Mosaad Khadr

Water Resources Management in the Context of Drought

(An Application to the Ruhr river basin in Germany)

Vorwort (Hrsg.)

Trockenperioden sind Naturereignisse, welche sich über einen längeren Zeitraum und eher schleichend entwickeln. Die frühzeitige Wahrnehmung stellt daher eine wesentliche Voraussetzung für ein angemessenes Handeln dar. In der vorliegenden Arbeit wird als Beispielregion die Ruhr in Nordrhein-Westfalen gewählt, wobei der Fokus auf den Talsperren des Ruhrverbandes liegt. Die Ruhr wird seit der Industrialisierung stark zur Wasserversorgung des Ruhrgebietes genutzt, wobei der saisonale Ausgleich durch die Bewirtschaftung der Talsperren sichergestellt wird. Trockenperioden kehren, wie die Analyse zeigt, auch an der Ruhr in regelmäßigen Abständen wieder und zwingen die Entscheidungsträger die Abgabesteuerung an der saisonalen und witterungsgegebenen Situation, dem Bedarf sowie dem Speicherfüllungsgrad und der wahrscheinlichen weiteren Entwicklung zu orientieren. Dies gilt insbesondere für extreme Ereignisse.

Die vorliegende Dissertation beschäftigt sich mit der Fragestellung, wie das wasserwirtschaftliche Management auf Perioden von Trockenheit bzw. Wassermangel reagieren kann. Dabei wird eine einfach zu handhabende Methode entwickelt, um auf der einen Seite die Intensität der Trockenheit zu klassifizieren und auf der anderen Seite eine Grundlage für eine Entscheidungsfindung zu schaffen. Der gewählte SPI Index (Standardized Precipitation Index) basiert hierbei ausschließlich auf den in der Vergangenheit gemessenen Niederschlägen. Weiterhin wird auf die Vorhersehbarkeit von Trockenheit mittels SPI Index eingegangen. Eine Vorhersage wird unter Verwendung eines ARMA Modells (Auto Regressive Moving Average) entwickelt. Zudem wird ein stochastisches Simulationsmodell für monatliche Talsperrenzuflüsse aufgestellt.

Die vorliegende Arbeit zeigt, dass der SPI Index über mehrere Monate akkumuliert werden kann, so dass sich der Grad der Trockenheit sowohl auf eine Region als auch auf eine Dauer beziehen lässt. Bei der Analyse mittels SPI Index muss allerdings beachtet werden, dass eine Trockenperiode nur relativ zu der Vorgeschichte in der untersuchten Region als trocken

bezeichnet wird. Die vorgestellte Methode ist unabhängig von Regionen, da sie auf diesen Relativbetrachtungen gegenüber dem langjährigen Verlauf basiert.

Wuppertal, April 2011   *Andreas Schlenkhoff*

BERGISCHE UNIVERSITÄT WUPPERTAL

WATER RESOURCES MANAGEMENT IN THE CONTEXT OF DROUGHT
(AN APPLICATION TO THE RUHR RIVER BASIN IN GERMANY)

Vom Fachbereich D (Abteilung Bauingenieurwesen)
der Bergischen Universität Wuppertal

genehmigte

Dissertation

zur Erlangung des akademischen Grades
DOKTOR-INGENIEUR (Dr.-Ing.)

von
M. Sc. Mosaad Bayoumi Desoky Ahamed Khadr
aus Tanta - Ägypten

Acknowledgments

Abstract

During the last decades water resources managers are facing severe challenges all over the world and the trends of increasing temperature and decreasing precipitation intensify this situation. Climate change is a major global challenge facing water resources managers. Rising global temperatures will lead to an intensification of the hydrological cycle, resulting in dryer dry seasons and wetter rainy seasons, and subsequently heightened risks of more extreme, longer and frequent floods and droughts. Drought is considered by many to be the most complex but least understood of all natural hazards, affecting more people than any other hazard. Drought is a natural hazard temporarily affecting almost every region in the world. The main target of this thesis is to provide some analyses and to evolve appropriate and interdisciplinary tools and techniques for drought characterization and for enhanced management of water resources systems during drought periods. The proposed methodologies are applied to the Ruhr river basin as a case study.

In this thesis, the climate change in the Ruhr river basin has been investigated using a set of data containing precipitation, temperature and inflow. All data series have been subjected to homogenization procedure. The data homogenization is described in detail. Yearly and seasonal trend analyses have been performed on all data series using the Mann- Kendall test. The frequency distributions of warm/cold days and very/extremely wet days have been examined using percentile indices.

Results of the hydrological analysis showed that a significant increase in the mean temperature is considered over all time scales in the study area. The occurrence of warm days in both winter and summer has a significant increase while the occurrence of cold days in both seasons showed a similar proportion of significant decrease. These results give evidence that the winter becomes warmer and the summer becomes hotter.

Results of the precipitation analysis give evidence on a significant increase in winter precipitation while the increases in summer and the annual precipitation were statistically insignificant. The number of consecutive dry days displayed decreasing tendencies in winter while there is no indication of statistically significant change in the summer. Analysis of very & extremely wet days showed that the main identified trends are an increase of the very wet days in the winter. For the inflow analysis, the results showed that there is a significant increase in winter inflow while the increases in summer and annual inflow were found to be statistically insignificant. Correlation calculations, which have been applied to the data series, showed that variations of streamflow from year to year were much more strongly related to precipitation changes than to temperature changes; this is corresponding to actual common results in hydrological research.

Drought is a normal, recurrent feature of climate and is a complex phenomenon and generally viewed as a sustainable and regionally extensive occurrence of below-average natural water availability either in the form of precipitation, river runoff or groundwater. The meteorological drought in the Ruhr river basin has been investigated using the Standardized Precipitation Index (SPI). The Standardized Precipitation Index aims to provide a concise overall picture of drought, regardless of the actual probability distribution of the observed cumulative amounts of rainfall for a given time scale. By applying the SPI methodology, the obtained results indicated that the drought randomly affected the Ruhr river basin and several drought events occurred during the period under study. Results also indicated that although the significant positive trend in winter precipitation drought visited the Ruhr basin in both summer and winter and that the most severely event was in the winter. Trends in SPI data series have been examined using the Mann-Kendall test. Results of trend analysis indicated that the proportion of drought condition has changed insignificantly during the period under study the Ruhr catchment.

Since the calculations and the analysis of the standardized precipitation index (SPI) are complex and not so easy to be done with several precipitation time series, software with a friendly and interactive graphical user interface (GUI) for SPI calculations and analysis has been developed in MATLAB environment. The main objectives of the program are: calculation of the SPI values for a given precipitation data series; detection whether a drought event exists in a data series using several time steps and classifications of the drought events according to its intensity (moderate, severe, extreme). The developed program makes the analysis of the SPI easier compared with the program which is used by the National Drought Mitigation Center (USA), rather than the developed program has more possibilities.

Drought forecasting is an essential tool for implementing appropriate mitigation measures in order to reduce negative impacts of drought on water resources systems. The SPI index has been used as a drought indicator for drought forecasting due to its many advantages compared to other drought indices. The capability of the Auto Regressive Integrated Moving Average (ARIMA) model in drought forecasting has been investigated using the correlation methods of Box and Jenkins and the AIC and SBC structure selection criteria. ARIMA models are, in theory, the most general class of models for forecasting a time series which can be stationarized by transformations such as differencing and logging. Validation of the forecasting models has been carried out by comparing SPI values computed on observed precipitation and the corresponding forecasts. Results showed a fairly good agreement between observations and forecasts, as it has also been confirmed by the values of some performance indices. Results also showed that, the good fitting of stochastic models such as ARIMA to hydrologic time series such as SPI time series could result in better tool that can be used for water resource planning.

The forecasting of the standardized precipitation index (SPI) using stochastic models, such as ARIMA, is a complex procedure to be performed on several SPI data series. There are many statistical software packages, like SAS and SPSS, which are used for time series forecasting. In this thesis, a software package, containing ARIMA and multiplicative Seasonal Auto Regressive Integrated Moving Average (SARIMA) models, has been developed. This program has several advantages compared with the other statistical software, when ARIMA model is considered. First of all is simplicity compared with other programs which need an experienced user. The user puts limits for the model's parameters then the program optimizes these parameters to detect the best candidate model. The predicted results using the developed software have been compared with the observed data and with the predicted values obtained by using the well known software SPSS in case of ARIMA and SARIMA models. The results of the calibration showed good agreement between the forecasted values using the developed program and those which were obtained using the software SPSS with reasonable accuracy.

Stochastic simulation of hydrologic time series has been widely used for solving various problems associated with the planning, management and operational purposes for several decades. In this thesis, the stochastic streamflow generation model of Thomas-Fiering and a Monte Carlo simulation model have been applied to generate synthetic monthly inflow scenarios for four reservoirs in the Ruhr river basin. New method has been proposed to preserve the statistical parameters of the random part in the Thomas-Fiering model. Comparison of the main statistical parameters such as mean, standard deviation and skewness has been done for both historical and generated data by the proposed models. The results showed that, the generated data series have successfully preserved the historical statistical parameters of streamflow. The results showed also that, the Thomas-Fiering model has preserved the correlation coefficient between consecutive months. Thus, the Thomas-Fiering model was suitable to be used for producing inflow scenarios needed for the optimization model and stochastic simulation model presented in this thesis.

Reservoir operation is a complex problem that involves many decision variables, multiple objectives as well as considerable risk and uncertainty. In addition, the conflicting objectives lead to significant challenges for operators when making operational decisions. Reservoir operation for an optimal use of available water during prolonged periods of drought has always been a primary concern for water management. Using Genetic Algorithm, Pattern Search and Gradient-Based methods, an optimization model has been developed for the operation of reservoirs during normal periods and drought periods as well. The reservoir Bigge has been presented as case study. Two objective functions have been considered, then a weighted approach has been adopted to convert the multiple objectives problem into a single objective problem so that the user can specify the priorities by giving a specified weight for each function. Several scenarios for low inflow periods have been attempted. Each scenario has its assumptions for monthly inflow and monthly demand.

The evaluation of the model has been carried out using the driest year in the available historical records. The monthly inflow of this year has been considered as an input to the optimization model. Results of the evaluation demonstrated that, the developed model is beneficial. Results also showed that the developed model with its several scenarios and the suggested optimization approaches could be helpful for the real life operation of the reservoir.

In reservoir management practices, a simulation model can be used as a valuable planning tool to evaluate the impact of changes to the system's configuration or operational objectives. The desired generation or release scheduling can be checked using inflow forecasting in order to satisfy the entire set of operational constraints. At real-time operation stage, a simulation tool can be used to quickly check operational alternatives due to emergency events or planning and real-time incongruence. Fuzzy set theory plays an important role in dealing with uncertainty when making decisions in reservoirs operation. In this thesis, an example of the collective use of stochastic models and adaptive network-based fuzzy inference system (ANFIS) for reservoir operation and simulation has been presented. ANFIS provides a method for fuzzy modeling to learn information about the data set that best allow the associated fuzzy inference system to trace the given input/output data. The applicability and capability of the ANFIS model have been investigated through the use of a set of data in the Ruhr reservoirs system. The historical data are time of year (months), inflow, reservoir storage, SPI index and reservoir release. The historical data sets have been divided into two independent sets to train and to test the constructed models.

Two main models have been developed. In both models the set of input include time of year, storage, inflow and Standardized Precipitation Index (SPI). The output of the first model is the release during the next month; on the other hand, the output of the second model is the release of the current month. Predicted release and observed release values have been evaluated using several evaluation criteria. Results of the evaluation showed that the ANFIS models are accurate and consistent in different subsets. In order to demonstrate that the effect of using SPI index as input, two ANFIS models have been developed and investigated; one with SPI as input variable, another without. It has been found that the model which contains SPI as input variable has consistently superior performance compared with the one without SPI index. The results showed that the ANFIS models provide reliable reservoir release prediction for the current and the next month. Results showed also that the proposed approach could be a good tool for the evaluation of release for a specified month and could be also a helpful reference guide to the operator during making decisions.

Historical records demonstrate that droughts are causing potential impacts. The risk of these potential impacts depend on the type of water demand, how these demands are met and the corresponding water supplies available to meet these demands. These impacts could be categorized into domestical, agricultural, environmental, industrial and recreational impacts. The Ruhr basin is exposed to drought hazard rather frequently. Results of drought analysis in the Ruhr basin demonstrate that severe and extreme events occurred in 1959, 1976, 1996, 2003 and 2007.

Preparing an efficient drought management plan is the best way to reduce drought impacts. These impacts could be continued to several weeks or months even after the drought event. In this thesis, a drought management plan is proposed and the procedures of this plan have been applied to case studies. The analysis of the case studies showed that the implementation of the actions of each stage of drought stages is very important to address drought impacts and to prevent reservoir from being drained. Analysis of the case studies showed also that the use of a transition probability matrix can be a useful guide for decision makers during dry periods. In this study the actions which have been qualified only are monthly release of the Bigge reservoir.

x

Deutsche Kurzfassung

Während der letzten Dekaden hat die Wasserwirtschaft global mit zunehmend schwierigeren Herausforderungenin der Wasserbewirtschaftung zu tun. Eine besondere Herausforderung stellt der klimatische Trend zu abnehmenden Niederschlägen und steigenden Temperaturen dar, der in einigen bereits heute ariden und semi-ariden Regionen besonders stark ausgeprägt ist. Der als Klimawandel bezeichnete Prozess stellt damit zusätzliche Anforderungen an die Bewirtschaftung der Wasserressourcen dar. Generell sollte mit steigenden Temperaturen zwar auch der Wasserkreislauf intensiviert werden, was aber nicht nur häufigere und höhere Niederschläge bedeutet, sondern auch Perioden längerer Trockenheit einschließt und damit auch das Risiko von Extremen wie Wassermangel oder Hochwasser erhöht. Trockenheit wird von vielen Forschern und Wasserwirtschaftlern - auch wegen der unscharfen Genese - als eine der komplexesten und am wenigsten verstandenen aller Naturgefahren bezeichnet, die zudem weit mehr Menschen als alle anderen Gefahren betrifft. Trockenheit ist eine Naturgefahr, die zeitweise in fast allen Regionen der Erde auftreten kann. Das Hauptziel dieser Dissertation ist es, neben einigen Analysemethoden dem Wasserwirtschaftler, Werkzeuge und Techniken für den Umgang mit Trockenheit zur Verfügung zu stellen, die es ermöglichen, auf extreme Ereignisse frühezeitig reagieren zu können. Die vorgeschlagene Vorgehensweise wurde beispielhaft auf das Einzugsgebiet der Ruhr in Nordrhein-Westfalen angewendet.

In dieser Dissertation wurden zunächst Klimaänderungen im Einzugsgebiet der Ruhr anhand von Datenaufzeichnungen über Niederschlag, Temperatur und Talsperrenzuflüssen untersucht. Alle untersuchten Zeitreihen wurden homogenisiert und die verwendete Methode wurde im Detail beschrieben. Trendanalysen wurden sowohl für jährliche als auch für saisonale Zeitreihen mit Hilfe des Mann-Kendall-Tests vorgenommen. Dabei wurden Häufigkeitsverteilungen von warmen und kalten Tagen sowie von feuchten und sehr feuchten Tagen anhand ihrer Perzentile dargestellt. Die Ergebnisse der hydrologischen Analyse zeigten einen signifikanten Anstieg der mittleren Temperatur über alle Zeitskalen. Das Auftreten von warmen Tagen sowohl im Winter als auch im Sommer hat ebenfalls einen signifikanten Anstieg zu verzeichnen, während die Anzahl der kalten Tage eine signifikante Verringerung erfuhr. Insgesamt kann anhand der Temperaturaufzeichnungen klar dargelegt werden, dass die Winter wärmer geworden sind und die Sommer heißer.

Die Niederschlagsaufzeichnungen ergaben einen Beleg für einen signifikanten Anstieg der Niederschläge im Winter, während für den Sommer keine signifikante Aussage getroffen werden konnte. Die Anzahl von aufeinanderfolgend trockenen Tage ging im Winter zurück, während im Sommer ein statistisch signifikanter Trend nicht festgestellt werden konnte. Die Analyse von feuchten und sehr feuchten Tagen zeigte hauptsächlich ein Ansteigen im Winterhalbjahr. Die Zuflüsse zu den Talsperren stiegen ebenfalls nur im Winter, während für den Sommer und für das hydrologische Jahr keine signifikanten Anstiege gefunden werden konnten. Korrelationsberechnungen zeigten, dass die Zuflüsse deutlich stärker mit den Niederschlägen als mit der Temperatur korrelieren, was den allgemeinen Erwartungen entspricht.

Trockenheit ist eine normale, wiederkehrende Eigenschaft des Klimas und wird als eine zeitlich anhaltende, regional ausgeprägte Wasserverfügbarkeit bezeichnet und kann sich auf Indikatoren wie Niederschlag, Abfluss oder Grundwasservorrat beziehen. Die meteorologische Trockenheit – auch als Niederschlagsdefizit bezeichnet – wurde für das Einzugsgebiet der Ruhr anhand des Standardized Precipitation Index (SPI) analysiert. Der SPI stellt ein Maß (Index) dar, mit welchem ein zusammenfassendes Urteil über den Grad der Trockenheit angegeben werden kann, ohne über eine mögliche Verteilungsfunktion der beobachteten Niederschlagssummen für eine Zeitskala Annahmen treffen zu müssen. Bei der Anwendung der SPI-Methode auf das Einzugsgebiet der Ruhr zeigte sich, dass Trockenperioden zeitlich zufällig verteilt aufgetreten und dass innerhalb der analysierten Zeitreihen von 1965 bis 2008 eine Reihe von Trockenperioden aufgetreten sind. Die Ergebnisse zeigten auch, dass obwohl die Niederschläge im Winter im Laufe der Jahre zugenommen haben, Trockenperioden sowohl im Sommer als auch im Winter auftreten und dass die größte Trockenheit der letzen 30 Jahre im Winter aufgetreten ist. Trenduntersuchungen des SPI anhand des Mann-Kendall-Tests zeigten, Keine signifikante Veränderung.

Da die Berechnung und Interpretation des SPI insbesondere bei der Analyse mehrerer Zeitskalen recht zeitintensiv ist, wurde ein Softwareprogramm mit graphischer Eingabemaske (GUI) in der MATLAB-Umgebung (The MathWorks, Inc.) entwickelt. Die Hauptziele dieser Software waren eine einfache Oberfläche zu schaffen, um folgende Analysen durchführen zu können: a) Berechnung des SPI für gegebene Niederschlagszeitreihen, b) Bestimmung, ob die Zeitreihe eine Trockenperiode enthält und c) Berechnung und Klassifizierung der Intensität der Trockenheit. Gegenüber dem häufig genutzten Programm des US National Mitigation Center ist die Anwendung des hier vorgestellten Tools komfortabler.

Die Vorhersehbarkeit von Trockenheit ist eine wesentliche Voraussetzung für einen angemessenen und rechtzeiten Umgang mit solchen Ereignissen. Eine Vorhersage würde insbesondere die Talsperrenbewirtschaftung oder mögliche landwirtschaftliche Bewässerungsstrategien erheblich vereinfachen und optimieren lassen. In dieser Dissertation wurde der SPI als Indikator für eine Vorhersage benutzt. Die Vorhersage selbst wurde mit Hilfe eines ARIMA-(Auto Regressive Integrated Moving Average) Modells durchgeführt. Die Anpassung des Modells wurde anhand der Korrelationsmethode nach Box und Jenkins vorgenommen. Die Auswahl des Modells orientierte sich an den beiden Kriterien AIC und SBC. ARIMA Modelle sind die am meisten eingesetzten Modelle für die Vorhersage von Zeitreihen, wobei die erforderliche Stationarität durch Transformationen wie Differenzieren oder Logarithmieren erreicht werden kann. Die Überprüfung des so entwickelten Vorhersagemodells wurde anhand der vorhandenen Zeitreihe vorgenommen. Des SPI-Vorhersagemodell zeigte dabei eine ausreichend gute Anpassung.

Die Vorhersage des SPI mittels stochastischer Modelle, wie zum Beispiel mit ARIMA, stellt einen komplexen Berechnungsvorgang dar. ARIMA-Modelle sind mittlerweile aber in einigen Standard-Statistik-Programmen wie SAS oder SPSS enthaltet werden. In dieser Dissertation wurde allerdings ein Softwarepaket entwickelt, welches sowohl ARIMA-Modelle als auch saisonale ARIMA-Modelle zur Verfügung stellt und dem Benutzer die wesentlichen Arbeitsschritte interaktiv bereitstellt.

Stochastische Simulationen von hydrologischen Zeitreihen werden seit vielen Jahren erfolgreich angewendet. In dieser Dissertation wurde die Generierung von Zuflusszeitreihen für vier Talsperren im Einzugsgebiet der Ruhr auf Basis von Monatswerten mit Hilfe der Thomas-Fiering Modells und des Monte Carlo-Modells durchgeführt. Dabei wurde ein neuer Ansatz für die Erhaltung der statistischen Parameter bei der Zufallszahlengenerierung verwendet. Damit konnten neben den Hauptmomenten der Verteilungsfunktion auch die Korrelation zwischen aufeinander folgenden Monaten erhalten werden, was unter anderem für die Analyse von Summenwerten und die Optimierung der Talsperrenbewirtschaftung unerlässlich ist.

Die Bewirtschaftung von Talsperren ist ein komplexer Prozess, welcher neben unterschiedlichen Zielen, eine Reihe von Randbedingungen und Unsicherheiten zu berücksichtigen hat. In Trockenzeiten stellt die variable Bewirtschaftung von Talsperren häufig die einzige Option für Handlungsalternativen dar. Für die Ermittlung einer fiktiven, aber möglichen optimalen Bewirtschaftung und für unterschiedliche Szenarien von Trockenperioden und Bedarfssituationen wurden die Optimierung Methoden verwendet Genetic Algorithm, Patterns Search und Gradient-Based und auf die Biggetalsperre beispielhaft angewendet. Die Ziele wurden in einer gewichteten Funktion zusammengeführt und anschließend unter Berücksichtigung der Randbedingungen einer Optimierung zugeführt. Das Ergebnis kann als Referenz für die tatsächliche Bewirtschaftung verwendet werden, bei der die Entscheidung unter unsicherer Prognose über den weiteren Verlauf der Trockenheit getroffen werden muss.

Für die Talsperrenbewirtschaftung kann ein Simulationsmodell ein sehr hilfreiches Werkzeug für die Entscheidungsfindung sein. Der beabsichtigte Abgabeplan kann zum Beispiel mit Hilfe der Vorhersagemodelle der Monatszuflüsse abgesichert werden. Ebenso lassen sich unterschiedliche Abgabestrategien zeitnah vergleichen. In dieser Dissertation wurde die oben beschriebene stochastische Modellierung mit einem Fuzzy-Logic Ansatz kombiniert. Dieser Ansatz basiert auf adaptiven Netzwerken und wird als ANFIS bezeichnet. Die Anwendung wird wiederum anhand der Biggetalsperre untersucht. Als Eingabe wurden der Monat, der Zufluss und die Speicherfüllung sowie der SPI-Index verwendet. Als Entscheidung wurde die monatliche Abgabe definiert. Für die Anpassung wurden die Zeitreihen in zwei Teile getrennt und als Trainings- und Testdatensatz benutzt.

Zwei unterschiedliche Modelle wurden dabei entwickelt einmal für die Abgabe im laufenden und zum anderen für die Abgabe in den kommenden Monaten. Die Modelle wurden anhand der historischen Aufzeichnungen validiert. Das so entwickelte Modell war gut in der Lage, die historische Bewirtschaftung nachzuvollziehen. Zudem wurden die Modelle sowohl mit, als auch ohne SPI Index entwickelt. Es konnte gezeigt werden, dass die Nutzung des SPI-Index zu wesentlich realistischeren Entscheidungen führt.

Historische Aufzeichnungen zeigen, dass Trockenperioden einen erheblichen Einfluss auf das Einzugsgebiet und die Talsperrenbewirtschaftung haben. Das Risiko von möglichen negativen Einflüsse steht im Zusammenhang mit den unterschiedlichen Nutzungsansprüchen. Im Ruhreinzugsgebiet sind die Hauptnutzungen die Trink- und Brauchwasserversorgung, die Erfüllung von Umweltqualitätsnormen im Fließgewässer, die Befriedigung von Ansprüchen aus touristischen Aktivitäten und eingeschränkt die Landwirtschaft. Die Ruhr war im Untersuchungszeitraum relativ häufig von Trockenperioden betroffen. Extreme Trockenperioden traten 1959, 1976, 1996, 2003 und 2007 auf.

Die Anfertigung und Vorhaltung von Bewirtschaftungs- und Handlungsoptionen im Falle einer Trockenheit ist wichtig für Entscheidungen über die Talsperrenbewirtschaftung. In dieser Arbeit wurden unterschiedliche Teilaspekte der Bewirtschaftung von Talsperren im Falle von Trockenheit aufgezeigt und beispielhaft für die Jahre 1976, 1996 und 2003 in einem Entscheidungs-Unterstützungs-Modell zusammengeführt. Es konnte gezeigt werden, dass der aufgestellte modellhafte Bewirtschaftungsplan die damaligen Expertenentscheidungen untermauert.

Table of Contents

List of Figures

List of Tables

Abbreviations

ACF	Autocorrelation Function
AIC	Akaike Information Criterion
ADI	Aggregated Drought Index
ANFIS	Adaptive Neural Fuzzy Inference Systems
ANN	Artificial Neural Network
ARIMA	Auto Regressive Integrated Moving Average
BIC	Bayes Information Criterion
CDF	Cumulative Distribution Function
CMI	Crop Moisture Index
DP	Dynamic Programing
DMP	Drought Management Plan
EA	Evolutionary Algorithm
FIS	Fuzzy Inference Systems
GA	Genetic Algorithm
GUI	Graphical User Interface
IPCC	Intergovernmental Panel on Climate Change
kWh	Kilowatt-hour
LP	Linear Programing
MAD	Mean Absolute Deviation
NDMC	National Drought Mitigation Center
PACF	Partial Autocorrelation Function
PDSI	Palmer Drought Severity Index

RACF	Residuals Autocorrolation Function
RPACF	Residuals Partial Autocorrelation Function
RMSE	Root of Mean Square Error
SBC	Schwarz Bayesian Criterion
SNHT	Standard Normal Homogeneity Test
SPI	Standardized Precipitation Index
SQDV	Sum of Squared Deviation of Releases from Demands
SWSI	Surface Water Supply Index
UNEP	United Nations Environment Programme
WMO	World Meteorological Organization

Chapter 1

Introduction

1.1 General

Water resources play an important role in most of human's activities. During the last decades water resources managers are facing severe challenges all over the world and the trends of increasing temperatures and decreasing precipitation intensify this situation. Climate change is a major global challenge facing water resources managers. Climate change permeates all aspects of our lives, from the food we eat, to the water we drink, to the places we can live. Rising global temperatures will lead to an intensification of the hydrological cycle, resulting in dryer dry seasons and wetter rainy seasons, and subsequently heightened risks of more extreme, longer and frequent floods and droughts.

Drought has been a major concern of mankind for centuries. It is considered by many to be the most complex but least understood of all natural hazards, affecting more people than any other hazard. Drought is a complex phenomenon and it is generally viewed as a sustainable and regionally extensive occurrence of below-average natural water availability either in the form of precipitation, river runoff or groundwater. Drought may also be referred as an interaction and combination between physical processes and human activities (Changnon and Easterling, 1989). Such processes are extremely stochastic in nature and, thus, problematical for reliable prediction. However, cumulative experience from scientific investigations of recent decades is indicating that given a certain period of time in a given locale, the occurrence of an uncertain event such as drought, becomes a certainty.

Drought is considered as a fruitful field for research in several science disciplines. Drought and its consequences must be recognized and estimated in advance for all planning and management efforts in water resources. Comprehensive planning for developing optimal strategies to deal with drought situations is becoming an increasingly important subject of concern to researchers and water manager in order to protect the affected community from the adverse effects of drought. Drought has direct impacts on municipal water resources management, thus water resources decision makers must be prepared to anticipate such situations and accept the challenges and complications that are involved in dealing with drought related problems.

Drought monitoring has much to offer to water decision making. Drought monitoring, the ability to assess the current conditions and predict future drought development are a key to any water resources management plan during drought periods. The main purpose of any drought monitoring system is to identify various drought indices to provide information to resources manager and system operators. The indicators that are used to derive drought indices are precipitation, snow pack, streamflow and reservoir storage. A drought index value is typically a single number, far more useful than raw data for decision making (NDMC, 2006). Although none of the major indices is inherently superior to the rest in all circumstances, some indices are better suited than others for certain uses. Some of the widely used drought indices are the Palmer Drought Severity Index (PDSI), Crop Moisture Index (CMI), Standardized Precipitation Index (SPI) and Surface Water Supply Index (SWSI).

Drought forecasting plays an important role in the mitigation of impacts of drought on water resources systems. Traditionally, statistical models based on time series methods have been used for hydrologic drought forecasting (Kim and Valde´s, 2003). One of the basic deficiencies in mitigating the effects of drought is the inability to forecast drought conditions reasonably well in advance by either a few months or seasons. Accurate drought forecasts would enable optimal operation of irrigation systems. Panu (Panu and Sharma, 2002) reported that the ARMA models, pattern recognition techniques, physically based models using Palmer drought severity index (PDSI), standardized precipitation index (SPI), a moisture adequacy index involving Markov chains, or the notion of conditional probability, seems to offer a potential to develop reliable and robust forecasts.

Reservoir operation, especially during drought periods, is a complex problem that involves many decision variables, multiple objectives as well as considerable risk and uncertainty. In addition, the conflicting objectives lead to significant challenges for operators when making operational decisions. Traditionally, reservoir operation is based on heuristic procedures, embracing rule curves and subjective judgments by the operator. This provides general operation strategies for reservoir releases according to the current reservoir level, hydrological conditions, water demands and the time of the year (Hakimi-Asiabar et al.). Preparing efficient drought management plan is the best way to reduce drought impacts which could be continued to several weeks or months even after drought events. Simulation models become the most commonly used method for monitoring, planning and managing drought. Simulation models can be used for evaluating drought plans before a drought's onset.

1.2 Description of the Study Area

The River Ruhr catchment area covers 4485 km² and forms the largest reservoir system in Germany with a total storage capacity of 464,1 million m³ (Morgenschweis et al., 2003). The mean annual runoff at the mouth of the Ruhr is 2.4 Billion m^3. The Ruhr basin contains 16 hydroelectric power plants and 110 pumping stations. The name of this region was derived from the name of the river Ruhr. The Ruhr River Association (Ruhrverband) is the manager of this reservoir system. The major tasks of the Ruhr Association are: to provide drinking water; and to supply local industry with process water within the so-called Ruhr district, which is one of the most densely populated and industrialized areas in Europe (Morgenschweis et al., 2003). As shown in figure 1.1, about 50 % of the withdrawal are exported to neighboring catchments.

It is due to this highly densely populated area (about 2.13 million inhabitants) that water consumption per unit of area is approximately seven times higher than the average consumption in the Federal Republic of Germany (Khadr et al., 2009). Special measures are therefore necessary in order to guarantee the supply of drinking water and the disposal of wastewater (Morgenschweis et al., 2003). Maniak (Maniak and Renz, 1977) reported that in 1977, 70 % of the water demand of the Rhenish-Westphalian industrial zone was covered by the Ruhr and this percentage increased in dry periods. In times of extreme droughts it increases up to the 1.6 fold of the annual average.
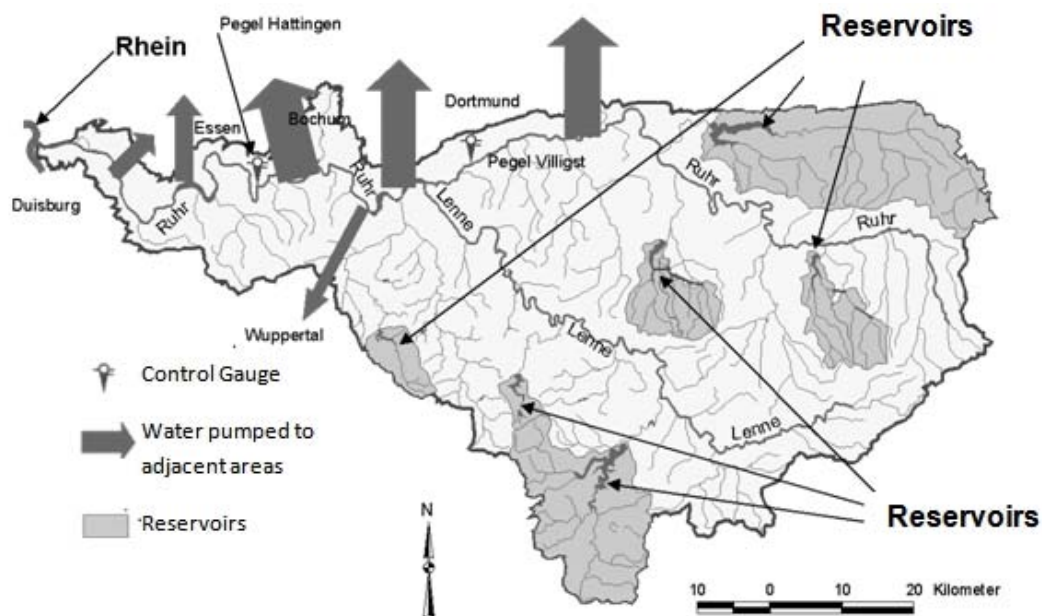
Figure 1.1: Water management system in the Ruhr drainage basin.
Source (after (Brudy-Zippelius, 2003)).

1.3 Objectives and Organization of the Dissertation

The overall objective of this research effort is to evolve appropriate and interdisciplinary tools and techniques for drought characterization and for enhanced management of water resources systems during drought periods. The developing tools will have four major: 1) Climate change; 2) Drought monitoring and forecasting; 3) Reservoir operation during drought and 4) drought management plan. The developed tools are demonstrated by an application to the Ruhr river basin as case study.

The thesis is structured into 11 chapters, a reference list and annexes.

In Chapter 1 the introductory background on the main theme, the research, study area and objectives of the thesis. The various chapters that follow systematically to analyze different issues on the basis of these objectives.

In Chapter 2, an investigation of climate change in the Ruhr River basin is presented using a set of measured data containing precipitation, temperature and inflow records. All data are subjected to a homogenization procedure; the data homogenization is described in detail. Yearly and seasonal trend analyses are performed on all data series using the Mann-Kendall test. The frequency distributions of warm & cold days and very & extremely wet days are also examined using percentile indices.

Chapter 3 deals with the temporal and spatial characteristics of meteorological drought in the Ruhr river basin using the Standardized Precipitation Index (SPI). The Standardized Precipitation Index aims to provide a concise overall picture of drought, regardless to the actual probability distribution of the observed cumulative amounts of rainfall for a given time scale.

Chapter 4 deals with drought forecasting, which is an essential tool for implementing appropriate mitigation measures in order to reduce negative impacts of drought on water resources systems. The capability of the Auto Regressive Integrated Moving Average (ARIMA) model in drought forecasting is investigated using the correlation methods of Box and Jenkins and the AIC and SBC structure selection criteria.

Chapter 5 deals with the stochastic simulation of hydrologic time series which has been widely used for solving various problems associated with the planning and management and operational purposes. The stochastic streamflow generation model of Thomas-Fiering and the Monte Carlo simulation model are applied to synthetically generate monthly inflow scenarios for four reservoirs in the Ruhr river basin. These scenarios are then used in the optimization and simulation models of reservoir operation.

In chapter 6, an optimization model is proposed for reservoir operation during prolonged periods of drought using Genetic Algorithm, Pattern Search and Gradient-based method. The Bigge reservoir is presented as case study. Several scenarios for low inflow period are attempted. Each scenario has its assumptions for monthly inflow and monthly demand. Evaluation of the developed model has been carried out using the driest year in the available historical records.

In chapter 7, an example of the collective use of stochastic models and Adaptive Neural Network-based Fuzzy Inference System (ANFIS) for reservoir operation and simulation is presented. The applicability and capability of the ANFIS model are investigated by the use of a set of data in the Ruhr reservoirs system, Germany. The historical data are inflow, reservoir storage, the SPI index and reservoir release. The historical data are divided into two independent sets, one set to train and the other to test the constructed models. Two main models are developed. In both models the set of input includes the time of the year, storage, inflow and Standardized Precipitation Index (SPI). The output of the first model is the release during the next month; on the other hand, the output of the second model is the release of the current month. Predicted release values and observed release values are evaluated using several common evaluation criteria.

In chapter 8, a drought management plan is proposed and the procedures of this plan are applied to the case study.

Chapter 9 presents summaries and conclusions of the research. It also outlines recommendations for further research.

The Appendices provide supplementary information to the materials presented above. Appendix A present a graphical user interface (GUI) to monitor and analyze meteorological drought using the Standardized Precipitation Index (SPI). Appendix B presents software with a friendly graphical user interface (GUI) for meteorological drought forecasting. The developed GUI contains ARIMA and multiplicative Seasonal Autoregressive Integrated Moving Average (SARIMA) model.

Chapter 2

Study of Climate Change in the Ruhr River Basin Concerning the Occurrence of Drought

2.1 Climate is changing

Climate change is a real and growing problem for the world. It is a complex phenomenon that alters the whole environment in which humans live. Global climate change will have profound implications for the quality of life of hundreds of millions of people (Hübler et al., 2008). In the last few years, climate change has become one of the most heavily researched subjects in science. There is no doubt that the increase in mean global surface temperature by 0.6 ± 0.2 ^{0}C over the 20th century (IPCC, 2001)[1] is not only a result of climate variability but of enhanced emission of greenhouse gases due to human activities (Menzel and Bürger, 2002). From the recent Intergovernmental Panel on Climate Change fourth assessment report (IPCC, 2007), little doubt remains that the climate system has warmed in recent decade (Steele-Dunne et al., 2008).

Changes in climate variability and extremes of weather have received increased attention in the last few years. Understanding changes in climate variability and climate extremes is made difficult by interactions between the changes in the mean and variability (IPCC, 2001). Such interactions vary from one variable to another one depending on the statistical distribution of these variables. For example, the distribution of temperatures often resembles a normal distribution where non-stationarity of the distribution implies changes in the mean or variance. In such a distribution, an increase in the mean leads to new record high temperatures (Figure 2.1.a), but a change in the mean does not imply any change in variability.

Figure 2.1.b shows that the range between the hottest and coldest temperatures does not change. An increase in variability without a change in the mean implies an increase in the probability of both hot and cold extremes as well as the absolute value of the extremes (Figure 2.1.b). Increases in both the mean and the variability are also possible (Figure 2.1.c), which affects the probability of hot and cold extremes, with more frequent hot events with more extreme high temperatures and fewer cold events. Other combinations of changes in both mean and variability would lead to different results.

---

[1] Report of the Intergovernmental Panel on Climate Change
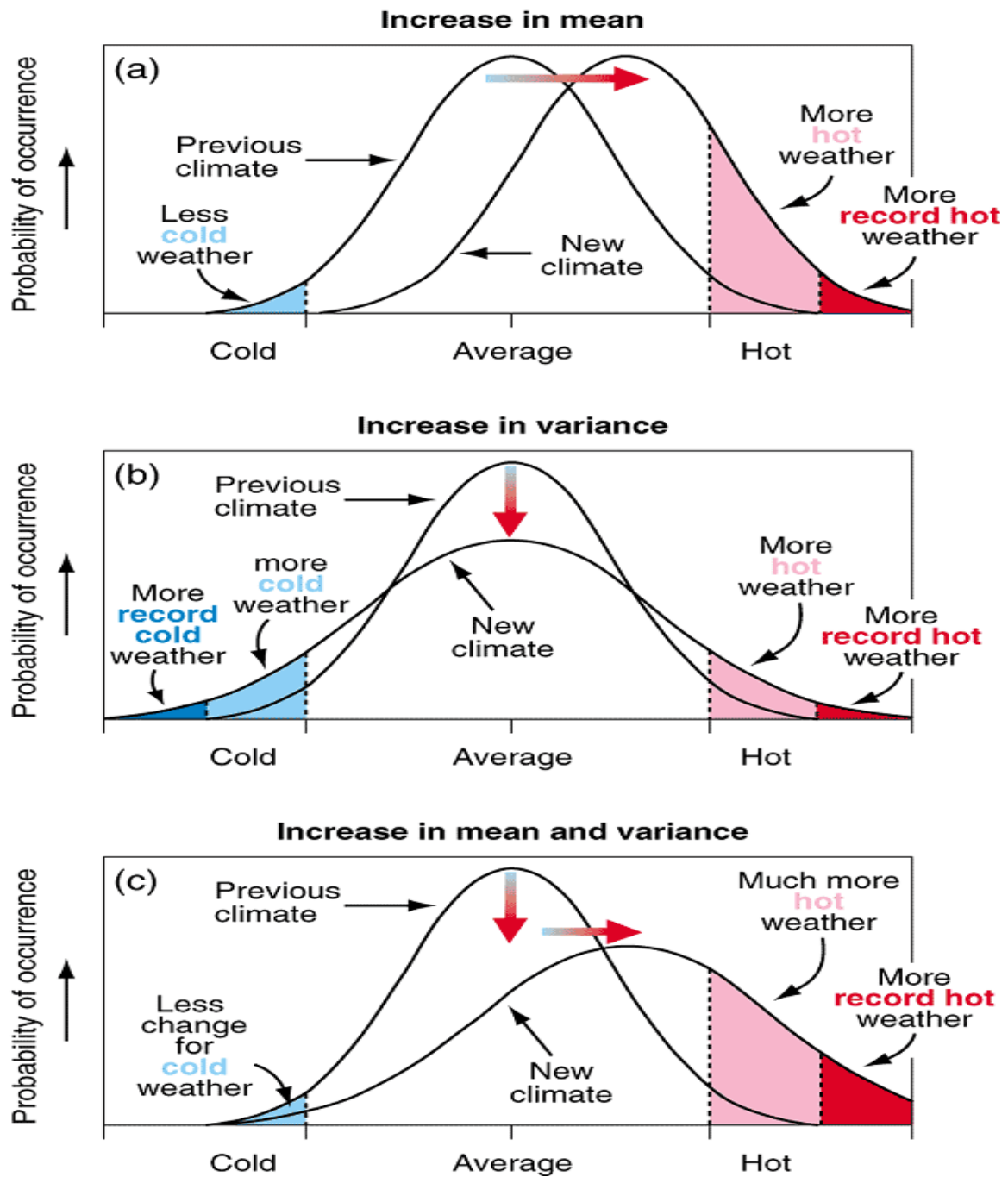
Figure 2.1: Schematic showing the effect on extreme temperatures

(a) the mean temperature increases, (b) the variance increases and (c) when both the mean and variance increase for a normal distribution of temperature.
(Source: IPCC Third Assessment Report: http://www.grida.no/climate/ipcc_tar/wg1/088.htm)

2.2 Background

The average climate experienced over long periods, such as temperature, wind and rainfall patterns, has changed many times in response to natural variability and natural causes such as volcanic activity. However, according to the Intergovernmental Panel on Climate Change (IPCC, 2001), since the industrial revolution, anthropogenic causes are playing an important role primarily due to the combustion of fossil fuels, agriculture and land-use changes (*e.g.* deforestation), which has increased the atmospheric concentration of aerosols and greenhouse gases. New evidences suggest that most of the warming observed over the last 50 years is attributable to human activities

Historical records show that climate has been changing on different time scales. During the last century, a global steady warming trend occurred from the late 1890s through the 1940s followed by a minor cooling trend in the late 1940s and the 1950s (Leemans and Cmmer, 1991). Precipitation plays an important role in the global energy and water cycle. Exact information about precipitation amounts reaching the land surface is of special importance for fresh water assessment and management related to agriculture land use, hydrology and risk reduction of flood and drought (Schneider et al., 2008).

The importance of assessing trends in weather extremes is often emphasized. The principal reason is that extreme weather conditions related to temperature, precipitation, storms or other aspects of climate, can cause loss of life, severe damage and large economic and societal losses (Moberg, 2006). Using results from a number of workshops held in data-sparse regions and high-quality station data supplied by numerous scientists world wide, seasonal and annual indices for the period 1951–2003 were gridded (Alexander et al., 2006). Widespread significant increase in temperature extremes for the period 1951–2003, especially those related to daily minimum temperatures is evident that warming is apparent in all seasons. Precipitation changes have been much less coherent than temperature changes, but annual precipitation has shown a widespread significant increase.

Alexander (Alexander, 2005) studied the variability of temperature and precipitation in the European alps since 1500, unlike temperature, precipitation variation over the European Alps showed no significant low-frequency trend and increased uncertainty back to 1500. Results showed that the years 1540, 1921 and 2003 were very likely the driest in the context of the last 500 years. Groisman et al (Groisman, 2005) found disproportionate changes during the past decades in heavy and very heavy precipitation compared to the change in the annual and/or seasonal precipitation. Their results indicate an increasing probability of intense precipitation.

Moberg (Moberg and Jonesa, 2005) reported that, there are some coherent patterns of climate changes in parts of Europe over the 20th century. The most outstanding feature for precipitation is that winter precipitation increased significantly at several stations, both regarding the mean precipitation intensity and moderately strong events; at the same time, the length of dry spells in winters also generally increased. The length of dry periods also increased (insignificantly) in summer, but there are few significant changes in summer precipitation amounts. Warming trends dominate in the study region over the 20th century as a whole, both in winter and summer and both for the cold and warm tails of the temperature distribution. When analyzing the two century halves separately, there is evidence for markedly different behavior in the warm and cold tails of the temperature distribution and also strong differences between winter and summer. Winter temperatures warmed in the second half of the century, with the largest changes in the cold tail for daily minimum temperatures. There is much less evidence for widespread warming in summer in the same period.

Hundexha (Hundexha and Bardossy, 2005) investigated the evolution of daily extreme precipitation and temperature across Western Germany from 1958 to 2001. The results obtained indicated that both the daily minimum and maximum extreme temperatures have increased over the investigation period, with the degree of change showing seasonal variability. On an annual basis, the change in the daily minimum extreme temperature was found to be greater than that of the daily maximum extreme temperature. The daily extreme heavy precipitation has shown increasing trends both in magnitude and frequency of occurrence in all seasons except summer, where it showed the opposite trend. Beck (Beck et al., 2004) studied the extreme daily precipitation events and droughts in Germany. Results gave an evidence of an increase in both frequency and intensity of extreme precipitation events in the 20th century.

Morgenschweis (Morgenschweis et al., 2007) reported that the Ruhr River Association analyzed long time series (1927-2005) of calculated areal precipitation with the aid of statistical methods to identify trends. Results showed that there is an increase in the winter precipitation and no trend was noted for the summer precipitation. Youmin Chen (Chen and Buerger) reported that, the mean precipitation in the Ruhr basin will increase, but its frequency will slightly decrease and its intensity will significantly increase.

The main objective of this chapter is to study the change of climate in the Ruhr river basin. A set of data, containing precipitation, temperature and inflow records, has been used to investigate to perform the required hydrological analysis.

2.3 Data and Methodology

2.3.1. Data Collection

The database includes 13 stations with mean daily temperature and/or precipitation series starting at 1961 has been established. The data include 4 stations with mean daily inflow series. The inflow time series present the inflow to the main reservoirs in the Ruhr river basin namely, Bigge reservoir, Moehne reservoir, Henne reservoir and Sorpe reservoir. Source of data is the Ruhrverband (Ruhr River Association). Figure 2.2 shows the location of each station used in the study. All time series were checked to find out all missing data. Table 2.1 and table 2.2 contain information about stations, covered period and the fraction of missing data.

2.3.2 Analysis of Hydrological Time Series

Records of rainfall and river flow form suitable data sequence can be studied by the methods of time series analysis. The tools of this specialized topic in mathematical statistics provide valuable assistance to engineers in solving problems involving the frequency of occurrences of major hydrological events (Shaw, 1994). In particular, when only a relatively short data record is available, the formulation of a time series model of those data can enable long sequences of comparable data to be generated to provide the basis for better estimates of hydrological behavior. In addition, the time series analysis of rainfall, evaporation, runoff and other sequential records of hydrological variables can assist in the evaluation of any irregularities in those records. Cross-correlation of different hydrological time series may help in the understanding of hydrological processes.

Tasks of time series analysis include:

(1) Identification of the several components of a time series,

(2) Mathematical description (modeling) of different components identified.

If a hydrological time series is represented by $X_1,$ $X_2,$ $X_3,$ ..., $X_t,$ ..., then symbolically, one can represent the structure of the X_t by:

$$X_t \iff [T_t, P_t, E_t].$$

Where T_t is the trend component, P_t is the periodic component and E_t is the stochastic component (for more details see chapter 5). The first two components are specific deterministic features and contain no element of randomness. The third, stochastic, component contains both random fluctuations and the self-correlated persistence within the data series. These three components form a basic model for time series analysis.

2.3.3 Trend Component

This may be caused by long-term climatic change or, in river flow, by gradual changes in a catchment's response to rainfall owing to land use changes. Sometimes, the presence of a trend cannot be readily identified.

2.3.3.1 Methods of Trend Identification:

Different statistical methods, both nonparametric tests and parametric tests, for identifying trend in time-series are available in the literature. Two methods are commonly used for identifying the trend in the hydrologic time series. These two methods are The Mann-Kendall test and the Linear Regression Method. In this study the Mann-Kendall test has been applied to identify the trend in the time series. This method will be discussed briefly in the following section.

(1) Mann-Kendall Test

The Mann–Kendall nonparametric test is an effective tool for analyzing change trend. It is one of the most common non-parametric rank-based statistical tests which are used in hydrological studies (Yue et al., 2002). Mann–Kendall test is simple, robust and can cope with missing values and values below a detection limit. To identify the change in temperature, precipitation and streamflow for several time scales (year/season/month), the probabilistic parameter has been studied at 0.05, 0.10 and 0.20 field significance level (Storch and Zwiers, 2001).

The Mann-Kendall test is based on the test statistic S defined as follows:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_i - x_j) \qquad (2.1)$$

Where x_j are the sequential data values, n is the length of the data set and

$$sgn(\theta) = \begin{cases} 1 & if\ \theta > 0 \\ 0 & if\ \theta = 0 \\ -1 & if\ \theta < 0 \end{cases} \qquad (2.2)$$

Mann (1945) and Kendall (1975) have documented that when, the statistic S is approximately normally distributed with the mean and the variance as follows:

$$E(S) = 0 \qquad (2.3)$$

$$V(S) = \frac{n(n-1)(2n+5) - \sum_{p=1}^{q} t_p(t_p-1)(2t_p+5)}{18} \qquad (2.4)$$

Where n = number of data,

t_p = the number of ties for the pth value (number of data in the pth group),

q = the number of tied values (number of groups with equal values/ties).

The standardized Mann-Kendall test statistic Z_{MK} is computed by:

$$Z_{MK} = \begin{cases} \dfrac{S-1}{\sqrt{Var(s)}} & S > 0 \\ 0 & S = 0 \\ \dfrac{S+1}{\sqrt{Var(s)}} & S < 0 \end{cases} \qquad (2.5)$$

The standardized MK statistic Z follows the standard normal distribution with mean of zero and variance of one. The hypothesis that there has not trend will be rejected if

$$|Z_{MK}| > Z_{1-\alpha/2} \qquad (2.6)$$

Where $Z_{1-\alpha/2}$ is the value read from a standard normal distribution table with α being the significance level of the test.

2.3.4 Missing Data Calculation

In the used data series there are only missing values in precipitation records. The calculation of the missing data is performed for the daily time series displaying gaps. The procedure of calculation considers the linear regression between the series with gaps (*Y*) and the reference series (*X*) (Santos and Henriques, 1999; Simolo et al., 2009). The correlation coefficient between each two stations is calculated. Then for a data series *Y*, the higher correlation coefficient is selected, then the corresponding data series *X* was selected and finally the missing values in the *Y* series are calculated. Calculations involved series with larger number of available data where the linear model can give good estimates of statistical parameters (mean and variance in the extended series). The equation used to calculate the missing values is:

$$Y_i = B_0 + B_1 X_i \qquad (2.7)$$

Where the regression parameters B_0 and B_1 were calculated by the least squares method as follow:

$$B_1 = \frac{(\sum XY)/n - \bar{X}\,\bar{Y}}{S_x{}^2} \qquad (2.8)$$

$$B_0 = \bar{Y} - B_1\bar{X} \qquad (2.9)$$

Table 2.1: Names of stations, covered period and proportion of missing data of the time series of temperature and precipitation records

Station	Ref. No. on figure 2.2	Temperature covered period	Temperature missing data (%)	Precipitation covered period	Precipitation missing data (%)
Biggetalsperre	1	1961-1995	0	1960-2007	8.31
Mohnetalsperre	2	1961-1995	0	1960-2007	8.22
Sorpetalsperre	3	1961-2007	0	1960-2007	8.32
Hennetalsperre	4	1961-1995	0	1960-2007	8.03
Listertalsperre	6	1961-1995	0	1960-2007	0
Drolshagen_Bleche	8	1961-1995	0	1960-2007	0.28
Willertshagen_Volmehof	9	1961-1995	0	1960-2007	0.28
Ennepetalsperre	14	1961-1995	0	1960-2007	13.68
Neuhaus	23	1961-1995	0	1960-2007	0.46
Essen_Kettwig	27	1961-1995	0	1960-2007	0.28

Station	Ref. No. on figure 2.2	Temperature covered period	Temperature missing data (%)	Precipitation covered period	Precipitation missing data (%)
Essen_Steele	30	1961-1995	0	1960-2007	0.82
Hagen	35	1961-1995	0	1960-2007	0.28
Versetalsperre	5	1961-2007	0	1960-2007	0.28

Table 2.2: Names of stations, covered period and proportion of missing data of the time series of inflow records

Station	Inflow covered period	Inflow missing data (%)
Biggetalsperre	1967-2008	0
Möhnetalsperre	1967-2008	0
Hennetalsperre	1967-2008	0
Sorpetalsperre	1967-2008	0
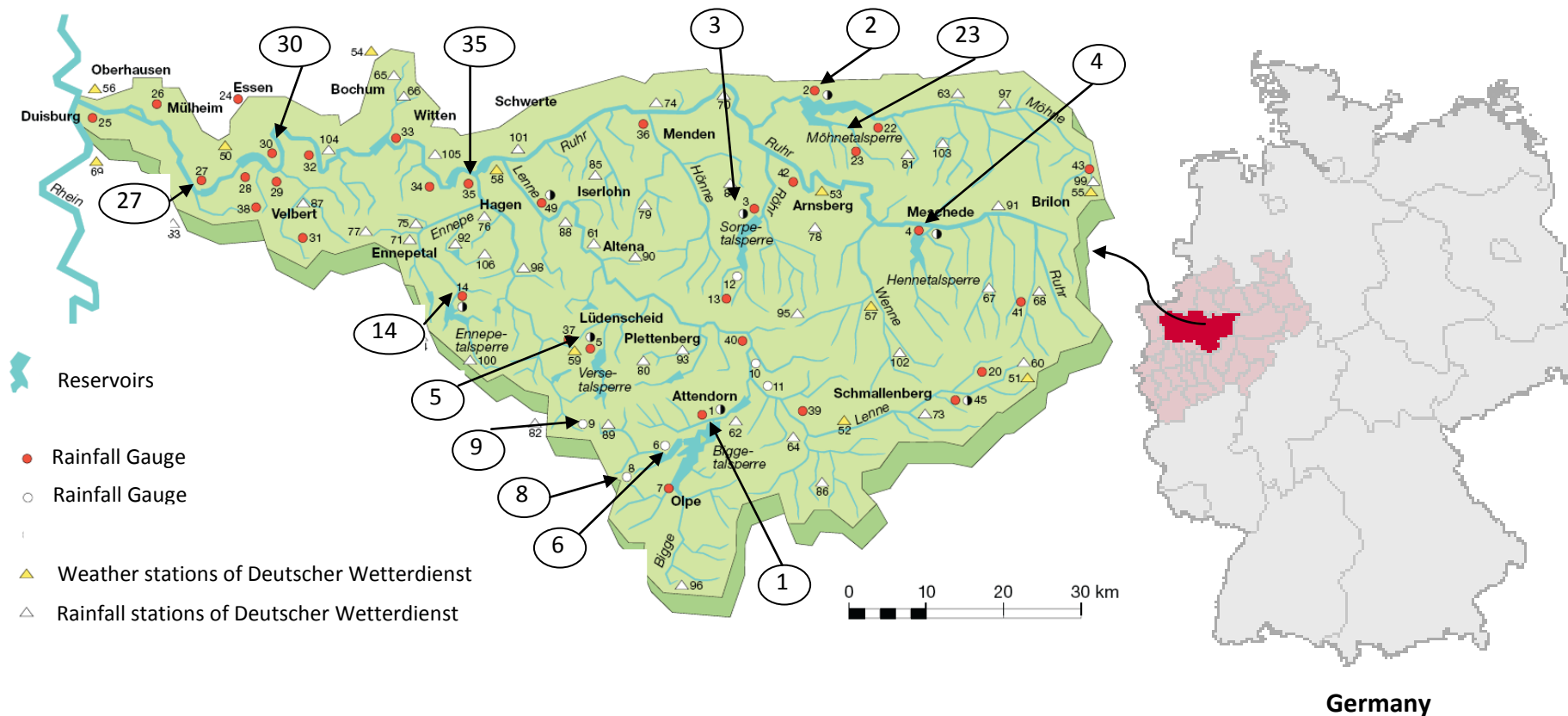
Figure 2.2: Location of stations used in the study (Temperature and Precipitation)

Source: http://www.talsperrenleitzentrale-ruhr.de/daten/internet/veroeffentlichungen/ruhrwassermenge_2006.pdf

2.3.5 Homogeneity testing

Climate data can be used to generate a enormous deal of information about the atmospheric environment which affects all aspects of human endeavor (Aguilar et al., 1998). Many factors affect on the quality and reliability of the data obtained from the meteorological stations. Time series of precipitation are influenced by the location of the gauge, the tool and method used and the observation quality and the time series might gain inhomogeneous structure (Dikbas et al., 2010). For this reason, the reliability and quality of the data to be used in the climate analysis be tested statistically. When the time series have a homogenous structure, tt can be stated that the observation time series is a reliable climatic series.

A homogeneous climate time series can be defined as one where variations are caused only by variations in weather and climate (Keiser and Grieffiths, 1997). If a precipitation or a temperature time series is homogeneous, all variability and changes of the series can be considered due to the atmospheric processes.

Most long-term climatic time series have been affected by a number of non-climatic factors that make these data unrepresentative of the actual climate variation occurring over time (Aguilar et al., 1998). These factors include changes in instruments, observing practices, station locations, formulae used to calculate means and station environment. Some changes cause sharp discontinuities while other changes, particularly change in the environment around the station, can cause gradual biases in the data. All of these inhomogeneities can bias a time series and lead to misinterpretations of the studied climate. It is important, therefore, to remove the inhomogeneities or at least determine the possible error they may cause.

There exist many methodologies for detecting homogeneity of climatological time series. These methods can be grouped into two categories, direct or indirect methods, depending on the availability or use of station history files known as metadata. Direct methods use metadata and indirect methods use a variety of statistical and graphical techniques to determine inhomogeneities (Peterson et al., 1998). The indirect homogeneity tests of a climatic time series could be classified into two groups; absolute tests and relative tests. The absolute tests depend on the use of a single station's records, whereas relative tests depend on the use of neighboring stations' data that are supposedly homogeneous (Karabork et al., 2007). Some relative homogeneity tests which do not require homogeneous reference series have become available (Albert, 2004; Szentimrey, 1999).

The main purpose of this section is to determine a reliable climatic series for the climate analysis which is the target of this chapter. First, the missing values of the meteorological time series are completed (see section 2.3.4) , then the homogeneity will be tested. In this study, the two groups of the indirect method are applied to all data series to test its homogeneity.

2.3.5.1 Absolute Homogeneity Tests

The most common tests which could be used to test the departure of homogeneity of a given time series are the Standard Normal Homogeneity Test (SNHT) for a single break (Alexandersson, 1986), the Buishand range test (Buishand, 1982), the Pettitt test (Pettitt, 1979) and the Von Neumann ratio test (Von Neumann, 1941). All four tests suppose under the null hypothesis that the annual values X_i of the testing variable X are independent and identically distributed. Under the alternative hypothesis, the SNHT, the Buishand range and the Pettitt test assume that a step-wise shift in the mean -a break- is present (Yesilirmak et al., 2009). The fourth test, the Von Neumann ratio test, assumes under the alternative hypothesis that the series is not randomly distributed. This test is not location specific, which means that it does not give information on the year of the break. In this study, the Buishand range test and Von Neumann ratio test have been applied to all time series for both precipitation and temperature.

2.3.5.1.1 Buishand Range Test:

The Buishand Range test can be used for testing homogeneity of the data (Buishand, 1982). The test is based on the rescaled adjusted partial sums for a time series X_i as follow:

$$S_k^* = \sum_i^k (X_i - \bar{X})^2 \qquad , \ k\text{=}1, 2,\text{…………,} N \quad , S_0^* = 0 \qquad\qquad (2.10)$$

When a time series is homogeneous the values of S_k^* will fluctuate around zero, because no systematic deviations of the X_i values with respect to their mean will appear. If a break is present in year K, then S_k^* reaches a maximum (negative shift) or minimum (positive shift) near the year $k = K$. Rescaled adjusted partial sums are obtained by dividing the S_k^* by the sample standard deviation:

$$S_k^{**} = {S_k^*}\big/{S_d} \qquad , \ k\text{=}1, 2,\text{…………,} N \qquad\qquad (2.11)$$

Where S_d is the standard deviation,

$$S_d = \sum_i^N (X_i - \bar{X})^2/N \qquad\qquad (2.12)$$

  A statistic which is sensitive to departures from homogeneity is:

$$Q = \mathop{max}_{0 \ll k \ll N} | S_k^{**} | \qquad\qquad (2.13)$$

High values of Q are an indication for a change in level. Critical values for the test-statistic can be found in (Buishand, 1982).

Figure 2.3 and figure 2.4 present the results of Buishand Range Test for both precipitation and temperature time series for all stations which have been used in this study. The results indicate that the precipitation time series at 3 stations are not homogeneous since the test statistic exceeds the critical value of 1.52 at the 95 % confidence level (Figure 2.3.a). On the other hand the results of the Buishand Range Test for temperature and inflow time series show that all stations have homogenous data series (figure 2.3.b and figure 2.4).
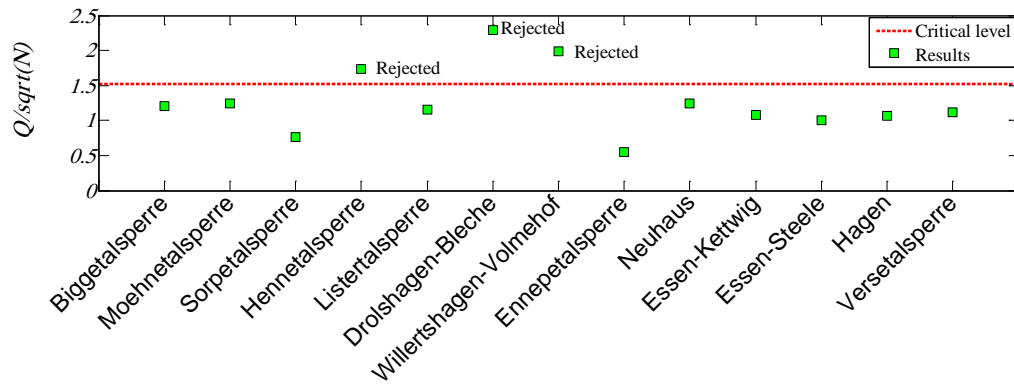
Figure 2.3.a: Results of the Buishand Range Test – Precipitation data series (1961 - 2007).
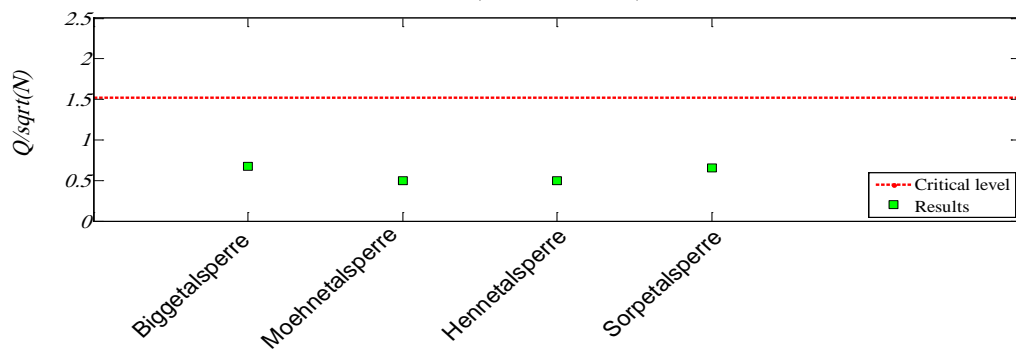
Figure 2.3.b: Results of the Buishand Range Test – Inflow data series (1967 - 2007).
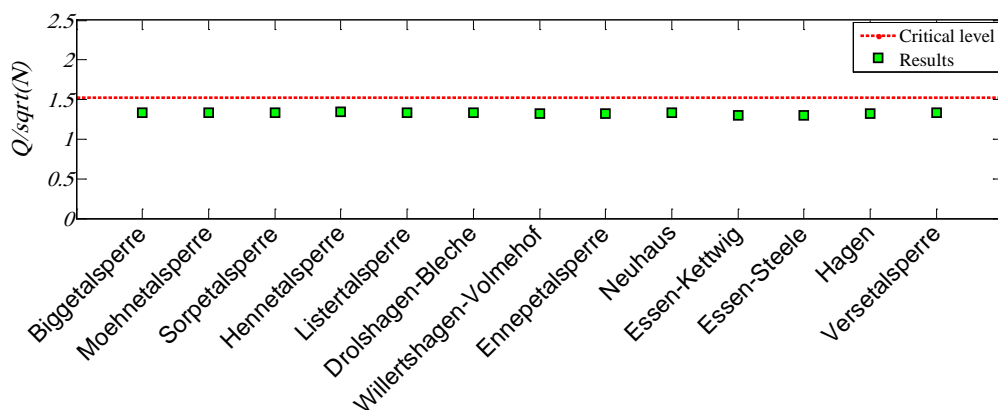
Figure 2.4: Results of the Buishand Range Test – Temperature data series (1961 - 2007).

2.3.5.1.2 Von Neumann ratio test:

The well-known Von Neumann ratio is defined as:

$$N_V = \sum_{i=1}^{n-1}(X_i - X_{i+1})^2 / \sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad (2.14)$$

In which $\bar{X}$ stands for the average of the Xi's. If the sample contains a break, then the value of N_v tends to be lower than this expected value (Buishand, 1982). If the sample has rapid variations in the mean, then values of N_v may rise above 2 (Sahin and Cigizoglu, 2010). Only this test does not give information on the year of break. The results of the Von Neumann ratio test (figures 2.5 and 2.6) indicate that the precipitation time series at 2 stations are not homogeneous since the values of N_v are lower than the critical level.
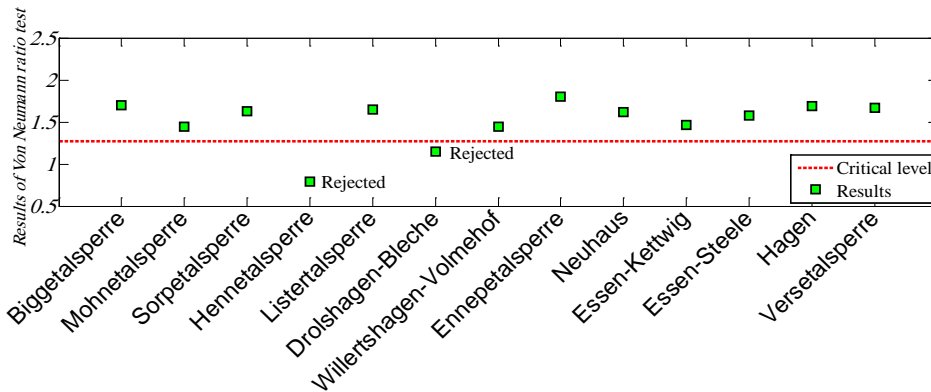
Figure 2.5.a: Results of Von Neumann ratio test – Precipitation data series (1961-2007)
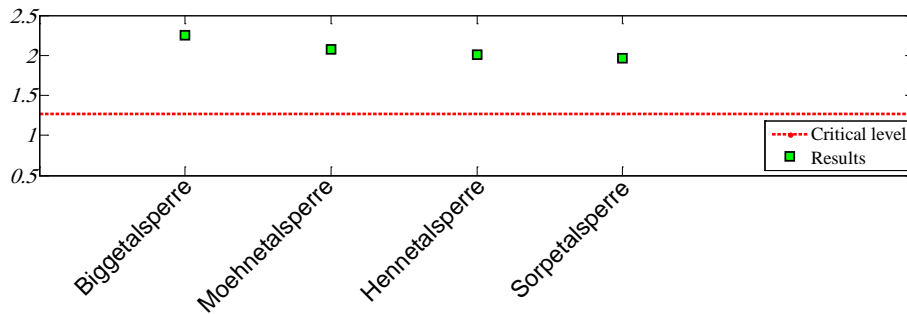
Figure 2.5.b: Results of Von Neumann ratio test – Inflow data series (1967 - 2007)
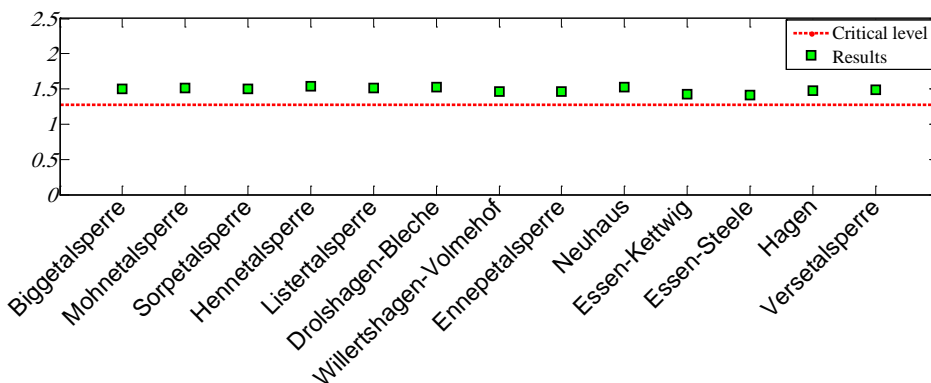
Figure 2.6: Results of Von Neumann ratio test – Temperature data series (1961 - 2007)

2.3.5.2 Relative Homogeneity Tests

In this study, the relative homogeneity of precipitation has been examined using the double mass curve test which is a commonly used data analysis approach. The theory behind double mass curves is that by plotting the cumulation of two quantities, the data will plot as a straight line and the slope of this line will represent the constant of proportionality between the two quantities (Albert, 2004). A break in slope indicates a change in the constant of proportionality (Reddy, 2005). The main purpose of these curves is to check the consistency of data over time and to identify changes in trends by changes in the slope as shown in figure 2.7.

For example, let Aand B are two neighboring stations in the same region. Suppose that in a specific year the amount of precipitation at station A lies above the historical average. Then it is expected that the annual amount of B is also higher than the historical mean. Because of this correlation it is possible that in a specific region the number of significant values is much larger than the expected number under the null hypothesis. Figure 2.7 presents a sample of the double mass curve for the precipitation data series. The X axis presents the reference station (Listertalsperre), Y axis presents other stations (Biggetalsperre, Hennetalsperre and Drolshagen-Bleche) and Z axis presents the time in years. Both of Hennetalsperre and Drolshagen-Bleche provided significant break in slope as Figure 2.7.

The time at which a change occurred is the most significant information that obtained when a break in slope is provided. Once the date in which the change occurred is known, one can study the historical record of the gauging station to see if any changes or sampling methods have been documented. To compare the distributions of the data series for all stations, the empirical cumulative distribution function plots of the data have been plotted. Figure 2.8 presents a sample of the distributions of the data series of Biggetalsperre, Hennetalsperre and Listertalsperre stations. It is notable from Figure 2.8 that the station Hennetalsperre has different distribution. The same approach has been applied to the temperature data series and results showed that all station did not provide any break in slope as shown in figure 2.9. The cumulative distributions of temperature records of all stations (figure 2.10) have approximately the same behavior and seem to be parallel to each other because all records are strongly correlated but the stations do not have the same statistical properties.
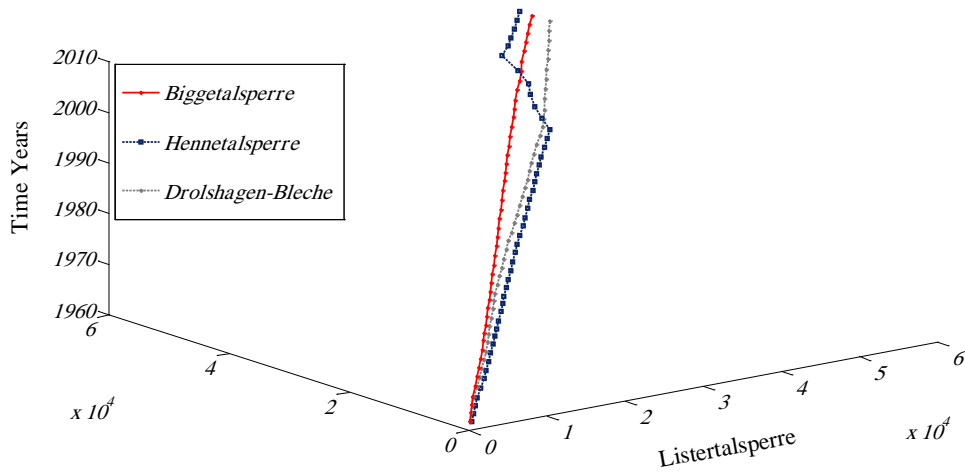
Figure 2.7: Cumulative precipitation for station Listertalsperre vs. cumulative precipitation for the other stations (1961 - 2007).
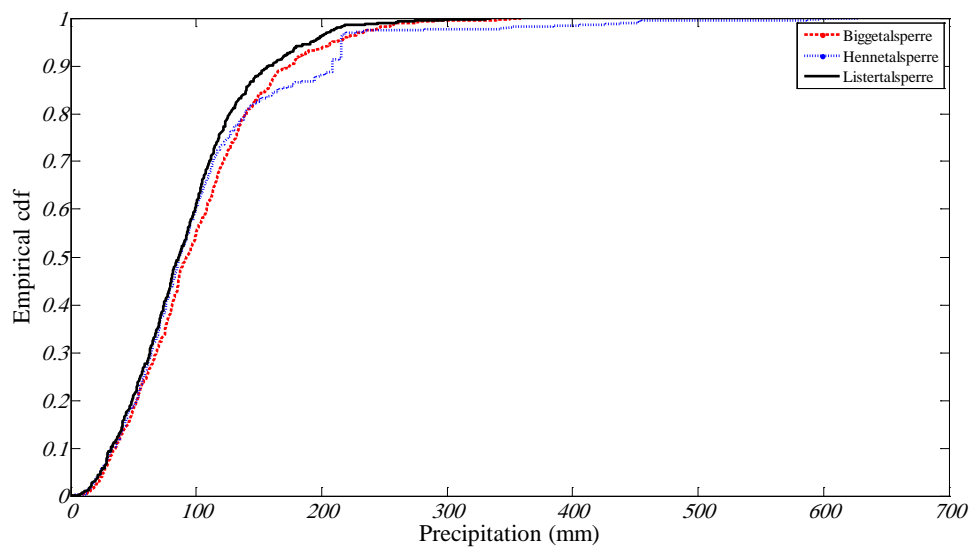
Figure 2.8: Empirical cumulative distribution function (CDF) for the precipitation data series (1961-2007)
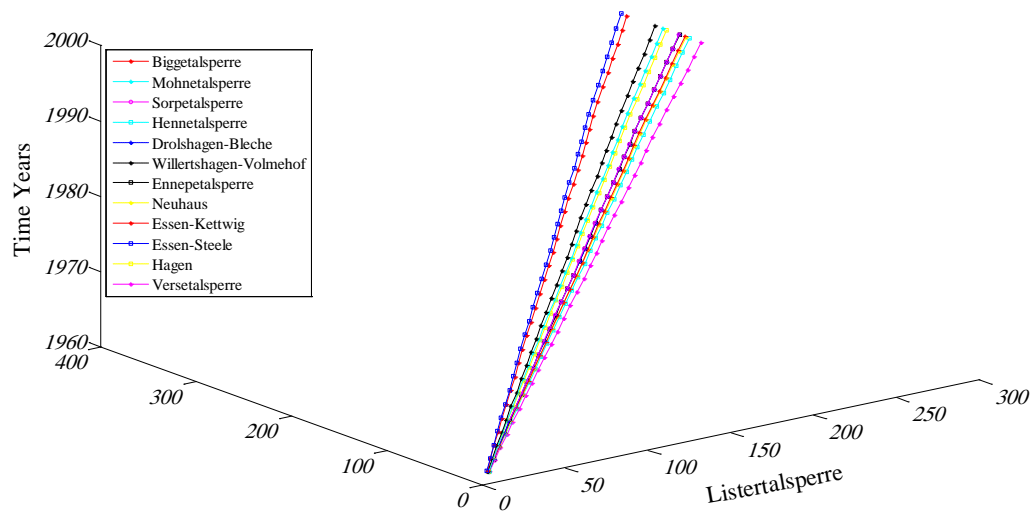
Figure 2.9: Empirical cumulative distribution function (CDF) for the temperature data series (1961-2007)
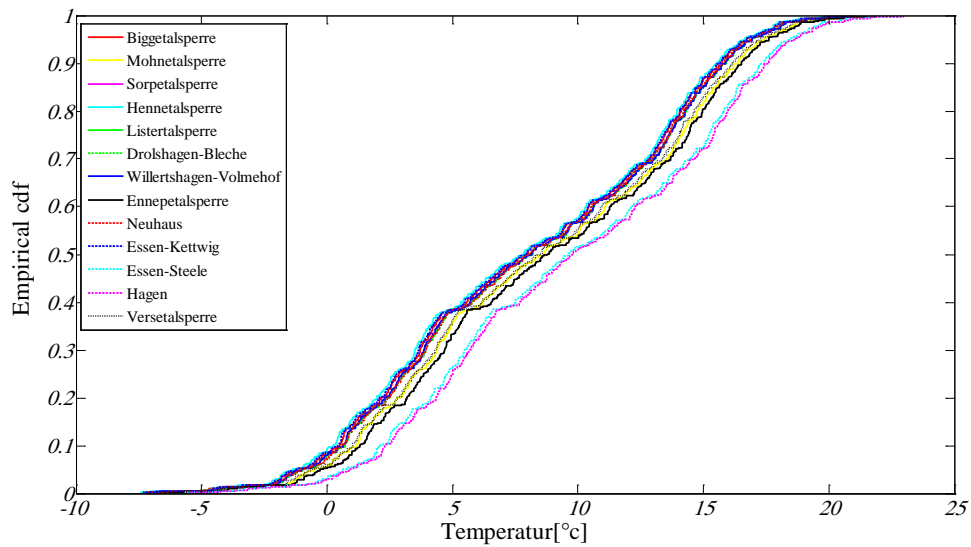
Figure 2.10: Empirical cumulative distribution function (CDF) for the temperature data series (1961-2007)

2.4 Temperature Analysis

In this study 13 stations (table 2.1) with daily air temperature records have been used to investigate the trend of the temperature in the Ruhr river basin. Results of correlation between the stations showed that there is high correlation (≈1) among the stations. In order to analyze the temperature records, several time scales have been considered (months, winter, summer and annual) to study the behavior of the temperature trend for these time scales. Also the temperature records have been classified to mean, minimum and maximum records for the considered time scales to examine the trends of the extreme values, occurrence of warm days and occurrence of cold days for both the winter and the summer. In many climate data series, a trend may be exist only for a specified part of the time series which is called local trend. This may be at the beginning, midst, or at the end of the data series. So, the partial time series approach was provided to examine the trend for different time periods

2.4.1 Mean Daily Temperature

The mean values of a given temperature data series are basic climatic parameters that are widely used in the studies of the climate change. The data series of the temperature as mentioned before are daily records. The mean values of the months, the winter and the summer have been computed within the study period. The results of the different time scales have been analyzed by applying the Mann-Kendall trend test. Confidence levels of 80 %, 90 % and 95 % were taken as thresholds to classify the significance of positive and negative precipitation trends. Trends at significance below the 80 % confidence level were not considered.

