ABSTRACT

The increasing inability of domestic maize producers to meet the domestic consumption needs has been of major concern in Zimbabwe leading to the rise in maize imports. The purpose of this study was to come up with a model for predicting maize yield based on the correlation between yield and vegetation indices. Using vegetation indices derived from Landsat 8 satellite images, multi-collinearity and col-linearity models were developed to come up with the relationships between vegetation indices and crop yield from 2013 to 2017. The normalised difference vegetation index and soil adjusted vegetation index were found to have a high correlation to crop yield making it possible to estimate crop yield with up to 95% accuracy. This study definitively confirms the possibility of using Vegetation indices to predict crop yield, further studies are required to determine the requirements for national applications and implications of these estimates in national economic planning.

2 MATERIALS AND METHODS

The study area is UZ farm, formerly known as Thornpark Estate Farm which is located in Mashonaland central province, in Mazowe district (Figure 3.1). The farm falls into natural farming region 1 of Zimbabwe which receives mean annual rainfall of 815mm and a range of 750mm to 1000mm and is considered an area of high yield potential.

2.1METHODOGY

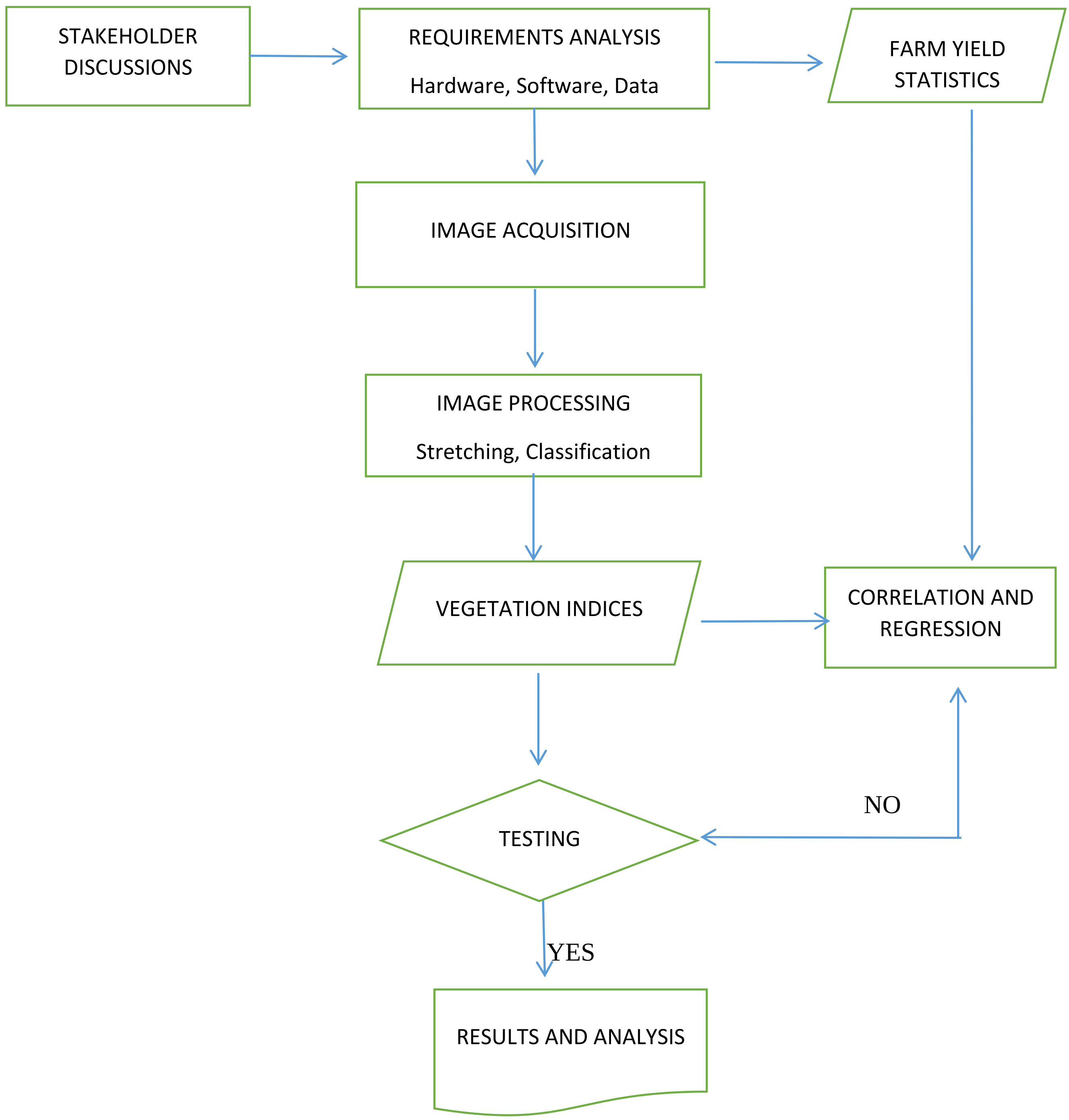

FIGURE 1: FLOWCHART OF METHODOLOGY

Understanding the needs of the users and how best to meet them requires the carrying out of a requirements analysis.The requirements analysis for this project was divided into two major parts, the first part being the determination of the functional requirements of the system. This is where all the needs of the various stakeholders were determined and brought into agreement with what is possible. This was done through informal discussions with officials from Agricultural, Technical and Extension Services(AGRITEX), the body responsible for monitoring of crops and generating crop statistics.The second part of the requirements analysis was to come up with the data, hardware and software requirements necessary to meet the functional requirements. This was done by studying other similar systems and reviewing similar projects both proposed and operational.

