Varietal discrimination and optimal yield prediction of the common dry bean (*Phaseolus vulgaris* L.) grown under different watering regimes using multi-temporal hyperspectral data

Perushan Rajah

A thesis submitted in the fulfilment for the degree of Master of Science in Environmental Science in the School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal

July 2015

Pietermaritzburg

South Africa

Abstract

The common dry bean (*Phaseolus vulgaris* L.) is considered to be amongst the most important food legumes in the world. Due to precision agriculture and in order to attain optimal pre and post-harvest management strategies, precise and timely crop discrimination and yield prediction are of vital importance. Hyperspectral remote sensing in concert with contemporary discriminant analysis and regression analysis, provides the perfect method for on-farm varietal discrimination and yield prediction. This study utilised a combined method of integrating partial least squares discriminant analysis (PLS-DA) on hyperspectral data to determine the optimal period for common dry bean on-farm discrimination. Additionally, a combined method of integrating sparse partial lest squares regression (SPLSR) analysis on hyperspectral data was employed to determine the optimal period for common dry bean on-farm yield predictions. With experimental plots based on irrigated and rain-fed watering treatments, ground-based hyperspectral datasets were collected at three key phenological stages. The three key phenological stages were: (1) germination and stand establishment (two weeks after seed sowing); (2) branching and rapid vegetative growth (six weeks after seed sowing) and (3) flowering and pod development (ten weeks after seed sowing). Results showed that PLS-DA in conjunction with hyperspectral data was able to accurately discriminate between different common dry bean varieties. The growth stages which were selected for optimal varietal discrimination were the branching and rapid vegetative growth stage, with an overall accuracy of 80%, and the flowering and pod development, with an overall accuracy of 100%. Results similarly showed that SPLSR in conjunction with hyperspectral data was able to accurately predict yield of all three common dry bean varieties. Yield prediction models which were developed using data from the flowering and pod formation proved to provide optimal results for all the Caledon (R^2= 0.60 and RMSE= 0.011, 7.48% of the mean) and Ukulinga (R^2= 0.80 and RMSE= 0.013, 7.85% of the mean) varieties. The Gadra variety yield was best predicted at the branching and rapid vegetative growth stage (R^2= 0.74 and RMSE= 0.01, 6.7% of the mean). The study provides valuable information with regard to the optimal period during the life cycle of the common dry bean that should be utilised for discriminatory as well as yield prediction efforts using hyperspectral remote sensing. This as a result, would prove to be invaluable for on-farm and after farm decision making, management practices, field crop sensor development and crop monitoring systems.

Preface

This study was conducted in the School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa, from February 2014 to July 2015, under the supervision of Doctor John Odindi and Doctor Elfatih Mohamed Abdel-Rahman.

I declare that the work presented in this thesis has never been submitted in any form to any other institution. This work represents my original work except where due acknowledgements are made.

Perushan Rajah Signed: ______________ Date: ______________

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

1. Dr. John Odindi Signed: ___________ Date: _______________

2. Dr. Elfatih M. Abdel-Rahman Signed: ____________ Date: _______________

Declaration

I Perushan Rajah, 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 institution.
3. This thesis does not contain other person's 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:

    a. Their words have been re-written and the general information attributed to them has been referenced.
    b. Where their exact words have been used, their writing has been placed in italics 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: _____________

Dedication

To my dear family and friends – for all the support and unwavering faith in me and the pursuit of my Master of Science.

Family is not an important thing. It's everything – Michael J Fox

Acknowledgements

This study would not have been possible without the support and guidance of many individuals who contributed in a substantial manner throughout my Masters programme.

I would like to extend my gratitude to the University of KwaZulu-Natal, for allowing me to pursue my Masters degree. Special thanks to the School of Agricultural, Earth and Environmental Sciences, this study would not have been possible without the funding opportunity that was afforded to me.

To my supervisors, Dr. John Odindi and Dr.Elfatih Mohamed Abdel-Rahman, thank you for believing in me, for the guidance, mentorship and support throughout the duration of my Masters programme. I am grateful to Dr. Kabir Peerbhay for his encouragement and valuable contribution to the statistical analysis of my study. To my co-authors on papers emerging from this study, Prof. Onisimo Mutanga and Prof. Albert Modi, I am very grateful for the scientific contribution.

To my friends, the support and love that was shown towards me during tough times will never be forgotten. Thank you to all my colleagues in the Geography Department for providing an environment which was conducive to research and learning.

I am indebted to my dear mum Shamla, dad Sam and brother Laneel for the love, encouragement and unwavering faith in myself in the pursuit of this Masters degree. A special thank you to my grandparents, Radha and Anthony Pillay, this one is for you.

Lastly, a quote from my favourite evolutionary biologist and role model:

“By all means let's be open-minded, but not so open-minded that our brains drop out” – Sir Richard Dawkins.

Table of contents

List of tables

List of figures

Chapter one

General introduction

1.1 Introduction

The common dry bean (*Phaseolus vulgaris* L.*)* is one of the most important human food legumes in the world (Rubiales et al., 2014). Consumed worldwide, the dry bean is particularly popular in developing countries as a vital source of protein, vitamins and minerals (Fourie, 2002). Whereas it is a common crop in several parts of the African continent, it is a particularly important food crop in eastern and southern Africa, where consumption commonly surpasses 50kg per person per annum (Wortmann et al., 1998). Approximately 32% of the continents bean output is produced in Southern Africa (Wortmann et al., 1998).

Due to the value of the dry bean as a source of food to the local and global economy, reliable on-field mapping is valuable in the growth and production process. Varietal discrimination and yield estimation are particularly important in the overarching goal of precision agriculture and sustainable crop management. Variation within common dry bean varieties can occur due to difference in micro-climate, soil characteristics, phenology and other stresses related to among others, disease and pest infestation and fertilizer treatments. Varietal discrimination on the other hand, could be useful for herbicide application, irrigation regimes and labour-force optimisation. Furthermore, varietal discrimination forms a critical basis for yield estimation, crop destruction risk assessment, quantitative censuses, site variety matching and crop area expansion (Apan et al., 2004; Fortes and Demattê, 2006; Galvão et al., 2005; Mingwei et al., 2008).

Estimation of common dry bean yield is vital for handling, as varieties mature at different times during the cropping season and are characterised by different yield loads. Furthermore, yield estimation is valuable in determining possible production output and therefore commercial profitability. At local and national levels, yield estimation is critical for planning for shortage or surplus.

Commonly, vegetation identification is based on unique phenology and visual percentage cover which are used to identify crop varieties (Govender et al., 2007; Schmidt and Skidmore, 2001). However, these procedures are often labour exhaustive, costly and time consuming, as they require, among others, complicated taxonomical statistics, collateral and ancillary data analysis (Adam and Mutanga, 2009). Whereas use of aerial photography has commonly been adopted, it is often considered costly, due to a high cost per unit area.

Recently, other remotely sensed datasets have emerged as relatively cheaper, practical and economical means to estimate vegetation's bio-physical parameters (Wilson et al., 2014). The use of remotely sensed datasets in vegetation mapping allows for the non-destructive collection of data. Furthermore, possibility for recurrent sampling allows for short-term and long-term data storage, useful for both temporal and multi-temporal varietal identification (Gomez-Casero et al., 2010). According to Wilson et al. (2014) and Govender et al. (2009), remotely sensed datasets offer viable promise for economical, non-destructive and efficient means for estimating biophysical parameters, crop discrimination, classification and yield estimation.

Whereas adoption of remotely sensed datasets has recently gained popularity in mapping agricultural fields (Thenkabail et al., 2004; Zhang et al., 2012), discrimination of varieties has to date been hindered by lack of remotely sensed data of suitable spatial and spectral resolution (Shaw and Burke, 2003; Zwiggelaar, 1998). Spectral discrimination of species, based on the commonly used multispectral data is often challenging as species may produce similar spectral signatures from datasets (Cochrane, 2000; Sobhan, 2007). Characteristically, multispectral remote sensing systems typically use a parallel array of sensors, which are used to differentiate radiation in a small amount of broad wavelength bands (Govender et al., 2007). Therefore, multispectral sensors dilute the electromagnetic spectrum into a small number of broad bands, often inadequate to discriminate individual crop varieties (Govender et al., 2007).

The emergence of hyperspectral sensors of high spectral and spatial resolution has elevated novel prospects for discriminating vegetation species and crop varieties (Schmidt and Skidmore, 2001; Sobhan, 2007). A number of studies (Koedsin and Vaiphasa, 2013; Schmidt and Skidmore, 2001; Wilson et al., 2014) have for instance, achieved success in discriminating vegetation species based on fresh leaf reflectance and field canopy reflectance. Hyperspectral sensors acquire narrow adjacent spectral bands across the visible, near infra-red, mid-infrared and thermal infrared segments of the electromagnetic spectrum (Cody, 2007; Govender et al., 2007). This eliminates any form of averaging or dilution, which could occur to spectral data and therefore promote possibility for crop varietal discrimination.

To date, most studies that have discriminated crop varieties and to estimate yield have been based on once-off data collection at a specific stage during the growing season. However, Sobhan (2007) notes that effective discrimination of varieties and yield estimation might change during the plant's life cycle. Consequently, selection of the optimum phenological stage, based on continuous monitoring, may enhance the spectral separability of varieties, and therefore effective varietal discrimination and yield estimation.

Hyperspectral remote sensing provides a huge amount data which are often difficult to process. Hence, it is essential to execute classification algorithms that are able to utilise the vast amount of data to effectively discriminate crop varieties. One such algorithm is the partial least squares (PLS) (Wold et al., 2001). The algorithm is a proficient multivariate statistical approach that deals with the intricacies related to high dimensional datasets by performing dimension reduction and classification concurrently (Peerbhay et al., 2013a). PLS seeks to reduce the dataset to include only the most important wavebands and combine them into a few components, thus reducing model over-fitting and the exclusion of insensitive wavebands (Abdel-Rahman et al., 2014). Although many studies have used PLS for regression based applications, very few have used it in the discrimination realm (Humphreys et al., 2008; Panneton et al., 2011). Herein lies the opportunity to evaluate the utility of partial least squares, discriminant analysis (PLS-DA) and ground-based hyperspectral remote sensing in classifying different varieties of common dry bean.

According to Sun (2000) progress in the estimation of yield can be divided into a number of stages (1) pre 1940s, where qualitative research by comparing meteorological conditions and crop yield was proposed; (2) during the 1950s, statistical and botanical physiology advanced at a rapid rate, and regression models concerning crop yield and weather conditions were applied; (3) in the 1960s, dynamic simulation models were favoured; (4) in the 1970s, researchers started to use remote sensing methods to predict global crop yield, which ultimately elevated yield estimation knowledge to a higher level and (5) the current stage of improvement, which involves the combination of remote sensing and GIS.

On the other hand, Prasad et al. (2006) notes that numerous methods like neural network, autoregressive state space models, exponential linear crop growth algorithm and numerical crop yield models have been used in crop yield estimation. However, these methods have achieved moderate success. Using remotely sensed data, efforts have been made to develop numerous indices such as, normalized difference vegetation index (NDVI), vegetation condition index (VCI) and temperature condition index (TCI) (Prasad et al., 2006). Although these indices have produced moderately accurate and reliable yield models, factors such as pests, disease, irrigation regimes and anthropogenic activities can often cause variations in predicted crop yield. Consequently, timely remotely sensed yield estimation soon after planting to just before the harvesting of the crop is necessary for reliable prediction.

As afore-mentioned, ground-based surveys and estimates have commonly been used to determine crop area and production (Craig and Atkinson, 2013). Such estimations generally involve complete censuses, farmer reported sample survey, point samples observed data, area frame systems and administrative data. The Southern African Development Community (SADC) countries make use of an arrangement of subjective (extension officers/growers assessments) and objective procedures which involve direct measurement (Craig and Atkinson, 2013). In most SADC countries, crop estimation has been limited to cereals and other major crops, with very few countries, due to cost, expanding crop estimation to minor agricultural commodities.

Remote sensing however, negates a number of these costs involved with traditional methods of crop yield estimation. Although remote sensing has a long history of use on crop prediction and assessment which dates back to the early 70s, technological advancements in sensor superiority and accessibility and processing advances make it ideal for reliable crop yield estimation. Remotely sensed data are commonly used as auxiliary variables in regression or standardisation estimators and occasionally in confusion matrices.

Recently, statistical methods for yield estimation using remotely sensed data have shown that the enumeration of small samples and use of appropriate analytical techniques can significantly reduce the cost of the data collection and increase yield estimation accuracy (Craig and Atkinson, 2013). In examining the link between crop yield and hyperspectral data for predicting yield, a number of methods are commonly used. Moran et al. (1997) identified two key approaches, single or multiple time-integrated indices like normalized vegetation index (NDVI), which is used to estimate crop yield by empirical regression equations and indices, such as leaf area index (LAI) used as estimators of plant physical characteristics. These approaches are used as input data into crop growth and agro-meteorological models (Serrano et al., 2000).

Recently, statistical methods such as piecewise linear regression, non-linear quasi-newton multivariate optimization (Prasad et al., 2006) and neural network techniques (Panda et al., 2010) have been used with limited success. In the recent past, sparse partial least squares regression (SPLSR) has shown promise for yield estimation. According to Abdel-Rahman et al. (2014), due to the high dimensionality of hyperspectral datasets, the new SPLSR is valuable for mining hyperspectral data during the yield estimation model development processes. A unique benefit of SPLSR is that it selects significant variables (wavebands) for estimating the feature of interest (yield), which makes it an exclusive procedure when analysing high dimensional hyperspectral datasets (Abdel-Rahman et al., 2014; Chun and Keles, 2010). Therefore, a further purpose of this study was to determine common dry bean yield estimate using hyperspectral remotely sensed data based on the entire crop cycle using SPLSR. Specifically, the study aimed at determining the most ideal growth period for reliably estimating the common dry bean yield.

1.2 Aims and objectives

The study sought to pursue two aims, firstly to evaluate the effectiveness of non-imaging hyperspectral remote sensing in discriminating common dry bean varieties.

Objectives:

- To determine the potential of hyperspectral data to discriminate among three different varieties of common dry bean using partial least squares – discriminant analysis (PLS-DA).
- To determine the optimal stage within the growth cycle for discriminating the common dry bean varieties using hyperspectral data.

Secondly, to determine the optimal stage within the common dry bean's growth cycle for yield estimation using ground-based hyperspectral data

Objectives:

- To estimate the common dry bean yield using hyperspectral data and partial least squares regression (SPLSR) at bi-weekly intervals throughout the life cycle

- To determine the stage in the life cycle of common dry beans that is optimal for yield predictions, based on hyperspectral data and SPLSR.

Chapter two

Discrimination of common dry bean varieties using ground-based hyperspectral data

This chapter is based on:

Rajah, P., Odindi. J., Abdel-Rahman, E. M., Mutanga, O., and Modi., A, 2015: Varietal discrimination of common dry bean (*Phaseolus vulgaris* L.) grown under different watering regimes using multi temporal hyperspectral data, Journal of Applied Remote Sensing, 9, 096050-1. [DOI: 10.1117/1.JRS.9.096050]

Abstract

Globally, the common dry bean varieties (*Phaseolus vulgaris* L.) are regarded as valuable food crops. Due to diverse on-farm and post-harvest management requirements, quick, reliable and cost-effective varietal discrimination is critical for optimal management during growth and after harvesting. The large number of valuable wavelengths that characterize hyperspectral remotely sensed datasets in concert with emerging robust discriminant analysis techniques offer great potential for on-farm dry bean varietal discrimination. In this study, an integrated approach of Partial Least Squares Discriminant Analysis (PLS-DA) on hyperspectral data was used to determine the bean's optimal timing for on-farm varietal discrimination. Based on experimental plots under irrigated and rain-fed watering regimes, hyperspectral data was collected at three major phenological stages. Data at each stage were first used to generate PLS-DA models to determine Variable (wavebands) Importance in the Projection (VIP) and the VIP bands used to generate VIP conditioned PLS-DA models. The study identified six weeks (branching and rapid vegetative growth) and ten weeks (flowering and pod development) after seed sowing as optimal stages for varietal discrimination. The study offers insight into the optimal period to discriminate dry bean varieties using spectroscopy, valuable for on-farm and after farm management and crop monitoring sensor development.

