FUZZY AND CONCEPTUAL-FUZZY MODELLING OF COMPLEX RIVER SYSTEMS WITH SCARCE DATA: CASE OF LETABA RIVER

Zacharia Katambara

A thesis submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfillment of the requirements for the degree of Doctor of Philosophy.

Johannesburg, 2011.

DECLARATION

I declare that this thesis is my own unaided work. It is being submitted to the Degree of Doctor of Philosophy to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination to any other University.

…………………………………………………………………………………………………….

………………………...day……………………………...year…………………..………………

ABSTRACT

The investment in water infrastructure on a number of river systems in South Africa and many other regions of the world so as to meet the ever growing demands for water over the last few decades, has not been matched by the implementation of adequate hydrometric data collection and water-use accountability practices. This has resulted in complex rivers systems with scarce data. A typical example in South Africa is the Letaba River system.

The main objective of this research was to investigate the applicability of fuzzy inference based and hybrid fuzzy inference-conceptual modelling approaches to highly developed and complex river systems with scarce data using Letaba River as a case study. For completeness, a standalone conceptual model was included and three models were therefore studied; a fuzzy inference, a hybrid fuzzy inference-conceptual, and a standalone conceptual model.

The evaluation of the modelling showed that:

- The models simulate better the flows at those locations of the river system that were impacted less by human activities.
- The fuzzy inference model was found to be a black box although it obtained the best statistical performance in modelling flow in those locations highly impacted by human activities.
- The conceptual model reproduced the main natural catchment and water resource development processes and systems reasonably well.
- The hybrid fuzzy-conceptual model performed comparably to the fuzzy model and also represented the catchment and water resource development processes in a manner comparable to that of the conceptual model. This suggests that the hybrid may be the better model to apply in situations where simulation accuracy and adequate representation of the catchment processes and water resource development system are required.

The study recommends that:

- Further studies on the use of the hybrid fuzzy inference-conceptual modelling approach need to be undertaken with the aim of improving both statistical simulation performance and system representation in the reality of scarce data.
- Deliberate initiatives need to be undertaken to improve collection and management of hydrometric and water use data in the Letaba River system and other data-scarce systems.

DEDICATION

In the memory of my lovely parents Saimon Katambala Mhamilawa and Atwimilye Msabaha Semutonyole who were called by God at different times during this study period.

My wife Osmunda Oigen Mwanyika, our family doctor whose love is the cord linking me and our lovely kids Emmanuel Luhuvilo and Ezra.

The family of Prof. Burton L. M. Mwamila who saw me not in the way I was, but what I could be in life and thereafter made it happen.

Glory and honour to the Almighty God, the provider.

ACKNOWLEDGMENTS

The journey that has led to this achievement has been a long one, wrought with several challenges which if it were not for the divine hand of the Almighty God, as well as input from several great people, I would most certainly have found it impossible to proceed.

Among these great people is Dr. John G. Ndiritu, my supervisor and a godly man who first of all accepted me as his student. Further, the encouragement, support and guidance he offered to me was significantly characterised by patience that has made me strongly believe that, for every existing approach a more suitable one may exist if more effort, proper knowledge and skills are applied towards its improvement. All these have culminated into this report. I am very grateful and I consider myself a lucky person to have been supervised by him.

I would like to thank the government of Tanzania through Mbeya Institute of Science and Technology, my employer, for the opportunity to pursue this study and the support received is highly appreciated.

I am very grateful to the members of staff in the School of Civil and Environmental Engineering, at the University of the Witwatersrand for the opportunity to study. I am also thankful to the library staff for their support, the security officers for maintaining peace on campus, the residence life for the shelter, Café Dulcé for the coffee, the bus service provider for making the spins safe and enjoyable and other service providers for bringing their services within a working distance.

The National Research Foundation, the Department of Water Affairs of South Africa, South African Weather Services, Letaba Water User Association, LQ Consulting, Department of Home Affairs of South Africa and Kruger National Park are acknowledged for the various ways in which they supported this study.

To exceptional individuals, many thanks to Paulo Kagoda a friend by nature and a brother by character, Mthokozisi Ncube a friend who prefers to offer help even when he is in a difficult situation, Dr. A. Ilemobade a very kind person who is ready to assist even without an appointment, Prof A. Taigbenu a generous man whose advise provided direction towards my excelling, Dr. K. Li a man who advises one to be specific and maintain standards, Dr. Catherine Namono a friend exemplified by deeds, Mr. Jean Boroto a man who encourages one not to panic but to remain focussed, Jeanmarie Kileshye a friend who encourages one to just keep on going, Jeanmarc Kahinda a friend who inspires with confidence, James Adewumi a friend who shares, Olanrewaju a friend who encourages hardworking, Valentine Katte a friend who encourages one to trust God in everything you do, Hammony Musiyaria a friend who encourages patience, Bolanle Ikotune a friend who believes nothing is impossible and Mrs D. Simelane a friend whose advice is laced with humour. While it is not possible to mention all of them, I appreciate all those who made this happen. God bless you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table **page**

LIST OF PUBLICATIONS

Katambara, Z. and Ndiritu, J. (2009) A fuzzy inference system for modelling streamflow: Case of Letaba River, South Africa, Physics and Chemistry of the Earth, Parts A/B/C, 34(10-12), pp. 688-700.

Katambara, Z. and Ndiritu, J. (2010) A hybrid conceptual-fuzzy inference streamflow modelling for the Letaba River system in South Africa, Physics and Chemistry of the Earth, Parts A/B/C, 35(13-14): 582-595.

1 INTRODUCTION

1.1 BACKGROUND

Many river systems have hydraulic structures built across them that serve various purposes. Some of these structures are for monitoring the flows for management purposes, some are for storing water for the purpose of managing the demands and some are for storing flood flows or for recreation (Labadie, 2004). Typically, the water demands include supply to meet irrigation requirements, municipal/domestic supply, industrial supply and ecological requirements and these demands are growing (McKenzie and Craig, 2001). However, the construction of large facilities to store more water has declined due to various reasons. In Australia, Cui and Kuczera (2003) report this to be the result of inadequate funds while in South Africa there are only a few suitable sites available to accommodate these larger storage facilities (Ndiritu, 2005). Therefore, in these situations, a deliberate effort towards the operation of the existing systems in a manner that seeks to maximise resource utilization is paramount.

In meeting demands, use is made of streamflow in rivers as this is the most common and easiest mode of conveyance of water released from a reservoir (Katambara and Ndiritu, 2009). In a situation where the flows in the river are maintained by releases made from an upstream dam during dry periods, the flowing water has to meet demands as it flows downstream and also meet the minimum flows requirements necessary to sustain the ecosystem. In such circumstances, streamflow modelling that is capable of informing how the releases, the abstractions and meteorological conditions (rainfall and evaporation) impact the flow is a valuable aid to the operation of such river systems.

In South Africa, catchment water is increasingly becoming limited (Dye and Croke, 2003) and the demands for water are growing. Yet the infrastructure (e.g. dams, storage weirs, agricultural facilities) in these catchments is highly developed to the extent that major expansion is infeasible. The operation of such river systems is characterised by inadequate recording of water used and sparse meteorhydrological monitoring. As a result, the catchments are complex and have scarce data. Using existing models to model such systems is hampered by inadequate data and the complexity of the river systems. For this matter, therefore, new modelling approaches that may be capable of handling complex systems in a manner that incorporates uncertainties and at the same time attempt to improve the understanding on the poorly understood processes are preferable.

Uncertainties resulting from catchment modelling is a combination of the structural uncertainty and natural variability (Loucks et al., 2005). The structural uncertainty emanates from the model setup, while parameter uncertainty results from not only estimation techniques but also differences in scale. The natural variability is concerned with the spatial and temporal variability of the input values. Traditional uncertainty analysis utilizes probability theory that requires distributional assumptions concerning random variables to be generated (Katambara and Ndiritu, 2007). For instance, Kagoda and Ndiritu (2008) applied the Bayesian inference in modelling extreme rainfall in South Africa in a study that incorporated uncertainty in the analysis. The estimates obtained from Bayesian method were higher than those obtained from Regional Storm Index Method (RSIM) for higher return periods. Lloyd and Atkinson (2001) applied ordinary kriging and indicator kriging in assessing uncertainty in estimates of elevation. It was observed that indicator kriging with locally adaptive indicator threshold provided a more accurate guide to uncertainty on local basis than ordinary kriging. Elfeki (2006) coupled Markov chain, a stochastic technique with numerical groundwater flow and transport models applied in the Central Rhine Meuse Delta in the Netherlands. The probability distributions used in the coupled Markov chain where generated from the available information. Hence precision of the model depends on the amount of the data available to generate the distributions. The drawbacks with these techniques are the amount of data required to estimate the distribution and computational difficulties arising from multiple convolutions in the usual case of dealing with several non-normal dependent random variables (Bardossy et al., 1995; Tayfur et al., 2003).

Modelling approaches based on fuzzy logic and their extended applications in hybrid modelling are gaining momentum with respect to modelling systems with inadequate data. Such applications include fuzzy multi-objective optimization in construction projects (Afshar and Fathi, 2009), ranking multi-criterion river basin planning (Raja and Kumar, 1998), sustainable rangeland management (Azadi et al., 2007) and assessment of the population risk of flood disaster (Jiang et al., 2008). Considering that fuzzy inference based approaches have been noted to be highly promising (Beven, 2004), particularly with regard to their ability to deal with inadequate data while inferring catchment behaviour, it is appealing to consider their application in modelling complex systems that are data scarce as evidenced by limited information on how the complex systems are being operated.

In recent years fuzzy inference system has been introduced and applied in the water related studies. Enthusiasm for the approach has increased due to its ability and potential to deal with scarce and vague data. As such, the poorly understood processes can be modelled while incorporating uncertainties resulting from the data, model structure and parameters in a manner that involves less intensive computations than traditional modelling techniques (Lohani et al., 2006). Considering the fact that the approach can be linked to a conceptual or physically based model to develop hybrid models, it is reasonable to suggest that its use in modelling complex river systems with scarce data has not been exhaustively explored. The coupling of fuzzy logic model to conceptual or physically based model provides a basis for spearheading the application of fuzzy logic and hybrid models in modelling more complex real systems since the approach has been shown to be capable of dealing with complex system with inadequate data and can be said to have the potential for use in developing future modelling procedures (Lauzon and Lence, 2008, Lohani et al., 2006).

In their review, Jacquin and Shamseldin (2009) indicated that fuzzy inference systems can be used as effective tools for river flow forecasting, even though their application is rather limited in comparison to the popularity of neural networks models. However, it was found that there are several unresolved issues requiring further attention before more clear guidelines for the application of fuzzy inference systems can be given. These involve the selection of suitable input variables, selection of the appropriate number of rules, rule construction, removal of unnecessary rules and appropriate model calibration approach that allows retaining of the interpretation of the fuzzy rules with a hydrological basis. Therefore, the use of fuzzy inference approach in this study is considered as an attempt to contribute to the unresolved issues.

In South Africa just like in other regions in the world, there are many catchments that have multiple water demands and highly developed water utilization infrastructure. Often, water utilization in these catchments is not monitored closely and hydrological data measurements are mostly inadequate. As a consequence, modelling such catchments for improving resource utilization is complex and inherently imprecise. This study uses the Letaba River system; a system developed to the extent that there is no room for expansion (DWAF, 2004), as a case study to test the applicability of fuzzy and hybrid conceptual-fuzzy modelling to complex data scarce river systems. For completeness, a stand-alone conceptual modelling is also applied to the same system. Within the system, the Tzaneen Dam is the largest and the most downstream dam and supplies municipal demand to the town of Tzaneen and neighbouring communities. The releases from Tzaneen Dam are also meant to supplement the flows in the Letaba River during periods of low flows to meet irrigation demands and environmental water requirements further downstream in the Kruger National Park, a habitat of sensitive ecological species. The imprecise rule of thumb applied in determining water releases from storage weirs in the river and poorly recorded water abstractions by canals and pumping complicate the modelling of the system. Furthermore, the catchment processes resulting from the physical characteristics of this complex system that has among other features an alluvial aquifer, are themselves complex.

1.2 PROBLEM STATEMENT

The demand for water in the Letaba River system is higher than the amount that the system can comfortably supply. Moreover, the infrastructure for water utilisation is highly developed while the accounting for the water uses and the scale of meteorhydrological monitoring is inadequate and not proportional to the extent at which the infrastructure is developed. These factors and others result in a complex river system that is data scarce. Modelling a complex system may produce simulations that are uncertain and as such shed little light on the poorly understood catchment processes. For this reason, efficient ways of managing the resource are highly imperative. If limited effort is applied to improve operation, how will the sustainability of resource use and the environment be achieved? This study seeks to find out how well fuzzy inference and hybrid fuzzy inference-conceptual modelling approaches can model complex data-scarce river systems.

1.3 OBJECTIVES

1.3.1 Main objective

The main objective of the study is to find out the applicability of fuzzy inference in modelling hydrological systems and its extended applications in hybrid modelling for highly developed and complex river systems with scarce data that are common in South Africa and other regions in the world.

1.3.2 Sub-objectives

- To develop a fuzzy inference model, apply it to model the Letaba River system and evaluate its performance.
- For comparison purposes, develop a conceptual model, apply it to model the Letaba River system and evaluate its performance.
- Develop a hybrid conceptual-fuzzy inference model, apply it to model the Letaba River system and evaluate its performance.
- To compare the capabilities of the three models in modelling Letaba River and make inferences into their ability to model other complex hydrological systems with scarce data.

1.4 LAYOUT OF THE THESIS

This thesis is organised into ten chapters as follows:

Introduction: This part of the thesis provides an overview introduction, statement of the problem and objectives.

Catchment modelling: This part of the thesis reviews catchment modelling in general. Descriptions of five models subjectively selected based on their data/information requirements and their capability in modelling complex systems is given.

Fuzzy inference and hybrid conceptual-fuzzy inference: This part of the thesis reviews the fuzzy inference and describes their extended application in hybrid modelling. Several studies that have applied this approach are discussed in this chapter.

Letaba River system: This chapter describes the Letaba River system; a typical example of a complex river system. The characteristics of the system, the operation, water allocation and the data available are also described.

Fuzzy inference modelling: This chapter describes fuzzy theory and discusses the different types of fuzzy inference systems including rule based inference and the clustering based inference. The different types of clustering algorithms, their advantages and disadvantages are also discussed.

Application of fuzzy inference modelling to the Letaba river system: This chapter is about the application of the fuzzy inference in modelling the Letaba river system. Preliminary attempts to improve the modelling are presented and the model performance is discussed.

Conceptual modelling of the Letaba river system: This chapter describes the conceptual modelling approach, the development of a conceptual model and its application in the Letaba river system. A discussion on the results including the models' representation of the real system is done. While the complexity is limited by the available data, the model's capabilities and limitations are also discussed.

Hybrid modelling of the Letaba river system: This part of the thesis is about the hybrid conceptual-fuzzy modelling approach, its development and application in modelling the Letaba river system. A discussion on the model's suitability and its performance is done.

Comparison of fuzzy inference, conceptual and hybrid models: This particular chapter compares the fuzzy inference, conceptual and hybrid models.

Conclusions and recommendations: This particular part of the thesis summarises the study with regard to the set objectives and makes conclusions and recommendations for further work.

2 CATCHMENT MODELLING OF COMPLEX RIVER SYSTEMS CHARACTERIZED BY DATA SCARCITY

2.1 INTRODUCTION

Hydrological models can be classified into three main categories namely black box, conceptual and physically-based (Chen and Adams, 2006, Ndiritu and Daniell, 1999, Rajurkar et al., 2004) regardless of their structural diversity. Black box models, as the name suggests are data driven models that do not include any perceived understanding of the processes being modelled. The main principle behind these models is the linear or non-linear mapping of input datasets to the expected output dataset. Chen and Adams (2006) termed these models universal approximators. Black box models are preferred in situations where there is inadequate data and the processes to be modelled are poorly understood. Examples of black box models are those based on the heuristic approaches which include artificial neural networks (ANNs) and fuzzy logic. A detailed description of fuzzy logic-based models has been presented in Chapter 5 and Chapter 6.

Conceptual models represent the perceived and known catchment processes as storages that are linked to each other in a simplified manner (Ndiritu and Daniell, 1999). These models have the potential to be used for evaluating various scenarios (e.g. land-use impact on hydrological processes and nutrients transport) based on the relationship that exists between the model parameters and physical catchment descriptors and offer the opportunity for improved performance through model structural adjustment (Chen and Adams, 2006, Fenicia et al., 2008). In addition, conceptual models are not computationally intensive and can easily be linked to an automatic optimisation algorithm for calibration purposes. The soil moisture accounting and routing (SMAR) model (O'Connell et al., 1970), NAM model (Nielsen and Hansen, 1973) HBV model (Bergström and Singh, 1995) and FLEX mode (Fenicia et al., 2006) all fall under the category of conceptual models.

The structure of physically-based models is deduced from physical phenomena. A pure physics based model attempts to represent the complex reality of the processes while utilizing physical field measurements to derive relationships between the variables. Unlike black box models and conceptual models that need to be calibrated, physical models are not meant to be calibrated. The large amount of information required has resulted in the limited application of physical models in the solving real world problems (Ndiritu and Daniell, 1999). The absence of sufficient data to setup a physically based model often necessitates for the model to be calibrated in order to obtain some representative parameters. Examples of physically-based models include the MIKE11 (Havnø et al., 1995), HEC-RAS (Brunner, 1998), and ACRU (Schulze, 1989).

Successful application of models is dependent on several issues and the selection of a particular model needs to be based on the availability of the data required by the model, the hydrological processes that needs to be modelled, the nature of the output required and in some cases the cost involved (Cunderlik, 2003, Beven, 2004).

2.2 APPLICATION OF MODELLING TO COMPLEX DATA SCARCE RIVER SYSTEMS

Modelling complex river systems with scarce data requires approaches that can optimally utilise the available data and at the same time attempt to improve the understanding of poorly known hydrological processes. Several catchment models with different complexity and data requirements have been developed for various purposes. The suitability of a particular model depends on the task at hand, data available, the scale and the time step of the input and output variables (Beven, 2004). With these factors in mind, this section, based on published studies of their applications, attempts to evaluate the suitability of the Pitman model (Pitman, 1973), ACRU (Schulze, 1989), Soil Water And Routing (SMAR) (Tan and O'Connor 1996), the coupled MIKE SHE and MIKE 11 model and the Soil and Water Assessment Tool (SWAT) (Arnold and Fohrer, 2005) as tools for modelling complex systems with data scarcity concerns. Considering the abundance of models that have been developed, the selection of these models for review is subjective as there are no strong reasons beyond the fact that the selected models are commonly used in South Africa and other regions in the world.

2.2.1 Pitman model

The Pitman model shown in Figure 2.1 is a calibrated deterministic model developed in 1973 (Pitman, 1973) to simulate the rainfall-runoff transformation process in a form suitable for the water resources assessment and has become one of the most widely used monthly rainfall-runoff models within Southern Africa (Hughes, 2004) and other regions. The Pitman model has been applied in a number of countries including South Africa (Pitman, 1973), Pakistan (Abulohom, 1997), Botswana (Hughes et al., 2006) and Zambia (Ndiritu, 2009), Tanzania and United States of America (Gan et al., 1997b). The basic form of the model has been preserved but subsequent versions have been re-coded by the original author and others. Additional components and functions have been added. The most recent upgrade was meant to improve model's representation of the highly developed South Africa catchments (Bailey 2008).

a. Main components

The main components of the Pitman model (Figure 2.1) are the rainfall distribution function, the interception function, the surface runoff function, the soil moisture storage, the runoff function, evaporation from the moisture store, runoff delays and the groundwater function (recharge and discharge). Several assumptions have been made in the development of the model:

- *Rainfall distribution function (RDF):* The rainfall is assumed to be controlled by parameter RDF set up such that the model execution time steps result in 4 model executions per month.

- *Interception function:* The quantity of water intercepted depends on a seasonally varying parameter and is related to the amount of rainfall and the conceptual storage.

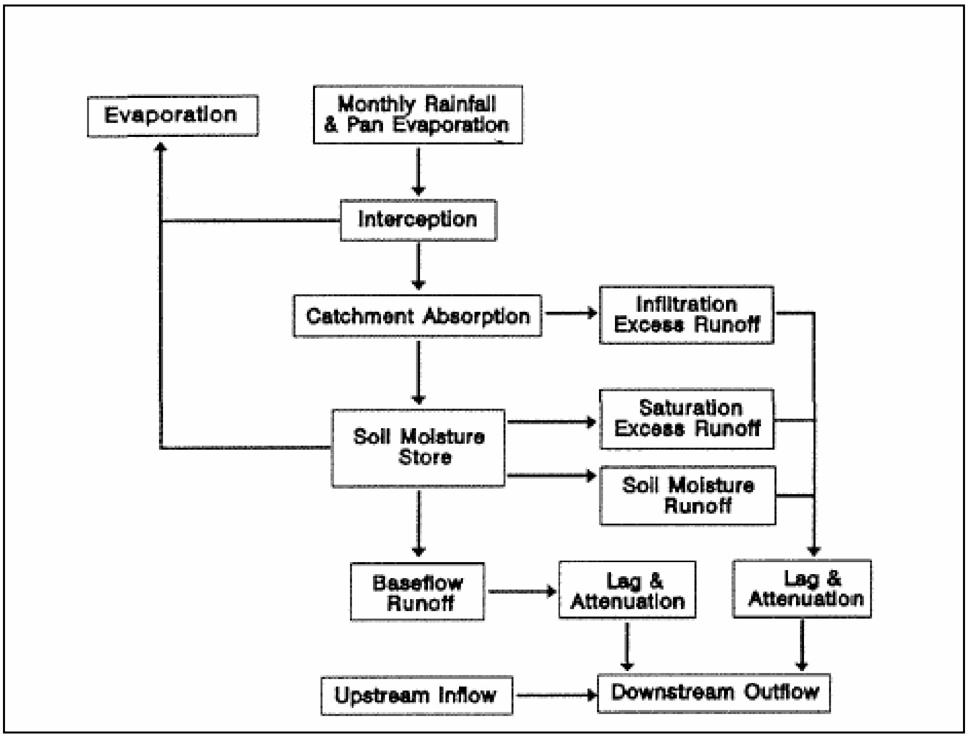


Figure 2.1: Flow diagram to illustrate the structure of the monthly Pitman model (Pitman, 1973)

- *Surface runoff function:* This function involves four parameters, two of which are based on a seasonally varying triangular distribution of the absorption rate. The third and fourth parameters define the symmetry and the portion of the catchment which is impervious.

- *Soil moisture storage and runoff functio*n: The runoff is assumed to be a result of the non-intercepted rainfall and excess storage from the upper soil zone. However, the moisture storage is assumed to be controlled by a nonlinear relationship between the runoff and storage through a single parameter. The runoff, however, is assumed to cease when the storage level fall below some threshold.
- *Evaporation from the moisture store:* The evaporation is assumed to be controlled by two parameters. The first parameter is responsible for the effective evaporation and the second one is responsible for the scaling of the vegetation type covering the basin. The potential evaporation serves as the variable.
- *Runoff delays and lags:* This component involves two parameters. The generated runoff from the soil zone is assumed to occur at different rates; hence, two lagging parameters together with the Muskingum routing equation are used for each soil zone.
- *Groundwater recharge and discharge components:* The two groundwater processes are represented by two functions. The first function is the recharge function which considers the existence of a minimum value of the

moisture storage below which recharge will not occur. The second function represents the groundwater discharge and it assumes one dimensional flow and is based on the reduction of the complexity of the basin spatial geometry to a simple geometric arrangement.

b. Scale of modelling and data requirements

The Pitman model is a monthly time step model that uses monthly potential evaporation (pan evaporation), cumulative monthly rainfall and the catchment area.

2.2.2 ACRU (Agricultural Catchments Research Unit) Model

The Agricultural Catchments Research Unit (ACRU) model was developed to simulate agrohydrological processes within Southern Africa (Schulze, 1989). ACRU is a physical, conceptual, multi-layer soil water budgeting model that operates at a daily time step. It is considered as a multipurpose model capable of simulating several water resources assessments issues, design flood estimation (Chetty and Smithers, 2005), irrigation demand, crop yield, assessment of hydrological and land use impact on water resources (Everson, 2001, Jewitt et al., 2004) and shallow groundwater systems.

a. Main components of ACRU model

The main operational modules in the ACRU model (Figure 2.2) are the soil water budgeting and potential evaporation and also the dynamic time or cyclic change and all the modules can be simulated as point, lumped or distributed and even GIS linked. At a distributed scale, the model recognizes the existence of units/sub-catchment called cells (not exceeding 30 km^2) which are joined together to represent a catchment. The model has been developed to be applicable in the simulation of runoff, reservoir status, sediment yield, irrigation demand and supply, land use, climate change and crop yield.

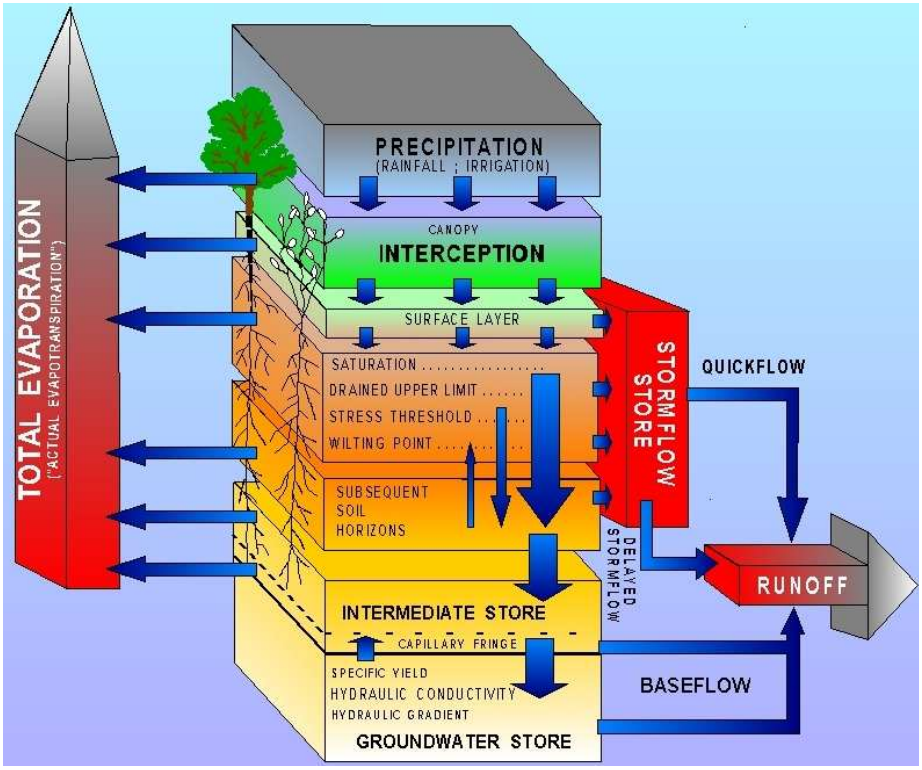


Figure 2.2: Schematic representation of the ACRU water budget (Schulze, 1995)

b. Data Requirements

In this regard, the ACRU model is considered as a multilevel model, designed to accommodate a hierarchy of different levels of available input data hence has the ability to optimally use the available information. The most commonly used data is at a daily time step. However, when this is not available, monthly or annual values can be used. In these cases, the annual or monthly values are disaggregated into daily values by the use of Fourier analysis. In addition, daily data can be disaggregated into lower time scales for use in routines which require finer scaled data. In some cases, data generation is done by the model itself. While this generated data is capable of representing the historical extreme events, it can also be used for long term planning purposes. Most of the parameters used are estimated from physical catchment characteristics. These include saturated hydraulic conductivity, bulk density, percentage content of clay, silt, sand and organic matter.

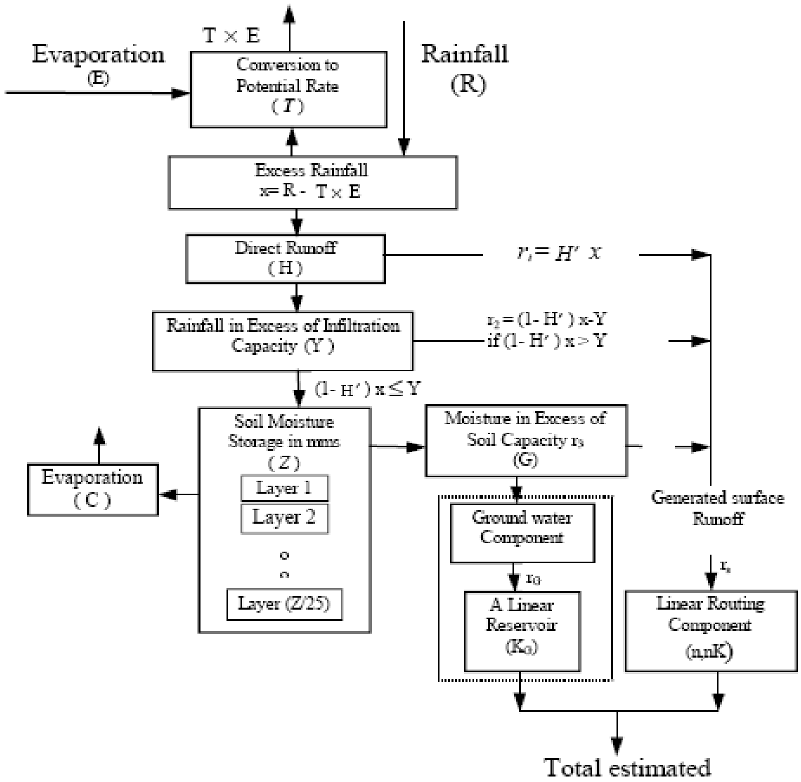
2.2.3 Soil Moisture Accounting and Routing (SMAR) Model

The earliest version of SMAR model was known as the 'layers model' introduced in the 1970s (Tan and O'Connor 1996). The SMAR model is a lumped rainfall-runoff model of the conceptual, quasi-physical type and it has undergone several developments and modifications at the University of Ireland (Shamseldin and O'Connor, 2001).

a. Main components in SMAR model

The SMAR model (Figure 2.3) consists of two modules namely the water budget module and the routing module as explained in the next subsections.

- *Water-budget module:* The catchment is assumed to be composed of a set of horizontal soil layers. The water storage in the catchment is augmented by rainfall and is depleted by potential evaporation at a rate that depends on the available water storage in the soil layers. The water-budget module produces three generated runoff components which are: direct runoff, runoff in excess of infiltration and runoff in excess of the soil storage capacity. A groundwater weighting parameter is used to subdivide the excess of the soil storage component into the groundwater runoff and the subsurface runoff.
- *Routing module:* The routing module has a groundwater runoff component and a surface runoff component. The generated surface runoff is routed through a two parameter distribution model and the groundwater runoff generated is routed through a single linear reservoir having a single storage coefficient parameter. For each time step, the total outflow from these components constitutes the river discharge.



Parameter Description
Z- the combined water storage depth of the layers
T- a parameter (less than unity) which converts the given evaporation to potential.
C- evaporation decay parameter, facilitating the lower evaporation rates from the deeper layers
H- the direct runoff coefficient
Y- the maximum infiltration capacity
n- the shape parameter of the Nash gamma function model, a routing parameter
nK- the scale parameter of the Nash gamma function model, a routing parameter
G- the ground water weighting parameter
K_G- the storage coefficient of the linear reservoir, a routing parameter

Figure 2.3: Schematic diagram of the SMAR Model and a summary description of its parameters (Shamseldin and O'Connor, 2001)

b. Scale and data required

The model has nine calibrated parameters and can operate at hourly or daily time steps. The inputs to the model for each time step are areal average rainfall and estimated potential evaporation or the pan evaporation data. The main model outputs are the river discharge and estimated actual evaporation.

2.2.4 MIKE models (MIKE SHE and MIKE 11)

The MIKE SHE modelling system is derived from the SHE hydrological model (Thompson et al., 2004). MIKE SHE is a deterministic, fully distributed and physically based model. The capability of the model to simulate both surface water (McMichael and Hope, 2007) and groundwater (Demetriou and Punthakey, 1999) related processes with precision that can be equated to the capability of those models that have been developed to simulate either surface water or groundwater alone. This makes the MIKE SHE model a powerful tool. Most of the time, the stream width is smaller than the grid size and this impairs the MIKE SHE model's ability to accurately simulate stream flow. Therefore, in practice, the MIKE 11 model (Thompson et al., 2004), a river network modelling system is coupled to MIKE SHE model so as to account for the bi-directional interaction between the watershed hydrological processes and the river hydrodynamics. The open model interface (OpenMI) that enables the linking processes of the two models is called MIKE Zero which ensures that less effort is required during the linking process.

a. Main components of the MIKE SHE and MIKE 11 models

The main components of the MIKE SHE model include the overland flow, vertical flow, evapotranspiration and interception (Thompson et al., 2004). The physical processes that can be simulated by the model are shown in Figure 3.4. In the MIKE 11 model, the main component involved is the streamflow hydrodynamics for the river network and also the physical characteristics of the river that include the hydraulic structures. All these components are described below.

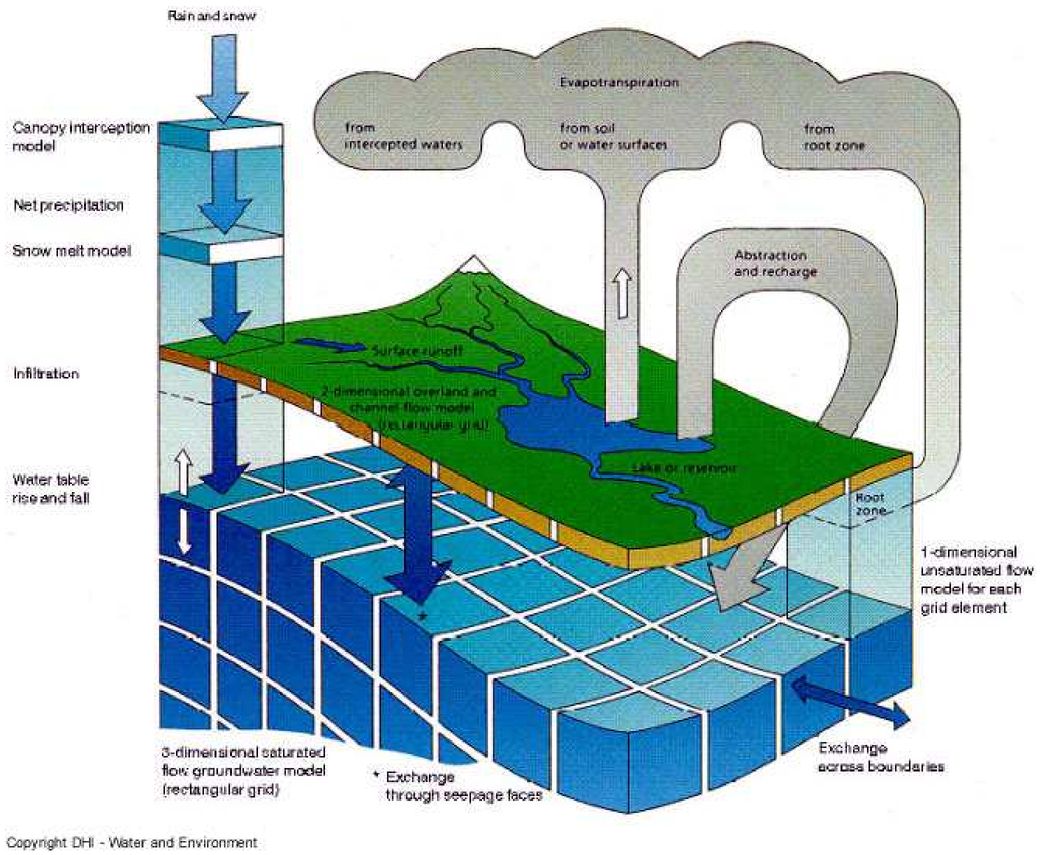


Figure 2.4: Schematic representation of the MIKE SHE model (Refsgaard et al., 1995)

- *Overland flow:* The overland flow is simulated by solving the diffusive wave approximation in two horizontal dimensions.

- *Vertical flow:* The vertical flow is simulated for two conditions, saturated and unsaturated states. For the saturated states, the zone is described by the three-dimensional Boussinesq equation while for the unsaturated zone the one-dimensional Richards equation is used.

- *Evapotranspiration and canopy interception:* In order to quantify the net evapotranspiration and canopy interception, the MIKE SHE utilizes the techniques developed by Kristensen and Jensen (1975).
- *Stream flow hydrodynamics:* The channel flow and the water level are evaluated by a complete formulation of the one-dimensional Saint Venant equations.
- *Coupling component:* This is an essential component that allows for the accurate description of the dynamic processes of the interaction between streamflow and groundwater while incorporating all the complex river/stream branch system, flood plain as it is explicitly defined in MIKE 11.

b. Scale and data required for the coupled models

- *Hydrological data:* This includes hydraulic conductivity, specific yield of unconfined aquifer, cross-section of river channel, Manning coefficients, structures (weirs, dams and diversion), stream flow, groundwater level, DEM, land cover and the river network. All this information is required for every grid and the spatial resolution should be as fine as possible.
- *Meteorological and water user data:* These sets of data include rainfall, temperature, surface water abstraction and reference potential evapotranspiration. The resolution of the data can range from less than hour

to more than a day. However, a finer resolution is preferable although this is not always possible.

2.2.5 Soil and Water Assessment tool (SWAT) Model

The Soil and Water Assessment Tool (SWAT) model is a result of continuous effort of nearly 30 years of model development started by United States Department of Agriculture Research Service (USDA-ARS) to predict the impact of land management practices on water, sediment, and agricultural chemical yields in large ungauged basins (Arnold et al., 1995, Arnold and Fohrer, 2005). Therefore, the model was meant to be used without the need for calibration, to make use of readily available input data, to be computationally efficient and to be capable of simulating long periods suitable for evaluating the effects of management changes (Arnold and Fohrer, 2005).

a. Main components of the SWAT model

The model attempts to realistically simulate the major hydrologic components and their interactions. These hydrology processes as discussed in Arnold and Fohrer (2005) include: (i.) surface runoff estimated using the SCS curve number or Green–Ampt infiltration equation; (ii.) percolation modelled with a layered storage routing technique combined with a crack flow model; (iii) lateral subsurface flow; (iv) groundwater flow to streams from shallow aquifers; (v) potential evapotranspiration by the Hargreaves, Priestley–Taylor; and Penman–Monteith methods; (vi) snowmelt;

(vii) transmission losses from streams; and (viii) water storage and losses from ponds.

