The potential for using remote sensing to quantify stress in and predict yield of sugarcane (*Saccharum* spp. hybrid)

Elfatih Mohamed Abdel-Rahman

A thesis submitted to the Faculty of Science and Agriculture, at the University of KwaZulu-Natal, in fulfilment of the requirements for the degree of Doctor of Philosophy in Environmental Sciences

January 2010
Pietermaritzburg
South Africa

Abstract

South Africa is the leading producer of sugarcane in Africa and one of the largest sugarcane producers in the world. Sugarcane is grown under a wide range of climatic, agronomic, and socio-economic conditions in the country. Stress factors such as water and nutrient deficiencies, and insect pests and diseases are among the most important factors affecting sugarcane production in the country. Monitoring of stress in sugarcane is therefore essential for assessing the consequences on yield and for taking action of their mitigation. The prediction of sugarcane yield, on the other hand is also a significant practice for making informed decisions for effective and sound crop planning and management efforts regarding e.g., milling schedules, marketing, pricing, and cash flows. In South Africa, the detection of stress factors such as nitrogen (N) deficiency and sugarcane thrips (*Fulmekiola serrata* Kobus) damage and infestation are made using traditional direct methods whereby leaf samples are collected from sugarcane fields and the appropriate laboratory analysis is then performed. These methods are regarded as being time-consuming, labour-intensive, costly, and can be biased as often they are not uniformly applied across sugarcane growing areas in the country. In this regard, the development of systematically organised geo-and time-referenced accurate methods that can detect sugarcane stress factors and predict yields are required. Remote sensing offers near-real-time, potentially inexpensive, quick and repetitive data that could be used for sugarcane monitoring. Processing techniques of such data have recently witnessed more development leading to more effective extraction of information. In this study the aim was to explore the potential use of remote sensing to quantify stress in and predict yield of sugarcane in South Africa.

In the first part of this study, the potential use of hyperspectral remote sensing (i.e. with information on many, very fine, contiguous spectral bands) in estimating sugarcane leaf N concentration was examined. The results showed that sugarcane leaf N can be predicted at high accuracy using spectral data collected using a handheld spectroradiometer (ASD) under controlled laboratory and natural field conditions. These positive results prompted the need to test the use of canopy level hyperspectral data in

predicting sugarcane leaf N concentration. Using narrow NDVI-based vegetation indices calculated from Hyperion data, sugarcane leaf N concentration could reliably be estimated.

In the second part of this study, the focus was on whether leaf level hyperspectral data could detect sugarcane thrips damage and predict the incidence of the insect. The results indicated that specific wavelengths located in the visible region of the electromagnetic spectrum have the highest possibility of detecting sugarcane thrips damage. Thrips counts could also adequately be predicted for younger sugarcane crops (4–5 months).

In the final part of this study, the ability of vegetation indices derived from multispectral data (Landsat TM and ETM+) in predicting sugarcane yield was investigated. The results demonstrated that sugarcane yield can be modelled with relatively small error, using a non-linear random forest regression algorithm.

Overall, the study has demonstrated the potential of remote sensing techniques to quantify stress in and predict yield of sugarcane. However, it was found that models for detecting a stress factor or predicting yield in sugarcane vary depending on age group, variety, season of sampling, conditions at which spectral data are collected (controlled laboratory or natural field conditions), level at which remotely-sensed data are captured (leaf or canopy levels), and irrigation conditions. The study was conducted in only one study area (the Umfolozi mill supply area) and very few varieties (N12, N19, and NCo 376) were tested. For practical and operational use of remote sensing in sugarcane monitoring, the development of an optimum universal model for detecting factors of stress and predicting yield of sugarcane, therefore, still remains a challenging task. It is recommended that models developed in this study should be tested – or further elaborated – in other South African sugarcane producing areas with growing conditions similar to those under which the predictive models have been developed. Monitoring of sugarcane thrips should also be evaluated using remotely-sensed data at canopy level; and the ability of multispectral sensors other than Landsat TM and ETM+ should be tested for sugarcane yield prediction.

Preface

The research work described in this thesis was carried out in the School of Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, from August 2005 to November 2009, under the supervision of Prof. Fethi B. Ahmed (School of Environmental Sciences, University of KwaZulu-Natal; UKZN, South Africa) and Dr. Maurits van den Berg, (South African Sugarcane Research Institute; SASRI).

I would like to declare that the research work reported in this thesis has never been submitted in any form to any other university. It therefore represents my original work except where due acknowledgments are made.

Elfatih Mohamed Abdel-Rahman _____ January 2010

As the candidate's supervisors, we certify the above statement and have approved this thesis for submission.

1. Prof. Fethi B. Ahmed Signed:_____ Date:_____

2. Dr. Maurits van den Berg Signed:_____ Date:_____

Declaration 1 – Plagiarism

I, Elfatih Mohamed Abdel-Rahman, declare that:

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at 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, then:
    a. Their words have been re-written but the general information attributed to them has been referenced
    b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References section.

Signed_____

Declaration 2 – Publications and manuscripts

1. **Abdel-Rahman**, E. M. and Ahmed, F. B., 2008. The application of remote sensing techniques to sugarcane (*Saccharum* spp. Hybrid) production: a review of the literature. *International Journal of Remote Sensing*, 29, 3753–3767.

2. **Abdel-Rahman**, E. M., Ahmed, F. B. and van den Berg, M., 2008. Imaging spectroscopy for estimating sugarcane leaf nitrogen concentration. *Proceedings of SPIE Remote Sensing for Agriculture, Ecosystems, and Hydrology X Conference*, V-1 – V-12.

3. **Abdel-Rahman**, E. M., Ahmed, F. B., van den Berg, M. and Way, M. J., 2008. Preliminary study on sugarcane thrips (*Fulmekiola serrata*) damage detection using imaging spectroscopy. *Proceedings of South African Sugar Technologists' Association*, 81, 287–289.

4. **Abdel-Rahman**, E. M., van den Berg, M., Way, M. J., Ahmed, F. B. and Sewpersad, C., 2009. Using spectroscopic data sets to predict numbers of thrips (*Fulmekiola serrata*) in sugarcane. *Proceedings of South African Sugar Technologists' Association*, 82, 441– 445.

5. **Abdel-Rahman**, E. M., van den Berg, M. Way, M. J., and Ahmed, F. B., 2009. Handheld spectrometry for estimating thrips (*Fulmekiola serrata*) incidence in sugarcane. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, IV-268 – IV-271.
6. **Abdel-Rahman**, E. M., Ahmed F. B, van den Berg, M. and Way, M. J., 2009. Potential of spectroscopic data sets for sugarcane thrips (*Fulmekiola serrata* Kobus) damage detection. *International Journal of Remote Sensing*, In Press.

7. **Abdel-Rahman**, E. M., Ahmed, F. B. and van den Berg, M., 2009. Estimation of sugarcane leaf nitrogen concentration using *in situ* spectroscopy. *International*

Journal of Applied Earth Observation and Geoinformation, Remote Sensing for Africa: Special Issue, In Press.

8. **Abdel-Rahman,** E. M. **(**In Preparation). Estimation of sugarcane leaf nitrogen concentration using normalised ratio indices generated from EO-1 Hyperion hyperspectral data.

9. **Abdel-Rahman**, E. M. (In Preparation). Random forest regression for sugarcane yield prediction based on Landsat TM and ETM+ derived spectral vegetation indices.

All the conceptual and experimental work, analysis of data, and preparation of above publications and manuscripts were accomplished by the candidate, Elfatih M. Abdel-Rahman, under the supervision of Prof Fethi B. Ahmed and Dr. Maurits van den Berg. Mr. Mike J. Way, in addition to giving logistical support for the determination of thrips damage and infestation, provided comments on the analysis of the data and assisted in editing some of the manuscripts as a SASRI research project member.

Dedication

To my beloved father, Mohamed Abdel-Rahman Ahmed, much-loved mother Masturah Omar Mohamed, and dearly loved wife, Tamador Musa Abdelrahman, for their constant encouragement and prayers for my success

Acknowledgements

First and foremost I thank Allah for giving me health, power and support, and for making it possible for me to complete my PhD thesis.

My research project was undertaken in conjunction between the University of KwaZulu-Natal (UKZN), the South African Sugarcane Research Institute (SASRI) and the University of Khartoum (U of K). I would like to thank these institutions for giving me the opportunity to read for a PhD. Their financial support for living allowance, field work, satellite imageries, and laboratory analyses was also appreciated.

I thank my supervisors Prof. Fethi B. Ahmed (UKZN) and Dr. Maurits van den Berg (SASRI) for their committed guidance and assistance during the course of this study. You taught me how to be a researcher, I learnt how to critically and scientifically comment on and review research work. I also would like to thank my research project members from SASRI: Mike Way, Ingrid Mthembu, Dr. Stuart Rutherford, Prof. Abraham Singels, Prof. Jan Meyer and Dr. Neil Miles for their comments and advises during my research project meetings. SASRI extension specialist, Mr. Marius Adendorff, I cannot forget your support, helpfulness and hospitality in so many ways while I was collecting data in Umfolozi mill supply area. Mr. Franko Sokolic (formerly UKZN staff), you taught me how to be spatial analyst; you were always ready to help when I was stuck in ArcMap. Your assistance in configuring the ASD sensor was also appreciated.

Sampling in sugarcane fields is not an easy task and is considered by field experts as risky. I am grateful to Cyril Cele, Nitesh Poona, Innocent Shezi and Tholang Mokhele who helped me achieve my goals in the field. I am thankful to Isaac Abboy, Susan Sherriff and Samantha Govender for arranging safe and comfortable transport and accommodation, as well as organising other logistics related to my field work.

I'm also indebted to Dr. Ibrahim Eldeen (University of Khartoum, Sudan), Dr. Hafiz Ahmed Abdelgadir (formerly postdoctoral fellow, UKZN), Helmut Neumann (Satellite

Application Centre, South Africa), Dr. Yvette Everingham (James Cook University, Australia), Dr. Mohamed Elfadel (Sudan University of Science and Technology), GeoTerraImage Company (South Africa), Prof. Mary Martin (University of New Hampshire), Dr. Ron Wehrens (Radboud University Nijmegen), Dr. Micheal Gebreslasie (Medical Research Council, South Africa), Nikki Sewpersad (SASRI), Suveshnee Muneien (UKZN), Kerry Philp (UKZN), Elhadi Adam (UKZN), and Prof Di Scott (UKZN) who were supportive and helpful during my PhD study. Dr. Solomon Tesfamicheal (postdoctoral fellow, UKZN), I was always afraid to give you a graph or a map to comment on. Thanks for your steadfast support as well as academic and social interactions.

My gratitude goes to R Development Core Team for their very powerful open source packages for statistical analysis. Special thanks go to Dr. Riyad Ismail (Sappi forests, South Africa) who introduced me into the world of statistical script writing in R.

My gratitude goes to my family: my mother and father, brothers, sisters and all relatives for the love and committed support. I am sorry for not seeing you for more than four years.

Finally, I am grateful to my wife Tamador. I sincerely say, thank you for your patience and carrying the burden of looking after our kids. It has helped me focus on my PhD work. My achievement as well as the kids' excellence is your reward. My son Mubark and my daughter Eithar, in the future you will understand why I would go to work during weekends and holidays while you were bringing your toys to play.

Table of contents

List of figures

List of tables

CHAPTER ONE

General introduction

1.1 Background

Sugarcane (*Saccharum* spp. hybrid) is grown in the tropics and subtropics (Grof and Campbell, 2001) between latitudes 37^{o} N and 31^{o} S. Over half of the world's sugar supply is derived from sugarcane and the crop is of great importance in the agricultural sectors and in the general economy of many developing countries.

In South Africa, the sugarcane crop is grown under a wide range of climatic, agronomic, and socio-economic conditions (Bezuidenhout and Singels, 2007a). Stress factors such as water and nutrient deficiencies, insect pests, and diseases are among the most important factors affecting sugarcane production in the country. These stress factors can influence the functioning of the crop and consequently limit its growth (Baret *et al*., 2007; Jones and Schofield, 2008). Detection of stress levels to which sugarcane is subjected is therefore essential for assessing the effects on yield and for taking action to mitigate these effects (Estep *et al*., 2004; Baret *et al*., 2007). The prediction of sugarcane yield, on the other hand is also a significant practice for making informed decisions for effective and sound crop planning and management (Bastiaanssen and Ali, 2003; Doraiswamy *et al*., 2004; Zhang *et al*., 2005; Everingham *et al*., 2009). Sugarcane crops are normally grown at different age groups across relatively large areas compared to other field crops. This makes the use of traditional methods of monitoring stress in the crop, and predicting its yield a challenging task. The traditional methods of crop stress monitoring and yield prediction are also labour-intensive, time-consuming, expensive, and tend to be subjective (Jongschaap and Booij, 2004).

Process-based growth model (e.g., Canesim in South Africa) approaches for forecasting sugarcane yield could also result in inaccurate estimates because of the low spatial coverage of the weather stations that provide input for these models. Yield estimates from growth models may also vary considerably from the final yield at harvest (van den Berg *et al*., 2009), because these models are based on weather data only, without considering stress caused by other factors such as nutrient deficiency or pests and diseases (Bezuidenhout and Singels, 2007b). Therefore, real-time methods of monitoring

sugarcane stress and predicting its yield are required to: (1) improve our understanding of the dynamics of sugarcane production systems, (2) serve as an early warning system, enabling more timely corrective or preventive actions to be taken, and (3) provide benchmarks against which changes in production performance could be measured. In this context, remote sensing can play a unique role because of its capability of providing rapid, synoptic, relatively inexpensive and near-real-time data over large areas (de Boer, 1993; Lillesand and Kiefer, 2001; Kumar *et al*., 2003; Aronoff, 2005).

The application of remote sensing in sugarcane production has been explored and was found to be potentially useful in: (1) mapping sugarcane areas in e.g., Australia, Brazil, South Africa, Thailand, India, and Guadeloupe, (2) identifying thermal age groups in e.g., South Africa and Australia, (3) discriminating varieties in e.g., South Africa, Brazil, Australia, and the USA, (4) monitoring water and nutritional stress in e.g., the USA and Australia, (5) detecting insect damage and diseases in Australia, and predicting yield in Brazil, Australia, South Africa, and Reunion Island. See Chapter 2 of the present study for a detailed literature review. The monitoring of nutrient stress (e.g., Nitrogen), and pest and diseases in sugarcane using remotely-sensed data has only been investigated under experimental conditions. On the other hand, the application of remote sensing in predicting yield of sugarcane has only been examined using linear regression models and normalisation methods; so far, in South Africa, without any reliable results. Linear models assume a linear relationship between the dependent (sugarcane spectral information) and the independent (yield) variables, which might not always be the case. Despite their simplicity, the linear models might not explain the high variance in sugarcane yield because of the complex spatio-temporal pattern of sugarcane spectral features. The more complicated multiple non-linear relationships could give more flexibility than the simpler linear models (Motulsky and Christopoulos, 2003) and provide better yield estimates.

The challenge for the researchers in using remote techniques for sugarcane production in South Africa would be to quantify stress (e.g., nitrogen deficiency, thrips damage and

infestation) under on-farm conditions and to provide relatively more accurate models for yield prediction.

Because of time-constrained it is not possible to address all the prevailing stress factors in one study, only two very important stress factors *viz*., nitrogen deficiency, and sugarcane thrips damage and infestation were investigated in the present study.

1.2 Sugarcane nitrogen status

Nitrogen (N) plays a vital role in many biophysiological processes in plants (Clevers and Jongschaap, 2003; Fageria, 2009) and is a part of some important compounds in plants such as enzymes, nucleic acids, proteins, and chlorophyll (Marschner, 1993; Jarmer *et al*., 2003). It is estimated that about 80 million tonnes of N fertilizers are applied every year worldwide by farmers to sustain high crop yields (Epstein and Bloom, 2005). Sugarcane production is one of the intensive agricultural production systems in the world that requires large amounts of N fertilizers because of the high biomass produced by the crop (Wiedenfeld, 1995; Muchow *et al*., 1996). The optimum amount of N to be applied for sugarcane production depends on factors such as growth cycle (start and length), climatic conditions, cultivar, and the physical, chemical and biological properties of the soil (Meyer *et al*., 1986; Wood *et al*., 1996). In South Africa, for example, the optimum recommended N amount for sugarcane production is 60–120 kg N/ha, depending primarily on soil type, crop type (cane crop and ratoon), and sugarcane variety (SASRI, 2000). N deficiency reduces chlorophyll content, leads to short stalks, and lower yield in sugarcane (Rice *et al*., 2006), but excessive N can negatively affect cane quality and increase the susceptibility of the crop to some insects (Atkinson and Nuss, 1989; Rice *et al*., 2006).

The assessment of crop N status is important for making decisions regarding the application of optimum N amounts in the field. Conventionally, sugarcane N status is examined using direct methods of collecting leaf samples from the field followed by

analytical assays in the laboratory (Roth *et al*., 1989). However, these methods do not meet the need for real-time and non-destructive strategies of monitoring crop N status (Zhu *et al*., 2008). Therefore, methods that make use of the advantages of remote sensing, particularly hyperspectral remote sensing (i.e., information on many, very fine, contiguous spectral bands), are needed for diagnosing crop N status.

1.3 Sugarcane thrips

Insect pests are among the most yield reducing factors for most field crops. The losses in the yields of crops because of pests and diseases before harvest were estimated to be about 20% worldwide (Evans, 1993). There are many insects that attack sugarcane and they are reported as being the most significant factors in the economics of sugarcane production in most sugarcane growing regions (Leslie, 2004). Most of these insects are local species that invaded sugarcane from the surrounding habitats (Pemberton and Williams, 1969). However, some, for example sugarcane thrips: *Fulmekiola serrata* Kobus (Thysanoptera: Thripidae), have recently spread into new sugarcane growing regions. Thrips incursion in South Africa was first detected in 2004 in the Umfolozi mill supply area (Way *et al*., 2006) and now is present throughout the sugarcane production areas in the country (Leslie, 2006). The thrips insect feeds on the spindle leaves of sugarcane causing yellowish patches on the leaves (Figure 1.1). Nymphs pierce the central leaf roll and suck the plant sap. Younger crops are relatively more likely to be infested by the insect (Keeping *et al*., 2008). Severe leaf damage might reduce the growth of the crop and probably influence yield, particularly during the summer period in South Africa (Leslie, 2005). Thus, the monitoring and control of thrips is of critical importance.

Currently, thrips in South Africa is only being monitored in a limited number of fields in the Umfolozi mill supply area (van den Berg *et al*., 2009) using the traditional field survey methods, because of (i) a very short period of peak incidence of thrips which would require a large number of well-trained people to conduct the field surveys for monitoring the pest, and (ii) the lengthy process of determining thrips incidence in the

laboratory. The field survey methods are subjective, time-consuming, and expensive. Complementary methods that can provide repetitive, fairly accurate, unbiased means for monitoring thrips damage and infestation are needed. Remote sensing, particularly spectroscopy, offers advanced techniques that can provide such needed thrips monitoring protocols.

Figure 1.1. Typical damage to the sugarcane spindle leaves caused by thrips. Source: Leslie (2005).

1.4 Aim and objectives

The main aim of this study was to explore the potential use of remote sensing to quantify stress in and predict yield of sugarcane in the Umfolozi area, South Africa. Rather than conducting a comprehensive study addressing all aspects of sugarcane production, three contrasting, yet very relevant, broad objectives *viz*., nitrogen detection, thrips monitoring, and yield prediction were selected to explore the strengths and weakness of remote sensing techniques in sugarcane monitoring. Each of these broad objectives had its own specific objectives, which were:

1. To evaluate the use of spectroscopic data in predicting leaf nitrogen concentrations in sugarcane under controlled conditions,
2. To explore the use of *in situ* spectroscopic data to estimate sugarcane leaf nitrogen concentrations,
3. To investigate the potential of imaging spectroscopy in predicting sugarcane leaf nitrogen concentrations using Hyperion data,
4. To examine the potential of spectroscopic data to detect sugarcane thrips damage at leaf level,
5. To investigate the potential use of spectroscopic data in predicting thrips counts at leaf level, and
6. To assess the utility of spectral vegetation indices derived from multispectral Landsat TM and ETM+ data sets in predicting sugarcane yields.

1.5 Scope of the study

This study examines the potential use of remote sensing techniques to quantify stress in and predict yield of sugarcane in South Africa. Two stress factors were investigated: nitrogen and sugarcane thrips. The use of the techniques in monitoring crop yields, which are affected by these stress factors, was subsequently evaluated.

The usefulness of hyperspectral data to predict sugarcane leaf N content was evaluated using a handheld spectrometer under laboratory and field conditions. In a follow-up study, the usefulness of a spaceborne spectrometer (Hyperion) was also evaluated for estimating sugarcane leaf N concentration. Thrips damage and incidence, on the other hand, were also investigated using a handheld spectrometer only. These two stress factors were chosen because nitrogen is the most important nutrient for sugarcane production and sugarcane thrips is a recent pest incursion in South Africa, and any control strategy will benefit from identifying and monitoring this arthropod, particularly in "hotspots". The utility of multispectral data sets for forecasting sugarcane yields was then assessed using two spaceborne Landsat sensors (TM and ETM+).

1.6 Description of the study area

1.6.1 General

South Africa is the world's 10th biggest sugarcane producer (Hassan, 2008) with an estimated mean annual production of about 20 million tonnes of sugarcane (SASA, 2009). Sugarcane production is one of the major contributors to employment, particularly in rural areas, to sustainable development, and to the national economy of the country (SASA, 2009). There are 14 mill supply areas in the country, and these extend from the Eastern Cape province along the coast and midlands of KwaZulu-Natal to Mpumalanga (Figure 1.2). The country has about 38,200 registered growers of sugarcane who employ a large number of people. South African sugarcane production generates an annual income from the sale of sugar and molasses of about US$ 1 billion (SASA, 2009). The contribution of South African sugar production to the national gross domestic product (GDP) was estimated to range between 0.5 and 0.7% (McCarthy, 2007). The milling season normally stretches from April to December. Rainfed, supplementary irrigated, and fully irrigated sugarcane crops are managed by growers in small-scale operations and growers in commercial operations. Umfolozi mill supply area in KwaZulu-Natal was selected as a study area because the situation and challenges facing the mill are largely representative of other mills and because of the existence of the studied stress factors with reasonable variation in the area (e.g. thrips).

.

1.6.2 Umfolozi mill supply area

The Umfolozi mill supply area (28.43° S, 32.18° E) covers an area of 20,565 ha in KwaZulu-Natal (KZN) province of South Africa (Figure 1.3). The long-term annual mean precipitation in Umfolozi ranges from above 1400 mm near the coast to 800 mm inland. Average monthly temperature ranges from 19 °C (inland) to 22 °C (at the coast) and frost is very rare. In addition to sugarcane, commercial forest plantations (mainly *Eucalyptus spp.*) are very common in the area. Similar to the rest of the sugarcane

growing areas in South Africa, there are two types of growers in Umfolozi: large-scale and small-scale. Most of the farms in the large-scale growing sector have full or supplementary irrigation (irrigated or supplementary irrigated growing sectors, respectively) and those farms who cultivated under high rainfall condition towards the coast do not have any kind of irrigation (rainfed growing sector). Growers of the small-scale growing sector cultivate completely under rainfed conditions.

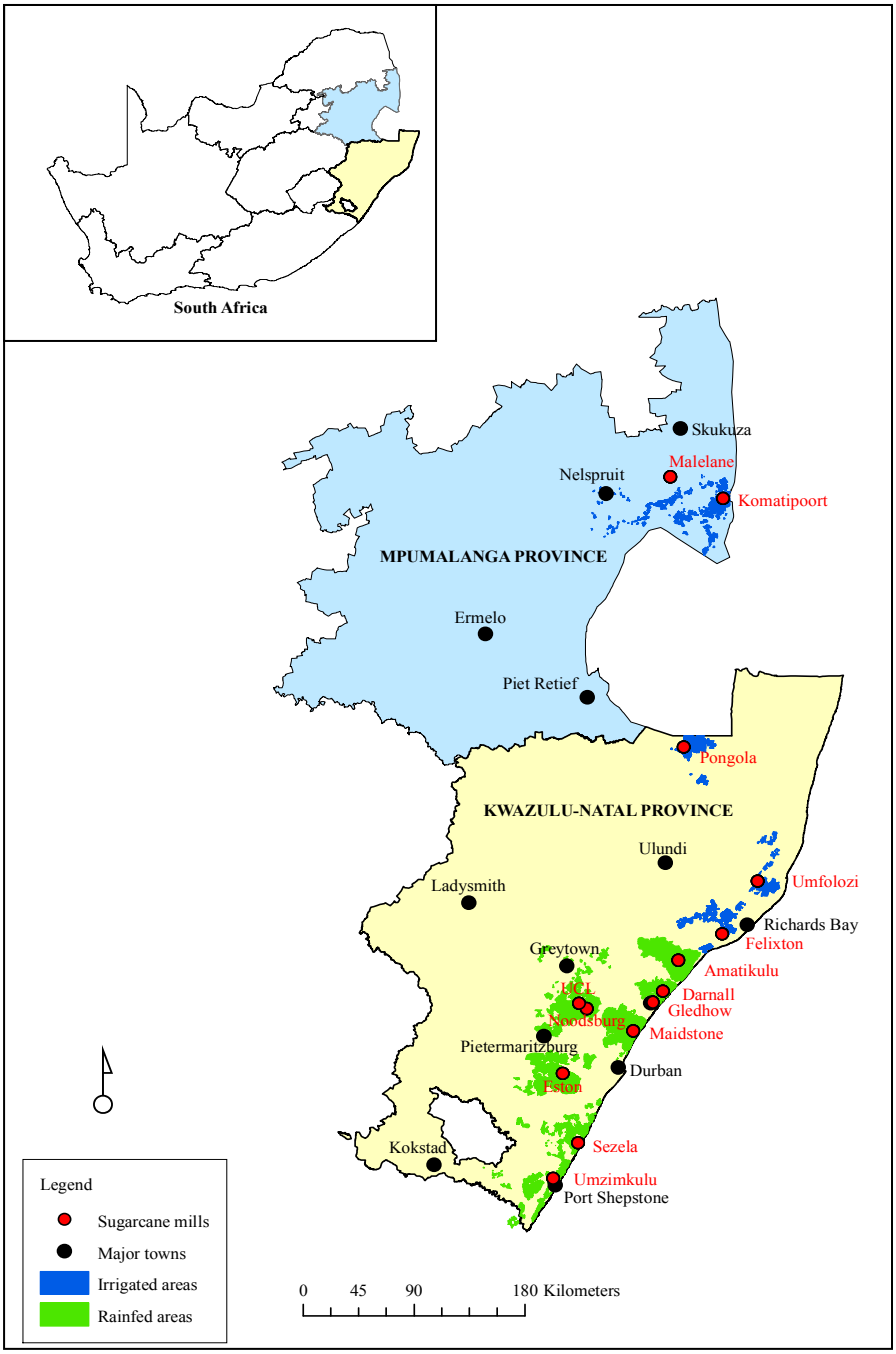

Figure 1.2. Sugarcane production areas in South Africa

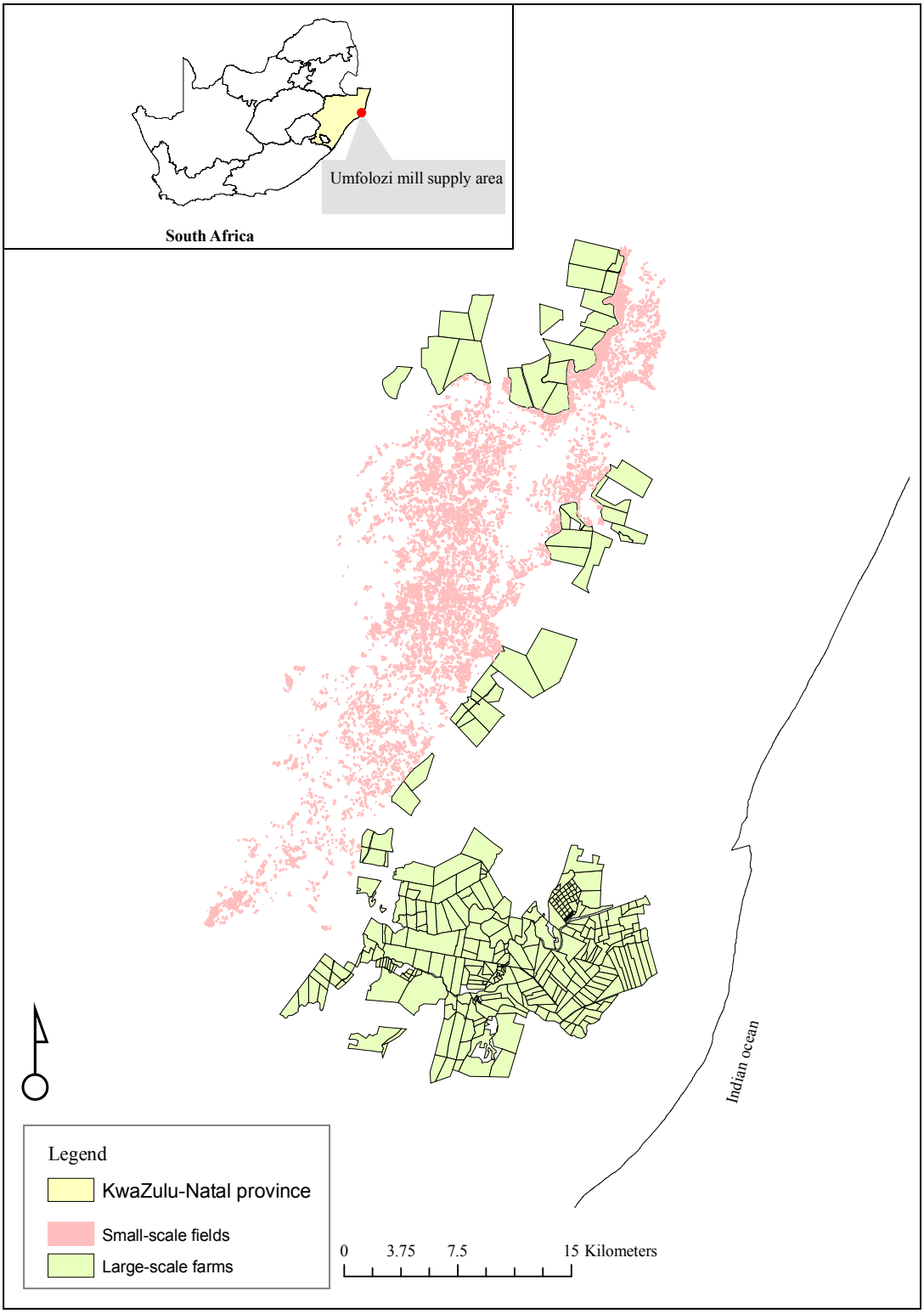

Figure 1.3. Location of study area in KwaZulu-Natal (KZN) province of South Africa.

1.7 Outline of the thesis

This thesis is organised as a set of research papers addressing each of the objectives listed in Section 1.5. The papers are either published in or submitted to peer-reviewed journals, published in peer-reviewed proceedings, or under preparation for submission to international journals. The thesis consists of 9 chapters. Chapter 2 contains a detailed literature review to the relevant application of remote sensing in sugarcane production. Specific relevance to the objectives of this study is given in Section 2.5 (monitoring sugarcane nutritional status, health, and condition) and Section 2.6 (sugarcane yield prediction).

Chapter 3 contains an examination of the potential use of hyperspectral data in detecting sugarcane leaf N content under laboratory conditions. Spectral vegetation indices are generated to assess their utility in predicting N content using simple linear relationships.

Chapter 4 contains the findings of an investigation into the potential use of *in situ* hyperspectral data in estimating sugarcane leaf N content. In this chapter the work presented in Chapter 3 is extended to assess the use of spectroscopic data in estimating sugarcane leaf N content under natural field conditions.

The usefulness of imaging spectroscopy in modelling sugarcane leaf N content is explored in Chapter 5. A random forest regression algorithm was employed to reduce the redundancy in hyperspectral data and to predict leaf N with a small subset of spectral variables derived from Hyperion data.

Chapter 6 contains the results of evaluating the potential use of spectroscopic data in detecting sugarcane thrips damage at leaf level. This chapter determines if different thrips-damaged leaves (healthy, low damage, medium damage, and severe damage) reflect significantly differently in the spectral region 400–2500 nm. The chapter also contains the results of an examination of whether certain wavebands have more discriminatory power in distinguishing between thrips-damaged classes than others.

Chapter 7 contains an assessment of using spectroscopic data to predict sugarcane thrips incidence (i.e counts) at leaf level. An evaluation is made of the use of the random forest algorithm as a feature selection method and the ability of partial least squares (PLS) regression to predict thrips numbers from subsets of selected spectral wavelengths.

Chapter 8 provides an evaluation of the utility of different vegetation indices derived from multispectral data sets (Landsat TM and ETM+) for predicting sugarcane yield. The random forest regression algorithm is implemented to test whether multi-date vegetation indices can predict sugarcane yield measured in fresh mass (t ha^{-1}).

A synthesis of the study is given in Chapter 9. The findings are summarised and conclusions are derived from the preceding chapters. Some relevant recommendations for future research are outlined. A special focus is directed towards the operational use of remote sensing techniques in sugarcane production.

A single reference list is provided at the end of the thesis.

CHAPTER TWO

Literature review

This chapter is based on

Abdel-Rahman, E. M. and Ahmed, F. B., 2008. The application of remote sensing techniques to sugarcane (*Saccharum* spp. hybrid) production: a review of the literature. *International Journal of Remote Sensing*, 29, 3753–3767.

Abstract

Remote sensing techniques provide timely, up-to-date and relatively accurate information for the management of sugarcane crops. This article reviews the literature on the application of remote sensing to sugarcane agriculture and highlights the challenges and opportunities pertinent to the success of this application. The aim of the review was to provide accurate and fundamental information relating the spectral properties of sugarcane to its agronomic, health, and nutritional status characteristics that would be of importance to sugarcane farmers and farm managers. The applications of the remote sensing techniques in sugarcane agriculture have been undertaken with particular emphasis on sugarcane classification, areal extent mapping, thermal age group identification, varietal discrimination, crop health, nutritional status monitoring, and yield prediction. It can be concluded that by selecting appropriate spatial and spectral resolution as well as suitable processing techniques for extracting sugarcane spectral information, remotely-sensed data should find use with satisfactory results in sugarcane agriculture in all areas of application.

2.1 Introduction

Sugarcane (*Saccharum* spp. hybrid) is a tall-growing perennial plant that is cultivated in the tropical and subtropical regions of the world, between latitudes 37°N and 31°S, primarily for its ability to store high concentrations of sucrose in the internodes of the stem (Grof and Campbell, 2001). Sugarcane has a sucrose content of 10–18% and a fibre content of 10–15% at harvest. The stems or stalks develop from buds and are ready for harvesting 10–24 month later. It is essentially a plant of the warm tropics and grown best when frequent heavy rainfall is interspersed with bright sunshine. It is very sensitive to temperature: below 15 °C growth is very slow, and growth ceases when the temperature exceeds 35 °C. The optimum temperature range for sugarcane growth is 20–30 °C. There are many factors affecting sugarcane production such as choice of sugarcane variety, climatic and soil conditions, and availability of water. The most important of these factors is water availability (Inman-Bamber and Smith, 2005).

Although the crop is grown primarily as a source of sugar, and over half of the world's sugar supply is derived from sugarcane, several by-products are produced from crushing sugarcane at sugar mills. These include bagasse, molasses, fibre cake, and cane wax (OGTR, 2004). However, some countries, such as Brazil and Reunion, sometimes use the crop basically for the production of alcohol as fuel (Xavier *et al*., 2006). Sugarcane is an important component of the economy in many countries in the tropics and subtropics. Many of these are developing countries, where sugar contributes substantially to the economy as well as to the calorie component of the diet (Graham *et al*., 2005). The crop provides employment not only to agricultural workers in the field but also to industrial workers in the sugar factories. Sugarcane is grown in 107 countries, and the largest producers are Brazil, India, and China, accounting for more than 50% of world production. Most sugar-producing countries have enacted legislation to protect their domestic industries, including the USA (Alvarez and Polopolus, 2002).

This article reviews the research results concerning the application of remote sensing techniques in sugarcane agriculture. The present study discusses how sugarcane canopies

interact with sunlight, and reviews the literature on sugarcane classification and mapping. The use of remote sensing in the identification of sugarcane varieties and the monitoring of sugarcane health, condition, and nutritional status are discussed. The prediction of sugarcane yield is reviewed. The study also highlights the challenges and opportunities pertinent to the successful application of remote sensing in sugarcane production.

2.2 Light interaction with sugarcane canopies

When light interacts with sugarcane canopies, as with any land surface, it is partly reflected, absorbed, and/or transmitted. The reflectance spectra of crop canopies are a combination of the reflectance spectra of the plants and the underlying soil (Guyot, 1990). In general, the spectral response of the sugarcane plant depends on five factors: canopy architecture, foliar chemistry, agronomic parameters, the geometry of data acquisition, and atmospheric conditions. Despite the absence of articles dealing specifically with sugarcane spectral properties at the orbital level (Fortes and Demattê, 2006), some authors, in their work on other aspects, have discussed how light interacts with the sugarcane canopies (Lee-Lovick and Kirchner, 1991; Gers, 2003a; Apan *et al*., 2004b; Galvão *et al*., 2005, 2006; Almeida *et al*., 2006; Fortes and Demattê, 2006). Others have related spectral information to some sugarcane agronomic variables using handheld instruments (Simões *et al*., 2005; Singels *et al*., 2005; Smit and Singels, 2006; Tejera *at al*., 2007).

The most important factor affecting the optical properties of the sugarcane canopy is its geometrical structure. The relationship between sunlight reflectance and sugarcane canopy architecture has been reported in the literature. Differences in the canopy architecture produce a higher reflectance for planophile (medium erect foliage) than erectophile (erect foliage) sugarcane plants (Galvão *et al*., 2005, 2006; Fortes and Demattê, 2006). Tejera *et al*. (2007) reported recently that the amount of light intensity penetrating the sugarcane canopy and interacting with the lower leaves is related to the canopy architecture.

The spectral response of sugarcane is also influenced by the pigments in the leaves, such as chlorophyll *a* and *b*, carotene, xanthophyll, and anthocyanins (Guyot, 1990; de Boer 1993; van der Meer *et al*., 2003). Other foliar nutrients can affect sugarcane spectral behaviour either directly by absorption (Kumar *et al*., 2003; Mutanga *et al*., 2004; Galvão *et al*., 2005) or indirectly through their influence on plant physiological processes that affect the vegetative development (Muchow *et al*., 1996; Luther and Carroll, 1999; Zhang *et al*., 2000; Isa *et al*., 2006) and interfere with sugarcane spectral response (Singels *et al*., 2005; Tejera *at al*., 2007). Nitrogen deficiency, for example, changes the whole reflectance spectrum (Guyot, 1990; Zhao *et al*., 2005). The water content of the leaves also influences the sugarcane spectral response producing absorption bands at specific wavelengths (e.g., 980 nm and 1205 nm).

Sugarcane spectral behaviour is also affected by different agronomic parameters. The most important of these factors is the leaf area index (LAI). Canopies with a high LAI reflect much more than a canopy with a medium or low LAI (Simões *et al*., 2005; Fortes and Demattê, 2006). However, a higher LAI canopy reduces the light radiation reaching the mature leaves down the stalk (Tejera *at al*., 2007).

As there has been no specific research on how sugarcane canopies interact with light, detailed studies on these aspects are needed for a better understanding of sugarcane spectral response. The results of such studies could help researchers to develop accurate models describing, for example, the separability of sugarcane varieties, estimations of foliar nutrients in sugarcane farms, and sugarcane biophysical characteristics. Plant physiologists and breeders could use these models to identify parameters and develop cultivars that use light efficiently to maximise sugarcane yield.

2.3 Sugarcane classification and areal extent mapping

Identification of sugarcane areas and accurate forecasts of the crop acreage are needed for crop yield estimation and other management purposes. Lee-Lovick and Kirchner (1991) studied the spectral signature of sugarcane in Bundaberg, Australia, by implementing

data from the Landsat Thematic Mapper (TM) sensor and their results demonstrated that data for bands 1, 2, and 3 (visible blue, green, and red, respectively) consistently fell into a narrow reflectance range and would therefore be more useful in identifying sugarcane as a crop rather than assessing crop condition. Tardin *et al*. (1992) in the Furnas region in Brazil were able to discriminate sugarcane crop from coffee and citrus plants using Landsat TM imagery with a high level of classification accuracy (95%).

Johnson and Kinsey-Henderson (1997) assessed the utility of satellite imagery from SPOT (*Système Pour I'Observation de la Terre*) and ERS-1 (European Remote Sensing Satellite 1) radar for accurately detecting land use in the Australian sugar industry and found that SPOT data were able to detect land use in cane-growing regions. They concluded that the SPOT and satellite radar data collected simultaneously are an appropriate data combination for producing efficient and accurate land use assessment in the lower Herbert catchment, Australia, and this conclusion is applicable to other sugarcane-growing regions. Narciso and Schmidt (1999) tested Landsat TM for identification, classification, and estimation of sugarcane areas in the Eston district, South Africa. The results demonstrated the feasibility of using Landsat TM satellite imagery for identifying and classifying sugarcane. Estimates of sugarcane areas derived from the classification of TM data were within 5% of the total area recorded by growers on 46 farms. Nevertheless, the results of the research conducted in South Africa by Gers (2004) revealed that remote sensing applications would not replace the existing methods and procedures for sugarcane area mapping in the foreseeable future. However, in conjunction with existing methodologies and information sources, significant benefits can be realised: in particular, the use of remote sensing to facilitate the identification of small-scale fields' expansion by more detailed mapping and the identification of the sugarcane inventory throughout the milling season. Hadsarang and Sukmuang (2000) in Thailand visually interpreted Landsat TM imagery to map and estimate sugarcane areas. They were able to separate sugarcane from other crops, and delineated the areas covered by cane using false colour composite in the district level at 1: 50 000 map scale.

Gers and Schmidt (2001) used SPOT 4 satellite imagery to monitor sugarcane-harvested areas grown on small-scale grower sector at Umfolozi, South Africa. Their findings clearly indicated that the supervised classification algorithm for the imagery distinguished between standing sugarcane and harvested plots. In India, Rao *et al*. (2002) estimated sugarcane acreage using Indian Remote Sensing Satellites (IRS) and Landsat TM data sets. They performed supervised classification and found that sugarcane, 100 days after planting, was separable from other classes. Markley *et al*. (2003) adopted SPOT 5 and Landsat Enhanced TM Plus (ETM+) imagery to calculate harvested sugarcane areas in Australia. Their findings demonstrated that by using the change detection method, the harvested area from satellite imagery was within 2% of the area calculated by other means such as harvester global positioning system (GPS) tracking and manual area calculation. Bégué *et al*. (2004) reported that sugarcane fields that were harvested and ploughed could easily be mapped with accuracy of >90% in Guadeloupe. Masoood and Javed (2004) concluded that sugarcane area can be predicted with a reasonable level of accuracy a couple of months before harvest by using an area forecasting model developed in Pakistan.

Tulip and Wilkins (2005) attempted further work on the spectral separation of sugarcane and trash as well as attempted to determine whether spectra from mixtures of billet sugarcane, leaf, and tops could be used to examine surface coverage proportions. Their research used two different instruments to collect spectra: a PP system UniSpec® portable spectrometer and a field portable analytical spectral devices® (ASD, 2005) spectrometer. The spectral unmixing technique was performed, and the results showed that billet sugarcane, leaf, and tops were spectrally separable for the sugarcane variety they investigated. In addition, with known end member spectra for billet sugarcane, leaf, and tops and sufficient sample areas, accurate coverage estimates could be made.

Xavier *et al*. (2006) classified sugarcane crops in Brazil using a coarse-resolution Moderate Resolution Imaging SpectroRadiometer (MODIS) sensor by developing a multi-temporal Enhanced Vegetation Index (EVI). The results of the unsupervised

classification showed that sugarcane can be distinguished from soybean, peanuts, water bodies, urban areas, and natural and planted forests. However, it was difficult to differentiate sugarcane from pasture, which has similar temporal–spectral behaviour to sugarcane. It was concluded that the confusion of pasture with sugarcane could be minimised using images from higher spatial resolution sensors integrated with the sugarcane classification procedure. Ancillary data, such as local official crop statistics and spatial distribution of sugar and/or alcohol plants, may help to improve sugarcane classification. Based on their study results, a supervised classification for the sugarcane areas should be pursued.

The thermal age group (also called growing degree-days) is the phenological stage of the crop. It is called thermal time (heat accumulation over calendar time) because a certain phenological stage is highly correlated with thermal time or thermal time window (phyllochron) when using a base temperature (Tb). Schmidt *et al*. (2000) were able to identify sugarcane thermal age groups by using digital multispectral video (DMSV) imagery. Gers (2003b, 2004) concluded that the sugarcane thermal age groups could be easily distinguished using Landsat ETM+ sensor reflectances (Figure 2.1). However, some concerns were raised regarding sugarcane thermal age group identification that should be taken into consideration for future research such as accurate digital field boundary, soil water depletion level, rainfall and irrigation water applied, climatic data, fertilizer information, and image atmospheric correction (Gers, 2004).

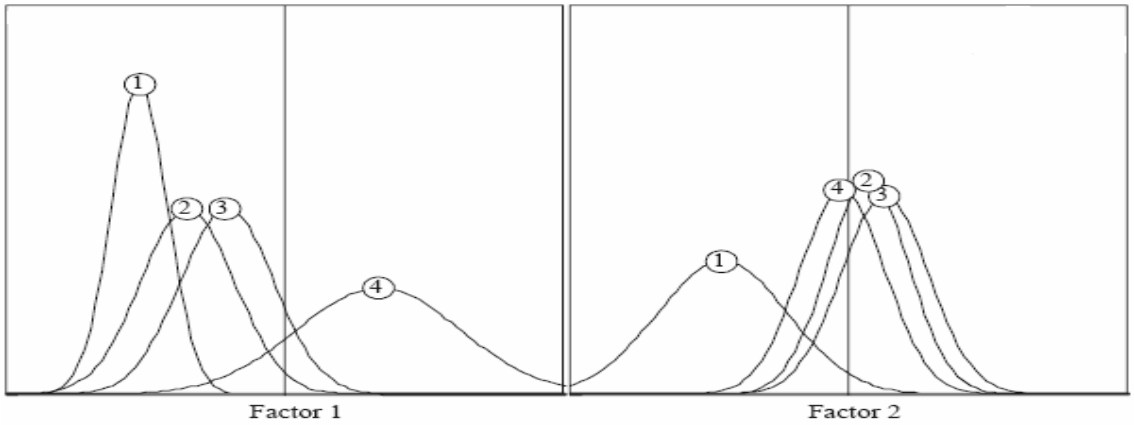

Figure 2.1. At-Landsat ETM+ reflectance of four sugarcane thermal age groups against first and second variance–covariance principal component factors. Numbers in circles indicate the thermal age group for each factor (Source: Gers, 2003b).

In Australia, Everingham *et al*. (2007a) employed three classification procedures *viz*., discriminate analyses, random forest and support vector machine to distinguish among nine sugarcane cycles (i.e., the number of times that the plant has grown back after harvest) using Hyperion data. They found that support vector machine and random forest methods outperformed the traditional discriminate analyses approaches. The highest overall accuracies based on pixel classification were 62.3% using penalised linear discriminant analysis, 80.4% using random forest, and 83.9% using support vector machine. Whereas the highest overall accuracies based on the per field classification strategy were 77.1% employing penalised linear discriminant analysis, 97.6% employing random forest and 100% employing support vector machine.

With the exception of that of Everingham *et al*. (2007a), the above-mentioned studies either implemented conventional techniques such as visual interpretation to identify sugarcane crop areas or used coarse-resolution sensors (e.g., MODIS) to discriminate sugarcane from other plantations, and that resulted in separating sugarcane from only easily separable classes (e.g., forests, water bodies). In addition, the use of an unsupervised classifier for most of the previously mentioned studies led to difficulties in discriminating sugarcane crops from some land cover classes (e.g., pasture). Although the results of these studies produced reasonable results for sugarcane area estimation and thermal age group identification, more work is needed to improve the results.

2.4 Varietal identification

The mapping of sugarcane varieties is very important for yield prediction and assessment of risk of crop damage. The use of remote sensing in identification of sugarcane varieties has been applied using multispectral remotely-sensed data in South Africa (Schmidt *et al*., 2000, 2001; Gers, 2004) and Brazil (Fortes and Demattê, 2006; Galvão *et al*., 2006) and hyperspectral remotely-sensed data in Australia (Apan *et al*., 2004a), Brazil (Galvão *et al*., 2005, 2006), and the USA (Johnson *et al*., 2005).

Schmidt *et al*. (2000, 2001) noted the use of DMSV was able to distinguish between different sugarcane varieties in South Africa. However, Gers (2003a, 2004) showed that by analysing at-sensor reflectances from Landsat ETM+ across all investigated thermal age groups no significant differences were found among the top five sugarcane varieties in the Umfolozi mill supply area. Fortes and Demattê (2006) used Landsat ETM+ data to discriminate among four sugarcane varieties (Figure 2.2) by analysing individual spectral bands and spectral vegetation indices. They developed discriminant equations derived from the multiple relationships between the spectral bands and the vegetation indices and the specific characteristics of the varieties. The results showed that sugarcane varieties could be identified with 93.6% certainty. However, Galvão *et al*. (2006) were not able to differentiate among sugarcane varieties using NOAA/AVHRR (National Oceanic and Atmospheric Admission/Advanced Very High Resolution Radiometer), SPOT-5 and CCD/CBERS-2 (Charge-Coupled Detector/China–Brazil Earth Resources Satellite), Landsat ETM+ and TERRA/ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) sensors that were simulated from Hyperion data using their filter functions.

By using EO-1 Hyperion data, Apan *et al*. (2004a) attempted to evaluate the discrimination of five important Australian sugarcane varieties by conversion of the sensor radiance into surface reflectance values and by the calculation of some ratios of reflectance and several vegetation indices potentially sensitive to chlorophyll content, leaf water, and lignin-cellulose. Using multiple discriminant analysis (MDA), it was possible to distinguish between the sugarcane varieties with classification accuracies for the sample pixels of 97% and 74% for five and eight varieties, respectively. However, for the classification of the entire image the accuracy was reduced to only 46%, and that was attributed to the high number of classes used and the many confounding factors pertaining to crop management regime, growth stage, and background features. Galvão *et al*. (2005), by using similar techniques, were able to distinguish among five Brazilian sugarcane varieties.

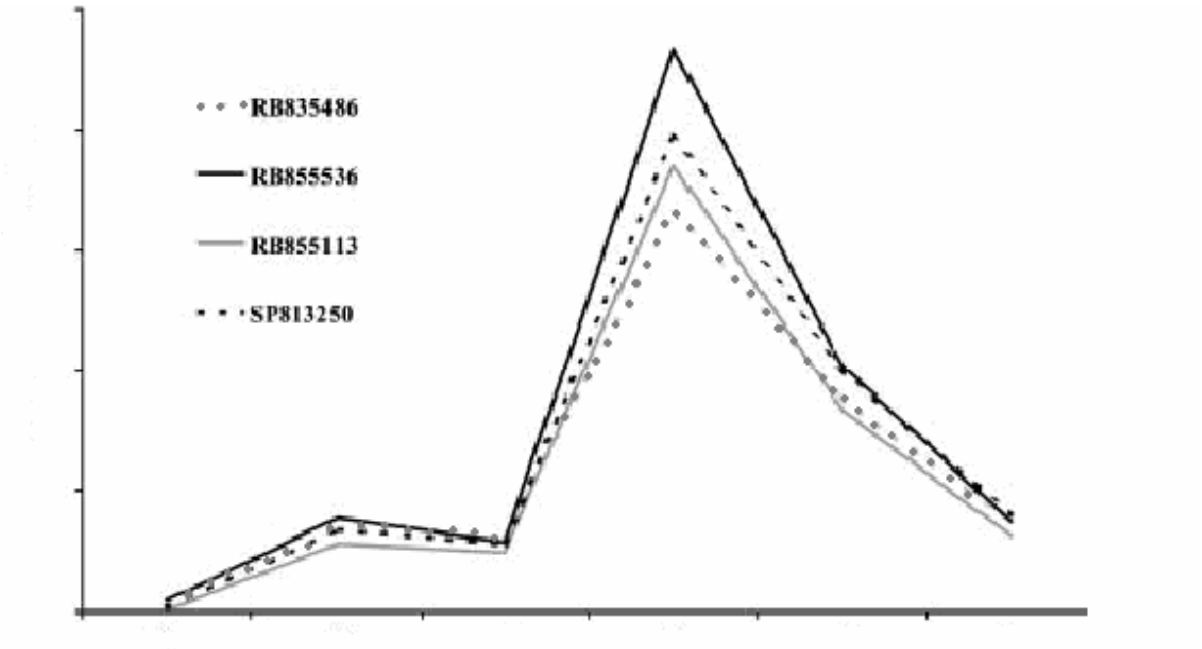

Figure 2.2. Spectral curves of four sugarcane varieties obtained by Landsat 7 ETM+ (Source: Fortes and Demattê, 2006).

In the USA, Johnson *et al*. (2005) were able to identify and distinguish seven sugarcane varieties: five commercial cultivars, one noble cane, and one wild cane. Their research was conducted by collecting hyperspectral reflectance data using a spectrometer under natural light conditions and by analysing plant pigments and performing conical discrimination and discriminant analysis.

Recently, Everingham *et al*. (2007a) evaluated the use of Hyperion data to discriminate among nine sugarcane varieties in Australia. Their research compared different classification techniques (i.e., discriminant analyses, random forest, and support vector machines) to differentiate among the varieties. The relatively new approaches of random forest and support vector machine performed better compared to discriminant analyses methods. The overall accuracy of the classification employing the pixel classification method was 79.7%, 87.5%, and 90.0% when penalised linear discriminant analysis, random forest, and support vector machine were used, respectively. On the other hand, the overall accuracy was enhanced (100% using random forest, 98.8% using support vector machine, and 85.7% using penalised linear discriminant analysis) when the per field classification approach was applied.

The use of fine spatial resolution sensors (e.g., IKONOS, QuickBird platform) for discriminating sugarcane varieties has not yet been tested. By using hyperspectral remote sensing the sugarcane varieties could possibly be discriminated at a high level of accuracy. However, the images from hyperspectral sensors are expensive to obtain and do not cover vast areas. The fusion of fine-resolution sensor data with hyperspectral imagery could result in a new and useful product by taking advantage of the relatively large coverage and high spatial resolution of the fine sensors and the very high spectral resolution of hyperspectral sensors, providing more accurate discrimination models for sugarcane varieties.

