ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

KNOWLEDGE DISCOVERY FROM SATELLITE IMAGES FOR DROUGHT MONITORING

GETACHEW BERHAN DEMISSE

A THESIS SUBMITTED TO IT DOCTORAL PROGRAM
ADDIS ABABA UNIVERSITY

PRESENTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN INFORMATION TECHNOLOGY (INFORMATION SYSTEMS)

ADDIS ABABA, ETHIOPIA
JUNE 2013

Knowledge Discovery from Satellite Images for Drought Monitoring

Getachew Berhan Demisse

Addis Ababa University
June 2013

Addis Ababa University

School of Graduate Studies

This is to certify that the thesis prepared by Getachew Berhan Demisse, entitled: Knowledge Discovery from Satellite Images for Drought Monitoring and submitted in fulfillment of the requirement for the Degree of Doctor of Philosophy in Information Technology (Information Systems) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

External Examiner: Dr. Ying Xie Signature _____ Date 22/08/2013

Internal Examiner: Dr. Fekade Getahun Signature _____ Date 22/08/2013

Principal Advisor: Dr. Tsegaye Tadesse Signature TSEGAYE TADESSE Date 22/08/2013

Co-Advisor: Dr. Solomon Atnafu Signature _____ Date 22/08/2013

Co-Advisor: Dr. Shawndra Hill Signature _____ Date 22/08/2013

Dr. Salehu Anteneh 16/10/2013

Chair of Department or Graduate Program Coordinator

Abstract

Knowledge Discovery from Satellite Images for Drought Monitoring

Getachew Berhan Demisse
Addis Ababa University, 2013

Drought is one of the most important challenges facing the planet. When it happens, it usually results in serious economic, environmental, and social crises. Despite the growing number of freely available biophysical, climate, and satellite data for characterizing and modeling drought, research efforts have been constrained to using only meteorological point data, such as the amount of rainfall, for drought monitoring information. This point data is insufficient for representing diversified ecosystems, and the data has coarse resolution levels (limited spatial coverage). Researchers also have limited tools for data retrieval and integration for improved drought identification and modeling, which usually results in a time delay for information to reach decision makers. Taking this into account, this dissertation research has three objectives: 1) identify the most relevant attributes for efficiently implementing drought monitoring, 2) develop a new approach for extracting knowledge from satellite imageries for improved identification and prediction of drought, and 3) evaluate the new approach for national and regional drought prediction applications. Using an exploratory research approach and modeling research method, different data collection and analysis techniques were executed using knowledge discovery in a database approach. The data mining models developed using artificial neural network and regression tree models were able to predict DroughtObject with accuracy of 0.70 – 0.95 correlation coefficients, in one- to four months' time lag. The developed DroughtObject model was evaluated for its application in showing drought severity and food deficit status. There were positive relationships between DroughtObject products and crop yield data up to 0.91 R^2 values. The results confirmed that the model can directly be used by those who are

currently responsible for drought monitoring and risk management. The new concept developed in this research was prototyped and demonstrated in an easy-to-use approach, with a focus on demonstrating the concept of DroughtObject characterization and identification from a group of pixels. This demonstration also revealed possible future system developments.

This dissertation research could help decision makers use advanced satellite technology for effective drought monitoring and early warning systems in various regions. Combined with proper policies, these systems can help to prevent famine and starvation in food-insecure regions. Up to now, satellite technologies have been used primarily in areas of meteorological applications. In this research, the main emphasis is on mining knowledge from satellite images for drought risk assessment and saving the lives of individuals who are affected by recurring droughts. The findings of this research can help decision makers take timely and appropriate actions to save lives in drought-affected areas using advanced satellite technology.

Keywords: Drought Monitoring; Drought Prediction; Geo-spatial Information; Knowledge Discovery; Satellite Image

Dedication

To my family and my parents

Acknowledgements

First of all I would like to thank the Almighty God, who helped me to succeed in this dissertation work with His power and love. He was with me through the ups and downs and helped me to achieve this success, and I give the glory to Him. The Bible says, "But we have this treasure in earthen vessels, that the excellency of the power may be of God, and not of us" (II CORINTHIANS 4: 7). Nothing is impossible to God!

I wish to thank the Ethiopian Government, specifically the Ministry of Education, for providing me sponsorship, and Addis Ababa University, IT Doctoral Program, for all financial, academic and technical support in pursuing my study.

I would like thank my principal supervisor, Dr. Tsegaye Tadesse, for his supervision, dedicated help and advice throughout this dissertation work. He has been a true inspiration throughout my research period. He taught me how to think critically and confidently in the research I was doing. He encouraged me to publish and work hard without losing patience.

I would like to thank my co-supervisor Dr. Solomon Atnafu for his professional guidance, from the planning of the research work up to the write-up. I appreciate his dedicated help on all the academic and administrative matters throughout the research work.

I would like to thank my co-supervisor Dr. Shawndra Hill for her technical advice, guidance and support during this PhD work. She taught me punctuality and how to work with deadlines. I will always remember our weekly Skype meetings and her dedicated help in shaping my research directions. She encouraged me to submit to the top journals and covered the cost of publishing an article.

I would like to express my great appreciation for my three supervisors, Dr. Tsegaye Tadesse, Dr. Solomon Atnafu and Dr. Shawndra Hill, for their coordinated work and dedication. I was really lucky to work with a team that was always ready to

help and work hard. They have helped me to expand the breadth of my thesis by providing me insights into their areas of expertise and research directions.

I would like to thank Dr. Dejene Ejigu and Dr. Salehu Anteneh for their administrative and academic support with everything from course work to proposal writing up to the final thesis. They really helped me in getting administrative support as well as the research idea for this dissertation work.

I would like thank the United States Geological Survey, specifically Dr. Gabriel Senay, Dr. James Vogelmann, Dr. Michael Hayes and Ann Fiedler for their dedication and help in sponsoring and facilitating me for my extended stay in the United States. My special thanks also go to the various U.S. government entities, especially the University of Nebraska-Lincoln (UNL), for helping with the logistics of my stay in the United States.

I would like to thank Karin Callahan and Chris Poulsen for their support, especially in the GIS and remote sensing data analysis phase of this dissertation work.

I would like to express my great appreciation and thanks to Deborah Wood for giving me her precious time from her extremely busy schedule for editing the final version of this thesis.

I express my deepest gratitude and thanks to Dr. Assefa Zebene, Hayal, Abere, Meskerem, Konjit, Zablon Adane, Kasahun, Zenebe, Abraham, Getachew, Eleni, Israel, Lala, Jalo and all fellowship members at UNL for their unlimited support during my stay in Nebraska. Without their support, my stay would have been impossible.

I am happy to express my great appreciation to the entire staff of Addis Ababa University and the University of Nebraska-Lincoln and the National Drought Mitigation Center, with whom I had the privilege of working during my studies.

My everlasting gratitude goes to my loving kids, family, relatives and friends, who always encouraged me and wished me success.

Table of Contents

List of Figures

List of Tables

Abbreviations

A.D.	Anno Domini – "in the year of our lord"
AA	Afro-alpine vegetation
ACB	Acacia-Commiphora woodland and bush land proper
ACB/RV	Acacia wooded grassland of the Rift Valley
AD	Anisotropic diffusion
AI	Artificial Intelligence
AIC	Akaike's Information Criterion
AMCIS	Americas Conference on Information Systems
AMO	Atlantic Multi-decadal Oscillation Index
AMSU	Advanced Microwave Sounding Unit
ANN	Artificial neural network
AOI	Area of interest
AVHRR	Advanced Very High Resolution Radiometer
B.C.	Before Christ
BMI	Bhalme–Mooley Index
CART	Classification and regression tree
CCD	Cold-cloud-duration
CNDVI	Cumulative NDVI
CSA	Central Statistical Authority
CSPRO	Census and Survey Processing System
CTW	Combretum-Terminalia woodland and wooded grassland
DAF	Dry evergreen Afro-Montane Forest and Grassland complex
DEM	Digital elevation model
DRMFSS	Disaster Risk Management and Food Security Sector
DSS	Desert and Semi-desert Scrubland
EB	Ericaceous Belt
EMA	Ethiopian Mapping Agency
ENSO	El Nino–Southern Oscillation
EROS	Earth Resources Observation and Science
ES	Expert system
ESRI	Environmental Systems Resource Institute
FEWS NET	Famine Early Warning Systems Network
FLV/MFS	Freshwater marshes and swamps, Floodplains and Lake Shore vegetation
FLV/OW	Freshwater lakes - open water vegetation
FNEA	Fractal net evolution approach
GIS	Geographic information systems
GO	Governmental organizations
GPP	Gross primary productivity
GTS	Global Telecommunication System
GUI	Graphical user interface
GWR	Geographically weighted regression
ITCZ	Inter-tropical convergence zone
KDD	Knowledge discovery in database
LST	Land Surface Temperature
MAF	Moist Evergreen Afro-Montane Forest
MEI	Multivariate ENSO Index

MOWCATL	Minimal Occurrences With Constraints and Time Lags
NAO	North Atlantic Oscillation
NASA	National Aeronautic and Space Administration
NDVI	Normalized Difference Vegetation Index
NDVI_ib	Base NDVI
NGO	Non-governmental organizations
NIR	Near-infrared
NMA	National Meteorological Agency
NMSA	National Meteorological Services Agency
NOAA	National Oceanic and Atmospheric Administration
OLS	Ordinary Least Square
PDO	Pacific Decadal Oscillation
PDSI	Palmer drought severity index
PNA	Pacific North American Index
REAR	Representative Episodal Association Rule
REST	Representational State Transfer
RF	Rainfall
RFE	Rainfall estimate
SDNDVI	Standardized deviation of Normalized Difference Vegetation Index
SLV/OW	Salt Lake open water vegetation
SLV/SSS	Salt Pans, Saline/brackish and Intermittent wetlands and Salt-lake Shore Vegetation
SPI	Standard precipitation index
SSM/I	Special Sensor Microwave/Imager
SST	Sea surface temperatures
SVI	Standardized vegetation index
SWSI	Surface water supply index
TAMSAT	Tropical Applications of Meteorology using Satellite
TRF	Transitional Rain Forest
UML	Unified modeling language
UNEP	United Nation Environmental Program
USGS	United States Geological Survey
veg_ethiopia	Ecosystem type of Ethiopia
VegDRI	Vegetation Drought Response Index
VegOut	Vegetation Outlook
VIF	Variance Inflation Factor
VIs	Vegetation indices
VTCI	Vegetation Temperature Condition Index
WFP	World Food Programme
WGG	Wooded grassland of the Western Gambela region
WHC	Water holding capacity

Chapter 1 Introduction

1.1 Knowledge Discovery in Database

Knowledge is power [1] and very important for solving human challenges [1, 2]. Nonaka [3] defined knowledge as justified true belief, and it is created and organized by the flow of information. Useful knowledge for solving human challenges is produced from the collection of useful information, which occurs in the process of interaction between people sharing their experiences and information sources [4]. In this process, information is a flow of messages, while knowledge is created and organized by the flow of information, anchored on the commitment and beliefs of its holder [3]. This understanding emphasizes an essential aspect of knowledge that relates to human action [3, 4].

Nonaka [3] described two types of knowledge: explicit and tacit. Explicit knowledge is a type of knowledge that is transmittable in formal, systematic language. Tacit knowledge has a personal quality, which makes it hard to formalize and communicate, and it is deeply rooted in action, commitment, and involvement in a specific context.

This research is concerned with explicit knowledge [5], specifically with the patterns observed from data, where this pattern can easily be understood by humans and validated by test data with some degree of certainty [6]. The practice of finding useful patterns in data has been given a variety of names, including data mining [6], knowledge extraction[7], information discovery [8], information harvesting [7, 9], data archaeology [8], and data pattern processing [7, 9, 10]. All of these terms describe a robust computational technique known as Knowledge Discovery in Database (KDD) [9]. Fayyad et al. [7] defined KDD as the overall process of discovering useful knowledge from data, including data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge, and proper interpretation of the results of mining.

The KDD process has been described in several articles [7, 11, 12]. From these articles, it is possible to divide the KDD process into four major steps: developing and understanding the background domain, data preprocessing, data mining, and knowledge construction. Background analysis is usually done to get an understanding of the application domain. Data pre-processing includes data selection, data cleaning, and data reduction. Data mining includes choosing the data mining task and technique and applying the technique to search for interesting patterns. Knowledge construction includes interpreting the mined patterns, often through visualization, and consolidating the discovered knowledge, either by incorporating the knowledge into a computational system or through documenting the knowledge and reporting it to interested parties [10]. Data mining, which is the application of specific algorithms for extracting patterns (models) from data, is an important step in the KDD process [7].

Data mining is an interdisciplinary field, combining artificial intelligence, computer science, machine learning, database management, data visualization, mathematical algorithms, and statistics [1]. It is a tool and technology for KDD, which provides different methodologies for decision making, problem solving, analysis, integration, learning and innovation [1, 9]. For example, scientific understanding of complex geographic problems often depends on the discovery, interpretation, and presentation of the spatial patterns [13]. Guo et al. [13] indicated that identifying such patterns becomes ever more challenging as powerful data collection and distribution techniques produce geographic datasets of unprecedented size in many application and research areas. These datasets are not only large in data volume but also characterized by a high number of attributes or dimensions [14]. A spatial dataset consists of two major components: a spatial location (longitude and latitude) and a set of variables (descriptive data about a location) [13]. These datasets have to exist to be considered as spatial dataset. In such a context, it is extremely challenging to effectively and efficiently detect and understand relationships and patterns in voluminous and high-dimensional data [10]. For such scenarios, data mining is a tool for distilling data into information or facts implied by the data [1, 15-19]. In this study, for characterizing drought from satellite images and

developing a model for discovering knowledge, recent developments in data mining technologies were used.

1.2 Drought Modeling

Drought is a recurring extreme climate event characterized by below-normal precipitation over a period of months to years, and it is a temporary dry period, in contrast to the permanent aridity in arid areas [20]. Drought is manifested in erratic and uncertain rainfall distribution in rainfall-dependent areas. Frequent and severe droughts have become one of the most important natural disasters. In sub-Saharan Africa, for example, drought is causing serious economic, social and environmental crises [21]. Its effect is manifested by creating uncertainties in agricultural economies [22].

Modeling drought is crucial for its management and predictions. In this study, modeling is the process of discretizing the geographic phenomenon [23] of drought in a space and time dimension [24-26]. The onset, duration and severity of droughts are often difficult to determine, and their characteristics may vary significantly from one region to another [27], which makes the modeling task challenging. One of the most important attributes for modeling drought is rainfall records [28, 29].

Africa has a long history of rainfall fluctuations of varying lengths and intensities [30]. Past studies showed differences in behavior of rainfall trends in Africa, at different spatial and temporal scales. These studies also demonstrated a decrease in rainfall in eastern Africa between 2003 and 2008 [31] where drought and famine situations were periodically reported [32]. The causes of these droughts usually remain unknown, which makes modeling and information system development challenging.

Many studies have shown that the recent Sahel droughts resulted primarily from a southward shift of the warmest sea surface temperatures (SSTs) and associated Inter-Tropical Convergence Zone (ITCZ) [33], and the steady warming in the Indian Ocean [20, 34]. ITCZ is the area encircling the earth near the equator where winds originating in the northern and southern hemispheres come together; it is

formed by vertical motion largely appearing as convective activity of thunderstorms driven by solar heating [35]. Other studies [36, 37] also indicated that reduced vegetation cover and surface evaporation may have provided a positive feedback that enhanced and prolonged drought.

Drought has a particularly negative impact on agricultural production in the eastern African region because most of the area's agriculture is dependent on rainfall [38], thus underscoring the need for simple and improved drought monitoring and early warning information systems. The conventional approach to drought monitoring and early warning systems using ground-based data collection is tedious, time consuming and difficult [39]. In recent years, remote sensing data has increasingly been used for monitoring agro-climatic conditions, the state of the agricultural fields, and vegetation cover and for estimating crop yield in various countries. For example, the Advanced Very High Resolution Radiometer (AVHRR) Normalized Difference Vegetation Index (NDVI) data has been used in vegetation monitoring, crop yield assessment and forecasting [40-44]. The National Oceanic and Atmospheric Administration (NOAA) AVHRR series satellite data provide a long-term record of NDVI data that can be used in the prediction of crop yields [39]. Prediction of crop yields is an important part of drought monitoring systems, as it is one of the most important factors for ensuring the food security of a given region.

For drought and food security monitoring, the Famine Early Warning System Network (FEWS NET) uses data from two operational remote sensing products on a near-real time and spatially continuous basis [45]. These data include dekadal (ten days) AVHRR NDVI images produced by the National Aeronautic and Space Administration (NASA) and rainfall estimates (RFE) produced by the Climate Prediction Center of NOAA. For the African continent, blended satellite RFE images are produced by NOAA at 10 km spatial resolution [46].

Currently there are more than ten satellite rainfall products at different spatial and temporal resolutions [47, 48]. The RFE data for this research was obtained from Tropical Applications of Meteorology using Satellite (TAMSAT), which is produced by the TAMSAT group at Reading University in the United Kingdom

[48]. For the present research, the three-month Standard Precipitation Index (SPI) was calculated from this RFE data.

The three-month SPI of a given month uses the precipitation total of the previous two months and also the precipitation total of the specific month of interest. For example, the SPI of September uses the precipitation total of July, August and September itself. The three-month SPI is selected in this research because it gives optimum estimation of short-term drought condition information and provides a seasonal estimation of precipitation [49].

In this research, in addition to the three-month SPI, which was obtained from RFE data, other biophysical, climate and satellite data were used for characterizing and modeling drought. These datasets were combined using KDD computational techniques. Using this computational technique, the state of the art of drought identification, modeling and predictions in one to four months' time periods were demonstrated. Some of the results were also published in reputable journals [50, 51], a book chapter [52] and conference proceedings [53, 54].

1.3 Motivation

Because of climate change and uncertainties in future weather conditions, drought information extraction research has gotten the attention of both politicians and scientists [21]. The motivation for the research reported here was threefold: 1) to contribute to drought risk-management efforts to monitor drought, which is one of the top challenges to society, 2) to exploit the huge dataset from biophysical, climate and satellite sources for drought monitoring research and risk-management, and 3) to make this research applicable to real-life problem solving in Ethiopia, where recurrence of drought is common.

The knowledge discovery and information extraction analysis in this research was tested on real-life problem-drought monitoring. Most regions of the world are affected by drought from time to time. Drought has had and will continue to have some of the greatest impacts on climate. Drought has brought ancient civilizations to their knees and led to famine, food scarcity, and loss of lives and property [55].

Evidence suggests that the areas affected by drought are increasing globally [56]. Many climate predictions anticipate a further increase in drought-affected regions in the future [55]. Therefore, designing drought monitoring information is one of the most important issues that needs to be researched for the world community [57, 58], especially sub-Saharan African countries such as Ethiopia [21, 59]. Huge datasets from biophysical, climate and satellite observations are extremely useful for this problem solving research.

Despite the enormous economic, environmental, and societal impacts of drought, there has never been a coordinated and integrated drought information system. Given the importance of this extreme form of climate variability, it is critical that drought should be studied appropriately and its information system designed properly, with the hope of better understanding its nature and anticipating its occurrence on short and long time scales [55].

Since drought is closely related to climate change [20, 60], its behavior is unexpected. This makes designing information systems for drought monitoring challenging [57]. Fortunately, in this research, KDD can see the dimensions of both the expected and unexpected aspect of the attribute under investigation [7, 10], such as drought prediction in space-time dimensions.

The main purpose of this study was to design a new information system by integrating the outputs from different research efforts using KDD, the new computational technique, for the purpose of reducing the impact of drought on the environment and people. It is also hypothesized that it is possible to identify, track and model drought as a spatial object for its improved prediction and management. DroughtObject here is the geographic phenomenon of drought characterized by a group of pixels, where these pixels can be segmented into separate regions with defined spatial location and attributes [61]. This concept is demonstrated in Figure 1.1. The assumption in this research was that it is possible to identify, model and predict drought incidence by combining its key attributes with KDD.

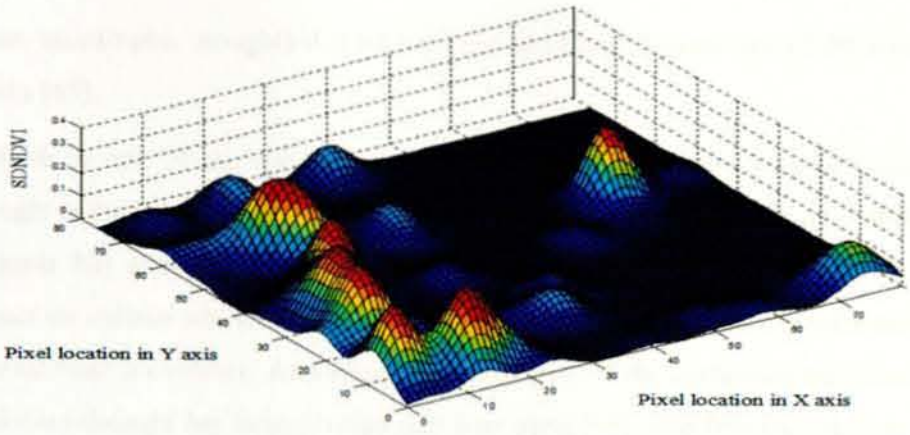


Figure 1.1: Hypothesized graphical representation of DroughtObject identification and modeling using 3D representation. The graph surfaces represent relatively low SDNDVI deviation in the blue areas and the highest deviations in the deep red areas. High deviation from historical mean means the occurrence or incidence of drought spatially.

1.4 Problem Statement

Drought is one of the most important challenges to humans [20]. When it happens, it usually results in serious economic, environmental, and social crises [21, 55]. In the past, drought was perceived as an extreme event in climatic systems, whereas in reality it needs to be recognized as a normal occurrence [57], and its recurrence is inevitable [55]. Drought has occurred many times in different parts of the world and will continue to occur, and its adverse consequences are likely to increase because of global warming and climatic change [20].

Between 1967 and 1992, droughts affected 50% of the 2.8 billion people who suffered from all natural disasters [62]. In the past, out of a total of 3.5 million people killed by disasters, 1.3 million were lost because of direct and indirect impacts of droughts [62, 63]. Nearly 50% of the world's most populated areas are highly vulnerable to drought [22]. More importantly, almost all of the major agricultural lands are located in these areas [64].

A report compiled by the United Nations Environmental Program (UNEP) and World Bank indicated that about 180 million people in Africa live in drought-prone areas, and 50 million people are threatened with starvation in case of rain failure [22]. This report further indicated that from 1965 to 1999, a total of 330 drought incidences were reported in Africa and caused about 880,000 deaths. In 1984/1985

example, in some districts there are up to three meteorological stations whereas in other districts there is only one station. Limited tools for data retrieval and integration for improved drought identification and modeling have often resulted in delayed delivery of information to decision makers.

In the past, computational techniques such as KDD were underutilized for integrating the different data sources in drought risk-management applications. Past efforts to use the data mining aspect of KDD on drought-related research include Harms et al. [67, 68], Tadesse et al. [69-71], and Brown et al. [72]. KDD, the new computational technique, has interesting data integration and knowledge delivery methods. Unfortunately, past research in using KDD for drought monitoring applications was very limited. In the present study, attempts have been made to demonstrate the state of the art of KDD application for improved drought monitoring information system design and development.

1.5 Research Questions and Objectives

To address the above-stated drought monitoring challenges, drought identification and characterization from satellite images were examined using an object-oriented approach. Here the assumption is that the episode or spatial extent of drought depends on different parameters or attributes (e.g., amount of rainfall, land cover, elevation, etc.). These relevant attributes can be included in the analysis to improve identification and characterization of drought. After including the most relevant attributes, the basic processing units of image analysis are image segments (groups of pixels characterizing drought), not single pixel [23, 61]. Based on this context, this study has the following research questions:

1. What are the most relevant attributes for drought monitoring?
2. How is drought modeled as spatial object for improved identification and prediction?
3. How is DroughtObject identified on satellite imagery evaluated using ground observation data?

To address these questions, the main objective of this research was to investigate and develop a new spatial object modeling approach for detecting and predicting drought using satellite images. In line with this, the specific objectives were to:

1. identify the most relevant attributes for efficiently implementing drought monitoring,
2. develop a new approach for extracting knowledge from satellite imageries for improved identification and prediction of drought, and
3. evaluate the new approach for national and regional drought prediction applications.

Advanced satellite technology products with high temporal resolution are cost effective and can serve to detect the onset of a drought, its duration and magnitude. Such information can help decision makers to take appropriate actions in a timely manner, as well as reduce the impact of drought and mitigate its adverse effects on the environment and the society.

1.6 Research Framework

There are "two research paradigms in information systems (IS): behavioral science and design science paradigms" [73]. As described by Hevner et al. [73], the behavioral science paradigm seeks to develop and verify theories that explain or predict human or organizational behavior, whereas the design science paradigm extends the boundaries of human and organizational capabilities by creating new and innovative artifacts. From these two IS research paradigms, this PhD study used the design science research paradigm and framework. The research framework was chosen because of the nature of this research, which aims at developing a new approach for extracting knowledge from satellite imageries for improved identification and prediction of drought. Gregor [74] indicated that "design theory" gives explicit prescriptions for constructing an artifact and mainly answers the question of how to do something.

Design science inherently is a problem solving process [73, 75]. In its problem solving process, Hevner et al. [73] presented seven guidelines on how to conduct design science research: design as artifact, problem relevance, design evaluation, research contribution, research rigor, design as a search process, and communication of research. These guidelines have been adapted for this research, and the seven steps have been condensed into four: identification (which is design as artifact), modeling (which includes problem relevance, design evaluation, research contribution, and research rigor), prediction (which includes design as a research process) and communication with stakeholders. The artifact for the process of knowledge discovery from satellite imageries is presented in Figure 1.2. Detailed discussions of the steps are presented in the following subsections.

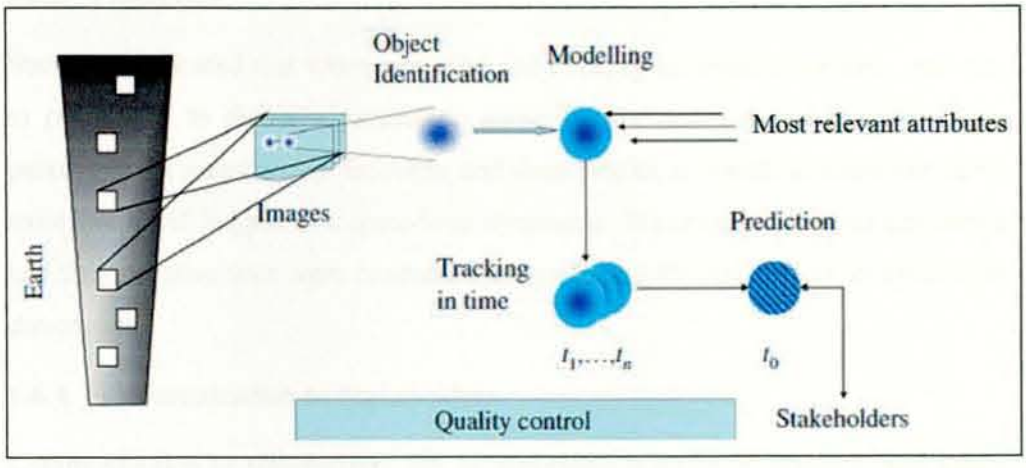


Figure 1.2: An artifact for the process of knowledge discovery from satellite imageries (adopted from Stein et al. [25]).

1.6.1 Identification

Identification of DroughtObject requires the conversion of raster pixel values to objects. In remote sensing this means a segmentation of imagery followed by classification of the individual imagery pixels. Typically, identification is done by applying a segmentation routine, in which both the object and the uncertainty are modeled [24]. The basic assumption here is that drought pixels have similar values, which are different from non-drought pixel values in the vicinity.

1.6.2 Modeling

The modeling steps developed by Jakeman et al. [76] were followed in this research. According to their recommendation, there are about ten steps for developing a scientific model: define model purpose, specify modeling context, conceptualize and specify data, select model features, determine model structure and parameter values, choose performance criteria and algorithm, identify model structure and parameter values, verifications, quantification of uncertainty, and model evaluation [24, 76]. These steps were followed in our DroughtObject modeling process.

1.6.3 Prediction

Stein [24] indicated that when modeling and tracking an object over time, one way to proceed is to define a parametric curve for the centroid and possibly other parameters of membership functions and then predict at which location the curve most likely will happen in a space-time dimension. These functions were developed and then iterative tests were conducted for predicting DroughtObject in space-time dimensions.

1.6.4 Communication to Stakeholders

Communication to stakeholders can be expressed in different ways; it may range from visualization to assessment of cost and benefit of the research output [24]. Hevner et al. [73] indicated that design-science research must be presented both to technology-oriented as well as management-oriented audiences. Technology-oriented audiences need sufficient detail to enable the described artifact to be implemented and used within an appropriate organizational context. This enables practitioners, such as decision makers, to take advantage of the benefits offered by the artifact, and it enables researchers to build a cumulative knowledge base for further extension and evaluation. It is also important for such audiences to understand the processes by which the artifact was constructed and evaluated. This establishes repeatability of the research project and builds the knowledge base for further research by design-science researchers in information systems. The outputs

of this research were presented using cartographic technology or maps so that decision makers can make intelligent decisions using recent satellite technology products.

In this research, the hypothesis is that it is possible to model, monitor and predict drought by integrating satellite and ground-based data. The conceptual frame of this hypothesis is presented in Figure 1.3. This figure presents the conceptual framework for understanding, executing and evaluating the new object-oriented information extraction process using a design science paradigm in IS. The information obtained using this framework can help decision makers in setting an action plan that is likely to save drought victims in food-insecure areas.

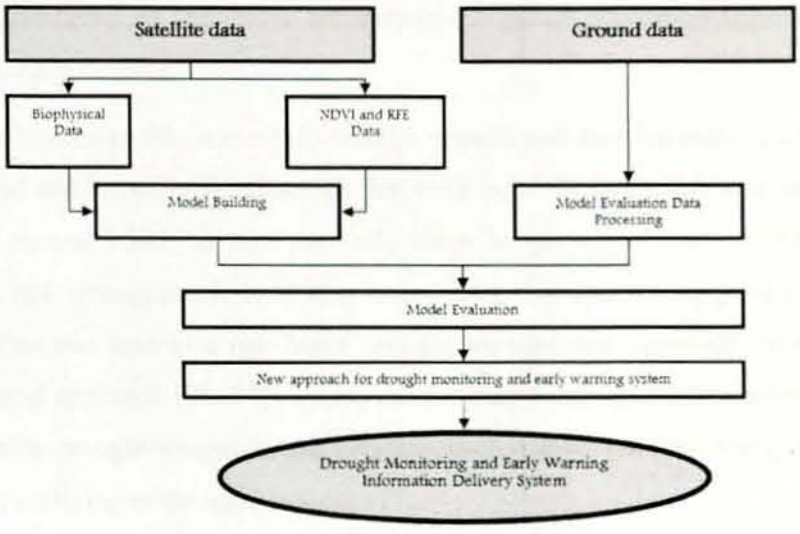


Figure 1.3: Flow chart showing the major steps in knowledge discovery from satellite images for near real time drought monitoring and early warning systems. The last arrow in this figure is a broken line, because the information delivery system will be addressed in future research.

1.7 Research Contribution

The contribution of this research will be the applicability of the method developed for drought monitoring information systems, and also the extension of the state-of-the-art KDD problem solving approach. These two contributions are described in this subsection.

KDD, which is the nontrivial process of identifying valid, novel, potentially useful data [7], has an enabling role in managing and understanding patterns in drought risk-management. From the actual applicability of the method for drought monitoring information systems, this research develops concept implementation algorithms that enable drought monitoring institutions to improve drought risk management. These include:

1. integrating different attributes from biophysical, climate and satellite data sources,
2. extracting relevant data/information from drought episodes for near real time decision making and/or prediction,
3. developing an applicable and easy-to-use model evaluation approach using crop-yield data.

The developed algorithms can help drought experts and decision makers understand the spatial and temporal relationships that exist between DroughtObject and global climate events, which in turn can help them to get up-to-date information for drought risk management. It is also hoped that this kind of frequently updated information can lead to a risk-based drought management approach rather than a crisis-based approach [21, 57]. This kind of drought management eventually leads to proactive drought mitigation and gives decision makers the opportunity to relieve the most suffering at the least expense [77].

In extending the state of the art of KDD, this research contributes in two major aspects: 1) KDD is extended in the drought monitoring domain for practical problem solving, and 2) KDD in this research is used in spatial context, whereas past efforts were focused on non-spatial data or non-spatial context [10].

In implementing the steps in the KDD process in the drought monitoring domain, all the steps, namely data selection, pre-processing, transformation, and data mining using geo-information data [10], were demonstrated. Data mining in this case is only one step of the KDD process. It involves the application of techniques for distilling data into information or facts implied by the drought data. KDD in this research is the higher level process of obtaining facts through data mining and

distilling this information into knowledge to be used by decision makers at different levels. This generally requires a human-level intelligence to guide the process and interpret the results based on pre-existing knowledge [10]. The data miner, as indicated by Gahegan et al. [78], is the critical interface between the syntactic knowledge or patterns generated by machines and the semantic knowledge required by humans for reasoning about the real world. In this case, a group of pixels was identified and segmented from the input image on the real ground. Here, the KDD process does not seek any arbitrary pattern from a database; rather, it seeks only those that are interesting. These patterns are valid (a generalizable pattern, not simply a data anomaly), novel (unexpected), useful (relevant) and understandable (can be interpreted and distilled into knowledge) [7]. These interesting concepts are practically demonstrated in the present research using drought as a domain.

The second important aspect, showing the practical applicability of KDD in spatial context, means that DroughtObject has two important components, spatial and non-spatial [13, 23] attributes (e.g., drought severity status). The spatial component of DroughtObject is the spatial location (longitude and latitude) or the geometric aspect of the data, which is usually challenging to represent and manage in the database. Although the spatial aspect is challenging from data management perspectives, it can add value to the data and make it easier for decision makers to visualize and implement the information [10].

For drought monitoring applications of KDD, non-spatial data is not sufficient. The reason for this is that drought data by its nature has a spatial location on earth. Shekhar et al. [79] indicated that in the KDD analysis, if the object of interest has a spatial component, standard KDD techniques with non-spatial data are not sufficient. Specific reasons include the nature of geographic space, the complexity of spatial objects and relationships as well as their transformations over time, the heterogeneous and sometimes ill-structured nature of geo-referenced data, and the nature of geographic knowledge. Taking this fact into account, the KDD analysis in this research was done using spatial data for practically visualizing and managing drought risk-management. In the past there were limited efforts to use spatial data in KDD analyses [10].

In general, the overall research output is expected to contribute to improving drought monitoring information systems and preventing social and economic crises in drought-prone areas. The method developed here can be directly implemented in drought monitoring related risk management and mitigation efforts. The concepts developed in this research were also prototyped as a startup for future full system development.

1.8 Scope of the Study

This study is delimited by research approach, study area and dataset inputs. The most usual way to classify the research approach is to distinguish the underlying philosophical assumptions guiding the overall research. There are three different philosophical assumptions: positivist, interpretive and critical research. According to Myers [80], a positivist assumption is that reality is objectively given, and it can be described by reference to the measurable properties that are independent of the researcher. An interpretive paradigm, on the other hand, considers that reality can only be accessed through social constructions, such as shared meanings. Critical research makes the assumption that social reality is historically constituted and that people are constrained in their actions by various forms of cultural and political domains. From these three different paradigms, this PhD research is delimited by the positivist philosophical paradigm, with the assumption that reality can be described objectively by reference to the measurable properties of the geographic phenomenon of drought. Furthermore, in this philosophical paradigm, the design science research framework was used for explicitly modeling drought episodes in space time dimensions.

In the context of the study site, the KDD drought modeling experiment was executed in the ecosystems of Ethiopia. With coarse generalization, Ethiopia has about nine different ecosystems (using vegetation type as a classification parameter) [81]. This was decided because of the unavailability of ecosystem data representing the diversified ecosystem types of Ethiopia. The result of this modeling experiment can be generalized with optimum accuracy to the study area,

and it was also assumed that the results may be generalized and applied to other similar ecosystems in other parts of the world.

The dataset used for the modeling experiment has a spatial resolution of 8km, which is a coarse resolution for capturing the diversified ecosystems of the country. The 8km spatial resolution imagery data was used because the major NDVI data for modeling drought was produced by NOAA with this resolution standard to the global application. To fit this standard, all the data inputs gathered from different sources were downgraded to this resolution for fitting the model implementation experiment in Mapcubist software.

The drought monitoring experiment was done only for the growing season of June to October. An assessment of the overall growing seasons of Ethiopia revealed about four different growing seasons in the country. The current research focused on the June to October growing season, since this season covers the highest spatial extent of Ethiopia. It was also assumed that the modeling and information system design for the remaining three growing seasons can be developed using the approaches developed in the June to October growing season. The modeling experiment also included only biophysical, climate and satellite datasets and does not include any other socio-economic and infrastructure factors. Therefore, the research was focused on the use of geo-spatial information for drought monitoring for improved drought mitigation.

Drought monitoring in this research is in the context of early awareness and prediction of drought in one- to four-month periods. Research on predictions was carried out on a monthly period of time, since that is the timeframe in which the information is most helpful to decision makers for their appropriate risk management and mitigation activities. Therefore, the highest temporal resolution of the information produced is one month.

This research is also limited to presenting the concept for identifying, modeling and predicting drought using the KDD approach. In this process, the different data mining tools available were not compared for their accuracy level for distilling the pattern for the actual knowledge construction.

1.9 Thesis Outline

This thesis is presented in monograph style with the order of introduction, review of literature, related works, methodology, results and discussions, concept implementation and future works. A preceding chapter is also assumed to be a prerequisite of the subsequent chapter and serve as a building block of the whole thesis. Therefore, one chapter may not be complete if the preceding chapter is missing. Accordingly, the seven chapters are presented in the following order.

Chapter 1 presents the introduction and provides a brief background of KDD and the approaches to characterizing drought as spatial object, as well as introducing our new approach to identify, model and predict drought as spatial object. This chapter also includes the research questions, objectives and outline of the thesis.

In Chapter 2, a detailed review of literature relevant for drought and drought monitoring information systems is presented. Chapter 3 presents related works with the present research, the state of the art of drought monitoring, explores the different methods applied, and identifies the research gap. In Chapter 4, the detailed methodology of the dissertation is presented for the use of satellite imageries in detecting and identifying drought as spatial object.

Chapter 5 presents the results of the research and puts the thesis into context by presenting the actual identification, modeling and predicting of DroughtObjects. The results obtained are also discussed, and the chapter provides a necessary preamble to the investigations of how to model drought in the study areas. The model evaluation results are also discussed in this chapter.

In Chapter 6 (implementation and discussion), the new concepts tested and discussed in the preceding chapters are practically implemented. The concepts discussed are implemented using Matlab software. The implementation codes are demonstrated with real world data and with the focus of demonstrating the concept developed in the thesis.

Finally, Chapter 7 presents conclusions and future works, including a synthesis of the findings of the thesis and a discussion of the implications of the results for

modeling drought as spatial object using satellite technologies. The implications and overall outcomes of the research for the decision makers are also discussed in this chapter.

Chapter 2 Literature Review

The literature review in this chapter highlights the state of the art of design science research in information systems and drought monitoring. In the first subsection, the review presents the nature of design science in information systems. The second subsection presents definition and types of drought, and sets the boundary of this dissertation research. In the third subsection, the concept of DroughtObject modeling is presented. In subsection four, object-oriented image analyses with the selected methods are discussed.

2.1 Design Science Research in Information Systems

Information systems are implemented in organizations to improve the efficiency and effectiveness of organizations' objectives [73, 75, 82-84]. In line with this, the role of research in information systems is to acquire knowledge and understanding that enables the development and implementation of technology-based solutions to previously unsolved and important business problems [73]. Acquiring such types of knowledge involves two complementary but distinct research paradigms, behavioral science and design science [85]. Hevner [73] noted that the behavioral-science paradigm has its roots in natural science research methods. It seeks to develop and justify theories (i.e., principles and laws) that explain or predict organizational and human phenomena surrounding the analysis, design, implementation, management, and use of information systems. On the other hand, the design-science paradigm has its roots in engineering and the sciences of the artificial [86].

Differentiating behavioral science and design science research, Hevner et al. [73] noted that behavioral science addresses research through the development and justification of theories that explain or predict phenomena related to the identified business need; and design science addresses research through the building and evaluation of artifacts designed to meet the identified business need. The goal of

behavioral science research is truth and the goal of design science research is utility. Within this context, the design-science research paradigm is proactive with respect to technology, and it focuses on creating and evaluating innovative IT artifacts that enable organizations to address important information-related tasks. The behavioral science research paradigm is reactive with respect to technology in the sense that it takes technology as given. Hevner et al. [73] further noted that one issue that must be addressed in design science research is differentiating routine design or system building from design research. The difference is in the nature of the problems and solutions. Routine design is the application of existing knowledge to organizational problems, such as constructing a financial or marketing information system using best practice artifacts (constructs, models, methods, and instantiations) existing in the knowledge base. On the other hand, design-science research addresses important unsolved problems in unique or innovative ways or solved problems in more effective or efficient ways. The key differentiator between routine design and design research is the clear identification of a contribution to the archival knowledge base of foundations and methodologies.

Design-science basically is a problem solving paradigm and seeks to create innovations that define the ideas, practices, technical capabilities, and products through which the analysis, design, implementation, management, and use of information systems can be effectively and efficiently accomplished [73, 87].

Walls et al. [83] indicated that design is both a process (set of activities) and a product (artifact) and it describes the world as acted upon (processes) and the world as sensed (artifacts). The design process here is a sequence of expert activities that produces an innovative product (i.e., the design artifact) [73]. There are two design processes and four design artifacts produced by design-science research in IS [85]. The two processes are build and evaluate, and the artifacts are constructs, models, methods, and instantiations [73, 85].

Hevner et al. [73] indicated that purposeful artifacts are built to address heretofore unsolved problems and they are evaluated with respect to the utility provided in solving those problems. Hevner et al. [73] and March and Storey [75] described

that constructs provide the language in which problems and solutions are defined and communicated. Models use constructs to represent a real world situation or the design problem and its solution space. Models aid problem and solution understanding and frequently represent the connection between problem and solution components enabling exploration of the effects of design decisions and changes in the real world. Methods define processes. They provide guidance on how to solve problems, that is, how to search the solution space. These can range from formal, mathematical algorithms that explicitly define the search process to informal, textual descriptions of best practice approaches, or some combination. Instantiations show that constructs, models, or methods can be implemented in a working system. They demonstrate feasibility, enabling concrete assessment of an artifact's suitability to its intended purpose. They also enable researchers to learn about the real world, how the artifact affects it, and how users appropriate it.

The innovative or basic contribution of this PhD study is that it recognized drought as spatial object and accordingly modeled drought using its most relevant attributes. The concept of modeling drought using an object-oriented approach was proposed by Rulinda et al. [27], and it is a new approach for characterizing and modeling drought. For this purpose, the artifacts (constructs, models, methods, and instantiations) were developed for improved drought monitoring. The concept of object identification from remote sensing data was also demonstrated by past research, such as Stein [24], Stein et al. [25] and Benz et al. [61]. These concepts were used in the current study for designing a drought monitoring information system (from data selection up to knowledge construction) using design processes and artifacts principles from a design science research approach.

2.2 Definition and Types of Drought

Drought is an adverse environmental phenomenon that influences almost all aspects of society [57]. It is a normal feature of climate and its occurrence is inevitable [77, 88]. Drought is defined as "the naturally occurring phenomenon that exists when precipitation has been significantly below normal recorded levels, causing serious hydrological imbalances that adversely affect land resource production systems"

[20, 89]. Drought is also defined as a prolonged abnormally dry period when there is not enough water for users' normal needs, resulting in extensive damage to crops and loss of yields [55]. These definitions of drought are conceptual definitions and are the basis for the operational definition. The operational definition of drought focuses on identifying the beginning, end, spatial extent and severity of the drought in a given area, and it is based on scientific reasoning [20]. The analysis is done using hydro-meteorological information and is beneficial in developing drought policies, early warning monitoring systems, mitigation strategies and preparedness plans [90].

There are three types of drought, namely meteorological drought, agricultural drought and hydrological drought [20, 91]. Meteorological drought is a period of months to years with below-normal precipitation, and it is often accompanied by above-normal temperatures [20]. Agricultural drought is a period with dry soils that results from below-average precipitation, intense but less frequent rain events or above-normal evaporation, all of which lead to reduced crop production and plant growth [20, 91]. While meteorological drought is usually defined by a precipitation deficiency over a pre-determined period of time; agricultural drought is defined more commonly by the lack of availability of soil water to support crop and forage growth than by the departure of normal precipitation over some specified period of time [91]. Hydrological drought is normally defined by deficiencies in surface and subsurface water supplies relative to average conditions at various points in time through the seasons [20]. These three types of drought gradually contribute to socioeconomic drought, which involves an imbalance between supply and demand for economic goods such as water, livestock forage, hydroelectric power, and others that are dependent on precipitation [91]. The three drought types are diagrammatically demonstrated and presented in Figure 2.1a. To explicitly set the boundary, the focus of this research is on agricultural drought analysis and early warning information systems.

In the modeling process, drought is characterized by three main aspects: intensity, duration, and spatial coverage [20]. Dai [20] noted that intensity is the degree of the precipitation, soil moisture, or water storage deficit; it may include consideration of

the severity of the associated impacts. Drought typically lasts for several months to a few years, but extreme drought can persist for several years, or even decades for so-called mega-drought [92]. The frequencies of agricultural drought in Ethiopia [66] are presented in Figure 2.1(b and c). These maps show the general pattern of drought frequencies in the country. The frequency values show the increase of drought frequency in the contour values as one moves west to east and south to north.

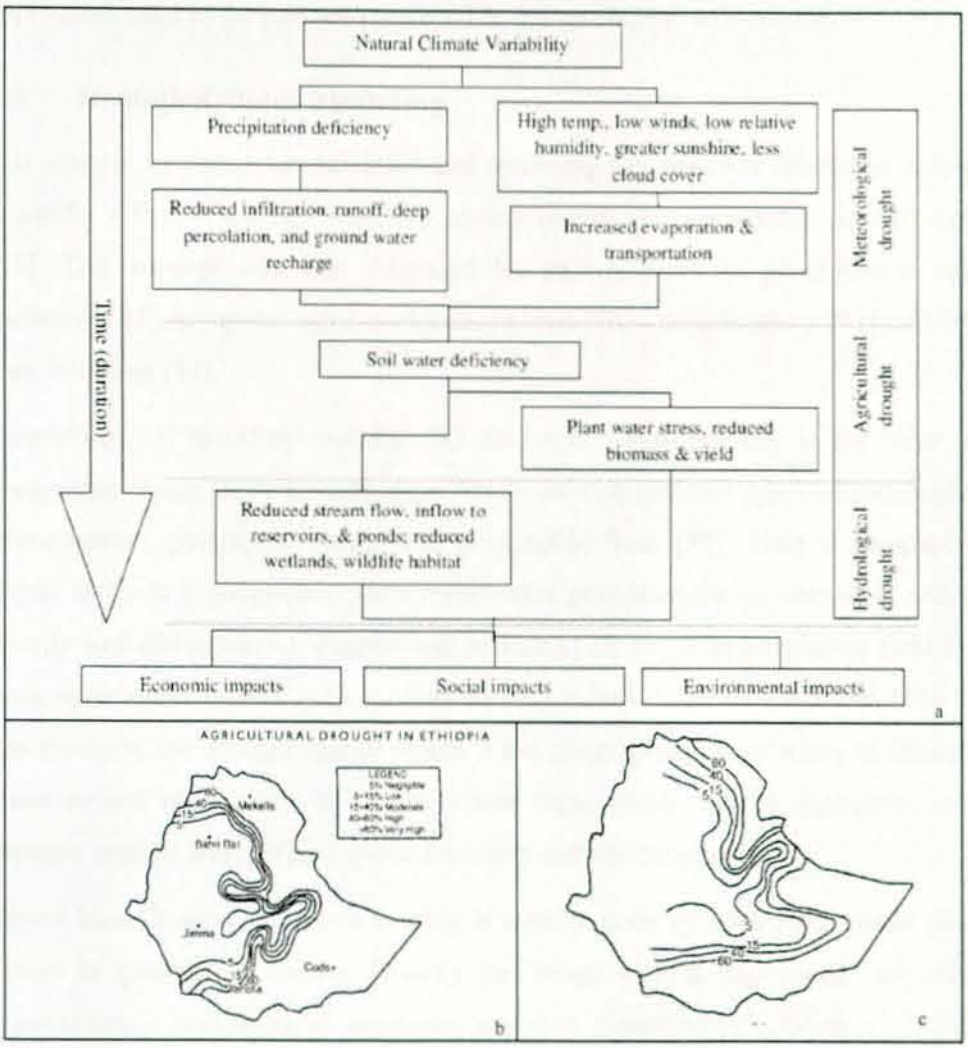


Figure 2.1: Diagrammatic representation of (a) drought classification (adapted from UNISDR [91], (b) frequency of agricultural droughts in Ethiopia during the first rainy seasons (February – May), and (c) second rainy seasons (July - September) (adapted from NMSA [66]).

Drought is a natural phenomenon, and it may happen at any time and at any place as an episode. Therefore, there must be an advanced information system that helps to mitigate the negative impact of drought on both human life and the environment. This mitigation effort has to be supported by the available technology and research. Drought research needs to be based on a long historical record that leads to the compilation of historical data on climatic, oceanic and other environmental fluctuations. With this in mind, drought spatial object modeling and the different approaches used in the past are presented in the subsequent subsections.

2.3 DroughtObject Modeling

The concept of object identification and modeling has been the subject of a long scientific effort to convert remotely sensed image into geographic phenomenon [25]. This concept was also discussed by past research as geographical data modeling [23], computer science object-oriented (OO) programming [93] and OO data modeling [94].

Goodchild [23] described that the OO notion of object identity is the same as geographic object entity identification. There are two different types of geographic phenomenon: geographic object and geographic field [95]. Here a geographic object entity is a geographic phenomenon that populates the concern area and is usually well distinguished, discrete and bounded [23, 95]. The geographic field is a geographic phenomenon, with continuous data values in the concern area [95]. In this research, the drought spatial object is the geographic phenomenon of drought characterized by a group of pixels, where these pixels can be segmented into separate regions with defined spatial locations and attributes [61].

Object identification in remote sensing is usually done by converting raster pixel values to geographic objects. Usually the image is first segmented, providing approximately homogeneous segments, and then classified [25]. Stein et al. [25] noted that various procedures for image segmentation are well documented, and include procedures based on mathematical morphology, edge detection and identification of homogeneity in one band or in a set of bands. Classification

routines include statistical routines, such as k-nearest neighbor classifiers, and increasingly fuzzy classification methods.

The objective of many remote sensing studies is to identify objects that have an ontological representation on the earth's surface. These objects can have different meanings, and they can be of various types and shapes. Commonly, a segmentation procedure is first applied, identifying homogeneous sets of pixel values, in one or more bands [25]

In this research, the object is in the context of object-oriented modeling [94]. This is based on the basic principle that an object has two characteristics: state and behavior [93, 94]. In this case, state is the attribute or information contained by an object and the behavior is the set of actions that an object performs to be perceived [93]. In identifying and modeling the DroughtObject using satellite images, state means the actual reflectance attributes (digital numbers registered by the sensor of satellites as pixel values or any index values, such as NDVI); behavior means when the DroughtObject happens on the ground, plants start dying or stop photosynthesizing (red band of the spectrum not used by plants and reflected back to the satellite sensor) and as a consequence the NDVI value is low. In the long run this can result in reduced crop yield [91, 96].

There are two key questions to be asked that help in identifying the state and behavior of any object: 1) "what possible states (attributes) can a given DroughtObject be in?", and 2) "what possible behavior (actions) can this object perform when it happens?" [93]. In this research, the geographic object that we are interested in is agricultural drought. Agricultural drought usually results in reduced biomass and yield [55].

The concept of identifying and modeling drought as an object is new [27]. Rulinda et al. [27] further indicated that "a next step in drought modeling is an approach focusing on spatial object and this kind of object can be built from different temporal resolution images". In remote sensing, objects are identified and subsequently classified on the basis of pixel information, and the objects are subsequently tracked in time, during which their behavior may be governed by

external factors (most relevant attributes) that also have to be identified and quantified [25].

In remotely sensed images, a pixel or group of pixels with similar spectral reflectance are considered to characterize the objects of interest. Remote sensing object classification methods usually consider texture information of features on the earth. The pixels identified as having the same texture are grouped and the groups are considered as objects [61]. These objects can represent physical features on earth, such as roads, parcels, or water bodies. When these physical features are classified based on texture, they are considered to be physical objects [61]. One step ahead of this basic concept, this research identifies virtual objects by using vegetation stress during drought incidence for identifying virtual drought spatial objects (extreme drought, severe drought, moderate drought, near normal and above optimum drought classes) on the real ground.

2.4 Object-Oriented Image Analysis

Benz et al. [61] explained principal strategies of object-oriented image analysis. They noted that the basic processing units of object-oriented image analysis are segments and image objects, and not single pixels. Advantages of object-oriented analysis include meaningful statistic and texture calculation, an increased uncorrelated feature space using shape (e.g., length, number of edges, etc.) and topological features (neighbor, super-object, etc.), and the close relation between real-world objects and image objects. This relation improves the value of the final classification and cannot be fulfilled by common, pixel-based approaches. Initial segmentation relies on low-level information, e.g., the pixel values and basic features of the intermediate image objects. Image objects are contiguous regions in an image. In their study, Benz et al. [61] distinguish between image object primitives and objects of interest. Only objects of interest match real-world objects, e.g., the building footprints or whole agricultural parcels. Object primitives are usually the necessary intermediate step before objects of interest can be found by segmentation and classification process. According to Benz et al. [61], the smallest image object is one pixel. In practical terms, to identify any geographic

phenomenon on the planet earth, the object on the ground should be bigger than the pixel size (spatial resolution).

For defining the minimum mapping unit, Jensen and Cowen [97] recommended a minimum of four pixels on the actual ground. This means that the satellite sensor spatial resolution should be one-half the diameter of the smallest object of interest. Cowen et al. [98] also indicated that to identify a given geographic phenomenon that is 5m wide (in diameter), the minimum spatial resolution (pixel size) of high quality imagery without haze or other problems is 2.5 x 2.5m pixels. The DroughtObject identification and modeling was also done based on these scientific recommendations. For this purpose, a number of image pre-processing steps were also implemented.

The image analysis in this research has two major steps: image pre-processing using anisotropic diffusion (AD) and image segmentation. These steps are described in the following subsections.

2.4.1 Image Pre-processing Using AD

AD is a technique aiming at reducing image noise without removing significant parts of the image content, typically edges, lines or other details that are important for the interpretation of the image [99, 100]. Diffusion algorithms remove noise from an image by modifying the image via a partial differential equation. Modifying the image according to isotropic (uniformly in all orientations) diffusion is filtering an image with a Gaussian filter.

