

https://dspace.org/

Rainfall-runoff modeling for real-time ecological reserve implementation.

Halwiindi, Mazunda

2010-02-03T09:01:07Z
http://hdl.handle.net/10539/7496

Downloaded from a DSpace repository. https://dspace.org/

RAINFALL-RUNOFF MODELING FOR REAL-TIME ECOLOGICAL RESERVE IMPLEMENTATION

Mazunda Halwiindi

A Research Report submitted to the Faculty of Engineering and the Built Environment of the University of the Witwatersrand in partial fulfillment of the requirement for the award of Master of Science (MSc) in Engineering.

September, 2009

DECLARATION

I declare that this research report is my own, unaided work. It is submitted for the Degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

_____

(Signature of Candidate)

ABSTRACT

Real-time ecological reserve implementation systems are essential for preventing deterioration, or facilitating the restoration of riverine ecosystems. These conditions are a result of modified flow regimes that alter essential facets of flow, and which are in turn due to water resource development projects and other activities undertaken along river courses. The essential facets of flow for ecological reserve implementation are presently seen as the low or base flows, the small increases in flow, referred to as freshes, and the small to medium floods. Large floods, which may not be impounded, are not considered because they are unmanageable. The objective therefore is to bring back certain facets of the original natural flow regime that are essential for the proper functioning of the riverine ecosystem. These facets rely strongly on the natural variability of river flow. To accomplish this objective, there is need of effecting optimized reservoir releases that mimic natural variability of flow, in this way to satisfy both ecosystem needs as well as human needs for water. Therefore defining an approach that will provide trigger information to release the required flow becomes imperative.

The scope of this project is confined to the study of high flows or flood events in a sub-catchment of the Thukela River in an attempt to build a model that can be used to forecast an impending flood event and hence trigger appropriate releases. The models are for making a 1 day forecast of flow from rainfall at two

rain gauge stations i.e. Heartsease and Monks Cowl stations. The study employs 3 methods: regression analysis, a simple empirical model and Artificial Neural Networks (ANN) to build the models. The flood events were selected from stream flow records of a stream gauge at Instream Flow Requirement (IFR) station V1H010 in the little Thukela River.

The results obtained from the study show that the ANN performed better than the other two models which yielded unsatisfactory results for prediction of flood events. ANN analysis produced a coefficient of determination of 0.60 with a correlation coefficient of 0.78. Results obtained from regression analysis were 0.49 as coefficient of determination and 0.70 for correlation coefficient. Analysis results from the empirical model showed the worst performance of the three models with a coefficient of determination of 0.36 and a correlation coefficient of 0.60. The results bring forth the need to further analyze data using more powerful models in order to achieve better results than those obtained from ANN. The analysis also indicates that the recommended ecological reserve implementation from a reserve determination study of the selected catchment cannot even be met by the natural series and are therefore most likely invalid.

ACKNOWLEDGEMENTS

A number of people have contributed to make this project possible. First and foremost I would like to thank Professor Chris James of the University of the Witwatersrand for the provision of finance and other facilities, through the Water Research Commission, that enabled me to carry out this study. I would also like to express my thanks to Professor Dennis Hughes from Rhodes University for his input and the raw data. Last but not least my sincere gratitude goes to my supervisor Dr John Ndiritu for all the guidance, effort and perseverance that finally made this report possible.

TABLE OF CONTENTS

DECLARATION i
ABSTRACT ii
TABLE OF CONTENTS v
LIST OF FIGURES viii
LIST OF TABLES ix
CHAPTER 1 1
1. INTRODUCTION 1
1.1 Problem Statement 5
1.2 Objective 5
CHAPTER 2 6
2. LITERATURE REVIEW 6
2.1 Rainfall- runoff Modeling 6
2.2 Environmental Flow Methodologies 9
2.2.1 Hydrological Methodologies 9
2.2.2 Hydraulic Rating Methods 10
2.2.3 Habitat Simulation or Microhabitat Modeling Methodologies 10
2.2.4 Holistic Methodologies 11
2.3 Choice of Hydrological Method 18
CHAPTER 3 19
3. STUDY AREA 19
3.1 General Description of the Area 19
3.2 Rainfall 20
3.3 Water Resource Developments and Tributaries 21
3.4 IFR 3– Little Thukela 22
CHAPTER 4 25
4. METHODS AND DATA ACQUISITION 25
4.1 Relevant Data Sets 26

4.2 Stream-Flow Data 27
4.3 Rainfall Data 29
4.4 Required Conditions for Satisfaction of the Reserve 29
CHAPTER 5 33
5. REGRESSION ANALYSIS 33
5.1 Performance Criteria 33
5.2 Model Development 38
5.2.1 Simple Regression Analysis 38
5.2.2 Multiple Regression Analysis 39
5.3 Results Analysis and Discussion 40
CHAPTER 6 45
6. EMPIRICAL MODEL 45
6.1 Development of the Model 46
6.2 SCE-UA Algorithm 49
6.3 Analysis of Results and Discussion 52
CHAPTER 7 55
7. ARTIFICIAL NEURAL NETWORKS 55
7.1 Introduction 55
7.2 Network Topology 56
7.3 Architecture of ANN 57
7.3.1 Weighting Factors 59
7.3.2 Summation Function 59
7.3.3 Transfer Function 60
7.3.4 Scaling and Limiting 60
7.3.5 Output Function (Competition) 61
7.3.6 Error Function and Back-Propagated Value 61
7.3.7 Learning Function 62
7.4 Training an Artificial Neural Network 62
7.4.1 Learning Parameters 64
7.5 Development of the Artificial Neural Network (ANN) Model 66
7.6 Results Analysis and Discussion 68

CHAPTER 8 74
8. COMPARISON OF MODELS 74
CHAPTER 9 76
9. CONCLUSIONS AND RECOMMENDATIONS 76
9.1 Conclusions 76
9.2 Recommendations 77
10. REFERENCES 79

LIST OF FIGURES

Figure 3-1: Overview of Study Area 20
Figure 3-2: Study Area Showing IFR Sites 24
Figure 4-1: Location of IFR 3 28
Figure 4-2: Time-series of Rainfall and Flow 29
Figure 5-1: Comparison of Observed and Simulated Flows Data from Multiple Regression 44
Figure 6-1: Observed and calibrated stream-flow time series of empirical model 54
Figure 7-1: A Simple Neural Network Diagram. 57
Figure 7-2: A three-layer feed forward ANN 58
Figure 7-3: Simple Error Back-propagation ANN 62
Figure 7-4: Comparison of Observed and Simulated Flows for ANN Training (10/8/1965 to 2/2/1967) 70
Figure 7-5: Scatter Plot of Observed Versus Simulated Flow for ANN Training .. 71
Figure 7-6: Streamflow Error Plot for ANN Training 71
Figure 7-7: Comparison of Observed and Simulated Flows of the for ANN Testing (10/8/1965 to 2/2/1967) 72
Figure 7-8: Scatter Plot of Observed Versus Simulated Flow for ANN Testing ... 72
Figure 7-9: Streamflow Error Plot for ANN Testing 73

LIST OF TABLES

Table 4-1: Rainfall and Stream-Flow Measuring Stations 27
Table 4-2: Required Conditions to Satisfy the IFR 30
Table 4-3: Number of Months in which IFR Events are Satisfied 31
Table 5-1: Single Regression (one antecedent day) Analysis Results for the Monks Cowl Data 41
Table 5-2: Single Regression (one antecedent day) Analysis Results for the Heartsease Data 42
Table 5-3: Multi Regression (one antecedent day for each) Analysis Results for the Heartsease and Monks Cowl Data 42
Table 5-4: Multi-regression analysis results with seven antecedent days' rainfall for both Monks Cowl and Heartsease 43
Table 6-1: Number of Datasets Used for calibration 48
Table 6-2: Algorithmic parameters for the SCE algorithm 51
Table 6-3: Parameters of calibrated empirical model. 52
Table 6-4: Empirical model Calibration 53
Table 7-1: Performance Statistics of an ANN Model Developed Using Complete Dataset 68
Table 7-2: Qualitative Guidelines for Assessing the Adequacy of Stream-flow Estimates 69

CHAPTER 1

1. INTRODUCTION

Dams and reservoirs have, through the ages, provided tremendous social and economic benefits to mankind. Most notable of the benefits are making available a source of water for domestic supply, irrigation, industrial use and power generation. In the endeavor to manage water to meet the various human needs, however, the water needs of freshwater species and ecosystems have been largely neglected. The ecological consequences have been catastrophic (IUCN 2000, Pringle et al., 2000, Stein et al., 2000, Baron et al., 2002). River flow regime alteration which is a result of reservoir operation has been acknowledged as one of three leading causes, along with non-point source pollution and invasive species, of the endangerment of aquatic animals (Richter et al., 1997a, Pringle et al., 2000). Freshwater ecosystem services and products which add remarkable value to society have been severely compromised as well (Postel and Carpenter 1997, IUCN 2000).

Natural freshwater ecosystems are to a large extent influenced by specific facets of natural hydrologic variability. Of particular importance are seasonal high and low flows, and occasional floods and droughts (Stanford et al., 1996, Poff et al., 1997, Richter et al., 1997b). A river's flow regime is now recognized as a ''master

variable'' that drives variation in many other components of a river ecosystem, e.g., fish populations, floodplain forest composition, nutrient cycling, in both direct and indirect ways (Sparks 1995, Walker et al., 1995, Poff et al., 1997). The extraordinary species richness and productivity characteristic of freshwater ecosystems is strongly dependent upon and attributable to the natural variability of their hydrologic conditions. But variability runs counter to the dominant goals of water resource management (Holling and Meffe 1996). Traditional water management has generally sought to dampen the natural variability of river flows to attain steady and dependable water supplies for domestic and industrial uses, irrigation, navigation, and hydropower, and to moderate extreme water conditions such as floods and droughts. For instance, by storing water in reservoirs, water managers capture high flows during wet years or seasons to supplement water supplies at drier times, thereby maximizing the reliability of water supplies and certain economic benefits each year.

When natural variability in river flows is altered too much, marked changes in the physical, chemical, and biological conditions and functions of natural freshwater ecosystems can be expected. When changes to natural flow regimes are excessive, causing a river ecosystem to degrade toward an altered character, the costs are high to both biodiversity and society (Postel and Carpenter 1997, IUCN 2000, WCD 2000). Costs in terms of soil erosion, land loss through bank collapse and consequent reduction in the life-span of in-channel dams, loss of valued species, blooms of pest species, loss of fisheries, increasing levels of water

pollution and linked health problems, loss of river features and habitats, loss of recreational and spiritual values of water systems, loss of river resources for riparian peoples reliant on them for subsistence, etc have profound economic and social implications.

“In recent years the value of dams to human society has been questioned. Over the last two decades, the multiple values of natural ecosystems to human society and the environmental consequences of dams have become more widely understood” (Acreman et. al., 2000). Significant insight has been achieved by river scientists about ecosystem maintenance requirements through a study process known as in-stream flow assessments (IFAs). In-stream flow assessments (IFAs) address how much and which specific temporal characteristics of the original flow regime of a river should continue to flow downstream in order to maintain specified features of the “riverine ecosystem” (Arthington et al., 1992). “An environmental flow requirements study (EFR) produces descriptions of possible modified hydrological regimes for the river, the EFA or environmental water allocation(s) of each modified regime linked to a predetermined objective in terms of the ecosystem’s future condition” (King & Tharme 1994).

Estimating ecosystem flow requirements requires input from an interdisciplinary group of scientists familiar with the habitat requirements of native biota (i.e., species, communities) and the hydrologic, geomorphic, and biogeochemical

processes that influence those habitats and support primary productivity and nutrient cycling (Swales and Harris 1995, King and Louw 1998). In South Africa, expert assessment workshops have been adopted for the purpose of defining necessary flows to support desired future conditions of riverine ecosystems (King et al., 2000). During these workshops, interdisciplinary participants draw upon existing data, research results, ecological and hydrological models, and professional judgment in developing initial targets for ecosystem flow requirements (King and Louw 1998). At the global level, a wide variety of tools and methods is being used to prescribe ecosystem flow requirements, the approaches of which are evolving rapidly (Tharme 1996, Arthington and Zalucki 1998, Bragg and Black 1999, Railsback 2001, Tharme 2003).

Scales at which IFAs are undertaken vary widely, from entire large river basins that include a regulated main channel and/or several regulated tributaries, to a flow restoration project for a single flow-impacted river reach or even for a single fish species. “Different methodologies are appropriate at each particular spatial scale as well as in relation to typical project constraints including the time frame for assessment, availability of data, technical capacity and finances” (Tharme 1996; Arthington and Zalucki 1998, Arthington and Pusey 2003). Methodologies for that reason range from rapid, reconnaissance-level approaches for regional, national or basin-wide water resources planning, to resource intensive methodologies for highly exploited, individual river sites subject to multiple uses, or single species of high conservation significance.

1.1 Problem Statement

The most difficult issue related with the releasing of managed floods is when to initiate a release and how to establish the downstream objective at any specific moment in time, which determines how much to release.

Hughes et al. (1997) discussed the issue of triggering high flow events in the context of the Daily IFR model. The original Daily IFR model used a system of looking forward 10 days in the input time series to identify a suitable trigger for a high flow release. This is clearly of little value in implementation unless some form of forecasting model with a lead time of not less than 1 day is also made available to the reservoir operational staff.

The major issue that still remains unresolved and for which no suitable methods have been identified is the approach for triggering the high flow releases in order to achieve the Reserve requirement with a minimum volume of release (Hughes 2006).

1.2 Objective

The objective of the study is to evaluate the rainfall method suggested by Hughes (2006) as a possible approach to provide trigger information to effect optimal reservoir flood releases.

CHAPTER 2

2. LITERATURE REVIEW

2.1 Rainfall- runoff Modeling

Rainfall-Runoff (R-R) models and generally hydrologic models are simplified, conceptual representations of part of the hydrologic cycle. Models are primarily used for research, hydrologic design and prediction. Their application is very much dependent on the purposes for which the modeling is done. Simulation and prediction using hydrologic models aids decision makers in planning and operation of hydrological systems such as in real-time flood forecasting and warning, estimating flood frequencies, flood routing and inundation prediction, impact assessment of climate and land use change and integrated watershed management.

Abbott and Refsgaard (1996) classified hydrologic models into three types based on the runoff generation process. These classes are deterministic, stochastic and joint stochastic- deterministic models. Deterministic models are further classified as empirical (black box), lumped conceptual (grey box) and distributed physically based (white box) models.

Empirical models are developed using measured time series instead of mathematical expressions that describe the physical processes in a catchment.

Some empirical models are statistically based using statistical methods such as Autoregressive Integrated Moving Average (ARIMA). Others are based on the unit hydrograph model, which is a linear regression on excess precipitation. The third group of empirical models is data-driven using methods such as evolutionary algorithm, artificial neural networks, nearest neighbour method, model trees, support vector machines, etc.

Lumped conceptual models employ equations that are semi empirical, but with a physical basis. The parameters and variables represent average values over the entire catchment, disregarding spatial variability. These models cannot usually be assessed from field data alone, but through calibration. This approach relates the forcing data, mainly precipitation inputs, to system outputs (streamflow) without considering the spatial processes, patterns and organization of the characteristics governing the processes. These limitations cause difficulties (Beven, 2001), and therefore a number of complex lumped R-R models have been developed (Fleming 1975; Singh 1995; Singh and Frevert 2002a, b). One disadvantage lumped models present is that they are generally designed to simulate the streamflow just at the watershed outlet and thus are not suitable for estimating flow at some interior locations in a river Basin (Refsgaard 1996).

Physically-based distributed models processes are represented by one or more partial differential equations with equations and parameters that are distributed in space. Beven (1985) and Smith et al. (2004) outlined benefits of distributed

models as the possibility of considering spatially variable inputs and outputs, assessment of pollutants and sediment transport, and also analyzing the hydrological response at ungauged basins. However, the parameterization, calibration, and error correction of these complex models still present many outstanding questions. Several studies to justify the use of distributed models versus lumped ones have been conducted (Loague and Freeze, 1985; Bell and Moore, 1998; Michaud and Sorooshian, 1994). The estimation of the excessive parameters within distributed models is the main source of uncertainty in these models.

