Mining Drought from Remote Sensing Images

Musaningabe Rulinda Coco
March, 2007
Mining Drought from Remote Sensing Images

by

Musaningabe Rulinda Coco

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: Geoinformatics.

Thesis Assessment Board

Thesis advisor       Dr. Ir. W. Bijker (First Supervisor)
                     Prof. Dr. A. Stein (Second Supervisor)
Assessment Board Chair Prof. Dr. J.L. van Genderen (chair)
Thesis examiners     Dr. D.E. van de Vlag
Disclaimer

This document describes work undertaken as part of a programme of study at the International Institute for Geo-information Science and Earth Observation (ITC). All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.
Abstract

Remote sensing advancements offered new opportunities to access timely information for drought monitoring. In traditional NDVI based drought monitoring, the common practice is to apply crisp classification techniques, based on vegetation stress intensity values. Such approach is less effective for monitoring low vegetation stress, which, when extended over a long period of time can cause severe drought. Drought monitoring from images often require time series of images that are difficult to process using traditional image processing methods. The extraction and analysis of large amount of data can be achieved with the use of image mining techniques.

In this study, the main objective is to develop a framework that uses image mining techniques to monitor drought. The proposed framework considers both vegetation stress intensity and duration to characterize drought from remote sensing images. As a case, the drought that occurred in East Africa during the period of December 2005 to February 2006, is taken and analyzed. The data used for this study are NOAA-AVHRR derived NDVI images. A fuzzy membership function is used to characterize drought on space. The patterns of these droughts objects on space are extracted and tracked over time. The extracting and tracking algorithms are automated to analyze a series of images. The results obtained formed the gained knowledge about drought in space and time. This knowledge is used for further understanding the behavior of drought characteristics based on the vegetation conditions.

The proposed approach allows an easy monitoring of drought in space and time. Such model can be used to monitor other vague phenomena, which are characterized based on intensity and duration parameters. This research is promising for data mining processes applied on vague phenomenon monitoring.

Keywords

Remote sensing, spatial data mining, spatial data quality, vague phenomenon, uncertainties, fuzzy set theory, NOAA-AVHRR NDVI images, drought monitoring, East Africa
Abstract
Contents

Abstract i

List of Figures vii

List of Tables ix

Acknowledgements xi

1 Introduction 1
  1.1 Background ................................................. 1
  1.2 Problem statement ........................................... 2
  1.3 Research objective ........................................... 2
  1.4 Research questions .......................................... 2
  1.5 Overview of the methodology ............................... 3
  1.6 Structure of the thesis ..................................... 4

2 Literature review 5
  2.1 Defining drought ............................................. 5
    2.1.1 Types of drought ....................................... 6
    2.1.2 Sequence of types of drought [4] ....................... 7
  2.2 Drought monitoring ......................................... 8
    2.2.1 Introduction ............................................. 8
    2.2.2 Remote sensing indices used in drought monitoring .... 9
## Contents

2.2.3 Monitoring drought from NDVI .................................. 10

2.3 Study Area .......................................................... 10

2.3.1 East Africa ......................................................... 10

2.3.2 Drought in the East Africa ....................................... 12

2.4 Dataset description .................................................. 14

3 Methodology ............................................................. 19

3.1 Data mining .......................................................... 21

3.1.1 Data mining models and techniques .............................. 21

3.1.2 Data mining in remote sensing .................................. 23

3.1.3 Data mining for drought monitoring ............................. 23

3.2 Spatial data quality .................................................. 24

3.3 Fuzzy set theory ...................................................... 25

3.3.1 Fuzzy sets .......................................................... 26

3.3.2 Types of Fuzzy membership functions .......................... 26

3.4 Pixel vs object based RS analysis .................................. 26

3.5 Data analysis .......................................................... 28

3.5.1 Understanding drought from images ............................ 29

3.6 Mining Tasks .......................................................... 30

3.6.1 Extracting process ................................................ 32

3.6.2 Tracking process .................................................. 33

3.7 Space-time analysis ................................................... 33

3.8 validation ............................................................. 34

3.9 Software packages used .............................................. 34

4 Results ................................................................. 35

4.1 Understanding drought from the images ............................ 35

4.2 Characterizing drought over space .................................. 35
## List of Figures

1.1 Overview of the methodology .............................................. 3

2.1 Sequence of drought impacts [4] ........................................ 7

2.2 General Map of East Africa [14] ........................................ 11

2.3 Land cover of East Africa in 2000, modified from [7] ............ 12

2.4 Drought in East Africa: December 2005 and February 2006. From darker brown: severe drought to yellow, less severe drought . . . 13

2.5 Situation in Somalia, October 2005 (photo from UNICEF) . . . 13

2.6 Current NDVI images ..................................................... 15

2.7 Corresponding long-term mean NDVI images ....................... 16

3.1 Overview of the proposed RS drought monitoring approach . . . 20

3.2 Types of fuzzy membership functions ................................. 27

3.3 Illustration of the type of the membership function selected . . 30

3.4 Overview of the mining tasks and expected outcomes ............ 31

4.1 NDVI deviation images ................................................... 36

4.2 Variation of total number of positive, zero and negative pixels . 37

4.3 Histograms of deviation images ........................................ 38

4.4 Images obtained after applying the defined membership function with $a = -40$ and $b = 0$. Drought membership values range from 0.025 (lowest value) to 1(highest value) ......................... 40

4.5 Tracked objects variation over time for the five locations . . . 44
List of Figures

4.6  Drought coverage, based on temporal continuity  . . . . . . . . . 46
4.7  Tracked objects based on temporal continuity on locations 4 and 5 48
4.8  Drought coverage area using a fuzzy approach  . . . . . . . . . 48
4.9  Drought severity expansion: Results vs. Reports  . . . . . . . . 50
4.10 Drought severity in December: Results vs. Reports  . . . . . . 51
4.11 Drought severity in February: Results vs reports  . . . . . . . 52
List of Tables

2.1 Data characteristics .................................................... 17

3.1 Common types of data mining tasks and techniques (Modified from Sankar et.al.[16]) ............................................. 22

4.1 Table of extracted drought objects ................................... 41
4.2 Tracked objects at location(83,180) ................................. 42
4.3 Tracked objects at location(71,122) ................................. 42
4.4 Tracked objects at location (44,230) ............................... 43
4.5 Tracked objects at location (131,202) ............................. 43
4.6 Tracked values at location(166,204) ............................... 43
4.7 Table of tracked objects based on their temporal continuity ... 47
4.8 Drought coverage with fuzzy approach ............................ 49
4.9 Drought coverage with crisp approach ............................. 49
Acknowledgements

I first address my deep and sincere thanks to Dr. Ir. Wietske Bijker and Prof. Dr. Alfred Stein for their professional supervision throughout this work. With them, I learnt not only how to develop and implement scientific ideas but mostly how to think as a researcher.

I also take this opportunity to express my deep thanks to Dr. Arta Dilo, for her unconditional guidance. I thank her for all the discussions we had together; her sense of details often required me to reconsider many times each and every word I used to ensure better meanings.

My deep appreciations go also to various ITC staff members and students who helped me to understand various aspects of drought modeling. Special thanks go in this regard to Mr. Alemseged Tamiru Haile.

I can't forget to thank NUFFIC through the CGIS-NUR/ITC Project for the financial support. Thanks for trusting me and granting me the scholarship to pursue MSc studies.

Last but not least, warm thanks to all my GFM2 classmates at ITC and officemates in Rwanda; to my family and to all my great friends all over the World. May God bless you all.
Acknowledgements
Chapter 1

Introduction

1.1 Background

The advancements in satellite technology have led to the increase of amount of data collected and stored; Images are now being collected covering large areas of earth surface, often at high temporal and spatial resolutions. These images can be used for a range of different purposes but little has been done to analyze and to extract all possible information from them. The human mind capacity does not allow to easily consider each image individually, analyze it and develop relationships with other images of varying time steps. Hence, issues of processing large sets of remotely sensed data using automatic techniques become important solutions to consider.

Although geospatial natural phenomena are represented in a GIS as crisp entities, most of them present in reality uncertainties. This is mainly due to their poor descriptions and vague extents. geospatial data quality has become an important issue within the frame of revision and updating of spatial information\textsuperscript{2}. Progress in this respect has been made, with researches conducted in the last years to address the issue of defining, handling and storing vague objects \textsuperscript{2}\textsuperscript{6}\textsuperscript{21} and their monitoring in space and time \textsuperscript{10}\textsuperscript{26}.

Drought is considered by many to be the most complex but least understood of all natural hazards, affecting more people than any other hazard \textsuperscript{28}. It is a creeping natural hazard, characterized by a slow-onset, that is a normal part of climate for virtually all regions of the World. Drought onset, duration and severity are often difficult to determine and the impacts are largely non-structural and spread over a larger geographical area than are damages from other natural hazards. The non-structural characteristic of Drought impacts has hindered the development of accurate, reliable, and timely estimates of occurrence. For the nations and regions to make progress in reducing the serious
1.2 Problem statement

consequences of drought, they must improve their understanding of the hazard and the factors that influence vulnerability. It is important for drought-prone regions to better understand their drought climatology and establish comprehensive and integrated drought information system. The lack of timely available ground observations for monitoring drought impacts has led to the exploitation of remote sensing images. Data from sensors such as the MODerate resolution Imaging Spectrometer (MODIS) onboard TERRA satellite and the Advanced Very High Resolution Radiometer (AVHRR) onboard NOAA satellites are used in different regions of the world [24] in the process of monitoring drought. The use of remote sensing data is especially recommended in regions where ground observation data density is poor.

