

VISIBE AND INFRARED SPECTRAL CHARACTERISATION OF CHINESE CABBAGE (*BRASSICA RAPA L. SUBSPECIES CHINENSIS*), GROWN UNDER DIFFERENT NITROGEN, POTASSIUM AND PHOSPHORUS CONCENTRATIONS

B. B. Mokoatsi^{1*}, S. G. Tesfamichael¹, H. Araya², & M. Mofokeng²

¹Dept. Of Geography, Environmental Management and Energy Studies, University of Johannesburg, Johannesburg 2092, South Africa – b.mokoatsi@gmail.com

²Agricultural Research Council (ARC), Roodeplaat Vegetable and Ornamental Plant Institute, Pretoria 0001, South Africa - ArayaH@arc.agric.za

KEYWORDS: Chinese cabbage, spectral measurements, fertilisers, gradient boosting model, random forest model.

ABSTRACT:

There is a need to intensify research efforts on improving productivity of indigenous vegetables in South Africa. One research avenue is operationalizing remote sensing techniques to monitor crop health status. This study aimed at characterising the spectral properties of Chinese cabbage (*Brassica Rapa L. subspecies Chinensis*) grown under varying fertilizer treatments: nitrogen (0 kg/ha, 75 kg/ha, 125 kg/ha, 175 kg/ha and 225 kg/ha), phosphorus (0 kg/ha, 9.4 kg/ha, 15.6, 21.9 kg/ha and 28.1 kg/ha) and potassium (0 kg/ha, 9.4 kg/ha, 15.6 kg/ha, 21.9 kg/ha and 28.1 kg/ha). Visible and infrared spectral measurements were taken from a total of 60 samples inside the laboratory. Contiguous spectral regions were plotted to show spectral profiles of the different fertilizer treatments and then classified using gradient boosting and random forest classifiers. ANOVA revealed the potential of spectral reflectance data in discriminating different fertiliser treatments from crops. There was also a significant difference between the capabilities of the two classifiers. Gradient boost model (GBM) yielded higher classification accuracies than random forest (RF). The important variables identified by each model improved the classification accuracy. Overall, the results indicate a potential for the use of spectroscopy in monitoring food quality parameters, thereby reducing the cost of traditional methods. Further research into advanced statistical analysis techniques is needed to improve the accuracy with which fertiliser concentrations in crops could be quantified. The random forest model particularly requires improvements.

1. INTRODUCTION

Environmental degradation, food insecurity and malnutrition are increasingly becoming a concern globally and in South Africa. South Africa needs to ensure a healthy agricultural industry that contributes to the gross domestic product (GDP), food security, social welfare, job creation and ecotourism; while adding value to natural resources. Sustainable agriculture has been promoted as an alternative to conventional farming systems. Precision agriculture has been adopted by some producers to increase efficiencies of fertiliser and water inputs and to optimise crop growth and product quality (Blignaut et al., 2014). Precision farming uses regular, detailed soil and leaf mineral analyses which are the basis of precise fertiliser recommendations.

As per the South African Year Book published by the Department of Government Communications and Information System (GCIS, 2015), the agricultural industry has been greatly influenced by diverse climatic conditions in various regions. Commercial farmers who produce 95% of South Africa's food are heavily dependent on fertilisers to maintain yield levels (Goldblatt et al., 2009; Blignaut et al., 2014; GCIS, 2015). This results in roughly 60% of the cropland area in South Africa being moderately to severely acidic in the topsoil, while 15% of the cropland area is affected by subsoil acidity (Goldblatt et al., 2009; Blignaut et al., 2014; GCIS, 2015). In addition, approximately 1.3 million hectares of croplands are under irrigation to ensure year-round food supply in various climatic regions (GCIS, 2015). Therefore, water scarcity and soil degradation are pressing issues affecting crop production.

