ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF COMPUTER SCIENCE

# Analyzing the Effect of High Resolution Satellite Images for Drought Prediction

By

Yonatan Getachew

A thesis submitted to the school of Graduate studies of Addis Ababa University in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science

April 16, 2013

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF COMPUTER SCIENCE

# Analyzing the Effect of High Resolution Satellite Images for Drought Prediction

By: Yonatan Getachew

Advisor: Dr. Solomon Atnafu

Name and Signature of members of the Examining Board:

| Name | Signature |
|---|---|
| 1. Dr. Solomon Atnafu, Advisor | ______________________________ |
| 2. ______________________________ | ______________________________ |
| 3. ______________________________ | ______________________________ |
| 4. ______________________________ | ______________________________ |

# Acknowledgement

First and for most I thank God for his support in achieving my dream throughout my life

Without the love and care of my beautiful Yoeyea everything would have been a sad story. I would like to thank u my dear, for your heart felt love and encouragement in showing me the right way to achieve my goals. So I just wish I could do the same till the end of time for your education and life.

I also take this opportunity to express a deep sense of gratitude to Mother who have given me her valuable support and guidance, which helped me in completing this task through various stages.

I take this opportunity to express my profound gratitude and deep regards to my guide Dr. Solomon Atnafu for his exemplary guidance, monitoring and constant encouragement throughout the course of this thesis. In addition I would like to give my profound appreciation for Dr. Tsegeye Taddese and Ato Getachew Birhanu for their unlimited support; it is their help and guidance which carried me a long way in this research.

# Abstract

Drought is a condition of moisture deficit sufficient to have an adverse effect on vegetation. When such event occurs it leaves a devastating effect on the community. This is mainly due to the lack of prediction and mitigation efforts. Thus predicting this event should be carefully considered. Although, there are many attempts to predict drought using satellite images, none have come across in using a higher resolution satellite images. These type of images have a significant role in developing a better model and a more reliable drought prediction system.

This study aims in analyzing the effect of using higher spatial resolution satellite images in drought prediction systems. The first step we followed towards this objective is to consider the for drought indicating attributes that were identified in earlier research for the study area (Ethiopia). Out of this attributes, the satellite image (showing the actual condition on the ground) shows a higher coloration to drought incidents. The second step is to process the acquired data and produce the appropriate model. In this study, producing the model is typically generating models. The models, if satisfied, will help in classifying a condition as dry, wet or normal.

Finally, we have produced a system that uses the model developed to predict drought. The system produces a map showing the actual conditions on the ground. After going through all the steps specified, we showed that higher resolution satellite images help in producing a more accurate model as compared to the previous studies on the same study area.

Key words:

Drought Prediction, Satellite Image for Drought Prediction, Normalized Difference Vegetation Index (NDVI), Standard Seasonal Greenness (SSG).

# Acronyms

**AMO**: Atlantic Meridional Mode

**DEM**: Digital Elevation Model

**LC:** Land Cover

**MEI:** Multivariate ENSO Index

**NAO:** North Atlantic Oscillation

**NDVI:** Normalized Vegetation Index

**PDO**: Pacific Decadal Oscillation

**PNA:** Pacific North American Index

**SG**: Seasonal Greenness

**SPI**: Standard Precipitation Index

**SSG**: Standardized Seasonal Greenness

**USGS**: United States Geological Survey

**VEG_Ethiopia** : Vegetation Ethiopia

**WHC**: Water Holding Capacity

# Table of Contents

**Page No.**

# Chapter One: Introduction

Almost every aspect of our live depends on plants. They feed us, cloth us, absorb carbon dioxide, provide us with oxygen, and give us building materials and medications. When drastic changes occur to the vegetation around us, our health, economy, and environment are all affected.

Twenty-seven years ago, for instance, thousands of people starved when the vegetation in the Welo and Tigray regions of Ethiopia dried up during an extended drought. Similarly prolonged drought in the horn of Africa has brought an immediate cause of the severe food crisis affecting around twelve million people in parts of Kenya, Ethiopia, Djibouti and Somalia [30].

To predict such events prior to their occurrence different attempts were made, for instance, in Ethiopia drought assessment and monitoring efforts have been based on conventional methods that rely on the availability of meteorological data, which is very tedious and time consuming to collect. Moreover, meteorological data and weather information analysis and dissemination is also a challenge as it requires time and lots of resources.

In recent years, data from satellites such as Normalized Difference Vegetation Index (NDVI) has been used for drought monitoring [2]; however this technique considered a spatial resolution of 8km which makes it difficult to get detailed and more precise information. Bering in mind most of the farm lands in Ethiopia are individualized and small plots of lands are dominant, high resolution image is required to study the local environmental changes. Considering this farm land sizes of the local farmers in Ethiopia, the use of low resolution satellite images may have limitations to predict an environmental degradation and would result with inadequate information. That is why this research will use higher resolution satellite images (250m spatial resolution) to predict drought incidents and see the effect of high-resolution images in decision making.

## 1.1 Statement of the Problem

Previous studies have provided us with a mechanism to oversee drought incidents prior to their happenings and the significance of these works is unquestionable. Specifically [2] have showed that it is possible to predict drought using satellite images. Furthermore it has also showed that it is possible to characterize drought as an object.

Although many works used satellite images for drought prediction [1, 2, 31, 32], due to the lack of accumulated historical data most of them used low resolution satellite images. Hence the effect of using high resolution satellite images to predict drought has never been tested and the extent of precession or improved prediction is not known. Recent technological developments enabling us to acquire high resolution satellite image has now created the opportunity of exploring this effect for drought prediction.

Considering the condition in Ethiopia where the farmlands are localized and small plots are prevalent, exploring the effect of high-resolution satellite images becomes more critical. Hence this study will try to see the effect of using a higher resolution satellite for a better and detailed prediction.

## 1.2 Objectives

The general objective of this research is to analyze and determine the effect of using higher resolution (250m resolution) satellite images from the MODES satellite to predict drought and design a drought prediction system.

Furthermore, this research will include the following specific objectives:

1. Review related work on satellite image based drought perdition systems.
2. Identify the peculiar characteristics of drought and environmental changes in the Ethiopian condition.
3. Identify and acquire appropriate high resolution satellite images to the intended study area.
4. Compare results with previous works and develop a prototype which will enable end users with minimal computer skills to predict drought.
5. Conduct experiments to draw conclusions.
6. Draw conclusions and recommendations.

## 1.3 Scope

The scope of this research is to apply a higher resolution image and come up with detailed information regarding drought prediction and environmental changes. Since, previous works have already been done on metrological data analysis we will not engage on such activity due to limitation of time and emphasis of our work.

Although, the principles employed in this study could be applied to different study sites, the prediction and the satellite images analysis will be on Ethiopian conditions only.

## 1.4 Methodology

To achieve the objectives stated above we believe that a deeper understanding of the problem is required. The second step is acquiring data which is an input for the experiment. The third stage of the experiment is processing the acquired data and spatial and extent correction. Prior to experimentation we will develop a model that will be used to conduct the experiment; then finally we will show the validity of this research by comparing the result with the previous research.

### 1.4.1. Literature Review

This phase is one of the crucial steps that were taken to get a deeper understanding of drought prediction using satellite imagery in combination with other metrological data which will contribute to a precise prediction. In doing so, we have reviewed similar researches that attempted to model drought with lower resolution satellite imagery.

### 1.4.2. Data acquisition

The data needed for this research is biophysical and climate data. To make our judgment unbiased at the end of the experiment we have accumulated a similar data with the other related work [2], in an attempt to see the effect of introducing higher resolution imagery.

### 1.4.3. Data Processing

The acquired data for this experiment need to be organized in such a way that they would fit spatially, for this purpose we will correct the spatial extent of the images. In addition, the NDVI value will be converted to SSG (Standardized Seasonal Greenness), which will show the vegetative conditions on the ground.

### 1.4.4. Model Development and Experimentation

Having all the pre-processed data at hand, the next stage is to train the machine in order to produce models. These models will be letter used to map the relationship between drought indicating attributes and predict future events. The model will be prepared using the cubist regression tree software.

In the experimentation phase we predict future seasonal greenness of the study site and weight them against the actual values exhibited on the ground. This will again be weighed against other research which was done in the same study site using lower resolution satellite image.

The rest of the document is organized as follows: The second chapter is the related work section; this section will discuss the major research works done in drought prediction and indicate the basis that motivated this research. Here we will focus on satellite based approaches as they are highly relevant to this paper.

The third chapter was prepared to discuss the design of drought prediction system; here we will detail the plan that is implemented to conduct this research. In addition this section will discuss the main steps engaged up to the creation of the model.

Following the design of drought prediction system, the fourth chapter will include the implementation; in this section of the paper we will produce a map showing the drought conditions in Ethiopian context. The fifth chapter is the evaluation of the model. At this stage we will test the predicted drought events against the actual values exhibited. This is done to confirm the validity of the research. In the final chapter we will conclude the findings of our work and give recommendations for future works.

# Chapter Two: Literature Review

Different studies have attempted to look for ways to predict when and where extreme weather events will happen, and how severe they will be. For some extreme events, like tornadoes, hurricanes, and floods, researchers have discovered patterns and methods that allow us to watch and even predict bad weather or hazardous conditions shortly before they happen.

Drought is a different kind of extreme weather event. Unlike tornadoes, hurricanes, and floods, drought doesn't have a clear beginning or ending. It is hard to tell when a drought begins and ends [23]. But, by watching various “indicators” of drought, like water levels in streams, soil moisture, or the amount of rainfall an area has received, we can keep track of droughts.

Predicting drought helps the concerned party to make informed decisions. For example, for farmers, the early prediction of drought helped them shape their production in a way that they can adapt to this event or even migrate if situations are worse, and for the government prediction helps prioritize aid and support to be allocated and know the extent of it before it leaves its devastating effect. Such a system not only helps the government in achieving a better distribution of resources but also helps saving lives.

## Earlier drought prediction approaches

Different efforts have been exhibited in predicting drought event prior to its happening [1,2]; in this research we will try to discuss these approaches in their chronological order.

In earlier days the cause of drought was believed to be the lack of rainfall. Thus, in an attempt to predict drought the normal trend was to measure the amount of rainfall in that particular area. After doing so the current value will be weighed against the historical data and if anomalies occur that was believed to be an indication of drought or flooding. For the purpose of discussing researches related to this trend we will refer them as “traditional climate based approaches”.

Recently, researchers have devised a new way to predict drought. This approach utilizes satellites to take periodical image of the ground. Following this image specialists try to predict drought comparing the historical images to what is seen on the ground. This approach totally relies on what is observable on the ground. In this study we will call this scheme “Satellite Image based approach”.

Before discussing the third technique of drought prediction, we will try to shed light on some of the contributing factors that are believed to cause drought.

It is already known that drought occurs when enough rain does not fall into the ground. However, specific to that surrounding different attributes or indicating factors could be accounted for causing drought. For example, the water holding capacity of a particular soil in that region would play a great role in leaving surrounding wet or dry.

Incorporating the major contributing factors for that specific region, a new approach was created. Here we will refer to such an approach a "Hybrid approach", since it integrates the rain fall data and other attributes that are believed to be the major contributing factor for the cause of drought along with what is observed on the actual ground (satellite image).

In the sections to come we will highlight some of the efforts attempted in each of the three drought prediction schemes.

### 2.1.1 Traditional Climate Based Approach

In the early days of drought prediction primary source of data has been meteorological data. Climate-based drought indices are often used to support drought planning decisions and to trigger mitigating actions in different parts of the world [27, 28].

The main data source of climate based indexes is metrological stations. These stations only provide point data and thus estimation methods like kringing should be applied to calculate the values in between stations. But since weather stations are barely available in sparsely populated areas, the calculation would be even difficult and inaccurate.

Climate indexes typically characterize the intensity of dryness as compared to the long-term average or normal condition and are usually calculated from one or more of the following variables: rainfall, temperature, snow pack, stream flow, soil water holding capacity, and other water supply indicators [1].

In the traditional climate based approach different drought indicating indexes could be used. Indexes such as, Stream flow Drought Index (SDI), Standardized Precipitation Index (SPI), Standardized Runoff Index (SRI), Palmer Drought Severity Index (PDSI), and Surface Water Supply Index (SWSI). How these indexes are calculated and their significance relative to this work is detailed in the coming sections.

The SPI was first developed in [7] to quantify the precipitation deficit for multiple time scales. Computed on the availability of water resources on the vicinity relative to the historical precipitation that area receives, the SPI's time scale reflects the impact of drought on the availability of the different water resource(Here water resource could mean: water reservoirs, rainfall, etc.).

The key advantage using SPI is that it can be calculated for different time scales, showing the severity and period of drought on the water supply. In addition the calculation of SPI is simple, as long as a reliable and long term rainfall data is accumulated (usually a minimum of 25 years to be exact).

As suggested in [7], the SPI calculation can also be applied to other water variables, such as soil moisture, snowpack, stream flow, reservoir and groundwater. The SDI developed in [24] and the SRI developed in [25] has computation procedures very similar to that of SPI. The difference between SDI and SRI is that the SDI uses observed stream flow data, while the SRI uses simulated runoff data from hydrological models [26].

The Palmer Drought Severity Index (PDSI) developed in [9], is widely used to measure the severity of agricultural and general drought . The U.S. Department of Agriculture uses PDSI to determine when to grant emergency drought assistance [27]. The PDSI is based on the water balance concept and takes into account temperature, precipitation, soil moisture, runoff and other climate and hydrological properties. Simply stated, PDSI uses temperature data to calculate potential evapotranspiration (PE) from the Thornthwaite method, and then uses precipitation and PE as inputs to compute basic hydrological cycle components, such as evapotranspiration, soil moisture, and runoff [26].

### 2.1.2 Satellite Image Based Approach

As technological innovations start to emerge, decision makers in many countries started to use remotely sensed data or data from satellite sensors. These can provide continuous datasets that can be used to detect the onset of a drought as well as its duration and magnitude [29].

Before stating some of the works related to remotely sensed drought prediction schemes, this paper tries to identify the three major kinds of drought, namely Metrological Drought, Hydrological Drought, and Agricultural Drought.

Figure 2.1 was taken from [19] to show these different kinds of drought.