1.2 Problem statement

In traditional NDVI based drought monitoring, the common practice is to apply crisp classification techniques, based on vegetation stress intensity values. Such approach is less effective for monitoring low vegetation stress, which, when extended over a long period of time can cause severe drought.

1.3 Research objective

The main objective of this study is to develop a framework that uses image mining techniques to monitor drought. The proposed framework considers both vegetation stress intensity and duration to characterize drought from remote sensing images.

1.4 Research questions

In order to achieve the above stated objective, the following questions have been addressed:

1. How can drought be understood from RS data?
2. Which function can best characterize drought?
3. How to define and extract drought objects?
4. How can drought be tracked over time?
5. How can drought be analyzed over space and time?
Chapter 1. Introduction

The Severe drought that occurred in East Africa during the period of the end of the year 2005 to the beginning of the year 2006 are taken as a case study. NOAA-AVHRR derived $NDVI$ images covering the entire study area are used. In this study, the following assumptions were made:

1. The region is vegetated and mainly rain-fed;
2. Vegetation stress is assumed to be caused by prevailing climatic drought conditions;
3. Drought covered area in the region is large enough to be identified from the low resolution (64 $Km^2$) NOAA-AVHRR derived $NDVI$ images used.

1.5 Overview of the methodology

![Overview of the methodology](image)

The research methodology adopted to conduct the study, as illustrated in Figure 1.1, involved four major steps from data acquisition to space-time analysis of drought objects. After the data has been acquired, we calculated the $NDVI$ deviation values for each location of the nine dekadals and produced the deviation maps. A function has been selected to characterize drought severity over space based on the vegetation response; the uncertainty derived from its vague
description was considered and addressed using the fuzzy set theory. Drought objects were defined and extracted from the images, to later be tracked over time. The patterns of the objects tracked over time constituted the knowledge gained, which was later used to further understand the characteristic behavior of drought. A space-time analysis, to understand drought characteristics over space and time, was lastly conducted.

1.6 Structure of the thesis

The thesis report is structured in the following way: Chapter two provides with a literature review of the concepts of drought and its monitoring. It also describes the study area and the dataset used in this study. The third chapter details the methodology used. Chapter four presents the results obtained and discusses them, while chapter chapter five discusses the overall proposed framework. The last chapter of this thesis report concludes the study and give some recommendations for further studies in the field.
Chapter 2

Literature review

This chapter is divided into three sections: the first section reviews the main concepts of drought and its monitoring. It briefly list the most commonly used remote sensing indices used in drought monitoring. The second section introduces the study area and further relates drought to it. This chapter ends with the section describing the dataset used in this study.

2.1 Defining drought

Drought is a weather-related natural disaster; It is defined as an extended period of abnormally dry weather that causes water shortages and crop damage. It is a creeping natural hazard characterized by a slow-onset and is a normal part of climate for virtually all regions of the World. It is a temporary aberration and its characteristics may vary significantly from one region to another. Drought should not be viewed as merely a physical phenomenon or natural event; Its impacts on society result from the interplay between a natural event—less precipitation than expected resulting from natural climatic variability—and the demand people place on water supply [4]. From a cause - and - effect perspective, drought is usually defined as a deficiency of precipitation over a long period of time that results in a shortage of surface/ground water and a reduction of plant growth. Other climatic factors such as high temperature, high wind and low relative humidity are often associated with it in many regions of the World and can significantly aggravate its severity.

Drought definitions are mainly divided into two kinds: conceptual and operational. The conceptional definitions of drought help the people understand the concepts of drought for example: “Drought is a protracted period of deficient precipitation resulting in extensive damage to crops, resulting in loss of yield.[4]” The conceptual definitions of drought may also be important for establishing drought policies. Operational definitions help people identify the beginning, end, and degree of severity of a drought[4]. To determine the beginning of drought, operational definitions specify the degree of departure from the aver-
2.1. Defining drought

age of a climatic or environmental variable. This is usually done by comparing the current situation to the historical average, often based on a long time (20–year to 30–year) period of record. The threshold identified as the beginning of a drought is usually established somewhat arbitrarily, rather than on the basis of its precise relationship to specific impacts [4]. Operational definitions of drought are dependent on perspectives; from a meteorological perspective to a socio-economic perspective. The following subsection briefly discusses the types of drought on a perspective basis and the sequence of their impacts.

2.1.1 Types of drought

- **Meteorological drought** is defined usually on the basis of the degree of dryness — in comparison to some ‘normal’ or average amount — and the duration of the dry period [4]. These definitions must be regarded as region specific since the atmospheric conditions that result in deficiencies of precipitation are highly dependent on climatic regimes, hence vary from region to region.

- **Agricultural drought** links various characteristics of meteorological or hydrological drought to agricultural impacts, focusing on precipitation shortages, differences between actual and potential evapotranspiration, soil water deficits, reduced ground water or reservoir levels, and so forth [4]. Plant water demand depends on prevailing weather conditions, biological characteristics of the specific plant, its stage of growth, and the physical and biological properties of the soil. A good definition of agricultural drought should be able to account for the variable susceptibility of crops during different stages of crop development, from emergence to maturity. Deficient topsoil moisture at planting may hinder germination, leading to low plant populations per hectare and a reduction of final yield. However, if topsoil moisture is sufficient for early growth requirements, deficiencies in subsoil moisture at this early stage may not affect final yield if subsoil moisture is replenished as the growing season progresses or if rainfall meets plant water needs.

- **Hydrological drought** is associated with the effects of periods of precipitation shortfalls on surface or ground water supply. The frequency and severity of hydrological drought is often defined on a watershed or river basin scale [4]. Hydrological droughts usually lag the occurrence of meteorological and agricultural droughts as it takes longer time to occur. As an example, a precipitation deficiency may result in a rapid depletion of soil moisture that can easily affect crop production, but the impact of this deficiency on reservoir levels may affect hydroelectric power production only if the precipitation deficiency persist for a longer period.

- **Socio-economic** definitions of drought associate the supply and demand of some economic good with elements of meteorological, hydrological, and agricultural drought [4]. It differs from the aforementioned types of drought.
because its occurrence depends on the time and space processes of supply and demand to identify or classify droughts. Socioeconomic drought occurs when the demand for water exceeds supply as a result of a weather-related shortfall in water supply.

### 2.1.2 Sequence of types of drought [4]

The impacts associated with different types of drought further emphasize their difference. The agricultural sector is usually the first to be affected when drought occurs, because of its heavy dependency on stored soil water. During extended dry periods, soil water can be rapidly depleted. If dry periods continue to extend, then sectors of activities dependent of other sources of water will begin to feel the effects of the shortage. Those relying on surface and subsurface water for example, are usually the last to be affected. A short-term drought that persists for 3 to 6 months may have little impact on these sectors, depending on the characteristics of the hydrologic system and water use requirements. When precipitation returns to normal, the sequence is repeated for the recovery of surface and subsurface water supplies. Soil water reserves are replenished first, followed by streamflow, reservoirs and lakes, and lastly ground water. Drought impacts may diminish rapidly in the agricultural sector because of its reliance on soil water, but linger for months or even years in other sectors dependent on
2.2 Drought monitoring

2.2.1 Introduction

The United Nation Environmental Program (UNEP) with the World Bank reported [22] that about about 180 million people in Africa live in drought prone areas and 50 million people are threatened to starve in the case of rain failure. They reported also that during the period of 1965 to 1999, 330 droughts occurred in Africa in total and caused 880,000 deaths. The understanding of drought and the use of appropriate methods to extract information contained in the existing large volume of data, are important in the processes of monitoring drought; These processes involve in general monitoring precipitation, soil moisture, evapotranspiration, biomass and water levels [3]. Each process is briefly introduced below:

- Drought originates from a deficiency of the amount of precipitation. Methods for monitoring precipitation often make use of meteorological satellites data to detect clouds and cloud temperatures; Precipitation is estimated based on Cold Cloud duration. Another method is based on interpolating rain gauge measurements or ground-based rain radar data.

- Monitoring soil moisture is performed using Radar data. When the soil is dry, the plant does not receive enough water for its growth and if the conditions persist, the plants dry.

- Monitoring evapotranspiration is performed from energy balance with thermal remote sensors. Evapotranspiration is directly linked to plant growth. A reduction of evapotranspiration leads to loss of growth. This process depends also on the stage and type of crop.

- Monitoring biomass is performed through the use of indices such as the Normalized Difference Vegetation Index (NDVI). Biomass reduction affects the plant growth and its measures can be used as a drought indicator.

- Monitoring water levels is usually performed with Digital Elevation models (DEM). A reduction of surface or subsurface water levels can be an indicator of drought.
2.2.2 Remote sensing indices used in drought monitoring

Looking at the sequence of drought impacts, we observe that vegetation condition is the first feature on earth surface to be affected by drought; hence, we can conclude that it constitutes an interesting parameter to monitor drought from its impacts, at early stage using remote sensing images. RS indices have been developed and used for the past decades to monitor drought from vegetation response; among them we can list:

1. The Normalized Difference Vegetation Response (NDVI). First suggested by Tucker in 1979 as an index of vegetation condition, it is calculated as:

\[ NDVI = \frac{\lambda_{NIR} - \lambda_{RED}}{\lambda_{NIR} + \lambda_{RED}} \]

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

It is the most commonly used vegetation index and reflects vegetation vigor, percent green cover and biomass. NDVI is a nonlinear function that varies between -1 and +1, and undefined when NIR and RED are zero. NDVI values for vegetated land generally range from about 0.1 to 0.7, with values greater than 0.5 indicating dense vegetation and values lesser than 0.1 indicating no vegetation (barren area, rock, sand or snow). The main limitations of NDVI are that:

(a) It uses only two bands and is not very sensitive to influences of soil background reflectance at low vegetation cover;

(b) It has a lagged response to drought because of a lagged vegetation response to developing rainfall deficits due to residual moisture stored in the soil. Previous studies have shown that NDVI lags behind antecedent precipitation by up to 3 months. The lag time is dependent on whether the region is purely rainfed, fully irrigated, or partially irrigated. The greater the dependence on rainfall the shorter the lag time [9].