As an advancement to the AD concept, Perona and Malik [99] replaced the classical isotropic with anisotropic diffusion equation, and qualitatively, the effect of anisotropic diffusion is to smooth the original image while preserving brightness discontinuities [101]. Perona and Malik [99] suggested a new definition of scale-space and introduced a class of algorithms that realize their method using a diffusion process. They indicated that the scale-space technique generates coarser resolution images by convolving the original image with a Gaussian kernel. This approach has a major drawback in that it is difficult to accurately obtain the locations of the semantically meaningful edges at coarse scales. In their research,

they suggested a new definition of scale-space, and introduced a class of algorithms that realize it using a diffusion process. The diffusion coefficient is chosen to vary spatially in such a way as to encourage intra-region smoothing in preference to inter-region smoothing. As the region boundaries in their approach remain sharp, they obtain a high quality edge detector that successfully exploits global information. The detail of the algorithm is presented in the authors' detailed work [99] and it is applicable for our drought virtual object analysis. The concept of anisotropic diffusion is also described by Black et al. [101] and Yu and Acton [102]. The diffusion algorithms removed noise from the Mapcubist output image by modifying the image via a partial differential equation. This was actually applied to reduce the image noise before segmentation and object extraction.

2.4.2 Image Segmentation

Image segmentation is defined as "the process of completely partitioning a scene into non overlapping regions in scene space (image space)" [103]. In addition to pixel information, texture and shape information are useful for object identification and segmentation [104]. In image analysis and classification, much information is contained in the relationship between adjacent pixels, including texture and shape information, which allows for identification of individual objects as opposed to single pixels [104]. Such an object-oriented approach allows the user to apply locally different strategies for analysis. Incorporating both spectral information (tone, color) as well as spatial arrangements (size, shape, texture, pattern, association with neighboring objects) comes closer to the way humans interpret information visually from different remotely sensed products. Franklin et al. [105] found that the incorporation of texture in addition to spectral information increased classification accuracy on the order of 10–15%. In this case, pixels are aggregated into image objects by segmentation, which is defined as the division of remotely sensed images into discrete regions or objects that are homogenous with regard to spatial or spectral characteristics [106]. Homogenous, in this case, refers to the fact that the within-object variance is less than the between-object variance [107].

Image segmentation is appealing for remote sensing applications because human vision tends to generalize images into homogenous areas. Research into image segmentation is not new. The works of Haralick et al. [108], for example, date back to the 1970s, and several methods now exist. They can be broadly categorized into measurement–space-guided spatial clustering, single-linkage region growing, spatial clustering, hybrid–linkage region growing, centroid–linkage region growing, and split and merge methods [109], or more simply, into edge-based and area-based algorithms [110]. Reed and Wechsler [111] used a filter-based approach to segment texture images, while Haddon and Boyce [112] incorporated edge detection into their segmentation algorithm. Other approaches include a probability-based image segmentation approach [113] and a fractal net evolution approach (FNEA), which is a multifractal approach [114]. With the FNEA, images are segmented at different scales, which adds a scale hierarchy to the analysis [115, 116]. Such a multi-resolution analysis using image segmentation is driven by remotely sensed data as well as expert knowledge, leading to a better understanding of the image content because image information is fractal in nature [107].

Many image segmentation methods are based on crisp relationships between or within the individual regions to be segmented. In many cases, however, these relationships may vary across the image because of noise, uneven illumination, limited spatial resolution, partial occlusions, and other phenomena [117]. The fuzzy connectivity segmentation approach takes these uncertainties into consideration. Rather than defining crisp relations, it attempts to describe the segmentation task with fuzzy rules, such as if two regions have about the same gray-value and if they are relatively close to each other in space, then they likely belong to the same object [117, 118]. A framework for such a reasoning approach is called fuzzy logic [117]. In the case of DroughtObject identification and modeling, there is a gradual change from drought pixels to non-drought pixels. These transitional regions on the actual ground can be captured if we use a fuzzy segmentation approach. Therefore, it makes sense to use fuzzy image segmentation in our current context.

2.4.3 Fuzzy Set Theory and Image Segmentation

In fuzzy connectivity segmentation, the basic assumption is that the image pixels seem to hang together when forming an object [117]. The hanging togetherness property is then described using fuzzy logic. There are a number of past studies on fuzzy connectivity [119-123].

Rosenfeld [119] indicated that topological relationships among parts of a digital picture, such as connectedness and surroundedness, play an important role in picture analysis and description. This paper generalizes these concepts to fuzzy subsets, and develops some of their basic properties. Rosenfeld [120] also presented another study on the fuzzy geometry of image subsets. This study summarizes efforts done on such fuzzy geometric concepts.

Bloch [121] presented associations between fuzzy sets and mathematical morphology by proving equivalence between two concepts: degree of connectedness for fuzzy sets, and connection cost for gray-level mathematical morphology. Dellepiane and Fontana [122] modified the traditional concept of fuzzy connectedness and described the concept, which extends the basic ideas to gray-level objects.

In another study, Udupa et al. [124] presented a novel methodology and a system that can be routinely used for segmenting and estimating the volume of multiple sclerosis lesions via dual-echo fast spin-echo magnetic resonance imagery. Based on several evaluation studies, the authors conclude that the methodology is highly reliable and consistent. Udupa et al. [125] also introduced the notion of relative connectedness that overcomes the need for a threshold and that leads to more effective segmentations. The central idea is that an object gets defined in an image because of the presence of other co-objects. Each object is initialized by a seed element. An image element c is considered to belong to that object with respect to whose reference image element c has the highest strength of connectedness. In this fashion, objects compete among each other utilizing fuzzy connectedness to grab membership of image elements. In this study, Udupa et al. [125] presented a theoretical and algorithmic framework for defining objects via relative

connectedness and demonstrated utilizing the theory that the objects defined are independent of reference elements chosen as long as they are not in the fuzzy boundary between objects. An iterative strategy is also introduced wherein the strongest relative connected core parts are first defined and iteratively relaxed to conservatively capture the more fuzzy parts subsequently.

Sonka et al. [117] noted that object definition in images may be considered to consist mainly of two related tasks: recognition and delineation. Recognition is the process of determining roughly the whereabouts of the object in the image. Delineation, on the other hand, is a process that defines the precise spatial extent and composition of the object in the image.

Udupa and Samarasekera [118] described a theory and algorithms for fuzzy connected object definition, treating a given image as a fuzzy subset of the set of spatial image elements (pixels) that make up the image. This approach consists of defining a local fuzzy relation called affinity on the set of pixels. The affinity value assigned to a pair of pixels is based on how close the pixels are spatially and in terms of intensity and intensity-based properties. A global fuzzy relation called connectedness is defined on the set of pixels based on affinity. The connectedness value assigned to a pair of pixels is the strength of the strongest of all paths from c to d. The strength of a path is simply the smallest affinity along the path. It was shown that fuzzy connectedness is a similitude relation and that the fuzzy components defined by this relation are an appropriate choice for characterizing objects in images. It was also shown that in spite of its enormous combinatorics, fuzzy component extraction can be done computationally elegantly via dynamic programming. This is an established method that is currently routinely utilized in several imaging applications [124, 126, 127]. This method is used in our DroughtObject segmentation.

2.4.4 Artificial Intelligence System

Artificial Intelligence (AI) is the science and engineering of intelligent machines, with the aim of replicating human-level intelligence in a machine [128]. It is related to the similar task of using computers to understand human intelligence, but AI

does not have to confine itself to methods that are biologically observable. AI is a broad field that includes specialty areas, such as problem solving and planning, expert systems, natural language processing, robotics, computer vision, learning, genetic algorithms, and neural networks.

Stair and Reynolds [129] described the specialty areas. According to Stair and Reynolds [129], problem solving and planning deals with systematic refinement of goal hierarchy, plan revision mechanisms and a focused search of important goals. Expert systems deal with knowledge processing and complex decision-making problems. Natural language processing areas include automatic text generation, text processing, machine translation, speech synthesis and analysis, grammar and style analysis. Robotics deals with the controlling of robots to manipulate or grasp objects and using information from sensors to guide actions. Computer vision deals with intelligent visualization, scene analysis, image understanding and processing and motion derivation. Learning deals with research and development in different forms of machine learning. Genetic algorithms are adaptive algorithms that have inherent learning capability and are used in search, machine learning and optimization. Neural networks deal with simulation of learning in the human brain by combining pattern recognition tasks, deductive reasoning and numerical computations. For the actual DroughtObject predictions, AI specialty areas, such as expert system, artificial neural network (ANN), and theoretical modeling approaches were assessed.

An expert system (ES) is a computer program that mimics the human reasoning process, which relies on logic, belief, rules of thumb, opinion, and experience [130-132]. Basheer and Hajmeer [132] indicated that in rule-based ES, the knowledge and experience of the expert are coded as a series of IF–THEN rules. An inference engine traverses the rules in the stored knowledge base, draws conclusions, and provides the user with a solution (such as a medical diagnosis). Unlike ANNs, ESs suffer from serious limitations, mainly their hypersensitivity to incomplete and noisy data [131]. Moreover, some human knowledge cannot be expressed explicitly in terms of rules [132]. These and several other limitations have stimulated the exploration of ANNs for data modeling [133]. ESs work in sequential manner,

whereas ANNs are parallel data manipulators, and sometimes viewed as a reverse approach to ESs [134]. Other differences relate to information processing, connectivity, self-learning ability, fault tolerance, and relation to neurobiology [131]. In order to capture the desirable features of both systems, ESs and ANNs are integrated into one hybrid system [135]. In this system, ANNs learn the hidden rules from the examples, and the ES extracts them in explicit forms, thus roughly simulating a whole-brain operation [134].

The decision to use ANNs, ESs, or theoretical modeling for an arbitrary problem depends primarily on the availability of both the theory explaining the underlying phenomena and the data, as presented in Figure 2.2 [136]. Basheer and Hajmeer [132] indicated that for a problem with abundant data but unclear theory, ANNs can be a perfect tool. Conversely, when both the data and theory are inadequate, the human expert's opinion should be sought, followed by coding this knowledge into a set of rules using ES techniques. Finally, when the problem is rich in both data and theory, it is possible to derive a physical (theoretical) model in which the data are used for model calibration and verification. Basheer and Hajmeer [132] further noted that overlapping areas are added to Figure 2.2 to characterize some other common scenarios, such as sufficient data and limited theory (i.e., a partially understood phenomenon), that dictate the need for a semi-empirical model. Alternatively, when both theory and data are abundant but a physical model is hard to formulate, the modeler may resort to empirical modeling such as ANNs. It is also obvious that Figure 2.2 supports hybrid ANN–ES systems (along the ANN–ES borderline) [132].

In identifying and tracking drought spatial object, there are huge datasets from different satellite sensors and limited theories [58]. According to Basheer and Hajmeer [132] and Rumelhart et al. [136], and also from our practical observations, it is appropriate to use ANN for DroughtObject exploration and predictions. After getting a clear understanding of the DroughtObjects from satellite pixels, it is quite reasonable to experiment with the theoretical modeling approaches, such as regression tree data mining technique [137].

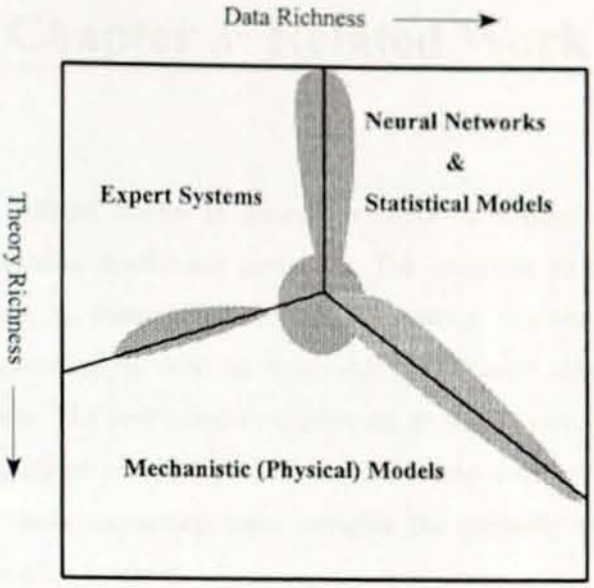


Figure 2.2: Suitability of modeling technique in relation to data and theory richness (adapted from Basheer and Hajmeer [132]).

Chapter 3 Related Work

In this chapter, related works in drought monitoring researches and drought monitoring information system are presented. The objective of the chapter is to assess past efforts in drought monitoring information system and to identify research gaps in coming up with an improved and efficient drought monitoring information system. The first subsection presents drought monitoring approaches. The second subsections presents past research in using data mining for drought monitoring, and lastly subsection three presents the currently available drought monitoring information systems.

3.1 Drought Monitoring Approaches

In drought monitoring, there are two major approaches: (i) climate-based, and (ii) remote sensing based (satellite-derived vegetation indices (VIs)) approaches. Various studies have also demonstrated the relationships between climate variables (e.g., precipitation) and satellite-derived VIs [49, 138, 139]. These two approaches are discussed in the following subsections.

3.1.1 Climate-Based Drought Monitoring Approaches

Climate-based drought monitoring approaches fall into two categories: 1) climate variable rainfall only, and 2) indices-based. These two approaches are discussed in this subsection.

3.1.1.1 Rainfall-Based Drought Monitoring Approaches

The rainfall-based drought monitoring approach is a direct approach using rainfall records from meteorological stations and satellites to estimate the rainfall attribute. Verdin et al. [59] reviewed background materials on food security assessment, drought monitoring, flood monitoring, and changing climate and food security issues with special focus on Ethiopia. They indicated that the Climate Hazards Group at the University of California, Santa Barbara, in 2005 completed a study of

rainfall variability in Ethiopia for the period 1960–2004, with a focus on the key growing seasons. Monthly data from 162 stations of the National Meteorological Service Agency were supplemented with observations from the Global Historical Climate Network and archives of the FAO and FEWS NET. In all, 186 stations within Ethiopia, and 373 in neighboring countries, were used. Block kriging was used to interpolate the station data and create spatial averages for zonal administrative units. The administrative units were then grouped into four regions based on similarity of their rainfall patterns. Figure 3.1 illustrates these regions along with the pattern of time-series of seasonal rainfall totals in the country. In this map, seasonal plots of rainfall for the country as a whole are also shown, for March–May ('Belg'), June–September ('Kiremt') and March–September periods.

Examination of Figure 3.1 shows that nationally, "Kiremt" rains have been quite consistent since 1960, with 7-year trends staying within ±50 mm of the long-term mean of 760 mm. "Belg" rains, on the other hand, have fallen off consistently since 1996. This decrease is seen to carry through in the graph of national rainfall totals for the full March–September season, though the downward turn is less dramatic. Turning to the regions, the annual rains in the northwest are seen to be most stable. The southwest, on the other hand, shows a steady decline throughout the entire period examined. Fortunately, rains are still abundant and there are no adverse implications for crops. The northeast and southeast give cause for concern. In the former case, we see dryness since 1996, and in the latter, dryness persists since about 1980. They also indicated that station data, satellite precipitation estimates, reanalysis fields and lake levels all suggest recent dryness in eastern Africa, primarily during the "Belg" (March–May) season. The associated circulation changes appear consistent with the expected response to diabatic heating in the southeastern Indian Ocean.

Recently, Ephrem and Meissner [140] analyzed the relationships between NOAA satellite rainfall estimates (RFE) and National Meteorological Agency (NMA) rainfall records. They indicated that there is higher correlation (r) during summer and spring seasons and weak correlation during winter. The correlation values are higher for the months with higher rainfall and lower during the period of little rain.

They also indicated that an in-depth study on the algorithm of RFE is needed to know exactly why it shows this kind of seasonal variation of correlation values. The good correlation between the two datasets during the main rainy seasons shows that the RFE data can be used to analyze the spatial and temporal extent of climatic variability occurring in large parts of the country during recent decades. Finally, they concluded that the two datasets are well correlated during the important rainy seasons, summer and spring. As a result, RFE images are reliable enough to be used for timely spatio-temporal analyses of disasters in times characterized by late inception, dry spells and early cessation of rainfall. The analysis of RFE images is a key tool for an advanced early warning system in the country. Ephrem and Meissner [140] note that the RFE images are very useful for a timely showing of rainfall stress in areas where there are few meteorological stations. Based on this recommendation, the three-month SPI of the current research was calculated from RFE data.

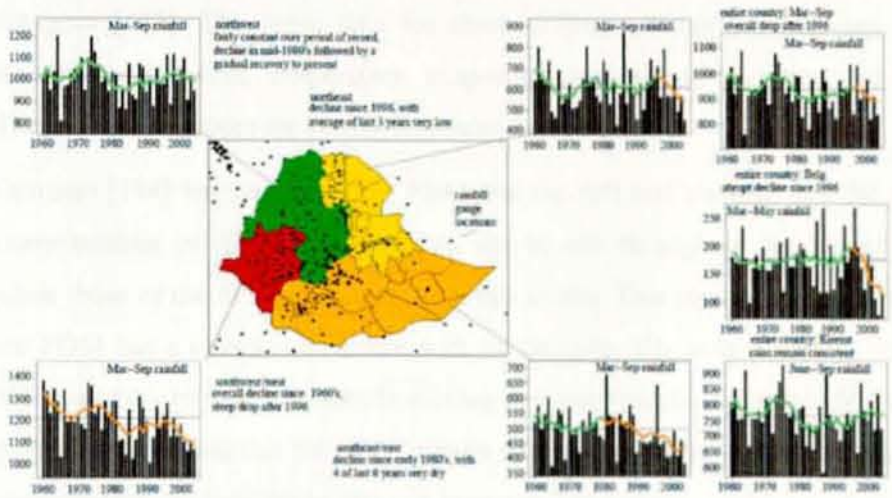


Figure 3.1: Time-series trends of seasonal rainfall totals for Ethiopia. The four colored maps are the four categories of rainfall pattern. The polygon maps are zones of Ethiopia and the black dots are rainfall gauge locations (adopted from Verdin et al. [59]). The green color for the northwestern part of the country is fairly constant over the period of record, with a decline in mid-1980 followed by a gradual recovery to present. The yellow color in the northeast part of the country shows decline since 1996, with an average of 2003, 2004 and 2005 very low. The golden color in the southeast/east part of the country shows decline since the early 1980s. The red color in southwest/west part of the country shows overall decline since the 1960s and a steep drop after 1996.

3.1.1.2 Climate Indices Based Drought Monitoring Approaches

Drought indices measure the departure from the local normal condition of a moisture variable based on its historical distribution [20]. These indices integrate various hydrological, meteorological, evapotranspiration, runoff and other parameters into a single number and give a comprehensive picture for decision making [20, 141, 142].

A number of climate indices are available for monitoring drought at optimum accuracy for assessing and responding to drought incidence. Among the various indices, the Palmer Drought Severity Index (PDSI) [143] and the Standardized Precipitation Index (SPI) [28] are extensively used. These drought indices are designed to provide a concise overall picture of droughts. They are often derived from massive amounts of hydro climatic data and are used for making decisions on water resources management and water allocations for mitigating the impact of droughts [142]. The input data for these different drought monitoring indices include precipitation, temperature, evapotranspiration, soil moisture and runoff. These climatic indices are briefly presented in Appendix 1.

Guttman [144] has compared the PDSI and the SPI and showed that the spectral characteristics of the PDSI vary from site to site throughout the United States, while those of the SPI do not vary from site to site. This research also showed that the PDSI has a complex structure with an exceptionally long memory, while the SPI is an easy to interpret, simple moving average process. From this finding, it is possible to conclude that the SPI is a more appropriate precipitation index to use for characterizing and validating drought from satellite sources.

Ntale and Gan [145] analyzed and modified the properties of three drought indices: the PDSI, the BMI (Bhalme–Mooley Index) and the SPI. They suggested that precipitation alone could explain most of the variability of East African droughts. Finally, Ntale and Gan [145] indicated that among the three indices, the SPI is more appropriate for monitoring East African droughts because it is more easily adapted to the local climate, has modest data requirements, can be computed at almost any

time scale, provides relatively consistent power spectra spatially, has no theoretical upper or lower bounds, and is easy to interpret.

The data inputs of the indices-based drought monitoring approaches are conventional meteorological networks, which are sparse and often report with significant delays [146]. Consequently, the requirements of drought early warning have inspired creative uses of remote sensing, numerical modeling and GIS to adapt traditional methods of drought monitoring.

3.1.2 Remote Sensing Based Drought Monitoring Approaches

Remote sensing is the science and art of obtaining data and/or information about an object, area or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area or phenomenon under investigation [147]. In acquiring remote sensing data, depending on the platforms used, there are two major categories: airborne remote sensing and space-borne remote sensing. Airborne remote sensing is carried out using aircraft with specific modifications to carry sensors (cameras), with a flight height of 100m up to 40km [147]. Space-borne remote sensing is carried out using satellites positioned at orbit about 150km and above in space [147]. This research mainly uses satellite remote sensing products for drought monitoring.

The primary interest of earth observing satellites in an environmental context is to study the role of terrestrial vegetation in large global process with the goal of understanding how the earth functions as a system [148]. This requires an understanding of the global distribution of vegetation types as well as their biophysical and structural properties and spatial variation [148]. Vegetation indices (VIs) are used to extract such information. VIs are spectral transformations of two or more satellite bands designed to enhance the contribution of vegetation properties and allow reliable spatial and temporal intercomparisons of terrestrial photosynthetic activity [148]. Several studies have been done, including Peters et al. [149], Lacaze and Berges [150], Wang and Li [151], and Rulinda et al. [152].

Peters et al. [149] indicated that reliance on weather data alone is not sufficient to monitor areas of drought, particularly when these data can be untimely, sparse, and

incomplete. Their research demonstrated that the Standardized Vegetation Index (SVI) is a very effective drought monitoring approach during the growing season. Their overall findings also showed that the SVI, along with other drought monitoring tools, is useful for assessing the extent and severity of drought at a spatial resolution of 1 km. The SVI is capable of providing a near-real-time indicator of vegetation condition within drought regions, and more specifically areas of varying drought conditions.

Wang and Li [151] developed the Vegetation Temperature Condition Index (VTCI) drought monitoring approach for using Terra–MODIS reflective and thermal infrared data, and used it to monitor drought at a study area in the southern Great Plains of the United States. They indicated that because meteorological data (such as precipitation and land-surface air temperature) collected by surface observation stations often possess poor spatial resolution, especially in remote regions with difficult access and in some developing countries, remotely retrieved NDVI and Land Surface Temperature (LST) data may provide a valuable source of information for monitoring drought. In this article, it was indicated that LST is a good indicator of the energy balance at the earth's surface because it is one of the key parameters in the physics of land-surface processes on regional and global scales. In this research, the concepts of NDVI and LST are combined and expressed as Vegetation Temperature Condition Index (VTCI). VTCI is not only related to NDVI changes in the region, but also related to LST changes of pixels with a specific NDVI value.

Lacaze and Berges [150] presented a comparison of NDVI data derived from two satellite sensors (i.e., SPOT-VEGETATION and MSG-SEVIRI) for the whole continent of Africa. Their findings suggest that because of higher temporal frequency, MSG-SEVIRI data can be used to obtain improved NDVI products through better removal of cloud-contaminated pixels. Rulinda et al. [152] also proposed a validation method for the MSG SEVIRI derived NDVI values with regard to vegetation health from ground observations. They recommended that NDVI can reliably be used for drought monitoring.

In the past, there were various efforts to monitor drought using remote sensing approaches [27, 149, 150, 152, 153]. Recently, Rulinda et al. [27] indicated that drought can be studied as a spatial object, an approach that had not yet been explored. Therefore, the main focus of the current research is in characterizing, identifying and modeling drought as spatial object by using the available remotely sensed data.

3.2 Application of Data Mining in Drought Monitoring

Data mining is a technique that uses a variety of data analysis tools to discover patterns and relationships of physical variables [6]. This technique has shown promise in multiple disciplines, bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issues of information extraction from large databases [154, 155]. Studies in ecological research have also introduced data mining techniques and found that it is a powerful tool in addressing complex ecological problems handling both numeric and categorical data [156]. Although drought effects on vegetation result from complex atmospheric and biophysical phenomena, data mining could provide mechanisms for understanding drought characteristics in space and time [67, 69]. These studies illustrate the potential of data mining for drought analysis and prediction.

In data mining, a model is a high level description of a dataset that can be subdivided into two categories: descriptive and prescriptive [157]. The descriptive category is used for summarizing the data in a convenient and concise way while the predictive model allows researchers to make statements about the populations from which the data is drawn or to likely be future values [10]. Data mining methods often ignore the appropriateness of the model for the data, namely the goodness of the fit, and usually find the best model in a given class of models. It is important to determine the classes of models that best fit the data, and in order to determine the appropriate class of models for specific data, it is important to understand the data [10].

Data mining is part of a larger process called knowledge discovery. Essentially, data mining discovers patterns and relationships hidden within large amounts of data. Data mining may be considered as advances in statistical analysis and modeling techniques to find useful patterns and relationships [158]. Data mining algorithms are also useful for data automation that is designed to allow users to create intelligent datasets by discretizing or converting data from many formats into compatible and user friendly formats. This process of automation makes it easier to execute different functions of the algorithms for the desired output without human interference and within a relatively short time. Thus, data mining tools can answer questions that traditionally were too time-consuming to resolve. These tools search databases for hidden patterns and find predictive information that experts may miss because it lies outside their expectations.

Recent developments in computing have provided the basic infrastructure for fast data access as well as many advanced computational methods for extracting information from large quantities of data. These developments have created a new range of problems and challenges for data analysts, as well as new opportunities for intelligent systems in data analysis. According to Thearling [159], data mining techniques are the result of a long process of research and product development. Thearling [159] indicated that this evolution began when business data were first stored on computers, continued with improvements in data access, and, more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining algorithms represent techniques that have recently been implemented as mature, reliable, understandable tools that are consistently outperforming older statistical methods in commercial applications [159]. Studies by Bigus [160] and Cabena et al. [154] show that data mining tools can also be used in predicting future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. Similarly, data mining algorithms and models such as decision trees, associations, clustering, classification, regression, sequential patterns, and time series forecasting have the potential to identify drought patterns and characteristics

[69]. For example, time series data mining can be applied in monitoring patterns of drought events.

One of the challenges in understanding drought is the large volume of data for numerous climate and hydrologic variables and indices, and the variety of spatial and time scales for which these data exist. Data mining is a technique that might help solve this problem, improving drought monitoring by identifying spatial and temporal patterns of drought characteristics [69]. It is a recent technology with great potential for identifying the most important information in databases. The information that can be derived through data mining can be used in making knowledge-driven decisions to reduce the impacts of drought through better monitoring. The already available algorithms in data mining increase efficiency in decision making and allow decision makers to make optimal choices for planning and preparedness. These tools are useful for proactive management of drought and improving the reliability of drought predictions [69]. Past studies on this domain include Tadesse et al. [69, 70], Rulinda [26], and Rulinda et al. [27].

Tadesse et al. [69] assessed the relationships between oceanic/atmospheric indices and drought, and identified the drought episodes within a certain time lag of the occurrences of oceanic and atmospheric indices so that the product could be used by decision makers in Nebraska, USA. In this study, the use of data mining techniques is introduced to find associations between drought and several oceanic and climatic indices that could help users in making knowledgeable decisions about drought responses before the drought actually occurs. They used two time-series data mining algorithms to find the relationships of drought and oceanic/atmospheric indices by considering time lags of their occurrences. These algorithms are the Representative Episodal Association Rule (REAR) and the Minimal Occurrences With Constraints and Time Lags (MOWCATL). The REAR algorithm converts the time-series data into discrete representations and generates association rules. The pre-processing of the time-series data for the REAR algorithm begins by discretizing (segmenting into groups of records) the data.

In another study, Tadesse et al. [70] introduced a prototype vegetation stress map called VegPredict that depicts vegetation conditions several weeks in advance and demonstrates the potential use of data mining for drought research. In this study, they present a data mining approach to modeling vegetation stress due to drought and mapping its spatial extent during the growing season. Rule-based regression tree models are generated to identify relationships between satellite-derived vegetation conditions, climatic drought indices, and biophysical data, including land-cover type, available soil water capacity, percent of irrigated farm land, and ecological type. The data mining method builds numerical rule-based models that find relationships among the input variables.

In Tadesse et al. [69, 70], the focuses were in using data mining tools for integrating drought-related attributes, such as vegetation stress, climate and biophysical data, and relating these datasets to satellite images. There were also other efforts in applying image mining techniques to the huge datasets available from different sources. These include Rulinda [26] and Rulinda et al. [27], who introduced the use of image mining for drought modeling.

Rulinda [26] developed a framework that uses image mining techniques to monitor drought by considering both vegetation stress intensity and duration. In this research, after the data was acquired, the NDVI deviations values were calculated for each location of the study area, and a function was selected to characterize drought severity over space, based on vegetation response. Then drought pixels were defined and extracted from the analyzed images. One of the research limitations identified by Rulinda [26] was that their study was conducted based on pixel-level analysis and remained at low information level. It was further recommended that the pixel-level definition of objects be replaced by a higher object-based level for facilitating processing tasks and interpretation of results.

In another study, Rulinda et al. [27] attempted to improve the early detection of drought using MSG SEVIRI data in eastern Africa. They used an image mining approach to handle the large amount of data used in the processing of hourly NDVI images to obtain drought indicator metrics. They also recommended that a next step

in drought modeling is to develop an approach focusing on spatial objects. To do so, objects have to be built from collected images.

3.2.1 ANN for DroughtObject Prediction

ANN is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information [161]. The human brain provides proof of the existence of massive neural networks that can succeed at those cognitive, perceptual, and control tasks in which humans are successful [162]. This natural behavior of biological neurons led to the derivation of a novel structure of the information processing system in a computing environment. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example [162].
The main objective of ANN-based computing is to develop mathematical algorithms that will enable ANNs to learn by mimicking information processing and knowledge acquisition in the human brain [132]. ANN-based models are empirical in nature; however, they can provide practically accurate solutions for precisely or imprecisely formulated problems and for phenomena that are only understood through experimental data and field observations.

Ashish et al. [163] noted that ANNs are computational mathematical models that emulate some of the observed properties of biological neural systems and draw on the analogies of adaptive biological learning. An ANN is composed of a number of interconnected processing elements that are similar to biological neurons. These processing elements are joined by weighted connections that are analogous to synapses in the human brain.

Abraham [164] indicated that there are two approaches that can be used to configure the neural networks so that the application of set inputs produces the desired set of outputs. These approaches involve either setting the weights manually or teaching the neural network to learn certain patterns. The latter approach is the most commonly used, and the network learns the pattern by itself and accordingly

updates its weights. Abraham [164] further noted that at a high level, the tasks performed using neural networks can be classified as those requiring supervised or unsupervised learning. In supervised learning, a teacher is available to indicate whether a system is performing correctly, to indicate a desired response, to validate the acceptability of a system's responses, or to indicate the amount of error in system performance. This is in contrast with unsupervised learning, where no teacher is available and learning must rely on guidance obtained heuristically by the system examining different sample data or the environment. A concrete example of supervised learning is provided by classification problems, whereas clustering provides an example of unsupervised learning. Because of the nature of the problem to be solved by the current research, which is the prediction of drought spatial objects classes (severe drought, drought, near normal and above optimum), the supervised learning method is preferred. Supervised learning in an ANN typically occurs by example through training or exposure to a known set of input and corresponding output data. The training algorithm adjusts the connection weights through an iterative procedure in which the error is minimized.
Usually, the back-propagation algorithm is used in layered feed-forward ANNs [162]. Rumelhart et al. [162] further noted that the artificial neurons are organized in layers and send their signals forward, and then the errors are propagated backward. The network receives inputs by neurons in the input layer, and the output of the network is given by the neurons on an output layer. There may be one or more intermediate hidden layers. The back-propagation algorithm uses supervised learning, which means that the experimenter provides the algorithm with examples of the inputs and outputs that the network is expected to compute, and then the error (difference between actual and expected results) is calculated.

Past research, such as Bischof et al.[165], Paola and Schowengerdt [166], Blackard and Dean [167], Giraudel [168], and Ashish et al.[163], has shown the superiority of ANNs to some of the classical statistical methods in various problems, including classification challenges. ANNs are commonly used for segmentation and

classification purposes and are recommended for problems where data diversity is large [169] or complex [170].

Kumar et al. [171] indicated that ANN is recommended for three reasons. First, the application of a neural network does not require a priori knowledge of the underlying process. Second, all the existing complex relationships between various aspects of the process under investigation need not be known. Third, ANNs are data driven when compared to conventional approaches, which are model driven. Furthermore, Kumar et al. [171] noted that neural networks offer several advantages over conventional approaches. The most important aspect is their ability to develop a generalized solution to a problem from a given set of examples and to combine and adapt to changing circumstances with exposure to new variations in the problem. This attribute of generalization permits them to be applied to a variety of problems and to produce valid solutions even when there are errors in training data or in the description of the problem. For these advantages, the ANN models were used for exploring DroughtObject monthly prediction in the current research. Past researches, such as Kim and Valdes [172], Mishra and Desai [62], Morid et al. [173], and Marj and Meijerink [174] used ANN for modeling and predicting drought.

Kim and Valdes [172] presented a conjunction model in order to improve forecast accuracy for regional droughts. In their research, the proposed conjunction model is based on dyadic wavelet transforms and neural networks. Neural networks have shown great ability in modeling and forecasting nonlinear and nonstationary time series in a water resources engineering, and wavelet transforms provide useful decompositions of an original time series. The wavelet-transformed data aids in improving the model performance by capturing helpful information on various resolution levels. Neural networks were used to forecast decomposed sub-signals in various resolution levels and reconstruct forecasted sub-signals. The performance of the conjunction model was measured using various forecast skill criteria. Their results also confirmed that the conjunction model significantly improves the ability of neural networks for forecasting the indexed regional drought.

Mishra and Desai [62] compared linear stochastic models, recursive multi-step neural network and direct multi-step neural network for drought forecasting. Their models were applied to forecast droughts using standardized precipitation index (SPI) series as drought index. The results obtained from their models show that recursive multi-step approach is best suited for one month-ahead prediction. When longer lead time of four months is considered direct multi-step neural network approach outperforms recursive multi-step and autoregressive moving average models. They also showed that the performance of the autoregressive moving average models provides good result up to two-month lead time but inferior in comparison to direct multi-step neural network approach.

Morid et al. [173] examined the utility of ANN approach for medium and long-term forecasting of both the likelihood of drought events and their severity. The indices used for their research are the Effective Drought Index (EDI) and the Standard Precipitation Index (SPI). The forecasts are attempted using different combinations of past rainfall, EDI and SPI drought indices in preceding months and climate indices, such as Southern Oscillation Index (SOI) and North Atlantic Oscillation (NAO) index. A number of different ANN models for both EDI and SPI with the lead times of one to 12 months have been tested at several rainfall stations in their study area. The structure of the model inputs (previous rain and drought indices) does not vary with the lead time, which makes the models very convenient for the operational purposes. They recommended that their final forecasting models can be utilized by drought early warning systems.

Marj and Meijerink [174] related a forecast of NDVI with two climatic signals by using ANN analysis. In their research, the applied ANN is a feedforward multiple neural network. The inputs of the model are the climatic signals Southern Oscillation Index (SOI) and North Atlantic Oscillation (NAO). In order to forecast NDVI with ANN, the normal method was used for the recent period, and for evaluation, the moving window method was used for a longer (18 years) period.

Their results showed that in spring, NDVI can be predicted using ANN, with the input of SOI and NAO indices of the preceding (one year) spring period.

3.2.2 Regression Tree for DroughtObject Prediction

Regression tree performs induction by means of an efficient recursive partitioning algorithm [175]. The choice of the test at each node of the tree is usually guided by a least squares error criterion. Regression trees obtain good predictive accuracy on many domains, and they provide interpretable models of the data and have low computational demands (both running time and storage requirements) [137, 175, 176].

Classification and regression tree (CART) analysis is an increasingly popular form of statistical analysis available through widely used statistical packages [176-178]. Lawrence and Wright [176] noted that CART operates by recursively splitting the data until ending points or terminal nodes are achieved using preset criteria. CART therefore begins by analyzing all explanatory variables and determining which binary division of a single explanatory variable best reduces deviance in the response variable [176, 177]. For each portion of the data resulting from this first split, the process is repeated, continuing until homogeneous terminal nodes are reached in a hierarchical tree [176].

Lawrence and Wright [176] described the result of the CART analysis as a dichotomous decision or classification tree. Each path through the tree, defined by a series of dichotomous splits, specifies the conditions that lead to a most probable class. The tree, therefore, might be viewed as a series of rules that can be used for unknown observations to predict likely class membership. CART also performed well in drought prediction analysis [71, 72].

Brown et al. [72] used regression tree in CART to analyze the historical data in the training database and generate the three seasonal, rule-based, piecewise linear regression VegDRI models. The rules and instances option available in Cubist were utilized for their model development. The Cubist models are composed of an

unordered set of rules, with each rule having the syntax, "if x conditions are met then use the associated linear regression model."

Tadesse et al. [71] used regression tree in CART to 18 years of historical dataset in the training database for each of the targeted, seasonal bi-weeks analysis of VegOut model. In VegOut the regression tree was used to analyze the historical data in the training database and generate rule-based, piecewise linear regression models for each biweekly period. In their research, the regression tree algorithm in CART 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. This algorithm produces a series of rule-based models from this partitioning. Each rule set has a corresponding multivariate linear regression equation that can be used to predict the target value. Tadesse et al. [71] also described that the regression tree models in their analysis can account for nonlinear relationships between predictive and target variables through a series of regression equations associated with different rule sets.

In DroughtObject modeling and prediction, regression tree models were used as white-box theoretical modeling, following ANN DroughtObject exploration and in the actual monthly predictions. The rules generated from this analysis were also implemented on Mapcubist software for pattern recognitions and spatial analysis of the DroughtObjects.

3.3 Existing Drought Monitoring Information Systems

The purpose of this subsection is to assess the currently available drought monitoring information systems and present some related works with a focus in coming up with an efficient drought early warning information system. The most related works with this PhD research include U.S. Drought Monitor (USDM) [179], Vegetation Drought Response Index (VegDRI) [72], Vegetation Outlook (VegOut) [71], Drought Mitigation Decision Support System of NASA [180], Early Warning

Information systems in South Africa, Famine Early Warning Systems Network (FEWSNET) [45], and Early Warning and Disaster Risk Monitoring Directorate of Disaster Risk Management and Food Security Sector (DRMFSS) of Ethiopia [181]. The USDM is produced in partnership with numerous agencies including NOAA, the U.S. Department of Agriculture, and the National Drought Mitigation Center (NDMC) [179]. Svoboda et al. [182] described that the USDM, operational since 1999, is developed using a hybrid approach that considers a number of variables such as short- and long-term climate-based drought indicators, hydrologic indices, and remote sensing information. The USDM map provides a general assessment of drought conditions across the nation. The USDM map is updated weekly and has a spatial resolution at the approximate scale of a climate division. Climate divisions are administrative construct of the NOAA that can vary in area from tens to hundreds of km^2 across the conterminous U.S. The coarse spatial resolution of the USDM limits its utility for drought response and mitigation at a more localized level. The USDM maps are released at the website http://droughtmonitor.unl.edu.

To overcome the spatial resolution limitations of USDM, VegDRI has been introduced in 2006 [179] for U.S. VegDRI is based on monitoring drought-induced vegetation stress [72]. VegDRI integrates traditional climate-based drought indicators and satellite-derived vegetation index metrics with other biophysical information to produce a 1 km map of drought conditions that can be produced in near-real time. The results from this product demonstrate that more spatially detailed drought patterns can be characterized and monitored in the 1 km VegDRI maps, compared to the commonly used USDM map. This provides more localized drought information, which is currently at a county to sub-county scale [72, 179]. VegDRI product is being released at the website http://www.drought.unl.edu/MonitoringTools/VegDRI.aspx. The limitation of VegDRI product is that the information is being released at near real-time basis and the information is at pixel level (low information level).

The other drought product for U.S. is VegOut. VegOut tool provides accurate early warning drought prediction [71]. VegOut integrates climate, oceanic, and satellite-based vegetation indicators to identify historical patterns between drought and

vegetation conditions indices and predict future vegetation conditions based on these patterns at multiple time steps (two, four and six-week outlooks) [68]. Even thought, the tool is looking into the future and has prediction capability of the future drought condition, the same as VegDRI product, the information is produced at pixel level (low information level).

Drought Mitigation Decision Support System of NASA is producing an integrated system solution (ISS), which is designed to incorporate the NASA earth science model, Global Modeling and Assimilation Office (GMAO), climate and hydrologic prediction models, and decision analysis modules into the Drought Monitor-Decision Support System (DM-DSS) [180]. The main objective of DM-DSS is to develop seasonal predictive capacity for the Drought Monitor-Decision Support System (DM-DSS), using Earth science models and satellite products. An extended framework of this tool is presented in Figure 3.2. This tool is under development and it is not possible to get detail information.

Early warning information systems in South Africa uses decile rainfall, the water satisfaction index (WSI), the NOAA Normalized Difference Vegetation Index (NDVI), and other crop- and rangeland-based models for their drought early warning information. Their drought early warning system remains dormant until the need arises mainly due to limited resources available to run the system [183]. There is lack of integrated tool and coordination between the institutions involved in South African drought monitoring information system [184].

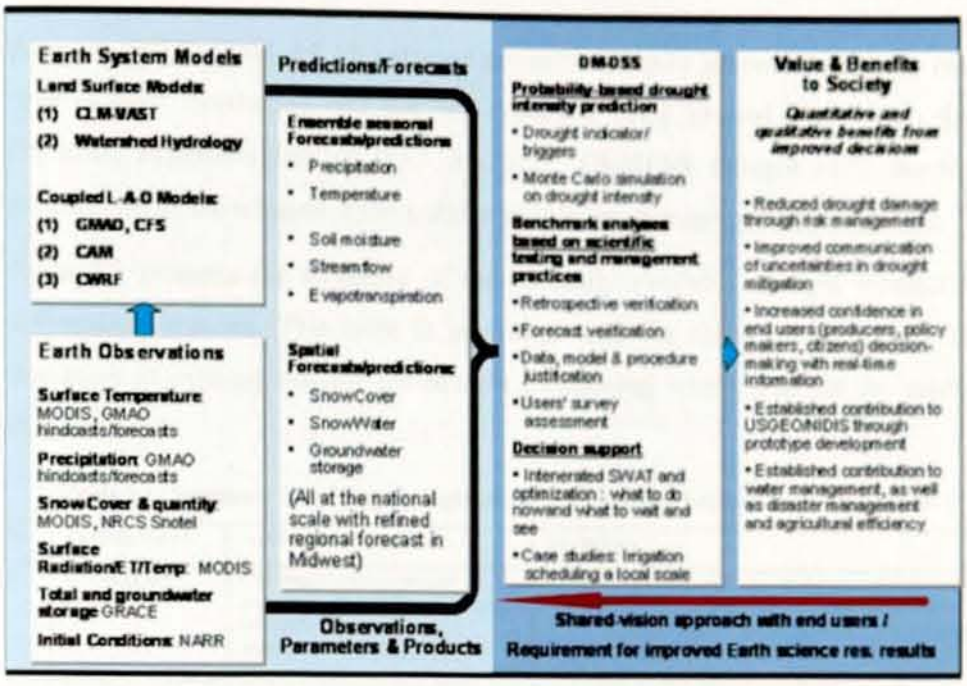


Figure 3.2: An extended drought monitoring DM-DSS framework of NASA[180].

Famine Early Warning System Network (FEWS NET) produces drought monitoring products for Ethiopia. FEWS NET was started in 1985 [65], after the worst drought in Ethiopia [66]. The FEWS NET data portal provides access to spatial data, satellite imagery, and other data and graphic products in support of the FEWS NET activities in Ethiopia and other African countries [185]. The limitation of FEWS NET product is the coarse resolution of the products. FEWS NET products are being released at the website http://www.fews.net/pages/country.aspx?gb=et. FEWS NET for Ethiopia is closely working with Ethiopian government, specifically DRMFSS.

DRMFSS has one directorate, which is responsible for the general management and overall coordination of the early warning disaster risk monitoring and response to disasters at national level. The directorate works in collaboration with regional, zonal and woreda (district) level offices. The offices are organized under Bureaus of Agriculture at regional levels with similar set ups at zonal and woreda levels. The case teams under the directorate use LEAP (Livelihood, Ethiopia, Assessment, Protection) software for drought episode assessment in coordination to World Food

Program (WFP) [186]. LEAP software products can give a general guide to real-time drought monitoring and not yet evaluated using ground observation data (Personal communication at WFP in 2012). DRMFSS drought early warning products are being released at the website http://www.dppc.gov.et/Pages/maps.html.

Table 3.1 presents the summary of the currently available drought monitoring information systems. This table is presented with the focus on presenting the limitation of existing systems for drought monitoring, which lead for the current PhD research.

Table 3.1: Summary of existing drought monitoring approaches and their limitations

No	Existing System	Achievements	Limitations
1	USDM [179]	The map provides a general assessment of drought conditions across the nation and has a temporal resolution of one week.	• Coarse spatial resolution at the approximate scale of a climate division (tens to hundreds of km^2). • The coarse spatial resolution of the USDM limits its utility for drought response and mitigation at a more localized level.
2	VegDRI [72].	Real time drought monitoring.	• The system output is at pixel level (low information content). • The model was not evaluated using ground observation data.
3	VegOUT [71]	Six weeks drought predictions.	• The system output is at pixel level (low information content). • The model was not evaluated using ground observation data.
4	Drought Mitigation Decision Support System of NASA [180]	Seasonal predictive capacity for the Drought Monitor-Decision Support System (DM-DSS), using Earth science models and satellite products	• Not yet functional and it is not possible to get detail information.
5	Early warning information systems in South Africa [183, 184].	Provide drought information when the need arise.	The information is not continuously available for users and hence lack of complete information.
6	FEWS NET [45]	The FEWS NET data portal provides access to spatial data, satellite imagery, and other data and graphic products	The FEWS NET product is a coarse spatial resolution.
7	DRMFSS [181]	The offices uses LEAP software to produce real-time drought severity status in Ethiopia	LEAP drought monitoring product is not evaluated using ground observation data.

Chapter 4 Materials, Methods and Analysis

4.1 Introduction

From the two research paradigms (behavioral and design science) [187, 188] in information systems, this PhD research employed the design science paradigm with an exploratory research approach. This is because the research was mainly focused on exploring, identifying key attributes, and investigating possible relationships of the identified attributes with DroughtObject for its improved characterization and modeling.

The method employed for the exploratory investigation of DroughtObject was modeling. As defined and described by Jenkins [189], modeling is a closed deterministic system, in which all of the independent and dependent variables are known and included in the model. The reason for selecting the modeling research approach was that drought by its nature is virtual (i.e., it is not a physical phenomenon that can be observed on the actual ground). In this kind of scenario, some attributes (such as lack of precipitation) have to be used for its characterization and representation using scientific modeling approaches. This representation has to use some mathematical representation or modeling [189]. In modeling DroughtObject dependent attributes using the identified key independent attributes, a regression data analysis technique ([190, 191]) was used.

Using the exploratory research approach and modeling research method, different data collection and analysis techniques were employed with a KDD approach. This chapter presents the KDD process followed in extracting drought information from satellites.

KDD has clearly defined steps and processes [192]. In this study, these steps were divided into four: 1) data selection, 2) data pre-processing (which includes data pre-processing and transformation), 3) data mining (including interpretation), and 4)

knowledge construction [10]. Each of these steps is diagrammatically presented in Figure 4.1.

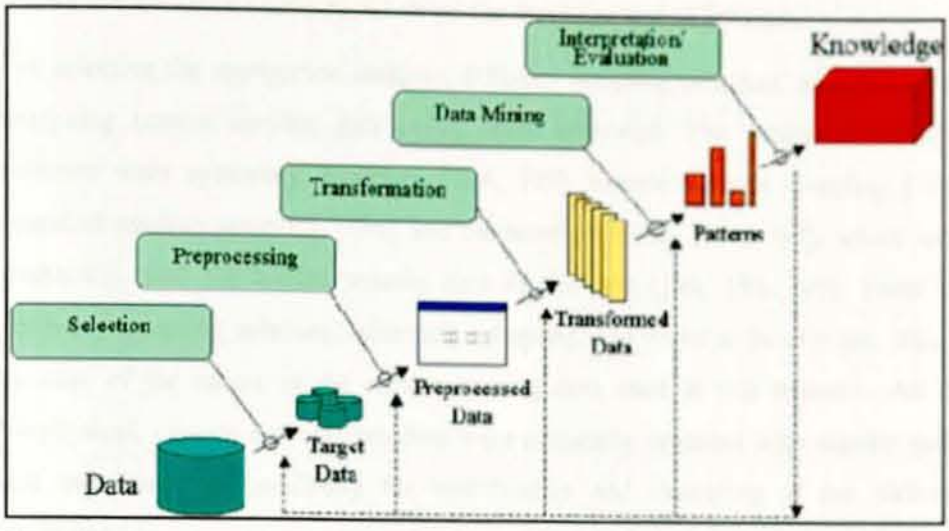


Figure 4.1: Major steps in KDD for finding patterns from data (adapted from Hamilton [192]).

4.2 Study Area

The study area for this research was Ethiopia. Ethiopia occupies the interior of the Horn of Africa, stretching between 3° and 14° N latitude and 33° and 48° E longitude, with a total area of 1.13 million km^2 [193].

Ethiopia is located in the tropics, and variations in altitude have produced a variety of microclimates. Mean annual rainfall ranges from 2000 mm over some pocket areas in the southwest highlands to less than 250 mm in the lowlands. In general, annual precipitation ranges from 800 to 2200 mm in the highlands (>1500 meters above sea level) and varies from less than 200 to 800 mm in the lowlands (<1500 meters above sea level)[66, 193]. Rainfall also decreases northward and eastward from the high rainfall pocket area in the southwest [66].

4.3 Sampling Scheme

The sample frame for this research was determined based on the spatial resolution of Tropical Applications of Meteorology using Satellite (TAMSAT) rainfall data. This was used because agricultural drought is highly related to the amount of

precipitation. TAMSAT data has a spatial resolution of 10km. Using this spatial resolution, a total of 11261 pixels cover the spatial extent of Ethiopia.

For selecting the appropriate samples, different sampling schemes, specifically for analyzing remote sensing data [194], were reviewed. The reviewed sampling schemes were systematic sampling [194, 195], simple random sampling [194], stratified random sampling [196] and clustered sampling [194, 197], which were frequently used for remote sensing data in the past [194, 198, 199]. From the reviewed sampling schemes, systematic sampling was found to be relevant. This is because of the nature of the remote sensing data used in this research. All the biophysical, climate and satellite data were originally acquired with regular grids, and there was no possibility for stratification and clustering of the different attributes.

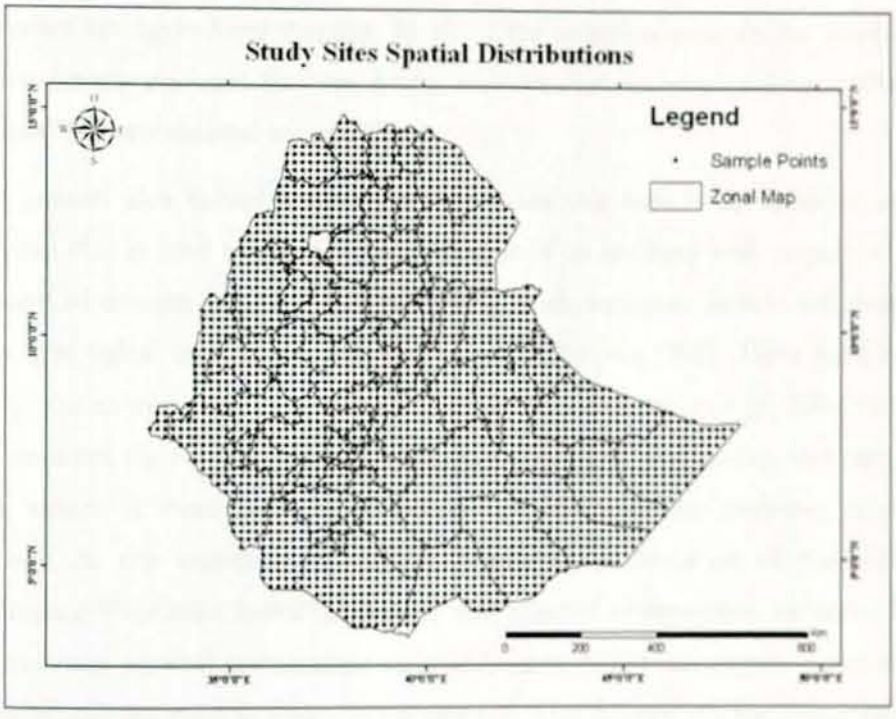


Figure 4.2: Map showing the distributions of the 2812 sample points in Ethiopia.

Using a systematic sampling approach, the sampling points were spaced at a regular intervals over the study area. From the 11261 pixels covering all of Ethiopia, starting at the north corner of the country, every other pixel was selected for the current study. Here the assumption was that these pixels can capture the maximum

information from the heterogeneous ecosystems of Ethiopia. This has resulted in a total of 2812 sample points covering the whole country (Figure 4.2). This means that one observation was sampled for almost every 20km distance, and this can capture heterogeneous elevations, land covers, ecosystems and soil types of Ethiopia.

4.4 Attribute Selection and Data Analysis

To identify the most relevant attributes for drought monitoring, detailed reviews of past research and document analysis of possible sources of qualitative evidence (both direct observations of drought monitoring institutes and archival records) were conducted. Currently, hundreds of attributes are available for drought monitoring (Appendix 1). To select the most relevant attributes, two criteria were used: 1) availability of the attribute for modeling (cost or affordability), and 2) relevance for agricultural drought. In all of the selection process, the attributes selected were the ones that are freely available for experiment from different national and international organizations.

The general idea behind attribute relevance analysis here is to compute some measure that is used to quantify the relevance of an attribute with respect to the concept of drought modeling and monitoring. Such measures include information gain, gini index, uncertainty, and correlation coefficients [200]. There have been many studies in machine learning on attribute relevance analysis [6, 200-202]. In this research, the potential attributes were identified from past studies, and then the data values of these potential attributes were extracted for modeling drought episode. In this experimental analysis, Standardized Deviation of Normalized Difference Vegetation Index (SDNDVI) was selected as dependent variable. This variable was selected as dependent variable because NDVI can capture agricultural drought with optimum accuracy in the past [49, 71, 72, 149]. All the data obtained from the potential attributes were iteratively selected and statistically tested for their relationships to the dependent attribute. For this statistical exploration, Ordinary Least Square (OLS) was used (equation 1). The experiment was implemented on ArcGIS 10 [203] and Matlab software Version 7.9.0 [204].

$$y_{OLS} = a_0 + \sum_k a_k x_{ik} + e_i \qquad \textbf{Equation 1}$$

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 e_i is the error.

