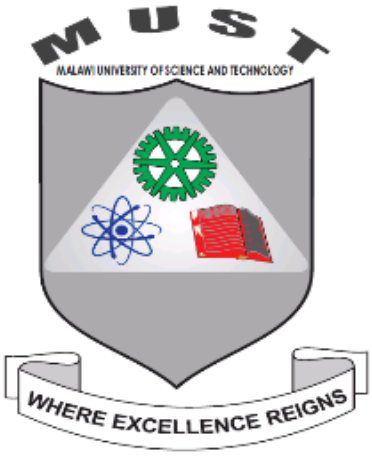


FORECASTING MAIZE YIELD USING REMOTE SENSING

CHRISTOS LIVILA
Thyolo, Malawi

Dissertation submitted to Ndata School of Climate and Earth Sciences of the Malawi University of Science and Technology in partial fulfilment of the requirements for the Bachelor of Science degree in Earth Sciences

July 2021

DISSERTATION APPROVED BY:

SUPERVISOR:
Mrs. E.G. Chimimba Signature: _____

ASSESSOR:
Title, Initials, Name Signature: _____

DISCLAIMER

This document describes work undertaken as part of a program of study at Ndata School of

Climate and Earth Sciences of the Malawi University of Science and Technology. All views and opinions

COPYRIGHT STATEMENT:

This copy of the dissertation has been supplied on condition that anyone who consults it is understood to recognize that its copyright rests with its author and that no quotation from the dissertation and no information derived from it may be published without acknowledgement.

ABSTRACT

The study aimed at evaluating the relationship between NDVI and maize yield at Makande prison farm in Thyolo district. The data used in the study were remotely sensed images, and crop yield statistics of years 2019 to 2021. Images used were obtained using the Sentinel-2 satellite constellation, which is part of the Copernicus program, and were acquired using QGIS and were processed. NDVI was used with an empirical regression model to estimate maize yield. The predicted yield was correlated with the actual yield using linear regression. A correlation coefficient of 0.99, R^2= 98.8%, and root mean error square and mean absolute error of 0.1779 and 0.119 in that order were established. The findings illustrated a very strong correlation between NDVI and maize yield. The study recommends that the use of more satellite images and yield data for not less than 6 years will help to increase the prediction accuracy.

Keywords: *NDVI, regression analysis, Sentinel 2*

ACKNOWLEDGEMENTS

I would like the express my gratitude to my supervisor Mrs. E.G. Chimimba for her guidance and positive criticism through the research work. Appreciation should also go to: my parents who financially assisted in the funding of the project; Mr. H. Sawasawa, Mr. D. Mbeya, and Mr. S Jha for further guidance and consultations during the study; Mr. Chigalu from Makande prison for his assistance during the data collection; and friends for their warm academic company and encouragement.

I sincerely give the Almighty God all glory for granting me divine health, strength, and determination to undertake this study.

TABLE OF CONTENTS

1. INTRODUCTION 1
   1.1. Background 1
   1.2. Problem Statement 4
   1.3. Objectives 5
   1.4. Research question 5
   1.5. Scientific significance 5
2. LITERATURE REVIEW 6
   2.1. Background of the use of NDVI in yield forecasting 6
   2.2. Factors that affect crop yield 7
3. METHODOLOGY 9

3.1. Research design 9
3.2. Study area 9
3.3. Data Collection 10
3.4. Data analysis 11
3.5. Accuracy assessment 13
4. RESULTS AND DISCUSSION 14
4.1. Crop yield data 14
4.2. NDVI images 14
4.3. Estimated yield 15
4.4. Comparison of estimated and actual yields 15
5. CONCLUSION AND RECOMMENDATIONS 19
6. REFERENCES 20

ACRONYMS/ABBREVIATIONS

GDP: Gross Domestic Product

GNDVI: Green Normalized Difference Vegetation Index

MAE: Mean Absolute Error

NDVI: Normalized Vegetation Index

NDWI: Normalized Difference Water Index

NIR: Near-Infrared

OSAVI: Optimized Soil Adjacent Vegetation Index

SCP: Semi-Automatic Classification Plugin

SWIR: Short-wave Infrared

QGIS: Quantum Geographic Information System

RMSE: Root Mean Square Error

VCI: Vegetation Condition Index

VI: Vegetation Index

LIST OF FIGURES

Figure 1: Flow diagram of the methodology 9
Figure 2: Map of the study area, Makande prison farm 10
Figure 3: NDVI images for 2019, 2020, and 2021 15
Figure 4: Linear correlation between estimated yields and actual yield 16

LIST OF TABLES

Table 1: Mean and standard deviation of NDVI 12
Table 2: Actual annual yield 14
Table 3: Estimated maize yield 15
Table 4: Estimated yields and actual yields 16
Table 5: Results from a regression analysis 17

1. INTRODUCTION

World's population is increasing rapidly and at the same time, essential natural resources, such as land and water, are declining in both quantity and quality due to such factors as competition with industrial and urban demands. The food for the growing world population is largely dependent on agriculture and its production. Ezeh, Bongaarts, & Mberu, (2012), reported that agricultural activities constitute a major part of the African economies, absorbing 70% of the workforce, and account for 30% of GDP. As such, increasing food production becomes the focus of research in most developing countries. Maize planting area and yield information are crucial in Africa as maize plays a central role in fulfilling the staple food requirements since it accounts for more than 50% of the energy requirements in the continent (Dorward & Chirwa, 2011).

To satisfy the ever-increasing food demand due to the forecasted doubling of Africa's population by 2050 (Ezeh, Bongaarts, & Mberu, 2012), a better mechanism that supports the monitoring and forecasting of food security is urgently needed. National trade, agricultural policies, and timely response to food shortages rely on accurate and timely maize production data (Chivasa, Mutanga, & Biradar, 2017).

1.1. Background

According to Reynolds et al., (2000), the conventional method of maize crop yield estimation would lead to poor crop yield assessment and crop area estimation, which generally involves data collection for crop and yield estimation based on ground-based field visits and reports that are often found to be subjective, time-consuming and errors due to incomplete ground observation. Burke & Lobell (2007) added that the accuracy and precision of such methods are affected by many factors such as; expert opinion, quality of sampling kit used, sampling methods, sampled locations, road access, data analysis methods used, and self-reported data where farmers tend to over-report yield and crop area. Furthermore, results from field-based surveys are difficult to scale to large areas.

