WHEAT YIELD FORECAST USING REMOTE SENSING AND GIS IN EAST ARSI ZONE, ETHIOPIA

BY

AKLILU FIKRE REDA

A Thesis Submitted to the School of Graduate studies in Partial Fulfillment of the Requirements for the Degree of Master of Science in Remote Sensing and Geographic Information Systems (GIS)

Addis Ababa University
Addis Ababa, Ethiopia
June, 2015

WHEAT YIELD FORECAST USING REMOTE SENSING AND GIS IN EAST ARSI ZONE, ETHIOPIA

BY

AKLILU FIKRE REDA

A Thesis Submitted to the School of Graduate studies in Partial Fulfillment of the Requirements for the Degree of Master of Science in Remote Sensing and Geographic Information Systems (GIS)

Under the guidance of

Dr. K.V.SURYABHAGAVAN
Asst. Professor and Prof. M. Balakrishnan
Addis Ababa University, Addis Ababa

Addis Ababa University
Addis Ababa, Ethiopia
June, 2015

Addis Ababa University

School of Graduate Studies

This is to certify that the thesis prepared by AKLILU FIKRE REDA, entitled: *WHEAT YIELD FORECAST USING REMOTE SENSING AND GIS IN EAST ARSI ZONE, ETHIOPIA* and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Remote Sensing and GIS complies with the regulations of the university and meets the accepted standards with respect to the originality and quality.

Signed by the Examining Committee:

Session Chair Prof. Tigistu Haile Signature ______________Date_________

Examiner Dr. Gtachew Berhan Signature ____________Date_________

Examiner Dr. Dereje Ayalew Signature ____________Date__________

Advisor Dr. K.V. Suryabhagavan Signature ____________Date__________

Chairman, Department

Dr. Seifu Kebede Signature _____________Date_________

ACKNOWLEDGEMENTS

I am very much grateful to my advisors, Dr. K.V. Suryabhagavan and Prof. M. Balakrishnan for their collaboration in sharing knowledge and valuable comments, guidance and encouragement throughout the period of this study.

I am also thankful to the Central Statistical Agency (CSA) for provision of financial support to complete my thesis work in addition to the scholarship opportunity.

I thank Shemeles Mulugeta for his help in the analysis of the data and all staff of CSA for their unreserved and tireless support, sharing of knowledge and invaluable advice in every aspect during my study period.

My colleague, Abiy Wegderes and Birhanu Bergello has helped me in a number of ways and deserve many thanks. I am grateful to all my family members, my wife Meaza and my son Noah Aklilu. Your encouragement is unforgettable.

Last but not least, I would like to extend my appreciation to my M.Sc. batch mates for their team spirit and cooperation during the whole course work.

There are many other individuals, who have helped me in one way or the other during my M.Sc. study period. I am not mentioning the names of each and every one of them; I thank all of them which helped me to successfully complete this task.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS i

TABLE OF CONTENT ii

LIST OF TABLESE vi

LIST OF FIGURES vii

LIST OF APPENDICES viii

LIST OF ABBREVIATIONS ix

ABSTRACT xi

CHAPTER ONE xi

INTRODUCTION 1

1.1 Background 1

1.2 Statement of the problem 3

1.3 Objective 4

1.4 Significance of the study 4

1.5 Limitation of the study 5

1.6 Organization of the study 5

CHAPTER TWO 6

LITERATURE REVIEW 6

2.1 Theoretical framework 6

2.1.1 Satellite imagery for agricultural application 7

2.1.1.1 High spatial resolution satellite imagery 7

2.1.1.2 Low spatial resolution satellite imagery 7

2.1.2 Remote sensing derived variables for agricultural application 8

2.1.2.1 Vegetation index (VI) 9

2.1.2.2 Water Requirement Satisfaction Index (WRSI) 9

2.1.2.3 Weather variables 10

2.2 Crop yield forecast in the world 11

2.3 Crop yield forecast in Ethiopia 14
2.4 Wheat crop 17
2.5 Agricultural periods in Ethiopia 17
2.5.1 Shortest rainy season 18
2.5.2 Longest rainy season 18
CHAPTER THREE 20
MATERIALS AND METHODS 20
3.1 Description of the study area 20
3.1.1 Population 20
3.1.2 Temperature 20
3.1.3 Rainfall 21
3.1.4 Cropping condition 22
3.1.5 Soil 23
3.2 Source of data, and software's used. 24
3.2.1 Satellite imagery, available models and ancillary data 24
3.2.1.1 Rainfall estimate (RFE 2.0) 24
3.2.1.2 SPOT VEGETATION (VGT) 25
3.2.1.3 Water requirement satisfaction lndex 25
3.2.1.4 Pansharpened spot 5 imagery 26
3.2.1.5 Ancillary dataset 26
3.2.1.6 Official yield statistics 26
3.2.1.7 Materials 27
3.2.1.7.1 SPIRIT Software 28
3.2.1.7.2 Statistical Package For Social Science (SPSS) 29
3.3 Data processing and analysis 30
3.3.1 Selection of date of satellite imagery 31
3.3.2 Preprocessing of satellite images 33
3.3.2.1 Projection of the data 33

3.3.2.2 Extracting region of interest 34
3.3.2.3 Filtering of the study area 34
3.3.3 Wheat crop masking 34
3.3.3.1 Image interpretation 34
3.3.3.2. Accuracy assessment 36
3.3.3.3. Crop mask data derivation 38
3.3.4. Preparing independent variables using mask data 39
3.3.5 Multiple linear regression analysis 41
3.3.6 Multicollinearity 43
3.3.6.1 Effects of Multicolinearity 43
3.3.6.2 Multicolilnearity Diagnostics 44
3.3.6.2.1 Variance Inflation Factor (VIF) 44
CHAPTER FOUR 45
RESULTS 45
4.1 Correlation analysis of different indices with wheat yield 45
4.1.1 Correlation of NDVI with wheat yield 45
4.1.2. Correlation of rainfall with wheat yield 46
4.1.3. Correlation of WRSI with wheat yield 47
4.1.4 Correlation of ETa with wheat yield 48
4.2 Multiple Linear Regression Model for Yield Forecasting 49
4.2.1 Estimation of the model Parameters: 49
4.2.2 Test for Significance of Regression 49
4.2.3 Tests on Individual Regression Coefficients 51
4.3 Contrasting of conventional crop yield prediction with the developed model 52
4.4 Wheat crop forecast for the year 2014 54
CHAPTER FIVE 56
DISCUSSION 56

CHAPTER SIX 58

CONCLUSION AND RECOMMENDATIONS 58

6.1. Conclusion 58

6.2. Recommendations 59

Reference 60

LIST OF TABLES

Table 3.1 Area, Production and yield of cereal crops for private peasant holding for Meher season 2013/14. 22

Table 3.2 Trends of Wheat Crop Yield (2004-2013) ---27

Table 3.3 Summary of equipment and materials used for data collection and analysis. . 28

Table 3.4 Accuracy assessment table 38

Table 3.5 Table showing observed yield and independent variables.............................. 42

Table 4.1: Correlation coefficient of dependent variable and independent variables-----45

Table 4.2 Wheat yield forecast model variance analysis (ANOVA)............................... 51

Table 4.3 Variance Inflation Factor (VIF) between NDVIa and REE 2.0------------------51

Table 4.4 Parameter estimates for the wheat forecast model... 52

Table 4.5 Wheat production level of the year 2014 for East Arsi zone........................... 55

LIST OF FIGURES

Figure 2.1 Seasonal calendar of crops 19
Figure 3.1 Location map of the study area. 20
Figure 3.2 Monthly temperature vs Rainfall Distribution of the study area 21
Figure 3.3 Soil map of the study area 23
Figure 3.4 Trend of Wheat crop yield (2004 – 2013) 27
Figure 3.5 Methodological flow chart 30
Figure 3.6 SPOT VGT image of Ethiopia,1st Decade of June 2004 32
Figure 3.7 RFE 2.0 image of Ethiopia, Mean of 2003(June - September) 32
Figure 3.8 ETa imagery of september,2010 33
Figure 3.9 PET imagery for 2004 mean annual 33
Figure 3.10 SPOT image of the study area. 35
Figure 3.11 Interpreted image of the study area 36
Figure 3.12 Random points generated for accuracy assessment. 37
Figure 3.13 Crop mask data for wheat 39
Figure 3.14 Crop coefficient of wheat in different stage(Planting-Flowering) 40
Figure 3.15 NDVI value for the month of july,2004 41
Figure 3.16 Mean ETa for the month of June,2004 41
Figure 3.17 PET for the month of June, 2004. 41
Figure 3.18 RFE for the month ofJune, 2004 41
Figure 4.1 Graph showing yield and NDVIa 46
Figure 4.2 Graph showing yield and rainfall 47
Figure 4.3 Graph showing yield and WRSI 47
Figure 4.4 Graph showing yield and ETa 48
Figure 4.5 Graph showing yield and ETa total 48
Figure 4.6 Comparison between the wheat yield estimated by the spectro agrometeorological model and the observed yields for the study area 50
Figure 4.7. Comparison between wheat yield estimated by the model and the observed yield. 54
Figure 4.8 Wheat yield forecast map of 2014. 55

LIST OF APPENDICES

Appendix 1 Sample GPS reading for accuracy assessement……………………………64

Appendix 3 Accuracy assessemnt matrix result ………………………………………..65

LIST OF ABBREVIATION

AGRISTARS – Agriculture and Resource Inventory Surveys through Aerospace Remote Sensing

ANOVA-Analysis of Variance

AMSU - Advanced Microwave Sounding Unit

asl - Above Sea level

AVHRR – Advanced Very High Resolution Radiometer

CPSZ – Crop Production System Zone

CSWB - Crop Specific Water Balance

CV- Coefficient of Variation

CYMFS- Crop Yield Monitoring and Forecasting System

DEM- Digital Elevation Model

ECMWF- European Center for Medium Range Weather Forecast

ETA- Actual Evapotranspiration

fAPAR – Fraction of Absorbed Photo synthetically Active Radiation

FAO – Food and Agriculture Organization

FAS – Foreign Agriculture Service

FEWSNET – Famine Early Warning System

GAC – Global Area Coverage

GDP- Gross Domestic Product

GIS – Geographic Information System

GLAM – Global Agriculture Monitoring

IFPRI – International Food Policy Research Institute

JRC- Joint Research Cycle

LAC – Local Area Coverage

LACIE – Large Area Crop Inventory Experiment

LEAP- Lively hood Early Assessment Protection

MERIS - Medium Resolution Imaging Spectrometer

MIR – Middle Infrared

MODIS - Moderate Resolution Imaging Spectroradiometer

MSS- Multi Spectral Scanner

MVC- Maximum Value Composite

NASA - National Aeronautics and Space Administration

NDVIa – Average Normalized Difference Vegetation Index

NDVI- Normalized Difference Vegetation Index

NMA- National Meteorological Agency

NOAA – National Oceanic and Atmospheric Administration

NWP – Numerical Weather Prediction

PET – Potential Evapotranspiration

P Value – Probability Value

Qt/ha - Quintal per Hectare

RFE – Rainfall Estimate

RMSE – Root Mean Square Error

RS- Remote Sensing

RVI – Ratio of Vegetation Index

SPSS- Statistical Package for Social Science

USAID - United States Developmental Aid

WFP- World Food Program

WRSI – Water Requirement Satisfaction Index

ABSTRACT

Ethiopia's agriculture is involving substantial variations in crops grown across in different regions and agro ecological zones. The core crop season is the long rainy season, with harvests between Novembers to February. Five major cereals (teff, wheat, maize, sorghum and barley) are the major agricultural crops of the country, accounting for about three-quarters of total area cultivated and 29% of agricultural GDP in 2005/06 (i.e. 14% of the country total GDP). This thesis reports the development of an operational agrometeorological yield model for wheat crop derived from time series data of SPOTVEGETATION, actual and potential evapotranspiration, rainfall estimate and satellite data for the years 2004−2013. Official grain yield data maintained by the Central statistical Agency of Ethiopia was used to validate the strength of the indices in explaining the yield. Crop masking at crop land area was applied and refined by using agroecological zones suitable for the crop of interest. Correlation analyses were used to determine associations among crop yield, spectral indices and agrometeorological variables for wheat crop of the long rainy season (Meher). Indices with high correlation with wheat yield were identified. Average Normalized Difference Vegetation Index (NDVIa) and rainfall have high correlation of wheat yield with 96% and 89%, respectively. That means there variables are positively strong related with wheat yield. Multicollinearity was assessed among the independent variables. Many studies reported that linear regression modeling is the most suitable method to produce yield predictions by using remote sensing derived indicators together with bioclimatic information. Accordingly, NDVIa was used to the regression models with P- value of less than 0.0069 at 95% confidence level were derived. Very encouraging results were obtained by the model (r^2 =0.93, RMSE= 0.99 quintal /ha and 16.01% CV). The result of ANOVA table shows that there is a linear relationship between the dependent and independent variables. Policy makers need accurate and timely information on crop production and for that matter multiple linear regression model and remote sensing method are appropriate approach for crop yield forecasting. This study has revealed that crop yield forecasting is possible using remote sensing and GIS, and this method can be used as a modern tool for similar analysis. For future studies, more researches can be done using longer period of time serious data to enhance the model results.

Keywords: NDVI, Remote *sensing*, RFE 2.0, SPOT VEGETATION**,** Wheat yield.

CHAPTER ONE

INTRODUCTION

1.1 Background

Ethiopia's crop agriculture is complex, involving substantial variation in crops grown across the country's different regions and agroecologies. Smallholders account for 96% of total area cultivated and generate the key share of total production for the main crops. The core crop season is the June to September, with harvests between November to February. Five major cereals (teff, wheat, maize, sorghum, and barley) are the core of Ethiopia's agriculture and food economy, accounting for about three-quarters of total area cultivated and 29% of agricultural GDP in 2005/06 (i.e. 14 % of the country total GDP).(Alemayehu *et.al*, 2010). Crop yield forecasting before harvest is necessary especially in regions characterized by climatic uncertainties. The yield estimates projected before the actual production is required for taking various policy decisions. Hence, early assessment of crop yield is necessary, particularly in countries that depend on rain-fed agriculture as their main source of economy. It enhances timely provision of information for use in food security (Sawasawa, 2003), so as to help the administration to take essential precaution to control famine, especially during the years of natural calamities.

At present, Ethiopia has two methods of monitoring and forecasting crop yield in advance of harvest. The first method is the conventional method of data collection, where personal views about the crop conditions are collected from the stakeholders by the Central Statistical Agency (CSA). This is the body charged with providing the Ministry of Agriculture and Rural Development (MoARD) with pre-harvest crop yield information. These estimates are highly subjective and dependent on the agenda of stakeholders but are widely used by decision makers (Greatrex, 2012). The second is an operational crop yield monitoring and forecasting system that run by the Ethiopian National Meteorological Agency in conjunction with the European Union Joint Research Council (JRC) and the Food and Agricultural Organization (FAO) of the United Nations. It combines Geographical Information System (GIS), the FAO Crop-Specific Water Balance (CSWB) model; the JRC Crop Production System Zones database (CPSZ) and meteorological information from numerical models provided by the European Centre for Medium-range Weather Forecasts (ECMWF).

Additional and independent real-time Normalized Difference Vegetation Index (NDVI) satellite data from SPOT VEGETATION is also incorporated, using a specific crop mask to concentrate the analysis only on agricultural areas. This crop mask was derived from a static CPSZ database, which itself was created using survey data. Therefore, the NDVI data were used to look at plant conditions throughout the season. The technique performed well in trials and is easily applied over Ethiopia (Rojas *et al.*, 2005), but has several potential shortcomings. The first is its reliance on an empirical CSWB model rather than a processed based crop simulation model. The second potential problem is that, ECMWF modeled rainfall data significantly underestimate or overestimate precipitation that might introduce bias into the CYMFS results (Greatrex, 2012).

At present, Remote sensing (RS) data is being used in different parts of the world to estimate crop yield (Wang *et al.*, 2005). Remote Sensing and the derived vegetation indices is considered as a potential tool to improve simulations in real-time. NDVI and has been used as an indicator of the vigor of vegetative activity as represented by indirectly, observable chlorophyll activity (Hastings and Emery, 1992). Low values of NDVI are associated with lack of vegetation, dormant states of existing vegetation or stress caused by drought, over-irrigation or diseases (Hastings, 2005).

The agro-meteorological models introduce information about solar radiation, temperature, air humidity, and soil water availability, while the spectral component introduces information about crop management, varieties, and stresses not taken into consideration by the agro meteorological models (Potdar, 2002). Some cereal crops grown in rain-fed conditions where rainfall distribution parameters in space and time need to be incorporated into crop yield models in addition to vegetation indices deduced from remote-sensing data. Such hybrid models show a higher correlation and predictive capability than simple models.

The use of spectral data was studied extensively by using satellite imagery after the launch of the first civil earth observation satellite. A large range of satellite sensors provide regularly with data covering a wide spectral range (from optical through microwave) and these data are acquired from various orbits and in different spatial and temporal resolutions viz., high resolution and low resolution imagery. However, only since the growing availability of low resolution satellite images from the meteorological satellite series (NOAA) National Oceanic and Atmospheric Administration and (AVHRR) Advanced Very High Resolution Radiometer, in the early 1980s, have similar analyses been extended to large areas, including many countries in arid and semiarid climates. Low resolution systems have much better synoptic view and temporal revisit frequency compared to high resolution sensors (Rembold *et al*., 2013).

In Ethiopia, the conventional methods of production forecasting can be improved by substituting the conventional method by RS and GIS. Therefore, this research address the development of an operational agro meteorological yield model for wheat farming in East Arsi zone of Oromiya Region of Ethiopia using spectral index; the Normalized Difference Vegetation Index (NDVI) derived from SPOT-VEGETATION, meteorological data obtained from Rainfall estimate (RFE 2.0) model and Official figures produced by CSA yield data.

Wheat is one of the major cereal crops grown in the Ethiopian highlands, which lie between $6^{o} - 8^{o}$ N and $35^{o} - 42^{o}$ E, at altitudes ranging from 1500 to 3000 m. The most suitable areas for wheat production, however, fall between 1900 and 2700 m. In the highlands, rainfall distribution is bimodal and ranges between 600−2000 mm/annum. The rainy season is divided into the short rains (belg) falling from February to April and the main rains (meher) falling from June to September. At present, wheat is produced solely under rainfed conditions. Currently, about 60% of the wheat area is covered by durum and 40% by bread wheat. The estimate for 1967 consisted of about 15% for bread wheat and 85% for durum, (Hailu Gebre-Mariam, 2003).

1.2 Statement of the problem

The conventional technique of crop yield forecast is practiced in many countries by collecting a data based on field visits and reports. These process of reporting are subjective, costly, time consuming and prone to large errors due to incomplete ground observation that leads to poor crop yield assessment and delay in reporting appropriate actions to be taken, (Greatrex, 2012). Crop-weather models had also been used for crop monitoring and yield forecasting before the advent of remote-sensing method, like NDVI (Rojas, 2006).

Remote Sensing can provide accurate and fast information in crop production estimation and hence most studies have established that there is correlation between NDVI, agro meteorological data, green biomass and yield (Rojas, 2006).