2.5 Monitoring sugarcane nutritional status, health, and condition

2.5.1 Detection of nutrient and water deficiencies

Nutrient and water deficiencies in sugarcane may result in reduced cane tonnage particularly in ratoon crops. Nitrogen (N) and potassium (K) concentrations in sugarcane leaves were detected by Jackson *et al*. (1981) in the USA using a radiometer mounted on a four metre aluminium pole. Their research indicated that the infrared/red ratios measured over a plot with adequate nitrogen and potassium were significantly higher than those measured over nitrogen- and potassium-deficient plots. It was concluded that nitrogen deficiency can be readily detected, and could have been detectable earlier when visual symptoms of deficiency were less evident, whereas K deficiency was more difficult to detect using this technique. Schmidt *et al*. (2000, 2001) reviewed the use of a DMSV sensor for identifying changes in sugarcane nutrition following different application rates of fertilizers to the test plots but were not able to detect nutrient deficiencies in sugarcane.

Regarding water stress, Jackson *et al*. (1981) tested the ability of a radiometer to detect water-deficient plots. Their results showed that the infrared/red ratios were a useful indicator of water deficiency in sugarcane. Schmidt *et al*. (2000) reported that the DMSV sensor had the potential to identify, from the sugarcane canopy, different levels of

moisture stress at an early stage of development. An investigation by Lee-Lovick and Kirchner (1991) in Australia concluded that the sugarcane spectral signatures from Landsat TM bands 4, 5, and 7 (near infrared and shortwave infrared bands) were much wider and offered more scope for indicating sugarcane condition; that is reflectance levels were low, indicating a high water content and therefore a relatively healthy, vigorously growing crop. Bands 5 and 7 were highly correlated: therefore, crop condition correlations should be with bands 4, 5, and 7 individually, or 4 with 5, or 4 with 7.

Wiegand *et al*. (1996) tried to detect and map salt stress in sugarcane crops using SPOT HRV (High Resolution Visible) spectral observations together with ground truth and image unsupervised procedures. They concluded that the method could be used to quantify and map variations in weighted electrical conductivity of the root zone and yield.

Yang *et al*. (1997) investigated the estimation of evapotranspiration using Landsat TM data for sugarcane fields based on the concept of vegetation index/temperature trapezoid (VITT). They showed that the Landsat TM data, along with the VITT concept, can provide a practical means for estimating the evapotranspiration of a sugarcane field at a local scale.

Sugarcane nutrient deficiencies are, with fair results, detectable in plants under trial conditions using handheld radiometers. The shortcoming of the above-mentioned studies is that they did not attempt earlier to detect nutrient deficiencies at the field level to monitor the nutritional status of the sugarcane crop so that the nutrient-deficient areas could be corrected by applying the correct nutrients and suitable doses. Similar work has been undertaken for water stress at research plots. However, some attempts have been made on a large scale for estimating evapotranspiration. The sugarcane producers need to know how much water is used by the sugarcane plantation to meet their stakeholders' legislation and rules; therefore, investigators should use simulation models of water estimates that incorporate remotely-sensed data to provide the sugarcane producer with accurate water use estimation models and algorithms.

2.5.2 Disease detection

Disease management is an important issue in maintaining the competitive advantages in the sugar industry. Apan *et al*. (2004b) addressed the problem of orange rust (*Puccinia kuehnii*) disease in Australia and evaluated several narrow-band vegetation indices from EO-1 Hyperion imagery. Forty spectral vegetation indices were calculated focusing on spectral bands related to leaf pigments, leaf internal structure, and leaf water content. Discriminant function analysis was then performed to select an optimum set of indices based on their correlation with the discriminant function. The results demonstrated that Hyperion data can be used to detect orange rust disease in sugarcane crops. The results also verified that areas with the sugarcane orange rust disease showed differences in spectral reflectance signatures and could be discriminated from non-diseased areas, at certain wavelengths.

Although sugarcane is prone to many diseases and pests, only the investigation by Apan *et al*. (2004b) has been undertaken to detect and delineate rust-infested areas. The research into detection of orange rust disease by using hyperspectral remote sensing achieved promising results. However, more work is needed for the detection of sugarcane pests and diseases. Some insects cause serious problems for sugarcane production, but the visual symptoms on sugarcane plants are not very clear. Investigators need to look at any physiological changes in sugarcane due to insect infestation that can be monitored and detected by remote sensing, using the capability of hyperspectral sensors to produce indices and models that could be used for delineating disease-infested areas.

2.6 Sugarcane yield prediction

Remotely-sensed data and techniques have a large potential for use in the prediction of sugarcane yield [tonnes of stalks per hectare (t ha^{-1})]. In Brazil, Rudorff and Batista (1990) used satellite data from Landsat MSS (Landsat 4) and an agro-meteorological model for yield prediction. Their research illustrated that sugarcane yield can be estimated by spectral data alone, agro-meteorological data alone, and also by integrating

spectral data that were transformed to a vegetation index with the agro-meteorological model. The yield estimated by the vegetation index model alone or the agro-meteorological model alone gave poor results compared to the combined (vegetation index and agro-meteorological data) model. It was suggested that techniques to estimate some crop biophysical parameters (e.g., LAI) through spectral data, as used in the agro-meteorological model, be further investigated. It was also suggested that atmospheric correction procedures and vegetation indices that are referenced to a soil line should be used for more accurate prediction models for sugarcane yield. Lee-Lovick and Kirchner (1991), using data from Landsat TM sensor for prediction of sugarcane yield, obtained no correlation between yields and reflectance for any band within a scene in Bundaberg, Australia, indicating that short-term variations in the canopy condition were independent of previous or future stalk growth. The maximum coefficient of determination (R^2) was only 0.25, and most values of R^2 were below 0.10. The Agricultural Research Council (ARC, 2000a,b) investigated the use of coarse-resolution satellite imagery from NOAA/ AVHRR and fine-resolution DMSV for prediction of sugarcane yield in the Pongola mill supply area, South Africa, and found that there was no significant correlation between either estimated or recorded yield and median normalised difference vegetation index (NDVI) derived from DMSV. However, for NDVI derived from AVHRR, the results showed significant correlation with yield at mill level but no correlation at farm level. Gers (2003a, 2004) has conducted research in the Umfolozi mill supply area by using a moderate spatial resolution sensor (Landsat ETM+). Principal component analysis on sugarcane yield and at-satellite reflectance values was performed, and the data across all tested sugarcane thermal age groups were analysed, but no significant relationship was found between yield and spectral characteristics of sugarcane. It was recommended that alternative approaches be investigated. These may include the derivation of crop parameters from remote sensing, such as LAI, which can be obtained from NDVI.

In Reunion Island, Bappel *et al*. (2005) reported that LAI can be estimated ($R^2 = 0.86$) using multi-date NDVI values derived from SPOT 4 and 5 data sets in an exponential relationship. The estimated LAI values resulted in better forecasts of sugarcane yield when used as inputs for a process-based growth model. R^2 was 0.67 when measured and

estimated sugarcane yields were compared (Root Mean Square Error of Prediction: RMSEP = 12.8 t ha^{-1}). Everingham *et al*. (2005) compared the use of a process-based crop growth model (APSIM) and a forecasting approach based on remotely-sensed data in predicting sugarcane yields in Queensland, Australia. For remote sensing models, they tested if a principal component index combined from integrated NDVI derived from multi-date AVHRR data could be related to final sugarcane yield. The highest correlation of predicting sugarcane yield was 0.91. Furthermore, they demonstrated that this approach has potential for operational yield forecasting in Australia. Overall, the remote sensing models produced more accurate yield forecast than APSIM model.

In Mozambique, using Monteith's efficiency model with coarse-resolution MODIS fraction photosynthetically active radiation (fPAR) data and temporal unmixing fPAR profiles of segments for sugarcane yield assessment, the investigation of Ha (2005) did not lead to accurate yield estimates, probably because the variation in yield came from using different sugarcane varieties. Simões *et al*. (2005), who used field spectroscopy with equivalent bands to Landsat ETM+, found that spectral variables B3 (red band), the ratio vegetation index (RVI), NDVI, and soil adjusted vegetation index (SAVI) were highly correlated (Correlation of Coefficient: $r > 0.79$) with sugarcane yield and number of stalks per metre (NPM). The best regression models were verified for RVI, LAI, and NPM, which explained 97% of yield variation and 99% of total sugarcane biomass variation. Ueno *et al*. (2005), by using Landsat TM to predict yields for every sugarcane-growing field in Japan and by using different types of vegetation indices, found low prediction accuracy because of the relatively large (one ha) field size used. However, the correlation of coefficient (r) values reached more than 0.6, which increased the predictive capacity of the model used. It was suggested that the prediction accuracy could be improved considerably by the use of high-resolution satellite imagery.

Almeida *et al*. (2006) attempted to forecast sugarcane yield in Brazil using ASTER and Landsat ETM+ sensors. The authors developed a method for enhancing the specific spectral responses of vegetation constraints, reducing the spectral dimensions with prioritisation of information and weighting of the parameters related to the foliar area.

The data processed through these steps were reduced to a single image, from which the mean digital number (DN) per cultivated sugarcane area was calculated, and the image DNs were subsequently transformed to cane yield (t ha^{-1}) through normalisation. The results showed that this method proved to be efficient in yield prediction and was more accurate than traditional methods used in the Brazilian sugarcane agro-industry. It was recommended that further research in tropical terrains such as Brazil should consider issues about how variables such as topography, sugarcane variety, soil type, and spectral and spatial resolution might affect the quality of the correlation and the forecast of sugarcane yield.

The limitation of the above-mentioned studies is that no operational model for predicting sugarcane yield has been produced, and most of the studies have been undertaken at research level. Moreover, with exception of Bappel *et al*. (2005), no study implements the combination of sugarcane growth simulation models with sugarcane biophysical parameters derived from remotely-sensed data for providing relatively accurate estimates of sugarcane yield. In addition, testing of fine-resolution sensors for estimates of sugarcane yield is needed, especially for large fields.

2.7. Overall challenges and opportunities

Although considerable progress has been made in sensor development and radiometric correction applied to agricultural remote sensing, there are still challenges to be met. First, there is no general agreement on the critical spectral regions to use for agriculture and the sensor specifications for a dedicated, orbiting agricultural sensor. Second, advances in radiometric correction have been focused primarily on on-board radiometric reference sources, such as lamps and solar illuminated diffuser panels. However, calibration of radiometers is only the first step in providing the surface reflectance required by agricultural models and algorithms (Moran *et al*., 2003).

For the potential use of remote sensing for making day-to-day decisions on sugarcane management, there is overall importance to the sugarcane growers of timely, up-to-date, and accurate data. Despite the recent development of new sensors and processing techniques, very few sugarcane farmers presently have access to regular images of their farms, and even when they do, slow turnaround of the processed product continues to be a problem.

Despite these shortcomings, there is no doubt that remote sensing technology will permeate many aspects of sugarcane farming in the future. Grower acceptance will increase as products with higher spatial, spectral, and temporal resolution become more efficient, affordable, and cost-effective. By selecting appropriate spatial and spectral resolution, and suitable processing techniques for extracting sugarcane spectral information, the remotely-sensed data should find use in sugarcane agriculture in all areas of applications with satisfactory results.

From a research perspective, however, there are several overarching and interrelated challenges that must be dealt with to advance remote sensing beyond today's largely qualitative applications for sugarcane crop management (Pinter *et al*., 2003). The first challenge deals with sugarcane nutritional status. Some research in the past has been undertaken to estimate foliar nitrogen and potassium concentrations in experimental fields using handheld spectrometers, with promising results. The use of spectral observations from polar orbiting hyperspectral satellites (e.g., Hyperion) for at-field foliar nitrogen estimation is a challenge for estimating sugarcane foliar nitrogen distribution over large areas, which could be integrated into further planning and management of sugarcane farms.

Second, the attacks on sugarcane crops by many pests and diseases that significantly reduce its yield warrant serious investigation. Detection and delineation of disease-infected (e.g., orange rust) sugarcane areas using hyperspectral remotely-sensed data have been attempted but more work is needed to investigate new insect invaders (e.g., thrips in South Africa). Such work should take advantage of handheld field

spectroradiometers and polar orbiting hyperspectral sensors (e.g., Hyperion) as well as the power of sophisticated spatial data processing and analysis software to detect and model sugarcane pest and disease outbreaks and damage. A third research challenge is that in some countries (e.g., South Africa) there are Water Acts and Legislation which levies certain land uses (e.g., commercial forestry, sugarcane) according to their potential for reducing stream flows. No research has yet been carried out on estimating water requirements for sugarcane using process-based models that use remotely-sensed data as input parameters.

The fourth challenge deals with understanding and being able to investigate the techniques that estimate some sugarcane biophysical parameters [e.g., LAI, APAR (Absorbed Photosynthetically Active Radiation)] through spectral data; and the possibility of using these parameters for sugarcane growth simulation model (e.g., CANESIM) to enhance their accuracy which should be studied further. The use of fine spatial resolution sensors (e.g., IKONOS, 2005; NASA, 2006; SIC, 2007) will benefit prediction of sugarcane yield in small fields. Moreover, researchers could use microwave remotely-sensed data (RADARSAT-2, 2006) in wet tropical regions to reduce the problems of cloud cover and to provide accurate information on the physical structure of sugarcane fields. In addition, the use of radar may complement multispectral remotely-sensed data through data fusion.

Fifth, although some work has been undertaken on the prediction of sugarcane yield using remotely-sensed data, there is still a challenge to develop accurate operational sugarcane yield-estimating models that take advantage of improvements in sensor characteristics and processing as well as analysis techniques.

Acknowledgements

The candidate is very grateful to the anonymous reviewers for their useful comments and suggestions.

CHAPTER THREE

Estimation of sugarcane leaf nitrogen concentration using a handheld spectrometer under laboratory conditions

This chapter is based on

Abdel-Rahman, E. M., Ahmed, F. B. and van den Berg, M., 2008. Imaging spectroscopy for estimating sugarcane leaf nitrogen concentration. *Proceedings of SPIE Remote Sensing for Agriculture, Ecosystems, and Hydrology X Conference*, V-1 – V-12.

Abstract

Spectroscopy can provide real-time high throughput information on growing crops. The spectroscopic data can be obtained from handheld, spaceborne, and airborne sensors. Such data have been used for assessing the nutritional status of some field crops (e.g., maize, rice, barely, wheat, and potato). In this study a handheld FieldSpec® 3 spectroradiometer in the 350–2500 nm range of the electromagnetic spectrum was evaluated for its use to estimate sugarcane leaf nitrogen (N) concentrations. Sugarcane leaf samples from one variety *viz*., N19 of two age groups (4–5 and 6–7 month) were subjected to spectral and chemical measurements. Leaf reflectance data were collected under controlled conditions, and leaf N concentrations were obtained using an automated combustion technique (Leco TruSpec® N). The spectral data were transformed to their first-order derivative and continuum-removed spectra. The potential of spectroscopic data for estimating sugarcane leaf N status was evaluated using univariate correlation analysis method with the transformed reflectance data. The variables that presented high correlation with N concentration were used to develop SR-based (Simple Ratio) and NDVI-based (Normalised Difference Vegetation Index) indices. Simple linear regression was then used to select a model that yielded the highest R^2. These were the SR (744, 2142) index for the 4–5 month old sugarcane crop and the NDVI (2200, 2025) index for the 6–7 month old cane crop, with R^2 of 0.74 (Root Mean Square Error of Prediction: RMSEP = 0.120; 7.90% of the mean) and 0.87 (RMSEP = 0.100; 5.87% of the mean), respectively.

Keywords: spectroscopy; hyperspectral; sugarcane; nitrogen estimation

3.1 Introduction

Nitrogen (N) is an essential element for crop growth, development, and yield and is often a limiting nutrient in agricultural production. Farm managers aim to identify crop N status and apply the appropriate amount of fertilizers for optimal yield and N use efficiency (Marschner, 1993; Richards, 2000; Osborne *et al*., 2002; Zhao *et al*., 2003). Substantial amounts of N fertilizer are necessary for commercial sugarcane production because of the large biomass produced by sugarcane crops (Wiedenfeld, 1995; Muchow *et al*., 1996; Thorburn *et al*., 2005), but excessive amounts negatively affect cane quality and increase susceptibility to pests (Atkinson and Nuss, 1989; Marschner, 1993). Therefore, efficient estimation of plant N status and appropriate N fertilizer management are essential for sugarcane production. The South African Sugarcane Research Institute (SASRI) recommends fertilizer rates ranging between 60 and 120 kg N/ha, depending primarily on soil type and sugarcane variety. SASRI advices the growers to take leaf samples for N analysis to get information about the nutritional status of their crops and to correct any kind of N deficiency by fertilization (SASRI, 2003). However, the methods used for estimating leaf N content are laborious, pricy, and time-consuming. A rapid and efficient method for estimating sugarcane N concentration is therefore necessary.

The use of spectroscopy techniques to estimate the nutrient status of growing crops could save time, and reduce the cost associated with sampling and analysis. Spectroscopy (also known as “spectrometry” and “hyperspectral remote sensing”) is a technology that entails acquiring data in narrow (<10 nm) and contiguous spectral bands throughout the ultraviolet, visible, and infrared portions of the spectrum (Bakker, 2001; Lillesand and Kiefer, 2001; Baltsavias, 2002; Hansen and Schjoerring, 2003; Jiang *et al*., 2004a; Kerekes and Schott, 2007). These narrow spectral bands enable the detection of small spectral features that might otherwise be masked within the broader bands of multispectral scanner systems. Researchers have evaluated the use of spectroscopy techniques for estimating N status of some crops such as corn (e.g., Strachan *et al*., 2002), wheat (e.g., Oppelt and Mauser, 2004), sorghum (e.g., Zhao *et al*., 2005), rice (e.g., Nguyen and Lee, 2006; Nguyen *et al*., 2006), potato (e.g., Jain *et al*., 2007), and

barely (e.g., Jørgensen *et al*., 2007) by determining the appropriate wavelength or combination of wavelengths that characterise N deficiency. Their work resulted in satisfactory results of leaf or canopy N estimates. In a study with sugarcane (Jackson *et al*., 1981) under trial conditions, it was found that there was a detectable N deficiency where the infrared/red ratios measured over sugarcane plots with adequate N were significantly higher than those measured over N-deficient plots. A recent literature review (Abdel-Rahman and Ahmed, 2008) indicated that no work has been done to estimate sugarcane N at farmers' fields using spectroscopic data from handheld, airborne, or space borne sensors. The aim of the present study was to evaluate the use of spectroscopic data to predict leaf N content in sugarcane.

3.2 Material and methods

3.2.1 Leaf sample collection

In this study, twenty sugarcane fields - from 17 farms - were selected for leaf sampling (from large-scale and small-scale grower sectors). All fields selected were grown with sugarcane variety N19 of 4–5 and 6–7 month age crops at the moment of sampling (10 fields representing each age group), during the first week of December 2007 (i.e., late spring). Variety N19 was selected because it is the most common variety in the study area, representing about 39% of the total crop (South African Sugar Association: SASA, 2007, unpublished industry database). Samples were taken from two sites in each field (total number of samples = 40, representing 20 samples for each age group). The crops sampled presented no water stress and no visible disease or pest damage symptoms. The third fully expanded leaf from the top was excised for spectral (5 leaves per sample) and chemical (20 leaves per sample) analyses. Midribs were removed and discarded as suggested by SASRI (2003) and Muchovej *et al.* (2005). Hence, only leaf blades were used for spectral and chemical analyses. The leaves collected for chemical analysis were dried, whereas the leaves collected for spectral measurements were placed in plastic bags and kept cool until they were taken to the laboratory, where they were stored at -18^{o} C

for about a week before analysis. One sample was excluded from the analysis because it was dried mistakenly before the spectral measurements, leaving 20 samples of the 4–5 month age group and 19 samples of the 6–7 month age group for further analysis.

3.2.2 Leaf spectral measurements

Before the spectral measurements were made, the leaf samples collected were exposed to indoor laboratory conditions for two hours to allow them to be free from any water droplets. A FieldSpec® 3 spectroradiometer (Analytical Spectral Device: ASD, 2005) was used to collect relative reflectance data from each leaf sample. The ASD is a compact, non-imaging, battery powered, portable spectrometer, which uses a fibre optic cable for light collection and a notebook computer for data logging. The spectral range is 350–2500 nm with a resolution of 1.4 nm in the 350–1000 nm range and 2 nm in the 1000–2500 nm range and a data collection time of 0.1 second per spectrum. The spectral measurements were carried out under room temperature (23–25 ^{o}C). The fibre optic cable with 1º field of view was pointed 0.1 m above the sample which was laid on a black rubber platform with about zero reflectance to account for any background reflectance (Qifa and Jihua, 2003; Chen *et al*., 2007). A 50 W halogen lamp mounted on a triangular tripod and positioned at the nadir position 0.4 m and 45° from the sample surface was used as the only light source. All spectral measurements were made relative to a spectralon white reference panel, which was measured before and after the spectral measurements of sugarcane leaves. For each measurement, 20 scans were performed by the ASD, and the spectral data were averaged over the measurement. The bottom, middle, and top upper surface of each leaf was measured so that the spectrum of a sample (10 blades) was an average of 30 spectra.

3.2.3 Chemical analysis

The leaf samples were oven-dried at 75 °C for 24 hours, then ground using a leaf grinder, and oven-dried again at 75 °C for 24 hours to ensure that water content was about zero.

Leaf powder, 0.11–0.25 gram per a sample, was then taken and put in a tin foil to be placed into the sample carousel of TruSpec N® instrument (Leco, 2006) which uses a thermal combustion technique to measure leaf N concentration (%).

3.2.4 Spectral transformation

3.2.4.1 First-order derivative spectra

The primary sugarcane leaf spectra were transformed using the first-order derivative transformation technique described in ASD (2005). This transformation technique is widely used by researchers (Mauser and Bach, 1994; de Jong, 1998; Sims and Gamon, 2002; Huang *et al*., 2004; Thomas *et al*., 2007) to reduce the effect of sample geometry and roughness on the spectral reflectance measurements and to locate the positions of absorption features and inflection points on the spectra. Moreover, derivatives transformation may swing with greater amplitude than the primary spectra. The clearly distinguishable derivatives are especially useful for separating out peaks of overlapping wavelengths (de Jong, 1998; ASD, 2005).

3.2.4.2 Continuum-removed spectra

Continuum removal normalises the primary reflectance spectra to allow comparison of individual absorption features from a common baseline (Kokaly, 2001). The continuum is a convex hull fitted over the top of a spectrum to connect local spectral maxima (Schmidt and Skidmore, 2003; Mutanga and Kumar, 2007). Two local start and end points on specific absorption features *viz*., $R_{460–530}$ and $R_{550–750}$ were defined. These absorption features were used by Mutanga *et al*. (2003, 2004) as chlorophyll absorption features related to foliar N concentration. The continuum-removed reflectance is obtained using the formula described by Mutanga *et al*. (2004):

$$R'_{(\lambda i)} = R_{(\lambda i)} / R_{c(\lambda i)} \quad (3.1)$$

Where:

$R'_{(\lambda i)}$ = Continuum-removed reflectance

$R_{(\lambda i)}$ = the reflectance value for each wavelength i in the absorption feature

$R_{c(\lambda i)}$ = the reflectance level of the continuum line (convex hull) at the corresponding wavelength i.

This formula was performed using ENVI software (Environment for Visualising Images: ENVI, 2006).

3.2.5 Statistical analysis

One-way analysis of variance (ANOVA) was used to test whether the two age groups could be discriminated by their spectral features, thereby permitting their separation for analysis. The first-order derivative of reflectance values in the range between 400 and 2500 nm and the continuum-removed reflectance at two ranges of the visible portion of the spectrum *viz.*, $R_{460-530}$ and $R_{550-750}$ were used for this task.

Statistical analyses were conducted using the SPSS package (SPSS, 2006). Univariate correlations were calculated on first-order derivatives of reflectance as well as continuum-removed reflectance and used to evaluate linear relationships between sugarcane leaf N concentration and reflectance values. The wavelengths at which strong positive or negative relationships ($r \geq 0.6$ or $r \leq -0.6$) were found were used to generate 2-wavelengths vegetation indices, *viz*., SR-based and NDVI-based indices at each of the other wavelengths from 400 to 2500 nm for first-order derivatives of the leaf reflectance, and from 460 to 530 nm and 550 to 750 nm for the continuum-removed reflectance (Equations 3.2 and 3.3) as suggested by Zhao *et al*. (2003, 2005), hence:

$$\text{SR-based} = R_{\lambda 1i} / R_{\lambda 2i} \quad (3.2)$$

$$\text{NDVI-based} = (R_{\lambda 1i} - R_{\lambda 2i}) / (R_{\lambda 1i} + R_{\lambda 2i}) \quad (3.3)$$

Where:

$R_{\lambda 1i}$ = the first-order derivative of the reflectance or continuum-removed reflectance at the i^{th} wavelength at which there was strong correlation ($r \geq 0.6$ or $r \leq -0.60$) between the first-order derivative of the reflectance or continuum-removed reflectance and N concentration

$R_{\lambda 2i}$ = the first-order derivative of the reflectance at the other i^{th} wavelength between 400 to 2500 nm or continuum-removed reflectance at 460 to 530 and 550 to 750 nm.

Univariate simple linear regression analysis was carried out on the vegetation indices developed in this study to explore any possible relationships between these vegetation indices and sugarcane leaf N content. The coefficients of determination (R^2) values were ranked and models that produced the highest three values were recorded for each spectral transformation method.

A preliminary validation of the predictive models was conducted using a leave-one-out cross validation method (Efron, 1979), whereby one-by-one an observation was left out, the regression model re-calibrated and then used to predict the N content of the excluded leaf sample. R^2 values as well as RMSEP when measured and estimated N values were compared in one-to-one relationship were recorded to test the predictive ability of the developed models.

3.3 Results

3.3.1 Foliar nitrogen concentration

The foliar concentration of N varied between the two age groups. Table 3.1 shows the descriptive statistics of the N data set used in the analysis.

Table 3.1: Descriptive statistics of the N concentration (%) for the leaf samples measured in the laboratory

Age group	Number of samples	Mean	Confidences (CL) of mean (95%)	Minimum	Maximum
4–5 month	20	1.52	0.41	1.24	2.05
6–7 month	19	1.70	0.11	1.28	2.02

3.3.2 Discriminating age groups using spectral signature

The mean relative reflectance spectrum of the sugarcane leaves is shown in Figure 3.1, describing a typical vegetation spectrum (Kumar *et al*., 2003). The results of one-way ANOVA test (Figures 3.2 and 3.3) showed that the two age groups exhibited significantly different spectral responses (*p* values were less than 0.0001 at most wavelengths). Consequently, the subsequent analyses were done for each age group separately.

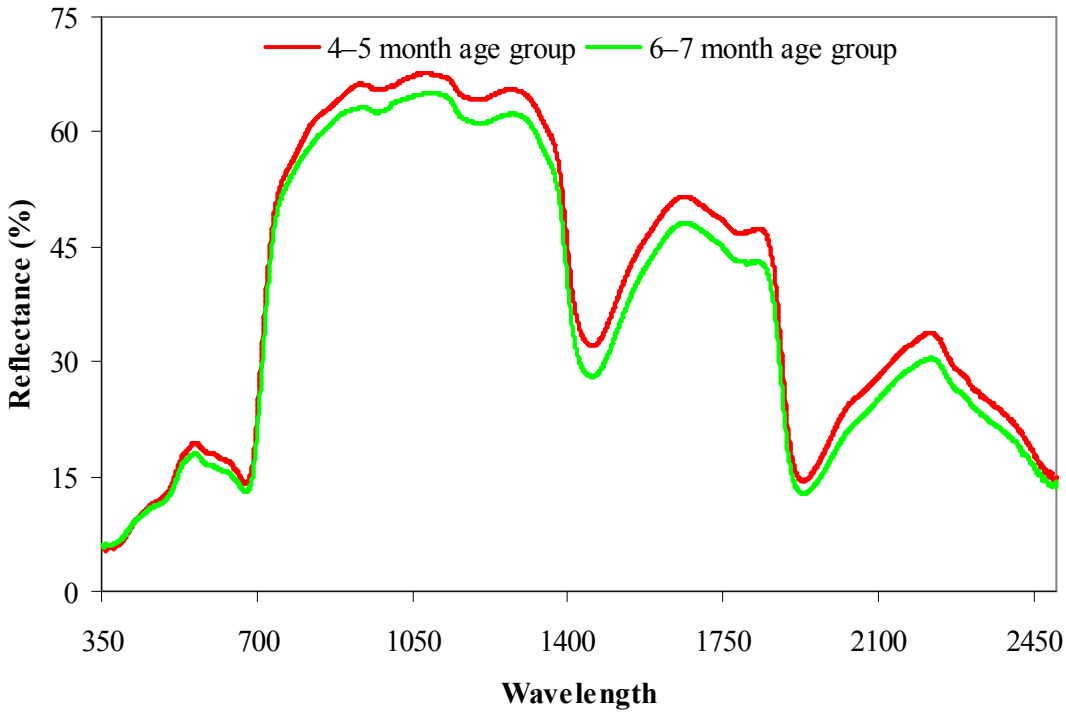

Figure 3.1. Spectral characteristics of sugarcane leaves (average of two age groups) measured under controlled conditions.

3.3.3 First-order derivative of spectra

The results of the univariate correlation test showed that the leaf first-order derivative of reflectance was significantly correlated with N concentration in some wavelengths ($p \leq 0.05$). Figure 3.4 shows the correlation coefficient at wavelengths between 400 and 2500 nm of the electromagnetic spectrum for the two age groups. The numbers of the wavebands that exhibited higher correlation ($r \geq 0.60$ or ≤ -0.60) were 19 for the younger age group (4–5 month) and 6 for the older age group (6–7 month).

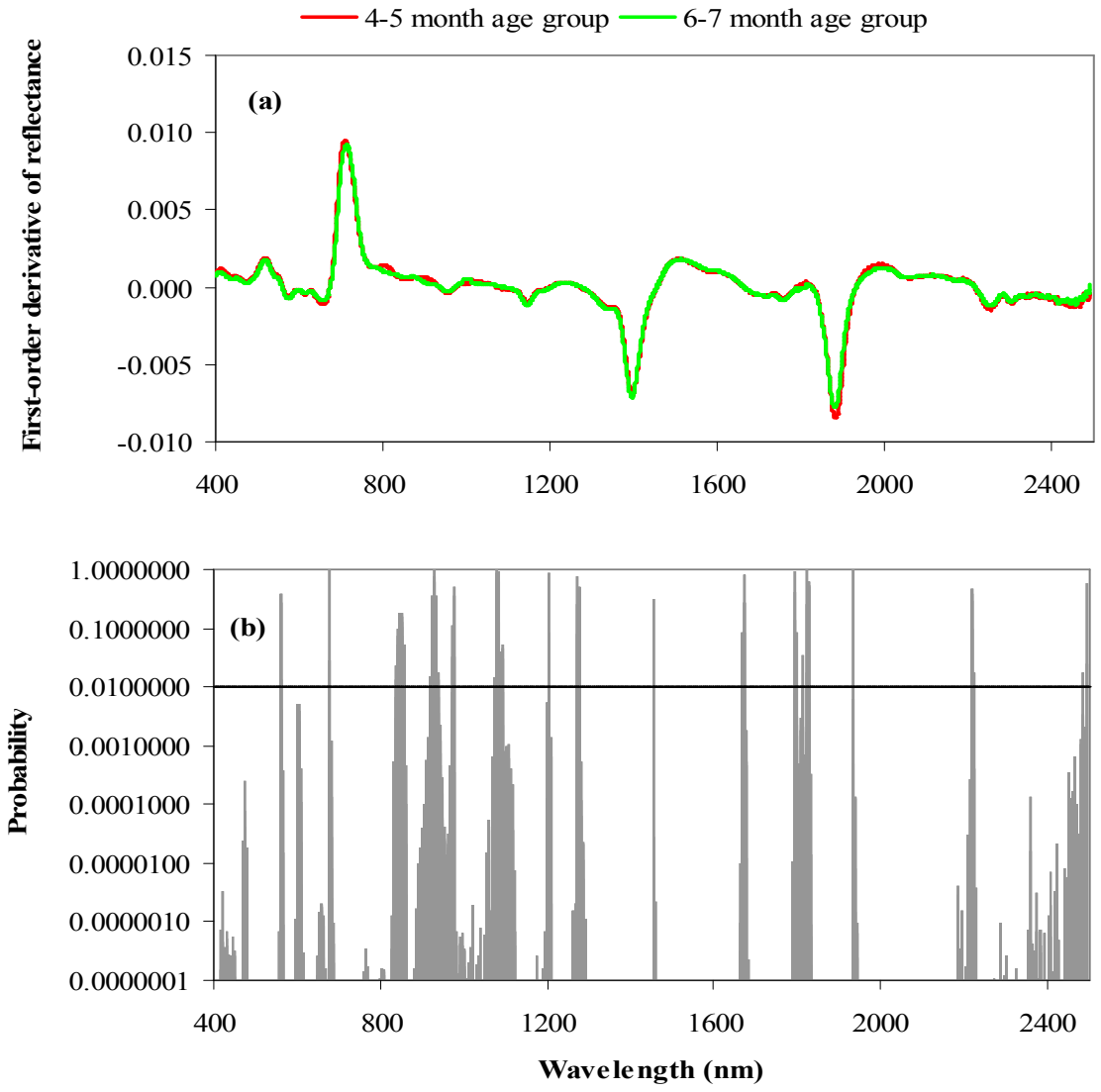

Figure 3.2. Mean first-order derivative of spectra of sugarcane leaves measured under controlled conditions (a) and results of one-way ANOVA test showing wavelengths where first-order derivative of reflectance between age groups were significant (b).

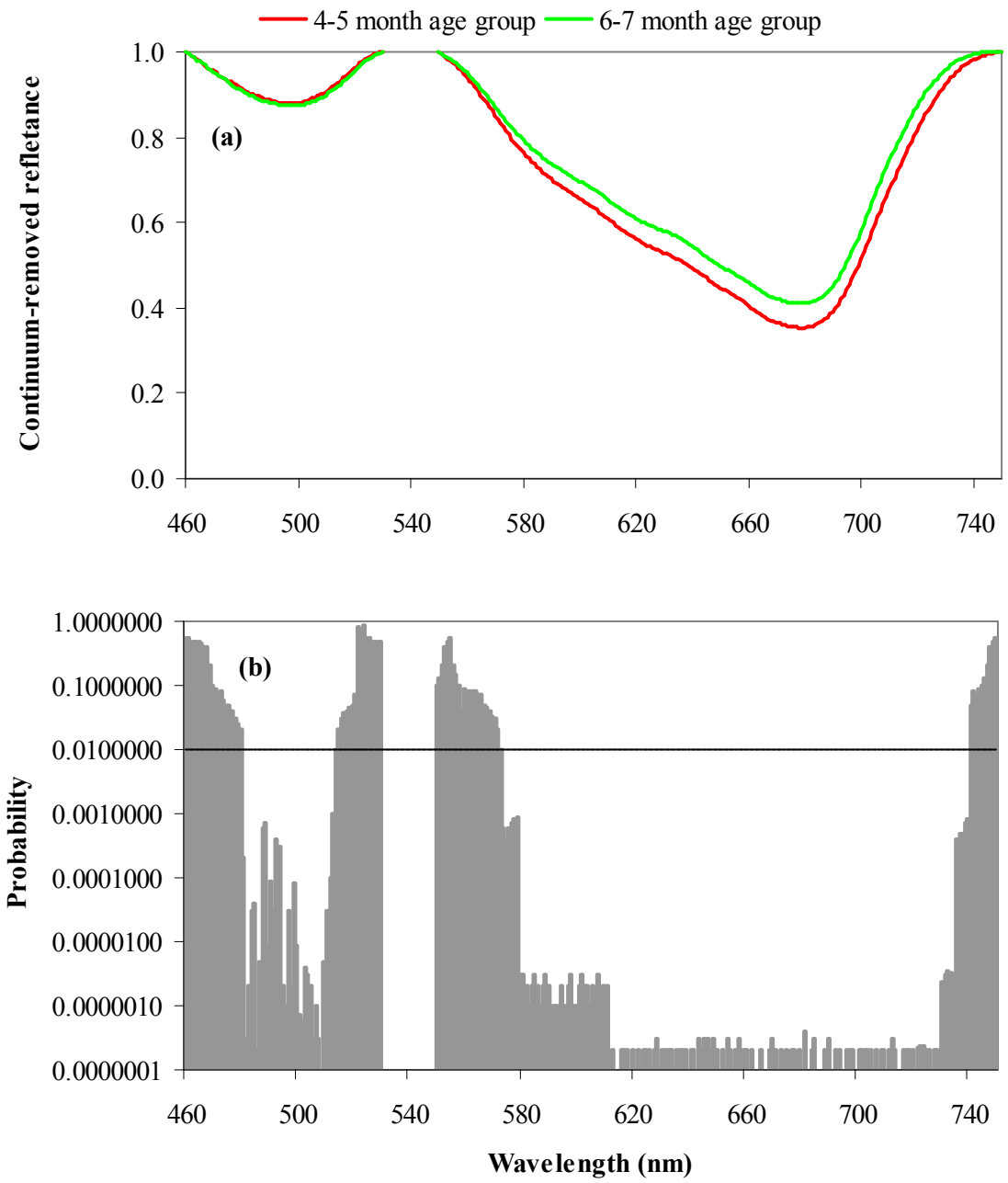

Figure 3.3. Mean continuum-removed reflectance of sugarcane leaves measured under controlled conditions (a) and results of one-way ANOVA test showing wavelengths where continuum-removed reflectance between two age groups were significant (b).

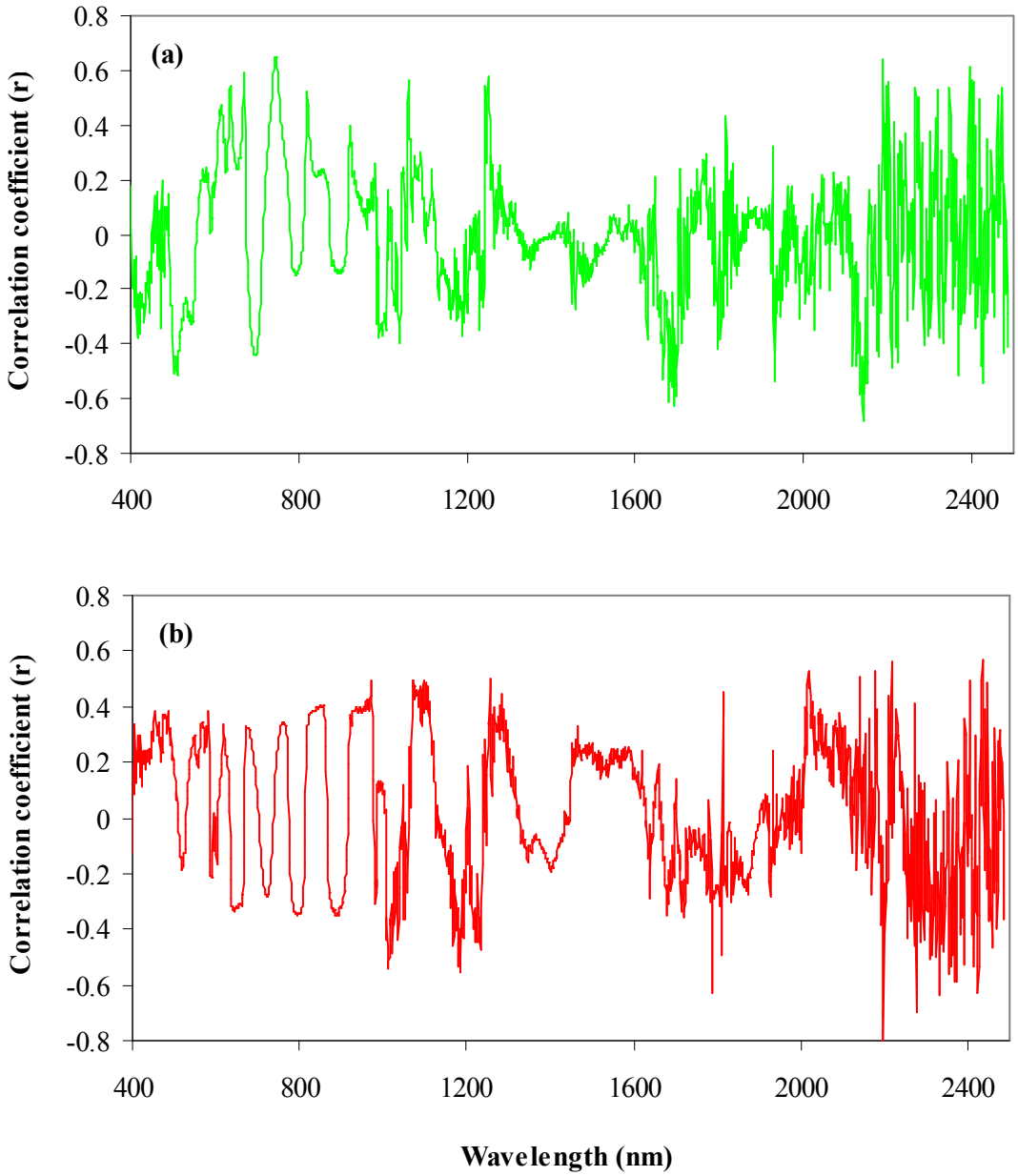

Figure 3.4. Relationship between mean first-order derivative of reflectance and sugarcane leaf N concentration for two sugarcane age groups: (a) 4–5 month and (b) 6–7 month.

Table 3.2 shows the models that produced the highest R^2 values and their regression terms. The vegetation indices included in the models for estimating leaf N concentration for the sugarcane crops of 4–5 month were generated from first-order derivative of reflectance at wavebands 744 and 746 nm which are in the visible near infrared (VNIR) region of the electromagnetic spectrum and 2142 nm which is in the middle infrared

region. In addition, 744 and 746 nm are in the red edge portion (670–780 nm) which is the region between the end of the visible and the beginning of near infrared of the spectrum. While models for the 6–7 month crops were generated from vegetation indices included wavebands 2025, 2029, 2200 and 2277 nm which are in the middle infrared region of the electromagnetic spectrum.

Table 3.2: Vegetation indices generated from first-order derivative of reflectance that yielded the highest association (R^2) with sugarcane leaf N concentration for the two studied age groups (n = 20 and 19 for 4–5 and 6–7 month age groups, respectively)

Age group	Vegetation index	Coefficient	R^2	*p*	RMSEC
4–5 month	Intercept	0.901			
	SR (744, 2142)	0.119	0.734	0.001	0.110
	Intercept	–0.179			
	NDVI (2142, 744)	–2.569	0.721	0.001	0.113
	Intercept	0.043			
	NDVI (2142, 746)	–2.345	0.721	0.001	0.113
6–7 month	Intercept	0.935			
	NDVI (2200, 2025)	–1.565	0.869	0.001	0.093
	Intercept	2.285			
	SR (2200, 2277)	0.923	0.866	0.001	0.094
	Intercept	0.986			
	NDVI (2200, 2029)	–1.503	0.865	0.001	0.094

RMSEC = Root mean square error of calibration

3.3.4 Continuum-removed spectra

Figure 3.5 shows the results of the univariate correlation test between continuum-removed reflectance and leaf N concentration. There were 22 wavelengths for the younger crops (4–5 month) and 1 wavelength for the older crops (6–7 month) at which the continuum-removed reflectance showed strong relationship ($r \geq 0.60$ or $r \leq -0.60$)

with sugarcane leaf N concentration. These wavelengths fall in the range between 726 and 748 nm for both age groups.

Table 3.3 demonstrates the results of the simple linear regression models that yielded the three highest R^2 values for both age groups. These models included vegetation indices calculated from continuum-removed reflectance at wavelengths 461, 523, 524, 525, 526, 527, 737, 738, and 748 nm.

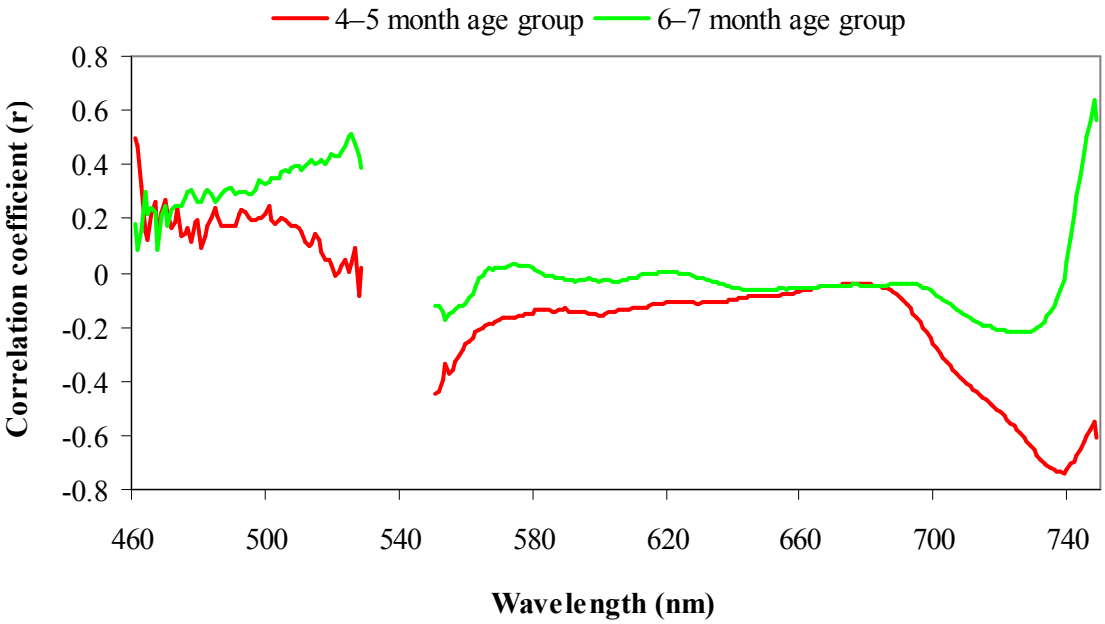

Figure 3.5. Relationship between mean continuum-removed reflectance and sugarcane leaf N concentration for two age groups.

3.3.5 Validation

The result of leave-one-out cross validation procedure is given in Figures 3.6 and 3.7, showing one-to-one relationships between predicted and measured leaf N contents. Based on R^2 and RMSEP values when measured and estimated N concentrations were compared. SR (744, 2142) and SR (2200, 2277) which were generated from first-order derivatives of reflectance showed the best predictive ability compared to other models.

Table 3.3: Vegetation indices generated from continuum-removed reflectance that yielded the highest association (R^2) with sugarcane leaf N concentration for the two studied age groups (n = 20 and 19 for 4–5 and 6–7 month age groups, respectively)

Age group	Vegetation index	Coefficient	R^2	*p*	RMSEC
4–5 month	Intercept	18.872			
	SR (738, 461)	–17.739	0.676	0.001	0.122
	Intercept	19.121			
	NDVI (737, 527)	–18.052	0.619	0.001	0.132
	Intercept	19.330			
	NDVI (737, 526)	–18.202	0.618	0.100	0.133
6–7 month	Intercept	34.282			
	SR (748, 525)	–23.291	0.384	0.100	0.302
	Intercept	25.175			
	SR (748, 524)	–23.202	0.377	0.100	0.304
	Intercept	19.659			
	SR (748, 523)	–17.702	0.369	0.100	0.405

RMSEC = Root mean square error of calibration

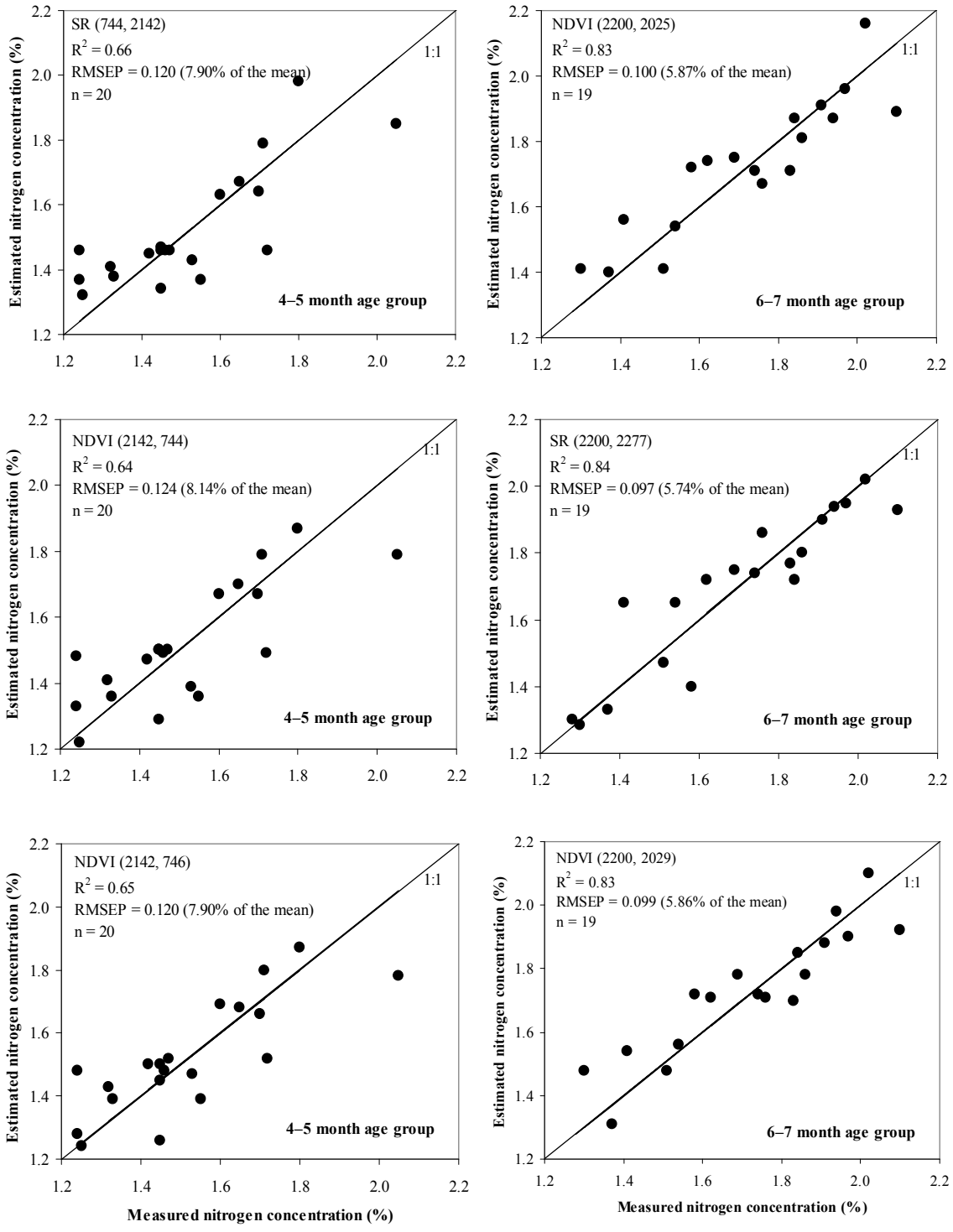

Figure 3.6. One-to-one relationships between measured and estimated sugarcane leaf N concentration (%) for the sample data sets using leave-one-out cross validation method for regression models developed from first-order derivative of reflectance.

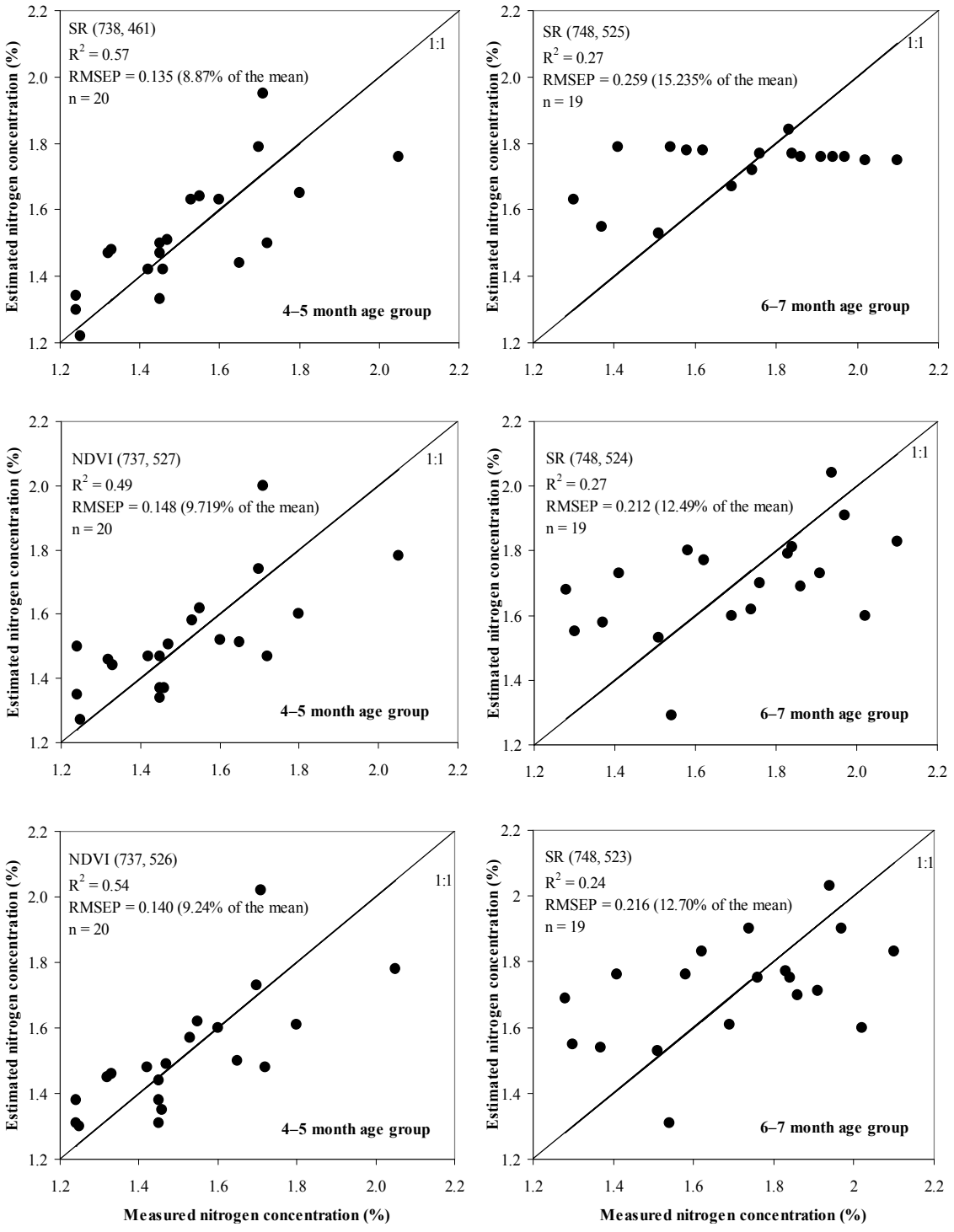

Figure 3.7. One-to-one relationships between measured and estimated sugarcane leaf N concentration (%) for the sample data sets using leave-one-out cross validation method for regression models developed from continuum-removed reflectance.