While there is a need for more sustainable farming methods, there is also a need to expand South Africa's food resources and enhance food quality assessment techniques. Remote sensing is one method that can improve the effectiveness of crop production management and has shown considerable potential to monitor food quality. The basic principle of remote sensing is that all materials, due to difference in their chemical composition and inherent physical structure, absorb, scatter, reflect and emit electromagnetic energy in distinctive patterns at specific wavelengths (Reddy, 2008; Ortenberg, 2009; Elmasry et al., 2012). This unique characteristic is called a spectral signature. Each material has a distinctive spectral signature that is indicative of its chemical composition and other characteristics (Reddy, 2008; Elmasry et al., 2012).

Traditionally, agricultural remote sensing used multispectral imagery. With advances in sensor technology over the past two decades, the introduction of hyperspectral remote sensing imagery to agriculture provided more opportunities for field level information extraction (Yao et al., 2012). Hyperspectral remote sensing acquires information about objects in several (usually hundreds) narrow, contiguous wavelengths of the electromagnetic radiation (Carroll et al., 2008; Huang & Asner, 2009; Jensen, 2014; Alparone et al., 2015). Several studies have successfully used hyperspectral remote sensing to monitor crop nutrients (Anawar et al., 2012; Feng et al., 2015; Basso et al., 2016), detect weeds (Eddy et al., 2013; Shapira et al., 2013), and manage diseases and pests (Huang et al., 2013).

Limited research has been conducted on Chinese cabbage characteristics – structure (van Averbeke et al., 2007), growth period and environmental conditions (Kalisz, 2011) as well as optimal fertiliser inputs for maximum yields (Ahmad et al., 2014; Li et al., 2015). Thus, there is limited-to-nil research conducted

to analyse the spectral properties of Chinese cabbage relative to various fertiliser treatments. The aim of this paper was, therefore, to characterise the spectral properties of Chinese cabbage grown under varying concentrations of Nitrogen, Potassium and Phosphorus. Specific objectives are (1) to investigate the performance of spectroscopy in discriminating fertiliser concentrations, and (2) to compare the performances of gradient boosting and random forest classifiers in categorising fertiliser treatments from spectra.

2. METHODOLOGY

2.1 Study Area

Chinese cabbage was cultivated at the Roodeplaat Vegetable and Ornamental Plant Institute of the Agricultural Research Council (ARC-VOPI). The institute is located approximately 25 km north-east of Pretoria, as shown in figure 1. Pretoria, South Africa's capital city, is governed by the City of Tshwane Metropolitan Municipality. The main geological formations around Roodeplaat are the Daspoort, Timeball Hill and Magaliesberg formations from the Pretoria Group. The region is characterised by ridge and valley topography. Prominent ridges include the Daspoort rant, Piemeefrant, Bronberg and Magaliesberg. Across the ARC-VOPI is the Roodeplaat Dam which is fed by four streams – Roodeplaatspruit, Pienaars River, Moreletaspruit and Hartbeesspruit. Roodeplaat normally receives about 573 mm of rain per year, with most rainfall occurring during summer. It receives the lowest rainfall (0 mm) in June and the highest (11 mm) in January. The average midday temperatures for Roodeplaat range from 18.3° C in June to 27.5° C in January. The region is the coldest during July when the mercury drops to 1.7° C on average during the night. January is the hottest month with average maximum temperatures reaching 30° C.