Current trends in flood forecasting are moving away from the conventional simple deterministic forecasts which use hydrographs toward probabilistic forecasts, which include prediction uncertainty. Probabilistic forecasts specify a certain probability distribution function of the predicted value. The predictive probability in a probabilistic forecast is a numerical measure of the confidence of the intensity of a flood event, based on all meteorological or hydrological information utilized in the forecasting process (Krzysztofowicz, 2001). Deterministic models are based on cause-effect relationships and usually are described by mathematical equations. Stochastic models are based on the premise that relationships often cannot be expressed in simple or complex cause-effect mathematical forms. Instead, the "effect" variables are observed and their properties investigated by using methods of stochastic processes and mathematical statistics (Yevjevich 1974).

2.2 Environmental Flow Methodologies

Tharme (2003) has documented four relatively distinct types of environmental flow methodologies. These are (1) hydrological, (2) hydraulic rating, (3) habitat simulation, and (4) holistic methodologies.

2.2.1 Hydrological Methodologies

These represent the simplest set of techniques where, at a desktop level, hydrological data in the form of naturalized, historical monthly or average daily flow records are analyzed to derive standard flow indices which then become the recommended environmental flows. Commonly, the EFR is represented as a proportion of flow, often termed the 'minimum flow'.

This minimum flow is intended to maintain river health, fisheries or other highlighted ecological features at some acceptable level, usually on an annual, seasonal or monthly basis. In a few instances, secondary criteria in the form of catchment variables, hydraulic, biological or geomorphological criteria are also incorporated.

Hydrological methodologies are generally used mainly at the planning stage of water resource developments, or in situations where preliminary flow targets and exploratory trade-offs are required (Arthington et al., 1998; Tharme 2003).

2.2.2 Hydraulic Rating Methods

Hydraulic methods use changes in simple hydraulic variables, such as wetted perimeter or maximum depth, usually measured across single, flow-limited river cross-sections (e.g. riffles) as a surrogate for habitat factors known or assumed to be limiting to target biota. Environmental flows are determined from a plot of the hydraulic variable against discharge, commonly by identifying curve breakpoints where significant percentage reductions in habitat occur with decreases in discharge (Jowett 1997). It is assumed that ensuring some threshold value of the selected hydraulic parameter at a particular level of altered flow will maintain aquatic biota and thus, ecosystem integrity. These relatively low-resolution hydraulic techniques have been superseded by more advanced habitat modeling tools, but they continue to be improved and are still being applied, often as part of holistic approaches.

2.2.3 Habitat Simulation or Microhabitat Modeling Methodologies

Habitat simulation methods also make use of hydraulic habitat-discharge relationships, but provide more detailed, modeled analyses of both the quantity and suitability of physical stream habitat available to target biota under different discharges. Environmental flow recommendations are based on the integration of hydrological, hydraulic and biological response data (Jowett 1997). Flow-related changes in physical microhabitat are modeled in various hydraulic programs, typically using data on depth, velocity, substratum composition and cover and, more recently, complex hydraulic indices (e.g. benthic shear stress). This

information is collected at multiple cross-sections within each study river reach. Simulated available habitat data are then linked with habitat suitability index curves representing the habitat conditions used (or preferred) by target fish or invertebrate species (or life-history stages). The resultant outputs, in the form of habitat-discharge curves for specific biota, or extended as habitat time and exceedence series, is used to derive optimum environmental flows for a species or life history stage. The habitat simulation modeling package PHABSIM (Bovee 1982; Stalnaker et al., 1994) is the pre-eminent model of this type widely used in the USA and many other countries. A modified version of this package known as RHYHABSIM is often used to develop minimum environmental flows for New Zealand Rivers (Jowett 1997).

2.2.4 Holistic Methodologies

Over the past decade, river ecologists have increasingly made the case for a broader approach to the definition of environmental flows to sustain and conserve the river ecosystem (Arthington and Pusey 1993; Sparks 1995; Poff et al., 1997). From the conceptual foundations of a proposed “holistic approach” (Arthington et al., 1992) at least 16 so called “holistic methodologies” (Tharme 1996, 2003) have been developed and applied in Australia, South Africa and the UK. This type of approach is founded on the assumption that if certain features of the natural hydrological regime can be identified and adequately incorporated into a modified flow regime, then, all other things being equal, the extant biota and functional integrity of the ecosystem should be maintained (King & Tharme

1994). Reasoning along similar lines, Sparks (1995) suggested that “rather than optimizing water regimes for one or a few species, a better approach is to try to approximate the natural flow regime that maintained the entire panoply of species.”

Holistic methodologies aim to address the water requirements of the entire riverine ecosystem rather than the needs of only a few taxa (usually fish or invertebrates). These methodologies are underpinned by the concept of the “natural flow paradigm” (Poff et al., 1997), and basic principles guiding river corridor restoration.

They share a common objective– to maintain or restore the biophysical components and ecological processes of in-stream and groundwater systems, floodplains and downstream receiving waters (e.g. terminal lakes and wetlands, estuaries and near-shore marine ecosystems).

Ecosystem components that are commonly considered in holistic assessments include geomorphology, hydraulic habitat, water quality, riparian and aquatic vegetation, macro-invertebrates, fish, and other vertebrates having some dependency upon the river and its riparian habitats (e.g. amphibians, reptiles, birds, mammals). The flow requirements of each of these components can be evaluated using data derived from field studies, modeling and desktop techniques (refer to Arthington & Zalucki 1998 for a review of such methods

particularly those used in Australia), and/or by accessing expert opinion. The requirements of each component for particular volumes and timing of flows are then incorporated into the EFA recommendations.

Holistic methodologies currently represent around 8% of the global total of some 200 environmental flow methods (Tharme 2003). Although predominantly developed and used in South Africa and Australia, recently holistic approaches have begun to attract growing international interest in both developed and developing regions of the world, with strong expressions of interest by some 12 countries in Europe, Latin America, Asia and Africa (Tharme 2003). At least 16 methodologies based on the holistic principles now exist, and they can be described (Arthington et al., 1998) as either 'bottom-up' methods (designed to 'construct' a modified flow regime by adding flow components to a baseline of zero flows), or 'top-down' methods addressing the question: "How much can we modify a river's flow regime before the aquatic ecosystem begins to noticeably change or becomes seriously degraded"?

The South African Building Block Methodology or BBM (King et al., 2002) was the first structured approach of this type. It began as a bottom-up method, more recently incorporating the Flow Stress-Response Method. In this modified form it is legally required for intermediate and comprehensive determinations of the South African Ecological Reserve (Tharme 2003). Other bottom-up

methodologies include several expert and scientific panel methods developed and applied in Australia (Arthington 1998 and Cottingham *et al*., 2002).

Environmental flow assessments may include evaluation of a range of other mitigation measures, for example, how to restore longitudinal and lateral connectivity by providing fish passes or altering the configuration of levee banks on a floodplain. Management of storage water levels may also be examined and recommendations made on the benefits of more, or less, stable water levels. Some of the holistic methodologies described above (e.g. the Flow Restoration Methodology) also take into consideration the influence of threatening processes and disturbances unrelated (or less directly related) to flow regulation, and advise on possible mitigation measures such as riparian and habitat restoration, or the management of invasive vegetation and fish.

Over 50 countries now use environmental flow assessment as a water management tool. Moreover, the requirement to provide environmental flows to protect and restore river ecosystems is increasingly appearing in national legislation in Australia as part of recent water reforms (Commonwealth of Australia 1996; Arthington & Pusey 2003), in South Africa associated with the new water laws (King *et al.,* 2003; Tharme 2003), and in Europe in response to the European Water Directive.

Ecologically sustainable water management protects the ecological integrity of affected ecosystems while meeting intergenerational human needs for water and sustaining the full array of other products and services provided by natural freshwater ecosystems. Ecological integrity is protected when the compositional and structural diversity and natural functioning of affected ecosystems is maintained (Richter et al., 2003a).

The natural flow-regime paradigm postulates that the structure and function of a riverine ecosystem, and the adaptations of its constituent riparian and aquatic species, are dictated by the pattern of temporal variation in river flows. In ecological terms, the primary components of a flow regime are the magnitude, frequency, seasonal timing, predictability, duration and rate of change of flow conditions. From an evolutionary perspective, extreme events (floods and droughts) exert primary selective pressure for adaptation because they often represent sources of mortality. In the context of adaptation to flow regimes, a lexicon of flow regime parameters would be the following:

a. Magnitude: the amount of water moving past a fixed location per unit time. The larger (or smaller) the magnitude of a flood (or drought), the greater the expected physical impact.
b. Frequency: the number of events of a given magnitude per time interval (e.g. per year). For a given river or stream, frequency is typically related inversely to magnitude.

c. Duration: the period of time associated with a particular flow event expressed in terms of number of days a flood or drought lasts.
d. Timing: the date during the year that flood or drought occurs, often derived from long-term flow records.
e. Predictability: the degree to which flood or drought events are auto-correlated temporally, typically on an annual cycle. Predictable events might also be correlated with other environmental signals (e.g. rainfall events, seasonal thermal extremes, sudden increases or decreases in flow).

Scudder (1980) was one of the first to suggest managed floods as a viable development strategy. Since then various experiments with managed floods have been made, primarily in Africa. The restoration of the floodplain ecosystem and traditional farming systems which can be achieved by managed floods releases has been demonstrated on the Waza-Logone floodplain in Cameroon through ecological and socio-economic surveys.

Following an economic valuation of the products and functions of the Hadejia-Nguru wetlands in Nigeria, the Hadejia-Jamare River Basin Development Authority has experimented with dam releases to augment the annual flood. In the Senegal basin, the value of flood recession farming to rural livelihoods has been acknowledged and a programme of managed flood releases has been implemented. The size and timing of managed flood releases from the

Pongolapoort dam on the Phongolo River in South Africa have been defined by a participatory process, involving local farmers, fishermen and other stakeholders.

In the past managed floods have rarely, been considered as an integral component at the design stage of dams. An exception is the Itezhi-tezhi reservoir on the Kafue River in which additional storage was set aside specifically for managed flood releases. However, releases have not been entirely successful because of institutional problems between conflicting users. The concept of managed floods is being considered for some new dams, such as that proposed for Grand Falls/Mutonga on the Tana River in Kenya.

Hydrologic simulation modeling has advanced rapidly and computerized models have become essential tools for understanding human influences on river flows and designing ecologically sustainable water management approaches. Such models are capable of performing simultaneous calculations of all the many influences on water flows, even in complex river systems. They can be used to evaluate river flow changes expected under proposed water management approaches, such as increased future human demands and associated operation of water infrastructure. Because short-term hydrologic conditions such as extreme low flows or floods can have tremendous ecological influence, it is highly desirable and increasingly feasible to develop hydrologic simulation models that operate on daily (or shorter) time steps.

2.3 Choice of Hydrological Method

One of the most important issues related to this study was the need to use data that was easily acquired in order to provide an approach as simple as possible for predicting an impending flood by reservoir operators. Two methods have been proposed that can be used to provide trigger information to reservoir operators: the rainfall method (Hughes et al., 2006) and the rate of rise criterion (Hughes et al., 1997). The former is the subject of this study.

The rainfall method depends on correlating the amount of flood at the IFR site caused by antecedent rainfall conditions in the previous day(s) in the catchment. With this method, reservoir operators would receive rainfall data on a daily basis from several gauges in the catchment and use this data as a decision criterion for effecting the releases.

The rate of rise criterion provides a probability matrix that identifies the likelihood of getting an event given a certain rate-of-rise of the flow hydrograph at the beginning of the event.

In maintaining the aim to keep the modelling simple, the three methods that were used in this study are: regression analysis, empirical modelling and artificial neural networks (ANN) modelling.

CHAPTER 3

3. STUDY AREA

3.1 General Description of the Area

The Tugela River also known as Thukela is the largest river in KwaZulu-Natal Province, South Africa. The river originates in the Drakensberg Mountains near Bergville, Mont-aux-Sources, itself being the source of tributaries of two other major South African rivers, the Orange River and the Vaal River, and flows 947 metres down the Tugela Falls. From the Drakensberg range, the river meanders eastwards for 520 km through the KwaZulu-Natal midlands before flowing into the Indian Ocean about 95 km north of Durban. The total Thukela catchment area is approximately 29,100 km². Major tributaries include the Little Thukela, Klip, Bushman's, Sundays, Mooi and Buffalo Rivers (the latter being the largest). Land uses in the catchment are mainly rural subsistence farming and commercial forestry. The figure below shows the location of the study area and the Thukela River.

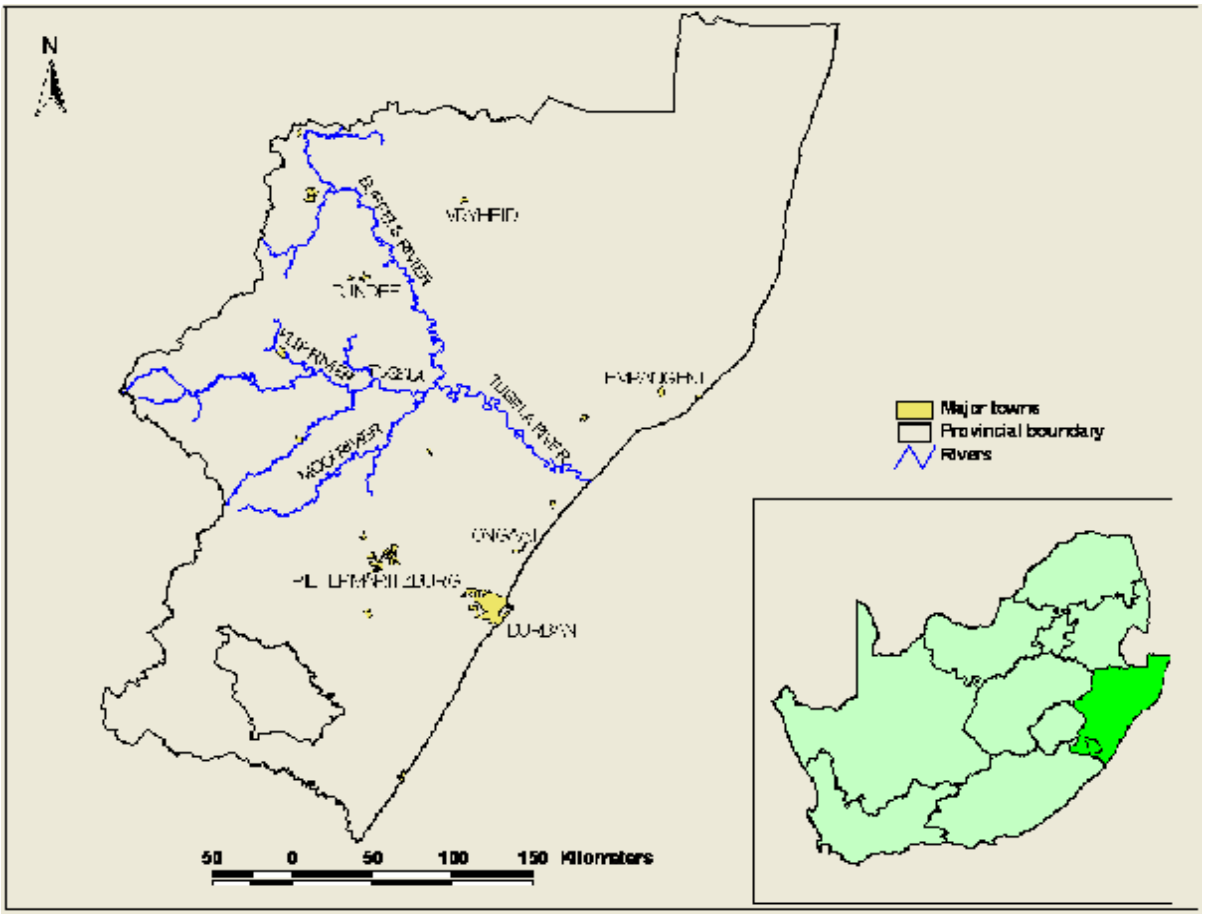


Figure 3-1: Overview of Study Area

3.2 Rainfall

The Thukela Catchment has a Mean Annual Runoff (MAR) of 3 865 x 10^6 m^3. The MAR expressed as average unit runoff is about 133 mm which is equivalent to 16% of the Mean Annual Precipitation (MAP) of 840 mm. However, MAP varies from more than 1 500 mm in the Drakensberg to 500 mm and less in the dry central regions of the basin.