Figure 2.11 is a plot of the mean monthly temperatures for station Sorpetalsperre. In general there is an increase in the monthly temperature. The Mann-Kendall test shows that there is a significant trend in the mean monthly temperature at confidence levels of 95 % and the estimated trend is 0.0036 ^{0}C per month. Results of winter data series (figure 2.12) show that there is a significant positive trend in the winter mean temperature at confidence levels of 95 % with estimated trend equal to 0.041 ^{0}C per year.

Figure 2.13 illustrates the summer mean temperatures for station Sorpetalsperre. As shown in the figure, the mean temperature is increasing during the summer. The increase of the winter mean temperature is more pronounced than the summer mean temperature. Results of Mann-Kendall test show that there is a significant trend in the mean summer temperature at confidence levels of 95 % and the estimated trend is 0.037 ^{0}C per year. When the annual time scale is considered, the result (figure 2.14) shows significant increase in the mean annual temperature with an increase of 0.039 ^{0}C per year. Results of the mean temperature analysis showed that for the four examined time scales (months, winter, summer and annual) there is in general a significant increase in the mean temperature within the study period. Table 2.3 presents the results of Mann-Kendall test of the above mentioned time scales.

Table 2.3: Result of Mann-Kendall test (Trend analysis) - Mean temperature

Data series	T- value	Estimated trend	Confidence level		
			80 %	90 %	95 %
Monthly data series	2.38	+ 0.0036 ^{0}C per month	-	-	Yes
Winter data series	3.48	+ 0.04146 ^{0}C per year	-	-	Yes
Summer data series	4.035	+ 0.03721 ^{0}C per year	-	-	Yes
Annually data series	4.355	+ 0.03961 ^{0}C per year	-	-	Yes
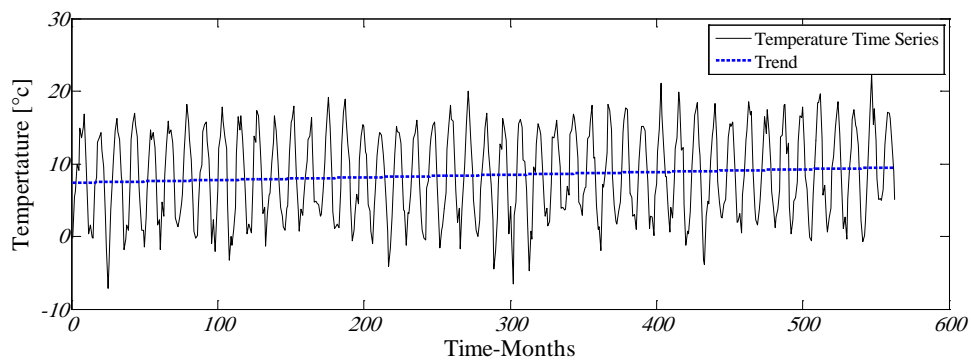
Figure 2.11: Fluctuations and trends of mean daily temperature for station Sorpetalsperre (Months time scale 1961-2007)
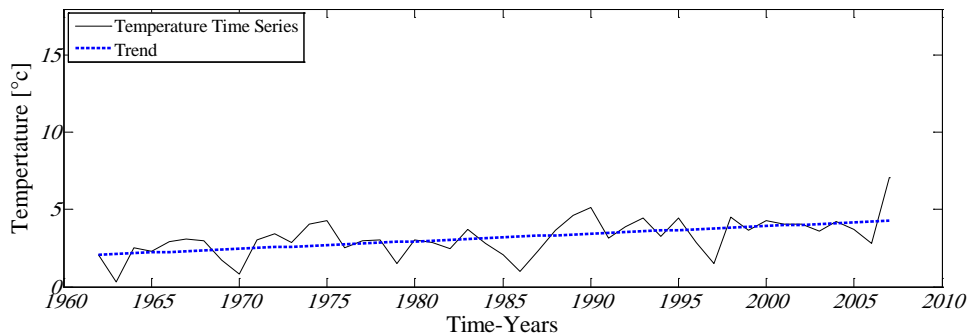
Figure 2.12: Fluctuations and trends of mean daily temperature for station Sorpetalsperre (Winter time scale 1961-2007).
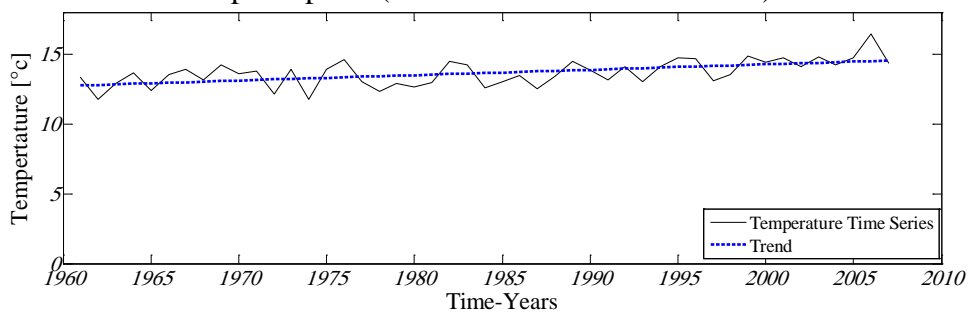
Figure 2.13: Fluctuations and trends of mean daily temperature for station Sorpetalsperre (Summer time scale 1961-2007).
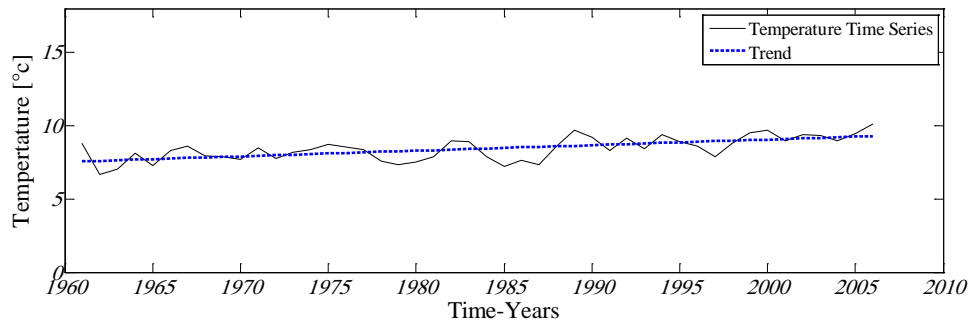
Figure 2.14: Fluctuations and trends of mean daily temperature for station Sorpetalsperre (Annual time scale 1961-2007).

2.4.2 Maximum and Minimum Mean Daily Temperature

Results of the min and max of mean daily temperature analysis in the Ruhr basin are in good agreement with the analysis done by Brazdil in 1994 (see (Heino, 2004)). The analysis which was performed over the min and max of T_{mean} ($T_{mean,min}$ - $T_{mean,max}$) data series has detected an increase in in the winter (Figure 2.15). The trend test shows that this increase is significant at 90 % confidence level. On the other hand, an increase in $T_{mean,min}$ and $T_{mean,max}$ in the summer is detected as shown in Figure 2.16. This increase in $T_{mean,max}$ is significant at 95 % confidence level but the increase in $T_{mean,min}$ is significant at 90 % confidence level. This means that the winter became warmer and the summer became hotter. Results of the annual analysis (Figure 2.17) show that the increase of $T_{mean,max}$ is more significant than the increase in $T_{mean,min}$. Results of the Mann-Kendall test are shown in table 2.4.

Table 2.4: Result of Mann-Kendall test (Trend analysis) – $T_{mean,min}$ and $T_{mean,max}$

Data series		T- value	Estimated trend	Confidence levels		
				80 %	90 %	95 %
Winter	$T_{mean,min}$	1.467	+ 0.0488 ^{0}C per year	-	Yes	No
	$T_{mean,max}$	1.344	+ 0.0393 ^{0}C per year	-	Yes	No
Summer	$T_{mean,min}$	1.94	+ 0.0394 ^{0}C per year	-	Yes	No
	$T_{mean,max}$	2.73	+ 0.0472 ^{0}C per year	-	-	Yes
Annual	$T_{mean,min}$	0.557	+ 0.0203 ^{0}C per year	No	-	-
	$T_{mean,max}$	2.54	+ 0.046 ^{0}C per year	-	-	Yes
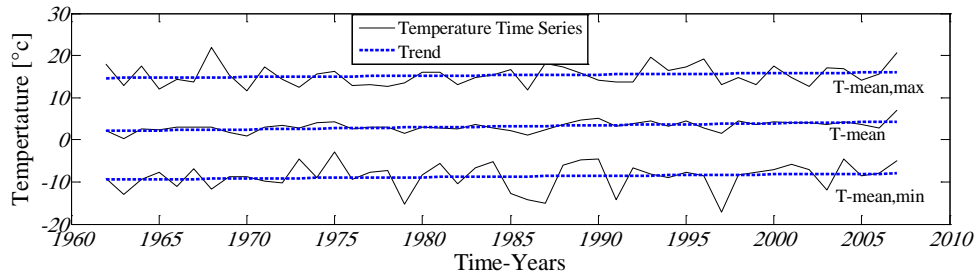
Figure 2.15: Fluctuations and trends of T_{mean}, $T_{mean,min}$ and $T_{mean,max}$ for station Sorpetalsperre (winter-1961-2007)
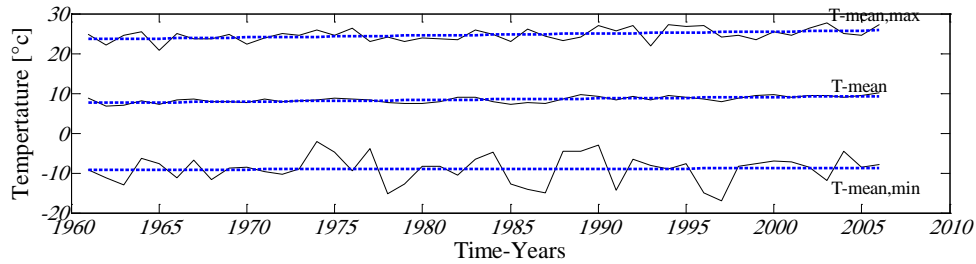
Figure 2.16: Fluctuations and trends of T_{mean}, $T_{mean,min}$ and $T_{mean,max}$ for station Sorpetalsperre (summer-1961-2007)
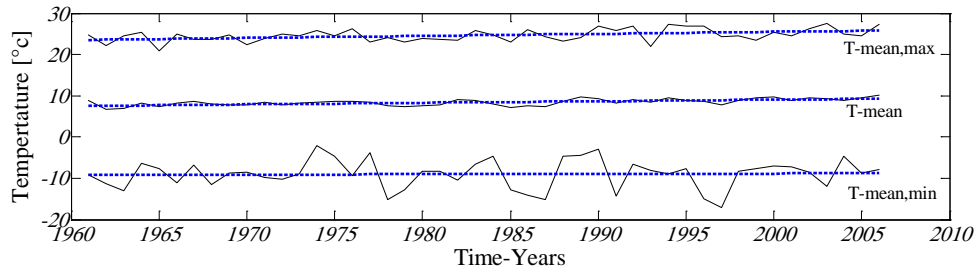
Figure 2.17: Fluctuations and trends of T_{mean}, $T_{mean,min}$ and $T_{mean,max}$ for station Sorpetalsperre (Annual -1961-2007)

2.4.3 Warm and Cold Days

The occurrence of cold and warm days is very useful for the detection of changes of climate. Trends of the occurrences of cold and warm days in both winter and summer have been examined based on Percentile indices (Alexander, 2005; Tank et al., 2005). The Percentile indices are *Tw10* %, *Tw90* %, *Ts10* % and *Ts90* % to detect cold/winter, warm/winter, cold/summer and warm/summer respectively. The result of the Percentile indices shows a significant positive trend in the occurrence of warm days during summer (figure 2.18) and negative trend in the occurrence of cold days in the summer as well (figure 2.19). On the other hand, a strong positive trend has been detected in the winter warm days (figure 2.20). And a strong negative trend in the winter cold days has been also detected (figure 2.21). This means that the number of warm days increased in both the summer and in the winter as well.
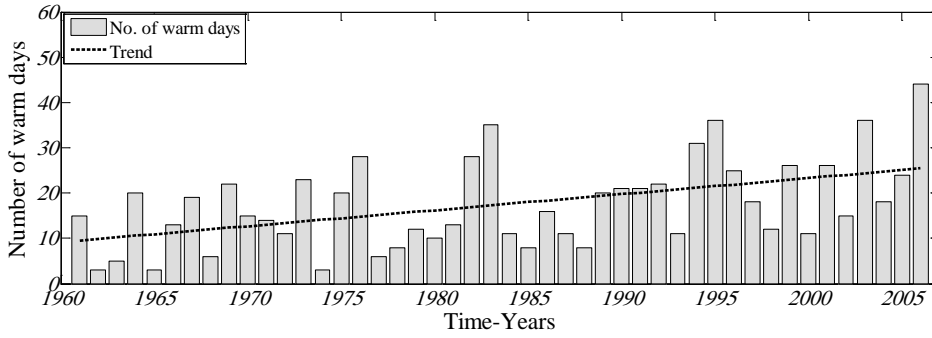
Figure 2.18: Time series of occurrence of warm days for station Sorpetalsperre (Summer time scale 1961-2007).
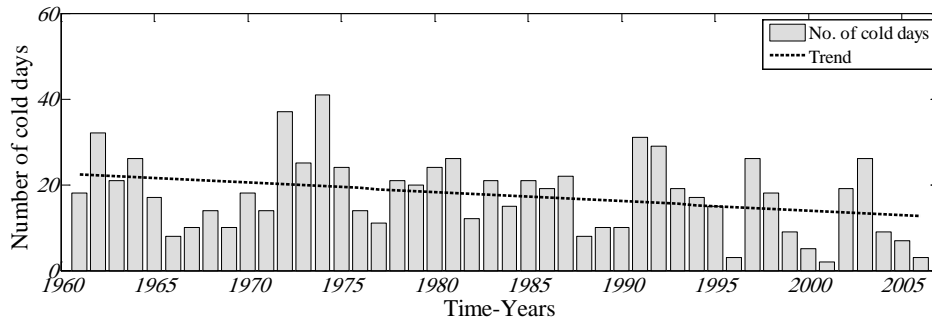
Figure 2.19: Time series of occurrence of cold days for station Sorpetalsperre (Summer time scale 1961-2007).
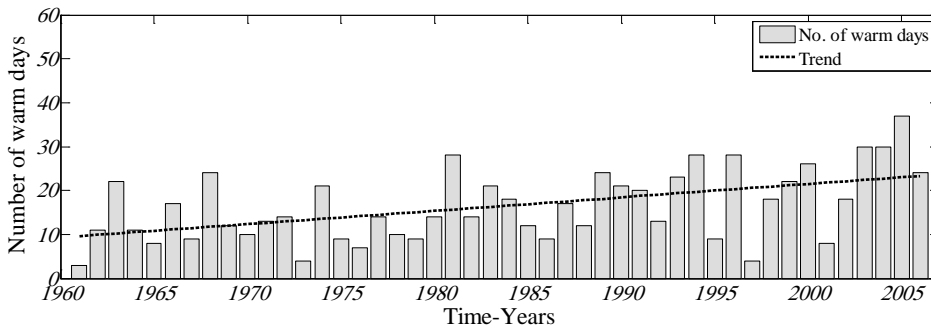
Figure 2.20: Time series of occurrence of warm days for station Sorpetalsperre (Winter time scale 1961-2007).
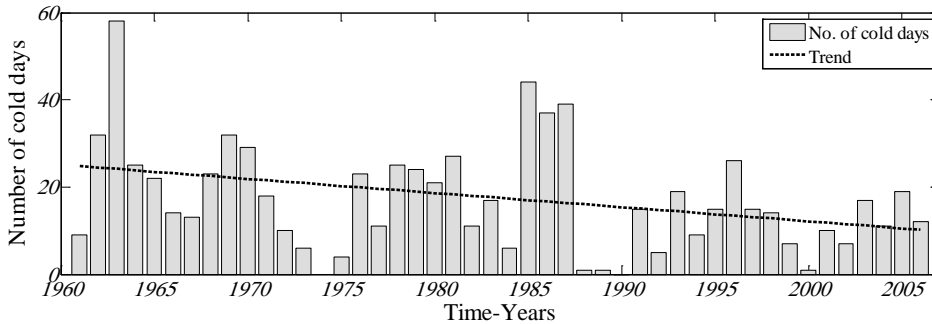
Figure 2.21: Time series of occurrence of cold days for station Sorpetalsperre (Winter time scale 1961-2007).

2.5 Precipitation Analysis

Precipitation over a catchment is the most important climatic factor for hydrological response. The precipitation data series which have been used in this study are from 1960 to 2007. Table 2.5 displays the correlation coefficient between stations. In order to study the behavior of the change in the amount of precipitation within the study period, the amount of precipitation has been calculated over several time scales (months, winter, summer and annual). The non-parametric Mann-Kendall test has been applied to the several time scales to distinguish the significance of the trend and to find the corresponding estimated trend. In many time series the global trend within a specified period is insignificant, but if the same time series is divided into more than one part the results might be different. So in this study, the time series of the several time scales was divided into several parts, then the trend of each part was individually examined to classify the several periods within the study period.

Table 2.5: Correlation factor between stations: Precipitation (1961-2007)

Station No. in Table (1)	1	2	3	4	6	8	9	14	23	27	30	35	39
1	1												
2	0.72	1											
3	0.81	0.86	1										
4	0.65	0.72	0.72	1									
6	0.96	0.74	0.84	0.67	1								
8	0.82	0.81	0.78	0.67	0.86	1							
9	0.84	0.7	0.76	0.63	0.86	0.87	1						
14	0.82	0.76	0.85	0.64	0.82	0.71	0.82	1					
23	0.76	0.9	0.82	0.69	0.79	0.87	0.77	0.73	1				
27	0.74	0.68	0.71	0.58	0.75	0.74	0.82	0.77	0.72	1			
30	0.75	0.68	0.72	0.59	0.76	0.74	0.82	0.78	0.73	0.94	1		
35	0.86	0.78	0.85	0.67	0.87	0.81	0.88	0.88	0.83	0.84	0.86	1	
39	0.91	0.76	0.87	0.67	0.92	0.8	0.85	0.88	0.8	0.77	0.78	0.92	1

2.5.1 Distribution Changes and Trends

The analysis of precipitation data series shows that different precipitation trend patterns occurred in the Ruhr basin in the study period. Figure 2.22 displays the fluctuations of the summation of the monthly precipitation for station Listertalsperre. Trend test shows a significant positive trend in the monthly precipitation, while the winter, summer and the annual data series have an insignificant positive trend as shown in Figure 2.23, 2.24, 2.25 respectively. To study the behavior of the trend of each month individually, the amount of precipitation for each month was calculated, then the Mann-Kendall test has been applied. Results of the Mann-Kendall test are shown in table 2.6.

Table 2.6: Result of Mann-Kendall test (Trend analysis)

Data series	Mean (mm)	T-value	Estimated trend (mm) per year	Confidence level
Months	100	1.45	0.0194	80 %
Winter (Nov.-Apr.)	650	1.1	1.87	-
Summer (May-Oct.)	550	0.84	1.008	-
Annual	1200	0.825	1.985	-
November	115	-0.346	-0.242	-
December	136	0.142	0.170	-
January	126	1.022	0.903	-
Febraury	91	1.679	1.049	90 %
March	101	1.537	0.774	80 %
April	75	-0.595	-0.290	-
May	77	-0.079	-0.053	-
June	91	-0.951	-0.421	-
July	100	0.488	0.236	-
August	89	-0.231	-0.111	-
September	92	0.951	0.410	-
October	98	0.222	0.122	-
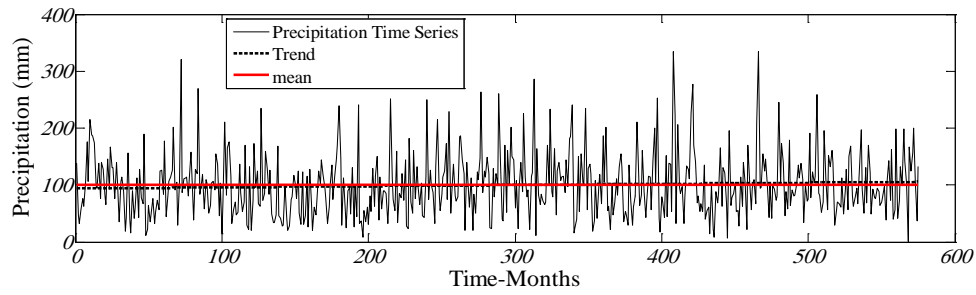
Figure 2.22: Trend analysis of the monthly precipitation for station Listertalsperre (Reference period 1960-2007)
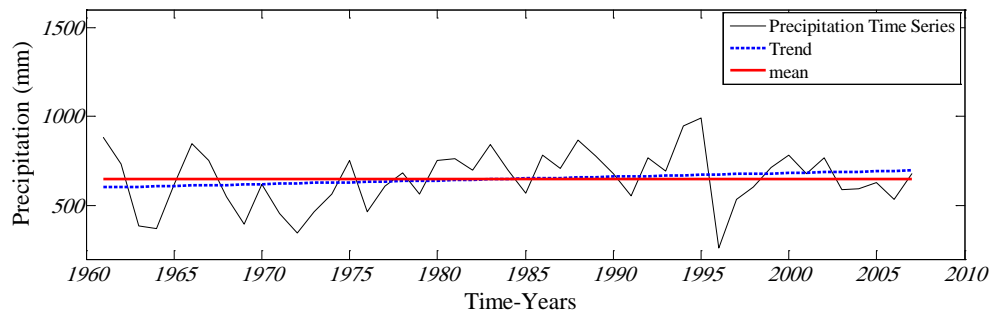
Figure 2.23: Trend analysis of the winter precipitation for station Listertalsperre (Reference period 1960-2007)
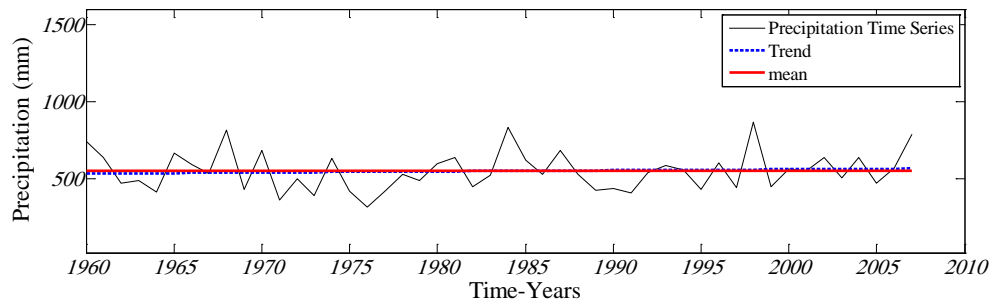
Figure 2.24: Trend analysis of the summer precipitation for station Listertalsperre (Reference period 1960-2007)
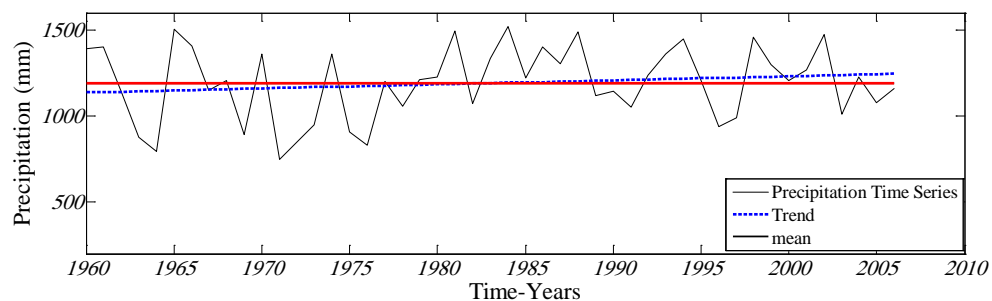
Figure 2.25: Trend analysis of the annually precipitation for station Listertalsperre (Reference period 1960-2007)

Morgenschweis (Morgenschweis et al., 2007) examined the trend of the precipitation in the Ruhr basin in the period 1927-2005. Results obtained show an increase in winter precipitation with a significant trend at 98 % confidence level (figure 2.26). Results show also insignificant trend for the summer precipitation. In fact there is no conflict between these results and the results shown in Figure 2.23 and table 2.7. It is well known that the length of the hydrological time series has a greet affect on the results of any trend test and the length of the data series in the two studies is not the same. Furthermore, it is very clear from figures 2.25 and 2.26 that the frequency of the precipitation within the period 1961-2005 is more or less the same and this was expected.

As mentioned before, the global trend for a given data series may present a significant /insignificant increase/decrease within the study period. But locally, if the data series is divided into several parts the data series may contain local insignificant /significant decrease/increase and vice versa. Figure 2.27 displays an application for the pervious approach. When the winter data series was divided into two parts (1960-1995, 1995-2007), results showed that an significant increase in the winter precipitation has taken place within the period 1961-1995. Significant positive trend is detected (95 %) in the first part and an insignificant negative trend in the second part as shown in Figure 2.28.
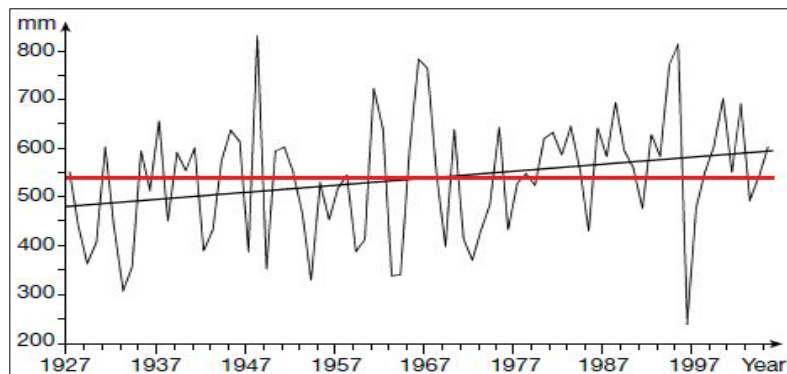
Figure 2.26: Trend analysis of the annually precipitation in the Ruhr basin
(Reference period 1927-2005) (Morgenschweis et al., 2007)
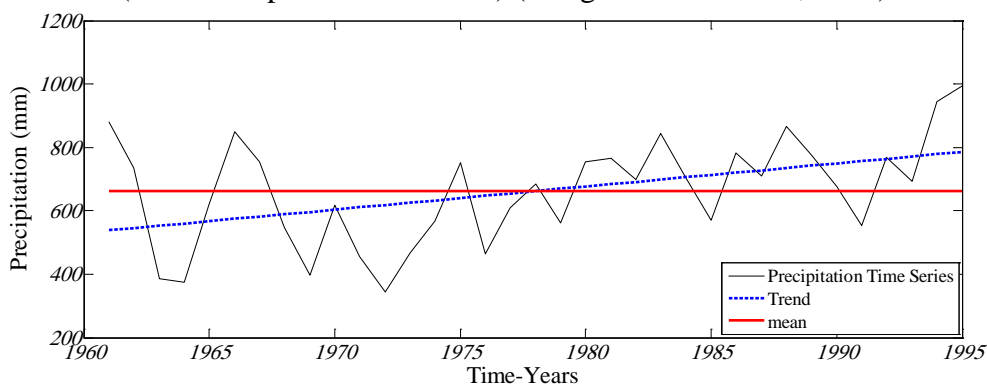
Figure 2.27: Trend analysis of the winter precipitation for station Listertalsperre
(Reference period 1960-1995)
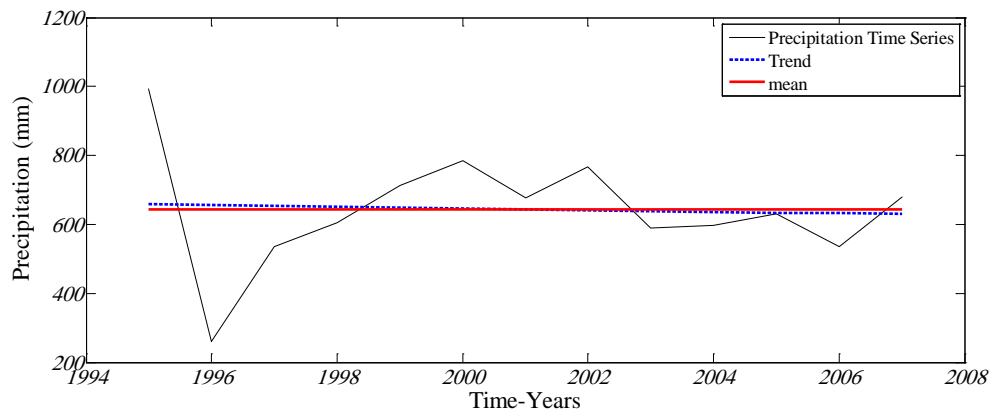
Figure 2.28: Trend analysis of the winter precipitation for station Listertalsperre
(Reference period (1997-2007)

2.5.2 Days with No Precipitation

Days with no precipitation have been detected and examined and the trend has been
calculated. The days which have been detected are the consecutive days. Five time
scales were examined namely 3 days, 7 days, 14 days, 21 days and 28 days with no
precipitation and this has been applied to the winter and the summer data series. In this
study the day without precipitation is defined as the day within the amount of
precipitation less than 0.10 mm. Results obtained in this part show an indication
towards negative insignificant trends during the winter in the 3 days ,7 days and 14
days data series (figures 2.29, 2.30, 2.31 and table 2.7). Results of 21 days data series
show few individual events within the study periods, also for 28 days data series only
the year 2007 contains this type of consecutive dry days and this event are in April and
May 2007. The month April in this year had an extreme drought event as shown later in
chapter 3.
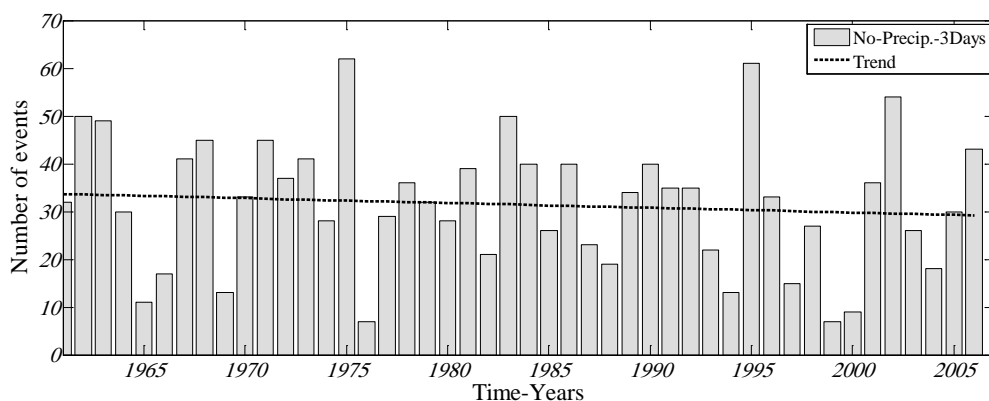
Figure 2.29: Time series of occurrence of no precipitation days for station
Listertalsperre (3 days time scale–Winter. 1960-2007).
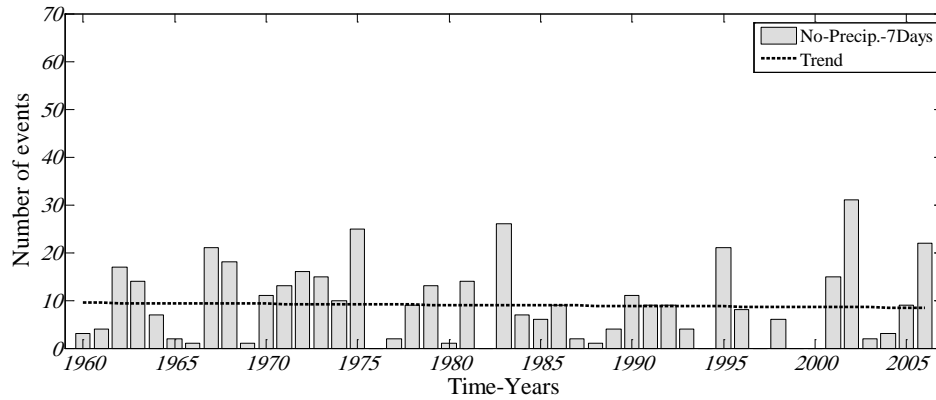
Figure 2.30: Time series of occurrence of no precipitation days for station Listertalsperre (7 days time scale–Winter.1960-2007).
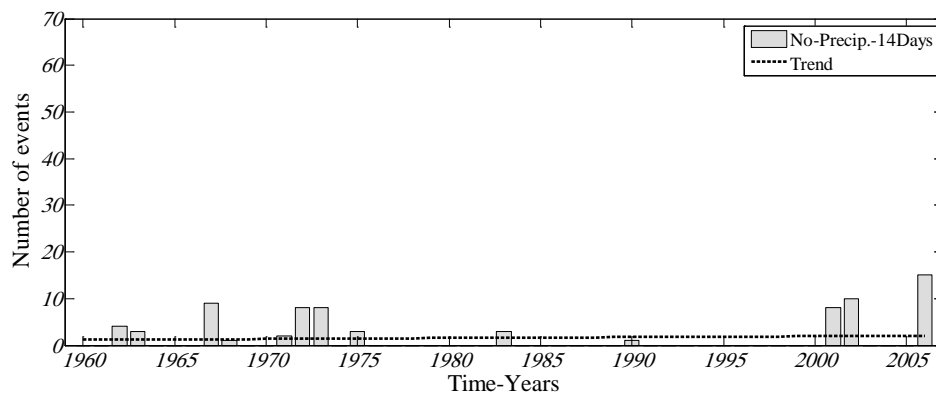
Figure 2.31: Time series of occurrence of no precipitation days for station Listertalsperre (14 days time scale–Winter.1960-2007).

Table 2.7: Result of the Mann-Kendall test (Trend analysis)– days with no precipitation

Data series		T- value	Estimated trend	Confidence levels		
				80 %	90 %	95 %
Winter	3 days	-0.972	-0.167 day/year	-	-	-
	7 days	-0.883	-0.058 day/year	-	-	-
	14 days	-0.84	-	-	-	-
Summer	3 days	0.017	-	-	-	-
	7 days	0.231	-	-	-	-
	14 days	0.328	-	-	-	-

2.5.3 Frequency Distribution of Very and Extremely Wet Days

The occurrence of very wet and extremely wet days gives evidence about the change in the intensity of the precipitation during the study period. Very and extremely wet days have been detected based on the precipitation indices $PR95$ % and $PR99$ %. Number of days (per year/season/month) with precipitation amount above a site specific threshold value for very and extremely wet days, were calculated as the 95th ($PR95$ %) and 99th ($PR99$ %) percentile of the distribution of daily precipitation amounts at days with 1 mm in the 1961–2007 baseline period. Let $PR_{w,j}$ be the daily precipitation amount at a wet day w (precipitation$\geq$ 1 mm) in period j and let $PRn95$ be the 95th percentile of precipitation at wet days in the 1961–2007 baseline period. Then the very wet days with $PR_{w,j}> PRn95$ are counted. The extremely wet days are calculated likewise.

Results of trend analysis for the winter and the summer precipitation show that there are insignificant trends in vey and extremely wet days in the summer (figures 2.32, 2.33).For the winter, an insignificant trend in the extremely wet days data series is obtained, while a positive significant trend is detected in the very wet days data series as shown in figures 2.34and 2.35.
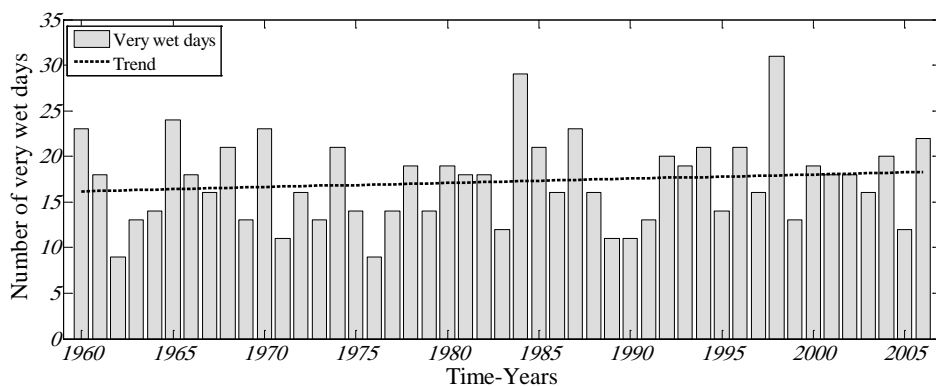
Figure 2.32: Time series of occurrence of very wet days for station Listertalsperre (Summer time scale 1960-2007).
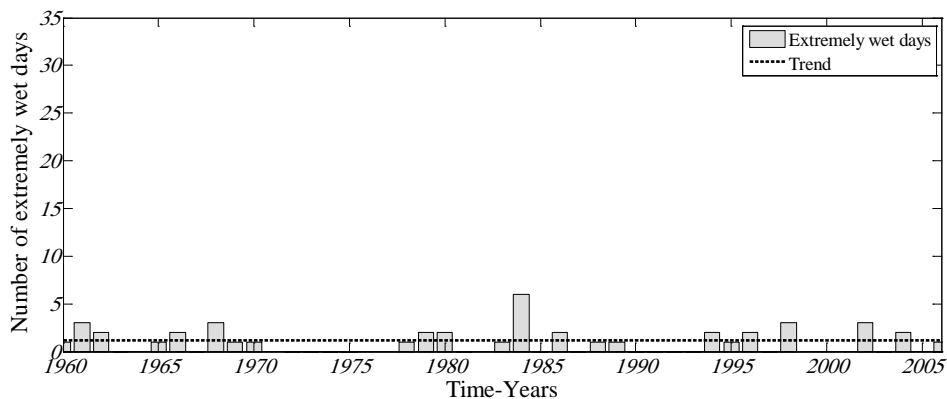
Figure 2.33: Time series of occurrence of extremely wet days for station Listertalsperre (Summer time scale 1960-2007).
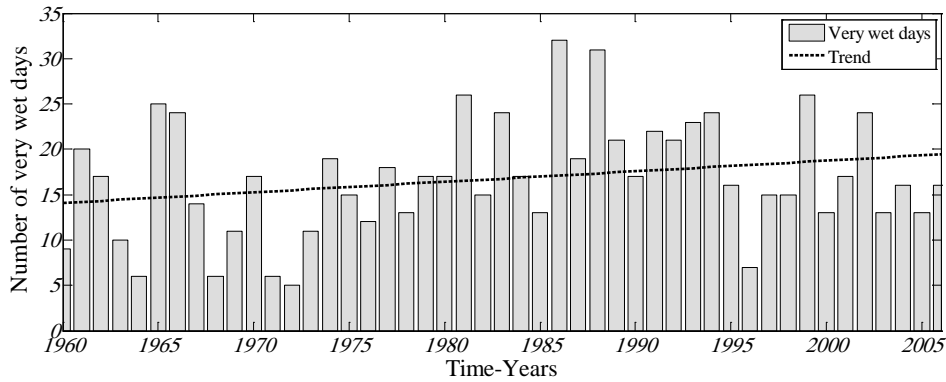
Figure 2.34: Time series of occurrence (days) of very wet days for station
Listertalsperre (Winter time scale 1960-2007).
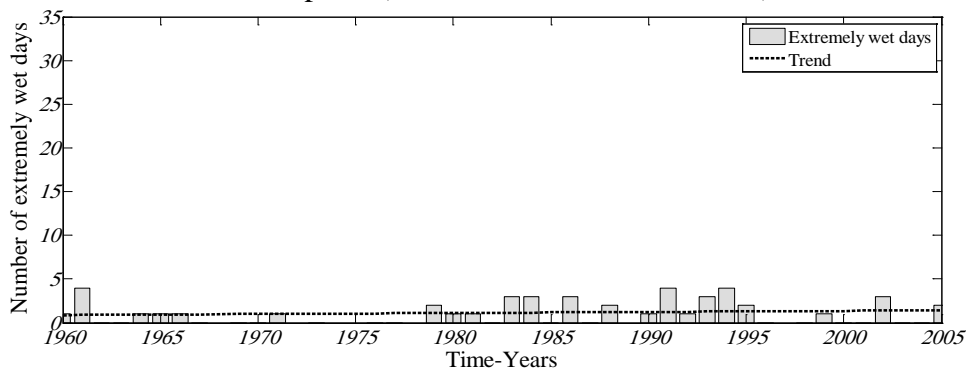
Figure 2.35: Time series of occurrence (days) of extremely wet days for station
Listertalsperre (Winter time scale 1960-2007).

2.6 Inflow Analysis

In this section, trends in inflow, that integrates the influence of atmospheric variables over a watershed, at the annual, seasonal and monthly time scales for the periods of records 1967-2008 are analyzed for the Ruhr River basin. Presumably, if consistent changes are observed in point measurements of precipitation and air temperature, these should also be reflected to some degree in streamflow at a watershed scale. As a spatially integrated variable streamflow is more appealing for detecting regional trends than point measurements of precipitation or temperature which is highly variable in space and time (Yan et al., 2007).

Data are daily inflow time series at the main four reservoirs in the Ruhr basin (table 2.2). High correlation between the inflow time series (monthly, seasonally) has been detected. For the summer and annual time scales (figures 2.37, 2.38) no significant trends have been detected; however for the winter time scale (Figure 2.36) a significant positive trend has been detected within the study period at 80 % confidence level. It is worth mentioning that when the winter data series were divided, a significant trend has been detected within the period 1967-1994 at 90 % confidence level. Results of the Mann-Kendall test are shown in table 2.8.

Table 2.8: Result of Mann-Kendall test (Trend analysis) - inflow

Data series	T- value	Estimated trend	Confidence levels		
			80 %	90 %	95 %
Months	0.61	0.00208 M.m^3 per year	-	-	-
Winter	1.314	0.764 M.m^3 per year	Yes	No	No
Summer	-0.2817	-0.067 M.m^3 per year	-	-	-
Annual	0.64	0.77 M.m^3 per year	-	-	-
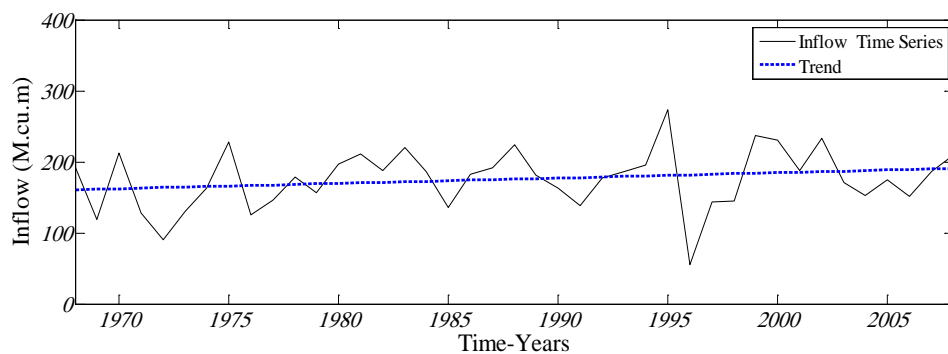
Figure 2.36: Fluctuations and trends of the inflow in the Ruhr basin
(Bigge Reservoir- Winter time scale. 1967-2008)
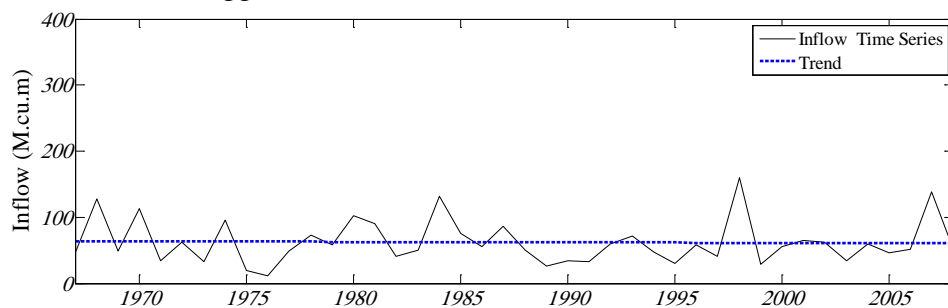
Figure 2.37: Fluctuations and trends of the inflow in the Ruhr basin
(Bigge Reservoir- Summer time scale. 1967-2008)
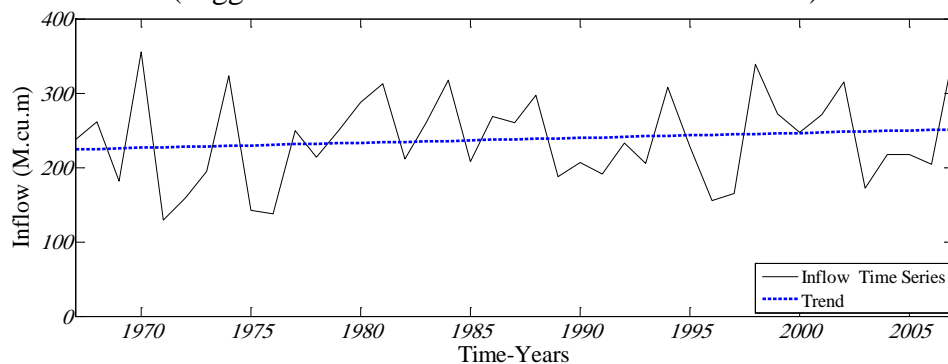
Figure 2.38: Fluctuations and trends of the inflow in the Ruhr basin
(Bigge Reservoir-Annual time scale. 1967-2008)

2.7 Correlation between Precipitation, Temperature and Inflow

Temperature for station Sorpetalsperre, precipitation for station Listertalsperre and inflow into the Bigge reservoir (Biggetalsperre) have been correlated together for daily, monthly, winter, summer and annually time scales. Coefficients of correlation are shown in Figure 2.39. The daily correlation between the precipitation and the inflow are calculated with different lags and one-lag correlation is the best one.

The correlation coefficient between temperature and precipitation is negative and relatively low except the coefficient correlation in the winter which is positive and relatively low. This would indicate that through all time scales except the winter as temperatures increase, precipitation tends to decrease. The positive correlation between temperature and precipitation could be interpreted as increasing temperatures could increase atmospheric water vapor, thus producing conditions conducive for increased rainfall. The correlation between the temperature and the inflow through all time scales except the winter is negative and relatively high in case of monthly and summer time scales (-0.56, -0.48 respectively). This may be due to the fact that with high temperatures, evapotranspiration would increase thus reducing streamflow.

Since precipitation is the driving force for all streamflow in the Ruhr basin, high positive correlation coefficients between precipitation and inflow through all time scales are expected. Variation in streamflow from year to year is found to be much strongly related to precipitation changes than to temperature changes and this is a common result in hydrological researches (Krasovskaia, 1995; Limbrunner, 2001). Precipitation and inflow are strongly correlated for winter, summer and annually data series and the correlation is significant as shown in Figure 2.39. Values of the correlation coefficient for the winter, the summer and the annual data series are (0.88), (0.92) and (0.90) respectively.
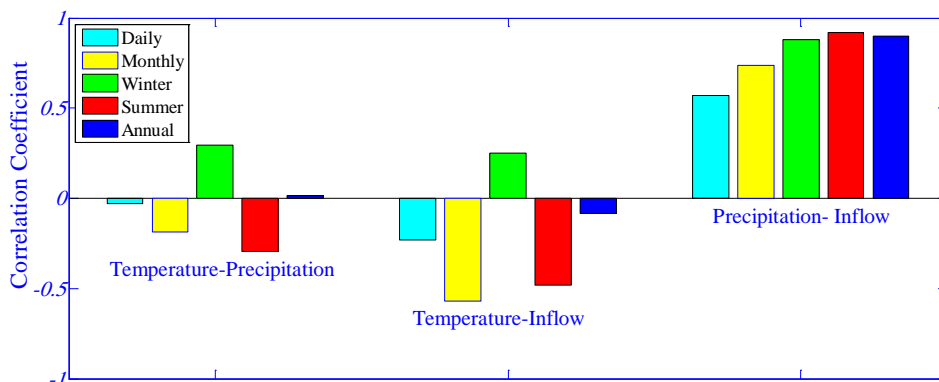
Figure 2.39: Correlation between Temperature, precipitation and Inflow

Temperature /Precipitation   :   Reference Period 1961-2007

Temperature /Inflow   :   Reference Period 1967-1995

Precipitation/Inflow   :   Reference Period 1961-2007

2.8 Conclusion

In this chapter, a study of climate change in the Ruhr basin has been presented using a set of data series for temperature, precipitation and inflow. Homogeneity of data series has been examined using several homogeneity tests. These tests, which have been used, included absolute and relative homogeneity tests. Results of homogeneity tests showed that some individual stations have inhomogeneous time series.

The findings regarding temperature and precipitation analysis are in good agreement with results obtained by Yeshewatesfa (Hundexha and Bardossy, 2005), Morgenschweis (Morgenschweis et al., 2007) and (Beck et al., 2004). After examining 68 years of precipitation and temperature data and 62 years of streamflow data, some statistically significant trends have been identified. There are slight increases in all three variables (streamflow, precipitation, temperature) in the Ruhr River basin.

Between 1961 and 2007, the results showed that a significant increase in the mean temperature over all time scales. The occurrence of warm days in both winter and summer has significant increase (95% confidence level) while the occurrence of cold days in both seasons showed a similar proportion of significant decrease. These results give evidence that the winter becomes warmer and the summer becomes hotter

Results of precipitation analysis gave evidence of a significant increase in the winter precipitation while the increases in the summer and the annual precipitation were statistically insignificant. The number of consecutive dry days displayed decreasing tendencies in winter while there is no indication of statistically significant change in the summer. The index PR95 % (very wet days), PR99 % (extremely wet days) have been introduced in this study to explore the supposed amplified response of the extreme precipitation events relative to the change in total amount. The main identified trends are an increase in the very wet days during winter.

For the inflow data series, obtained results showed that there is a significant increase in the winter inflow while the increases in the summer and the annual inflow were found to be statistically insignificant. Correlation calculations which have been applied to the data series showed that variations in streamflow from year to year were much more strongly related to precipitation changes than to temperature changes and this is a common result in hydrological researches.

Chapter 3

3. Analysis of Meteorological Drought in the Ruhr Basin by Using the Standardized Precipitation Index

3.1 Background

Drought is considered by many researchers to be the most complex but least understood of all natural hazards, affecting more people than any other hazard (Sivakumar et al., 2005). Drought is one of the major weather related disasters which is persisting over months or years. It can affect large areas and may have serious environmental, social and economic impacts. These impacts depend on the severity, duration, and spatial extent of the precipitation deficit, but also and to a large extent on the socio-economic and environmental vulnerability of affected regions (European Commission, 2008b). When thinking of natural hazards, droughts are often perceived by society to play a less dominant role compared to floods. Unlike the effects of a flood which can be immediately seen and felt, droughts build up rather slowly, creeping and steadily growing (Lehner et al., 2001).

As drought is a slowly developing phenomenon, only indirectly affecting human life, its impacts are often underestimated in financially well off regions such as Europe (Stahl, 2001). Droughts often result in heavy crop damage and livestock losses, disrupt energy production and hurt ecosystems. Drought mortality is concentrated in developing countries, while absolute economic losses are largest in developed regions. Drought is a major natural hazard affecting large areas and millions of people every year. The World Meteorological Organization (WMO) reported that in the 25 years from 1967 to 1991 about 1.4 billion people were affected by drought and 1.3 million people were killed due to the direct and indirect cause of drought (Obasi, 1994).

Drought differs from other natural hazards based on several specific features such as (Wilhite, 2005) :

- its unpredictability, due to its medium to long time scale of occurrence, while seasonal weather forecasts still present low levels of confidence and accuracy;

- its slow and progressive onset, cumulative through time, with events being detectable only when they are already occurring and impacts identified when drought has already become quite severe;

- its widely distributed impacts, affecting several components of the hydrological cycle and many economic sectors of human activity and persisting for a considerable period after the event itself has terminated.

The severe heat wave which started in Europe in June 2003 and continued through July until mid-August, raising summer temperatures up to 30 % higher than the seasonal mean in Celsius degrees over a large portion of the continent, extending from the Czech Republic to northern Spain and from Italy to Germany (UNEP, 2004). Extreme maximum temperatures of 35°C to 40°C were repeatedly recorded in July and to a larger extent in August in most of the southern and central countries from Germany to Turkey. A recent commonly study performed by the European Commission and Member States estimates the costs of droughts in Europe over the last thirty years to at least 100 billion Euro. The drought of 2003 in Central and Western Europe has been responsible for an estimated economic damage of more than 12 billion Euro (European Commission, 2008b). Table 3.1 presents some details about European droughts since 1970.

Table 3.1: Drought events in Europe 1970-2003. after (Lloyd-Hughes, 2002)).

Year	Region	Characteristics
1971	Most of Europe	Extremely dry year. The minimum rainfall in Spain during 30 years. An intensive summer drought in Poland. The water level in the Rhine reached lowest value since 1818.
1972	USSR	Lowest river levels for 50-80 years.
1973	North and east Europe	Very dry spring in eastern UK; low winter rain/snowfall in Austria, Germany and Czechoslovakia.
1974	Scandinavia, France, Holland	Dry spring in Norway , Denmark, Holland, Austria. 9 week spring drought in Sweden and low rainfall April-August in France.
1975	North and east Europe	Dry winter in eastern Europe. Summer rainfall in Sweden the lowest on record; October rainfall in Belgium lowest on record.
1976	Northern and eastern Europe	Severe drought in SE England, some parts in France and UK. Hot dry summer following a dry winter. Record rainfall deficits. Surface water and groundwater deficits. Low rainfall in Netherlands, Denmark, Norway, Sweden, and Scotland. Severe drought in some parts of Germany.
Year	Region	Characteristics
1977	UK	Dry summer from May to August. Scotland (mid) – driest summer since 1868.N Ireland - seventh successive summer with below average rainfall.
1988-92	Most of Europe	Anomalous circulation pattern caused rainfall deficits over a large area interspersed with short wet periods. Insured

		losses due to subsidence estimated at £600 m for UK alone
1990-95	Spain, Portugal	Prolonged drought across the entire Iberian peninsular. Water supplies in Seville were cut for up to 12 hours per day during 92-93. Hydroelectric power suspended 94-95.
1992-93	Bulgaria, Hungary	Very hot, dry summer 1992.Continued with below average rainfall to October 1993.Severe loss of agricultural production in Bulgaria. Worst drought in USSR for 10 years.
1995	Ireland, UK, Norway, Sweden	Hot, dry summer and autumn. Dry soil. Impact on surface water supplies but not groundwater. Low temperature, little winter snow in Nordic countries.
1995-96	Germany	Extremely dry winter in some parts of Germany
1996	Bulgaria	Hot, dry summer across whole country.
1999	Finland	Hot, dry summer in southern Finland. Very low water levels both in rivers and groundwater formations.
2003	Much of continental Europe	Many deaths from unusually prolonged high temperatures. Forest fires, subsidence, power cuts and agricultural losses.

3.2 Drought Definitions

Drought has no universal definition. Drought definitions reflect many disciplinary perspectives and therefore incorporate different biological, physicall and socioeconomic variables in their definitions. Most of drought definitions are region specific, reflecting differences in climatic characteristics. For this reason, it is usually difficult to transfer definitions derived for one region to another (AMS, 2009). Labedzki reported that Wilhite uncovered in the early 1980s more than 150 published definitions of drought (Labedzki, 2007; NDMC, 2006). The definitions reflect differences in regions, needs and disciplinary approaches.

Beran (Beran and Rodier, 1985) summarized that, in any case it is evident that the notion of drought is relative but its chief characteristic is a decrease of water availability in a particular period and over a particular area rather than a general decrease of water availability. Drought affects all components of the water cycle from a deficit in soil moisture reduced groundwater recharge and levels and to low streamflow or dried up rivers. It is a reoccurring and worldwide phenomenon, with spatial and temporal characteristics that vary significantly from one region to another (Khadr et al., 2009).

In general, drought gives an impression of water scarcity due to insufficient precipitation, high evapotranspiration and over-exploitation of water resources or combination of these parameters (Bhuiyan, 2004). The primary cause of a drought is the lack of precipitation over a large area and an extensive period of time; this type is called meteorological drought (Tallaksen and Lanen, 2004). This water deficit propagates to the hydrological cycle and gives rise to different types of droughts.

3.3 Classification of Drought

All types of drought originate from a deficiency of precipitation (Wilhite and Glantz, 1985). Droughts can be classified in four major categories:

3.3.1 Meteorological Drought

Meteorological drought, also termed climatological drought, is commonly based on precipitation's departure from normal average over a certain period of time and region, since deficiency of precipitation is highly variable from region to region.

3.3.2 Hydrological Drought

This type is associated with the deficiency of water on surface or subsurface due to shortfall in precipitation. Although all droughts have their origin from deficiency in precipitation, hydrological drought is mainly concerned about how this deficiency affects components of the hydrological system such as soil streamflow, moisture, groundwater and reservoir levels.

3.3.3 Agricultural Drought

This links several characteristics of both meteorological and hydrological drought to agricultural impacts, focusing on precipitation shortages, differences between actual potential evapotranspiration, soil, soil water deficits and reduced groundwater or reservoir levels. Plant water demand depends on prevailing weather conditions, biological characteristics of the specific plant, its stage of growth and the physical and biological properties of the soil.

3.3.4 Socio-Economical Drought

It is associated with the demand and supply aspect of economic goods together with elements of meteorological, hydrological and agricultural drought. This type of drought mainly occurs when the demand for an economic good exceeds its supply due to weather related shortfall in water supply.

3.4 Time Sequence of Drought Impacts

The sequence of impacts associated with meteorological, agricultural and hydrological droughts highlights its differences. When drought event begins, the first to suffer is usually the agricultural sector because of it is heavily dependence on stored soil water (Hisdal and Tallaksen, 2000). The latter can be rapidly depleted over extended dry periods. If no precipitation period continues, then people will begin to feel the effects of the shortage. Those who rely on surface water (i.e., reservoirs and lakes) will suffer first and those who rely on subsurface water (i.e., groundwater) are usually the last to be affected. Although, groundwater users, often the last to be affected by drought during its onset, they are the last to experience a return to normal water supply levels. Obviously, the length of the recovery period is a function of the intensity of the drought, its duration and the quantity of precipitation received following the drought period.

As schematically illustrated in figure 3.1, a drought event is caused by a certain meteorological situation, for instance a persisting anticyclone/ high pressure system. Associated with the prevailing dry and warm weather, a meteorological drought with a rainfall deficit develops. The rainfall deficit and the high evapotranspiration reduce the soil water content, which might cause an agricultural drought if it occurs during the growing season. Due to the precipitation deficit in the catchment, streamflow decreases until it is only fed by groundwater and finally the groundwater reservoirs will also deplete. Consequently, hydrological droughts lag the occurrence of atmospheric droughts and depending on the season and the crop also the occurrence of agricultural drought. Water in hydrological storage systems such as surface and groundwater reservoirs is often used for multiple and competing purposes, e.g. flood control, irrigation, recreation, hydropower, navigation or wildlife habitat, further complicating the sequence and quantification of impacts (Wilhite, 2005). When the demand exceeds the supply, a socio-economic drought occurs.
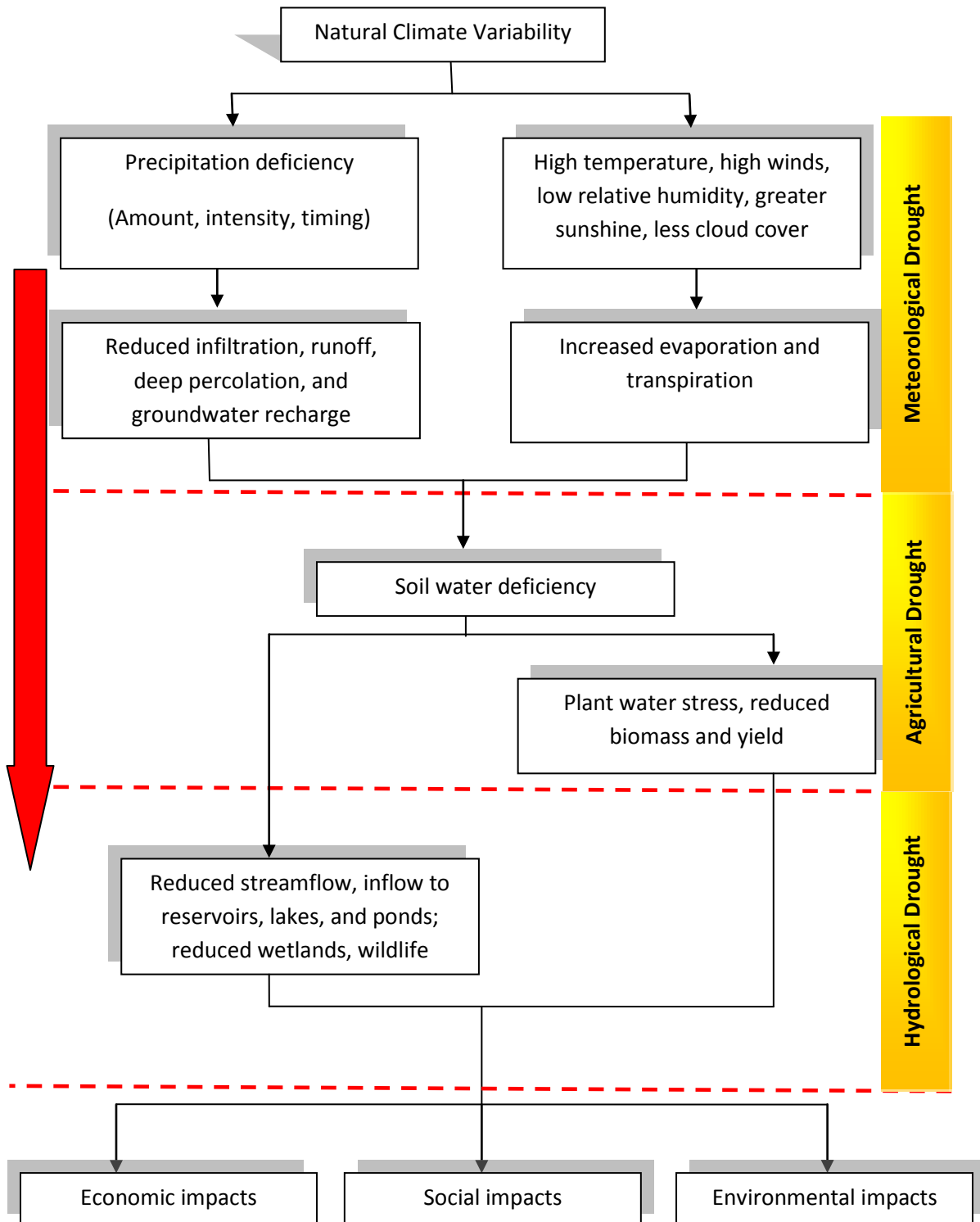
Figure 3.1: The sequence of drought impacts associated with meteorological, agricultural and hydrological drought. After (Wilhite, 2009).

3.5 Drought Indices

Drought indices assimilate thousands of data on rainfall, snowpack, streamflow and other water supply indicators into a comprehensible big picture. A drought index value is typically a single number, far more useful than raw data for decision making (NDMC, 2006). There are several indices that measure how much precipitation for a given period of time has deviated from historically established norms. Although none of the major indices is inherently superior to the rest in all circumstances, some indices are better suited than others for certain uses. Table 3.2 presents some of the widely used drought indices including Palmer Drought Severity Index (PDSI), Crop Moisture Index (CMI), Standardized Precipitation Index (SPI) and Surface Water Supply Index (SWSI).

Table 3.2: Different drought indices and their pros and cons (after (Awass, 2009))

Index	Pros	Cons	Developed by
PDSI/ PHDI	Non-dimensional, widely accepted specially in USA	Arbitrary threshold, may lag emerging droughts by several months less well suited for mountainous or of frequent climatic extremes	Palmer 1965
SPI	Identifies emerging droughts months sooner than the PDSI, Limited data input, can provide early warning of drought and help assess drought severity	Arbitrary threshold,	McKee et al. 1995
CMI	Identifies potential agricultural droughts.	Not a good long-term drought monitoring tool	Palmer 1965
SWSI	Representative measure of water availability across a basin, region,	It is difficult to compare SWSI values between basins or regions	Shafer and Dezman 1982

3.5.1 Standardized Precipitation Index

3.5.1.1 Definition of the Standardized Precipitation Index (SPI)

Standardized precipitation index (SPI) is based on an equi-probability transformation of aggregated monthly precipitation into a standard normal variable. In practice, computation of the index requires the fitting of a probability distribution to aggregated monthly precipitation series (e.g. k= 3, 6, 12, 24 months, etc), computing the non-exceedance probability related to such aggregated values and defining the corresponding standard normal quantile as the SPI. McKee (McKee et al., 1993) assumed an aggregated precipitation gamma distribution and used a maximum likelihood method to estimate the parameters of the distribution. SPI has advantages of statistical consistency and the ability to describe both short-term and long-term drought impacts through the different time scales of precipitation anomalies (Cancelliere et al., 2007). Its limitation is that it relies on one input. In general, different studies have indicated the usefulness of the SPI to quantify different drought types (Vicente-Serrano and L´opez-Moreno, 2005).

Precipitation is a climatological phenomenon more difficult to study than temperature, because it is discontinuous with some days receiving no precipitation, while other days receive abundant amounts of precipitation. For this reason, the basic measurement period for many precipitation studies is the total precipitation for each month. The Standardized Precipitation Index (SPI) is a tool developed by McKee (McKee et al., 1993) for the purpose of defining and monitoring local droughts. It was conceived to identify drought periods and the severity of droughts, at multiple time scales. Shorter or longer time scales may reflect lags in the response of different water resources to precipitation anomalies.

McKee (McKee et al., 1993) reported that a drought event occurs if the SPI is continuously negative and reaches an intensity of -1.0 or less. The event ends when the SPI becomes positive. Each drought event, therefore, has a duration defined by its beginning and the end, and an intensity for each month the event continues. The accumulated magnitude of the negative values of the SPI during a drought event can be considered as drought magnitude. The Standardized Precipitation Index (SPI) is an index widely used for drought monitoring purposes. Since its computation requires the preliminary fitting of a probability distribution to monthly precipitation aggregated at different time scales, the SPI value for a given year and a given month will depend on the particular sample of observed precipitation data adopted for its estimation and in particular on the sample size. Furthermore, the presence of a trend in the underlying precipitation will adversely affect the estimation of parameters and the computation of SPI (Cancelliere and Bonaccorso, 2009).

The calculation of SPI requires that there is no missing data in the time series. The data record length is required to be at least 30 years (Wu et al., 2001). A number of advantages arises from the use of the SPI index (Cacciamani et al., 2007). First of all, the index is simple and is only based on the amount of precipitation so that its evaluation is rather easy. Also the SPI index can be computed for multiple time scales (i.e., 1, 2, 3, . . . 72 months), thus allowing the comparison between different time periods. This can be an excellent communication tool to the public and to policy makers (Wilhite et al., 2000). In addition, these various time scales can be useful in assessing effects on different components of the hydrologic system (e.g., streamflow, reservoir levels and groundwater levels). McKee (McKee et al., 1993) used the classification system shown in table 3.3 to define dry and wet events.

The Standardized Precipitation Index aims to provide a concise overall picture of drought, regardless to the actual probability distribution of the observed cumulative amounts of rainfall for a given time scale (Gbete and Soumaila, 2007). It consists in realizations of standard Gaussian distribution with mean zero and variance one obtained by applying appropriate transformation to each of the observed cumulative amount of precipitation. But one should notice that applying the inverse of the cumulative probability function of the standard Gaussian distribution to the actual cumulative probability function of each observed amount of precipitation fails to give Gaussian deviates as precipitation data may include many zeros corresponding to period with no precipitation. In this study, there is no zero monthly precipitation for the covered period. SPI was applied on 1, 3, 6, 9, 12and 24 month time scales.

Table 3.3: Classification of drought based on the SPI index

SPI	Classification
2 or more	Extremely wet
1.5 to 1.99	Very wet
1 to 1.49	Moderately wet
0.99 to -0.99	Near normal
-1 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
-2 and less	Extremely dry

3.5.1.2 Computation of the SPI Index

McKee (McKee et al., 1993) developed the Standardized Precipitation Index (SPI) for the purpose of defining and monitoring drought. Among others, the Colorado Climate Center, the Western Regional Climate Center and the National Drought Mitigation Center use the SPI to monitor current states of drought in the United States. The nature of the SPI allows an analyst to determine the rarity of a drought or an anomalously wet event at a particular time scale for any location in the world that has a precipitation record.

In the most cases, the Gamma distribution is the distribution that best models observed precipitation data. The density probability function for the Gamma distribution is given by the expression (Cacciamani et al., 2007):

$$g(x) = \frac{1}{\beta^\alpha \, \Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}} \qquad\qquad For\ x>0 \qquad\qquad (3.1)$$

Where:

$\alpha > 0$ is a shape parameter

$\beta > 0$ is a scale parameter

x > 0   x is the precipitation amount

$\Gamma(\alpha)$ is the Gamma function and defined by:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy \qquad\qquad\qquad (3.2)$$

Computation of the SPI involves the fitting of a gamma probability density function to a given frequency distribution of precipitation totals for a station. The alpha and beta parameters of the gamma probability density function are estimated for each station, for each time scale of interest (1month, 3 months, 12 months, 48 months, etc.) and for each month of the year.

After estimating coefficient alpha and beta the density of probability function *g(x)* is integrated with respect to *x* and we obtain an expression for cumulative probability *G(x)* that a certain amount of rain has been observed for a given month and for a specific time scale.

$$G(x) = \int_0^x g(x)dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x x^{\alpha-1} e^{\frac{-x}{\beta}} dx \qquad\qquad (3.3)$$

The Gamma function is not defined by *x*= 0 and since there may be no precipitation the cumulative probability becomes:

$$H(x) = q + (1-q)G(x) \qquad\qquad\qquad (3.4)$$

Where q is the probability of no precipitation. The cumulative probability is then transformed into a normal standardized distribution with null average and unit variance from which we obtain the SPI index. The above approach, however, is neither practical nor numerically simple to use if there are many grid points or many station on which to calculate the SPI index. In this case, an alternative method was described in (M.V.K. Sivakumar) using the technique that converts the cumulative probability into a standard variable Z.

The SPI Index is then defined as:

$$Z = SPI = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \qquad \text{for } 0 < H(x) < 0.5 \qquad (3.5)$$

$$Z = SPI = +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \qquad \text{for } 0.5 < H(x) < 1 \qquad (3.6)$$

Where:

$$t = \sqrt{\ln\left[\frac{1}{\left(H(x)\right)^2}\right]} \qquad \text{for} \quad 0 < H(x) < 0.5 \qquad (3.7)$$

$$t = \sqrt{\ln\left[\frac{1}{\left(1 - H(x)\right)^2}\right]} \qquad \text{for} \quad 0.5 < H(x) < 1 \qquad (3.8)$$

Where x is precipitation, H(x) is the cumulative probability of precipitation observed and c_0, c_1, c_2, d_1, d_2 and d_3 are constants with the following values:

$c_0 = 2.515517$ $\qquad\qquad$ $c_1 = 0.802853$ $\qquad\qquad$ $c_2 = 0.010328$

$d_1 = 1.432788$ $\qquad\qquad$ $d_2 = 0.189269$ $\qquad\qquad$ $d_3 = 0.001308$

3.6 Data Collection and Methodology

The data set which has been used in this chapter is the same data which has been used in chapter 2. For more details, please see chapter 2, section 2.3.

In this study a program called "SPI_Analysis" was used to calculate and analyze the SPI values. This program has been developed by the author as a part of this thesis. Details about this program are presented in appendix A.

3.7 Drought Occurrences and Analysis

3.7.1 SPI Index of Consecutivel Months

In this study the overall meteorological drought vulnerability in the Ruhr river basin has been assessed by reconstructing historical occurrences of droughts at varying time steps and drought categories with the SPI approach. The basic idea is that this can be a guide to the decision makers in the Ruhr basin to develop strategies of water resources management in the context of drought. The SPI index is applied to long-term precipitation data at 13 stations for the period 1960-2007 (January 1960 to December 2007) (See chapter 2 for more details).

