PAN AFRICAN UNIVERSITY INSTITUTE FOR BASIC SCIENCES TECHNOLOGY AND INNOVATION

Use of Earth Observation Data and Artificial Neural Networks for Drought Forecasting: Case Study of Narumoro Sub-Catchment.

By

Joseph Mutuga Kigumi

A thesis submitted to Pan African University in partial fulfillment of the requirement for the degree of Master of Science in Civil Engineering (ASAL and Environmental Management option)

September 2014

DECLARATION

This thesis is my original work and has not been presented for a degree/diploma/certificate in any other University.

……. ……

Signature **Date**

Joseph Mutuga Kigumi

CE300-003/2012

This thesis has been submitted for examination with my approval as University Supervisors

Prof. M.O. Nyadawa

Jaramogi Oginga Odinga University of Science & Technology

…… …….

Signature **Date**

Prof. P.G. Home

Jomo Kenyatta University of Agriculture and Technology

…… …….

Signature **Date**

ACKNOWLEDGMENT

I wish to acknowledge all those who made this possible; the African Union for the scholarship through Pan African University, and the PAUISTI fraternity. I also acknowledge my supervisors Prof. M.O. Nyadawa of Jaramogi Oginga Odinga University of Science & Technology and Prof. P.G. Home of Jomo Kenyatta University of Agriculture and Technology for guiding me through the research and my wife Mary and our son Jayden who patiently supported me, bearing many hours of my absence as I prepared this work. Above all I acknowledge God, without whom none of this would have been.

DEDICATION

This work is dedicated to my son Jayden Kigumi. May he grow up to be a fine man who loves God and his country.

ABSTRACT

Droughts are a major problem in Kenya especially in the Arid and Semi-Arid Lands (ASALs) where they are frequent and causes a great deal of suffering and loss. Drought monitoring and forecasting requires extensive climate and meteorological data which is usually largely missing in developing countries or not available in the required spatial and temporal resolutions. Use of the readily available remotely sensed an alternative to observed data for drought monitoring, is faced by many challenges as to the utility and its applicability at the local sub-catchment level.

This study examined the use Tropical Rainfall Measuring Mission (TRMM) precipitation estimates in meteorological drought monitoring alongside stream flow modelling for drought identification using Artificial Neural Networks (ANN). Monthly TRMM data was downscaled from its original 0.25° x 0.25° resolution to 1km x 1km resolution using NDVI (Normalized Differential Vegetation Index) from the SPOT VEGETATION program. Using TRMM and observed data meteorological droughts were identified by SPI (Standardized Precipitation Index), and hydrological drought using the threshold method. Both downscaled and original resolution TRMM were found to detect similar meteorological drought pattern and number of drought months as the observed data. TRMM data was found to be suitable for streamflow modeling since the identified hydrological drought pattern was similar to the precipitation pattern. ANN was found to be effective in modelling TRMM-streamflow relationship where the model was able to reproduce a similar flow pattern as the observed streamflow at a 6 month lead time. However ANN model was found to underestimate high flow peaks and overestimate low flows. The best TRMM-streamflow relationship model was found at 1 month lag with correlation coefficient of 0.79 and regression coefficient of 0.87, the least performance was at 6 month lead time where the correlation coefficient was 0.53 and R^2 of 0.43. For medium forecasting e.g. 3 month lead time, time lagged TRMM was found to produce better results that using a combination of TRMM and flow. This is based on R^2 of 0.72 and 0.54 for TRMM only input and TRMM + flow input respectively. It was concluded that TRMM whether downscaled or at the original resolution can be used to monitor

meteorological drought in Narumoro sub-catchment, and that ANN can be used to simulate flow for hydrological drought forecasting using TRMM as input with similar skill to that of ground observed data sets.

TABLE OF CONTENTS

List of Tables

List of Figures

LIST OF ACRONYMS

ANN	-	Artificial Neural Network
SPI	-	Standardized Precipitation Index
TRMM	-	Tropical Rainfall Measuring Mission
DEM	-	Digital Elevation Model
UNEP	-	United Nations Environmental Program
UNCCD	-	United Nations Convention to Combat Drought and Desertification
ASAL	-	Arid and Semi-Arid Land
SRTM	-	Shuttle Radar Topography Mission
NGA	-	National Geospatial-Intelligence Agency
NASA	-	National Aeronautics and Space Administration
SMAC	-	Simplified Method for Atmospheric Corrections
FDC	-	Flow Duration Curve

CHAPTER 1: Introduction

1.1. Background

Droughts are extreme climate events occurring over land and persisting over periods of months or years. They are temporary and can occur in any climatic region even in the wet and humid zones, and are characterized by below average precipitation (Dai, 2010). In addition to the precipitation deficit that has persisted over long period of time, other factors like temperature, high winds, low relative humidity, characteristics of the rainfall such as onset, duration, intensity and distribution, also play a significant role in the occurrence and characteristics of droughts (Mishra and Singh, 2010).

It is difficult to have a precise definition of droughts due to the different hydro meteorological variables, socioeconomic factors and also the water use needs in different regions. The definitions can be conceptual which involves using relative terms like a drought is a long dry period or operational where an attempt to define the drought characteristics like onset, severity and termination of the drought period is done (Mishra and Singh, 2010; Belal et al., 2012).Many definitions therefore exists depending on who is defining or why the drought is being defined. However the UN Convention to Combat Drought and Desertification (UNCCD) definition of droughts as: “that naturally existing phenomena that exist when the precipitation has been significantly lower than the recorded levels leading to hydrological imbalances that adversely affects the land resources producing systems” (Mishra and Singh, 2010) seem to be all inclusive incorporating even the impacts of the drought events. WMO (2006) emphasizes the need to differentiate droughts from seasonal aridity which is a well-defined dry season and aridity which is a permanent feature of climate.

This study investigated the hydrological and meteorological droughts in Narumoro sub-catchment. It adopted a definition of meteorological drought as occurring when the received precipitation is less than a preset threshold, similarly, hydrological droughts are taken to be when the stream flow is less than a threshold level.

Owing to the definition problems stated above, droughts are usually categorized according to the variable impacted for easier understanding and study. These categories are meteorological (characterized by deficit, less than normal precipitation), agricultural droughts (less moisture such that crop growth or its output is hindered), hydrological droughts (resulting into less stream flow or ground water levels falling) and economic droughts (when water deficit hinders some economic activity or social welfare) (Dai, 2010; Mishra and Singh, 2010; UNEP, 2006). Other literature suggests new droughts including ground water droughts and vegetative droughts.

Droughts have several impacts which can be termed as direct and indirect impacts. Direct impacts includes reduced crop, range land, and forest productivity; increased fire hazard; reduced water levels; increased livestock and wildlife mortality rates; and damage to wildlife and fish habitat. Indirect impacts include reduced income for farmers and agribusiness, increased prices for food and timber, unemployment, reduced government tax revenues because of decreased expenditures, increased crime, foreclosures on bank loans to farmers and businesses, migration, and disaster relief programs (Wilhite et al., 2007). Droughts cause untold suffering to communities hence it is important to study them and improve drought monitoring.

Kenya and especially its Arid and Semi-Arid Land (ASALs) which constitute about 90% of its total land surface, suffers from recurring droughts that have continued to retard development. Kenyan ASALs are poorly developed compared to the rest of the country a fact that can be directly attributed to the frequent and devastating droughts that this lands suffers.

Droughts are some of the hardest of all natural hazards to quantify, analyze and monitor, first because they have slow onset and progress slowly and over large areas. The impacts due to droughts are only realized long after the droughts have already set in, it is therefore important to have a good drought monitoring system to detect occurrence and quantify the drought events. This will allow for proper drought mitigation actions to be effected. Droughts indices have been used to monitor and quantify droughts, their intensity and severity. Some of the indices used includes Palmer drought severity index (PDSI), standard drought index

(SDI), and standard runoff index (Dai, 2010).These indices are data intensive and requires long term data of high quality and sufficient temporal - spatial resolution. However most regions heavily impacted by droughts especially in the developing nations like Kenya, such data is largely missing or unavailable in the desired density thus greatly hindering the computation, application and usefulness of such indices in the regions where they are most needed (Duan & Bastiaanssen, 2013). Setting up and operating monitoring stations (e.g. rain gauges and river gauging stations) is expensive and it is impossible to set up the observation stations in sufficient density as would be required for effective drought monitoring. An alternative to such indices and data is therefore needed remotely sensed data provides an alternative for observed data, this data is available for most regions in the world and at good spatial and temporal resolution (Vicente-Serrano et al., 2012).

One of the key advantage of remotely sensed data, and in this case satellite observed data is the readily accessibility of the data as most of the data sets are distributed free of charge. The other advantage is that they are available for the entire (or most) of the world rendering them very appropriate for regional studies and monitoring, they are also available in very fine temporal resolutions including real-time. The disadvantages of these data is that they are available in coarse spatial resolutions that renders them unsuitable for local studies, most of the method of acquiring the satellite data involves indirect observation of the phenomena rather than direct measuring of the phenomena. These disadvantages require the satellite data to be downscaled to resolutions good enough for local study and to be compared to or calibrated with ground observed data. For example, among the most widely used precipitation data is data obtained from the Tropical Rainfall Measuring Mission (Funk et al., 2010) which is available in spatial resolution of 0.25° x 0.25° which is quite coarse catchment level study.

Drought monitoring and prediction can also be done by applying physical models which depend on catchment characteristics. Most of the models are hugely data intensive, requiring many parameters, and long term historical data for calibration and validation. These models, for example. Global Circulation Models are expensive to acquire, difficult to learn and needs highly skilled and specialized

personnel to operate. As such they are mainly used by large international organizations to carry out drought monitoring on regional scale. They are inappropriate and difficult to apply to small catchments. Theoretical models including statistical models are what is mainly used to model droughts in small catchments. These theoretical models require some fore knowledge about the catchment for them to be applied appropriately. This fore knowledge is sometimes inaccurate and inadequate leading to inaccurate results. Artificial Neural Networks (ANN) are a new type of theoretical models which are able learn the relationship between parameters within a system without requiring intensive data or fore knowledge of the system being modelled (Imrie et al., 2000). The application of ANN have been contested by some researchers while others have reported positive feedback on their application.

It is important for studies to be done that show the utility and appropriateness of applying remotely sensed data and ANN in drought monitoring. TRMM data is a fairly new technology having been launched in 1998, it is important that these technologies are applied in as many different areas as possible to compare their performance.

1.2. Statement of the Problem

There is a challenge of monitoring droughts due scarcity of data and that the available solution of using satellite observed data requires downscaling and verification with the ground observed data what is normally referred to as “ground truthing”. Downscaling of TRMM data has been done for some large catchments, but very few studies have been done for small catchments. Also the downscaling of TRMM data has been done mostly for comparison with rain gauge data and very few compares drought indices computed from rain gauge data with those from TRMM. It is important to check the utility of TRMM data in drought monitoring and prediction, especially so since TRMM is a new technology having been introduced in 1998 and having more case studies in its usage will go a long way in improving its use in drought monitoring and in hydrological modeling among other uses.

ANN has not enjoyed a wide spread usage in drought monitoring and prediction as have other methods like physical modelling and regression analysis. Many questions still arise ANN utility considering its “black box model” nature. It is there important to examine the quality of hydrological model that can be developed using ANN and relating that to drought modelling and prediction.

1.3. Objectives

1.1.1. Main Objective

To develop and evaluate a hydrological drought early warning Artificial Neural Network (ANN) model based on Tropical Rainfall Measuring Mission rainfall estimates.

1.1.2. Specific Objectives

1. To evaluate the utility of Tropical Rainfall Measuring Mission precipitation data to characterize meteorological drought as compared to rain gauge observed rainfall data in the Narumoro sub-catchment.
2. To characterize hydrological drought in the Narumoro sub-catchment
3. To develop and evaluate hydrological Artificial Neural Network drought prediction model for Narumoro sub-catchments based on Tropical Rainfall Measuring Mission precipitation data.

1.4. Research Questions

1. Is Tropical Rainfall Measuring Mission precipitation estimates data capable of identifying meteorological droughts to the same skill as rain gauge observed data in the Narumoro river sub-catchment?
2. What are the characteristics of the meteorological drought in the Narumoro river sub-catchment?
3. Can Artificial Neural Networks model adequately the hydrological drought conditions in the Narumoro river sub-catchment using Tropical Rainfall Measuring Mission precipitation estimates data?

1.5. Justification

Kenya frequently suffers from droughts that greatly affect the economy and the livelihoods of her people, the loss of life and property can be prevented if an early warning system that is reliable and easy to use is in place. This study is a major

step towards ensuring that such a system can be set up by providing the prerequisite information. Narumoro sub-catchment was selected due to availability of both rainfall and streamflow data within a small sub-catchment. The sub-catchment is also located in rainfall scarce area which suffers from droughts hence increases chance of drought identification during the study.

There have been many successful cases on use of ANN and earth observation data in predicting droughts. There is no documented study that have been done in the Narumoro sub-catchment to evaluate the usefulness of such a technology. This is in spite of the fact that Kenya is not only water scarce but also suffer from very frequent and often devastating drought events. This study is crucial in making such a determination which could result into a better approach to drought management in the Kenya and other developing countries.

There are also several criticism to ANN use in hydrology as compared to physical models (Imrie et al., 2000). This study will add to body of knowledge available in this area hence helping to shed light on the usefulness and the challenges of applying ANN and earth observation data in drought prediction.

1.6. Scope

The study limited its analysis to within the Narumoro River sub catchment area. The skill and the usefulness of the developed model was determined through statistical parameters like regression coefficient (R2). The study concentrated on hydrological drought analysis and identification based on the threshold method and development of the ANN model. For the purpose of downscaling of TRMM, satellite data was collected from a region much larger than the study area so as to get enough data points for regression analysis.

CHAPTER 2: Literature Review

2.1.Introduction

They are two ways of defining droughts, conceptual and operational. Conceptual definitions are non-analytical observation of the indicative phenomena like weather conditions usually resulting into statements like “it is very dry this year”. Operational definitions, on the other hand, attempt to identify the onset, severity, and termination of drought periods. Generally operationally defined droughts can be used to analyze drought frequency, severity, and duration for a given return period Droughts are generally classified into four categories including 1) Meteorological droughts, 2) Hydrological droughts, 3) Agricultural droughts, and 4) social-economic droughts (Mishra and Singh, 2010).

Meteorological drought is as result of having precipitation below long term mean for a period of time over a region. Agricultural drought results when there is a decline in the available soil moisture as compared to the long term mean and the consequent crop failure. This decline in the soil moisture could be a result of reduced precipitation or other factors like increased precipitation. Hydrological droughts results when the available water resources (ground and surface) goes below the amount required to meet the various water needs. It can also be described as the scenario where the available water resources goes below the long term mean (Mishra and Singh, 2010; Tallaksen et al., 2009; Dai, 2010). There is generally a causal relationship between the various drought categories, with all of them tracing their origin from precipitation deficit.

Generally hydrological drought is manifested in the reduction of the available water resources for example reduction in streamflow, fall in reservoirs heights, and decrease ground water depth to below normal levels, this reduction can however result from many other reasons rather than just reduction in precipitation. The causes of low flow can be classified as natural causes for example reduction in precipitation, increase in temperatures resulting into higher evaporation, and changes in the precipitation patterns or human influenced causes like increased demand of the water resources, and land use changes resulting into changes in the runoff/infiltration rates. While studying / monitoring hydrological droughts, it is

important to distinguish between periods of low flows which are normal, river regime changes over the seasons and lowest river flow in the year may be higher than a seasonal drought flow.

2.2.Drought Identification and Monitoring

Drought identification and monitoring has been achieved through the use of drought indices which are used to identify and quantify drought occurrence. These indices can be defined as prime variables for assessing the effect of drought and defining various drought parameters which includes intensity, severity, duration and spatial extent (Mishra and Singh, 2010). Some of the commonly used drought indices includes: 1) Palmer Drought Severity Index (PDSI), Palmer Hydrological Drought index (Palmer,1995) , Percent of Normal, Crop Moisture Index (CMI), Standardized Precipitation Index (SPI), Evapotranspiration Deficit Index, Water Requirement Satisfaction Index, Soil Moisture Anomalies, Soil Moisture Deficit Index, NDVI (Mishra and Singh, 2011). Normalized Difference Vegetation index (NDVI) (Caccamo, et al., 2011), 2011), VTCI (Vegetation Temperature Condition Index), Vegetation Condition Index (VCI). In this study SPI was used to identify and characterize meteorological droughts while the threshold method (or Theory of Runs) was used to characterize hydrological droughts.