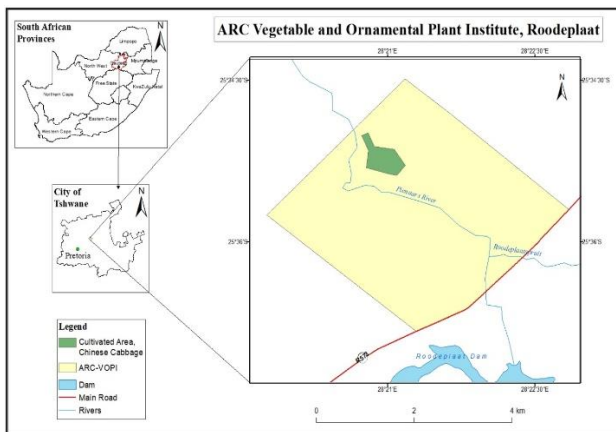


Figure 1. Location of the Study Area

2.2 Sampling

Chinese cabbage was systematically cultivated across a demarcated field, as shown in figure 2. Seedling and transplanting were performed in a controlled environment (glass house) prior to cultivation. Variable-rates of nitrogen, phosphorus and potassium (N: P: K) fertiliser were applied to Chinese cabbage, as detailed in table 1. Since K is not available to the crop immediately, it was applied before planting. Whereas, N was applied on top-soil, then irrigated to percolate. This N: P: K trial was replicated 4 times in different blocks (demarcations). There was little-to-nil within-block variation as the cultivation was strictly controlled. A drip-irrigation system was used to water the field at variable-rates in the morning. The first harvest

was carried out on 2nd February 2017, after 3 months of cultivation. Crops were harvested from 3 inner rows and not along the boundaries of the block. The crops were bagged and labelled in a systematic manner according to their N: P: K treatment levels.



Figure 2. Sampling technique for Variable-Rate Fertilizer Treatments

Treatment	Nutrient Applied (kg/ha)		
	N	P	K
T ₁ (control)	0	0	0
T ₂	75	9.375	9.375
T ₃	125	15.625	15.625
T ₄	175	21.875	21.875
T ₅	225	28.125	28.125

Table 1. NPK Concentrations per Treatment Level

2.3 Data Analysis

2.3.1 Processing: the spectral measurements of each treatment group were averaged to overcome individual scan variations. Noisy bands were removed at several regions, namely 340 nm – 494 nm, 603.6 nm – 663.9 nm, 920.7 nm – 1051.6 nm and 2123.1 nm – 2503.4 nm. Reflectance curves were subsequently plotted to identify spectral regions which can distinguish the five treatment levels, as shown in figures 3 to 5.

2.3.2 Random Forest Classification: is a bagging method that employs recursive partitioning to divide the data into many homogeneous subsets called trees (Abdel-Rahman *et al.*, 2013). Each tree is independently grown to its maximum size based on a bootstrap sample from the training data set without pruning. In each tree, the model randomly selects a subset of variables to determine the split at each node (Abdel-Rahman *et al.*, 2013). The 'randomForest' package in the R software for statistical analysis was utilised. However, a leave-one-out cross validation method was applied instead of out-of-bag partitioning, owing to the small dataset used in the experiment. Top-20 important bands were selected, and the model was re-run using the selected bands to improve classification accuracy and subsequently assessed by confusion matrix and overall statistics.

2.3.3 Gradient Boosted Classification: is a supervised method and it assumes availability of a set of training samples (Nowakowski, 2015). It has two phases of processing: training and testing. The common approach to the training stage of boosting methods is to build a strong classifier from iteratively selected weak classifiers. In each iteration, every weak classifier is evaluated on weighted training data and a classification error

is provided (Nowakowski, 2015). The weak classifier which produces the smallest error is added to the resulting strong classifier with computed weight. The performances of the two classifiers were compared based on accuracy and stability. The ‘gbm’ package in the R software for statistical analysis was utilised. A leave-one-out cross validation method (available in caret package) was applied instead of partitioning the dataset. Post classification, top-20 important bands were selected and the model re-run only with selected bands to improve classification accuracy. Classification accuracy was assessed by confusion matrix as well as the Kappa coefficient and confidence interval.

3. RESULTS

A single-factor analysis of variance (ANOVA) was performed with 95% confidence interval. Overall, the difference between treatment levels was insignificant ($p = 0.628869$). Significant differences were found only at the two identified regions which could clearly discriminate treatment levels. Region A ($p = 3.96E-14$) had highly significant differences compared to Region B ($p = 0.009206$). Figures 3 shows the spectral profile of the 5 treatment levels, with visible discrimination of treatment levels shown in figures 4 and 5. This analysis was followed by a comparison of two classifiers: random forest and gradient boosted classification.