3.4 Discussion

The results of the present study suggest that information on the N status of sugarcane crops can be obtained from spectroscopic data. Specifically, sugarcane leaf N information is contained in the visible (400–700 nm), red edge (670–780 nm), and middle infrared (1300–2500 nm) regions of electromagnetic spectrum as demonstrated by the result of correlation and regression analyses. The visible and red edge portions are known pigment (chlorophyll) absorption features (Kumar *et al*., 2003) which one would expect to present a strong relationship with N (Zhao *et al*., 2005). For example, Mutanga *et al.* (2004) also found a good relationship between leaf N (in pasture) and reflectance in the red edge portion. However, Stroppiana *et al.* (2006) indicated a weak relationship between N contents and reflectance of single wavebands in the red edge portion for rice crops at canopy level. On the other hand, Jackson *et al*. (1981) found that the ratio of radiance at red (630–690 nm) and near infrared (760–900 nm) was useful for detecting nutrient (including N) in sugarcane.

The middle infrared wavelengths 1680, 1691, and 1693 nm were among those at which the first-order derivative of reflectance was well correlated with sugarcane leaf N. These bands are within ±10 nm of known N absorption feature at 1690 nm (Kumar *et al*., 2003). The results of this study exhibited some wavebands between 2025 and 2200 nm that are among the bands at which first-order derivative of reflectance had a good relation with N. This result confirms other findings such as those of Bortolot and Wynne (2003) which indicated that wavelengths between 2024 and 2204 nm were suited for predicting fresh green leaf N concentration. The present study also indicated that bands 2277 and 2279 were among the best bands at which first-order derivative of reflectance were well correlated with N concentration. These bands are within ±10 nm of the band 2268 nm at which the first-order derivative of reflectance was found to have a good relationship with vegetation N concentration (White *et al.,* 2000). High correlation coefficients (r) were also observed in this study at bands 2334, 2395, 2396, and 2427 nm. Three of these bands (2334, 2395, and 2396 nm) were within 4 nm of those reported by other authors (Luther and Carroll, 1999; White *et al.*, 2000; Mutanga *et al*., 2004). N content is related to the

absorption of electromagnetic spectrum in the middle infrared region as a result of the stretching and bending vibrations of the C–H, N–H, C–O, and O–H bonds in the plants' tissues (Meyer, 1983; Kumar *et al*., 2003).

The present study showed different predictive models for sugarcane leaf N concentration for the two age groups when first-order derivative transformation was used. The models for 4–5 month sugarcane crops involved combinations of first-order derivative of reflectance at the red edge and middle infrared regions. However, models for the 6–7 month age group involved combinations of first-order derivative of reflectance at only the middle infrared region. This is an interesting topic for further investigation. Further studies on estimating sugarcane N concentration using hyperspectral data should seek to explain these (apparent) inconsistencies and to find an optimum combination of wavebands to compile a universally applicable index. Since the leaf samples were stored for a week before the spectral measurements were taken, further studies should also look at the possible changes on leaf N content that might occur during the storage period. Models generated from first-order derivative of reflectance showed better results compared to those developed from continuum-removed reflectance. However, continuum removal was reported by other authors (e.g., Mutanga *et al*., 2003, 2004) to be a good transformation method that enhances N spectral features. On the other hand, the present study explored a procedure of developing vegetation indices rather than using the primary or transformed reflectance of a single wavelength. Several studies reported that reflectance ratios and other vegetation indices were superior to reflectance at a single wavelength value in predicting leaf N status in different crops in spite of inconsistencies in the wavelengths used in calculating vegetation indices (Tarpley *et al*., 2000; Read *et al*., 2002; Qifa and Jihua, 2003; Zhao *et al*., 2003, 2005; Jain *et al*., 2007). However, some wavelengths that included in vegetation indices generated from continuum-removed reflectance (e.g., SR 748, 525) were relatively closer to the end points on the convex hull. Consequently, spectral features at those wavelengths might be auto-correlated. That might have caused some loss in the spectral information and reduced the degrees of freedom when R^2 values were calculated. Further studies should look at

relating sugarcane leaf N content to continuum-removed variables (e.g., BD: Band Depth).

The relatively low error values (ranged between 5.74% and 8.14% of the mean measured N concentration) when models developed from first-order derivative of reflectance were validated have shown the potential to use *in situ* spectral data at leaf or canopy level to estimate sugarcane leaf N concentration. This study investigated the potentiality of spectroscopy for predicting leaf N concentration for one sugarcane variety (N19) in two age groups, since there are many varieties in different age groups in the study area. Further studies should test the utility of spectroscopic data in estimating N content for other varieties and age groups.

3.5 Conclusions

This study was undertaken to test the potential of spectroscopy to estimate sugarcane leaf N concentration. Narrow-wavelength vegetation indices developed from first-order derivative of reflectance were used in simple linear regression analyses to select vegetation indices that can explain most variations in sugarcane leaf N concentration. The results of the study indicated that spectroscopic data contain information on the N status of sugarcane crops. The SR (744, 2142) index accounted for 74% (RMSEP = 7.90% of the mean) of leaf N concentration for sugarcane of 4–5 months old and the NDVI (2200, 2025) index explained 87% (RMSEP = 5.87%) of leaf N variability of sugarcane of 6–7 months old. These indices are a combination of wavelengths at the red edge (744 nm) and middle infrared (2025, 2142, and 2200 nm) regions of the electromagnetic spectrum.

Overall, the results of this study are an important step towards the hyperspectral remote sensing and mapping of sugarcane N content at field level. Further studies are necessary under natural environmental conditions in the field with handheld, airborne, and spaceborne sensors, and should include other sugarcane varieties.

Acknowledgement

The candidate thanks the SASRI extension specialist, Mr Marius Adendorff, for his helpfulness in identifying sample fields. Many thanks to Cyril Cele, Nitesh Poona, Innocent Shezi, and Tholang Mokhele for their keen assistance in field data collection; and Mrs Tamador M. Abdelrahman for her support in spectral data collection.

CHAPTER FOUR

Estimation of sugarcane leaf nitrogen concentration using a handheld spectrometer under field conditions

This chapter is based on

Abdel-Rahman, E. M., Ahmed, F. B. and van den Berg, M., 2009. Estimation of sugarcane leaf nitrogen concentration using *in situ* spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, Remote Sensing for Africa: Special Issue, In Press.

Abstract

The aim of this study was to explore the use of *in situ* spectroscopy for estimating sugarcane leaf nitrogen (N) concentration. Leaf spectral reflectance was measured using a field spectroradiometer in the 350–2500 nm range from sugarcane variety N19 in crops of 6–7 months age under on-farm conditions. Laboratory-determined leaf N concentrations of the samples taken ranged from 1.00–1.55%. Vegetation indices based on simple ratio (SR) *viz*., SR (743, 1316), SR (743, 1317), and SR (741, 1323) generated from first-order derivative of leaf reflectance showed the best correlation with leaf N concentration, with R^2 values of 0.76, 0.75, and 0.74, respectively. The root mean square errors of prediction (RMSEP) using a leave-one-out cross validation method were 0.089% (6.85% of the mean) for SR (743, 1316), 0.092% (7.08% of the mean) for SR (743, 317), and 0.084% (6.46% for the mean) for SR (741, 1323). These results suggest that the *in situ* spectroscopy has potential use in predicting sugarcane leaf N.

Keywords: *in situ*; spectroscopy; hyperspectral; sugarcane; nitrogen estimation

4.1 Introduction

Nitrogen (N) is an important determinant in sugarcane productivity and quality (Wiedenfeld, 1995). Optimum application of N fertilizers to sugarcane crops has been studied extensively in every cane-growing region of the world and has been found to depend on growth cycle (start and length), climatic conditions, cultivar and physical, chemical, and biological properties of the soil (Meyer *et al*., 1986; Wood *et al*., 1996). Suboptimum N application implies lower yields, but excessive N can lead to poorer cane quality and increased vulnerability to certain pests (Atkinson and Nuss, 1989; Marschner, 1993), apart from the waste of expensive fertilizers. Sugarcane accumulates most of the N during the first growth stages, up to canopy closure (Thompson, 1991; Wiedenfeld, 1995). In subsequent growth stages, most N in new leaves is provided by senescing leaves lower in the canopy.

Many growers verify crop N status early in the season by leaf sample analysis while N deficiency can still be remedied. However, sampling of representative sites, pre-treatment, transport to the laboratory, the analysis itself, and communication of results to the grower can be a costly, labour-intensive, and lengthy process. The use of remote sensing techniques could provide rapid, timely, and cost-effective methods for N estimation by offering non-destructive monitoring and diagnosis methods that help with site-specific N application in wide areas (Thenkabail *et al*., 2000; Jongschaap and Booij, 2004).

Spectroscopy provides a very large number of spectral bands measured across the electromagnetic spectrum (350–2500 nm) (Jiang *et al*., 2004a; Numata *et al*., 2008). Spectroscopic data at leaf level in the visible and infrared regions of the electromagnetic spectrum is based on the principle that light interacts with the plant and influences the reflectance properties of a green leaf (van der Meer, 2003). This enables the quantification of chemical attributes of the leaves. For example, spectral absorption features have been used successfully to estimate foliar N concentration (Tarpley *et al*., 2000; White *et al*., 2000; Zhao *et al*., 2005; Kruse *et al*., 2006; Jain *et al*., 2007).

White *et al*. (2000) selected the 2028 nm wavelength to form a model for estimating vegetation leaf N concentration from *in situ* spectroscopic data. Read *et al*. (2002) found the highest correlation between cotton leaf N content and reflectance at a ratio of reflectance between 700 and 710 nm. Furthermore, Kruse *et al*. (2006) indicated that the reflectance ratio 706 to 760 nm was the best index of predicting leaf N concentration of bent grass. Additionally, Zhao *et al*. (2005) who investigated the relationship between leaf N concentration and reflectance of sorghum crops reported that first-order derivative of reflectance in the red edge centred at 730 or 740 nm yielded R^2 of 0.73.

As a first step towards the evaluation of remote sensing in estimating sugarcane N status, Abdel-Rahman *et al*. (2008a) used spectroscopic data collected under controlled (indoor) conditions. They found that a vegetation index based on normalised difference vegetation index equation *viz*., NDVI (2200, 2025) could explain 86.9% of leaf N concentration in sugarcane crops of 6–7 months old. There is, however, no information as to whether *in situ* spectroscopy at leaf level can be used to predict sugarcane leaf N content. Thus, this study aims to explore the use of *in situ* spectroscopic data for estimating sugarcane leaf N concentration.

4.2 Material and methods

4.2.1 Sampling procedure

Eleven sugarcane fields (from five farms) aged 6–7 months of sugarcane variety N19 in the large-scale sector were selected for spectral measurements and leaf samples collection. Variety N19 was selected because it is the most common variety in the study area, representing about 39% of the total crop (South African Sugar Association: SASA, 2007, unpublished industry database). The fields were sampled during the first week of March 2008. The third fully expanded leaf from the top was sampled for spectral measurements (five leaves per sample) and chemical analyses (20 leaves per sample). Two to three points in each field were sampled to comprise 25 sample points in total.

4.2.2 Leaf spectral measurements

Reflectance data of sugarcane leaves were collected using a FieldSpec® 3 spectroradiometer (Analytical Spectral Devices: ASD, 2005). The ASD uses a fibre optic cable for light collection and a notebook computer for data logging. The spectral range is 350–2500 nm with a resolution of 1.4 nm in the 350–1000 nm range and 2 nm in the 1000–2500 nm range. The leaf spectral data were collected by pointing the fibre optic cable with 1º field of view about 5 cm above the leaves on a clear sunny day between 11:00 am and 2:30 local time (Greenwich Mean Time; GMT +2). Care was taken not to cast a shadow over the leaves while taking the spectral measurements. Relative reflectance spectra were calculated by dividing leaf radiance by reference radiance from a spectralon white reference panel for each wavelength. The upper surface of the bottom, middle and top of each leaf was measured so that the spectrum of a sample point (five leaves) was an average of 15 spectra.

4.2.3 Chemical analysis

The midribs of the leaf samples that were collected for chemical analysis were removed from the leaf blades and discarded according to South African Sugarcane Research Institute (SASRI) recommended procedures (SASRI, 2003). The leaf blades were oven-dried at 75 °C for 24 hours, then ground using a leaf grinder, and oven-dried again at 75 °C for 24 hours to ensure that water content was about zero. Leaf powder, 0.11–0.25 gram per a sample, was then taken and put in a tin foil to be placed into the sample carousel of TruSpec N® instrument (Leco, 2006) which uses a thermal combustion technique to measure leaf N concentration (%).

4.2.4 Spectral transformation

4.2.4.1 First-order derivative spectra

First-order derivatives of leaf spectra, calculated according to an equation provided by ASD (2005) were used in the analysis rather than the reflectance spectra *per se*. Derivatives involve the calculation of the slope of the spectrum (rate of change of reflectance with wavelength). Derivatives are useful for reducing the variability in the spectral reflectance due to sample geometry, surface roughness etc., and for locating the positions of absorption features and inflection points on the spectra as demonstrated by de Jong (1998) and ASD (2005).

4.2.4.2 Continuum-removed spectra

Continuum removal normalises the primary reflectance spectra to allow comparison of individual absorption features from a common baseline (Kokaly, 2001). The continuum is a convex hull fitted over the top of a spectrum to connect local spectral maxima (Schmidt and Skidmore, 2003; Mutanga and Kumar, 2007). Two local start and end points on specific absorption features *viz*., $R_{460–530}$ and $R_{550–750}$ were defined. These absorption features were used by Mutanga *et al*. (2003, 2004) as chlorophyll absorption features related to foliar N concentration. The continuum-removed reflectance is obtained using the formula described by Mutanga *et al*. (2004):

$$R'_{(\lambda i)} = R_{(\lambda i)} / R_{c(\lambda i)} \qquad (4.1)$$

Where:

$R'_{(\lambda i)}$ = Continuum-removed reflectance

$R_{(\lambda i)}$ = the reflectance value for each wavelength i in the absorption feature

$R_{c(\lambda i)}$ = the reflectance level of the continuum line (convex hull) at the corresponding wavelength i.

4.2.5 Data analysis

Univariate correlations were calculated on first-order derivatives of leaf reflectance as well as continuum-removed reflectance and used to evaluate linear relationships between sugarcane leaf N concentration and reflectance values. Spectral regions between 1355 and 1420, 1813 and 1940, and 2470 and 2500 nm which are known water absorption features (Kumar *et al*., 2003; Mutanga *et al*., 2004; ASD, 2005) were removed from first-order derivatives of reflectance and excluded from the analysis due to excessive noise (Abdel-Rahman *et al*., 2008a). The wavelengths at which the slope of the reflectance and continuum-removed reflectance had strong positive or negative relationships ($r \geq 0.6$ or $r \leq -0.6$) were used to generate SR-based (Simple Ratio) and NDVI-based (Normalised Difference Vegetation Index) indices at each of the other wavelengths from 400 to 2500 nm for first-order derivatives of leaf reflectance, and from 460 to 530 nm and 550 to 750 nm for the continuum-removed reflectance (Equations 4.2 and 4.3):

$$\text{SR-based} = R_{\lambda 1i} / R_{\lambda 2i} \quad (4.2)$$

$$\text{NDVI-based} = (R_{\lambda 1i} - R_{\lambda 2i}) / (R_{\lambda 1i} + R_{\lambda 2i}) \quad (4.3)$$

Where:

$R_{\lambda 1i}$ = the first-order derivative of reflectance or continuum-removed reflectance at the i^{th} wavelength at which there was strong correlation ($r \geq 0.6$ or $r \leq -0.60$) between the first-order derivative of the reflectance or continuum-removed reflectance and N concentration

$R_{\lambda 2i}$ = the first-order derivative of reflectance at the other i^{th} wavelength between 400 to 2500 nm or continuum-removed reflectance at 460 to 530 and 550 to 750 nm.

Simple linear regression was carried out on the vegetation indices generated in this study to develop models that could be used to predict sugarcane leaf N content. The coefficients of determination (R^2) values resulting from the simple linear regressions were ranked for each spectral transformation method and the indices that gave the highest three values were recorded. A preliminary validation of the predictive models was conducted using a leave-one-out cross validation technique (Efron, 1979), whereby

observations were left out one-by-one, the regression model was re-calibrated and then used to predict the N content of the excluded leaf sample. Values of R^2 and root mean square error for the prediction (RMSEP) were recorded to test the performance of the models developed.

4.3 Results

4.3.1 Descriptive statistics

Sugarcane leaf N concentration for the samples used in this study averaged 1.30% and ranged from 1.00 to 1.55% (Standard deviation = 0.16%). Figure 4.1 shows the mean relative reflectance of sugarcane leaves measured under field conditions.

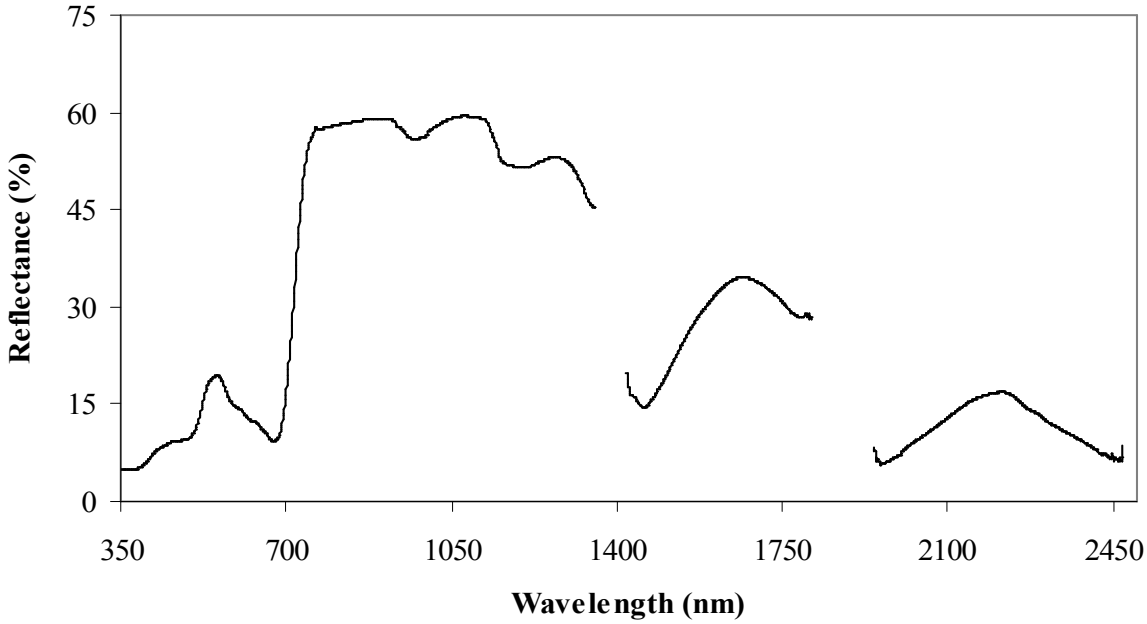

Figure 4.1. Average spectral reflectance of sugarcane leaves measured under field conditions. Spectral features between 1355 and 1420, 1813 and 1940, and 2470 and 2500 nm were removed due to excessive noise.

4.3.2 Relationships between transformed spectral reflectance and leaf N concentration

Wavebands in the ranges 418–481 nm, 551–608 nm, 697–749 nm, and wavebands centred at 1266, 2142, and 2243 nm with first-order derivatives of reflectance and/or

continuum-removed reflectance presented a significantly ($p \leq 0.05$) strong positive or strong negative correlation with leaf N concentration (Figures 4.2 and 4.3).

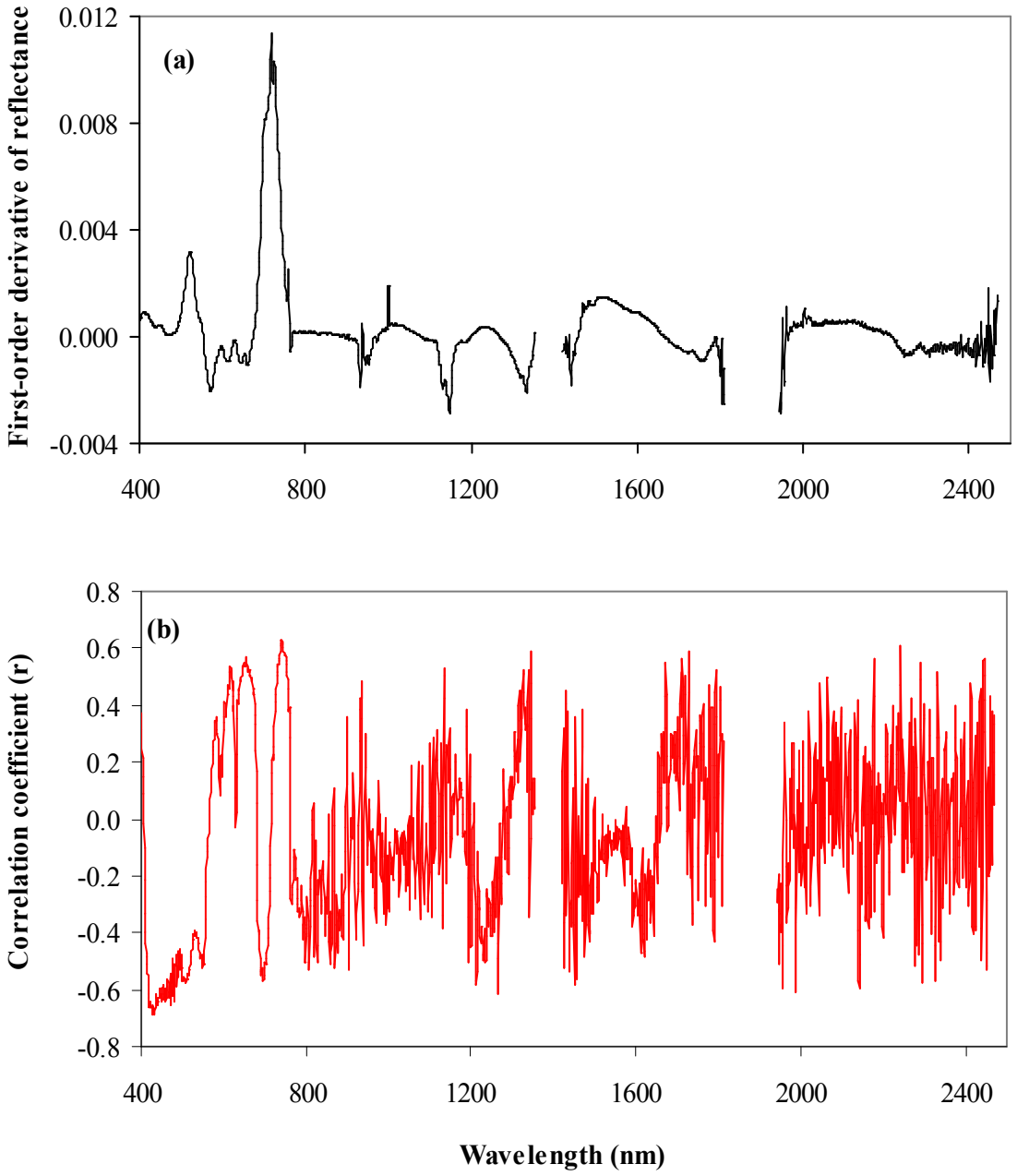

Figure 4.2. Average first-order derivative of reflectance measured under field conditions (a) and its relationship with sugarcane leaf N concentration (b). Spectral features between 1355 and 1420, 1813 and 1940, and 2470 and 2500 nm were removed due to excessive noise.

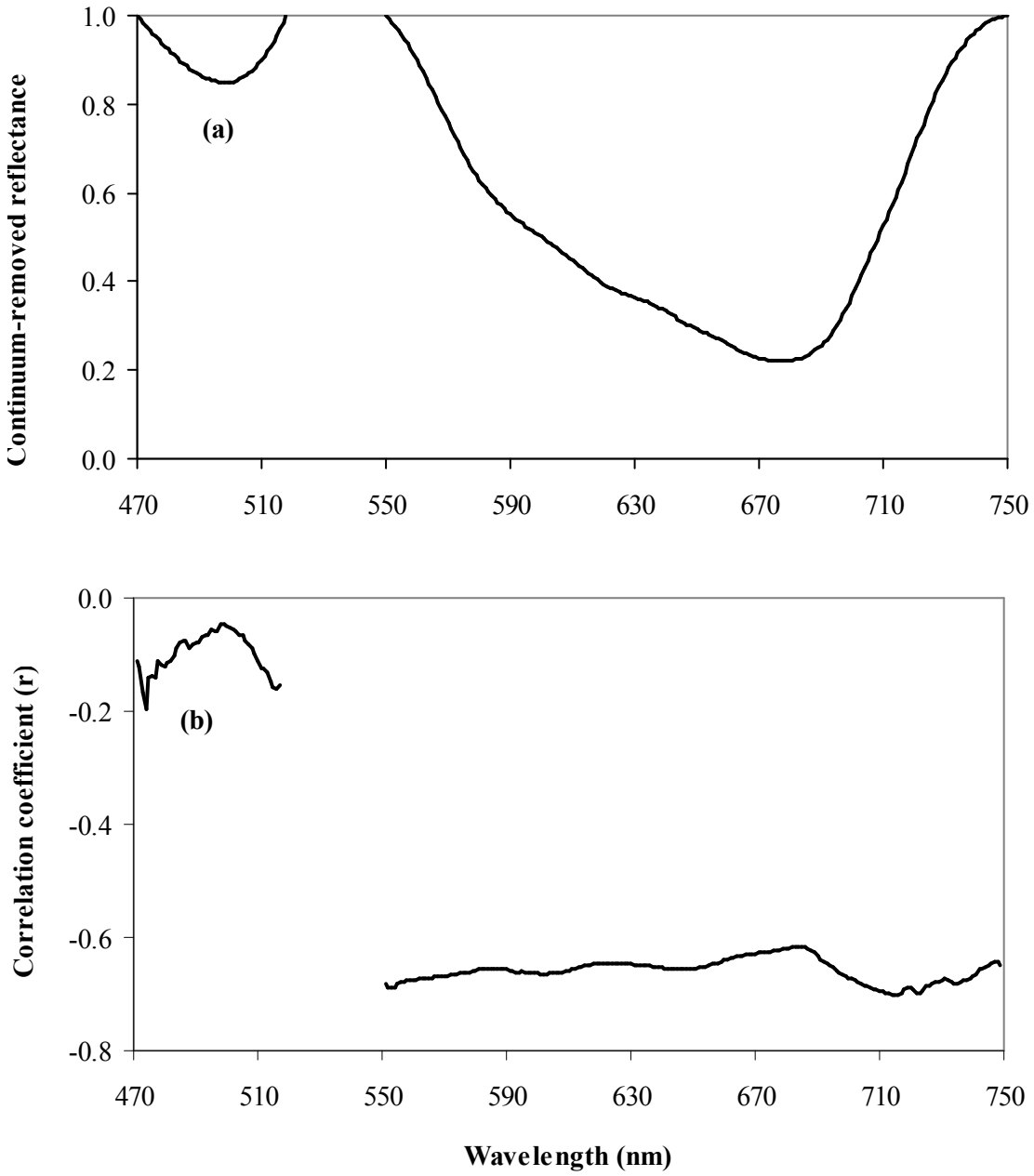

Figure 4.3. Average continuum-removed reflectance measured under field conditions (a) and its relationship with sugarcane leaf N concentration (b).

Figure 4.4 shows regression terms and scatter plots for indices: SR (743, 1316), SR (743, 1317), and SR (741, 1323) that were developed from the first-order derivative of reflectance and SR (725, 723), SR (725, 724), and NDVI (724, 725) that were generated from the continuum-removed reflectance. These vegetation indices yielded the highest R^2

values and the lowest root mean square error of calibration (RMSEC). The indices were generated from wavelengths in the red edge (741–743 nm) and middle infrared (1316–1323 nm) regions.

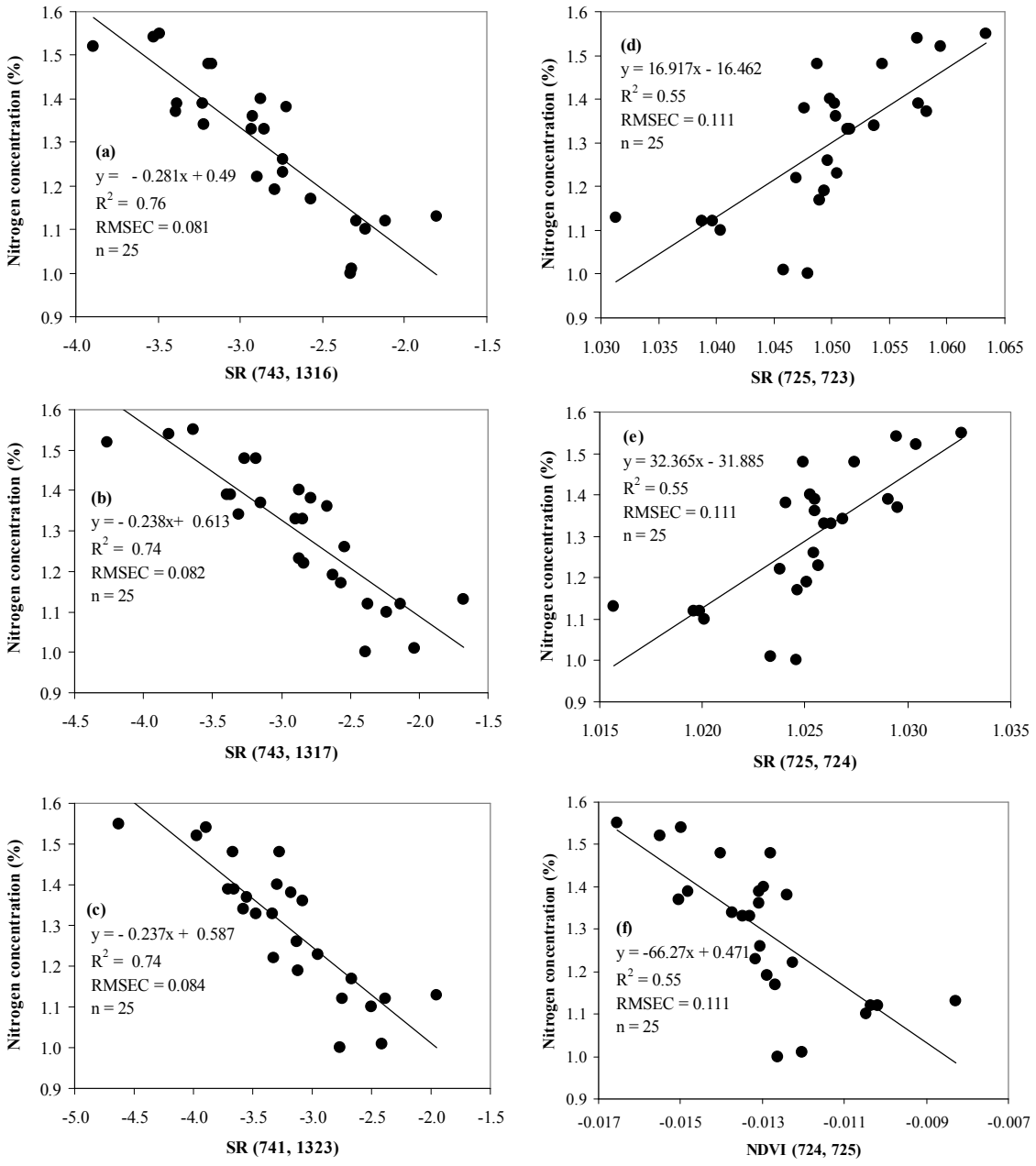

Figure 4.4. Relationships between nitrogen concentration (%) and best-modified vegetation indices developed from: (a), (b), and (c) first-order derivative of reflectance, and (d), (e), and (f) continuum-removed reflectance.

4.3.3 Validation

The result of the leave-one-out cross validation procedure is given in Figure 4.5. Low RMSEP values on one-to-one relationships between predicted and measured leaf N concentrations as well as high R^2 values of validation were obtained showing the predictive ability of the indices SR (743, 1316), SR (743, 1317), and SR (741, 1323).

4.4 Discussion

Spectral features at some wavebands in the ranges 418–481 nm, 551–608 nm, 697–749 nm, and others centred at 1266, 2142, and 2243 nm showed high correlation with N content. This result is in accordance with those of Yoder and Pettigrew-Crosby (1995), Kokaly and Clark (1999), Luther and Carroll (1999), Tarpley *et al*. (2000), Read *et al*. (2002), Bortolot and Wynne (2003), Kumar *et al*. (2003), Zhao *et al*. (2003), Nguyen *et al*. (2006) and Jain *et al*. (2007) who reported similar regions in the electromagnetic spectrum to be useful for estimating N concentration or discriminating non-deficient N plants from deficient ones. Moreover, most of these wavebands are within the visible (400–700 nm) and red edge (670–780 nm) regions which are known spectral features of pigments in vegetation (Asner, 2008). Chlorophyll is the main pigment which governs the spectral features in the visible and red edge regions (Blackburn, 1998; Kumar *et al*., 2003) and it contains ~ 6% N (Asner, 2008). Thus chlorophyll and N tend to be correlated with one another (Costa *et al*., 2001; Oppelt and Mauser, 2004; Jain *et al*., 2007). Therefore, chlorophyll spectral properties can be associated directly with N. Other wavebands are in the middle infrared region (1300–2500 nm) and are known to be N absorption features due to overtones of vibrational absorptions of the C–H, N–H, C–O, and O–H bonds (Kumar *et al*., 2003). On the other hand, the wavelengths 1316, 1317, and 1323 which are among the wavelengths that form the vegetation indices which yielded the highest R^2 are in disagreement with the results reported by Wenjiang *et al*. (2004), who found wavebands in the 1200–1300 nm region to be quite well related to foliar N content of winter wheat crops.

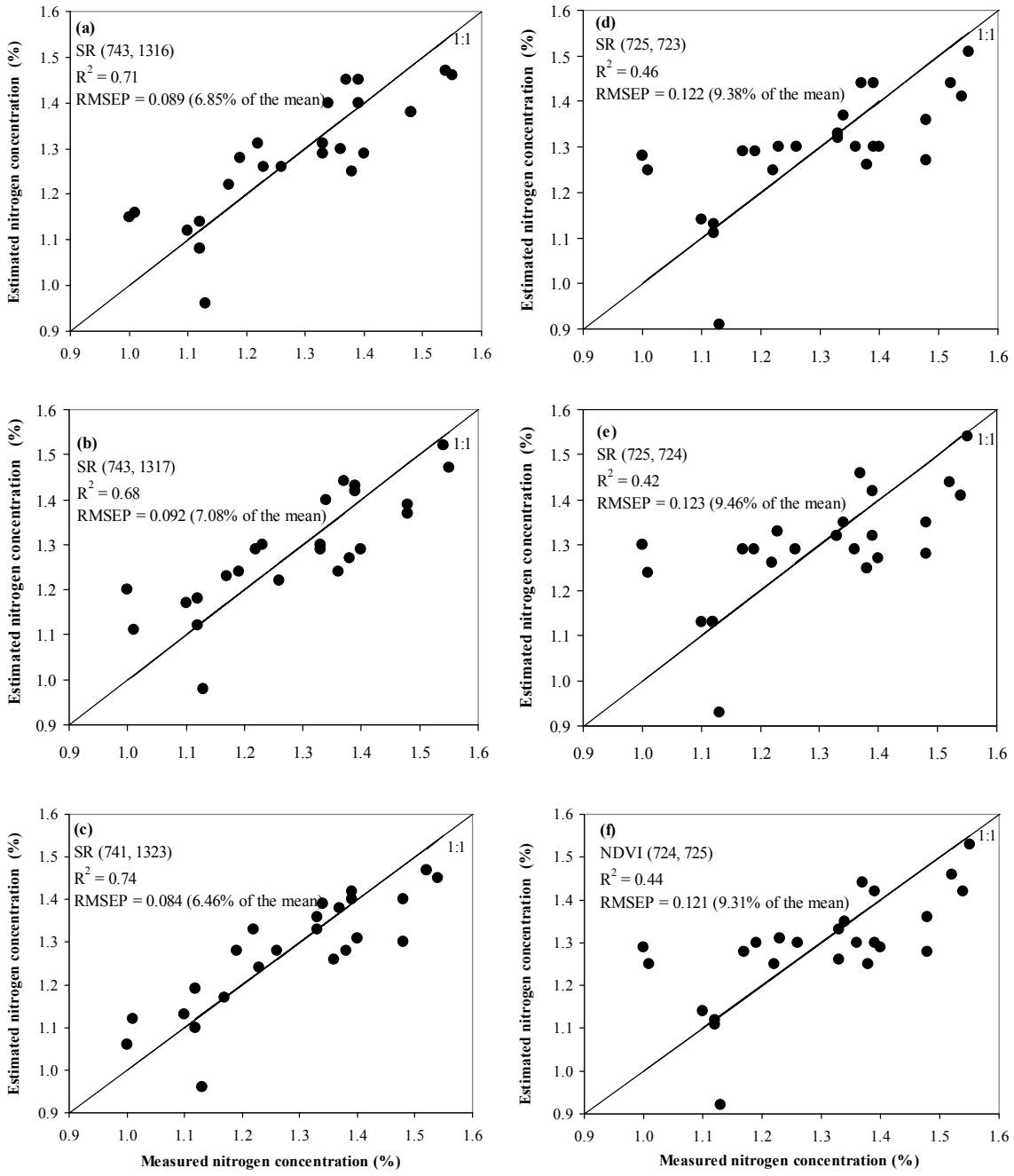

Figure 4.5. One-to-one relationships between measured and predicted sugarcane leaf nitrogen concentration (%) for the sample data set (n = 25) using leave-one-out cross validation method for the modified vegetation indices from: (a), (b), and (c) first-order derivative of reflectance, and (d), (e), and (f) continuum-removed reflectance.

It is noted that the present study reported different vegetation indices for estimating sugarcane leaf N concentration from the ones suggested by Abdel-Rahman *et al.* (2008a). This seems to be due to the fact that the spectroscopic data used for the two studies were collected under different conditions. The wind, solar angle, light intensity, and the geometry of the leaves in the field cause leaf reflectance to be quite different from that measured under controlled (indoor) conditions. This can result in different spectral features for predicting N content under field and controlled conditions.

The vegetation indices developed from the first-order derivatives of leaf reflectance showed better relationship with N content than those developed from continuum-removed reflectance. This result is in line with the findings of Zhao *et al*. (2005) who reported that first-order derivative spectra resulted in an improved relationship between leaf reflectance and sorghum leaf N content compared to the original spectra.

4.5 Conclusions

From the results, it can be concluded that:

1. The SR (743, 1316) vegetation index generated from first-order derivatives of reflectance yielded the highest R^2 value of 0.76 with sugarcane leaf N concentration.
2. The study tested one variety of one age group, and further studies should examine more varieties and age groups and validate the vegetation indices developed in this study with more samples from independent test data.

Overall, the study showed the potential value of *in situ* spectroscopy at leaf level to estimate sugarcane leaf N content. This means that there is a need to explore the usefulness of *in situ* spectroscopy at canopy level using handheld, airborne, and spaceborne sensors.

Acknowledgement

This study was funded by the University of KwaZulu-Natal in South Africa and the South African Sugarcane Research Institute (SASRI).

CHAPTER FIVE

Estimation of sugarcane leaf nitrogen concentration using EO-1 Hyperion data

This chapter is based on

Abdel-Rahman, E. M., In Preparation. Estimation of sugarcane leaf nitrogen concentration using normalised ratio indices generated from EO-1 Hyperion hyperspectral data.

Abstract

Nitrogen (N) is one of the most limiting nutrients for sugarcane production. Conventionally, sugarcane N status is examined using direct methods such as collecting leaf samples from the field followed by analytical assays in the laboratory. These methods do not offer real-time, quick, and non-destructive strategies for monitoring sugarcane N status. Methods that take advantage of remote sensing, particularly imaging spectroscopy, can present reliable techniques for predicting sugarcane leaf N content. The aim of this study was to investigate the potential of imaging spectroscopy in predicting sugarcane leaf nitrogen concentration using Hyperion data. To achieve this, two Hyperion images were captured from fields of 6–7 month old sugarcane variety N19. The machine learning random forest algorithm was used as a feature selection and regression method to analyse the spectral data. A multiple linear stepwise regression was also examined to predict concentration of sugarcane leaf N after the reduction of the redundancy in hyperspectral data. The results showed that sugarcane leaf N concentration can be predicted using both non-linear random forest regression (R^2 = 0.67; RMSEP = 0.024%; 1.32% of the mean) and multiple linear regression (R^2 = 0.71; RMSEP = 0.036%; 1.98% of the mean) models. It was concluded that concentration of sugarcane leaf N can be predicted with relatively high accuracy after the reduction of the redundancy in Hyperion data.

Keywords: imaging spectroscopy; Hyperion; sugarcane; nitrogen estimation

5.1 Introduction

Nitrogen (N) worldwide is one of the most limiting nutrients for crop production (Blumenthal *et al*., 2001; Zhu *et al*., 2008; Fageria, 2009). The growth of sugarcane as a semi-perennial crop depends highly on the application of N fertilizers. The application of an optimum amount of N results in higher sugarcane biomass production and higher N content in the plant tissues (Blumenthal *et al*., 2001). However, excessive N application may promote adverse toxicity symptoms and increase susceptibility to certain pests (Atkinson and Nuss, 1989; Rice *et al*., 2006). N is a component of amino and nucleic acids, enzymes, vitamins, and chlorophyll in the plant (Jarmer *et al*., 2003; Fageria, 2009). Hence, N content of the crop stand indicates the nutritional status of the crop and the expected quality of the final product (Jarmer *et al*., 2003). N nutrition has been a main focus of agronomic practices in sugarcane producing areas for many years, because soil N content and availability can change rapidly depending on environmental (e.g., rainfall and temperature) and crop management (e.g., tillage and irrigation) factors (Rice *et al*., 2006). The element is also important for the quality of the crop. On the other hand, any loss of N from sugarcane farms contributes to the contamination of surface and underground water (Blumenthal *et al*., 2001). Field managers now face the new challenge of striking a balance between the need to increase sugarcane yields and quality and reduce environmental impacts of excessive N applications (Ma *et al*., 2009). There is therefore a need for accurate and quick field-wide N estimation that could assist in making decisions regarding the application of optimal amounts of N in the right place at the right time.

Methods that take advantage of remote sensing, particularly the use of hyperspectral data, can provide non-destructive, cost-effective, and near-real-time monitoring routines for the estimation of sugarcane N content (Tarpley *et al*., 2000). The term "hyperspectral" which is used interchangeably with "spectroscopy", "spectrometry", and "spectroradiometry" refers to the remote sensing technique that enables capturing of spectral data in many narrow (<10 nm) and contiguous wavebands across ultraviolet, visible, and infrared regions of the electromagnetic spectrum (Lillesand and Kiefer, 2001;

Borengasser *et al*., 2008; Nansen *et al*., 2009). Hyperpsctral data can be related to foliar nitrogen concentration (e.g., Mutanga *et al*., 2004; Wenjiang *et al*., 2004; Abdel-Rahman *et al*., 2008a) due to its sensitivity in detecting any small changes in the vegetation biochemical content (Kumar *et al*., 2003). Hyperspectral data can be acquired from handheld (e.g., Analytical spectral device; ASD), airborne (e.g., Airborne Visible/Infrared Imaging Spectrometer; AVIRIS and Hyperspectral Mapper; HyMap) and spaceborne (EO-1 Hyperion) sensors and used as a monitoring tool for scheduling N applications. The use of both airborne and spaceborne spectrometers could enable the estimation of within field leaf N variation and contribute to the concept of precision farming.

Hyperspectral data are extremely large and of high dimensionality (Shen, 2007) because of the quasi-continuous spectra captured (Jiang *et al*., 2004a). Many hyperspectral features are redundant due to the strong correlation between wavebands that are adjacent (Shen, 2007; Demir and Ertürk, 2008). Therefore, the analysis of hyperspectral data is complex and needs to be simplified by selecting the most relevant features. A random forest algorithm which was developed by Breiman (2001) has recently been used as a feature selection method to reduce the redundancy in hyperspectral data (e.g., Chan and Paelinkx, 2008; Ismail, 2009; Abdel-Rahman *et al*., 2009a). The usefulness of the technique was also explored to predict (e.g., Ismail, 2009) or classify (Pal, 2005; Gislason *et al*., 2006; Everingham *et al*., 2007b) features of interest using spectroscopic data.

Selected variables as input in the predictive regression models when using hyperspectral data can be the original or transformed spectral reflectance values or vegetation indices based on these spectral values. Vegetation indices are useful in reducing variations due to irradiance, canopy geometry, and shading and in minimising the effect of soil background on the canopy reflectance (Jackson and Huete, 1991). Studies have shown that canopy level vegetation indices yielded better relationships for estimating leaf nitrogen content than reflectance from single wavebands (Jain *et al*., 2007; Zhu *et al*., 2008).

Studies have shown that canopy level hyperspectral data can be successfully used to estimate nitrogen content in various field crops such as wheat (e.g., Hansen and Schjoerring, 2003; Jarmer *et al*., 2003; Oppelt and Mauser, 2004; Wenjiang *et al*., 2004; Tilling *et al*., 2007; Zhu *et al*., 2008), rice (e.g., Inoue and Peñuelas, 2001; Yang, 2001; Xue *et al*., 2004; Nguyen and Lee, 2006; Nguyen *et al*., 2006; Zhu *et al*., 2008), maize (e.g., Osborne *et al*., 2002; Strachan *et al*., 2002), potato (e.g., Jain *et al*., 2007) and cotton (e.g., Read *et al*., 2002).

Hansen and Schjoerring (2003) found that a two-band (440 and 573 nm) vegetation index at canopy level based on a normalised difference vegetation index (NDVI) equation explained 56% of the variance in wheat leaf nitrogen concentration at a fertilizer trial. Jarmer *et al*. (2003) reported that the visible region (400–700 nm) of the spectrum had the highest influence on a partial least squares (PLS) regression model developed from canopy reflectance at the complete spectrum range (400–2500 nm) for predicting wheat nitrogen status. On the other hand, Oppelt and Mauser (2004) used AVIS data to estimate wheat leaf nitrogen content and found that chlorophyll absorption integral (CAI) index at wavebands 600 and 735 nm yielded R^2 of 0.78.

Wenjiang *et al*. (2004) demonstrated that canopy reflectance between 1000 and 1140 nm can be related to foliar nitrogen content in wheat (R^2 = 0.91) at the reviving growth stage, whereas Tilling *et al*. (2007) showed that canopy chlorophyll content index (CCCI) can explain 76% of wheat nitrogen status. In field experiments, Zhu *et al*. (2008) pointed out that ratio vegetation indices at 870 and 660 nm as well as 810 and 660 nm at canopy level were well correlated to leaf nitrogen accumulation in wheat (R^2 = 0.85) and rice (R^2 = 0.66). A relatively similar result was obtained by Xue *et al*. (2004) who concluded that a reflectance ratio index at 810 and 560 nm yielded R^2 of 0.85 for predicting rice canopy nitrogen accumulation. Inoue and Peñuelas (2001) pointed out that the spectral range of 520–570 nm had high influence on multiple linear regression models for estimating N content in rice. Furthermore, Nguyen and Lee (2006) illustrated that PLS regression models developed from rice canopy reflectance at wavelengths between 300 and 1100 nm can be used to estimate leaf nitrogen status at different growth stages (R^2 = 0.87). In

another study, Nguyen *et al*. (2006) found that PLS regression models generated from logarithm of reflectance at canopy level at the same wavebands range (300–1100 nm) used by Nguyen and Lee (2006) produced R^2 of 0.86 for estimating rice nitrogen status.

For maize crop, Osborne *et al*. (2002) reported that a multiple linear regression model consisted of canopy reflectance values at eight wavelengths (600, 610, 625, 700, 805, 875, 975, and 980 nm) yielded R^2 of 0.81 for predicting maize leaf nitrogen concentration, whereas, Strachan *et al*. (2002) recommended that canopy reflectance at red edge position can explain 81% of maize leaf nitrogen variability. For a potato crop, Jain *et al*. (2007) indicated that a canopy reflectance ratio index at 750 and 710 nm produced R^2 of 0.55 for predicting potato leaf nitrogen content. Conducting pots experiment, Read *et al*. (2002) found that cotton leaf N content was best correlated with canopy reflectance at 755 and 695 nm.

The results of the above-mentioned studies demonstrated the potential use of canopy level spectroscopic data in detecting field crops' N status. However, there was inconsistency in the spectral features used to construct models for estimating crop N status. That might be due to the different experimental conditions under which the above-mentioned studies were undertaken. With the exception of that of Jarmer (2003), these studies investigated the estimation of crop N content together with other factors such as plant density, sowing date, water stress, and cultivars. Other reason could be due to different spectral regions or vegetation indices that were employed in the above-cited studies to explore the use of spectroscopic data in predicting crop N content. In addition, with the exception of e.g., Oppelt and Mauser (2004), the above-mentioned studies used handheld sensors to collect spectral data to monitor crop N status. More studies to investigate the use of airborne or spaceborne sensors in detecting N status of crops under on-farm conditions are needed.

For sugarcane, Abdel-Rahman *et al*. (2008a, 2009d) have examined the potential of leaf level spectroscopic measurements for estimating nitrogen content under laboratory and *in situ* conditions. They found that a vegetation index based on NDVI (2200, 2025 nm)

equation developed from first-order derivative of the reflectance yielded R^2 of 0.87 for predicting N content under laboratory conditions, whereas under *in situ* conditions the simple ratio based index, SR (743, 1316) produced R^2 of 0.76. The above studies did not explore the use of spectroscopic data for estimating sugarcane leaf nitrogen concentration at canopy level. Hence, the objective of this study was to investigate the potential of imaging spectroscopy in predicting sugarcane leaf nitrogen concentration using Hyperion data at canopy level.

5.2 Materials and methods

5.2.1 Image acquisition and preprocessing

Two Hyperion images were captured (path 167/ 79) on 27 November and 1 December 2007 covering some sugarcane fields in large-scale and small-scale sectors. Hyperion has 242 wavebands in the 400–2500 nm spectral range with 10 nm spectral resolution and 30 m spatial resolution. The images were provided as Hyperion level 1B data which were calibrated to at-sensor radiance (W m^{-2} sr^{-1} μm^{-1}). Radiance images were atmospherically corrected and transformed to canopy reflectance using FLAASH (Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes) algorithm built in ENVI software (Environment for Visualising Images: ENVI, 2006). Only 196 wavebands out of 242 were calibrated: wavebands 8–57 and 79–224. These were in 427–925 nm (Visible Near Infrared: VNIR) and 933–2396 nm (Shortwave Infrared: SWIR) regions of the electromagnetic spectrum.

In order to reduce the variability in spectral reflectance due to sample geometry, surface roughness etc., and to locate the positions of absorption features on the spectrum, the atmospherically corrected images were transformed to their first-order derivatives (de Jong, 1998; Rezaei *et al*., 2008) based on the following equation:

$$d_{(\lambda i)} = (R_{\lambda(j+1)} - R_{\lambda(j)}) / \Delta_{\lambda} \quad (5.1)$$

Where $d_{(\lambda i)}$ is the first-order derivative of reflectance at a wavelength i, midpoint between wavelength j and j+1, $R_{\lambda(j)}$ is the reflectance at the j waveband, $R_{\lambda(j+1)}$ is the reflectance at the j+1 waveband, and Δ_{λ} is the difference in wavelengths between j and j+1.

Hyperion images were then geometrically registered to a Landsat TM image (7 May 2001), which was already georeferenced (Universal Transverse Mercator, South 36 zone). A nearest neighbour algorithm was performed to resample Hyperion images into their initial spatial resolution (30 m). A root mean square error (RMSE) of less than a pixel was considered to be an indicator of good geometric correction (Ferencz *et al*., 2004). Sugarcane canopy reflectance spectra from single pixels were extracted for analysis using the collected ground truth points.

5.2.2 Field data collection and chemical analysis

Twelve fields of large-scale growing sector, and six fields of small-scale growing sector, all having 6–7 month old sugarcane variety N19, were visited for leaf sample collection within 10 days of image acquisition. Variety N19 was selected as it is the most common variety in the study area, representing about 39% of the total crop (South African Sugar Association: SASA, 2007, unpublished industry database). The third fully expanded leaf from the top was sampled for chemical analysis (20 leaves per sample). Two to three points in each field were sampled to make a total of 27 sample points in the large-scale growing sector and 15 in the small-scale growing sector (total of 42 sample points). However, due to cloud cover on both images only 28 spectra in total at 28 sampling points were extracted from cloud-free pixels and used for the analysis (n = 28).

The midribs of the collected leaf samples were removed from the leaf blades and discarded according to South African Sugarcane Research Institute (SASRI) recommended procedures (SASRI, 2003). The leaf blades were oven-dried at 75 °C for 24 hours, then ground using a leaf grinder, and oven-dried again at 75 °C for 24 hours to ensure that water content was about zero. Leaf powder, 0.11–0.25 gram per a sample, was then taken and put in a tin foil to be placed into the sample carousel of TruSpec N®

instrument (Leco, 2006) which uses a thermal combustion technique to measure leaf N concentration (%).

5.2.3 Data analyses

Spectral regions between 1356 and 1466 nm and 1810 and 2012 nm which are associated with strong water absorption features and show high levels of noise (Curran, 1994; Smith *et al*., 2003; Thenkabail *et al*., 2004; Abdel-Rahman *et al*., 2008a) were excluded from the analysis. Hence, only 163 out of 196 calibrated wavebands were used in this study.

5.2.3.1 Random forest ensemble

A machine learning algorithm known as "random forest" was examined to predict sugarcane leaf nitrogen concentration using at-canopy reflectance data. Random forest uses recursive partitioning to divide the data into many homogenous subsets called regression trees (*ntree*) and then averages the results of all trees. Each tree is independently grown to maximum size based on a bootstrap sample from the training data set (approximately 70%) without any pruning. In each tree, random forest uses randomness in the regression process by selecting a random subset of variables (*mtry*) to determine the split at each node (Breiman, 2001). In each tree, the ensemble predicts the data that are not in the tree (the out-of-bag: OOB data) and by calculating the difference in the mean square errors between the OOB data and data used to grow the regression trees. The random forest algorithm gives an error of prediction called OOB error of estimate for each variable (Breiman, 2001; Maindonald and Braun, 2006; Prasad *et al*., 2006; Palmer *et al*., 2007). This error produces a measure of the importance of the variables by comparing how much OOB error of estimate increases when a variable is permuted, whilst all others are left unchanged (Archer and Kimes, 2008). Simply stated, "*variable importance is evaluated based on how much worse the prediction would be if the data for that variable were permuted randomly*" (Prasad *et al*., 2006). Moreover, random forest algorithm performs well when a large number of input variables are analysed to build a model using a few number of samples (Breiman, 2001).