After identifying the key attributes for DroughtObject modeling, detailed exploratory analyses were conducted for understanding the relationship of these variables to SDNDVI. Linear regression analysis scatter plots were used to examine the relationship of the identified attributes with SDNDVI. Emphasis was given to the strong relationship between the identified key variables and the dependent variable (SDNDVI). During these analyses, variables that were redundant and had non-significant relation with the dependant variables were excluded. Experiments were also done on the different models for their statistical significance using OLS in ArcGIS 10 software [203]. In this process, the adjusted R^2 value, Akaike's Information Criterion (AIC) [203, 205, 206], Variance Inflation Factor (VIF) [203, 207, 208], Moran's I index [203, 209], the coefficient sign (+/-) and magnitude of each explanatory variable's relationships values were used as criteria for attribute relevance. When the coefficient is positive, it means that the explanatory variable has a positive relationship with DroughtObject. A large positive coefficient is a strong relationship. Probabilities were also computed for each coefficient to evaluate whether or not the identified key attribute was helping the DroughtObject model. Here the assumption was that effective explanatory variables should be statistically significant. When the identified variables were not statistically significant (both their coefficient and constant), then they were not helping the DroughtObject model, and they were dropped from the model. At the beginning of the analysis, floristic_region data was included, but after some experimenting, it was realized that it has no relationship (with R^2 value < 0.01) with drought episodes, and the data was excluded from the model. For other attribute, such as DEM, the R^2 was 0.095, AIC was 8012.92 and spatial autocorrelation (Global Moran's I index) was 0.58 ($p < 0.01$). The interpretation for this is that DEM alone can explain about 10% of the variability of DroughtObject. The spatial autocorrelation also showed that there is statistically significant clustering of the

OLS residuals in this prediction process. The null hypothesis is that there is clustering of the residuals and this hypothesis cannot be rejected. This means that all the key attributes that explain SDNDVI are not included in the model. This is a logically meaningful statistical output in that only DEM is included from the other potential attributes collected from previous research. Each of these attributes was tested using the same procedure as DEM attribute.

In overall attribute selection process, the AIC, VIF and Moran's I index were very important criteria. The AIC was used as a model performance by including a given potential attribute; the VIF as parameter for controlling the duplication of the information of the potential attribute with previously selected attributes; and Moran's I index as a parameter for controlling absence of key attributes and for avoiding misspecification of the DroughtObject model. The three parameters used for selecting the relevant attributes in the whole experimental process are presented in Figure 4.3 for the subsequent modeling experiment. In this figure, floristic region (flor_Rgion) is not connected with the relevance metrics, since this attribute was not related with the dependent attribute. Using these criteria, a total of 11 attributes were identified for modeling DroughtObject. The lists, acronyms, data type, format, sources and references of these attributes are presented in Table 4.1. These attributes were also classified into biophysical, climatic and satellite sources and they are described in the following subsections.

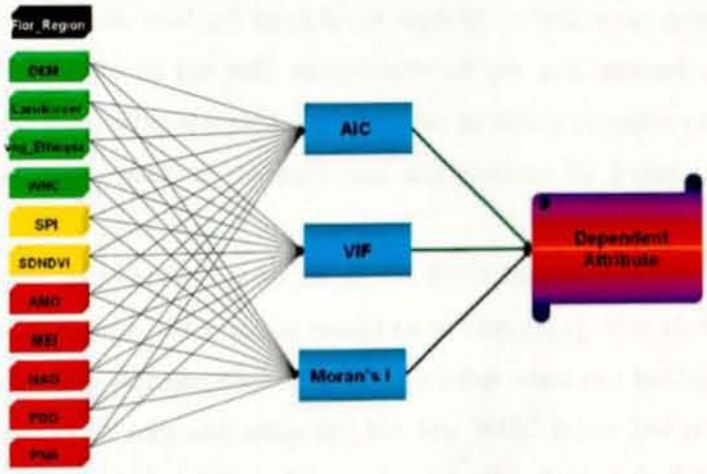


Figure 4.3: Diagrammatic illustration of the iterative process for selecting relevant attributes.

Table 4.1: Attributes identified for DroughtObject modeling experiment.

No.	Attribute	Acronym	Type	Format	Source	Reference (s)
1	Standardized Deviation of Normalized Difference Vegetation Index	SDNDVI	Satellite	Raster	NOAA AVHRR	[185, 210]
2	Digital Elevation Model	DEM	Biophysical	Raster	USGS	[211]
3	Soil Water Holding Capacity	WHC	Biophysical	Raster	USGS	[211]
4	Ecological Regions (Ecosystems of Ethiopia represented by veg_Ethiopia)	veg_Ethiopia	Biophysical	Vector	Ecodiv.org	[212]
5	Land Cover	Landcover	Biophysical	Raster	ESA	[213, 214]
6	Three Month Standard Precipitation Index	SPI_3month	Climate	Raster	IRI	[47, 215]
7	Pacific Decadal Oscillation	PDO	Oceanic/Atmospheric	Point data	NOAA	[216]
8	Atlantic Multi-decadal Oscillation Index	AMO	Oceanic/Atmospheric	Point data	NOAA	[216, 217]
9	North Atlantic Oscillation	NAO	Oceanic/Atmospheric	Point data	NOAA	[216, 218, 219]
10	Pacific North American Index	PNA	Oceanic/Atmospheric	Point data	NOAA	[216]
11	Multivariate ENSO Index	MEI	Oceanic/Atmospheric	Point data	NOAA	[216, 220]

4.4.1 Biophysical Attributes

The biophysical attributes include water holding capacity (WHC), land cover, digital elevation model (DEM) and ecosystem type of Ethiopia (veg_ethiopia). These attributes are described in the following subsections.

4.4.1.1 Water Holding Capacity (WHC)

WHC data is the specific ability of a particular type of soil to hold water against the force of gravity. Different types of soils have different capacity for holding soil moisture. For example, sand soil has a lower capacity to hold water compared to a clay soil. The nature of the soil, composition of the soil, amount of organic component and size of the soil particles determine its ability to retain water. Water molecules are held closely to the individual soil particles by forces of cohesion [221].

The WHC data was obtained from the USGS-Earth Resources Observation and Science (EROS) Center with a spatial resolution of 1km [211]. This attribute in the current research was included with the 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 agricultural drought (vegetation condition). Churkina et al. [222] also indicated that WHC represents the potential of the soil to hold moisture and make it available to plants, which exerts

control over plant growth and influences the sensitivity and response of vegetation to agricultural drought.

4.4.1.2 Land Cover

Land cover is the physical material at the surface of the earth [214]. This attribute is included in the DroughtObject model to represent the different monthly 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 (Medium Resolution Imaging Spectrometer) mosaics for the year 2009 with spatial resolution of 300m [213]. This data was found to be one of latest, and highest spatial resolution data available. For Ethiopia, a total of 14 different land cover classes were identified in [213]. The different land cover class representations of the original raster data and the label explanations are presented in Table 4.2.

Table 4.2: Land cover data raster representation and label explanation [213]

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

4.4.1.3 Digital Elevation Model (DEM)

A digital elevation model (DEM) is a digital model or 3D representation of a terrain's surface [223]. The DEM data was obtained from EROS with a spatial resolution of 1km [211]. The DEM was included in the DroughtObject model to

account for the diversified elevations of Ethiopia (ranging from -120m below sea level to 4600 above sea level), which are assumed to have great influence on the spatial extent of agricultural drought.

4.4.1.4 Vegetation Type of Ethiopia (veg_ethiopia)

Vegetation type is a general term for the type of plant life of a region; it refers to the ground cover provided by plants. It is broader than the term flora, which refers exclusively to species composition [224]. Ecosystems of Ethiopia are mainly determined based on vegetation types [81]. Using vegetation type as major attribute, Ethiopia has about fourteen major ecosystems [81, 212]. The data was obtained from the Royal Danish Academy of Sciences [212]. The ecosystem data (veg_Ethiopia) was included in the DroughtObject model 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 descriptions of this data are presented in Table 4.3.

Table 4.3: Ecosystem of Ethiopia raster data representation and label explanation [212].

No.	Raster attribute value	Label descriptions
1	1	Desert and semi-desert Scrubland (DSS)
2	2	Acacia-Commiphora woodland and bush land 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)

4.4.2 Climate Attributes

4.4.2.1 SPI

Because this study was primarily focused on the effect of precipitation deficits on the spatial patterns of vegetation health and vigor (agricultural drought), the SPI was selected, representing climatic data input. For this research, the rainfall climatic data for the SPI derivation was obtained from satellite rainfall estimates (RFE). RFE was used because the number of rain gauges in Ethiopia is very small, and they are unevenly distributed.

Currently there are more than ten satellite rainfall products at different spatial and temporal resolutions [47, 48]. The RFE data for this research was obtained from TAMSAT, which is produced by the TAMSAT group at Reading University in the United Kingdom [48].

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. 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 [48].

The TAMSAT method is based on the assumption that cold cloud-top temperatures of tropical storms identify raining clouds. These temperatures are obtained from Meteosat thermal-infrared images. The length of time that a satellite pixel is colder than a given temperature threshold is then summed over 10 days to create cold-cloud-duration (CCD) images. The parameters of this relationship are found through calibration using rain gauge data [47]. The detailed descriptions of these products are presented in Dinku et al. [47] and Dinku et al. [48].

For this PhD research, RFE products from 1983 to 2006, which were accumulated to 10-daily totals, were obtained from the International Research Institute for Climate and Society [215] with spatial resolution of 10km. Then these dekadal (10 days) values were summed and monthly values were taken for DroughtObject modeling.

4.4.2.2 Teleconnections and Oceanic Indices

Teleconnection in atmospheric science refers to anomalies being related to each other at large distances of thousands of kilometers [225]. Climate scientists use teleconnection to correlate or make teleconnection patterns to describe relationships in the variability of large-scale features of the atmospheric circulation as well as tropical and extra-tropical precipitation and temperature relationships, especially those related to the El Nino–Southern Oscillation (ENSO) [225, 226].

Trenberth et al. [225] noted that ENSO is characterized by variations in the temperature of the surface of the tropical eastern Pacific Ocean warming or cooling known as El Niño and La Niña, respectively, and air surface pressure in the tropical western pacific (the Southern Oscillation). The two variations are coupled: the warm oceanic phase, El Niño, accompanies high air surface pressure in the western Pacific, while the cold phase, La Niña, accompanies low air surface pressure in the western Pacific. Using these principles, the physical modeling of the inter-annual variability of sea surface temperature (SST) over the eastern tropical Pacific Ocean revealed the predictability of ENSO [227]. ENSO predictability then led to potential predictability of seasonal climate over many tropical and some extra-tropical regions. Other studies have also indicated that northern summer rainfall in the Sahel responds partly to ENSO fluctuations [228]. Research devoted to the twentieth-century Sahel drought focused heavily on the impact of regional and global SST anomalies on inter-decadal time scales [229].

Atmospheric circulation patterns (high pressure and anticyclonic patterns) have been shown to exert influence on the occurrence of droughts [230]. 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 [231]. Thus, understanding such relationships improves drought monitoring [71].

There are various oceanic indices to account for temporal and spatial relationships between ocean-atmosphere dynamics and climate (i.e., teleconnection patterns), especially in the case of predicting droughts that are located in different parts of the world. In this research, for investigating the impact of oceanic-atmospheric

dynamics and climate-drought interactions (i.e., teleconnection patterns), the Atlantic Multi-decadal Oscillation Index (AMO) [216, 217], Pacific Decadal Oscillation (PDO) [216], North Atlantic Oscillation (NAO) [216, 218, 219], Multivariate ENSO Index (MEI) [216, 220], and Pacific North American Index (PNA) [216] were selected and included in DroughtObject modeling. The statistical relationships and the strength of these attributes to the DroughtObject model were also tested.

4.4.3 NOAA AVHRR Satellite Attribute

NDVI data was obtained from NOAA AVHRR [232]. NOAA is owned by the U.S. government. The sensor on board NOAA missions that is relevant for earth observation is the AVHRR. The datasets are 10-day composites of daily data (red, near-infrared (NIR)), mapped to a global equal area projection [185]. There are three 10-day composites per month; the first is for days 1 through 10, the second is for days 11 through 20, and the third is for the remaining days. The data contain NDVI, a highly correlated parameter to surface vegetation, derived from the visible and near IR channel reflectance [185, 210]. This pathfinder dataset has gone through many stages of calibration and correction [233].

From NOAA AVHRR historical data for this study, a total of 24 years of time series (1983–2006) of monthly composited 8km AVHRR NDVI data was used to calculate the vegetation related metrics. The monthly NDVI data was obtained by summing (the three dekadal data) from the previously compiled dekadal data. The three dekadal data were summed, since it was believed that the effect of agricultural drought can easily be observed in the overall monthly NDVI values. This dataset is also one of the major dynamic data, and the standardized deviation of the NDVI (SDNDVI) value of a specific time period from its historical mean value is assumed to represent agricultural drought [49, 71, 149].

4.5 Data Pre-processing and Transformation

The data pre-processing step in KDD includes all the methods applied to the data to make them fit for the study [7]. The data obtained from different sources were

checked for accuracy and precision. In the data pre-processing step, data cleaning (noise removal) and data reduction, including transformations and projections to their respective geographical locations, were conducted.

Originally, the data were found in different formats and spatial resolutions, and all of them were resampled to 8km so that they could geometrically fit with other datasets during analysis and model development. In the entire modeling experiment in this research, there were two major types of attributes: static and dynamic. Static attributes were the ones that do not change in the modeling experiment. These attributes were DEM, land cover, ecosystem type (veg_Ethiopia) and WHC. Their values remain the same in the modeling process. On the other hand, the dynamic attribute values change in the modeling experiment for each monthly period. These attributes were three-month SPI, SDNDVI, PDO, AMO, NAO, PNA and MEI. The values of the dynamic variables change for each monthly period during the modeling experiment.

4.5.1 SPI Data Pre-processing and Transformation

The three-month SPI was obtained from RFE data, which were blended with rainguage rainfall measurements in Ethiopia. Using this RFE image data, a total of 2812 point data were extracted for Ethiopia (almost one point data for every 20km). Using these point data, the SPI was calculated.

The SPI was 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 [49]. Since the SPI is equal to the Z-value of the normal distribution, McKee et al. proposed a seven-category classification for the SPI: extremely wet ($z > 2.0$), very wet (1.5 to 1.99), moderately wet (1.0 to 1.49), near normal (-0.99 to 0.99), moderately dry (-1.49 to -1.0), severely dry (-1.99 to -1.5), and extremely dry (<-2.0).

Computation of the SPI involves fitting the Gamma probability density function to a given frequency distribution of precipitation totals for a station. The probability density function of the gamma distribution can be expressed in terms of the gamma function parameterized in terms of a shape parameter. The alpha and beta

parameters of the Gamma probability density function were estimated for each point location, for each time scale of interest (1, 3, 6, 9, 12 months), and for each month of the year. The Gamma distribution was defined by its frequency or probability density function using equation 2 [145, 234]:

$$g(x)=\frac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{\frac{-x}{\beta}} \qquad \textbf{Equation 2}$$

where $g(x)$ is probability function, x is the precipitation amount, $\Gamma(\alpha)$ is the gamma function, α and β are shape and scale parameters, respectively.

Maximum likelihood solutions were used to optimally estimate α and β using equation 3 [235].

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

where n is the number of precipitation observations.

The resulting parameters were then used to find the cumulative probability of an observed precipitation event for the given month and time scale for the station in question. Since g(x) is undefined for x=0 and a precipitation distribution may contain zeros, the cumulative probability is obtained using equation 4 [235]:

$$H(x)=q+(1-q)G(x) \qquad \textbf{Equation 4}$$

where q is the probability of a zero and $G(x)$ is the cumulative probability of the incomplete gamma function. If m is the number of zeros in a precipitation time series, then q can be estimated by m/n. By applying Equation 4, errors are eventually introduced to parameters a and b of the Gamma distribution. These errors depend on the number of months with null precipitation (x=0), and they are evident only for the 1-month precipitation. For larger time scales (e.g., 3-month, 6-month, etc.), the probability of null precipitation is usually zero [145].

After its computation, the cumulative probability, H(x), is transformed to the standard normal random variable Z with mean equal to zero and variance of one, which is the value of the SPI. SPI value is more easily obtained computationally

using an approximation that converts cumulative probability to the standard normal random variable Z using equation 5 and 6 [236].

$$Z = SPI = -\left(t - \frac{C_0 + C_1 t + C_2 t^2}{1 + d_1 t + d_2 t^2 + {}_{d_1} t^3} \right) \text{ for } 0 < H(x) <= 0.5 \qquad \textbf{Equation 5}$$

$$Z = SPI = \left(t - \frac{C_0 + C_1 t + C_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \text{ for } 0.5 < H(x) <= 1 \qquad \textbf{Equation 6}$$

where $t = \sqrt{\ln \frac{1}{(H(x))^2}}$ for $0 < H(x) <= 0.5$ and $t = \sqrt{\ln \frac{1}{(1.0 - H(x))^2}}$ for $0.5 < H(x) <= 1$. Estimations are done for the variables C_0=2.515517, C_1 = 0.802853 1, C_2 = 0.010328, d_1 = 1.432788, d_2 = 0.189269 and d_3 = 0.001308.

The SPI was incorporated to represent meteorological dryness during the growing season. Because this PhD study was primarily focused on the effect of precipitation deficits on the spatial patterns of vegetation health and vigor, the analysis has been restricted to a time period roughly aligned with the growing season (June to October).

From the SPI equation, it was observed that the three-month SPI of a given month would use the precipitation total of the previous two months and also the precipitation total of that specific month of interest. For example, the SPI of September uses the precipitation total of July, August and September itself. In the current research, the three-month SPI was used as a model input for getting the optimum estimation of short-term drought conditions.

4.5.2 NDVI Data Pre-processing and Transformation

The NDVI data preprocessing and transformation include image pre-processing, coordinate system creation, geo-referencing each individual image and re-sampling the image for matching NOAA AVHRR 8km data. A sample workflow for this process is presented in Figure 4.4.

Historical NOAA AVHRR data was obtained from the FEWS NET website [32]. The metadata for this image was well documented and accordingly the pre-processing was done. The data was in 8-bit format and according to the metadata,

water and cloud pixels were with value 255 and bad NDVI with value 253. Taking these into consideration, the image values were converted to an NDVI value based on the recommendation of the metadata using an ILWIS 3.7 software script as NDVI.mpl: = maplistcalculate ("@1/250").

A NOAA AVHRR image was found for all of Africa and a sub-map of Ethiopia was re-sampled from this image. For this purpose, a new geo-reference with GeoRef corners in ILWIS 3.7 software was created. Using this new geo-reference, all the NOAA AVHRR historical image data was re-sampled and the process was automated using a script in ILWIS software. This re-sampled image data was used as input to the subsequent DroughtObject analysis.

A total of 24 years' time series (1983–2006) of monthly composited 8km AVHRR NDVI data was used to calculate the vegetation-related metrics. The modeling period from 1983 to 2006 was used because NOAA AVHRR data was fully available in these time periods. From the year 2006 onward, there were missing NOAA AVHRR data in the growing season, June to October.

The monthly NDVI data was obtained by summing the three dekadal data from the previously compiled dekadal data. The three dekadal data were summed because it was believed that the effect of the drought could easily be observed on the overall monthly NDVI values. A script was written in ILWIS 3.7 software to automate the data extraction process.

The NDVI data calculated at monthly time periods was integrated for the monthly growing periods using equation 7. In order to calculate the cumulative NDVI (CNDVI), the starting month of the vegetation growth was taken as June and the end of the growing season was taken as October. Then, the CNDVI was calculated for all the growing months (June to October) for all 24 years in the time series.

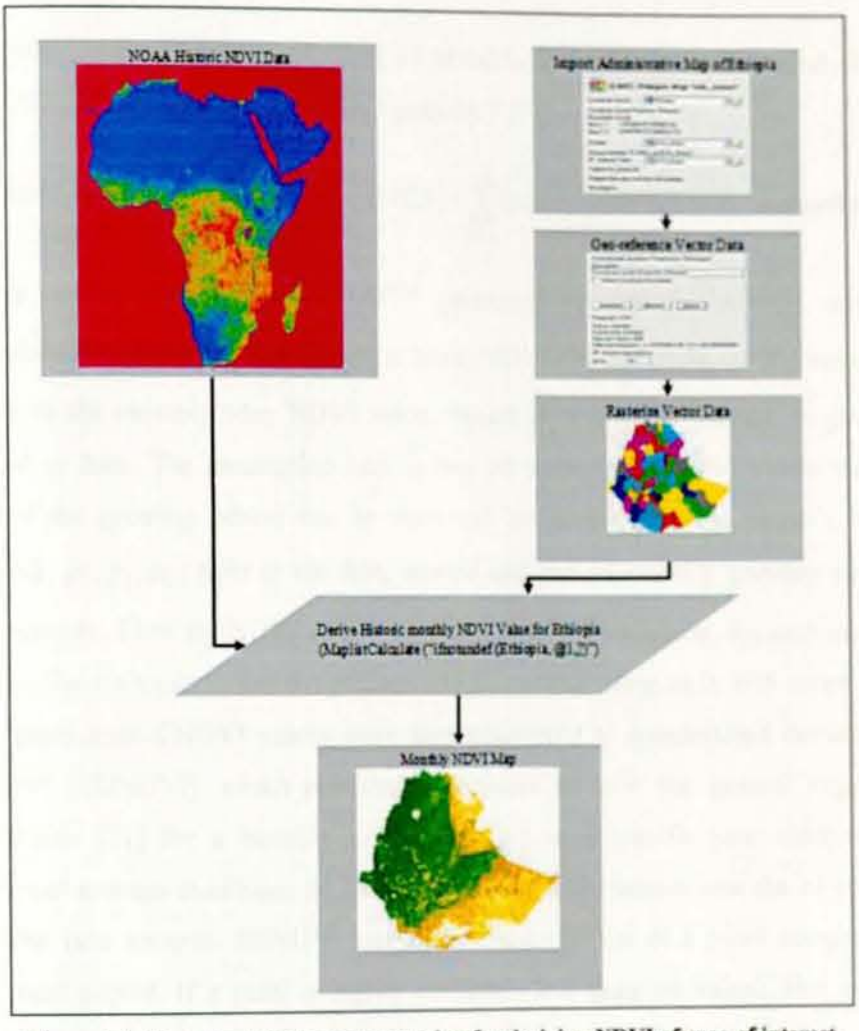


Figure 4.4: Flow chart for image processing for deriving NDVI of area of interest.

In this analysis, CNDVI represents the accumulated NDVI above a baseline (i.e., latent or background NDVI) within each monthly growing period. To monitor vegetation conditions, sequential NDVI values across the growing season were summarized using CNDVI, which can be used as a general proxy for vegetation performance (i.e., gross primary productivity (GPP) [72, 237]). Here, the CNDVI represents the integrated NDVI above a baseline latent NDVI value (representative of the non-vegetative background signal from soil and non-photosynthetic litter) from the start of the growing season to a specific time during the year [71].

High CNDVI values reflect high green biomass conditions, whereas low CNDVI values reflect lower biomass levels [72]. In this study, the CNDVI values were

calculated at the 8km pixel level for all of Ethiopia for the growing period of June to October from 1983 to 2006 using equation 7 [71].

$$CNDVI = \sum_{P_1}^{P_n} NDVI_i = \sum_{P_1}^{P_n} (NDVI_i - NDVI_b) ... + \sum_{P_{n-1}}^{P_n} (NDVI_i - NDVI_b) \qquad \textbf{Equation 7}$$

where $CNDVI$ is cumulative NDVI (seasonal greenness), $NDVI_i$ is monthly normalized vegetation index, $NDVI_b$ is latent NDVI (in this study, $NDVI_b$ value was taken as the monthly May NDVI value, which is before the start of the growing period in June. The assumption here is that all background NDVI values that are out of the growing period can be removed by subtracting this month's NDVI values). p_1, p_2, p_n, refer to the first, second and end of monthly growing periods, respectively. Then the NDVI values were extracted for each year, for each growing month (June–October), and the process was automated using an ILWIS script.

The pixel-level CNDVI values were then converted to standardized deviation of CNDVI (SDNDVI), which provided a measure of how the general vegetation conditions [71] for a monthly growing period in a specific year compared to historical average conditions for that same period in the season over the 24 years of satellite data records. SDNDVI can show the condition of a pixel compared to historical period. If a pixel is highly deviated (low negative value), that pixel is showing the drought condition.

SDNDVI values were calculated at monthly time steps during the growing season for each year using equation 8 [149].

$$SDNDVI = \frac{CNDVI_i - \overline{CNDVI}}{\sigma} \qquad \textbf{Equation 8}$$

where $SDNDVI$ is standardized deviation of NDVI, CNDVI is the cumulative NDVI at a particular period within a growing month for a specific year, $\overline{CNDVI}$ is the 24-year historical mean cumulative seasonal NDVI for the same period, and σ is the standard deviation of the historical record.

After DroughtObject was well characterized using SDNDVI, for each monthly growing period (June to September), point data were extracted using the point data

extraction tool of ArcGIS 10 [203] for the historical 24-year period. These point data were used for the model building experiment.

4.5.3 Oceanic Indices Pre-processing and Transformation

As indicated above, the oceanic indices in this research include PDO, AMO, NAO, PNA and MEI. These indices were found to have a constant value for a monthly time period. These values were changed to grid images using the create constant raster tool of ArcGIS 10 [203]. This tool creates raster in which the output values are based on a constant value of the monthly indices obtained from the NOAA [216] website. Since the values were floating points, the constant raster were multiplied by 100 to convert them to integer value. This was done to remove the decimal values for Mapcubist software during model implementation and knowledge construction (since Mapcubist accepts only integer values). After this preprocessing, point data were created from each oceanic index for all of the 2812 points selected for model input for the whole 24-year time period.

4.5.4 Geo-referencing and Data Extraction

All the biophysical, climate and satellite data were preprocessed and structured in such a way that they fit geometrically. For this purpose, one standard data frame was created on ArcMap [203] and all the data were converted to this standard data frame. This data frame was defined based on NOAA NDVI metadata, which was described by the FEWS NET website [32]. The data frame was defined with projection: Albers, false easting: 0.00, false northing: 0.00, central meridian: 20, standard parallel_1:-19, standard parallel 2: +21, latitude of origin: 1 and linear unit: meter. All the data were then added to one data frame, which had this standard coordinate system, and then exported with the defined data frame spatial reference as GRID data format with cell size resolution of 8km. The spatial resolutions of all the data were converted to 8km, because the main modeling data input NDVI was obtained from NOAA AVHRR, with a spatial resolution of 8km for the whole globe [32].

After preprocessing and geometrically fitting all the data inputs from biophysical, climate and satellite data sources, a 2812-point map was created on ArcMap [203]

and overlaid on the datasets. First, the static data (DEM, land cover, WHC, veg_Ethiopia) were extracted and added to the point map data as an attributes using the batch processing tool of the ArcMap data extraction tool [203]. These static datasets were used for the whole model building process.

For the dynamic variables, three-month SPI, SDNDVI, AMO, PDO, NAO, MEI and PNA (monthly values for the growing months (June–October) from 1983 to 2006) were extracted and added to the point map of the 2812 points. For the 24-year time period, a total of 67488 records (five months of a year, and for a total of 24 years) were extracted and added to the point map of the 2812 points for the model building experiment.

4.5.5 Assessing the Relationships between RF and NDVI

In the modeling process, the two most important attributes were the NDVI and RF data. To observe the relationships between NDVI and RF, a total of twenty-one 2 x 2 degree grids were produced (Figure 4.5). The historical average RF gauge data and NOAA AVHRR NDVI data inside these grids were calculated and their relationships were assessed. This analysis was done to assess the temporal and spatial strength of the relationships between these two important attributes.

To observe the relationship between RF and NDVI, the RF recorded by all stations inside the grids were averaged from 1982 to 2004 and an average point data was generated. The same procedure was followed for the NOAA AVHRR NDVI values of the 2 x 2-degree grids. The RF patterns inside these grids were also analyzed separately. From this analysis, different patterns were observed for different parts of the country. All the patterns observed are presented in the result section.

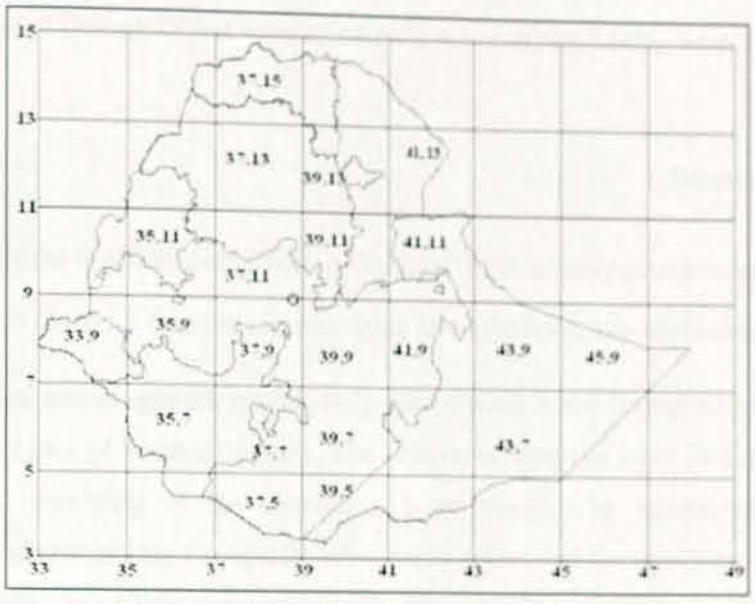


Figure 4.5: Map of Ethiopia with 2 x 2 degree grids.

4.6 DroughtObject Modeling using Data Mining

In this study, the two data mining approaches, ANN and regression tree, were used. These modeling approaches are described in the following subsections.

4.6.1 ANN Model

In the ANN modeling, the network was used for predicting one- to four-month SDNDVI values (agricultural drought) using the identified 11 key attributes of DroughtObject. In this analysis, DroughtObject attributes were received as input signals and were aggregated based on an input function I_i. Then the function I_i generated an output DroughtObject severity extent based on an output transfer function O_i. This conceptual representation of DroughtObject in the ANN model is presented using equation 9 [162].

$$I_i = \sum W_{ij} O_i + \phi_i \qquad \textbf{Equation 9}$$

where I_i is input function characterizing DroughtObject key attributes, O_i is output function characterizing DroughtObject severity extent output of unit i, W_{ij} is connection weight between unit i and j and ϕ_i = bias of unit i.

The output of DroughtObject severity extent i was computed using equation 10 [162].

$$O_i = \frac{1}{1+e^{I_i}} \qquad \textbf{Equation 10}$$

where O_i is the function characterizing DroughtObject severity extent output of unit i, and I_i is the input function characterizing DroughtObject key attributes.

As a typical neural network model [162], the network in the DroughtObject model had three types of processing units. The first units were the input DroughtObject attributes (consisting of the identified 11 attributes). The second units were functions characterizing DroughtObject severity extent and the third units were the hidden units (the weights characterizing the DroughtObject attributes and the output DroughtObject severity extent). The weights associated with these connections (w) constitute what a neural net knows and determine the output from an arbitrary input from the environment [162, 238]. Using this principle, a feed-forward network with an appropriate weight (w) was used to model the relationships between the key attributes and the target DroughtObject. In this process, the network learning algorithm searched through the space of w for a set of weights offering the best fit with the training sample data. From the available learning algorithms in ANN, a back propagation algorithm is an effective learning technique that is capable of exploiting regularities and exceptions in training sample data [238]. This algorithm was also used for drought forecasting [62] in the past, and interesting results were obtained. For these reasons, a back propagation algorithm was used in DroughtObject exploratory analysis and predictions.

In using the back propagation algorithm in the DroughtObject model, the key attributes X_i were fed into the input layer, and an output value of DroughtObject Y_i was then compared with the actual (or desired) output D_i by calculating the squared error $((y_{ij} - d_{ij})^2, 1 \le i \le n,)$ at each output unit. Output differences were summed up to generate an error function E (equation 11) [238].

$$E = \sum_{i=1}^{i} \sum_{j=1}^{n} \frac{(y_{ij} - d_{ij})^2}{2}$$ **Equation 11**

The overall objective here was to minimize E by changing the w, so that all input vectors are correctly mapped to their corresponding output vectors [238].

To build the models for predicting DroughtObject, each network was processed through three stages: (1) the training stage, where the network was trained to predict future data based on the historical 24 years data; (2) the testing stage, where the network was tested to stop training or to keep in training (to avoid overfitting); and (3) the evaluation stage, where the network ceased training and was used to forecast future data and calculate different measures of error. The input nodes are the previous lagged observations while the outputs are the predictions for the future monthly SDNDVI values. Hidden nodes with appropriate non-linear transfer functions were used to process the information received by the input nodes.

The ANN model in this experiment was based on multiple outputs, where several nodes are included in the output layer, and each output node represents one time step to be predicted. For the five plant-growing periods, the prediction started with June and ended with September. June data has the possibility of predicting July, August, September and October (four months); the last month, September, can predict only one month ahead (October).

To get the appropriate ANN models, four steps were followed: assembling the data for the actual training, creating the network object, training the network, and simulating response to new inputs (model testing). For training and testing, a total of 67488 records from the 24 years of historical data were used. Of these, 80% were used for training and 20% for testing the models. The data split, 80-20% (80% for training and 20% for testing) was decided based on past similar research, where improved result was found for 80-20% split compared to other splits, such as 50-50%, 60-40%, and 70-30% [239]. The training and test datasets were randomly selected using the RANDBETWEEN function in Microsoft Excel. After generating the random number, sorting was done to split the training and test datasets. The Neural Network Toolbox on Matlab v7.9.0 [204] was used to build the ANN

models. All the discussed concepts were prototyped and are practically demonstrated in the implementation section of this thesis.

4.6.2 Regression Tree Model

In the regression tree modeling, Classification and Regression Tree (CART) algorithm Cubist software [137], originally developed by Breiman et al. [240], was used. The default splitting rule in CART is the Gini index (or Gini impurity measure), a measure of heterogeneity. The Gini impurity measure at node t is defined using equation 12 [241].

$$g(t) = \sum p(j)p(i)$$

Equation 12

where p(j) and p(i) are the probability of class j and i at node t. The Gini index ranges from 0 to 1; it is equal to 0 if all observations in a node belong to the same class, and it is 1 when different class sizes at the node are equal [240, 242].

In CART software, a maximal tree was grown and then pruned back to obtain the optimal tree by determining the lowest misclassification errors. The maximal tree is always overfit, because it represents all idiosyncrasies of the learning dataset. Error rates from the trees were combined to yield estimated error rates for the nodes in the maximal tree. This allowed for determination of error rates for trees of different sizes, and gave an indication of the optimal tree size [242]. In addition, to prevent overfitting of the tree, splitting was stopped when a terminal node had less than ten cases [137].

The software produces three parameters for evaluation of the output model: the average error, relative error and correlation coefficients. Average error was the error for estimating the average value. The relative error magnitude was the ratio of the average error magnitude to the error magnitude that would result from always predicting the mean value; for useful models, this should be less than 1. The correlation coefficient measures the agreement between the cases' actual values of the target attribute and those values predicted by the model [137].

The rules and instances options available in Cubist were utilized for DroughtObject model development. Some of the available options in the actual model development are presented in Figure 4.6 from GUI of Cubist v202. The Cubist models are

composed of an unordered set of rules, with each rule having the syntax "if x conditions are met then use the associated linear regression model".

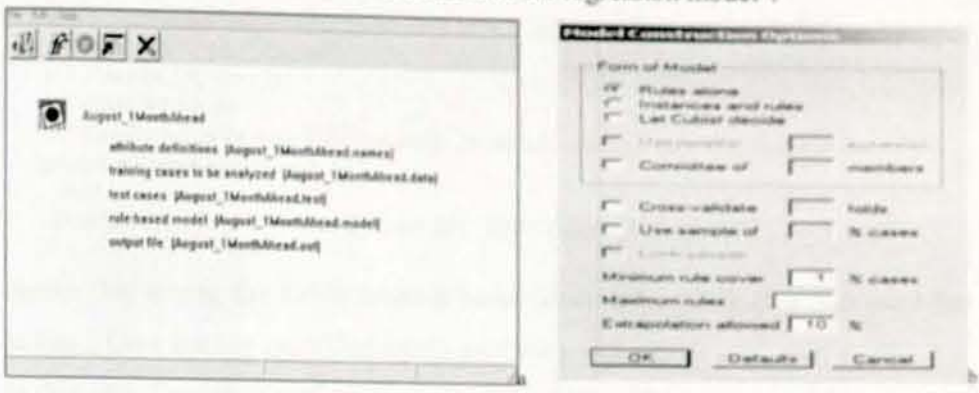


Figure 4.6: Regression tree, (a) GUI of Cubist software, and (b) its model development options.

4.6.3 Regression Tree Rule Interpretation

During the model development, the SDNDVI and SPI attributes were multiplied by 100 to remove the small decimal values. This was done because during the model implementation, Mapcubist does not accept the decimal values. Therefore, the values were standardized, which range from -4 to +4 in terms of standardized values (-400 to 400 after multiplied by 100). For instance, the value 100 DroughtObject in our modeling is similar to 1 SDNDVI or SPI values.

In this research, drought severity (represented by SDNDVI and referring to agricultural drought severity) has seven classes. These classes are similar to the SPI drought classes [28, 29] and have been adapted for vegetation condition to indicate levels of agricultural drought [71]. These seven classes are presented in Table 4.4.

Table 4.4: Drought severity classes using SDNDVI values [71].

No.	SDNDVI Value	Vegetation Condition	Drought Severity Classes
1	-2.0 and less	Extreme Stress	Extreme Drought
2	-1.0 to -2.0	Severe 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

A number of trials with different options were done before selecting a given rule. The criteria for selection of rules were the accuracy level (quantified by correlation

coefficient) and amount of both absolute and relative errors. The interpretation for this rule:

Rule 1: [124 cases, mean -59.0, range -251 to 163, est. err 62.0]
if
3 Month SPI <= -20
SDNDVI <= 0
Land cover in 14 (rain fed croplands), 20 (mosaic cropland (50-70%) / vegetation (grassland/shrub land/forest)
then
***DroughtObject** = 57.6 + 0.83 3 Month SPI - 0.023 DEM + 0.15 SDNDVI*

means that among the 53990 training cases (since 20% of the data was used for testing), there are 124 cases that satisfy all three conditions.

In this case, DroughtObject ranges from -251 to 163 with an average value of -59.0. The regression tree model finds that the target value of these or other cases satisfying the conditions can be modeled by the following formula:

DroughtObject = 57.6 + 0.83 3 Month SPI - 0.023 DEM + 0.15 SDNDVI.

During this modeling, the estimated error is 62.0, and for a case covered by this rule, 3 Month SPI has the greatest effect on DroughtObject, DEM a lesser effect, and SDNDVI the least effect. Each rule generated during the modeling was interpreted this way, and whenever a case satisfied all the conditions, the linear model was used for predicting the value of the target attribute. If two or more rules applied to a case, then the values were averaged to arrive at a final prediction.

4.6.4 DroughtObject Model Implementation

The regression tree model developed with Cubist software was implemented using Mapcubist software. This software was developed at EROS. During the process of model implementation, a mask file was prepared for the spatial extent of Ethiopia. This file was prepared in such a way that all areas inside the Ethiopian boundary have a pixel value 1 and all areas outside the 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. Then the developed models were applied for each pixel using all key attributes of the DroughtObject model.

As indicated in the previous sections, there were two types of input data: static and dynamic. The static data DEM, land cover, WHC, and veg_Ethiopia remained the same throughout the modeling process (their values remain constant for each pixel). The dynamic attributes SPI_3month, SDNDVI, PDO, AMO, NAO, PNA and MEI were prepared for each monthly period for their target predictions. For example, for the month of June, these dynamic variables were prepared and used for predicting July, August, September and October. Then the regression tree DroughtObject models were applied for each pixel using the static and dynamic attributes. For this experiment, the drought years 1984, 2002 and 2011 were used.

During the implementation, Mapcubist software was run from the command line on a Windows XP operating system (Figure 4.7). Accordingly, the developed linear regression equations associated with the rules of the input values were applied to the input values in order to calculate the DroughtObject values for each pixel.

Figure 4.7: Running Mapcubist software on command line on a Windows XP operating system

Besides the regression tree modeling output files, two files were required for running Mapcubist software. These files were .names and .icases. The .names file was the list of both the dependent attribute (SDNDVI) and explanatory attributes (the 11 key attributes) with their corresponding data type for the DroughtObject model (Figure 4.8a). The .icases file (Figure 4.8b) was the list of both the dependent attribute and explanatory attributes with the same order as the .names file, and its main purpose was to present the image types and their band numbers for Mapcubist software.

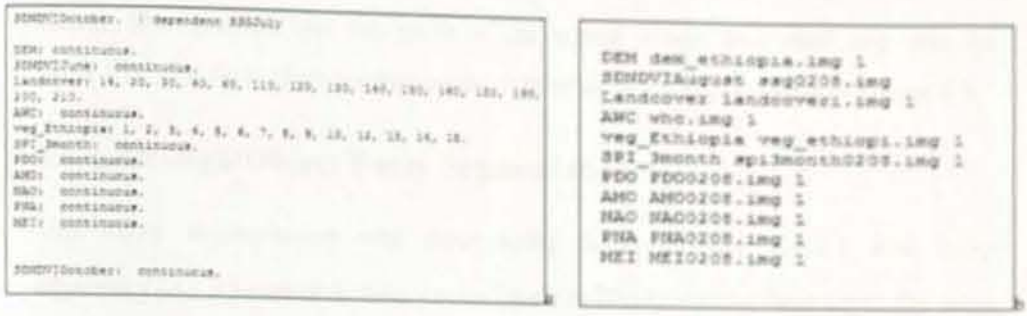


Figure 4.8: Sample file types for (a) .inames and (b) .icase

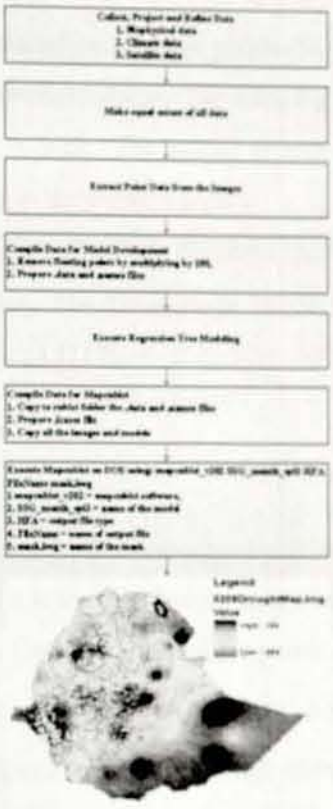


Figure 4.9: Flow chart for regression tree model implementation.

The .names, .icases, regression model outputs, static attributes, and dynamic attributes were collected and put in the same folder as Mapcubist software. This software was run on the DOS "mapcubist_v202 MonthlyPredictionModel HFA DroughtMap mask.img" command line. In this command line, mapcubist_v202 was Mapcubist software version 202, MonthlyPredictionModel was the name of the regression tree model, HFA was an ERDAS image (.img) format for the model

output, DroughtMap was the name of the output image, and mask.img was the mask file. The flow chart for this model implementation is presented in Figure 4.9.

4.7 DroughtObject Fuzzy Segmentation

The fuzzy segmentation was done using fuzzy affinity $\psi[c,d]$ and fuzzy adjacency $\mu(c,d)$ concepts. Fuzzy rules (such as if two regions have about the same gray value and if they are relatively close to each other in space, then they likely belong to the same object [117, 118]) were used in this experiment.

In this analysis the basic assumption was that pixels (SDNDVI pixels) seem to hang together when forming a DroughtObject. The hanging togetherness property was then represented using fuzzy logic. The spatial relationships were determined for each pair of image elements in the entire image.

The fuzzy adjacency $\mu(c,d) \in [0,1]$ of two elements (pixels) c, d was determined by the fuzzy adjacency function. The 2D drought (SDNDVI) fuzzy pixel adjacency was obtained using equation 13 [117].

$$\mu(c,d)=\begin{cases} \frac{1}{1+k_1\sqrt{\sum_{i=1}^{n}(c_i-d_i)^2}} & \text{if } \sum_{i=1}^{n}|c_i-d_i|\le n, \\ 0 & \end{cases} \qquad \textbf{Equation 13}$$

where c and d are pixel values, k_1 is a non-negative constant, c and d are any pixels in the SDNDVI image, and n is the total sample size (the image size).

The affinity function $\psi[c,d]$ was determined for pixels c, d that are fuzzy adjacent (i.e., which have adjacency value $\mu(c,d) \neq 0$). Fuzzy connectedness μ_ψ is a global fuzzy relationship that assigns every pair of image elements c and d a value in the interval [0, 1] based on the affinity values ψ along all possible paths between these two image elements. The elements c and d are not expected to be nearby. They are connected by a path $\pi = \left(c^{(n)},\ldots,c^{(N)}\right)$ of pixels, with $c = c^{(0)}$ and $d = c^{(N)}$. Each pair of consecutive pixels is characterized by a fuzzy affinity $\psi\left(c^{(n)},c^{n+1}\right) 0 \le n \le N-1$ [117]. For each path, its strength was defined as the minimum affinity value of all pair wise consecutive elements on the path, so the

strength of the entire path is defined by the strength of its weakest local connection, and quantified by equation 14 [117].

$$\psi'(\pi) = \min_{0 \le n \le N-1} \Psi\left(c^n, c^{n+1}\right) \qquad \textbf{Equation 14}$$

Many different paths may connect two pixels c and d. Let M denote the set of all paths joining c, d. Then the fuzzy connectedness was derived using equation 15 [117].

$$\mu_{\psi(c,d)} = \max_{\pi \varepsilon M} \Psi'(\pi) \qquad \textbf{Equation 15}$$

The local fuzzy relation is called fuzzy affinity, denoted by $\psi[0,1]$ [117], and represents a strength of hanging togetherness of nearby image elements. The affinity is a function of the spatial distance between two fuzzy adjacent image elements considering their image intensities or other image-derived properties. As such, any image I can be represented by a pair, $I = [C, F]$[117]. In our context, C represents the image domain (SDNDVI values, which range from -400 to +400) and F represents local image properties (the value contained by the target pixels). Then, $f(c)\varepsilon[0,1]$ represents a normalized image property (feature) associated with pixel c. The fuzzy affinity $\psi[c,d]$ is a function of fuzzy adjacency $\mu(c,d)$, pixel properties $f(c)$, $f(d)$, and in spatially variant cases of c and d, and in this case was obtained using equation 16 [117, 118].

$$\Psi(c,d) = \frac{\mu(c,d)}{1 + k_2 \left| f_{(c)} - f_{(d)} \right|} \qquad \textbf{Equation 16}$$

where μ is the fuzzy adjacency defined by equation 13 and k_2 is a non-negative constant.

DroughtObject identification from images in this study has two major steps: segmentation (identification) and classification. The segmentation step was the process of determining roughly the whereabouts of the object in the image, whereas classification, was the process that defines the precise spatial extent and composition of the DroughtObject in the image.

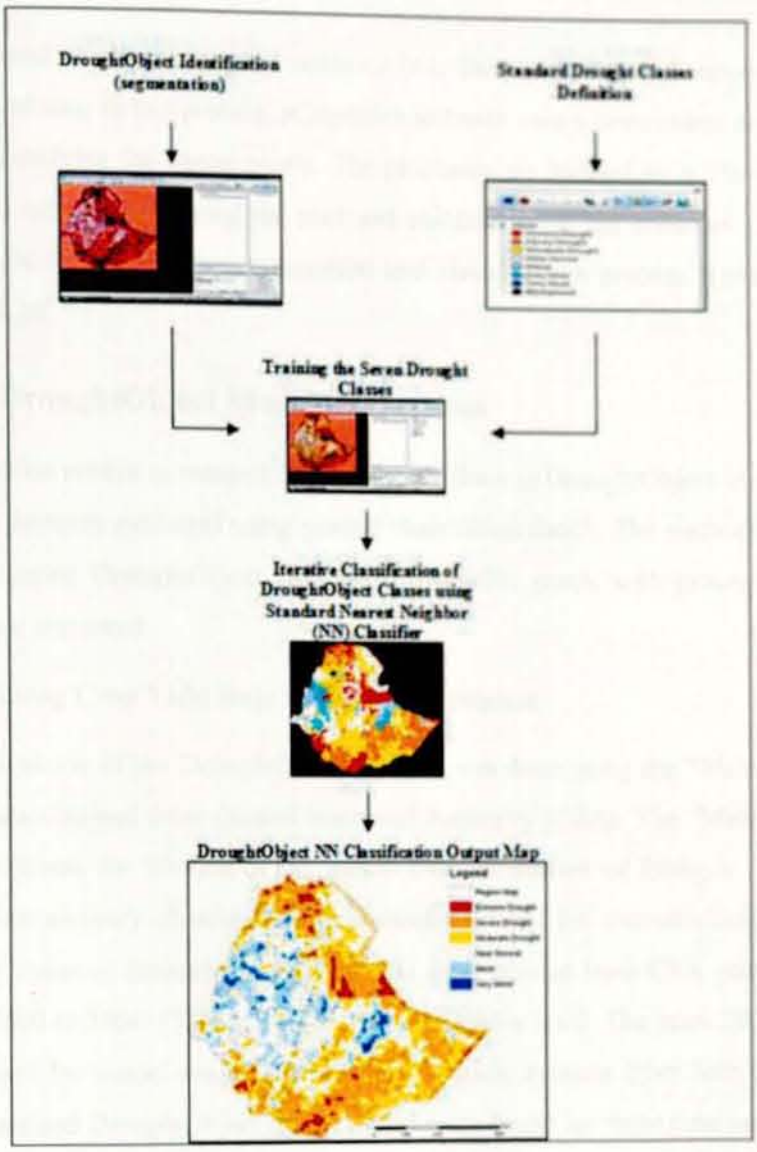


Figure 4.10: Major steps in DroughtObject identification and classification.

eCognition Developer software [243] was used for this experiment. In this experiment, the segmentation algorithm groups similar and adjacent pixels together into polygons or image objects and uses a bottom-up, low-to-high value region growing approach. In this multi-resolution segmentation algorithm, a seed pixel was selected from within the image, and then neighboring pixels with similar values were consolidated into objects based on the smallest growth of heterogeneity. The segments (DroughtObjects) were then classified through fuzzy

logic based object metrics and statistics [61, 243] such as pixel values, texture, shape, and size. In this process, eCognition software uses a hierarchical process tree when classifying the image pixels. The processes are defined by a 'Parent-Child' concept, with parents being the root and children being the branches [243]. The work flow for the image segmentation and classification process is presented in Figure 4.10.

4.8 DroughtObject Model Evaluations

This section relates to research question three (how is DroughtObject identified on satellite imagery evaluated using ground observation data?). The methods followed for evaluating DroughtObject identified on satellite pixels with ground observed yields are presented.

4.8.1 Using Crop Yield Data for Model Evaluation

The evaluation of the DroughtObject product was done using the "Meher" season yield data obtained from Central Statistical Authority (CSA). The "Meher" season yield accounts for 90-95% of the annual crop production of Ethiopia [244]. The yield data anomaly observed in this season is crucial for characterizing the food security status of Ethiopia. Crop yield data was obtained from CSA yearly reports (from 2000 to 2006) [245] at the zonal administrative level. The years 2000 to 2006 were used for model evaluation because complete datasets from both CSA crop yield data and DroughtObject model output were found for these time periods. The yield data was also obtained using standard statistical data collection and analysis by CSA of Ethiopia within these time periods.

After standard statistical data processing, CSA [245] produced the yield data at the zonal (district) administrative level. The zonal administrative level in Ethiopia is next to the regional administrative level. Currently there are 72 zones in the country. From the 72 zones, 41 were selected for the current study (Figure 4.11). The zones were selected based on two criteria: 1) availability of yield data in the years 2000 to 2006 for the "Meher" season, and 2) location of the zones in the

"Meher" crop growing district where annual cereal crops are grown (in the growing months June to October).

From the 41 selected zones, the cereals yield data was considered for DroughtObject evaluation. According to CSA [245], cereals are the major food crops both in terms of the area in which they are planted and the volume of production obtained in the selected zones. Cereals are also produced in larger volumes compared to other crops because they are the principal staple crops in these zones.

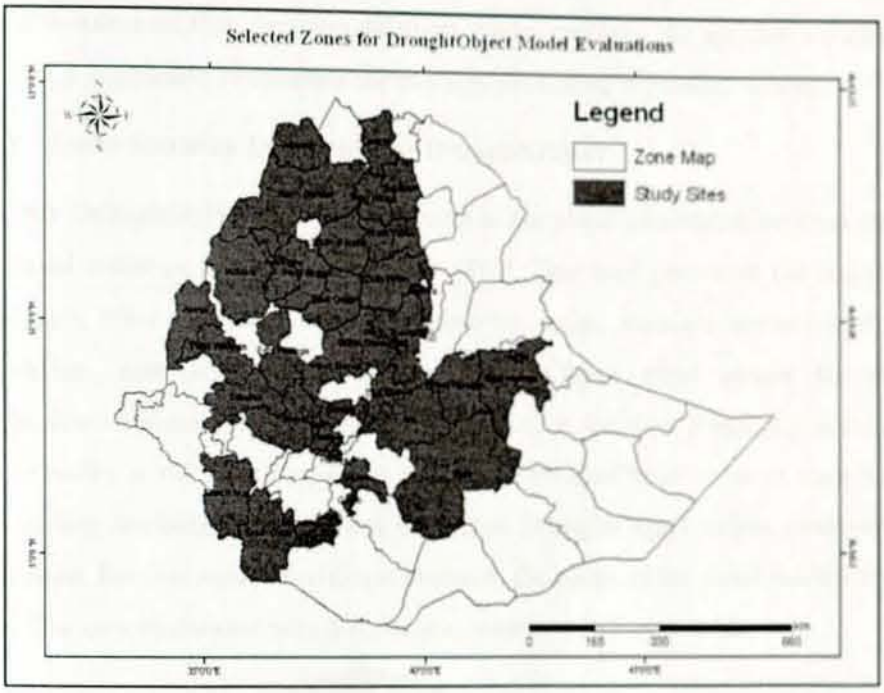


Figure 4.11: Spatial distribution of zones (districts) on the map of Ethiopia.

The yield data that was obtained from CSA was assessed for its suitability for DroughtObject model evaluation. A trend of yield data increase was observed for almost all crop yields assessed. Analysis detected a trend toward crop yield temporal increase due to farm management practices; this has to be detrended in order to use the data for model evaluation.

The detrending of yield data was done in two steps. The first step was fitting a smooth curve through the yield statistics. This kind of fitting is described by Wu et al. [246], where the detrending is the operation of removing the trend. The second

step was calculating the actual detrended yield. Matlab software v 7.9 [204] was used for this deterending analysis.

Detrending in this research is the operation of removing the crop yield increasing trend from 2000 to 2006, and after detrending, what remains is the variability of the residue of the crop yield data. It was believed that determining trend and implementing detrending operations are important steps in our data analysis. This is because the detrended crop yield data define a more meaningful variability of the yield variation caused by drought (weather variability) through the study periods. It was also assumed that the DroughtObject model captured the weather variability and that it is possible to correlate the two attributes using regression analysis.

4.8.2 Zonal Statistics Extraction for DroughtObject

The raw DroughtObject data was extracted at the zonal administrative level using the zonal statistics tool of ArcGIS 10.0 [203]. This tool generated the majority, maximum, mean, median, minimum, minority, range, standard deviation of the population, sum and variety of the DroughtObject pixel values for each administrative district. In this output, the majority is the most frequently occurring and minority is the least frequently occurring DroughtObject value in each zone. The variety attribute is the number of unique DroughtObject values observed in each zone. For this model evaluation research, the mean of the zonal statistics was used. The overall data extraction process is presented in Figure 4.12.