The calibrated approach was selected over the modelling approach because it allows for better results in a localized situation like the study area. For application in larger areas it is easier to use the modelling approach. Crop yield data for five years, i.e. 2012/2013, 2013/2014, 2014/2015, 2015/2016 farming seasons, was collected from the farm managers. These yield statistics were used in establishing the mathematical relationship between yield and calculated vegetation indices through regression models.The Landsat 8 images downloaded had a cloud cover of up to 30%. Landsat 8 images used were chosen because they were readily available for the entire period of study thus allowing for a uniformity of input data. The Landsat 8 images had a spatial resolution of 30m which is quite suitable for the study because the farm has a total area of 1 636 hectares.Each image was stretched in ILWIS software using the linear stretch in order to enhance the visual appearance of the images before classification was done. Supervised classification for the current season (date??) was done using training data collected from a visit to the farm. For the previous growing years google earth images from the archives were used. Image classification was done using the ILWIS software. The bands to be used for the calculation of the vegetation indices were merged in QGIS software using the *merge tool* to come up with a singe multi-band image. The Vegetation indices were calculated in QGIS using the raster calculator, Equations (1) to (3) show the formulas used to calculate the vegetation indices. The resulting vegetation indices were then exported to ILWIS for analysis.

With the VIs and the actual yield for the years the Excel data analysis tool was used to develop a multi-collinearity regression model for predictive analysis of crop yield. The multi-collinearity regression model assigns weights to the different indices based on their predictive power and based on their predictive power some indices may be removed from the model. The multi-collinearity regression model was selected because vegetation indices are derived quantities which are based on mostly the same measured quantities implying that they themselves are correlated, in addition to being related to the respondent variable (Franke, 2010). The equations were tested by plugging in variables calculated for the study area but from different years, i.e. years that were not used as part of the model.

3.RESULTS AND ANALYSIS

3.1 IMAGE PROCESSING

The Landsat 8 bands 6,5,4 were merged in QGIS and represented as a single pseudo colour composite. The image below shows a sample colour composite for the season 2014/2015 after clipping to the study area extent, the ones for the other seasons, along with the shape file of the study area that was used as a mask can be viewed in appendix c.