Key words: Discrimination, Hyperspectral data, Partial Least Squares Discriminant Analysis, Variable Importance in the Projection.

2.1 Introduction

The common dry bean (*Phaseolus vulgaris* L.) is a popular food leguminous plant globally. In eastern and southern Africa, it has been identified as the second most valuable source of human nutritive protein and the third most vital energy source (Wortmann et al., 1998). In South Africa, DAFF (2010) notes that a mean annual production of approximately 58 000 tons per year was recorded in the previous 10 years. According to Sobhan (2007) the common dry bean is generally grown for subsistence and commercially in all of the country's provinces.

Up to 75% of the common dry bean consumed in South Africa is produced locally. Consequently, due to the crop's socio-economic value, an understanding of the different factors that can be used to expedite effective management decisions is required. Varietal discrimination is particularly valuable in the management of the crop. The value of varietal discrimination includes diseases and pest control, irrigation, fertilizer treatment regimes, allocation of harvesting schedules, optimization of field labour, crop acreage and trade decisions (Apan et al., 2004; Fortes and Dematte, 2006; Galvao et al., 2005; Mingwei et al., 2008;).

Conventional methods for discriminating crop varieties involve taxonomical information, visual percentage cover and other relevant ancillary data (Adam and Mutanga, 2009). However, these techniques commonly require technical expertise and are often costly and time consuming (Govender et al., 2007; Schmidt and Skidmore, 2001). Recent advancements in remote sensing offer an opportunity to reliably and cost effectively discriminate the common dry bean varieties (Kumar et al., 2005; Sobhan, 2007; Van Niel and McVar, 2004; Wilson et al., 2014). The emergence of hyperspectral remotely sensed data, due to increased spectral resolution, particularly holds great promise for field level species and varietal discrimination (Galvao et al., 2005; Govender et al., 2007; Kumar et al., 2005; Van Niel and McVar, 2004; Zhang et al., 2012). However, high data dimensionality and natural multi co-linearity that characterize hyperspectral data often impede effective statistical analysis and classification. According to Koedsin and Vaiphasa (2013) eliminating data redundancy while conserving valuable spectral data is critical in a discrimination process.

Recent literature shows that emerging classification algorithms can be used to reduce spectral dimension and therefore valuable in determining useful wavebands in hyperspectral data-sets (Peerbhay et al., 2013a; Peerbhay et al., 2013b). Abdel-Rahman et al. (2014) note that the partial least squares (PLS) technique has particularly emerged as valuable in reducing redundancy in hyperspectral data. According to Dorigo et al. (2007) and Wold et al. (2001) the

PLS method is a proficient multivariate statistical method addressing complexities linked to high dataset dimensionality by performing dimension reduction and classification concurrently. This reduces over-fitting as well as the exclusion of insensitive wavebands that preserve characteristics of random noise and background effects (Cho et al., 2007; Dorigo et al., 2007; Wold et al., 2001).

A number of studies have demonstrated the value of PLS in increasing the value of hyperspectral datasets. These include predicting swiss chard yield (Abdel-Rahman et al., 2014) forecasting of grain produce and protein content in wheat and barley (Hansen et al., 2002) as well as the estimation of green grass biomass (Cho et al., 2007). However, very few studies have used the technique to discriminate between different crop varieties. Furthermore, most studies that have used partial least squares-discriminant analysis (PLS-DA) to distinguish vegetation species have been based on once-off data collection at a specific stage throughout the growing season (Panneton et al., 2011; Peerbhay et al., 2013a). However, Sobhan (2007) notes that effective discrimination of species could change during the plant's life cycle. Consequently, selection of the optimum phenological stage, based on continuous monitoring, may facilitate reliable spectral separability of species or varieties.

The common dry bean provides a suitable crop with notable growth stages. Nuland and Schwartz (1989) note that the common dry bean can be categorized into four major phenological stages:

1) Germination and stand establishment,
2) Rapid vegetative growth,
3) Flowering and pod development and;
4) Pod fill and maturation.

These phenological stages can be generally grouped as vegetative phases (1 and 2) and reproductive phases (3 and 4). These unique stages indicate the common dry bean's physiological and biophysical changes and offer a valuable guideline for multi-temporal sampling intervals for varietal discrimination. In this regard, this study sought to determine the value of ground-based hyperspectral datasets using PLS-DA to discriminate common dry bean varieties at different growth stages.

2.2. Materials and methods

2.2.1 Study area

Field data was collected at the Ukulinga training and research farm at the University of KwaZulu-Natal, Pietermaritzburg, South Africa (Latitude: $30^{0}24'S$, Longitude: $29^{0}24'E$, Altitude: 800m). The area is characterized by a mean 800 mm annual precipitation and 18ºC temperature (Makanda et al., 2012). Summers are warm with increased rainfall and winters are mild with occasional frost. The soil which is mainly derived from shale, is loamy (41.8% sand, 37.1% silt, 21.1% clay) and fine-textured (Makanda et al., 2012).

2.2.2 Planting material

Three certified and disease free varieties (Caledon, Ukulinga and Gadra) were used in this study. Caledon has white while the Ukulinga and Gadra varieties have red speckled seeds. Debouck and Hidalgo (1986) note that the common dry bean varieties are characterized by either determinate or indeterminate growth. Determinate growing varieties have stems which end in well-developed inflorescence, typified by a large main stem, plant height between 30-50cm and a short flowering period with pods maturing simultaneously (Debouck and Hidalgo, 1986). Indeterminate growing varieties on the other hand are characterized by in-ability to climb, continued vegetative growth, even during flowering and short branches in relation to the main stem (Debouck and Hidalgo, 1986). The Caledon and Ukulinga varieties exhibit indeterminate while the Gadra variety exhibit determinate growth (Kornegay et al., 1992).

2.2.3 Growth stages and multi- temporal data collection

In this study, only three out of the four aforementioned major growth stages were considered for data collection. The fourth growth stage (pod fill and maturation) was not included in the analysis as the canopies of the bean varieties were senescing and the canopy reflectance was dominated by soil background reflectance. Spectral data were collected after two weeks, six weeks and ten weeks after planting, representing the emergence and early vegetative growth, branching and rapid vegetative growth and flowering and pod formation stages, respectively.

2.2.4 Experimental design and treatments

Experimental treatment on the three varieties involved two watering regimes; rain-fed and irrigation laid out in a split-plot trial using randomized complete block design with three replicates. Drip irrigation was initially applied once a week, then twice a week during the intermediate growth phase and once a week close to maturity. Approximately 23 mm of water was applied during each irrigation episode and the rain-fed plots used as control treatment.

The irrigation treatments were assigned to the main plots, while the varieties of common dry bean were assigned to sub-plots. The experimental plots were treated with a pre-emergence herbicide and bean seeds sown on 31 March 2014 on plots (5 X 1 m) at inter-row and intra-row spacing of 0.75 m and 10 cm, respectively. A combined fertilizer (NPK 2:3:2) was applied on the experimental plots with an amount of 600 kg ha^{-1}. A post-emergence herbicide was sprayed every two weeks to control weeds throughout the growing period. Other agricultural practices like pests and disease control were carried out according to optimal practice recommendations.

2.2.5 Hyperspectral data collection

Hyperspectral data was collected five times (bi-weekly), with the first measurements recorded two weeks after planting and the last, a week before harvesting. Reflectance spectra at canopy level were collected under clear sky settings between 10:00-14:00 local time using the non-imaging Field Spec 3 spectroradiometer (Analytical spectral devices, 2008). The spectroradiometer measures spectral data in the 350-2500 nm region of the electromagnetic spectrum and has a sampling interval of 1.4 nm in the 350-1000 nm spectral region and a 2 nm spectral interval in the 1000-2500 nm range.

Reflectance measurements were taken at nadir (25^o) 50 cm above the plant canopy. This yields a field of view of approximately 23 cm in diameter on the ground, large enough to cover desired dry bean canopy. Each experimental plot was sub-divided into six subplots, with three varieties replicated three times in the experiment. Eighteen spectral measurements were taken from each subplot, totalling 162 spectra for the three varieties (54 spectra per variety per watering regime). The eighteen spectral measurements in each subplot were recorded at three different locations in the three middle experimental rows. Six measurements were taken at each of the three positions in the experimental row, the first just after the outer plants, the second in the middle of the experimental row and the third near the end of the experimental row. This was done twice per experimental plot, totalling six data collection points for each watering regimes with three replicates (n = 18). Spectral reflectance measurements were made comparative to a spectralon white reference panel. A reflectance measurement was taken from the reference panel before and after every five minutes of measurement to adjust for changes in atmospheric condition and/or sun irradiance. For each reflectance measurement collected, twenty scans were processed internally by the spectroradiometer and raw spectral measurements averaged to obtain a more reliable spectral reflectance measurement. The averaged spectra were then interpolated every 10 nm in a user specified spectral range using the ViewSpec Pro spectral software interpolation technique (Analytical spectral devices, 2008).

2.2.6 Statistical analysis

Prior to statistical analysis, spectral data in 350–399 nm, 1355–1420 nm, 1810–1940 nm and 2470–2500 nm ranges of the electromagnetic spectrum were excluded from the analysis due to high noise associated with these wavelength ranges (Abdel-Rahman et al., 2014; Curran, 1994; Tian et al., 2011).

2.2.7 Partial least squares discriminant analysis

Partial least squares discriminant analysis (PLS-DA) was utilised in order to classify the common dry bean varieties. The technique is based on the classical PLS regression method to construct predictive models (Wold et al., 2001). PLS regression decreases data dimensionality in a process of variable response's relation to the predictor variables. However, in PLS-DA, the response variable (the common dry bean variety) expresses class membership. PLS uses the created spectral matrices to produce eigenvectors serving to explain variance within spectral data and relationship with the response variables (Mutanga et al., 2004; Wolter et al., 2008).

Wold (2001) notes that owing to the increase in correlation of variables in PLS models, it is imperative to identify the precise amount of components, to decrease the risk of over-fitting. Therefore, by selecting an optimal number of components yielding the highest accuracy, model performance is tested for each component. Cross validation has commonly been used in the past and is recognized as an accurate technique to test the models significance for every PLS component (Dorigo et al., 2007; Wold, 2001). In this study, the parameters of the PLS-DA models were optimized by using a tenfold cross-validation which was based on the training data set of the interpolated spectrum range (400-2500 nm). The process of optimization was methodical and involved adding each component progressively to the PLS-DA model. Each optimization process produced a cross-validated (CV) error and the process was repeated until the addition of components to the PLS-DA model failed to provide a decline in CV error. The PLS-DA model with the lowest CV error rate according to the number of components based on the training data was then used in the classification of the test data set.

Perez-Enciso and Tenenhaus (2003) note that for a PLS model to generate significant classification accuracies, prior-selection of variables based on the variable weight in the projection (VIP) score is required. The VIP computes the weight of each wavelength range by generating scores functioning as a measure of importance for each the wavelength range defined (Chong and Jun, 2005). The VIP scores range from 0 upwards. Wavelength ranges having a VIP value greater than one are selected, since this indicates a strong classification potential. The VIP method calculates a ranked score for each wavelength range based on the equation:

$$\mathrm{VIP}_k = \sqrt{K \sum_{a=1}^{A} \left[(q_a^2 \ t_a^T \ t_a)(w_{ak} / \ w_k^2) / \sum_{a=1}^{A} \ (q_a^2 \ t_a^T \ t_a) \right]} \tag{1}$$

Where VIP_k is the calculated importance of the *k*th wavelength range based on the generated PLS-DA model using the number of components (a) which yielded the lowest CV error. W_{ak} is the equivalent loading weight of the *k*th waveband in the *a*th PLS-DA component. *K* is the total number of wavebands and t_a ,w_a, and q_a , are the *a*th column vectors (Alfanador, 2014; Peerbhay et al., 2013a;).

In this study, the important wavelength ranges with a value greater than one after the initial PLS-DA models were used to generate VIP based PLS-DA models. The same process was followed in the cross validation at ten-fold to produce VIP based PLS-DA models.

2.2.8 Accuracy assessment

The final 54 samples were divided into n=39 (70%) and n=15 (30%) as training and test data sets respectively. Using Tanagra data mining software (Tanagra project, 2008) the number of samples allocated to the training and test data sets were approximately equal. Lindstrom et al. (2011) note the importance of equity and balance to test and train hyperspectral data for optimization and classification of the PLS-DA process. To determine classification accuracy, matrices were generated and users, producers and overall accuracies computed. Furthermore, two components of disagreement (allocation and quantity) were calculated. These components provide disagreement based on the information contained within the matrix (Pontius and Milones, 2011). According to Pontius and Millones (2011) allocation and quantity disagreements are the most reliable measures to classify accuracy.

2.3 Results

The spectral characteristics for the three varieties at the three growth stages are shown in Figure 1.1, from the figure it is evident that the reflectance for each of the three varieties at the first growth stage differs from the reflectance in the second and third growth stages. All varieties are most visually separable at the later growth stages throughout the entire spectral range, except for Caledon and Ukulinga (Figure1.1c and f). At the second growth stage, all the varieties could be visually separated in the visible, near infrared and a small portion of the shortwave infrared portion of the electromagnetic spectrum for the irrigated treatment (Figure1.1b). The rain-fed treatment showed visual spectral differences between the varieties for the entire wavelength range of the optical spectrum, except for Caledon and Ukulinga (Figure1.1b and e). There were no clear spectral variations (visually inspected) between the varieties during the first growth stage for both treatments (Figure 1.1a and d).

