transform Near-infrared spectroscopy
GDP - Gross Domestic Product
GHG - Green House Gas
GIS - Geographic Information Systems
GPS - Global Positioning System
ICFR - Institute for Commercial Forestry Research
ISTA - International Seed Testing Association
ML - Machine learning
MP-AES - Microwave plasma atomic emission spectroscopy
MSC - Multiplicative Scatter Correction
NIRS - Near Infrared Spectroscopy
NMR - nuclear magnetic resonance
PCA - Principal Component Analysis
PLS - Partial Least Squares
PLS-DA - Partial Least Squares Discriminant Analysis
PQI - Plant Quality Index
QC - Quality control
 R^2 - Coefficient of determination
RFID - Radio Frequency Identification
RMSEP - Root Mean Square Error of Prediction
RPD - Ratio of performance to Deviation
SANSOR - South African National Seed Organization
SEC - Standard Error of Calibration
SECV - Standard Error of Cross-Validation
SMSE - Root Mean Squared Error
SNV – Support Vector Normalisation
TZ - Tetrazolium test

UV - Ultraviolet spectroscopy

Vis-NIR - Visible Near Infrared

General Introduction

Forestry is defined by various range of activities which is establishment of tree plantations process by planting or seeding and silviculture for the purpose of forest regeneration (Onyekwelu *et al.*, 2011). It adopts typical agricultural practices, i.e., species selection and optimal growth performance. Forests cover approximately twenty million hectares globally. For instance, South African forestry has an export value of R38.4 billion, produced from 1,212,383 ha of commercial forestry plantations producing up to 13.4 million tonnes of forest products, such as timber, wood chips and bark extract. This makes a GDP contribution of 9.8% to agriculture and 4.9% to manufacturing and indirectly employs over half a million South Africans. Tree plantations in South Africa consist exclusively of exotic tree species, i.e., *Eucalyptus*, *Acacia*, and *Pinus*. *Eucalyptus* species are the fastest-growing commercial tree varieties worldwide. Their adoption and development in the South African Forestry sector have facilitated the increase in the production of timber, stem poles, pulp, paper and represent a significant proportion of the global market (Pallett, 2005). Furthermore, oils from certain *Eucalyptus* varieties are extracted for their medicinal and therapeutic properties (Shyam, 1989). The Institute for Commercial Forestry Research (ICFR) is engaged in the development of clonal cuttings, new hybrid varieties, and protocols for improved rooting ability and hedge management (Beck-Pay and Laing, 2019). Quality plants produced from seeds or clonal cuttings are important to meet the demands of commercial *Eucalyptus* forestry plantations.

Production of quality seedlings in nurseries is a critical step to supply seedlings with a high sowing success rate or infield germination and establishment (Riikonen and Luoranen, 2018). Hence, seed germination and vigour assays are conducted before sowing or seedling production (Wang *et al.*, 2020). Seed germination rate and vigour are influenced by physiological composition and environmental factors but can be negatively affected by physiological changes which occur due to long-term storage and improper storage conditions (Rao *et al.*, 2017; Walsh *et al.*, 2000). Therefore, seed germination and vigour are important and sensitive attributes requiring non-destructive, accurate, high throughput, and frequent assessment.

Conventional methods used to test seed quality include traditional seed germination assays, tetrazolium seed germ assays, conductivity assays, accelerated aging test

and low-temperature germination (Jia *et al.*, 2016; Landis *et al.*, 2010). Both agricultural and forestry industries require high-quality seeds and seedlings to a guarantee high rate for field establishment and silvicultural practice. Therefore, there is a requirement for new method development to be used at nurseries for monitoring the production of high-quality seedlings from seed to planting sites to address continuously changing needs of the forestry sector (Grossnickle and MacDonald, 2018).

Near-infrared reflectance spectroscopy (NIRS) is considered to be a viable method for seed quality assessment, while offering relatively low analysis costs and provide high throughput capability. The NIR region ranges from 780 to 2500 nm ($\sim 12500 - 4000 \text{ cm}^{-1}$). The basic concept of NIRS involves illuminating samples with light within the NIR region and then creating an absorbance profile from the light reflected to the detectors at different wavelengths (Budic-Ieto *et al.*, 2011; Li *et al.*, 2020). This profile captures the energy difference associated with the stretching of different covalent bonds and creates an NIR profile of the sample. This can then be processed via machine learning algorithms, where calibration models are developed to predict the properties of unknown samples. NIRS is regarded as an indirect analysis method of chemical and physical properties of samples, with little to no sample preparation required. Collection of sample spectra that capture overtone and combination vibrations from specific functional groups are correlated to sample properties using chemometric methods (Zhang *et al.*, 2020). A study by Al-Amery *et al.* (2018) reported that seed germination and seed vigour for soybeans can be reliably predicted using NIR. A feasibility study by Esteve Agelet *et al.* (2012) shows that NIR can detect viability and damaged soybeans and corn kernels. These studies encouraged similar applications to assess forestry seeds. Near-infrared reflectance spectroscopy provides real-time fast and non-destructive monitoring methods and offers optimal seedling quality and plant nutrient monitoring (Prananto *et al.*, 2020). It can also provide reliable clone purity identification (Wan-hong *et al.*, 2019).

Applications of NIRS have been used to assess plant nutrient content. However, this can be challenging due to various factors, including the sample material, NIR region selection techniques, spectral pre-processing and modelling approaches (Prananto *et al.*, 2020). Routine macronutrients measured as part of a plant analysis panel include nitrogen (N), phosphorus (P), potassium (K), sulphur (S), calcium (Ca) and

magnesium (Mg), while micronutrients include iron (Fe), zinc (Zn), manganese (Mn), and copper (Cu). Macronutrients were reported to be more reliably modelled than micronutrients and are further impacted by the age of the sample and its milled particle size (Van Maarschalkerweerd and Husted, 2015, and Prananto *et al.*, 2020). These elements are measured on analysers such as Microwave plasma atomic emission spectroscopy (MP-AES) or Induction coupled plasma optical emission spectroscopy (ICP-OES). These spectrophotometers are expensive to operate and maintain, and consequently have expensive analysis costs. Cheaper analysis options will allow for greater samples analysed and subsequently greater awareness of hedge nutrient levels.

Identification and or classification of *Eucalyptus* varieties can be done using molecular techniques such as DNA sequencing or fingerprinting (Lu *et al.*, 2021). This technique is robust and largely accurate. However, it requires stringent sample preparation, involves sophisticated laboratory analysis and is not financially viable at scale (Lu *et al.*, 2021). The literature indicated various applications for species classification from molecular analysis to NIR analysis such as those described by Abasolo *et al.* (2013) and Wan-hong *et al.* (2019), where they classified species and hybrids of *Corymbia*, *E. pellita*, *E. urophylla*, *E. wetarensis* and *E. tereticornis*. Therefore, it is plausible that similar success could be achieved at the nursery level using hybrids of importance in South Africa forestry.

Quality control procedures for seeds and hedges at forestry nurseries are of paramount importance to the holistic sustainability of plantations and forestry. The assessment of seed germination and vigour, hedge nutrient content and hedge clonal purity requires the most improvement. These assessments are generally hindered by destructive sampling resulting in seed loss or decreasing the potential cutting number and mother plant recovery; expensive laboratory costs; human error in analysis and variation in sample mass and sample pre-processing procedures for outsourced chemical analysis. Thus, this study focused on the clones of *Eucalyptus* because it is the major genus planted in South Africa. Through evaluating NIRS for the analysis, this study aims to develop calibration models that can serve as the benchmark for developing future models, thereby contributing to precision forestry in South Africa.

The aim of the study was to investigate the feasibility of NIRS for different applications, i.e., for monitoring nutrient content, clone purity, determination of seed germination rates and vigour, and to compare the accuracy of the NIR models with the reference data.

The research objectives were as follows:

- Development of NIR models to predict seed germination rate of *Eucalyptus dunnii*, *Eucalyptus nitens*, *Eucalyptus benthamii*, and *Eucalyptus macarthurii*.
- Development of NIR models to estimate nutrition levels in nursery seedlings.
- Development of NIR screening methods to monitor varietal purity of clonal cuttings of nursery *Eucalyptus* material.

This dissertation has been written in the form of five chapters. Chapter 1 is a current review of the literature related to this study, followed by Chapters 2, 3, and 4 as experimental chapters, and Chapter 5, which is the dissertation overview of the study. The thesis was formatted per the guidelines of the University of KwaZulu-Natal Discipline of Plant Pathology.

References

- Abasolo, M., Lee, D.J., Raymond, C., Meder, R. and Shepherd, M. (2013). Deviant near-infrared spectra identify *Corymbia* hybrids. *Forest Ecology and Management*, 304, pp.121-131.
- Al-Amery, M., Geneve, R.L., Sanches, M.F., Armstrong, P.R., Maghirang, E.B., Lee, C., Vieira, R.D. and Hildebrand, D.F. (2018). Near-infrared spectroscopy used to predict soybean seed germination and vigour. *Seed Science Research*, 28(3), pp.245-252.
- Ambrose, A., Kandpal, L.M., Kim, M.S., Lee, W.H. and Cho, B.K. (2016). High-speed measurement of corn seed viability using hyperspectral imaging. *Infrared Physics and Technology*, 75, pp.173-179.
- Budić-Leto, I., Gajdoš Kljusurić, J., Zdunić, G., Tomić-Potrebuješ, I., Banović, M., Kurtanjek, Ž. and Lovrić, T. (2011). Usefulness of near-infrared spectroscopy

- and chemometrics in a screening of the quality of dessert wine Prošek. *Croatian Journal of Food science and Technology*, 3(2), pp.9-15.
- Jia, S., Yang, L., An, D., Liu, Z., Yan, Y., Li, S., Zhang, X., Zhu, D. and Gu, J. (2016). Feasibility of analysing frost-damaged and non-viable maize kernels based on near-infrared spectroscopy and chemometrics. *Journal of Cereal Science*, 69, pp.145-150.
- Landis, T.D., Dumroese, R.K. and Haase, D.L. (2010). The container tree nursery manual: Seedling processing, storage, and out planting. vol. 7. *USDA forest Service Agriculture Handbook*, 674, pp.192.
- Li, X., Zhang, L., Zhang, Y., Wang, D., Wang, X., Yu, L., Zhang, W. and Li, P. (2020). Review of NIR spectroscopy methods for non-destructive quality analysis of oilseeds and edible oils. *Trends in Food Science and Technology*, 101, pp.172-181.
- Lu, W., Arnold, R.J., Wang, C., Lin, Y., Luo, J., Meder, R. and Yang, Y. (2021). Defining *Eucalyptus urophylla* with its hybrid and the rules of genetic recombination using near-infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 29(4), pp.236-244.
- Prananto, J.A., Minasny, B. and Weaver, T. (2020). Near-infrared (NIR) spectroscopy as a rapid and cost-effective method for nutrient analysis of plant leaf tissues. *Advances in Agronomy*, 164, pp.1-49.
- Riikonen, J. and Luoranen, J. (2018). Seedling production and the field performance of seedlings. *Forests*, 9(12), pp.740.
- Van Maarschalkerweerd, M. and Husted, S. (2015). Recent developments in fast spectroscopy for plant mineral analysis. *Frontiers in Plant Science*, 6, pp.169.
- Wang, Y., Peng, Y., Zhuang, Q. and Zhao, X. (2020). Feasibility analysis of NIR for detecting sweet corn seeds vigour. *Journal of Cereal Science*, 93, pp.102977.
- Wan-hong, L., Jie, Q. and Jian-Zhong, L. (2019). Application of NIRS for discrimination of *Eucalyptus* hybrids. *Spectroscopy and Spectral Analysis*, 39(5), pp.1451-1457.
- Yizhi Wang, Jianying Xiang, Yu Tang, Wanjun Chen and Yuanjie Xu. (2021): A review of the application of near-infrared spectroscopy (NIRS) in forestry. *Applied Spectroscopy Reviews*, 57(4), pp.300-317.

Zhang, T., Fan, S., Xiang, Y., Zhang, S., Wang, J. and Sun, Q. (2020). Non-destructive analysis of germination percentage, germination energy and simple vigour index on wheat seeds during storage by Vis/NIR and SWIR hyperspectral imaging. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 239, pp.118488

Chapter 1

The applications of near infrared spectroscopy in forestry seedling production: A review

Abstract

The most critical component of nursery seedling production is the use of high-quality seeds. Production of high-quality seeds ensures that the farming system can start with seeds that demonstrated high vigour, are pure of species and are pathogen free. Seed quality influences the production of high-quality seedlings in nurseries. Nurseries are required to monitor several factors such as human error (e.g., resulting in seed species contamination, temperature, irrigation, and the level of nutrients as these factors can influence the seedling's survival and establishment in field conditions. Different methods and tools, which are time-consuming, expensive and can have errors are used to maintain and assess the quality of seeds, seedling production, and pure species identification. Near-infrared reflectance spectroscopy (NIRS) is a tool that is rapid, low-cost, and user-friendly. The NIRS approach requires initial effort to develop the calibration models, but once a NIRS model is developed it can reduce the analysis time for sample materials to seconds. NIRS can be the technology of the future for seed quality analysis. It is cost and time efficient for large and small nurseries to ensure high-quality seedlings for any farming system.

1.1 Introduction

Forestry is fundamental to global ecosystem conservation, as it provide protection to soil and aids in CO₂ fixation (Miura *et al.*, 2015). For millennia, forestry has had a vital role in society, from the provision of basic raw materials for wood products, medicine, and ecotourism (Wang *et al.*, 2021). The world's wood and wood product demand is growing. Therefore, optimisation in forestry practices is required to meet this demand (Keil and Griffin, 1994). One major challenge is improving plantation productivity, considering the limitations of the site, site quality assessment and changes in climate. Exotic species like *Eucalyptus* have become the most utilised and planted hardwoods

around the world for the past years, because they are fast growing rate, with excellent wood and fibre properties (Keil and Griffin, 1994).

Productivity improvement is also influenced by matching the variety of tree crops with the site it is afforested on. The health of the forest is an important concept for sustainable forest management (Xulu, 2019). Seeds are the fundamental input in agriculture and are essential in producing many seedlings (Sangvikar and Wadje, 2012). High-quality seeds result in healthier seedlings, which produce healthier and productive forests (Rahman and Cho, 2016). The quality of the seeds is influenced by the physical and biochemical properties, which influence germination and subsequent quality control scoring (Bradbeer, 1988).

The South African seed industry has increased over the years and constitutes of approximately 107 companies that are members of the South African National Seed Organization (SANSOR) (DAFF, 2017). The success of plant production is the result of high-quality seeds, and the use of clone hybrids that are propagated using superior parent plants through cuttings for producing uniform, high-yield seedlings in nurseries for forest establishment (Azevedo *et al.*, 2021).

The methods for determining seed quality can be complex, however, advancements in technology offers new techniques, which include an ultraviolet spectroscopy, the Raman spectroscopy, a fluorescence spectroscopy, nuclear magnetic resonance (NMR), hyperspectral imaging and near infrared spectroscopy (NIRS). NIRS is broadly attractive as a multipurpose analysing system. It can be an efficient method to assess seed and seedling quality, plant and soil nutrient composition by decreasing sample preparation and analysis time. It is non-destructive, low cost, reliable, clean, and rapid which are all attractive traits for forestry and nursery quality control (QC) analysis techniques (Wang *et al.*, 2021). Improving forestry and nursery QC analysis techniques can result in the plantation of seedlings with higher survival rates, promoting successful establishment of plantations (Grossnickle and MacDonald, 2018).

1.2 Background of forestry and its importance

Forestry includes relationships between people and the resources that the forest provides. The development and maintenance of planted forests is referred to as silviculture (Pallett, 2005; Van Wyk *et al.*, 2012). Forestry plays a significant role in the global ecosystems by providing services such as CO₂ fixation and environmental preservation (Wang *et al.*, 2021). Forests are also planted for products such as wood pulp, sawn wood products, poles, railway sleepers' long, mining poles, tannin extract and other wood products with R38.4 billion revenue for the forestry industry through its numerous products (Statistic SA, 2020). About 31% of the world's land surface is covered by forest plantations (Wang *et al.*, 2021), with 10 million hectares annually estimated for deforestation between the years 2015–2020 (FAO, 2020). Forests host approximately three-quarters of the world's terrestrial biodiversity and help combat climate change (Wang *et al.*, 2021; FAO, 2020). In addition to typical commercial products, forestry is a source of firewood, biofuels, and medicines for more than 1 billion people (Wang *et al.*, 2021).

In South Africa, the introduction of exotic tree species with much faster growth rates when compared to indigenous trees was advantageous to the timber industry and marked the first phase of forest species domestication (Van Wyk *et al.*, 2012). South Africa is considered one of the largest pulp-producing countries globally and is recognised as a significant producer and supplier of pulp for manufacturing printing, writing, packaging, and tissue papers (PAMSA, 2002). The area used for forest planting in South Africa is 1% (1,212,383 ha). Forestry is a large global industry, and South Africa has the largest and oldest plantation assets in the Southern Hemisphere, with over 1.5 million hectares of plantation consisting of both softwood and hardwood species. The South African forestry industry produces various products such as medicinal extracts, sawn timber, mining poles, charcoal, poles, lumber, woodchips, pulp, paper, and cellulose-based products (Van Wyk *et al.*, 2012).

Challenges for forest production are uncontrolled weeds affecting uniform stands and diseases that affect tree quality (Van Wyk *et al.*, 2012). While forests fix carbon dioxide, it must be noted that agricultural and forestry greenhouse gas (GHG) emissions contribute approximately 2.5% of South Africa's emissions (FSA, 2020).

Therefore, any improvement in forestry production efficiency can help reduce these emissions.