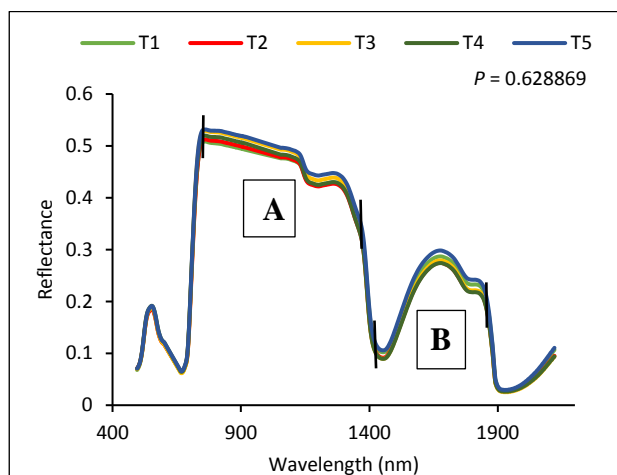


Figure 3. Reflectance Curves of Different Treatment Levels

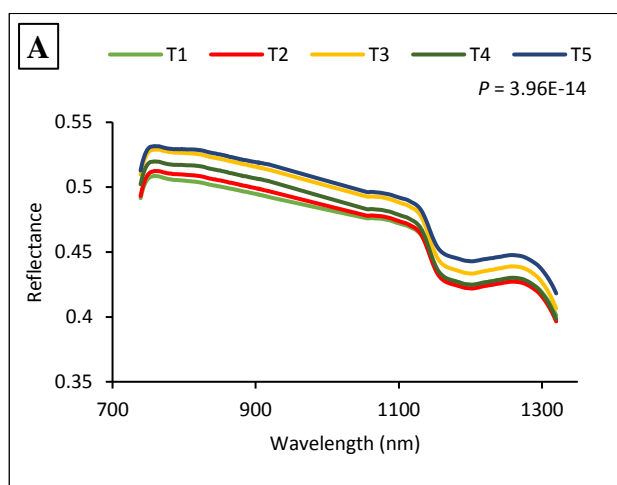


Figure 4. Region A Discriminating Treatment Levels

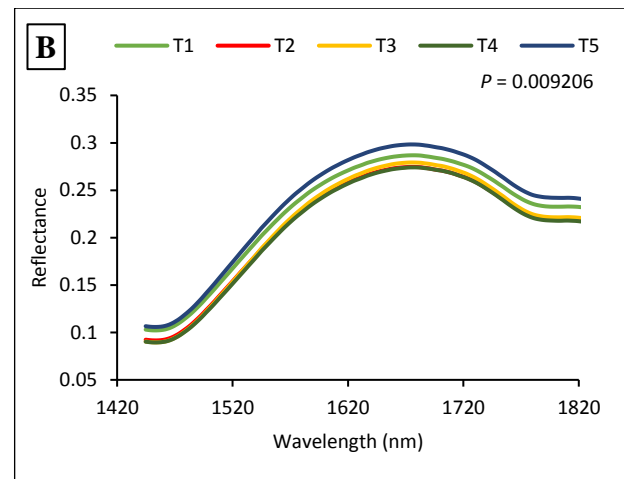


Figure 5. Region B Discriminating Treatment Levels

3.1 Gradient Boosted Classification

The GBM model produced a satisfactory classification accuracy of 70% (Kappa = 62.5%; 95% confidence interval (CI) = 0.5679, 0.8115) using the full spectrum. Figure 6 shows a matrix of the classification. The class T₁ had the lowest accuracy (25%), while T₅ produced the highest (75%). The T₁ and T₅ classes were misclassified more with each other than other classes. This implies that the control group (no fertiliser added) had similar reflectance to the group with the highest fertiliser concentration. If the two groups are not distinguishable, it could be deduced that T₅ is equivalent to not applying fertiliser to the crop.