The occurrences in varying drought categories at 1, 3, 6, 9, 12 and 24 month time steps have been analyzed. The SPI values have been calculated for the total period and also for a specific month. Figure 3.2 through figure 3.4 illustrate the SPI values based on 1, 3, and 6 months time steps respectively. Appearance of drought is defined when SPI is negative and its intensity comes -1.0 or lower. Several drought events have been detected. These events have also different durations. As mentioned before the duration of an event is defined as the time between the zero crossings that bound the events.
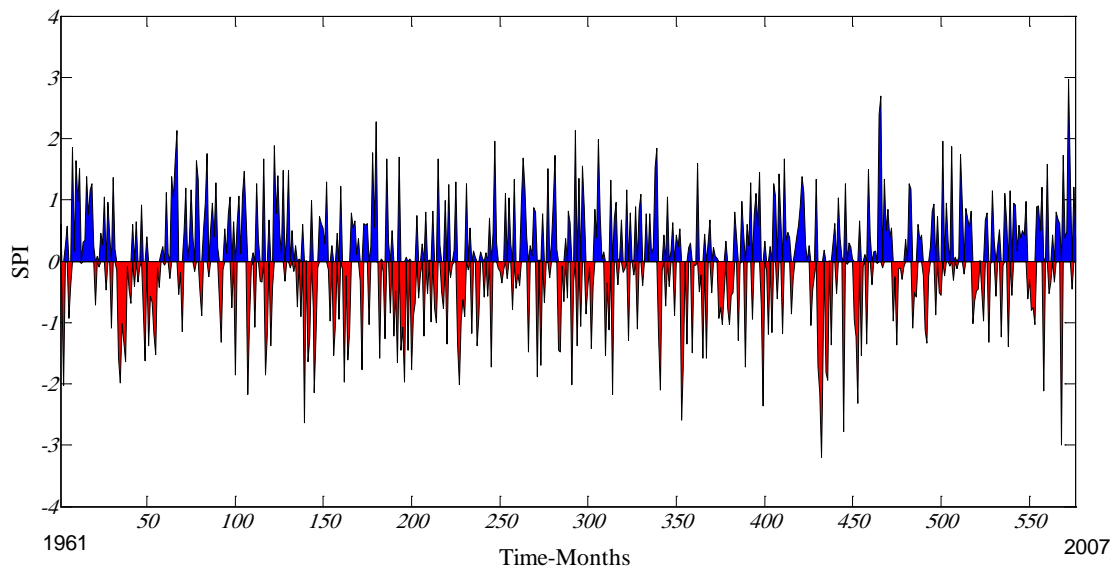
Figure 3.2: SPI time series based on the total monthly precipitation in the Ruhr River Basin (1960-2007) (One month time step – SPI_1) (Station Sorpetalsperre)
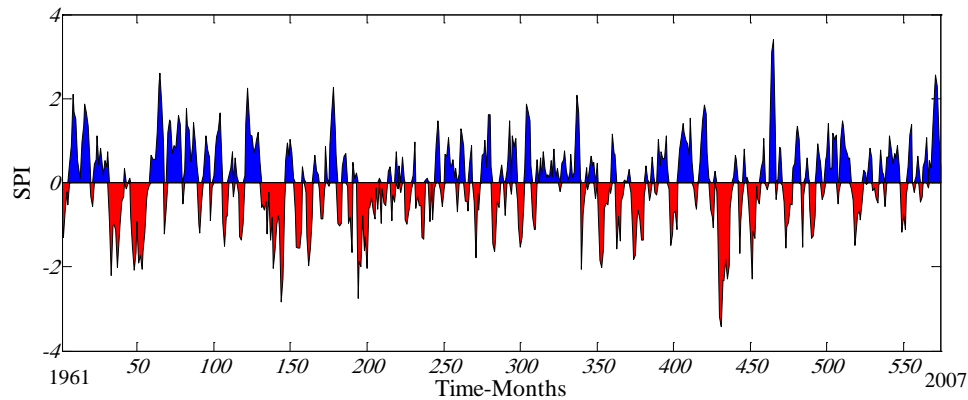
Figure 3.3: SPI time series based on the total monthly precipitation in the Ruhr River Basin (1960-2007) (Three months time step – SPI_3) (Station Sorpetalsperre)
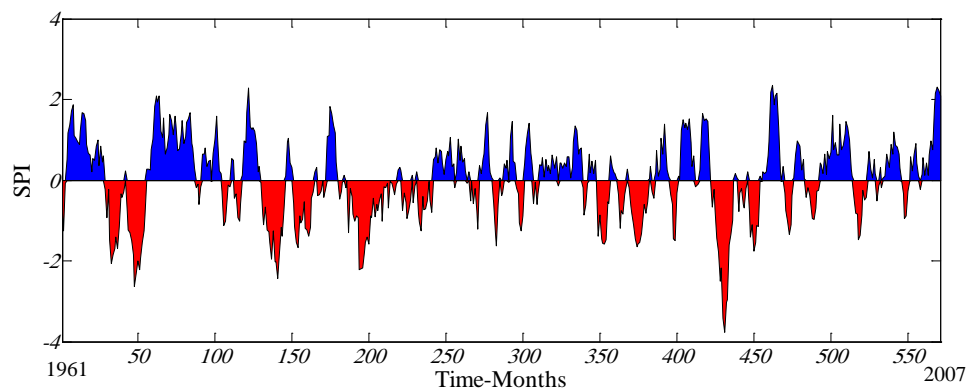
Figure 3.4: SPI time series based on the total monthly precipitation in the Ruhr River Basin (1960-2007) (Six months time step – SPI_6) (Station Sorpetalsperre)

Figure 3.5 illustrates the accumulated magnitude of the negative values of the SPI based on 1, 3 and 6 months time scale. The figure can be used as a guide for the selection of the driest years and to compare also between different droughts. As shown in the figure, several years (such as 1964, 1976 and 1996) exposed to sever drought.

Based on an analysis of stations across the Ruhr basin, results showed that SPI defines mild drought in 31.9 % of the time, moderate drought in 8.33 % of the time, severe drought in 5.5 % of the time and extreme drought in 1.3 % of the time. Because the SPI is standardized, these percentages are expected from a normal distribution of SPI. The 1.3 % of SPI values within the "extreme drought" category is a percentage that is typically expected for an "extreme" event (NDMC, 2006). This standardization allows determining the rarity of a current drought, as well as the probability of the precipitation necessary to end the current drought. Figure 3.6 shows that the probability of the occurrence of a dry or a wet event, according to the category, is approximately the same. The percentage of an event is the sum of the percentage of all similar events through the covered period (1960-2007).
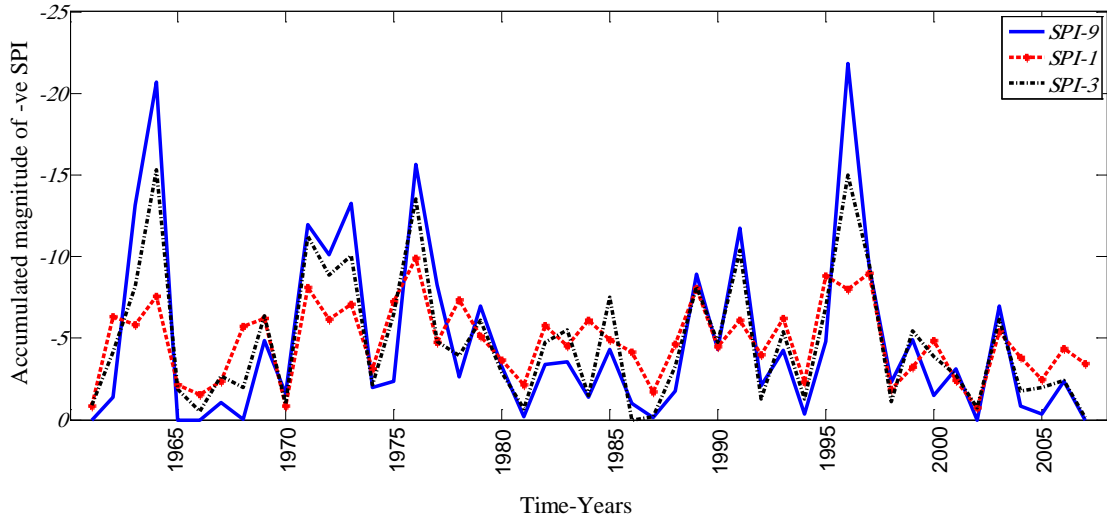
Figure 3.5: Accumulated magnitude of the negative values of the SPI
(Station Sorpetalsperre)
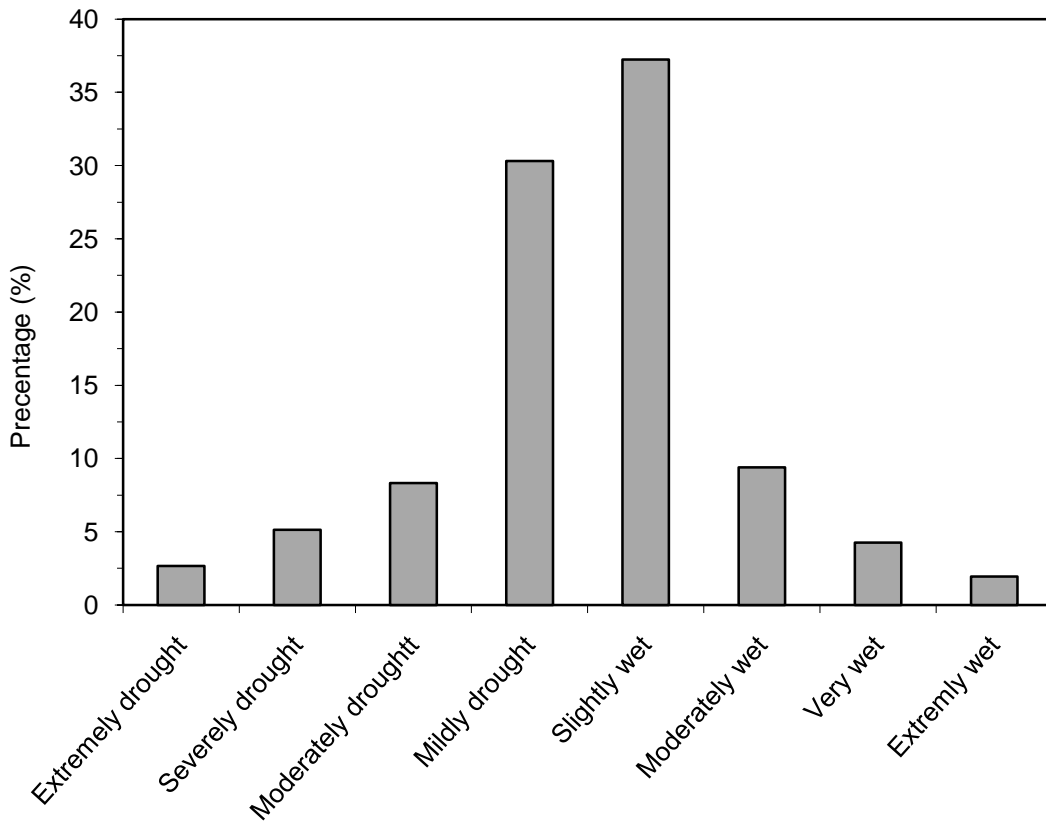
Figure 3.6: Percentage of dry and wet events based on one month SPI values
(Station Sorpetalsperre)

3.7.2 SPI Index of a Specified Month

The SPI values for the months January and April have been calculated based on 1 month time step as shown in figure 3.7 and figure 3.8. SPI values based on 3 months time step (quarter of a hydrological year) as shown in figure 3.19 through figure 3.12. Results show that drought occurred in both summer and winter although there is a significant increase in the winter precipitation as reported by Morgenschweis (Morgenschweis et al., 2007) (figure 2.26). The most extremely drought event in the basin was during winter as shown in figures 3.9 and 3.13. Figures 3.13 and 3.14 present the time series of SPI data values based on 6 months time step (winter and summer of a hydrological year). It is clear from the two figures that several severely and extremely drought events occurred in the Ruhr basin and the drought event in the winter of the hydrological year 1995-1996 was the most extremely event.
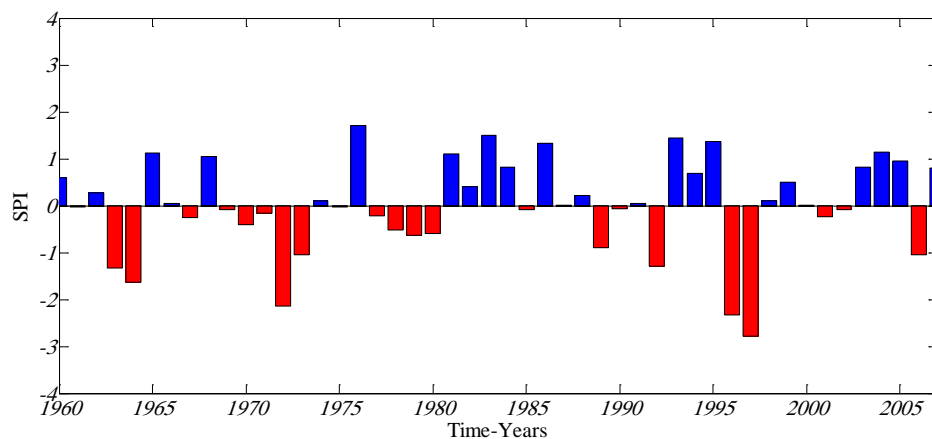
Figure 3.7: Drought severity index values representative of the Ruhr River Basin based on one month SPI values (SPI_1_January) (Station Sorpetalsperre)
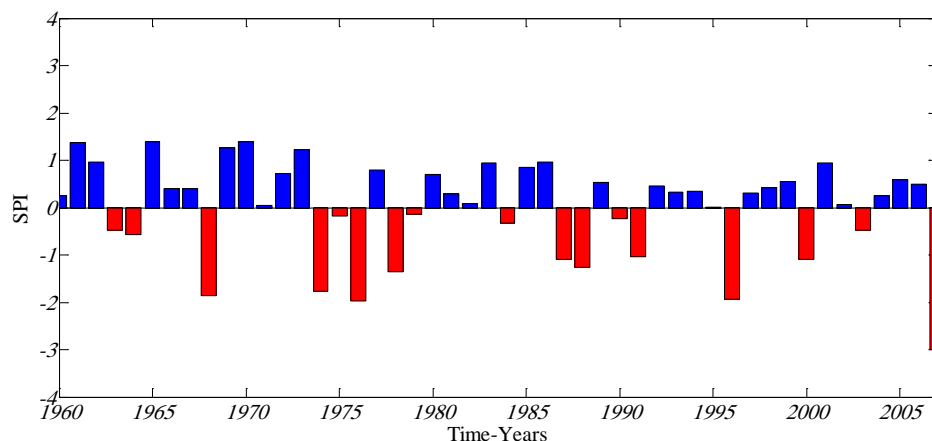
Figure 3.8: Drought severity index values representative of the Ruhr River Basin based on one month SPI values (SPI_1_April) (Station Sorpetalsperre)
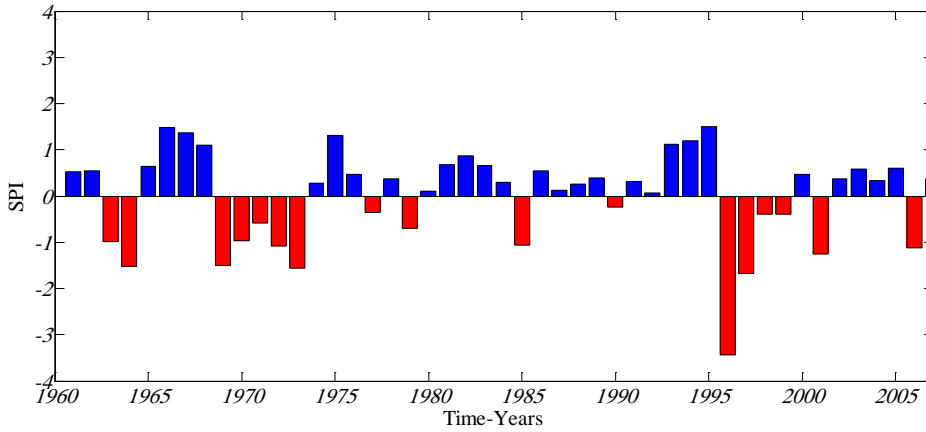
Figure 3.9: Drought severity index values representative of the Ruhr River Basin based on SPI values of Nov., Dec. and Jan. (SPI_3_Jan) (Station Sorpetalsperre)
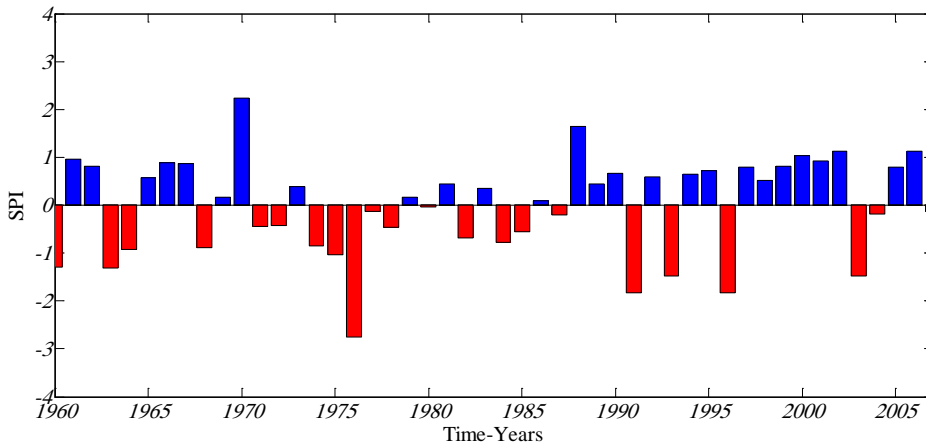
Figure 3.10: Drought severity index values representative of the Ruhr River Basin based on SPI values of Feb., Mar. and Apr. (SPI_3_Apr) (Station Sorpetalsperre)
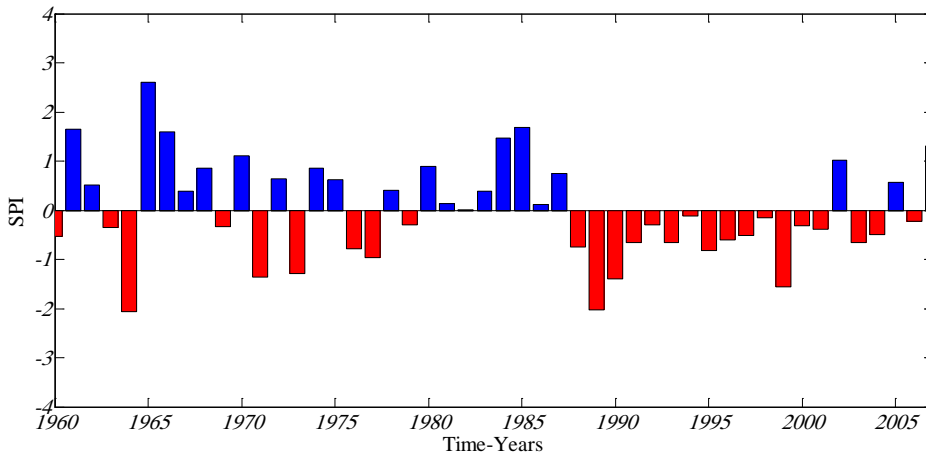
Figure 3.11: Drought severity index values representative of the Ruhr River Basin based on SPI values of May, Jun. and Jul. (SPI_3_Jul) (Station Sorpetalsperre)
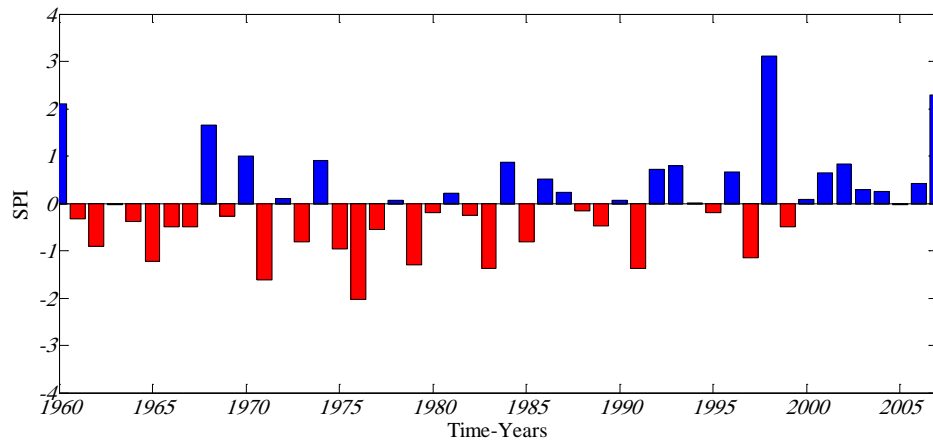
Figure 3.12: Drought severity index values representative of the Ruhr River Basin based on SPI values of Aug., Sep. and Oct. (SPI_3_Oct) (Station Sorpetalsperre)
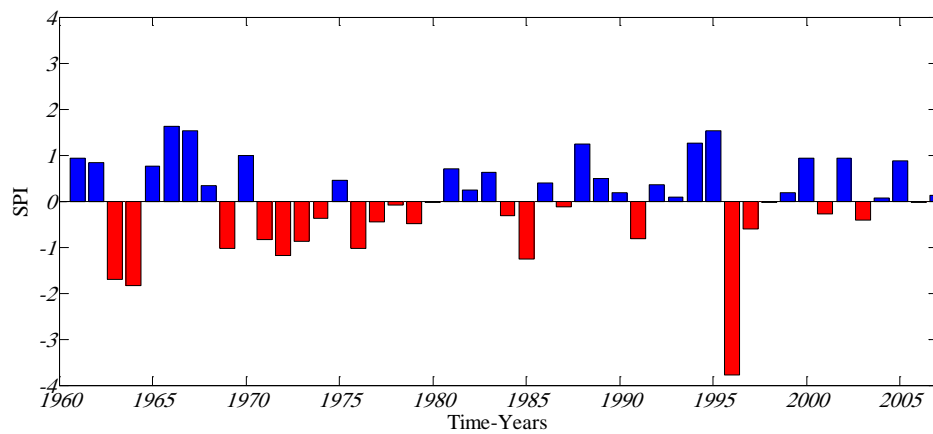
Figure 3.13 : Drought severity index values representative of the Ruhr River Basin based on SPI values of Nov., Dec., Jan., Feb., Mar. and Apr. (SPI_6_April) (Winter) (Station Sorpetalsperre)
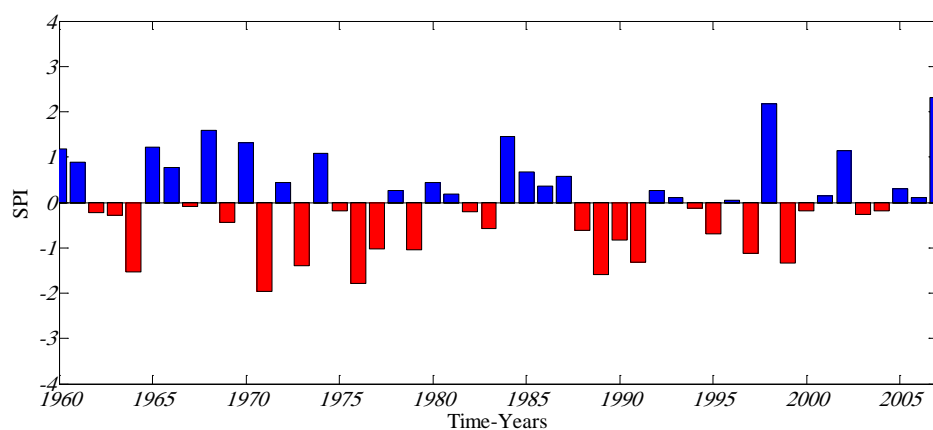
Figure 3.14: Drought severity index values representative of the Ruhr River Basin based on SPI values of May, Jun., Jul., Aug., Sep. and Oct. (SPI_6_Oct) (Summer) (Station Sorpetalsperre)

3.7.3 Probability of Drought Occurrence in the Ruhr Basin

The occurrence in varying drought categories at 3, 6, 9, 12-month and 2-year time steps has been analyzed. The aim was to identify drought events at comparable time steps based on their occurrence frequencies. Figure 3.15 (a through d) presents   percentages of drought occurrence expressed at multiple-time steps for varying drought severity categories. Each percentage is obtained by taking the ratio of drought occurrence in each time step to the total drought occurrence in the same time step and drought category.

(a)

(b)

(c)

(d)
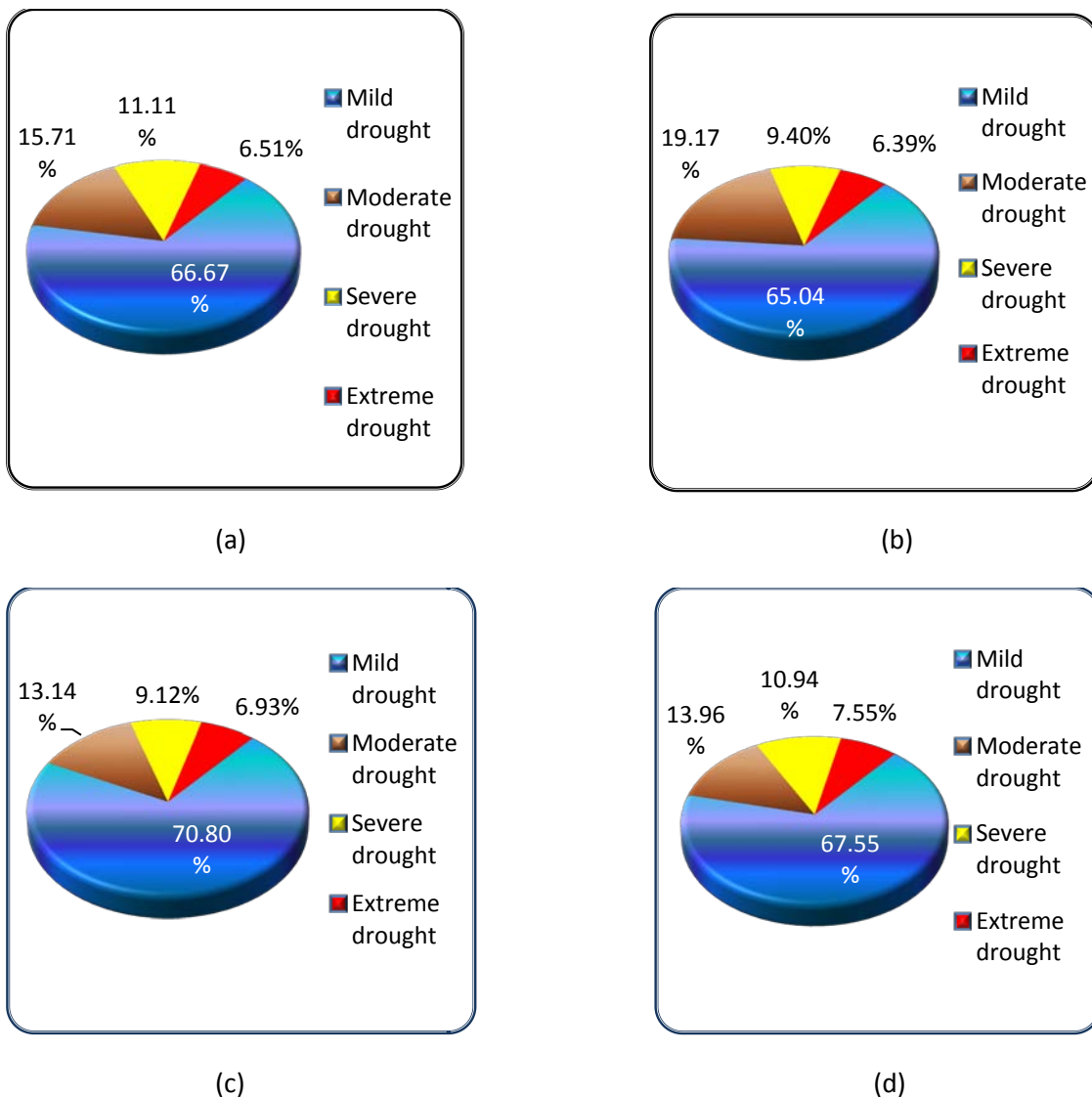
Figure 3.15: Drought occurrence in the Ruhr basin (Station Sorpetalsperre) at different drought categories and time steps:

(a) SPI_3   (b) SPI_6   (c) SPI_9   (d) SPI_12

3.7.4 Trend of SPI Index

To examine the trends in the SPI data series the Mann-Kendall test has been applied to SPI_3, SPI_6, SPI_9, SPI_12, SPI_3_Jan, SPI_3_Apr, SPI_3_Jul, SPI_3_Oct, SPI_6_Apr and SPI_6_Oct. Table 3.4 summarizes the results of the Mann-Kendall test. It shows that there, is a positive significant trend in the data series of SPI_3 for the months April and October and insignificant trend in SPI_3 for the month July. The significant positive trend in the SPI_3 for months April and October can be explained as due to the decrease of the number of drought events. This can be clearly noticed from figures 3.10 and 3.12 that from 1986 to 2007 the number of drought events is small compare with wet events. Conversely, for the month July (negative trend near 80 % confidence level) during the same period the number of wet events is small compared with drought events

Table 3.4: Results of Mann-Kendall test (1961-2007)

SPI Category	T-value	Type of trend
SPI_3	1.0024	Insignificant
SPI_6	0.899	Insignificant
SPI_9	0.876	Insignificant
SPI_12	1.057	Insignificant
SPI_3_Jan	-0.165	Insignificant
SPI_3_Apr	1.67	Significant at confidence level of 90 %
SPI_3_Jul	-1.27	Insignificant
SPI_3_Oct	2.05	Significant at confidence level of 95 %
SPI_6_Apr	0.59	Insignificant
SPI_6_Oct	-0.22	Insignificant

3.7.5 Number of Drought Events

Many drought events which occurred in the Ruhr river basin have been detected. Table 3.5 table and 3.6 present the severely and the extremely drought events respectively based on 1, 3, 6, 9, 12 and 24 months of SPI values. The two tables show the advantage of using several time steps when applying the SPI approach. For example if the SPI values are calculated based on one month time step, the detected event might be a drought event which cannot be detected if the SPI is calculated based on 3 months time step and vice versa. A practical example for this is shown in table 3.6, when SPI_1 has been applied; the drought event which occurred in April 2007, which was a very dry month, has been detected. But with SPI_3 this even has not been detected. Also as shown in table 3.6 there was an extremely drought event in the winter of the hydrological year 1996 (SPI_3 for month February), this event has been detected by using SPI based on 3 months time step and did not appear in the results of SPI based on one month time step. The results in tables 3.5 and 3.6 show that the Ruhr river basin received severely and extremely drought events in the period 1960-2007.

3.8 Conclusion

The aim of the chapter was to assess the overall meteorological drought vulnerability in the Ruhr basin by reconstructing historical occurrences of drought at several time steps and drought categories with the SPI approach. By applying the SPI approach, the obtained results indicated that the drought randomly affect the Ruhr river basin. Several drought events occurred during the period under study. Results also indicate that inspite of a significant positive trend in winter precipitation drought visited the Ruhr basin in both summer and winter and that the most severe event was in the winter. Trends in SPI data series indicated that the proportion of the Ruhr catchment drought condition has changed insignificantly during the period under study.

Results and the conclusion reached in this study can be an essential step toward addressing the issue to drought vulnerability in the Ruhr river basin and will be used as a guide for water resources management in the Ruhr river basin during droughts.

At the end it is worth to be mentioned that in reality extreme drought events in the last decades presented no severe challenges to the water supply of the Ruhr district due to the reservoir system existing in the Ruhr catchment basin.

Table 3.5: Severe drought events according to several time steps Severely Drought Events (Station Sorpetalsperre)

Severe drought events														
SPI_1			SPI_3			SPI_6			SPI_9			SPI_12		
Year	Month	Value	Year	Month	Value	Year	Month	Value	Year	Month	Value	Year	Month	
1962	10	-1.58	1964	1	-1.52	1963	3	-1.84	1963	4	-1.94	1963	5	-1.61
1962	11	-1.98	1964	3	-1.65	1963	4	-1.69	1963	6	-1.71	1963	7	-1.76
1963	2	-1.64	1964	5	-1.91	1963	6	-1.69	1964	4	-1.55	1963	8	-1.61
1964	1	-1.62	1964	6	-1.72	1964	3	-1.50	1964	5	-1.98	1963	9	-1.53
1964	7	-1.53	1964	8	-1.61	1964	4	-1.85	1964	6	-1.99	1963	12	-1.56
1968	4	-1.86	1969	1	-1.51	1964	9	-1.93	1964	10	-1.76	1964	1	-1.56
1969	9	-1.85	1971	10	-1.63	1964	10	-1.54	1964	11	-1.91	1964	6	-1.92
1971	9	-1.63	1972	12	-1.54	1971	9	-1.53	1964	12	-1.52	1964	12	-1.98
1972	12	-1.54	1973	1	-1.56	1971	10	-1.96	1971	9	-1.51	1971	10	-1.77
1973	6	-1.97	1973	2	-1.56	1972	3	-1.78	1971	10	-1.80	1971	11	-1.55
1973	8	-1.61	1973	8	-1.98	1973	2	-1.53	1971	11	-1.60	1971	12	-1.72
1974	4	-1.78	1973	9	-1.54	1973	3	-1.68	1971	12	-1.74	1972	5	-1.80
1975	2	-1.57	1975	12	-1.66	1976	10	-1.78	1972	4	-1.86	1972	6	-1.79

SPI_1			SPI_3			SPI_6			SPI_9			SPI_12		
Year	Month	Value	Year	Month	Value	Year	Month	Value	Year	Month	Value	Year	Month	Value
1975	12	-1.66	1976	5	-1.86	1976	11	-1.55	1972	5	-1.64	1973	8	-1.89
1976	4	-1.98	1976	6	-1.98	1977	1	-1.58	1973	8	-1.76	1973	9	-1.95
1976	8	-1.77	1976	8	-1.61	1983	12	-1.63	1973	9	-1.54	1976	7	-1.53
1980	5	-1.72	1982	9	-1.78	1989	9	-1.56	1976	6	-1.51	1976	8	-1.68
1982	7	-1.88	1983	9	-1.64	1989	10	-1.59	1977	1	-1.61	1976	9	-1.87
1982	9	-1.69	1985	2	-1.53	1991	7	-1.65	1977	2	-1.58	1976	10	-1.81
1985	10	-1.53	1989	6	-1.85	1991	8	-1.58	1977	3	-1.55	1976	11	-1.58
1989	6	-1.65	1989	8	-1.58	1991	9	-1.53	1989	11	-1.51	1976	12	-1.52
1989	11	-1.50	1990	5	-1.57	1993	8	-1.49	1991	9	-1.81	1977	5	-1.65
1990	5	-1.58	1991	4	-1.83	1995	12	-1.79	1995	12	-1.77	1977	7	-1.56
1990	7	-1.59	1991	5	-1.72	1996	7	-1.62	1996	9	-1.90	1992	1	-1.60
1992	5	-1.72	1996	4	-1.83	1997	11	-1.76	1998	2	-1.63	1996	2	-1.87
1995	10	-1.73	1996	6	-1.97	1997	12	-1.53				1996	11	-1.67
1996	3	-1.82	1997	1	-1.67							1996	12	-1.57
1997	11	-1.54	1999	7	-1.56							1997	1	-1.55

SPI_1			SPI_3			SPI_6			SPI_9			SPI_12		
Year	Month	Value	Year	Month	Value	Year	Month	Value	Year	Month	Value	Year	Month	Value
			2000	6	-1.54							1997	11	-1.56

Table 3.6: Extreme drought events according to several time steps (Station Sorpetalsperre)

Extremely Drought Events

SPI_1			SPI_3			SPI_6			SPI_9			SPI_12		
Year	Month	Value	Year	Month	Value	Year	Month	Value	Year	Month	Value	Year	Month	Value
1960	3	-2.03	1962	11	-2.20	1963	2	-2.06	1963	5	-2.21	1964	7	-2.16
1968	11	-2.18	1963	3	-2.01	1964	5	-2.64	1964	7	-2.40	1964	8	-2.39
1971	7	-2.63	1964	2	-2.08	1964	6	-2.32	1964	8	-2.81	1964	9	-2.23
1972	1	-2.14	1964	7	-2.06	1964	7	-2.01	1964	9	-2.34	1964	10	-2.21
1978	11	-2.01	1971	9	-2.04	1964	8	-2.22	1972	1	-2.25	1964	11	-2.53
1984	3	-2.02	1972	2	-2.83	1971	12	-2.02	1972	2	-2.45	1972	1	-2.10
1986	2	-2.17	1972	3	-2.26	1972	1	-2.03	1972	3	-2.72	1972	2	-2.49
1988	5	-2.10	1976	4	-2.76	1972	2	-2.45	1976	10	-2.78	1972	3	-2.35
1989	5	-2.58	1976	10	-2.03	1976	7	-2.21	1976	11	-2.12	1972	4	-2.16
1993	3	-2.36	1988	6	-2.07	1976	8	-2.19	1976	12	-2.08	1977	1	-2.45
1995	11	-2.15	1989	7	-2.02	1976	9	-2.17	1991	10	-2.02	1977	2	-2.00
1995	12	-3.20	1995	12	-3.20	1996	1	-2.50	1996	1	-2.30	1977	3	-2.06

SPI_1			SPI_3			SPI_6			SPI_9			SPI_12		
Year	Month	Value	Year	Month	Value	Year	Month	Value	Year	Month	Value	Year	Month	Value
1996	1	-2.33	1996	1	-3.43	1996	2	-2.17	1996	2	-2.39	1996	3	-2.41
1997	1	-2.78	1996	2	-2.32	1996	3	-3.42	1996	3	-2.56	1996	4	-2.86
1997	9	-2.32	1996	3	-2.33	1996	4	-3.78	1996	4	-3.03	1996	5	-2.98
2006	6	-2.11	1996	5	-2.29	1996	5	-3.02	1996	5	-2.90	1996	6	-3.00
2007	4	-2.99	1997	9	-2.30	1996	6	-2.96	1996	6	-3.64	1996	7	-2.72
									1996	7	-2.97	1996	8	-2.43
									1996	8	-2.23	1996	9	-2.82
												1996	10	-2.07

Chapter 4

Meteorological Drought Forecasting Using Stochastic Models

4.1 Theoretical Basis of Time Series Analysis

4.1.1 Definition of Time Series

A Time Series is a sequence of observations taken sequentially in time (Box et al., 2008). Mostly these observations are collected at equally spaced and discrete time intervals. A single time series or more specifically a univariate time series is the time series that has only one variable upon which observations are made then. A basic assumption in any time series analysis or modeling is that some aspects of the past pattern will continue to remain in the future. Also under this set up, often the time series process is assumed to be based on past values of the main variable but not on explanatory variables which may affect the variable system. So the system acts as a black box and we may only be able to know about 'what' will happen rather than 'why' it happens. Thus, if time series models are put to use, say, for instance for forecasting purposes, then they are especially applicable in the 'short term'. Here it is tacitly assumed that information about the past is available in the form of numerical data.

4.1.2 Missing Data

If some values are missing, they should be replaced by a theoretically defensible algorithm. If the time series have too much missing data, it may not be amenable to time series analysis (Yaffee and Magee, 2000). If the series does not have too much missing observations, it may be possible to perform some missing data analysis, estimation and replacement.

4.1.3 Sample Size

As a rule, the series should contain enough observations for proper parameter estimation (Yaffee and Magee, 2000). There seems to be no hard and fast rule about the minimum size. Some authors say that at least 30 observations are needed; others 50 and others indicate that there should be at least 60 observations. Ideally, at least 50 observations should be available for performing time series analysis as propounded by Box and Jenkins who were pioneers in time series modeling.

4.1.4 Stationarity

Time series may be stationary or non-stationary. A time series is said to be stationary when its statistical properties such as mean, variance, autocorrelation, etc. are all constant over time. Most statistical forecasting methods are based on the assumption that the time series can be rendered approximately stationary (i.e., "stationarized") through the use of mathematical transformations.

Non-stationary time series are characterized by random walk, drift, trend or changing variance. It should be prepared for statistical modeling; series are transformed to stationary either by taking the natural log or by taking a difference, or by taking residual from a regression. If the series can be transformed to stationarity by differencing, it is known as difference-stationary. If the series can be transformed to stationarity by detrending it, then we say that the series is trend-stationary.

4.2 The Nature and Use of Forecasts

4.2.1 Forecasting Definitions and Objectives

Forecasting can be defined as estimation of future trends by examining and analyzing available data. Making good forecasting is not always easy (Montgomery et al., 2008). Forecasting is an important problem that covers several fields including business, industry, government, economics, environmental science, medicine, social science, politics and finance. The importance of forecasting is well understood (Yaffee and Magee, 2000). Forecasting problems could be classified into three groups, short-term, medium-term and long-term forecasting. Short-term forecasting problems involve predicting events only a few time periods (days, weeks, months) into the future. Medium-term forecasts extend from one to two years into the future and long-term forecasts extend beyond that by many years. Short-term forecasting is usually applied to time series which do not change dramatically very quickly and the statistical methods are very useful in this case.

4.2.2 Basic Methodology of Forecasting

Despite the wide range of problems that require forecasts, there are two general types of forecasting techniques as reported in (Yaffee and Magee, 2000). The first is the qualitative method and the second is the quantitative method. Qualitative forecasts are often used in situations where there is little or no historical data on which to base the forecast. Quantitative forecasting methods use historical data and the forecasting model summarizes patterns in the data to express a statistical relationship between the previous and current values of the variable. Then the model is used to project the patterns in the data into the future. The most formal and widely known quantitative forecasting techniques are: forecasting based on historical data (naive methods- moving average, exponential smoothing, trend analysis and decomposition of time series) and associative forecasting (simple regression-multiple regression-econometric modeling)

4.3 Forecasting Using Stochastic Models

4.3.1 Definition of Stochastic Models

In probability theory, a stochastic process is the counterpart to a deterministic process or deterministic system. Instead of dealing with only one possible "reality" of how the process might evolve under time, in a stochastic or random process there is some indeterminacy in its future evolution described by probability distributions (Saglam, 2008). This means that even if the initial condition is known, there are many possibilities the process might go to, but some paths are more probable and others less.

A model that describes the probability structure of a sequence of observations is called a stochastic process (Box et al., 2008). Stochastic models, which are often known as time series models, have been used in scientific, economic and engineering applications for the analysis of time series. Time series modeling techniques have been shown to provide a systematic empirical method for simulating and forecasting the behavior of uncertain hydrologic systems and for quantifying the expected accuracy of the forecasts. In this study, linear stochastic models known as ARIMA and multiplicative Seasonal Auto Regressive Integrated Moving Average (SARIMA) models are used to forecast meteorological droughts.

4.4 Forecasting of the SPI Index Using ARIMA and SARIMA Models

4.4.1 Background Information on Drought Forecasting

The SPI has been developed for the purpose of defining and monitoring droughts. The global climate change in recent years is likely to enhance the frequency of droughts. While much of the weather that we experience is brief and short-lived, drought is a more gradual phenomenon, slowly affecting an area and tightening its grip with time. In severe cases, drought can last for many years, and can have devastating effects on agriculture and water supplies. It is very difficult to determine when a drought begins or ends. A drought can be short, lasting for just a few months, or it may persist for years before climatic conditions return to normal.

Drought forecasting plays an important role in the mitigation of impacts of drought on water resources systems. Traditionally, statistical models have been used for hydrologic drought forecasting based on time series methods (Kim and Valde´s, 2003). One of the basic deficiencies in mitigating the effects of drought is the inability to forecast drought conditions reasonably well in advance by either few months or seasons. Yevjevich as reported in (Dracup, 1991) was among the first at attempting a prediction of properties of droughts using the geometric probability distribution, defining a drought of k years as k consecutive years when there are no adequate water resources. Saldariaga (Saldariaga and Yevjevich, 1970) continued the development of run theory, incorporating concepts of time series analysis in formulations to predict drought occurrence.

Rao, and G. P. (Rao and Padmanabhan, 1984) investigated the stochastic nature of yearly and monthly Palmer's drought index (PDI) series and to characterize them via valid stochastic models which may be used to forecast and to simulate the PDI series. The monthly and annual PDI series were analyzed in their study. Sen (Sen, 1990) derived exact probability distribution functions of critical droughts in stationary second order Markov chains for finite sample lengths on the basis of the enumeration technique and predicted the possible critical drought durations that may result from any hydrologic phenomenon. Kendall (Kendall and Dracup, 1992) proposed a drought event generator using alternating renewal–reward model.

Moye (Moyé and Kapadia, 1994) developed a pertinent probability distribution based on difference equations to forecast drought of prespecified duration and average drought length of desired period. Loaiciga (Loaiciga and Leipnik, 1996) modeled the occurrence of drought events by the renewal processes. Lohani (Lohani and Loganathan, 1997) used PDSI in a non-homogenous Markov chain model to characterize the stochastic behavior of drought and based on these drought characterizations an early warning system was used for drought management. Chung (Chung and Salas, 2000) used low-order Discrete Auto Regressive Moving Average (DARMA) models for estimating the occurrence probabilities of drought events.

Kim (Kim and Valde´s, 2003) used PDSI as drought parameter to forecast drought in the Conchos River basin in Mexico using conjunction of dyadic wavelet transforms and neural network. There has been considerable research on modeling for various aspects of drought, such as the identification and prediction of its duration and severity.

It is rather easy to sense that a drought has set in, particularly during a cropping season. There is a need to develop methods and techniques to forecast the initiation/ termination point of droughts. The ARMA models, pattern recognition techniques, physically based models using Palmer drought severity index (PDSI), standardized precipitation index (SPI), a moisture adequacy index involving Markov chains, or the notion of conditional probability, seems to offer a potential to develop reliable and robust forecasts towards this goal (Panu and Sharma, 2002).

Such research efforts would be of considerable importance in mitigating the impacts of droughts. The stochastic models presented in this paper are based on SPI as drought index. The SPI is used in this study because of several reasons. The primary reason is that SPI is based on rainfall alone, so that drought assessment is possible even if other hydro-meteorological measurements are not available. The SPI is also not adversely affected by topography, it is defined over various timescales and this allows it to describe drought conditions over a range of meteorological, hydrological and agricultural applications.

4.4.2 ARIMA Model

4.4.2.1 Definition of ARIMA Model

Auto Regressive Integrated Moving Average (ARIMA) model was advanced by Box and Jenkins in 1960s for forecasting a variable (Box and Jenkins, 1970), hence this model is also known as Box-Jenkins model. Its appropriate use requires long time series data. Box and Jenkins introduced the concept of seasonal non-seasonal ARIMA models for describing a seasonal time series and also provided an iterative procedure for developing such models.

ARIMA models are, in theory, the most general class of models for forecasting a time series which can be stationarized by transformations such as differencing and logging. In fact, the easiest way to think of ARIMA models is as fine-tuned versions of random-walk and random-trend models. The fine-tuning consists of adding lags of the differenced series and/or lags of the forecast errors to the prediction equation, as needed to remove any last traces of autocorrelation from the forecast errors.

The acronym ARIMA stands for "Auto-Regressive Integrated Moving Average". Lags of the differenced series appearing in the forecasting equation are called "auto-regressive" terms, lags of the forecast errors are called "moving average" terms and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series (Ghafoor and Hanif, 2005). Random-walk and random-trend models, autoregressive models and exponential smoothing models (i.e., exponential weighted moving averages) are all special cases of ARIMA models.

4.4.2.2 Description of ARIMA Representation

In general, a non-seasonal ARIMA model is characterized by the notation ARIMA (p, d, q), where "p" is the number of autoregressive terms, "d" is the number of non-seasonal differences and "q" is the number of lagged forecast errors in the prediction equation.

In ARIMA parlance, TS is a linear function of past actual values and random shocks. For instance, given a time series process (Y_t), a first order auto-regressive process is denoted by ARIMA (1,0,0) or simply AR(1) and is given by:

$$Y_t = \mu + \varphi_1 * Y_{t-1} + \varepsilon_t \qquad\qquad (4.1)$$

where the auto regressive coefficient is denoted by φ "phi".

and a first order moving average process is denoted by ARIMA (0,0,1) or simply MA(1) and is given by:

$$Y_t = \mu - \theta_1 * \varepsilon_{t-1} + \varepsilon_t \qquad\qquad (4.2)$$

Where θ, the coefficient of the lagged forecast error, is denoted by the Greek letter "theta".

Alternatively, the model ultimately derived, may be a mixture of these processes and of higher orders as well. Thus a stationary ARMA (p, d, q) process is defined by the equation:

$$\varphi_p(B)\nabla^d Y_t \ = \ \theta_q(B)\varepsilon_t \qquad\qquad (4.3)$$

Where $\varepsilon_{t's}$ are independently and normally distributed with zero mean and constant variance σ^2 for t = 1, 2,…..n.

4.4.2.3 Description of Seasonal ARIMA Representation

Identification of relevant models and inclusion of suitable seasonal variables are necessary for seasonal. The Seasonal ARIMA i.e. ARIMA (p, d, q) (P, D, Q)$_s$ model, as reported in (Shumway and Stoffer, 2000), is defined by:

$$\varphi_p(B)\ \emptyset_P(B_s)\ \nabla^d \nabla_s^D\ \ Y_t = \Theta_Q(B_s)\ \theta_q(B)\ \varepsilon_t \qquad\qquad (4.4)$$

Where:

$$\varphi_p(B) = 1 - \varphi_1 B-....-\varphi_p B^p \qquad\qquad (4.5)$$

$$\theta_q(B) = 1-\theta_1 B-...-\theta_q B^q \qquad\qquad (4.6)$$

$$\emptyset_P(B_s) = 1-\emptyset_1 B^s-...-\emptyset_P B^{sP}\ , \qquad\qquad (4.7)$$

$$\Theta_Q(B^s) = 1-\Theta_1 B^s-...-\Theta_Q B^{sQ} \qquad\qquad (4.8)$$

Where B is the backshift operator (i.e. $B\ y_t= y_{t-1}$, $B^2\ y_t = y_{t-2}$ and so on),'s' is the seasonal lag and 'ε_t' is the sequence of independent normal error variables with mean 0 and variance σ^2. $\emptyset$ and φ are respectively the seasonal and non-seasonal auto regressive parameters. Θ and θ are respectively seasonal and non-seasonal moving average parameters. p and q are orders of non-seasonal auto regression and moving average parameters respectively whereas P and Q are that of the seasonal auto regression and moving average parameters respectively. Also d and D denote non-seasonal and seasonal differences respectively.

4.4.2.4 The Art of ARIMA Model Building

(i) Identification

Identification of the general form of a univariate model involves two steps. Within the **first step** the data series is analyzed for stationarity and normality (Brocklebank and Dickey, 2003). There are two kinds of stationarity, viz., stationarity in 'mean' and stationarity in 'variance'. A cursory look at the graph of the data and structure of autocorrelation and partial correlation coefficients may provide clues for the presence of stationarity. Appropriate differencing of the series is performed (if necessary) to achieve stationarity and normality. Stationarity in variance could be achieved by some modes of transformation, say, log transformation. This is applicable for both seasonal and non-seasonal stationarity. Thus, if 'Y_t' denotes the original series, the non-seasonal difference of first order is:

$$X_t = Y_t - Y_{t-1} \qquad\qquad (4.9)$$

Followed by the seasonal differencing (if needed)

$$Z_t = X_t - X_{t-s} = (Y_t - Y_{t-1}) - (Y_{t-s} - Y_{t-s-1}) \qquad\qquad (4.10)$$

In the **second step** the temporal correlation structure of the transformed data is identified by examining its autocorrelation (ACF) and partial autocorrelation (PACF) functions (Box and Jenkins 1976) and to find the initial values for the orders of non-seasonal and seasonal parameters, *p, q*, and *P, Q*.

They could be obtained by looking for significant autocorrelation and partial auto-correlation coefficients. If second order autocorrelation coefficient is significant, then an AR (2), or MA (2) or ARMA (2) model could be tried to start with. This is not a hard and fast rule, as sample autocorrelation coefficients are   poor estimates of population autocorrelation coefficients. Still they can be used as initial values while the final models are achieved after going through the stages repeatedly. Usually up to order 2 for p, d, or q are sufficient for developing a good model in practice.

(ii) Estimation

At the identification stage several models are tentatively chosen that seem to provide statistically adequate representations of the available data. Then precise estimates of parameters of the model are to be obtained by least squares as advocated by Box and Jenkins. Standard computer packages like SAS, SPSS are available for finding the estimates of relevant parameters using iterative procedures.

(iii) Diagnostics

Different models can be obtained for various combinations of AR and MA individually and collectively (Khattree and Rao, 2003). The best model is obtained with following diagnostics:

(a) Low Akaike Information Criteria (*AIC*)/ Bayesian Information Criteria (BIC)/ Schwarz-Bayesian Information Criteria (*SBC*)

AIC is given by:

$$AIC = (-2 \log L + 2 \ m) \tag{4.12}$$

Where $m = p + q + P + Q$ and L is the likelihood function. Since -2 *logL* is approximately equal to $\{n \ (1+log \ 2\pi) + n \ log \ \sigma^2\}$ where σ^2 is the model *MSE*, and *AIC* can be written as $AIC=\{n \ (1+log \ 2\pi) + n \ log \ \sigma^2 + 2 \ m\}$and because the first term in this equation is a constant, it is usually omitted while comparing between models. As an alternative to *AIC*, sometimes *SBC* is also used which is given by:

$$SBC = \log \ \sigma^2 + (m \ log \ n) \ /n. \tag{4.13}$$

(b) Plot of residual's ACF

Once the appropriate ARIMA model has been fitted, one can examine the goodness of fit. If the fitted model is adequate, the residuals should be approximately white noise. So, we should check if the residuals have zero mean and if they are uncorrelated. The key instruments are the time plot, the ACF and the PACF of the residuals. The theoretical ACF and PACF of white noise processes take value zero for lags $J \neq 0$, so if the model is appropriate most of the coefficients of the sample ACF and PACF should be close to zero. In practice, we require that about the 95 % of these coefficients should fall within the non-significance bounds.

(c) Tests for Residual Normality

Any graph suitable for displaying the distribution of a set of data is suitable for judging the normality of the distribution of a group of residuals. The most common types are; Histogram and Normal probability plots.

The histogram is a frequency plot obtained by placing the data in regularly spaced cells and plotting each cell frequency versus the center of the cell. The histogram is not be the best choice for judging the distribution of residuals If the sample sizes of residuals are generally small (<50). The normal probability plot should produce an approximately straight line if the points come from a normal distribution. Small departures from the straight line in the normal probability plot are common, but a clearly "S" shaped curve on this graph suggests a bimodal distribution of residuals. Breaks near the middle of this graph are also indications of abnormalities in the residual distribution. Figure 4.1 illustrate the steps o the ARIMA model.
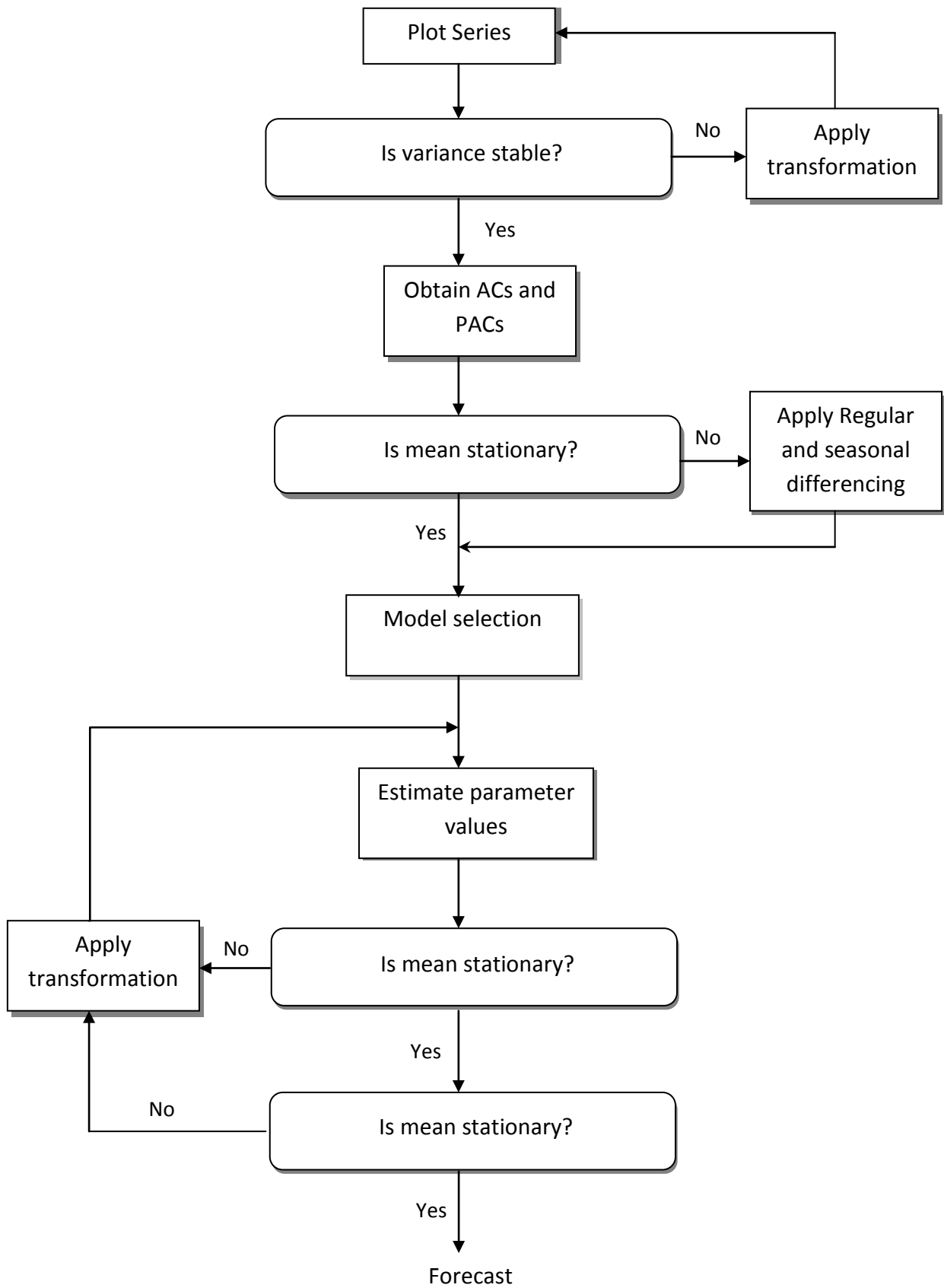
Figure 4.1: Box-Jenkins modeling approach
after (Box and Jenkins, 1970)

4.4.3   Development of an ARIMA Model to fit the SPI _3 Time Series

4.4.3.1 Computation of the Standardized Precipitation Index SPI_3

The frequency of drought events was calculated using the Standardized Precipitation Index (SPI). Figure 4.2 shows a sample of calculated SPI_3 time series. All details about the methodology and the results were presented in chapter 3
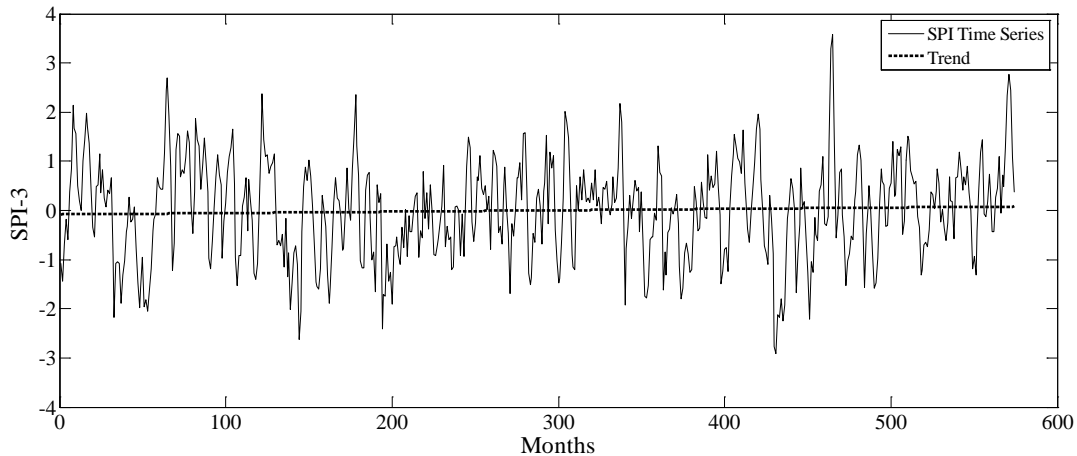
Figure 4.2: A sample of SPI time series

Time series model development consists of three stages identification, estimation, and diagnostic checking (Box and Jenkins, 1970). The identification stage involves transforming the data (if necessary) to improve the normality and the stationarity of the time series and determining the general form of the model to be estimated. During the estimation stage the model parameters are calculated. Finally, diagnostic checks of the model are performed to reveal possible model inadequacies and to assist in selecting the best model.

The data set from 1961 to 2007 were used for model development for SPI_3, SPI_6, and SPI_9, SPI_12 and SPI_24 series. The models have been developed for SPI_3, SPI 6, SPI 9, SPI 12, and SPI 24. For illustration, two examples are described briefly for SPI_3 and SPI_6. Details about SPI_9, SPI_12 and SPI_24 are presented in Appendices. The model identified for SPI_3 is ARIMA model, and for SPI_6, SPI_9, SPI_12 and SPI_24 is SARIMA model.

4.4.3.2 Model Identification

The ACF and PACF have been estimated for SPI_3, as shown in figure 4.3, and figure 4.4 respectively. The ACF and PACF show the series is stationary. The ACF is damping out in sine-wave manner with significant spikes at the first two lags. The first four values are significant in PACF, which indicates the process can be modeled as a combination of both AR and MA processes. Alternative ARIMA models have been identified by considering the ACF and PACF graphs of the SPI series. This indicates a possible ARIMA ($p, 0, q$) model with $p = 1$–4 and $q = 1$–3. All the combination have been tried to determine the best model out of these candidate models. The model with the minimum Akaike Information Criterion (*AIC*) and minimum Schwarz Bayesian Criterion (*SBC*) was selected as best fit model. Usually the model with the smallest AIC has residuals, which resemble white noise (Makridakis and Wheelwright, 1978). Table 4.1 presents a Comparison of AIC and SBC for the selected candidate models. It is clearly from table 4.1 that the model ARIMA(3,0,2) is the one with min *AIC*, but the two models ARIMA(1,0,3) and ARIMA(3,0,2) are examined to compare between the results of the two models.
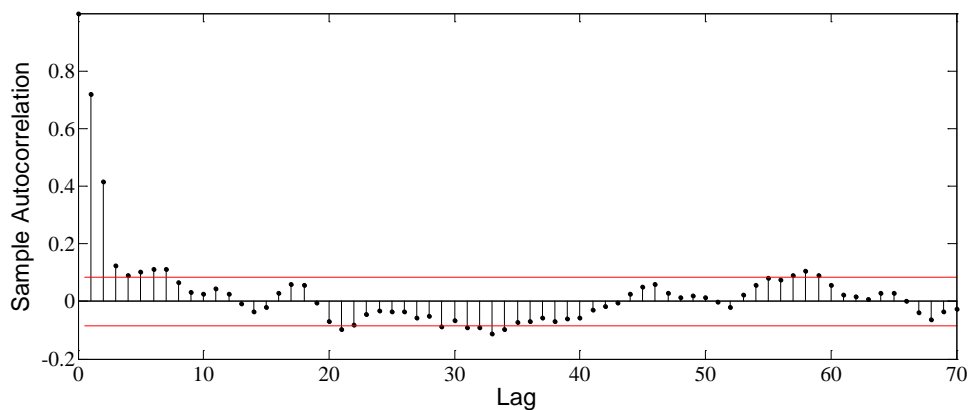
Figure 4.3: ACF plot used for the selection of candidate models for SPI_3 series
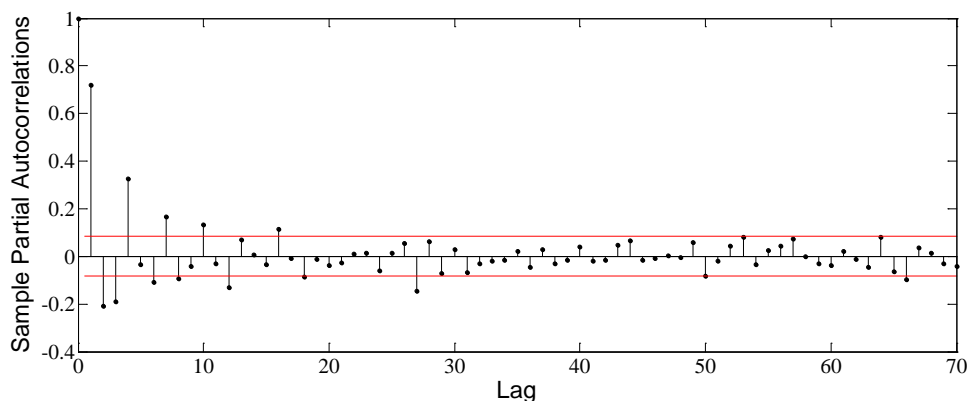
Figure 4.4: PACF plot used for the selection of candidate models for SPI_3 series

Table 4.1: Comparison of AIC and SBC for the selected candidate models

SPI series	Model	AIC	SBC
	ARIMA(1,0,2)	1056.32	1073.73
	ARIMA(1,0,3)	1056.11	1077.88
	ARIMA(3,0,1)	1144.70	1166.47
	ARIMA(3,0,2)	1055.48	1081.59
	ARIMA(3,0,2)	1058.20	1088.67
	ARIMA(4,0,0)	1110.23	1131.99
SPI-3	ARIMA(4,0,1)	1111.87	1137.98
	ARIMA(4,0,2)	1057.04	1087.52
	ARIMA(4,0,3)	1058.93	1093.75
	ARIMA(4,1,2)	1150.93	1181.39
	ARIMA(5,0,0)	1111.58	1137.70
	ARIMA(5,0,1)	1112.72	1143.19
	ARIMA(5,0,2)	1058.96	1093.79

4.4.3.3 Parameters Estimation

After the identification of the model using the AIC and SBC criteria, estimation of parameters is done. During the estimation stage, model estimates were calculated simultaneously for AR and MA parameters. Model estimates were made using the procedure outlined by Box and Jenkins (Box and Jenkins, 1970). Preliminary estimates of the parameters were computed from the ACF of the series developed in the identification stage. These preliminary estimates were then used as the starting values in an optimization algorithm for nonlinear least squares that minimize the residual sum of squares. The values of the parameters are shown in table 4.2. Model parameters have been calculated using a licensed software package of the statistical program SPSS.

Table 4.2: Statistical parameters of ARIMA (1,0,3), and ARIMA (3,0,2)

	Auto regressive parameters	Moving average parameters	Residual sum of squares	Residual variance
ARIMA (1,0,3)	0.8271	- 0.1069 - 0.1418 0.6135	207.94	0.3638
ARIMA (3,0,2)	0.07437 - 0.0258 0.1217	- 0.8832 - 0.8548	206.98	0.3627

4.4.3.4 Diagnostic Check

As mentioned before two models have been selected to compare between their results, namely ARIMA (1, 0, 3) and ARIMA (3, 0, 2). The models have been identified and the parameters have been estimated, then the model verification is concerned with checking the residuals of the model to see if they contain any systematic pattern which still can be removed to improve on the chosen ARIMA. For a good forecasting model, the residuals left over after fitting the model should be white noise. This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. For this purpose, the various correlations up to 70 lags have been computed. Also the histogram and the normal probability plot of the residuals have been drawn to check if the residual came from normal distribution or not.

The Ljung–Box test, which is commonly used in auto regressive integrated moving average (ARIMA) modeling, has been applied to the residuals of the fitted ARIMA models. The Ljung–Box test is a type of statistical test of whether any of a group of autocorrelations of a time series is different from zero. Instead of testing randomness at each distinct lag, it tests the "overall" randomness based on a number of lags, and is therefore a portmanteau test.

     I.   ARIMA (1, 0, 3)

(RACF) the residual ACF function and (RPACF) the residual PACF function should be calculated to determine whether residuals are white noise. If some of the RACF or some of the RPACF are significantly different from zero, this may indicate that the present model is inadequate. The ACF and PACF of residuals of the model ARIMA (1,0,3) are shown in figure 4.5 and 4.6 respectively. As shown in figures 4.5 and 4.6, most of the values of the RACF and RPACF are within confidence limits except very few individual correlations appear large compared with the confidence limits, which is expected among 70 lags.

The figures indicate no significant correlation between residuals. Histogram of residuals for SPI_3 is shown in Figure 4.7. The histogram shows that the residuals are normally distributed. This signifies residuals to be white noise. The graph of the cumulative distribution for the residual data normally appears as a straight line when plotted on normal probability paper as shown in figure 4.8 (Chow et al. 1988). The figure shows that, the normal probability plot of the residuals look fairly linear thus the normality assumptions of the residuals hold (Durbin 1960).

All results of the Ljung–Box test indicated a failure to reject the null hypothesis that a series of residuals exhibits no autocorrelation.
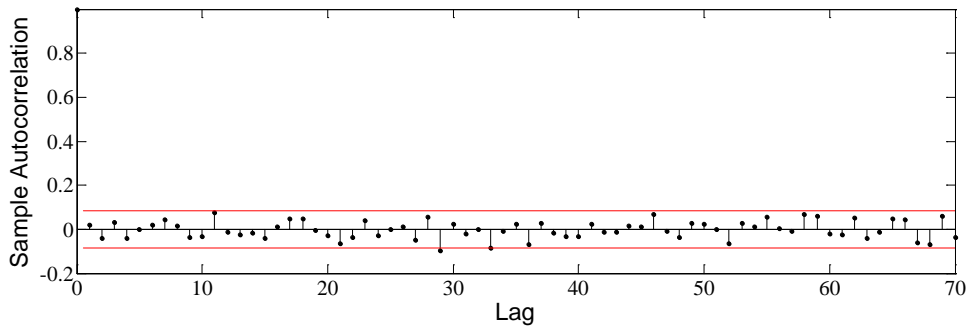
Figure 4.5: ACF plot used for diagnostic check of the model ARIMA (1, 0, 3)
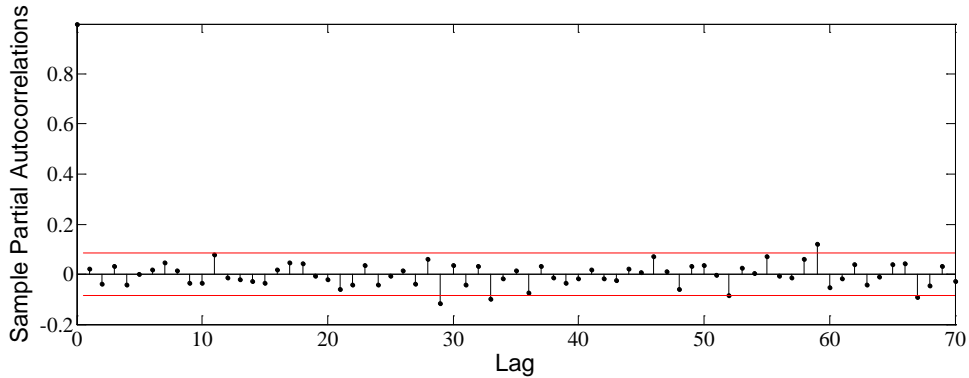
Figure 4.6: PACF plot used for Diagnostic Check of the model ARIMA (1, 0, 3)
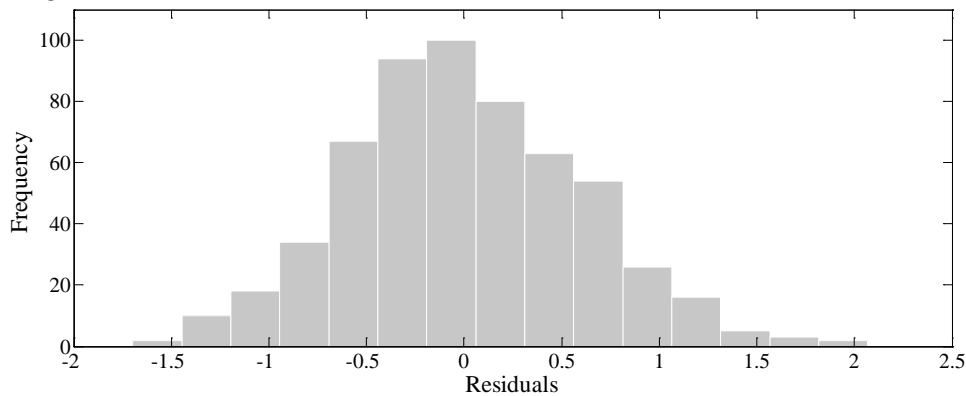
Figure 4.7: Histogram of the residuals – ARIMA (1, 0, 3)
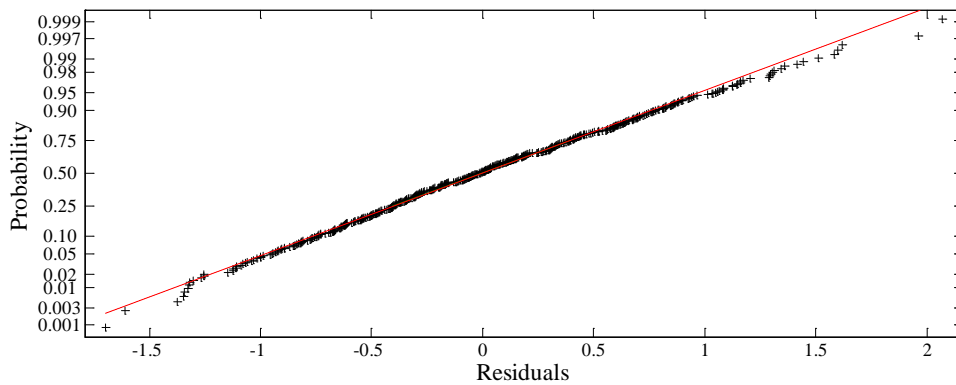
Figure 4.8: Normal probability plot of the residuals- ARIMA (1, 0, 3)

II.   ARIMA (3, 0, 2)

(RACF) the residual ACF function and (RPACF) the residual PACF function of the model ARIMA (3,0,2) are shown figure 4.9 and 4.10 respectively. There is no big difference between the plots of the model ARIMA (1,0,3) and the model ARIMA (3,0,2). Also as shown in figures 4.9 and 4.10, most of the values of the RACF and RPACF are within confidence limits except very few individual correlations appear large compared with the confidence limits. The figures indicate no significant correlation between residuals. Histogram of residuals for SPI_3 in figure 4.11 shows that the residuals are normally distributed. This signifies residuals to be white noise. The graph of the cumulative distribution for the residual data normally appears as a straight line as shown in figure 4.12.
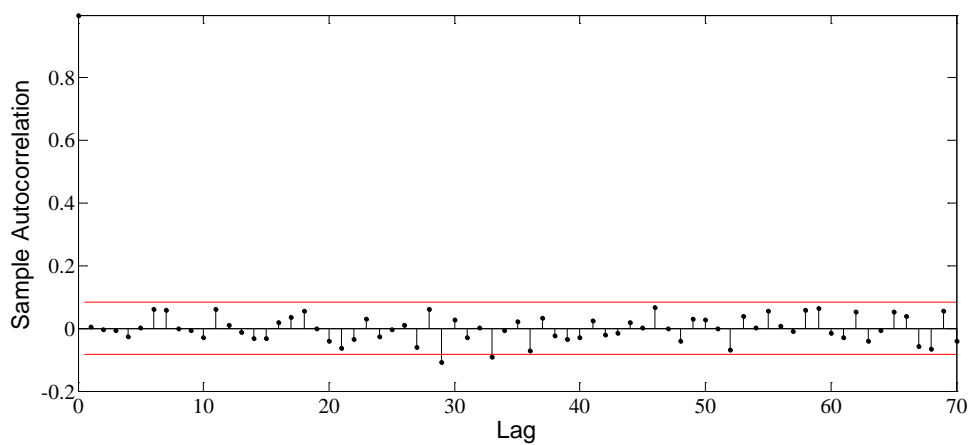
Figure 4.9: ACF plot used for Diagnostic Check of the model ARIMA (3, 0, 2)
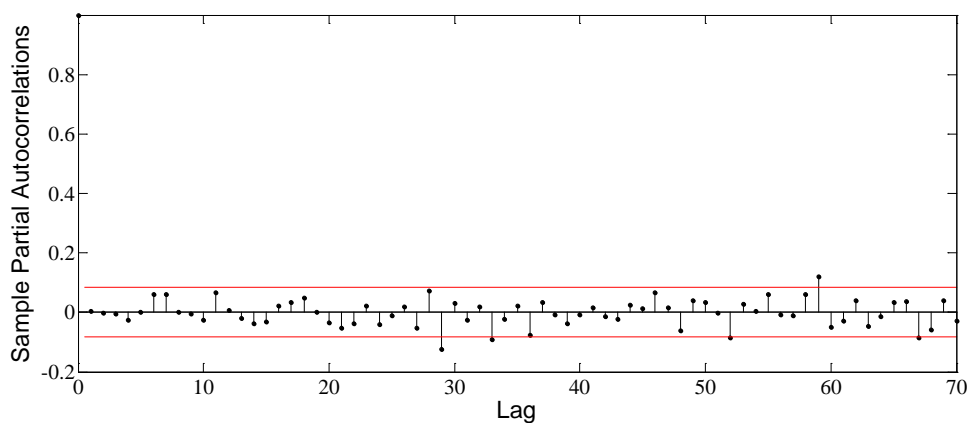
Figure 4.10: PACF plot used for Diagnostic Check of the model ARIMA (3, 0, 2)
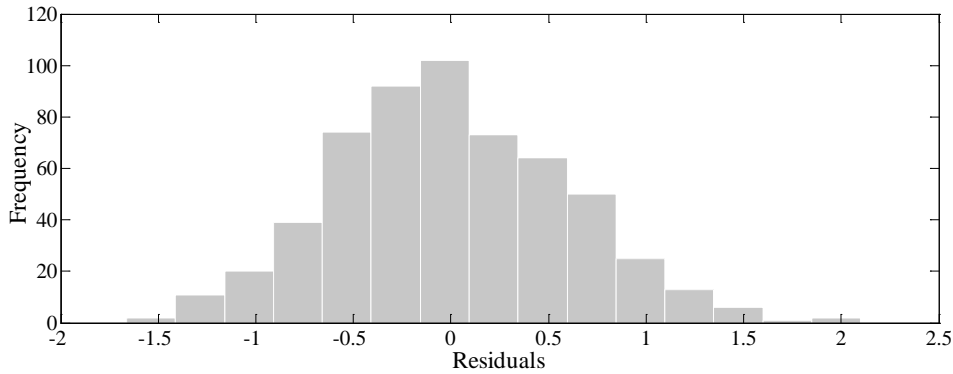
Figure 4.11: Histogram of the residuals – ARIMA (3, 0, 2)
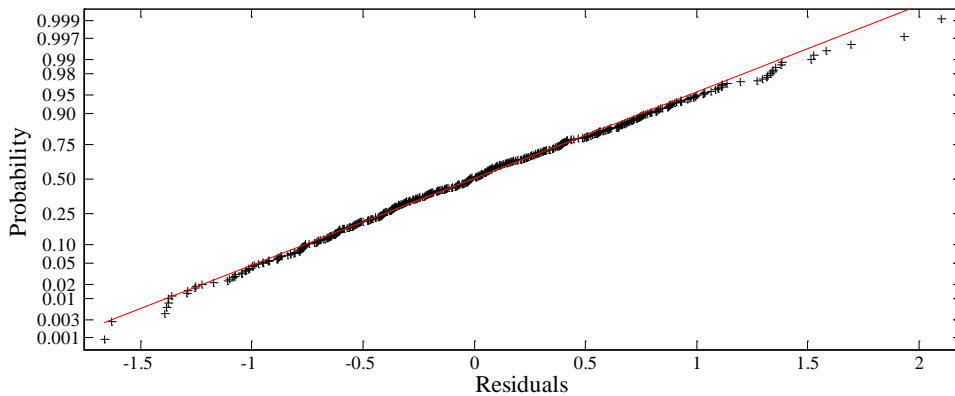
Figure 4.12: Normal probability plot of the residuals- ARIMA (3, 0, 2)

4.4.3.5 Drought Forecasting From Selected Models

ARIMA models are developed basically to forecast the corresponding variable. There are two groups of forecasts, namely the sample period forecasts and post-sample period forecasts. The first group is used to develop confidence in the model and the second to generate genuine forecasts for use in planning and other purposes. The ARIMA model can be used to yield both groups of forecasts. The forecast has been done for 1-month lead-time using the best models from historical data. Results of forecasting of the ARIMA model (1, 0, 3) are shown in figures 4.13, and figure 4.14. Figure 4.14 is a zoom window for the months from 550 to the end of the SPI time series, and this zoom window was taken from figure 4.13. Results of forecasting of the ARIMA model (3, 0, 2) are shown in figures 4.15, and figure 4.16. Figure 4.16 is a zoom window for the months from 550 to the end of the SPI time series. It can be clearly observed that the forecasted values of the SPI follow the calculated values closely. To evaluate the model, basic statistical properties have been compared between observed and forecasted data. The results, as shown in table 4.3, show that predicted values preserve the basic statistical properties of the observed series.