1.3 Seedling production for forestry

According to Lavender (1984) “cultivation applications are the physiological guidelines that are essential to produce plants with maximum survival and growth potential”. In forestry, cultivation includes proper seed handling to ensure young seedlings are spaced and watered appropriately to encourage seedling growth and the provision of essential nutrients through the use of fertiliser to enhance seedling growth (Lavender, 1984). Even if all these factors are maintained, the seed quality can have a greater influence on growth outcomes.

The first step in seedling production involves the qualitative assessment of seeds. The total percentage of seed germination can be regarded as a proxy for seed viability and vigour. Seed germination rate refers to the percentage of seeds that sprouted roots and developed into seedlings (Bareke 2018). Seed vigour refers to the rate of germination and the accessible energy reserves in the seed to allow for faster germination (Sivakumar *et al.*, 2002). Alternatively, seed vigour consists of attributes that determine the potential of how rapid seedlings develop under irregular environment conditions (Marcos-Filho, 2015). Seed vigour can be affected by maturity, age, mechanical injury, disease infection, pre-harvest weather, and storage environment (Wen *et al.*, 2018). Seed vigour and seed quality is evaluated using various techniques, such as vitality index, accelerated ageing tests, cold test, genetic purity tests, and tetrazolium tests (Guan *et al.*, 2013; Wang *et al.*, 2021). While there were no universally accepted test methods for seed vigour for all seed types, currently accepted methods are considered to be the standard germination test and accelerated ageing test. Seed vigour is more complex than seedling vigour, therefore, seedling vigour trends are extrapolated from seed vigour (Guan *et al.*, 2013). Standardised laboratory germination methods are criticised for not predicting field performance (Shaban, 2013; Sivakumar *et al.*, 2002). This is because field conditions can be highly variable and therefore not easily replicated or predictable. This is one of many challenges facing seed testing for optimal seedling production.

1.4 Seedling germination

Seed germination is defined as the emergence of the radicle through the seed coat and represents the observed germination under near-optimal environmental conditions (Davies *et al.*, 2015). Seed vigour is determined by the sum of seed properties, including the carbohydrate, fats, minerals, and proteins, which control the performance and activity of the seed or seed lot during germination and seedling emergence (Wen *et al.*, 2018). Seed vigour has attributes that contribute to rapid seedling development under variable field conditions (Marcos-Filho, 2015). Seed vigour is affected by maturity, age, mechanical injury, disease infection, pre-harvest weather, and storage environment.

Traditional methods of assessing seed quality are highly dependent on seed germination test which is based on the understanding that higher germination corresponds to higher seed vigour (Davies *et al.*, 2015; Wen *et al.*, 2018). A study published by Yasmin *et al.*, (2019), who investigated the use of Fourier transform near-infrared reflectance spectroscopy (FT-NIRS) as tool to discriminate viable and non-viable triploid watermelon seeds from three different cultivars with four-year (natural ageing) storage under controlled conditions (Yasmin *et al.*, 2019). Near infrared reflectance spectroscopy has been widely accepted as tool capable of improving agricultural practices and potentially having a place in forestry as well (Wang *et al.*, 2021).

1.5 Seedling quality assessment procedures

The quality of the seedlings is the result of sowing high-quality seeds, which have high vitality determined by various evaluation methods. In addition, the seedling performance is dependent on the growth performance of the tree species (Gregorio *et al.*, 2010). Seedlings can be considered high quality if they meet certain morphological and physiological characteristics (Duryea, 1984). The Plant Quality Index (PQI) is one of the methods that includes both morphological and physiological attributes, including plant age, root plug (plug integrity), and plant health (leaf colour and no signs of disease). The morphological characteristics used to measure quality seedlings are shoot heights, root mass, stem diameter, root mass, and shoot-to-root ratio (Duryea, 1984). A minimum score for a plant lot is a PQI of 60 and a maximum score of 100 (Van Wyk *et al.*, 2012). The advantages of high-quality seedlings include lower

plantation maintenance, low mortality, high timber recovery and quality as well as shorter rotation period (Gregorio *et al.*, 2010). Producing quality seedlings ensures the success of seedling stock establishment and plantation productivity.

1.6 Morden technology

Keeping up with major scientific developments and incorporating modern technology is indispensable for enhancing efficiency and sustainability in forestry. The most important part of precision forestry is the use of modern technologies, which provide highly repeatable measurements and optimized processes to initiate, cultivate and harvest trees. It provides valuable information with links between resource managers, the environment, producers, and public policy makers (Dyck, 2001). Precision forestry focuses on the incorporation of new technology to advance current assessment methods and holistic forestry practices (Pallet *et al.*, 2005).

The most well-known and widely used modern technology tools are remote sensing (CIR, Airborne Laser Scanner) and satellite imagery. Remote sensing technologies involve the use of visible, near-infrared, and short-wave infrared images (Singh *et al.*, 2022). These technologies use GPS to obtain accurate coordinates for drone flight that aid in preparations for new forest planting, navigation systems and geographic information systems (GIS) working with vast amounts of data in the form of maps, images, digital products, GPS, text data, and tabular data that can be obtained from multiple sources (Kovcsov and Antalov, 2010). Research has shown that remote sensing can be used for real-time canopy chemistry and plant function traits (Singh *et al.*, 2022). Real-time scanners such as radio frequency identification (RFID) and field mapping forest structures with long-term monitoring that also provide data on tree height, canopy eminence and profile, trunk profile, timber volume estimation, and other aerially assessed properties most using laser and NIR to scan (Arsner *et al.*, 2011). Satellite imagery technologies allow the extraction of data from high-resolution imagery, multispectral and hyperspectral images. Remote sensing provides an avenue for rapid, accurate, and cost-effective analysis, which has revolutionized various industries in terms of monitoring and analysis (Kovcsov and Antalov, 2010).

1.7 Near-infrared spectroscopy (NIRS)

NIRS is a screening tool that utilises the interaction of bonds in the NIR to calibrate prediction models, which then predict the properties of a sample. The radiation from the electromagnetic range of NIR spectroscopy is between 780-2500 nm (Lopez *et al.*, 2013; Huang *et al.*, 2008). It is found between the short and longer wavelengths of visible light and mid-infrared (MIR) light. It is characterized by the presence of overtones and combination bands (Lopez *et al.*, 2013). Absorption bands appear because of the vibration activity in the infrared range, resulting in overtones, combinations, and fundamental vibrations (Tigabu and Oden, 2004; Ciurezak, 2001). The NIR region consists of overtones and combinations of bands, due to vibrations of specific functional groups, i.e., CH, OH, NH, and SH bonds (Dyrby *et al.*, 2002) (Table 1.1). These functional groups are major components of biological material (Tigabu and Oden, 2004). In addition to chemical information on the NIR spectra, physical properties can also affect the estimation of properties such as seed bulk density and seed weight (Velasco *et al.*, 1998; Font *et al.*, 1999).

Near-infrared reflectance spectroscopy is a recognised and established tool for rapid quantitative and qualitative analysis of a variety of materials for many chemical and physicochemical properties (Tigabu and Oden, 2004). This can be attributed to the use of machine learning algorithms that predict information about the sample from its NIR spectrum (Lopez *et al.*, 2013). Among the first recognised industrial transformations using NIRS was by Karl Norris using grains and seeds in the 1960's (Esteve and Hurburgh, 2014). Near-infrared spectroscopy technology has a wide application, and it is used in food science, seed quality inspection, and technology research, to name a few (Singh *et al.*, 2022). In modern technology, NIRS analysis is a widely used method because of its advantages of being non-destructive, rapid, low cost, requiring minimal or optional sample preparation and can have high precision (Caporaso *et al.*, 2018; Brown *et al.*, 2020). The performance of NIRS is dependent on the accuracy of the calibration model developed and can become comparable to the wet chemistry analysis methods used to develop the model (Esteve and Hurburgh, 2014).

Table 1.1: An Infrared spectral fraction, showing stretching overtones in wavelengths region between 780 and 2500 nm (12,500–4000 cm^{-1} , a wavenumber) (Francisco *et al.*, 2017).

Group	Aliphatic hydrocarbons	Aromatic hydrocarbons	Carboxylic acid	Amines.	Water
Frequency range (cm^{-1})	9100–7800 (overtone of CH-stretching)	ca. 9000 (overtone of CH-stretching)	ca. 6900 (overtone of CH-stretching)	7000–6500 (overtone of NH-stretching)	7500–6400 (overtone of OH-stretching)
	7700–6900 (combination)	7300–6900 (combination)	ca. 5250 (overtone of CO-stretching)	5200–4500 (combination)	5400–4900 (combination)
	6300–5500 (overtone of CH-stretching)	ca. 6000 (overtone of CH-stretching)	4900–4600 (combination)		
	5000–4100 (combination)	4700–4000 (combination)			

NIRS can perform best under laboratory conditions compared to field conditions due to the interference of external factors such as the humidity, temperature, solar radiation, and sample orientation.

NIR spectra consist of broad wavebands resulting from overlapping absorbances that allow the detection of organic and biological materials (Xia *et al.*, 2019). During NIR sample scanning, radiation interacting with a sample can be absorbed, transmitted, or reflected to the detector (Esteve and Hurburgh, 2014). The sample can be presented as a ground powder, a solid (seed), or a fresh sample (leaf), although the fresh leaf material has the disadvantage of moisture that can alter spectra reading.

The intensity of the reflected spectra is affected by sample type, and the absorption process affects the shape of reflected spectra. Different NIR spectroscopy

measurement modes are suitable for various applications (Nicolai *et al.*, 2007). Karl Norris developed instruments and used computers to demonstrate that the NIR region is particularly useful for quantitative analysis, especially of agricultural samples (Hurburgh, 2014; Xia *et al.*, 2019). Researchers have focused on determining seed viability rather than seed vigour using tools such as NIRS, hyperspectral imaging (HSI) or multispectral imaging (MSI), Raman spectroscopy (RS), infrared thermography (IRT), and soft X-ray imaging (Xia *et. al.*, 2019).

1.8 The use of NIR

Several research studies have shown that NIR spectroscopy has enormous potential to help monitor and improve seed lot viability, vitality, or purity. This includes detection of defective seeds and non-viable seeds from seed lots using single seed NIRS, incorporated with advanced chemometric algorithms based on multivariate analysis (Schmitt *et al.*, 2014). Analytical information expressed in NIR spectra is not easy to detect and select. Extraction of useful information quickly and efficiently from spectral data sets requires chemometric algorithms to extract information for quantitative or qualitative analysis (Lestander and Oden, 2002).

Several studies show that NIRS in combination with multivariate analysis is suitable for characterizing groups of similar biological materials, such as uninfected and infected wheat kernels (Dowell *et al.*, 1998; Ridgway and Chambers, 1998; Tigabu and Oden, 2002); healthy and insect-infested seeds of *Cordia africana*; discriminating between dead and viable seeds of *Pinus sylvestris* L. (Lestander and Oden, 2002); analysis of empty and viable seeds of *P. patula schiede* and *deppe* by Tigabu and Oden, 2003); and identifying healthy and insect-damaged seeds of *Albizia schimperiana* olive (Tigabu and Oden, 2003).

Tigabu and Oden (2004) showed the feasibility of NIRS as a rapid, convenient and non-destructive tool for seed vigour analysis on *P. patula*. Near-infrared transmission spectra ($\log 1/T$) collected from individual seeds and Soft Independent Modelling of Class Analogy (SIMCA) and PLS-DA (Partial Least Squares Discriminant Analysis) were used to assess the NIR spectra. The NIR absorption bands correlated with lipids,

proteins and other compounds were also considered for classifying unaged and aged seeds (Tigabu and Oden, 2004).

The main goal of Tigabu *et al.*, (2017) was the improvement in seed set performance for fast, constant germination and seedling establishment of juniper trees in the nursery. They also evaluated the potential of visible and near-infrared (VIS + NIR) spectroscopy for the detection of insect-damaged seeds of *Juniperus procera* on a dry basis, and healthy as well as damaged seeds were also distinguished accurately. VIS + NIR Spectroscopy distinguished healthy and damaged seeds with an error of 10%. It has also been shown to identify the characteristic gene quality of seed lots (Tigabu *et al.*, 2005).

1.9 Clone identification

Eucalyptus plantations in China are an example of the success of clonal forestry. In the 1970s only about 300,000 ha of *Eucalyptus* plantations were cultivated, increasing slowly to 400 000 ha around 1980 and 700,000 ha during 1990's (Wu, 2019). The advantage of using vegetative propagation has been long recognized (for 1500 years) (Baolong and Taigang, 1988). The development and establishment of clonal-based forestry is phenomenally successful for various hardwood species, especially tropical and subtropical *Eucalyptus* trees, due to their fast vegetative reproduction and short rotation for instance in Brazil, about 50% of *Eucalyptus* plantations are of clonal origin (Wu, 2019). Two main companies promoting clonal forestry are New Zealand Forest Genetics and Australia Arbor Gen Australasia, each producing approximately 20 new clones from field trials each year (Wu, 2019).

The Dongmen Project (1982-1989) demonstrated the use of hybridisation and selection of *Eucalyptus* species clones, which enabled the *Eucalyptus* plantations in China to increase to 1.5 million ha in 2005. By 2012, it reached 3.7 million ha with 17 clones used to establish 2.6 million ha (Huang *et al.*, 2012). The cloned forest established from the same starting material consists of ramets that are genetically identical to each other. Traditional plantations from seed-based seedlings can consist of trees that are the same species, but each tree can be genetically different from others due to the combination of their parents' genes. Hence, it can be stated that

uniformity is a characteristic of a genetic control of tree plantation, while variation is a characteristic of natural plantations (Mushtaq *et al.*, 2017).

Rooting from stem cuttings was unsuitable for economically important forest tree species (Mushtaq *et al.*, 2017). Somatic embryogenesis is the process where embryos are formed from somatic cells without any genetic fusion (Haines and Woolaston, 1991). This clonal method is beneficial for large-scale production for genetic improvement. Another clonal method is tissue culture, where there is an aseptic *in vitro* culturing of tree tissues, organs, or whole plants grown in a controlled environmental condition (Thorpe, 2007). Clonal plantations can make a major contribution to the development of wood-based industries, with clonal technology also supported by an improved package of forestry techniques, where they offer great opportunities for substantial improvements in forestry plantation productivity and significant increases in plantation timber quality (Mushtaq *et al.*, 2017).

Da Silva *et al.*, (2021) examined Hevea gum clones to recognise and distinguish the leaves at different developmental growth stages and to distinguish between healthy and unhealthy adult leaves infected by *Microcyclus ulei* (South American leaf blight) using visible to infrared (Vis-NIR) along with chemometric methods. They used dry ground leaf samples from 14 tree rubber clones (*H. brasiliensis*), to develop classification models (Da Silva *et al.*, 2021).

1.10 Impact and future work

Several feasibility studies have been conducted to pave the way for further investigations for NIR possible use in the near future. With NIR universally used in the pharmaceutical industries, it makes sense that it can be used in agriculture to supplement wet chemistry for rapid results and monitor crucial aspects of quality seed and seedling production. NIR can assist in establishing silviculture in a small turnaround time, reducing the input cost for the forestry and agricultural industry as well as ensure high-quality products. Therefore, NIR reflectance spectroscopy can be used for nursery production.

