ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF COMPUTER SCIENCE

Drought Monitoring and Prediction with higher Temporal Resolution Satellite Images

By: Samuel Sisay Tadesse

Advisor: Solomon Atnafu (PhD)

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF THE ADDIS ABABA UNIVERSITY IN PARTIAL FULFILLMENT FOR THE DEGREE OF MASTERS OF SCIENCE IN COMPUTER SCIENCE

March 2013

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF COMPUTER SCIENCE

Drought Monitoring and Prediction with higher Temporal Resolution Satellite Images

By: Samuel Sisay Tadesse

ADVISOR: Solomon Atnafu (PhD)

APPROVED BY

EXAMINING BOARD:

1. Dr. Solomon Atnafu, Advisor _____
2. _____ _____
3. _____ _____

Acknowledgement

Above all, I would like to thank God who always helps me in every aspect of my life.

I am heartily thankful to my Co-advisor, Getachew Birhan, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. This thesis would not have been possible without his help. I acquired the data used in the study through him and I am grateful for that too.

I owe my deepest gratitude to Dr. Tsegaye Tadesse, University of Nebraska-Lincoln, National Drought Mitigation Center, U.S.A, for his help in building the models I used.

It is a great honor to work with my advisor Dr. Solomon Atnafu, who is like a father figure and helps people with patience. He has helped me in understanding the subject, not missing the objective and most importantly connected me to people who are experts in the area.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the Thesis.

Abstract

Attributed to climatic change and uncertainty of weather conditions, drought has become a recurrent phenomenon in many parts of the world and particularly in eastern African region as most of its agriculture is dependent up on rainfall. In most areas of the region drought monitoring and prediction is done based on meteorological data which is very tedious and time consuming to gather and even the stations from where the data are collected are not evenly distributed and it is impossible to get data representing all the areas. And this issue brought us the need for better tools for monitoring and predicting drought in the area. Our study focuses on monitoring and prediction of agricultural drought. Studies has indicated integration of remotely sensed data such as NDVI images with some other relevant environmental attributes that influence vegetation condition increases the capability of predicting future vegetation conditions. The purpose of our study is constructing models, using higher temporal resolution data (dekadal data), by integrating Satellite vegetation, Climate, Oceanic and Biophysical data, for case of Ethiopia. The models are then used in monitoring and prediction of drought with higher accuracy. A 24 year historical data from the years 1983 to 2006, with spatial resolution of 8 km, for the selected attributes were used in constructing the model. The models were validated on test data and we got up to 99% prediction accuracy but when the predicted period is getting far the prediction accuracy decreased. The models make dekadal predictions and requires only for 10-days data to make the predictions this gives us the advantage of getting early update on drought situation. The models were implemented for the years 1984 and 2002 where severe drought occurred in Ethiopia. And we got similar results as at the time. Our work was compared with a recent study by Getachew et al. [19] where they built models, for Ethiopia case, using the same type of data as we used but with lesser temporal resolution (monthly average data). The comparison was done to check if using dekadal data brings a significant increase in prediction accuracy of drought over using monthly average data. The comparison confirmed that using data with higher temporal resolution, dekadal data, gives us higher prediction accuracy. So decision makers and other stake holders in the area are encouraged to try this model as it gives us greater prediction accuracy and early update on drought condition.

Key words: Drought monitoring, Drought prediction, Dekadal data, Monthly average data, Model

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AMO:	Atlantic Multi-decadal Oscillation Index
AVHRR:	Advanced Very High Resolution Radiometer
CMI:	Crop Moisture Index
CART:	Classification and Regression Tree
DEM:	Digital Elevation Model
DEV_{NDVI}:	Deviation of NDVI
DRMFSS:	Disaster Risk Management and Food Security Sector
EMA:	Ethiopian Mapping Agency
ENSO:	El Nino–Southern Oscillation
EVI:	Enhanced Vegetation Index
FEWS:	Famine Early Warning Systems
IDW:	Inverse Distance Weighting
MEI:	Multivariate ENSO Index
MERIS:	Medium Resolution Imaging Spectrometer
MODIS:	MODerate resolution Imaging Spectroradiometer
MSG:	Meteosat Second Generation
NAO:	North Atlantic Oscillation
NDMC:	National Drought Mitigation Center
NDVI:	Normalized Difference Vegetation Index
NOAA:	National Oceanic and Atmospheric Administration
OLS:	Ordinary Least Square
PNA:	Pacific North American Index
PDO:	Pacific Decadal Oscillation

PDSI: Palmer Drought Severity Index

RFE: RainFall Estimate

SG: Seasonal Greenness

SPI: Standardized Precipitation Index

SSG: Standardized Seasonal Greenness

SWSI: Surface Water Supply Index

TAMSAT: Tropical Applications of Meteorology using Satellite

USGS-EROS: United States Geological Survey center for Earth Resources Observation and Science

VCI: Vegetation Condition Index

Veg_ethiopia: Vegetation Type of Ethiopia

VI: Vegetation Index

WHC: Water Holding Capacity

1. INTRODUCTION

1.1. Background

Drought is a weather-related natural disaster. It is the world's costliest natural disaster, causing an average $6–$8 billion in global damage annually and collectively affecting more people than any other form of natural disaster [25, 43]. It is characterized by a slow-start and is a normal part of climate for virtually all regions of the world [23].

Drought is difficult concept to define [1], because what may be considered a drought in one region would certainly not be considered drought in another. For example, in Bali six days without rain is considered as a drought but not in Libya where annual rainfall less than 180 mm. In the most general sense, drought originates from a deficiency of precipitation over an extended period of time usually a season or more resulting in a water shortage for some activity, group, or environmental sector.

Drought cannot be viewed solely as a physical phenomenon or natural event since its impact results from the relation between a natural event and demands on the water supply; it is usually defined both conceptually and operationally. Conceptual definitions generally help people understand the concept of drought. One conceptual definition can be: Drought is a prolonged period of deficient precipitation resulting in extensive damage to crops, resulting in loss of yield [6]. These conceptual definitions may also be important in establishing drought policy. For example declarations of exceptional drought are based on science-driven assessments. On the other hand operational definitions help define the start, severity, and end of droughts. They can also be used to analyze drought frequency, severity, and duration for a given historical period. No single operational definition of drought works in all circumstances, and this is a big part of why policy makers, resource planners, and others have more trouble planning for drought than they do for other natural disasters. To determine the beginning of drought, operational definitions specify the degree of departure from the average of precipitation or some other climatic variable over some time period. This is usually done by comparing the current situation to the historical average, often based on a 30-year period of record [65].

There are four major types of drought. These include: meteorological drought, hydrological drought, agricultural drought and socioeconomic drought [24].

Meteorological drought is defined usually on the basis of the degree of dryness (in comparison to some "normal" or average amount) and the duration of the dry period. Definitions of meteorological drought must be considered as region specific since the atmospheric conditions that result in deficiencies of precipitation are highly variable from region to region [6, 1].

Hydrological drought refers to a persistently low discharge and/or volume of water in streams and reservoirs, lasting months or years. Hydrological droughts are usually related to meteorological droughts, and their recurrence interval varies accordingly. Changes in land use and land degradation can affect the magnitude and frequency of hydrological droughts [6, 1].

Agricultural drought is a situation in which the amount of soil moisture no longer meets the plants' needs. It links various characteristics of meteorological (or hydrological) drought to agricultural impacts, focusing on precipitation shortages, differences between actual and potential evapotranspiration, soil water deficits, reduced groundwater or reservoir levels, and so forth [6, 1]. *Plant water demand and thus general health depends on prevailing weather conditions, biological characteristics of the specific plant, its stage of growth, and the physical and biological properties of the soil* [26].

Socioeconomic definitions of drought associate the supply and demand of some economic good with elements of meteorological, hydrological, and agricultural drought. It differs from the aforementioned types of drought because its occurrence depends on the time and space processes of supply and demand to identify or classify droughts. Socioeconomic drought occurs when the demand for an economic good exceeds the supply as a result of a weather-related shortfall in water supply [6, 1]. For example drought may result in significantly reduced hydroelectric power production because power plants were dependent on streamflow rather than storage for power generation. Figure 2.1 shows the sequence and impacts of the drought types.

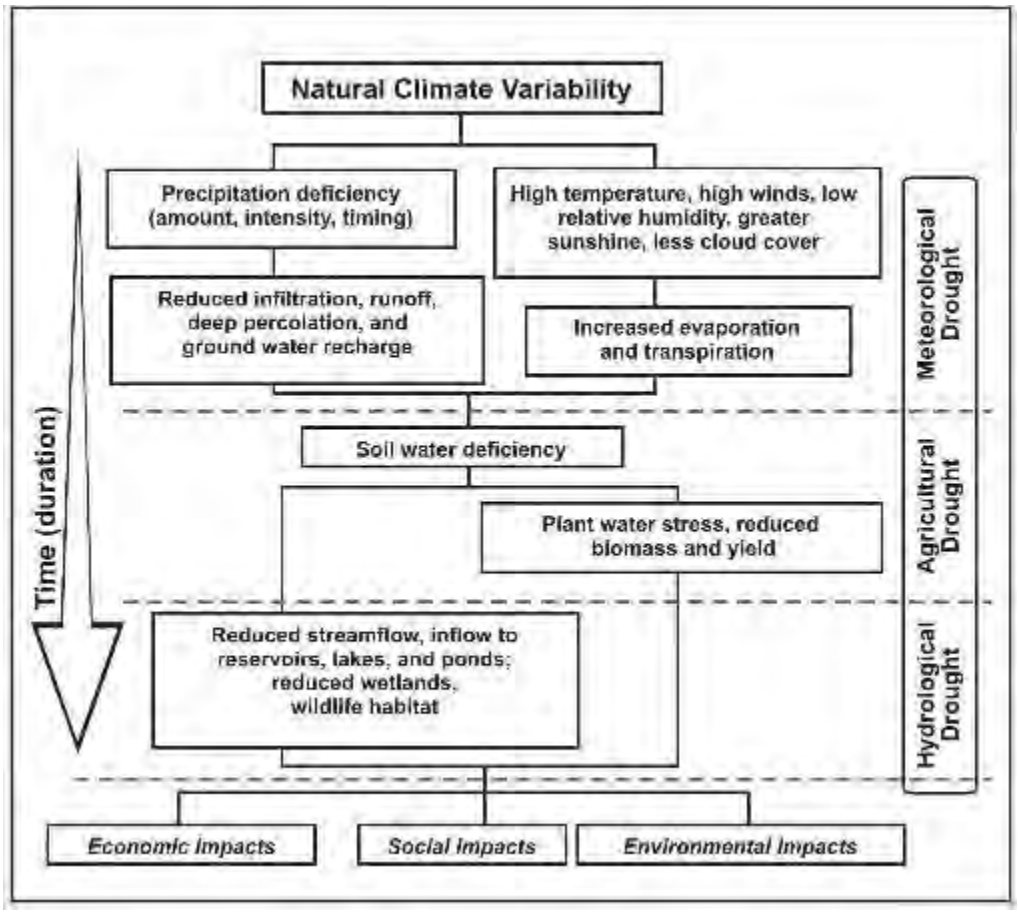


Figure 1.1: Sequence of droughts and their impacts (Source: National Drought Mitigation Center, University of Nebraska-Lincoln, U.S.A.)

Drought has a negative impact on one country's economy. For example in consequences of drought on the economy of Ethiopia from 1973 to 1986 in [36] shows decline in agricultural production which led to increases in food imports, reductions in agriculturally based manufacturing activities and substantial declines in agricultural (and thus in total) exports. Drought events in our country and elsewhere has occurred frequently and caused a variety of costly losses and this has brought the requirement of better tools for monitoring and predicting general vegetation conditions.

In this work we focus on agricultural drought. Among the technologies that are used currently, for monitoring vegetation condition, one is remote sensing. *Remote sensing is the science of acquiring information about the Earth's surface without actually being in contact with it* [4].

This technology can be applied in vegetation health assessment, crop type classification, crop yield estimation, mapping of soil characteristics, and drought monitoring.

Vegetation monitoring using remote sensing has been there for more than 20 years. Monitoring vegetation drought usually requires a large amount of temporal data. These huge data are collected from different satellites which capture the imagery data in different time intervals. The collected data are then preprocessed and analyzed for monitoring and predicting drought.

Vegetative Index (VI), the comparison of reflectance values at different wavelengths, is commonly used to determine vegetation health. The most common vegetation index used is the Normalized Difference Vegetation Index (NDVI).

Integration of remotely sensed vegetation data such as NDVI images with some other relevant environmental variables that influence vegetation holds increases the capability of predicting future vegetation conditions [22, 49]. Using such integration is getting applied in drought monitoring and prediction for different regions of the world. The purpose of our study is constructing a model which can be used in monitoring and prediction of drought with higher accuracy by integrating remotely sensed with some environmental data. Here we take one recently conducted study [19] and try to improve on it. Getachew et al. integrated different remotely sensed, climate, and environmental data for monitoring and predicting drought [19]. The data used in his study had a spatial resolution[1] of 8km and temporal resolution[2] of monthly average. Monthly average data was derived by summing up the three dekadal data (10-day composites of daily data) in a month and taking the average. Results from the study shown higher accuracy (up to 95%) in predicting drought. So in our study we attempt to improve the prediction accuracy even better by using the same type of data as in [19] but using data with higher temporal resolution i.e., dekadal data. Here we also prove whether using data with higher temporal resolution brings us significant increase in accuracy over using monthly average data.

1.2. Statement of the Problem

One of the obstacles of efficient agricultural practices is detecting vegetation health condition at the right time. Assessing vegetation health is critical in ensuring good agricultural productivity. Attributed to climatic change and uncertainty of weather conditions, drought has become a

recurrent phenomenon in many parts of the world and particularly in eastern African region as most of its agriculture is dependent up on rainfall [53, 19].

In the case of Ethiopia, lack of clearly summarized and up to date data on vegetation health condition has been observed as a critical problem. Until recently the efforts of drought monitoring and prediction in the country has been based on conventional methods, that rely on availability of metrological data, which is very tedious and time consuming to gather[11, 19]. And this way of drought monitoring and prediction needs to be improved.

These days, using remote sensing for drought monitoring and prediction is common. In Ethiopia the idea of using it for drought monitoring and prediction is a new area of study. And our study tries to contribute to this area. We will integrate some environmental variables and remotely sensed data with climate data to build a model which can be used in monitoring and predicting drought and see if accuracy of the information we get is improved relative to results obtained in [19].

In the study, we try to answer the following question:

1) Does using data with higher temporal resolution give better prediction information on drought condition?
2) Can using dekadal data prediction give better results compared to monthly average in monitoring drought?

Here we compare our results with results obtained in [19] which used monthly average data.

1.3. Objectives

General Objective

The general objective of the study is to evaluate the drought prediction and monitoring accuracy using models generated with higher temporal resolution satellite data.

Specific Objective

The specific objectives are listed as follows

- Collecting and preprocessing required data to build the model
- Building drought prediction model
- Validate the proposed model
- Implementing the model
- Develop a system for predicting drought
- Draw conclusions and recommendations

1.4. Scope and Limitation of the Study

The study area covers only Ethiopia. A total of 2812 sample points (stations) were selected randomly, each distanced 20 km approximately. And our only focus in the study is the growing season which starts on June and ends at October.

1.5. Data and Methods

Data collection

Four types of imagery data were used in the study: Satellite, Climate, Biophysical and Oceanic data. The data for the 2812 sample points were collected from different sources. Satellite vegetation data, NDVI, were gathered from National Oceanic and Atmospheric Administration (NOAA). Climate data, Rainfall Estimate Data (RFE), was collected from Tropical Applications of Meteorology using Satellite (TAMSAT). Biophysical data were collected from United States Geological Survey center for Earth Resources Observation and Science (USGS-EROS) and Royal Danish academy of Sciences. And oceanic index data were collected from Global Change Master Directory of NOAA.

Data preprocessing and transformation

After the imagery data collection, the data were preprocessed to make them ready for model development. And then coordinate systems of the all the imagery data were transformed to the same coordinate system so that all the data fit geometrically.

Developing a model

Using huge amount of historical data set a model consisting of certain rules were developed which were then used later in predicting and monitoring drought.

Model Validation

The developed model's prediction accuracy is assessed using test data.

Implementing the model

The developed models are then implemented using input data. When the models are implemented, the user decides the time period (dekadal or 10-days period in the growing season) for which drought is predicted.

Drought prediction system development

A Drought prediction system is then developed. The system accepts input data (images), preprocessed to a certain level, computes the level of drought and shows the result.

Software tools used

I. MATLAB

MATLAB is a high-level language and interactive environment that enables you to perform computationally intensive tasks faster than with traditional programming languages such as C, C++, and Fortran. With it you can perform algorithm development, data visualization, data analysis etc. our proto type system is developed using Matlab.

II. ILWIS

ILWIS is a software application which integrates GIS and Remote Sensing options. ILWIS integrates image, vector and thematic data in one unique and powerful desktop package. ILWIS will provide you a wide range of features including import/export, digitizing, editing, analysis and display of data, as well as production of quality maps. In satellite imagery it will be used for creation of histograms, color composites, sampling and classification, filtering, multi-band statistics. In our study it is used in calculating Standardized Seasonal Greenness (SSG).

III. ArcGis

ArcGis is a geographic information system (GIS) for working with maps and geographic information. It is used for: creating and using maps; compiling geographic data; analyzing mapped information; sharing and discovering geographic information; using maps and geographic information in a range of applications. In our study it is used for point values extraction, constant raster map creation, interpolation, and other purposes.

IV. ERDAS

ERDAS is a remote sensing image processing application. Here it was used to create equal extent for maps with different resolution.

V. Cubist

Cubist is a tool used for generating rule-based predictive models from data. Cubist models predict numeric values. Here cubist is used to develop the models which are used for model prediction.

VI. MapCubist

MapCubist software is a tool developed at the USGS Center for EROS and in our study it is used for implementing the models developed by Cubist i.e., generating drought prediction maps.

1.6. Significance of the Study

The study proves whether using data with higher temporal resolution (i.e., 10-days composite data) in a model construction, which monitors and predicts drought, increases the prediction accuracy or not and making accurate predictions if using dekadal data increases the models prediction accuracies. The study also encourages other researchers in the country to contribute in the area.

1.7. Document Organization

The remaining part of this document is organized in the following way. Literature review and related work are dealt in Chapter two. Chapter three discusses about the study area and data, Chapter four focuses on drought prediction model. Drought prediction system comes in Chapter

five then Chapter six covers discussion of results. And finally in Chapter seven conclusion and future work is presented.

2. LITERATURE REVIEW AND RELATED WORK

2.1. Literature Review

Drought monitoring has the objective of warning about a possible incoming drought, providing adequate information for an objective drought declaration and for avoiding severe water shortages.

Different drought indices and methods have been proposed since the '60s to identify and monitor drought events. Some of these indices describe meteorological drought which is based on precipitation series, while others are oriented to refer to hydrological or agricultural drought or water shortages in urban water supply systems. Drought indices incorporate thousands of bits of data on rainfall, snowpack, streamflow, and other water supply indicators into a comprehensible big picture. A drought index value is typically a single number, far more useful than raw data for decision making.

The expected outcome is the characterization of the meteorological and hydrological drought periods in the historical record in each geographical unit. The correct drought characterization provides decision makers with a measurement of the abnormality of historical weather variability and its effects on a region. None of the indices is inherently superior to the rest in all circumstances; but some indices are better suited than others for certain uses. There are two known drought monitoring approaches: 1.Conventional approach, and 2. Remote sensing-based approach.

I. Conventional approach of drought monitoring

Conventional approaches are also known as traditional or ground based approaches.

This approach includes indices for monitoring meteorological, hydrological, and agricultural drought. Among the indices, Meteorological Drought indices include Palmer Drought Severity Index (PDSI), Standardized Precipitation Index (SPI), Rainfall Anomaly Index (RAI), and others. And some of the Hydrological Drought indices are Total Water Deficit Traditional (TWDT), Palmer Hydrological Drought Severity index (PHDS), and Surface Water Supply

Index (SWSI) whereas agricultural drought indices include Crop Moisture Index (CMI), Palmer moisture anomaly index (z-index), and Computed Soil Moisture index.

However, these indices have major drawbacks which are lack of spatial details and they are also dependent on data collected at ground stations which sometimes are sparsely distributed affecting the reliability of the drought indices [32, 38].

Some of the indices are discussed below.

Palmer Drought Severity Index (PDSI)

Palmer Drought Severity Index (PDSI) is the first comprehensive drought index developed by W.C. Palmer in 1965 [68]. The PDSI is calculated based on precipitation and temperature data, as well as the local Available Water Content (AWC) of the soil. The PDSI responds to weather conditions that have been abnormally dry or abnormally wet. When conditions change from dry to normal or wet, for example, the drought measured by the PDSI ends without taking into account streamflow, lake and reservoir levels, and other longer-term hydrologic impacts [62]. It is a soil moisture algorithm calibrated for relatively homogeneous regions. The Palmer Index varies roughly between -6.0 and +6.0. Table 2.1 shows the palmer classification.

Table 2.1: Classification of drought using PSDI values [68]

Palmer Classifications	
4.0 or more	extremely wet
3.0 to 3.99	very wet
2.0 to 2.99	moderately wet
1.0 to 1.99	slightly wet
0.5 to 0.99	incipient wet spell
0.49 to -0.49	near normal
-0.5 to -0.99	incipient dry spell
-1.0 to -1.99	soft drought
-2.0 to -2.99	moderate drought
-3.0 to -3.99	severe drought
-4.0 or less	extreme drought

Standardized Precipitation Index (SPI)

The Standardized Precipitation Index (SPI) is a tool which was developed primarily for defining and monitoring drought. It is an index based on the probability of precipitation for any time scale. The SPI was formulated by T. B. McKee et. [60]. SPI allows to determine the rarity of a drought at a given time scale (temporal resolution) of interest for any rainfall station with historic data. It assigns a single numeric value to a precipitation which can be compared across regions with markedly different climates. Originally the SPI was calculated for 3-, 6-, 12-, 24-, and 48-month time scales. Table 2.2 shows the classification of drought using SPI values.

Table 2.2: Classification of drought using SPI values [60]

SPI Values	
2.0+	extremely wet
1.5 to 1.99	very wet
1.0 to 1.49	moderately wet
-.99 to .99	near normal
-1.0 to -1.49	moderately dry
-1.5 to -1.99	severely dry
-2 and less	extremely dry

Crop Moisture Index (CMI)

Crop Moisture Index (CMI) uses a meteorological approach to monitor week-to-week crop conditions. It was developed by W. C. Palmer in 1968 from procedures within the calculation of the PDSI [69]. Unlike PDSI the CMI was designed to evaluate short-term moisture conditions across major crop-producing regions is not intended to assess long-term droughts. It is based on the mean temperature and total precipitation for each week within a climate division, as well as the CMI value from the previous week. The CMI responds rapidly to changing conditions, and it is weighted by location and time so that maps, which commonly display the weekly CMI across the United States, can be used to compare moisture conditions at different locations.

The Surface Water Supply Index (SWSI)

The Surface Water Supply Index (SWSI) (pronounced "swazee") was developed by B.A. Shafer and L.E. Dezman to complement the Palmer Index for moisture conditions across the state of Colorado [14]. SWSI was designed to be an indicator of surface water conditions and described the index as “mountain water dependent”, in which mountain snowpack is a major component.

The objective of the SWSI was to incorporate both hydrological and climatological features into a single index value resembling the Palmer Index for each major river basin in the state of Colorado. The SWSI value is centered on zero and has a range between -4.2 and +4.2. The SWSI has been used, along with the Palmer Index, to trigger the activation and deactivation of the Colorado Drought Plan. One of its advantages is that it is simple to calculate and gives a representative measurement of surface water supplies across the state.

II. Remote sensing-based approach of drought monitoring

Remote sensing is done by sensing and recording energy reflected from a target and processing, analyzing, and applying that information [4]. For a sensor to collect and record energy reflected or emitted from a target, it must be placed on a stable platform which is separated from the target being observed. Platforms for remote sensors may be stationed on the ground, on an aircraft, or satellite. In space, remote sensing is done commonly from satellites revolving around the earth. In this work we focus on satellites for remote sensing.

Satellites measure energy intensities (radiances) at several wavelengths of the electromagnetic spectrum, which is energy from sunlight. This information is useful because everything including the ground, the oceans, the atmosphere, clouds, rain, vegetation, cities, people, etc. absorbs energy at certain wavelengths and emits energy at other wavelengths.

Remote Sensing technology in its current state of art can help in predicting, mitigating and monitoring of drought. Data from various satellites can be utilized for this purpose irrespective of the perspective that a researcher has towards drought, whether it is agricultural, meteorological or hydrological. It enables to understand the manifestations of drought in a larger area more directly than through conventional methods, and of all, in less time consuming manner [13, 38].

In [65] it is indicated that satellite based vegetation monitoring plays an important role in drought monitoring and early warning, since vegetation condition reflects the overall effect of rainfall, soil moisture, weather and agricultural practices. Due to this fact, our focus is on vegetation condition monitoring and prediction. Using remote sensing, condition of vegetation

can be assessed by interpreting the reflectance values at various wavelengths of energy. Wavelengths are measured in micrometers (μm) or nanometers (nm). One μm equals 1,000 nm. The visible region of the electromagnetic spectrum is from about 400 nm to about 700 nm. The green color associated with plant health has a wavelength that centers near 500 nm. Figure 2.1 shows visible region of the electromagnetic spectrum.

Wavelengths longer than those in the visible region and up to about 25 μm are in the infrared region.

The infrared region nearest to that of the visible region is the near infrared (NIR) region. Both the visible and infrared regions are used in agricultural remote sensing.

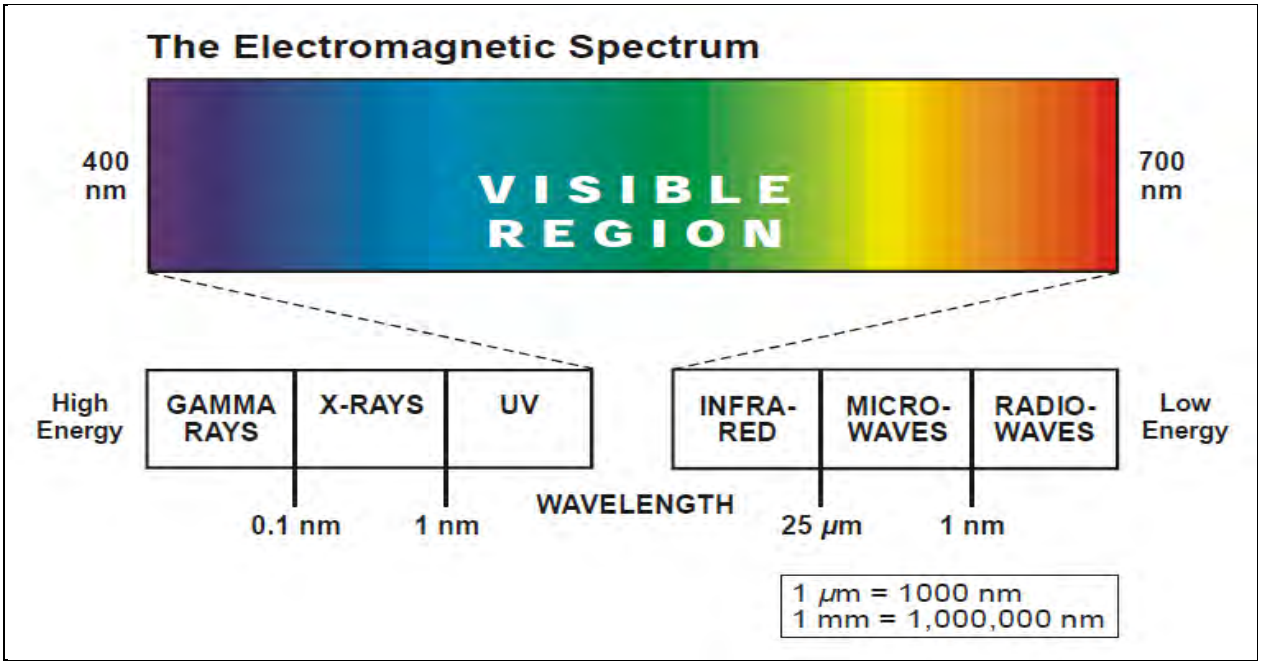


Figure 2.1: The visible region of the spectrum ranges from about 0.4 μm to 0.7 μm

Vegetation indices, the comparison of reflectance values at different wavelengths, are commonly used to determine vegetation health. A vegetation index is a gauge of plant health, productivity, and density. Plant leaves absorb visible light for use in photosynthesis, and they reflect near-infrared light. A light-measuring sensor looking at a healthy plant sees very little reflected visible light (red region) and lots of near-infrared light. Unhealthy or stressed vegetation increase its reflection in the red region of the spectrum and decrease in the infrared region of the spectrum.

The more healthy green leaves a plant has, the stronger the difference between the amount of visible and near-infrared light it reflects, and the higher the vegetation index.

The most common vegetative index is known as NDVI. NDVI is an index, or scale, of vegetation conditions based on differences in the amount of visible (red) and near-infrared light reflected from plants on Earth's surface.

NDVI values range from +1.0 to -1.0. Very low NDVI values (0.1 or less) usually show areas of barren rock, sand, or snow. Moderate NDVI values (0.2 to 0.5) denote sparse vegetation such as shrubs and grasslands or senescing crops. And high NDVI values (0.6 to 0.9) correspond to dense vegetation such as that found in temperate and tropical forests or crops at their peak growth stage [5].
Equation (2.1) shows how to calculate NDVI,

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}} \qquad \text{Equation (2.1)}$$

Where $_{\mathrm{NIR}}$ and $_{\mathrm{RED}}$ are the reflectance in the near infrared and visible (red) bands respectively.

The NDVI by itself does not reflect drought or non-drought conditions. But the effects of severity of drought on vegetation can be defined as the deviation of current NDVI values from their corresponding long-term mean NDVI values [52]. For each location, the deviation (DEV_{NDVI}) is calculated as the difference between the NDVI of the current time step and its corresponding long-term mean. Equation (2.2) shows the formula to calculate $\mathrm{DEV}_{\mathrm{NDVI}}$.

$$DEV_{NDVI} = NDVI_{i,m} - NDVI_{mean,m} \qquad \text{Equation (2.2)}$$

Where $NDVI_{i,m}$ is the current NDVI value i for each location for time step m and $NDVI_{mean,m}$ is the long-term mean NDVI for the corresponding calendar time step m for each location.