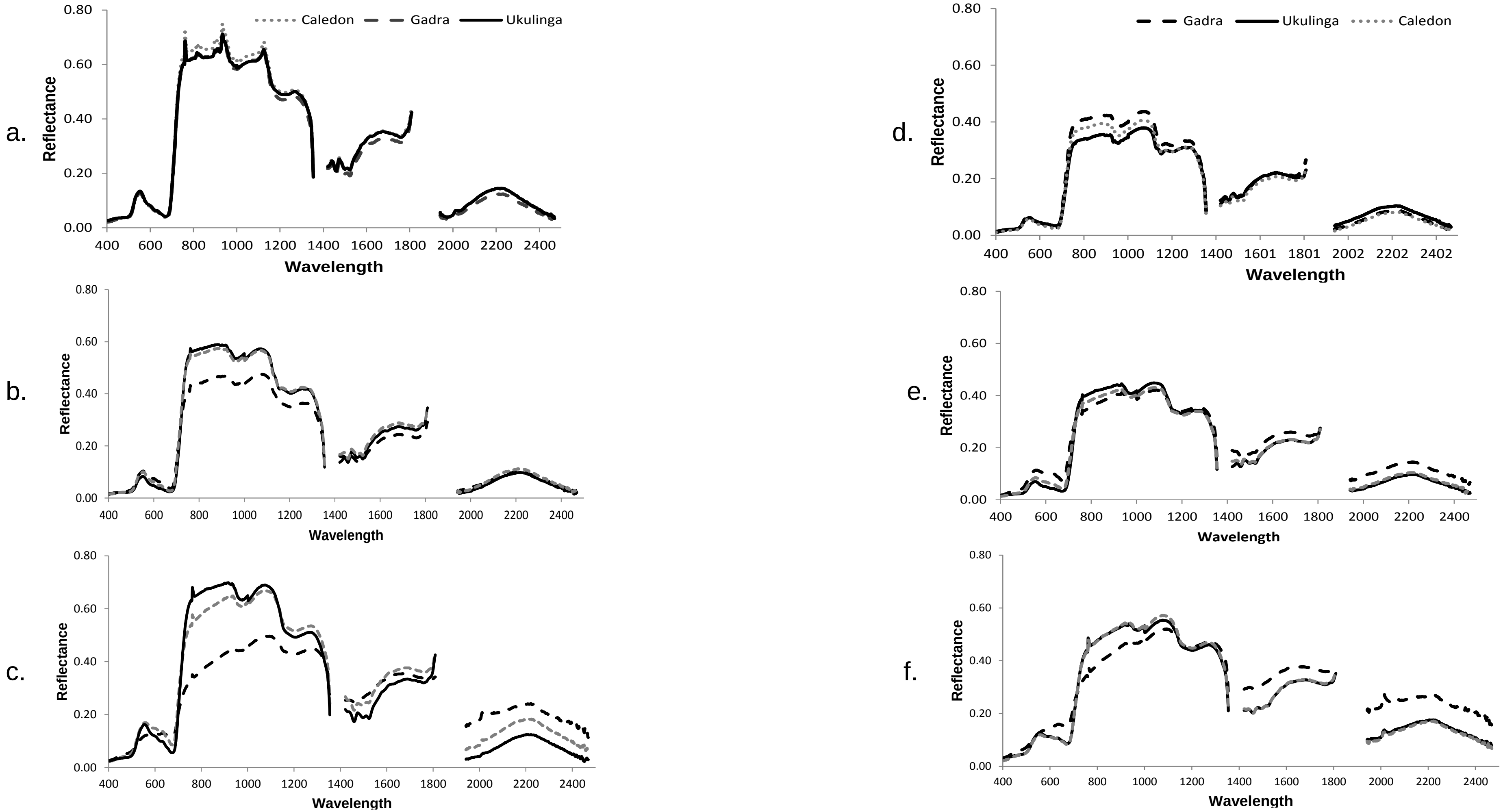

Figure 1.1: Average spectra of common dry bean canopies measured under experimental treatment conditions: (a) Irrigated treatment at the germination and stand establishment growth stage; (b) Irrigated treatment at the branching and rapid vegetative growth stage; (c) Irrigated treatment at the flowering and pod development growth stage; (d) Rain-fed treatment at the at the germination and stand establishment growth stage; (e) Rain-fed treatment at the branching and rapid vegetative growth stage; (f) Rain-fed treatment at the flowering and pod development growth stage. The wavelengths were averaged at 10 nm. Spectral wavebands between 350 and 399 nm, 1355 and 1420 nm, 1810 and 1940 nm and 2470 and 2500 nm were removed due to noise.

2.3.1 PLS-DA using the full spectral range (400–2500 nm)

2.3.1.1 Emergence and early vegetative growth stage

Continuous adding of components to the PLS-DA model for the irrigated plot decreased the Cross Validation (CV) error rate until a threshold was reached (Figure1.2). There was a moderate decrease in CV error in the 1st component (error of 70.4%) to the 15th component (error of 61.8%). The lowest CV error was obtained when 4 components were used (error of 44.87%). Subsequently, the model fluctuated until reaching 10 components, where it stabilized through to the 15 components.

A similar trend was observed with the rain-fed plots (Figure1.2), where the addition of components to the PLS-DA model showed an overall reduction in CV error. The lowest CV error was reached using 2 components (33.8%). The PLS-DA model for the irrigated plot exhibited an overall classification accuracy of 67% and an allocation and quantity disagreement of 13 and 20% respectively, with individual common dry bean variety user's and producer's accuracies ranging from 40 to 100% (Table 1.1). The PLS-DA model showed an overall classification accuracy of 60% and an allocation and quantity disagreement of 20% each for the rain-fed plot. Individual variety user's and producer's accuracies ranged from 20 to 100% (Table 1.1).

2.3.1.2 Branching and rapid vegetative growth stage

The addition of components followed the same trend as the emergence and early vegetative growth stage. The irrigated plot produced the lowest CV error when 3 components were used (12.83%) (Figure1.2), which stabilised from 9 to 15 components. Using PLS-DA model, the irrigated plot generated an overall accuracy of 80% with an allocation and quantity disagreement of 13 and 7% respectively, with user's and producer's accuracies between 60% and 100% (Table 1.1). The rain-fed plot showed the lowest CV error rate when 4 components were used (8.53%). The CV error stabilised from use of 7 to 15 components. The results of the rain-fed plot PLS-DA model produced an overall classification accuracy of 100% and an allocation and quantity disagreement of 0%, with individual variety user's and producer's accuracies at 100%.

2.3.1.3 Flowering and pod formation

The addition of components yielded the same trend as the emergence and early vegetative and the branching and rapid vegetative growth stages, with the irrigated plot leading to the lowest CV error when 4 components were used (4.13%) (Figure 1.2). The CV error stabilised from 9 components onwards, until 15 components. The PLS-DA model of the irrigated plot produced an overall accuracy of 100% with an allocation and quantity disagreement of 0% and individual variety user's and producer's accuracies of 100% (Table 1.1). The rain-fed treatment yielded the lowest CV error when 5 components were used (11.43%). The CV error rate stabilized from 12 to 15 components. The rain-fed PLS-DA model showed an overall accuracy of 100% with an allocation and quantity disagreement of 0% and individual variety user's and producer's accuracies at 100% (Table 1.1).

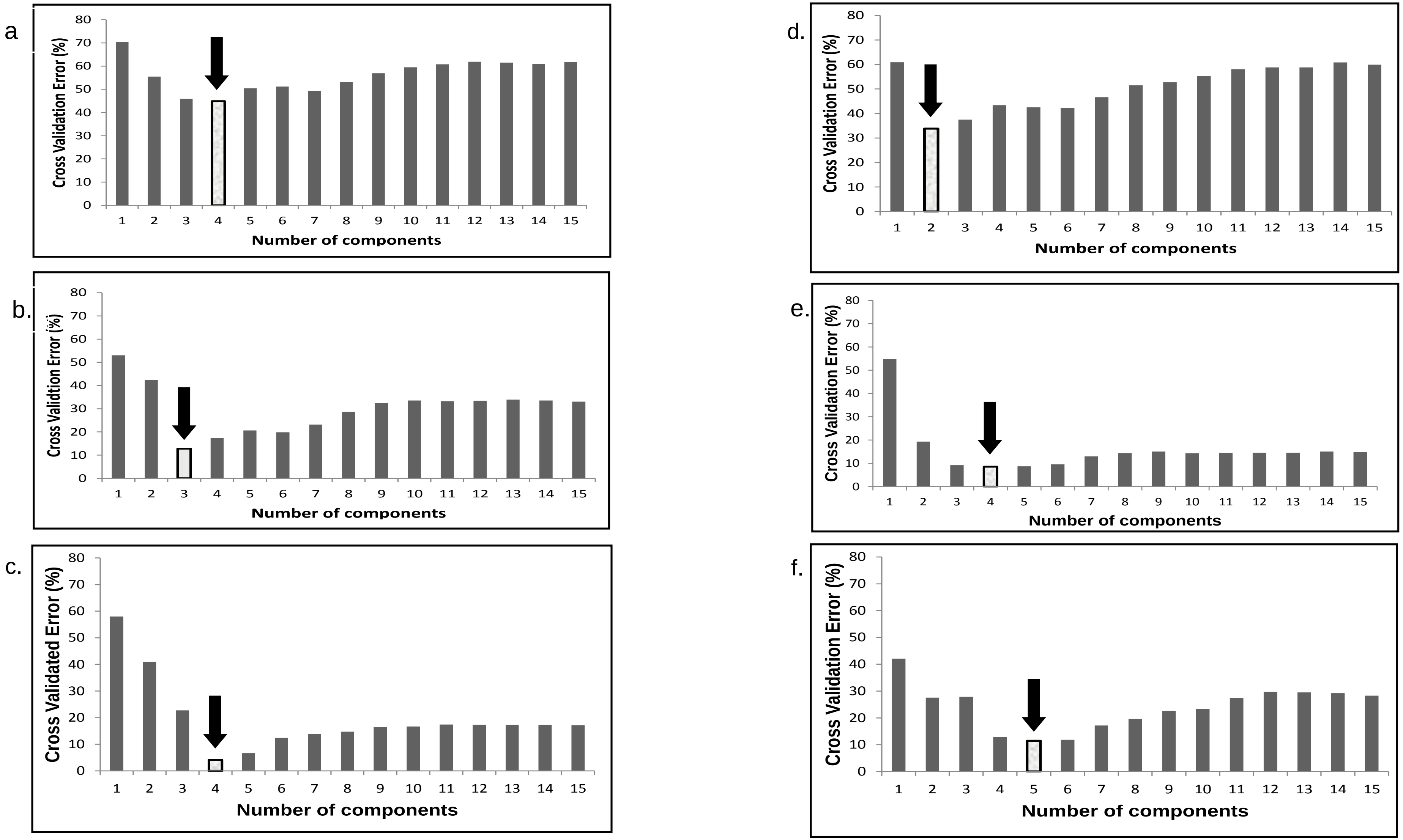

Figure 1.2: Analysis of each PLS-DA components power to discriminate varieties using the complete spectral range (400-2500 nm) for the three growth stages for both treatments. The lowest possible error rate based on the training data was computed via ten-fold cross validation. The optimal number of components which yielded the lowest error is highlighted with an arrow in the bar graph. Emergence and early vegetative growth (a=Irrigated, d= rain-fed); Branching and rapid vegetative growth (b=irrigated, e=rain-fed) and Flowering and pod (c=irrigated, f=rain-fed).

Table 1.1: Summary of discrimination results of the three common dry bean varieties based on the PLS-DA algorithm using the full spectral range (400-2500 nm) at the three growth stages for the two treatments. Emergence and early vegetative growth stage= A; Branching and rapid vegetative growth stage = B; and Flowering and pod formation stage = C. 1 and 2 indicate the irrigated and the rain-fed treatments, respectively.

	User's accuracy (%)	Producer's accuracy (%)	Overall accuracy (%)	Allocation disagreement (%)	Quantity disagreement (%)
A1					
Gadra	62.5	100			
Caledon	75	60	67	13	20
Ukulinga	66.6	40			
A2					
Gadra	50	60			
Caledon	71.42	100	60	20	20
Ukulinga	50	20			
B1					
Gadra	83.33	100			
Caledon	80	80	80	13	7
Ukulinga	75	60			
B2					
Gadra	100	100			
Caledon	100	100	100	0	0
Ukulinga	100	100			
C1					
Gadra	100	100			
Caledon	100	100	100	0	0
Ukulinga	100	100			
C2					
Gadra	100	100			
Caledon	100	100	100	0	0
Ukulinga	100	100			

2.3.2 Variable importance in the projection (VIP)

2.3.2.1 Emergence and early vegetative growth stage

Figure 1.3 explains and identifies the position of the optimal set of wavebands as determined by the VIP method for both the irrigated and the rain-fed treatments. The VIP method for the irrigated treatment placed importance on wavebands located within the visible (400-723 nm), near infrared (723-1400 nm) and shortwave infrared (1400-2500 nm) portions of the electromagnetic spectrum (Table 1.2). There was prevalence of wavebands located in the blue region as well as the green region of visible light (Figure1.3a). In the rain-fed treatment, there was no VIP in the visible portion of the electromagnetic spectrum. However, there was importance placed at the 870-1110 nm in the near infrared and 1990-2490 nm in the shortwave infrared region of the spectrum (Table 1.2).

2.3.2.2 Branching and rapid vegetative growth

The VIP method for the irrigated treatment did give weight on the wavelength ranges located in the visible portion of the spectrum. However, there was weight given to the 1660-2500 nm range within the shortwave infrared region (Table 1.2). The rain-fed treatment gave weight to the wavelength ranges situated in the visible portion of the spectrum with 19 wavebands identified in the visible portion of the spectrum (Figure1.3e). Furthermore, weight was high for spectral ranges located in the near infrared (701-1400 nm) and shortwave infrared (1400-2500 nm) section of the spectrum (Figure1.3b and e).

2.3.2.3 Flowering and pod formation

Figure1.3 shows that the VIP method for the irrigated treatment placed waveband importance on 23 bands in the visible part of the spectrum. Wavelength range weight was also high at 880-1150nm and 2140-2480 nm in the near infrared and shortwave infrared segments of the spectrum, respectively (Table 1.2). A similar pattern was visible in the rain-fed treatment. Fifteen wavebands were highlighted at the 400-670 nm range in the visible portion and the 2070-2500 nm range in the shortwave infrared region (Figure1.3e and f).

Table 1.2: Summary of weight importance in the projection (VIP) throughout the electromagnetic spectrum (400-2500nm) for irrigated and rain-fed treatments, for the three common dry been growth stages.

Treatment	**Irrigated**			**Rain-fed**		
Growth stage	**Visible (400-700 nm)**	**Near infrared (701-1400 nm)**	**Shortwave infrared (1400-2500 nm)**	**Visible (400-700 nm)**	**Near infrared (701-1400 nm)**	**Shortwave infrared (1400-2500 nm)**
Emergence and early vegetative	VIP (16)	VIP (32)	VIP (34)	None	VIP (25)	VIP (51)
Branching and rapid vegetative	None	None	VIP (85)	VIP (19)	VIP (40)	VIP (44)
Flowering and pod formation	VIP (23)	VIP (28)	VIP (35)	VIP (24)	VIP (15)	VIP (44)

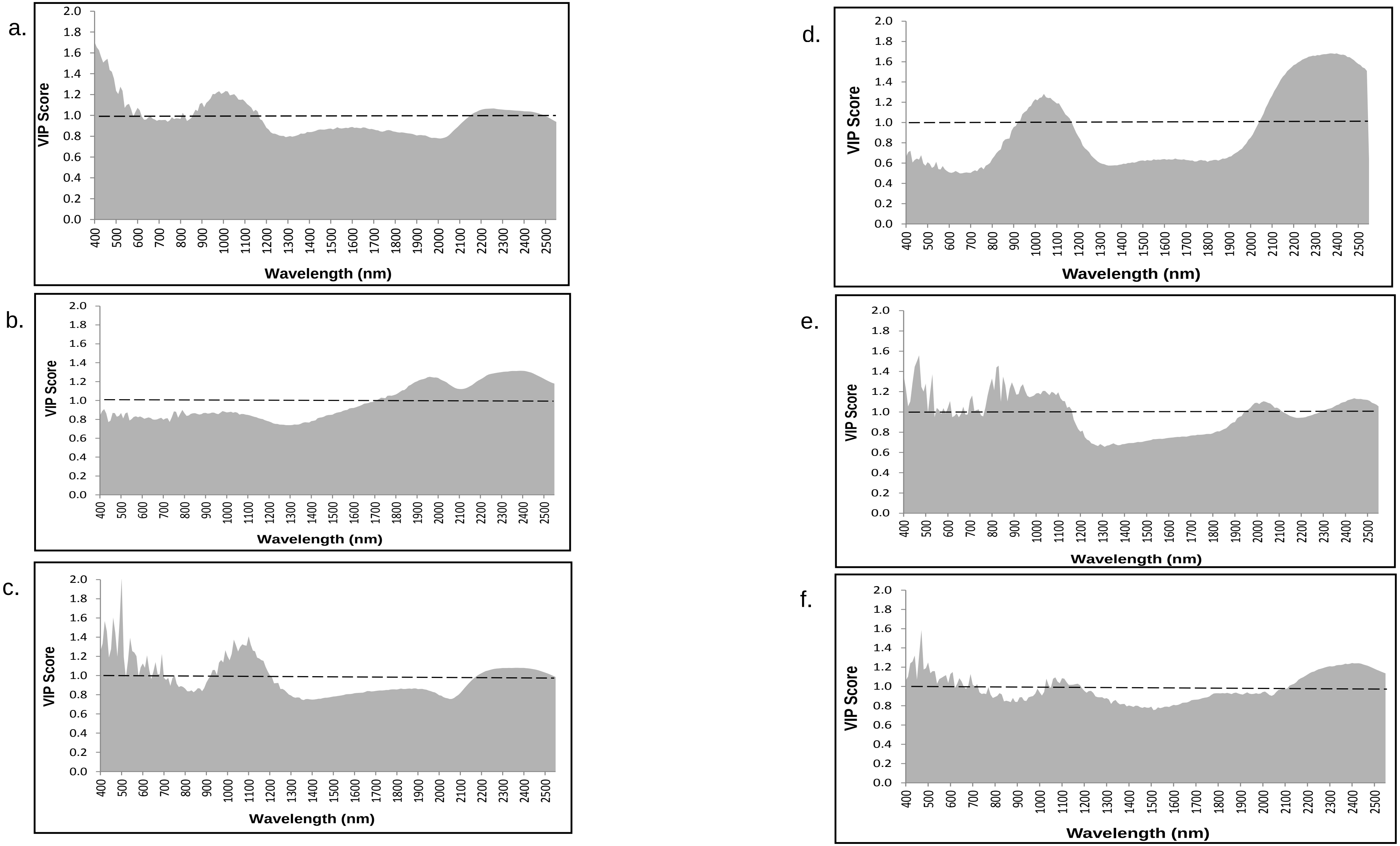

Figure 1.3: Waveband weight selected by the variable weight in the projection (VIP) technique of PLS-DA for three growth stages of three common dry bean. Spectrum bands (400-2500 nm) are displayed as VIP scores. Significant wavebands are those which achieved VIP scores larger than one. Emergence and early vegetative growth stage (a=Irrigated, d= rain-fed); Branching and rapid vegetative growth stage (b=irrigated, e=rain-fed) and Flowering and pod formation stage (c= irrigated, f=rain-fed).