Table 4.3: Statistical properties of ARIMA (1,0,3), and ARIMA (3,0,2) Results

Model	Mean of the Calculated SPI	Mean of the forecasted SPI	Standard deviation of the Calculated SPI	Standard deviation of the forecasted SPI	RMSE
ARIMA (1,0,3)	-2.6132e-005	-1.9711e-004	1.0009	0.8056	0.6023
ARIMA (3,0,2)	-2.6132e-005	-3.3537e-005	1.0009	0.7996	0.6009
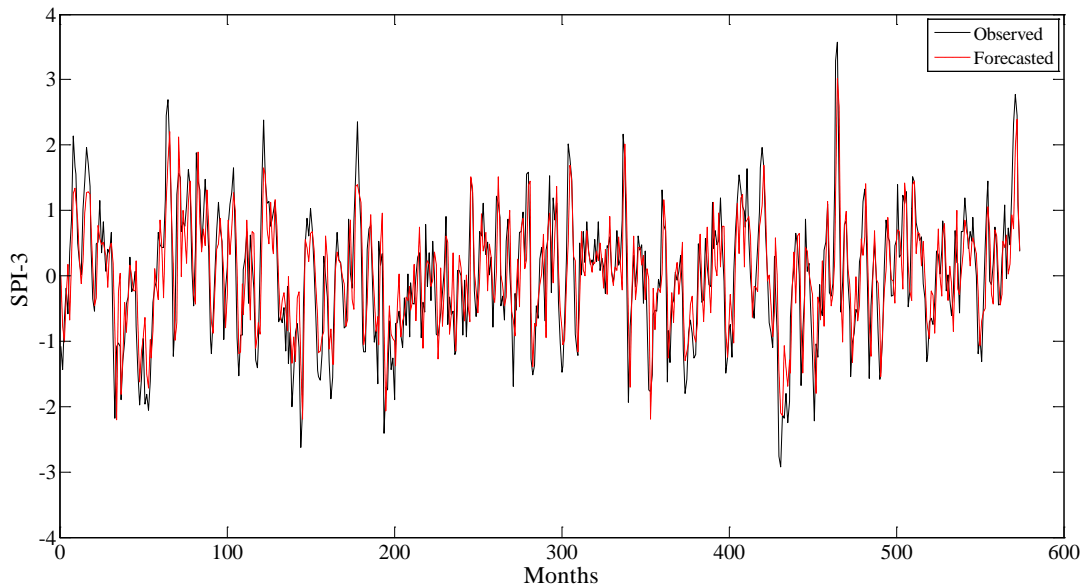
Figure 4.13: Comparison of calculated SPI with forecasted SPI using ARIMA (1, 0, 3)
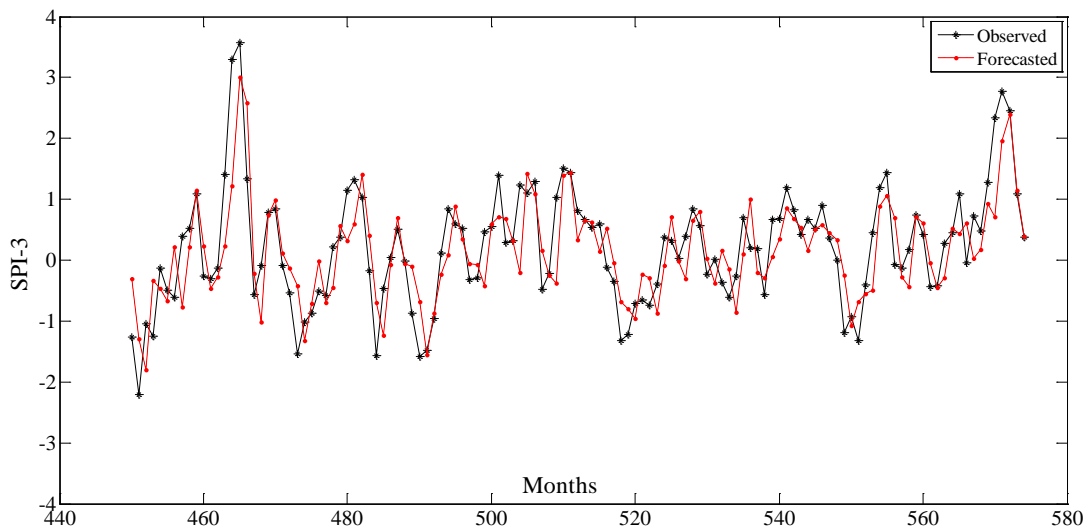
Figure 4.14: Comparison of calculated SPI with forecasted SPI using ARIMA (1, 0, 3) (From month 550 to the end of the time series)
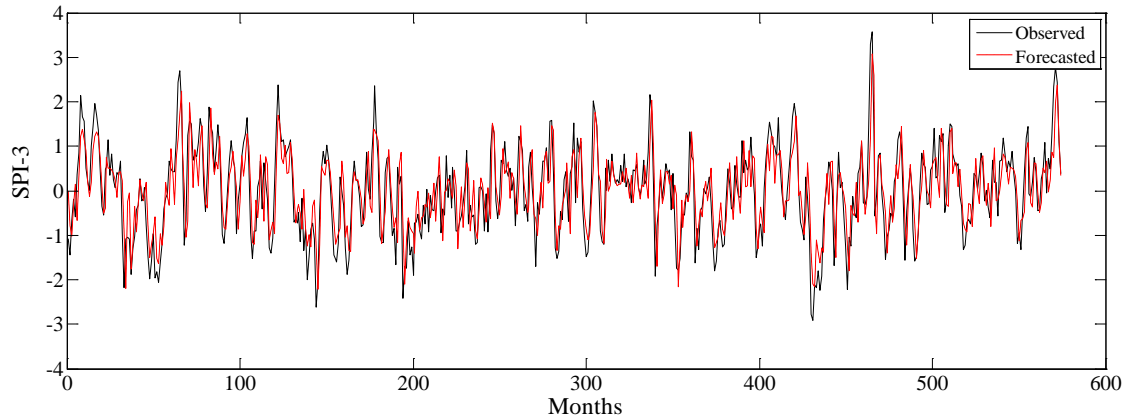
Figure 4.15: Comparison of calculated SPI with forecasted SPI using ARIMA (3, 0, 2)
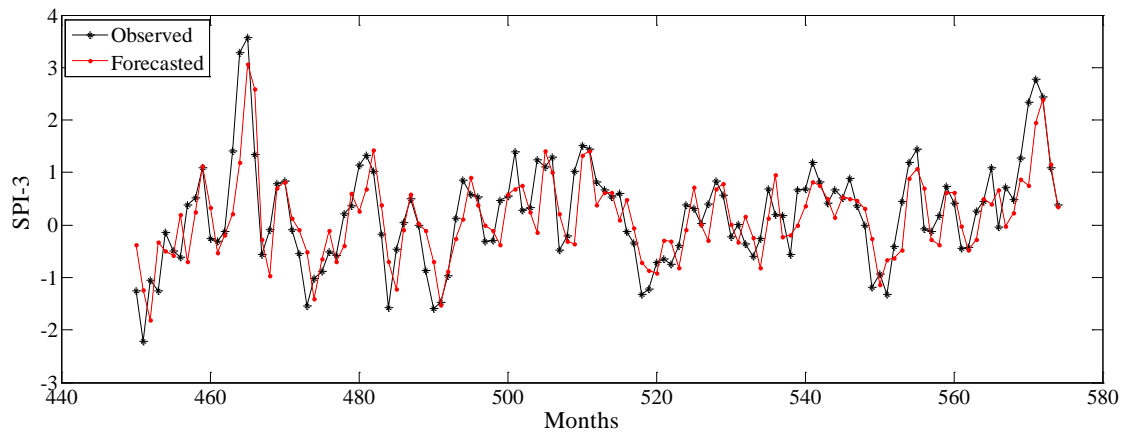
Figure 4.16: Comparison of calculated SPI with forecasted SPI using ARIMA (3, 0, 2)
(From month 550 to the end of the time series)

Table 4.4 presents a comparison of calculated SPI with forecasted SPI for both ARIMA (1,0,3) and ARIMA(3,0,2). Because of the long length of the SPI time series, only the last 21 months have been selected as shown in table 4.4. For illustration, for example in ARIMA (1,0,3), the SPI_3 in the second column is the calculated SPI for the total period. The second column was used to estimate the statistical parameters of the model (see table 4.3) then the forecasted values have been calculated which are in the third column. In the fourth column the new SPI time series has been selected. This time series is the same as the one which is in the second column except the last three values. This was done to forecast these three values to compare between the results. In this case new statistical parameters were estimated, then the forecasted values have been calculated as shown in fifth column in the table. It is clear from the fourth and the fifth columns that there is no big difference between the calculated and the forecasted values of the last three months. Comparison of the two models ARIMA (1,0,3) and ARIMA (3,0,2) demonstrated that there is no clear difference between the results of the two models.

As shown in table 4.4, the last three values in the 5th column which have been forecasted based on the data in the 4th column are very closely to the observed data in the second column. It is worth mentioning that this is not a special feature for the last three values and this has been confirmed by applying the ARIMA model to different lengths of the SPI time series.

Table 4.4: Comparison of calculated SPI with forecasted SPI

| | ARIMA(1,0,3) | | | | ARIMA(3,0,2) | | | |
	SPI_3	Forecasted values	SPI_3	Forecasted values	SPI_3	Forecasted values	SPI_3	Forecasted values
554	1.187	0.8851	1.187		1.187	.89597	1.187	
555	1.442	1.0576	1.442		1.442	1.07650	1.442	
556	-.078	0.6938	-.078		-.078	.70360	-.078	
557	-.132	-0.2769	-.132		-.132	-.27598	-.132	
558	.181	-0.4388	.181		.181	-.37286	.181	
559	.740	0.7102	.740		.740	.61969	.740	
560	.424	0.6148	.424		.424	.61423	.424	
561	-.440	-0.0456	-.440		-.440	-.03075	-.440	
562	-.428	-0.4519	-.428		-.428	-.47829	-.428	
563	.269	-0.2899	.269		.269	-.27457	.269	
564	.455	0.5279	.455		.455	.50029	.455	
565	1.088	0.4330	1.088		1.088	.39878	1.088	
566	-.047	0.6170	-.047		-.047	.67127	-.047	
567	.726	0.0286	.726		.726	-.02137	.726	
568	.480	0.1797	.480		.480	.23417	.480	
569	1.283	0.9356	1.283		1.283	.86749	1.283	
570	2.331	0.7128	2.331		2.331	.74812	2.331	
571	2.776	1.9661	2.776		2.776	1.95123	2.776	
572	2.447	2.4000		2.3997	2.447	2.38402		2.38303
573	1.082	1.1514		1.1084	1.082	1.15480		1.09348
574	0.382	0.3971		0.4207	0.382	0.34435		0.35730

4.4.4   Development of an ARIMA Model to Fit the SPI _6 Time Series

4.4.4.1 Model Identification

Figure 4.17 shows the SPI_6 time series. The ACF and PACF have been estimated for SPI-6, as shown in figure 4.18, and figure 4.19 respectively. The ACF is damping out with mixture of sine and exponential curve. The first value is significant in PACF which indicates an AR (1) as non-seasonal part of the model. Also in the PACF, there are significant spikes presented near lag 6, 12 and 18 which indicates a SARIMA model.

Alternative SARIMA models were identified by considering the ACF and PACF graphs of the SPI series. This indicates a possible SARIMA $(p, d, q)(P,D,Q)_s$ models with p = 1:4 , d=0:1, q = 1:4, P=1:4, D=0:1 and Q=1:4. All the combinations were examined to determine the best model out of these candidate models. The model that gives the minimum Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC) is selected as best fit model. Table 4.5 presents a Comparison of AIC and SBC for the selected candidate models. It is clearly from table 4.5 that the model SARIMA (1, 0, 3)$(1,0,3)_6$ is the one with min AIC .
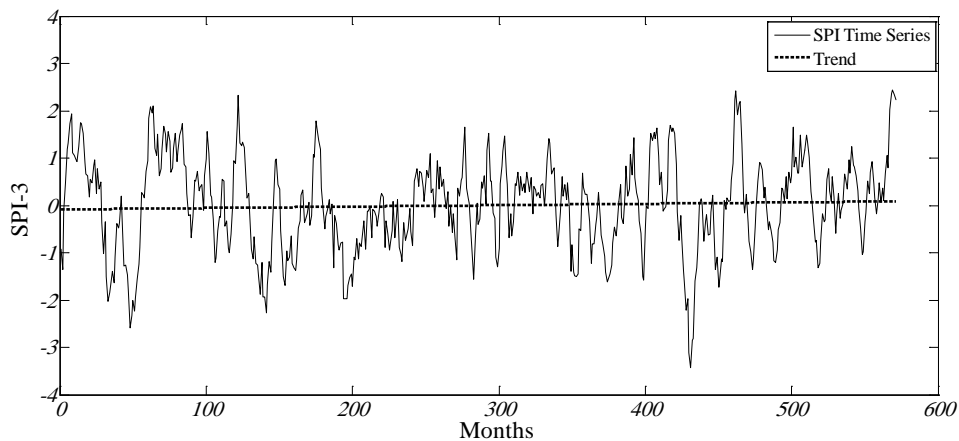
Figure 4.17: SPI_6 time series
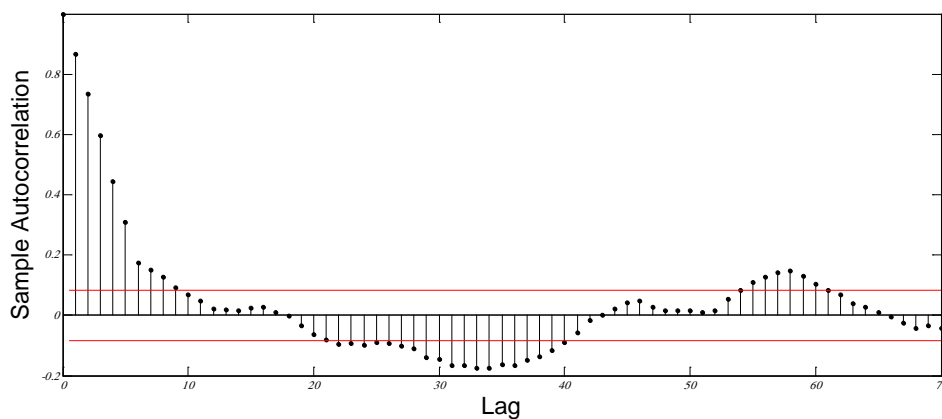
Figure 4.18: ACF plot used for the selection of candidate models for SPI_6 series
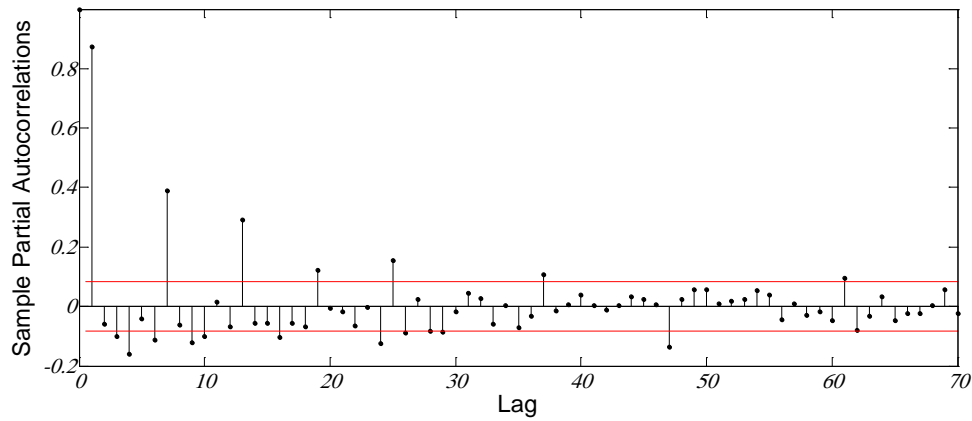
Figure 4.19: PACF plot used for the selection of candidate models for SPI_6 series

Table 4.5: Comparison of AIC and SBC for the selected candidate models

SPI series	Model	AIC	SBC
SPI-6	ARIMA(1,0,0)	812.51	821.20
	ARIMA(1,0,1)	812.82	825.86
	ARIMA(1,0,0) (1,0,0)$_6$	704.38	717.42
	ARIMA(1,0,0) (2,0,0)$_6$	645.53	662.92
	ARIMA(1,0,0) (3,0,0)$_6$	635.47	657.21
	ARIMA(1,0,0) (4,0,0)$_6$	621.32	647.40
	ARIMA(1,0,0) (1,0,1)$_6$	598.44	615.83
	ARIMA(1,0,0) (2,0,1)$_6$	600.33	622.06
	ARIMA(1,0,0) (3,0,1)$_6$	600.65	626.74
	ARIMA(1,0,0) (4,0,1)$_6$	601.37	631.80
	ARIMA(1,0,0) (1,1,1)$_6$	739.11	756.46
	ARIMA(1,0,0) (2,1,1)$_6$	680.64	702.32
	ARIMA(1,0,0) (3,1,1)$_6$	674.92	700.94
	ARIMA(1,0,0) (4,1,1)$_6$	657.58	787.94
	ARIMA(1,0,1) (1,1,1)$_6$	738.508	760.19
	ARIMA(1,0,3) (1,0,3)$_6$	581.33	620.46
	ARIMA(2,0,0) (0,0,0)$_6$	812.45	825.50
	ARIMA(2,0,0) (1,0,0)$_6$	705.56	722.95
	ARIMA(2,0,0) (1,0,1)$_6$	600.93	622.12
	ARIMA(2,0,1) (2,0,2)$_6$	593.34	628.11
	ARIMA(2,0,2) (2,0,2)$_6$	622.42	661.55

4.4.3.2 Parameters Estimation

After the identification of the model using the AIC and SBC criteria, estimation of parameters was done. During the estimation stage, the auto regressive and moving average parameters have been calculated simultaneously for the nonseasonal part of the model (AR and MA) and also for the seasonal part of the model (SAR and SMA) as well. The values of the parameters are shown in table 4.6.

Table 4.6: Statistical parameters of the model SARIMA $(1,0,3)(1,0,3)_6$

	Non-seasonal parameters	Seasonal parameters	Residual Sum of Squares	Residual Variance
SARIMA $(1,0,3)(1,0,3)_6$	AR 1 = 0.9533 MA1 = - 0.0495 MA2 = - 0.0227 MA3 = - 0.1879	SAR 1 = -0.9376 SMA1 = - 0.2314 SMA2 = 0.7747 SMA3 = 0.0114	89.66	0.155

4.4.4.3 Diagnostic Check

As mentioned before in table 4.5 the model SARIMA $(1, 0, 3)(1,0,3)_6$ has been selected as the one with min AIC. The model has been identified and the parameters have been estimated, the model verification is concerned with checking the residuals of the model. As mentioned before in the SPI_3 ARIMA models, for a good forecasting model, the residuals left over after fitting the model should be white noise. This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. For this purpose, the various correlations up to 70 lags have been computed. The histogram Also and the normal probability plot of the residuals have been drawn to check if the residual came from normal distribution or not.

The ACF and PACF of residuals of the model SARIMA $(1, 0, 3)(1,0,3)_6$ are shown in figure 4.20 and 4.21 respectively. As shown in figures 4.20 and 4.21, most of the values of the RACF and RPACF lies within confidence limits except very few individual correlations appear large compared with the confidence limits, which is expected among 70 lags. The figures indicate no significant correlation between residuals. Histogram of residuals for SPI_6 is shown in Figure 4.22. The histogram shows that the residuals are normally distributed. This signifies residuals to be white noise. The graph of the cumulative distribution for the residual data normally appears as a straight line as shown in figure 4.23. The figure show the normal probability plot of the residuals look fairly linear, the normality assumptions of the residuals hold (Durbin 1960).
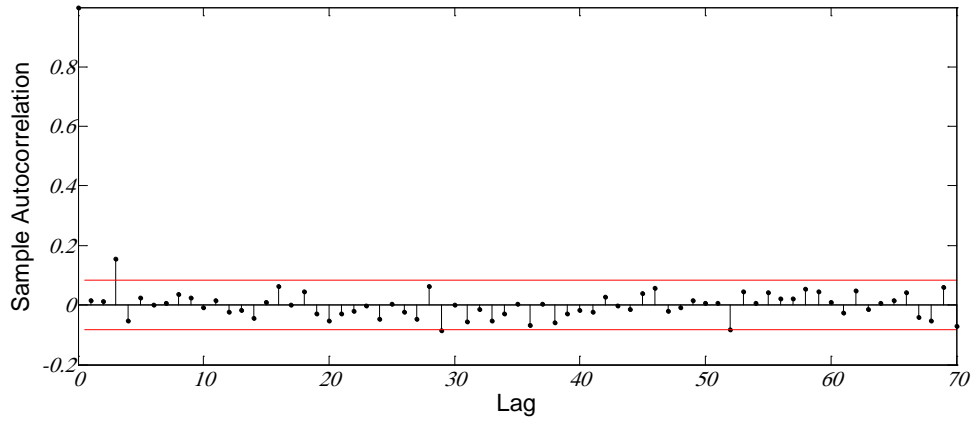
Figure 4.20: ACF plot used for Diagnostic Check of the model
SARIMA $(1, 0, 3)(1,0,3)_6$
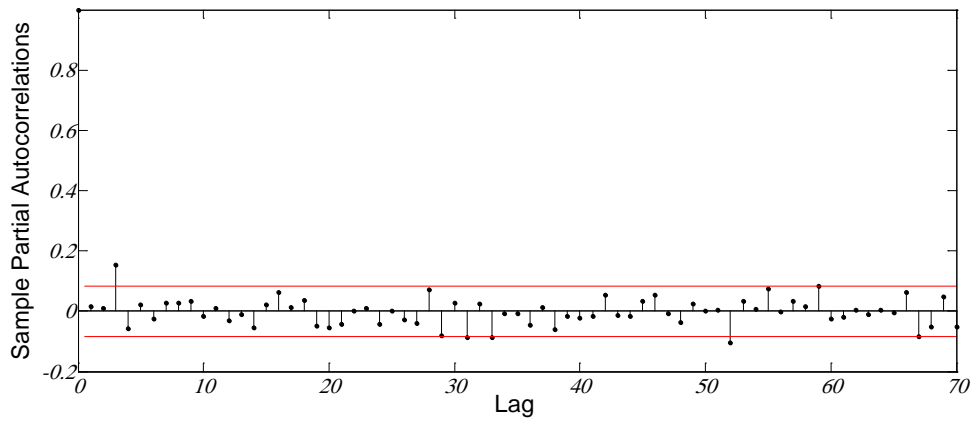
Figure 4.21: PACF plot used for Diagnostic Check of the model
SARIMA $(1, 0, 3)(1,0,3)_6$
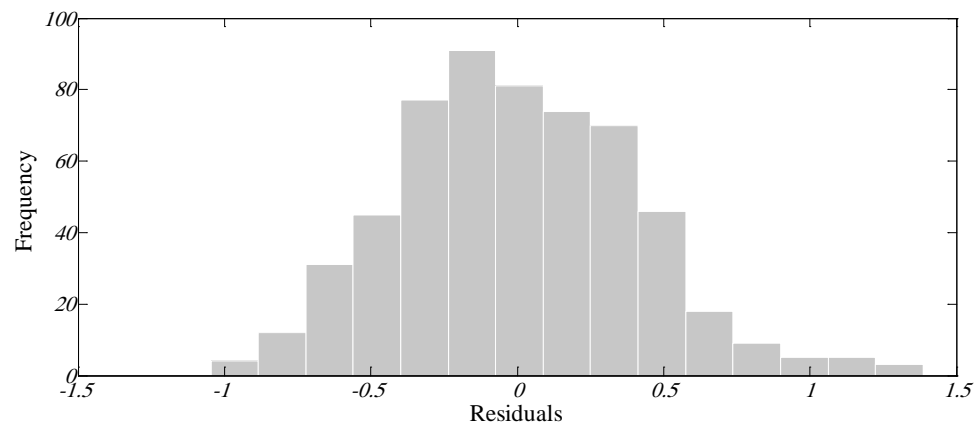
Figure 4.22: Histogram of the residuals – SARIMA $(1, 0, 3)(1,0,3)_6$
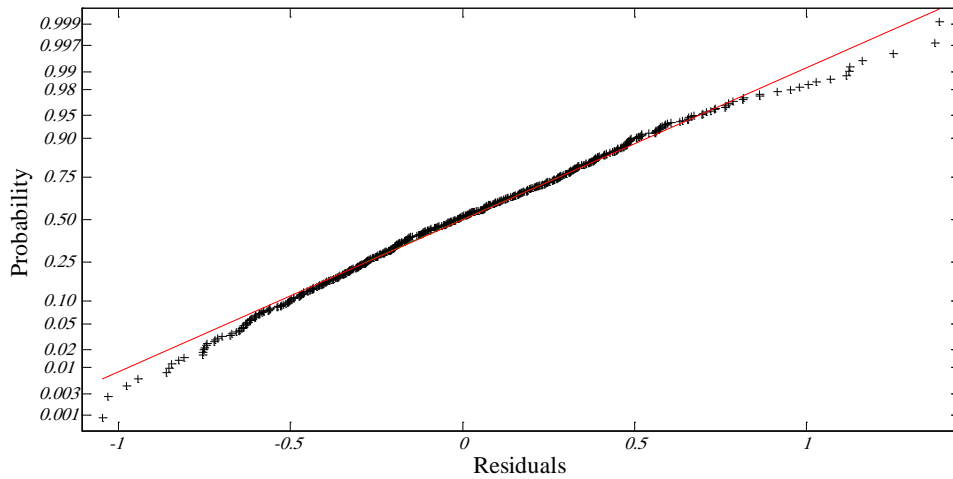
Figure 4.23: Normal probability plot of the residuals -SARIMA $(1, 0, 3)(1,0,3)_6$

4.4.4.4 Drought Forecasting with Selected Models

Results of forecasting of the SARIMA $(1, 0, 3)(1,0,3)_6$ are shown in figures 4.24 and figure 4.25. Figure 4.25 is a zoom window for the months from 550 to the end of the SPI time series, and this zoom window was taken from figure 4.24.

It is observed that, the forecasted values of the SPI follow the calculated values very closely. Basic statistical properties are compared between observed and forecasted data for one month ahead time. The results show that forecasted values preserves the basic statistical properties of the observed series (table 4.7).

Table 4.7: Statistical parameters of the model SARIMA $(1, 0, 3)(1,0,3)_6$

Model	Mean of the Calculated SPI	Mean of the forecasted SPI	Standard deviation of the calculated SPI	Standard deviation of the forecasted SPI	RMSE
SARIMA $(1, 0, 3)(1,0,3)_6$	1.9264e-004	-8.7054e-004	1.0010	0.9151	0.398
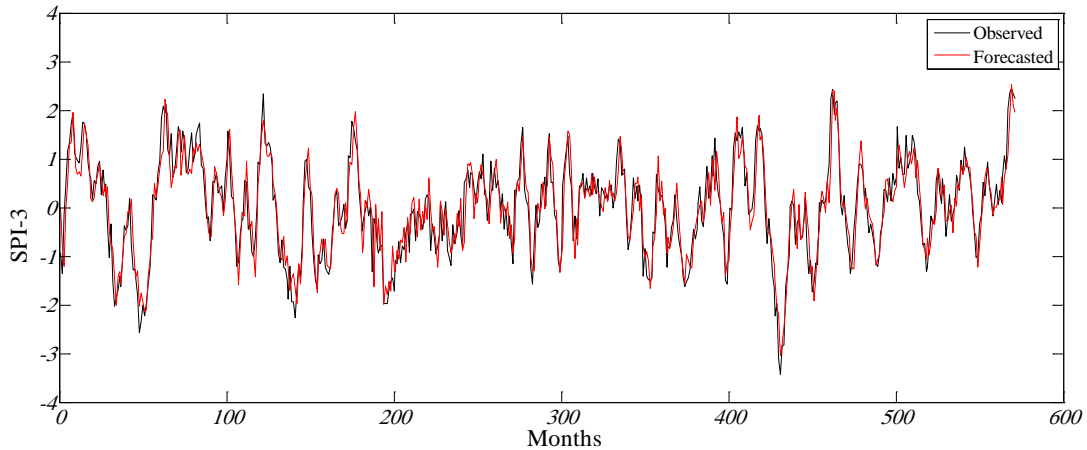
Figure 4.24: Comparison of calculated SPI with forecasted SPI using SARIMA $(1, 0, 3)(1,0,3)_6$
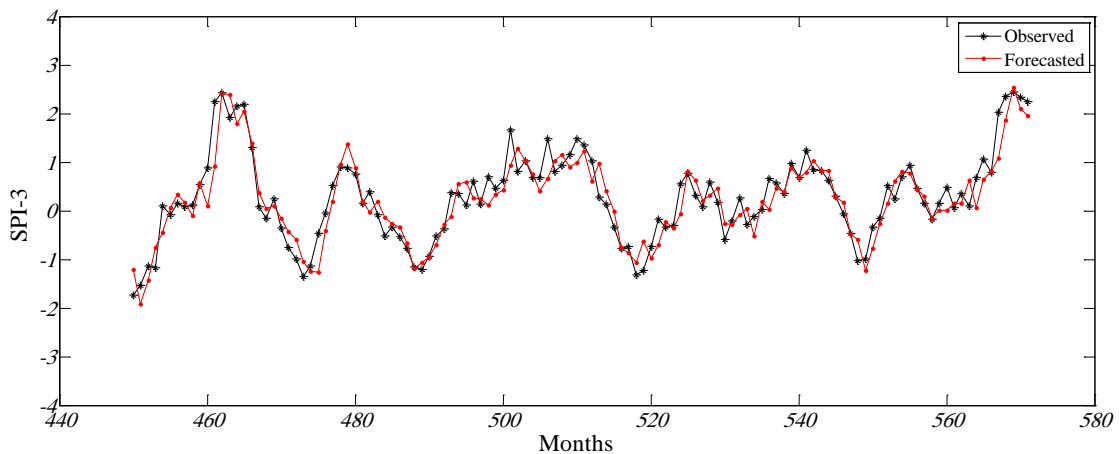
Figure 4.25: Comparison of calculated SPI with forecasted SPI using SARIMA $(1, 0, 3)(1,0,3)_6$ (From month 550 to the end of the time series)

Table 4.8 presents a comparison of calculated SPI_6 with forecasted SPI_6 for the model SARIMA $(1, 0, 3)(1,0,3)_6$. Because of the long length of the SPI time series, only the last 21 months were. The first column in the table is the number of months and the second column is the calculated SPI_6 for the total period. The second column was used to estimate the statistical parameters of the model (see table 4.7), then the forecasted values have been calculated as in third column. In the fourth column the new SPI_6 time series was selected. This time series is the same as the one which was used in the second column except the last three values. This means that the length of the new series is shorter than the original one by three months. This was done to forecast these three values to compare the results. In this case new statistical parameters were estimated then the forecasted values have been calculated as shown in fifth column in the table. It is clear from the fourth and the fifth columns that there is no big difference between the calculated and the forecasted values of the last three months.

Table 4.8: Comparison of calculated SPI with forecasted SPI

SARIMA $(1, 0, 3)(1,0,3)_6$				
	SPI_6	Forecasted values	SPI_6	Forecasted values
552	0.51	.16295	0.51	
553	0.25	.61353	0.25	
554	0.70	.80345	0.70	
555	0.93	.76410	0.93	
556	0.46	.44820	0.46	
557	0.16	.29139	0.16	
558	-0.18	-.15362	-0.18	
559	0.15	.00975	0.15	
560	0.48	.01814	0.48	
561	0.06	.14788	0.06	
562	0.36	.14961	0.36	
563	0.10	.62072	0.10	
564	0.69	.06699	0.69	
565	1.06	.63987	1.06	
566	0.80	.81163	0.80	
567	2.02	1.07914	2.02	
568	2.36	1.85954	2.36	
569	2.44	2.53403		2.54764
570	2.34	2.09435		2.18563
571	2.24	1.95113		1.78978

4.5 Conclusion

Drought monitoring and forecasting are essential tools for implementing appropriate mitigation measures in order to reduce negative impacts. Drought forecasting remains a difficult but vitally important task for hydrometeorologists and water resources managers. The availability of forecasts of drought indices, and of the related confidence intervals for a given site, could be a helpful tool to the decision making process for drought mitigation.

In this Study, the SPI index has been used as a drought indicator for drought forecasting due to its many advantages over other drought indices. This study has investigated the capability of ARIMA and SARIMA models in drought forecasting using the correlation methods of Box and Jenkins and the AIC and SBC structure selection criteria. Validation of the forecasting models has been carried out by comparing SPI values computed on observed precipitation and the corresponding forecasts. The results showed a fairly good agreement between observations and forecasts, as it has also been confirmed by the values of some performance indices. Evaluation of models showed that the results seem to be better for higher SPI series (SPI 6, SPI 9,..., and SPI 24) and this may be due to increase in filter length which reduces the noise more effectively. Finally, the good fitting of stochastic models such as ARIMA and SARIMA models to hydrologic time series such as SPI time series will result in a better tool that could be used for water resource planning within the basin.

Chapter 5

Stochastic Simulation of Monthly Streamflow

5.1 Introduction

Stochastic simulation of hydrologic time series has been widely used for solving various problems associated with the planning and management of water resources systems for several decades (Kim et al., 2004). Typical examples are the determination of a reservoir capacity, evaluations of adequacy and reliability of a reservoir for a given capacity, evaluation of adequacy of a water resource management strategy under various potential hydrologic scenario, and evaluation of the performance of an irrigation system under uncertain irrigation water distributions (Salas and Frevert). Stochastic simulation of hydrologic time series such as streamflow is typically based on mathematical models and a number of models have been suggested in (Singh and Frevert, 2001). Using one type of model or another for a particular case at hand depends on several factors such as, physical and statistical characteristics of the process under consideration, data availability, the complexity of the system and the overall purpose of the simulation study. Given the historical record, one would like the model to reproduce the historical statistics. This is why, a standard step in streamflow simulation studies is to determine the historical statistics. Once a model has been selected, the next step is to estimate the model parameters, then to test whether the model represents reasonably well the process under consideration and finally to carry out the needed simulation study (Singh and Frevert, 2001).

Time series of streamflow is an essential information for planning, design and operation of many water resources systems. However, in most instances, time series of flow records at the location of interest are limited. Therefore, the use of available historic streamflow may be insufficient for obtaining reliable estimate of flow statistics (Juran and Arup, 2007). In the event of non availability of a long series of historical streamflow record, generation of the data series is of utmost importance. Classical stochastic models, such as the Thomas-Fiering model (Altunkaynak et al., 2005; Phien and Ruksasilp, 1981), auto regressive moving average (ARMA) models (Box and Jenkins, 1970) are generally used for synthetic streamflow generation. Water resource planners must consider streamflow variability to provide effective long-term planning and management. Incorporating this variability has traditionally been achieved through generation of stochastic streamflow. Stochastic simulation of streamflow represents reasonable alternate streamflow comparable to observed data available in a river basin. These observed data are typically limited in time, limiting variability in the stochastic streamflow, particularly concerning the frequency of extremes (Prairie and Rajagopalan, 2007).

In analyzing streamflow and rainfall sequences many hydrologists regard it as a realization of a stochastic process (Ismail et al., 2004). The generated data sequences, particularly monthly time series such as streamflow or rainfall are widely used in water resources planning and management to understand the variability of future system performance. Stochastic data generation aimed at generating synthetic data sequences that are statistically similar to the observed data sequences. Therefore, the generated data is important for more accurate solutions of various complex planning, design and operational problems in water resources development.

Methods for design and operation of water supply reservoirs are usually deal with the basis of time intervals of one month which keeps the computing work within reasonable limits and appears to produce results of reasonable accuracy (Treiber and Schultz, 1976). By replacement of the observed time series by synthetically generated time series of a predetermined length, which are used for reservoir design, it became possible to make statistical statements on the reliability of the reservoir to fulfill a certain demand.

In this study two models have been used to generate monthly inflow time series, namely the Thomas-Fiering model and the Monte-Carlo simulation model. The following section presents some details about these models.

5.2. Description of Models

5.2.1 Thomas-Fiering Model

The first model that appeared in the hydrology literature for the generation of synthetic monthly flow sequences is that due to Thomas & Fiering in 1962 (Sen, 1978). Basically, this model is of a markovian nature with periodic parameters, namely, the monthly means, standard deviations and the lag-zero cross correlations between successive months. In its simplest form the model consists of twelve regression equations, one for each month. The method of Thomas and Fiering implicitly allows for the non-Stationarity observed in monthly inflow data (Singhal et al., 1980).

For the Thomas-Fiering model, synthetic monthly series is generated with the following recursive relationship:

$$Q_{i+1} = \bar{Q}_{j+1} + b_j \left(Q_i - \bar{Q}_j \right) + t_i * S_{j+1} * \left(1 - r_j^2 \right)^{1/2} \qquad (5.1)$$

Where:

Q_i = the inflow during the i month record from the start of the synthetic sequence.

Q_{i+1} = the inflow during the (i+1) month.

$\bar{Q}_j$ = the mean monthly inflow during the j month with a repetitive cycle of 12 months.

$\bar{Q}_{j+1}$ = the mean monthly inflow during the month (j+1).

b_j = the regression coefficient for estimating the flow in the month j+1 from the month j.

t_i = a normal random deviate with mean equal to zero and unit variance.

S_{j+1}= the standard deviation of the inflow in the month j+1.

r_j = the correlation coefficient between the inflows of the j and j+1 month.

Equation 5.1 is a linear regression model where the inflow in any month is a linear function of the inflow in the preceding month. The sequence of the inflow generated by Equation 5.1 possesses the same general statistical properties as those representing natural inflow.

The log transformed historical monthly streamflow data could be used to generate synthetic monthly streamflow using the Thomas-Fiering model, as the log-transformed data were found to be normally distributed. The use of log transformed streamflow data has the advantage of eliminating the negative flows that occur occasionally when untransformed streamflow are used in the model (Juran and Arup, 2007; Maass et al., 1970).

5.2.2 Monte Carlo Simulation

Simulation is a technique of performing sampling experiments on the model of the system (Ubeda and Allan, 1994). Stochastic simulation is experimenting with the model over time and includes sampling stochastic variates from probability distributions. Monte Carlo Simulation is a technique which has had a great impact in many different fields of computational science (Huber, 1997). The Monte Carlo method is any method which solves a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties (Weisstein1). The method is useful for obtaining numerical solutions to problems which are too complicated to solve analytically.

Monte Carlo methods use random numbers generated from a variety of distributions. Efficient generators have been developed for the most commonly used distributions (e.g. uniform, Gaussian, and exponential) and general techniques (e.g. inversion) are available for arbitrary distributions (Garcia and Wagner, 2006). In many simulation applications, it would be necessary to generate random values that are similar to existing data. This can be done by resampling from the original data. Another method is to fit a parametric distribution from one of the families of the most common distributions, and then generate random values from the selected distribution. However, choosing a suitable family can sometimes be difficult. In this study, three statistical distributions have been used, namely Gamma distribution, Pearson distribution and Johnson system of distributions.

5.2.2.1 Gamma Distribution

In probability theory and statistics, the gamma distribution is a two-parameter family of continuous probability distributions. It has a scale parameter θ and a shape parameter *k*. If *k* is an integer then the distribution represents the sum of *k* independent exponentially distributed random variables, each of which has a mean of θ (which is equivalent to a rate parameter of θ^{-1}) (Wikipedia1). When used to describe the sum of a series of exponentially distributed variables, the shape factor represents the number of variables and the scale factor is the mean of the exponential distribution. This is apparent when the profile of an exponential distribution with mean set to one is compared to a gamma distribution with a shape factor of one and a mean of one.

A random variable *X* that is gamma-distributed with scale θ and shape *k* is denoted:

$$X \sim \Gamma(k, \theta) \; or \; X \sim Gamma\,(k, \theta) \tag{5.2}$$

The probability density function of the gamma distribution can be expressed in terms of the gamma function parameterized in terms of a shape parameter *k* and scale parameter θ. Both *k* and θ are positive values. The equation defining the probability density function of a gamma-distributed random variable x is

$$f(x; k, \theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^{k}\,\Gamma(k)} \quad for \; x > 0 \; and \; k, \theta > 0 \tag{5.3}$$

The gamma distribution has long been used to model many natural phenomena, including daily, monthly and annual streamflow as well as flood flows (Bobée and Ashkar, 1991). In order to generate monthly inflow using Gamma distribution, the first step is to fit a gamma distribution to a given data series using the maximum likelihood estimation and then find the parameters of the selected distribution. Once one gets the parameters of gamma distribution, a new data series could be generated.

5.2.2.2 Pearson and Johnson Systems of Distribution

The statistician Karl Pearson devised a system, or family, of distributions that includes a unique distribution corresponding to every valid combination of mean, standard deviation, skewness, and kurtosis. If the sample values for each of these moments from data are computed, it is easy to find the distribution in the Pearson system that matches these four moments and to generate a random sample (Mathworks1).The Pearson system embeds seven basic types of distribution together in a single parametric framework (Weisstein3). It includes common distributions such as the normal and t distributions, simple transformations of standard distributions such as a shifted and scaled beta distribution and the inverse gamma distribution. Statistician Norman Johnson devised a different system of distributions that also includes a unique distribution for every valid combination of mean, standard deviation, skewness, and kurtosis.

5.3 Application to Actual Streamflow Data

5.3.1 Applications and Data

For testing the performance of the models which have been used in this study, the historical records of monthly inflow of four reservoirs have been used. Inflow data series used in this study were obtained from Ruhr River Association (Ruhrverband). The inflow time series present the inflow to the main reservoirs in the Ruhr river basin namely, Bigge reservoir (Biggetalsperre), Moehne reservoir (Möhnetalsperre), Henne reservoir (Hennetalsperre), and Sorpe reservoir (Sorpetalsperre). The considered period is from January 1967 to December 2008. All time series were checked to find out all missing data.

Figure 5.1 presents the monthly inflow data series used in this study. Figure 5.2 shows another important graph, called a box plot with the statistics of a given data series. It shows, on a graph, the minimum and maximum values, the median value and the top and bottom quartiles for a given set of data. They consist of a box, which surrounds the middle half of the data, containing a line where the median value is. In addition, there are two lines stretching from each end of the box. The extents of these lines are the minimum and maximum data values of the set. Also the mean was added as a small square inside the box.
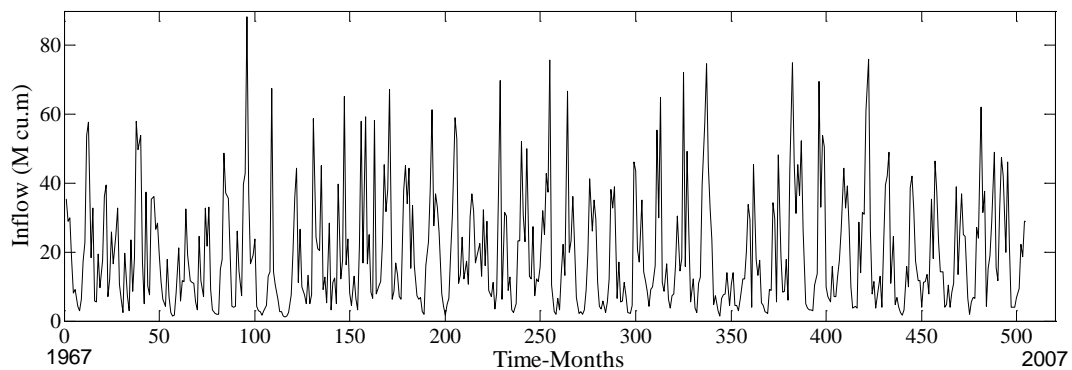
Figure 5.1: Observed monthly inflow (M. cu.m) - Bigge Reservoir
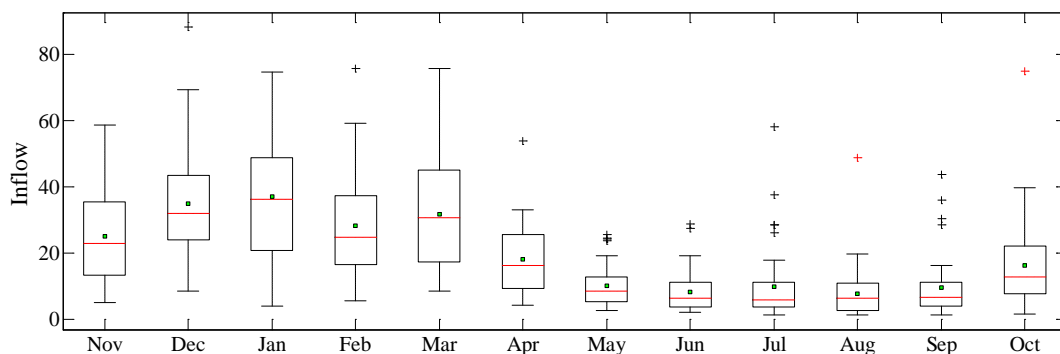
Figure 5.2: Box plot of monthly inflow time series (M. cu.m) - Bigge Reservoir

5.3.2 Stochastic Generation of Streamflow Series

5.3.2.1 Generation of monthly streamflow series Using Thomas-Fiering Simulation

Monthly streamflow data have been generated by using Thomas-Fiering model. Initially a known streamflow of any month (say, December) along with the mean and standard deviation of historical streamflow for that month were fed to equation 5.1. The output produced by this equation is the streamflow of the succeeding month.

As shown in equation 5.1, this equation contains a random part which has a great effect on the statistics of the generated streamflow especially the skewness of the generated inflow (figure 5.3).
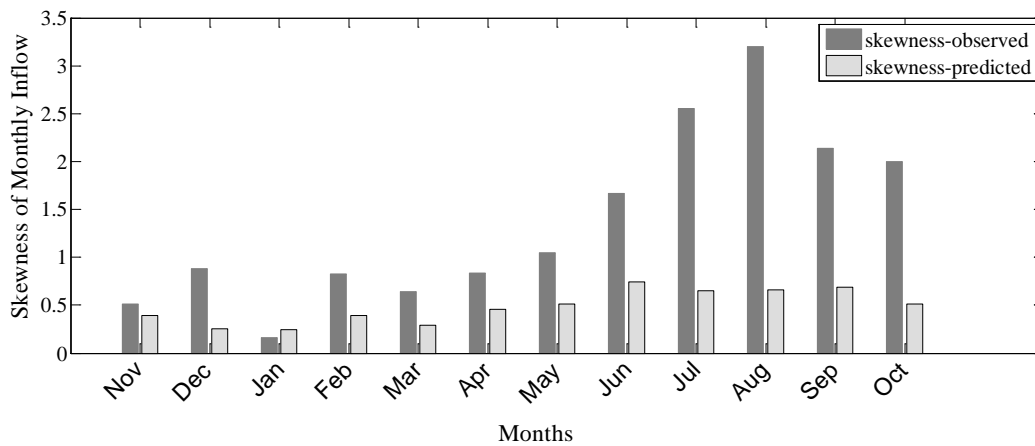
Figure 5.3: Comparison between skewness of observed and generated inflow (with a non specific random part)

The primary model did not preserve the skewness coefficient, and it should be modified to meet the statistical requirement. The preservation could be ensured by several ways. In this study a new method was developed to preserve the statistical parameters of the historical data. The idea behind that, is to generate random numbers (random part in the model) that have the same statistical properties of the random part in the observed streamflow data series and this method has been applied using Monte Carlo model. After adjusting of the random part in the model, the basic statistics such as mean, standard deviation, skewness and correlation coefficient etc. between the historical and generated streamflow have been computed and compared. Obtained results show a harmonization between statistical properties of observed and generated inflow time series as shown in figure 5.4 and table 5.1.

In this study 1000 years of synthetic streamflow has been generated using the selected model. It is worth to be mentioned that if the generated streamflow became negative, then it was replaced with the minimum observed streamflow the month. Another way to avoid negative values could be done by using log transformation (Maass et al., 1970).
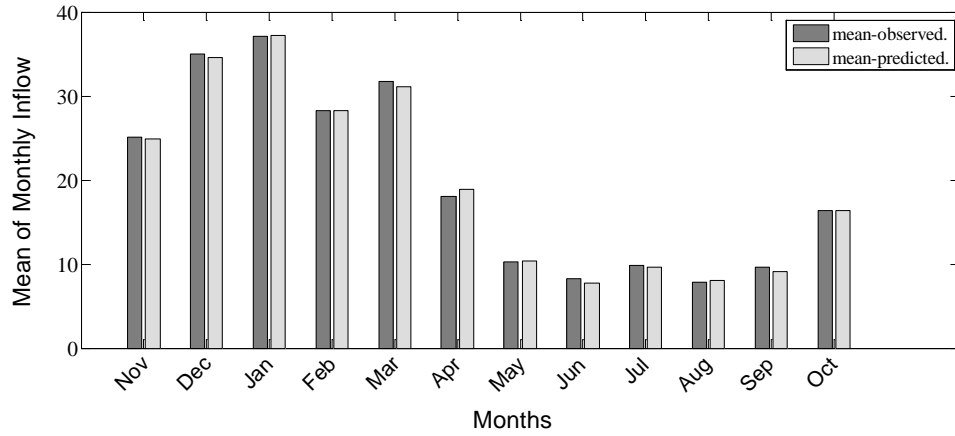
(a)
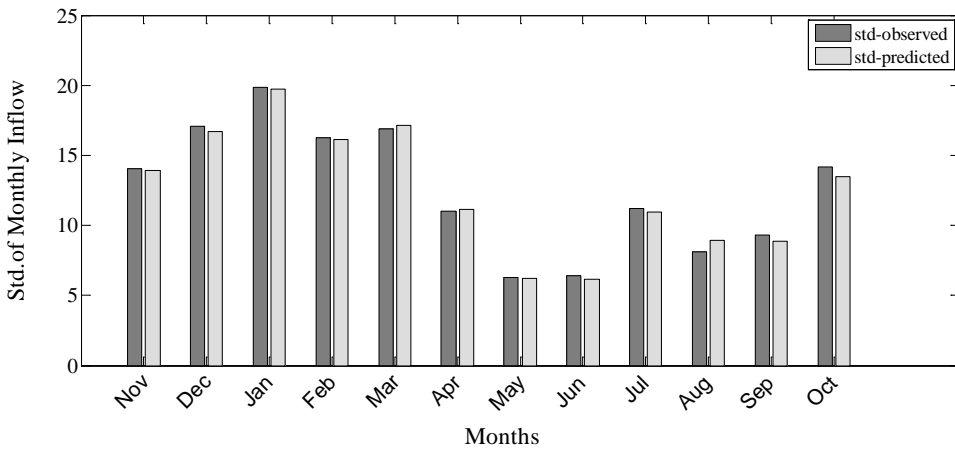
(b)
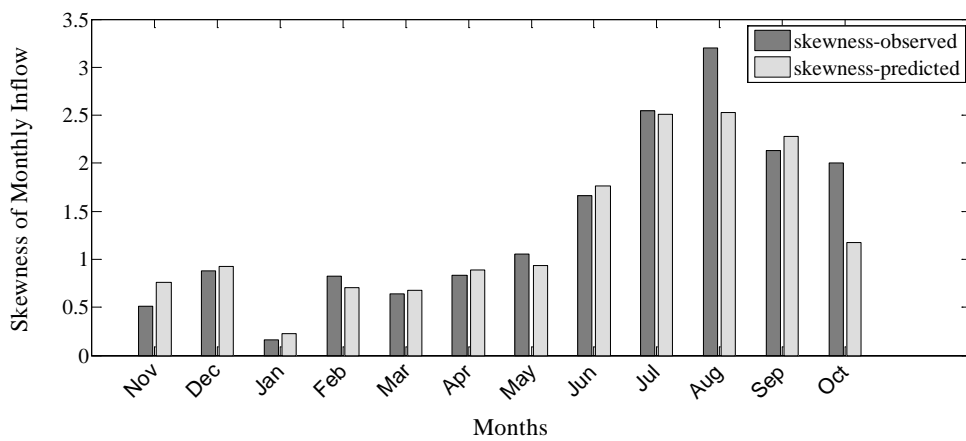
(c)

Figure 5.4: Comparison of the statistics of historical and synthetic monthly streamflow
using Thomas Fiering Model. (Bigge Reservoir)

(a) Mean   (b) Standard deviation   (c) Skewness

Table 5.1: The basic statistics of historical and synthetic annual streamflow using Thomas Fiering model-(Bigge Reservoir)

Properties / Month		November	December	January	February	March	April	May	June	July	August	September	October
mean	Obs.	25.09	35.00	37.13	28.31	31.70	18.02	10.24	8.27	9.88	7.80	9.59	16.38
	Pred.	26.41	36.10	36.88	30.03	32.47	19.35	10.46	7.65	10.12	8.09	9.07	16.38
std.	Obs.	14.04	17.10	19.86	16.23	16.90	11.01	6.28	6.44	11.23	8.13	9.31	14.20
	Pred.	13.75	17.80	20.15	16.77	17.97	11.31	6.24	5.88	11.58	7.89	9.06	14.18
skewness	Obs.	0.51	0.88	0.16	0.83	0.64	0.84	1.05	1.66	2.55	3.20	2.13	2.00
	Pred.	0.75	1.06	0.22	0.67	0.58	0.87	0.99	1.96	2.38	1.87	2.49	1.33
median	Obs.	22.92	32.03	36.40	24.69	30.68	16.13	8.43	6.31	5.97	6.36	6.70	12.69
	Pred.	24.65	33.11	35.78	27.27	28.96	17.91	8.80	5.60	4.95	5.19	5.70	13.48

Properties \ Month		November	December	January	February	March	April	May	June	July	August	September	October
kurtosis	Obs.	2.43	3.87	2.10	3.29	2.78	3.77	3.22	5.33	10.03	16.55	7.31	8.19
	Pred.	3.36	4.19	2.07	2.82	2.41	3.83	3.51	6.99	8.63	7.41	9.82	5.01
1st Q	Obs.	13.36	24.00	20.83	16.47	17.36	9.22	5.42	3.84	3.75	2.57	4.00	7.60
	Pred.	16.00	22.77	19.32	16.29	17.10	10.48	5.82	3.74	3.51	2.41	3.56	4.87
3rd Q	Obs.	35.48	43.48	48.90	37.32	45.23	25.71	12.70	11.06	11.06	10.80	11.14	22.21
	Pred.	34.55	45.56	52.13	39.98	44.98	26.47	13.92	9.23	10.78	10.89	10.38	23.09

Figure 5.5 presents a comparison between the observed and the generated monthly inflow for Bigge reservoir. The empirical cumulative distribution function plots of the data are shown in figure 5.6. It is notable from figure 5.6 that the observed and the generated monthly inflow have the same distribution.
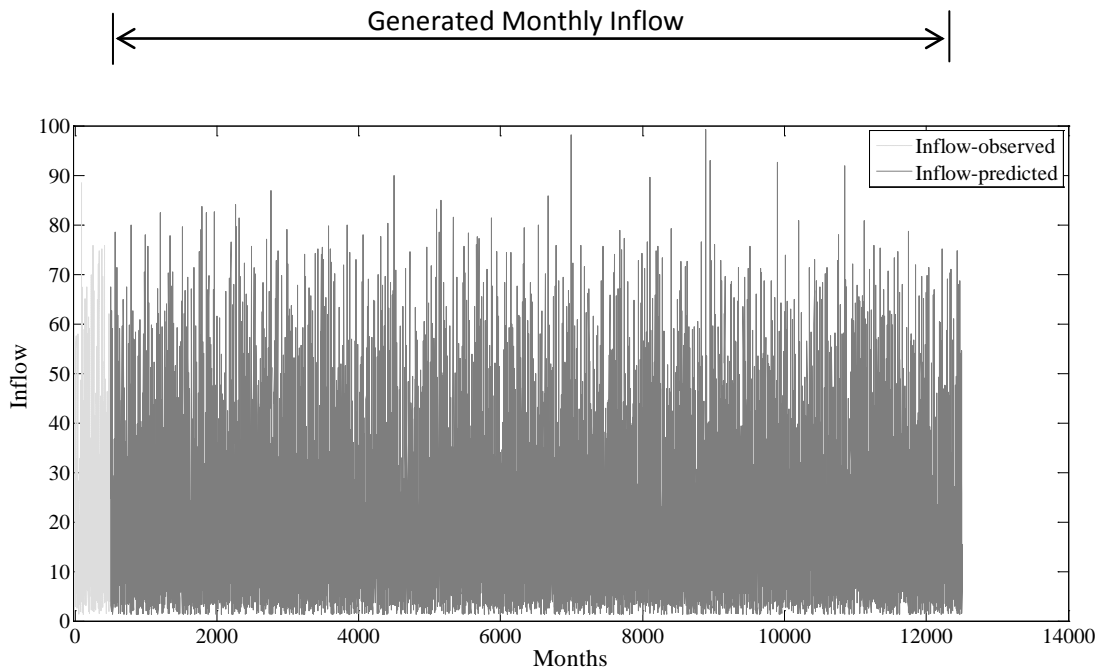
Figure 5.5: Comparison of the observed and generated monthly inflow using Thomas-Fiering model (Bigge Reservoir)
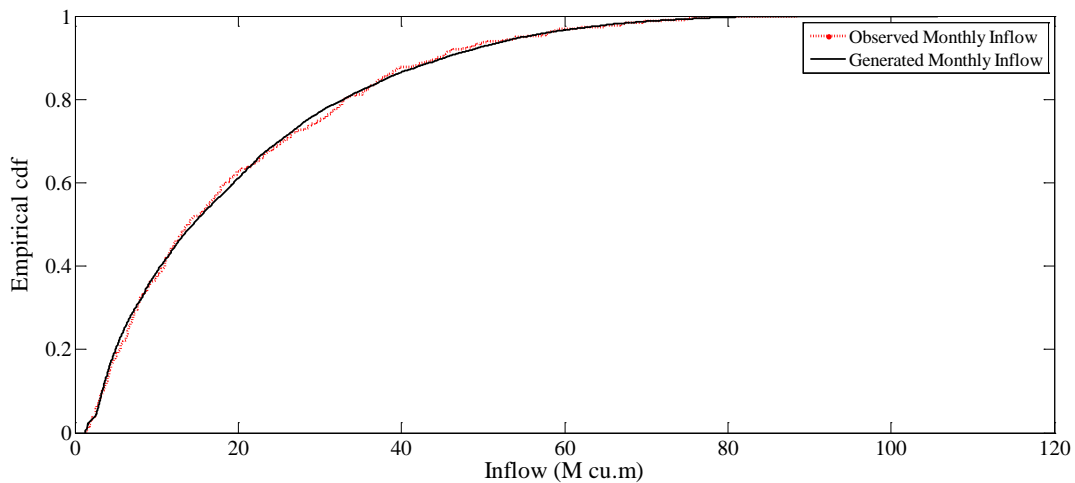
Figure 5.6: Empirical cumulative distribution function (CDF) for the observed and generated monthly inflow using Thomas-Fiering model (Bigge Reservoir)

5.3.2.2 Generation of monthly streamflow series using Monte Carlo Simulation.

Monthly streamflow data have been generated using Monte Carlo Simulation by applying several approaches namely Gamma distribution, Pearson distribution, and Johnson system of distributions. The Gamma distribution was fitted to the observed monthly time series then the parameters of the distribution were calculated then a 1000 year of synthetic streamflow has been generated. Results of this approach show that, this method preserves the mean and the standard deviation but it does not preserve the skewness coefficient as shown in figure 5.8.
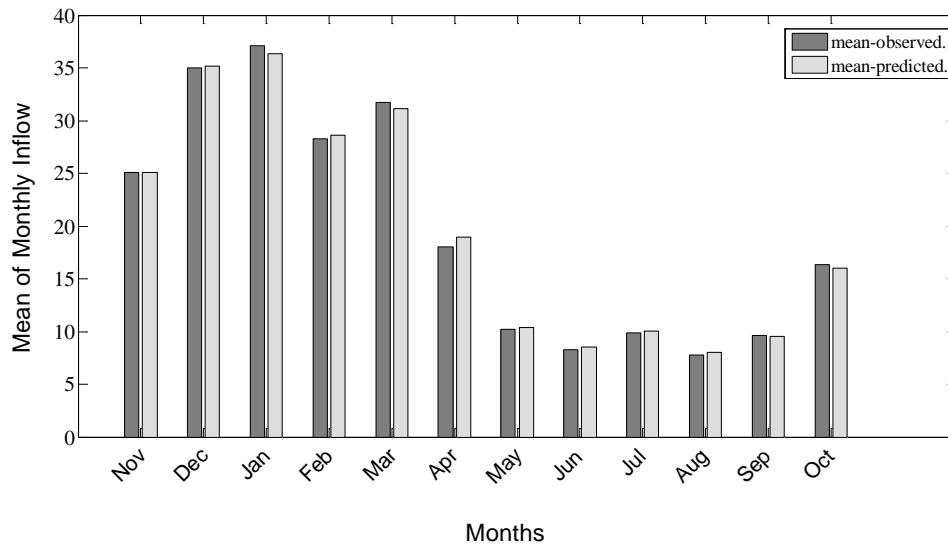
Figure 5.7: Comparison between mean of observed and generated Inflow
(Gamma Distribution) (Bigge Reservoir)
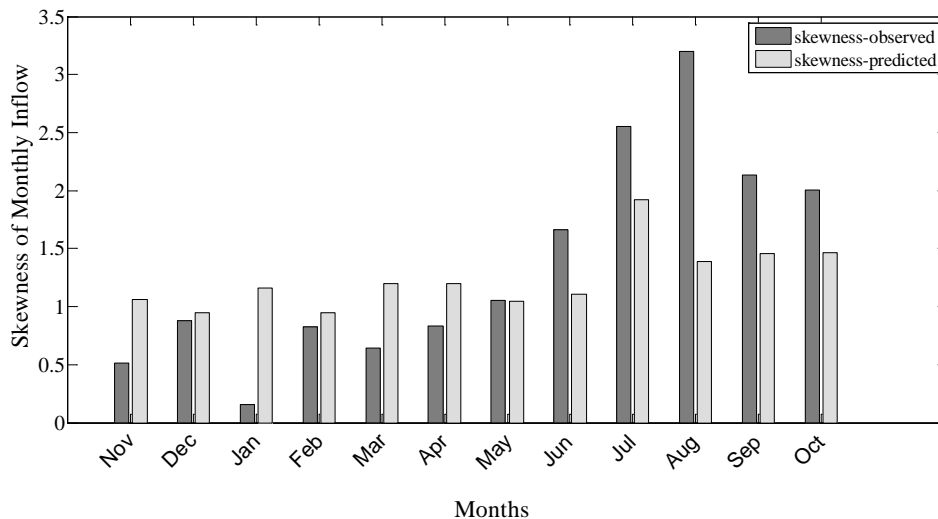
Figure 5.8: Comparison between skewness of observed and generated inflow
(Gamma Distribution) (Bigge Reservoir)

For adequately modeling, one must have a distribution which would appropriately preserve the skewness inherent in the hydrologic data series. The result of the Pearson system showed that this approach is more capable than the other suggested approaches (Gamma distribution and Johnson Systems) to preserve the statistical properties of the observed data series because it covers a wide range of distribution shapes, including both symmetric and skewed distributions.

To generate a synthetic streamflow from the Pearson distribution that closely matches the observed data, simply four sample moments have been computed (mean, standard deviation, skewness, kurtosis) and those moments have been treated as distribution parameters. Then one of the distributions within the Pearson system which matches the combination was selected.

Streamflow series generated have been compared with the observed series on the basis of statistics properties (figure 5.9 and table 5.2). Figure 5.10 present a comparison between the observed and the generated monthly inflow for Bigge reservoir. The empirical cumulative distribution function plots of the data are shown in figure 5.11. It is notable from figure 5.11 that the observed and the generated monthly inflow have the same distribution.

In this study, several run tests have been carried out. The sequences of the same length as the historical data series at each reservoir have been also generated and the three main descriptors, i.e. the mean, standard deviation and skewness coefficient for each month have been computed for each sequence. It was found that the values of these descriptors were closer to those of the historical sequence when only one sequence with larger sample was generated. A size of 1000 was found to produce satisfactory results, and that is the reason for its use.
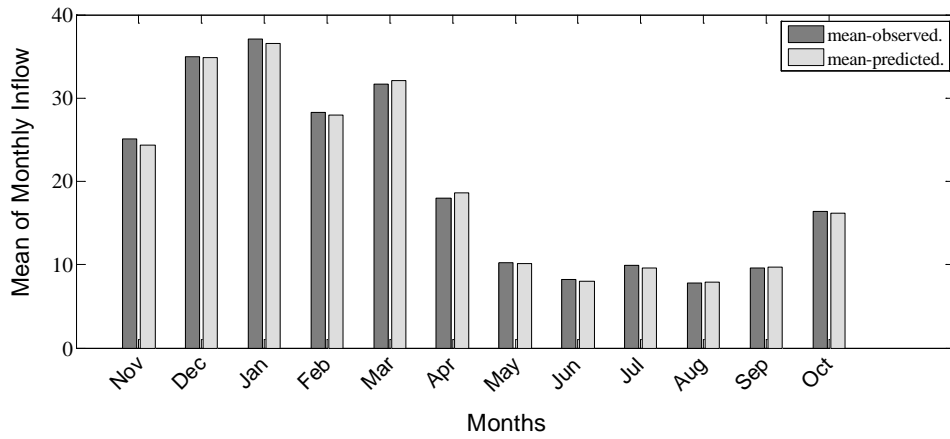
(a)
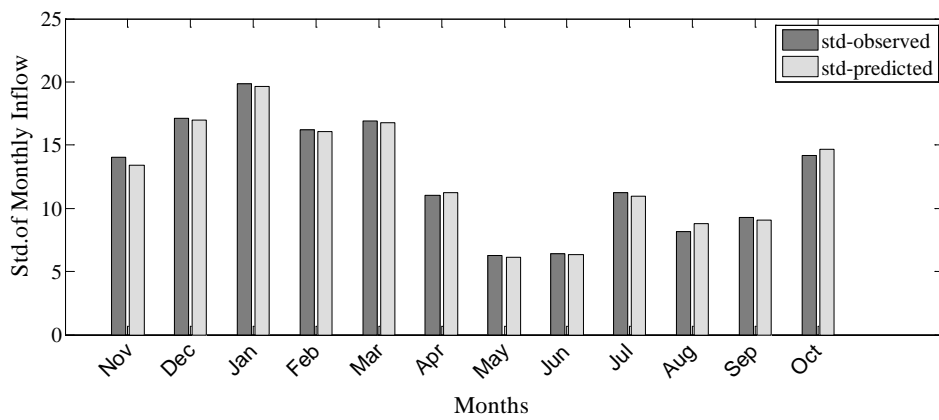
(b)
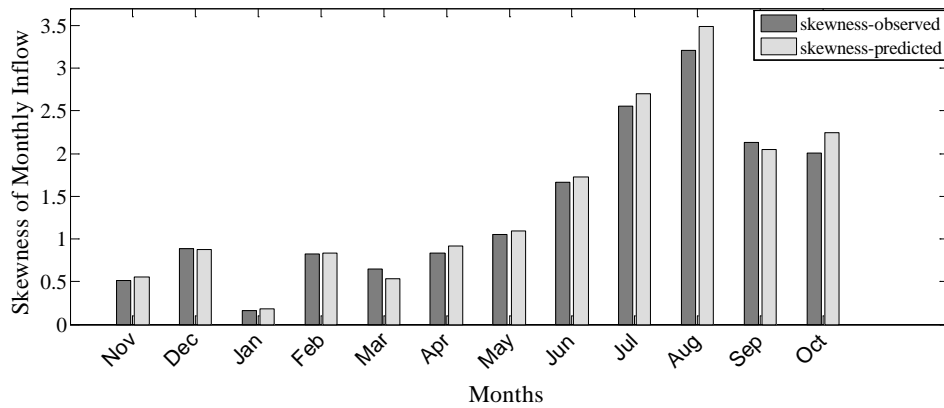
(c)

Figure 5.9: Comparison of the statistics of historical and synthetic monthly streamflow
using Monte Carlo simulation (Bigge Reservoir)

(a) Mean   (b) Standard deviation   (c) Skewness

Table 5.2: The basic statistics of historical and synthetic annual streamflow using Monte Carlo simulation (Bigge Reservoir)

Properties	Month	November	December	January	February	March	April	May	June	July	August	September	October
mean	Obs.	25.09	35.00	37.13	28.31	31.70	18.02	10.24	8.27	9.88	7.80	9.59	16.38
	Pred.	24.38	34.88	36.58	27.92	32.08	18.66	10.12	8.01	9.59	7.89	9.67	16.18
std.	Obs.	14.04	17.10	19.86	16.23	16.90	11.01	6.28	6.44	11.23	8.13	9.31	14.20
	Pred.	13.39	17.01	19.65	16.04	16.75	11.22	6.12	6.32	10.98	8.79	9.09	14.64
skewness	Obs.	0.51	0.88	0.16	0.83	0.64	0.84	1.05	1.66	2.55	3.20	2.13	2.00
	Pred.	0.56	0.87	0.19	0.83	0.54	0.92	1.09	1.73	2.70	3.49	2.04	2.25
median	Obs.	22.92	32.03	36.40	24.69	30.68	16.13	8.43	6.31	5.97	6.36	6.70	12.69
	Pred.	22.36	31.98	35.85	24.79	29.43	16.79	8.16	5.25	4.38	4.21	5.28	10.79

Properties / Month		November	December	January	February	March	April	May	June	July	August	September	October
kurtosis	Obs.	2.43	3.87	2.10	3.29	2.78	3.77	3.22	5.33	10.03	16.55	7.31	8.19
	Pred.	2.54	3.70	2.12	3.28	2.53	3.73	3.30	5.38	11.08	17.86	6.96	9.73
1st Q	Obs.	13.36	24.00	20.83	16.47	17.36	9.22	5.42	3.84	3.75	2.57	4.00	7.60
	Pred.	13.53	21.83	20.80	15.15	18.24	10.26	5.08	3.55	3.53	3.39	3.89	6.30
3rd Q	Obs.	35.48	43.48	48.90	37.32	45.23	25.71	12.70	11.06	11.06	10.80	11.14	22.21
	Pred.	33.04	45.37	51.02	37.16	43.92	24.82	13.22	9.82	10.42	8.20	11.68	20.73
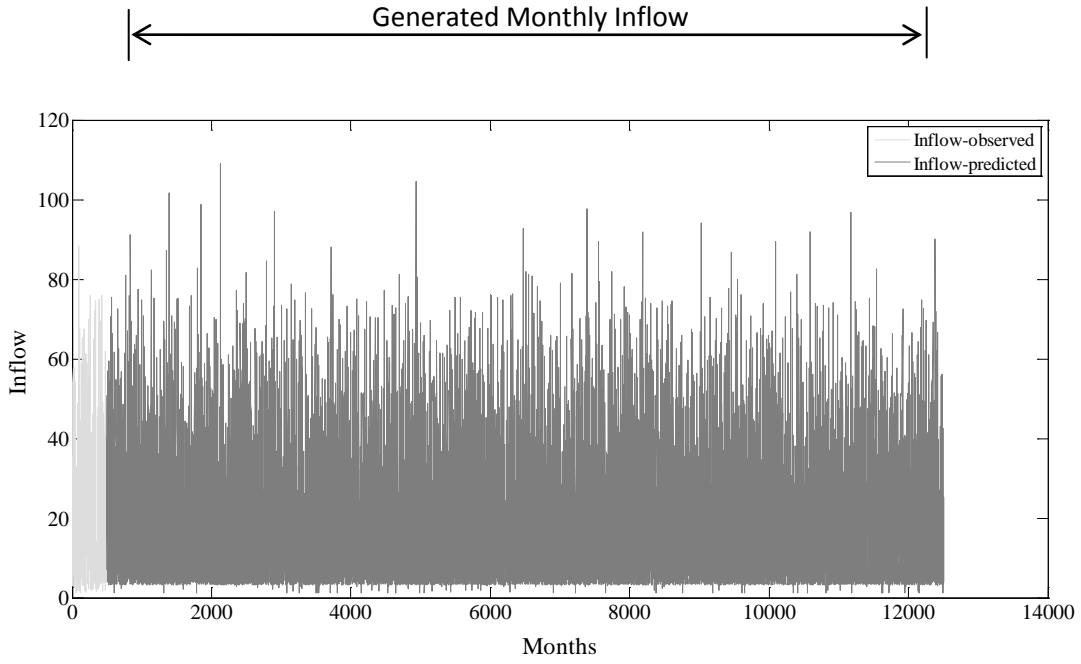
Figure 5.10: Comparison of the observed and generated monthly inflow using Monte Carlo simulation - (Bigge Reservoir)
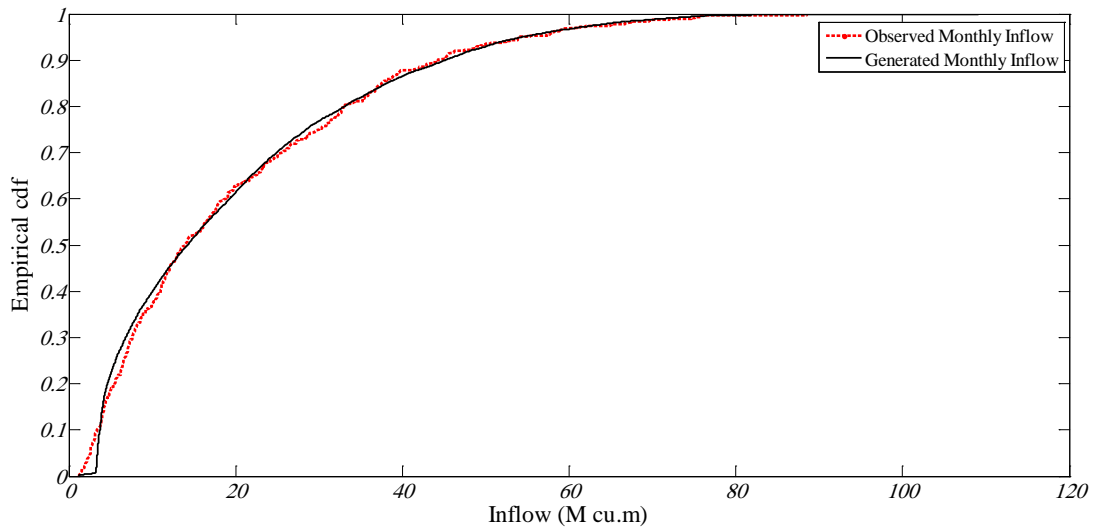
Figure 5.11: Empirical cumulative distribution function (CDF) for the observed and generated monthly inflow Monte Carlo simulation (Bigge Reservoir)

5.3.3 Comparison between the Results of theThomas-Fiering Model and the Monte Carlo Simulation Model.

To assess the performance of each model, the negative values, correlation between consecutive months and Skewness coefficient of the twelve monthly streamflow have been examined. These are shown in table 5.3 and table 5.4. All these parameters have been calculated for the historical data series and from a 1000-year generated sequence (the generated sequence in a month consisted of 1000 values representing the streamflow for that month in 1000 consecutive years). The values presented in table 5.3 are average values of 10 runs.