2.2.1. Standardized Precipitation Index (SPI)

SPI is one the most commonly adopted drought index across the world because it is easy to compute and use, suitable for drought comparison across different regions and time spans due its standardization. Due to its probabilistic nature, SPI is adequate for carrying out drought risk analysis (Serinaldi et al., 2009). It was adopted by World Meteorological Organization as the recommended index to be used all over the world for drought monitoring (Hayes et al., 2010). The index is computed by fitting a probability distribution on aggregated monthly precipitation series (e.g. k = 3, 6, 12, 24 months), and by computing the corresponding non-exceedance probabilities and standard normal quantiles, the latter defined as the SPI series (Serinaldi et al., 2009). When the SPI values are continuously negative, then the period of negative values is drought period, (Mishra and Singh, 2010); (Serinaldi et al., 2009). Other than just identifying dry periods, SPI is also applied

to classify wet periods. Table 1 shows categorization of climatic conditions using SPI, probabilities ΔP associated with drought class are also given.

Table 1: Wet and drought period classification according to the SPI index (Serinaldi et al., 2009)

Index value	**Class**	**Probability**	**ΔP**
SPI ≥2.0	Extremely wet	0.977–1.000	0.023
1.5 ≤ SPI < 2.0	Very wet	0.933–0.977	0.044
1.0 ≤ SPI < 1.5	Moderately wet	0.841–0.933	0.092
-1.0 < SPI < 1.0	Near normal	0.159–0.841	0.682
-1.5 < SPI ≤ -1.0	Moderate drought	0.067–0.159	0.092
-2.0 < SPI ≤ -1.5	Severe drought	0.023–0.067	0.044
SPI ≤ -2.0	Extreme drought	0.000–0.023	0.023

Severally, when the SPI value is less than -1, a drought period is identified. This is because SPI values between +1 to -1 are taken to be near normal conditions and therefore not defined as drought periods. This approach has been adopted by several authors including (Serinaldi et al., 2009),

2.2.2. Threshold Drought Index

A threshold value can be used to classify the drought condition. The use of a threshold to identify and characterize drought was introduced by Yevjevich (cited by Pandey et al., 2008) and is also known as the theory of runs . The selection of the threshold value depends on various factors including the type of deficit been investigated, and nature of the river. It may reflect a well-defined flow quantity like mean, proportion of the mean, percentile or a specified reservoir yield. For perennial rivers, the threshold is usually set as 70-90 percentile of the flow, this threshold for an intermittent and ephemeral rivers with may be zero and therefore certain level of exceedance is usually adopted based on a monthly flow duration curve (Pandey et al., 2008). The threshold can be set to a constant throughout the period e.g. Tallaksen et al. (2009) who used a time invariant but spatially variant

threshold or can be set to be variable varying with the seasons e.g. depending on water use needs (Nyabeze, 2004). The varying threshold level is adapted to detecting low flows during both the high and low flows, low flows however may not be necessarily drought flows as some seasons records low flows. Lows flows detected by varying threshold may be referred to as stream flow variations rather droughts.

Based on this theory of runs, a drought period is defined as the period when the flow is below the threshold level (Tallaksen et al., 2009), crossing the threshold level signifies onset or cession of drought event, drought length (Duration is the time period between two consecutive crossing), and the area below the threshold level gives the total deficit (Motalleb et al., 2014; Smakhtin, 2001). The threshold drought identification enables the study of droughts based on their key characteristics of drought duration D (days, months, seasons or years), drought severity S (= total deficit), and magnitude M in terms of average deficit or the intensity (= S/D) (Pandey et al., 2008; Smakhtin, 2001).

In their study Pandey et al. (2008) adopted a threshold based on a exceedence level of a monthly flow duration curve derived for every month of study arguing that it presents a more realistic drought pattern than an steady annual truncation level, Motalleb et al., (2014) used a constant threshold level based on 70% exceedence of daily flow duration curve, (Sung & Chung, 2004) also applied 70% exceedance but both daily and monthly flow duration curve were used to derive a varying and a fixed (constant) threshold level respectively.

As reported by Motalleb et al. (2014), threshold method had been used in several studies including frequency analysis of drought characteristics, studies of the at-site statistical characteristics of multi-year stream flow droughts, analysis of drought duration and deficit volume series selected from the daily stream flow records analysis of the link between drought occurrence and atmospheric circulation patterns , and study of spatial variability in temporal trends of stream flow drought characteristics.

2.3.Drought Characterization

Drought characterization involves the understanding of the nature of drought events that occurs in a certain area. This will include the understating of the magnitude, intensity, spread and the impact of the drought event, it involves the study of the historical drought occurrences in the study area and the impacts of such drought events had. Various ground based and remotely sensed data sets are used to both identify and quantify the drought events and their impacts. Drought indices are used to identify droughts and their characteristics (Serinaldi et al., 2009). For the purposes of this study, SPI and the threshold methods shall be used to identify and characterize droughts in the study area.

Based on SPI, the following drought characteristics can be identified: 1) drought length (or duration) L, defined as the number of consecutive intervals (months) where SPI remains below the threshold value -1; 2) mean SPI value Z, defined as the mean of SPI values within a drought period; 3) minimum SPI value Zmin, defined as the minimum SPI value within a drought period; and 4) drought areal extent, defined as the percentage of area where the index exhibits values less than the considered threshold. (Serinaldi et al., 2009).

Using the theory of runs, droughts can be characterized by their lengths (drought duration), and accumulated deficit, defined as the sum of single deficits, i.e. the deviations of the variable from the threshold, over drought duration (Bonaccorso et al., 2013), start of the drought period, drought intensity which is the total deficits divided by the drought period, and minimum flow during the drought period (Yahiaoui et al., 2009.).

2.4.Tropical Rainfall Measuring Mission (TRMM)

The Tropical Rainfall Measuring Mission is a joint mission between the NASA and Japan space agency (JAXA), the satellite was launched in November 1997 and have been providing high quality rainfall estimation data since then. The primary TRMM instruments are the Precipitation Radar (PR) (the first and only rain radar in space) and the TRMM Microwave Imager (TMI), a multi-channel passive microwave radiometer, which complements the PR by providing total

hydrometeor (liquid and ice) content within precipitating systems. The Visible Infrared Scanner (VIRS) is used to provide the cloud context of the precipitation and is used to connect microwave precipitation information to infrared-based precipitation estimates from geosynchronous satellites. TRMM processing algorithms combine information from these instruments and provide the finest scale (0.25° ×0.25°) precipitation estimate currently available from space (Fang et al., 2013; Immerzeel et al., 2009).

TRMM data is released in two datasets, 3B42 which is produced by merging passive microwave data from several low orbit satellites with infrared (IR) data collected by the international constellation of geosynchronous Earth orbit satellites (GEO-IR) and PR active microwave data, this data is available at 3-hour interval with a 9-hour lag (Duncan & Biggs, 2012) and 3B43 which is produced by refining 3B42 using several algorithms and ground based data. 3B43 dataset is released monthly and is the dataset that was used in this study.

2.4.1. Downscaling of TRMM data

The TRMM data have been found to correlate well with observed data, however the spatial resolution of the data (0.25° X 0.25°), equivalent to 28 km X 28 km at the equator is too coarse for small catchment hydrological study, (Duan & Bastiaanssen, 2013) reports of studies that have shown variability of up to 36% between two rain gauges within a 4 X 4 km grid. Due to this variability, (Duan & Bastiaanssen, 2013) concludes that downscaling of the TRMM data to finer spatial scale is crucial for obtaining accurate results. In their study they reduced the TRMM grid to 1km X 1km grid which “can be compared better with individual rain gauges” and offers a good resolution for studies at basin and regional scales.

Several techniques have been utilized in downscaling the precipitation including use of NDVI (Normalized Difference Vegetation Index) (Immerzeel et al., 2009), and a combination of NDVI and DEM (Jia et al., 2011). This have been based on the apparent relationship between the NDVI and DEM with precipitation.

2.4.2. Criticism of TRMM

Satellite based rainfall estimates are only estimates of the precipitation and not actual observations, as such their require validation using ground based data (e.g. from rain gauges) so as assess their accuracy (Duncan & Biggs, 2012). In their study (Duncan & Biggs, 2012) compared the accuracy of TRMM and ground based rainfall data from rain gauges and found that the TRMM overestimated the rainfall amounts, rainy days, and had false detection of extreme events. Consequently, they concluded that TRMM is not useful for hydrological studies at the region/ catchment level. Also Almazroui, 2011) reports the same issue of over estimation of rainfall amounts but notes further that during the wet season TRMM, tends to underestimate the rainfall by a small margin which the author corrected to march TRMM to Observed data, concluding that TRMM data is applicable to various hydrological applications. This reported difference between TRMM and observed rain gauge data is more in the daily estimates, than in the monthly estimates which portrays a better correlation. However, Duncan & Biggs (2012) working in Nepal and Almazroui (2011) working in Saudi Arabia agree that TRMM data shows trends that are very similar to those of observed data.

It is important to note that these two studies Duncan & Biggs, (2012)and Almazroui, (2011) used TRMM data at coarse spatial scale of 0.25 0X 0.250 (about 28km X 28km at the equator) which other studies have indicated to be coarse for such studies since significant Rainfall variability will be observed in such a huge area and a single rain gauge located in a such a grid cannot explain the rainfall characteristic of the entire area covered by the grid (Duan andBastiaanssen, 2013).

2.5.Artificial Neural Networks

2.5.1. Introduction

Artificial Neural Network (ANN) is a network of interconnected simple processing units known as neurons styled after the human brain neurons, the neurons are arranged in layers depending on the characteristics of the neurons. Each neuron is connected to other neurons in adjacent layers but not to those in the same layers, the strength of this connection is called “weights” and is what is

adjusted by the training algorithm until the input matches the expected output (Srinivasulu and Jain, 2006). They can also be defined as computational and mathematical model based on the functions of the biological neurons (Gao et al., 2010; Majumder et al., 2010), they are capable of identifying patterns and complex relationships between sets of input and output by leaning and recognizing patterns from historical data (Gao et al., 2010; Khan et al., 2010; Majumder et al., 2010).

ANN technology have been used to solve many problems in hydrology, some of the areas of applications includes forecasting water consumption, flow estimation, water management and even in climate variability (Khan et al., 2010). In most of these studies, ANN have been reported to have performed better than the classical regression models (Khan et al., 2010, 2010; Majumder et al., 2010). ANN has also been found to have many advantages over other modelling techniques including the ability to handle large amounts of noisy data in especially situations where the physical relationships between the dataset are unknown, such datasets could be linear or non-linear (Aqil et al., 2007). Additionally neural networks results into faster model development and improved model performance. Khan et al., (2010) used an ANN to develop an adaptive model that learns from the historical irrigation water allocation data and predicts the future allocation needs, in this study they highlights some of the advantages of the ANN to include 1) ANN are massively parallel allowing them to gain high speed performance in decision making, 2) ANN can be applied in situations that complex decisions needs to be made. ANN are classified depending on several attributes including learning procedure (supervised, non-supervised), topography (single layer, multi-layer), connection type (static/feed-forward, dynamic/feedback). Supervised ANN uses expected output to learn the pattern of the input-output relationship, non-supervised learns from data clusters without associating to the output. The multi-layer perceptron (MLP) is the most commonly used ANN in hydrology (Mutlu et al., 2008), it consists of neurons known as perceptron, the perceptron outputs a single output from a multiple real value inputs by forming a linear combination according to its input weights and passing the result through an non-linear activation function (Mutlu et al., 2008). The neurons are arranged in layers, a

basic MLP will have three layers comprising of an input layer, a hidden layer and an output layer. The input layer receives the input set from outside world and passes it to the hidden layer for processing (Adamowski and Sun, 2010).

Development of ANN follows the basic rules that 1) information must be processed at many single elements known as nodes, 2) Signals are passed between nodes through connection links and each link has an associated weight that represents its connection strength, and 3) Each of the nodes applies a nonlinear transformation called as activation function to its net input to determine its output signal.

While the number of neurons in the input and output layers are determined by the number of input and output variables in the modelled system, the number of neurons in the hidden layer is a critical consideration since the performance of the model is dependent on them, too many neurons will result into over fitting and too much computational time, on other hand fewer than needed neurons will hidden the opportunity for the neural network to capture the intricate relationships between indicator parameters and the computed output parameters.

The MLP though popular has several drawbacks 1) it is highly sensitive to the selection of the initial weight values and may give performance that differ significantly under different application, 2) During training the models are sometimes trapped at the local error minima preventing them from reaching the global minima (Cigizoglu, 2008).

2.5.2. Development of ANN Models

Srinivasulu and Jain, (2006) and Walczak and Cerpa, (2003) identifies the following steps as the key steps in the development of the ANN model: 1) selection of the model training and validation datasets, 2) output and input variables identification, 3) data preprocessing including normalization, 3) selection of ANN network to be adopted, including number of layers and number of neurons in the hidden layer, 4) Training of the model, and 5) validation of the model. The input factors can be determined using various processes including auto / cross correlation, or partial autocorrelation of the data.

Selection of the ANN input parameters is one of the most important steps in the development of the ANN model, however in most hydrological applications of ANN, this stage is largely ignored or not done properly (Bowden et al., 2005), it is important to select data that is relevant to the modelled output to avoid issues brought about by irreverent data including long periods of training, poor results as the ANN spends too much resources on the irreverent portions of the input, and the fact that it is easier to understand a simpler model than a complex one, more input complicates the model.

The selection involves the use of analytical procedures to select a subset of the data set that which result into a best model based on established criteria. In their paper Bowden et al., (2005) list five processes commonly applied in hydrology in identification of input values to use in the ANN, these methods are 1) Relying on the prior knowledge of the system been modelled, from this prior knowledge the researcher identifies inputs that are important to the modelled process, this approach requires good understanding of the modelled system. This approach is subjective and may vary from case to case as it depends on the researcher's judgment, 2) methods based on linear cross-correlation, though the method has been applied in several studies and can be used when the modelled system is not well known, it has limitation in that it can only detect linear relationship between the input and output hence may result in rejection of inputs that are related to the output but in a non-linear manner, 3) use of heuristic approaches , 4) methods that extracts information contained in the trained ANN , and 5) a combination of various methods from the identified four.
In determination of the number of hidden neurons, Srinivasulu and Jain, (2006) used trial and error method where the number of neurons were varied from 1 – 20 until an optimum number of neurons was determined.

Measured data is required for model training, this data is divided into two groups, one for model training and the other for model testing. First the ANN is trained to represent the relationship and the processes within the measured data set, once adequately trained the model is then used to output values given input set of the

testing measured data. This generated output is compared with the output of the measured testing data set (Mutlu et al., 2008).

The ANN will be trained using error backpropagation which works by iteratively changing the networks interconnecting weights such that the overall error (between modelled values and measured outputs) is progressively reduced. The performance of the model in the testing phase will be done using RMSE, CE, and R2.

Levenberg-Marquardt algorithm is best suited for function approximation problems where the number of weights is less than a few hundred and where the approximation must be very accurate they are other ANN training algorithms that are suited for large networks and pattern recognition functions including BFGS Quasi-Newton, Resilient Backpropagation, Scaled Conjugate Gradient, Conjugate Gradient with Powell/Beale Restarts, Fletcher-Powell Conjugate Gradient, Polak-Ribiére Conjugate Gradient, One Step Secant, Variable Learning Rate Backpropagation (Beale et al., 2014). Levenberg-Marquardt algorithm has been used in several hydrological ANN studies including (Naresh et al., 2011) who developed a neural network prediction model that forecasted quite accurately ten days inflows of two years ahead and generated synthetic series of ten days inflows that preserved the key statistics of the historical ten days inflows, (Kişi, 2007) who compared Back-Propagation Algorithm, Conjugate Gradient Algorithm, Levenberg–Marquardt, and Cascade Correlation Algorithm to develop a streamflow forecasting model of North Platte River in the United States, he reported that Levenberg–Marquardt produced better forecasting results than the other training algorithms. Other application of Levenberg-Marquardt algorithm in hydrology includes in storage-yield models e.g. (Adeloye and Munari, 2006), ground water level forecasting e.g. (Daliakopoulos et al., 2005), and reservoir inflow forecasting e.g. (Coulibaly et al., 2000). Therefore Levenberg–Marquardt algorithm was used to train the ANN in this study.

2.5.3. Criticism of Neural Networks

One of the major criticism of neural networks is their inability to generalize, an ANN network is unable to detect accurately values that are outside its training

data sets (Imrie et al., 2000). This inability to generalize can be addressed in a variety of ways including ensuring the correct number of neurons are used in the hidden layer (having too many neurons will result in model over fitting, too few will result into the model not learning all the possible relationship), using a large dataset that encompasses are the likely data points, and using ANN methodologies like early stopping (Beale et al., 2014).