Several modifications and addition of other components have been done in order to improve the model's applicability to a range of catchments with varying levels of complexity. These include the calibration routines (manual or automatic) (Eckhardt and Arnold, 2001), modelling of the uncertainties (Benaman and Shoemaker, 2004, Whittaker, 2004), interfacing SWAT with other environmental and economic models (Sophocleous and Perkins, 2000, Sophocleous et al., 1999, Qiu and Prato, 2007a, Qiu and Prato, 2007b, Qiu and Prato, 2001, Qiu and Prato, 1998) and its suitability when applied in data scarce catchments (Ndomba et al., 2008).

b. Scale and data required by the SWAT

The model's data requirements depend largely on the type of the process being simulated. It should be noted, however, that the efforts towards improved model performance that have been done in the past have resulted in increased data requirements over time. As such, with respect to the modelling of most processes, substantial amounts of data including the hydro-meteorological data, digital elevation model, soil type and land use/land cover are required. The model operates at a daily time step (Arnold and Fohrer, 2005).

2.2.6 Summary

The 5 models described above include the Pitman model, ACRU model, Soil moisture accounting and routing model, MIKE SHE-MIKE 11 and Soil and Water Assessment Tool. Since its development in 1973, the Pitman model has been modified to improve its representation of the highly developed South Africa catchments (Bailey, 2008). It is the most likely tool that would be selected for the assessment of surface water availability in southern Africa (Hughes, 2004). Although the Pitman model is a calibration model, the fact that it operates at a monthly time step means that for those tasks for which a daily time step catchment operation is required, the model will be deemed less suited than those models developed for application at daily time steps. In addition, aggregating daily input data into monthly values has, as a consequence, an associated loss of information (Salas et al., 2005). On the other hand, the ACRU model has a large number of parameters that are estimated from the catchment's physical characteristics (e.g. soil information, land cover, land use information and hydraulic conductivity) (Schulze, 1995). Hence, a large amount of data is required and consequently, its application in practical situations is limited due to the difficulties in obtaining the field parameters (Jewitt and Schulze, 1999). In addition, there are difficulties in understanding the ACRU model and as such its use can be challenging, particularly for new modellers.

The SMAR model is a relatively simple calibrated conceptual model and works well in wet catchments where the temporal and spatial variability is not significant.

However, in semi-arid regions, a wide range of flow scenarios is experienced and this may restrict the performance of the model (Gan et al., 1997a). With regard to the model's representation of complex systems, major improvement to the model would be needed so as to adequately incorporate these complexities.

The coupled MIKE models (MIKE SHE and MIKE 11) conceptualize almost all the watershed's hydrological and hydrodynamic processes and it is a robust distributed, physically based model. The amount of information required to represent all these concepts at grid scale is large (Jain et al., 1992) yet the observation/measurement is not always done at grid scale. Although an interpolation technique can be applied to estimate the various state variables for each grid, the relevance of those estimates is doubtful since there is a high variability of the processes in semi-arid and arid regions. However, calibration can improve the model results although this makes the modelling effectively conceptual and not physically-based. For the models to be adequately used, long records are required, a drawback that has resulted in its limited application. In addition, the expertise needed to setup the model is high and may require the user to undertake some tailored courses as a prerequisite.

The Soil and Water Assessment Tool is a conceptual model developed to predict the impact of land management practices on water, sediment, and agricultural chemical yields in large ungauged basins. Since its development, the SWAT model has undergone a number of modifications over time largely as a consequence of special requirements and provisions have been put in place to make it adaptable to more modification to meet future modelling requirements. Regardless of all these modifications, the model requires a large amount of information which may not be easily available in some data scarce catchments. For instance, the effect of coarse digital elevation data resolution result into uncertainty (Chaubey et al., 2005) and the soil data resolution required for adequate model performance depends on the accuracy of the results being targeted (Geza and McCray, 2008). The use of the SWAT model for operational purposes requires personnel who are experienced in modelling (Bärlund et al., 2007) and may require some customization when used in various hydrological conditions (Ndomba et al., 2008). Therefore, the application of SWAT model in complex river systems with scarce data may require some customization which may not guarantee the effort and expertise applied.

Generally, some of the models discussed above require substantial amount of data, are difficult to learn and use, do not adequately represent the complexity of the river system and operate at a monthly time step. In addition, attempts to modify any of the models may require a lot of effort and there is no guarantee that this may improve the quality of the model simulations. For this reason, it is imperative to identify other avenues that can be used to adequately simulate complex systems with scarce data. The performance of these alternative approaches would however need to be assessed to ensure their suitability.

3 FUZZY INFERENCE AND HYBRID CONCEPTUAL-FUZZY INFERENCE MODELLING

Since, the mathematical representation of systems involves some simplification and the use of assumptions (Bardossy et al., 1995), it is, therefore, reasonable to suggest that mathematical models inadequately represent the entire system being modelled and this inadequate representation can be a source of significant uncertainty. Heuristic approaches, like artificial neural networks (ANNs) and fuzzy logic provide an avenue that is flexible (Gershon, 1987, Lohani et al., 2006) simply because these map the input data space to the output space obviating the need for most of the assumptions associated with conceptual model development for instance. It is, however, important to note that with limited and/or noisy training data sets, ANNs may give results which are inconsistent and meaningless (Tayfur et al., 2003). Zimmermann (2001) noted that as the complexity of the system increases, the ability of the ANN to make precise and yet relevant statements about the system diminishes until some point beyond which the precision and relevance become mutually exclusive. To accommodate these complexities, a suggestion was put forward regarding the application of the fuzzy set theory in modelling complex systems. Fuzzy based approaches are highly promising considering the fact that they possess the potential of being applied in a variety of modelling problems (Beven, 2004). Fuzzy set theory provides a framework for dealing with uncertainty and vagueness at various levels of modelling even in situations where the data available is inadequate (Schulz and Huwe, 1997, Xiong et al., 2001, Zimmermann, 2001). Fuzzy inference is more logical and scientific in describing the properties of objects as well as relationships that are not completely known. For this reason, fuzzy logic is considered as one of those data mining techniques that can extract comprehensible rules from data as evidenced by Bessler et al. (2003) who used fuzzy logic to generate a set of reservoir control rules that gave the best operating policy. In addition, fuzzy logic can easily be linked to a numerical or mathematical model to form a hybrid model.

3.1 FUZZY INFERENCE APPLICATIONS TO CATCHMENT MODELLING

Fuzzy Logic was described by Zadeh (1965) as an extension of the classical set theory and this theory has been applied in many fields including hydrology. Xiong et al. (2001) tested the ability of the Takagi-Sugeno fuzzy linear system (TS1), the simple average method (SAM), the weighted average method (WAM), and neural networks (ANNs) of inheriting the merits of five models by combining their respective model outputs. The models used were the simple linear model (SLM), the linear perturbation model (LPM), the linear varying gain factor model (LVGFM), the constrained linear systems with a single threshold (CLS-T) and the soil moisture accounting and routing procedure (SMAR). Thereafter, a comparison of the results from the combined models was carried out. A good basis for comparison was achieved by the use of concurrent data. The performances of the combination methods showed that the WAM, NNM and TS1 behaved similarly during the calibration while during the verification period the TS1 performed better with most of correlation coefficient values being higher than 0.7. Xiong et al. (2001) noted that a better performance of TS1 could be achieved if the model is not over-parameterized. On the other hand, when the number of parameters is less than optimal, the flexibility of the model is limited and as such may not achieve a better performance. No suggestion was given on how to achieve an optimal number of parameters.

Schulz and Huwe (1997) applied fuzzy logic to express imprecision of model parameters of steady state water flow in unsaturated soils. These uncertainties were attributed to the subjective estimation of parameters such as saturated/unsaturated hydraulic conductivities, water retention curves and also the boundary conditions such as precipitation or evapotranspiration rates. A sensitivity analysis showed a strong dependency of the resulting values of the membership function on the shape of the membership functions of the input parameters and the range of each parameter. For the case of soil water pressure, the uncertainty increased with the number of model parameters and the distance from the known groundwater level.

In order to overcome the problems of data scarcity and computational difficulties such as convergence and numerical instability that might occur, Tayfur et al (2003) applied a fuzzy logic algorithm to model runoff induced sediment transport from bare soil surfaces. Six bare slopes of 5.7%, 10%, 15%, 20%, 30% and 40% were tested with four different rainfall intensities of 32, 57, 93 and 117 mm/h and a constant infiltration rate for each run of 5.3 mm/h. The input data used was the precipitation and slope while the sediment load was the output. Both the precipitation intensity (r) and slope (S) data were classified into three subsets, low (r< 40 mm/h), high (40 < r < 80 mm/h) and very high intensity (r > 80 mm/h) and mild (S < 10%), step (10 < S < 20%) and very steep slope (S > 20%) respectively.

The results from the fuzzy logic algorithm were also compared with the results from an artificial neural network and a physically based model whose basis was the one dimension kinematic wave approximation. The fuzzy logic algorithm used triangular membership functions to classify the input and the output data. Results showed that the fuzzy model performed better than the other approaches in predicting sediment loads from mild slopes under high and very high rainfall intensities and from very steep slopes under low rainfall intensities.

Two rainfall runoff models based on the Takagi-Sugeno fuzzy inference system were developed by Jacquin and Shamseldin (2006). The performances of these two models were thereafter compared with the simple linear model (SLM), Linear Perturbation Model (LPM) and Nearest Neighbour Linear Perturbation Model (NNLP). The first model was intended to take into account the effect of changes in the prevailing moisture content and the second model was intended to incorporate the seasonality effect in the analysis. The rainfall index was used as input to the first model and is given by:

$$RI_i = G^a \sum_{j=1}^{L} P_{i-j+1} h_j^a \qquad 2.11$$

where RI_i represents the rainfall index value at time step i, P_j is the precipitation measurement at time step j, L is the memory length of the catchment, G^a is the gain factor of the simple linear model (SLM) and h_j^a is j^{th} ordinate of the discrete pulse response function of the SLM. In the second model, the input was a normalised time of the year given as:

$$t^n = \frac{t}{365} \qquad 2.12$$

where t is the Julian day of the year and the normalised time varies from 1/365 to 1 for 1st January to 31st December respectively for a non-leap year.

The results showed that the performance of the TS fuzzy inference models were better than the SLM for the same input data. However, the second TS fuzzy model's efficiency values were generally not as good as the SLM model efficiency values. When compared to NNNLP and LPM, the TS fuzzy inference models performed better in some catchments than the NNLP and LPM. This suggests that increasing the complexity of the model structure does not always improve the performance of a fuzzy model. Rather, the type of information incorporated into the model may have a greater impact on the model performance (Jacquin and Shamseldin, 2006). These observations imply that simple fuzzy logic models have the ability to incorporate uncertainties even with little information and as such, can be used as an alternative for modelling poorly understood and non-linear catchment processes.

Altunkaynak and Åžen (2007) applied fuzzy logic in modelling water level time series of Lake Van located in Turkey. These series showed significant fluctuations, hence, creating a potential risk from the floods for the area downstream of the lake. Water level forecasting was considered as a viable option in reducing the risk of the potential damage that could result from the occurrence of floods downstream. Altunkaynak and Åžen (2007) applied TS fuzzy logic to model the water level in the lake using the historical water level time series and used rainfall as input data. As an outcome, fuzzy logic produced a mean absolute error of 2.09 % while autoregressive integrated moving average with exogenous input (ARMAX) produced a mean absolute error of 3.67 %.

Though the application of fuzzy inference system in water resources is relatively new, it has generated considerable enthusiasm (Lohani et al., 2006) which can be attributed to its ability to create a framework to deal with data which is incomplete, ill-defined and inconsistent. Moreover, it is less computationally intensive than stochastic approaches and it can deal with non-linear problems where the processes cannot be expressed explicitly in mathematical form. Unlike most stochastic approaches applied to water resources modelling problems that involve the subjective selection and use of suitable distributions, this is not the case with fuzzy logic based approach. There is a growing body of evidence that suggests that fuzzy approaches to the solution of real problems are an effective alternative to traditional methods (Altunkaynak and Åžen, 2007).

3.2 HYBRID FUZZY INFERENCE APPLICATIONS TO CATCHMENT MODELLING

In the last decade, hybrid modelling has begun gaining momentum, though it was introduced in the 1990’s (Jakeman and Hornberger, 1993). Van Lith et al. (2002) in a comparative study, investigated and compared the use of fuzzy clustering, genetic algorithms and neuro-fuzzy as three approaches for identifying hybrid submodels. While the neuro-fuzzy approach was observed to be sensitive to the initial values, the genetic algorithm was considered as a suitable alternative to fuzzy clustering owing to its self-adapting capability of identifying the optimal model structure. In the study by Seibert and McDonnell (2002) the incorporation, using fuzzy inference, of soft data from an experimentalist into a three box conceptual model resulted in a better goodness-of-fit. Xiong et al. (2001), on the other hand, used fuzzy inference as one of three techniques to combine the outputs of five different rainfall-runoff models. The combination of the various model outputs was premised on their being better suited for modelling different aspects of the rainfall-runoff process. The study showed that the fuzzy inference technique could achieve the most satisfactory results if it is not over-parameterised. Schulz and Huwe (1997) demonstrated that the structure of a conceptual model can be replaced with fuzzy logic variables in a study in which fuzzy logic variables were used in the Darcy-Buckingham equation. However, none of these studies considered the use of the fluxes to link the conceptual model stores and the fuzzy inference model.

Quantifying model parameter uncertainty has been addressed by several researchers using a number of approaches such as the use of the Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992) and the Bayesian Inference approach (Kagoda and Ndiritu, 2008). An alternative hybrid fuzzy-mechanistic approach for incorporating uncertainties was suggested by Lauzon and Lence (2008). This approach modifies the parameter values of mechanistic models whose parameters were originally fixed such that the parameter values are defined by a function that enables their values to possess meaningful relationships with catchment processes. Lauzon and Lence (2008) applied this approach in a rainfall-runoff model and an algal concentration model. The study noted that the approach is straightforward and has the potential of being applied in environmental and natural resource modelling. With respect to interpretability and transparency, fuzzy logic is highly regarded as a suitable technique for developing hybrid models (Van Lith et al., 2002).

4 LETABA RIVER SYSTEM

4.1 INTRODUCTION

Several complex river systems with scarce data exist in many parts of the world including South Africa. The Letaba River system (Figure 4.1) being one such system has been selected for analysis using conceptual- and fuzzy inference-based approaches. In this chapter, a description of the physical characteristics of the system, the information available and the operation of the system is provided. It is appropriate to describe these aspects of the river system as it is these that contribute to the complexity of the river system and as such would form some basis for understanding the system whose water resource needs to be optimally and sustainably utilized for the benefit of the current and future generations.

The Letaba River system is located in the north-eastern part of South Africa and is within the Luvhuvu/Letaba Water Management Area. The basin covers an area of 13 669 km^2 with the eastern part of the catchment being a low veld that drops to an altitude slightly below 450 m while the western part is a mountainous region that rises to an altitude exceeding 2000 m.

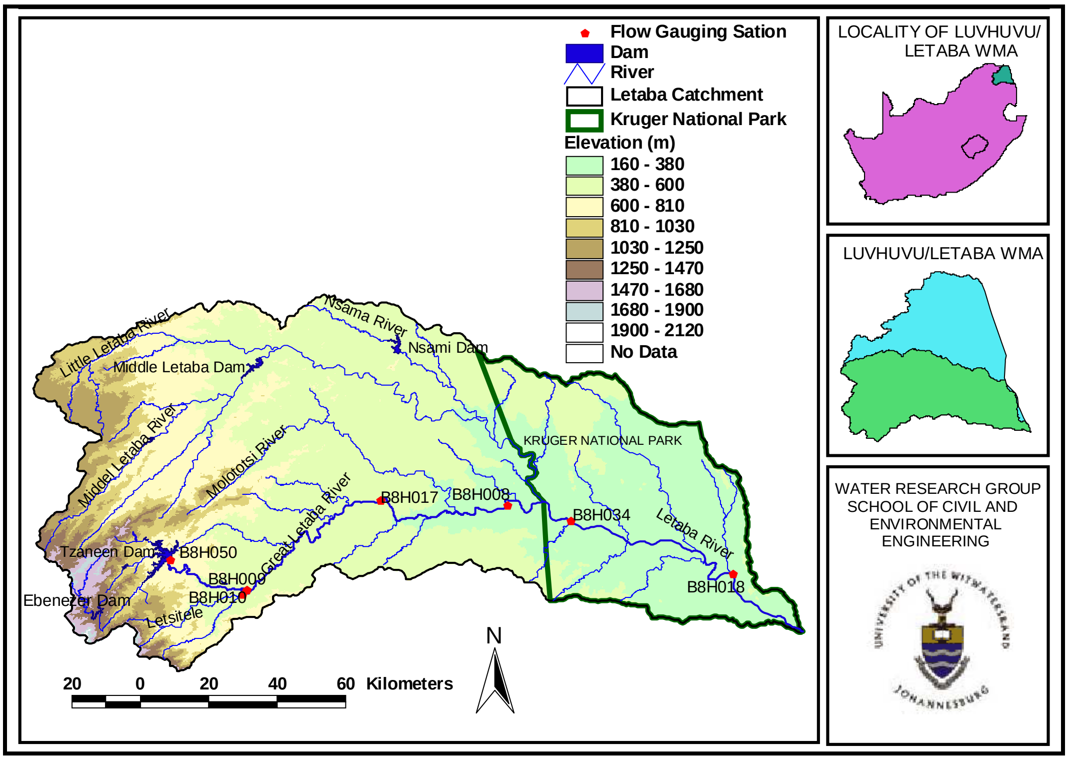


Figure 4.1: Location of the Letaba River system

4.2 PHYSICAL CHARACTERISTICS

The main physical characteristics of the Letaba River system include the river network (Figure 4.1), hydraulic structures (Figure 4.2), alluvial riverbed and the riparian vegetation existing along the flood plain (Figure 4.4).

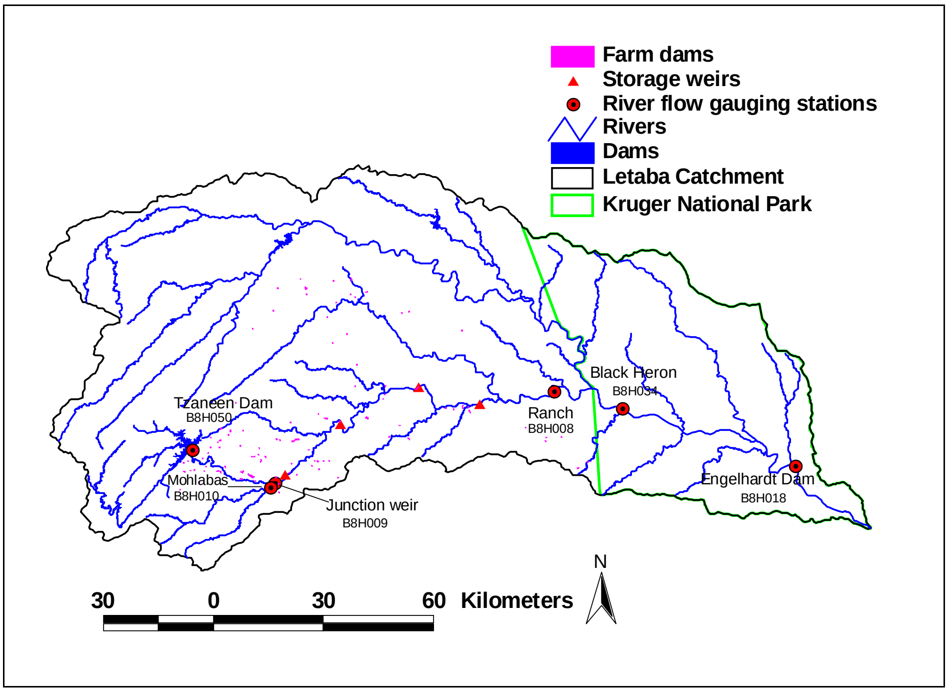


Figure 4.2: Hydraulic structures in the Letaba River system

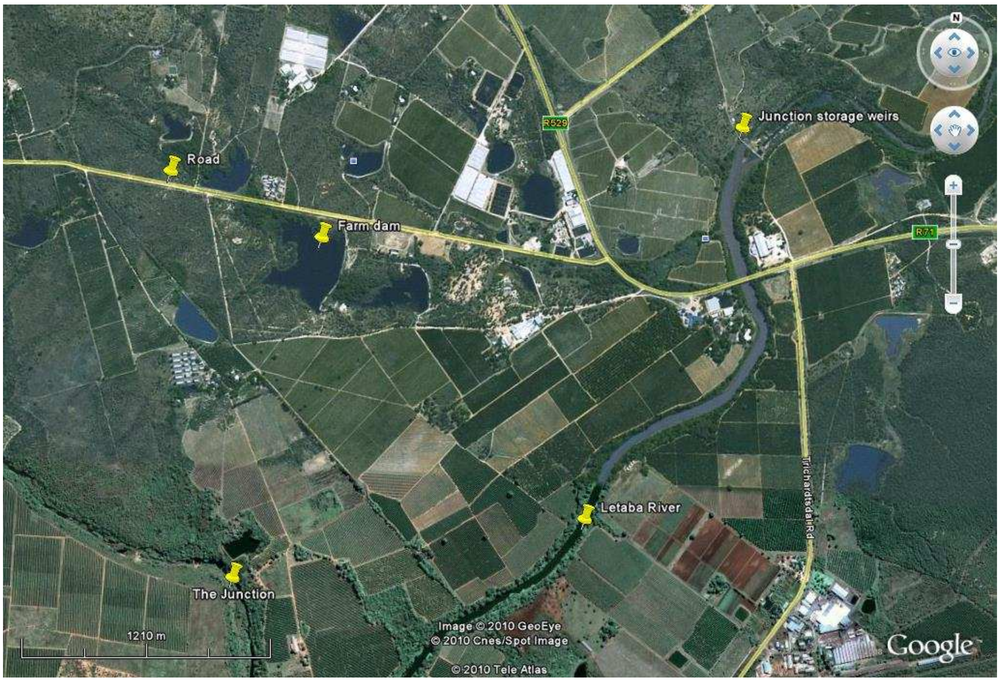


Figure 4.3: An aerial view of a section of the Letaba River system. (Google Earth, 31 March 2010)

4.2.1 River network

The main river downstream of the Tzaneen Dam is the Letaba River and the water released from Tzaneen Dam flows a distance of about 150 km before getting into the Kruger National Park (KNP). Several tributaries contribute to the flows in the Letaba River. With the exception of the Letsitele River, a perennial river which joins the Letaba River at Mahlabas, 34 km downstream of Tzaneen Dam, the contributions from other tributaries are mainly during the wet season.

4.2.2 Hydraulic structures

Downstream of the Tzaneen Dam, several structures that serve a number of different purposes have been constructed at a number of locations across the river. The structures (Table 4.1) can be classified into four categories:

- *Storage/release weirs (S/RW):* These structures are used to store water that is used during the periods of low flows.
- *Flow gauging weirs (FGW):* These are real time flow gauging points of varying accuracy as discussed in a Section 4.6.2. The width of the lowest level of the weir (LLW) is an essential consideration; the smaller it is, the greater the accuracy with which low flow measurement is done.
- *Abstraction/diversion weir (A/DW):* These structures are used for diverting water from the river into the canals that supply water to users who are far from the River.

- *Check points (CP):* This is a low bridge (drift) which is used by a water bailiff from the Letaba Water User Association to check the depth of flow so as to decide whether or not to release water from the storage weirs and/or to request for more water to be released from Tzaneen Dam.

Table 4.1: Structures existing along the Letaba River downstream of Tzaneen Dam and their respective functions (Figure 4.2).

S/N	Name	Code	FGW	S/RW	A/DW	CP	LLW (m)	Situation
1	Tzaneen Dam	B8H050	✓				3.000	Good
2	Junction weir	B8H009	✓				7.625	Good
3	Mohlabas	B8HO10	✓				7.625	Good
4	Prieska	B8H017	✓	✓	✓		156.175	Silted
5	The Ranch	B8H008	✓			✓	6.005	Good
6	Black Heron	B8H034	✓				9.940	Very good
7	Engelhardt	B8H018	✓	✓			326.320	Water Hyacinths
8	Nodweni Dam	-		✓				Good
9	Jasi	B8R010		✓				Good
10	Junction	B8R008		✓				Good
11	Letaba North	-			✓			Good
12	N & N Canal	-			✓			Good
13	Lower bridge	-				✓		Good
14	Yamona	-		✓				Good

4.2.3 Riverbed and riparian vegetation

The riverbed of the Letaba River consists mainly of sandy material whose thickness is up to 10 m and the floodplain extends up to 500 m and is mainly in the lower reaches of the river basin (DWAF, 2006a). The slope of the riverbed is steep in the upper reaches and mild towards the downstream. In addition, some vegetation exists within the flood plains and river banks. The transpiration of this vegetation is one of the process by which water is diverted from the streamflow. The density of the vegetation decreases towards the downstream (Figure 4.4 and Figure 4.5) and this can be attributed to reduced water availability downstream and/or the reduction in the concentration of agrochemicals (e.g. nitrates, phosphorous) in the flowing water. The source of these agrochemicals is the fertilizers used in the farms adjacent to the upstream reaches of the rivers. Water pollution resulting from silt, sediment and agrochemicals has been reported by Gyedu-Ababio (2005). The growth of aquatic alien weeds like hyacinths is a consequence of the increase in nutrient load in the flows (Figure 4.6).

Figure 4.4: The density of the riparian vegetation upstream side of the catchment (viewed on 18 June 18, 2008 on Google Earth)

Figure 4.5: The density of the riparian vegetation downstream side of the catchment (viewed on 18th June 2008 on Google Earth)

Figure 4.6: The Engelhardt Dam with water hyacinths covering the edge

4.3 WATER DEMAND AND UTILIZATION

Several water demands are imposed on the Letaba River system and these are reflected in the annual water allocation which sums up to 130 Mm^3. This value has been derived from several allocations including the irrigation water allocation which is 103.9 Mm^3, the combined domestic and industrial water allocation of 7.13 Mm^3 and an ecological water requirement of 18.9Mm^3 (DWAF,2006b). The annual historical yield (1:68 years) of Tzaneen Dam is 74.6 Mm^3 (DWAF, 2006b). It is clear that the sum of the various water demands outstrips the yield of Tzaneen Dam and as such cannot be satisfied. This highlights the need for the implementation of strategies that would ensure optimal operation of the river system so as to enhance the water resources sustainability.

4.4 OPERATION OF THE SYSTEM

Decision making with respect to the current operation of the Letaba River system is mostly done on an on-going basis at near real time, based on the rule of thumb. The operation aims at making water available to satisfy vital demands during periods of low flow while providing as much water during normal conditions. Both scheduled and unscheduled abstractions are made from the river from time to time with the scheduled abstractions based on values agreed upon amongst the stakeholders who own water rights while unscheduled abstractions are done when the water available in the river is considered to be more than what is required by the downstream users. It is mainly the farmers who execute these unscheduled abstractions.

During the rainy periods, the releases from Tzaneen Dam are reduced and the demands are largely met by the incremental flows. When the amount of water available in the river is more than the total requirements of the downstream users, the farmers abstract the excess flood water (unscheduled abstraction) and store this, together with the water from scheduled abstractions, in farm dams. During dry periods, water requirements are supplied with water released from the Tzaneen Dam as well as with water from the storage weirs which supplement the flows in the Letaba River. Restrictions are imposed on commercial users (e.g. irrigators) as a means of preventing total collapse of the system. Farmers tend to supplement their supply by using water from individually owned boreholes. It is worth noting, however, that the amount of groundwater abstracted using these privately owned boreholes are not monitored.

In addition to the releases made from Tzaneen Dam, the water stored in the weirs located in the upper reaches of the river are considered to be only for emergency purposes and as such is only released when there is a need to supplement the water supplied by the incremental flows as well as from the storage weirs in the lower reaches of the river (Venter, 2008).

Within the Letaba River system, the actual supply to the irrigation fields is mainly from the farm dams and the main irrigated crop is the citrus and according to the irrigation water requirement values obtained from the SAPWAT and PLANWAT (van Heerden et al., 2009) database set for the Limpopo area (Figure 4.7). The river has been demarcated into three river reaches (i.e. first, second and third river reach) as explained in Section 6.1. Figure 4.7 shows time-series of the amount of demand for citrus plantation in the two upper reaches. The difference in the series for the two respective reaches is attributed to the difference in areas irrigated by water from the first and second river reaches.

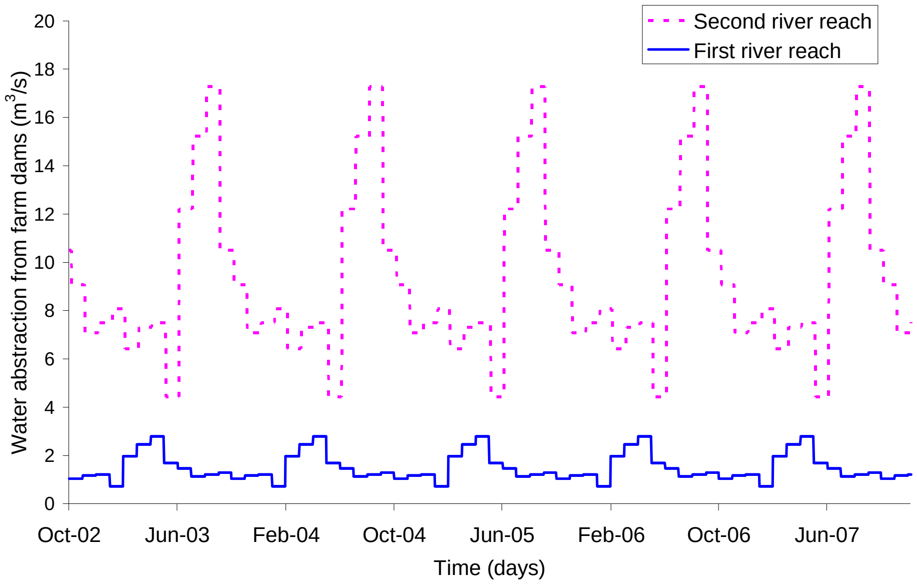


Figure 4.7: Typical water abstraction values derived from citrus demand for the Limpopo (Van Heerden et al., 2008)

4.5 FLOW CHARACTERISTICS

With respect to the flow gauging stations on the river downstream from the Tzaneen dam, the following gauging stations exist: B8H050 (Tzaneen Dam), B8H009 (Junction weirs), B8H010 (Lestitele inflows), B8H008 (The Ranch) and B8H034 (Black Heron) in that order (Figure 4.2). A plot of the flow measurements taken at these flow gauging stations is shown in Figure 4.8. The plots reveal that the pattern of flows observed in the upper river reach depart significantly from what can be considered a natural flow pattern. This is the consequence of human intervention with respect to the operation of the Tzaneen Dam and the intermittent operation of the storage weirs and abstractions from surface and groundwater within the floodplain.

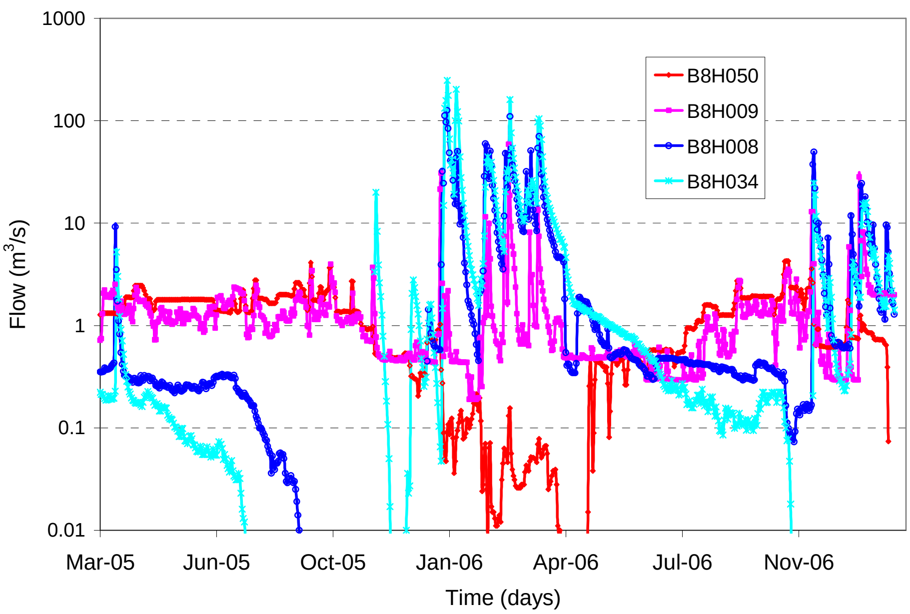


Figure 4.8: Typical flows along the Letaba River

4.6 DATA AVAILABLE

The data used in this study was sourced from DWAF, the Letaba Water User Association and the South African Weather Services (SAWS). The data includes streamflow, rainfall, evaporation, groundwater level, surface water abstraction. With the exception of the surface water abstraction which is available at monthly time steps, all the other data is available as daily values. Following is a description of the data that was used in this study.

4.6.1 Rainfall data

Although there is a rainfall database for South Africa (Lynch, 2003), the record length extends only from 1900 up to 2000 and as such data from this database was not considered for the modelling purposes. This is because of the floods that occurred in the year 2000 within the southern African region whose magnitudes transfigured some of the region's rivers' characteristics. It is reasonable to expect that this had an impact on the rainfall-runoff transformation processes within the region's catchments, so for this reason, it was not deemed feasible to use the rainfall database. It was assumed that an insight into the current hydrological situation within the catchment would be reflected better in the data observed and recorded after the 2000 floods. For this reason, the modelling was done using data recorded after the year 2000. Consequently, the rainfall data used was obtained from the Department of Water Affairs and the South Africa Weather Service (SAWS). Figures 4.9 and Figure 4.10 show the temporal distribution of the rainfall data and spatial distribution of the gauging stations respectively.

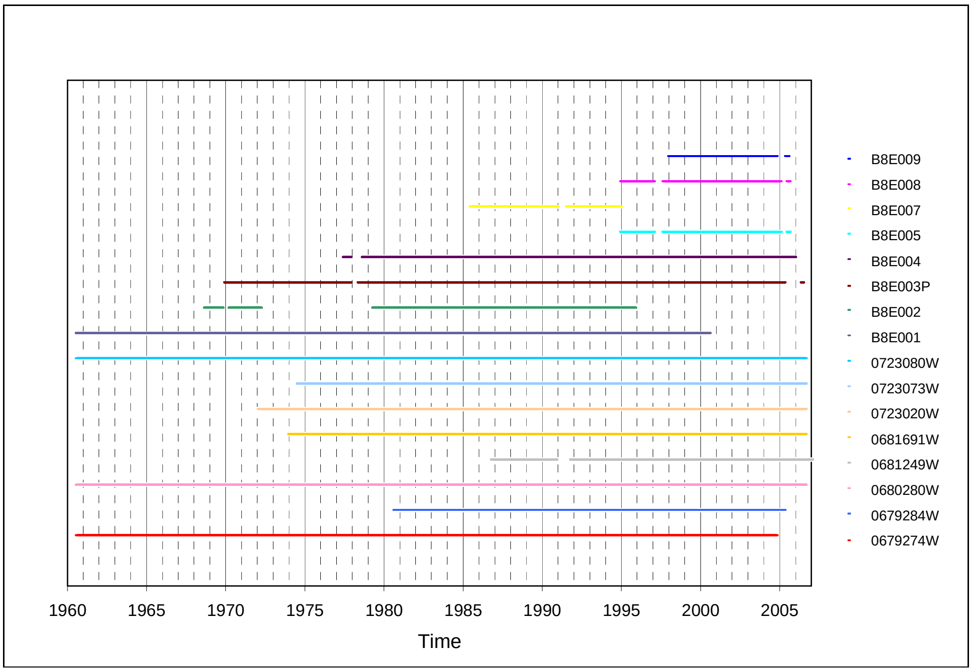


Figure 4.9: Temporal distribution of the rainfall data

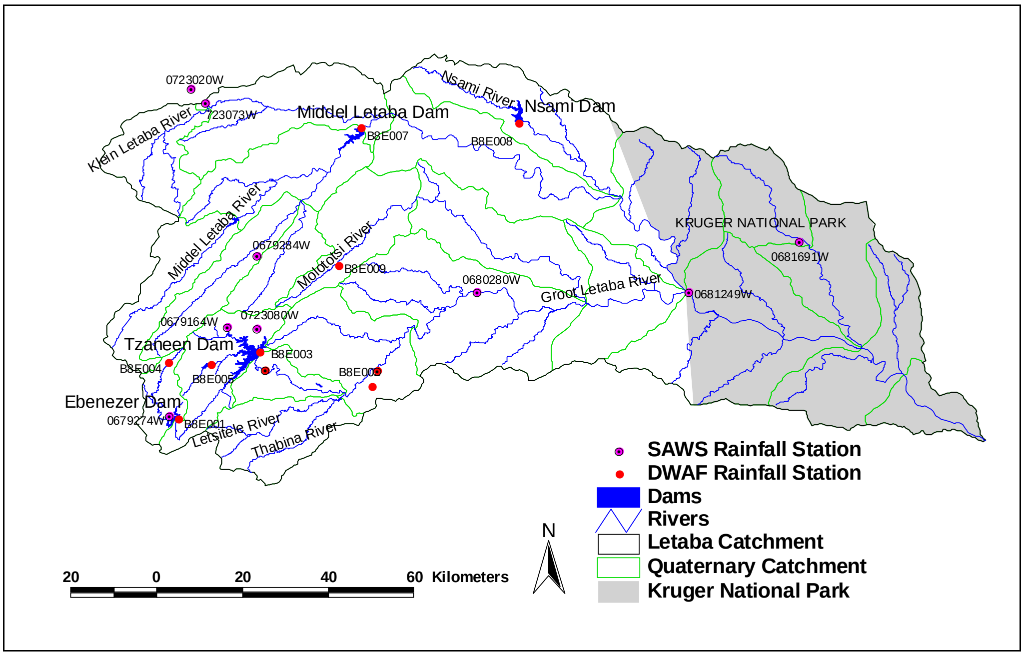


Figure 4.10: Spatial distribution of the rainfall gauging stations

The temporal distribution of the rainfall gauges suggests that the rain gauges owned by the SAWS have longer records than those owned by DWAF. Some of the rain gauges owned by DWAF have records ending in 1995.

4.6.2 Streamflow data

Along the Letaba River, there are six river gauging stations with data and one of gauging station along the Letsitele River. The temporal distribution of the data is shown in Figure 4.11, while the spatial distribution of the gauging stations is shown in Figure 4.12.