The *randomforest* library (Liaw and Wiener, 2002) developed in the R package for statistical analysis (R Development Core Team, 2008) was employed to implement the random forest algorithm. The two parameters, *mtry* and *ntree*, were optimised based on the root mean square error of prediction (RMSEP). The *ntree* values were tested at 500, 1000, 1500, and 2000 based on Prasad *et al*. (2006), while the *mtry* was evaluated at all the possible 163 wavebands.

The wavebands were ranked according to their importance in predicting sugarcane leaf nitrogen content as determined by the random forest algorithm. The 10 most important wavelengths were used to generate narrow NDVI-based vegetation indices including all possible two-band combinations (10 x 10). Random forest algorithm was re-calibrated in terms of *ntree* and *mtry* and used to predict the nitrogen concentration. The vegetation indices were ranked according to their importance, and a forward selection function (Kohavi and John, 1997; Guyon and Elisseeff, 2003) was implemented to determine the least number of vegetation indices that can predict nitrogen content with greater accuracy. Random forests were iteratively fitted; at each iteration new forests were developed after including the vegetation indices in the model one-by-one (it starts with the most important ones). The nested subset of vegetation indices with the lowest RMSEP was chosen as optimal vegetation indices for leaf nitrogen estimation. RMSEP was used instead of the OOB error of estimate, because the latter is biased and could not be used to evaluate the overall error of the selection procedure (Díaz-Uriarte and de Andrés, 2006).

5.2.3.2 Stepwise multiple linear regression

Stepwise multiple linear regression was also performed to test the utility of the generated vegetation indices in predicting sugarcane leaf nitrogen concentration in a linear relationship. To avoid the overfitting problem, only the five most important vegetation indices in estimating sugarcane leaf N concentration that were ranked by random forest were used. Stepwise regression has been widely employed to relate remotely-sensed data

to vegetation variables (Kumar *et al*., 2003; Bortolot and Wynne, 2003; Zhao *et al*., 2003; Mutanga *et al*., 2004; Jain *et al*., 2007).

5.2.3.3 Validation

A leave-one-out cross validation method (Kohavi, 1995) was applied to assess the performance of the predictive models developed. This was achieved by dividing the data into *k* samples (*k* = total number of samples used for the analysis) and then the samples were removed one-by-one. The model was calibrated *k* times using all *k* samples, except for the omitted one, and used to predict sugarcane leaf nitrogen concentration on this excluded sample. One-to-one relationships between measured and predicted nitrogen values were established and coefficients of determination (R^2) as well as RMSEP values were calculated.

5.3 Results

5.3.1 Descriptive statistics

Leaf N concentration for sugarcane leaf samples used in this study averaged 1.82% and ranged from 1.36 to 2.23% (Standard deviation = 0.26). Figure 5.1 shows the mean reflectance and its first-order derivative of sugarcane canopies measured using spaceborne EO-1 Hyperion.

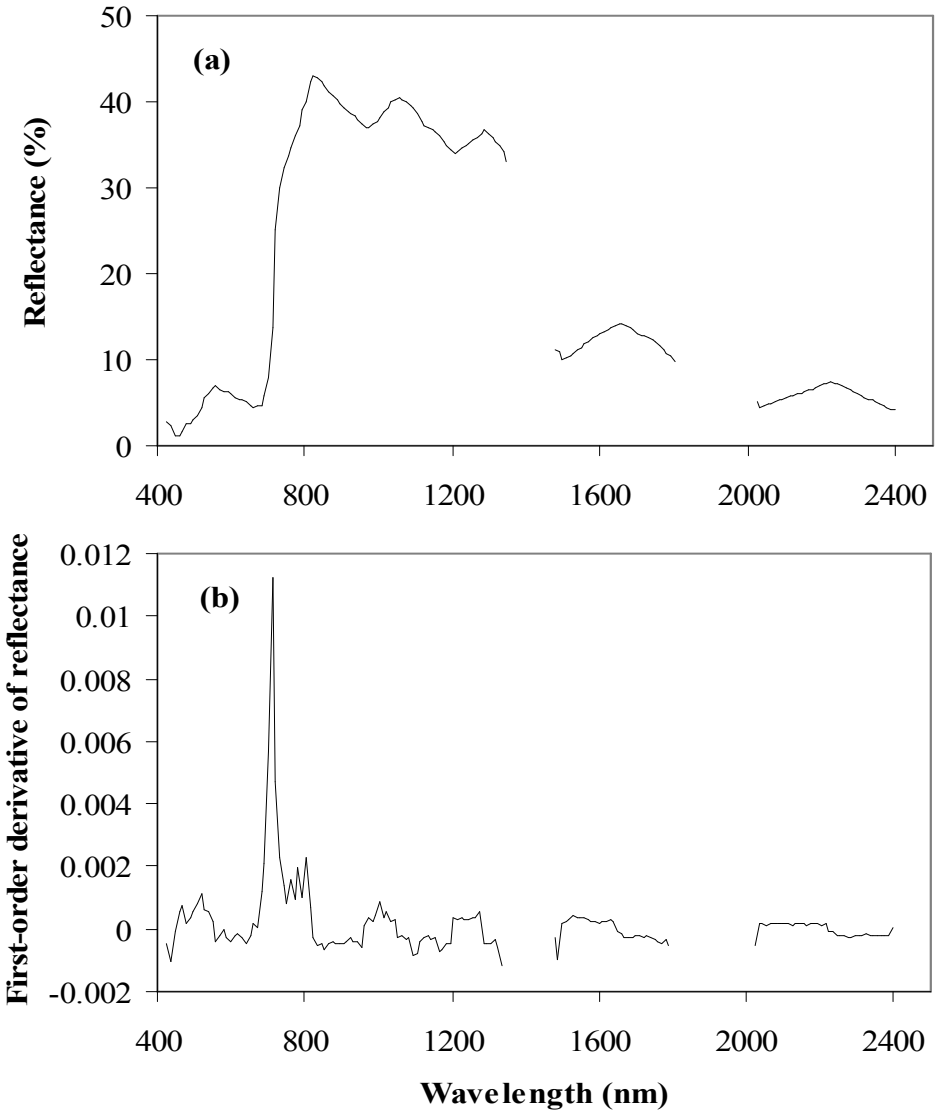

Figure 5.1. Average spectral characteristics of sugarcane canopies measured using Hyperion sensor: (a) reflectance and (b) first-order derivative of reflectance. Spectral features between 1356 and 1466 nm and 1810 and 2012 nm were removed due to excessive noise.

5.3.2 Optimisation of random forest regression models

Figure 5.2 shows the optimal *ntree* and *mtry* values for random forest regression models. The optimum *mtry* values are higher than the default value (i.e., *mtry* =1/3 of the total number of the variables) that was suggested by Liaw and Weiner (2002). The results

indicate that changes in random forest parameters (*ntree* and *mtry*) influence the error of validation.

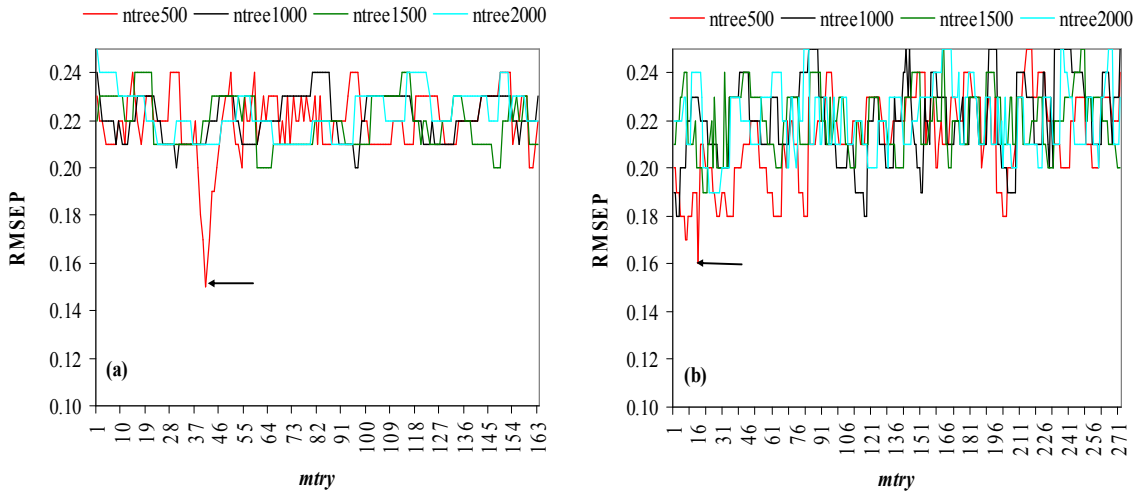

Figure 5.2. Random forest parameters (*ntree* and *mtry*) optimisation using RMSEP: (a) random forest model run with first-order derivative of reflectance and (b) random forest model run with NDVI-based vegetation indices. The arrows show the optimal *mtry* value at a definite *ntree* value.

5.3.3 Random forest predictive models

The 10 most important wavelengths in predicting sugarcane leaf N content as ranked by the random forest algorithm are shown in Figure 5.3. These wavelengths are located in the visible and middle infrared regions of the electromagnetic spectrum. A subset of 21 spectral vegetation indices shown in Figure 5.4 resulted in the lowest RMSEP for estimating sugarcane leaf N concentration when forward selection function was performed.

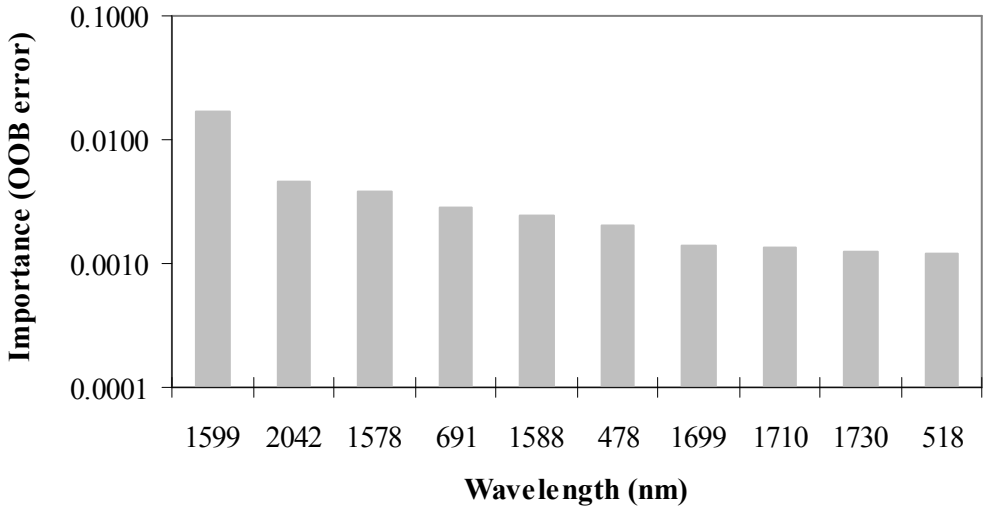

Figure 5.3. The ranking importance of the ten wavebands used to generate NDVI-based indices.

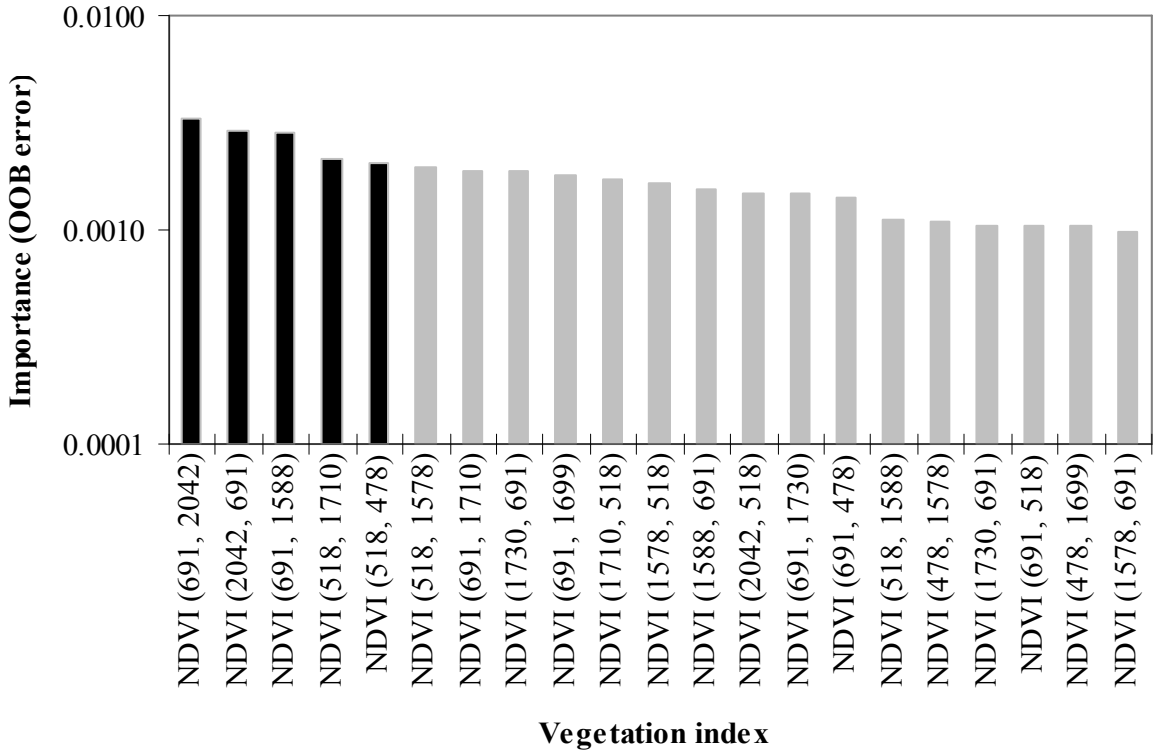

Figure 5.4. A subset of 21 NDVI-based vegetation indices selected by forward selection function and their rankings according to their importance in predicting sugarcane leaf N concentration. The black bars show the 5 most important vegetation indices used for stepwise multiple linear regression.

Table 5.1 demonstrates the result of random forest predictive models. When all generated vegetation indices were used, the random forest regression model accounted for 61% of the variability in sugarcane leaf N content. On the other hand, when utilising the subset of 21 vegetation indices, the random forest regression model explained 67% of leaf nitrogen variability. With the exception of three vegetation indices *viz.*, NDVI (518, 478), NDVI (691, 478) and NDVI (691, 518), every other vegetation index was generated from a waveband in the visible and another waveband in the middle infrared regions of the spectrum. These three vegetation indices were generated from wavebands in the visible region only.

Table 5.1: R^2 values for random forest predictive models developed. See Figure 5.4 for the subset of vegetation indices

Vegetation index	R^2	RMSEC
All developed vegetation indices	0.61	0.024
21 most important vegetation indices	0.67	0.022

RMSEC = root mean square error of calibration

5.3.4 Stepwise linear regression model

Stepwise linear regression using the five most important vegetation indices resulted in the selection of a model with two variables for predicting sugarcane leaf N concentration. Table 5.2 illustrates the selected vegetation indices, regression terms, R^2 values, and root mean square error of calibration (RMSEC) for the developed linear model. The selected vegetation indices were generated from wavelengths in the visible (478 and 518 nm), red edge (691 nm), and middle infrared (2042) regions of the electromagnetic spectrum.

5.3.5 Validation

The results of the leave-one-out cross validation method are shown in Figure 5.5. The non-linear models developed employing random forest algorithm showed a RMSEP of 0.16% (8.79% of the mean) using all generated vegetation indices and 0.024% (1.32% of

the mean) using the 21 most important vegetation indices. The linear model developed by performing stepwise linear regression yielded a RMSEP of 0.036% (1.98% of the mean). The performance of the non-linear random forest predictive model of 21 vegetation indices (Figure 5.5b) was better compared to other developed models.

Table 5.2: Regression terms and R^2 values for predictive models developed employing stepwise multiple linear regression method. The subset of 5 vegetation indices (Figure 5.4) were used

Selected vegetation index	Coefficient	R^2	*p*	RMSEC
Intercept	2.449			
NDVI (2042, 691)	0.594			
NDVI (518, 478)	–0.447	0.71	0.001	0.15

RMSEC = root mean square error of calibration

5.4 Discussion

The results have shown that N status in sugarcane crop can be predicted using imaging spectroscopic data. The developed models have illustrated that considerable information on sugarcane N status is contained in the visible range which is associated with chlorophyll (Kumar *et al*., 2003) and in the middle infrared range which is related to proteins in the plant (Curran, 1994).

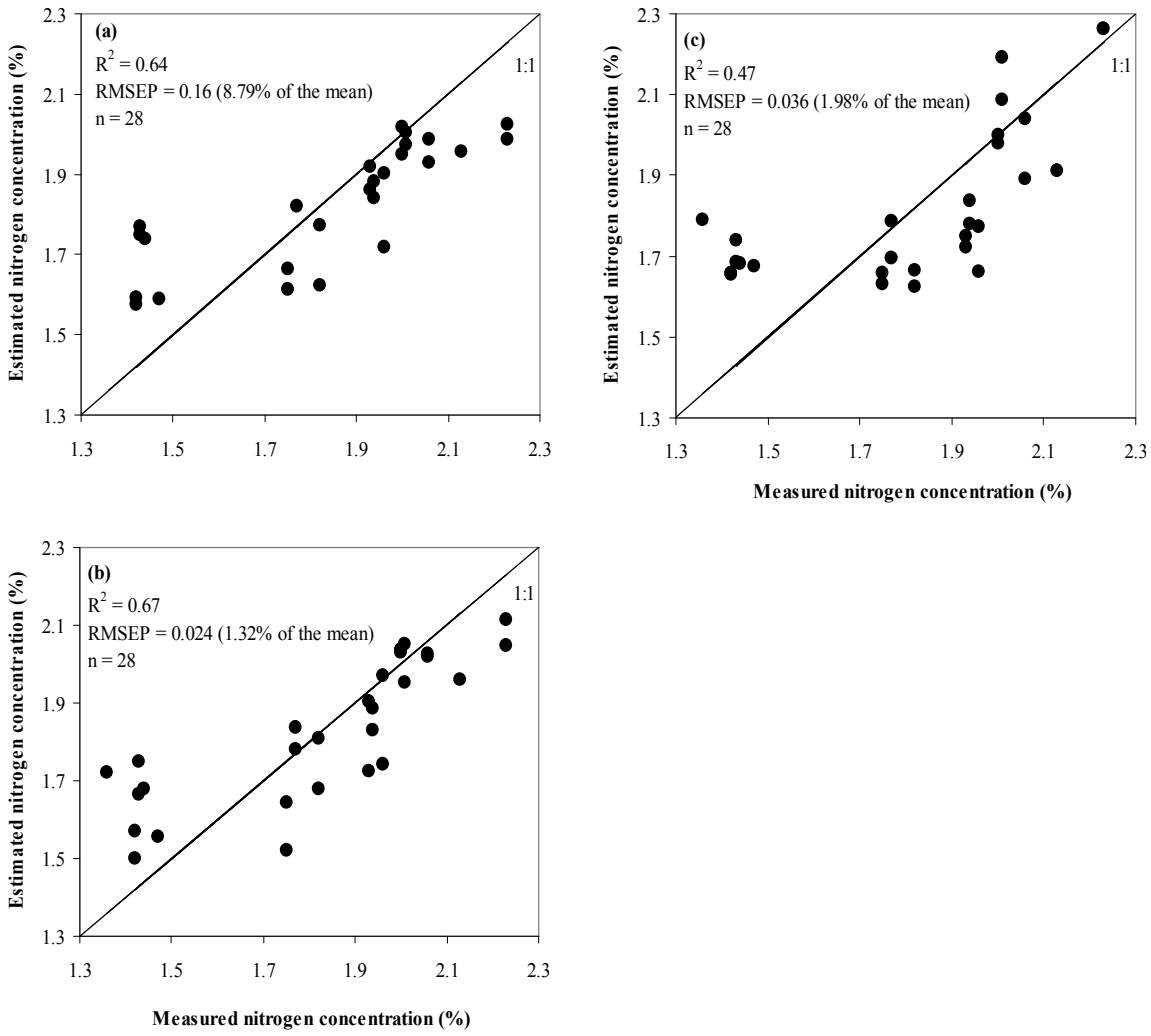

Figure 5.5. One-to-one relationships between measured and predicted sugarcane leaf nitrogen concentration (%) for the sample data set using leave-one-out cross validation method: (a) random forest regression model developed using all generated NDVI-based indices, (b) random forest regression model developed using a subset of 21 NDVI-based indices, and (c) multiple linear regression model including two NDVI-based indices developed using stepwise method.

Specifically, the 10 most important wavelengths (478, 518, 691, 1578, 1588, 1599, 1699, 1710, 1730, and 2042 nm) that were selected in this study and used to develop NDVI-based vegetation indices for predicting sugarcane leaf N concentration are within ±10 nm from known N absorption features reported in other studies. These are 480 nm (Smith *et al*., 2003), 470 nm (Serrano *et al*., 2002), 510 nm (Smith *et al*., 2003; Xue *et al*., 2004), 516 nm (Graeff and Claupein, 2003), 520 nm (Inoue and Peñuelas, 2001), 522 nm (Yoder

and Pettigrew-Crosby, 1995), 683 and 687 nm (Nguyen *et al*., 2006), 689 and 691 nm (Cho and Skidmore, 2006), 692 nm (Alchanatis *et al*., 2005), 695 (Read *et al*., 2002), 700 nm (Osborne *et al*., 2002), 1580 nm (Ferwerda *et al*., 2005), 1599 nm (Serrano *et al*., 2002), 1693 nm (Curran, 1989), 1730 (Kumar *et al*., 2003; Martin *et al*., 2008), 2048 nm (Serrano *et al*., 2002), and 2054 nm (Kokaly, 2001; Bortolot and Wynne, 2003). Furthermore, waveband 691 nm was included in most of the 21 vegetation indices that are shown in Figure 5.4 as a useful spectral feature for sugarcane N estimation. This wavelength (691 nm) also formed the NDVI (2042, 691) index which was selected by a stepwise procedure to construct a linear model for predicting sugarcane leaf N concentration (Table 5.2). The waveband centred at 691 nm was one of the spectral features used by Cho and Skidmore (2006) for calculating the red edge position when analysing Hyperion data for estimating maize N content. The red edge is the inflection point in the vegetation spectra between low reflectance in the visible region and low absorbance in the near infrared (Curran *et al*., 1990) that occurs between 670 and 780 nm. The reflectance in this inflection point is well related to chlorophyll content (Munden *et al*., 1994; Shafri *et al*., 2006; Kempeneers *et al*., 2008). Since there is a strong relationship between chlorophyll and N (Yoder and Pettigrew-Crosby, 1995; Oppelt and Mauser, 2004; Jain *et al*., 2007), reflectance at the red edge position is a good indicator of N status in the plant (Mutanga *et al*., 2004; Cho and Skidmore, 2006; Abdel-Rahman *et al*., 2008a). In addition, other wavelengths reported in this study in the range 1578–2042 nm are related to plant N content by their relation to the absorption of electromagnetic energy due to the action of the energy transition of the molecular vibration (rotation, bending, and stretching) of the C–H, N–H, C–O, and O–H bonds in the plants (Curran, 1989; Kumar *et al*., 2003).

However, other studies (e.g., Oppelt and Mauser, 2004; Wenjiang *et al*., 2004; Zhu *et al*., 2008) have shown different spectral regions for estimating crop N status compared to those reported in the present study. These studies either did not examine the spectral features at the full range of the electromagnetic spectrum (400–2500 nm) or they explored the use of a specific vegetation index only (e.g., CAI) in predicting crop N status. Specifically, the present study reports on different vegetation indices for sugarcane

leaf N estimation from those developed by Abdel-Rahman *et al*. (2008a, 2009d). That could be due to differences in the sensors characteristics used or because of different scales under which the spectral data were acquired. Abdel-Rahman *et al*. (2008a, 2009d) employed a handheld sensor with re-sampled 1 nm spectral resolution data, however, the present study used spectral data from a spaceborne sensor *viz.*, Hyperion with 10 nm bandwidth. Furthermore, Abdel-Rahman *et al*. (2008a, 2009d) collected spectral data at leaf level whereas the present study utilised canopy level spectral data. Jacquemoud *et al*. (1995) and Delalieux *et al*. (2008) suggested that spectral characteristics of vegetation at leaf and canopy levels are not comparable due to the complexity of the canopy architecture, soil background, atmospheric interferences, and differences in the illumination because of sensor-sun geometry. In this context, plant biochemical contents should be modelled firstly at leaf level and then at canopy level using different regression models when remotely-sensed data are employed (Jacquemoud *et al*., 1995).

The redundancy in hyperspectral data used in this study was reduced using the selection function of the machine learning algorithm "random forest". The successful use of this ensemble for predicting sugarcane leaf N concentration with a subset of a few vegetation indices demonstrated its utility as a feature selection method (Lawrence *et al*., 2006). Furthermore, the variability in sugarcane leaf N concentration was better explained in a linear relationship using a subset of two vegetation indices. However, The RMSEP (1.32% of the mean) of the non-linear random forest predictive model developed using a subset of 21 vegetation indices was lower than the one produced by the multiple linear predictive model (1.98%) of two vegetation indices. Generally speaking, the predictive models developed in this study overestimated leaf N concentration for the samples with low N content and underestimated N concentration of the sample with relatively high N content. Samples of low N content were collected from fields in the small-scale growing sector while samples of high N content were collected from fields in the large-scale growing sector. In addition, fields in the small-scale growing sector are grown under dry land conditions, whilst growers in the large-scale growing sector usually have full or supplementary irrigation. This kind of discrepancy could have affected sugarcane spectral properties when N concentration was estimated. Other factors such as damage by insect

pests and diseases might have also confounded the relationship between sugarcane spectral features and N content. Since the predictive models did not yield any kind of systematic bias that could be used to adjust the values of predicted N content, future work should study sugarcane N estimation in each growing sector (i.e., small-scale and large-scale) separately.

This study used data of one sugarcane variety of one age group, and since there are many varieties of different age groups grown in the study area, further studies should assess the utility of imaging spectroscopy in predicting sugarcane N content in other varieties of different age groups. In the small-scale growing sector, the fields are relatively small and elongated. Given the spatial resolution of Hyperion (30 X 30 meter), future work should explore sugarcane N estimation in the large-scale growing sector only, if remotely-sensed data are to be acquired using Hyperion or any other sensor of the same or larger pixel size.

5.5 Conclusions

The results of this study have shown that there is potential for using imaging spectroscopy for predicting sugarcane leaf N concentration. The random forest regression algorithm was deemed to be a robust method for reducing the redundancy in the complex hyperspectral data. The estimation of sugarcane leaf N concentration was relatively accurate after the reduction of the redundancy in the hyperspectral data. Further research work should study sugarcane N prediction in only the large-scale growing sector if remotely-sensed data are to be acquired from Hyperion or other sensor(s) of the pixel size like Hyperion or larger.

Overall, the successful use of imaging spectroscopy in estimating sugarcane leaf N concentration can assist in making decisions regarding site-specific applications of N fertilizers and thus contribute to the concept of precision farming.

Acknowledgments

The candidate thanks the South African Sugarcane Research Institute (SASRI) and the University of KwaZulu-Natal in South Africa for funding this study. Thanks are extended to SASRI extension specialist, Mr Marius Adendorff, for his keen help in identifying the sample fields. Many thanks to Cyril Cele, Nitesh Poona, Innocent Shezi, and Tholang Mokhele for their support in the field data collection. Gratitude is extended to the R development core team for their very powerful open source packages for statistical analysis.

CHAPTER SIX

Detection of sugarcane thrips (*Fulmekiola serrata* Kobus) damage using leaf level spectroscopic data

This chapter is based on

1. **Abdel-Rahman**, E. M., Ahmed, F. B., van den Berg, M. and Way, M. J., 2008. Preliminary study on sugarcane thrips (*Fulmekiola serratea*) damage detection using imaging spectroscopy. *Proceedings of South African Sugar Technologists' Association*, 81, 287–289.

2. **Abdel-Rahman**, E. M., Ahmed F. B, van den Berg, M. and Way, M. J., 2009. Potential of spectroscopic data sets for sugarcane thrips (*Fulmekiola serrata* Kobus) damage detection. *International Journal of Remote Sensing,* In Press.

Abstract

Sugarcane thrips was detected in South African sugarcane in 2004. Since then it has become widespread in South Africa. The South African Sugarcane Research Institute (SASRI) conducts field surveys to monitor this pest, but this is time-consuming and costly. As a first step towards evaluation of remote sensing for thrips monitoring, a preliminary experiment and analysis at leaf level were conducted using a handheld field spectroradiometer covering the 350 nm to 2500 nm range of the electromagnetic spectrum to detect sugarcane thrips damage. Reflectance spectra of sugarcane leaves with different levels of thrips damage, from two popular varieties (N19 and N12), were measured and statistically analysed using one-way analysis of variance, sensitivity analysis, and canonical discriminant analysis. The results of the analyses showed that there were significant differences in spectral reflectance and derived variables used in the study at the different levels of thrips damage. The red edge region of the visible portion gave the highest significant differences and levels of separability among the damage classes. It is hypothesised that this might be associated with chlorophyll and nitrogen deficiencies induced by thrips.

Keywords: sugarcane thrips; spectral reflectance; spectroscopy

6.1 Introduction

Sugarcane thrips, *Fulmekiola serrata* Kobus (Thysanoptera: Thripidae) was recorded for the first time in South Africa at Umfolozi, in 2004 and is now present throughout the industry (Leslie, 2005, 2006; Way *et al*., 2006). This insect is a recognised sugarcane pest in many countries including India, China, Indonesia, Philippines, Mauritius, Barbados, Dominica Republic, and Guyana. Thrips is a small insect (2–3 mm) that feeds on the young spindle leaves of sugarcane and other grasses causing chlorotic patches on leaves. Symptoms of sugarcane thrips damage are usually noticeable on the open leaves. Nymphs gather in the central leaf roll and are responsible for most of the damage. The nymphs pierce the leaf tissues with their mouthparts and suck the plant sap, usually from the upper parts of the young leaves. The damaged central leaf roll shows yellowish-white blotches when it unfurls. Afterwards, the patches turn light or dark yellow and become irregular patches or stripes, which may cover most of the seriously damaged leaves (Leslie, 2005, 2006). Another symptom is that leaf tips often remain stuck together. The effects of damage are manifested mainly on young cane plants (Mound and Teulon, 1995). Leslie (2005) stated that serious leaf damage may slow down growth and therefore affect yield, particularly during drier periods. In China, for example, several species of thrips (the major species being *F. serrata*) reportedly caused a reduction in sugarcane yield of up to between 10 and 15% (Leslie, 2005).

The South African Sugarcane Research Institute (SASRI) applies direct methods such as scouting and checking sugarcane leaves in the field for monitoring sugarcane thrips. However, these methods are time-and labour-intensive and hence very costly. Methods that take advantage of remote sensing could possibly provide near-real-time and cost-effective strategies for thrips monitoring. As a first step, this study aimed to examine the potential of spectroscopic data sets to detect sugarcane thrips damage at leaf level. This was done by analysing the spectral reflectances of different thrips-damaged sugarcane leaves, identifying narrow spectral bands or combinations of bands that are most sensitive to variation in thrips damage and gaining insights in the relationship between sugarcane thrips damage and the biochemical composition of the crop.

6.2 Applications of remote sensing for crop pests and diseases monitoring

Spectral reflectances of leaves, or canopies, in both the visible and near infrared regions of the electromagnetic spectrum provide information on their structure, physiology, and chemistry (Kumar *et al*., 2003; Jørgensen *et al*., 2006).

A recent review by Abdel-Rahman and Ahmed (2008) showed that applications of remote sensing in sugarcane agriculture have mostly focused on areal extent mapping, thermal age group classification, varietal discrimination, yield prediction, and monitoring water and nutritional status. The utility of remote sensing for pest and disease monitoring for sugarcane has only been examined by Apan *et al*. (2004b) who conducted a study on rust disease mapping in Australia by evaluating several narrow-band vegetation indices using EO-1 Hyperion data. Their results verified that areas with sugarcane orange rust disease showed differences in spectral reflectance signature and can be discriminated from non-diseased areas using combinations of wavebands at the visible near infrared (VNIR) and shortwave infrared regions of the electromagnetic spectrum.

Applications of remote sensing in other crops, for rapid detection of pest damage or disease, also include the use of handheld optical devices (Lorenzen and Jensen, 1989; Malthus and Madeira, 1993; Parker *et al*., 1995; Riedell and Blackmer, 1999; Yang and Cheng, 2001; Zhang *et al*., 2002; Sudbrink *et al*., 2003; Moshou *et al*., 2004; Yang *et al*., 2005; Mirik *et al*., 2006a; Xu *et al*., 2007; Mirik *et al*., 2007) and airborne sensors (Sudbrink *et al*., 2003; Huang *et al*., 2007).

Lorenzen and Jensen (1989) observed significant changes in reflectance measured by spectroradiometer (Licor Li 1800) of barley leaves infected with cereal powdery mildew in the visible region of the spectrum. Parker *et al*. (1995) compared visual estimates of severity of cereal leaf diseases (*Septoria tritici* and *Erysiphe graminis*) with measured severities that were determined using image captured on acetate sheets. They found that the visual estimates differed substantially from the estimates derived from the image. Moshou *et al*. (2004) classified images captured by a spectrograph to separate wheat

canopies infected by yellow rust disease from healthy ones. The images were normalised to reflectance and four wavelengths centred at 543 ±10, 630 ±10, 750 ±10 and 861 ±10 nm were selected as discriminant variables. The results showed classification accuracy up to 99% by using a neural network based classifier. On the other hand, Huang *et al*. (2007) found that the level of yellow rust disease in wheat can be related to a photochemical reflectance index (R^2 = 0.97) and a disease index (R^2 = 0.91) that were generated from airborne hyperspectral images. The damage to wheat by insect pests has also been studied fairly extensively. Riedell and Blackmer (1999), for example, noted that leaf reflectance in the 625–635 nm and 680–695 nm ranges, as well as the normalised total pigment to chlorophyll *a* ratio index (NPCI), were significantly correlated with total chlorophyll concentrations in wheat damaged by greenbugs and Russian wheat aphids.

In addition, Mirik *et al*. (2007) reported that canopies of wheat infested by Russian wheat aphids had significantly lower reflectance in the near infrared region and higher reflectance in the visible range of the spectrum when compared with non-infested canopies. Yang *et al*. (2005) studied stress in wheat canopies caused by greenbug infestation using multispectral radiometry. Their results demonstrated that reflectance from a band centred at 694 nm, and vegetation indices derived from bands centred at 800 and 694 nm were the most sensitive to identifying the damage due to greenbug infestation.

In another study, Mirik *et al*. (2006a) measured wheat canopies' reflectance using a hyperspectral field spectroradiometer and a digital camera, and they illustrated that damage sensitive spectral indices ($DSSI_1$ and $DSSI_2$), simple ratio (SR) index, and normalised difference vegetation index (NDVI) were strongly related to damage by greenbug. For rice crops, Yang and Cheng (2001) showed that canopy reflectance spectra acquired by spectroradiometer from six levels of infestation by plant hoppers were clearly discriminated, especially in the region of 734–925 nm of the spectrum. Furthermore, Malthus and Madeira (1993) discovered detectable changes in first-order derivative of reflectance over the visible and near infrared regions in field bean leaves infected by *Botrytis fabae*. For a cotton crop, Sudbrink *et al*. (2003) utilised reflectance

data from spectroradiometer at leaf level as well as from an aerial image, at canopy level, to identify variability in plant growth and correlation with larval densities of beet armyworm (BAW) and cabbage looper. Their results showed that BAW-damaged leaves reflected lower than the healthy ones when leaf level spectral data were analysed. Using canopy level observations, BAW infestations appeared to be associated with lower NDVI values. The study was unable to detect cabbage looper damage. In studies on tomato plants, Zhang *et al*. (2002) analysed canopy spectral data acquired using a field spectroradiometer and found that spectral bands of green to red and near infrared regions of the spectrum were related to late blight infection. Xu *et al*. (2007) noted that tomato leaves' reflectance captured using a spectrometer decreased significantly with increasing severity of leaf miner between 800 and 1100 nm and increased in severity for individual bands at 1450 and 1900 nm. Moreover, reflectance at 1450 and 1900 nm was highly correlated with severity of miner infestation (R^2 = 0.98 and 0.96, respectively).

6.3 Materials and Methods

6.3.1 Field sampling and categorisation

Leaf samples used in this study were collected in January and February 2007 as part of a survey by SASRI for sugarcane thrips identification and monitoring. The survey team targeted only large-scale growing sector's fields with different varieties from coastal and inland farms. Five spindle samples were randomly taken from each field with five leaves in each sample. The leaf samples collected were placed in plastic bags and kept cool in a fridge for five days until they were taken to the laboratory, where they were stored at –18 ^{o}C for about a month before analysis. In order to avoid confounding effects, leaves of the same age (3–4 months) after planting or ratooning were selected. For this study, only leaf samples from two varieties *viz*., N19 and N12 were selected. Variety N19 was chosen because it is the most common variety in the study area representing about 39% of the total crop (South African Sugar Association: SASA, 2007, unpublished industry

database), whereas N12 is the most widely grown variety in the rainfed regions of the South African sugar industry (SASRI, 2006).

The collected leaf samples were visually categorised into four damage classes using a similar approach as described by Yang and Cheng (2001) and Xu *et al*. (2007): healthy, and damage classes that were low, medium, and severe. In the case of N12, the low damage class was not present among the leaf samples. After the categorisation, 24 leaves were taken from each damage class for spectral measurements.

6.3.2 Leaf spectral measurements

Before analysis, after removal from the deep-freezer, the categorised leaf samples were kept indoors for three hours to allow the leaves to be free from any water droplets. A FieldSpec® 3 spectroradiometer (Analytical Spectral Device: ASD, 2005) was then used to collect relative reflectance data from each leaf sample under natural outdoor conditions. The ASD is a compact, battery powered, portable spectrometer which uses a fibre optic cable for light collection and a notebook computer for data logging. The spectral range is 350–2500 nm with a resolution of 1.4 nm in the 350–1000 nm range and 2 nm in the 1000–2500 nm range. The data collection time is 0.1 second per spectrum.

The leaf spectral data were collected by pointing the fibre optic cable with 1° field of view at a distance of 0.1 m above the leaf on a clear sunny day between 11:00 am and midday local time (Greenwich Mean Time; GMT +2). The instrument was optimised and calibrated before the first measurement and after every five minutes onwards to adapt to the changing atmospheric conditions (Luther and Carroll, 1999; ASD, 2005). Reflectance spectra, relative to a spectralon white reference panel were calculated by dividing leaf radiance by radiance from the reference panel for each wavelength. For each measurement, 25 scans were performed by the ASD, and the spectral data were averaged over the measurement. Two measurements per leaf were taken and the resulting data were averaged over the leaf.

6.3.3 Data analysis

6.3.3.1 One-way analysis of variance (ANOVA)

The study tested the hypothesis that there is a significant difference in the sugarcane leaf mean reflectance among different thrips damage classes at the 400 nm to 2500 nm range of the electromagnetic spectrum. This was tested using one-way analysis of variance and a post hoc-Tukey test (SPSS, 2006) was used to assess the occurrence of significant differences among thrips damage classes. The spectral regions between 1355 and 1450 nm, 1800 and 1950 nm, and 2420 and 2500 nm which are known water absorption features (Tian *et al*., 2001; Kumar *et al.*, 2003; Pu *et al.*, 2004; Mutanga *et al*., 2004; ASD, 2005; Datt *et al*., 2006) were excluded from the analysis and removed from the spectra for causing excessive noise. In order to identify the most responsive bands amongst those at which significant differences in reflectance between different thrips damage levels occur, reflectance difference and sensitivity analysis were performed following the formula described by Cibula and Carter (1992), Carter (1994), and Riedell and Blackmer (1999):

$$\text{Reflectance difference} = (R_{\lambda h} - R_{\lambda i}) \quad (6.1)$$

$$\text{Sensitivity reflectance} = (R_{\lambda h} - R_{\lambda i}) / R_{\lambda h} \quad (6.2)$$

Where $(R_{\lambda h} - R_{\lambda i})$ is the spectral reflectance curve calculated by subtracting the mean reflectance of the healthy sugarcane leaves at each wavelength from the mean reflectance of the low, medium, and severely thrips-damaged leaves.

The coefficient of variation (CoV) was also calculated at each responsive wavelength for the reflectance values across all the thrips damage classes. The coefficient of variation is the ratio of the standard deviation to the mean reflectance (Patel *et al*., 2001; SPSS, 2006).

6.3.3.2 Canonical discriminant analysis

A multiple discriminant analysis, also termed "canonical discriminant analysis", was applied over the spectral data set to achieve multigroup maximum separability and dimensionality reduction. The technique involves deriving linear combinations of two or more variables that discriminate best among the groups defined *a priori* (Johnson and Wichern, 2002; Hair *et al*., 2006). Discrimination is achieved by calculating the weight for each independent variable to maximise the variance between the groups, relative to the within-group variance. The procedure generates a discriminant function (or, for more than two groups, a set of discriminant functions) consisting of linear combinations of the predictor variables that provide the best discrimination between the groups (SPSS, 2006). The spectral variables considered in this procedure included: (1) the primary relative reflectance values (*R*) and the first-order derivative of reflectance at the blue (B; 478 nm), green (G; 569 nm), red (R; 671 nm), near infrared (NIR; 864 nm), shortwave infrared-1 (SWIR-1; 1649 nm), and shortwave infrared-2 (SWIR-2; 2204 nm) regions of the electromagnetic spectrum as they represent the main spectral changes when light interacts with leaves (Kumar *et al*., 2003); and (2) a set of spectral vegetation indices (Table 6.1) potentially sensitive to changes in chlorophyll, leaf water, and lignin-cellulose, as reported in the literature. A stepwise method of Wilks' Lambda was performed to select the best variables among the primary reflectance, first-order derivative of reflectance and the spectral vegetation indices. Wilks' Lambda is a ratio of the within-class sum of squares to the total sum of squares, and it ranges from zero to one. Values close to zero indicate strong group differences, and values close to one indicate non-significant differences. Probability levels of 0.05 and 0.1 were used as criteria to include or remove variables, respectively, from the stepwise procedure.

Table 6.1: Spectral vegetation indices used in the canonical discriminant analysis

Name	Formula*	Reference
1. Simple Ratio (SR) 750/705	R_{750}/R_{705}	Gitelson and Merzlyak (1994)
2. Modified simple ratio (MSR) 705/445	$(R_{750}-R_{445})/(R_{705}-R_{445})$	Sims and Gamon (2002)
3. Normalised difference (ND) 750/705	$(R_{750}-R_{705})/(R_{750}+R_{705})$	Gitelson and Merzlyak (1994)
4. Normalised difference (ND) 800/680	$(R_{800}-R_{680})/(R_{800}+R_{680})$	Sims and Gamon (2002)
5. Modified chlorophyll absorption in reflectance index (MCARI)	$[(R_{700}-R_{670})-0.2(R_{700}-R_{550})]$ (R_{700}/R_{670})	Daughtry *et al.* (2000)
6. Transformed chlorophyll absorption in reflectance index (TCARI)	$3[(R_{700}-R_{670})-0.2(R_{700}-R_{550})$ $(R_{700}/R_{670})]$	Haboudane *et al.* (2002)
7. Structure-insensitive pigment index (SIPI)	$(R_{800}-R_{445})/(R_{800}-R_{680})$	Peñuelas *et al.* (1995)
8. Photochemical reflectance index (PRI)	$(R_{531}-R_{570})/(R_{531}+R_{570})$	Gamon *et al.* (1992)
9. Pigment specific simple ratio (chlorophyll *a*) (PSSRa)	R_{800}/R_{680}	Blackburn 1998
10. Pigment specific simple ratio (chlorophyll *b*)	R_{800}/R_{635}	Blackburn (1998)
11. Disease–water stress index-1 (DWSI-1)	R_{800}/R_{1660}	Apan *et al.* (2004b)
12. Disease–water stress index-3 (DWSI-3)	R_{1660}/R_{680}	Apan *et al.* (2004b)
13. Moisture stress index (MSI)	R_{1600}/R_{820}	Hunt and Rock (1989)
14. Red edge index (REI)	$d_1(R_{720}- R_{690}) / \Delta$‡	Kumar *et al.* (2003)

* R_{750} = Reflectance at wavelength 750 nm, other R's are defined similarly

‡ d_1 = First-order derivative of reflectance and Δ the difference in wavelengths between 720 and 690

6.4. Results

The mean reflectance spectra of sugarcane leaves as affected by different levels of thrips damage are shown in Figure 6.1. These spectra can be regarded as typical vegetation spectra (see e.g., de Boer, 1993).

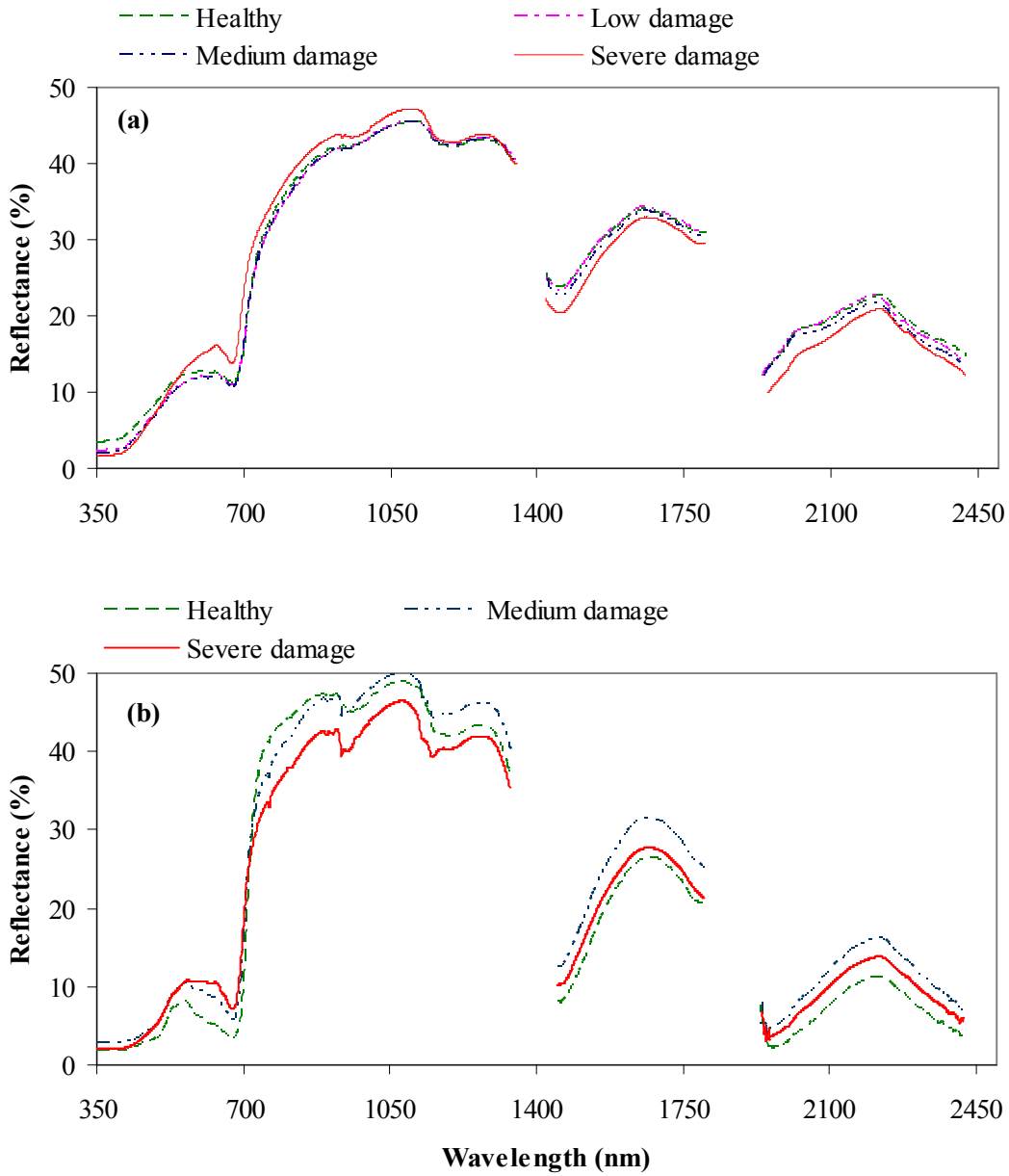

Figure 6.1. Average spectral characteristic (between 350 and 2500 nm) of sugarcane leaves with different thrips damage levels: (a) variety N19 and (b) variety N12. Spectral features between 1355 and 1450, 1800 and 1950, and 2421 and 2500 nm were removed due to excessive noise.

6.4.1 One-way ANOVA

Results of one-way ANOVA (Figure 6.2) show significant differences between the classes of thrips damage in some portions of the visible (400–700 nm) region of the spectrum and in particular in the so-called red edge region (690–720 nm) for both tested varieties. Variety N12 also shows significant differences in the middle infrared region of the spectrum, and specifically in some bands between 1954 nm and 1975 nm (Figure 6.2b).

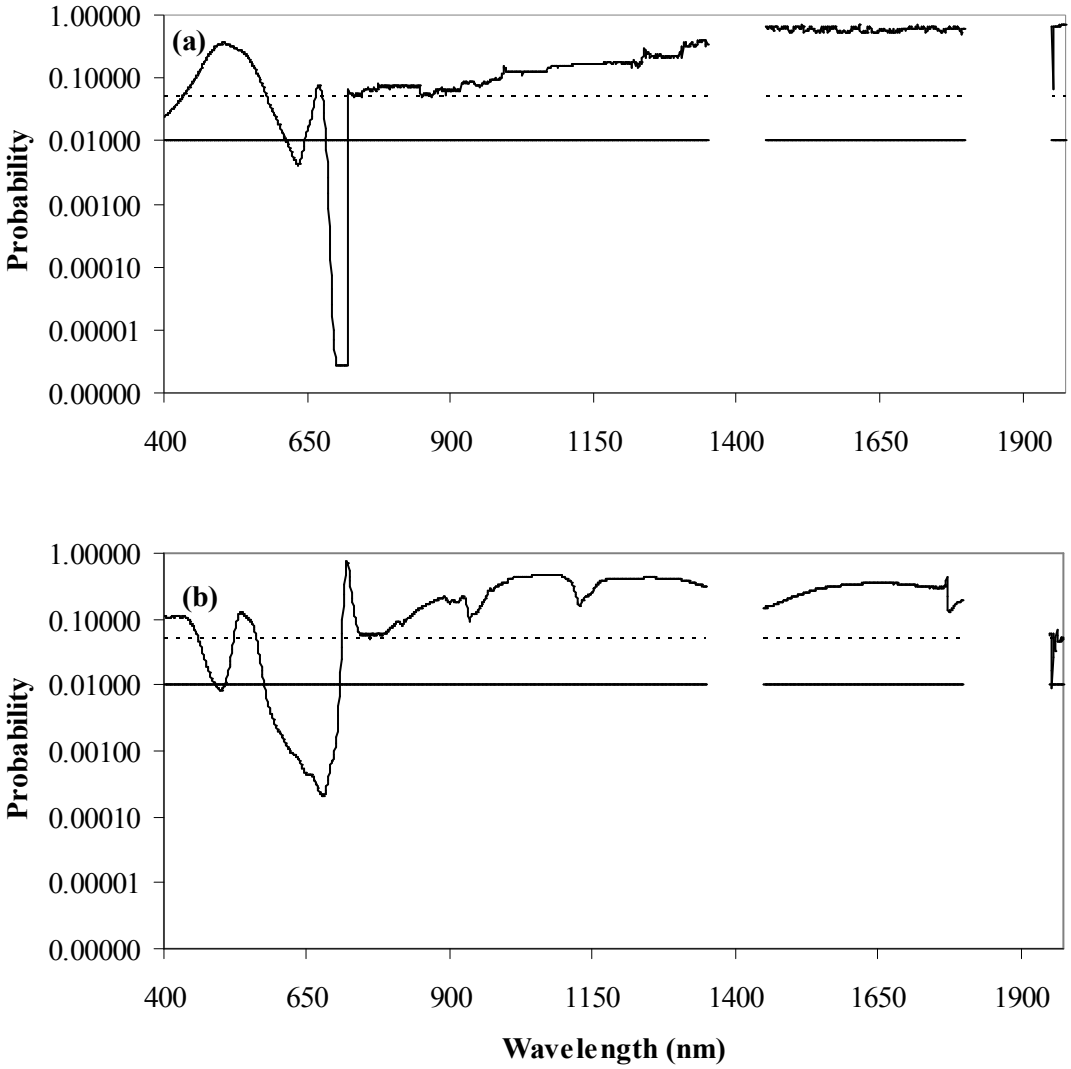

Figure 6.2. Results of one-way ANOVA showing level of significance of differences in reflectance between leaves with different levels of thrips damage at visible and middle infrared wavelengths. Spectral features between 1355 and 1450, and 1800 and 1950 were removed due to excessive noise. Horizontal solid and dashed lines show 99% and 95% confidence limits, respectively: (a) variety N19 and (b) variety N12.

The results of the Tukey test show that the healthy sugarcane leaves reflect significantly ($p < 0.05$) higher than severely damaged leaves for several bands in the blue (400–500 nm) region of the electromagnetic spectrum for variety N19 (Figures 6.2a, 6.3a, and Table 6.2a); but reverse results are shown for variety N12 (Figures 6.2b, 6.3b, and Table 6.2b). In the green region (500–660 nm) of the electromagnetic spectrum, variety N19 shows no significant differences between thrips damage classes in most bands; however, towards and within the red region the severely damaged leaves reflect significantly ($p < 0.05$) higher than medium damaged leaves (Figures 6.2a, 6.3a, and Table 6.2a). Variety N12 does show significant differences at some bands within the green region (Figure 6.2b), in all cases due to higher reflectance by severely damaged leaves than by medium and healthy leaves (Figure 6.1b). The most outstanding differences for both varieties are presented in the red (660–700 nm) and red edge of the spectrum, where the severely damaged leaves reflect significantly higher than low, medium and healthy leaves (Figures 6.2b, 6.3b, and Table 6.2b).