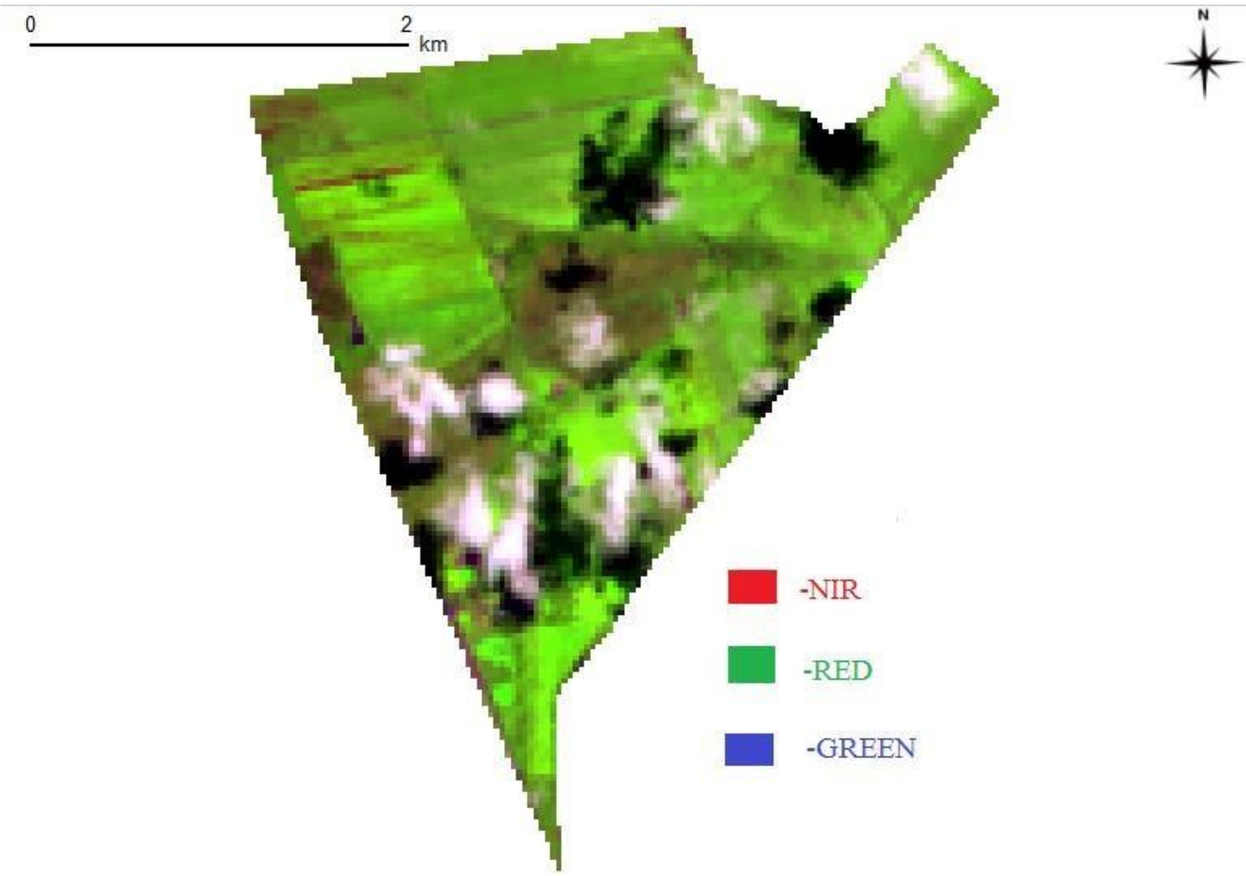

Figure 2: Extracted colour composite

The raster image of the study area was exported to ILWIS for classification several algorithms were tried but the *minimum distance mahalanobis* algorithm is the one that resulted in an output that best matched the crop distribution on the ground.Figure 4 below shows a sample classification result for the 2014/2015 season. According to the farm manager there is no crop rotation that has been practised at the farm during the chosen study period so the sample set used for the classification was the same for all images.