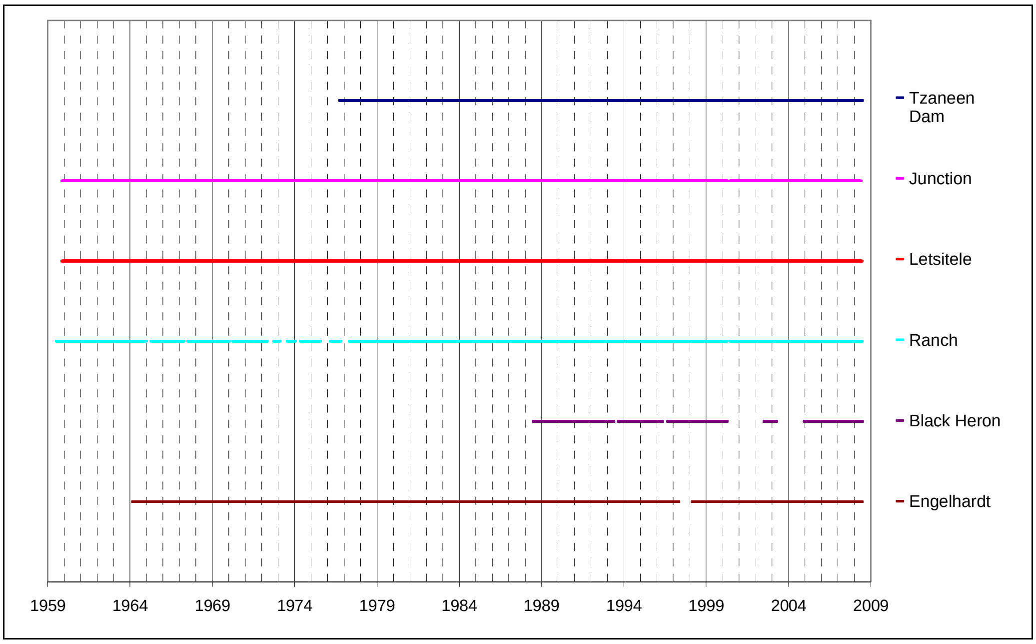


Figure 4.11: Temporal distribution of the streamflow data

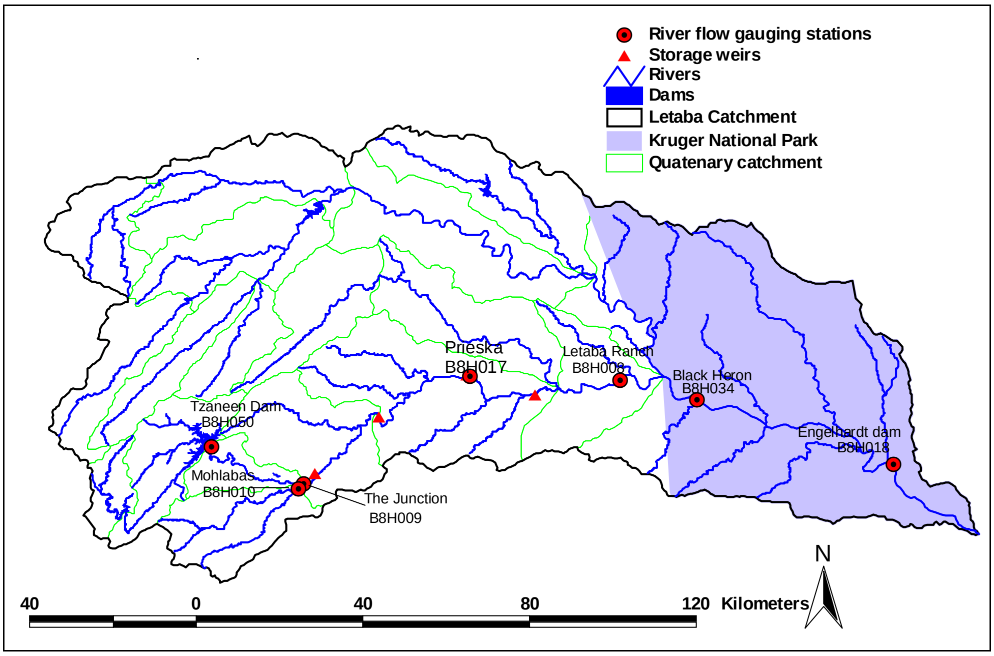


Figure 4.12: Spatial distribution of the streamflow gauging stations.

The current situation with respect to the flow gauging stations indicates that:

- Releases from Tzaneen Dam are measured by gauge station B8H050 that operates on real time basis and the flow record is available.

- The gauging weir located at the Prieska (B8H017) operates on real time basis has been observed to indicate zero flows even when water is released through a tunnel back to the river downstream. These releases are not measured; hence the data from this gauging station does not accurately reflect the actual flows. It is also possible that the rating curve of the gauging weirs has been altered by silt deposition as evidenced in Figure 4.13. The lowest level of the weir is wide and as such low flow records are unlikely to be accurate.

Figure 4.13: A silted weir located at the Prieska

- The gauging weir located at the Ranch (Figure 4.14) is suitable for low flows but the flange of the weir does not extend up to the right bank of the river (as observed on the picture). This may affect the weir's ability to accurately measure high flows.

- The flow gauging station located at the Black Heron has a lot of missing data even though the temporal scale suggests a long record. The design of the station is suitable for measuring low and high flows, is free from silting and is environmentally friendly since there is a fish ladder that facilitates the movement of fish from downstream to upstream (Figure 4.15).

Figure 4.14: A flow gauging station located at the Ranch

Figure 4.15: A flow gauging station located at Black Heron

4.6.3 Evaporation data

The evaporation data available in the catchment is from pan evaporation records. The spatial distribution of class A-pans and class S-pans within the catchment is shown on Figure 4.16. The temporal distribution of the pan evaporation data is shown in Figure 4.17.

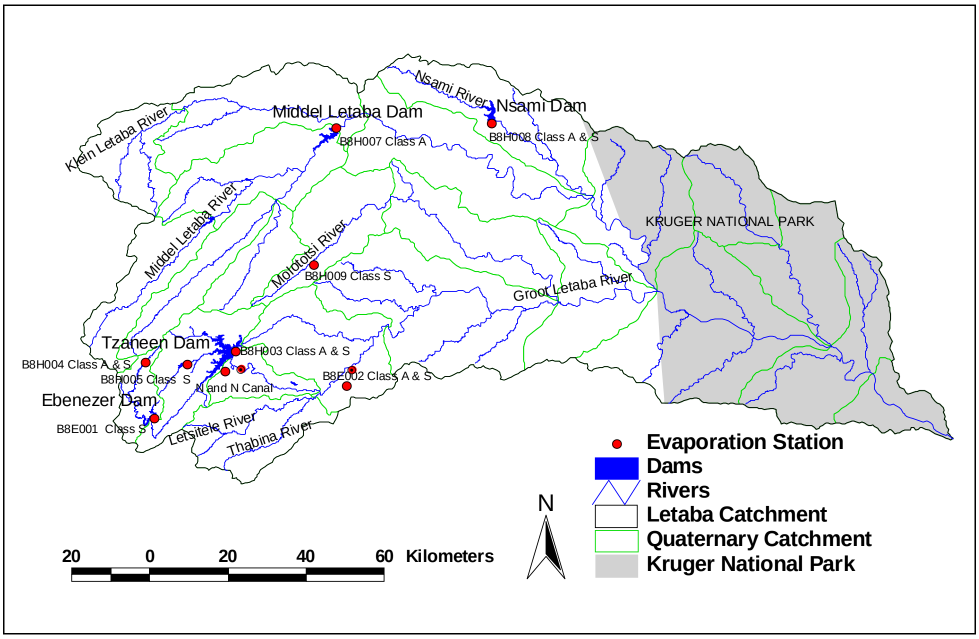


Figure 4.16: Spatial distribution of the evaporation pan

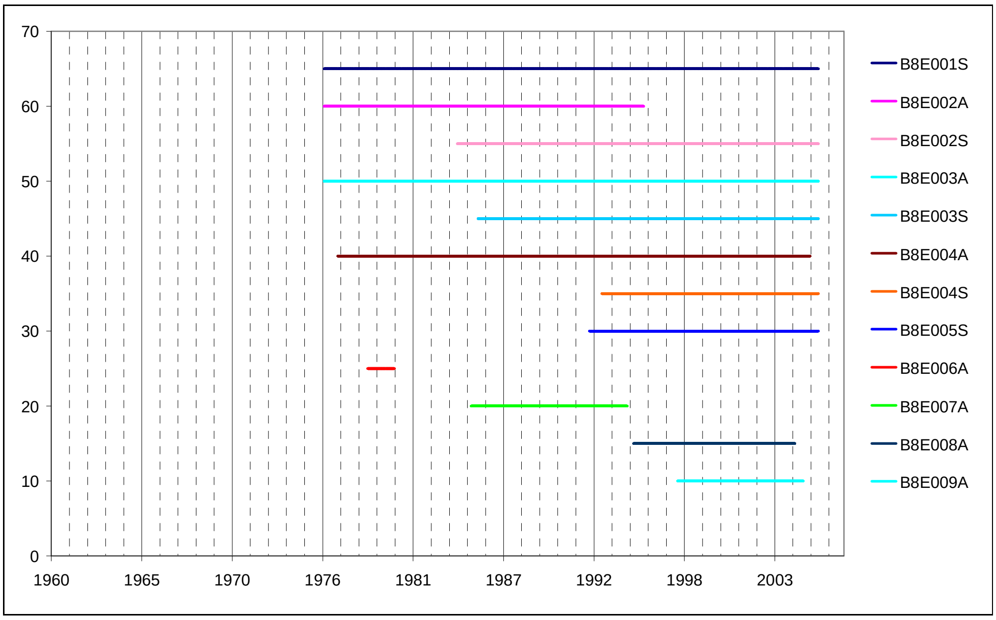


Figure 4.17: Temporal distribution of the pan evaporation records

Some stations (B8E002, B8E003 and B8E004) have both class A and S pans, while others have only one type of pan. A comparison of the evaporation pan records with the streamflow records reveals that a majority of pan evaporation records are shorter than streamflow records.

4.6.4 Groundwater

The groundwater monitoring network started in 2005 and several observation points were established in the catchment. The monitoring points are not close to Letaba River and the eastern part of the catchment does not have any groundwater monitoring points. Figure 4.18 shows the spatial distribution of the groundwater monitoring points. There are no missing records in the data sets.

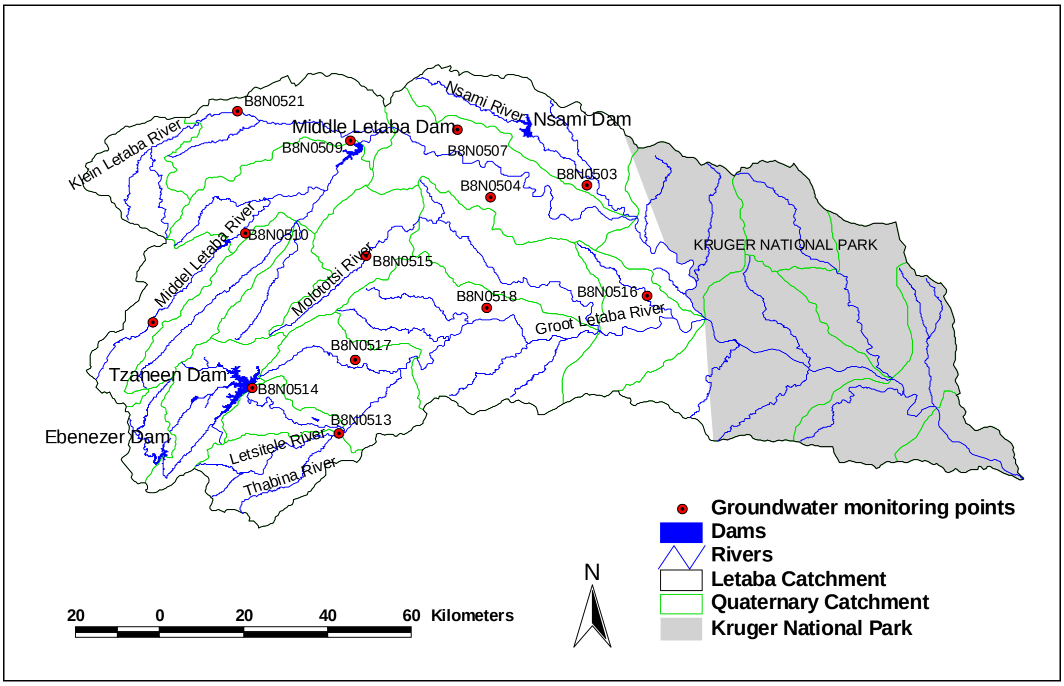


Figure 4.18: Spatial distribution of the groundwater monitoring points

4.6.5 Surface water abstraction

Water is abstracted from the river using either diversion canals or pumps. The available information starts for 2002 up to 2008 and is recorded and presented as monthly abstraction values. The monthly abstraction values are the sum of the scheduled and unscheduled abstractions made from the river. As stated in section 4.4, the unscheduled abstractions are done during the periods when the farmers, whose farm dams are not full, consider the amount of water available in the river to be in excess of the downstream users' requirements. For this study, the monthly values were disaggregated into daily values by dividing with the monthly values by the number of days.

4.7 HARD AND SOFT DATA

The information presented in this Chapter can be classified into two groups: hard and soft data. Hard data is the tangible (quantitative) information that is represented by numerical values. In this case, this would be the information given in Sections 4.3, 4.6.1, 4.6.2 and 4.6.3. On the other hand, soft data is the information that is intangible, more descriptive and cannot be expressed directly into quantitative values. In this chapter, all the data with the exception of the data discussed in sections 4.3, 4.61, 4.62 and 4.63, is soft data. The soft data is mainly obtained from the system operators and experimentalists (Seibert and McDonnell, 2002) and improves the understanding of the system particularly with respect to how the system works (Silberstein, 2006). Therefore, soft data is one of the resources that the modeler relies on when conceptualizing the hydrological processes and human activities within a system. The modeler can then rely on the hard data to verify the assumptions underlying the proposed conceptualization (Lyon et al., 2006). In support of the use of soft data in addition to hard data, Silberstein (2006) noted that it is better to be less right for the right reasons that to be right for the wrong reasons. For this reason, this study uses both the soft and hard data in the modelling process.

5 FUZZY INFERENCE MODELLING

5.1 FUZZY INFERENCE

Experts in various fields normally attempt to solve problems by drawing on the experience gained over several years. This experience informs model development in a number of aspects. For instance, the simplifying assumptions underlying the mathematical representation of hydrological systems are usually based on the hydrologist's understanding of the relative impact of the individual catchment processes that contribute to a hydrological response. Expert systems, much like the human experts, have been developed and contain a knowledge base – the experience in the form of rules/guidelines that can be used to analyze a problem in a manner similar to that of a human expert. The processes undertaken by such an expert system with respect to formulation of a solution to a particular problem are based on fuzzy logic and the process of determining the unknown relationships between variables in a given instance using what is already known about them in general is called inference. The fact that fuzzy logic can be used in this manner; having a knowledge base from which solutions representing expert judgment can be drawn, greatly increases the prospects for its application in various fields. This is what motivated its consideration for use in this study. A detailed discussion of fuzzy logic and fuzzy inference is therefore merited and as such is done in the following sections.

5.2 FUZZY THEORY AND DEFINITIONS

There are differences between classical set theory and fuzzy set theory even though fuzzy set theory is based, in part, on the classical set theory. This section defines the common and essential terms used in fuzzy logic literature.

5.2.1 Degree of membership

In classical set theory, sharp boundaries exist and elements can only belong to one set. Fuzzy set theory, on the other hand, allows for an element to belong to one or more sets. Fuzzy logic is based on the theory of sets in which the elements *u,* in a particular set A are characterised by a function. The function provides a means of determining whether or not a particular element belongs to a given set. Evaluation of the function value associated with a particular element gives a validity value called the degree of membership ranging from 0 and 1. If the degree of membership for a particular element is 0 with respect to some set, this indicates that the element does not belong to the particular set at all, while a value of 1 indicates that the element belongs to the particular set only. A single element can therefore belong to more than one set and it is the degree of membership that indicates the extent to which the particular element belongs to the particular set. The degree of membership for every element is defined through a generalized characteristic membership function given as (Kasabov, 1996):

$$\gamma_A(u): U \rightarrow [0,1] \quad 5.1$$

where U is the universal set and A is a subset. In cases where the universal set is discrete, a membership function can be defined by fine sets as:

$$A = \gamma(u_1)/u_1 + \gamma(u_2)/u_2 + \ldots + \gamma(u_n)/u_n \quad 5.2$$

where the / separates the degree of membership $\gamma(u_2)$ from the elements of the universal set $u_i \in U$ and the + stands for union.

In order to apply fuzzy logic in the solution of a problem, problem definition has to be done in fuzzy terms and Kasabov (1996) termed this process as conceptualization in fuzzy terms. In hydrology, the time-series of commonly encountered variables such as rainfall, temperature, abstraction and discharge are real numbers which in fuzzy language are called crisp values. In fuzzy terms, these crisp values can be expressed in linguistic terms that take on fuzzy labels or fuzzy propositions that have linguistic meanings. Examples of these fuzzy propositions are LOW, HIGH and MEDIUM and could include numerical tags (e.g. LOW FLOW ranges from 0.0 to 3.0 m^3/s). The true values of the fuzzy propositions are obtained from the membership function and the values have a greyness truth (neither true nor false). Linguistic variables can be quantitative (LOW, HIGH) or qualitative (CERTAINITY, TRUTH). Hence, fuzzy quantization is the process of representing variables as a set of linguistic values.

5.2.2 Representing fuzzy variables

The fuzzy variables can be quantized into subsets with fuzzy labels by using standard functions. This requires two parameters: (i) the number of fuzzy labels (subsets) and (ii) the form of the membership functions for each label to be defined before the quantization procedure is carried out. Several standard functions can be used to determine the degree membership; these include single valued function, triangular function, trapezoidal function, S-function, Z-function and the Gaussian function. Currently there are no guidelines for the selection of any of these functions, although the linear functions (trapezoidal and triangular function) have been used mostly in water related modelling due to the low computational intensity associated with their use (Hundecha et al., 2001).

5.2.3 Fuzzy clustering

A cluster can be defined as a bunch of items (e.g. data, materials) that possess similar properties or characteristics and are different from properties or characteristics of items in other clusters. Given that fuzzy set theory allows for an element to belong to one or more sets, it follows that an item can belong to more than one fuzzy cluster. Therefore, fuzzy clustering is considered as a process of grouping data into several potentially overlapping clusters and the quantifiable value that determines the extent to which an item belongs to a particular cluster is also called degree of membership.

5.3 TYPES OF FUZZY INFERENCE

The fuzzy inference procedure consists of the mapping of an input dataset onto an output dataset based on fuzzy rules or clustering techniques. Both techniques require the definition of two common parts, the antecedence and consequence part as described in the next subsections.

5.3.1 Rule-based fuzzy inference

The rule based fuzzy inference consists of several components that make up the antecedent and consequence parts. The input dataset, fuzzification, rule generation (fuzzy rule base) and mapping (fuzzy inference) constitute the antecedence part while the deffuzzification and output dataset constitute the consequence part. Figure 5.1 shows a schematic diagram of the fuzzy inference based on rules.

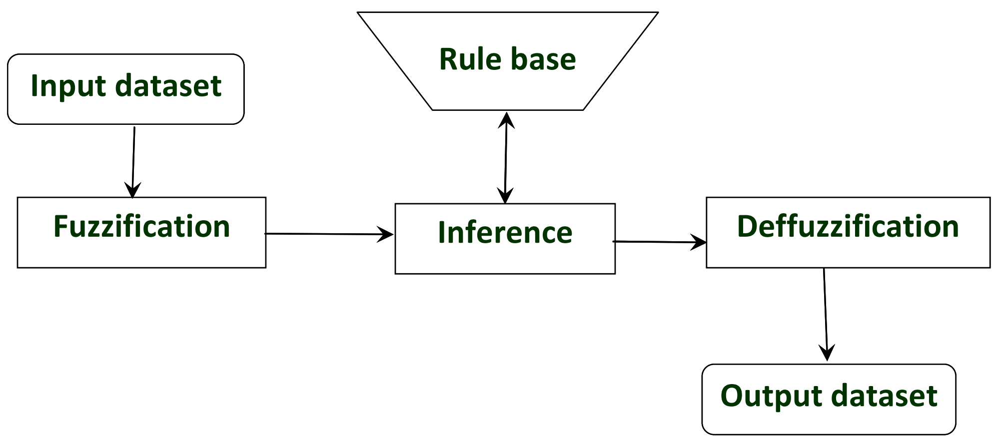


Figure 5.1: Typical flow chat for rule based fuzzy inference

The input dataset consists of either crisp or linguistic values. For typical hydrological problems, the datasets mainly consist of time series of observed data and linguistic values are not commonly used.

1) Fuzzification

Each subset (A_i) that exists is defined by a membership function that represents a linguistic variable (e.g. high, medium, low). Using the fuzzifier based on the membership function, the degree of membership for each input crisp value is determined with respect to the respective subsets. Depending on the degree of membership which takes on values in the range of [0, 1], the subset or subsets to which each crisp value belongs will be determined. For example, Figure 5.2 shows four triangular membership functions for rainfall. A crisp value of 12 mm of rainfall, based on the definition of the membership function in this particular example, can be seen to belong to two subsets A2 and A3 with the degrees of membership (α_{A1} and α_{A2}) to the respective subsets of 0.8 and 0.2. It is important to note that for a given crisp value, the sum of its degrees of memberships with respect to the various subsets should add up to 1.

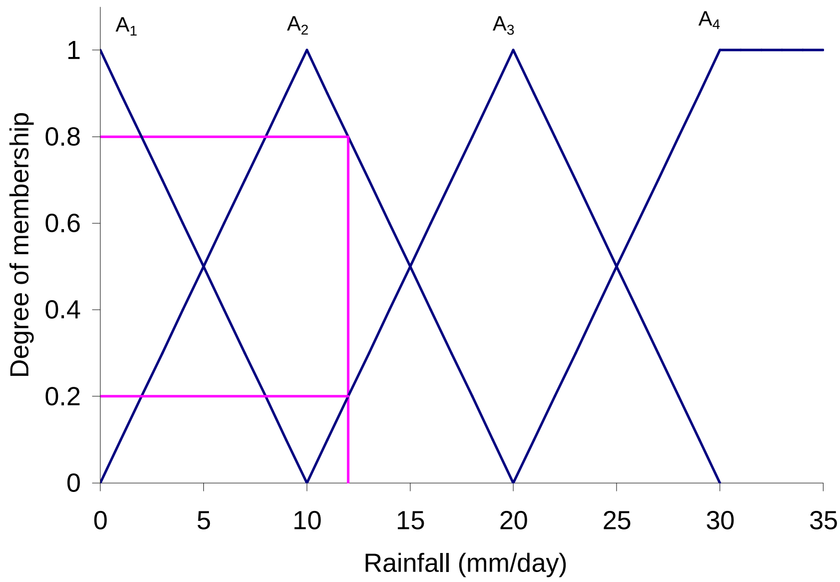


Figure 5.2: Triangular membership functions for rainfall

2) Rule base

Based on the subsets (A_j) and the input variables, rules are developed to cover the whole spectrum of the possible outcomes and they are stored in the fuzzy rule base. Typical fuzzy rules were introduced by Mamdani in 1974 (Mamdani and Assilian, 1975, Kasabov, 1996) and a detailed explanation is given below.

- *Mamdani's fuzzy rules*

The outcome (consequence) from the Mamdani's fuzzy rule is a fuzzy variable. In cases where a crisp value is required, a deffuzzification process has to be done as described in the next section. A typical Mamdani's fuzzy rule is of the following form:

R_1: **IF** x_1 is A_{11} ***AND*** x_2 is A_{12} ***AND*** x_2 is A_{12} ***AND… AND*** x_m is A_{1m} **THEN** y_1 *is* B_1

R_2: **IF** x_1 is A_{21} ***AND*** x_2 is A_{22} ***AND*** x_3 is A_{22} ***AND… AND*** x_m is A_{2m} **THEN** y_2 *is* B_2

…

R_i:**IF** x_j is A_{ij} ***AND*** x_j is A_{ij} ***AND*** x_j is A_{ij} ***AND… AND*** x_m is A_{nm} **THEN** y_i *is* B_i

…

R_n: **IF** x_1 is A_{n1} ***AND*** x_1 is A_{n2} ***AND*** x_1 is A_{n2} ***AND… AND*** x_m is A_{nm} **THEN** y_n *is* B_n

where the index *i,* (*1 to n*) represents the rules contained in the rule base.

As earlier stated, each fuzzy rule consists of two parts, the antecedent, and the consequence part and the word THEN is what demarcates the two parts. On the antecedent part of the rules, the input variables are connected by either an AND or an OR. The consequence part consists of the outputs. The connector AND represents the minimum value of degree of membership and OR represents the maximum value of the degree of membership that satisfies the use of the rule.

Further improvements in the definition or formulation of fuzzy rules have been developed including, incorporating degrees of confidence to them, gradual fuzzy rules, generalized production rules, recurrent fuzzy rules and the Takagi-Sugeno fuzzy rules (Takagi and Sugeno, 1985, Kasabov, 1996). Examples of rules incorporating these improvements are outlined below:-

- Fuzzy rules with confidence degrees

A fuzzy rule containing a confidence factor has the form:

R_i: **IF** x_j is A_{ij} ***AND*** x_j is A_{ij} ***AND*** x_j is A_{ij} ***AND… AND*** x_m is A_{nm} **THEN** y_i *is* B_i *(CF=0.9)*

where CF is the confidence factor of the validity of the conclusion.

- Gradual fuzzy rules

Gradual fuzzy rules are similar to the Mamdani's fuzzy rules (Kasabov, 1996) although they do not use fuzzy values for the variables in the rules. Instead, they use fuzzy representation of gradual properties. These rules have been applied mostly to modelling social, political and economic systems. When used, they significantly reduce the overall number of rules required to describe a problem while still spanning the whole range of input space.

- Generalized production rules

The generalized production rules consider the different inputs on the antecedence part of a rule to have differing importance (degree of importance), noise tolerance and also sensitivity factors. Hence, relative coefficients of importance, noise tolerance and sensitivity are introduced in the generalized production rules including the confidence factors.

- Recurrent fuzzy rules

Recurrent fuzzy rules are similar to the Mamdani's fuzzy rules as they contain two or more antecedence values and the consequence part. The main difference is that recurrent fuzzy rules use previous time step values from consequences either from the same rule or another rule to generate the current consequences. The general form of the recurrent fuzzy rule is given as:

R_i: **IF** x_j is A_{ij} ***AND*** x_j is A_{ij} ***AND*** x_j is A_{ij} ***AND… AND*** x_m is A_{nm} ***A****ND* $y_i(t-1)$ *is* B^i **THEN** $y_i(t)$ *is* B_i

Also, a set of recurrent fuzzy logic rules may include internal fuzzy variables (hidden variables) and can be shown as:

R_i: **IF** x_j is A_{ij} ***AND*** x_j is A_{ij} ***AND*** x_j is A_{ij} ***AND… AND*** x_m is A_{nm} ***AND*** $z_i(t-1)$ *is* C_i **THEN** $z_i(t)$ *is* C_i

R_i: **IF** x_j is A_{ij} ***AND*** x_j is A_{ij} ***AND*** x_j is A_{ij} ***AND… AND*** x_m is A_{nm} ***AND*** $z_i(t)$ *is* C_i **THEN** $y_i(t)$ *is* B_i

where the $z_i(t)$ is an internal variable and $y_i(t)$ is the output of the consequence variable.

- Takagi-Sugeno's fuzzy rules

Takagi and Sugeno (1985) introduced a type of fuzzy rules that have an antecedent part similar to Mamdani's rules but differing on the consequence part. The Takagi-

Sugeno's fuzzy rules use a function at the consequence part and can be represented as:

R_i: R_i: **IF** x_j is A_{ij} ***AND*** x_j is A_{ij} ***AND*** x_j is A_{ij} ***AND... AND*** x_m is A_{nm} **THEN** $(y_i=a_{i1}+a_{i2}x_1+a_{i3}x_2+...+a_{ij}x_m)$

where a_{ij} is the coefficient of the output function. The consequence function (y_i) is linear. The introduction of these rules made fuzzy inference more versatile in dealing with complex nonlinear relationships including mathematical analysis.

For a given system with two input variables and one fuzzy output variable, if the three variables (the two inputs and the single output variable) have three subsets then the possible number of rules is shown in Table 5.1. The number of rules increases exponentially with increase in the number of inputs and subsets (e.g. for *n* input variables and *m* subsets for each variable the total number of rules is n^m). Hence, rule-based fuzzy models are prone to rule explosion and this makes it difficult to specify the entire model from expert knowledge alone (See and Openshaw, 1999). However, it should be noted that when the fuzzy model encounters an input dataset, it is only those rules for which the input dataset matches the antecedent part that are fired/activated into the inference system.

Table 5.1: Possible number of rules for fuzzy system with three variables and each having three subsets.

Input 1	Input 2	Output 1	Input 1	Input 2	Output 1
Low	Low	Low	Medium	Medium	High
Low	Low	Medium	Medium	High	Low
Low	Low	High	Medium	High	Medium
Low	Medium	Low	Medium	High	High
Low	Medium	Medium	High	Low	Low
Low	Medium	High	High	Low	Medium
Low	High	Low	High	Low	High
Low	High	Medium	High	Medium	Low
Low	High	High	High	Medium	Medium
Medium	Low	Low	High	Medium	High
Medium	Low	Medium	High	High	Low
Medium	Low	High	High	High	Medium
Medium	Medium	Low	High	High	High
Medium	Medium	Medium			

3) Inference

In the rule-based fuzzy logic approach, inference is the process of mapping input the dataset onto the respective output. This is handled by the fuzzy inference engine - the central part of fuzzy inference. The inference engine picks only the suitable rules from the fuzzy rule base and learns how to map the input data space onto an output data space. This means that only those rules which match the input data space are fired each time the data is triggered. The input space is normally wider (more variables) than the output space and a chain of processes are executed while mapping an exact value at the consequence part. The mapping process framework based on theory of sets, provides an opportunity to incorporate uncertainty in the input datasets in a manner that is less limiting than those conventional methods which are based on statistical approaches (Kasabov, 1996). The inference system involves either multiple inputs and single output (MISO), single input single output (SISO) or multiple inputs and multiple outputs (MIMO).

4) Deffuzzification

Deffuzzification is the process of transforming the output fuzzy values into their respective crisp values.

5) Output data

The output data is the crisp (numerical) value that results from the contributions of the rules from deffuzzification.

6) Observations on rule based fuzzy inference

The observation made indicates that the rule based fuzzy inference requires many rules to completely define the entire span of the input data space (Katambara and Ndiritu, 2007) and this number of rules increases exponentially with the number of rules and subsets. Though the use of Takagi-Sugeno fuzzy inference can result in the complete definition of the input space using a lower number of rules than those required by Mamdani's fuzzy inference system, the resulting reduction is not very significant, particularly with respect to complex problems (Katambara and Ndiritu, 2009). For example, for an input data space of four input variables with each variable consisting of four subsets, 256 rules (4^4) are required. Although only few rules are fired at a time, the rules require storage space. Another drawback associated with rule-based fuzzy inference, is the subjectivity associated with the definition of subsets. If the definition of the subsets is done subjectively, this may lead to under-parameterization which might limit the models flexibility or over-parameterization which might increase the computation time which does not necessarily guarantee improved performance as was noted by Xiong et al (2001). Fuzzy cluster based inference was independently developed (Dunn, 1973, Chiu, 1994, Yager and Filev, 1994) and was found to have advantages over the rule-based fuzzy inference (Lohani et al., 2006).

5.3.2 Clustering based fuzzy inference

In contrast to rule based fuzzy system, cluster-based fuzzy inference approach, as the name suggests involves the grouping of the input datasets into clusters. A cluster is defined by its center and it consists of a set of data whose properties or characteristics are similar while differing from the properties or characteristics of the datasets in other clusters. Incorporating fuzziness in defining clusters allows input data to belong to more than one cluster. Cluster-based approaches have been applied in many fields such as organising, categorising and compressing data and even model building (Rao and Srinivas, 2006, Samhouri et al., 2009, Demirli et al., 2003, Zhao et al., 2009). Some modelling activities involve the use of either large datasets or inadequate or imprecise information. In such circumstances, an effective grouping technique of the data is required to enable the capturing of the concise representation of the behavioural characteristics of the system. Based on the selected criteria, the individual clusters identified are considered to be enough to support the fuzzy inference operations. Several fuzzy clustering approaches are currently available including the C-means clustering algorithm, the mountain clustering algorithm and the subtractive clustering algorithm. The details of these clustering approaches are given below.

1) Fuzzy C-means clustering algorithm

Fuzzy C-means (FCM) clustering relies on the understanding that one data point can belong to more than one cluster. FCM was first introduced in the 1930's and has since undergone several improvements (Bezdek, 1981, Bezdek et al., 1987). This method has been used in many fields including image recognition and data analysis.

The FCM clustering algorithm is based on the minimization of the objective function (Hathaway and Bezdek, 1986):

$$J = \sum_{K=1}^{n} \sum_{i=1}^{c} {\alpha_k^i}^m \left\| x_k - v_i \right\|^2 \tag{5.3}$$

where n is the number of data points to be clustered, c is the number of clusters, x_k is the k'th multi-dimensional data point, v_i is i'th *multi*-dimensional cluster centre,

α_k^i is the degree of membership of the *k*'th data in the *i*'th cluster and *m* is the weight factor greater that unity.

The membership degree is given as (Hathaway and Bezdek, 1986):

$$\alpha_k^i = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{(m-1)}}} \tag{5.4}$$

and the cluster centre is given as (Hathaway and Bezdek, 1986):

$$v_i = \frac{\sum_{i=1}^{n} \alpha_k^{i\,m} . x_i}{\sum_{i=1}^{n} \alpha_k^{i\,m}} \tag{5.5}$$

The termination criterion for the iteration is when the maximum difference of the degree of membership between two successive iterations is less than certain predefined value. In order to use this technique, initialization of the cluster centres needs to be done. Initialization of the cluster centres presupposes the existence of prior knowledge of the input datasets if an appropriate selection of the cluster centres is to be done and yet this prior knowledge may not always exist or be sufficient.

2) Mountain clustering method

The mountain clustering algorithm was introduced by Yager and Filev (1994). This method uses the following three processes:

(1) Division of the data space into grid points distributed either evenly or unevenly across the data space. The intersection of these grids constitutes the class centres.

(2) Construction of the mountain functions at every grid point. The mountain function (membership function) represents the density of data at grid point. The value of the mountain function at a point is given as (Yager and Filev, 1994):

$$m(v)_j = \sum_{i=1}^{n} \exp\left(- \frac{\|v - x_i\|}{2\sigma^2} \right) \qquad 5.6$$

where x_i is the *i* data point and σ is an application specific constant and v is the grid point.

(3) In order to select the next cluster center, the effect of the prior center is eliminated by subtracting a scaled Gaussian function from center at c_j given as (Yager and Filev, 1994):

$$m(v)_{j+1} = m(v)_j - m(c) \exp\left(- \frac{\|v - c_j\|}{2\beta^2} \right) \qquad 5.7$$

This process continues until a sufficient number of clusters are identified. The performance of mountain clustering is affected by the dimension of the problem.

Furthermore, the number of computations required increases exponentially with the dimension of input data since the mountain function has to be evaluated at each grid point in the data space.

3) Subtractive clustering method

The subtractive clustering method, introduced by Chiu (1994), can be considered as similar to mountain clustering; with the difference arising from the use of a density (potential) function which is computed at each data point instead of a grid point. Hence, each data point is considered to be a potential candidate for the cluster center. The amount of computation required is significantly reduced and is a function of the amount of data used; not the dimension of the data set. It has been applied in several fields including job sequencing (Demirli et al., 2003) extraction of fuzzy rules for nonlinear system modelling (Eftekhari and Katebi, 2008) and fast training of radial basis function neural networks (Sarimveis et al., 2003).

The potential of each data point is estimated by (Chiu, 1994):

$$P_i = \sum_{j=1}^{n}\sum_{k=1}^{q}\exp\left(-\frac{4}{r_a^2}\left\|x_{i,k} - x_{j,k}\right\|^2\right) \qquad 5.8$$

where P_i is the potential of i^{th} data point, whose center is at $x_{i,k}$, j is the counter for cluster center and r_a is a positive constant defining the neighbourhood range of the cluster. The cluster center with the highest potential is selected as the first cluster center x_{c1}.

The next cluster center is obtained as the point with the highest potential after penalising the previous cluster center and points in the neighbourhood. The expression is given as (Chiu, 1994):

$$P_{i+1} = P_i - P_{c1} \exp\left(-\frac{4}{(\eta r_a)^2} \left\| x_{i,k} - x_{j,k} \right\|^2 \right) \qquad 5.9$$

where η is the quash factor which is set to a value greater than 1 to prevent obtaining closely spaced clusters. The suggested values for η and r_a are $1.25 \leq \eta \geq 1.5$ and $0.15 \leq r_a \geq 0.30$ (Demirli et al., 2003). The obtained cluster centre is checked for the minimum distance given as (Chiu, 1994):

$$d_{min} / r_a + P_i / P_{c1} \geq 1 \qquad 5.10$$

where d_{min} is the minimum distance between the computed centre with other centres. If the cluster centre does not fulfil the above condition, its potential is set to zero and the data point with the next highest potential P_i is selected as the new possible cluster centre. This data point is also checked for the same condition (equation 5.10). Clustering ends when the following condition is fulfilled (Chiu, 1994):

$$P_i < \varepsilon P_{ci} \qquad 5.11$$

where ε is the rejection ratio.

The identified cluster centers are considered to have the potential to represent the system's behaviour and the contribution that each cluster has to the inferred output is determined by its degree of membership. The advantages of using this technique include a significantly reduced number of rules required that automatically reduce the computation time and storage space required. It should be noted that unlike rule-based fuzzy, in subtractive cluster-based fuzzy inference, all the rules are fired every time a computation is performed.

A comparison of the clustering techniques is given in Table 5.2.

Table 5.2: Comparison of clustering techniques

	Fuzzy C-means clustering	**Mountain clustering**	**Subtractive clustering**
Prior knowledge	Requires prior knowledge to locate cluster centres	Does not require prior knowledge to locate cluster centres	Does not require prior knowledge to locate cluster centres
computation	Fair and dependent on the number of cluster centres identified	Intensive computation increases that with the dimension of the data	Less than mountain clustering and increases with the number of data points
Memory storage	Fair	Requires large memory to store information for each grid	Required memory is less than for mountain clustering.
Remarks	The use of the approach in absence of prior knowledge is a challenge.	Although it does not require prior knowledge of the cluster centres, it may not be suitable in cases where the dimension of the data is large	Suitable for the modelling since the computation increases with the amount of data available and no prior knowledge is required.