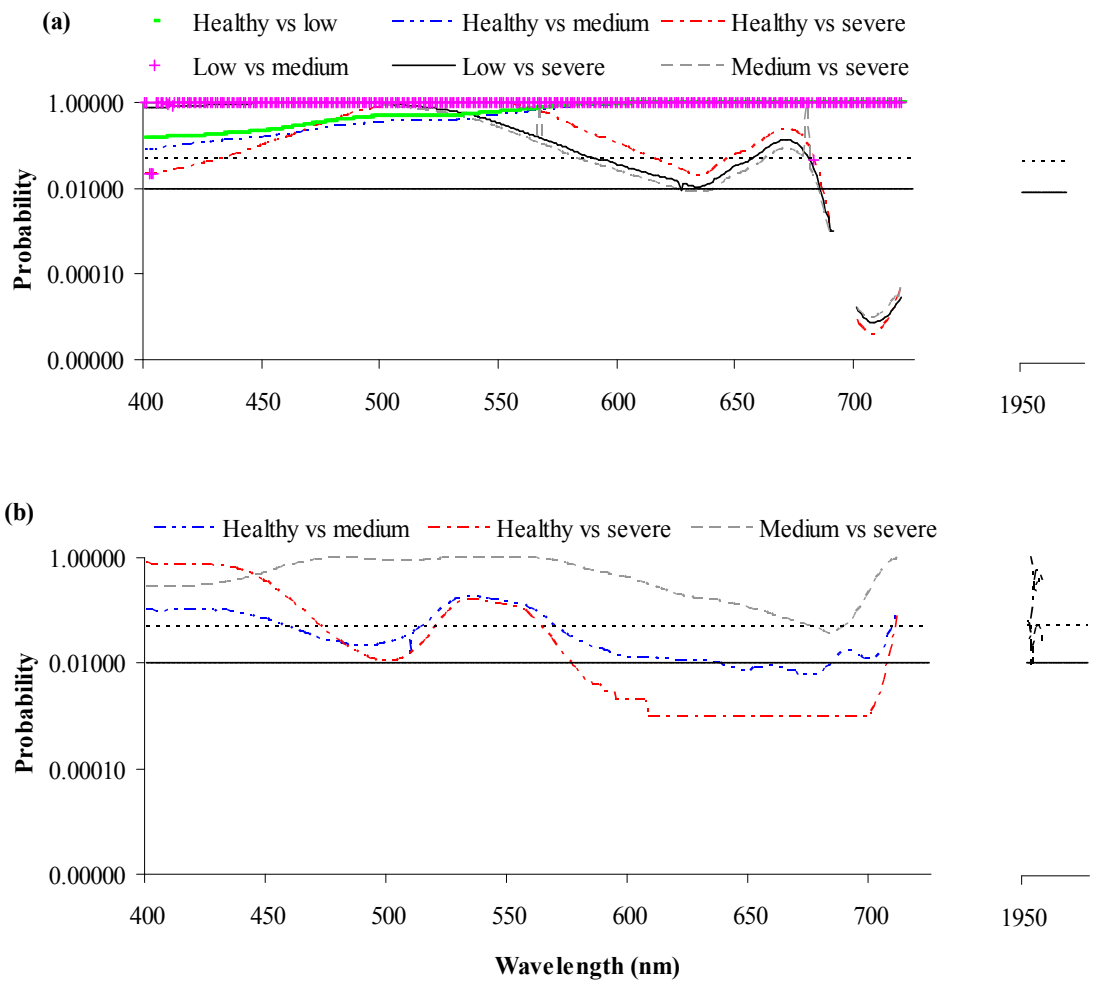

Figure 6.3. Results of Tukey test showing levels of significance of differences between reflectance of leaves with different levels of thrips damage (healthy, low, medium and severe) at visible, red edge and some middle infrared wavelengths. Spectral features between 1355 and 1450, and 1800 and 1950 were removed due to excessive noise. Horizontal solid and dashed lines show 99% and 95% confidence limits, respectively: (a) variety N19 and (b) variety N12.

6.4.2 Sensitivity analysis

Figures 6.4 and 6.5 show the results of the sensitivity analyses. The black arrows in these figures show the most responsive wavelengths to thrips damage. Reflectance difference maxima and minima located at 407 and 700 nm for N19 and at 514, 636, and 700 nm for N12 as well as reflectance sensitivity maxima and minima located at 400, 635, 680, and 697 nm for N19 and at 465, 502, 675, 693, and 1955 nm for N12 were found to be most

strongly affected by thrips damage. The coefficients of variation for those wavelengths are shown in Table 6.3. Higher coefficient of variation results in greater variability and discriminatory power of the particular wavelength.

Table 6.2: Frequency table of the wavelengths at regions described by Kumar *et al*. (2003) where reflectance differences between different severity scales (healthy, low, medium, and severe) of thrips are significant: (a) variety N19 and (b) variety N12

(a)

Wavelength (nm)	Description	Number of bands	H-L	H-M	H-S	L-M	L-S	M-S
400–500	Blue	32	0	0	32	0	0	0
500–660	Green	81	0	0	32	0	57	81
660–700	Red	26	0	0	18	0	20	21
690–720	Red edge	11	0	0	11	0	11	11
700–1300	Near infrared	0	0	0	0	0	0	0
1300–2500	Middle infrared	0	0	0	0	0	0	0

(b)

Wavelength (nm)	Description	Number of bands	H-M	H-S	M-S
400–500	Blue	40	40	28	0
500–660	Green	119	107	115	0
660–700	Red	39	42	40	0
690–720	Red edge	11	11	11	11
700–1300	Near infrared	0	0	0	0
1300–2500	Middle infrared	18	18	2	0

H = Healthy, L = Low, M = Medium, S = Severe

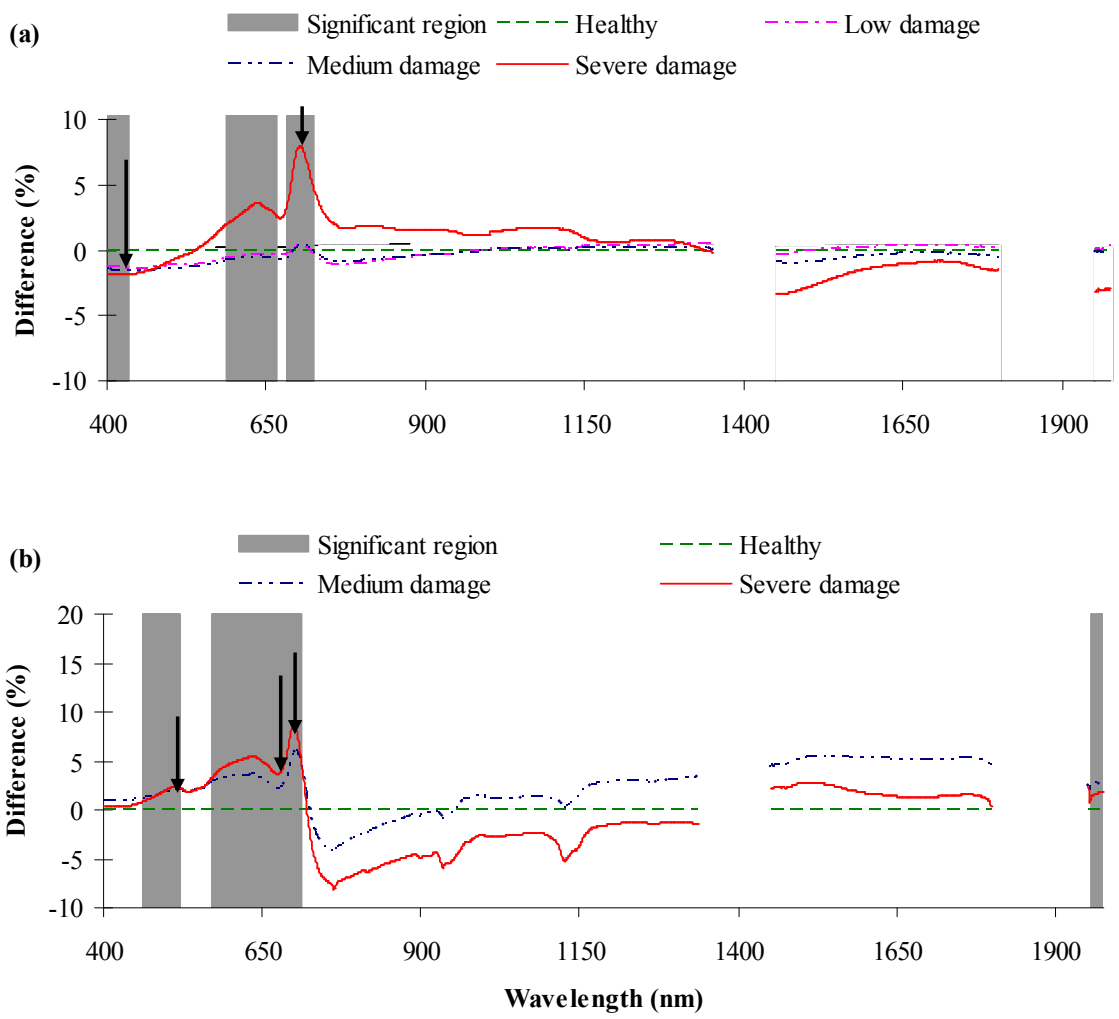

Figure 6.4. Results of sensitivity analysis, the reflectance difference was computed by subtracting mean reflectance of healthy leaves from that of low, medium and severe damaged leaves. Spectral features between 1355 and 1450, and 1800 and 1950 were removed due to excessive noise: (a) variety N19 and (b) variety N12.

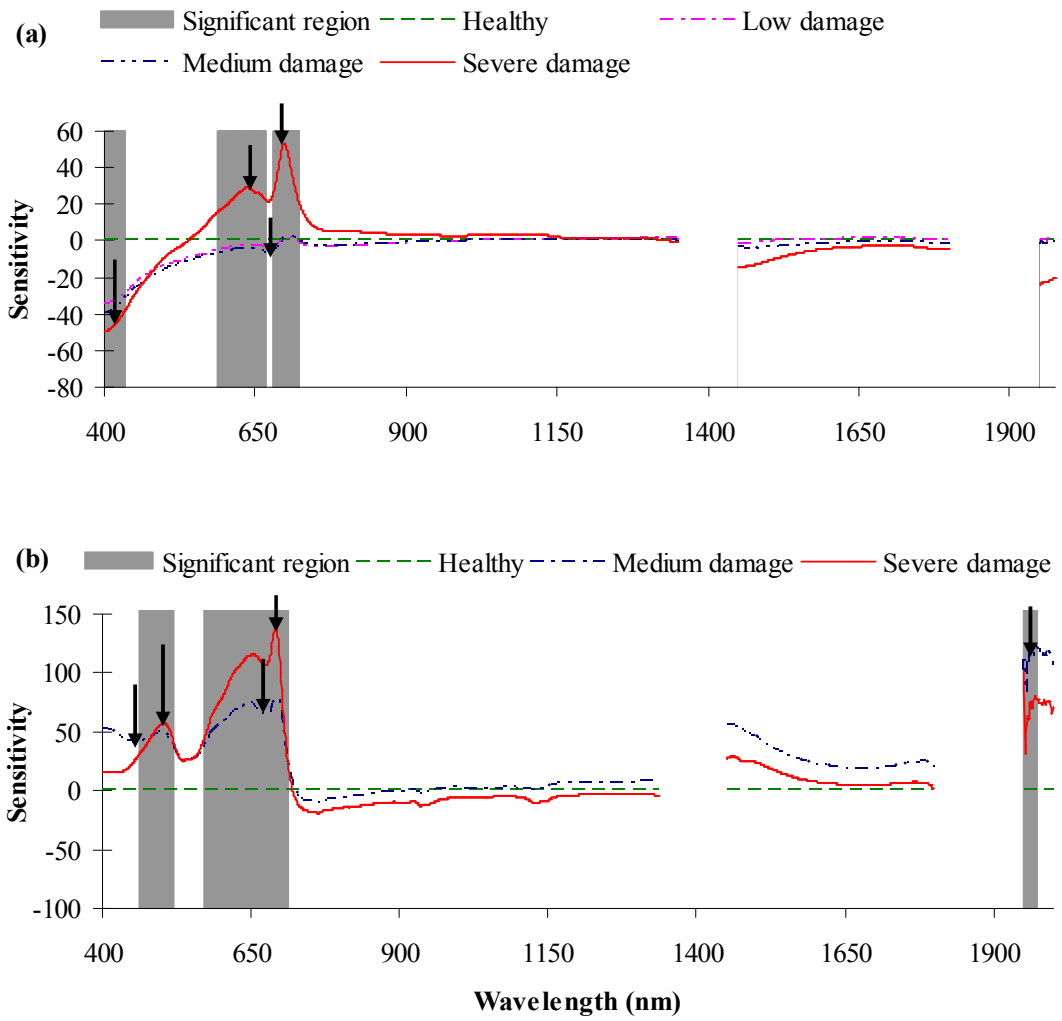

Figure 6.5. Results of sensitivity analysis, the reflectance sensitivity to damage was computed by dividing the reflectance difference by mean reflectance of healthy leaves. Spectral features between 1355 and 1450 and 1800 and 1950 were removed due to excessive noise: (a) variety N19 and (b) variety N12.

Table 6.3: Coefficient of variation (CoV) of the selected responsive wavelengths to thrips damage

Variety N19			Variety N12		
Wavelength (nm)	Description	CoV	Wavelength (nm)	Description	CoV
400	Blue	44.29	465	Blue	20.16
407	Blue	43.97	502	Green	23.25
635	Green	18.60	514	Green	21.68
680	Red	20.78	636	Green	31.69
697	Red edge	23.05	675	Red	30.39
700	Red edge	22.08	693	Red edge	36.21
			700	Red edge	28.80
			1955	Middle infrared	34.75

6.4.3 Canonical discriminant analysis

The results of the canonical discriminant analysis showed that the overall discrimination between the thrips damage classes was highly significant ($p < 0.001$) for both varieties. For N19, six spectral variables: the red edge index (REI), simple ratio (SR), modified simple ratio (MSR), photochemical reflectance index (PRI), and the first-order derivative of reflectance at NIR, and SWIR-2 regions were selected to compose three functions of the canonical discriminant model (Table 6.4: Functions 1a, 2a, and 3). For N12, four variables *viz*., REI, SR, DWSI-1 (Disease−water stress index-1), and relative reflectance at SWIR-2 were chosen to build two discriminant functions (Table 6.4: Functions 1b and 2b). For both varieties, 90% or more of the variance was explained by the first discriminant function, and in both cases, the REI had the highest factor structure coefficient (Table 6.4). Table 6.5 shows the means of the canonical scores, reflecting the nature of each discriminant function. The results in this table indicate that the first function discriminated mostly between healthy leaves and leaves of the other damage classes for both varieties, whereas the second function discriminated mostly between the severely damaged leaves and leaves of the other damage classes for N19 and between

medium damaged leaves and leaves of other damage levels for N12. The third function of N19 differentiated between the medium damaged leaves and leaves of the other damage scales.

Table 6.4: Coefficient matrix, representing the discriminant functions. See Table 1 for the meaning of the abbreviations used

Variety 19				**Variety N12**		
Variable	Function			Variable	Function	
	1a	2a	3		1b	2b
REI	0.758	–0.025	0.239	REI	0.806	–0.429
d_1(SWIR-2)	–0.075	0.552	–0.360	SR	0.644	0.408
SR	0.298	0.522	0.297	DWSI-1	0.254	0.556
MSR	0.387	0.500	–0.199	SWIR-2	–0.116	–0.537
PRI	0.460	0.478	0.170			
d_1(NIR)	0.021	0.415	0.741			

Table 6.5: Means of canonical scores of each discriminant function, for each of the damage classes of the two varieties

Variety 19				**Variety N12**		
Damage class	Function			Damage class	Function	
	1a	2a	3		1b	2b
Healthy	**3.376**	–0.605	0.001	Healthy	**3.790**	0.222
Low	–0.248	0.740	0.186	Medium	–0.908	**–0.749**
Medium	–0.266	0.709	**–0.190**	Severe	–2.882	0.527
Severe	–2.862	**–0.844**	0.003			

Figure 6.6 shows the scores for the two canonical discriminant functions and group centroids. The positioning of the canonical scores showed a gradient from the severe damage class, with low and medium damage class, to the healthy class. On the other hand, the results indicated the difficulty of discriminating between the low and medium damage classes.

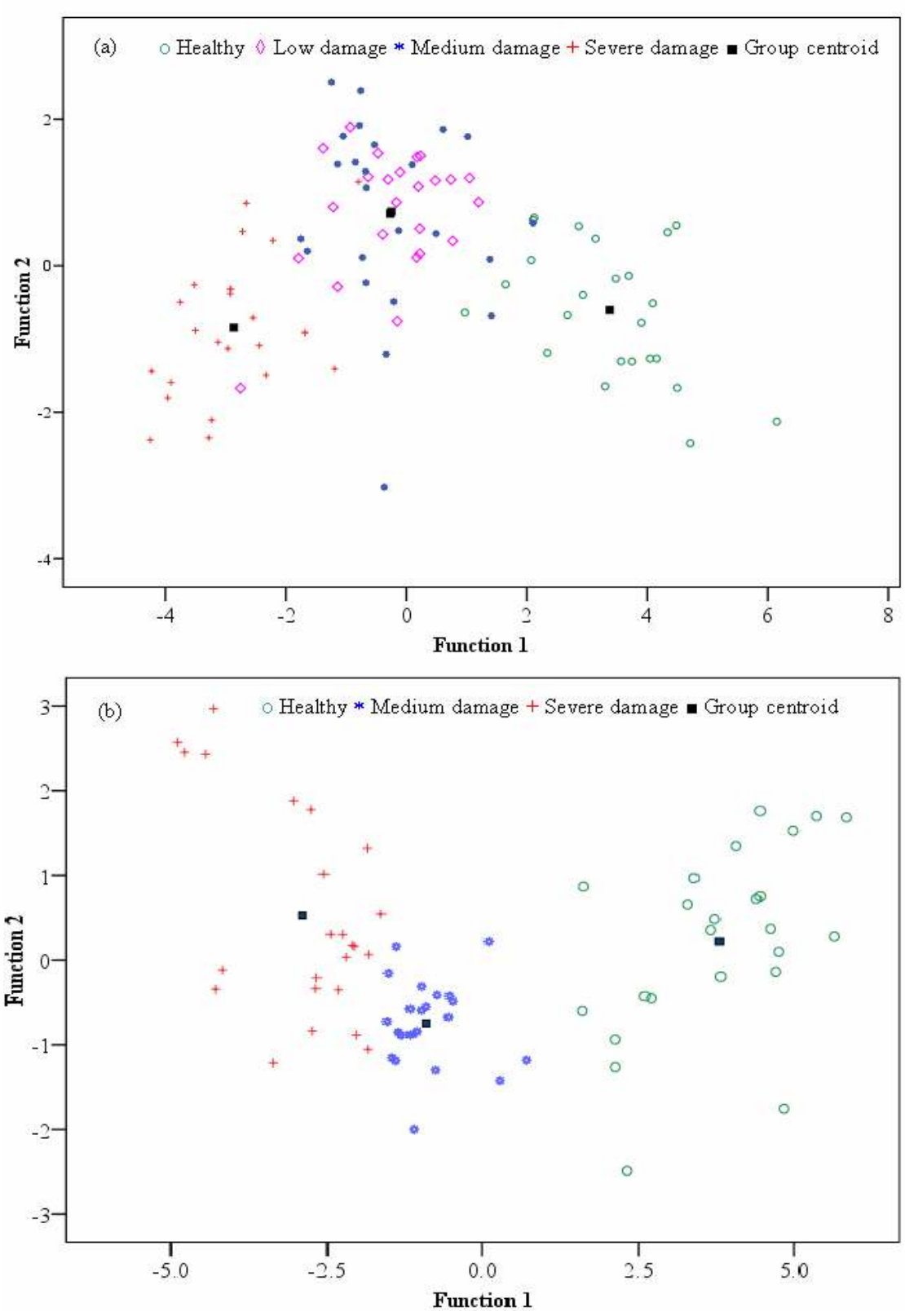

Figure 6.6. Scatter plot of canonical scores of the first two discriminant functions: (a) variety N19 and (b) variety N12.

6.5 Discussion

The motivation for this study was to investigate the potential of detecting spectral reflectance differences to discriminate between sugarcane thrips-damaged leaves of two varieties.

The results show that differences in thrips-damaged sugarcane leaves are associated with significant differences in spectral reflectance. Within the blue region of the spectrum (400–500 nm), the statistical differences between thrips damage classes of variety N12 could be related to the blue peak overlaps with the absorption of caretenoids (Sims and Gamon, 2002). Furthermore, thrips-damaged leaves are accompanied by a breakdown of chlorophyll, which would have caused reduced absorption in the blue region of the electromagnetic spectrum (de Boer, 1993). The unexpected results of relatively higher reflectance in this region associated with healthy leaves for variety N19 (Figure 6.1a) could possibly be explained by the slight differences in the crop age (3–4 months) and growth conditions of sample fields from which the analysed leaves were collected. These differences could affect the measured spectral reflectance of the leaves.

In the green region (500–660 nm), the relatively higher reflectance of healthy leaves compared to the medium and severely damaged leaves of variety N19 is expected since the green region is characterised by relatively higher reflectance due to chlorophyll content (Kumar *et al.*, 2003; Coops *et al.*, 2006; Jiao *et al.*, 2006). For variety N12, the study shows unexpected results as reflectance from sugarcane leaves increased with increasing severity of thrips damage (high reflectance with severe damage and low with healthy) in the green region, which could not be well explained.

However, the most outstanding differences were found in the red region (660–700 nm), where for both varieties, the reflectance increased with the increasing severity of thrips damage (high reflectance with severe damage and low with healthy), presumably manifesting the very strong absorption by chlorophyll in this region of the

electromagnetic spectrum (Malthus and Madeira, 1993; Kumar *et al.*, 2003; Mutanga *et al*., 2003; Kiang *et al*., 2007; Cho, 2007).

While the present study acknowledges major differences, these findings are also in conformity with the findings of Lorenzen and Jensen (1989) who showed that the detection of infected barely leaves by powdery mildew disease was more reliable in the visible region and the red portion was among the most responsive variables because of chlorophyll deficiency caused by the disease. Moreover, most of the vegetation indices (SR, MSR, and PRI) that were selected as a result of the canonical discriminant functions are known indices for chlorophyll estimation (Gamon *et al*., 1992; Gitelson and Merzlyak, 1994; Sims and Gamon, 2002). However, while plants damaged by pests and diseases often contain lower chlorophyll levels, Kolb *et al*. (1991) found that foliar chlorophyll content (mg cm^2) of thrips-damaged seedlings of sugar maples was the same as that of undamaged seedlings, because the feeding punctures by thrips in the leaves influenced water content more than chlorophyll content. Thrips-damaged leaves significantly increased the total reflectance in the visible region (Figure 6.1), which is in conformity with the findings of Riedell and Blackmer (1999) who reported that wheat leaves infested by aphids reflected significantly higher in the visible region as opposed to non-infested leaves.

Thrips-damaged leaves for both varieties were accompanied by a shift in the inflection point of red edge position (690–720 nm). The results show highly significant differences between thrips damage classes in this region. The high levels of discrimination provided by selecting REI for both varieties reinforce the point of the importance of the red edge in discriminating among thrips damage classes. This is also in agreement with the findings of Riedell and Blackmer (1999) who highlighted that some wavebands around the red edge portion were well related to chlorophyll deficiency caused by greenbug and Russian aphid on wheat leaves. On the other hand, Malthus and Madeira (1993) and Apan *et al*. (2004b) found that the red edge region was relatively poor in discriminating between leaves infected by *Botrytis fabae* and non-infected ones and between sugarcane plantations damaged by rust disease from non-diseased, respectively. The red edge is a

known means for estimating foliar chlorophyll or nitrogen content (Clevers and Jongschaap, 2003; Kumar *at al*., 2003; Hecker, 2003; Mutanga, 2004; Cho and Skidmore, 2006; Nguyen and Lee, 2006; Cho, 2007). Since there is strong nitrogen-chlorophyll relationship (Yoder and Pettigrew-Crosby, 1995; Nguyen and Lee, 2006), the significant high reflectance of the severely damaged leaves in the red edge position could be due to nitrogen deficiency because of thrips damage.

This study showed significant results in some middle infrared bands in discriminating severely damaged leaves from healthy leaves of variety N12, specifically bands 1954 and 1955 nm. In addition, other wavelengths centred at NIR, and SWIR-2, and DWSI-1 were among the selected variables of canonical discriminant models to discriminate among thrips damage classes. All these spectral variables are sensitive to moisture content (de Boer, 1993; Kumar *et al*., 2003; Apan *et al*., 2004b). Their selection could be related to loss of moisture in sugarcane leaves due to lesions or ruptures caused by thrips (Way *et al*., 2006). This is comparable to the results of Xu *et al*. (2007) who indicated that some middle wavelengths were highly correlated with severity of leaf miner damage on tomato crops because of water loss due to collapse of the tomato leaves' cell structure. However, in the present study the leaf samples were stored for a bit long time (a month) before the analysis and then kept indoors for a while to allow the leaves to be free from any water droplets before the spectral measurements were taken. That might have also caused moisture loss in the leaves. Future studies should take these points into consideration.

The relative positioning of the thrips damage classes in canonical space (Figure 6.6) provided an insight into the relationship among the damage classes. While the thrips damage classes were positioned according to their gradient of thrips damage, the study showed that it is difficult to separate the low damage class from the medium damage class. A similar challenge was found by Sudbrink *et al*. (2003) who reported difficulty in identifying visual cabbage looper damage on cotton crops using spectroscopic data before the severe damage stage. However, the difficulty of discriminating between low and medium damage classes might be due to the bias (if any) that may occur when the low

and medium classes were visually categorised. In other words there were no thresholds (e.g., percent of damage on the leaves) drawn between damage classes and that might have made the separation between the low and medium classes relatively inaccurate.

Obviously, many of the discrepancies between the results of the current study and other findings may be attributed to the fact that different crops and/or pests or diseases present different damage patterns and mechanisms; whereas concurrence does not necessarily indicate similar cause-effect relations. The author believes that this study contributes to a better understanding of the relations between thrips damage and spectral characteristics of sugarcane leaves, but further research is needed to investigate how spectral characteristics are affected at canopy level.

6.6 Conclusions

This research has shown that there were detectable changes in spectral reflectance over the visible and middle infrared regions in thrips-damaged leaves. The red edge position within the visible region shows the highest possibility of detecting sugarcane thrips damage. The higher reflectance in the middle infrared region of the spectrum in thrips-damaged leaves could be related to the collapse in sugarcane leaf tissues and loss of leaf moisture due to thrips damage. Although no single spectral variable has been selected to discriminate among all thrips damage classes, relative reflectance at wavelengths located at 697 and 700 nm (for N19) and at 636, 675, 693, 700, and 1955 nm (for N12) as well as the red edge index have the greatest discriminatory power. Both varieties showed very similar results in the red edge region where the responsive bands for N19 and N12 (697 and 693 nm, respectively) are 3 nm apart, while 700 nm was a responsive band for both varieties. Moreover, the red edge index was the selected variable as a result of discriminant analysis for both varieties.

The results demonstrate the potential of extending the investigation to canopy level using airborne or spaceborne hyperspectral imaging for detecting damage or biophysiological changes in sugarcane caused by sugarcane thrips.

Acknowledgments

The candidate thanks SASRI extension specialist, Mr Marius Adendorff, for providing the leaf samples. Thanks to SASRI's thrips survey team and technicians for their keen help in leaf samples storage. Many thanks to Mr. Franko Sokolic for his assistance in the configuration of the field spectroradiometer. Thanks go to the anonymous reviewers for their useful comments.

CHAPTER SEVEN

Estimation of thrips (*Fulmekiola serrata* Kobus) incidence in sugarcane using leaf level spectroscopic data

This chapter is based on

1. **Abdel-Rahman**, E. M., van den Berg, M., Way, M. J., Ahmed, F. B. and Sewpersad, C., 2009. Using spectroscopic data sets to predict numbers of thrips (*Fulmekiola serrata*) in sugarcane. *Proceedings of South African Sugar Technologists' Association*, 82, 441– 445.
2. **Abdel-Rahman**, E. M., van den Berg, M. Way, M. J. and Ahmed, F. B., 2009. Handheld spectrometry for estimating thrips (*Fulmekiola serrata*) incidence in sugarcane. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, In Press.

Abstract

A handheld FieldSpec® 3 spectroradiometer was employed at leaf level in the 350–2500 nm range of the electromagnetic spectrum to develop a monitoring technique for the serious sugarcane thrips pest (*Fulmekiola serrata* Kobus (Thysanoptera: Thripidae). Sugarcane leaf samples from variety N19 of 4–5 and 6–7 months were subjected to spectral analysis, and then nymphs and adults thrips were counted in the samples. Leaf reflectance measurements were transformed to first-order derivative of reflectance and analysed to evaluate estimation of the pest populations. Partial least squares (PLS) regression was used for the analysis. Predictions were adequate for nymph numbers in the younger (4–5 months) crops during summer (December) period (R^2 = 0.75 and Root Mean Square Error of Prediction: RMSEP = 1.25; 47.3% of the mean). During autumn (March) period, the correlations were high *viz*., R^2 = 0.92 for estimating populations of adults (RMSEP = 0.5; 45%) and 0.86 for nymphs and adults combined (RMSEP = 0.68; 40% of the mean) in the younger crops. Estimating thrips abundance for the older sugarcane crops (6–7 months) was less accurate. It was inferred that counts of sugarcane thrips can be estimated with spectroscopy for younger sugarcane crops at leaf level. Models differ according to thrips life stage assessed and the season at which the estimation takes place.

Keywords: sugarcane thrips; incidence; monitoring; spectroscopy; PLS

7.1 Introduction

Sugarcane thrips, *Fulmekiola serrata* Kobus (Thysanoptera: Thripidae) is a recent pest in the South African sugarcane, and it was positively identified in 2004 (Way *et al.*, 2006), It is now widely spread throughout the sugarcane growing regions in the country (Keeping *et al.*, 2008). Leaf necrosis caused by sugarcane thrips occurs due to puncturing of the leaf surface (Kirl, 1997 in Way *et al.*, 2006). Younger crops tend to present higher thrips populations (Keeping *et al.*, 2008). Thrips activity in younger sugarcane crops causes leaf tips to become tied or bound together. Injury symptoms on the open leaf blades are often confused with heat scorch, nutrient deficiency, water stress and/or herbicide damage (Leslie and Donaldson, 2005)**.**

The monitoring and control of thrips is of paramount importance but relatively expensive, subjective, and labour-intensive. Complementary methods that are relatively accurate, potentially inexpensive, and capable of proving up-to-date information on the pest are investigated.

Remote sensing has the potential to provide cost-effective and repetitive techniques for monitoring and detecting insect pest outbreaks (Mirik *et al.*, 2007). Spectroscopic (hyperspectral) data are characterised by many narrow and continuous wavelengths (Lillesand and Kiefer, 2001). In order to predict any feature of interest from these large sets of wavebands, one needs to collect many observations (samples) as possible to avoid expected overfitting problem. However, there are many logistics and constraints that make it difficult to collect many such observations. Therefore, researchers seek techniques and methods that could be used to reduce the redundancy and co-linearity in the complex hyperspectral data without losing relevant information to the features of interest. The random forest algorithm is a relatively new method that has been used by some authors (e.g., Chan and Paelinkx, 2008; Ismail, 2009) for such a purpose.

The random forest ensemble, a machine learning algorithm, was first developed by Breiman (2001). The algorithm employs recursive partitioning to grow many trees (*ntree*)

and then averages the results. Each tree is independently grown to maximum size based on a bootstrap sample from the training data set without any pruning. In a random forest algorithm, each node is split using the best among a subset of predictors, randomly selected at that node (*mtry*) (Breiman, 2001). The ensemble predicts data that are not in the trees (out-of-bag; OOB data) and by averaging OOB predictions from all trees, random forest algorithm gives an error of prediction called the OOB prediction error for each variable (Breiman, 2001; Maindonald and Braun, 2006). Random forest algorithm produces a measure of importance of the variables by comparing how much the OOB prediction error increases when a variable is removed, whilst all others are left unchanged (Archer and Kimes, 2008). One drawback of the random forest algorithm in selecting variables from the spectroscopic data is that the selected relevant wavebands could still be auto-correlated (Strobl *et al.*, 2008), especially with the very fine spectral resolutions produced by handheld sensors.

Partial least squares (PLS) regression (Wold, 1995) overcomes the co-linearity problems (Huang *et al.*, 2004) that might be encountered with the selected narrow spectral bands. One shortcoming with PLS regression in spectroscopic data analysis is the identification of the spectral region(s) that have relatively higher influence on PLS regression models (Huang *et al.*, 2004). Martin *et al.* (2008) recommended that the PLS regression coefficients can be normalised by the average spectral reflectance at all input wavelengths. Spectral regions that show high values of normalised coefficients indicate the influence of such regions on the calibrated PLS regression models.

7.2 Applications of remote sensing for monitoring the incidence of crops pests

The use of spectroscopy for monitoring insect pests damage and diseases in some field crops is given by Abdel-Rahman *et al.* (2009c). For monitoring of insect incidence, Yang *et al.* (2005) used a multispectral handheld radiometer to predict greenbug density in wheat canopies. The authors found that reflectance from bands centred at 694 and 630 nm as well as a ratio vegetation index (RVI) were the most sensitive spectral features to detect greenbug infestation, with correlation coefficients (r) between the greenbug

density and reflectance at 694 nm, 630 nm, and RVI exceeding 0.90. Mirik *et al.* (2006b) found that damage sensitive spectral index$_2$ ($DSSI_2$) generated from hyperspectral data exhibited a high correlation ($R^2 = 0.76$) with aphid density in wheat canopies. A slightly lower correlation ($R^2 = 0.67$) was found when $DSSI_2$ was derived from multispectral data. On the other hand, Mirik *et al.* (2007) reported that the structure insensitive pigment index (SIPI) explained 91% of variance in density of Russian wheat aphid whereas normalised difference vegetation index (NDVI) showed no correlation. Genc *et al.* (2008) investigated the feasibility of using hyperspectral data to detect three life stages of sunn pest (*Eurygaster integriceps* Put.) densities in wheat canopies. It was found that correlations (R^2) between the densities of the three life stages and narrow-band NDVI values were higher than 0.899. For cotton crops, Sudbrink *et al.* (2003) analysed reflectance data at leaf level as well as from an aerial image (at canopy level) to identify any association between larval densities of beet armyworm (BAW) and narrow-band NDVI values. Their results revealed a high and significant negative correlation between BAW hits and NDVI values on two different dates ($r = –0.386$ and $–0.77$).

In a study to investigate the potential of the spectroscopy for detecting thrips damage in South African sugarcane, Abdel-Rahman *et al.* (2008b; 2009c) found that the reflectance of severe and medium thrips-damaged leaves was significantly low and, they could easily be discriminated from healthy ones in the visible region for N19 and N12 varieties. There is, however, no information as to whether spectroscopic measurements on sugarcane leaves can be correlated with sugarcane thrips numbers to provide reliable strategies based on early detection of thrips incidence before much damage is done. Hence, the objective of this study was to investigate the potential use of leaf level spectroscopic data in predicting the incidence of sugarcane thrips *Fulmekiola serrata* (Kobus).

7.3 Materials and methods

7.3.1 Leaf sample collection

Leaf spindles were collected from 4–5 and 6–7 month old commercial sugarcane crops, of variety N19 during two periods (Table 7.1). Variety N19 was selected because it is the most common variety in the study area, representing about 39% of the total crop (South African Sugar Association: SASA, 2007, unpublished industry database). Randomly, 2–3 samples (five leaf spindles each) were collected from each field from spots of about a 5 metre radius each. Table 7.1 shows total numbers of sampled fields and samples collected. Leaf samples were placed in plastic bags and stored in a fridge (at ca +5 ^{o}C) for five days until they were taken to the laboratory, where they were stored at –18 ^{o}C for a week before spectral measurements were taken.

Table 7.1: Number of sample fields and samples collected during the two sampling periods for the two assessed age groups

	December 2007 (summer)		**March 2008 (autumn)**	
	4–5 month age group	6–7 month age group	4–5 month age group	6–7 month age group
Number of fields	13	8	7	8
Number of samples	27	16	20	22

7.3.2 Leaf spectral measurements

Before taking spectral measurements, leaf samples were exposed to indoor laboratory conditions for two hours to allow free water droplets to evaporate. A FieldSpec® 3 spectroradiometer (Analytical Spectral Device: ASD, 2005) was used to collect relative reflectance data from each leaf in the samples. The ASD is a compact, non-imaging, battery powered portable spectrometer, which uses a fibre optic cable for light collection and a notebook computer for data logging. The spectral range is 350–2500 nm with a

resolution of 1.4 nm in 350–1000 nm range and 2 nm in 1000–2500 nm range and a data collection time of 0.1 second per spectrum. Spectral measurements were carried out under room temperature (23–25 ^{o}C). A fibre optic cable with 1^{o} field of view was directed 10 cm above the sample which was laid on a black platform with zero reflectance. A 50 watt halogen lamp was used (40 cm away and oriented at 45° from the sample) as the only light source. All spectral measurements were made relative to a spectralon white reference panel, which was measured before and after the spectral measurements of sugarcane leaves. For each measurement, 20 scans were performed by ASD, and the spectral readings were averaged. The bottom, middle, and top upper surface of each leaf on both sides of midrib was measured so that the spectrum of a sample (5 leaf spindles) was an average of 30 spectra.

7.3.3 Determination of thrips numbers

Based on Way (2008), the process involved gently separating the individual leaf spindles and rinsing them, while agitating them in warm salty water to facilitate the removal of insects from between folds of the leaf spindles. Thrips were filtered through fine muslin cloth (1–2 mm) that was spread onto a laminated grid for counting adult thrips and nymphs under a microscope.

7.3.4 Data analysis

7.3.4.1 Spectral transformation

Primary sugarcane leaf spectra were transformed to their first-order derivatives as suggested by Sims and Gamon (2002), Zhao *et al.* (2005), Thomas *et al.* (2007), and Abdel-Rahman *et al.* (2008a) to reduce the effects of multiple scattering of radiation due to sample geometry, surface roughness etc. and to locate positions of absorption features and inflection points on the spectra. In addition, derivative spectra may swing with greater amplitude than primary spectra. Derivatives are also useful for separating out peaks of overlapping wavebands (de Jong, 1998; ASD, 2005).

7.3.4.2 Statistical analyses

A so-called random forest regression algorithm was first used to reduce the number of variables in the spectroscopic data set while preserving maximum relevant information for thrips incidence. Wavelengths in the 400–2500 nm range of the electromagnetic spectrum were used as input variables, except the regions between 1355 and 1450 nm, 1800 and 1950 nm, and 2420 and 2500 nm which are known noise features (Kumar *et al.* 2003; Mutanga *et al.* 2004; ASD 2005). The *randomforest* library (Liaw and Wiener, 2002) developed in the R package for statistical analysis (R Development Core Team, 2008) was employed to implement random forest regression algorithm. PLS regression was then used to predict the population of nymphs and adults individually as well as combined for each age group with the 2.5%, 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20%, 22.5%, and 25% most important wavelengths as ranked by the random forest algorithm. In the combined age groups data, the 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, and 50% most important wavelengths were examined. The objective was to allow not more than 25% and 50% of wavelengths to be included on PLS models for estimating thrips populations in each age group and the combined age group data set, respectively, in order to reduce the redundancy in the spectroscopic data.

The optimum number of spectral variables and the appropriate number of components (factors) that can be included in each PLS regression model were determined based on root mean square errors of prediction (RMSEP) values with a leave-one-out cross validation method. For each model, the first minimum RMSEP value was used to indicate a suitable number of components rather than the absolute minimum RMSEP value to avoid overfitting with so many components (Smith *et al.*, 2003; Nguyen *et al.*, 2006; Mevik and Wehrens, 2007). PLS models were calibrated to predict the number of nymphs and adults thrips individually as well as combined. Influence of each wavelength in PLS regression models was evaluated based on the method recommended by Martin *et al.* (2008).

7.3.4.3 Validation

A preliminary validation of each model was conducted using a leave-one-out cross validation method. R^2 and absolute as well as relative RMSEP values were recorded to test the predictive ability of the developed models. In addition, models that were generated from the data set collected during December 2007 were validated using the data collected during March 2008 to test the reliability of December models using independent test data.

7.4 Results

Figure 7.1 shows descriptive statistics for the number of nymphs, adults, as well as nymphs and adults combined. Samples that showed extreme values were identified as outliers and excluded from the analysis. Mean thrips numbers are shown in Table 7.2.

The importance of each wavelength in estimating thrips numbers as a result of applying the random forest algorithm is shown in Figures 7.2 and 7.3. The importance of a wavelength differed depending on thrips life stage, age group of sugarcane, and period of sampling (summer or autumn).

Based on RMSEP values, results showed that 2.5% of the most important wavelengths were suitable for predicting thrips numbers for each age group individually, and 5% for combined age groups data (Figures 7.4 and 7.5), but not for estimating nymph populations in younger (4–5 months) sugarcane crops during summer (December 2007), where 5% of the most important wavelengths yielded the lowest RMSEP value. Figure 7.6 shows the appropriate numbers of components (factors) depending on the RMSEP values.

Table 7.3 shows a summary of results for PLS regression models generated in this study. The highest R^2 value was 0.75 for estimating nymph and nymph+adult counts in

sugarcane crops of 4–5 months old during the summer period. During the autumn period, the highest R^2 was 0.92 for estimating adult numbers in the same younger age group.

The relative influence of the most important wavelengths in PLS models' calibration is shown in Figures 7.7 and 7.8. Results indicate high variation in the influence of spectral wavelengths on estimation of thrips numbers. Both positive and negative relationships between first-order derivative of reflectance and thrips numbers have been shown in the near infrared (NIR; 700–1300 nm) and middle infrared (MIR; 1300–2500 nm) regions of the spectrum. The localised peaks in these two spectral regions were labelled on Figures 7.7 and 7.8. In the visible region (VIS; 400–700 nm) there were very few wavebands at which there was slightly positive or negative correlation between first-order derivative of reflectance and thrips counts during the summer and autumn periods.

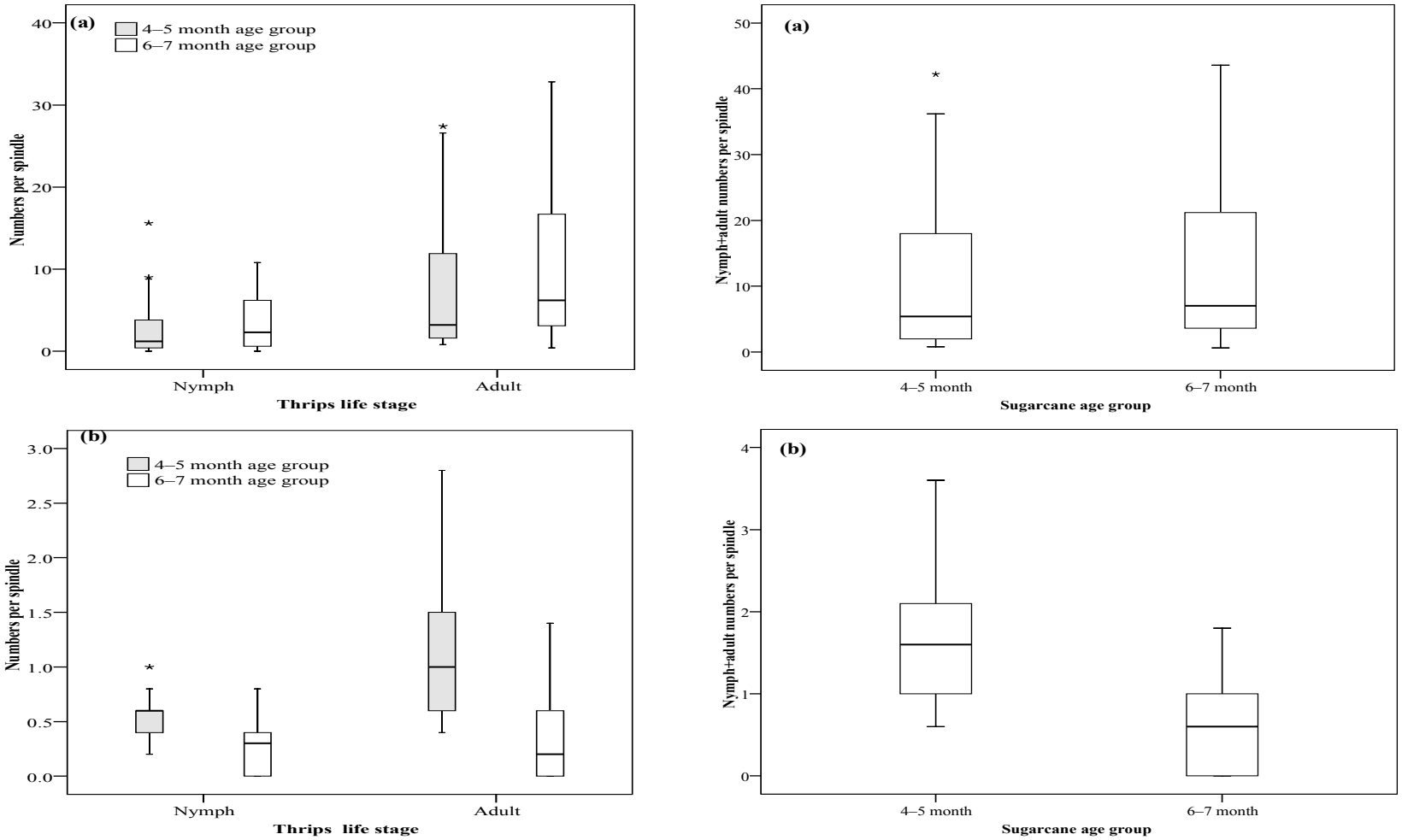

Figure 7.1. Nymph, adult, and nymph+adult numbers observed at the sample fields: (a) December 2007; (b) March 2008. The asterisks show the outliers.

Table 7.2: Mean number of nymphs, adults and nymphs+adults per spindle on the samples after removing outliers as shown in Figure 7.1

	December 2007 (midsummer)		**March 2008 (autumn)**	
	4–5 month age group	6–7 month age group	4–5 month age group	6–7 month age group
Nymphs	2.64	3.68	0.51	0.27
Adults	7. 31	10.56	1.19	0.37
Nymphs+adults combined	9.85	14.24	1.70	0.60

Leave-one-out cross validations for the models are given in Figures 7.9 and 7.10. Results of validating models that have been developed from the data set collected during summer (December 2007) using the data set collected during autumn (March 2008) as independent test data are shown in Figure 7.11. The lowest RMSEP value for predicting abundance of nymphs in the younger sugarcane crops (4–5 months) was 1.25 (47%) during summer and 0.15 (29.4%) during autumn (Figures 7.10 and 7.11).

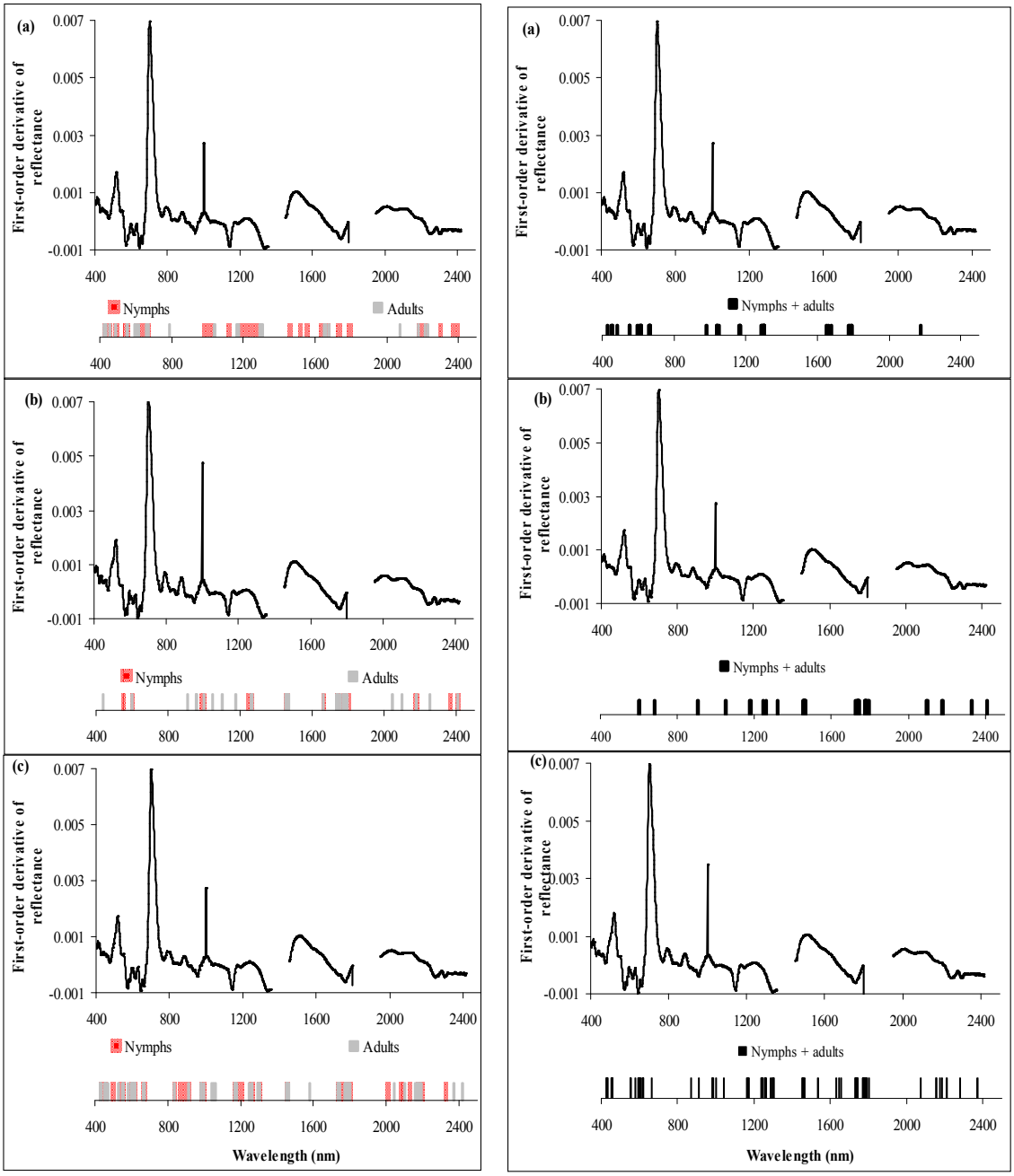

Figure 7.2. First-order derivative of reflectance for sugarcane leaves collected during December 2007. Spectral features between 1355 and 1450, 1800 and 1950, and 2420 and 2500 nm were removed due to excessive noise: (a) 4–5 month age group, (b) 6–7 month age group, and (c) combined age groups data. The bars show the most important spectral regions for predicting number of nymphs, adults, and nymphs+adults combined as a result of applying the random forest algorithm.

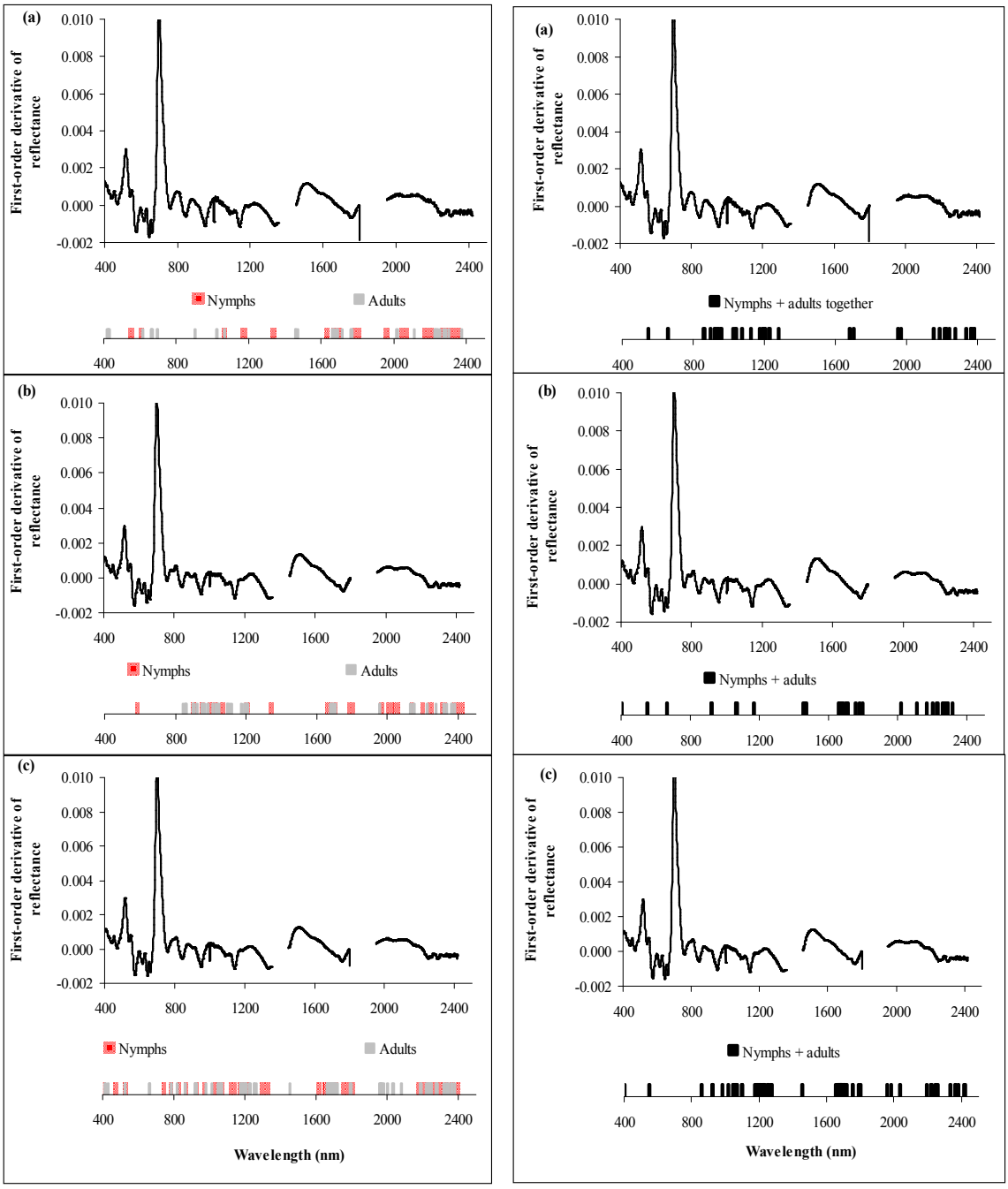

Figure 7.3. First-order derivative of reflectance for sugarcane leaves collected during March 2008. Spectral features between 1355 and 1450, 1800 and 1950, and 2420 and 2500 nm were removed due to excessive noise: (a) 4–5 month age group, (b) 6–7 month age group, and (c) combined age groups data. The bars show the most important spectral regions for predicting number of nymphs, adults, and nymphs+adults combined as a result of applying the random forest algorithm.