3.3 Water Resource Developments and Tributaries

Although there are a few large dams and numerous smaller ones in the Thukela River System, they are mainly located in the upper reaches of the Thukela River itself and in some of its tributaries. For the most part, the Thukela River remains comparatively unregulated. Water resource developments within the catchment are generally small and relate primarily to the needs of individual towns. The largest components of existing water development infrastructure are those associated with four inter-basin transfer schemes:

The upper reaches of the Thukela River, upstream of the confluence with the Bushmans River, includes the towns of Bergville, Ladysmith, Colenso and Weenen. The Klip River is the main tributary in this area. This area is the source of water for the Thukela-Vaal Transfer Scheme, which, *inter alia*, transfers water to the Vaal River system. The transfer capacity of this scheme represents a large portion (about 30%) of the water resources available in the Upper Vaal Water Management Area, which is the economic heart of South Africa.

The catchment of the Little Thukela River, a tributary of the Thukela River, is characterised by large irrigation requirements (36 million m^3/a). Other water uses are insignificant. The only significant dam in this area is the small Bell Park Dam. The upper areas of the Little Thukela are located in a nature reserve and areas adjacent to the natural reserve have developed rapidly into popular tourist resorts. The Bushmans River rises in the Drakensberg Mountain range and flows

in a north-easterly direction past the town of Estcourt to join the Thukela River near the town of Weenen. The Sundays River flows in a south-easterly direction from the eastern escarpment to its confluence with the Thukela River near the Bushmans River confluence.

The Buffalo River is the main northern tributary of the Thukela River and flows in a southeasterly direction from the eastern escarpment (Newcastle area) to its confluence with the Thukela River near Nkandla.

The Mooi River rises in the Drakensberg Mountains and flows parallel to the Bushmans River in a north-easterly direction to join the Thukela River near Muden. The predominant land use in the catchment is commercial agriculture and there is large-scale irrigation of pastures and summer cash crops. The transfer scheme situated at Mearns can transfer water to the Mgeni River system.

3.4 IFR 3– LITTLE THUKELA

This site which is located close to the confluence with the Thukela, but before the last major tributary joins the Little Thukela. It is within quaternary catchment V13E where there are two DWAF gauges i.e. V1H010 which is situated further upstream at the outlet of quaternary V13C, and V1H039 situated at the outlet of quaternary V13A. Gauged V1H010 records which extend back to 1965 appear to

be relatively natural except with regard to low flows which has too many zero flows.

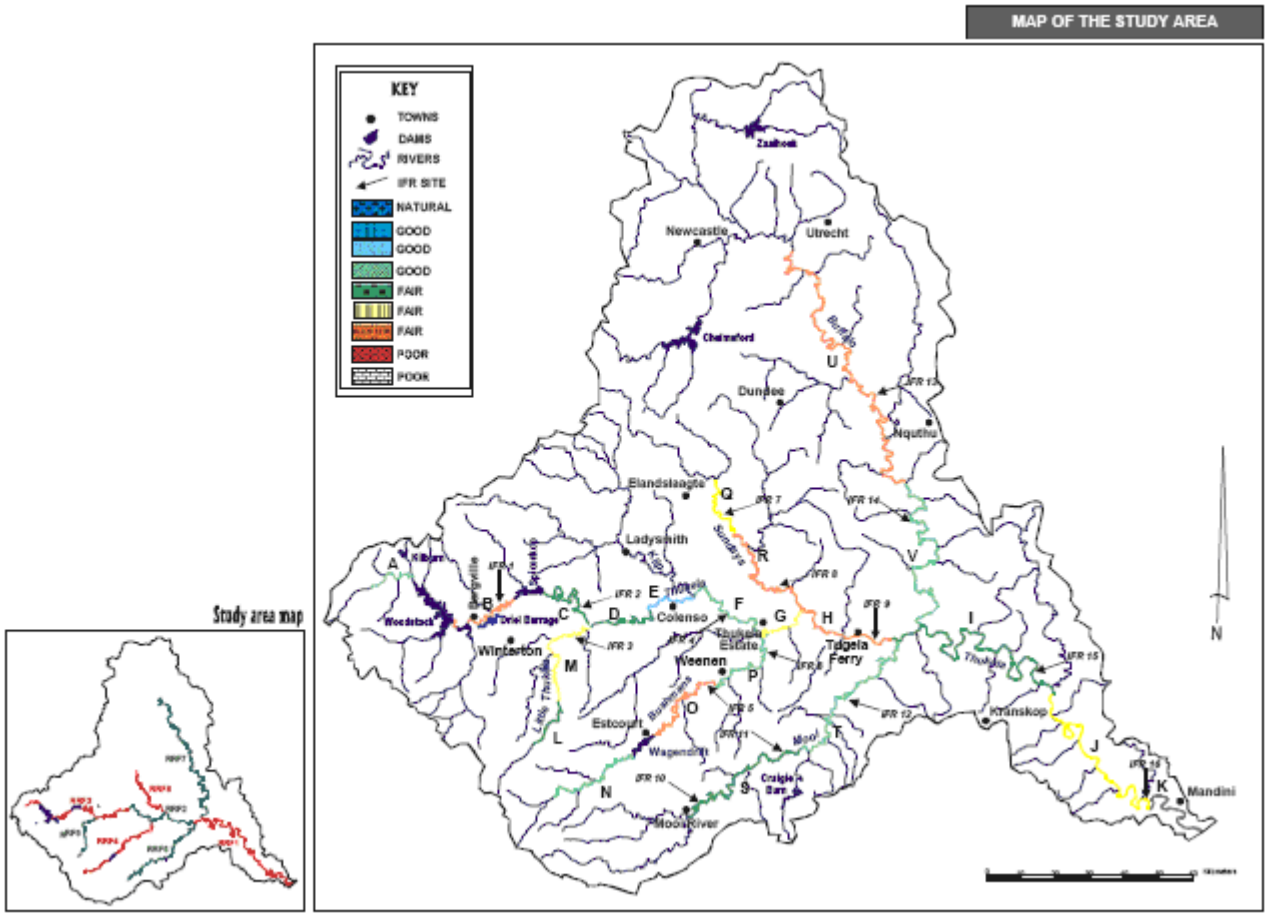


Figure 3-2: Study Area Showing IFR Sites

CHAPTER 4

4. METHODS AND DATA ACQUISITION

The most important aspects that need consideration in a bid to construct a good model of any process is to understand the physical elements that, to a greater or lesser extent, have a bearing on the phenomenon under consideration and the way each element affects the process. Stream-flow is affected by among other variables; antecedent rainfall, antecedent flow, soil moisture content, catchment area, vegetation cover, soil type, infiltration, evaporation, catchment terrain, agriculture practices, built environment developments in the area, etc. This study as mentioned above is aimed at establishing an approach that uses data that is easily obtainable. Rainfall is the most easily obtainable variable and consequently this study will focus on generating streamflow from catchment rainfall. The other elements will be ignored in this study in order to come up with a model that uses basic data i.e. rainfall. The effect of a rainfall event to drive streamflow diminishes with passing days and therefore the models considered will obtain antecedent wetness based on the rain that fell on the previous seven days. The seven days period has been chosen arbitrarily as the period within which rainfall events may have impact on streamflow. Three methods will be used to construct the models namely regression analysis, an empirical method and artificial neural networks (ANN).

One way of finding out if a suggested model is likely to produce satisfactory results is to analyze historical data. In this study, daily rainfall data from the South African Weather Service (SAWS) of the Monks Cowl (0267693W) and Heartsease (0299900W) gauges has been used as a possible trigger for releases to be made to satisfy the Reserve requirement at (or close to) IFR 3 site represented by stream-flow gauge V1H010 on the Little Thukela River. The flows used do not represent natural historical conditions, but are naturalized flows that can be used to give an indication of the range of events that actually occurred in this part of the Thukela basin. The choice of IFR 3 was based on the advantage of using a long period of record of hydrological data for both rainfall and stream flow and the rain gauges data which appears to be relatively natural.

The Department of Water Affairs and Forestry (DWAF) manages the repository of all hydrological data in South Africa. Hydrological and meteorological data obtained from the Department of Water Affairs and Forestry (DWAF) and the South African Weather Service (SAWS) respectively were used as inputs for the development of the models.

4.1 Relevant Data Sets

Necessary datasets for the project were IFR stream-flow from a streamflow gauge and rainfall data from two raingauges in the catchment with their record periods as shown in Table 4-1 below:

Table 4-1: Rainfall and Stream-Flow Measuring Stations

Data	Station No.	Location	Available Record
Rainfall	0267693W	Monks Cowl	1962-2006
Rainfall	0299900W	Heartsease	1927-2006
Streamflow	V1H010A01	Little Thukela	1965-2004

4.2 Stream-Flow Data

There are several sources of hydrological information for the Thukela River and its tributaries. First there are a large number of DWAF stream-flow gauges within the Thukela basin, all of which are impacted to various degrees by upstream influences (abstractions, stream flow reduction activities, reservoir operation, return flows etc). The records associated with them all have different lengths, differing degrees of accuracy, and are able to measure high flows to different extents.

Naturalized daily flows (1965-2004) for river gauge VH0101 (IFR site 3) on the Little Thukela River were used for this study and are available on the Department of Water Affairs and Forestry (DWAF) website. These flows were generated using the patching model (Hughes and Shange 2004). Peak flows, however, are somewhat excessively high in certain years while the lowest flows have been estimated with low confidence (Hughes and Shange 2004). The location of IFR 3 site is shown in Figure 4-1.

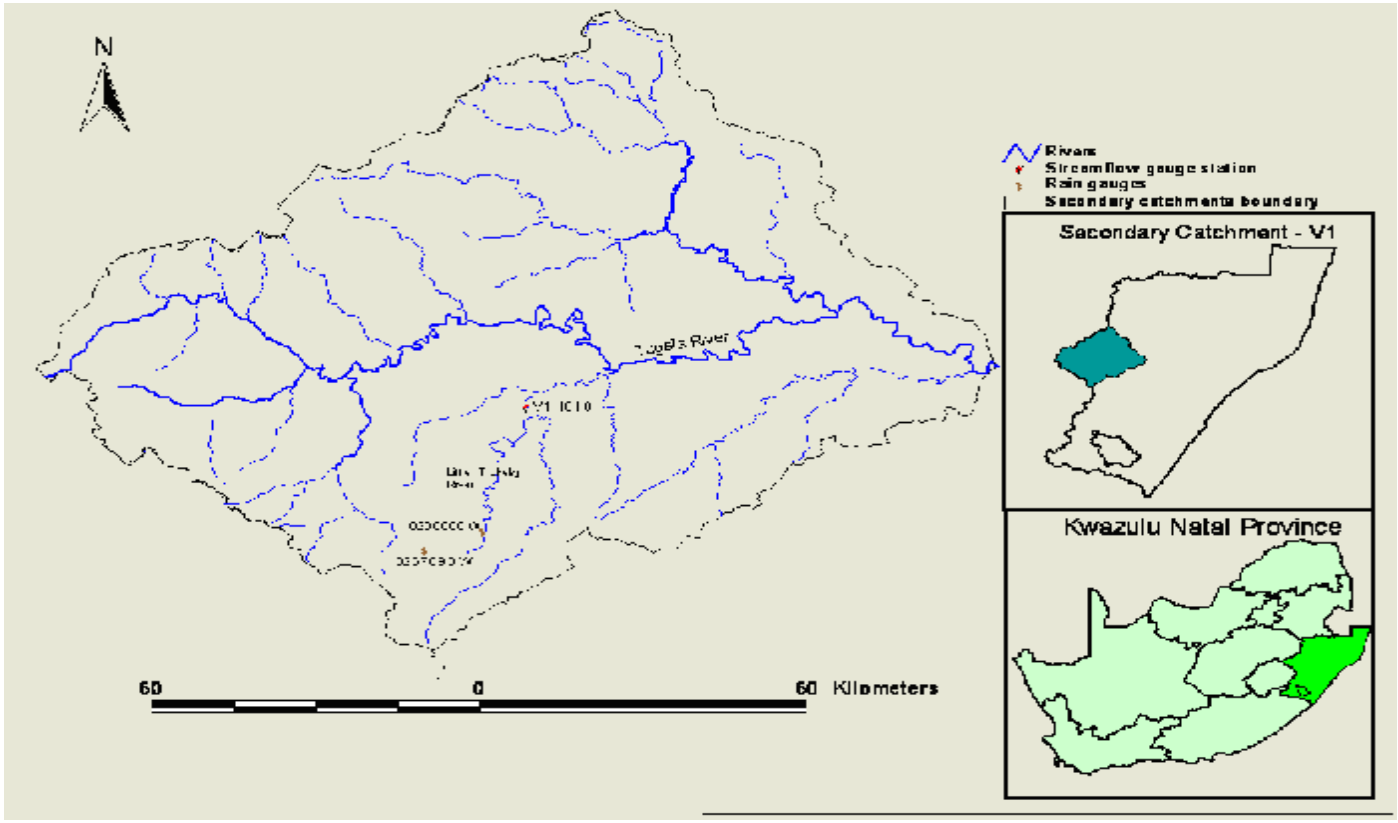


Figure 4-1: Location of IFR 3

4.3 Rainfall Data

The rainfall method of flow prediction is based on the fact that river flow variation is driven by antecedent rainfall that has occurred in the river catchment. Part of a time series of data for a typical rain season for the months of December 1966 to march 1967 is shown in Figure 4-2 below. Here, rainfall data time series for Monks Cowl and Heartsease rain gauges have been plotted together with the flow time series data to show how antecedent rainfall varies with or drives resulting flow events. Antecedent rainfall is represented by the sum of the rainfall recorded during the preceding seven days to the high flow event.

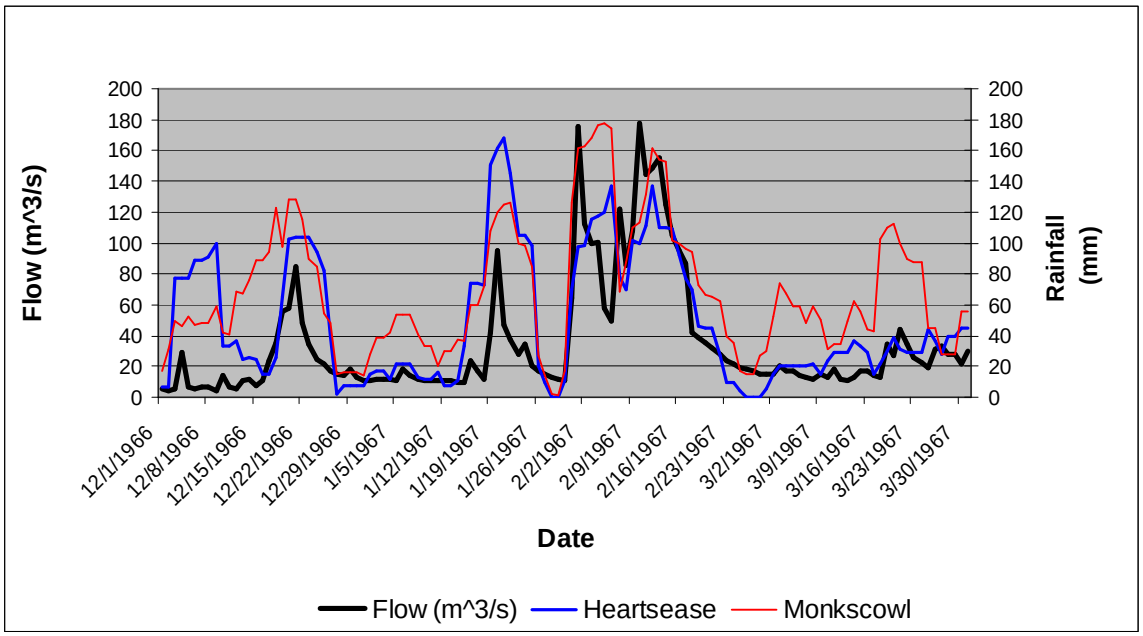


Figure 4-2: Time-series of Rainfall and Flow

4.4 Required Conditions for Satisfaction of the Reserve

Table 4-2 below presents required conditions to satisfy high flow IFR requirements as recommended in a recent ecological reserve study (Hughes et

al., 2004). The required conditions include a flow range, duration in days and frequency in a month or period in years.