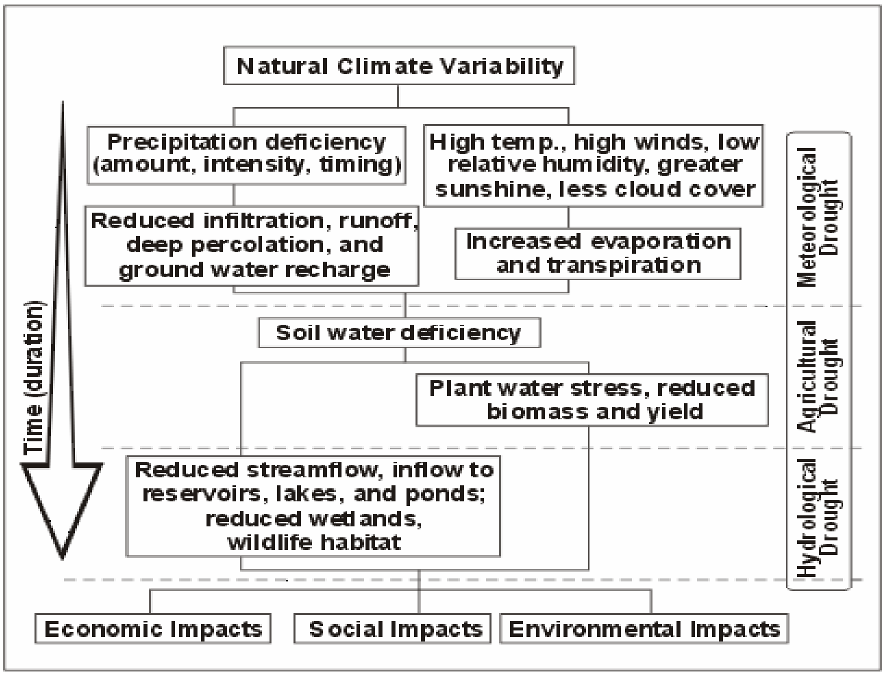



Figure 2.1: Influence of precipitation deficiency and other factors on drought development (National Drought Mitigation Centre)

The first type of drought is “Meteorological drought”; it is generally defined by comparing the rainfall in a particular place and at a particular time with the average rainfall for that place [14]. Meteorological drought leads to a depletion of soil moisture and this almost always has an impact on crop production.

When we define drought this way, we only consider the reduction in rainfall amounts and don't take into account the lack of crop production, lack of water on water reservoirs. Such drought could be easily determined by analyzing the historical precipitation index on the vicinity, and thus satellite imagery showing the vegetative conditions play a less significant role. Thus, to predict Metrological drought the traditional drought prediction scheme is efficient.

The second type of drought is “Hydrological drought”; it is associated with the effect of low rainfall on water levels in rivers, reservoirs, lakes and aquifers [14]. Hydrological droughts usually are noticed some time after meteorological droughts. First precipitation decreases and, some time after that, water levels in rivers and lakes drops.

Hydrological drought affects uses which depend on the water levels. Changes in water levels affect ecosystems, hydro-electrical power production and recreational, industrial and urban water use.

The traditional rain fall monitoring scheme combined with a periodical measurement of lakes and water reserves depth is efficient in predicting hydrological drought. This is true because this kind of drought is ought to occur only when the water levels are below normal or less than the regular trend.

The third category of drought is “Agricultural drought”. Agricultural drought occurs when there is not enough water available for a particular crop to grow at a particular time. This drought doesn’t depend only in the amount of rainfall, but also on the correct use of that water.

Agricultural drought is typically seen after meteorological drought (when rainfall decreases) but before a hydrological drought (when the water level in rivers, lakes and reservoirs decreases).

We can imagine a period of low rainfall where water is used carelessly for irrigation and other purposes. Under these circumstances, the effect of the drought becomes more pronounced than it was before.

Predicting agricultural drought has always been a challenge. One explanation for this is that agricultural droughts are a slow-onset, or "creeping," natural disaster, developing over months and years, and frequently they exist before it is realized. In general sense we can forecast such events by observing what is on the ground, this is done by comparing the historical NDVI , “greenness” value to the current vegetative condition .

In an attempt to predict the environmental response towards drought, [15] used only NDVI data for predicting drought. Although this research is mainly focusing on the effects of environmental change (drought) on animal populations, the approach used for prediction is solely based on NDVI values acquired from satellites.

### 2.1.3 Hybrid approach

Drought prediction in different study sites differ for its approach, since some of the contributing factors differ in many study sites and some of them are not even applicable in other sites.

The Hybrid approach incorporates other relevant indicators that contributes to the occurrence of drought with satellite Images.

The next section deals with the related works and the specific attributes that are chosen for this research.

# Chapter Three: Related Work

A new drought monitoring technique called the Vegetation Drought Response Index (VegDRI) was introduced in [1]. This technique integrates historical climate data and satellite-based earth observations with other biophysical information (for example, land cover, land use, and soils) to produce a 1 km–resolution indicator of the geographic extent and intensity of drought stress on vegetation. This approach builds on the traditions from both the climate and remote sensing communities for drought monitoring and utilizes new data mining analysis techniques to identify historical climate-vegetation relationships related to the drought phenomenon. The overall methodology taken in this work is indicated using Figure 3.1 taken from [1]

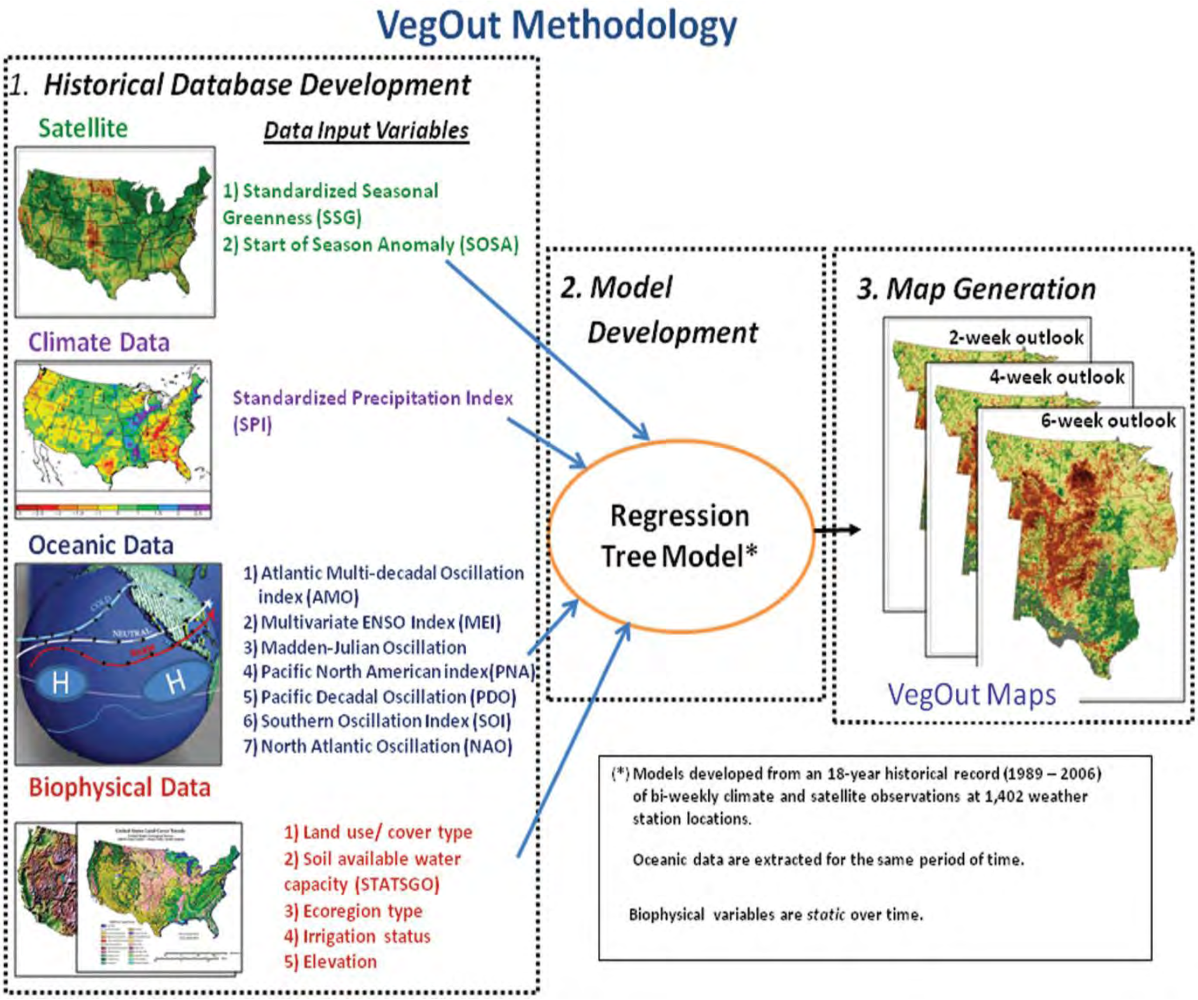



Figure 3.1: A New Integrated Approach for Monitoring Drought Stress in Vegetation Methodology

A research which relates to this study is the work in [2], which was conducted for the whole of Ethiopia. Although, the main aim of the work was to characterize drought as an object, this study follows the same approach with our work, in that it implemented a hybrid approach, incorporating eleven attributes.

Another thing that made our interest compelling is the fact that [2] used a very low resolution (8 km) satellite image. The work in [2] implemented two different approaches in attempt to compare the two methods. These are Artificial Neural Network (ANN) and Regression tree approach.

The main aim of this research is to predict drought using a hybrid approach by incorporating drought indicating indices in the prediction mechanism and see the significance of using a 250m spatial resolution satellite Image. In the next section of this document we will discuss how the study site was chosen and the drought indicating attributes with their relation to drought in the study site.

### Study Site

When predicting drought the attribute selection processes and all other experimental procedures considered needed to be tailored for that test bed. The study site in [2] was Ethiopia; Ethiopia has diversified climates, on average Ethiopia receives 1089 mm of precipitation annually. There are three seasons in Ethiopia. From September to February is the long dry season known as the “bega”; this is followed by a short rainy season, the “belg”, in March and April. May is a hot and dry month preceding the long rainy season “kremt”, in June, July and August [3].

Nowadays the “Keremt“ season is extending from August to October. Following the work in [2], this research will be conducted in the rainy seasons including the two months September and October to see the anomaly of rain fall in the region and predict whether a drought is approaching or not.

## 3.1 Brief Overview of Drought Indicating Attributes

In the coming sections this paper will try to highlight the significance of incorporating the eleven drought indicating attributes.

### Water Holding Capacity (WHC)

It is highly understood that water is the major limiting factor of plant growth; accordingly if plants do not receive adequate rainfall or irrigation, plant growth will be highly affected. It is inevitable that the absence of rainfall or irrigation for a prolonged period of time sufficient to deplete soil moisture and injure plants will cause drought. If the severity of dryness in the soil exceeds the ability of the plant's roots to absorb water, drought will occur. Water holding capacity is defined as the amount of water that can be held in a particular type of soil against the pull of gravity [16]. The weaker the WHC the lower the amount of water available for plants to use. That is why WHC is included as one attribute to indicate drought occurrence.

## Land Cover

Conversion of natural habitats to human dominated landscapes has led to worldwide deterioration of ecosystems [6]. In these regard the physical material at the surface of the earth is termed as land cover. Land cover shows how green or how sandy material covers the surface of the earth , i.e. on one hand the greener the local area the higher the precipitation and thus causing a wetter environment on the other hand the sandier the land coverage the drier the surrounding, consequently the higher likelihood of drought event to occur.

## Digital Elevation Model (DEM)

The Digital Elevation Model (DEM) of a certain area describes the 3D mapping of the surrounding. In general, highland areas are considered to be wet and low land areas are presumably drier. In our situation this is significantly true, for example if we consider one of the tallest mountain in Ethiopia, mount Ras-Dashen and its surrounding it have a relatively wetter environment, on the other end the Dallol depression i.e. the lowest point in the country , is also the driest. Taking this in to account DEM was included to indicate the occurrence of drought event.

## Vegetation type of Ethiopia (Veg_Ethiopia)

Plants life could be classified as wet land plants or arid land plants. Specific to drought on arid lands of the world plant life have been classified as: 1) *drought escaping,* by completing their life cycle in a very short time when growing conditions are favorable, after which they become dormant for the remainder of the year; 2) *drought evading,* by remaining small or restricting growth when moisture is limiting; 3) *drought enduring,* plants which may grow very little or not at all for an entire year or more, yet remain alive to renew growth when rain arrives; and 4) *drought resisting,* by withstanding arid conditions by accumulating water in the plant as a stored reserve. There are many drought resisting plants in grasslands of the world that renew growth following dormancy even though there is no apparent soil water available

It is highly anticipated that drought is the major cause of the extinction of plant life. This is to mean that, if drought is to occur in a specific surrounding the plant life is likely to suffer abundantly. In another perspective, if the surrounding is filled with plant of arid regions the likelihood of drought to occur is less, and thus the type of plant life could be one of the scaling factors for a drought occurrence, even if the surrounding is dry.

As discussed above the response of plant life towards lack of rain fall varies greatly, as some plants tend to survive a prolonged time without water. Taking this in to consideration, the vegetation type of Ethiopia (Veg_Ethiopia) was included to indicate drought as it occurs.

### Oceanic Index

Covering three fourth of the earth and absorbing twice as much of the suns radiations than the land cover, oceans are believed to highly influence the earth's climate system. Transporting as much heat currents as the atmosphere, oceans play a significant role in shaping the atmospheric conditions on the earth. But since these oceans are closer to the land mass the effect of ocean tides is much more localized to the atmosphere. Thus, the oceanic tides that affect some region may not even participate in shaping the climate of other land mass.

Following this, our study carefully examines specific oceanic indexes that contribute much to affect the climate of the study area. In doing so five oceanic indexes (i.e. PDO, PNA, MEI, AMO and NAO) are believed to have a direct contribution to the climate of the study area. The description on the selected oceanic Index is located on the data archive site [12].

### MODIS Satellite Data Inputs

MODerate-resolution Imaging Spectro radiometer (MODIS) on board NASA's Terra satellite provides -series of Normalized Difference Vegetation Index (NDVI) observations that can be used to examine the dynamics of the growing season or monitor phenomena such as droughts. Here we will highlight the core data input, NDVI which serves as the basis for the prediction of drought. As this data gathered from the satellite shows the actual events (vegetation change) on the ground, it is understood have a high correlation to a drought event. In the paragraphs to come we will shade some light on how this NDVI value is derive and how relevant this index is to a drought event.

The light emitted from the sun has different spectral regions. In our specific study we can classify these spectral regions in to two based on their significance to the plants photosynthesis process, namely photo-synthetically active radiation and near-infrared spectral region.

Live green plants absorb solar radiation in the photo-synthetically active radiation (PAR) spectral region, which they use as a source of energy in the process of photosynthesis [8]. The main reason for plants to absorb this light is for its vital importance as a source of energy to drive their photosynthesis process.

On the contrary leaf cells have also evolved to reflect and transmit solar radiation in the near-infrared spectral region (NIR). In doing so plant leaves appear dark to scatter light energy in the near infrared – spectral domain since it only carries approximately half of the total incoming solar energy, and this energy level is not sufficient to be useful to synthesize organic molecules. A strong absorption at these wavelengths would only result in overheating the plant and possibly damaging the tissues.

Considering the work in [40] that introduced the index (NDVI) which shows the greenness of a surrounding by taking the average of the two wave lengths (i.e. absorbed and reflected by plants). This is computed using the equation proposed in [40] (*Equation 3.1*)

$$NDVI = \frac{NIR - PAR}{NIR + PAR} \qquad \textit{Equation: 3.1}$$

**PAR** Value of Photo-synthetically Active Radiation
**NIR** Value of Near-Infrared Radiation

Note that this NDVI value is retrieved from the data source in the form of a tagged image file (.tif) format by which it contains the stretched NDVI value at each pixel points. The NDVI value gives a measure of the vegetative conditions on the ground over wide areas. Dense vegetation shows a value between 0 and 1, and areas with arid vegetative conditions identified by negative values greater than -1.