2.3.3 PLS-DA using the optimal subset of VIP wavelength ranges

The PLS-DA models which were generated using the VIP wavebands of the irrigated and rain-fed treatments for the three growth stages showed a consistent overall accuracy (Table 1.3). However, there was a slight difference in overall accuracy of the last growth stage using all wavelength ranges (overall accuracy of 100% for both irrigated and rain-fed treatments) (Table 1.1), when compared to the overall accuracy of 93% obtained when only using the VIP wavelength ranges for both irrigated and rain-fed treatments) (Table 1.3). Furthermore, user's and producer's accuracies decreased for singular common dry bean varieties for the second and third growth stages (Table 1.3), when using VIP-based PLS-DA models. There was little difference in overall accuracy achieved when comparing PLS-DA models using all wavebands and PLS-DA models using only the VIP wavelength ranges. Generally, the user's and producer's accuracies increased as the common dry bean matured. Although Gadra showed a 100% producer's accuracy at all growth stages, Caledon rose from 60%, 80% and 100% in the first, second and third growth stages, respectively. A similar trend was seen for the Ukulinga variety, which increased from 40, 60 to 100% (Table 1.1).

2.4 Discussion

This study explored the prospect of field-based hyperspectral remote sensing to discriminate three varieties of the common dry bean (*Phaseolus vulgaris* L.) under rain-fed and irrigation regimes. The study demonstrated the prospect of hyperspectral data and PLS-DA in discriminating the common dry bean varieties. The use of PLS-DA provided a unique framework to integrate hyperspectral data and to determine the valuable wavebands through dimension reduction for discriminating the common dry bean varieties. This study confirms results by Peerbhay et al. (2013a) who found that the PLS-DA can handle a large number of wavelength ranges, has the ability to successfully discriminate between plantation forest species that are spectrally similar and has the ability to discriminate between data showing a high correlation between variables.

Table 1.3: Summary of discrimination results of the three common dry bean varieties based on the PLS-DA algorithm using the VIP wavebands at the three growth stages for the two treatment plots. Emergence and early vegetative growth stage = A; Branching and rapid vegetative growth stage = B; and Flowering and pod formation stage = C. 1 indicates the irrigated treatment and 2 indicates the rain-fed treatment.

	User's accuracy (%)	Producer's accuracy (%)	Overall accuracy (%)	Allocation disagreement (%)	Quantity disagreement (%)
A1					
Gadra	75	60			
Caledon	66.67	40	60	20	20
Ukulinga	50	80			
A2					
Gadra	60	60			
Caledon	55.56	100	60	13	27
Ukulinga	100	20			
B1					
Gadra	80	80			
Caledon	100	80	80	13	7
Ukulinga	66.67	80			
B2					
Gadra	100	80			
Caledon	100	100	93	0	7
Ukulinga	83.33	100			
C1					
Gadra	100	100			
Caledon	83.33	100	93	0	7
Ukulinga	100	80			
C2					
Gadra	100	100			
Caledon	83.33	100	93	0	7
Ukulinga	100	80			

2.4.1 Discrimination using all spectral bands

In this study, PLS-DA was moderately successful (67% for irrigated and 60% for rain-fed) in discriminating common dry bean varieties during the emergence and early vegetative growth stage. At the branching and rapid vegetative growth stage, discrimination accuracies increased to 80% for irrigated and 100% for the rain-fed water supply. Compared to the emergence and early vegetative growth stage, PLS-DA was more efficient to classify certain varieties. The lower discrimination accuracies for the emergence and early vegetative growth stage, compared to the latter two growth stages can be attributed to low vegetative cover after emergence and hence a dominant background soil reflectance. The increase in discrimination accuracy at the branching and rapid vegetative growth stage can be attributed to a range of factors including plant leaf angular orientation, duration of growth, leaf volume and leaf pigment concentration (Singh et al., 2013). Generally, the PLS-DA model achieved higher classification accuracies at all irrigated treatment growth stages.

The flowering and pod formation growth stage for the irrigated treatment reached the highest classification accuracies (100%). According to Zwigelaar (1998) spectral characteristics of plant canopies are determined by their physical properties and chemical composition. In consistency with Misra et al. (2012) a plant species or variety's growth stage and vigour is valuable in varietal discrimination. Lucas et al. (2008) identified light absorption and scattering at different wavelengths as major factors influencing the variation in spectral reflectance within a plant species. These differences can be determined with hyperspectral remotely sensed data, and ultimately resulted in the favourable accuracies obtained with by PLS-DA models.

2.4.2 Discrimination using wavelength ranges selected by the VIP method

Due to inability of PLS-DA to determine the most optimal bands to be used in the final model for varietal discrimination, the VIP method was used. In contradiction to Peerbhay et al. (2013a) and Wolter et al. (2008) adoption of the VIP method to remove irrelevant bands did not increase overall accuracy. Overall accuracy remained constant for all treatments and growth stages besides the irrigated treatment of the first growth stage, showing a slight decrease (7%) when the VIP wavelength ranges were used. Although Zwiggelaar et al. (1998) suggests that by selecting specific wavelength ranges for discrimination, optimum discrimination accuracy could be obtained. They note that these could be dependent on the crop varieties under consideration. This can therefore be used to explain why the VIP wavelength ranges did not improve overall accuracy when used to generate further PLS-DA models with only VIP wavebands.

According to Misra (2012) knowledge on crop phenology is valuable in varietal discrimination. Since the growth stages of the common dry bean varieties could be distinguished, reliable discrimination could be achieved by using both VIP and all wavebands at the later stages of growth.

2.5 Conclusions

Using the full spectrum range, the common dry bean varieties were discriminated with a relatively higher accuracy at the earlier growth stage under irrigated conditions as opposed to rain-fed conditions. PLS-DA models were successful in discriminating the common dry bean varieties from the first growth stage. Generally, accuracies increased with successive temporal sampling intervals. It can therefore be concluded that hyperspectral field-based remotely sensed data, using PLS-DA can be used to discriminate the common dry bean varieties. To date, most species and varietal discrimination studies have failed to account for multi-temporal variability. This study therefore investigated the implication of multi-temporal phenology on spectral separability. It is concluded that the intermediate growth stage provides above average results while the late growth stage provides the most reliable discriminatory results. This study shows that field-based hyperspectral remote sensing can be used to discriminate common dry bean varieties at various stages of growth. Whereas this study was conducted on small experimental plots, results in this study offer the basis for larger commercial agricultural applications using aerial or satellite platform sensors developed from the highlighted spectral ranges.

Chapter three

Prediction of common dry bean yield using ground-based hyperspectral data

This chapter is based on:

Rajah, P., Odindi., J., Abdel-Rahman, E. M., and Mutanga, O., (in Review). Determining optimal phenological stage for predicting the common dry bean (*Phaseolus vulgaris* L.) yield using field spectroscopy, Journal of Integrative Agriculture.

Abstract

Agricultural crop yield prediction is critical worldwide. Precise and early prediction of common dry bean yield provides valuable information on among others generation of relevant food policies, on-farm and after farm planning and pricing and marketing. Therefore, this study sought to investigate the use of multi-temporal ground-based hyperspectral data, acquired at major phenological stages, to predict three common dry bean varieties yield grown under irrigated and rain-fed watering regimes. Canopy-level hyperspectral data were collected from the Caledon, Ukulinga and Gadra varieties at three distinct growth stages and Sparse partial least squares regression (SPLSR) was used for data analysis. Results indicated that variations in yield could be described at specific growth stages within the two watering regimes. Generally, with the exception of the Gadra variety under irrigation, models established using data collected from the flowering and pod development stage were more accurate, compared to models derived from preceding growth stages. This study highlights the prospect of phenological ground-based hyperspectral data in predicting common dry bean yield under different watering regimes. The study provides a valuable basis for large scale yield prediction of common dry bean using upgraded airborne or space-borne hyperspectral sensors.

Key words: Common dry bean, yield, watering regimes, hyperspectral data, sparse partial least squares regression.

3.1 Introduction

Globally, the common dry bean (*Phaseolus vulgaris* L.) is regarded as an important food crop, rich in protein and other dietary benefits (Beninger and Hosfield, 2003; DAFF, 2010). In Africa, numerous varieties of common dry bean are an essential food and an imperative constituent of household and national incomes (Siddiq et al., 2010). In South Africa, approximately 58 000 tons per year, on 56 000 hectares, has been recorded under irrigated and rain-fed conditions in the past 10 years (DAFF, 2013). Therefore, due to the common dry bean's socio-economic value, reliable yield prediction is important for pre-harvest and post-harvest planning. According to Monteiro et al. (2012), such prediction is valuable for among others, on-farm crop treatment, determination of potential yield, commercial profitability, forecast for shortage or surplus, food policy formulation and marketing.

Recently, remotely sensed datasets have been widely adopted for crop yield estimation and forecasting. Specifically, multispectral data have been adopted in among others, early estimation of soybean yield using canopy based reflectance measurements (Ma et al., 2001), to investigate the relationship between rice crop spectral characteristics and its yield (Nuarsa et al., 2011) and sugarcane yield forecasting (Abdel-Rahman et al., 2012; Almeida et al., 2006). However, Goa et al. (2000) note that the fewer broad and discrete spectral bands that characterize multispectral data commonly lead to saturation, particularly on crops with high biomass or leaf area index (LAI). Since most common dry bean varieties have relatively high LAI (Monteiro et al., 2012), adoption of multispectral data for yield estimation and forecasting may be inappropriate (Abdel-Rahman et al., 2014; Monteiro et al., 2012).

Numerous studies eg, Abdel-Rahman et al. (2012), Almeida et al. (2006) and Ma et al. (2001), have suggested the adoption of hyperspectral remotely sensed data for crop yield estimation. However, these studies have ignored the potential of multi-temporal phenological crop cycle reflectance differences for reliable yield estimation. Whereas the common dry bean’s distinct phenological cycle offers an opportunity to determine the optimum stage for predicting yield using hyperspectral data, the large number of narrow and contiguous bands and therefore high data dimensionality and redundancy commonly impedes selection of the most valuable bands for modelling and therefore yield prediction (Abdel-Rahman et al., 2014; Lillesand et al., 2004). Previous studies such as Fu et al. (2014) and Nguyen et al. (2006) have sought to overcome these limitations by using partial least squares regression (PLS) analysis. According to Huang et al. (2004), PLS regression has the ability to overcome collinearity and over-fitting problems while optimizing available information (Mulla, 2013). However, PLS does not identify the most valuable wavebands for modelling. Consequently, the sparse partial least squares regression (SPLSR) has been identified as an innovative technique that can be used for processing hyperspectral data and dealing with the aforementioned PLS limitations (Chung and Keles, 2010). According to Abdel-Rahman et al. (2014) the superiority of SPLSR lies in its imposition of sparsity solution during data transformation as well as selection of applicable variables for predicting the feature of interest. Using hyperspectral canopy reflectance measurements, this study sought to determine the optimum phenological stage to predict yield for the three common dry bean varieties grown under irrigation and rain-fed conditions using the SPLSR.

3.2. Materials and methods

3.2.1 Study area

Experimental plots were established at the Ukulinga training and research farm, University of KwaZulu-Natal, Pietermaritzburg, South Africa (Latitude: 30^{o}24S, Longitude: 29^{o}24'E, Altitude: 800m). Mean annual temperature and precipitation at the site are 18^{o}C and 800mm, respectively (Makanda et al., 2012). Summers are moderately warm with increased rainfall while winters are temperate with infrequent frost. Soils in the area are of a fine texture and are derived from shales.

3.2.2 Planting material and experimental design

Three certified common dry bean cultivars (Caledon, Ukulinga and Gadra) were used in this study. The Caledon is a white bean while the Ukulinga and Gadra are red speckled beans. The common dry bean exhibit determinate and indeterminate bush growth. Determinate dry bean bush grows to a height of between 30 and 50 cm with stems ending in well-developed inflorescence. It is also characterized by a larger main stem and a short flowering period with pods maturing simultaneously (Debouck and Hidalgo, 1986). Indeterminate dry bean bush growth exhibits in-ability to climb, continued vegetative growth throughout flowering and shorter branches in comparison to the main stem (Debouck and Hidalgo, 1986). The Gadra variety exhibits determinate growth while Ukulinga and Caledon exhibit indeterminate growth (Kornegay et al., 1992).

Treatments were laid out in split-plot experiment using randomized complete block design with three replicates. The irrigation treatments were assigned to the main blocks, while the common dry bean cultivars were assigned to sub-plots. Seeds were sown at an inter-row spacing of 0.75 m and intra-row spacing of 10 cm. The experimental plots were treated with a pre-emergence herbicide and a combined fertilizer (NPK 2:3:2) applied at a proportion of 600 kg ha^{-1}. Post emergence herbicide was sprayed every two weeks to control weeds throughout the growing period. Other agricultural practices like disease and pest control were carried out in accordance to optimal guidelines.

To determine the influence of different watering regimes on yield prediction, half of the experimental plots were irrigated using the drip system. This was applied once a week after planting, twice a week during intermediate growth and once a week close to maturity. Approximately 23 mm of water was applied for each irrigation session. The remaining rain-fed plots were used as the control treatment. Spectral data was collected two, six, ten and thirteen weeks after planting. This represents: (1) the emergence and early vegetative growth, (2) branching and rapid vegetative growth, (3) flowering and pod formation, and (4) pod filling and maturation stages, respectively. However, only data from the first three stages of growth were considered for analysis as the plant canopy was senescing and reflectance measurements dominated by the soil background at the fourth development stage.

3.2.3 Hyperspectral and common dry bean yield data collection

Hyperspectral data were collected to coincide with the three aforementioned phenological stages. A non-imaging Field Spec 3 spectroradiometer (Analytical Spectral Devices: ASD, 2005) was used to collect canopy reflectance spectra under clear sky between 10:00 and 14:00 local time. The spectroradiometer measured spectral reflectance within the 350-2500nm range and has a sampling interval of 1.4nm for the 350-1000nm region and 2nm spectral interval for the 1000-2500nm range. The data was then interpolated to a 1nm spectra interval (Analytical Spectral Devices: ASD, 2005).