Table 4-2: Required Conditions to Satisfy the IFR

Month	Required Flow Range (m^3/s)	Required Duration (days)	Required Frequency
December	10- 24	2	3 in a month
January	25- 39	3	2 in a month
February	125- 250	5	1 in a month
March	25- 39	3	1 in 4 years

Source: Hughes et al., 2004

Different methods were used to arrive at flow requirement/ IFR (quantity component of the Ecological reserve) for each site. A method derived from the standard building block methodology (BBM, King and Louw, 1998) and the Downstream response to Imposed Flow Transformation (DRIFT; Brown and King 2000) was used to establish high flows, while the Flow-Strength response method (FS-R) was used to obtain low flows. Specialists that study fish, invertebrate and riparian vegetation estimate the functions of floods described as flood classes and identify the sizes of the events using the hydraulic cross sections, sediment transport modeling, photos and videos of known flows and interactions with geomorphologies. The number of events of each flood class under natural conditions are then determined. The number of events is subsequently recommended depending on the objectives for the Ecological Reserve Class (ERC).

Table 4-3 below shows the number of months out of the complete flow record (1965-2004) that the high flow IFR requirements were satisfied at a river gauging station located close to the IFR site.

Table 4-3: Number of Months in which IFR Events are Satisfied

Month	No. of Months with Satisfied Events
December	9
January	3
February	1
March	15

Evidently the IFR wasn't satisfied for most of the 39 year period except for the month of March which exhibits more than the required number of events. This finding questions the validity of the recommended IFR requirements for the site and suggests that these may need to be reassessed.

Due to the need to try a simple model first and finding how well it works before imposing more rigorous tests on it, it was decided to ignore the other requirements i.e. the required number of events in a month and the duration of the required flow. Only the possibility of predicting the range of flow has been investigated in the study. Use of one day events with the required flow range was

adopted to test the validity of the rainfall method to forecast floods with a 1 day lead time. The whole set of data available consisting of 14,295 datasets (days) was used in the analysis. This constitutes rainfall and stream flow data between 08/10/1965 and 26/11/2004.

CHAPTER 5

5. REGRESSION ANALYSIS

Regression analysis is a statistical tool for the investigation of relationships between variables. Usually, the objective is to ascertain the causal effect of one variable upon another. Regression analysis is used for explaining or modeling the relationship between a single variable Y, called the response, output or dependent variable, and one or more predictor, input, independent or explanatory variables, $X_1 \ldots X_p$. When p=1, it is called simple regression but when p>1 it is called multiple regression or sometimes multivariate regression. When there is more than one Y, then it is called multivariate multiple regression. Data on the underlying variables of interest is assembled and regression is employed to estimate the quantitative effect of the causal variables upon the variable that they influence.

5.1 Performance Criteria

Performance of the suggested models in this study was evaluated using the coefficient of determination, the mean absolute error, the root mean square error, correlation coefficient and the Nash-Sutcliffe model efficiency coefficient (Nash and Sutcliffe 1970). “Statistical significance” of the estimated relationships, i.e. the degree of confidence that the estimated relationship was close to the true relationship was also investigated.

The coefficient of determination (R^2) is useful for measuring the strength of a linear relationship between variables or how well a model fits the data determined from the formula:

$$R^2 = 1 - \frac{F}{F_0} \qquad \textbf{(5-1)}$$

where

$$F = \sum_{n=1}^{N} (Q_{obs}(i) - Q_{cal}(i))^2; \qquad \textbf{(5-2)}$$

$$F_0 = \sum_{n=1}^{N} (Q_{obs}(i) - Q_{ave})^2 \qquad \textbf{(5-3)}$$

where

F_0 is the initial variance for the flows and

F is the residual model variance.

N is the total number of data sets,

Q_{obs} and Q_{cal} are the observed and simulated (forecast) flows respectively

Q_{ave} is the mean value of the observed flows.

The higher the value of R^2, the better is the forecasting performance. If R^2 is equal to 1, it implies that the forecast replicates observations 100% of the time. If all forecast values are equal to the long-term observed mean, R^2 would assume

a value of 0. If R^2 is less than 0, however, it implies the forecast is worse than the long-term observed mean.

The accuracy of the model simulation can also be evaluated using the mean absolute error (MAE) that gives a quantitative indication of the model error in terms of a dimensioned quantity. The mean absolute error is a quantity used to measure how close forecasts or predictions are to the eventual outcomes:

$$MAE = \frac{1}{N}\sum_{n=1}^{N}\left|(Q_{obs}(i) - Q_{cal}(i))\right|; \qquad \textbf{(5-4)}$$

MAE describes the difference between the model simulations and observations in the units of the variable (Legates & McCabe, 1999).

To quantify the errors with the same units as the quantity being estimated, the Root Mean Square Error (RMSE) is used and is defined as:

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}\left[Q_{obs}(i) - Q_{cal}(i)\right]^2\right]^{\frac{1}{2}}; \qquad \textbf{(5-5)}$$

The RMSE is most useful when large errors are particularly undesirable.

The MAE 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 MAE;

the greater the 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 coefficient of correlation which is formulated as:

$$r = \left\{ \frac{\sum_{i=1}^{N} (Q_{obs}(i) - Q_{ave})(Q_{cal}(i) - \bar{Q})}{\left[\sum_{i=1}^{N} (Q_{obs}(i) - Q_{ave})^2 \right]^{\frac{1}{2}} \left[\sum_{i=1}^{N} \left(Q_{cal}(i) - \bar{Q} \right)^2 \right]^{\frac{1}{2}}} \right\} \qquad \textbf{(5-6)}$$

where $\bar{Q}$ is the mean of the simulated values, and measures the strength and the direction of a linear relationship between two variables.

A correlation greater than 0.8 is generally described as strong, whereas a correlation less than 0.5 is generally described as weak. These values can vary based upon the type of data being examined. The correlation coefficient is not a measure of the predictive capabilities of the model since it is sensitive to outliers and spurious data.

The Nash-Sutcliffe model efficiency coefficient which is used to assess the predictive power of hydrological models is defined as:

$$ce = 1 - \frac{\sum_{t=1}^{T}(Q_{obs}^{t} - Q_{m}^{t})^{2}}{\sum_{t=1}^{T}(Q_{obs}^{t} - \bar{Q}_{obs})^{2}} \qquad \textbf{(5-7)}$$

where:

ce is the coefficient of determination

Q_{obs} is observed discharge,

Q_{m} is modeled discharge,

Q_{obs}^{t} is observed discharge at time t,

$\bar{Q}_{obs}$ is mean of observed discharges.

Nash-Sutcliffe efficiencies can range from $-\infty$ to 1. An efficiency of 1 (ce=1) corresponds to a perfect match of modeled discharge to the observed data. An efficiency of 0 (ce=0) indicates that the model predictions are as accurate as the mean of the observed data, whereas an efficiency less than zero ($-\infty$<ce<0) occurs when the observed mean is a better predictor than the model.

The above tests were applied to both the complete dataset and the datasets composed of only flood events for each month considered in this study i.e. December to march. The criteria used to decide on what event was large for the individual months was based on Table 4-2.

5.2 Model Development

5.2.1 Simple Regression Analysis

Single regression analysis was employed to investigate the correlation between the Monks Cowl and Heartsease rainfall datasets. This was to check whether there was a strong correlation between them that would allow using only one of them to predict floods. Alternatively data from both raingauges would be used if results show that the correlation was high meaning that the model developed would be improved by the use of both datasets. This will again be confirmed by running regression analyses for each rainfall data set and both datasets in a multi regression analysis.

Correlation between Heartsease rainfall data and IFR site 3 (V1H010) flow data was checked followed by Monks Cowl rainfall data. This involved running flow data separately with one day's antecedent rainfall data for each raingauge station in single-regression analyses. The models referred to here that include rainfall data from only one raingauge are represented by:

$$Flow_{(j)} = K\,x\,\text{Rainfall}_{(j-1)} \qquad \textbf{(5-8)}$$

where:

Flow $_{(j)}$ = flow on the j^{th} day

Rainfall $_{(j-1)}$ = rainfall on previous day to j, i.e. rainfall during the past 24 hours.

K = compound coefficient representing other factors that have impact on streamflow.

Following these analyses it was necessary to establish whether by using datasets from both raingauges the correlation results between rain and flow could be improved. This called for the use of multi-regression analysis.

5.2.2 Multiple Regression Analysis

Two options of antecedent days' rainfall datasets are suggested and used in the analyses as outlined below:

i. One day's antecedent rainfall data from both gauges, i.e. Heartsease and Monks Cowl as the independent variables, ran simultaneously with flow data, from the flow gauge V1H010 as the dependent variable, in a multi-regression analysis. This analysis informed whether it would be advantageous to use both rain gauge datasets. The model considered here is represented by the following equation:

$$Flow_{(j)} = K_1 x R_{(j-1)MC} + K_2 x R_{(j-1)H} \qquad \textbf{(5-9)}$$

where:

Flow $_{(j)}$ = flow on the j^{th} day

R $_{(j-1)\ MC}$= rainfall on previous day to j, i.e. rainfall during the past 24 hours at Monks Cowl.

$R_{(j-1)H}$= rainfall on previous day to j, i.e. rainfall during the past 24 hours at Heartsease.

K = compound coefficient representing other factors that have impact on streamflow.

ii. Seven days' antecedent rainfall for both data sets ran in a multi-regression analysis.

$$Flow_{(j)} = \sum_{i=1}^{7} \left(K_{(i)MC}\, x\, R_{(j-i)MC} + K_{(i)H}\, x\, R_{(j-i)H} \right) \qquad \textbf{(5-10)}$$

where:

$Flow_{(j)}$ = flow on the j^{th} day

$R_{(j-1)MC}$ = rainfall on previous day to j, i.e. rainfall during the past 24 hours at Monks Cowl

$R_{(j-1)H}$ = rainfall on previous day to j, i.e. rainfall during the past 24 hours at Heartsease

$K_{(i)}$ = compound coefficient representing other factors that have impact on stream flow.

5.3 Results Analysis and Discussion

The correlation coefficient was treated as a measure of the extent to which two measurement variables vary. More precisely, it was used to examine each pair of measurement variables to determine the relationship between the two

measurement variables. This involved determining whether large values of one variable tended to be associated with large values of the other (positive correlation); whether small values of one variable tended to be associated with large values of the other (negative correlation); or whether values of both variables tended to be unrelated (near correlation zero). The correlation coefficient obtained for the analysis between Monks cowl and Heartsease data was 0.44 which signifies a positive poor correlation between the two datasets. This implies that the use of both rainfall datasets may yield better results for modeling purposes than using either set alone.

Statistics for testing model performance developed by using flow data and one day's antecedent rainfall data for Monks Cowl are shown in Table 5-1 below. The models produce poor results for all months including the complete dataset. The coefficients of determination and correlation coefficients are too low to give reliable predictions.

Table 5-1: Single Regression (one antecedent day) Analysis Results for the Monks Cowl Data

Statistic	Whole Dataset	Dec	Jan	Feb	Mar
Coefficient of Determination, R^2	0.15	0.03	0.18	0.07	0.02
Mean Absolute Error (m^3/s), MAE	17.71	11.05	11.79	33.60	7.09
Root Mean Square Error (m^3/s), RMSE	2.98	1.82	1.65	4.12	1.83
Correlation Coefficient, r	0.28	0.40	0.43	0.27	0.15
Efficiency Coefficient, E	0.08	0.17	0.18	0.07	0.02

Table 5-2 below shows results for the model developed using Heartsease data. The results were no better than those produced from the previous Monks Cowl data.

Table 5-2: Single Regression (one antecedent day) Analysis Results for the Heartsease Data

Statistic	Whole Dataset	Dec	Jan	Feb	Mar
Coefficient of Determination, R^2	0.16	0.08	0.08	0.11	0.003
Mean Absolute Error (m^3/s), MAE	19.48	11.41	11.29	33.25	7.59
Root Mean Square Error (m^3/s), RMSE	3	1.82	1.59	4.10	1.83
Correlation Coefficient, r	0.24	0.39	0.27	0.33	0.05
Efficiency Coefficient, E	0.17	0.15	0.08	0.11	0.002

Despite the results still being poor, multi-regression analysis combining Heartsease and Monks Cowl data (Table 5.3) yielded slightly better results than either of the above models. For this reason it is prudent to use both sets of data to develop a multi-regression model.

Table 5-3: Multi Regression (one antecedent day for each) Analysis Results for the Heartsease and Monks Cowl Data

Statistic	Whole Dataset	Dec	Jan	Feb	Mar
Coefficient of Determination, R^2	0.20	0.17	0.20	0.12	0.03
Mean Absolute Error (m^3/s), MAE	22.34	11.05	9.94	33.40	1.82
Root Mean Square Error (m^3/s), RMSE	2.96	1.82	1.60	4.10	7.00
Correlation Coefficient, r	0.30	0.41	0.44	0.33	0.17
Efficiency Coefficient, E	0.09	0.16	-0.06	0.11	0.03

Results from the model consisting of seven antecedent days' rainfall data for both gauge stations were apparently much better as seen in Table 5.4. Coefficients of determination values ranged from 0.40 to 0.49 except for the month of February which exhibited very good result at 0.96. The explanation to this improved result may have been due to the data used that was very limited and comprised of only nineteen sets. Coefficients of correlation range from 0.64 to 0.70 except for the value of 0.98 for February. From the correlation coefficients the results show that there is good correlation between the modeled and observed data. However it must be remembered that this statistic does not show the predictive capability of the model as mentioned before and therefore the model may not necessarily produce satisfactory predictions of flow.

Table 5-4: Multi-regression analysis results with seven antecedent days' rainfall for both Monks Cowl and Heartsease

Statistic	Whole Dataset	Dec	Jan	Feb	Mar
Coefficient of Determination, R^2	0.49	0.40	0.44	0.96	0.43
Mean Absolute Error (m^3/s), MAE	21.60	7.92	7.23	6.64	5.72
Root Mean Square Error (m^3/s), RMSE	2.56	1.73	1.42	1.76	1.53
Correlation Coefficient, r	0.70	0.64	0.65	0.98	0.66
Efficiency Coefficient, E	0.488	0.41	0.43	0.97	0.43
Bias					

Figure 5-1 below shows part of the data for the observed and simulated flows data for the analysis done using Monks Cowl and Heartsease data with seven days antecedent rainfall. It is observed from this graph that the model apparently over estimates low flows and underestimates high flows. Due to the fact that this

study is aimed at predicting high flows, the model developed in this way may miss the cue to make reservoir releases as it will still record a lower forecast of a flood event. This renders the model of limited use for this purpose.

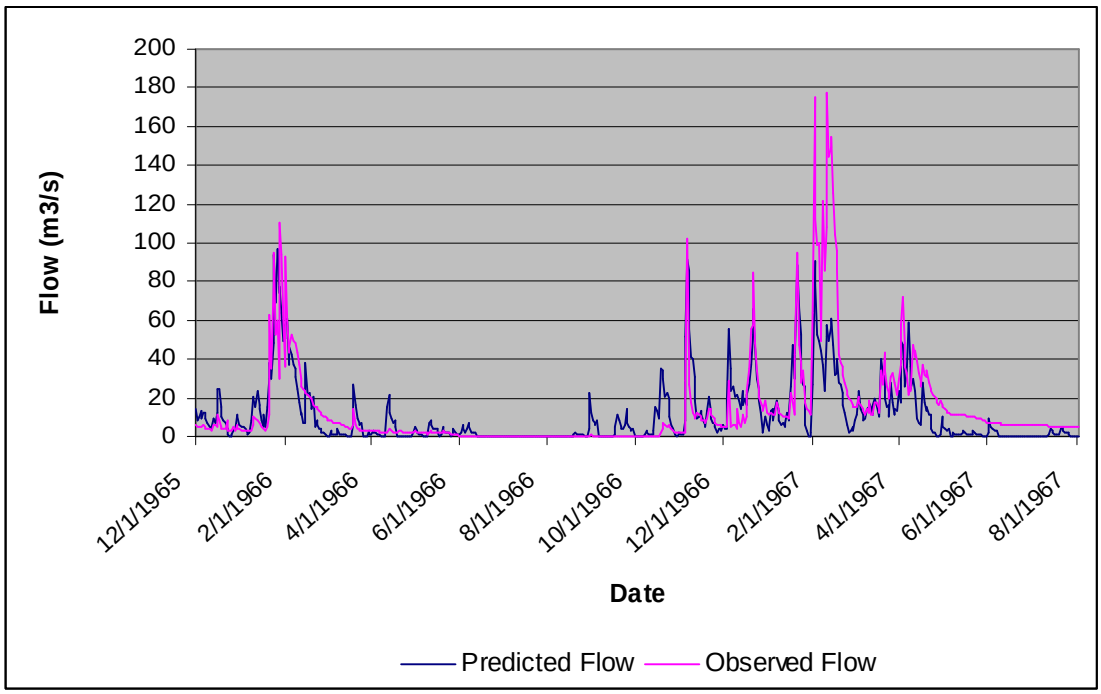


Figure 5-1: Comparison of Observed and Simulated Flows Data from Multiple Regression

CHAPTER 6

6. EMPIRICAL MODEL

Flood forecasting requires a powerful model for simulating complicated rainfall-runoff processes. Generally, “hydrological processes are usually assumed to be nonlinear and therefore need to be represented by nonlinear equations consisting of a number of members and an intercept” (Tsykin, 1985). Conceptual rainfall-runoff models (CRRM) are the tools most commonly applied to represent watershed rainfall runoff relationships. These models are designed for simulating the physical processes that are subject to the hydrologic cycle. These models include a large number of parameters that cannot be measured directly but have to be estimated on the basis of a calibration process i.e. minimizing an objective function (OF). The accuracy of their calculations depends on how the relevant parameters are defined. The main goal for calibration is therefore to assign optimized values for the different parameters in the model. The search for the minimum of the objective function (OF) in the case of CRRMs is, however, somewhat complex.