The commonly used ground-based survey methods have quality and quantity-related problems in the crop yield information generated (Chivasa, Mutanga, & Biradar, 2017). For example, during the 2008/09 season in Malawi, in the cropland area, there was an overestimation by 30% using the conventional ground-based survey method (Dorward & Chirwa, 2011). Jayne and Rashid (2010) reported how surplus estimates in maize by the Malawi government caused maize prices to reach record levels when the estimated surplus

was not achieved. Rukuni, et al. (2006), reported that the cause of the late importation of maize by the Zi

The application of remote sensing and satellite imagery technologies in crop production estimation have proven to provide quick, reliable, and timely (and potentially real-time) and objective data on crop growth at relatively minimum cost (Chivasa, Mutanga, & Biradar, 2017; Adeniyi, Szabó, Tamás, & Nagy, 2020). Moreover, it provides an effective tool for measuring yield-relevant crop parameters, e.g. based on the leaf area indices and normalized difference vegetation index (Karst, et al., 2020). The interest in using satellite remotely sensed data today for crop monitoring and crop production forecasting has tremendously increased due to its potential to produce data synoptically, with more spatial coverage, potentially at the global scale (Adeniyi, Szabó, Tamás, & Nagy, 2020). Remotely sensed data can be used as a stand-alone methodology in estimating crop production yields or integrated with the household sample survey to develop a more robust methodology for estimating crop production yield. Remote sensing and imagery have also uncovered an opportunity for early yield estimation for crops, more precise acreage estimation, and generation of crop production maps which can be used for the strategic establishment of grain storage facilities in the country (Mwanaleza, 2016; Karst, et al., 2020).

1.1.1. Vegetation indices used in remote sensing

Multi-spectral remotely sensed data, in particular different derived vegetation indices (VIs), can represent biomass and crop conditions. Many indices have been developed by the use of remote sensing data, and some of the indices used are normalized difference vegetation index (NDVI) and vegetation condition index (VCI) (Lv, 2013); normalized difference water index (NDWI), and normalized difference red edge index (NDRE) (Karst, et al., 2020); optimized soil adjacent vegetation index (OSAVI) and green normalized difference vegetation index (GNDVI) (Iskandar, 2020).

The NDVI stands out among the vegetation indices and is regarded as an all-purpose index. It is simple to calculate, has the best dynamic range, and has the best sensitivity to changes in vegetation health and cover (Shiba, Mabaso, Dlamini, & Singwane, 2020). Green healthy vegetation reflects light in the near-infrared portion of the spectrum and absorbs red light. NDVI ranges from -1 to 1, where positive values indicate green vegetation and are derived from red and near-infrared bands of remotely sensed images (Rouse, Haas, Schell, & Deering, 1973).

1.1.2. Correlating NDVI and crop yield

NDVI data has been used extensively in vegetation monitoring, crop yield assessment, and forecasting (Hayes, O'Rourke, Terjung, & Todhunter, 1982; Benedetti & Rossinni, 1993; Johnson, 2014). The NDVI was initially proposed by Rouse et al. (1973). It is closely correlated with green biomass and leaf area, and this is one of the most widely used indices for agriculture monitoring. Lykhovyd, (2020) reported the successful use of the NDVI data for operational monitoring of the conditions of such crops as wheat, cotton, rice, and corn. He also claimed quite a high reliability and accuracy of a simple linear regression model of "NDVI-yield" inter-relationship for every studied crop and suggested that NDVI is very good for crop productivity forecasting. The corn production estimations based on the NDVI based model with the coefficient of regression R = 0.75 were performed by Lewis et al. (1998), who claimed that the NDVI-based simple regression models are very encouraging as reported by (Lykhovyd, 2020).

The concept behind NDVI explains that plants' chlorophyll absorbs sunlight, which is captured by the red light region of the electromagnetic spectrum, whereas a plant's spongy mesophyll leaf structure creates considerable reflectance in the near-infrared region of the spectrum (Lv, 2013). For this reason, greener and dense vegetation has low red light reflectance and high near-infrared reflectance, and thus high NDVI values. On the other hand, near-zero and negative values of the index indicate non-vegetated surface features such as rock, soil, water, ice, and clouds (US Geological Survey, 2010).

In short, the index is sensitive to changes in plant biomass and leaf size, which varies for forest structural and crop types. Therefore, monthly NDVI values of crops, as well as the changes during the growing season, indirectly measures crop health and growing conditions. As the growing season progresses, from December to May, crops become greener and bigger, and that translates to an increasing monthly mean NDVI. Near the end of the growing season, the mean NDVI for crops is always greater and the increase of monthly NDVI varies depending on how healthy and dense the crops are.

1.2. Problem Statement

In most African countries, studies on crop yield estimation using NDVI have been conducted (Burke & Lobell, 2007; Rojas, 2007; Tychon et al., 2008; Chivasa et al., 2017; Lambert et al., 2018; Shiba et al., 2020). The results obtained from these studies have proved to be

successful and encouraging, providing crucial information to the government and other agricultural stakeh

Though this is the case, in Malawi, relatively few studies on the relationship between remotely sensed NDVI measurements and maize yield have been conducted (Mwanaleza, 2016). Many of these were done on a very large scale. Since they cover very large areas, low-resolution images were used resulting in generalization of the crop condition and yield estimates (Sawasawa, 2003). Studies conducted with the use of high spatial resolution NDVI multispectral images in predicting maize yield at a small scale are scanty. According to Staggenborg & Taylor, (2000), at a small scale agricultural production is a result of complex environmental stresses including farmer's management. These have a great effect on the final yield, which may not be detected with low-resolution satellite images. This study, therefore, proposed to estimate maize yield by applying remotely sensed NDVI measurements, and establish the relationship between estimated yields and actual yields at Makande prison farm using high-resolution Sentinel-2 images.

1.3. Objectives

1.3.1. Main Objective

- The primary objective of the study is to establish the relationship between remotely sensed NDVI measurements and maize field yields.