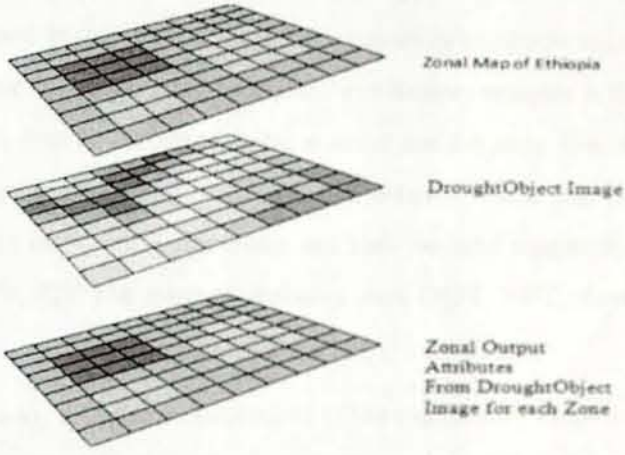


Figure 4.12: Data extraction for DroughtObject zonal statistics (adapted from ESRI [203]).

Chapter 5 Results and Discussions

The results and discussions in this chapter are presented based on the order of the research questions. The answer to the first research question was found by reviewing and analyzing past research for identifying the key attributes for DroughtObject modeling (see Chapters 2, 3 and 4). The key attributes identified from past research were statistically tested for their relationship with the DroughtObject target variable (see subsection 5.1). The results of the second research question and the actual modeling of DroughtObject are presented in subsection 5.2. The evaluation of DroughtObject identified on satellite images using crop yield data, which is a response to the third research question, is presented in subsection 5.3. In each subsection, detailed discussions of the results are presented. Finally, subsection 5.4 presents application of DroughtObject models for assessing the 2011 drought status in Ethiopia. A way to use the developed models for 2012 and future years' drought predictions is also recommended.

5.1 DroughtObject Model Attributes Selection

To select the key attributes, all attributes that were identified from past research were statistically explored for their relationships with the new DroughtObject model developed in this study. Since 11 explanatory attributes were identified, it was not possible to present all the statistical explanatory analysis in this subsection. For this reason, four key attributes were selected and the steps followed for all key attributes are presented using these sample attributes. These four attributes were also assumed to be the most significant and also the most repeatedly used by past research [49, 71, 72]. The selected attributes were DEM, WHC, three Months SPI and SDNDVI.

In an iterative way, different combinations of the explanatory variables were tested. Some of the iterative combinations are presented in Table 5.1. In all of the combinations, the expected signs of the coefficients have been checked. All of the

explanatory variables were found to have coefficient values as expected. For instance, as the three-month SPI values decrease, the SDNDVI values have to decrease; conversely, as the three-month SPI values increase, the SDNDVI values have to increase. This is because both values are showing the drought and non-drought situation of a given location as confirmed by past research [49, 71, 72]. Therefore, these attributes can be reliably used for representing DroughtObject.

The 3D scatter plots are presented in Figure 5.1. The blue points are the scatter points. The 3D scatter surface represents the two predictor independent variables and the dependent variable DroughtObject. Interestingly, all of the 3D scatter plots showed a strong statistical relationship between the two predictors and the dependent variable DroughtObject (Figure 5.1a–f). The best scatter plot was observed for Figure 5.1d (using SDNDVI August, August three-Months SPI attributes for predicting DroughtObject). These attributes are the two major attributes obtained from NDVI and rainfall estimates showing the drought severity in space-time dimensions. Detailed investigations are also presented on these two attributes in subsections 5.1.1 and 5.1.2.

The R^2 values were used as a model fitness criterion. From the various combinations, the R^2 ranged from 0.11 to 0.85. This means that these models can explain from 11% to 85% of the DroughtObject variability. It was also realized that as the number of attributes increase the R^2 was found to increase. The highest R^2 (i.e., 0.85) was found when SDNDVI August was included in the various combinations. SDNDVI of August 1984 alone can also explain about 82% of the variability of SDNDVI of September 1984. This is expected because the NDVI value of the previous month (August) is correlated with the following predicted month (September).

The Akaike's Information Criterion (AIC) ranges from 6797 to 8017. The AIC model selection criterion is based on the fact that there is no single model that represents the whole truth or complete information about the phenomenon under investigation, and models only approximate reality [205]. Based on this principle, Akaike [206] developed a relative index for getting a model that would best

approximate reality using some set of explanatory variables. Since it is a relative value (the smaller the value, the better the model is) for comparing models [206, 247], the AIC in this research was used to compare the models for relatively better representation of the DroughtObject. Interestingly, the least AIC value was obtained when all the four attributes were combined for modeling DroughtObject (Table 5.1). This indicates the strong relationships of the explanatory variables with the dependent variable. As the number of explanatory variables increases, DroughtObject variability is captured in the model.

Table 5.1: Experimental result of key attributes selection for modeling DroughtObject.

No.	Independent attribute (s)	Dependent attribute	Adjusted R^2	AIC
1	DEM, WHC	SDNDVI September	0.11	8006.2
2	DEM, SDNDVI August	SDNDVI September	0.83	6869.7
3	DEM, August 3 Month SPI	SDNDVI September	0.19	7939.3
4	SDNDVI August, August 3 Month SPI	SDNDVI September	0.84	6811.6
5	SDNDVI August	SDNDVI September	0.82	6886.7
6	WHC, August 3 Month SPI	SDNDVI September	0.09	8016.5
7	WHC, SDNDVI August	SDNDVI September	0.82	6887.8
8	DEM, WHC, August 3Month SPI	SDNDVI September	0.19	7938.7
9	DEM, WHC, August 3Month SPI, SDNDVI August	SDNDVI September	0.85	6797.2

The statistical analysis output of the model from the four explanatory attributes (DEM, WHC, August three-Month SPI, SDNDVI August) is presented in Table 5.2. As it can be observed, all the coefficients are statistically significant. The Variance Inflation Factor (VIF) value, which is a measure of the severity of multi-colinearity in regression analysis [207], was also found to be low in all the attributes. Multicollinearity here is the statistical phenomenon in which two or more predictor variables in the regression model are highly correlated [207, 208]. Obrien [208] indicated that usually a VIF value greater than 10 shows the existence of multicollinearity. As a guide, ESRI [203] recommended that the VIF value should be less than 7.5 for GIS data. The maximum VIF value recorded in the DroughtObject model using the four attributes is 1.89, which is much less than the recommended value (Table 5.2). Therefore, using the four key attributes, DroughtObject at September 1984 is:

$$DroughtObject = 10.62 - 0.0098DEM + 0.96\ SDNDVI\ August + 0.1179\ WHC + 0.122\ August\ 3Month\ SPI.$$

Table 5.2: Exploratory data analysis output of DroughtObject model

No.	Key attribute	Coefficient	Standard Error	t-statistic	Probability	VIF value
1	Intercept	10.62	2.62	4.05	0.00006	—
2	DEM	-0.0098	0.0024	4.09	0.00005	1.89
3	SDNDVI August	0.96	0.0173	55.29	0.0000	1.29
4	WHC	0.1179	0.035	3.36	0.0008	1.59
5	August 3Month SPI	0.122	0.015	8.08	0.0000	1.09

In the entire modeling experiment in this research, there are two major types of attributes: static and dynamic. Of the dynamic attributes, three-month SPI and SDNDVI are very important. The three-month SPI represents the climate variability and is one of the important independent attributes. The SDNDVI represents the agricultural drought extent and was taken as a dependent attribute. The three-month SPI was obtained from rainfall data sources (from ground observation meteorological stations and satellite rainfall estimates) and SDNDVI was obtained from NOAA AVHRR NDVI data.

Before the detailed modeling experiment, the rainfall (RF) pattern of the Ethiopia was assessed and spatially demarcated. This helps for practical implementation of the models. The strengths of correlation of the two major attributes (RF and NDVI) were also investigated in space-time dimensions to determine the reliability of the models. For this purpose, section 5.1.1 presents the RF pattern and 5.1.2 presents the strength of correlations of RF and NDVI in space-time dimensions in Ethiopia.

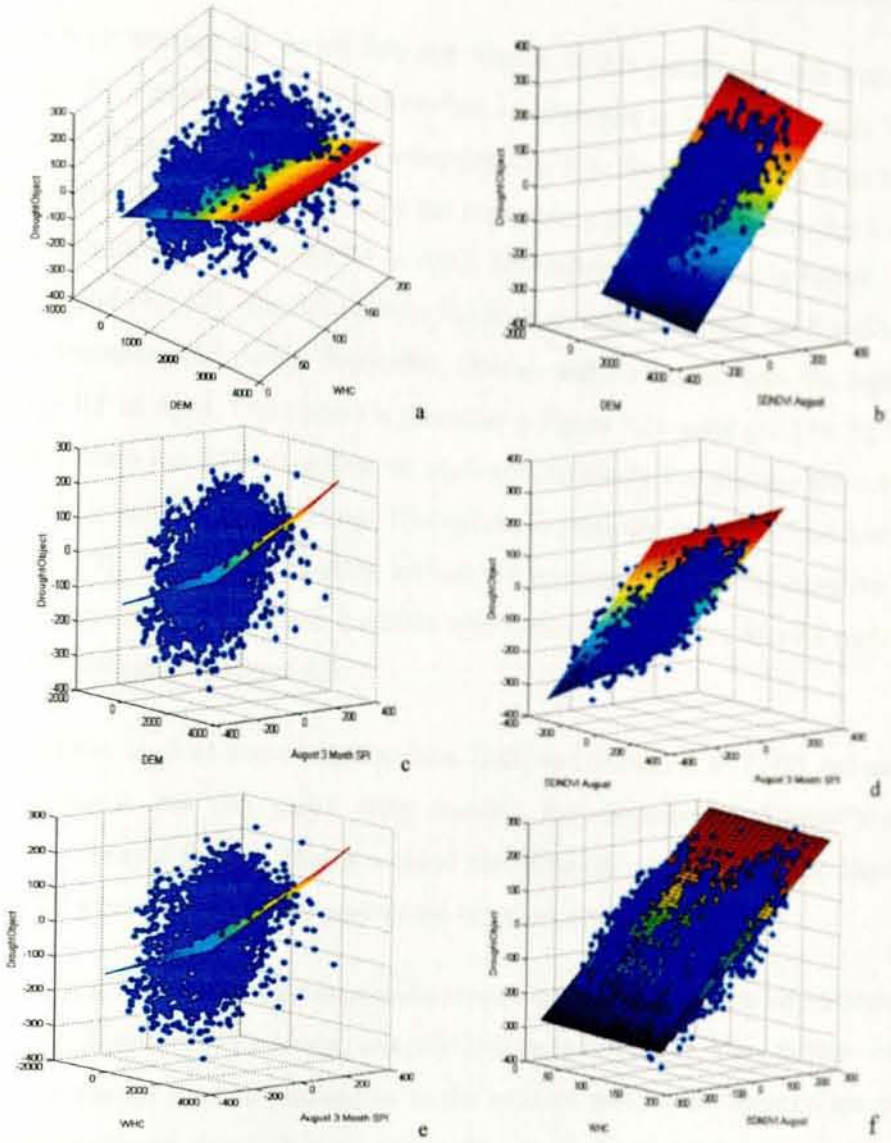


Figure 5.1: 3D scatter plots showing the correspondence between predictor attributes (a) DEM-WHC, (b) DEM-SDNDVI August, (c) DEM-August 3 Month SPI, (d) SDNDVI August- August 3 Month SPI, (e) WHC- August three-Month SPI, and (f) WHC- SDNDVI August for predicting the dependent variable SDNDVI of September 1984.

5.1.1 Rainfall Pattern in Ethiopia

For practically implementing the DroughtObject model, the rainfall pattern in Ethiopia was analyzed and spatially demarcated. In this study, the overall monthly average RF pattern analysis showed that there is a difference between the different grids. From this analysis, four major patterns were identified. The first pattern

shows high average RF during July and August. In this pattern, the rain starts in March with a gradual increase and reaches its maximum in July and August. This pattern is presented in Figure 5.2a using grid [35, 11]. The second pattern has high average RF during July and August and a secondary RF during March, April and May, with the highest average RF in April. This pattern is presented in Figure 5.2b using a grid [41, 13]. The third pattern has high average RF during April and May and a secondary RF during September, October and November, with the highest average RF in April. This pattern is presented in Figure 5.2c using grid [39, 7]. The fourth pattern has RF throughout the year, with relatively low average RF during December, January and February. This pattern is presented in Figure 5.2d using a grid [35, 7]. The overall monthly average RF analysis for Ethiopia using the 21 grids is presented in Appendix 2 and the spatial distributions of the four RF patterns are also presented in Figure 5.3.

Past studies, such as Sileshi and Demaree [248] and Shanko et al. [249], indicated that Ethiopia has two major rainy seasons: June–September ("Kiremt") and February–May ("Belg"). This is a broad classification, and they did not classify spatially where the two major seasons are in the country.

Gissila et al. [250] classified three rainy seasons in Ethiopia: (i) February/March to May, (ii) June/July to September and (iii) October to November. They further noted that the annual rainfall distribution in the western part of the country has one rainfall maximum during July or August. This is similar to rainfall pattern one in our current research. Gissila et al. [250] also indicated that in the eastern part of the country, the rainfall has a bimodal distribution, with either a long-rain/short-rain pattern (similar to coastal equatorial East Africa) or two seasons of roughly equal length and amplitude. Rainfall pattern two (Figure 5.2b and Figure 5.3) agrees with this classification.

A FEWS NET [244] study showed that April–May RF totals could explain 50% of the variance of the long-cycle water requirement satisfaction index, revealing that this is a critical stage when rainfall deficits can negatively impact yields of crops

harvested in September–December. A similar study by Verdin et al. [59] noted that the RF seasons for Ethiopia as a whole are March–May (locally called "Belg"), and June–September (locally called "Kiremt"). They also indicated that most of the rains in Ethiopia come in the period March to September, with a pause in many parts of the country around the end of May or beginning of June. The "Belg" rains come in March–May, and the "Kiremt" rains in June–September. Neither FEWS NET [244] nor Verdin et al. [59] indicated the spatial locations of these two major seasons.

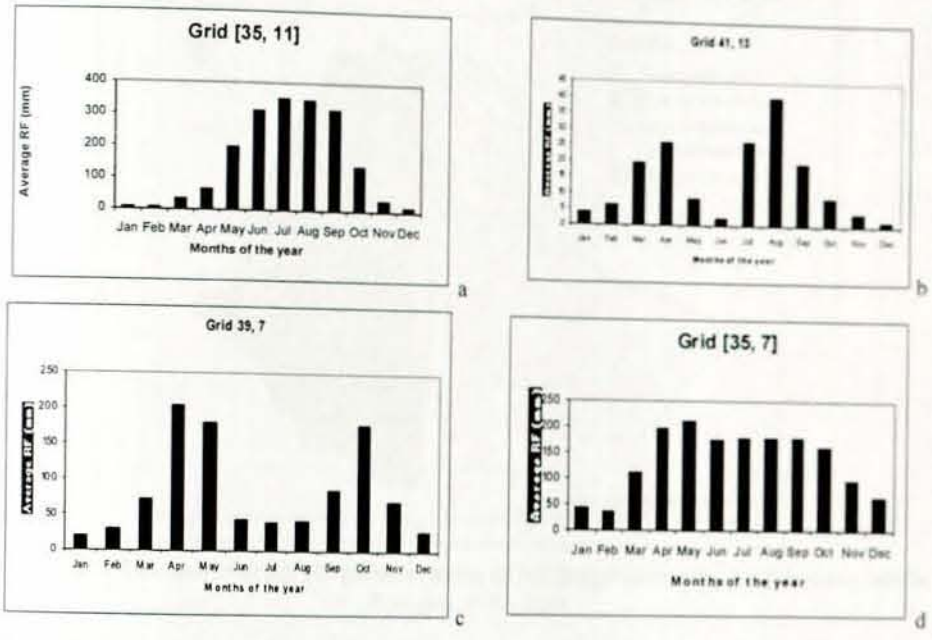


Figure 5.2: Average RF for Ethiopia: (a) pattern 1, (b) pattern 2, (c) pattern 3 and (d) pattern 4.

Diro et al. [251] identified three main rough rainfall regimes. Regime A in the western part of the country has a monomodal rainfall pattern (February to November). Regime B (comprising central and eastern parts) has two rainy periods: February to May ("Belg") and June to September ("Kiremt"). Regime C in the south and southeast has two distinct rainy periods: March to May (long) and October to November (short). In Diro et al. [251], the RF pattern for regime A is similar to RF pattern one (Figure 5.2a), the RF pattern for regime B is similar to RF pattern two (Figure 5.2b), and the RF pattern for regime C is similar to RF pattern

three (Figure 5.2c). But they did not include or describe RF pattern four in the southwestern part of the country (Figure 5.2d).

Results from Sileshi and Demaree [248], Shanko et al. [249], Gissila et al. [250], Verdin et al. [59], FEWS NET [244] and Diro et al. [251] are not refined enough to develop a practical drought monitoring system in Ethiopia. Compared to these coarse classifications, our approach classified the country temporally and spatially (Figures 5.2 and 5.3).

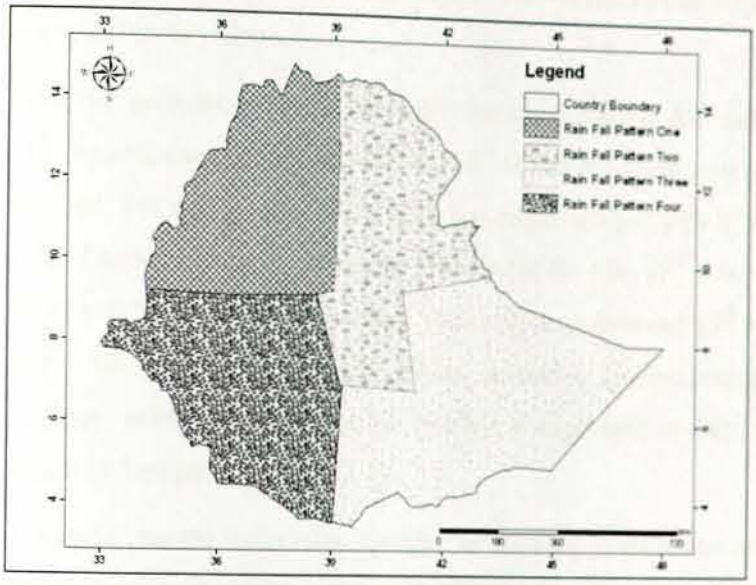


Figure 5.3: The four average RF patterns using all RF gauges inside the 2 x 2 degree grids in Ethiopia, 1982–2004.

5.1.2 Relationships between RF and NDVI in Space-Time Dimensions

Since the relationship between precipitation and vegetation status varies within growing season months, Pearson correlation analyses between RF and NDVI were done for all months of the year separately. The correlation coefficients and their p-values (the probability of rejecting the null hypothesis that there is no correlation between RF and NDVI) were obtained for each of these monthly values. In all the statistical tests in the study, $\alpha = 0.05$ was used as the cutoff for rejection of the null hypothesis.

Figure 5.4 shows the linear regression analysis and scatter plots for the average RF and NDVI for the years 1982–2004. All the monthly correlations between RF and

NDVI were assessed (Table 5.3). For almost all the months, the level of significance for coefficient of NDVI in predicting RF values was significant (at $p<0.05$). The assessed level of significance for constant (intercepts of the models) showed that there were differences in the months (Table 5.3). For June, August and September, the intercepts of the models were found to be significant ($p<0.05$); all other months were not significant. For the month of July, our expectation was that the intercept of the model is significant, since it is one of the rainy months. Unexpectedly, it was found to be not significant. This needs further investigation with actual ground observations.

Overall, from the analysis of the relationship between average RF and average NDVI for 12 months, there were relatively high R^2 values for rainy months (Figure 5.4), as expected. For dry months, there were low relationships (with R^2 values of less than 0.5). The maximum R^2 values recorded were for May ($R^2 = 0.7613$) and September ($R^2 = 0.7698$) and the minimum values were in February ($R^2 = 0.2172$) and November ($R^2 = 0.2095$). The R^2 values showed a gradual increase from February to May, with a maximum value in May; it decreased in July, and then increased again in September (Figure 5.4).

The result showed that the main rainy months, which start in mid-June and end in mid-September, have strong relationship with NDVI values (p-value < 0.05). From these analyses, it was found that the decrease or increase in RF value could be tracked using NDVI values.

The overall pattern indicates that the correlation between RF and NDVI is low during dry months. It starts to increase when the rain starts, and reaches maximum when plants are in vigorous growing stages. It was also observed that the impact of RF on vegetation does not occur instantaneously, but is cumulative of the past months. The significant values indicate a positive response of vegetation to precipitation, which can be captured by NDVI-RF models. The R^2 values in June to September, 0.60 to 0.77, respectively, means from 60% to 77% of the variation in RF can be explained by NDVI.

A study by Ji and Peters [49] indicated that vegetation has its strongest response to moisture availability during the peak greenness period. The authors also found that the impact of precipitation on vegetation does not occur instantaneously, but is cumulative. In most cases, precipitation occurring in one month does not strongly affect vegetation in that month, but the response is notable over periods longer than one month. Our results showed the same as their study in that there were gradual increases in the R^2 values from the beginning of the rainy month to the peak rainy month.

Detailed studies by Dai et al. [56] have determined that the peak response of NDVI to precipitation was with 14–25 days of delays, and the response duration was 22–61 days during the growing season. Yang et al. [252] found the NDVI response lag to precipitation was five to seven weeks. Other studies by Rundquist et al. [253] found the lag time of NDVI response to precipitation was approximately one month, and more precisely Wang et al. [254] showed a lag of four to eight weeks. Since the present study was only using average monthly NDVI and RF values, the response of vegetation to RF in the context of time lag was not assessed. This is a subject for future research, specifically on analyzing the time lag effect of drought (vegetation stress) caused by lack of RF.

Table 5.3: Statistical comparison of average RF and NDVI.

No.	Month	Standard error for constant	Standard error for coefficient	t-calculated for constant	t-calculated for coefficient	Level of significance for constant	Level of significance for coefficient
1	January	7.491	23.627	-0.334	2.717	0.742	0.013
2	February	9.24	32.022	0.200	2.356	0.843	0.029
3	March	15.449	51.446	0.279	3.986	0.783	0.001
4	April	26.086	73.97	0.030	4.297	0.976	0.000
5	May	18.339	48.09	-2.029	7.987	0.056	0.000
6	June	33.49	90.007	-2.608	5.526	0.017	0.000
7	July	48.421	134.108	-1.119	4.433	0.276	0.000
8	August	34.838	87.913	-2.536	7.384	0.020	0.000
9	September	20.251	47.588	-2.348	8.177	0.029	0.000
10	October	23.194	53.378	0.733	2.833	0.449	0.010
11	November	16.516	40.156	-0.086	2.302	0.932	0.032
12	December	8.402	22.870	-1.137	3.508	0.269	0.002

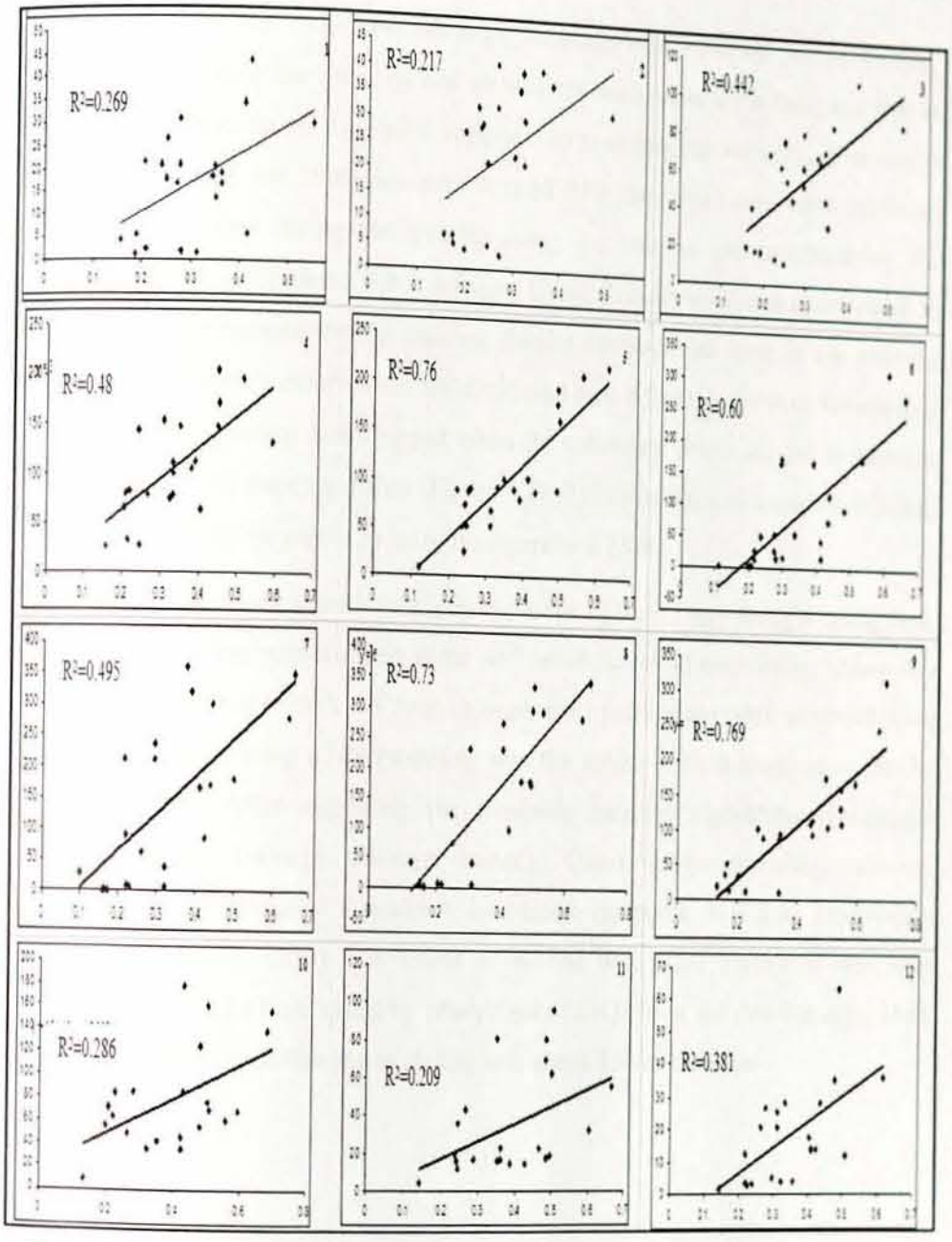


Figure 5.4: Scatter plots showing the correspondence between the average NDVI and RF from 1982 to 2004. The X-axis is for average NDVI and the Y-axis is for average RF in mm. The numbers (1–12) show the months January to December.

5.2 DroughtObject Modeling Using Data Mining

5.2.1 Exploring ANN DroughtObject Models

ANN modeling experiments using all the training, validation, testing and data for predicting the target dependent attribute are presented in Figure 5.5. As it can be

observed, the highest model performance accuracy was obtained on the training data. This is because the network was already trained on the same data, and this in line with our expectation. It is also important to note that the network, after it was being created, was not 100% accurate in predicting the target dependent attribute. This was controlled during the training using the training halt option using the avoid overfitting in Matlab v 7.9.0 software [204]. The network was constrained in such a way that training on the training dataset continues as long as the training reduces the network's error on the validation dataset. After the network memorized the training set, training was stopped when the validation errors started to increase as compared to the training errors (Figure 5.6). This was done to avoid overfitting, which is a challenging aspect in learning algorithms [204].

In Figure 5.6, June one-month prediction (predicting July 1984 drought using June data) of the training was stopped at the 44th epoch to avoid overfitting. Using this model exploration approach, all June to September predictions were explored. One of the challenges during ANN modeling was the length of time required to get the network model. After exploring the available training algorithms (Conjugate gradient (traincgf, traincgp, traincgb, trainscg), Quasi-Newton (trainbfg, trainoss), and Levenberg-Marquardt (trainlm)) in Matlab software v 7.9.0, Levenberg-Marquardt (trainlm) [255] was found to be the best. This algorithm was also recommended for its high speed by MathWorks [204]. Even for this fast algorithm, the minimum time to finish the modeling was about 15–20 minutes.

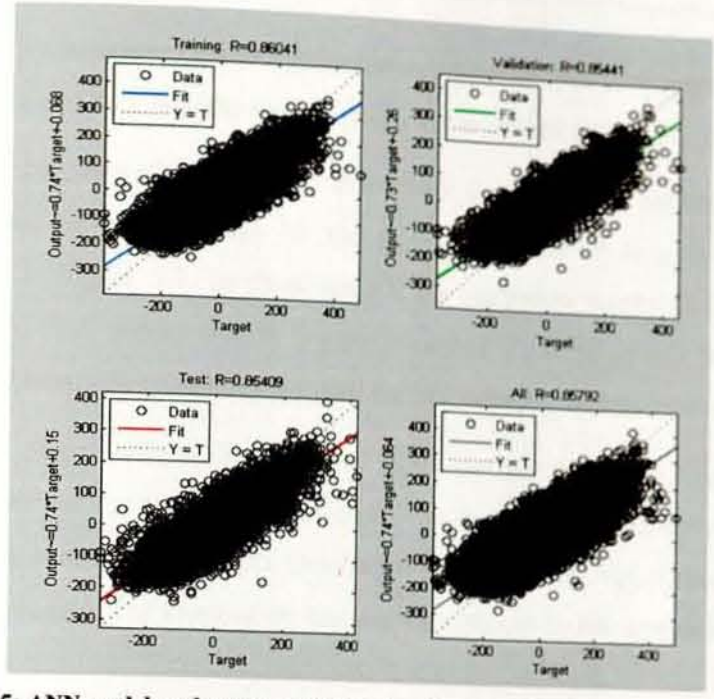


Figure 5.5: ANN model exploratory analysis using June one-month prediction. The X-axis presents the model predicted values and the Y-axis presents the observed values.

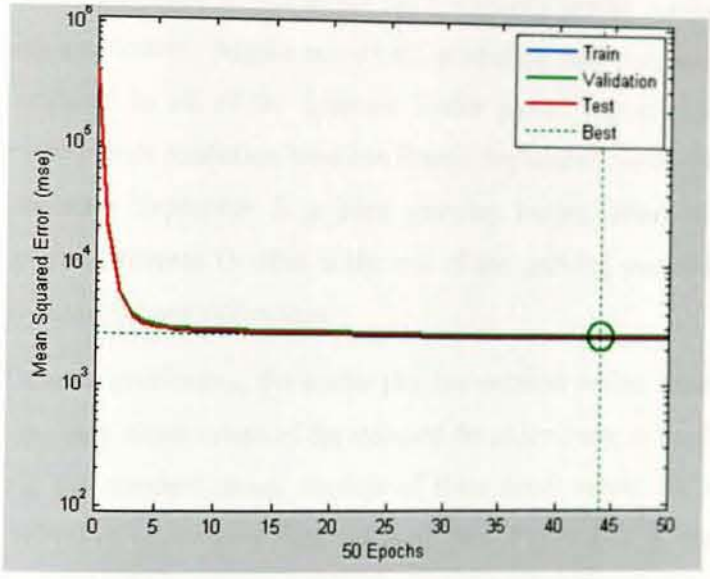


Figure 5.6: Best validation performance of ANN at 44th epoch.

5.2.2 Cross-Validation of DroughtObject ANN Models

The learned parameters from the data in the training dataset were subjected to those parameters in the test dataset. Then the prediction qualities of the models were

evaluated by comparing the predictions with the target data. This approach usually gives an idea of how well the models generalize to unseen data points [204].

Using the 24 years of data, one- to four-month DroughtObject prediction models' performance on test datasets for the growing months of June to October is presented in Figure 5.7. In these figures, the regression scatter plots (observed versus predicted) and the best fit line are presented. To come up with the best fit, a number of iterations were analysed until the best network models were obtained.

As can be observed from Figure 5.7 (a –j), for the June prediction model, the best fit was obtained for one-month prediction. The four-month prediction of June data (i.e., using June data to predict October) was found to be higher than the three-month prediction. The expectation was that there should be low correlation between the two-month data. Therefore, this needs further investigation.

Both July one-month and two-month predictions showed best fit, and as expected, the three-month prediction of this month has a relatively low fit compared to these two time lag predictions. August one-month prediction was also found to be the best fit compared to all of the assessed scatter plots. August two-month and September one-month prediction have low fitness. September one-month prediction was low, because September is a plant growing month, where the NDVI is relatively higher, whereas October is the end of the growing season (dry month) and the two datasets have differences.

In all the October predictions, the scatter plot has extreme outlier values. This was caused by the very small values of the standard deviation (near to zero) of October data. During the standardization, because of these small values, for some pixels very high SDNDVI values were observed as outliers (Figure 5.7d, g, i and j).

At each iteration, the correlation coefficient (r) (which is a measure of the accuracy level of the models) was assessed. The summary of these accuracy levels for different time lag predictions is presented in Table 5.4. The highest correlation coefficient was attained by August one-month prediction (predicting September using August data). This is in agreement with our expectation. Both of these months are vigorous growing periods and can be captured by NDVI data. The

lowest correlation was obtained by June three-month prediction (using June data to predict September). This is also in line with our expectation in that June is drier and the start of the growing months while September is a vigorous plant growth period, and the NDVI data for the two are different.

Table 5.4: Prediction accuracy of the networks on the test dataset.

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

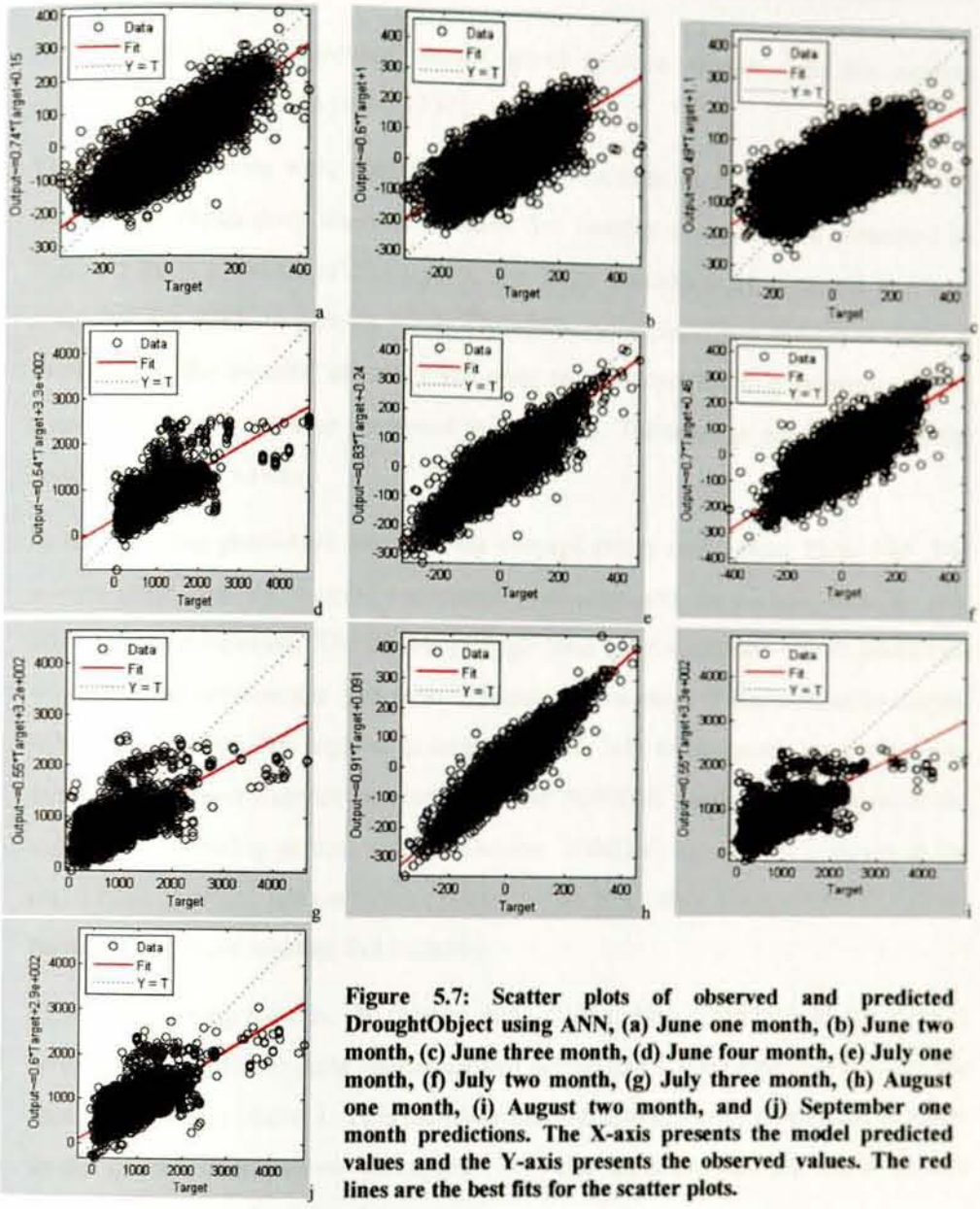


Figure 5.7: Scatter plots of observed and predicted DroughtObject using ANN, (a) June one month, (b) June two month, (c) June three month, (d) June four month, (e) July one month, (f) July two month, (g) July three month, (h) August one month, (i) August two month, and (j) September one month predictions. The X-axis presents the model predicted values and the Y-axis presents the observed values. The red lines are the best fits for the scatter plots.

5.2.3 DroughtObject Modeling Using Regression Tree

As in the ANN modeling approach, in cross-validating the DroughtObject models in the regression tree, the data were randomly split into two sets: a training set and a test set. The learned parameters from the data in the training set were subjected to those parameters of the test dataset, and the quality of the predictions on the test set

was evaluated. This approach usually gives an idea of how well the models generalize to unseen data points [137].

The 24 years of data were used for developing the time lag prediction models. The scatter plot exploratory analysis of these five months of models are presented in Figure 5.8a – j. In all of these plots, the same patterns were obtained as those observed for ANN in section 5.2.2. Therefore, no explanations are given here. A summary of the models' accuracy and error terms across the five growing months (June to October) are also presented in Table 5.5. The average error here is in the unit of SDNDVI values.

In the time lag prediction analysis, the average errors range from 22 to 190. The lowest error was for August one-month prediction and the highest was for July three-month prediction. The lowest average error for August one-month prediction was achieved because the next month, September, is more or less similar to August SDNDVI values. The highest average error in July three-month prediction was attributed to the difference between the June SDNDVI value, which occurs at the start of the growing period, and the October SDNDVI value, which occurs at the end of the growing season when plants have stopped their photosynthetic activity because they have reached full maturity.

In all the growing months, the relative error value (which is the ratio of the average error magnitude to the error magnitude that would result from always predicting the mean value) was below 1. This indicates that the average error magnitude is lower in the overall observations. Rulequest [137] also indicated that for useful models the relative errors should be less than one.

The highest correlation coefficient was achieved by August one-month (prediction of September SDNDVI values using August SDNDVI values) prediction. This is in agreement with our expectation in that both of these months are vigorous plant growing months and there is a similarity in their SDNDVI values.

The lowest correlation value was by June three-month prediction. This is in line with our expectation because June is the start of the growing months and August is a vigorous growing month, and the two months have different SDNDVI values.

With the same pattern as the ANN model, the June four-month prediction (predicting October using June data) was found to be higher than the June three-month prediction (predicting September using June data). Overall, low prediction accuracy was observed when the month of prediction was far from the predictor month. This is also in agreement with our expectations.

Prediction accuracy comparisons of ANN and regression tree data mining tools are presented in Figure 5.9. It can be observed that both tools have the same accuracy. This is because the training and test datasets were found to have well clustered data and easy to be classified. Such type of data clustering is described by Jain et al. [256] for different domains. Therefore, it is possible to use either ANN or regression tree in future drought prediction experiments.

The basic statistical properties were also compared between observed and predicted data values for the one- to four-month predictions based on Z-test for the means (Table 5.6). All the monthly prediction Z-cal values were found to be less than Z-critical table values (±1.96 for two tailed value at a 5% significance level), with the exception of August two-month and September 1-month prediction (i.e., for the month of October). This means that with the exception of August two-month and September one-month predictions, there is no significant difference between the mean values of observed and predicted data values.

Table 5.5: Summary of validation results for June-October regression tree models.

No.	Months	Evaluation on test data		
		Average error	Relative error	Correlation coefficient
1	June one month	37.5	0.49	0.85
2	June two month	46.488	0.61	0.77
3	June three month	51.748	0.67	0.71
4	June four month	179.753	0.57	0.77
5	July one month	27.255	0.36	0.92
6	July two month	39.691	0.51	0.84
7	July three month	189.925	0.59	0.75
8	August one month	22.184	0.29	0.95
9	August two month	186.404	0.58	0.75
10	September one month	180.224	0.57	0.77

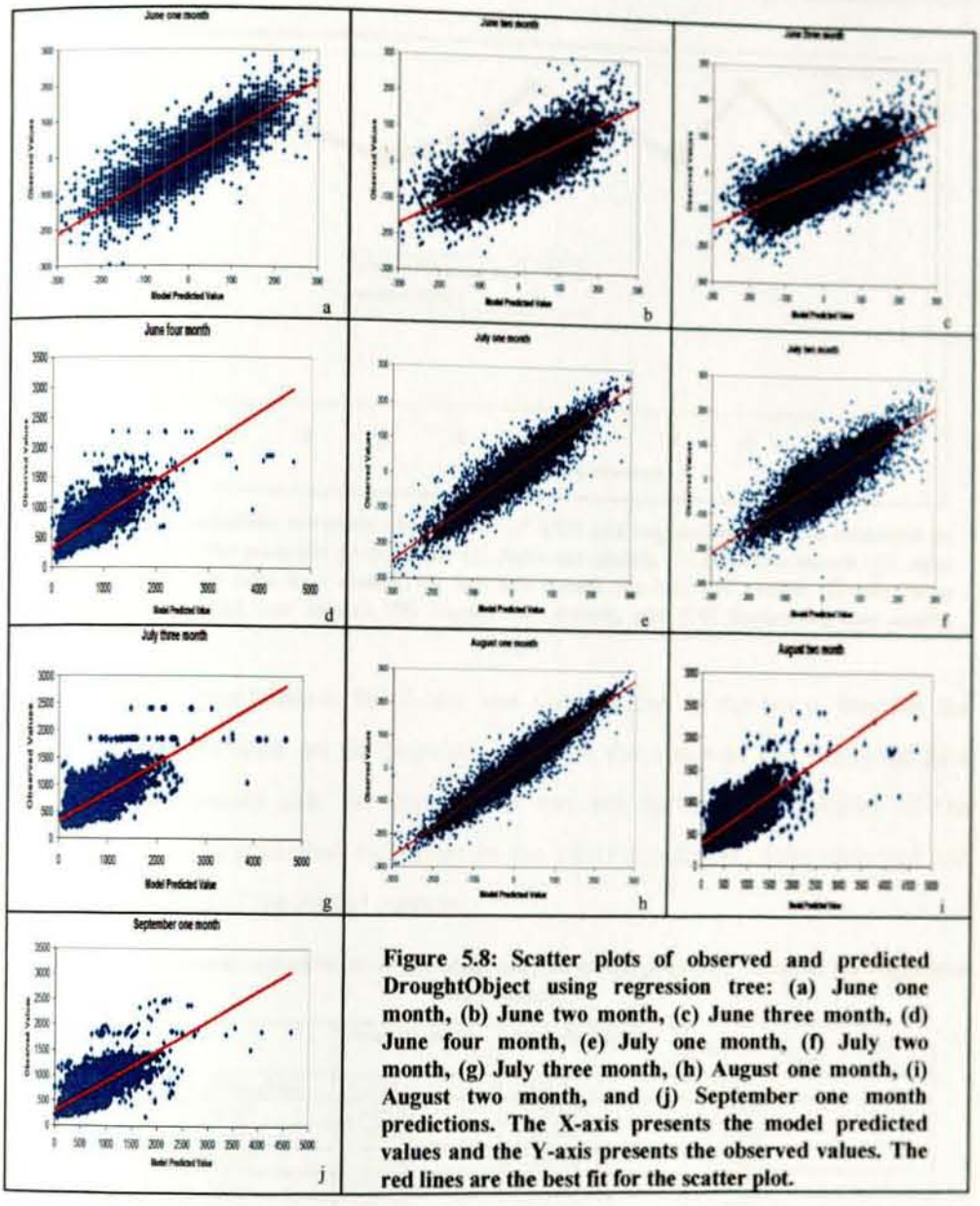


Figure 5.8: Scatter plots of observed and predicted DroughtObject using regression tree: (a) June one month, (b) June two month, (c) June three month, (d) June four month, (e) July one month, (f) July two month, (g) July three month, (h) August one month, (i) August two month, and (j) September one month predictions. The X-axis presents the model predicted values and the Y-axis presents the observed values. The red lines are the best fit for the scatter plot.

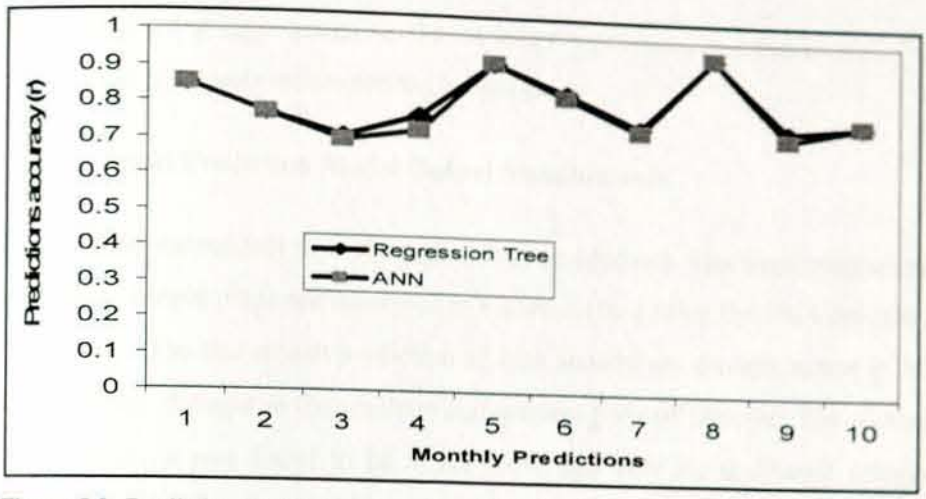


Figure 5.9: Prediction accuracy comparison of ANN and regression tree. The numbers on the X-axis are the monthly predictions: (1) June one month, (2) June two month, (3) June three month, (4) June four month, (5) July one month, (6) July two month, (7) July three month, (8) August one month, (9) August two month, and (10) September one month predictions.

In the means comparison, the Z-test was used instead of the t-test, because the comparison was done on the population of both the observed and predicted data values. This means that the comparison was not between the samples of the observed and the predicted, but rather on the whole population of the observed and predicted values of the model outputs.

Table 5.6: Statistical comparisons of the observed and model predicted values of the regression tree models.

No.	Models	Mean Observed	Mean Predicted	Decision $Z_{[cal]} < 1.96$
1	June one month ahead	0.2525	2.373	0.169
2	June two month	2.384033	2.078194	0.208551
3	June three month	1.523748	2.366955	-0.58469
4	June four month	710.9969	712.2986	-0.20616
5	July one month	2.038964	2.543363	-0.31571
6	July two month	3.435608	3.773627	-0.2175
7	July three month	720.7159	708.4185	1.920804
8	August one month	2.829	3.723647	-0.54913
9	August two month	721.8606	707.4495	2.242656
10	September one month	724.8788	711.6075	2.096091

The errors in the models in general are attributed to the low resolution of the satellite imageries (8km) and the high heterogeneity of the ecosystems of Ethiopia. From our practical experience, within a short range of distance, the elevation, vegetation type, and soil conditions vary greatly. These attributes were represented by 8km resolution DEM, veg_Ethiopia, land cover, and WHC. These errors can be

improved by using high spatial resolution imagery products, and this is one of the future research projects recommended by this study.

5.2.4 Drought Prediction Model Output Visualizations

The experimental models from the regression tree approach were implemented, and the models' output maps are presented in Figure 5.10a–j using the 1984 drought as a case study. The one-month prediction of June showed the drought extent in July with moderate drought in the southern and western parts of Ethiopia. The northern part of Ethiopia was found to be in the moist and very moist drought category (Figure 5.10a). The two-month prediction (for the month of August) showed extreme drought classes in the northern and southwestern parts of Ethiopia (Figure 5.10b). This means that during this month, these places were found to have highly deviated SDNDVI values compared to the historical mean of the 24 years. During this month, the central and eastern parts of the country were found in the moist classes. The three-month prediction of the June model (i.e., for the month of September) showed that there was moderate drought all over the country except the northeastern part of Ethiopia. The northeastern part of the country was found to be moist by this model prediction. The four-month June prediction model showed that there were moderate droughts in the south and southeastern parts of the country. The model accuracies of the June two- and three-month predictions were found to be low (Table 5.5), and accordingly the maps produced have fewer accuracies than the June one-month prediction.

The July one-month prediction map (for the month of August) showed that northern, eastern and southern parts of Ethiopia were in the severe drought classes. The central part of Ethiopia was in the moist and very moist classes by this prediction (Figure 5.10e). The two-month prediction (i.e., for the month of September) showed only the northeastern part of Ethiopia in the extreme to severe drought class and the south and southeastern parts in the moderate drought class. The central part of the country was moist according to this model (Figure 5.10f). The four-month prediction map showed the northeastern and western parts of Ethiopia in the moist drought class and the eastern part of the country in the near-

normal drought class (Figure 5.10g). The three-month July prediction has a low accuracy model (Table 5.5), and the map also does not reflect the reality during October 1984.

The August one-month prediction (for the month of September) clearly showed the extreme drought in northern, eastern and southern parts of Ethiopia. This model showed high vegetation stress throughout the country with a few exceptions in the central and western parts of Ethiopia (Figure 5.10h). This model had the highest accuracy of all models (Table 5.5) and also the best fit in the scatter plot analysis (Figure 5.8h). Therefore, this model is highly reliable and also showed the extreme to severe drought situation described by the literature [65, 66, 257-259]. The August two-month and September one-month (for the month of October) showed a pattern similar to the July three-month prediction map (Figure 5.10i and j).

Overall, as it can be observed from the maps, the best models with higher accuracies produced the best patterns. As indicated above, the highest accuracy from the regression tree model was achieved by the August one-month prediction (Figure 5.8h and Table 5.5). The best prediction map was also obtained using this model (Figure 5.10h).

As the prediction month length increases, the pattern observed from the maps also decreases, showing the drought severity. Overall, the one-month predictions were found to be the best. The last growing month (October) in all of the monthly predictions showed no change compared to the previous drought severity status. These reflect the reality that this month is the end of the growing month and is usually dry. Therefore, no SDNDVI deviation is expected in this month.

From the above analysis, it is clear that the drought prediction models developed can confidently be used for predicting future drought situations with one- to four-month time lags with the accuracy levels presented in Table 5.5. The experimental maps of drought severity status prediction for the 1984 drought in Ethiopia matched the reality described by different articles and reports. Therefore, these models can be used for any future drought status predictions in the case study area. Using these models, a result similar to the accuracy level presented in Table 5.5 can be

obtained. For the actual implementation of these models, the required models ***.icase***, ***.names***, and historical ***.data files*** were also produced, and they are ready for the model implementation.

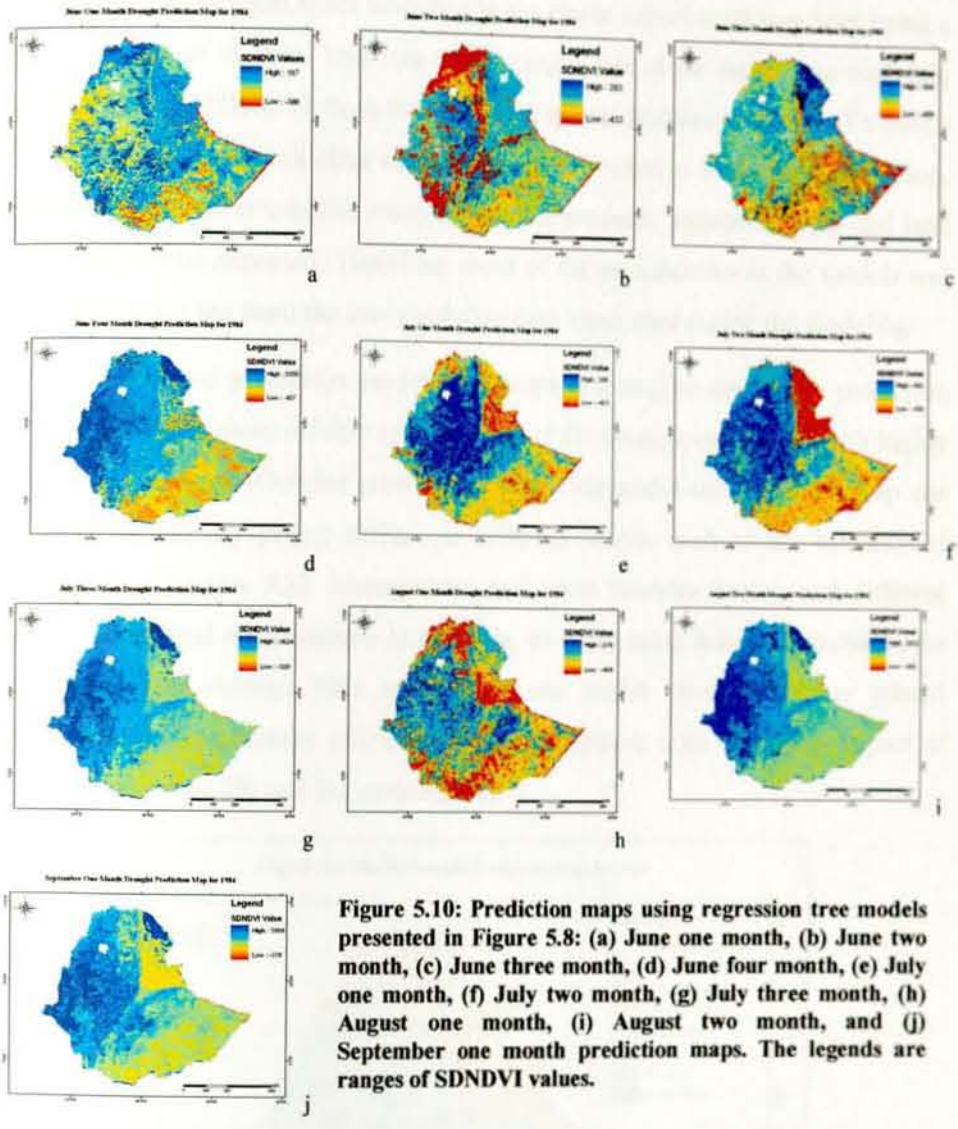


Figure 5.10: Prediction maps using regression tree models presented in Figure 5.8: (a) June one month, (b) June two month, (c) June three month, (d) June four month, (e) July one month, (f) July two month, (g) July three month, (h) August one month, (i) August two month, and (j) September one month prediction maps. The legends are ranges of SDNDVI values.