In recent years, different researchers have tried to develop different optimization models to calibrate rainfall-runoff models. “Local” or “global” search methods can be used to solve optimization problems. The local search methods such as gradient-based methods and direct search methods have been widely applied in

water resources management (Yeh, 1985). The advantages of these methods are that they are effective and efficient when applied for optimization of convex, single extremum functions. However, for more complex functions they may produce a local optimal solution. In fact, there are many functions such as non-convex, non-differentiable and multi extrema functions, which cannot be effectively solved by local methods. It requires more robust optimization techniques to find the global optimum solution of complex problems. Duan et al. *(*1992) presented a global optimization method called the shuffled complex evolution SCE-UA that has become the most popular calibration method among hydrologists.

The Shuffled Complex Evolution (SCE-UA) method is a general purpose global optimization evolutionary programming technique which combines the strength of the "simplex search" with the "concept of controlled random search", "competitive evolution" and "complex shuffling" (Duan et al., 1994). It conducts an efficient and robust search of the parameter space and has been widely applied in calibrating various conceptual models (e.g. Duan et al., 1994; Yapo et al., 1998; Madsen 2000; Eckhardt and Arnold, 2001; Brath et al., 2004).

6.1 Development of the Model

Rainfall run-off processes are best represented by nonlinear models. Regression analysis in Chapter 5 showed that combining data from the two raingauge stations improved results and therefore the empirical model proposed below

includes rainfall data from the two gauge stations. One important parameter that could have been used in combination with antecedent day's rainfall was antecedent flow, more especially for the previous day as this may improve the model significantly. Antecedent streamflow would however not be applicable for purposes of forecasting flows because the real time historic streamflow would not be the natural one but one with all the human impacts that (in the first place) lead to the need for the implementation of IFR. The model below that simply relates flow to antecedent rainfall has therefore been suggested:

$$Flow_{(j)} = \sum_{i=1}^{7} \left(K_{(i)MC}\, x\, R_{(j-i)MC}{}^{P_{(i)}} + K_{(i)H}\, x\, R_{(j-i)H}{}^{P_{(i)}} \right) + C \qquad \textbf{(6-1)}$$

Where:

Flow $_{(j)}$ = flow on the j^{th} day

R $_{(j-1)MC}$ = rainfall on previous day to j, i.e. rainfall during the past 24 hours at Monks Cowl

R $_{(j-1)H}$ = rainfall on previous day to j, i.e. rainfall during the past 24 hours at Heartsease

K $_{(i)MC}$ = compound coefficient representing other factors that have impact on stream flow for Monks Cowl

K $_{(i)H}$ = compound coefficient representing other factors that have impact on stream flow for Heartsease

C= is a constant

P($_{i(i)}$= parameters that introduces nonlinearity in the equation

The objective of this study was to analyze high flows and therefore the high flows for the months which are the subject of the study were extracted. The high flows were considered to be those which fall within the recommended range that satisfies the IFR for a particular month. Together with these high flows, for each value of flow corresponding rainfall data for seven antecedent days were also extracted obtaining 15 model inputs i.e. one for flow and seven for Monks Cowl and another seven for Heartsease.

Model calibration was carried out using the SCE-UA to obtain optimized values for the parameters of the equation i.e. $K_{(i)MC}$, $K_{(i)H}$, $P_{(i)}$ (i=1 to 7) and C. Table 6-1 below shows the number of datasets used in the analyses.

Table 6-1: Number of Datasets Used for calibration.

Whole Dataset	Dec	Jan	Feb	Mar
7146	179	43	20	45

Ten randomly initialized calibrations were carried out. Out of the 10 sets of parameter lists produced, the one with the smallest objective function was selected for use for generation of simulated flow data.

The objective function applied for this empirical model is

$$Minimize \sum \left(\left| simf_t - observedf_t \right| \right) \qquad \textbf{(6-2)}$$

where:

$simf_t$ is the simulated flow

$observedf_t$ is the observed (recorded) flow

6.2 SCE-UA Algorithm

The SCE-UA method requires the initial selection of a "population" of points distributed randomly throughout the feasible parameter space. The population is then partitioned into several "complexes", each containing *(2n + 1)* points, where *n* is the number of parameters to calibrate. Each complex evolves independently according to a "reproduction" process that, in turn, uses the Simplex Method (Nelder & Mead, 1965). At periodic stages, the entire population is shuffled and points are reassigned to new complexes formed so that the information gained by the previous complexes is shared. The evolution and the shuffling steps continue until prescribed convergence criteria are reached. In the current study all the parameters and stop criteria controlling SCE-UA action are set at the suggested values indicated in Duan *et al. (*1994). Further detailed explanation of the method is given in Duan *et al. (*1992, 1993, and 1994).

A brief description of algorithm steps is given here:

1. **Generate Sample**

A sample of points, i.e. parameter sets or sets of decision variables are randomly generated from the feasible parameter space. For each parameter set the

objective function value is calculated. The initial sample has the size s = pr where p is the number of complexes and r is the number of points in each complex.

2. Rank Points

Sort the s points in order of increasing objective function value so that the first point represents the point with the smallest objective function value (best point) and the last point represents the point with the largest objective function value.

3. Partition into Complexes

Partition the s points into p complexes, such that the first complex contains every p (k-1) + 1 ranked point, the second complex contains every p(k-1) + 2 ranked point, and so on, where k=1,2…r.

4. Evolution

A sub-complex of size q is formed from the complex by randomly choosing q points. A triangular probability distribution is used for assigning the probability of a point to be included in the sub-complex (larger probability for points with smaller objective function value). The sub-complex is evolved according to the simplex algorithm (Nelder and Mead, 1965). Each complex is evolved β times.

5. Complex Shuffling

Combine the points in the evolved complexes into a single sample of s points and return to step 2.

Steps 2-5 are repeated until one of the criteria for termination is satisfied.

The SCE algorithm constitutes different algorithmic parameters, which must be chosen carefully. Ranges and recommended values for these parameters have been established through studies undertaken by Duan et al. (1994) (Table 6-2).

Table 6-2: Algorithmic parameters for the SCE algorithm

Parameter	Description	Range	Recommended Value
p	Number of complexes	$p \geq 1$	
r	Number of points in a complex	$r \geq 2$	2n+1
q	Number of points in a sub-complex	$2 \leq q \leq r$	n+1
β	Number of evolution steps taken by each complex before shuffling	$\beta \geq 1$	2n+1
p_{min}	Minimum number of complexes required in the population	$1 \leq p_{min} \leq p$	p

(n is the number of decision variables)

The number of complexes p is the most important algorithm parameter. Generally, a large value of p will give a higher possibility of converging into the global optimum but it requires a larger number of model evaluations, and vice versa. In the application by Madsen (2003) two complexes in SCE-UA provided a reasonable compromise between robustness and computing time.

6.3 Analysis of Results and Discussion

Table 6-3 below show final parameter values obtained for the proposed model. Parameters progressively reduce in value from p1 to p7 (Monks Cowl parameters) and p8 to p14 (Heartsease parameters) in line with the notion that the impact of rainfall on the magnitude of stream flow diminishes as the lag increases.

Table 6-3: Parameters of calibrated empirical model.

Data/ Parameter	Complete Dataset	December	January	February	March
p1	0.2821	0.5723	0.863	0.822	0.7144
p2	0.355	0.548	0.5519	0.8915	0.5458
p3	0.2858	0.8064	0.5654	0.9019	0.628
p4	0.4471	0.6196	0.4867	0.561	0.4398
p5	0.4479	0.4822	0.4975	0.5793	0.4791
p6	0.4288	0.3991	0.4641	0.5023	0.4069
p7	0.3179	0.3639	0.3799	0.3301	0.3478
p8	0.133	0.767	0.7917	0.7779	0.8928
p9	0.4303	0.8888	0.6827	0.7666	0.7998
p10	0.4372	0.6442	0.6053	0.5111	0.5954
p11	0.5091	0.36	0.5673	0.6455	0.4648
p12	0.3977	0.5589	0.5616	0.5209	0.4937
p13	0.3988	0.4235	0.4215	0.4401	0.4208
p14	0.3524	0.277	0.2996	0.3589	0.3546
p15	0.6818	0.2786	0.5945	0.7528	0.5816
p16	0.8037	0.4816	0.7123	0.9836	0.7447
p17	0.75	0.541	0.5922	1.1013	0.497
p18	0.6183	0.287	0.6924	0.9951	0.6408
p19	0.6879	0.4593	0.6135	0.9765	0.7478
p20	0.3728	0.1298	0.6886	0.8491	0.5529
p21	0.8779	0.4844	0.6945	0.8961	0.9966
p22	1.2874	1.8096	1.3312	1.363	1.9714
Objective Function	89015.54	659.7228	249.9003	375.8999	354.4198

The parameters obtained were used to generate time series of stream flow using the model and then they were analyzed together with the dataset of the observed flow to assess model adequacy. The results are tabulated below.

Table 6-4: Empirical model Calibration

Statistic	Complete Dataset	Dec	Jan	Feb	Mar
Coefficient of Determination, R^2	0.36	0.02	0.13	0.09	0.02
Mean Absolute Error (m^3/s), MAE	281	23.13	23.36	115.27	48.53
Root Mean Square Error (m^3/s), RMSE	277	222	269	6.77	3.14
Correlation Coefficient, r	0.60	0.13	0.36	0.31	0.14
Efficiency Coefficient, E	0.36	-0.88	0.13	-5.81	-10.50

It is clear that the results for all the sets show weak performance on the basis of all five criteria and therefore rendering the model unsatisfactory for predicting flow.

From the graph below, it is observed that generally the model underestimates high flows and that there is however good correspondence of seasonal high and low flows.

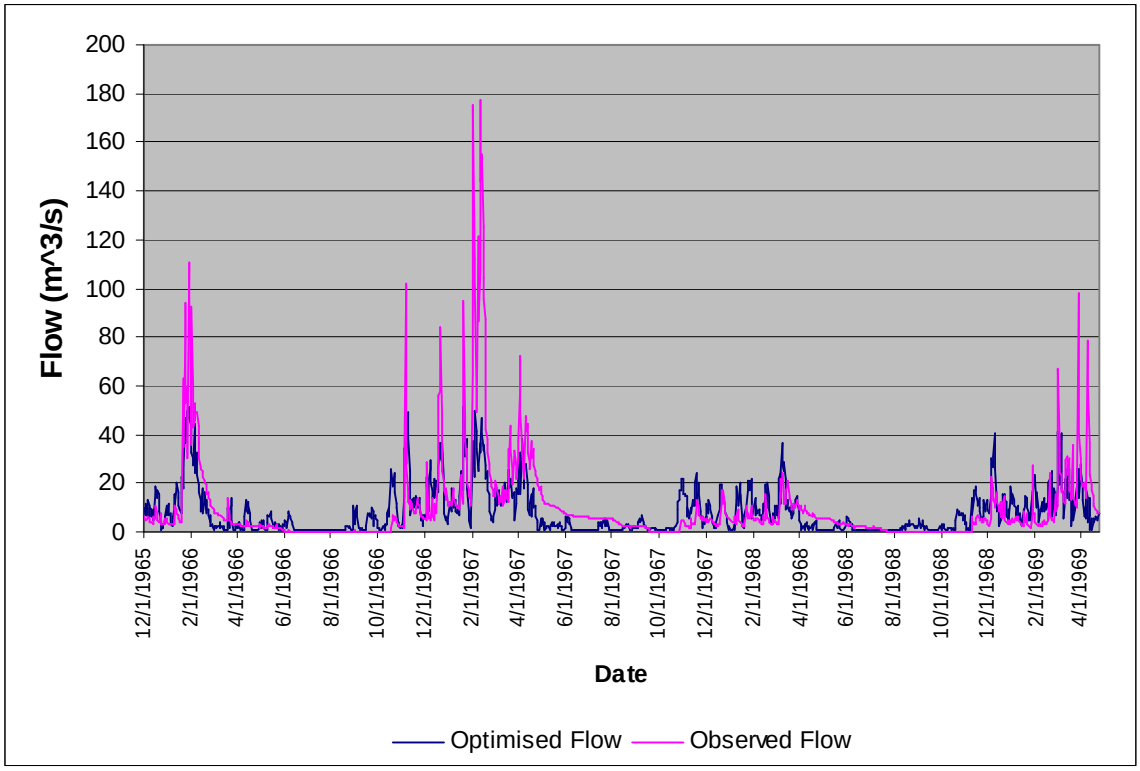


Figure 6-1: Observed and calibrated stream-flow time series of empirical model

CHAPTER 7

7. ARTIFICIAL NEURAL NETWORKS

7.1 Introduction

An artificial neural network (ANN), often just called a "neural network" (NN), is a mathematical model or computational model based on biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.

In more practical terms neural networks are non-linear statistical data modeling tools that can be used to model complex relationships between inputs and outputs or to find patterns in data. ANNs are general-purpose techniques that are increasingly being used for nonlinear data-driven rainfall–runoff modeling.

These networks are similar to biological neural networks in the sense that functions are performed collectively and in parallel by the units, rather than there being a clear delineation of subtasks to which various units are assigned.

ANNs are gaining popularity for use in many fields due to the following capabilities (Jain and Singh, 2003; Zealand, *et al*., 1999; Jain *et al*., 1999):

a) They can learn relationships between input and output variables even with unknown underlying physical laws
b) They use simple mathematical equations
c) They adapt to solutions over time
d) They are easy to use once they have been trained
e) They work well even when the training sets are incomplete

7.2 Network Topology

An ANN involves a network of simple processing elements (neurons) which can exhibit complex global behavior, determined by the connections between the processing elements and element parameters. The original inspiration for the technique was from examination of the central nervous system and the neurons (and their axons, dendrites and synapses) which constitute one of its most significant information processing elements. In a neural network model, simple nodes (called variously "neurons", "neurodes", "PEs" ("processing elements") or "units") are connected together to form a network or layers of nodes — hence the term "neural network". Nodes are classified as input, hidden or output layer nodes depending on their location in the network and function. Input layer nodes receive information from external sources, while output layer nodes transmit information out of the neural network. Hidden layer neurons act as the computational nodes, providing means of communication between input nodes

and other hidden layer or output nodes. The number of nodes in the input layer is equal to the number of independent variables entered into the network while the number of output nodes corresponds to the number of variables to be predicted. While a neural network does not have to be adaptive per se, its practical use comes with algorithms designed to alter the strength (weights) of the connections in the network to produce a desired signal flow.

7.3 Architecture of ANN

Artificial neural networks (ANN) are created by simple clustering of artificial neurons. This clustering occurs by creating layers which are then connected to one another. An example of ANN architecture is shown in Figure 7-1 below.