## Climate Data Inputs

Drought is an extended period of months or years when a region notes a deficiency in its water supply, whether surface or underground water [17]. Generally, this occurs when a region receives consistently below average rainfall. Subsequently, the climate data attributed for this research is rainfall, since the lack of rain fall could be the direct cause of drought.

It should be noted that this rainfall data could not be used in its crude form for the prediction of drought, since it will be very difficult to compare rainfall events for two or more different areas in terms of drought, as drought is really a "below-normal" rainfall event. And what is "normal rainfall" in one area can be surplus rainfall in another area.

To accommodate this The Standardized Precipitation Index (SPI) is used for monitoring drought. It allows us to determine the rarity of a drought at a given time scale of interest for any rainfall station with historic data. It can also be used to determine periods of anomalously wet events.

Mathematically, the SPI is based on the cumulative probability of a given rainfall event occurring at a station. The historic rainfall data of a station is fitted to a gamma distribution, as the gamma distribution has been found to fit the precipitation distribution quite well. Therefore, based on the historic rainfall data, we can then tell the probability of the rainfall to be less than or equal to a certain amount.

SPI can effectively represent the amount of rainfall over given time scale with an added advantage to an indication of what this amount is in relation to the normal condition at that locality, thus leading to the definition of whether a station is experiencing drought or not.

### Summary

Although, many works have been conducted to predict drought events, none were done to assure the quality of the prediction mechanism. The recent trend to predict drought is to use the hybrid approach, following this satellite Images showing the ground conditions are integrated with other contributing factors. The need for including these factors in the study site has been clearly discussed in this chapter. Among the eleven attributes the one showing a higher correlation to drought is the satellite image.

At the present time, satellite images are being made with higher resolution. In this regard, satellites are providing images with a resolution up to 40 centimetres on the ground, but advantage of higher resolution satellite Image have never been quantified. This research is conducted to show the effect of higher resolution satellite images in drought prediction.

# Chapter Four: Design of Drought Prediction System

In this chapter, a comprehensive overview of the design of the drought prediction system with higher resolution satellite imagery and the materials that were used to conduct the research is presented. This design first starts from data gathering, this stage is mainly concerned with the acquisition of the appropriate data to be used for drought prediction system. The second stage is data pre-processing, this stage extracts the data from the African extent to the study site (Ethiopia), in addition NDVI value generation and coordinate system transformation are included. The third stage is data processing, at this point the drought indicating index SSG will be computed, also included at this stage is the point data generation; this is the step where we convert all the eleven drought indicating attributes to a point data. The Fourth stage is the model generation, generating the model is necessary to predict drought event and hence the output of this stage is used in the prediction step. Finally, comes the prediction, where we have developed an extension for ARCGIS which is used for prediction of drought events ahead of time. Before discussing the detailed paths taken by this research we would like to briefly highlight the design of the drought prediction system in Figure 4.1.

## 4.1 Data selection and acquiring

There are different situations that indicate the existence of drought, and thus could be used for the prediction of such an event. For example the lack of rain fall, the setting of the ground (elevation, and surrounding environment), the wind movement and oceanic tide are among the few. Considering all this and others, the work in [2] selected eleven attributes which characterizes best the study site.

In order to coincide the data inputs of this research with [2] and judge weather a higher resolution satellite image would result in a better prediction scheme we have also used the same eleven attributes. These attributes could be classified into two: Static Data and Dynamic Data

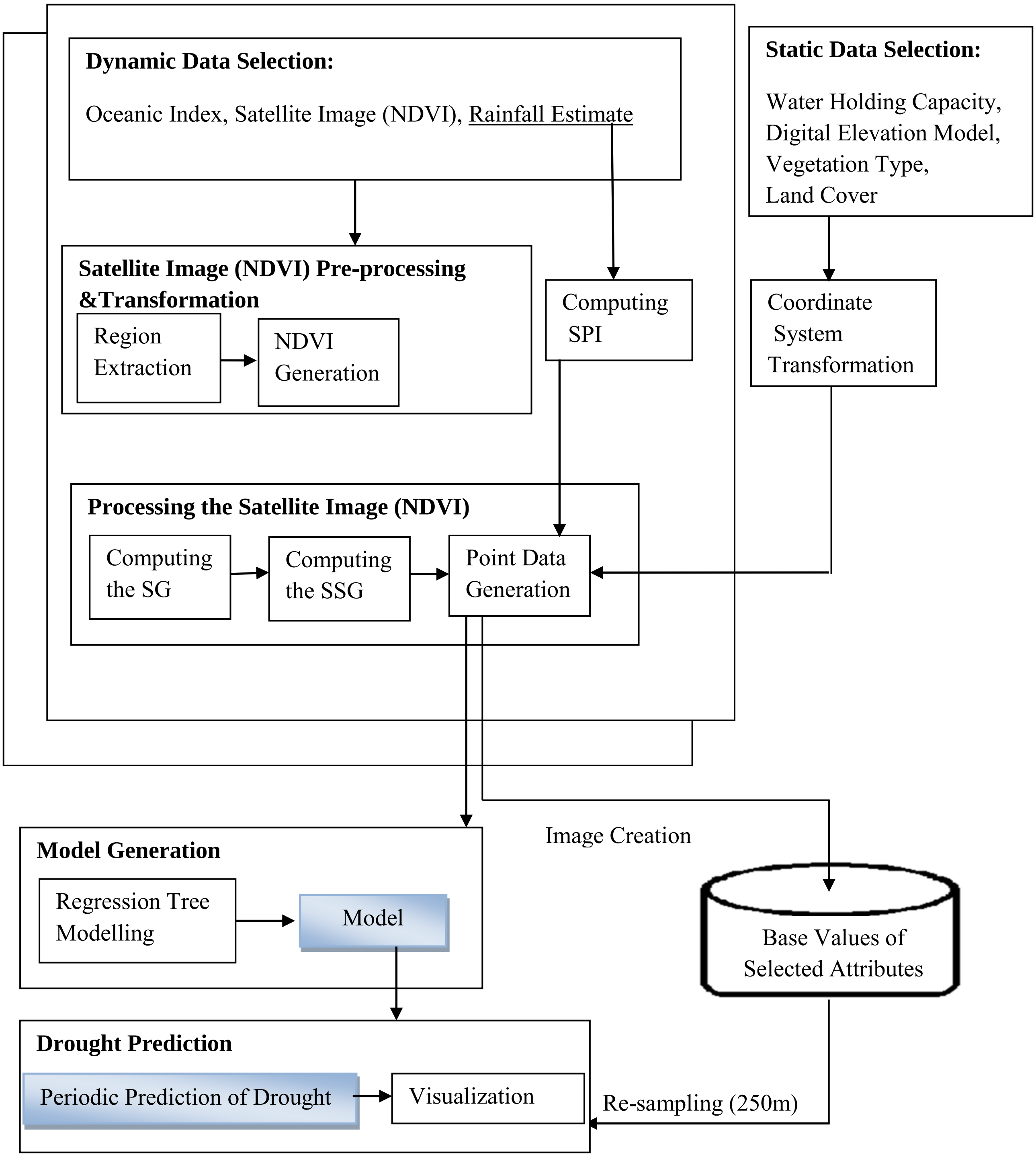



Figure 4.1: Design of drought prediction system

The static data inputs are termed as their value is consistent (non-changing) for the period of the research. Static data inputs included in this research are Digital Elevation Model (DEM), Water Holding Capacity (WHC), VEGetation type of ETHiopia (VEG_ETH), and Land Cover. These are called static since their value is not subjected to change.

On the other hand the dynamic data inputs are termed so since their value continues to change over a period of time. The dynamic data inputs included are: Oceanic Index (contains Five attributes), Standard Participation Index (SPI), Standardized Seasonal Greenness (SSG). These values change every month and values are carefully integrated in the research.

Whether static or dynamic the data inputs for this research which indicate the physical features on the ground could be collectively termed as biophysical data inputs. Next we will indicate the data inputs with a brief description of their relationship with drought as follows:

### Data Inputs

While selecting data inputs relevant to drought incident, we have incorporated both biophysical and climate data. Here as it will be discussed in the next sections all the eleven attributes are not obtained from one data source with the same resolution and geographical reference, thus data re-sampling and pre-processing were done accordingly. In the coming subsections we will describe the need to incorporate each of these attributes in relation to drought occurrence along with their data source.

Figure 4.2 presents the two categories of data inputs used in this work..

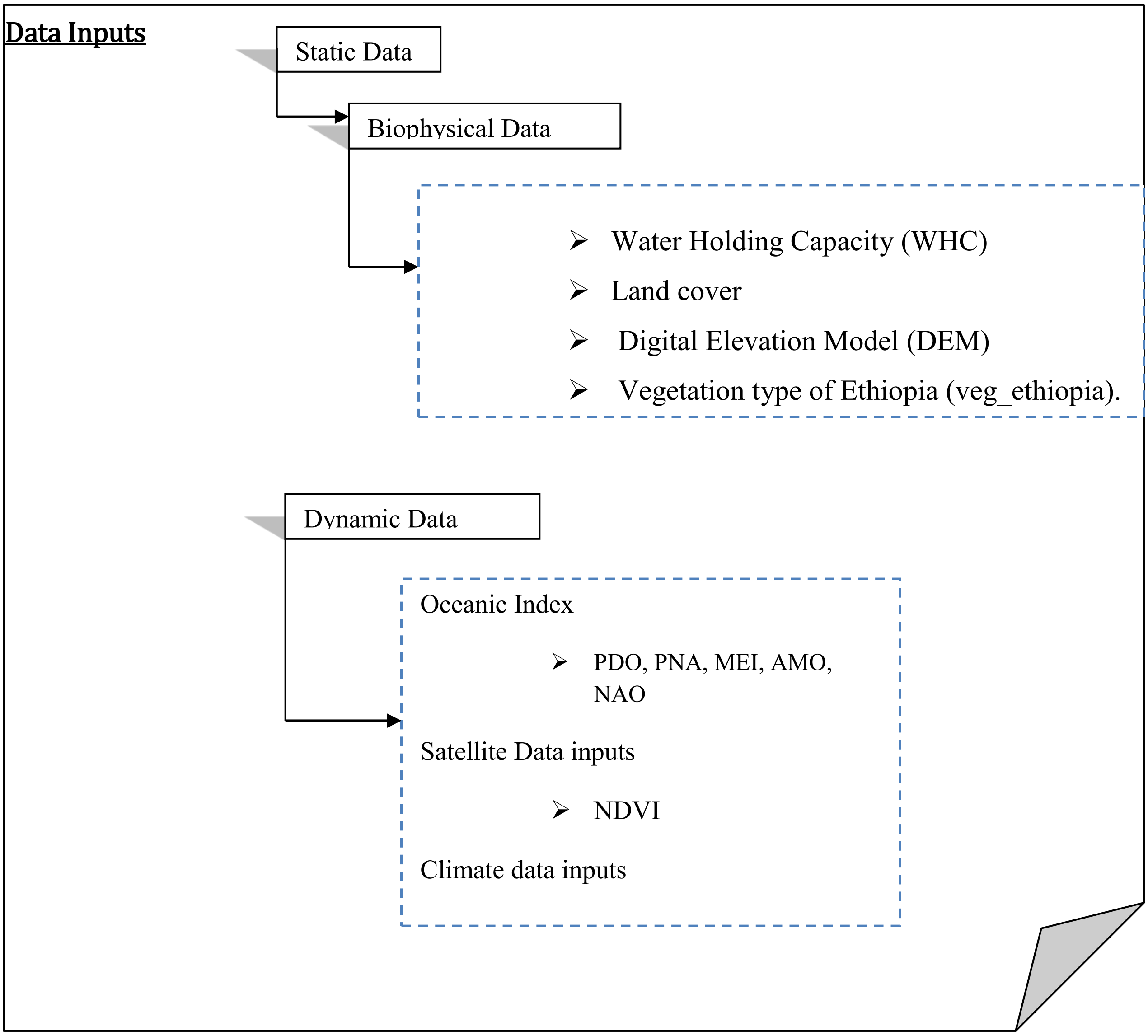



Figure: 4.2: The two categories of data inputs

### Biophysical Data Inputs

Some physical features on the ground like vegetation type, and water holding capacity are highly correlated to drought incident as they indicate moisture or the aridity on the ground. In this essence, the study has incorporated physical features such as: water holding capacity (WHC), land cover, digital elevation model (DEM) and vegetation type of Ethiopia (veg_ethiopia).

### Water Holding Capacity (WHC)

The WHC data was obtained from United States Geological Survey's National Center for Earth Resources Observation and Sciences (EROS). As the data was in 1Km resolution, re-sampling the data back to 250m in which the seasonal greenness was calculated was believed to have no effect and this data was used in its original format.

### Land Cover

The Land cover data was obtained from Global Land Cover Facility (GLCF) with a spatial resolution of 1km. This data was used in its original format since re-sampling any data from 1 km to 250m is not believed to add value to the original data.

### Digital Elevation Model (DEM)

DEM was obtained from USGS-Earth Resources Observation and Science (EROS) Centre with a spatial resolution of 1km. Then the data was projected and extracted to the Ethiopia extent. As discussed above the data was used in its original format since data re-sampling will not add any value.

### Vegetation type of Ethiopia (Veg_Ethiopia)

Veg_Ethiopia was obtained from Royal Danish Academy of Sciences. Originally, the data was in shape file format. For the purpose combining the data in the model development, the data was converted to raster data format.

### Oceanic Index

Five oceanic index are believed to have a direct contribution to the climate of the study area, (i.e. PDO, PNA, MEI, AMO and NAO). The description on the selected oceanic Index is located on the data archive site [12].

### MODIS satellite data inputs

The Historical NDVI images with a 250m spatial resolution from the Terra's satellite are used to monitor vegetation condition. This data can be downloaded from the USGS website at the following link [4].

### Climate Data Inputs

Integrating both ground and satellite data, the Tropical Applications of Meteorology using SATellite (TAMSAT) produces a higher quality compared to others, thus we have included this as our source for the rain fall data.

## 4.2 Data Pre-Processing and Transformation

After the acquisition of the data the step to follow is to pre-process the data for the construction of the model. To calculate NDVI values first we extracted the data only to the study site, then re-sampling and projections to their respective geographical locations followed. The overall step is indicated in Figure 4.3.

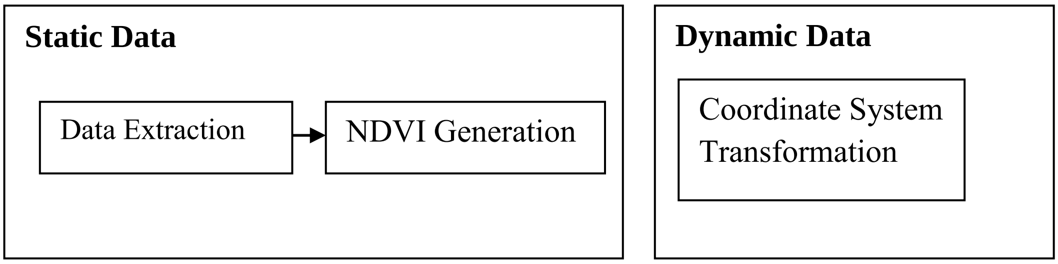



Figure 4.3: Data Pre-processing

### Image Pre-processing and Transformation

The calculated NDVI values are typically very noisy and are affected by a number of phenomena including cloud contamination and atmospheric perturbations. Thus data smoothing was required. To smooth the data a weighted least-squares approach proposed in [39] was used. The following paragraphs indicate how this data smoothing technique operates.