6 APPLICATION OF FUZZY INFERENCE MODELLING TO LETABA RIVER

6.1 ADAPTING FUZZY INFERENCE MODELLING

However complex a model may be, it is still a simplified representation of the catchment processes of systems that are usually complex (Abbott et al., 1986). In a river system like the Letaba River system with highly developed water infrastructure and inadequate water resource use data, the application of black box (data based) modelling approaches capable of inferring the expected flow characteristics at various points of interest from the input data offer an alternative to the more traditional models described in Section 2.2 or similar ones.

In semi-arid regions, the spatial and temporal variability of the catchment processes is high with the consequence that satisfactory modelling of the processes in such catchments is challenging. Moreover, modelling is made more challenging if there are concerns with respect to the reliability of the measured information such as those from the flow gauging stations. A consideration of these challenges and the spatial variability of the physical characteristics of such catchments suggest that effective modelling can be achieved if the system is divided into units so as to reduce variability within a modelling unit. Such division also needs to consider the locations of data measurement in order to enable verification of model performance for each unit. On this basis, the river system has been demarcated into three river reaches with the first reach covering the section from Tzaneen Dam to the Junction weir, the second reach covering the section Junction to the Ranch and the third section covering the Ranch to Black Heron (Figure 6.1). The first and second river reaches are characterised by the existence of storage weirs, and human activities such as the operation of storage weirs and water abstractions, while the third river reach is characterised by the existence of an alluvial aquifer and insignificant human activities.

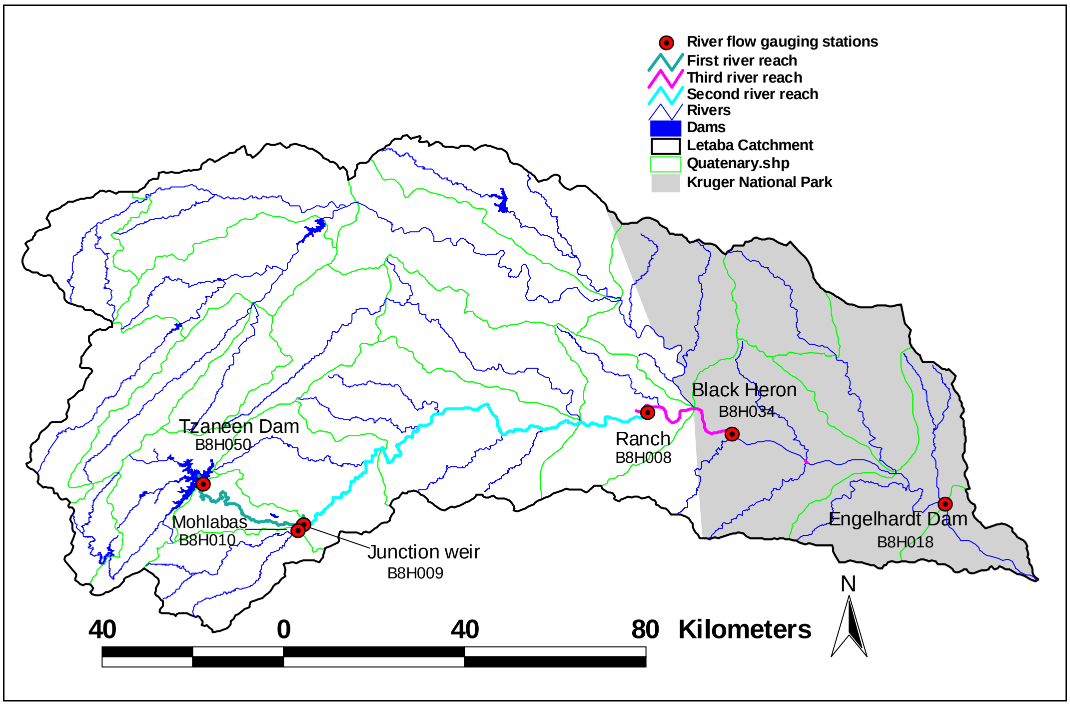


Figure 6.1: Demarcation of the river reaches

Of the several clustering techniques discussed in section 5.3, the subtractive clustering algorithm was noted to have several advantages over the other algorithms as indicated in Table 5.2 and the Takagi-Sugeno inference approach observed to offer more advantages than the Mamdani fuzzy inference approach

(Section 5.3). For this reason, the subtractive clustering algorithm and Takagi-Sugeno inference approach were selected for application as they offered the best prospects for efficiently carrying out the computational analysis required for this study.

6.2 METHODOLOGY OF THE FUZZY INFERENCE MODELLING

6.2.1 Model development

During model setup, the flow measured flowing out of the upper (first) river reach (see Figure 6.1) was considered as inflow into the adjacent (second) downstream river reach. In the same vein, inflow into the third rich reach was the flow observed at the downstream end of the second river reach. The model, operating at a daily time step, was designed to estimate the daily streamflow ($q_{sim,i}$) at the downstream end of the respective river reach given the rainfall (R_i), upstream inflow ($q_{in,i}$), the lagged moving average evaporation ($\hat{e}_i$) (Section 6.2.5), abstraction ($q_{abs,i}$) along the reach and contribution from Letsitele tributary ($q_{tri,i}$) to the flow in the second river reach. The other consideration is the rule of thumb based operation of the dam or storage weirs (Section 4.4), and is based on the inspection of the depth of flow (current flow) observed at selected points on the river. The impact of such a decision is only evident the next day (next day's flow). Therefore the current day's inflows from the upper river reach $q_{in,t}$, and the one day lagged outflows $q_{obs,t-1}$ have been considered to encapsulate rule of thumb based operation. Including the one day lagged outflows $q_{obs,t-1}$ however makes the model most applicable to forecasting applications. The impact of these human activities is limited only to the first and second river reaches. For the third river reach, there are insignificant human induced processes, however the existence of the alluvial aquifer tends to delay the flow as such merited foe the use of previous time step flows. In addition, this previous time step is meant to incorporate the human induced processes; the operation of the storage weirs whose response is mostly noticeable at the next flow gauging station downstream after a day. Also, several modelling attempts indicated some improvement when one day lagged rainfall in used and it is likely that the storage weirs and alluvial aquifer retains the runoff for a day. Although more comprehensive input data optimization was possible, it was considered better to allow the use of inputs that will allow for a fairer comparison of the 3 modelling approaches. Moreover, it was possible to include rainfall with additional lags, the inclusion of the two rainfall inputs in addition to the rainfall transformation described in Section 6.2.5 was considered adequate. The inclusion of rainfall inputs with longer lags could however be the subject of further studies. Thus, the model was of the form:

$$q_{sim,t} = f\left(q_{in,t}, q_{in,t-1}, R_i, R_{t-1}, \hat{e}_t, q_{abs,t}, q_{out,t-1}\right) \qquad 6.1a$$

$$q_{sim,t} = f\left(q_{in,t}, q_{in,t-1}, R_t, R_{t-1}, \hat{e}_t, q_{abs,t}, q_{tri,t}, q_{tri,t-1}, q_{out,t-1}\right) \qquad 6.1b$$

$$q_{sim,t} = f\left(q_{in,t}, q_{in,t-1}, R_t, R_{t-1}, \hat{e}_t, q_{out,t-1}\right) \qquad 6.1c$$

for the first, second and the third river reach respectively.

Xiong et al. (2001) applied the Takagi-Sugeno fuzzy inference to combine results from five models and noted that the model can obtain satisfactory results if it is not over- or under-parameterised but no guidelines were given on how to achieve an optimal number of parameters. In this study, the optimal number of parameters has been attained iteratively by gradually increasing the number of cluster centres and monitoring the fuzzy inference model performance. As the number of cluster was increased, it reached a level where there was insignificant improvement in the model performance. Several trials runs were done, out these, the optimum number of cluster centres obtained for the river reaches were 10, 10 and 5 for the first, second and third river reach respectively. For uniformity, 10 clusters were used for all the river reaches. The data that was used for model calibration and verification was data observed in the period starting March 2002 to April 2008.

The input variables in equation 6.1 can be denoted by a general variable z as:

$$\{z_{t,1}, z_{t,2},....z_{t,k}.....z_{t,q}\} = \{q_{in,t}, q_{in,t-1}, R_t, \hat{e}_t,...., q_{out,t-1}\} \qquad 6.2$$

where $z_{i,k}$ is the k^{th} input variable, k (1 to q) is the counter for the dimension of the dataset, q is the number of input variables, t (1 to n) is the counter of the input data points and n is the number (days) of data points. The data points are normalized to range between 0 and 1 using the function:

$$x_{t,k} = \frac{(z_{t,k} - z_{\min,k})}{(z_{\max,k} - z_{\min,k})} \qquad 6.3$$

where $z_{min,k}$ and $z_{max,k}$ is the minimum and maximum value of the k^{th} variable.

The potential of each data point was obtained using equation 5.8 and the data point with the highest potential was selected as the first cluster center. The subsequent potentials were obtained using equation 5.9. While in this study the values for η and r_a were subjectively set to 1.25 and 0.15 respectively, preliminary analysis was done as an attempt to identify the impact of varying the two parameters and has been reported in Section 6.2.5.

After locating the cluster centers, a Gaussian function was used to determine the degree of membership ($DOM_{i,m}$ ($m=1$ to NC)) of every input data point using equation 6.4 (Chiu, 1994):

$$DOM_{t,m} = \exp\left(-\frac{4}{r_a} \sum_{k=1}^{q} \left(x_{t,k} - c_{m,k} \right) \right) \qquad 6.4$$

where $c_{m,k}$ is the cluster center. The sum of the degrees of membership of any given point for all the cluster centers was then obtained as:

$$DOMSUM_t = \sum_{m=1}^{NC} DOM_{t,j} \qquad 6.5$$

where NC is the number of clusters in consideration.

Each cluster centre is associated with a function of the form (Chiu, 1994):

$$y_{t,j} = a_{0,j} + a_{1,j}x_{1,t} + \ldots + a_{q,j}x_{q,t} \qquad 6.6$$

where $a_{0,j}$, $a_{1,j}$..., $a_{q,j}$ are the coefficients. It is these coefficients, based on their values, which determine the relationship between the input datasets including the rainfall and the respective magnitude of simulated streamflow. The simulated streamflow values, $q_{sim,i}$ are determined by using the weighted average method as follows (Chiu, 1994):

$$q_{sim,t} = \sum_{m=1}^{NC}\left(y_{t,m} \times \frac{DOM_{t,m}}{DOMSUM_t} \right) \qquad 6.7$$

The simulated and the observed streamflow series are used to obtain the root mean square error (RMSE) given as:

$$RMSE = \sqrt{\frac{\left(q_{obs,t} - q_{sim,t}\right)^2}{n}} \qquad 6.8$$

where $q_{obs,i}$ is the observed flows and is equal to $q_{out,i}$. Calibration of the model is therefore effected by minimizing the value of the RMSE by varying the coefficients in equation 6.6. The shuffled complex evolution algorithm (SCE-UA) (Duan et al., 1992) was used to calibrate the model. Table 6.1 shows the search that was used during the calibration process.

Table 6.1: Typical parameter search range

	a_o	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
Upper	1	1	1	1	1	1	1	1	1
Lower	-1	-1	-1	-1	-1	-1	-1	-1	-1

6.2.2 Model implementation

The fuzzy inference model discussed in Section 6.2.1 above, the conceptual model discussed in Chapter 7 including the hybrid conceptual-fuzzy model discussed in Chapter 8 have been developed and coded by the author. The programming language used is Delphi® 2006 available in the Borland Developer Studio 2006 meant for educational non-commercial use.

6.2.3 Statistical model performance measures

Several model evaluation techniques exist and some studies including (Freer et al., 2004, Moriasi et al., 2007) have noted that different performance measures may be required to adequately assess model performance. This study applied four performance measures: the Nash-Sutcliffe efficiency (NSE), correlation coefficient (CCoef), percent bias (PBIAS), and the root mean square (RSR). The Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) is a normalized statistic that compares the relative magnitude of the residual variance to the observed variance. By doing so, it gives an indication of how well the observed and simulated series fits to 1:1 line and is mathematically defined as:

$$NSE = 1 - \left[\sum_{i=1}^{n} \left(q_{obs_i} - q_{sim\ i} \right)^2 / \sum_{i=1}^{n} \left(q_{obs_i} - \overline{q}_{obs_i} \right)^2 \right] \qquad 6.9$$

where $q_{\text{obs i}}$ is the i^{th} observed flows and, $q_{\text{sim i}}$ is the i^{th} simulated flows, $\overline{q}_{obs,i}$ is the mean of the observed flows and *n* is the total number of observations in consideration.

The second measure, the correlation coefficient (*CCoef*) is based on the Pearson product moment correlation coefficient of the simulated and the observed flow series and is obtained as:

$$CCoef = \left[\sum_{i=1}^{n} \left(q_{obs_i} - \overline{q}_{obs\ i} \right)\left(q_{sim_i} - \overline{q}_{sim_i} \right) / \sqrt{ \sum_{i=1}^{n} \left(q_{obs_i} - \overline{q}_{obs\ i} \right)^2 \sum_{i=1}^{n} \left(q_{sim_i} - \overline{q}_{sim_i} \right)^2 } \right]$$

6.10

where $\overline{q}_{sim}$ is the mean of the simulated values.

PBIAS is the third measure and it measures the averaged tendency of the simulated series to be larger or smaller than their observed series. Positive and negative values give an indication of whether the model underestimates bias and overestimates bias respectively. PBIAS (Moriasi et al., 2007) is obtained as:

$$PBIAS = \left[\sum_{i=1}^{n} \left(q_{obs_i} - q_{sim\ i} \right) * 100 / \sum_{i=1}^{n} \left(q_{obs_i} \right) \right] \qquad 6.11$$

The forth measure used is the root mean square error observed standard deviation ratio (RSR) and it incorporates the benefits of error index statistics and includes a scaling factor of the standard deviation (Moriasi et al., 2007) of the observed series. The value varies from zero to a large positive value where zero indicates an optimal value. The RSR is obtained as:

$$RSR = \left[\left[\sqrt{\sum_{i=1}^{n}\left(q_{obs\,i} - q_{sim\,i}\right)^2}\right] \Big/ \left[\sqrt{\sum_{i=1}^{n}\left(q_{obs\,i} - \overline{q}_{obs\,i}\right)^2}\right]\right] \qquad 6.12$$

Moriasi et al. (2007) recognized that watershed models are powerful tools for management purposes as they are capable of simulating the processes associated with the water movement. Noting the absence of a comprehensive guide for model evaluation, Moriasi et al. (2007) undertook a study to establish several model performance ratings. These values attempt to measure the accuracy of simulated data compared to measured flow using statistics. The study recommended that for streamflow simulations, a model could be judged as satisfactory if the Nash-Sutcliff efficiency was greater than 0.5, the root mean square error observed standard deviation ratio (RSR) was less than 0.7 and the absolute value of the percent bias (PBIAS) was less than 25%. Details of the performance ratings are given in Table 6.2 and have been adopted for use in this study.

Table 6.2: Performance rating (Moriasi et al., 2007)

Performance Rating	RSR	NSE	PBIAS%
Very good	0 ≤ RSR ≤ 0.5	0.75 < NSE ≤ 1.0	PBIAS < ± 10
Good	0.5 < RSR ≤ 0.6	0.65 < NSE ≤ 0.75	10 < RSR ≤ 15
Satisfactory	0 < RSR ≤ 0.7	0.55 < NSE ≤ 0.65	15 < RSR ≤ 25
Unsatisfactory	RSR > 0.7	NSE < 0.50	RSR > 25

6.2.4 Fuzzy inference model setup

It has been stated in Section 6.1 that the Letaba River has been demarcated into three river reaches for the purpose of accounting for spatial and temporal variations. For each river reach, a model was set up to simulate flow observed at the downstream end of the respective reach. In addition, the model setup for simulation of flow observed at the downstream end of the second river reach was run using, as an input, the output of the model setup to simulate flow in the first river reach. This procedure of using the output of simulated flow values of the first river reach as an input of the model for the second river reach is referred to hereafter as a linked model for the first and second river reach. In a similar manner, a linked model for the first, second and third river reaches was setup such that simulated outflows from the first and the second river reaches were used as inputs of the models of the adjacent river reaches downstream. This was meant to allow for the evaluation of the model's performance with respect to the individual river reaches as well as at an integrated level.

6.2.5 Correlation analysis of input and output variables

Figure 6.2 shows the correlation between the elements of the input dataset without any transformation. The correlation between the one day lag rainfall data (R_{t-1}) and the outflow ($q_{out,t}$), with values found to be 0.48, 0.32 and 0.25 for the first, second and third river reach respectively (Figure 6.2) is observed to be slightly higher than the correlation between rainfall data (R_t) and the outflow ($q_{out,t}$) for which the values of 0.12, 0.15 and 0.12 corresponding to the first, second and third river reach respectively, are obtained. The difference in the values suggests that the storage weirs and alluvial aquifer retain the flow for one day. A similar observation was made for the inflow into second river reach including the contribution from the Letsitele River, the only gauged tributary that flows into the second river reach. With correlation values of 0.46 and 0.53, the one day lag ($q_{tri,t-1}$) of the inflows were observed to be better correlated to the outflow ($q_{out,t}$) than the inflows without a one day lag whose corresponding values were 0.45 and 0.47 (Figure 6.5). The insignificant human activities as well as the non-existence of the storage weirs in the third river reach suggests that the presence of an alluvial aquifer may impacted flows and as such is responsible for this lag. Therefore, the existence of the storage weirs at various locations along the river and the presence of the underlying alluvial aquifer significantly impact the flows in the Letaba River.

The abstraction data that was obtained for this study was recorded as a monthly abstraction totals. To obtain the daily values used in the study, the monthly totals were divided by the number of days in the month. Considering the derived daily abstraction values and daily evaporation data, their influence on the flows was observed to be less significant compared to that of the rainfall and inflows (Figure 6.2). There is loss of information associated with the process of aggregating daily values into monthly values, therefore, the derived daily abstraction values do not contain enough information to be strongly correlated with the outflows. The poor correlation values between evaporation and outflow for all the three reaches do not manifest the effect of evaporation at a daily time step. It is certain that at a longer time scale (monthly or annual), higher correlations can be expected.

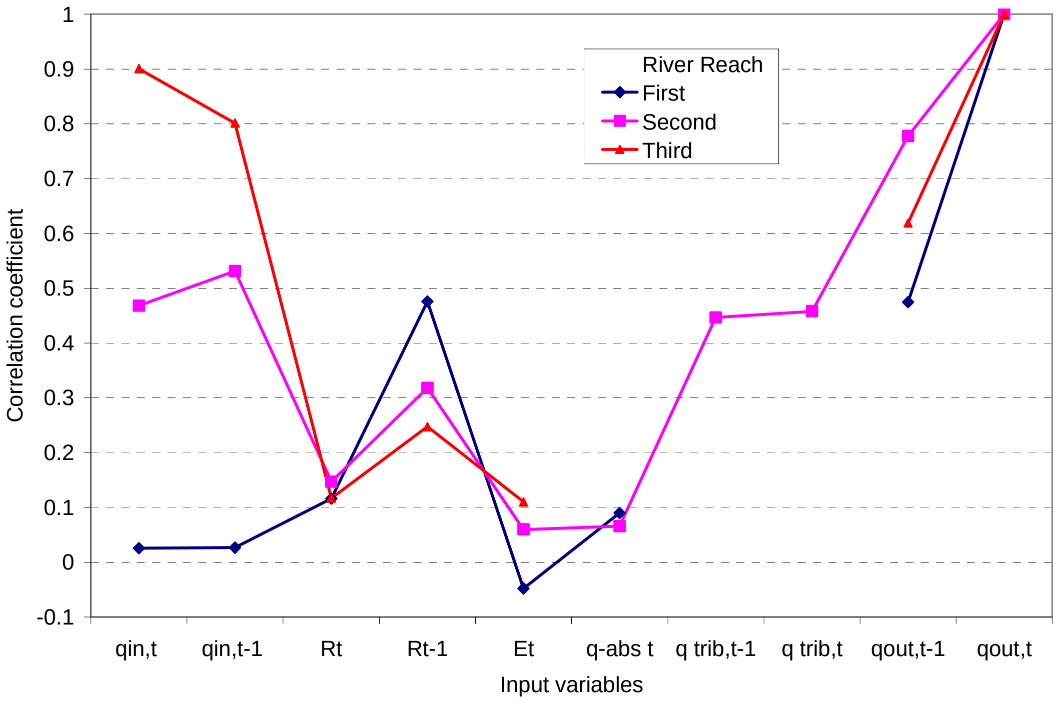


Figure 6.2: Correlation coefficients between input variables and the observed outflow

6.2.6 Preliminary fuzzy modelling tests and improvements

(a) Influence of storage weirs

The existence of storage weirs along the river tends to impound some of the flows thus reducing the amount of flow observed downstream of the weirs. In a situation where an alluvial aquifer also exists, the effects of the alluvial materials are significant, particularly at the start and end of the rainy season. Delays in the onset of channel flows downstream are one of the major effects of the presence of an alluvial aquifer (Newman et al., 2006, Boroto and Gorgens, 2003). As such, the effects of the storage weirs and the alluvial material were implicitly considered in the modelling. Considering the distance between the Ranch weir (B8H008) and the closest upstream storage weir is considerably large as shown in Figure 4.2, the influence that the alluvial materials may have a more significant impact on the flow than the storage weir and therefore needs to be considered. Figure 6.3 illustrates what happens during a rainfall event for a river section with a storage weir adjacent to a flow gauging weir.

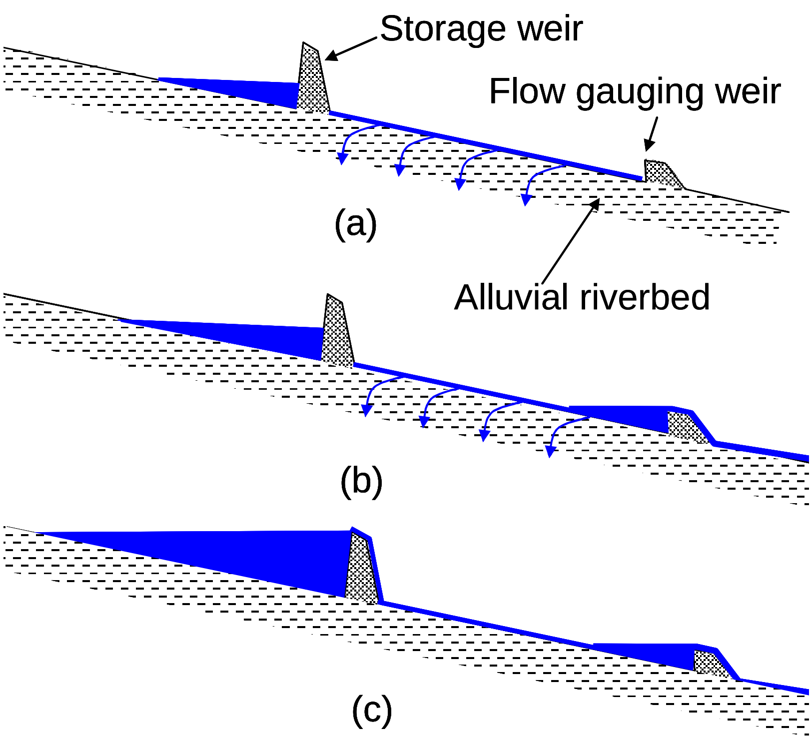


Figure 6.3: Influence of the storage weirs and alluvial material

When an amount of rainfall is received, if the volume generated between the storage weir and flow gauging weirs is not enough to flow over the gauging weir, no response will be noticed downstream of the gauging station. This phenomenon is represented by Figure 6.3 (a) and it shows that all the runoff generated in this case contributes to the storages of the weirs and the alluvial aquifer. When the amount of rainfall received, results in a flow volume that is enough to fill the gauging weir storage with some flowing over but does not fill and flow over the storage weir then some response is observed downstream of the gauging weir although the response is small. In this instance, the streamflow response is from the catchment area between the storage weir and the flow gauging weir and is represented by Figure 6.3 (b). However, in the instance when the rainfall is more and the volume of runoff generated exceeds the storage requirements of both weirs and the alluvial aquifer then a significant response in flow is observed. This phenomenon in which the weirs and the alluvial aquifer are saturated is represented by Figure 6.3 (c). This description of the impact of the storage weirs and the alluvial aquifer was found to be compatible with the observed relationship between measured streamflow and rainfall (Figure 6.4) and resulted in the decision to transform the observed rainfall based on the sort of response it was expected to generate for the fuzzy modelling. This transformation allowed for the influence of the storage weirs and alluvial aquifer to be implicitly incorporated in the modelling process. Prior to several trial runs done using the transformed rainfall values obtained by Equation 6.19, the fuzzy modelling obtained much poorer results without this transformation.

$$R_i = \begin{cases} 0 & Rain_i < 15mm \\ Rain_i / 10 & 15 \le Rain_i < 28mm \\ Rain_i & Rain_i \ge 28mm \end{cases} \qquad 6.19$$

where R_i is the modified rainfall series and $Rain_i$ is the observed rainfall series.

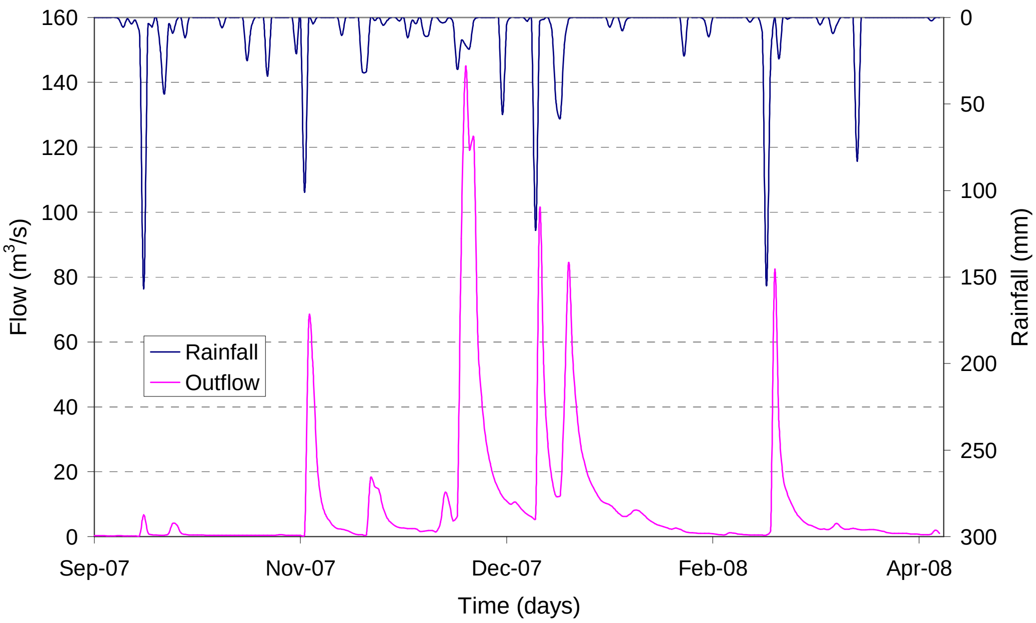


Figure 6.4: Stream flow response due to rainfall received

This transformation has been applied to the first and second river reaches since it is in these reaches that storage weirs exist.

(b) Pan evaporation data

The influence that the pan evaporation data has on the flow was noted during the trial runs made. It was therefore found necessary to use a 6-days delayed moving average so as improve the simulation particularly with respect to the low flows. Several trial runs with 2-, 3-, 4-, 5-, and 6-days delayed moving average pan evaporation series were made. Those moving average series that represented an average evaporation of more days were observed to result in better performance than those series whose values represented fewer days, therefore, those series of 6-day moving average was selected for use.

(c) Location of the cluster centers

Considering how the cluster-based fuzzy model functions, it was considered possible that different cluster centres may be more closely linked to the different processes being modelled in the river system. A simple way to infer into this was the location of the identified cluster centers in the dataset. This was done by obtaining the percent rank of the cluster centers in the data space. The second river reach was selected for this as its characteristics were regarded to be the most representative of the three reaches. The plot for the ranks is shown in Figure 6.5. While an average percent rank of 0.49 was observed for the inflow, three clusters were observed to have values less than the average for the inflow into the second river reach for both with ($q_{in,t}$), and without a day lag ($q_{in,t-1}$). Half of the ten clusters indicated that the outflow values are above the average percent rank of 0.39. Majority of the values obtained for the cluster centers for rainfall are zero, with only one cluster center being greater than zero. All the values of the cluster center locations for evaporation data and derived abstraction values have been found to be relatively high. For evaporation data, a minimum rank of 0.44 and maximum of 0.59 was observed, while a smaller range for abstraction estimates with a minimum value of 0.61 and a maximum value of 0.68 was found. Considering the contribution from the Letsitele River, a wider range of the rank has been observed with a minimum of 0.09, maximum of 0.8 and an average of 0.27. From Figure 6.5 there is no evidence linking any cluster centres to particular processes more than any other.

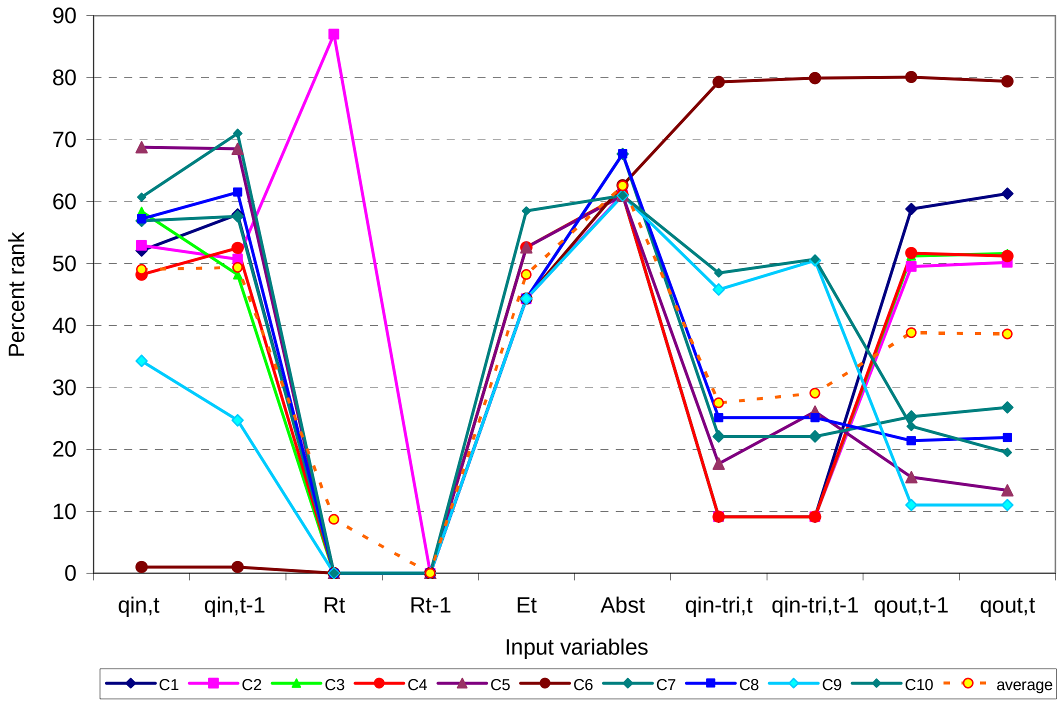


Figure 6.5: Rank for the cluster center of each input variable for the second river reach.

(d) Influence of radius of the hypersphere and quash factor on the model coefficients

It was perceived that the variation of the radius of the hypersphere and quash factor may have an influence on the model performance in simulating particular characteristics of the hydrograph. Therefore, the variation of these values was analysed and Figure 6.6 shows the average values of the coefficients for ten clusters for each value of radius of the hypersphere (Equations 5.8 and 5.9). It was observed that the variation of the coefficient values did not correspond with the influence of the radius of hypersphere on the model coefficients and the degree of membership. This was done for the second river reach as it was assumed to be a representative of the river reaches. No discernable trend of the variation in radius with the average values of model coefficients was observed and no relationship could be detected between the quash factor values and the average value of the coefficients.

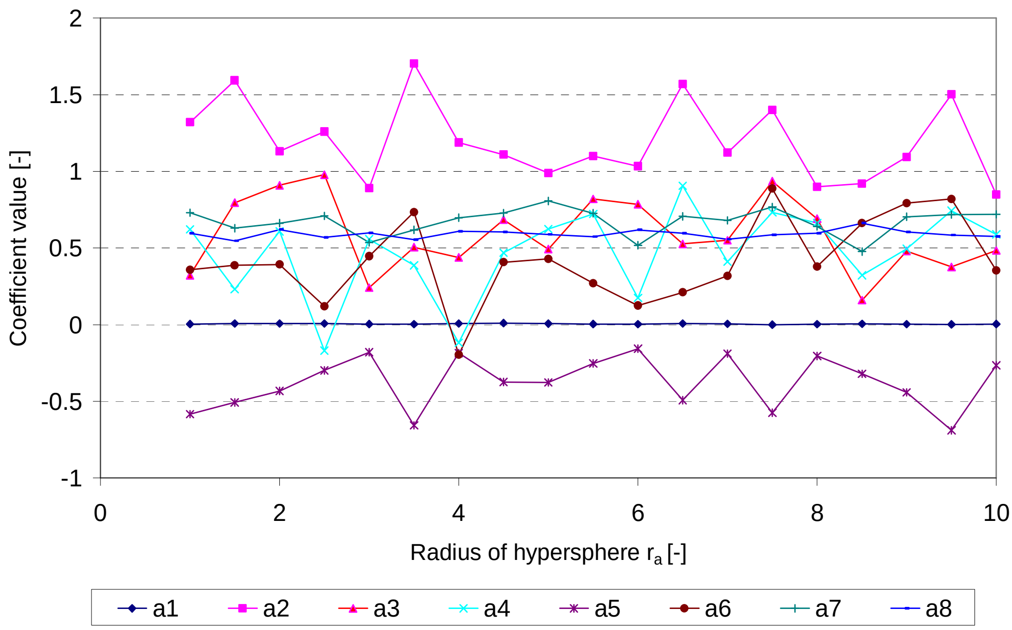


Figure 6.6: Variation of model coefficients with radius of hypersphere r_a for the second river reach.

(e) Influence degree of membership on the simulated flows

The values obtained using Equation 6.6 were all used in Equation 6.7 based on the ratio of degree of membership to the sum of the degrees of memberships. This ratio is referred to as weight. The possibility that some of the weights favour particular component of the simulated hydrograph was investigated. Figure 6.7 shows the variation of the weights with the flows. The weights do not significantly vary among the clusters. However, the variation of the weights with the simulated flow is more pronounced in cases where there is a sudden increase in flow corresponding to a uniform reduction in the degrees of membership for all the clusters. Therefore, the nonexistence of any noticeable trends as a result of the variation of the weights suggests that all the clusters work together in obtaining a particular simulated value for all the components of the streamflow hydrograph.

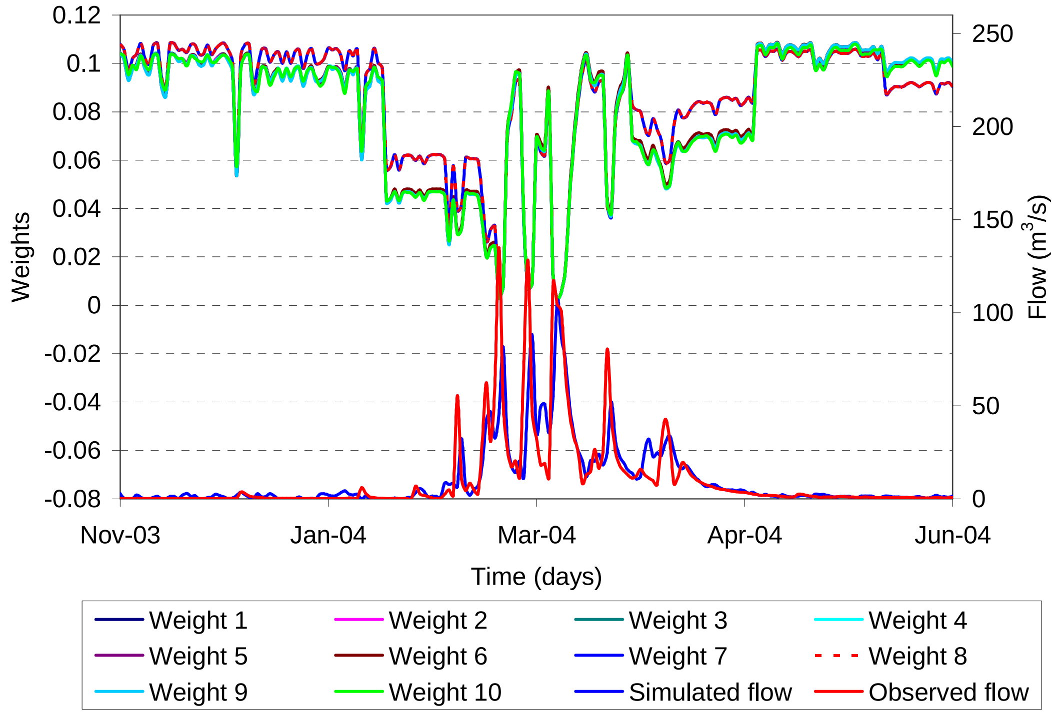


Figure 6.7: Variation of the degree of membership for the 10 clusters of the second river reach.

The various model components discussed in the previous sections indicate that all components work together to produce a particular value. Therefore, all the foregoing analysis seems to suggest that fuzzy inference is a black box model.

6.3 RESULTS AND DISCUSSION OF THE FUZZY MODELLING

Several attempts done to improve the model simulation have been discussed in Section 6.2.5 and all the evidence from the analysis conducted in that section does not support the suggestion that fuzzy inference model is not a black box model. This section evaluates the performance of the model based on four performance measures - the Nash-Sutcliffe efficiency (NSE), correlation coefficient (CCoef), percent bias (PBIAS), and the root mean square (RSR) - all of whose details are given in Section 6.2.2 and the performance statistics are given Table 6.3. The performance of the fuzzy inference model improved downstream; an observation attributed to the reduced system complexity downstream resulting from the reduced human activities. The model simulation results for the calibration and verification periods for the three reaches are shown in Figure 6.8 through to Figure 6.13. In addition, the simulation (calibration and verification) of the linked (combined) reaches discussed in Section 6.2.3 was done for the second and the third river reaches and the results are presented in Figure 6.13 and Figure 6.14 (the first and second linked) and Figure 6.15 and Figure 6.16 (all the three reaches are linked).