When DEV_{NDVI} is negative, it indicates below normal vegetation conditions and therefore suggests prevailing drought conditions; the greater the negative value, the greater the magnitude of drought severity.

In addition, other VIs has been developed to improve on the NDVI for vegetation monitoring. These include the enhanced vegetation index (EVI) [12], the wide dynamic range vegetation index (WDRVI) [8], the normalized difference water index (NDWI) [15, 42], and the vegetation condition index (VCI) [33].

Some of the vegetation indices are discussed below.

Enhanced Vegetation Index (EVI)

The Enhanced Vegetation Index (EVI) is an 'optimized' index designed to enhance the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring by de-coupling the canopy background signal and a reduction in atmosphere influences.

EVI is computed following this equation:

$$EVI = G \times \frac{(NIR - RED)}{(NIR + C1 \times RED - C2 \times Blue + L)} \qquad \text{Equation (2.3)}$$

where NIR/red/blue are atmospherically-corrected or partially atmosphere corrected (Rayleigh and ozone absorption) surface reflectances, L is the canopy background adjustment that addresses non-linear, differential NIR and red radiant transfer through a canopy, and C1, C2 are the coefficients of the aerosol resistance term, which uses the blue band to correct for aerosol influences in the red band. The coefficients adopted in the MODIS-EVI algorithm are; L=1, C1 = 6, C2 = 7.5, and G (gain factor) = 2.8. The value of this index ranges from -1 to 1. The common range for green vegetation is 0.2 to 0.8. The EVI is more responsive to canopy structural variations whereas NDVI is chlorophyll sensitive. The two VIs complement each other in global vegetation studies and improve upon the detection of vegetation changes and extraction of canopy biophysical parameters.

Vegetation Condition Index (VCI)

Vegetation Condition Index (VCI) is an indicator of the status of the vegetation cover as a function of the NDVI minimum and maximum encountered for a given ecosystem over many years. VCI is defined as:

$$VCI_j = (NDVI_j - NDVI_{min}) / (NDVI_{max} - NDVI_{min}) * 100 \qquad \text{Equation (2.4)}$$

Where, $NDVI_{max}$ $NDVI_{min}$ is calculated from long-term record for a particular month and j is the index of the current month.

The condition of the ground vegetation presented by VCI is measured in percent. The VCI values between 50 to 100 percent indicate optimal or above normal conditions whereas VCI values close to 0 percent reflects an extreme dry month.

2.2. Related Work

Drought monitoring is challenging because of the wide spatial and temporal variability exhibited by each drought event. No single method can be used to adequately characterize and monitor drought. Drought monitoring techniques should be adapted to capture the time, location, and sector-specific characteristics of drought. Improvements in drought monitoring techniques and products, such as higher spatial, temporal resolution and the timely delivery of information in a variety of accessible formats increase the value and relevancy of the information available to decision makers, and thereby enhance and support drought response and mitigation activities.

Since the 1980s, many studies have capitalized on the synoptic, timely, and spatially continuous characteristics of remotely sensed data gathered by satellites to analyze and monitor vegetation conditions over large areas [17, 41, 44, 48, 57, and 58].

Various studies have demonstrated the utility of satellite measurements for observing and monitoring drought and provide analyses of the relationships between climate variables (e.g., precipitation) and satellite-derived VIs [48, 49]. T. R. McVicar, and P. N. Bierwirth investigated the utility of satellite data as a drought assessment tool for the 1997 drought in Papua New Guinea and found a strong correlation (r^2=0.809) between accumulated rainfall and an integrated measurement of surface temperature (T_s) and NDVI over meteorological stations [61, 64]. Getachew proposed application of NDVI and SPI parameters to monitor drought at national scale, for Ethiopia. The study showed that, using near real-time MSG and historical NOAA AVHRR NDVI data with some calibration and validation, it is possible to identify and predict drought incidences [18]. Significant correlations between monthly NDVI and the SPI during the growing season over four states in the U.S. central plains were shown in [49]. Even though the study was based on spatially averaged NDVI and SPI data, they found NDVI to be an affective

indicator of moisture and vegetation condition [49]. And Sharma proposed a spatial data mining method for monitoring drought using temporal NDVI and rainfall in India. The VCI, calculated from NDVI was overlaid with rainfall data to identify drought areas [56].

Even though satellite-based observations have proven very useful for detecting vegetation condition anomalies (apparent declines in vegetation health), the specific cause for the vegetation stress may not always be determined solely from the remotely sensed data. A number of natural (for example, drought, flooding, fire, pest infestation, and hail damage) and anthropogenic (for example, land cover/land use conversion) events can produce these anomalies [21, 28, 33, 34, and 37]. Thus, remote sensing data provides a means to monitor detailed spatial patterns of vegetation conditions, but it is often difficult to distinguish drought-related vegetation stress from vegetation changes caused by these other drivers without additional information. As a result, the integration of coarser-resolution climate data and higher-resolution satellite-based vegetation observations provides an alternative approach to better monitor and characterize the spatial extent, intensity, and local variability of drought's affect on vegetation conditions [41] and also this integration of remotely sensed data with other relevant environmental variables (e.g., climate, ocean observations) that influence vegetation holds considerable potential for improving our capabilities to predict future vegetation conditions, as demonstrated in [22] and [49].

The development of methods to predict vegetation conditions from satellite based observations and other ancillary information has been limited [22, 49, and 65]. A new drought monitoring methodology called the Vegetation Drought Response Index (VegDRI), which integrates historical climate data and satellite-based earth observations with other biophysical information to produce a 1 km–resolution indicator of the geographic extent and intensity of drought stress on vegetation was developed by Jesslyn F.Brown [41]. The VegDRI methodology represents a new approach to drought monitoring by integrating traditional climate-based drought index information and satellite-based NDVI measures of vegetation conditions with several biophysical characteristics. This index specifically targets the effects of drought on vegetation by considering the general vegetation conditions as observed by satellite and the level of dryness experienced for a given location. Additional environmental characteristics are also represented in this

approach given the different climate-vegetation response relationships that can occur for different land cover types, soil types, and land use practices.

L. Ji and A. J. Peters developed a vegetation greenness forecasting model to predict future NDVI values for crops and grasslands in the United States Great Plains using 1 km AVHRR NDVI data and time-lagged precipitation and temperature information [49, 65]. C. C. Funk and M. E. Brown used AVHRR NDVI observations in combination with observed precipitation and relative humidity to project NDVI changes one to four months into the future for Africa [22, 65].

Tsegaye et al. presented a data mining approach for modeling vegetation stress due to drought and mapping its spatial extent during the growing season [64]. Rule-based regression tree models were generated from the study which identifies relationships between satellite-derived vegetation conditions, climatic drought indices, and biophysical data. The method enabled to provide predictions of vegetation conditions farther into the growing season based on earlier conditions, since it can be applied iteratively with input data from previous time period. J.C. Eidenshink stated that data mining technique maximizes the information contained in traditional drought indicators and integrates it with satellite-based greenness measures from AVHRR processed at the USGS-EROS Data Center [40, 64]. Tsegaye et al. developed a new approach, vegetation outlook, for predicting general vegetation condition patterns across large areas over USA [65]. Vegetation Outlook predicts SSG measure, which represents a general indicator of relative vegetation health. It predicts SSG values at multiple time steps (two to six weeks into the future) based on the analysis of historical patterns (each 1 km grid cell) of satellite, climate, oceanic and biophysical data. Getachew et al. developed models by integrating climate, satellite, biophysical and oceanic indices data to monitor and predict drought in Ethiopia [19]. The results from the study showed that the models perform well, up to 95% prediction accuracy for some cases.

Generally, the studies showed that integrating remotely sensed data with other relevant environmental variables like climate, oceanic and biophysical observations has increased the capabilities of predicting vegetation conditions.

Most of the works are done for countries where drought is not that much of a problem compared to Africa. And in Ethiopia only few studies have been conducted until recently. Among the most

recent works one is done by Getachew et al. [19]. They used monthly average data to determine vegetation condition by believing that the effect of drought can easily be observed on the overall monthly average. And in our study we try to improve on his work by increasing the drought prediction accuracy even better. We will use data with higher temporal resolution (dekadal data) to build a drought monitoring and prediction model and see if prediction accuracy has improved compared to using monthly average data.

3. THE STUDY AREA AND THE DATA USED

In this Chapter we deal with the following topics. In section 3.1 the study area is specified. Section 3.2 deals with data selection and description and finally section 3.3 talks about data preprocessing and transformation.

3.1. Study Area

The study area for this study covers the whole of Ethiopia. Ethiopia occupies the interior of the horn of Africa stretching between 3^0 and 14^0 N latitude and 33^0 and 48^0 E longitude, with a total area of 1.13 million km^2 [30].

3.2. Data description

Four types of attributes are used in the study including climate, satellite, biophysical, and oceanic attributes.

A total of 2812 sample points were selected randomly from the whole Ethiopia (each point at an approximate of 20 km). Figure 3.1 below shows the selected sample points.

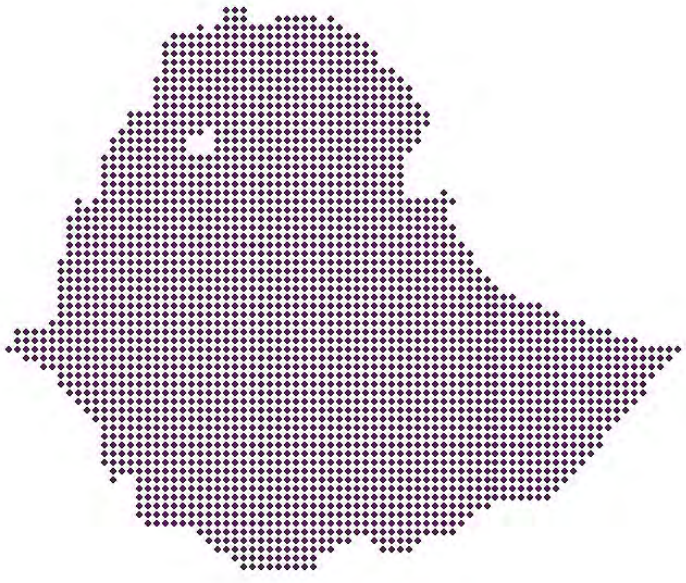


Figure 3.1: Selected sample points of Ethiopia

A 24 year historical data, for all the attributes, between the years 1983 and 2006 were used. The data on these years were found to be complete and easily accessible. The collected data are first preprocessed and then used in model construction.

Climate, satellite and oceanic attributes have dynamic values while biophysical attributes have constant values through the whole process.

The specific attributes used here were selected and used first by Getachew et al. [19] where they used Ordinary Least Square (OLS) to identify the key attributes used in drought prediction using the following equation.

$$y_{OLS} = a_0 + \sum_k a_k x_{ik} + \in_i \qquad \text{Equation (3.1)}$$

where y_{OLS} is the droughtObject, a_0 is the intercept, x_{ik} is the value of the predictor variable, a_k is the coefficient for the predictor variable, and $\in_i$ is the error.

SSG, which is derived from NDVI, a measure that shows general vegetation condition was selected as dependent variable in this work and it is discussed later briefly. This variable was selected as dependent variable due to the fact that NDVI can capture agricultural drought with optimum accuracy [19]. After identifying SSG as dependent variable, he then identified potential explanatory variables by reviewing available theory [50] and consultation with experts at DRMFSS and EMA (nationally) and NDMC (internationally).

The attribute used to build the model include SPI, SSG, landcover, veg_ethiopia, WHC, DEM, AMO, PNA, MEI, PDO, NAO. SSG and SPI were the most important attributes. These key attributes are listed in Table 3.1 below.

Table 3.1: Attributes identified for droughtObject modeling after theory analysis [19].

No	Attribute	Acronym	Unit	Type	Format
1	Standardized Seasonal Greenness	SSG	Unit less	Biophysical	Raster
2	Digital Elevation Model	DEM	Meter	Biophysical	Raster
3	Soil Water Holding Capacity	WHC	kg/m^3	Biophysical	Raster

4	Vegetation of Ethiopia	veg_Ethiopia	Nominal (numbers represent attributes)	Biophysical	Vector
5	Land Cover	Landcover	Nominal (numbers represent attributes)	Biophysical	Raster
6	Three Month Standard Precipitation Index	SPI_3month	Unit less	Climate	Raster
7	Pacific Decadal Oscillation	PDO	Unit less	Oceanic index	Point data
8	Atlantic Multi-decadal Oscillation Index	AMO	Unit less	Oceanic index	Point data
9	North Atlantic Oscillation	NAO	Unit less	Oceanic index	Point data
10	Pacific North American Index	PNA	Unit less	Oceanic index	Point data
11	Multivariate ENSO Index	MEI	Unit less	Oceanic index	Point data

3.2.1. Satellite Vegetation data

Normalized Difference Vegetation Index (NDVI)

NDVI data was obtained from NOAA's AVHRR sensor. Decadal (10-day) composite of NDVI images were obtained from Famine Early Warning System (FEWS). FEWS used AVHRR data to produce dekadal composite NDVI images. These are three 10-days composites of datasets per month; organized as day 1 to 10, day 11 to 20, and day 21 to 30. The reason for using 10-day composites of daily data is, NDVI values tend to be affected by factors such as atmospheric effects, clouds, soil effects, anisotropic effects and spectral effects. This effect is lowered by forming composite images from daily images [20], obtaining 7-, 10-, 15- or 30-day composite images. Vegetation changes happen typically at 10 to 15 days time [71].

A total of 24 years time series (1983-2006) of dekadal composites of 8km NDVI data is used to calculate the Standardized Seasonal Greenness (SSG), which shows the general condition of vegetation.

3.2.2. Climate data

Rainfall Estimate (RFE)

The RFE data was obtained from TAMSAT, which is produced by the TAMSAT group at Reading University in United Kingdom [19]. TAMSAT generates 30-year time series of rainfall estimates for Africa based entirely on data from the Meteosat satellite calibrated against local gauges. These estimates are combined with the observations from over 600 rain gauges [19]. The result is a unique, high-quality dataset, which has been shown to be more accurate than any other long-term, satellite-based time series data [63].

RFE products from year 1983 to 2006, which were accumulated to 10-daily totals, were obtained from International Research Institute for Climate and Society [39] with spatial resolution of 10km. These data were used to calculate SPI which allows to determine the rarity of a drought at a given time scale (temporal resolution) of interest for any rainfall station with historic data. SPI is discussed briefly in section 3.3.

3.2.3. Oceanic indices

The oceans play a significant role in shaping the complex nature of weather and climate, through interactions with the atmosphere. Since the positions of high and low pressure centers over various parts of the world, these interactions produce recognizable, repetitive, and alternative weather that affect vegetation conditions. These patterns can occur very far away from their oceanic triggers (i.e., teleconnection) and are a natural part of climate [55]. Teleconnections represent anomalies related to each other at large distances of thousands kilometers [46].

Oceanic indices have been widely used in weather and climate forecasts because they collect quantitative data about current conditions that can be used to project how the complex relationships with climate and drought conditions evolve [58]. In this work, for investigating impact of oceanic-atmospheric dynamics and climate-drought interactions (i.e., teleconnection patterns) the following oceanic indices were selected: Atlantic Multi-decadal Oscillation Index (AMO) [27, 51], Pacific Decadal Oscillation (PDO) [51], North Atlantic Oscillation (NAO) [45, 51, and 52], Multivariate ENSO Index (MEI) [47, 51], and Pacific North American Index (PNA) [51].

A. Atlantic Multi-decadal Oscillation Index (AMO)

The Atlantic Multi-decadal Oscillation (AMO) is a mode of variability occurring in the North Atlantic Ocean and which has its principal expression in the sea surface temperature (SST) field.

B. Pacific/North American (PNA) Index

The PNA is the dominant mode of low-frequency climate variability in the northern hemisphere's extra-tropical regions.

C. Multivariate ENSO Index (MEI)

MEI is calculated based on six main observed variables over the tropical Pacific. These six variables are: sea-level pressure, zonal and meridional components of the surface wind, sea surface temperature, surface air temperature, and total cloudiness fraction of the sky. A positive MEI is associated with El Niño (warm ocean phase) and negative values indicate La Niña (cold ocean phase) conditions.

D. The Pacific Decadal Oscillation (PDO) Index

The PDO index is defined as the leading principal component of North Pacific monthly sea surface temperature variability pole ward of 20N. The positive values show the warm phase of the North Pacific sea surface temperature, and the negative values show the cold phase.

E. North Atlantic Oscillation (NAO) Index

The NAO index is defined as the normalized pressure difference between a station on the Azores and one on Iceland [45]. Since it is computed based on stations to the north (Iceland) and south (Azores) of the middle latitude westerly flow, it could be considered as a measure of the strength of these winds. The positive values show strong mid-latitude westerly flow while the negative values show the weak mid-latitude westerly flow [45].

3.2.4. Biophysical Data

Four Biophysical attributes were included in the study. These include Water Holding Capacity (WHC), Land Cover, Digital Elevation Model (DEM), and Vegetation Type of Ethiopia.

i. Water Holding Capacity (WHC)

Water Holding Capacity (WHC) was included with assumption that when soil has high WHC it can resist drought severity and when soil has low WHC it has low resistance to drought severity. These properties of the soil eventually affect the vegetation condition. The WHC data was obtained from USGS- EROS Center with a spatial resolution of 1km [67].

ii. Land Cover

Land cover is the physical material at the surface of the earth [35]. This attribute is included for representing the different NDVI signals and climate-agricultural drought responses that are exhibited by different land cover types (such as crop land versus forest land). The land cover data is derived by an automatic and regionally-tuned classification of a time series of global MERIS mosaics for the year 2009 with spatial resolution of 300m [31]. This data was found to be one of the latest with high spatial resolution. For Ethiopian extent a total 14 different land cover classes were identified. Land cover classes represented on the original raster data and their explanation are presented in Table 3.2 below.

Table 3.2: Land cover classes with description [31, 19]

No	Raster attribute value	Label descriptions
1	14	Rainfed croplands
2	20	Mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%)
3	30	Mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20-50%)
4	40	Closed to open (>15%) broadleaved evergreen or semi-deciduous forest (>5m)
5	60	Open (15-40%) broadleaved deciduous forest/woodland (>5m)
6	110	Mosaic forest or shrubland (50-70%) / grassland (20-50%)
7	120	Mosaic grassland (50-70%) / forest or shrubland (20-50%)
8	130	Closed to open (>15%) (broadleaved or needleleaved, evergreen or

		deciduous) shrubland (<5m)
9	140	Closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses)
10	150	Sparse (<15%) vegetation
11	180	Closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged soil - Fresh, brackish or saline water
12	190	Artificial surfaces and associated areas (Urban areas >50%)
13	200	Bare areas
14	210	Water bodies

iii. Digital Elevation Model (DEM)

Digital Elevation Model is a digital model or 3D representation of a terrain's surface of the planet earth [72]. The DEM data was obtained from USGS-EROS Center with a spatial resolution of 1km [67]. The DEM was included to account for the diversified elevations of Ethiopia (that ranges from -120m to 4600 above sea level), which is assumed to have great influence on the spatial extent of agricultural drought.

iv. Vegetation Type of Ethiopia (veg_ethiopia)

Vegetation type is a general term for the type of plant life of a region referring to the ground cover provided by plants. Ecosystems of Ethiopia are mainly determined based on vegetation types [70]. Using vegetation type as major attribute, Ethiopia has about fourteen major ecosystems [70, 29]. The data was obtained from Royal Danish Academy of Sciences [29]. The ecosystem data (veg_Ethiopia) was included to account for the different response of the different vegetation types (such as semi-arid desert vegetation and evergreen afromontane forest), which assumed to have great influence on the spatial extent and severity of agricultural drought. The raster attribute value and label description of this data is presented in Table 3.3 below.

Table 3.3: Ecosystem of Ethiopia (based on vegetation type) raster data classes with description [19]

No	Raster attribute value	Label descriptions
1	1	Desert and Semi-desert Scrubland (DSS)
2	2	Acacia-Commiphora woodland and bushland proper (ACB)
3	3	Acacia wooded grassland of the Rift Valley (ACB/RV)

4	4	Wooded grassland of the Western Gambela region (WGG)
5	5	Combretum-Terminalia woodland and wooded grassland (CTW)
6	6	Dry evergreen Afro-Montane Forest and Grassland complex (DAF)
7	7	Moist Evergreen Afro-Montane Forest (MAF)
8	8	Transitional Rain Forest (TRF)
9	9	Ericaceous Belt (EB)
10	10	Afro-alpine vegetation (AA)
11	12	Freshwater lakes - open water vegetation (FLV/OW)
12	13	Freshwater marshes and swamps, Floodplains and Lake Shore vegetation (FLV/MFS)
13	14	Salt Lake open water vegetation (SLV/OW)
14	15	Salt Pans, Saline/brackish and Intermittent wetlands and Salt-lake Shore Vegetation (SLV/SSS)

The overall assumption in this study is that each attribute contribute to the complex nature of the vegetation growth. But these attributes alone may not be enough to explain the vegetation dynamics. The eleven attributes (from climate, satellite and biophysical data's) to be integrated are considered to be the most important parameters for Ethiopia case in contributing to vegetation monitoring and prediction.

3.3. Data Pre-processing and Transformation

Originally the imagery data used in the study were found with different spatial resolution and therefore they were re-sampled to 8km, since SSG (the dependent variable in the model to be built) was acquired with spatial resolution of 8km. The coordinate systems of the other attributes were also changed to that of NDVI data (source for SSG) so that they fit geometrically.

For a total of 2812 samples (which is almost one point sample on every 20 km) satellite image data, for each attribute, were collected. Then point values were extracted from the images and used in model building.

The detailed preprocessing done on the individual attributes are discussed below.

RFE Data

RFE point data were extracted from RFE images. A total of 2812 point data were extracted for Ethiopia. Then the data was exported to excel. And using Egor software the excel data were changed to a format which makes it ready for input to SPI program. SPI is derived from RFE data and it can effectively represent the amount of rainfall over a given time scale, with the advantage that it provides not only information on the amount of rainfall, but that it also gives an indication of what this amount is in relation to the normal, thus leading to the definition of whether a station is experiencing drought or not. Mathematically, SPI is based on the cumulative probability of a given rainfall event occurring at a station. It is calculated by fitting historical precipitation data to a Gamma probability distribution function for a specific time period and location, and transforming the Gamma distribution to a normal distribution with a mean of zero and standard deviation of one [19].

SPI was incorporated to represent meteorological dryness during the growing season (June to October). The analysis has been restricted to a time period of roughly aligned to the growing season since the effect of precipitation deficits on the spatial patterns of vegetation health is our primary focus.

SPI can be interpreted at different time scales (3-month, 6-month, 9-month, 12-month SPI and so on) and for our case, 3-month SPI is preferred because it reflects short- and medium-term moisture (drought) conditions and provides a seasonal estimation of precipitation[3,19]. The 3-month SPI provides a comparison of the precipitation over a specific 3-month period with the precipitation totals from the same 3-month period for all the years included in the historical record. In other words, a 3-month SPI at the end of August compares the June-July-August precipitation total in that particular year with the June-August precipitation totals of all the years.

To calculate the 3-month SPI a C code program, which uses monthly RFE data, was obtained from National Drought Mitigation Center, University of Nebraska-Lincoln. The C code was automated so that dekadal data can be used. Then SPI was calculated for each dekad of the growing season. The results are then ready to be used in model building. And raster images were

generated for each dekad periods using an inverse distance weighting (IDW) interpolation method in ArcGis, which is used in model implementation later.

NDVI Data

NDVI image was found for the whole of Africa and sub map of Ethiopia was re-sampled from this image.

A geo-reference was created in ILWIS software to do the re-sampling. The data was found in 8 bit format and according to the metadata, water and cloud pixels were with value 255 and bad NDVI with value 253. So these values were removed and NDVI image values between the range -1 and 1 was obtained by dividing the raw NDVI values by 255 using ILWIS software.

For monitoring vegetation conditions these sequential NDVI image values across the growing season are summarized to a metric such as seasonal greenness (SG) (also referred to as time-integrated NDVI), which can be considered as the general proxy for vegetation performance (i.e., gross primary productivity [GPP]) [16, 41]. The SG represents the accumulated NDVI above a baseline "latent" NDVI value from the start of the growing season (SOS) to a specified time during the year. The baseline NDVI value represents the non-vegetated background signal from soil and non-photosynthetic plant waste. SG values were calculated for the whole Ethiopia for each 10-days period of the growing season in 24-year time series. Equation 3.2, adopted from [65], shows how SG is calculated:

$$SG = \sum_{P1=SOS}^{p_n=EOS} NDVI_i = \sum_{p_1}^{p_n} \left(NDVI_i - NDVI_b\right) ... + \sum_{p_{n-1}}^{p_n} \left(NDVI_i - NDVI_b\right) \qquad \text{Equation (3.2)}$$

where P_1, P_2, P_3 … P_n refer to the 10-day periods; P_1 is the time period for the historical median SOS, and P_n is the time period for the historical median end of the growing season (EOS). The $NDVI_b$ is the NDVI value at a base line. The month May's NDVI value (before the start of the growing period in June) was taken as $NDVI_b$ value by taking the assumption that all background NDVI values which are out of the growing period values can be removed by subtracting this month's NDVI values. High SG values reflect high green biomass conditions, whereas low SG values reflect lower biomass levels.

After calculating SG, standardized seasonal greenness (SSG) was calculated from pixel-level SG values. SSG provides a measure to compare general vegetation conditions for a 10-days period in a specific year with historical average conditions for that same period in the growing season over the 24-year satellite data.

SSG values were calculated at dekadal time steps during the growing season for each year using a standardization formula [9, 65] which is shown in equation 3.3 below:

$$SSG = \frac{SG_i - \overline{SG}}{\sigma} \qquad \text{Equation (3.3)}$$

where SG_i is the seasonal greenness at a particular period within a growing season for a specific year, $\overline{SG}$ is the 24-year historical mean seasonal greenness for the same period, and σ is the standard deviation of the historical record. SSG representing agricultural drought severity is classified in to seven categories as in SPI [19, 65]. These seven classes are presented in Table 3.4.

Table 3.4: Drought severity classes using SSG values [19]

No	SSG Value	Vegetation Condition	Drought Severity Classes
1	-2.0 and less	Extreme Stress	Extreme Drought
2	-1.0 to -2.0	Sever Stress	Severe Drought
3	-1.0 to -0.5	Poor Vegetation	Moderate Drought
4	-0.5 to 0.5	Fair (Near Normal)	Near Normal
5	0.5 to 1.0	Good Vegetation	Moist
6	1.0 to 2.0	Very Good Vegetation	Very Moist
7	2.0 and greater	Excellent Vegetation	Extreme Moist

Both the SG and SSG values for each dekad were calculated using ILWIS software. Serious of long codes were written for calculating both values. After calculating SSG, point data were extracted from the images using ArcGis. These point data are then used for model building. And from those point data raster images were generated for each dekad period SSG values, using an inverse distance weighting (IDW) interpolation method in ArcGis. The raster images were then used in model implementation.

Oceanic Indices

A single oceanic index value was obtained for each month in the growing season of the 24 years (1983-2006). For building the model one month oceanic index value is used three times for the three dekads in a month. This is due to SSG value (which is dependent variable in the model) being computed for each dekad. And the single oceanic index value is used for the whole 2812 sample points in a dekad for model building. Constant raster maps were created in ArcGis using these single monthly values which are then used in model implementation.

Biophysical data

Biophysical attributes point data were extracted from their images. In the model building, since biophysical attributes are considered static, values of the 2812 points are not changed for the 24 years. Here a raster map was also created to be used in model implementation phase.

4. DROUGHT PREDICTION MODEL

4.1. Drought prediction model development

The preprocessed data were then used in model construction. In this study, regression tree analysis technique was applied for building the drought prediction model. Regression tree techniques have the ability to identify complex historical relationships between the suite of climate-, oceanic-, and satellite-based parameters and static biophysical attributes that can be used to predict future vegetation conditions (i.e., SSG). The basic concept is that this statistically based approach searches for and identifies similar patterns in the climate, oceanic, biophysical and satellite records to those of a specific prediction period to base the vegetation condition (or SSG) predictions[65].

A commercial CART algorithm called Cubist [54, 7] was used to analyze the historical data and generate rule-based, linear regression drought prediction models for each dekad. CART algorithm performs a binary, recursive partitioning process that splits the initial set of training observations (root or parent node) into two child nodes that each contains a subset of more homogeneous training observations. This process is repeated, further subdividing the training data into pairs of child nodes until the partitioning process is terminated by user-defined criteria.

Cubist builds a model containing one or more rules, where each rule is a conjunction of conditions associated with a linear expression. The meaning of a rule is that, if a case satisfies all the conditions, then the linear expression is appropriate for predicting the dependent (target) value.

In building the model two essential files (names file and data file) and one optional file (test file) were required. The data file provides information on the training cases that Cubist uses to construct a model whereas the names file defines the attributes used to describe each case, and test file (unseen cases in model construction) is used to evaluate the model. For the study all the three files were created. Figure 4.1 below shows the interface for cubist.