1.3.2. Specific Objectives

- To calculate maize yield estimate using an empirical regression model.
- To correlate estimated yields and field yields.

1.4. Research question

- What is the relationship between remote sensed NDVI and crop yield at the field level?

1.5. Scientific significance

This study has helped in the establishment of a significant relationship between remotely sensed NDVI images and crop yields at the field level. The outcome will help to state if the methodology can provide timely and accurate crop yield forecasting results. This provides crucial information for all agricultural stakeholders to make informed decisions and leads to

efficient resource allocation. The results will also be a resource to other people carrying out similar studie

2. LITERATURE REVIEW

2.1. Background of the use of NDVI in yield forecasting

NDVI tends to measure plant greenness, a direct measurement of chlorophyll content and photosynthetic activity, which is a basic but reliable way to gauge plant health and biomass, or yield (US Geological Survey, 2010; Léger, 2014). It is most effective beginning in the pre-peak season of crop growth and continuing through harvest. During the early stages of the growing season, when most crops are just starting to show shoots, NDVI tends to produce less useful information as it does not measure enough surface area (Léger, 2014).

According to Lv (2013), NDVI derived from remotely sensed images (1.1×1.1 km resolution), was first applied by Benedetti and Rossini (1993) in assessing and forecasting grain yield. They conducted a case study in Emilia Romagna in Italy with a simple linear regression model from 1986 to 1989. The scholars used integrated NDVI during the critical development stage of wheat as an independent variable and the predicted wheat yields had greater than 10% but less than 19% deviation from the official wheat data. Labus and his team (2002) studied NDVI (16×16 km resolution) during the growing season of wheat, and estimated wheat yield at regional and farm scales in Montana from 1989 to 1997. The study found a strong relationship between wheat yield and integrated NDVI throughout the whole growing season (R^2=75.3%) and at the grain filling stage (R^2=69.7%).

Importantly, both studies indicated that the early-season NDVI values alone were not consistent estimators for yields. This was an important finding because it suggested that biomass is not a determinant of final grain yield. The findings also suggested that a large biomass accumulation early in the season is not a determining factor of final grain yield. A study conducted by Benedetti & Rossinni (1993), concluded that early vegetation development before the grain filling stage contributes only to plant structure but not grain production. Pinter et al. (1981) suggested that the incorporation of VIs, like NDVI, performs better when taken at specific growth stages (e.g. at tussling or flowering stage). These VIs are measured by remote sensors ranging from handheld to satellite-based with a diverse spatial, temporal, and spectral resolution, which affects the final output.

2.2. Factors that affect crop yield

Huang et. al (2013), noted that increased yields in cereal are mainly the result of greater inputs of fertilizer, water and pesticides, new crop species, and the improvement of

management over the last decades. For all developing countries, modern varieties accounted for 21% of

Few studies to date (Sawasawa, 2003; Prasad, Chai, Singh, & Kafatos, 2006; Tychon, Balaghi, Eerens, & Jlibene, 2008; Lv, 2013) have combined NDVI variables with surface temperature, precipitation, management, and soil moisture in estimating crop yields. Prasad et al. (2006), examined crop yields for the state of Lowa from 1982 to 2001 for 19 years. An average of NDVI (8×8 km resolution) throughout the growing season was used as the input to the model, which produced R^2=.78 for the corn. They took into account most of the parameters, which control the crop yield. They concluded that the method can be used to predict crop yield for other crops as well as corn and soybeans. Based on data obtained before harvest, crop yield can be predicted with acceptable accuracy. They also added that studies can be done in other countries as well, where crop production is primarily dependent on weather and climatic conditions. Most importantly, their results demonstrated that remotely sensed data can be used to forecast crop yield with acceptable accuracy before harvesting (Chivasa, Mutanga, & Biradar, 2017).

Several studies in different parts of the world have examined the relationship between NDVI and maize yield (Battude, et al., 2016; Chivasa, Mutanga, & Biradar, 2017; Rojas, 2007; Tychon, Balaghi, Eerens, & Jlibene, 2008). The results from these studies showed that the use of satellite imagery to estimate maize yield offers possible and cost-effective results. The studies concluded that no single indicator is valuable always and everywhere, and its performance depends on many factors, e.g. climate, soils, management, crop type, growing season, or limitation of given device or sensor (Johnson, 2014; Chivasa, Mutanga, & Biradar, 2017; Héctor, et al., 2020).

On the contrary, studies in Malawi using remote sensing in agriculture are limited and have led not only to poor insights into the productivity and yield gaps of these systems but also missed opportunities to improve food security through early warning systems (Burke & Lobell, 2007). A limited effort has been made in estimating yield in staple food crops like maize. The scarcity of studies linking to maize yield in Malawi was the motivation for this study.

3. METHODOLOGY

3.1. Research design

The study involved the use of quantitative and qualitative data analyses. The quantitative data analysis was aimed at evaluating the relationship between the predicted yield and actual yield. The qualitative analysis focused on analyzing NDVI images for maize yield.