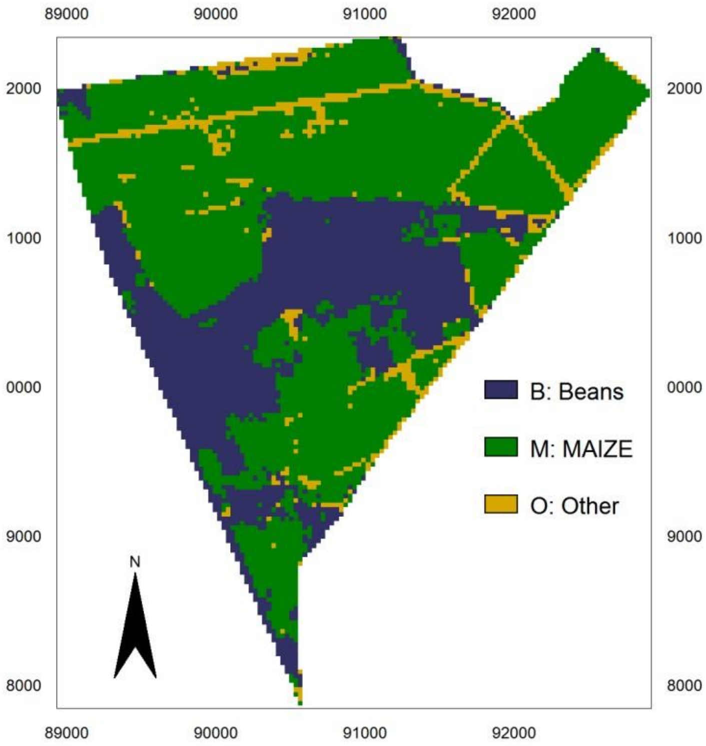

Figure 3: Crop types

After the classification was done, the vegetation indices were calculated for all the growing seasons, that is from 2012/2013 upto and including 2016/2017.The classification output was validated by a site visit.

3.2Normalised Difference Vegetation Index

The calculation of NDVI results in a raster image whose values range from -1 to +1. The negative values generally have no ecological value so they were taken as no data areas , they may be a result of clouds since the images being used had a cloud cover range between 0 to 30%. Higher NDVI values are associated with highly photo synthetically active vegetation. Low NDVI values represent a more even reflectance in both the red and NIR ranges, this corresponds to very little photosynthetic activity, or in the case of water just very little NIR reflectance.