The drought severity maps for the years 1984 and 2002 are presented in Figure 5.11 and Figure 5.12, respectively, using standard drought severity classes defined by McKee et al. [28], McKee et al. [29] and Tadesse et al. [71]. As it can be observed, using these standard drought severity classes, the 1984 drought was highly severe compared to the 2002 drought in Ethiopia that covered most of the country. The

2002 drought spatially observed in the southeastern and northeastern parts of the country was moderate to severe (Figure 5.12). The extreme drought class was not observed in 2002.

Inherently, the DroughtObject modeling in the above experiments was done using a spatial resolution of 8km. This was done because one of the major data modeling attributes of the SDNDVI from NOAA has a spatial resolution of 8km. To match SDNDVI NOAA data, all other attributes were converted to 8km spatial resolution. Within 8km spatial resolution, many different elevations, vegetation types and land covers should be expected. Therefore, most of the uncertainties in the models and prediction maps are from the low resolution data input used during the modeling.

From the assessed prediction models and maps, the August one-month prediction model and map is more reliable and can be used for drought monitoring with higher accuracy in the June–October growing periods. This model and prediction map can be used directly by policy makers at different levels, such as the Ministry of Agriculture, Disaster Risk Management and Food Security Sector, and different non-governmental organizations in Ethiopia, to make more informed decisions for saving drought victims. This information can enable these and other related organizations to undertake efficient drought mitigation, with minimum impact of drought on human life and the environment.

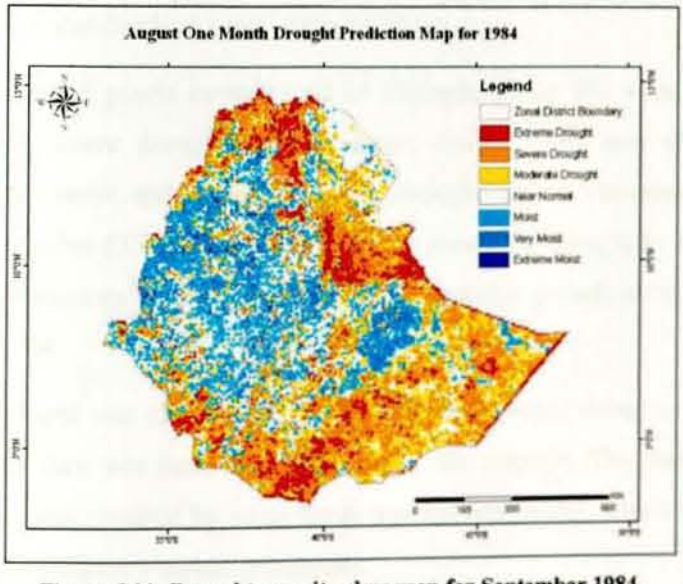


Figure 5.11: Drought severity class map for September 1984.

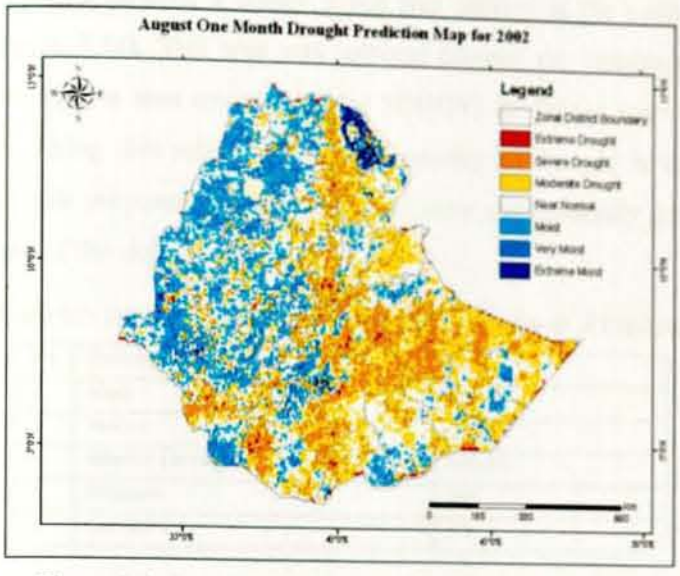


Figure 5.12: Drought severity class map for September 2002.

5.2.5 DroughtObject Extractions

For DroughtObject extraction, the drought severity extent in September 1984 was used (Figure 5.13). The descriptive statistics of all the pixel values of this month's drought severity image are presented in Table 5.7. The values are SDNDVI values. These values were multiplied by 100 and the values that are presented in the table should be divided by 100 to get the actual standardized values. Therefore, a value of 100 means 1 standardized value of the SDNDVI.

From the assessed pixels covering all of Ethiopia, about 8% were in extreme drought, 25% severe drought, 16% moderate drought, 30% near normal, 13% moist, 7% very moist, and 1% extreme moist drought classes. The assumption here is that the negative SDNDVI are revealing the prevailing drought in the area and the positive deviations are showing the healthy vegetation growth in the area during September 1984.

The 1984 drought was also reported to be one of the worst droughts in Ethiopia [66], and this data was used for demonstrating the concept. The map used here (Figure 5.13) was obtained by using the regression tree model, which is presented in Figure 5.8h (August 1-month prediction).

To extract drought objects, a subset image was created in the northern part of Ethiopia (Figure 5.14). This area was selected because the maximum SDNDVI value and the highest area coverage of the SDNDVI deviations were observed in this location. Using this subset image, the intensity of drought in the area was assessed. For this purpose, all the pixel values were systematically generated in a point grid map of the site.

Table 5.7: Descriptive statistics for SDNDVI values for all of Ethiopia.

No.	Description	Value
1	Mean	-59.8915
2	Median	-58.5
3	Standard Deviation	166.6987
4	Minimum	-435
5	Maximum	279

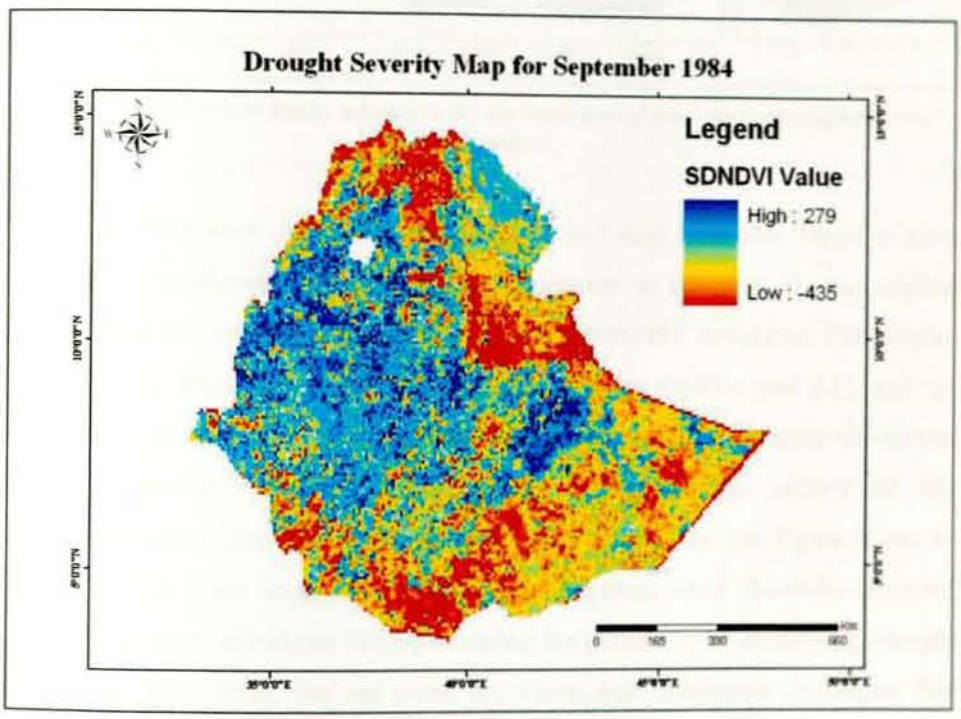


Figure 5.13: Drought severity extent in September 1984.

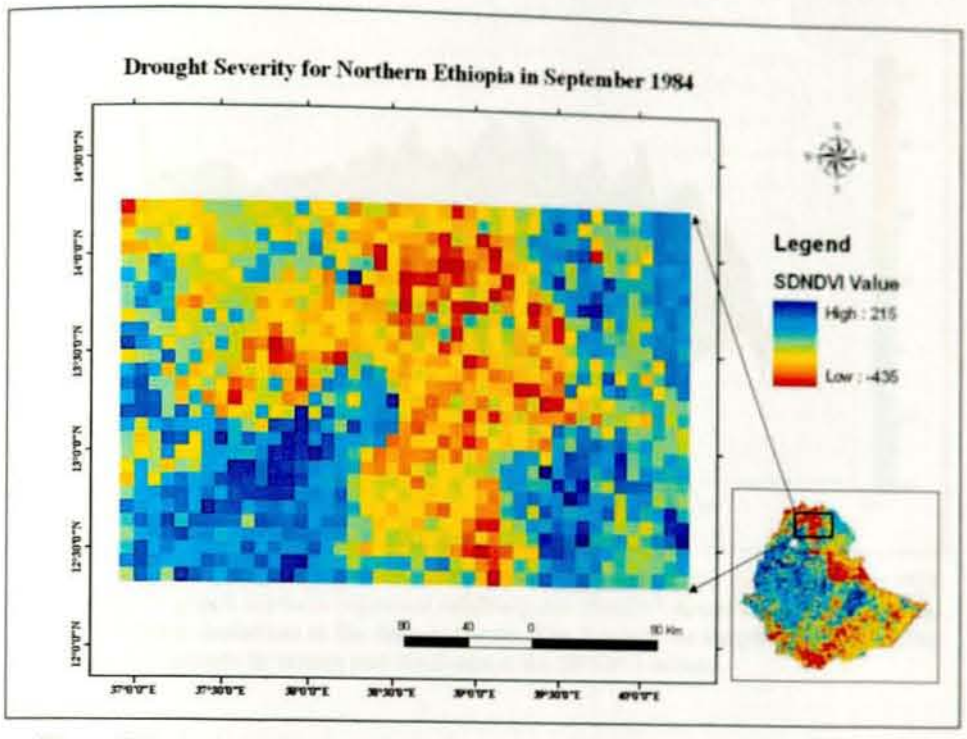


Figure 5.14: A subset image selected in the northern part of Ethiopia for DroughtObject extraction.

The pixels that were obtained from the point grid map were also found to have some patterns showing DroughtObject distribution in the area. In the original image, there are ranges of negative to positive SDNDVI deviations. For display purposes, the negative deviations were multiplied by negative one (-1), and the original negative deviation values became positive and the positive deviations became negative. The 3D representation of the spatial pattern of the DroughtObjects in the area is presented in Figure 5.15. From this figure, it can be observed that some local maximum deviation values were found to decrease gradually to local minimum values, revealing the possibility of identifying drought objects. In this figure, the red areas are pixels with maximum deviations that gradually decrease to the yellow color, and the blue colors show positive deviations of the pixels (in this case, with optimum red band use for plant photosynthesis activity or conversion of light energy into chemical energy during September 1984). After observing these patterns, the drought objects were then extracted using the concepts of local maxima and local minima SDNDVI deviation values.

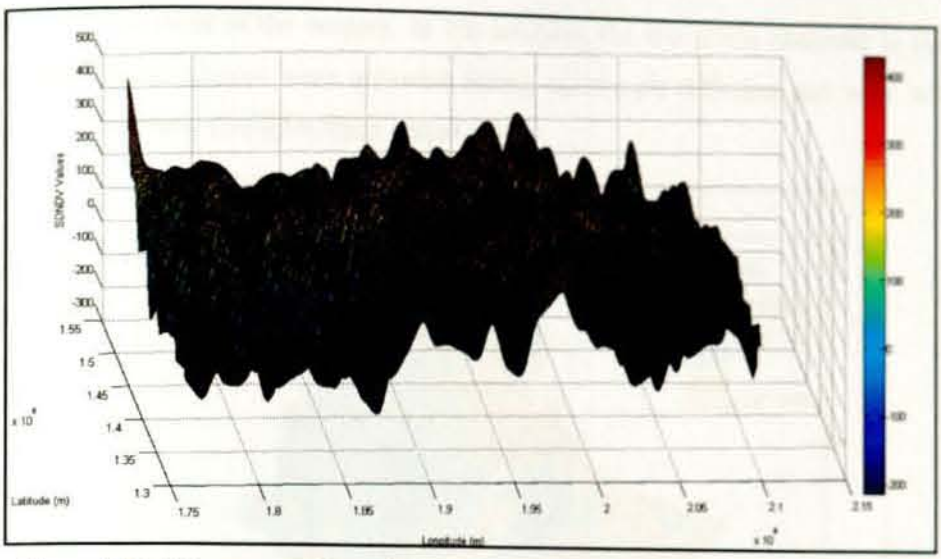


Figure 5.15: 3D representation of SDNDVI deviation values for the September 1984 drought. The graph surfaces represent relatively low SDNDVI deviation in the blue areas and the highest deviations in the deep red areas. The X axis is the longitude in meters, the Y-axis is the latitude in meters and the Z-axis is the SDNDVI values.

The actual DroughtObject extraction for the entire country was done after diffusing the image using non-directional anisotropic diffusion. The diffused image of the September 1984 drought map is presented in Figure 5.16. For this diffusion, a low pass convolution filter with kerned size 3x3 was used. As can be observed from the figure, the DroughtObject patterns from high to low severe drought areas are clearly identified.

Using the anisotropic diffused image, DroughtObject segmentation was executed using the eCognition Developer 8.7 [260] fuzzy segmentation tool. This classification approach works using a membership function, where a pixel's value is determined by whether it is closer to one class than another. The segmented and classified map is presented in Figure 5.17.

As can be observed from Figure 5.17, most of the extreme drought classes were observed in the northern part of the country. Almost all of Tigray and Afar regions were highly affected by the extreme drought class. The severe and moderate drought classes were observed in the northern, eastern and southern parts of the country. The near normal, moist and very moist classes were observed in the central

and western parts of the country. In this analysis, the few pixels observed in the extreme moist classes were removed during anisotropic diffusion and were not found in the final DroughtObject output map.

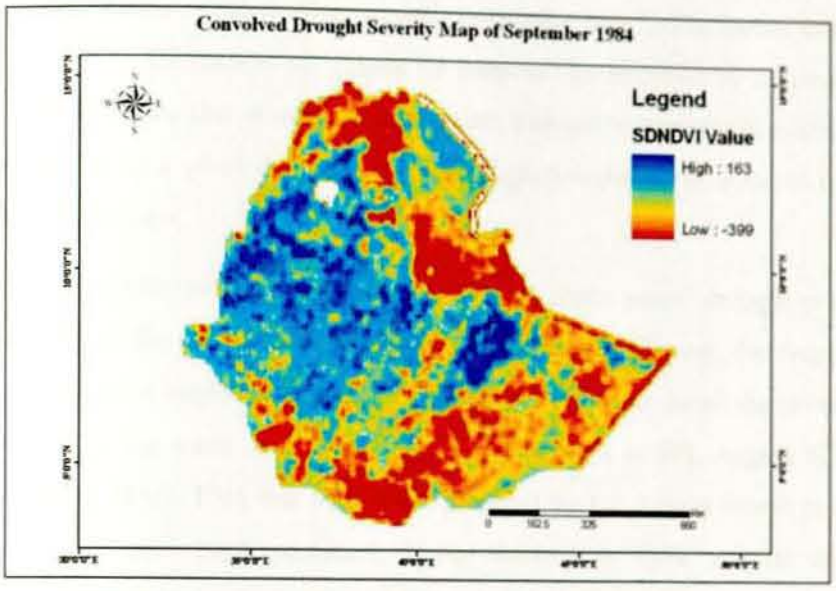


Figure 5.16: Anisotropic diffused image of September 1984 drought.

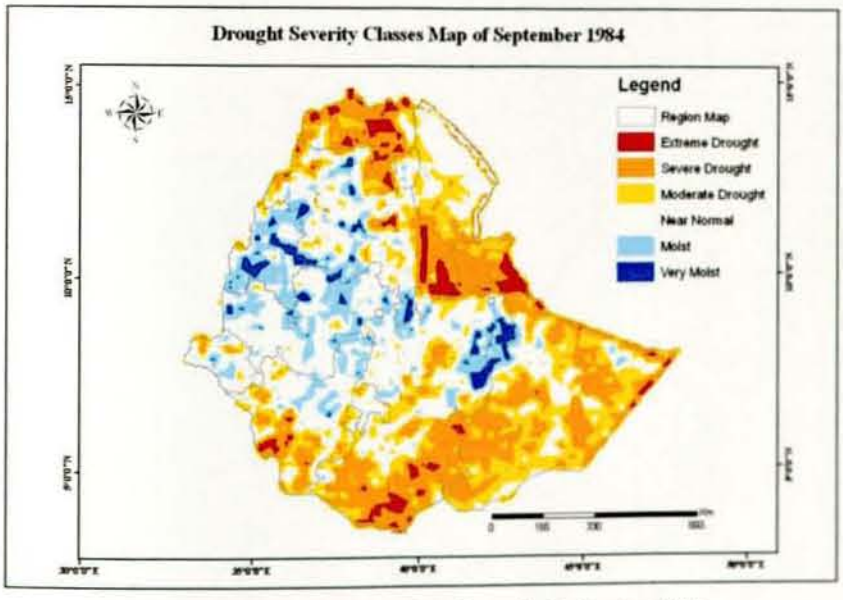


Figure 5.17: Drought severity classes in September 1984

5.3 DroughtObject Model Evaluation

This section addresses research question three (how is DroughtObject identified on satellite imagery evaluated using ground observation data?). The model evaluation here is done in the context of "fitness for purpose" as described by Jakeman et al. [76]. This is to say that in one way or another, drought monitoring is related to the food security of a given district, and the drought monitoring tool has to take this reality into account.

The model evaluation was done using a total of seven years' drought prediction model outputs for the years from 2000 to 2006. For this purpose, the August one-month prediction model was run for the month of September for all the seven years. In addition to the static data, all the dynamic data, such as SPI, August SDNDVI, PDO, AMO, NAO, PNA and MEI, were prepared for the August month prediction during the model implementation. Using these input data, a total of seven DroughtObject maps were produced (Figure 5.18). These images were combined with zonal administrative data of Ethiopia for zonal statistical average data extraction.

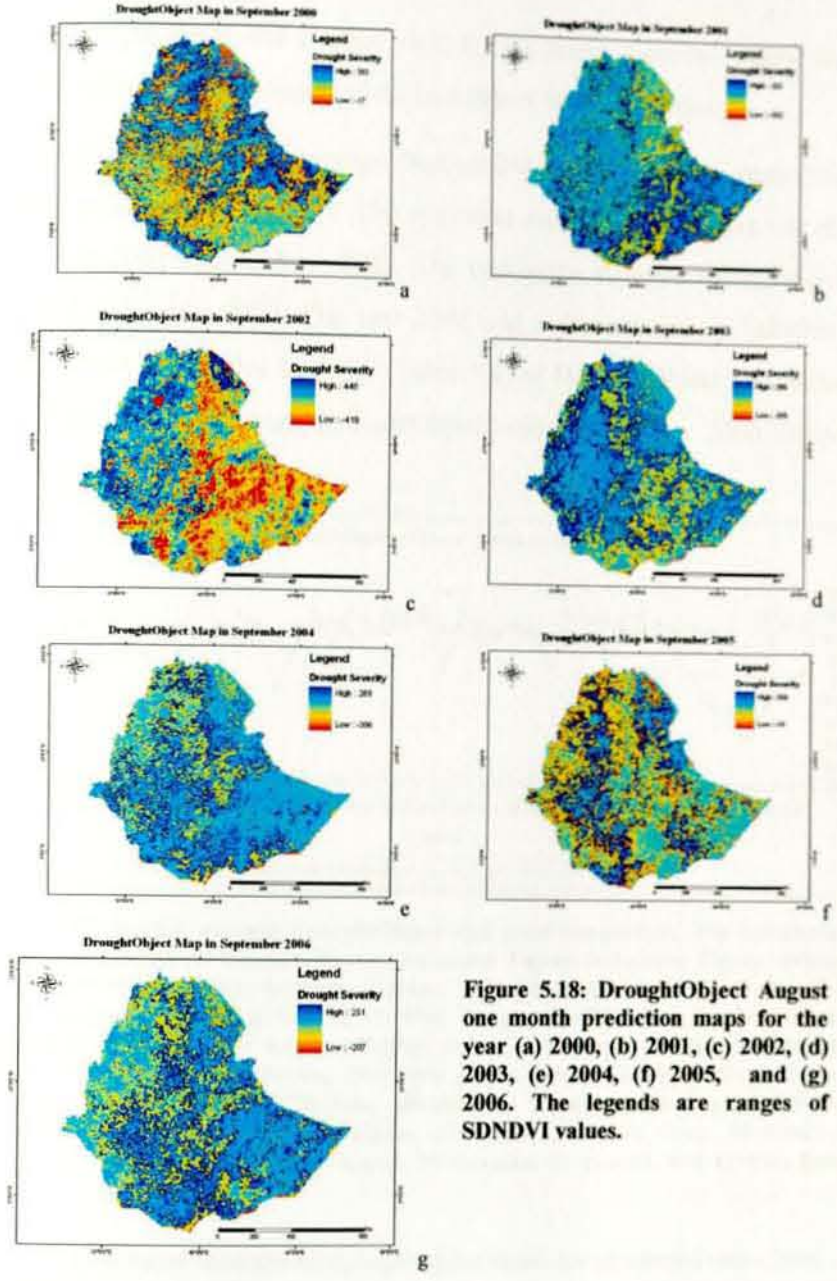


Figure 5.18: DroughtObject August one month prediction maps for the year (a) 2000, (b) 2001, (c) 2002, (d) 2003, (e) 2004, (f) 2005, and (g) 2006. The legends are ranges of SDNDVI values.

5.3.1 DroughtObject Versus Cereal Yield Data Exploratory Analysis

The comparison of average DroughtObject and average cereals yield (Qt/ha) from 2000 to 2006 is presented on a 2-dimensional line graph in Figure 5.19. The x-axis shows the zone numbers, y1-axis average DroughtObject and y2-axis average yield data. As can be observed from the graph, there is a spatial relationship between

average DroughtObject and average yield for the selected zones. Except for a few zones, the increase and decrease of the two values were correlated.

The descriptive statistics for average DroughtObject values for all zones from 2000 to 2006 is presented in Table 5.8. The minimum mean DroughtObject was observed for the year 2002 followed by 2001. The maximum standard deviation was also observed for the year 2002. The year 2002 was a drought year in Ethiopia [261], and the overall descriptive statistics (Table 5.8) of DroughtObject also reflects this reality. The maximum mean DroughtObject was for the year 2000 followed by 2004.

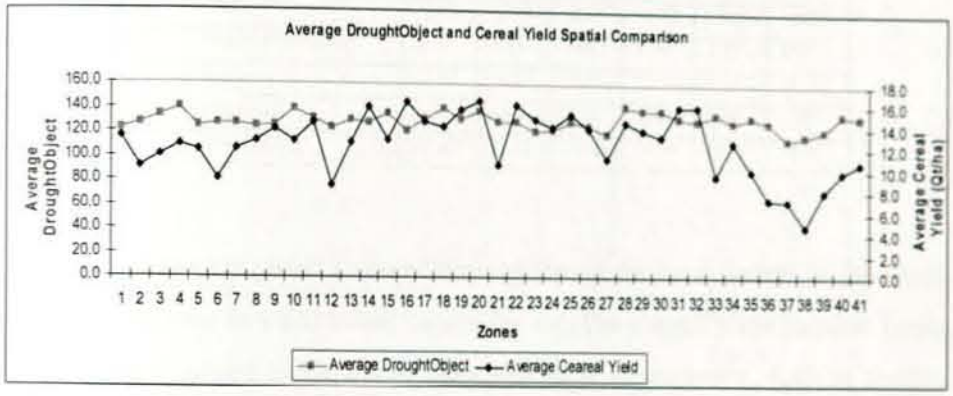


Figure 5.19: Spatial average DroughtObject and yield comparison. The horizontal axis numbers are Zones (1=Western Tigray, 2=Central Tigray, 3=Eastern Tigray, 4=Southern Tigray, 5=North Gonder, 6=South Gonder, 7=North Wollo, 8=South Wollo, 9=North Shewa (Region 3), 10=East Gojam, 11=West Gojam, 12=Wag Himra, 13=Awi, 14=East Wellega, 15=Oromia Zone Amhara Region, 16= West Wellega, 17=Ilubabor, 18=Jimma, 19=West Shewa, 20=East Shewa, 21=North Shewa Oromia Region, 22=Arsi, 23=West Harerge, 24=East Harerge, 25=Bale, 26=Metekel, 27=Asosa, 28=Gurage, 29=Hadiya, 30=Kembata-Alaba-Tembaro, 31=Sidama, 32=Gedio, 33=South Omo, 34=Bench Maji, 35=Yem, 36=Amaro, 37=Burji, 38=Konso, 39=Derashe, 40=Hareri , and 41=Dire Dawa).

Table 5.8: Descriptive statistics for DroughtObject values for all selected zones, 2000 – 2006.

No.	Descriptive Statistics	Years						
		2000	2001	2002	2003	2004	2005	2006
1	Mean	123.25	130.69	-1.43	143.06	136.27	116.23	125.02
2	Median	125.09	132.00	13.71	147.27	139.88	111.03	124.05
3	Standard deviation	25.54	18.07	45.79	17.53	20.56	23.54	22.65
4	Range	114.37	93.91	189.59	83.20	76.46	91.33	94.88
5	Minimum	74.96	61.29	-105.89	98.20	95.74	72.38	80.69
6	Maximum	189.32	155.20	83.70	181.40	172.20	163.71	175.57

The average yield data in space and time dimensions were explored. For this purpose, cereal yield average values (both the overall and individual averages of Teff, barley, wheat, maize, sorghum, millet and oat) were explored. The descriptive statistics for cereal are presented in Table 5.9. The cereal yield data varies both temporally and spatially. The spatio-temporal average yield analysis output is also presented in Figure 5.20.

Table 5.9: Descriptive statistics for cereal yield for all zones, 2000 – 2006.

No.	Descriptive Statistics	Years					
		2000	2001	2003	2004	2005	2006
1	Mean	10.92	12.73	11.70	12.14	13.00	13.77
2	Median	11.45	13.52	12.39	12.47	13.25	14.26
3	Standard Deviation	2.72	2.83	3.48	3.19	3.32	3.55
4	Range	12.62	12.75	15.87	12.74	15.32	15.17
5	Minimum	4.85	4.73	2.11	5.35	4.64	4.75
6	Maximum	17.47	17.48	17.98	18.09	19.96	19.93

All the average crop yields assessed showed almost the same pattern in that there is a gradual increase in yield value, except for oat. The overall yield increase through the years is assumed to be due to improved farm management, such as fertilizer, pesticide applications and improved farming systems through agricultural extensions. The challenge now is how to use these yield data, which are increasing gradually from year to year, with DroughtObject data, which is assumed to increase when there is surplus production (favorable weather conditions) and decrease when there is a deficit in crop yield (due to unfavorable weather conditions). In other words, in the DroughtObject evaluation using the crop yield data, the expectation was that when there is high deviation in SDNDVI values (a decrease in DroughtObject value), there is a decrease in yield value. This way it is possible to observe drought incidence on the ground. The crop data that we assessed had a gradual increase in yield from 2000 to 2006. This means there is a trend in the crop yield data. If we simply test the correlation between DroughtObject data and crop data without removing the inherent trend in crop yield data, unexpected and/or misleading results can be produced. Therefore, we first removed the trend and then

tested the correlation between the two attributes. For this purpose, we detrended the crop yield data.

Detrending is removing linear trends from regularly sampled time-domain input-output data [262]. This data processing operation helps to estimate more accurate linear models because linear models cannot capture arbitrary differences between the input and output signal levels. After detrending, the relationships between DroughtObject and crop yield data were analyzed.

The detrended average crop yields for cereal (Teff, sorghum, maize, barley and wheat) are presented in Figure 5.21. This is the average temporal yield for all zones for each of the crops. Since the year 2002 was missing, the average values of 2001 and 2003 were used for 2002.

Because the year 2002 was missing in the yield data, it was not included in the statistical comparisons. The cereals yield data showed that the years 2003 to 2006 were below average value (zero); for all other years, the cereal yield was above average (Figure 5.21a). Teff yield showed the yearly variation clearly. The years 2003, 2005 and 2006 had below-average values (Figure 5.21b). Sorghum yield was observed to be below average from 2003 to 2006 (Figure 5.21c). Maize was observed to be below average for 2003 and 2004. Barley yield was observed to be below average for the years 2000, 2003 and 2004. Wheat yield was found to be below average for the years 2000 and 2005.

The detrended crop yield data for all the zones selected for this study were assessed for their correlation with DroughtObject data. The correlation and scatter plots for some of the selected zones (one zone per assessed region) are presented in Figure 5.22. The highest correlation was obtained for the Western Hararge Zone of Oromia region. The lowest correlation coefficient was found for Harari region (r^2=0.01, scatter plot not presented).

The analysis results showed that there is a relationship between DroughtObject and crop yield data. The decrease or increase of DroughtObject values in a space-time dimension can show the deficit or increase of yield data with optimum level of certainty.

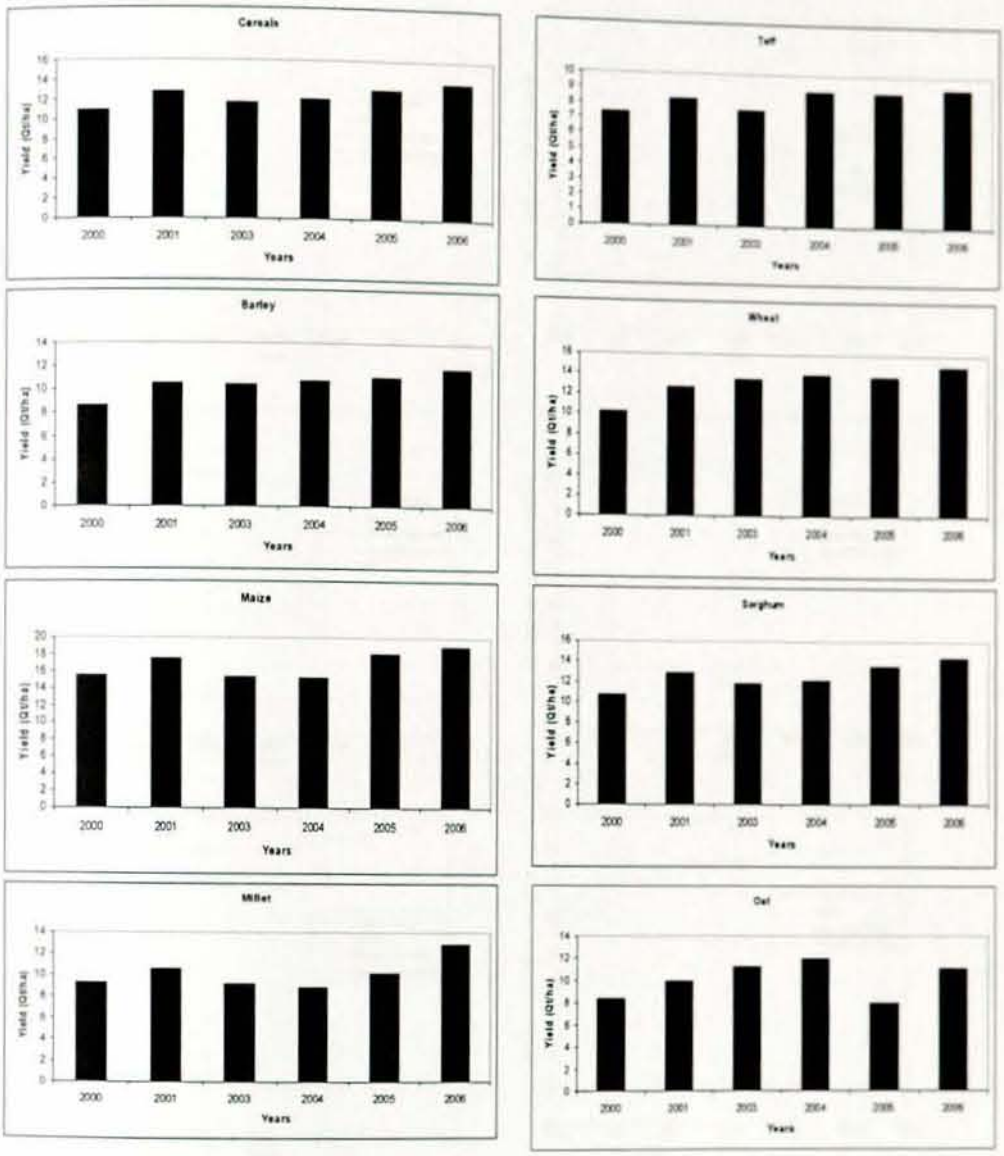


Figure 5.20: Crop yield pattern for the past 6 years in the studied zones.

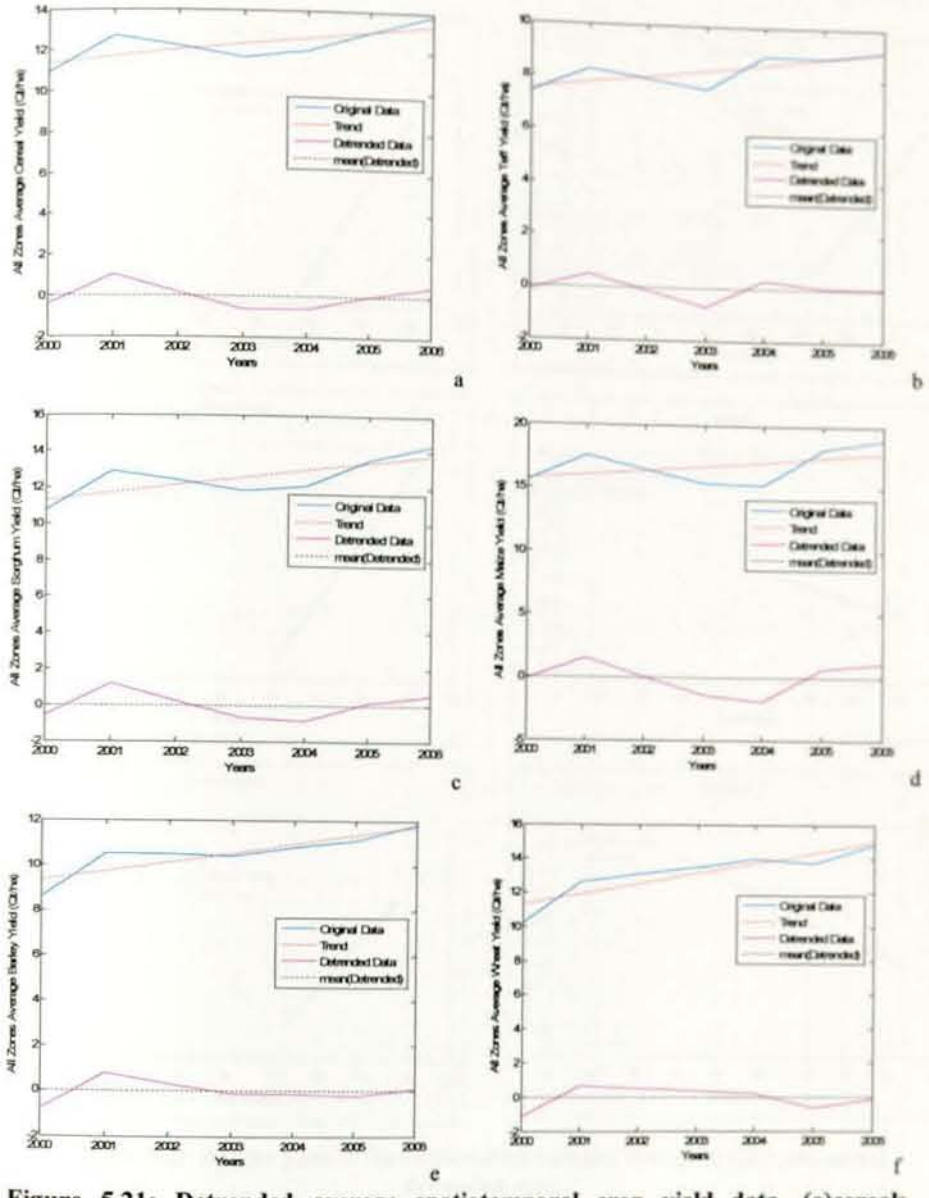


Figure 5.21: Detrended average spatiotemporal crop yield data, (a)cereals, (b)Teff, (c) sorghum, (d) maize, (e) barley, and (f) wheat.

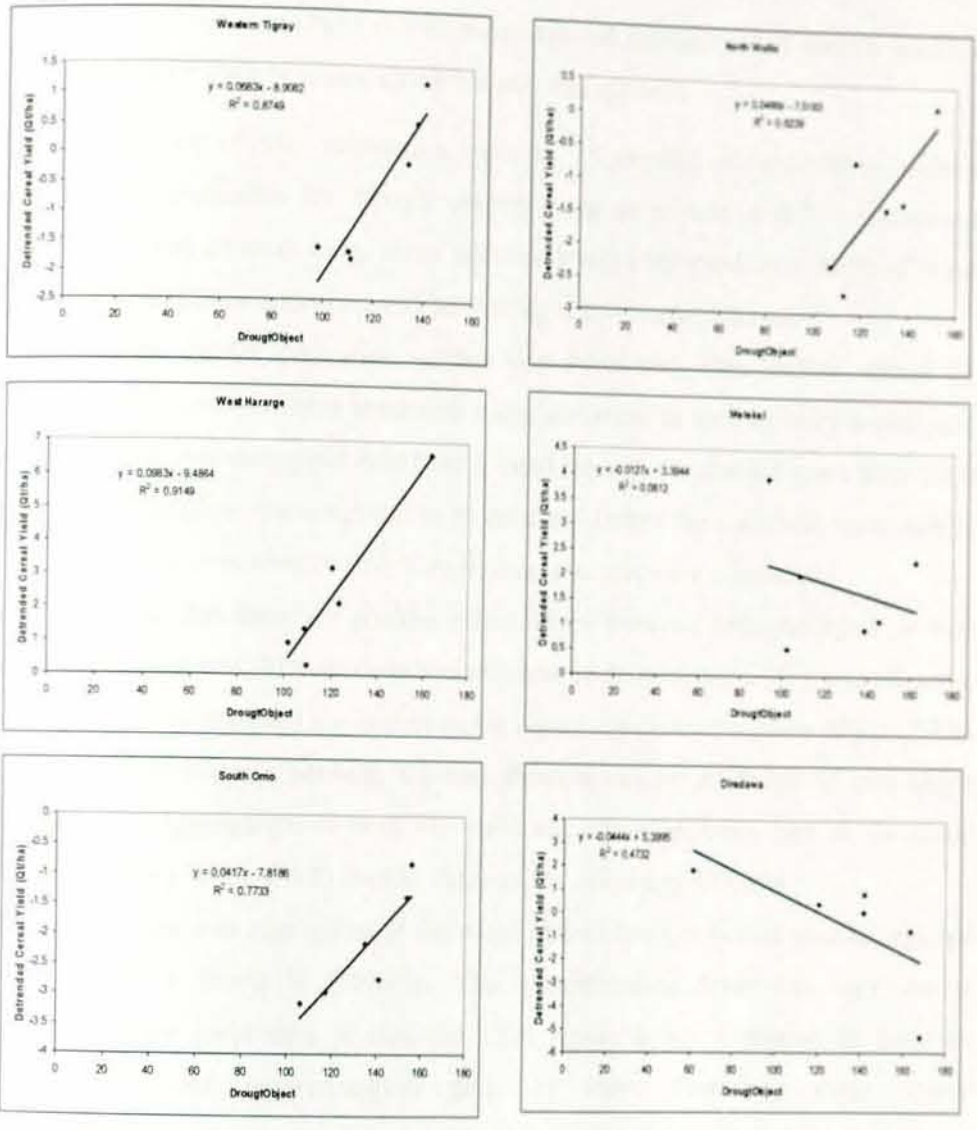


Figure 5.22: Scatter plots of the relationships between DroughtObject and cereal detrended yield.

5.3.2 Summary on DroughtObject Evaluation

In summary, interesting results were found in evaluating the DroughtObject product. Even without any statistical analysis, the DroughtObject model output in Figure 5.18a-g clearly shows that the year 2002 has highly deviated pixels compared to other years from 2000 to 2006. The model output showed where the highly deviated pixel and group of pixels are spatially located. Decision makers

who are involved in drought risk-management and mitigation can directly use this information for their resource allocation and management.

The objectives of this subsection were to 1) develop an evaluation method specifically applicable for drought models using crop yield, and 2) evaluate the DroughtObject product using zonal (district level) crop yield data. Both of these two objectives were addressed and interesting results were obtained.

A simplified model evaluation method was developed. This method started by exploring the two attributes separately using statistical data exploratory techniques. It was found that crop yield data have a trend through the studied years from 2000 to 2006. Therefore, the trend had to be removed before the statistical relationships analysis. After removing the trend, regression analysis were conducted.

It was found that there are positive relationships between DroughtObject product and crop yield data. The maximum coefficient of determination (R^2) was observed for Western Hararge and the minimum for Harari region for the years 2000 to 2006. The weak correlation between the two datasets can be attributed to two major reasons: 1) aggregation of both the yield and DroughtObject data at the zonal administrative level; and 2) the two datasets are inherently different.

The yield data was aggregated at the zonal district level, which is next to regional administrative levels in Ethiopia. This administrative level has very coarse resolution for yield data in that the CSA report is not homogenous from an agricultural and agro-ecological point of view. Therefore, these coarse administrative level averages cancel local variability of yield and DroughtObject data. One future research project on evaluating the DroughtObject model for food security and drought monitoring is to use high resolution administrative level crop yield data, such as the woreda administrative level (the next administrative resolution up from the zone administrative level in Ethiopia).

Interestingly, even though the DroughtObject and crop yield datasets were from different sources and were produced for different purposes, there were strong relationships. It was observed that the increase or decrease of SDNDVI values can indicate the yield increase or deficit in the studied districts for the "Meher" growing season. Therefore, it is possible to say that the DroughtObject model product can be

used directly by all governmental and nongovernmental organizations involved in drought monitoring and food security assessment.

5.4 Application of the DroughtObject Model for Assessing the 2011 Ethiopian Drought

This section demonstrates the applicability of DroughtObject models for assessing drought episodes with a dataset outside the modeling years. The models were built for a total of 24 years (from 1983 to 2006). The model developed using the 24 years of data from biophysical, climate and satellite data sources was used for assessing the 2011 drought in Ethiopia. Both the static and dynamic data inputs were used as described in the previous subsections. The static data (DEM, Land cover, Veg_Ethiopia and WHC) were directly taken from the modeling data in the previous experiments.

The dynamic data inputs, which include the oceanic indices, SPI, and SDNDVI data, were derived for the year 2011. The oceanic indices AMO, MEI, NAO, PDO and PNA were obtained from the NOAA website [216] as point data for each monthly period from June to September 2011. Using these point data, constant raster maps were produced using the ArcGIS 9.3 Create Constant Raster Tool [203]. For each oceanic index, four different constant raster maps (for the four months June–September) were produced. For the five oceanic indices, a total of 20 constant raster maps were created.

The SDNDVI was obtained from eMODIS NDVI data. eMODIS NDVI data has a spatial resolution of 250m and temporal resolution of six pentadal composite (10 day updated every five days) for each month [185]. The eMODIS NDVI data were stretched (mapped) linearly (to byte) values from [-1, 1] to [0, 200] when obtained from USGS database. These values were converted to the original NDVI value as $NDVI = (value - 100)/100$, where $value$ is the stretched NDVI value ([0, 200]). Then to obtain the monthly NDVI values, the average of the pentadal values were obtained for June–September. October was not included here because it was not used for the prediction models.

The cumulative NDVI (CNDVI) values for the four months were obtained by subtracting the May 2011 base NDVI (NDVI_ib) as described in equation 7. After getting the CNDVI for each monthly value, the standardized deviation of NDVI (SDNDVI) was obtained using equation 8. The SDNDVI values obtained from the above analysis were used as model input for the monthly DroughtObject prediction models.

The SPI data was obtained from eMODIS with spatial resolution of 10km and temporal resolution of dekadal (10 days). The three-month SPI data was used in this analysis. The data was also obtained in BIL (band interleaved by line) format, which is a filing scheme for storing the actual pixel values of an image in a file. The .bil SPI data was 8 bit format.

According to a description of the SPI meta data from FEWS NET [185], the .bil images were compiled in unsigned 8 bit format. The SPI negative values were stored in reverse order starting at the top end of the possible values range for each data file. Since the 8 bit .bil files have a value range 0–255, the -1 values were stored as 255, -2 values were stored as 254, etc. Therefore, to convert the 8 bit .bil data into ArcInfo grid image SPI data, the formula:

$out\_grid\_SPI = con(in\_grid\_SPI >= 128, in\_grid\_SPI - 256, in\_grid\_SPI)$ was implemented in ArcGIS 9.3 [203]. Then all the three-month SPI images obtained for June, July, August and September were used for the monthly prediction models in the DroughtObject models.

All the static and dynamic attributes were subsetted and the same spatial extent was obtained for the datasets. After getting same spatial extent for all the datasets, they were also converted to same spatial resolution (8000m). Then Mapcubist was used for implementing the models in DOS command as *mapcubist_v202...modelName..HFA..outpuMapName.img..maskFileName.img*. The overall workflow for these data integrations is presented in Figure 5.23.

Using the regression tree DroughtObject models (Figure 5.8), the 2011 drought prediction maps were produced and the models' output maps are presented in Figure 5.24a–j. The one-month prediction of June showed the drought severity in the northeastern, eastern and southeastern parts of Ethiopia. The central, western,

southwestern and northern parts of Ethiopia were found to be in the moist and very moist drought category (Figure 5.24a). The two and three months' prediction (for the month of August and September, respectively) also showed a pattern similar to the one-month June prediction map (Figure 5.24b and c). The four-month June prediction model showed that there were vegetation stresses in the eastern and southeastern parts of the country. The model accuracies of the June two, three, and four months' predictions were found to be low (Table 5.5), and accordingly the maps produced have high uncertainties compared to the June one-month prediction.

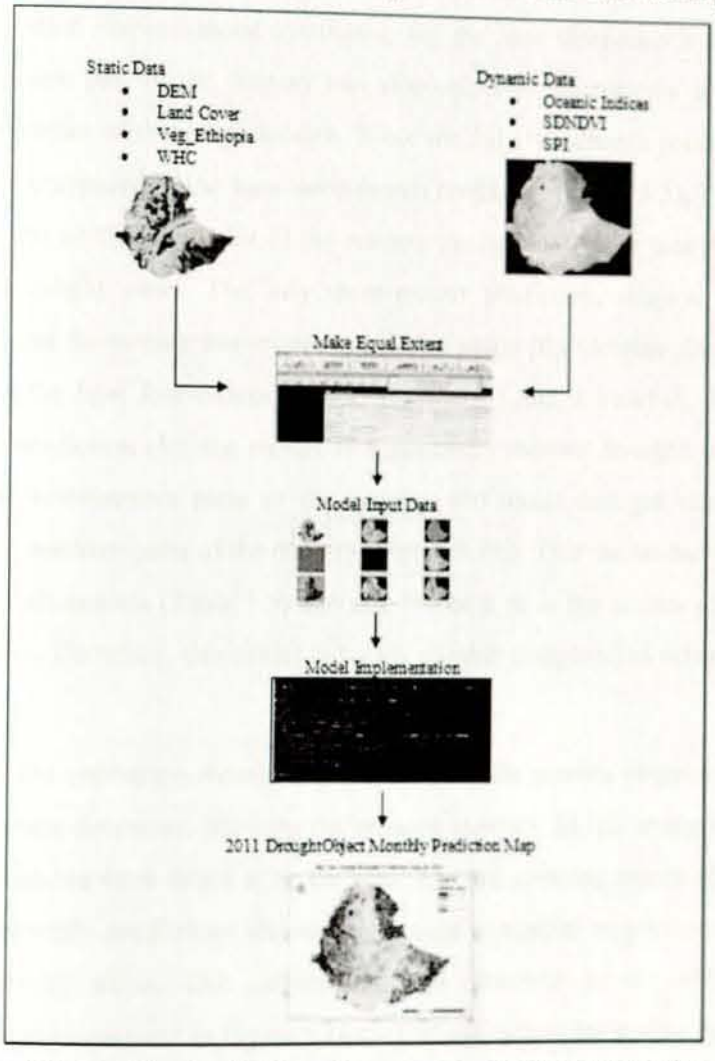


Figure 5.23: Workflow in 2011 drought assessment using DroughtObject model.

The July one-month prediction map (for the month of August) showed that there were extreme vegetation stresses in the eastern and southeastern parts of Ethiopia.

These areas were also reported to be drought-affected areas during 2011 by the Ethiopian government (personal communication with senior expert at Disaster Risk Management and Food Security Sector of Ethiopia). The central and southern parts of Ethiopia were also found to be below average condition during this month. The northeastern part of Ethiopia, which was stressed during July (Figure 5.24a), was in the moist category in August. This may be due to summer rain improvements at the end of July and August. The July two-month prediction map (for the month of September) showed the reverse of the June three-month prediction: the July two-month prediction showed moist conditions, but the June three-month showed that the northeastern part of the country was stressed. This discrepancy is due to the high uncertainties of these two models. Since the July two-month prediction has a high r value compared to the June three-month prediction (Table 5.5), it is possible to say that the northeastern part of the country during September was relatively in the moist drought class. The July three-month prediction, August two-month prediction, and September one-month prediction maps (for October drought) were the same as the June four-month prediction (Figure 5.24c, i, j and d). The August one-month prediction (for the month of September) showed drought stress in the eastern and southeastern parts of the country and moist drought classes in the western and northern parts of the country (Figure 5.24i). This model had the highest accuracy of all models (Table 5.5) and also the best fit in the scatter plot analysis (Figure 5.8h). Therefore, this model is highly reliable compared to other models in this analysis.

Overall, as the prediction month length increases, the pattern observed from the prediction maps decreases, showing the drought severity. In this analysis, the one-month predictions were found to be the best. The last growing month (October) in all of the monthly predictions showed no change compared to previous historical drought severity status. This pattern was also observed for the 1984 drought severity status assessment in Figure 5.10a – j. These reflect the reality that October is the end of the growing season and is usually dry. Therefore, no SDNDVI deviation is expected in this month.

The DroughtObject model prediction SDNDVI values were compared with the actual satellite-measured SDNDVI values (Figure 5.25). The highest correlation coefficient (R^2) was obtained for the August one-month prediction, followed by the June one-month prediction. Unexpectedly, the R^2 value for the July one-month prediction was found to be low (Figure 5.25e). Surprisingly, the July four-month prediction R^2 was also greater than both the July one-month and two-month prediction values and needs further investigation. With the exception of July, as prediction months increase, the R^2 values were observed to be decreasing, and this is in line with our expectations.

The 2011 drought severity status analysis showed that the DroughtObject model, which was developed from the 24-year historical data from biophysical, climate and satellite sources, can be used for predicting future drought episodes. This experiment has been done by integrating different temporal and spatial resolution images of satellite products. For this purpose, the required modeling implementation files ***.icase***, ***.names***, and historical ***.data files*** were also produced, and they are ready for the model implementation. These models will also be used for assessing and predicting 2012 and subsequent years' drought severity status.

This research was done using low spatial (8000m pixels) and temporal (on monthly bases) resolution data. The observed uncertainties were as expected because of the high heterogeneity of Ethiopian ecosystems. Future research plans are to repeat the modeling approach using high spatial and temporal resolution data and assess how much information gain can be achieved.

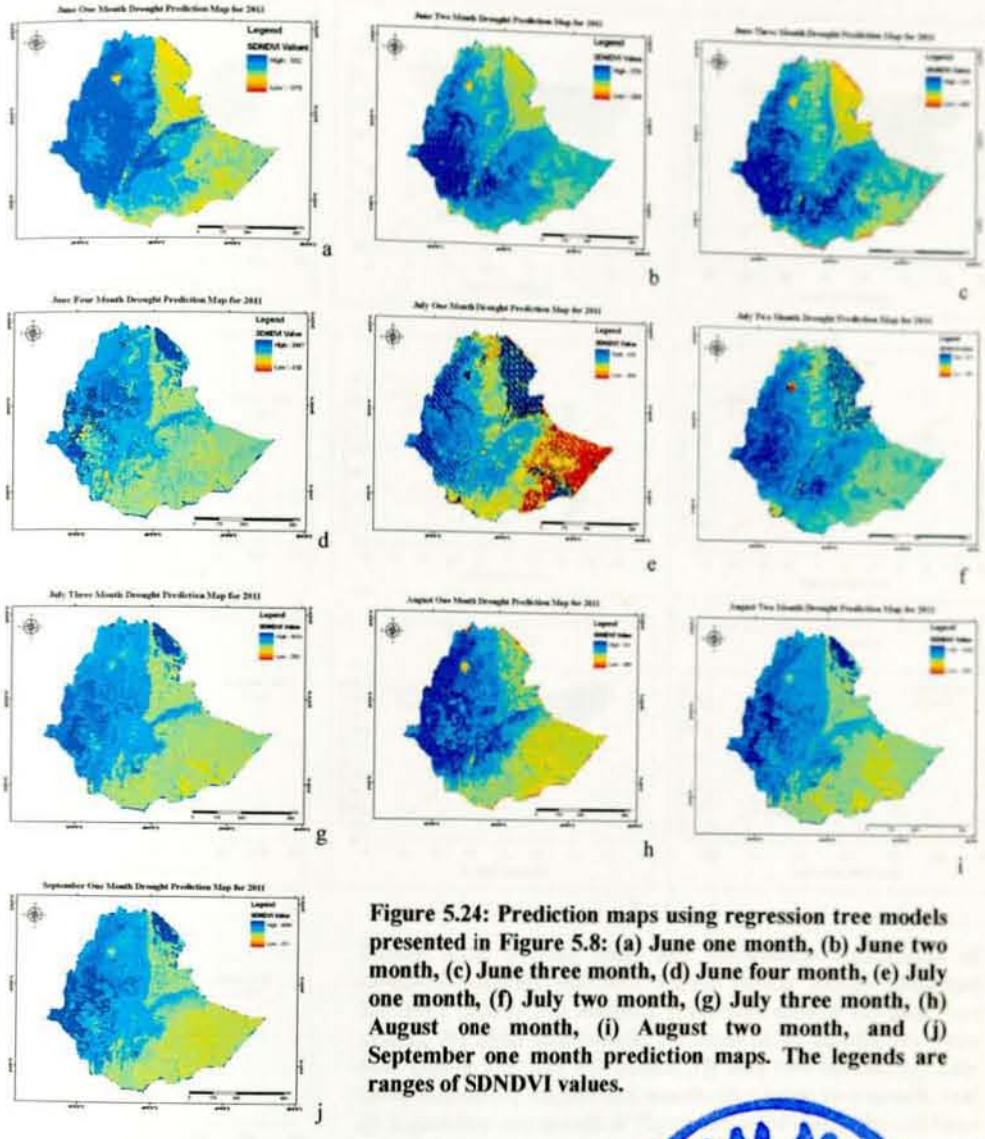


Figure 5.24: Prediction maps using regression tree models presented in Figure 5.8: (a) June one month, (b) June two month, (c) June three month, (d) June four month, (e) July one month, (f) July two month, (g) July three month, (h) August one month, (i) August two month, and (j) September one month prediction maps. The legends are ranges of SDNDVI values.