2.6.Summary

From the review of the literature that has been done so far, it is clear that use of artificial neural networks and remote sensing data offers a solution in the study of drought hazards and disasters in Africa, a continent which suffers a great deal from frequent droughts and also lacks an adequate climatic data observation network.

Remotely sensed variables have been found to compare well with the observed data and artificial neural networks models have been found to be adequate alternatives to the traditional regression models.

2.7.Research Gaps

There is little work done to study hydrological droughts monitoring and forecasting using Tropical rainfall Measuring Mission precipitation estimates and Artificial Neural Networks in Kenya that make use of remotely sensed data. Further, no other study has been done to develop and evaluate an artificial neural network model for the hydrological drought based on downscaled Tropical Rainfall Measuring Mission precipitation estimates for the Narumoro sub-catchment.

CHAPTER 3: Methodology

3.1.Introduction

In the carrying out of the study, the following key steps were followed:

1. Collection of meteorological data and remote sensing data.
2. Data preprocessing
3. Downscaling of TRMM data
4. Meteorological drought characterization
5. Hydrological drought characterization
6. ANN development

3.2.Study Area and Datasets

3.2.1. Study Area

The study area comprises the upper catchment areas of Narumoro River, The catchment areas defined by the catchment areas of the Narumoro River Gauging Stations (RGS) as shown in the Figure 1. The RGS is located at 37.025, 0.1653 and the Rainfall Station is located at 31.15, -0.1667 (Longitude, Latitude)

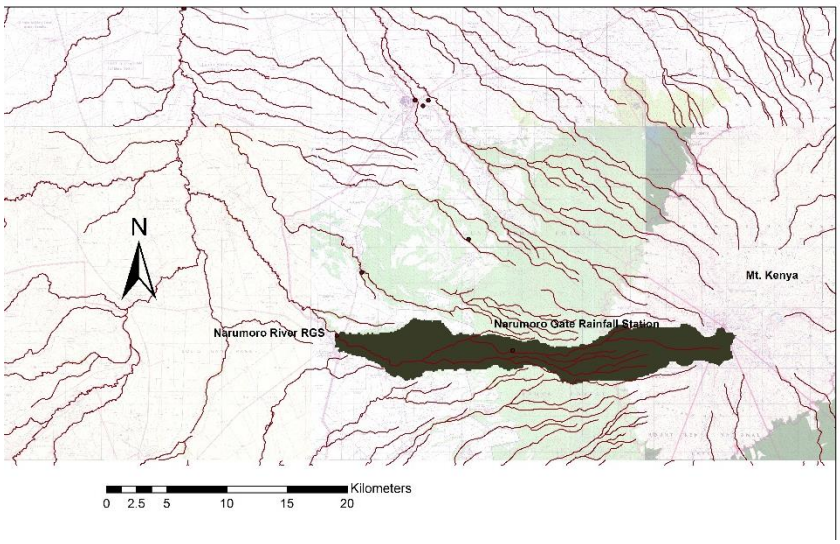

Figure 1: Study Area

Remote sensing and ground observed data was collected for the study area. Ground observed data was collected from Water Resources Management Authority and consisted of daily rainfall and daily streamflow data. A list of the meteorological data collected is listed in table 2. Remote sensing data collected included Tropical Rainfall Measuring Mission precipitation data, digital elevation

model data from the Shuttle Radar Topography Mission (SRTM), and Normalized Difference Vegetation Index (NDVI) images from SPOT VEGETATION satellite. The remote sensed data was collected for a larger area than project site to provide enough data sets for regression analysis procedure of downscaling TRMM, the extent of data collection area is as shown in Figure 2 and defined by coordinates (0.5, -0.5, 36.5,36.5).

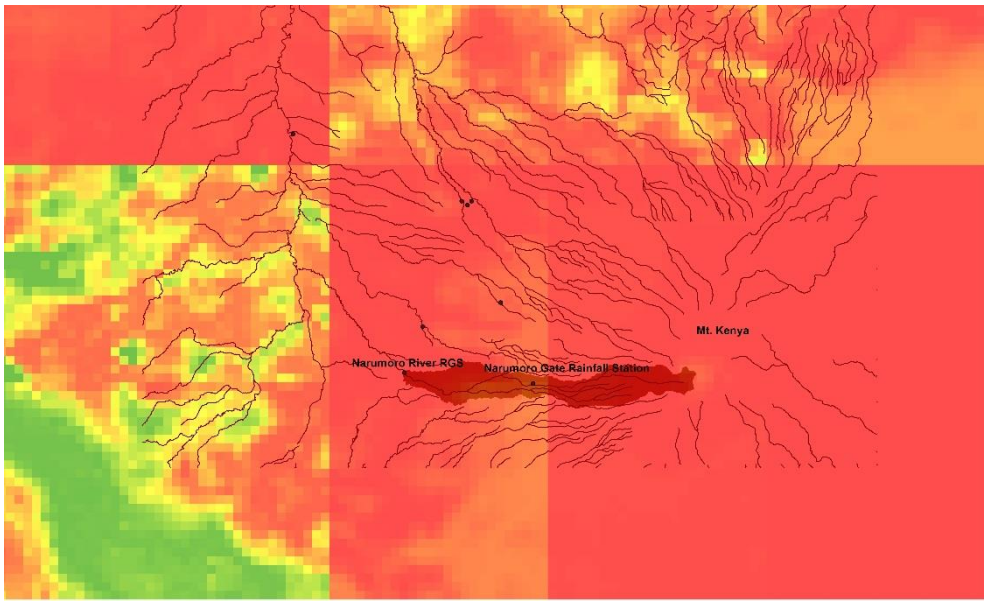

Figure 2: A downscaled TRMM image of January 2001 showing the data collection extent.

In the following sections, collected datasets (i.e. TRMM data, DEM, SPOT VGT) are described.

3.2.2. Tropical Rainfall Measuring Mission Precipitation Estimates

Tropical Rainfall Measuring Mission precipitation data is readily and freely available from the TRMM website (NASA, 2014) where a "simple subset wizard" facilitated searching and downloading data for the study area. TRMM 3B43 version7 was used in this study, the data is distributed in netCDF (Network Common Data Form) format (Unidata, 2014). TRMM 3B43 v7 has a temporal resolution of one month and a spatial resolution of 0.25^0 x 0.25^0. It is a measure of precipitation rate and is reported in mm/hr. The data was downloaded for the period from January 2000 to November 2011 (131 months), this was to correspond to the available ground based data. A total of 131 netCDF files were downloaded, for each month of the study period. To facilitate analysis the netCDF

files were converted into raster images using ESRI's ArcGIS (ESRI, 2014), the raster images were then converted to ASCII GRID format which is an accessible text format. So ultimately, the TRMM data was available in a text format that can be easily imported into data processing software like MATLAB.

3.2.3. Digital Elevation Model

"The Shuttle Radar Topography Mission (SRTM) is an international project spearheaded by the National Geospatial-Intelligence Agency (NGA) and the National Aeronautics and Space Administration (NASA). By a specially modified radar system that flew onboard the Space Shuttle Endeavour during an 11-day mission in February of 2000, the SRTM obtained the most complete high-resolution Digital Elevation Models (DEM) on a near-global scale from 56°S to 60°N. The resolution of the cells of the source data is normally for one arc second; however, 1″ (approx. 30 m) data have only been released for the United States territory; for other world regions, only three-arc second (approx. 90 m) data are available. Therefore, the DEM used in this study is the three-arc-second data" (Jia et al., 2011). This data was downloaded and used to demarcate the catchment area.

3.2.4. Spot Vegetation

The VEGETATION instrument (VGT), on board the SPOT 4 and SPOT 5 satellites has four spectral bands: blue (0.43–0.47 μm), red (0.61– 0.68 μm), Near Infrared (NIR, 0.78–0.89 μm) and Short Wave Infrared (SWIR, 1.58–1.74 μm). The red and NIR bands are used to calculate NDVI ([NIR−RED]/ [NIR+RED]). The spatial resolution of the imagery is 1 km at nadir and the 2200 km swath width allows daily imaging of about 90% of the equatorial regions, the remaining 10% being imaged the following day. At latitudes higher than 35° (North and South), all regions are observed daily. The synthesized pre-processed S10 NDVI product was used in this study, which is a geometrically and radio metrically corrected 10-day composite image. The 10 day periods were defined from the 1st to the 10th, from the 11th to the 20th and from the 21st to the end of each month. Atmospheric corrections were performed using the Simplified Method for Atmospheric Corrections (SMAC) (Immerzeel et al., 2009) Each image contains the maximum NDVI value for the 10 days. Data from January 2000 to November 2013 was collected for the study area, a total 468 images were collected. The data is downloaded in form of raster images (SPOT-VGT) from the product website

(VITO, 2013). This raster images were converted into ASCII GRID format which is easier to import into data process and manipulation software. For every month the maximum value of the three images present in each monthly data were selected to form a dataset that was used in the study. NDVI value was computed using Equation 1 (VITO, 2013).

$$NDVI = 0.01 * P - 1 \quad \ldots\ldots (1)$$

Where P is the pixel value of the Spot-Vegetation Image

3.2.5. Meteorological Data

Daily streamflow data and rainfall data were collected for various stations. The data collected is summarized in Table 2. Missing values were filled using linear interpolation of the daily data. The daily data was then aggregated into monthly totals by summing up the daily values.

Table 2: Meteorological Data Collected

Name of Station	Data	Start Year	End Year	Comments
NANYUKI DISTRICT AGRIC OFFICE	Rainfall	2000	2004	Discarded(To many missing data records & very short span)
NANYUKI M.O.W	Rainfall	1965	2012	
Narumoru Gate	Rainfall	1970	2010	
5BC02 NARO MORU	Discharge	1947	2011	

3.3. Downscaling of TRMM Data

Downscaling of TRMM data was done for the period January 2000- November 2011 as influenced by the available historical observed data records.

3.3.1. Establishment of relationship between TRMM and NDVI

The relationship between rainfall and vegetation was established for NDVI values per TRMM pixel. This was done by degrading the 1 km NDVI image to 28 km pixels in order to match the TRMM image. The image was exported from arcGIS to an ASCII file and imported into MATLAB. Using MATLAB a relationship between NDVI and TRMM is determined by fitting a polynomial through the data. As the relationship is thought to be not universal, this relationship is determined for every month separately. This was done by plotting the TRMM values against the mean NDVI values, with NDVI on the horizontal axis and TRMM rainfall on the vertical axis. The relationship was established for 3, 2 and 1 order polynomial as defined in Equations 2, 3, , and 4 respectively and the best fit polynomial based on R^2 was selected for establishing the relationship.

$$p_n = a * NDVI^3 + b * NDVI^2 + c * NDVI + d \quad \ldots\ldots (2)$$

$$p_n = a * NDVI^2 + b * NDVI + c \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

$$p_n = a * NDVI + b \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Where a, b, c, and d are the parameters of the polynomials.

3.3.2. Downscaling Model

This second order relation derived in the former paragraph is used in the downscaling model. For every NDVI pixel, the corresponding TRMM value (P_n) was determined using the analytical relation established (for which the coefficient change by month). This value was used as a weighing function P_n / P_{28} in the downscaling procedure, then for every TRMM pixel the mean P_n value of 28 pixels is determined from linear averaging (P_{28}). Using Equation 5, the downscaled 1 km TRMM image was calculated:

$$TRMM_{1Km} = \frac{P_n}{P_{28}} * TRMM_{28Km} \quad \ldots\ldots\ldots (5)$$

3.3.3. Retrieving Dataset for the Study

From the down scaled image, TRMM data for the various drought identification studies were obtained. For meteorological drought studies, TRMM values were extracted for the 1km pixel within which the target rain gauge was located, this

was classified as the downscaled rainfall equivalent to the observed rain gauge data. The values are obtained as precipitation rate (mm/hr) and are converted into total monthly rainfall. For hydrological droughts, the total monthly rainfall over the catchment area of the considered river gauging station was obtained. Similarly TRMM at the original unscaled resolution was obtained for the corresponding rain gauges and catchment areas.

3.4.Identification of Meteorological Drought

Meteorological droughts were analyzed using the standardized precipitation index (SPI), the index was computed for both the ground observed data and the for the TRMM data. SPI was computed for 3, 6, 9, 12, and 24 month aggregation periods to give SPI3, SP6, SPI9, SPI12, SPI24 respectively ,using an SPI program developed and distributed by (National Drought Mitigation Center , 2014) which uses monthly dataset as inputs and outputs the SPI values. The computed SPI was used to identify and derive the characteristics of the identified droughts between year 2000 and 2011, this time period was adopted as depending on the availability of TRMM and observed data. TRMM data is only available from year 1998 and consistent rain gauge data for the stations is available from 2000 up to 2011. The computed SPI values from the three datasets (downscaled TRMM, original TRMM and rain gauge data) were analyzed for similarity using statistical test (e.g. covariance), graphing techniques and comparison of the characteristics of the identified droughts considering the drought period, drought onset and drought end, drought severity and drought classes. Based on this tests and comparisons, conclusions were made on the effectiveness of the three datasets in drought identification in Narumoro river sub-catchment.

3.5.Identification of Hydrological Drought

The threshold method also known as the theory of runs was used to identify and characterize hydrological droughts. The first step involved determination of the threshold level (truncation level) below which a flow was considered a drought flow. The Flow Duration Curve (FDC) method was used to obtain the threshold with flows exceeded 70% of the time considered as drought flows. The FDC curve was generated using the long-term observed historical discharge data which run

for at least 30 years for each of the considered river gauging station. The lows flows were however determined for the period running from 2000 due to the limitation of the available data.

Seasonal runoff coefficients were computed at monthly and annual time intervals based on the observed flow data and the corresponding TRMM data for the catchment. The seasonal runoff coefficients were computed using Equation 6.

$$seasonal\ Runoff\ Coefficient = {}^{Total\ Flow}/_{Total\ TRMM} \quad \cdots\cdots (6)$$

Consequently the mean, minimum and maximum seasonal runoff coefficient were computed for the entire study period. Using the runoff coefficients and TRMM, flows were estimated and regression analysis done between the observed and estimated flows. Based on the regression model, low flows were estimated. Statistical tests were used to test the similarity of the estimated low flows and the observed low flows.

3.6.Developing ANN Model

3.6.1. Selection and Preparation of the datasets

The objective of the model development is to predict streamflow that can be used to forecast hydrological droughts in the catchment. The model inputs will be either lagged flow, lagged values of TRMM or both TRMM and flow. The projected output will always be streamflow. Six cases were determined to be used:

Case 1 used the TRMM as input and the corresponding flow as the expected output. This case was to simply check the utility of current month TRMM to estimate the current month stream flow. This relationship can be expressed as in Equation 7.

$$TRMM_0 \rightarrow FLOW_0 \text{ ---- } (7)$$

NB: the subscript denotes the time lag in month of the variable from the matched flow.

Case 2 TRMM and flow with a one month time lag was used to model current flow. This relationship can be modeled as in Equation 8;

$$TRMM_{-1} + FLOW_{-1} \rightarrow FLOW_0 \text{ ---- (8)}$$

Case 3 used TRMM and flow with a 3 month time lag as shown in Equation 9

$$TRMM_{-3} + FLOW_{-3} \rightarrow FLOW_0 \text{ ---- (9)}$$

Case 4 was similar to case 3 but used TRMM alone without the flow as shown in Equation 10

$$TRMM_{-3} \rightarrow FLOW_0 \text{ ---- (10)}$$

Case 5 used TRMM only at a six month lag time as shown in Equation 11

$$TRMM_{-6} \rightarrow FLOW_0 \text{ ---- (11)}$$

And finally case 6 which used TRMM and flow at 6 months lag as shown in Equation 12.

$$TRMM_{-6} + FLOW_{-6} \rightarrow FLOW_0 \text{ ---- (12)}$$

For each case a dataset was prepared involving the inputs and the corresponding output. Where the inputs were two, a 2 X 139 matrix was formed and 1 X 139 matrix where the input was just one. The output was only one throughout thus 1 X 139 matrix was created in all instances. A dataset for the purpose of this description will refer to a two element dataset, composed of the input (s) and the output. Each dataset was subdivide into two groups, one group composed of data from January 2000 up to September 2010, was used for the ANN model development (Design dataset) whereas the rest October 2010 to November 2011 was put aside for assessing the ANN model after development (Test Dataset). The workflow of neural network design involves 7 key stages including: 1) Collect data; 2) Create the network; 3) Configure the network; 4) Initialize the weights and biases; 5) Train the network; 6) Validate the network (post-training data); 7) Use the network.

Creating the network involved determining the number of hidden layers, the network training algorithm, and the number of neurons in the hidden layer. Configuration of the network involves arranging the network so as to make it compatible with the problem been solved as defined by the example data. Training

process involves the adjustment of the network parameters (weights and biases) to optimize performance of the network.