A time series smoothing technique developed in [39] is famous for its satellite data smoothing approach. This technique was applied to the NDVI composites for the years 2001 through 2006.It follows weighted least-squares linear regression approach to temporal NDVI smoothing to more efficiently reduce contamination in the NDVI signal. This technique uses a moving window operating on temporal NDVI to calculate a regression line. The window is moved one period at a time, resulting in a family of regression lines associated with each point; this family of lines is then averaged at each point and interpolated between points to provide a continuous temporal NDVI signal [39].

In general, the smoothing algorithm effectively corrects these erroneous NDVI values based on characteristics of the valid NDVI curve. Following this the NDVI values with poor quality due to clouds

Converting the original image between -1 and 1 was necessary since original image acquired was stretched between 0- 200 and any value above was considered as a bad value (i.e. values up to 255). Following the recommendation of USGS [40] the NDVI value was calculated using ARC MAP10 by the map-algebra utility under spatial analyst tool according to *Equation 4.2* proposed in [40]

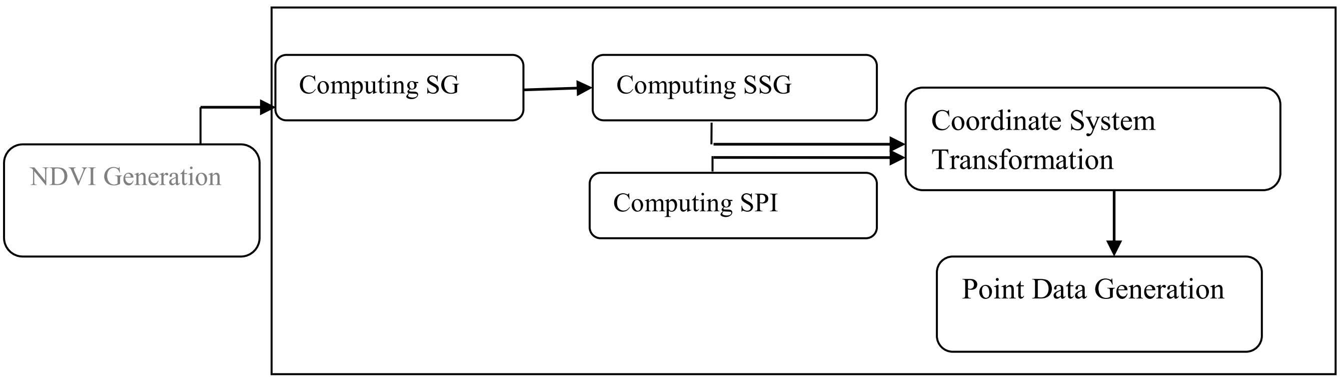
$$NDVI = (Value - 100) / 100 \qquad \textit{Equation } 4.2$$

According to [40] the images were 10 day composites, giving us only dakadal information. Since the research is based on Monthly NDVI values we have summed the three dakadal images to create monthly NDVI value. This was done using the Arc GIS 10 Map algebra tool with the *Equation 4.3* proposed in [2].

$$NDVI_{June} = NDVI_{June\_dakad1} + NDVI_{June\_dakad2} + NDVI_{June\_dakad\ 3} \qquad \textit{Equation 4.3}$$

### 4.3 Data Processing

The data that was gathered and organized now is ready to pass through the data processing phase. This stage begins by computing the Seasional Greeenness (SG) for the satellite images and continues to the computation of the Standardized Seasional Greenness (SSG). The SSG values along with the SPI and the rest of the nine attributes are then transformed to the same coordinate system so that they can fit togetther for prediction. The over all step of the data processing is indicated in Figure 4.6.




Figurq 4.6: steps of the Data Processing

## Climate Data Processing

## SPI Calculation

The main reason for the idea of creating a generalized Standardized Precipitation index by [7] was to create a standard way to over look the rainfall data and tell whether the surrounding is receiving the expected precipitation relative to the previously collected rainfall estimates. This Index primarily depends on the historical data and thus we have used the 25 years cumulative rain fall data.

Apart from its popularity SPI is based on precipitation only. It can be used on a variety of time scales, which allows it to be useful for both short-term agricultural and long-term hydrological applications. Generally speaking a drought event occurs any time the SPI is continuously negative and reaches an intensity of -1.0 or less. The event ends when the SPI becomes positive. Each drought event, therefore, has a duration defined by its beginning and end, and intensity for each month that the event continues.

The general steps taken in SPI calculation is highlighted as follows. First, we need to build a frequency distribution from the historical precipitation data at the study site for a specified period (1, 3, 6, 9, 12, 24 or 48 months). Then, a theoretical probability density function (i.e. gamma distribution) is fitted to the empirical distribution of precipitation frequency for the selected time scale. At last, an equi-probability transformation is applied from the fitted distribution to the standard normal distribution [7]. The detailed description of each step along with the Equation for SPI calculation is indicated as follows.

First, the gamma distribution is calculated by its frequency or probability density function using *Equation 4.4* proposed in [20] :

$$G(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}} \qquad \textit{Equation 4. 4}$$

Where:

$g(x)$: is probability function,
$x$ : is the precipitation amount,
$\Gamma(\alpha)$ : is the gamma function,
$\alpha$ : is a shape parameter and
$\beta$ : is scale parameter.

And parameters $\alpha$ and $\beta$ satisfy $\alpha > 0$ and $\beta > 0$.

For $\alpha > 0$ the gamma function is defined by *Equation* 4.4.1 as proposed in [20]:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha - 1} e^{-x}\, dx \qquad \textit{Equation } 4.4.1$$

Then, following the recommendation in [7] to optimize the shape and scale parameters ($\alpha$ and $\beta$) the maximum likelihood solution was used, which is described in [7] as (*Equation 4.5*):

$$\hat{\alpha} = \frac{1}{4A}\left(1 + \sqrt{\frac{4A}{3}}\right), \hat{\beta} = \frac{\hat{x}}{\hat{\alpha}} \text{ and } A = \ln\left(\hat{x}\right) - \frac{\sum_0^n \ln(x)}{n} \qquad \textit{Equation 4.5}$$

Where $n$ is number of precipitation observations.

Considering a condition where the time scale is short for example in the "Bega" season , the participation value X will have a Zero value, the above probability function is undefined .Thus, formulae 3.5 is modified as follows to include the condition where x= 0. (*Equation 4.6* as proposed in [7]).

$$H(x) = q + (1-q)G(x) \qquad \textit{Equation 4.6}$$

Where: q is the probability of no rain fall on the specified time scale.

Finally, the probability distribution H(x) is transformed into standardized normal distribution Z with the average equal to 0 and standard deviation 1; hence SPI is number of Standard deviations left (drought) or right (wet) from 0.

Z is calculated based on the cumulative probability distribution's value (i.e. for H(X) > 0.5 and for H(X) < 0.5). This is implemented using *Equations* 4.7 & 4.8 as proposed in [21].

$$Z = SPI = -\left(t - \frac{C_0 + C_1 t + C_2 t^2}{1 + d_1 t + d_2 t^2 +_{d_3} t^3}\right) \text{ For } 0 < H(x) <= 0.5 \qquad \textit{Equation 4.7}$$

$$Z = SPI = \left(t - \frac{C_0 + C_1 t + C_2 t^2}{1 + d_1 t + d_2 t^2 +_{d_3} t^3}\right) \text{ For } 0.5 < H(x) <= 1 \qquad \textit{Equation 4.8}$$

Where $t = \sqrt{\ln \frac{1}{(H(x))^2}}$ for $0 < H(x) <= 0.5$ and $t = \sqrt{\ln \frac{1}{(1.0 - H(x))^2}}$ for $0.5 < H(x) <= 1$.

$C_0$=2.515517,

$C_1 = 0.802853$

$C_2 = 0.010328$

$d_1 = 1.432788$

$d_2 = 0.189269$ and

$d_3 = 0.001308$

Table 4.1 shows the different classes of SPI proposed in [7].

Table 4.1: Standard Precipitation Index (SPI) Values and the associated drought Categories [7]

| SPI Values | Drought Category |
|---|---|
| **SPI>2.00** | Extremely wet |
| **1.50 to 1.99** | Very wet |
| **1.00 to 1.49** | Moderately wet |
| **0 to 0.99** | Mildly wet |
| **0 to -0.99** | Mild Drought |
| **-1.00 to -1.49** | Moderate Drought |
| **-1.50 to -1.99** | Severe Drought |
| **SPI<-2.00** | Extreme Drought |

While choosing the time scales in the SPI calculation we have observed to show a short term drought conditions [22] and thus we have used the three month time scale in the SPI calculation. During the calculation the whole process was automated using open source software.

### Seasonal Greenness Calculation

As the name indicates the seasonal greenness is a measure of growing season vegetation vigour or performance. This value is derived from NDVI. Typically the SG represents the collective NDVI above a non-vegetated surrounding from the start of the growing season to a specified time during the year. High SG values reflect high green biomass conditions, whereas low SG values reflect lower biomass levels [11].

The historical SSG to be calculated needs to have an appropriate SPI and the rest of the attributes to predict drought. But due to the lack of the RFE data (which is an input for SPI calculation) after the year 2006 we have constrained from calculating SSG after the year 2006. *Equation* 4.9 from [1] is used to calculate the SG values.

$$SG_m = \sum_{S_m}^{E_m} NDVI_m - NDVI_b \qquad \textit{Equation 4.9}$$

Where:

**$NDVI_b$**: the latent NDVI value = the NDVI value before the start of the growing month

**SGm** : Seasonal Greenness for a Month

**$NDVI_m$**: Monthly accumulated NDVI value

**$S_m$** : Start of growing month

**$E_m$** : End of growing month

### Standardized Seasonal Greenness (SSG) Calculation

In an attempt to standardize the SG to get a value which could give a measure of how the general vegetative condition on the ground is , we have calculated the SSG value. In doing so this value compares the specific year seasonal greenness to the 6 year historical mean seasonal greenness value. SSG is calculated using *Equation* 4.10 from [1] as follows:

$$SSG = \frac{SG - SG_{mean}}{\sigma} \qquad \textit{Equation 4.10}$$

Where: SG is the Seasonal Greenness

$SG_{mean}$ is the average/Mean of the six years Seasonal Greenness

σ is the Standard deviation.

### Point Data Generation

All the relevant attributes that contribute to the occurrence of drought need to be converted to a point value. A total of 2812 points were selected that are spatially distributed to cover the whole of Ethiopia. Accordingly, every attribute's value at each of the 2812 point coordinates will be organized and overlapped together for the creation of the model. For this purpose we have chosen a coordinate system (i.e. Clarke_1866_Albers) using the projection attribute, false easting: 0.00, false northing: 0.00, central meridian: 20, standard parallel_1:-19, standard parallel 2: +21, latitude of origin: 1 and linear unit: meter.

Since the oceanic index values were acquired as a single value for a month period throughout the whole of Ethiopia, a constant raster needed to be created covering the extent of Ethiopia. Using the Arc Map create constant raster tool we have created the constant raster with Clarke_1866_Albers coordinate system.

Having all the attributes in an image format, we have transformed the coordinate system of all the attributes to Clarke_1866_Albers so that all the images will fit geometrically to represent the same point value. Then we have overlaid a point data mask that composes the sample 2812 points with the same coordinate system as the rest of the attributes. Then the point values along with the stations locations (XY coordinates) were calculated. The whole process was automated using the batch processing.

### 4.4 Model Generation

For this study we have chosen the Cubist Regression tree software for creating the model. Thus, the data that we have as a point data needs to be organized based on the specification of this tool. To build a model using cubist software three different files are essential.

The first essential file is the names file (e.g. **`june1MonthAhead.names`**) that defines the attributes used to describe each case. For this specific research there are two important subgroups of attributes. The first type is the enumerated types, where the values in the actual data file are predetermined and given explicitly in the names file. These attributes are Land-cover and veg_Ethiopia, thus the values they can assume are explicitly indicated. The other type of attributes is those which have variable values. Such attributes are defined as continuous.

The second essential file, the application's data file (e.g. **`june1MonthAhead.data`**), which provides information on the *training* cases that Cubist will use to construct a model. The entry for each case consists of one or more lines that give the values for all explicitly-defined eleven attributes. Values are separated by commas and the entry for each case is optionally terminated by a period. Anything on a line after a vertical bar is ignored, for the data file we have dedicated 80% of the total data.

The third file we have created for checking the accuracy of the model, is the test file (e.g. **`june1MonthAhead.test`**). This file is the same as the .data file, but only contains 20% of the data. Following the Cubist software we have kept all the file names are kept the identical. A screen shot of cubist software is presented in Figure 4.7.

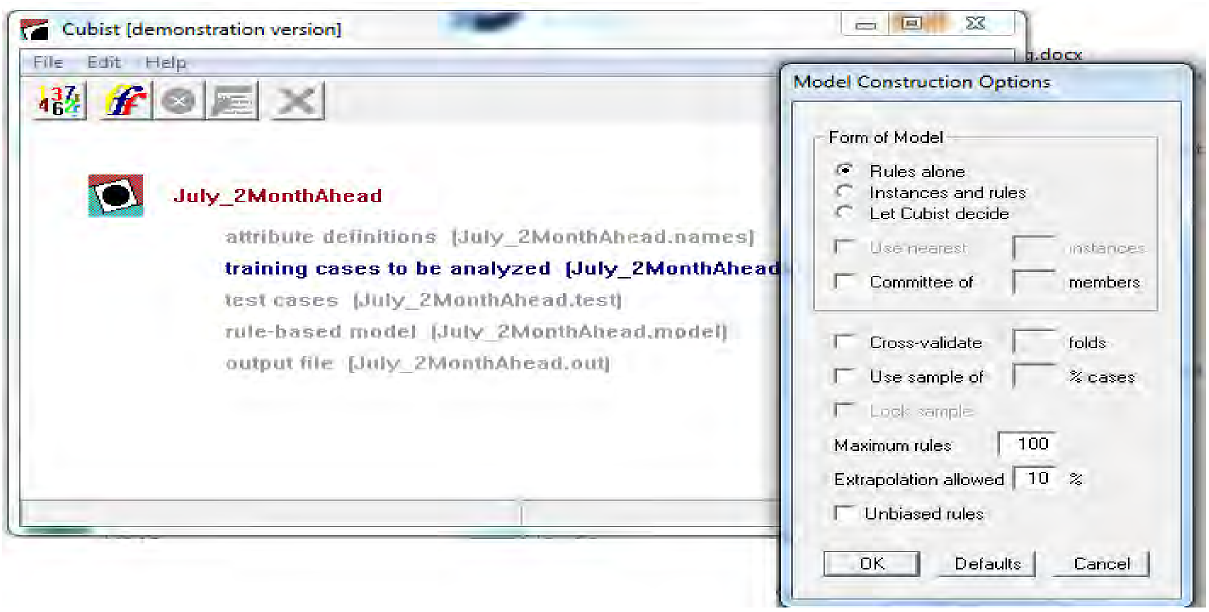



Figure 4.7: Screenshot of Cubist software - generating the model

## 4.5 Model Implementation

After all the necessary files are properly organized the cubist software is loaded with the .data file for each model to be generated. Then, using the build model an appropriate model was built. When building the models we have chosen the rules to be generated along with the instances.