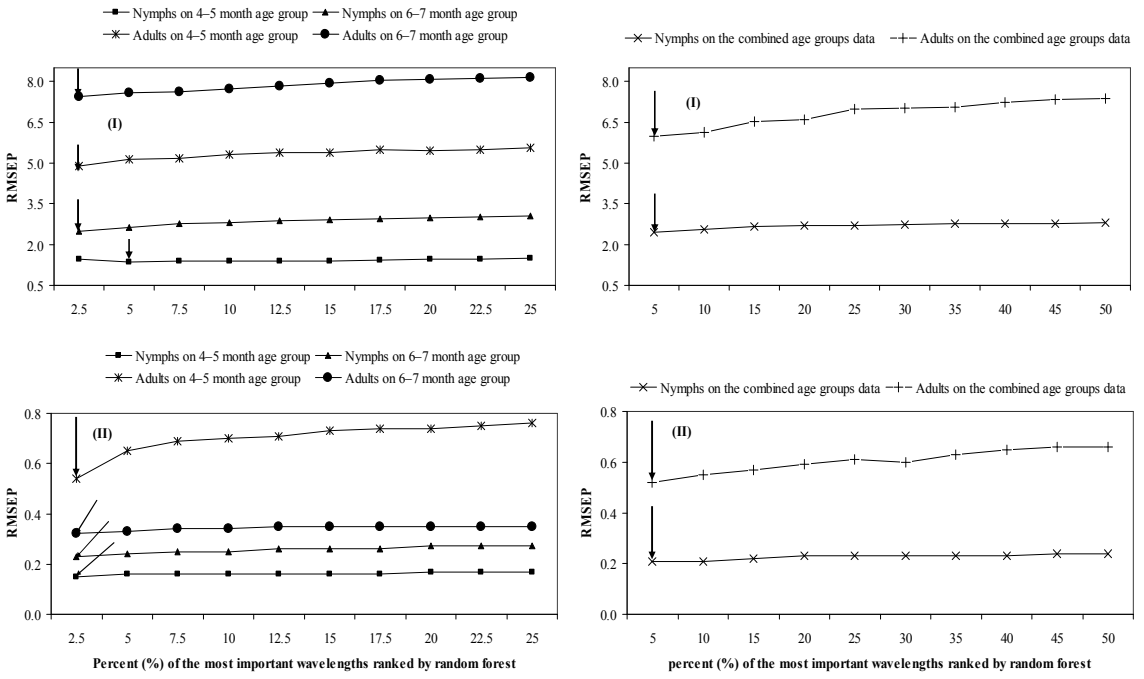

Figure 7.4. RMSEP values as a result of PLS analysis using the most important wavelengths in estimating nymph and adult numbers in 4–5 and 6–7 month age groups as well as the combined age groups data: (I) December 2007 and (II) March 2008. The arrows show the suitable percentages for the wavelengths that were used for developing the PLS regression models.

7.5 Discussion

Results show that it is possible to predict thrips numbers in younger sugarcane crops (4–5 months old). Keeping *et al.* (2008) reported that sugarcane thrips infests mostly younger sugarcane crops. Despite the low thrips counts (1.19 adults and 1.70 nymphs+adults combined per spindle) in younger crops (4–5 months old) during autumn, the present study also shows that, it could be possible to estimate their abundance. During autumn (March), sugarcane crops of 4–5 months old were 1–2 months old during summer (December) and they might have been heavily infested by the insects at that very young age. This might have caused damage and/or physiological changes in sugarcane leaves which could be detectable using spectral properties during autumn. This result is in accordance with the findings of Genc *et al.* (2008) who reported that younger life stage (3rd instars) of sunn pest numbers had less correlation as opposed to older life stages (4th

instars and adults) with narrow-wavelength vegetation indices in wheat canopies. Difficulty of estimating thrips population for older sugarcane crops (6–7 months old) might be due to tolerance of this age group to thrips infestation.

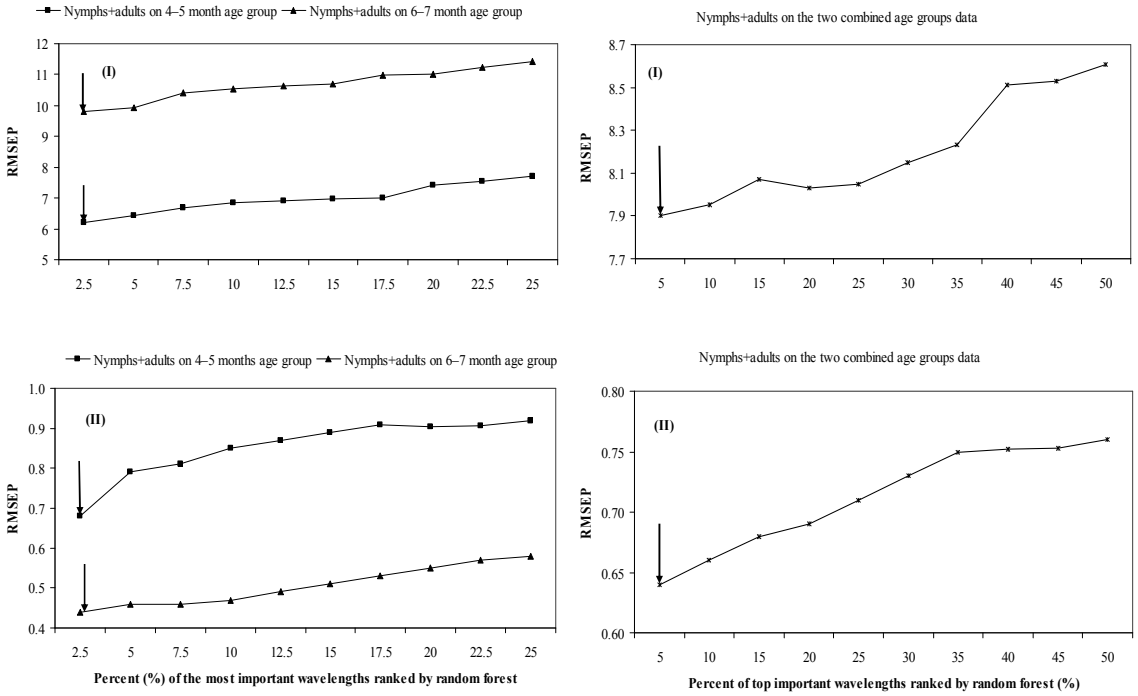

Figure 7.5. RMSEP values as a result of PLS analysis using the most important wavelengths in estimating nymph+adult numbers in 4–5 and 6–7 month age groups as well as the combined age groups data: (I) December 2007 and (II) March 2008. The arrows show the suitable percentages for the wavelengths that were used for developing the PLS regression models.

Results also show relative difficulties of predicting adult numbers as opposed to nymph numbers at all studied age groups in the summer period. This finding needs more investigation in future research work. Predictions of thrips abundance in combined age groups data set were difficult. This may have been due to tolerance of 6–7 month old sugarcane crops to thrips infestation. It also could be due to significant differences of the spectral characteristics between studied age groups (Abdel-Rahman *et al.*, 2008a) which might have caused discrepancies when the insect counts were being estimated with combined life stages on the combined age groups data.

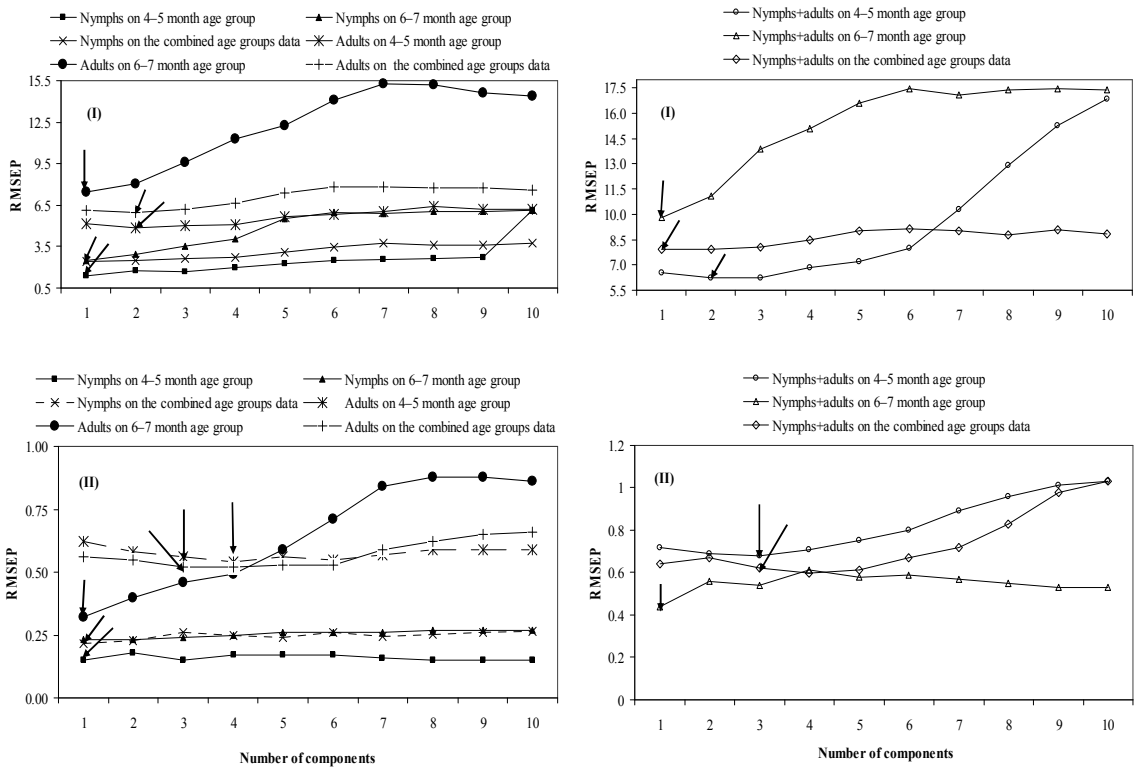

Figure 7.6. RMSEP values using leave-one-out cross validation at different numbers of components (factors) for estimating nymph, adult, and nymph+adult counts on 4–5 month and 6–7 month age groups as well as the combined age groups data. (I) December 2007 and (II) March 2008. The arrows show the suitable number of components for each PLS regression model.

Table 7.3: Summary results of partial least squares (PLS) regression models that were developed to predict nymph, adult, and nymph+adult numbers

	December 2007 (midsummer)		
	Nymphs	Adults	Nymphs+adults
Age group	R^2	R^2	R^2
4–5 month	0.75	0.72	0.75
6–7 month	0.56	0.48	0.49
Combined age groups data	0.57	0.66	0.58
	March 2008 (autumn)		
	Nymphs	Adults	Nymphs+adults
Age group	R^2	R^2	R^2
4–5 month	0.38	0.92	0.86
6–7 month	0.40	0.63	0.60
Combined age groups data	0.42	0.82	0.81

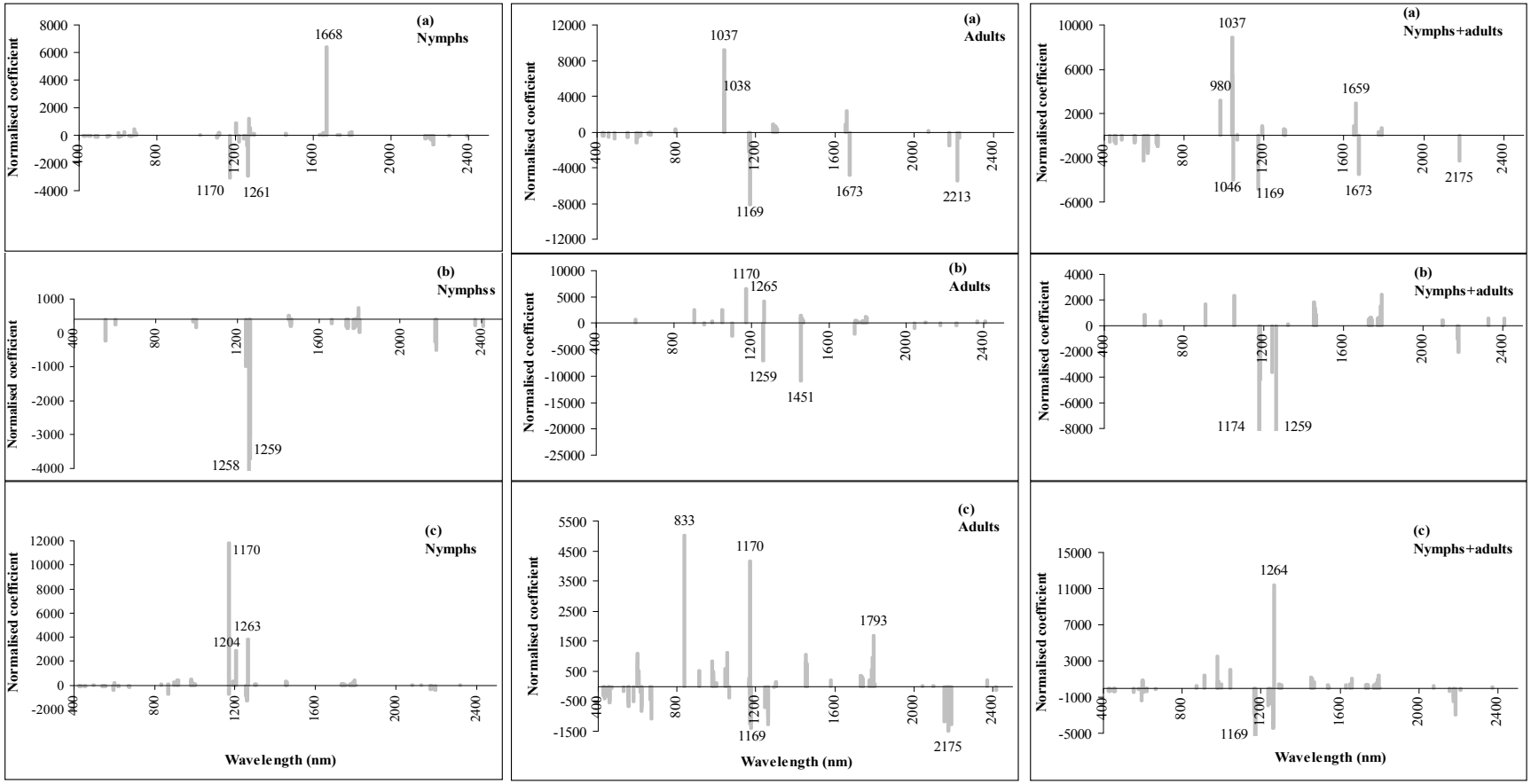

Figure 7.7. Calibration equation coefficients normalised by average first-order derivative of reflectance in December 2007: (a) 4–5 month age group, (b) 6–7 month age group, and (c) the combined age groups data. The labelled spectral features are the peak regions with high influence on the PLS regression models.

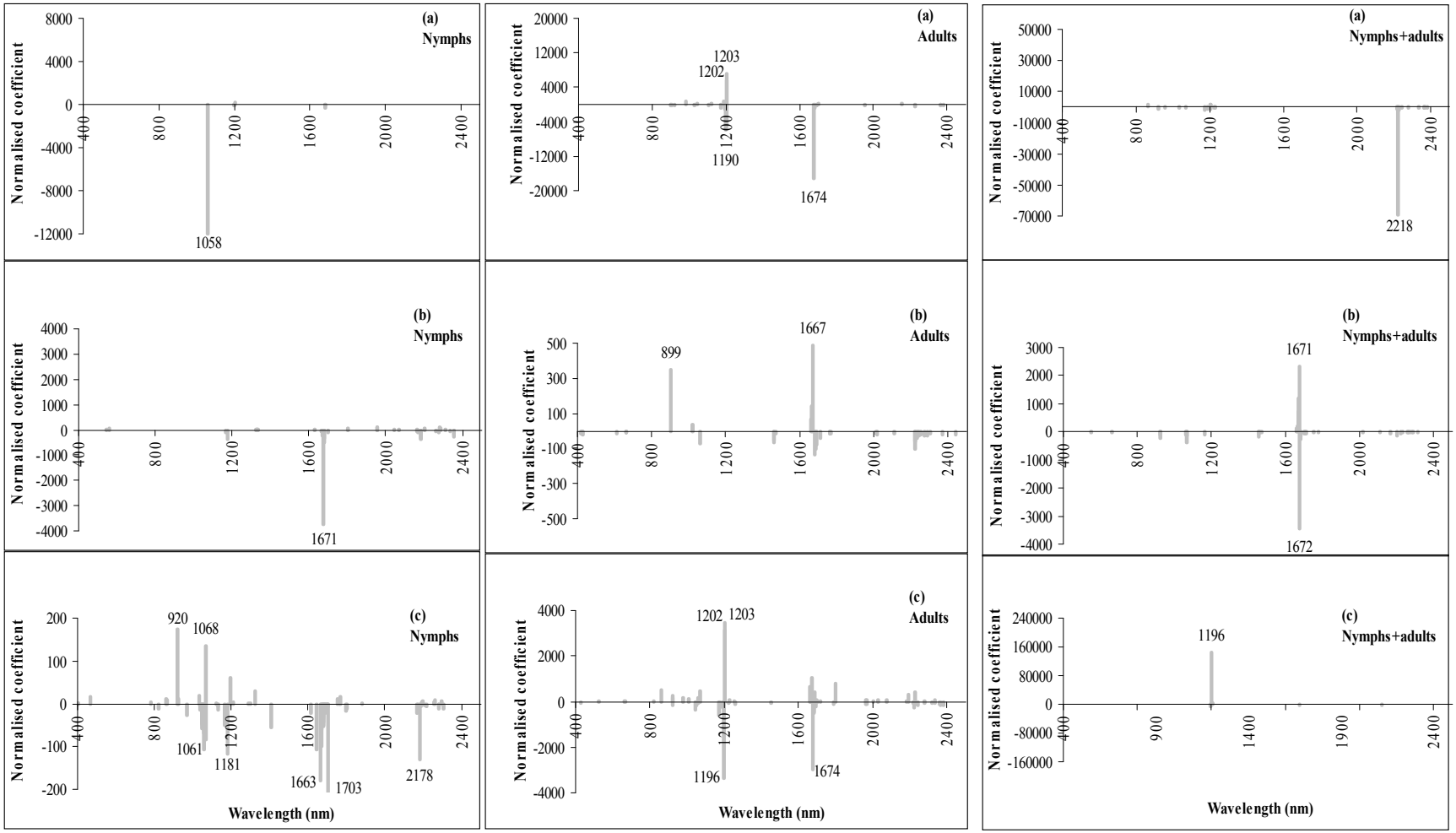

Figure 7.8. Calibration equation coefficients normalised by average first-order derivative of reflectance in March 2008: (a) 4–5 month age group, (b) 6–7 month age group, and (c) the combined age groups data. The labelled spectral features are the peak regions with high influence on the PLS regression models.

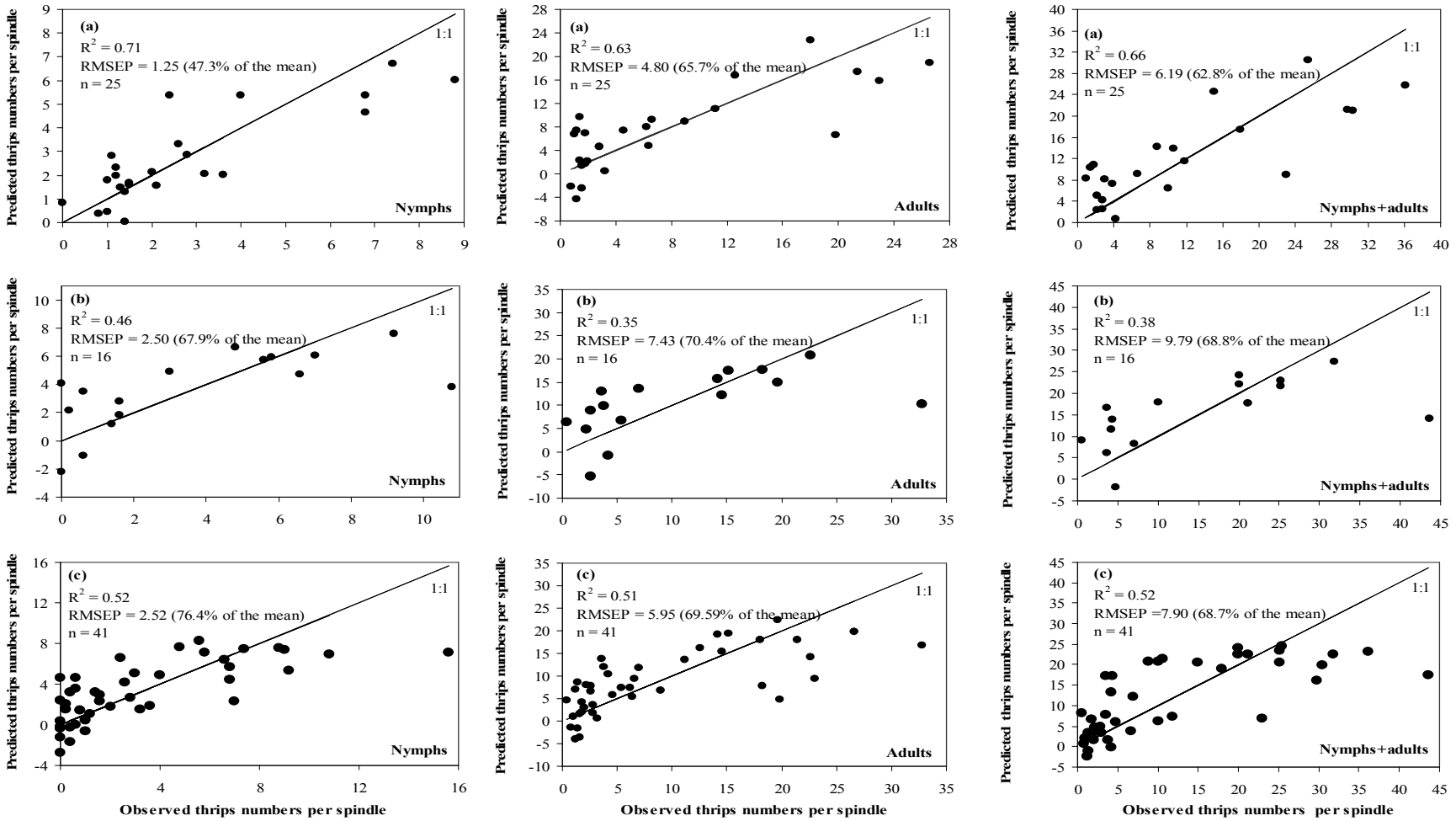

Figure 7.9. Observed versus predicted nymph, adult, and nymph+adult numbers for the sample data set collected during December 2007 using leave-one-out cross validation method for all PLS regression models: (a) 4–5 month age group, (b) 6–7 month age group, and (c) the combined age group data.

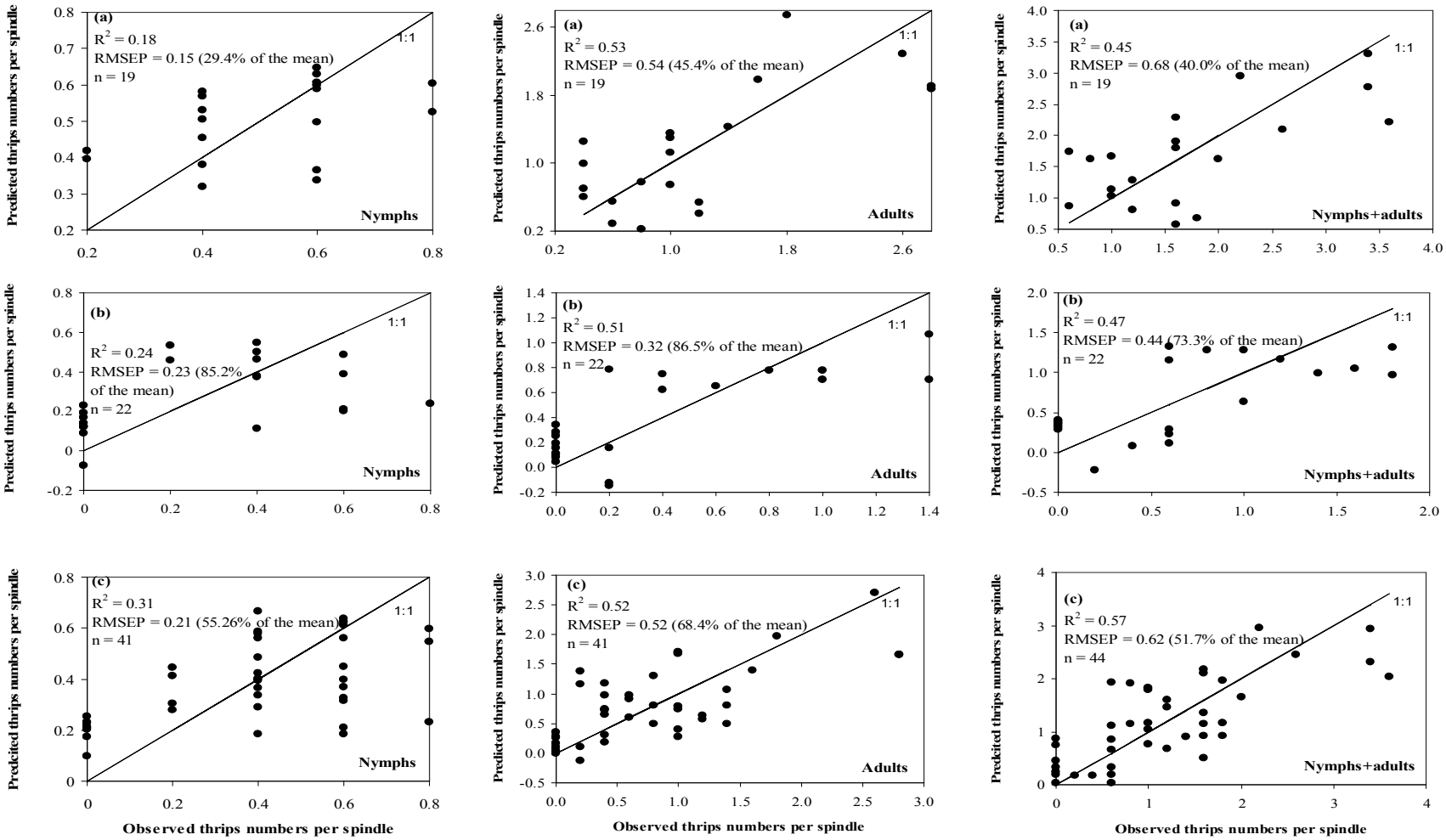

Figure 7.10. Observed versus predicted nymph, adult, and nymph+adult numbers for the sample data set collected during March 2008 using leave-one-out cross validation method for all PLS regression models: (a) 4–5 month age group, (b) 6–7 month age group, and (c) the combined age groups data.

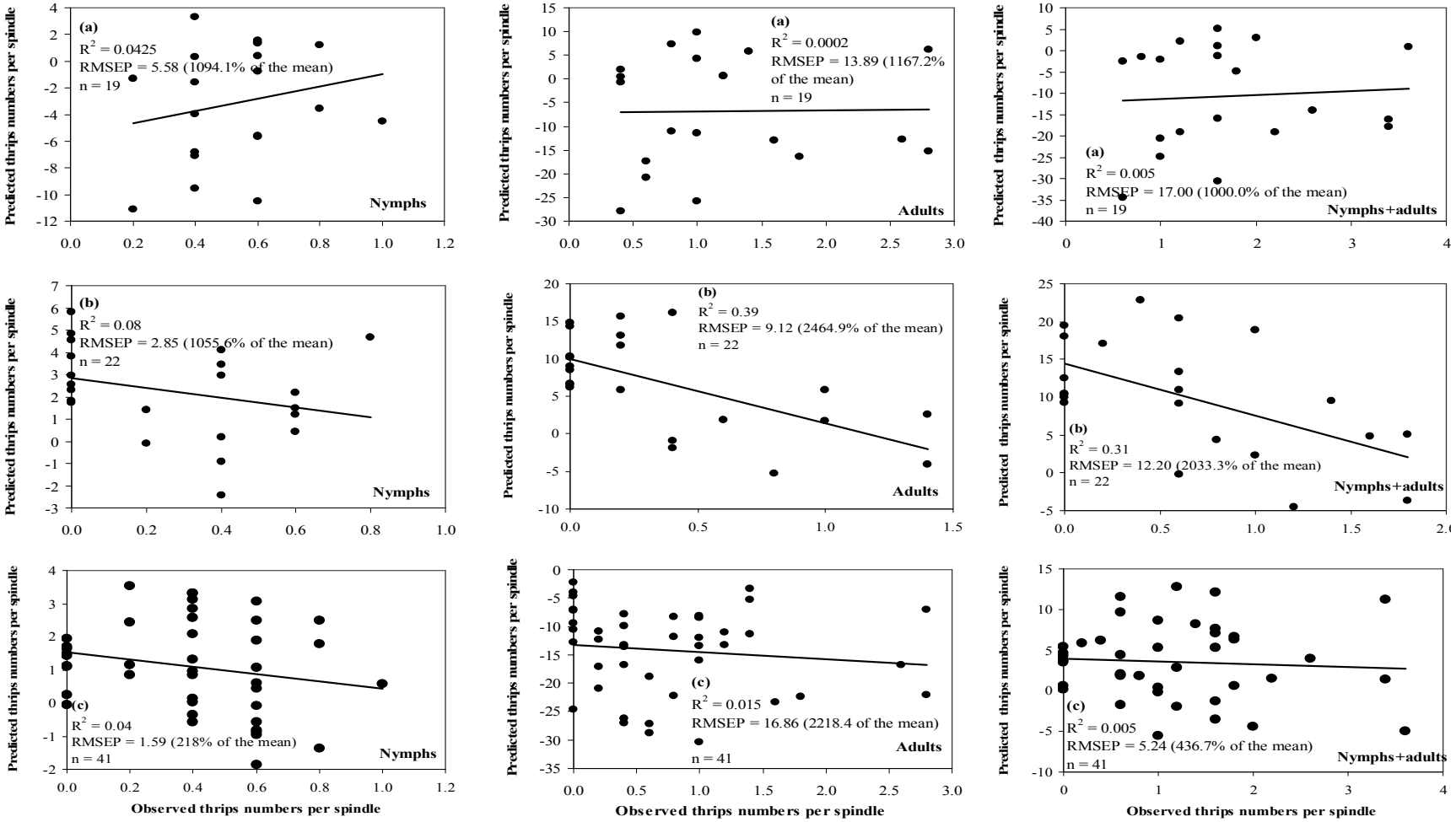

Figure 7.11. Observed versus predicted nymph, adult, and nymph+adult numbers for March 2008 independent data as test data. PLS regression models developed from December 2007 training data set were used for the prediction: (a) 4–5 month age group, (b) 6–7 month age group, and (c) combined age groups data.

Most of the influential wavelengths on the models for estimating thrips populations are within the near infrared (NIR) and middle infrared (MIR) regions and very few wavelengths are within the visible (VIS) region. This result is in disagreement with Abdel-Rahman *et al.* (2008b; 2009c) who found that the VIS region was the most responsive region for detecting sugarcane thrips damage at leaf level. Nonetheless, Abdel-Rahman *et al.* (2008b; 2009c) purposely categorised thrips-damaged leaves into four damage classes where VIS was the detectable region. On the other hand, in the present study no attention has been paid to the thrips damage, and most of leaf samples which were used in this study could be in the healthy category that was employed by Abdel-Rahman *et al.* (2008b; 2009c), hence the VIS region had a low influence on the models. In addition, when leaf samples were collected for this study during both sampling periods (summer and autumn) there were no visible symptoms of damage. Therefore, invisible regions (NIR and MIR) had the most influences on the models.

A negative relationship between first-order derivative of reflectance at some NIR wavelengths (860, 1046, 1058, 1061, 1169, 1170, 1174, 1181, 1190, 1196, 1258, 1259, 1261, and 1451 nm) and thrips numbers is expected, since NIR spectral features affect water content and structure of cells in plants (de Boer, 1993; Kumar *et al.*, 2003) and leaves ruptures or lesions caused by thrips might cause water loss. This result is in accordance with Kolb *et al.* (1991) who observed that pear thrips affected sugar maples' water content and Sudbrink *et al.* (2003), Datt *et al.* (2006), and Mirik *et al.* (2007) who found that insect infestations affected water content of plants and reduced reflectance in the NIR region. However, moisture losses could also happen during storage period of the leaf samples and while keeping the leaves indoors to allow water droplets to evaporate.

Wavelength centred at 1663 nm was also negatively associated with nymph numbers in combined age groups data set for autumn (March, 2008). This was unexpected since 1663 wavelength is within +3 nm from a known moisture-sensitive wavelength (1660) that has been reported by Apan *et al.* (2004b) as being an important spectral feature in separating orange rust-diseased from non-diseased sugarcane canopies. Other wavelengths (1671, 1672, 1673, 1674, 1703, 2175, 2178, 2213, and 2218 nm) had a negative relationship

with thrips abundance. These wavelengths are within ±10 nm from known nitrogen absorption features centred at 1671 nm (Mutanga *et al.*, 2003), 1693 nm (Abdel-Rahman *et al.*, 2008a), 2173 nm (Martin *et al.* 2008), 2180 nm (Kumar *et al.*, 2003), 2204 nm (Bortolot and Wynne, 2003), and 2210 nm (Yoder and Pettigrew-Crosby, 1995). This might imply that sugarcane thrips could cause nitrogen deficiencies.

The models developed using summer (December) observations performed very poorly when applied to autumn (March) conditions. This might be due to numerous factors, such as (i) the difference in the physiological age of the crops between the two sampling periods (ii) the difference in thrips counts between the two sampling periods (iii) the effect of the confounding factors (if any) could be different between the two sampling periods.

Since the PLS regression method was used for the development of the predictive models, and very few components were selected for each model, the researcher does not expect any overfitting problem. The results of this study can be used to monitor thrips incidence. In addition, the study also provides relative information about the effect of thrips on sugarcane.

The study showed different models for estimating thrips population during summer and autumn. Further research work should look for an optimum model to work as a commonly appropriate index for estimating sugarcane thrips abundance. The results of this study should also be validated using spectral data from different varieties at different areas to test the performance of the developed models. Since there might be some confounding factors that could cause similar effects to thrips infestation, further research should investigate the effect of thrips incidence on sugarcane leaves' and canopies' spectral characteristics under trials conditions.

The inconsistency of findings of this study and the results of some other studies might be due to fact that different insect pests and diseases cause diverse symptoms, while the agreement does not indicate direct cause-effect associations.

7.6 Conclusions

There is potential to use the spectroscopy technique at leaf level to detect the presence of the sugarcane thrips pest. Thrips counts could be estimated in sugarcane of 4–5 month old crops. There were difficulties in estimating thrips populations in the older sugarcane crops (6–7 month old). Time of field sampling and knowledge of physiological age of the plants are important when acquiring remotely-sensed data. The study also provided relative information about the effect of thrips on sugarcane.

Acknowledgments

The University of KwaZulu-Natal (South Africa) and the South African Sugarcane Research Institute are thanked for providing funding. Mr. Marius Adendorff, Cyril Cele, Nitesh Poona, Innocent Shezi, and Tholang Mokhele are thanked for their assistance in the field.

CHAPTER EIGHT

Estimation of sugarcane yield using Landsat TM and ETM+ data sets

This chapter is based on

Abdel-Rahman, E. M., In Preparation. Random forest regression for sugarcane yield prediction based on Landsat TM and ETM+ derived spectral vegetation indices.

Abstract

Traditionally, yield predictions are made using direct field survey methods. Remote sensing offers attractive techniques for predicting sugarcane yield. This study investigated the use of vegetation indices derived from multi-date Landsat TM and ETM+ data in predicting sugarcane yield (t ha^{-1}) in South Africa. To explore the utility of the spectral vegetation indices, sugarcane fields of two popular sugarcane varieties (N19 and NCo376) were selected and harvested when the crops were 11–12 months old to measure their yield. The random forest algorithm was used to develop the predictive models. The results showed that sugarcane yield can be estimated for variety NCo376 under rainfed (R^2 = 0.88, RMSEP = 2.17; 2.30% of the mean) and irrigated (R^2 = 0.92, RMSEP = 6.75; 29.38% of the mean) conditions while yield estimates for variety N19 can only be predicted under irrigated scenarios (R^2 = 0.81, RMSEP = 10.56; 12.39% of the mean). It was recommended that the random forest predictive models developed in this study should be tested in the other South African sugarcane growing areas under similar conditions to evaluate their performance.

Keywords: Keywords: multispectral**;** Landsat; sugarcane; yield prediction; random forest

8.1 Introduction

Crop yield is an important measure that integrates the cumulative effect of biotic and abiotic factors (Pinter *et al*., 2003; Fuqin and Guoliang, 1991). The effect of these factors on yield can vary among fields, across fields, and between growing seasons (Bullock, 2004). Researchers have shown that early prediction of crop yield is important for planning and decision making (Ray *et al*., 1999; Sawasawa, 2003; Doraiswamy *et al*., 2004; Ferrio *et al*., 2005; Kogan *et al*., 2005; Zhang *et al*., 2005; Liu *et al*., 2006; Everingham *et al*., 2007b; Everingham *et al*., 2009). Traditionally, yield predictions are made using direct field survey methods. These methods are considered to be time-consuming, labour-intensive and costly. Similarly, yield prediction models that use weather data are also associated with a number of problems related to the spatial distribution of the weather stations (Doraiswamy *et al*., 2005; Kastens *et al*., 2005). Therefore, effective alternative methods that provide good estimates of crop yield are sought and could be highly beneficial.

Remote sensing provides spectral information on crop biophysical characteristics that help in assessing crop vigour and growth which are related to crop yield (Singh *et al*., 2002). In addition, remote sensing can provide near-real-time, consistent, synoptic, and inexpensive data that can be valuable in yield prediction (Maselli *et al*., 2000; Thenkabail, 2003). Using remote sensing techniques, crop yield can be estimated at large spatial coverage and with archived remotely-sensed data the prediction models can be validated using yield data from the preceding cropping seasons (Lobell *et al*., 2005). A single-date image may not be suitable for predicting crop yield because the prevailing conditions in the field could change at any time during the growing season. Multi-date spectral information can capture the seasonal crop cycle and most likely predict crop yield accurately as opposed to single-date spectral data during the growing season (Dash and Curran, 2007).

Remote sensing studies have shown that it is possible to obtain relationships between spectral information extracted from multispectral data and yield in field crops such as

sugarcane (see Abdel-Rahman and Ahmed, 2008), maize (Lewis *et al*., 1998; Kastens *et al*., 2005; Kogan *et al*., 2005; Mkhabela *et al*., 2005; Salazar *et al*., 2008), wheat (Idso *et al*., 1980; Smith *et al*., 1995; Labus *et al*., 2002; Singh *et al*., 2002; Thenkabail, 2003; Bullock, 2004; Enclona *et al*., 2004; Jiang *et al*., 2004b; Kastens *et al*., 2005; Vicente-Serrano *et al*., 2006; Ren *et al*., 2008), rice (Sawasawa, 2003), cotton (Ray *et al*., 1999), millet and sorghum (Maselli *et al*., 2000), and soybean (Kastens *et al*., 2005). With the exception of e.g., Singh *et al*. (2002), Thenkabail (2003), and Enclona *et al*. (2004), the majority of these studies investigated crop yield prediction at regional and district levels using vegetation indices (e.g., Normalised Difference Vegetation Index: NDVI) derived from coarse-resolution sensors such as Advanced Very High Resolution Radiometer (AVHRR on-board NOAA). With the exception of Jiang *et al*. (2004b), who employed an artificial neural network model with back propagation algorithm, studies have applied simple and multiple linear regression models to estimate crop yield. Notwithstanding their simplicity, yield prediction models based on linear relationships might not fit sugarcane yield data well. This might be attributed to the complex spatio-temporal pattern of sugarcane spectral characteristics (Bégué *et al*., 2008). The more complicated multiple non-linear relationships could give more flexibility than simpler linear models (Motulsky and Christopoulos, 2003) and explain high variance of sugarcane yield. Linear models, on the other hand, assume a linear relationship between spectral input variables and yield, which might not always be the case.

The keen interest in crop yield prediction using remotely-sensed data is to test the utility of many spectral vegetation indices that are related to plant pigments, water, and other foliar chemical components. However, one would need to know the relevant spectral indices to predict crop yield. In addition, there is a need to collect a reasonable number of field observations in order to avoid the expected overfitting problem. This problem is a result of using a few field samples as opposed to a large number of input variables in the regression models, which subsequently leads to instability in the prediction models (Reunanen, 2003; Faber and Rajkó, 2007). Given the difficulties and logistics related to the cost, time, and accessibility of the field data collection, researchers have commonly used statistical methods that do not encounter any overfitting problems and variable

selection methods that can efficiently be used to select a few input variables in the regression model. There are relatively new approaches such as boosting and the random forest algorithms that can be used for both purposes, where the selection process of variable is included during the development of the regression models.

Random forest, a machine learning algorithm, is a relatively new approach that was developed by Breiman (2001) as a classification and regression method. It is robust against overfitting, missing observations, and noise in the data (Breiman, 2001; Prasad *et al*., 2006; Palmer *et al*., 2007). The random forest algorithm can be used for a dual purpose: developing non-linear regression models and selection of variables. This algorithm has been more widely used in remote sensing applications as a classifier (e.g., Pal, 2005; Gislason *et al*., 2006; Lawrence *et al*., 2006; Everingham *et al*., 2007a; Chan and Paelinckx, 2008) than a prediction model (e.g., Abdel-Rahman *et al*., 2009b; Ismail, 2009) and it has been found to perform well.

8.2 Rationale

In South Africa, the use of remote sensing in predicting sugarcane yield (tonnes of fresh stalk per hectare: t ha^{-1}) has been studied by the Agricultural Research Council (ARC, 2000a) using NDVI derived from NOAA/AVHRR, ARC (2000b) employing Digital Multispectral Viedo (DMSV) data, and Gers (2003a) who utilised multi-date NOAA NDVI and Landsat ETM+ data. ARC (2000 a,b) found no correlation between NDVI and sugarcane yield at farm level, but there were significant correlations between NOAA-NDVI and yield at the level of mill supply area. Gers (2003a) performed correlation matrix principal component analyses between yield and NDVI or infrared index at four age groups individually and at combined age groups data set. The highest coefficient of determination (R^2) values were 0.047 using NOAA-NDVI, 0.52 using NDVI based on ETM+, and 0.98 using infrared index based on ETM+ (using three observations only). In conclusion, none of the work that was carried out in South Africa to estimate sugarcane yield employing remote sensing at farm and field levels showed any reasonable or reliable results. More studies are still needed: to explore the utility of other spectral

vegetation indices in predicting sugarcane yield; and to apply other sophisticated statistical techniques. Thus, the objective of this study was to investigate the utility of spectral vegetation indices derived from multispectral Landsat TM and ETM+ data sets in predicting sugarcane yield (t ha^{-1}) using the machine learning random forest regression algorithm.

8.3 Materials and Methods

8.3.1 Image acquisition and preprocessing

Three cloud-free mosaiced images were successfully captured (path 167/ 79 and 80) covering large-scale growing sector only, i.e. two Landsat 5 TM (21 April and 11 August 2007) and one Landsat 7 ETM+ SLC-off (2 December 2007). The Landsat 7 ETM+ image was acquired instead of the Landsat 5 TM one because the Landsat 5 TM sensor had a battery problem on October 2007 which resulted in a suspension of operational imaging until March 2008 (Loveland *et al*., 2008). Additionally, Landsat 7 ETM+ was chosen because of its similarities to Landsat 5 TM in its spatial and spectral characteristics.

The satellite images were first geometrically registered to a mosaiced Landsat TM image (7 May 2001), which was already georeferenced (Universal Transverse Mercator, South 36 zone). Nearest neighbour algorithm was performed to resample the images used in the current study into their initial pixel size (30 m x 30 m). A root mean square error (RMSE) of less than a pixel was considered to be an indictor of good geometric correction (Ferencz *et al*., 2004).

Atmospheric correction eliminates the effects of the atmospheric interferences across the image and between images that are being captured at different sites and/or dates. Subsequently, a Chavez's Cost model (Chavez, 1996) which is a modified image-based dark object subtraction model was applied to perform radiometric and atmospheric

corrections on the images. As a result, the Landsat images had a surface reflectance factor of between 0 and 1 for all visible and infrared bands.

8.3.2 Spectral vegetation indices

Spectral vegetation indices that were sensitive to chlorophyll, other foliar components, and leaf water content were tested to predict sugarcane yield (Table 8.1). As suggested by Thenkabail (2003), in order to avoid the edge effects of field boundaries, only pixels that were completely within the fields were included in the extraction of the spectral vegetation indices values. Additionally, depending on the field size, pixel values were extracted from at least two polygons (2 x 2 pixels each). These polygons were considered as sample observations (n) when yield prediction models were developed using profiles of multi-date spectral vegetation indices. The multi-date spectral vegetation indices profiles were generated by averaging the extracted values from the three images (Mkhabela *et al*., 2005). These profiles were limited to sugarcane crops that have fully developed canopies (greater than 5 months in age) to avoid the effects of soil background (Maselli *et al*., 2000).

8.3.3 Field data collection

Fields in the large-scale growing sector were surveyed within a week of image acquisition to update information on management practices and to get information on stalk maturation so as to predict the harvesting date of each field. To reduce the effect of rainfall variation, the study area was divided into three distinct strata that consisted of rainfed (costal) farms, supplementary irrigated (midland) farms, and completely irrigated (inland) farms. Then 55 sugarcane fields of varieties N19 (15 fields in the rainfed sector, 13 fields in the supplementary irrigated, and 11 fields in the completely irrigated) and NCo376 (6 fields in the rainfed sector and 10 fields in the completely irrigated) were selected from all strata and surveyed. Varieties N19 and NCo376 were purposively selected because they are the most common varieties in the study area, representing about

39% and 16% of the total crop in South Africa, respectively (South African Sugar Association: SASA, 2007, unpublished industry database). Moreover, variety NCo376 is used as a reference variety in all modelling work of the South African Sugarcane Research Institute (SASRI). There were no cultivations of variety NCo376 in the supplementary irrigated growing sector in the sample data sets. The sample fields were manually harvested when the crops were 11–12 months old, and measured yield data (tonnes of fresh stalk per hectare: t ha^{-1}) were recorded.

8.3.4.1 Random forest ensemble

A machine learning algorithm known as random forest was explored to predict sugarcane yield (t ha^{-1}). The random forest algorithm uses recursive partitioning to grow many regression trees (*ntree*) and then averages the results. Each tree is independently grown to maximum size based on a bootstrap sample from the training data set (approximately 70%) without any pruning. In each tree, the random forest algorithm uses randomness in the regression process by selecting a random subset of variables (*mtry*) to determine the split at each node (Breiman, 2001). In each tree, the ensemble predicts the data that are not in the tree (the out-of-bag: OOB data) and, by calculating the difference in the mean square errors between the OOB data and data used to grow the regression trees, the random forest algorithm gives an error of prediction called the OOB error of estimate for each variable (Breiman, 2001; Maindonald and Braun, 2006; Prasad *et al*., 2006; Palmer *et al*., 2007). This error produces a measure of the importance of the variables by comparing how much the OOB error of estimate increases when a variable is permuted, whilst all others are left unchanged (Archer and Kimes, 2008). Simply stated, "*variable importance is evaluated based on how much worse the prediction would be if the data for that variable were permuted randomly*" (Prasad *et al*., 2006).

The *randomforest* library (Liaw and Wiener, 2002) developed in the R package for statistical analysis (R Development Core Team, 2008) was employed to implement the random forest algorithm. The two parameters, *mtry* and *ntree*, were optimised based on the root mean square error of prediction (RMSEP). The *ntree* values were tested at 500,

1000, 1500, and 2000 based on Prasad *et al*. (2006), while the *mtry* was evaluated at all the possible 24 values (number of tested spectral vegetation indices).

Table 8.1: Spectral vegetation indices used in the analysis

Index	Formula	Description	Reference
1. SR	R_{NIR}/R_{RED}	Simple ratio: related to changes in amount of green biomass, pigment content, and leaf water stress.	Jordan (1969)
2. NDVI	$(R_{NIR}-R_{RED})/(R_{NIR}+R_{RED})$	Normalised difference index: related to changes in amount of green biomass: pigment content and water stress.	Rouse *et al.* (1973)
3. SAVI	$\frac{(R_{NIR}+R_{RED})(1+L)}{(R_{NIR}+R_{RED}+L)}$	Soil adjusted vegetation index: reduces soil brightness in vegetation reflectance.	Huete (1988)
4. Sqrt SR	$(R_{NIR}/R_{RED})^{1/2}$	Square root of simple ratio: sensitive primarily to the green leaf area or green leaf biomass.	Tucker (1979)
5. MSR	$\frac{(R_{NIR}/R_{RED}-1)}{(R_{NIR}/R_{RED})^{1/2}+1}$	Modified simple ratio: makes the relationships between the index and biophysical parameters non-linear.	Chen (1996)
6. NDVIgreen	$(R_{NIR}-R_{GREEN})/(R_{NIR}+R_{GREEN})$	Uses the green reflectance and helps determine nitrogen influences from the green colour of the leaf.	Gitelson *et al.* (1996)
7. GDI	$R_{NIR}+R_{RED}+R_{GREEN}$	Green difference index: non-normalised index, hence usable when the impact of factors such as slope and aspect is not pronounced.	Gianelle and Vescovo (2007)
8. GRDI	$(R_{GREEN}-R_{RED})/(R_{GREEN}+R_{RED})$	Green red difference index: visible light normalised index.	Gianelle and Vescovo (2007)

Table 8.1: (*Continued*.)

Index	Formula	Description	Reference
9. VI	$R_{NIR}-R_{RED}$	Vegetation index: sensitive to the green leaf material or photosynthetically active biomass in the plant canopy.	Tucker (1979)
10. TNDVI	$(NDVI+0.5)^{1/2}$	Transformed normalised difference vegetation index: sensitive to the green leaf material or photosynthetically active biomass in the plant canopy.	Tucker (1979)
11. NLI	$(R^2_{NIR}-R_{RED})/(R^2_{NIR}+R_{RED})$	Non-linear index: makes the non-linear relationship with surface parameters linear.	Goel and Qin (1994)
12. RDVI	$(R_{NIR}-R_{RED})/(R_{NIR}+R_{RED})^{1/2}$	Re-normalised difference vegetation index: linearises relationships with surface parameters that tend to be non-linear.	Roujean and Breon (1995)
13. MNLI	$\frac{(R^2_{NIR}-R_{RED})(1+L)}{(R^2_{NIR}+R_{RED}+L)}$	Modified non-linear vegetation index: it is an improved version of NLA, and it also considers the merit of SAVI.	Gong *et al.* (2003)
14.NDVI*SR	$(R^2_{NIR}-R_{RED})/(R_{NIR}+R^2_{RED})$	Attempts to combine the merits of NDVI with those of SR.	Gong *et al.* (2003)
15.SAVI*SR	$\frac{(R^2_{NIR}-R_{RED})}{(R_{NIR}+R_{RED}+L)\,R_{RED}}$	Attempts to combine merit of SAVI with that of SR.	Gong *et al.* (2003)
16. NDWI	$(R_{NIR}-R_{SWIR1})/(R_{NIR}+R_{SWIR1})$	Normalised difference water index: sensitive to changes in liquid water content of vegetation canopies.	Gao (1996)

Table 8.1: (*Continued*.)

Index	Formula	Description	Reference
17. MSI	R_{SWIR1}/R_{NIR}	Moisture stress index: sensitive to changes in liquid water content of vegetation canopies.	Vogelmann and Rock (1988)
18. R/ G	R_{RED}/R_{GREEN}	Red to green ratio: helps to determine chlorophyll and anthocyanin pigment contents.	Gamon and Surfus (1999)
19. NIR / G	R_{NIR}/R_{GREEN}	Near infrared to green ratio: helps to determine chlorophyll and carotenoid pigment contents.	Almeida and de Souza Fillo (2004)
20. BC	R_{NIR}/R_{SWIR2}	Biochemical compounds: sensitive to the biochemical compounds in vegetation.	Almeida and de Souza Fillo (2004)
21. BC1	$(R_{NIR}/R_{SWIR1})\ R_{SWIR2}$	Biochemical compounds one: sensitive to the biochemical compounds in vegetation.	Almeida and de Souza Fillo (2004)
22.EVI	$2.5[(R_{NIR}-R_{RED})/(R_{NIR}+6R_{Red}-7.5R_{BLUE}+1)]$	Enhanced vegetation index: enhances sensitivity in high biomass canopies (like sugarcane canopies).	Gitelson *et al.* (1996)

Table 8.1: (*Continued.*)

Index	Formula	Description	Reference
23. PCA_1	A multivariate statistical technique that selects non-correlated linear combinations of variables in an n-dimensional space	Principal component analysis for all derived vegetation indices. It helps in removing any possible redundant spectral information and to concentrate the pertinent vegetation information into the top PCs.	Almeida and de Souza Fillo (2004)
24. YI	PCA_1*NDVI	Yield index: it enhances signs of crop damage and works as a mask where areas of high biomass have values of PCA and non-vegetative areas have zero values.	Almeida *et al.* (2006)

R_{BLUE}, R_{GREEN}, R_{RED}, $R_{NIR,}$ R_{SWIR1} and R_{SWIR2} are surface reflectance at blue (band 1), green (band 2), red (band 3), near infrared (band 4), shortwave infrared 1 (band 5), and shortwave infrared 2 (band 7) of Landsat 5 TM+ or Landsat 7 ETM+. For the SAVI and MNLI, L is a canopy background adjustment factor (a correlation factor for soil line between red and near infrared reflectance) set at 0.5.

Sugarcane yield was estimated for variety N19 at each growing sector individually by using a holdout method (Kohavi, 1995). This was done by randomly dividing the original data set into training (70%) and test (30%) samples. In each model, the random forest algorithm was calibrated on the training data (n = 42 for rainfed, 37 for supplementary irrigated, and 30 for completely irrigated growing sector) and was then used to predict sugarcane yield on the test data (n = 18 for rainfed, 16 for supplementary irrigated, and 13 for completely irrigated growing sector). For variety NCo376, the entire data set in each growing sector was used for training due to the limited number of samples (n = 19 for rainfed, and 21 for completely irrigated growing sector). A leave-one-out cross validation method (Kohavi, 1995) was then applied to assess the performance of the model. This was achieved by dividing the data into *k* samples (*k* = total number of samples used for the analysis) and then removing the samples one-by-one. The random forest algorithm was calibrated *k* times using all *k* samples, except for the omitted one, and used to predict sugarcane yield on this excluded sample. For the combined varieties and/or growing sectors data sets, 30% random samples from each variety and/or growing sector were randomly selected and combined to comprise 30% of the combined data sets to ensure that the training and test samples were representative for all varieties and/or growing sectors.

To evaluate the performance of the developed random forest regression models in estimating sugarcane yield, one-to-one relationships between measured and predicted yield values were established and RMSEP, bias, and R^2 were calculated.

8.3.4. 2 Selection of variables

Selection of variables was carried out for the models that showed reasonable calibrated coefficient of determination values of only $R^2 > 0.60$. The importance of the variables (spectral vegetation indices) in sugarcane yield prediction as measured by the random forest algorithm was ranked, and a forward selection function (Kohavi and John, 1997; Guyon and Elisseeff, 2003) was implemented to determine the least number of spectral vegetation indices that could predict sugarcane yield with greater accuracy. Random

forests were iteratively fitted; at each iteration new forests were developed after including the spectral vegetation indices in the model one-by-one (it starts with the most important variable). The nested subset of variables with the lowest RMSEP is chosen as optimal variables for yield estimation. RMSEP was used instead of OOB error of estimate, because the latter is biased and could not be used to evaluate the overall error of the selection procedure (Díaz-Uriarte and de Andrés, 2006).