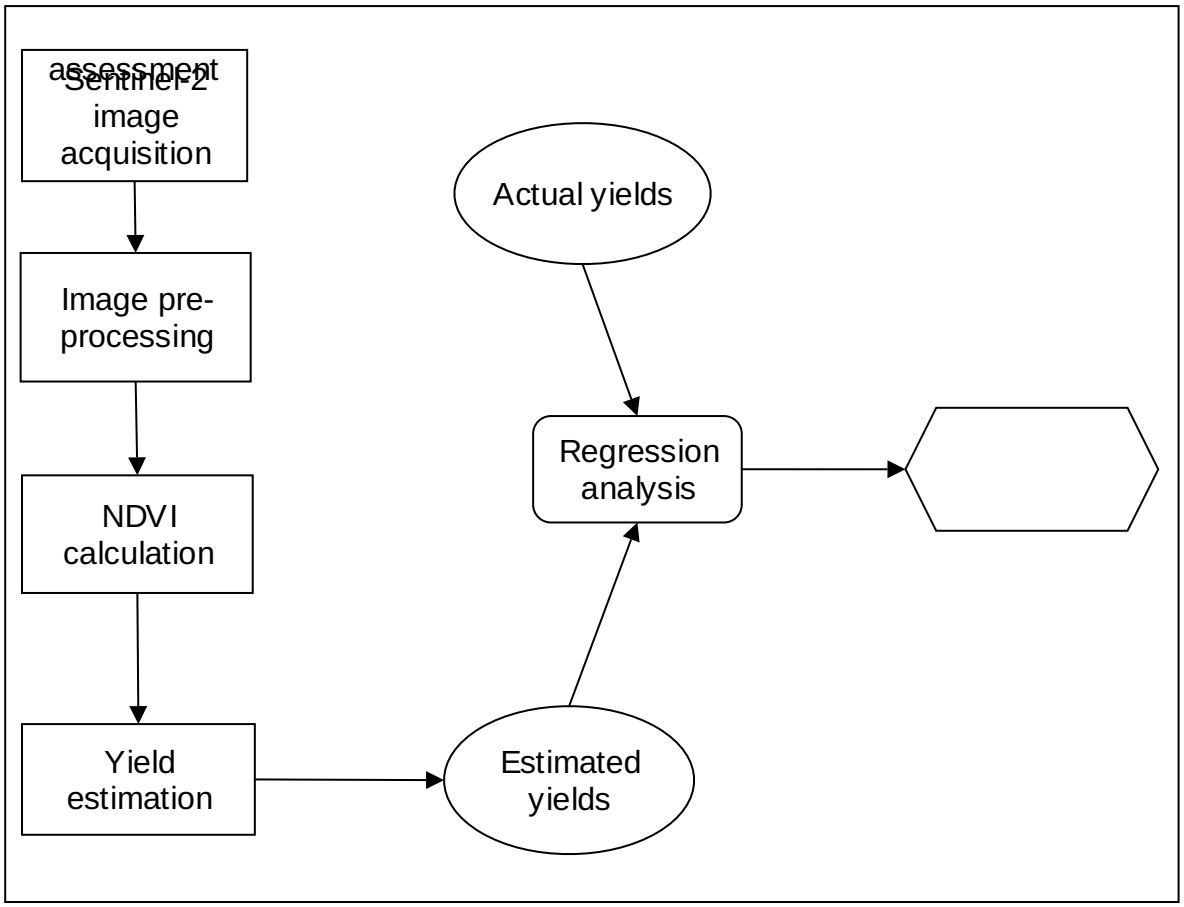


Figure 1: Flow diagram of the methodology

3.2. Study area

Makande Prison Farm is found in Chimaliro Traditional Authority in Thyolo district. It is located at 15°57'56.89"S and 35°15'37.91"E, 21 km East of Thyolo township. It is 14 hectares in size. The area has red-brown, residual clay soils on top of its geology. These soils are derived from the breakdown of ferromagnesian constituents of gneisses and diorite dikes. Black cotton clays and sandy clay soils overlay the red-brown residual clay soils. Lateritic concretions are also common in the soils. Its underlain by migmatic gneisses and granulites is referred to as the Malawi Basement Complex (Geological Survey Department, 1973). Figure 2 shows the map of the study area, Makande prison farm.

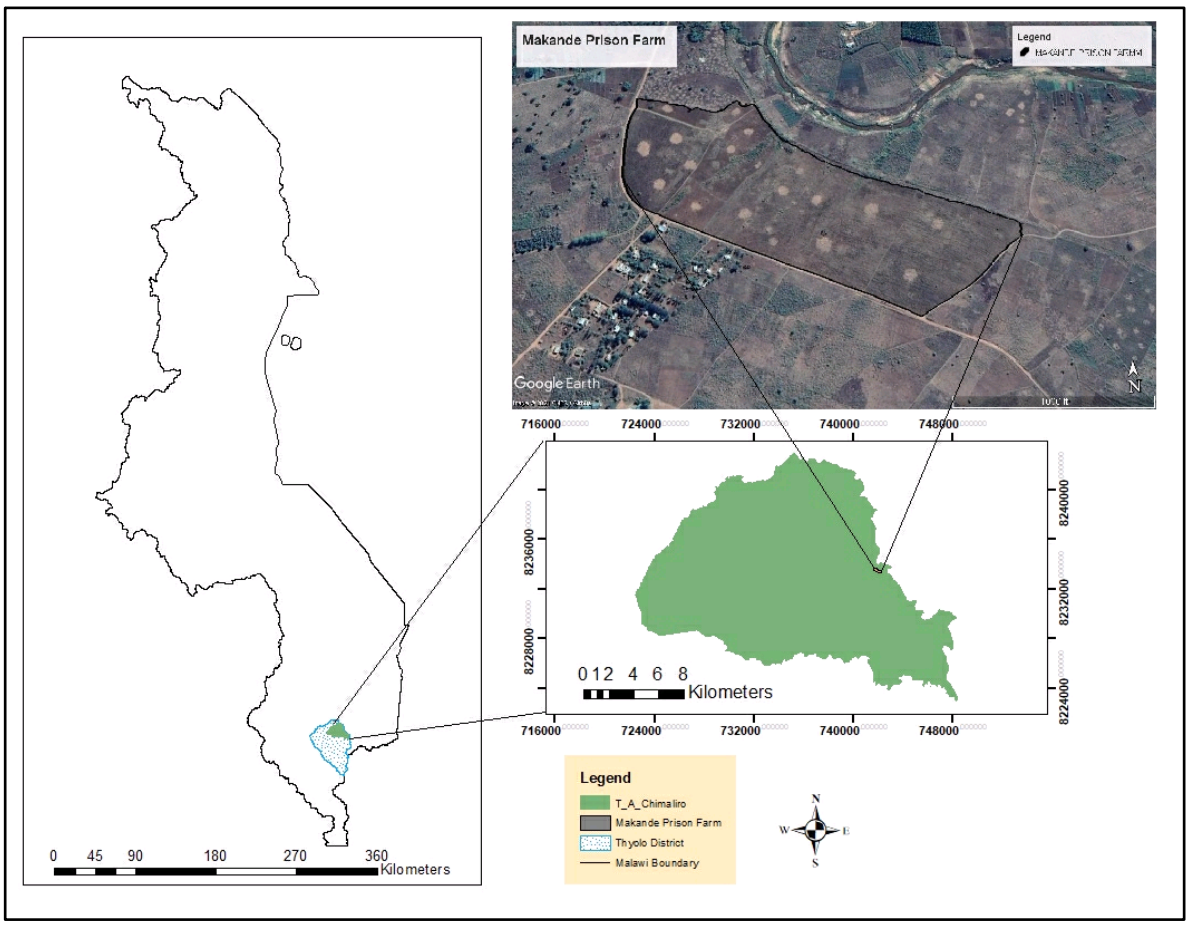


Figure 2: Map of the study area, Makande prison farm

3.3. Data Collection

The study involved the use of satellite image data for predicting crop yield and actual crop yield data for checking the accuracy of the prediction method. The images used were from mid to end of February. This is when the maize was tasselling. Sellers, (1985); Ngie & Ahmed, (2017); Lykhovyd, (2020), recommended that this is the best time to predict maize yield.