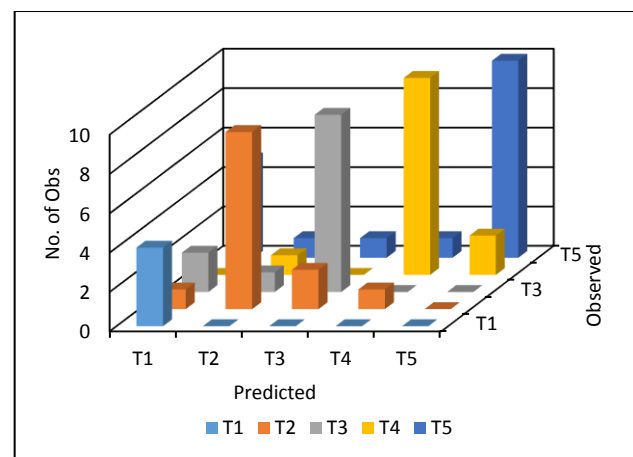


Figure 6. Classification Matrix for GBM Model using Full Spectra

Top 20 ranked variables identified by the GBM model are shown in figure 7. These bands are distributed across the range of the spectrum. The wavelength with the highest importance (10.87%) lies in the infrared region (1895.4 nm), but outside of the red-edge inflection point (690 – 730 nm). This can be attributed to the previously reported importance of the infrared region in vegetation studies.

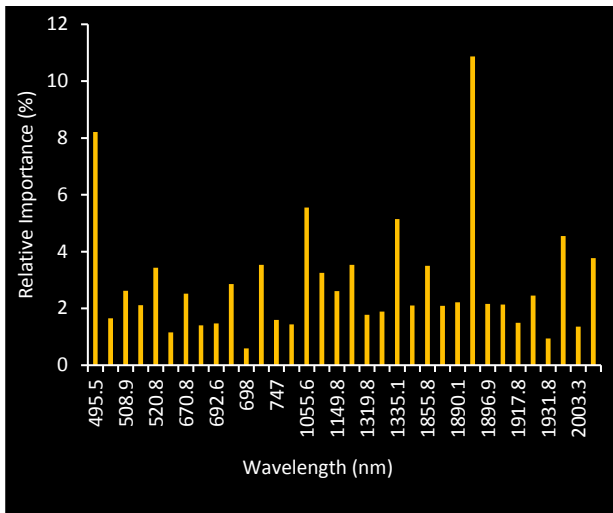


Figure 7. Important Variables Identified by GBM Model using Full Spectra

Classification accuracy of the GBM model improved with the use of the important variables, as seen in Figure 8. The overall accuracy improved to 88.33% (Kappa = 85.42%; 95% CI = 0.7743, 0.9518). The confidence interval is also excellent, considering the worst case would be 77% accurate. The class T₁ produced a 100% accuracy, followed by T₅ with 83% accuracy.

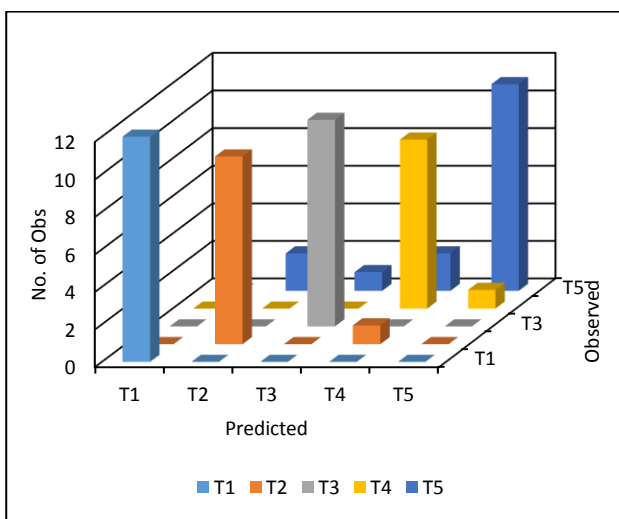


Figure 8. Classification Matrix for GBM using Important Variables

3.2 Random Forest Classification.

The RF model yielded a poor classification accuracy of 32% using the full spectra (Kappa = 15%). As seen in figure 9, only one class (T₅) had a classification accuracy above 50%. All the other classes produced similarly poor accuracies. Although random forest is favoured in many studies for its superiority, the model did not perform to the expected standards. One reason for this unexpected result could be the size of the sample used. Random forest was used with a small sample of only 60 leaves, making it 12 samples per treatment level.