Typically a good model is ought to have more than one rule for a dispersed data. In our case, average of twenty rules was produced. In addition, all rules were generated based on all the eleven attributes. The promising result acquired is briefly discussed in the Model Evaluation and Testing section below.

### Model Evaluation and Testing

The primary purpose of creating a model is to generate rules. That indicates a condition which must be met in order to classify a particular case into an appropriate class. In technical terms, a rule is composed of a linear formulae and whenever a case satisfies all the conditions in that formulae; a value is assigned to it (a predicted value is assigned) [13]. In general these rules are created to foresee the patters in the historical data and predict future events.

Although, the number of rules that are appropriate for a particular data set is difficult to quantify, in general, the existence of more than one rule is an indication of a superior model. As we can check in the created model this research has twelve rules on average, whose linear formulae is derived from all the eleven attributes. This alone can show the correlation between the attributes chosen for this research.

To judge whether the model generated is appropriate, the main thing that needs to be considered is the accuracy of the prediction. In order to prove the accuracy of the model prior to using it, we have predicted values of Standardized Seasonal Greenness (SSG) which are already known, i.e. being at June ,2002 we have predicted the SSG values up to four months ahead (July, August, September, October). This way we can see the true positives and true negatives (i.e. *hits*), and the false positives and the false negatives (i.e. *misses)*. This can be all aggregated to a numerical value and create a correlation coefficient index. Ideally this index needs to be close to 100%, if the model hits outweigh the Misses.

Here, as the prediction extends to the future , i.e. when we try to predict three of four months ahead instead of one or two month ahead, the value of the correlation coefficient decreases significantly. One of the things that could indicate that this research is accurate is the accuracy of the prediction scheme. Thus, in Figures 4.8-4.11 we have compared the result of our work that uses high-resolution satellite images to the work in [2]. On the Left is the model of our work and on the right is the model of the work in [2].

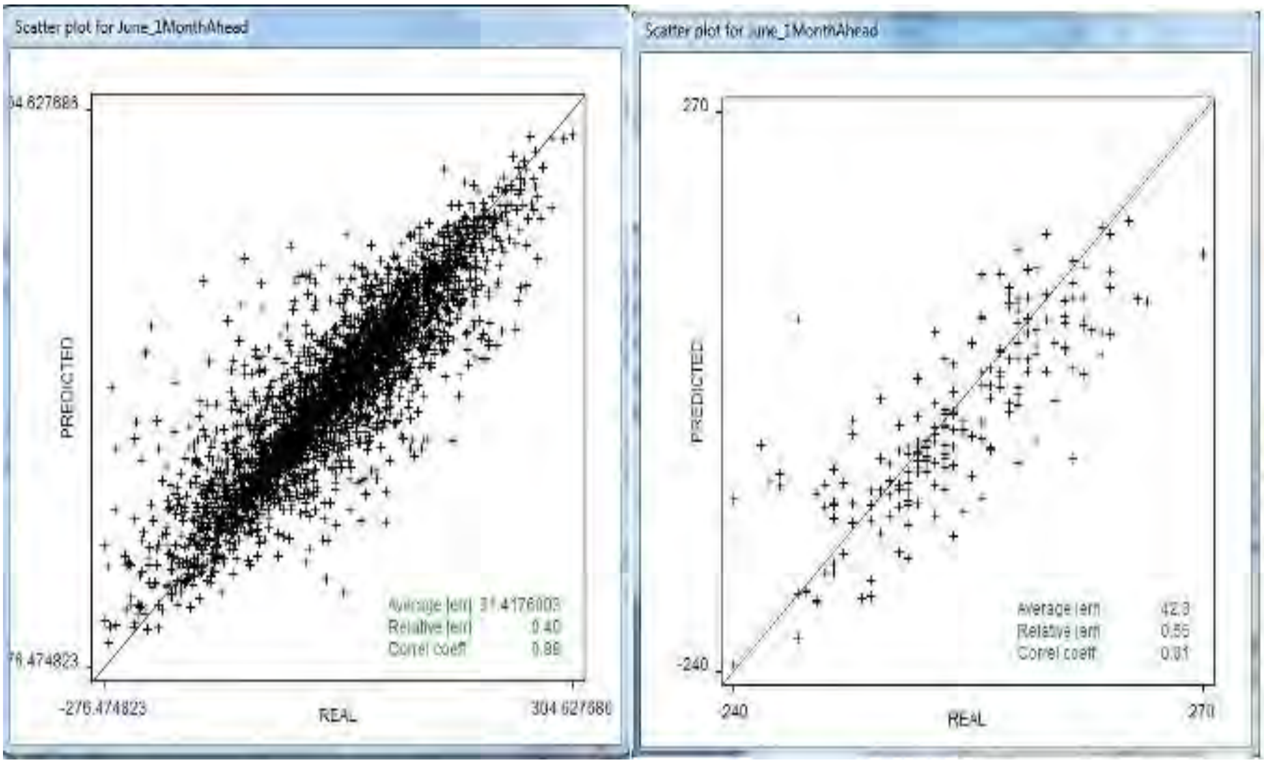



Figure 4.8: June one month ahead prediction

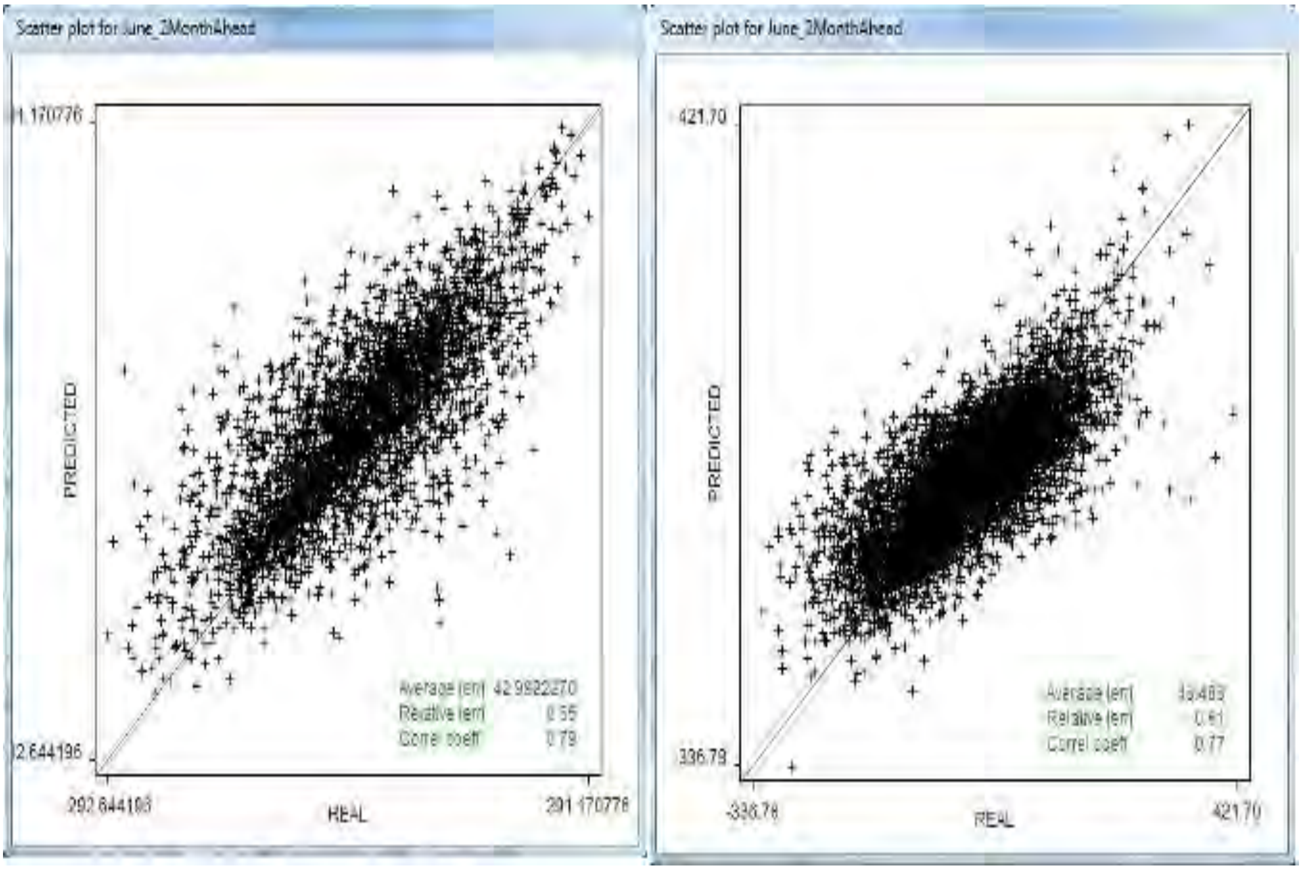



Figure 4.9: June two months ahead prediction

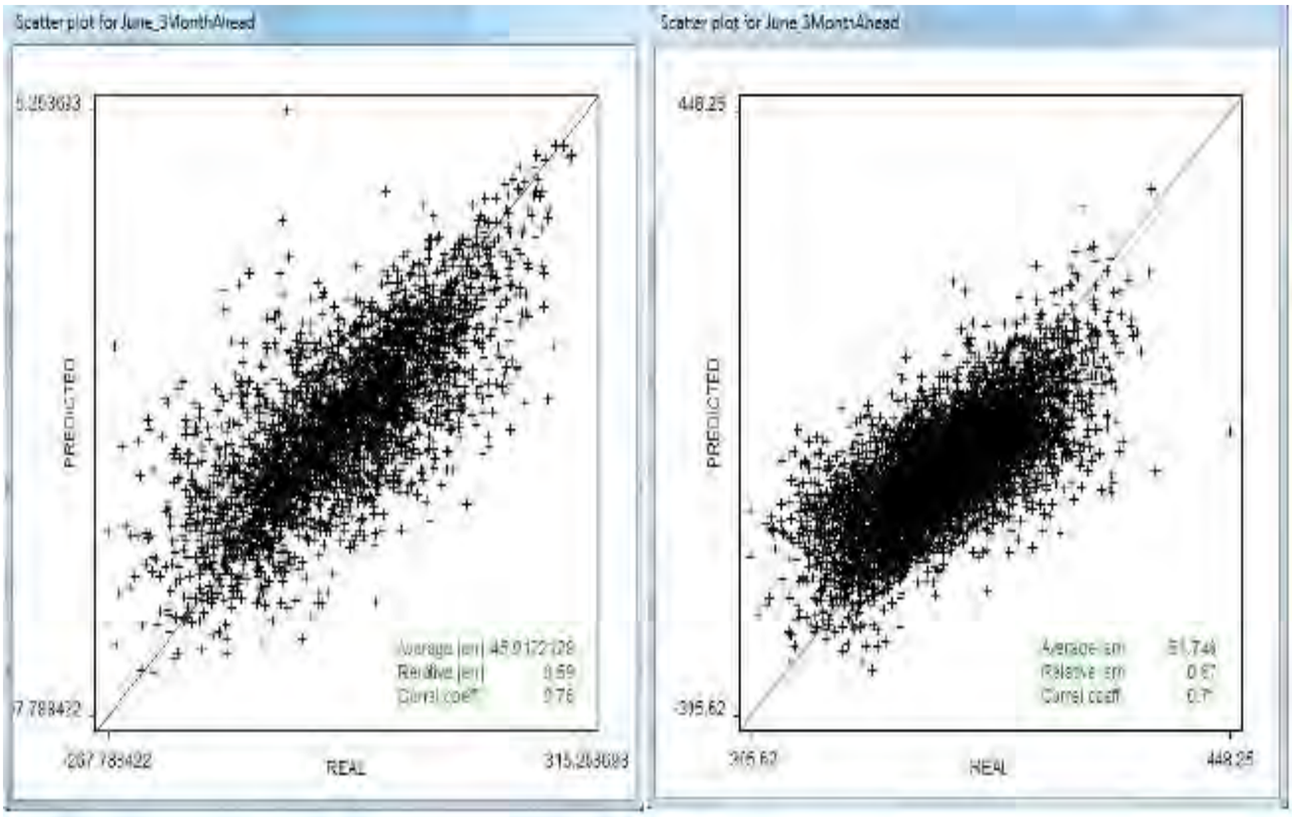



Figure 4.10: June three months ahead prediction

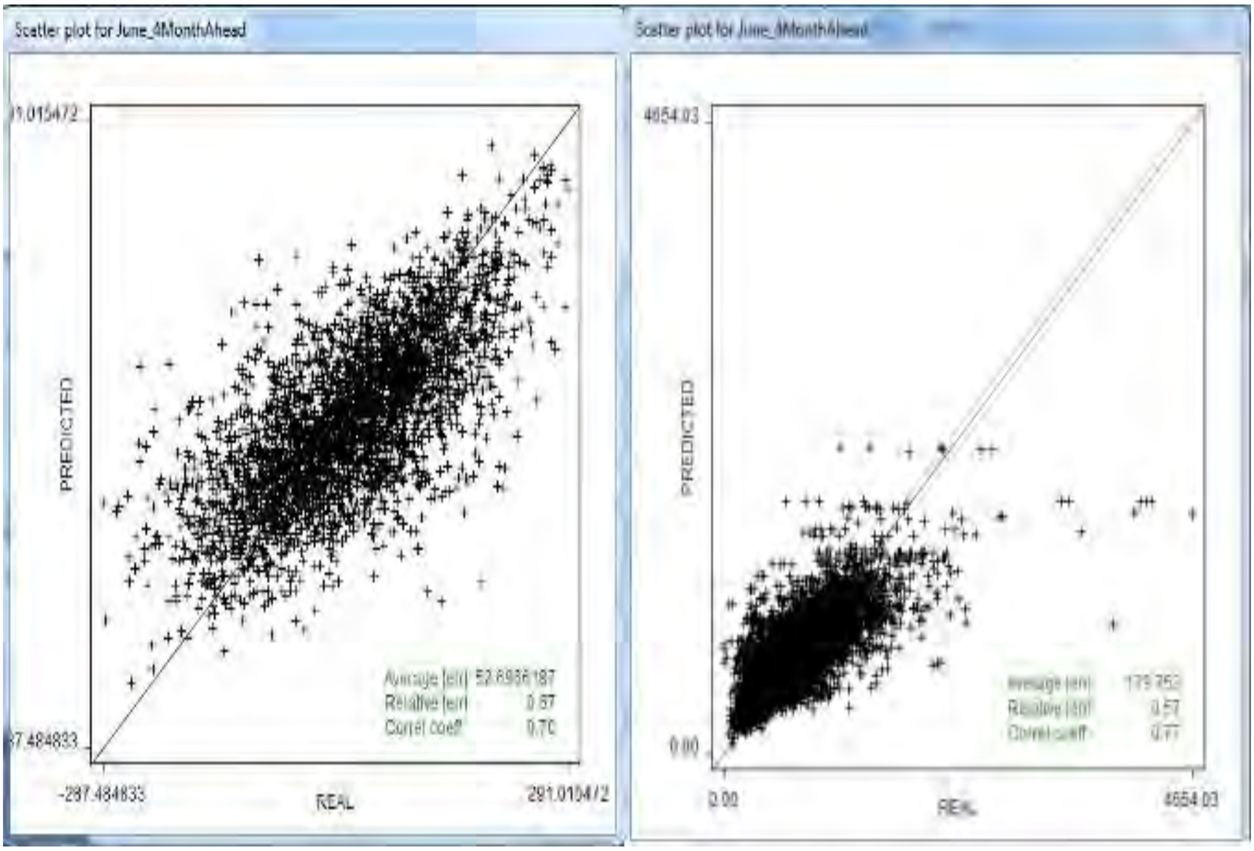



Figure 4.11: June four months ahead prediction

For the sake of simplicity we have only included the June month prediction and left out other growing months from Figure 4.8 - 4.11

There are two things to be noted before analyzing the models. First, this work have used a high resolution Images (specifically 250 Metter resolution Images) compared to the work in [2], which used a relatively lower resolution (8km resolution) satellite images for the prediction. Second, this research could only acquire six years of historical data in total, compared to the twenty five years historical data in [2], Our model could have been improved much if a longer period of historical data could be available to use as in [2].However, unavailability of historical data for our case has been a limitation.