2. The vegetation Condition Index (VCI), also developed by Kogan, shows how close the NDVI of the current month is to the minimum NDVI calculated from the long-term record. The VCI is calculated as the following:

\[ VCI_j = \frac{(NDVI_j - NDVI_{min})}{(NDVI_{max} - NDVI_{min})} \times 100 \] (2.1)

where \( NDVI_{max} \) and \( NDVI_{min} \) are calculated from the long-term record for that month (or week) and \( j \) is the index of the current month (week) [24]. In the context of estimating degrees of drought severity, VCI values below 50% are considered. The VCI value close to zero percent reflects an extremely dry month, when the NDVI value is close to its long-term minimum. Low VCI values over several consecutive time intervals point to drought development.
The Water Supplying Vegetation Index ($WSVI$) has been used to monitor drought related vegetation stress in China [15] using NOAA satellites data. It is based on the fact that, in drought conditions, the $NDVI$ values derived from satellite data fall below normal. At the same time, the crop canopy temperature as measured by the same satellite rise above normal. Both effects are related to available water supply, and both measures can be combined in one index to obtain a sensitive measure of drought conditions. It is calculated as:

$$WSVI_j = \frac{NDVI_j}{BT_j}$$ (2.2)

where $j$ is the current observation and $BT$ is the brightness temperature measured on the thermal band (4) of NOAA-AVHRR sensor.

### 2.2.3 Monitoring drought from NDVI

The $NDVI$ by itself does not reflect drought or non-drought conditions. But the effects of severity of drought on vegetation can be defined as the deviation of current $NDVI$ values from their corresponding long-term mean $NDVI$ values. For each location, this deviation ($DEV_{NDVI}$) is calculated as the difference between the $NDVI$ of the current time step and its corresponding long-term mean.

$$DEV_{NDVI} = NDVI_{i,m} - NDVI_{\text{mean},m}$$ (2.3)

Where $NDVI_{i,m}$ is the current $NDVI$ value $i$ for each location for time step $m$ and $NDVI_{\text{mean},m}$ is the long-term mean $NDVI$ for the corresponding calendar time step $m$ for each location. When $DEV_{NDVI}$ is negative, it indicates below-normal vegetation conditions and therefore suggests prevailing drought conditions; The greater the negative departure, the greater the magnitude of drought severity. In general, the departure from the long-term mean $NDVI$ is effectively more than just a drought indicator, as it would reflect the conditions of healthy vegetation in normal and wet months/years. This indicator is widely used in drought studies.

### 2.3 Study Area

#### 2.3.1 East Africa

East Africa is a vast region of Africa that encompasses many countries, usually divided geographically into three subregions. The first, the Great Lakes Region, includes Uganda, Kenya, Tanzania, Rwanda, and Burundi. The second, the Horn of Africa, includes Ethiopia, Eritrea, Sudan, Djibouti, and Somalia.
Chapter 2. Literature review

And the last subregion is the *Indian Ocean islands* of Comoros and Seychelles. These distinctions are made based on different types of vegetation, availability of water, and topography in the three subregions. The vegetation of East Africa is very diverse as seen on Figure 2.3. The Horn of Africa has more desert and semidesert regions and the Great Lakes Region has more forests and woodlands. Both subregions have some mountains, savanna land, and steppe. Lakes and rivers constitute main sources of water availability in East Africa. However, there are some subregions of East Africa with a lot of lakes and rivers and others where they are scarce. The other source of water availability is rainfall. East Africa experiences its rainy seasons in boreal spring and autumn, centered around April–May and October–November; the spring rains being more abundant and the autumn rains more variable [8]. However, as is the case in other regions of Africa, rainfall has declined overall in East Africa since the 1960’s. This has caused periodic drought and famine [13]. Figure 2.2 shows the general map of East Africa. Figure 2.3 shows the land cover of Africa in the year 2000.
2.3. Study Area

2.3.2 Drought in the East Africa

East Africa is a region prone to extreme climate conditions such as drought and floods. Drought has severe negative impacts on key socioeconomic sectors such as agriculture production, which is the mainstay of economic activities in the region. For that reason, the development of early warning systems to monitor and mitigate drought impacts has become an important issue [1].

The US National Aeronautics and Space Administration (NASA) Earth Observatory (EO) reported that for the whole East Africa, 2005 was anything but a normal year. The long rainy season (April-May) produced little rain, and the short rainy season (October-November) failed altogether. Severe drought affected the region during the end of year 2005 and the beginning of year 2006 and the Famine Early Warning System Network (FEWSNET) reported the situation whereby a serious food security concern was raised; The drought has resulted in crop failure, pasture degradation and water shortage, and by the end of January 2006, millions of people were in need of food aid in the region. Figure 2.4 [29] [20] shows drought affected area in East Africa. Drought severity is indicated with colors from dark brown (severe drought) to yellow (less severe drought) in December 2005 and in February 2006. Figure 2.5 shows tents in a camp for persons displaced by the drought in the town of Wajid in the southern Bakol Region, Somalia. Two main aspects motivated our the choice for the East Africa region as the study area:
Figure 2.4: Drought in East Africa: December 2005 and February 2006. From darker brown: severe drought to yellow, less severe drought

Figure 2.5: Situation in Somalia, October 2005 (photo from UNICEF)
2.4 Dataset description

- The region is mainly rain-fed; vegetation conditions depend heavily on rainfall. In this case, drought affects vegetation conditions with a short lag time.

- Remote sensing images are of great importance in monitoring natural phenomena at the East Africa regional level as it can facilitate data collection and processing routines, as well as obtaining results in an improved timely manner. Some organizations and projects such as the Famine Early Warning System (FEWS) and the Preparation for the Use of Meteosat Second Generation in Africa (PUMA) project, exploit or promote the exploitation and development of RS based methods to monitor drought in African regions.

2.4 Dataset description

The primary set of data used in this study is composed of eighteen \( NDVI \) images acquired from the FEWS Africa Data Dissemination Service (FEWS/ADDS). These images are freely downloadable via the FEWS/ADDS website \(^1\). The \( NDVI \) product is derived from the data collected by the National Oceanic and Atmospheric Administration (NOAA) satellites, and processed by the Global Inventory Monitoring and Modeling Studies group (GIMMS) at the NASA. Positive \( NDVI \) values have been extracted to track vegetation conditions and those values have been linearly stretched from 0 to 255. Water and cloud pixels are valued 255, and bad \( NDVI \), 253. The pixel size of each image is \( 8\text{Km} \times 8\text{Km} \). Maximum value compositing has been used to compute 10-days \( NDVI \) for each pixel location. Eighteenth of the 10-days \( NDVI \) images have been considered and subsets have been created, for the period of December 1\(^{st}\) 2005 to February 28\(^{th}\) 2006 for our study: nine of them are current maximum 10-days compositing \( NDVI \) images, as seen on Figure 2.6 and nine others are their corresponding long-term mean compositing \( NDVI \) images, as seen on Figure 2.7. The long-term mean \( NDVI \) images are calculated for the period of 1982 to 2005.

Table 2.1 summarizes the characteristics of the input data set. Each subset image is of size (363 rows \( \times \) 313 columns).

\(^1\)http://earlywarning.cr.usgs.gov/adds/datatheme.php
Chapter 2. Literature review

Figure 2.6: Current NDVI images
Figure 2.7: Corresponding long-term mean NDVI images
Table 2.1: Data characteristics

<table>
<thead>
<tr>
<th>Rows/Columns Coordinated system description</th>
<th>Pixel size</th>
<th>Data value parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of rows:</strong> 363. <strong>Number of columns:</strong> 313</td>
<td>Projection: Albers Equal Area Conic</td>
<td>x-direction: 8000m. y-direction: 8000m</td>
</tr>
</tbody>
</table>
2.4. Dataset description
Chapter 3

Methodology

Methods have been developed and applied to monitor drought from vegetation response on RS images. Most of them monitor vegetation stress intensity has an indicator of drought; i.e. high intensity of stress is associated with drought. In reality, the duration of low vegetation stress intensity on a location over a long period of time, can also cause drought; hence constitutes an indicator of drought. In our approach, we propose methods to monitor drought from NOAA-AVHRR derived $NDVI$ images, considering not only the intensity of vegetation stress on a location, but also its duration over a period of time. The proposed approach makes use of image mining techniques to address the issues of large dataset processing. Fuzzy set theory is used to address the issue of uncertainties related to the definition of drought characteristics monitored from vegetation response.