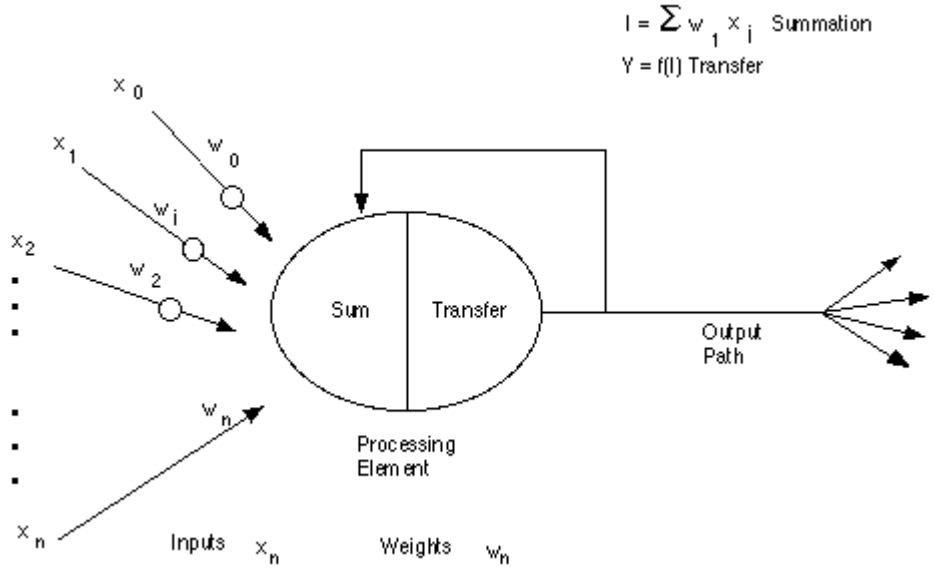


Figure 7-1: A Simple Neural Network Diagram.

In Figure 7-1 various inputs to the network are represented by the mathematical symbol, x (n). Each of these inputs is multiplied by a connection weight represented by w (n). In the simplest case, these products are simply summed, fed through a transfer function to generate a result, and then output.

Basically, all artificial neural networks have a similar structure or topology as shown in 7.2. In the structure some of the neurons interface to the real world to receive its inputs. Other neurons provide the real world with the network's outputs. All the rest of the neurons are hidden.

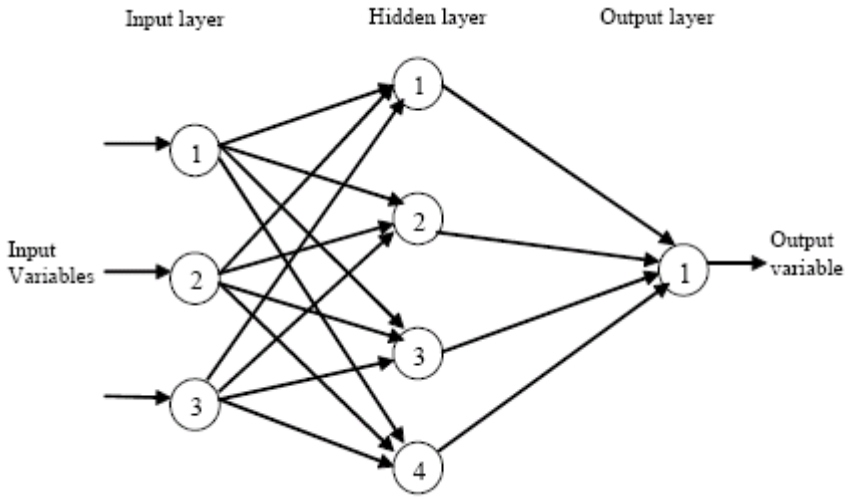


Figure 7-2: A Three-layer Feed Forward ANN

Most applications require networks that constitute at least the three normal types of layers i.e. input, hidden, and output. However there are useful networks which contain only one layer, or even one element. The layer of input neurons receives the data either from input files or directly from electronic sensors in real-time

applications. The output layer sends information directly to the outside world, to a secondary computer process, or to other devices such as a mechanical control system. Between these two layers can be many hidden layers. These internal layers contain many of the neurons in various interconnected structures. The inputs and outputs of each of these hidden neurons simply go to other neurons.

7.3.1 Weighting Factors

When a neuron receives simultaneous inputs, each input has its own relative weight which gives the input the impact that it needs on the processing element's summation function. This makes some inputs more important than others so that they have a greater effect on the processing element as they combine to produce a neural response. Weights are adaptive coefficients within the network that determine the intensity of the input signal as registered by the artificial neuron. The intensity can be modified subject to use of various training sets, network's specific topology or through its learning rules.

7.3.2 Summation Function

The first step in a processing element's operation is to compute the weighted sum of all of the inputs and the corresponding weights (vectors). The total input signal is the product of these two vectors.

$$NET_j = \sum_{i=0}^{n} w_{ji} x_i \qquad \textbf{(7-1)}$$

where x_i and w_{ji} are the input and weight terms respectively.

The input and weighting coefficients can be combined in many different ways before passing on to the transfer function. In addition to a simple product summing, the summation function can select the minimum, maximum, majority, product, or several normalizing algorithms. The specific algorithm for combining neural inputs is determined by the chosen network architecture and paradigm.

7.3.3 Transfer Function

The transfer function is the result of the summation function which is almost always the weighted sum that has been transformed to a working output through an algorithmic process. In the transfer function the summation total can be compared with some threshold to determine the neural output. For a sum value that is greater than the threshold value, the processing element generates a signal where as if the sum of the input and weight products is less than the threshold, no signal is generated. Both types of response are significant. The threshold, or transfer function, is generally non-linear. Linear functions are not very useful because they are limited due to the output that is simply proportional to the input.

7.3.4 Scaling and Limiting

Scaling simply multiplies a scale factor with the transfer value, and then adds an offset. Limiting ensures that the scaled result does not exceed an upper or lower

bound. This limiting is in addition to the hard limits that the original transfer function may have performed.

7.3.5 Output Function (Competition)

Each processing element outputs a signal to one or more other neurons which is normally directly equivalent to the transfer function's result. Some topologies may however modify the transfer result to incorporate competition among neighboring processing elements. Neurons compete with each other, inhibiting processing elements unless they have great strength. Competition can occur at one or both of two levels determining which artificial neuron will be active, or provides an output and secondly determining which processing element will participate in the learning or adaptation process.

7.3.6 Error Function and Back-Propagated Value

The calculated difference between the current and desired output is called raw error which is then transformed by the error function to match particular network architecture. The raw error may be used directly, squared while retaining its sign, cubed or modified by other paradigms to fit specific purposes. The artificial neuron's error term is typically propagated into the learning function of another processing element sometimes called the current error. The current error is typically propagated backwards to a previous layer. Yet, this back-propagated value can be either the current error, the current error scaled in some manner (often by the derivative of the transfer function), or some other desired output

depending on the network type. Normally, this back-propagated value, after being scaled by the learning function, is multiplied against each of the incoming connection weights to modify them before the next learning cycle.

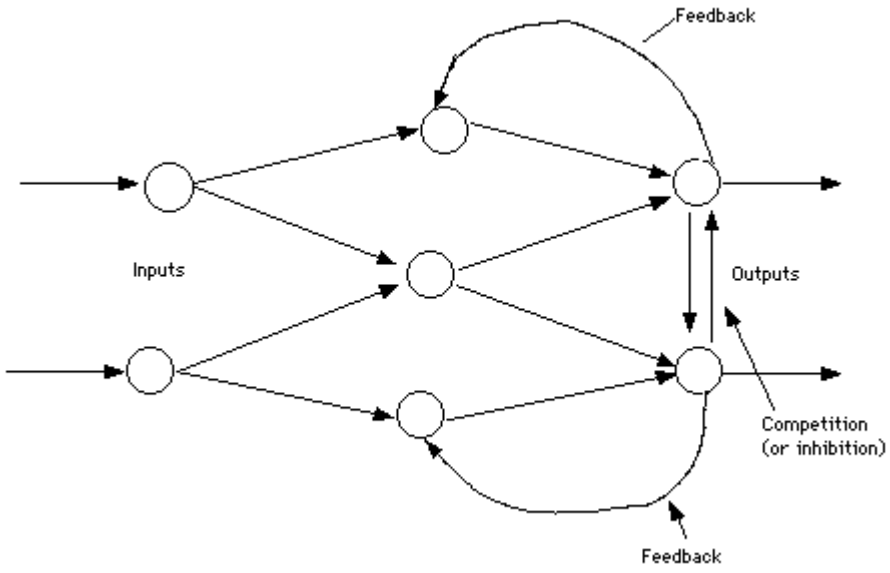


Figure 7-3: Simple Error Back-propagation ANN

7.3.7 Learning Function

The learning function modifies the variable connection weights on the inputs of each processing element according to some neural based algorithm to achieve some desired result, the process of which is called the adaption function or learning mode. There are two types of learning: supervised and unsupervised.

7.4 Training an Artificial Neural Network

Training of ANNs is accomplished by applying optimization algorithms that adjust the network weights and the neuron biases thereby reducing the error in the

network (De Vos and Rientjes, 2005). In essence, training determines the set of weights and thresholds such that for any input signal the ANN output is as close to the desired output as possible. To start this process, the initial weights are chosen randomly. Then, the training, or “learning”, is initiated. It is assumed that the neural network has no prior knowledge about the problem before being trained. When the network weights are changed, the data transfer through the ANN changes and the network performance changes. Multilayer feed-forward neural networks, like other non-linear estimation methods, have the disadvantage of either under-fitting (where too much hidden nodes fit the noise) or over-fitting (where there are insufficient hidden nodes thus failing to detect irregularities in the data set). Under-fitting produces excessive bias in the model outputs whereas over-fitting produces excessive variance. To avoid over-fitting and under-fitting, a stop training approach is used. The most popular stopping criterion involves a trade-off between training time and generalization error.

The general procedure involves splitting the available data into three parts:

i. a **training set**, used to determine the network weights ;
ii. a **cross validation set**, these are separate data sets used during the training process to estimate the network performance and decide when the training is to stop; and
iii. a **testing data** set, independent sets of data, not used in training or validation and are used to verify the effectiveness of the stopping criterion and to estimate the expected performance of the ANN.

During training, the output predicted by the network is compared with the actual (desired) output and the mean squared error (MSE) between the two is calculated. As more and more data are presented to the network, the results keep on improving until a suitable weight combination is found and the prediction error of the testing data is minimized. At this stage the ANN is considered trained.

There are two approaches to training: supervised and unsupervised. Supervised training, which is common in water resources applications, involves a mechanism of providing the network with the desired output either by manually grading the network's performance or by providing the desired outputs with the inputs. Unsupervised training is where the network has to make sense of the inputs without outside help. Unsupervised training is used to perform some initial characterization on inputs.

7.4.1 Learning Parameters

7.4.1.1 Learning Rate

The learning rate determines the absolute size of the weight change during learning and limits or expands the extent of weight adjustments in a training cycle. A high learning rate reacts quickly to input changes, and can make networks unstable if the rate is too high-the changes can be too extreme and cripple the network’s prediction ability. However, if the learning rate is too low, the network training time is substantially increased. A high learning rate is useful to

accelerate learning until the weight adjustments begin to plateau. However, the high learning rate increases the risk that the weight search jumps over a minimum error condition, which could jeopardize the integrity of the network and cause back-propagation learning to fail.

7.4.1.2 Momentum Factor

The momentum factor describes the proportion of the weight change that is added to each subsequent weight change. Low momentum causes weight oscillation and instability, preventing the network from learning. High momentum factor cripples network adaptability. For stable back-propagation, the momentum factor should be kept less than unity. Momentum factors close to unity are needed to smooth error oscillations when they occur. During the middle of training, when steep error slopes often occur, a small momentum factor is optimal, whereas towards the end of training a large momentum factor is desirable.

7.4.1.3 Training Tolerance

This is the margin of error permitted when training target values are compared with the values generated by the network during supervised training. A training tolerance factor of zero is the most desired since it indicates that the network values exactly match the target values. The higher the training tolerance factor is, the more inaccurate the neural network will be. ANNs generally use more parameters than conventional statistical methods and are therefore susceptible to

overtraining when too much data is presented to the network. The network is over-trained when the mean squared error increases as the network trains at predicting the test values. This indicates that the network's ability to recognize new patterns and generalize unknown data sets is hampered. The simplest method of correction to the overtraining phenomenon is to train the model with only part of the data and use the rest to check the network's performance.

7.5 Development of the Artificial Neural Network (ANN) Model

Development of an ANN model is done using the following framework:

i) Selection of data to be used for training and testing of the model.

ii) Selection of input-output variables. Decide what the neural network is to accomplish.

iii) Selection of the network architecture.

iv) Determination of the optimum number of neurons in the hidden layer,

v) Training of the ANN model.

vi) Testing of the model using selected performance evaluation statistics.

Data used with ANN analysis was the same data used in the previous methods of analysis. This included the complete dataset and the monthly datasets constituting extracted high flow data with corresponding seven days antecedent rainfall. The ANN model developed in this study was a Multilayer Perceptron (MLP) which is commonly used and also ideal for regression problems. The model consists of an input layer, two hidden layers and an output layer consisting

of one output neuron. For the Multilayer Perceptron, one hidden layer is usually sufficient to solve most problems. Good performance with respect to convergence and central processing time has been reported by Sahoo and Ray (2006) by using the hyperbolic tangent activation function; hence the choice for the hidden layers. The input data were fourteen (14) sets of rainfall data and a single output of flow. To avoid de-normalization of outputs, the output layer was provided with a linear activation function so that the output range was between -∞- and +∞. Twenty percent (20%) of the data was set aside for cross validation.

NeuroSolutions software automatically scales the data such that the training data lies within the range [-0.9-0.9]. The data is scaled to the range [- 1-1] to allow for values beyond the range for which the network was trained. The importance of scaling the data to this range (normalization) is to avoid one predictor (input) dominating others due to their different scale and units and cover different ranges.

To develop an 'optimum' network, the neurons (processing elements) in the hidden layer were increased from 2 to 14. The maximum number of fixed iterations (epochs) was set at 1500 during training. The optimum network is one which yields the lowest MSE on the training datasets. The training was carried out with a momentum factor of 0.7 and a learning rate of 0.1 in the hidden layer.

Performance criteria used to assess model adequacy was outlined in chapter 5 and will not be repeated here.

7.6 Results Analysis and Discussion

The most commonly used statistics in literature for performance evaluation of ANN models have been used here and the results are shown in Table 7-1. It is very important to note that results from analyses of data for individual months in Table 7-1 are not reliable as the software used could not produce good results due to being too low on training data. It is critical that training data used is adequate so as to develop a model that has high enough assurance. It must be pointed out here that due to the same data limitation mentioned above, it was not feasible to set out sample data sets for testing performance for the individual months.

Table 7-1: Performance Statistics of an ANN Model Developed Using Complete Dataset

Data/ Statistics	Complete Data		December	January	February	March
	Training	Testing	Training	Training	Training	Training
Coefficient of Determination, R^2	0.601	0.523	0.741	0.908	0.948	0.142
Mean Absolute Error (m^3/s), MAE	210.276	85.805	7.184	2.502	2.325	6.921
Root Mean Square Error (m^3/s), RMSE	2.637	2.171	1.294	0.894	11.810	1.684
Correlation Coefficient, r	0.775	0.723	0.861	0.953	-0.975	0.377
Efficiency Coefficient, E	0.601	0.523	0.741	0.908	0.948	0.136

Adequacy of a model for a particular application depends on the application for which the model is intended. Chiew and McMahon (1993) developed qualitative

guidelines for assessing the adequacy of stream flow estimates as shown in Table 7-2. It is worth noting that the bias statistic has been left out from this table.

Table 7-2: Qualitative Guidelines for Assessing the Adequacy of Stream-flow Estimates

Level of Accuracy	Performance Coefficients
Perfect	ce ≥ 0.93 or cd ≥ 0.97 or
Acceptable	ce ≥ 0.80 or cd ≥ 0.90 or
Generally Satisfactory	ce ≥ 0.60

Source: Chiew and McMahon (1993)

The ANN model produced results that are generally satisfactory for forecasting on the complete dataset, in line with the criteria above, but very poor results for some individual months. Performance on training data is better than that on testing data as expected. The correlation coefficient, a commonly used statistic provides information on the strength of a linear relationship between the observed and the simulated values. According to Imrie *et al.,* 2000, a correlation coefficient of less than 0.7 signifies an inadequate model and more than this acceptable. The correlation coefficient obtained is over 0.7 confirming that the model is acceptable for prediction.

Figure 7-4 and Figure 7-7 present flow time series of part of the simulation and actual observed datasets for the training and testing sets respectively. From the graphs it is evident that high flows are over estimated generally but mostly on the training data than on the testing data. The fit is not very close between the two data sets as observed also from the statistics but could be considered for forecasting with caution. The full dataset nevertheless yields better results than the high flow extracts' results from individual months. Figure 7-5 and Figure 7-8 provide scatter plots together with the line plot of model simulation versus the observed data for training and testing data respectively. The error plots of the same are shown in Figure 7-6 and Figure 7-9. It can be seen that maximum errors for training and testing data are 150 and 85 m^3/s respectively.