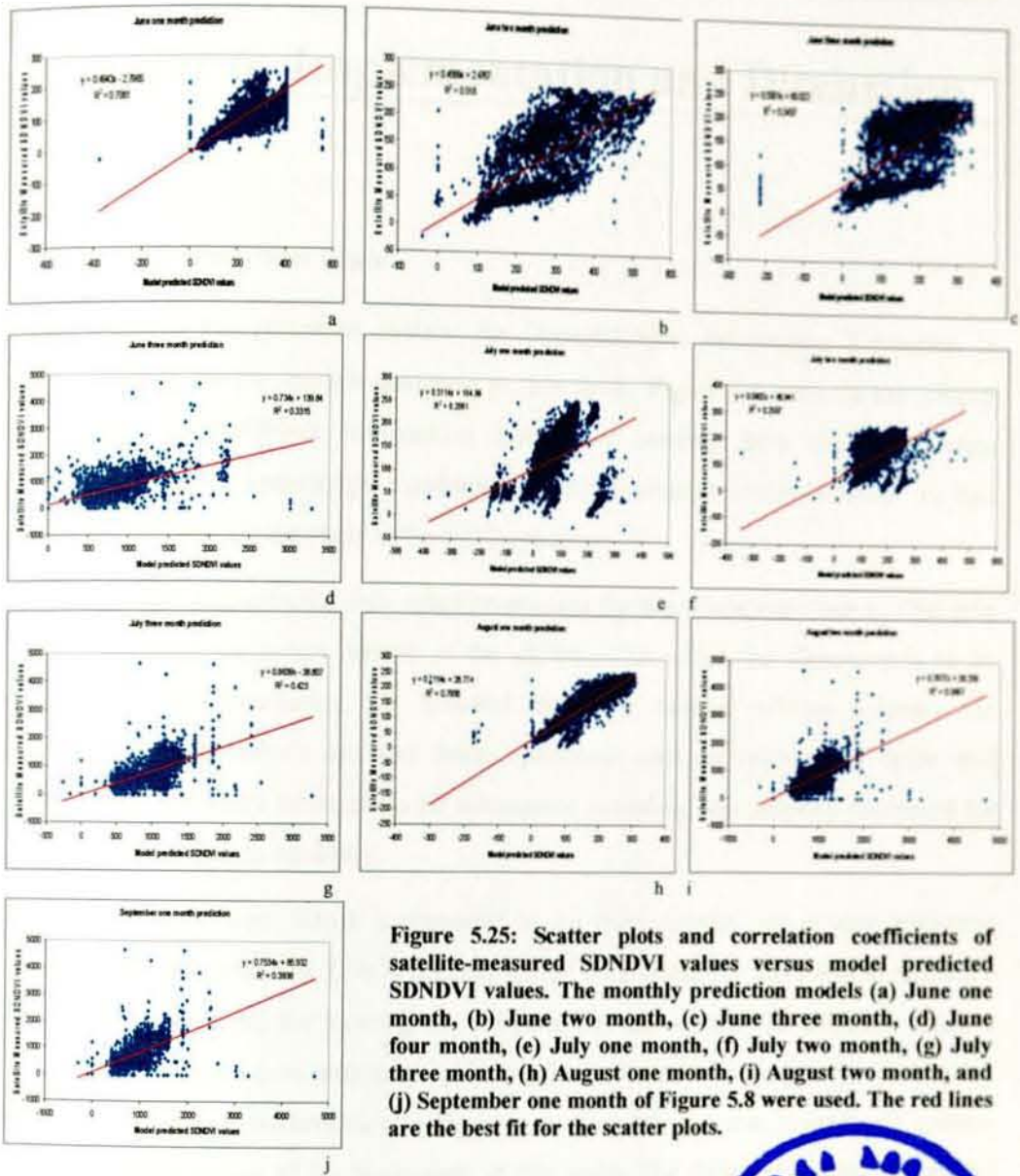


Figure 5.25: Scatter plots and correlation coefficients of satellite-measured SDNDVI values versus model predicted SDNDVI values. The monthly prediction models (a) June one month, (b) June two month, (c) June three month, (d) June four month, (e) July one month, (f) July two month, (g) July three month, (h) August one month, (i) August two month, and (j) September one month of Figure 5.8 were used. The red lines are the best fit for the scatter plots.

Chapter 6 Implementation and Discussion

6.1 Implementation Plan

In this section, the prototype system for DroughtObject Information Extraction is presented based on the models designed in this work. Figure 6.1 presents the overall process of DroughtObject Information Extraction, starting from the actual data preprocessing up to knowledge construction and information communication. In this figure, the columns are the steps in the KDD process.

The first column presents the data selection process for the whole experiment. This step is the creation of the target dataset of the attributes on which the discovery is to be performed. After structuring the required data, the second column presents the preprocessing step, which includes basic operations such as removal of noise and collecting the necessary information for subsequent modeling, was properly processed for both static and dynamic variables.

The transformation step, which is presented in the third column, was a very important step in spatial data analysis. This is because spatial data inherently have spatial reference (longitude and latitude) for locating the object in the real world. The different datasets coming from different data sources should have the same spatial reference for the overlay analysis and logical mathematical computations. For this purpose, coordinate system transformation was one of the basic steps at this stage. The different datasets were also standardized (such as SDNDVI) for logical mathematical computations.

The fourth column presents the rule and data mining step. Two data mining techniques were used at this step, based on theoretical recommendations as discussed in the review section. First, the relationships between independent attributes (explanatory attributes) and dependent attributes were explored using ANN. Then the required rules were tested and developed using the cubist regression tree data mining technique. The machine learning and prediction experiments were also exercised at this step.

The fifth column presents the actual information extraction and visualization step. The outputs of the data mining patterns were interpreted at this step. For this purpose, standard cartographic output means (maps) were used. The model implementation outputs were also evaluated at this step.

In all these processes, recent developments in computations, such as geographic information systems (GIS) and knowledge discovery in database (KDD), were used. The objective of this chapter was to develop a new intelligent system-based concept for drought information extraction and demonstrate the new concept developed in this dissertation research.

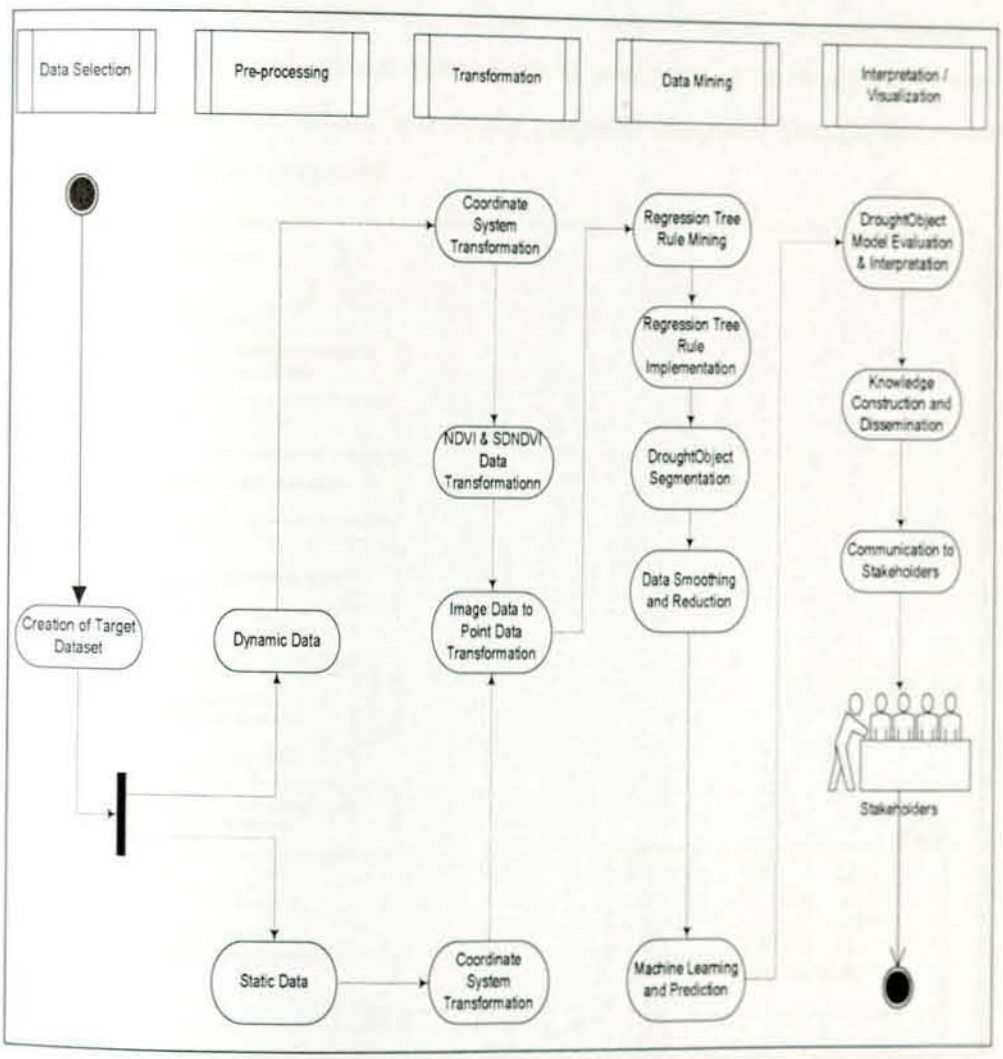


Figure 6.1: Generalized DroughtObject Information Extraction System work flow.

The new concept developed in this research is the identification of DroughtObject from a group of pixels characterizing drought and published in different articles [50-54]. The identification of DroughtObject from group of pixels involved a number of scientific investigations, including the identification of the key attributes characterizing drought, modeling drought using these key attributes, and evaluating the reliability of the models. The implementation and discussions chapter focuses on the concept demonstration of DroughtObject characterization and machine learning and reporting the target objects for the end users. The steps involved specifically in the image analysis part of this prototype presentation are presented in Figure 6.2. In the following subsections, Drought Information Extraction Framework, system use case scenario diagram description, experimental analysis and concept demonstrations, evaluation of the Drought Information Extraction System, discussions, and finally proposed Integrated Drought Information System Framework are presented.

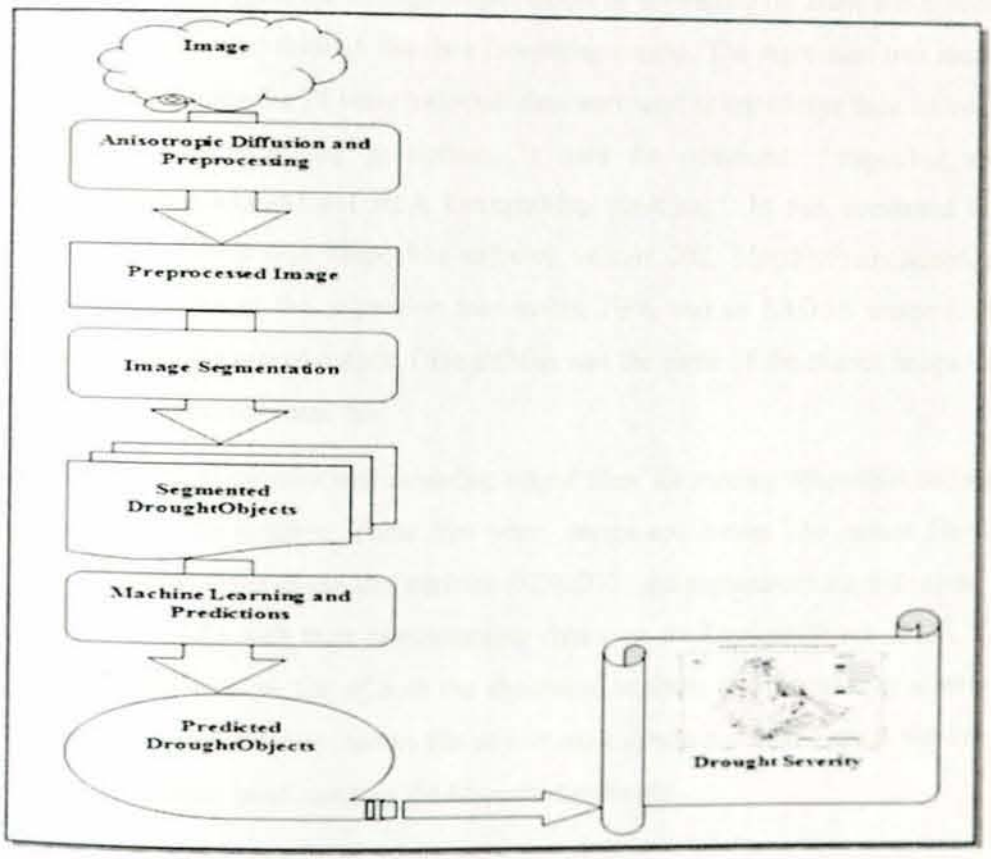


Figure 6.2: Work flow for demonstrating the major steps followed in DroughtObject information extraction and prediction prototype system.

6.2 Drought Information Extraction Framework

Figure 6.3 presents Drought Information Extraction design, which is a four-tier framework including Data Processing Component, Knowledge Base Triggering Component, DroughtObject Identification Component and Knowledge Construction and Presentation Component. Each designed component is briefly described in the following subsections.

6.2.1 Data Processing Component

This component accepts both static and dynamic raw data and processes for the subsequent engines. The data interoperability is included in the system for integrating the different input attributes for DroughtObject modeling.

6.2.2 Knowledge Base Triggering Component

This engine triggers the DroughtObject model by combining the static and dynamic datasets accepted through the data processing engine. The regression tree models developed using the 24 years historical data were used as knowledge base for one to four months time lag predictions. It uses the command: "mapcubist_v202 MonthlyPredictionModel HFA DroughtMap mask.img". In this command line, mapcubist_v202 was Mapcubist software version 202, MonthlyPredictionModel was the name of the regression tree model, HFA was an ERDAS image (.img) format for the model output, DroughtMap was the name of the output image, and mask.img was the mask file.

Besides the regression tree modeling output files, for running Mapcubist software two files were required. These files were: .names and .icases. The .names file was the list of both the dependent attribute (SDNDVI) and explanatory attributes (the 11 key attributes) with their corresponding data type for DroughtObject model. The .icases file was the list of both the dependent attribute and explanatory attributes with the same order as .names file and its main function was to present the image types and their band numbers for Mapcubist software.

6.2.3 DroughtObject Identification Component

This component has two major modules: anisotropic diffusion and image segmentation. The algorithm for anisotropic diffusion is also called Penora–malik diffusing[263] (Appendix 3) and it is the technique used for removing noise from digital images without blurring edges with a constant diffusion coefficient. The fuzzy image segmentation module (Appendix 4) is the actual DroughtObject creation module using standard drought severity classes defined by Mckee et al. [28], McKee et al. [29] and Tadesse et al. [71].

6.2.4 Knowledge Construction and Presentation Component

The main task of this component is displaying and presenting the analysis output for the system user. This functionality is also important for extracting the relevant patterns for further analysis. It has three major modules: publish server services, prepare REST Protocol and Integrated Service Interface.

The Integrated Service Interface serves as an information for delivering drought data/information request from different clients. At the backend it has integrated services server (ISS), which is installed at the drought prediction engine. ISS here is designed to allow multiple searches and data/information delivery. The end product of the search from the client side is displayed on GoogleEarth platform for locating the data/information on the real world context display.

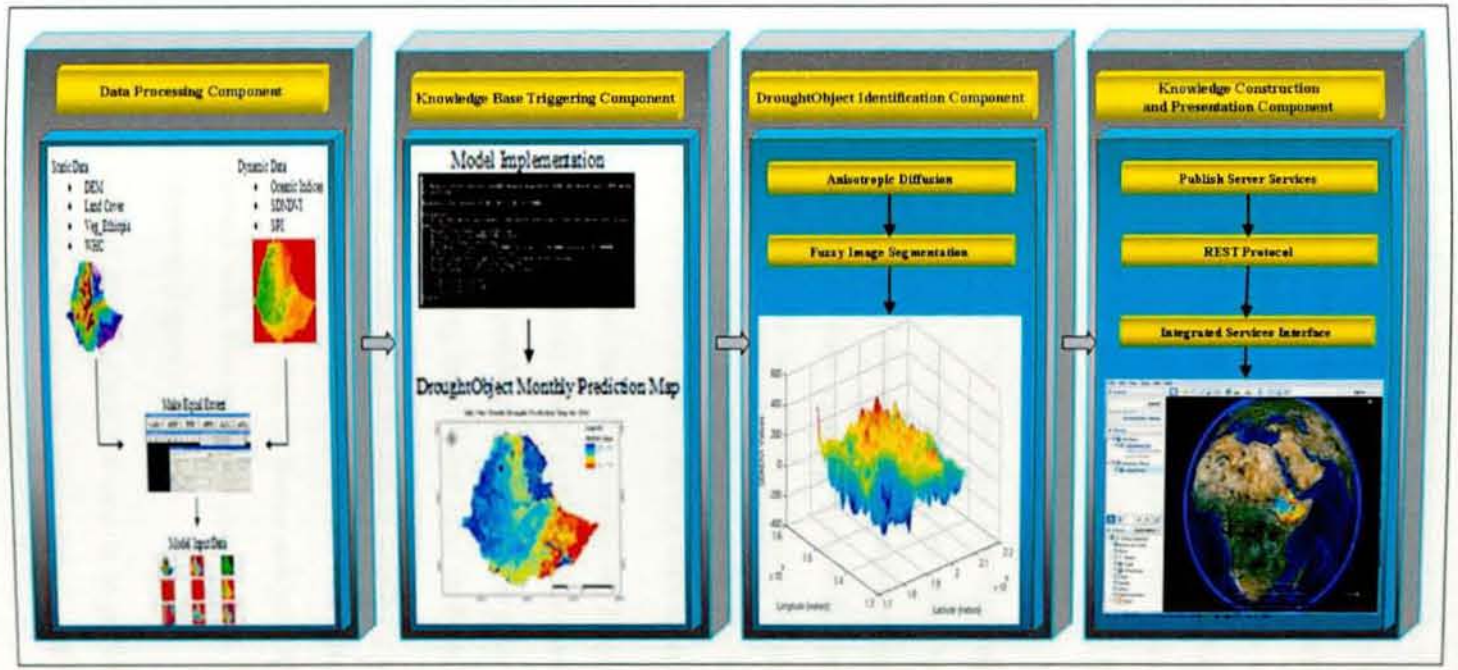


Figure 6.3 Drought Information Extraction Framework.

6.3 System Use Case Diagram Description

To demonstrate the DroughtObject Information Extraction concept, unified modeling language (UML) was used. The UML use case diagram for the prototype is presented in Figure 6.4. This figure is used to depict the functionality and behavior of the system from an external point of view. There are four use cases: InputData, AnalyseData, StoreData and ViewData/Information for DroughtObject Information Extraction System. The use cases here present the functional requirements of the system at macro level, which capture the intended behavior of the overall system. These use cases were selectively taken from Drought Information Extraction Framework (Figure 6.3) for clearly demonstrating the system design and architecture.

The behaviors of InputData use case are expressed by performing the task of accepting spatial data. The input data can either be directly forwarded for geo-database storage or used for analysis. The actors directly involved with this use case are data collector and data analyst (Figure 6.4). The inputData use case has two include: PredictDrought and ProducePredictionMaps (Figure 6.5a). Include use case relationship in these use cases is that the two use cases (PredictDrought and ProducePredictionMaps) are included as a subroutine in the inputData use case.

InputData use case also has three extensions: BiophysicalData, ClimateData and SatelliteData (Figure 6.5a). These additional use cases store a dataset in a different format compared to InputData use case, and the data interoperability is included in the system for integrating the different input attributes for DroughtObject modeling.

The main task of ViewOutput use case is displaying and presenting the direct input data or analysis output for the system user. This functionality is also important for extracting the relevant patterns for further analysis. All three actors involved in this use case interact with this use case (Figure 6.4). This use case also has four different extensions (Figure 6.5b).

The main task of AnalyseData use case is to perform DroughtObject extraction (such as anisotropic diffusion and segmentation), and it is the core of the developed

system in this research. The actor involved in this use case is the data analyst or the researcher (Figure 6.4). Four use cases were also included in this use case: InputData use case for temporary and/or permanent data storage, ViewModelOutputGraph, ViewPredictionMap, and ViewTabularData (Figure 6.5c).

The main task of StoreData use case is to store the different data inputs coming from direct data entry or analysis output in a structured and logical way for future retrieval. The actors interacting with this use case include data collector and data analyst (Figure 6.4). The four use cases (InputData, PredictionMap, DroughtReport and PredictionTabulardata) were also included in this use case (Figure 6.5d).

In addition to brief use case relationships, the class diagram for DroughtObject Information Extraction System is presented in Figure 6.6. In this diagram, there are five blocks with dashed line. The first block represents the interface of DroughtObject Information Extraction System. The interface of DroughtObject Information Extraction System defines the service offered by all classes, and it is presented at the top left corner. The packages, which groups similar computations are presented at the top of the blocks. The second block presents six classes (for all processes from data input up to compilation), which are included under ILWIS (Integrated Land and Water Information System) package. The third block presents two classes (for data storage and retrieval), and they are included under SQL Server package. The fourth block presents four classes (for all processes from data analysis up to pattern extraction), and they are represented under Matlab package. The fifth block presents five classes (for all processes from knowledge construction up to service integration), and they are represented under ArcGIS Server package.

In Figure 6.6, inheritance and dependency relationships of each classes is also presented. The inheritance relationships are presented using solid lines, which are drawn from child classes (the class inheriting the behaviors) with a closed and unfilled arrow head pointing to the super classes. The dependency relationships of the classes are shown with dashed arrow, which mean that the client class make use of the supplier in the computation processes [264]. Access dependency relationships of the packages are also presented at the top of the figure between the

packages, which include the ability of one package to access the output of the other package.

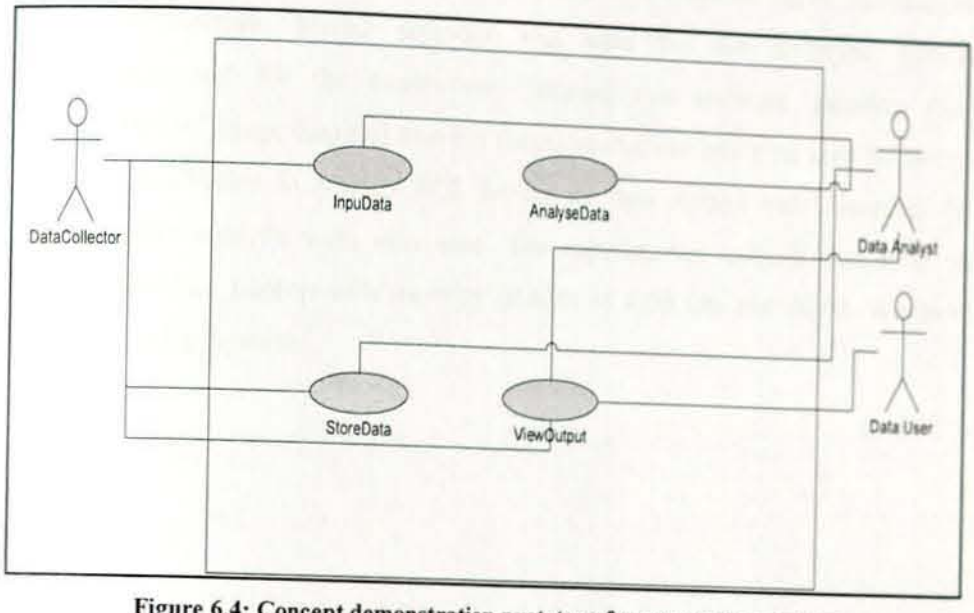


Figure 6.4: Concept demonstration prototype for use case interactions

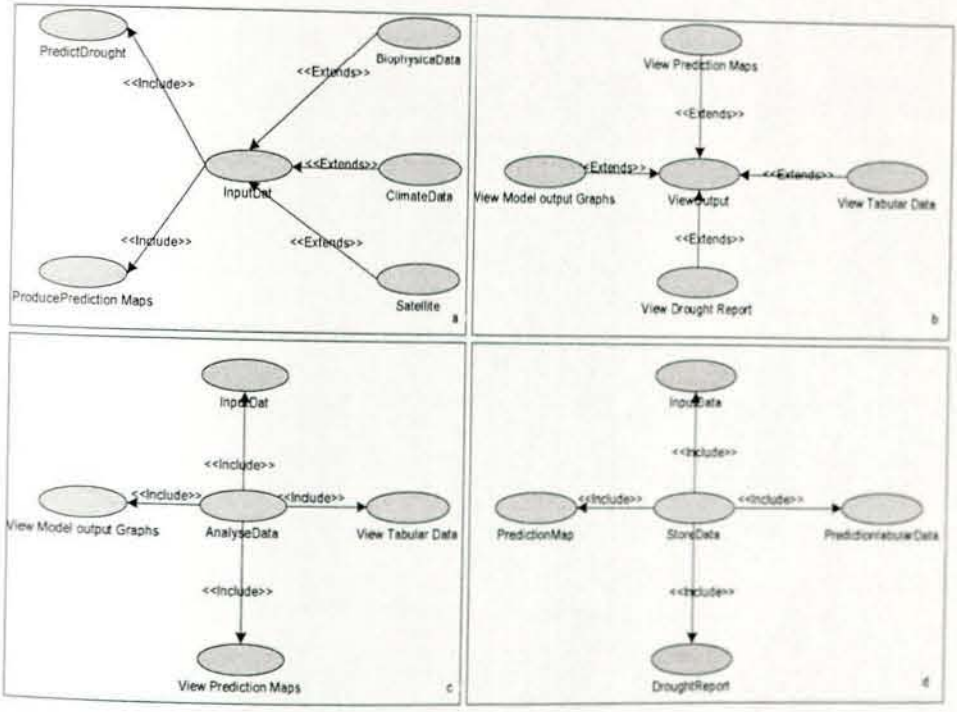


Figure 6.5: Details of use cases shown in Figure 6.4.

The class diagram in Figure 6.6 is presented for demonstrating how the DroughtObject Information Extraction System was planned for extracting pattern from the image data and it is presented as a proof of concept for future full-fledged system architecture. Matlab software was used for the prototype system development and for the experiment, because this software supports easy manipulation of image data and also has functions that can easily be used for image analysis. In addition to Matlab, SQL Server for data storage and Mapcubist for image pattern analysis were also used. The experimental analysis was done on Toshiba Satellite Laptop with memory (RAM) of 4.00 GB and 64-bit Windows Vista Operating System.

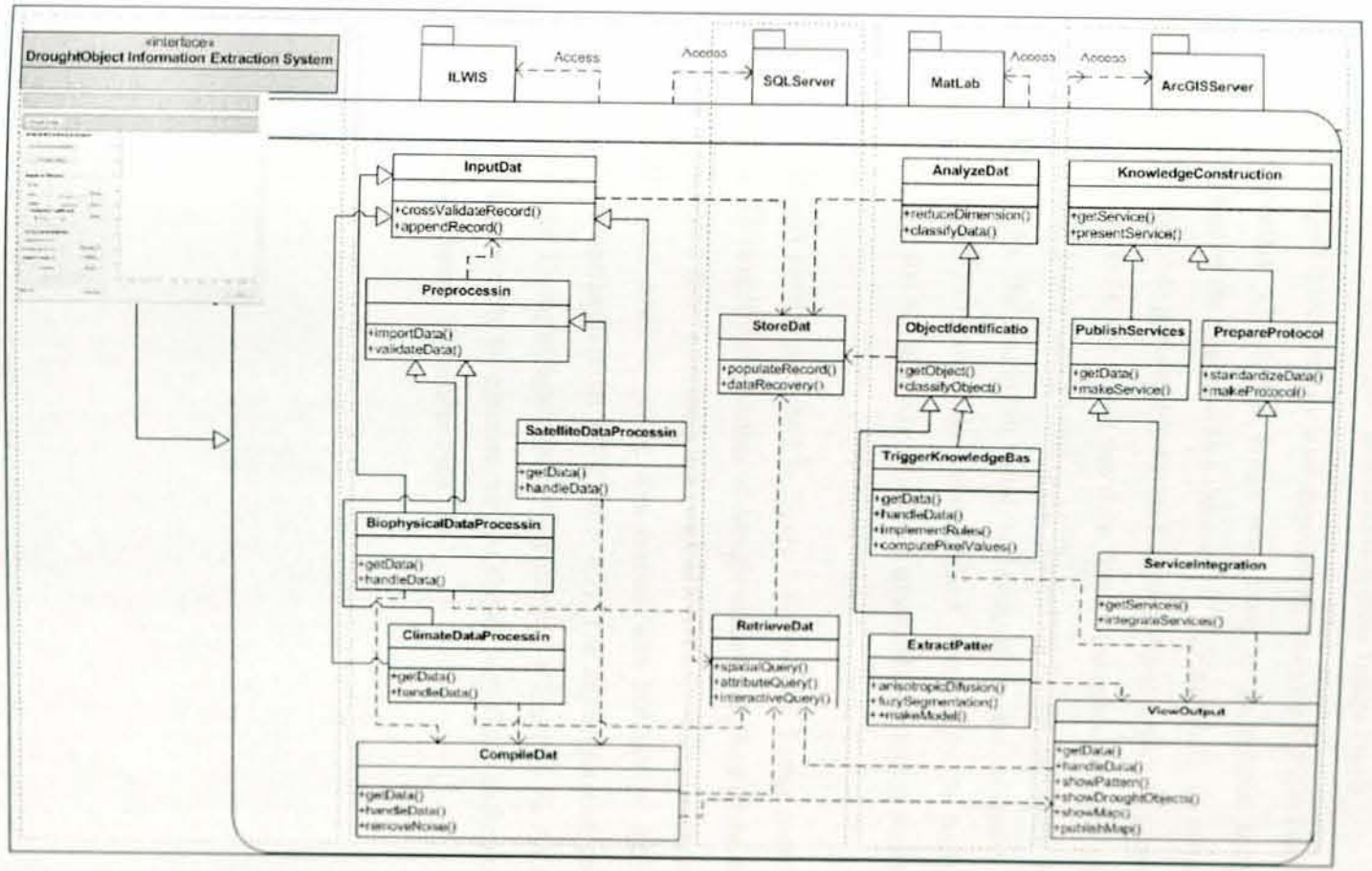


Figure 6.6: Class diagram of the DroughtObject Information Extraction System.

6.4 Experimental Analysis and Concept Demonstrations

6.4.1 Process of DroughtObject Identification from Image Pixels

During the DroughtObject identification experiment, when the SDNDVI image was displayed in Matlab v7.9.0, the image was found to be blurred because of background values of the image (with a value of 10^{-38}) (Figure 6.7). To manage this blurring effect, the code presented in Figure 6.8a was implemented. As presented in the implementation code, the search was done first for the minimum value of all the pixels (background pixels with a value of 10^{-38}) and then for the maximum value of all the pixels. Then, the maximum value was assigned for the minimum values (with a value of 10^{-38}) to remove the blurring effect. Accordingly, the background becomes white, and the actual image with the original color contrast is presented in Figure 6.8b.

The DroughtObject pattern presented in Figure 6.8b was also further tested using contour values. To simplify the number of drought classes, only four of the drought classes are presented here: extremely dry (with SDNDVI <= -0.2), moderately dry (with SDNDVI >-0.2 AND <= -0.05), near normal (with SDNDVI > -0.05 AND <=0.1) and above optimum (with SDNDVI > 0.1). The implementation code for contouring the four DroughtObject classes is presented in Figure 6.9a. As can be observed from Figure 6.9b, the extreme drought classes were found in the northern part of the country during September 1984.

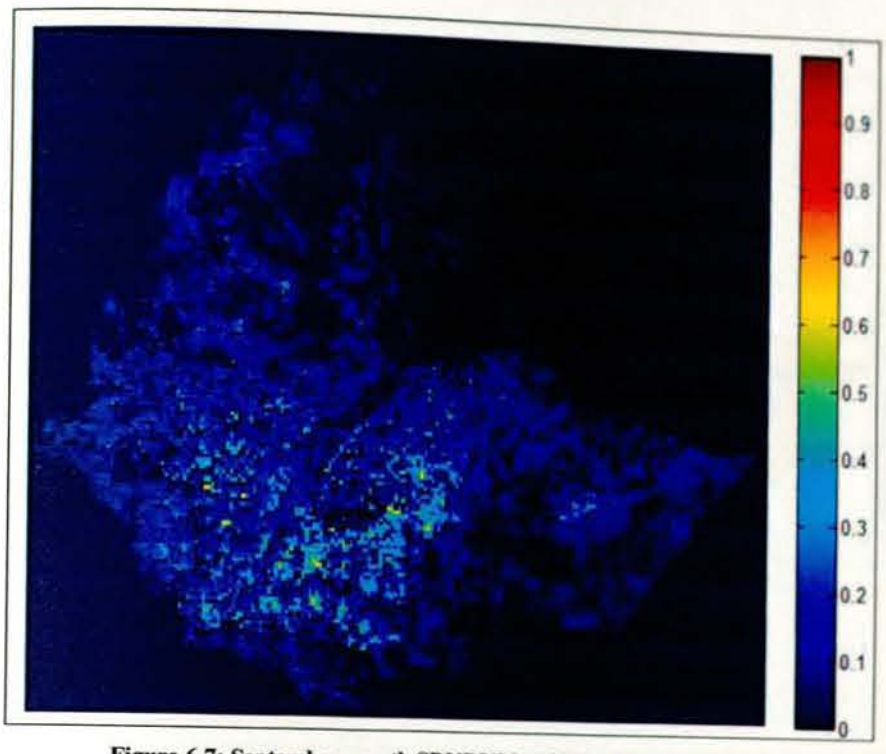
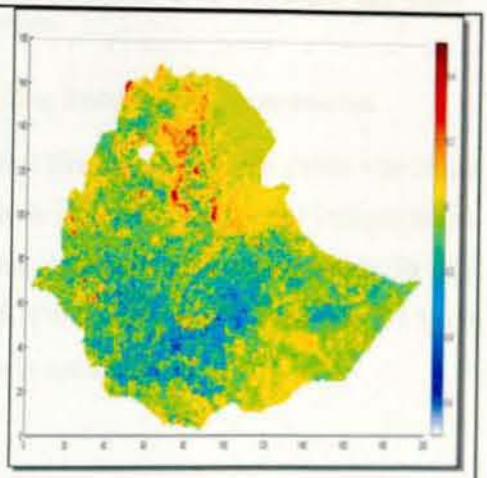


Figure 6.7: September month SDNDVI in 1984 on Matlab software

```
%The following code read the SDNDVI image and adjust the
background blurring effect

 img=imread('SDNDVI8409.tif');
[r,c]=size(img);
m=min(min(img));
ma=max(max(img));
for i=1:r
   for j=1:c
      if(img(i,j)==m)
         img(i,j)=ma;
      end
   end
end
img=-1*img;
img=im2double(img);
img2=img;
for i=1:r
   img(i,:)=img2(r-i+1,:);
end
   [r,c]=size(img);
[x,y]=meshgrid(1:c,1:r);
surf(x,y,img);
```

a

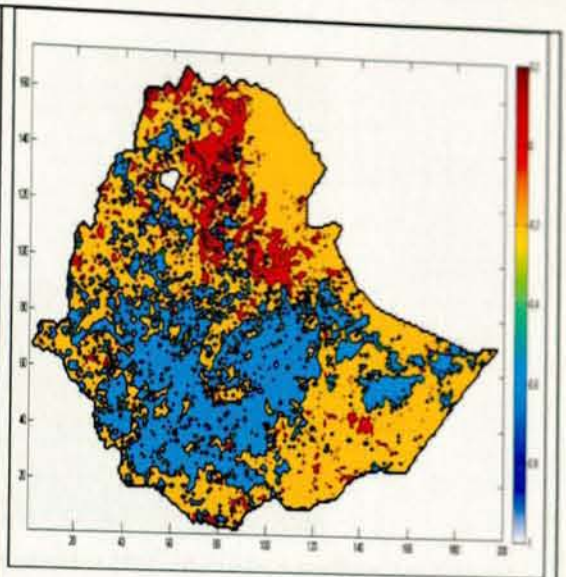


b

Figure 6.8: A sample of droughtObject characterization and identification: a) implementation Matlab code for removing blurring effect, and b) DroughtObject patterns after removing background blurring effect.

```
%Matlab implementation code for contouring
[r,c]=size(img);
m=min(min(img));
ma=max(max(img));
for i=1:r
  for j=1:c
    if(img(i,j)==m)
      img(i,j)=ma + 0.001;
    end
  end
end
mi=min(min(img));
img=-1*img;
img=im2double(img);
[r,c]=size(img);
for i=1:r
  for j=1:c
    if (img(i,j)>=0.2)
      img(i,j)=0.2;
    elseif ((img(i,j)<0.2)&&(img(i,j)>=0.05))
      img(i,j)=0.05;
    elseif ((img(i,j)<0.05)&&(img(i,j)>=-0.1))
      img(i,j)=-0.1;
    elseif ((img(i,j)<-0.1)&&(img(i,j)>=-0.6))
      img(i,j)=-0.6;
    else
      img(i,j)=-1;
    end
  end
end
[x,y]=meshgrid(1:c,r:-1:1);
[c,h]=contourf(x,y,img);
set(gca,'XTick', 0:5:c);
set(gca,'YTick', 0:5:r);
axis equal;
```

a



b

Figure 6.9: Contouring drought classes a) Matlab implementation code, b) spatial distributions of the four drought classes in September 1984. The x-axis and y-axis presented in b are pixel locations from the origin.

6.4.2 DroughtObject Identification Using Fuzzy Image Segmentation

The concept of fuzzy DroughtObject identification from image pixels was tested using four pixels, which were obtained from the northeastern part of Ethiopia (since there were highly deviated pixels at this location). The sample pixels for this experiment are presented in Figure 6.10. Using these pixels, algorithm 5.1 [118] was implemented for fuzzy DroughtObjects extraction.

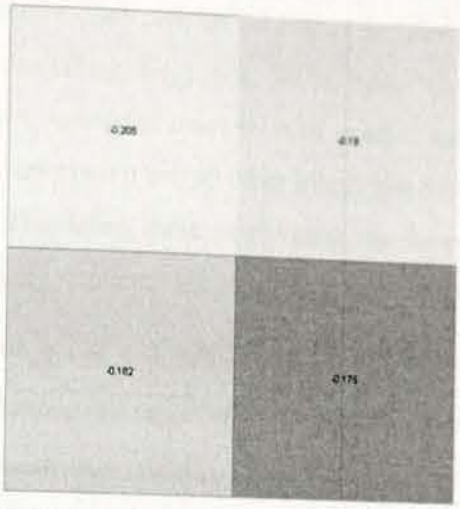


Figure 6.10: NOAA SDNDVI of September 1984 selected from northeastern Ethiopia.

Algorithm 5.1: Fuzzy DroughtObjects extraction from the four experimental pixels[117].

1. *Define a seed-point c in the input image.*
2. *Form a temporary queue Q and a real-valued array* f_c *with one element* $f_c(d)$ *for each pixel d,*
3. *For all pixels* $d \varepsilon C$ *, initialize array* $f_c(d) := 0$ *if* $d \neq c$; $f_c(d) := 1$ *if* $d = c$.
4. *For all pixels* $d \varepsilon C$ *for which fuzzy pixel adjacency* $\mu_{\Psi(c,d)} > 0$ *, add pixel d to queue Q.*
5. *While the queue Q is not empty, remove pixel d from queue Q and perform the following operations:*

$$f_{\max} := \max_{e \varepsilon C} \min(f_c(e), \psi(d,e))$$

if $f_{\max} > f_c(d)$ *then*

$f_c(d) := f_{\max}$

for all pixels g *for which* $\psi(d,g) > 0$ *, add* g *to queue Q*

endif

6 *endwhile*

7 *Once the queue Q is empty, the connectedness map* (C, f_c) *is obtained*

In algorithm 5.1, the concept of queue is used in the context of data structure, in which the pixels in the image are kept in order and the principal operations on the image are the addition of pixels to the rear terminal position and removal of entities from the front terminal position. This makes the queue a First-In-First-Out (FIFO) data structure. The equations in chapter 4 on fuzzy adjacency (equation 13) and fuzzy affinity (equation 16) were used here. The affinity map $\Psi(c,d)$ was produced by executing the Algorithm 5.1.

From Figure 6.10, it can be observed that the highly deviated pixel is the first pixel. For convenience of describing the pixels, let us name the pixels in clockwise direction, the first pixel A, second pixel B, third pixel C and fourth pixel D. The seed pixel (highly deviated pixel) and all other image pixels with their attributes are presented in Figure 6.11a. Using these pixel values, the fuzzy affinity values were calculated using the fuzzy segmentation algorithm (Appendix 4).

The SDNDVI in Figure 6.11a was represented by a pair $I = (C, f)$, in which C represents the image domain (all arrays of the image) and f represents local image properties (connectedness map values). Then, $f_{(c)\in[0,1]}$ represents a normalized image property (feature) associated with pixels of C. To connect the two pixels A and B in Figure 6.10, there were two possible links (either to pass directly A to B-clockwise direction or A to C to D to B-anti-clockwise direction). In this case, a path is the number of links between two pixels and a rectilinear distance was used. Each link that can be produced is the affinity between successive pixels and has got strength (affinity value).

The pixels c and d are not expected to be nearby [117], and in our context $f_{(d)}$ represented the SDNDVI values of every pixel. Along a path there are a number links, and the strength of a path is the strength of the smallest link. For each path, its strength is defined as the minimum affinity value of all pair wise consecutive elements on the path, so the strength of the entire path is defined by the strength of its weakest local connection. There can be different paths between two pixels, and in the process we strove to get the strongest path, which finally produced the connectedness map or affinity map. This concept is demonstrated in Figure 6.11.

In Figure 6.11, we have four links $(\Psi(A,B), \Psi(A,C), \Psi(C,D), \Psi(B,D))$. The calculation for each link is presented in Table 6.1. After getting the fuzzy affinity values, the links were assigned; they are presented in Figure 6.11b. Finally, using this approach, we were able to produce the fuzzy connectedness map, which was ready for applying reasonable threshold value and segmentation or object identification as described in Sonka et al. [117] and Udupa and Samarasekera [118].

Starting from a seed-pixel A and determining the fuzzy connectedness $\mu_{\psi}(c,d_i)$ to every other pixel d_i in the image domain C, assigning the corresponding connectedness value to every pixel, the resulting image was a fuzzy connectedness map representing the degree of connectedness of every pixel in the image with the seed-pixel A. Any degree of connectedness in the range [0, 1] is possible. The connectedness map with this concept is presented in Figure 6.12.

A very strong connectedness was denoted by 1, no connectedness by 0. By thresholding the connectedness map with an appropriate value, only pixels with a certain pre-determined minimum degree of connectedness to the seed-pixel remain. Thresholding the connectedness map yields the segmentation result [117]. By implementing this concept, the segmented image map for our experimental site is presented in Figure 6.13.

Table 6.1: Fuzzy affinity values calculation for the four pixels (let $k_1 = 1$ and $k_2 = 1$). The explanations of the equations used for deriving the values were presented in the method section.

Link	Adjacency $\mu(c,d)=\begin{cases}\frac{1}{1+k_1\sqrt{\sum_{i=1}^{n}(c_i-d_i)^2}} & \text{if } \sum_{i=1}^{n}\lvert c_i-d_i\rvert\le n,\\ 0\end{cases}$	*Link values (fuzzy affinity)* $\Psi(c,d)=\frac{\mu(c,d)}{1+k_2\lvert f_{(c)}-f_{(d)}\rvert}$
$\Psi(A,B)$	1	0.985
$\Psi(A,C)$	1	0.978
$\Psi(C,D)$	1	0.993
$\Psi(B,D)$	1	0.985

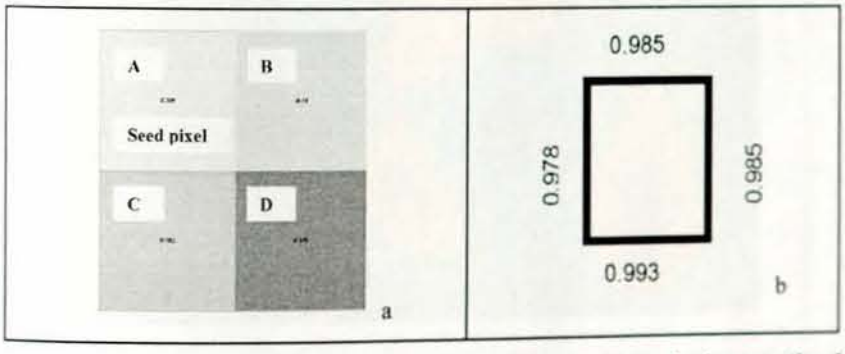


Figure 6.11: Identified seed pixel (a) and fuzzy affinity values calculated using equation 16 (b).

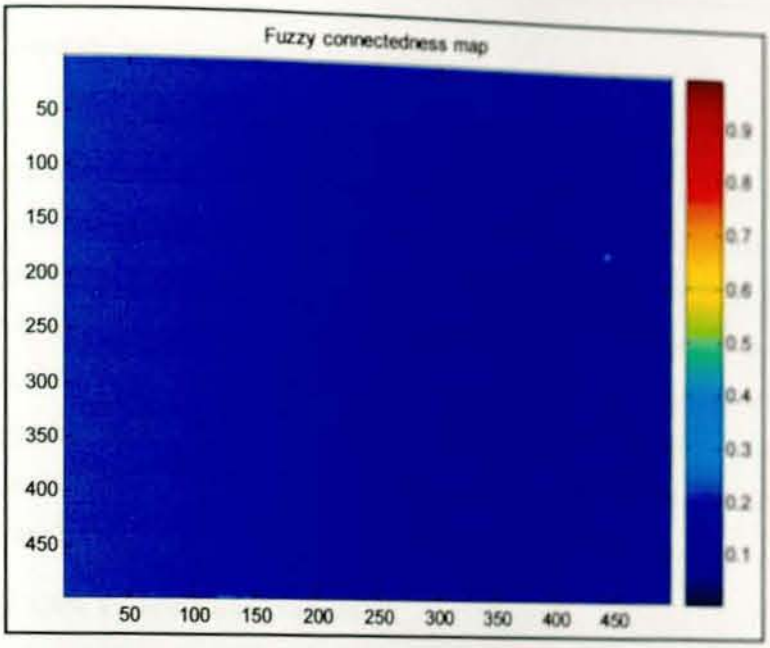


Figure 6.12: Fuzzy connectednesses map of all SDNDVI pixels in the experimental site.

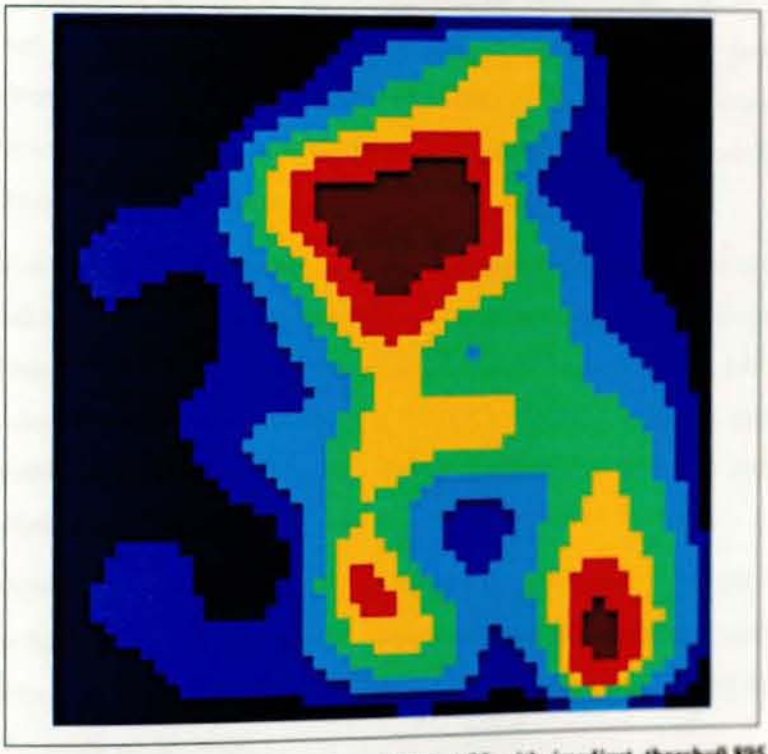


Figure 6.13: Segmented image map for selected AOI with imadjust thresh=0.895 and imagesc(z).

6.4.3 System Implementation and Usages

In this subsection, the user interface and overall architecture of Drought Information Extraction System is presented. The prototype graphical user interface demonstration was done on Matlab software v7.9.0 [204]. The main objective of this prototype presentation was demonstrating the concept of DroughtObject characterization and identification from a group of pixels and also to present the startup for future system development for the actual implementation of the developed concept. Therefore, the graphical user interface (GUI) part of the experiment presents ANN training and testing of drought pixels and subsequent image analysis for characterizing and identifying DroughtObject.

6.4.3.1 ANN Training and Prediction

After characterizing and identifying drought spatial object, the experiment focus was on predicting drought in one to four months' time periods. The 24 years of historical data were used for training and testing the networks for these predictions. There were four major steps in the training process of the networks: 1) assemble the training data, 2) create the network object, 3) train the network, and 4) simulate the network response to new inputs.

For the assembly of the training data, the attributes were ordered in one column of different rows and the training data were split into input example data (explanatory independent attributes) and target data (dependent attribute, i.e., target SDNDVI to be predicted). The network object was created by training it until it learned the relationship between the examples inputs (independent attributes and target dependent attributes).

The ANN analysis part of the GUI presents two generalized command buttons: process backpropagation and process testing. The process backpropagation button represents the training mode of operation and the process testing button represents the using mode of the ANN.

A supervised training method was used for this experiment. The reason for using supervised training instead of unsupervised training is that during the experimental training, the expected outputs are already known (the DroughtObject). In this kind

of scenario (when expected outputs are known), both MathWorks [204] and Heaton [265] suggested that the appropriate training method is supervised training. The work flow in the DroughtObject training experiment is presented in Figure 6.14. In this training process, the neural network adjusts the values in the weight matrix based on the differences between the anticipated output and the actual output. A sample of the learning curve with learning rate of 0.04 is presented in Figure 6.15. In this training experiment, a backpropagation algorithm was used (Appendix 5).

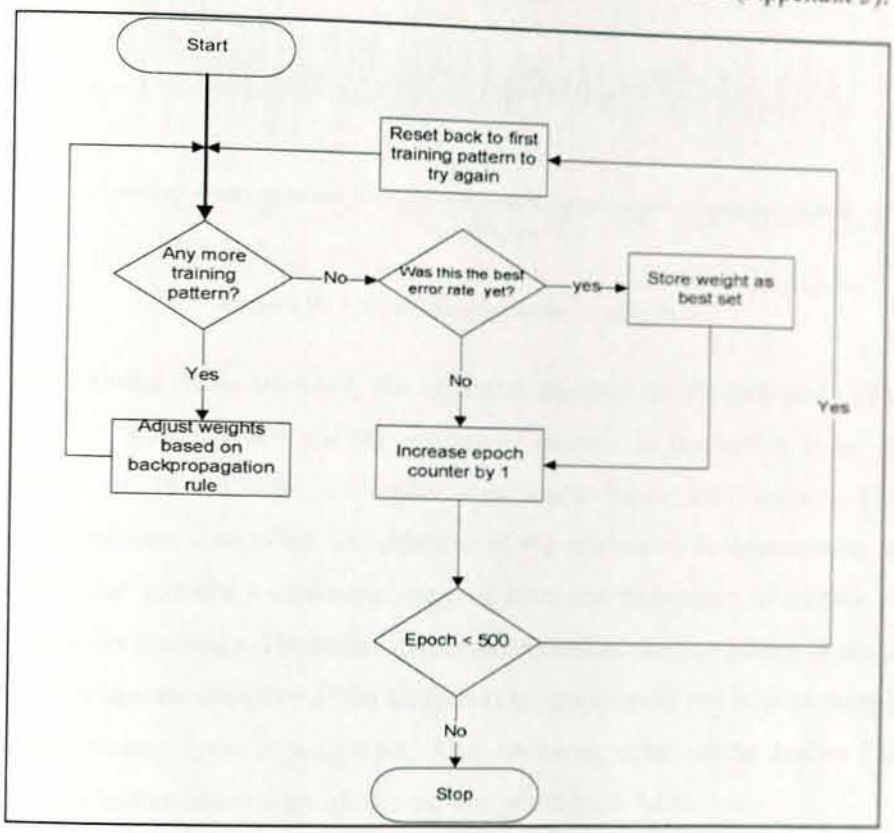


Figure 6.14: Workflow in supervised DroughtObject training.

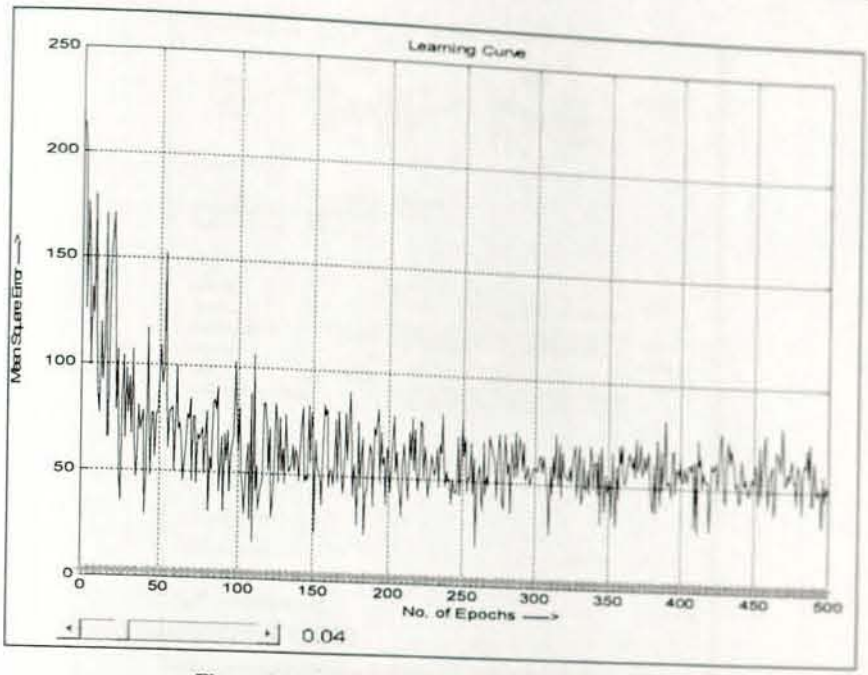


Figure 6.15: ANN learning curve on training data.