8.4 Results

8.4.1 Sugarcane yield (t ha^{-1})

Table 8.2 shows the descriptive statistics of sugarcane yield for the tested varieties in each growing sector. Low yield was observed in fields cultivated with variety NCo376 under completely irrigated conditions. Both varieties showed higher mean yield in the rainfed sector.

Table 8.2: Descriptive statistics of sugarcane yield data (t ha^{-1}). There were no cultivations of variety NCo376 in the supplementary irrigated sector

Variety	Sector	Mean	Standard deviation	Minimum	Maximum
N19	Rainfed	93.60	23.53	34.02	128.41
	Supplementary irrigated	78.83	10.35	55.00	110.00
	Completely irrigated	75.70	37.02	05.00	133.00
NCo376	Rainfed	94.41	06.27	90.00	106.40
	Completely irrigated	22.97	25.74	04.80	70.00

8.4.2 Optimisation of random forest regression models

The results of random forest parameters (*ntree* and *mtry*) optimisation are shown in Table 8.3 and an example of how these parameters were selected is presented in Figure 8.1. The optimal *ntree* and *mtry* numbers vary according to the variety and growing sector. All models showed *mtry* values higher than the suggested value (i.e., *mtry* =1/3 of the total number of the variables) by Liaw and Weiner (2002), except for the N19 variety grown in the completely irrigated growing sector. Overall, the results showed that changes in random forest parameters (*ntree* and *mtry*) affect the error of prediction.

Table 8.3: Summary of random forest prediction models developed in this study to predict sugarcane yield using all the tested spectral vegetation indices

Variety	Sector	*mtry*	*ntree*	R^2	RMSEC
N19	Rainfed	17	500	0.25	19.63
	Supplementary irrigated	12	500	0.18	09.87
	Completely irrigated	22	500	0.61	24.14
NCo376	Rainfed	21	1000	0.84	02.42
	Completely irrigated	21	500	0.85	9.71
N19	Combined across all sectors	19	1000	–0.26	30.51
NCo376	Combined across all sectors	16	500	0.23	31.00
N19+NCo376	Rainfed	18	500	0.23	17.81
	Completely irrigated	4	500	0.19	37.91
	Combined across all sectors	3	500	0.19	29.53

RMSEC = root mean square error of calibration

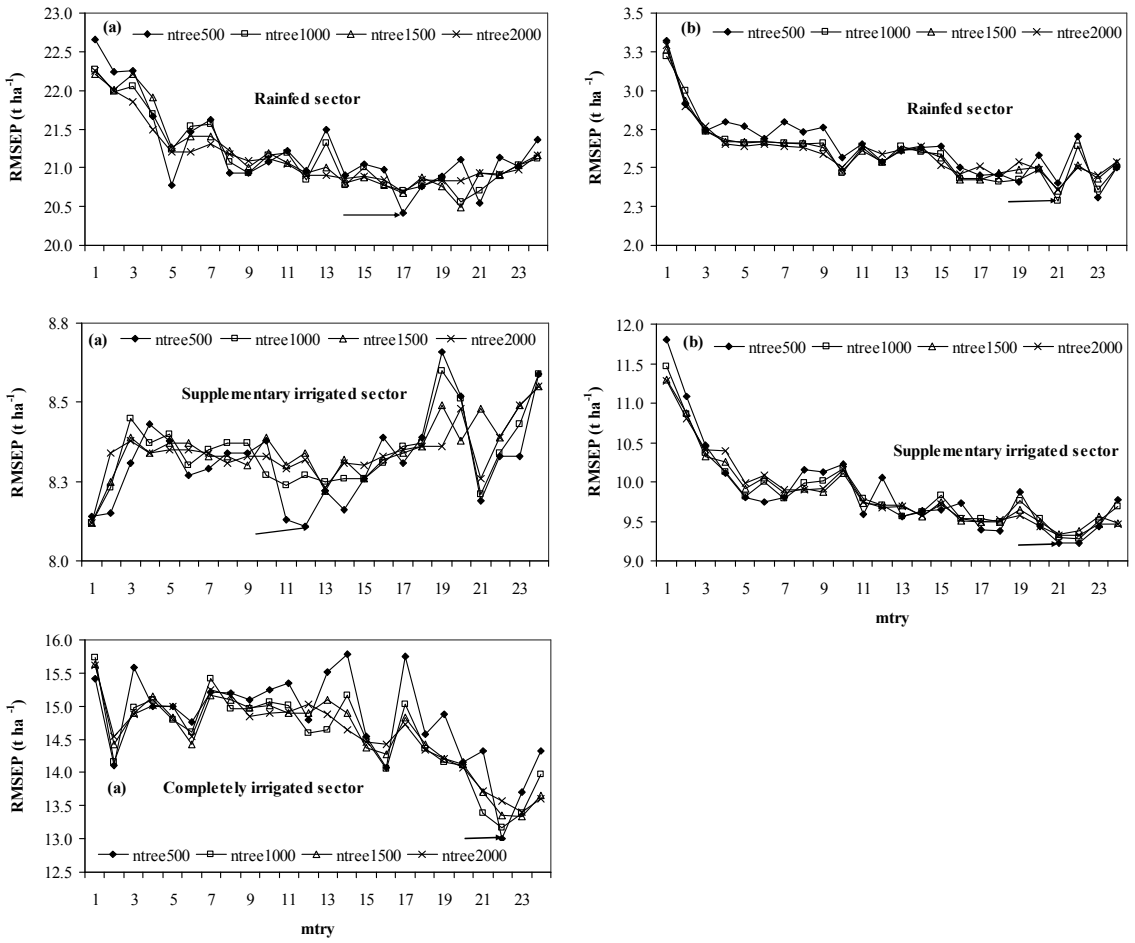

Figure 8.1. An example of random forest parameters (*ntree* and *mtry*) optimisation using RMSEP: (a) variety N19 and (b) variety NCo376. The arrows show the optimal *mtry* value at a definite *ntree* value. There were no cultivations of variety NCo376 in the supplementary irrigated sector.

8.4.3 Random forest prediction models

Table 8.3 summarises the results for the different random forest models when using all the spectral vegetation indices (n = 24) as input variables. High R^2 values were obtained for (i) variety N19 grown under completely irrigated conditions (0.61), (ii) variety NCo376 cultivated under rainfed conditions (0.84), and (iii) variety NCo376 under irrigated conditions (0.85). The values of RMSEC imply that random forest regression models that yielded the three highest R^2 values have fitted NCo376 yield data better than

N19. For all the combined variety and/or growing sector data sets, the result showed low R^2 values.

The validation of each random forest regression model is presented in Figures 8.2 and 8.3. Models that explained high amount of yield variance ($R^2 > 0.60$) showed the lowest RMSEP% and bias.

8.4.4 Selection of variables

Figure 8.4 shows the ranking of individual spectral vegetation indices according to their importance in predicting sugarcane yield for variety N19 (grown under irrigated conditions) and NCo376 (grown under rainfed and irrigated conditions). These are shown for the prediction models that yielded R^2 values of greater than 0.60 (Table 8.3). The importance of a spectral vegetation index in predicting sugarcane yield is dependent on variety and growing sector. In addition, the spectral vegetation indices that are sensitive to the biochemical components in the vegetation (BC and BC1) were ranked among the most four important spectral vegetation indices in all the considered models.

Figure 8.5 shows the results of function of forward selection of variables. A subset including two and three spectral vegetation indices resulted in lowest RMSEP for estimating varieties N19 and NCo376 yields, respectively (Table 8.4). Consequently, these models were run using only the selected spectral vegetation indices (Table 8.4). The results of calibrating and validating the new selected random forest regression models were highly improved (Figure 8.6) because of increased R^2 and reduced RMSEP and bias values.

Table 8.4: Summary of random forest predicted models developed using the selected spectral vegetation indices only

Variety	Sector	Selected vegetation indices	R^2	RMSEC
N19	Completely irrigated	BC, PCA_1	0.81	16.86
NCo376	Rainfed	NDVI*SR, NDVIgreen, EVI	0.88	02.15
	Completely irrigated	EVI, BC1, PCA_1	0.92	06.86

RMSEC = root mean square error of calibration.

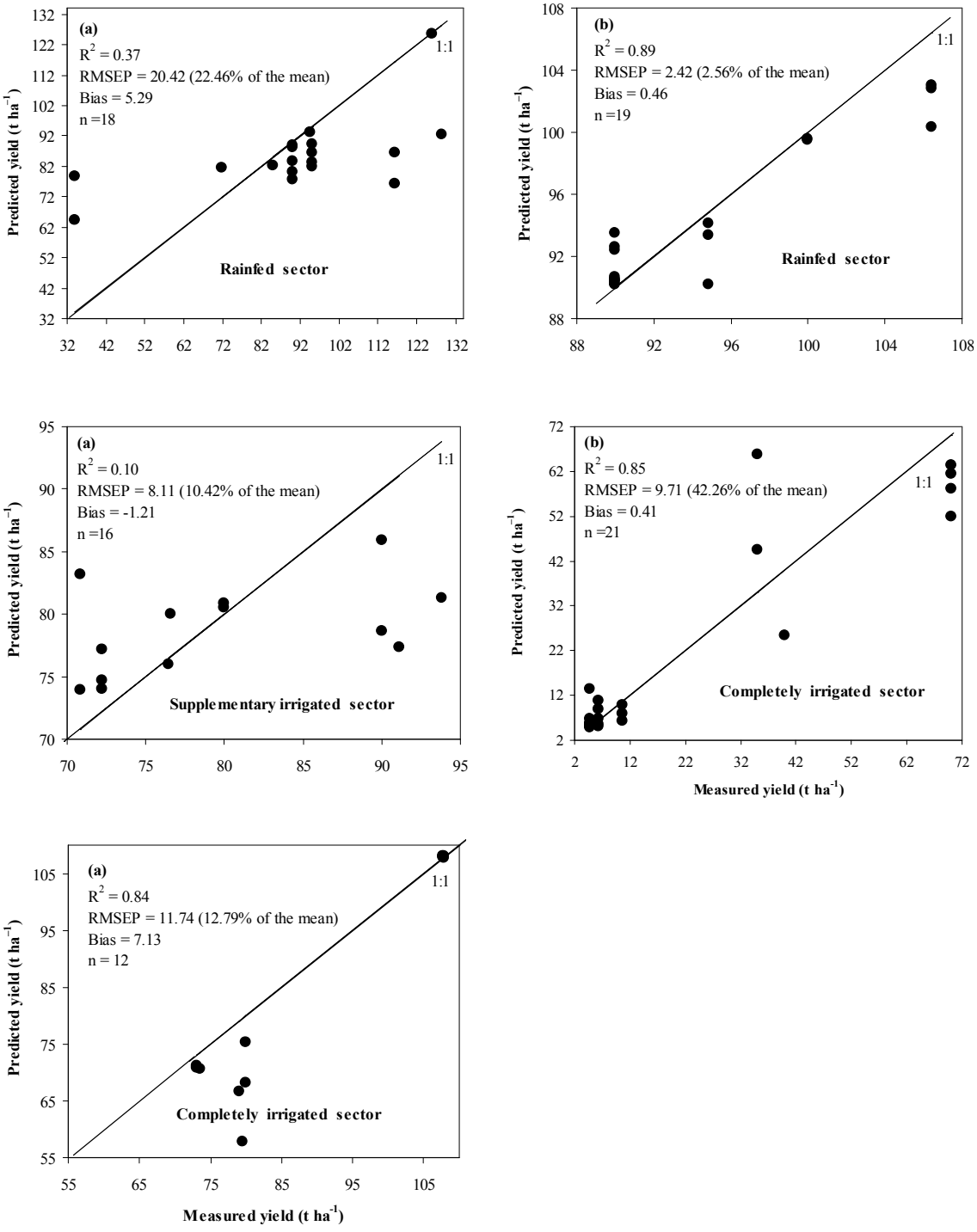

Figure 8.2. One-to-one relationships between measured and predicted sugarcane yield to validate the random forest prediction models developed using all the spectral vegetation indices: (a) variety N19 and (b) variety NCo376. Variety N19 models were validated using a holdout sample (30%), whereas variety NCo376 models were validated using a leave-one-out cross validation method. There were no cultivations of variety NCo376 in the supplementary irrigated sector.

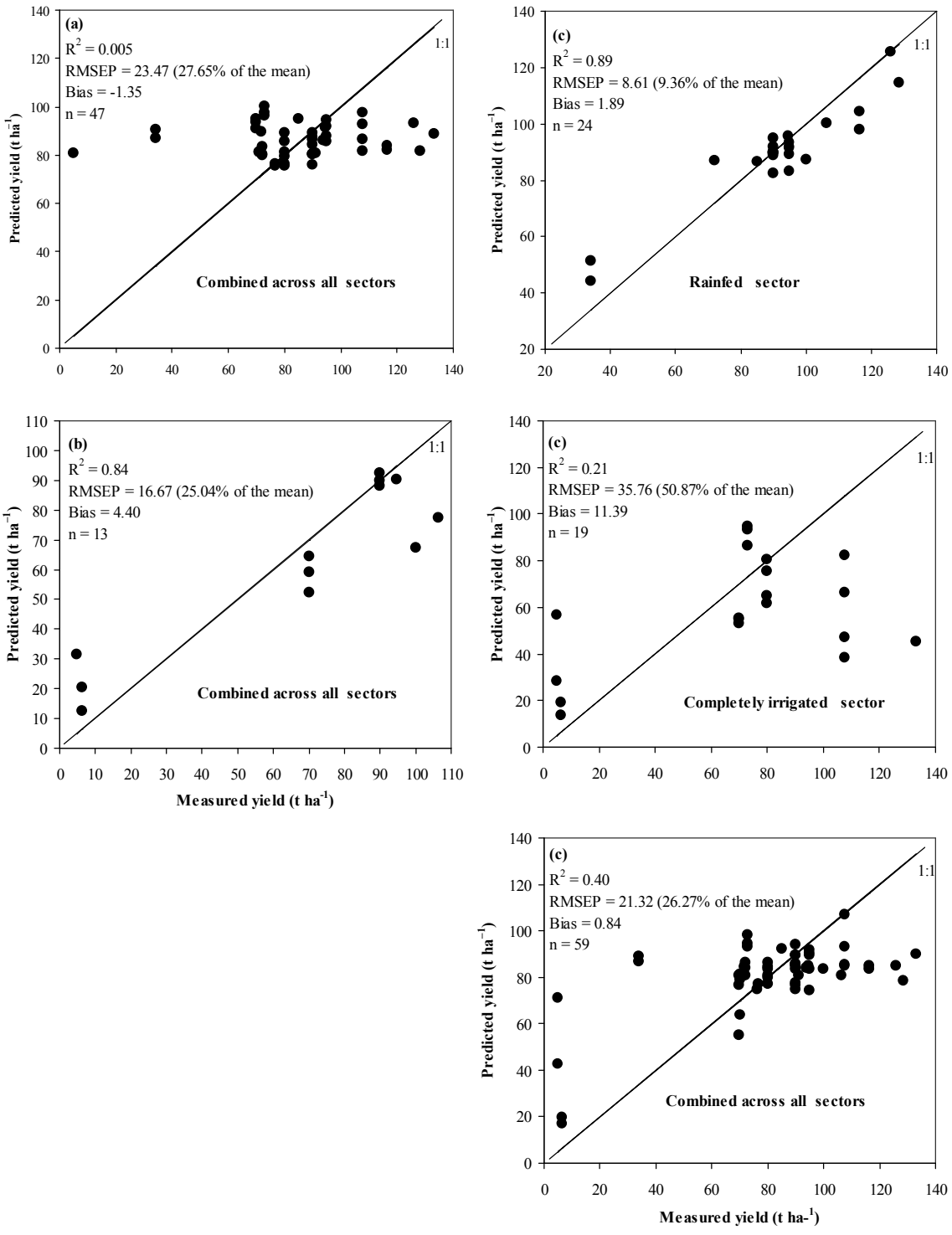

Figure 8.3. One-to-one relationships between measured and predicted sugarcane yield to validate the random forest predictive models developed using all the spectral vegetation indices with a combined holdout sample (30%): (a) variety N19, (b) variety NCo376, and (c) varieties N19 and NCo376 combined. There were no cultivations of variety NCo376 in the supplementary irrigated sector.

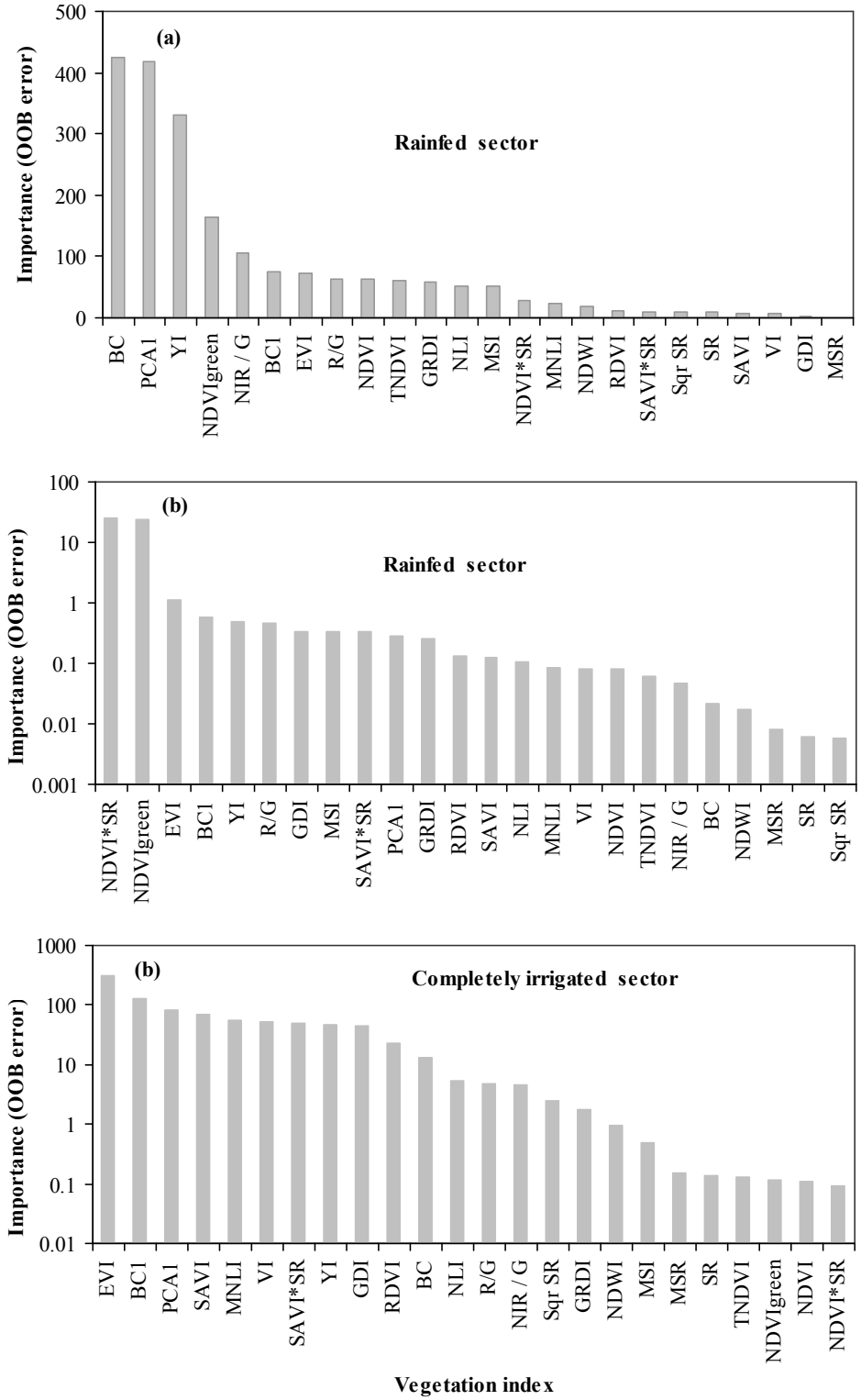

Figure 8.4. The ranking of spectral vegetation indices according to their importance in predicting sugarcane yield using random forest regression models that yielded $R^2 > 0.60$: (a) variety N19 and (b) variety NCo376.

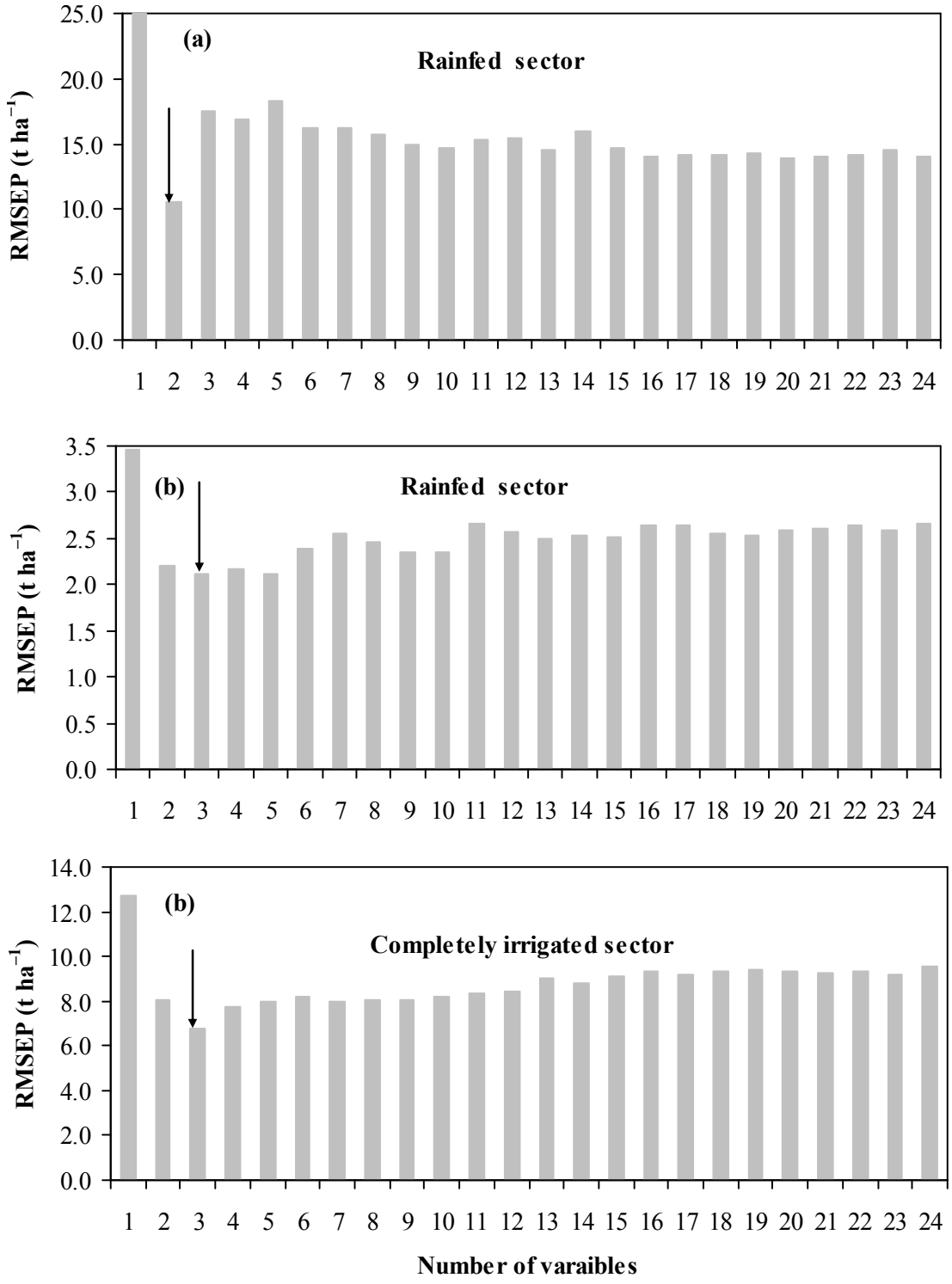

Figure 8.5. Selection of variables using forward selection function. The resultant RMSEP is shown: (a) variety N19 and (b) variety NCo376. The arrows show the minimum number of spectral vegetation indices that resulted in the lowest RMSEP.

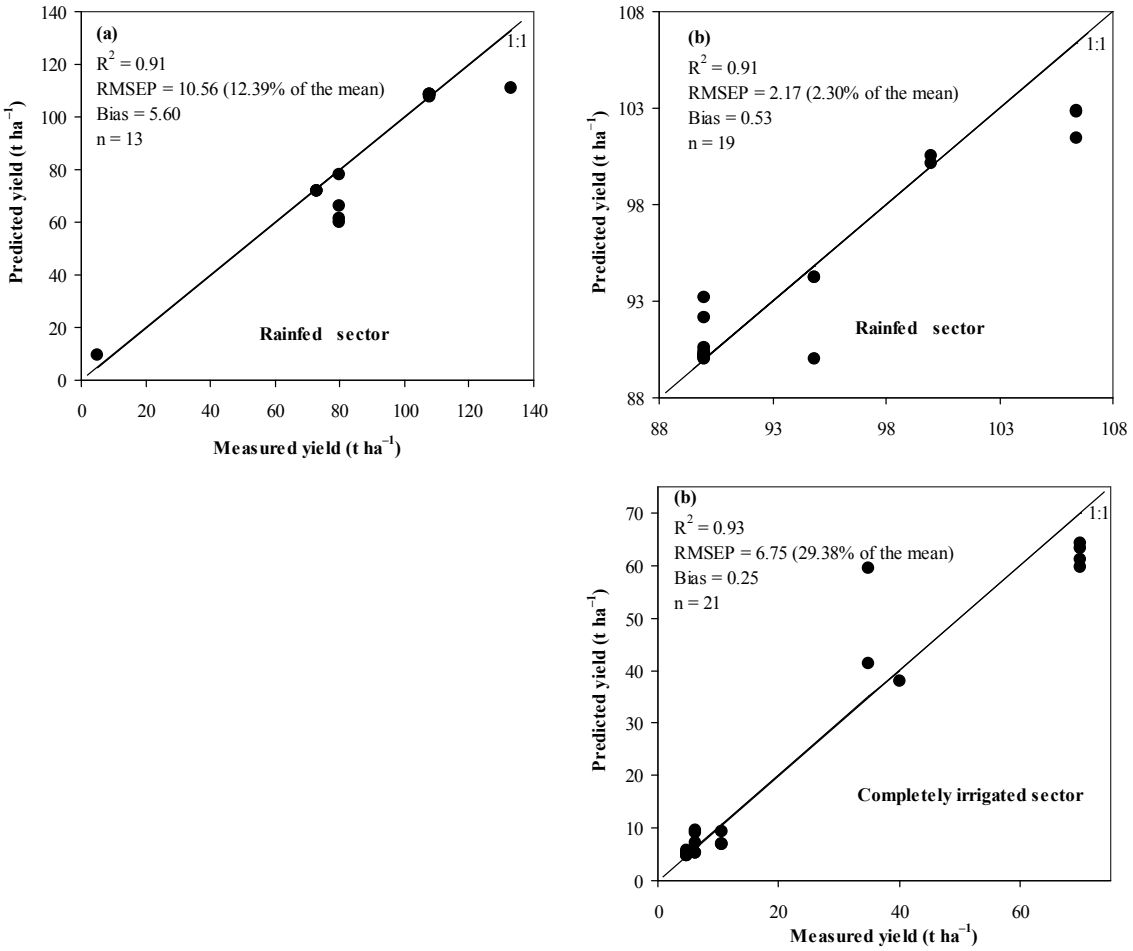

Figure 8.6. One-to-one relationships between measured and predicted sugarcane yield to validate the random forest prediction models developed using the selected spectral vegetation indices only: (a) variety N19 and (b) variety NCo376. Variety N19 model was validated using a holdout sample (30%), whereas variety NCo376 models were validated using a leave-one-out cross validation method.

8.5 Discussion

In order to use one measurement for evaluating the use of the random forest algorithm in this study, RMSEP was used instead of OOB error, as the latter introduces some bias when used in the assessment of selection of variables (Díaz-Uriarte and de Andrés, 2006). The results showed that the RMSEP decreased with increased *mtry* for most of the developed random forest prediction models. However, Liaw and Weiner (2002) recommended that the optimum number of *mtry* should be one third of the total number of the input variables.

The present study showed that sugarcane yields can be estimated for variety N19 under completely irrigated conditions and NCo376 under completely irrigated and rainfed conditions using the random forest regression algorithm. However, under supplementary irrigated and rainfed conditions, variety N19 yields cannot be estimated using spectral vegetation indices derived from Landsat. Possible reasons for the poor performance of the variety N19 prediction models could be due to management strategies (since some of the variety N19 fields in the supplementary and rainfed growing sectors were badly managed and hence infested by weeds). This may have resulted in high green material detected by the remote sensor and relatively low actual sugarcane yield recorded. This could have made the prediction of sugarcane yield difficult using the tested spectral vegetation indices. Another possible reason might be the severe and early lodging of variety N19 crops (SASRI, 2006) in the higher rainfall growing sectors (rainfed and supplementary irrigated), thus modifying the canopy architecture (Almeida *et al*., 2006). This can make the measurement of sugarcane canopy spectral reflectance inaccurate, therefore the yield will be predicted imprecisely. The prediction of sugarcane yields in the combined varieties and/or growing sectors was very inaccurate. Many authors reported that the prediction of sugarcane yield is dependent on variety and agro-climatic conditions (Rao *et al*., 2002; Ha, 2005; Mkhabela *et al*., 2005; Bezuidenhout and Schulze, 2006; Everingham *et al*., 2007b).

The finding of Almeida *et al*. (2006), who successfully estimated sugarcane yield using a single-date Landsat ETM+ data for one variety which was grown under one soil type, confirms the usefulness of Landsat data in estimating cane yield. However, Almeida *et al*. (2006) employed a three-step normalisation method rather than regression models. On the other hand, Lee-Lovick and Kirchner (1991) and Gers (2003a) reported no relationships between sugarcane yield and spectral vegetation indices based on Landsat sensors. The former mentioned that sugarcane canopy moisture content seemed to dominate the spectral properties and mask the long-term stalk development. Whereas the latter recommended that the weak relationship between sugarcane yield and vegetation indices based on Landsat ETM+ could be due to the unsuccessfulness of the approach used. Gers (2003a) tested the utility of multi-date NDVI and infrared indices only by using a simple principal component regression analysis, whereas the present study employed multivariate machine learning method for predicting sugarcane yield. This could suggest that yield of sugarcane crop which is characterised by complex spatio-temporal patterns (Bégué *et al*., 2008) is difficult to estimate using simple linear relationship procedures.

The improvement in the yield prediction (Table 8.4 and Figure 8.6) as a result of using the random forest selection function reveals a good performance of this ensemble as a powerful prediction and feature selection algorithm (Lawrence *et al*., 2006). The selected spectral vegetation indices for both sugarcane varieties are either sensitive to the amount of green materials (PCA_1, EVI, NDVI*SR) or to other biochemical compounds (BC, BC1 and NDVI green) in the plants (Gitelson *et al*., 1996; Gong *et al*., 2003). Spectral vegetation indices that are responsive to the green pigments and to internal cell structure are excellent indictors for vegetation quantity and status (Salazar *et al*., 2008; Kumar *et al*., 2003; Kalubarme *et al*., 2003; de Boer, 1993). Furthermore, EVI enhances the sensitivity when estimating high biomass canopies like sugarcane plantations (Gitelson *et al*., 1996). NDVI*SR is the only index among the selected variables for predicting variety NCo376 yields under rainfed conditions. This index is also a trait of leaf water content which is a major component of the weight of fresh sugarcane stalk.

The relatively high RMSEC and RMSEP in estimating sugarcane yields under irrigation conditions (Table 8.4 and Figure 8.6) could be due to the relatively high variance (standard deviation) of yield in this growing sector (Table 8.2). This could imply that when RMSEC values were calculated the low or high yield values affected the result since the calculation is based on the mean deviations. Furthermore, the bias as shown in the one-to-one relationship between measured and predicted yields is relatively low (Figure 8.6). Therefore, the developed models in this study can be tested at other sugarcane areas in South Africa to evaluate their performance under similar growing conditions. In this study, the yield estimation of two varieties only was assessed, whereas there are many varieties grown in the study site.

Further research work should investigate the effect of crop class (plant or ratoon) on yield prediction using remote sensing techniques. On the other hand, as the average field size in the study area is large (~ 6.5 ha), fine-resolution sensors should also be tested for sugarcane yield prediction. The possibility of integrating climatic factors (e.g., rainfall, temperature) and satellite-based spectral vegetation indices to enhance the performance of the prediction models should be investigated. However, vegetation indices derived from multispectral data can be asymptotically saturated at a certain biomass density or leaf area index (LAI) level (Gao *et al*., 2000). Almeida *et al*. (2006) mentioned that the broadband vegetation parameters can also enhance the unneeded information that bears no relation to vegetation. Mutanga and Skidmore (2004) on the other hand, found that narrow-band vegetation indices can accurately estimate grass biomass at full canopy cover, hence hyperspectral data should also be assessed for estimation of sugarcane yields.

8.6 Conclusions

This study showed that it is possible to estimate variety NCo376 yields, grown under rainfed and irrigated conditions. Variety N19 yields, grown under irrigated conditions only, can be predicted. These findings confirm that prediction of sugarcane yields can be made using multi-date vegetation indices based on Landsat TM and ETM+. The machine

learning algorithm, random forest, proved to be a powerful method of regression analysis and selection of variables. Using this ensemble, sugarcane yield estimation was relatively simple with a minimum number of spectral features that produce the best prediction accuracy.

Acknowledgments

The candidate thanks the South African Sugarcane Research Institute (SASRI) and the University of KwaZulu-Natal in South Africa for funding this study. Thanks are extended to SASRI extension specialist, Mr. Marius Adendorff, for his keen help in identifying the sample fields and providing the yield data. Thanks go to Cyril Cele, Nitesh Poona, Innocent Shezi, and Tholang Mokhele for their support in the field data collection. Gratitude is extended to the R development core team for their very powerful open source packages for statistical analysis.

CHAPTER NINE

Sugarcane nitrogen status, thrips damage and infestation, yield, and remote sensing: A synthesis

9.1 Introduction

Monitoring of stress in sugarcane and early estimates of its final production are of paramount importance. In South Africa, the monitoring of stress factors is made using the traditional methods of field samples collection. These methods are expensive and time-consuming. Sugarcane yield estimates from the current Canesim forecasting system used in South Africa (Bezuidenhout and Singels, 2007a), on the other hand, may vary considerably from the final yield at harvest (van den Berg *et al*., 2009), because they are based on weather data only, without considering stress caused by nutrient deficiency or pests and diseases. This might negatively impact on the milling schedules and other strategic plans, and hamper mitigating action. Near-real-time monitoring of sugarcane nutrient status and pest and disease infestations, and relatively accurate yield estimates are needed.

The aim of this study was to explore the potential use of remote sensing to detect factors of stress in and to predict yield of sugarcane in South Africa. The specific objectives were: (1) to evaluate the use of spectroscopic data in predicting leaf nitrogen concentrations in sugarcane under controlled conditions, (2) to explore the use of *in situ* spectroscopic data to estimate sugarcane leaf nitrogen concentrations, (3) to investigate the potential of imaging spectroscopy in predicting sugarcane leaf nitrogen concentrations using hyperspectral EO-1 Hyperion data, (4) to examine the potential of spectroscopic data to detect sugarcane thrips damage at leaf level, (5) to investigate the potential use of spectroscopic data in predicting thrips counts at leaf level, and (6) to assess the utility of spectral vegetation indices derived from multispectral Landsat TM and ETM+ data sets in predicting sugarcane yields.

All of the above objectives have been achieved in this study.

9.2 Summary of findings

The potential use of spectroscopy in estimating concentration of sugarcane leaf N was addressed in this study (Chapter 3) by making measurements under laboratory conditions. Variety N19 of two age groups (4–5 months and 6–7 months) was assessed. Spectroscopic data were collected using a FieldSpec® 3 ASD sensor. The spectral data were transformed to their first-order derivatives and continuum-removed and correlated with leaf N content to evaluate the strength of each waveband in detecting sugarcane N status. Wavebands at which spectral features showed high correlations ($r \geq 0.6$ or $r \leq -0.6$) were used to generate vegetation indices based on SR and NDVI which were then employed to estimate leaf N concentration. The SR (744, 2142) and NDVI (2200, 2025) generated from first-order derivatives of reflectance yielded the best relationships with sugarcane leaf N concentrations for the 4–5 month crops ($R^2 = 0.74$; RMSEP = 8.29% of the mean) and the 6–7 month crops ($R^2 = 0.87$; RMSEP = 6.18% of the mean), respectively. The good performance of these models showed the usefulness of spectroscopic data in detecting sugarcane N status. These positive findings prompted the interest in testing the utility of *in situ* spectroscopy in predicting sugarcane leaf N concentrations.

Field spectroscopy is based on collecting spectral data under natural field conditions as described in Chapter 4. A FieldSpec® 3 ASD sensor was used to acquire leaf level *in situ* spectral data from sugarcane variety N19 of 6–7 month crops. The utility of the collected spectral data in predicting sugarcane N status was tested by performing the same procedures as described in Chapter 3. The SR (743, 1316) index generated from first-order derivative of reflectance showed the best relationship ($R^2 = 0.76$; RMSEP = 6.85% of the mean) with sugarcane leaf N content. This prediction model is different from the one developed under laboratory controlled conditions for estimation of sugarcane leaf N. The difference was attributed to the fact that the spectroscopic data were collected under different conditions. Wind, solar angle, light intensity, and the geometry of the leaves in the field make leaf reflectance quite different from that measured under controlled (indoor) conditions. The potential value of *in situ* spectroscopy at leaf level to estimate

sugarcane leaf N content was shown. This prompted the need to explore the usefulness of *in situ* spectroscopy at canopy level using handheld, airborne, and spaceborne sensors.

N estimation using spaceborne spectrometers enables the site-specific application of N fertilizers and can be employed as a management tool to reduce the cost of fertilizer application. The potential use of EO-1 Hyperion data in predicting concentrations of sugarcane leaf N was tested (Chapter 5). Two Hyperion scenes were acquired on 27 November and 1 December 2007 from large-scale and small-scale growing sectors, respectively. The sample fields in both sectors were grown with variety N19 of 6–7 month sugarcane. The hyperspectral Hyperion data were transformed to the first-order derivative of reflectance. A random forest ensemble was employed to reduce the redundancy in the complex hyperspectral Hyperion data and to predict concentrations of sugarcane leaf N. Stepwise multiple linear regression was also tested after the reduction of the redundancy in the hyperspectral data. The utility of NDVI-based vegetation indices involved all possible waveband combinations among 10 most important bands in predicting N content that were ranked by random forest was evaluated. Non-linear random forest prediction model consisted of 21 NDVI-based indices showed R^2 of 0.67 (RMSEP = 1.32% of the mean), while linear multiple prediction model that included NDVI (2042, 691) and NDVI (518, 478) indices yielded R^2 of 0.71 (RMSEP = 1.98% of the mean). The relatively high prediction accuracy of the developed models demonstrated the potential of hyperspectral Hyperion data for assessing sugarcane N status. Non-linear random forest prediction models yielded higher accuracy for N estimation compared to multiple linear models. However, the models generated in this study overestimated concentrations of sugarcane leaf N for the samples of low N content and underestimated the values of samples with a relatively high N content. This might be attributed to the effect of other confounding factors (e.g., insect pests damage and diseases) on sugarcane spectral properties when N concentration was estimated.

The usefulness of hyperspectral data in detecting damage due to sugarcane thrips, *Fulmekiola serrata* Kobus (Thysanoptera: Thripidae), was examined in Chapter 6. Leaf level spectroscopic data were collected under natural sunlight conditions from spindle

leaf samples of two sugarcane varieties *viz*., N19 and N12 of 4–5 months. A handheld FieldSpec® 3 ASD was used to collect the spectral data. Spindle leaf samples were categorised into four thrips damage classes (healthy, low damage, medium damage and severe damage). One-way ANOVA was performed to test the occurrence of significant differences between these damage classes in spectral reflectance data in 400 to 2500 nm range. In order to identify the most responsive wavelengths among those at which reflectance was significantly different between damage levels, sensitivity analysis was performed. Furthermore, canonical discriminant analysis was applied to evaluate the utility of different vegetation indices reported in the literature in discriminating among thrips damage classes. The results presented showed that there are significant differences between the classes of thrips damage in some portions of the visible (400–700 nm) region of the spectrum and in particular in the red edge region (690–720 nm) for both varieties. The results of sensitivity analysis demonstrated that reflectance difference maxima and minima located at 407 and 700 nm for N19 and at 514, 636, and 700 nm for N12 as well as reflectance sensitivity maxima and minima located at 400, 635, 680, and 697 nm for N19 and at 465, 502, 675, 693, and 1955 nm for N12 were the most strongly affected by thrips damage. In addition, it was revealed that the red edge index has the greatest discriminatory power among levels of thrips damage for both varieties. It was difficult to discriminate between healthy and low damage classes. It is hypothesised that sugarcane thrips may induce N and water deficiencies and reduce chlorophyll content in sugarcane leaves.

In Chapter 7 of this thesis, the findings were reported of investigations aimed at assessing the potential use of leaf level spectroscopy for estimating thrips incidence in sugarcane. A handheld FieldSpec® 3 ASD was employed to collect the spectral measurements. The analysis was conducted using variety N19 spindle leaves of 4–5 and 6–7 month old sugarcane collected during December 2007 (summer) and March 2008 (autumn). The random forest algorithm was employed to reduce the redundancy in the spectroscopic data. Partial least squares regression was then used to develop models for predicting numbers of adult, nymph and nymph+adult thrips. It was shown that the predictions were adequate for nymphs population in the younger (4–5 months) crops during the summer

period (R^2 = 0.75 and RMSEP = 47.3% of the mean). For the autumn period, the correlations were high *viz*., R^2 = 0.92 for estimating the number of adults (RMSEP = 45% of the mean) and 0.86 for nymphs+adults combined (RMSEP = 40% of the mean) in the younger crops. However, the accuracy of predicting abundance of sugarcane thrips was low in older sugarcane crops (6–7 months). Prediction models of thrips counts differ according to thrips life stage assessed, sugarcane age group, and the season at which the estimation takes place. It is inferred that thrips could cause water and N deficiencies in sugarcane leaves.

The potential use of multi-date vegetation indices derived from multispectral Landsat TM and ETM+ data for predicting sugarcane yield was assessed in Chapter 8. Three landsat images (two Landsat 5 TM and one Landsat 7 ETM+) were acquired on 21 April, 11 August, and 2 December 2007, respectively. Two popular sugarcane varieties (NCo376 and N19) were selected from the large-scale growing sector only. The random forest algorithm was used to develop yield prediction models. Sugarcane yield (t ha^{-1}) was measured in sample fields under three different irrigation conditions (completely irrigated, supplementary irrigated, and rainfed). It was illustrated that sugarcane yield can be estimated for variety NCo376 grown under rainfed (R^2 = 0.88, RMSEP = 2.30% of the mean) and irrigated (R^2 = 0.92, RMSEP = 29.38% of the mean) conditions, while yield estimates for variety N19 can be predicted only for irrigated conditions (R^2 = 0.81, RMSEP = 12.39% of the mean). Models developed for yield forecasting when the data from different varieties and/or irrigation conditions were combined did not perform well. It was concluded that weed infestation in some fields of variety N19 could be one of the main reasons that influenced the spectral features of sugarcane canopies when yield was modelled.

9.3 Conclusions

The main aim of this study was to examine the potential use of remote sensing techniques in quantifying stress and predicting yield of sugarcane. The results reported in Chapters 3 to 7 show that there is potential for using hyperspectral data to detect sugarcane N status

and to monitor thrips damage and infestation in sugarcane. The possible use of multispectral data in predicting sugarcane yield has also been demonstrated (Chapter 8). The findings reported in this study contribute to the research in general and to the feasibility of applying remote sensing technologies in monitoring sugarcane. The final concluding remarks were based on the following findings from the different objectives addressed in this study:

1. Information on sugarcane N is contained in the visible (400–700 nm), red edge (670–780 nm), and middle infrared (1300–2500 nm) regions of the electromagnetic spectrum and no spectral features in the near infrared region were reported in this study to be related to sugarcane leaf N content (chapters 3 to 5).
2. Models for estimating status of sugarcane N depend on the age of the crop, conditions under which spectral data are collected (controlled laboratory or natural field conditions), and the level at which the spectral data are collected (leaf or canopy levels) (Chapters 3 to 5).
3. The red edge position showed the highest potential of detecting thrips damage in sugarcane (Chapter 6).
4. It is difficult to discriminate between low and medium thrips-damaged leaves using spectroscopic data (Chapter 6).
5. Incidence of thrips in sugarcane can be predicted using leaf level spectroscopic data. Models of estimation depend on the age of sugarcane, thrips life stage, and the season in which the spectral data are collected (summer or autumn) (Chapter 7).
6. Sugarcane thrips might cause N and water deficiencies and reduce chlorophyll content in sugarcane leaves (Chapters 6 and 7).
7. Sugarcane yield could be predicted using multi-date vegetation indices derived from multispectral data. Models of prediction depend on sugarcane variety and irrigation conditions (Chapter 8).
8. Weed infestation could be one of the main factors that influence the relationship between spectral properties of sugarcane and its yield (Chapter 8).

9. The machine learning random forest algorithm is deemed to be a powerful feature selection method that can be used to reduce the redundancy in the complex hyperspectral data sets (Chapters 5 and 7).

Overall, this study developed prediction models for detecting status of sugarcane N, thrips damage and incidence, and forecasting sugarcane yield. It was found that models for detecting a stress factor or predicting yield in sugarcane vary depending on age group, variety, season of sampling, conditions at which spectral data are collected (controlled laboratory or natural field conditions), level at which remotely-sensed data are captured (leaf or canopy levels), and irrigation conditions. The study was conducted in only one study area (the Umfolozi mill supply area) and very few varieties (N12, N19, and NCo 376) were tested, thus any extrapolation of the results to other sugarcane varieties and/or sugarcane growing areas should be made with caution. For practical and operational use of remote sensing in sugarcane monitoring, the development of an optimum universal model for detecting factors of stress and predicting yield of sugarcane, therefore, still remains a challenging endeavour.

9.4 Recommendations

Future research could lie in the validation and further elaboration of the models developed in this study in other South African sugarcane producing areas with growing conditions similar to Umfolozi mill supply area, in monitoring sugarcane thrips at field level, in understanding of sugarcane nutritional elements, and the operational applications of remote sensing techniques in sugarcane production. In this context, the following recommendations could be considered for future research work:

1. Models developed in this study should be tested in other South African sugarcane producing areas under similar conditions using more sample data, provided that a reliable geodatabase is established. This allows researchers to select sugarcane varieties, age groups, and irrigation conditions of interest.

2. Detection of sugarcane thrips damage and its incidence should be evaluated at canopy level employing handheld, airborne, and spaceborne sensors (e.g., Hyperion).
3. Further research should investigate the effect of thrips incidence and damage under trial conditions to account for the influence of other confounding factors that might cause effects similar to thrips infestation and/or damage.
4. The ability of multispectral sensors other than Landsat TM and ETM+ (e.g., SPOT, RapidEye, IKONOS, QuickBird, LREye, Sumbandilasat) should be tested for prediction of sugarcane yield. These sensors provide finer spatial resolution compared to Landsat and capture within field variability relatively more accurately. This may result in relatively low error in estimates of sugarcane yield.
5. The utility of hyperspectral data in predicting sugarcane yield should be evaluated. The narrow sugarcane spectral features that are captured by hyperspectral sensors might accurately estimate sugarcane yield, particularly where there are dense sugarcane canopies.
6. The ability of hyperspectral data in predicting nutrients other than N could also be assessed. Sugarcane production and quality can also be influenced by other nutrients such as phosphorous (P) and potassium (K). Hyperspectral data could give accurate estimates of these elements and provide imperative information on sugarcane nutritional status.

9.5 The practical and operational use of remote sensing techniques in sugarcane production

Unlike other field crops, many sugarcane varieties of different age groups are grown in the same area during the growing season. For example, in April a 5-month old ratoon crop can stand adjacent to a 15-month old plant crop. Hence, the spectral and spatio-temporal characteristics of sugarcane are very complex. Furthermore, sugarcane production systems in some regions like South Africa are very dynamic and characterised by diverse climatic conditions (rainfall, temperature, soil, management etc.). These variations and complexity make the development of a single universal model based on

remotely-sensed data for detecting sugarcane stress factors (e.g., N deficiency and thrips damage) and predicting its yield quite a difficult task. Different stress factors may cause similar biophysiological changes, in addition to the visible symptoms thereof on sugarcane crops. Yellowish patches on sugarcane leaves, for example, may be due to thrips damage, N deficiency, disease infection, etc. These confounding effects may be detectable using spectral features in the same region of the electromagnetic spectrum. Spectral features on the red edge position of the electromagnetic spectrum, for instance, could be used to detect insect damage, disease infection, N deficiencies, and so forth. For practical and operational use the exact cause of the spectral changes within a specific region (e.g., red edge) of the spectrum must be identified and, consequently, the specific stress factor must be detected in order to apply the appropriate agricultural input (insecticides, herbicides, N fertilizers, etc.) to control the cause of the stress.

Given the difficulty of developing a universal model to detect each stress factor separately on the one hand, and the high cost per unit area of the hyperspectral data that are captured from airborne (e.g., AVIRIS) and spaceborne (Hyperion) sensors on the other hand, the future practical and operational applications of remote sensing techniques in sugarcane production could involve the use of multispectral data. This could be done in the near future, after the successful launch of the new specific-centric South African Sumbandilasat microsatellite which carries a multispectral sensor, with 6.25 m spatial resolution and 6 spectral bands (Main *et al*., 2008; Mostert *et al*., 2008). Two of its bands (i.e., the red edge band and xanthophyll band) have been strategically placed given an *a priori* relationship between the vegetation spectral features in these two bands and vegetation status (Main *et al*., 2008). Multispectral data are relatively inexpensive, accessible, and do not require complex preprocessing and processing techniques. In this regard, the use of multispectral data should be operationalised and implemented to provide informed and near-real-time management practises such as:

- Estimation of area under cane: the broader multispectral wavebands should mask the spectral differences between some sugarcane varieties, close age groups, seasonality, and other vegetation and/or crops and sugarcane etc. Hence the

multispectral data can be used on a routine basis for mapping areal extent of sugarcane.

- Monitoring harvest: Mutispectral data could identify and map sugarcane fields of standing crops that need to be harvested, trashed fields, green harvested fields, and fields of bare soil that may have been burnt.
- Detection and mapping of stress factors that cause distinct and clear symptoms on sugarcane crops (e.g., rust disease).
- Detection of areas of anomalies: in this case multispectral data could be used to detect areas in the field with anomalies (within field variations). Then, by conducting field visits, the cause of the anomaly in the field could be identified and the appropriate agricultural input could be applied.
- Enhancing of the accuracy of the process-based growth models that are routinely used for sugarcane production forecasts (e.g. Canesim): Spectral information extracted from multispectral data can be very useful for this.

REFERENCES

Abdel-Rahman, E. M. and Ahmed, F. B., 2008. The application of remote sensing techniques to sugarcane (*Saccharum* spp. Hybrid) production: a review of the literature. *International Journal of Remote Sensing*, 29, 3753–3767.

Abdel-Rahman, E. M., Ahmed, F. B. and van den Berg, M., 2008a. Imaging spectroscopy for estimating sugarcane leaf nitrogen concentration. *Proceedings of SPIE Remote Sensing for Agriculture, Ecosystems, and Hydrology X Conference*, V-1–V-12.

Abdel-Rahman, E. M., Ahmed, F. B., van den Berg, M. and Way, M. J., 2008b. Preliminary study on sugarcane thrips (*Fulmekiola serrata*) damage detection using imaging spectroscopy. *Proceedings of South African Sugar Technologists' Association*, 81, 287–289.

Abdel-Rahman, E. M., van den Berg, M., Way, M. J., Ahmed, F. B. and Sewpersad, C., 2009a. Using spectroscopic data sets to predict numbers of thrips (*Fulmekiola serrata*) in sugarcane. *Proceedings of South African Sugar Technologists' Association*, 82, 441– 445.

Abdel-Rahman, E. M., van den Berg, M., Way, M. J. and Ahmed, F. B., 2009b. Handheld spectrometry for estimating thrips (*Fulmekiola serrata*) incidence in sugarcane. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, IV-268–IV-271.

Abdel-Rahman, E. M., Ahmed F. B., van den Berg, M. and Way, M. J., 2009c. Potential of spectroscopic data sets for sugarcane thrips (*Fulmekiola serrata* Kobus) damage detection. *International Journal of Remote Sensing*, In Press.

Abdel-Rahman, E. M., Ahmed, F. B. and van den Berg, M., 2009d. Estimation of sugarcane leaf nitrogen concentration using *in situ* spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, Remote Sensing for Africa: Special Issue, In Press.

Alchanatis, V., Schmilovitch, Z. and Meron, M., 2005. In-Field assessment of single leaf nitrogen status by spectral reflectance measurements. *Precision Agriculture*, 6, 25– 39.

Almeida, T. I. R. and de Souza Filho, C. R., 2004. Principal component analysis applied to feature-oriented band ratios of hyperspectral data: a tool for vegetation studies. *International Journal of Remote Sensing*, 25, 5005–5023.

Almeida, T. I. R., de Souza Filho, C.R. and Rossetto, R., 2006. ASTER and Landsat ETM + images applied to sugarcane yield forecast. *International Journal of Remote Sensing*, 27, 4057–4069.

Alvarez, J. and Polopolus, L. C., 2002. Domestic and international competition in sugar markets. EDIS document SC 021, Gainesville, FL: Department of Food and Resource Economics, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences (University of Florida, USA).

Apan, A., Held, A., Phinn, S. and Markley, J., 2004a. Spectral discrimination and classification of sugarcane varieties using EO-1 Hyperion hyperspectral imagery. *Proceedings of the 25th Asian Conference on Remote Sensing*, available online at: www.gisdevelopment.net/aars/2004/hyper/ acrs2004_a1001.asp (accessed 14 January 2007).

Apan, A., Held, A., Phinn, S. and Markley, J., 2004b. Detection of sugarcane 'orange rust' disease using EO-1 Hyperion hyperspectral imagery. *International Journal of Remote Sensing*, 25, 489–498.

ARC, 2000a. Evaluation of NAOAA-AVHRR satellite information for the estimation of sugarcane yields at a macro scale. Agricultural Research Council Report No. GWA\A\2000\11 (Pretoria, South Africa: ARC).

ARC, 2000b. Evaluation of aerial remote sensing using a digital multispectral video sensor (DMSV) for field monitoring and management of sugarcane. Agricultural Research Council Report No. GW/a/2000/12 (Pretoria, South Africa: ARC).

Archer, K. J. and Kimes, R. V., 2008. Empirical characterisation of random forest variable importance measures. *Computational Statistics and Data Analysis*, 52, 2249–2260.