In Ethiopia, such studies were mostly done at regional / national levels covering large areas using low-resolution imagery and Abiy (2014) had conducted a similar study in south Tigray Zone for maize crop. Very little has been done at zonal level.

In this research I try to fill this gap by developing a model which is used to forecast wheat yield for the year 2014 for east Arsi zone using remote sensing and GIS. The research problem has been identified based on the recommendation of Abiy (2014) to be tasted in other crop and different agro ecology zones.

1.3 Objective

The general objective of this research is to develop a wheat yield forecast model for East Arsi Zone of the Oromia Regional state, Ethiopia using Remote Sensing and GIS techniques.

Specific objectives of this research are:

- To develop a model for wheat crop yield forecast and to compare the result with other similar studies.
- To evaluate the accuracy levels of remote sensing of the new model for determine the accuracy level.
- To test the model for predicting wheat yield for the year 2014.

1.4 Significance of the study

Remote sensing could be used for crop growth monitoring and yield estimation. With the development of satellites, remote sensing images provide access to spatial information of features and phenomena on an almost real-time basis. Therefore, this present study uses the potential of remote sensing to identifying crop classes and estimating crop yield. It can also identify and provide information on spatial variability and permit more efficiency in crop forecasting for further research. The research output is expected to be important in improving the agricultural forecast of the country by shifting from crop yield forecasting based on stakeholders and experts assessment (conventional approach), which is proofed by researchers as highly subjective, time consuming and costly, to remote sensing supported approach. This attempt also serve the Central Statistical Agency for future plan using for modeling of crop forecast applying remote sensing and GIS. In the academic area, it will serve as a reference for further research on crop forecast modeling.

1.5 Limitation of the study

Limitation of the study was lack of better resolution satellite imageries especially for rainfall estimation and crop mask data derivation. Limited software available for time series image analysis (difficulty to shift when it fails to compute) was also the major limitation encountered during the implementation of this research. The study area is east Arsi zone and hence may lack to represent the characteristic of other zones of Oromiya region.

1.6 Organization of the study

Including this introductory chapter, this thesis is organized into six chapters. Each chapter has its own heading and few subheadings. Chapter one gives background information about the study. Chapter two presents a review of literature where theoretical and related literatures are reviewed. Chapter three, deals in detail, description of the study area, data acquisition and software packages and data processing and analysis method. Chapter four presents results of the study, and here correlation of different indices with wheat yield and a multivariate regression model for yield forecast are presented. Chapter five deals with discussion of the result, where the relevance of the findings are discussed in relation with related literatures. Chapter six presents conclusion and recommendations of the research output for consideration of various stakeholders and government agencies for implementation.

CHAPTER TWO

LITERATURE REVIEW

2.1 Theoretical framework

The crop annalist production estimates are derived by synthesizing information from a wide array of sources, including local reports, field surveys, climate data, and exploiting a variety of tools such as crop models and data visualization utilities. However, it is the remotely sensed earth observations that enable production forecasts to be comprehensive, global, unbiased, timely and cost-effective (Becker-Reshef *et al*., 2010). Agricultural analysis using remote sensing data requires knowledge about plants photosynthesis activity. Green plants have a unique spectral reflectance influenced by their structure and composition. The proportion of radiation reflected in different parts of the spectrum depends on the state, structure and composition of the plant. In general, healthy plants and dense canopies will reflect more radiation especially in the near infrared region of the spectrum.

In the visible part of the spectrum (0.4 m – 0.7 m), plants absorb light in the blue (0.45 m) and red (0.6 m) regions and reflect relatively more in the green portion due to the presence of chlorophyll. High photosynthetic activity will result in lower reflectance in the red region and high reflectance in infrared region of the spectrum. In, cases where plants are subjected to moisture stress or other conditions that hinder growth, chlorophyll production will decrease. This in turn leads to less absorption in the blue and red bands. In the near-infrared portion of the spectrum (0.7 – 2.5 m), green plants reflectance increases to 40 – 60%. Beyond 1.3 m, there are dips in the reflectance curve due to absorption by water in the leaves, more free water result in less reflectance (George and Hanuschak, 2010).

Chlorophyll does not absorb all wave length of sunlight. It absorbs blue and the red wave lengths while green light is reflected. The reflection of visible radiation is mainly a function of leaf pigments, whereas the near infrared (NIR) is reflected by the internal mesophyll structure of the leaf. As the leaves dry out or as the plant ripens or senescence or become diseased or cells die, there is reduction in the chlorophyll pigments. This results in general increase in reflectance in the visible spectrum and a reduction in a reflectance in the middle infrared (MIR) portion of the spectrum due to cell deterioration. Thus, the spectral response of a crop canopy is influenced by the plant health, percentage of groundcover, growth stage, stress condition and canopy structure. These reflected wave lengths may be detected by a sensor positioned above the crop. As a result of the above mechanism, healthy vegetation will show high value of reflectance in the NIR and low values in the visible spectrum. In the visible region, leaf reflectance is lower than soil reflectance whereas in the NIR leaf reflectance is higher than soil reflectance. This behavior is useful for explaining the utility of these reflectance measurements in agricultural applications and for the separation of crops from soil (Atzberger, 2013).

2.1.1 Satellite imagery for agricultural application

Spectral reflectance properties were used to monitor green vegetation. At present, a large range of satellite sensors provide regular data covering a wide spectral range (from visible to micro wave) and these data are acquired from various orbits and in deferent spatial, spectral and temporal resolution (Alzberger, 2013).

2.1.1.1 High spatial resolution satellite imagery

High spatial resolution satellite imagery was in use for analyzing and identifying crop variability in the mid of 1970's due to availability of landsat MSS/TM, which is characterized by high spatial resolution and relatively low temporal resolution. Hence their revisit time is long. Today, imagery of high spatial and radiometric fidelity is available throughout the world with the launch of IKONOS satellite which is commercial with high performance. This imagery is ideal for cropland analysis. Information derived and analyzed image can be seamlessly brought in to GIS as part of the crop management decision support structure (ESRI, 2013).

A combination of high-resolution imagery with microclimate weather data powerful agro economic modeling and assessment tools can be developed for improved field and sub field management. High resolution imageries can be used to progressively reduce the limitation created by mixed nature of low-resolution pixels. However, the low-resolution sensors have better synoptic view, large swath width and temporal revisit frequency than high resolution sensors (Rembold *et al*., 2013).

2.1.1.2 Low spatial resolution satellite imagery

Low resolution satellite imagery has been extensively used for crop monitoring and yield forecasting for over 30 years and plays an important role in a growing number of operational systems. The combination of their high temporal frequency with extended geographical coverage generally associate with low costs per area unit makes these images a convenient choice at both national and regional scales. Several qualitative and quantitative approaches can be clearly distinguished, going from the use of low resolution satellite imagery as the main predictor of final crop yield to complex crop growth models where remote sensing-derived indicators play different roles, depending on the nature of the model and on the availability of data measured on the ground. Vegetation performance anomaly detection with this resolution images continues to be a fundamental component of early warning and drought monitoring systems at the regional scale. For applications at more detailed scales, the limitations created by the mixed nature of low resolution pixels are being progressively reduced by the higher resolution offered by new sensors, while the continuity of existing systems remains crucial for ensuring the availability of long time series as needed by the majority of the yield prediction methods used today (Rembold *et al*., 2013).

Low resolution satellite images essentially refers to optical sensors in the reflective domain (*i.e.*, from the visible to the short-wave infrared: 400–2,500 nm) and with a spatial resolution between 250 m and several kilometers. Most of the early studies showed the use of different sensors of the NOAA and AVHRR series for agricultural analysis. These images are typically available at the national and multinational level with a 1-km resolution (LAC or Local Area Coverage) and, at the continental and global level, with a 4.6-km resolution (GAC or GLOBAL Area Coverage) or below. It was only at the end of the 1990s that the French–Belgian–Swedish satellite (SPOT) was equipped with a 1-km resolution sensor for vegetation monitoring at the global scale called VEGETATION. In addition, several so-called medium resolution sensors (maximum 250 m) have become operational since the year 2000, amongst the best known are the MODIS and MERIS sensors belonging to the TERRA/AQUA and ENVISAT platforms, respectively. All the low and medium resolution sensors that have proven their validity for land surface observation and vegetation analysis normally also find their applications in agriculture (Rembold *et al*., 2013).

2.1.2 Remote sensing derived variables for agricultural application

According to FAO (2010), for deriving information from RS data for agricultural application, a large number of spectral analysis and the spatial arrangement of the pixels, *i.e.*, the texture of the image, even at coarse resolution have been developed. Besides the spectral signature, useful information can also be retrieved by analyzing the temporal signature properties of vegetation. It is worthwhile to investigate the relationship between different indices and yield from different literatures to apply for this research.

2.1.2.1 Vegetation index (VI)

Vegetation Indices (VIs) are mathematical combinations or ratios of mainly red, green and infrared spectral bands designed to find the functional relationships between crop characteristics and RS observations. Vegetation indices are strongly modulated by the interaction of solar radiation with plant photosynthesis and thus are indicative of the dynamics of biophysical properties related to the plant status. But, during the early crop development stages, the effects of soil reflectance influence the values of some vegetation indices for the detection of crop stress. The most commonly used vegetation index include ratio vegetation index (RVI) and Normalized difference vegetation index (NDVI) (FAO, 2010).

Vegetation Indices (VIs) have been developed that have specific features concerning the range of vegetation cover. Vegetation Indices have been used since the Landsat satellite became operational. Vegetation indices provide information on the state of vegetation on the land surface. Vegetation is the result of a complex relation between the land and land use, and provides a means of monitoring and estimating changes over time.

This is the simplest form of ratio-based vegetation indices calculated through the use of infrared and the Red band of the electromagnetic spectrum. The VI seems to be more affected by the noise present in the image due to the atmospheric condition (Sawasawa, 2003).

Normalized Difference Vegetation Index (NDVI) is another form of ratio based vegetation index but the difference between IR and Red band as a ratio of the summation of the two bands. It is computed as follows:

NDVI= IR-R/IR+R

Out of all VIs, NDVI stands out and is regarded as an all-purpose index. This vegetation index is the most widely used and well understood vegetation index. Through this normalization, the values are scaled between –1 and +1 (Sawasawa, 2003).

2.1.2.2 Water Requirement Satisfaction Index (WRSI)

Water Requirement Satisfaction index (WRSI) is an indicator of crop performance based on the availability of water to the crop during the growing season. This index is a geospatial model developed by Food and Agricultural Organization for use with satellite data to monitor water supply and demand for rainfed crop throughout the growing season. It is also a crop performance index based on the availability of water in the soil.

WRSI = (AET / WR) 100 where, WR was calculated from the Penman- Monteith reference crop evapotranspiration (ETo) using the crop coefficient (Kc) to adjust for the growth stage of the crop: WR = ETo (Kc) (Legesse and Suryabhagavan, 2014). Currently, crop moisture stress on crops can be monitored using WRSI, which is a satellite based crop performance index. This index indicates the extent to which water requirement of the crop has been satisfied in the growing season (Tewlde, 2012). Technically, WRSI is the ratio of actual seasonal crop evapotranspiration (ETac) to the seasonal crop water requirement, which is the same as the potential crop evapotranspiration (PETc).

Originally developed by FAO, WRSI has been adapted and extended by USGS in a geospatial application to support FEWS NET monitoring requirements. As a monitoring tool, the crop performance indicator can be assessed at the end of every 10-day period during the growing season (Geo WRSI v2.0 manual).

Crop water balance can be estimated based on the onset and termination of the rain season, where the onset of rain is recognized at least 25 mm of rainfall in first decade followed by a minimum of 20 mm rainfall for two consecutive decades (Senay and Verdin, 2003). WRSI is a geospatial model developed by the Food and Agricultural Organization (FAO) for use with satellite data to monitor water balance for rainfed crop throughout the growing period (Tewelde, 2012).

2.1.2.3 Weather variables

Weather variables affecting crop yield are rainfall, temperature and solar radiation. These variables are in extricable by interlinked and is often difficult to observe individual effects in the field.

Rainfall is one of the elements of weather, which is an indicator of crop and weather relationship. Even if rainfall is generally beneficial, high intensity of it can negatively affect a crop through erosive runoff. When a plant is water stressed, it will affect by limitation of transpiration through restricting photosynthesis, resulting in less growth and yield (Osborne, 2004).

Timely and accurate rainfall estimation is of great importance when forecasting crop yields and real time rainfall observations. Rain gauge networks have traditionally provided a simple and in expensive method for daily and dekedal rainfall estimation. In recent years, these have been complemented by the development of precipitation radar networks, satellite rainfall estimates and output from numerical weather prediction models, which have been particularly successful in increasing the temporal and spatial resolution of the estimates (Novella and Thiaw, 2012).

To monitor drought and rainfall of certain area, specifically rainfall estimate (RFE) was carried out. The panchromatic and infrared sensors plus daily rainfall observation algorithms were used to produce rainfall estimate at a scale of 0.1^{o}. It is a combination of one rain gauge rainfall data and three satellite rainfall data set inputs. The inputs are GOES precipitation Index which is 4 km and half hours of spatial and temporal resolution; respectively, special sensor micro wave imagery with 15 km spatial and 4 times per day (6 hours) temporal resolution and advanced micro wave sounding unit is also 37 km spatial and 5 day temporal resolution of satellite derived rainfall data sets and Global Telecommunication Station. Global Telecommunication Station is a station rainfall data with un even spatial and from minute to daily temporal resolutions rainfall data sets (Novella and Thiae, 2012).

Another most important indicator of the relationship between crop and weather is solar radiation. In tropics, if a crop is not water limited, yield will be higher in cloudless season than in the wet season. As radiation is directly used for photosynthesis. In addition plants are often sensitive to heat stress during certain development stage. The effect of temperature is primarily day the development stage (Greatrex, 2012).

A comparison and validation results of all of the satellite products at a 10 day time scale over the Sahel has revealed that the regionally calibrated TAMSAT and RFE 2.0 had higher skill (Jobard *et al*., 2011).

2.2 Crop yield forecast in the world

According to (Sawasawa, 2003), the primary goal of yield forecasting in certain area is to give scientific, precise, sound and independent forecast of crop yield as early as possible by considering the effect of weather and climate during the growing season. Forecasts are made before the entire crop has been harvested where as estimates are made after the crop has been fully harvested.

There are different approaches of yield forecasting. The first one is the method of crop status evaluation by experts, which is the conventional method. This method is often subjective, costly, time consuming and are prone to large errors due to incomplete ground observation, which leads poor crop yield assessment and crop area estimation. As observations and measurements are made throughout the growing season such as tiller number, spikelet number and their fertility percentage of damage from fungi and pests etc, the data gained in these method can be used for yield forecast using regression method (Reynolds *et al*., 2000).The second method is RS and crop simulation models. In this method, RS images provide spatial information at global scale and real features and phenomena on earth. This method has potential of both estimating crop yield and identifying crop classes (Greatrex, 2012).

According to (Becker-Reshef *et al*., 2010), in the early of 1970s landsat 1 system started with the development of satellite to monitor agriculture activity due to the occurrence of unanticipated sever wheat shortage in Russia to drew attention of the function of timely and accurate prediction of world food supplies. The United States Development Aid together with NASA and NOAA initiated the large area inventory experiment. The aim of the experiment is to improve domestic and international crop forecasting approach.

In the early of 1980s, the program of (AGRISTAS) Agriculture and Resource Inventory Surveys through Aerospace Remote Sensing and (NOAA-AVHRR) sensor Advanced Very High Resolution Radiometer, allowing daily global monitoring where initiated. Recently, NASA and USDA Foreign Agriculture Service (FAS) have initiated the Global Agricultural Monitoring Service (GLAM) project. However, there are also currently other regional to global operational agricultural monitoring systems providing critical agricultural information like USAID Famine Early Warning System, and UNFAO Global Information early Warning System. At global level, crop production forecasts regularly, timely and objectively provided by combination of USDA FAS with GLAM system (Atzberger, 2013). When these modes are applied at regional level they cannot fully simulate the different crop growing conditions within the region. Such problems are commonly solved by integration of agro meteorological model with RS (George and Hanuschak, 2010).

The AVHRR satellite derived NDVI data were used for wheat yield forecast in a region in Italy and derived a simple linear regression model for wheat yield estimates and forecast based on NDVI images during the grain season. The result was validated with the official data by showing a good correlation between the two (Benedetti and Rossini, 1993). Similarly, the NDVI data derived from the AVHRR platform was used to estimate cereal production in Mediterranean African countries and found a good result too (Rembold and Maseli, 2004). However, it has been argued that RS might not be suitable in developing countries due to their stratified agricultural system and very small farm size. Meanwhile the increased availability of high spatial resolution makes this technique possible and as an alternative for yield forecast.

According to Rembold *et al.* (2013), yield forecast can be obtained by the type of crop grown and use of NDVI data at specific period, which depends on the climatic conditions of the area. The limitation of yield/NDVI regression for the aforementioned studies, was linked to environmental characteristics of specific geographic areas or due to availability of large and homogenous datasets of low resolution data and it is difficult in extending locally calibrated forecasting methods to other scale or other areas.

Where the crop area is not known, the NDVI /yield relationship does not provide information on final crop production and also this relation makes use of under specific conditions such as stable crop area over the observed period. In many cases, the predictive power of remotely sensed indicators can be improved by adding independent meteorological (bio climatic) variables in the regression model. Several bioclimatic and RS based indicators have proven to be highly correlated with yield for certain crops in specific areas. These variables can be either measured directly (like rainfall coming from synoptic station) or by satellites (such as rainfall estimates) or can be the result of other models like ETa (actual evapotranspiration) or soil moisture (Rembold *et al.*, 2013).

Rainfed based crop performance can be assessed using WRSI. The bioclimatic variables introduce information about solar radiation, temperature, air humidity and soil water availability, while spectral component introduce information about crop management, varieties and stresses not taken in to consideration by the agro meteorological models. Such hybrid models show higher correlation and prediction capability than the models using RS indicators only. For crop forecasting, satellite derived point specific rainfall estimates were input in to a crop water balance model to calculate WRSI. When WRSI values were regressed with historical yield data, relatively high skill yield forecasts can be made even when the crops are at their early stages of growth (Sawasawa, 2003; Senay and Verdin, 2003).

It is possible to conduct operational maize yield forecasts using Crop Normalized Difference Vegetation Index (CNDVI) derived from SPOT VEGETATION and ETa from the FAO CSWB model. Crop Normalized Difference Vegetation Index CNDVI showed to improve the spectral signal of maize crop areas when compared with the simple spatially averaged NDVI using the general crop mask. CNDVI proved to be a simple and valid method for NDVI extraction with low resolution satellite images and highly fragmented high resolution land-cover classes. Due to this prediction capability it is possible to obtain an early forecast using the CNDVI and ETa accumulated from planting decade to the end of the flowering phonological phase. A more accurate estimate will be possible when the maize crop cycle reaches the end using the CNDVI and ETA accumulated for the whole length of the maize crop cycle. It is possible to have reliable predictions 3−4 months earlier than the official estimates provided by national authorities and based on traditional field sampling surveys. As the time-series of the yield data was limited, some reservations for the model must be made, until a longer series of yield data will be available. The simplicity of the proposed regression yield model should allow an operational implementation in developing countries (Rojas, 2006).