Spectral reflectance measurements were collected at nadir-looking 25^{o} angle at 50cm beyond the bean canopy. The ground field of view was approximately 23cm in diameter, adequate enough to cover a cluster of bean plants. Each experimental unit (plot) was sub-divided into six quadrats, with three varieties replicated three times. Eighteen spectral reflectance measurements were collected from each quadrat, totalling 162 spectra for the three varieties (54 spectra per variety per watering regime). The eighteen spectral measurements in each quadrat were recorded at three different positions in the three middle experimental rows. Six measurements were taken at each of the three positions in the experimental row; the first just after the outer bean plants, the second at the middle of the experimental row and the third near the end of the experimental row. This was done twice per experimental unit, totalling six data collection points for each watering regimes with three replicates (n = 18). A reflectance reading was taken from a spectralon white reference panel before and after every five minutes of measurement to adjust for any variations in atmospheric condition and sun irradiance. For each reflectance measurement, twenty scans were processed internally by the spectroradiometer.

Raw spectral measurements were averaged (n = 18 in each quadrat) to obtain a more reliable spectral reflectance measure. All common dry bean plants within a quadrat, which contained a single bean variety were harvested and weighed to obtain fresh yield output per quadrat.

3.2.4 Data analysis

The reflectance spectra were interpolated every 10 nm in a user specified spectral range using ViewSpec Pro spectral software interpolation technique (Analytical Spectral Devices: ASD, 2008). Spectral data at 350–399, 1355–1420 nm, 1810–1940 nm and 2470–2500 nm regions of the electro-magnetic spectrum (EMS) associated with high noise level (Abdel-Rahman, 2014; Curran, 1994; Zhao et al., 2007) were excluded from the analysis, therefore, only 211 wavebands were retained for analysis. Common dry bean yield predictive models were derived for each of the first three common dry bean phenological stages (i.e stages 1, 2, and 3) using SPLSR method.

3.2.4.1 Sparse partial least squares regression (SPLSR)

As aforementioned, the vast number of contagious bands and therefore data dimensionality and redundancy are often a major limitation in hyperspectral data analysis. Consequently, Chun and Keles (2010) developed the sparse partial least squares regression (SPLSR), a multivariate technique based on partial least squares (PLS). SPLRS differs from other commonly used regression techniques as it executes a sparsity solution throughout the data transformation stage and chooses the appropriate variables for estimating the feature of interest (Abdel-Rahman et al., 2014). SPLSR also decreases the noise contained in unrelated variables, hence can be used as a variable selection technique. SPLSR repeatedly constructs an orthogonal set of latent components from the predictor variables which have best covariance with the response variable. This low-dimensional depiction of the data is then used to fit a linear regression model (Kramer and Sugiyama, 2011). Consequently, SPLSR promotes variable selection within the course of dimension reduction (Chun and Keles, 2010). The ability of SPLSR to simultaneously promote variable selection within the course of dimension reduction makes SPLSR an innovative method for investigating high dimensional hyperspectral data. Hence, integrating hyperspectral remotely sensed data with SPLSR offers a unique and effective technique for the common dry bean yield prediction.

According to Chun and Keles (2010) and Chung and Keles (2010), SPLSR only requires two parameters that need to be optimised, the "*eta*", which signifies the sparsity threshold parameter and "*k*" which signifies the amount of concealed components. The sparsity threshold parameter of "*eta*" should characteristically range between 0 and 1, while "*k*" is an integer which depends chiefly on the amount of explanatory variables as well as samples size. A cross validation method is used in order to optimise these two parameters. As recommended by Chung and Keles (2010), by using the leave one-out cross-validation technique, we searched for optimum "*eta*" between 0.1 and 0.9 and for "*k*" between 1 and 10 to reduce data dimensionality and redundancy and to determine fewer and most valuable spectral wavelengths.

Additionally, the SPLSR algorithm computes bootstrapped confidence intervals of the original coefficients for the chosen variables based on the bootstrapped samples. The algorithm then makes use of the confidence intervals to rectify the coefficients and extract the applicable variables based on the correct coefficients. Variables which have greater correct coefficients values are more beneficial for estimating the response variable. In this study, the SPLS library in the R statistical packages version 3.1.2 (R development Core team, 2015) was used to implement the SPLSR algorithm.

3.2.4.2 Validation

Model performance was assessed by using the leave one-out cross-validation technique (Richter et al., 2012). This was done by separating the data into *n* samples, where *n* was the total amount of sample instants used in this study (18). The *n* samples were then disregarded one by one. The models from different growth stages were then trained *n* times using *n*-1 samples, and then tested on the disregarded one. A one-to-one relationship between the measured common dry bean yield and predicted yield were then scrutinized. Root mean square error (RMSE) and the coefficient of determination (R^2_{loocv}) were calculated in order to evaluate the certainty of the SPLSR models. RMSE was calculated using the following formula:

$$\text{RMSE: } \sqrt{\frac{\sum(\hat{Y}i - Y)2}{n}} \tag{1}$$

Where Y and Ŷ are measured and predicted common dry bean yield, correspondingly and *n* is sample size (McCuen et al., 2006).

3.3 Results

3.3.1 SPLSR correct coefficients (loadings)

Figure 1 shows the correct SPLSR coefficient (loadings) which were derived from the influence of common dry bean canopy spectral features at each wavelength for the three growth stages in the irrigated and rain-fed treatments. Wavelengths which resulted in zero values were not considered for predicting common dry bean yields.

At the germination and stand establishment stage, the number of selected wavebands for the Gadra variety was 40 and 48 for irrigated and rain-fed respectively, hence a decrease of hyperspectral data dimensionality by more than 80% for the irrigated and more than 75% for the rain-fed treatments (Figure2.1a). 3 and 33 wavebands for the Ukulinga variety were selected for the irrigated and rain-fed treatments, respectively (Figure2.1b), hence a dimensionality reduction by more than 80% for irrigated and rain-fed treatments. A high data dimensionality reduction was observed for the irrigated and rain-fed Caledon treatments with 3 and 6 wavebands selected, respectively (Figure2.1c).

At the branching and rapid vegetative growth stage, the number of selected wavebands were 23 for irrigated and 165 for rain-fed Gadra variety (Figure2.1d). This represents a data dimensionality reduction of 89% and 22% for the two treatments, respectively. The number of selected wavebands for the Ukulinga variety was 126 for the irrigated and 47 for the rain-fed treatments (Figure2.1e), corresponding to 40% and 81% reduction in data dimensionality in the two treatments. 44 and 28 wavebands for the Caledon variety were selected in the irrigated and rain-fed treatments, respectively (Figure2.1f), a data dimensionality reduction of 80% for the rain-fed and 86% for the irrigated treatments.

At the flowering and pod development, 8 wavebands were selected for the irrigated treatment and 11 for rain-fed (Figure2.1g), hence a dimensionality reduction of more than 90% for both treatments. For the Ukulinga variety, 40 and 32 wavebands were selected for irrigated and rain-fed treatments, correspondingly (Figure2.1h), more than 80% reduction in data dimensionality for both treatments. 30 and 24 wavebands were selected for irrigated and rain-fed treatments, translating to a data dimensionality reduction of more than 85% on both treatments for the Caledon variety (Figure2.1i).

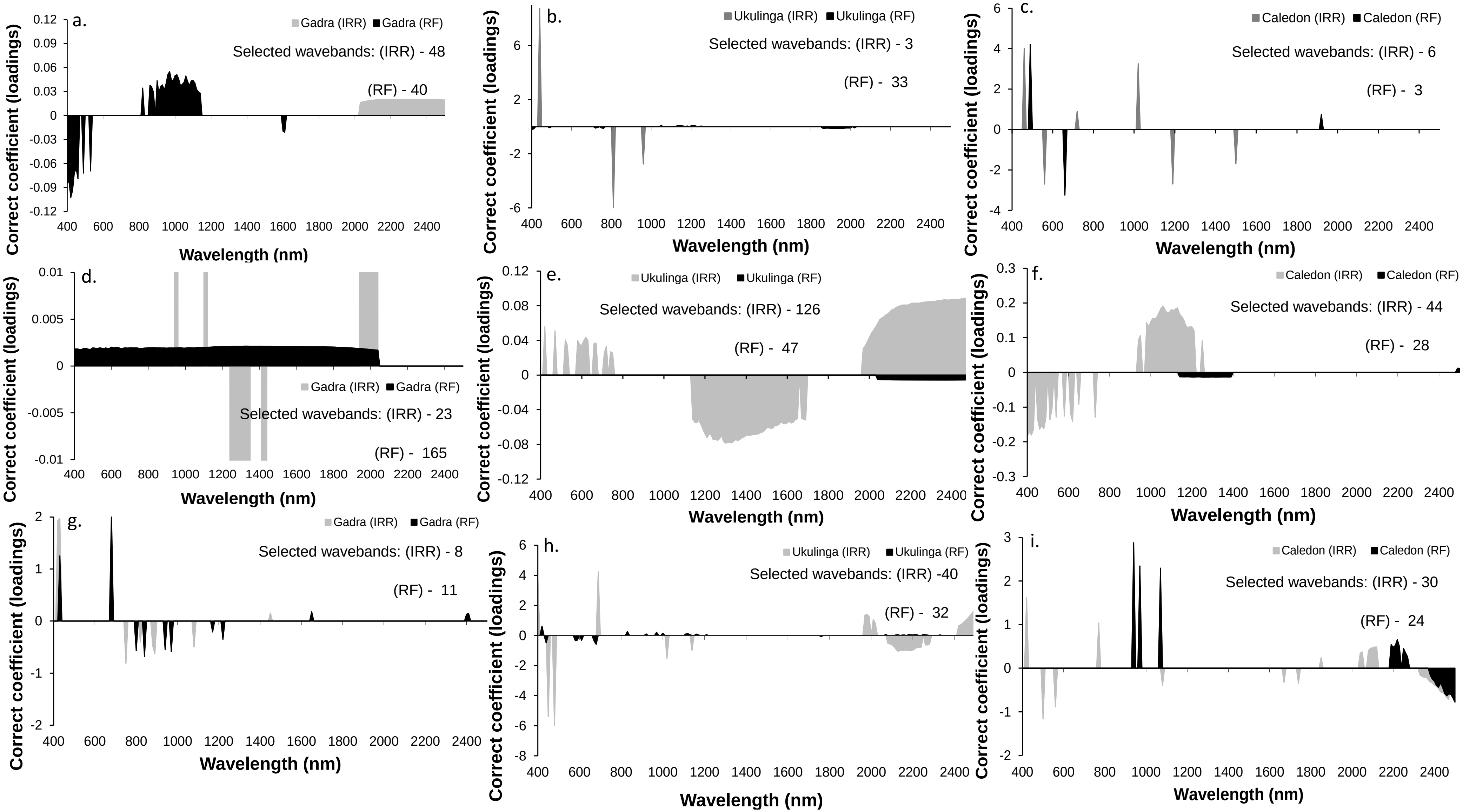

Figure 2.1: The correct SPLSR coefficients (loadings) for the selected wavelengths as determined by bootstrapped confidence intervals for all growth stages; where a, b and c = 1st growth stage (germination and stand establishment); d, e and f = 2nd growth stage (branching and rapid vegetative growth); and g, h and i = 3rd growth stage (flowering and pod development). IRR- Irrigated, RF. Rain-fed. Emergence and early vegetative growth (a, b, c), Branching and rapid vegetative growth (d, e, f), Flowering and pod formation (g,h,i).

Figure 2.2 displays the frequency at which each waveband was selected in all yield predictive models. The most selected wavebands are located at near-infrared and shortwave infrared regions of the EMS.

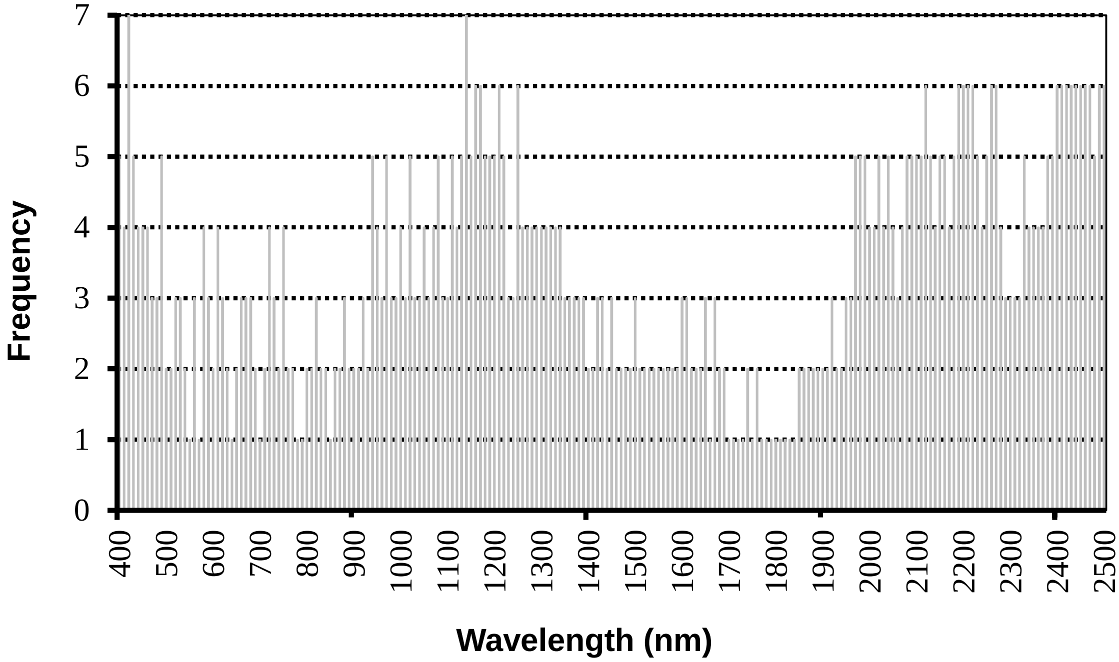

Figure 2.2: Frequency of wavebands as selected by sparse partial least squares regression (SPLSR) models for all common dry bean varieties at all phenological stages.

3.3.2 SPLSR models

A synopsis of the results achieved from the SPLSR models is shown in Table 2.1. The optimal number of components varied between 2 and 5 for each prediction model and respective growth stages of the common dry bean. Table 2.1 also shows the optimum number for both "*eta*" and "*k*" for the prediction models, as well as the number of selected wavebands (*n*) for each prediction model.

The lowest R^2 training value (0.49) for the irrigated Gadra variety was observed at the germination and stand establishment stage (Table 2.1). An R^2 training value of 0.49 was produced for the irrigated Ukulinga variety when SPLSR was trained at the first growth stage, this proved to be the lowest for the irrigated Ukulinga treatment. For the Caledon variety, both the irrigated and rain-fed treatments produced the lowest R^2 training values of 0.26 and 0.41, respectively at the germination and stand establishment stage (Table 2.1). The highest R^2 training value was seen for the irrigated Gadra (0.84) when SPLSR was trained at the branching and rapid vegetative growth. Contrastingly, the lowest R^2 training value of the rain-fed Gadra variety was observed at this growth stage (0.45) (Table 2.1). The rain-fed Ukulinga treatment produced the lowest R^2 training value for the rain-fed treatment of 0.45 (Table1). The rain-fed Gadra produced an R^2 training value of 0.79 during the flowering and pod development stage, which was the highest for the rain-fed Gadra treatment (Table 2.1). The highest R^2 training values were produced for both the irrigated and rain-fed Ukulinga variety, as 0.94 and 0.96 were observed during the flowering and pod development stage. Similarly, both the irrigated and rain-fed Caledon treatments produced the highest R^2 training values, with values of 0.86 and 0.60 correspondingly (Table 2.1).