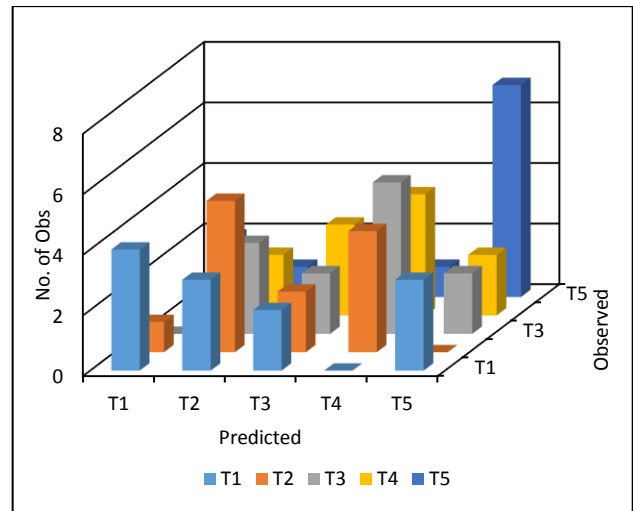


Figure 9. Classification Matrix for RF Model using Full Spectra

The RF model identified different important variables from those identified by the GBM model. No bands in the blue and green regions of the spectrum were identified as important, as shown in figure 10. The important variables are instead distributed across the red, near-infrared, infrared regions of the spectrum. The variable with the highest influence (100%) lies in the mid-infrared region (1959.5 nm).

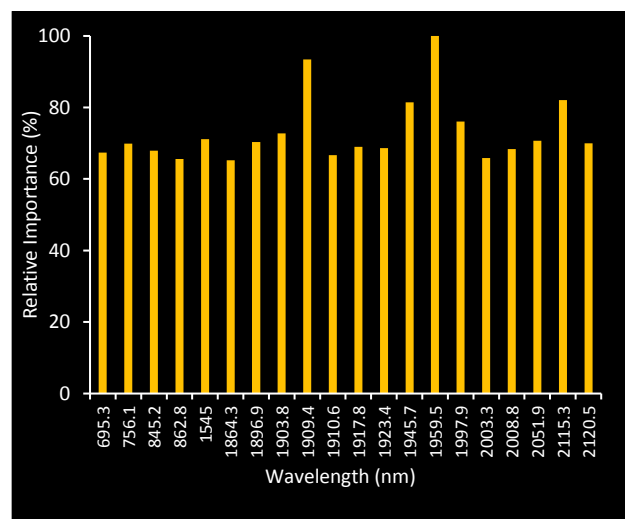


Figure 10. Important Variables Identified by RF Model using Full Spectra

Using the important variables, the accuracy of the RF model increased accuracy to 37% (Kappa = 21%). However, there was more variation in the class prediction accuracies, as shown in figure 11. The T₁ class was classified with the highest accuracy (41.7%). Again, T₁ was misclassified more with T₅ than other classes.