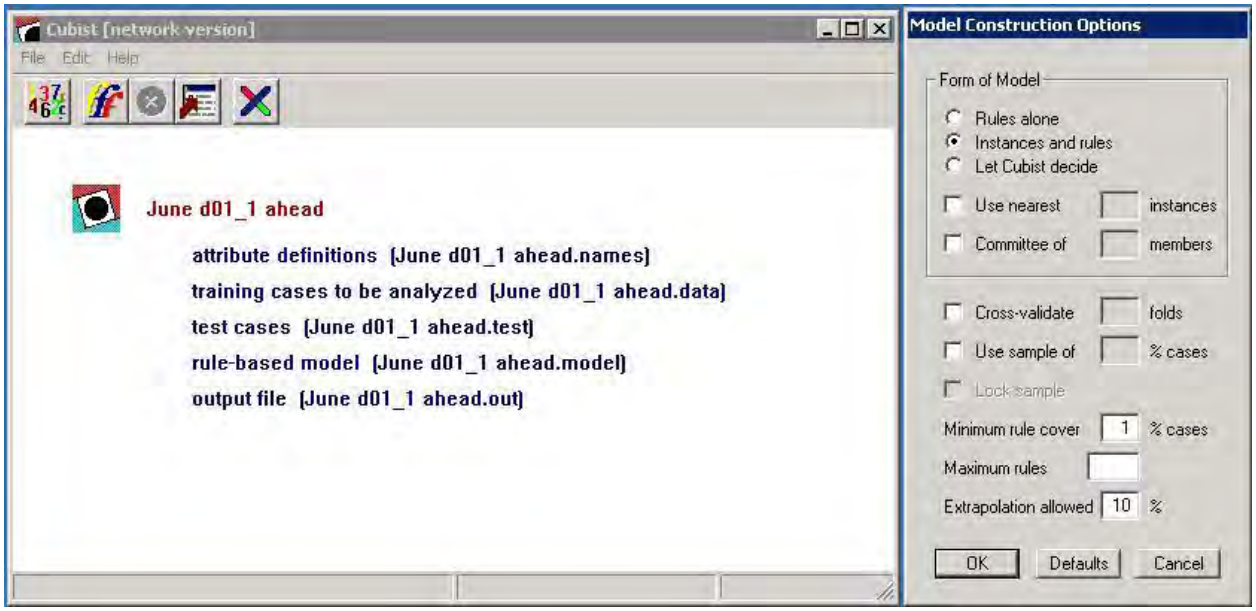


Figure 4.1: Interface of Cubist software

The drought prediction model was developed by using historical values of SPI, DEM, Veg-Ethiopia, WHC, Landcover, PDO, AMO, PNA, MEI, and NAO as independent attributes while SSG values were used in two ways. First, the "current" dekad (e.g., June 1) values were used as independent attributes to set the "initial" vegetation conditions from which prediction is calculated. Secondly, the next dekad (e.g., June 2) values for which predictions are calculated were used as dependent (target) attribute to establish the historical patterns of vegetation condition evolution over this interval in the model. Thus, the actual observed values of the SSG are used to "initialize" the prediction of future values of the SSG. This type of approach was used in [65].

Each dekadal period in the growing season having 2812 points, a total of 67488 point data from the 24 years were extracted for each variable (attribute) and used in the model building. In model building process, first the data were randomly split into two sets: training dataset (80%) and a test dataset (20%). Accordingly, a total of 53990 records were used for training and 13498 records for testing then all attribute values were multiplied by 100 to make them integer. This was done to make it suitable for model implementation later, since MapCubist (a tool to implement the model) accepts only integer values. The values are divided by 100 after model implementation.

After model building four files were produced (model, pred, out and set files). The model file is the rule-based model produced by Cubist. Pred contains the actual and predicted target values for any test cases. While out file is the report produced when a model is generated and set file is the setting used for the last model.

The following rules provide an example of the rule based models generated by the Cubist algorithm:

Cubist [Release 2.02]

Read 53451 cases (12 attributes) from July d02_1 ahead.data

Model:

Rule 1: [1190 cases, mean -170.4, range -370 to 0, est err 36.6]

if

July_SSG_dekad_2<= -60

then

July_SSG_dekad_3 = -147.6 - 0.023 DEM - 0.43 AMO + 0.1 WHC

Rule 2: [1009 cases, mean -154.1, range -360 to -20, est err 41.7]

if

July_SSG_dekad_2<= -20
Landcover in {20, 30, 40, 60, 120, 130}
veg_Ethiopia in {4, 5, 7, 8}
PDO <= 74
PNA > -89

then

July_SSG_dekad_3 = -54.5 - 0.06002 July_SPI_dekad_2+ 2.64 July_SSG_dekad_2- 0.51 PDO + 1.34 AMO - 0.07 PNA - 0.004 DEM

Rule 3: [1007 cases, mean -128.1, range -380 to 30, est err 38.0]

if

July_SSG_dekad_2<= -20
veg_Ethiopia in {4, 5, 7, 8}
PDO > 74
PNA > -89

then

July_SSG_dekad_3 = 95.8 - 0.74 NAO + 2.09 July_SSG_dekad_2- 0.56 PDO + 0.78 AMO - 0.014 DEM - 8e-005 July_dekad_2x

Rule 4: [1984 cases, mean -116.9, range -380 to -10, est err 30.9]

if

DEM > 1278
July_SSG_dekad_2> -60
July_SSG_dekad_2<= -20
veg_Ethiopia in {1, 2, 3, 6, 9, 10, 12, 13, 14, 15}
AMO > -9.6

then

July_SSG_dekad_3 = -42.8 + 3.23 July_SSG_dekad_2+ 0.00031 July_SPI_dekad_2+ 0.13 NAO + 0.07 PDO + 0.14 WHC - 0.07 PNA + 0.29 AMO - 0.007 DEM

Rule 5: [574 cases, mean -115.6, range -330 to -10, est err 32.9]

if

July_SSG_dekad_2<= -20
Landcover in {14, 110, 140, 150, 200}
veg_Ethiopia in {4, 5, 7, 8}
PNA > -89

then

July_SSG_dekad_3 = -106 + 2.5 July_SSG_dekad_2+ 0.038 DEM - 0.12 NAO + 0.1 PNA + 0.1 WHC - 8e-005 July_dekad_2x

Rule 6: [1028 cases, mean -101.1, range -320 to 10, est err 24.9]

if

DEM > 551
DEM <= 1278
July_SSG_dekad_2<= -20
Landcover in {14, 20, 40, 60, 130, 200}
veg_Ethiopia in {1, 2, 3, 6, 9, 10, 12, 13, 14, 15}
AMO > -9.6

then

July_SSG_dekad_3 = -27.5 + 0.01982 July_SPI_dekad_2+ 2.31 July_ssg_dekad_2x

Rule 7: [616 cases, mean -88.7, range -380 to 40, est err 40.8]

if

DEM <= 551
July_SSG_dekad_2> -60
July_SSG_dekad_2<= -10
Landcover in {14, 20, 40, 60, 130, 200}
veg_Ethiopia in {1, 2, 3, 6, 9, 10, 12, 13, 14, 15}

then

July_SSG_dekad_3 = -86.4 + 0.02727 July_SPI_dekad_2+ 2.47 July_SSG_dekad_2+ 0.055 DEM + 0.33 WHC + 0.13 PNA - 0.29 AMO

Rule 8: [3984 cases, mean -85.9, range -390 to 80, est err 23.5]

if

July_SSG_dekad_2<= -10

Landcover in {30, 110, 120, 140, 150, 180, 190, 210}

veg_Ethiopia in {1, 2, 3, 6, 9, 10, 12, 13, 14, 15}

July_SPI_dekad_2<= 96

AMO > -9.6

then

July_SSG_dekad_3 = -16.2 + 0.05991 July_SPI_dekad_2+ 2.69 July_SSG_dekad_2+ 0.006 DEM - 0.05 MEI + 0.04 PNA - 0.16 AMO - 0.03 NAO + 0.02 PDO

Rule 9: [1911 cases, mean -84.9, range -340 to 30, est err 26.9]

if

DEM > 551

July_SSG_dekad_2<= -10

Landcover in {14, 20, 40, 60, 130, 200}

veg_Ethiopia in {1, 2, 3, 6, 9, 10, 12, 13, 14, 15}

AMO <= -9.6

then

July_SSG_dekad_3 = 25.1 + 2.98 July_SSG_dekad_2+ 3.56 AMO + 0.33 MEI + 0.00024 July_SPI_dekad_2- 0.11 PNA + 0.1 NAO - 0.013 DEM + 0.09 PDO + 0.04 WHC

Rule 10: [1484 cases, mean -84.8, range -330 to 60, est err 24.6]

if

DEM > 1258

July_SSG_dekad_2> -60

July_SSG_dekad_2<= -10

Landcover in {30, 110, 120, 140, 150, 180, 190, 210}

veg_Ethiopia in {1, 2, 3, 6, 9, 10, 12, 13, 14, 15}

AMO > -9.6

then

July_SSG_dekad_3 = -36.8 + 3.28 July_SSG_dekad_2 + 0.00084 July_SPI_dekad_2+ 0.07 PDO - 0.08 MEI + 0.11 WHC + 0.18 AMO

Setting number of nearest neighbors to 3

Evaluation on training data (53451 cases):

Average |error|	22.0
Relative |error|	0.28

Correlation coefficient 0.95

*** Ignoring cases with unknown or N/A target value

Evaluation on test data (13365 cases):

Average |error| 22.2

Relative |error| 0.29

Correlation coefficient 0.95

For example **Rule 10** is interpreted as follows: If the data associated with a case met the threshold criteria for the attributes DEM, July_SSG_dekad_2, Landcover, veg_Ethiopia, AMO and were represented by one of the ten Veg_Ethiopia classes and one of the 8 classes of Landcover, as identified by Cubist, then the above multivariate linear regression equation was applied to calculate a drought prediction value (i.e., July_SSG_dekad_3). If two or more rules applied to a case, then the individual predictions from each regression equation were averaged to arrive at the final predicted value.

The remaining models are incorporated in a cd which can be found in the department.

4.2. Model Validation

The value of predictive models lies in their ability to make accurate predictions. It is difficult to judge the accuracy of a model by measuring how well it does on the cases used in its construction; the true predictive accuracy can be evaluated by using separate test cases which were not used to build the models. And the number of test cases used to evaluate the models should be large enough. Performance of the model on new cases is much more informative.

In evaluating the predictive accuracy of the model, 20% of the dataset (13498 cases) were used for testing and the rest 80% (53990 cases) for training randomly.

Correlation coefficient (r) metrics was used for validating our models. It measures the agreement between the cases' actual values of the target attribute and those values predicted by the model. The correlation coefficient was computed with Cubist software. The prediction accuracy on the test cases for all dekadal predictions are shown in Table 4.1.

Table 4.1: Models Prediction accuracies on test data

No	Model	Target Month	Correlation coefficient (r)
1	June d1_1 dekad ahead prediction	June dekad2	0.98
2	June d1_2 dekad ahead prediction	June dekad3	0.92
3	June d1_3 dekad ahead prediction	July dekad1	0.84
4	June d1_4 dekad ahead prediction	July dekad2	0.79
5	June d1_5 dekad ahead prediction	July dekad3	0.75
6	June d1_6 dekad ahead prediction	August dekad1	0.71
7	June d1_7 dekad ahead prediction	August dekad2	0.67
8	June d1_8 dekad ahead prediction	August dekad3	0.63
9	June d1_9 dekad ahead prediction	September dekad1	0.60
10	June d1_10 dekad ahead prediction	September dekad2	0.54
11	June d1_11 dekad ahead prediction	September dekad3	0.54
12	June d1_12 dekad ahead prediction	October dekad1	0.50
13	June d1_13 dekad ahead prediction	October dekad2	0.47
14	June d1_14 dekad ahead prediction	October dekad3	0.44
15	June d2_1 dekad ahead prediction	June dekad3	0.98
16	June d2_2 dekad ahead prediction	July dekad1	0.91
17	June d2_3 dekad ahead prediction	July dekad2	0.86
18	June d2_4 dekad ahead prediction	July dekad3	0.82
19	June d2_5 dekad ahead prediction	August dekad1	0.78
20	June d2_6 dekad ahead prediction	August dekad2	0.73
21	June d2_7 dekad ahead prediction	August dekad3	0.69
22	June d2_8 dekad ahead prediction	September dekad1	0.66
23	June d2_9 dekad ahead prediction	September dekad2	0.59
24	June d2_10 dekad ahead prediction	September dekad3	0.59
25	June d2_11 dekad ahead prediction	October dekad1	0.56
26	June d2_12 dekad ahead prediction	October dekad2	0.52
27	June d2_13 dekad ahead prediction	October dekad3	0.49
28	June d3_1 dekad ahead prediction	July dekad1	0.93
29	June d3_2 dekad ahead prediction	July dekad2	0.90
30	June d3_3 dekad ahead prediction	July dekad3	0.86
31	June d3_4 dekad ahead prediction	August dekad1	0.82
32	June d3_5 dekad ahead prediction	August dekad2	0.77
33	June d3_6 dekad ahead prediction	August dekad3	0.73
34	June d3_7 dekad ahead prediction	September dekad1	0.69
35	June d3_8 dekad ahead prediction	September dekad2	0.63
36	June d3_9 dekad ahead prediction	September dekad3	0.63
37	June d3_10 dekad ahead prediction	October dekad1	0.59

No	Model	Target Month	Correlation coefficient (r)
38	June d3_11 dekad ahead prediction	October dekad2	0.55
39	June d3_12 dekad ahead prediction	October dekad3	0.51
40	July d1_1 dekad ahead prediction	July dekad2	0.92
41	July d1_2 dekad ahead prediction	July dekad3	0.90
42	July d1_3 dekad ahead prediction	August dekad1	0.86
43	July d1_4 dekad ahead prediction	August dekad2	0.81
44	July d1_5 dekad ahead prediction	August dekad3	0.77
45	July d1_6 dekad ahead prediction	September dekad1	0.74
46	July d1_7 dekad ahead prediction	September dekad2	0.66
47	July d1_8 dekad ahead prediction	September dekad3	0.66
48	July d1_9 dekad ahead prediction	October dekad1	0.62
49	July d1_10 dekad ahead prediction	October dekad2	0.57
50	July d1_11 dekad ahead prediction	October dekad3	0.53
51	July d2_1 dekad ahead prediction	July dekad3	0.95
52	July d2_2 dekad ahead prediction	August dekad1	0.92
53	July d2_3 dekad ahead prediction	August dekad2	0.88
54	July d2_4 dekad ahead prediction	August dekad3	0.83
55	July d2_5 dekad ahead prediction	September dekad1	0.80
56	July d2_6 dekad ahead prediction	September dekad2	0.72
57	July d2_7 dekad ahead prediction	September dekad3	0.73
58	July d2_8 dekad ahead prediction	October dekad1	0.68
59	July d2_9 dekad ahead prediction	October dekad2	0.64
60	July d2_10 dekad ahead prediction	October dekad3	0.59
61	July d3_1 dekad ahead prediction	August dekad1	0.98
62	July d3_2 dekad ahead prediction	August dekad2	0.95
63	July d3_3 dekad ahead prediction	August dekad3	0.91
64	July d3_4 dekad ahead prediction	September dekad1	0.87
65	July d3_5 dekad ahead prediction	September dekad2	0.79
66	July d3_6 dekad ahead prediction	September dekad3	0.80
67	July d3_7 dekad ahead prediction	October dekad1	0.76
68	July d3_8 dekad ahead prediction	October dekad2	0.71
69	July d3_9 dekad ahead prediction	October dekad3	0.67
70	August d1_1 dekad ahead prediction	August dekad2	0.99
71	August d1_2 dekad ahead prediction	August dekad3	0.96
72	August d1_3 dekad ahead prediction	September dekad1	0.93
73	August d1_4 dekad ahead prediction	September dekad2	0.87
74	August d1_5 dekad ahead prediction	September dekad3	0.87
75	August d1_6 dekad ahead prediction	October dekad1	0.83

No	Model	Target Month	Correlation coefficient (r)
76	August d1_7 dekad ahead prediction	October dekad2	0.79
77	August d1_8 dekad ahead prediction	October dekad3	0.75
78	August d2_1 dekad ahead prediction	August dekad3	0.99
79	August d2_2 dekad ahead prediction	September dekad1	0.97
80	August d2_3 dekad ahead prediction	September dekad2	0.90
81	August d2_4 dekad ahead prediction	September dekad3	0.90
82	August d2_5 dekad ahead prediction	October dekad1	0.87
83	August d2_6 dekad ahead prediction	October dekad2	0.82
84	August d2_7 dekad ahead prediction	October dekad3	0.78
85	August d3_1 dekad ahead prediction	September dekad1	0.99
86	August d3_2 dekad ahead prediction	September dekad2	0.93
87	August d3_3 dekad ahead prediction	September dekad3	0.93
88	August d3_4 dekad ahead prediction	October dekad1	0.89
89	August d3_5 dekad ahead prediction	October dekad2	0.86
90	August d3_6 dekad ahead prediction	October dekad3	0.81
91	September d1_1 dekad ahead prediction	September dekad2	0.96
92	September d1_2 dekad ahead prediction	September dekad3	0.97
93	September d1_3 dekad ahead prediction	October dekad1	0.94
94	September d1_4 dekad ahead prediction	October dekad2	0.91
95	September d1_5 dekad ahead prediction	October dekad3	0.87
96	September d2_1 dekad ahead prediction	September dekad3	0.96
97	September d2_2 dekad ahead prediction	October dekad1	0.94
98	September d2_3 dekad ahead prediction	October dekad2	0.91
99	September d2_4 dekad ahead prediction	October dekad3	0.87
100	September d3_1 dekad ahead prediction	October dekad1	0.99
101	September d3_2 dekad ahead prediction	October dekad2	0.97
102	September d3_3 dekad ahead prediction	October dekad3	0.94
103	October d1_1 dekad ahead prediction	October dekad2	0.99
104	October d1_2 dekad ahead prediction	October dekad3	0.97
105	October d2_1 dekad ahead prediction	October dekad3	0.99

Cubist also computes the average and relative errors. The average error is the error for estimating the average value whereas relative error is the ratio of the average error magnitude to the error magnitude that would result from always predicting the mean value; for useful models, this value should be less than 1. Sample evaluation on test data is presented in Figure 4.2 below. Here cases with undefined or N/A target value are ignored.

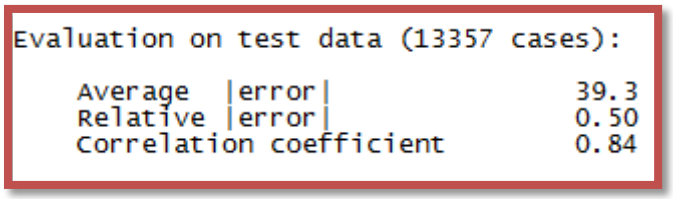


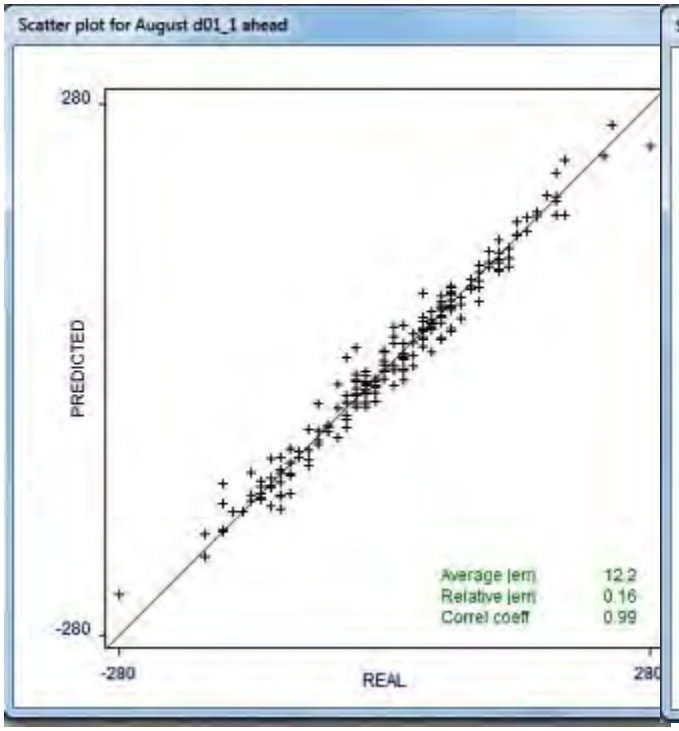
Figure 4.2: Sample evaluation on test cases

Evaluation on the test data generally shows that prediction model works greatly for near prediction periods (*r* value up to 0.99) and the *r* values gets lower continuously (*r* value up to 0.44) when the predicted dekad period is far from the dekad period which is used to make the prediction.

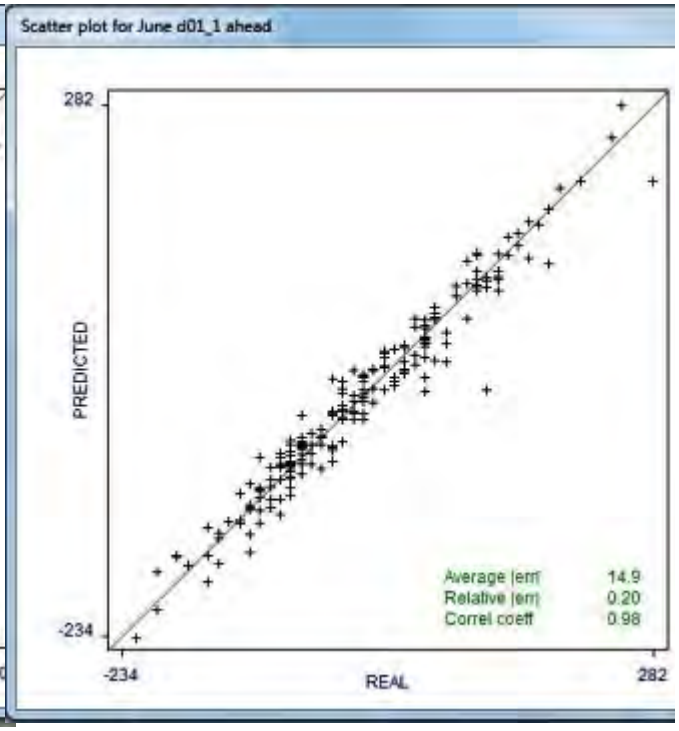
For every dekad period making the prediction the first two predictions (two dekad ahead predictions) has *r* values greater than or equal to 0.9 and the *r* values are greater or equal to 0.71 for the first six prediction periods. Far from the first six prediction periods the *r* values get very low to the point where it loses the prediction accuracy. This is due to the dekad period, which is used to make the prediction, being so far from the predicted period and existence of the two data in different parts of the growing season. For instance, June dekad one data can be used to predict, with certain accuracy, starting from June dekad 2 (one dekad ahead) with *r* value of 0.98 up to August dekad 1 (six dekads ahead) which has *r* value of 0.71. If we try to predict values far from that (greater than six dekad ahead) the *r* values goes down to a level where it can't give good result. This is because June is drier as it is beginning of growing season while August and September are wet and we find healthy plan growth in that period. So this leaves as with 69 models out of the 105 which are relevant in making drought predictions.

One other thing that can be seen from Table 4.1 is September dekad 2 and 3 has the same prediction values all the time. This may be due to September dekad two and three get the same amount of rainfall generally.

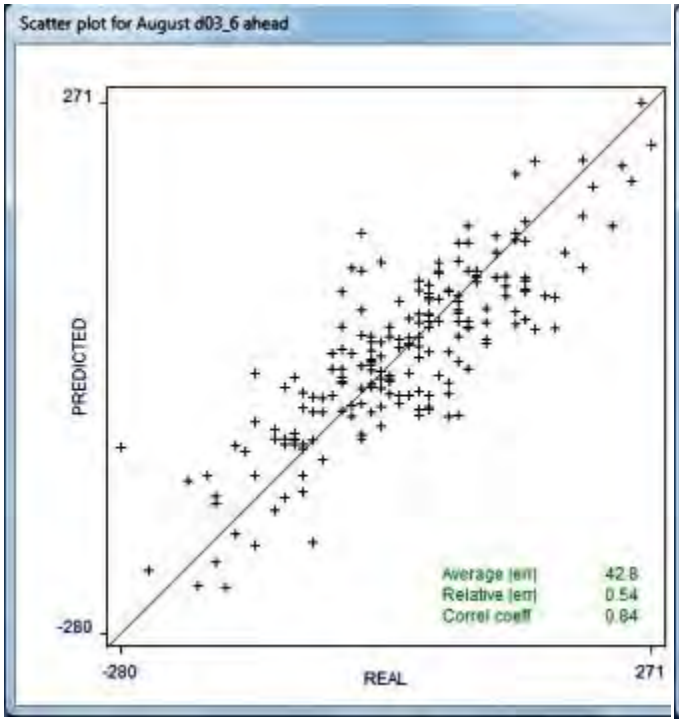
Scatter plots of some selected dekadal periods which shows the correlation between the actual and predicted values are displayed in Figure 4.3 below.


Scatter plot for August d01_1 ahead
280
PREDICTED
-280
-280
REAL
280
Average |err| 12.2
Relative |err| 0.16
Correl coeff 0.99

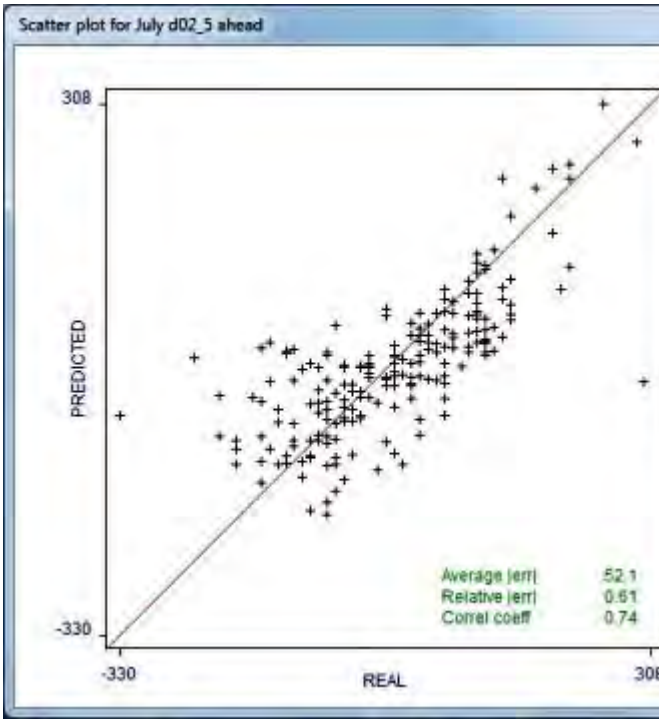
(a)


Scatter plot for June d01_1 ahead
282
PREDICTED
-234
-234
REAL
282
Average |err| 14.9
Relative |err| 0.20
Correl coeff 0.98

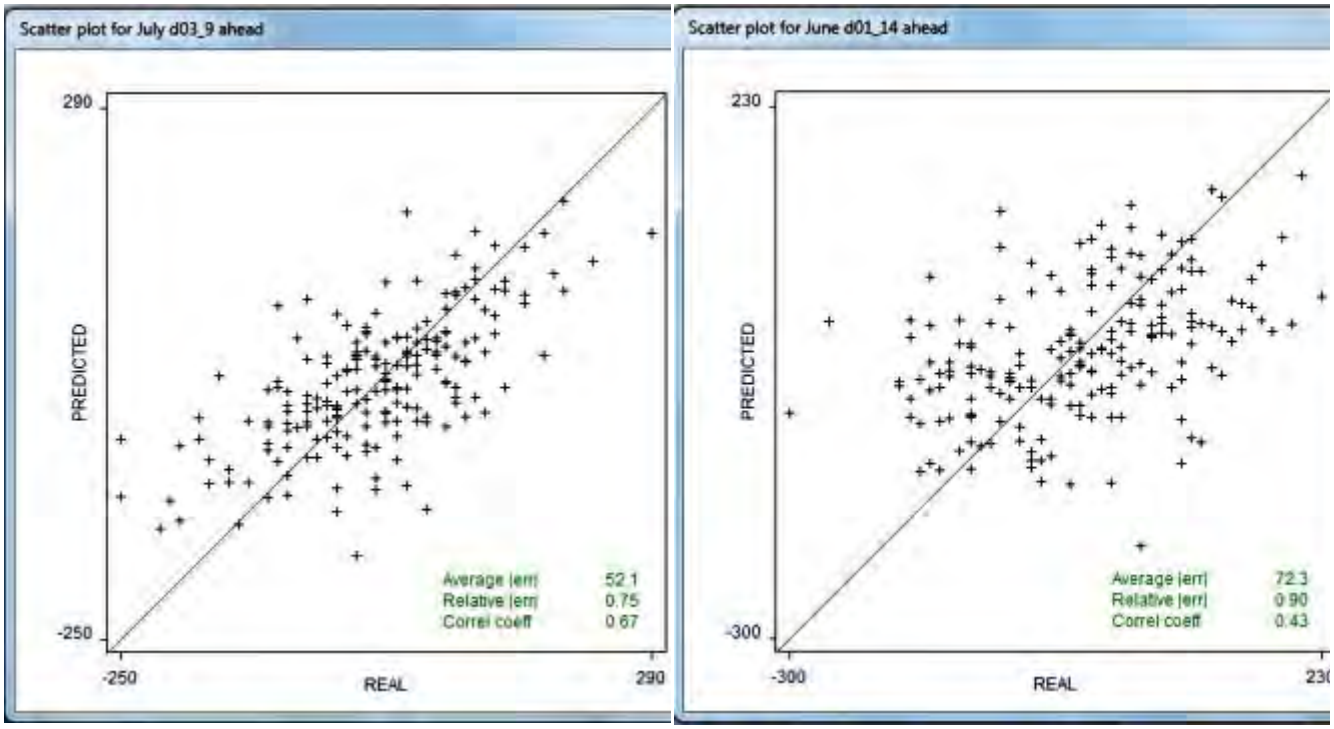
(b)


Scatter plot for August d03_6 ahead
271
PREDICTED
-280
-280
REAL
271
Average |err| 42.8
Relative |err| 0.54
Correl coeff 0.84

(c)


Scatter plot for July d02_5 ahead
308
PREDICTED
-330
-330
REAL
Average |err| 52.1
Relative |err| 0.61
Correl coeff 0.74

(d)



(e) (f)

Figure 4.3: Scatter plots of real versus predicted values

5. THE DROUGHT PREDICTION SYSTEM

5.1. Architecture of the drought prediction system

Architecture of the drought prediction system is shown in Figure 5.1 below.