The correlation coefficients are tabulated in Table 4.2. We can also note that the correlations coefficients drop highly when we try to predict 3 – 4 months ahead of time.

Table 4.2: Table relating correlation coefficients between the two researches

| Months | June | | July | | August | | September | |
|---|---|---|---|---|---|---|---|---|
| **Study** | **A** | **B** | **A** | **B** | **A** | **B** | **A** | **B** |
| One Month Ahead | 88 | 81 | 95 | 92 | 97 | 95 | 97 | 65 |
| Two Month Ahead | 79 | 77 | 89 | 84 | 91 | 75 | - | - |
| Three Month Ahead | 76 | 71 | 83 | 75 | - | - | - | - |
| Four Month Ahead | 70 | 77 | - | - | - | - | - | - |

**Study A:** Result of using the high resolution satellite Images (250m resolution) in the prediction of drought.

**Study B:** Based on the result of the model developed in [2] that used low resolution satellite images(8Km resolution)

# Chapter Five: The Drought Prediction System

In this research the essence of building a model is for predicting drought with a higher accuracy. But since the end users of this research are decision makers and climatologists, the output needs to be presented in an informative way that shows the actual event predicted on the ground. To do so we have produced a map showing the drought affected areas of the region.

For the purpose of predicting the seasonal greenness two years have been chosen (i.e. the years 2002 and 2006). The first reason for choosing these years was availability of the SSG with the same resolution; this was essential for cross checking the accuracy of the results. The second reason being the fact that, the years 2002 and 2006 have exhibited relatively severe droughts in the study area. Finally, since the work in [2] has done the prediction for these two time periods, will be convenient for us to make a proper comparison with [2] and judge which is more accurate.

## 5.1 Data Processing

### Static Data Preparation

The static attributes (WHC, VEG_ETHIOPIA, DEM, LandCover) were given a constnt value since it assumed to have constant values throuout the six years of duration. Thus, this images were created only once and repitidely used in every prediction. Althought the static data values assumed the same values throughout the cource of the six year period, their values at each of the choosen 2812 points were distinct. The steps taken to convert this values to Image format (i.e. **`xxx.img`** file) will be discussed in the following paragraphs.

The first step was to extract the point data of each sample points in that specific year and month.This values were stored in a single MS Excel file along with the coordnates of each sample point. The .xls (Excel file) was then imported to MSAccess and exported to a dbase (.dbf) format. The reason to convert the data to dbf format was simply for the arc map software (that was later used to create the Image file) to recognize the data .

The second step was to import the point data to arc map and create a raster image. This step mainly involves interpolating values between points so that a continious raster image will be produced. The conversion of the point values to a map was done using the inverse distance weighting algorithm (IDW). This algortinm was choosen as it is a quick and deterministic interpolator which is exact and will produce a more precise estimation around the data locations [41]. Following the creation of the raster image the third step was to export the raster image to an imagine-image format(**`xxx.img`** file ), here the resolution and the output coordinates were specified, This step was done using the Arc MAP export data tool.

Fillally, the extent of the data was trimmed to remove the fragnet of the data that extended out of the standard choosen dataframe. As these static images contained less detailed information (compared to the dynamic data) , this images were encoded using 16 bit integer.

### Oceanic Index Preparation

The first step in producing the image file is to multiply and roundoff the values encountered at each month by 100. This was done to properly understand the significance of these values since this values are very small. Then, this value is rounded to an integer value since the software used to produce the map of the predicted monthes (Mapcubist software) only accepts images with values at each pixel encoded by an integer value. This step was done using MS Excel.

The second step before the data exporting to an imagine image is correction of the resolution.This stage ensures that all the image values point to the same location in the study site. Please note that this is not cutting the image to the Ethiopian extent, since the data fromat created to the raster image's resolution was alerady for the Ethiopian extent. The third step was to create a constant raster out of the integer value generated by the first step. Creating a raster image was necessary since it serves as an intermidery to create the Imagine Image (**`xxx.img`**).

The fourth step was specifying the coordinate system of the raster image to a standard data frame that the rest of the attributes follow. The coordinate system is : Clarke_1866_Albers using the projection attribute , false easting: 0.00, false northing: 0.00, central meridian: 20, standard parallel_1:-19, standard parallel 2: +21, latitude of origin: 1 and linear unit: meter. This step was done by the ARC GIS software using the create raster tool box, and applying the constant raster. Please note that the resolution of the Image is corrected to (250 meters )at this stage.Then exporting the raster image to an imagine image format follows.

Finally, to coincide all the ocenic index images with the other six attributes the ocenic index images were trimmed to Ethiopia's extent. This was done using the Erdas 8.5 tool. In doing so all the ocenic index files were encoded using 16-bit signed integer value, since 16-bit is sufficient for the detail this images posses.

### SSG and SPI Preparation

The SPI and SSG values at each of the choosen 2812 points were distinct. The steps taken to convert this values to image format will be detailed in the following paragraphs.The first step was to extracct the point data of each sample points in that specific year and month. This values were stored in a single MS Excel file along with the coordnates of each stations. The .xls(Excel file) was then imported to MS.Access and exported to a dbase(.dbf) format.The reasion to convert the data to dbf format was simply for the arc map software (later used to create the Image file) to recognize the data .

The second step was to import the point data (SPI or SSG value) to arc map and create a raster image. This step mainly involves interpolating values between points so that a continious raster image will be produced. The conversion of the point values to a map was done using the inverse distance weighting algorithm, this algortinm (IDW) was choosen as it is a quick and deterministic interpolator which is exact and can produce accurate predictions around data locations.

Following the creation of the raster image the data was exported to an imagine image , here the resolution and the output coornates were specified, this step was done using the Arc MAP export data tool.Finally, the extent of the data was trimmed to remove the segment of the data that extended out of the standard choosen dataframe. As the SSG and SPI images contained more detailed information the image was encoded using 32 bit integer.

## 5.2. Map Generation

To compare two works which predicts drought is to produce a map predicting the drought extent. Then after having the actual values exhibited in that area accuracy of the model could be overlooked. In order to produce a map showing the extent of drought in the region all the contributing attributes need to be changed to an image format. Then, all the eleven images needed to coincide pixel to pixel (i.e. they needed to have the same spatial resolution). This step was required due to the diverse spatial resolutions of the eleven attributes. Although it wouldn't add any additional value to the original data, for the purpose of overlapping images as desired, we have increased the spatial resolution of all the other attributes to a standard 250m resolution. Finally a standard frame was created to place one image on top of another and the extent of the images was trimmed. The values to be predicted are tabulated in table 5.1.

Table 5.1: The values to be predicted in this research

| Month | Model | Predicted value |
|---|---|---|
| June | June One Month Ahead | July |
| | June Two Month Ahead | August |
| | June Three Months Ahead | September |
| | June Four Months Ahead | October |
| July | July One Month Ahead | August |
| | July Two Month Ahead | September |
| | July Three Month Ahead | October |
| August | August One Month ahead | September |
| | August Two Month Ahead | October |
| September | September One Month Ahead | October |

**Prototype for Prediction**

Drought prediction, starting from the extraction of the NDVI values to the SSG generation and then to the production of the predicted image requires the extensive use of different tools. In this work we have sited this problem and tried to create a bridge for the users to predict drought with a limited resources.

One of the major tools that were used in the data processing stage is ARC MAP. But, this does not mean that any individual will be able to predict drought using this tool. Once the data is processed and the model is produced using the cubist software, the rest is prediction. For prediction, we have used the Map-Cubist software. To avoid the need to go back and forth between software to predict drought we have developed an extension on the ARC GIS Explorer.

This tool first accepts the eleven attributes contributing to the drought event. Then the user locates the model that consists a set of rules which will be used for drought prediction. Finally, the system produces a prediction map. The resulting map will then be displayed on the ARC GIS Explorer software. The system administration then uploads the produced map on a server from which it will be accessible for end users (mostly decision makers). The steps in predicting drought using the prototype are depicted in the Figure 5.1 using a use case diagram.

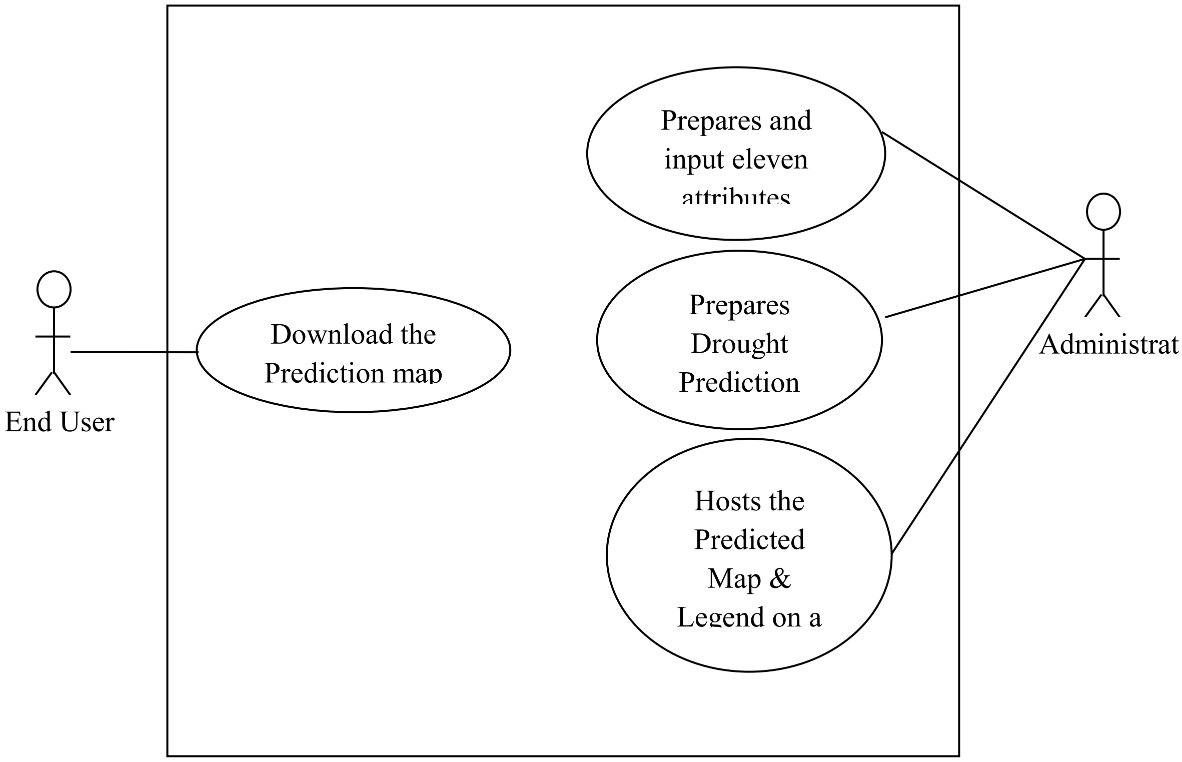



Figure 5.1: Use Case Diagram for the Drought Prediction System

In Figure 5.2 we have extended our use-case diagram to demonstrate the flows of the steps used by the prototype used in this work. Thus, we will present the activity diagram for the drought prediction system.

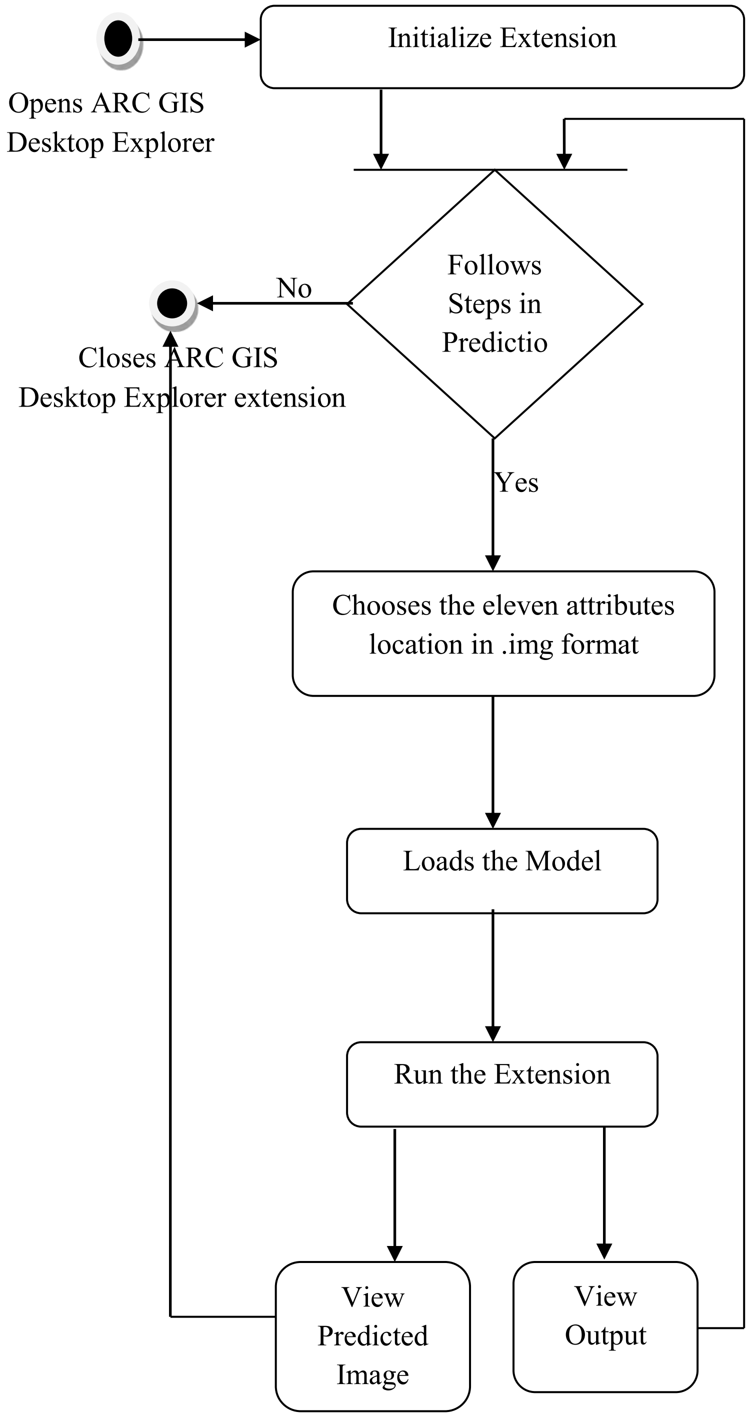



Figure 5.2: Activity Diagram of the Drought Prediction system prototype

To understand the process of drought prediction using our system, we presented some screen shots below.