Table 2.1: Coefficients of determination (R^2) values and number of components for training SPLSR models for common dry bean yields prediction (*n* refers to the number of selected wavebands, IRR – Irrigated, RF – Rain-fed)

Bean variety	Treatment	Growth stage	Number of components	R^2 train
Gadra	IRR	1	3 (*n*= 48, eta = 0.7, *k* =1)	0.49
		2	3 (*n*= 23, eta = 0.9,*k*=7)	0.84
		3	3 (*n*= 8, eta = 0.7,*k*=4)	0.74
	RF	1	4 (*n*= 40, eta = 0.7,*k*=4)	0.51
		2	3 (*n*= 165, eta = 0.8,*k*=1)	0.45
		3	4 (*n*= 11, eta = 0.9,*k* = 5)	0.79
Ukulinga	IRR	1	3 (*n*= 3, eta = 0.4,*k*=9)	0.49
		2	4 (*n* = 126, eta =0.6,*k*=3)	0.63
		3	6 (*n*= 40, eta = 0.8, *k* =6)	0.94
	RF	1	3 (*n*= 33, eta = 0.8,*k*=4)	0.64
		2	2 (*n*= 47, eta = 0.9, *k* =1)	0.45
		3	5 (*n*= 32, eta = 0.3,*k*=6)	0.96
Caledon	IRR	1	3 (*n*= 6, eta = 0.5,*k*=1)	0.26
		2	3 (*n*= 44, eta = 0.7,*k*=4)	0.70
		3	5 (*n*= 30, eta = 0.6,*k*=5)	0.86
	RF	1	3 (*n*= 3, eta = 0.9, *k* =6)	0.41
		2	3 (*n*= 28, eta = 0.9,*k*=2)	0.46
		3	3 (*n*= 24, eta = 0.9, *k* =4)	0.60

3.3.3 Models validation

Figure 2.3, 2.4 and 2.5 present the performance of the SPLSR prediction models. The vital statistical indicators (R^2_{loocv} and RMSE) propose that predictive models at different growth stages performed differently for each of the specific common dry bean varieties.

The SPLSR model results showed that all three common dry bean varieties could not be reliably predicted using interpolated hyperspectral reflectance data collected two weeks after seed sowing (Figure 2.3). The irrigated and rain-fed Gadra variety treatments produced R^2_{loocv} values of 0.34 and 0.25, respectively. The Caledon obtained values of 0.33 and 0.30 for both treatments, respectively while for Ukulinga, the R^2_{loocv} values were 0.02 for irrigation and 0.21 for rain-fed treatments (Figure 2.3a, b and c). The R^2_{loocv}values for all the irrigated treatments increased at the second growth stage, compared to the first growth stage. The Gadra variety produced an R^2_{loocv} value of 0.74, which was the highest R^2_{loocv} achieved in its three growth stages (Figure 2.4a). The only increase in the rain-fed treatment was seen with the Caledon variety, where the R^2_{loocv} value increased to 0.23 at the second growth stage (Figure 2.4f). There was a decrease in R^2_{loocv} in the irrigated treatment for the Gadra variety at the third growth stage (Figure 2.5a). The Caledon and Ukulinga varieties had an increase in R^2_{loocv} for the irrigated treatment (Figure 2.6b and c). Caledon and Ukulinga achieved high R^2_{loocv} values at this stage compared to earlier stages. Additionally, all three varieties of common dry bean achieved the highest R^2_{loocv} values at the last growth stage of rain-fed treatment, 0.30 for Gadra, 0.77 for Caledon and 0.32 for Ukulinga (Figure 2.5d, e and f).

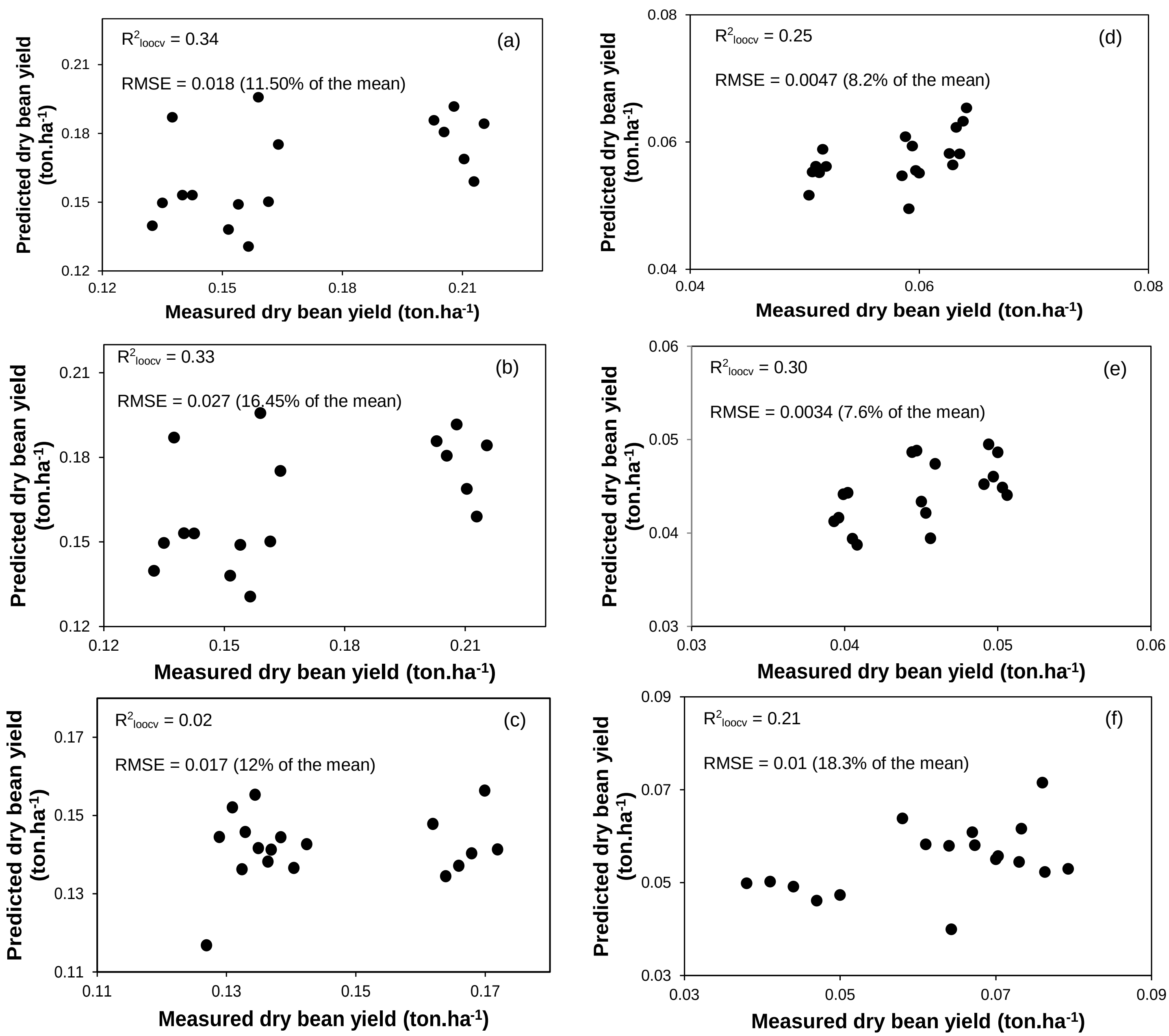

Figure 2.3: One-to-one relationship between measured and predicted common dry bean yield for validating SPLSR models using a leave one-out cross-validation method for the 1st growth stage (germination and stand establishment). Where (a) Gadra irrigated; (b) Ukulinga irrigated; (c) Caledon irrigated; (d) Gadra rain-fed; (e) Ukulinga rain-fed and (f) Caledon rain-fed.

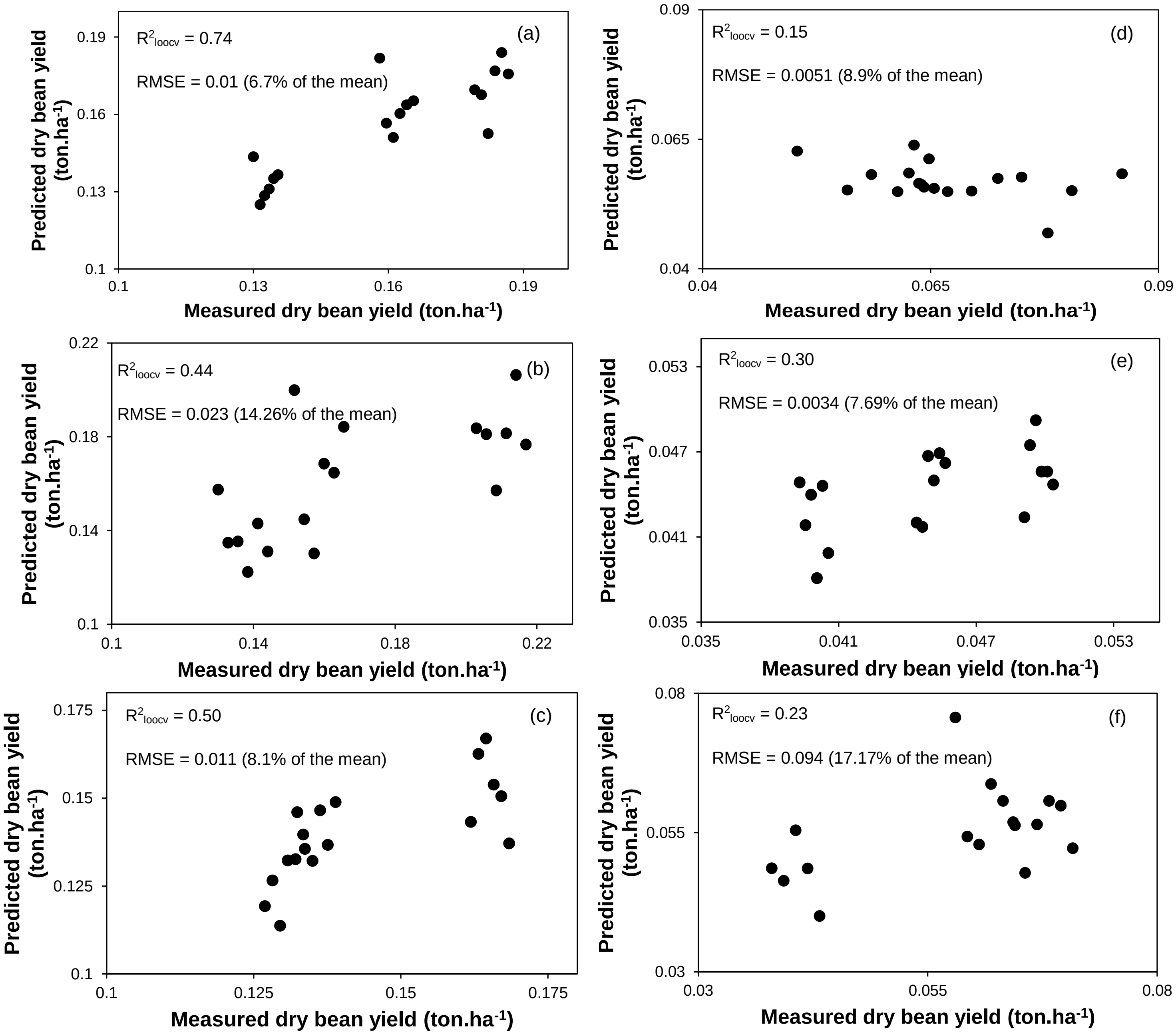

Figure 2.4: One-to-one relationship between measured and predicted common dry bean yield for validating SPLSR models using a leave one-out cross-validation method for the 2nd growth stage (branching and rapid vegetative growth). Where (a) Gadra irrigated; (b) Ukulinga irrigated; (c) Caledon irrigated; (d) Gadra rain-fed; (e) Ukulinga rain-fed and (f) Caledon rain-fed.

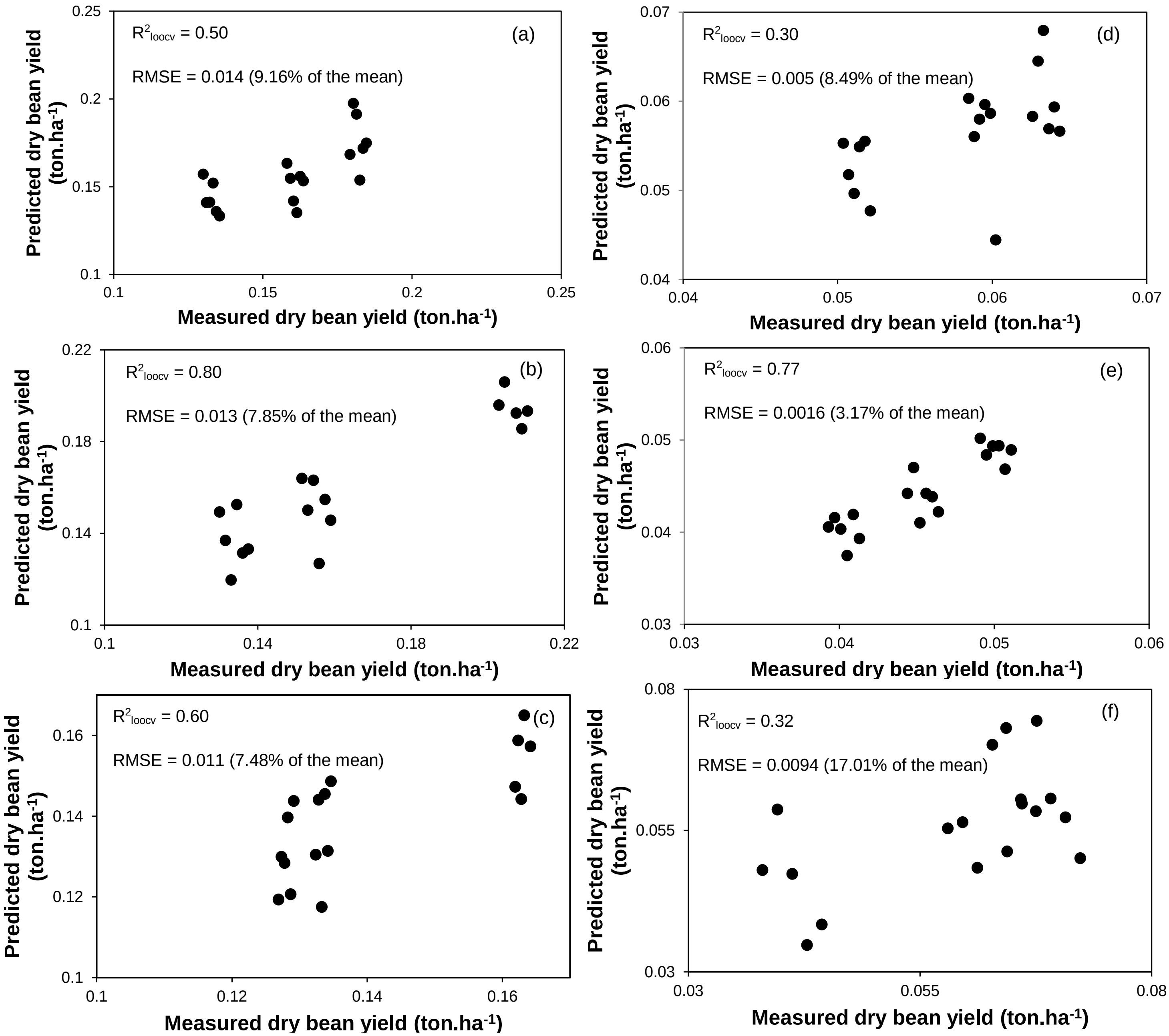

Figure 2.5: One-to-one relationship between measured and predicted common dry bean yield for validating SPLSR models using a leave one-out cross-validation method for the 3rd growth stage (flowering and pod development). Where (a) Gadra irrigated; (b) Ukulinga irrigated; (c) Caledon irrigated; (d) Gadra rain-fed; (e) Ukulinga rain-fed and (f) Caledon rain-fed.