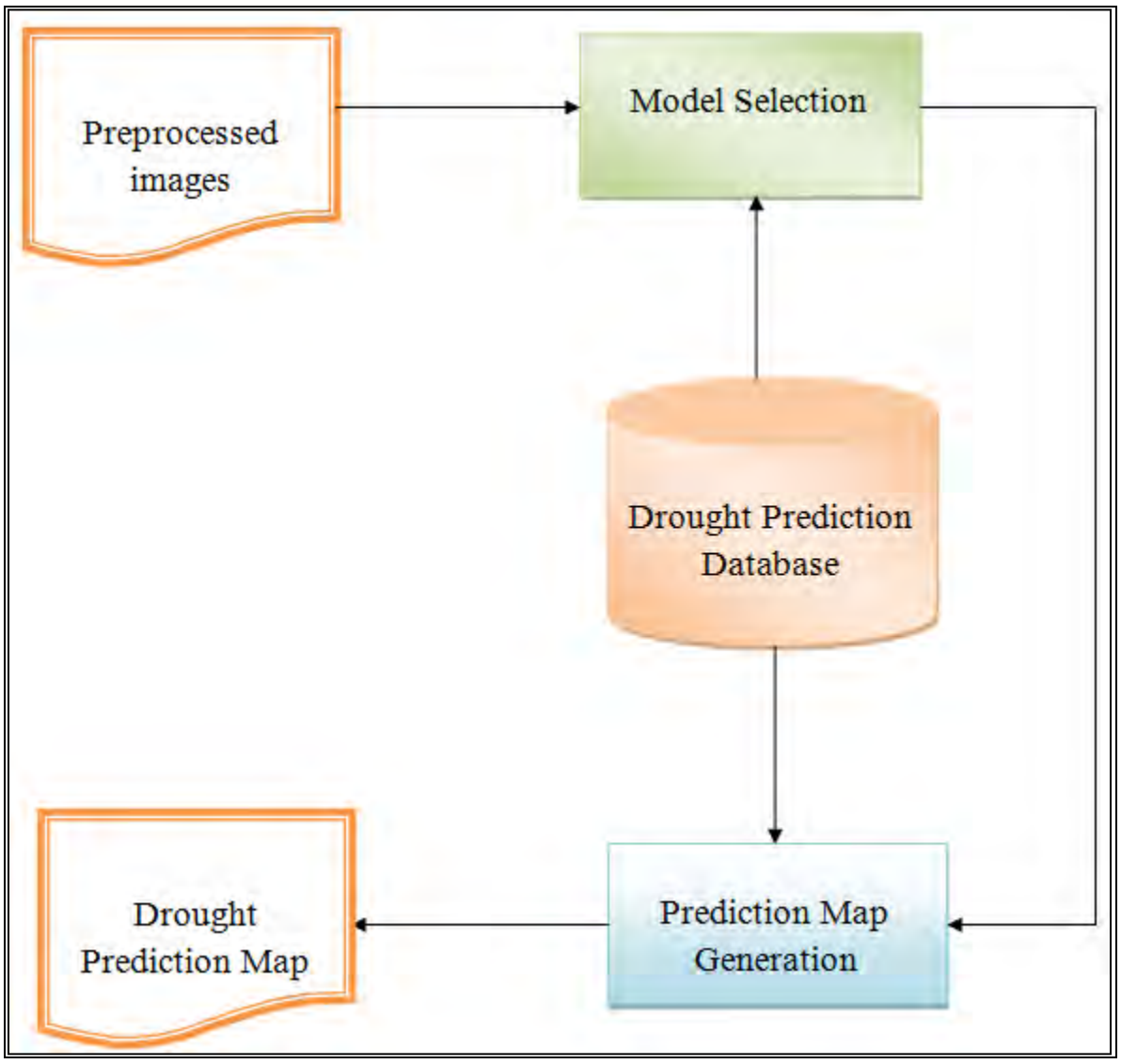


Figure 5.1: Architectural design of drought prediction system

Preprocessed Image

The user input data is a preprocessed imagery data. Preprocessed images of all the eleven attributes are made input. The raw data for the eleven attributes are found in different formats. Six of the attributes (NDVI, RFE, Veg_ethiopia, Landcover, DEM, and WHC) are images obtained from different sources and have different spatial resolutions. The rest five attribute (oceanic indices) are point data and have single values for each month in the growing season. To generate drought prediction maps all the input data should be changed to imagery data. The procedures in preprocessing the data include: First, the two most important attribute, SPI and SSG, values needs to be calculated from the raw data RFE and NDVI respectively and maps should be created for both attributes. This was done by first extracting the point data from RFE and NDVI images and calculating the SPI and SSG values. C code was written and used to calculate SPI and ILWIS software for calculating SSG. Then using IDW tool of ArcGis, maps are created for both SPI and SSG. Second, a constant raster map was created for the oceanic indices by using ArcGis, as oceanic indices have point data. Third, coordinate systems of the data were changed to the same coordinate system using ArcGis so that all the images fit geometrically. And finally, the data were made to have the same extent using ERDAS software with coordinates ulx 1361186, uly 1620208, lrx 2921186, lry 292208 where ulx, uly represent upper left x and y coordinates and lrx, lry represent lower right x and y coordinates. The preprocessing was done manually and it is a very tiresome process. If a preprocessing component is added it reduces lot of the tedious work and it should be looked on in the future. When a user inputs preprocessed image the data are stored in a drought prediction database.

Model Selection

The developed models are incorporated in the drought prediction database. After making the inputs (preprocessed images), the user then selects the time period for which drought is to be predicted. Based on user's selection appropriate model is selected automatically to be applied on the input data in generating the drought prediction maps.

The models may contain one or more rules, where each rule is a conjunction of conditions associated with a linear expression. The meaning of a rule is that, if a case satisfies all the conditions, then the linear expression is appropriate for predicting drought. If two or more rules are applied to a case, then the individual predictions from each regression equation are averaged

to arrive at the final predicted value. The models are best for making predictions as far as six dekad ahead predictions.

Prediction Map Generation

The selected model was then implemented on the input data by using MapCubist software tool to generate drought prediction maps.

The images for each dekad period were overlaid, since all of them have same spatial resolution. During implementation, values of all the eleven attributes for each pixel were considered to determine which rule(s) applied and the corresponding linear regression equations associated with the rules were applied to the input data values to calculate the drought prediction image value for each pixel in the study area.

MapCubist software is also incorporated in the Drought prediction database. After model is selected MapCubist is triggered through DOS command and drought prediction map is generated.

Drought Prediction Database

The drought prediction database holds static i.e., biophysical data. The biophysical data including Veg_ethiopia, DEM, WHC and Lancover values were considered constant which does change over the 24 years period and the user is not expected to input images for biophysical data. Dynamic data including satellite vegetation, climate, and oceanic indices data are accepted from user and integrated with the static data in the database. Then appropriate model is selected and MapCubist is triggered to implement the model on the input data.

Drought Prediction Map

The drought prediction map generated after the model implementation is presented to the user. On the upper right corner of the map range of values for SSG is displayed. The values were multiplied by 100 in the implementation so they need to be divided by 100 after implementation. The SSG values present agricultural drought severity classes (Table 3.4). These classes shows where drought occurred and its' severity. The prediction map helps the user to easily identify the areas where drought is occurring and to take action.

5.2. Development Environment

For developing the system Matlab 7.1 and MapCubist 2.2 were used. In the system, Matlab was used to allow users to input preprocessed imagery data, then storing the data in the database and finally triggering MapCubist. MapCubist then implements one of the models on the input data for generating drought prediction map. After the drought prediction map is generated it can be displayed on ERDAS or ArcGis.

5.3. Models incorporated in the system

The developed models are the most essential part of the drought prediction system. 69 of the 105 models having correlation coefficient >= 0.71 on the test data are incorporated in the system database. The user is required to input time period of the input data and the prediction period for which drought prediction map is generated. Then based on user selection appropriate model is selected by the system to be implemented on the input data.

5.4. Model Implementation

MapCubist software developed at the USGS Center for EROS was used to implement the models i.e., prediction maps generation.

As indicated in the data preprocessing and transformation section raster images were generated for use in the implementation of the model. SSG and SPI images were produced for each dekadal period while for oceanic indices monthly constant raster map was created (one oceanic index image being used with three dekadal images in a month) and a static raster image was created for biophysical attributes since the values doesn't change(the same static image is used with every dekadal image). The images for each dekad period were then overlaid, since all of them have spatial resolution of 8 km.

The rules from the built model were then applied to the images input data for each dekadal period. During implementation of our model to the images data, the values of all the input attributes for each pixel were considered to determine which rule(s) applied and the corresponding linear regression equation(s) associated with the rule(s) was (were) applied to

input data values to calculate the drought prediction image value for each pixel across the study area.

MapCubist software was executed from DOS command window. It was executed using the following command

"MapCubist_v202 ModelName HFA PredictionMap Mask (optional)"

Where MapCubist_v202 is version of MapCubist, ModelName is name of the regression tree model, HFA is an ERDAS image (.img) format for the output, PredictionMap is the output prediction image, and Mask is the mask file. The use of MapCubist is shown in Figure 5.2.

The prediction maps shows whether there is agricultural drought or not. For example using June dekad one image we can predict June dekad two, June dekad three, July dekad one … and October dekad three.

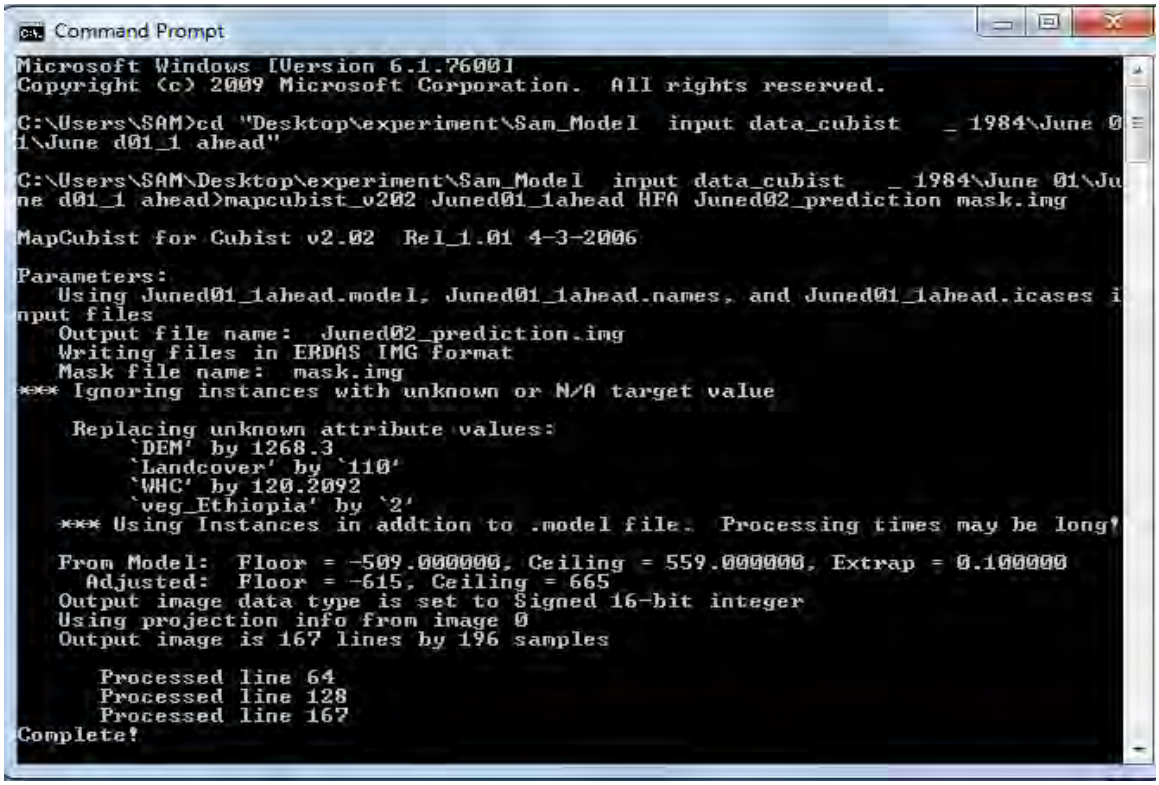


Figure 5.2: Model implementation using MapCubist

During the implementation, a mask file was created for the spatial extent of Ethiopia in such a way that all areas inside Ethiopian boundary have a pixel value 1 and areas outside Ethiopian boundary have a value 0. This was done to reduce the processing time. A mask pixel value of 0 indicates that pixel should not be processed and a mask value of 1 indicates that the pixel should be processed.

5.5. User interface of the system

The system only accepts preprocessed images as input. Here the dynamic data including SSG, SPI, and the oceanic indices (PDO, PNA, MEI, NAO, and AMO) are accepted from the user. As explained earlier biophysical data is static and it is already stored in the database. The user then selects dekadal period of the input data and the prediction period. After all the inputs are acquired, the system allows the user to run. When the user clicks on the Run button all the acquired data (dynamic data) is integrated with the static data and then MapCubist software inside the database is triggered. The Mapubist software implements the selected model on the input images to generate drought prediction map. Figure 5.3 below shows the user interface for drought prediction system.

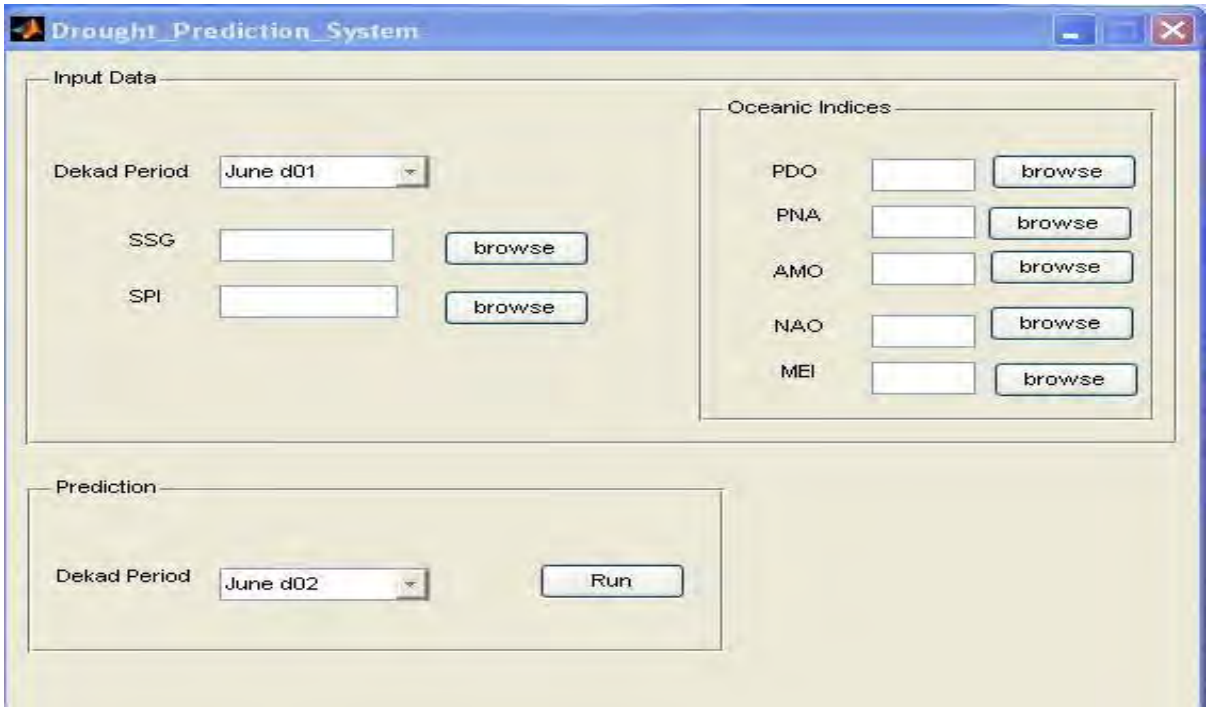


Figure 5.3: User interface of drought prediction system

5.6. Sample output

Below Figure 5.4 shows sample output produced after the drought prediction map is generated by MapCubist.

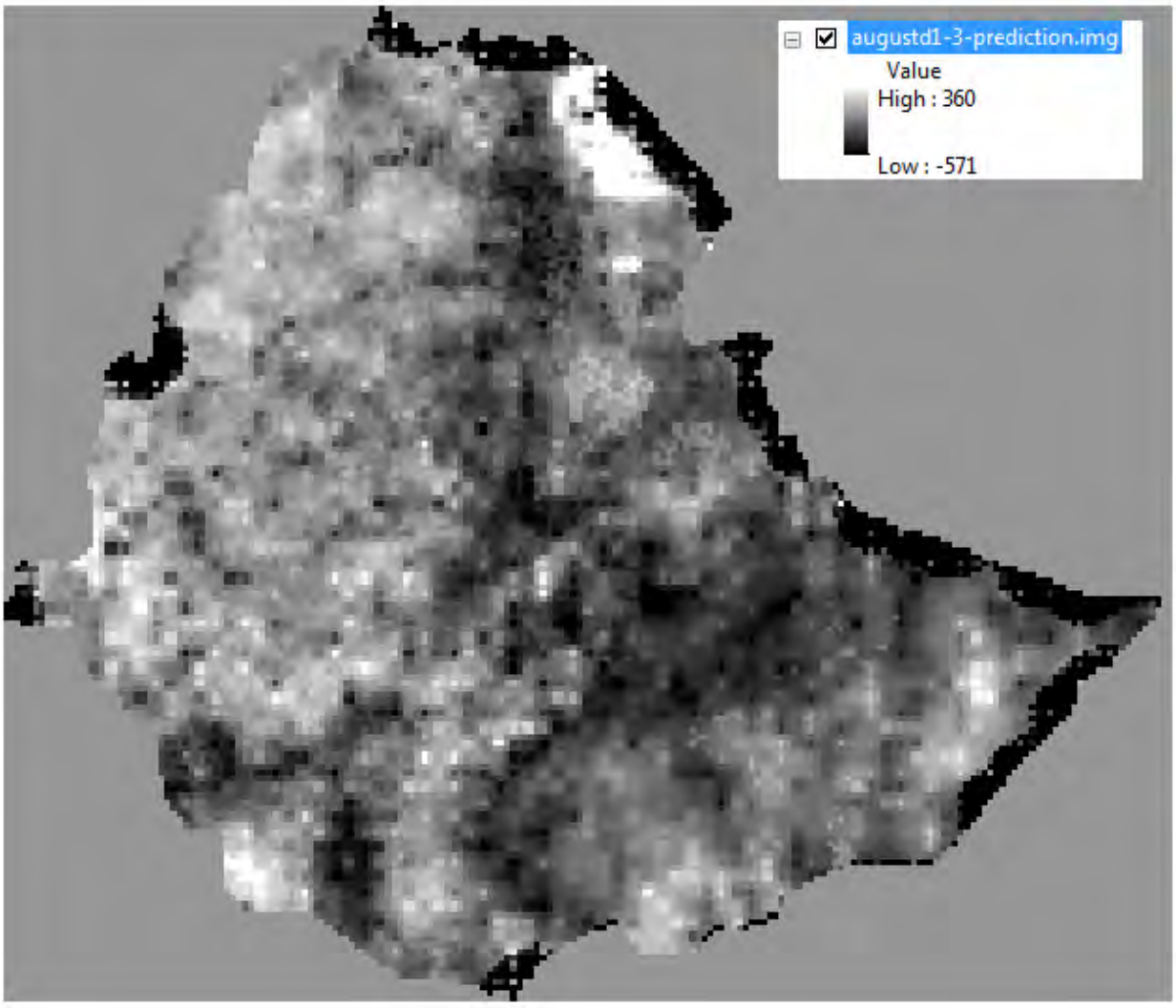


Figure 5.4: Sample drought prediction map

The sample output map shows a drought prediction generated by the selected model on the input data. The range values are divided by 100 since the input values were first multiplied by 100 to remove decimal points since the MapCubist software which implements the model only accepts integer values. Drought severity classes for SSG values are presented in Table 3.4.

Figure 5.4 can be presented in a better way as in Figure 5.5 by adding color effects using ArcGis properties.

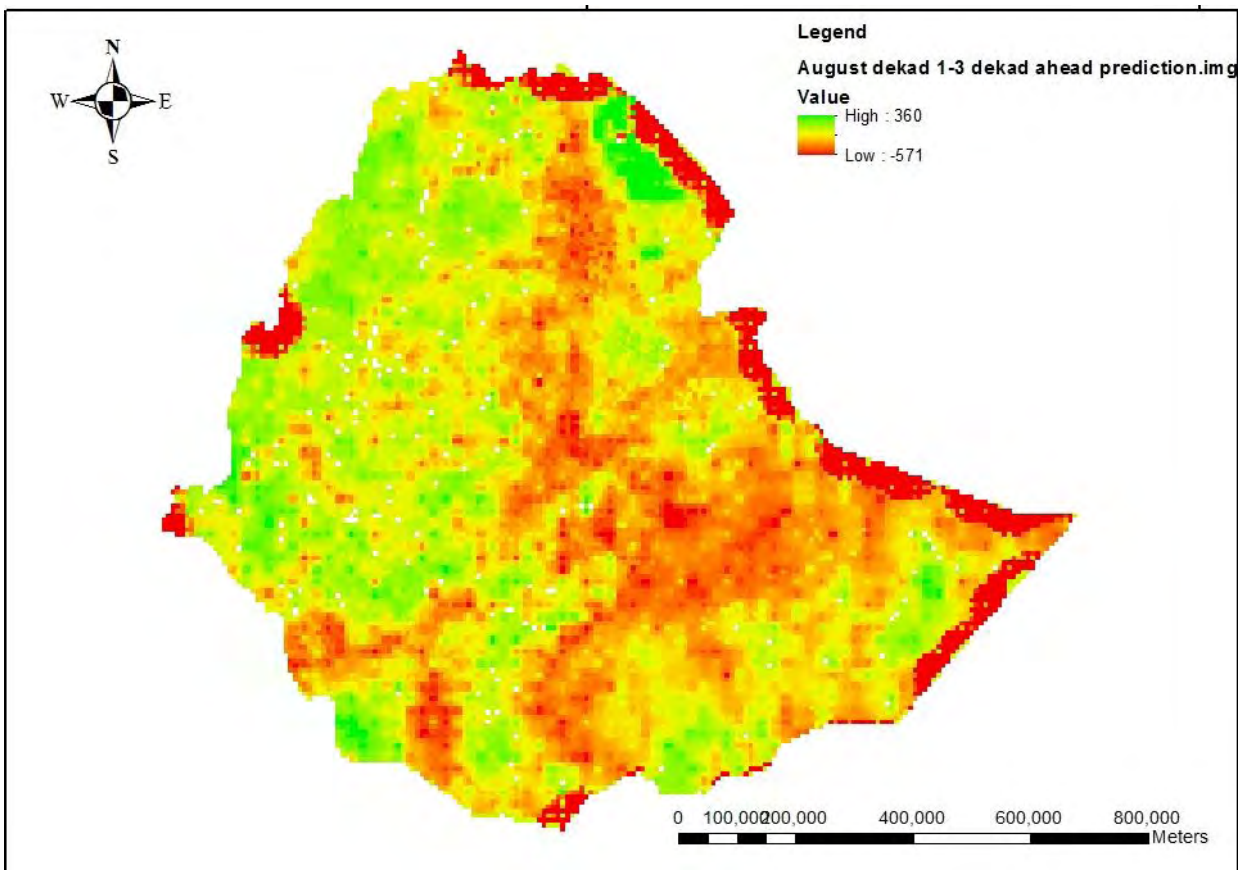


Figure 5.5: Sample drought prediction map with color texture

6. DISCUSSION OF RESULTS

The model was implemented for the years 1984 and 2002 where severe drought occurred in Ethiopia.

The 1984 drought known as the Great Famine was the worst famine in recent Ethiopian history. At the time five Ethiopian provinces including Gojjam, Eritrea, Hararghe, Tigray, and Wollo all received record low rainfalls. The severe famine in 1983-85 centered on Tigray and northern Wollo, which further affected Eritrea, Gondor and northern Shewa [2]. In northern Tigray alone 400,000 has died due to the drought. The UN estimated that around 1,000,000 died from the whole country. And in the 2002 drought one of the worst-affected areas was the Somali region, centering on the shinille zone.

The developed models are implemented using data of the years 1984 and 2002. After implementation of the models a series of dekadal period prediction maps were generated. The prediction maps show where drought has occurred and its severity. As indicated earlier our focus is on the growing season (June to October), so we can create the prediction maps by implementing 69 of the 105 models in Table 4.1 since predictions far from the first six prediction periods have very low *r* values as discussed in the model validation section.

It is impossible to show the results for all the 69 prediction models therefore for demonstrate the results of September prediction maps using August data is presented. All the three August dekad periods data and models were used to predict the three September dekads. To predict September dekad one we have used - August dekad 1 - 3 dekad ahead, August dekad 2 - 2 dekad ahead, and August dekad 3 - 1 dekad ahead prediction models. For predicting September dekad two, we have used - August dekad 1 - 4 dekad ahead, August dekad 2 - 3 dekad ahead, and August dekad 3 - 2 dekad ahead prediction models. And for predicting September dekad three, we have used - August dekad 1- 5 dekad ahead, August dekad 2 - 4 dekad ahead, and August dekad 3 - 3 dekad ahead prediction models. The reason for selecting August models to make the predictions is that all the three dekad models has higher *r* values i.e., August dekad 1 - 3 dekad ahead prediction (0.93), August dekad 2 - 2 dekad ahead prediction (0.97), August dekad 3 - 1 dekad ahead prediction (0.99), August dekad 1 - 4 dekad ahead prediction (0.87), August dekad 2 - 3 dekad

ahead prediction (0.90), August dekad 3 - 2 dekad ahead prediction (0.93), August dekad 1- 5 dekad ahead prediction (0.87), August dekad 2 - 4 dekad ahead prediction (0.90), and August dekad 3 - 3 dekad ahead prediction (0.93).

The prediction maps for the three September dekad periods of the year 1984 and 2002 are presented below in Figure 6.1 and Figure 6.2 respectively.

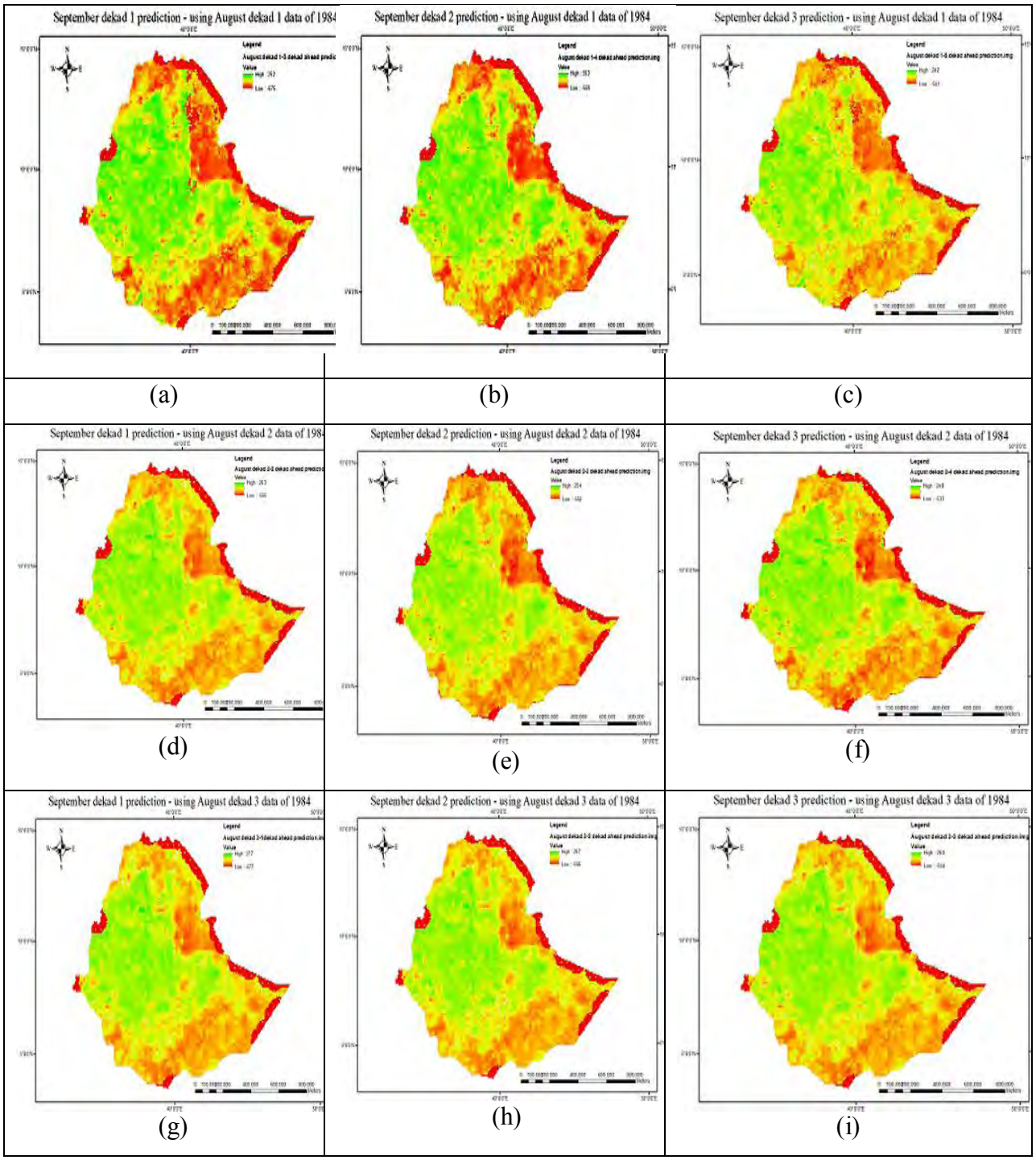


Figure 6.1: September Prediction Maps using the three August dekads data and models for 1984

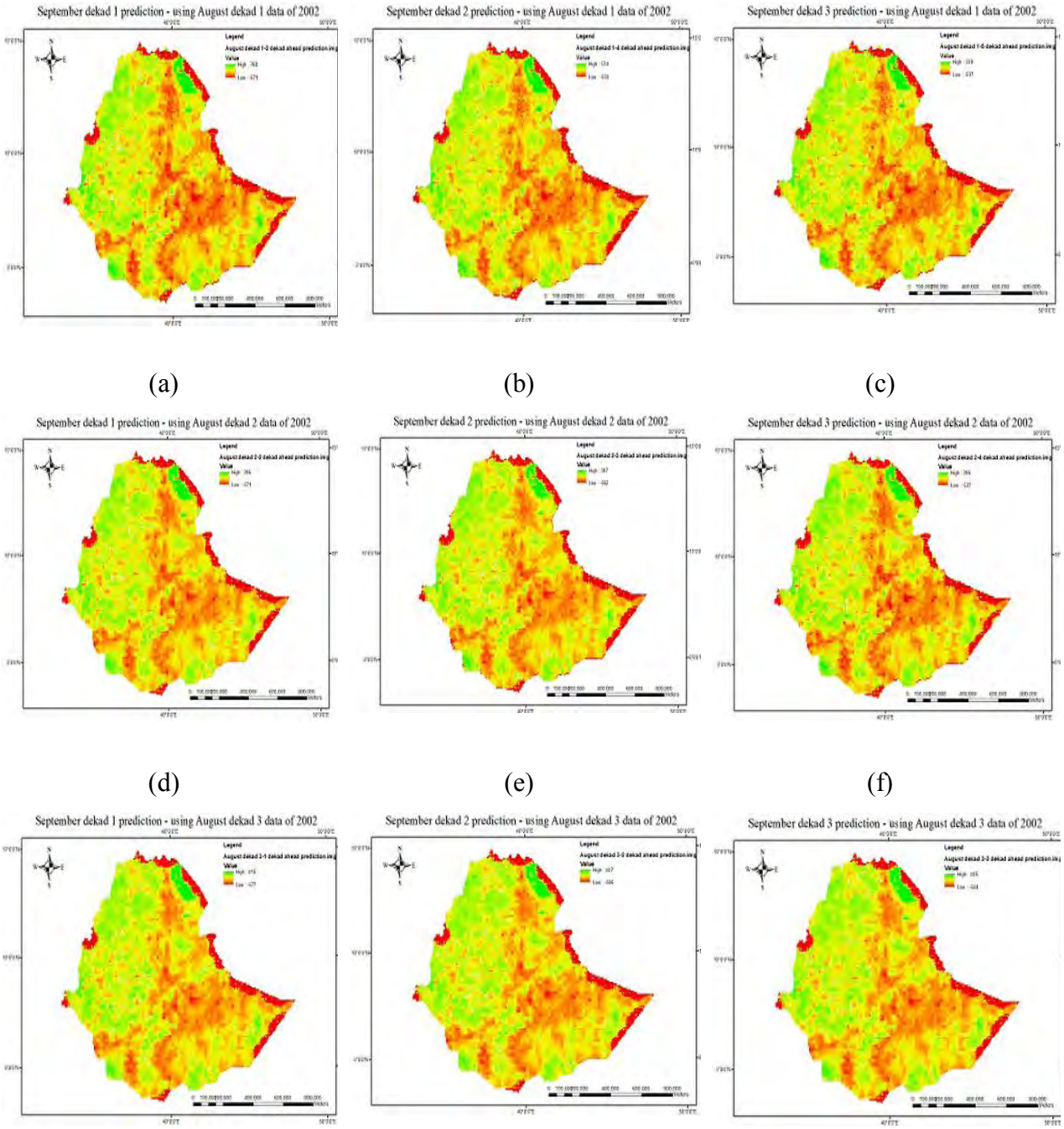


Figure 6.2: September Prediction Maps using the three August dekads data and models for 2002

In both Figure 6.1 and 6.2 the images in the first raw represent prediction maps of September dekad one to three created using August dekad one prediction models (August dekad 1 - 3 dekad ahead, August dekad 1 - 4 dekad ahead and August dekad 1- 5 dekad ahead prediction) whereas second raw images represent images created using August dekad two prediction models (August

dekad 2 - 2 dekad ahead, August dekad 2 - 3 dekad ahead, and August dekad 2 - 4 dekad ahead prediction) and the third raw images represent images created using August dekad three prediction models (August dekad 3 - 1 dekad ahead, August dekad 3 - 2 dekad ahead and August dekad 3 - 3 dekad ahead prediction).