1.11 References

- Agelet, L.E. and Hurburgh Jr, C.R. (2014). Limitations and current applications of Near Infrared Spectroscopy for single seed analysis. *Talanta*, 121, pp.288-299.
- Baolong, H. and Taigang, L. (1988). Preliminary study on cultural history of Chinese Fir. *Journal of Nanjing Forestry University*, 31(02), pp.54.
- Bareke T. (2018). Biology of seed development and germination physiology. *Advance in Plants and Agriculture Research*. 8(4), pp.336–346.
- Boswell V.R. (1961). The importance of seed, Yearbook of agriculture.
- Bradbeer J.W. (1988). Seed Dormancy and Germination. Boston, MA, Springer US (United States).
- Caporaso N, Whitworth M.B., and Fisk I.D. (2018). Near-Infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains, *Applied Spectroscopy Reviews*, 53(8), pp. 667-687.
- Craig Norris. (2014). Forestry Facts: Wattle rust disease News and Views publication
- Da Silva, A.L.D., Alves Filho, E.G., Silva, L.M.A., Tavares, O.C.H., Pereira, M.G., de Campos, T., and da Silva, L.M. (2021). Near-infrared spectroscopy to rapidly assess the rubber tree clone and the influence of maturation and disease on the leaves. *Microchemical Journal*, 168, pp.106478.
- Davies, R., Di Sacco, A. and Newton, R. (2015). Germination testing: procedures and evaluation. *Technical Information Sheet_13a. Royal Botanic Gardens, Kew.*
- Department of Agriculture, Forestry and Fisheries. (2017). South African grain seeds market analysis report.
- Dowell F.E., Throne J.E., and Baker J.E. (1998). Automated non-destructive detection of internal insect infestation of wheat kernels by using near-infrared reflectance spectroscopy. *Journal of Economic Entomology*. 91, pp. 899–904.
- Dyck, B. (2003). Precision forestry—The path to increased profitability. In *Proceedings of the Second International Precision Forestry Symposium. University of Washington, Seattle, WA* Vol. 2003, pp. 3-8.
- Dyrby M., Engelsen S.B., Nørgaard L., Bruhn M., and Lundsberg-Nielsen L. (2002). “Chemometric quantitation of the active substance (containing $C\equiv N$) in a pharmaceutical tablet using near-infrared (NIR) transmittance and NIR FT-Raman spectra.” *Applied Spectroscopy*, 56, pp. 579-585.
- Esteve Agelet L., and C.R. Hurburgh. (2014). Limitations and current applications of Near Infrared Spectroscopy for single seed analysis. *Talanta* 121, pp. 288-299

- FAO, Global Forest Resources Assessment (2020): Main Report; FAO: Rome.
- Forestry South Africa (FSA), Annual Report 2020
- García-Sánchez, F., Galvez-Sola, L., Martínez-Nicolás, J.J., Muelas-Domingo, R. and Nieves, M. (2017). Using near-infrared spectroscopy in agricultural systems. *Kyprianidis KG, Skvaril J. Developments in near-infrared spectroscopy. London: Intech Open*, pp.97-127.
- Grossnickle, S.C. and MacDonald, J.E., (2018). Seedling quality: history, application, and plant attributes. *Forests*, 9(5), pp.283.
- Guan Y.J., Jhen-Jia H., Wung Z., Zhu S.J., Wang J.C., and Knapp A. (2013) Time series regression analysis between changes in kernel size and seed vigour during the developmental stage of sh2 sweet corn (*Zea mays* L.) seeds. *Sci. Hortic. Amsterdam* 154, pp. 25–30.
- Huang, H., Yu, H., Xu, H., *et al.* (2008). Near-infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: a review. *Journal Food Engineering*. 87 (3), pp 303–313.
- ISTA (2009). International Rules for Seed Testing. International Seed Testing Association, Switzerland.
- Kováčsová, P. and Antalová, M. (2010). Precision forestry—definition and technologies. *Šumarski list*, 134(11-12), pp.603-610.
- Lavender, D.P. (1984). Plant Physiology and Nursery Environment: Interactions Affecting Seedling Growth, Chapter 14 In Duryea. Mary L., and Thomas D. Landis (eds.). *Forest Nursery Manual: Production of Bareroot Seedlings*. Martinus Nijhoff/Dr W.
- Lee H., Jeon Y., Lee Y., and Lee G. (2017). Large-scale screening of intact tomato seeds for viability using near-infrared reflectance spectroscopy (NIRS). *Sustainability*, 9:618.
- Lestander T.A., and Odén P.C. (2002). Separation of viable and non-viable filled Scots pine seeds by differentiating between drying rates using single seed near-infrared transmittance spectroscopy. *Seed Science and Technology*. 30, pp.383-392.
- López, A., Arazuri, S., García, I., Mangado, J. and Jarén, C. (2013). A review of the application of near-infrared spectroscopy for the analysis of potatoes. *Journal of agricultural and food chemistry*, 61(23), pp.5413-5424.

- Marcos-Filho J. (2015). Seed vigour testing: an overview of the past, present, and future perspective. *Scientia Agricola* 72, pp. 363-374.
- Mushtaq, T., Banyal, R., Mugloo, J., Mushtaq, T. and Aziz, M.A. (2017). Clonal forestry: An effective technique for increasing the productivity of plantations. *SKUAST Journal of Research*, 19(1), pp.22-28.
- Neupane R.K., and Gotame T.P. (2019). A Classification of Agronomic Crop Species based on Use Values. Proceedings of a National Workshop on Working Crop Groups, Kathmandu (Conference Paper).
- Nicolia B.M., Beullens k., Bobelyn E. (2007). Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biological Technology* 46(2), pp. 99-118.
- Nikolaeva, M.G. (1977). Factors controlling the seed dormancy pattern. *Physiology and Biochemistry of Seed dormancy and Germination*.
- Pallett, R.N. (2005). Precision forestry for pulpwood re-establishment silviculture. *Southern African Forestry Journal*, 203(1), pp.33-40.
- Rahman A., and Cho B.K. (2016). Assessment of seed quality using non-destructive measurement techniques: a review. *Seed science research* 26, pp.285-305.
- Ridgway C., and Chambers J. (1998). Detection of insects inside wheat kernels by NIR imaging. *Journal of Near Infrared Spectroscopy*. 6, pp.115-119.
- Robert Neil Pallett. (2005). Precision forestry for pulpwood re-establishment silviculture, *Southern African Forestry Journal*, 203(1), pp. 33-40.
- Sangvikar V.R., and Wadje S.S. (2012). In vivo testing of plant extracts against seed-borne pathogens. *International Research Journal of Biological Sciences* 1(6), pp.1-4.
- Shaban, M. (2013). Study on some aspects of seed viability and vigour. *International Journal of Advanced Biological and Biomedical Research*, 1(12), pp.1692-1697.
- Sidhu H.K., Haagenson D.M., and Wiesenborn D.P. (2012). Non-destructive analysis of single plant canola (*Brassica napus*) seeds using near infra-red spectroscopy. ASABE Meeting Presentation. 12-1337308.
- Sivakumar V., Parthiban K.T., Singh B.G., Gnanambal V.S., Anandalakshmi R., and Geetha S. (2002). Variability in drupe characters and their relationship to seed germination in teak (*Tectona grandis* L.F.). *Silvae Genetics* 51, pp.232-237.

- Tigabu M., and Odén P.C. (2002). Multivariate classification of sound and insect-infested seeds of a tropical multipurpose tree, *Cordia Africana*, with near-infrared reflectance spectroscopy. *Journal of Near Infrared Spectroscopy* 10, pp.45-51.
- Tigabu M., and Oden P.C. (2003). Discrimination of viable and empty seeds of *Pinus patula* Schiede and Deppe with near-infrared spectroscopy. *New Forests* 25, pp. 163-176.
- Tigabu M., Oden P.C., and Lindgren D. (2005). Identification of seed source and parents of *Pinus sylvestris* L. using Visible near-infrared reflectance spectra and multivariable analysis trees. 19, pp.468-476.
- Tigabu, M., Daneshvar, A., Karimidoost, A. and Odén, P.C. (2017). Flotation techniques to improve the viability of *Juniperus polycarpos* seed lots. *Journal of Forestry Research*, 28(2), pp.231-239.
- Thorpe, T.A. (2007). History of plant tissue culture. *Molecular biotechnology*, 37(2), pp.169-180.
- Van Wyk, A., Brink, A., Van Huyssteen, G. and Jansen, J. (2012). Tree farming guidelines for private growers. *Cascades: Sappi Forests*.
- Van Wyk, B. and Gericke, N. (2000) People's Plants. A Guide to Useful Plants of Southern Africa. Briza Publications, Pretoria, 352.
- Wang, Y., Xiang, J., Tang, Y., Chen, W. and Xu, Y. (2021). A review of the application of near-infrared spectroscopy (NIRS) in forestry. *Applied Spectroscopy Reviews*, pp.1-18.
- Wen D., Hou H., Meng A., Meng J., Xie L., and Zhang C. (2018) Rapid evaluation of seed vigour by the absolute content of protein in seed within the same crop. *Scientific Report*, 8, pp.55-69.
- Wu, H.X. (2019). Benefits and risks of using clones in forestry—a review. *Scandinavian Journal of Forest Research*, 34(5), pp.352-359.
- Xia Y., Xu Y., Li J., Zhang C., and Fana S. (2019). Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artificial Intelligence in Agriculture*, 1, pp.35-47.
- Yasmin J., Raju Ahmed M., Lohumi S., Wakholi C., Kim M.S., and Cho B.K. (2019). Classification Method for Viability Screening of Naturally Aged Watermelon Seeds Using FT-NIR Spectroscopy. *Sensors* 19(1190), pp.1-14.

Chapter 2

Screening of *Eucalyptus* species for seed germination rate and vigour using near-infrared reflectance spectroscopy

Abstract

Seed germination and vigour are important characteristics that impact the success rate of hedges and the vitality of emerging plants. However, these seed assessments are destructive, prone to interference by changes in incubation conditions, and limited to a small sample size. This problem is more serious in the propagation of *Eucalyptus* hedges in nurseries, which feeds into expansive plantation establishments associated with South African Forestry. Near Infrared Spectroscopy (NIRS) was proposed as a viable tool for these assessments as it can allow for a rapid prediction of seed germination and vigour rate in a non-destructive and cost-efficient manner. Hence, this study aimed to evaluate the feasibility of NIRS in predicting *Eucalyptus* seed germination and seed vigour. Species included in the study were *E. nitens*, *E. dunnii*, *E. benthamii* and *E. macarthurri*, which were accessed from 191 seed lots. Seed germination and vigour were determined through seed germination assays in 96-well culture plates for 10 days and used as the reference data for model development and validation. The results indicate that NIRS can predict the seed germination with high accuracy in most of the species tested (RMSEP: 0.357 – 73.5%, R^2 : 58.46 – 95.42%) and collectively as mixed species model (RMSEP: 11.4 – 18.8%, R^2 : 73.68 – 84.99%). Also, the results for the seed vigour indicate that NIRS can predict with accuracy for the species tested (RMSEP: 5.68 – 126, R^2 : 78.91 - 83.57%) and collectively as a mixed species model (RMSEP: 18.8, R^2 : 71.37%). These developed NIR models showed promise for enhancing *Eucalyptus* seed germination and vigour test. This study was the first to report using NIR as a viable tool for seed germination and vigour in nursery operations prioritising *Eucalyptus* plantations.

2.1 Introduction

The success of plantations depends on high-quality seeds, which ensure healthy forests, vigorous seedlings, and efficient forestry operations while supporting industry sustainability (da Costa, 2021; Reed *et al.*, 2022). Therefore, the sustainability of the *Eucalyptus* plantations depends on producing healthy seedlings to replace felled trees or nursery stocks.

Seed viability is the first assessment conducted in establishing nursery hedges, which then get propagated and planted to establish a forest. Therefore, knowing the proportion of viable seed lot determines the number of seeds sown to fill a bed or tunnel. Current germination tests involve sowing seeds in 100 well seedling trays and incubating them for 10-15 days. Thereafter, the germination rate of the seed lot is determined and used for propagation. However, this typically results in underplanting as *Eucalyptus* seed lots can have more than 100 000 seeds per kg, indicating that the standard method for evaluating seed germination is not extensive enough. Furthermore, these methods are time-consuming, use trays which take up space, which are not well insulated against variation in environmental conditions and are consequently prone to error in assessing seed germination (Bareke 2018). Therefore, new methods for seed germination need to be evaluated.

Seed vigour, defined as “the sum of those properties that determine the activity and performance of seed lots of acceptable germination under various environmental conditions” (Marcos, 2015; ISTA, 2015), is vital for achieving desired yields and plant densities. Vigorous seeds germinate rapidly and develop into healthy seedlings, but factors such as storage conditions and harvest protocol can reduce rate of seed vigour. Current vigour testing methods include physiological, biochemical, and physical tests. Standard laboratory germination tests evaluate seeds under optimal conditions, while stress tests like cold and accelerated ageing assess performance under adverse conditions (Milosevic *et al.*, 2010). Biochemical methods, such as tetrazolium testing, enzyme activity analysis, and advanced techniques like protein electrophoresis and high-performance liquid chromatography, provide additional insights (Kumar *et al.*, 2018). However, these methods are labor-intensive, time-consuming, and often destructive, limiting their practicality for large-scale seed quality assessments. Additionally, no single test can reliably predict seed performance, highlighting the need for innovative approaches.

Near-infrared spectroscopy is a promising non-destructive analytical technology widely used in the pharmaceutical industry and is increasingly applied to seed, soil, and food analysis. NIRS uses absorbance patterns in the 780–2500 nm range to detect interactions between chemical bonds and profile samples, and their properties (Abbas *et al.*, 2020). This can be used to generate NIR models capable of rapid, cost-effective, and highly accurate assessments of samples, such as *Eucalyptus* seeds seed lots (Caporaso *et al.*, 2018). NIRS is a valuable analytical tool, the model development is aligned with statistical algorithms that are applied to correlate spectral signatures (Hein and Parrish, 2020). The NIR spectra is made up of overlapping wavelengths, therefore, multivariate statistical techniques are used like Principal Component Analysis, Partial Least Squares etc. Also, this in association with statistic concepts including used of Coefficient of Determination, Root Mean Square Error of Prediction, Standard Error of Calibration (SEC), Standard Error of Cross-Validation (SECV), Ratio of Performance to Deviation (RPD) (Xai *et. al.*, 2019 and Hein and Parrish, 2020).

Although of informative value, current seed vigor testing methods are limited by their labor-intensive, time-consuming, and often destructive nature. Therefore, a study was established to evaluate the feasibility of using NIRS as a rapid and non-destructive tool to predict *Eucalyptus* seed germination and vigour. This advances seed assessment methods by evaluating new technologies, like NIRS.

2.2 Materials and methods

2.2.1 Sample identification and sub-sampling

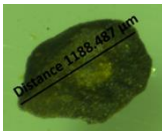
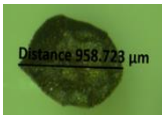
The seed lots used in this study were collected and stored over different years by the Institute for Commercial Forestry Research (ICFR) in their seed bank. Species used in this study included several *Eucalyptus* varieties of seeds of interest to the South African forestry industry *E. macarthurii*, *E. dunnii*, *E. nitens*, and *E. benthamii* (Table 2.1).

2.2.2 Seed preparation and spectral acquisition

Seed samples of test varieties (2 g) were transferred into quartz vials (diameter = 20 mm) and were left overnight to climatise to the temperature of 19°C and its ambient

humidity. The samples were analysed using an integrating sphere on the Bruker MPA I FT-NIR spectrometer (Bruker Optics, USA) at the ICFR Forestry Research and Analytical Laboratory (FRAL). NIRS operation and spectral acquisition were managed using OPUS Lab software (v.6.5, Bruker, USA). Spectra were acquired in the range of 14963 cm⁻¹ to 3574 cm⁻¹ with a resolution of 8 cm⁻¹. Each sample was scanned in triplicate, with hand inversion of the vials between scans to enhance NIR sample variability. To ensure high-quality spectral signatures, each spectrum represented an average of 32 scans.

Table 2.1: Description of *Eucalyptus* seed lots used.

Species	Number of seed lots	Collection date	Seed lot mass (g)	Seed shape
<i>E. macarthurii</i>	13	2013-2020	0.161	
<i>E. dunnii</i>	13	2017-2019	0.210	
<i>E. nitens</i>	14	2022	0.193	
<i>E. benthamii</i>	130	2003-2019	0.105	

The collection date represents the year(s) in which seeds were harvested. The seed lot mass represents the cumulative mass of 300 seeds for the respective *Eucalyptus* varieties. Pictures of the seeds were provided as qualitative descriptors of the seed dimension.

2.2.3 Assessment of seed germination and vigour

Seed germination assessments were conducted according to ISTA (2010) with modifications and adaptations to accommodate feasibility. One hundred seeds were sampled from the quartz vials just after they were scanned on the Bruker MPA NIRS. The seeds were surface sterilized by soaking them in a 0.5% bleach solution for 3 min and rinsed with deionised water (Lindsey *et al.*, 2017). The surface sterilized seeds were placed on sterile 96-well micro-titre plates containing 100 μL of water agar (1.5%

bacteriological agar) per well, at a density of 1 seed per well. The plates were then placed in zip-lock bags and incubated at $25 \pm 1^\circ\text{C}$ with a 16:8-hour light-dark cycle. The plates were observed at day 3, 7 and 10 intervals for germination. All germination assessments were conducted in duplicate and then expressed as an average of the replicates. Seed vigour was determined as the cumulative sum of the percentage of seeds that germinated at day 3, 7 and 10 intervals. Was calculated by formular adding each frequency from a frequency distribution table to the sum of its predecessors.

2.2.4 NIR model development and validation

NIR calibration models for seed germination and vigour were created using the OPUS software (v. 6.5, Bruker, USA). Wavelengths of interest were selected using Principal Component Analysis (PCA), prior model development. Calibration and validation datasets were created by splitting the data in a 9:1 ratio, respectively. The Quant-2 module was used to create partial least squares (PLS) regression models by calibrating the NIR seed spectra with the reference seed germination or vigour data. In each iteration of model optimisation, all spectral pre-processing options available were selected, viz: first derivative, second derivative, vector normalisation (SNV), Multiplicative Scatter Correction (MSC) or a combination of them. The accuracy of developed models was assessed using the root mean squared error of prediction (RSMEP) from the validation dataset and the R^2 was used as a qualitative assessment of the trend between the predicted and actual germination values. All models were optimised to have the lowest RSMEP.

2.3 Results

This section presents the performance and predictive capabilities of NIRS models developed to assess *Eucalyptus* seed germination and vigour. The models were built to explore the relationship between seeds NIR spectral signatures and the predictability of seed germination or vigour, assessed by the models RMSEP. These findings provide insights into the potential of NIR spectroscopy as a rapid, non-destructive tool for evaluating *Eucalyptus* seed quality.

2.3.1 NIR model performance for predicting *Eucalyptus* seed germination

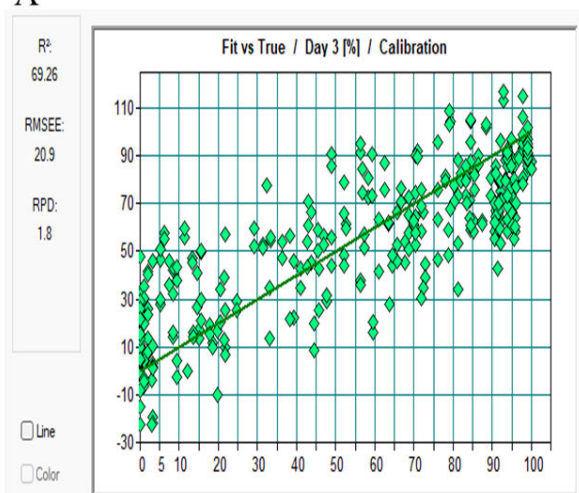
Modelling of the *Eucalyptus* seed germination percentage was accomplished for all varieties for day 3, 7 and 10 intervals (Table 2.2 – 2.6, respectively). Global seed

germination models showed higher levels of prediction error but had a good correlation ($R^2 > 70\%$) with the reference germination values. The spectral regions associated with seed germination were found within $3965.1 - 10283.1 \text{ cm}^{-1}$ and mostly concentrated around $4200 - 8900 \text{ cm}^{-1}$. Within these regions, spectral data was commonly processed using SNV or first derivative to acquire optimised calibration and validated models with the lowest RMSEP (Figure 2.1).