3.4 Discussion

This study sought to determine the optimum phenological stage to predict yield for the three common dry bean varieties grown under irrigation and rain-fed conditions using the SPLSR. Generally, apart from the Gadra variety, yield prediction at the flowering and pod development stage outperformed models developed at the branching and rapid vegetative growth and germination and stand establishment stages in both treatments. This is consistent with Abdel-Rahman et al. (2014), who established a better performance of SPLSR model for predicting Swiss chard yield at a later stage in the crop growth cycle. Aparicio et al. (2000) also reported that maize yield model accuracy, based on hyperspectral vegetation indices, increased with the development of the crop. The economic produce of the common dry bean (seeds) is formed during the pod development stage, which is the stage during which most of the photosynthate is used for seeds filling (Chandhla, 2001). Hyperspectral data acquired during pod development stage are therefore expected to model common dry bean yield more accurately compared to other growth stages. During other growth stages, an efficient photosynthetic system (i.e., green leaf area index - GLAI) might not necessary have a direct relationship with common dry bean yield components (seed numbers and weight). However, the Gadra yield prediction under irrigation was better at the branching and rapid vegetative growth stage, which could be attributed to its determinate growth habit. According to Repinski et al. (2012), common dry bean varieties which exhibit a determinate growth habit allow the terminal shoot meristem to switch from a vegetative to a reproductive stage, which ultimately results in terminal inflorescence. However, varieties which exhibit an indeterminate growth habit possess a terminal meristem which remains constantly vegetative and regulates production and reproductive growth (Repinski et al., 2012). Hence, common dry bean with a determinate growth habit experience shorter flowering periods, earlier maturation and earlier maximum leaf volume. This could have ultimately resulted in the Gadra variety under irrigated conditions showing optimum time for yield prediction during the branching and rapid vegetative growth stage.

A single multivariate regression method (SPLSR) was used at three different growth stages of the common dry bean. Predicted yield was compared at varying growth stages under different watering regimes. The high R^2_{loocv} values and comparatively low RMSE values achieved by the models for the Gadra and Ukulinga varieties under irrigation conditions (Figure 2.4 and 2.5) validated the capability of ground-based multi-temporal hyperspectral data and the SPLSR algorithm for predicting common dry bean yield. However, the Caledon variety attained low R^2_{loocv} values and high RMSE across all three growth stages when grown under rain-fed conditions. This confirms findings by Kancheva (2003) who explains that bio parameters (treatment in this case) can explain spatial and temporal crop development variability. What further underlined this finding by Kancheva (2003) was the variability in RMSE values which were produced for each common dry bean variety in both treatments.

Ustin et al. (2009) note that bands located at the visible segment of the EMS correspond with leaf pigment contents such as chlorophyll, while bands at the near and shortwave infrared segments correspond with plant cell structure and leaf water content, respectively. Subsequently, with regard to band selection of SPLSR (correct coefficients loadings), no clear trend was seen in terms of band selection in any of these regions of the EMS for all bean varieties at all the three studied growth stages. However, Figure 2.1 shows that bands at the near-infrared and shortwave infrared regions of the EMS were frequently selected in all models, indicating the effect of water treatments on common dry bean yield. Near-infrared and shortwave infrared are known water spectral characteristics in the vegetation (Lillesand et al., 2004).

The number of components in the SPLSR models fell within the two to six range and in accordance with Abdel-Rahman et al. (2014), was considered appropriate for overcoming the over-fitting problem that may be associated with a high number of components. The highest R^2_{loocv} and lowest RMSE values for the irrigated treatment was obtained at the 2nd growth stage for the Gadra variety (Figure 2.4a) and the 3rd growth stage for both the Ukulinga and Caledon varieties (Figure 2.5b and c). Generally, the higher R^2_{loocv} and lower RMSE values were achieved in the final growth stage (Figure 2.5b, c, and e). As all varieties were sown on the same date, differences in growth at similar treatment can be used to explain variability in R^2_{loocv} and RMSE. Overall, the Gadra and Ukulinga varieties performed better (R^2_{loocv} values of 0.74 and 0.80 and RMSE values of 0.010% and 0.013% of the mean, respectively), hence more reliable than the Caledon (R^2_{loocv} value of 0.60 RMSE values of 0.011% of the mean). Although Weber et al. (2012) suggest that there may be weak associations between yield crop and canopy hyperspectral measurements due to environmental background noise, this study shows that under irrigation conditions, yield prediction models of reliable accuracy can be achieved for the Gadra and Ukulinga common dry bean varieties. This is further underlined by lower performance of predictive models under rain-fed conditions (R^2_{loocv} values of 0.15 (Gadra); 0.30 (Ukulinga) and 0.23 (Caledon); and RMSE values of 0.0051%; 0.094% and 0.0034% of the mean, respectively).

In cases of low prediction accuracy, particularly in the rain-fed treatment, Darvishzadeh et al. (2008) note that canopy level measurements can influence reflectance due to, among others, (LAI), structure of the canopy, foliage structure and soil background. According to Kancheva (2003), soil background significantly influence spectral reflectance characteristics of vegetation. Consequently, as the common dry bean varieties differed in phenology, and therefore varying leaf volume and background soil noise, it is plausible that the poor prediction accuracies achieved at certain growth stages can be attributed to canopy/background soil reflectance mix, an observation consistent with Mulla (2013).

This study suggests that different common dry bean variety yields can be modelled using ground-based hyperspectral measurements. However, it is worth mentioning that certain varieties could be modelled more accurately at different growth stages. For example the irrigated Gadra variety was more accurately modelled at the branching and rapid vegetative growth stage as compared to the Ukulinga and Caledon varieties, which were more accurately modelled at the flowering and pod development growth stage. The rain-fed treatment showed that all the three common dry bean varieties could be most accurately modelled during the flowering and pod development growth stage.

3.5 Conclusions

The results of this study indicate the following:

1. The optimum period for predicting common dry bean yield using ground-based hyperspectral data was at the flowering and pod development stages for the Ukulinga and Caledon varieties and the branching and rapid vegetative growth stage for Gadra variety.
2. The SPLSR algorithm is a valuable technique for modelling the common dry bean yield under both irrigated and rain-fed watering regimes.
3. R^2_{loocv} values for common dry bean yield prediction obtained under irrigation treatment were considerably higher than those obtained under rain-fed condition across the majority of growth stages.

This study provides the optimum phenological stage for predicting the common dry bean yield and offers an advanced method for yield prediction. However, it is important to note that this study was conducted under experimental conditions, hence yield prediction models which apply to on-farm conditions using airborne or space-borne hyperspectral data are required. This study could be used as a benchmark for optimum temporal sampling intervals for either airborne or space-borne hyperspectral data collection for common dry bean yield prediction models. As aforementioned, resulting models could then be used to make informed decisions on pricing, marketing, handling and planning for future surpluses or shortages of common dry bean produce grown under both rain-fed and irrigation.

Chapter four

Objectives reviewed and conclusions

4.1 Introduction

The focus of this study was to examine the utility of multi-temporal hyperspectral groundbased remotely sensed data in varietal discrimination and yield estimation of the common dry bean (*Phaseolus vulgaris* L*.)* under different irrigation regimes. In this chapter, the aims and objectives established in the introductory chapter (Chapter 1) are reviewed against the findings. The chapter also highlights the major conclusions and recommends potential opportunities for future research.

4.2 Aims and objectives reviewed

Objective (1):

To determine the potential of ground-based hyperspectral data to discriminate among three different varieties of common dry bean using partial least squares –discriminant analysis (PLS-DA).

Previous studies on crop discrimination using remotely sensed data have been hindered by spatial and spectral limitations. This study explored the potential of ground-based multi-temporal hyperspectral data to discriminate among three common dry bean varieties. Whereas studies such as Moriotto et al. (2013) and Wilson et al. (2014) have shown the potential of hyperspectral remote sensing for crop discrimination, however data dimensionality and redundancy, that characterise hyperspectral data has limited its wide adoption. This study adopted the partial least squares discriminant analysis (PLS-DA) to reduce data dimensionality and redundancy and to discriminate among three common dry bean varieties. Results showed that adoption of PLS-DA on hyperspectral remotely sensed data could be used to reliably discriminate among the three common dry bean varieties. Furthermore, discrimination accuracies increased at each of the three specific growth stages which were marked for temporal data sampling.

Objective (2):

To determine the optimal period within common dry bean's life cycle for varietal discrimination using ground-based hyperspectral remotely sensed data.

Ra (2013) underlines the importance of crop discrimination as it is described as an important step for development and management for monitoring of crops systems. Moreover, precise crop identification and discrimination allows for government sectors and researchers to complete nation-wide crop inventory in an economically viable and non-destructive way (Wilson et al., 2014). Gomez-Casero et al. (2010) notes that crop varieties, at different phonological stages, are commonly characterised by unique spectral reflectance characteristics. Previous studies using hyperspectral remote sensing have failed to account for the implication of distinct phenological cycles in varietal discrimination. This study isolated three common dry bean's distinct growth stages and sought to determine the optimal period for varietal discrimination. Results identified the final growth stage (flowering and pod development) as the optimal growth stage for discriminating among the common dry bean varieties. Using all and selected wavebands, the growth stage produced overall accuracies of 93 and 100% for both irrigated and rain-fed treatments, respectively. This finding provides valuable insights on optimal periods for adoption of remotely sensed data for the common dry bean varietal discrimination and other vegetation species in general.

Objective (3):

To estimate the common dry bean's yield at different phenological cycles using the partial least squares regression (SPLSR) on ground-based hyperspectral data.

Timely detection and management of limitations related to crop yield indicators could assist in increasing yield as well as subsequent profit (Panda, 2010). The capability of accurately predicting field crops such as the common dry bean permits producers, buyers, and economic agencies to make decisions regarding crop management, pricing and market availability (Ma et al., 2001). Previous hyperspectral data treatment techniques such as partial least squares regression (PLSR) and piecewise linear regression have achieved relative success in yield prediction. However, these techniques fail to mechanically choose significant predictors in the model (Chun and Keles, 2010). In contrast, SPLSR executes a sparsity solution during the process of data transformation and simultaneously selects the appropriate variables for predicting the feature of interest. This unique ability of SPLSR separates it from previous hyperspectral treatment techniques thus providing a distinctive advantage when dealing with hyperspectral remotely sensed data. Hence, to achieve this objective, SPLSR was used to analyse hyperspectral data at different growth stages of the common dry bean. Generally, it was found that the irrigated common dry bean varieties outperformed the rain-fed varieties at all growth stages. Accordingly, this study proves that ground-based hyperspectral remotely sensed data have the potential to predict yield of the common dry bean for different varieties using SPLSR.

Objective (4):

Determine which stage in the life cycle of common dry bean is optimal for yield prediction based on ground-based hyperspectral data.

The value of timeliness is emphasised by Atzberger (2013), who states that it plays a major role in fundamental agricultural statistics and monitoring systems. Traditional yield prediction techniques are time consuming and labour intensive (Drummond et al., 2003), which inevitably impacts timeliness of data for statistical and monitoring purposes. Earth observation techniques such as hyperspectral remotely sensed data allow for timely, efficient, robust and cost effective yield prediction. Thus, in order to determine the optimal growth stage for common dry bean yield prediction, reflectance at different stages of growth of three varieties and ultimate yield were compared. The Ukulinga and Caledon varieties were best predicted at the branching and rapid vegetative growth stage, while the Gadra variety was best predicted at the flowering and pod development stage on irrigated treatment. The Gadra and Ukulinga varieties generated higher R^2 values, (0.74 and 0.80) while the Caledon variety generated a lower R^2 value (0.58) on irrigated treatments. In contrast, the three common bean varieties were best predicted at the flowering and pod development stage for the rain-fed treatment. Under this treatment, only the Ukulinga variety generated a higher R^2 value (0.77), while the Gadra and Caledon varieties generated 0.30 and 0.32 R^2 values, respectively. Consequently, from these results it can be concluded that the optimal period for predicting the common dry bean yield is at the flowering and pod development stage for the Ukulinga and Caledon varieties and the branching and rapid vegetative growth stage for the Gadra variety.

4.3 Conclusions

The aim of this study was to assess the potential of ground-based hyperspectral remotely sensed data in discriminating between common dry bean (*Phaseolus vulgaris* L*.)* varieties and predicting yield of the common dry bean varieties. This study has shown that ground-based hyperspectral remotely sensed data can be used to discriminate and predict yield for the common dry bean varieties. This conclusion is based on aforementioned research questions; can ground-based hyperspectral remotely sensed data be used in discriminating common dry bean varieties?

When PLS-DA was used in conjunction with the common dry bean's spectral reflectances, overall accuracies of more than 80% were achieved beyond the branching and rapid vegetative growth stage. At these stages, producer's and user's accuracies ranged from 60 to 100% and 80 to 100%, respectively. It can therefore be concluded that ground-based hyperspectral remotely sensed data can be used to discriminate among the common dry bean varieties.

What is the optimal period within the common dry bean's growth cycle for yield estimation?

Using SPLSR yield models, there was a variability in optimal period (growth stage) for yield prediction. The Gadra variety yield was best predicted at the branching and rapid vegetative growth stage, while the yield of Ukulinga and Caledon varieties was best predicted at the flowering and pod development stage. This finding can be attributed to differences in spectral reflectance at the identified phenological stages (Gomez-Casero et al., 2010). Based on these findings, it can be concluded that ground-based hyperspectral remotely sensed data have immense potential in the discrimination and prediction of the common dry bean. In this study, the combination of multi-temporal and high spectral resolution provided by ground-based hyperspectral measurements allowed robust investigation of the common dry bean spectral characteristics.

These research findings are particularly valuable for determination of useful bands for satellite image acquisition and mapping to broader spatial scales like larger farms and commercial production units. Hence, future research which seeks to discriminate between field crop varieties or predict yield should attempt to achieve this on a wider scale so as to investigate the potential for national applications. Future research should also attempt to discriminate between and predict yield of other varieties of the common dry bean which were not utilised in this research. The production of a universal model for yield prediction should also be investigated with regards to future endeavours as this would mitigate the need for climatic based models.

References

Abdel-Rahman, E.M., Mutanga, O., Odindi, J., Adam, E., Odindo, A., and Ismail, R., 2014: A comparison of partial least squares (PLS) and spare partial least squares regression for predicting yield of Swiss Chard grown under different irrigation water sources using hyperspectral data, Computers and Electronics in Agriculture, 106, 11-19.

Abdel-Rahman, E. M., Ahmed, F. B. and Ismail, R., 2012: Random forest regression for sugarcane yield prediction in Umfolozi, South Africa based on Landsat TM and ETM+ derived spectral vegetation indices. In Sugarcane: Production, Cultivation and Uses. J. F. Goncalves and K. D. Correia (eds). NOVA Science Publishers, Inc. Hauppauge NY, USA. 257–284.

Adam, E., and Mutanga, O., 2009: Spectral discrimination of papyrus vegetation (*Cyperus papyrus* L.) in swamp wetlands using field spectrometry, Journal of Photogrammetry and Remote Sensing, 64, 612-620.

Alfanador, N.L., 2014: Important Variable Selection in Partial Least Squares for Industrial Process Understanding and Control, PhD thesis, Radboud University Nijmegen, Netherlands.

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.

Analytical Spectral Devices: ASD, 2005: Handheld Spectroradiometer: User Guide Version 4.05, Analytic Spectral Devices, Suite A Boulder, USA.

Analytical Spectral Devices: ASD., 2008: Handheld Spectroradiometer: User Guide Version 4.05, Analytic Spectral Devices, Suite A Boulder, USA.