The red color in the maps shows areas where drought has occurred and the green ones shows moist areas.

In the maps in Figure 6.1 for the three September dekad predictions, using August data of 1984, the prediction maps have values ranging from -8.99 to 2.77. The values show drought classes ranging from extreme drought to extreme moist.

The three prediction maps of September dekad one has value ranges of (-8.75 to 2.52) for August dekad 1 - 3 dekad ahead, (-8.65 to 2.63) for August dekad 2 - 2 dekad ahead, and (-8.77 to 2.77) for August dekad 3 - 1 dekad ahead predictions. In all the three prediction maps an extreme drought is seen for almost all parts of north, east, south and south east parts of Ethiopia. For August dekad 1 - 3 dekad ahead and August dekad 3 - 1 dekad ahead predictions some parts of the west are in moderate drought but for August dekad 2 - 2 dekad ahead prediction the west is in relative moist condition. In August dekad 1 - 3 dekad ahead moderate drought is seen around the center.

For September dekad two prediction the three prediction maps has value ranges as follows, for August dekad 1 - 4 dekad ahead prediction (-8.99 to 2.53), for August dekad 2 - 3 dekad ahead prediction (-8.52 to 2.54), and for August dekad 3 - 2 dekad ahead prediction (-8.66 to 2.67). August dekad 1 - 4 dekad ahead and August dekad 3 - 2 dekad ahead prediction maps has almost the same maps as September dekad one predictions (August dekad 1 - 3 dekad ahead and August dekad 3 - 1 prediction maps respectively) except in August dekad 1 - 4 dekad ahead prediction the moderate drought around the center has changed to near normal condition. In August dekad 2 - 3 dekad ahead prediction an increase in the level of drought is seen for parts of Ethiopia where extreme drought has occurred in September prediction maps.

In September dekad three prediction, the three prediction maps has value ranges of (-8.41 to 2.42) for August dekad 1 - 5 dekad ahead , (-8.33 to 2.48) for August dekad 2 - 4 dekad ahead, and (-8.44 to 2.64) for August dekad 3 - 3 dekad ahead predictions. The drought level has slightly increased in all three prediction maps for different parts of Ethiopia i.e., in August dekad

1 - 5 dekad ahead prediction map on north and east parts, for August dekad 2 - 4 dekad ahead prediction on south and east, and for August dekad 3 - 3 dekad ahead prediction on south east part. In August dekad 1 - 5 dekad ahead, some part of the south and around the center, level of drought gets to moderate drought. While for August dekad 2 - 4 dekad ahead prediction moderate drought areas of the west get to near normal. And in August dekad 3 - 3 dekad ahead prediction apart from south east the rest remained the same as September dekad 1 and 2 prediction maps (August dekad 3 -1 dekad ahead and August Dekad 3 – 2 dekad ahead prediction maps).

Generally in all the three dekad predictions of September the maps show that there was an extreme drought in the north, east, south and south eastern part of the country for the year 1984. And the west was in normal condition and there was moderate drought on minor part of the center. The real situation at the time also tells as that there was an extreme drought in the north (Tigray and Wollo, Gonder), and south east (hararghe). North shewa was also affected by the drought and the model prediction conformed by indicating that there was drought around the center.

On the edge of the maps darker red color is seen. This is caused when predictions are made for missing values. This might not be the only cause as only few points on those positions have missing values and there needs to be a further investigation to solve the problem. The extreme moist values that were seen in the prediction maps are outlier values since there was no part of the country with extreme moist condition.

The three September dekad prediction maps, using August 2002 data, in Figure 6.2 has values ranging from -8.77 to 8.34. The values show drought classes ranging from extreme drought to extreme moist.

The three prediction maps of September dekad one has value ranges of (-8.71 to 3.60) for August dekad 1 - 3 dekad ahead, (-8.71 to 3.96) for August dekad 2 - 2 dekad ahead, and (-8.77 to 4.16) for August dekad 3 - 1 dekad ahead predictions.

For September dekad two prediction the three prediction maps has value ranges as follows, (-8.59 to 8.34) for August dekad 1 - 4 dekad ahead prediction, (-8.62 to 3.87) for August dekad 2 - 3 dekad ahead prediction, and (-8.66 to 4.07) for August dekad 3 - 2 dekad ahead prediction.

And in September dekad three prediction, the three prediction maps has value ranges of (-8.37 to 3.38) for August dekad 1 - 5 dekad ahead, (-8.37 to 3.96) for August dekad 2 - 4 dekad ahead, and (-8.44 to 4.05) for August dekad 3 - 3 dekad ahead predictions.

All the prediction maps from all August dekad predictions have almost the same pattern. The pattern shows an extreme drought on east and most of south east part of Ethiopia. The drought is also seen on some parts of northern Ethiopia. Most of the west part remained in moist condition. The real situation 2002 shows that there was an extreme drought in the south east (Somali region) which is also predicted with our models.

Also for 2002, on the edge of the maps darker red color is seen. And again recommendation follows as in 1984. And the extreme moist values in the prediction maps are outlier values.

Linear plots, for both 1984 and 2002, showing the correlations between actual and predicted September SSG values are presented below. The SSG values are indicators which tell whether there exists agricultural drought or not.

Here in the Figures 6.3, 6.4, and 6.5 below, September prediction values using August dekad 1 predictions are presented with the actual values in linear plots. In all comparisons for dekadal predictions only the first 100 of each actual and predicted value are displayed on the linear plots.

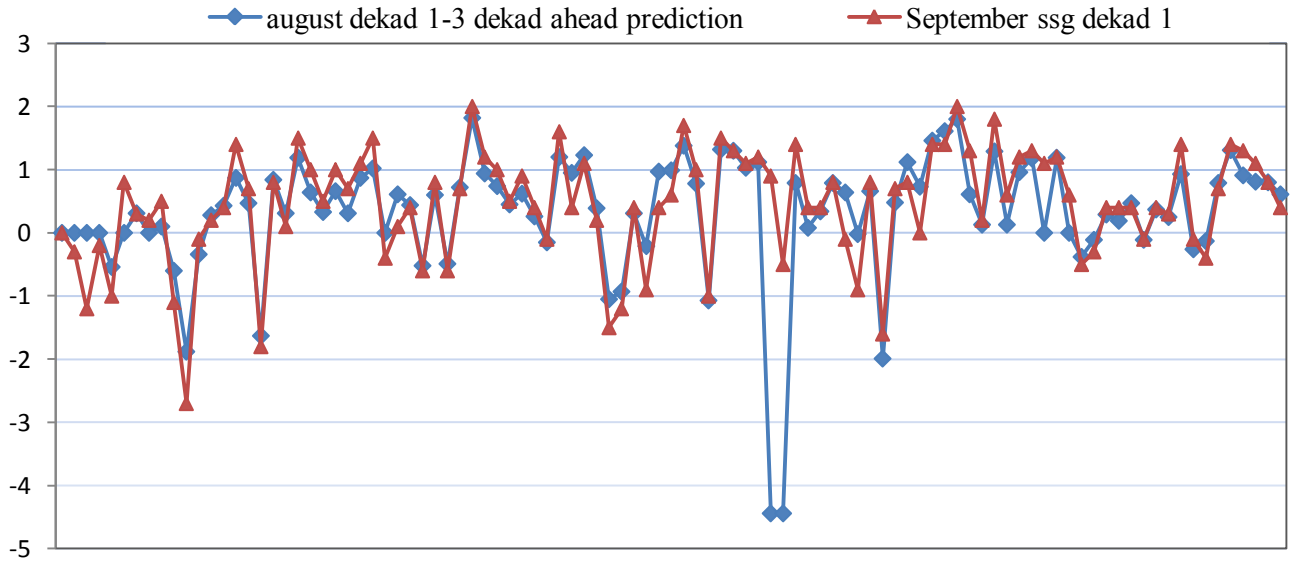


Figure 6.3: September dekad 1 prediction using August dekad 1 - 3 dekad ahead prediction

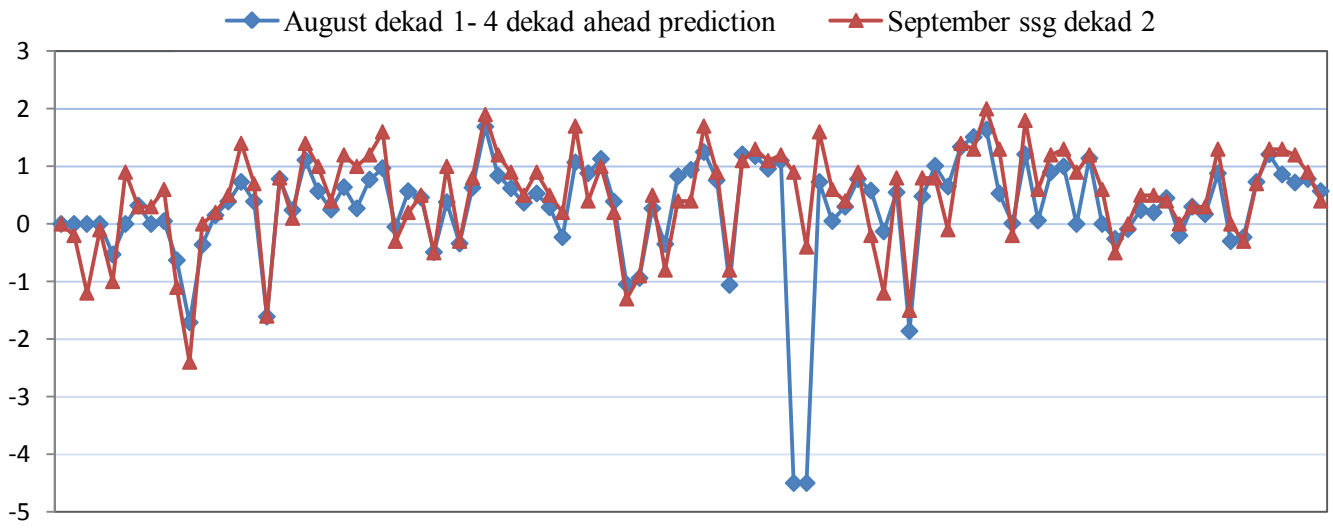


Figure 6.4: September dekad 2 prediction using August dekad 1 - 4 dekad ahead prediction

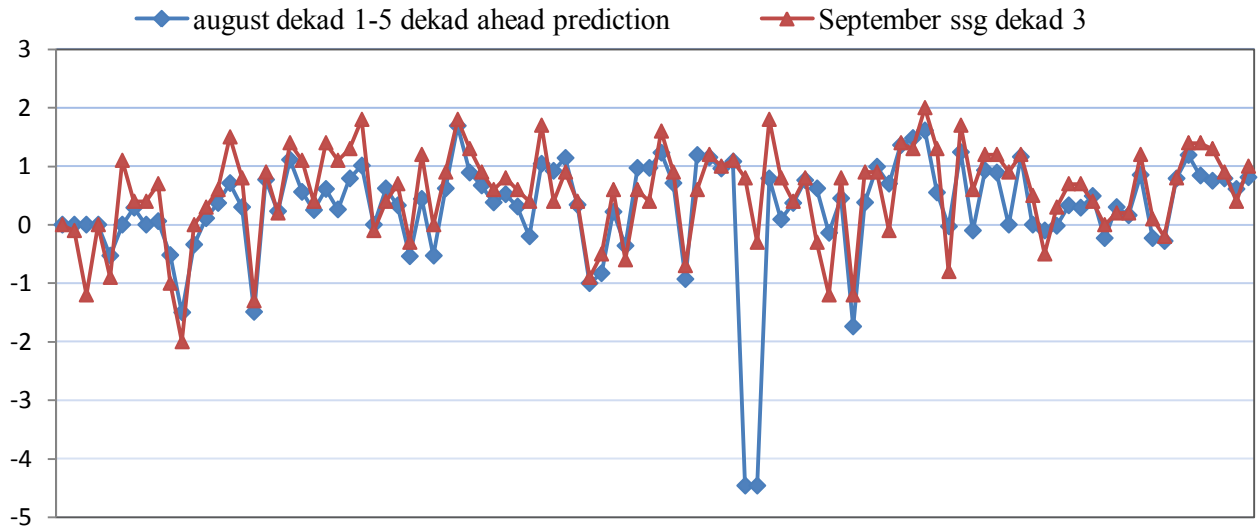


Figure 6.5: September dekad 3 prediction using August dekad 1 - 5 dekad ahead prediction

All the three prediction values for August dekad 1 highly fit with the actual SSG values, August dekad 1 - 3dekad ahead prediction having the highest accuracy. August dekad 1 - 5 dekad ahead prediction values deviate more from the actual values compared to the rest two predictions, this happens because it is relatively far from the predicted period. At two points, on the linear plots, for all the three predictions outlier values has occurred and this is due to undefined values initially in the input data.

In the Figures 6.6, 6.7, and 6.8 next, September prediction values using dekadal predictions for August dekad 2 predictions are presented.

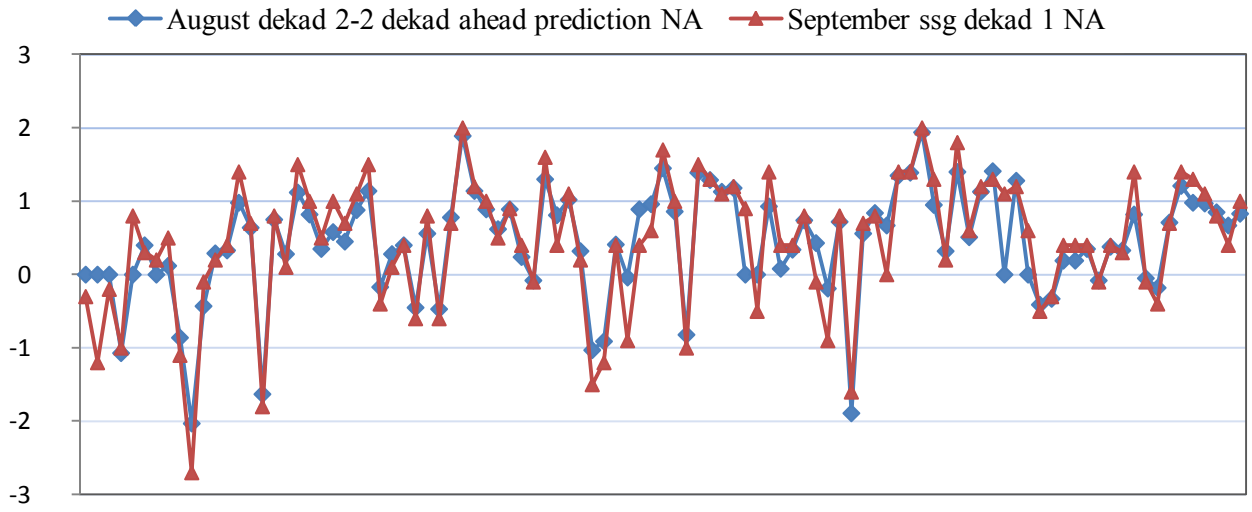


Figure 6.6: September dekad 1 prediction using August dekad 2 - 2 dekad ahead prediction

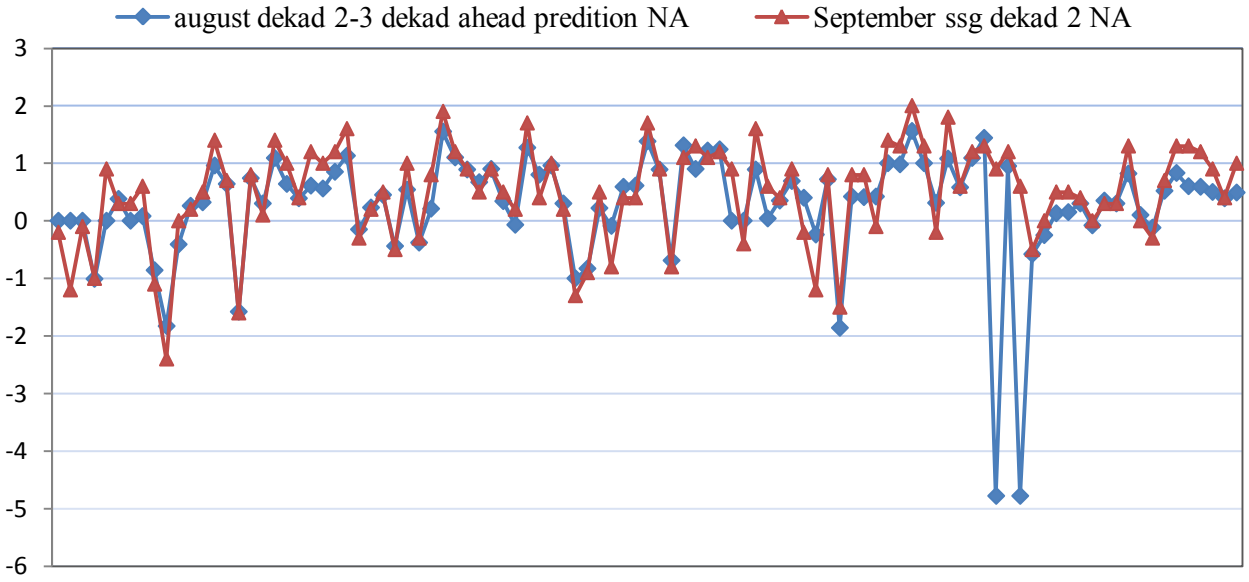


Figure 6.7: September dekad 2 prediction using August dekad 2 - 3 dekad ahead prediction

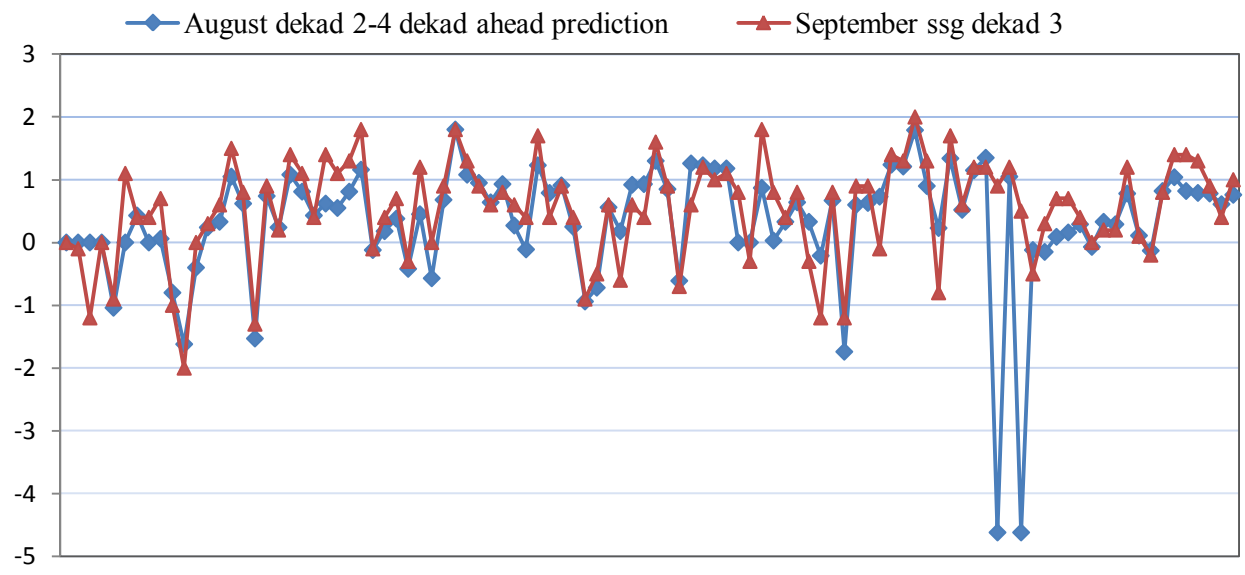


Figure 6.8: September dekad 3 prediction using August dekad 2 - 4 dekad ahead prediction

As in August dekad 1 predictions dekad 2 prediction values highly fit with the actual SSG values pattern. August dekad 2 - 2 dekad ahead prediction has the highest accuracy. August dekad 2 - 3 dekad ahead and August dekad 2 - 2 dekad ahead predictions has highly deviated outlier values.

And also in the following Figures 6.9, 6.10, and 6.11 September prediction values using dekadal predictions for August dekad 3 predictions are presented.

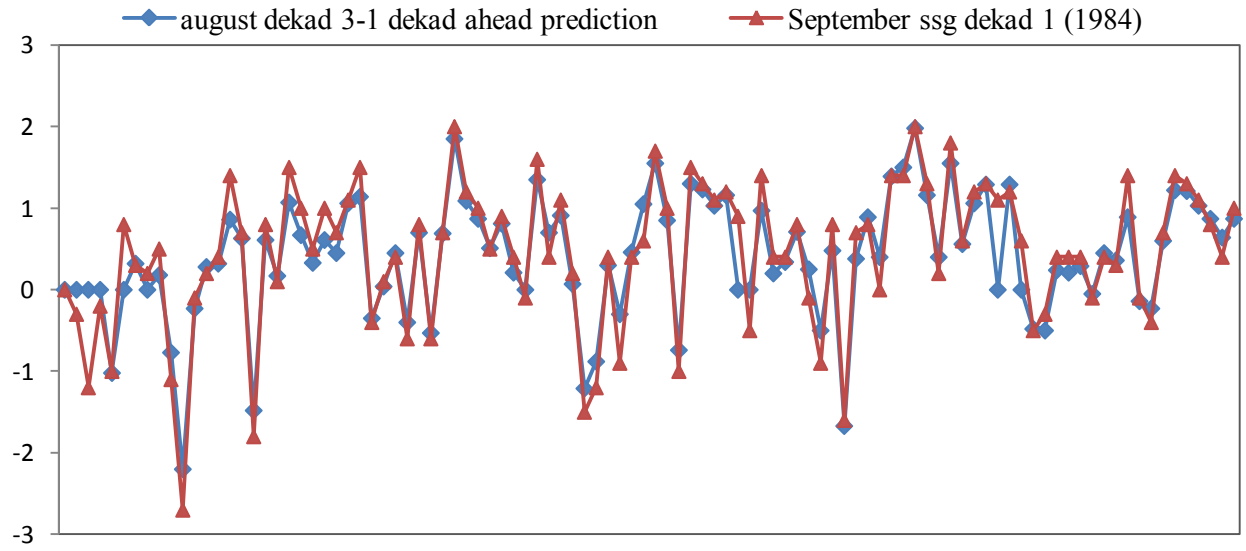


Figure 6.9: September dekad 1 prediction using August dekad 3 - 1 dekad ahead prediction

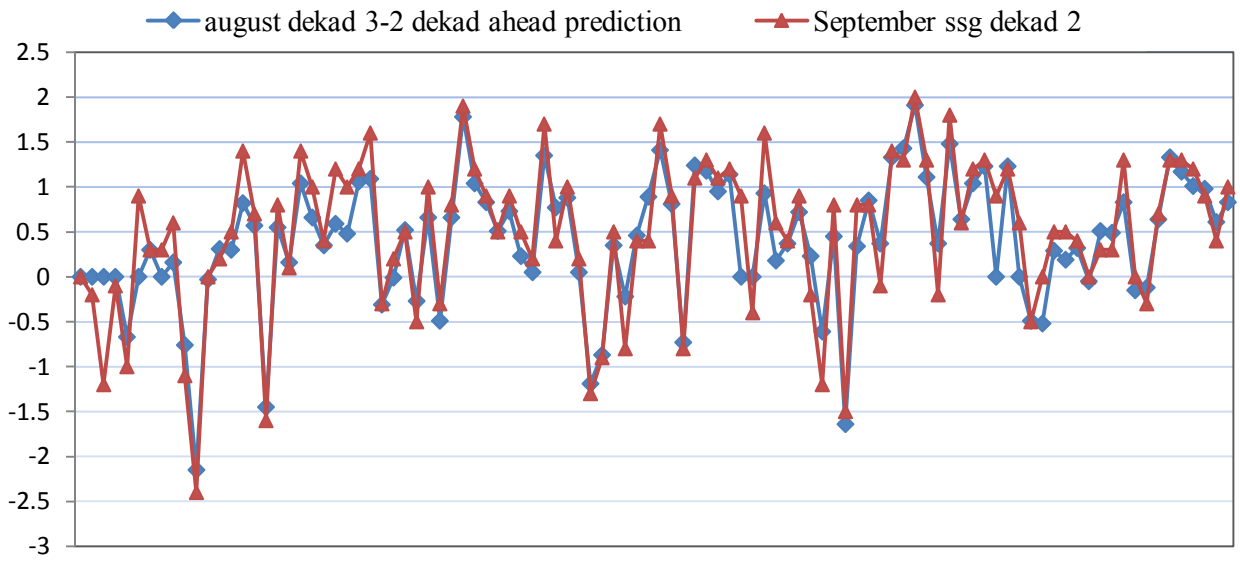


Figure 6.10: September dekad 2 prediction using August dekad 3 - 2 dekad ahead prediction model

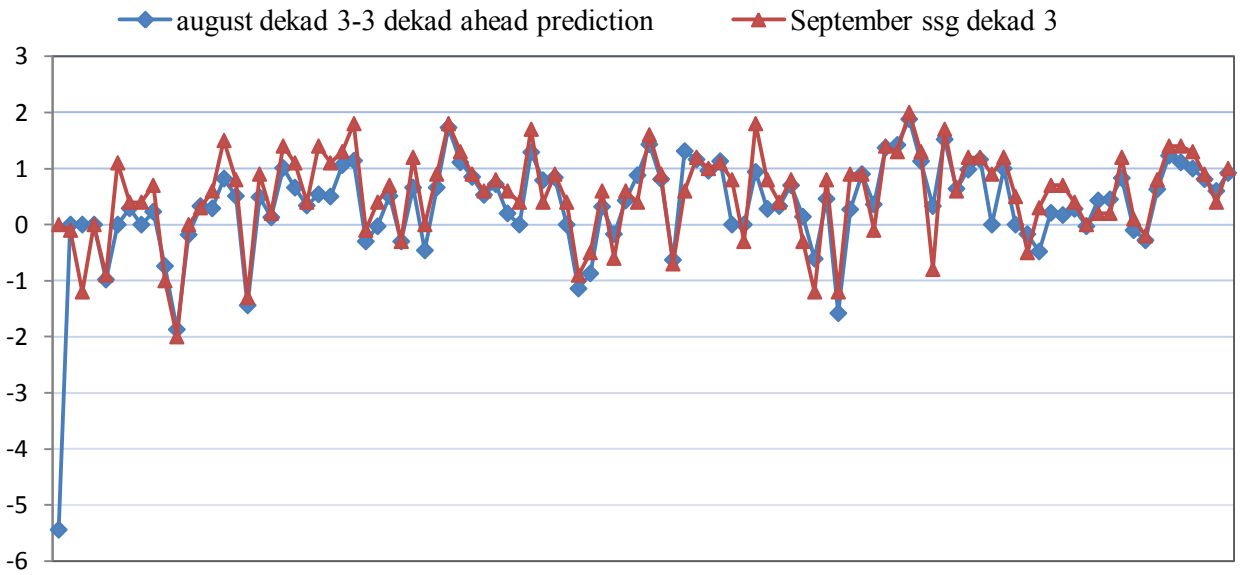


Figure 6.11: September dekad 3 prediction using August dekad 3 - 3 dekad ahead prediction model

This dekad predictions has better prediction accuracy compared to August dekad 1 and 2 predictions. An outlier value is seen only in August dekad 3 - 3 dekad ahead prediction. August dekad 3 - 1 dekad ahead prediction has the highest prediction accuracy.