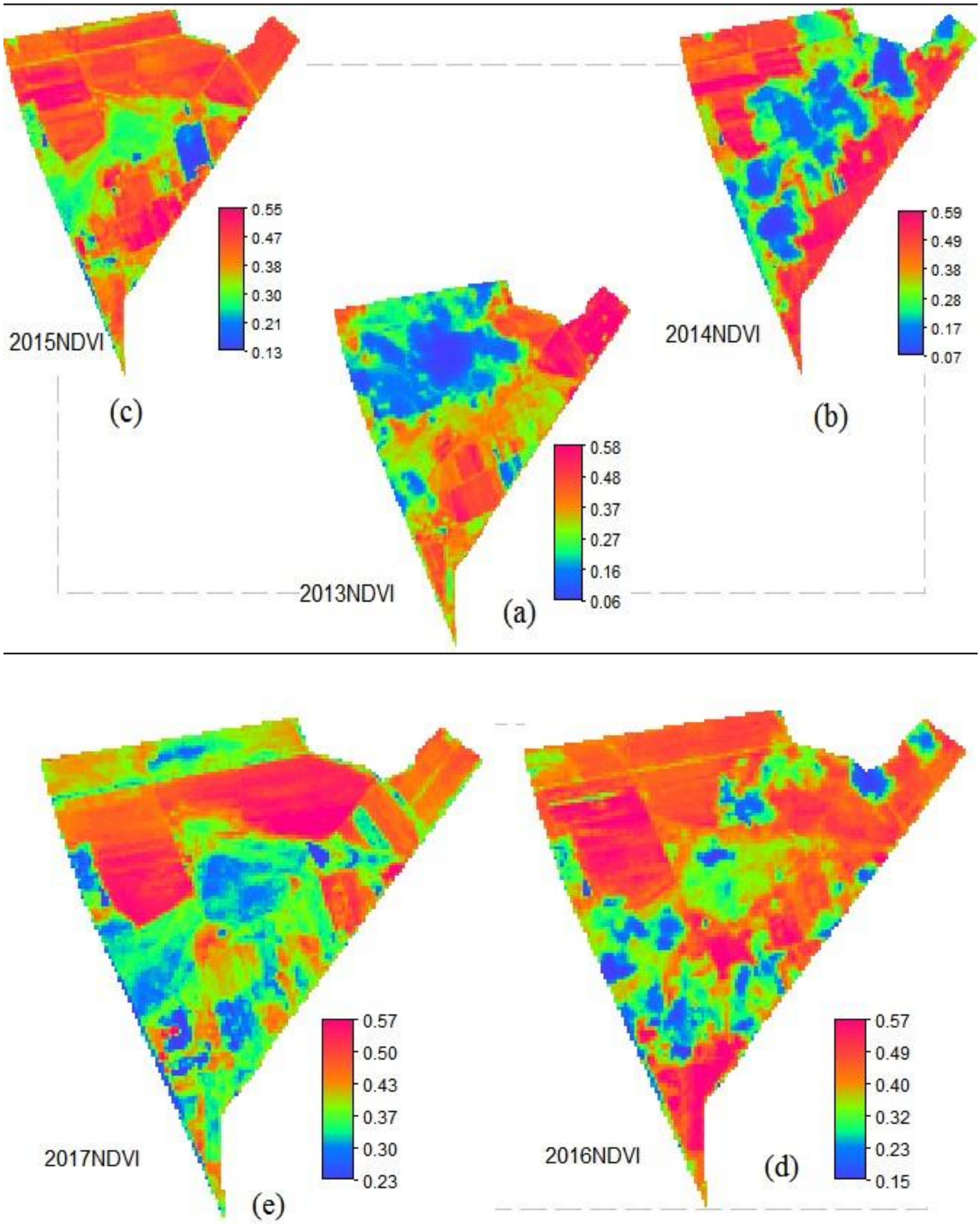

Figure 4: Images (a) to (e) show the NDVI maps for each of the 5 growing seasons under study

The figure 4 (a) to (e) shows the calculated NDVI maps from the 2012/2013 season to the 2016/2017 season. From these images the maximum NDVI values ranged from0.55 to 0.59, the minimum NDVI value for the areas classified as maize was 0.37. The lowest average NDVI for the maize pixels was 0.51. The standard deviation ranged from 0.09 to 0.1. These values of NDVI are consistent with the findings by Soris-Ruiz et al (2004) the averages of his study were 0.56. NDVI values in this range indicate moderate to high yield potential.These results indicate that at the VT stage there was a moderate to dense distribution of photosynthetic plants. These NDVI values indicate a high probability of a good yield.

The same process was followed for the calculation of RVI and SAVI maps and the results were as expected in agreement with the NDVI calculations

4 YIELD PREDICTION USING NDVI, SAVI and RVI MODELS

The linear regression equations were developed in order to come up with a predictive model for crop yield. The correlation coefficient between crop yield and NDVI was 0.802 which is high and shows that there is a very strong relationship between NDVI values and crop yield. The resulting regression equation was

, $y = 5421.2x - 1648.1$………….*equation 4.1*

Where   y = crop yield

X=NDVI

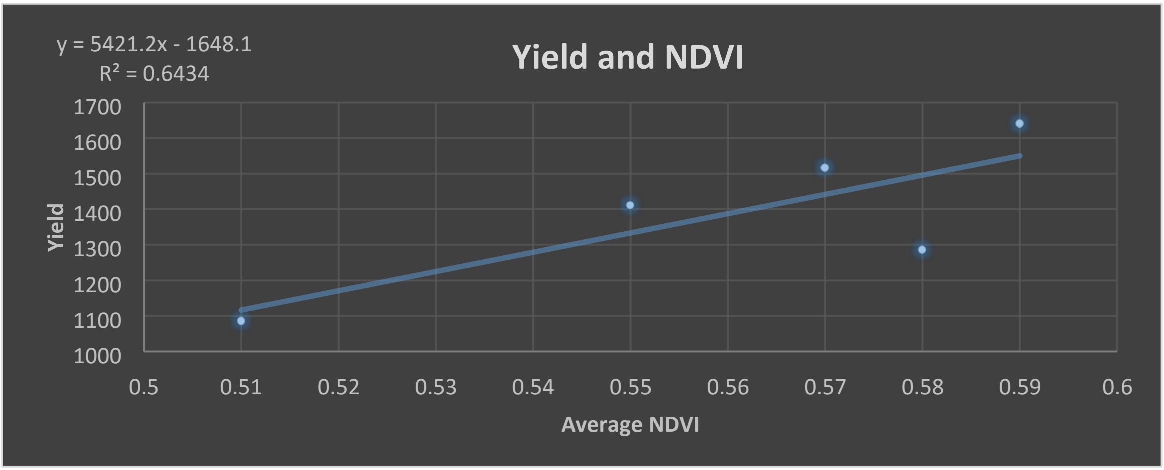

Figure 5: Graph of yield against NDVI

The correlation between yield with SAVI and NDVI was done using the same mathematical model and similar graphs were generated for the two models.

4.2YIELD PREDICTION USING MULTI-COLLINEARITY REGRESSION

Several multi-collinearity regression models were performed to relate Yield with NDVI, SAVI, and RVI together and in pairs. The various results showed that the multi-collinearity model of yield against SAVI and NDVI was the most statistically significant, with an F value well below the 0.15 threshold. It also explained over 90% of the variation in Yield.