The present chapter describes all the steps involved in the proposed drought monitoring approach. It is divided into three main parts: The first part provides with an overview of the concepts used in developing the approach, namely: data mining, in section 3.1, spatial data quality, in section 3.2, fuzzy set theory, in section 3.3, and finally object versus pixel based analysis, in section 3.4. The second part of this chapter describes the methods used in the proposed monitoring approach. The first section 3.5 of this second part presents the methods used to analyze the images, to select the function that characterizes drought on the images and to define drought objects. The other section 3.6 of this part presents the methods use to apply the function on the images, extract and track drought objects, and finally the methods used to analyze the results on the space-time dimension. Figure 3.4 shows the overview of the proposed drought monitoring approach and the sequences of tasks. Section 3.8 describes the methods used to validate the results obtained. The last part of this chapter is the section 3.9 that lists the software packages used in this study.
Figure 3.1: Overview of the proposed RS drought monitoring approach
Chapter 3. Methodology

3.1 Data mining

Data mining can be defined as the analysis of often large observational data sets to reveal hidden relationships and summarize the data in novel ways that are both understandable and useful to the users (modified from [5]). Data mining techniques are developed in the fields where huge amount of data are collected and need to be processed. Although recently named as such, the history of data mining goes back to 1970’s by the development of many experts system applications mainly in the fields of medicine and defense [17]. Nowadays, these techniques are being more and more adopted in remote sensing processing methods, due to the advancements in remote sensing data collection technologies. Data mining is a field developed by encompassing principles and techniques from statistics, machine learning, numeric search and scientific visualization to accommodate the new types of data types and data volumes being generated [11]. Although the data mining tasks vary, the underlying principle remains the discovery of unknown information from large data sets. Three major technological factors are at the origin of the development of data mining techniques [17]:

- The growing amount of collected data led to the development of mass storage devices.
- The problem of accessing huge amount of data has led to the development of advanced and improved processors.
- The need for automating the tasks involved in data retrieval and processing has led to the development of advanced statistical and machine learning algorithms.

This section briefly introduces the concepts of and presents the existing data mining models and techniques. It elaborates also on data mining related to remote sensing images, while discussing the application on mining drought.

3.1.1 Data mining models and techniques

A model is by definition a high-level, global description of a data set [5]. They can be subdivided into two categories: descriptive and predictive. The first category is used for summarizing the data in a convenient and concise way while the predictive models allow to make statements about the population from which the data were drawn or about likely future values [11]. Data mining methods often ignore the appropriateness of model of the data, namely the goodness of fit [17]. In order to find the best model in a given class of models, it is important to determine the class of models that best fits the data [18], and in order to determine appropriate class of models for specific data, it is important to understand the data. The table 3.1 briefly list some of the common data existing mining tasks and their description.
Table 3.1: Common types of data mining tasks and techniques (Modified from Sankar et.al.(16))

<table>
<thead>
<tr>
<th>TASK</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Association rule discovery</td>
<td>Describes association relationship among different attributes. The origin of the association rules is in market basket analysis. The \textit{a priori} algorithm provided one early solution which was improved by subsequent algorithms.</td>
</tr>
<tr>
<td>Clustering</td>
<td>Maps a data item into one of several clusters.</td>
</tr>
<tr>
<td>Classification</td>
<td>Classifies data item into one of several predefined categorical classes. Used for predicting data in several fields such as, in atmospheric data mining.</td>
</tr>
<tr>
<td>Sequence analysis</td>
<td>Models sequential patterns, like time-series data. It aims at modeling the process of generating the results or to extract and report deviation and trends over time.</td>
</tr>
<tr>
<td>Regression</td>
<td>It maps a data item to a real-valued prediction variable. It is used in different prediction and modeling applications.</td>
</tr>
<tr>
<td>Summarization</td>
<td>It provides with a compact description for a subset of data. Summarization functions are often used in interactive data analysis, automated report generation and text mining.</td>
</tr>
<tr>
<td>Dependency modeling</td>
<td>Describes significant dependencies among variables.</td>
</tr>
</tbody>
</table>
3.1.2 Data mining in remote sensing

Data mining in RS image databases is similar to automated image processing. Their difference lies on the fact that in mining case, often a very large amount of data has to be processed, compared to a single or a few images in image processing. Image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the images. Image mining is more than just an extension of data mining to image domain. It is an interdisciplinary endeavor that draws upon expertise in computer vision, image processing, image retrieval, data mining, machine learning, database, and artificial intelligence. Remote sensing images present characteristics such as dependency within observations, data uncertainty, non-linearity, non-stationary, high levels multivariate and space-time dependency make them different than other type of data. For this reason, common mining methods, originally designed for relational data structure and image data can not be used for remote sensing images.

Since data mining is a data-driven analysis approach, issues like the nature and quality of data are of great importance. In the case of remotely sensed data, more issues such as the quality, in terms of efficiency, of processing algorithms and the capacity of storage devices are of concern, as remotely sensed data can be of hundreds of terabyte. E.g., in the case of monitoring changes; current and long-term (can vary from 20 to 100 years) time period data is needed.

3.1.3 Data mining for drought monitoring

Drought is a complex phenomenon to study. Its complexity is due to the followings:

- It combines information from different sectors: not only climatic issues are considered but their impacts on the vegetation on the ground, water, etc. It also involves social factors such as the population, and economic factors such as crop production in an area.

- It is a spatio-temporal phenomenon: its severity is based on how long the conditions remain unfavorable and is highly dependent on specific regions.

- Huge amount of data is needed: drought often affects large regions, and/or during a long period of time; hence a large amount of collected and archived data is often required to understand it.

- Multi-scale: Sometimes, a combination of larger and smaller scales data is required (ex: Administrative land use maps vs continental land cover maps)
3.2 Spatial data quality

- Multi-resolution: a combination of higher and lower resolution data can also be required (e.g., data from a coarse satellite to be combined with high resolution land use data)

- Multi-temporal data: data required is collected at time interval such as NDVI data to monitor vegetation health.

- Multi-sensors data: data required is collected by different sensors; e.g., temperature data by thermal sensors and soil data collected by radar sensors.

- Multi-dimensional data: 2D and 3D data are required; 2D land cover images to monitor land cover changes, combined with Digital Terrain Models (DTM) to integrate the contribution of the terrain on drought.

With all these aspects, it is timely impossible to study drought from RS images and ground observations non automatically. Data mining techniques have already been explored in the context of monitoring drought. T. Tadesse et.al [23] proposed a data mining approach to monitor vegetation stress due to drought, over the U.S. Central Plains. They proposed an integration of satellite, climate, and biophysical data and developed the Vegetation Drought Response Index (Veg-DRI) model. Sharma A. [19] proposed a spatial data mining method to monitor drought using temporal NDVI and rainfall in India. The VCI, calculated from NDVI was overlaid with rainfall data to identify drought areas.

3.2 Spatial data quality

With current technology as an enabler, society’s reliance on more precise and accurate spatial data is rapidly increasing. Interest by mankind in accurate and precise data, including spatial data has always been high, but interest in high quality data has been especially pronounced. Today, especially in the medical and military sciences, practitioners routinely demand ever more exact data quality. Fortunately recent advances in electronic technology enable practitioners to meet this increasing demand. Spatial data has not been exempted from these trends and demands for increased spatial data quality are apparent. The elements of spatial data quality, as defined by the International Cartographers Association (ICA) [12] commission, are described below:

**Lineage** means reporting a description of the source material from which the data were derived, including all transformations involved in producing the final digital file. The lineage is mainly aimed at ensuring quality of metadata for appropriate future use.

---

1 For more information: [http://www.ibimet.cnr.it/Case/MTPprogramme/data/doc/4ClimateAnalysisTools/DataMiningApproach.pdf](http://www.ibimet.cnr.it/Case/MTPprogramme/data/doc/4ClimateAnalysisTools/DataMiningApproach.pdf)
Chapter 3. Methodology

Positional accuracy is still a major concern in the electronic age and constitute the second element of spatial data quality. A positional accuracy report should include the degree of compliance to the spatial registration standard. It must also consider the effect of the quality of all transformation performed on the data, and report the result of any positional accuracy testing performed on the data.

Attribute accuracy can be interpreted as the ability to measure attributes class accurately at location, or for a well-defined feature. A report on attribute accuracy shall include the degree of compliance to the attribute definition.

Completeness is defined as the relationship between the objects represented and the abstract universe of all similar objects. The quality report shall include information about selection criteria, definitions used and others relevant mapping rules. In particular, the report shall describe the exhaustiveness of a set of features.

Logical consistency describes the number of features, relationships and attributes that have been correctly encoded in accordance with the integrity constraints of the feature data specification. The report on logical consistency shall describe the fidelity of relationships encoded in the data structure of the digital spatial data.

Semantic accuracy describes the number of features, relationships, or attributes that have been correctly encoded in accordance with a set of feature representation rules. Semantic accuracy refers most to the pertinence of the meaning of the geographic object rather than to its geometrical representation.

Temporal accuracy is the seventh element of spatial data quality. Temporal information describes the date of observation, the types of updates and validity periods of spatial spatial data.

3.3 Fuzzy set theory

The Fuzzy set theory has been introduced for the first time by Prof. Zadeh Lofti, in 1965. It provides a conceptual framework for solving knowledge representation and classification in an ambiguous environment, where traditional crisp set theory could not be applied. The main difference between crisp and fuzzy sets can be characterized by means of a membership function [25]. In crisp sets, the membership function can only output 2 values (yes, no) or (0, 1). This means that an element of a crisp set can only belong to one group, for which he has a membership grade of 1. The fuzzy set theory softens this constraint and allows for partial membership such that an element of a set may simultaneously hold several non-zero membership grades for different groups. In brief, the fuzzy set theory allows greater flexibility than the crisp approach.
3.4. **Pixel vs object based RS analysis**

3.3.1 **Fuzzy sets**

A *fuzzy set* is a set whose boundaries are characterized by transition zones. To be able to define a *fuzzy set*, values of threshold and dispersion as well as a selection of an appropriate membership function is required.