3.3.1. Crop yield data

Data was collected from the field officer of Makande prison farm. Four bags/ha of both NPK and Urea were applied; totaling 122 bags of fertilizer each year. Termites and fore-army worms have been attacking the farm usually during the tasselling stage, which is also the critical point in maize production. In all the years, all farm management activities were timely done thus planting, weeding, and fertilizer application. The pests were treated on time in the years. For crop yield forecasting the most important variables are planting dates, pest and disease control, fertilizer application, and weeding dates (Hoefsloot, et al., 2012). Table 1 shows the annual maize yield.

3.3.2. Satellite Data

The study was based on remotely sensed data derived from the Sentinel-2 satellite constellation, which is part of the Copernicus program. Sentinel-2 images, comprising of both 2A and 2B satellites, were acquired from the Copernicus Open Access Hub. The multi-spectral images composed of 13 bands provided spectral information in the visible, red-edge, NIR, and SWIR electromagnetic spectrum (Battude, et al., 2016). Ten spectral bands provided valuable data for vegetation monitoring, of which four had a spatial resolution of 10 m and six of 20 m (Chivasa, Mutanga, & Biradar, 2017; Lambert, Traoré, Blaes, Baret, & Defourny, 2018). Red and NIR Bands were used for calculating NDVI, and these were represented by band 04 and band 08 respectively (Chivasa, Mutanga, & Biradar, 2017). Sentinel-2 images were used in this study because of the high temporal resolution of 5 days, and high spatial resolution. The temporal resolution increased the probability of acquiring cloud-free images; as such it resolved the pixel contamination challenge. Consequently, yield monitoring in highly fragmented agricultural systems like Africa is very possible, with

these high spatial resolution sensors (Chivasa, Mutanga, & Biradar, 2017; Karst, et al., 2020).

3.4. Data analysis

The objectives of this study were achieved by using satellite images and crop yield data. This section describes how these, satellite images and crop yield data, were used to achieve the objectives of the study.

3.4.1. Image pre-processing

Automatic image masking to remove clouds and shadows from the images using an SCP in QGIS Software was the first task that was performed. Then the unwanted parts of the image were clipped out using the 'clip raster by mask layer'. This was done because the acquired image covered a vast area.

3.4.2. NDVI calculation

NDVI for the images was calculated using a raster calculator in QGIS. The calculator applies the formula below, developed by (Rouse, Haas, Schell, & Deering, 1973).

$$NDVI = (NIR - RED) / (NIR + RED)$$

The NDVI values showed how green the canopy cover is and how dense the crops were populated. High NDVI values indicated high crop density and greener canopy cover whereas low values indicated low crop density and less green canopy cover. Mean and standard deviation for the NDVI were extracted. The mean NDVI was used in yield estimation. Table 1 shows the NDVI values extracted from the satellite images that were used for estimating crop yield.

Table 1: Mean and standard deviation of NDVI

Year	Mean	St. dev
2019	0.2731	0.048872
2020	0.1862	0.050001
2021	0.3121	0.0469

The moderate mean values found in this exercise show that the vegetation in most parts of the farm was not very green, with a small part of the farm having greener vegetation.

3.4.3. Yield estimation

The NDVI mean values from the Sentinel 2 image within the total area of the farm were used for analysis as a statistical sample. In this study, the empirical regression model was used to calculate the estimated yield:

$$Y = -13.098(x)^2 + 23.302(x) - 2$$

Where: x = average NDVI corresponding to the maize at the tasseling stage, and Y = Yield (ton/ha) (Soria-Ruiz & Fernandez-Ordoñez, 2017). The total estimated yield was calculated by multiplying the estimated value with the total number of hectares of the farm, 14 hectares. The results are shown in Table 4, in the Results and Discussion section.

3.5. Accuracy assessment

Linear regression analysis was performed to correlate the estimated maize yield with the actual maize yield, with confidence level set to $\alpha = 95\%$. Linear regression analysis was done using Microsoft excel.

4. RESULTS AND DISCUSSION

The study used satellite data to forecast crop yield and actual maize yield for checking the accuracy of the methodology. This section shows and explains the results obtained from the analyses that were done during the study.

4.1. Crop yield data

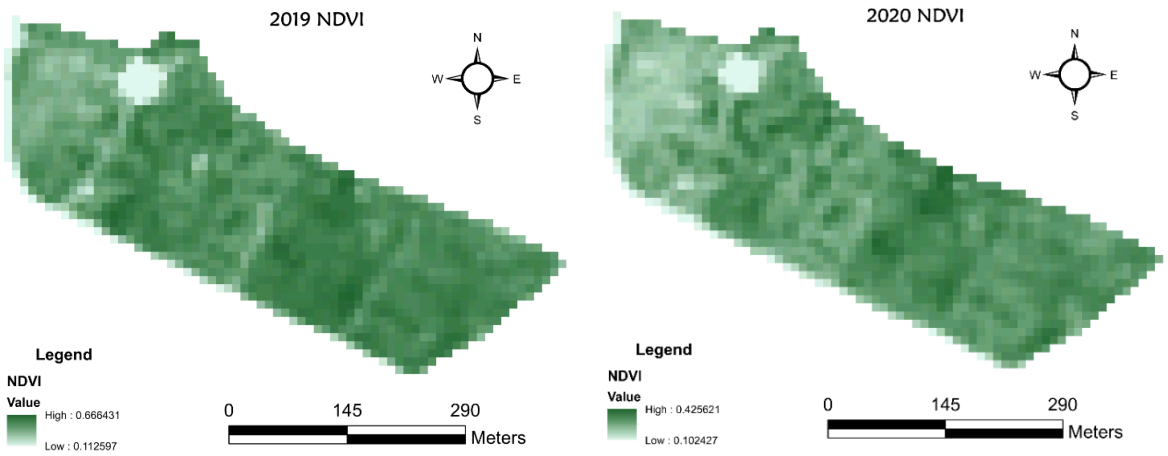
Table 2 shows the actual annual maize yields that were recorded at Makande prison farm from 2019 to 2021. The simple analysis indicated that the highest yields (47.5 tones) were obtained in 2021, and the lowest yields (22.5 tones) were obtained in 2020.