After the data is collected, it is preprocessed and divided in groups before been subjected to the model. Both the input and the targets are normalized to prevent the saturation of the training function as part of the data preprocessing procedures. Figure 3 shows typical ANN blocks comprising of the pre and post processing blocks and the ANN itself.

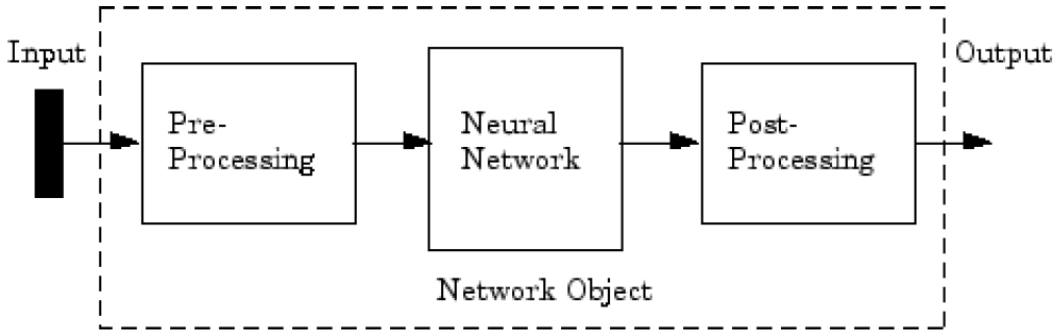

Figure 3: Neural Network showing the pre and post processing blocks (Beale et al., 2014)

The model development dataset is then subdivided into three groups for training, testing and validating the model. The training dataset is used for computing the gradient and updating the network weights and biases, the second dataset is the validation set. It is the error on the validation set that is monitored during the training process, network weights and biases are saved when the error on the validation set is at its minimum. The test dataset is the third group, it is not used during the training process but it is used to test the performance of the model developed. MATLAB has various inbuilt functions for sub-dividing the datasets into the three groups, *dividerand* function was used to divide the data randomly into the groups based on a ration 0.7: 0.15: 0.15 (Training: Testing: Validation).

The next step involves creation of the network, the MATLAB function *feedforwardnet* was used to create a network. The network created has no inputs or weights defined. The configuration of the network follows where the inputs are defined and in the initialization, the model weights and biases are initialized. The

network is then trained using a training Algorithm. Figure 4 shows a basic multilayer ANN with 10 neurons in the hidden layer that was used in the study.

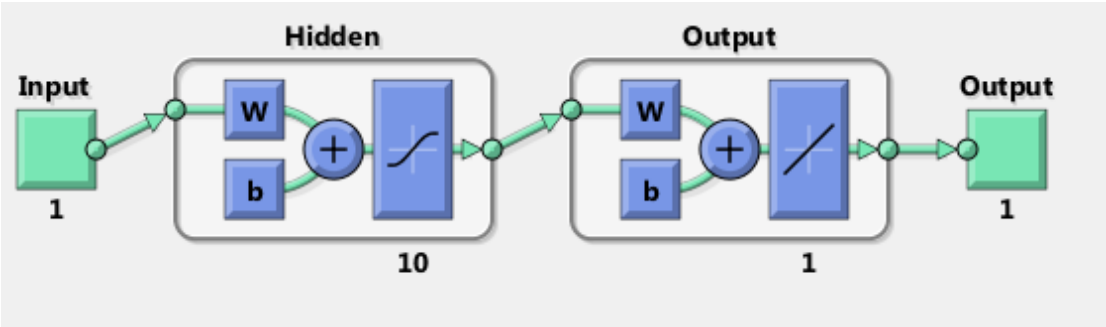

Figure 4: Developed Default ANN Model

The parameters of the developed ANN models are:

1 hidden layer with 10 neurons, Levenberg–Marquardt was used as the training algorithm, Mean square error (MSE) was used as the performance indicator. The performance indicator is monitored during the training session, the training is stopped when no further reduction in the MSE is achieved when the model is applied to the validation dataset.

3.7.Monitoring the performance of the ANN model after Training

The performance of the training, validation and testing process is plotted and the curves examined. If there is no significant difference between the testing and validation curves, then the training is taken to be sufficient. This performance plot is used to detect issues of over fitting or under fitting of the datasets. The over fitting error results when the network memorized the training examples, but it has not learned to generalize to new situations, therefore error on the training set is driven to a very small value, but when new data is presented to the network the error is large (Beale et al., 2014). Problems of over fitting or under fitting are resolved by adjusting the number of the hidden neurons, when few neurons are used under fitting results and when more than necessary neurons are used, over fitting results. Since it is impossible to know firsthand the number of adequate neurons to use in the network, regulation and early stopping of the training process is used to ensure good model generalization.

The method of early stopping was used in this study. It generally involves stopping the training process when no more reduction in MSE is achieved when

the model is presented with validation dataset. To ensure generalization using the early stopping method, several approaches can be used including retraining the ANN several times, each time using different initial weights and biases, the ANN with the least MSE is then adopted or several ANN can be developed and trained, the average of their outputs is then adopted for each input, this is particularly helpful in small noisy datasets. The approach of training several models and using the average of their input was adopted in this study. Several models were created and trained for each case, up to eight best models based on MSE and R Squared were selected. The R Squared is computed by making regression plots of outputs vs targets for the training, validation and testing datasets. Figure 5 shows a typical regression plot for the training, testing and validation stages of the model development. The regression coefficient R indicates the strength of the relationship, it ranges from 0 to 1 with one indicating the strongest relationship. The dashed line in the plot shows the ideal scenario when output = target. Based on the regression plots, the generalization capability of the model is evaluated, presence of outliers may indicate data points whose range was not represented in the training phase and the model could be extrapolating the value, if this is the case then the training dataset needs to be revised to include all the required data set range. The closer to 1 the R squared is, the better the model is. Based on the evaluation the model can be adjusted by adding or removing the hidden neurons, revising the training datasets and retraining the model to achieve the required performance levels. (Beale et al., 2014).

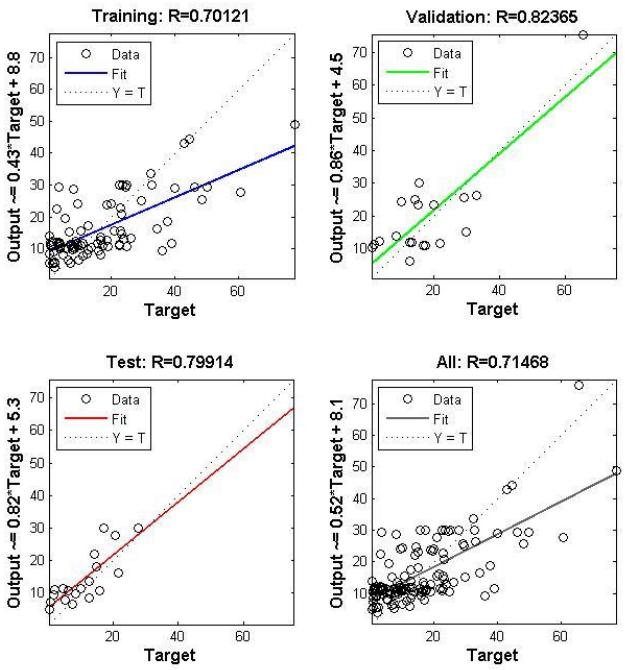

Figure 5: Regression plots for training, validation, and testing datasets

The other check adopted for selecting the best model is plotting the model training performance which indicates the rate of reduction of MSE for the training, validation and testing datasets. If the training and validation curves have no major differences, then the model is regarded as having been well trained and is unlikely to suffer from over fitting. A typical performance curve can be seen in Figure 6.

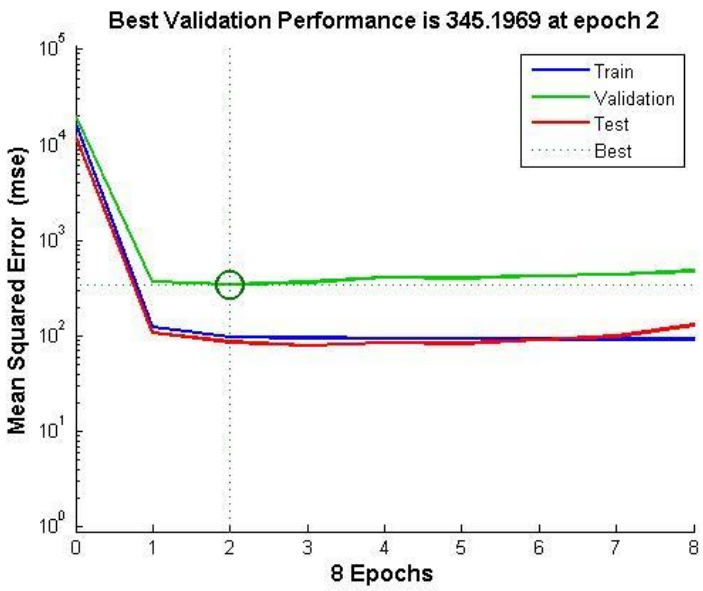

Figure 6: ANN Training Monitoring Performance Plot

CHAPTER 4: Analysis and Results

4.1.Meteorological Droughts Analysis

4.1.1. Narumoro Gate station

Figure 7 shows the comparison of SPI 3 computed from downscaled and original TRMM, and rain gauge observed data.

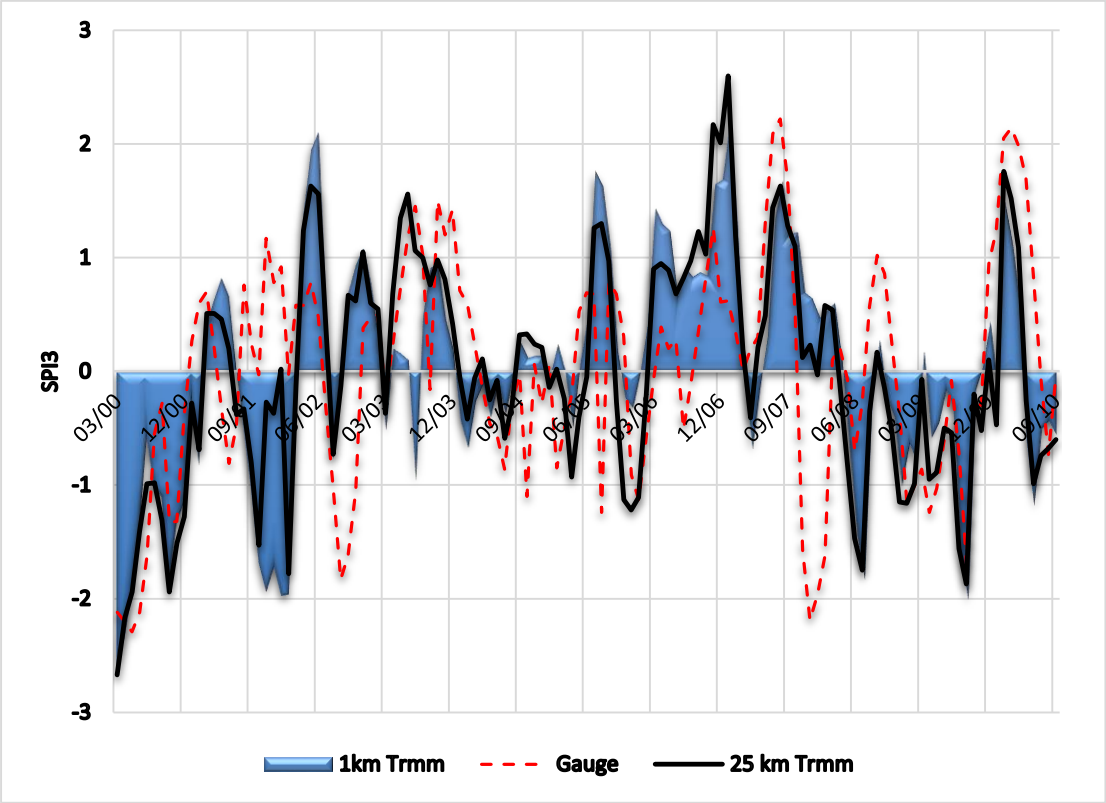

Figure 7: SPI 3 comparison at Narumoro gate station

The downscaled TRMM computed SPI compares well with those from the non-downscaled TRMM. In regard to ground observed data, there is a tendency toward following the pattern of SPIs computed from TRMM data. Disregarding SPI values greater than -1 which corresponds to non- drought periods, about 11 drought events can be identified. Of the identified droughts 6 are jointly detected by TRMM and ground observed data, four are detected only by ground data and 2 were detected by TRMM data only. Only two of the four droughts detected by the ground based data only can be regarded as major droughts and either closely followed or preceded a case of false detection by TRMM. Table 3 shows the droughts identified using SPI 3, it shows the drought onset and ending, drought

period and intensity of the drought which is the mean SPI during the drought period.

Table 3: Drought identification using SPI 3

Narumoro Gate Drought Identification SPI3											
Gauge				1KM TRMM				25KM TRMM			
Start	End	Period	Intensity	Start	End	Period	Intensity	Start	End	Period	Intensity
3/00	***12/00***	***10***	***-1.42***	***3/00***	***2/01***	***12***	***-1.37***	***3/00***	***2/01***	***12***	***-1.43***
5/01	7/01	3	-0.54	8/01	3/02	8	-1.38	7/01	12/01	6	-0.61
10/01	10/01	1	-0.03	**8/02**	**8/02**	**1**	**-0.77**	2/02	3/02	2	-0.95
3/02	3/02	1	-0.03	3/03	3/03	1	-0.52	**8/02**	**9/02**	**2**	**-0.43**
7/02	**11/02**	**5**	**-1.17**	7/03	7/03	1	-0.96	3/03	3/03	1	-0.37
3/03	3/03	1	-0.21	1/04	8/04	8	-0.30	1/04	3/04	3	-0.18
9/03	9/03	1	-0.16	1/05	1/05	1	-0.03	5/04	8/04	4	-0.30
4/04	10/04	7	-0.49	4/05	4/05	1	-0.49	1/05	1/05	1	-0.14
12/04	12/04	1	-0.28	11/05	1/06	3	-0.19	3/05	6/05	4	-0.41
2/05	4/05	3	-0.52	3/07	5/07	3	-0.30	10/05	2/06	5	-0.79
8/05	8/05	1	-1.24	5/08	8/08	4	-1.04	4/07	4/07	1	-0.41
12/05	3/06	4	-0.67	10/08	2/09	5	-0.61	***1/08***	***1/08***	***1***	***-0.03***
7/06	8/06	2	-0.31	4/09	10/09	7	-0.75	4/08	8/08	5	-0.90
11/07	***2/08***	***4***	***-1.84***	1/10	1/10	1	-0.12	10/08	11/09	14	-0.79
5/08	7/08	3	-0.35	5/10	9/10	5	0.02	1/10	1/10	1	-0.47
12/08	11/09	12	-0.78								
7/10	9/10	3	-0.29								
TOTALS		**62**			**TOTALS**	**61**			**Totals**	**62**	

The highlighted events shows corresponding drought events with SPI values less than -1. Similar color indicates a drought event detected by all the three datasets, the red font indicates cases of detection by downscaled TRMM only. Apart from one case (drought starting 8/05) where TRMM did not appropriately detect the drought event, it can be concluded that both the downscaled and TRMM at original resolution, exhibits a good level of drought detection compared to the ground observed data when SPI3 is used to classify droughts in Narumoro sub catchment.

Figure 8 shows the comparison of SPI 6 computed from TRMM and rain gauge observed data sets.