Let \( S \) represent a universe of discourse composed of elements denoted \((s)\). A *fuzzy subset* \( G \) of \( S \) is determined by a membership function \( \mu_G \), which assigns a membership grade with \([0,1]\) to each element \( s \). The membership grade can be expressed by [25]:

\[
\mu_G : [0,1] 
\]

(3.1)

Let \( \mu_G(s) \) denotes the grade of membership of \( s \) in a fuzzy subset \( G \) that can be expressed as:

\[
G = \sum_i \mu_G(s_i) / s_i 
\]

(3.2)

and in the continuous case, \( G \) becomes

\[
G = \int \mu_G(s) / s ds 
\]

(3.3)

where the symbol ‘/’ represents the link between the value of \( s \) and its corresponding membership grade \( \mu_G(s) \) in the fuzzy subset \( G \).

3.3.2 **Types of Fuzzy membership functions**

There exists different types of membership function. Some examples are *boolean*, *bell-shaped*, *triangular*, *Gaussian*, *trapezoidal*, or a function with a central core region and upper and lower transition zones with different width. Figure 3.2 illustrates a few examples of types membership function types. The *triangular* shape type of membership function has a single core value, and linear dispersion slopes; while the *bell-shaped* type of membership function has curved dispersions. The *trapezoidal* shape type of membership function has an interval of values on its core region. The parameters of these membership functions can be selected with statistical methods or based on field knowledge.

3.4 **Pixel vs object based RS analysis**

In remote sensing fields, a pixel-based analysis takes into account spectral information to process the images. For example in image classification process,
Figure 3.2: Types of fuzzy membership functions
a pixel based approach results in pixels associated to classes or clusters, based on their bands characteristics. This approach is suitable for some applications such as soil mapping or DTM generation, but is not applicable for applications such as land use mapping, where other information than spectral characteristics is needed. In brief, we can say that pixel-based analysis takes into consideration the physical properties of the observations. A pixel value can often be a mixture of spectral reflectance of different elements (vegetation, soil, water, etc), which together give a mixed reflectance value for that pixel. When assigning pixels to classes, mixed pixels cause problems. A common method of solving this problem is the fuzzy classification, which assigns to pixels membership values for the pre-defined classes.

An object-based analysis considers groups of pixels, which may or may not have similar spectral reflectance, representing a pre-defined object of interest. RS objects 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. Objects can represent physical features on the earth (building blocks, corn field, water body, etc,) when classified based on texture, hence can be regarded as physical objects. The concept of object can also be extended to non physical features on the ground, hence can be called 'virtual geographic objects'. The 'virtual geographic objects' can be defined as measurements having geographical information but do not represent physical features on the earth; rather, their definitions are based on some attributes of physical features. 'Virtual geographical objects', are for example: air pollution, drought or any other event. Objects are either crisp or vague. Their vagueness can be related to their location or their definition and can depend on observation scale. E.g a tree is a crisp object with well defined boundaries while a forest, which is a group of trees, have vague boundaries: there is no crisp location to identify the beginning of a forest. Some virtual objects present also transition zones at their boundaries and threshold values are often required for their definitions. Segmentation methods are mostly used to generate crisp objects, based on texture information. Methods for handling vague objects within a GIS mostly use the principle or triangulation to address the gradual spatial changes of membership function values of the transition zone.

### 3.5 Data analysis

Data Analysis is an important stage of research projects. It can be defined as the transformation of data to extract useful information. There are two approaches for data analysis: the qualitative approach and the quantitative approach. Both of them have been used. The qualitative approach adopted here involved visual interpretation of images and the quantitative approach involved numerical measurements.
Visual image interpretation is the most intuitive way to extract information from remote sensing images. In this context we applied human vision methods such as spontaneous recognition, e.g. recognizing water bodies as we know their location on the images. Interpretation elements have also been used, e.g. tone to recognize areas affected most than others. The output of the analysis helped us understanding drought on the images.

3.5.1 Understanding drought from images

As described in section 2.2.3, NDVI by itself does not indicate drought or non-drought. But the effects of severity of drought on vegetation can be defined as the deviation of current NDVI values from their corresponding long-term mean NDVI values. For each location, this deviation ($DEV_{NDVI}$) has been calculated in order to characterize drought from the images.

$$DEV_{NDVI}: R^2 \rightarrow NDVI_{range}$$

Each long-term mean NDVI image has been subtracted from its corresponding current NDVI image as per the formula 2.3. The resulting images constituted the inputs to our mining tasks.

Drought related uncertainties

Drought, observed from vegetation stress using ($DEV_{NDVI}$) values, refers to a fact about vegetation conditions; This fact is an interpretation of quantitative measurements: drought is characterized on images from ($DEV_{NDVI} < 0$) values. But the severity of the effects of drought on vegetation depend on the degree of the negative departure from 0 and on the duration of drought conditions over a period of time. Measurements of drought parameters are subject to the inaccuracies of measuring instruments, of definition of a class, and of data models. The definition of the class ‘drought’ is vague: there exists no crisp point of measurement of when one can start talking about drought; This attribute uncertainty leads to the positional uncertainty of drought extent. More over, the measurements used in this study — (NDVI and $DEV_{NDVI}$) — contain uncertainties in the way that they are measures in the form of maximum and sum of observations, respectively, on locations.

Function definition

The fuzzy set theory has been applied to model the uncertainties of the effects of drought on vegetation, related to the degree of departure from 0. The followings have been considered to select the type of membership function to be used:

1. Only ($DEV_{NDVI} < 0$) values range has been considered as the range of interest to quantify the effects of drought on vegetation.
2. The slope of the dispersion of the function is assumed to be linear.
3.6 Mining Tasks

The core of function is expected to be an interval of \((\text{DEV}_{\text{NDVI}})\) values.

4. The lower and upper limits of the transition (TW) width can be adjusted by the experts, based on field knowledge or statistical methods.

Considering the above mentioned statements, a one-sided, trapezoidal-shaped membership function, as defined by the equation 3.4 has been found the most appropriate to be used in this study.

\[
\mu(x) = \begin{cases} 
1 & \forall x : \text{min} \leq x \leq a \\
\frac{b-x}{b-a} & \forall x : a \leq x \leq b \\
0 & \text{otherwise}
\end{cases}
\]  

(3.4)

Figure 3.3 illustrates the shape of the selected membership function.

3.6 Mining Tasks

This section explains the tasks performed to mine drought from the NOAA-AVHRR derived NDVI images. The mining tasks proposed in this approach were selected based on their capabilities to provide with relevant information to monitor drought from RS images, from vegetation stress intensity and duration. Figure 3.4 illustrates the individual processes and expected output of the mining tasks developed. All the mining tasks were automated to facilitate the processing. The first task was the application of the membership function on the images. This process allowed to characterize drought over space based on vegetation stress intensity. The second process, extraction, allowed to facilitate the tracking processed of drought objects. Two tracking processes, namely the tracking of drought objects based on their intensity at locations, and the tracking of drought objects based on their temporal continuity at locations, were developed. The output of the two tracking processes constituted the knowledge gained to perform a space-time analysis. The space-time analysis allowed to characterized drought severity based on intensity and duration. The following sections explain each mining process in details.
Figure 3.4: Overview of the mining tasks and expected outcomes
3.6. Mining Tasks

Application of the membership function on images

The application of a membership function on a remote sensing image with \( n \) spectral bands, for \( k \) number of classes, results in \((n \times k)\) number of output images. The fuzzy representation of geographical classes in a layer with each pixel having its own value give rise to the following kind of fuzzy partition matrix

\[
\begin{pmatrix}
\mu(x_1) & \mu(x_2) & \ldots & \mu(x_p) \\
\mu(x_1) & \mu(x_2) & \ldots & \mu(x_p) \\
\vdots & \vdots & \ddots & \vdots \\
\mu(x_1) & \mu(x_2) & \ldots & \mu(x_p)
\end{pmatrix}
\]

(modified from Wang [27])

Where \( p \) is the number of pixels, \( k \) is the number of predefined classes and \( \mu \) is the membership function.

In our case, \( n = 1 \) and \( k = 1 \).

On the \( DEV_{NDVI} \) images, drought severity has been characterized as a function \( \eta \) of \([0,1]\) on \( R^2 \) defined:

\[
\mu : DEV_{NDVI_{range}} \rightarrow [0, 1] \quad (3.5)
\]

\[
DEV_{NDVI} : R^2 \rightarrow DEV_{NDVI_{range}} \quad (3.6)
\]

\[
\eta : drought_{sev} = \mu \circ DEV_{NDVI} \quad (3.7)
\]

\[
\eta : R^2 \rightarrow [0, 1] \quad (3.8)
\]

The function implementation has been automated. The automated algorithm was designed to apply the function on all the images specified, and the expert user was prompted to enter the values of the dispersion width limits.

3.6.1 Extracting process

The extraction process depends on how drought objects have been defined. In remote sensing, an object often refers to a group of pixels having similar textural information. The object can then be defined based the texture and extracted using segmentation techniques. An object can also be defined based on spectral characteristics. Fire is an example; pixels with high spectral reflectance in the thermal bands of MSG-SEVIRI sensor were identified as possible fire objects in a study conducted to model forest fire in Spain [17]. In our
case, a similar approach has been adopted, due to the fact that the spectral characteristics of the images constituted the only basis for identifying drought from its impacts on vegetation conditions; Pixels with $\eta(x) > 0$ values have been identified as possible drought objects. These objects could have been grouped together to constitute a “region object” for example, but in our case, we kept our definition of object at the lowest information level: pixel-level objects. An algorithm has been automated to extract all possible drought objects from the images. Information about all the pixels $x$ with values $\eta(x) > 0$ were extracted and recorded in a file. The algorithm automatically generates a file and record the following information for each drought objects: location, membership value, and corresponding image identifier.