Table 2: Actual annual yield

Year	Yield (ton/ha)	Total Yield (ton)
2019	3.06	42.85
2020	1.61	22.5
2021	3.39	47.5

4.2. NDVI images

Figure 3 shows the NDVI image for the Makande prison farm extracted from the satellite images for 2019, 2020, and 2021. The greener parts indicated high maize canopy cover whereas the whitish-gray parts indicated low vegetative cover. Statistical data extracted from the images are shown in table 2.



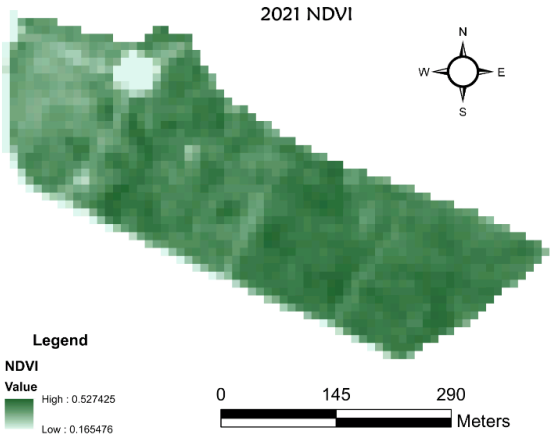


Figure 3: NDVI images for 2019, 2020, and 2021

4.3. Estimated yield

The estimated yields from the NDVI were in tons/ha as illustrated in table 3. The highest yields were predicted in the year 2021, and the lowest yields in 2019. This indicated that low mean NDVI values give low yield estimates as compared with high mean NDVI values which give high yield estimates.

Table 3: Estimated maize yield

Year	Estimated yield(tons/ha)	Total estimated yield(tones)
2019	3.086	43.204
2020	1.584	22.176
2021	3.696	51.744

4.4. Comparison of estimated and actual yields

A simple comparison analysis was done to check the differences between the actual yield and the estimated yield and the results produced are shown in table 4.

Table 4: Estimated yields and actual yields

Year	Estimated yield(tones)	Actual yield(tones)	Difference %
2019	43.204	42.85	-0.00354
2020	22.176	22.5	0.00324
2021	51.744	47.5	-0.04244

Table 4, shows the differences in percentage between the estimated yields and the actual yields which are less than 1% in all the years. The negative difference percentages indicated an over-estimation and the positive difference percentage indicated under-estimation. These differences are likely due to several contamination sources that can confound the potential relationship between NDVIs and maize yield. For instance, cloud and atmospheric-moisture contamination can influence the NDVI signal (Huang, Wang, Li, Tian, & Pan, 2013). These results demonstrated the potential of the estimation method used in the study.

For the performance analysis of the estimated yields by NDVI, estimated yields were compared with the actual yields by linear regression analysis, thus estimating the RMSE, MAE, and correlation coefficient. RMSE results showed that the results from the model are more precise.

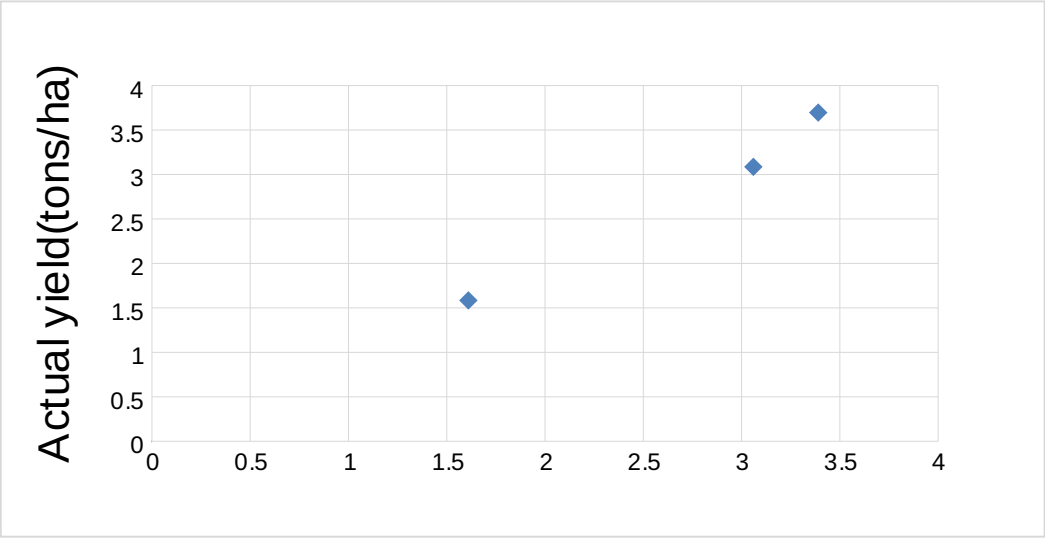


Figure 4: Linear correlation between estimated yields and actual yield

Figure 4 shows the correlation between estimates from maize canopy cover and actual maize yield. The canopy calculated through the NDVI index was closely related to the yield of maize. This is attributed to the fact that the green pigment in maize leaves is largely responsible for its yields through photosynthetic activities (Ngie & Ahmed, 2017). This means that the greener the canopy cover, the higher the yield. Resulting in making maize yield have a direct response to canopy cover.