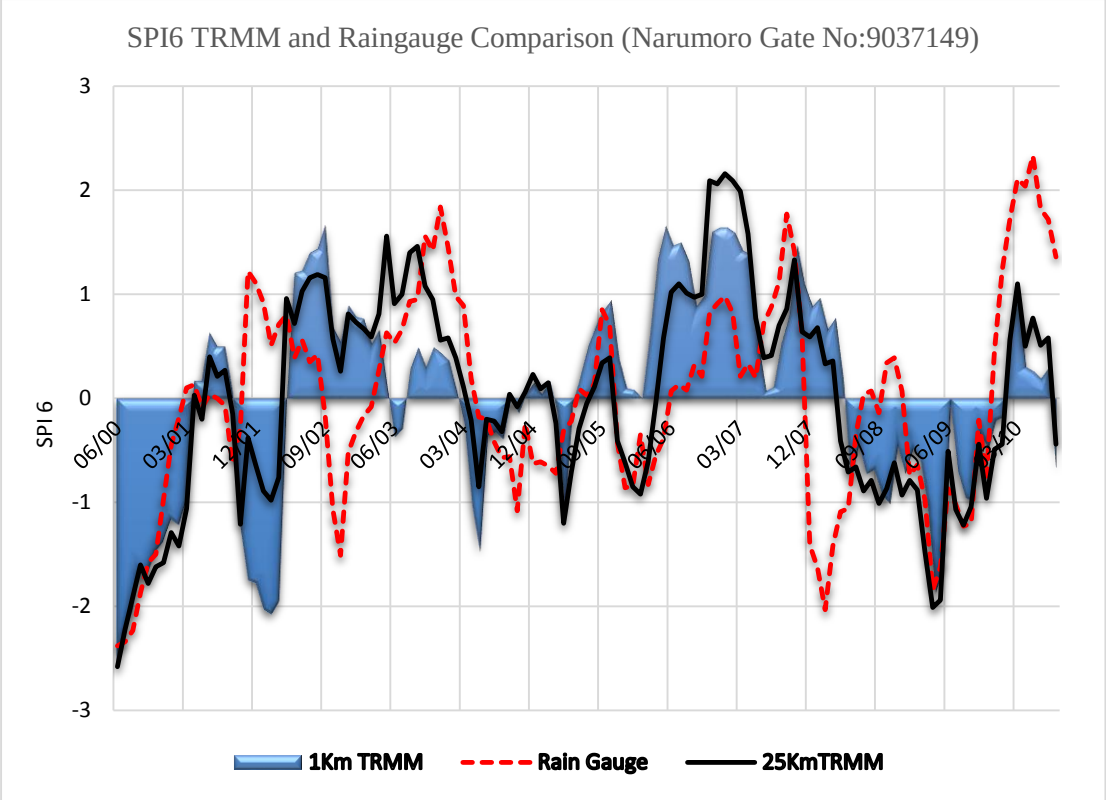

Figure 8: SPI6 Comparison

It can be seen that SPI computed from downscaled and original TRMM generally follows a similar pattern, this is also the case with SPI from rain gauge data which exhibits peaks and troughs that corresponds to peaks and troughs from TRMM SPIs. 8 drought events can be detected, of these only four are co-detected by TRMM and ground based data, two are detected by TRMM only and 2 are detected by rain gauge data only. SPI6 computed from the original non-downscaled TRMM tends to agree more with SPI6 from rain gauge data than does SPI from downscaled TRMM. From this results, it appears that for good drought detection using SPI6, downscaled and non-downscaled TRMM should be used together since their results complement each other. Table 4 shows drought identification based on SPI 6, it shows similar number of months with SPI below 0 for rain gauge data and the original TRMM, and downscaled TRMM had less number of months where SPI was negative. As can be seen from the table, all the significant drought events were sufficiently detected by the three datasets except the drought starting 12/7 though portions of its duration are.

Table 4: SPI 6 drought identification

Narumoro Gate Drought Identification SPI6											
1KM TRMM				GAUGE				25KM TRMM			
Start	End	Period	Intensity	Start	End	Period	Intensity	Start	End	Period	Intensity
				6/00	**2/01**	**9**	**-1.5**				
6/00	**3/01**	**10**	**-1.63**	5/01	5/01	1	-0.09	**6/00**	**3/01**	**10**	**-1.71**
10/01	*4/02*	*7*	*-1.54*	8/01	10/01	3	-0.33	5/01	5/01	1	-0.20
6/03	7/03	2	-0.33	9/02	3/03	7	-0.55	9/01	3/02	7	-0.72
3/04	8/04	6	-0.63	5/04	5/05	13	-0.49	4/04	8/04	5	-0.36
10/04	10/04	1	-0.14	11/05	5/06	7	-0.61	10/04	10/04	1	-0.08
3/05	5/05	3	-0.43	12/07	6/08	7	-1.28	3/05	7/05	5	-0.50
5/08	***1/10***	***21***	***-0.72***	9/08	9/08	1	-0.14	11/05	4/06	6	-0.58
8/10	**9/10**	**2**	**-0.95**	***1/09***	***11/09***	***11***	***-1.01***	***4/08***	***1/10***	***22***	***-0.92***
				9/10	**9/10**	**1**	**-1.04**	**8/10**	**9/10**	**2**	**-0.73**
Totals		**52**		**Totals**		**60**		**Totals**		**59**	

It can therefore be concluded that SPI 6 computed from TRMM can be used to monitor drought in the study area almost as effectively as rain gauge data would.

Figure 9 shows drought identification based on SPI 9 which is SPI computed at a 9 month aggregation period. Apart from isolated cases, the pattern of peaks and trough is similar for all the datasets. Only two drought events with SPI less -1 are not jointly detected by TRMM and rain gauge data. However the SPI values are becoming increasingly different across the different SPI aggregation periods. Table 5 shows identification of the droughts based on SPI 9, it can be seen that significant droughts with SPI values greater than -1 are well detected by both TRMM and rain gauge data. Although TRMM derived SPI are less than those of rain gauge data by some margin. There is a general underestimation of low values of SPI and over estimation of high SPI values computed from TRMM when compared to those computed from rain gauge data, this could be as result of TRMM over estimating rainfall.

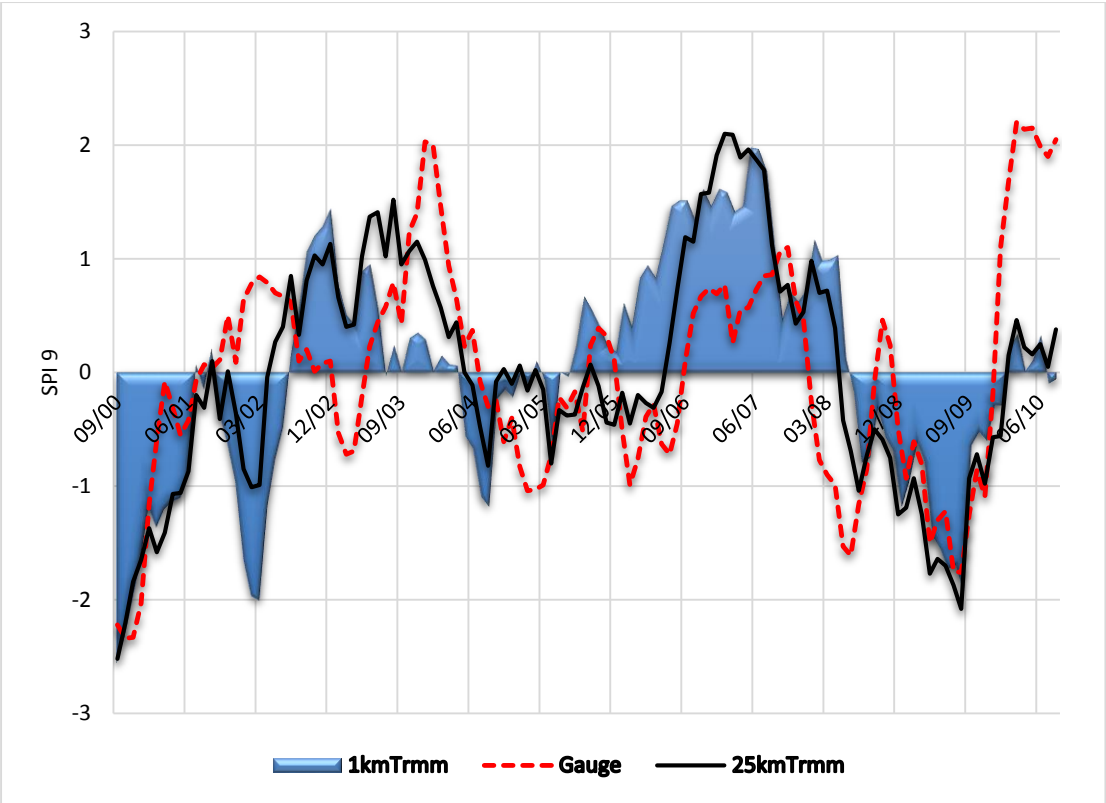

Figure 9: SPI 9 Comparison

Table 5: Drought identification using SPI9

Narumoro Gate Drought Identification SPI9											
1KM TRMM				GAUGE				25KM TRMM			
Start	End	Period	Intensity	Start	End	Period	Intensity	Start	End	Period	Intensity
9/00	**6/01**	**10**	**-1.54**	**9/00**	**7/01**	**11**	**-1.10**	**9/00**	**8/01**	**12**	**-1.34**
8/01	8/01	1	-0.15	1/03	4/03	4	-0.54	10/01	10/01	1	-0.41
10/01	*6/02*	*9*	*-1.11*	7/04	8/05	14	-0.53	12/01	4/02	5	-0.65
7/03	7/03	1	-0.03	1/06	8/06	8	-0.59	6/04	9/04	4	-0.37
5/04	11/04	7	-0.59	1/08	9/08	9	-0.91	11/04	11/04	1	-0.10
1/05	1/05	1	-0.12	**12/08**	**12/09**	**13**	**-1.06**	1/05	1/05	1	-0.16
3/05	4/05	2	-0.37					3/05	8/05	6	-0.36
6/05	6/05	1	-0.04					10/05	6/06	9	-0.29
6/08	**1/10**	**20**	**-0.87**					**5/08**	**1/10**	**21**	**-1.06**
Total		**52**		**Total**		**59**		**Total**		**60**	

Table 6 and Figure 10 gives the results of drought identification by SPI 12, like in the previous cases, the identified SPI pattern is largely the same and use of both downscaled and non-downscaled TRMM will be complementary and improve the skill of drought monitoring.

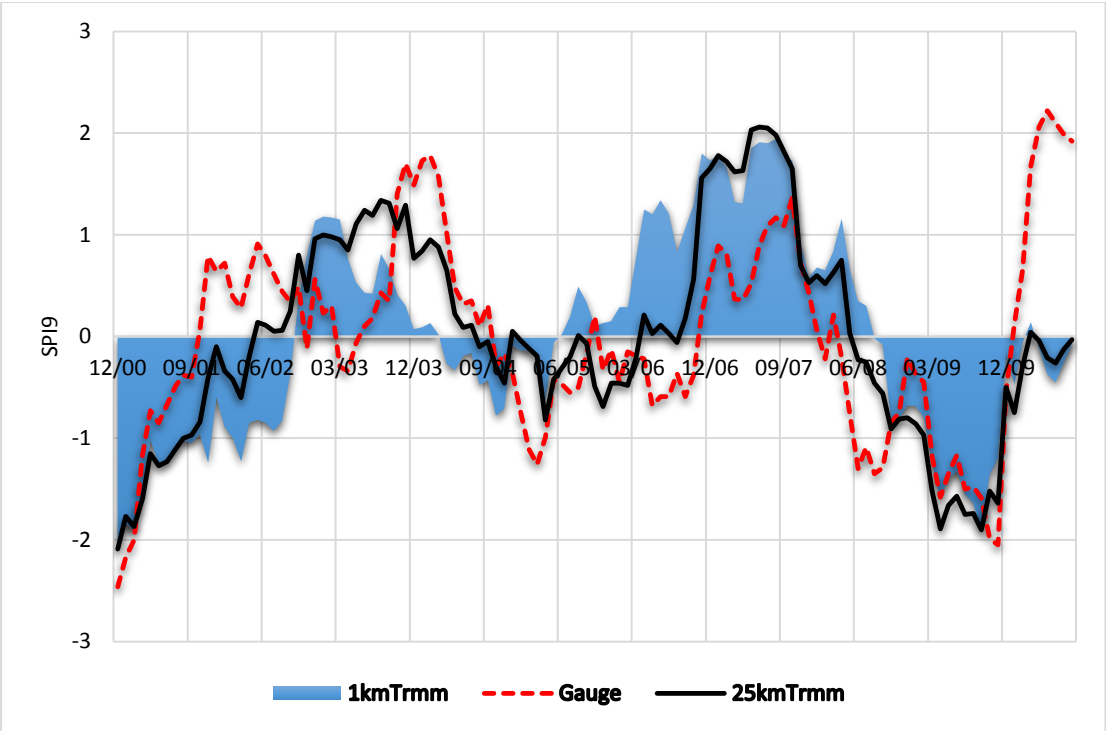

Figure 10: SPI12 Drought characterization,

Table 6: Drought Identification by SPI 12

Narumoro Gate Drought Identification SPI12											
1KM TRMM				GAUGE				25KM TRMM			
Start	End	Period	Intensity	Start	End	Period	Intensity	Start	End	Period	Intensity
12/00	9/02	22	-1.12	12/00	9/01	10	-1.13	12/00	4/02	17	-1.00
8/08	2/10	19	-0.94	11/02	11/02	1	-0.12	8/04	11/04	4	-0.23
				3/03	5/03	3	-0.23	1/05	7/05	7	-0.30
				10/04	9/05	12	-0.59	9/05	3/06	7	-0.41
				11/05	10/06	12	-0.39	8/06	8/06	1	-0.06
				2/08	2/08	1	-0.22	6/08	2/10	21	-1.08
				4/08	12/09	21	-1.10	4/10	8/10	5	-0.13
Totals		**41**		**Totals**		**60**		**Totals**		**62**	

Figure 11 and Table 7 shows the results of SPI 24 drought identification, it is noted that non-downscaled TRMM and downscaled TRMM falls out of sync with each other though the general pattern is the same. This like in the previous cases is an indication of the abilities of remotely sensed data to detect drought with almost the same skill as the ground observed data. The major difference between the TRMM data sets and the rain gauge is that TRMM generally overestimates the

high precipitation and at times underestimate the low values. Also detectable is a slight time lag between TRMM and rain gauge data.

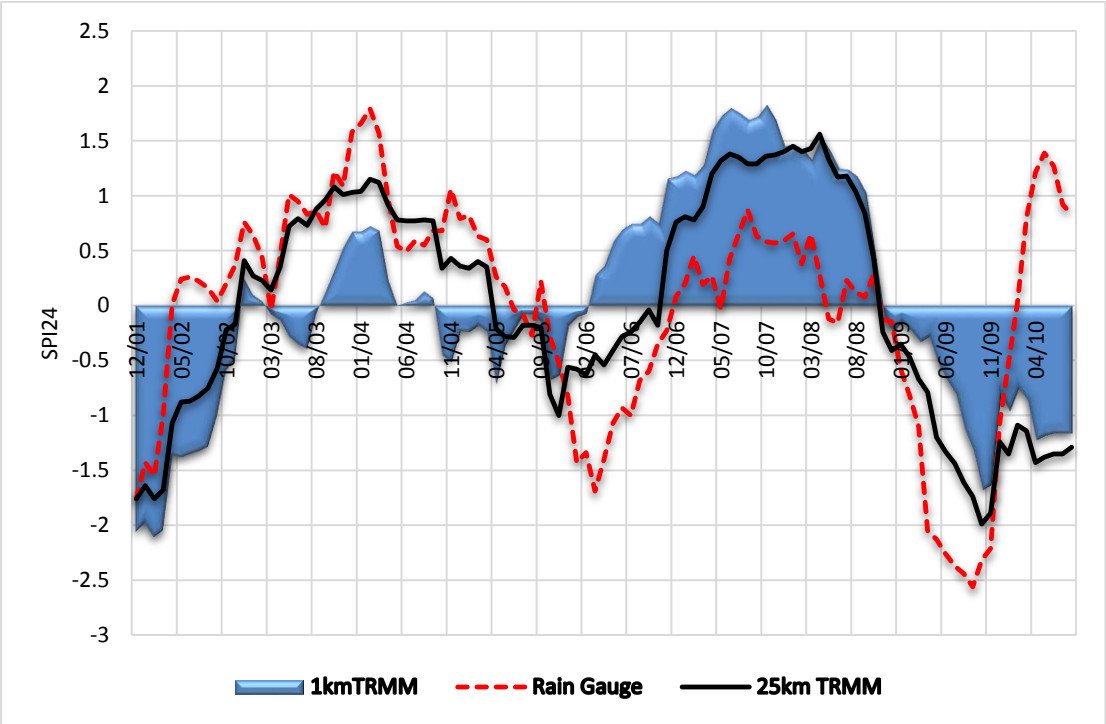

Figure 11: SPI 24 Comparison.

Table 7: Drought identification using SPI 24

Narumoro Gate Drought Identification SPI24											
1KM TRMM				GAUGE				25KM TRMM			
Start	End	Period	Intensity	Start	End	Period	Intensity	Start	End	Period	Intensity
12/01	11/02	12	-1.41	12/01	3/02	4	-1.44	12/01	11/02	12	-1.02
3/03	8/03	6	-0.22	3/03	3/03	1	-0.04	4/05	10/06	19	-0.38
10/04	2/06	17	-0.30	6/05	8/05	3	-0.13	11/08	8/10	23	-1.18
11/08	8/10	23	-0.85	10/05	11/06	14	-0.88				
				5/07	5/07	1	-0.03				
				5/08	6/08	2	-0.14				
				11/08	1/10	15	-1.51				
Total		58		Total		40		Total		54	

Table 8 shows a summary of identified drought months for various drought classes using SPI from TRMM and rain gauge data. For all the SPI aggregation periods, the number of the droughts months is similar. The maximum difference in the total drought months as identified by TRMM and rain gauge data is 3 months which is obtained at SPI 12.