### 3.6.2 Tracking process

Tracking drought objects constituted an important process in order to determine drought characteristics, based on vegetation stress severity and duration. In this study, we devised two ways of tracking drought objects: based on location, we first tracked intensity, and second we tracked temporal continuity. The first tracking process has been developed to provide with relevant information that would allow to understand the influence of vegetation stress intensity on characterizing drought severity. The second tracking process has been developed to provide with relevant information in order to understand the influence of temporal continuity of objects on locations, in the process of characterizing drought severity. In the first process, we provided the system with locations of interest and the system returned corresponding drought objects in each image. In the second process, all the pixels presenting a minimum temporal continuity of the whole period of study were tracked.

### 3.7 Space-time analysis

A space-time analysis of geographical data is an analysis performed on a four dimensional space — three are spatial coordinates $(x,y,z)$ and one is temporal coordinate — in which any event or physical object is located. The patterns resulting from the mining tasks constituted the knowledge gained and this knowledge was used to understand the spatial and temporal characteristics of drought. The questions considered for the analysis were:

1. What are the impacts of the definition of drought on its space-time characteristics?

2. What is the impact of the vegetation cover in characterizing drought?

In order to give an answer to the first question, five different locations were provided to the system and their drought objects on each image were returned.
3.8 validation

On each image, the drought object value was used as an indicator of drought severity at the selected locations. The results were used to analyze the space-time characteristics of drought. Drought objects have also been tracked on a temporal continuity basis. The results obtained were used to study the spatio-temporal characteristics of drought. The two observations were then compared to understand the influence of the definition of drought on its space-time characteristics. To answer the second question, the observations obtained from the answers to the first questions were analyzed based on the vegetation cover type of the region.

3.8 validation

All the results obtained from the space-time analysis of drought were compared with the reports describing drought on the region and the land cover map of East Africa. The reports did not provide with any quantitative information, hence the validation was done by visualization.

3.9 Software packages used

The software packages used for this study are:

ILWIS 3.3 was used for visual exploration of original images and to compute the deviation $DEV_{NDVI}$ maps. The ‘map calculation’ built-in function was used to perform this task.

ArcGIS/ArcMap has been used for visual exploration of image outputs.

PYTHON programming language has been used to automate the mining tasks. The core functions implemented can be found in Appendix.
Chapter 4

Results

This chapter presents the results obtained from all the processes involved in the proposed drought monitoring approach. They are presented and discussed against the traditional crisp approach in four sections. Section 4.2 describes and discusses the results of the processes implemented to characterize drought from vegetation response over space. Section 4.3 describes and discusses the results of the extraction process and the section 4.4 describes and discusses the results obtained from the tracking methods developed. Finally, the section 4.5 describes and discusses the results of the space-time analysis.

4.1 Understanding drought from the images

In the process of understanding drought from the images, we first calculated the deviation values. Figure 4.1 shows the images obtained. On these images, we observed a range of negative to positive deviations. We assumed that the region is mainly rain-fed and that the vegetation stress is caused by prevailing drought conditions. From these assumptions, we stated that pixels showing negative deviation values on the images are drought affected pixels. Figure 4.2 illustrates the variation of the total number of positive, zero and negative ($DEV_{NDVI}$) pixels on the nine deviation images through time and their difference. The number of pixels with values ($DEV_{NDVI} = 0$) remains more or less constant through time. From the fourth dekadal, we observe a jump of the total number of ($DEV_{NDVI} < 0$) pixels from around 4,000 to more than 5,000.

4.2 Characterizing drought over space

The deviation values computed constituted the inputs to characterize drought based on vegetation conditions. The trapezoidal membership function selected was applied on all the images.
4.2. Characterizing drought over space

Figure 4.1: NDVI deviation images
In order to select the dispersion width limits, we proceeded as following:

- We first fixed the upper limit $b$ to 0. This was motivated by the assumption made in defining the function based on the negative range of deviation values $3.5$.

- In the previous section, it is also mentioned that threshold values to determine drought from the negative departure of a normal condition is often made arbitrary than on the basis of scientific knowledge. Field knowledge is then applied in the process of selecting a threshold value. In our case, since no field knowledge was available, the lower limit $a$ has been selected somehow arbitrary, based on the visualization of the histograms of the deviation images and on the reports of drought occurrence in the region. The reports $[29]$ $[20]$ mentioned severe drought occurrence for the period of December 2005 to February 2006. We expected then in all the images, a significant amount of drought affected pixels, but with different intensities, as illustrated in the figure $2.4$. We then performed a sensitivity analysis by varying the value of $a$, and comparing the results with the description. We selected somehow arbitrary the value of $a = -40$ to test our method.

Figure 4.3 shows the histograms of the nine deviation images. The reports had limited accuracy, in a way that they described drought at particular time interval different from the time interval of the remote sensing images used.
4.2. Characterizing drought over space

Figure 4.3: Histograms of deviation images
More over, they were in form of newsletters, with no quantitative measures. Figure 4.4 illustrates the results acquired after applying the membership function on each of the nine images, with the values of $a = -40$ and $b = 0$. We observe from the images severe drought on the south-eastern region of Lake Victoria; this pattern is constant on all the images. The distribution of drought severity intensity on an image is gradual in most cases; we observe however abrupt changes of severity intensity at locations such as near water bodies or near pixels originally with missing values. Other abrupt changes may be explained by the presence of non-vegetated area (bare soil) in drought affected areas. In order to proceed with the study, we assumed the followings:

- Vegetation stress observed from the images after applying the membership function is caused by prevailing drought conditions.
- The variation of the intensity of vegetation stress observed reflects drought severity intensity variation.

In the traditional approach, where the value ($-40$) was taken as the drought threshold value, less drought coverage has been observed on the images, with no variation of intensity. The use of a membership function allowed to characterize drought on space, taking into consideration its gradual dispersion.

### 4.3 Extracting objects

Extraction of drought on the images was performed at pixel-level, as by the description of drought objects. The extraction function characterized drought objects as pixels $x$ with values $\eta(x) > 0$. The automated function extracted a total of 415,603 drought objects from all the nine images. The results were stored in a table. Each drought object extracted was recorded with its corresponding drought intensity value. Table 4.1 shows the first lines of the table storing extracted drought objects information. The result of the extraction was used for further processing. This algorithm, used in a crisp classification process of drought, extract a total of 105,376 drought objects, all with the same intensity level, $1$. We observed that with the crisp approach, a total of 310,227 objects were left out of the analysis; hence the crisp approach lost information about a huge number of objects that can potentially be used in the process of studying drought. The use of a membership function showed that more objects are participating in the process of drought modeling.

### 4.4 Tracking objects

The tracking processes developed aimed at allowing us to make use of the different intensity values of the objects, in the process of monitoring drought. In
4.4. Tracking objects

Figure 4.4: Images obtained after applying the defined membership function with $a = -40$ and $b = 0$. Drought membership values range from 0.025 (lowest value) to 1 (highest value)
Table 4.1: Table of extracted drought objects

<table>
<thead>
<tr>
<th>Image ID</th>
<th>x-location</th>
<th>y-location</th>
<th>Membership value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.075</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.15</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0.15</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0.025</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0.15</td>
</tr>
<tr>
<td>0</td>
<td>11</td>
<td>0</td>
<td>0.25</td>
</tr>
<tr>
<td>0</td>
<td>14</td>
<td>0</td>
<td>0.15</td>
</tr>
<tr>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0.025</td>
</tr>
<tr>
<td>0</td>
<td>18</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>0</td>
<td>19</td>
<td>0</td>
<td>0.15</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>0</td>
<td>0.025</td>
</tr>
<tr>
<td>0</td>
<td>22</td>
<td>0</td>
<td>0.125</td>
</tr>
<tr>
<td>0</td>
<td>23</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>0</td>
<td>24</td>
<td>0</td>
<td>0.225</td>
</tr>
<tr>
<td>0</td>
<td>26</td>
<td>0</td>
<td>0.125</td>
</tr>
<tr>
<td>0</td>
<td>28</td>
<td>0</td>
<td>0.25</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
developing the tracking processes of drought, we described two approaches: tracking based on intensity values at specified locations and tracking based on temporal continuity. In the first approach, five different locations were selected, on vegetated area. The results of drought objects tracked at each location are presented in different tables. Table 4.2 shows the output of the first tracking algorithm applied on location (83, 180). The algorithm returned high valued drought objects for the selected location from the fourth dekadal to the last one. The same algorithm has been applied to the second selected location (71, 122) and the results are shown on Table 4.3. The algorithm returned drought objects for the selected location during all the dekadals with the highest valued ones at fifth dekadal and the last dekadal. Table 4.4 shows the results of the tracked objects at location (44, 230). No drought object was tracked from the fifth dekadal to the eighth dekadal. Table 4.5 shows the results of the tracked objects at location (131, 202). The algorithm returned drought objects for all the dekadals, with the highest valued ones from the third dekadal to the last one. Table 4.6 shows the results of the tracked objects at location (166, 204). The algorithm returned low valued drought objects ($\eta(x) < 2$) for all the dekadals.

Figure 4.5 compares the results obtained from the tracking processes applied on the five location, with the maps reporting drought in East Africa, for the third and seventh dekadals.