Table 5: Results from a regression analysis

Regression Statistics	
Multiple R	0.994035133
R Square	0.988105845
Adjusted R Square	0.976211691
Standard Error	0.167643944
RMSE	0.177940065
MAE	0.119333333

The evaluation of the accuracies was noted through the lower RMSE and MAE, and the increased R^2 values for all the years. According to Montgomey, et al., (2012) multiple R shows how the two variables being analyzed are correlated, whereas R^2 shows the ability of the dependent variable, NDVI, to explain the independent variable which in this case is the maize yield. The RMSE is the standard deviation of the prediction error while MAE shows the average of all errors in the measurements (Montgomey, Peck, & Geoffrey Vining, 2012; Poovammal & Sellam, 2016).

As illustrated in table 5, multiple R (0.99403) is the correlation coefficient illustrating the existence of a strong linear relationship between the forecasted yield and actual yield. R^2 (0.988) indicated that a very high more than 98% of maize yield can be explained by NDVI. The MAE (0.11933) and the RMSE (0.17794) indicated that canopy cover generates a prediction model with a small error. Therefore, with the statistics, it is worth concluding that maize yield has a direct response to canopy cover. The canopy cover calculated through the NDVI was closely related to the yield of maize hence making it an excellent indicator of maize yields. This was in agreement with Ngie & Ahmed, (2017), who stated that spectral

vegetation indices that are responsive to the green pigments are excellent indicators for vegetation quant

The statistical analyses applied in the study are in line with the conclusions made by Lykhovyd, (2020) that the NDVI-based simple regression models are very encouraging. Indeed, the estimated yields from satellite-based crop yield predictions can be as reliable as actual yields as stated by Benedetti & Rossinni, (1993); Burke & Lobell, (2007); Sivarajan, (2011); Huang, et al., (2013); Huang, et al., (2014); Battude, et al., (2016); Chivasa, et al., (2017); Soria-Ruiz & Fernandez-Ordoñez, (2017); Lobell, et al., (2018); Lykhovyd, (2020).

5. CONCLUSION AND RECOMMENDATIONS

The study aimed at establishing the relationship between yield estimation based on canopy cover (NDVI) and actual yield data. Using the sentinel-2 satellite images, NDVI was calculated which was used for predicting maize yields. Based on the regression analysis, it can be concluded that there is a very strong relationship between NDVI and crop yield. The results from the accuracy assessment have proven how effective the methodology is, thus, the methodology is very effective and reliable for maize yield forecasting.

Even though the overall strength of the relationship between predicted and observed yields was good, a more robust study would need to be conducted with the use of more satellite images and yield data for not less than 6 years. A similar study can also be conducted in either September or October, thus in irrigated maize fields, to reduce the cloud cover effect which reduces the accuracy of the results.

6. REFERENCES

Adeniyi, O. D., Szabó, A., Tamás, J., & Nagy, A. (2020).*Wheat Yield Forecasting Based on Landsat NDVI and SAVI Time Series.* Preprints.

Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Sicre, M. C., . . . Demarez, V. (2016). Estimating Maize Biomass and Yield over Large Areas Using High Spatial and Temporal Resolution Sentinel-2 like Remote Sensing Data. *Remote Sensing of Environment*, 668-681.

Benedetti, P., & Rossinni, P. (1993). On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in Emilia Romagna. *Remote Sens. Environ.*, 311-326.

Burke, M., & Lobell, D. B. (2007). Satellite-Based Assessment of Yield Variation and Its Determinants in Smallholder African Systems. *Proceedings of the National Academy of Sciences*, 2189–2194.

Chivasa, W., Mutanga, O., & Biradar, C. (2017). Application of remote sensing in estimating maize grain yield in heterogeneous African agricultural landscapes: a review. *International Journal of Remote Sensing*, 6816-6845.

Copernicus Open Access Hub. (2018). Retrieved from https://scihub.copernicus.eu

Dorward, A., & Chirwa, E. (2011). The Malawi Agricultural Input Subsidy Programme: 2005/06 to 2008/09. *International Journal of Agricultural Sustainability 9*, 232–247.

Ezeh, A. C., Bongaarts, J., & Mberu, B. (2012). Global Population Trends and Policy Options. *The Lancet 380*, 142–148.

Geological Survey Department. (1973). *The Geology of the Thyolo Area.* Zomba: The Government Printer.

Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of crop biophysical characteristics. *Plant Physiology*, 165–173.

Hayes, J. T., O'Rourke, P. A., Terjung, W. E., & Todhunter, P. E. (1982). YIELD: A numerical crop yield model of irrigated and rainfed agriculture. *Publications in Climatology*, 35.

Héctor, G.-M., Héctor, F.-M., Roberto, A.-H., Abdul, K.-G., Leonardo, T.-C., Oscar, M.-V. R., & Vázquez-Peña, M. A. (2020). Corn Grain Yield Estimation from Vegetation Indices Canopy Cover, Plant Density, and a Neural Network Using Multispectral and RGB Images Acquired with Unmanned Aerial Vehicles.*Agriculture*, 1-24.

Hoefsloot, P., Ines, A., van Dam, J., Duveiller, G., Hansen, J., & Kayitakire, F. (2012). *Combining crop models and remote sensing for yield prediction: Concepts,*

applications and challenges for heterogeneous, smallholder environments. Ispra: European Union

Hoogenboom, G., Liang, S., & Fang, H. (2011). Integration of MODIS LAI and vegetation index products with the CSM-CERES-Maize model for corn yield estimation. *Intern. J. Remote Sens*, 1039-1065.

Huang, J., Dai, Q., Wang, H., & Han, D. (2014). Empirical Regression Model Using Ndvi, Meteorological Factors For Estimation Of Wheat Yield In Yunnan, China. *11th International Conference on Hydroinformatics.* New York: City University of New York.

Huang, J., Wang, X., Li, X., Tian, H., & Pan, Z. (2013). Remotely Sensed Rice Yield Prediction Using Multi-Temporal NDVI Data Derived from NOAA's-AVHRR.

Iskandar, A. (2020).*NDVI Imaging in Plantation Management.* Britannica.

J, H., X, W., X, L., H, T., & Z, P. (2013). Remotely Sensed Rice Yield Prediction Using Multi-Temporal NDVI Data Derived from NOAA's-AVHRR.