Table 8: Summary of Drought Classes

SPI		Extremely Dry	Severely Dry	Moderately Dry	**Total**
	$TRMM_{1km}$	3	9	9	**21**
	Gauge	5	7	11	**23**
SPI 3	*$TRMM_{28km}$*	2	8	9	**19**
	$TRMM_{1km}$	5	7	11	**23**
	Gauge	4	7	10	**21**
SPI 6	*$TRMM_{28km}$*	3	6	11	**20**
	$TRMM_{1km}$	3	8	10	**21**
	Gauge	4	8	8	**20**
SPI 9	*$TRMM_{28km}$*	3	7	9	**19**
	$TRMM_{1km}$	1	7	14	**22**
	Gauge	4	4	11	**19**
SPI 12	*$TRMM_{28km}$*	1	12	5	**18**
	$TRMM_{1km}$	3	3	13	**19**
	Gauge	8	3	9	**20**
SPI 24	*$TRMM_{28km}$*	0	8	15	**23**

Table 9 shows the correlation coefficients between SPI values computed from TRMM and those computed from rain gauge data. The original non-downscaled TRMM computed SPI values have better correlation with the rain gauge data computed SPI values than the downscaled TRMM. Coefficient of determination (R^2) is similarly better between non downscaled TRMM and Rain gauge data computed SPIs than when using the downscaled TRMM. From these results, meteorological drought identification based on TRMM is better at higher SPI aggregation levels than at lower levels.

Table 9: Correlation Coefficients and R^2 of TRMM SPI with Rain gauge SPI values

	Correlation		**R^2**	
SPI	*$TRMM_{1km}$*	*$TRMM_{28km}$*	*$TRMM_{1km}$*	*$TRMM_{28km}$*
SPI 3	0.37	0.57	0.14	0.32

SPI 6	0.35	0.58	0.13	0.34
SPI 9	0.37	0.6	0.14	0.36
SPI 12	0.42	0.66	0.17	0.43
SPI 24	0.34	0.62	0.11	0.39

Drought detection using TRMM whether downscaled or non-downscaled was found to be similar to that of rain gauge data detection in that drought occurrence was jointly detected in most instances by the three datasets. However TRMM data would either underestimate or overestimate the peaks / troughs of the identified droughts. This is similar to the results obtained by Hongwei et al., (2012) who reported that SPI 1 and SPI 3 identified droughts with similar skill to that of rain gauge data but with overestimation or underestimation of the drought peaks. However, the R2 values obtained by Hongwei et al., (2012) are much better ranging from 0.5 for most of the rain gauge stations analyzed though they also reported some stations with R2 less than 0.5. Li, et al., (2012) reported R2 values of between 0.5 – 0.83 when comparing TRMM data with rain gauge data, they also reported a similar observation to one made here of underestimation / overestimation in other instances of peaks and troughs of rain gauge data series by TRMM data. Zhang & Jia, (2013) reported a much larger variation of the R2 ranging from 0.7 – 0.20 when they correlated precipitation condition index to SPI.

4.2.Hydrological Droughts Analysis

4.3.1. Narumoro River

Hydrological droughts at Narumoro River were identified and analyzed using the theory of runs. The threshold was determined from the Flow Duration Curve (FDC) and was selected as flow exceeded 70% of the time, according to (Yahiaoui et al., 2009.) 70% - 90% exceedence are considered reasonable for drought extreme value analysis in perennial rivers. Any flow below the threshold was classified as drought flow, for Narumoro this flow is 0.353 m^3/s. Figure 12 shows the FDC curve of the river.

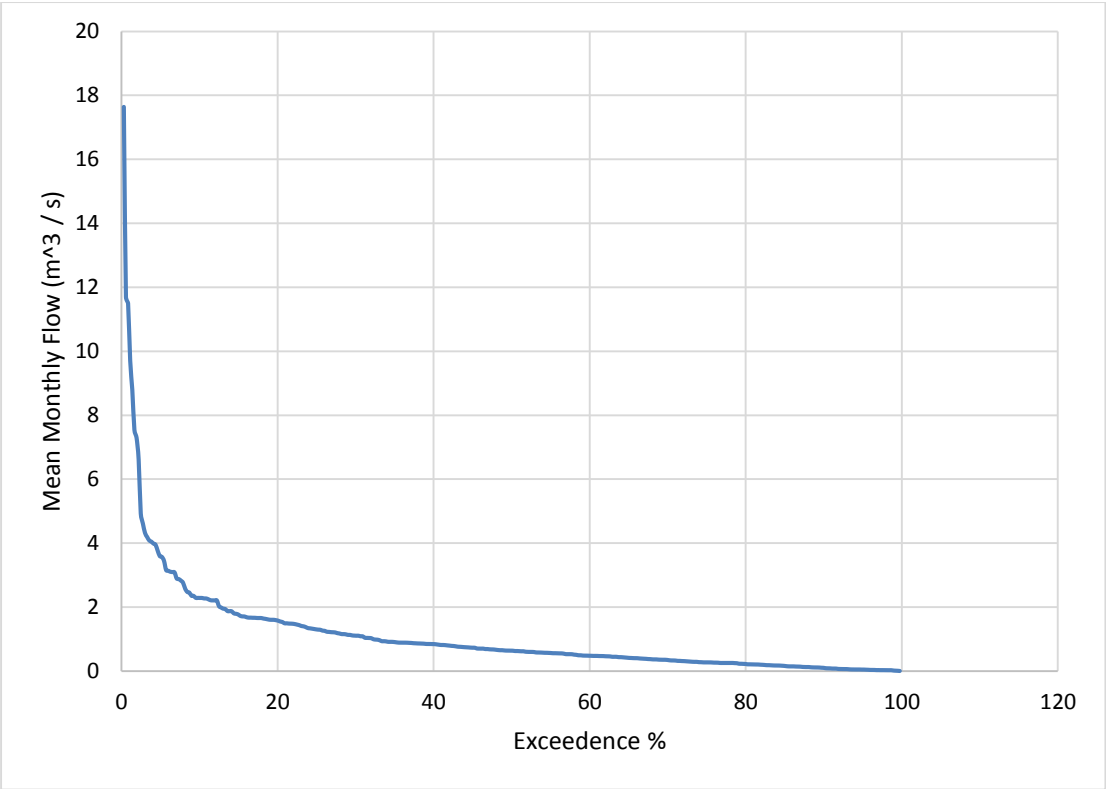

Figure 12: Narumoro River Flow Duration Curve

The figure 14 shows the identified drought based on the average monthly flows as determined at the River Gauging Station (RGS).

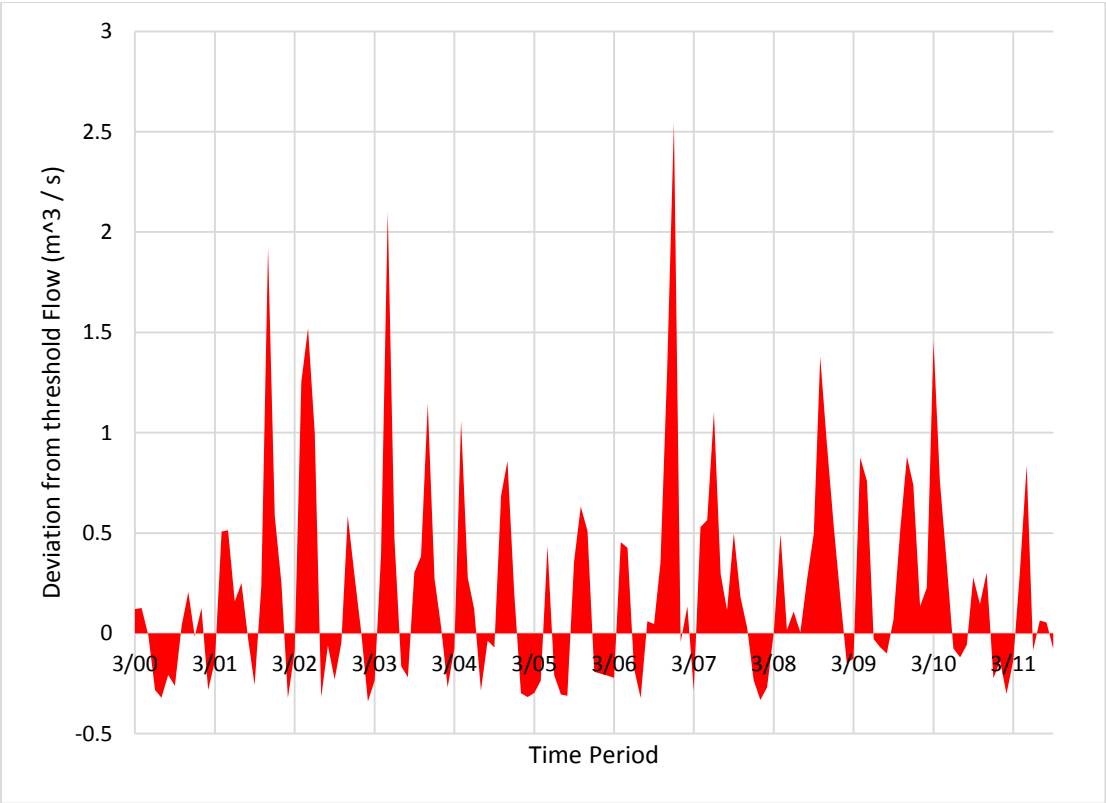

Figure 13: Hydrological Droughts from 2000 - 2011

Characteristics of the identified droughts were determined based on theory of runs, Table 10 lists the start and end times of drought events, drought lengths, accumulated deficit (summation of all deficits within the drought length), drought intensity(which is accumulated deficit divided by drought length) and the lowest flow recorded during the drought event. Figure 14 shows the distribution of low flows in the sub catchment during the study period, flows below 0.1m3 / s are most commonly low monthly flows

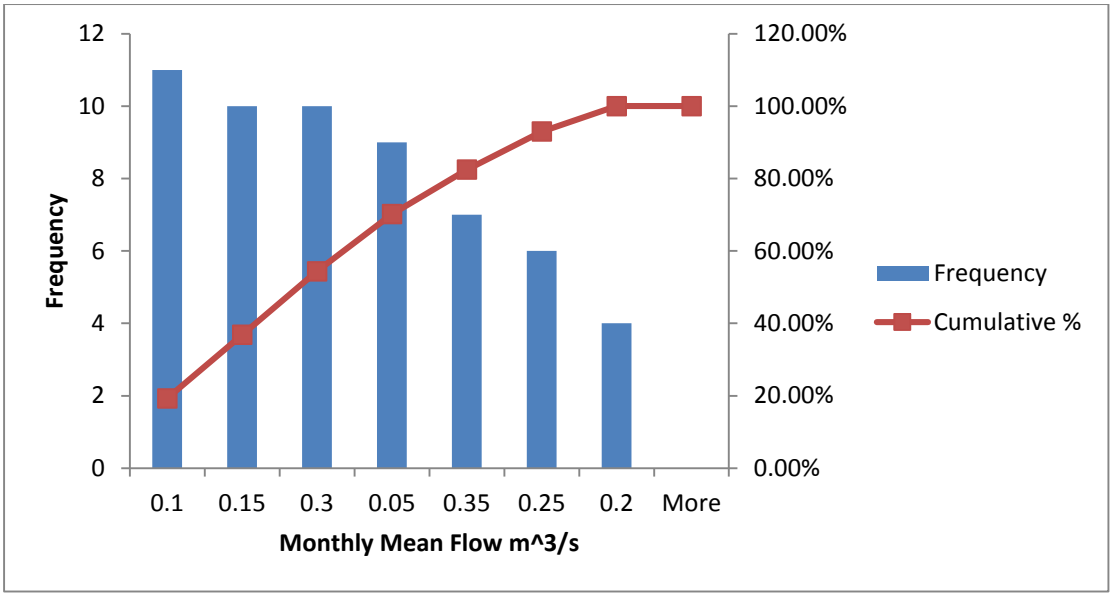

Figure 14: Distribution of low flows

Table 10: Hydrological Droughts Characteristics

Start	End	Length (Months)	Accumulated Deficit (m^3/s)	Intensity (m^3/s)	Lowest Flow (m^3/s)	Recovery Period (m^3/s)
May-00	Sep-00	5	-1.08	-0.22	0.03	14
Dec-00	Dec-00	1	-0.02	-0.02	0.34	1
Feb-01	Mar-01	2	-0.42	-0.21	0.07	1
Aug-01	Sep-01	2	-0.29	-0.14	0.09	2
Feb-02	Mar-02	2	-0.42	-0.21	0.03	1
Jul-02	Oct-02	4	-0.66	-0.16	0.03	2
Feb-03	Mar-03	2	-0.58	-0.29	0.01	2
Jul-03	Aug-03	2	-0.38	-0.19	0.13	2
Feb-04	Mar-04	2	-0.35	-0.17	0.08	1
Jul-04	Sep-04	3	-0.39	-0.13	0.07	1
Jan-05	Apr-05	4	-1.15	-0.29	0.03	13
Jun-05	Aug-05	3	-0.83	-0.28	0.04	2
Dec-05	Mar-06	4	-0.83	-0.21	0.13	2
Jun-06	Jul-06	2	-0.49	-0.25	0.03	4
Jan-07	Jan-07	1	-0.04	-0.04	0.31	1
Mar-07	Mar-07	1	-0.29	-0.29	0.07	1

Dec-07	Feb-08	3	-0.83	-0.28	0.02	6
Feb-09	Mar-09	2	-0.27	-0.14	0.20	1
Jun-09	Aug-09	3	-0.20	-0.07	0.25	2
Jun-10	Aug-10	3	-0.25	-0.08	0.23	1
Dec-10	Mar-11	4	-0.80	-0.20	0.05	2
Jun-11	Jun-11	1	-0.08	-0.08	0.27	2
Sep-11	Sep-11	1	-0.08	-0.08	0.28	

Figure 15 shows that the most droughts starts in June and February while ending mostly in march, August and September. The drought pattern is very responsive to the precipitation pattern meaning we can use precipitation to model streamflow.

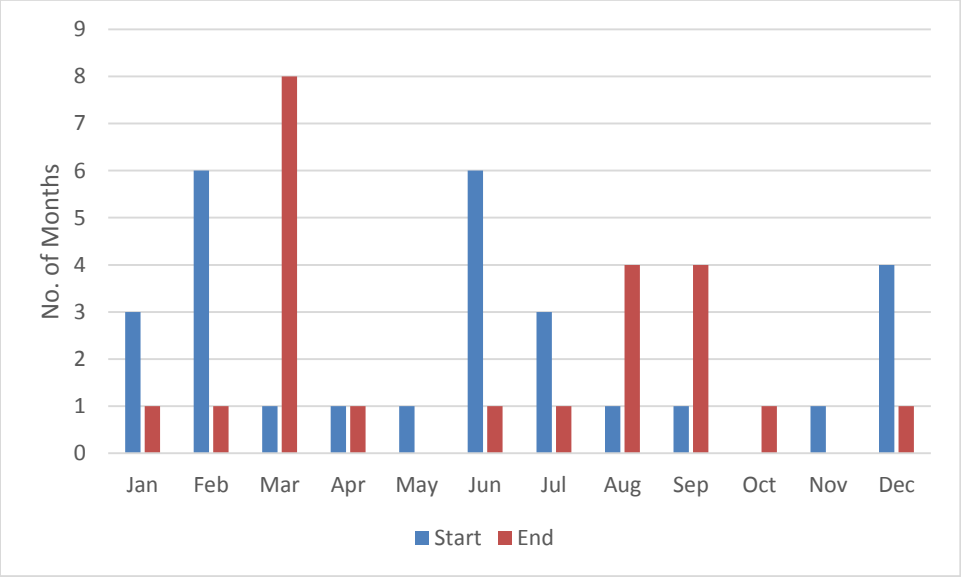

Figure 15: Droughts onset and cessation

4.4.Artificial Neural Networks (ANN) Models

Six ANN case studies were developed. The cases were defined by the inputs that were used to predict streamflow.