Table 1: Summary Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.999457017
R Square	0.908914329
Adjusted R Square	0.996742986
Standard Error	13.56796795
Observations	4

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	169378.9482	84689.47411	460.0444736	0.0032949529
Residual	1	184.0897543	184.0897543		
Total	3	169563.038			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-2056.286966	187.2039002	-10.98421007	0.057798389	-4434.93805	322.3641179
0.74	-1436.773504	524.7360508	-2.738088039	0.022923569	-8104.177199	5230.63019
0.58	8303.376068	552.1317653	15.03875812	0.042269712	1287.876817	15318.87532

$y = -2056.29 + (-1436.77)x + 8303.38z$ *…..equation 4.3*

Where y=yield

X=SAVI

Z=NDVI

4.3 Testing of equations

Table 2: Summary of results

Vegetation Index	Season	Value	Predicted Yield	Actual Yield	% Accuracy
NDVI	2007/2008	0.75	2316.1	2417.8	95
NDVI	2011/2012	0.64	1722.4	1821.468	96
SAVI	2007/2008	0.94	2316.1	2105.906	91
SAVI	2011/2012	0.84	1722.4	1651.416	96

The positive relationships between vegetation indices and crop yield are also in agreement with the findings from similar studies done in 2004 in Hungary and Mexico. The drop in accuracy of the SAVI model as the values go higher is because the SAVI loses sensitivity to changes in the state of vegetation in the higher values. The average percentage discrepancy within these findings is 5.5 % which is in agreement with the findings of a study by Ferenz et al (2004) in Hungary the average for that study was 8%. Another similar study by Soria-Ruiz et al(2004) in central Mexico resulted in an average overestimation of 9.6%.The multi-collinearity model was also tested using the same data. The average for the multi-collinearity model dropped to 6.5 , this is a result of combining the errors in the separate vegetation indices.

Table 3: summary of results

Season	NDVI	SAVI	Actual yield	Predicted yield	%Accuracy
2007/2008	0.75	0.94	2316.1	2105.906	91
2011/2012	0.64	0.84	1722.4	1651.416	96

The gap between the predicted and expected yields can be attributed to loss of yield due to adverse conditions that may have occurred after the prediction like excess or lack of rainfall, loss of yield under soil salinity. Also there are errors inherent in the vegetation indices themselves that affect their representation of true conditions, for example sensor errors.

Table 4: Summary of l Vegetation indices for all the years under study

Season	NDVI	RVI	SAVI

2012/2013	0.58	0.37	0.74
2013/2014	0.59	0.30	0.84
2014/2015	0.57	0.34	0.80
2015/2016	0.51	0.39	0.76
2016/2017	0.55	0.35	0.77

5.CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

Optical remote sensing techniques in particular are well suited for agricultural applications, because the techniques are able to provide information on the actual status of crops at different growth stages via their spectral signatures.This study resulted in the following findings.

a) Vegetation Indices derived from the spectral reflectance of green vegetation can be used for the classification of different crop types on a farm.

b) The relationship between vegetation indices and crop yield can be represented by a mathematical model.

c) Linear regression models were derived and tested, the result was that the models were generally accurate upto 95% but that the models become less sensitive to changes in crop condition the higher they go so the accuraccy dropped to about 90% for very high values of the Vegetation Indices

d) Conditions within regions like soil reflectance and soil salinity affect the accuracy of the model and therefore should be accounted when choosing which predictive model to use , for example the SAVI derived model in more suitable for areas with highly reflective soils.

e) Several multi-collinearity models were derived by combining various vegetation indices according to their predictive powers in an attempt to smoothen the predictive equation but the resultant accuracy dropped due to the effects of combining the errors from the different indices.

5.2 Recommendations

a. Due to the cloudy conditions that are common in the rain season, the use of several satellite that image the areas at different times will increase the chances of obtaining cloud free images but if that fails softwares such as Fmask can be used as an alternative.

b. While use of models based only on vegetation indices which assume similar climatic conditions over regions will give acceptable results, combining these models with climate data and soil salinity data would improve the accuracy of the model.

c. Coming up with these regional predictive models for the whole of Zimbabwe would make it possible to estimate with an acceptable accuracy the expected production and put measures in place to deal with shortages or surplus.

d. Automating the system would also make it easier for the laymen to be able to predict yield without needing to rely on the government produced statistics.