The performance statistics shown in Table 6.3 suggest that the values of CCoef improved towards the lower river reach and ranged between 0.720 and 0.923 for the calibration and also between 0.47 and 0.95 for verification periods respectively. The maximum values being obtained for the third river reach and the minimum values obtained for the first river reach (Table 6.3). For the first river reach, the model performed better during calibration than verification. The model failed to attain the same or a higher performance levels when linked than when the individual reaches were modelled in isolation as values of 0.74 and 0.813 were obtained for the connected second and third river reaches respectively. The model's performance on the upper reach influences the performance in the lower river reaches through error propagation. The general observation of the model performance based on *CCoef* indicates that the model performed better during the calibration phase than during the verification phase, with the exception of the third river reach. When a better performance is observed during the verification phase, it is likely that the calibration processes managed to obtain a generally suitable set of parameters. The upper river reaches are characterised by the existence of hydrological and non-hydrological processes, some of which act simultaneously and consequently increase the complexity of the system.

Table 6.3: Performance of the fuzzy inference model based on RMSE

River Reach	River	Calibration				Verification			
	reach	CCoef	NSE	PBIAS	RSR	CCoef	NSE	PBIAS	RSR
First	1st	0.72	0.51	-8.9	0.70	0.47	0.08	-3.47	0.83
Second	2nd	0.80	0.63	-9.43	0.60	0.79	0.56	-13.95	0.52
Third	3rd	0.92	0.85	3.96	0.38	0.95	0.90	6.95	0.21
				Connected					
Second	1st - 2nd	0.76	0.58	8.4	0.63	0.74	0.48	-9.8	0.57
Third	1st - 3rd	0.85	0.72	4.04	0.53	0.81	0.66	6.66	0.40

Considering the values of the PBIAS obtained for all the river reaches (Table 6.3), the model generally overestimates flow. In some incidences, the statistics obtained for the Black Heron suggest that the model underestimated the flows during the calibration and verification of linked and individual reaches. For the second river reach, the values of the PBIAS obtained ranged between -13.95% and 8.40% for the individual and linked river reaches respectively. When the best performance value of PBIAS is considered for all the reaches during calibration and verification phase, the values obtained are -3.47 and 3.96 being for the first and third river reaches respectively. Inline with the *CCoef*, the NSE values obtained indicated some improvement towards the downstream. During the calibration phase and for the individual reaches, the NSE values obtained ranged from 0.507 (for the first reach) to 0.851 (for the third river reach), while for the linked reaches the values ranged between 0.76 and 0.85 for the second and third river reaches. During the verification phase, the model performed unsatisfactorily with a NSE value of 0.08 obtained for the first river reach, while the rest of the NSE values ranged between 0.48 and 0.90. However, the general values obtained for the NSE indicate that the model performed satisfactorily in simulating the flows. The values of the root mean square error and observed standard deviation ratio obtained for all the river reaches ranged between 0.21 and 0.83 and the best (small) value was obtained for the third river reach.

In general, and based on the performance ratings in Table 6.2, very good simulations have been obtained in the third river reach because of the insignificant human activities and performances ranging from very good to good have been noticed for the simulation obtained for the second river reach.

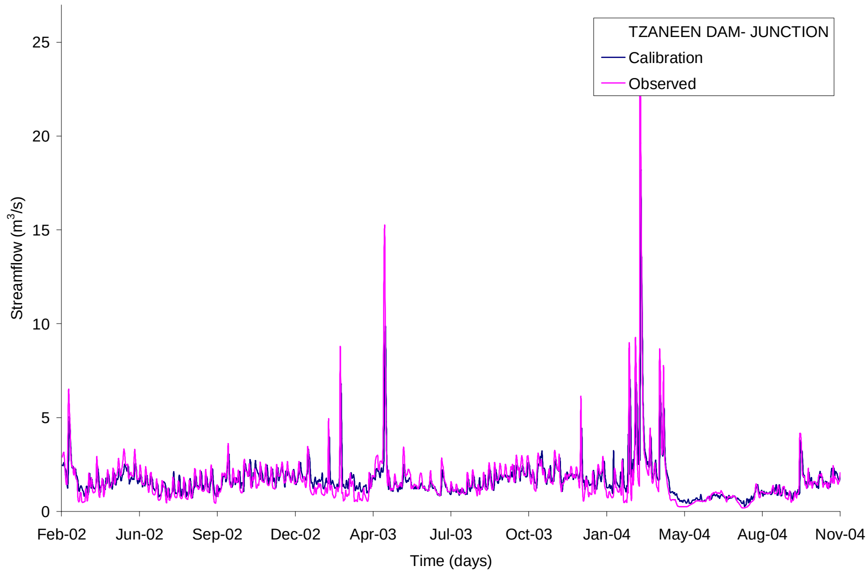


Figure 6.8: Calibration and observed flow at the first river reach

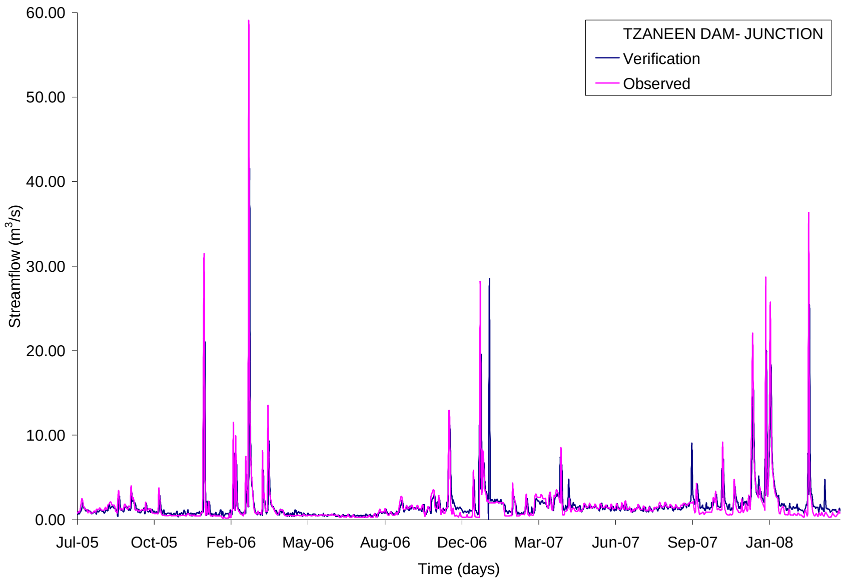


Figure 6.9: Verification and observed flow at the first river reach

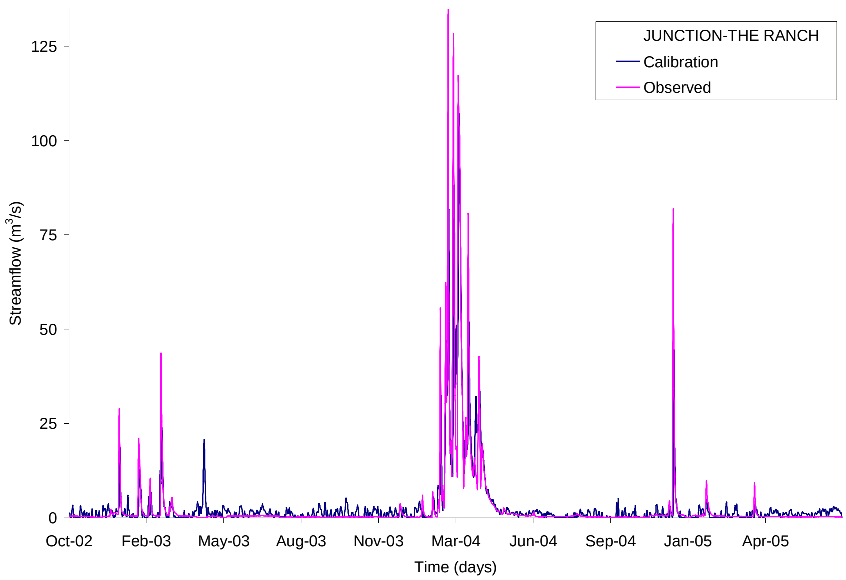


Figure 6.10: Calibration and observed flows at the second river reach

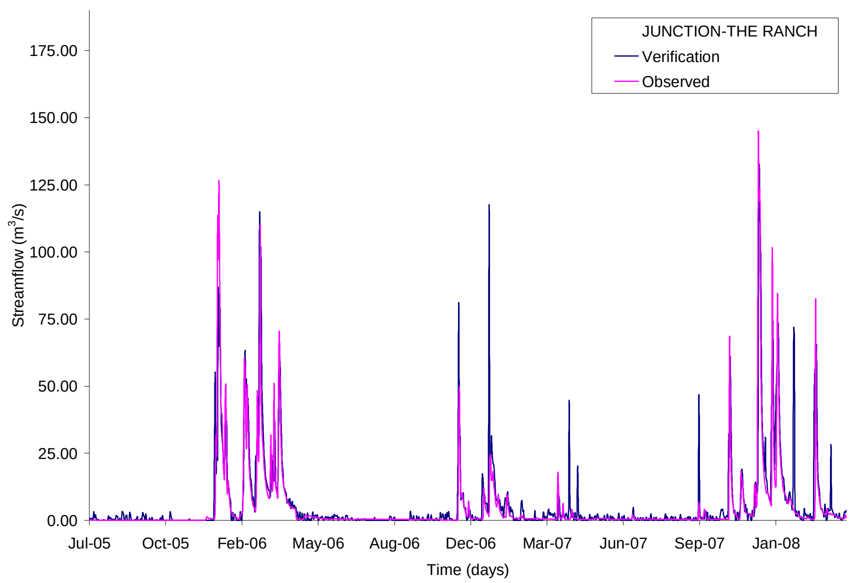


Figure 6.11: Verification and observed flows at the second river reach

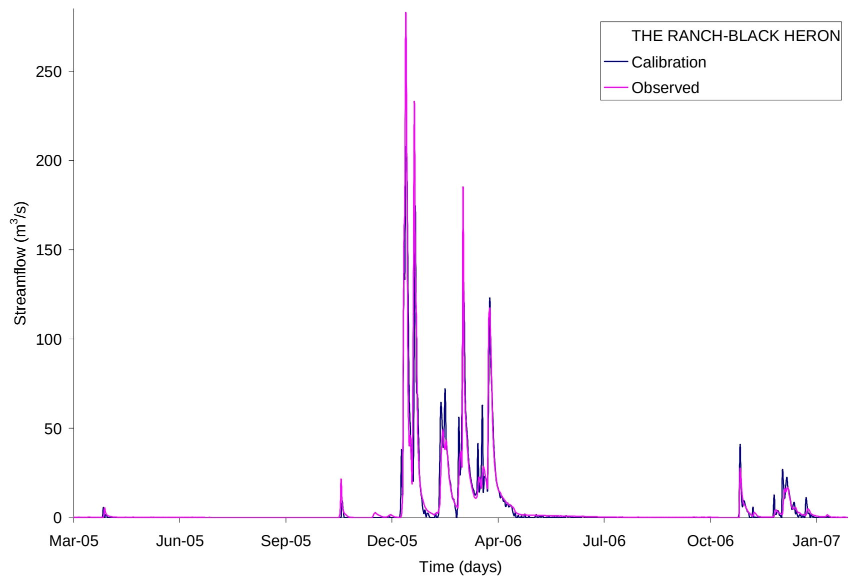


Figure 6.12: Calibration and observed flows at the third river reach

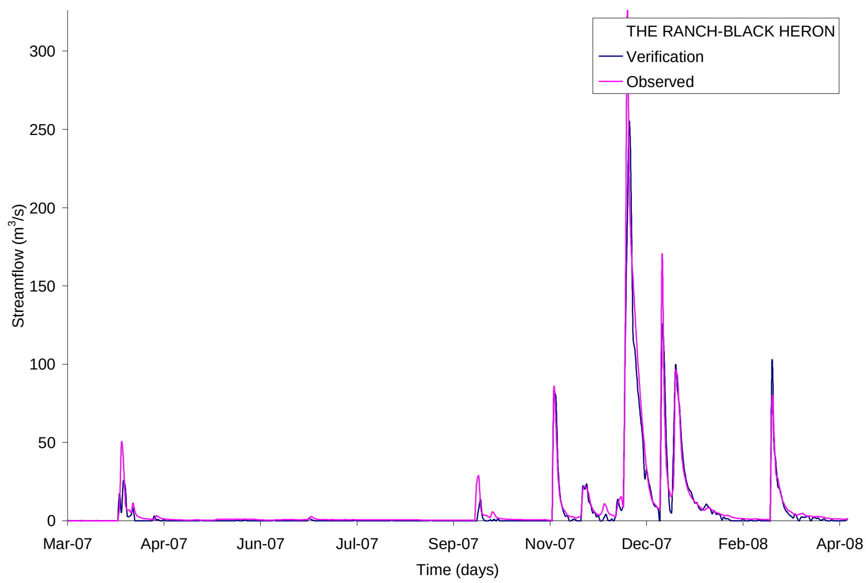


Figure 6.13: Verification and observed flows at the third river reach

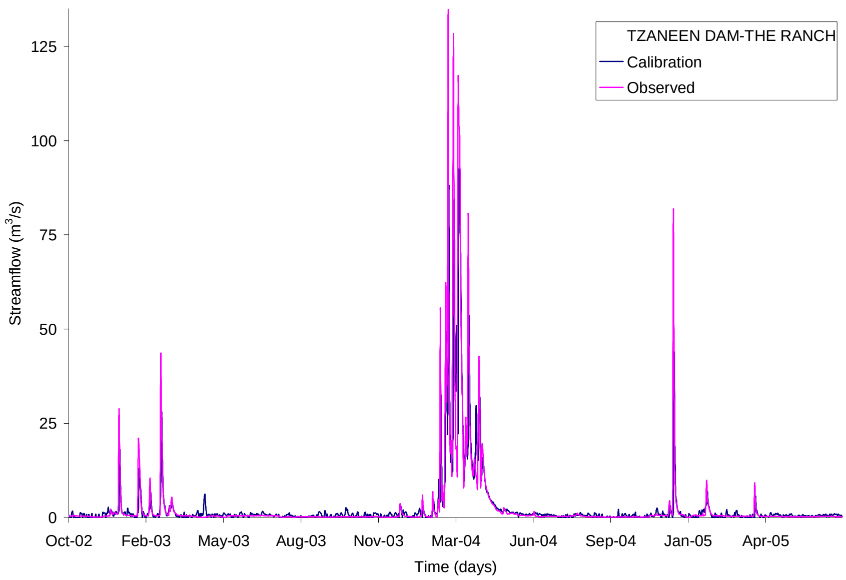


Figure 6.14: Observed and calibration of the combined flows at the second river reach

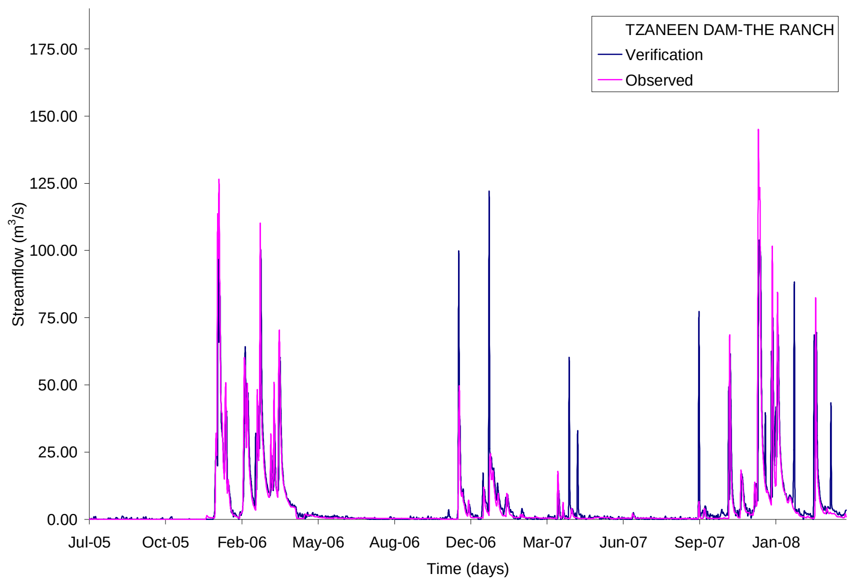


Figure 6.15: Observed and verification of the combined flows at the second river reach

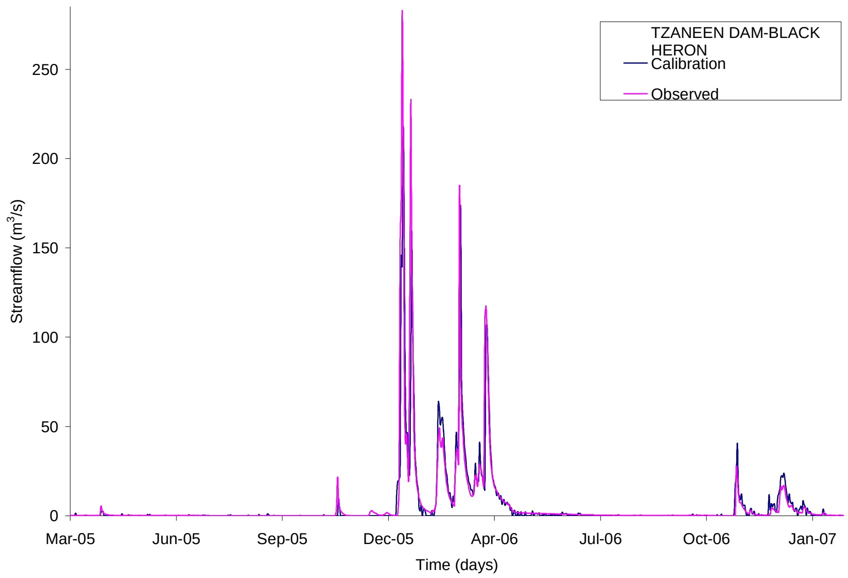


Figure 6.16: Observed and calibration of the combined flows at the third river reach

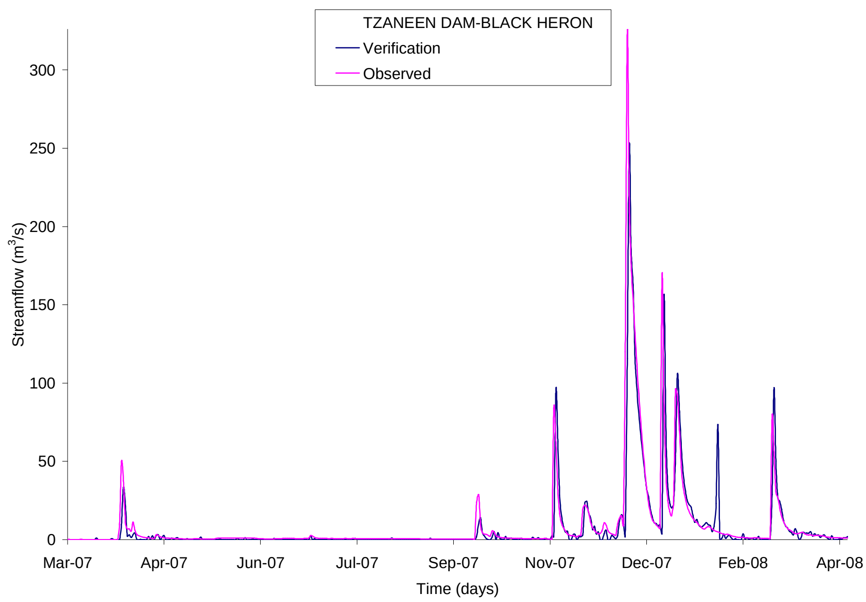


Figure 6.17 Observed and verification of the combined flows at the third river reach

Fuzzy inference has been applied to model the flows in the Letaba River system. Although the simulations suggest that the performance was satisfactory, the need to shed some light on the main catchment processes in the Letaba River system would go a long way in improving the understanding of the system. For this reason, a conceptual model, which allows for the representation of the major catchment processes, was developed for purposes of comparison with the fuzzy inference model. A description of this conceptual model follows in the next chapter.

7 CONCEPTUAL MODELLING OF THE LETABA RIVER SYSTEM

7.1 CONCEPTUAL MODELLING

Fuzzy inference was used in modelling flows in the Letaba River system and the results of this are presented in the Chapter 6. However, it is reasonable to expect that better simulations may be obtained if the known hydrological processes are explicitly incorporated into the fuzzy inference model. The conceptual model described in this chapter was therefore relied upon to provide a good understanding of the major hydrological processes in the catchment which would subsequently be incorporated in the fuzzy inference model. The development and application of the conceptual model presented in this chapter is also meant to compare a conceptual modelling approach (that incorporates an understanding of catchment processes) and the fuzzy modelling which did not provide any evidence of being more than be a black box in Chapter 6. The incorporation of conceptual modelling into the fuzzy inference model has been done in the next chapter.

The modelling of complex systems like the Letaba River system (as described in Chapter 4) is more challenging than that of a natural catchment and needs to make valuable use of the available soft data and consider limitations in availability of hard data. For verification of the conceptualization, use is made of the hard data from field measurements and reported findings of other studies. What follows next (Section 7.2) is the description of the development of the conceptual model. The developed conceptual model is very specific to the activities taking place in the Letaba River system. The general principles applied here can however be more generally applied. The discussion of the results obtained from its application is done in Section 7. 3.

The fuzzy inference model applied is a black box (Chapter 6), the need to improve the knowledge on the poorly understood catchment processes merited the development of a conceptual model that represent the main catchment processes.

7.2 DEVELOPING CONCEPTUAL MODELS

7.2.1 Description of the main processes and their modelling

In catchments, water is held in various parts and conceptual models attempt to represent these parts using a set of conceptual storages, while the various hydrological processes are modelled using algebraic functions (Pilgrim and Bloomfield, 1980). The hydrological processes are complex and the existence of human induced processes only serves to complicate the modelling of the processes even more.

The initial rainfall is intercepted by the vegetation and this intercepted water is only available for evaporation. The rainfall that is not intercepted by the vegetation contributes to the soil moisture; thus it influences the soil moisture status. The soil moisture contributes to the subsurface flow and evaporation. When the amount of excess rainfall is higher than the amount required to saturate the catchment soil, the surplus water contributes to the overland flow. The overland and subsurface flows contribute to the storage weirs and farm dams. Since the operation of the system is done in a manner that suggests that the abstractions are done from the storage weirs to farm dams, the releases made from Tzaneen Dam are meant to supplement the amount of water supplied from the storage weirs. As the released water from Tzaneen Dam flows downstream, some of it is abstracted depending on the scheduled amount and the remaining water flows is for the Kruger National Park. The abstraction done from the farm dams to the farms is mainly for irrigation purposes and the main crop is citrus. Table 7.1 shows a brief description of the main processes, the corresponding data availability and how data limitations impact on the modelling.

The building of the conceptual model recognises the identified subsystems; interception, soil, storage weirs, farm dams and channel flow as inter-linked conceptual storages. Figure 7.1 shows the model structure for the first and second river reaches while Figure 7.2 shows the model structure used for the third river reach.

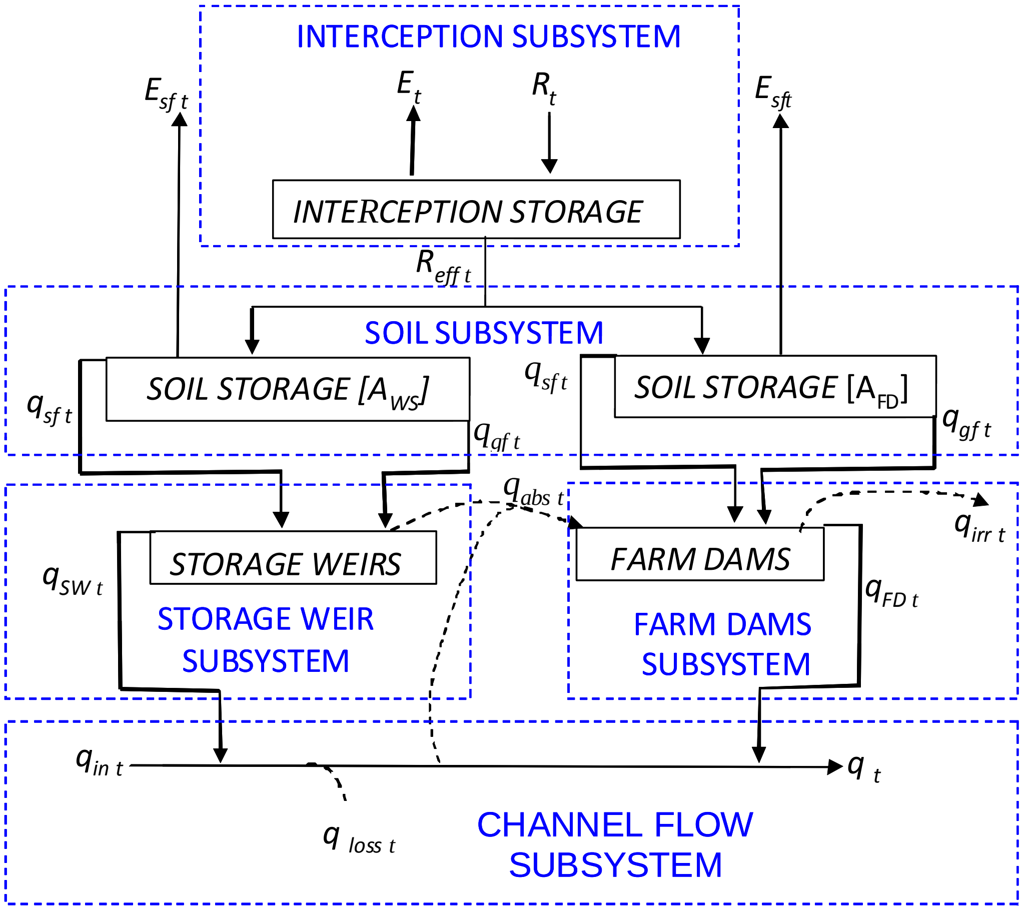


Figure 7.1: Schematic diagram for the conceptual model for the first and second river reaches

(where E_t is the evaporation from the interception storage, $E_{sf\,t}$ is the evaporation from the soil surface, R_t is the rainfall received, $R_{eff\,t}$ is the effective rainfall, $q_{sf\,t}$ is the overland flow, $q_{gf\,t}$ is the flow contribution from the soil, $q_{FD\,t}$ is the contribution from the farm dams, $q_{sw\,t}$ is the contribution from the storage weirs, $q_{loss\,t}$ is the transmission losses, $q_{in\,t}$ is the inflow into the river, q_t is outflow from the river reach

$q_{abs\,t}$ is the abstraction into the farm dams and $q_{irr\,t}$ is the abstraction made from the farm dams into the farms)

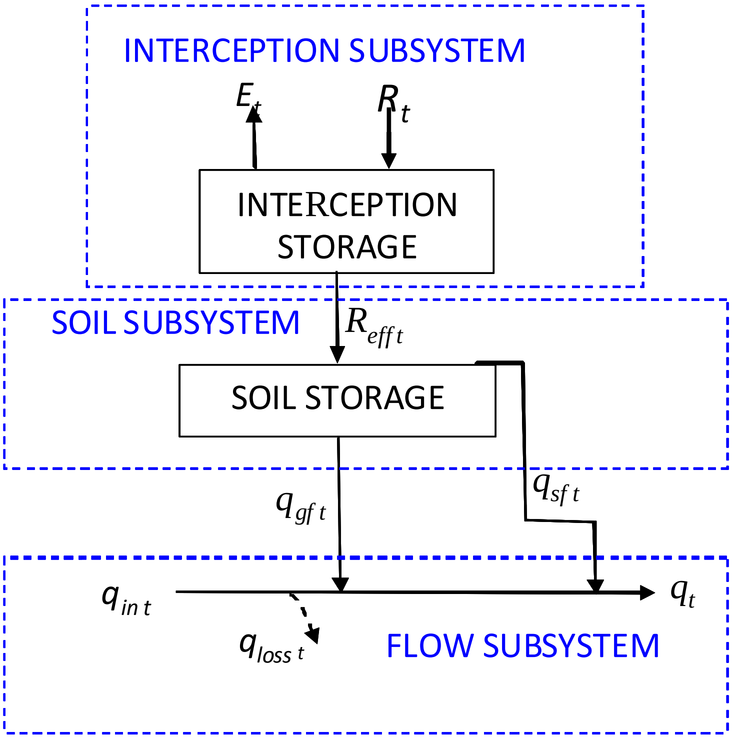


Figure 7.2: Schematic diagram for the conceptual model for the third river reach.

Table 7.1: A brief description of the processes and data availability

Storages	**Processes**	**Data available**	**Description**
Interception storage	Evaporation and contribution to soil moisture	Pan evaporation and rainfall	Rainfall is intercepted; the excess rainfall contributes soil moisture, while the intercepted water contributes to evaporation. The capacity [mm] is calibrated
Soil storage	Overland flow, subsurface flow and evaporation	No data is available	The excess rainfall contributes to the soil moisture that generates overland flow and subsurface flow into the storage weirs and farm dams. The capacity [mm] is calibrated
Storage weirs	Evaporation, inflow from the soil, channel flow contribution, and abstraction	Only monthly abstraction data is available and no data is available for the total capacity (lumped storage has been assumed)	The status of the storage weirs depend on the inflow (overland flow and subsurface flow), evaporation and abstraction The capacity [m^3] is calibrated
Farm dams	Evaporation, inflow from the soil, abstraction	Only monthly abstraction data is available, no data is	The status of the farm dams depend on the inflow (overland flow and

Storages	Processes	Data available	Description
	from the weirs and river into farm dams and abstraction from farm dams into the farms.	available for the total capacity (lumped storage has been assumed) and irrigation demands.	subsurface flow), evaporation, abstraction from the weirs and river into dams and abstraction from farm dams into the farms. The capacity [m^3] is calibrated
Channel storage	Stream flow losses and stream flow	Only stream flow (inflow and outflow) is available and there is no data on the streamflow losses	The inflow is contributed by releases from the upstream storage and runoff.
Other issues			The releases made from Tzaneen Dam are meant to supplement the flows and are not stored in the storage weirs.

a.) Interception subsystem

The interception subsystem involves the interception of rainfall by vegetation. When a certain amount of rainfall R_t, is received in the catchment, a portion of the rainfall is intercepted by the vegetation and this water is only available for evaporation. The capacity of the interception storage is determined by calibration and represents the total amount available for evaporation. The excess (effective) rainfall $R_{e,t}$, is given as:

$$R_{e,t} = \begin{cases} 0 & if\, R_t - e_t + S_{int,t-1} \le S_{intmax} \\ R_t + S_{int,t-1} - e_t - S_{intmax} & if\, R_t - e_t + S_{int,t-1} > S_{intmax} \end{cases} \quad 7.1$$

where S_{intmax} is the maximum capacity of the interception storage, $S_{int,t-1}$ is the storage status at time step *t-1*. The water in the interception storage is lost through evaporation and is given as (Beven and Kirkby, 1979):

$$e_t = \begin{cases} E_{POT,t} & if\;\; E_{POT,t} \le S_{int,t} \\ E_{POT,t} \times \dfrac{S_{int,t}}{S_{intmax}} & if\;\; E_{POT,t} > S_{int,t} \end{cases} \quad 7.\ 2$$

where e_t is the actual evaporation, S_{intmax} is the calibrated capacity of the storage, $S_{int,t-1}$ is the storage status at time step *t* and $E_{POT,t}$ is the potential evaporation and is given as:

$$E_{POT,t} = \varpi E_{pan,t} \quad 7.\ 3$$

where $E_{pan,t}$ is the pan evaporation value and ω is the pan coefficient. The pan evaporation values significantly vary and this have been also reported in other studies (Chiew et al., 1995, Sumner and Jacobs, 2005). The use of known pan coefficient values is not practical and as such the appropriate pan coefficient value was determined through calibration as the value obtained this way was considered more realistic. The inputs into the interception subsystem are the pan evaporation and rainfall (R_t) and the outputs are actual evaporation and excess rainfall. The excess rainfall is then used as an input to the soil storage subsystem.

b.) Soil subsystem

The soil subsystem involves the movement of water from and into the soil layer, as evaporation, overland and subsurface flow. Excess rainfall flows into the soil conceptual storage and the amount required to saturate the soil layer depends on the antecedent soil moisture condition. The capacity of the soil conceptual storage is obtained through calibration. The water in the soil conceptual storage contributes to the flow and evaporation when an insufficient amount of water is available in the interception storage. The water balance in the soil storage is given as:

$$S_{soil,t} = \begin{cases} S_{soil,t-1} + R_{e,t} + e_{a,t} - q_{soil,t} & if\, S_{soil,t-1} + R_{e,t} + e_{a,t} - q_{soil,t} \le S_{soilmax} \\ S_{soil,t-1} + R_{e,t} + e_{a,t} - q_{soil,t} - q_{soilflow} & if\, S_{soil,t-1} + R_{e,t} + e_{a,t} - q_{soil,t} > S_{soilmax} \end{cases} \quad 7.4$$

where $S_{soil,t}$ is the status of the soil conceptual storage, $e_{a,t}$ is the evapotranspiration from the soil when an insufficient amount of water is available in the interception storage, $q_{soilflow,t}$ is the excess amount of water available when the soil storage is full (overland flow) and it contributes to the storage weirs and farm dams and $S_{soilmax}$ is the capacity of the soil conceptual storage. The subsurface flow from the soil contributes to the farm dams or storage weirs and is given as (Beven and Kirkby, 1979):

$$q_{soil,t} = S_{soil,t} e^{-ks} \quad 7.\ 5$$

When $q_{soil,t}$ is the flow from the soil storage. The evaporation from the soil is given as:

$$e_{a,t} = \begin{cases} 0 & if\ E_{POT,t} - e_t = 0 \\ (E_{POT,t} - e_t) \times \dfrac{S_{soil,t}}{S_{soilmax}} & if\ E_{POT,t} - e_t > 0 \end{cases} \quad 7.6$$

Flows from the soil, $q_{soil,t}$ and the $q_{soilflow,t}$ contributes to the storage weirs and the farm dams and is dependent on the area that drains into the farm dams or storage weirs at the respective time period. The catchment area that drains to the storage weirs and farm dams has been obtained through digitization.

c.) Farm dams subsystem

The farm dam subsystem involves the inflow and outflow from the farm dams. The existence of numerous farm dams and the lack of information on the size and shape of the farm dams lead to the decision to use a single lumped storage to represent the farms dams in the first and second reach. It was considered practical to model storage weirs similarly as their shapes and sizes were also not known precisely. The tetrahedral reservoir shape (Figure 7.3) applied by Schulze et al. (1989) was adopted. The maximum volume at full supply capacity is based on the catchment area and is given as:

$$S_{\max} = A \times H \quad 7.7$$

where S_{max} is the volume at full supply capacity of the facility, A is the catchment area that drains into the facility and H is the catchment depth of water that is enough to fill an empty facility.

The relationship between the storage depth (D) and the width (2B) and also between the storage depth (D) and the length L are given by Equation 7.8 and 7.9 respectively.

$$B = D \times tan\theta \qquad 7.8$$

$$L = \frac{D}{tan\beta} \qquad 7.9$$

where β and θ are the angles as shown in Figure 7.3.

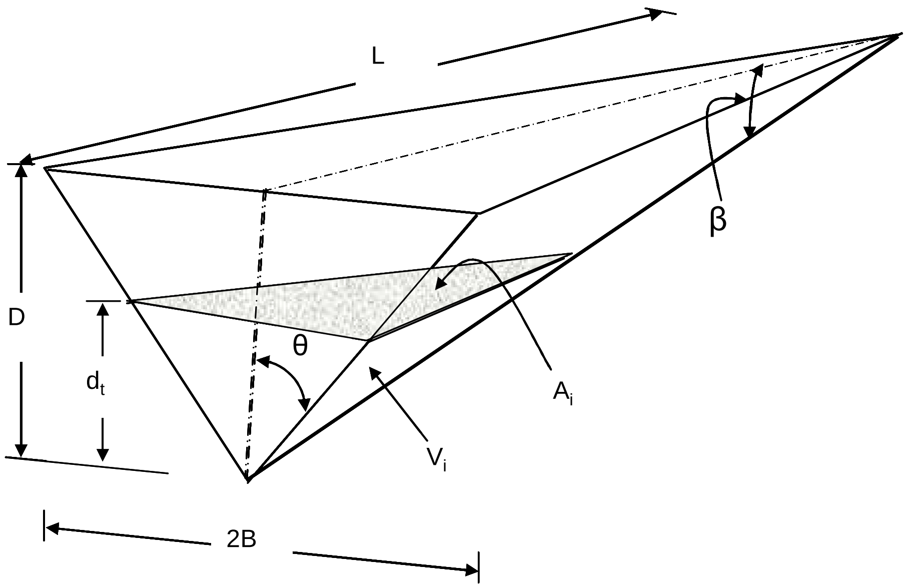


Figure 7.3: A typical shape of the storage weir/farm dam used in the model

The volume of the storage facility based on the dimensions is given as:

$$S_{\max} = \frac{1}{3} D \times B \times L \qquad 7.10$$

Substituting Equation 7.8 and 7.9 into Equation 7.10 gives the volume as

$$S_{\max} = \frac{1}{3} D^3 \frac{tan\theta}{tan\beta} = \frac{\gamma}{3} D^3 \qquad 7.11$$

The shape parameter γ for the storage facilities is given as:

$$\gamma = \frac{tan\theta}{tan\beta} \qquad 7.12$$

Therefore the capacity of the facility is given as:

$$S_{\max} = \frac{\gamma_{weir}}{3} D_{weir}{}^3 \qquad 7.13$$

The area covered by the storage facility is given by

$$A = B \times L \qquad 7.14$$

Substituting Equation 7.8 and 7.9 into Equation 7.14 gives the area as

$$A = \frac{tan\theta}{tan\beta} D^2 = \gamma_{weir} \times D_{weir}{}^2 \qquad 7.\ 15$$

The area covered by the storage facility can also be given as

$$A = \gamma \times D^2 \qquad 7.16$$

The depth at full supply capacity of the storage facility is obtained by combining Equation 7.13 and 7.16 giving:

$$D = \frac{3 \times S}{A} \qquad 7.17$$

The volume at any time step *t* is given as:

$$S_t = \frac{\gamma}{3} d_t^{\,3} \tag{7.18}$$

And the surface area is given as:

$$A_t = \gamma \times d_t^{\,2} \tag{7.19}$$

The shape parameters and the area at full supply capacity are calibrated.