2.3 Crop yield forecast in Ethiopia

In Ethiopia currently use highly qualitative methods to estimate crop yields within a season, thus it can be difficult to identify food shortages until well after harvest. The ability to forecast crop yield would be of enormous benefit to decision makers as it could allow a country to shift its food needs assessment to earlier in the growing season. This would allow more time for intervention by external organizations in the event of food shortages, or for governments to plan in the event of a food surplus (Haile, 2005).

Ethiopia is following two methods of monitoring and forecasting crop yield in advance of harvest. The first is an operational Crop Yield Monitoring and Forecasting System (CYMFS), run by the Ethiopian National Meteorological Agency (NMA) in conjunction with the European Union Joint Research Council (JRC) and the Food and Agricultural Organization (FAO). The system is discussed in detail in Rojas *et al.* (2005), but in short, it combines a Geographical Information System (GIS), the FAO Crop-Specific Water Balance (CSWB) model, the JRC Crop Production System Zones database (CPSZ) and meteorological information from numerical models provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). Additional and independent real-time Normalized Difference Vegetation Index (NDVI) satellite data from SPOT VEGETATION is also incorporated, using a specific crop mask to concentrate the analysis only on agricultural areas. This crop mask was derived from a static CPSZ database which itself was created using survey data, therefore the NDVI data was simply used to look at plant conditions throughout the season. The technique performed well in trials and is easily applied over Ethiopia (Rojas *et al.*, 2005), but has several potential shortcomings. The first is its reliance on an empirical CSWB model rather than a processed based crop simulation model. As discussed in Challinor *et al.* (2004), empirical crop simulation models are less likely to capture complex non-linear interactions between crop and climate. Teo (2006) also suggested that a CSWB model performed less than a process based model when forecasting groundnut yield in Gambia. This was because the CSWB model was unable to capture yield loss associated with short duration dry spells that occurred during flowering.

The second method used by the Ethiopian authorities to monitor crop yield is carried out by CSA. At the time of harvest for either cropping season, the CSA will invite key stakeholders such as farmers unions, NGOs and external organizations to a meeting to discuss how the season has developed. These stakeholders are asked to agree a percentage change in perceived crop yield from the previous season. For example, it might be decided that the crop yield in Tigray during 2009 is 80% of the crop yield in Tigray during 2008. This number is then multiplied by the previous year’s yield to give a pre-harvest estimate (Beyeyne, 2009).

These estimates can be highly subjective and dependent on the agenda of stakeholders, but are widely used by Ethiopian decision makers. According to Hoefsloot (2008 as cited in Tewlde yideg, 2012), another software called Livelihood Early Assessment Protection (LEAP) developed by World Food Program (WFP) has also used in Ethiopia for crop yield forecast in addition to the two mainly known approaches method earlier. In Ethiopia, crop yields are predicted by the amount of available water as compared to the crop water requirement for the growing season using LEAP software developed specifically for Ethiopian context. One of the goals of LEAP is to serve as a platform for calculation of weather based indices starting out with the calculation of a crop water balance indicator, WRSI. In addition, it uses relevant soil information from FAO digital soil map and topographical parameters derived from the GTOPO30 Digital Elevation Model (DEM).

Greatrex (2012) carried out research on Ethiopia using the crop simulation model called the General Large Area Model for annual crops (GLAM). General Large Area Model is a process based model designed to simulate tropical crop production in regions where there is an observed relationship between climate and crop yield and has been shown in studies to capture the crop/weather relationship at large scales and found that this model was shown to exhibit the correct sensitivity to climate and to perform reasonably when compared with observed crop yields. The limitations of the approach are extremely complex with many different varieties of crops used by farmers to adapt to their climate. Thus, the use of a single cultivar which leads to unrealistic results in some high altitude locations and the time a plant takes to develop to maturity and strongly dependent on growing season temperature. In the study, GLAMWHEAT was run using a single lowland wheat cultivar that needs relatively high temperatures in order to develop. This means that for high altitude, low temperature pixels, GLAMWHEAT took an excessively and unrealistically long time to develop and results in an unrealistic pattern. Lastly he recommended that an operational crop yield forecasting system needs to be able to work at a regional scale if it is to be of use to policy makers in order to substitute the existing conventional method which was proofed as subjective, time consuming, un reliable and have also a problem of timeliness.

2.4 Wheat crop

Flour milling takes the major share of wheat processing in which the main product is sold to grain based food manufacturers and some of them are sold to ingredient distributors. Apart from milling wheat in to flour, the crop product can be used to produce livestock feed, starch and ethanol. Another potential use of wheat bran is in bio-ethanol production as an alternative to fossil fuel that is facing depletion. Bran of wheat can be used to supplement wheat flour and baked products with cheap vitamin and other nutrients. Further, wheat bran is abel used to make livestock feed formulations and to improve milk yield (Chandiposha *et al*., 2013).

Use of wheat consumption is gradually increasing in Ethiopia, particularly in urban areas due to high population growth and changes in life styles. In most parts of the country, families prefer to use teff to make injera and sometimes to make porridge. Teff straw is an important source of animal fodder and has been shown to be more nutritious as animal feed than other grain by-products. Because of the price escalation of teff compared to wheat, most middle and lower class families are shifting to wheat consumption. Ethiopia is the second largest wheat producing country in Africa next to South Africa. Wheat is mainly grown in the central and south eastern highlands of the country during the longer rainy season (June to September) and harvested in October-November. Arsi, Bale, and parts of Shoa are considered the wheat growing belt of the country. Wheat is produced in large state-owned farms covering around 124,000 ha of land in Arsi and Bale zones. The Federal Government of Ethiopia is in the process to privatizing these farms. The remaining 92% (1,390,000 ha) of production of wheat is from small farms (Abu Tefera, 2013).

2.5 Agricultural periods in Ethiopia

Ethiopia is one of the most climatically complex countries in Africa. This is in part due to its topography as well as due to the country’s placement with respect to large scale weather patterns. There is a large amount of spatio-temporal complexity in Ethiopia’s seasonal rainfall cycle, which is in general defined by the progression of the Inter-Tropical Convergence Zone (ITCZ). The North and Mid-West of Ethiopia have a single long rainy season (Meher*)* with its maximum in late July. The rest of the country has a bimodal seasonal distribution made up of the short rainy season (Belg) and longest rainy season (Meher*)*, (MOA, 2007).

2.5.1 Shortest rainy season

This occurs primarily during February, March and April to correspond with the Belg rains. It generally consists of short season crops and out of the six staple food crops, short cycle maize is considered the most important. Although, the season is considered important as a hunger breaker (DPPC, 2003), the Belg season accounts for only 3% of the total production (Taffesse *et al.*, 2009). Due to this and as it is also difficult to estimate Ethiopian rainfall during January and February; it was decided to only concentrate on the Meher season.

2.5.2 Longest rainy season

This is the main agricultural season in Ethiopia and is dominated by teff, wheat and maize grown during the summer or rainy season. The wheat crop includes a mix of long and short cycle crops. Funk *et al*. (2005) describes the system in more detail: Meher crops are typically harvested in September or October. Short cycle Wheat and Teff are typically planted between June and July. Long cycle crops are planted during the Belg season and harvested following the Meher in late fall. Given sufficient agricultural inputs, they are often substantially (1.5−2.5 times) more productive than short cycle varieties planted during the Meher season. Long cycle crops contribute approximately 50% of national production, compared to 40−44% for 'short-cycle' (June−September) varieties. Unfortunately, the agricultural data provided for this thesis does not distinguish between short cycle and long cycle crops during the Meher season, but the existence of the two cultivars should be born in mind during the analysis. Figure 2.1 shows Seasonal calendars of crops.

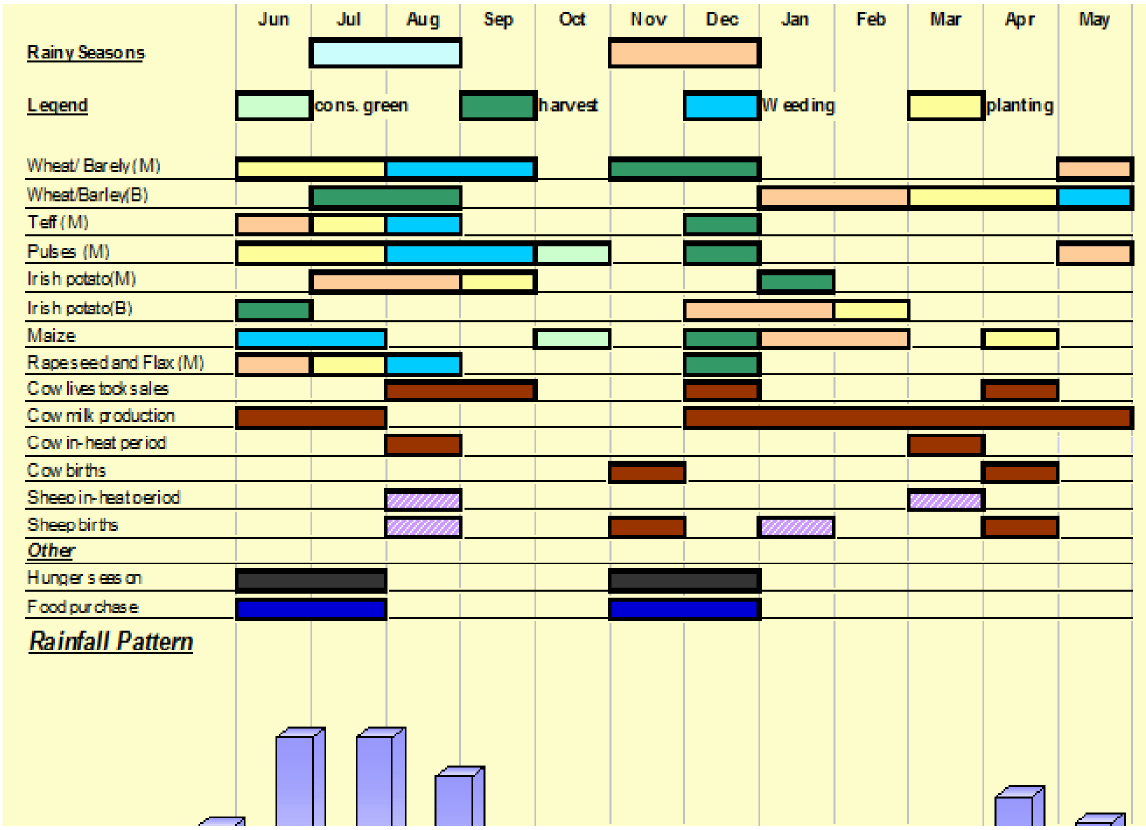


Figure 2.1 Seasonal calendars of crops in Ethiopia.
(Source: Livelihood profile for Arsi, 2007).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Description of the study area

East Arsi Zone is located in the central part of the Oromiya National Regional State. The zone bounded by latitude 6° 45'−8° 58' 00"N and longitude 38° 32'−40° 50' 00"E covering a total area of 23881 km² (Fig 3.1), which accounts for about 7 percent of the total area of the Oromiya Regional State of Ethiopia.

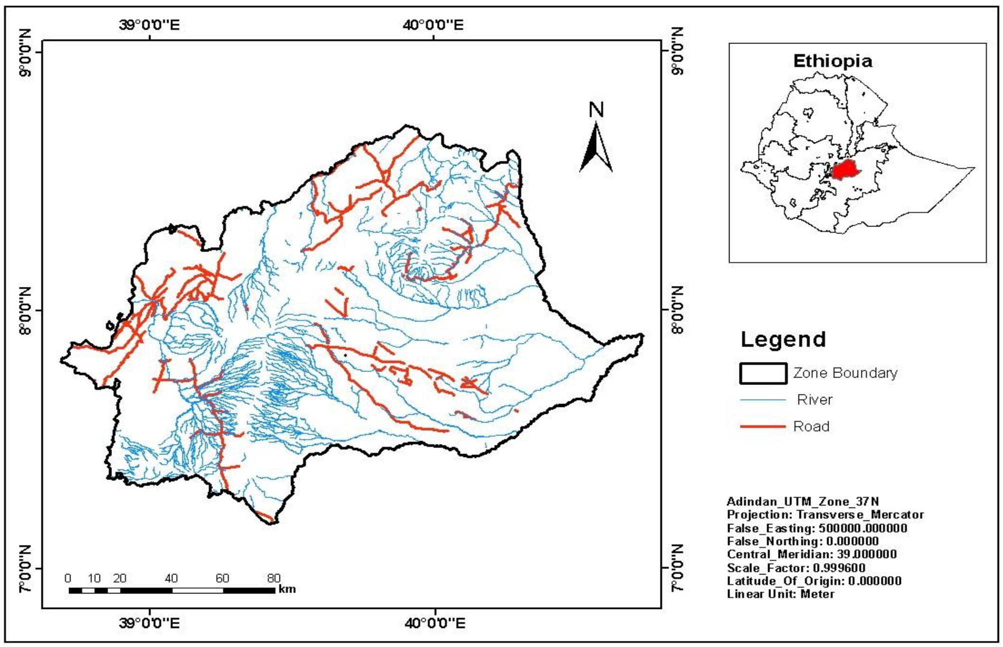


Figure 3.1 Location map of the study area

3.1.1 Population

Based on the 2007 Census of Ethiopia, East Arsi Zone has a total population of 2,637,657, of which 1,323,424 are men and 1,314,233 women, central statistical agency (CSA, 2007). According to 2014 projection the total population of the zone is 3,202,689 with 1,601,757 men and 1,600,932 women (http://www.csa.gov.et/index.php/2014-02-20-13-43-35/national-statistics-abstract/141-population)

3.1.2 Temperature

East Arsi Zone has five agro-climatically zones, mainly due to variations in the altitude. The variation in temperature provides wide opportunities for the production of different types of crops. The mean annual temperature of the Zone is between 20−25°C in the low land and 10−15°C in the central high land (Fig 3.2). However, there is a slight variation in temperature by months. February to May is the hottest months, while October to January is the coldest months in the study area.

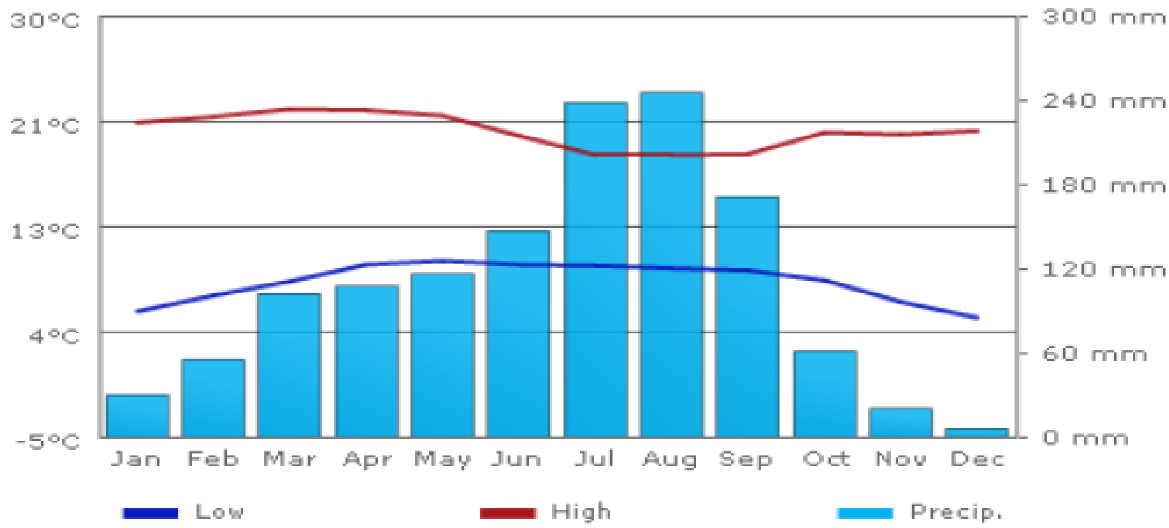


Figure 3.2 Monthly temperature *vs.* rainfall distribution of the study area.

3.1.3 Rainfall

The rainy season of the East Arsi Zone starts in March and extends in October with the highest concentration in June, July and August. The mean annual rainfall varies from 633.7 mm at Dera station (located at an altitude of 1680 m above sea level) to 1059.3 mm at Bekoji station (located at an altitude of 2760 m above sea level). Generally, East Arsi Zone receives abundant and well-distributed rainfall both in amount and season, which is a conducive for different type of vegetation growth and agricultural activities. On average, the Zone gets a monthly mean rainfall of 85mm and an annual mean rainfall of 1020 mm.

(http://oromiabofed.org/index.php?view=article&catid=44:home&id=48:establishm)

3.1.4 Cropping condition (Pattern)

The planting or sowing time of different crops varies depending on the onset and continuity of the rainfall. There are two distinctly known (Bi-modal) and traditionally used cropping seasons. Short cropping season is the one, which starts as soon as the last harvest of the previous long rainy season crop is over. Successful short rainy season crops are meant to leave the land for the second crop season and hence will be harvested around May, allowing enough time for land preparation and sowing of the long rainy season crops. The second cropping season is the long rainy season as it is practiced in most parts of the country and the study area too. As shown in Table 3.1, cereal crops grown in the study area include teff, wheat, maize, barley and sorghum (Dagnew Belay, 2007).

Table 3.1 Cereal crops of area, production and yield for season (2013/14) of the study area.

Cereal Crop	Area in Hectare	Production in quintal	Yield (quintal /hectare)
Teff	114,578.57	1,723,016.47	15.04
Barley	91,616.90	2,440,862.00	26.64
Wheat	192,232.38	5,823,929.72	30.30
Maize	69,916.08	1,814,061.50	25.95
Sorghum	34,308.19	903,760.87	26.34

(Source: CSA, 2014, Agricultural report).

3.1.5 Soil

The major soil types in Arsi Zone are: Chromic and Pellic Vertisols (30%) characteristics of water holding and heaviness for plowing during rainy seasons due to high clay content (Fig 3.3). Luvisols (13%) is good for agriculture with base saturation and weather able minerals and dominant on the high land parts of the Zone. Cambisols (23%) dominantly occur on the steep slopes and are often shallow or have many rock outcrops and those developed on gentler slopes, however, have good base saturation and fertility and can highly be used for agriculture. Other types of soils of the zone are Andosols (4%) having over 60% volcanic ash, or other vertic pyroclastic materials in silt, sand and gravel fraction occur in parts of the zone extending in coverage from lake Koka to lake Langano. Lithosols (6%) is another group of soil found in the Zone, has high base saturation and hence fertile. Fluvisols, which constitutes about 2% of the total soil coverage in the zone, is restricted to the lowland parts of Gololcha, Merti and Ziway Dugda districts of the Zone. Phaeozems, Orthic Solonchak, Calcic xerosols, Eutric regosols, Gypsic Yermosols, Mollic Gleysols and Orthic Acrisols are other soil groups found in the Zone, which together constitute about 11% of the total soils coverage with dominant of mollic Gleysols (4%). Generally, the fertility status of the soil of Arsi Zone is good and conducive for crop production.