Aronoff, S. (2005). Satellite-based sensors operating in the visible and infrared wavelengths. In *Remote Sensing for GIS Managers*, S. Aronoff (ed), pp. 167–196 (California, USA. ESRI Press).

ASD, 2005. Handheld Spectroradiometer: *User's Guide Version 4.05* (Suite A Boulder, USA. Analytical Spectral Devices, Inc.), 136 pp.

Asner, G. P., 2008. Hyperspectral remote sensing of canopy chemistry, physiology, and biodiversity in tropical rainforests. In *Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests*, M. Kalacska and G. A. Sanchez-Azofeifa (eds), pp. 261– 296 (London, UK. Taylor and Francis Publishers).

Atkinson, P. R. and Nuss, K. J., 1989. Associations between host-plant nitrogen and infestation of sugarcane borer, *Eldana saccharina* Walker (Lepidoptera: Pyralidae). *Bulletin of Entomological Research*, 79, 489–506.

Bakker, W. H., 2001. Multispectral scanners. In *Principles of Remote Sensing*, 2nd edition [CD-ROM], L. L. F. Janssen and G. C. Huurnmen, (eds), pp. 154–218, (ITC Educational Textbooks Series 2, Enschede, The Netherlands).

Baltsavias, E. P., 2002. Special section on image spectroscopy and hyperspectral imaging. *ISPRS Journal of Photogrammetry and Remote Sensing*. Editorial, 57, 169–170.

Bappel, E. Bégué, A. Martiné, J. F., Pellegrino, A. and Siegmund, B., 2005. Assimilation of a biophysical parameter estimated by remote sensing using SPOT 4 and 5 data into a sugarcane yield forecasting model. *Proceedings of International Society of Sugar Cane Technologists Congress*, XXV, 1–4.

Baret, F., Houlès., V. and Guèrif, M., 2007. Quantification of plant stress using remote sensing observation and crop models: the case of nitrogen management. *Journal Experimental Botany*, 58, 869–880.

Bastiaanssen, W. G. M. and Ali, S., 2003. A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. *Agriculture, Ecosystems and Environment*, 94, 321–340.

Bégué, A., Degenne, P., Pellegrino, A., Todoroff, P. and Baillarin, F., 2004. Application of remote sensing technology to monitor sugarcane cutting and planting in Guadeloupe (French West Indies). *Proceedings of International Congress Geomatica*, available online at: http://publications.cirad.fr/une_notice.php?dk=521009 (accessed 22 January 2007).

Bégué, A., Todoroff, P. and Pater, J., 2008. Multi-time scale analysis of sugarcane within-field variability; improved crop diagnosis using satellite time series?. *Precision Agriculture*, 9, 161–171.

Bezuidenhout, C. N. and Schulze, R. E., 2006. Application of seasonal climate outlooks to forecast sugarcane production in South Africa. *Climate Research*, 30, 239–246.

Bezuidenhout, C. N. and Singels, A., 2007a. Operational forecasting of South African sugarcane production: part 1 system description. *Agricultural Systems*, 92, 23–38.

Bezuidenhout, C. N. and Singels, A., 2007b. Operational forecasting of South African sugarcane production: part 2 system evaluation. *Agricultural Systems*, 92, 39–51.

Blackburn, G. A., 1998. Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. *International Journal of Remote Sensing*, 19, 657–675.

Blumenthal, J. M., Baltensperger, D. D., Cassman, K. G., Mason, S. C. and Pavlista, A. D., 2001. Importance and effect of nitrogen on crop quality and health. In *Nitrogen in the Environment: Sources, Problems, and Management*, R. F. Follet and J. L. Hatfield (eds), pp. 45–63 (London, UK. Elsevier).

Borengasser, M., Hungate, W. S. and Watkins, R. (2008). *Hyperspectral Remote Sensing: Principles and Applications* (New York, USA. Taylor and Francis Group), 119 pp.

Bortolot, Z. J. and Wynne, R. H., 2003. A method for predicting fresh green leaf nitrogen concentration from shortwave infrared reflectance spectra acquired at the canopy level that requires no *in situ* nitrogen data. *International Journal of Remote Sensing*, 24, 619–624.

Breiman, L., 2001. Random forests. *Machine Learning*, 45, 5–32.

Bullock, P. R., 2004. A comparison of growing season agro-meteorological stress and single-date Landsat NDVI for wheat yield estimation in west central Saskatchewan. *Canadian Journal of Remote Sensing*, 30, 101–108.

Carter, G. A., 1994. Ratios of leaf reflectance in narrow wavebands as indictors of plant stress. *International Journal of Remote Sensing*, 15, 697–703.

Chan, J. C. and Paelinckx, D., 2008. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotype mapping using

airborne hyperspectral imagery. *Remote Sensing of Environment*, 112, 2999–3011.

Chavez, P. S., 1996. Image-based atmospheric corrections: revisited and improved. *Photogrammetric Engineering and Remote Sensing*, 62, 1025–1036.

Chen, J. M., 1996. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing*, 22, 229–242.

Chen, L., Huang, J. F., Wang, F. M. and Tang, Y. L., 2007. Comparison between back propagation neural network and regression models for the estimation of pigment content in rice leaves and panicles using hyperspectral data. *International Journal of Remote Sensing*, 28, 3457–3478.

Cho, M. A. and Skidmore, A. K., 2006. A new technique for extracting the red edge position from hyperspectral data: the linear extrapolation method. *Remote Sensing of Environment*, 101, 181–193.

Cho, M. A., 2007. Hyperspectral remote sensing of biochemical and biophysical parameters. PhD Thesis, International Institute for Geo-information Science and Earth Observation, ITC, Enschede, The Netherlands.

Cibula, W. G. and Carter, G. A., 1992. Identification of a far-red reflectance response to ectomycorrhizae in slash pine. *International Journal of Remote Sensing*, 13, 925–932.

Clevers, J. P. G. W. and Jongschaap, R., 2003. Imaging spectroscopy for agricultural application. In *Image Spectrometry*, vol. 3, F. D. van der Meer and S. M. de Jong (eds), pp. 157–199 (London, UK. Kluwer Academic Publishers).

Coops, N. C., Goodwin, N., Stone, C. and Sims, N., 2006. Application of narrow-band digital camera imagery to plantation canopy condition assessment. *Canadian Journal of Remote Sensing*, 32, 19–32.

Costa, C., Dwyer, L. M., Dutilleul, P., Stewart, D. W., Luo, B. and Smith, D. L., 2001. Inter-relationships of applied nitrogen, spad and yield of leafy and non-leafy maize genotypes. *Journal of Plant Nutrition*, 8, 1173–1194.

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

Curran, P. J., 1994. Imaging spectrometry- its present and future role in environmental research. In *Imaging Spectrometry – a Tool for Environmental Observations*, J. Hill and J. Mégier (eds), pp. 1–23 (London, UK. Kluwer Academic Publishers).

Curran, P. J., Dungan, J. L. and Gholz, H. I., 1990. Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. *Tree Physiology*, 7, 33–48.

Dash, J. and Curran, P. J., 2007. Relationship between the MERIS vegetation indices and crop yield for the state of South Dakota, USA. *Proceedings of Envisat Symposium*, available online at: http://earth.esa.int/workshops/envisatsymposium/proceedings/posters/3P2/457801da.pdf (accessed 24 April 2009).

Datt, B., Apan, A. and Kelly, R., 2006. Early detection of exotic pests and disease in Asian vegetables by imaging spectroscopy. Rural Industries Research and Development Corporation Publication no. 05/170 (Kingston, Australia).

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., Brown de Colstoun, E. and Mcmurtrey, J. E., 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74, 229–239.

de Boer, Th. A., 1993. Botanical characteristics of vegetation and their influence on remote sensing. In *Land Observation by Remote Sensing*: *Theory and Applications,* vol. 3, H. J. Buiten, and J. G. P. W. Clevers (eds), 89–106 (Amsterdam, the Netherlands. Gordon and Breach Science Publishers).

de Jong, S. M., 1998. Imaging spectrometry for monitoring tree damage caused by volcanic activity in the Long Valley Caldera, California. *ITC Journal*, 1, 1–10.

Delalieux, S. Somers, B., Verstraeten, W. W., Keulemans, W. and Copping, P., 2008. Hyperspectral canopy measurements under artificial illumination. *International Journal of Remote Sensing*, 29, 6051–6058.

Demir, B. and Ertürk, S., 2008. Phase correlation based redundancy removal in feature weighting band selection for hyperspectral images. *International Journal of Remote Sensing*, 29, 1801–1807.

Díaz-Uriarte, R. and de Andrés, S. A., 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, available online at:

http://www.biomedcentral.com/content/pdf/1471-2105-7-3.pdf (accessed 24 April 2009).

Doraiswamy, P. C., Hatfield, J. L., Jackson, T. J., Akhmedov, B., Prueger, J. and Stern, A., 2004. Crop condition and yield simulations using Landsat and MODIS. *Remote Sensing of Environment*, 92, 548–559.

Doraiswamy, P. C., Sinclair, T. R., Hollinger, S. Akhmedov, B., Stern, A. and Prueger, J., 2005. Application of MODIS derived parameters for regional crop yield assessment. *Remote Sensing of Environment*, 97, 192–202.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7, 1–26.

Enclona, E. A. Thenkabail, P. S., Celis, D. and Diekmann, J., 2004. Within-field wheat yield prediction from IKONOS data: a new matrix approach. *International Journal of Remote Sensing*, 25, 377–388.

ENVI, 2006. Environment for Visualising Images. Release 4.3 (Boulder, USA. ITT industries, Inc)

Epstein, M. and A. J. Bloom, 2005. *Mineral Nutrition of Plants: Principles and Perspectives*, 2nd edition (Sunderland, Massachusetts, USA. Sinauer Associates, Inc), 400 pp.

Estep, L., Terrie, G. and Davis, B., 2004. Crop stress detection using AVIRIS hyperspectral imagery and artificial neural networks. *International Journal of Remote Sensing*, 25, 4999–5004.

Evans, L. T., 1993. *Crop Evolution, Adaptation and Yield* (New York, USA. Cambridge University Press), 512 pp.

Everingham, Y. L., Inman-Bamber, N. G., Thorburn, P. J. and McNeill, T. J., 2007b. A bayesian modelling approach for long leads sugarcane yield forecasts for the Australian sugar industry. *Australian Journal of Agricultural Research*, 58, 87–94.

Everingham, Y. L., Lowe, K. H., Donald, D. A., Coomans, D. H. and Markley, J., 2007a. Advanced satellite imagery to classify sugarcane crop characteristics. *Agronomy for Sustainable Development*, 27, 111–117.

Everingham, Y. L., Smyth, C. W. and Inman-Bamber, N. G., 2009. Ensemble data mining approaches to forecast regional sugarcane crop production. *Agriculture and Forest Meteorology*, 149, 689–696.

Everingham, Y., Inman-Bamber, G., Ticehurst, C., Barrett, D., Lowe, K. and Mcneill, T., 2005. Yield forecasting for marketers. *Proceedings of Australian Society of Sugar Cane Technologists Conference*, 1–10.

Faber, N. M. and Rajkó, R., 2007. How to avoid overfitting in multivariate calibration – The conventional validation approach and an alternative. *Analytica Chimica Acta*, 595, 98–106.

Fageria, N. K., 2009. *The Use of Nutrients in Crop Plants* (New York, USA. Taylor and Francis Group), 430 pp.

Ferencz, C., Bognár, P., Lichtenberger, J. Hamar, D. Tarcsai, G., Timár, G., Molnár, G., Pásztor, S. Steinbach, P. Székely, B. Ferencz, O. E. and Ferencz-Árkos, I., 2004. Crop yield estimation by satellite remote sensing. *International Journal of Remote Sensing*, 20, 4113–4149.

Ferrio, J. P., Villegas, D., Zarco, J. Aparicio, N., Araus, J. L. and Royo, C., 2005. Assessment of durum wheat yield using visible and near infrared reflectance spectra of canopies. *Field Crops Research*, 94, 126–148.

Ferwerda, J., Skidmore, A. K. and Mutanga, O., 2005. Nitrogen detection with hyperspectral normalized ratio indices across multiple plant species. *International Journal of Remote Sensing*, 26, 4083–4095.

Fortes, C. and Demattê, J. A. M., 2006. Discrimination of sugarcane varieties using Landsat 7 ETM+ spectral data. *International Journal of Remote Sensing*, 27, 1395–1412.

Fuqin, L. and Guoliang, T., 1991. Research on remote sensing-meteorological model for wheat yield estimation. *Asian Conference on Remote Sensing*, available online at: http://www.aars-acrs.org/acrs/proceeding/ACRS1991/Papers/AGV91-4.htm (accessed 28 April 2009).

Galvão, L. S., Formaggio, A. R. and Tisot, D. A., 2005. Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. *Remote Sensing of Environment*, 94, 523–534.

Galvão, L. S., Formaggio, A. R. and Tisot, D. A., 2006. The influence of spectral resolution on discriminating Brazilian sugarcane varieties. *International Journal of Remote Sensing*, 24, 769–777.

Gamon, J. A. and Surfus, J. S., 1999. Assessing leaf pigment content and activity with a reflectometer. *New Phytologist*, 143, 105–117.

Gamon, J. A., Peñuelas, J. and Field, C. B., 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment*, 41, 35–44.

Gao, B., 1996. NDWI– a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58, 257–266.

Gao, X., Huete, A. R., Ni, W. and Miura, T., 2000. Optical–biophysical relationships of vegetation spectra without background contamination. *Remote Sensing of Environment*, 74, 609–620.

Genc, H., Genc, L., Turhan, H., Smith, S. E. and Nation J. L., 2008. Vegetation indices as indicators of damage by the sum pest (Hemiptera: Scutelleridae) to field grown wheat. *African Journal of Biotechnology*, 7, 173–180.

Gers, C. J. and Schmidt, E. J., 2001. Using SPOT 4 satellite imagery to monitor areas harvested by small scale sugarcane farmers at Umfolozi. *Proceedings of the South African Sugar Technologists' Association*, 45, 158–164.

Gers, C. J., 2003a. Relating remotely sensed multi-temporal Landsat 7 ETM+ imagery to sugarcane characteristics. *Proceedings of the South African Sugar Technologists' Association*, 77, 313–321.

Gers, C. J., 2003b. Remotely sensed sugarcane phenological characteristics at Umfolozi South Africa. *Proceedings of the 2003 IEEE International Geoscience and Remote Sensing Symposium*, 1010–1012.

Gers, C. J., 2004. Applications of remote sensing in sugarcane agriculture at Umfolozi, South Africa. MSc thesis, School of Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa.

Gianelle, D. and Vescovo, L., 2007. Determination of green herbage ratio in grasslands using spectral reflectance. Methods and ground measurements. *International Journal of Remote Sensing*, 28, 931–942.

Gislason, P. O., Benediktsson, J. A. and Sveinsson, J. R., 2006. Random forest for land cover classification. *Pattern Recognition Letters*, 27, 294–300.

Gitelson, A. A. and Merzlyak, M. N., 1994. Spectral reflectance changes associate with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*, 143, 286–292.

Gitelson, A. A., Kaufman, Y. J. and Merzlyak, M. N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58, 289–298.

Goel, N. S. and Qin, W., 1994. Influences of canopy architecture on relationships between various vegetation indices and LAI and fPAR: a computer simulation. *Remote Sensing Reviews*, 10, 309–347.

Gong, P., Pu, R., Biging, G. S. and Larrieu, M. R., 2003. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 1355–1362.

Graeff, S. and Claupein, W., 2003. Quantifying nitrogen status of corn (*Zea mays* L.) in the field by reflectance measurement. *European Journal of Agronomy*, 19, 611–618.

Graham, P. H., Lennè, J. M. and Chatel, D., 2005. Sugarcane physiology: integrating from cell to crop to advance sugarcane production. *Field Crops Research*, 92, 115–117.

Grof, C. P. L. and Campbell, J. A., 2001. Sugarcane sucrose metabolism: scope for molecular manipulation. *Australian Journal of Plant Physiology*, 28, 1–12.

Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157–1182.

Guyot, G., 1990. Optical properties of vegetation canopies. In *Application of Remote Sensing in Agriculture*, M. D. Steven and J. A. Clark (eds), pp. 19–44 (London, UK. Butterworths).

Ha, N. T. T., 2005. Spatial and temporal unmixing of coarse resolution polar orbiter observations for sugarcane yield assessments: a seasonal integration method based on MODIS and MISR fPAR estimates in Xinavane, Mozambique. MSc

thesis, International Institute for Geo-information Science and Earth Observation, ITC, Enschede, the Netherlands.

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J. and Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, 81, 416–426.

Hadsarang, W. and Sukmuang, S., 2000. Utilization of Landsat-5 TM imagery for sugarcane area survey and mapping in Thailand. *Proceedings of the Asian Conference on Remote Sensing*, Available online at: www.gisdevelopment.net/aars/acrs/2000/ps2/ps214pf.htm (accessed 16 January 2007).

Hair, J. F., Black, W. C., Babin, B., Anderson, R. E. and Tatham, R. L., 2006. *Multivariate Data Analysis*, 6th edition (New Jersey, Pearson Education, Inc), 899 pp.

Hansen, P. M. and Schjoerring, J. K., 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86, 542–553.

Hassan, S. F., 2008. Development of sugar industry in Africa. *Sugar Tech*, 10, 197–203.

Hecker, J. H., 2003. Investigation of the relationship between chlorophyll concentration and high resolution data of *Phragmites australis* in heavy metal contaminated sites. MSc Thesis, International Institute for Geo-information Science and Earth Observation, ITC, Enschede, The Netherlands.

Huang, W., Lamb, D. W., Niu, Z. and Zhang, Y., 2007. Identification of yellow rust in wheat using *in situ* spectral reflectance measurements and airborne hyperspectral imaging. *Precision Agriculture*, 8, 187–197.

Huang, Z., Turner, B. J., Dury, S. J., Wallis, I. R. and Foley, W. J., 2004. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sensing of Environment*, 93, 18–29.

Huete, A. R., 1988. A Soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25, 295–309.

Hunt, E. R. and Rock, B. N., 1989. Detection of changes in leaf water content using near- and middle-infrared reflectance. *Remote Sensing of Environment*, 30, 43–54.

Idso, S. B. Pinter, P. J. Jackson Jr, R. D. and Reginato, R. J., 1980. Estimation of grain yields by remote sensing of crop senescence rates. *Remote Sensing of Environment*, 9, 87–91.

IKONOS, 2005. The project, available online at: www.ikonosheritage.org/project/ (accessed 29 January 2007).

Inman-Bamber, N. G. and Smith, D. M., 2005. Water relations in sugarcane and response to water deficits. *Field Crops Research*, 92, 185–202.

Inoue, Y. and Peñuelas, J., 2001. An AOTF-based hyperspectral imaging system for field use in ecological and agricultural applications. *International Journal of Remote Sensing*, 22, 3883–3888.

Isa, D.W., Hofman, G. and van Cleemput, O., 2006. Uptake and balance of fertilizer nitrogen applied to sugarcane. *Field Crops Research*, 95, 348–354.

Ismail, R., 2009. Remote sensing of forest health: the detection and mapping of *Pinus patula* trees infested by *Sirex noctilio*. PhD thesis, School of Environmental Sciences, University of KwaZulu-Natal, Pitermaritzburg, South Africa.

Jackson, R. D. and Huete, A., 1991. Interpreting vegetation indices. *Preventive Veterinary Medicine*, 11, 185–200.

Jackson, R. D., Jones, C. A., Uehara, G. and Santo, L. T., 1981. Remote detection of nutrient and water deficiencies in sugarcane under variable cloudiness. *Remote Sensing of Environment*, 11, 327–337.

Jacquemoud, S., Verdebout, J., Schmuck, G., Andreoli, G. and Hosgood, B., 1995. Investigation of leaf biochemistry by statistics. *Remote Sensing of Environment*, 58, 180–188.

Jain, N., Ray, S. S., Sinph, J. P. and Panigrahy, S., 2007. Use of hyperspectral data to assess the effects of different nitrogen applications on a potato crop. *Precision Agriculture*, 8, 225–239.

Jarmer, T. Lilienthal, H. and Udelhoven, T., 2003. Spectral determination of nitrogen content of wheat canopies. *Proceedings of 3rd EARseL Workshop on Imaging Spectroscopy*, 513–517.

Jiang, D. Yang, X., Clinton, N. and Wang, N., 2004b. An artificial neural network model for estimating crop yields using remotely sensed information. *International Journal of Remote Sensing*, 25, 1723–1732.

Jiang, X., Tang, L., Wang, C. and Wang, C., 2004a. Spectral characteristics and feature selection of hyperspectral remote sensing data. *International Journal of Remote Sensing*, 25, 51–59.

Jiao, H. B., Zha, Y., Gao, J., Li, Y. M., Wei, Y. C. and Huang, J. Z., 2006. Estimating of chlorophyll-a concentration in lake Tai, China using *in situ* hyperspectral data. *International Journal of Remote Sensing*, 27, 4267–4276.

Johnson, A. K. L. and Kinsey-Henderson, A. E., 1997. Satellite-based remote sensing for monitoring land use in the sugar industry. *Proceedings of the Australian Society of Sugar Cane Technology*, 19, 237–245.

Johnson, R. A. and Wichern, D. W., 2002. *Applied Multivariate Statistical Analysis*, 5th edition (New Jersey, USA. Pearson Education, Inc), 767 pp.

Johnson, R. M., Viator, R. P., Veremis, C. J., Richard, E. P. and Zimba, P., 2005. Discrimination of sugarcane varieties with hyperspectral reflectance measurements and plant pigment analysis. *Journal of the American Society of Sugar Cane Technologists*, 25, 111–112.

Jones, H. G. and Schofield, P., 2008. Thermal and other remote sensing of plant stress. *General and Applied Plant Physiology*, Special Issue, 34, 19–32.

Jongschaap, R. E. E. and Booij, R., 2004. Spectral measurements at different spatial scales in potato: relating leaf, plant and canopy nitrogen status. *International Journal of Applied Earth Observation and Geoinformation*, 5, 205–218.

Jordan, C. F., 1969. Derivation of leaf area index from quality of light on the forest floor. *Ecology*, 50, 663–666.

Jørgensen, R. N., Christensen, L. K. and Bro, R., 2007. Spectral reflectance at sub-leaf scale including the spatial distribution discriminating NPK stress characteristics in barely using multiway partial least squares regression. *International Journal of Remote Sensing*, 28, 943–962.

Jørgensen, R. N., Hansen, P. M. and Bro, R., 2006. Exploratory study of winter wheat reflectance during vegetative growth using three-mode component analysis. *International Journal of Remote Sensing*, 27, 919–937.

Kalubarme, M. H., Potdar, M. B., Manjunath, K. R. Mahey, R. K. and Siddhu, S. S., 2003. Growth profile based crop yield models: a case study of large area wheat yield modelling and its extendibility using atmospheric corrected NOAA AVHRR data. *International Journal of Remote Sensing*, 24, 2037–2054.

Kastens, J. H., Kastens, T. L., Kastens, D. L. A., Price, K. P., Martinko, E. A. and Lee, R., 2005. Image masking for crop yield forecasting using AVHRR NDVI time series imagery. *Remote Sensing of Environment*, 99, 341–356.

Keeping, M., Butterfield, M., Leslie, G. and Rutherford, S., 2008. Initial measurements for management of thrips. *The Link*, 17, 4–5.

Kempeneers, P., Zarco-Tejada, P. J., North, P. R. J., de Backer, S., Delalieux, S., Sepulcre-Cantó, G., Morales, F., van Aardt, J. A. N., Sagardoy, R., Coppin, P. and Scheunders, P., 2008. Model inversion for chlorophyll estimation in open canopies from hyperspectral imagery. *International Journal of Remote Sensing*, 29, 5093–5111.

Kerekes, J. P. and Schott, J. R., 2007. Hyperspectral imaging systems. In *Hyperspectral Data Exploitation: Theory and Applications*, Chein-I Chang (ed), pp. 19–46 (Chichester, England. John Wilely and Sons, Inc.).

Kiang, N., Y., Sifert, J., Govindjee and Blankenship, R. E., 2007. Spectral signature of photosynthesis. I. review of earth organisms. *Astrobiology*, 7, 222–251.

Kogan, F. Yang, B., Wei, G., Zhiyuan, P. and Xianfeng, J., 2005. Modelling corn production in China using AVHRR-based vegetation health indices. *International Journal of Remote Sensing*, 26, 2325–2336.

Kohavi, R. and John, G. H., 1997. Wrappers for feature subset selection. *Artificial Intelligence*, 97, 273–324.

Kohavi, R., 1995. A study of cross validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1137–1143.

Kokaly, R. F. and Clark, R. N., 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment*, 67, 267–287.

Kokaly, R. F., 2001. Investigating a physical basis for spectroscopic estimates of leaf nitrogen concentration. *Remote Sensing of Environment*, 75, 153–161.

Kolb, T. E., McCormick, L H. and Shumway, D. L., 1991. Physiological responses of pear thrips-damaged sugar maples to light and water stress. *Tree physiology*, 9, 401–413.

Kruse, J. K., Christians, N. E. and Chaplin, M. H., 2006. Remote sensing of nitrogen stress in creeping bentgrass. *Agronomy Journal*, 98, 1640–1645.

Kumar, L., Schmidt, K., Dury, S. and Skidmore, A., 2003. Imaging spectrometry and vegetation science. In *Image Spectrometry*, vol. 3, F. D. van der Meer and S. M. de Jong (eds), pp. 111–156 (London, UK. Kluwer Academic Publishers).

Labus, M. P., Nielsen, G. A., Lawrence, R. L. and Engel, R., 2002. Wheat yield estimates using multi-temporal NDVI satellite imagery. *International Journal of Remote Sensing*, 23, 4169–4180.

Lawrence, R. L., Wood, S. D. and Sheley, R. L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment*, 100, 356–362.

Leco, 2006. TruSpec N® analyser, FP-2000, (St Joseph Missouri, USA. Leco Corporation).

Lee-Lovick, G. and Kirchner, L., 1991. Limitations of Landsat TM data in monitoring growth and predicting yields in sugarcane. *Proceedings of the Australian Society of Sugar Cane Technology*, 13, 124–129.

Leslie, G. and Donaldson, R., 2005. Scorched and yellow leaves in sugarcane. *The Link*, 14, 1–1.

Leslie, G., 2004. Pests of sugarcane. In *Sugarcane,* 2nd edition, G. James (ed), pp. 78–100 (Oxford, UK. Blackwell Science Ltd).

Leslie, G. W., 2005. Thrips: a new pest of sugarcane in South Africa. *The Link,* 14, 2–3.

Leslie, G. W., 2006. Thrips. *The Link*, 15, 11–11.

Lewis, J. E., Rowland, J. and Nadeau, A., 1998. Estimating maize production in Kenya using NDVI: some statistical considerations. *International Journal of Remote Sensing*, 19, 2609–2617.

Liaw, A. and Wiener, M., 2002. Classification and regression by random forest. *R News*, 2/3, 18–22.

Lillesand, T. M. and Kiefer, R. W., 2001. *Remote Sensing and Image Interpretation*, 4th edition, (New York, USA. John Wiley and Sons, Inc.), 724 pp.

Liu, L., Wang, J., Bao, Y., Huang, W., Ma, Z. and Zhao, C., 2006. Predicting winter wheat condition, grain yield and protein content using multi-temporal EnviSat-ASAR and Landsat TM satellite images. *International Journal of Remote Sensing*, 27, 737–753.

Lobell, D. B., Ortiz-Monasterio, J. I., Asner, G. P., Naylor, R. L. and Falcon, W. P., 2005. Combining field surveys, remote sensing and regression trees to understand yield variation in an irrigated wheat landscape. *Agronomy Journal*, 97, 241–249.

Lorenzen, B. and Jensen, A., 1989. Changes in leaf spectral properties induced in barley by cereal powdery mildew. *Remote Sensing of Environment*, 27, 201–209.

Loveland, T. R., Bouchard, M. A., Irons, J. R. and Woodcock, C., 2008. Landsat science team meeting summary. *The Earth Observer*, 20, 23–27.

Luther, J. E. and Carroll, A. L., 1999. Development of an index of balsam fir vigour by foliar spectral reflectance. *Remote Sensing of Environment*, 69, 241–252.

Ma, L., Ahuja, L. R. and Bruulsema, T. W., 2009. Current status and future needs in modelling plant nitrogen uptake: A preface. In *Quantifying and Understanding Plant Nitrogen Uptake for Systems Modelling*, L. Ma, L. R. Ahuja and T. Bruulsema (eds), pp. 1–11 (New York, USA. Taylor and Francis Group).

Main, R., Cho, M., van Aardt, J. and Majeke, B., 2008. A comparison between sensors with different spectral resolutions, relative to the Sumbandila satellite, for assessing site quality differences, in *Eucalyptus grandis* plantations. *Proceedings of 14 Australasian Remote Sensing and Photogrammetry Conference*, available online at: http://hdl.handle.net/10204/3037

Maindonald, J. and Braun, J., 2006. *Data Analysis and Graphics Using R, an Example-Based Approach*, 2nd edition, R. Gill, B. D. Ripley, S. Ross, B. W. Silverman and M. Stein (eds), (Cambridge, UK. Cambridge University Press), 540 pp.

Malthus, T. J. and Madeira, A. C., 1993. High resolution spectroradiometry: spectral reflectance of field bean leaves infected by *Botrytis fabae*. *Remote Sensing of Environment*, 45, 107–116.

Markley, J., Raines, A. and Crossley, R., 2003. The development and integration of remote sensing, GIS and data processing tools for effective harvest management. *Proceedings of the Australian Society of Sugar Cane Technology*, 25, 5–5.

Marschner, H., 1993. *Mineral Nutrition of Higher Plants* (London, UK. Academic Press), 889 pp.

Martin, M. E., Plourde, L. C., Ollinger, S. V., Smith, M-L. and McNeil, B. E., 2008. A generalisable method for remote sensing of canopy nitrogen across a wide range of forest ecosystems. *Remote Sensing of Environment*, 112, 3511–3519.

Maselli, F., Romanelli, S. Bottai, L. and Maracchi, G., 2000. Processing of GAC NDVI data for yield forecasting in the Sahelian region. *International Journal of Remote Sensing*, 21, 3509–3523.

Masoood, M. A. and Javed, M. A., 2004. Forecast models for sugarcane in Pakistan. *Pakistan Journal of Agricultural Sciences*, 41, 80–85.

Mauser, W. and Bach, H., 1994. Imaging spectroscopy in hydrology and agriculture – determination of model parameters. In *Imaging Spectroscopy – a Tool for Environmental Observations*, vol. 4, J. Hill and J. Mégier (eds), pp. 261–283 (New York, USA. Springer Publishing Company).

McCarthy, J., 2007. Interactive report: generic economic and social impact of the sugar industry in the context of milling areas. The South African Sugar Industry.

Mevik, B. and Wehrens, R., 2007. The pls package: Principal components and partial least squares regression in R. *Journal of Statistical Software*, 18, 1–24.

Meyer, J. H., 1983. Rapid determination of nitrogen in cane leaves. *Proceedings of the South African Sugar Technologists' Association*, 57, 109–112.

Meyer, J. H., Wood, R. A. and Leibbrandt, N. B., 1986. Recent advances in determining the N requirement of sugarcane in the South African sugar industry. *Proceedings of the South African Sugar Technologists' Association*, 60, 205–210.

Mirik, M., Michels Jr, G. J., Kassymzhanova-Mirik, S. and Ellott, N. C., 2007. Reflectance characteristics of Russian wheat aphid (Hemiptera: Aphididae) stress and abundance in winter wheat. *Computers and Electronics in Agriculture*, 57, 123–134.

Mirik, M., Michels Jr, G. J., S. Kassymzhanova-Mirik, Elliott, N. C., Catana, V., Jones, D. B. and Bowling, R., 2006a. Using digital image analysis and spectral reflectance data to quantify damage by greenbug (Hemitera; Aphididae) in winter wheat. *Computers and Electronics in Agriculture*, 51, 86–98.

Mirik, M., Michels Jr, G. J., Kassymzhanova-Mirik, S., Elliott, N. C. and Catana, V., 2006b. Spectral sensing of aphid (Hemipters: Aphididae) density using field spectrometry and radiometry. *Turkish Journal of Agriculture and Forestry*, 30, 421–428.

Mkhabela, M. S., Mkhabela, M. S. and Mashinini, N. N., 2005. Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data derived from NOAA's-AVHRR. *Agriculture and Forest Meteorology*, 129, 1–9.

Moran, S., Fitzgerald, A., Rango, A., Walthall, C., Barnes, E., Bausch, W., Clarke, T., Daughtry, C., Everitt, J., Escobar, D., Hatfield, J., Havstad, K., Jackson, T., Nitchen, N., Kustas, W., Mcguire, M., Pinter Jr, P., Sudduth, K.., Schepers, J., Schmugge, T., Starks, P. and Upchurch, D., 2003. Sensor development and radiometric correction for agricultural applications. *Photogrammetric Engineering and Remote Sensing*, 69, 705–718.

Moshou, D., Bravo, C., West, J., Wahlen, S., Mccartney, A. and Ramon, H., 2004. Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks. *Computers and Electronics in Agriculture*, 44, 173–188.

Mostert, S., Steyn, H., Burger, H. and Bosman, H., 2008. Sumbandilasat–an operational technology demonstrator. *Acta Astronautica*, 63, 1273–1282.

Motulsky, H. and Christopoulos, A., 2003. *Fitting Models to Biological Data Using Linear and Non-linear Regression. A Practical Guide to Curve Fitting*, 4th version (San Diego, USA. GraphPad Software, Inc), 351 pp.

Mound, L. A. and Teulon, D. A. J., 1995. Thysanoptera as phytophagous opportunists. In *Thrips Biology and Management*, B. L. Parker, M. Skinner and T. Lewis (eds), pp. 3–20 (New York, USA. Plenum Press).

Muchovej, R. M., Newman, P. R. and Luo, Y., 2005. Sugarcane leaf nutrient concentrations: with or without midrib tissue. *Journal of Plant Nutrition*, 28, 1271–1286.

Muchow, R. C., Robertson, M. J., Wood, A. W. and Keating, B. A., 1996. Effect of nitrogen on the time-course of sucrose accumulation in sugarcane. *Field Crops Research*, 47, 143–153.

Munden, R., Curran, P. J. and Catt, J. A., 1994. The relationship between red edge and chlorophyll concentration in the Broadbalk winter wheat experiment at Rothamsted. *International Journal of Remote Sensing*, 15, 705–709.

Mutanga, O. and Kumar, L., 2007. Estimating and mapping grass phosphorus concentration in an African savanna using hyperspectral image data. *International Journal of Remote Sensing*, 28, 4897–4911.

Mutanga, O. and Skidmore, A. K., 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25, 1–16.

Mutanga, O., 2004. Hyperspectral remote sensing of tropical grass quality and quantity, PhD thesis, ITC, University of Wageningen, Enschede, The Netherlands.

Mutanga, O., Skidmore, A. K. and Prins, H. H. T., 2004. Predicting *in situ* pasture quality in Kruger national park, South Africa, using continuum-removed absorption features. *Remote Sensing of Environment*, 89, 393–408.

Mutanga, O., Skidmore, A. K. and van Wieren, S., 2003. Discriminating tropical grass (*Cenchrus ciliaris*) canopies grown under different nitrogen treatments using spectroradiometry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57, 263–272.

Nansen, C., Macedo, T., Swanson, R. and Weaver, D. K., 2009. Use of spatial structure analysis of hyperspectral data cubes for detection of insect-induced stress in wheat plants. *International Journal of Remote Sensing*, 30, 2447–2464.

Narciso, G. and Schmidt, E. J., 1999. Identification and classification of sugarcane based on satellite remote sensing. *Proceedings of the South African Sugar Technologists' Association*, 73, 189–194.

NASA, 2006. Science missions, available online at: http://science.hq.nasa.gov/ (accessed 16 January 2007).

Nguyen, H. T. and Lee, B., 2006. Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression. *European Journal of Agronomy*, 24, 349–356.

Nguyen, H. T., Kim, J. H., Nguyen, A. T., Nguyen, L. T., Shin, J. C. and Lee, B-W., 2006. Using canopy reflectance and partial least squares regression to calculate within-field statistical variation in crop growth and nitrogen status of rice. *Precision Agriculture*, 7, 249–264.

Numata, I., Roberts, D. A., Chadwick, O. A., Schimel, J. P., Galvão, L. S. and Soares, J. V., 2008. Evaluation of hyperspectral data for pasture estimate in the Brazilian Amazon using field imaging spectrometers. *Remote Sensing of Environment*, 112, 1569–1583.

OGTR, 2004. The biology and ecology of sugarcane (*Saccharum* spp. hybrids) in Australia. A resource document, Office of the Gene Technology Regulator, Department of Health and Ageing, Australia, available online at: www.ogtr.gov.au/pdf/ir/biologysugarcane. pdf (accessed 15 February 2006).

Oppelt, N. and Mauser, W., 2004. Hyperspectral monitoring of physiological parameters of wheat during a vegetation period using AVIS data. *International Journal of Remote Sensing*, 25, 145–159.

Osborne, S. L., Schepers, J. S., Francis, D. D. and Schlemmer, M. R., 2002. Detection of phosphorus and nitrogen deficiencies in corn using spectral radiance measurements. *Agronomy Journal*, 94, 1215–1221.

Pal, M., 2005. Random forest classifier for remote sensing classification, *International Journal of Remote Sensing,* 26, 217–222.

Palmer, D. S., O'Boyle, N. M., Glen, R. C. and Mitchell, B. O., 2007. Random forest models to predict aqueous solubility. J*ournal of Chemical Information and Modelling*, 47, 150–158.

Parker, S. R., Shaw, M. W. and Royle, D. J., 1995. The reliability of visual estimates of disease severity on cereal leaves. *Plant Pathology*, 44, 856–864.

Patel, J. K., Patel, N. M. and Shiyani, R. L., 2001. Coefficient of variation in field experiments and yardstick thereof – an empirical study. *Current Science*, 81, 1163–1164.

Pemberton, C. E. and Williams, J. R., 1969. Distribution, origins and spread of sugarcane pests. In: *Pests of Sugarcane,* J. R. Williams, J. R. Metcalfe, R. W. Mungomery and R. Mathes (eds), pp. 1–9 (Amsterdam, The Netherlands. Elsevier).

Peñuelas, J., Baret, F. and Filella, I., 1995. Semi-empirical indices to assess carotenoides/ chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, 31, 221–230.

Pinter, P. J., Hatfield, J. L., Schepers, S. J., Barnes, M. E., Moran, M. S., Daughtry, C. S. T. and Upchurch, R. D., 2003. Remote sensing for crop management. *Photogrammetric Engineering and Remote Sensing*, 69, 647–664.

Prasad, A. M., Iverson, L. R. and Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, 9, 181–199.

Pu, R., Foschi, L. and Gong, P., 2004. Spectral feature analysis for assessment of water status and health level in coast live oak (*Quercus agrifolia*) leaves. *International Journal of Remote Sensing*, 25, 4267–4286.

Qifa, Z. and Jihua, W., 2003. Leaf and spike reflectance spectra of rice with contrasting nitrogen supplemental levels. *International Journal of Remote Sensing*, 24, 1587–1593.

R Development Core Team, 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, available online at: http://www.R-project.org.

RADARSAT-2, 2006. A new era in synthetic aperture radar, available online at: www.radarsat2.info/ (accessed 29 January 2007).

Rao, P. V. K., Rao, V. V. and Venkataratnam, L., 2002. Remote sensing: a technology for assessment of sugarcane crop acreage and yield. *Sugar Technology International*, 4, 97–101.

Ray, S. S., Pokharna, S. S. and Ajai, 1999. Cotton yield estimation using agro-meteorological model and satellite-derived spectral profile. *International Journal of Remote Sensing*, 20, 2693–2702.

Read, J. J., Tarpley, L., Mckinion, J. M. and Reddy, K. R., 2002. Narrow-waveband reflectance ratios for remote estimation of nitrogen status in cotton. *Journal of Environmental Quality*, 31, 1442–1452.

Ren, J., Chen, Z., Zhou, Q. and Tang, H., 2008. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *International Journal of Applied Earth Observation and Geoinformation*, 10, 403–413.

Reunanen, J., 2003. Overfitting in making comparisons between variables selection methods. *Journal of Machine Learning Research*, 3, 1371–1382.

Rezaei, Y., Mobasheri, M. R. and Valadan Zoej, M. J., 2008. Unsupervised information extraction using absorption line in Hyperion images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII, 383–388.

Rice, R. W., Gibert, R. A. and Lentini, R. S., 2006. Nutritional requirements for Florida sugarcane. Institute of Food and Agricultural Sciences, University of Florida Document No. SS-AGR-228 (Florida, USA: IFAS, UF).

Richards, R. A., 2000. Selectable traits to increase crop photosynthesis and yield of grain crops. *Journal of Experimental Botany*, 51, 447–458.

Riedell, W. E. and Blackmer, T. M., 1999. Leaf reflectance of cereal aphid-damaged wheat. *Crop Science*, 39, 1835–1840.

Roth, G. W., Fox, R. H. and Marshall, H. G., 1989. Plant tissue tests for predicting nitrogen fertilizer requirements of winter Wheat. *Agronomy Journal*, 81, 502–507.

Roujean, J. and Breon, F., 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, 51, 375–384.

Rouse Jr., J. W., Haas, R. H., Schell, J. A. and Deering, D. W., 1973. Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings of Third Earth Resources Technology Satellite-1 Symposium*, I, 309–317.

Rudorff, B. F. T. and Batista, G. T., 1990. Yield estimation of sugarcane based on agro-meteorological-spectral models. *Remote Sensing of Environment*, 33, 183–192.

Salazar, L., Kogan, F. and Roytman, L., 2008. Using vegetation health indices and partial least squares method for estimation of corn yield. *International Journal of Remote Sensing*, 28, 175–189.

SASA, 2009. Industry overview. South African Sugar Association. Available online at: http://www.sugar.org.za/IndustryOverview43.aspx (accessed 6 August 2009).

SASRI, 2000, Recommendations for nitrogen (N). South African Sugarcane Research Institute, Information sheet.

SASRI, 2003. Leaf sampling, South African Sugarcane Research Institute, Information sheet.

SASRI, 2006. Varieties. South African Sugarcane Research Institute, Information Sheet.

Sawasawa, L. A., 2003. Crop yield estimation: integrating RS, GIS and management factors. A case study of Birkoor and Kortigiri Mandals-Nzamabad district, India. MSc Thesis, International Institute for Geo-information Science and Earth Observation, ITC, The Netherlands.

Schmidt, E. J., Gers, C. J., Narciso, G. and Frost, P., 2001. Remote sensing in the South African sugar industry. *Proceedings of the International Society of Sugar Cane Technologists*, 24, 241–245.

Schmidt, E. J., Narciso, G., Frost, P. and Gers, C. J., 2000. Application of remote sensing technology in the South African sugar industry: a review of recent research findings. *Proceedings of the South African Sugar Technologists' Association*, 74, 192–201.

Schmidt, K. S. and Skidmore, A. K., 2003. Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, 85, 92–108.

Serrano, L., Peñuelas, J. and Ustin, S. L., 2002. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: decomposing biochemical from structural signals. *Remote Sensing of Environment*, 81, 355–364.

Shafri, H. Z. M., Salleh, M. A. M. and Ghiyamat, A., 2006. Hyperspectral remote sensing of vegetation using red edge position techniques. *American Journal of Applied Sciences*, 3, 1864–1871.

Shen, S. S., 2007. Optimal band selection and utility evaluation for spectral systems. In *Hyperspectral Data Exploitation: Theory and Applications*, C-I. Chang (ed), pp. 227–243 (New York, USA. John Wiley and Sons, Inc.).

SIC, 2007. QuickBird satellite images and sensor specifications: satellite image corporation, available online at: www.satimagingcorp.com/gallery-quickbird.html (accessed 29 January 2007).

Simões, M. D. S., Rocha, J. V. and Lamparelli, R. A. C., 2005. Spectral variables, growth analysis and yield of sugarcane. *Scientia Agricola* (Piracicaba, Brazil), 62, 199–207.

Sims, D. A. and Gamon, J. A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and development stages. *Remote Sensing of Environment*, 81, 337–354.

Singels, A., Smit, M. A., Redshaw, K. A. and Donaldson, R. A., 2005. The effect of crop start date, crop class and cultivar on sugarcane canopy development and radiation interception. *Field Crops Research*, 92, 249–260.

Singh, R., Semwal, D. P., Rai, A. and Chhikara, R. S., 2002. Small area estimation of crop yield using remote sensing satellite data. *International Journal of Remote Sensing*, 23, 49–56.

Smit, M. A. and Singels, A., 2006. The response of sugarcane canopy development to water stress. *Field Crops Research*, 98, 91–97.

Smith, M-L., Martin, M. E., Plourde, L. and Ollinger, V., 2003. Analysis of hyperspectral data for estimation of temperate forest canopy nitrogen concentration: comparison between an airborne (AVIRIS) and spaceborne (Hyperion) sensor. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 1332–1337.

Smith, R. C. Adams, J. Stephens, D. J. and Hick, P. T., 1995. Forecasting wheat yield in Mediterranean-type environment from NOAA satellite. *Australian Journal of Agricultural Research*, 46, 113–125.

SPSS, 2006. SPSS for Windows. Release 15 (Chicago, USA. Statistical Package for Social Sciences; SPSS Inc.).

Strachan, I. B., Pattey, E. and Boisvert, J. B., 2002. Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance. *Remote Sensing of Environment*, 80, 213–224.

Strobl, C., Boulesteix, A-L., Kneib, T., Augustin, T. and Zeileis, A., 2008. Conditional variable importance for random forests. *BMC Bioinformatics*, available online at: http://www.biomedcentral.com/1471-2105/9/307 (accessed 30 March 2009).

Stroppiana, D., Boschetti, M., Brivio, P. A. and Bocchi, S., 2006. Remotely sensed estimation of rice nitrogen concentration for forcing crop growth models. *Italian Journal of Agrometeorology*, 3, 50–57.

Sudbrink, D. L., Harris, F. A., Robbins, J. T., English, P. J. and Willers, J. L., 2003. Evaluation of remote sensing to identify variability of cotton plant growth and correlation with larval densities of beet armyworm and cabbage looper (Lepidoptera: Noctuidae). *Florida Entomologist*, 86, 290–294.

Tardin, A. T., Assuncão, G. V. and Soares, J. V., 1992. Preliminary analysis of TM data for coffee, citrus and sugarcane discrimination in Furnas region, MG, Brazil. *Pesquisa Agropecuária Brasileira*, 27, 1355–1361.

Tarpley, L., Reddy, K. R. and Sassenrath-Cole, G. F., 2000. Reflectance indices with precision and accuracy in predicting cotton leaf nitrogen concentration *Crop Science*, 40, 1814–1819.

Tejera, N. A., Rodés, R. and Ortega, E., 2007. Comparative analysis of physiological characteristics and yield components in sugarcane cultivars. *Field Crops Research*, 102, 67–72.

Thenkabail, P. S., 2003. Biophysical and yield information for precision farming from near-real-time and historical Landsat TM images. *International Journal of Remote Sensing*, 24, 2879–2904.

Thenkabail, P. S., Enclona, E. A., Ashton, M. S., Legg, C. and de Dieu, M. J., 2004. Hyperion, IKNOS, ALI and ETM+ sensors in the study of African rainforests. *Remote Sensing of Environment*, 90, 23–43.

Thenkabail, P. S., Smith, R. B. and Pauw, E. D., 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, 71, 158–182.

Thomas, V., Treitz, P., Mccaughey, J. H., Noland, T. and Rich, L., 2007. Canopy chlorophyll concentration estimation using hyperspectral and lidar data for a boreal mixedwood forest in northern Ontario, Canada. *International Journal of Remote Sensing*, 29, 1029–1052.

Thompson, G. D., 1991. The growth of sugarcane variety N14 at Pongola. Mount Edgecombe, South African Sugar Association Experiment Station, Research Report Number 7.

Thorburn, P. J., Meier, E. A. and Probert, M. E., 2005. Modelling nitrogen dynamics in sugarcane systems: recent advances and applications. *Field Crops Research*, 92, 337–351.

Tian, Q., Tong, Q., Pu, R., Guo, X. and Zhao, C., 2001. Spectroscopic determination of wheat water status using 1650–1850 nm spectral absorption feature. *International Journal of Remote Sensing*, 22, 2329–2338.

Tilling, A. K., O'Leary, G. J., Ferwerda, J. G., Jones, S. D., Fitzgerald, G. J., Rodriguez, D. and Belford, R., 2007. Remote sensing of nitrogen and water stress in wheat. *Field Crops Research*, 104, 77–85.

Tucker, C. J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, 127–150.

Tulip, J. R. and Wilkins, K., 2005. Application of spectral unmixing to trash level estimation in billet cane. *Proceedings of the Australian Society of Sugar Cane Technology*, 27, 387–399.

Ueno, M., Kawamitsu, Y., Sun, L., Taira, E. and Maeda, K., 2005. Combined applications of NIR, RS and GIS for sustainable sugarcane production. *Sugarcane International*, 23, 8–11.

van den Berg, M., Singles, A., Armitage, R. M., Gillitt, C. G., Way, M. J. and McFarlane, S. A., 2009. South African sugarcane production and quality in the 2008–2009 milling season. *Proceedings of South African Sugar Technologists' Association*, 82, 30–49.

van der Meer, F. D., 2003. Basic physics of spectrometry. In *Image Spectrometry*, vol. 3, F. D. van der Meer and S. M. de Jong (eds), pp. 3−16 (London, UK. Kluwer Academic Publishers).

Vicente-Serrano, S. M., Cuadrat-Prats, J. M. and Romo, A., 2006. Early prediction of crop production using drought indices different time-scale and remote sensing data: application in Ebro Valley (north-east Spain). *International Journal of Remote Sensing*, 27, 511–518.

Vogelmann, J. E. and Rock, B., 1988. Assessing forest damage in high-elevation coniferous forests in Vermont and New Hampshire using Thematic Mapper data. *Remote Sensing of Environment*, 24, 227–246.

Way, M. J., 2008. Arthropods associated with sugarcane leaf spindles in South Africa. *Proceedings of South African Sugar Technologists' Association*, 81, 362–364.

Way, M. J., Leslie, G. W., Keeping, M. G. and Govender, A., 2006. *Fulmekiola serrata* (Kobus) (Thysanoptera: Thripidae), a new pest in southern African sugarcane. *African Entomology*, 14, 401–403.

Wenjiang, H., Jihua, W., Zhijie, W., Jiang, Z., Liangyun, L. and Jindi, W., 2004. Inversion of foliar biochemical parameters at various physiological stages and grain quality indicators of winter wheat with canopy reflectance. *International Journal of Remote Sensing*, 12, 2409–2419.

White, J. D., Trotter, C. M., Brown, L. J. and Scott, N., 2000. Nitrogen concentration in New Zealand vegetation foliage derived from laboratory and field spectrometry. *International Journal of Remote Sensing*, 21, 2525–2531.

Wiedenfeld, R. P., 1995. Effects of irrigation and N fertilizer application on sugarcane yield and quality. *Field Crops Research*, 43, 101−108.

Wiegand, C., Anderson, G., Lingle, S. and Escobar, D., 1996. Soil salinity effects on crop growth and yield: illustration of an analysis and mapping methodology for sugarcane. *Journal of Plant Physiology*, 148, 418–424.

Wold, S., 1995. PLS for multivariate linear modeling. In *Chemometric Methods in Molecular Design*, H. van de Waterbeemd (ed), pp. 195–218 (Weinheim, Gemany. VCH).

Wood, A. W., Muchow, R. C. and Robertson, M. J., 1996. Growth of sugarcane under high input conditions in tropical Australia. III. Accumulation, partitioning and use of nitrogen. *Field Crops Research*, 48, 223–233.

Xavier, A. C., Rudorff, B. F. T., Shimabukuro, Y. E., Berka, L. M. S. and Moreira, M. A., 2006. Multi-temporal analysis of MODIS data to classify sugarcane crop. *International Journal of Remote Sensing*, 27, 755–768.

Xu, H. R., Ying, Y. B., Fu, X. P. and Zhu, S. P., 2007. Near infrared spectroscopy in detecting leaf miner damage on tomato leaf. *Biosystems Engineering*, 96, 447–454.

Xue, L., Cao, W., Luo, W., Dai, T. and Zhu, Y., 2004. Monitoring leaf nitrogen status in rice with canopy spectral reflectance. *Agronomy Journal*, 96, 135–142.

Yang, Z., Rao, M. N., Elliott, N. C., Kindler, S. D. and Popham, T. W., 2005. Using ground-based multispectral radiometry to detect stress in wheat caused by greenbug (Homoptera: Aphididae) infestation. *Computers and Electronics in Agriculture*, 47, 121–135.

Yang, C. and Cheng, C., 2001. Spectral characteristics of rice plants infested by brown planthoppers. *Proceedings of National Science Council, Republic of China. Part B*, 25, 180–186.

Yang, C-M., 2001. Estimation of leaf nitrogen content from spectral characteristics of rice canopy. *Knowledge Bridge*, 18, 1–2.

Yang, X., Zhou, Q. and Melville, M., 1997. Estimating local sugarcane evapotranspiration using Landsat TM image and a VITT concept. *International Journal of Remote Sensing*, 18, 453–459.

Yoder, B. J. and Pettigrew-Crosby, R. E., 1995. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400–2500 nm) at leaf and canopy scales. *Remote Sensing of Environment*, 53, 199–211.

Zhang, M., Chen, R., Luo, J., Lu, J. and Xu, J., 2000. Analyses for inheritance and combining ability of photochemical activities measured by chlorophyll fluorescence in the segregating generation of sugarcane. *Field Crops Research*, 65, 31–39.

Zhang, M., Liu, X. and O'neill, M., 2002. Spectral discrimination of *Phytophthora infestans* infection on tomatoes based on principal component and cluster analyses. *International Journal of Remote Sensing*, 23, 1095–1107.

Zhang, P., Anderson, B. Tan, B., Huang, D. and Myneni, R., 2005. Potential monitoring of crop production using a satellite-based climate-variability impact index. *Agriculture and Forest Meteorology*, 132, 344–358.

Zhao, D., Reddy, K. R., Kakani, V. G. and Reddy, V. R., 2005. Nitrogen deficiency effects on plant growth, leaf photosynthesis, and hyperspectral reflectance properties of sorghum. *European Journal of Agronomy*, 22, 391–403.

Zhao, D., Reddy, K. R., Kakani, V. G., Read, J. J. and Carter, G. A., 2003. Corn (*Zea mays* L.) growth, leaf pigment concentration, photosynthesis and leaf hyperspectral reflectance properties as affected by nitrogen supply. *Plant and Soil*, 257, 205–217.

Zhu, Y., Yao, X., Tian, Y., Liu, X. and Cao, W., 2008. Analysis of common canopy vegetation indices for indicating leaf nitrogen accumulations in wheat and rice. *International Journal of Applied Earth Observation and Geoinformation*, 10, 1–10.