The subsystem modelling reflects the actual operation of the system, where all the abstractions are done from the storage weirs and the releases made from Tzaneen Dam are meant to supplement the storage weirs' supplies when insufficient water is available. The water from the Tzaneen Dam therefore does not contribute to the storage weirs. The farm dams in each reach have been lumped and represented by a single farm dam that the model attempts to simulate the inflow and outflow. The storage is depleted by the net evaporation process and abstraction of water to irrigate the farms while the inflow are from subsurface flow, overland flow, abstractions made from the storage weirs and directly from the river. The area of catchment that contributes water to these dams varies as some of the farm dams become full. In order to account for this, the lumped farm dam has been perceived as having 3 zones as shown in Figure 7.4. The bottom zone between 0 and d_{min}, represents situations were no farm dam has overflowed and there is therefore no contribution to river flow. The middle zone bounded by d_{min} and D with range ($D-d_{min}$) represents the situation where some of the farm dams are full and therefore spilling. The third zone is reached when the water level reaches the highest depth D when all the farm dams become full and cannot store more water. The whole catchment therefore contributes to flow.

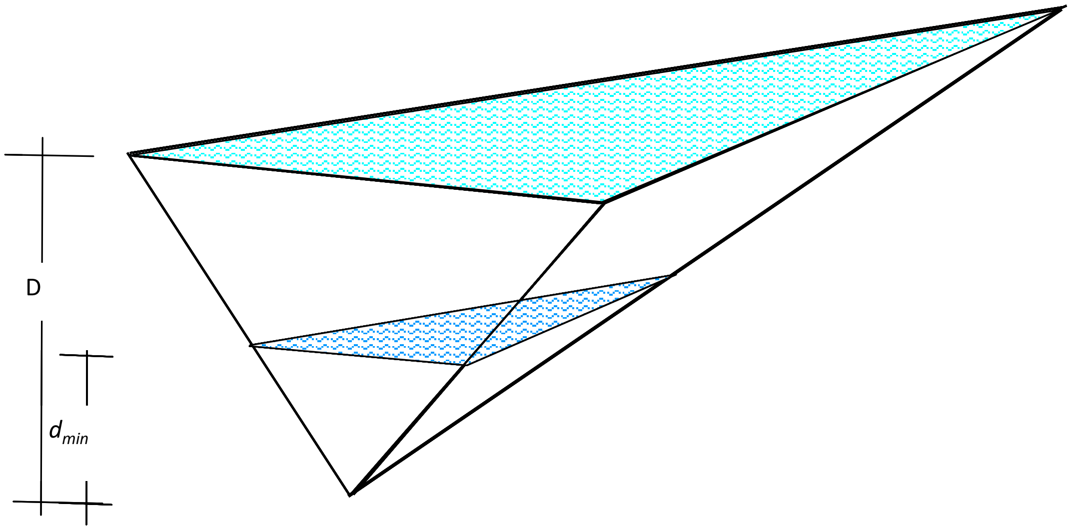


Figure 7.4: Storage zone of the farm dam

The water balance for the lumped farm dam is obtained as:

$$S_{daminter,t} = \begin{cases} 0 \quad if\ S_{dam,t-1} + q_{soil,t} \times \dfrac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} - e_{dam,t} + q_{abs,t} - q_{irr,t} - q_{damflow,t} < 0 \\ S_{dam,t-1} + q_{soil,t} \times \dfrac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} - e_{dam,t} + q_{abs,t} - q_{irr,t} - q_{damflow,t} \\ if\ 0 \leq S_{dam,t-1} + q_{soil,t} \times \dfrac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} - e_{dam,t} + q_{abs,t} - q_{irr,t} - q_{damflow,t} \leq S_{dam\max} \\ S_{dam\max}\ if\ S_{dam\max} < S_{dam,t-1} + q_{soil,t} \times \dfrac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} - e_{dam,t} + q_{abs,t} - q_{irr,t} - q_{damflow,t} \end{cases} \qquad 7.20$$

where $S_{daminter,t}$ is the status of the farm dam, A_{dam} is the area that drains into the farm dams, A_T is the total area of the river reach, $e_{dam,t}$ is the evaporation from the weir, S_{dammax} is the capacity of the farm dams and $q_{irr,t}$ is the amount of water abstracted from the farm dams for irrigation purposes. The evaporation from the farm dams is given as:

$$e_{dam,t} = 0.5 \times \left(A_{damsur,t} + A_{damsur,t-1}\right) \times E_{POT,t} \qquad 7.21$$

where $A_{damsur,t}$ and $A_{damsur,\ t-1}$ is surface area covered by water in the storage weir at time step t *and* t-1. *The status of the farm dam $S_{dam,t}$, is given as:*

$$S_{dam,t} = \begin{cases} 0 \quad if \ S_{dam,t-1} + q_{soil,t} \times \frac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} - e_{dam,t} + q_{abs,t} - q_{irr,t} - q_{damflow,t} < 0 \\ S_{dam,t-1} + q_{soil,t} \times \frac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} - e_{dam,t} + q_{abs,t} - q_{irr,t} - q_{damflow,t} \\ if \ 0 \le S_{dam,t-1} + q_{soil,t} \times \frac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} - e_{dam,t} + q_{abs,t} - q_{irr,t} - q_{damflow,t} \le S_{dam\max} \\ S_{dam\max} \ if \ S_{dam\max} < S_{dam,t-1} + q_{soil,t} \times \frac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} - e_{dam,t} + q_{abs,t} - q_{irr,t} - q_{damflow,t} \end{cases} \tag{7.22}$$

The outflow from the storage weir is given by:

$$q_{damflow,t} = \begin{cases} 0 & if \, d_{\min} > H \\ c \times H_t^{\ d} & if \, d_{\min} < H_t \le D \\ \left(q_{soil,t} + q_{soilflow,t}\right) \times \frac{A_{dam}}{A_T} & if \ H_t > D \end{cases}$$

The water balance for the storage weirs is given as:

$$S_{dam,t} = \begin{cases} 0 \quad if \ S_{dam,t-1} + q_{soil,t} \times \frac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} - e_{dam,t} + q_{abs,t} - q_{irr,t} - q_{damflow,t} < 0 \\ S_{dam,t-1} + q_{soil,t} \times \frac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} - e_{dam,t} + q_{abs,t} - q_{irr,t} - q_{damflow,t} \\ if \ 0 \le S_{dam,t-1} + q_{soil,t} \times \frac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} - e_{dam,t} + q_{abs,t} - q_{irr,t} - q_{damflow,t} \le S_{dam\max} \\ S_{dam\max} \ if \ S_{dam\max} < S_{dam,t-1} + q_{soil,t} \times \frac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} - e_{dam,t} + q_{abs,t} - q_{irr,t} - q_{damflow,t} \end{cases} \tag{7.20}$$

where $S_{dam,t}$ is the status of the storage weir, A_{dam} is the area that drains into the weirs, A_T is the total area of the river reach, $e_{dam,t}$ is the evaporation from the weir, $q_{damflow,t}$ is the outflow from the farm dam and $q_{irr,t}$ is the amount of water abstracted from the farm dams for irrigation purposes as given in Figure 4.7 (Section 4.4). The evaporation from the farm dams is given as:

$$e_{dam,t} = 0.5 \times \left(A_{dam,t} + A_{dam,t-1}\right) \times E_{POT,t} \qquad 7.21$$

where $A_{dam\ t}$ and $A_{dam\ t-1}$ is surface area covered by water in the storage weir at time step t and $t-1$. When the water level is within the second zone, the outflow is given by:

$$q_{damflow,t} = \begin{cases} 0 & if\, d_{\min} > H \\ c \times H_t{}^d & if\, d_{\min} < H_t \leq D \\ q_{soil,t} \times \dfrac{A_{dam}}{A_T} + q_{soilflow,t} \times A_{dam} & if\ H_t > D \end{cases} \qquad 7.22$$

Where c and d are parameters that take into account the outflow from the filled farm dams and they are calibrated.

d.) Storage weir subsystem

The storage weir subsystem represents the movement of water into and from the storage weirs. The subsystem is based on the actual operation of the system, where all the abstractions are done from the storage weirs and the releases made from Tzaneen Dam are only meant to supplement the supply from these weirs when the available water is insufficient or when the storage weirs are empty. Even when the weirs are empty, the water released from Tzaneen dam does not contribute to the storage weirs. A detailed discussion of the current system operation has been given in Section 4.4

The water balance for the storage weirs prior the consideration of the evaporation to be defined as:

$$S_{weirinter,t} = \begin{cases} 0 & if\, S_{weir,t-1} + q_{soil,t} \times \frac{A_{weir}}{A_T} + q_{soilflow,t} \times \frac{A_{weir}}{A_T} - q_{absweir,t} < 0 \\ S_{weir,t-1} + q_{soil,t} \times \frac{A_{weir}}{A_T} + q_{soilflow,t} \times \frac{A_{weir}}{A_T} - q_{absweir,t} & \\ \quad if\, 0 \le S_{weir,t-1} + q_{soil,t} \times \frac{A_{weir}}{A_T} + q_{soilflow,t} \times \frac{A_{weir}}{A_T} - q_{absweir,t} \le S_{weir\max} & \\ S_{weir\max} & if\, S_{weir\max} < S_{weir,t-1} + q_{soil,t} \times \frac{A_{weir}}{A_T} + q_{soilflow,t} \times \frac{A_{weir}}{A_T} - q_{absweir,t} \end{cases} \quad 7.23$$

where $S_{weirinter,t}$ is the status of the storage weir before considering evaporation, A_{weir} is the area that drains into the weirs, AT is the total area of the river reach, $e_{weir,t}$ is the evaporation from the weir and $q_{absweir,t}$ is the amount of water abstracted from the weir. The portion of the catchment area that drains into the storage weirs was determined by using GIS and is 92.44 km2 and 430.63 km2 for the first and second river reach respectively. A portion of water is lost from the weirs through evaporation and is given as:

$$e_{weir,t} = 0.5 \times \left(A_{weirsur,t} - A_{weirsur,t-1}\right) \times E_{POT,t} \quad 7.24$$

where $A_{weirsur,t}$ *and* $A_{weisurr,t-1}$ is the water surface area in the storage weir. The status of the storage weir, $S_{weir,t}$ is given as:

$$S_{weir,t} = \begin{cases} 0 & if\, S_{weir,t-1} + q_{soil,t} \times \frac{A_{weir}}{A_T} + q_{soilflow,t} \times \frac{A_{weir}}{A_T} - e_{weir,t} - q_{absweir,t} < 0 \\ S_{weir,t-1} + q_{soil,t} \times \frac{A_{weir}}{A_T} + q_{soilflow,t} \times \frac{A_{weir}}{A_T} - e_{weir,t} - q_{absweir,t} & \\ \quad if\, 0 \le S_{weir,t-1} + q_{soil,t} \times \frac{A_{weir}}{A_T} + q_{soilflow,t} \times \frac{A_{weir}}{A_T} - e_{weir,t} - q_{absweir,t} \le S_{weir\max} & \\ S_{weir\max} & if\, S_{weir\max} < S_{weir,t-1} + q_{soil,t} \times \frac{A_{weir}}{A_T} + q_{soilflow,t} \times \frac{A_{weir}}{A_T} - e_{weir,t} - q_{absweir,t} \end{cases} \quad 7.25$$

where $q_{weir,t}$ is the flow contribution from the filled storage weirs, $S_{weir,t}$ is the storage status of the weirs and $S_{weirmax}$ is the capacity of the storage weirs.

e.) Streamflow subsystem

The streamflow subsystem involves the movement of water within the stream based on the inflows and outflows. During periods of low flow, when flows within the river cannot meet the anticipated demands, releases are made from Tzaneen Dam to supplement the flows in the river. This, however, only happens when the water in the storage weirs has been used up and as such the abstractions are made directly from the river. The total abstraction from the river to the farm dams is therefore given by:

$$q_{abs,t} = q_{absweir,t} + q_{absriver,t} \qquad 7.26$$

where $q_{absriver,t}$ is the amount of water abstracted from the river as is available as data.

During dry periods and in absence of any lateral inflows, the inflow at the upstream end of the river reach is higher than the outflow. The difference between the two is attributed to the river losses that occur in between as demonstrated in Figure 7.5a for two gauging stations of the third river reach. Figure 4.2 shows the locations of the two stations. The observed inflows at the Ranch are higher than the observed outflows at the Black Heron due to the losses occurring within the reach as the impact of human activities on flows in this reach are negligible. In their study of ephemeral streams, Lane et al. (1971) found that the losses can be related to inflow and a similar approach has been used here for all the three river reaches. The difference between inflow and outflow has been used for the initial estimation of the losses (Equation 7.28). Figure 7.5b shows the relationship between the losses computed this way and the inflow.

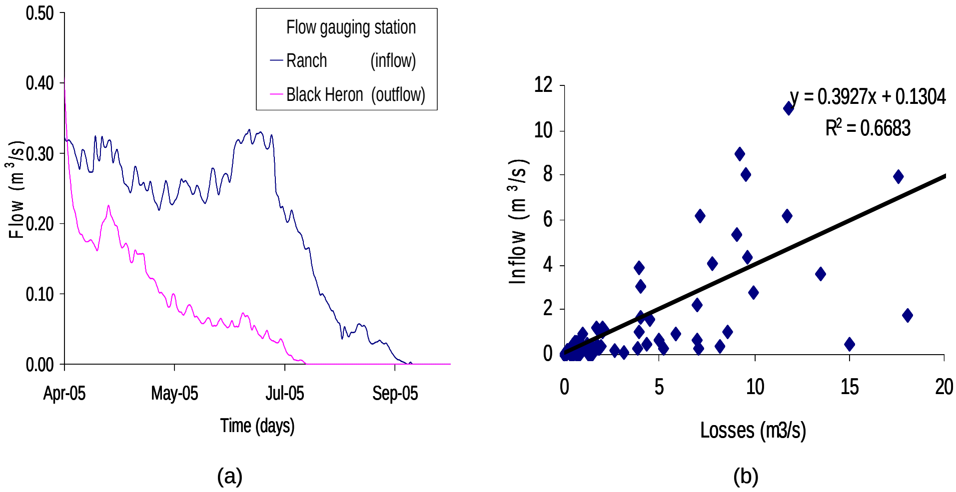


Figure 7.5: Third river reach (a.) Observed flows at the gauging stations observed at the Ranch and Black Heron, and (b) Correlation between the losses and inflow.

$$q_{loss,t} = \begin{cases} q_{in,t} - q_{out,t} & if\ q_{in,t} - q_{out,t} > 0 \\ 0 & if\ q_{in,t} - q_{out,t} \leq 0 \end{cases} \qquad 7.27$$

where $q_{loss,t}$, $q_{out,t}$ and $q_{in,t}$ are the losses, outflow and inflow at time t.

The satisfactory value of R^2 obtained (0.67) (Figure 7.5b) suggests that the losses in the model can be expressed as a linear function of the form:

$$q_{loss,t} = aq_{in,t} + b \qquad 7.28$$

where $q_{loss,t}$ is the losses at time space t, a and b are coefficients. It was assumed that the losses derived from Equation 7.28 represent the sum of all the losses that include evaporation and flow into the alluvial aquifer.

The total simulated outflow from each river reach is given as:

$$q_{sim,t} = q_{in,t} + q_{weir,t} + q_{damflow,t} - q_{loss,t} - q_{absriver,t} \qquad 7.29a$$

$$q_{sim,t} = q_{in,t} + q_{tri,t} + q_{weir,t} + q_{damflow,t} - q_{loss,t} - q_{abriver,t} \qquad 7.29b$$

$$q_{sim,t} = q_{in,t} + q_{soil,t} + q_{soilflow,t} - q_{loss,t} \qquad 7.29c$$

where Equation 7.29a is for the first river reach, Equation 7.29b is for the second river reach and Equation7.29c is for the third river reach respectively.

7.2.2 Parameters and calibration of the conceptual model

Table 7.2 shows the calibrated parameters for the two models. Only the catchment areas were determined from the GIS maps. The Shuffled complex evolution-University of Arizona (SCE-UA) algorithm (Duan et al., 1992) was used as the calibration algorithm, while the root mean square error and logarithmic transformed root mean square error were used as objective functions for all the river reaches.

Table 7.2: Summary of the calibrated model parameters

Parameter	Units	Description
S_{intmax}	[mm]	Capacity of the interception storage
$S_{soilmax}$	[mm]	Capacity of the soil storage
D_w	[mm]	Equivalent depth of lumped storage weirs based on catchment area draining to the weirs
A_w	[km^2]	Surface area of storage weirs at full supply capacity
D_f	[mm]	Equivalent depth of lumped farm dam storage based on catchment area draining to the farm dams
A_f	[km^2]	Surface area of the lumped farm dam storage when full
ω	[-]	Pan coefficient
k_s	[-]	Coefficient of the flow contributing area
a and b	[-]	Parameter of the loss function
β	[-]	Parameter for the irrigation demand*
c and e	[-]	Parameters for the flow contribution for the filled farm dams
S_{FDMIN}	mm	Lowest depth of the lumped farm dam storage below which flow contribution ceases

* to obtain actual demand from estimated monthly distribution

The shape parameter is not calibrated, but the optimal value is obtained through the calibration of the depth H of rainfall in Equation 7.7 and the depth of the storage facility D given in Equation 7.11.

The existence of equifinality is evident when hydrological model obtains many different parameter sets that reproduce outputs that are equally good (Beven, 2006; Savenije, 2001). This situation merited to the use for ten random runs for each model and optimal parameter sets were obtained when values of the objective function consisted of insignificant variations. Also, each objective function tends put more emphasis on a particular aspect of the hydrograph, as such, the RMSE and LOGE where used and favours high flows and low flows respectively.

7.3 RESULTS AND DISCUSSION OF THE CONCEPTUAL MODELLING

An attempt to find out how realistic the modelling was, is done by analysing the calibrated quantities of the modelled physical characteristics (subsystems) including the interception, soil, farm dams, storage weirs and the movement of water as losses, abstractions and outflows. It is however necessary to assess the calibration adequacy first and this is what follows next.

7.3.1 Calibration adequacy of the conceptual model

The insignificant variations in the values obtained for the objective function from 10 randomly initialized calibration runs (Figure 7.6) suggest that the model calibration was adequate. Model calibration was therefore considered to be adequate for all three river reaches.

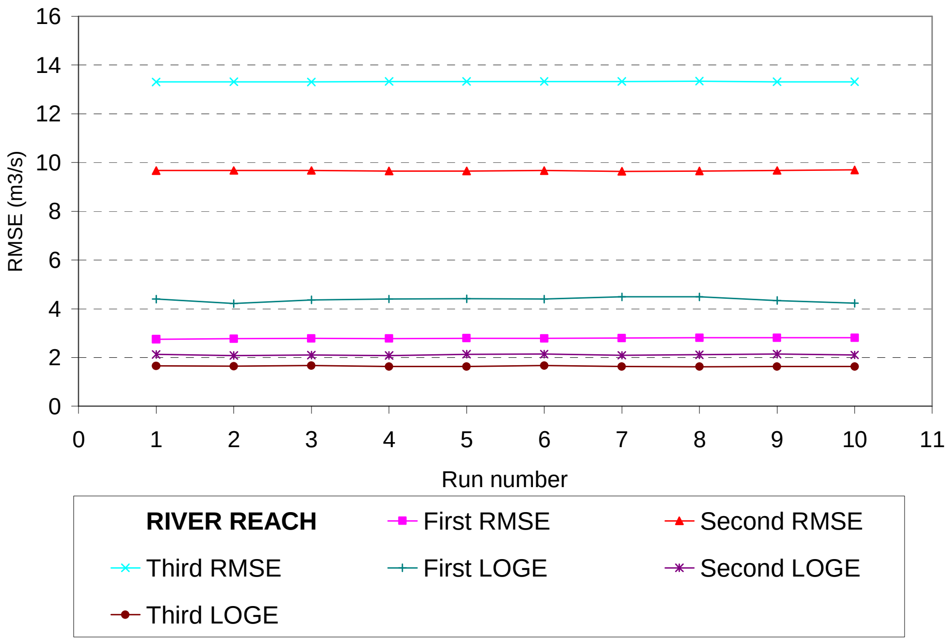


Figure 7.6: Typical values for the objective function for the various runs (where RMSE is the root mean square error and LOGE is the transformed logarithmic values of the RMSE)

7.3.2 Reproduction and assessment of catchment processes by conceptual models

The absence of data to represent the physical characteristics of the Letaba River system such as the capacities of the farm dam and storage weirs necessitates for some modelling procedures to be undertaken to account for these characteristics. Using a conceptual model, it may be possible to identify representative parameter values for some of these system characteristics. In addition, the ability to reproduce the catchment processes requires the identification of the various representative parameter values of the models. The obtained values and the simulations will give an indication of the probable ranges of magnitudes of various components of the system such as farm dam capacities and these parameter values are dependent on the objective function.

a. Calibrated Interception storage

Figure 7.7 shows the capacity of the interception storages for the river reaches based on the two objective functions. Generally, the values in Figure 7.7 indicate that when the root mean square error (RMSE) was used as the objective function, higher values for the storages were obtained than was the case with the transformed logarithmic (LOGE) as an objective function. The values range between 2.0 mm and 5.93 mm when the root RMSE is used as an objective function and when the LOGE is used, the values obtained range between 0.83 mm and 4.85 mm. Generally, the maximum value obtained for the interception storage in this study is slightly less than 6 mm. The value is comparable to what was reported in other studies (Klaassen et al., 1998, Liu et al., 2002, Granger and Gray, 1990, Qiu et al., 1998) since they reported values that are slightly higher than 7 mm.

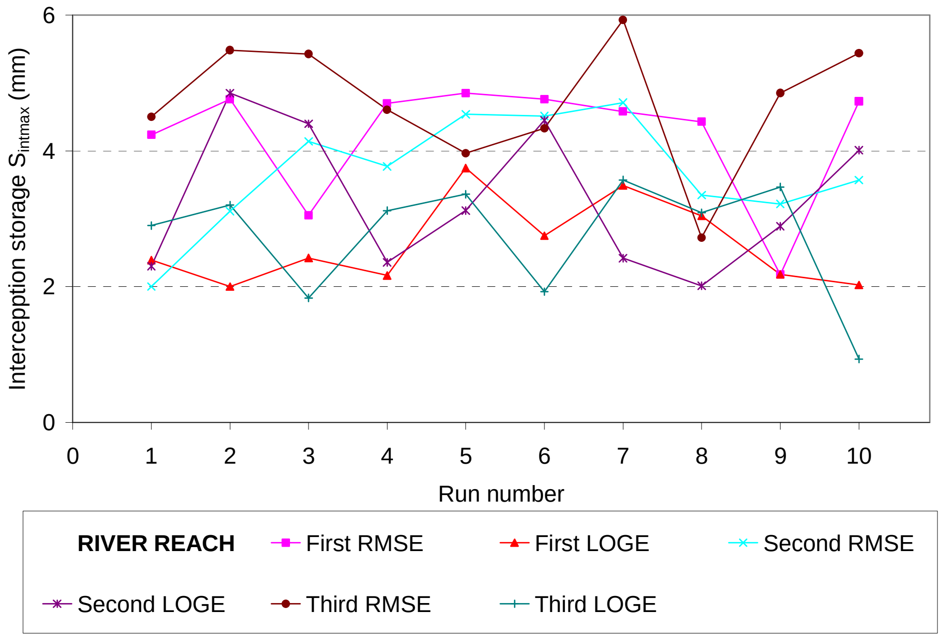


Figure 7.7: Capacity of interception storage for the conceptual model obtained for all the river reaches for 10 calibration runs

b. Calibrated soil storage

A portion of the catchment water is held by the soil and is represented by the soil storage. When the interception storage is full, the excess water flows into the soil storage and the water is available as subsurface flow, overland flow and direct evapotranspiration losses. Figure 7.8 shows the capacity of the soil storages for the river reaches based on the two objective functions. The values obtained in Figure 7.8 are lower than what was expected to be the catchment soil storage based on what was reported in other studies (e.g. Hughes and Sami, 1994, Tan and O'Connor, 1996, Fenicia et al., 2006). For instance, Fenicia et al. (2006) used the FLEX model in Luxembourg and obtained the upper soil storage to be slightly above 400mm. Tan and O'Connor (1996) applied SMAR model in different regions and the soil storage values obtained range from slightly above 100mm to 400mm. Hughes and Sami (1994) applied the Variable Time Interval (VTI) model in the Eastern Cape Province in South Africa and obtained the field capacity to be 12mm. This small value only suggests that the flow contribution does not last long and is inline with actual systems where it has been observed that the tributaries just flow for a short duration after the rainfall (Venter, 2008).

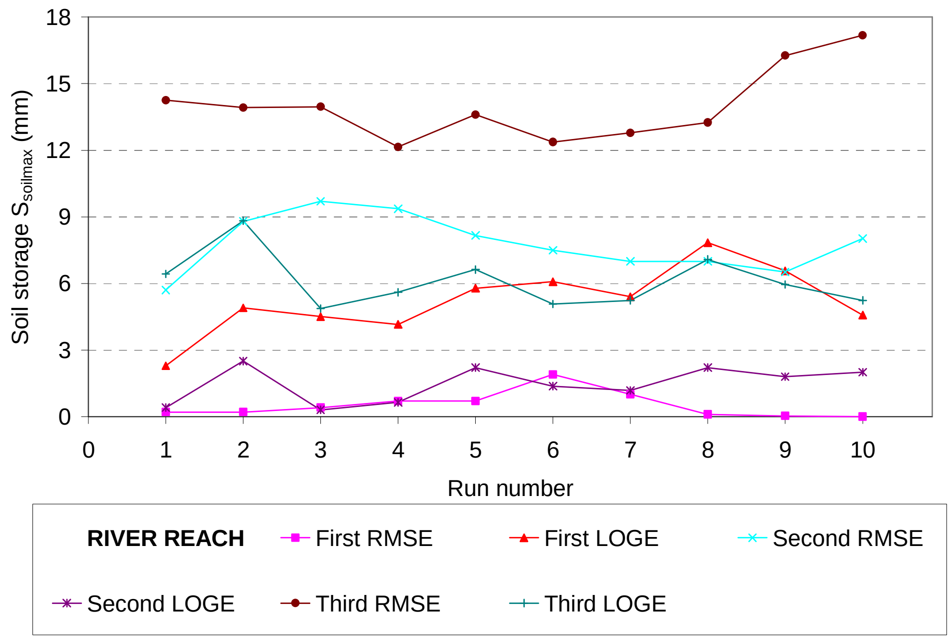


Figure 7.8: Capacity of soil storage for the conceptual model obtained for all the river reaches for 10 calibration runs

c. Calibrated capacities of storage weirs and farm dams

The calibration of the depth and surface area of the equivalent storage weir and farm dam volume has been used as a means of determining the depth and the capacity of the facilities for the conceptual lumped farm dam and storage weir. Figures 7.9 and 7.10 show the depth of the storage weirs and farm dams obtained from the first and second river reaches for 10 calibration runs, while Figures 7.11 and 7.12 show the capacities of the storage weirs and farm dams obtained for the first and second river reaches from 10 calibration runs.

The depth of the storage weirs obtained for the first and second river reaches varies (Figure 7.9). The values range from 0.03 m to 3.85 m for the first river reach and from 7.75 m to 18.42 m for the second river reach. Considering the two objective functions, the values obtained when RMSE is used as an objective function are lower than those obtained when the LOGE is used as an objective function. It is likely that RMSE is biased towards the high flow, implying it calibrated to less water storage than LOGE which favours low flow simulation. Considering the depth of the farm dams, the values obtained for the first and second river reaches significantly vary (Figure .7.10). For the first river reach, the obtained values range from 0.01 m to 1.51 m, while for the second river reach, the values range from 27.75 m to 286.49 m. The value of 0.01 obtained for the first river reach and the value of 286.49 m obtained for the second river reach do not represent a realistic situation, however, many of the values obtained do represent a realistic situation and currently there is no reason that can be linked to this. The values obtained when RMSE is used as an objective function and those obtained when LOGE is used as an objective function for the second river reach are within the same range. For the first river reach, the values obtained when LOGE is used as an objective function are higher than those obtained when RMSE is used as an objective function. This is attributed to influence of the objective function.

Figure 7.11 and Figure 7.12 show the capacities of the storage weirs and farm dams obtained for the first and second river reach respectively. The values show that more water is stored in the second river reach than in the first river reach. For the first river reach, the values range from 0.01×10^6 m^3 to 0.73×10^6 m^3, while for the second river reach the obtained values range from 2.35 $\times 10^6 m^3$ to 3.45 $\times 10^6 m^3$. Considering the capacity of the farm dams, the values obtained range from 0.01 $\times 10^6$ m^3 to 2.35 $\times 10^6$ m^3 for the first river reach and for the second river reach the values range from 17.07 $\times 10^6$ m^3 to 61.79 $\times 10^6$ m^3. The difference in the storage values between the first and second river reach is inline with the manner in which the weirs are operated (Section 4.4) and is evidence of the model's ability to represent the system in spite of the shortage of data and model's structural limitations.

The highest value obtained for the capacity of the farm dam (61.79 $\times 10^6 m^3$) as runoff storage capacity (dam volume divided by catchment area) is equal to 19.4 mm and it is within the range of values reported in a study done in the Bedford catchments in South Africa (Hughes and Sami, 1993). Hughes and Sami (1993) reported that the capacity of the farm dam can go up to 20 mm; a value that can absorb a substantial amount of the generated runoff.

Considering the effect of the storage weirs, these structures store some of the runoff generated and they are operated in a manner that is based on rule of thumb. The maximum storage obtained is $3.5x10^6$ m^3. The capacity of one of the storage weirs is $2.0x10^6$ m^3 (Venter, 2008) and the addition of other weirs may result in higher storage values close to the maximum obtained value; suggesting that the values are realistic.

When comparing the values obtained by the two objective functions, the root mean square error resulted in low farm dam capacities for the first river reach. However, the values obtained for the second river reach have insignificant differences. At the current stage, there are no reasons that can be linked to this discrepancy.

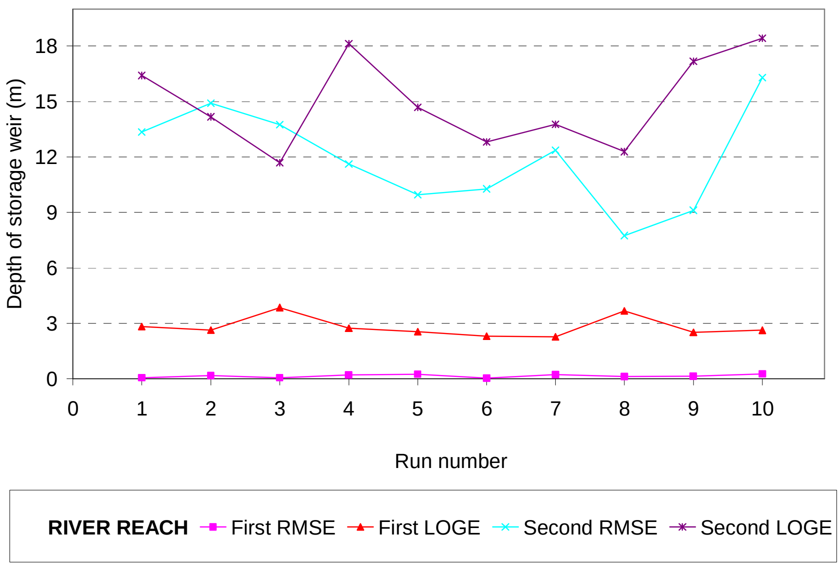


Figure 7.9: The calibrated depths of storage weirs obtained by the conceptual model for the first and the second river reaches for 10 calibration runs

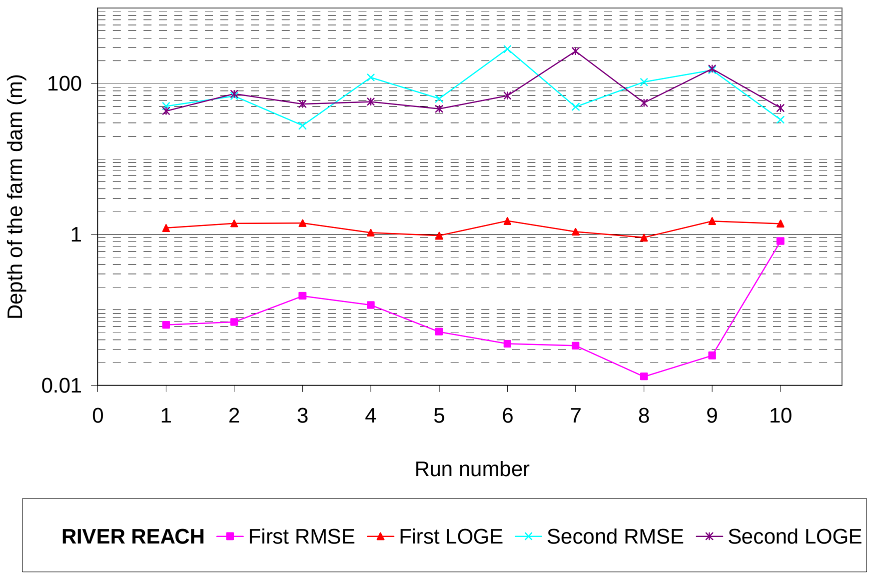


Figure 7.10: The calibrated depths of farm dam obtained by the conceptual model for the first and the second river reaches for 10 calibration runs

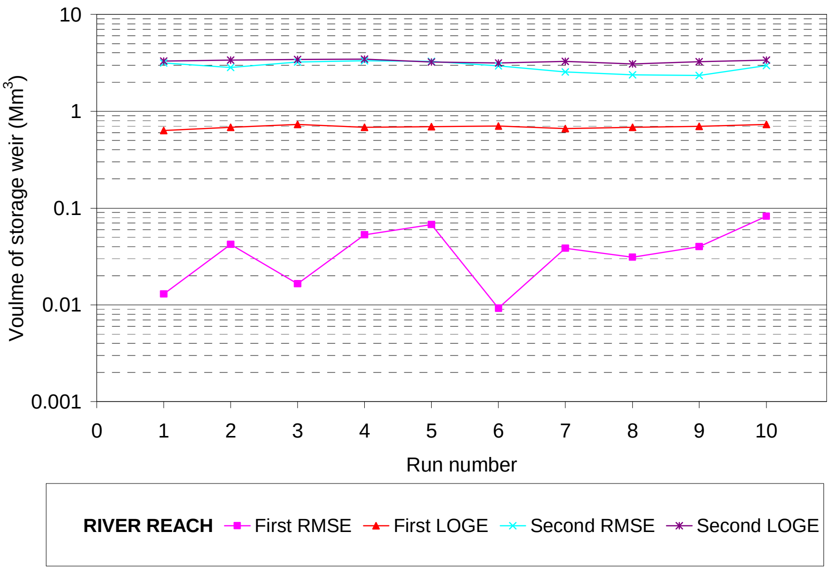


Figure 7.11: The calibrated volumes of storage weirs obtained by the conceptual model for the first and the second river reaches for 10 calibration runs

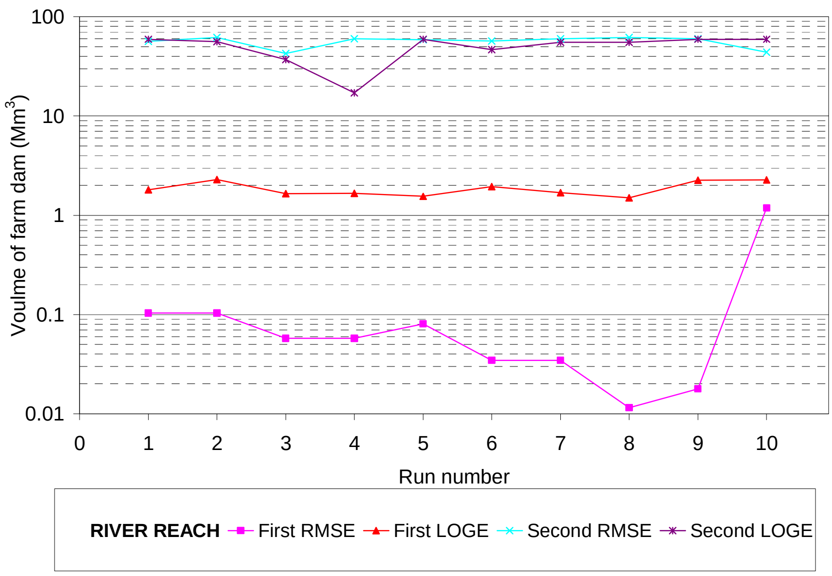


Figure 7.12: The calibrated volumes of farm dams and the total obtained by the conceptual model for the first and the second river reaches for 10 calibration runs.

d. Calibrated Stream flow loss parameters and simulated losses

The parameters values obtained for the linear flow loss function (Equation 7.28) for all the 10 runs vary significantly (Figures 7.13and 7.14). The majority of the slope values obtained for the third river reach are negative. The contribution to the flow in the river, particularly during rainy period, by the ungauged tributaries (Figure 6.1) is attributed to the negative values. Generally, the values obtained for the constant parameter of equation 7.28 for all the runs range from 1 to 8 (Figure 7.14). All the values obtained are positive and are in agreement with the hypothesis that losses take place as the water flows.