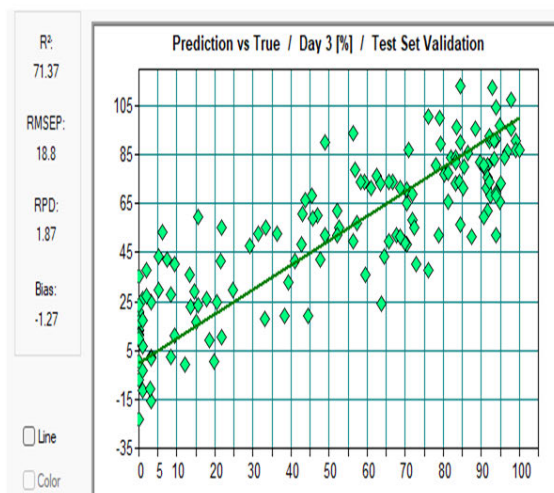
The performance metrics of day 3 germination models are shown in Table 2.2. Models of all varieties used absorbance values within $4242.9 - 10290.9 \text{ cm}^{-1}$ and required a first derivative, SNV, or a combination of both to develop optimised models. The RMSEP of the models ranged from $5.8 - 18.8\%$ and showed a good correlation ($R^2 > 80\%$) with the reference germination data. The most accurate (RMSEP: 5.8%) and well-correlated ($R^2 = 81.3\%$) model was developed for *E. macarthurri* despite having a low sample number ($n = 13$). To increase the variation, and subsequently complexity, captured by the model, all species were incorporated into the development of the day 3 global germination model, which showed slightly higher levels of prediction error (RMSEP: 18.8%) and lower levels of correlation ($R^2 = 71\%$) compared to the models developed for single varieties.

A

PLS Calibration plot

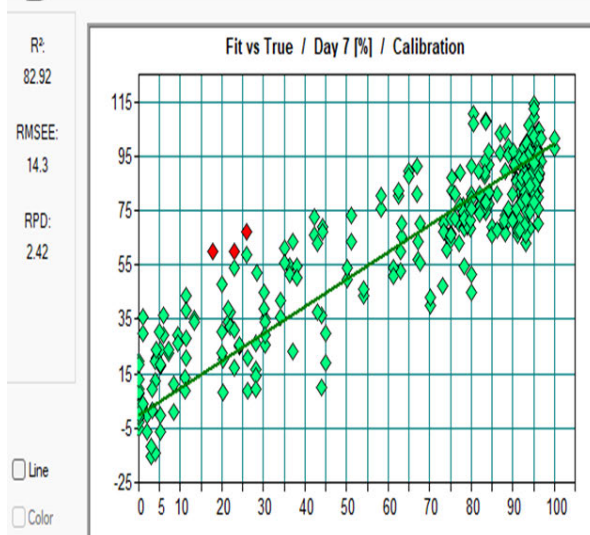


PLS Validation plot

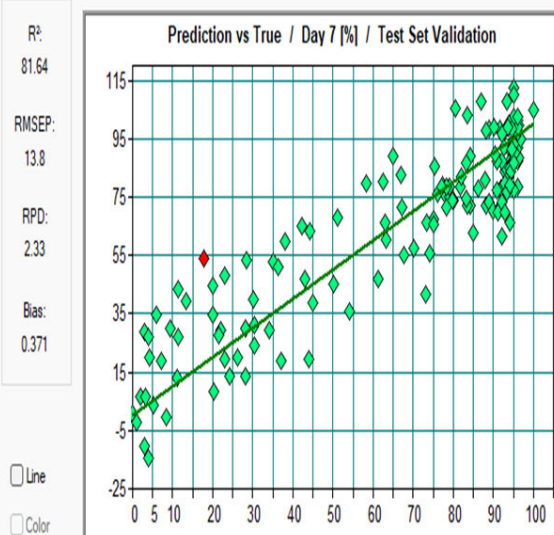


B

PLS Calibration plot



PLS Validation plot



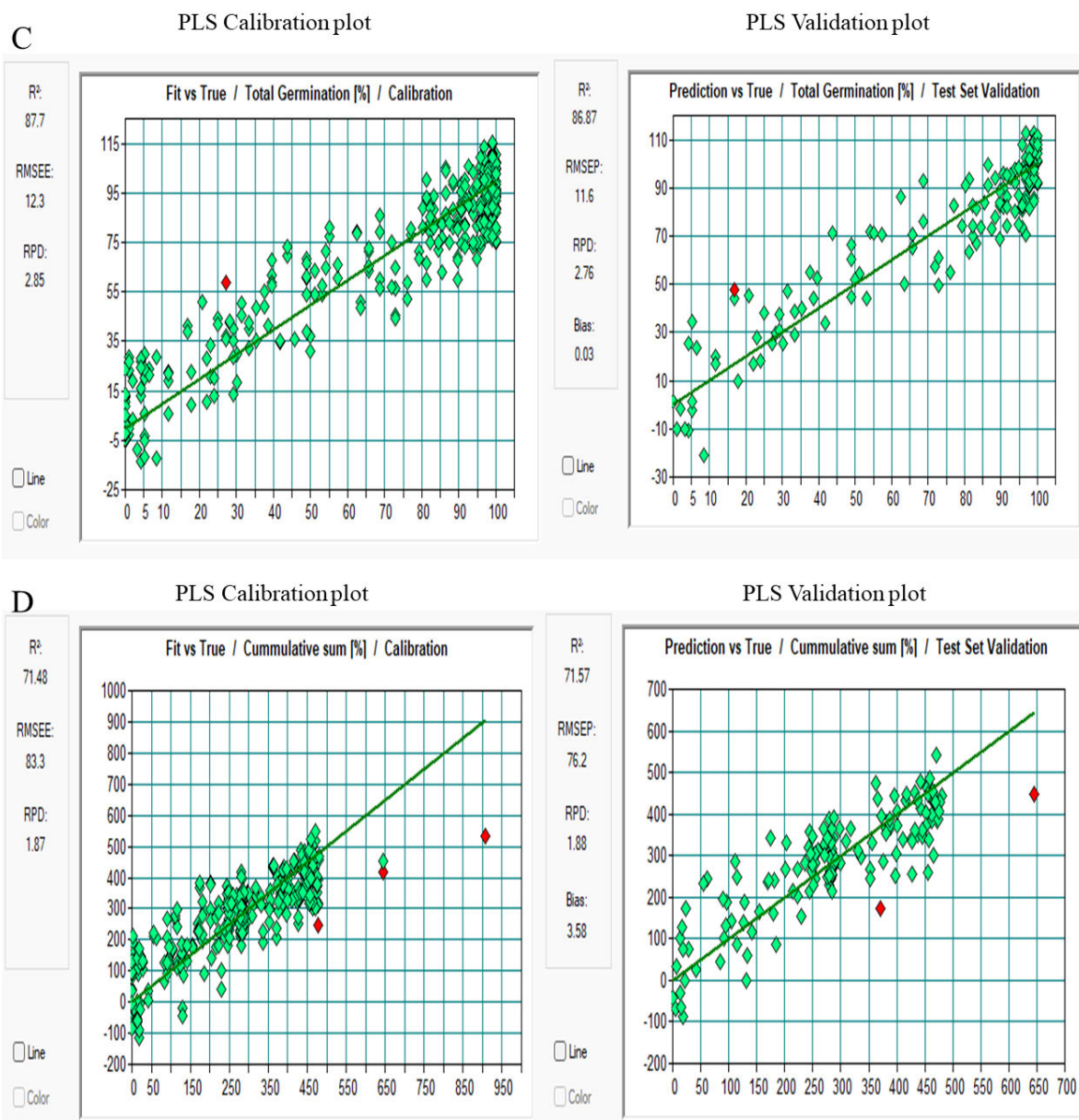


Figure 2.1: Representative plots comparing the NIR seed germination values between calibration and validation datasets, analysed using the global models for day 3, 7 and 10 intervals. The figure represents the germination plots for the day 3 global germination model and indicates the models performance statistics, viz. RMSEP, R^2 and RPD (ratio of performance difference). Data points in red were cautionary flagged as outliers by the OPUS software.

The performance of *Eucalyptus* spp. at day 3 germination models is described in Table 2.2. Models for all varieties used spectra within 4242.9 – 10290.9 cm^{-1} NIR region and required a first derivative, SNV, or a combination of both spectral pre-processing

algorithms to develop optimised models. The accuracy of the models had an error ranging from 5.8 – 18.8% and showed a good correlation ($R^2 > 80\%$) with the reference germination data. The most accurate (RMSEP: 5.8%) and well-correlated ($R^2 = 81.3\%$) model was achieved for *E. macarthurri* despite having a low sample number ($n = 13$). To increase the variation, and subsequently complexity, captured by the model, all species were incorporated into the development of a day 3 global model showing promising seed germination prediction accuracies (RMSEP: 18.8%) and correlation ($R^2 = 71\%$) despite being less accurate than the individual species models.

Table 2.2: NIR model performance of seed germinated in *Eucalyptus* species after 3 days of incubation.

Species	Sample number	Wavelengths of interest (cm^{-1})	Spectral pre-processing	RMSEP (%)	R^2
<i>E. nitens</i>	14 (42)	4242.9 – 10290.9	SNV	8.64	91.53
<i>E. benthamii</i>	130 (390)	4420.3 – 8894.5	First derivative	73.50	81.46
<i>E. dunnii</i>	13 (39)	4242.8 - 7506	First derivative + SNV	7.40	87.59
<i>E. macarthurri</i>	13 (39)	4242.8 - 5454	First derivative	5.80	80.13
Global model	175 (525)	6102 - 4242.8	First derivative	18.80	71.37

The values in parenthesis next to the sample number represent the spectra captured and used for model development. Wavelengths of interest were determined using PCA prior model development to select spectral regions that were associated with seed germination. The spectral pre-processing options included vector normalization (SNV) or the first derivative of the spectra. The global model represents the *Eucalyptus* seed germination model that was created using seed samples from all varieties scanned.

On day 7 (Table 2.3), germination models required a larger region of interest for model development ($3965.1 - 10283.1 \text{ cm}^{-1}$), which required SNV pre-processing. However, similar to the day 3 model had superior performance metrics for accuracy (RMSEP:

0.357 – 13.8%) and correlation (R^2 :0.5846 – 0.9218). The most accurate (RMSEP: 6.68%) and well-correlated ($R^2 = 0.92$) model was developed for *E. nitens* despite having a low sample number ($n = 14$). To increase the variation, and subsequently complexity, captured by the model, all species were incorporated into the development of day 7 global model showing promising seed germination prediction accuracies (RMSEP: 13.8%) and correlation ($R^2 = 81.6\%$) and was accurate similar to some individual species models.

Table 2.3: NIR model performance of seed germinated in *Eucalyptus* species after 7 days of incubation.

Species	Sample number	Wavelength of interest (cm ⁻¹)	Spectral pre-processing	RMSEP (%)	R ²
<i>E. nitens</i>	14 (42)	4003 – 9588.9	First derivative	6.68	92.18
<i>E. benthamii</i>	130 (390)	4242.9 – 9404.4	SNV	12.5	87.23
<i>E. dunnii</i>	13 (39)	3965.1 – 10283.1	SNV	2.83	58.46
<i>E. macarthurri</i>	13 (39)	4682.6 - 8208	SNV	0.357	77.32
Global model	175 (525)	7506 - 4242.8	First derivative	13.8	81.64

The values in parenthesis next to the sample number represent the total number of spectra captured and used for model development. Wavelengths of interest were determined using PCA prior model development to select spectral regions that were associated with seed germination. The spectral pre-processing options included vector normalization (SNV) or the first derivative of the spectra. The global model represents the *Eucalyptus* seed germination model that was created using seed samples from all varieties scanned.

The performance of *Eucalyptus* spp. at day 10 germination models, also referred to as the total germination models, is described in Table 2.4. All species models used spectra within 4242.9 – 9588.9 cm⁻¹ NIR region and required a first derivative, SNV of a combination of both for optimised models to be developed. The accuracy of the

models had an error ranging from 0.609 – 11.4% and showed a good correlation ($R^2 > 75\%$) with the reference germination data. The most accurate (RMSEP: 0.609%) and well correlated ($R^2 = 83\%$) model was achieved for *E. dunnii* despite having a low sample number ($n = 13$). The day 10 global model had seed germination accuracies, measured using the RMSEP, of 11.6% and an R^2 -value of 86%. This indicated that the global model had good predictive potential.

Table 2.4: NIR model performance of seed germinated in *Eucalyptus* species after 10 days of incubation.

Species	Sample number	Wavelength of interest (cm^{-1})	Spectral pre-processing	RMSEP (%)	R^2
<i>E. nitens</i>	14 (42)	4242.8 – 9403.7	SNV	4.96	95.42
<i>E. benthamii</i>	130 (390)	4003.7 – 9588.9	First derivative + SNV	11.4	90.18
<i>E. dunnii</i>	13 (39)	4597.7 - 7506	First derivative + SNV	0.609	83.35
<i>E. macarthurii</i>	13 (39)	4556.3 - 7506	First derivative + SNV	2.01	78.77
Global model	175 (525)	7506 - 4242.8	First derivative	11.6	86.87

The values in parenthesis next to the sample number represents the total number of spectra captured and used for model development. Wavelengths of interest were determined using PCA prior model development to select spectral regions that were associated with seed germination. The spectral pre-processing options included vector normalization (SNV) or the first derivative of the spectra. The global model represents the *Eucalyptus* seed germination model that was created using seed samples from all varieties scanned.

The day 10 model was further evaluated against an independent samples dataset (Table 2.5). The results of the predicted seed germination were close to that of the reference germination percentages. The RMSEP of the day 10 global seed germination model was found to be influenced by the variety of *Eucalyptus* scanned

(Figure 2.2). *Eucalyptus dunni* and *E. macarthurii* had the lowest associated RMSEP ($\leq 2.37\%$), whereas *E. benthamii* and *E. nitens* had the highest RMSEP of 22.75% and 32.95% respectively. This indicates that the model has better performance with *Eucalyptus dunni* and *E. macarthurii* seeds compared to the other varieties tested.

Table 2.5: Sample report for selected values produced by NIRS prediction of total germination percentage of *E. dunnii*, *E. nitens*, *E. macarthurii* and *E. benthamii* compared to the true values.

<i>Eucalyptus</i> variety	Reference germination (%)	Predicted germination (%)
<i>E. dunnii</i>	99.00	98.91
	99.00	99.06
	97.90	97.38
	100.00	97.20
	99.00	99.14
<i>E. nitens</i>	97.92	98.16
	97.92	100.30
	93.80	34.22
	95.80	39.69
	68.80	22.35
<i>E. macarthurii</i>	100.00	99.82
	99.00	101.92
	100.00	100.57
	86.50	81.56
	93.80	97.03
<i>E. benthamii</i>	96.90	32.51
	31.30	30.56
	43.80	56.78
	37.50	26.88
	79.20	54.21

Samples (n = 5) of each *Eucalyptus* variety used for the independent NIR germination assessment were randomly selected and analysed using the day 10 global seed germination model.

The accuracy of the day 10 global seed germination model was found to be influenced by the variety of *Eucalyptus* scanned (Figure 2.2).

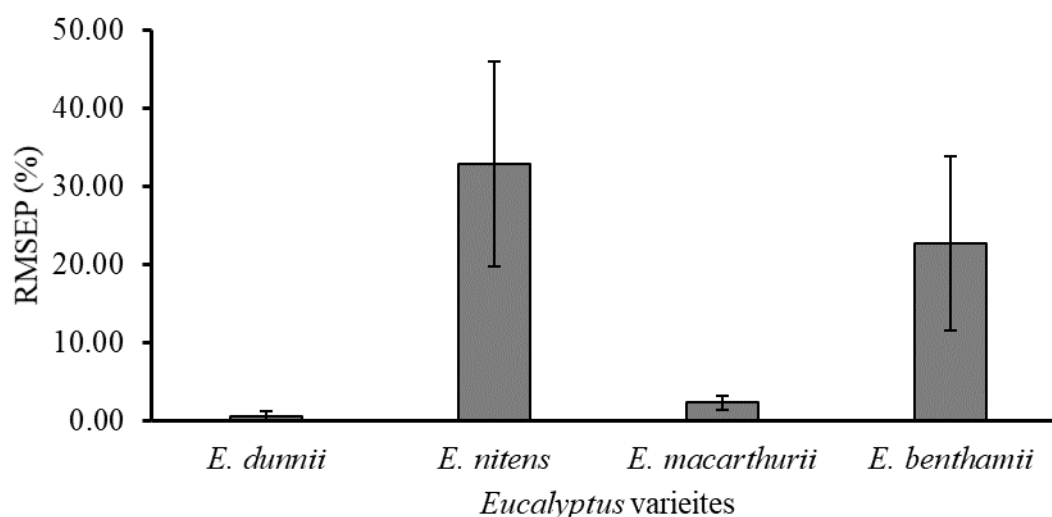


Figure 2.2: Level of accuracy of the day 10 global seed germination model assessed against independent samples of different *Eucalyptus* varieties.