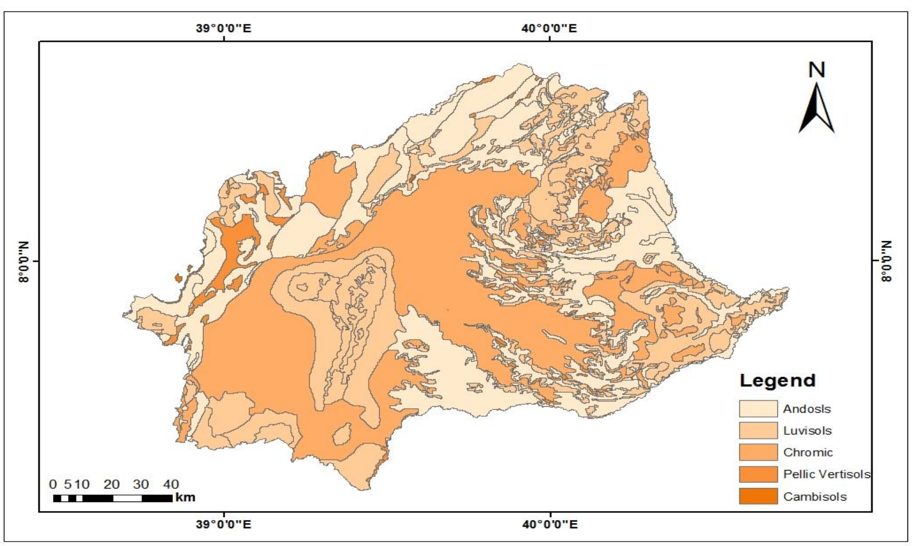


Figure 3.3 Soil map of the study area.

3.2 Source of data and software's used.

The data used in this study were collected both from primary and secondary sources. Primary data comprised of information captured from satellite imagery and field observations. Secondary data sources include published and unpublished materials such as books, topographic and thematic layers, journals, reports of Meteorological Agency and Central Statistical Agency as well as other publications and scientific works. To analysis these data sets, different softwares were also used.

3.2.1 Satellite imagery, available models and ancillary data

Different satellite imageries including models were collected from different sources to be used for the model building. Among them, rainfall estimates, SPOT VEGETATION, WRSI and pan sharpened SPOT 5 imagery were the major ones. These sources are dealt in detail as follows:

3.2.1.1 Rainfall estimate (RFE 2.0)

According to Helen (2012) rain gauge, radar and numerical weather prediction (NWP) networks are inadequate and insufficient for use in crop yield forecasting in Ethiopia. Therefore, the forecast should be based on satellite rainfall estimation. Washington *et al*. (2006) described that satellite rainfall data have the potential to be used as inputs to crop yield prediction models due to the good spatial coverage and high temporal resolution of the satellite information. This is particularly relevant for Africa where the ground based observational network is often sparse and not always well maintained.

Based on these findings, metrological satellites were used rather than synoptic information to derive meteorological data. Accordingly, the rainfall data used in this study were derived from RFE, which is a rainfall estimate product of NOAA's climate prediction. There exists two RFE versions (RFE 1.0 and RFE 2.0) produced with different methodologies. RFE 1.0 uses an interpolation method to combine Meteosat and Global Telecommunication system data and it is available for the period 1995−2000. RFE 2.0 uses additional techniques to better estimate precipitation and the data is available from 2001. Compared to other rainfall data like ECMWF, RFE shows a better estimation (Rijks *et al*., 2007). RFE 2.0 obtains the final daily rainfall estimation using a two part merging process. All satellite data were first combined using the maximum likelihood estimation method, and GTS station data were used to remove bias. This RFE 2.0 satellite rainfall estimate was used for the analysis of this research, was down loaded from http://earlywarning.usgs.gov/fews/africa/web/dwndailyrfe.php (Eerens et al., 2014).

3.2.1.2 SPOT VEGETATION (VGT)

The SPOT VEGETATION(VGT), launched in March 1998 on board of the SPOT 4 satellite, to monitor surface parameters with a frequency of about once a day on a global basis at a spatial resolution of 1 km was selected for NDVI derivation. Compared with the existing instruments of the same kind, in particular NOAA AVHRR, SPOT VEGETATION provides an enhanced radiometric resolution and above all a limited local distortion of about 0.3 km (Sawasawa, 2003; Rojas *et al*., 2005).

A high geometric accuracy is absolutely very important when a set of multi-temporal images is processed. This means that the same pixel, taken from images relative to different dates, must represent the same ground area, with a minimum shift. In this regard, SPOT-VGT images are much better (Rijks *et al*., 2007).

SPOT VEGETATION which is synthesized for Decadal (S10) images became regularly available from the first of January 2004 and can be used for NDVI analysis (2004−2013). The software called Spirit was used to analyze time series imagery. Analyzed using sprit soft ware done by combining status map with Flag VGT extract and the analysis was made based on that. Daily synthesis (S1) or ten-day synthesis (S10); were mosaics of acquired image segments, respectively for 24h periods and for the last 10 years. A Maximum Value Composite (MVC) synthesis can be delivered with several spatial resolutions. These three products are called S10 for 1 km data, S10.4 for 4 km^2 and S10.8 for a resolution of 8 km^2 (Eerens *et al*., 2014).

3.2.1.3 Water requirement satisfaction index

USGS/FEWSNET recently uses Geospatial WRSI crop model, which allows for localized crop modeling, monitoring and forecasting at sub national level, using locally available datasets as model inputs. Result of this model was also selected as one parameter in order to develop wheat forecast model. Water requirement satisfaction index for a season is based on the water supply and demand a crop experiences during a growing season.

Water requirement satisfaction index where calculated as the ratio of seasonal actual evapotranspiration (ETa) to the seasonal crop water requirement (WR): WRSI = (AET / WR) * 100. Water requirement were calculated from the Penman-Monteith potential evapotranspiration (PET) using the crop coefficient (Kc) to adjust for the growth stage of the crop: WR = PET * Kc. AET represents the actual (as opposed to the potential) amount of water withdrawn from the soil water reservoir ("bucket").

3.2.1.4 Pan sharpened spot 5 imagery

Pan sharpening is a technique that merges high resolution panchromatic data with medium resolution multi spectral data to create a multi spectral image with higher resolution features. Pan sharpening allows using panchromatic (pan) and multispectral (MS) images from different sources and sensors to produce pan sharpened color images (Leiliesand and Kiefer, 1998).

Data are usually representative of a range of ‘n’ bands and wavelength, such as visible or thermal infrared. It combines many colors so it is “pan” chromatic. A multi spectral image contains a higher degree of spectral resolution than a panchromatic image, while often a panchromatic image will have a higher spatial resolution than a multi spectral image. Therefore, this study used a pan sharpened image, which represented a sensor fusion between the multi spectral Landsat image and panchromatic image of SPOT, which gave the best of both image types (Leiliesand and Kiefer, 1994).

3.2.1.5 Ancillary dataset

There were other relevant data sets obtained from CSA, like shape file, developed for 2007 population and housing census cartography. These shape files were used to demarcate the study area boundary. Topo sheets from EMA were also used to check the geometric correction of the satellite imageries.

3.2.1.6 Official yield statistics

The production of quantitative yield estimates needs the calibration of the model with historical crop yield statistics (Rijks *et al*., 2007). Accordingly, historical grain yield data (2004−2013) available at the Zonal level were collected from CSA. The agricultural department of CSA provided the archive of wheat grain yield estimate (Table 3.2, Fig

3.4). The yield statistics were obtained from ground sample survey based on list frame approach.

Table 3.2 Trend of Wheat crop yield (2004–2013) in East Arsi Zone.

Year	Holders	Hectares	Production	QT/HA
2004	228,802.0	104,059.0	1,707,602.0	16.41
2005	268,753.0	149,391.0	2,987,505.0	20.00
2006	303,565.0	184,477.0	3,805,387.0	20.63
2007	325,234.0	190,525.6	3,330,402.9	17.48
2008	316,150.0	171,815.0	2,885,066.8	16.79
2009	259,204.0	124,896.1	2,053,569.7	16.44
2010	265,607.0	145,291.3	3,070,950.8	21.14
2011	302,897.0	199,852.9	4,385,513.4	21.94
2012	299,811.0	192,152.9	4,942,743.7	25.72
2013	332,347.0	209,218.9	5,115,026.7	24.45

(Source of data: CSA annual agricultural report (2014).

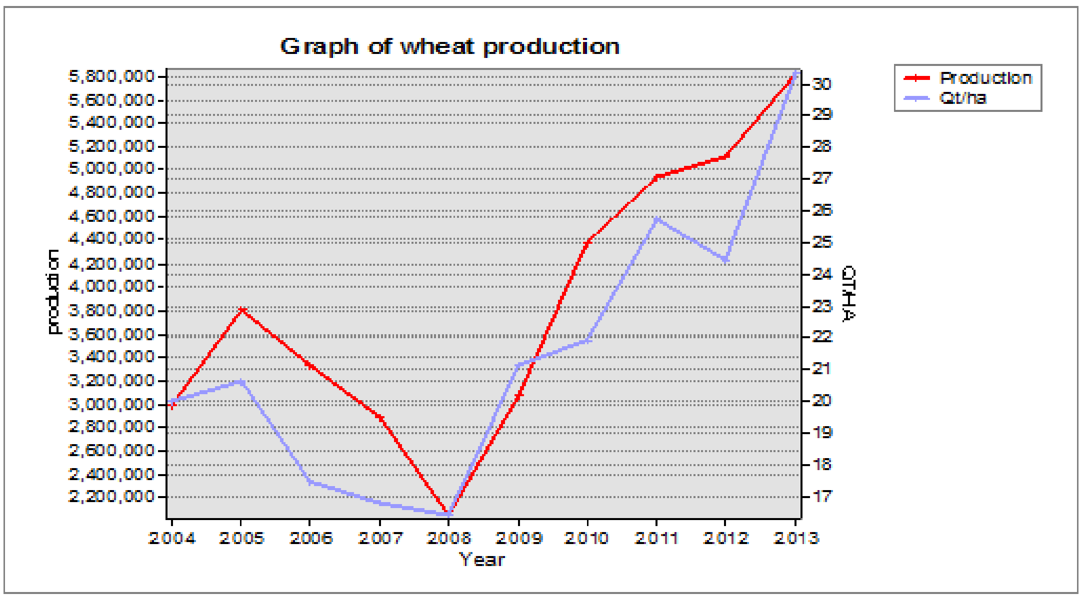


Figure 3.4 Trend of Wheat crop yield (2004–2013) in East Arsi Zone.
(Source of data: CSA annual agricultural report 2014).

3.2.1.7 Materials

Material includes equipments which are used to conduct the research. The equipments used for this research are summarized on (Table 3.3 and Fig 3.5).

Table 3.3 Summary of equipment and materials used for data collection and analysis.

No	Software used	Purpose
1	GPS(Global Positioning System)	For collecting of GCP points which will be created at random for the study area Using Arc GIS 10 Used mainly for accuracy assessment and Area Measurement GPS is from CSA
2	Erdas 10, ArcMap10,Spirit, MadCAT 3.3, VGT extract, LEAP 2.7.1, SPSS statistical tool, e- cognition 8.2	GIS & statistical software for image and vector processing and data analysis.
3	Google Earth	Used as supplementary for checking and correcting area of doubt about accuracy of the classification.

3.2.1.7.1 SPIRIT Software

Most of the available GIS and image processing software packages do not provide a complete range of built in functionalities to handle time series of images, and are not optimized for the needs of the crop monitoring community. They generally do not offer a flexible image processing environment that can be used by technicians and involved institutions to adapt the data analysis steps and generate additional and customized outputs. Therefore, we can conclude that none of the existing systems provides in one package the highly specific set of time series processing functions to assess crop and vegetation status, including temporal smoothing, computation of long term averages and anomalies, classification based on seasonal performance of vegetation, and production of the outputs traditionally used in crop monitoring bulletins (statistics, maps, graphs). For this reason, a flexible and user-friendly interface, targeting both national and international agriculture and food security experts, is highly desirable (Eerens *et al*., 2014).

SPIRITS is an integrated, modular software platform that aims at answering the requirements outlined above. This software is extensively documented and distributed freely for non-commercial use. SPIRITS works with 2D-flat binary image files. “2D-flat” means that the layers of a multi-spectral or multi-temporal set are stored as separate files. The ancillary information (“metadata”) must be present in an additional ASCII-file with the same name as the image but with extension “.hdr”. For the latter, the rules of the ENVI software

are followed, although new keywords have been added to provide more information on the image values, their scaling to physical units, the presence of flags (i.e. codes to label special features such as sea, clouds, snow), acquisition date and periodicity (at the moment only daily, ten daily, monthly and annual images are allowed) (Eerens *et al*., 2014).

3.2.1.7.2 **Statistical Package for Social Science *(*SPSS*)***

This package was first developed in the 1960s and was the first major attempt to provide software for the social scientist (Bryman and Cramer, 2005). The "Statistical Package for the Social Sciences" (SPSS) is a package of programs for manipulating, analyzing, and presenting data; the package is widely used in the social and behavioral sciences. Data analysis is only one part of the research process. Before using the analysis, the data obtained must be prepared for entry into SPSS.

- SPSS provides a powerful statistical analysis and data management system in a graphical environment based on the user interface facility.
- This program can be used to analyze data from surveys, tests observations, etc (Landau and Evertitt, 2003).
- It can perform a variety of data analyses and presentation functions, including statistical analysis and graphical presentation of data. Among its features some are listed below
- Descriptive statistics such as frequencies, central tendency, plots, charts, and lists; and sophisticated inferential and multivariate statistical procedures, such as analysis of variance (ANOVA), factor analysis, cluster analysis, and categorical data analysis.
- SPSS is particularly well-suited for survey research, though by no means is it limited to just this topic of exploration.

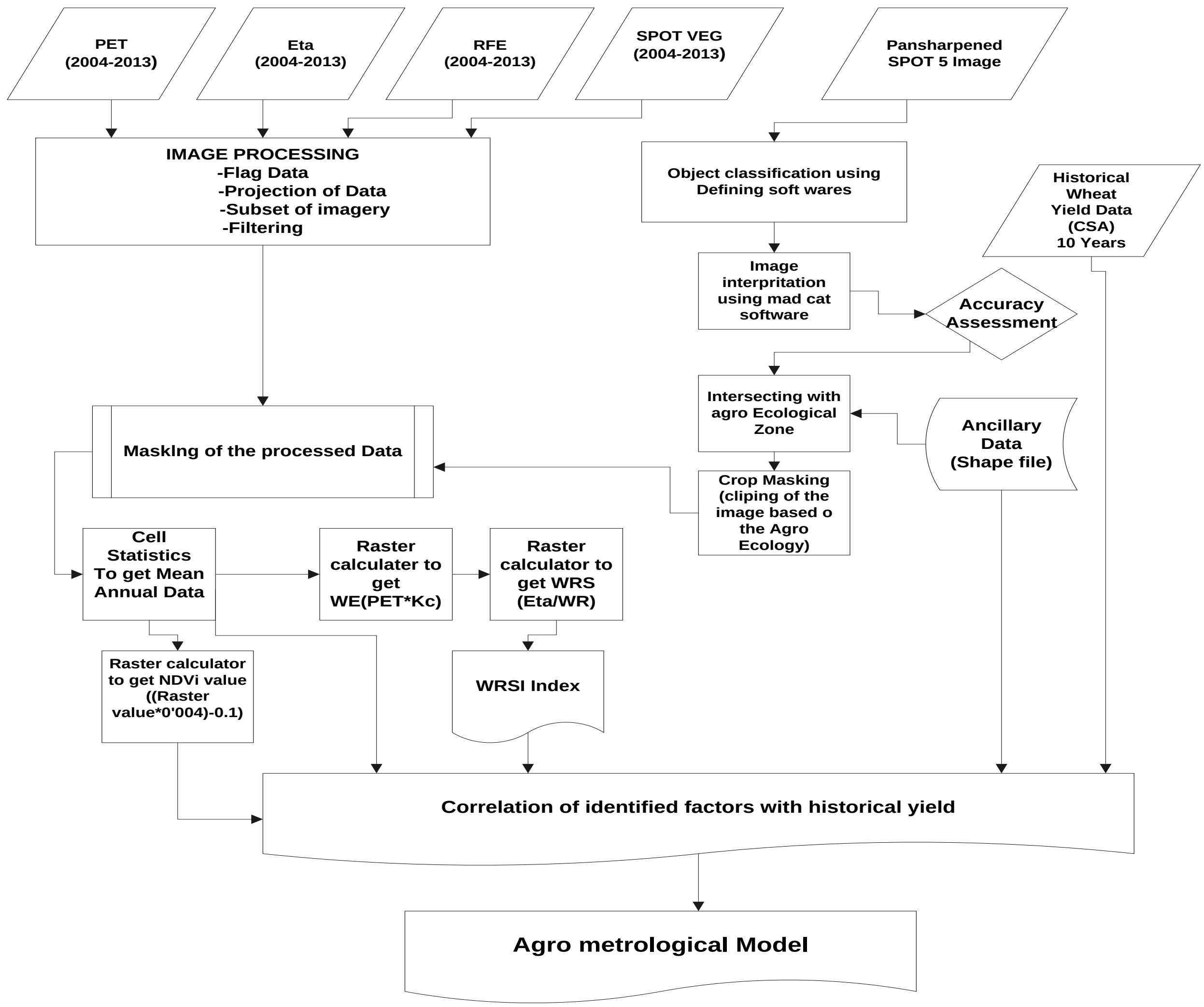


Figure 3.5 Methodological flow chart.

3.3 Data processing and analysis

There are two approaches of quantifying yield production using remote sensing. These are purely remote sensing approach (direct approach) and mixed approach, where additional bio-climatic predictor variables are used. In many cases, the predictive power of remotely sensed indicators can be improved by adding independent meteorological (bio climatic) variables in the regression models but in an area where there is stable crop area over the observed period and homogenously large area, a direct NDVI / production regression can be (1999, as cited in Rembold *et al.*, 2013) observed that the spatiotemporal rainfall distribution needs to be incorporated in to crop yield models, in addition to vegetation indices deduced from remote sensing data, to predict crop yield of different cereal crops grown in rainfed condition. Such hybrid models show higher correlation and predictive capabilities than the models using remote sensing indicators only as input variables. Therefore, due to these advantages for this research the mixed approach is used for wheat forecast.

In the spectro-agro meteorological approach, the spectral component introduces information about crop management, varieties and stresses while bio-climatic variables introduce information about solar radiation, temperature, air humidity and soil water availability (Rembold et al., 2013).

3.3.1 Selection of date of satellite imagery

The choice of the date of the image, was based on the analysis of the information given by the farmers on the date of planting and date of harvesting. The choice of date was performed in such a way that the image should coincide with the peak vegetation period of the farmers field (Rembold *et al*., 2013).