REFERENCES

Agriculture, M. O. F. (2008) 'Ministry of Agriculture, Zimbabwe: SECOND ROUND CROP AND LIVESTOCK', (April).

Anseeuw, W.,Kapuya, T. & Saruchera, D., 2012. Zimbabwe's agricultural reconstruction: Present Stage, ongoing projects and prospects for reinvestment.

Casley, D.J. & Kumar, K. 1988. The collection, analysis and use of monitoring and evaluation data. Johns Hopkins University Press: Baltimore, Maryland, USA

Carletto, G., Gourlay, S. & Winters, P. 2015. From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis. Journal of African Economies, 24(5), pp. 593-628.

Carpenter, C. G. et al. 1942. Production and Yield: The Agriculture estimating and reporting services of the United States : U.S Government printing office

Chipanshi, A. *et al.* (2015) 'Agricultural and Forest Meteorology Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape', *Agricultural and Forest Meteorology*. Elsevier B.V., 206, pp. 137–150. doi: 10.1016/j.agrformet.2015.03.007.

Cross, E. 2015.The importance of agriculture to Zimbabwe.The Zimbabwe independent.

Domenikiotis C, Spiliotopoulos M, Tsiros E, Dalezios NR (2004) Early cotton yield assessment by the use of NOAA/AVHRR derived Vegetation Condition Index (VCI) in Greece. Int J Remote Sensing(14),pp.2807–2819

Dorward, A. and Chirwa, E. (2010) 'A review of methods for estimating yield and production impacts A review of methods for estimating yield and production impacts', (December).

Flowerday, D. 1995.Corn growth stages with estimated calendar days and growing-degree units

Ferencz, C. *et al.* (2004) 'Crop yield estimation by satellite remote sensing', *International Journal of Remote Sensing*. Taylor & Francis Group , 25(20), pp. 4113–4149. doi: 10.1080/01431160410001698870.

Ferencz, C. *et al.* (2017) 'Crop yield estimation by satellite remote sensing', 1161(September). doi: 10.1080/01431160410001698870.

Franke, G. R. (2010) 'Multicollinearity', in *Wiley International Encyclopedia of*

Marketing. Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/9781444316568.wiem02066.

Go-, T. R. S. (2017) 'Methodology for Estimation of Crop Area and Crop Yield under Mixed and Continuous Cropping', (March).

Greaves, G. E. and Wang, Y. (2017) 'Yield response , water productivity , and seasonal water production functions for maize under deficit irrigation water management in southern Taiwan', *Plant Production Science*. Taylor & Francis, 1008, pp. 1–13. doi: 10.1080/1343943X.2017.1365613.

Govaerts, B.& Verhulst, N., 2010. The Normalised Difference Vegetation Index (NDVI) GreenSeeker TM Handheld Sensor: Toward Intergrated Evaluation of Crop management.p12

Hercules, A., 2011. Computational Methods for Agricultural Research::Advances and Applications. New York : Hershy

Henry, C & Taylor, A.D. 1952. The story of Agricultural economics in the United States: The Iowa State college press.

History, B. (no date) 'Timely and Accurate Crop Yield Forecasting and Estimation History and Initial Gap Analysis '.

Hoffman, S. (no date) 'U.S. Crop Production Forecasting & Estimation Methodology'.

Ji-hua, M.et al., 1999. Study on crop condition monitoring methods. *Archives*, pp.945-950.

Literature, S. O. F. (2016) 'SYNTHESIS OF LITERATURE AND Research on Improving Methods for Estimating Crop Area , Yield and Production under Mixed , Repeated and Continuous Cropping Global Strategy Working Papers', (5).

Manatsa, D. *et al.* (2011) 'Maize yield forecasting for Zimbabwe farming sectors using satellite rainfall estimates', *Natural Hazards*. Springer Netherlands, 59(1), pp. 447–463. doi: 10.1007/s11069-011-9765-0.

Makuvaro, V. 2014.Impact of climate change on smallholder farming in Zimbabwe, using a modelling approach

McNairn, I., 2014.Zimbabwe food security Brief.,(2), pp.1-4

Norman, D.W., Worman, F.D., Siebert, J.D. & Modiakgotla, E. 1995. The Farming Systems Approach to Development and Appropriate Technology Generation. FAO Publication: Rome.