September prediction values using August dekad 1 predictions for 2002 are presented with the actual values in linear plots below in the Figures 6.12, 6.13, and 6.14.

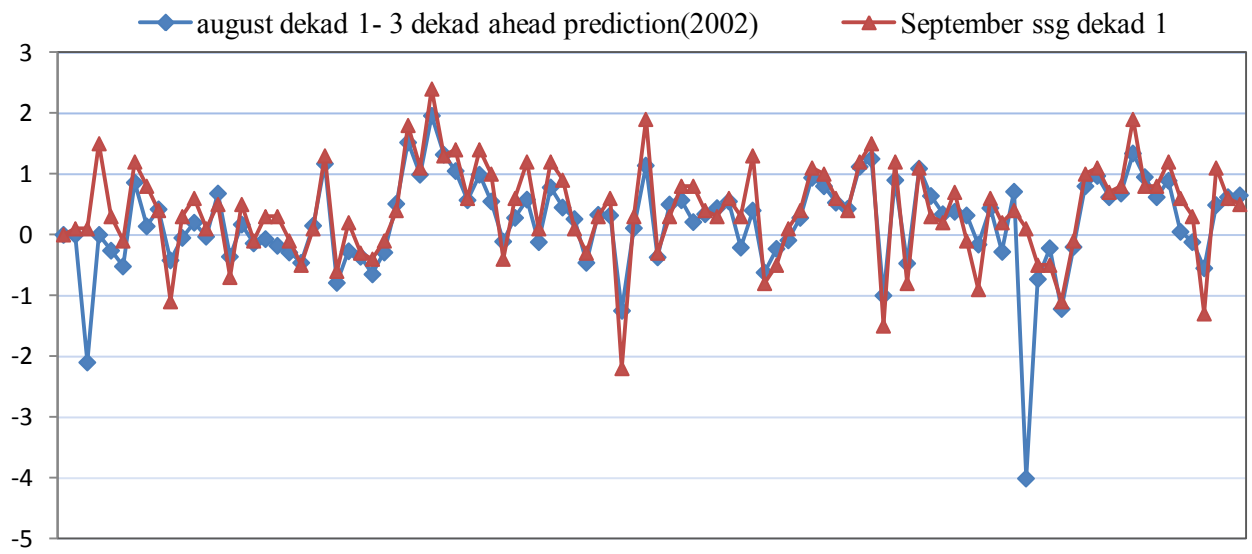


Figure 6.12: September dekad 1 prediction using August dekad 1 - 3dekad ahead prediction model for 2002

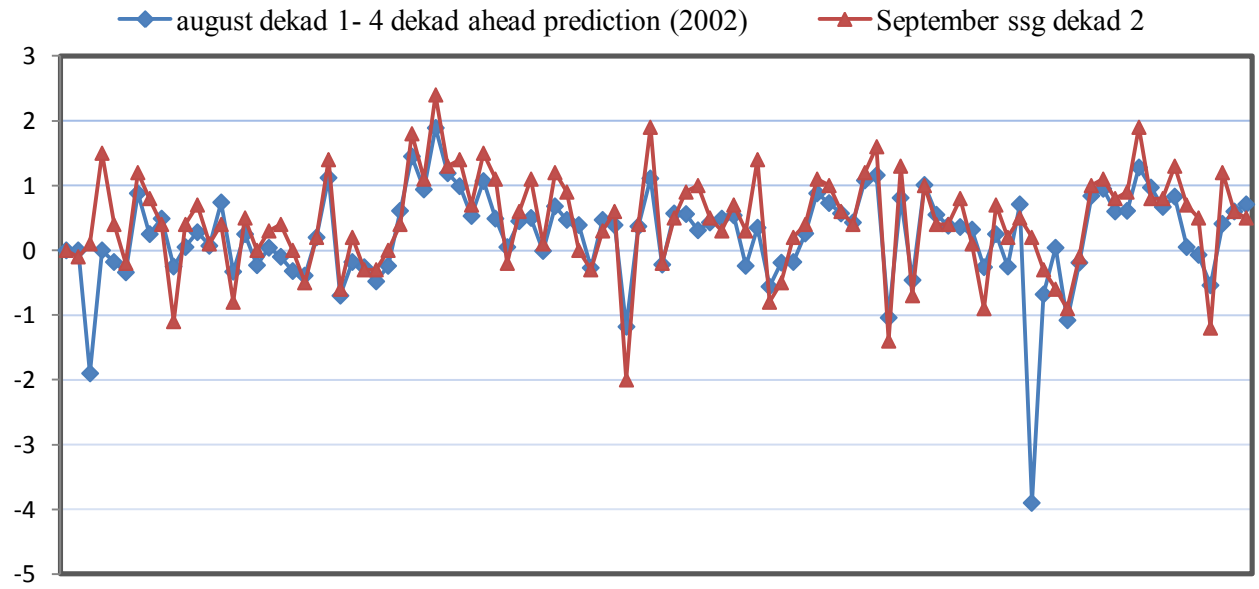


Figure 6.13: September dekad 2 prediction using August dekad 1 - 4 dekad ahead prediction model for 2002

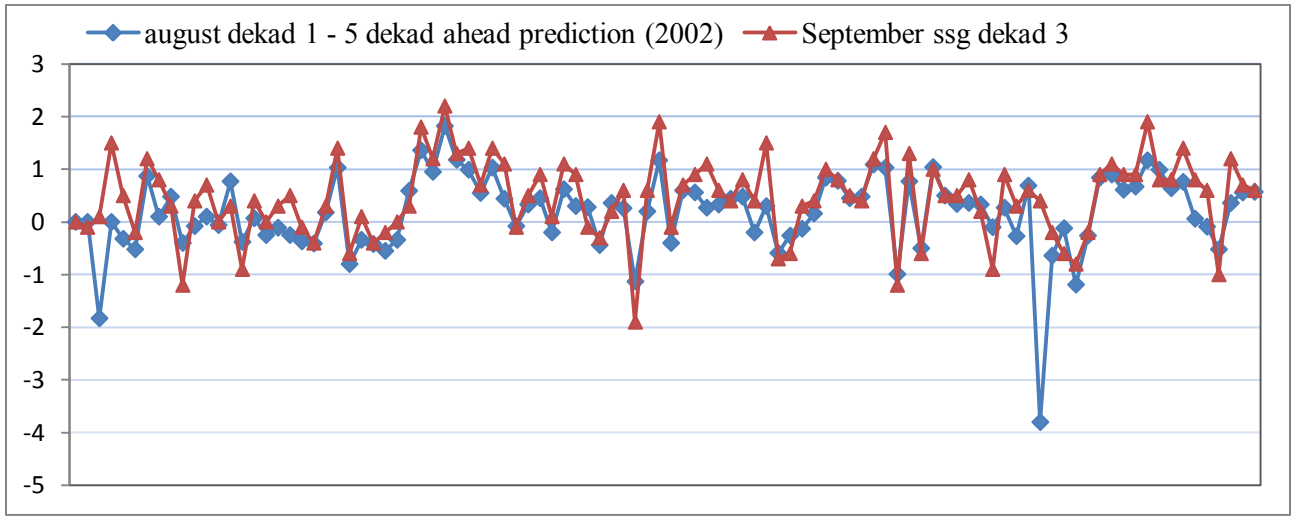


Figure 6.14: September dekad 3 prediction using August dekad 1 - 5 dekad ahead prediction model for 2002

All the three August dekad 1 prediction values in the linear plots above exceptionally fit with the actual values except for the outlier layer values they have. Even though the three predictions has different r values the prediction accuracy for all looks similar.

And here September prediction values using dekadal predictions for August dekad 2 predictions are presented in Figures 6.15, 6.16, and 6.17.

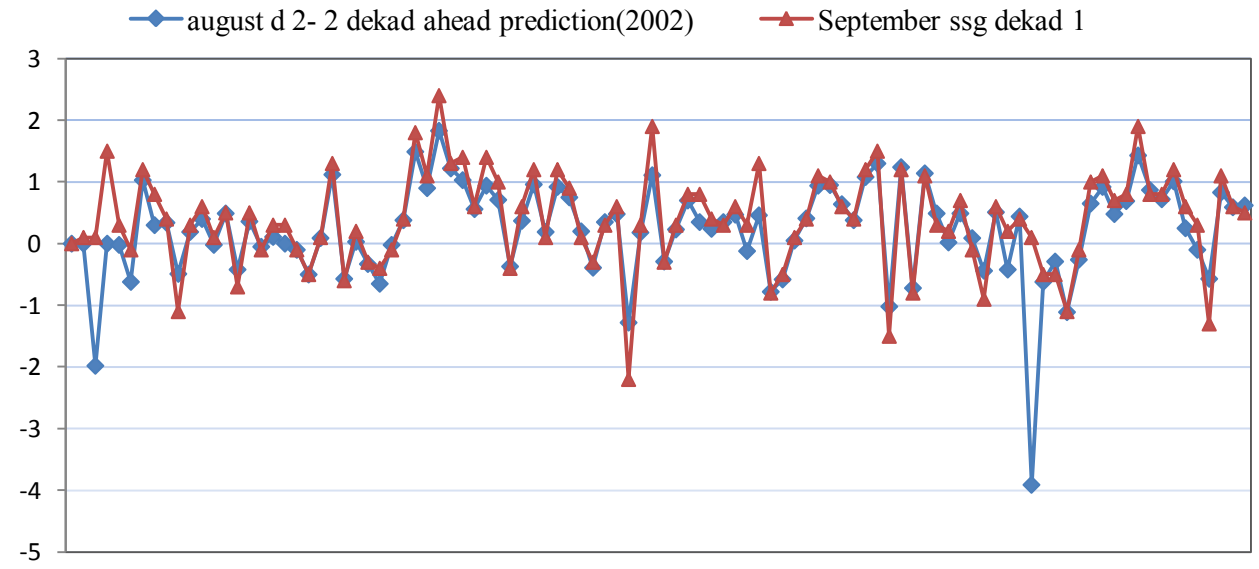


Figure 6.15: September dekad 1 prediction using August dekad 2 - 2 dekad ahead prediction model for 2002

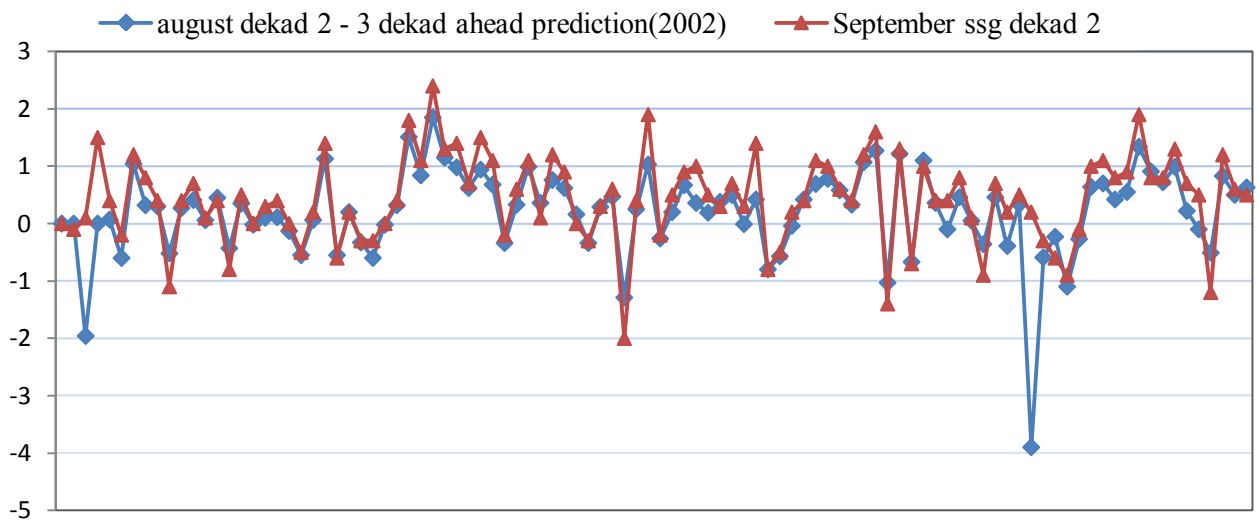


Figure 6.16: September dekad 2 prediction using August dekad 2 - 3 dekad ahead prediction model for 2002

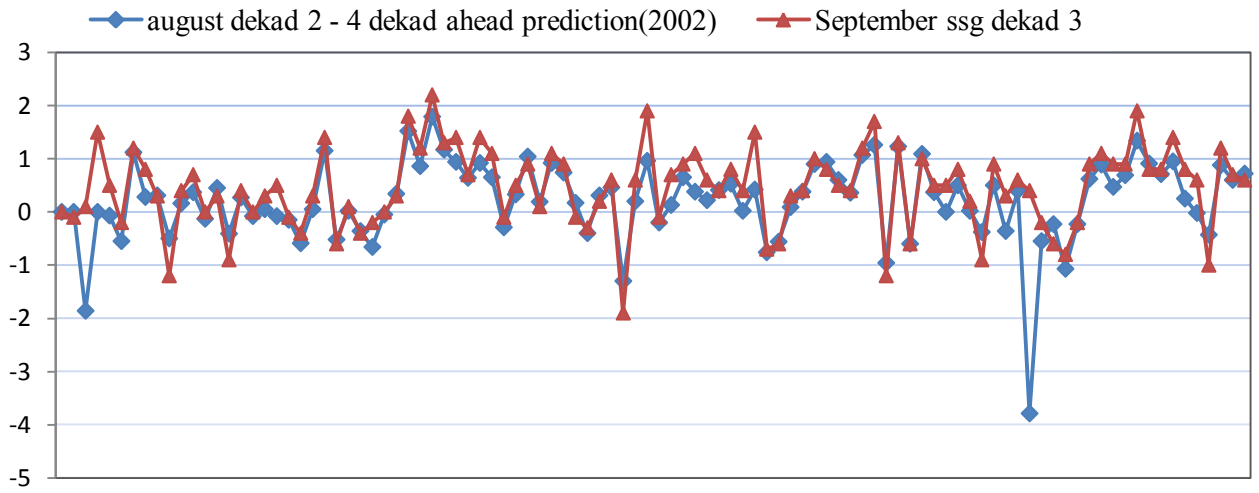


Figure 6.17: September dekad 3 prediction using August dekad 2 - 4 dekad ahead prediction model for 2002

August dekad 2 prediction values fit to the actual values even better than August dekad 1 predictions did. Same as in August dekad 1 the three predictions have different *r* values but the prediction accuracy for all models looks similar. And there exists few outlier values for the predictions.

Then September prediction values using dekadal predictions for August dekad 3 predictions are presented in Figures 6.18, 6.19, and 6.20.

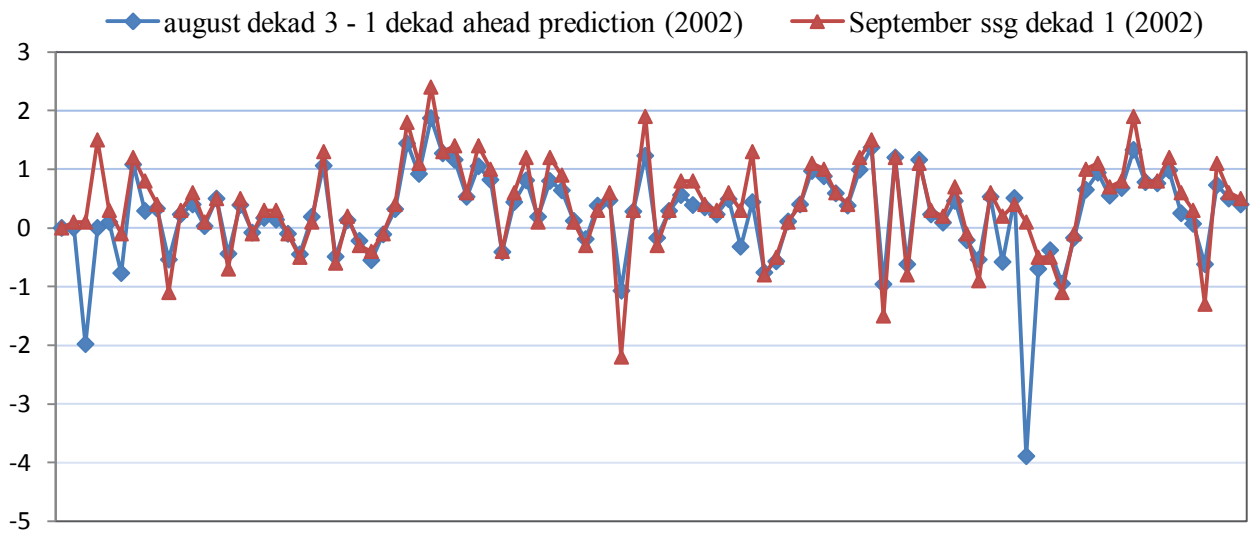


Figure 6.18: September dekad 1 prediction using August dekad 3 - 1 dekad ahead prediction model for 2002

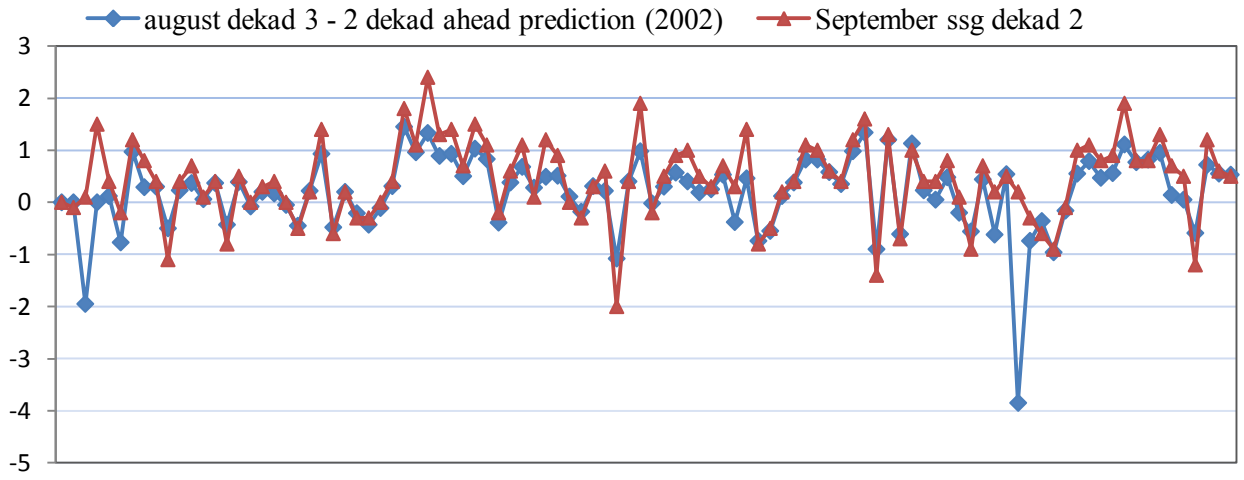


Figure 6.19: September dekad 2 prediction using August dekad 3 - 2 dekad ahead prediction model for 2002

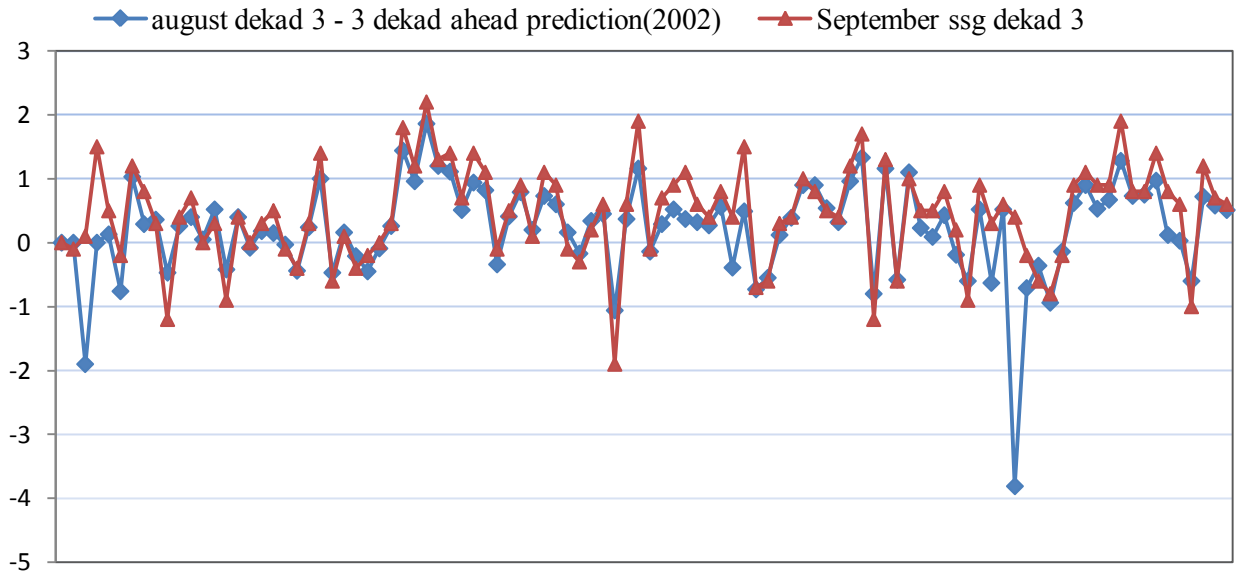


Figure 6.20: September dekad 1 prediction using August dekad 1 - 3dekad ahead prediction model for 2002

As in August dekad 1 and 2 predictions August dekad 3 prediction values fit to the actual values with exceptional accuracy, August dekad 3 – 1 dekad ahead prediction fitting best to the actual values compared to all August three dekads. Few outlier values are also seen in the predictions.

Comparison between dekadal prediction and monthly prediction

In this section we compare the prediction accuracy of the two prediction models. The monthly prediction models were developed by Getachew et al. [19]. In model construction, all the attributes used for the two models were the same except using values with different temporal resolution for SPI and SSG which are the most influential attributes of all. For dekadal prediction, dekadal (ten days) data were used while for monthly prediction monthly average was computed and used. And of course in dekadal prediction models construction, for the five oceanic index attribute values, each index value were used three times for the three dekads in a month as described in Chapter four. And for the two models same numbers of sample points (2812) from the same places (coordinates) were taken.

The monthly predictions has proven to be efficient, so the aim of the comparison is too see if dekadal predictions has better prediction accuracy compared to monthly predictions. Or is it a waste of time using data with higher temporal resolution (i.e., dekadal data)? The steps followed

to develop dekadal prediction models is very tedious compared to monthly prediction models, so if dekadal predictions doesn't improve the prediction accuracy then there is no need to use dekadal prediction models. And this will prove that using data with higher temporal resolution will not have an impact on the prediction accuracy of the models.

The plan for comparison is first comparing the correlation coefficient (*r*) values then comparing visually the linear plots produced for the two models by comparing the actual and predicted SSG values and finally comparing by checking whether the monthly prediction model has the ability to predict accurately using dekadal data.

Correlation coefficient (*r*) measures the agreement between the actual values and values predicted by the models. Correlation coefficient (*r*) for the monthly prediction on test data is presented in Table 6.1 below.

Table 6.1: Prediction accuracy on test data for monthly prediction

No	Models	Target month	Models Correlation-coefficient (r)
1	June one month prediction	July	0.85
2	June two month prediction	August	0.77
3	June three month prediction	September	0.70
4	June four month prediction	October	0.73
5	July one month prediction	August	0.92
6	July two month prediction	September	0.83
7	July three month prediction	October	0.74
8	August one month prediction	September	0.95
9	August two month prediction	October	0.73
10	September one month prediction	October	0.77

Correlation coefficient (*r*) for monthly prediction is compared to that of dekadal prediction in Table 4.1. As it can be seen from the two models for dekadal prediction all the first prediction values are always greater than monthly prediction. The highest *r* value for monthly prediction is 0.95 and 0.99 for dekadal prediction. The lowest *r* value for monthly prediction is 0.70 and for dekadal prediction it is 0.44. This shows that when we go far from the predicted period monthly prediction has accuracy way ahead than dekadal prediction. This difference is caused partly due to the fact that monthly prediction predicts as far as four periods ahead but dekadal prediction

can go as far as fourteen periods ahead. If the two models are compared for predicting four periods ahead, using dekadal prediction for any first four prediction periods *r* values are greater than or equal to 0.79 which is greater than 0.70, the lowest value for monthly prediction. But this comparison has its own limitation since four periods ahead in monthly prediction is four months away and in dekadal prediction it means four dekads ahead which is less than two months.

The next comparison is a visual comparison between the linear plots produced for both models. Here we check how much the predicted and the actual values of SSG fit on the linear plots. For monthly prediction, August 1 month ahead prediction (September prediction) has the highest *r* value and for dekadal prediction also using the three August dekad predictions to predict September has higher *r* values. So we used September prediction values (SSG values) of the years 1984 and 2002 for both models.

For dekadal predictions, the procedures followed to make the predictions are first using August dekad 1 we predicted August dekad 1 - 3, 4, and 5 dekad ahead predictions, which gave us September dekad 1, 2, and 3 prediction values then using August dekad 2 we predicted August dekad 2 - 2, 3, and 4 dekad ahead predictions, and this gave us another prediction for September dekad 1, 2, and 3. And finally using August dekad 3 we predicted August dekad 3 - 1, 2, and 3 dekad ahead predictions, which give as also different prediction for September dekad 1, 2, and 3. Each of the three dekads of August gives us different September dekad 1, 2, and 3 values. And for monthly prediction, using August 1 month ahead prediction we got September prediction values.

First, predicted values of both models for 1984 are compared with actual values.

Now the level of prediction accuracy for dekadal prediction is compared with the monthly prediction accuracy for September prediction of 1984 in Figure 6.21.

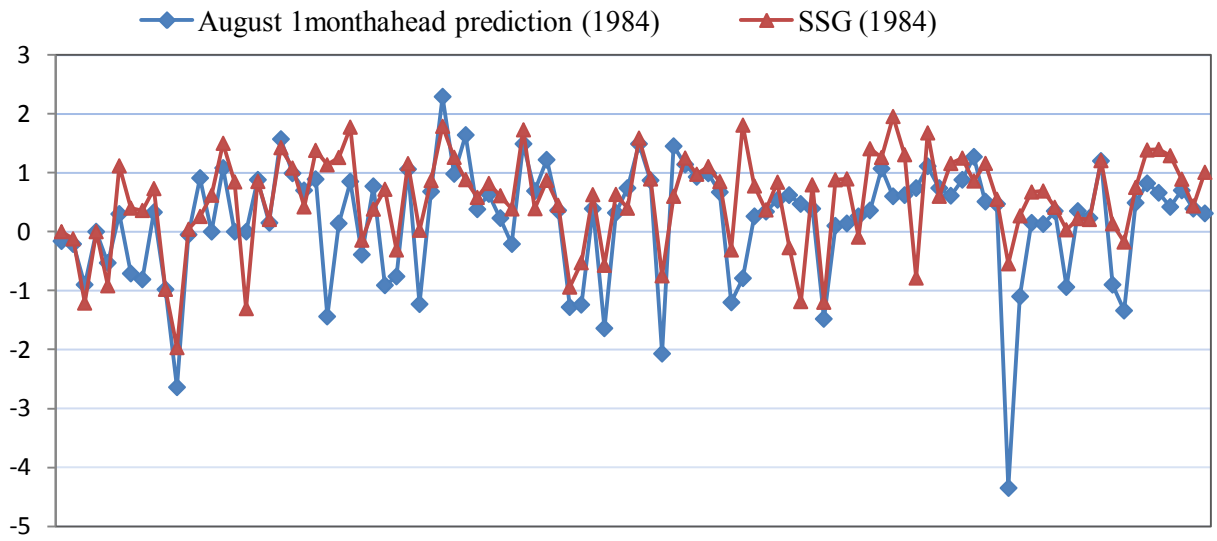


Figure 6.21: September prediction using August 1 month ahead prediction model for 1984

As it is shown in the linear plot the monthly prediction has also high accuracy, except it shows outlier values for some points. Here we compare visually the scatter plot results for the monthly prediction with each August dekad predictions separately.

First, the comparison is made with August dekad 1 predictions. When monthly prediction is compared to August dekad predictions it shows that August 1 month ahead prediction has wrongly deviated to negative values usually. August dekad 1-3 dekad ahead and August dekad 1-4 dekad ahead predictions has shown higher prediction accuracy even though they have lower *r* values(0.93 and 0.87 respectively) compared to August 1 month ahead prediction(0.95). Most of August dekad 1-5 dekad ahead prediction values better fit the actual values than August 1 month ahead prediction but for some values it is less. The visual comparison is made from Figure 6.22.

Secondly, August 1 month prediction was compared with August dekad 2 predictions. As it can be visualized from Figure 6.23 it shows that August dekad 2-2 dekad ahead prediction best fit the actual values than August 1 month ahead prediction. On August dekad 2-3 dekad ahead prediction most prediction values fit better to actual values than August 1 month ahead prediction. And for August dekad 2-3 dekad ahead prediction some values better fit than August 1 month ahead prediction and vise versa. The outlier values in August dekad 2-3 dekad ahead and August dekad 2-4 dekad ahead predictions are greater than August 1 month ahead

prediction. As indicated earlier August 1 month ahead prediction values deviate often towards negative wrongly.