Table 5.3: Comparison of model performance based on 1000-years generated sequences. (Skewness &negative values) (Bigge reservoir)

Month	Skewness coefficient		
	Historical data	Thomass Fering model	Monte Carlo simulation model
Jan.	0.16	0.239	0.173
Feb.	0.828	0.783	0.757
Mar.	0.644	0.631	0.671
Apr.	0.836	0.879	0.882
May.	1.051	0.893	1.025
Jun.	1.663	1.759	1.726
Jul.	2.55	2.499	2.29
Aug.	3.203	2.513	3.357
Sep.	2.135	2.365	1.898
Oct.	2	1.208	2.02
Nov.	0.51	0.737	0.49
Dec	0.88	0.987	0.9
Number of negative values		103	9

Table 5.4: Comparison of model performance based on 1000-years generated sequences. (Correlation coefficient) (Bigge reservoir)

Month	Correlation coefficient between consecutive months		
	Historical data	Thomass Fering model	Monte Carlo simulation model
Jan.	0.102	0.107	0.019
Feb.	0.013	-0.026	0.01
Mar.	0.014	0.067	-0.013
Apr.	0.076	0.06	-0.004
May.	0.043	0.031	0.025
Jun.	0.328	0.323	0.082
Jul.	0.215	0.244	0.041
Aug.	0.079	0.107	-0.01
Sep.	0.336	0.279	-0.014
Oct.	0.261	0.285	0.02
Nov.	0.681	0.696	-0.01
Dec	0.321	0.33	0.012

By inspecting the results shown in tables 5.3 and 5.4, it appeared that the model of Thomas-Fiering, with modifications to account for the preservation of the monthly skewness coefficients, and Monte Carlo simulation model, both seem to perform very well as far as the mean, standard deviation and skewness coefficient are to be reproduced. However, it should be noted that the model of Thomas-Fiering preserve the correlation coefficient between consecutive months. In the other hand the number of negative values obtained by Thomas-Fiering model is more than those by Monte Carlo simulation model. These negative data can be avoided using log transformation. But when a transformation is applied to the historical record to make it normal or to get avoid occurrence of negative values in generated sequences, then the model preserves the parameters of the transformed data, but not those of the historical sequence (Phien and Ruksasilp, 1981).

5.4 Detection of Dry Periods

As mentioned in previous chapters, the main goal of this thesis is to manage water resources during drought periods. So the next step after generating the monthly inflow is to detect dry periods. Several time steps have been applied, namely one year, two years, three years, four years and five years. In each time step the minimum summation of monthly inflow has been detected through the 1000 years of synthetic inflow. To illustrate this approach, a time step of three years is taken as an example. A moving window with 36 months width has been used to calculate the summation of generated monthly inflow. This window moves from the beginning to the end of the data series with one month lag. After that the minimum summation, i.e. the window with minimum summation, has been selected as the driest period. Results of this approach (figure 5.12) could be used as input to the Optimal Reservoirs Operation model presented in the next chapter.
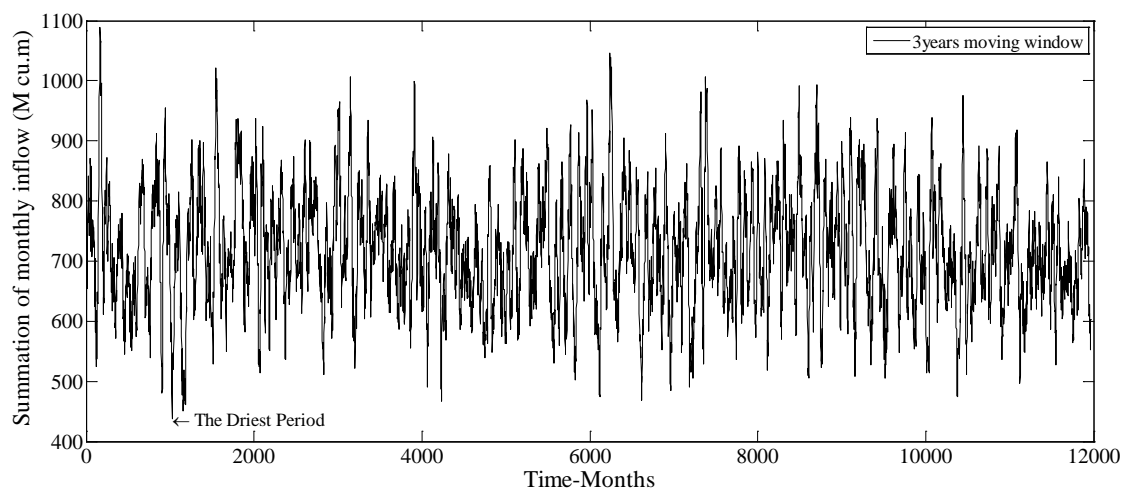
Figure 5.12: An example for detection of driest period using generated monthly inflow

5.5 Conclusion

Generating of streamflow using stochastic models is a very important process for water resources planning and operational purposes. In this study, the stochastic streamflow generation model of Thomas-Fiering and Monte Carlo simulation model have been applied to synthetically generate monthly inflow scenarios for four reservoirs in the Ruhr river basin. A new method has been applied to Thomas-Fiering model to preserve the statistical parameters of the historical data. Comparison of statistical parameters such as means, standard deviation and skewness for observed data and generated data from the used approaches were presented. The results showed that generated data have successfully preserved the historical statistical parameters of streamflow. Results also showed that, the Thomas-Fiering model has preserved the correlation coefficient between consecutive months. Thus, it can be said that the Thomas-Fiering model is suitable to be used for producing inflow scenarios needed for the optimization model presented in chapter 6 and stochastic simulation model presented in chapter 7.

Chapter 6

Reservoir System Optimization during Drought Events

6.1 Background

Reservoir operations involve flood control and drought management with the goal of minimizing adverse environmental impacts and securing water supply for a wide range of purposes and a diverse set of water (Tu et al., 2008). Optimizing reservoir operations may take into account many factors, such as water allocation, streamflow regulation, and real-time decision making regarding advanced scheduling of water releases and hydropower generation (Tu et al., 2008). Reservoir operating rules are used to determine water yield from a single-reservoir system or a multireservoir system under various hydrologic conditions. Reservoir operation involves many decision variables, multiple objectives as well as considerable risk and uncertainty (Wang et al., 2010). Various techniques are applied to improve the performance of reservoirs operation. These techniques include Linear Programming (LP), Nonlinear Programming (NLP), Dynamic Programming (DP), and Heuristic Programming such as Genetic algorithms, Fuzzy logic, and Neural Networks (Adeyem, 2009). The method chosen for any particular case depends mainly on: (i) the character of the objective function and whether it is known explicitly; (ii) the nature of the constraints; and (iii) the number of independent and dependent variables (Babu and Angira, 2001).

One of the simplest methods of optimization techniques is the linear programming (LP), which had been widely applied to several cases of reservoir operation problems. Some of the applications of LP in reservoir operation are suggested in Dorfman (Dorfman, 1962), Martin (Martin, 1987), Palmer (Palmer and Holmes, 1988), Randall (Randall et al., 1990), Mohan (Mohan and Raipure, 1992) and Mujumdar (Mujumdar and Teegavarapu, 1998). In the Dynamic Programming method (DP), multidecision problems are broken down into a sequence of separate, but interrelated, single-decision sub-problems. thus, complex problems can be solved by combining the solutions of the sub-problems to obtain the solution of the entire problem (Ferreira et al., 1996). It is well suited to deal with short-term operation (hourly or daily) when the hydrologic inputs and water demands are generally considered deterministic. In case of optimization of real-time operations for reservoir systems, the objective functions often consist of benefits and costs expressed as non-linear functions of storage and discharge. The two approaches used in the solution of non-linear programming (NLP) are direct or pattern search method and gradient-based optimization. However, as compared to linear programming and dynamic programming, the number of applications of nonlinear programming methods in water resources studies is relatively small (Reddy, 2006).

Pattern search finds a local minimum of an objective function by a method called polling (MathWorks9). The search starts at an initial point, which is taken as the initial point in the first step. Then the algorithm generate a pattern of points, typically plus and

minus the coordinate directions, times a mesh size, and center this pattern on the current point. An evaluation of the objective function at every point in the pattern is then provided, if the minimum objective in the pattern is lower than the value at the current point, then the poll is successful and the minimum point found becomes the current point. Then the mesh size is doubled and the algorithm proceeds to the first step. If the poll is not successful, then the mesh size is halved. If the mesh size is below a threshold, the iterations stop otherwise, the current point is retained and the algorithm proceeds at first Step.

In recent years, Genetic Algorithms (GAs) have become popular among researchers as a robust and general optimization technique (Namchaiswadwong et al., 2006). The results of employment of GAs to a wide variety of problems have indicated their potential in the application to water resource management. The genetic algorithm (GA) is one of the most promising techniques in that domain and has received a great deal of attention with regard to optimizing complex systems (Chen, 2003). GAs handle nonlinear optimization problems in efficient manner and it differs from traditional methods in number of ways (Goldberg, 1989). The concept of GAs was developed by Holland and his colleagues in the 1960s and 1970s (Konak et al., 2006). In GAs terminology, a solution vector is called an individual or chromosome. Chromosomes are made of discrete units called gens. Each gene controls one or more features of the chromosome. GAs operate with a collection of chromosomes called a population. The population is normally randomly initialized. As search evolves, the population includes fitter and fitter solutions, and eventually it converges, meaning that it is dominated by a single solution (Konak et al., 2006). GAs use two operators to generate new solutions from existing ones: crossover and mutation. The crossover operator is the most important operator of GAs.

The objective of this chapter is to build a model, which utilizes maximum available information, for optimization of reservoir operation by applying Genetic Algorithm, Pattern Search and gradient-based Approaches to the Ruhr reservoirs system in which utilization of multipurpose reservoirs are considered. The specific objectives of this chapter are:

1. To apply the suggested approaches to the multipurpose reservoir operation in the study area;

2. To compare between the outputs of the developed model and the historical records especially in case of drought events;

3. To evaluate GA performance against that of pattern search and gradient-based Approaches.

6.2 Development of a Reservoir Optimization Model in the Context of Drought

Optimization is a procedure of finding and comparing feasible solutions until no better solution can be found (Deb, 2001). In general, optimization problems can be classified into two groups; single objective and multiobjective problems. The main goal of single objective optimization problems is to define the minimum or the maximum value of an objective function, depending on the goal. The procedures for solving this kind of problem are gradient-based and heuristic-based search techniques.

Multiobjective optimization problems represent an important class of real-world search and optimization problems. Multi-objective optimization (MOP) refers to problems which includes several objectives that are expected to be fulfilled simultaneously. Burke (Burke and Landa Silva, 2006) reported that for multiobjective optimization problems, three broad typical approaches can be identified to deal with multiple objectives as follow:

1. Optimizing one objective at a time while imposing constraints on the other objectives,

2. Combining all objectives into a single objective,

3. Optimizing all objectives simultaneously.

Frequently in the first approach one objective is chosen as the dominating objective and the rest of the objectives are treated as constraints. In the second approach preferences for the objectives are established a priori while and the vector of objectives is scalarized into one objective by averaging the objectives using a weight vector. In the last approach, no preference information is considered or is available before the search. In terms of the number of solutions needed, it may be that only one solution is required or that a set of solutions should be presented to the decision-makers so that one of the solutions can be chosen. In the last case, this set of solutions should represent a trade-off among the different objectives. It is also commonly required that this set of solutions be as diverse as possible. Such diversity may be in terms of the solution space, the objective space or both, depending upon the problem domain.

In this study the approach of combining all objectives into a single objective has been used for developing the optimization model. Details about the procedures of the developed model are presented in the following section with an illustration to the reservoir Bigge.

6.2.1 Objective Functions

In order to deal with multiple objectives of the reservoir system, the approach of combining all objectives into a single objective is adopted to convert the multiple objectives problem into a single objective problem. The objective functions of the model are to maximize hydropower production subject to flood restrictions and to minimize the sum of squared deviations of releases from demands under release constraints and other physical and technical constraints. The model is formulated for monthly operation, as follows:

1) Minimize Sum of Squared Deviation of Releases from Demands,

$$Minimize\ SQDV = \sum_{i=1}^{12} \left(\sum D_{n,t} - \sum R_{m,t} \right)^2 \qquad (6.1)$$

Where;

$SQDV$: The sum of squared deviation of releases from demands;

$D_{n,t}$: The demands in period t in Mm^3, $n = 1,\ldots\ldots,n$ number of demands;

$R_{n,t}$: The releases in period t in Mm^3, $m = 1,\ldots\ldots,m$ number of releases

2) Maximize Annual Energy Production

$$Maximize\ E = \sum_{t=1}^{12} \left(\sum \left(P_m * R_{m,t} * H_{m,t} \right) \right) \qquad (6.2)$$

Where;

E : The annual energy produced in MkWh;

P_m : The power production coefficient (number of turbines = $1,\ldots, m$)

$H_{m,t}$:The net heads available to turbines (number of turbines = $1,\ldots, m$)

The objective functions presented in equations 6.1 and 6.2 are subject to some constraints as illustrated in the following section.

6.2.2 Constraints

The objective functions expressed by equations 6.1 and 6.2 are subjected to the following constraints:

    a) State transformation equation: continuity of inflow, storage, release and losses.

$$S_{t+1} = S_t + I_t - \sum R_{m,t} - Q_t - Q_{loss} \qquad \textit{For all t} \qquad (6.3)$$

S_t : Active reservoir storage at the beginning of period t in Mm^3;

I_t : The inflow to the reservoir during period t in Mm^3;

O_t, O_{loss} : Overflow and Losses from the reservoir in period t in Mm^3 .

The storage of the reservoir varies from the dead storage capacity to the maximum storage capacity according to the month under consideration. Also during the flood risk period between 1 November and 1 February, a flood control storage space of 32 million cubic meters is kept available, which is then released for refilling in the period between 1 February and 1 May.

$$S_{min,t} < S_t < S_{max,t} \qquad \textit{(see appendix D)} \qquad (6.4)$$

    b) Maximum power production limits

$$P_m * R_{m,t} * H_{m,t} < E_{max,m} \qquad \textit{For all m, t} \quad \textit{(see appendix D)} \qquad (6.5)$$

Where, $E_{max,m}$ is the maximum amounts of power in MkWh that can be produced by a turbine m in a time period t.

    c) Demands constraint

$$D_{n,t,min} < R_{m,t} < D_{n,t,max} \qquad \textit{(see appendix D)} \qquad (6.6)$$

Where, $D_{n,t,min}$ and $D_{n,t,max}$ are the minimum and maximum demands in a time period t .

Water quality requirement constraint

$$\sum(R_{m,t}) \geq MRWQ \qquad (6.7)$$

Where $\sum(R_{m,t})$ is the summation of releases in a time period t, $MRWQ$ is the minimum releases to meet downstream water quality requirement in Mm^3

    d) Steady state storage constraint

$$S_{13} = S_1 \qquad (6.8)$$

Under steady-state conditions for the storage, the storage at the end of last month of a year is to be equal to the initial storage at the beginning of first month of that year.

As mentioned before, in this study a weighted approach is adopted to convert the multiple objectives problem into a single objective problem. The user can specify the priorities by giving a specified weight for each function. As presented in equations 6.1 and 6.2, the two objective functions do not have the same units. To bring both the objectives into same units, the hydropower objective and *SQDV* objective are non-dimensionalized and the final fitness function for the model is as follow:

$$F = \frac{W_1}{W_1+W_2} * \sum_{t=1}^{12} \frac{(\sum E_{m,max} - \sum P_m * R_{m,t} * H_{m,t})}{\sum E_{m,max}} + \frac{W_2}{W_1+W_2} * \sum_{t=1}^{12} \frac{(\sum D_{n,t} - \sum R_{m,t})^2}{\sum D_{n,t}} \qquad (6.9)$$

Where W_1 and W_2 are constant weights to be chosen based on priority. So the final model is to minimize *F* in the above equation in duly satisfying the constraints.

6.2.3 Model Application Using Genetic Algorithm

To apply the optimization approaches to the formula presented in equation 6.9, several scenarios have been analyzed. The suggested scenarios are focused especially on dry periods as shown in table 6.1. The driest two years in the historical records are presented by scenarios13 and 14. Thomas-Fiering model has been used to generate inflow of 1000 years, and then the year with minimum summation of monthly inflow has been selected to be scenario number 15 in table 6.1.

The parameters used in applying the optimization model using GA have been selected after studying of how the variation in the output of the optimization model can be apportioned to these parameters in the input of the model. In the developed model, the GA parameters have been fixed for all scenarios. Crossover probability of 0.8; population size of 500 and generation size of infinity have been selected.

Once the GA parameters are fixed the model is run for any scenario of the selected scenarios in table 6.1. The optimization model has been applied to all scenarios for two sets of priority; hydropower as the priority and only *SQDV* as the priority. Figure 6.1 presents the results of scenario 1 (Base Case) and the priority is for the hydropower. The optimal annual hydropower is $E = 25.70$ MkWh and sum of squared deviation of releases from demands, *SQDV*, equals 1984. It is worth mentioning that the annual power produced by the reservoir Bigge is 22.60 MkWh (Ruhrverband-online-Report).

In case of the priority for *SQDV*, the obtained hydropower is $E = 22.97$ MkWh and *SQDV*= 387.24 (figure 6.2). It can be clearly observed for figures 6.1 and 6.3 that, if the reservoir is having *SQDV* as the only priority, it tends to keep the storage in the reservoir at high level and this due to the relatively small downstream demands, whereas for priority for hydropower this is reversed, which requires higher releases to produce optimal hydropower. As shown in figures 6.1 and 6.2, the storage in the beginning and the end of the optimization processes is the same. Figure 6.3 presents results of applying scenario 12 in case of the hydropower is the only priority.

Table 6.1: Description of the scenarios which are used in the optimization model

Scenario (1)	Base Case - Mean monthly inflow and initial storage=mean storage
Scenario (2)	Mean monthly inflow and initial storage=90 % of mean storage
Scenario (3)	Mean monthly inflow and initial storage=80 % of mean storage
Scenario (4)	Mean monthly inflow and initial Storage=70 % of mean storage
Scenario (5)	Mean monthly inflow-0.25 * std and mean Initial Storage
Scenario (6)	Mean monthly inflow-0.25 * std and initial storage=90% of mean storage
Scenario (7)	Mean monthly inflow-0.25 * std and initial storage=80 % of mean storage
Scenario (8)	Mean monthly inflow-0.25 * std and initial Storage=70 % of mean storage
Scenario (9)	Mean monthly inflow-0.50 * std and mean initial Storage
Scenario (10)	Mean monthly inflow-0.50 * std and initial storage=90 % of mean storage
Scenario (11)	Mean monthly inflow-0.50 * std and initial storage=80 % of mean storage
Scenario (12)	Mean monthly inflow-0.50 * std and initial storage=70 % of mean storage
Scenario (13)	Monthly inflow of the calendar year 1996
Scenario (14)	Monthly inflow of the calendar year 2003
Scenario (15)	The Year with minimum summation of monthly inflow generated using Thomas-Fiering model
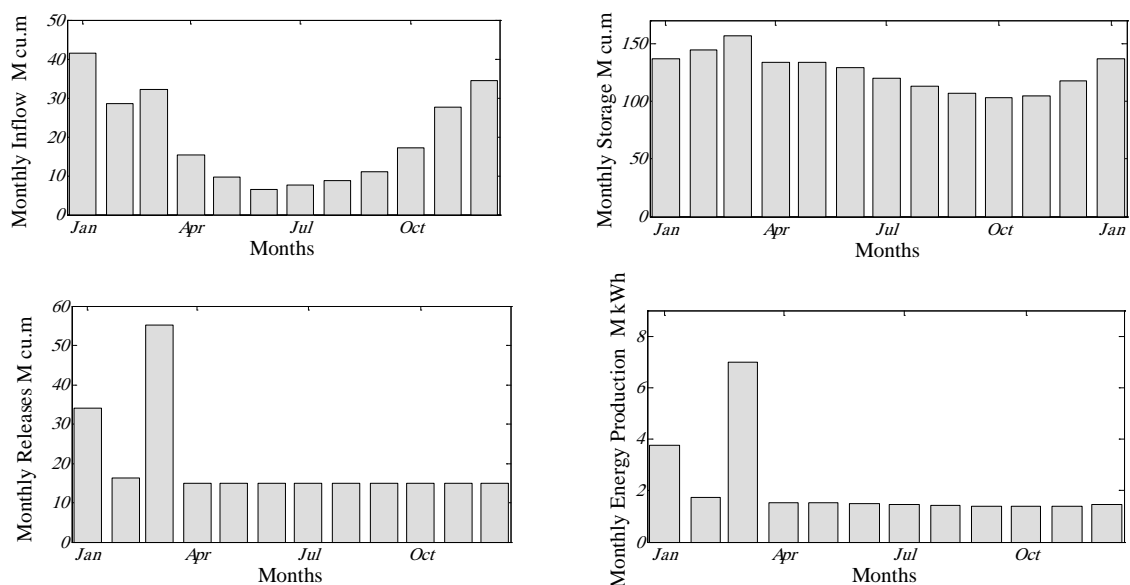
Figure 6.1: Optimal release policy - hydropower is the only priority (Scenario 1)
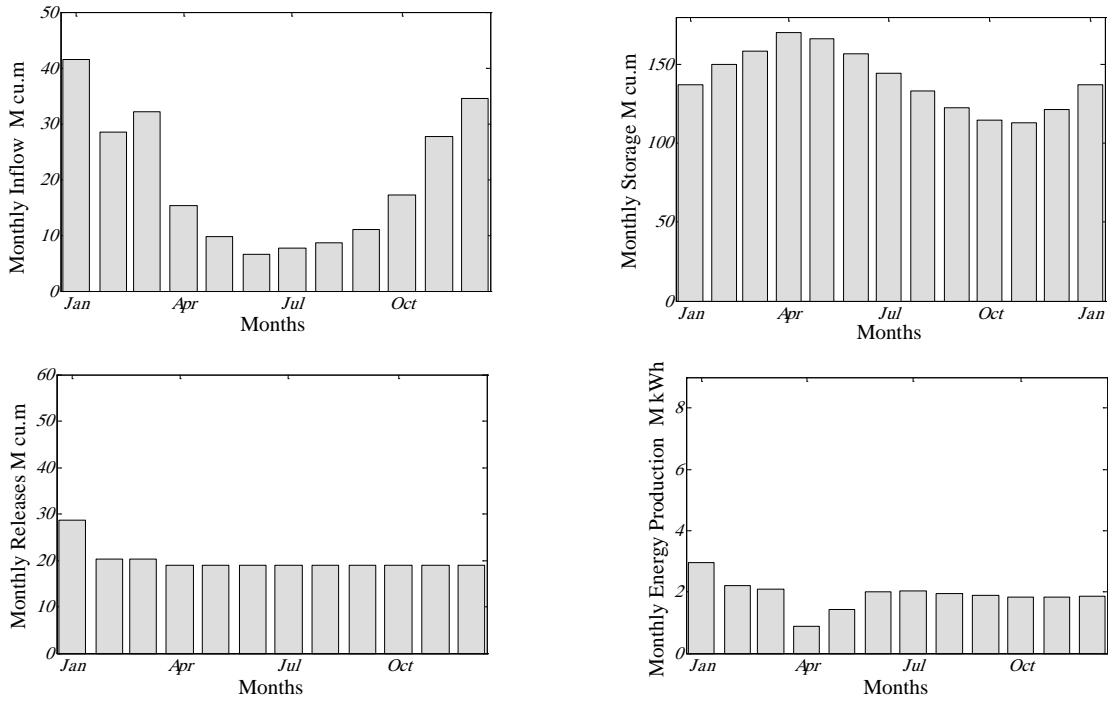
Figure 6.2: Optimal release policy - *SQDV* is the only priority (Scenario 1)
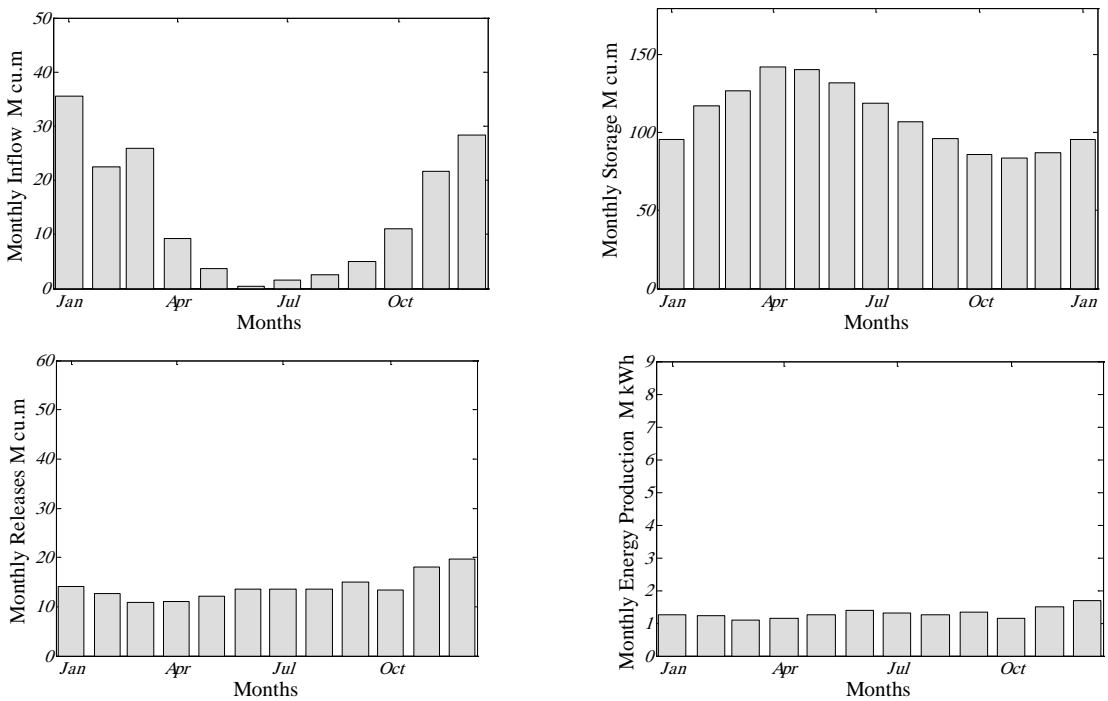
Figure 6.3: Optimal release policy - hydropower is the only priority (Scenario 12)

Results of scenario 15 for the two priorities are shown in figure 6.4 and 6.5 respectively. Result indicates that, there is no clear difference between the two priorities. This is due to the low inflow of this scenario, 101.46 M.m^3, which presents the driest year of 1000 generated years using Thomas-Fiering model. In this case the constraint of demand governs the model and this leads to a relatively low storage accompanied with small releases and thus the result are the same if either hydropower or *SQDV* is selected as the only priority.
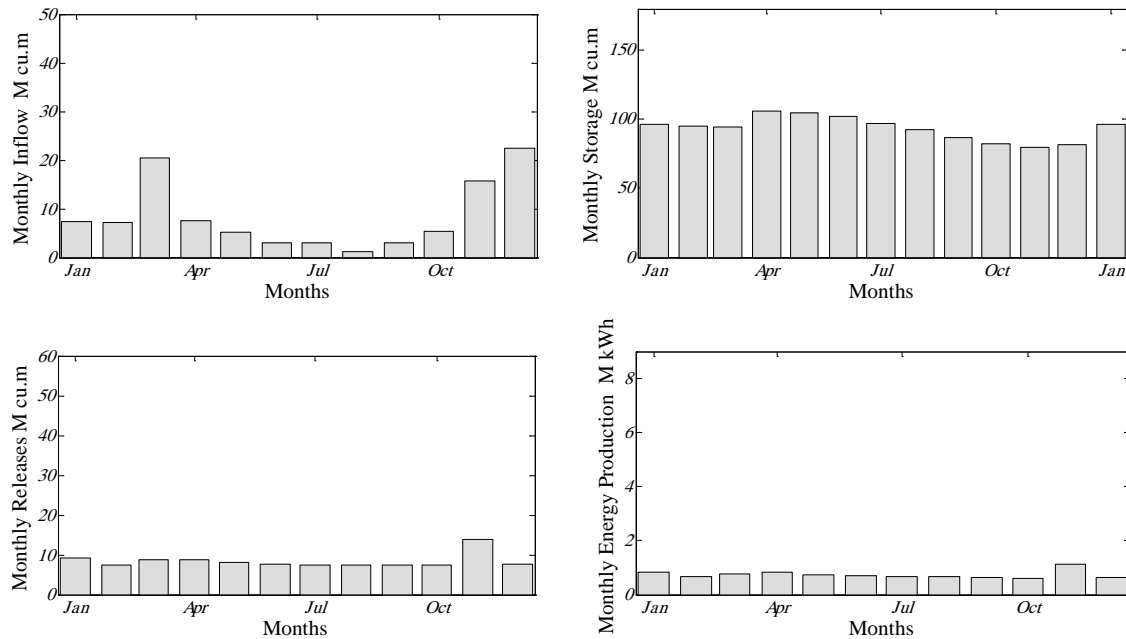
Figure 6.4: Optimal release policy - hydropower is the only priority (Scenario 15)
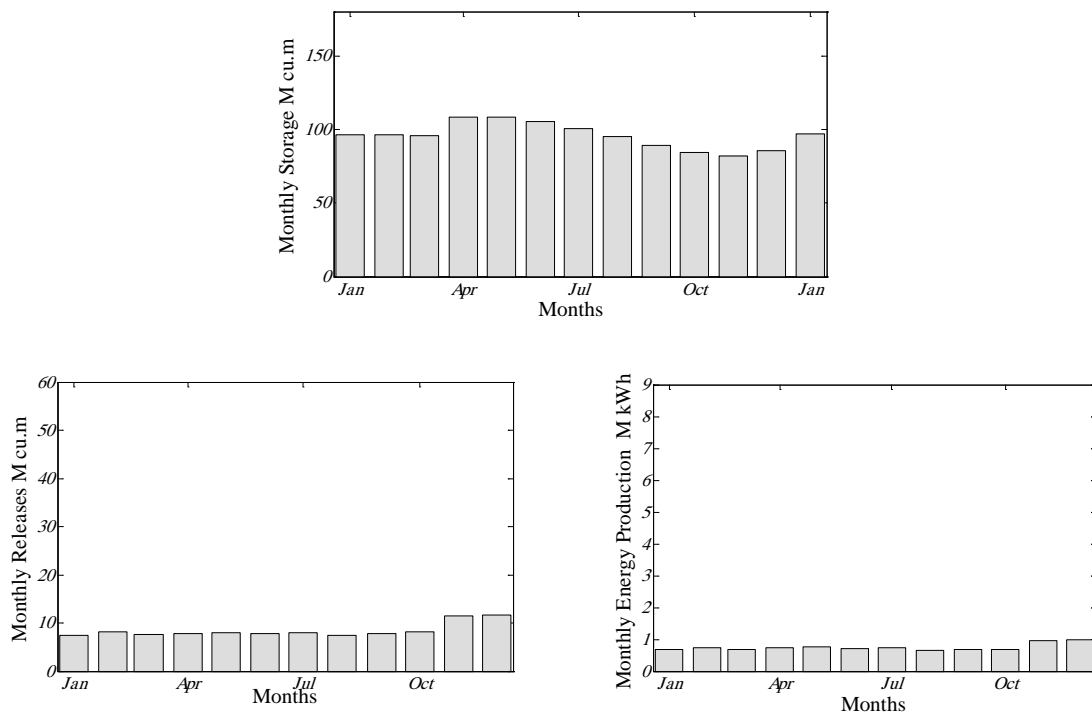
Figure 6.5: Optimal release policy - *SQDV* is the only priority (Scenario 15)

6.2.3.1 Comparison between the Results of the Developed Model and Actual Historical Data

To evaluate the output of the developed model, the driest year (the year with minimum summation of monthly inflow) in the available historical data series has been detected then the set of output of the optimization model was compared with those of historical data. The calendar year 1996 is the driest year in the available historical records with annual inflow 155.3 M.m^3, while for the reservoir Bigge, the mean annual inflow is 240.51 M.m^3 (figure 6.6).
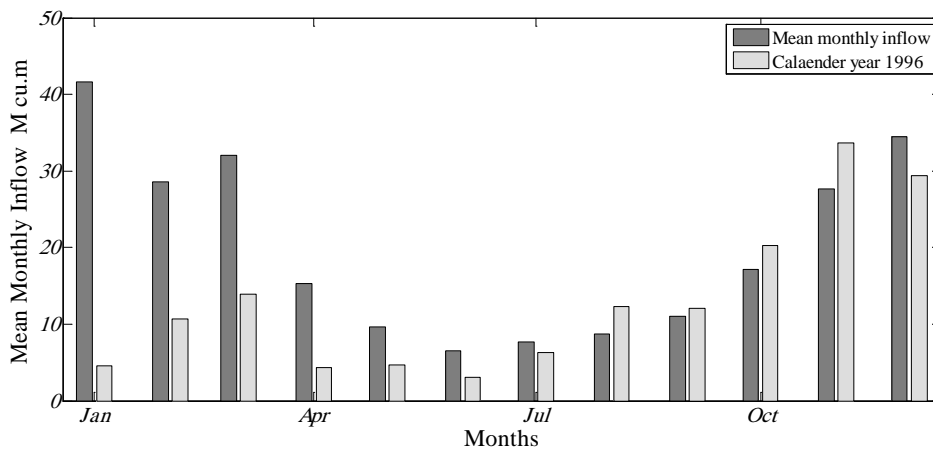
Figure 6.6: Comparison between mean monthly inflow and the monthly inflow of the calendar year 1996

For the historical calendar year 1996, inflow, storage of reservoir and releases are known variables. Using these data the monthly energy produced in MkWh for this year has been calculated. Then the historical inflow of the year 1996 and the initial storage was taken as an input for the optimization model. Both of monthly target release from the reservoir and target storage has been estimated based on the monthly demand to be met from the reservoir. The developed model in this study is designed so that demands in dry period are reduced up to 70 % of mean monthly demands and always there will be a 3 months of demands reserve in the reservoir in additional to the minimum storage required to meet water quality requirements downstream. For this inflow scenario, if the hydropower is selected as the only priority, then the optimal release polices obtained are shown in figure 6.7.

The optimal annual hydropower obtained using the optimization model is 15.36 MkWh, however the actual annual energy produced in this year is 13.24 MkWh. If the *SQDV* is selected as the only priority, then the annual hydropower is 13.67 MkWh.
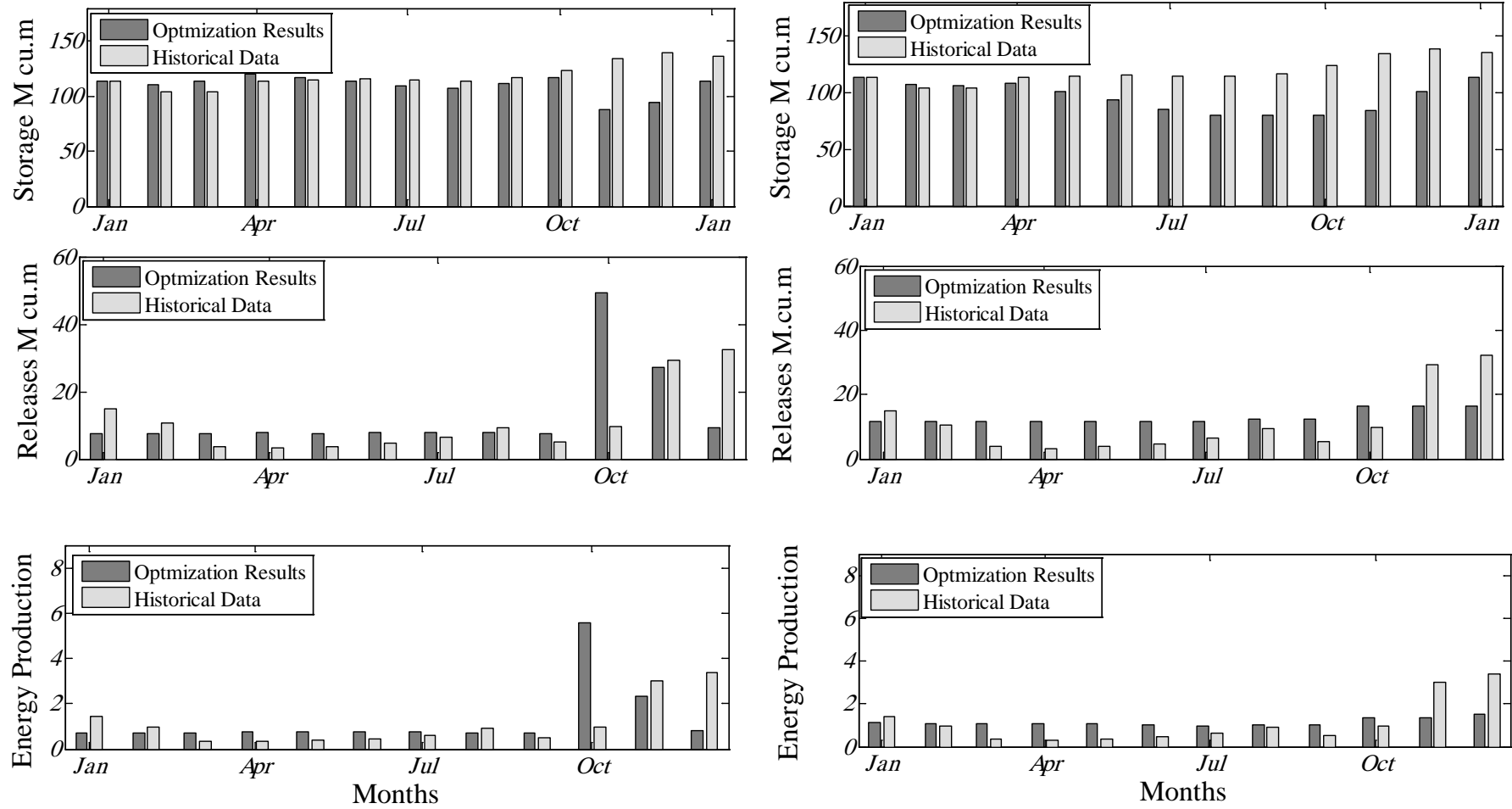
(a) hydropower is given priority   (b) *SQDV* is given priority

Figure 6.7: Comparison between results of the optimization model and historical records (Scenario 13 – calendar year 1996)

It can be clearly observed from figure 6.7 that for the developed model, the storage at the end of optimization period is the same as the initial storage however the historical data does not satisfy this constraint. The constraint related to the final storage was modified so that the final storage in the model equals this one in the historical data to investigate the effect of the final storage at the end of optimization period on results and also to present accurate comparison between the developed model and historical data. The optimal annual hydropower obtained using the modified model is 12.80 MkWh. But still there is a major difference between the modified model and historical data, which is the minimum monthly release. As shown in figure 6.7, historical releases during several months are less than 4 M.m^3, however in the developed model the allowable minimum release in case of dry periods is 7.5 M.m^3. After modifying this constraint in the model, the hydropower obtained using the modified model is 13.41 MkWh (figure 6.8).
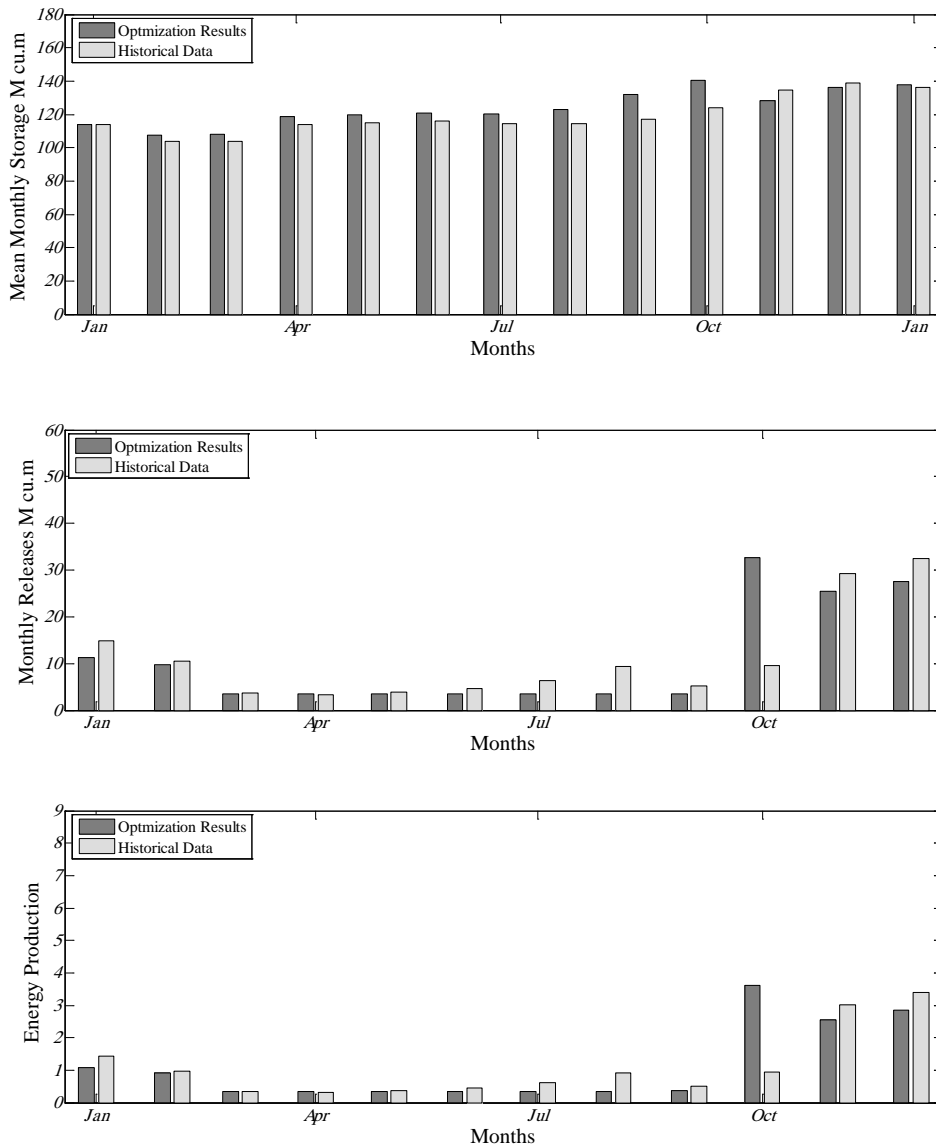
Figure 6.8: Comparison between results of the optimization model and historical records (scenario 13 after modification of final storage constraint)

6.2.3.2 Comparison between Alternative Optimization Methods

Optimization results presented in previous sections obtained using GA approach. In this section another two methods are performed, namely pattern search and gradient-based optimization. The two models are subject to the same constraints and optimize the same objective function with the same possibilities. Results of GA, pattern search and gradient-based optimization have been compared to those of historical data as shown in figures 6.9 and table 6.1. Results indicate that, when the priority is for the hydropower, the optimal annual hydropower E in case of using GA equals 15.36 MkWh, in case of using pattern search equals 16.58 MkWh and in case of using gradient-based optimization equals 13.81 MkWh
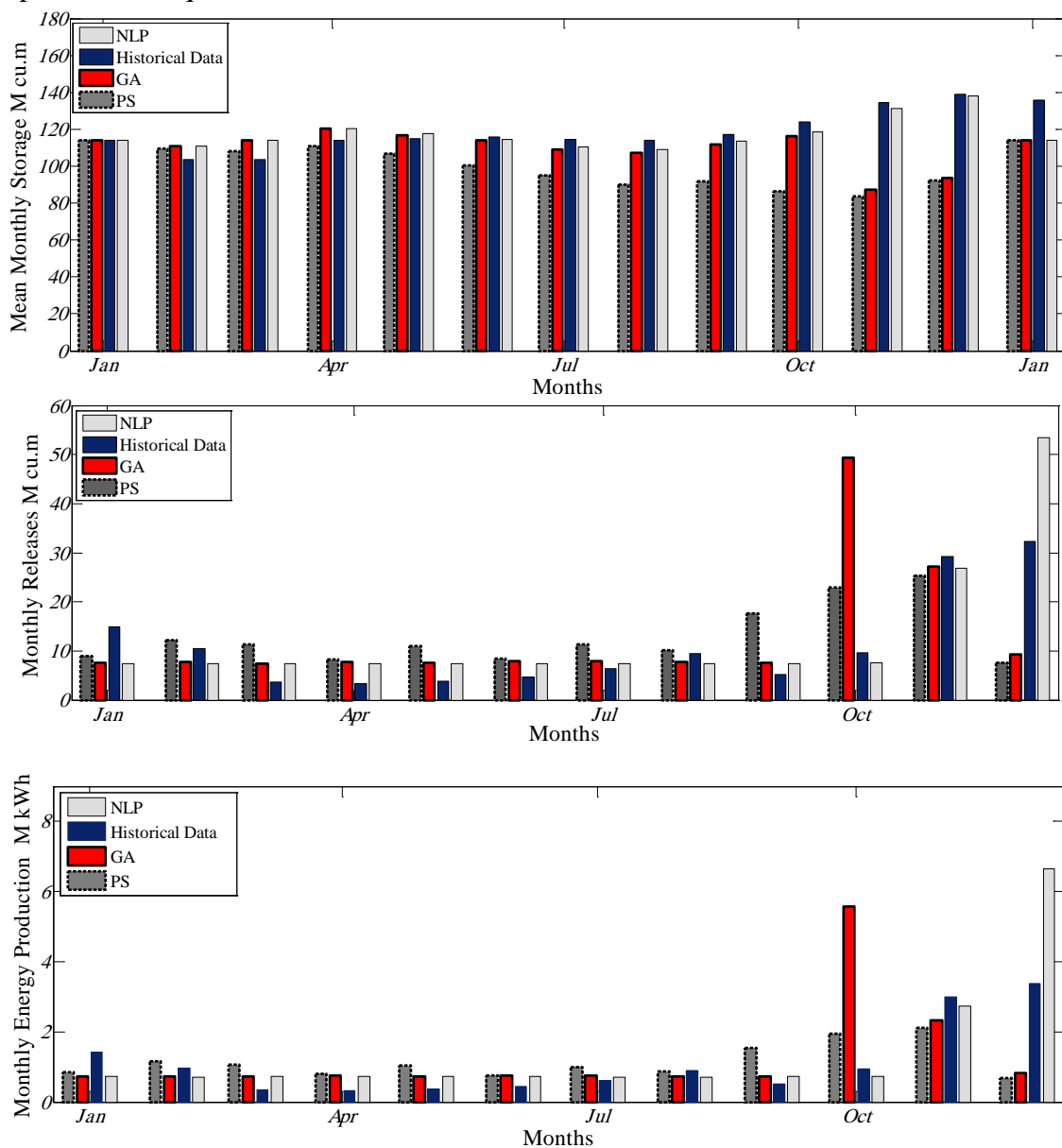
Figure 6.9: Comparison between Alternative Optimization Methods (scenario 13)

Table 6.2: Comparison between Alternative Optimization Methods (scenario 13)

	Storage				Release				Energy Production M kWh			
	GA	Gradient-based	Pattern search	Historical	GA	Gradient-based	Pattern search	Historical	GA	Gradient-based	Pattern search	Historical
January	113.89	113.89	113.89	113.89	7.63	7.50	8.89	14.85	0.73	0.72	0.85	1.43
February	110.87	111.00	109.61	103.65	7.67	7.50	12.14	10.55	0.73	0.71	1.15	0.97
March	113.83	114.13	108.10	103.72	7.50	7.50	11.41	3.70	0.72	0.72	1.07	0.34
April	120.30	120.60	110.66	114.00	7.79	7.50	8.35	3.27	0.76	0.74	0.79	0.31
May	116.89	117.49	106.69	115.11	7.59	7.50	11.05	3.81	0.74	0.73	1.03	0.37
June	113.93	114.61	100.27	115.92	7.92	7.50	8.39	4.65	0.76	0.72	0.76	0.45
July	109.09	110.19	94.96	114.36	8.00	7.50	11.33	6.44	0.75	0.71	1.00	0.62
August	107.36	108.97	89.90	114.20	7.74	7.50	10.13	9.43	0.73	0.71	0.87	0.91
September	111.88	113.73	92.03	117.04	7.58	7.50	17.62	5.26	0.72	0.72	1.54	0.51
October	116.42	118.35	86.53	123.89	49.36	7.50	22.98	9.57	5.58	0.73	1.95	0.95
November	87.36	131.15	83.85	134.62	27.26	26.85	25.36	29.23	2.32	2.72	2.12	3.00
December	93.81	138.00	92.20	139.10	9.29	53.47	7.67	32.38	0.82	6.65	0.67	3.39
Total					155.33	155.33	155.33	133.14	15.36	16.58	13.81	13.24

6.3 Conclusion

Reservoir operation for an optimum use of available water during prolonged periods of drought has always been a primary concern for water management. Using Genetic Algorithm, Pattern Search and Gradient-based method, optimization model has been developed for the operation of reservoir during normal periods and drought periods as well. The reservoir Bigge has been presented as the case study . Two objective functions have been considered then a weighted approach has been adopted to convert the multiple objectives problem into a single objective problem so that user can specify the priorities by giving a specified weight for each function. Several scenarios for low inflow period have been attempted. Each scenario has its assumptions for monthly inflow and monthly demand.

The optimization model developed in this study has been carried out using Genetic Algorithm and Direct Search toolbox and Optimization Toolboxes in MATLAB software. The obtained results showed that both of GA approach and Gradient-based approach provides higher benefits more than Pattern Search approach. Evaluation of the developed model has been carried out using the driest year in the available historical records. The monthly inflow of this year has been considered as an input to the optimization model. Results of evaluation demonstrated that the developed model with its several scenarios and the suggested optimization approaches could be a helpful guide for the real operation of the reservoir during drought events.

Chapter 7

Stochastic Simulation of Reservoir Operation Using Adaptive Neuro-Fuzzy Inference Systems

7.1 Background

In reservoir management practices, a simulation model can be used as a valuable planning tool to evaluate the impact of changes to the system's configuration or operational objectives. The desired generation or release scheduling can be checked using inflow forecasting in order to satisfy the entire set of operational constraints (Cicogna et al., 2009). At the real time operation stage, a simulation tool can be used to quickly check operational alternatives due to emergency events or planning and real-time incongruence (Cicogna et al., 2009).

McMahon (McMahon, 2009) reported that operational models have been broadly categorized as descriptive simulation, prescriptive optimization and hybrid simulation /optimization models involving elements of both. These categories can be classified as follows:

i.   Descriptive models which are used to simulate reservoir release decisions following predefined logical "if-then-else" operating rules, driven by input hydrologic data and subject to multiple constraints,

ii. Prescriptive optimization models employ mathematical programming techniques to solve for decision variables which maximize or minimize the value of an objective function which is subject to multiple constraints.

iii. Hybrid models which are primarily descriptive simulation models with piecewise optimization of specific aspects of predefined operating rules.

Each type of the described model has strengths and weaknesses with respect to specific operational planning and real-time water control applications. Descriptive simulation models are most useful for detailed analysis and evaluation of predefined operating rules. Several approaches that use fuzzy set theory to simulate reservoir operation have been described in the literature. These include fuzzy optimization techniques, fuzzy rule base systems, and combinations of the fuzzy approach with other techniques (Dubrovin et al., 2002) .

The fuzzy logic approach may provide a promising alternative to the methods used for reservoir operation modeling because the approach is more flexible and allows incorporation of expert opinions, which could make it more acceptable to operators (Panigrahi and Mujumdar, 2000). Applications can be found in the work of Chuntian (Chuntian, 1999), Panigrahi (Panigrahi and Mujumdar, 2000), and Shrestha (Shrestha et al., 1996). The fuzzy rule base could be constructed on the basis of expert knowledge or observed data. Approaches for deriving a rule base from observed data have been presented by Mohan (S.Mohan and Prasad, 2006), and Panigrahi (Panigrahi and Mujumdar, 2000).

Fuzzy Logic was initiated in 1965 by Lotfi A. Zadeh, professor for computer science at the University of California in Berkeley (Zadeh, 1973). Basically, Fuzzy Logic (FL) is a multivalued logic that allows intermediate values to be defined between conventional evaluations like true/false, yes/no, high/low, etc. Fuzzy logic variables may have a truth value that ranges between 0 and 1 and is not constrained to the two truth values of classic propositional logic notions like rather tall or very fast can be formulated mathematically and processed by computers, in order to apply a more human-like way of thinking in the programming of computers (Zadeh, 1973).

Fuzzy logic models, called fuzzy inference systems, consist of a number of conditional "if-then" rules. For the designer who understands the system, these rules are easy to write, and as many rules as necessary can be supplied to describe the system adequately. In fuzzy logic technique, unlike standard conditional logic, the truth of any statement is a matter of degree (Metaxiotis et al., 2003). Fuzzy inference systems rely on membership functions to explain to the computer how to calculate the correct value between 0 and 1. The degree to which any fuzzy statement is true is denoted by a value between 0 and 1.

The advantages of fuzzy logic are that calculation is straightforward and the model easy for the operator to understand due to its structure, which is based on human thinking. The system can also be easily modified when necessary (Dubrovin et al., 2002). The fuzzy rule based system utilizes the knowledge of a reservoir operator and avoids complex optimization procedure hence it may be more acceptable to the reservoir managers (S.Mohan and Prasad, 2006). The present study is aimed to present a new approach for long-term and short-term reservoir operation based on Artificial Neuro-Fuzzy Inference Systems (ANFIS). Different models have been developed for simulation of reservoirs operation. The procedure is illustrated through a case study of the system of Ruhr reservoirs in Germany.

7.2 Fundamental Fuzzy System for Reservoir Operation Model

In modeling of reservoir operation with fuzzy logic, the following distinct steps are followed (Panigrahi and Mujumdar, 2000 ; Shah, 2009):

- ➢ Selection and fuzzification of inputs, where the crisp inputs such as the inflow, reservoir storage and release are transformed into fuzzy variables,

- ➢ Fuzziness procedure and formulation of the fuzzy rule set, based on an expert knowledge base,

- ➢ Application of a fuzzy operator, to obtain one number representing the premise of each rule,

- ➢ Shaping of the consequence of the rule by implication,

- ➢ De-fuzzificationprocedure.

Details about these steps can be found in (Dubrovin et al., 2002; Panigrahi and Mujumdar, 2000; S.Mohan and Prasad, 2006; Shrestha et al., 1996).

7.3 Adaptive Neuro-Fuzzy Inference System

In recent years, there has been a growing trend in the use of fuzzy logic in combination with neuro-computing and genetic algorithms in many of the industrial and research applications because of their ability to deal with ill-posed and uncertain systems (Yegireddi and Kumar, 2008). An adaptive neuro-fuzzy inference system (ANFIS) is a fuzzy inference system formulated as a feed-forward neural network. Hence, the advantages of a fuzzy system can be combined with a learning algorithm (Venugopal et al., 2010).

Fuzzy systems present particular problems to a developer then rules have to be determined somehow. This is usually done by 'knowledge acquisition' from an expert. It is a time consuming process that is weighed down by many problems. A fuzzy set is fully determined by its membership function (Kablan, 2009). This has to be determined, for example if it is Gaussian then what are the parameters. The ANFIS approach learns the rules and membership functions from data. ANFIS is an adaptive network of nodes and directional links with associated learning rules. It is called adaptive because some, or all, of the nodes have parameters which affect the output of the node. These networks identify and learn relationships between inputs and has been the adaptive network of choice to be investigated in detail and used for high frequency forecasting and trading due to its high learning capability and membership function definition properties (Kablan, 2009).

The Sugeno model makes use of if rules to produce an output for each rule. It is similar to the Mamdani method in many respects. The first two parts of the fuzzy inference process, fuzzifying the inputs and applying the fuzzy operator, are exactly the same. The main difference between Mamdani and Sugeno is that in the Sugeno type rule outputs consist of the linear combination of the input variables plus a constant term; the final output is the weighted average of each rule's output. Adaptive neuro-fuzzy inference system mimics the operation of a Takagi–Sugeno–Kang (TSK) fuzzy system (Tang et al., 2005).

Fuzzy inference systems are composed of five functional blocks as given in Figure 7.1 as shown in the figure, the ANFIS model contains (Venugopal et al., 2010):

1. A rule base containing a number of if-then rules,

2. A database which defines the membership function,

3. A decision making interface that operates the given rules,

4. A fuzzification interface that converts the crisp inputs into "degree of match "with the linguistic values like high or low etc.,

5. A de fuzzification interface that reconverts to a crisp output.

The rule base in the Sugeno model has of the form:

$$If\ x\ is\ A_1\ and\ y\ is\ B_1\ then\ f_1 = p_1 * x + q_1 * y + r_1 \qquad (7.1)$$

$$If\ x\ is\ A_2\ and\ y\ is\ B_2\ then\ f_2 = p_2 * x + q_2 * y + r_2 \qquad (7.2)$$

Where x and y are predefined membership functions, A_i and B_i are membership values, p_i, q_i, and r_i are the consequent parameters that are updated in the forward pass in the learning algorithm, and f_i is the outputs within the fuzzy region specified by the fuzzy rule.
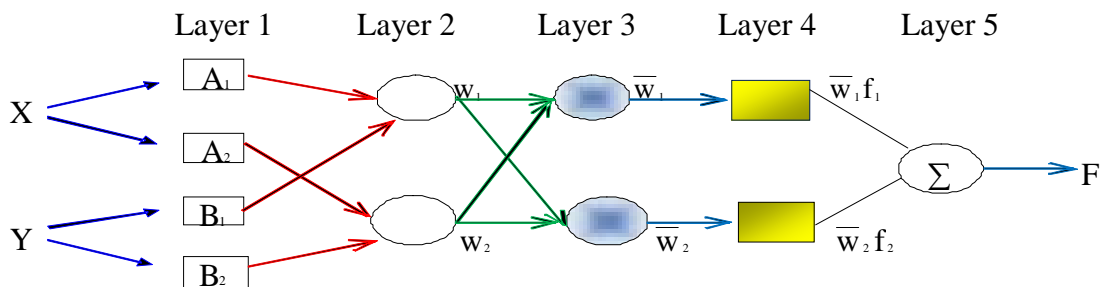
Figure 7.1: An ANFIS architecture for a two rule Sugeno system

Let the membership functions of fuzzy sets A_i and B_j, be μ_{A_i} and μ_{B_i} respectively. The five layers that integrate ANFIS are as follow:

Let the output of the ith node in layer l is denoted as $O_{1,i}$ then,

Layer 1: Every node i in this layer is an adaptive node with node function

$$Q_{1,i} = \mu_{A_i}(x) \quad \text{for i = 1, 2, or} \quad Q_{1,i} = \mu_{B_{i-2}}(y) \quad \text{for i = 3, 4} \tag{7.3}$$

where x (or y) is the input to the ith node and A_i (or B_{i-2}) is a linguistic labels.

Layer 2: This layer consists of the nodes labeled which multiply incoming signals and send the product out. Each node output represents the firing strength of a rule.

$$O_{2,i} = w_i = \mu_{A_i}(x)\,\mu_{B_i}(y) \quad \textit{for i = 1, 2} \tag{7.4}$$

Layer 3: In this layer, the nodes labeled N acts to scale the firing strengths to provide normalized firing strengths.

$$O_{3i} = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \; i = 1,2 \tag{7.5}$$

Layer 4: The output of layer 4 is comprised of linear combination of inputs multiplied by normalized firing strengths. This layer's nodes are adaptive with node functions.

$$O_{4i} = w_i\, f_i = w_i\,(p_i\, x + q_i\, y + r_i) \tag{7.6}$$

Where, w_i is the output of layer 3, and $\{p_i, q_i,\ r_i\}$ are the parameter set. Parameters of this layer are referred to as consequent parameters.

Layer 5: This layer consists of a single node, computes the final output as the summation of all incoming signals

$$O_{5i} = \sum_{i=1} \overline{w}_i\, f_i = \frac{\sum_{i=1} w_i f_i}{\sum_{i=1} w_i} \tag{7.7}$$

Layers represented by squares are adaptive and their values are adjusted when carrying out the system training. Layers represented by circles remain invariable before, during and after the training (Kablan, 2009).

7.4 Simulation of Reservoir Operation Using Adaptive Neuro-Fuzzy Inference Systems –ANFIS

7.4.1 Data Used in this Study

The methodology discussed in sections 7.3 and 7.4 has been used for modeling of operation of the system of Ruhr reservoirs. A database having monthly inflow, storage, release series starting at 1991 has been established. The time series of inflow present the inflow to the main reservoirs in the Ruhr river basin namely, Bigge reservoir, Moehne reservoir, Henne reservoir and Sorpe reservoir. Source of data is the Ruhrverband (Ruhr River Association). All time series were checked to find out all missing data. Table 7.1 presents a typical data sample of one year of the data which have been used in this study.

Table 7.1: Typical data sample for one year of used data

Year	Month	Inflow	Storage	Release
1990	11	38.27	102.12	11.73
1990	12	32.53	128.66	15.71
1991	1	38.94	145.47	52.33
1991	2	6.64	132.08	10.59
1991	3	16.99	128.12	6.33
1991	4	5.54	138.78	8.86
1991	5	5.85	135.46	7.88
1991	6	11.062	133.43	8.33
1991	7	7.250	136.15	11.96
1991	8	2.57	131.44	16.80
1991	9	2.15	117.21	16.15
1991	10	4.40	103.20	12.24

7.4.2 Methodology

In reservoir operation, a direct method for making a decision is to look at the historical data for similar cases and make a decision similar to the decision that was mad in those cases. One of the main operational goals in the management of reservoirs is to

determine a suitable release based on observation data and other conditions. Applications of fuzzy logic system which presented in the literature used storage and inflow as input to the fuzzy system and the output make a similar decision is the release during the same period.

In this study a new approach using Adaptive neuro-fuzyy inference system approach "ANFIS" has been applied. ANFIS has been used to extract the relation of time of year (months), storage, inflow, and Standardized Precipitation Index (SPI) and release variables and represent them as fuzzy if-then rules. The premise part of fuzzy if-then rules is months, inflow, storage, and SPI. The consequent part is the release. The SPI has been calculated based on streamflow data series which means that the drought index from the streamflow series has been used as one of measures for streamflow deficit. The distinctive feature of this method is that the drought management and monitoring would be effective because of the more realistic judgment on the drought severity (Yoo et al.).

Also in this study two main models for the simulation of reservoir operation have been developed using ANFIS. Each main model contains a set of sub-models. The set of input into the two main models contains time of year, storage, inflow, and Standardized Precipitation Index (SPI) with Alternative arrangement. The output of the first model is the release during the next month, which could be a helpful reference guide to the operator during dealing with decisions. On the other hand, the output of the second model is the release of the current month which could be a good tool for the evaluation of release for a specified month.

Each model from the two main models consists of two stages. At the first stage, operation rules are developed using fuzzy approach, then the developed fuzzy inference system "FIS" is an input to the ANFIS system. ANFIS uses a hybrid learning algorithm to identify parameters of Sugeno-type fuzzy inference systems. It applies a combination of the least-squares method and the back propagation gradient descent method for training FIS membership function parameters to emulate a given training data set.

At second stage, the operation of reservoirs is simulated for any required number of years using the final FIS developed by using ANFIS. Thomas-Fiering model is used to generate monthly inflow, and a Markov model is developed to forecast SPI index. The entire methodology applied in this study is shown in figure 7.2.
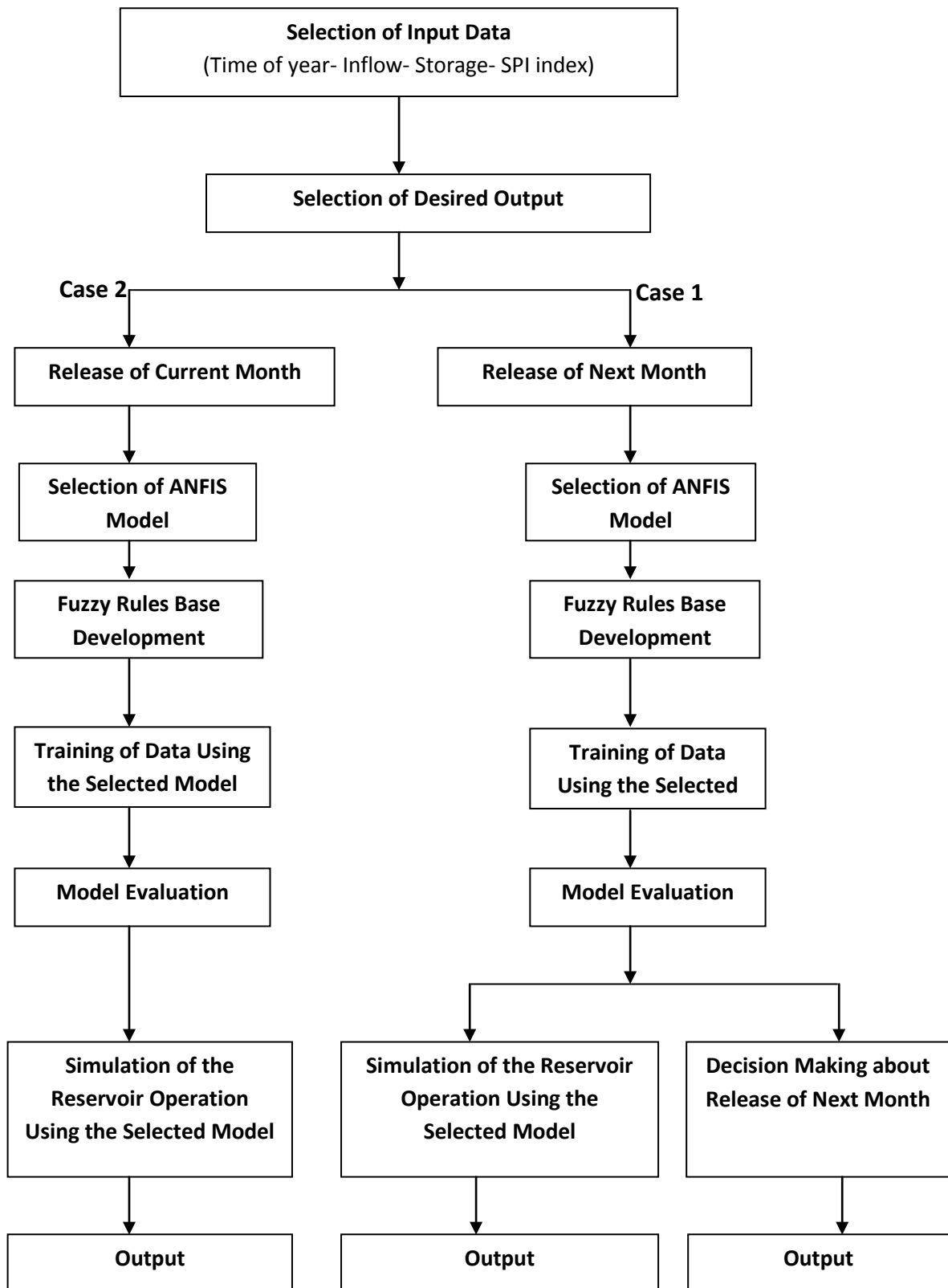
Figure 7.2: Flow diagram of ANFIS model that has been developed for reservoir operation and simulation

7.4.3 Modeling of Reservoir Operation – Case1: Release of next Month

The goal of this section is to describe and analyze the main models presented in figure 7.2. The first case is the case of release of the next month. In this case a main model is developed for reservoir operation and this model has been applied to the Ruhr reservoirs system. In order to illustrate the methodology, an application to Bigge reservoir will be presented in following sub sections.

7.4.3.1 Selection of Input Data

Input data consists basically of four variables, namely time of year (month), inflow into reservoir, storage volume, and the SPI index. As mentioned in section 7.4.2, in this study the SPI index is calculated based on streamflow data series. More details about the methodology of SPI index calculation has been presented in chapter 3. SPI index is calculated for different time scales (3, 6, 9, and 12 months). At this stage, Selection of Input Data, the SPI time scale is identified before going to the next stage. A sample of input data for Bigge reservoir is shown in figure 7.3.

7.4.3.2 Selection of ANFIS Model

After selection of SPI time scale, alternative sub-models have been identified by considering the input data series. In this study a set of six models for each SPI time scale has been developed as shown in table 7.2. This indicates that 24 models have been actually tested. All the combination have been tried to determine the best model out of these candidate models. In table 7.2, the letter I is an abbreviation for inflow, and S for storage volume. For illustration, a typical sample of input/output data of the model M_3 is shown in table 7.3.

Table 7.2: Description of the input of ANFIS-based learning models
Case1: Release of next month

Model	Input data of the selected model						
M_1	Month(i)	I(i)	S(i)	SPI(i)	SPI(i+1)		
M_2	Month(i)	I(i)	S(i)	SPI(i-1)	SPI(i)	SPI(i+1)	
M_3	Month(i)	I(i)	S(i-1)	S(i)	S(i+1)	SPI(i-1)	SPI(i)	SPI(i+1)
M_4	Month(i)	I(i)	S(i-1)	S(i)	SPI(i-1)	SPI(i)	SPI(i+1)	
M_5	Month(i)	I(i)	S(i+1)	SPI(i-1)	SPI(i)	SPI(i+1)	
M_6	Month(i)	I(i)	S(i+1)	SPI(i)	SPI(i+1)	SPI(i+2)	
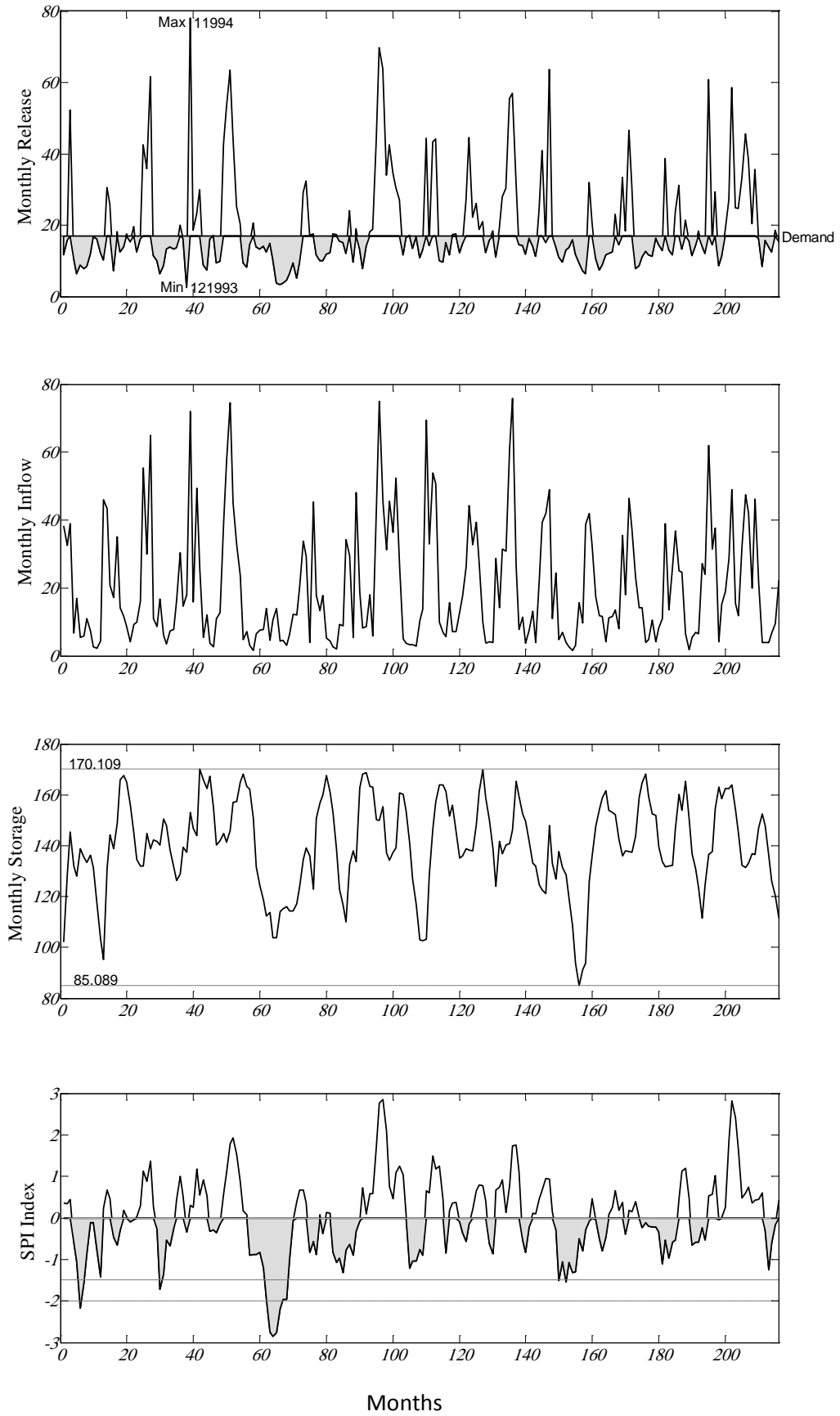
Figure 7.3: Sample of input data

Table 7.3: Typical sample of input and output data of the model M_3

Input								Output
Month (i)	Inflow (i)	Storage (i-1)	Storage (i)	Storage (i+1)	SPI (i-1)	SPI (i)	SPI (i+1)	Release (i+1)
11	46.05	103.21	95.36	131.16	-1.43	0.23	0.67	30.51
12	43.48	95.36	131.16	144.13	0.23	0.67	0.46	25.99
1	20.83	131.16	144.13	138.97	0.67	0.46	-0.47	7.27
2	17.17	144.13	138.97	148.87	0.46	-0.47	-0.66	18.23
3	35.20	138.97	148.87	165.84	-0.47	-0.66	-0.30	12.53
4	14.19	148.87	165.84	167.50	-0.66	-0.30	0.18	14.00
5	11.64	165.84	167.50	165.14	-0.30	0.18	0.02	17.60
6	8.60	167.50	165.14	156.15	0.18	0.02	-0.09	15.39
7	4.19	165.14	156.15	144.95	0.02	-0.09	-0.04	19.59
8	9.20	156.15	144.95	134.56	-0.09	-0.04	0.02	12.47
9	9.99	144.95	134.56	132.07	-0.04	0.02	0.29	16.11
10	16.17	134.56	132.07	132.13	0.02	0.29	1.12	42.62
11	55.37	132.07	132.13	144.88	0.29	1.12	0.88	35.92

7.4.3.3 Fuzzification of Inputs and ANFIS-Based Learning Models

The degree to which a particular measurement of inflow or storage is high, low or medium depends on how the fuzzy sets of high inflow/low storage are defined. This definition may arise from statistical data or neural clustering of historical data. In order to begin the training using ANFIS, an initial fuzzy inference system "FIS" is needed first. In the present study 42 years of historical data of inflow and 18 years of historical data of storage and release have been collected. From this data, 14 years of data have been used for building (training) the model and 4 years of data have been used to test the model on monthly basis. The long time series of monthly inflow has been used to perform the calculation of the standardized SPI because the calculation of SPI index should have at least 30 years of historical data (see chapter 3).