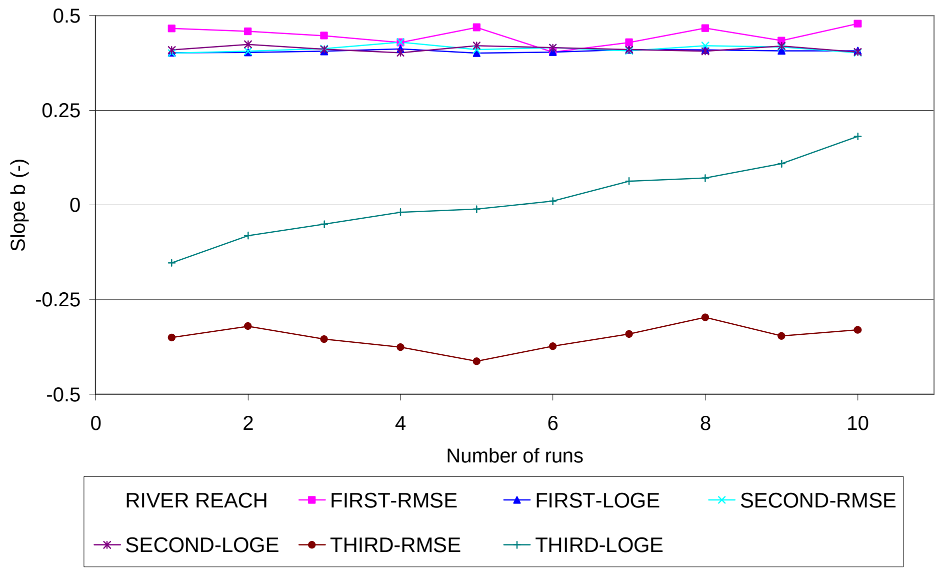


Figure 7.13: Variation of the slope b, of loss function parameter for the all reaches for 10 calibration runs

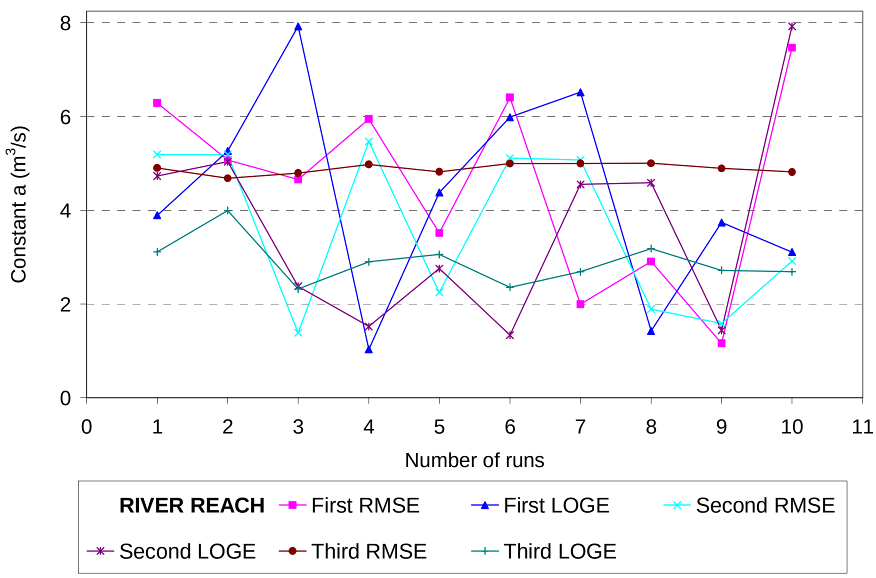


Figure 7.14: Variation of the constant a, of loss function parameter for the all reaches for 10 calibration runs

The simulated river flow losses are shown in Figures 7.15, 7.16 and 7.17. It can be observed that they decrease downstream with maximum values slightly less than 8 m^3/s, 7 m^3/s and slightly above 4 m^3/s for the first, second and third river reaches. It is likely that values of the losses also include the groundwater abstractions that are done within the alluvial aquifer. Although the function applied in the model is a linear function (Equation 7.28), the model managed to simulate the characteristics of the system's losses regardless of limited information. The obtained loss values are comparable to the values reported by Hughes and Sami (1992). Hughes and Sami (1992) obtained transmission losses of 4.2 m^3/s (1088900 m^3 in 72 hours) to an alluvial river bed in Bedford catchment. Considering the effects of the objective functions, the loss values obtained for the second river reach do not vary substantially but the variation is more pronounced in the values obtained for the first and third river reaches. The RMSE obtained higher values than the LOGE function which was attributed to the biasness of the respective objective functions; however the model simulations are realistic.

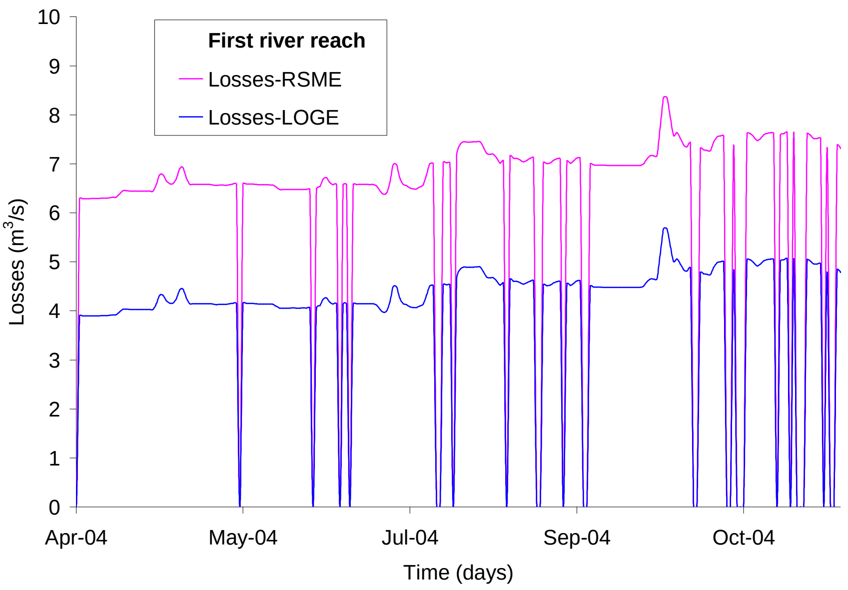


Figure 7.15: Simulated flow losses occurring along the first river reach during the calibration phase.

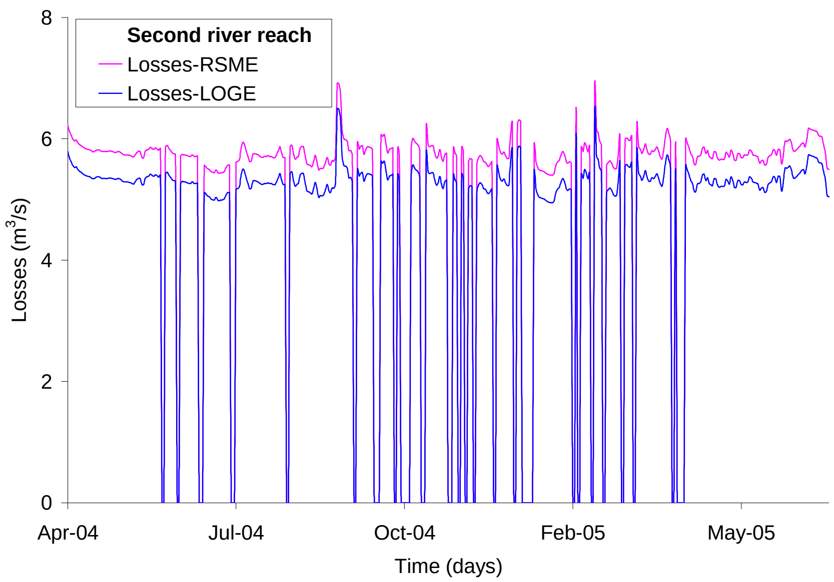


Figure 7.16: Simulated flow losses occurring along the second river reach during the calibration phase.

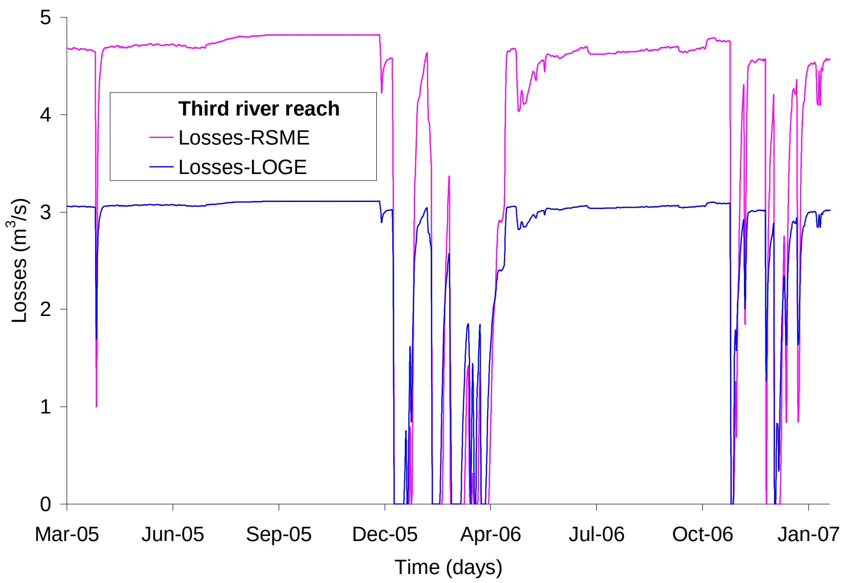


Figure 7.17: Simulated flow losses occurring along the third river reach during the calibration phase.

e. Simulated water use from farm dams

Farm dams are found in the first and second river reaches and the majority of the farm dams are used for irrigation purposes. Figures 7.18 through 7.21 show the simulated amounts of water abstracted from the reservoirs obtained during the model calibration and verification phase. For the first river reach, the simulated reservoir abstraction values obtained when the objective function used is LOGE were higher than those values obtained when RMSE was set as an objective function. On the other hand, the abstraction values obtained for the second river reach when the RMSE was set as an objective function were similar to those obtained when LOGE was set as an objective function. The difference is likely to be a result of the object function's tendency to favour low flow simulation. Although the simulated values are realistic, the influence that the soil storage has on the flow is insignificant when compared to the actual catchment soil storage. This is evidenced by low soil storage values of 5.7 mm obtained when RMSE was used as an objective function and 0.41 mm when LOGE was used as an objective function.

Generally, the obtained simulated reservoir abstraction values suggest that more abstraction activities happen in second river reach than in the first river. This is inline with what has been observed and also reported in DWAF (2006a). More irrigation activities have been reported along the second river reach than in the first. In addition, the use of groundwater to supplement the supply is supported by the fact that there are periods when the storage weirs and farm dams were simulated to be empty; suggesting, therefore, that the model managed to obtain realistic storage trajectories.

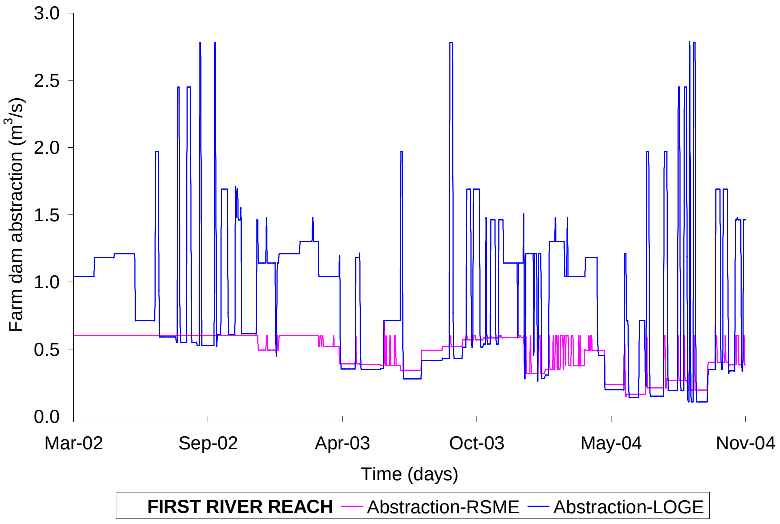


Figure 7.18: The farm dam abstraction series for the first river reach during the calibration phase

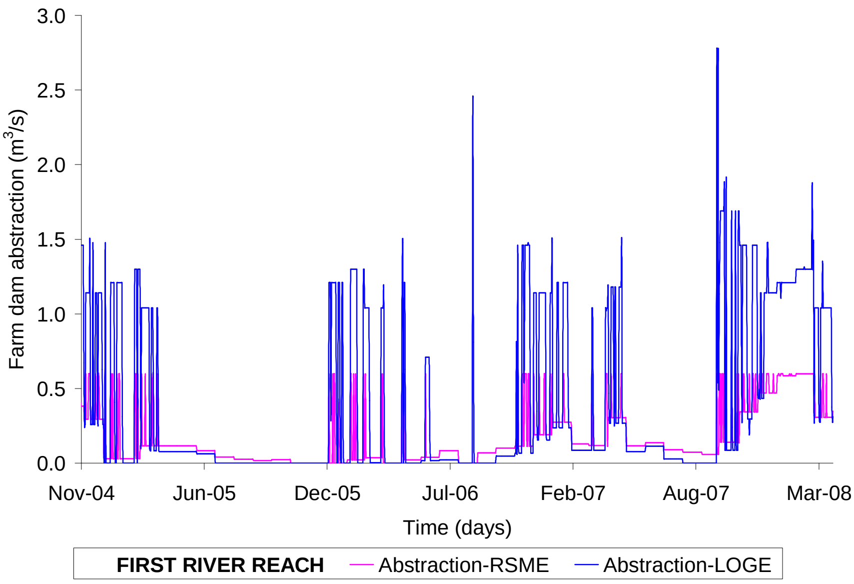


Figure 7.19: The farm dam abstraction series for first river reach during the verification phase

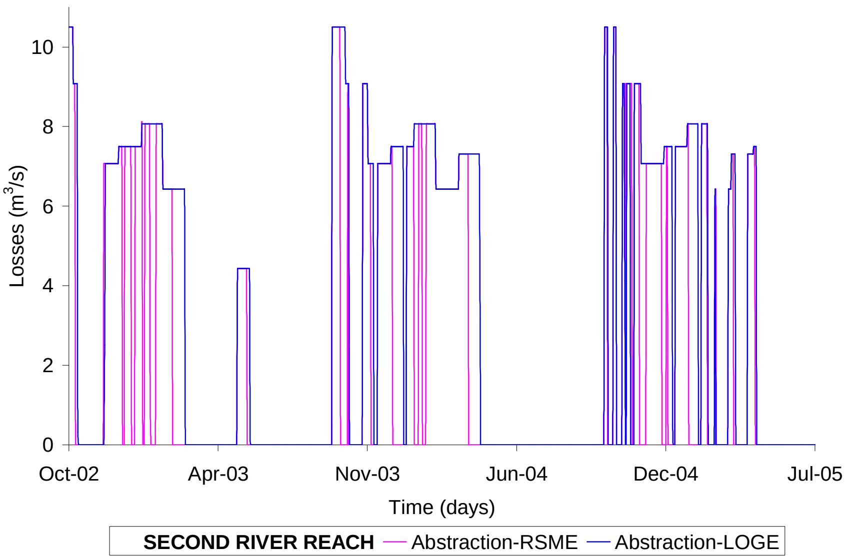


Figure 7.20: The farm dam abstraction series for second river reach during the calibration phase

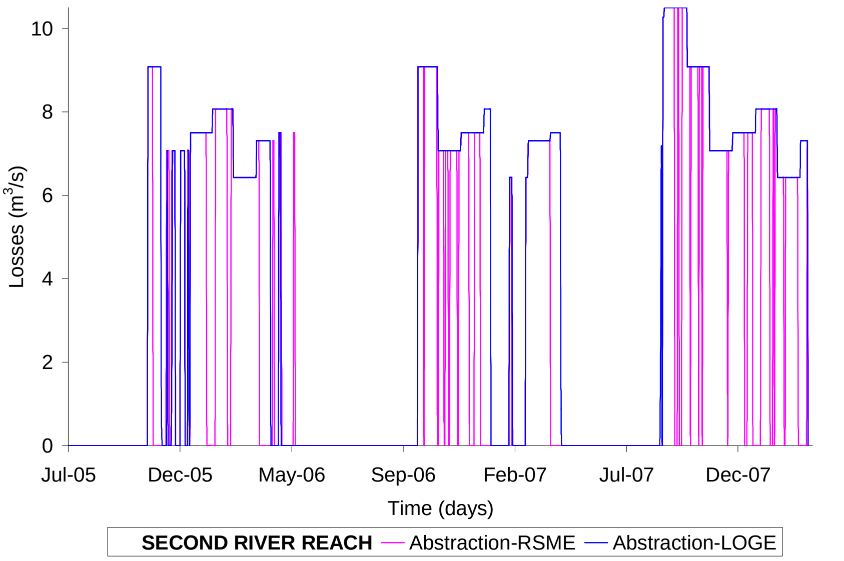


Figure 7.21: The farm dam abstraction series for second river reach during the verification phase

f. Simulated trajectories of Farm dams and storage weirs

The incorporation of the storage weirs and the farm dams in the conceptual model is based on the operation objectives that allow the Tzaneen Dam to supplement the amount of water available in the storage weirs. Therefore, the releases from the dam are assumed not to contribute to the storage; rather are abstracted from the river when the amount of water available in the storage weirs does not meet the anticipated demand (scheduled abstraction) for both reaches. Figure 7.22 and Figure 7.23 show the storage series (trajectories) of the storage weirs during the calibration and verification respectively and Figure 7.24 and Figure 7.25 show the modelled status of the farm dams during the calibration and verification respectively. During the calibration and verification phases, the trajectories indicate the existence of zero storage in the storage weirs. The water was abstracted directly from the river, and is inline with the operation of the system that intends for the releases made from the Tzaneen Dam only to supplement the flow. In addition, it is during the same period when restrictions are imposed on irrigation demands. The tendency of the RMSE to favour high flow is evidenced by the existence of longer periods in which the storage facilities are empty than the lengths of the periods associated with empty storages when LOGE is used. In the absence of the quantities abstracted from the groundwater, the empty status of the storage facilities may suggest that the supply was supplemented by the groundwater (Section 4.4). Therefore, it can be considered that the trajectories of the weirs and farm dams represent a realistic situation.

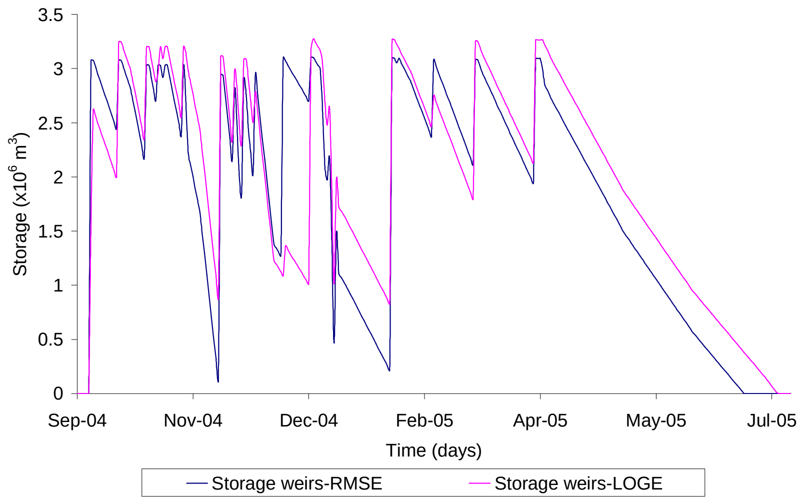


Figure 7.22: Typical storage trajectories of the storage weirs in the second river reach during calibration phase

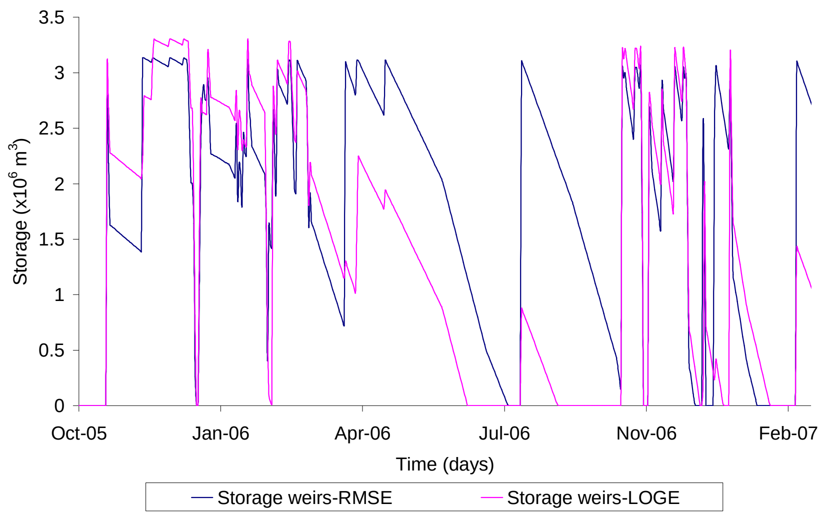


Figure 7.23: Typical storage trajectories of the storage weirs in the second river reach during verification phase.

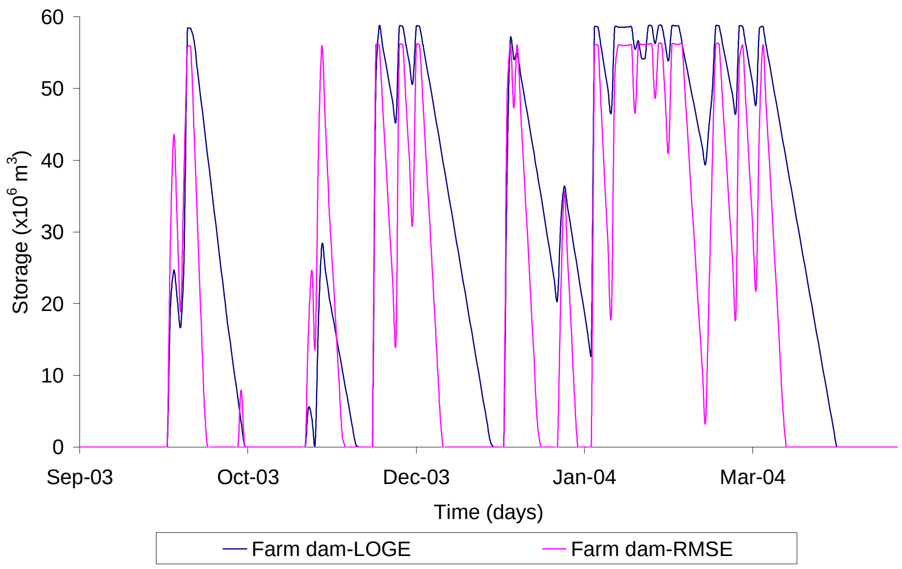


Figure 7.24: Typical storage trajectories of the farm dam in the second river reach during calibration phase.

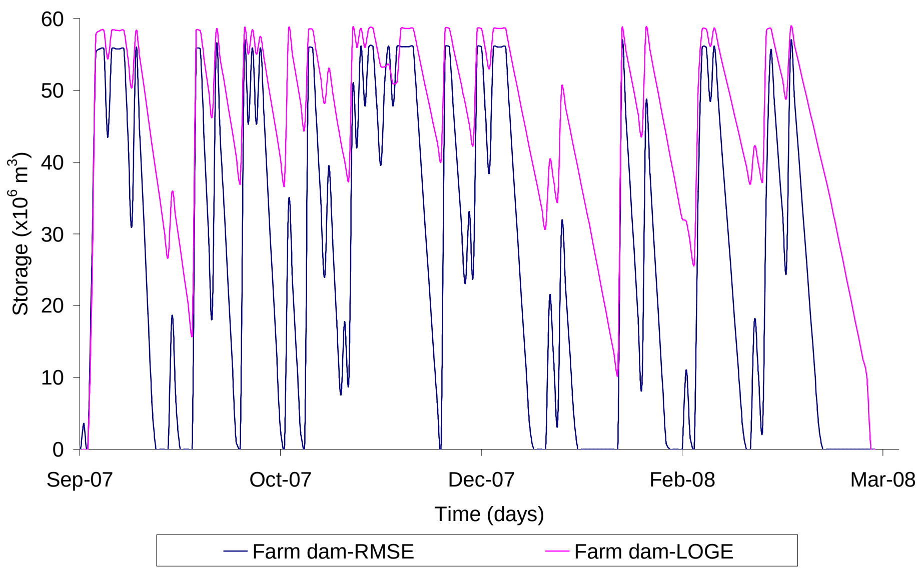


Figure 7.25: Typical storage trajectories of the farm dams in the second river reach during verification phase.

The discussion in the previous subsections is about the model structure. The various components of the models, interception storage, soil storage, farm dams and storage weirs are considered to represent the main catchments hydrological features realistically. The capacities of the storages are within the expected ranges and the water in the soil storage is the main contributor to evaporation than the water in the river. The zero values obtained for the farm dam storage trajectories indicate that farmers use groundwater to supplement the supply during the periods of low flows as discussed in section 4.4. Therefore, the reproduction of the catchment processes is considered realistic regardless of the insufficient data available to explicitly describe the system characteristics. The use of the soft data in conceptualising the system has been verified by the use of hard data, highlighting the need to use both types of data as part of the modelling process. The performance of the models in simulating daily streamflows is presented in the next Section.

7.3.3 Streamflow simulation performance of conceptual model

The simulation results indicate that the model performed better in simulating flows in the lower river reaches than the upper river reaches. Figure 7.26 through Figure 7.31 show the calibration and verification results of the models for all the three river reaches, while Figures 7.32 through Figure 7.37 show the same results on a logarithmic scale.

a. Conceptual model performance based on statistics

Statistically, the model performance based on *CCoef*, indicates an improvement towards the downstream (Table 7.3) for the two objective functions (RMSE and LOGE). During the calibration phase and when RMSE is used as an objective function, the values of the *CCoef* for the first, second and third river reach are 0.23, 0.52 and 0.86 respectively. For the verification phase and when the RMSE is used as an objective function, the values obtained are 0.09, 0.42 and 0.94 for the first, second and third river reach respectively. The satisfactory performance of the model in simulating the flows in the third river reach shows that the model is capable of satisfactorily reflecting the influence of the alluvial aquifer. The unsatisfactory performance in the first river reach can be attributed to the model's inability to adequately model the human induced processes such as the operation of the storage weirs and intermittent water abstraction. Also, the model structure has lumped storages representing weirs and farm dams whose overflows are likely not accurately represented. With respect to low flows, since most of the water is used for irrigation purpose, there might be some return flows which are not accounted for. Also, the loss component in the model has parameters which do not vary seasonally therefore during the period of low flows in winter period the losses are lower than those occurring in summer period.

The model generally underestimates the simulated values. The majority of the PBIAS values obtained are positive ranging from 1.45 % and 57.63 % during the calibration and verification phases for all the river reaches and based on the two objective functions (Table 7.3). Therefore, and based on the performance rating given in Table 6.1, the model's performance range from very good (1.45 %) to unsatisfactory (57.63 %).

A trend similar to that observed with the *CCoef* has been observed for the NSE values. While the NSE values obtained improved downstream, the NSE value obtained during the calibration phase are better than those obtained during verification phase. During the calibration phase based on the RMSE, the NSE values obtained are -2.22, 0.27 and 0.69 for the first, second and third river reach respectively. For the verification phase, based on the RMSE objective function, the NSE values obtained are -2.87, -0.14 and 0.78 for the first, second and third river reach respectively. Therefore, unsatisfactory model performances were obtained for the first and second river reaches and good and very good performances were obtained for the third river reach during the calibration and verification respectively (based on Table 6.2). A consideration of the NSE values obtained when LOGE is used as an objective function reveals that these were not better than those obtained when RMSE is used as an objective function.

The RSR values indicate a similar trend, with the average values found ranging from 0.334 to 2.50; with the very good values being for the third river reach. The general performance based on this model structure is satisfactory. The model performance in the upper reaches is significantly influenced by the complex natural catchment processes, human induced processes and the absence of data to adequately represent all the main processes. The model adequately attained a general set of parameters that enabled the satisfactory performance with respect to the simulation of the flows in the third river reach.

Due to the model's inability to satisfactorily simulate flow in the first river reach, it was considered unnecessary to use the model output for the first river reach as an input of the model setup to simulate flow in the second river reach as was the case with the linked models since the model output errors would be propagated.

Table 7.3: Performance of the conceptual model

River Reach	Phase	RMSE				LOGE			
		CCoef	NSE	PBIAS	RSR	CCoef	NSE	PBIAS	RSR
First	Calibration	0.23	-2.22	57.62	1.794	0.24	-5.24	11.17	2.50
	Verification	0.09	-2.87	12.54	1.789	0.11	-4.69	-8.64	2.17
Second	Calibration	0.52	0.27	4.00	0.839	0.5	0.25	13.15	0.851
	Verification	0.42	-0.14	-3	0.838	0.4	-0.12	1.45	0.832
Third	Calibration	0.86	0.69	2.16	0.57	0.86	0.68	26.75	0.584
	Verification	0.94	0.78	6.63	0.334	0.94	0.76	17.78	0.349

b. Conceptual model performance based on graphical plots

Generally the model underestimated the high flows for all the river reaches (Figures 7.32 through 7.37). When considering the simulations of the low flows, the model was observed to underestimate the low flow during the calibration and verification phases for the first river reach for both objective functions (Figure 7.32 and Figure 7.33). During the calibration of the second river reach, the model was observed to overestimate the low flows for both objective functions (Figure 7.34). However, during the verification phase of the second river reach, the model was observed to slightly overestimate the low flows in the early stages of the verification although the values obtained were close to the observed on the later stage of verification for both objective functions (Figure 7.35). For the third river reach, the low flow values obtained during the calibration and verification when the RSME was used as an objective function are higher than the observed values (Figure 7.36 and Figure 7.37). The use of LOGE as an objective function resulted in simulated low flow values that were close to the observed. These values indicated that the objective function, LOGE favours low flow simulation more than the objective function based on the RMSE. Furthermore, the difference in the simulated flows based on the two objective functions is more pronounced on the lower reaches than the upper reaches.

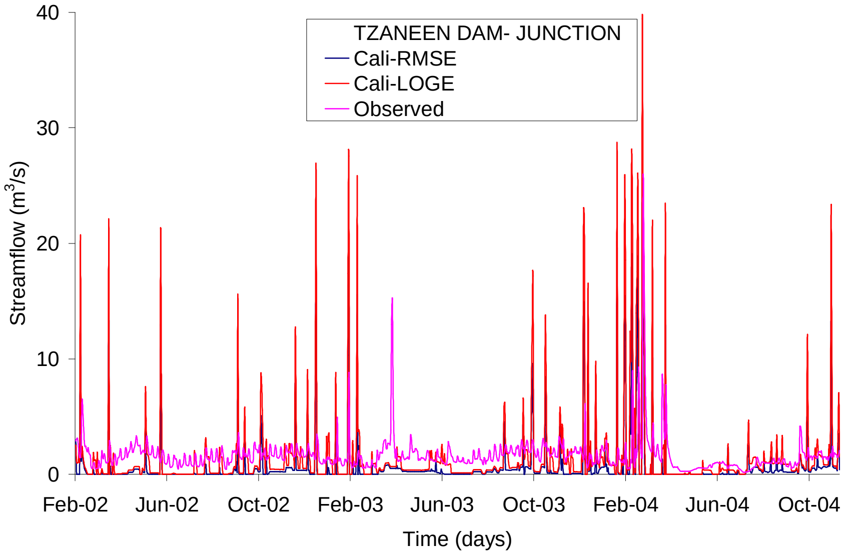


Figure 7.26: Observed and calibration of the conceptual model of the first river reach

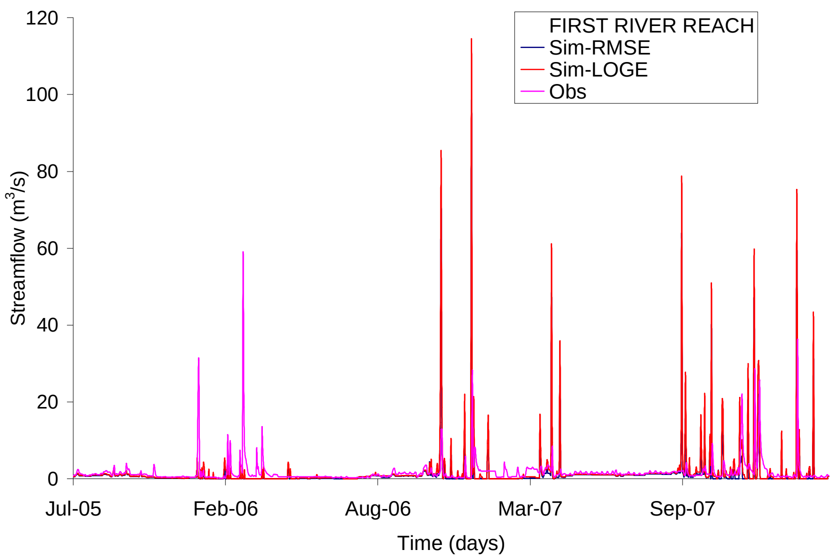


Figure 7.27: Observed and verification of the conceptual model of the first river reach

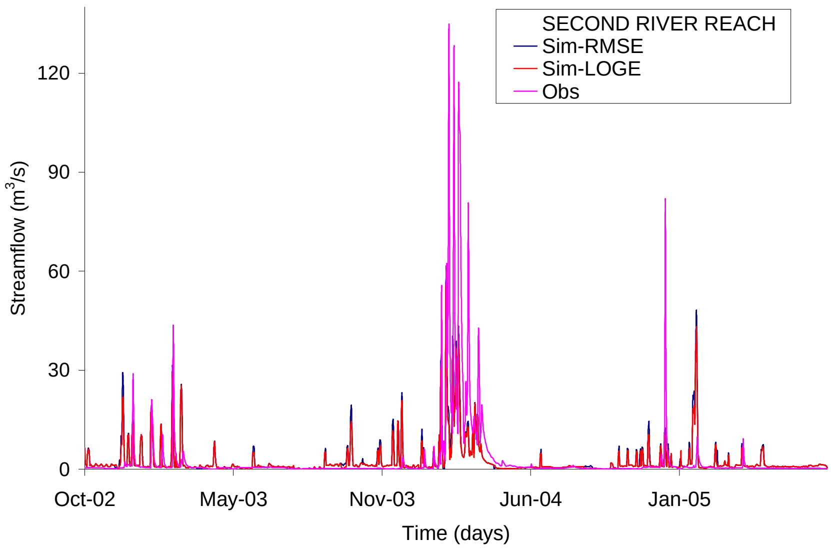


Figure 7.28: Observed and calibration of the conceptual model of the second river reach

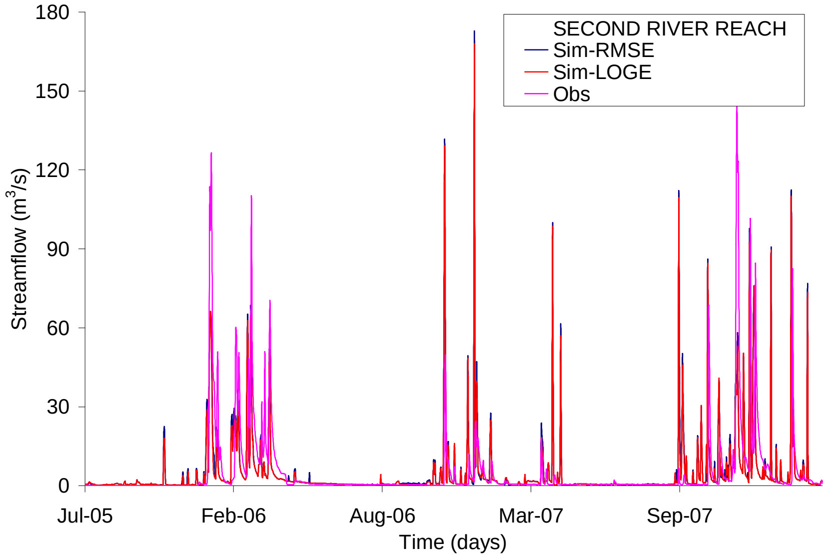


Figure 7.29: Observed and verification of the conceptual model of the second river reach

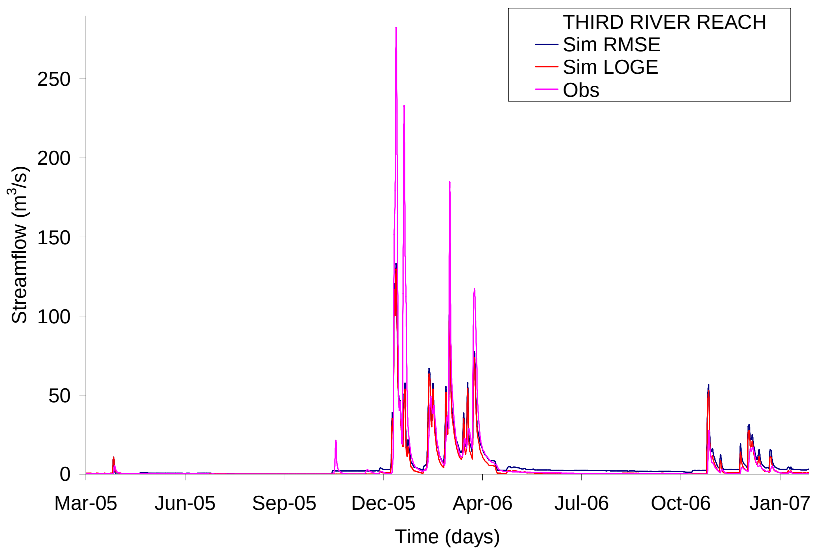


Figure 7.30: Observed and calibration of the conceptual model of the third river reach

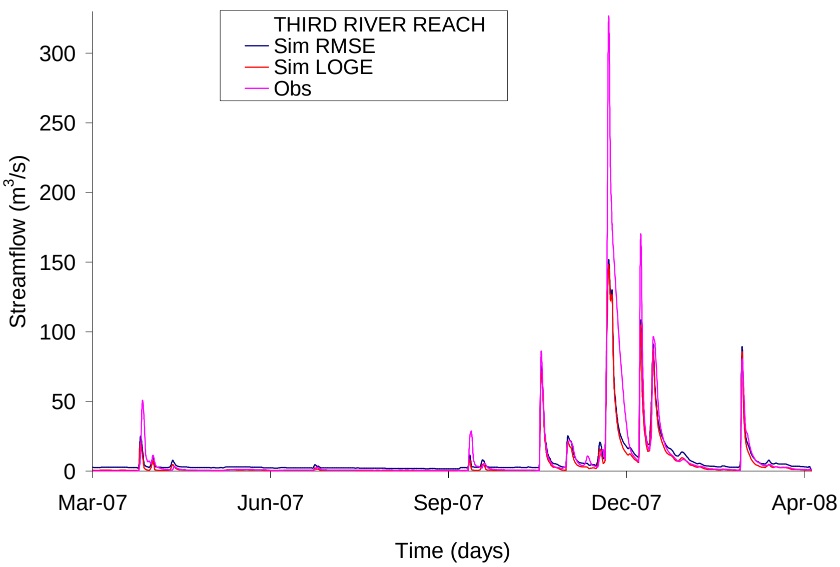


Figure 7.31 Observed and verification of the conceptual model of the third river reach

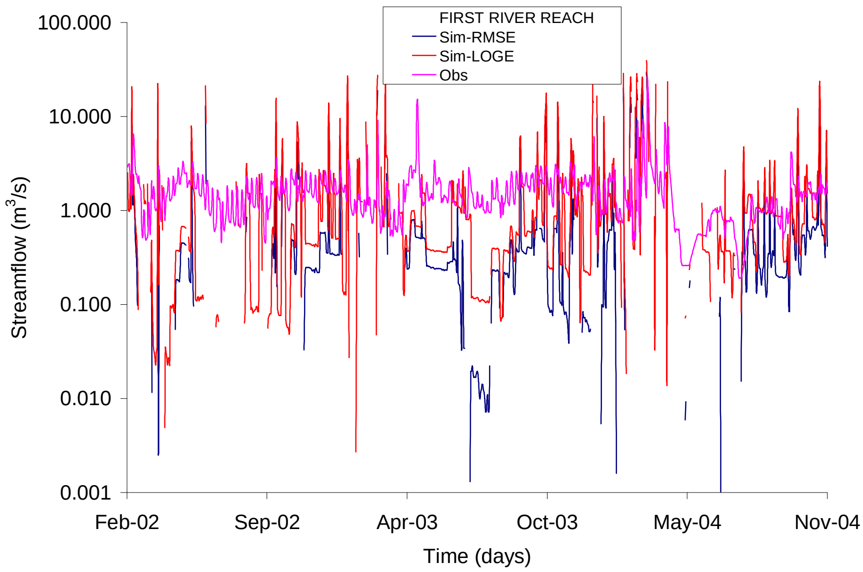


Figure 7.32: Observed and calibration of the conceptual model of the first river reach in a logarithmic scale.