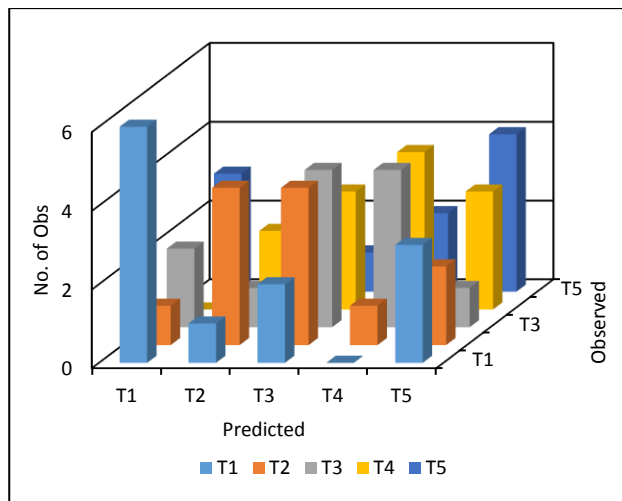


Figure 11. Classification Matrix for RF Model using Important Variables

4. CONCLUSION

It is evident that spectral reflectance data can discriminate slight differences between fertiliser treatment levels. However, the T₁ and T₅ classes were misclassified with one another more than with other classes. This was an unexpected yet interesting discovery, considering that T₁ was not treated with fertiliser (control) and T₅ was treated with the highest N: P: K input. Furthermore, gradient boosting model proved to have excellent predictive capability with significantly high accuracy levels (70% & 88%). The model identified about 20 important variables which extended across the entire range of the spectrum. The model's prediction accuracy improved by 18% when it was re-run using the important variables. On the other hand, random forest showed poor predictive capability with insignificant accuracy levels (32% & 37%). The important variables identified by RF model were not identical to those selected by GBM – they were distributed only between the red, mid-infrared and infrared regions of the spectrum. With the use of these important variables, classification accuracy of the RF model improved only marginally (5%). Unlike the GBM model, RF also does not provide the confidence interval (CI) in the results. Therefore, there is room for improvement of the RF classification model.

ACKNOWLEDGMENTS

A special thanks to the Department of Geography, Environmental Management and Energy Studies (University of Johannesburg) for providing the technical resources that supported this project. Gratitude is also extended to the National Research Foundation (NRF), for funding the project. The study was conducted in collaboration with the Vegetable and Ornamental Plants Institute (VOPI) of the Agricultural Research Council (ARC) of South Africa.

REFERENCES

Abdel-Rahman, E.M., Ahmed, F.B. & Ismail, R. 2013. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *International Journal of Remote Sensing*, 34(2):712-728.

Ahmed, A., Hue, N. & Radovich, T. 2014. Nitrogen release patterns of some locally made composts and their effects on the growth of Chinese cabbage (*brassica rapa*, *Chinensis* group) when used as soil amendments. *Compost Science & Utilization*, 22:199–206.

Alparone, L., Aiazzi, B., Baronti, S., & Garzelli, A. (2015). *Remote Sensing Image Fusion, Signal and Image Processing of Earth Observations*. New York: CRC Press.

Anawar, H.D., Garcia-Sanchez, A., Hossain, M.N. & Akter, S. (2012). Evaluation of health risk and Arsenic levels in vegetables sold in markets of Dhaka (Bangladesh) and Salamanca (Spain) by Hydride Generation Atomic Absorption Spectroscopy. *Bulletin of Environmental Contamination Toxicology*, 89:620-625.

Basso, B., Fiorentino, C., Cammarano, D. & Schulthess, U. (2016). Variable rate nitrogen fertilizer response in wheat using remote sensing. *Precision Agriculture*, 17:168–182.

Blignaut, J.N., De Wit, M.P., Knot, J., Midgley, S., Crookes, D.J., Drimie, S. & Nkambule, N.P. (2014). Sustainable agriculture: A viable option for enhanced food and nutritional security and a sustainable productive resource base in South Africa. An investigation. Baseline Review. Prepared for the Development Bank Southern Africa. Pretoria: ASSET Research.

Carroll, M.W., Glaser, J.A., Hunt, T.E. & Sappington, T.W. (2008). Use of spectral vegetation indices derived from airborne hyperspectral imagery for detection of European corn borer infestation in Iowa Corn Plots. *Journal of Economic Entomology*, 101.