**Step 1:** The administrator will go to the Add-Ins tab, in ARC GIS Explorer, and choose the prediction button as shown in Figure 5.3.

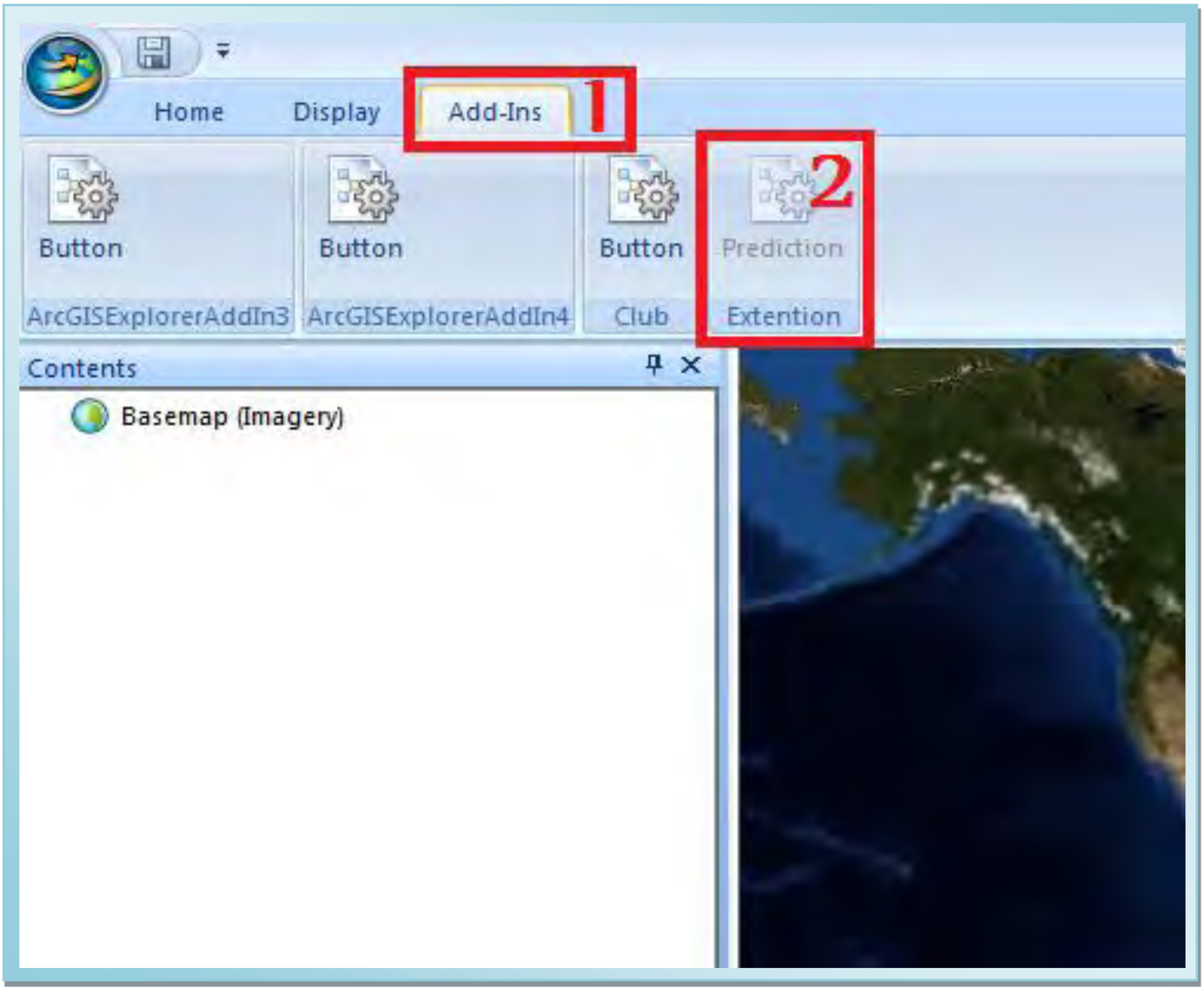



Figure 5.3: Screen shot of the ARC-GIS Explorer with our Extension

**Step 2**: The administrator needs to locate all the eleven attributes using the user interface provided as shown in Figure 5.4.

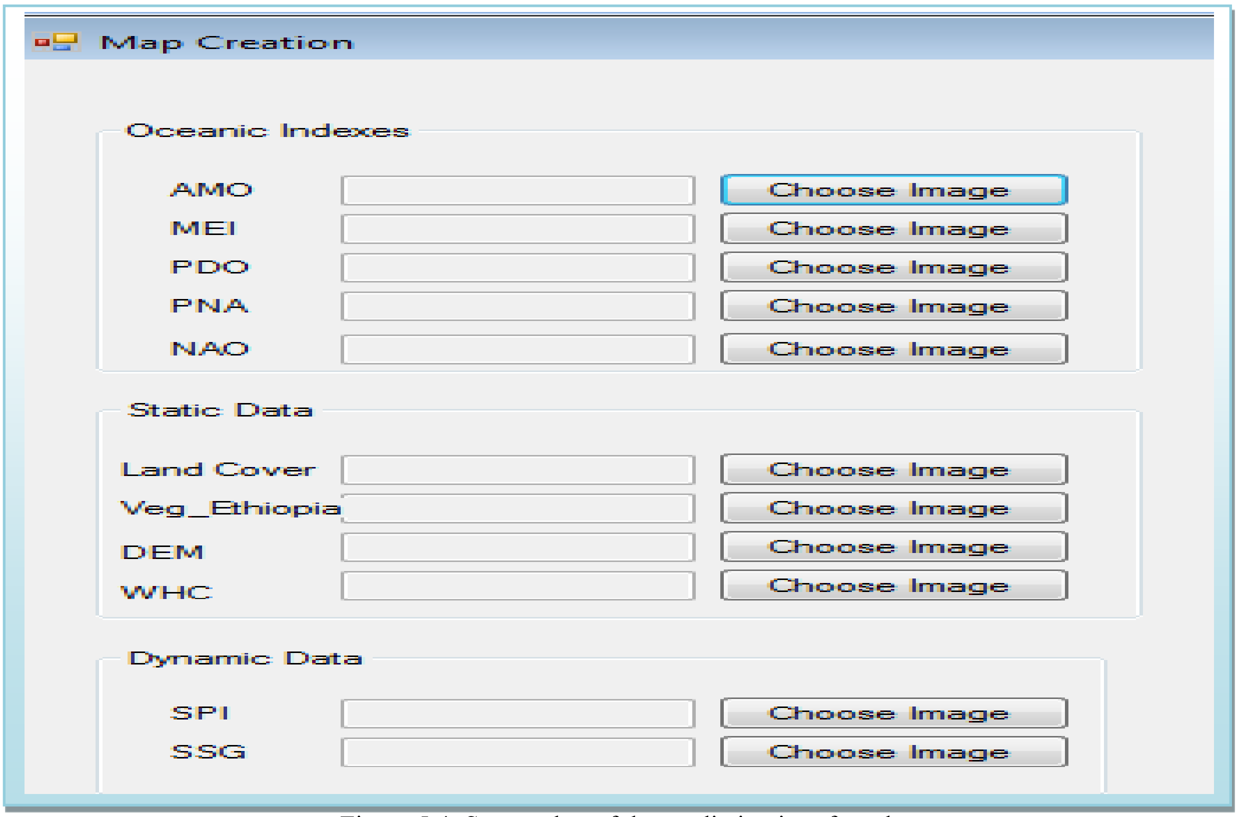



Figure 5.4: Screen shot of the prediction interface the system

**Step 3**: The model for that specific month is located using the locate model path button as shown in Figure 5.5.

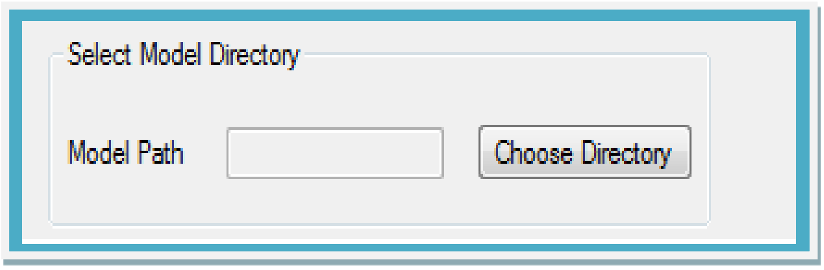



Figure 5.5: Screen shot of the system to select model

**Step 4:**

After selecting the model and having all the .img files as an input. The tool will hand over the task of prediction to the cubist software, where the predicted map will be produced. The image produced following the predicted values is shown in Figure 5.6

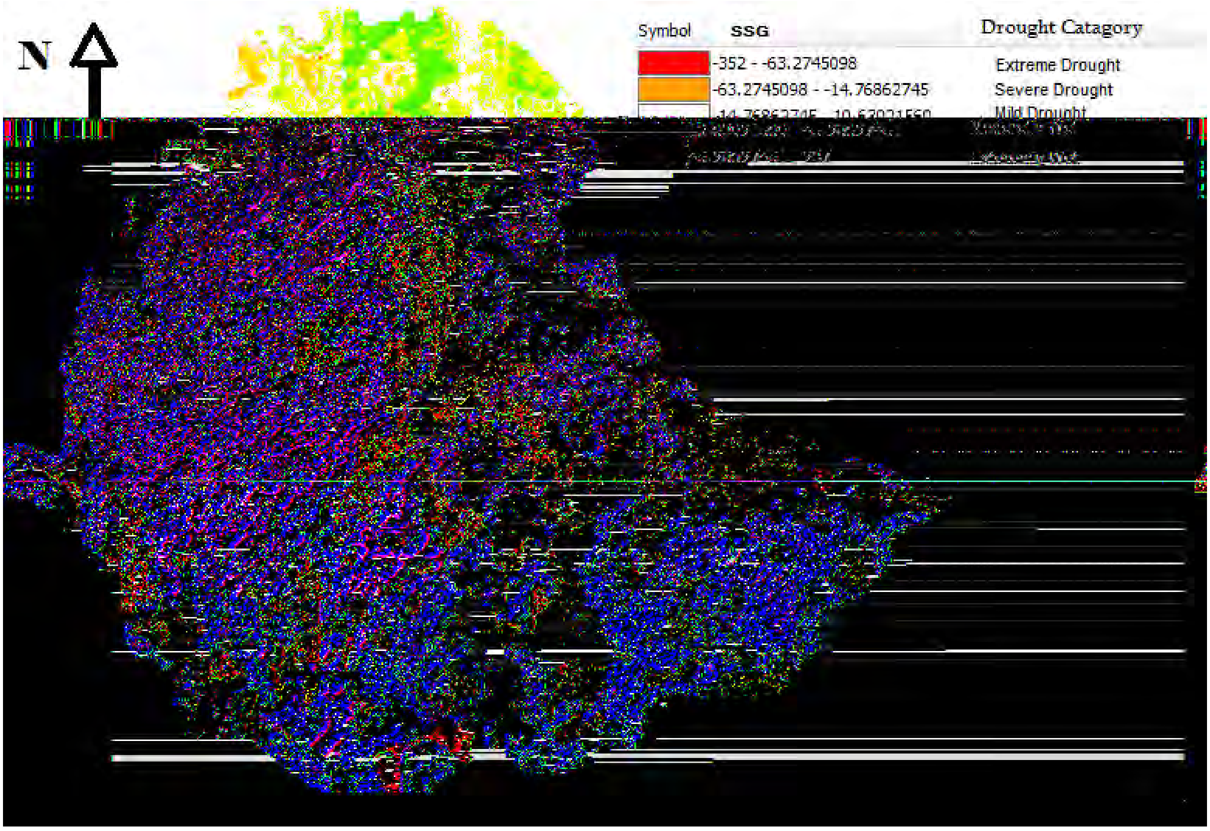



Figure 5.6: Prediction of the July 2002 Standardized Seasonal Greenness with the June 2002 data

**Summary**

Lower resolution satellite images are subsets of higher resolution satellite images. The significance of incorporating higher resolution satellite images in drought prediction is analyzed in this research. The accuracy of drought prediction system and the more detailed information it reveals, plays a significant role in getting more reliable information to decision makers. Hence, we can conclude that using higher resolution satellite Images, accuracy of drought prediction systems have been greatly improved. In addition a detailed prediction capability whose precision level is known is realized. The design of this system also considered the end-users, who may have minimal computer skills. Hence, the tool developed for predicting a drought is easy to use and is user friendly. Understanding the output of the prediction system is straightforward since it is colour coded map. Furthermore, the prediction of drought could be easily downloaded for quick reference from the host servers. This not only saves lots of resources but also helps decision makers acquire the desired prediction in a timely fashion.

# Chapter Six: Evaluation of the Prediction

The ultimate goal of this research is to enable end users with minimal knowledge and resource to predict drought prior to its occurrence with a better accuracy. Thus the model prepared needs to be tested. This document is prepared to give a high-level overview of our testing strategy for the drought prediction system prepared in experimentation phase. Its objective is to quantify the level of accuracy exhibited and compare the prediction capability of the high resolution image based prediction to that of the system with a lower resolution image.

Next, we will include the test plan, to highlight which section of the research will be subjected to testing and how the testing will be done and most importantly to signify what the test coverage will be. Here we will be detailing the test conditions and the expected results as well as the test pass criteria. Following this, we will incorporate the test case specification, where we will specify the test data used in running the test conditions identified in the test design specification .Finally will conclude by reporting the result of the testing procedure.

When developing this prototype we have constrained from using a different algorithm possibly a better one from the one used in [2]. In doing so we have avoided accuracy and efficiency issues along with other unforeseen software driven effects that could positively or negatively affect the outcome. Thus comparison of this work with other lower satellite resolution image works will not be affected due to algorithm efficiency. The main intent of this document is to test the quality of the predicted image relative to other prediction schemes. Though, quality could be measured in different metrics, speed accuracy, efficiency, memory, scalability, etc; we have considered only the accuracy test.