As shown in table 7.3, we have a set of input data and one output (reservoir release). FIS, fuzzy inference system, has been generated using fuzzy subtractive clustering to develop a set of rules and membership functions that models the data behavior. Then the generated FIS has been used as an initial FIS, initial conditions, for ANFIS training. The FIS has been then evaluated to obtained output data which is the predicted value of the release for the particular model. Figure 7.4 presents the developed FIS system using ANFIS system. Forecasted release values and observed release values for training period and test period are shown in figure 7.5, 7.6 respectively.
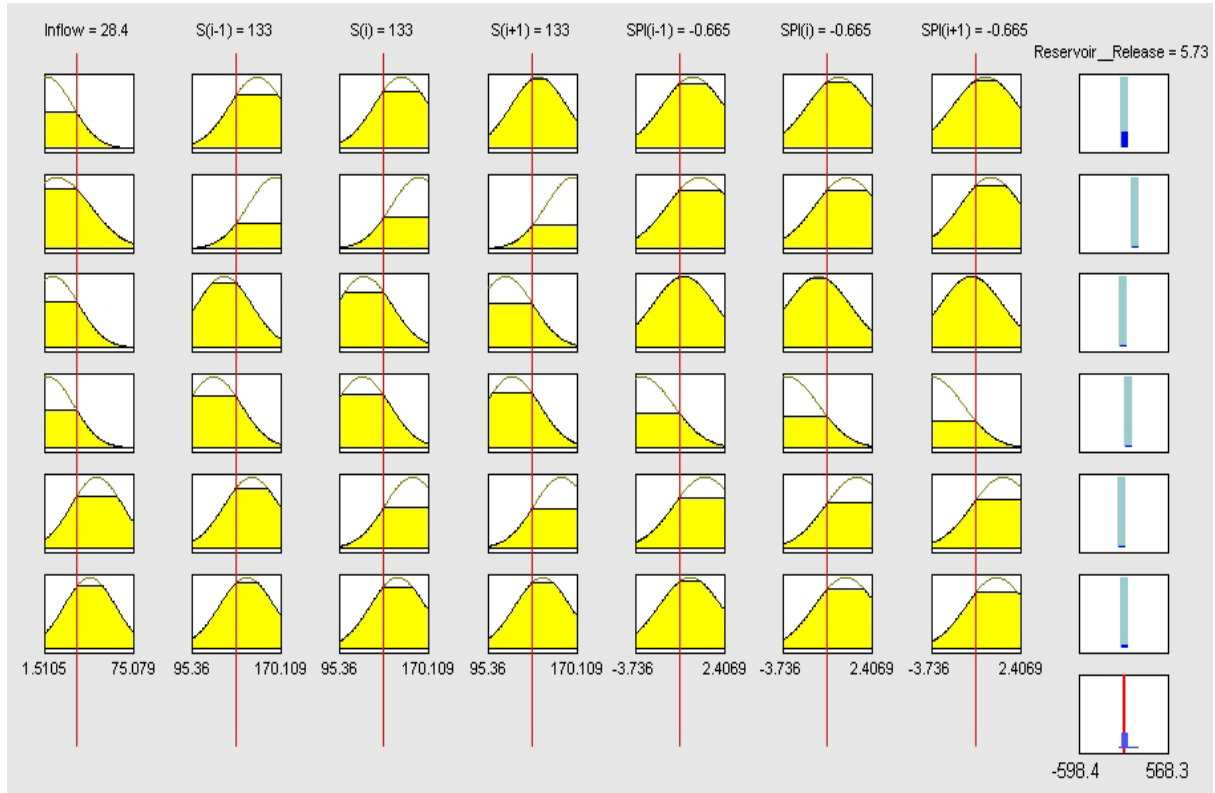
Figure 7.4: Fuzzy inference system "FIS" developed using ANFIS
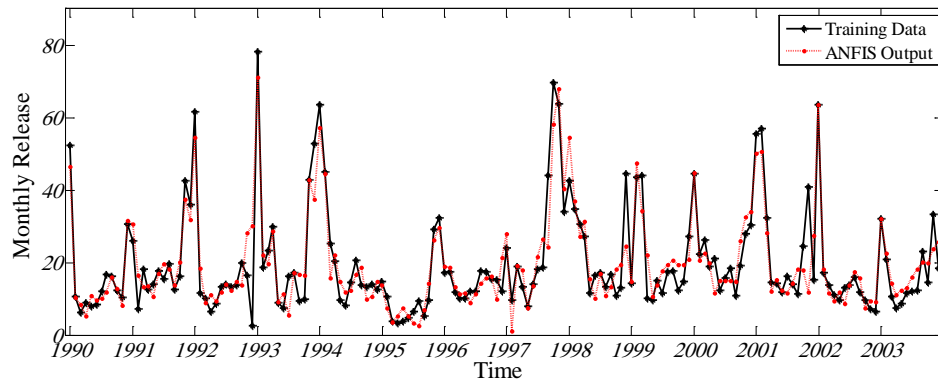
Figure 7.5: ANFIS output for reservoir release (training period).
Model M_3 / SPI_9
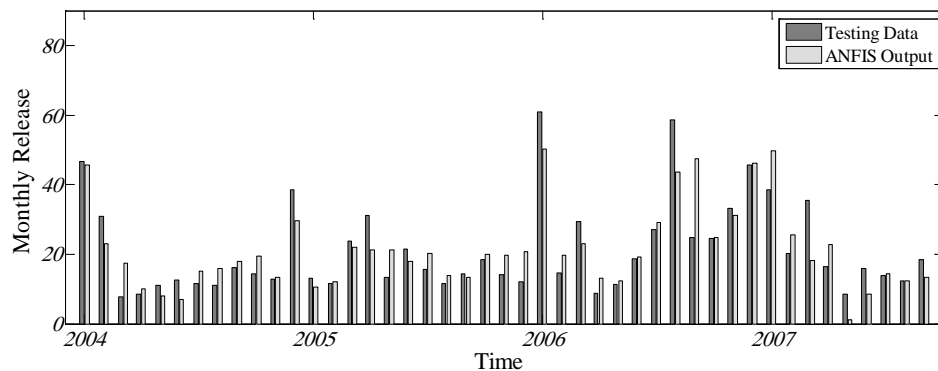
Figure 7.6: ANFIS output for reservoir release (test period)
Model M_3 / SPI_9

7.4.3.4 Model Evaluation

In order to evaluate and compare the forecasting performance of the ANFIS system, it is necessary to introduce forecasting evaluation criteria. In this study, four criteria include; Mean Absolute Deviations (*MAD*), R-squared, Root Mean Square Error and correlation coefficient have been used.

  i.   Mean Absolute Deviations (*MAD*)

The *MAD* measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures accuracy for continuous variables. Expressed the MAD is calculated as follow:

$$MAD = \frac{\sum_{i=1}^{n}|Ro_i - Rf_i|}{n} \tag{7.8}$$

Where Ro and Rf are the observed and forecasted reservoir releases, and n is the number of observations.

ii.   R-squared

In statistics, the coefficient of determination, R^2 is used in the context of statistical models whose main purpose is the prediction of future outcomes on the basis of other related information. The absolute fraction of variance, R^2, is calculated as follow:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Ro_i - Rf_i)^2}{\sum_{i=1}^{n}(Ro_i)^2} \qquad (7.9)$$

iii.   Root mean squared error (RMSE)

The *RMSE* is the square root of the variance of the residuals. It indicates the absolute fit of the model to the data–how close the observed data points are to the model's predicted values. Whereas R-squared is a relative measure of fit, *RMSE* is an absolute measure of fit. Lower values of *RMSE* indicate better fit. *RMSE* is calculated as follow:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Ro_i - Rf_i)^2}{n}} \qquad (7.10)$$

iv.   Correlation coefficient (Cr)

The correlation coefficient a concept from statistics is a measure of how well trends in the forecasted values follow trends in past actual values (historical releases). The correlation coefficient is calculated as follow:

$$C_r = \frac{\sum_{i=1}^{n} Ro_i \, Rf_i - \frac{(\sum Ro_i)(Rf_i)}{n}}{\sqrt{\left[\left(\sum_{i=1}^{n} Ro_i^2 - \frac{\left(\sum_{i=1}^{n} Ro_i\right)^2}{n}\right)\left(\sum_{i=1}^{n} Rf_i^2 - \frac{\left(\sum_{i=1}^{n} Rf_i\right)^2}{n}\right)\right]}} \qquad (7.11)$$

The *MAD* and the *RMSE* can be used together to diagnose the variation in the errors in a set of forecasts. The *RMSE* will always be larger or equal to the *MAD*; the greater difference between them, the greater the variance in the individual errors in the sample. If the *RMSE = MAE*, then all the errors are of the same magnitude.

The results of model evaluation for training data sets and test data sets are summarized in Table 7.4 and 7.5 respectively. It appears that the ANFIS models are accurate and consistent in different subsets, where most of the values of *RMSE* and *MAE* are smaller, and most of correlation coefficients and R^2 are also very close to unity. Results also indicate that for each time scale of SPI index, there is one model which has a minimum *MAE* and *RMSE*, and maximum R^2 and C_r. The model which has these advantages would be more accurate. It should be noted that the ranges of historical data which have been used for model development has clear effect on the model performance. The results of models evaluation might also suggest that the ANFIS has a great ability to learn from input–output patterns. The results demonstrate that the ANFIS can be successfully applied to establish models that could provide reliable release for the selected reservoirs.

Table 7.4: Model evaluation criteria in case of release of next month (training period)

Model	SPI_3				SPI_6				SPI_9				SPI_12			
	RMSE	R^2	MAD	C_r	RMSE	R^2	MAD	C_r	RMSE	R^2	MAD	C_r	RMSE	R^2	MAD	C_r
M_1	6.52	0.93	4.47	0.90	6.27	0.993	4.41	0.89	6.17	0.94	4.15	0.90	7.5	0.91	5.1	0.83
M_2	6.24	0.93	4.43	0.89	6.28	0.934	4.56	0.90	7.5	0.91	5.16	0.92	7.68	0.90	5.19	0.85
M_3	5.17	0.95	3.92	0.93	5	0.956	3.69	0.93	5.15	0.95	3.34	0.91	5.59	0.95	4.12	0.92
M_4	6.27	0.93	4.46	0.89	6.4	0.93	4.69	0.89	7.9	0.89	5.56	0.90	7.9	0.89	5.5	0.87
M_5	5.9	0.94	4.31	0.93	5.92	0.94	4.39	0.91	8.47	0.88	6.38	0.91	8.09	0.89	6	0.82
M_6	5.24	0.95	3.81	0.93	5.59	0.95	4.19	0.92	5.82	0.94	3.98	0.91	7.88	0.89	5.76	0.89

Table 7.5: Model evaluation criteria in case of release of next month (test period)

Model	SPI_3				SPI_6				SPI_9				SPI_12			
	RMSE	R^2	MAD	C_r	RMSE	R^2	MAD	C_r	RMSE	R^2	MAD	C_r	RMSE	R^2	MAD	C_r
M_1	7.13	0.92	5.36	0.81	8.95	0.86	6.5	0.73	8.21	0.89	5.7	0.77	8.88	0.87	6.60	0.72
M_2	7.57	0.91	5.63	0.81	8.40	0.886	6.213	0.76	8.189	0.89	5.71	0.70	8.91	0.87	6.72	0.73
M_3	7.82	0.90	5.75	0.79	9.14	0.865	7.12	0.72	7.04	0.92	5.209	0.84	7.31	0.91	5.65	0.83
M_4	7.59	0.91	5.93	0.81	8.80	0.875	6.609	0.75	8.88	0.87	6.27	0.74	8.82	0.88	6.75	0.73
M_5	7.31	0.91	5.66	0.82	8.40	0.886	6.60	0.77	8.05	0.89	5.64	0.80	7.59	0.91	5.80	0.81
M_6	6.82	0.93	5.10	0.86	8.13	0.89	6.37	0.78	7.72	0.90	5.492	0.82	7.82	0.90	6.048	0.81

7.4.3.5 Simulation of the Reservoir Operation Using the Selected Model

After development and evaluation of the selected model, the FIS system could be used for simulation of reservoir operation for any required number of years (figure 7.7). In order to illustrate the mechanism of the simulation process, we will take the model M_3 as an example and the process will be illustrated step by step as follow:

➢ In the model M_3, the set of input consists of 8 variables (see table 7.3). At any month t, it is required to predict reservoir release at the next month t+1.

➢ The inflow of month t+1 is unknown, and it could be generated by several models. In this study Thomas-Fiering Model has been used as an inflow generator (see chapter 7). Thomas-Fiering model is used to generate monthly inflow for the month t+1; this means that there are two known inputs (month, and inflow).

➢ From historical data, storages of previous, current, and next month are known. As mentioned before, the storage volume is the storage at the beginning of any month so storage at months t-1, t, and t+1 are known.

➢ After holding the previous steps, three input variables are remaining unknown, namely SPI index for months t-1, t, and t+1. SPI for month t-1 is calculated from historical data, and SPI for month t is calculated based on the generated inflow from Thomas-Fiering Model. SPI index for month t+1 is predicted using transition probability matrix.

Once the input data are available, the developed FIS system predicts the release and this process could be repeated for any number of months. The simulation model of a reservoir system is based on water balance of reservoirs. The output of the model (release) must satisfy the constraints of storage and demands. The simulation model subject to the following constraints:

Storage Continuity

$$S_{t+1} = S_t + I_t - R_t - O_t \quad \text{for all } t \qquad (7.12)$$

Where,

S_t : Active reservoir storage at the beginning of period t in Mm^3;

I_t : The inflow to the reservoir during period t in Mm^3;

O_t : Overflow from the reservoir during period t in Mm^3;

R_t : Reservoir release during period t in Mm^3.

➢ Storage Limits

$$S_{min} \leq S_t \leq S_{max} \quad for\ all\ t \qquad\qquad (7.13)$$

Where, S_{min} and S_{max} are the minimum and maximum active storage of the reservoir and these storage limits depend on the month i.e. each month has its storage limits.

➢ Demands constraint

$$R_{min} < R_t \qquad\qquad (7.14)$$

Where, R_{min} is the minimum demand in a time period t and this minimum demand is identified by the reservoir operator.

Figure 7.8 presents historical data compared with the results of simulation for a period of 15 years. In order to study the behavior of the reservoir storage, a period of 100 year has been simulated using the model M_4 and SPI_3.

In order to present long records, a data series of 500 years has been simulated using model M_4 & SPI_9. This process has been executed using CPU 3.0 GHz with 1 GB of Ram, and the required time needed to perform this process was 44 hours. Table 7.6 presents a sample data of the simulated records and this sample data contains the driest year in the simulated period. Results of simulation during the period of 500 simulated years, using the proposed model, showed that the minimum reservoir storage was 64 M.m^3. During the simulated period, the reservoir storage reached values less than 85 M.m^3 903 times with a percentage of 15 % of the simulated months and the minimum release was 5.0 M.m^3. It is worth to be mentioned that, in the available historical data, the minimum reservoir storage in was 53.1 M.m^3 in month December-1976 and the minimum release from the Bigge reservoir was 0.535 M.m^3 in month April-1979.
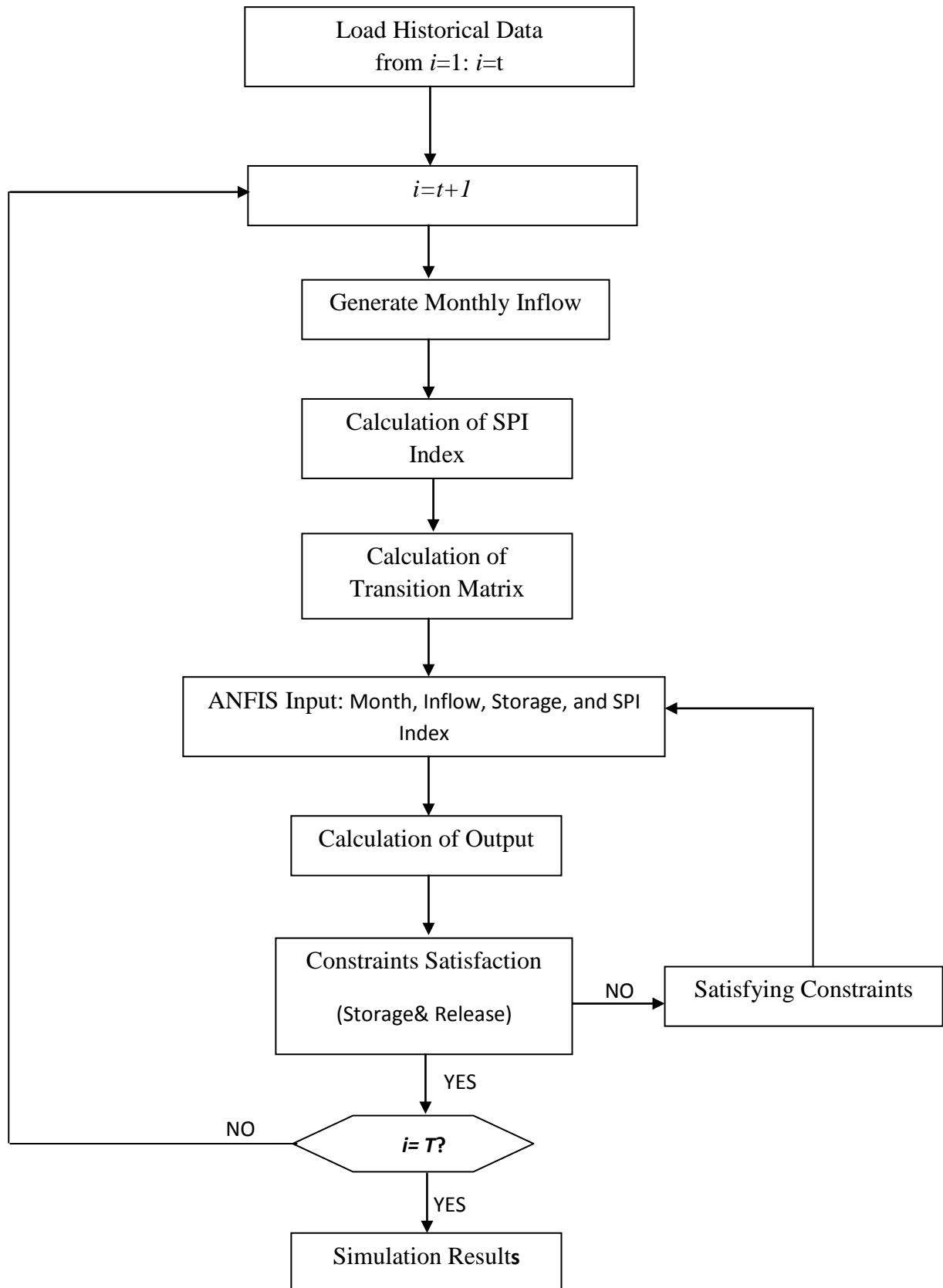
Figure 7.7: Flow diagram for the simulation of reservoir Operation
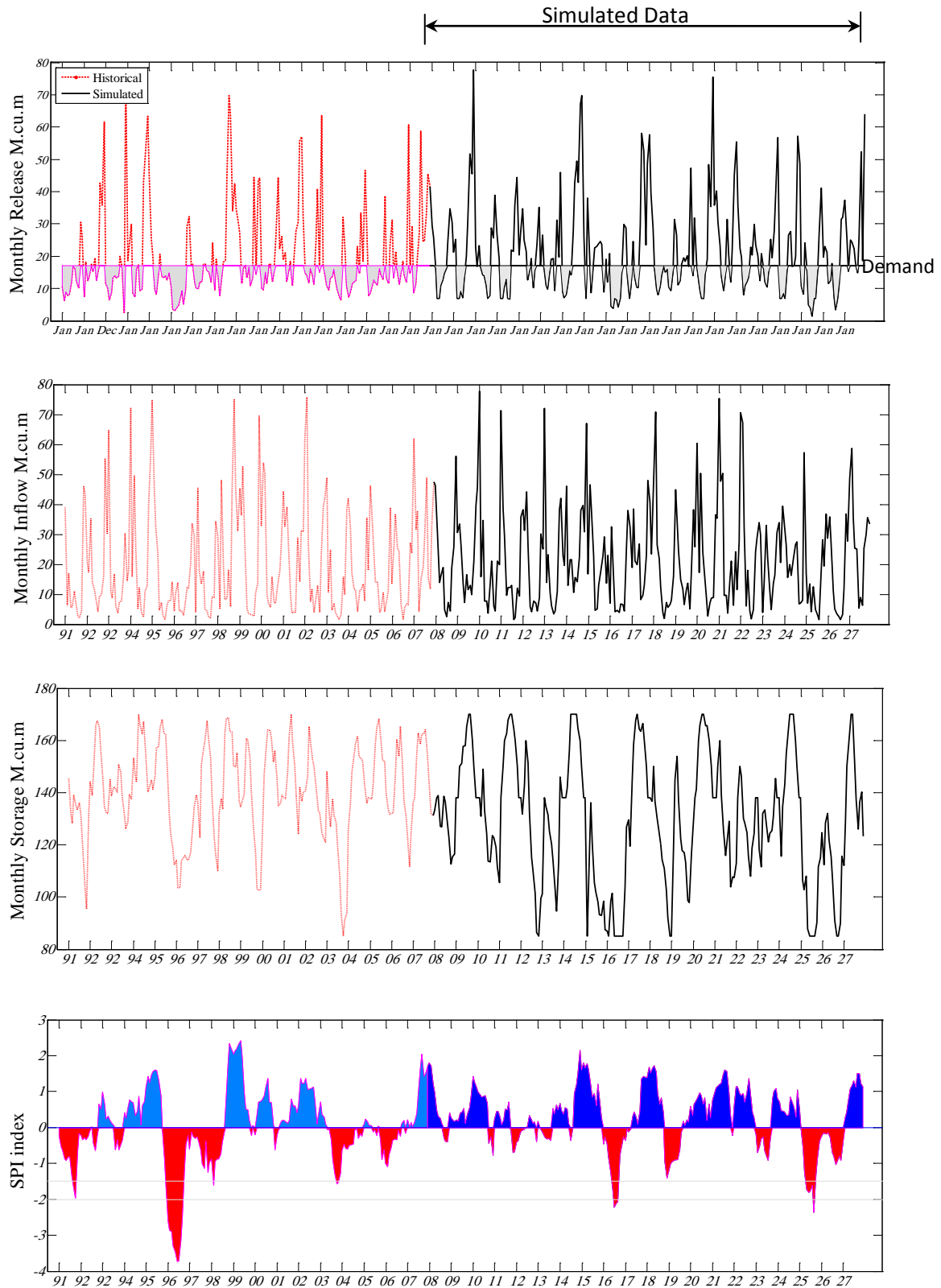
Figure 7.8: Comparison between historical and simulated data
(15 years simulation period)

Table 7.6: Sample of simulated data contains the driest year in 500 simulated years using model M_4 and SPI_9

Month (i)	Inflow(i)	Storage(i-1)	Storage(i)	SPI(i-1)	SPI(i)	SPI(i+1)	Releas(i+1)
1.00	33.03	124.93	138.00	0.24	0.16	0.00	20.17
2.00	8.52	138.00	131.71	0.16	-0.48	0.32	20.88
3.00	10.48	131.71	120.06	-0.48	-0.84	-0.56	12.47
4.00	5.91	120.06	109.67	-0.84	-0.99	-0.40	10.75
5.00	4.62	109.67	103.10	-0.99	-0.99	0.08	9.71
6.00	3.36	103.10	96.97	-0.99	-1.22	-1.39	11.99
7.00	1.89	96.97	90.61	-1.22	-1.32	-1.27	17.12
8.00	1.51	90.61	80.52	-1.32	-1.86	-1.77	12.17
9.00	6.66	80.52	75.00	-1.86	-2.51	-2.62	5.00
10.00	4.40	75.00	75.00	-2.51	-2.80	-2.10	5.70
11.00	12.12	75.00	74.40	-2.80	-2.60	-2.30	7.08
12.00	26.51	74.40	80.82	-2.60	-2.11	-2.34	12.50
1.00	16.15	80.82	100.25	-2.11	-2.29	-2.60	6.27
2.00	12.95	100.25	103.90	-2.29	-2.42	-2.44	9.05
3.00	25.52	103.90	110.58	-2.42	-2.30	-2.26	9.70
4.00	4.22	110.58	127.05	-2.30	-2.47	-2.80	6.80
5.00	28.27	127.05	121.57	-2.47	-1.80	-1.77	6.53
6.00	16.96	121.57	143.03	-1.80	-1.59	-1.67	12.14
7.00	19.71	143.03	153.46	-1.59	-1.02	-1.31	13.60
8.00	7.97	153.46	161.02	-1.02	-0.72	-0.08	19.60
9.00	7.60	161.02	155.39	-0.72	-0.61	-0.16	22.57
10.00	14.46	155.39	143.38	-0.61	-0.04	0.16	43.35
11.00	37.97	143.38	135.28	-0.04	0.59	-0.24	52.61

Driest year in simulated data

7.4.3.6 Decision Making about the Release of the Next Month

Real-time reservoir operation requires a quick system response for calculation and rational decision making using available monitored data (Khattree and Rao, 2003). A quick response to an operator request is of utmost importance for a real-time decision support system. Fuzzy inference system gives the best assistance for these issues by comparing the similarities of the current events and the historical data. One of the important features of the developed models is the ability to forecast the release of next month based on the inflow of current month, storage of next month and considering the accuracy of SPI forecasting using transition matrix probability.

In the present study, by using any model from the developed models in case of release of next month, the user need only to load the historical data, to identify inflow of current month and to identify storage (according to the selected model). Based on historical inflow data, the SPI is forecasted using transition matrix probability then the FIS system predicts the release of next month. By considering the value of the release confidence factor, the operator can decide on the actual release and the starting time for operation.

7.4.4 Modeling of Reservoir Operation–Case2: Release of Current Month

In this main model, assumptions and procedures are the same as the model in case 1, case of release of next month, except some differences. The main difference between this main model and the one presented in section 7.4.3 is the output. Also set of four models for each SPI time scale is developed as shown in table 7.7. Results of training and test of models, which presented in table 7.7, are shown in figure 7.9 and figure 7.10 respectively. Tables 7.8& 7.9 present result of model evaluation for both training and test period.

Table 7.7: Description of the input of ANFIS-based learning models
Case2: Release of current month

Model	Input data of the selected model						
M_1_1	Month(i)	I(i)	S(i)	SPI(i)	SPI(i+1)		
M_2_1	Month(i)	I(i)	S(i)	SPI(i-1)	SPI(i)	SPI(i+1)	
M_4_1	Month(i)	I(i)	S(i-1)	S(i)	SPI(i-1)	SPI(i)	SPI(i+1)
M_7_1	Month(i)	I(i)	S(i)	SPI(i)	SPI(i+1)	SPI(i+2)	
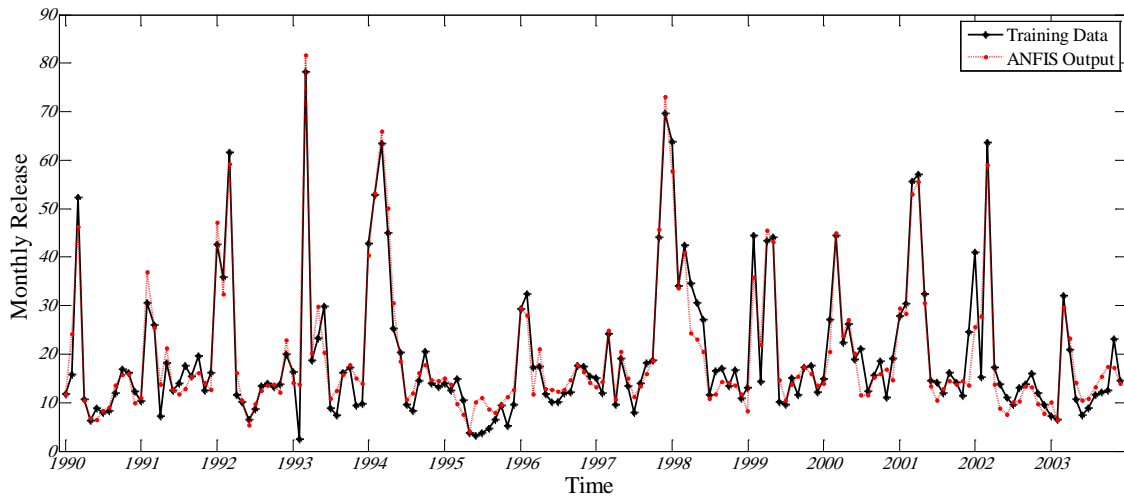
Figure 7.9 ANFIS output for reservoir release (training period).
Case2: Release of current month - Model M_1_1 / SPI_3
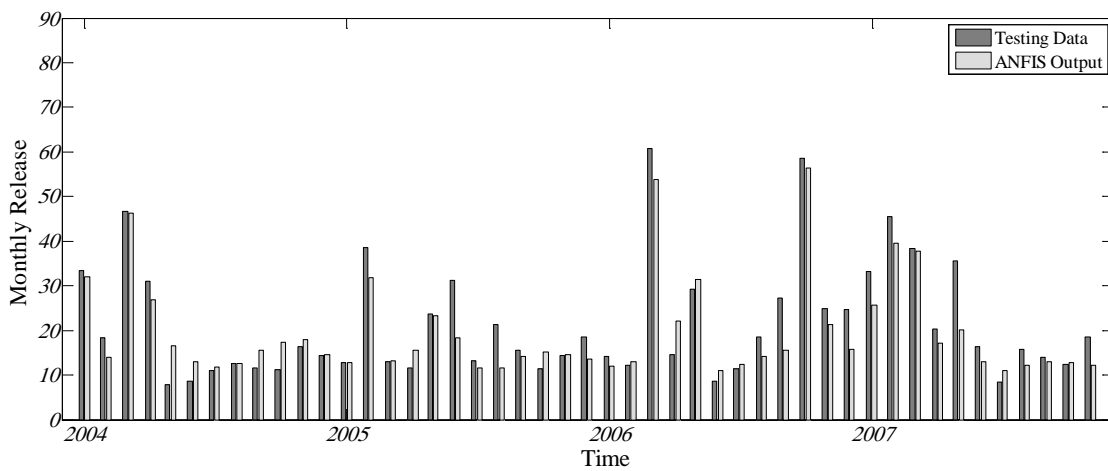
Figure 7.10: ANFIS output for reservoir release (test period). Case2: Release of current
month- Model M_1_1 / SPI_3

Table 7.8: Model evaluation criteria- case of release of current month (training period)

Model	SPI_3				SPI_6				SPI_9				SPI_12			
	$RMSE$	R^2	MAD	C_r	$RMSE$	R^2	MAD	C_r	$RMSE$	R^2	MAD	C_r	$RMSE$	R^2	MAD	C_r
M_1_1	3.97	0.97	2.88	0.96	3.44	0.98	2.46	0.95	3.46	0.98	2.45	0.96	3.45	0.98	2.42	0.92
M_2_1	3.91	0.97	2.90	0.96	5.37	0.95	4.11	0.93	5.07	0.95	3.8	0.93	5.52	0.95	4.38	0.92
M_4_1	3.69	0.98	2.75	0.97	3.62	0.98	2.7	0.97	5.58	0.95	4.71	0.92	5.47	0.95	4.22	0.92
M_7_1	4.38	0.97	3.38	0.95	5.36	0.95	4.16	0.93	5.22	0.95	4.09	0.93	5.4	0.95	4.33	0.93

Table 7.9: Model evaluation criteria- case of release of current month (test period)

Model	SPI_3				SPI_6				SPI_9				SPI_12			
	$RMSE$	R^2	MAD	C_r	$RMSE$	R^2	MAD	C_r	$RMSE$	R^2	MAD	C_r	$RMSE$	R^2	MAD	C_r
M_1_1	5.47	0.95	4.16	0.91	5.71	0.95	4.29	0.91	5.9	0.94	4.29	0.90	6.3	0.94	4.59	0.88
M_2_1	5.68	0.95	3.98	0.91	5.48	0.95	4.06	0.92	6.49	0.93	4.62	0.87	5.7	0.95	4.4	0.92
M_4_1	5.28	0.94	3.86	0.93	6.46	0.95	4.74	0.88	5.91	0.94	4.43	0.89	5.97	0.94	4.48	0.90
M_7_1	6.23	0.94	4.85	0.90	5.69	0.95	4.4	0.91	5.3	0.95	4.01	0.91	5.76	0.94	4.49	0.91

7.4.6 Studying the effect of using SPI index on Performance enhancement of Simulation Models

As mentioned before, the approach used in this study is a new approach. A set of models with different assumption have been applied. In order to investigate the effect of using SPI index, all suggested models have been applied to historical data but without using SPI index. Results of models evaluation indicate that using of SPI index has enhanced the performance of simulation models. Figure 7.11 and table 7.10 presents a comparison between candidate models for the two cases with and without SPI index.
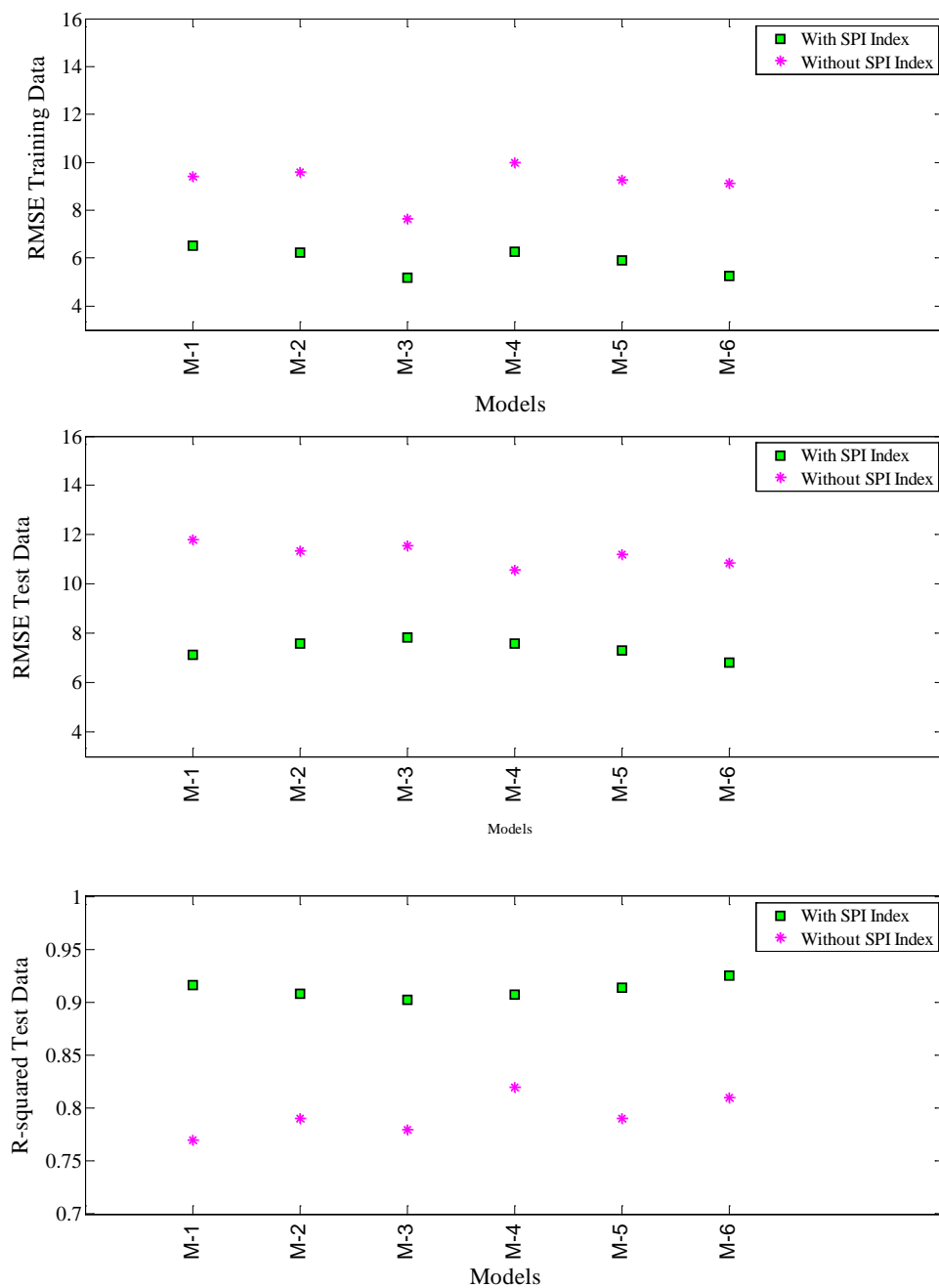
Figure 7.11: Comparison between evaluation criteria of candidate models for the two cases with& without SPI index - Case1" Release of next month"

Table 7.10: Comparison between evaluation criteria of candidate models for the two cases with& without SPI index
Case1" Release of next month" (test period).

Model	SPI_3			SPI_9			SPI_12			Without SPI		
	$RMSE$	R^2	MAD	$RMSE$	R^2	MAD	$RMSE$	R^2	MAD	$RMSE$	R^2	MAD
M_1	7.13	0.916	5.36	8.21	0.889	5.7	8.88	0.87	6.60	11.80	0.77	8.2
M_2	7.57	0.9077	5.63	8.189	0.889	5.71	8.91	0.87	6.72	11.34	0.79	8.02
M_3	7.82	0.902	5.75	7.04	0.9203	5.209	7.31	0.914	5.65	11.55	0.78	8.25
M_4	7.59	0.9073	5.93	8.88	0.87	6.27	8.82	0.875	6.75	10.58	0.82	7.81
M_5	7.31	0.914	5.66	8.05	0.89	5.64	7.59	0.9073	5.80	11.21	0.79	7.86
M_6	6.82	0.925	5.10	7.72	0.904	5.492	7.82	0.9016	6.048	10.84	0.81	7.98

7.5 Conclusion

In this study, an example of the collective use of stochastic models and ANFIS has been presented. Fuzzy set theory plays an important role in dealing with uncertainty when making decisions in reservoirs operation. ANFIS is a powerful fuzzy logic neural network, which provides a method for fuzzy modeling to learn information about the data set that best allow the associated fuzzy inference system to trace the given input/output data. In this study, the use of the adaptive network-based fuzzy inference system (ANFIS), to construct a model for reservoir operation, simulation of reservoir operation and decision making about reservoir release has been proposed. The applicability and capability of the ANFIS model have been investigated through the use of a set of data in the Ruhr reservoirs system, Germany. The historical data are inflow, storage, SPI index and release. The historical data sets have been divided into two independent sets to train and to test the constructed models.

Two main models have been developed. In both models the set of input include time of year, storage, inflow and Standardized Precipitation Index (SPI). The output of the first model is the release during the next month; on the other hand, the output of the second model is the release of the current month. Fuzzy Inference System has been prepared using Fuzzy logic toolbox in MATLAB and this system has been used as an input to ANFIS to obtain the final FIS. The FIS has been evaluated to obtained output data which is the predicted value of the reservoir release for the particular model. Predicted release values and observed release have been then evaluated using several evaluation criteria. Results of evaluation showed that the ANFIS models are accurate and consistent in different subsets, where most of the values of RMSE and MAE are smaller, and most of correlation coefficients and R^2 are also very close to unity.

In order to demonstrate the effect of using SPI index as input, two ANFIS models have been developed and investigated; one with SPI as input variable and another without. It has been found that, the model which contains SPI as input variable has consistently superior performance compared with the one without SPI index. Results obtained in this study showed that, the ANFIS models provide reliable reservoir release prediction for current and next month. Results also showed that the proposed approach could be a good tool for evaluation of release for a specified month and could be also a helpful reference guide to the operator during dealing with decisions.

Chapter 8

8. Drought Management Plan

8.1 Introduction

Water is one of the most important natural resources we need. There are several reasons for short supply of water. Some of these causes are over-allocation, over-use of water sources or a prolonged period of below normal precipitation, more commonly referred to as a drought. Drought is a natural hazard temporarily affecting almost every region in the world. The temporary shortage of water poses a great threat on nature, quality of life and economy. As drought is a slowly developing phenomenon, only indirectly affecting human life, its impacts are often underestimated in financially well off regions such as Europe (Stahl, 2001). Droughts often result in heavy crop damage and livestock losses, disrupt energy production and hurt ecosystems. Drought mortality is concentrated in developing countries, while absolute economic losses are largest in developed regions

Drought is a major natural hazard affecting large areas and millions of people every year. The World Meteorological Organization (WMO) estimated that in the 25 years from 1967 to 1991 about 1.4 billion people were affected by drought and 1.3 million people were killed due to the direct and indirect cause of drought (Obasi, 1994).

A recent study performed by the European Commission and Member States estimates the costs of droughts in Europe over the last thirty years to at least 100 billion Euro (European Commission 2007). The drought of 2003 in Central and Western Europe has alone been responsible for an estimated economic damage of more than 12 billion Euro (European Commission, 2008b).

Many countries and local municipalities have a drought plan or participate in a drought planning effort. The actual implementation of the plans varies, as soon as the drought begins to lessen, most efforts get shelved until the next drought happens. There should be a more concerted effort to keep the drought planning, preparation and mitigation going, especially during the wetter periods.

Historical records demonstrate that droughts are also causing potential impacts in Europe. The risks of these potential impacts depend on the type of water demands, how these demands are met and the corresponding water supplies available to meet these demands. These impacts could be categorized into environmental, social and economic impacts .

Analysis of meteorological drought in the Ruhr basin shows that the Ruhr basin is exposed to drought events rather frequently. Historical records of meteorological drought in the Ruhr basin demonstrate that several severe and extreme events occurred in 1932/34, 1947, 1959, 1976, 1996, 2003 and 2007. It is worth to be mentioned that in reality extreme drought events in the last decades presented no severe challenges to the water supply of the Ruhr district due to the reservoir system existing in the Ruhr catchment basin (Khadr et al., 2009). In this study a drought management plan was proposed for the Ruhr river basin in order to reduce the impacts of drought events.

8.2 Classification of Drought Impacts

Drought produces a complex combination of impacts that exten over many sectors. Drought impacts can be classified as follow (European Commission, 2008a; Rossi et al., 2007):

- ➢ Environmental impacts

    - Lack of feed and drinking water,

    - Mortality of fish species,

    - Damages to river life (flora, fauna),

    - Loss of biodiversity in terrestrial areas depending on the aquatic system,

    - Damage to landscape quality (dust, soil erosion and reduce vegetation coverage),

    - Forest fires risk,

    - Increase of salt concentration in streams, underground layers and irrigated areas.

- ➢ Social impacts

    - Inconveniences due to water system rationing,

    - Risk for health connected with increase of pollution concentration and discontinuous water system,

    - Impacts on way of living (unemployment, reduced saving capability, difficulty in personal care, reuse of water at home, street and cars washing prohibition, doubt on future),

    - Risks on public security due to more frequent fires (forests, pasture).

➢ Economic impacts

- Damage to agricultural production,

- Damage to forest production,

- Damage to fishing,

- Damage to industries connected with agricultural production,

- Damage to industries affected by hydroelectric energy reduction,

- Damage to reduced navigability of streams, rivers and canals,

- Damage to tourism sector due to the reduced water availability in water supply and water bodies.

8.3 Drought and Water Scarcity

The two terms 'water scarcity' and 'drought' are commonly used alternately, while they are quite different phenomena affected by water management practices and natural causes respectively (European Commission 2007). Water-scarcity is both a natural and a human-made phenomenon. It is defined as a situation where insufficient water resources are available to satisfy long-term average requirements. It refers to long-term water imbalances, where the availability is low compared to the demand for water, and means that water demand is more than the water resources exploitable under sustainable aspects. On the other hand, droughts represent the relevant temporary decrease of the average water availability, refer to important deviations from the average level of natural water availability and are considered natural phenomena (European Commission, 2008b). It is not possible to control the occurrence of droughts although the resulting impacts may be mitigated to a certain degree, namely through appropriate surveillance and management strategies previously planned in a Drought Management Plan "DMP".

8.4 Drought Management in the European Union (EU)

Drought is an issue affecting all EU countries in different ways: severe droughts were identified that have affected more than 800.000 km² of the EU's territory (37 %) and at least 100 million inhabitants (20 %) in recent years with different degrees of intensity (European Commission, 2008a). Austria, Belgium, Cyprus, Finland, France, Germany, Hungary, Italy, Lithuania, Malta, the Netherlands, Norway, Portugal, Spain and the United Kingdom have all been hit, but other European countries have also been severely affected by droughts (e.g. Slovenia, Greece and Romania). As for the economic impacts of drought at the EU level estimates suggest losses of 100 billion Euros over the past 30 years (European Commission, 2008a).

The European Commission (European Commission, 2008a) reported that it is difficult to establish common European indicators to describe droughts and define prolonged drought due to the complexity of drought variability according to climatic and geographic conditions. Therefore the European Commission mentioned that it is better to work on different parameters to be included in local or national indicators that could be calibrated and compared, when sufficient data is available. The presence or not of these parameters in local indicators will depend on their local relevance.

The member states of EU are using several indicators to identify and manage droughts. Spain, UK, Portugal, Italy, Finland Netherlands and France have presented drought indicators to describe droughts and identify prolonged drought. According to the examples of indicators presented by these Member States, there are two main types of indicators. The first type is used to prepare for an event and the second type is used to characterize the event when it happens (European Commission, 2008a). Each Member State uses the first, the second or a combination of both, according to its needs. In general, drinking water supply is the priority usage in most EU countries and a minimum volume should be provided to the population whatever the climatic conditions are. This priority could become an aggravating factor for drought during summer seasons. Its importance compared to drought issues should be evaluated on the following factors: number of inhabitants supplied, volume, amount of abstraction from surface waters as part of total drinking water abstraction etc. In the following section an example of the indicators and management plan, that are used in EU, is presented.

8.4.1 Drought Management in Spain

The Spanish indicator system has been recognized to assess the quantitative status of water resources in the different exploitation systems existing in each river basin district (Rossi et al., 2007). The Hydrological Indicators System (HIS) was elaborated using different parameters (inflow, outflow and storage of reservoir, streamflow river gauges and aquifer water level) for each exploitation system. These parameters are used to assess the quantitative status of water resources in each system, comparing the record achieved in a determined period that has a historical and representative mean value.

As an example for drought indicators in Spain, the status indicator "I_e" that is used in the Jucar river basin is calculated as follow:

$$I_e = \frac{1}{2}\left[1 + \frac{V_i - V_{med}}{V_{max} - V_{min}}\right] \; if \; V_i \geq V_{med} \qquad\qquad (8.1)$$

$$I_e = \frac{V_i - V_{min}}{2(V_{med} - V_{min})} \; if \; V_i < V_{med} \qquad\qquad (8.2)$$

where;

I_e Status indicator,

V_i Measured mean value for the analyzed period (one month, 3 accumulated months or 12 accumulated months)

V_{med} Mean value for the historical period,

V_{max} Maximum value for the historical period,

V_{min} Minimum value for the historical period.

The following four levels are used to characterize a drought situation:

Green level (stable situation) $I_e > 0.50$

Yellow level (pre-alert situation) $0.50 > I_e > 0.30$

Orange level (alert situation) $0.30 > I_e > 0.15$

Red level (emergency situation) $0.15 \geq I_e$

The bases for the drought in Spain plans were established as follow :

- Present indicators that will provide a quick drought status early enough to act according to the forecasts of the Plan,

- Provide knowledge of the resources system and its elements' capability to be strained during scarcity situations,

- Present structural and non-structural alternatives to reduce drought impacts, and adaptation according to the status indicator,

- Measure the cost of implementing measures,

- Adapt the administrative structure for its follow-up and coordination among the different Administrations involved (Ministry, regional governments, municipalities...),

- Discuss Plans, results and follow-ups with all interested parties, ensuring full public participation to avoid social conflicts.

Basin authorities have been able to particularize plans according to their specificities, declare the drought status according to the Hydrological Indicators System "HIS" threshold, and initiate measures included in the plan depending on the gravity of the phenomenon. Based on the HIS thresholds, monthly maps of the drought situation in the different management units within each Spanish basin are being developed (European Commission, 2008a). Other examples of drought plans and drought indicators for EU states, such as Portugal and France, are found in the report published by European Commission (European Commission, 2008a).

8.5 Developing a Drought Management Plan for the Ruhr Basin

8.5.1 Definition of a Drought Management Plan (DMP)

Drought preparedness and mitigation actions should be carried out by attempting to answer some key questions such as:

- How do the managers when there is a drought?

- Which institution is in charge to manage drought related problems?

- What type of measures have to be implemented and when?

- What type of tools can be adopted to assess the effectiveness of the implemented measures?

A drought management plan (DMP) is a document required to be prepared by a drought management setting out how to minimize the impact on communities of water shortages caused by drought. It should detail (Rossi et al., 2007)

- The principal activities and groups at risk,

- Criteria to identify drought vulnerable areas,

- Mitigation actions and programs that address the vulnerability faced by the service provider in continuing to provide water services during drought conditions,

- Criteria to compare alternative drought mitigation measures,

- Drought indicators for calamity declaration,

- Definition of the priority in water allocation under shortage conditions among different users (municipal, agricultural and industrial),

- Tools to improve stakeholders' participation and public awareness,

- List of actions to recover drought damage.

8.5.2 Stages of a Drought Management Plan

The drought management plan proposed in this study has three phases, which are sequentially invoked as conditions dictate. These three phases are Drought Watch, Drought Warning and Drought Emergency.

8.5.2.1 Drought Watch

Rainfall data functions as a preliminary indicator for all phases of drought conditions. There are several indices that measure how much precipitation for a given period of time has deviated from historical norms. The National Drought Mitigation Center in U.S. is using the Standardized Precipitation Index (SPI) to monitor moisture supply conditions. Many drought planners appreciate the SPI's versatility. Distinguishing traits of this index are that it identifies emerging droughts months sooner than the Palmer Index and that it is computed on various time scales. In this study the SPI index and the percentile indices are used to assess the drought severity. The percentile indices are applied to reservoir storage as shown in table 8.2-a. The percentile indices of each month are presented in table 8.2-b. Details about the methodology of the SPI index are presented in chapter 3. Table 8.1 defines drought intensities resulting from the SPI index. A drought watch is declared when any of the indices indicate a drought watch; however indication of one index alone does not mandate a declaration.

Table 8.1: Classification of drought stages based on the SPI index

Stages	1	2	3	4	5	6	7
SPI	> 2	1.5 to 1.99	1 to 1.49	0.99 to -0.99	-1 to -1.49	-1.5 to -1.99	-2 and less
Classification	Extremely wet	Very wet	Moderately wet	Near normal	Moderately dry	Severely dry	Extremely dry

Table 8.2-a: Storage Triggers

	stable situation	Drought Watch	Drought Warning	Drought Emergency
Storage range from	> 45 Percentile	25-45 Percentile	10-25 Percentile	less than 10 Percentile
DMP class	Green level (1)	Yellow level (2)	Orange level (3)	Red level (4)

Table 8.2-b: Storage Percentiles (Bigge reservoir)

Month	Storage Percentiles (M.m^3)			
	45 Percentile	25 Percentile	10 Percentile	5 Percentile
January	138.97	133.20	118.79	89.76
February	137.74	133.38	123.62	103.18
March	146.43	139.71	127.37	112.67

Month	Storage Percentiles (M.m^3)			
	45 Percentile	25 Percentile	10 Percentile	5 Percentile
April	160.16	154.15	139.40	124.61
May	163.92	156.32	141.50	122.40
June	161.54	153.60	135.26	128.44
July	155.41	144.43	135.22	120.99
August	145.14	136.93	125.67	110.95
September	134.64	122.76	116.93	100.42
October	127.85	117.25	102.32	92.47
November	132.13	112.68	96.30	86.24
December	131.79	118.10	100.23	82.49

8.5.2.2 Drought Warning

With perfect forecasting abilities, water managers exactly know when and what type of restrictions to implement (if that is the management option of choice) to minimize drought impacts. Several models are used for drought forecasting. One of these models is presented in chapter 4 with a reasonable accuracy. In this study the transition matrix is used as a tool for drought assessment. In this study it is supposed that a drought warning is declared when one of the following conditions is met;

- The drought event according to SPI values extended for more than one month,

- There is more than a 30 % probability that the SPI index of the next month lies between -1.5 and -2. This probability is calculated using Markov model as explained in the next section,

- The reservoir storage is less than the 25 percentile.

8.5.2.2.1 Transition matrix

Modern probability theory studies chance processes for which the knowledge of previous outcomes influences predictions for future experiments (Grinstead and Snell, 1997). In principle, when we observe a sequence of chance experiments, all of the past outcomes could influence our predictions for the next experiment. In 1907, A. A. Markov began the study of an important new type of chance process. In this process, the outcome of a given experiment can affect the outcome of the next experiment. This type of process is called a Markov chain.

A Markov chain can be defined as follows: We have a set of states, $S = \{S_1, S_2,.... S_r\}$. The process starts in one of these states and successively moves from one state to another. Each move is called a step. If the chain is currently in state S_i, then it moves to state S_j at the next step with a probability denoted by p_{ij}, and this probability does not depend upon which states the chain was before the current state. The probabilities p_{ij} are called transition probabilities. The process can remain in the same state and this occurs with the probability p_{ii}. An initial probability distribution, defined on S, specifies the starting state. Usually this is done by specifying a particular state as the starting state.

In general, the size of this transition probability matrix depends on the total number of possible outcomes. For the SPI index, the possible outcomes are the 7 condition states as shown in table 8.1, thus the size of the matrix is 7 x 7 for each month (table 8.3). These transition probabilities can be more conveniently arranged in the matrix form P as follows:

			To state at the next month t+1						
			1	2	3	4	5	6	7
$P =$	From state at any month t	1	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	P_{16}	P_{17}
		2	P_{21}	P_{22}	P_{23}	P_{24}	P_{25}	P_{26}	P_{27}
		3	P_{31}	P_{32}	P_{33}	P_{34}	P_{35}	P_{36}	P_{37}
		4	P_{41}	P_{42}	P_{43}	P_{44}	P_{45}	P_{46}	P_{47}
		5	P_{51}	P_{52}	P_{53}	P_{54}	P_{55}	P_{56}	P_{57}
		6	P_{61}	P_{62}	P_{63}	P_{64}	P_{65}	P_{66}	P_{67}
		7	P_{71}	P_{72}	P_{73}	P_{74}	P_{75}	P_{76}	P_{77}

Table 8.3: Transition probability matrix for months based on SPI_3

Febraury

January	Stage	1	2	3	4	5	6	7
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	3	0.000	0.125	0.375	0.500	0.000	0.000	0.000
	4	0.000	0.000	0.129	0.839	0.032	0.000	0.000
	5	0.000	0.000	0.000	0.500	0.000	0.250	0.250
	6	0.000	0.000	0.000	0.500	0.000	0.250	0.250
	7	0.000	0.000	0.000	0.000	0.000	0.000	1.000

March

Febraury	Stage	1	2	3	4	5	6	7
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	3	0.000	0.000	0.286	0.714	0.000	0.000	0.000
	4	0.029	0.000	0.147	0.794	0.029	0.000	0.000
	5	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	6	0.000	0.000	0.000	0.000	1.000	0.000	0.000
	7	0.000	0.000	0.000	0.000	0.000	0.333	0.667

April

March	Stage	1	2	3	4	5	6	7
	1	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	2	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	3	0.143	0.000	0.286	0.571	0.000	0.000	0.000
	4	0.000	0.000	0.031	0.844	0.094	0.000	0.031
	5	0.000	0.000	0.000	0.667	0.000	0.333	0.000
	6	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	7	0.000	0.000	0.000	0.333	0.333	0.333	0.000

May

April	Stage	1	2	3	4	5	6	7
	1	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	2	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	3	0.000	0.000	0.333	0.667	0.000	0.000	0.000
	4	0.000	0.056	0.139	0.722	0.028	0.056	0.000
	5	0.000	0.000	0.000	0.500	0.500	0.000	0.000
	6	0.000	0.000	0.000	0.000	0.000	0.500	0.500
	7	0.000	0.000	0.000	0.000	0.000	1.000	0.000

		June									July						
	Stage	1	2	3	4	5	6	7		Stage	1	2	3	4	5	6	7
May	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**June**	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.000	0.333	0.333	0.333	0.000	0.000	0.000		2	0.000	0.667	0.000	0.333	0.000	0.000	0.000
	3	0.333	0.167	0.000	0.500	0.000	0.000	0.000		3	0.000	0.500	0.500	0.000	0.000	0.000	0.000
	4	0.000	0.032	0.032	0.839	0.000	0.065	0.032		4	0.000	0.000	0.086	0.800	0.086	0.029	0.000
	5	0.000	0.000	0.000	1.000	0.000	0.000	0.000		5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	6	0.000	0.000	0.000	0.500	0.000	0.500	0.000		6	0.000	0.000	0.000	0.600	0.000	0.000	0.400
	7	0.000	0.000	0.000	0.000	0.000	1.000	0.000		7	0.000	0.000	0.000	1.000	0.000	0.000	0.000

		August									September						
	Stage	1	2	3	4	5	6	7		Stage	1	2	3	4	5	6	7
July	1	0.500	0.500	0.000	0.000	0.000	0.000	0.000	**August**	1	0.500	0.000	0.500	0.000	0.000	0.000	0.000
	2	0.000	0.000	1.000	0.000	0.000	0.000	0.000		2	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	3	0.250	0.000	0.250	0.500	0.000	0.000	0.000		3	0.000	0.000	0.375	0.625	0.000	0.000	0.000
	4	0.000	0.000	0.121	0.727	0.121	0.030	0.000		4	0.000	0.000	0.071	0.786	0.071	0.036	0.036
	5	0.000	0.000	0.000	0.667	0.000	0.333	0.000		5	0.000	0.000	0.000	0.600	0.000	0.200	0.200
	6	0.000	0.000	0.000	0.000	1.000	0.000	0.000		6	0.000	0.000	0.000	0.250	0.500	0.250	0.000
	7	0.000	0.000	0.000	0.000	0.000	1.000	0.000		7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

	October										**November**							
September	Stage	1	2	3	4	5	6	7	**October**	Stage	1	2	3	4	5	6	7
	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000		1	0.500	0.000	0.500	0.000	0.000	0.000	0.000
	2	0.000	0.000	0.000	0.000	0.000	0.000	0.000		2	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	3	0.167	0.167	0.000	0.667	0.000	0.000	0.000		3	0.000	0.000	1.000	0.000	0.000	0.000	0.000
	4	0.000	0.000	0.031	0.906	0.063	0.000	0.000		4	0.000	0.028	0.111	0.750	0.083	0.000	0.028
	5	0.000	0.000	0.000	0.500	0.250	0.000	0.250		5	0.000	0.000	0.000	0.600	0.400	0.000	0.000
	6	0.000	0.000	0.000	0.667	0.333	0.000	0.000		6	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	7	0.000	0.000	0.000	0.000	0.500	0.500	0.000		7	0.000	0.000	0.000	1.000	0.000	0.000	0.000

	December										**January**							
November	Stage	1	2	3	4	5	6	7	**December**	Stage	1	2	3	4	5	6	7
	1	0.000	0.000	1.000	0.000	0.000	0.000	0.000		1	0.000	0.000	1.000	0.000	0.000	0.000	0.000
	2	1.000	0.000	0.000	0.000	0.000	0.000	0.000		2	0.000	0.000	1.000	0.000	0.000	0.000	0.000
	3	0.000	0.000	0.333	0.667	0.000	0.000	0.000		3	0.000	0.000	0.500	0.500	0.000	0.000	0.000
	4	0.000	0.030	0.030	0.818	0.030	0.061	0.030		4	0.000	0.000	0.118	0.706	0.088	0.088	0.000
	5	0.000	0.000	0.000	0.600	0.400	0.000	0.000		5	0.000	0.000	0.000	0.750	0.250	0.000	0.000
	6	0.000	0.000	0.000	0.000	0.000	0.000	0.000		6	0.000	0.000	0.000	0.500	0.000	0.500	0.000
	7	0.000	0.000	0.000	0.000	1.000	0.000	0.000		7	0.000	0.000	0.000	0.000	0.000	0.000	1.000

8.5.2.3 Drought Emergency

A drought emergency is declared when there is a reasonable probability that, without the implementation of predefined measures to reduce water consumption, a prolonged drought period would cause the reservoirs to be drained. Historical records show that there are no two droughts which have identical characteristics; therefore no single probability profile could be identified in advance that would be applied to the declaration of drought emergency. The estimation of this probability is based on several items such as analysis of historical records, the pattern of dry period months, reservoirs system storage balances, water supply system, precipitation patterns and forecasting models.

8.5.3 Drought Response

In this section several rules are proposed for each phase of drought phases. Certain actions are to be implemented according to each phase of the successive phases.

8.5.3.1 Drought Watch

When a drought watch is declared, the following actions are to be implemented to prevent and prepare for a very dry stage. These actions are as follow:

- To increase public attention and to clarify the situation to consumers and request their cooperation in water conservation efforts,

- To reduce the water use; hotels and restaurants are urged to provide water only upon request,

- To prevent washing vehicles except at station with water recycling,

- To establish direct drought communication between the industrial sectors and the drought management planners,

- To implement industrial water reduction opportunities which are previously identified and to identify alternative water sources for emergency use for water dependent industries.

8.5.3.2 Drought Warning

When drought warning is declared, additional actions are to be implemented to prevent and prepare for extremely drought as follow:

- To make an appeal to the public not to abuse drinking water and to increase the water conservation efforts,

- To reduce the legal constraints in several steps to reduce the additional water supply from reservoirs (e.g. -10 %, -20 %, -25 % depending on economical and ecological restrictions),

- To forbid water use for washing vehicles and the filling of swimming pools,

- To eliminate the filling of public fountains and watering of public parks, gardens and other similar areas and to increase outdoor water restrictions such as street cleaning and other outdoor water uses.

8.5.3.3 Drought Emergency

When drought emergency is declared, additional actions to those in the previous two phases are to be implemented to prevent and prepare for possible loss of supplies and maximum reductions for all sectors. These actions are as follow:

- To reduce central water supply on some hours per day,

- To close high water consuming industries

- To maximize the use of alternative sources such as pumping stations,

- To import drinking water from other regions,

- To follow the steps of the emergency drought plan including the coordination between all sectors and the disaster preparedness commission.

8.5.4 Case Study

The main objective of this section is to present an illustration of the proposed drought management plan by applying it to an actual drought event from historical records. The negative values of the SPI have been aggregated, based on SPI_3 and SPI_6, to be used as an indicator for dry years during the period 1969-2007 as shown in table 8.4. From the values shown in table 8.4, the three hydrological years 1976, 1996 and 2003 have been selected as the driest years. For the selected years precipitation and storage data are available. The SPI index, based on one and three months time step, and storage percentiles for the reservoir Bigge have been calculated and shown in table 8.5. The SPI index has been calculated using two time steps because sometimes a drought event could be detected using a specified time step, but the same event could not be detected using another time step. Thus, using several time steps could be useful when applying the drought management plan. It is worth to mention that, the status indicator that is used in the Spanish DMP has been applied to the storage records then the results have been compared to the storage percentiles as shown in figure 8.1. Results of the comparison showed reasonable agreement between the two indicators. In the following sections the drought event in selected dry years will be discussed.

Table 8.4: Summation of negative values of the SPI (1969-2007) (Bigge reservoir)

Year	Summation of negative values of SPI_3	Summation of negative values of SPI_6	Year	Summation of negative values of SPI_3	Summation of negative values of SPI_6
1969	-6.37	-4.83	1989	-8.19	-8.91
1970	-0.98	-1.60	1990	-4.76	-4.31
1971	-11.23	-11.94	1991	-10.37	-11.73
1972	-8.85	-10.14	1992	-1.28	-2.09
1973	-10.07	-13.24	1993	-5.38	-4.25
1974	-2.12	-1.96	1994	-1.28	-0.37
1975	-6.50	-2.34	1995	-6.93	-4.82
1976	-13.52	-15.62	1996	-14.98	-21.81
1977	-4.82	-8.29	1997	-9.55	-9.42
1978	-3.91	-2.65	1998	-1.10	-2.28
1979	-6.10	-6.97	1999	-5.48	-4.93
1980	-2.90	-3.30	2000	-3.86	-1.52
1981	-0.69	-0.18	2001	-2.71	-3.14
1982	-4.72	-3.38	2002	-0.71	0.00
1983	-5.50	-3.54	2003	-6.14	-6.94
1984	-1.48	-1.37	2004	-1.77	-0.82
1985	-7.49	-4.32	2005	-1.97	-0.37
1986	0.00	-1.01	2006	-2.41	-2.36
1987	-0.21	-0.13	2007	-0.13	0.00
1988	-3.27	-1.74			

Table 8.5: Description of the drought events in the year 1976 (Bigge Reservoir)

Year	Month	SPI_3				SPI_1				Storage M.m³	Storage range	DMP class
		SPI_3	stage	DMP class	Probability of the stage of the next month	SPI_1	stage	DMP class	Probability of the stage of the next month			
1976	2	0.1	4	1	79 % to stage 4	-1.46	5	2	28% to 2, 57% to 4, 15% to 5	148.70	> 45 Percentile	1
1976	3	0.24	4	1	84 % to stage 4	-1.07	5	3	25% to 3, 50% to 4, 25% to 6	153.30	> 45 Percentile	1
1976	4	-2.76	7	4	100 % to stage 6	-1.98	6	2	100% to stage 4	158.10	< 45 & > 25 Percentile	2
1976	5	-1.86	6	4	50 % to stage 4, 50 % to 6	0.06	4	1	57 % to stage 4, 23 % to 5	152.80	< 25 & > 10 Percentile	3
1976	6	-1.98	6	4	60 % to stage 4, 40 % to 7	-1.45	5	2	22 % to stage 3, 67 % to 4	138.00	< 25 & > 10 Percentile	3
1976	7	-0.78	4	1	72 % to stage 4	0.04	4	1	52 % to stage 4, 17 % to 5	122.90	< 10 Percentile	4
1976	8	-1.61	6	3	25 % to 4, 50 % to 5, 25 % to 6	-1.77	6	2	50 % to stage 4, 50 % to 5	103.20	< 10 Percentile	4
1976	9	-1.29	5	3	50 % to 4, 25 % to 5, 25 % to 7	-0.9	4	1	70 % to stage 4, 10 % to 5	86.90	< 5 Percentile	4
1976	10	-2.03	7	3	100 % to stage 4	-0.68	4	1	72 % to stage 4	71.50	< 5 Percentile	4
1976	11	-0.62	4	1	82 % to stage 4	0.74	4	1	67 % to stage 4	58.00	< 5 Percentile	4
1976	12	-0.59	4	1	70 % to stage 4	-0.74	4	1	20 % to stage 3, 65 % to 4	53.10	< 5 Percentile	4
1977	1	-0.36	4	1	84 % to stage 4	-0.21	4	1	62 % to stage 3, 16 % to 5	69.40	< 5 Percentile	4
1977	2	-0.68	4	1		0.28	4	1		102.80	< 10 Percentile	4
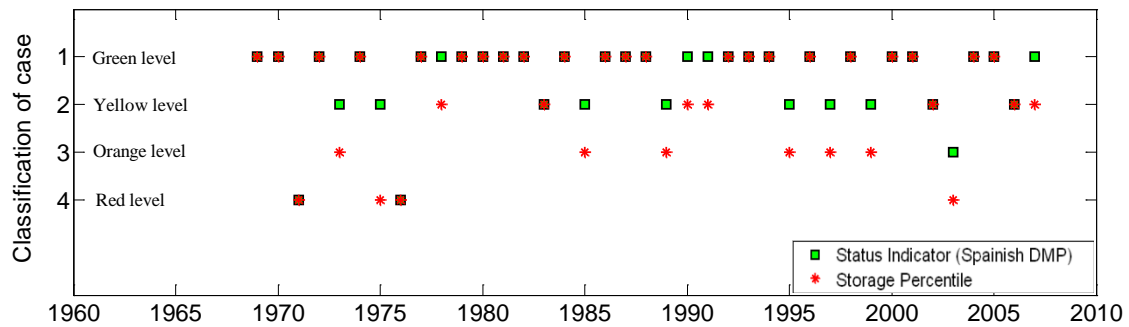
Figure 8.1: Comparison of the status indicator (Spanish DMP) and the storage percentile indicator which is proposed in this study (Bigge reservoir) (month December)

8.5.4.1 Case Study year 1976

8.5.4.1.1 Drought Watch

A shown in table 8.5, a drought watch is declared during the month February in 1976 using SPI_1 However, the storage percentiles and SPI_3 for the same month did not indicate a drought watch. With the declaration of drought watch (using SPI_1), all actions of this stage are to be implemented to prevent and prepare for very dry stage.

8.5.4.1.2 Drought Warning

The probabilities of transition from a specified stage during the current month to a specified stage in the next month are presented in table 8.5. When SPI_1 is considered for the month March in 1976, where drought watch is also declared, there is a probability that the stage of the next month will be also dry (25 % to stage 6) but this probability is less than 33 % and this indicates yellow level. The storage range and SPI_3 indicate non-drought condition (green level). According to the DMP, the drought warning is declared because the drought event extended to more than one month. All actions of this stage are to be applied to prepare for extremely drought.

8.5.4.1.3 Drought Emergency

As shown in table 8.5, the SPI_1 index indicates that the month April is a severe dry event (SPI_1 = - 1.98) and SPI_ 3 indicates an extreme dry event (SPI_3= -2.76). The storage percentile also indicates yellow level. SPI_3 stages show that there is high probability (100 % to stage 6) that next month will be extremely dry. These indicators lead to the declaration of the drought emergency. It is worth to mention that the probability value (100 %) was obtained because this type of drought only happened one time during this month through the study period, but the value of probability is not the only reason of drought emergency declaration. The drought emergency is declared because the drought event is continuing and there is a reasonable probability that without the implementation of predefined actions to reduce water consumption, a prolonged drought period would cause the reservoirs to be drained and this can be clearly notable from the storage data of the following months (table 8.5).

The beginning of a normal period (according to SPI_3) was in November 1976 where the SPI_3 started to be positive. The impacts of drought on the reservoir storage continued until March 1977 as shown in tables 8.5 and 8.6. Table 8.6 presents the actions that have been implemented during this period related to reservoir release. A comparison of the dry period 1976 with mean historical records is presented in figure 8.2.