4.3.1 Case 1: Considering the relationship between TRMM and streamflow occurring at the same month.

A two layer *feedforwardnet* ANN model with 10 neurons was designed and trained using Levenberg-Marquardt algorithm. Several models were developed and trained, seven models that posted best results as determined by R Squared and MSE we selected for further testing. Using dataset set aside for testing (from October 2010 to September 2011), the ANN model was used to simulate the values of flows corresponding to the TRMM of the same period. The simulated flows were compared to the observed flows and similarity tests were conducted. This general methodology was used for all the six cases.

Figure 17 shows the relationship between the simulated and the observed flows for case 1.

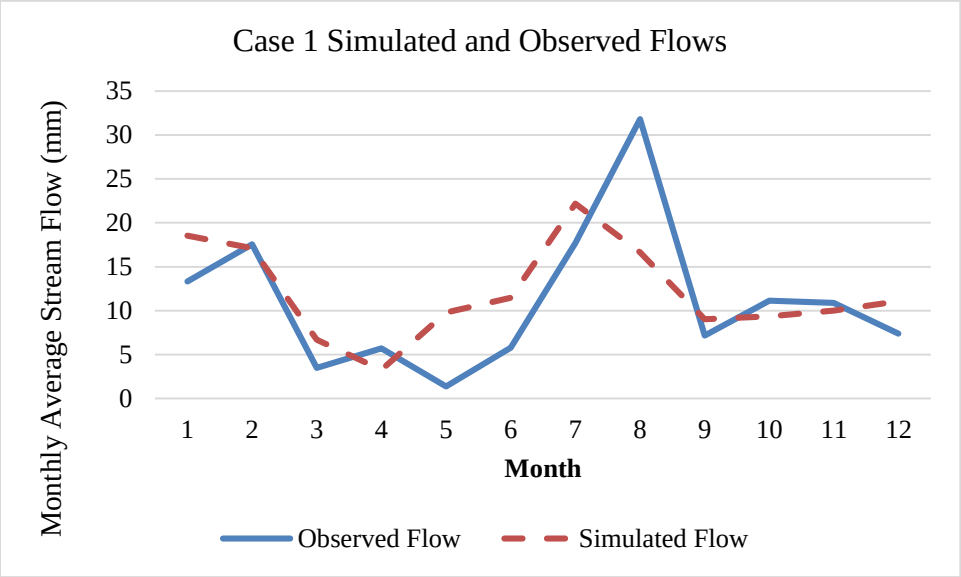

Figure 16: Simulated Flows and Observed Flows for Case 1

It can be noted from the Figure 16 above that the simulated flows closely follows the pattern of the observed flows. Due to the short observation record, it is difficult to derive enough drought events to characterize droughts identified from the simulated flows, it be deduced though, from the figure that the simulated flow is close to the average flow, slightly overestimating low flows and underestimating the high peaks.

Case 2: Forecasting flow based on previous month flow and TRMM

The ANN model was trained with a dataset composed of TRMM from the previous month, its corresponding flow and the TRMM of the current month. The simulated flows represent a 1 month forecast of the streamflow based on the

previous month flow, the previous month TRMM and the current month TRMM. This data is useful for monitoring purposes. As can be seen from Figure 18, the 1 month flow forecast follows a similar pattern to that of the observed flow indicating they are of similar character. Like in Case 1, this flow underestimate the high flows and slightly overestimate the low flows.

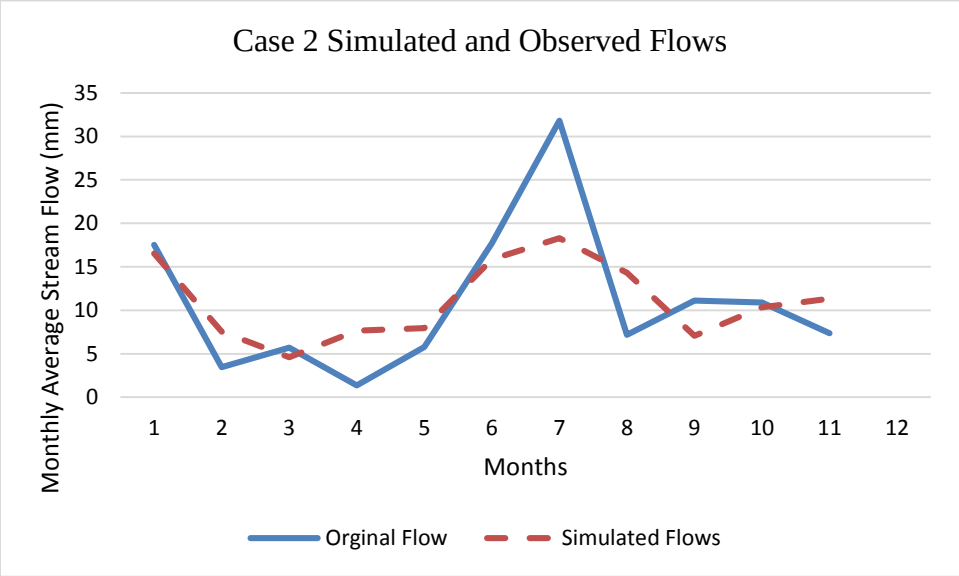

Figure 17: Simulated flows and observed flows for case2

Case 3: Forecasting flow based on previous 3 month flow and TRMM

The ANN Model was trained using a dataset composed of flows and corresponding TRMM from the previous 3 months and the TRMM of the current month to give a flow estimate with three month lead time. As can be seen from Figure 19 the simulated flows are similar to the observed flows. This dataset can be used a monitoring for the current situation or forecasting the flow with a 3 month lead time.

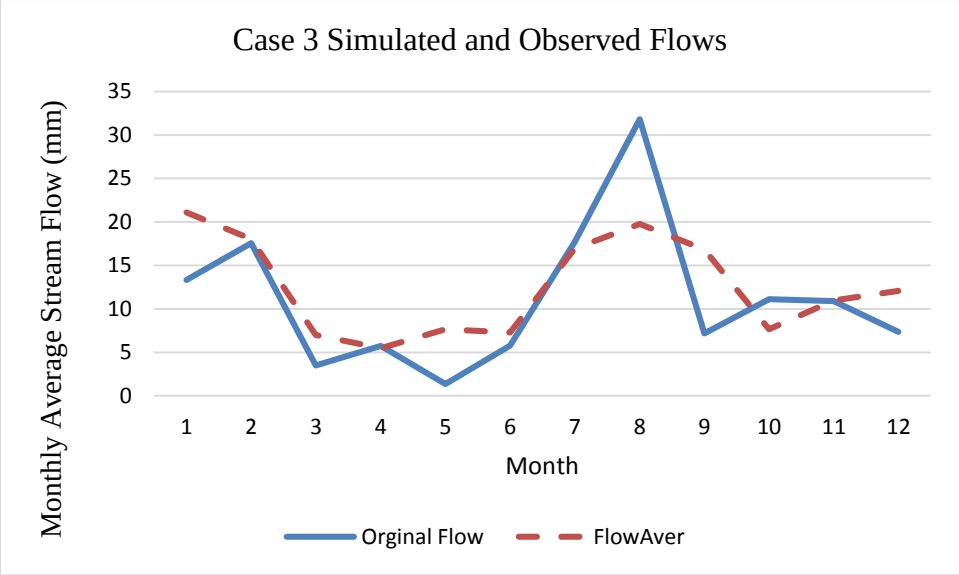

Figure 18: Simulated and Observed Flows for Case 3

Case 4: Forecasting flow based on previous 3 month TRMM only

The ANN Model was trained using a dataset composed of flows and corresponding TRMM from the previous 3 months to give a flow estimate with three month lead time. As can be seen from Figure 20 the simulated flows are similar to the observed flows. This dataset can be used a monitoring for the current situation or forecasting the flow with a 3 month lead time.

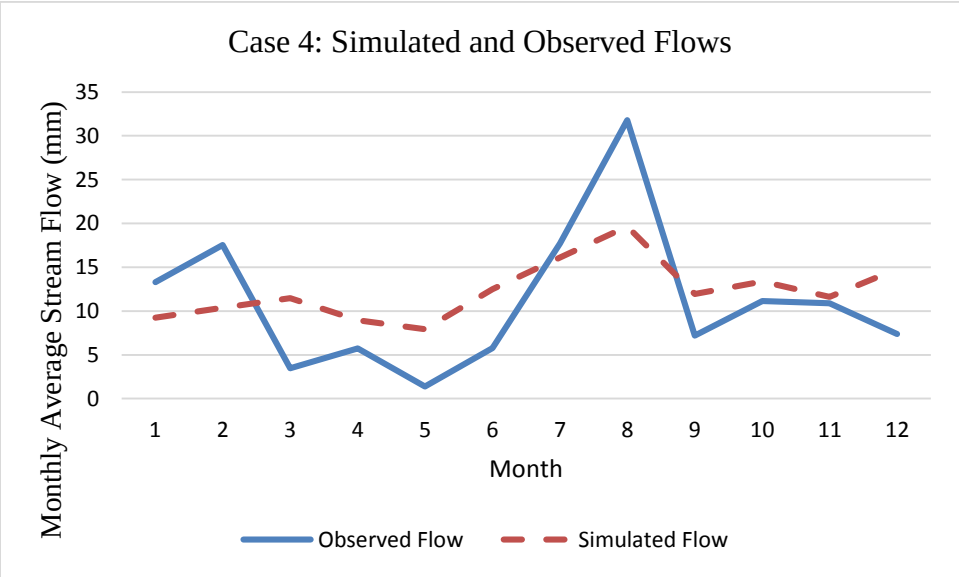

Figure 19: Simulated and Observed Flows for Case 4

Case 5: Forecasting flow based on previous 6 month TRMM only

From Figure 21, it can be seen the simulated flow based on previous 6 month TRMM only. The flow is similar to the observed flow in pattern but as with previous incidences, it overestimates the low flows and under estimate the higher flows.

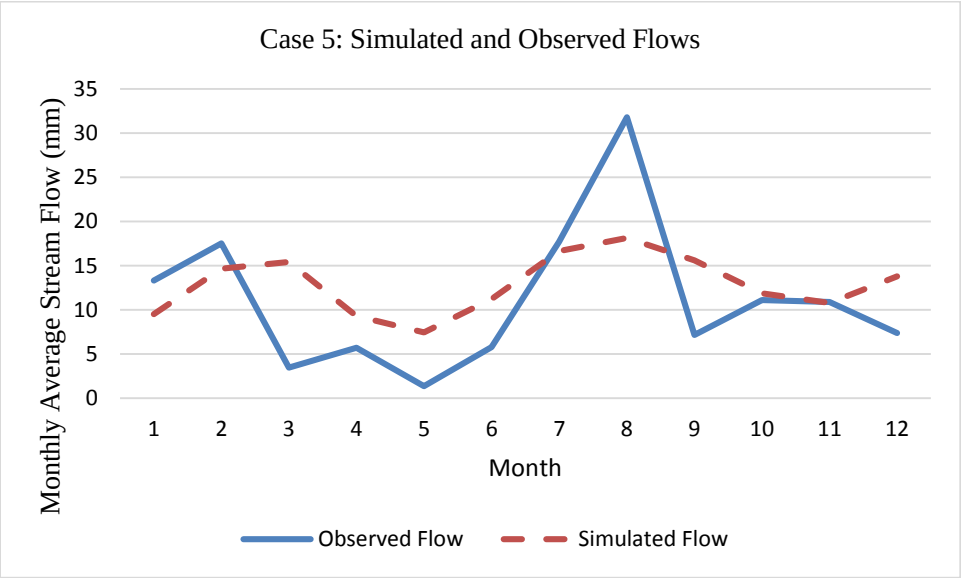

Figure 20: Simulated and Observed Flows for Case 5

It can also be observed that the simulated flow is adopting a smoother curve than that of the observed flow and does not pick local peaks and troughs. Therefore simulated flows using TRMM at 6 month lead time are more appropriate for giving a general "feel" of the future hydrological behavior of the river.

Case 6: Forecasting flow based on previous 6 month TRMM and Flow

Case 6 is similar to case 5 except that flow at 6 month lead time was added to the input variable of the ANN. Figure 22 shows the comparison of the observed and simulated flows for Case 6. It can be noted that the simulated flow appear to a have a slight forward lag, though the simulated flow pattern is much more similar to observed flow than in case 5.

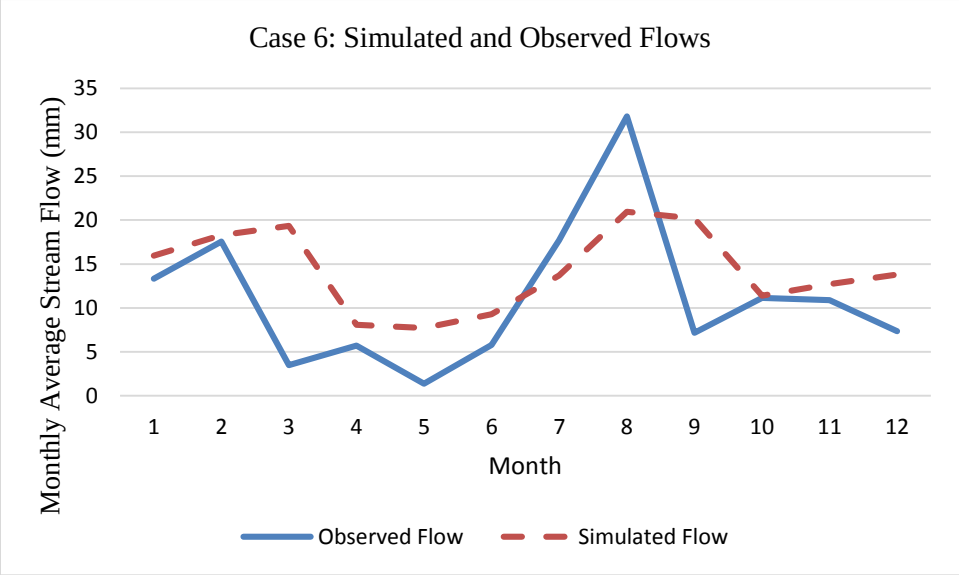

Figure 21: Simulated and Observed Flows for Case 6

Table 11 shows the statistical analysis results of observed flows versus the simulated flows. The results includes correlation coefficients, and coefficient of determination (R^2) while considering linear and non-linear regression. Non-linear regression using a non-linear 3rd order polynomial was found to represent the relationship between the observed and simulated flow better. Based on this results Cases 2, 5, and 4 performed better than the other cases. Case 5 and 4 used time lagged TRMM only as inputs, case 6 and case 3 were similar to Case 5 and 4 respectively but included time lagged flow as inputs. Case 2 performed very well despite using both TRMM and flow as inputs. It can be concluded that for longer term forecasting, time lagged TRMM performs better as input as compared to using both time lagged TRMM and flow, for short term forecasting for example 1 month, TRMM and Flow lagged at 1 month should be used.

Table 11: ANN Summary Statistics

Case	Correlation Coefficient	R^2 1st order	R^2 2nd order	R^2 3rd order
1	0.68	0.47	0.47	0.60
2	0.79	0.63	0.81	0.87
3	0.72	0.52	0.53	0.54

4	0.72	0.51	0.65	0.73
5	0.61	0.38	0.48	0.75
6	0.53	0.28	0.29	0.43

The ANOVA test was done on the simulated and observed flow to evaluate the variation of the simulated flows from the observed flow, Table 12 below shows the results of variation of the simulated flow from the observed flows for the various cases. The null hypothesis that the means of observed flow and simulated flow were similar was accepted. This result shows that the simulated flows are similar to the observed flows.

Table 12: ANOVA Table Summary

Case	*SS*	*df*	*MS*	*F*	*P-value*	*F critical*	*H_o Status*
1	5.864744	1	5.864744	0.119245	0.733134	4.30095	Accept
2	0.112613	1	0.112613	0.002347	0.961844	4.351244	Accept
3	12.76742	1	12.76742	0.253159	0.61986	4.30095	Accept
4	8.98049	1	8.98049	0.224913	0.639995	4.30095	Accept
5	18.62467	1	18.62467	0.466005	0.501956	4.30095	Accept
6	60.32846	1	60.32846	1.327962	0.261537	4.30095	Accept

The ANN model using TRMM as input was found to be adequate for streamflow prediction with up to six month lead time. This is similar to conclusions made by Li, et al., (2012) who concluded that TRMM data at monthly time scale is adequate for most hydrological applications. Jain & Kumar, (2007) used time lagged flow as inputs to ANN model, they reported best R^2 0.68 but on detrending the data, the R^2 improved to 0.8 during the training phase of the model. This result is comparable to the improvement in the R^2 when using higher order polynomials. Besaw, et al., (2008) reported a time lag in the predicted flow similar to that observed in case 6.

CHAPTER 5: Conclusions and Recommendation

This chapter summarizes the conclusions of this study and makes some recommendation on how the results could be applied or improved.