Nain. A.S, Dadhwal VK, Singh TP (2002) Real time wheat yields assessment using technology trend and crop simulation model with minimal data set. Current Science (82), pp.1255–1258

Nain A.S, Dadhwal VK, Singh TP (2004) Use of CERES-Wheat model for wheat yield forecast in central Indo-Gangetic Plains of India. J Agricultural Science (142), pp.59–70

Ouhbi, S.2010,Software Development Process – activities and steps,pp 1-42.

Pinter, P.J. et al., 2003. Remote Sensing for Crop Management. Photogrammetric Engineering & Remote Sensing, 69(6),pp.647-664

Practices, G. I. S. B. (2009) 'Gis-for-agriculture-ESRI2020.pdf', *Www.Esri.Com*, (June), pp. 1–30.

Poate, C.D. 1988. A review of methods for measuring crop production from smallholder producers. Experimental Agriculture, 24(1).pp1-14.

Rukuni, M. & Eicher C.K. 1994., Zimbabwe's agricultural revolution

Satir, O. and Berberoglu, S. (2016) 'Field Crops Research Crop yield prediction under soil salinity using satellite derived vegetation indices', *Field Crops Research*. Elsevier B.V., 192, pp. 134–143. doi: 10.1016/j.fcr.2016.04.028.

Setiyono, T., Nelson, A. and Holecz, F. (no date) 'Remote Sensing based Crop Yield Monitoring and Forecasting'.

Shanahan, J. F., Schepers, J. S. and Francis, D. D. (2001) 'Use of Remote-Sensing Imagery to Estimate Corn Grain Yield'.

Soria-ruiz, J., Fernández-ordóñez, Y. and Granados-ramírez, R. (2004) 'Methodology for prediction of corn yield using remote sensing satellite data in Central Mexico

Metodología para la estimación del rendimiento de maíz utilizando datos de sensores remotos satelitales en el Centro de México', pp. 61–78.

Study, A. C., Birkoor, O. F. and Sawasawa, H. L. A. (2003) 'Crop Yield Estimation : Integrating RS , GIS and Management Factors', (March).

Tadross, M. A., Hewitson, B. C. and Usman, M. T. (2005) 'The Interannual Variability of the Onset of the Maize Growing Season over South Africa and Zimbabwe', *Journal of Climate*, 18(16), pp. 3356–3372. doi: 10.1175/JCLI3423.1.

Singh, R. (1951) 'CROP YIELD ESTIMATION AND FORECASTING USING'.

Silleos,N.G. et al., 2006. Vegetation Indices: Advances made in Biomas Estimation and Vegetation Monitoring in the last 30 years.Geocarto International, 21(4),pp21-28

Vorovencii, I. (2014) 'ASSESSMENT OF NDVI FOR DIFFERENT LAND COVERS BEFORE AND AFTER', 7(1).

Wang, R., Cherkauer, K. and Bowling, L. (2016) 'Corn Response to Climate Stress Detected with Satellite-Based NDVI Time Series'. doi: 10.3390/rs8040269.

Yaakup, A. (2004) 'GIS as a Tool for Development Planning and Monitoring', *ePrints*, pp. 1–23.

ZimAsset, Command Agric inseparable | The Sunday Mail (no date). Available at: http://www.sundaymail.co.zw/zimasset-command-agric-inseparable/ (Accessed: 2 October 2017).

Annalie Fourie, 2009. *ESRI GIS.* [Online]
Available at: http://www.esri.com/industries/agriculture [Accessed 30 August 2017].

Jessica Wyland, 2009. *ESRI GIS.* [Online]
Available at: http://www.esri.com/industries/agriculture [Accessed 7 september 2017].

Pindula local knowledge, 2016. *Pindula.* [Online]
Available at: http://www.pindula.co.zw/Command_Agriculture_2016 [Accessed 30 August 2017].

Scoones, I., 2013. *zimbabweland.* [Online]
Available at:
https://zimbabweland.wordpress.com/2013/09/23/zimbabwes-agricultural -sector-goes-from-bread-basket-to-basket-case-or-is-it-again-a-bit-more-complicated/
[Accessed 7 sept 2017].

U.S.A Department of agriculture, 2017. *index mundi.* [Online]
Available at:
https://www.indexmundi.com/agriculture/?country=zw&commodity=corn&graph=domestic-consumption [Accessed 5 september 2017].