Department of Government Communications and Information System (GCIS), Republic of South Africa. (2015). *South Africa Year Book*. [Online]. Available from: <http://www.gcis.gov.za/content/resourcecentre/sa-info/yearbook>. Accessed on 22 May 2016.

Eddy, P.R., Smith, A.M., Hill, B.D., Peddle, D.R., Coburn, C.A. & Blackshaw, R.E. (2013). Weed and crop discrimination using hyperspectral image data and reduced bandsets. *Canadian Journal of Remote Sensing*, 39(6):481-490.

Elmasry, G., Kamruzzaman, M., Sun, D-W. & Allen, P. (2012). Principles and Applications of Hyperspectral Imaging in Quality Evaluation of Agro-Food Products: A Review. *Critical Reviews in Food Science and Nutrition*, 52(11):999-1023.

Feng, W., Guo, B-B., Zhang, H-Y., He, L., Zhang, Y-S., Wang, Y-H., Zhu, Y-J. & Guo, T-C. (2015). Remote estimation of above ground nitrogen uptake during vegetative growth in winter wheat using hyperspectral red-edge ratio data. *Field Crops Research*, 180:197–206.

Goldblatt, A., Kotze, I., Corcoran, B., Botha, M., Brinkcate, T. & Hawkins, H-J. (2009). *Agriculture: Facts and Trends*. South Africa: The World-Wide Fund (WWF-SA).

Huang, C. & Asner, G.P. (2009). Applications of remote sensing to alien invasive plant studies. *Sensors*, 9:4869–4889.

Huang, W., Luo, J., Zhang, J., Zhao, J., Zhao, C., Wang, J., Yang, G., Huang, M., Huang, L. & Du, S. (2012). *Crop Disease and Pest Monitoring by Remote Sensing*. Chapter 2, *Remote Sensing - Applications*, Dr. Boris Escalante (Ed.). China: InTech.

Jensen, J.R. (2014). *Remote Sensing of the Environment: An Earth Resource Perspective*. 2nd Edition. Harlow, Essex: Pearson.

Kalisz, A., Kostrzewa, J., Sekara, A., Grabowska, A. & Cebula, S. (2012). Yield and nutritional quality of several non-heading Chinese cabbage (*Brassica Rapa* var. *Chinensis*) cultivars with different growing period and its modelling. *Korean Journal of Horticultural Science and Technology*, 30(6):650-656.

Li, C., Du, C., Ma, F. & Zhou, J. 2015. Diagnosis of nitrogen status in Chinese cabbage (*Brassica rapa* *Chinensis*) using the ratio of amide II to amide I in leaves based on mid-infrared photoacoustic spectroscopy. *Journal of Plant Nutrition and Soil Science*, 178:888-895.

Nowakowski, A. 2015. Remote Sensing Data Binary Classification Using Boosting with Simple Classifiers. *Acta Geophysica*, 63(5):1447-1462.

Ortenberg, F. (2009). *Hyperspectral Sensor Characteristics: Airborne, Space-borne, Hand-Held, and Truck-Mounted; Integration of Hyperspectral Data with LIDAR*. Chapter 2: Hyperspectral Remote Sensing of Vegetation. London: CRC Press.

Reddy, M.A. (2008). *Textbook of Remote Sensing and Geographical Information Systems*. Hyderabad: BS Publications.

Shapira, U., Herrmann, I., Karnieli, A. & Bonfil, D.J. (2013). Field spectroscopy for weed detection in wheat and chickpea fields. *International Journal of Remote Sensing*, 34(17):6094-6108.

Van Averbeke, W., Tshikalange, T.E. & Juma, K.A. (2007). The commodity systems of *Brassica rapa* L. subsp. *Chinensis* and *Solanum retroflexum* Dun. in Vhembe, Limpopo Province, South Africa. *Water SA*, 33(3):349-353.

Yao, H., Tang, L., Tian, L., Brown, R.L., Bhatnagar, D. & Cleveland, T.E. (2012). *Using Hyperspectral Data in Precision Farming Applications*. New York: CRC Press.