2.3.2 NIR model performance for predicting seed germination vigour for *Eucalyptus* varieties

The level of performance of the *Eucalyptus* seeds to germinate, i.e., seed vigour, was modelled for all varieties tested, and a global seed vigour model was also developed. The performance of the seed vigour models is described in Table 2.6. All species models used spectra within 4242.9 – 9588.9 cm^{-1} NIR region and required a first derivative and SNV to develop optimised models. The accuracy of the models had an error ranging from 5.68 – 126% and showed good correlation ($R^2 > 0.71$) with the reference seed vigour data. The most accurate (RMSEP: 5.8%) and well-correlated ($R^2 = 80\%$) model was developed for *E. macarthurii*. However, high levels of prediction error were observed for *E. nitens* (RMSEP: 126%) and *E. benthamii* (RMSEP: 73.5%). The global model showed promising seed vigour prediction accuracies (RMSEP: 18.8%) and correlation ($R^2 = 71\%$) with the reference information. Its RMSEP values were slightly higher than *E. macarthurii* and *E. dunnii* but were largely lower than those achieved for *E. nitens* and *E. benthamii* models.

Table 2.6: NIR model performance of seed germinated in *Eucalyptus* species cumulative sum (vigour) after incubation.

Plant	Sample number	Wavelength of interest (cm ⁻¹)	Spectral pre-processing	RMSEP (%)	R ²
<i>E. nitens</i>	14 (42)	4242.9 – 9403.7	SNV	126	83.57
<i>E. benthemii</i>	130 (390)	4009.7-9588.9	First derivative	73.5	81.46
<i>E. dunnii</i>	13 (39)	6094.3 - 7506	SNV	5.68	78.91
<i>E. macarthurri</i>	13 (39)	5446.3 - 7506	First derivative	5.8	80.13
Global model	175 (525)	6102 - 4242.8	First derivative	18.8	71.37

The values in parenthesis next to the sample number represents the total number of spectra captured and used for model development. Wavelengths of interest were determined using PCA prior model development to select spectral region that are associated with seed vigour. The spectral pre-processing options included vector normalization (SNV) of first derivative of the spectra. The global model represents the *Eucalyptus* seed germination model that was created using seed samples from all species scanned.

2.4 Discussion

Evaluating seed germination characteristics, including germination rate, vigour, and viability, is essential for optimizing hedge propagation and ensuring a thriving plantation. Therefore, this study investigated the potential of NIRS as a rapid, non-destructive tool for predicting *Eucalyptus* seed germination and vigour. NIRS offers an efficient alternative to traditional germination tests by eliminating the need for destructive sampling and long incubation times by employing appropriate calibration, validation models, and machine learning algorithms. The findings of this study highlight the potential of NIRS in *Eucalyptus* seed quality management.

The 4000 – 9600 cm⁻¹ NIR region was used to develop the best-performing models for *Eucalyptus* seed vigour and germination, regardless of germination interval. This NIR region has been associated with bonds usually seen by NIR spectra (H-C, H-O, H-S

and H-N), which constitute the primary structure of organic materials (Cozzollino *et al.*, 2004). The biological and the organic materials are often analysed using NIR due to its chemometrics and since organic material has chemical composition of molecular with H-bonds (Cogdill *et al.*, 2004). The wavelength used by the seed germination models can be associated with the important elements required for germination. Kandpal *et al.* (2016) reported that the wavebands at 5236 cm^{-1} and 5181 cm^{-1} correlate with the carbohydrate content, and the peak around 5208 cm^{-1} for C=O is due to the peptide in the protein. Another study by Lohumi *et al.* (2015) identified that the wavelength at 5208 cm^{-1} is also related to CONH, referred to as an amide. At 6369 cm^{-1} , the N-H first overtone stretching vibration that represents protein content, at 4440.5 cm^{-1} range of O-H stretching, which represents starch content and another absorption band also presents near 4264 cm^{-1} , which was related to CH_2 , which is a functional group being assigned to lipids (Venkatesan *et al.*, 2020) indicating the feasibility of NIR to identify the specific chemical compound in seeds.

NIR models predicting seed germination were reasonably successful for most *Eucalyptus* species and the accuracy of models was found to improve with the duration of the incubation time (Table 2.2 – 2.6, respectively). The improvement in the model's accuracy at later germination intervals was not unexpected as the seeds were given more time to germinate, making them easier to model (Onofri *et al.*, 2014). Therefore, by identifying spectra associated with poor germination, the NIR model can guide the development of targeted interventions, such as optimizing seed storage conditions or selecting treatments that enhance vigour. Additionally, the accuracy of germination models was found to be influenced by the *Eucalyptus* varieties modelled. Afroze *et al.* (2021) determined that the germination of *Eucalyptus* seeds is variable between species; presumably, the seed germination potential of certain species or varieties is more amenable than others. This would explain the large level of error observed for the seed germination global models. On this basis, future investigations would evaluate models that distinguish seeds by species and then apply a species-specific germination regression model to predict their germination at certain milestones.

The germination regression model indicated a good correlation between spectra and reference germination data. R^2 values above 70% across all models and varieties indicates that the spectra captured essential information that influence seed germination rate. The findings on day 3 also indicated that the species-specific models

for *E. macarthurii*, *E. nitens* had better predictive capability. It can be inferred that the unique spectral characteristics of their seeds were reliably accounted for by the model. The capacity of global model prediction with an R^2 of 71% and an RMSEP of 18.8% suggests that it has potential, however, it needs more development. Global models are difficult to develop as they must account for confounding varietal-specific NIR signals or refine their wavelength selection to specific regions for seed germination. The use of Vis/NIR and SWIR hyperspectral imaging systems combined with the partial least squares regression algorithm was able to predict the germination percentage (GP), germination energy (GE), simple vigour index (SVI), soluble protein content and sugar content of aged wheat seeds (Zhang *et al.*, 2020). However, they would offer a more robust seed germination model once they are developed. Therefore, the global models warrant further development by increasing the sample number and varieties used to calibrate and validate the models.

The findings indicated that NIR can be used as a daily analysis tool for predicting seed germination rate and vigour. The result indicated that the species-specific model could work best as the level of variation and can be limited within the species. The global model could be the best solution as it represents the large variation of spectra based on multiple species used to develop the model. The prediction model used similar spectra pre-processing SNV, first derivative or combination. The region of interest is also similar to either species-specific or global model developed. The NIR models showed that it is independent to sample number, therefore a small number of samples were used for model development. Day 3 model predicted early germination of healthier seed lots referring to seed vigour and day 7 model showed that low vigour seed can later germinate, with day 10 model predicted the overall seed germination. However, the global model presented an option of a more robust prediction accuracy, due to larger variation and increased sample size, with a high error that still requires perfecting to be more accurate with minimal errors.

Seed vigour is regarded as a sensitive indicator for seed lot quality in comparison with standard germination, as it relates to field conditions and seed establishment (Egli and TeKrony, 1995). The vigour level is also reduced during the seed storage prior to the germination rate. (Venkatesan *et al.*, 2020). It is common for seed lots with the same standard germination to have different vigour levels resulting from variable seed production systems or storage conditions. The seed vigour model also showed a low

accuracy level compared to the single species model. This can be influenced by the number of different species used, as the variation increases with the sample number, the model can be highly utilised and increased accuracy can be archived on determining what percentage can germinate faster with the seed lot or species specifically.

2.5 Conclusion

This study demonstrated the feasibility of using NIRS to screen seeds prior to planting to estimate germination characteristics. While this modern technology shows promise, its commercial application requires the development of more comprehensive predictive models. These models should incorporate larger sample sizes with evenly distributed variation to improve accuracy and robustness. Additionally, validation in nursery conditions is necessary to assess practical implementation. Nevertheless, the findings indicate that NIRS screening has the potential to serve as a valuable tool for nursery operators, enabling them to optimize sowing rates and achieve more reliable seedling establishment for field out-planting.

2.6 References

- Abbas, O., Pissard, A. and Baeten, V. (2020). Near-infrared, mid-infrared, and Raman spectroscopy. In *Chemical Analysis of Food* (pp. 77-134). Academic Press.
- Abiyu, A., Teketay, D., Glatzel, G. and Gratzner, G. (2016). Seed production, seed dispersal and seedling establishment of two Afromontane tree species in and around a church forest: implications for forest restoration. *Forest Ecosystems*, 3(1), pp.1-10.
- Agelet, L.E. and Hurburgh Jr, C.R. (2014). Limitations and current applications of Near Infrared Spectroscopy for single seed analysis. *Talanta*, 121, pp.288-299.
- Bareke, T.J.A.P.A.R. (2018). Biology of seed development and germination physiology. *Advances in Plants and Agriculture Research*, 8(4), p.336.
- Davies, R., Di Sacco, A. and Newton, R. (2015). Germination testing: procedures and evaluation. *Technical Information Sheet_13a. Royal Botanic Gardens, Kew*.
- Dyrby, M., Engelsen, S.B., Nørgaard, L., Bruhn, M. and Lundsberg-Nielsen, L. (2002). Chemometric quantitation of the active substance (containing $C\equiv N$) in a

- pharmaceutical tablet using near-infrared (NIR) transmittance and NIR FT-Raman spectra. *Applied Spectroscopy*, 56(5), pp.579-585.
- Fan, Y., Ma, S. and Wu, T. (2020). Individual wheat kernels vigour assessment based on NIR spectroscopy coupled with machine learning methodologies. *Infrared Physics and Technology*, 105, p.103213.
- Gebeyehu, B. (2020). Review on: Effect of seed storage period and storage environment on seed quality. *International Journal of Applied Agricultural Sciences*, 6(6), pp.185-190.
- Gregorio, N.O., Herbohn, J.L. and Harrison, S.R. (2010). Guide to quality seedling production in smallholder nurseries. *Visayas State University funded by the Australian Centre for International Agricultural Research, Q-Seedling Project, College of Forestry and Natural Resources, Visca, Baybay City, Leyte, Philippines*.
- Grossnickle, S.C. and MacDonald, J.E. (2018). Seedling quality: history, application, and plant attributes. *Forests*, 9(5), pp.283.
- Hein, P. R. G., & Parrish, C. (2020). Near infrared spectroscopy in forestry: a review of applications and opportunities. *Journal of Near Infrared Spectroscopy*, 28(4), 231–246.
- Lindsey III, B.E., Rivero, L., Calhoun, C.S., Grotewold, E. and Brkljacic, J. (2017). Standardized method for high-throughput sterilization of Arabidopsis seeds. *Journal of Visualized Experiments*, (128), p.e56587.
- Loha, A., Tigabu, M., Teketay, D., Lundkvist, K. and Fries, A. (2006). Provenance variation in seed morphometric traits, germination, and seedling growth of *Cordia Africana* Lam. *New Forests*, 32(1), pp.71-86.
- Marcos Filho, J. (2015). Seed vigour testing: an overview of the past, present and future perspective. *Scientia Agricola*, 72(4), pp.363-374.
- Milivojević, M., Ripka, Z. and Petrović, T. (2018). ISTA rules changes in seed germination testing at the beginning of the 21st century. *Journal on Processing and Energy in Agriculture*, 22(1), pp.40-45.
- Miransari, M. and Smith, D.L. (2014). Plant hormones and seed germination. *Environmental and Experimental Botany*, 99, pp.110-121.
- Rahman, A., and Cho, B.K. (2016). Assessment of seed quality using non-destructive measurement techniques: a review. *Seed Science Research*, 26(4), pp.285-305.

- Roopa, S.V. and Wadje, S.S. (2012). In-vivo testing of plant extracts against seed borne pathogens. *International Research Journal of Biological Sciences*, 1, pp.1-4.
- Santos, M.M., Ataíde, G.D.M. and Souza, G.A.D. (2017). Germination of seeds of *Melanoxylon brauna* Schott. under heat stress: production of reactive oxygen species and antioxidant activity. *Forests*, 8(11), p.405.
- Shaban, M. (2013). Study on some aspects of seed viability and vigour. *International Journal of Advanced Biological and Biomedical Research*, 1(12), pp.1692-1697.
- Sivakumar, V., Parthiban, K.T., Singh, B.G., Gnanambal, V.S., Anandalakshmi, R. and Geetha, S. (2002). Variability in drupe characters and their relationship on seed germination in Teak (*Tectona grandis* L. f.). *Silvae Genetica*, 51(5-6), pp.232-236.
- Tigabu, M. and Odén, P.C. (2002). Multivariate classification of sound and insect-infested seeds of a tropical multipurpose tree, *Cordia Africana*, with near infrared reflectance spectroscopy. *Journal of Near Infrared Spectroscopy*, 10(1), pp.45-51.
- Tigabu, M. and Odén, P.C. (2003). Discrimination of viable and empty seeds of *Pinus patula* Schiede & Deppe with near-infrared spectroscopy. *New Forests*, 25(3), pp.163-176.
- Toumey, J.W. (1916). *Seeding and planting: a manual for the guidance of forestry students, foresters, nurserymen, forest owners, and farmers*. John Wiley & Sons, Incorporated.
- Van Wyk, A., Brink, A., Van Huyssteen, G. and Jansen, J. (2012). Tree farming guidelines for private growers. *Cascades: Sappi Forests*.
- Walsh, D.G.F., Waldren, S. and Martin, J.R. (2003). Monitoring seed viability of fifteen species after storage in the Irish threatened plant gene bank. In *Biology and Environment: Proceedings of the Royal Irish Academy* (Vol. 103, No. 2, pp. 59-67). Royal Irish Academy.
- Xia, Y., Xu, Y., Li, J., Zhang, C. and Fan, S. (2019). Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artificial Intelligence in Agriculture*, 1, pp.35-47.

Zhang, Tingting, Shuxiang Fan, Yingying Xiang, Shujie Zhang, Jianhua Wang, and Qun Sun. (2020). Non-destructive analysis of germination percentage, germination energy and simple vigour index on wheat seeds during storage by Vis/NIR and SWIR hyperspectral imaging. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 239: 118488.

Chapter 3

Evaluation of near-infrared reflectance spectroscopy for low-cost foliar nutrient analysis of *Eucalyptus* seedlings

Abstract

Eucalyptus clones are intensively propagated in nurseries and planted to establish commercial plantations. The nutritional status of *Eucalyptus* cuttings impacts their ability to establish viable hedges and subsequently impacts the success of plantation establishment. However, traditional assessments of plant nutrition are costly and destructive. Therefore, the aim of the present study was to evaluate if NIRS can be used as a rapid tool to analyse and help monitor *Eucalyptus* nutrient levels at the seedling production stage and at establishment of plantations. Foliar material was sampled from the field and pot trials that are part of a long-term nutrient depletion trial, samples at 1 year old coppice trees and pot trial samples, with a total of 382 samples and 1146 spectra excluding independent test data set. The spectra were captured using a Bruker MPA NIRS. Foliar nutrient assessment was carried out with an Agilent MP-AES 4200 and Leco CNS and used as the reference data to create NIR nutrient calibration models. The global models developed for key nutrients models were good and the accuracy rate was $R^2 > 89\%$ for *E. dunnii*, G×N and G×U. Global models for N had a highest value of $R^2 > 95\%$ and the lowest value was $R^2 > 56\%$ and the pre-processing was mostly between the first derivative, SNV, or a combination of both in all the elements tested. The study showed that NIRS could analyse some key nutrients in *Eucalyptus* varieties.

3.1 Introduction

Effective nutrient management in nurseries is essential for producing and propagating healthy seedlings or tree clones. These clones are then used to establish forests for commercial tree production (Rahman, 2006). The success of plantation establishment depends on the health of the seedlings or clones planted (Mohan and Sharma, 2005). Hence, nutritional assessment of clones or cuttings at the nursery and throughout plantation reestablishment is important (Nobre and Rodriguez, 2014).

Therefore, a need for rapid nutrient and cost-effective nutrient assessment methods would be viable for improving the monitoring of nutrient levels.

Key elements critical to the health and vitality of *Eucalyptus* species include nitrogen (N), phosphorus (P), and potassium (K), as well as calcium (Ca), copper (Cu), iron (Fe), sulphur (S), carbon (C), zinc (Zn), boron (B), magnesium (Mg), manganese (Mn), and sodium (Na) (Agustan *et al.*, 2021). These nutrients are monitored to assess the health of *Eucalyptus* seedlings. Nitrogen supports amino acid and protein production, as well as growth regulation. Phosphorus facilitates energy transfer (via ATP), nucleic acid synthesis, and high-energy bond formation, while potassium is essential for nucleic acids, stomatal function, enzyme activity, protein synthesis, photosynthesis, and cell growth. These elements are also critical for auxin production, root development, and immune responses, especially in cutting propagation (Humphreys *et al.*, 2007). Commonly, the analysis of these elements involves sample preparation for techniques like MPAES or ICP-OES, along with CNS analysis for measuring total carbon, nitrogen, and sulphur percentages.

Traditional methods for measuring the nutritional status of a plant are laborious and destructive, require expensive equipment, or have long turnaround times for results to be reported. Processes contributing to this include sample drying, milling, complete combustion, digestion and analysis on an ICP-OES system. Dried and milled samples are also used on a CNS analyser for total percentage of carbon, nitrogen, and sulphur. These common traditional methods have a long turnaround time. This suggests that it is not easy to determine the nutrients for all plants in the nursery. Furthermore, it is impractical to measure all nutrients for each plant to determine the likelihood of survival in the field (Grossnickle and MacDonald, (2018).

Near-infrared reflectance spectroscopy (NIRS) is a non-destructive and rapid form of analysis that can reduce the turnaround time, meaning the information can be delivered to forester or nursery personnel to attend to the seedlings needs. It is a spectroscopic technique that uses the vibrational signatures of covalent bonds in the NIR region with machine learning algorithm to predict properties of a sample (Reich, 2016). It can be useful because infrared light interacts with key nutrient elements, or the organic molecules that are associated with them. NIRS offers potential benefits for improving nursery operations by enabling frequent nutrient assessment.