Jackson, R., Slater, P., & Pinter, P. (1983). Discrimination of Growth and Water Stress in Wheat by Various Vegetation Indices through Clear and Turbid Atmospheres.*Remote Sensing of Environment 13*, 187–208.

Jayne, T. S., & Rashid, S. (2010). *The Value of Accurate Crop Production Forecasts. MSU International Development Working Paper 108.* Michigan: Michigan State University.

Johnson, D. M. (2014). An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, 116–128.

Karst, I. G., Mank, I., Traoré, I., Sorgho, R., Stückemann, K., Simboro, S., . . . Sauerborn, R. (2020). Estimating Yields of Household Fields in Rural Subsistence Farming Systems to Study Food Security in Burkina Faso. *Remote Sensing*, 1-20.

Labus, M. P., Nielsen, G. A., Lawrence, R. L., Engel, R., & Long, D. (2002). Wheat yield estimates using multi-temporal NDVI satellite imagery. *International Journal of Remote Sensing*, 4169-4180.

Lambert, M., Traoré, P. C., Blaes, X., Baret, P., & Defourny, P. (2018). Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sensing of Environment*, 647–657.

Léger, P.-M. (2014). The early explanatory power of NDVI in crop yield modelling. *International Journal of Remote Sensing*, 1-27.

Lobell, D. B., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., & Murray, S. (2018). *Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis in Uganda.* World Bank Group.

Lobell, D. B., Ortiz-Monsterio, J. I., Gurrola,, F. C., & Valenzuuela, L. (2007). Identification of saline soils v

Lv, X. (2013). *Remote sensing, Normalized Difference Vegitation Index, and crop yield forecatsing.* Urbana, Illinois: University of Illinois at Urbana.

Lykhovyd, P. (2020). Sweet Corn Yield Simulation Using Normalized Difference Vegetation Index and Leaf Area Index. *Journal of Ecological Engineering*, 228–236.

Montgomey, D. C., Peck, E. A., & Geoffrey Vining, G. (2012). *Introduction to linear regression analysis.* Hoboken: John Wiley & Sons, Inc.

Mwanaleza, E. J. (2016). Use of remote sensing and satellite imagery in estimation crop production: Malawi's experience. *Seventh International Conference on Agricultural Statistics* (pp. 4-26). Rome: Ministry of Agriculture, Irrigation and Water Development, Statistics Unit.

Ngie, A., & Ahmed, F. (2017). Estimation of Maize grain yield using multispectral satellite data sets (SPOT 5) and the random forest algorithm. *South African Journal of Geomatics*, 11-30.

Poovammal, E., & Sellam, V. (2016). Prediction of Crop Yield using Regression Analysis. *Indian Journal of Science and Technology*.

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

Reynolds, C. A., Petersen, M. S., Yitayew, M., Slack, D. C., Hutchison, C. F., & Huete, A. (2000). Estimating crop yields and production by integrating the FAO Crop Specific Water Balance model with real-time satellite data and ground-based ancillary data. *International Journal of Remote Sensing*, 3487-3508.

Rojas, O. (2007). Operational Maize Yield Model Development and Validation Based on Remote Sensing and Agro-Meteorological Data in Kenya. *International Journal of Remote Sensing*, 3775–3793.

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. *Third Earth Resources Technology Satelite-1 Symposium*, (pp. 309-317). Washington, D.C.

Rukuni, M., Munyki-Hungwe, M., & Matondi, P. M. (Eds.). (2006). *Zimbabwe's Agricultural Revolution Revisited.* Harare: University of Zimbabwe Publications.

Sawasawa, H. L. (2003). *Crop Yield Estimation: Integrating RS, GIS, and Management factors.* Enschede: Internationa Institute for Geo-Information Sciences, and Earth observation.

Sellers, P. J. (1985). Canopy Reflectance, Photosynthesis, and Transpiration. *International Journal of Rer*

Shiba, B. S., Mabaso, S. D., Dlamini, S. N., & Singwane, S. S. (2020). Remote Sensing for Sugarcane Crop Yield Estimation in Eswatini: Case of Lower Usuthu Smallholder Irrigation Project Sugarcane Farms. *International Journal of Agriculture, Forestry and Fisheries*, 19-27.

Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of Remote Sensing in Precision Agriculture: A Review. *Remote sensing*, 1-31.

Sivarajan, S. (2011). *Estimating Yield of Irrigated Potatoes Using Aerial and Satellite Remote Sensing.* Logan: Utah State University.

Soria-Ruiz, J., & Fernandez-Ordoñez, Y. M. (2017). Maize crop yield estimation with remote sensing and empirical models.

Staggenborg, S. A., & Taylor, R. K. (2000, 03 17). Predicting Grain Yield Variability with infrared images. *Proceedings of the 5th International Conference on Precision Agriculture, Bloomington* (pp. 1-6). Minnesota: American Society of Agronomy. Retrieved from Cabdirect Website: https://www.cabdirect.org/cabdirect

Tannura, M. A., Irwin, S. H., & Good, D. L. (2008). *Weather, technology, and corn and soybean yields in the U.S. Corn Belt*. Retrieved from University of Illinois at Urbana-Champaign, Department of Agricultural and Consumer Economics website: http://www.farmdoc.uiuc.edu/marketing/morr/morr_archive.html

Tucker, C. J. (1980). Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment 8*, 127–150.

Tychon, B., Balaghi, R., Eerens, H., & Jlibene, M. (2008). Empirical regression models using NDVI, rainfall, and temperature data for early prediction of wheat grain yields in Morocco. *International Journal of Applied Earth Observation*, 438-452.

UN country team in Malawi. (2021, 03 16). *Drone images help farmers to predict crop yields in Malawi.* Retrieved from Relief Web Site: https://www.reliefweb.int/

US Geological Survey. (2010). *What's New?* Retrieved from Greenness of the Conterminous U.S.: http://ivm.cr.usgs.gov/whatnew.php

Walker, P. J. (1989). *Famine Early Warning Systems.* London: Earthscan Publications.

Yahya, N., Mulat, S., & Getachew, M. (2019). *Technical manual on the use of Satellite Images in Ecological Studies.* Addis Ababa.