Accordingly, the study area planting date was found to be from the middle of June and this was cross checked with the Zone livelihood profile. This also states that June is planting date for wheat crop in the study area. Generally, wheat crop in East Arsi Zone is planted in June, growth in biomass occurs from July to August and flowering in September based on the information from interview with the local farmers.

Following this, SPOT-VGT NDVI Decadal images were downloaded from the website http://www.vito-eodata.be/PDF/portal/Application.html from the month of June up to September starting from 2004 to 2013 (ten years' time series data). After acquiring the data (which consists of several HDF layers joined in one ZIP file), the following steps were performed:

1. Extraction of the NDVI product and the so-called 'Status Map' using VGT Extract software. This software is freely available for download on www. agricab.info.

2. Applying the Status Map on the NDVI image in SPIRITS software using the 'Flag VGT NDVI' tool. After the above two processes were carried out, 120 decadal (S-10) images of SPOT VGT NDVI (Fig 3.6) were ready for further analysis with raster value ranging from 0 to 255.

In the same manner, RFE 2.0 satellite rainfall estimates were found in decadal at http://earlywarning.usgs.gov/fews/downloads/index.php? Region ID was downloaded from the month June up to September starting from 2004 to 2013 (ten years' time series data) with rainfall estimate ranging from 0 up to the maximum rainfall estimated (Fig 3.7).

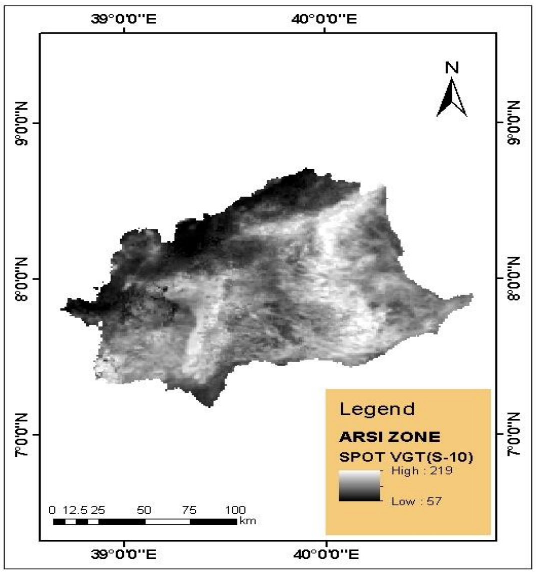


Figure 3.6 SPOT VGT image of Arsi, 1st Decade of June 2004.

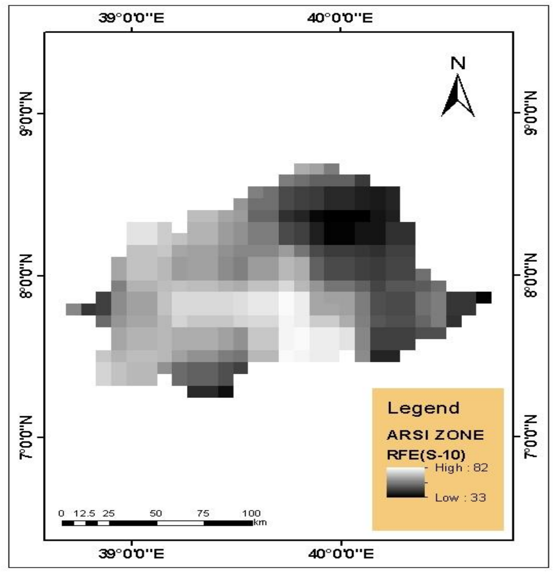
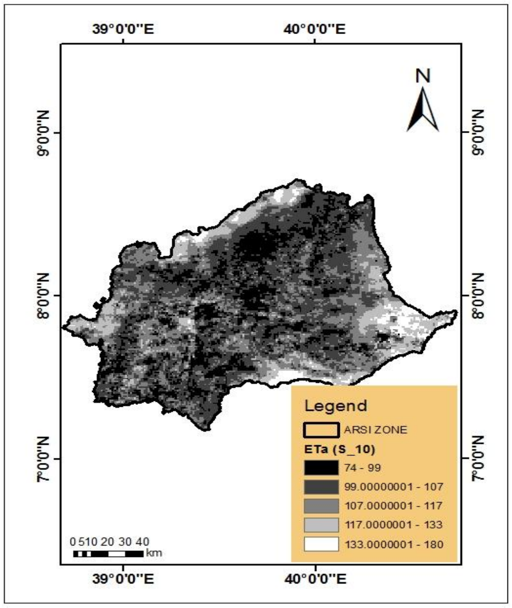


Figure 3.7 RFE 2.0 image of Arsi, Mean of 2004(June - September).

Actual ETa and PET were another input for the model computation, which were downloaded freely from EWSNET http://earlywarning.usgs.gov/fews/downloads/index.php? At monthly and annual level from June up to September of year 2004 up to 2013, respectively (Fig 3.8 and 3.9).



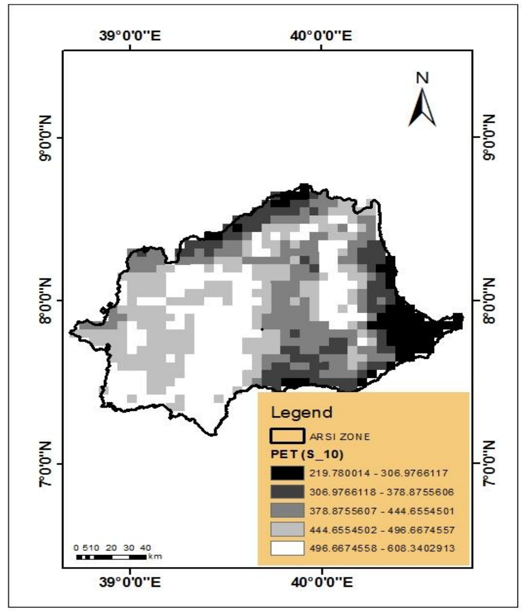


Figure 3.8 ETa imagery of September, 2010.   Figure 3.9 PET imagery (2004 mean annual).

3.3.2 Preprocessing of satellite images

A prerequisite for using time series of remote sensing data for agricultural application is atmospheric correction and geometric rectification of the dataset. In this study, the S-10 images were used. Atmospheric corrections for Zone were done on the images before they are delivered to users (Sawasawa, 2003). The products of SPOT VEGETATION acquired by MARS are 10-day NDVI synthesis (S10) images, obtained through Maximum Value Compositing (MVC). The images were corrected for radiometry, geometry and atmospheric effects and the same was carried out for RFE 2.0 results (Rojas *et al*., 2005). However, the satellite imageries I have acquired (spot 5, spot vegetation) were already georeferenced and radio metrically corrected. Therefore, other preprocesses were carried out such as projecting the layers, extracting region of interest and filtering.

3.3.2.1 Projection of the data

All data used in this study including satellite imageries of different source, topographic maps, thematic layers like road, towns were projected to the geographic coordinate system GCS WGS 1984 and datum of World Geodetic System 84 (WGS 84), ensuring consistency between datasets. Here topographic maps were used to check the geometric rectification of the satellite imageries.

3.3.2.2 Extracting region of interest

The first step, necessary for the preprocessing is extracting of all the available raster data with the shape file of the study area. This extraction of region of interest allows running any process in a relatively short time.

3.3.2.3 Filtering of the study area

In a filter operation, the values of each pixel is changed according to its value in the original image and the value of neighboring pixels. Some examples of applications when filtering is used are noise removal or for the visual enhancement of the image. The filter tool applies a user defined kernel to calculate new values for the central pixel using a mathematical operation on the original cell value and its neighbors. The software which I have used for the analysis of time series data (SPIRITS) includes a low pass filter tool which is designed to emphasize larger, homogenous areas of similar tone and reduce the smaller detail in an image (Eerens *et al*., 2014).

3.3.3 Wheat crop masking

One obstacle to successful modeling and prediction of crop yields using remotely sensed imagery is the identification of image masks. Image masking involves restricting an analysis to a subset of a crop land masking where all sufficiently cropped pixels are included in the mask, but the ideal approach would be to use crop specific masks. This would allow one to consider only information pertaining to the crop of interest (in this case wheat). However, this approach is not applicable to areas such as, East Arsi Zone where there is crop rotation. Therefore, crop masking at crop land area is acceptable for such areas (Rijks, 2007). For this reason, pan sharpened SPOT 5 imagery was used in order to carryout land cover classification using two classes (Agriculture and non agriculture) for the study area.

3.3.3.1 Image interpretation

Satellite images are like maps: they are full of useful and interesting information. They can show us how much a city has changed, how well our crops are growing, where a fire is burning, or when a storm is coming. To unlock the rich information in a satellite image, you need to, look for a scale, look for pattern, shapes, and texture defines the colors (including shadows). Find north, consider in your prior knowledge. These tips come from the Earth Observatory's writers and visualize who use them to interpret images daily. They will help you get oriented enough to pull valuable information out of satellite images (Holli

Riebeek, 2014). Pansharpened SPOT 5 image of 2006 was acquired from CSA and used for image classification using visual image interpretation (Fig 3.10). The pan sharpened SPOT 5 image is then processed in Defining software for object based classification.

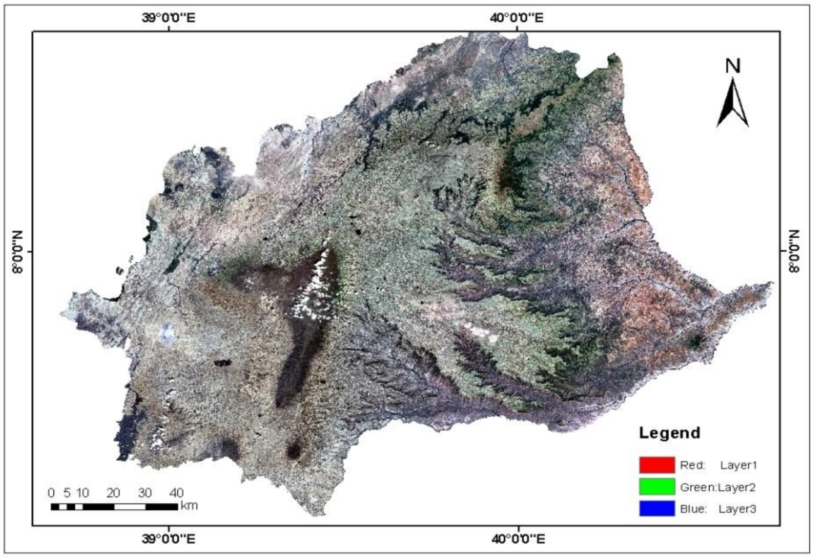


Figure 3.10 SPOT image of the study area.

Object based image analysis requires the creation of objects or separated regions in an image. One established way to do so is image segmentation. Depending on its application, different approaches exist for image segmentation ranging from very simple to highly sophisticated algorithm (Kindu *et al.*, 2013).

Multi resolution segmentation (MS) which was available in e-cognition developer 8.0 software was used for image segmentation. The MS algorithm was bottom up region merging technique starting with a single image object of one pixel and repeatedly merging them in several loops in pairs to larger units. It also optimizes the procedure that minimizes the average heterogeneity for a given number of objects and maximizes their homogeneity based on defined parameters. These parameters are scale, shape and compactness and defined through trial and error to successfully segment objects in an image (Kindu *et al.*, 2013). Using identified target, LULC classes object based classification was applied to a segmented image in order to assign a class to each of the segments using Mad

CAT software and the technique followed was visual interpretation. This approach attempts to assign objects that were generated through image segmentation in to a specific class of interest; agriculture and non-agriculture class (Fig 3.11 and) .

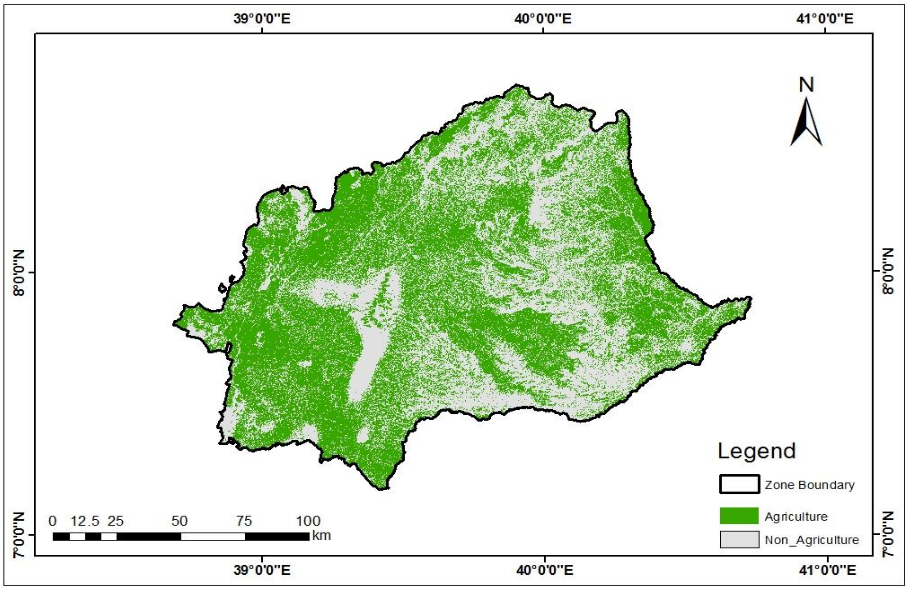


Figure 3.11 Interpreted image of the study area.

.

3.3.3.2. Accuracy assessment

Accuracy assessment is critical for a map generated from any remote sensing data. Error matrix is the most common way to present the accuracy of the classification results. Overall accuracy, user's and producer's accuracies, and the Kappa statistic were derived from the error matrices. The Kappa statistic incorporates the off diagonal elements of the error matrices and represents agreement obtained after removing the proportion of agreement that could be expected to occur by chance. For an accuracy assessment of a map being produced by object based image analysis (OBIA) the units to be tested are the image objects. Therefore, the OBIA classification is validated with the representation of the whole polygon majority class.

Accordingly, the above interpreted classes (i.e. agriculture and non-agriculture) were equally represented. The enough number of samples that represent the thematic classes and ensure good distribution across the map is important to test the attribute accuracy. Rule of thumb is 50 samples per map class or can be derived using the formula devised by Grenier et al. (2008).

$$n = \frac{B\pi(1-\pi)}{b2}$$

Where' n' is total sample, 'B' is determined from a chi square table with 1 degree of freedom multiplied by $1-\frac{\alpha}{k}$ where k is the number of classes ,

'a' is the precision error tolerance,

'b' is absolute precision of each cell ,

'π' is the proportion of the class.

Accordingly the sample size for the accuracy assessment was found to be 288, and 144 sample points were generated for each class. Then, these points were randomly generated for each class and their GPS reading was up loaded to GPS for the field accuracy assessment (Fig 3.12).

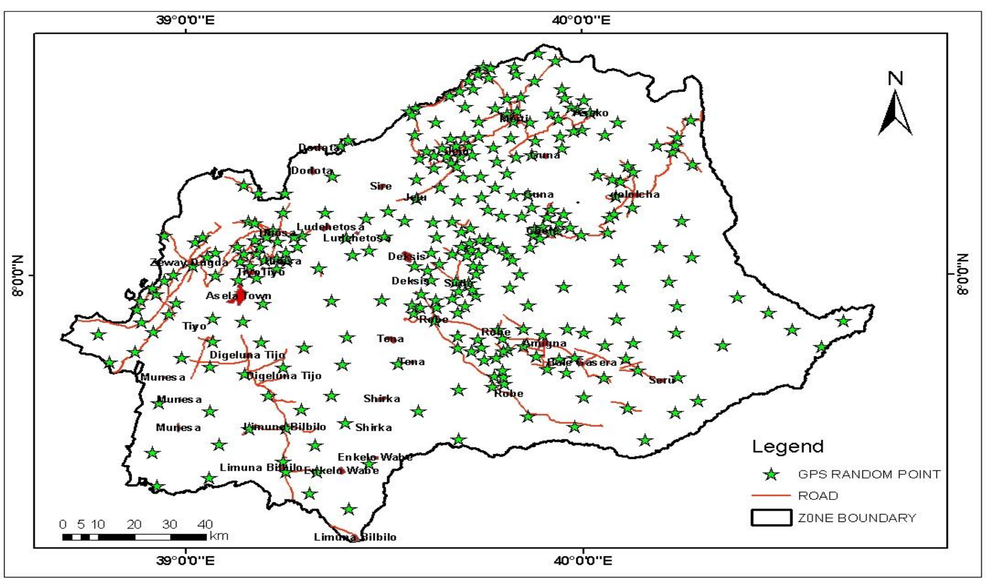


Figure 3.12 Random points generated for accuracy assessment.

These points were checked in two ways; those that were accessible and observed in the field and the second means was using Google Earth as a reference. Accordingly, the following error matrix (Table 3.4) for the 288 sample points is presented.

Table 3.4 Accuracy assessment

	Ground Truth data				
Map Data		Agriculture	Non Agriculture	Total	User Accuracy
	Agriculture	129	15	144	89.58
	Non Agriculture	22	122	144	83.72
	Total	151	137	288	
	Producer Accuracy	85.43	89.05		

The overall accuracy and kappa analysis were used to perform a classification accuracy assessment and accordingly over all accuracy of the data is 87.2% and kappa coefficient was computed which is 0.86 and from the result the interpretation can be taken as accurate result for further analysis. Detail calculation of user and producer accuracy can be referred from the Appendix 2.

3.3.3.3. Crop mask data derivation

Another input for the masking of the crop data is crop agro-ecology of the study area. This data were selected using an optimum elevation suitable for wheat growth according to (Hailu Gebre-Mariam, 2003), i.e. between elevations of l900 and 2700 m Crop mask data for Wheat presented in Figure 3.13.

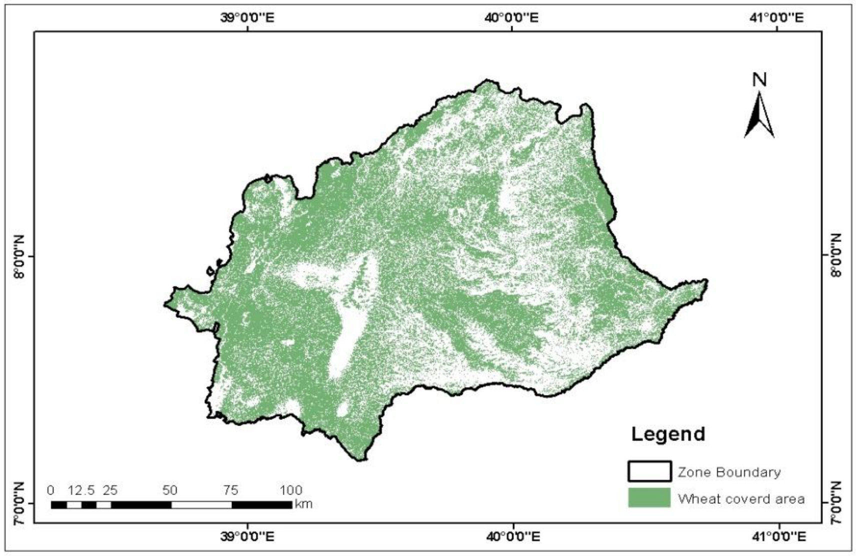


Figure 3.13 Crop mask data for Wheat.