And thirdly, the comparison is done with August dekad 3 predictions. All the predictions for August dekad 3 predictions better fit the actual values than August 1 month ahead prediction except on some points on August dekad 3-3 dekad ahead prediction as it can be seen in Figure 6.24. August dekad 3-1 dekad ahead prediction fit best to the actual values from all August dekad predictions.

Generally, from the comparisons we can see that August 1 month ahead prediction fits well with actual values but when it is compared with August dekad predictions it has lower prediction accuracy. Outlier values are usually greater for August dekad predictions and it needs to be looked on in the future.

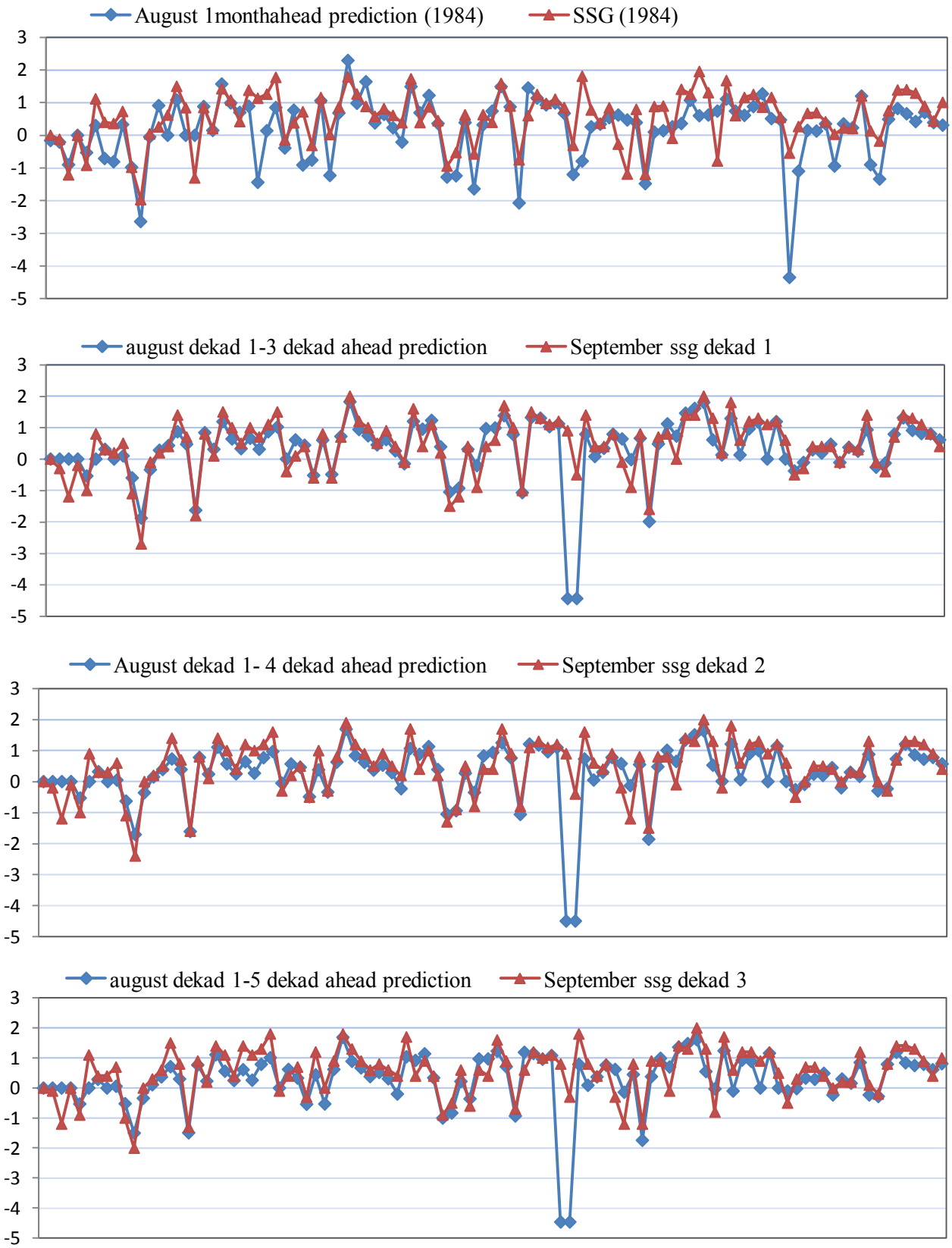


Figure 6.22: Comparison between August dekad 1 predictions and August 1 month prediction for 1984

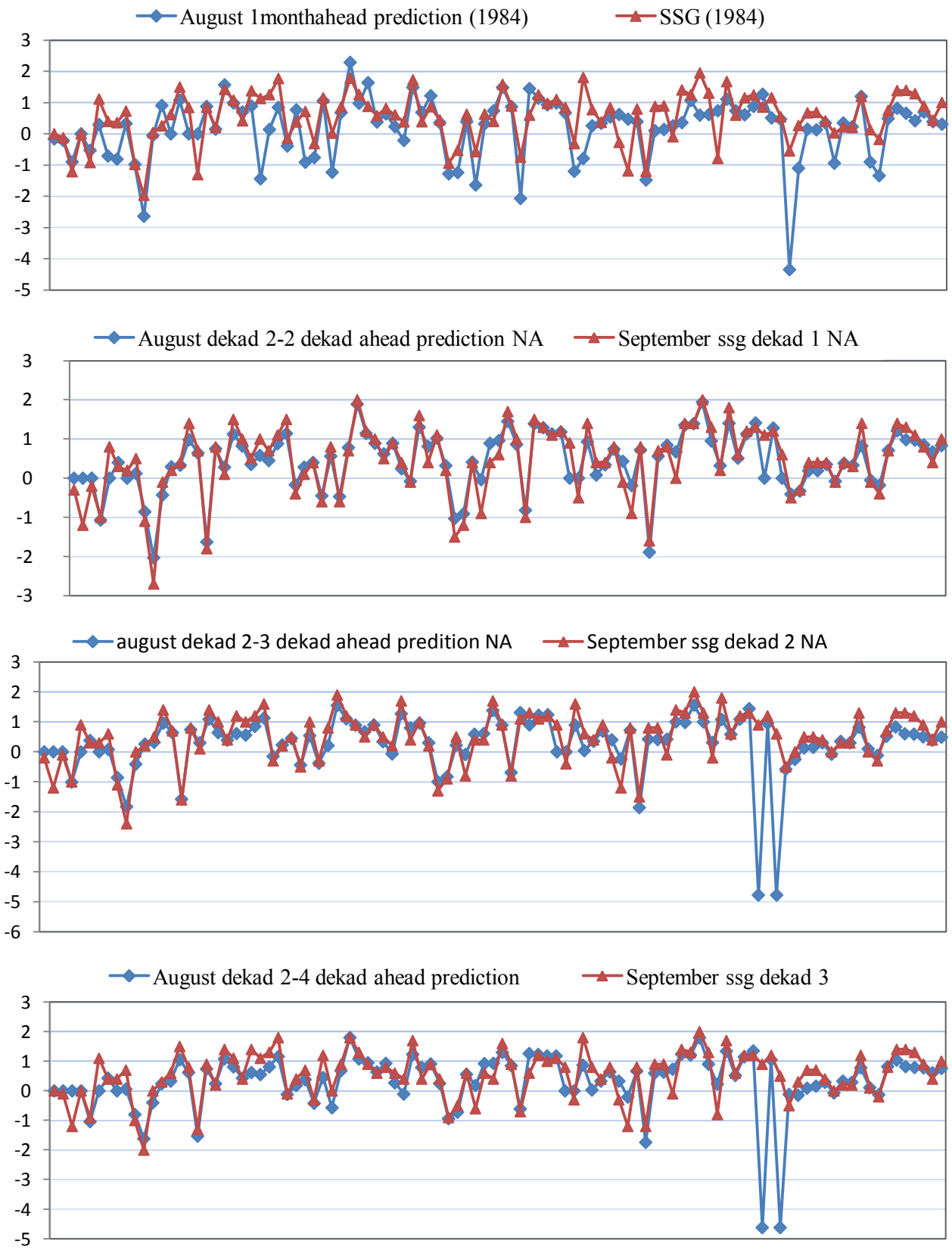


Figure 6.23: Comparison between August dekad 2 predictions and August 1 month prediction for 1984

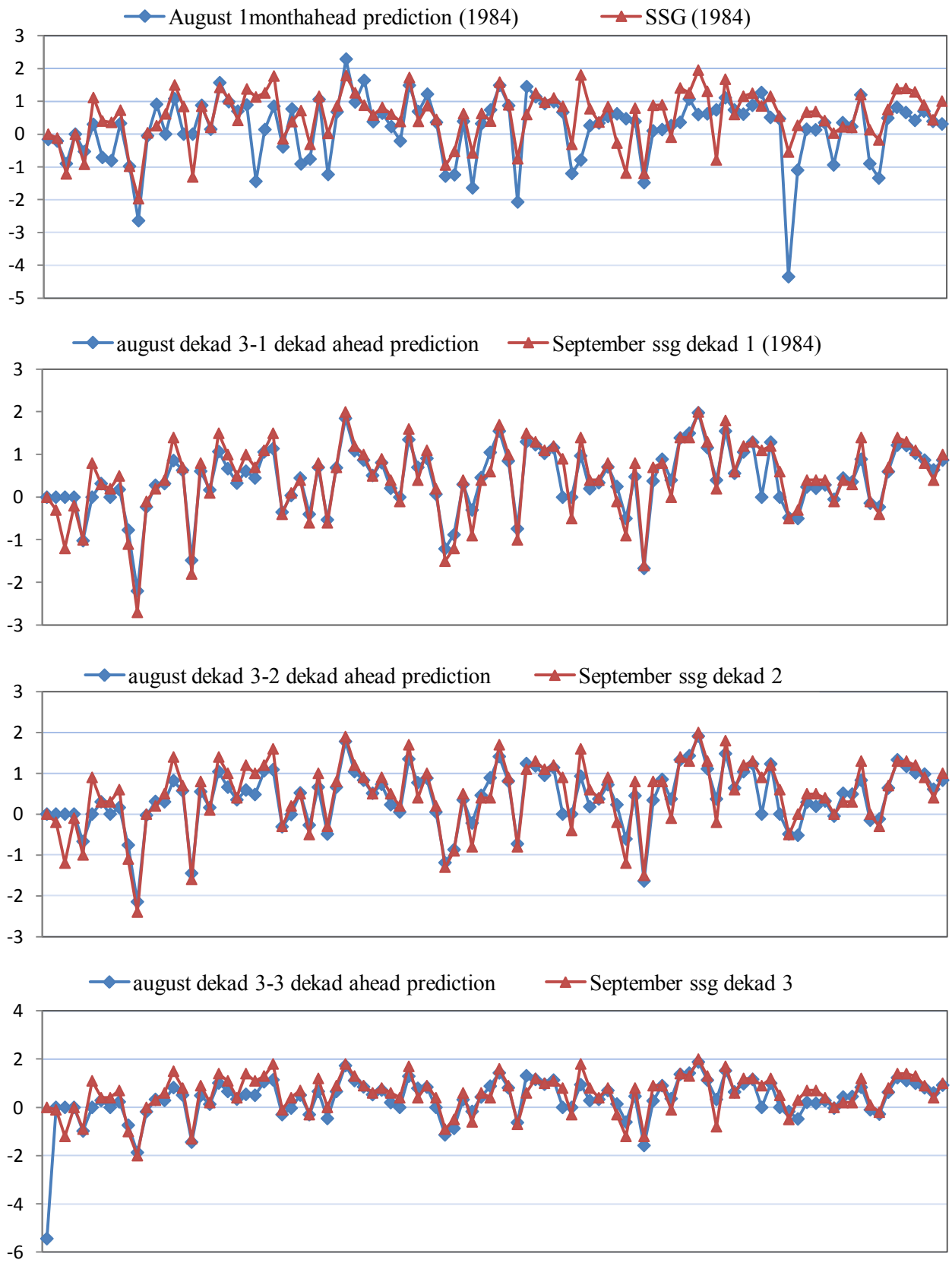


Figure 6.24: Comparison between August dekad 3 predictions and August 1 month prediction for 1984

Next, predicted values of both models for the year 2002 are compared with actual values.

The level of prediction accuracy for dekadal predictions is compared with the monthly prediction accuracy for September prediction of 2002 in Figure 6.25.

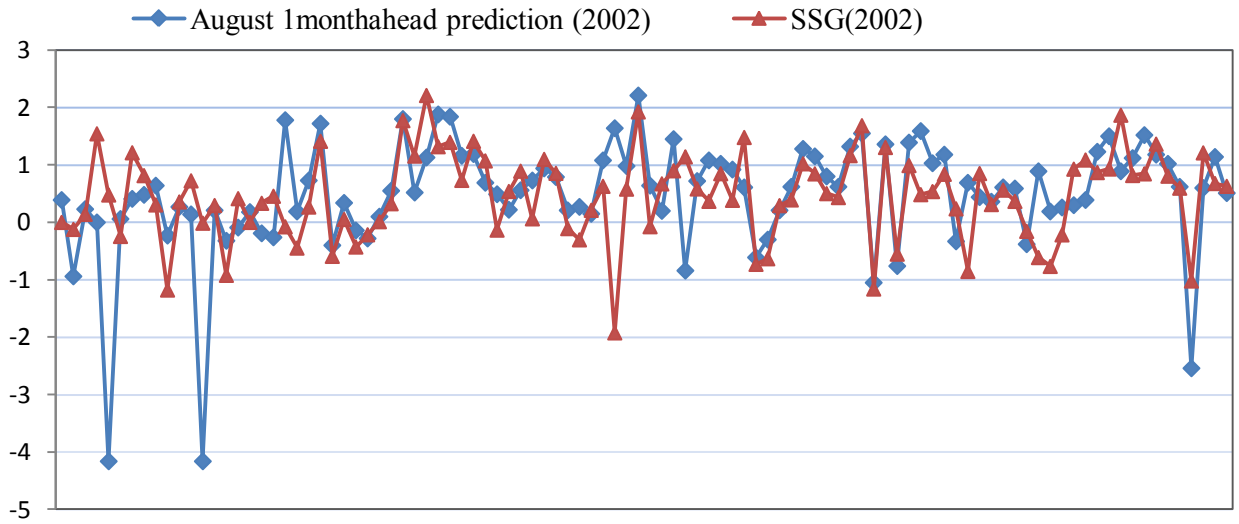


Figure 6.25: September prediction using August 1 month ahead prediction model for 2002

In the linear plots in Figure 6.26 both August 1 month prediction and August dekad 1 predictions has almost the same pattern, with almost similar data, and due to that similar prediction values are expected. August 1 month prediction slightly deviates from the actual values both to the positive and negative. And all the three August dekad 1 values better fit to the actual values than August 1 month prediction, August dekad 1-3 dekad ahead prediction having the highest accuracy.

As in August dekad 1 predictions in August dekad 2 also all the three predictions better fit to the actual values than August 1 month prediction as can be seen in Figure 6.27. August dekad 2 - 2 dekad ahead prediction has the highest accuracy as it is nearer to the prediction period.

And it is shown in Figure 6.28 all the three predictions in August dekad 3 better fit to the actual values than August 1 month prediction as in August dekad 1 and 2, August dekad 3 - 1 dekad ahead prediction having the highest accuracy.

Generally, all August dekad prediction values better fit the actual compared to August 1 month prediction but this time the difference is very small as the actual SSG values (August data) for both prediction models are almost similar and August 1 month prediction having prediction

accuracy with 0.95 *r* value. Compared to August 1 month prediction for 1984 the actual values fit better for August 1 month prediction of 2002 and this shows that monthly prediction predicts better for 2002 than 1984.

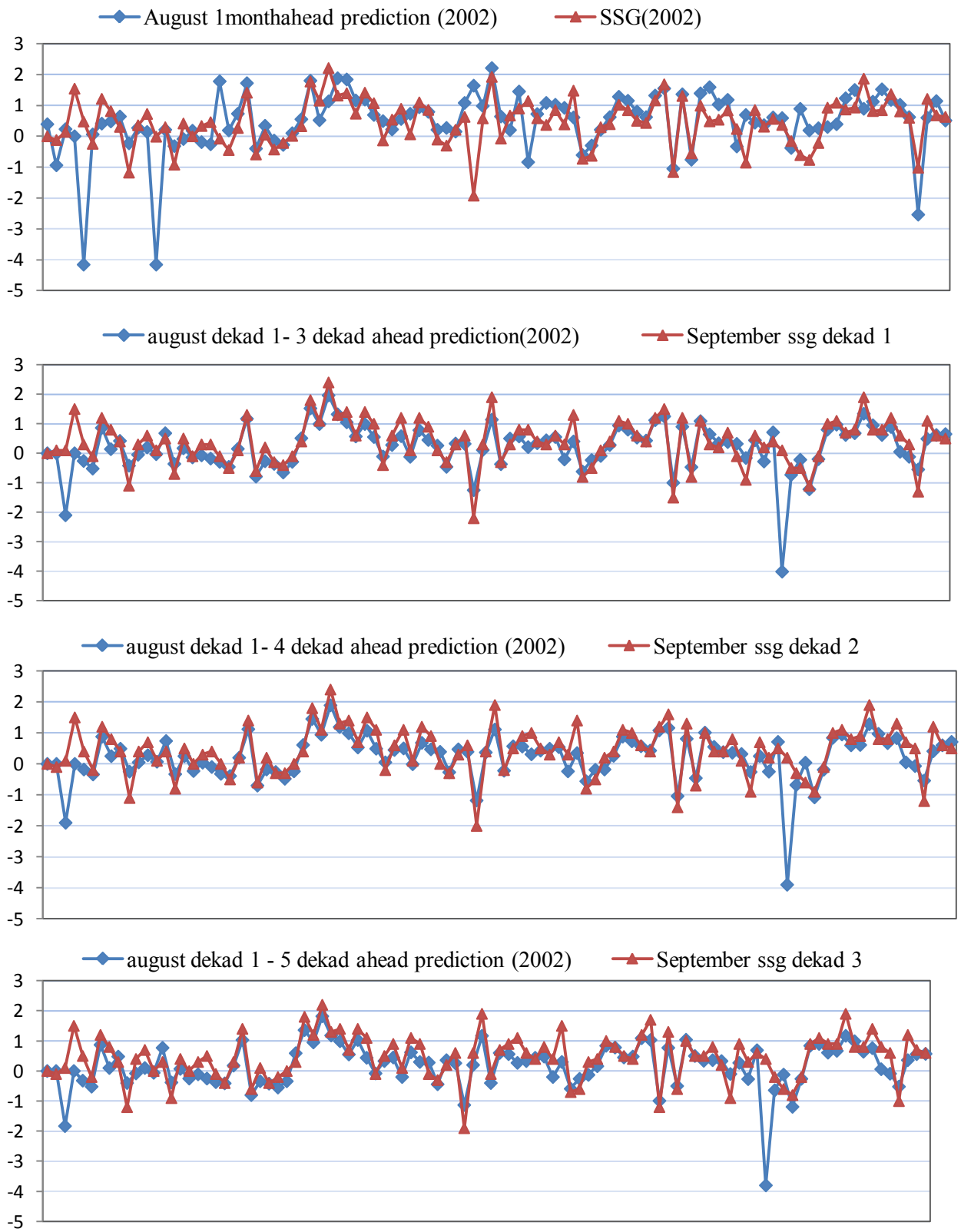


Figure 6.26: Comparison between August dekad 1 predictions and August 1 month prediction for 2002

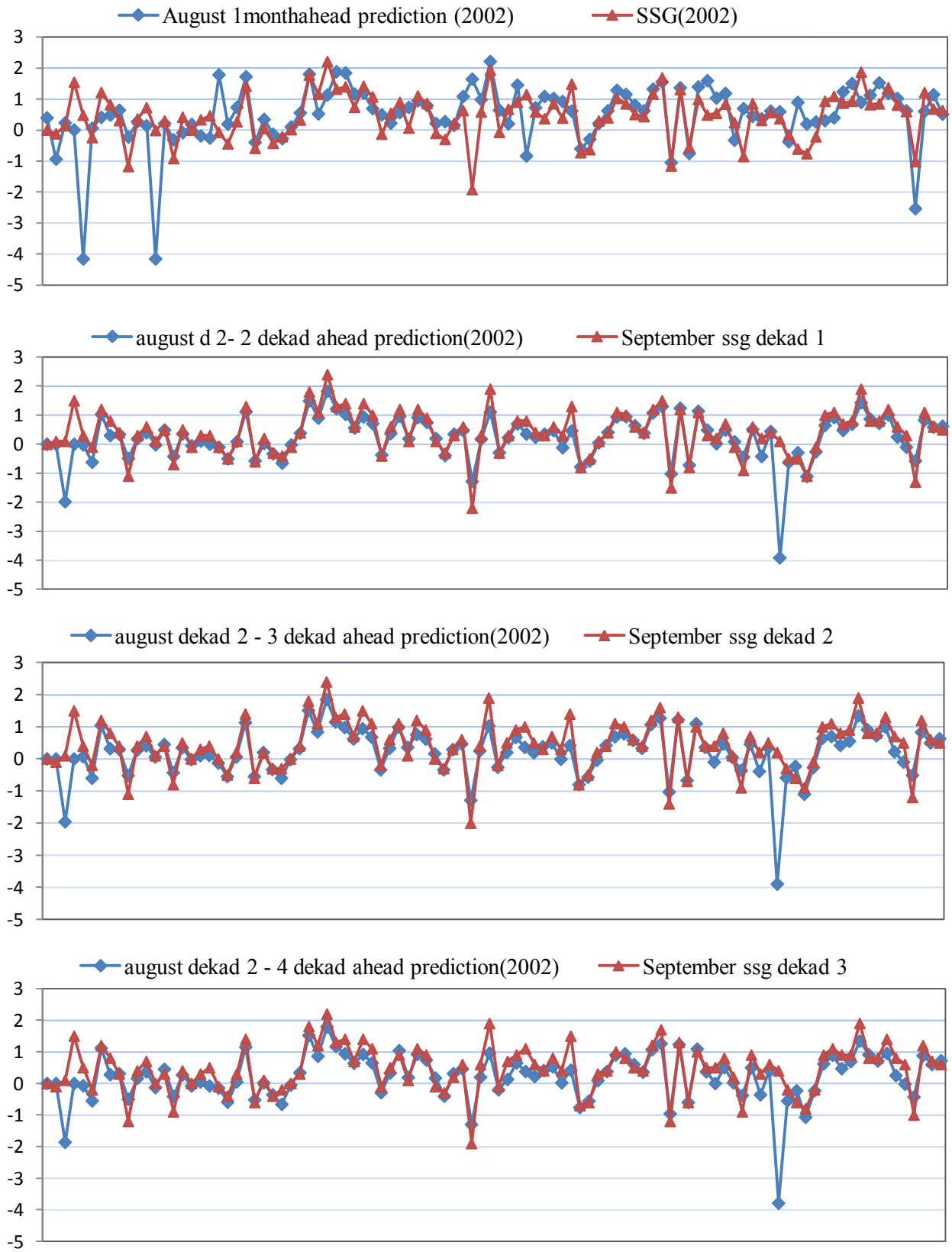


Figure 6.27: Comparison between August dekad 2 predictions and August 1 month prediction for 2002

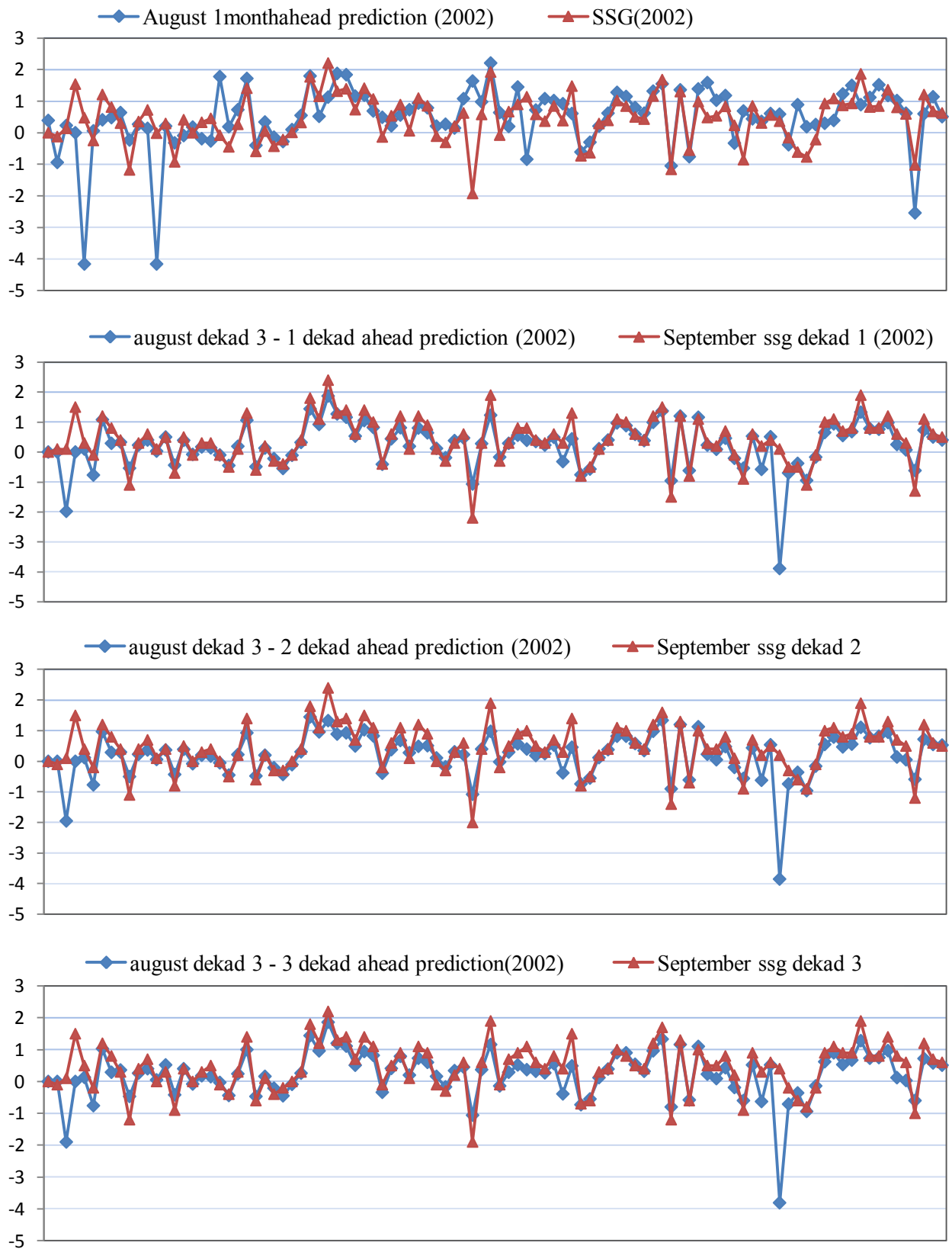


Figure 6.28: Comparison between August dekad 3 predictions and August 1 month prediction for 2002

The final comparison is done by checking whether the monthly prediction has the ability to predict using dekadal data. So here August dekad 3 data for both 1984 and 2002 is given as input for monthly prediction model and the resulted predictions are compared to August dekad 3-1 dekad ahead predictions.

September prediction maps both for August dekad 3 - 1 dekad ahead prediction and August 1 month prediction of 1984 is presented below in Figure 6.29.

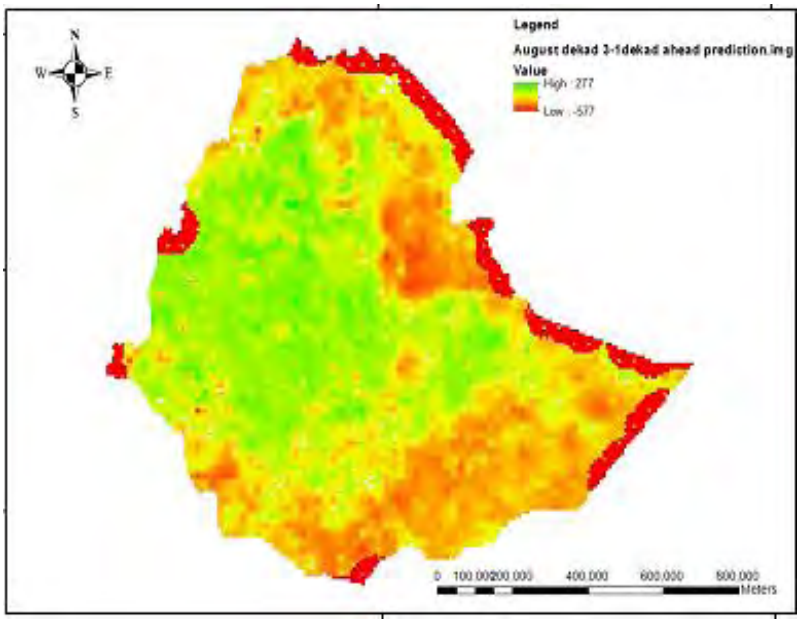


Figure 6.29: a. Dekadal prediction for September using August dekad 3 data of 1984

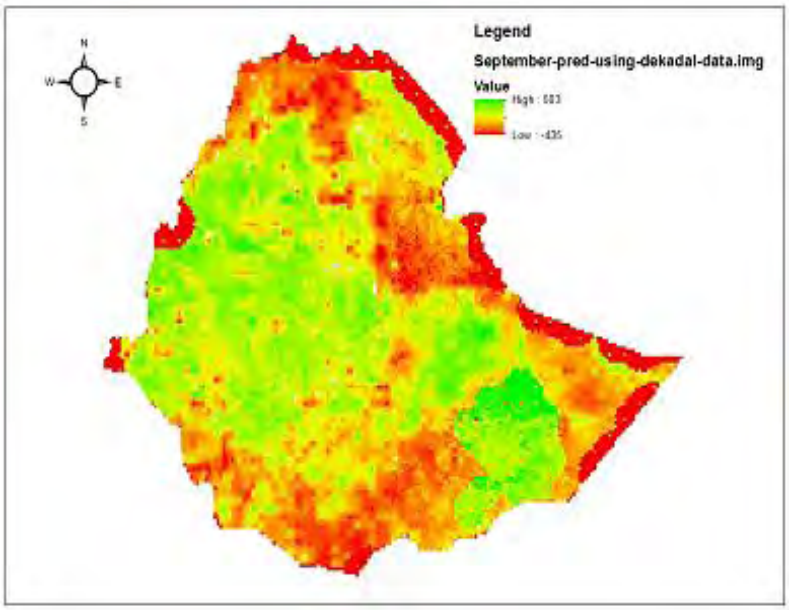


Figure 6.29: b. Monthly prediction for September using August dekad 3 data of 1984

As discussed earlier, August dekad 3 - 1 dekad ahead prediction values fits to actual values of September 1984 with higher accuracy. Therefore monthly prediction's prediction accuracy using dekadal data is assessed based on dekadal prediction in Figure 6.29 (a). As can be seen from the maps on some parts of south eastern Ethiopia for monthly prediction, Figure 6.29 (b), moist condition is shown which is wrongly predicted.