5.1 Conclusions

The results from this study can be used by water resources managers in planning water allocation and development of water resources projects. They are also useful for researchers interested in application of ANN and TRMM in hydrology and drought forecasting. The study concludes that:

Both downscaled and original TRMM are able identify droughts with similar characteristics to those identified by the rain gauge data. Similar number of total drought months are detected whether rain gauge data or TRMM data is used. The number of drought months of some drought classes vary slightly but are generally the same. TRMM downscaled or none downscaled datasets are able to detect droughts similar to rain gauge data but the TRMM data underestimates or overestimates the peaks of these droughts. None downscaled TRMM computed SPIs are more correlated to rain gauge data SPIs compared to the downscaled TRMM. Correlation coefficients ranges from 0.57-0.66 and 0.34-0.42 for non-downscaled TRMM and downscaled TRMM respectively. It is concluded that for meteorological drought monitoring, TRMM whether downscaled or at original resolution can be used, however using the two datasets together will be more beneficial as downscaled TRMM may detect some droughts that non-downscaled TRMM is unable to detect and vice versa.

Hydrological droughts in Narumoro river sub-catchment typically starts in the months of June and February and mostly cease in March. January, July and December are also a common drought onset months while August and September are common drought cession months. The most common drought flows ranges between 0.05 – 0.1 m^3/s and longest drought observed in the period lasted for 5 months. The drought pattern is similar to the precipitation pattern hence precipitation can be used to model stream and hydrological droughts.

ANN models can be used together with TRMM to predict streamflow with up to 6 months lead time in the Narumoro river sub-catchment with good accuracy. For streamflow forecasting with lead time of 3 or 6 months, time lagged TRMM as input is superior to combining both TRMM and time lagged flow. The relationship between the observed and simulated flow can be represented using a cubic polynomial. Using this relationship, a streamflow model at 1 month lead time using time lagged TRMM and flow had R2 of 0.86 which was the highest. The least R2 obtained was 0.43, which was of a 6 month lead time model using both TRMM and flow.

5.2 Recommendation

Following the study, the following recommendation are made:

1. Further studies to be conducted to especially compare the inter catchment transfer of ANN models. This will be important especially in solving the data problems encountered in the ungauged sites.
2. This study showed that for meteorological drought, either downscaled or original resolution Tropical Rainfall Measuring Mission data can be used, further studies should be done to evaluate if this result is reproducible to other small catchments.
3. ANN models can be developed and used by water managers to plan water resources use at the sub-catchment level.
4. The relationship between the observed and simulated flow to be further modelled. If a universal equation is obtained, it can be used to adjust the simulated flows to make to be closer to the observed increasing the accuracy of drought detection.

References

Aqil, M., Kita, I., Yano, A. & Nishiyama, S., 2007. A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behaviour of runoff. *Journal of Hydrology,* pp. 22-34.

Caccamo, G., Chisholm, L., Bradstock, R. & Puotinen, M. L., 2011. Assessing the sensitivity of MODIS to monitor drought in high biomass ecosystems. *Remote Sensing of Environment,* p. 2626–2639.

Collischonn, B., Collischonn, W. & Tucci, C. E. M., 2008. Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. *Journal of Hydrology ,* p. 207– 216.

Duan, Z. & Bastiaanssen, W. M., 2013. First results from Version 7 TRMM 3B43 precipitation product in combination with a new downscaling–calibration procedure. *Remote Sensing of Environment,* pp. 1-14.

Imrie, C., Durucan, S. & Korre, A., 2000. River flow prediction using artiacial neural networks: generalisation beyond the calibration range. *Journal of Hydrology,* pp. 138-153.

Mutiga, J. K. et al., 2010. Water Allocation as a Planning Tool to Minimise Water Use Conflicts in the Upper Ewaso Ng'iro North Basin, Kenya. *Water Resources Management,* p. 3939–3959.

Srinivasulu, S. & Jain, A., 2006. A comparative analysis of training methods for artificial neural network rainfall–runoff models. *Applied Soft Computing,* p. 295–306.

Walczak, S. & Cerpa, N., 2003. Artificial Neural Networks. In: *Encyclopedia of Physical Science and Technology (Third Edition).* s.l.:Academic Press, p. 631–645.

Adamowski, J. & Sun, K., 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology,* pp. 85-91.

Adeloye, A. & Munari, A. D., 2006. Artificial neural network based generalized storage–yield–reliability models using the Levenberg–Marquardt algorithm. *Journal of Hydrology,* 326(1-4), p. 215–230.

Almazroui, M., 2011. Calibration of TRMM rainfall climatology over Saudi Arabia during 1998–2009. *Atmospheric Research,* p. Atmospheric Research.

Belal, A.-A.et al., 2012. Drought risk assessment using remote sensing and GIS techniques. *Arab J Geosci.*

Besaw, L. E., Rizzo, D. M., Bierman, P. R. & Hackett, W. R., 2008. Advances in ungauged streamflow prediction using artificial neural networks. *Journal of Hydrology,* Volume 386, p. 27–37.

Bonaccorso , B., Peres , D. J., Cancelliere , A. & Rossi, G., 2013. Large Scale Probabilistic Drought Characterization Over Europe. *Water Resources Management Vol. 27,* p. 1675–1692.

Bowden, G. J., Dandy, G. C. & Maier, H. R., 2005. Input determination for neural network models in water resources applications. Part 1—background and methodology. *Journal of Hydrology,* p. 75–92.

Chiu, L. S., Liu, Z., Rui, H. & Teng, W. L., 2006. Tropical Rainfall Measuring Mission Data and Access Tools. In: *Earth Science Satellite Remote Sensing Vol. 2: Data, Computational Processing, and Tools.* Berlin: Springer-Verlag GmbH, pp. 202-218.

Cigizoglu, H., 2008. ARTIFICIAL NEURAL NETWORKS IN WATER RESOURCES. In: *Integration of Information for Environmental Security.* s.l.:Springer., pp. 115-148.

CIGIZOGLU, H., 2008. ARTIFICIAL NEURAL NETWORKS IN WATER RESOURCES. In: *Integration of Information for Environmental Security.* s.l.:Springer., pp. 115-148.

Coulibaly, P., Anctil, F. & Bobée, B., 2000. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *Journal of Hydrology,* 230(3-4), p. 244–257.

Dai, A., 2010. Drought under global warming: a review. *Interwiley Reviews: climate change.*

Daliakopoulos, I. N., Coulibaly, P. & Tsanis, I. K., 2005. Groundwater level forecasting using artificial neural networks. *Journal of Hydrology,* 309(1-4), p. 229–240.

Duncan, J. M. & Biggs, E. . M., 2012. Assessing the accuracy and applied use of satellite-derived precipitation estimates over Nepal. *Applied Geography,* pp. 626-638.

ESRI, 2014. *Arcgis.* [Online]
Available at: http://www.esri.com/software/arcgis
[Accessed 9 September 2014].

Fang, J. et al., 2013. Spatial downscaling of TRMM precipitation data based on the orographical effect and meteorological conditions in a mountainous area. *Advances in Water Resources 61,* p. 42–50.

Fleig, A. K., Tallaksen, L. M., Hisdal, H. & Demuth, S., 2006. A Global Evaluation of Streamflow Drought Characteristics. *Hydrology and Earth System Sciences vol 10,* pp. 535-552.

Funk, C., Ederer, G. & Pedreros, D., 2010. *TROPICAL RAINFALL MEASURING MISSION,* s.l.: s.n.

Gao, C. et al., 2010. Projected streamflow in the Huaihe River Basin (2010–2100) using artificial neural network. *Stoch Environ Res Risk Assess,* pp. 685-697.

ha y e s , M. c. h. a. e. l., sv o b o d a , M. r. k., Wa l l , N. c. o. l. e. & Wi d h a l M, M. l. i. s. s. a., 2010. *THE LINCOLN DECLARATION ON DROUGHT INDICES,* Lincoln: American Meteorological Society.

Hongwei, Z., Lijiuan, L. & Jiuyi, L., 2012. The evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) in drought monitoring in the Lancang River Basin. *Journal of Geographical Sciences,* 22(2), pp. 273-282.

Immerzeel, W., Rutten, M. & Droogers, P., 2009. Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula. *Remote Sensing of Environment,* p. 362–370.

Jain, A. & Kumar, M. A., 2007. Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing,* Volume 7, p. 585–592.

Jia, S., Zhu, W., Lű, A. & Yan, T., 2011. A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam Basin of China. *Remote Sensing of Environment,* p. 3069–3079.

Khan, S., Dassanayake, D. & Gabriel, H. F., 2010. AN ADAPTIVE LEARNING FRAMEWORK FOR FORECASTING SEASONAL WATER ALLOCATIONS IN IRRIGATED CATCHMENTS. *NATURAL RESOURCE MODELING,* pp. 325-355.

Kişi, Ö., 2007. Streamflow Forecasting Using Different Artificial Neural Network Algorithms. *JOURNAL OF HYDROLOGIC ENGINEERING,* 12(5), pp. 532-539.

Li, X.-H., Zhang, Q. & Xu, C.-Y., 2012. Suitability of the TRMM satellite rainfalls in driving a distributed hydrological model for water balance computations in Xinjiang catchment, Poyang lake basin. *Journal of Hydrology,* Volume 22, p. 426–427.

Majumder, M., Roy, P. & Mazumdar, A., 2010. A Generalized Overview of Artificial Neural Network and Genetic Algorithm. In: *Impact of Climate Change on Natural Resource Management.* s.l.:Springer Science+Business Media B.V, pp. 393-416.

Mathworks, 2014. *Neural Network Tool Users Guide.* s.l.:Mathworks Inc.

McKerchar, A. I., 2000. *Streamflow.* s.l.:National Institute of Water and Atmospheric Research.

Mishra, A. K. & Singh, V. P., 2010. A Review of DroughtConcepts. *Journal of Hydrology vol 391,* pp. 201-216.

Mishra, A. K. & Singh, V. P., 2011. Drought Modelling A review. *Journal of Hydrology 403,* pp. 157-175.

Motalleb , B., Saghafian, B., Mohammadi, K. & Siosemarde, M., 2014. Regional analysis of streamflow drought: a case study in southwestern Iran. *Environment and Earth Scincies,* p. 2955–2972.

Mutlu, E., Chaubey, I., Hexmoor, H. & Bajwa, S. G., 2008. Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed. *HYDROLOGICAL PROCESSES.*

Naresh, R., Sharma, V. & Yadav, D., 2011 . Stream flow forecasting using Levenberg-Marquardt algorithm approach. *International Journal of Water Resources and Environmental Engineering* , Vol. 3(1), pp. 30-40.

NASA, 2014. *TRMM.* [Online]
Available at: http://disc.sci.gsfc.nasa.gov/TRMM
[Accessed 9 September 2014].

National Drought Mitigation Center , 2014. *Program to Calculate Standardized Precipitation Index.* [Online]
Available at:
http://drought.unl.edu/MonitoringTools/DownloadableSPIProgram.aspx
[Accessed 9 July 2014].

Nyabeze, W. . R., 2004. Estimating and interpreting hydrological drought indices using a selected catchment in Zimbabwe. *Physics and Chemistry of the Earth 29,* p. 1173–1180.

Pandey , R. P., Mishra , S. K., Singh , R. & Ramasastri, K. S., 2008. Streamflow Drought Severity Analysis of Betwa River System (India). *Water Resoures Management,* p. 1127–1141.

Serinaldi, F., Bonaccorso, B., Cancelliere, A. & Grimaldi, S., 2009. Probabilistic characterization of drought properties through copulas. *Physics and Chemistry of the Earth,* p. 596–605.

Shaban, A., 2009. Indicators and Aspects of Hydrological Drought in Lebanon. *Water Resources Management vol 23,* pp. 1875-1891.

Smakhtin, V., 2001. Low flow hydrology: a review. *Journal of Hydrology 240 ,* p. 147–186.

Smakhtin, V. & Hughes, D., 2004. *Review, automated estimation and analyses of drought indices in south asia,* Colombo: International Water Management Institute.

Sung, J. H. & Chung, E.-S., 2004. Development of Streamflow Drought Severity-Duration-Frequency Curve Using Threshold Level Method.

Tallaksen, L. M., Hisdal, H. & Van Lanen, H. A., 2009. Space–time modelling of catchment scale drought characteristics. *Journal of Hydrology,* Volume 375 , p. 363–372.

The National Drought Mitigation Center, 2014. *Program to Calculate Standardized Precipitation Index.* [Online]
Available at:
http://drought.unl.edu/MonitoringTools/DownloadableSPIProgram.aspx
[Accessed 4 September 2014].

UN, 2006. *Global Survey of Early Warning,* s.l.: United Nations.

UNEP, 2006. *Early Warning Systems: A State of the Art Analysis and Future Directions. Division of Early Warning and Assessment (DEWA),* Nairobi: United Nations Environment Programme.

Unidata, 2014. *Network Common Data Form (NetCDF).* [Online]
Available at: http://www.unidata.ucar.edu/software/netcdf/
[Accessed 9 September 2014].

Vicente-Serrano, S. M. et al., 2012. Challenges for drought mitigation in Africa: The potential use of geospatial data and drought information systems. *Applied Geography*, pp. 471-486.

VITO, 2013. *SPOT Vegetation Programme.* [Online]
Available at: http://www.spot-vegetation.com/
[Accessed 9 September 2014].

Wilhite, D. A., Svoboda, M. D. & Hayes, M. J., 2007. Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. *Water Resour Manage* , Volume 21, p. 763–774.

WMO, 2006. *Drought Monitoring and Early Warning: Concepts, Progress and Future Challenges,* s.l.: World Meteological Organization.

WRMA, 2013. *Water Resources Management Authority.* [Online]
Available at: http://wrma.or.ke

Yahiaoui, A., Touaïbia, B. & Bouvier, C., 2009.. Frequency analysis of the hydrological drought regime. Case of oued Mina catchment in western of Algeria. *Revue Nature et Technologie,* pp. 3-5.

Zhang, A. & Jia, G., 2013. Monitoring meteorological drought in semiarid regions using multi-sensor microwave remote sensing data. *Remote Sensing of Environment,* Volume 134, pp. 12-23.

Appendices

ANN Model Script

```
% Solve an Input-Output Fitting problem with a Neural Network
% Script prepared by Mutuga Kigumi
% Created Mon Sep 08 16:03:52 EAT 2014
%
% This script assumes these variables are defined:
%
%   ALLtrmm - input data.
%   flow - target data.

inputs = Alltrmm;
targets = flow;

% Create a Fitting Network
hiddenLayerSize = 15;  % Number of the neurons in the hiden layer
net = fitnet(hiddenLayerSize); % Create Network with default
properties

% Choose Input and Output Pre/Post-Processing Functions
% 'removeconstantrows' removes constant rows that do not add value
% 'mapminmax' normalizes data to be between -1 and 1 appropriate
for
% model use
net.inputs{1}.processFcns = {'removeconstantrows','mapminmax'};
net.outputs{2}.processFcns = {'removeconstantrows','mapminmax'};


% Setup Division of Data for Training, Validation, Testing
net.divideFcn = 'dividerand';  % Divide data randomly
net.divideMode = 'sample';  % Divide up every sample
net.divideParam.trainRatio = 70/100; %ratios for dividing up the
data
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;

%Select the training alogarithim
net.trainFcn = 'trainlm';  % Levenberg-Marquardt

% Choose a Performance Function
net.performFcn = 'mse';  % Mean squared error
net.trainParam.goal = 0.1;

% define parameters controlling training rate
net.trainParam.mu =1;
net.trainParam.mu_dec = 0.8;
net.trainParam.mu_inc = 1.5;
net.trainParam.max_fail = 10;
net.trainParam.epochs = 1000;
net.trainParam.goal = 2;

% Choose Plot Functions
% For a list of all plot functions type: help nnplot
net.plotFcns = {'plotperform','plottrainstate','ploterrhist', ...
  'plotregression', 'plotfit'};
```

```
% Train the Network
[net,tr] = train(net,inputs,targets);

% Test the Network
outputs = net(inputs);
errors = gsubtract(targets,outputs);
performance = perform(net,targets,outputs)

% Recalculate Training, Validation and Test Performance
trainTargets = targets .* tr.trainMask{1};
valTargets = targets  .* tr.valMask{1};
testTargets = targets  .* tr.testMask{1};
trainPerformance = perform(net,trainTargets,outputs)
valPerformance = perform(net,valTargets,outputs)
testPerformance = perform(net,testTargets,outputs)

% View the Network
view(net)

% Plots
figure, plotperform(tr)
figure, plottrainstate(tr)
figure, plotregression(targets,outputs)

%End
```