3.3.4. Preparing independent variables using mask data

To determine the predictive capability of the independent variables, all variables were extracted with crop mask data for further correlation analysis and to identify highly correlated ones with the dependent variable (wheat yield).

The time series data (120 decadal) of NDVI have passed through image preprocesses in one go were ready for monthly maximum value compositing (MVC) and 40 monthly composited NDVI images were prepared. These monthly NDVI images were then extracted using the crop mask data to focus only on crop of interest then average NDVI, cumulated NDVI and Maximum NDVI value for each year was computed. The calculated value is in raster value, which ranges from 0 to 255 and needed to be changed to NDVI value. Thus, the formula, NDVI = (RAW*0.004) - 0.1, was run and the result were ready for correlation with wheat yield (Fig 3.15) (Eerens and Haesen, 2013). RFE 2.0 time series data of Decadal image was also composited at monthly level using MVC and were extracted with crop mask data and yearly average was computed from the extracted results for further analysis (Fig 3.18). The WRSI model is a ratio of seasonal actual crop evapotranspiration (ETa) to the seasonal crop water requirement, the same as the potential crop evapotranspiration (PETc). Here, wheat crop coefficient from LEAP software was adopted (Fig 3.14) for the phonological.

May and June – 0.3

July – 0.1.2

August – 1.2

September – 0.25

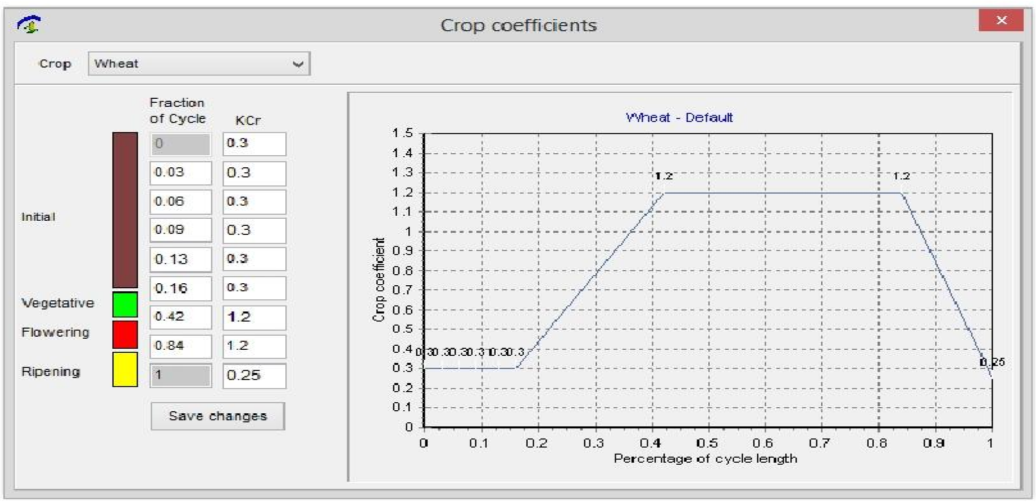
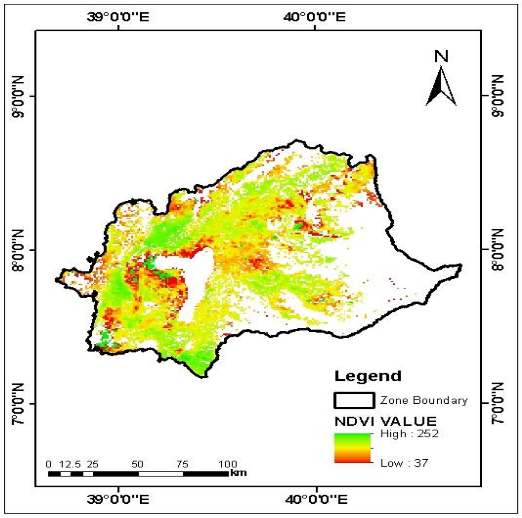


Figure 3.14 Crop coefficient of Wheat in different stage (Planting-Flowering). (Source: LEAP software)

Monthly ETa were multiplied by their respective coefficient and extracted using crop mask data and averaged in order to give ETac for each year. The same procedure was followed for water requirement and resulted in PETc. The ratio of ETac with PETc gave the WRSI and used for further analysis (Fig 3.16 and 3.17).



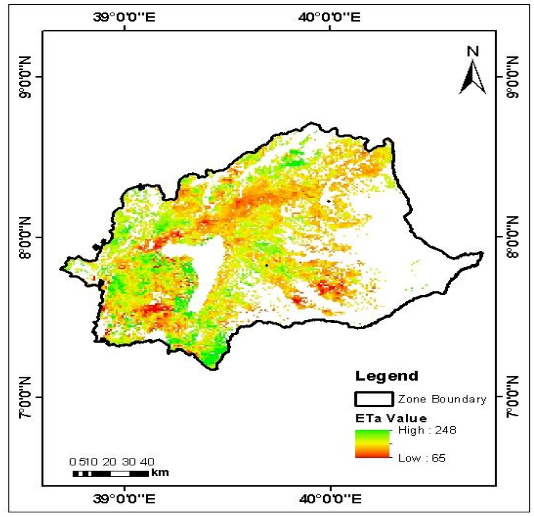
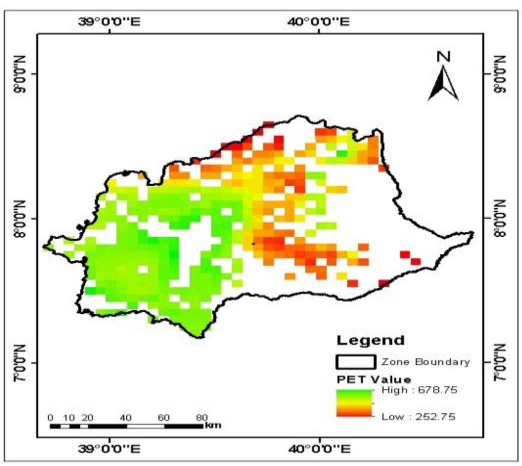


Fig 3.15 NDVI value for the month of. July, 2004

Fig 3.16 Mean ETa for the month of July, 2004



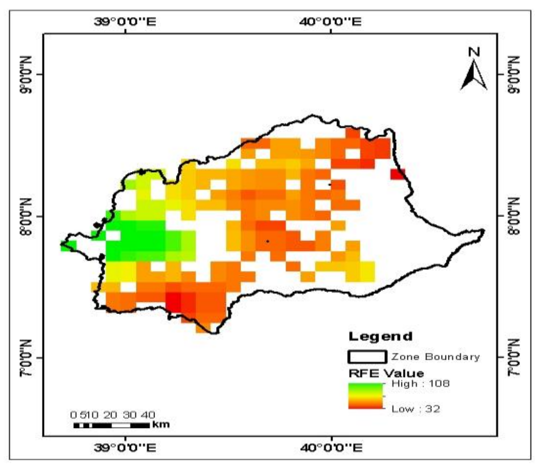


Figure 3.17PET for the month of June, 2004.

Figure 3.18 RFE for the month of June, 2004

.

3.3.5 Multiple linear regression analysis

Before correlating the indices with wheat yield, data quality control was carried out. This is a checking mechanism for the evaluation of collected data before it was used for model development. The only statistical way is the identification of "outliers" within the collected data.

Table 3.5 Table showing observed yield and independent variables.

No.	Year (Meher season)	Yield in Qt/ha	NDVIa	ETa	ETa Total	WRSI	RFE 2.0
1	2004	16.41	0.458	101.1	566	50.57	19.49
2	2005	20.00	0.524	92.8	366.5	45.97	33.71
3	2006	20.63	0.529	120.7	602.5	46.15	34.85
4	2007	17.48	0.512	115.6	570.5	37.8	27.58
5	2008	16.79	0.486	115.8	615	23.11	26.29
6	2009	16.44	0.478	108.8	563.5	18.88	21.39
7	2010	21.14	0.531	75.5	370	50.42	35.79
8	2011	21.94	0.544	119.1	575	29.92	36.73
9	2012	25.72	0.590	115.9	586	54.95	37.23
10	2013	24.45	0.563	96.4	565.5	54.29	36.92

To run Multiple Linear Regression we use the data of table 3.5. The objective of multiple linear regression analysis was to predict the single dependent variable by a set of independent variables. There were some assumptions in using this statistics:-

(a) the criterion variable was assumed to be a random variable

(b) there would be statistical relationship (estimating the average value) rather than functional relationship (calculating an exact value)

(c) there should be linear relationship among the predictors and between the predictors and criterion variable. Multiple regression analysis provides a predictive equation:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_n x_n + \varepsilon$$

Where, β_0= constant

$\beta_1, \beta_2, \ldots, \beta_n$ = beta coefficient or standardized partial regression coefficients (reflecting

the relative impact on the criterion variable)

$x1, x2, \ldots.. x\, n$ = scores on different predictors

The β's are the regression coefficients, representing the amount the dependent variable y changes when the corresponding independent changes 1 unit. The β_0 is the constant, where the regression line intercepts the y axis, representing the amount the dependent y will be when all the independent variables are 0. The standardized version of the β coefficients is the beta weights, and the ratio of the beta coefficients is the ratio of the relative predictive power of the independent variables (Linear regression analysis, Yan and Su 2009). The developed model predicts the average value of one variable (Y) from the value of another variable (X). The X variable is also called a predictor. Generally, this model is called a regression model.

3.3.6 Multicollinearity

The use and interpretation of a multiple linear regression model often depends on the estimates of the individual regression coefficients. Some examples of inferences that are frequently made include:

1. Identifying the relative effects of the independent variables,
2. Prediction and /or estimation , and
3. Selection of an appropriate set of variables for the model.

If there is no linear relationship among the independent variables, they are said to be Orthogonal. When there are near linear dependencies among the independent variables, the problem of **Multicolinearity** is said to exist (Montgomery and Peck, 1991).That means multicollinearity refers to the situation in which two or more independent variables in a multiple linear regression model are highly correlated.

3.3.6.1 Effects of Multicolinearity

Multicollinearity occurs when the predictors included in the linear model are highly correlated with each other (Yan and Su, 2009). In such cases the inferences based on the regression model can be misleading or erroneous. That means the presence of Multicolinearity has a number of potentially serious effects on the least squares estimates of the regression coefficients. The method of least squares will generally produce poor estimates of the individual model parameters. Regression models fit to data by the method of least squares when strong multicillinearity is present are notoriously poor prediction equations, and the values of the regression coefficients are often very sensitive to the data in the particular sample collected. (Montgomery and Peck, 1991).

3.3.6.2 Multicolilnearity Diagnostics

Several techniques have been proposed for detecting multicolinearity. Some of these are Examination of the correlation Matrix, Variance inflation factors, Eigensystem analysis (Condition number). We will now discuss Variance Inflation Factor (VIF).

3.3.6.2.1 **Variance Inflation Factor *(VIF)***

VIF is a commonly used indicator for multicollinearity and computed as the coefficient of determination or R^2 in the linear model that regresses X_j on other predictors($X_{1,}, X_{2, \ldots,} X_p$) in the model, for j=1,. . . ,p. It measures how large the variance of $\widehat{\beta_j}$ is relative to its variance when predictors are uncorrelated (Yan and Su, 2009). If there is multicollinearity between variables, then the correlation coefficient will be large. Strong multicollinearity between independent variables results in large variances and covariances for the least square estimators of the regression coefficients. This implies that different samples taken at the same levels of the dependent variable could lead to widely different estimates of the model parameters.

The formula of VIF is given by:

$VIF_j = C_{jj} = \left(1 - R_j^2\right)^{-1}$

Where $C_{jj,}$ is the j^{th} diagonal element of C, C = $(X'X)^{-1}$,

R_j^2 is the coefficient of determination obtained when x_j is nearly linearly dependent on some subset of the remaining p-1 independent variables.

If R_j equals zero (i.e., no correlation between x_i and the remaining independent variables), then VIF_j equals 1. This is the minimum value of variance inflation factor. For the multiple regression model it is recommended looking at the largest VIF value. Practical experience indicates that if any of the VIFs exceeds 5 or 10, it is an indication that the associated regression coefficients are poorly estimated because of multicollinearity. The VIF for each term in the model measures the combined effect of the dependencies among the independent variables on the variance of that term. (Montgomery and Peck, 1991).

CHAPTER FOUR

RESULTS

4.1 Correlation analysis of different indices with wheat yield

The first step to develop a model is to correlate the independent variables with the dependent variable and by observing the correlation result (coefficient of correlation, R^2, R square adjusted, P value, RMSE), the predictive capability of the independent variable is determined in addition the assumptions were checked and if it is acceptable then will be considered for model development. Based on our data, the SPSS output of the correlation coefficient among the variables is given on the following Table 4.1.

Table 4.1: Correlation coefficient of dependent variable and independent variables.

Variables	Yield in Qt/ha	NDVIa	ETa	ETa Total	WRSI	RFE 2.0
Yield in Qt/ha	1	0.965	-0.080	-0.074	0.613	0.889
NDVIa	0.965	1	0.018	-0.067	0.514	0.919
ETa	-0.080	0.018	1	0.845	-0.422	-0.109
ETa Total	-0.074	-0.067	0.845	1	-0.303	-0.251
WRSI	0.613	0.514	-0.422	-0.303	1	0.475
RFE 2.0	0.889	0.919	-0.109	-0.251	0.475	1

4.1.1 Correlation of NDVI with wheat yield

The relationship between NDVI and biomass enables the early estimation of crop yield, a first evaluation of the available NDVI data for area comprised in the wheat crop mask was made by computing the monthly NDVI averages of each year for the study area and correlating it with the annual cereal yield values. Accordingly, one variable was created when aggregating the NDVI values on a temporal scale. Table 4.1 shows that r =0.96 and Figure 4.1 shows the scatter plot of NDVIa versus wheat yield, (RMSE= 0.93, R^2= 0.96 and significant p–value= 0.00001).

The result indicates that, there is a strong positive linear relationship between NDVIa and wheat yield. This shows that, the NDVI average is a good indicator of crop yield forecast. Therefore, NDVIa is selected for model development.

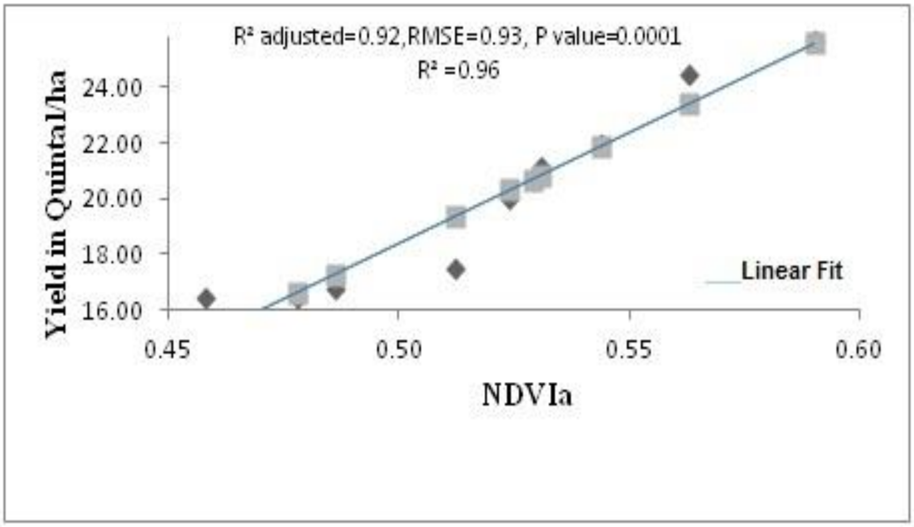


Figure 4.1 Graph showing wheat yield *vs.* NDVIa.

4.1.2. Correlation of rainfall with wheat yield

As it was observed in many findings, the spatiotemporal rainfall distribution needs to be incorporated into crop yield models, in addition to vegetation indices deduced from remote sensing data, to predict crop yield of different cereal crops grown in rain fed condition.

Accordingly rainfall derived from RFE 2.0 was computed for correlation with yield. Table 4.1 shows that r = 0.889 and scatter plot is presented in Figure 4.2 (RMSE = 1.62, R^2 = 0.79, p – value = 0.00058). The result indicate that there is a strong positive linear relationship between RFE 2.0 and wheat yield.

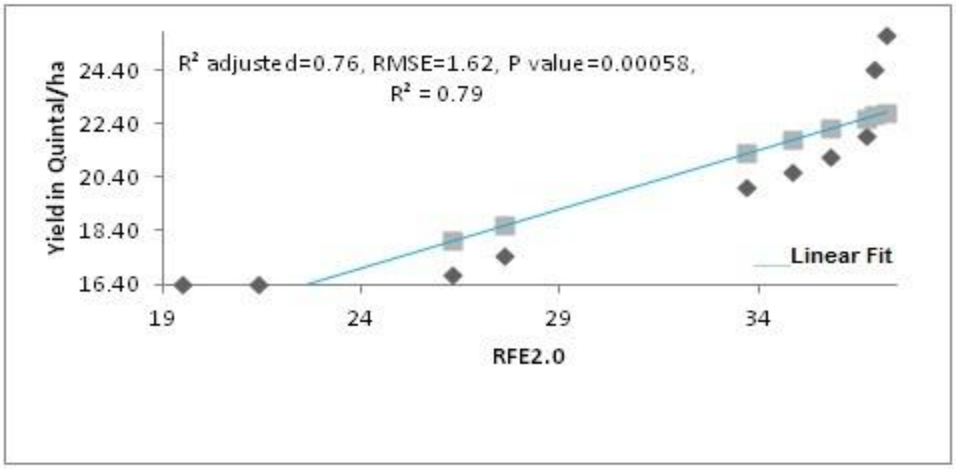


Figure 4.2 Graph showing the correlation between wheat yield and rainfall

From this, it can be observed that rainfall is a determinant factor for crop production for areas like east Arsi where rain fed agriculture is practiced and this variable is taken for model development in line with the previously identified NDVI average.

4.1.3. Correlation of WRSI with wheat yield

Studies proof that there is strong relation between WRSI and Yield. Accordingly, WRSI which was computed for the study area were bringing to the correlation analysis with yield. (Table 4.1) and (Fig 4.3) shows minimum value of correlation (0.61%), RMSE=2.8, R^2=0.38 and p value (0.06) of WRSI with wheat yield. Therefore for our data, even though there was medium correlation coefficient (r=0.61), WRSI and wheat yield variables do not have a Statistical significant relationship.