This study was initiated to evaluate if NIRS can be used to analyses and help monitor *Eucalyptus* nutrition level at the seedling production stage and at early reestablishment of plantation. The objective was to develop an NIR screening method to estimate nutrition levels in nursery seedlings. Currently such a tool using NIRS nature is lacking for the SA forestry industry.

3.2 Materials and methods

3.2.1 Sample collection and preparation

The samples used were part of a long-term nutrient depletion and fertiliser trial. Samples of 1 year old *E. dunnii* were collected from the field trials that have coppicing shrubs that were harvested every year. There were 4 treatments, fertilised plots (shrubs removed and fertiliser was applied), retained plots (shrubs retained in their plot), removed plots (all shrubs were removed from the plots) and 2RF plots (retained and fertilised). Other samples of 1 year old hybrids G×N and G×U were collected from a pot trial at ICFR nursery also a long-term nutrient depletion and fertiliser trial. Soil collected from removed plot (arenite soil, sandy stone, granite, gabbro, and Clan-shale) were used to establish the pot trial. Species used included *E. dunnii*, hybrids G×U, and G×N. A minus-one design was used, i.e., removal of single nutrient per pot to no fertiliser, as well as fully fertilised nutrient pot (all micro-, i.e., B, Mn, Na, Cu, Fe, Zn, and macronutrients, i.e., total N, P, K, Ca, Mg and total S). Foliar samples were from small, medium, and large trees, from each plot, compartment, field, and nursery, A total of 382 samples from 1 year old tree or seedling, from October to November 2019.

3.2.2 Assessment of foliar nutrient analysis

The plant samples were prepared as per the ICFR lab manual (Ramesar *et al.*, 2024). Foliar material was oven dried at 60°C for a week. Samples were then milled at 6000 rpm using Retch ZM 200 for particle size less than 0.05 mm. Plant nutrient analyses involved drying 1 g of milled sample at 105°C for 24 hours. Thereafter, the sample was ashed in a furnace at 500°C for 6 hours. The ash was then digested by boiling it in 5 ml of 1:1 hydrochloric acid HCl on a hot plate. The digest was allowed to cool to room temperature before being filtered through a single disk of Whatman 42 filter paper and made to 50 ml using deionised water. The filtrate was analysed on an Agilent MPAES

4200. This provided concentration data for Ca, Na, Mg, K, Fe, Ni, Cu, Zn, Mn, P and B. The foliar sample percentage composed of C, N and S measured using a Leco Trumac CNS analyser at ICFR (as per ICFR FRAL manual), whereby 0.5 g of sample was processed.

3.2.3 NIR spectra acquisition and model development

The spectral acquisition for all foliar milled samples was acquired by placing samples into quartz vials (diameter = 20 mm), filling to a third and were allowed to acclimatise overnight to the temperature of the NIRS lab (19°C) and its ambient humidity. The samples were scanned using the integrating sphere on the Bruker MPA I FT-NIR spectrometer (model: MPA I, Bruker, USA) at the ICFR Forestry Research and Analytical Laboratory (FRAL). The OPUS Lab software (v.6.5) (Bruker, USA) was used to control the NIRS and subsequent spectral acquisition. The acquired spectral range was from 14963 cm⁻¹ to 3574 cm⁻¹ with a resolution of 8 cm⁻¹. All samples were scanned in triplicate, with hand inversion of the vials to increase the NIR sample variation captured. Model development was performed according to section 2.2.4 described in Chapter 2.

3.3 Results

The predicted versus true results comparing the NIR nutrient analyses with MPAES and Leco CNS analyser of the *Eucalyptus* hybrids, G×N and G×U varieties showed that these parameters could be predicted relatively well. Uncertainty or error in the model predictions appear to be correlated with the number of samples or spectra used to develop the models and varies between *Eucalyptus* varieties.

3.3.1 NIR model performance for predicting foliar nutrient levels in *Eucalyptus* varieties

Modelling of the foliar nutrient of the *Eucalyptus* varieties was accomplished for key elements for plant growth, i.e., N, P and K. The model calibration and validation plots correlated well with the reference data (Figure 3.1). N appeared to have the most precise grouping of data points on both calibration and validation graphs. Graphs of P and K were spread around the theoretical 1:1 true versus predicted line, indicating promise for other elements being modelled using NIR. Table 3.1 shows the NIR foliar

nutrient models for *E. dunnii* and the hybrids, G×N and G×U. All N models showed good correlation ($R^2 > 89\%$) and accuracy ($RMSEP \leq 0.046\%$) for all *Eucalyptus* varieties. This element was the most amenable to accurate modelling using NIR compared to P and K concerning their RMSEP values. High RMSEP values were observed for P and K despite having good R^2 -values ($R^2 > 82\%$). The cause or association of the high RMSEP values was common between all models for P and K. Therefore, alluding to the modelled elements associated with the high RMSEP values.

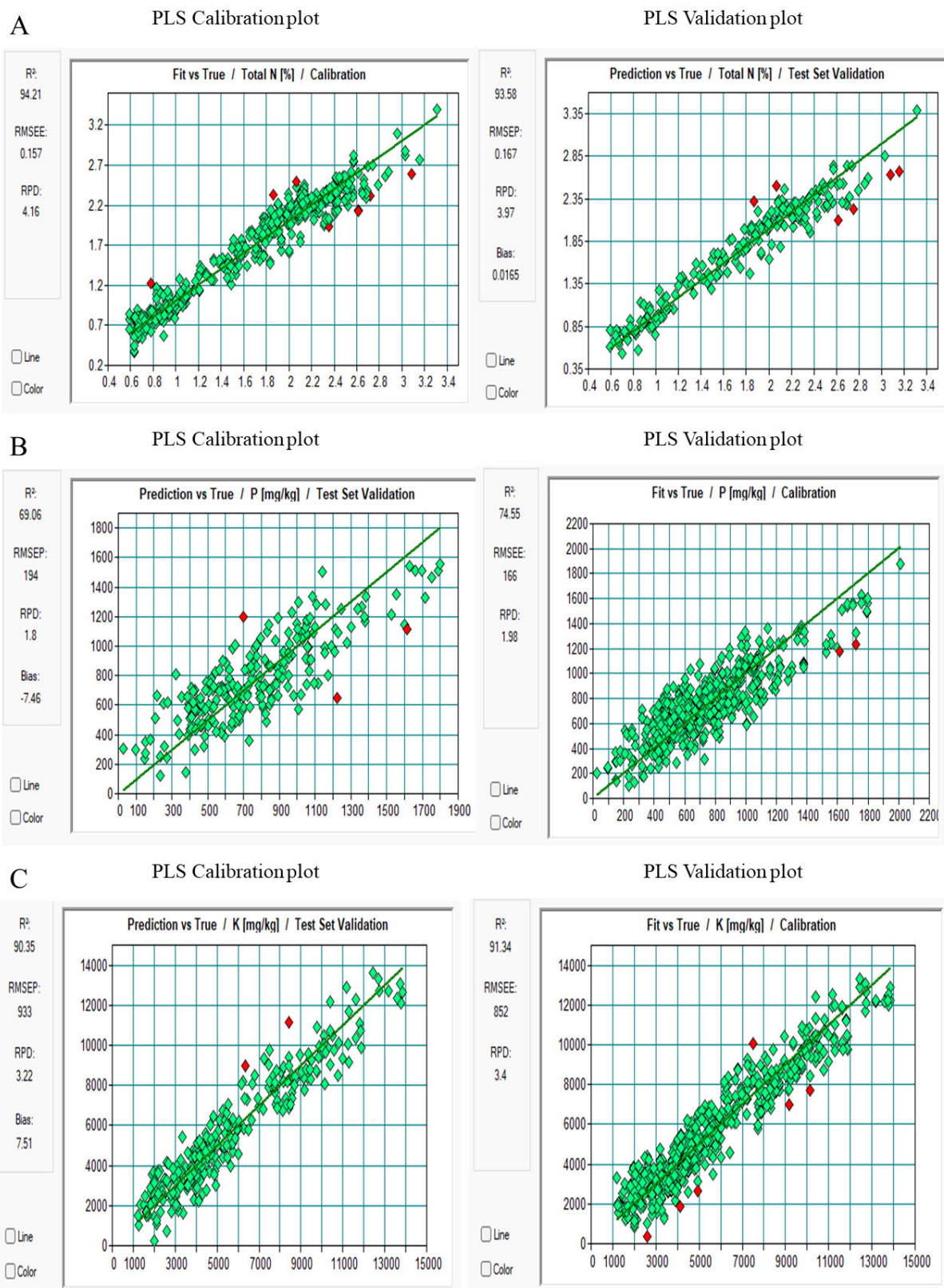


Figure 3.1: Representative plots of foliar nutrient data set, showing test validation data set A- Calibration and validation test for total (N) nitrogen, B - calibration and validation test for total phosphorus (P) and C - calibration and validation test for total potassium (K). The figure represents the true values vs predicted values and model indicating the

statistics for the PLS regression used, viz. correlation coefficient (R^2), root mean squared error (RMSE) of prediction (RMSEP) and ratio of performance difference (RPD). Data points in red were flagged as outliers by the OPUS software.

Table 3.1: NIR model for nutrient analysis using *Eucalyptus* varieties (*E. dunnii*, hybrid G×N, and G×U, comparing R^2 of global model and single species analysis.

Element	<i>E. dunnii</i>		G×N		G×U	
	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP
N (%)	89.74	0.046	92.74	0.045	99.20	0.017
P (mg/kg)	94.34	51.70	82.13	209	89.83	170
K (mg/kg)	95.69	3120	95.21	381	91.99	385

Nitrogen content was reported as the total amount of nitrogen per gram of sample, expressed as a percentage (%). Phosphorus (P) and Potassium (K) were reported as milligrams of element per kilogram of sample (mg/kg). The RMSEP values were, therefore, interpreted in the concentration unit of the respective element.

An independent test set analysis of the varietal nutrient models was performed to assess the models robustness. Figure 3.2 shows the correlation between true vs predicted nutrient values for NPK elements (Table A1 and A2 also enlighten the performance of the model, showing predicted values). A- Nitrogen element indicate high correlation for *E. dunnii* $R^2= 0.99$ and lowest $R^2= 0.51$ for G×U. B-Phosphorus only G×N had a high correlation of $R^2= 0.89$ and the lowest being for *E. dunnii* and G×U variety. C-Potassium G×N had a high correlation of $R^2= 0.87$ and the lowest being for *E. dunnii* and G×U variety. The correlation graphs show a good distribution of the data, but it was only limited to 5 samples that were used as independent test set.

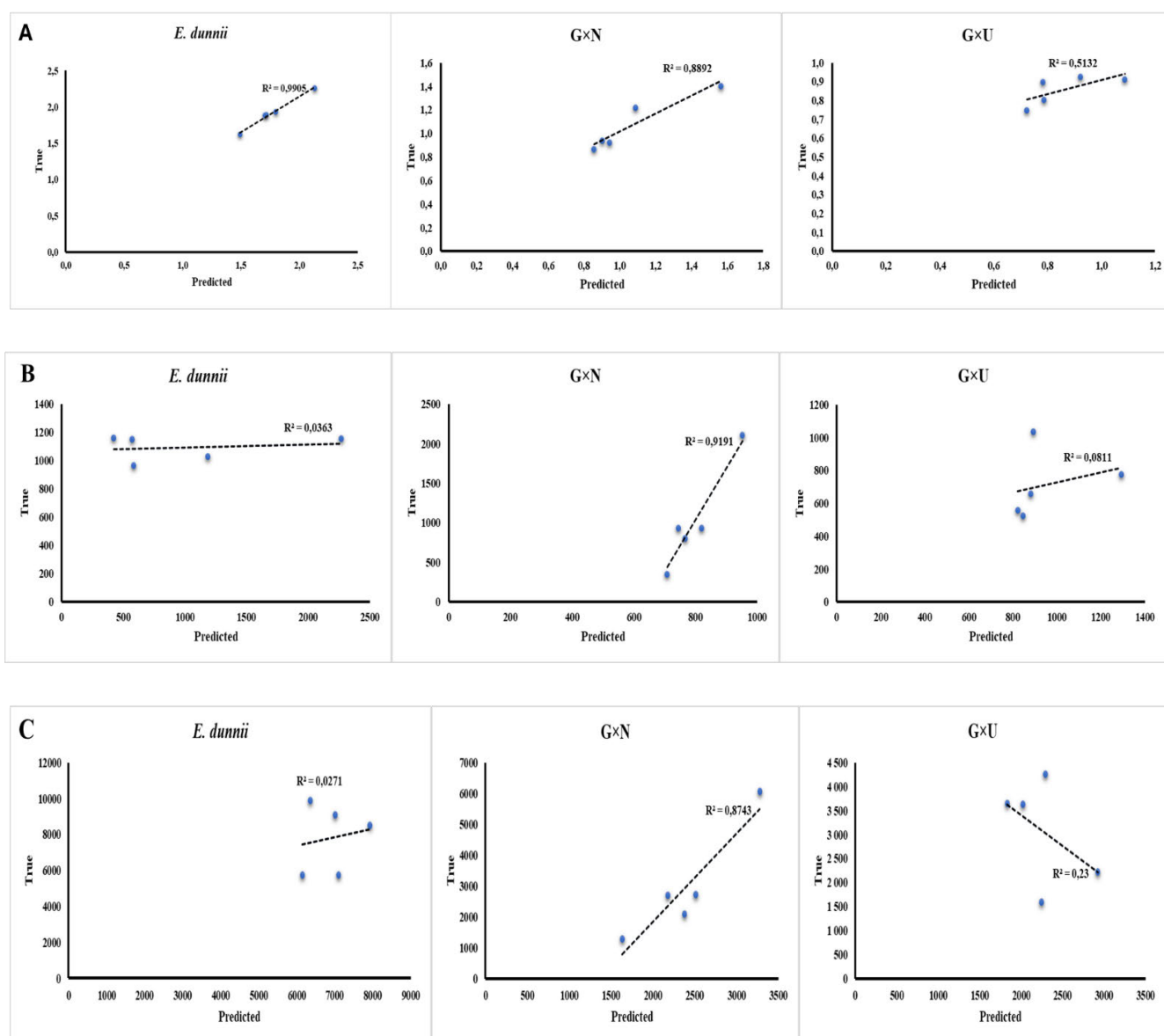


Figure 3.2: The correlation graphs set for true vs predicted values between nutrient models for (A) nitrogen, (B) phosphorus and (C) potassium on independent test for *E. dunnii*, G×N and G×U. Data was generated from an independent sample sets comprised of 5 samples per varietal model (Table 3.3).

3.3.2 NIR model performance of global *Eucalyptus* nutrient models

The global *Eucalyptus* nutrient models were developed to reduce the RMSEP of individual models and thereby consolidate the models. Furthermore, a consolidated model using spectra from each *Eucalyptus* variety could reduce the RMSEP for element prediction by increasing the sample size of model development.

The performance metrics of the *Eucalyptus* nutrient models are shown in Table 3.2. Models for all elements used absorbance values within 3942 – 10946.6 cm⁻¹ and required a first derivative, SNV, or a combination of both to develop optimised models. The RMSEP of the models for N, P and K varied as function of the element analysed. However, these models showed good correlation ($R^2 > 86.53\%$) with their reference chemistry data. In terms of accuracy, the order was N > P > K. The model RMSEP for N and P was lower than the varietal models; however, the K model was more accurate than *E. dunnii* and the G×U models. This alluded to the global model having improved performance from the individual varietal models.

Models for the macro elements, i.e., Na, Mg and Ca, were developed; however, they had high RMSEP values that ranged between 345 and 1360 mg/kg. These levels of error are too great for reliable quantification of these elements. The micro elements, i.e., Fe, Zn, and Cu and Mn, had relatively smaller RMSEP values, except for Mn. Their RMSEP ranged between 1.37 and 34.8 mg/kg. Manganese had the highest RMSEP (703 mg/kg) of all microelements. Boron had an acceptable RMSEP value of 18.60 mg/kg. The results show that not all elements model well in the NIR region, however, some, such as N, can be modelled effectively.

Table 3.2: NIR *Eucalyptus* global model performance metrics for foliar nutrients.

Elements	Wavelengths of interest (cm ⁻¹)	Spectral pre-processing	RMSEP	R ²
B (mg/kg)	8154 - 3965.2	SNV	18.60	65.58
Ca (mg/kg)	9403.7 - 4242.9	SNV	1360	91.77
Cu (mg/kg)	7506 - 4242.9	First Derivative	1.37	80.88
Fe (mg/kg)	9403.7 - 4242.9	SNV	34.8	83.61
K (mg/kg)	8377.7 - 4242.9	First Derivative + SNV	1310	90.35
Mg (mg/kg)	7860.9 - 4242.9	SNV	484	90.85
Mn (mg/kg)	9403.7 - 4242.9	First Derivative	703	87.42
Na (mg/kg)	8377.7 - 4242.9	SNV	345	87.73
P (mg/kg)	10946.6 - 3965.2	First Derivative	199	86.53
Total N (%)	9658.3 - 3942	First Derivative	0.163	95.39
Total S (%)	8894.6 - 4003.7	SNV	0.028	56.81
Zn (mg/kg)	8192.6 - 4003.7	SNV	9.62	71.72

The global model represents the elemental or nutritional levels modelled from foliar material acquired from 3 *Eucalyptus* varieties comprised of 382 samples which produced 1146 spectra. The values in parenthesis next to the sample number represents the total number of spectra captured and used for model development. Wavelengths of interest were determined using PCA prior model development to select spectral region that are associated with each element. The spectral pre-processing options included vector normalization (SNV), first derivative or a combination of both. Performance metrics include the root mean squared error of prediction (RMSEP) and the correlation coefficient (R^2) to assess the models accuracy and association with the reference data, respectively.