Linear plot of the two models in Figure 6.30 also shows that the monthly prediction, August 1 month ahead prediction, deviates a lot for some points.

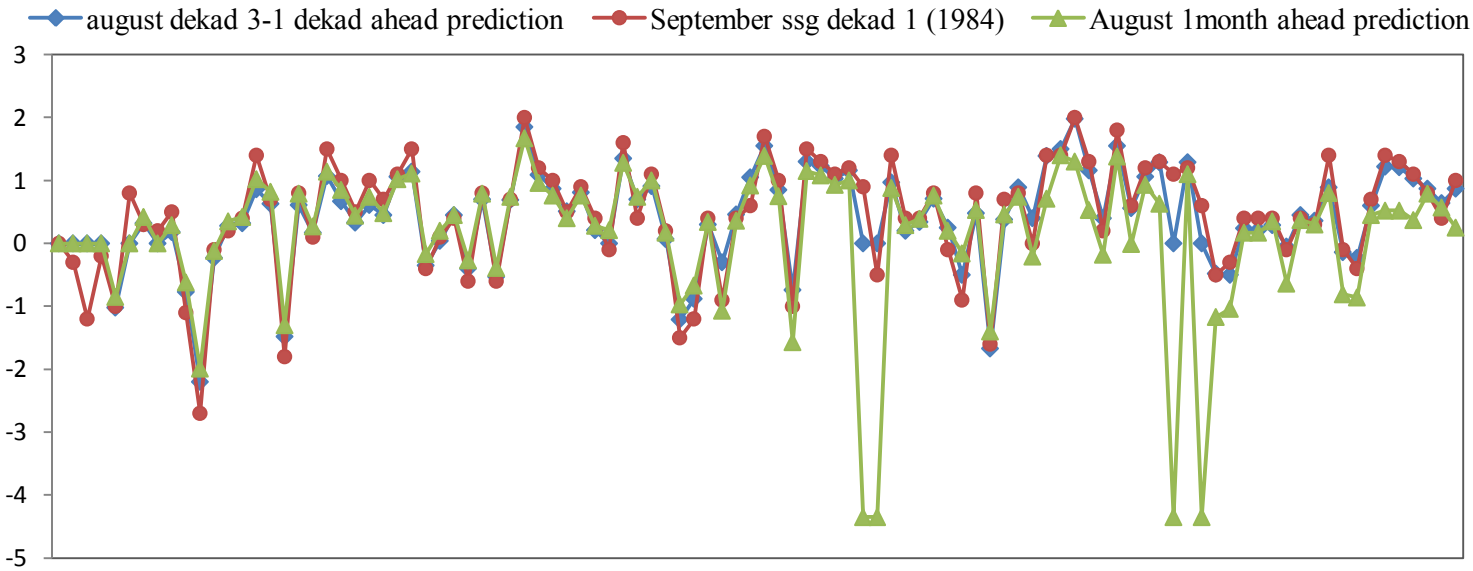


Figure 6.30: Dekadal prediction versus monthly prediction using dekadal data of 1984

For September 2002 also dekadal and monthly predictions, using dekadal data as input, are compared. From Figure 6.31 it can be seen that both of the prediction models produced almost similar prediction maps. This is caused because both the monthly and dekadal periods have almost similar actual SSG values as indicated earlier in the linear plots comparison.

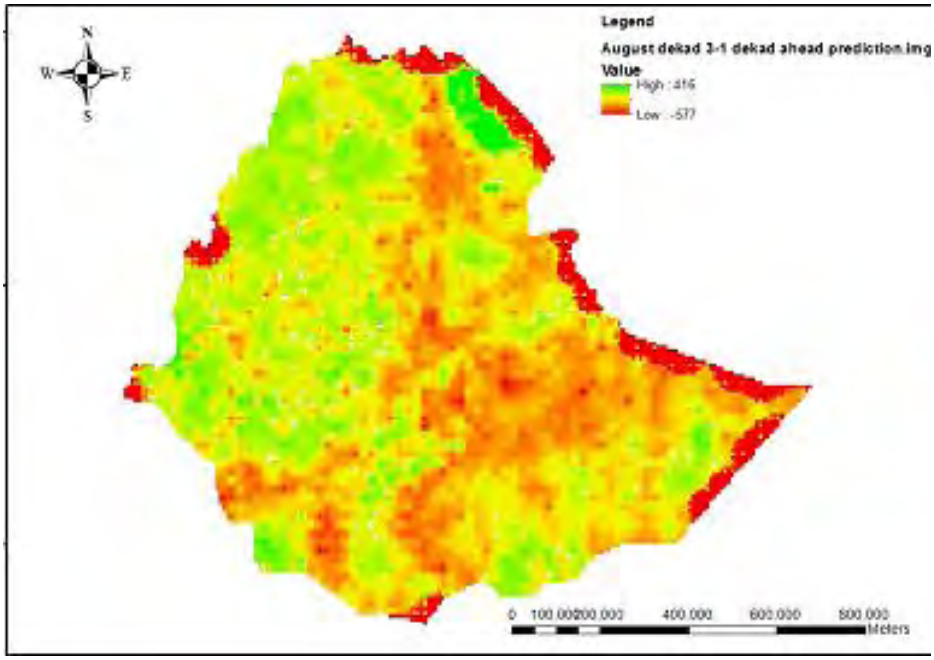


Figure 6.31: a. Dekadal prediction September using August dekad 3 data of 2002

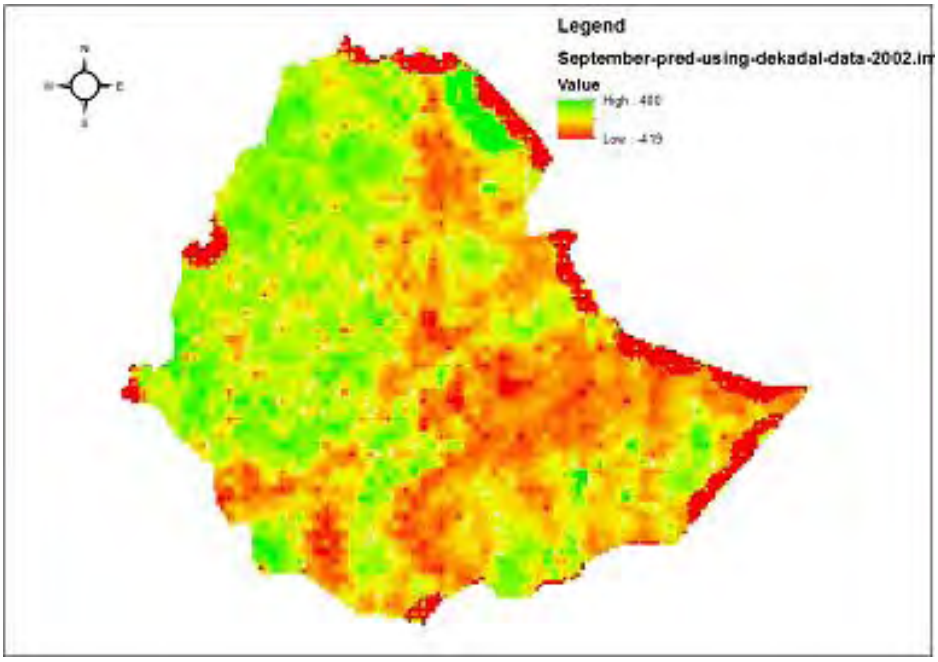


Figure 6.31: b. Dekadal prediction September using August dekad 3 data of 2002

The linear plot in Figure 6.32 also shows that the two models have almost same level of prediction accuracy.

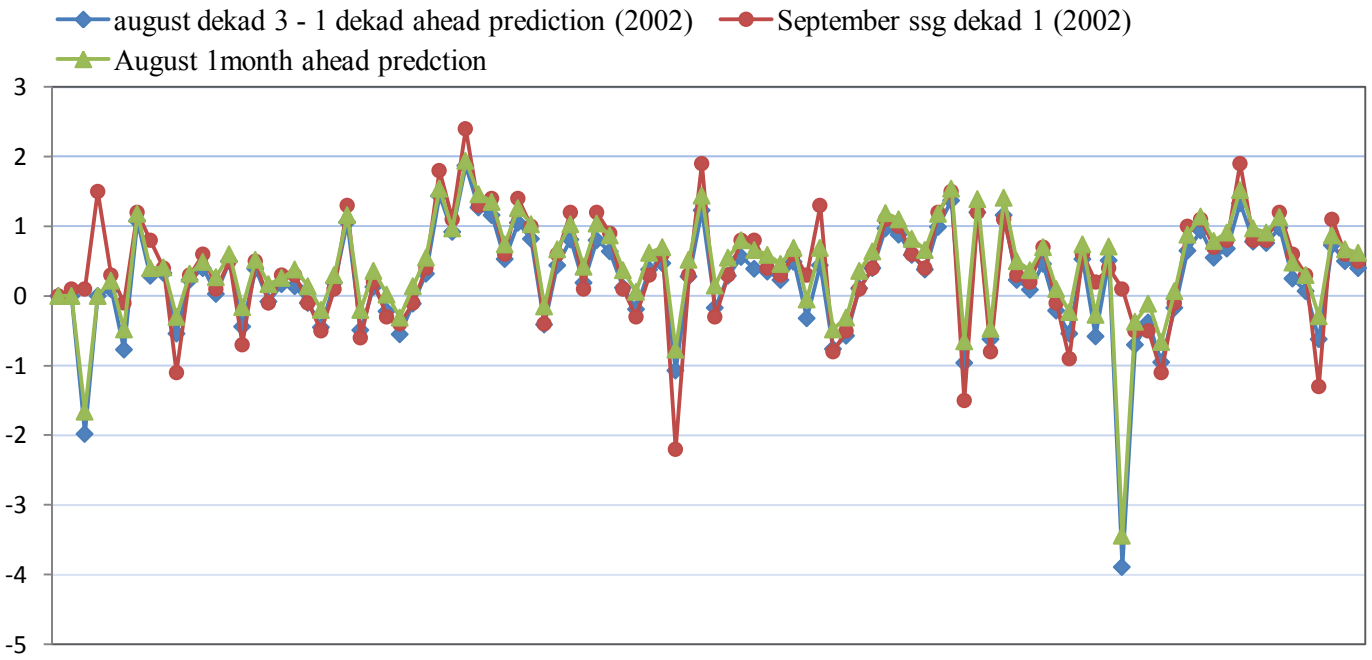


Figure 6.32: Dekadal prediction versus monthly prediction using dekadal data of 2002

From the comparisons of the two models for the years 1984 and 2002, using dekadal data as input, we can see that monthly prediction does not have the ability to predict correctly for the whole parts of Ethiopia unless the actual SSG values for the two periods are somehow similar. And we do not recommend monthly prediction models to be used for dekadal predictions as the results of bad predictions like the one in 1984 can create major catastrophes.

Generally, in all the three different ways of comparisons it is clearly visible that dekadal predictions have better prediction accuracies than monthly predictions. And of course developing the dekadal prediction models was important.

Apart from the above comparisons dekadal predictions has the advantage of requiring only for 10-days data to make predictions unlike monthly prediction which requires a full month data. This also helps in acquiring updated data on vegetation condition in every 10 days.

7. CONCLUSION AND FUTURE WORK

7.1. Conclusion

Drought has caused a lot of damages through the years in Ethiopia. It is the world's costliest natural disaster and affects more people than any other form of natural disaster. In 1984 alone, around 1,000,000 people have lost their lives because of the drought in Ethiopia. Drought is normal part of the climate virtually occurring in all parts of the world and has continued to cause problems. Ethiopia is one of the countries which are highly affected by drought recently. So this brought as the urgency to develop tools using recent technologies for monitoring and prediction of drought to help decision makers take actions to mitigate or reduce the damages caused by drought.

There have been some efforts in using satellite data for drought prediction and monitoring in Ethiopia in the past and our work has tried to contribute on it.

Researches proved that integration of remotely sensed data with other relevant environmental variables (e.g., climate, oceanic observations) that influence vegetation holds considerable potential for improving our capabilities to predict future vegetation conditions.

In our work we integrated four types of data including: climate, satellite, biophysical, and oceanic data. RFE data was selected representing climate data. NDVI was selected representing satellite vegetation data. Representing biophysical data four attributes including Water Holding Capacity, Land Cover, Digital Elevation Model, and Vegetation Type of Ethiopia were selected. And representing oceanic index five attributes including Atlantic Multi-decadal Oscillation Index, Pacific Decadal Oscillation, North Atlantic Oscillation, Multivariate ENSO Index, and Pacific North American Index were selected. The eleven attributes are considered to be the most important parameters in contributing to vegetation monitoring and prediction in Ethiopia.

A 24 year historical data, for all the attributes, between the years 1983 and 2006 were collected. This is done because the data on these years were complete and easily accessible.

After data collection, the data were preprocessed to make them ready for model development. And then they were transformed to coordinate system where all the data fit geometrically. In preprocessing, SPI showing whether a station is experiencing drought or not was calculated from RFE data, SSG indicating agricultural drought severity was derived from NDVI and raster maps

were created in ArcGis for all the attributes which were then used in model implementation. And in transformation all attributes spatial resolution were re-sampled to 8km, since NDVI (the source for SSG) was acquired with spatial resolution of 8km. And their coordinate systems were changed to that of NDVI so that they fit geometrically.

Once the data were preprocessed and transformed using Cubist, a commercial CART algorithm, the historical data were analyzed and rule-based linear regression drought prediction models were generated. MapCubist software was then used to generate prediction maps. The model was implemented for the years 1984 and 2002 where severe drought occurred in Ethiopia. The implemented models for 1984 showed the real situation at the time i.e., an extreme drought in the north (Tigray and Wollo, Gonder), and south east (hararghe - Somali region). And for 2002 also, the implemented model confirmed the real situation showing extreme drought in the south east (Somali region) which is also predicted with our models.

Then prediction accuracies of dekadal prediction and monthly prediction models were compared to answer the question we raised in the beginning i.e., "can using dekadal (10-day composites of daily data) prediction give better result compared to monthly average in monitoring drought?". Results of the comparison showed that dekadal predictions have better prediction accuracies than monthly prediction models.

Finally we conclude by encouraging decision makers and other stake holders in drought risk-management and mitigation to use the model as using the model brings significant change on the accuracy of the information we get on drought.

7.2. Future work

Even though our work has produced significant information gain over monthly prediction it has limitations. So in order to get optimum prediction accuracy, works should proceed on the following points.

The spatial resolution of the data used in the study was 8 km (1 pixel representing 8x8 km^2 area), which has very low quality. The work can be repeated on data with higher temporal resolution.

An Investigation can also be made on other relevant attributes which can contribute on prediction accuracy.

The preprocessing part was done manually and it was a very tedious process. So if a preprocessing component integrating ArcGis, ERDAS, ILWIS, and the drought prediction system is added it helps users of the system a lot.

References

1. Drought: http://www.drought.unl.edu/DroughtBasics/TypesofDrought.aspx, accessed on October 4, 2011
2. Famine in Ethiopia: http://www.en.wikipedia.org/wiki/1983-1985_famine_in_Ethiopia, accessed on January 7, 2012
3. SPI: http://www.in.gov/dnr/water/4864.htm, accessed on December 12, 2011
4. Remote sensing: http://www.pages.csam.montclair.edu/~chopping/rs/CCRS, accessed on September 15, 2011
5. NDVI: http://www.phenology.cr.usgs.gov/ndvi_foundation.php, accessed on Deember 13, 2011
6. Drought: http://www.ponce.sdsu.edu/three_issues_droughtfacts01.html, accessed on February 11, 2011
7. CART algorithm: http://www.rulequest.com/cubist-win.html, accessed on May 18, 2012.
8. A. A. Gitelson, "Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation," Journal of Plant Physiology, vol. 161, pp. 165–173, 2004.
9. A. J. Peters, E. A. Walter-Shea , L. Ji, V. Vina, M. Hayes, and M. D. Svoboda, "Drought Monitoring with NDVI-Based Standardized Vegetation Index," Journal of Photogrammetry and Remote Sensing, vol. 68, pp.71-75, 2002.
10. A. Peters, A. Elizabeth, Walter-Shea, J. Lei, A. Vliia, M. Hayes, and M. Svoboda, "Drought Monitoring with NDVI-Based Standardized Vegetation Index," Photogrammetric Engineering & Remote Sensing, vol. 68, pp. 71-75, 2002.
11. A. Prasad, R. Singh, V. Tare, and M. Kafatos, "Use of vegetation index and metrological parameters for the prediction of crop yield in India," International Journal of Remote Sensing, vol. 28, pp. 5207 - 5235, 2007.
12. A. R. Huete, K. Didan, Y. E. Shimabukuro, P. Ratana, S. R. Saleska, L. R. Hutyra, W. Yang, R. R. Nemani, and R. Myneni, "Amazon Rainforests Green-up with Sunlight in Dry Season," Geophysical Research Letters, vol. 33 , 2006.

13. A. Sharma, “Spatial data mining for drought monitoring: An approach using temporal NDVI,” Society, vol. 78: pp. 621-636, 2006.
14. B.A. Shafer and L.E. Dezman, “Development of a Surface Water Supply Index (SWSI) to assess the severity of drought conditions in snowpack runoff areas,” In: Proceedings of the Western Snow Conference, Reno, NV, vol. 19-23, April 1982, pp. 164- 175.
15. B. C. Gao, “NDWI—a Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space,” Remote Sensing of Environment, vol. 58, pp. 257–266, 1996.
16. B.C. Reed, J.F. Brown, D. VanderZee, T.R. Loveland, J.W. Merchant, and D.O. Ohlen, “Measuring phenological variability from satellite imagery,” Journal of Vegetation Science, vol. 5, pp. 703–714, 1994.
17. B. C. Reed, T. R. Loveland, and L. L. Tieszen, “An Approach for Using AVHRR Data to Monitor U.S. Great Plains Grasslands,” Geocarto International, vol. 11, pp.13-22, 1996.
18. B. Getachew, T. Tsegaye, A. Solomon, H. Shawndra, and T. Yitaktu, “Application of Normalized Difference Vegetation Index and Standard Precipitation Index Parameters to Monitor Drought at National Scale: The Case of Ethiopia,” Ethiopian Journal of Development Research, vol. 34, pp. 67 - 94, 2012.
19. B. Getachew, T. Tsegaye, A. Solomon, H. Shawndra, “Drought Information Mining from Satellite Images for Improved Climate Change Mitigation,” Information and Communication Technologies (WICT), 2012 World Congress on, pp. 197 - 202, 2012.
20. B. Holben, “Characteristics of maximum-value composite images for temporal AVHRR data,” International Journal of Remote Sensing, vol. 7, pp.1435–1445, 1986.
21. C. Domenikiotis, A. Loukas, and N. R. Dalezios, “The Use of NOAA/ AVHRR Satellite Data for Monitoring and Assessment of Forest Fires and Floods,” Natural Hazards and Earth System Science, vol. 3, pp. 115–128, 2003.
22. C. C. Funk and M. E. Brown, “Intra-seasonal NDVI Change Projections in Semi-arid Africa,” Remote Sensing of Environment, vol. 101, pp. 249-256, 2006.
23. C.M. Rulinda, W. Bijker, and A. Stein, “Image mining for drought monitoring in eastern Africa using Meteosat SEVIRI data” In: International journal of applied earth observation

and geoinformation : JAG, 12 (2010)Supp. 1 : Issue on Remote Sensing for Africa : Special Collection from AARSE pp. S63 S68, 2010.

24. D.A. Wilhite, and M.H. Glantz, "Understanding the Drought Phenomenon: The Role of Definitions," *Water International* vol. 10, pp. 111–120, 1985.
25. D. A . Wilhite, "Integrated climate monitoring for drought detection," In: Drought: A Global Assessment,hazards and Disasters Series, *Routledge*, London and New York, Vol. 1, *Ed. D.A. Wilhite*, pp. 145-158, 2000.
26. D. A . Wilhite, "Drought," in Encyclopedia of World Climatology, Oliver, J. (Ed.), New York, NY: Springer, pp. 338-340, 2005.
27. D. B. Enfield, A. M. Mestas-Nunez, and P. J. Trimble, "The Atlantic multidecadal oscillation and it's relation to rainfall and river flows in the continental U.S.," Geophysical Research Letters, vol. 28, pp. 2077-2080, 2001.
28. D. D. Breshears, N. S. Cobb, P. M. Rich, K. P. Price, C. D. Allen, R. G. Balice, W. H. Romme, J. H. Kastens, M. L. Floyd, J. Belnap, J. J. Anderson, O. B. Myers, and C. W. Meyer, "Regional Vegetation Die-off in Response to Global-change-type Drought," Proceedings of the National Academy of Sciences of the United States of America, vol. 102: pp. 5144–5148, 2005.
29. Ecodiv.org, "Atlas of the Potential Vegetation of Ethiopia," vol. 2011, P. Breugel, Ed.: Copyright Ecodiv.org, 2010.
30. EMA, National Atlas of Ethiopia. Addis Ababa: Ethiopian Mapping Authority (EMA), 1988.
31. ESA, "European Space Agency, Global Land Cover Map," vol. 2011: ESA, 2011.
32. F. J. Brown, and B. C. Reed, "A Prototype Drought Monitoring System Integrating Climate and Satellite Data," Percora, 15/Land Satellite Information IV/ ISPRS Commission I/ FIEOS 2002, 2002.
33. F. N. Kogan, "Remote Sensing of Weather Impacts on Vegetation in Non-homogeneous Areas," International Journal of Remote Sensing, vol.11, pp. 1405–1419, 1990.
34. F. N. Kogan, "Global Drought Watch from Space," Bulletin of the American Meteorological Society," vol. 78, pp. 621–636, 1997.
35. GLCF, "Global Land Cover Facility," vol. 2010: University of MARYLAND, 2010.

36. G. Kiros Fassil, “Economic Consequences of Drought, Crop Failure and Famine in Ethiopia, 1973-1986”, Issue 1 of Ethiopian journal of development research: Monograph series, Indiana University, 1990.
37. G. P. Asner, A. R. Townsend, and B. H. Braswell , “Satellite Observation of El Niño Effects on Amazon Forest Phenology and Productivity,” Geophysical Research Letters, vol. 27, pp. 981–984, 2000.
38. H. Legesse, “Drought Risk Assessment using remote sensing and GIS: A case study in southern zones, Tigray region, Ethiopia,” Master’s Thesis, Addis Ababa University, 2010.
39. IRI, "International Research Institute (IRI) for Climate and Society, IRI/LDEO Climate Data Library," vol. 2011: IRI, 2011
40. J.C. Eidenshink, “The 1990 conterminous U.S. AVHRR dataset,” Photogrammetric Engineering and Remote Sensing, vol. 58 , pp. 809- 813, 1992.
41. J. F. Brown, B. D. Wardlow, T. Tadesse, M. J. Hayes, and B. C. Reed, "The Vegetation Drought Response Index (VegDRI): A New Integrated Approach for Monitoring Drought Stress in Vegetation," GIScience and Remote Sensing, vol. 45, pp. 16–46, 2008.
42. J. Gu, F. Y. Brown, J. P. Verdin, and B. Wardlow, “A Five-Year Analysis of MODIS NDVI and NDWI for Grassland Drought Assessment over the Central Great Plains of the United States,” Geophysical Research Letters, vol. 34, pp.1–6, 2007.
43. J. Keyantash and J.A. Dracup, “The Quantification of Drought: An Evaluation of Drought Indices,” Bulletin of the American Meteorological Society, vol. 83, pp. 1167-1180, 2002
44. J. P. Malingreau, “Global Vegetation Dynamics: Satellite Observations Over Asia,” International Journal of Remote Sensing, vol. 7, pp.1121–1146, 1986.
45. J. W. Hurrell, "Decadal trends in the North Atlantic Oscillation and relationships to regional temperature and precipitation," Science, vol. 269, pp. 676-679, 1995.
46. K. E. Trenberth, G. W. Branstator, D. Karoly, A. Kumar, N. C. Lau, and C. Ropelewski, "Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures," journal of geophysical research, vol. 103, pp. 14291 - 1432, 1998.

47. K. Wolter and M. S. Timlin, "Measuring the strength of ENSO - how does 1997/98 rank?," Weather Forecasting, vol. 53, pp. 315-324, 1998.
48. L. Di, D. C. Rundquist, and L. Han, “Modeling Relationships between NDVI and Precipitation during Vegetative Growth Cycle,” International Journal of Remote Sensing, vol. 15, pp. 2121–2136, 1994.
49. L. Ji and A. J. Peters, "Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices," Remote Sensing of Environment, vol. 87, pp. 85 - 98, 2003.
50. M. l. J. Hayes, O. V. Wilhelmi, and C. L. Knutson, "Reducing Drought Risk: Bridging Theory and Practice," Natural Hazards Review, vol. 5, pp. 106 - 113, 2004.
51. NOAA, "National Oceanic and Atmospheric Administration, Climate Indices: Monthly Atmospheric and Ocean Time Series," vol. 2011: Earth System Research Laboratory, Physical Science Division, 2011.
52. P. D. Jones, T. Jonsson, and D. Wheeler, "Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and South-West Iceland," International Journal of Climatology, vol. 17, pp. 1433-1450, 1997.
53. P. Thorton, P. Jones, G. Alagarswamy, and J. Andresen, "Spatial variation of crop yield response to climate change in East Africa," Global Environmental Change, vol. 19, pp. 54-65, 2009.
54. Quinlan, J. R. 1993, C4.5 Programs for Machine Learning, San Mateo, CA: Morgan Kaufmann Publishers, 302 p. Rulequest, 2009.
55. R. Hanson, “The Rough Guide to Climate Change,” 2nd ed. New York, NY: Rough Guides Ltd., 374 p, 2008.
56. S. J. Goetz and S. D. Prince, “Remote Sensing of Net Primary Production in Boreal Forest Stand,” Agricultural and Forest Meteorology, vol. 78: pp. 149–179, 1996.
57. S. N. Goward, C. J. Tucker, and D. G. Dye, “North American Vegetation Patterns Observed with the NOAA-7 Advanced Very High Resolution Radiometer,” Vegetatio, vol. 64, pp. 3–14, 1985.
58. S. O. Los, G. J. Collatz, L. Bounoua, P. J. Sellers, and C. J. Tucker, "Global Interannual Variations in Sea Surface Temperature and Land Surface Vegetation, Air Temperature, and Precipitation," Journal of Climate, vol. 14, pp. 1535–1549, 2001.

59. S. Aditi, “Spatial data mining for drought monitoring: An approach using temporal ndvi and rainfall relationship,” Master’s Thesis, ITC, IIRS, 2006.
60. T. B. McKee, N. J. Doesken, and J. Kleist, “The relationship of drought frequency and duration to time scales,” Preprints, 8th Conference on Applied Climatology, 17-22 January, Anaheim, CA, pp.179-184, 1993.
61. T. R. McVicar, and P. N. Bierwirth, “Rapidly Assessing the 1997 Drought in Papua New Guinea Using Composite AVHRR Imagery,” International Journal of Remote Sensing, 22:2109–2128, 2001.
62. T. R. Karl and R. W. Knight, “Atlas of Monthly Palmer Hydrological Drought Indices (1931-1983) for the Contiguous United States,” Historical Climatology Series vol. 3-7, National Climatic Data Center, Asheville, NC, 1985.
63. T. Dinku, K. Asefa, K. Hilemariam, D. Grimes, and S. Connor, "Improving availability, access and use of climate information," in Weather, Climate and Water Bulletin, vol. 60, 2011.
64. T. Tadesse, J. F. Brown, and M. J. Hayes, “A New Approach for Predicting Drought-Related Vegetation Stress: Integrating Satellite, Climate, and Biophysical Data over the U.S. central Plains,” Journal of Photogrammetry and Remote Sensing, vol. 59, pp. 244-253, 2005.
65. T. Tadesse, 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, vol. 47, pp. In Press, 2010.
66. U. S. Panu and T. C. Sharmat, "Challenges in drought research: some perspectives and future directions," Hydrotogical Sciences-Journal—des Sciences Hydrologiques, vol. 47(S), pp. S19-S30, 2002.
67. USGS, "USGS-Earth Resources Observation and Science (EROS) Center-Elevation Data," vol. 2011, 2011.
68. W. C., Palmer, “Meteorological Drought,” Research Paper No. 45, United State. Department of Commerce Weather Bureau, Washington, D.C, 1965.
69. W. C. Palmer, Keeping track of crop moisture conditions, nationwide: The new crop moisture index, Weatherwise, vol. 21, pp. 156-161, 1968.

70. W. Zerihun, “Methods and Strategies for the Conservation of Forest Genetic Resources,” In: Proceedings of Forest Genetic Resources Conservation: Principles, Strategies and Actions. Addis Ababa: Support for Biodiversity Institute, German Technical Cooperation (GTZ), 1999.
71. Y. Julien, “Vegetation monitoring through retrieval of NDVI and LST time series from historical Databases,” PhD desertation, University of Valencia, 2008.
72. Z. Li, Q. Zhu, and C. Gold, “Digital terrain modeling: principles and methodology,” Boca Raton: CRC Press, 2005.

.

Declaration

I, the undersigned, declare that this thesis is my original work and has not been presented for a degree in any other university, and that all source of materials used for the thesis have been duly acknowledged.

Declared by:

Name: _____

Signature: _____

Date: _____

Confirmed by advisor:

Name: _____

Signature: _____

Date: _____

Place and date of submission: Addis Ababa, Ethiopia