Considering drought defined based on objects intensity levels only, from the
Table 4.4: Tracked objects at location (44, 230)

<table>
<thead>
<tr>
<th>Image ID</th>
<th>Membership value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.375</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td>0.175</td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Table 4.5: Tracked objects at location (131, 202)

<table>
<thead>
<tr>
<th>Image ID</th>
<th>Membership value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.675</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
</tr>
<tr>
<td>4</td>
<td>1.0</td>
</tr>
<tr>
<td>5</td>
<td>1.0</td>
</tr>
<tr>
<td>6</td>
<td>1.0</td>
</tr>
<tr>
<td>7</td>
<td>1.0</td>
</tr>
<tr>
<td>8</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 4.6: Tracked values at location (166, 204)

<table>
<thead>
<tr>
<th>Image ID</th>
<th>Membership value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td>0.15</td>
</tr>
<tr>
<td>5</td>
<td>0.15</td>
</tr>
<tr>
<td>6</td>
<td>0.2</td>
</tr>
<tr>
<td>7</td>
<td>0.175</td>
</tr>
<tr>
<td>8</td>
<td>0.15</td>
</tr>
</tbody>
</table>
4.4. Tracking objects

Figure 4.5: Tracked objects variation over time for the five locations
results, looking at the third and seventh dekadals, we observed that location 1 is not affected by drought on the first dekal, and locations 2, 3, 4, 5 are affected, with low intensity though. On the seventh dekal, locations 1, 2 and 4 showed to be respectively highly affected by drought compared to the first dekal, while location 5 remained affected with a low intensity. Location 3 showed not to be affected by drought. The results were compared with the drought maps and we observed a correlation between the trend of the variations obtained from the results and the ones observed on the maps: for the location 1, 2, 4 the severity increased and for the location 3, it decreased. The location 5 revealed a contrasting observation, though presenting a constant trend on both the results and the maps: on the map, it was shown to be severely affected by drought, while on the result, it showed less intensity of severity for both dekadals. This observation led us to conclude that the level of intensity itself is not sufficient to estimate drought severity at a location and at a particular time, based on the assumption that the results as well as the maps are accurate.

Note that with the crisp approach in this context of characterizing drought based on intensity level at a location, a different scenario was presented; For the first dekal, all the location were shown not to be affected by drought and on the seventh dekal, locations 3 and 5 were shown not to be affected by drought. This result was valid only for location 1 at the first dekal and for location 3 at the seventh dekal, after comparison with the reports. The result obtained from the crisp approach was found to be less accurate than the results obtained using a fuzzy approach.

The second tracking function was developed to track drought objects on locations based on their temporal continuity during the period of the nine dekadals. It has been developed to introduce the concept of temporal continuity of intensity, in the process of monitoring drought. The algorithm was automated to track all drought objects that are continuous at corresponding locations in all the images, for a minimum of nine dekadals. A total of 170,028 drought objects were tracked over time for all the images. Table 4.7 shows the first lines of the table generated by this second tracking algorithm. Figure 4.6 shows the distribution of drought objects (black points) tracked based on their temporal continuity, on the study area. Drought objects on the pre-selected locations 4 and 5 were also tracked by the second algorithm, as shown on Figure 4.7. Location 4 was identified with 3 successive low intensity drought objects and 6 successive high intensity drought objects. Location 5 was identified with all 9 drought objects, all of low intensity \((\eta(x) < 2)\); By defining drought severity based on temporal continuity, we concluded from the results that locations 4 and 5 were affected by drought during all the dekadals. This result was compared with the maps and the two locations showed to be affected by drought at corresponding dekadals (third and seventh). From these results, we observed that the temporal continuity of drought objects, can result in severe drought. The observation at location 5 indicated that the temporal continuity of low intensity drought objects on a location, can result in severe drought. In a crisp classification approach, this observation could not be made as location 5 was
4.4. Tracking objects

Figure 4.6: Drought coverage, based on temporal continuity
Chapter 4. Results

Table 4.7: Table of tracked objects based on their temporal continuity

<table>
<thead>
<tr>
<th>x-location</th>
<th>y-location</th>
<th>Image ID</th>
<th>Membership value</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>1</td>
<td>0.075</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>2</td>
<td>0.075</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>3</td>
<td>0.15</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>4</td>
<td>0.05</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>5</td>
<td>0.325</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>6</td>
<td>0.2</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>7</td>
<td>0.1</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>8</td>
<td>0.075</td>
</tr>
<tr>
<td>57</td>
<td>0</td>
<td>0</td>
<td>0.175</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

not identified as being affected by drought during all the dekadals.

4.5 Space-time analysis

The knowledge gained from the results of the previous tasks were used to perform a space-time analysis to further understand drought characteristics. In our study, we looked at answering the following questions:

1. What are the impacts of the definition of drought on its space-time characteristics?

2. What is the impact of the type of vegetation cover in characterizing drought?

Answers to these questions can help understanding the space-time characteristics of drought severity, given its definition and representation. To answer the first question, we first calculated the area covered by drought on each image, using the fuzzy and the crisp definitions of drought. An algorithm to compute the polygon areas per image was implemented and applied on each image of the two approaches. Table 4.8 shows the results of the area covered by drought in Km$^2$ on each image, using the fuzzy definition of drought. Figure 4.8 illustrates on the images the total area covered by drought, using a fuzzy approach. Table 4.9 shows the areas covered by drought using a crisp approach, with a the threshold value of $-40$ to characterize drought. In both cases, we observe variations of the sizes of the area covered by drought. By comparing them, we observe a significant difference in corresponding images; the areas computed from
4.5. Space-time analysis

Figure 4.7: Tracked objects based on temporal continuity on locations 4 and 5

Figure 4.8: Drought coverage area using a fuzzy approach
Chapter 4. Results

Table 4.8: Drought coverage with fuzzy approach

<table>
<thead>
<tr>
<th>Image ID</th>
<th>Area in Km²</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2 374 144</td>
</tr>
<tr>
<td>1</td>
<td>2 310 464</td>
</tr>
<tr>
<td>2</td>
<td>2 542 016</td>
</tr>
<tr>
<td>3</td>
<td>3 461 504</td>
</tr>
<tr>
<td>4</td>
<td>3 357 760</td>
</tr>
<tr>
<td>5</td>
<td>2 947 136</td>
</tr>
<tr>
<td>6</td>
<td>3 185 024</td>
</tr>
<tr>
<td>7</td>
<td>3 343 424</td>
</tr>
<tr>
<td>8</td>
<td>3 077 120</td>
</tr>
</tbody>
</table>

Table 4.9: Drought coverage with crisp approach

<table>
<thead>
<tr>
<th>Image ID</th>
<th>Area in Km²</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>55 744</td>
</tr>
<tr>
<td>1</td>
<td>46 464</td>
</tr>
<tr>
<td>2</td>
<td>45 888</td>
</tr>
<tr>
<td>3</td>
<td>1 418 560</td>
</tr>
<tr>
<td>4</td>
<td>1 051 584</td>
</tr>
<tr>
<td>5</td>
<td>715 136</td>
</tr>
<tr>
<td>6</td>
<td>936 896</td>
</tr>
<tr>
<td>7</td>
<td>1 189 952</td>
</tr>
<tr>
<td>8</td>
<td>963 392</td>
</tr>
</tbody>
</table>

The crisp definition of the function are smaller, hence indicate smaller drought coverage extent. This result shows how the estimation of the extent variation over space and time of drought can be affected by its definition. In the case drought is wrongly defined and characterized on images, the calculation of its extent results in wrong estimations, hence, misleading the model quantifying the variation of the extent of drought impacts, in a monitoring system.

The definition of drought does not only impact the model quantifying the change but also the model qualifying and characterizing the direction of change of drought in space and time. In a crisp approach, drought is defined based on drought objects intensity values only. In the proposed approach, using a membership function to define drought, drought is defined not only based on intensity values but on temporal continuity of intensity values also. This approach allows more expansion of severe drought in more possible directions than on the crisp approach, over space in time.

The direction of the expansion of the severity of drought from the third dekadal to the seventh dekadal, observed from our results of defining drought severity from vegetation stress intensity level, and from the maps were not similar. Figure 4.9 illustrates drought severity distribution from the maps and on the images obtained from our analysis. We notice a difference in the patterns of
4.5. *Space-time analysis*

![Figure 4.9: Drought severity expansion: Results vs. Reports](image)

Figure 4.9: Drought severity expansion: Results vs. Reports
the direction of expansion of drought severity in both corresponding dekadals. The images obtained from our analysis showed highest drought severity more in the south-east, around Tanzania, on the first image, and an expansion on the eastern direction on the second image, around Kenya. On the maps reporting drought occurrence in the East Africa for the period of study, high severity is indicated more around southern Somalia and southern Ethiopia, and expended on all directions, on the seventh dekadal. This led us to investigate the contribution of the characteristics of the types of land cover, to understand this difference.

In order to achieve this, we first compared each map and image resulted from our analysis, with the land cover map of Africa. We observed that area highly vegetated showed on our images higher drought severity intensity
4.5. Space-time analysis

Figure 4.11: Drought severity in February: Results vs reports
(around Kenya and Tanzania). Area with sparse vegetation (around southern Somalia and southern Ethiopia) showed on our images less drought severity intensity but high drought severity intensity on the maps from the reports. This observation is understandable as our function is based on NDVI values, hence, the presence of vegetation on the study area.

This brought us to observe that the definition of drought based on stress intensity levels, works only for vegetated area. For areas with sparse vegetation, such as the semi-arid region of Somalia and Ethiopia, the use of intensity level only was less effective. Drought, defined based on temporal continuity of vegetation stress intensities, constituted a better approach to monitor its impacts on the semi-arid regions. Hence, the combination of the two methods to quantify, qualify and model the direction of drought severity in space and time, is required for regions such as East Africa, with greatly diversified vegetation cover types.
4.5. Space-time analysis
Chapter 5

Discussions

This chapter discusses the proposed drought monitoring framework as a whole. The first section briefly discusses each step involved in the process of monitoring drought. The second section discusses the issues to consider for an adaptation to study other phenomena, presenting uncertainties in their definitions and extent.