3.3.3 Comparison between single variety and global model

The data comparison of global model performance and of single species model (*E. dunnii*, hybrid G×N, and G×U) for the nutrient element had a good correlation (R^2 82.13 – 99.2%) is observed within all the models developed. The performance metrics of the foliar nutrient models are shown in Table 3.2. Models of all varieties used absorbance values within 3942 – 10946.6 cm^{-1} and required a first derivative, SNV, or a combination of both to develop optimised models. The RMSEP of the models ranged from 0.03 – 703% and the NPK model showed a good correlation ($R^2 > 86.53\%$) with the reference wet chemistry data. The most accurate (RMSEP: 0.163%) and well-correlated ($R^2 = 95.39\%$) model was developed for Total N. To increase the variation, and, subsequently complexity, captured by the model, all nutrient elements were incorporated into the development of the mixed element global nutrient model, which showed slightly higher levels of prediction error (RMSEP: 703%) and lower levels of correlation ($R^2 = 56.81\%$) compared to the models developed individually for each *Eucalyptus* species

Table 3.3: Independent test dataset for N, P and K using the global model for samples of *E. dunnii*, hybrid G×N, and G×U.

Element	<i>E. dunnii</i>		G×N		G×U	
	True	Predicted	True	Predicted	True	Predicted
N (%)	1.707	1.88	1.0874	1.2168	0.78749	0.80191
	2.1296	2.26	0.94126	0.92187	0.92299	0.92404
	1.4911	1.62	0.85543	0.86442	0.78324	0.89637
	1.7193	1.89	0.90162	0.9383	0.72343	0.74604
	1.798	1.93	1.5632	1.4045	1.0871	0.91104
P (mg/kg)	1187.3	1031.33	819.62	929.79	846.57	525.23
	2267	1155.45	743.93	932.75	822.19	558.95
	586.19	969.20	765.14	800.09	1292.27	775.55
	571.59	1153.59	708.06	352.69	881.54	657.36
	423.57	1164.16	952.14	2113.6	892.6	1035.5
K (mg/kg)	7013	9105.30	1635.73	1284.3	2020.42	3632.4
	6362.9	9909.11	2512.44	2730.5	2247.69	1601.7
	6146.9	5760.77	2376.68	2103.2	4924.3	2217.9
	7101.9	5770.49	2182.1	2706.6	1830.32	3644.5
	7919	8534.90	3278.68	6065.4	2294.13	4262.8

3.4 Discussion

Assessment of the nutritional status of *Eucalyptus* cutting productions is crucial for successful propagation at nurseries and establishment of plantations. However, the major limitation to this is the cost of analysis, considering the scale at which hedges are propagated, and cuttings are planted in the field. NIRS is among the few viable tools to address the limitations of this type of assessment, namely as it can analyse high sample number, eliminate the destructive nature of samples and reduce analysis cost. The current study evaluates NIRS as the potential alternative in facilitating the rapid prediction of the nutritional status of *Eucalyptus* varieties.

Models for key nutrients (N, P and K) were modelled and showed a good correlation of reference information for *E. dunnii*, G×N and G×U ($R^2 > 89\%$). However, with

different degree of accuracy, this varies with the varieties of *Eucalyptus* and the element modelled. These *Eucalyptus* varieties have distinct biomolecular arrangements, making it more or less likely to measure the interaction or the elements within the foliar material (Barbosa *et al.*, 2016). This explains why the best-performing model for nitrogen (N) G×U variety, whereas *E. dunnii* had high error levels. The element's molecule state influences the model performance as NIR easily detects and measures it. The overall error level for P and K is very high, which can be associated with the molecular state of the elements (Table 3.2). Singh *et al.* (2022) demonstrated that the visible and near-infrared regions have high correlations with chemical compounds analysed in their study using a stepwise regression. However, insufficient sample sizes did not allow prediction for all elements, and they suggest implementing techniques for reducing instrument error (Singh *et al.*, 2022). Herbert and Schönau (1990) indicated that minimum value thresholds where nutrient deficiency is possible for N, P and K were 29, 1.4 and 7.5 g/kg respectively, while Neto *et al.* (2020) and Goncalves *et al.* (2008) suggested different minimum nutrient deficiency values requirements for specific cutting growth on cuttings production or in field plantation. Additional elements N, P and K were modelled at varying level of success and still require extensive analysis to ensure high accuracy. Generally, the highest level of error was observed for detecting potassium, followed by phosphorus and nitrogen. This is significant because potassium is an element that must be measured down to nearest milligrams based on its distinct availability to plant cells (Millner and Kemp, 2012). Considering the low sample numbers for developing single species models, the global models development was considered because they include more variation, resulting in better performance and avoiding limitations to one variety of *Eucalyptus* species. Global models are not limited to a single *Eucalyptus* variety. They include *E. dunnii*, G×N and G×U in one model.

Table 3.2 shows that other important elements in *Eucalyptus* including elements Ca, Mg, and Zn have good correlation and levels of accuracy, but Ca had a very high error rate. This was unexpected as calcium is associated with many biological molecules. Using global models ensures the level of accuracy will be consistent and the level of error compared to the levels of single species model. An advantage of a global model is that it is not susceptible to varietal impurity. Improving the model requires refining and testing different pre-processing options to enhance the model performance. Some

elements exist in soluble ionic states and the remainder in forms that interact with biomolecules, e.g., N in amino acid (Dietzen, 2018). There can be discrepancies in spectral acquisition of these elements because reference data measures the total nutrient in the foliar material. This can result in underestimation of the nutrient if the NIRS can only measure the nutrients that are interacting with the constituent of biomolecules.

In summary, the performance of the models that estimate nutrient levels are affected by the accuracy of the reference methods; the number and presentation of samples used to develop the models; the interactions between elements and organic molecules, i.e. whether they are in ionic or organic states, which determines their responsiveness to infrared light; and the pre-processing methods and the chemometric algorithms employed. Future work should aim to develop methods that are capable of measuring both inorganic and organically assimilated elements to provide better indications of plant nutritional status.

3.5 Conclusion

Considering the need for a rapid foliar nutrient assessment, the evaluated NIRS method of analysis showed positive results for certain important micro and macro elements. Certain elements were accurately predicted, which highlights that the model further development is required and possibly an increase in sample numbers can be considered to increase model performance for all parameters. Successful development of NIRS methods can contribute significantly to sustainable nursery seedling production.

3.6 References

- Barbosa, L.C.A., Filomeno, C.A. and Teixeira, R.R. (2016). Chemical variability and biological activities of *Eucalyptus* spp. essential oils. *Molecules*, 21(12), pp.1671
- Comino, F., Ayora-Cañada, M.J., Aranda, V., Díaz, A., and Domínguez-Vidal A. (2018). Near-infrared spectroscopy and X-ray fluorescence data fusion for olive leaf analysis and crop nutritional status determination. *Talanta*, 188, 676-684.
- Dietzen, D.J. (2018). Amino acids, peptides, and proteins. In *Principles and Applications of Molecular Diagnostics* (pp. 345-380). Elsevier.
- Gonçalves, J.D, Stape, J.L., Laclau, J.P., Bouillet, J.P. and Ranger, J. (2008). Assessing the effects of early silvicultural management on long-term site productivity of fast-growing eucalypt plantations: the Brazilian experience. *Southern Forests* 70(2): pp.105-18
- Grossnickle, S.C. and MacDonald, J.E., (2018). Seedling quality: history, application, and plant attributes. *Forests*, 9(5), pp.283.
- Herbert, M.A. and Schönau, A.P.G. (1990). Fertilising commercial forest species in southern Africa: Research progress and problems (Part 2). *South African Forestry Journal* 152: pp34-42
- Mohanani, C. and Sharma, J.K. (2005). Improvement of seedling production system in forestry sector and its impact on seedling health. *Kerala for Research Institute*, 11, pp.77-82.
- Nobre, S.R. and Rodriguez, L.C. (2014). Integrating Nursery and Planting Activities. *The Management of Industrial Forest Plantations: Theoretical Foundations and Applications*, pp.347-371.
- Neto, A.J.L., Neves, J.C., Martinez, H.E., Sousa, J.S., and Fernandes, L.V. (2020). Establishment of critical nutrient levels in soil and plant for *Eucalyptus*. *Revista Brasileira de Ciência do Solo* 44: e0190150
- Prananto, J.A., Minasny, B. and Weaver, T. (2020). Near infrared (NIR) spectroscopy as a rapid and cost-effective method for nutrient analysis of plant leaf tissues. *Advances in Agronomy*, 164, pp.1-49.
- Rahman, M.A. (2006). Clonal Forestry (Doctoral dissertation, Khulna University).
- Ramesar, D.R., Cooper, G.D., Naidoo, D.S., and Burgdorf, R.J. (2024). ICFR Forestry Research and Analytical Laboratory Analysis Manual of Standard Operating Procedures.

- Reich, G. (2016). Mid and Near Infrared Spectroscopy. In: Müllertz A., Perrie, Y., Rades, T. (eds) *Analytical Techniques in the Pharmaceutical Sciences. Advances in Delivery Science and Technology*. Springer, New York, NY. https://doi.org/10.1007/978-1-4939-4029-5_3
- 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, pp.100676.
- Sun, W., Zhang, X., Zhang, Z., and Zhu, R. (2017). Data fusion of near-infrared and mid-infrared spectra for identification of rhubarb. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 171, pp.72-79.
- Uchida, R. (2000). Essential nutrients for plant growth: nutrient functions and deficiency symptoms. *Plant Nutrient Management in Hawaii's Soils*, 4, pp.31-55.

Chapter 4

Evaluation of the ability of near-infrared spectroscopy to assess varietal purity of *Eucalyptus* hedges

Abstract

The accurate verification of clonal identity and purity amongst nursery hedges is a critical step in forestry industries to ensure uniformity, optimize productivity, and maintain genetic integrity. Traditional methods, such as morphological assessments and molecular techniques, while reliable, are often labour-intensive, time-consuming, and cost-prohibitive. The objective of the study is to develop an NIR screening method to monitor varietal purity of clonal cuttings of nursery production of *Eucalyptus* material. The foliar material of 4 *Eucalyptus* varieties (G×U400, G×NPP, GL222 and G×U488) were sampled at Top Crop Nursery (Cramond, Pietermaritzburg) from commercial hedges. The samples were dried, milled and presented to the NIRS for spectra acquisition and the reference data used for model calibration relied on the nursery label of the hedges used. The Ident on Opus software was used to develop a classification model to discriminate between the varieties. It showed the potential to differentiate between groups of *Eucalyptus* hybrids based on the 2-d and 3-d PCA analysis. The confusion matrix showed the reliability of the model to classify the clones accurately. The prediction statistics for accurate classification of the independent data test set for G×NPP clones has 100% accuracy, 100% precision and 100% specificity while G×U400 clones has 60% accuracy, precision of 0% and specificity of 75%, G×U488 clones accuracy is 92%, precision of 83.33% and specificity of 94.74%, GL222 accuracy is 68%, precision of 55.56% and specificity of 75%. With the independent testing, the model that was developed was able to identify unknown sample to correct groups with accuracy. This study demonstrated the feasibility of near-infrared reflectance spectroscopy (NIRS) as a rapid, cost-effective, and non-destructive alternative for assessing clonal purity. Also, this study highlights the potential of NIRS as a transformative tool for clonal verification in large-scale forestry operations, with implications for improving efficiency and reducing operational costs.

4.1 Introduction

Eucalyptus plantations are regarded as the most productive plantations due to growth rates dependent on genetics, silviculture, and climate (Binkley *et al.*, 2017). Properly managed plantations employ genetically identical clonal trees to maximize uniformity and growth. A clone is defined as any group of genetically identical units that are either unattached or temporarily attached together and can exist independently (Sebens, 1985; Resende *et al.*, 2012). The practice of clonal propagation via cutting practices enables the replication of desired traits such as growth rate, disease resistance, and wood quality in trees. However, maintaining varietal clonal purity is essential to preserve the genetic consistency of clones, which directly impacts the success of downstream forestry operations (Denison and Kietzka, 1993).

Several methods are currently utilised to determine varietal purity in forestry and each method has challenges and advantages (Aimers-Halliday *et al.*, 2003). These methods help to detect genetic variations or contamination that may compromise the performance and uniformity of clonal populations. The traditional method involves visual inspection of the physical characteristics of seedlings, such as leaf shape, tree height, bark texture, and growth patterns (Li *et al.*, 2022). Morphological assessments are easy to perform but can be subjective and may fail to detect subtle genetic differences for long-term performance and impact. Other methods include molecular-based techniques such as gene sequencing and fingerprinting. Despite being accurate, these methods are expensive and not viable at the nursery scale. This highlights the constraints of reliable assessment of clonal verification. Therefore, methods that are cost-effective, high-throughput and accurate need to be assessed or developed.

Near-infrared reflectance spectroscopy (NIRS) is a rapid, non-destructive and cost-effective analysis method that can revolutionize agriculture (Singh *et al.*, 2022). It utilizes the unique spectral signatures of the scanned samples and extrapolates or predicts the value of properties of interest using trained machine learning algorithms. This technology has been used to measure nutrient content, oil content and proteins and determine the age of *Eucalyptus* plant. It has been used to identify clonal varieties of rubber trees and type of oil content (Humphreys *et al.*, 2007; da Silva *et al.*, 2021). Therefore, it is plausible that NIRS can be used to develop algorithms capable of distinguishing different varieties of *Eucalyptus* for South African forestry nurseries.

This study was conducted to evaluate whether NIRS can be used to differentiate between different *Eucalyptus* varieties of importance to the South African Forestry Sector. Varieties of importance included G×U400, G×NPP, GL222 and G×U488 which were hybrids of *Eucalyptus grandis* (G), *E. urophylla* (U), *E. nitens* (N) and *E. longiostrata*. This study aims to contribute to the development of Precision Forestry at the nursery level in South Africa.

4.2 Materials and methods

4.2.1 Sample collection

Eucalyptus hedges were accessed at Top Crop Nursery (Crammond, Pietermaritzburg) between March 2022 and September 2022. Only foliar material was collected, and each leaf was cut at the petiole. A total of 35-40 leaves of *Eucalyptus* species were placed in brown paper bags for drying. Before oven drying, a leaf tip from each sample was collected, placed in Eppendorf tubes and stored at -4°C for future molecular identification using DNA techniques. The sampling took place at different tunnels located at the Top Crop Nursery.

4.2.2 Sample processing

The sampled foliar material was dried on arrival at the Institute for Commercial Forestry Research (ICFR), Forestry Research and Analytical Laboratory (FRAL) at 60°C for 72 hours. Thereafter, the samples were milled at 6000 rpm on a Retch ZM 200 mill (Germany) equipped with a 0.5 mm particle size screen. The milled samples were filled into a 20 mm diameter quartz vial for NIR spectral acquisition.

4.2.3 NIR model development

The NIR model development was carried out using IDENT package on OPUS Software to develop a clustering model for clone purity monitoring. The sample processing for spectral acquisition was conducted as described in Section 2.2.4.

4.3 Results

The NIRS analysis successfully demonstrated the ability to classify the tested *Eucalyptus* species according to their respective group.

4.3.1 Cluster analysis

Principal component analysis (PCA) using the first and second components was able to separate the samples into their species clusters based on their NIR spectra (Figure 4.1). This was achieved by plotting the distribution of the NIR data against PC1 and PC2 as the x-axis and y-axis. All the tested species had clear groupings. The G×U400 (green) and G×U488 (blue) clusters are located far away from other species, indicating they are highly different from one another.

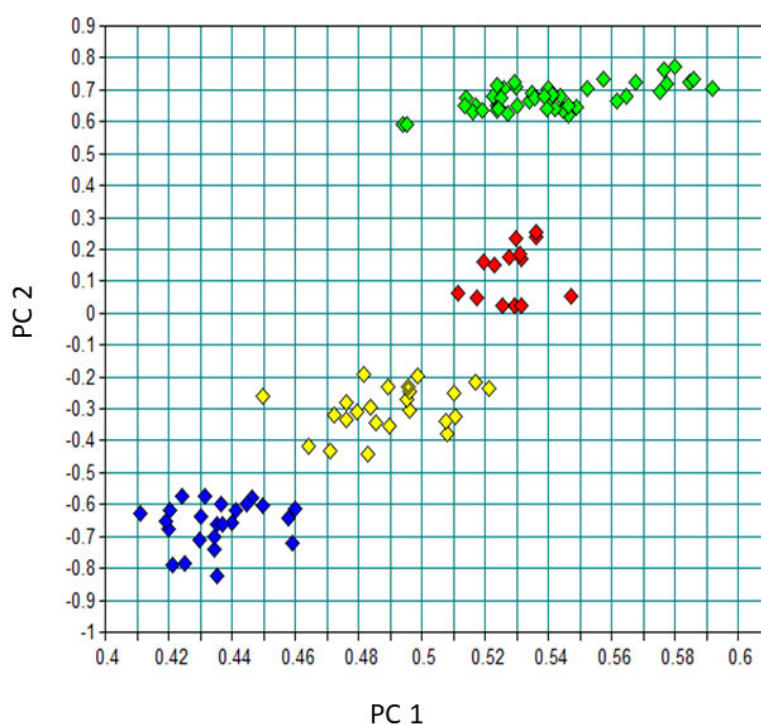


Figure 4.1: The 2-d PCA cluster presentation of data points are shown to be grouped to distinctive clusters. The figure shows the PCA, whereby the x and y-axis were PC 1 and PC 2, respectively. The varieties of *Eucalyptus* were colour coded as green for G×U400, red for G×NPP, yellow for GL222 and blue for G×U488.