Table 8.6: Comparison of the dry period 1976 with normal periods (Bigge reservoir)

	Month	Dry period (Case Study- year 1976)			Mean Values (1969-2008)			10 Percentile of release
		Year	Inflow (M.m^3)	Storage (M.m^3)	Release (M.m^3)	Inflow (M.m^3)	Storage (M.m^3)	Release (M.m^3)	
Dry Periods	1	1976	67.46	101.80	20.57	36.52	135.45	36.34	13.56
	2	1976	14.35	148.70	9.75	28.76	135.77	20.64	9.68
	3	1976	10.42	153.30	5.62	31.35	143.67	18.78	2.90
	4	1976	6.40	158.10	11.70	18.25	156.60	15.08	4.51
	5	1976	2.66	152.80	17.46	10.58	159.86	12.40	7.89
	6	1976	2.59	138.00	17.69	8.08	157.88	14.08	7.81
	7	1976	1.36	122.90	21.06	10.18	151.77	16.93	10.49
	8	1976	1.26	103.20	17.56	7.70	144.86	17.89	11.22
	9	1976	1.43	86.90	16.83	9.00	134.65	15.82	9.56
Transition Period	10	1976	2.39	71.50	15.89	15.62	127.84	18.80	10.40
	11	1976	7.66	58.00	12.56	25.77	124.99	22.90	11.54
	12	1976	20.09	53.10	3.79	35.26	127.87	27.50	10.66
	1	1977	35.88	69.40	2.48	36.52	135.45	36.34	13.56
Normal Period	2	1977	44.30	102.80	0.70	28.76	135.77	20.64	9.68
	3	1977	11.10	146.40	2.90	31.35	143.67	18.78	2.90
	4	1977	26.67	154.60	10.37	18.25	156.60	15.08	4.51
	5	1977	9.93	170.90	10.53	10.58	159.86	12.40	7.89
	6	1977	7.68	170.30	7.48	8.08	157.88	14.08	7.81

(a)

(b)

(c)

(d)

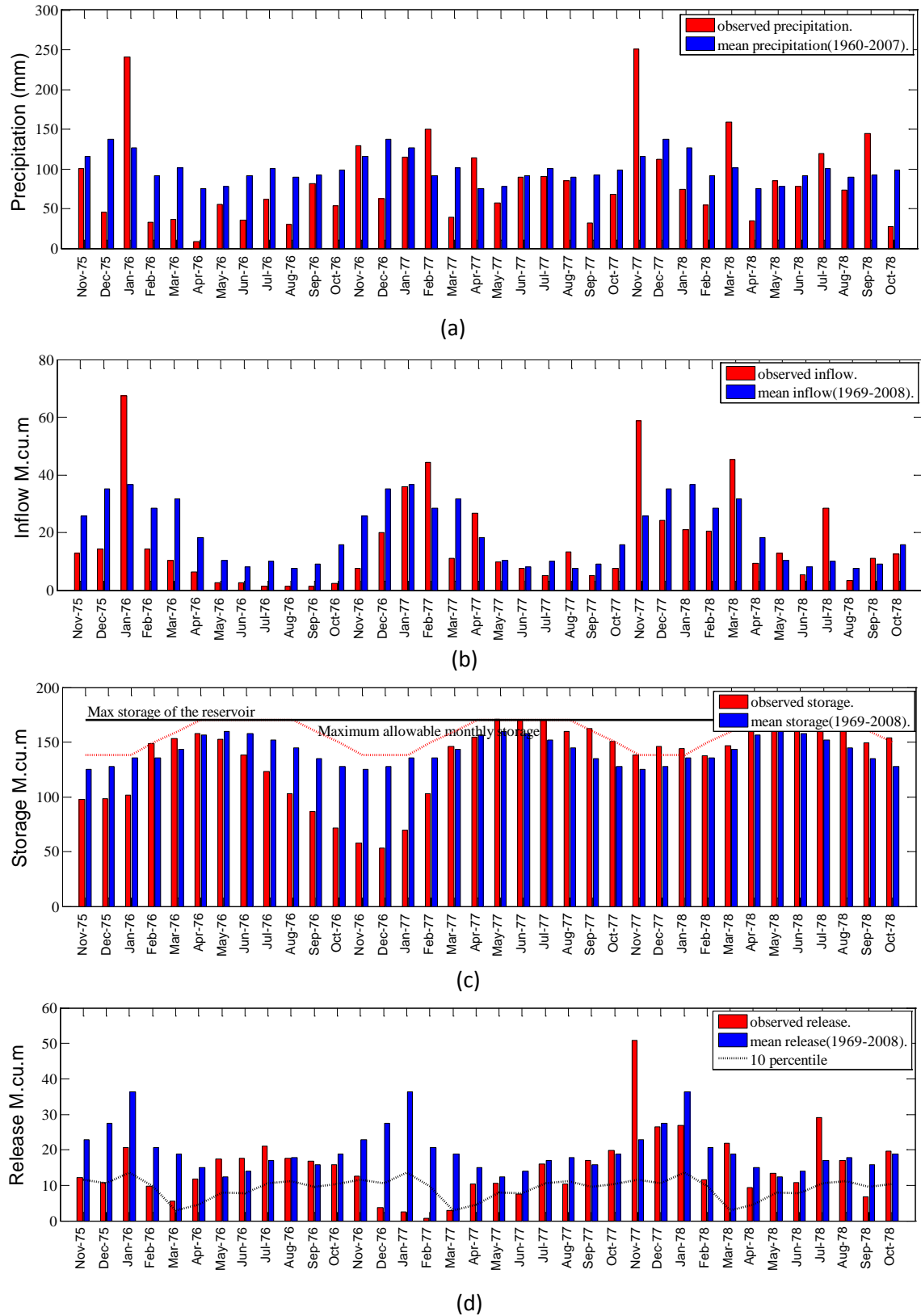**Figure 8.2:** Comparison of the dry period 1976 with mean historical records
(Bigge reservoir)

(a) Precipitation   (b) Inflow   (c) Storage   (d) Release

8.5.4.2 Case Study year 1996

8.5.4.2.1 Drought Watch

A shown in table 8.7, a drought watch is declared during the month October in 1995 using SPI_1 and the storage percentiles. However, SPI_3 for the same month did not indicate a drought watch. For this reason, the use of the transition probability matrix could be influential and effective when several time steps are used for SPI calculations. With the declaration of drought watch, all actions of this stage are to be implemented to prevent and prepare for very dry stages.

8.5.4.2.2 Drought Warning

The probabilities of transition from a specified stage during the current month to a specified stage in the next month are presented in table 8.7. When SPI_1 is considered for the month October, where drought watch was declared, there is a probability that the stage of the next month will be also dry (33 % to stage 6 + 33 % to stage 7) in addition to the storage range which indicates the warning case (yellow level). By this result the drought warning is declared. All actions of this stage are to be applied to prepare for extreme drought. This can be clearly noticed in table 8.8, the release of the reservoir during the month November was 14.04 $M.m^3$. This release is relatively small compared to the releases of the previous year (42.79 $M.m^3$) and the mean release of this month through the study period (22.90 $M.m^3$).

8.5.4.2.3 Drought Emergency

As shown in table 8.7, the SPI_1 index indicates that the month November has extremely dry event (SPI_1 = - 2.15) and there is high probability to the occurrence of this event as mentioned in the previous section. By using the SPI_1 and SPI_3 indices simultaneously in addition to the storage percentiles, the emergency stage will continue until the month May in 1996. When the SPI_1 is considered, the individual normal events which were in between dry periods were not an indication of the end of the dry period because normal rainfall of one month does not have notable effect on drought impacts.

This can be illustrated by the month February in the year 1996, SPI_1 indicates normal event but SPI_3 indicates an extremely dry event. The storage of this month was also less than 10 percentile. That reveals the importance of using SPI index with more than one time step. The emergency state ends when both SPI_1 and SPI_3 became positive in addition to a high probability that the next month will not have a dry event (table 8.7). The beginning of normal state is at the month July in 1996 where the SPI_1 started to be positive, the probability of a dry event during the next month became small and the type of state reached normal stage by September 1996.

Table 8.7: Description of the drought events in the year 1996 (Bigge reservoir)

| Year	Month	SPI_3				SPI_1				Storage M.m³	Storage range	DMP class
		SPI_3	stage	DMP class	Probability of the stage of the next month	SPI_1	stage	DMP class	Probability of the stage of the next month			
1995	10	-0.20	4	1	75 % to stage 4	-1.73	6	3	33 % to 4, 33 % to 6, 33 % to 7	124.2	< 45 & > 25 Percentile	2
1995	11	-0.99	4	1	82 % to stage 4	-2.15	7	4	67 % to 4, 33% to 6	118.5	< 45 & > 25 Percentile	2
1995	12	-3.20	7	4	100% to stage 7	-1.85	6	4	50 % to 5, 50 % to 7	112.4	< 25 & > 10 Percentile	3
1996	1	-3.43	7	4	100 % to stage 7	-2.33	7	4	33 % to 3, 33 % to 4, 33 % to 5	113.8	< 10 Percentile	3
1996	2	-2.32	7	4	33 % to stage 6, 67 % to 7	0.18	4	1	67 % to 4, 10 % to 7, 3 % to 6	103.6	< 10 Percentile	4
1996	3	-2.33	7	4	33 % to 4, 33 % to 5, 33 % to 6	-1.82	6	4	100 % to stage 6	103.7	<10 Percentile	4
1996	4	-1.83	6	4	50 % to stage 6, 50 % to 7	-1.94	6	4	100 % to stage 4	114.0	< 10 Percentile	4
1996	5	-2.29	7	4	100 % to stage 6	0.01	4	1	57 % to stage 4, 23 % to 5	115.1	< 10 Percentile	4
1996	6	-1.97	6	4	60 % to stage 4, 40 % to 7	-1.37	5	2	22 % to stage 3, 67 % to 4	115.9	< 10 Percentile	4
1996	7	-0.60	4	1	72 % to stage 4, 12 % to 5	0.30	4	1	52 % to stage 4, 17 % to 5	114.3	< 10 Percentile	4
1996	8	-0.20	4	1	78 % to stage 4	0.61	4	1	73 % to stage 4	114.2	< 10 Percentile	3
1996	9	0.14	4	1	90 % to stage 4	-0.53	4	1	70 % to stage 4, 10 to stage 5	117.0	< 25 & > 10 Percentile	3
1996	10	0.66	4	1	75 % tos tage 4	1.04	3	1	72 % to stage 4	123.89	< 45 & > 25 Percentile	2

Table 8.8 presents the actions that have been implemented during this period related to reservoir release. The Releases were reduced compared to mean values and this reduction is notable for month 3, 4, 5, 6, 7, and 8 in 1996. Releases during this months reached valus less than the 10 percentile of release as shown in table 8.8. To investigate the effect of this reduction on the storage, another scenario has been assumed for the relase during this dry period. the assumed releases equal to the difference between mean values and standard deviation of monthly releases(table 8.8). Results show that the the minimu storage using this scenario is 102 M.m^3 during August 1996.

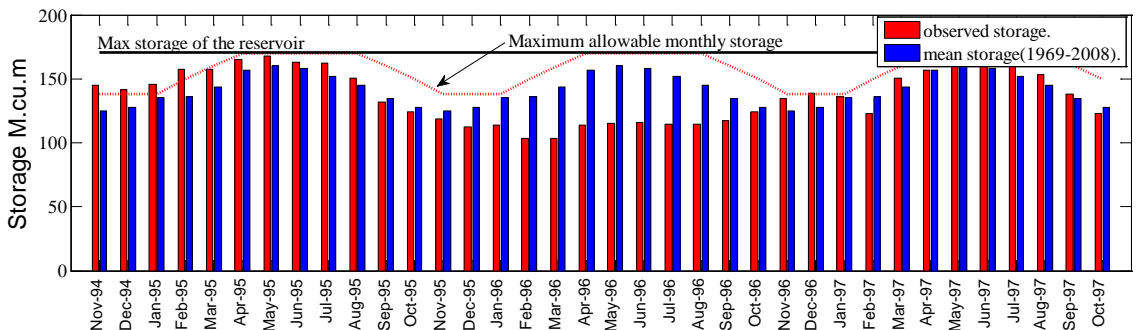**Table 8.8:** Comparison of the dry period 1996 with normal periods (Bigge reservoir)

|	Month	Year	Dry period (Case Study)			Mean Values (1969- 2008)			10 Percentile of release	Mean of release –std of relase
			Inflow (M.m^3)	Storage (M.m^3)	Release (M.m^3)	Inflow (M.m^3)	Storage (M.m^3)	Release (M.m^3)		
Dry Periods	10	1995	7.62	124.24	13.28	15.62	127.84	18.80	10.40	13.28
	11	1995	7.88	118.58	14.04	25.77	124.99	22.90	11.54	14.04
	12	1995	14.05	112.41	12.57	35.26	127.87	27.50	10.66	12.57
	1	1996	4.61	113.89	14.85	36.52	135.45	36.34	13.56	14.85
	2	1996	10.63	103.65	10.55	28.76	135.77	20.64	9.68	10.55
	3	1996	13.97	103.72	3.70	31.35	143.67	18.78	2.90	2.84
	4	1996	4.38	114.00	3.27	18.25	156.60	15.08	4.51	5.01
Transition Period	5	1996	4.62	115.11	3.81	10.58	159.86	12.40	7.89	7.58
	6	1996	3.08	115.92	4.65	8.08	157.88	14.08	7.81	8.87
	7	1996	6.27	114.36	6.44	10.18	151.77	16.93	10.49	9.70
	8	1996	12.26	114.20	9.43	7.70	144.86	17.89	11.22	9.31
Normal Period	9	1996	12.12	117.04	5.26	9.00	134.65	15.82	9.56	9.15
	10	1996	20.30	123.89	9.57	15.62	127.84	18.80	10.40	8.15
	11	1996	33.71	134.62	29.23	25.77	124.99	22.90	11.54	9.24
	12	1996	29.37	139.10	32.38	35.26	127.87	27.50	10.66	9.01
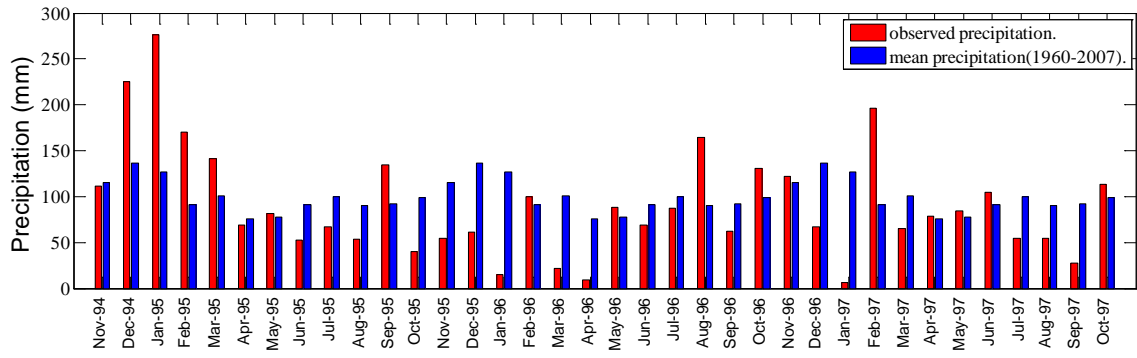
(a)
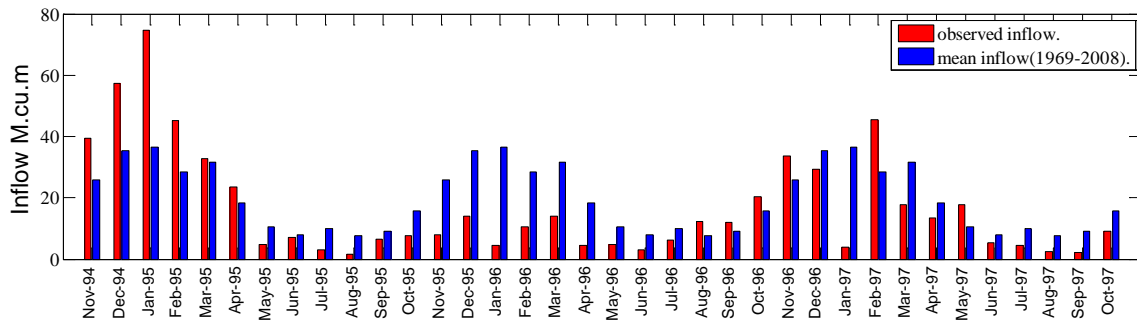
(b)

(c)
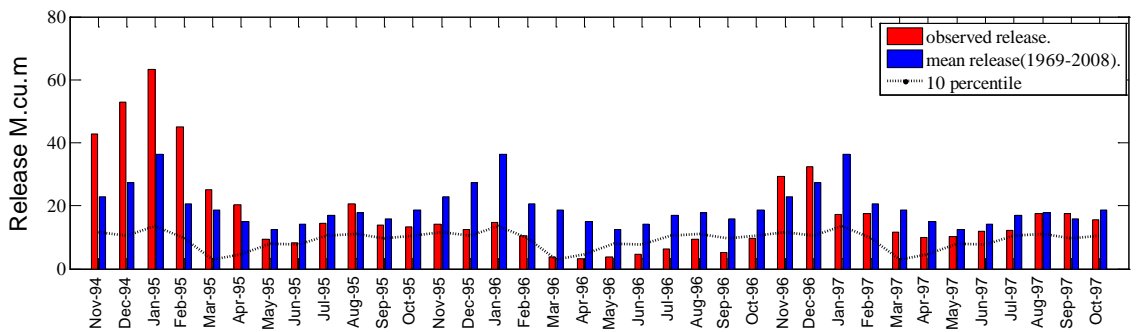
(d)

Figure 8.3: Comparison of the dry period 1996 mean historical records
(Bigge reservoir)

(a) Precipitation   (b) Inflow   (c) Storage   (d) Release

8.5.4.3 Case Study year 2003

8.5.4.3.1 Drought Watch

Table 8.9 shows that the drought watch is declared during the month February in 2003 using SPI_1 and the storage percentiles. However SPI_3 did not indicate a drought watch for the same month. The storage of this month is less than 25 percentilewhich declares drough warning.

8.5.4.3.2 Drought Warning and drought emergency

As shown in table 8.9, the storage percentile of month March 2003 declares the drought emergency (red level) and this DMP class continues until December 2003. SPI_3 of month April 2003 (-1.49) shows that there is a 50 % probability that the stage of the next month will be also dry. Stages of SPI_3 show that May 2003 was last month with drought stage, but it is clearly notable that some months have negative SPI_3 values that are more than -1. When SPI_1 is considered, only two months (August and November) provided moderately drought stage. Also when SPI_6 is considered, both July and August provided moderately drought stage. The values of SPI during this year show that there is no existence for severe or extremely drought but the negative values of SPI continued for several month. Also releases from reservoir were relatively high compared with the inflow during this year (table 8.10), thus the storage percentile declared the drought emergency.

Table 8.9: Description of the drought events in the year 2003 (Bigge reservoir)

Year	Month	SPI_3	stage	DMP class	Probability of the stage of the next month	SPI_1	stage	DMP class	Probability of the stage of the next month	Storage M.m^3	Storage range	DMP class
2003	2	-0.15	4	1	14 % to stage 3, 79 % to 4	-1.02	5	2	28 % to 2, 57 % to4, 14 % to 5	133.2	< 25 & > 10 Percentile	3
2003	3	-0.42	4	1	3 % to 3, 84 % to 4 , 9 % to 5	-0.64	4	1	12 % to 2, 69 % to 3, 9 % to 4	126.8	< 10 Percentile	4
2003	4	-1.49	5	3	50 % to 4, 50 % to 5	-0.48	4	1	6 % to 3, 72 % to 4, 9 % to 6	137.5	< 10 Percentile	4
2003	5	-1.27	5	3	100 % to 4	-0.45	4	1	11 % to 3, 57 % to 4, 23 % to 5	131.3	< 10 Percentile	4
2003	6	-0.76	4	1	8 % to 3, 80 % to 4	0.01	4	1	7 % to 2, 71 % to 4, 7 % to 5	128.6	< 10 Percentile	4
2003	7	-0.67	4	1	12 % to 3, 72 % to 4	-0.48	4	1	11 % to 3, 53 % to 4, 17 % to 5	119.4	< 5 Percentile	4
2003	8	-0.89	4	1	7 % to 3, 78 to 4	-0.98	4	1	10 % to 3, 73 % to 4	108.3	< 5 Percentile	4
2003	9	-0.47	4	1	3 % to 3, 90 % to 4	0.67	4	1	9 % to 3, 69 % to 4	93.8	< 5 Percentile	4
2003	10	0.29	4	1	11 % to 3, 75 % to 4	0.79	4	1	71 % to 2, 9 % to 3, 3 % to 5	85.1	< 5 Percentile	4
2003	11	0.23	4	1	3 % to 3, 82 % to 4	-1.32	5	2	100 % to 4	91.3	< 10 Percentile	4
2003	12	-0.03	4	1	12 % to 3, 70 % to 4	-0.03	4	1	20 % to 3, 65 % to 4	93.8	< 10 Percentile	4
2004	1	0.34	4	1	13 % to 3, 84 % to 4	1.15	3	1	14 % to 2, 71 % to 4	126.1	< 25 & > 10 Percentile	3
2004	2	0.82	4	1	15 % to 3, 79 % to 4	0.34	4	1	10 % to 3, 67 % to 4	136.10	< 45 & > 25 Percentile	2

Table 8.10: Comparison of the dry period 2003 with normal periods (Bigge reservoir)

|	Month		Dry period (Case Study)			Mean Values (1969- 2008)			10 Percentile of release
		Year	Inflow (M.m^3)	Storage (M.m^3)	Release (M.m^3)	Inflow (M.m^3)	Storage (M.m^3)	Release (M.m^3)	
Normal Period	11	2002	39.41	122.63	40.93	25.77	124.99	22.90	11.54
	12	2002	41.94	121.11	15.16	35.26	127.87	27.50	10.66
	1	2003	48.90	147.88	63.58	36.52	135.45	36.34	13.56
Dry Period	2	2003	10.95	133.21	17.28	28.76	135.77	20.64	9.68
	3	2003	24.53	126.87	13.82	31.35	143.67	18.78	2.90
	4	2003	4.71	137.59	10.99	18.25	156.60	15.08	4.51
	5	2003	6.90	131.31	9.60	10.58	159.86	12.40	7.89
	6	2003	3.91	128.61	13.08	8.08	157.88	14.08	7.81
	7	2003	2.56	119.44	13.70	10.18	151.77	16.93	10.49
	8	2003	1.58	108.30	15.99	7.70	144.86	17.89	11.22
	9	2003	3.10	93.88	11.90	9.00	134.65	15.82	9.56
	10	2003	15.74	85.09	9.52	15.62	127.84	18.80	10.40
	11	2003	9.81	91.31	7.28	25.77	124.99	22.90	11.54
	12	2003	38.78	93.84	6.49	35.26	127.87	27.50	10.66
Normal Period	1	2004	42.03	126.13	32.06	36.52	135.45	36.34	13.56
	2	2004	32.26	136.10	20.83	28.76	135.77	20.64	9.68
	3	2004	17.36	147.54	10.60	31.35	143.67	18.78	2.90
	4	2004	11.77	154.30	7.36	18.25	156.60	15.08	4.51

.
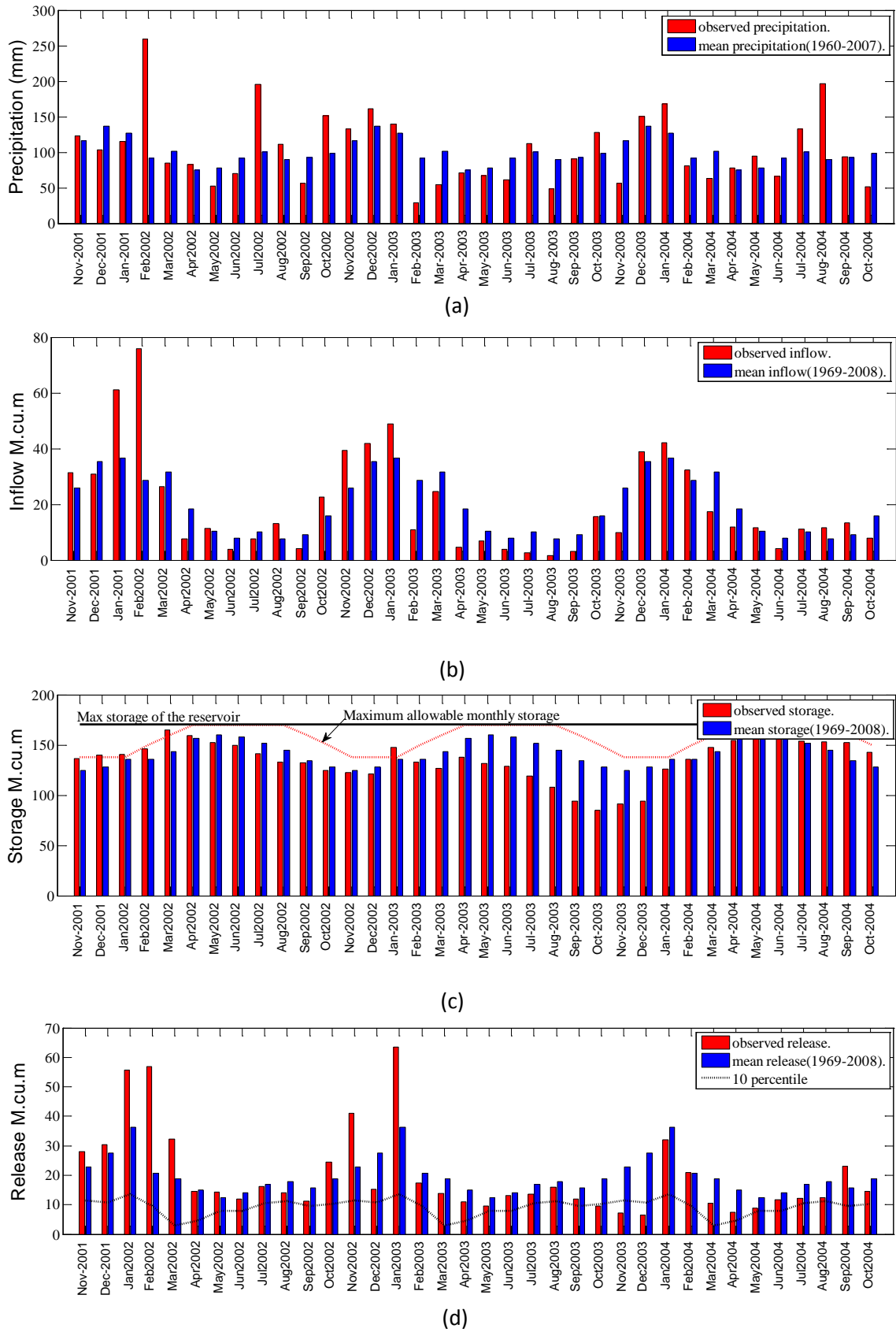
Figure 8.4: Comparison of the dry period 2003 mean historical records
(Bigge reservoir)
(a) Precipitation   (b) Inflow   (c) Storage   (d) Release

8.5.4.4 Comparative Analysis of Droughts of the years 1976, 1996 and 2003

As mentioned before, the accumulated magnitude of negative values of SPI have been calculated to be used as a measure for the selection of the driest years within the study period. The hydrological year 1996 was found to be the most critical year followed by 1976. The year 2003 was not so critical compared to 1976 and 1996. The drought events during 1976 and 2003 were in summer but in 1996 drought event was in winter.

Table 8.11 illustrates that inspite of the small sum of the monthly inflow during 1976 (only 20 % higher than 1996), the releases during 1976 exceeded those of 1996 (80 % higher than releases during 1996). The main reason behind this is that the water abstraction during 1976 was more than this of 1996 as shown in figure 8.5.

In 1976 the situation (summer drought) was not the same as in 1996 (winter drought). The releases were not decreased (to satisfy water demand), thus the storage of the reservoir reached minimum value of 53 M.m^3 in December 1976. On the other hand, the situation during 2003 was totally different. According to data of this year, as shown in table 8.10, table 8.11 and figure 8.6, the initial storage of this year was approximately as this of 1996 and the summation of the inflow was 1.79 times that of 1996, but despite this the reservoir reached a storage less than 1996 as shown in table 8.11.
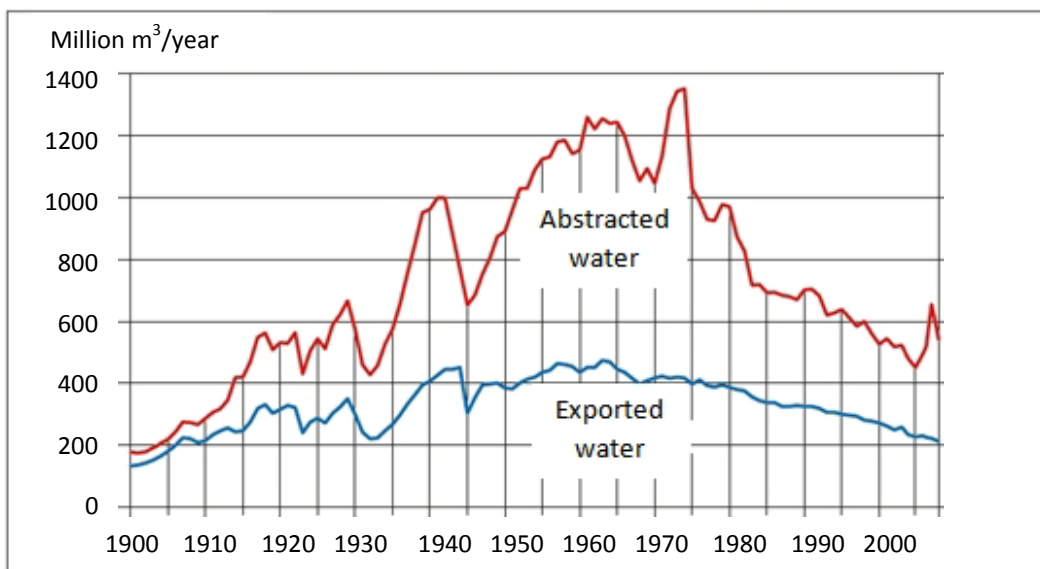

Figure 8.5: Annual abstracted and exported water in the Ruhr catchement area
between 1900 and 2009
Source(http://www.talsperrenleitzentrale-ruhr.de/veroeffentlichungen.html)

Table 8.11: Comparison of Drought between hydrological year 1976, 1996 and 2003

Hydrological Year	1976 (Summer drought)	1996 (Winter drought)	2003 (Summer drought)	Historical Mean (1969-2008)
Summation of monthly inflow (November: October) (Million.m^3)	137.39	114.17	204.23	237
Summation of monthly release (November: October) (Million.m^3)	177.09	98.14	235.55	237
Difference between inflow and release. D=Inflow_Release (Million.m^3)	-39.19	16.03	-31.32	0
Percentage of the difference between inflow and release (D/inflow) %	-28.9 %	14 %	-15.3 %	0 %
Minimum Storage (Million.m^3)	71.50 (reached a value of 53.1 in Dec. 76)	103.65	85.09	124.99
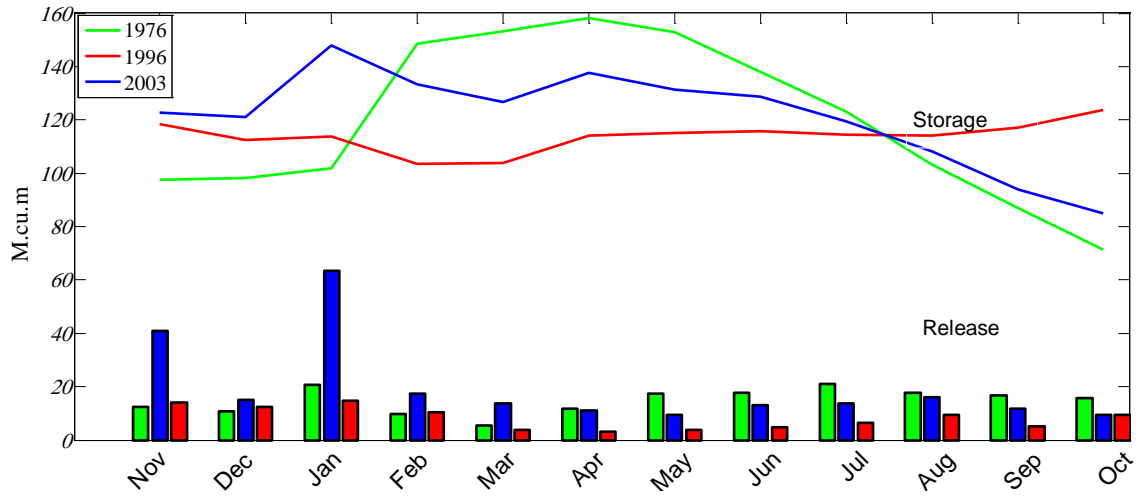
(a)
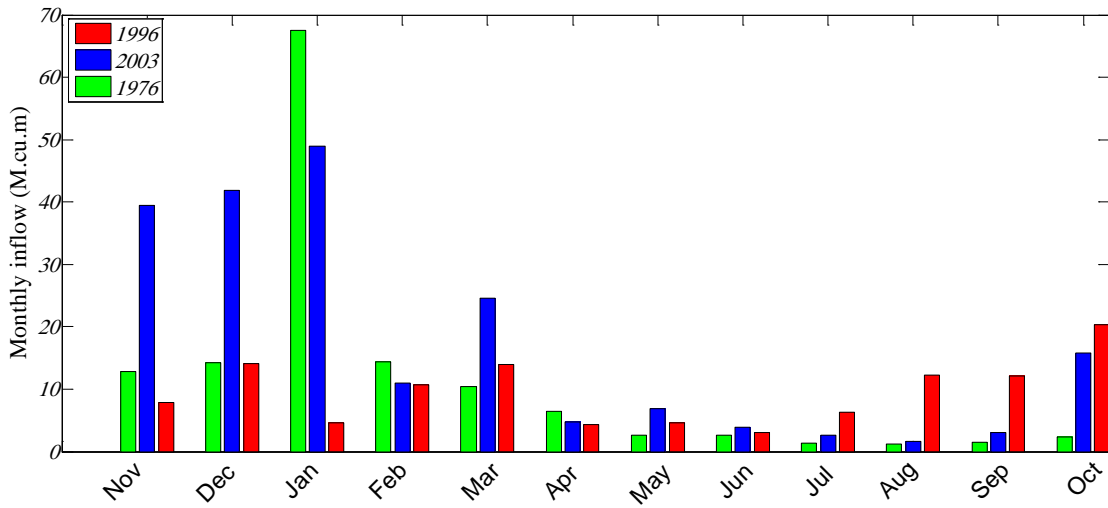
(b)
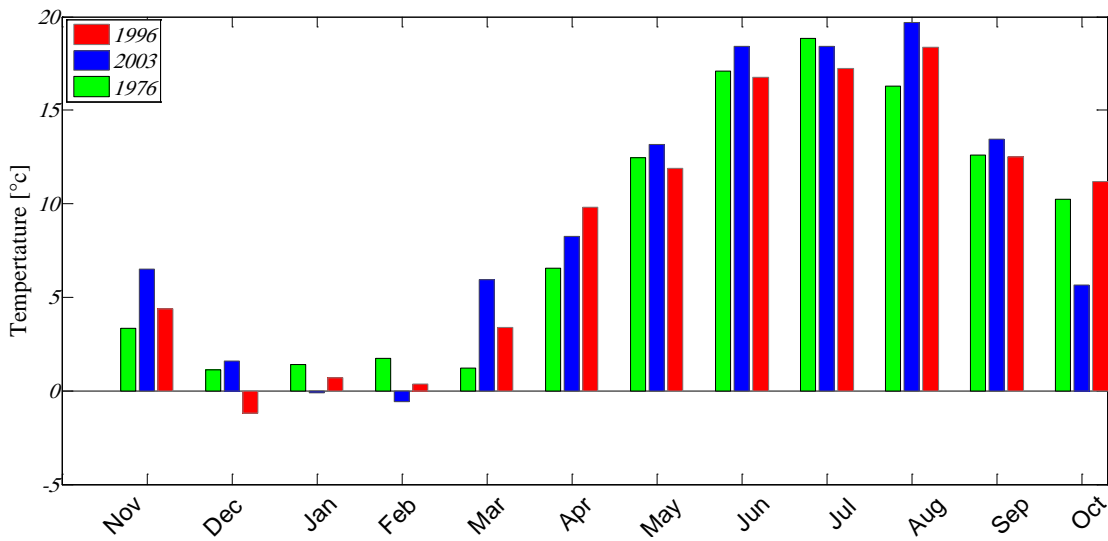
(c)

Figure 8.6: Comparison of Drought between year 1976, 1996 and 2003

(a) Storage and release   (b) Inflow   (c) Mean temperature

8.6 Conclusion

Drought differs from other natural hazards in several important points. Drought is a slow-onset and creeping phenomenon that makes it difficult to determine the onset and end of the event, its duration may range from months to years. The first and important evidence of drought is usually stated in precipitation records. The effects of a drought on streamflow and reservoirs may not be noticed for several weeks or months. Balancing the needs of all users of water supply during a drought periods can be difficult. Drought impacts mitigation represents one of the most challenging issues in water resources management, which can be successfully carried out by developing an efficient strategy in a Drought Management Plan (DMP). The DMP consists of planning, monitoring and implementation of planned and emergency actions to recover drought damage. Preparing an efficient drought management plan is the best way to reduce drought impacts. These impacts can continue to several weeks or months even after a drought event.

In this study a drought management plan "DMP" has been proposed for the monitoring and mitigation of drought in the Ruhr river basin. The proposed DMP consists of three stages namely Drought Watch, Drought Warning and Drought Emergency with several actions to be implemented during each stage. In the developed DMP, the indicators proposed to classify the drought classes are the SPI index and the storage percentiles in addition to the transition matrix of probability of the SPI index. In order to select a case study period from the historical records, the negative values of the SPI index have been aggregated based on one and three months time scale. The dry periods in 1976, 1996 and 2003 have been selected as case studies. The procedures of the developed DMP have been applied to the case studies. The proposed indicators, the SPI index and the storage percentiles, have been calculated during each month for the period under consideration. By applying the proposed DMP to the case studies, results showed the successful use of the SPI index based on several time scales simultaneously with the storage percentile in classifying the drought situation. Results also showed that the transition matrix of probability of the SPI index can be a useful guide for decision making during dry periods.

One of the main findings of this chapter is to emphasize that drought is a natural phenomenon, while water scarcity is both a natural and a human-made phenomenon. From that point the analysis of the three case studies 1976, 1996 and 2003 demonstrated that although 1996 was more critical than the 1976, the year 1995 was better managed. The reservoir storage during the year 1976 reached a value of 53.1 M.m^3 while the minimum storage of reservoir during the year 1996 was 103.65 M.m^3. On the other hand, the annual amount of water that was released from the reservoir in 1996 was 98.14 M.m^3 and 114.05 M.m^3 in 1976. These results emphasize the importance of implementing the actions of each stage of drought stages to address drought impacts and to prevent reservoir from being drained.

Finally, in order to assess risk and respond to drought, water suppliers have to establish a local drought management team. The drought management team needs to focus on three main goals: get to know the water supplies, improve water use efficiency and communicate, educate, and participate.

Chapter 9

Conclusions and Recommendations

In this chapter the main summaries and important conclusions are stated. It also provides some recommendations, poses open questions and suggests areas of future research.

9.1 Summary and Conclusions

The overall objective of this research effort was to study the change of climate in the Ruhr river basin and to evolve appropriate techniques and tools for drought characterization and for enhanced management of water resources systems during drought periods. The developing tools have four major components: 1) Climate change; 2) Drought monitoring and forecasting; 3) Reservoir operation during drought and 4) drought management plan. The developed tools have been demonstrated through an application to the Ruhr river basin as case study.

The major findings of this research are summarized as follows:

- ➤ After examining 68 years of precipitation and temperature data and 62 years of streamflow data, some statistically significant trends have been identified. There are slight increases in all three variables (streamflow, precipitation, temperature) in the Ruhr River basin. Between 1961 and 2007 results showed that over the study area a significant increase in the mean temperature over all time scales is considered. The occurrence of warm days in both winter and summer has a significant increase, while the occurrence of cold days in both seasons showed a similar proportion of significant decrease which is evident that the winter becomes warmer and the summer becomes hotter. Significant increase in the winter precipitation has been detected while the increases in the summer and the annual precipitation were statistically insignificant. The number of consecutive dry days displayed decreasing tendencies in winter while there is no indication of a statistically significant change in the summer. The index PR95 % (very wet days), PR99 % (extremely wet days) have been introduced in this study to explore the supposed amplified response of extreme precipitation events relative to the change in total amount. The main identified trends of very & extreme wet days were an increase in the very wet days in the winter. Results of inflow analysis showed that there is a significant increase in the winter inflow while the increases in summer and the annual inflow were found to be statistically insignificant.

➢ By applying the SPI methodology for drought monitoring, the obtained results indicated that the drought randomly affect the Ruhr river basin. Several drought events occurred during the period under study. Results indicated also that inspite of the significant positive trend in winter precipitation, drought visited the Ruhr basin in both summer and winter and that the most severely event was in the winter. Trends in SPI data series indicated that the proportion of the Ruhr catchment drought condition has insignificantly changed during the period under study. It is worth to be mentioned that in reality extreme drought events in the last decades presented no severe challenges to the water supply of the Ruhr district due to the reservoir system existing in the Ruhr catchment basin.

➢ The SPI index has been used as a drought indicator for drought forecasting due to its advantages over other drought indices. The capability of the ARIMA and SARIMA models in drought forecasting has been investigated using the correlation methods of Box and Jenkins and the AIC and SBC structure selection criteria. Results of the model evaluation showed a fairly good agreement between observations and forecasts, as it has also been confirmed by the values of some performance indices. The evaluation of the models showed that the results seem to be better for higher SPI series (SPI_6, SPI_9,..., and SPI_24) and this may be due to the increase in filter length which reduces the noise more effective. Results showed also that the good fitting of stochastic models such as ARIMA and SARIMA to hydrologic time series, such as SPI time series, could result in a better tool which can be used for water resource planning within the basin.

➢ A software package for meteorological drought forecasting has been developed. This package contains two main stochastic models, namely the Auto Regressive Moving Average (ARIMA) model and the Seasonal Auto Regressive Moving Average (SARIMA) model. The developed software has several advantages compared with other programs that are used for time series forecasting, such as its simplicity and ease of use. One of the advantages of the developed program is that the user does not need to try several candidate models in order to get the best model. The user needs only to put limits for the model's parameters then the program optimizes these parameters to detect the best model. The developed program has been calibrated using the well known software SPSS. Results of the calibration showed good agreement between the forecasted values using the developed program and those which obtained using the SPSS software with reasonable accuracy. The developed model has some limitations in order to obtain results with reasonable accuracy. Using of SPI model requires that there is no missing data in the time series and length data record is required to be at least 30 years. The developed model can be applied to any precipitataion records worldwide.

➢ The stochastic streamflow generation model of Thomas-Fiering and the Monte Carlo simulation model have been applied to generate monthly inflow data for four reservoirs in the Ruhr river basin. The statistical parameters such as means, standard deviation and skewness of the observed data and the generated data of the used approaches has been compared. The results showed that the generated data have successfully preserved the historical statistical parameters of streamflow. The results showed that the Thomas-Fiering model has also preserved the correlation coefficient between consecutive months. Thus, the Thomas-Fiering model has been used for producing inflow data needed for reservoir optimization and simulation models presented in this thesis.

➢ An optimization model has been developed using Genetic Algorithm (GA), Pattern Search and Gradient-based method for reservoir operation during normal periods and drought periods as well. The Bigge reservoir has been presented as case study. Two objective functions have been considered, then a weighted approach has been adopted to convert the multiple objectives problem into a single objective problem, so that the user can specify the priorities by giving a specified weight for each function. Several scenarios for low inflow periods have been attempted. The obtained results showed that both the GA approach and Gradient-based approach provide higher benefits than the Pattern Search approach. The evaluation of the developed model has been carried out using the the driest years in the available historical records. The monthly inflow of this year has been considered as input to the optimization model. Results of the evaluation demonstrated that the optimization model is beneficial. Results also showed that the developed model with its several scenarios and the suggested optimization approaches could be a helpful guide for the real operation of the reservoir.

➢ An example of the collective use of stochastic models has been presented. The use of the adaptive network-based fuzzy inference system (ANFIS) to construct a model for reservoir operation, simulation of reservoir operation and decision making about reservoir release has been proposed. The applicability and capability of the ANFIS model have been investigated by using a set of data of the Ruhr reservoir system, Germany. The used data are time of year (months), inflow, storage, Standardized Precipitation Index (SPI) and release. The historical data sets have been divided into two independent sets to train and test the constructed models. Two main models have been developed. The output of the first model is the release during the next month; on the other hand, the output of the second model is the release of the current month. Predicted release values and observed release have been evaluated using several evaluation criteria.

Results of the evaluation showed that the ANFIS models are accurate and consistent in different subsets, where most of the values of RMSE and MAE are smaller and most of the correlation coefficients and R^2 are also very close to unity. The effect of using the SPI index as input has been examined using two ANFIS models; one with SPI as input variable, another without. It has been found that the model which contains SPI as input variable has consistently superior performance compared with the one without SPI index. Results obtained in this study showed that the ANFIS models provide reliable reservoir release prediction for the current and the next month, and the proposed approach could be a good tool for the evaluation of the release for a specified month and could also be a helpful reference guide to the operator during decision making.

➢ Preparing an efficient drought management plan is the best way to reduce drought impacts. These impacts can continue several weeks or months even after a drought event. A drought management plan (DMP) has been proposed for monitoring and mitigation of drought in the Ruhr river basin. The proposed DMP consists of three stages namely Drought Watch, Drought Warning and Drought Emergency with several actions to be implemented during each stage. The proposed DMP has been applied to three years as case studies. The analysis of the case studies showed that the implementation of the actions of each stage of the drought management plan is very important to address drought impacts and to prevent reservoirs from being drained. The analysis of the case studies also showed, that the use of transition probability matrix can be an useful guide for decision makers during dry periods.

9.2 Recommendations for Further Study

To conclude this thesis, the following recommendations are made for further work that could lead to enhanced drought management performance:

- To develop a drought mapping system to monitor meteorological drought using the Standardized Precipitation Index (SPI) and interpolation methods.

- To study the occurrence probabilities, return periods and risk of meteorological drought events in the Ruhr river basin.

- To develop a drought early warning system using the Standardized Precipitation Index (SPI) as tool where this is the approach that the European Commission, Joint Research Centre proposed for drought forecasting over Europe.

- To develop a model for predicting the transition from a drought class of severity to another using the Standardized Precipitation Index (SPI) as drought indicator.

- To use the optimization model proposed in this study to develop a model for optimal multipurpose-multireservoir operation during drought in the Ruhr basin.

- To develop a simulation model for operating a multireservoir system during drought in the Ruhr basin using the approach proposed in this study

- To develop a drought management decision support system for short-term and long-term management.

References

(AMS), A.M.S., 2009. AMS Policy Statement on Meteorological Drought.

Adeyem, J.A., 2009. Application of Differential Evolution to Water Resources Management, Tshwane University of Technology.

Aguilar, E., Auer, I., Brunet, M., Peterson, T.C. and Wieringa, J., 1998. Guidance on Metadata and Homogenization, from http://cccma.seos.uvic.ca/ETCCDMI/docs/guide-to-metadata-homogeneity.doc.

Albert, J.M., 2004. Hydraulic Analysis and Double Mass Curves of the Middle Rio Grande From Cochiti to San Marcial, New Mexico, Colorado State University, Fort Collins, Colorado.

Alexander, L.V., 2005. Global Observed Changes in Daily Climate Extremes of Temperature and Precipitation. Journal of Geophysical Research, 111, D05109.

Alexander, L.V. et al., 2006. Global observed changes in daily climate extremes of temperature and precipitation. J. Geophys. Res., 111(D5): D05109.

Alexandersson, H., 1986. A Homogeneity Test Applied to Precipitation Data. Journal of Climatology, 6(6): pp. 661-675.

Altunkaynak, A., Özger, M. and Şen, Z., 2005. Regional Streamflow Estimation by Standard Regional Dependence Function Approach. Journal of Hydraulic Engineering © ASCE, 131(Issue 11): pp. 1001-1006

Awass, A.A., 2009. Hydrological Drought Analysis-occurrence, severity, risks: the case of Wabi Shebele River Basin,Ethiopia, University of Siegen- Germany, Siegen.

Babu, B.V. and Angira, R., 2001. Optimization of Nonlinear functions using Evolutionary Computation, Proceedings of 12th ISME Conference, Chennai,India, January, 10–12, pp. pp. 153-157

Beck, C., Grieser, J. and Rudolf, B., 2004. Extreme Daily Precipitation Events and Droughts in Germany. EMS Annual Meeting Abstracts, 2004 European Meteorological Society, 1

Beran, M.A. and Rodier, J.A., 1985. Hydrological aspects of drought, Studies and Reports in Hydrology 3. Unesco / International Hydrological Programme ; Unesco, Studies and reports in hydrology ; 39

Bhuiyan, C., 2004 Various Drought Indices for Monitoring Drought Condition In Aravalli Terrain of India, Geo-Imagery Bridging Continents XXth ISPRS Congress, Istanbul, Turkey.

Bobée and Ashkar, 1991. *The gamma distribution and derived distributions applied in hydrology*. Littleton Colo., Water Resources Press.

Box, G.E.P. and Jenkins, G.M., 1970. Time Series Analysis; Forecasting and Control. Holden-Day series in time series analysis. Holden-Day, San Francisco,, xix, 553 p. pp.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C., 2008. Time series analysis : Forecasting and Control. Wiley series in probability and statistics. John Wiley, Hoboken, N.J., xxiv, 746 p. pp.

Brocklebank, J.C. and Dickey, D.A., 2003. SAS for forecasting time series. J. Wiley, New York, x, 398 p. pp.

Brudy-Zippelius, T., 2003. Wassermengenbewirtschaftung im Einzugsgebiet der Ruhr: Simulation und Echtzeitbetrieb, Karlsruhe UniversitY, Karlsruhe.

Buishand, T., 1982. Some Methods for Testing the Homogeneity of Rainfall Records. Journal of Hydrology 58: pp. 11-27.

Burke, E.K. and Landa Silva, J.D., 2006. The influence of the fitness evaluation method on the performance of multiobjective search algorithms. *European Journal of Operational Research* 169 pp. 875-897.

Cacciamani, C., Morgillo, A., Marchesi, S. and V. Pavan, M., 2007. Monitoring and Forecasting Drought on a Regional Scale: Emilia-Romagna Region, 62. Springer Netherlands.

Cancelliere, A. and Bonaccorso, B., 2009. Uncertainty analysis of the Standardized Precipitation Index in the presence of trend, Hydrology Days 2009, Colorado State University.

Cancelliere, A., Mauro, G.D., Bonaccorso, B. and Rossi, G., 2007. Drought Forecasting Using the Standardized Precipitation Index. Water Resources Management 21(5): pp. 17-22.

Changnon, S.A. and Easterling, W.E., 1989. Measuring Drought Impacts: The Illinos Case1. JAWRA Journal of the American Water Resources Association, 25(1): 27-42.

Chen, L., 2003. Real Coded Genetic Algorithm Optimization of Long Term Reservoir Operation. Journal of the American Water Resources Association, 39(5): pp. 1157-1165.

Chen, Y. and Buerger, G., Precipitation Downscaling with Expanded Downscaling (EDS) Methodin the Ruhr Catchmentin Germany GermanyYouminChenEarth Sciences Centre, Gothenburg University, Sweden - BuergerPotsdam Institute for Climate Impact Research, GermanyEmail , from: http://www.pik-potsdam.de/avec/peyresq2003/posters/youmin_chen.pdf.

Chow, V., Maidment, D. and Mays, L., 1988. Applied hydrology. Mcgraw-Hill Book Company, New York.

Chung, C.-h. and Salas, J.D., 2000. Drought Occurrence Probabilities and Risks of Dependent Hydrologic Processes. Journal of Hydrologic Engineering, 5(3): pp. 259-268.

Chuntian, C., 1999 Fuzzy optimal model for the flood control system of the upper and middle reaches of Yangtze River. Hydrological Sciences Journal 44(4): pp. 573 - 585.

Cicogna, M.A., Fontane, D.G., Hidalgo, I.G. and Lopes, J.E., 2009. Multireservoir Simulation Using Multipurpose Constraints and Object-Oriented Software Design. In: S. Steve (Editor). ASCE, pp. 508.

Deb, K., 2001. Multi-objective optimization using evolutionary algorithms. Wiley; 1 edition (June 27, 2001).

Dikbas, F., FIRAT, M., KOC, A.C. and GÜNGÖR, M., 2010. Homogeneity Test for Turkish Temperature Series, BALWOIS 2010, Ohrid, Republic of Macedonia - 25, 29 May 2010.

Dorfman, R., 1962. "Mathematical models: The multi-structure approach," in Design of water resources systems. Edited by Maass, A. et al, Harvard University Press, Cambridge, Mssachusetts.

Dracup, J.A., 1991. Drought monitoring Stochastic Hydrology and Hydraulics, Volume 5(0931-1955 (Print) 1435-151X (Online)): pp. 261-266.

Dubrovin, T., Jolma, A. and Turunen, E., 2002. Fuzzy Model for Real-Time Reservoir Operation. Journal of Water Resources Planning and Management, 128(1): pp. 66-73.

Durbin, J., 1960. The fitting of time series models. review of the international institute of Statistics, 28: pp. 233-140.

European Commission, J.R.C., 2008a. Drought Management Plan report. From http://ec.europa.eu/environment/water/quantity/pdf/dmp_report.pdf.

European Commission, J.R.C., 2008b. European drought observatory.

Ferreira, A., Pardalos, P. and Gengler, M., 1996. An introduction to parallel dynamic programming, Solving Combinatorial Optimization Problems in Parallel. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 87-114.

Garcia, A.L. and Wagner, W., 2006. Generation of the Maxwellian inflow distribution. Journal of Computational Physics, 217(2): pp. 693-708.

Gbete, D. and Soumaila, M., 2007. Analysis of drought in Burkina Faso by using Standardized Precipitation Index, The Pyrenees International Workshop on Statistics, Probability and Operations Research SPO 2007 Jaca, September 12-15, 2007.

Ghafoor, A. and Hanif, S., 2005. Analysis of the Trade Pattern of Pakistan: Past Trends and Future Prospects. Journal of Agriculture & Social Sciences, 1(4): pp. 346-349

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning Addison-Wesley Professional; 1 edition (January 11, 1989).

Grinstead, C.M. and Snell, J.L., 1997. Introduction to Probability. American Mathematical Society, 510 pp.

Groisman, P.Y., 2005. Trends in Intense Precipitation in the Climate Record. Journal of Climate, 8: pp. 1326-150.

Hakimi-Asiabar, M., Ghodsypour, S.H. and Kerachian, R., Deriving operating policies for multi-objective reservoir systems: Application of Self-Learning Genetic Algorithm. Applied Soft Computing, 10(4): pp. 1151-1163.

Heino, R., 2004. Progress in the Study of Climatic Extremes in Northern and Central Europe Climatic Change, 42(1): pp. 151-181.

Hisdal, H. and Tallaksen, L.M., 2000. Drought Event Definition, Technical Report to the ARIDE project No.6, Department of Geophysics, University of Oslo, P.O. Box 1022 Blindern,N-0315 Oslo, Norway.

Huber, T., 1997. Introduction to Monte Carlo Simulation

Hübler, M., Klepper, G. and Peterson, S., 2008. Costs of climate change: The effects of rising temperatures on health and productivity in Germany. Ecological Economics, 68(1-2): pp. 381-393.

Hundexha, Y. and Bardossy, A., 2005. Trends in Daily Precipitation and Temperature Extremes Across Western Germany in the Second Half of the 20th Century. International Journal of Climatology 25: pp.189-1202.

IPCC, 2001. The IPCC Third Assessment Report, Climate Change 2001.

IPCC, 2007. Summary for Policymakers,Robert T. Watson, Daniel L. Albritton, Terry Barker,et.al,Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom

Ismail, N.A., Harun, S. and Yusop, Z., 2004. Synthetic Simulation of Streamflow and Rainfall Data Using Disaggregation Models.

Juran , A.A. and Arup, K.S., 2007. Artificial neural network model for synthetic streamflow generation. Water Resour Manage, 21(6): pp. 1015- 1029.

Kablan, A., 2009. Adaptive Neuro-Fuzzy Inference System for Financial Trading using Intraday Seasonality Observation Model. World Academy of Science, Engineering and Technology, 58: pp. 479-488.

Karabork, M.C., Kahya, E. and Komuscu, A.U., 2007. Analysis of Turkish precipitation data: homogeneity and the Southern Oscillation forcings on frequency distributions. Hydrological Processes, 21(23): pp. 3203-3210.

Keiser, D.T. and Grieffiths, J.F., 1997. Problems Associated with Homogeneity Testing in Climate Variation Studies: A Case Study of Temperature in the Northern Great Plains, USA. International Journal of Climatology, 17: pp. 497-510.

Kendall, D.R. and Dracup, J.A., 1992. On the generation of drought events using an alternating renewal-reward model Stochastic Hydrology and Hydraulics, 6(1): pp. 55-68.

Khadr, M., Morgenschweis, G. and Schlenkhof, A., 2009. Analysis of Meteorological Drought in the Ruhr Basin by Using the Standardized Precipitation Index. International Conference on Sustainable Water Resources Management (SWRM2009), Amsterdam - Netherland.

Khattree, R. and Rao, C.R., 2003. Statistics in industry. Handbook of statistics,. Elsevier, Amsterdam Boston, xxi, 1187 p. pp.

Kim, B.S., Kim, H.S. and Seoh, B.H., 2004. Streamflow simulation and skewness preservation based on the bootstrapped stochastic models. Stochastic Environmental Research and Risk Assessment, 18(6): pp. 386-400.

Kim, T.-W. and Valde´s, J.B., 2003. Nonlinear Model for Drought Forecasting Based on a Conjunction of Wavelet Transforms and Neural Networks. Journal of Hydrologic Engineering, 8(6): pp. 319-328.

Konak, A., Coit, D.W. and Smith, A.E., 2006. Multi-objective optimization using genetic algorithms. Reliability Engineering and System Safety 91: pp. 992–1007.

Krasovskaia, I., 1995. Quantification of the stability of river flow regimes. Hydrological Scienc, 40(5): pp. 587-598.

Labedzki, L., 2007. Estimation of Local Drought Frequency in Central Poland Using the Standardized Precipitation Index SPI. Irrigation and Drainage, 56: pp. 67-77.

Leemans, R. and Cmmer, W.P., 1991. The IIASA Database for Mean Monthly Values of Temperature, Precipitation, and Cloudiness on a Global Terrestrial Grid.

Lehner, B., Henrichs, T., Döll, P. and Alcamo, J., 2001. EuroWasser.

Limbrunner, J.F., 2001. Climate elasticity of streamflow in the United States. Water Resources Research, 37(6): pp. 1771-1781.

Lloyd-Hughes, B., 2002. The Long-Range Predictability of European Drought, University College London, London.

Loaiciga, H.A. and Leipnik, R.B., 1996. Stochastic Renewal Model of Low-Flow Streamflow Sequences Stochastic Hydrology and Hydraulics, 10(1): pp. 65-85.

Lohani, V.K. and Loganathan, G.V., 1997. An Early Warning System For Drought Management Using the Palmer Drought Index. Journal of the American Water Resources Association, 33(6): pp.1375-1386.

M.V.K. Sivakumar, a.D.A.W., DROUGHT PREPAREDNESS AND DROUGHT MANAGEMENT.

Maass, A. et al., 1970. Design of Water-Resource Systems : New Techniques for Relating Economic Objectives, Engineering Analysis, and Governmental Planning [Hardcover], Harvard University Press, Cambridge.

Macal, C.M., 2005. Model Verification and Validation, Workshop on "Threat Anticipation: Social Science Methods and Models", The University of Chicago and Argonne National Laboratory.

Makridakis, S.G. and Wheelwright, S.C., 1978. Forecasting : Methods and Applications. The Wiley/Hamilton series in management and administration. Wiley, Santa Barbara, Calif., xvi, 713 p. pp.

Maniak, U. and Renz, F.W., 1977. Optimal operation of the szstem of Ruhr Reservoirs. International Asscociation on Waterpollution Research, 10: 301/311.

Martin, Q.W., 1987 Optimal Daily Operation of Surface- Water Systems. Journal of Water Resources planning and Management, ASCE, 113 (4): pp. 453-470.

MathWorks9, Global Optimization Toolbox

Mathworks(1), Generating Data Using Flexible Families of Distributions.

McKee, T.B., Doesken, N.J. and Kleist, J., 1993. The relationship of Drought Frequency and Duration to Time Scales, 8th Conf. Applied Climatology, Anaheim, California, 17-22 January 1993.

McMahon, G.F., 2009. Models and Realities of Reservoir Operation. Journal of Water Resources Planning and Management, 135(2): pp. 57-59.

Menzel, L. and Bürger, G., 2002. Climate Change Scenarios and Runoff Response in the Mulde Catchment (Southern Elbe, Germany). Journal of Hydrology, 267(1-2): pp. 53-64.

Metaxiotis, K., Psarras, J. and Samouilidis, E., 2003. Integrating fuzzy logic into decision suppport systems: current research and future prospects. Information Management & Computer Security, 11(2): pp. 53-59.

Moberg, A., 2006. Indices for Daily Temperature and Precipitation Extremes in Europe Analyzed for the Period 1901–2000. Journal of Geophysical Research, 111(D22106).

Moberg, A. and Jonesa, P.D., 2005. Trends in Indices for Extremes in Daily Temperature and Pecipitation in Central and Western Europe. International Journal of Climatology, 25: pp. 1149-1171.

Mohan, S. and Raipure, D.M., 1992. Multi Objective Analysis of Multi Reservoir

systems". Journal of Water Resource Planning and Management, ASCE, 118(4): pp. 356-370.

Montgomery, D.C., Jennings, C.L. and Kulahci, M., 2008. Introduction to Time Series Analysis and Forecasting. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, N.J., xi, 445 p. pp.

Morgenschweis, G., Brudy-Zippelius, T. and Ihringer, J., 2003. Operational Water Quantity Management in a River Basin. Water Science and Technology 48(10): pp. 111-118.

Morgenschweis, G., Strassen, G. , Patzke, S. and Schwanenberg, D., 2007. Estimation of the Impact of Possible Climate Change on the Management of the Reservoirs in the Ruhr Catchement Basin. Report from Annual Report Ruhrwassermenge 2006. Ruhrverband Essen (2007). pp. 32-50

Moyé, L.A. and Kapadia, A.S., 1994. Predictions of Drought Length Extreme Order Statistics Using Run Theory. Journal of Hydrology, 169(1-4): pp. 95-110

Mujumdar, P.P. and Teegavarapu, R., 1998. A Short-Term Reservoir Operation Model for Multicrop Irrigation. Hydrologic Sciences Journal, 43 pp. 479-494.

Namchaiswadwong, K., Bhaktikul, K. and Kongjun, T., 2006. Optimization of Multiple Reservoir Releases Using Genetic Algorithms: Case Study of Mae Klong River Basin, Thailand 7th International Micro Irrigation Congress, Sept 10 - 16 2006, PWTC, Kuala Lumpur.

NDMC, N.D.M.C., 2006. Defining Drought: Overview. National Drought Mitigation Center, University of Nebraska–Lincoln.

Obasi, G.O.P., 1994. Wmos Role in the International Decade for Natural Disaster Reduction. Bulletin of the American Meteorological Society, 75(9): pp.1655-1661.

Palmer, R.N. and Holmes, K.J., 1988. Operational Guidance During Droughts: Expert System Approach Journal of Water Resources Planning and Management, ASCE,, 114(6): pp. 647-666.

Panigrahi, D.P. and Mujumdar, P.P., 2000 Reservoir Operation Modelling with Fuzzy Logic. Water Resources Management **14**: pp. 89-109.

Panu, U.S. and Sharma, T.C., 2002. Challenges in Drought Research: Some Perspectives and Future Directions. Hydrotogical Sciences, 47(S): pp. S19-S30.

Peterson, T.C. et al., 1998. Homogeneity adjustments of in situ atmospheric climate data: A review. International Journal of Climatology, 18(13): pp. 1493-1517.

Pettitt, A.N., 1979. A non-parametric approach to the change-point detection. Applied Statistics, 28

(2): pp. 126-135.

Phien, H.N. and Ruksasilp, W., 1981. A Review of Single-Site Models for Monthly Streamflow Generation. Journal of Hydrology, 52(1-2): pp. 1-12.

Prairie, J. and Rajagopalan, B., 2007 Stochastic Streamflow Generation Incorporating Paleo-Reconstruction, Proceedings the World Environmental & Water Resources Congress, Tampa, FL.

Randall, D., Houck, M.H. and Wright, J.R., 1990. Drought management of existing water supply system. Journal of Water Resources Planning Management, ASCE, 114(1): pp. 1-20.

Rao, A.R. and Padmanabhan, G., 1984. Analysis and Modeling of Palmer's Drought Index Series Journal of Hydrology, Volume 68(1-4): pp. 211-229

Reddy, J.R., 2005. A Textbook of Hydrology . Laxmi Publications.

Reddy, M.J., 2006. Swarm Intelligenca and Evolutionary Computation for Single and Multiobjective Optimization in Water Resources systems, Indian Institute of Science.

Rossi, G., Vega, T. and Bonccorso, B., 2007. Methods and Tools for Drought Analysis and Management. Springer; 1 edition (December 3, 2007), 418 pp.

Ruhrverband-online-Report, (Ruhr Association), Biggetalsperre , Ruhrverband, Essen, Germany from:
     http://www.ruhrverband.de/fileadmin/pdf/presse/gewaesser/biggetalsperre.pdf

S.Mohan and Prasad, M.A., 2006. Fuzzy Logic Model for Multi Reservoir Operation, Advances in Geosciences(A 5-Volume Set) - Volume 1 Solid Earth (SE) (pp 117-126)

Saglam, Y., 2008. Introduction to Probability Theory for Graduate Economics, from http://home.uchicago.edu/~hickmanbr/uploads/chapter5_1.pdf.

Sahin, S. and Cigizoglu, H.K., 2010. Homogeneity Analysis of Turkish Meteorological Data set. Hydrological Processes, 24(8): pp. 981-992.

Salas, J.D. and Frevert, D.K., Stochastic Analysis Modeling and Simulation (SAMS) , Colorado State University http://www.sams.colostate.edu/index.html.

Saldariaga, J. and Yevjevich, V., 1970. Application of Run-Lengths to Hydrologic Series.

Santos, M.J. and Henriques, R., 1999. Analysis of the European Annual Precipitation Series. Technical Report to the ARIDE project No.3: supplement to Work Package 2 Hydro-meteorological Drought Activity 2.4 Regional Drought Distribution Model Water Institute, DSRH - Av. Almirante Gago Coutinho 30 1049 - 066 Lisbon, Portugal.

Schneider, U., Fuchs, T., Meyer-Christoffer, A. and Rudolf, B., 2008. Global Precipitation Analysis Products of the GPCC Global Precipitation Climatology Centre (GPCC) ,Deutscher Wetterdienst, Offenbach a. M., Germany

Sen, Z., 1978. A mathematical model of monthly flow sequences. Hydrological Sciences-Bulletin-des Sciences Hydrologiques, 23(2).

Sen, Z., 1990. Critical Drought Analysis by Second-Order Markov Chain. Journal of Hydrology, 120(1-4): pp. 83-202.

Shah, V.A., 2009. Adaptive Neuro-Fuzzy Inference System for Effect of Wall Capacitance in a Batch Reactor. Advances in Fuzzy Mathematics , Research India Publications, 4(1): pp. 69-75.

Shaw, E.M., 1994. Hydrology in Practice. Chapman & Hall, 2-6 Boundary Row, London SE1 8HN, UK.

Shrestha, B.P., Duckstein, L.E. and Stokhin, Z., 1996. Fuzzy Rule-Based Modeling of Rreservoir Operation Journal of Water Resources Planning Management 122(4): pp. 262-269.

Shumway, R.H. and Stoffer, D.S., 2000. Time Series Analysis and its Applications. Springer texts in statistics. Springer, New York, xiii, 549 p. pp.

Simolo, C., Brunetti, M., Maugeri, M. and Nanni, T., 2009. Improving Estimation of Missing Values in Daily Precipitation Series by a Probability Density Function-Preserving Approach. International Journal of Climatology, 30(10): pp. 1564-1576.

Singh, V.P. and Frevert, D.K., 2001. Mathematical Models of Small Watershed Hydrology and Applications, 972 pp.

Singhal, M.K., Singhal, H.S.S. and Maheshwari, J.B., 1980 Mathematical Modelling of Inflows to the Matatila Reservoir. Hydrological Forecasting - (Proceedings of the Oxford Symposium, April 1980 IAHS-AISH Publ. no. 1 29.

Sivakumar, M.V.K., Motha, R.P. and Das, H.P., 2005. Natural Disasters and Extreme Events in Agriculture. Springer Berlin Heidelberg.

Stahl, K., 2001. Hydrological Drought - A Study Across Europe. Ph.D Thesis Thesis, University of Freiburg, Freiburg.

Steele-Dunne, S. et al., 2008. The impacts of climate change on hydrology in Ireland. Journal of Hydrology, 356(1-2): pp. 28-45.

Storch, H.v. and Zwiers, F.W., 2001. Statistical Analysis in Cimate Research. Cambridge University Press, Cambridge, U.K. ; New York, x, 484 p. pp.

Szentimrey, T., 1999. Multiple Analysis of Series for Homogenization (MASH), Proc. Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary, pp. pp. 27-46.

Tallaksen, L.M. and Lanen, H.A.J.V., 2004. Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater.

Tang, Y.F., Zhang, Z.Y. and Fan, G.Q., 2005. [Identification of rhubarb samples based on IR spectra by using Takagi-Sugeno fuzzy systems]. Guang Pu Xue Yu Guang Pu Fen Xi, 25(4): pp. 521-4.

Tank, A.M.G.K., Koennen, G.P. and Selten, F.M., 2005. Signals of Anthropogenic Influence on European Warming as Seen in the Trend Patterns of Daily Temperature Variance. International Journal of Climatology, 25: pp. 1-16.

Treiber, B. and Schultz, G.A., 1976. Comparison of Required Reservoir Storages Computed by the Thomas-Fiering Model and the 'Karlsruhe Model' Type A and B. Hydrological Sciences-Bulletin, XXI, 1 (3): pp. 177-185.

Tu, M.-Y., Hsu, N.-S., Tsai, F.T.-C. and Yeh, W.W.-G., 2008. Optimization of Hedging Rules for Reservoir Operations. Journal of Water Resources Planning and Management, 134(1): pp. 3-13.

Ubeda, J.R. and Allan, R.N., 1994. Stochastic Simulation and Monte Carlo Methods Applied to the Assessment of Hydro-Thermal Generating System Operation. Top, 2(1): pp. 1-24.

UNEP, U.N.E.P., 2004. Impacts of Summer 2003 Heat Wave in Europe.

Venugopal, C., Devi, S.P. and Rao, K.S., 2010. Predicting ERP User Satisfaction—an Adaptive Neuro Fuzzy Inference System (ANFIS) Approach Intelligent Information Management **2**: pp. 422-430

Vicente-Serrano, S.M. and L´opez-Moreno, J.I., 2005. Hydrological Rresponse to Different Time Scales of Climatological Drought: An Evaluation of the Standardized Precipitation Index in a Mountainous Mmediterranean Basin. Hydrology and Earth System Sciences Discussions, 2: pp.1221-1246.

Von Neumann, J., 1941. Distribution of the ratio of the mean square successive difference to the variance. Annals of Mathematical Statistics, 13: pp. 367-395.

Wang, L. et al., 2010. Development of an integrated modeling system for improved multi-objective reservoir operation. Frontiers of Architecture and Civil Engineering in China, 4(1): 47-55.

Weisstein(1), E.W., Monte Carlo Method, from MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/MonteCarloMethod.html

Weisstein(3), E.W., Pearson System, from MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/PearsonSystem.html.

Wikipedia(1), Gamma distribution, From Wikipedia, the free encyclopedia :http://en.wikipedia.org/wiki/Gamma_distribution.

Wilhite, D.A., 2005. Drought and Water Crises : Science, Technology, and Management Issues. Taylor & Francis, Boca Raton, xxiv, 406 p. pp.

Wilhite, D.A., 2009. Defining Drought: The Challenges for Early Warning Systems, Inter-Regional Workshop on Indices and Early Warning Systems for Drought, Nebraska - USA.

Wilhite, D.A. and Glantz, M.H., 1985. Understanding the Drought Phenomenon: the Role of Definitions. Water International, 10(3): pp. 111-120.

Wilhite, D.A., Sivakumar, M.V.K. and Wood, D.A., 2000. Early Warning Systems for Drought Preparedness and Drought Managementx, Proceedings of an Expert Group Meeting. Lisbon, Portugal, September 5-7. World Meteorological Organization, Geneva, Switzerland.

Wu, H., Hayes, M.J., Weiss, A. and Hu, Q., 2001. An evaluation of the Standardized Precipitation Index, the China-Z Index and the statistical Z-Score. International Journal of Climatology, 21(6): pp. 745-758.

Yaffee, R. and Magee, M., 2000. Time Series Analysis and Forecasting with Applications of SAS and SPSS. Academic Press, San Diego:.

Yan, J., Chenghu, Z. and Weiming, C., 2007. Streamflow Trends and Hydrological Response to Climatic Change in Tarim Headwater Basin. Journal of Geographical Sciences, 17(1): pp. 511-61.

Yegireddi, S. and Kumar, A., 2008. Geoacoustic Inversion Using Adaptive Neuro-Fuzzy Inference System. Comput Geosci, 12: pp. 513-523.

Yesilirmak, E., Akcay, S., Dagdelen, N., Gurbuz, T. and Sezgin, F., 2009. Quality Control and Homogeneity of Annual Precipitation Data in Buyuk Menderes Basin, Western Turkey. Fresenius Environmental Bulletin, 18(9A): pp. 1748-1757.

Yoo, W.-H., Kim, H.-S. and Seoh, B.-H., Applications of Standardized Precipitation Index to Streamflows, from: http://kfki.baw.de/conferences/ICHE/2000-Seoul/pdf/201/PAP_221.PDF.

Yue, S., Pilon, P. and Cavadias, G., 2002. Power of the Mann-Kendall and Spearman's Rho Tests for Detecting Monotonic Trends in Hydrological Series. Journal of Hydrology, 259(1-4): pp. 254-271.

Zadeh, L.A., 1973. Outline of a New Approach to Analysis of Complex Systems and Decision Processes. Ieee Transactions on Systems Man and Cybernetics, Smc3(1): pp. 28-44.

Appendix A

Software for the calculation and Analysis of the Standardized Precipitation Index

A.1 Possibilities of SPI _Analysis

As presented in chapter 3, the calculation and the analysis of the standardized precipitation index (SPI) are complex and not so easy to be done with several precipitation time series. The developed program named "SPI_Analysis" is a program to calculate and analyze the standardized precipitation index (SPI). The main objectives of the program are:

Calculation of the SPI values for a given precipitation data series of a length of least 30 years with daily records.

Detection whether a drought event exists in data series.

Classification of the drought events according to its intensity (moderate- severe- extreme).

The program is easy to use for a given data series. The user only needs to prepare the data series to meet the required format.

A.2 Mathematical Core of SPI_Analysis

McKee (McKee et al., 1993) developed the Standardized Precipitation Index (SPI) for the purpose of defining and monitoring drought. Among others, the Colorado Climate Center, the Western Regional Climate Center and the National Drought Mitigation Center use the SPI to monitor current states of drought in the United States. The nature of the SPI allows an analyst to determine the rarity of a drought or an anomalously wet event at a particular time scale for any location in the world that has precipitation records. In most cases, the Gamma distribution best models observational precipitation data. Details about the SPI methodology are presented in chapter 3.