Apan, A., Held, A., Phinn, S. and Markley, J., 2004: Spectral discrimination and classification of sugarcane varieties using EO-1 Hyperion hyperspectral imagery. In Proceeding of Asian Conference on Remote Sensing, Chiang Mai, Thailand, 2004.

Aparicio, N., Villegas, D., Casadesus, J., Araus, J.L and Royo, C., 2000: Spectral vegetation indices as non-destructive tools for determining durum wheat yield, Agronomy Journal, 92, 83-91.

Atzberger, C., 2013: Advances in Remotesensing of agriculture: context description, existing operational monitoring systems and major information needs, Remote Sensing, 5, 949-981.

Beninger, C.W, Hosfield, G.L., 2003: Antioxidant activity of extracts, condensed tannin fractions, and pure flavonoids from *Phaseolus vulgaris* L. seed coat colour genotypes, Journal of Agricultural Food Chemistry, 51, 7879-7883.

Chandhla, J., 2001: Optimisation of dry bean (*Phaseolus vulgaris* L.) seed production under greenhouse conditions, MSc thesis, University of Pretoria, South Africa.

Chun, H., and Keles, S., 2010: Sparse partial least squares regression for simultaneous dimension reduction and variable selection, Journal of the Royal Statistic Society, 72, 3-25.

Chung, H., and Keles, S., 2010: Sparse partial least squares classification for high dimensional data, Statistical Applications in Genetics and Molecular Biology, 9, 1544-6115.

Cho, M.A., Skidmore, A.K., Corsi, F., Van Wieren, S.E., Sobhan, I., 2007: Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression, International Journal of Applied Earth Observation and Geoinformation, 9, 414–424.

Chong, I.G., and Jun, C.H., 2005: The performance of some variable selection methods when multicollinearity is present, Chemometrics and Intelligent Laboratory Systems, 78, 103-112.

Cochrane, M.A., 2000: Using vegetation reflectance variability for species level classification of hyperspectral data, International Journal of Remote Sensing, 21, 2075-2087.

Cody, R., 2007: Spectral Resolution, Geography 581, available online at http://web.pdx.edu/~jduh/courses/Archive/geog481w07/Students/Cody_SpectralResolution.pdf (accessed September 2014).

Craig, M., and Atkinson, D., 2013: A literature review of crop area estimation, United Nations Food and Agriculture Organization, available online at http://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/GS_SAC_2013/Improving_methods_for_crops_estimates/Crop_Area_Estimation_Lit_review.pdf (accessed July 2014).

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

Darvishzadeh, R., Skidmore, A., Atzberg, C., and van Wieren, S., 2008: Estimation of vegetation LAI from hyperspectral reflectance data: Effects of soil type and plant architecture, International Journal of Applied Earth Observation and Geoinformation 1, 10, 358-373.

Department of Agriculture, Forestry and Fisheries, 2010:Dry Beans: Production guideline, available online at http://www.nda.agric.za/docs/Brochures/prodGuideDryBeans.pdf (accessed October 15, 2014).

Department of Agriculture, Forestry and Fisheries., 2013: Trends in the Agricultural Sector, available online athttp://www.econostatistics.co.za/TrendsinSAAgriculture13.pdf (accessed October 15, 2014).

Debouck, D., and Hidalgo, R., 1986: Morphology of the Common Bean Plant *Phaseolus Vulgaris*, Cali, Columbia.

Dorigo, W.A., Zurita-Milla, R., de Wit, A.J.W., Brazile, J., Singh, R., and Schaepman, M.E., 2007: A review on reflective remote sensing and data assimilation techniques for enhanced agro-ecosystem modelling, International Journal of Applied Earth Observation, 9, 165-193.

Drummond, S.T., Sudduth, K.A., Joshi, A., Birrell, S.J., and Kitchen, N.R., 2003: Statistical and neural methods for site-specific yield prediction, American Society of Agricultural Engineers, 46, 5-14.

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

Fourie, D., 2002: Distribution and severity of bacterial diseases on dry beans (*Phaseolus vulgaris* L.) in South Africa, Journal of Phytopathology, 150, 220-226.

Fu, Y., Yang, G., Wang, J., Song, X., and Feng, H., 2014:Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements, Computers and electronics in Agriculture, 100, 51-59.

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

Goa, X., Huete, A.R., Ni, W., and Feng, H., 2000: Optical-biophysical relationships of vegetation spectra without background contamination, Remote Sensing of Environment, 74, 609-620.

Gomez-Casero, M.T., Castillejo-Gonzalez, I.L., Garcia-Ferrer, A., Pena-Barragan, J.M., Jurado-Exposito, M., Garcia-Torres, L., and Lopez-Granados, F., 2010: Spectral discrimination of wild oat and canary grass in wheat fields for less herbicide application, Agronomy for Sustainable Development, 30, 689-699.

Govender, M., Chetty, K., and Bulcock, H., 2007: A review of hyperspectral remote sensing and its application in vegetation and water resource studies, Water S.A., 33, 145-151.

Govender, M., Dye, P.J., Weiersbye, I.M., Witkowski, E.T.F., and Ahmed, F., 2009: Review of commonly used remote sensing and ground-based technologies to measure plant water stress, Water S.A., 35, 741-752.

Hansen, P.M., Jørgensen, J.R., and Thomsen, A., 2002: Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression, The Journal of Agricultural Science, 139, 307-318,.

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.

Humphreys, J.R., Wapstra, J.M.O., Harbard, J.L., Davies, N.W., Griffin, A.R., Jordan, G.J., and Potts, B.M., 2008: Discrimination between seedlings of *Eucalyptus globulus* E. nitens and their F1 hybrid using near-infrared reflectance spectroscopy and foliar oil content, SilvaeGenetica, 57, 4-5.

Kancheva, R., 2003: Main principles in vegetation spectrometric studies, 50 years University of Mining and Geology “St. Ivan Rilski”, Geology and Geophysics, 46, 351-353.

Koedsin, W., and Vaiphasa, C., 2013: Discrimination of tropical mangroves at the species level withEO-1 Hyperion data, Remotes Sensing, 5, 3562-3582.

Kornegay, J., White, J.W., and de la Cruz, O., 1992: Growth habit and gene pool effects on inheritance of yield in common bean, Euphytica, 62, 171-180.

Kumar, P.V., Ramakrishna, Y.S., Bhaskara-Roa, D.V., Sridhar, G., Srinivasa-Roa, G., and Roa, G.G.S.N., 2005: Use of remote sensing for drought stress monitoring, yield prediction, and varietal evaluation in castor beans, International Journal of Remote Sensing, 26, 5525-5534.

Kramer, N., and Sugiyama, M., 2011: The degrees of freedom of partial least squares regression, Journal of the American Statistical Association, 106, 697-705.

Lillesand, T., Kiefer, T., and Chipman, J., 2004: Remote sensing and Image Interpretation, 5, John Wiley and Sons, New York, USA.

Lindström, S.W., Geladi, P., Jonsson, O., and Pettersson, F., 2011: The importance of balanced data sets for partial least squares discriminant analysis: classification problems using hyperspectral imaging data, Journal of Near Infrared Spectroscopy, 19, 233-241.

Lucas, R., Bunting, P., Paterson, M., and Chisholm, L., 2008: Classification of Australian forest communities using aerial photography, CASI and HyMapdata, Remote Sensing of Environment, 112, 2088–2103.

Ma, B.L., Dwyer, L.M., Costa, C.C., Cober, E.R., Morrison, M.J., 2001:Early prediction of soybean yield from canopy reflectance measurements, Agronomy Journal, 93, 1227-1234.

Makanda, I., Derera, J., Tongoona, P., and Sibiya, J., 2012: Genetic and GGE biplot analyses of sorghum germplasm for stem sugar traits in Southern Africa, African Journal of Agricultural Research, 72, 212-223.

McCuen, R.H., Knight, Z., and Cutter, A.G., 2006: Evaluation of the Nash-Sutcliffe efficiency index, Journal of Hydrologic Engineering, 11, 597-602.

Mingwei, Z., Qingbob, Z., Zhongxinb, C., Jia, L., Yong, Z., and Chongfa, C., 2008: Crop discrimination in Northern China with double cropping systems using Fourier analysis of time-series MODIS data, International Journal of Applied Earth Observation and Geoinformation, 10, 476-485.

Misra, G., 2012: Mapping Specific Cops and their Phenology – Multi Sensor and Temporal Approach, MSc Thesis, University of Twente, Netherlands.

Monteiro, P.F.C., Filho, R.A., Xavier, A.C., and Monteiro, R.O.C., 2012: Assessing biophysical variable parameters of bean crop with hyperspectral measurements, Scientia Agricola, 69, 87-94.

Moran, M.S., Y. Inoue, and E.M. Barnes, 1997: Opportunities and limitations for image-based remote sensing in precision crop management, Remote Sensing of Environment, 61, 319-346.

Mariotto, I., Thenakbail, P.S., Huete, A., Slonecker, E.T., and Platonov, A., 2013: Hyperspectral versus multispectral crop-productivity modelling and type discrimination for the HyspIRI mission, Remote Sensing of Environment, 139, 291-305.

Mulla, D.J., 2013: Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps, Biosystems Engineering, 114, 358-371.

Mutanga, O., Skidmore, A., and Prins, H., 2004: Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features, Remote Sensing of Environment 89, 393-408.

Nuarsa, I.W., Nishio, F., and Hongo, C., 2011: Relationship between rice spectral and rice yield using MODIS data, Journal of Agricultural Science, 3, 80-88.

Nuland, D.S., and Schwartz, H.F., 1989: A Visual description of the common bean plant four major growth periods,Reports of Bean Improvement Cooperative and National Dry Bean Council Research Conference, 32, 16-17.

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.

Panda, S.S, Ames, D.P, and Panigrahi, S., 2010: Application of vegetation indices for agricultural crop yield prediction using neural network techniques, Remote Sensing, 2, 673-696.

Panneton, B., Guillaume, S., Samson, G., and Rodger, J.M., 2011: Discrimination of corn from monocotyledonous weeds with ultraviolet (UV) induced fluorescence, Applied Spectroscopy, 65, 10-19.

Peerbhay, K., Mutanga, O., and Ismail, R., 2013a: Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu–Natal, South Africa, ISPRS Journal of Photogrammetry and Remote Sensing, 79, 19-28.

Peerbhay, K., Mutanga, O., and Ismail., 2013b: Investigating the capability of few strategically placed Worldview-2 multispectral bands to discriminate forest species in KwaZulu-Natal, South Africa, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 307-316.

Prasad, A.K., Chai, L., Singh, R.P., and Kafatos, M., 2006: Crop yield estimation model for Iowa using remote sensing and surface parameters, International Journal of Applied Earth Observation and Geoinformation, 8, 26-33.

Pérez-Enciso, M., and Tenenhaus, M., 2003: Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach, Human Genetics, 112, 581-592.

Pontius, R.G., and Millones, M., 2011: Death to kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment, International Journal of Remote Sensing, 32, 4407-4429.

Ra, E.A.M., 2013: A Dissertation on Crops Discrimination Researches using SAR Data, High Spatial Studies Institute Mario Gulich, Universidad Nacional de Cordoba, Argentina.

R Develpoment Core Team, 2015. R: A language and environment for statistical computing, R Foundation for statistical computing, Vienna, Austria.

Repinski, S.L., Kwak, M., and Gepts, P., 2012: The common bean growth habit gene PvTFL1y is a functional homolog of Arabidopsis TFL1, Theoretical and Applied Genetics, 124, 1539-1547.

Richter, K., Atzberger, C., Hank, T.B., and Mauser, W., 2012: Derivation of biophysical variables from earth observation data: Validation and statistical measures, Journal of Applied Remote Sensing, 6, 1-23.

Rubiales, D., 2014: Phaseolus: A new world gift to mankind – Why are common beans so common?, The journal of the International Legume Society, 2, 3-58.

Schmidt, K. S. and Skidmore, A. K., 2001: Exploring spectral discrimination of grass species in African rangelands, International Journal of Remote Sensing, 22, 3421-3434.

Serrano L., Filella, I., and Penuelas, J., 2000: Remote sensing of biomass and yield of winter wheat under different nitrogen supplies, Crop Science, 40, 723-731.

Shaw, G.A., and Burke, H.K., 2003: Spectral imaging for remote sensing, Lincoln Laboratory Journal, 14, 3-28.

Siddiq, M., Ravi, R., Harte, J.B., and Dolan, K.D., 2010: Physical and functional characteristics of selected dry bean (*Phaseolus vulgaris* L.) flours, Food Science and Technology, 43, 232-237.

Singh, V., Jose, J., Dalwai, A., and Kalubarme, M.H., 2013: Geo-informatics and remote sensing applications for village level crop inventory in Gujarat State, India, Asian Journal of Geoinformatic, 13, 31-38.

Sobhan, I., 2007: Species discrimination from a hyperspectral perspective, PhD thesis, Wageningen University, Netherlands.

Sun, J., 2000: Dynamic monitoring and yield estimation of crops by mainly using the remote sensing technique in China, Photogrammetric Engineering and Remote Sensing, 66, 645-650.

Tanagra Project., 2008: http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html, (accessed August 22, 2014).

Thenkabail, P.S., Enclona, E.A., Ashton, M.S., and van der Meer, B., 2004: Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications, Remote sensing of the Environment, 91, 354-376.

Tian, Q., Tong, Q., Pu, R., Guo, X., and Zhao, C., 2011: Spectroscopic determination of wheat water status using 1650-1850 nm spectral absorption features, International Journal of Remote Sensing, 22, 2329–2338.

Ustin, S.L., Gitelson, A.A., Jacquemound, S., Schaepman, M., Asner, G.P., Gamon, J.A., and Zarco-Tejada, P., 2009: Retrieval of foliar information about plant pigment systems from high resolution spectroscopy, Remote sensing of Environment, 113, 567-577.

Van Niel, T.G., and McVicar, T.R., 2004: Determining temporal windows for crop discrimination with remote sensing: a case study in south-eastern Australia, Computers and Electronics in Agriculture, 45, 91-108.

Wilson, J.H., Zhang, C., and Kovaes, J.M., 2014: Separating crop species in North-eastern Ontario using hyperspectral data, Remote Sensing, 6, 925-945.

Weber, V.S., Araus, J.L., Cairns, J.E., Sanchez, C., Melchinger, A.E., and Orsini, E., 2012: Prediction of grain yield using reflectance spectra of canopy and leaves in maize plants grown under different water regimes, Field Crops Research, 128, 82-90.

Wold, S., Sjostrom, M., and Eriksson, L., 2001: PLS-regression: A basic tool of chemometrics, Chemometrics and Intelligent Laboratory Systems, 58, 109-130.

Wolter, P.T., Townsend, P.A., SturtevantB.R., and Kingdon, C.C., 2008: Remote sensing of the distribution and abundance of host species for spruce budworm in Northern Minnesota and Ontario, Remote Sensing of Environment, 112, 3971-3982.

Wortmann, C.S., Kirkby, R.A., Eledu, C.A., and Allen, D.J., 1998: Atlas of Common Bean (*Phaseolus vulgaris* L.) Production in Africa, International Centre for Tropical Agriculture (CIAT), Cali, Columbia.

Zhao, D., Reddy, K.R., Kakani, V.G., Read, J.J., and Koti, S., 2007: Canopy reflectance in cotton for growth assessment and lint yield prediction, European Journal of Agronomy, 26, 335–344.

Zhang, H., Lan, Y., Suh, C.P., Westbrook, J.K., Lacey, R., and Hoffmann, W.C., 2012: Differentiation of cotton from other crops at different growth stages using spectral properties and discriminant analysis, American Society of Agricultural and Biological Engineers, 55, 1623-1630.

Zwiggelaar, R., 1998: A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops, Crop Protection, 17, 189-206.