In the training mode of ANN, the algorithm requires the determination of the number of hidden layers and the number of neurons in the hidden layer. The decision on the number of layers was made based on Heaton's [265] recommendation that, when the objective of the analysis is to approximate any function that contains a continuous mapping from one finite space to another, one hidden layer is enough. The author further recommended that two hidden layers can be used when the objective of the analysis is to approximate any smooth mapping to any accuracy level of the output. After reviewing other similar studies [266, 267], two hidden layers were used in our analysis (Figure 6.16).

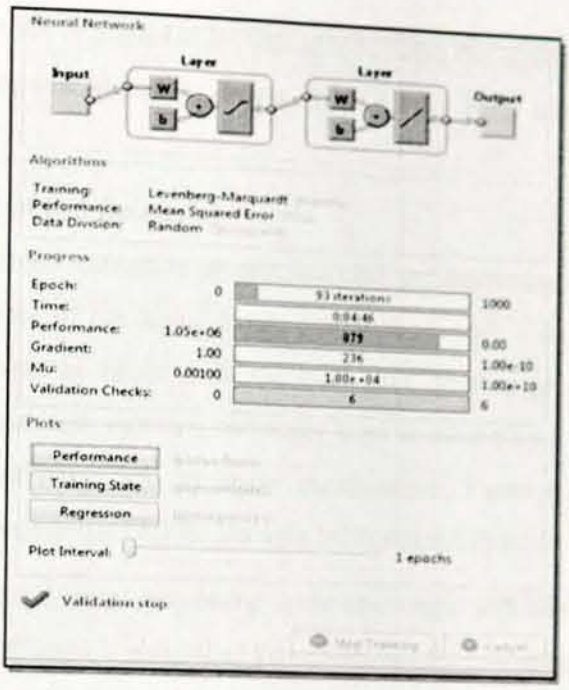


Figure 6.16: A screen shot from ANN backpropagation training.

Several studies have been done on the number of neurons in a given hidden layer for better prediction accuracy [62, 268-272]. The challenge here is that using too few neurons in the hidden layers results in underfitting, and using too many neurons in the hidden layers results in overfitting [265]. Underfitting here is defined as too few neurons in the hidden layers to adequately detect the DroughtObject pattern from image pixels; and overfitting occurs when the neural network has so much information processing capacity that the limited amount of information contained in the training set of DroughtObject pixels is not enough to train all of the neurons in the hidden layers. For this experiment in the GUI, 20 neurons were used (since relatively good performances were found compared to below 20 and above 20) for each hidden layer, and the network was trained for 500 epochs using a back propagation algorithm with learning rate of 0.04.

To implement the using mode of the ANN, required functions such as readData and SaveToFile were developed. In this mode, when a thought pattern is detected at the input, its associated output becomes the current output; and if the input pattern does not belong to the thought list of input pattern, then the firing rule is used to determine whether to fire or not. The algorithm and implementation code for the

using mode of ANN is presented in Appendix 6. Using this algorithm, the network was tested on an unseen sample dataset. The GUI for ANN is also presented in Figure 6.17.

6.4.3.2 DroughtObject Extraction

In the DroughtObject extraction process, the GUI and interaction design approach keeps the sequence of the digital image processing in such a way that using the system user selects an image from the local hard drive or geo-database. After accessing the input data, there are two major steps in this digital image processing: 1) anisotropic diffusion, and 2) image segmentation. Figure 6.17 presents the sample screen shots of the GUI for Drought Information Extraction System.

The first step in this image processing is the anisotropic diffusion. The algorithm for anisotropic diffusion is also called Penora–malik [263] diffusing. The algorithm and implementation code for this diffusion is presented in Appendix 3. It is the technique for removing noise from digital images without blurring edges with a constant diffusion coefficient. To load the image, for anisotropic diffusion, browse image button can be clicked and it can display the selection preference dialog list with the lists of the images to be displayed (Figure 6.18a). Then the image is selected and displayed in the GUI, as presented in Figure 6.18b. After loading the image on the workspace, if the user accepts the default values of number of iterations=20, delta_t= 1/7, kappa=45, the intended diffusion can be produced. These values are obtained from the requirement of anisotropic diffusion function, which needs at least four arguments (ANISOTROPICDIFF(IMAGE, NUM_ITER, DELTA_T, KAPPA, OPTIONS). In this argument, IMAGE is any gray scale image with MxN matrices of pixels ready for diffusion; NUM_ITER is the number of iterations that the algorithm has to iterate; DELTA_T is integration constant (0 <= delta_t <= 1/7) (usually this parameter is set to its maximum value 1/7); KAPPA is gradient modulus threshold that controls the conduction; and OPTIONS are two optional arguments, such as conduction coefficient functions proposed by Perona and Malik[99]. These two options are represented by the two radio button as opt1 and opt2.

To demonstrate the concept, drought severity status image of Ethiopia in September 1984 was used. Then, using this image (Figure 6.19a), if the user enters the appropriate inputs and input image, the diffusion algorithm is triggered and the output image look like the one presented in Figure 6.19b.

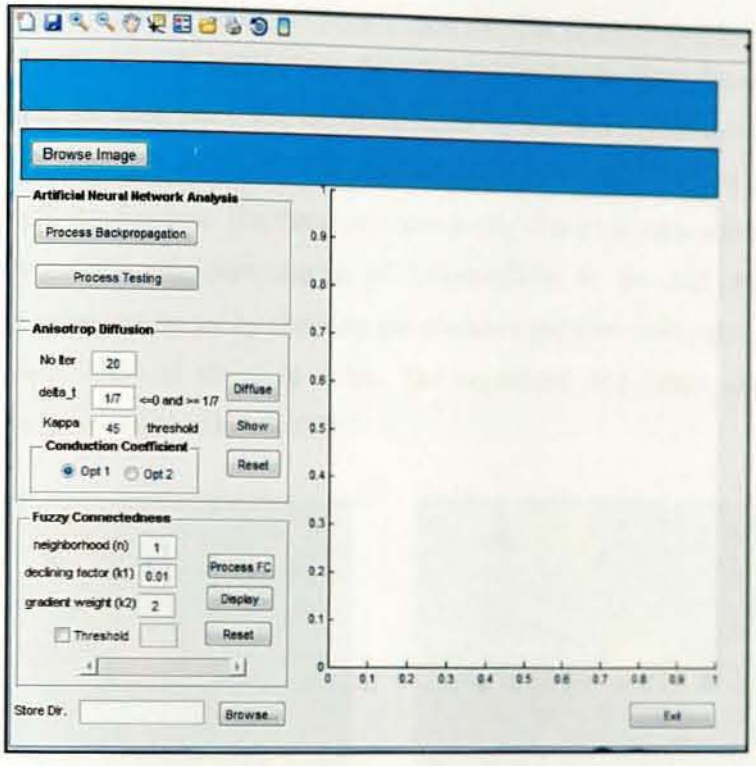


Figure 6.17: A screen shot from Drought Information Extraction System.

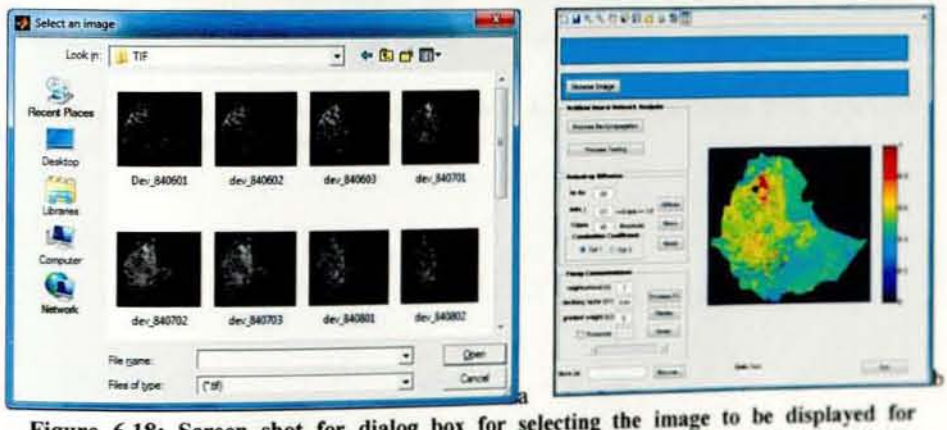


Figure 6.18: Screen shot for dialog box for selecting the image to be displayed for anisotropic diffusion: a) list of images in the directory, b) selected image displayed in the GUI of droughtObject Information Extraction System.

After anisotropic diffusion, the fuzzy segmentation can be implemented by clicking on the Process FC (fuzzy connectedness) button. This button triggers the algorithm and implementation code in Appendix 4. The default values neighborhood (n) = 4, declining factor (k_1) = 0.01, and gradient weight (k_2) = 2 can be given. The neighborhood n here is the total sample size (number of pixels involved in the neighborhood analysis). In 2D image, the value 4 represents the four adjacent pixels neighboring the seed pixel. The declining factor k_1 is a non-negative constant for non-binary adjacency of pixels. The gradient weight k_2 is a non-negative constant in fuzzy affinity function. The threshold value in this GUI is the value which shows a pre-determined minimum degree of connectedness to the seed-pixel. The threshold value can be set by checking the checkbox and then sliding the slider to the required level of threshold value. The segmented AOI image using this procedure is presented in Figure 6.20.

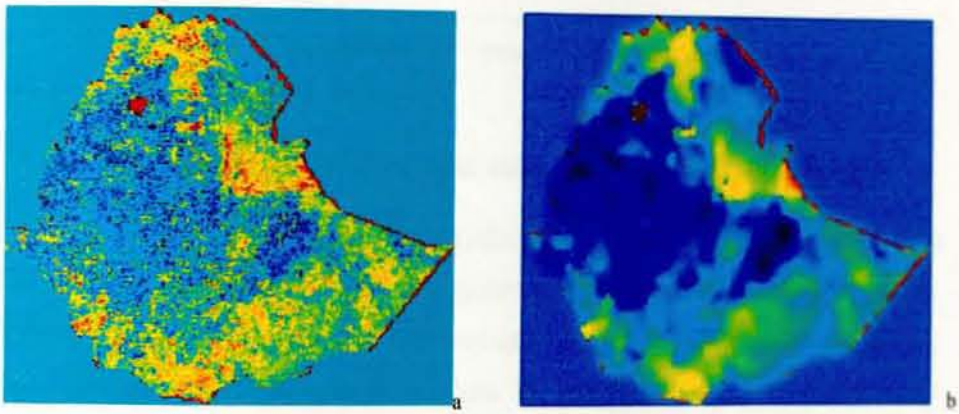


Figure 6.19: Drought severity status image of Ethiopia in September 1984: a) original image, b) anisotropic diffused image.

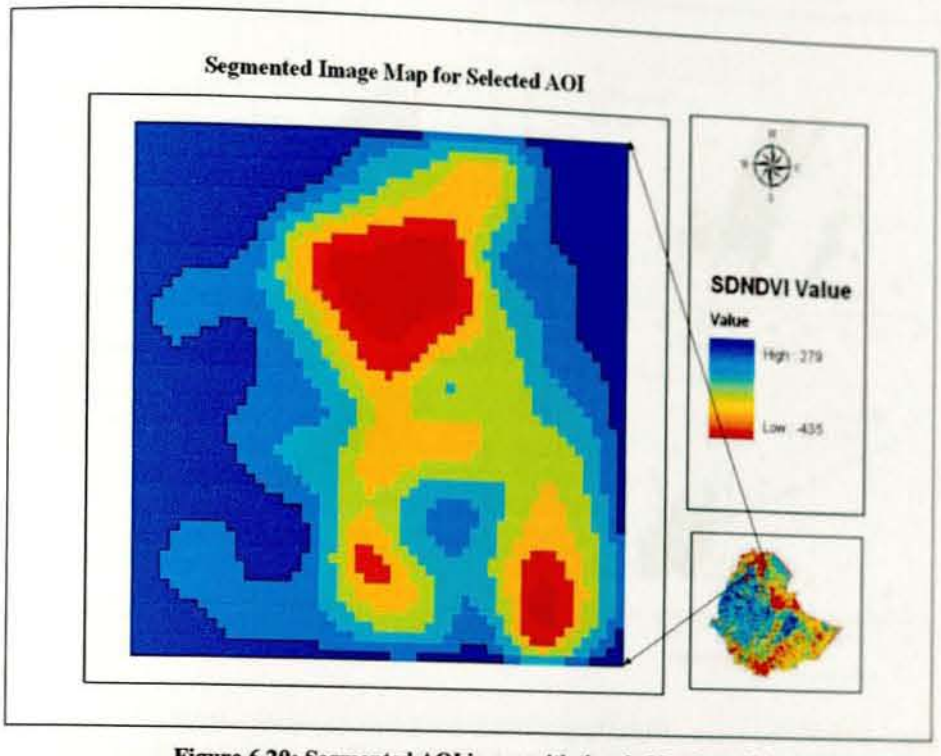


Figure 6.20: Segmented AOI image with threshold value of 0.8.

6.5 Evaluation of the Drought Information Extraction System

For evaluating Drought Information Extraction System, the 2011 drought year in Ethiopia was used as case study. The system prediction was run for the four months (July, August, September and October) using only one month prediction model) and then compared with measured values. The scatter plots are presented in Figure 6.21. The highest correlation coefficient (R^2) was obtained for September and the lowest for August month. The highest R^2 in September month was achieved because the model has highest accuracy. The lowest R^2 in August was unexpected and needs further investigations.

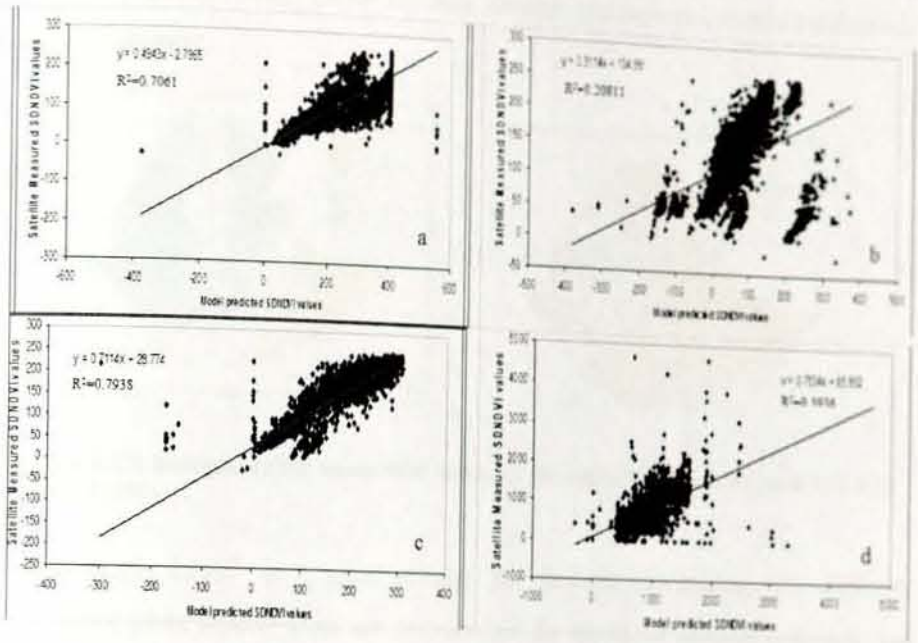


Figure 6.21: Scatter plots and correlation coefficients of satellite measured SDNDVI values versus intelligent system predicted SDNDVI values, (a) July, (b) August, (c) September, and (d) October. The x-axis is model predicted SDNDVI value and the y-axis is satellite measured SDNDVI value.

6.6 Discussions

Past efforts for drought monitoring tools include LEAP (Livelihood, Ethiopia, Assessment, Protection) software [186], WINDISP software [273], Vegetation Drought Response Index (VegDRI) model [72], and Vegetation Outlook (VegOut) model [71]. Specifically, LEAP is drought indexing software designed specifically for the local Ethiopian context, and it was initiated by the World Food Program (WFP) [186]. The software acts as a platform for the calculation of weather-based indices, starting out with the calculation of a water balance indicator (WRSI) designed by the Food and Agriculture Organization of the United Nations (FAO) and used for many years. The LEAP software uses input datasets, such as crop yield, soil, rainfall, actual evapotranspiration, NDVI and livelihood for monitoring drought [45, 186]. The livelihood baseline data is used to calculate the number of beneficiaries of food aid for each district [186]. Currently, institutions such as NMA [274], WFP VAM (Vulnerability Analysis and Mapping) [275] and

DRMFSS [181] are using LEAP for their drought assessment information delivery to their customers.

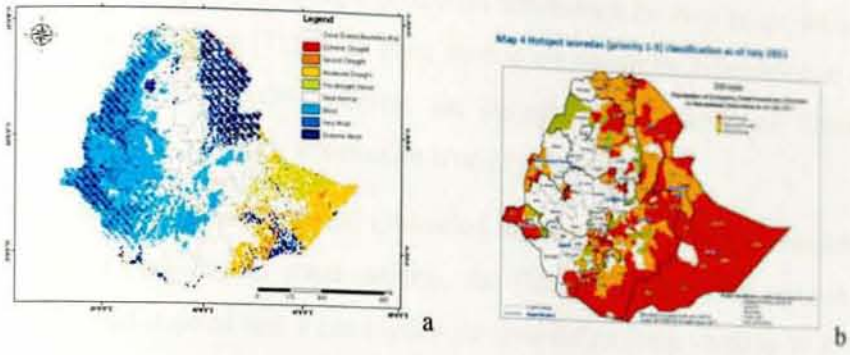


Figure 6.22: Drought extent assessment maps: a) DroughtObject product, and b) LEAP product [181].

Even though LEAP is a good startup for drought monitoring, the different parameters (data inputs) were not coordinated for assessing drought episodes. The software products can give a general guide to real-time drought monitoring and not yet evaluated using ground observation data (Personal communication at WFP in 2012).

Samples of a DroughtObject product and a LEAP software product are presented in Figure 6.22 a and b, respectively. As can be observed from these figures, the DroughtObject product precisely showed the spatial location of the drought episode with standard drought classes, and the LEAP product was showing a coarse drought situation at district level with three classes (first priority, second priority and third priority emergency relief intervention for humanitarian actions). The DroughtObject information product was produced using a one-month prediction model of June, whereas the LEAP product was produced with a real time drought assessment of July. This means that DroughtObject product is a one month-ahead prediction of drought extent. It is compelling that the DroughtObject model has advanced information delivery means compared to LEAP software in terms of temporal and spatial resolutions (Figure 6.22 a and b).

The VegDRI model integrates traditional climate-based drought indicators and satellite-derived vegetation index metrics with other biophysical information to

produce drought conditions in near-real time basis [72]. On the other hand, the VegOut model produces drought prediction information for two- to six-week time periods into the future [71]. Both these models are presenting information at a low information level (pixels), whereas the DroughtObject Information Extraction System produces at a high information level or object level.

There was no exhaustive model evaluation for VegDRI and VegOut models [71, 72, 179]. Compared to these models, the DroughtObject model product was evaluated and showed that it can explain the food deficit (crop yield) up to R^2 value of 0.91.

In the current research, DroughtObject identification from satellite image was demonstrated using simple and easy-to-use algorithms and implementation codes. The developed concepts were implemented on GUI-based presentation, which may be the base for future full system development. The developed algorithms were also tested and improved results were obtained. The output of this new concept is helpful in extracting the freely available satellite images for drought mitigation application at different levels of decision making. For this purpose, a full fledged Integrated Drought Information System Framework is recommended and presented in the following subsections.

6.7 Proposed Integrated Drought Information System Framework

The main focus of this dissertation research was on developing DroughtObject modeling using its key attributes. The mathematical modeling experiments were clearly demonstrated and evaluated. The integrated drought information system and information dissemination experiment is one of the future research agenda identified from the current experiment. For this experiment, the framework and design presented in Figure 6.23 is proposed. In this framework, the Drought Information Extraction discussed in subsection 6.2 is considered as one module and the most important component of the system. The components of this proposed system are described in the following subsections.

❖ The Near-Real-Time Environmental Data Sensor

The environmental near-real-time data source proposed in this system is Meteosat Second generation (MSG) [276]. MSG is the new European system of geostationary meteorological satellites together with the associated infrastructure.

The advanced Spinning Enhanced Visible and Infrared Imager (SEVIRI) radiometer onboard of the MSG series of geostationary satellites enables the Earth to be scanned in 12 spectral channels from visible to thermal infrared at 15 minute intervals. From these 12 channels, Drought Information System uses channel 1 and 2 for detecting drought episodes. These two visible channels are well known from similar channels of the AVHRR instrument flown on NOAA satellites and can be used in combination to generate vegetation indices (such as SDNDVI) [276].

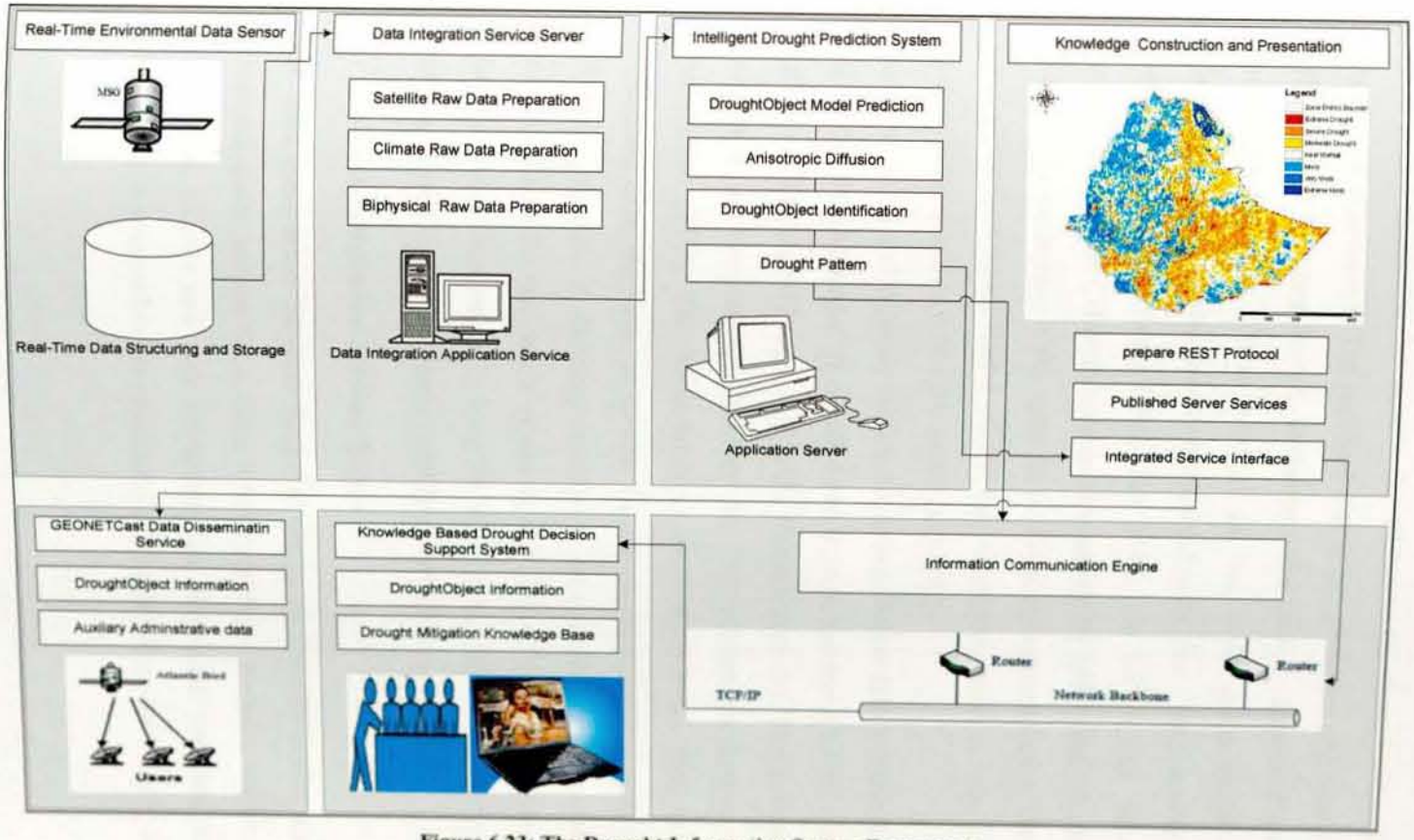


Figure 6.23: The Drought Information System Framework.

The raw data can be directly received via the EUMETSAT Multicast Distribution System (EUMETCast) and currently IT Doctoral Program has already installed fully functional data receiving station for this system. The data is being received from geostationary position of Meteosat-8, which continuously scans the Earth surface and transmits the data to the EUMETSAT Primary Ground Station in Darmstadt (Germany). The received data is pre-processed and rectified into a so called Level 1.5 data-format; furthermore the data is compressed and split into small data packages. These packages are sent to the uplink station in Usingen (Germany) and are subsequently transmitted to the Hot Bird-6 satellite (combined with some other services) [276]. For Africa, the EUTELSAT Atlantic Bird-3 satellite carries the C-Band dissemination service for MSG data. The format of the C-Band dissemination is the same as for HotBird-6 dissemination. The data is re-transmitted to Atlantic Bird-3 via the Fucino ground station in Italy. This C-Band covers the frequencies 3.70 GHz to 4.2 GHz [276].

For structuring and properly storing the near-real-time data, a software tool called the "MSG File Manager" is installed and is made functional for the data receiving station at IT Doctoral Program of Addis Ababa University. This software has been developed at International Institute for Geo-information Science and Earth Observation (ITC), The Netherlands to move the relevant data, based on the acquisition date and time, to specific folders on the processing system's attached USB-terabyte disk. A directory structure is automatically created under the main directory where the raw data is moved to, with Year – Month – Day sub-directories. The tool also creates entries in a log file if data is missing and optionally it can send warning emails if large problems (such as system failures) occur.

- **Data Integration Services Server**

The data Integration Services Server is similar to Data Processing Component, discussed under section 6.2.1. Here it has three major modules: satellite raw data preparation, climate raw data preparation and biophysical raw data preparation. This server is responsible for processing all the data from static and dynamic data

sources. The data interoperability is included in the system for integrating the different input attributes for Drought Information System.

❖ **Intelligent Drought Prediction System**

The Intelligent Drought Prediction System Component includes DroughtObject Prediction, Anisotropic Diffusion and DroughtObject Identification Modules, which were discussed under section 6.2.2, 6.2.3 and 6.4.3.1.

❖ **Knowledge Construction Component**

The main task of this component is displaying and presenting the analysis output for the system user. This functionality is also important for extracting the relevant patterns for further analysis. It has three major modules: prepare REST (representational state transfer) Protocol, publish server services, and Integrated Service Interface.

- **Prepare REST Protocol**

REST protocol is simpler to use than the well-known SOAP (Simple Object Access Protocol) approach, which requires writing or using a provided server program (to serve data) and a client program (to request data) [277]. Therefore, REST protocol is used for DroughtObject Information System for distributing the data/information to intended clients (via REST requests).

- **Publish Server Services**

The service here is DroughtObject model prediction products of one-to-four month's predictions of drought products represented as GIS resource that the central server is making available to the product users on a network. This service is published in such a way that enables the geo-processing capability, which allows clients to run a model on the server and see the results in a map service. Then, clients can consume those services over World Wide Web (WWW).

- **Integrated Services Interface**

This interface serves as an information for delivering drought data/information request from different clients. At the backend it has integrated services server (ISS), which is installed at the drought prediction engine. ISS here is designed to allow

multiple searches and data/information delivery. The end product of the search from the client side is displayed on GoogleEarth platform for locating the data/information on the real world context display.

❖ **Information Communication Engine**

The Information Communication Engine is responsible for making available and automatic uploading of published server services to client side users. There are two options for these data uploading services: 1) sending alert messages for clients, who are already registered users, and 2) uploading information for GEONETCast TelliCast Multicast Distribution System [278].

▪ **Sending Alert Message**

This option can be managed by sending alert messages to registered users, which can be government authorities or any other stakeholders' interested to get the data. This is web-based information dissemination by uploading data/information on a dedicated web server for clients to refer to it regularly. This can be supported by Knowledge Based Drought Decision Support System, which includes past efforts in managing drought episode disasters and other auxiliary administrative data for managing drought disasters. Therefore, this service is for those who are registered to get the service from Drought Information System. It also includes past Drought Mitigation Knowledge Base as a baseline information.

▪ **GEONETCast Data Dissemination Service**

The GEONETCast Data Dissemination Service (uploading data/information for GEONETCast dissemination service) is supported in coordination with the existing data delivery systems of GEONETCast. This system is already existing system, which is a near real-time global environmental information delivery system [278].

The data providers transmit to users through satellites using a multicast, access-controlled, broadband capability. This system works with the principle that data providers are sending their contributions to an uplink site from where the data is up-linked to a commercial satellite and broadcast to users. The data are then relayed to other commercial satellites to increase the footprint coverage. The conceptual data dissemination framework is presented in Figure 6.24 [278].

There are three GEONETCast Network Centers (GNCs) participating covering the whole globe: 1) EUMETSAT with EUMETCast, NOAA with GEONETCast Americas, and the Chinese Meteorological Administration (CMA) with FENGYUNCast. For this service, the African continent is being served by EUMETCast C-Band data dissemination. Currently, IT Doctoral Program has got official user name and password for getting the near-real time data from EUMETSAT data dissemination services. In the future, we negotiate with EUMETSAT for delivering DroughtObject product to be disseminated with this service. Furthermore, future research is also recommended for covering the DroughtObject Prediction System covering the whole Africa and integrates the product in EUMETCast data dissemination system.

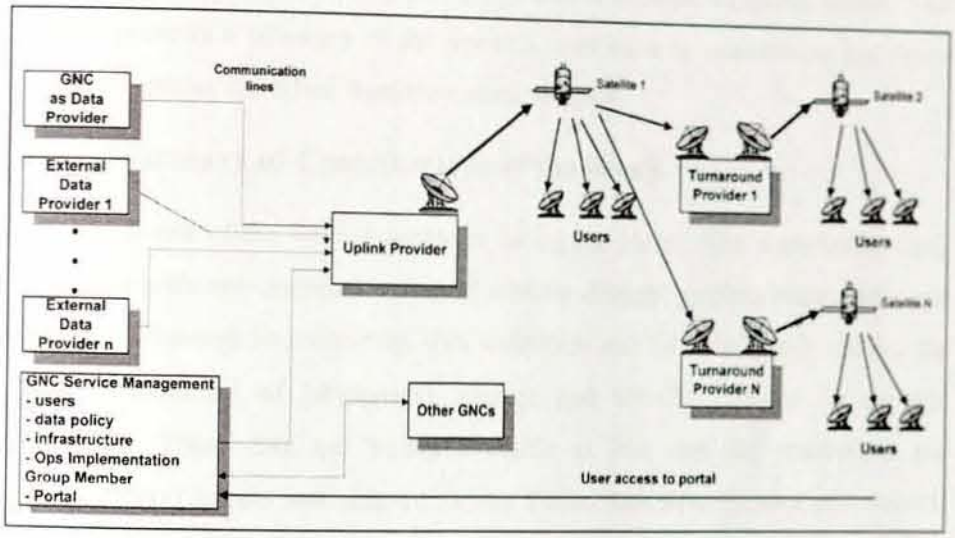


Figure 6.24: Detailed GEONETCast Network Centers framework for environmental data delivery [278].

Chapter 7 Conclusions and Future Works

This PhD research was started with the hypothesis that it is possible to identify, characterize and model drought as spatial object. To prove this hypothesis, the research problem was approached by selecting the appropriate IS research paradigm: design science research framework. Design science theory gives explicit prescriptions for constructing an artifact and mainly answers the question of how to do something, which allows the researcher to clearly analyze and find the solution for the target research problem. Accordingly, this theoretical framework was used for the identification, modeling and prediction of drought as spatial object. This chapter presents a summary of the research contributions, conclusions and future research agendas identified from the current research.

7.1 Summary of Contributions of the Work

Drought is one of the leading problems facing the planet. This dissertation thesis aims to provide new approaches to deal with the drought problem better, faster and cheaper. Advances in computing, data collection and GIS tools have enabled the recent exploration of biophysical, climate and satellite datasets for drought prediction. These data are mostly available at low cost for researchers and environmental hazard assessments. In this thesis, data from various new sources were combined, to make predictions for drought (as defined by SDNDVI) up to 4 months in advance. The developed drought prediction models were evaluated and it was confirmed that the models can be used for drought mitigation applications. These findings have implications for the benefactors of such predictions. It was demonstrated that in a relatively straight forward way, the Ethiopian government might combine data to make better predictions for drought than they have in the past. This helps decision makers at different levels to better plan for potential food insecurity and helps them to mitigate drought related risks. In addition, the results have much broader impact for farmers that can now attempt to plan based on the

current research drought predictions and for firms that hope to provide insurance for these farmers.

In the past, most of the drought identification and modeling research has been at a low information level or pixel level. In this research, drought identification, modeling and prediction were tested at a high information spatial object level (group of pixels characterizing drought spatial object). The major contributions of this research are 1) identifying the key attributes of drought spatial object for its characterization and modeling as a spatial object, 2) integrating different attributes identified from different disciplines (biophysical, climate and satellite data sources) using a KDD approach for an improved drought monitoring information system, 3) extracting relevant drought episode data/information for near real time and/or prediction, 4) developing a practical and easy-to-use model evaluation approach using crop-yield data, and 5) practically demonstrate design science problem solving approach.

Key attributes for characterization and modeling drought: A total of 11 attributes were identified from past research analysis. Each of these identified attributes was statistically tested for its relationship with drought spatial object. These attributes are the basic attributes frequently used by past research. The dataset from these attributes are both extremely large and comprehensive. This dissertation research is the first work that combines and compares the efficacy of such data for the study site's context. Other attributes can also be added to these attributes for improved modeling and prediction of DroughtObject in the future. Using these 11 attributes, we were able to predict DroughtObject one to four months' time lag with optimum accuracy.

Integrating different attributes identified from different disciplines: Currently we have huge datasets from different disciplines, which can be used for characterizing and modeling DroughtObject. The challenge is how to integrate the different datasets to convert them to useful knowledge for decision makers. In this research, the state of the art of KDD was demonstrated for data selection, pre-processing, transformation, data mining, interpretation and knowledge extraction for drought monitoring. This can help in exploiting the huge datasets available from

biophysical, climate and satellite sensors for drought monitoring information system.

Extracting relevant drought episode data/information for near real time and/or prediction: The DroughtObject prediction model developed in this research was tested and was able to produce the envisaged predictions with high accuracy. This model can be used directly by those organizations that are currently working on drought monitoring and drought risk management in Ethiopia. The drought prediction models developed showed drought episodes up to four months in advance. To our knowledge, this is the first work that reports drought episode using biophysical, climate and satellite data. This is much better than the real-time drought assessment approach using LEAP software both by DRMFSS of Ethiopia and WFP.

The DroughtObject prediction model was tested for year 1984, 2002 and 2011 Ethiopian droughts. Using this model, it was confirmed that it is possible to predict the drought episodes in advance in one- to four months in advance. This prediction capability has great impact for the decision makers who are involved in drought mitigation efforts.

Practical and easy-to-use model evaluation approach using crop-yield data: The DroughtObject model developed was evaluated with an independent dataset for its applicability in "fitness for purpose" context. The ultimate purpose of drought monitoring is to assure food security of a given district. Therefore, the drought monitoring tool has to take this reality into account. Food security is also in one way or another related to the amount of crop yield produced in a given district. In the past, there was no clearly defined drought monitoring model evaluation approach. This research developed an easy-to-use model evaluation approach and tested the DroughtObject model. It was found that the model can reliably be used for monitoring both drought severity status and also yield deficit of a district during time of drought with reliable accuracy. This is the first work that links drought episode with crop yield deficit in the Ethiopian context.

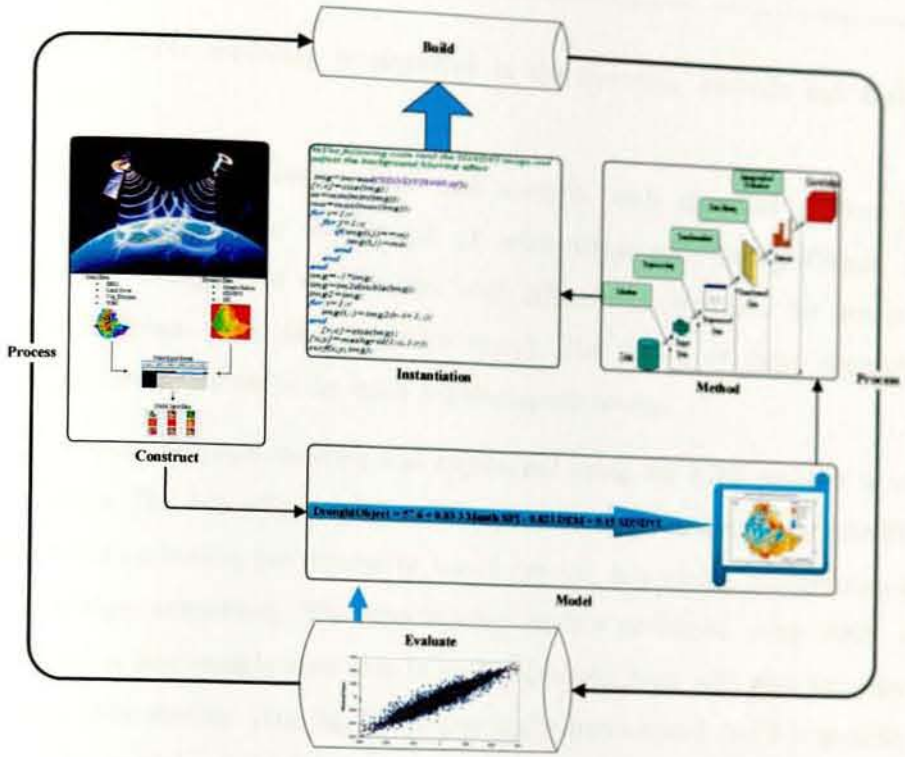


Figure 7.1: Design science problem solving process in the current research.

7.2 Conclusions

In this research, we investigated the experimental and practical application of DroughtObject identification, modeling and prediction. The three research questions posed in this work were: 1) what are the most relevant attributes for drought monitoring, 2) how is drought modeled as a spatial object for improved identification and prediction, and 3) how is DroughtObject identified on satellite imagery evaluated using ground observation data? All of these questions were addressed and interesting results were found.

For the first research question, detailed research reviews were conducted and a total of 11 key attributes were identified for the identification, modeling and prediction of DroughtObject in space-time dimensions. These attributes are DEM, SDNDVI, three-month SPI, ecosystem (veg-Ethiopia), land cover, WHC, PDO, AMO, NAO, PNA and MEI. The details of each attribute and why they were selected for

DroughtObject modeling is presented in the materials, methods and analysis chapter.

After a detailed research review and analysis, each identified attribute was statistically tested for its strength of relationships with DroughtObject. The statistical strengths of relationships with different combinations for predicting DroughtObject were also iteratively tested. The details of these exploratory analyses are presented in the result and discussions section.

The second research question was approached using the KDD problem solving approach. The data obtained from the key attributes were passed through the KDD steps (data selection, pre-processing, transformation, data mining, interpretation and knowledge extraction). The data mining models developed using ANN and regression tree models were able to predict DroughtObject with high accuracy in one to four months' time lag. It was practically demonstrated that it is possible to identify, model and predict DroughtObject using its key attributes in a space-time dimension (Figure 7.2a and b). This helps improve drought risk management and provides decision makers with information before drought can affect human life and the environment.

The third research question was approached by developing a practical model evaluation method using ground observed crop yield data. The DroughtObject model that was developed was evaluated for its application in showing the drought severity and food deficit status of a given district (zone administrative level). Interesting evaluation results were found and it was confirmed that the model can be used directly by those who are responsible for drought monitoring and risk management.

7.3 Future Works

In light of climate change and uncertainties in future weather conditions, research that comes up with simple and accurate drought monitoring methods is increasingly getting attention. A number of studies have also characterized drought for its improved modeling and prediction at low information level. Our current research has demonstrated an object-oriented identification, modeling and prediction of drought at high information level. In this experiment, a number of future research agendas were identified for the improved modeling and management of drought episodes.

The research output currently can be used directly by the decision makers at different levels in Ethiopia. How to reach these decision makers, and the appropriate media for the dissemination of the model output, is the immediate research agenda to be addressed.

Currently, huge datasets are available from SPOT satellite and high temporal resolution from Meteosat satellite (which can be updated every 15 minutes). Data filtering and automating the process for model input is required for using these datasets. The required level of temporal resolution of the satellite products for drought monitoring has to be determined for improved resource utilization. These all require high to medium level data management research and infrastructure for efficient implementation of the current research output.

The current research was conducted using data with spatial resolution of 8km. This spatial resolution is relatively low for characterizing heterogeneous environment. The errors in the models are also attributed to this low spatial resolution of the data inputs and the high heterogeneity of the ecosystems of Ethiopia. From our practical experience, within a short range of distance, the elevation, vegetation type, and soil conditions vary greatly. Future research can use the method developed here with high spatial resolution for improved model accuracy and assess the amount of information gain by using the high spatial resolution data.

The model evaluation results also showed optimum accuracy (R^2 values) ranging from 0.01 to 0.91. This difference in the accuracy of the model is again due to low administrative resolution (zonal administrative level) crop yield data used for the model evaluation. Evaluating the model developed here using high resolution administrative data (woreda or kebele administrative levels) helps to improve the accuracy, which further improves drought mitigation measures.

Future research may also assess the efficiency, effectiveness and applicability of DroughtObject Information Extraction System compared to LEAP, VegDRI and VegOut for identifying and modeling drought episodes. Efficiency here is same as Rothermel and Harrold's [280] definition of the measurement of the computational cost, which may also determine the practicality of a technique. Effectiveness can also be measured by comparing DroughtObject Information Extractor with LEAP, VegDRI and VegOut for identifying and modeling drought episodes. This future research can be practical after the DroughtObject Information Extraction System is deployed as a working system for organizations who are currently working on drought monitoring.

In this thesis, the basic groundwork has been laid for better drought prediction in general. In the Ethiopian case, the approach is tested using a large dataset that were compiled from different sources. Organizations and future researchers can use data from biophysical, climate and satellite sources that are valuable for predicting drought at a relatively low cost. The potential impact is already great in that these data and methods used can be used almost immediately to help better predict drought. In an immediate future work, it is recommended to extend and refine the prediction models to incorporate spatial attributes from other biophysical, climate and satellite data, and appropriate spatial statistical methods towards improving predictive performance and the time at which predictions can be made reliably. In addition, it is also recommended to design a decision making tool with a GUI to help decision makers to make better decisions, especially for offering drought related insurance to Ethiopian farmers.

References

[1] S. Liao, "Knowledge management technologies and applications-literature review from 1995 to 2002," *Expert Systems with Applications*, vol. 25, pp. 155 - 164, 2003.

[2] J. F. Courtney, "Decision making and knowledge management in inquiring organizations: toward a new decision-making paradigm for DSS," *Decision Support Systems*, vol. 31, pp. 17 - 38, 2001.

[3] I. Nonaka, "A Dynamic Theory of Organizational Knowledge Creation," *Organizational Science*, vol. 5, pp. 14 - 37, 1994.

[4] C. Hor, "Extracting Knowledge From Substations for Decision Support," *IEEE TRANSACTIONS ON POWER DELIVERY*, vol. 20, pp. 595 - 602, 2005.

[5] Z. Dienes and J. Perner, "A theory of implicit and explicit knowledge," *BEHAVIORAL AND BRAIN SCIENCES*, vol. 22, pp. 735–808, 1999.

[6] H. Han and M. Kamber, *Data mining: concepts and techniques*, 2 ed. San Francisco: Morgan Kaufmann Publishers, an imprint of Elsevier, 2006.

[7] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," *Communication of the ACM*, vol. 39, 1996.

[8] J. Jackson, "DATA MINING: A CONCEPTUAL OVERVIEW," *Communications of the Association for Information Systems*, vol. 8, pp. 267 - 296, 2002.

[9] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "Knowledge Discovery and Data Mining: Towards a Unifying Framework," *KDD-96 Proceedings. AAAI (www.aaai.org)*, pp. 82 - 88, 1996.

[10] H. J. Miller and J. Han, *Geographic Data Mining and Knowledge Discovery*: Taylor & Francis, 2001.

[11] J. Han, M. Kamber, and A. K. H. Tung, "Spatial clustering methods in data mining: A survey," in *Geographic Data Mining and Knowledge Discovery*, H. J. Miller and J. Han, Eds. London: Taylor and Francis, 2001, pp. 188-217.

[12] F. Qi and A.X. Zhu, "Knowledge discovery from soil maps using inductive learning," *International Journal of Geographical Information Science*, vol. 175, pp. 771 - 79, 2003.

[13] D. Guo, M. Gahegan, A. M. MacEachren, and B. Zhou, "Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach," *Cartographic and Geographic Information Science*, vol. 32, pp. 113–132, 2005.

[14] D. Guo, M. Gahegan, D. Peuquet, and A. MacEachren, "Breaking down dimensionality: An effective feature selection method for high-dimensional clustering," in *Clustering High Dimensional Data and its Applications. Presented at the Workshop on 2003 the Third SIAM International Conference on Data Mining*. San Francisco, California, USA, 2003.

[15] S. S. Anand, D. A. Bell, and J. G. Hughes, "EDM: a general framework for data mining based on evidence theory," *Data and Knowledge Engineering*, vol. 18, pp. 189 - 223, 1996.

[16] M. Cannataro, D. Talia, and P. Trunfio, "Distributed data mining on the grid," *Future Generation Computer Systems*, vol. 18, pp. 1101 - 1112, 2002.

[17] S. Y. Chen and X. Liu, "The contribution of data mining to information science," *Journal of Information Science*, vol. 3, pp. 550-558, 2004.

[18] Z. Chen and Q. Zhu, "Query construction for user-guided knowledge discovery in database," *Journal of Information Sciences*, vol. 109, pp. 49 - 64, 1998.

[19] C. M. Chiu, "Towards integrating hypermedia and information systems on the web," *Information and Management*, vol. 40, pp. 165–175, 2003.

[20] A. Dai, "Drought under global warming: a review," *Advanced Review, National Center for Atmospheric Research*, vol. 2, pp. 45 - 65, 2011.

[21] T. Tadesse, M. Haile, G. Senay, D. Brian, W. Cody, and L. Knutson, "The Need for Integration of Drought Monitoring Tools for Proactive Food Security Management in Sub-Saharan Africa," *Natural Resources Forum*, vol. 32, pp. 265-279, 2008.

[22] UNEP, *United Nations Environment Programme, Climate Change and Variability in the Sahel Region: Impacts and Adaptation Strategies in the Agricultural Sector*, 2006.

[23] M. F. Goodchild, "Geographical Data Modeling," *Computer & Geosciences*, vol. 18, pp. 401 - 408, 1992.

[24] A. Stein, "Modern developments in image mining," *Science in China Series E. Technological Science*, vol. 51, pp. 13 -25, 2008.

[25] A. Stein, N. A. S. Hamm, Ye, and Qinghua, "Handling uncertainties in image mining for remote sensing studies," *International Journal of Remote Sensing*, vol. 30, pp. 5365 - 5382, 2009.

[26] C. Rulinda, "Mining drought from remote sensing images," in *Geoinformatics*. Enschede, The Netherlands: International Institute for Geo-Information Science and Earth Observation (ITC), 2007, pp. 68.

[27] C. M. Rulinda, W. Bijker, and A. Stein, "Image mining for drought monitoring in eastern Africa using Meteosat SEVIRI data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 12S, pp. 63-68, 2010.

[28] T. B. McKee, N. J. Doesken, and J. Kleist, "The relationship of drought frequency and duration to time scales," *Paper presented at the 8th Conference on Applied Climatology, 17 - 22 January*, pp. 179-184, 1993.

[29] T. B. McKee, N. J. Doesken, and J. Kleist, "Drought Monitoring with Multiple Time Scales," *Preprints, 9th Conference on Applied Climatology, January 15–20, Dallas, Texas*, pp. 233-236, 1995.

[30] S. Nicholson, "Recent rainfall fluctuations in Africa and their relationship to past conditions over the continent," *The Holocene*, vol. 4, pp. 121-134, 1994.

[31] S. Swenson and J. Wahr, "Monitoring the water balance of lake Victoria, east Africa, from space," *Journal of Hydrology*, vol. 370, pp. 123-176, 2009.

[32] FEWSNET, "United Nations' FEWSNET (Famine Early Warning Systems Network)," vol. 2011: Famine Early Warning Systems Network, 2009.

[33] A. Giannini, R. Saravanan, and P. Chang, "Oceanic forcing of Sahel rainfall on interannual to interdecadal time scales," *Science*, vol. 302, pp. 1027–1030, 2003.

[34] S. M. Hagos and K. H. Cook, " Ocean warming and late twientieth-century Sahel drought and recovery," *Journal of Climate*, vol. 21, pp. 3797–3814, 2008.

[35] NOAA, "JetStream - Online School for Weather, Inter-Tropical Convergence Zone," vol. 2012: National Weather Service, 2007.

[36] A. Giannini, M. Biasutti, and M. M. Verstraete, " A climate model-based review of drought in the Sahel: desertification, the re-greening and climate change," *Glob Planetary Change*, vol. 64, pp. 119–128, 2008.

[37] N. Zeng, J. D. Neelin, K. M. Lau, and C. J. Tucker, "Enhancement of interdecadal climate variability in the Sahel by vegetation interaction," *Science*, vol. 286, pp. 1537 - 1540, 1999.

[38] 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.

[39] 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.

[40] J. Hayes, P. O'Rourke, W. Terjung, and P. Todhunter, "A feasible crop yield model for world wide international food production," *International Journal of Biometeorology*, vol. 26, pp. 239 - 257, 1982.

[41] R. Benedetti and Rossinni, "On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in Emilia Romagna," *Remote Sensing of the Environment*, vol. 45, pp. 311 - 326, 1993.

[42] N. Quarmby, M. Milnes, T. Hindle, and N. Silcos, "The use of multi temporal NDVI measurements from AVHRR data for crop yield estimation and prediction," *International Journal of Remote Sensing*, vol. 14, pp. 199 - 210, 1993.

[43] L. Ungani and F. Kogan, "Drought monitoring and corn yield estimation in southern Africa from AVHRR DATA," *Remote Sensing of the Environment*, vol. 63, pp. 219 -232, 1998.

[44] F. Kogan, A. Gitelson, E. Zakarin, L. Spivak, and L. Lebed, "AVHRR-based spectral vegetation index for quantitative assessment of vegetation state and productivity: calibration and validation," *Photogrammetric Engineering and Remote Sensing*, vol. 69, pp. 899 - 906, 2003.

[45] FEWSNET, "Famine Early Warning Systems Network," vol. 2011: FEWSNET, 2011.

[46] G. Senay and J. Verdin, "Characterization of yield reduction in Ethiopia using a GIS-based crop water balance model.," *Can. J. Remote Sens.*, vol. 29, pp. 687–692, 2003.

[47] T. Dinku, P. Ceccato, E. Grover-Kopec, M. Lemma, S. J. Connor, and C. F. Ropelewski, "Validation of satellite rainfall products over East Africa's complex topography," *International Journal of Remote Sensing*, vol. 28, pp. 1503–1526, 2007.

[48] 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.

[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] B. Getachew, T. Tsegaye, A. Solomon, H. Shawndra, and T. Yitaktu, "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.

[51] B. Getachew, H. Shawndra, T. Tsegaye, and A. Solomon, "Using Satellite Images for Drought Monitoring: A Knowledge Discovery Approach," *Journal of Strategic Innovation and Sustainability*, vol. 7, pp. 135 - 153, 2011.

[52] B. Getachew, T. Tsegaye, A. Solomon, and H. Shawndra, "Drought Monitoring in-Food-Insecure Areas of Ethiopia by Using Satellite Technologies," in *Experience of Climate Change Adaptation in Africa, Climate Change Management, DOI: 10.1007/978-3-642-22315-0_11*, W. L. Filho, Ed., 2010, pp. 1145 - 1162.

[53] B. Getachew, A. Solomon, and H. Anat, "Knowledge Discovery from Satellite Images for Drought Monitoring in Food Insecure Areas," in *Americas Conference on Information Systems (AMCIS) 2010 Conference Proceeding*: AMCIS, 2010.

[54] B. Getachew, T. Tsegaye, A. Solomon, and H. Shawndra, "Drought Object Modeling for Climate Change Mitigation: The Case of Ethiopia," in *Building Software in the Developing World, Proceedings of the 1st EICTF October 27, 2011*. Addis Ababa: Ethiopian ICT Forum, 2011, pp. 59.

[55] D. Wilhite, *Drought and water crisis: science, technology and management issues*: Taylor & Francis Group, 2005.

[56] A. G. Dai, K. E. Trenberth, and T. T. Qian, "A global dataset of Palmer Drought Severity Index for 1870-2002: Relationship with soil moisture and effects of surface warming," *Journal of Hydrometeorology*, vol. 5, pp. 1117-1130, 2004.

[57] U. S. Panu and T. C. Sharma, "Challenges in drought research: some perspectives and future directions," *Hydrological Sciences-Journal—des Sciences Hydrologiques*, vol. 47(S), pp. S19-S30, 2002.

[58] M. Hayes, O. V. Wilhelmi, and C. L. Knutson, "Reducing Drought Risk: Bridging Theory and Practice," *Natural Hazards Review*, vol. 5, pp. 106 - 113, 2004.

[59] J. Verdin, C. Funk, G. Senay, and R. Choularton, "Climate science and famine early warning," *Philosophical Transactions*, vol. 360, pp. 2155–2168, 2005.

[60] E. Meze-Hausken, " Contrasting climate variability and meteorological drought with perceived drought and climate change in northern Ethiopia," *Climate Research*, vol. 27, pp. 19-31, 2004.

[61] U. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen, "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," *Journal of Photogrammetry & Remote Sensing*, vol. 58, pp. 239-258, 2004.

[62] A. K. Mishra and V. R. Desai, "Drought forecasting using feed-forward recursive neural network," *ecological modelling*, vol. 198, pp. 127 - 138, 2006.

[63] G. O. P. Obasi, "WMO's role in the international decade for natural disaster reduction," *Bulletin of American Meteorological Society*, vol. 75, pp. 655–1661, 1994.

[64] USDA, *Major world crop areas and climatic profiles, World Agricultural Outlook Board*: United States Department of Agriculture (USDA), Agricultural Handbook, 1994.

[65] C. Funk and J. Verdin, "Real-Time Decision Support Systems: The Famine Early Warning System Network," *Hydrology*, vol. In Press, 2009.

[66] NMSA, *(National Meteorological Services Agency) Assessment of drought in Ethiopia, Meteorological Research Reports Series, No. 2*. Addis Ababa, Ethiopia, 1996.

[67] S. K. Harms, J. Deogun, and T. Tadesse, "Discovering sequential association rules with constraints and time lags in multiple sequences," in *Proc. the 13th International Symposium on Methodologies for Intelligent Systems, Lyon, France, June 27-29*, vol. 2366, M. S. Hacid, Z. W. Ras, D. A. Zighed, and Y. Kodratoff, Eds. Berlin: Foundations of Intelligent Systems, Lecture Notes in Artificial Intelligence, Springer, 2002, pp. 432- 441.

[68] S. Harms, T. Tadesse, and B. Wardlow, "Algorithm and Feature Selection for VegOut: A Vegetation Condition Prediction Tool," J. Gama, Ed.: Springer-Verlag Berlin Heidelberg, 2009.