Figure 4.2 shows the distinct clustering of data points into four groups by a 3-D PCA plot. The relative inter-cluster distances indicate the level of dissimilarity between groups. GL222 (yellow) is farther from G×U400 (green) and G×U488 (blue) than from G×NPP (red), suggesting that GL222 and G×NPP might be more alike. G×U400

(green) and G×U488 (blue) are clusters that are furthest apart even though they are genetically most similar as they are both G×U crosses.

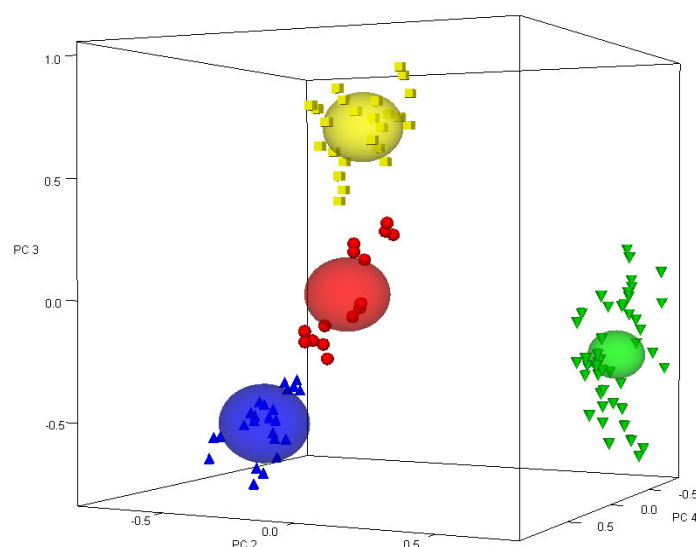


Figure 4.2: The 3-D PCA presentation of the sphere of clustered data set. The coloured coded as green for G×U400, red for G×NPP, yellow for GL222 and blue for G×U488.

The NIR could discriminate the samples of clonal purity according to their representative groups (Figure 4.2). This shows that NIRS analysis using the Bruker OPUS IDENT module can be effective in the classification of the clones in their respective group using spectral readings of their foliar material.

4.3.2 Independent evaluation of model reliability

The independent test set to establish the model capability to predict clone purity on samples is shown in Table 4.1. This shows that the model can correctly predict to identity G×NPP, GL222 and G×U488. However, G×U400 identity was predicted poorly.

Table 4.1: The presentation of confusion matrix of independent data testing.

		Actual			
Predicted		G×NPP	G×U 400	G×U 488	GL222
	G×NPP	5	0	0	0
	G×U 400	0	0	1	4
	G×U 488	0	1	5	0
	GL222	0	4	0	5

The confusion matrix in Table 4.2 suggests that the model was effective at predicting well.

Table 4.2: The presentation of how the method performed with the actual data to predict the unknown samples and model robustness.

	G×NPP	G×U 400	G×U 488	GL222
True Positive (%)	20.00	0.00	20.00	20.00
False Negative (%)	0.00	20.00	4.00	16.00
False Positive (%)	0.00	20.00	4.00	16.00
True Negative (%)	80.00	60.00	72.00	48.00
Totals (%)	100	100	100	100

The model performance parameters show that the model was reliable in distinguishing between the classes.

Table 4.3: *Eucalyptus* varietal-wise prediction statistics.

	G×NPP	G×U 400	G×U 488	GL222
Accuracy (%)	100.00	60.00	92.00	68.00
Precision (%)	100.00	0.00	83.33	55.56
Specificity (%)	100.00	75.00	94.74	75.00

4.4 Discussion

The study investigated whether NIR spectroscopy can be used to differentiate between different clones of *Eucalyptus* species by developing a model to categorise or group the varieties based on their spectral signatures. The work aimed to address the incorrect labelling of clones in nursery hedges, which cause issues at the propagation and forest establishment phase.

The near-infrared spectra of *Eucalyptus* species leaves possessed significantly different properties that allowed them to be grouped in separate clusters using PCA (Figure 4.1 and Figure 4.2). This suggests that the groups (GL222, G×NPP, G×U400, and G×U488) have unique characteristics that are distinguishable in the principal component space and that due to sample composition variation and differences a model can be developed to discriminate between samples. Although the PCA grouping did not guarantee 100% accuracy, but it provided robust information that classification model can be developed successful. Lu *et al.* (2021) described that NIR PCA analysis could distinguish between hybrids of *E. urophylla*, which motivated that the genetic variation between *Eucalyptus* clones was measurable using NIR and warranted that further development of the varietal purity models.

The NIR categorical model was successful in correctly assigning the different clones to their expected categories, with the exception of G×U400 (Table 4.2). Comparing the accuracy, precision and specificity, the model showed the highest levels of performance when categorising/identifying G×NPP and G×U488 (Table 4.2), whereas G×U 400 and GL222 had the lowest performance metrics and generally identified as each other, suggesting the possibility of contamination between the hedges or samples taken for model development. This emphasises the importance of the development of clonal or varietal purity models. For reliable categorical models to be

developed, molecular genotyping *Eucalyptus* samples using PCR or molecular fingerprinting is proposed to prevent incurred sample assignment and ensure that the correct information is used to calibrate the spectra (Huong and Huan, 2023; Altememmi *et al.*, 2024). Additionally, the performance of the models is expected to improve by increasing the number of samples used for both calibration and validation, thereby associating more spectral variation of the samples to their assigned variety class (Ramolha *et al.*, 2018).

Overall, prediction results showed potential for positive identification, indicating that NIRS can be used to develop successful variety classification or identification models (Lu *et al.*, 2019). This is supported by findings that near-infrared reflectance spectroscopy could rapidly identify rubber tree clones and discriminate leaves at different stages (young or adult) of development (da Silva *et al.*, 2021). Furthermore, this indicates there are more factors to consider that influence the performance of categorical NIR models such as organ/tissue scanned and the age of the plant. More work is required to develop a robust model by incorporating more commercial clonal *Eucalyptus* varieties and using molecular testing to validate the identity of samples used to develop calibration models and for independent test sets.

4.5 Conclusion

The project was initiated to address the need for the industry to rapidly and frequently monitor clonal purity during production to minimise error that can arise in the nursery operations. NIRS model can be develop for daily use by the industry. Based on the success of this work, further research is recommended to refine spectral models and validate the approach across diverse genetic backgrounds and environmental conditions.

4.6 References

- Aimers-Halliday, J., and Burdon, R. D. (2003). Risk management for clonal forestry with *Pinus radiata*—analysis and review. 2: technical and logistical problems and countermeasures. *New Zealand Journal of Forestry Science*, 33(2), pp. 181-204.
- Al-tememmi, N.J., Al-judy, N.J. and Al-zubaidi, L.A., (2024). Genetic Diversity and Polymorphism of *Eucalyptus* L' Species by ITS Sequence. *Iraqi Journal of Science*, 65(6), pp.3040-3046.
- Binkley, D., Campoe, O.C., Alvares, C., Carneiro, R.L., Cegatta, Í. and Stape, J.L. (2017). The interactions of climate, spacing and genetics on clonal *Eucalyptus* plantations across Brazil and Uruguay. *Forest Ecology and Management*, 405, pp. 271-283.
- da Silva, A.L D., Alves Filho, E.G., Silva, L.M.A., Tavares, O.C.H., Pereira, M G., de Campos, T. and da Silva, L.M. (2021). Near infrared spectroscopy to rapid assess the rubber tree clone and the influence of maturation and disease at the leaves. *Microchemical Journal*, 168, 106478.
- Denison, N.P. and Kietzka, J.E. (1993). The development and utilisation of vegetative propagation in Mondi for commercial afforestation programmes. *South African Forestry Journal*, 166(1), pp.53-60.
- Humphreys, J.R., O'Reilly-Wapstra, J.M., Harbard, J.L., Davies, N.W., Griffin, A.R., Jordan, G.J., and Potts, B.M. (2008). Discrimination between seedlings of *Eucalyptus globulus*, *E. nitens* and their F hybrid using near-infrared reflectance spectroscopy and foliar oil content. *Silvae Genetica*, 57(1-6), pp. 262-269.
- Huong, B.T.M. and Van Huan, H., 2023. Identification of DNA barcode sequence of hybrid *Eucalyptus* UP99 (*E. urophylla* x *E. pellita*) and UP95 (*E. urophylla* x *E. pellita*) to identify plant varieties. *Journal of Forestry Science and Technology*, 8(2), pp.036-046.
- Li, H., Zhang, X., Li, Z., Wen, J., and Tan, X. (2022). A review of research on tree risk assessment methods. *Forests*, 13(10), pp. 1556.
- Lu W.H., Qi J. and Luo J.Z. (2019). Application of NIRS for discriminant of *Eucalyptus* hybrids. *Spectroscopy Spectral Anal.* 39: 1451–1457.

- Lu, W., Arnold, R.J., Wang, C., Lin, Y., Luo, J., Meder, R. and Yang, Y. (2021). Defining *Eucalyptus urophylla* with its hybrid and the rules of genetic recombination using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 29(4), pp.236-244.
- Ramvalho, F.M., Andrade, J.M. and Hein, P.R. (2018). Rapid discrimination of wood species from native forest and plantations using near infrared spectroscopy. *Forest Systems*, 27(2), pp. e008 (9 pages).
- Rezende, G.D.S., de Resende, M.D.V. and de Assis, T.F. (2013). *Eucalyptus* breeding for clonal forestry. In *Challenges and Opportunities for the World's Forests in the 21st Century* (pp. 393-424). Dordrecht: Springer Netherlands.
- Sebens, K.P. (1985). Coexistence of clones, clonal diversity, and the effects of disturbance. pp. 357-398 in J.B.C. Jackson, L.W. Buss, and R.E. Cook, eds. *Population biology and evolution of clonal organisms* (17th eds). Yale Univ. Press, New Haven, CT.
- 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, pp.100676.

Chapter 5

Thesis overview

5.1 Introduction

Forestry is a significant contributor to global productivity and a major source of income for South Africa. The major issue with different assessment of seed germination rate and seed vigour, hedge nutrient content levels and hedge clonal purity requires vigorous improvement. The sample destruction are the contributing factors in seed loss or decrease in cutting production and mother plant recovery; also, expensive laboratory analysis costs; human error in analysis and variation in sample size and sample pre-processing procedures. Sample analyses are time-consuming and delay in results can cause economic loss to the producer due to either non-viable seeds can be sold to the nurseries, production of poor state seedlings or the use of unnecessary fertiliser to revive the seedlings to a good nutritional state, and the identification of wrongly planted seedling can be costly to foresters at harvest.

This work aimed to evaluate near-infrared reflectance spectroscopy (NIRS) to increase nursery efficiency and reduce costs to allow nurseries to produce optimal seedlings. The research objectives of the study were to develop NIR models capable of: (1) predicting seed germination rate and vigour of *Eucalyptus dunnii*, *E. nitens*, *E. benthamii*, and *E. macarthurii*; (2) estimating nutrition levels in nursery seedlings, and (3) screening for varietal purity of clonal cuttings. At each objective milestone, NIR models showed promising levels of performance and motivates for further model development.

5.2 Research objectives and major findings

Therefore, this study compartmentalised each process as a strategic objective to improve the use of NIRS and as described in each chapter. A summary of the research objectives and major findings of each chapter is described below.

Chapter 2: Screening of *Eucalyptus* species for seed germination and vigour using near-infrared reflectance spectroscopy

- Multiple seed lots samples of *Eucalyptus dunnii*, *E. nitens*, *E. benthemii*, and *E. macarthurii* showed that an increase in sample variation increases the accuracy of the NIR model.

- Developing seed germination rate NIR prediction model for a single species produced suitable models but an increase in sample size will also increase the reliability and reduce the error of the model.
- With NIR, it is possible to train calibration models and develop test validation models that produce predictions that are close to the reference values obtained from the traditional testing methods. These observations can assist seed producers and nurseries to rapidly determine the germination rate of seed lots.

Chapter 3: Evaluation of near-infrared reflectance spectroscopy for low-cost foliar nutrient analysis of *Eucalyptus* seedlings

- The NIR screening model used to estimate nutrition levels in nursery seedlings using foliar material was successful.
- The cultivation of major varieties of forestry seedlings can be monitored easily and efficiently using NIR to monitor the nutrient levels and determine the required nutrients in real-time prediction reports using the NIR on *E. dunnii*, hybrids G×U, and G×N.
- A global nutrient elemental model was developed, but single-element models were shown to be more accurate.
- The models were successfully created using the OPUS software using chemometrics such as PLS regression.

Chapter 4: Evaluation of the ability of near-infrared spectroscopy to assess varietal purity of *Eucalyptus* hedges

- The NIR model was developed using foliar materials of different *Eucalyptus* species to create a classification model for the tested species of hedge clonal purity in the nursery.
- The NIR model developed could discriminate between the samples of G×U400, G×NPP, GL222 and G×U488.
- The NIRS model developed together with the PCA was able to categorise the samples according to their respective species, but signs of contamination were

seen for lower performance. This can assist the industry to minimise human error and keep the production lines with pure species.

5.3 Recommendations and future research

This study demonstrated that NIR can be used to monitor different aspects of seedling production in the future for the industry. The capability of NIR to estimate or predict seed germination rate was as reliable as the standard germination testing. This is effective for the seed producers as to ensure the consumers the reliability of the seed lots and can give nurseries the vigour level of the seed lots to be used for seedling production. NIR was also observed to be able to estimate nutrient levels, which is important for the production of healthy seedlings for the consumer and a successful establishment in the field plantation. Pure quality clone production is a vital step in the nursery as the clone carries the genetic desired for that specific species or is environmentally compatible, and yield levels are being enhanced based on the breeder's goal. The NIR models require further development to improve accuracy and robustness.

This study showed that NIRS could be a valuable tool for optimising and supporting the South African commercial forestry industry by improving the operational or quality control capabilities at nurseries. Using NIRS has the potential to overcome many financial hurdles that limit extensive analysis of seed lots and nursery hedges. Furthermore, this contributes to improving access to precision forestry in the country by improving productivity at the nursery level, which impacts the downstream processes associated with it. Therefore, NIRS should be considered when improving forestry operations and warrants further development into a tool for routine analysis.

Appendices

Appendix A: Near-infrared reflectance spectroscopy for low-cost foliar nutrient analysis of forestry seedlings

Table A1: Independent test dataset for NPK on Global model and single species (*E. dunnii*, hybrid G×N, and G×U).

Element	<i>E. dunnii</i>		G×N		G×U	
	True	Predicted	True	Predicted	True	Predicted
N (%)	1.707	1.88	1.0874	1.2168	0.78749	0.80191
	2.1296	2.26	0.94126	0.92187	0.92299	0.92404
	1.4911	1.62	0.85543	0.86442	0.78324	0.89637
	1.7193	1.89	0.90162	0.9383	0.72343	0.74604
	1.798	1.93	1.5632	1.4045	1.0871	0.91104
P (mg/kg)	1187.3	1031.33	819.62	929.79	846.57	525.23
	2267	1155.45	743.93	932.75	822.19	558.95
	586.19	969.20	765.14	800.09	1292.27	775.55
	571.59	1153.59	708.06	352.69	881.54	657.36
	423.57	1164.16	952.14	2113.6	892.6	1035.5
K (mg/kg)	7013	9105.30	1635.73	1284.3	2020.42	3632.4
	6362.9	9909.11	2512.44	2730.5	2247.69	1601.7
	6146.9	5760.77	2376.68	2103.2	4924.3	2217.9
	7101.9	5770.49	2182.1	2706.6	1830.32	3644.5
	7919	8534.90	3278.68	6065.4	2294.13	4262.8

The true versus predicted nutrient data was produced from an independent sample set comprised of 5 plants per varietal nutrient model.

Table A2: NIR report of N, P and K values predicted for different *Eucalyptus* hybrids.

Element	E. dunnii			G×N			G×U		
	True	Predicted	Single	True	Predicted	Single	True	Predicted	Single
N (%)	1.707	1.88	0.78538	1.0874	1.2168	16,719	0.78749	0.80191	0.93055
	2.1296	2.26	0.96797	0.94126	0.92187	16,069	0.92299	0.92404	0.9208
	1.4911	1.62	0.73072	0.85543	0.86442	0.7651	0.78324	0.89637	12.186
	1.7193	1.89	0.83743	0.90162	0.9383	10,104	0.72343	0.74604	17.018
	1.798	1.93	0.98376	1.5632	1.4045	13,597	1.0871	0.91104	1.311
P (mg/kg)	1187.30	1031.33	892.99	819.62	929.79	2109.00	846.57	525.23	911.00
	2267.00	1155.45	822.56	743.93	932.75	2122.10	822.19	558.95	691.30
	586.19	969.20	490.27	765.14	800.09	1948.90	1292.27	775.55	1730.80
	571.59	1153.59	723.54	708.06	352.69	2717.90	881.54	657.36	1279.10
	423.57	1164.16	571.89	952.14	2113.60	1970.20	892.60	1035.5	1380.10
K (mg/kg)	7013.00	9105.30	1534.80	1635.73	1284.30	1955.10	2020.42	3632.40	1947.70
	6362.90	9909.11	1419.90	2512.44	2730.50	1947.70	2247.69	1601.70	2112.10
	6146.90	5760.77	4295.20	2376.68	2103.20	2112.10	4924.30	2217.90	2605.90
	7101.90	5770.49	2035.90	2182.10	2706.60	2605.90	1830.32	3644.50	4350.30
	7919.00	8534.90	1862.20	3278.68	6065.40	8832.80	2294.13	4262.80	5507.30