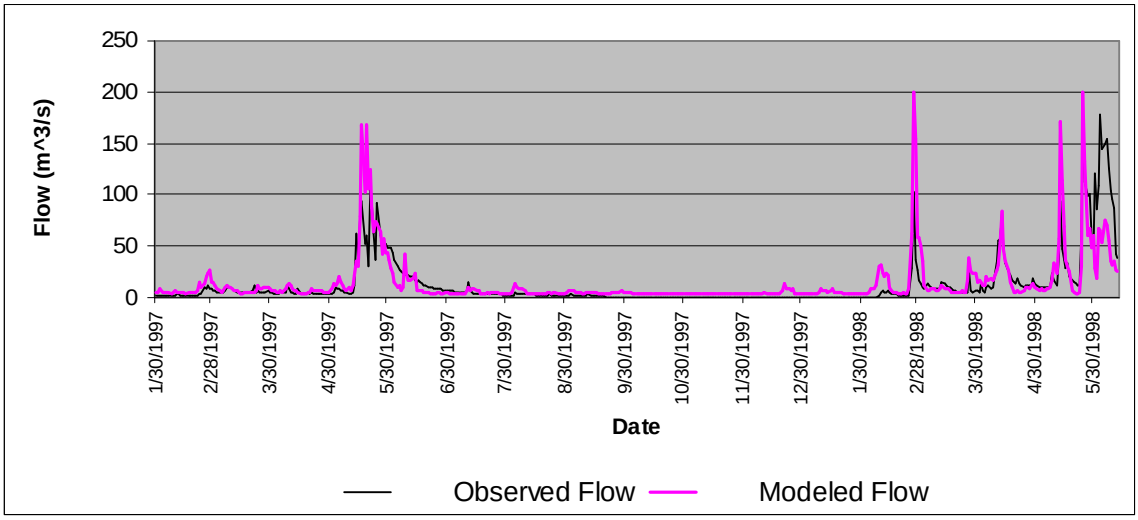


Figure 7-4: Comparison of Observed and Simulated Flows for ANN Training (10/8/1965 to 2/2/1967)

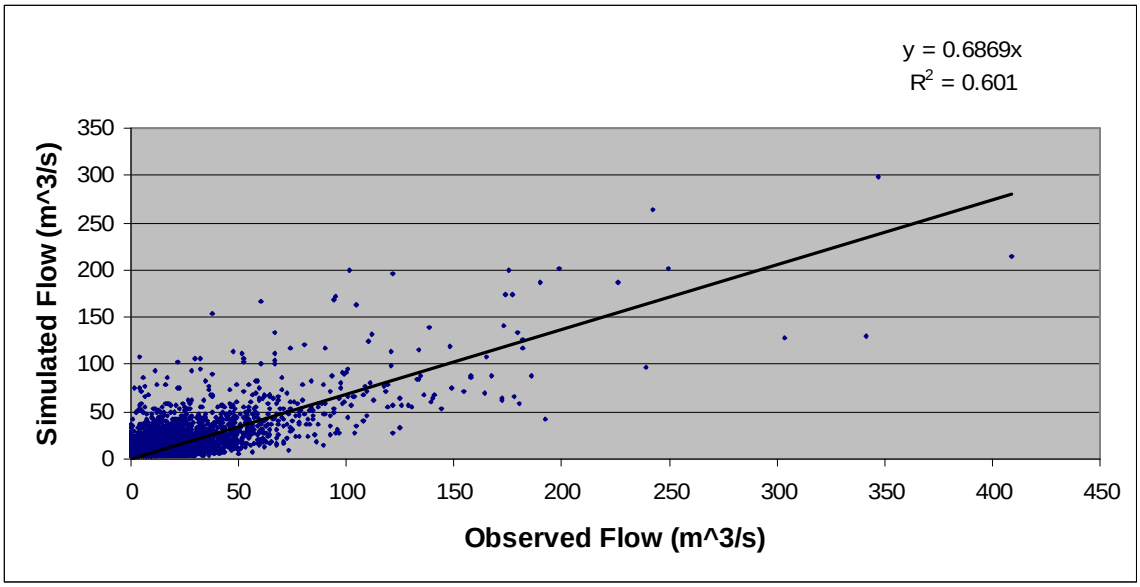


Figure 7-5: Scatter Plot of Observed Versus Simulated Flow for ANN Training

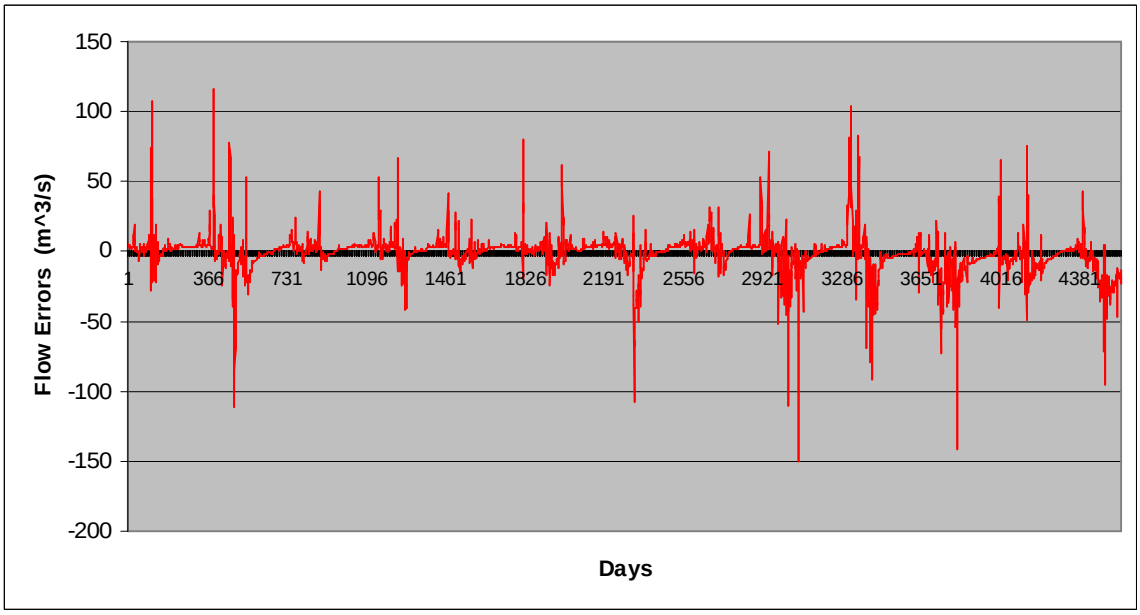


Figure 7-6: Streamflow Error Plot for ANN Training

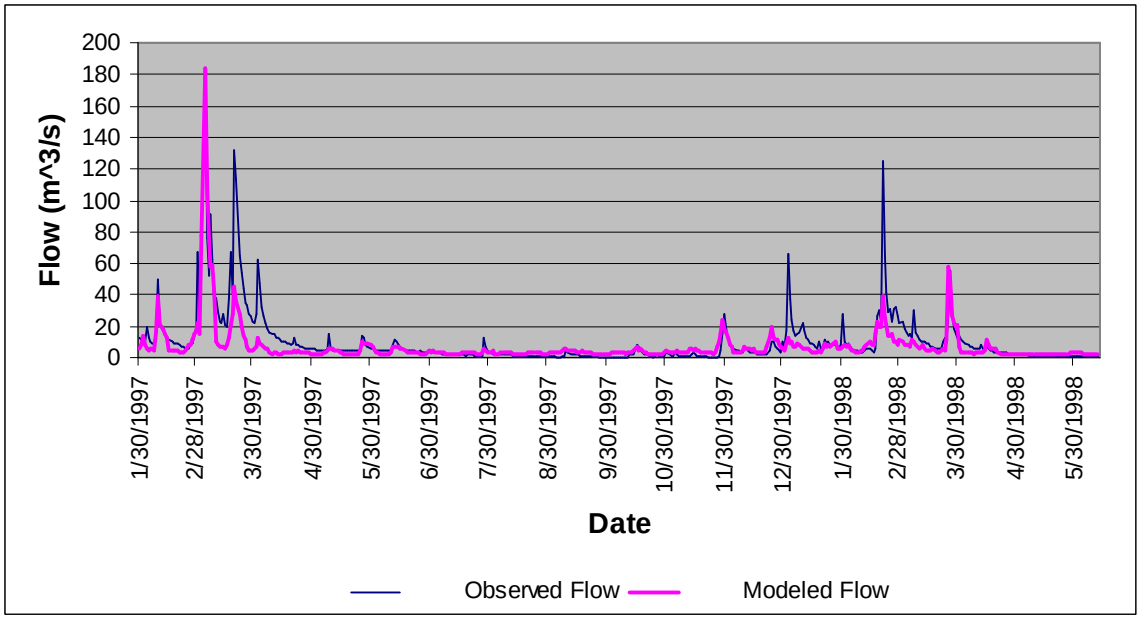


Figure 7-7: Comparison of Observed and Simulated Flows of the for ANN Testing (10/8/1965 to 2/2/1967)

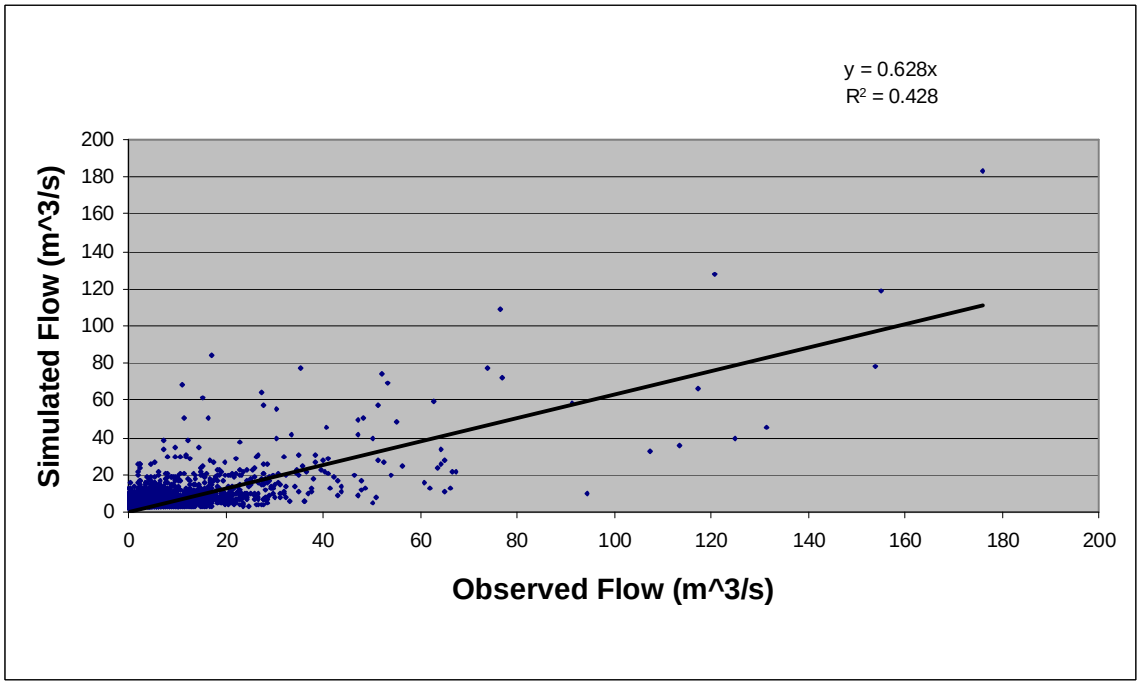


Figure 7-8: Scatter Plot of Observed Versus Simulated Flow for ANN Testing

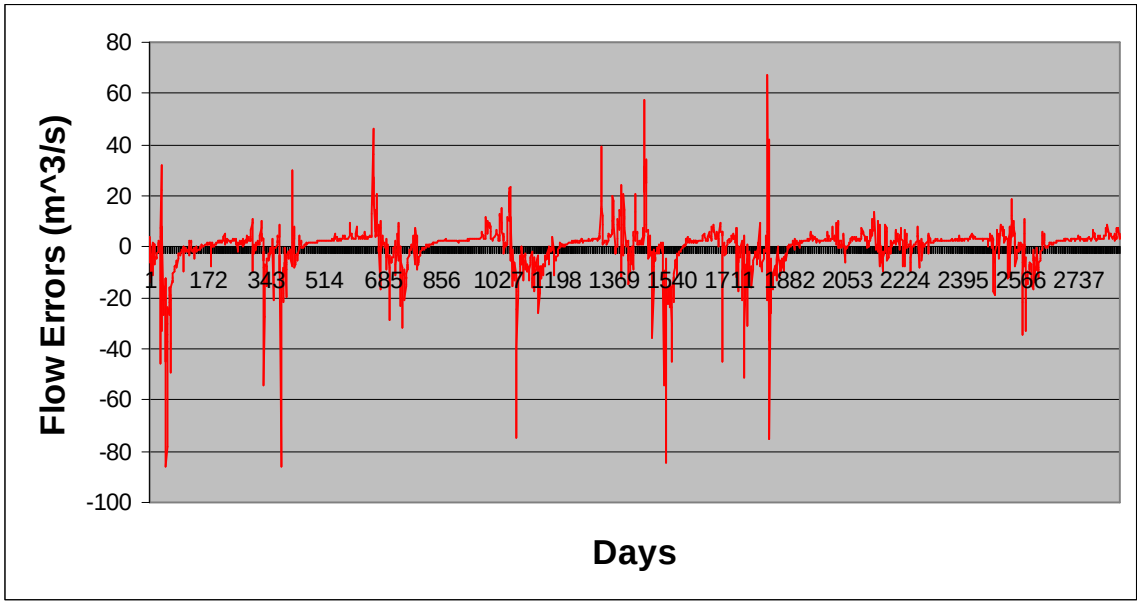


Figure 7-9: Streamflow Error Plot for ANN Testing

CHAPTER 8

8. COMPARISON OF THE THREE MODELS

This study has investigated and tried to establish an approach that would be used to predict high flows by use of easily obtainable data i.e. rainfall using the three models described in Chapters 5, 6 and 7. Consistency in the way the three models are tested and evaluated for performance is an important aspect in order for the models to be compared without bias. To achieve this objective the data used for calibration (training and cross validation for the ANN) has been the same for the three models. However, limitation to this requirement was experienced with respect to validation. Only for ANN was validation of the model performed. ANN software sets aside 20% of the whole dataset for validation. The method of evaluation for performance reporting has been identical for the three models.

Of the three models, the ANN yielded the best results with a coefficient of determination at 0.601 and a correlation coefficient of 0.775. Regression analysis using the seven antecedent days of rainfall followed with a coefficient of determination of 0.49 and a correlation coefficient of 0.70. The empirical model proved to be the least versatile with a coefficient of determination at 0.36 and a correlation coefficient of 0.60. The empirical model was expected to have performed better than the linear multi- regression model because it is a non

linear model and rainfall- runoff is best modeled using non linear models. It could probably perform better if another objective function had been applied.

It is clear from the evaluation of model performance criteria that only the ANN model produces generally satisfactory results with a *ce* of 0.601. An attempt is made here to arrive at a reasonable conclusion as to whether the model may be applicable for the intended use. Rainfall runoff is influenced by so many factors as outlined in chapter 4 (paragraph 1). These factors may only be roughly estimated and in line with Imrie *et al.,* 2000 who recommend that any model that produces a correlation coefficient above the threshold of 0.7 it is acceptable, it may be reasonable to adopt this model for the intended purpose. The ANN would however need to be subjected to more tests before adopting it for the task.

CHAPTER 9

9. CONCLUSIONS AND RECOMMENDATIONS

9.1 Conclusions

The objective of this study was to develop an approach for predicting impending flood events that would have happened on a particular day at an IFR site under natural un-impacted flow conditions to enable reservoir operators to make optimum reservoir releases that will satisfy both the ecological water requirements and the multiple reservoir water needs for other uses.

Three methods were used to develop models for prediction i.e. regression analysis, which is one of the most highly used methods for prediction, an empirical model, which is a nonlinear model known for modeling rainfall-runoff well and artificial neural networks (ANN), which are versatile systems that are known for their ability to model nonlinear systems.

The analyses carried out suggest that ANN is capable of providing a good model that would be used to predict flow using rainfall as input data. Application of the model on rainfall data resulted in a coefficient of determination of 0.601 with a correlation coefficient of 0.78. Regression analysis and the empirical model produced unsatisfactory results.