[69] T. Tadesse, D. Wilhite, S. Harms, M. Hayes, and S. Goddard, "Drought Monitoring Using Data Mining Techniques: A Case Study for Nebraska, USA," *Natural Hazards*, vol. 33, pp. 137–159, 2004.

[70] T. Tadesse, J. Brown, and M. Hayes, "A new approach for predicting drought-related vegetation stress: Integrating satellite, climate, and biophysical data over

the U.S. central plains," *ISPRS Journal of Photogrammetry & Remote Sensing*, vol. 59, pp. 244– 253, 2005.

[71] T. Tadesse, B. D. Wardlow, M. J. Hayes, and M. D. Svoboda, "The Vegetation Outlook (VegOut): A New Method for Predicting Vegetation Seasonal Greenness," *GIScience and Remote Sensing*, vol. 47, pp. 25 - 52, 2010.

[72] 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.

[73] A. Hevner, S. T. March, J. Park, and S. Ram, "Design Science in Information Systems Research," *MIS Quarterly*, vol. 1, pp. 75-105, 2004.

[74] S. Gregor, "The Nature of Theory in Information Systems," *MIS Quarterly*, vol. 3, pp. 611-642, 2006.

[75] S. T. March and V. C. Storey, "Design Science in the Information Systems Discipline: An Introduction to the Special Issue on Design Science Research," *MIS Quarterly*, vol. 32, pp. 725-730, 2008.

[76] A. Jakeman, R. Letcher, and J. Norton, "Ten iterative steps in development and evaluation of environmental models," *Environmental Modelling Software*, vol. 21, pp. 602-614, 2006.

[77] D. A. Wilhite, "Drought as a natural hazard: Concepts and definitions," in *Drought: A global assessment, Natural Hazard and Disasters Series*, D. Wilhite, Ed. UK: Routledge Publishers, 2000, pp. 3--18.

[78] M. Gahegan, M. Wachowicz, M. Harrower, and T. M. Rhyne, "The integration of geographic visualization with knowledge discovery in databases and geocomputation," *Cartography and Geographic Information Systems*, vol. 28, pp. 29 - 44, 2001.

[79] S. Shekhar, C. T. Lu, and P. Zhang, "A unified approach to detecting spatial outliers," *GeoInformatica*, vol. 7, pp. 139 - 166, 2003.

[80] M. D. Myers, *Qualitative Research in Business and Management*. London: Sage Publications, 2009.

[81] W. Zerihun, *Methods and Strategies for the Conservation of Forest Genetic Resources. In: Sue Edwards, Abebe Demissie, Taye Bekele and Gunther Haas (Eds): Proceedings of Forest Genetic Resources Conservation: Principles, Strategies and Actions*. Addis Ababa: Support for Biodiversity Institute, German Technical Cooperation (GTZ), 1999.

[82] M. L. Markus, A. Majchrzak, and L. Gasser, "A Design Theory for Systems that Support Emergent Knowledge Processes," *MIS Quarterly*, vol. 26, pp. 179 - 212, 2002.

[83] J. G. Walls, G. R. Widmeyer, and O. A. E. Sawy, "Building an Information System Design Theory for Vigilant EIS," *Information Systems Research*, vol. 3, pp. 36 - 59, 1992.

[84] J. Iivari, "A Paradigmatic Analysis of Information Systems as a Design Science," *Scandinavian Journal of Information Systems*, vol. 19, pp. 39 - 64, 2007.

[85] S. T. March and G. Smith, "Design and Natural Science Research on Information Technology," *Decision Support Systems*, vol. 15, pp. 251 - 266, 1995.

[86] H. A. Simon, *The Sciences of the Artificial*, 3 ed. Cambridge, MA: MIT Press, 1996.

[87] P. J. Denning, "A New Social Contract for Research," *Communications of the ACM*, vol. 40, pp. 132 - 134, 1997.

[88] N. J. Rosenberg, "North American Droughts," in *Westview Press, Boulder, CO.*, 1978, pp. 177.

[89] UNCCD, *United Nations Convention to Combat Desertification, Article 1*. Bonn, Germany: United Nations, 1999.

[90] V. Smakhtin and D. Hughes, *Review, Automated estimation and analysis of drought indices in south Asia, W.P. 83*: Sri Lanka, International Water Management Institute, 2004.

[91] UNISDR, *United Nations Secretariat of the International Strategy for Disaster Reduction, Drought Risk Reduction Framework and Practices: Contributing to the Implementation of the Hyogo Framework for Action*. Geneva, Switzerland., 2009.

[92] D. W. Stahle, F. K. Fye, E. R. Cook, and R.D. Griffin, "Tree-ring reconstructedmegadroughts overNorth America since AD 1300," *Climate Change*, vol. 83, pp. 133–149, 2007.

[93] T. Budd, *Understanding object-oriented programming with Java*: Reading, Mass: Addison Wesley, 2000.

[94] M. Worboys, M. Hilary, M. Hearnshaw, and D. Maguire, "Object-Oriented Data Modelling for Spatial Databases," *International Journal of Geographical Information Systems*, vol. 4, pp. 369-383, 1990.

[95] R. A. Deby, R. A. Knippers, Y. Sun, M. C. Ellis, M. J. Kraak, M. J. Weir, Y. Georgiadou, M. M. Radwan, C. J. Westen, W. Kainz, and E. J. Sides, *Principles of Geographic Information Systems*. Enschede, The Netherlands: ITC, 2001.

[96] C. Tucker, "Red and photographic infrared linear combinations for monitoring vegetations," *Remote Sensing of Environment*, vol. 8, pp. 127-150, 1979.

[97] J. R. Jensen and D. Cowen, C., "Remote Sensing of Urban/Suburban Infrastructure and Socio-Economic Attributes," *Photogrammetric Engineering & Remote Sensing*, vol. 65, pp. 611-622, 1997.

[98] D. Cowen, C.J. Jensen, R.G. Bresnahan, D. Ehler, D. Traves, X. Huang, C. Weisner, and H. E. Mackey, "The Design and Implementation of an Integrated GIS for Environmental Applications," *Photogrammetric Engineering & Remote Sensing*, vol. 61, pp. 1393-1404, 1995.

[99] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transaction Pattern Analysis and Machine Intelligence*, vol. 12, pp. 629–639, 1990.

[100] G. Sapiro, *Geometric partial differential equations and image analysis*: Cambridge University Press, 2001.

[101] M. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust Anisotropic Diffusion," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 7, pp. 421-432, 1998.

[102] Y. Yu and S. Acton, "Speckle Reducing Anisotropic Diffusion," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 11, pp. 1260-1270, 2002.

[103] J. Schiewe, *Segmentation of high resolution remotely sensed data concepts, applications and problems*. Ottawa: Symposium on Geospatial Theory, Processing and Applications, 2002.

[104] N. Thomas, C. Hendrix, and R. G. Congalton, "A comparison of urban mapping methods using high-resolution digital imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 69, pp. 963- 972, 2003.

[105] S. E. Franklin, R. J. Hall, L. M. Moskal, A. J. Maudie, and M. B. Lavigne, "Incorporating texture into classification of forest species composition from airborne multispectral images," *International Journal of Remote Sensing*, vol. 21, pp. 61- 79, 2000.

[106] S. Ryherd and C. E. Woodcock, "Combining spectral and texture data in the segmentation of remotely sensed images," *Photogrammetric Engineering and Remote Sensing*, vol. 62, pp. 181- 194, 1996.

[107] A. Laliberte, A. Rango, K. Havstad, J. Parish, R. Beck, R. McNeely, and A. Gonzalez, "Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico," *Remote Sensing of Environment*, vol. 93, pp. 198- 210, 2004.

[108] R. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, pp. 610- 621, 1973.

[109] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Computer Vision, Graphics and Image Processing*, vol. 29, pp. 100- 132, 1985.

[110] T. Blaschke and J. Strobl, "What's wrong with pixels? Some recent developments interfacing remote sensing and GIS," *GeoBIT/GIS*, vol. 6, pp. 12 - 17, 2001.

[111] T. R. Reed and H. Wechsler, "Segmentation of textured images and gestalt organization using spatial/spatial frequency representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 1 - 12, 1990.

[112] J. F. Haddon and J. F. Boyce, "Image segmentation by unifying region and boundary information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 929-948, 1990.

[113] A. A. Abkar, M. Sharifi, and N. Mulder, "Likelihood-based image segmentation and classification: A framework for the integration of expert knowledge in image classification procedures," *International Journal of Applied Earth Observation and Geoinformation*, vol. 2, pp. 101- 119, 2000.

[114] M. Baatz and A. Schaepe, *Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation*, vol. XII. Germany, Wichmann: Heidelberg, 2000.

[115] C. Burnett and T. Blaschke, "A multi-scale segmentation/object relationship modelling methodology for landscape analysis," *Ecological Modelling*, vol. 168, pp. 233-249, 2003.

[116] G. J. Hay, P. Dube, A. Bouchard, and D. J. Marceau, "A scale-space primer for exploring and quantifying complex landscapes," *Ecological Modelling*, vol. 153, pp. 27- 49, 2002.

[117] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 3rd Edition ed. United States of America: Thomson, 2006.

[118] J. K. Udupa and S. Samarasekera, "Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation," *Graphical Models and Image Processing*, vol. 58, pp. 246-261, 1996.

[119] A. Rosenfeld, "Fuzzy Digital-Topology," *Information and Control*, vol. 40, 1979.

[120] A. Rosenfeld, "The fuzzy geometry of image subsets," *Pattern Recognition Letters*, vol. 2, pp. 1-317, 1984.

[121] I. Bloch, "Fuzzy connectivity and mathematical morphology," *Pattern Recognition Letters*, vol. 14, pp. 483-488, 1993.

[122] S. Dellepiane and F. Fontana, "Extraction of intensity connectedness for image processing," *Pattern Recognition Letters*, vol. 16, pp. 313-324, 1999.

[123] J. Udupa, J. Tian, D. Hemmy, and P. Tessier, "A Pentium PC Based Craniofacial 3D Imaging and Analysis System," *Journal of Craniofacial Surgery*, vol. 8, pp. 333-339, 1997.

[124] J. Udupa, L. Wei, S. Samarasekera, Y. Miki, M. van Buchem, and R. Grossman, "Multiple Sclerosis Lesion Quantification Using Fuzzy Connectedness Principles," *IEEE Transaction Medical Imaging*, vol. 16, pp. 598-609, 1997.

[125] J. K. Udupa, K. Punam, and Saha, "Relative Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation," *IEEE*

TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol. 24, pp. 1485-1500, 2002.

[126] S. Samarasekera, J. K. Udupa, Y. Miki, and R. Grossman, "A New Computer-Assisted Method for Enhancing Lesion Quantification in Multiple Sclerosis," *Journal of Computer Assisted Tomography*, vol. 21, pp. 145-151, 1997.

[127] Y. Miki, R. Grossman, J. Udupa, M. van Buchem, L. Wei, M. Philips, U. Patel, J. McGown, and D. Kolson, "Differences Between Relapsing Remitting and Chronic Progressive Multiple Sclerosis as Determined with Quantitative MR Imaging," *Radiology*, vol. 210, pp. 769-774, 1999.

[128] R. Brooks, "Intelligence without representation," *Artificial Intelligence*, vol. 47, pp. 139-159, 1991.

[129] R. Stair and G. Reynolds, *Principles of Information Systems*. United States of America: Course Technology, Cengage Learning, 2010.

[130] R. E. Plant and N. D. Stone, *Knowledge-based Systems in Agriculture*. New York: McGraw-Hill, 1991.

[131] L. Fu, *Neural Networks in Computer Intelligence*. New York: McGraw-Hill, 1995.

[132] I. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of Microbiological Methods*, vol. 43, pp. 3–31, 2000.

[133] D. Z. Sui, "An initial investigation of integrating neural networks with GIS for spatial decision making," *In: Proceedings GIS/LIS 92*: San Jose, CA, 1992.

[134] M. Nelson and W. T. Illingworth, "A Practical Guide To Neural regression models," *In: Proceedings of Intelligent Engineering Nets*. Reading, MA: Addison-Wesley, 1990.

[135] A. Kandel and G. Langholz, *Hybrid Architectures For Intelligent Systems*. Boca Raton, FL.: CRC Press, 1992.

[136] D. E. Rumelhart, R. Durbin, R. Golden, and Y. Chauvin, "Backpropagation: the basic theory," *In: Backpropagation: Theory, Architecture, and Applications*, D. E. Rumelhart and C. Yves, Eds.: Lawrence Erlbaum, NJ, 1995, pp. 1–34.

[137] Rulequest, "An Overview of Cubist," vol. 2011, 2011. [Online]. Available: http://www.rulequest.com/cubist-win.html.

[138] L. Yang, B. K. Wylie, L. L. Tieszen, and B. C. Reed, "An analysis of relationships among climate forcing and time integrated NDVI of grasslands over the U.S. northern and central Great Plains," *Remote Sensing of the Environment*, vol. 65, pp. 25–37, 1998.

[139] 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.

[140] G. B. Ephrem and B. Meissner, "Spatio-temporal analyses of correlation between NOAA satellite RFE and weather stations' rainfall record in Ethiopia," *International Journal of Applied Earth Observation and Geoinformation*, vol. 12S, pp. S69–S75, 2010.

[141] B. Narasimhan and R. Srinivasan, "Development and evaluation of soil moisture deficit index (SMDI) and evapotranspiration deficit index (ETDI) for agricultural drought monitoring," *Agricultural and Forest Meteorology*, vol. 133, pp. 69-88, 2005.

[142] K. Henry, H. Ntale, and T. Y. Gan, "DROUGHT INDICES AND THEIR APPLICATION TO EAST AFRICA," *INTERNATIONAL JOURNAL OF CLIMATOLOGY*, vol. 23, pp. 1335–1357, 2003.

[143] W. C. Palmer, "Meteorological drought. U.S. Department of Commerce, Weather Bureau, Research Paper, vol. 45," 1965.

[144] N. Guttman, B., "Comparing the Palmer drought index and the standardized precipitation index," *JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION*, vol. 34, pp. 113 - 121, 1998.

[145] H. Ntale and T. Gan, "Drought indices and their application to east Africa," *International Journal of climatology*, vol. 23, pp. 1335 - 1357, 2003.

[146] R. Washington, M. Harrison, and D. Conway, "African climate report," University of Oxford, Oxford, UK. 2004.

[147] ITC, *Principle of remote sensing. International Institute for Geo-information Science and Earth Observation*, 3 ed. Enschede, 2004.

[148] A. Huete, K. Didan, T. Miura, P. Rodriguez, X. Gao, and G. Ferreira, "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," *Remote Sensing of the Environment*, vol. 83, pp. 195-213, 2002.

[149] 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.

[150] B. Lacaze and J. Berges, *Contribution of Meteosat Second Generation (MSG) to drought early warning*. Trier, Germany, 2005.

[151] Z. Wang and X. Li, "Using MODIS Land Surface Temperature and Normalized Difference Vegetation Index products for monitoring drought in the southern Great Plains, USA," *International Journal of Remote Sensing*, vol. 25, pp. 61-72, 2004.

[152] C. Rulinda, W. Bijkera, and A. Steina, "The chlorophyll variability in Meteosat derived NDVI in a context of drought monitoring," *Procedia Environmental Sciences-Science Direct*, vol. 3, pp. 32–37, 2011.

[153] Z. Wang and X. Li, "Using MODIS Land Surface Temperature and Normalized Difference Vegetation Index products for monitoring drought in the southern Great Plains," *USA. International Journal of Remote Sensing*, vol. 25, pp. 61-72, 2004.

[154] P. H. Cabena, R. Stadler, J. Verhees, and A. Zanasi, *Discovering Data Mining: From Concept to Implementation*. New Jersey: IBM, 1998.

[155] R. Groth, *Data Mining: A Hands-On Approach for Business Professionals*. New Jersey: Prentice Hall, 1998.

[156] G. De'ath and K. E. Fabricius, "Classification and regression trees-a powerful yet simple technique for ecological data analysis," *Ecology*, vol. 8, pp. 3178- 3192, 2000.

[157] D. Manila and P. Smyth, *Principles of Data Mining*, vol. 1: The MIT Press, 2001.

[158] H. Edelstein, "Data mining: Exploiting hidden trends in your data," vol. 2: DB2magazine, Spring, 1997.

[159] K. Thearling, "An Introduction to Data Mining," 2001.

[160] J. P. Bigus, *Data Mining with Neural Networks: Solving Business Problems from Application Development to Decision Support*. New York: McGraw-Hill, 1998.

[161] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *BULLETIN OF MATHEMATICAL BIOPHYSICS*, vol. 5, 1943.

[162] D. Rumelhart, G. Hinton, and R. Williams, "Learning internal representation by backpropagating errors," *Nature*, vol. 323, pp. 533–536, 1986.

[163] D. Ashish, R. McClendon, and G. Hoogenboom, "Land-use classification of multispectral aerial images using artificial neural networks," *International Journal of Remote Sensing*, vol. 30, pp. 1989-2004, 2009.

[164] A. Abraham, "Artificial Neural Networks," in *Handbook of Measuring System Design*, P. Sydenham and R. Thorn, Eds. Oklahoma State University, Stillwater, OK, USA: John Wiley & Sons, Ltd. ISBN: 0-470-02143-8, 2005, pp. 901-908.

[165] H. Bischof, W. Schneider, and A. Pinz, "Multispectral classification of Landsat images using neural networks," *IEEE Transactions on Geosciences and Remote Sensing*, vol. 30, pp. 482–490, 1992.

[166] J. Paola and R. Schowengerdt, "A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery," *International Journal of Remote Sensing*, vol. 16, pp. 3033–3058, 1995.

[167] J. Blackard and D. Dean, "Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables," *Computer and Electronics in Agriculture*, vol. 24, pp. 131–151, 1999.

[168] J. Giraudel, "A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination," *Ecological Modelling*, vol. 146, pp. 329–339, 2001.

[169] J. Lee, R. Weger, S. Sengupta, and R. Welch, "A neural network approach to cloud classification," *IEEE Transactions on Geosciences and Remote Sensing*, vol. 28, pp. 846–855, 1990.

[170] K. Yang, T. Koike, and B. Ye, "Improving estimation of hourly, daily and monthly solar radiation by importing global data sets," *Agricultural and Forest Meteorology*, vol. 137, pp. 43–55, 2006.

[171] D. Kumar, S. Raju, and T. Sathish, "River Flow Forecasting using Recurrent Neural Networks," *Water Resources Management*, vol. 18, pp. 143–161, 2004.

[172] T. Kim and J. Valdes, "A nonlinear model for drought forecasting based on conjunction of wavelet transforms and neural networks," *Journal of Hydrologic Engineering*, vol. 8, pp. DOI: 10.1061/(ASCE)1084-0699(2003)8:6(319), 2003.

[173] S. Morid, V. Smakhtinb, and K. Bagherzadehc, "Drought forecasting using artificial neural networks and time series of drought indices," *International Journal of climatology*, vol. 27, pp. 2103 - 2111, 2007.

[174] A. F. Marj and A. Meijerink, "Agricultural drought forecasting using satellite images, climate indices and artificial neural network," *International Journal of Remote Sensing*, vol. 32, pp. 9707 - 9719, 2011.

[175] L. Torgo, "Functional models for regression tree leaves," *Machine learning International Conference*, pp. 385 - 393, 1997.

[176] R. L. Lawrence and A. Wright, "Rule-Based Classification Systems Using Classification and Regression Tree (CART) Analysis," *Photogrammetric Engineering & Remote Sensing*, vol. 67, pp. 1137 - 1142., 2001.

[177] W. N. Venables and B. D. Ripley, "Modern Applied Statistics with S-Plus," Second Edition ed. New York: Springer-Verlag, 1997, pp. 548.

[178] R. L. Lawrence and W. J. Ripple, "Fifteen years of revegetation of Mount St. Helens: A landscape-scale analysis," *Ecology*, vol. 81, pp. 2742 - 2752, 2000.

[179] NDMC, "National Drought Mitigation Center: Home WebSite," vol. 2011, 2011.

[180] NASA, "Developing Seasonal Predictive Capability for Drought Mitigation Decision Support System, National Aeronautics and Space Administration (NASA)," vol. 2012, F. Verter, Ed.: NASA, 2012.

[181] DRMFSS, "Disaster Risk Management and Food Security Sector (DRMFSS)," vol. 2011. Addis Ababa, 2011.

[182] M. Svoboda, D. LeComte, M. Hayes, R. Heim, K. Gleason, J. Angel, B. Rippey, R. Tinker, M. Palecki, D. Stooksbury, D. Miskus, and S. Stephens, "The Drought Monitor," *Bulletin of the American Meteorological Society*, vol. 83, pp. 1181 - 1190., 2002.

[183] A. Wilhite and M. D. Svoboda, *Drought Early Warning Systems in the Context of Drought Preparedness and Mitigation, Expert Group Meeting on Early Warning*

Systems for Drought Preparedness and Drought Management. Lincoln, Nebraska USA: National Drought Mitigation Center, 2000.

[184] M. Rouault and Y. Richard, "Intensity and spatial extension of drought in South Africa at different time scales," *African Journal Online*, vol. 29, pp. 489 - 500, 2003.

[185] FEWSNET, "Normalized Difference Vegetation Index, Product Documentation," vol. 2011: FEWS NET, 2011.

[186] P. Hoefsloot, *LEAP version 2.11 for Ethiopia, User manual*: World Bank and World Food Program, 2009.

[187] D. Straub, G. David, and B. Marie-Claude, "The IS World Quantitative, Positivist Research Methods Website," vol. 2012, D. Galletta, Ed., 2004.

[188] J. H. Hair, R. E. Anderson, R. L. Tatham, and W. C. Black, *Multivariate Data Analysis*: Prentice Hall, 1995.

[189] A. M. Jenkins, "Research Methodologies and MIS Research," *In: Research Methods in Information Systems*, E. Mumford, Ed. Amsterdam, Holland: Elsevier Science Publishers B.V., 1985, pp. 103-117.

[190] A. S. Fotheringham, M. E. Charlton, and C. Brunsdon, "Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis," *Environment and Planning*, vol. 30, pp. 1905 - 1927, 1998.

[191] W. S. Cleveland and S. J. Devlin, "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association*, vol. 83, pp. 596 - 610, 1988.

[192] H. Hamilton, "Computer Science: Knowledge Discovery in Databases," vol. 2011, 2009.

[193] EMA, *National Atlas of Ethiopia*. Addis Ababa: Ethiopian Mapping Authority(EMA), 1988.

[194] R. G. Congalton, "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data," *Remote Sensing of the Environment*, vol. 37, pp. 35 - 46, 1991.

[195] K. M. Wolter, "An Investigation of Some Estimators of Variance for Systematic Sampling," *Journal of the American Statistical Association*, vol. 79, pp. 781-790, 1984.

[196] S. V. Stehman, "Estimating the Kappa Coefficient and its Variance under Stratified Random Sampling," *Photogrametric engineering and remote sensing*, vol. 62, pp. 401-407, 1996.

[197] J. E. Cohen, "The Distribution of the Chi-Squared Statistic Under Clustered Sampling from Contingency Tables," *Journal of the American Statistical Association*, vol. 71, pp. 665-670, 1976.

[198] R. M. Hord and W. Brooner, "Land-Use Map Accuracy Criteria," *Photogrammetric Engineering and Remote Sensing*, vol. 42, pp. 671 - 677, 1976.

[199] K. Fitzpatrick-Lins, "Comparison of sampling procedures and data analysis for a land-use and land-cover map," *Photogrammetric Engineering Remote Sensing*, vol. 47, pp. 343 - 351, 1981.

[200] J. Han and M. Kamber, *Concept Description: Characterization and Comparison*, 1998.

[201] Y. Yao, "Information-Theoretic Measures for Knowledge Discovery and Data Mining," in *Entropy Measures, Maximum Entropy Principle and Emerging Applications*, vol. 119, Karmeshu, Ed.: Springer Berlin Heidelberg, 2003, pp. 115-136.

[202] M. Pechenizkiy, S. Puuronen, and A. Tsymbal, "Does Relevance Matter to Data Mining Research?," in *Data Mining: Foundations and Practice Studies in*

Computational Intelligence, vol. 118, T. Y. Lin, Y. Xie, A. Wasilewska, and C. J. Liau, Eds.: Springer Berlin Heidelberg, 2008, pp. 251-275.

[203] ESRI, *ArcGIS ArcMap Guide*. Redlands, CA: Environmental Systems Resource Institute (ESRI), 2011.

[204] MathWorks, *MATLAB Neural Network Toolbox User's Guide*: Natick, MA: The MathWorks, Inc., 2009.

[205] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Annals of Mathematical Statistics*, vol. 22, pp. 79 - 86, 1951.

[206] H. Akaike, *Information theory as an extension of the maximum likelihood principle. In: Second International Symposium on Information Theory*. Budapest: Akademiai Kiado, 1973.

[207] R. A. Stine, "Graphical Interpretation of Variance Inflation Factors," *American Statistical Association*, vol. 49, pp. 53 - 56, 1995.

[208] R. M. Obrien, "A Caution Regarding Rules of Thumb or Variance Inflation Factors," *Quality & Quantity*, vol. 41, pp. 673 - 690, 2007.

[209] P. A. P. Moran, "Notes on Continuous Stochastic Phenomena," *Biometrika*, vol. 37, pp. 17-23, 1950.

[210] B. N. Holben, "Characteristics of maximum-value composite images from temporal data," *International Journal of Remote Sensing*, vol. 7, pp. 1417-1434, 1986.

[211] USGS, "USGS-Earth Resources Observation and Science (EROS) Center-Elevation Data," vol. 2011, 2011.

[212] Ecodiv.org, "Atlas of the Potential Vegetation of Ethiopia," vol. 2011, P. Breugel, Ed.: Copyright Ecodiv.org, 2010.

[213] ESA, "European Space Agency, Global Land Cover Map," vol. 2011: ESA, 2011. [Online]. Avaiable: http://due.esrin.esa.int/globcover/.

[214] GLCF, "Global Land Cover Facility," vol. 2010: University of MARYLAND, 2010.

[215] IRI, "International Research Institute (IRI) for Climate and Society, IRI/LDEO Climate Data Library," vol. 2011: IRI, 2011.

[216] NOAA, "National Oceanic and Atmospheric Administration, Climate Indices: Monthly Atmospheric and Ocean Time Series," vol. 2011: Earth System Research Laboratory, Physical Science Division, 2011.

[217] 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.

[218] J. W. Hurrell, "Decadal trends in the North Atlantic Oscillation and relationships to regional temperature and precipitation," *Science*, vol. 269, pp. 676-679, 1995.

[219] 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.

[220] K. Wolter and M. S. Timlin, "Measuring the strength of ENSO - how does 1997/98 rank?," *Weather Forecasting*, vol. 53, pp. 315-324, 1998.

[221] NASA, "Global Soil Water Holding Capacity Dataset from UNEP/GRID," in *Global Change Master Directory*, vol. 2011, M. Holland, Ed.: USA.gov, 2011.

[222] G. Churkina, S. W. Running, and A. L. Schloss, "Comparing Global Models of Terrestrial Net Primary Productivity (NPP): The Importance of Water Availability," *Global Change Biology*, vol. 5, pp. 46–55, 1999.

[223] Z. Li, Q. Zhu, and C. Gold, *Digital terrain modeling: principles and methodology*. Boca Raton: CRC Press, 2005.

[224] I. Friis, D. Sebsebe, and P. Breugel, *Atlas of the Potential Vegetation of Ethiopia*, vol. 58: Biologiske Skrifter (Biol.Skr.Dan.Vid.Selsk.), 2010.

[225] 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.

[226] A. Simmons, J. Wallace, and G. Branstator, "Barotropic wave propagation and instability and atmospheric teleconnetion patterns," *Journal of Atmospheric Sciences*, vol. 40, pp. 1363 - 1392, 1983.

[227] M. A. Cane, S. E. Zebiak, and S. C. Dolan, "Experimental forecasts for El-Niño," *Nature*, vol. 321, pp. 827 – 832, 1986.

[228] S. E. Nilcholson and J. C. Selato, "The influence of La Nina on African raifall," *INTERNATIONAL JOURNAL OF CLIMATOLOGY*, vol. 20, pp. 1761–1776, 2000.

[229] H. Paeth and P. Friederichs, "Seasonality and time scales in the relationship between global SST and African rainfall," *Climate Dynamics*, vol. 23, pp. 815–837, 2004.

[230] K. Stahl and S. Demuth, "Linking Stream flow Drought to the Occurrence of Atmospheric Circulation Patterns," *Hydrological Science Journal*, vol. 42, pp. 15–34, 1999.

[231] 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.

[232] NOAA, "Global Vegetation Index Products," vol. 2009: NOAA, 2009.

[233] P. M. Smith, S. Kalluri, S. Prince, and R. Defries, "The NOAA NASA Pathfinder AVHRR 8km land data set," *Photogrametric engineering and remote sensing*, vol. 63, pp. 12-31, 1997.

[234] V. T. Chow, R. D. Maidment, and W. L. Mays, *Applied Hydrology*: McGraw-Hill, 1988.

[235] C. D. Edwards and T. B. McKee, "Characteristics of 20th century drought in the United States at multiple time scales," Department of Atmospheric Sciences, Colorado State University 634, 1997.

[236] UNL, "Weekly SPI Users Manual," National Drought Mitigation Center, University of Nebraska-Lincoln 2000.

[237] 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.

[238] K. Y. Tam and M. Y. Kiang, "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions," *Management Science*, vol. 38, pp. 926 - 947, 1992.

[239] S. Gopal, C. E. Woodcock, and A. H. Strahler, "Fuzzy Neural Network Classification of Global Land Cover from a 1 degree AVHRR Data Set," *Remote Sensing of Environment*, vol. 67, pp. 230-243, 1999.

[240] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Wadsworth International Group: Belmont, California, 1984.

[241] A. S. Laliberte, E. Fredrickson, and A. Rango, "Combining Decision Trees with Hierarchical Object-oriented Image Analysis for Mapping Arid Rangelands," *Photogrametric engineering and remote sensing*, vol. 73, pp. 197 - 207, 2007.

[242] D. Steinberg and P. Colla, *CART – Classification and Regression Trees: Supplementary Manual for Windows*. San Diego, California: Salford Systems, 1997.

[243] Trimble, *eCognition Developer 8.7.1*. München: Trimble Germany GmbH, Trappentreustr. 1, D-80339, 2012.

[244] FEWSNET, "Estimating Meher crop production using rainfall in the long cycle region of Ethiopia, June 21, revised October 6, Addis Ababa 2003. [Online]. Available: http://www.fews.net/docs/Publications/1000330.pdf".

[245] CSA, "Annual Agricultural Sample Survey," vol. 2011. Addis Ababa: Central Statistical Agency of Ethiopia (CSA), 2011.

[246] Z. Wu, N. E. Huang, S. R. Long, and C. Peng, "On the trend, detrending, and variability of nonlinear and nonstationary time series," *The National Academy of Sciences of the USA, PNAS*, vol. 104, pp. 14889 – 14894, 2007.

[247] M. J. Mazerolle, "Improving data analysis in herpetology: using Akaike's Information Criterion (AIC) to assess the strength of biological hypothesis," *Amphibia - Reptilia*, vol. 27, pp. 169 - 180, 2006.

[248] Y. Sileshi and G. R. Demaree, "Rainfall variability in the Ethiopian and Eritrean highlands and its links with the Southern Oscillation Index," *Journal of Biogeography*, vol. 22, pp. 945–952, 1995.

[249] D. Shanko, P. I. J. Camberlin, and Climatol., 1373–1378., "The effect of the southwest Indian Ocean tropical cyclones on Ethiopian drought," *International Journal of Climatology*, vol. 18, pp. 1373–1378, 1998.

[250] T. Gissila, E. Black, D. I. F. Grimes, and J. M. Slingo, "SEASONAL FORECASTING OF THE ETHIOPIAN SUMMER RAINS," *INTERNATIONAL JOURNAL OF CLIMATOLOGY*, vol. 24, pp. 1345 – 1358, 2004.

[251] G. T. Diro, D. I. F. Grimes, E. Black, A. ONeill, and E. Pardo-Iguzquiza, "Evaluation of reanalysis rainfall estimates over Ethiopia," *International Journal of Climatology*, vol. 29, pp. 67 – 78, 2009.

[252] W. Yang, L. Yang, and J. W. Merchant, "An assessment of AVHRR/ NDVI-ecoclimatological relations in Nebraska, USA," *International Journal of Remote Sensing*, vol. 18, pp. 2161–2180, 1997.

[253] B. C. Rundquist, J. A. Harrington, and D. G. Goodin, "Mesoscale satellite bioclimatology," *Professional Geographer*, vol. 52, pp. 331– 334, 2000.

[254] J. Wang, P. M. Rich, and K. P. Price, "Temporal response of NDVI to precipitation and temperature in the central Great Plains, USA," *International Journal of Remote Sensing*, vol. 24, pp. 2345–3364, 2003.

[255] M. T. Hagan and M. B. Menhaj, "Training Feedforward Networks with Marquardt Algorithm," *IEE TRANSACTIONS ON NEURAL NETWORKS*, vol. 5, pp. 989-993, 1994.

[256] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data Clustering: A Review," *ACM Computing Surveys*, vol. 31, pp. 264-323, 1999.

[257] J. Mattsson and R. Anders, "The Recent Droughts in Western Ethiopia and Sudan in a Climatic Context," *Environmental Security*, vol. 20, pp. 172-175, 1991.

[258] M. H. Glantz and W. K. Richard, "Drought as a Constraint to Development in Sub-Saharan Africa," *JSTOR. Ambio*, vol. 14, pp. 334-339, 1985.

[259] L. Olsson, "Drought, Desertification and Market Failure in the Sudan," *JSTOR. Ambio*, vol. 22, pp. 395-403, 1993.

[260] ERDAS, *ERDAS Field Guide*, 5th ed. Atlanta, Georgia: ERDAS®, Inc., 1997.

[261] FAO, "Food and Agriculture Organization of the United Nations, Ethiopia: drought-hit farmers receive emergency aid," FAO, 2003.

[262] FAO, "Production variability and losses," in *Special: Agroclimatic concepts*, vol. 2011, R. Gommes, Ed.: Sustainable Development Department (SD), Food and Agriculture Organization of the United Nations (FAO), 1999.

[263] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *Proceedings of IEEE Computer Society Workshop on Computer Vision*, pp. 16–22, 1987.

[264] J. Arlow and I. Neustadt, *UML and the Unified Process, Practical Object Oriented Analysis and Design*. Boston: Pearson Education Limited, 2002.

[265] J. T. Heaton, *Introduction to Neural Networks for Java*, Second Edition ed: jeffheaton, 2008.

[266] G. Cybenko, "Approximation by super positions of a sigmoidal function," *Mathematics of Control Signals, and Systems*, vol. 2, pp. 303–314, 1989.

[267] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359 - 366, 1989.

[268] R. Lippmann, "An introduction to computing with neural nets," *ASSP Magazine, IEEE*, vol. 4, pp. 4 - 22, 1987.

[269] F. S. Wong, "Time series forecasting using backpropagation neural networks," *Neurocomputing*, vol. 2, pp. 147–159, 1991.

[270] Z. Tang and P. A. Fishwick, "Feedforward neural nets as models for time series forecasting," *ORSA Journal of Computing*, vol. 5, pp. 374 - 385, 1993.

[271] D. Fletcher and E. Goss, "Forecasting with neural networks: an application using bankruptcy data," *Information & Management*, vol. 24, pp. 159 - 167, 1993.

[272] B. L. Zhang and Z. Y. Dong, "An adaptive neural-wavelet model for short term load forecasting," *Electric Power Systems Research*, vol. 59, 2001.

[273] I. Charlier and J. Lewis, *USER'S MANUAL, WINDISP 4.0, Map and Image Display and Analysis Software*. Rome: FAO Global Information and Early Warning System, 1999.

[274] NMA, "National Meteorological Agency," vol. 2011. Addis Ababa: NMA, 2011.

[275] WFP, "World Food Programme (WFP)." Addis Ababa: WFP, 2011.

[276] EUMETSAT, *MSG - IN - ORBIT - IN USE*, MSG.02 Version 1 ed. Darmstadt: EUMETSAT, 2005.

[277] ESRI, "Working with ArcGIS Server," vol. 2012: Environmental Systems Research Institute, Inc., 2012.

[278] L. Wolf and M. Williams, "GEONETCast—Delivering Environmental Data to Users Worldwide," *IEEE SYSTEMS JOURNAL*, vol. 2, pp. 401 - 405, 2008.

[279] A. R. Hevner, "A Three Cycle View of Design Science Research," *Scandinavian Journal of Information Systems*, vol. 19, pp. 87 - 92, 2007.

[280] G. Rothermel and M. J. Harrold, "Analyzing Regression Test Selection Techniques," IEEE Transaction on Software Engineering, vol. 22, pp. 529 - 551, 1996.

[281] J. K. Titlow, *A precipitation-based drought index for the Delaware River basin*, vol. 40: Thornthwaite Associates and University of Delaware, 1987.

[282] H. N. Bhalme and D. A. Mooley, "Large scale droughts/floods and monsoon circulation," *Monthly Weather Review*, vol. 108, pp. 1197–1211, 1980.

[283] 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*, 1982.

[284] A. G. Repin, "Terra Library, Geography Research Assistant," vol. 2012, A. G. Repin, Ed., 2010.

[285] NASA, "National Aeronautics and Space Administration, MODIS Satellite," vol. 2012: NASA, 2012.

[286] MindSitesGroup, "The GeoCommunity Marketplace," vol. 2012, 2012.

Appendices

Appendix 1: Climatic Indices and Satellite Data for Drought Monitoring

1.1 Climatic Indices

PDSI

The PDSI [143] is probably still one of the most complex drought indices in use today, and it is also one of the few that allows a direct comparison of index values between different climatological regions. It is a soil moisture algorithm calibrated for relatively homogeneous regions, and is based on moisture inflow, outflow and storage [281]. Details about the PDSI algorithm are presented in [142].

In describing the actual drought situation in a given region, Palmer [143] considered a drought to be established when the index is less than or equal to −1.0, and a wet spell is established when the index is greater than or equal to 1.0. A drought is considered to have certainly ended when the index reaches the "near normal" category, which lies between −0.5 and +0.5, or when the index returns to zero.

BMI

The Bhalme–Mooley index (BMI) was developed by Bhalme and Mooley [282] for assessing drought intensity using a precipitation parameter. The computational details of the BMI and the PDSI are generally similar, with just a few differences. Bhalme and Mooley replaced the moisture index in Palmer's algorithm with a simpler monthly rain index computed from the rainfall data only. Therefore, the BMI models the percentage departure of P from the long-term averages using an algorithm similar to that of the PDSI [142].

SPI

McKee et al. [28] defined the Standard Precipitation Index (SPI) as the number of standard deviations that the observed cumulative rainfall at a given time scale would deviate from the long-term mean. As a single numeric value, the SPI can be compared across regions with markedly different climates.

Since the cumulative precipitation may not be normally distributed, McKee et al. [28] transformed the data approximately to the normal domain to standardize the drought index. The time scale of the SPI is also flexible, which is an attractive feature because it is possible to experience wet conditions at one time scale but dry conditions at another simultaneously. Edwards and McKee [235] selected a 3-month SPI for a short-term drought index, a 12-month SPI for an intermediate-term drought index, and a 48-month SPI for a long-term drought index. In principle, the

SPI index is based on the probability of precipitation for any time scale. The detailed algorithm of SPI described in [142].

According to McKee et al.'s [28] definition of drought categories, SPI ≥ 2.0 means an extremely wet spell, $1.99 \geq$ SPI ≥ 1.0 means very to moderately wet, $0.99 \geq$ SPI ≥ -0.99 means near normal, $-1.0 \geq$ SPI ≥ -1.99 means moderately to severely dry, and SPI < -2.0 means an extremely dry spell. A drought event is considered to have occurred any time the SPI is continuously -1.0 or less. The event ends when the SPI becomes positive. Therefore, each drought event has a well-defined duration [142].

As a single numeric value, the SPI can be compared across regions with markedly different climates. Ntale and Gan [145] indicated that SPI is more suitable for monitoring droughts and it has modest data requirements and is easy to interpret. The SPI is an index based on the probability of recording a given amount of precipitation, and the probabilities are standardized so that an index of zero indicates the median precipitation amount (half of the historical precipitation amounts are below the median, and half are above the median). The index is negative for drought, and positive for wet conditions. As the dry or wet conditions become more severe, the index becomes more negative or positive [145]. SPI can also be used to determine the magnitude of a drought in a locality. Drought magnitude is the duration of drought with negative SPI deviation expressed in time period of month(s) [28].

SWSI

The Surface Water Supply Index (SWSI) [283] is primarily developed as a hydrological drought index with an intention to replace PDSI for areas where local precipitation is not the sole or primary source of water. It is calculated based on monthly non exceedence probability from available historical record of reservoir storage, stream flow, snow pack, and precipitation. SWSI is good measure to monitor the impact of hydrologic drought and industrial water supplies, irrigation and hydroelectric power generation. However, there is a time lag before precipitation deficiencies are detected in surface and subsurface water sources. As a result, the hydrological drought is out of phase with agricultural drought. Because of this, the SWSI is not a suitable indicator for agricultural drought [141].

1.2 List of potential drought monitoring attributes from satellite earth observation systems as of 2012 [185, 284-286].

No.	Satellite name	Type/Owner	No	Satellite name	Type/Owner
1	Aqua (EOS PM-1)	Earth Observing System	59	CBERS-1	Brazilian Space Agency (AEB)
2	Aura	Earth Observing System	60	CBERS-2	AEB
3	GRACE	Earth Observing System	61	CBERS-2B	AEB
4	Jason 1	Earth Observing System	62	SAC-A	Argentina Space Agency (CONAE)
5	Ocean Surface Topography Mission	Earth Observing System	63	SAC-B	CONAE
6	Orbview-2	Earth Observing System	64	SAC-C	CONAE
7	Orbiting Carbon Observatory	Earth Observing System	65	SAC-D	CONAE
8	Terra (EOS AM-1)	Earth Observing System	66	SAOCOM	CONAE
9	Disaster Monitoring Constellation (6 different satellites as of 2012)	Commercial	67	SPOT	Centre National d'Etudes Spatiales (CNES)

No	Satellite name	Type/Owner	No	Satellite name	Type/Owner
10	EROS A & B	Commercial	68	TOPEX/Poseidon	CNES
11	FORMOSAT-2	Commercial	69	Envisat	European Space Agency (ESA)
12	IKONOS	Commercial	70	ERS	ESA
13	QuickBird	Commercial	71	Oceansat-2 23	Indian Space Research Organisation (ISRO)
14	RapidEye (5 identical satellites as of 2012)	Commercial	72	IMS-1 28	ISRO
15	SPOT 4 & 5	Commercial	73	Cartosat-2A 28	ISRO
16	TerraSAR-X & TerraDEM-X	Commercial	74	CARTOSAT-2	ISRO
17	COSMO-SkyMed (4 identical satellites as of 2012)	Commercial	75	IRS P5 (CARTOSAT-1)	ISRO
18	WorldView-1	Commercial	76	IRS P4 (Oceansat 1)	ISRO
19	WorldView-2	Commercial	77	IRS P3]	ISRO
20	GeoEye-1	Commercial	78	IRS P2	ISRO
21	GOES 9	GOES (Geostationary Operational Environmental Satellite)	79	IRS 1D	ISRO
22	GOES 10	GOES	80	IRS 1C	ISRO
23	GOES 12	GOES	81	IRS P1	ISRO
24	NOAA-15	National Polar-orbiting Operational Environmental Satellite System (NPOESS)	82	IRS 1B	ISRO
25	NOAA-16	NPOESS	83	IRS 1A	ISRO
26	NOAA-17	NPOESS	84	MOS-1 (Momo-1)	JAXA (former NASDA)
27	NOAA-18	NPOESS	85	MOS-1b (Momo-1b)	JAXA
28	NOAA-19	NPOESS	86	JERS-1 (Fuyo-1)	JAXA
29	GMS-1 / Himawari-1	NPOESS	87	ADEOS (Midori)	JAXA
30	GMS-2 / Himawari-2	NPOESS	88	ADEOS II (Midori II)	JAXA
31	GMS-3 / Himawari-3]	NPOESS	89	GOSAT (Ibuki)	JAXA
32	GMS-4 / Himawari-4	NPOESS	90	ALOS (Daichi) PALSAR	JAXA
33	GMS-5 / Himawari-5	NPOESS	91	AVNIR-2 and PRIS	JAXA
34	Landsat 1	Landsat program	92	Lapan-TUBsat	Lembaga Penerbangan dan Antariksa Nasional (LAPAN INDONESIA)
35	Landsat 2	Landsat program	93	TIMED (Thermosphere Ionosphere Mesosphere Energetics and Dynamics)	NASA
36	Landsat 3	Landsat program	94	TOPEX/Poseidon	NASA
37	Landsat 4	Landsat program	95	Upper Atmosphere Research Satelli	NASA
38	Landsat 4	Landsat program	96	NOAA-4	NOAA
39	Landsat 5	Landsat program	97	Vanguard 2	Project Vanguard
40	Landsat 6	Landsat program	98	BelKA	National Academy of Sciences of Republic of Belarus
41	Landsat 7	Landsat program	99	Elektro-L	Russian Federal Space Agency (Roscosmos)
42	Munin	Swedish National Space Board	100	Monitor-E	Roscosmos
43	MetOp	Meteosat	101	Resurs-DK1	Roscosmos
44	Meteosat 5	Meteosat	102	RADARSAT-1	RADARSAT series
45	Meteosat 6	Meteosat	103	RADARSAT-2	RADARSAT series
46	Meteosat 7	Meteosat	104	Kompsat-2	South Korea
47	Meteosat 8	Meteosat	105	Theos-1	THAILAND
48	Meteosat 9	Meteosat	106	Meteor 1 series	Meteor series
49	TIROS-1	Weather	107	Meteor 2 series	Meteor series
50	TIROS-2	Weather	108	Meteor 3 series	Meteor series
51	TIROS-3	Weather	109	FY-1 series	FY (Feng Yun) series
52	TIROS-4	Weather	110	FY-2 series	FY (Feng Yun) series
53	TIROS-5	Weather	111	FY-3A	FY (Feng Yun) series
54	TIROS-6	Weather			
55	TIROS-7	Weather			
56	TIROS-8	Weather			
57	TIROS-9	Weather			
58	TIROS-10	Weather			

Appendix 2: Rain Fall Pattern for the 2 x 2 Degree Grids

Appendix 3: Anisotropic Diffusion Algorithm and Implementation Code

```
%% The following code is used to implement the SDNDVI Image Anisotropic diffusion
%Developed in University of Nebraska-Lincoln
%October 2 2011
strc2='0'; %This to hold 0
dir1='C:\DroughtObject\'; %The path for raw data
fmt='.tif'; %The format of the raw data
pwd % PWD-displays the current working directory.
%-----------------------------------------------------------------------
for i=6:10 % The months from June to October
  for j=1:3 %The dekada periods from June to October
    strc1='SDNDVI82';
    stri=num2str(i);
    strj=num2str(j);
    strc=strcat(dir1,strc1,strc2,stri,strc2,strj,fmt);%Concatenate
    u=imread(strc);%Reads each image
    num_iter = 10;% Number of iterations
    delta_t = 1/5;%integration constant
    kappa = 45;  %Gradient modulus threshold that controls the conduction
    option = 1;  %Conduction coefficient functions proposed by Perona & Malik
    ad = anisotropicdiff(u,num_iter,delta_t,kappa,option); %Execute Anisotropic diffusion
    r = mat2gray(ad); %Convert from mat to gray image
    dir2='C:\DroughtObject\denoised\'; %The directory where to write the image
    output=strcat(dir2,strc1,strc2,stri,strc2,strj,fmt);%Concatenate
    imwrite(r,output,'tif'); %Write the output in the directory
  end
end
```

Appendix 4: Fuzzy DroughtObject Segmentation Algorithm

```
%% Fuzzy connectedness segmentation implementation
%Read image and scale to [0,1]
I=double(imread('E:\DroughtObject\8409.tif'))./255;
disp(sprintf('Image size: %d x %d',size(I,1),size(I,2)));
%1. Compute adjacency
n=4;
k1=0.25;
A=adjacency(I,n,k1);
%2. Compute affinity
k2=2;
K=affinity(I,A,k2);
%3. Select seed points
S=zeros(size(I));
S(13:17,31:34)=1; %Extremely dry
S(16:16,31:31)=1; %Moderately dry
S(47:51,45:48)=1; %Near normal
S(31:34,35:38)=2; % above optimum
%4. Show seeds overlayed on image
I_rgb=repmat(I,[1,1,3]); %make rgb image (required by imoverlay)
figure(1)
image(I_rgb)
imoverlay(S,S>0); %requires image in range [0,1]
colorbar
title('Seed regions');
%5. Compute FC
disp(sprintf('Processing...'));
FC=irfc(S,K); %iterative relative FC
figure(2) %show resulting FC-map
imagesc(FC,[0,1])
colorbar
title('Fuzzy connectedness map');
thresh=0.9; %6. Threshold value
figure(3) %show the 0.9-connected component overlayed on original image
image(I_rgb)
imoverlay(FC,FC>thresh);
title(sprintf('Fuzzy connected component at level %.2f',thresh));
figure (4)
image(I_rgb)
r = mat2gray(FC);
dir2='E:\DroughtObject\Segmented';
strc1='8409_segmented.tif';
fmt='.tif';
output=strcat(dir2,strc1,fmt);
imwrite(r,output,'tif');
```

Appendix 5: An Excerpt from Backpropagation Algorithm

```
%%Backpropagation Algorithm Implemenation Code
clc
clear all
close all
disp('Beginning the Training phase of ANN...');
E=100; %initially assumed error to continue a while loop
d=0;
%Read Training Data getdata();
[a,b]=readData();
[ar,ac]=size(a);   %ac-is number of data used as input vector
[br,bc]=size(b);   %bc-corresponding no of outputs
          %ar- & br-no of descriptors per input- & output-vectors
 %Adjust Weights & Bias
    for k=l-1:-1:1
       for i=1:no(k)
          for j=1:no(k+1)
             delw(i,j,k)=N*del(j,k+1)*Y(i,k)+0.4*delw(i,j,k);
             W(i,j,k)=W(i,j,k)+delw(i,j,k);
          end
       end
    end

    for k=2:l
       for i=1:no(k)
          delb(i,k)=N*(-1)*del(i,k);
          B(i,k)=B(i,k)+delb(i,k);
       end
    end
 end
%-----------------------------------------------------------------
 err(d)=E;
  if rem(d,100)==0          %plot after every 100 epochs
    plot([1:d],err,'blu-',[1:d],Emax,'gre-');

    pause(0.05);

  end;
end
%End of Training Loop
toc;
save weightmatrix W B data;
save norms m1 v1 m2 v2 no l;
plot(err),title('Learning Curve'),xlabel('      No. of Epochs ------> ');
ylabel('Mean Square Error ------> ');
hold on;
plot([0:d],Emax,'-*gre');
grid on;
%End of Training Loop
```

Appendix 6: An Excerpt from Using Mode of ANN Algorithm

```
%=============================================
%This code implements the basic back propagation of error learning algorithm
%=============================================
clear all;
clc
close all;
%Reads the test vectors from files
[a,b]=readData();
ain=a;
[ar,ac]=size(a);
%Loads the stored weights after training & structure of the ANN
load weightmatrix W B;
load norms no l;
f=1;
% Normalizes the test inputs and outputs
for i=1:no(1)
   m1(i)=mean(a(i,:));
   v1(i)=std(a(i,:));
   a(i,:)=(a(i,:)-m1(i))/v1(i);
end;
b1=b;
for i=1:no(l)
   m2(i)=mean(b(i,:));
   v2(i)=std(b(i,:));
   b(i,:)=(b(i,:)-m2(i))/v2(i);
end;
for c=1:ac
   for i=1:no(1)
      Y(i,1)=a(i,c);
      V(i,1)=a(i,c);
   end
   for k=1:l-2
      for j=1:no(k+1)
         w=W(1:no(k),j,k);
         y=Y(1:no(k),k);
         V(j,k+1)=y'*w;
         Y(j,k+1)=logsig(V(j,k+1)-B(j,k+1));
      end
   end
   k=l-1;
.
.
for z=1:size(b,1)
figure;
plot(x,b(z,:),'o-r',x,O(z,:),'blu-o');
xlabel('Number of descriptors'),ylabel('Descriptors/Features');
legend('Actual Plot','Obtained Curve');
title('Feature vector of predicted image');
end
```

List of Publications

1. Peer Reviewed Journals

1) Getachew, B., Tadesse, T., Solomon, A., Shawndra, H., Yitaktu, T. (2012). 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, No. 1, pp. 67 - 94.
2) Getachew, B., Shawndra, H., Tadesse, T., Solomon, A. (2011). Using Satellite Images for Drought Monitoring: A Knowledge Discovery Approach, Journal of Strategic Innovation and Sustainability, Vol. 7, No. 1 , pp135 – 153, available at http://www.na-businesspress.com/jsis/berhanweb.pdf.
3) Getachew, B., Shawndra, H., Tadesse. T., Solomon, A. (2012) Drought Prediction System for Improved Climate Change Mitigation IEEE Transactions on Geosciences and Remote Sensing, vol. xx, pp. xx, 2013, In Press.

2. Peer Reviewed Book Chapter

4) Getachew, B., Tadesse, T., Solomon, A., Shawndra, H. (2010). Drought Monitoring in- Food-Insecure Areas of Ethiopia by Using Satellite Technologies, W. Leal Filho (ed.), Experience of Climate Change Adaptation in Africa, Climate Change Management, DOI: 10.1007/978-3-642-22315-0_11, pp. 1145 – 1162, available at http://link.springer.com/chapter/10.1007%2F978-3-642-22315-0_11.

3. Peer Reviewed Proceedings

5) Getachew, B., Solomon, A., Tadesse, T., Shawndra, H. (2011). Drought Object Modeling for Climate Change Mitigation: The Case of Ethiopia, Proceedings of the 1st EICTF, October 27, 2011, Addis Ababa , Ethiopia, available at http://www.eitpa.org.et/EICTF/Downloads/Proceedings.pdf.
6) Getachew, B., Solomon, A., Anat, H. (2010). Knowledge Discovery from Satellite Images for Drought Monitoring in Food Insecure Areas, Americas Conference on Information Systems 2010 Conference Proceeding, available at http://aisel.aisnet.org/amcis2010/257/.
7) Getachew, B., Tadesse, T., Solomon, A., Shawndra, H. (2012). Drought Information Mining from Satellite Images for Improved Climate Change Mitigation, 2012 World Congress on Information and Communication Technologies, 2012, available at http://rpsonline.com.sg/rps2prod/wict2012/html/154.xml.

Declaration Sheet

This Dissertation is my original work, has not been presented for a degree in any other university and all sources of material used for the Dissertation have been duly acknowledged.

Getachew Berhan Demisse

As an advisor and co-advisors, we confirm to the best of our knowledge.

TSEGAYE TADESSE		
Dr. Tesgaye Tadesse	**Dr. Solomon Atnafu**	**Dr. Shawndra Hill**