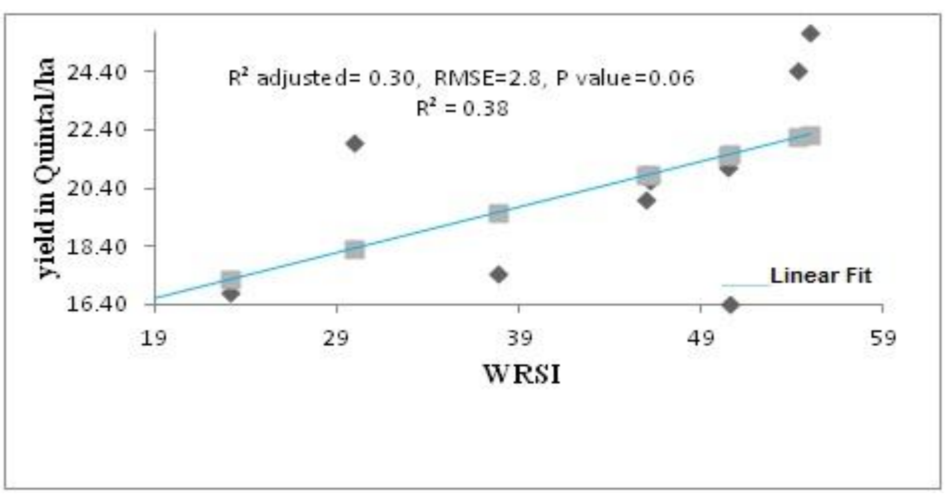


Figure 4.3 Graph showing the correlation between wheat yield and water Requirement satisfaction index.

4.1.4 Correlation of ETa with wheat yield

Evaluation of ETa variables was made by computing the cumulated and average ETa for the whole cycle. Table 4.1, Figure 4.4 and 4.5 indicates the correlation of different ETa variables with wheat yield.

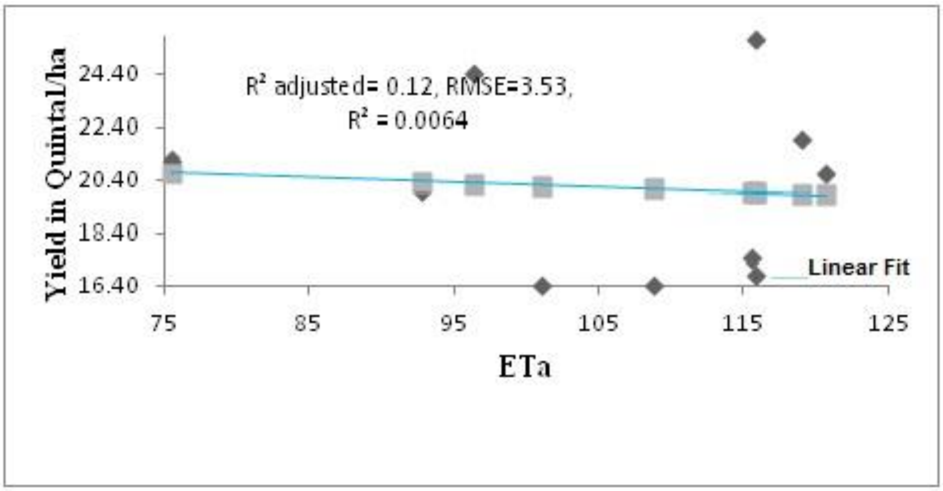


Figure 4.4 Graph showing the correlation between wheat yield and Actual Evapotranspiration.

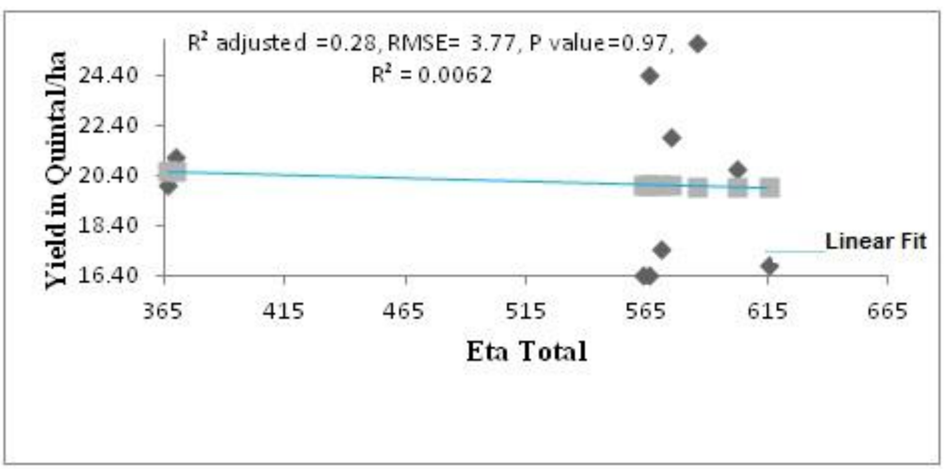


Figure 4.5 Graph showing the correlation between wheat yield and Actual Evapotranspiration total

The ETa result (Table4.1 and Fig 4.4) shows a week value of correlation (-0.078), RMSE=3.77, R^2=0.0062 and p value (0.97) of ETa with wheat yield.

From the correlation result we observed that there is no significant correlation between both average and cumulated Eta. Therefore, this variable cannot be considered for model development.

4.2 Multiple Linear Regression Model for Yield Forecasting

A regression model that involves more than one independent variable is called a multiple regression model. Multiple regression model was discussed in detail in sub-section 3.3.5. As can be noted from the result of correlation matrix (Table 4.1), three of the five variables did not show a significant relationship with yield at 5% level of significance. In this regard, Eta, Eta Total and WRSI were not significant at 5% level of significance and were excluded from the multiple linear regression models. In fact, the corresponding p-values for each of these variables were greater than 0.5.

4.2.1 Estimation of the model Parameters:

From the result of section 4.1, we know that the two variables (NDVIa and RFE 2.0) out of the five variables have a significant relationship with the dependent variable, Yield and they are selected for model development. Therefore, for our data, we have one dependent and two independent variables.

4.2.2 Test for Significance of Regression

The test for significance of regression is a test to determine if there is a linear relationship between the dependent and any of the independent variables.

Among the five independent variables: variables derived from remote sensing and climatic variables, which satisfies the assumptions for multiple linear regression model with significant P value at 95 % confidence level were derived for the model development.

Normalized Difference Vegetation Index Average, which is a result of monthly maximum value composite (MVC) averages of NDVI from the planting date to the end of the crop cycle gives a correlation coefficient of 0.96 with significant P value of 0.0001 at 95 % confidence level. The second highly correlated independent variable was rainfall with a correlation coefficient of 0.89 with significant P value of 0.00058 at 95% confidence level. While others like ETa (total), which has a correlation value of -0.079 and WRSI (r =0.61) with a P value of > 0.93, which is beyond the acceptable range at 95 % of confidence level were rejected from the model development.

Hence the two most correlated variables (NDVIa and rainfall) with the dependent variable (yield) are selected to create a multiple linear regression model. As many studies on crop forecast states that linear regression modeling is the most common method to produce yield predictions by using remote sensing derived indicators together with bio climatic information.

Wheat yield data and data derived from the different indices were prepared for multiple linear regression analysis. The Statistical Package for Social Science (SPSS) software was used to build a multiple linear regression model using the two most correlated variables.

As a result of all the above processes, the model highly correlated variables (NDVIa and Rainfall) were used to develop a model. This model was validated based on its Coefficient of determination (R^2), root mean square error (RMSE) and coefficient of variation (CV) as shows in (Fig 4.6).

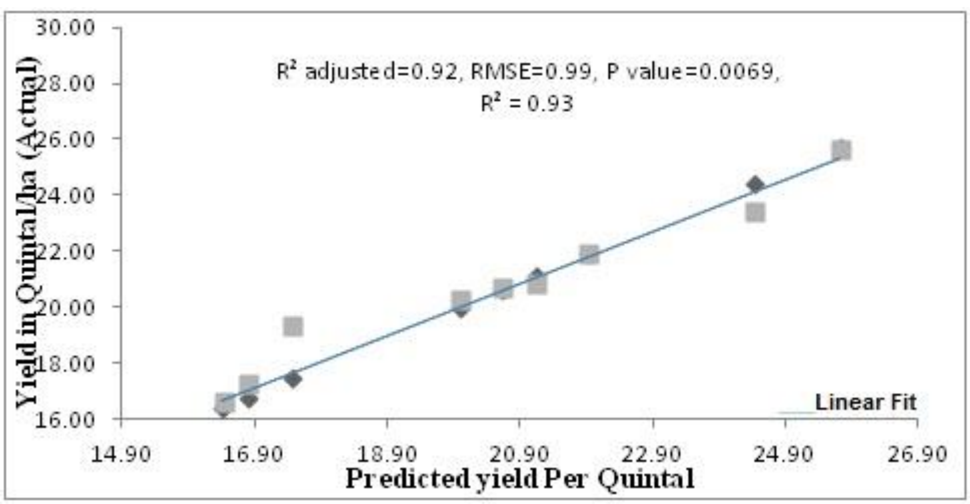


Figure 4.6. Comparison between the wheat yields estimated by the agro meteorological Model and the observed yields for the study area

When we see the overall fit of the model by examining the plot of the actual yield per hectare against the predicted yield per hectare, it reveals that, most points lie fairly close to the 45° line (exact prediction line). The R square value of the model is 0.93; R square adjusted is 0.92 with root mean square error of 0.99 quintal per hectare. The P value of the model is

0.00069 at 95 % confidence level and it is a good evidence that at least one independent variable appears to predict wheat yield in quintal with a significant value. By observing this P value, it is unclear which independent variable is the very good predictor and which is poor. The analysis of variance as shown in Table 4.2 state that wheat yield forecast model has an observed significance probability (Prob>F) of 0.0001, which is significant at 0.05 level. Since the $p<0.0001$, we conclude that Yield is related to NDVIa and/or RFE 2.0.

Table 4.2 Wheat yield forecast model variance analysis (ANOVA)

Model	**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Regression	93.205	2	46.602	47.330	0.000
Residual	6.892	7	0.985		
Total	100.097	9			

Degree of freedom (df):- is the number of values in the final calculation of a statistic (estimated value) that are free to vary.

Table 4.3 Variance Inflation Factor (VIF) between NDVIa and REE 2.0

Variance Inflation Factor (VIF)	
Constant	0
NDVIa	6.456
RFE 2.0	6.456

From Table 4.3, the Variance Inflation Factor (VIF) of NDVIa and RFE 2.0 is 6.456 for each. In addition we have already checked that on Table 4.1 the correlation between these two variables is 0.919 which is significant. This shows that there is a multicollinearity problem. Multicollinearity is discussed in section 3.6 in detail. In practice multicollinearity is high if the VIF is greater than 5. Therefore we have to delete one of the two variables. For this research NDVIa is selected for the model and RFE2.0 is removed from the model.

4.2.3 Tests on Individual Regression Coefficients

These tests are helpful in determining the value of each of the dependent variable in the model. Thus this is a test of the contribution of x_j given the other independent variables in the model. Table 4.4 contains the output of individual regression coefficient test.

Table 4.4 Parameter estimates for the wheat forecast model

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error			
1	(Constant)	-21.930	4.053	Beta	-5.412	.001
	NDVIa	80.595	7.750	.965	10.399	.000

a. Dependent Variable: yield

Table 4.4 shows parameter estimates of the model which reveals that NDVIa has a significant Probability value. Therefore, from the result of table 4.4 the values of the intercept (constant term) and the estimate of NDVIa are -21.93 and 80.595 respectively. The p-values of the intercept and NDVIa are less than 0.01. Therefore, NDVIa significantly affects yield. I.e. A unit change in NDVIa brings 80.595 unit change in Yield.

Therefore the model is given by:

$$\textbf{Wheat yield (qt/ha)} = \mathbf{-21.93 + 80.595 * NDVI_{average}}$$

According to the Central Statistical Agency Agricultural Production report coefficient of variation of Wheat Yield is 17.7% and it in the acceptable range of validation value:

4.3 Contrasting of conventional crop yield prediction with the developed model

The Central Statistical Agency (CSA) of Ethiopia estimate, crop production yearly bases data. The Agency has increased the number of data based on stakeholders. With the objective to keep up and improve the data quality in terms of reliability and accuracy. Since then, the Annual Crop Production Forecast Survey conducted included researchers (sampled households, development agents, chairperson of the rural kebele, community leaders and observations from highly qualified professionals from CSA and FAO) as ultimate statistical unit on collecting the data.

Gommes (2001) stated that any approach other than stake holders' assessment for crop yield forecast and which will avoid bias and ensure at least a reasonable degree of consistency from year to year and from place to place should be preferred.

When comparing the conventional and remote sensing yield forecast, the remote sensing approach minimizes subjectivity. According to CSA report, the forecast data which is a result of conventional approach reveals a coefficient of variation of 17.7% and it is a subjective approach. But the remote sensing supported model shows 16.01% with acceptable degree of confidence (95%) and significant probability value.

In addition the prediction result of the remote sensing supported approach can be provided at the beginning of October considering September as a flowering stage of the wheat crop while the conventional method data release calendar is mostly in December and it includes all cereal crops. This shows that the timeliness issue can be addressed by using the remote sensing supported approach in a better way than the conventional approach even though we did not consider all cereals which CSA has covered both in my research.

The other advantage of the remote sensing supported approach is it provides location information in that after the forecast is made, the result can be verified by taking GPS reading and navigates to the areas. Therefore this approach creates an opportunity to exactly indicate which areas have high and low yield and in a tangible manner which the conventional approaches lucks badly.

Therefore, it is clear that wheat yield prediction using remote sensing and GIS improves both quality and timeliness of the data more over it minimizes subjectivity considerably. Remote sensing supported approach has also a capacity to demonstrate areas (lower administrative areas) where there is relatively high, medium and low production and this makes intervention very easy for the decision makers as it was proofed by this research and the previous related work.

The following (Fig 4.7) illustrates the comparison between the conventional yield estimates with the Remote Sensing support approach.

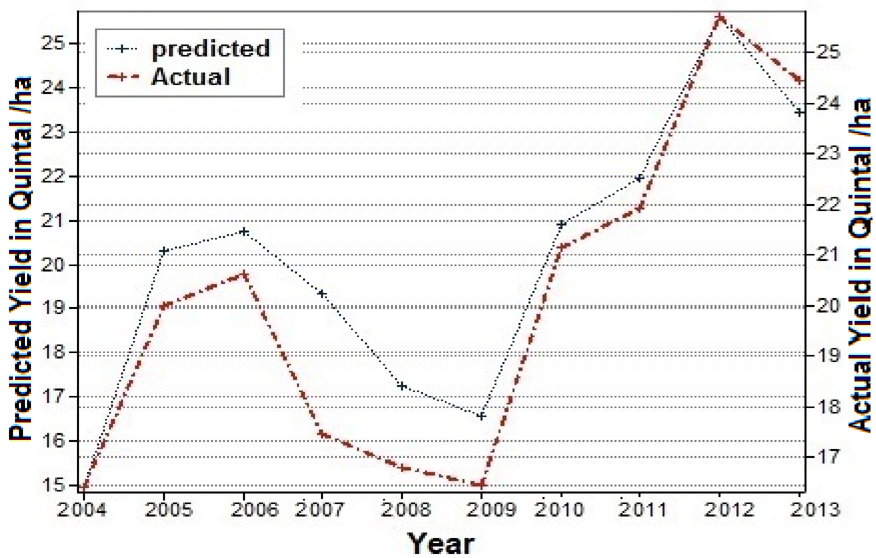


Figure 4.7. Comparison between wheat yield (quintal/ha) Estimated by the model and the observed yield.

4.4 Wheat crop forecast for the year 2014

Using the developed remote sensing model, the 2014 wheat crop forecast was made. Accordingly, an average of 38.9 quintal per/ha was expected with 65 quintal per/ha as the highest and 13 quintal per/ha as the lowest. Table 4.5 and Figure 4.8 show the productivity level of wheat crop for the year 2014 in the study area. More than 46.71% of the area covered with wheat yield 33−44 quintal/ha while 27.43% of the area falls which 45–65 quintal/ha production as the level one category of the zone. The rest (25.9% of the area) is within the range of 13–32 quintal/ha production.

According to the results obtained, areas to the north west part of the Zone such as Hitosa, Lude Hitosa, Tiyo, Asela Zuria, Digelina Tijo and in Munesamore district (about 27.43% of the areas) were falls in productivity of 45 – 65quintal/ha. The central part of the Zone, Robe, Tena, Shirka and Bale Gesera weredas which accounts (46.71% of the area) falls in 33−44 quintal/ha. The rest (25.9% of the area) is within the range of 13–32 quintal/ha productivity. Hence, the north western part of the Zone was more productive than other part of the study area.

Table 4.5 Wheat production level of the year 2014 for East Arsi zone

Wheat productivity level	Quintal per/ha	Area coverage in (%)
I	45–65	27.43%
II	33–44	46.71%
III	13–32	25.9%

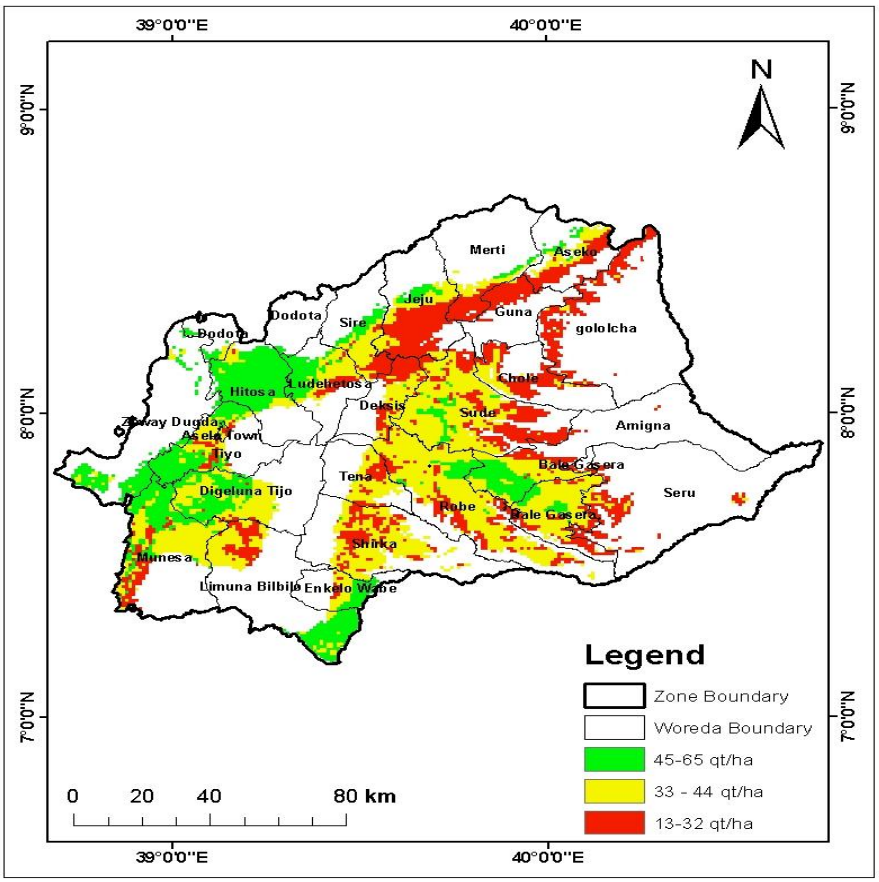


Figure 4.8 Wheat yield forecast map of 2014.