The importance of ecologically sustainable water management can never be over-emphasized. As human demands on the world's available freshwater supplies continue to grow, it is essential that the available limited water resources be carefully planned, managed and efficiently operated to meet the water needs of the present and future generations. To accomplish this objective, water managers need some form of decision support systems. Although results of this study have not resulted in the successful development of an ANN model with excellent predicting capability, this research has shown that Artificial Neural Networks (ANN) still have great potential of being used as a decision support system.

The ability to simulate river flows quickly and accurately is central in water resources management operations. Nonlinear rainfall-runoff models provide a basis for representing hydrological processes. This study has demonstrated that it is likely that, with further refinement, Artificial Neural Networks (ANN) could be used to provide reservoir operators with a tool for real time implementation of the high flow components of In-stream Flow Requirements (IFR).

9.2 Recommendations

The model presented in this study only sets down a foundation for further and thorough research on the enhancement of and applicability of ANN and empirical modelling to both model rainfall-runoff processes and the envisaged real time implementation of IFR in the Thukela River basin. The ANN model could be

investigated further using other training algorithms and different topologies to improve the results.

The ANN software used for this study was the educator version of Neurosolution which is subject to certain limitations. Upgrading to a more powerful version such as the Consultant or Professional version would provide more capabilities for modeling purposes and may result in improved performance. The predictive capability of using the real-coded genetic algorithm (RGA) and the self-organizing map (SOM) has been found to be superior to those trained using BPA (Srinivasulu and Jain, 2006). Other or a combination of training algorithms could therefore produce better results compared to the backpropagation learning algorithm used here.

Despite the fact that it was not intended to include as one of the objectives of this research to assess the practicality of the set in-stream flow requirements, it is worth noting that this study has found that the set in-stream flow requirements recommended in the ecological reserve implementation study are unachievable for IFR 3 in the Thukela river catchment. It is therefore imperative that a reassessment of these recommendations be done.

10. REFERENCES

ABBOTT, M.B. and REFSGAARD J.C. (Eds.) 1996 Distributed hydrological modelling. Kluwer Academic Pub.

ACREMAN M.C., FARQUHARSON F.A.K., MCCARTNEY M.P., SULLIVAN C., CAMPBELL K., HODGSON N., MORTON J., SMITH D., BIRLEY M., KNOTT D., LAZENBY J., WINGFIELD R. and BARBIER E.B. 2000. Managed Flood Releases from Reservoirs: Issues And Guidance. Report to DFID and the World Commission on Dams. Wallingford, UK, Centre for Ecology and Hydrology. 86 pp.

ARTHINGTON A.H., KING J.M., O'KEEFE J.H., BUNN S.E., DAY J.A., PUSEY B.J., BLUHDORN D.R. and THARME R. 1992. Development of an Holistic Approach for Assessing Environmental Flow Requirements of Riverine Ecosystems. In J.J. Pigram & B.P. Hooper eds. Proceedings of an International Seminar and Workshop on Water Allocation for the Environment. Armidale, USA, Centre for Water Policy Research, University of New England. pp. 69-76.

ARTHINGTON A.H. and PUSEY B.J. 1993. In-Stream Flow Management In Australia: Methods, Deficiencies & Future Directions. Australian Biologist, 6: 52-60.

ARTHINGTON A.H. and PUSEY B.J. 2003. Flow Restoration and Protection in Australian Rivers. River Research and Applications, 19: 377-395.

ARTHINGTON A.H. and ZALUCKI J.M. eds. 1998. Comparative Evaluation of Environmental Flow Assessment Techniques: Review of Methods. LWRRDC Occasional Paper Series 27/98. Canberra, Land and Water Resources Research and Development Corporation. 141 pp.

BARON, J. S., N. L. POFF, P. L. ANGERMEIER, C. N. DAHM, P. H. GLEICK, N. G. HAIRSTON, R. B. JACKSON, C. A. JOHNSTON, B. D. RICHTER, and A. D. STEINMAN. 2002. Meeting ecological and societal needs for freshwater. Ecological Applications 12:1247–1260.

BELL, V.A. and MOORE, R.J. 1998. “A grid-based distributed flood forecasting model for use with weather radar data: Part 1. Formulation”. Hydrol. Earth System Sci. 2(2-3), 265-281.

BEVEN, K.J., 1985, Chapter 13, Distributed Models. In: Anderson, M.G., Burt, T.P. (Eds.), Hydrologic Forecasting, Wiley, Mew York.

BEVEN, K.J. 2001, Rainfall-runoff modeling – The primer, Wiley: Chichester.

BOVEE K.D. 1982. A Guide to Stream Habitat Analysis Using the In-Stream Flow Incremental Methodology. In-stream Flow Information Paper 12. Fort Collins, US Department of Fisheries and Wildlife Service. 248 pp.

BRAGG, O. M., and A. R. BLACK. 1999. Anthropogenic Impacts on the Hydrology of Rivers and Lochs. Stage 1 Report: Literature Review and Proposed Methods. University of Dundee, Dundee, UK.

BRATH, A., MONTARANI, A. and TOTH, E., 2004, Analysis of the Effects Of Different Scenarios of Historical Data Availability on The Calibration Of a Spatially-Distributed Hydrological Model. Journal of Hydrology 291: 232-253.

BROWN, C. and KING, J. Environmental Flow Assessments for Rivers, 2000 . A summary of the DRIFT process. Southern Waters information Report No. 01/00.

CHIEW, F.H.S. and McMAHON, T.A. 1993 Assessing the quality of catchment streamflow yield estimates, Australian Journal of Research.

COMMONWEALTH OF AUSTRALIA 1996. The National Strategy for the Conservation of Australia's Biological Diversity. Department of the Environment, Sport and Territories: Canberra.

COTTINGHAM P., THOMS M.C. and QUINN G.P. 2002. Scientific Panels and their Use in Environmental Flow Assessment in Australia. Australian Journal of Water Resources, 5: 103-111.

DE VOS, N.J. and RIENTJES T.H.M. 2005 Constraints of artificial neural networks for rainfall-runoff modeling: trade-offs in hydrological state representation and model evaluation, Hydrology and Earth System Sciences, vol. 9 pp 313 -321.

DUAN, Q., V. K. GUPTA, and SOROOSHIAN. S, 1993, A shuffled Complex Evolution Approach for Effective and Efficient Global Minimization, J. Optim. Theory Appl.

DUAN, Q., SOROOSHIAN, S. and GUPTA, V., 1992, Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models. Water Resources Research 38(4): 1015-1031.

DUAN, Q., SOROOSHIAN, S. and GUPTA, V., 1994, Optimal Use Of The SCE-UA Global Optimization Method For Calibrating Watershed Models. Journal of Hydrology 158: 265- 284.

ECKHARDT, K. and ARNOLD, J. G., 2001, Automatic Calibration Of A Distributed Catchment Model. Journal of Hydrology 251: 103-109.

FLEMING, G. 1975, Computer Simulation techniques in hydrology. Elsevier, New York.

HOLLING, C. S., and G. K. MEFFE, 1996. Command and Control and the Pathology of Natural Resource Management. Conservation Biology 10:328–337.

HUGHES, DA, O'KEEFFE, J, SMAKHTIN, V and KING, J 1997 Development Of An Operating Rule Model To Simulate Time Series Of Reservoir Releases For In-stream Flow Requirements. Water SA, 23(1), 21-30.

HUGHES, D.A.and SHANGE, S., 2004, Thukela Water Project Decision Support Phase: Reserve Determination Model, Appendix A, Vol. 1, pp A2- A12.

HUGHES, DA, 2006, Development, Testing And Installation Of A Real-Time Ecological Reserve Implementation Method For The Thukela River: Deliverable F: Assessment of flood release capabilities and requirements and review of existing methods, Water Research Commission Project: K5/1582.

IMRIE, C.E., DURUCAN, S. and KORRE, A. 2000 River flow prediction using artificial neural networks: generalisation beyond the calibration range, Journal of Hydrology, vol. 233, pp 138 – 153.

IUCN (INTERNATIONAL UNION FOR THE CONSERVATION OF NATURE). 2000. Vision For Water And Nature: A World Strategy For Conservation And Sustainable Management Of Water Resources In The 21st Century. International Union for the Conservation of Nature, Gland, Switzerland and Cambridge, UK.

JAIN, S.K. and SINGH, V.P. 2003 Water Resources Systems Planning and Management, Elsevier, Amsterdam, pp 858.

JAIN, S.K., DAS, A. and SRINIVASARAN, D.K. 1999 Application of ANN for Reservoir inflow Prediction and Operation, Journal of Water Resources Planning and Management, 5, pp 263-271.

JOWETT I.G., 1997, In-stream Flow Methods: A Comparison of Approaches. Regulated Rivers: Research & Management, 13: 115-127.

KING, J., BROWN, C and SABET, H, 2000. A scenario-based holistic approach to environmental flow assessments for rivers. Rivers Research and Application.

KING J.M. and LOUW D. 1998. Instream Flow Assessments for Regulated Rivers in South Africa Using the Building Block Methodology. Aquatic Ecosystem Health and Restoration, 1: 109- 124.

KING J.M. and THARME R.E. 1994. Assessment of the Instream Flow Incremental Methodology (IFIM) and Initial Development of Alternative Instream Flow Methodologies for South Africa. Water Research Commission, Report No. 295/1/94. Pretoria, SA. 590 pp.

KING J.M., THARME R.E. and DE VILLIERS M.S. eds. 2002. Environmental Flow Assessments for Rivers: Manual for the Building Block Methodology. Water Research Commission Technology Transfer Report No. TT131/00. Pretoria, SA, Water Research Commission. 340 pp.

KING, J., BROWN, C and SABET, H. 2003, A scenario-based holistic approach to environmental flow assessments for rivers. Rivers Research and Application. (In press).

KRZYSZTOFOWICZ, R. 2001: The case for probabilistic forecasting in hydrology, Journal of Hydrology, 249, 2-9.

LEGATES, D. R. and MCCABE, G. J., 1999. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydro-climatic model validation. Water Resource. Res. 35(1), 233- 241.

LOAGUE, K and FREEZE,R.A. 1985. A comparison of rainfall- runoff modeling techniques on small upland catchments. Water Resources. Res., 21-248.

MADSEN, H., 2000, Automatic Calibration Of A Conceptual Rainfall-Runoff Model Using Multiple Objectives. Journal of Hydrology 235: 276-288.

MADSEN, H., 2003, Parameter Estimation in Distributed Hydrological Catchment Modelling Using Automatic Calibration with Multiple Objectives. Advances in Water Resources 26(2): 205-216.

NASH, J.E. and SUTCLIFFE, J.V. 1970, River flow forecasting through conceptual models; Part 1- a discussion of principles, Journal of Hydrology, vol. 10, pp 282 -290.

NELDER, J. A. and MEAD, R., 1965, A Simplex Method for Function Minimization. Computer Journal 7(4): 308-313.

POFF N.L., ALLAN J.D., BAIN M.B., KARR J.R., PRESTEGAARD K.L., RICHTER B.D., SPARKS R.E. and STROMBERG J.C. 1997. The Natural Flow Regime, A Paradigm For River Conservation And Restoration. BioScience, 47: 769-784.

POSTEL, S. and CARPENTER, S., 1997. Freshwater Ecosystem Services. Pages 195–214 In G. C. Daily, Editor. Nature’s services: societal dependence on natural ecosystems. Island Press, Washington, D.C., USA.

PRINGLE, C. M., FREEMAN, M. C. and FREEMAN, B. J. 2000, Regional effects of hydrologic alterations on riverine macrobiota in the New World: tropical–temperate comparisons, BioScience 50:807–823.

RAILSBACK, S. 2001. In-stream Flow Assessment Methods: Guidance for Evaluating In-Stream Flow Needs in Hydropower Licensing. Electric Power Research Institute, Palo Alto, California, USA.

REFSGAARD J. C., 1996, Terminology, modeling Protocol and classification of hydrological model codes, in Distributed Hydrological Modeling edited by Michael B. Abbott and J. C. Refsgaard, Water Science and technology Library.

RICHTER, B. D., BRAUN, D. P., MENDELSON, M. A. and MASTER, L. L., 1997a. Threats to imperiled freshwater fauna. Conservation Biology 11:1081–1093.

RICHTER, B. D., BAUMGARTNER, J. V. WIGINGTON, R. and BRAUN, D. P., 1997b. How Much Water Does A River Need? Freshwater Biology 37:231–249.

RICHTER B. D, MATHEWS R, HARRISON D.L and WIGINGTON R. 2003a. Ecologically sustainable water management: managing river flows for ecological integrity. Ecol Appl. 13: 206–24.

RICHTER BD, MEYER JL, LUTZ K, and WARNER A. 2003b. Specifying water flow requirements to support river health. Proceedings of the 2003 Georgia Water Resources Conference; 23–24 April 2003; University of Georgia, Athens, GA. Athens, GA: University of Georgia Institute of Ecology. p 457–59.

SAHOO, G.B. and RAY, C. (2006) Comment on “Flow forecasting for a Hawaii stream using rating curves and neural networks”. [Journal of Hydrology 317 63–80].

SCUDDER, T. 1980. River basin development and local initiatives in savannah environments. In: Harris, D.R., (Ed) Human ecology of savannah environments. Academic Press.

SINGH, V.P., 1995, Computer models of watershed hydrology, Water Resources Publications, 1130pp.

SINGH, V. P. and FREVERT, D. K, 2002a, Mathematical models of large watershed hydrology, Water Resources Publications, 891pp.

SINGH, V.P. and FREVERT, D. K. 2002b, Mathematical models of large watershed hydrology and applications, Water Resources Publications, 950pp.

SMITH, M. B., SEO, D., KOREN, V. I., REED, S. M., ZHANG, Z., DUAN, Q., MOREDA, F. and CONG, S. 2004, The distributed model intercomparison project (DMIP): motivation and experiment design. J. Hydrol. (Amst.), 298, 4-26 ER.

SPARKS, R. E., 1995. Need For Ecosystem Management Of Large Rivers And Their Floodplains. BioScience 45:169–182.

SRINIVASULU, S. and JAIN, A. 2006, A comparative analysis of training methods for artificial neural network rainfall-runoff modeling, J. Applied Soft Computing, 6(3), 295-306.

STALNAKER C., LAMB B.L., HENRIKSEN J., BOVEE K. and BARTHOLOW J. 1994. The In-stream Incremental Methodology. A primer for IFIM. Biological Report, 29. Washington, DC, U.S. Department of the Interior, National Biological Service.

STANFORD, J. A., J. V. WARD, W. J. LISS, C. A. FRISSELL, R. N. WILLIAMS, J. A. LICHATOWICH, and C. C. COUTANT. 1996. A general protocol for restoration of regulated rivers. Regulated Rivers 12:391–414.

STEIN, B. A., L. S. KUTNER, and J. S. ADAMS. 2000. Precious heritage: the status of biodiversity in the United States. Oxford University Press, New York, New York, USA.

SWALES S. and HARRIS J.H. 1995. The Expert Panel Assessment Method EPAM: a new tool for determining environmental flows in regulated rivers. In D.M. Harper & A.J.D. Ferguson, eds., The Ecological Basis for River Management. Chichester, UK, John Wile & Sons. pp. 125- 134.

THARME, R E. 1996. Review of International Methodologies for the Quantification of the In-stream Flow Requirements of Rivers. Draft report. Pretoria: Water Research Commission.

THARME R.E. 2003. A Global Perspective on Environmental Flow Assessment: Emerging Trends in the Development and Application of Environmental Flow Methodologies for Rivers. River Research and Applications, 19: 397-442.

TSYKIN, E.N. 1985. Multiple Nonlinear Statistical Models for Rainfall Simulation and Prediction. Journal of Hydrology, 77: 209-226.

WALKER, K. F., SHELDON, F., and PUCKRIDGE. J. T., 1995. A Perspective on Dry land River Ecosystems. Regulated Rivers 11: 85–104.

WCD (World Commission on Dams), 2000, Dams and development: a new framework for decision-making. Earthscan, London, UK.

YAPO, P. O., GUPTA, H. V. and SOROOSHIAN, S., 1998, Multi-Objective Global Optimization for Hydrologic Models. Journal of Hydrology 204: 83-97.

YEH, W., 1985, Reservoir Management and Operations Models: A State-Of-The-Art Review. Water Resources Research 21(12): 1797-1818.

YEVJEVICH, V. 1974, Systematization of flood control measures, Journal Hydraulic Division, ASCE, HY 11, November, 1974.

ZEALAND, C.M., BURN, D.H. and SIMINOVIC, S.P. 1999 Short term streamflow forecasting using artificial neural networks, Journal of Hydrology, vol., 214, pp 32 – 48.