A.3 How to Start an Application

The SPI_Analysis program is valid for applying to any precipitation data series in any basin. User can start the program by double clicking the SPI_Analysis icon with left mouse button. The SPI_Analysis title screen is shown in figure A.1. Once the program is started, the main screen appears. In the File-menu the user can select the option "Load Data" by one click on the button "Load Data". Then the user has to select a data file (figure A.2) which must satisfy the required format. The requirements for a data series to be analyzed and the required format of the input file are presented in section A.4.
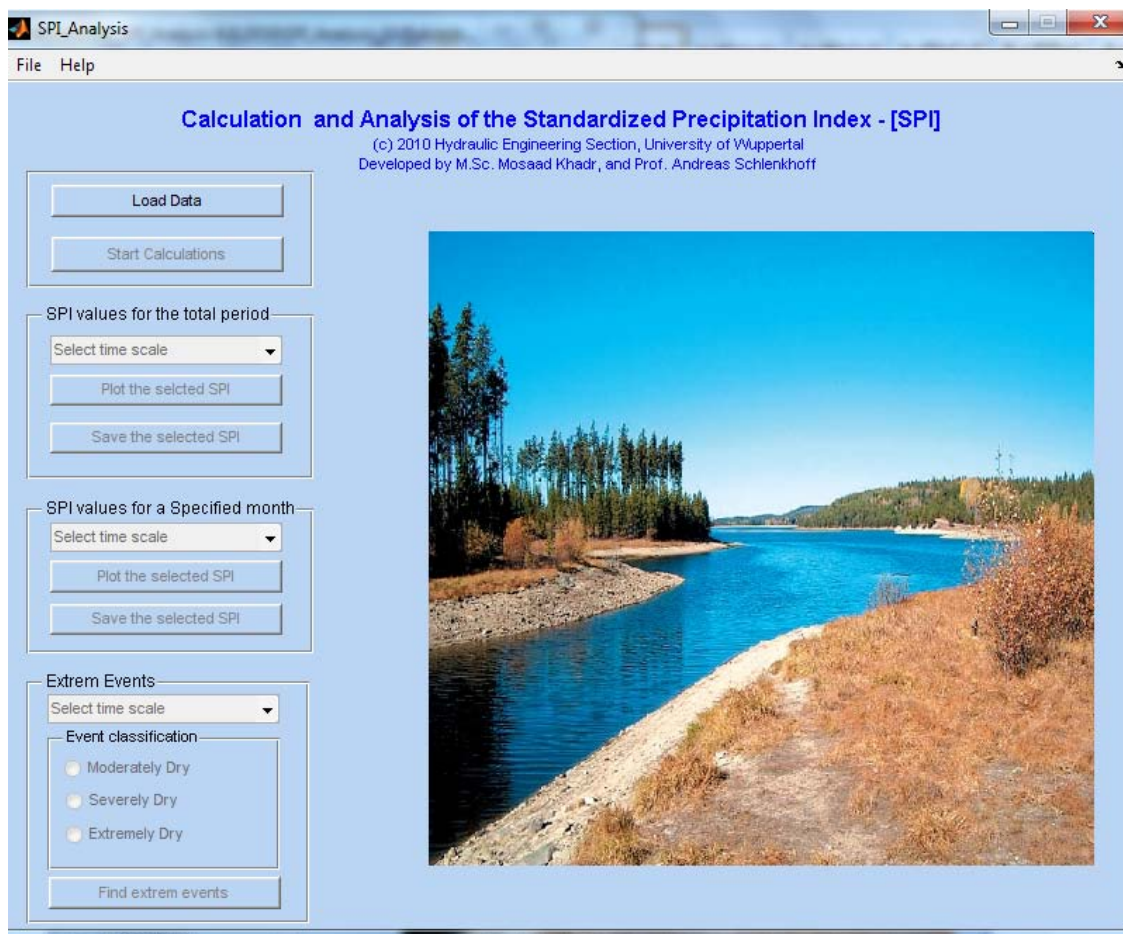
Figure A.1: SPI_Analysis initial screen
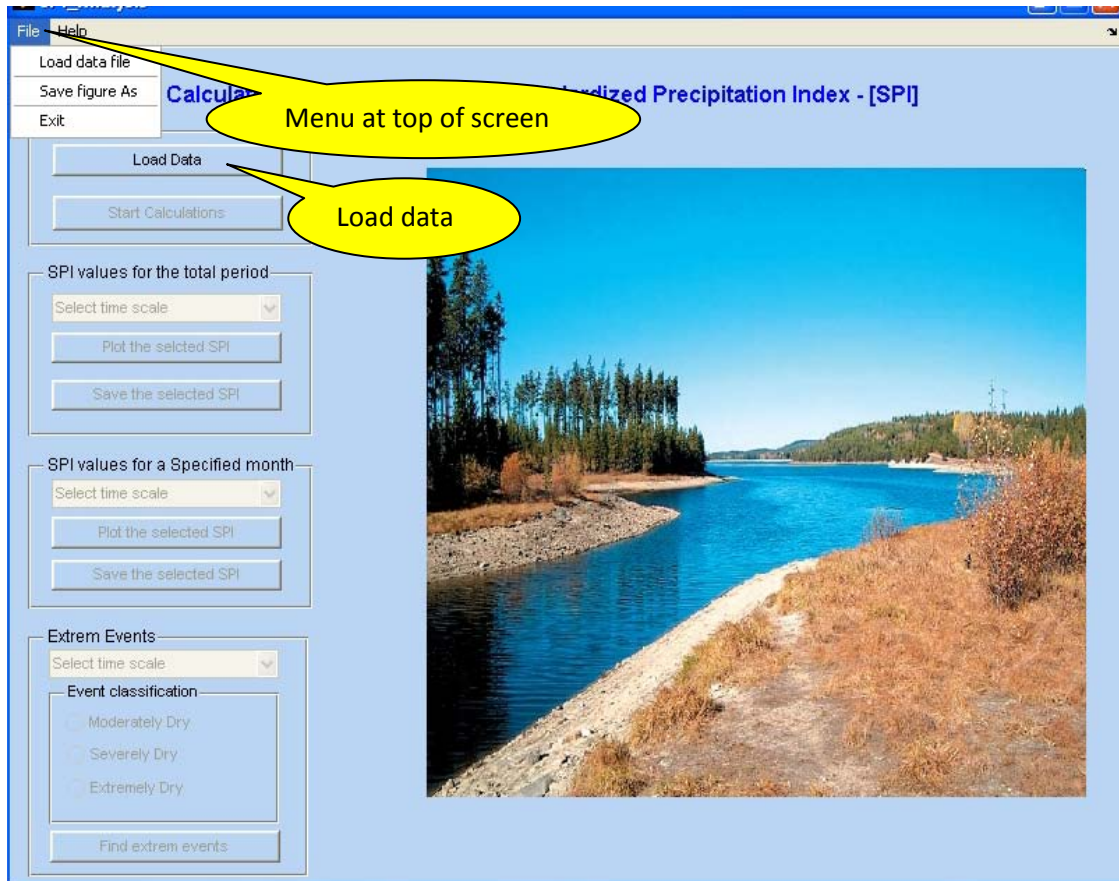
Figure A.2: SPI_Analysis title screen to select type of data

A.4. Required Information Content of Data Series

Each data series in the input file must satisfy the requirements in order to avoid any error. The requirements are:

The SPI_Analysis can only read data series in Microsoft Excel (.xls)

Each precipitation data series consists of two columns. The first one is the date with the required format "yyyymmdd"(see figure A.3) and the second one is the observed data (mm)

Missing data are not allowed, any missing data leads to an error.

The accepted data are daily or monthly observations.

Above requirements must be satisfied for all data series which the user wants to analyze.

Figure A.3: Sample of input data

A.5 Calculation of the SPI Index

A.5.1. Define Input Data Series

After choosing the precipitation data series, the button "Start Calculation" will be available as shown in figure A.4. One click on this button and all calculations required for SPI analysis will be done and the pop-up menu" Select time scale" will be available as shown in figure A.5. The user has two options: the first one is to analyze the SPI index for the total period (consecutive months) or for a specified month.

A.5.2. SPI Index for Consecutive Months

From the pop-up menu "Select time scale" under the title "SPI values for the total period "the user can choose one of six time scales, namely SPI_1, SPI_3, SPI_6 and SPI_9, SPI_12 and SPI_24 as shown in figure A.6. After selection of time scale, all calculation will be done for the selected time scale and then the user can plot the results by one click on the button "Plot Results". For example if the user selects the SPI_3 time scale the results will be as shown in figure A.7. Also the user can save the obtained results by clicking the button" Save Results" then will get message with the location of the saved results. A sample of saved results is shown in figure A.9. The first column presents the years, the second one presents the months and the third one presents the SPI values.
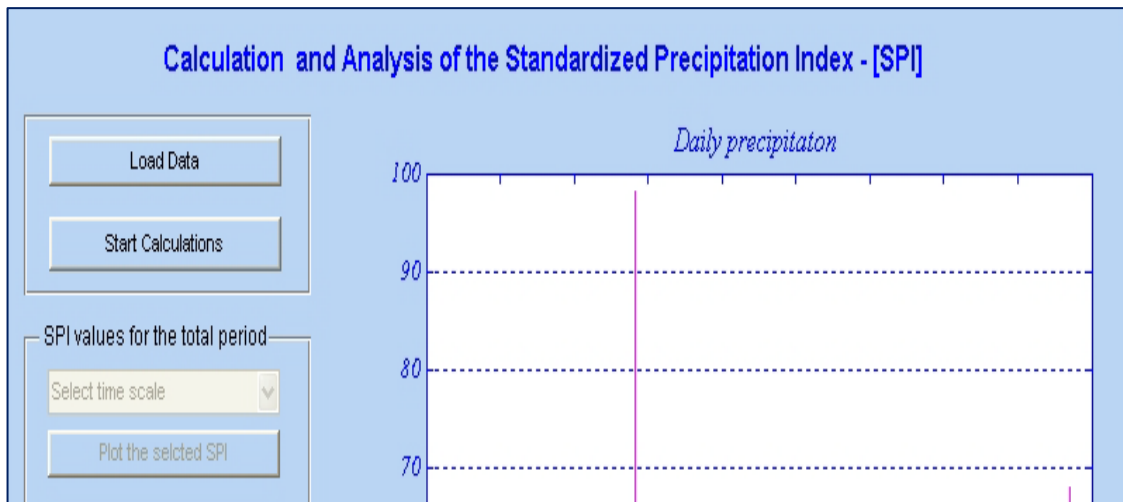
Figure A.4: Start calculations of the SP index
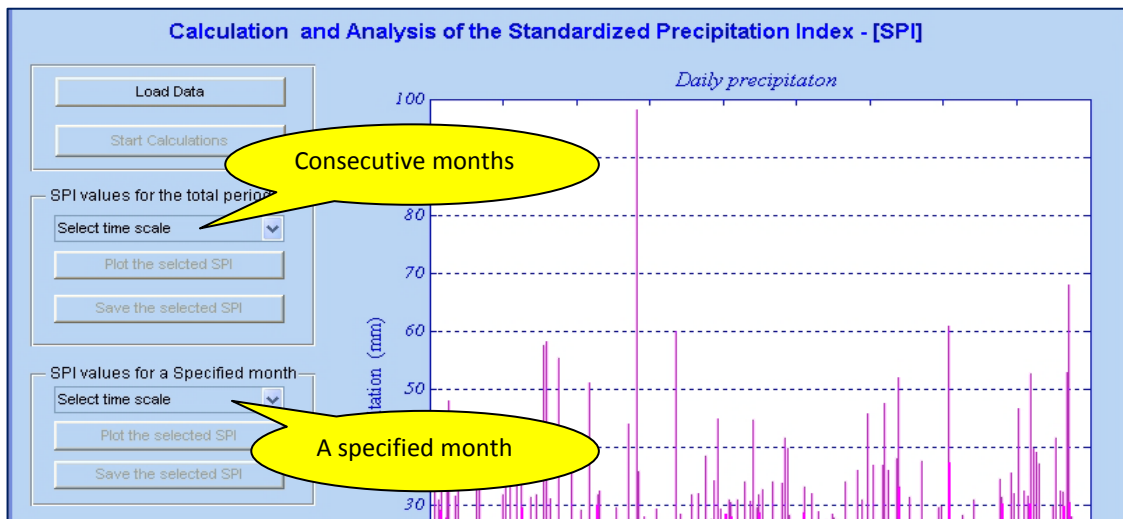
Figure A.5: Pop-up menu to select time scale
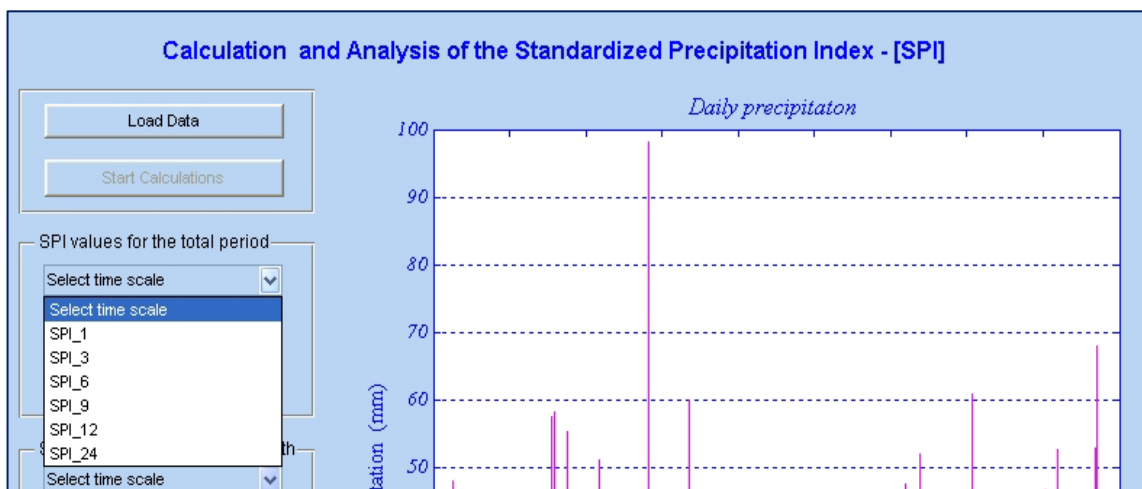
Figure A.6: Pop-up menu to select time scale (consecutive months)
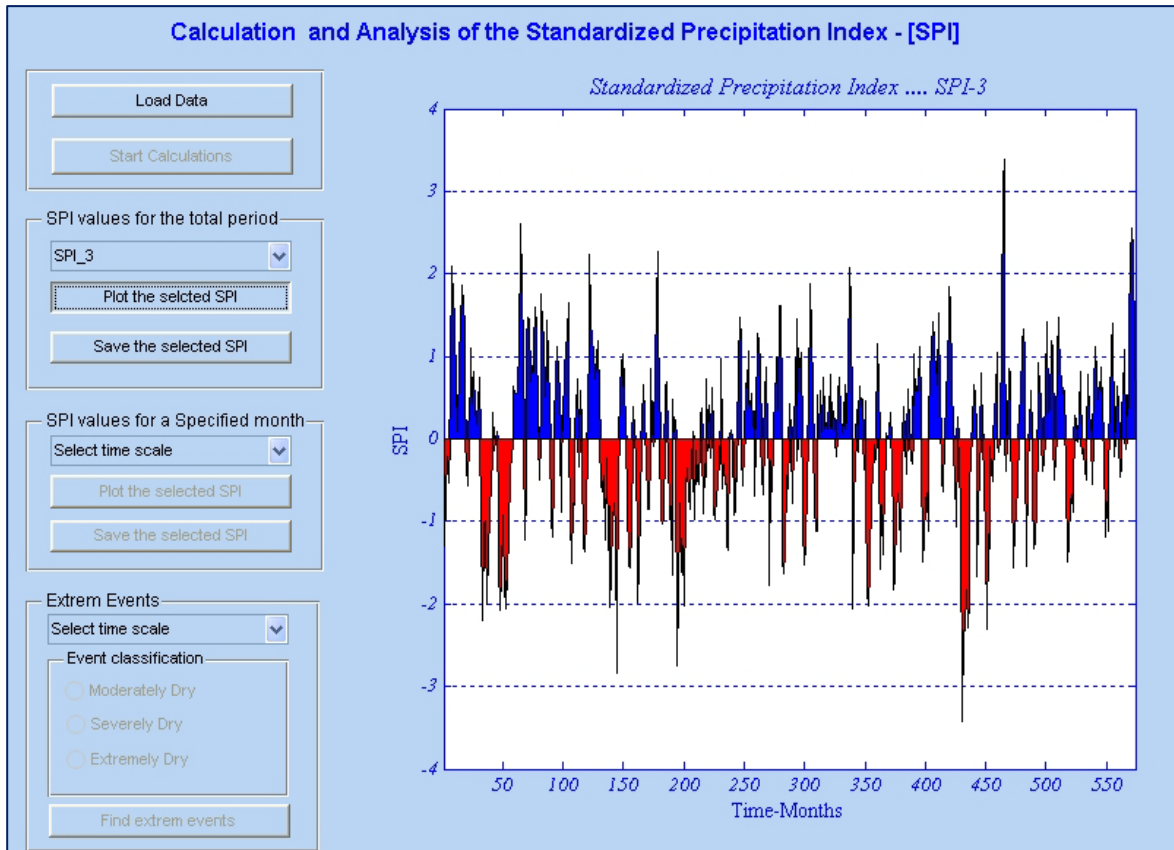
Figure A.7: SPI time series based (Three months time step – SPI_3)

Figure A.8: Results of SPI calculations (Three months time step – SPI_3)

A.5.3. SPI Index for a Specified Month

From the pop-up menu "Select time scale" under the title" SPI values for a specified month" the user can choose one of 72 scales i.e. 6 time scales for each month (figure A.9). After selection of time scale, all calculation will be done for the selected time scale and then the user can plot the results by one click on the button "Plot Results". For example if the user selects the SPI_3_Jan time scale (SPI_3 for the month January) the results will be as shown in figure A.10. The user also can save the obtained results by clicking the button" Save Results" then will get message with the location of the saved results. A sample of saved results is shown in figure A.11. The first column presents the years; the second one presents the corresponding SPI values.
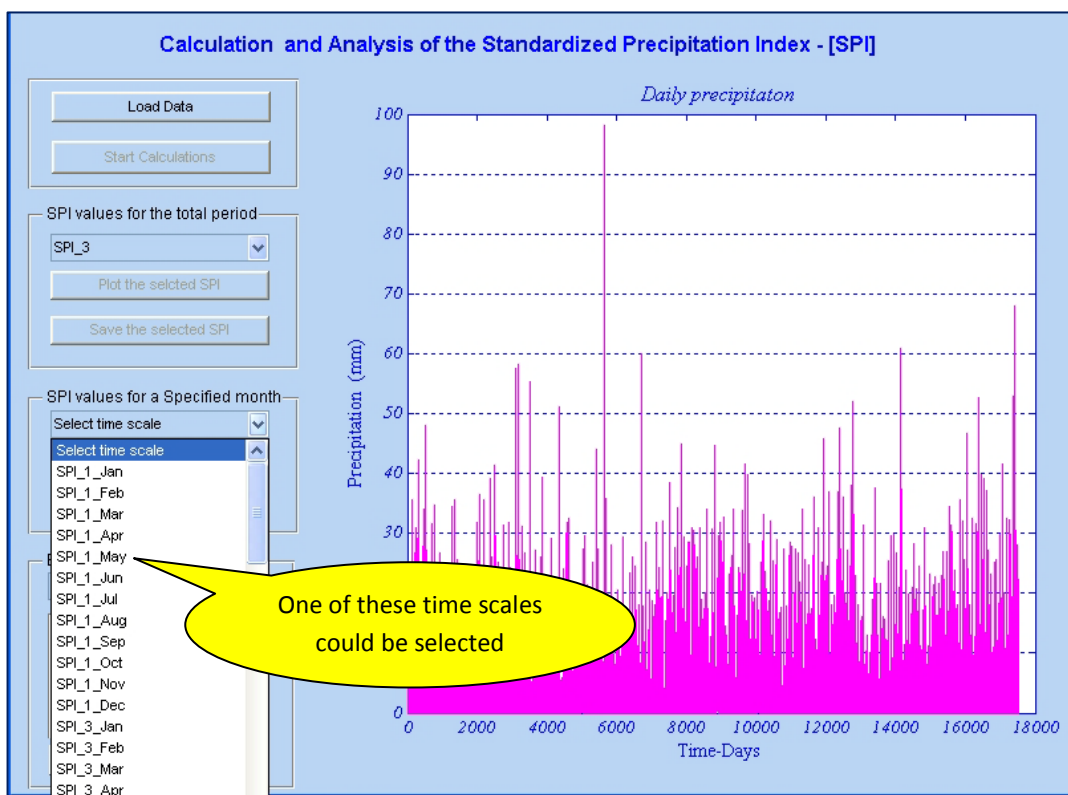
Figure A.9: Pop-up menu to select time scale (A specified month)
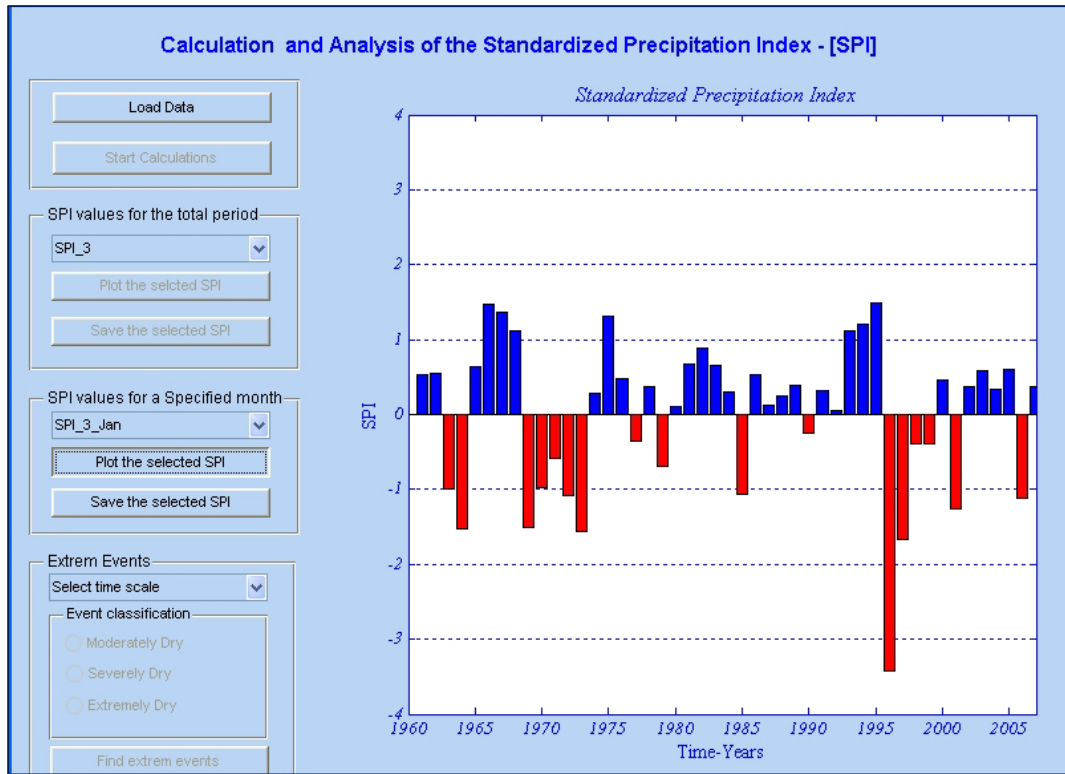
Figure A.10: Drought severity index values representative based on three months SPI values SPI-3-Jan. (November, December and January)

Figure A.11: Results of SPI calculations(Three months time step – SPI_3_Jan)

A.5.4 Detection of Extreme Events

The classification of drought events based on the SPI index. Under the title "Extreme Events" the user has to select one of the time scales. Once the user selects the required time scale, the three radio buttons (Moderately Dry- Severely Dry –Extremely Dry) will be available and the user has to check one of it in order to active the button "Find Extremes". Once the user clicks the button "Find Extremes", the results will be saved automatically and a message contains the location of the saved results will appear. A sample of results is shown in figure A.12. The first column presents the years, the second one presents the month in which the selected event happened and the third column presents the corresponding SPI value.

	A	B	C	D
1	Drought Events			
2	Year	Month	SPI value	
3	1960	4	-1.29	
4	1962	12	-1.02	
5	1963	2	-1.11	
6	1963	4	-1.32	
7	1964	9	-1.17	
8	1965	10	-1.22	
9	1967	9	-1.19	
10	1969	11	-1.29	
11	1969	12	-1.37	
12	1971	5	-1.22	
13	1971	7	-1.36	
14	1972	1	-1.08	
15	1973	3	-1.11	
16	1973	7	-1.28	
17	1975	4	-1.03	
18	1976	9	-1.29	
19	1979	10	-1.30	
20	1979	11	-1.34	
21	1983	8	-1.45	
22	1983	10	-1.37	
23	1985	1	-1.06	
24	1985	3	-1.30	
25	1985	11	-1.12	
26	1985	12	-1.12	
27	1990	7	-1.39	

Drought_events / Sheet2 / Sheet3

Figure A.12: Detection of drought events

Appendix B

Software Package for Meteorological Drought Forecasting Using Stochastic Models

B.1 Introduction

Occurrences of droughts all over the world are natural phenomena. Droughts represent an increasing hazard in many countries. Consequently, it is of the utmost importance to utilize efficient methods for drought events forecasting in order to assess and reduce such natural water hazards to the minimal or manageable level. In this study a software package named Drought_Forecasting has been developed.

The linear stochastic models ARIMA and multiplicative Seasonal Auto Regressive Integrated Moving Average (SARIMA) model have been used to forecast droughts based on the procedure of model development. The models have been applied to forecast droughts using standardized precipitation index (SPI) series in Ruhr river basin. The predicted results using the best models have been compared with the observed data and with the predicted results obtained by using the well known software SPSS. The predicted results show reasonably good agreement with the actual data with reasonably accuracy.

B.2 Possibilities of Drought_Forecasting

As presented in chapter 5, the forecasting of the standardized precipitation index (SPI) using stochastic models is not so easy to be done with several SPI data series. There are many software packages which are used for time series forecasting. One of these programs is the SPSS package which was used in drought forecasting in the previous chapter. The SPSS package has many tools, not only time series forecasting, which need an experienced user. However one of the advantages of the developed program, which named Drought_Forecasting, is its simplicity. The main objectives of the program are:

Forecasting of the SPI values for a given SPI data series using ARIMA model.

Forecasting of the SPI values for a given SPI data series using SARIMA model.

The program is easy to use; the user needs only to prepare the data series to meet the required format.

B.3 Mathematical Core of the developed program

The Drought_forecasting program contains mainly tow models; the Auto Regressive Integrated Moving Average (ARIMA) model and the Seasonal Auto Regressive Integrated Moving Average (SARIMA) model. Both of the two models assume the time series is stationary. With the appropriate modification, nonstationary series can also be studied with the two models (See chapter 4). Details about ARIMA and SARIMA model are presented in chapter 4.

B.4 Applications of the Program

The Drought_forecasting is valid for applying to any SPI data series. The user can start the model by double clicking the Drought_forecasting icon with lift mouse button. The Drought_forecasting title screen is shown figure B.1. In the File-menu the user can select the option "Load Data" by one click on the button "Load Data". Then the user has to select a data file which must have the required format. Section B.5 presents the requirements for a data series to be analyzed and the required format of the input file.

Figure B.1: Drought_Forecasting initial screen

B.5. Required Information Content of Data Series

Each data series in the input file must satisfy the requirements in order to avoid any error. The requirements are:

The Drought_forecasting can only read data series in Microsoft Excel (.xls),

Each SPI data series consists of one column,

Missing data are not allowed and any missing data leads to an error.

Above requirements must be satisfied for all data series which the user wants to analyze.

In the following section an illustration will be presented using SPI_6 time series which has been used in the previous chapter. As presented before in the previous chapter, the ARIMA model has been applied to SPI_3, however for SPI_6 time series the SARIMA model has been applied.

B.6 Model Identification (SPI_6)

After loading the data series (SPI_6), the ACF and PACF buttons will be visible then the user can plot both ACF and PACF. The next step is to determine whether the series is stationary or not by considering the graph of ACF (figure B.2). If a graph of ACF of the time series values either cuts off fairly quickly or dies down fairly quickly, then the time series values should be considered stationary. If a graph of ACF dies down extremely slowly, then the time series values should be considered non-stationary. If the series is not stationary, it can often be converted to a stationary series by differencing. That is, the original series is replaced by a series of differences. The next step is to check the PACF diagram (figure B.3) to find out the significant spikes which could give an initial estimation for the suitable model. If the significant spikes are for n consecutive lags only, then ARIMA model could be applied. If these significant spikes for n consecutive lags and there are another significant spikes at lags *k, 2k, ..., ik* then this is an indication to apply seasonal ARIMA model (SARIMA).
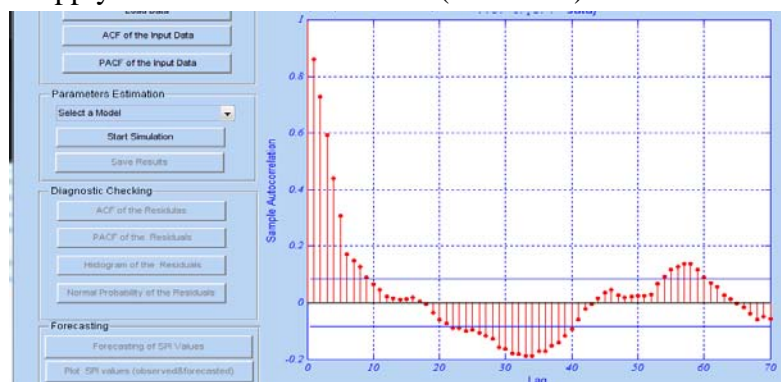
Figure B.2: ACF plot used for the selection of candidate models for SPI_6 series
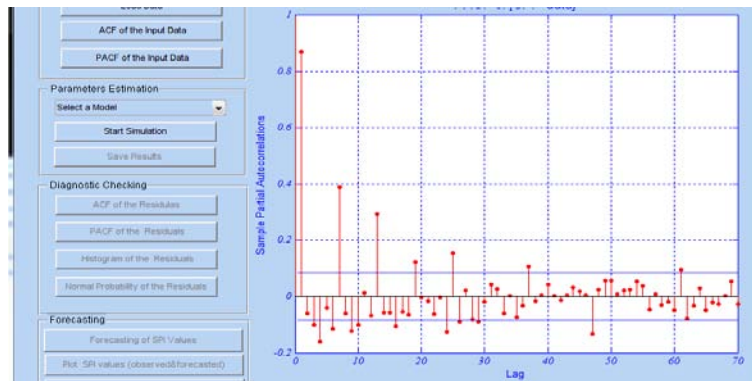
Figure B.3: PACF plot used for the selection of candidate models for SPI_6 series

B.7 Estimation and Optimization of the Parameters (Non-Seasonal and Seasonal Parameters, p, q, and P, Q)

As shown in figure B.2, the ACF is damping out with mixture of sine and exponential curve. The first value is significant in PACF which indicates an AR (1) as non-seasonal part of model. Also in the PACF, there are significant spikes presented near lag 6, 12 and 18 which indicates a SARIMA model. Alternative SARIMA models were identified by considering the ACF and PACF graphs of the SPI series. This indicates a possible SARIMA $(p, d, q)(P,D,Q)_s$ models with $p = 1:5$, $d = 0:1$, $q = 1:5$, $P =1:5$, $D = 0:1$ and $Q = 1:5$. All the combinations were examined to determine the best model out of these candidate models.

One of the most famous problems in the use of stochastic models, such as ARIMA or SARIMA, is that of the optimization of its parameters, i.e. the finding of the best model. This always is done by trial and error, i.e. to try several models with different parameters to find the model with minimum Low Akaike Information Criteria (AIC). In the developed software, an optimization method was applied to find out the best model for a given range for each parameter as shown in the following section.

B.7.1 Seasonal Model SARMA *(p, q) (P, Q)$_s$*

After selecting SARIMA model from the pop-up menu "select a model" the button "start simulation" will be visible (figure B.4). One click to this button, an input dialog will appear as shown in figure B.4. The user could easily put range for each parameter of the following parameter:

The auto regressive parameter *p* from 1 to 5, the moving average parameter q from 1 to 5, the seasonal auto regressive parameter *P* from 1 to 5, the seasonal moving average parameter *Q* from 1to 5 and the seasonal period *S*=6

The SARIMA model must contain at least one non-zero parameter in the seasonal part.

B.7.2 Estimation of the Model Parameters

When the user clicks the button "save results" all results will be automatically saved and the user get a message defines the location of the saved results. All results saved in one file (.xls). This file contains the best values of the parameters *(p,q,P,Q, AIC)* and also contains the obtained residuals of the model as shown in figure B.5. Results showed that the model SARIMA $(1,0,5)(1,0,1)_6$, with *AIC=557*, is the best one for the selected input range of the model's parameters.
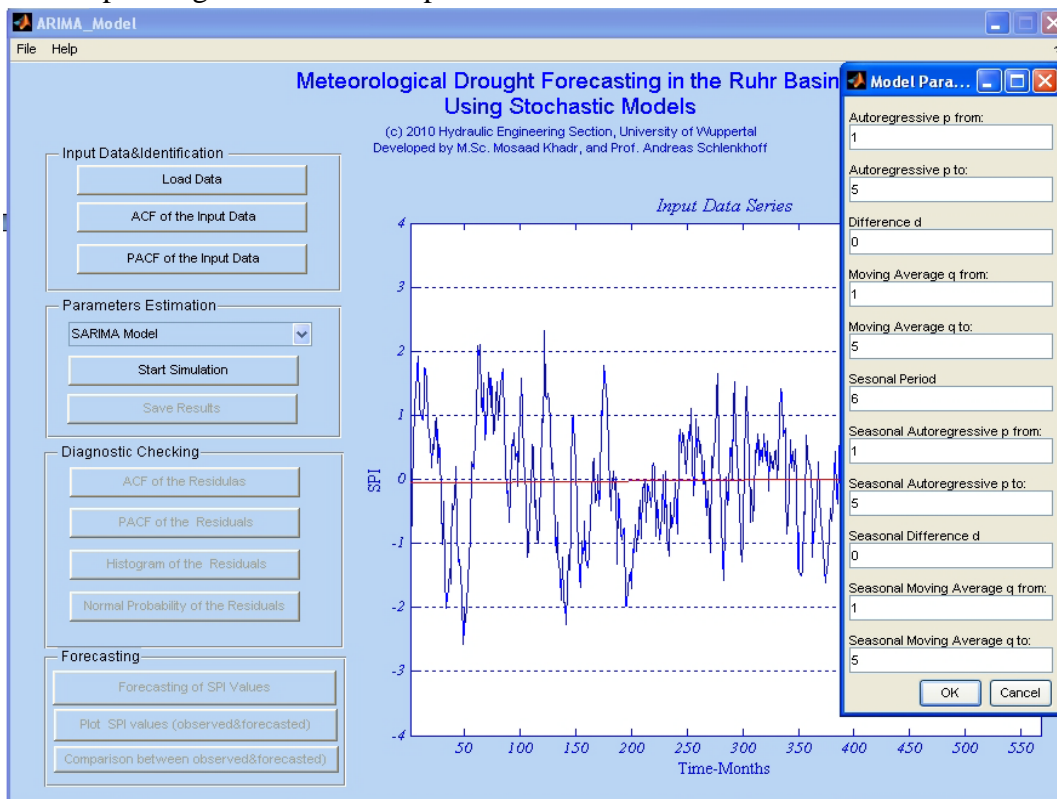
Figure B.4: Model Parameters – SARIMA Model
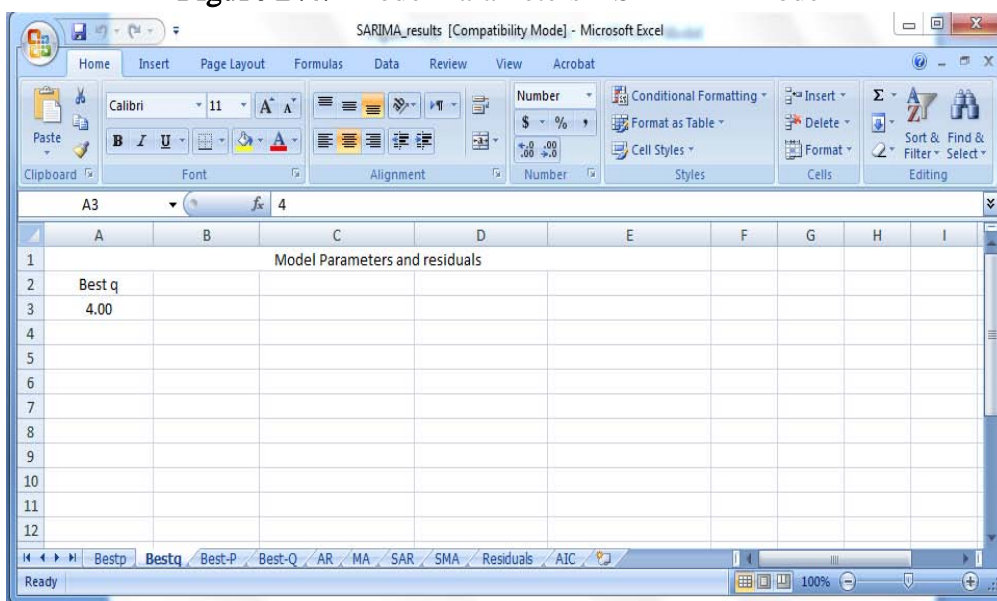
Figure B.5: Results of SARIMA model- Model Parameters.

B.8 Diagnostic Checking

Once the appropriate model has been fitted, the user can examine the goodness of fit. If the fitted model is adequate, the residuals should be approximately white noise. The theoretical ACF and PACF of white noise processes take value zero for lags $J \neq 0$, so if the model is appropriate most of the coefficients of the sample ACF and PACF should be close to zero. In practice, we require that about the 95 % of these coefficients should fall within the non-significance bounds as shown in figure B.5 and figure B.6 respectively. For an adequate model, the histogram of residuals should show that the residuals are normally distributed. This signifies residuals to be white noise (figure B.7). Also one of the important tests is the cumulative distribution for the residual data. For an adequate model, the graph of the cumulative distribution for the residual data normally appears as a straight line when plotted on normal probability paper as shown in figure B.8 (Chow et al., 1988; Durbin, 1960).
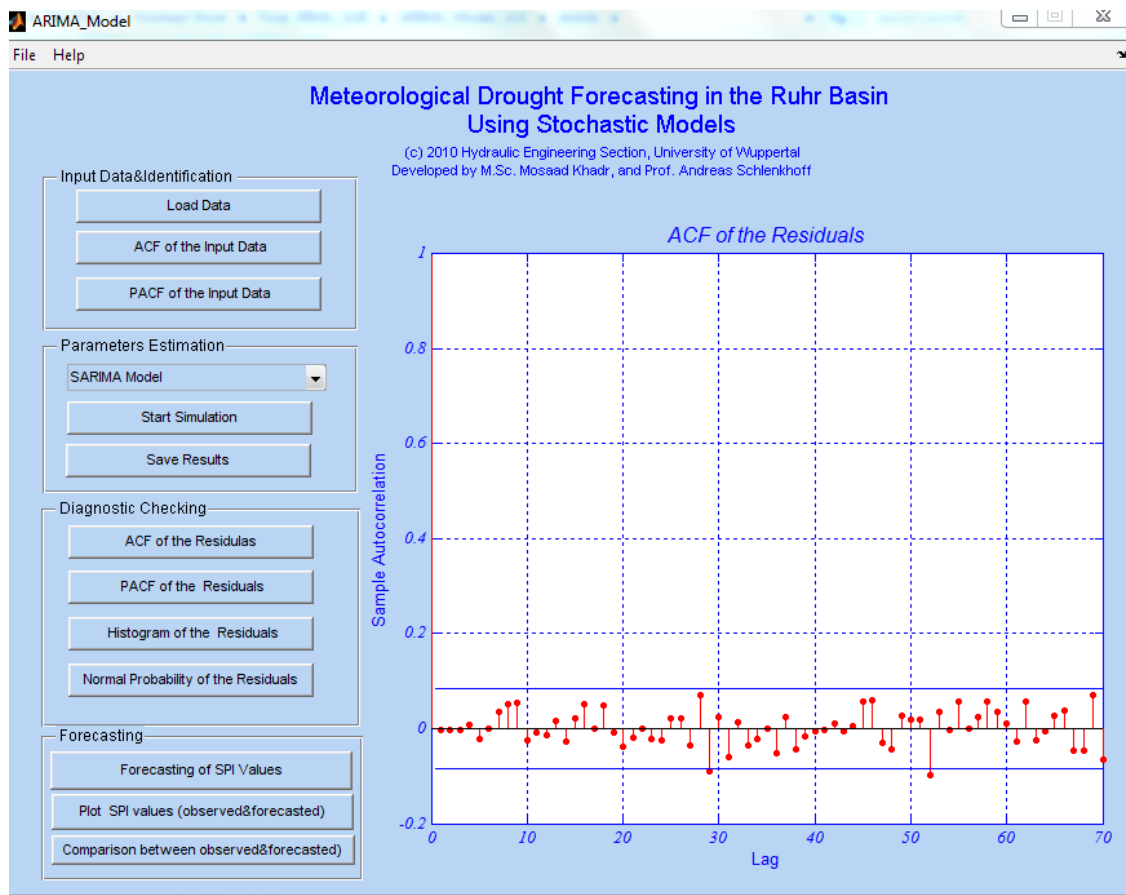
Figure B.5: ACF plot used for Diagnostic Check of the selected model
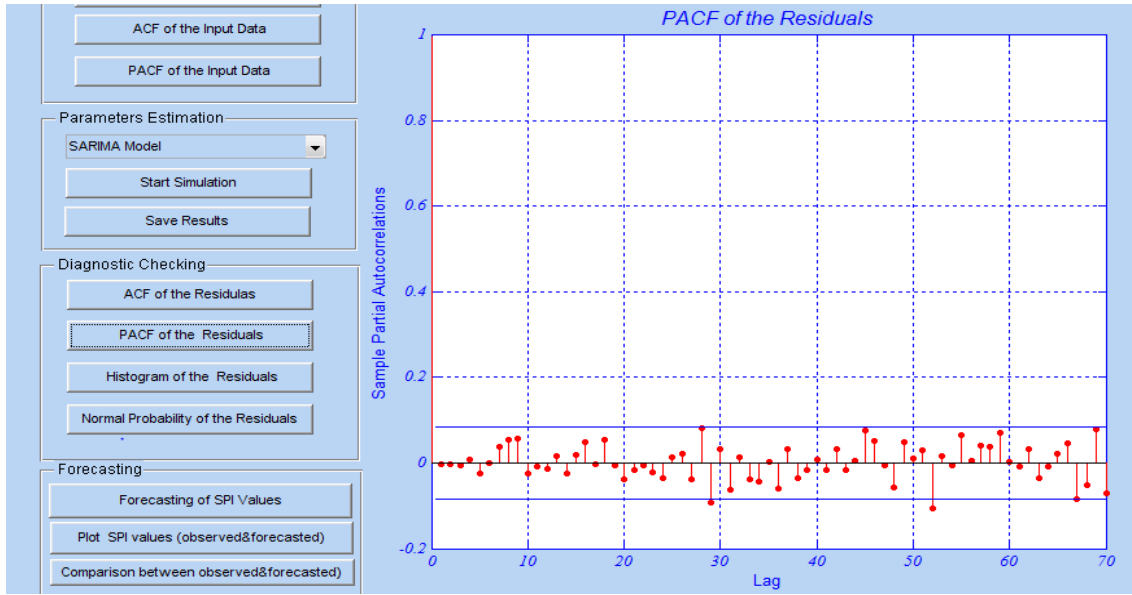
Figure B.6: PACF plot used for Diagnostic Check of the selected model
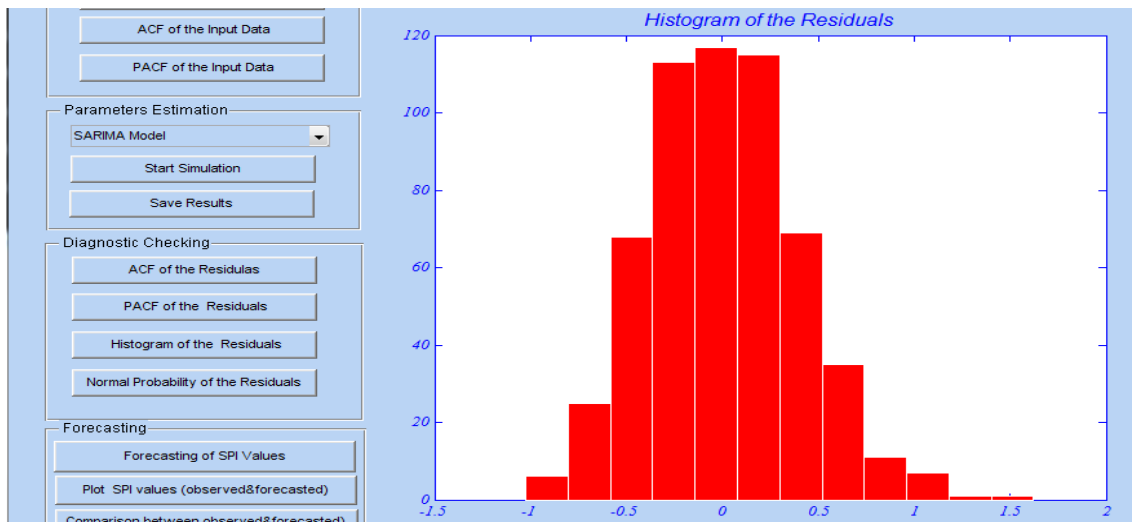
Figure B.7: Histogram of the residuals of the selected model
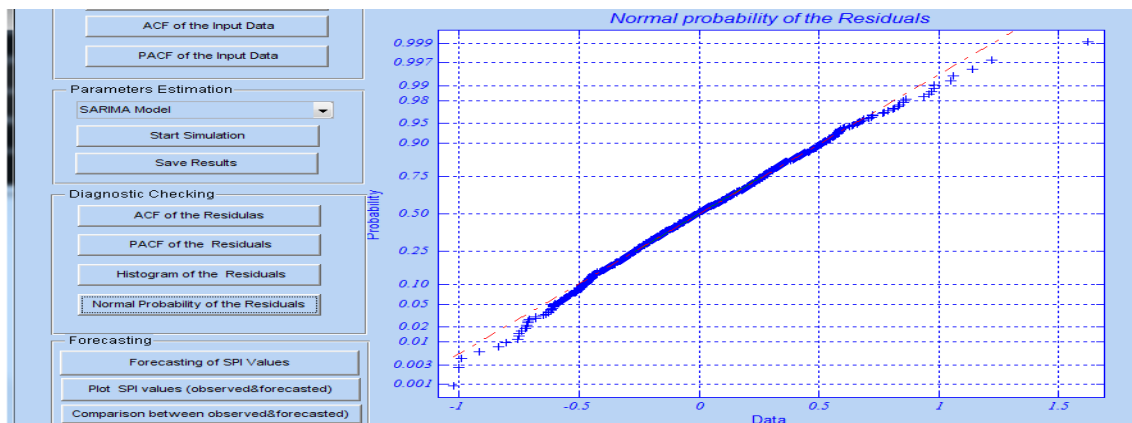

Figure B.8: Normal probability plot of the residuals of the selected model

B.9 Forecasting of the SPI Index from Selected Models

After finishing the previous steps, the user can easily forecast SPI values by clicking the button "Forecasting of SPI values". Then the user has to put the number of the required forecasted values. The forecasted values could be plot by clicking the button "Plot SPI values"(figure B.9). Also the user can compare between the observed and the forecasted values of selected model by clicking the button "comparison between observed & forecasted" (figure B.10).
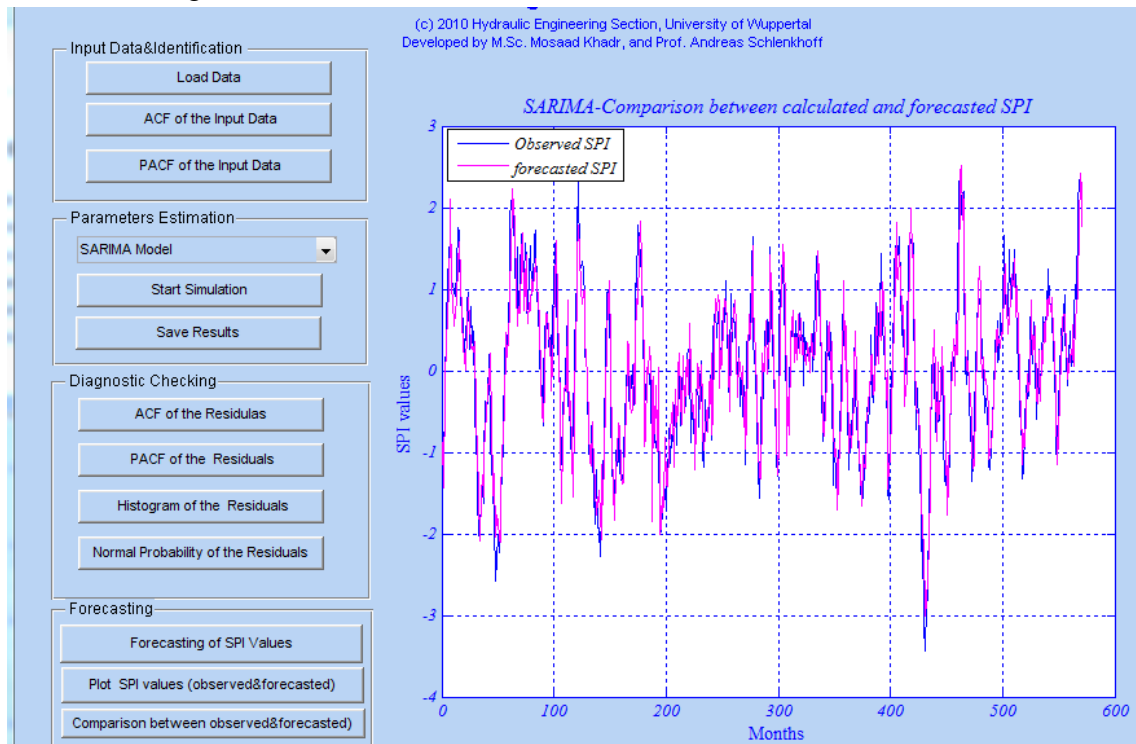
Figure B.9: Comparison of calculated SPI with forecasted SPI
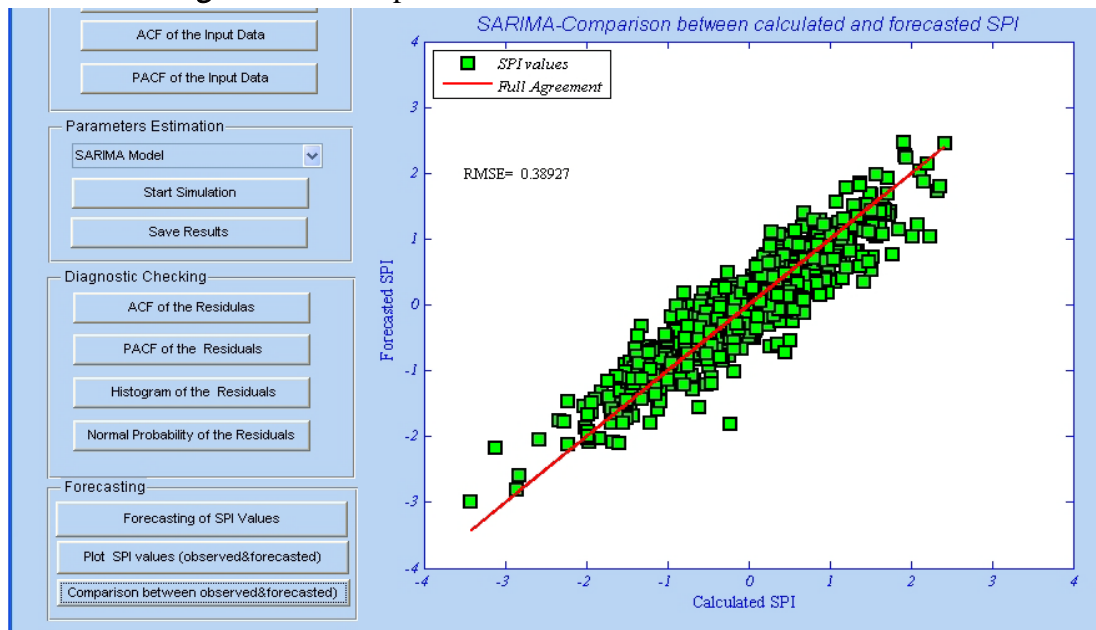
Figure B.10: Comparison of calculated SPI with forecasted SPI

B.10 Calibration and Accuracy of the Developed program (Drought_Forecasting)

B.10.1 Models Verification &Validation

Drought_Forecasting program contains two models, namely ARIMA model and SARIMA model. Model verification and validation (V&V) are essential parts of the model development process if models to be accepted and used to support decision making (Macal, 2005). It is the testing of the calibrated model against the additional set of field data preferably under different to further examine the range of validity of the calibrated model and this was done for the two models. Verification is done to ensure that (Macal, 2005); the model is programmed correctly; the algorithms have been implemented properly; the model does not contain errors, oversights, or bugs; the specification is complete and mistakes have not been made in implementing the model

But now becomes the critical question; is the program itself has a degree of accuracy or not. And this will be discussed in details in the following section.

B.10.2 Accuracy of the Developed program

Study of the accuracy of developed program (Drought_Forecasting) is very important to develop meaningful judgment. The extent of decision maker being wrong or right with regards to the obtained results of the program is greatly influenced by the accuracy of the program. In this section the relative accuracy has been examined by comparing the results of the program by the results of the well known computer program SPSS. The two models (ARIMA & SARIMA) have been investigated.

B.10.2.1 Comparison between the results of the ARIMA Model obtained by Drought_Forecasting& SPSS.

The ARIMA model has been fitted to the SPI_3 data series by using the developed program Drought_Forecasting and the software SPSS. In order to find the best model by using the software SPSS several trials must be done (See chapter 7), but by using the Drought_Forecasting the user can get directly the best model.
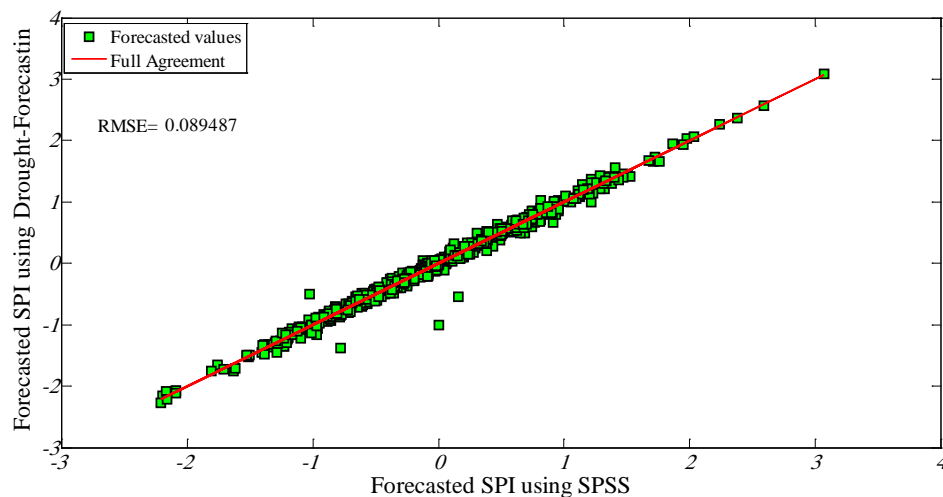
After several trails using SPSS program, the ARIMA (3, 0, 2) model has been found to the best one from the examined models. By using the Drought_Forecasting considering the same data series, the ARIMA (1, 0, 6) model has been found to be the best model. Table B.1 presents the estimated statistical parameter of the two models, table B.2 presents a comparison between the Statistical properties of the results obtained by the two models.

Table B.1: Comparison between statistical parameters

Model	AIC	Auto regressive parameters	Moving average parameters	Residual Sum of Squares	Residual Variance
-**SPSS**- ARIMA (3,0,2)	1055.48	0.07437 - 0.0258 0.1217	- 0.8832 - 0.8548	206.98	0.3627
Drought_Forecasting ARIMA (1, 0, 6)	1047.143	0.90	-0.02 -0.06 0.65 0 0.03 0.06	202.628	0.3536

Table B.2: Comparison between Statistical properties

| Model	Mean of the Calculated SPI	Mean of the forecasted SPI	Standard deviation of the Calculated SPI	Standard deviation of the forecasted SPI
-**SPSS**- ARIMA (3,0,2)	-2.6132e-005	-3.3537e-005	1.0009	0.7996
Drought_Forecasting ARIMA (1,0,6)	-2.6132e-005	-0.0030	1.0009	0.7998

Figure B.11: Comparison of forecasted SPI_3 values using the SPSS program and forecasted SPI_3 values using the developed program "Drought_Forecasting"

B.10.2.2 Comparison between the results of the SARIMA Model obtained by Drought_Forecasting& SPSS.

The ARIMA model has been fitted to the SPI_6 data series by using the developed program Drought_Forecasting and the software SPSS. After several trails using SPSS program (See chapter 5), the SARIMA $(1,0,3)(1,0,3)_6$ model has been found to the best one from the examined models. By using the Drought_Forecasting considering the same data series, the SARIMA $(1,0,5)(1,0,1)_6$ model was found to be the best model. Table B.3 presents the estimated statistical parameter of the two models and table B.4 presents a comparison between the statistical properties of the results obtained by the two models.

Table B.3: Statistical parameters of ARIMA model

Model	AIC	Auto regressive parameters	Moving average parameters	Seasonal Auto regressive parameters	Seasonal Moving average parameters	Residual Sum of Squares	Residual Variance
-**SPSS**- SARIMA $(1,0,3)(1,0,3)_6$	581.33	0.9533	- 0.0495 - 0.0227 - 0.1879	-0.9376	- 0.231 0.7747 0.0114	89.66	0.155
Drought_Forecasting SARIMA $(1,0,5)(1,0,1)_6$	557.92	1.124	o.1511 0.1531 -0.0199 0.2218 0.1403	-0.1124	0.804	86.067	0.1517

Table B.4: Statistical properties of ARIMA (3, 0, 2), and ARIMA (1, 0, 6) Results

Model	Mean of the Calculated SPI	Mean of the forecasted SPI	Standard deviation of the Calculated SPI	Standard deviation of the forecasted SPI
-**SPSS**- SARIMA $(1,0,3)(1,0,3)_6$	1.9264e-004	-8.7054e-004	1.0010	0.9151
Drought_Forecasting SARIMA $(1,0,5)(1,0,1)_6$	1.9264e-004	-0.0195	1.0010	0.9164
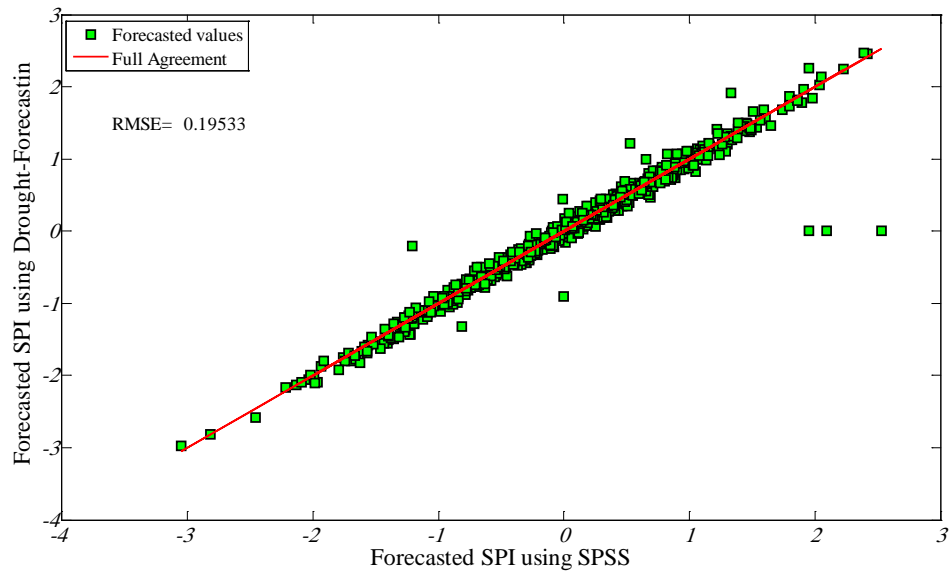
Figure B.12: Comparison of forecasted SPI_6 values using the SPSS program and forecasted SPI_6 values using the developed program "Drought_Forecasting"

Appendix C: Results of SPI Forecasting (SPI_12 and SPI_24)
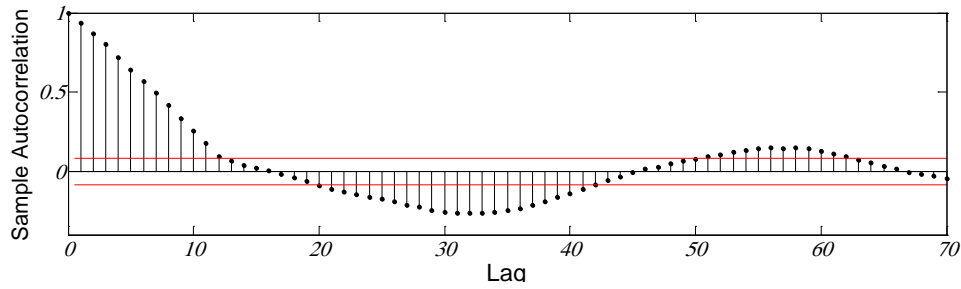
C.1 SPI_12

Figure C.1.1: ACF plot used for the selection of candidate models for SPI_12 series
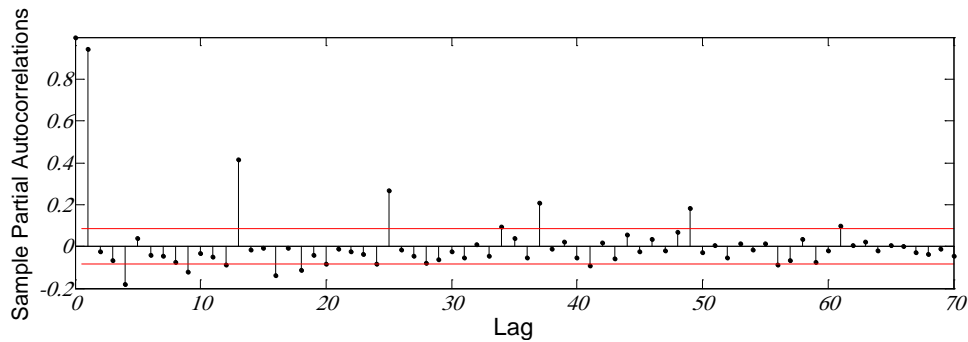

Figure C.1.2: ACF plot used for the selection of candidate models for SPI_12 series
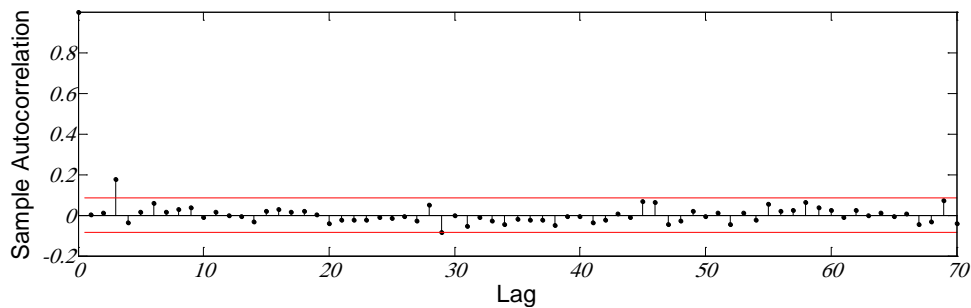

Figure C.1.3: ACF plot used for Diagnostic Check of the model SARIMA $(1, 0, 3)(1,0,3)_{12}$
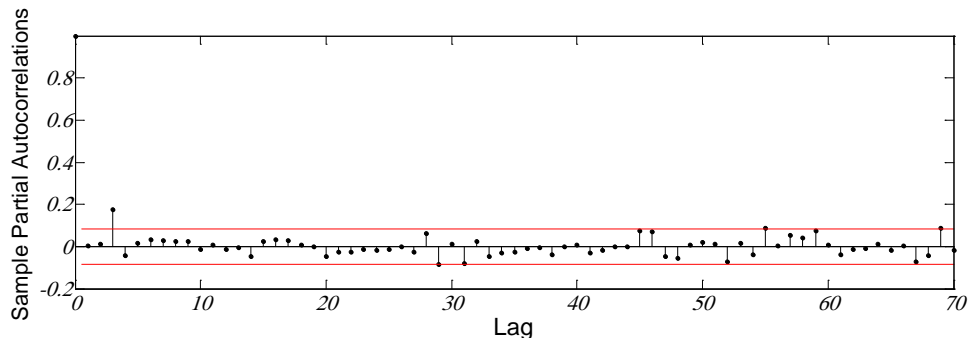
Figure C.1.4: PACF plot used for Diagnostic Check of the model SARIMA $(1, 0, 3)(1,0,3)_{12}$
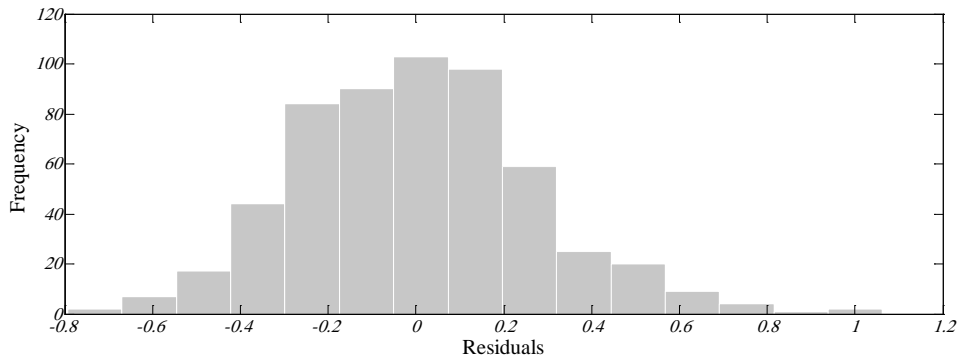
Figure C.1.5: Histogram of the residuals – SARIMA $(1, 0, 3)(1,0,3)_{12}$
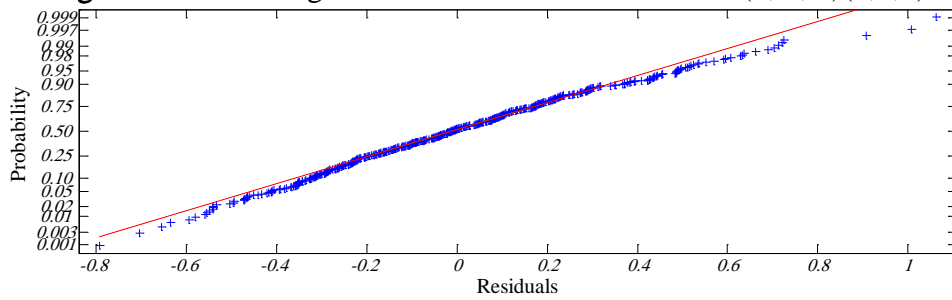
Figure C.1.6: Normal probability plot of the residuals– SARIMA $(1, 0, 3)(1,0,3)_{12}$

Table C.1.1: Statistical parameters of the model SARIMA $(1, 0, 3)(1,0,3)_{12}$

Model	Mean of the Calculated SPI	Mean of the forecasted SPI	Standard deviation of the calculated SPI	Standard deviation of the forecasted SPI	RMSE
SARIMA $(1, 0, 3)(1,0,3)12$	-1.7699e-005	-2.9264e-004	1.0010	0.9418	0.273
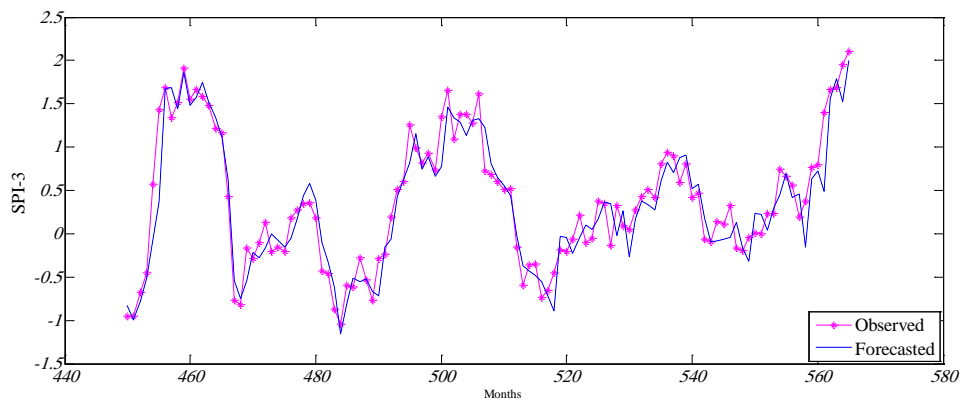
Figure C.1.7: Comparison of calculated SPI with forecasted SPI
SARIMA $(1, 0, 3)(1,0,3)_{12}$
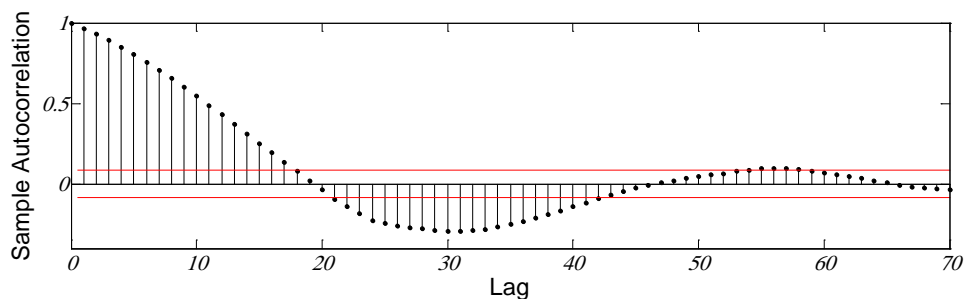
C.2 SPI_24

Figure C.2.1: ACF plot used for the selection of candidate models for SPI_12 series
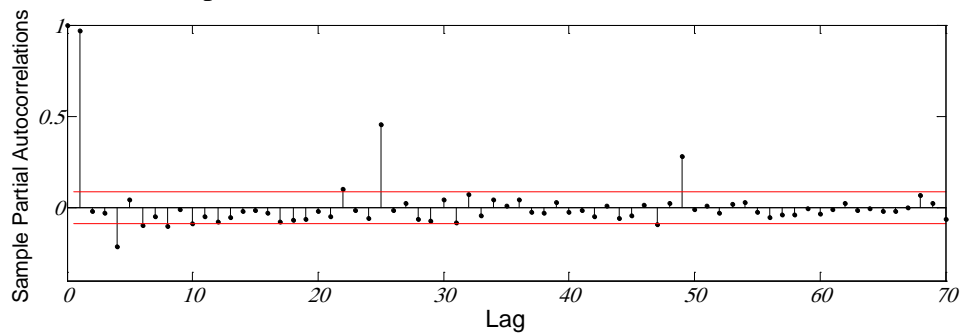

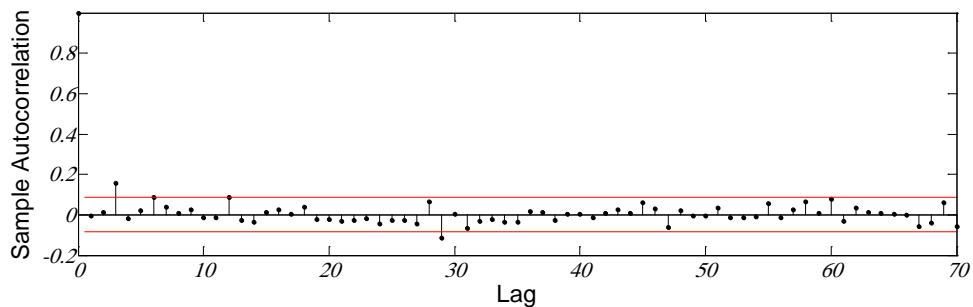**Figure C.2.2:** ACF plot used for the selection of candidate models for SPI_12 series

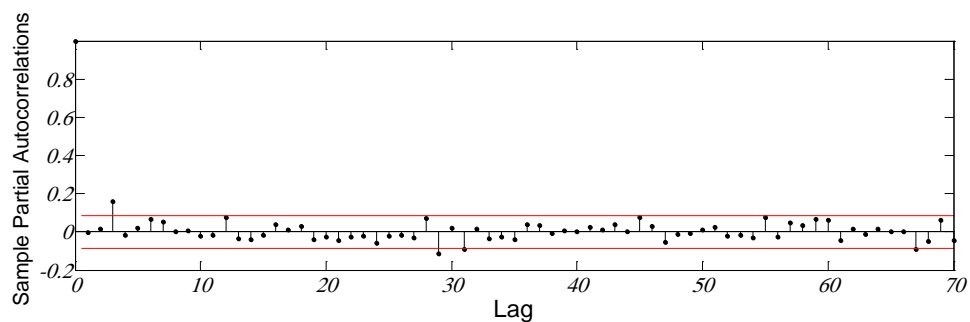**Figure C.2.3:** ACF plot used for Diagnostic Check of the model
SARIMA $(1, 0, 0)(6,0,0)_{24}$

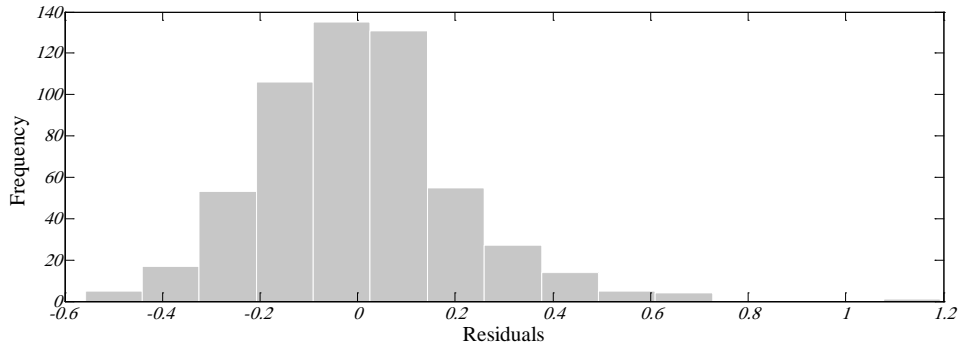**Figure C.2.4:** PACF plot used for Diagnostic Check of the model
SARIMA $(1, 0, 0)(6,0,0)_{24}$

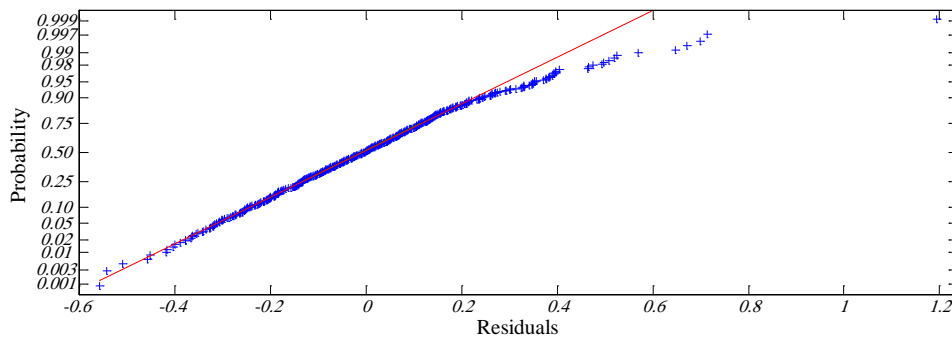**Figure C.2.5:** Histogram of the residuals – SARIMA $(1, 0, 0)(6,0,0)_{24}$

Figure C.2.6: Normal probability plot of the residuals– SARIMA $(1, 0, 0)(6,0,0)_{24}$

Table C.2.1: Statistical parameters of the model SARIMA $(1, 0, 3)(1,0,3)_{24}$

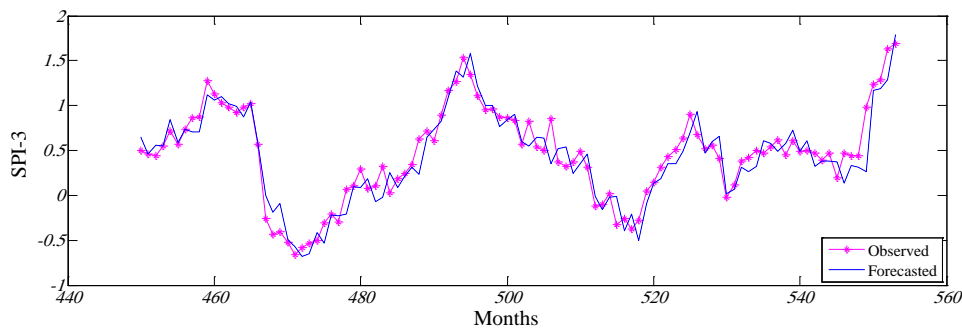| Model	Mean of the Calculated SPI	Mean of the forecasted SPI	Standard deviation of the calculated SPI	Standard deviation of the forecasted SPI	RMSE
SARIMA (1, 0, 3)(1,0,3)24	3.6166e-005	6.0101e-004	1.0009	0.9684	0.2029

Figure C.2.7: Comparison of calculated SPI with forecasted SPI
SARIMA $(1, 0, 0)(6,0,0)_{24}$

Appendix D: Sample of Input Data of Scenario Number 1 in the Optimization Model

Month	Inflow M.m^3	R_{T_max1} M.m^3	R_{T_max2} M.m^3	R_{T_max3} M.m^3	R_{T_max4} M.m^3	S M.m^3	S_{min} M.m^3	S_{max} M.m^3	Demand M.m^3	E_{max} MkWh
1	41.58	25.92	25.92	25.92	3.89	137.00	80.00	138.00	15.00	11.39
2	28.56	25.92	25.92	25.92	3.89		80.00	150.00	15.00	11.39
3	32.07	25.92	25.92	25.92	3.89		80.00	160.00	15.00	11.39
4	15.29	25.92	25.92	25.92	3.89		80.00	170.00	15.00	11.39
5	9.69	25.92	25.92	25.92	3.89		80.00	170.00	15.00	11.39
6	6.54	25.92	25.92	25.92	3.89		80.00	170.00	15.00	11.39
7	7.68	25.92	25.92	25.92	3.89		80.00	170.00	15.00	11.39
8	8.70	25.92	25.92	25.92	3.89		80.00	170.00	15.00	11.39
9	11.04	25.92	25.92	25.92	3.89		80.00	160.00	15.00	11.39
10	17.15	25.92	25.92	25.92	3.89		80.00	150.00	15.00	11.39
11	27.71	25.92	25.92	25.92	3.89		80.00	138.00	15.00	11.39
12	34.50	25.92	25.92	25.92	3.89		80.00	138.00	15.00	11.39
1						137.00				