5.1 On extending for complete drought study

- **Selection of parameters and their indices**: in this study, vegetation itself has been considered, using NDVI values, to monitor drought from remote sensing images. Drought has been identified on the images but a combination of other parameters contributing in drought occurrence along with the selection of more specific indices, depending on the type of drought under study, should be considered.

- **Selection of images and their time interval, bands and integration of non-RS information**: In our study, a limited number of NOAA-AVHRR derived NDVI images as been used (December 2005 to February 2006), to test the methods and the coarse spatial resolution of the images ($64\text{Km}^2$) More images and better spatial resolution of images is recommended.

- **Selection of appropriate function to characterize drought**: In this study, a one-sided trapezoidal shaped membership function has been selected. Other more appropriate function could also be used, after detailed analysis of drought in the region. The selection of the function can be based on field knowledge or exploratory analysis. One of the limitation of the fuzzy set theory applied in this context of drought monitoring, is that the core values of the drought membership function should be defined in a crisp way. This issue is not discussed here as it is a general aspect of the fuzzy set theory.
5.2. On extending for other vague phenomenon

- **Definition and construction of vague objects**: In this study, a pixel-based analysis was performed due to the lack of both knowledge of appropriate level of information required and methods to define drought objects at higher level, construct and use them. The pixel-level definition of objects could better be replaced by a higher object-based level, for facilitating processing tasks and interpretation of results.

- **Extracting algorithm**: In this study, the pixels with positive drought membership function were extracted and stored in a table. This algorithm can be modified to extract different levels of objects definitions' information. The use of techniques to handle vague objects are suggested to be considered at this level.

- **Tracking algorithm**: The performance of the first algorithm to track objects based on intensity was satisfactory. But the second algorithm can be enhanced by specifying more tracking criteria for temporal continuity; The consideration of the variation of intensity and the sequence of their occurrence in time (ascendant, descendent, missing values etc) should help improving the results.

- **Space-time analysis**: The application of the knowledge gained was used to answer some questions. However, the knowledge gained was not sufficient to perform further analysis such as the influence of other parameters (wind, temperature, precipitation, etc) on drought characteristics. The algorithm developed to compute the area covered by drought on images, is based on polygon area calculation, hence inherits its limitations.

5.2 On extending for other vague phenomenon

In the context of applying the proposed approach to study another vague phenomenon such as air pollution for example, the developed architecture can be used as a framework, given the following:

- **Selection of index or parameters to consider**: In the definition of a vague object, is it important to identify the parameters describing the phenomenon. Some vague phenomena are a result of a combination of factors. It is important to identify these factors and to quantify them. The type of phenomenon will lead to the selection of appropriate observations or indices to use.

- **Selection of images, bands and integration of non-RS information**: The selection of images, bands and other non-RS should be performed with necessary knowledge to do so. After the selection, often the data, from different sources, must be transformed to be compatible. The errors or uncertainties resulting from the transformations should be known, quantified if necessary, and considered in the next processing steps.
• **Selection of the appropriate function to characterize the phenomenon**: An appropriate data model should be made in order to ensure the completeness quality of the data model. Appropriate functions can be selected based on field knowledge or on data exploration.

• **Definition and construction of vague objects, often a combination of parameters**: The definition of the object and the construction of the objects are main characteristics of the ideal mining process proposed, to ensure efficient extraction and tracking processes of objects of interest.

• **Extracting algorithm**: The proposed extracting algorithm can be used in other applications, given that the objects are defined at required level of information.

• **Tracking algorithm**: In our study, the tracking algorithms definitions were related to drought monitoring. Similar tasks can be found in other applications, but the algorithm can be improved by the settings of characteristics specific to the application field.

• **Space-time analysis**: Given the results of the extracting and tracking algorithms, space-time analysis can be performed, but the tasks should be modified depending on the phenomenon under study. The variation of the size of a given vague spatio-temporal phenomenon can be of interest. But is has to be considered also that the algorithm is based on polygon area calculation.
5.2. On extending for other vague phenomenon
Chapter 6

Conclusions

6.1 Conclusions

The objectives of this study have been achieved and research questions answered. An approach, using image mining techniques to monitor drought from remote sensing images, has been developed and implemented. It aimed at monitoring drought as a vague phenomenon as characterized by the severity and duration of vegetation stress. The drought that occurred in East Africa for the period of December 2005 to February 2006 has been taken as a case study. NOAA-AVHRR derived NDVI images have been used.

The proposed approach included the processes of drought definition, drought characterization, drought objects extraction and tracking. Most of the processes have been automated using Python programming language to facilitate the processing of multiple images. In the first process of the proposed approach, drought has been defined based on vegetation stress as the deviation of the current NDVI values from their corresponding long-term mean values. The NOAA-AVHRR derived NDVI images have been processed to derive deviation values at each location. A trapezoidal membership function has been defined to characterize drought on space, assuming a linear dispersion of its impacts on vegetation. In this study, a drought object has been defined at the pixel-level in each image as being a pixel having a positive membership value. A total of 415,603 objects have been extracted and stored in a table. Two tracking algorithms have been developed and implemented to track drought objects and the results were used as inputs to characterize drought based on intensity and duration of vegetation stress. The first algorithm aimed at tracking drought objects at selected locations, based on intensity and the second algorithm was developed to track drought objects at locations, based on their temporal continuity.
From the results, it has been observed that drought intensity itself successfully characterized drought on regions with relatively high vegetation cover density. The use of temporal duration characteristics of extracted objects allowed us to characterize drought on regions of sparse vegetation cover. The proposed approach showed to be more effective than the traditional crisp approach, to characterize drought on diversified vegetated area.

6.2 Recommendations

For further studies on monitoring drought using this approach, we recommend to first define drought impacts on vegetation condition based on field knowledge in order to ensure the completeness of the data model. We also recommend the acquisition of ground truth for quantitative validation of the results.

We have learnt that the use of a membership function effectively characterize drought in regions with diversified vegetation cover. However, the function defined and used assumed a linear dispersion of drought impacts on vegetation. We recommend further studies in order to select a function type based on scientific knowledge about the relationship between drought and vegetation stress, on regions of interest.

Further work is also recommended in order to identify other factors contributing to the occurrence of drought and to integrate them into the model. Depending on the data to be integrated, the processes should be adapted.
Bibliography


Appendix A

Main functions automated

This section contains only the core parts of the code written to automate the tasks, using python.

A.1 Applying membership function

def membership( a, b):
    ###############Images previously converted into txt files#########
    #####################READ ALL IMAGES#################################
    while True:
        file_to_read= image_input
        thisimage= readfile( file_to_read)
        if not thisimage:
            break
    #######################CONVERT IMAGES TO LISTS#########################
    image= []
    for foo in range( len( thisimage)):
        imageline= thisimage[foo].split( )
        for m in range( len( imageline)):
            imageline[m]= float( imageline[m])
        image.append( imageline)
    images.append( image)

    ############APPLY MEMBERSHIP FUNCTION ON IMAGE####################
    for i in range( len( images)):
        image= images[i]
        pixel_values= []
for j in range( len( image)):  
    for k in range( len( image[0])):  
        pixel_value = image[j][k]  
        pixel_values.append( pixel_value)  
minimum= min( pixel_values)  
for j in range( len( image)):  
    for k in range( len( image[0])):  
        x= image[j][k]  
        if minimum<= x and x<= a:  
            images[i][j][k]= 1  
        elif a<x and x<b:  
            images[i][j][k]= ( b- x)*1.0/( b- a)  
        else:  
            images[i][j][k]= 0

##########WRITE RESULTS TO FILE################

for i in range( len( images)):  
    image= images[i]  
    outlines= []  
    for j in range( len( image)):  
        row= image[j]  
        outline= ''  
        for k in range( len( row)):  
            pixel= row[k]  
            outline+ = str( str( pixel))+ " "  
            outline+ = "\n"  
        outlines.append( outline)  
    writefile( outlines, "membership"+ str( i)+ ".txt")  
print "End.\n"  
return

A.2 Extraction of drought objects

def extract( ):  

    ################READ IMAGE FILES################

print "extract drought objects"  
images= []  
while True:  
    file_to_read= image_input  
    thisimage= readfile( file_to_read)
A.3 Tracking of drought objects

A.3.1 Tracking intensity

def track():
    while True:
        col= input( "x- location:")
        if col= = - 1:
            break
        row= input( "y- location:"
A.3. Tracking of drought objects

print "Values for x- location", col, "y- location", row
print "Time\tValue"
for image in range( len( images)):
    value= images[image][row][col]
    if value>= 0.025:
        print image, "\t", images[image][row][col]
    else:
        print "None"
return

A.3.2 Tracking temporal continuity

def track_cont( ):
    images= []
    outlines= []
    counter= 0
    len_images= len( images)
    row= len( images[0])
    col= len( images[0][0])

    for j in range( row):
        for k in range( col):
            pixel_values= []
            for i in range( len_images):
                pixel= images[i][j][k]
                pixel_values.append( pixel)

            if 0.0 not in pixel_values:
                outline= str( k)+ "\t"+ str( j)+ "\t"+ str( i)+ "\t"+ str( pixel_values[i])
                outlines.append( outline)

    ######WRITE RESULTS TO FILE################################
    writefile( outlines, "tracked.txt")
    return