CHAPTER FIVE

DISCUSSION

The model developed as a result of this research has utilized official crop yield data of Ethiopia government as collected and analyzed by Central Statistical Agency of Ethiopia and different variables derived from remotely sensed parameters (NDVIa and Rainfall).

A correlation analysis was carried out for all remotely sensed parameters and variables with high correlation coefficient with significant P value were selected for model development. Accordingly Normalized Difference Vegetation Index (NDVI) average and rain fall (REF 2.0) were found to be highly correlated with yield and applied for model building using multiple linear regression model which is proved as the best model for crop forecast. Based on Variance Inflation Factor (VIF) result. VIF of more than 5 is excluded. NDVIa is included in the model and RFE 2.0 is excluded from the model.

According to Sawasawa (2003) among all the NDVI parameters, only maximum NDVI was retained as significant variable for field level yield prediction in the stepwise multiple regression, explaining 25.0% of the yield variability suggesting that NDVI is the best parameter for yield prediction and this result also coincides with the research output. Mean while this research found out that NDVIa is retained in the model, explaining 93% of yield variability.

Regarding the related research conducted, Even though the crop is different, we have tried to compare our result with different research conducted. Rojas (2006) conducted a maize yield forecast in Kenya for a maize crop and accordingly cumulative NDVI with 87% correlation was found to be the most correlated parameter to estimate maize crop as like this research output.

According to Rembold *et al*. (2013) in many cases, the predictive power of remotely sensed indicators can be improved by adding independent meteorological (bio climatic) variables in the regression model. Several bioclimatic and remote sensing based indicators have proven to be highly correlated with yield for certain crops in specific areas. In this research too, we have proved that remote sensing indicator (NDVI) was highly correlated with yield and helped much for the high predictive capability of the model.

Satellite-derived vegetation and drought monitoring tools including ETa anomaly, NDVI, and merged (satellite and ground observation) RFE products are currently used by USGS FEWS NET for agricultural drought monitoring and food security status assessment in Africa. These products are expected to capture the weather variability and its impact on crop yield. In this study, a process to evaluate and assess the potential use of satellite-derived evapotranspiration anomaly using crop yield data was developed. This process includes exploring the variables (e.g., ETa and crop yield) using geospatial exploratory and statistical techniques. In this study the explored variables are NDVIa and crop yield.

The result of the model of this paper showed high prediction capability (Root mean square error =0.99 and R^2 =93%). When this result is compared with Abiy's 2014 south Tigray zone maize yield forecast, it shows the possibility of crop yield forecast using Remote Sensing data and it reveals that the magnitude (Root mean square error =1.41 and R^2 =88%) of the result is almost the same.

In relation to the independent variables highly correlated with yield, in this research Normalized Difference Vegetation Index Average (NDVIa) shows highly correlated (r=0.96) while rainfall (r=0.89) is second. But in Abiy, 2014 rainfall is highly correlated (r=0.84) followed by (r=0.79), Normalized Difference Vegetation Index Average (NDVIa). This shows that parameters for yield prediction differ from one agro ecological zone to other hence our model considers different factors in different correlation result.

The other similarity regarding factors or parameters derived from Remote Sensing data, in this research Water Requirement Satisfaction Index (WRSI) and Actual Evapotranspiration (Eta) were not correlated with yield and the same was true for Abiy's 2014.

In this research paper, for the final model, according to the Statistics result, NDVI is selected while rainfall is rejected based Variance Inflation Factor (VIF) result, but in Abiye's (2014) paper NDVIa and Rainfall are included in the final model.

In general the findings of this research, which is a second crop studied after Abiye's (2014) maize crop yield forecast research, demonstrates a clear potential of Agro metrological parameters for wheat yield forecasting in East Arsi zone.

CHAPTER SIX

CONCLUSION AND RECOMMENDATIONS

6.1. Conclusion

Policy makers need accurate and timely information on crop production and areas as soon as possible at the lower administrative level. Such information should be available before the harvest so that preparation can be made which is the very cause of forecasting crop yield using different approaches.

The major objective of this study was to develop a model for wheat crop using remote sensing and GIS. Accordingly crop statistical data as a dependent variable and different predictor variables derived from remotely sensed imageries were computed and those variables with higher correlation and significant P value are selected for model development. The analysis result confirmed that NDVIa and rainfall of study area have good correlation r=0.96 and r=0.89, respectively with significant P value confirms the result.

Based on these correlation results, agrometerological yield forecasting using a multiple linear regression was made using a table of data containing yields as a dependent and a series of agro meteorological and remote sensing variables which have high correlation with the yield. The developed agrometerological model has a predictive capability of 0.93 with RMSE 0.92 quintal per hectare and it is a very appropriate result. It can be stated that with the demonstrated yield forecasting methods an adequately accurate forecast can be given using remotely sensed data in an area like east Arsi zone where there is fragmented plots of land.

Using the regression model developed for the study area, wheat yield forecast is possible roughly well before the date of the harvest. Wheat yield forecast map of the year 2014 was also prepared using the developed model and an average result of 38.9 quintal per hectare was forecasted showing the north West part of the zone having high productivity per hectare and can be used by the decision makers to identify relative productive areas prior to harvest at the lower administration level.

Generally it is possible to conduct wheat yield forecast using NDVI derived from SPOT VEGETATION and for areas similar to east Arsi zone.

6.2. Recommendations

Based on the encouraging results of this research output, the developed model can be checked in areas other than east Arsi after a procedures stated in the methodology of the paper is followed meanwhile more research and broader testing is necessary. As an initial effort, this application of the meteorological yield model appears promising. Some further effort is necessary to operationalize the results of this research which includes:

- A relatively longer period of time series data should be analyzed in order to reach to operational application.
- Expertise from different disciplines (biologists, Agronomists etc.) should be involved to end up with a more sounding result.
- For future research, I recommend that by including some variables like soil and it is better to use other models such as Polynomial regression model, Non-linear regression model instead of depend on Multiple Linear Regression Model.

Further investigation should be done to identify more factors that contribute to yield variability and application of remote sensing and GIS in an advanced way.

References

Abiy Wgderes (2014). Maize Yield Forcasting in South Tigray, M.Sc. Thesis Addis Ababa University. Addis Ababa, Ethiopia.

Alemayehu Seyoum, Paul.D, Sinafikeh Asrat (2010). Crop production in Ethiopia: Regional Pattern and Trends. International Food policy Research Institute. Addis Ababa, Ethiopia.

Allbed, A., Kumar, L. and Sinha, P. (2014). Mapping and modeling spatial variation in soil Salinity in the Al Hassa Oasis based on remote sensing indicators and regression techniques. *Remote Sens*. **6**: 1137–1157.

Atzberger, C. (2013). Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. *Remote Sensing*. **5**: 949–981.

Abu Tefera (2013).Assessments of Commodity and Trend issues. USDA Foreign Agricultural Service. Addis Ababa, Ethiopia.

Becker-Reshef, I., Justice, C.O., Sullivan, M., Vermote, E.F., Tucker, C., Anyamba, A., Small, J., Pak, E., Masuoka, E. and Schmaltz, J. (2010). Monitoring global croplands with coarse resolution Earth observation: the global agriculture monitoring (GLAM) project. *Remote Sens.* **2**: 1589–1609.

Beyene, E. G. and Meissner, B. (2009). Spatio-temporal analyses of correlation between NOAA satellite RFE and weather stations' rainfall record in Ethiopia. *Applied Earth Observation and Geoinformation*. **12**(Supplement 1): S69–S75.

Benedetti, R. and Rossini, 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*. **45**: 311–326.

Bryman, A., Cramer, D. (2005). Quantitative Data Analysis with SPSS 12 and 13. published in the Taylor & Francis e-Library, 270 Madison Avenue, New York, NY 10016, USA.

Challinor, A.J., Wheeler, T.R., Craufurd, P.Q., Slingo, J.M., Grimes, D.I.F., 2004. Design and optimization of a large-area process-based model for annual crops. Agric. For. Meteorology. 124, 99−120.

Chandiposha, Chagonda Ignatius, Makuvaro. V (2013).Utilization of Common grain crops in Zimbabwe. Department of Agronomy, Midlands State University, P.Bag 9055,Gweru, Zimbabwe.

CSA (2013). Annual Agricultural Report. Central Statistical Agency, Addis Ababa, Ethiopia. http://www.csa.gov.et/index.php/2013-02-20-13-43- d35/national-statistics abstract/141- population accessed on 01/17/2015

Dagnew Belay (2007). Assessment of Causes and Extent of Land Degradation in Hashenge Catchment, Southern Tigray, Ethiopia. MSc Thesis, Addis Ababa University, Addis Ababa, Ethiopia.

Demeke, M. (2012). Analysis of Incentives and Disincentives for Maize in Ethiopia. Technical Notes Series, MAFAP, FAO, Rome, Italy.

Dereje, G. and Eshetu, A. (2011). Crops and Agroecology Zones of Ethiopia, Ethiopian Institute of Agricultural Research, Addis Ababa, Ethiopia.

DPPC (2003). Early Warning System Monthly Report, Page 1 AREAS OF PRIORITY ACTION The Reassessment Teams came up with Additional Number of Needy People

Eerens, H., Haesen, D., Rembold, F., Urbano, F., Tote, C. and Bydekerke, L. (2014). Image time series processing for agricultural monitoring. Environment. *Mod.Software*. **53**: 154–162.

Eerens, H. and Haesen, D. (2013). Software for the Processing and Interpretation of Remotely Sensed Image Time Series. Vito, Belgium, pp288.

Environmental Systems Research Institute (ESRI).(2013). http://proceedings.esri.com/library/userconf/proc00/professional/papers/pap601/p601.htm accessed on 02/12/2015.

FAO (2010). FAO Global Information and Early Warning System on Food and Agriculture Special Report. Food and Agriculture Organization, Addis Ababa, Ethiopia.

George, A. and Hanuschak, S. (2010). Timely and Accurate Crop Yield Forecasting and Estimation, History and Initial Gap Analysis.

Gommes, R. (2001). An introduction to the Art of Agro meteorological Crop Yield Forecasting Using Multiple Regression. Unpublished Document, FAO, Dhaka, Bangladesh, pp38.

Grenier, M., Labrecque, S., Benoit, M. and Allard, M. (2008). Accuracy assessment method for wet land object based classification. **In** : *Proceedings of the XXXVIII-4c1*. ISPRS, Aug 5-8,2008.Calgary,Canada.

Greatrex, H. (2012). The Application of Seasonal Rainfall Forecasts and Satellite Rainfall Estimates to Seasonal Crop Yield Forecasting for Africa. PhD Disertation, University of Reading, Reading, UK.

Ministry of Agriculture (MOA) (2007). Livelihood Profile of Tigray Region, Ethiopia. Irob Mountain Livelihood Zone. MOA, Addis Ababa, Ethiopia,5pp.

Haile, M., (2005). Weather patterns, food security and humanitarian response in sub-Saharan Africa.

Hailu Gebre-Mariam (2003). Wheat Production and Research in Ethiopia. Addis Ababa, Ethiopia.

Hastings, D.A. and Emery, W. J. (2005). The advanced very high resolution radiometer (AVHRR): A brief reference guide. *Photogram. Eng. Remote Sensing*. **58**: 1183–1888.

Helen Greatrex (2012). The Application of Seasonal Rainfall Forecasts and Satellite Rainfall Estimates to seasonal crop yield forecasting for Africa. The University of Reading, Reading, UK.

Holli.R (2014).How to Interpret Satellite Image: Five Tips and strategies

Jobard, I., F. Chopin, J. C. Berges, and R. Roca (2011), An intercomparison of 10-day satellite precipitation products during the west African monsoon, Int. J. Remote ens., 32(9), 2353-2376.

Kindu, M., Thomas, S., Teketay. D. and Thomas, K. (2013). Landuse/Landcover change Analysis using object based classification approach in Munesa Shashemene Landscape of the Ethiopian highlands. *Remote Sens*.**5**: 2411–2435.

Legesse, G and K. V. Suryabhagavan (2014). Remote sensing and GIS based agricultural drought assessment in East Shewa Zone, Ethiopia. *Tropical ecology*. 55, 349−363.

Landau, S., Evertitt B.S.A (2003).Handbook of Statistical Analyses using SPSS. London

Lillesand,T. and Kiefer, R.W. (1994). Remote Sensing and Image Interpretation. 3rd., John Wiley & Sons, Inc.

Montgomery, D., Peck, E.A. (1991). Introduction to Linear Regression Analysis. World Scientific Publishing Co. Pte. Ltd. University of Central Florida, USA.

Novella, N.S. and Thiaw, W.M. (2012). African rainfall climatology version 2 for famine early warning systems. *Applied Meteorology and Climatology*. **52**: 588−606.

Potdar, (2002). The agro-meteorological models introduce information about solar radiation, taken into consideration by the agro-meteorological models.

Rembold, F., Atzberger, C., Savin, I. and Rojas, O. (2013). Using Low Resolution Satellite Imagery for Yield Prediction and Yield Anomaly Detection. *Remote Sens*. **5**: 1704–1733.

Rojas, O. (2006). Operational maize yield model development and validation based on Remote sensing and agrometereological data in Kenya. **In:** *Proceedings of Remote* Sensing Support *to Crop yield Forecast and Area Estimates Workshop*.

Rojas, O., Rembold, F., Royer, A. and Negere, T. (2005). Real time agrometereological crop yield monitoring in eastern Africa. *Agron.Sustain.Dev*.**25**: 63–77.

Rembold, F. and Maseli, F. (2004). Estimating inter annual crop area variation using multi Rijks, resolution satellite sensor images. *Int. J. Remote Sensing*. **25**: 2641–2647.

Reynolds, C.A., Yitayew, M., Slack, D.C., Hutchinson, C.F., Huete, A., Petersen, M.S., (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 21, 3487–3508.

Rijks, O., Massart, M., Rembold, F., Gommes, R. and Leo, O. (2007). Crop and rangeland monitoring in eastern Africa. **In**: *Proceedings of the 2nd International workshop*, pp.95–104. Nairobi, Kenya.

Sawasawa H.L.A. (2003). Crop Yield Estimation: Integrating Remote Sensing, GIS and Management Factors: A Case Study of BIRKOOR and HORTGIRI MANDALS Nizambad District, India. M.S.c. Thesis, ITC, Enschede, The Netherlands.

Senay, G.B. and Verdin, J. (2003). Characterization of yield reduction in Ethiopia using a GIS based crop water balance model. Remote Sensing. **29**: 687–692.

Tafesse, T. (2009) 'A review of Ethiopia's water sector policy, strategy and program' in Assefa, T. (ed.), Digest of Ethiopia's National Policies, Strategies and Programs, Addis Ababa.

Tewelde Yideg Atakilti (2012). Assessing the Potential of GeoNetCast Earth Observation and Insitu Data for Drought Early Warning and Monitoring in Tigray, Ethiopia. M.Sc. Thesis, University of Twenty, Enschede, the Netherlands.

Wang, E., Robertson, M.J., Hammer, G.L., Carberry, P.S.,Holzworth, D., Meinke, H., Chapman, S.C., Chapman, J.N.G.,Hargreaves, J.N.G., Huth, N.I., McLean, G., 2005. Development of a generic crop model template in the cropping system modelAPSIM. Eur. J. Agron. 18,121−140.

Yan, X., Su, X.G. (2009). Linear Regression Analysis –Theory and Computing. World Scientific Publishing Co. Pte. Ltd. University of Central Florida, USA.

Appendix 1: Sample GPS readings for accuracy assessment

Agriculture class

S.No	Easting	Northing	MAP CLASS
1	534704	809679	Agriculture
2	554704	819879	Agriculture
3	608337	856097	Agriculture
4	610759	842541	Agriculture
5	606132	866033	Agriculture
6	616476	860267	Agriculture
7	625815	851804	Agriculture
8	574273	891625	Agriculture
9	514915	882753	Agriculture
10	516014	868420	Agriculture
11	519894	883414	Agriculture
12	522594	897516	Agriculture
13	525812	895969	Agriculture
14	574386	899602	Agriculture
15	537152	886680	Agriculture
16	520962	983771	Agriculture
17	494286	898025	Agriculture
18	508424	884228	Agriculture
19	496876	884465	Agriculture
20	494289	882413	Agriculture

Non Agriculture Class

S.No	Easting	Northing	MAP CLASS
1	676711	859839	Non agricultural
2	576016	828069	Non agricultural
3	527902	832011	Non agricultural
4	523262	843022	Non agricultural
5	608131	832462	Non agricultural
6	586798	865587	Non agricultural
7	636143	837243	Non agricultural
8	609912	898299	Non agricultural
9	540629	842997	Non agricultural
10	593036	884830	Non agricultural
11	487821	868353	Non agricultural
12	475952	864243	Non agricultural
13	502007	887486	Non agricultural
14	564133	941613	Non agricultural
15	562172	940582	Non agricultural

APPENDIX 2: Accuracy assessment matrix result

REPORT

```
          -----------------------------------------
Image File: c:/Aklilu folder/Arsi_unsup.img
User Name: user
Date       : Thu Apr 02 13:11:10 2015
```

ERROR MATRIX

Classified Data	Reference Data: Unclassified	Class 1	Class 2	Row Total
Unclassified	0	0	0	0
Class 1	0	129	15	144
Class 2	0	22	122	144
Column Total	0	151	137	288

----- End of Error Matrix -----

ACCURACY TOTALS

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Unclassified	0	0	0	---	---
Class 1	151	144	129	85.43%	89.58%
Class 2	137	144	122	89.05%	84.03%
Totals	288	288	251		

Overall Classification Accuracy =   87.2%

----- End of Accuracy Totals -----

```
KAPPA (K^) STATISTICS
---------------------

Overall Kappa Statistics = 0.8627



Conditional Kappa for each Category.
------------------------------------

    Class Name            Kappa
    ----------            -----
  Unclassified           0.0000
       Class 1           0.8624
       Class 2           0.9280

           ----- End of Kappa Statistics -----
```

D E C L A R A T I O N

I hereby declare that the thesis entitled. *AGROMETEREOLOGICAL WHEAT YIELD FORECAST MODEL USING REMOTE SENSING AND GIS IN EAST ARSI ZONE, ETHIOPIA***,** has been carried out by me under the supervision of Dr. K. V. Suryabhagavan and Prof. M. Balakrishnan, School of Graduate studies, Addis Ababa University, Addis Ababa during the year 2015 as a part of Master of Science program in Remote Sensing and GIS. I further declare that this work has not been submitted to any other University or Institution for the award of any degree or diploma.

AKLILU FIKRE REDA

Signature: ________________________

Addis Ababa University

Addis Ababa

Date: June, 2015

C E R T I F I C A T E

This is certified that the thesis entitled. *WHEAT YIELD FORECAST USING REMOTE SENSING AND GIS IN EAST ARSI ZONE, ETHIOPIA*. Is a bona fed work carried out by Aklilu Fikre Reda under my guidance and supervision. This is the actual work done by Aklilu Fikre Reda for the partial fulfillment of the award of the Degree of Master of Science in Remote Sensing and GIS from Addis Ababa University, Addis Ababa, Ethiopia.

Dr. K. V. Suryabhagavan
Assistant Professor

Signature: ________________________

Department of Earth Science

Addis Ababa University