The term accuracy, this term is used to describe systems and methods that measure, estimate, or predict. In all these cases, there is some parameter we wish to know the value of. This is called the true value, or simply, truth [36]. In our case the truth is the seasonal greenness. The image produced in predicting the drought event provides a measure of the estimated seasonal greenness, that we want to be as close to the true value as possible. Accuracy is one way to describe the error that can exist between these two values.

Accuracy is sometimes confused with precession. Although, these terms could be used interchangeably in non-technical settings, we have specific definitions for each. Accuracy of a measurement system is the degree of closeness of measurements of a quantity to the true value, where as the precision of a measurement system, also called reproducibility, is the degree to which repeated measurements under unchanged conditions show the same / exact results [37].

In our case, since predictions are being made the degree of precision could be very poor, but the accuracy could be a good measure. Considering this we have included the accuracy as one major yardstick for measuring the quality of the model produced. In the next paragraphs we will briefly detail how we have measured these parameters and will report the result of each measurement.

Higher value should be given for the accuracy to compare two drought prediction systems, as it reflects the validity of these works. To measure the accuracy of these models we have used the root mean square error estimating index. Before discussing the proceedings of this index it would appropriate to give a proper introduction of the root mean square error estimating index.

The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed [38]. RMSD is a good measure of accuracy, but only to compare different forecasting errors within a dataset and not between different ones, as it is scale-dependent. These individual differences are also called residuals, and the RMSD serves to aggregate them into a single measure of predictive power.

Assuming that we have different data points the RMSD works as follows: For each data point, RMSD subtracts observed or true value ($X_{obs}$) from the predicted value ($X_{pred}$) at that specific data point and measures the distance between these values. Since positive and negative distances should complement each other the differences are squared and summed. Then these values are averaged. Finally, the square root of the average value is calculated and is called the RMSD value.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} (X_{obs,i} - X_{pred,i})^2}{n}}$$

*Equation 6.1*

Following the suggestion in that defined Equation 6.1, [42] we have calculated the RMSD for 2812 data points at the study site. For this purpose, we have developed software that calculates the accuracy difference between two researches using the above formulae.

In Figure 6.1, the screen mock-up of the testing software is shown. Using this software we can compare two prediction mechanisms against their true values. The output of the software has RMSD value and error percentage value. In this manner a lower RMSD indicates a much closer (Accurate) value as compared to the higher value, and a lower error percentage shows a precise and accurate prediction.

Using the developed software we have experimented on the 2002 data and have come up with a very promising result. The screen mock-up of the software is indicated in Figure 6.1.

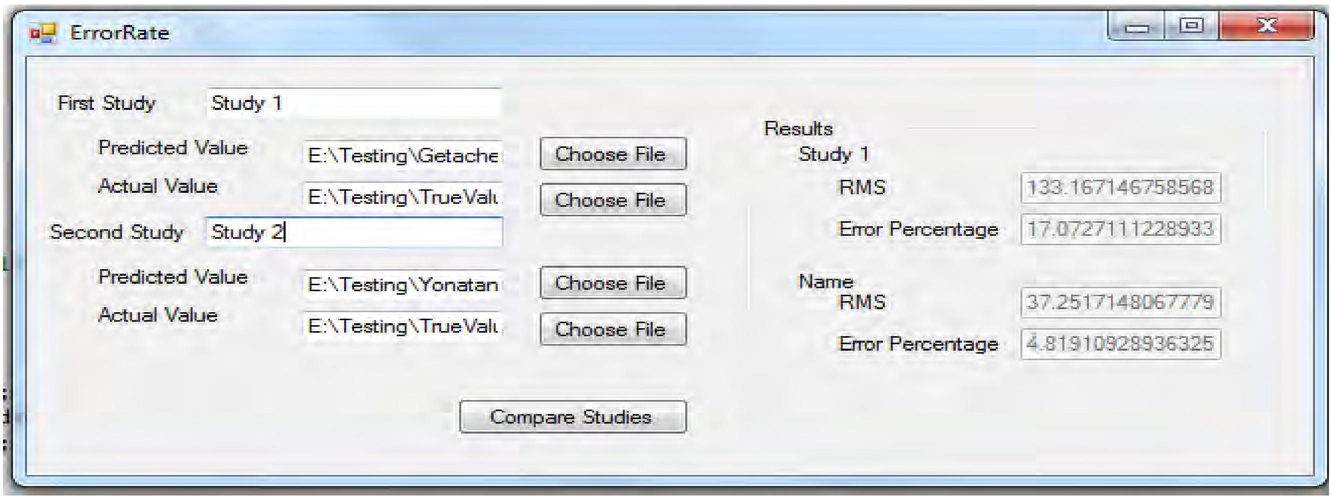



Figure 6.1: Error rate RMSE calculator

**Where**

Study 1: is drought prediction using low resolution satellite Image

Study 2: is drought prediction using high resolution satellite Image

As can be shown on Figure 6.1, based on the result of the RMSE, the previous work has shown an error percentage of 17.07, where as our work has an error percentage of 4.82. We have come down to compare the results of the prediction mechanism, in which we can clearly see the level of accuracy presented with this research. The accuracy mainly is accounted for the higher resolution image incorporated in this work. Thus it can be concluded that the effect of higher resolution satellite image increases the accuracy of drought prediction systems.

# Chapter Seven: Conclusion

Drought is one of the devastating natural events on earth. Although mitigating its occurrence should be a long time plan, drought prediction can play a role in dealing with drought for a short term basis. To predict drought analyzing what is observed on the ground plays a significant role. As it is the actual fact on the vicinity. In addition other attributes helps determine the possible occurrence of drought before a visible effect on the ground is exhibited. Combining these facts prediction could be eased.

Especially in developing countries where information dissemination on the ground site is a challenging task, knowing the occurrence of drought event beforehand by using an automated tool will play a significant part in minimizing its devastating effects.

Though there had been several efforts to predict drought, the lack of high resolution satellite images had been a challenge to design a system with high accuracy rate of drought prediction. We are now at a point where we start to question the accuracy of drought prediction attempts, since a lot of resources will be allocated for its mitigation effort. This work is an ideal attempt that has already proven that drought prediction could be more accurate by implementing a higher resolution satellite images.

Our work has gone through many stages towards the prediction of drought with higher accuracy. First, we have acquired a 250m resolution satellite image data. Then the ten other attributes that were identified for drought prediction were used. Second, we have pre-processed the high resolution satellite image and transformed it to generate the NDVI values. Using the NDVI values we have computed the Seasonal Greenness (SG) and Standardized Seasonal Greenness (SSG) values for the study area, Ethiopia. The SSG is then converted to a point data values for the selected 2812 points. In the Third stage, the model is generated and the current eleven attributes are integrated in a drought prediction system. For the purpose of easier understanding the colour coded map was generated that depicts the drought prediction status from one to four months. Finally, the accuracy of the drought prediction system is tested for correctness.

Furthermore, the result of prediction using high resolution satellite images is compared with a previous work that used low-resolution satellite images and the result shows that our work has improved the prediction accuracy.

The First step towards developing a more precise drought prediction engine is using a longer period of historical data. With this we train the engine so that future events could be easily predicted more accurately. This could be used to generate better models. Thus, the longer the accumulated historical data the higher the accuracy of the prediction system. In an attempt to tune up this research to a higher precession rate it is our belief that the longer historical data would play a better role as future work. In addition predicting unforeseen future drought events can be considered as a future work.

Finally, for the purpose of speed, we could implement parallel processing in the prediction system can enhance the speed of the prediction process.

# References

1. Brian D. Wardlow, Tsegaye Tadesse, and Michael J. Hayes. The Vegetation Drought Response Index (VegDRI):A New Integrated Approach for Monitoring Drought Stress in Vegetation ,Bellwether Publishing .45(1) :16–46, 2008.
2. Getachew Berhan , Shawndra Hill , Tsegaye Tadesse , Solomon Atnafu , Using Satellite Images for Drought Monitoring:A Knowledge Discovery Approach , Journal of Strategic Innovation and Sustainability vol. 7(1) 2011 , 135-153,2011.
3. Encyclopedia Britannica Online, s. v. "Ethiopia," Avaliable on: http://www.britannica.com/EBchecked/topic/194084/Ethiopia, Last accessed: February 2, 2012.
4. USGS earlywarning , eMODIS NDVI and Anomalies 250m Avaliable on : http://earlywarning.usgs.gov/fews/africa/web/imgbrowsc2.php?extent=nazd, Last accessed: January 2,2012
5. Wikimedia Foundation, Inc Breeding for drought stress tolerance, Avaliable on : http://en.wikipedia.org/wiki/Breeding_for_drought_stress_tolerance, Last accessed: December 20, 2011
6. Lotze HK, Lenihan HS, Bourque BJ, Bradbury RH, Cooke RG, Depletion, degradation, and recovery potential of estuaries and coastal seas. 1806–1809, 2006.
7. McKee, T.B., N.J. Doesken, and J. Kleist. Drought Monitoring with Multiple Time Scales, Preprints, 9th Conference on Applied Climatology, Dallas, Texas, Pages :233–236, 1995.
8. Wikimedia Foundation , Normalized Difference Vegetation Index Wikipedia, Avaliable on : http://en.wikipedia.org/wiki/NDVI, Last accessed January 1, 2012
9. Palmer, W. C., 1965. Meteorological Drought. Res. Paper No. 45, Weather Bureau, Washington, D.C., 58 pp.
10. Ravi Prakash Gutpa , Remote Sensing Geology Published by: Springer Verlag, Berlin-Heidelberg-New York-Tokyo 365 (Vol -) 1991 Page 572, 1991
11. Tadesse, T., B.D. Wardlow, M.J. Hayes, and M.D. Svoboda. The Vegetation Outlook (VegOut): A New Method for Predicting Vegetation Seasonal Greeness. GIScience and Remote Sensing 47:In Press, 2010
12. NOAA. National Oceanic and Atmospheric 2011.Administration, Climate Indices: Monthly Atmospheric and Ocean Time Series [Online]. Available by Earth System Research Laboratory, Physical Science Division Avaliable : http://www.esrl.noaa.gov/psd/data/climateindices/list/ Accessed : January 6, 2012
13. Rulequest Research , An Overview of Cubist, Avaliable : http://www.rulequest.com/cubist-win.html, Last Accessed January 12, 2012

14. Marta Moneo and Ana Iglesias(PhD) - Universidad Politécnica de Madrid, España last published: 2004-05-12 , Types of drought Avaliable : http://www.atmosphere.mpg.de/enid/26b.html, Last Accessed Feburarty 21, 2012.
15. Nathalie Pettorelli , Jon Olav Vik, Atle Mysterud Jean-Michel Gaillard, Compton J. Tucker, Nils Chr. Stenseth Using the satellite-derived NDVI to assess ecological responses to environmental change .Volume 20, Issue 9, Pages 503–510 ,September 2005
16. Dbxsoul , What is water holding capacity of soil? : Available on : http://wiki.answers.com/Q/What_is_water_holding_capacity_of_soil , Last Accessed : Feburary 2 , 2012.
17. Wikimedia Foundation , Drought ,Avaliable : http:// en.wikipedia.org/wiki/Drought , Last Accessed December 20, 2011
18. Michael J. Hayes (PhD), Drought Indices ,Avaliable : http://www.civil.utah.edu/~cv5450/swsi/indices.htm, Last Accessed December 12, 2011
19. Andrej Ceglar , Drought indices , Biotechnical faculty, University of Ljubljana January 12 , 2012.
20. Ntale, H. K., and Gan, T.Y., 2003, Drought indices and their application to East Africa, Int J. of Climatology, Royal Meteorological Society, 23, p. 1335-1357.
21. M. Abramowitz and I. A. Stegun , June 1965, Handbook of mathematical functions, http://people.math.sfu.ca/~cbm/aands/frameindex.htm, January 19, 2012.
22. Ji and Peters, Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices,2003, Remote Sensing of Environment , vol. 87, no. 1, pp. 85-98, 2003
23. Cancelliere, A., Di Mauro, G., Bonaccorso, B. and Rossi, G Investigating the potential of NAO index to forecast drought in Sicily, 2007
24. Nalbantis and Tsakiris Assessment of Hydrological Drought Revisited Springer Science and Buisness Media B.V, 2008.
25. Shukla and Wood, use of standardized runoff index for characterizing hydrologic drought , Geophysical Research letters, 35(2) , 2008
26. Drought Analysis under Climate Change by Application of Drought Indices and Copulas Wen Yang, 2010
27. Hayes, Drought Monitoring using Standardized Precipitation Index: A case study for the state of Karnataka, India, 2003
28. Keyantash, John, John A. Dracup, The Quantification of Drought: An Evaluation of Drought Indices. 2002
29. Thiruvengadachari& Gopalkrishana, 1993, An integrated PC environment for assessment of drought Inter. J. Rem. Sen., 14 (17), 3201-3208.

30. http://www.guardian.co.uk/global-development/interactive/2011/jul/04/somalia-hornofafrica-drought-map-interactive,  Last Accesed Date : Nov :09/11/2011
31. Ahmad Fatehi Marja & Allard M. J. Meijerink,  Agricultural drought forecasting using satellite images, climate indices and artificial neural network
32. Alireza Shahabfar & Azadeh Ramesh ,Construction of Drought Early Detection System,2010.
33. http://earthobsarvatory.nasa.gov/Features/MeasuringVegetation, Last Accessed Date :Oct 01,2011
34. Tucker, C. J,  Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ, 1979.
35. Kanellou, C. Domenikiotis, E. Tsiros &  N.R. Dalezios , Satellite-based Drought Estemation in Thessaley:vol 23 No, 24, 2008.
36. Steven W.Smith PhD, A scienat and engineer guide to digital signal processing, http://www.dspguide.com/ch2/7.htm, Last Accessed Date: sept,01,2012
37. "The international vocabulary of metrology- Basic and  general concept & associated terms"(VIM), 3rd Edition, JCGM 200:2008
38. wekipedia root mean square deviation, http://en.wikipedia.org/wiki/Root-mean-square_deviation, Last Accessed Date: Jan, 2013
39. Swets, D.L., Reed, B.C., Rowland, J.D. & Marko, S.E. (1999). A weighted least-squares approach to temporal NDVI smoothing
40. Rouse, J. W., R. H. Haas, J. A. Schell, and D. W. Deering (1973). Monitoring vegetation systems in the Great Plains with ERTS, Third ERTS Symposium, NASA SP-351 I, 309-317.
41. Shapard Donald, A tow-Dimensional interpolation function for irruglarly-spaced data" Proceedings of the 1968 ACM National Conference pp 517-524.
42. J.Scott Armstrong and Fred Collopy  Error Measures for Generalizing About Forecasting Methods : Empirical Comparisons. International Journal of forcasting8: pp 69-80. 1992

Declaration

I, Undersigned, declare that this thesis is my original work and has not been presented for degree in any other university, and that all sources of material used for the thesis have been acknowledged.

Declared by:

Name: **Yonatan Getachew**

Signature: ________________________

Date ________________________

Confirmed by advisor:

Name: **Solomon Atnafu (PhD)**

Signature: ___________________________

Date ___________________________

Place and date of submission: Addis Ababa University, April 16, 2013.