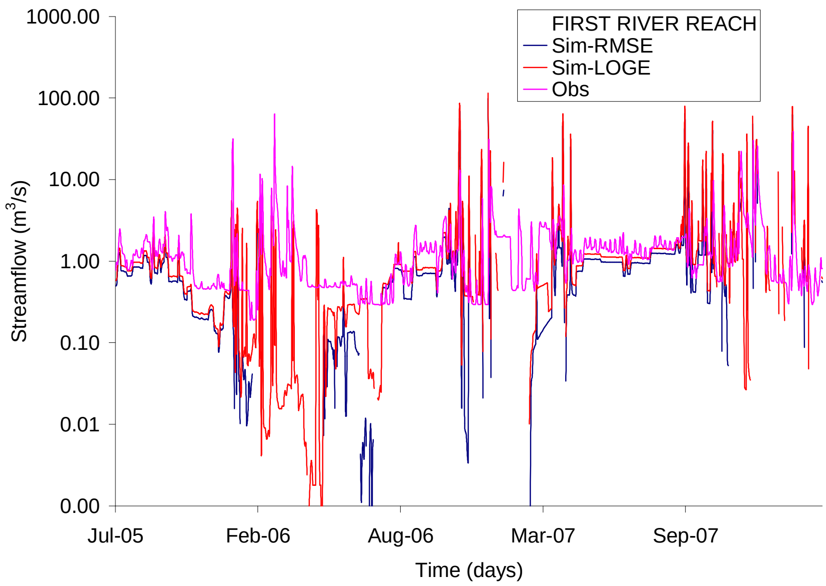


Figure 7.33: Observed and verification of the conceptual model of the first river reach in a logarithmic scale.

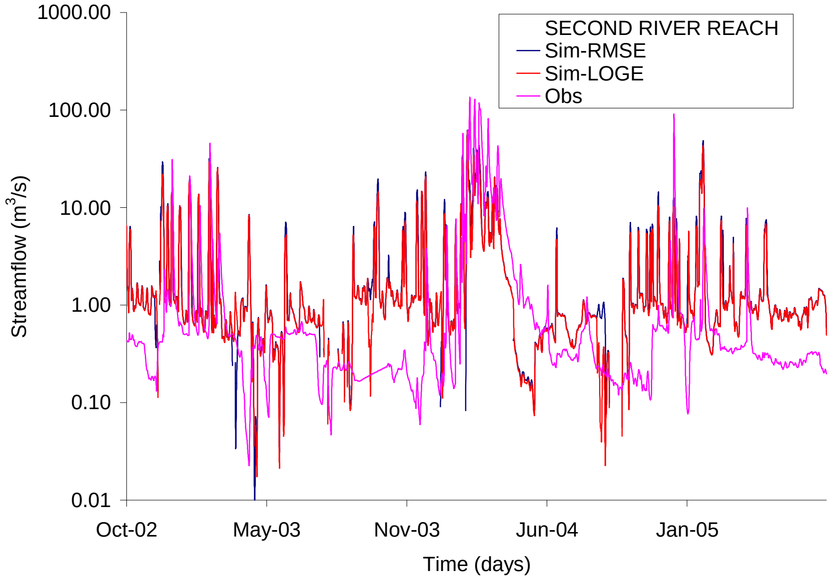


Figure 7.34: Observed and calibration of the conceptual model of the second river reach in a logarithmic scale.

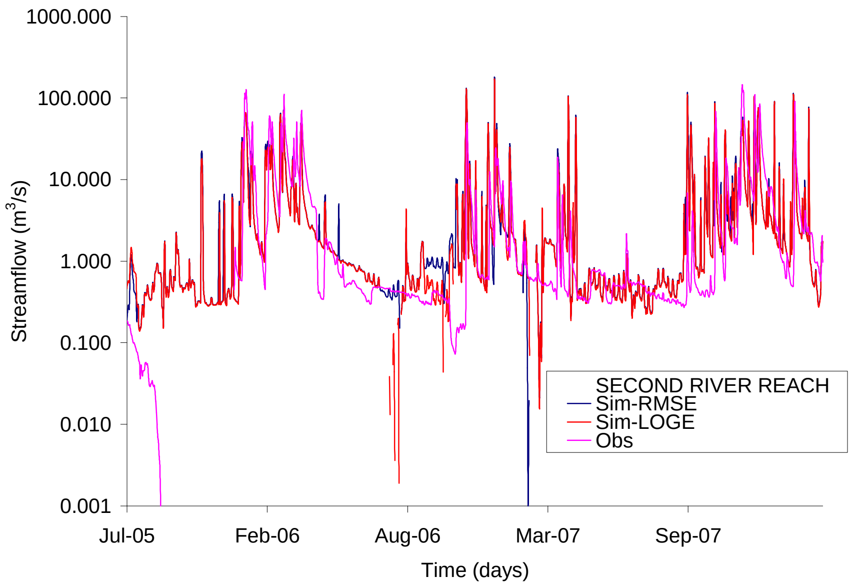


Figure 7.35: Observed and verification of the conceptual model of the second river reach in a logarithmic scale.

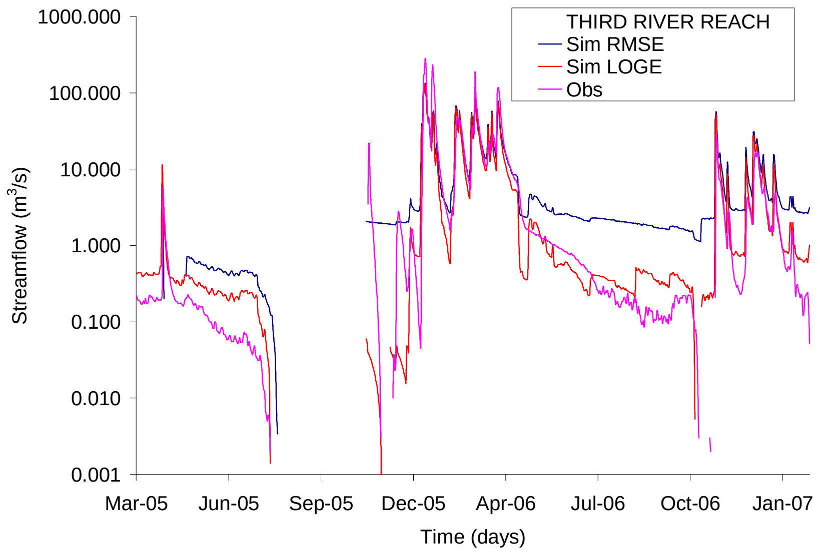


Figure 7.36: Observed and calibration of the conceptual model of the third river reach in a logarithmic scale.

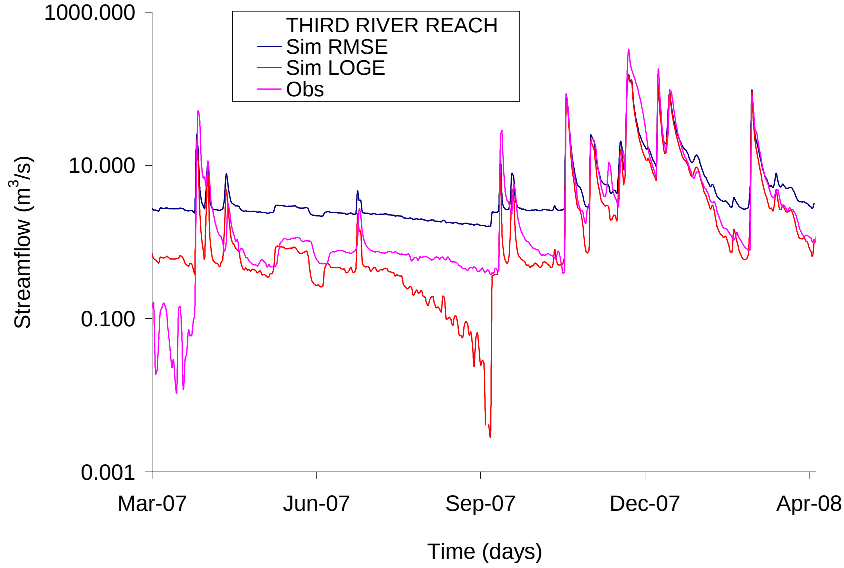


Figure 7.37: Observed and verification of the conceptual model of the third river reach in a logarithmic scale.

A discussion of the development and application of the hybrid conceptual-fuzzy inference model follows in the next Chapter and a comparison of the three modelling approaches is done in Chapter 9.

8 HYBRID CONCEPTUAL-FUZZY INFERENCE MODELLING FOR THE LETABA RIVER SYSTEM

8.1 HYBRID MODELLING

Standalone fuzzy inference and conceptual models have been used to model the Letaba River system (Chapter 6 and Chapter 7), however, a hybrid conceptual-fuzzy inference model that is based on conceptual and fuzzy inference model has been perceived to acquire the good traits from the conceptual and the fuzzy model that may outperform either of the standalone models.

The hybrid conceptual-fuzzy inference model combines the conceptual model structures developed in Chapter 7 and the fuzzy inference model approach developed in Chapter 6. The decision to undertake the development of the hybrid model was merited by the fact that both the fuzzy inference approach and the conceptual modelling approach have strengths and drawbacks associated with them. The fuzzy inference model component, just like any other data mining based technique, attempts to satisfy accuracy in the simulations while making no attempt to achieve comprehensible relations between its parameters and the system's features as both cannot be achieved simultaneously (Babovic et al., 2002, Tsukimoto, 2005). The conceptual model applied here, on the other hand, attempts to comprehensibly represent the catchment processes as detailed in Section 7.3.2 although its inability to satisfactorily map the input data to the output was poorer than that of the fuzzy model as the streamflow simulation results revealed. The fuzzy inference model component attempts to improve the accuracies in the simulation in a manner that implicitly takes care of the poorly understood processes and uncertainties in a fuzzy fashion and thereby obtained evidently better simulations than the stand-alone conceptual model.

This chapter is about the development and application of a hybrid model to simulate flow in the Letaba River system. The development of the hybrid model is described in Section 8.2 followed by a discussion of the results in section 8.3.

8.2 DEVELOPING HYBRID CONCEPTUAL-FUZZY INFERENCE MODELS

An analytical consideration of the conceptual model suggests that, each particular flux within a system has an influence on the flows but the significance of this influence varies from one flux to another. However, since the precise nature of the influence of the various fluxes is not certain, it is reasonable to suggest that the final output of a conceptual model is expected to inherit the uncertainties resulting from the different stages of the modelling process. The model calibration procedure implemented in this study influences the magnitude of the model parameter errors as part of implicit compensation process where the model structure error, input data errors are compensated through the adjustment of the model's parameters (Ewen and Parkin, 1996). With this mind, the flow simulation resulting from the conceptual component $q_{s,t}$, are perceived to represent the main hydrological processes. The effect that results from the rule of thumb operation of the dam or storage weirs has been incorporated as discussed in Section 6.2.1. Therefore the structure of the hybrid model can be represented as Figure 8.1:

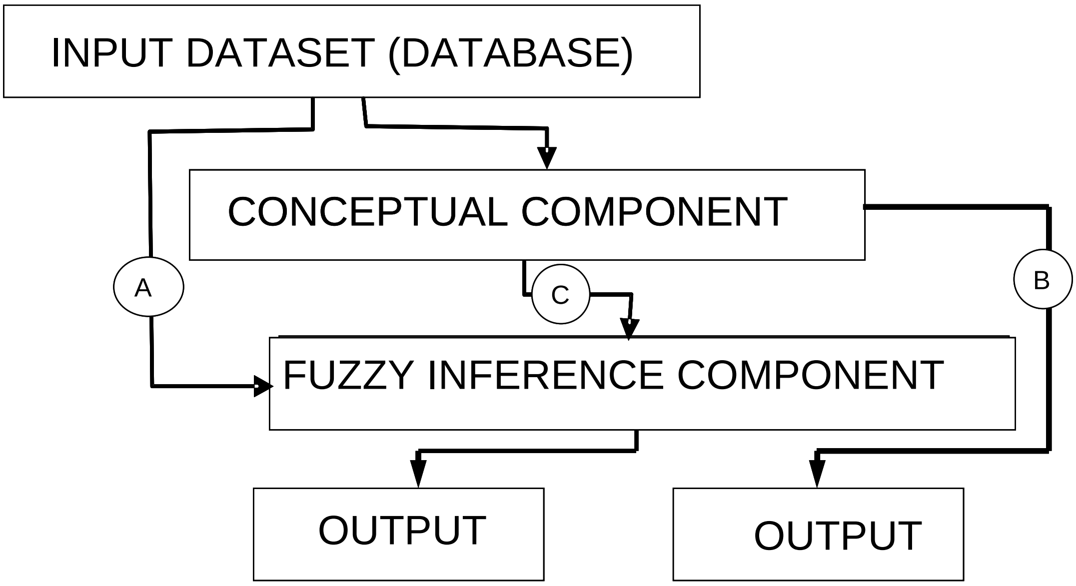


Figure 8.1: A schematic diagram for the hybrid model

(where A represents data $q_{in,t}$, $q_{out,t-1}$ and $q_{tri,t}$; B represents the other conceptual parameters and intermediate simulations such as trajectories, flow losses; and C represents the conceptual flow $q_{s,t}$).

The linkage between the conceptual model and the fuzzy inference model is represented by Equation 8.1. The respective fuzzy models therefore took the form of Equations 8.1 a-c for the first, second and third river reaches respectively.

$$q_{simt,i} = f\left(q_{in,t}, q_{s,t}, q_{out,i-1}\right) \qquad 8.1a$$

$$q_{simt,i} = f\left(q_{in,t}, q_{tri,t}, q_{s,t}, q_{out,i-1}\right) \qquad 8.1b$$

$$q_{simt,i} = f\left(q_{in,t}, q_{s,t}, q_{out,i-1}\right) \qquad 8.1c$$

where $q_{sim,t}$ is the simulated flow, $q_{in,t}$ is the flow into the reach, q_t is the outflow from the conceptual component, $q_{out,i-1}$ is the lagged observed outflow. The detailed description of the fuzzy inference model is given in Chapter 6 and in Katambara and Ndiritu (2009). The SCE-UA algorithm (Duan et al., 1992) has been used to calibrate the model.

By comparing equation 6.1 and 8.1, the role of the conceptual modelling becomes clear but it seems that the stand-alone conceptual model in Chapter 7 is actually solving a different and more difficult modelling problem (as it does not use the observed flows anywhere). It is however valuable to know that with the conceptual modelling in the hybrid model, all the data transformations that were done for the stand-alone fuzzy model are not required thereby demonstrating that the hybrid modelling is physically much more realistic. It was decided to include the lagged observed flows in the hybrid modelling as this had also been included in the standalone fuzzy model. Further analysis that excludes the lagged observed flows thereby making the model more versatile need to be investigated at a later stage. The results and the discussion then follows.

8.3 RESULTS AND DISCUSSION OF THE HYBRID MODEL

The application of the hybrid model aims at conceptually representing the hydrological process while simultaneously improving the accuracies in the simulations. The various processes including the calibration adequacy, ability to reproduce the catchment processes flow simulations are therefore discussed in the following sub-sections.

8.3.1 Calibration adequacy of the hybrid model

The adequacy of the calibration process has been attained by conducting several model runs in such a manner that the values obtained for the objective function remains the same (Figure 8.2).

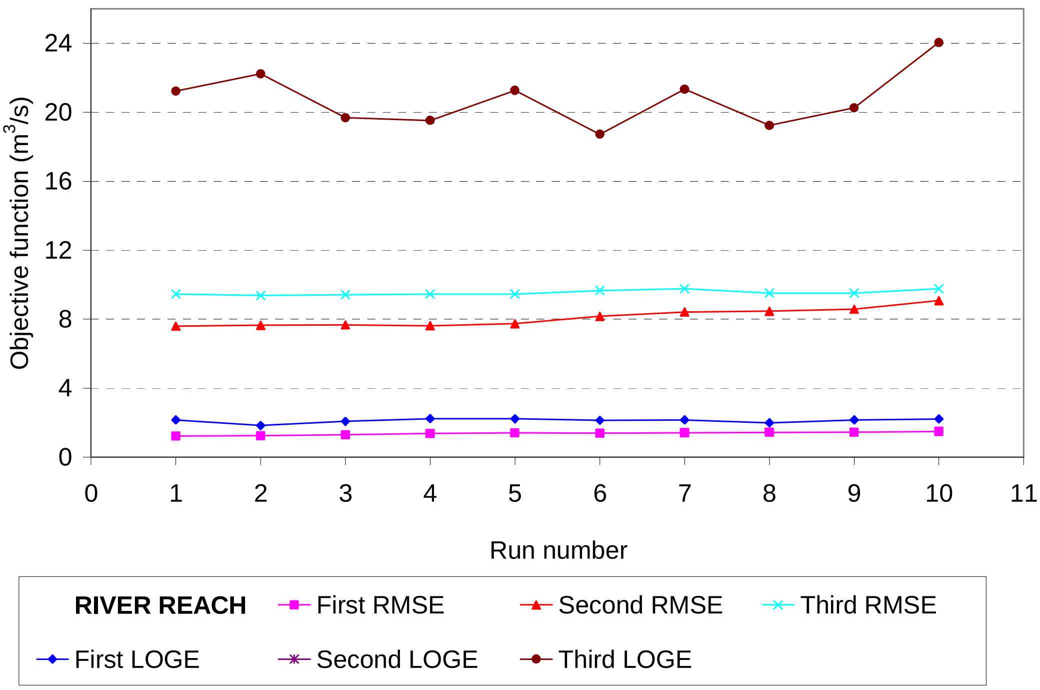


Figure 8.2: Typical values of the objective function obtained for the hybrid model for the various randomly initialised calibration runs

8.3.2 Reproduction of Catchment Processes by hybrid models

The absence of data to represent the physical characteristics of the Letaba River system necessitated the modelling of some of these characteristics such as the capacities of the farm dam and storage weirs so as to obtain simulated estimates of these capacities. The conceptual model component was used to identify some of these representative parameter values. The need for increased accuracy in the simulations necessitated the inclusion of the fuzzy inference model components in the hybrid. While the simulated time series (e.g. storage trajectories, abstraction etc) as well as the calibrated capacities of the farm dam and storage weirs indicate the probable ranges of magnitudes of the various components of the system and the parameter values are dependent on the objective function, they have been compared to the reported values.

a. Calibrated interception storage by hybrid model

The initial catchment process that takes place after rainfall is interception and is modelled as a calibrated conceptual storage. Figure 8.3 shows the capacities of the interception storages obtained when ten best hybrid model calibration are performed on each of the three river reaches, while Figure 8.4 shows the capacities of the interception storages obtained for the connected river reaches when ten calibration runs are made. The variation of the values in the Figures 8.3 and Figure 8.4 indicates that the variation of the capacities in the first river reach is insignificant when compared to the other reaches. Higher capacities were obtained for lower reaches than the upper river reaches suggesting that the model managed to obtain values that agree with what the literature suggests i.e. that higher losses occur in the lower reaches where high temperatures are experienced (DWAF, 2004). The model managed to achieve estimates that have insignificant differences when linked and also when it was not linked. The range of values obtained by the model are realistic and are comparable (as an order of magnitude) to the values reported in other studies (e.g. Klaassen et al., 1998, Liu et al., 2002, Granger and Gray, 1990, Qiu et al., 1998) where values that are slightly above 7 mm were reported.

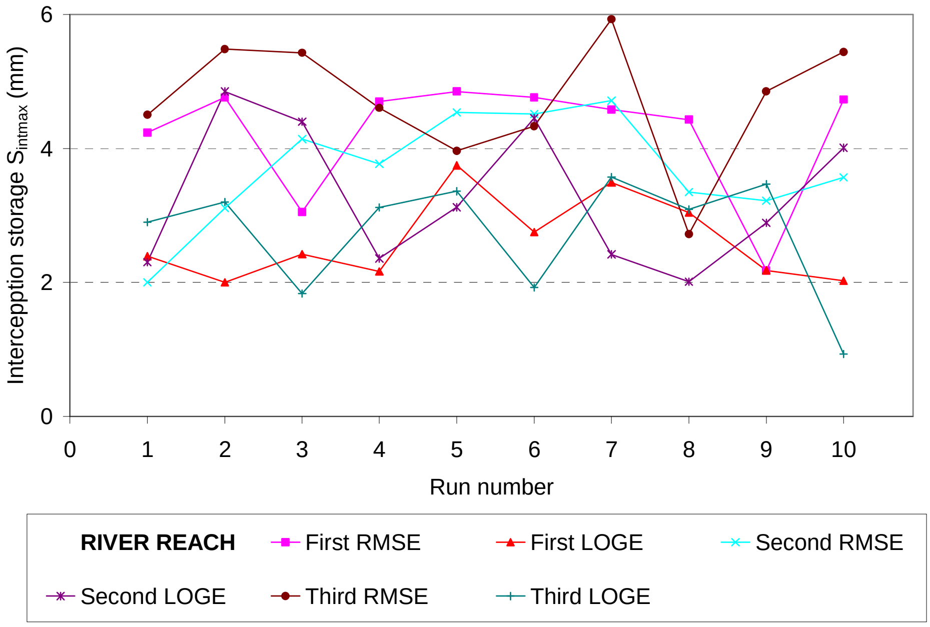


Figure 8.3: Values for the interception storage for the unconnected hybrid model obtained for all the river reaches after model 10 runs

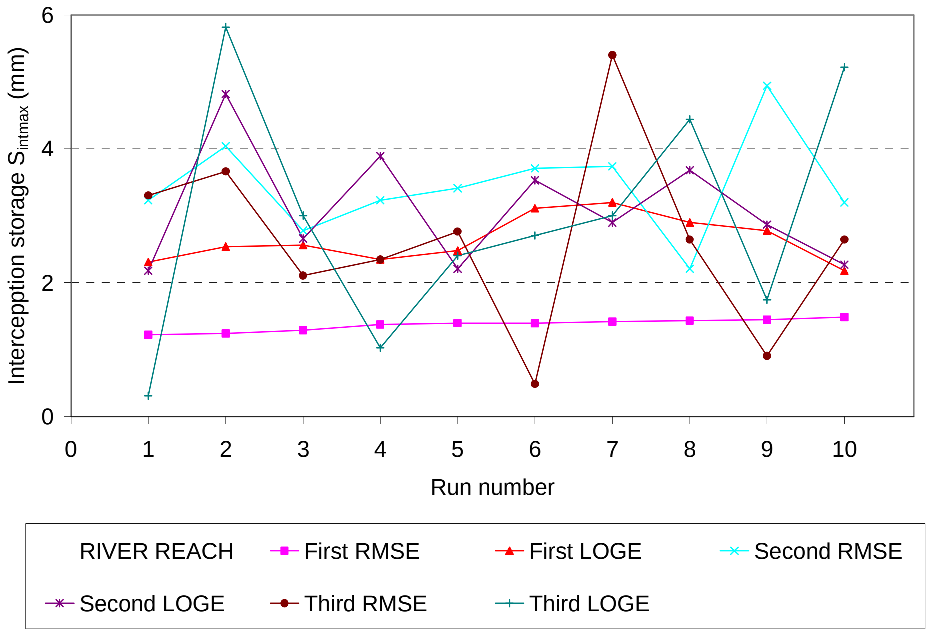


Figure 8.4: Typical values for the connected interception storage for the hybrid model obtained for all the river reaches after model 10 runs

b. Calibrated soil storage by hybrid model

The soil stores water and as such contributes to the amount of water stored in the catchment. The water that contributes to the soil storage is excess rainfall remaining after the interception storage is filled. This water is available as subsurface flow, overland flow and direct evapotranspiration losses. Figure 8.5 and Figure 8.6 show the capacity of the soil storages for the river reaches based on the two objective functions for the stand-alone and connected river reaches. The maximum values obtained in Figure 8.5 and Figure 8.6 are lower than what was expected based on what was reported in other studies (e.g. Tan and O'Connor, 1996, Fenicia et al., 2006) where the soil storage was found to be as high as 400 mm (section 7.3.2). The obtained soil storage values are low than those reported in other studies, however the soil storage contributes to the evapotranspiration and subsurface flow.

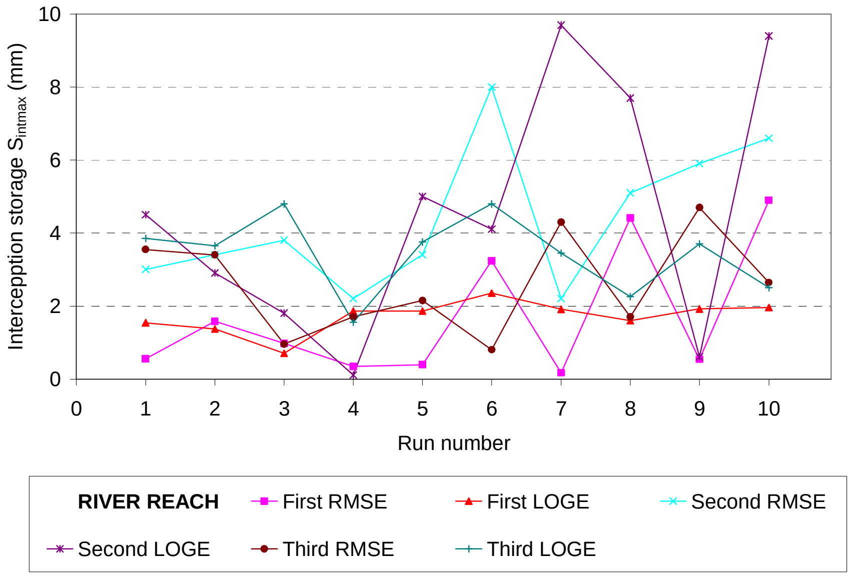


Figure 8.5: Capacity of soil storage for the stand-alone model obtained for all the river reaches for 10 calibration runs

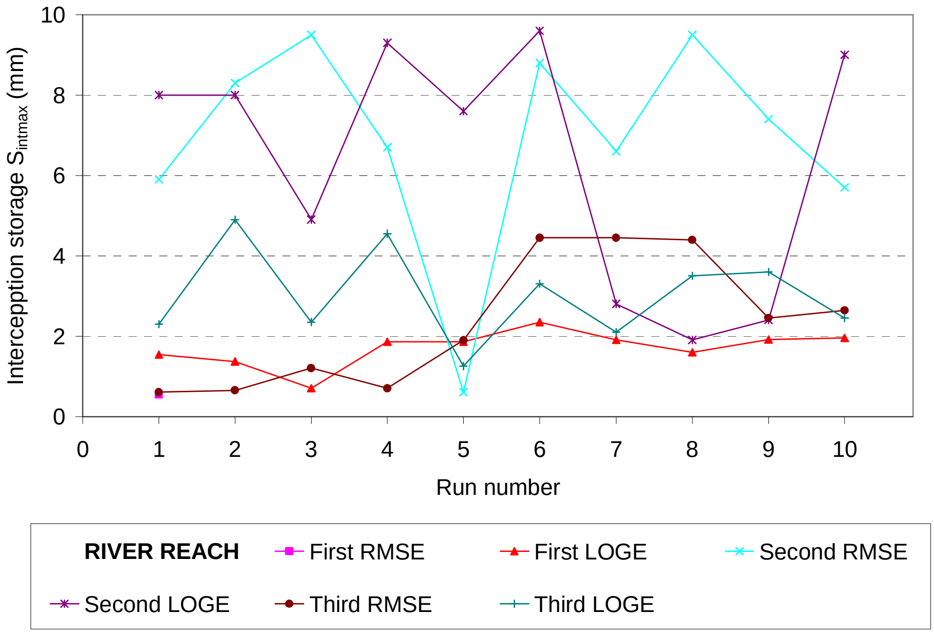


Figure 8.6: Capacity of soil storage for the connected model obtained for all the river reaches for 10 calibration runs

c. Calibrated capacities of storage weirs and farm dams by the hybrid model

A similar approach to that used in Chapter 7 for determining the capacity of the farm dams and storage weirs was used. Figure 8.7 and Figure 8.8 show the capacities of the storage weirs and farm dams obtained for the individual river reaches for 10 calibration runs. The values obtained indicate that the capacity of the storage weirs is higher for the second river reach than for the first river reach with values ranging from $0.2\text{x}10^6$ m^3 to 0.7 $\text{x}10^6$ m^3 for the first river reach and from 0.11 $\text{x}10^6$ m^3 to 3.4 $\text{x}10^6$ m^3 for the second river reach. The capacities of the farm dams in the first river reach range from $0.1\text{x}10^6$ m^3 to $2\text{x}10^6$ m^3 and from 6.7 $\text{x}10^6$ m^3 to 63 $\text{x}10^6$ m^3 for the second river reach. These capacity values obtained for all runs are comparable to what has been obtained in Chapter 7 suggesting that the model attained realistic values.

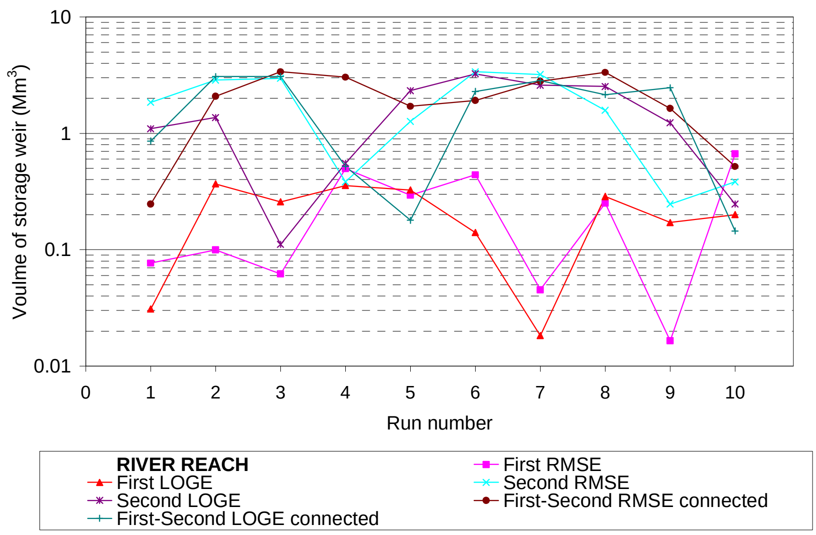


Figure 8.7: Typical values for the storage capacity of the weirs for the hybrid model

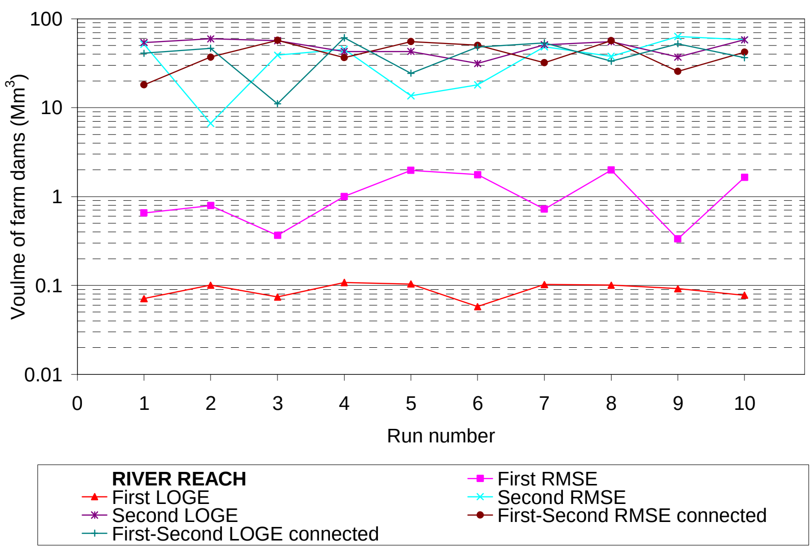


Figure 8.8: Typical values for the storage capacity of farm dams for the hybrid model

d. Calibrated Stream flow loss parameters by the hybrid model

The parameters values obtained for the linear flow loss function (Equation 7.28) for all the 10 runs are not of the same value (Figures 8.9 and 8.10). The values of the slopes obtained range from 0.4 to 0.5. A significant variation was observed in the values of the constant with values ranging from -5 to 7.5. Although, currently there are no strong reasons to explain these variations other than the influence of fuzzy inference model component's attempt to achieve more accurate values of the flow simulations.

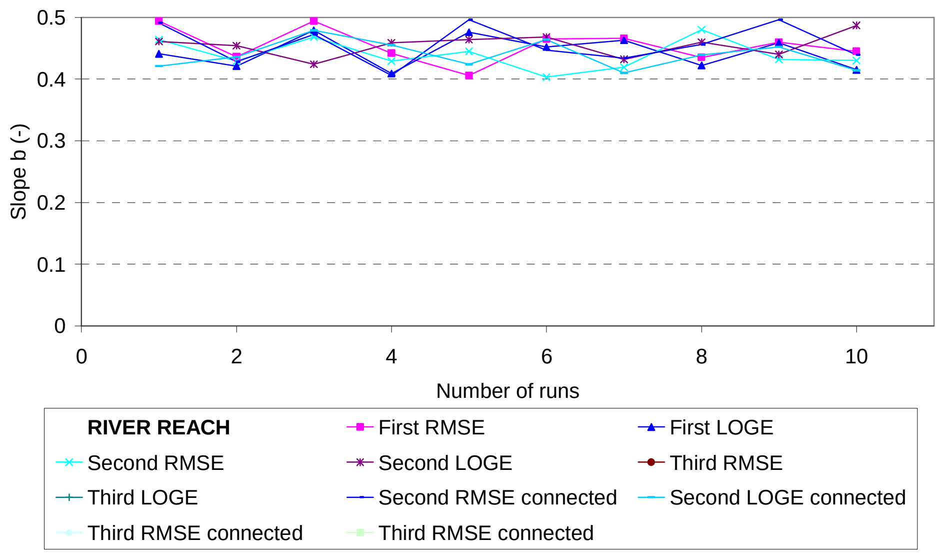


Figure 8.9: The slope values obtained for all the connected and unconnected river reaches for the hybrid model

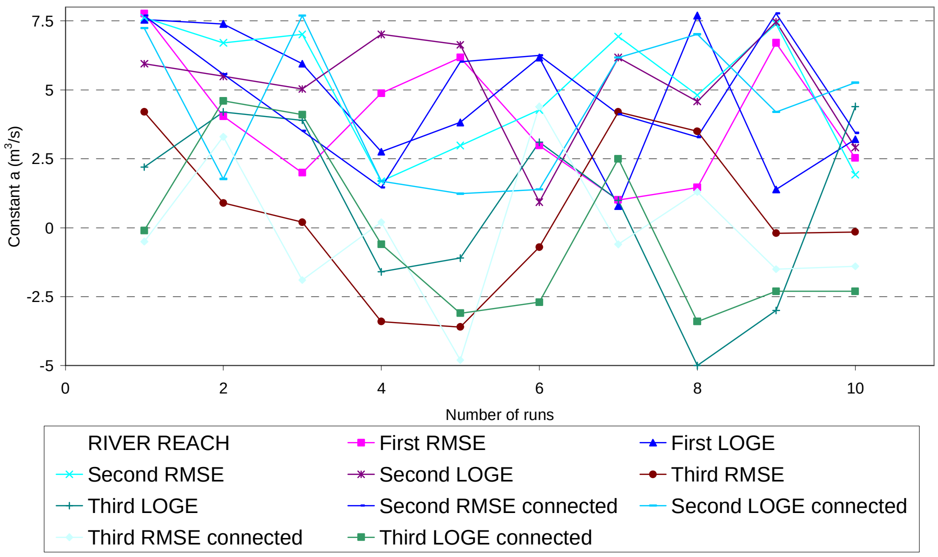


Figure 8.10: The constant values obtained for all the connected and unconnected river reaches for the hybrid model

e. Simulated water use from farm dams by hybrid model

Part of the catchment area of the first and second river reaches drains into farm dams. These farm dams are mainly for irrigation purposes. Figures 8.11 through 8.14 show the simulated values obtained for the amount of water abstracted from the reservoirs during the calibration and verification phase. The values obtained suggest that the two objective functions obtained values of nearly the same value. For those simulated water abstractions that are zero, it is reasonable to expect that farmers make use the groundwater to supplement the irrigation requirement suggesting that the simulations are realistic.

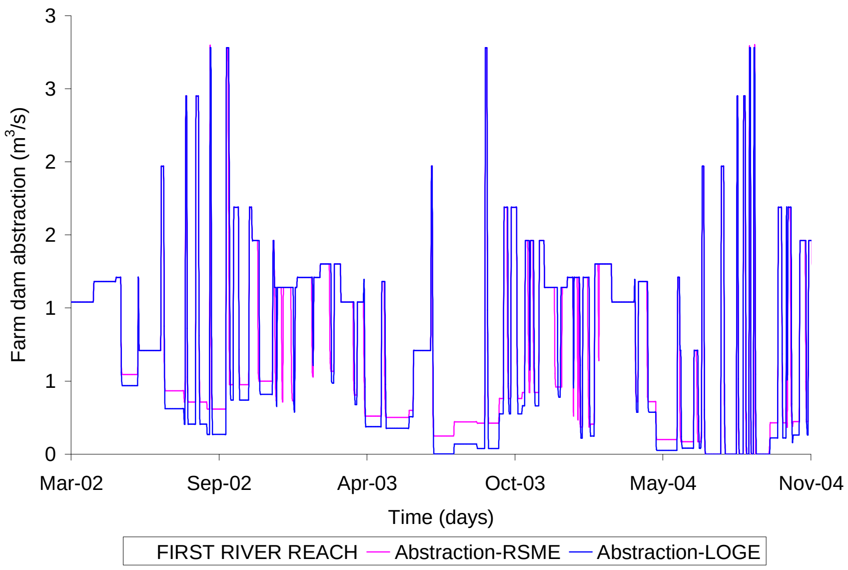


Figure 8.11: The abstraction series obtained by the hybrid model for the first river reach during the calibration phase

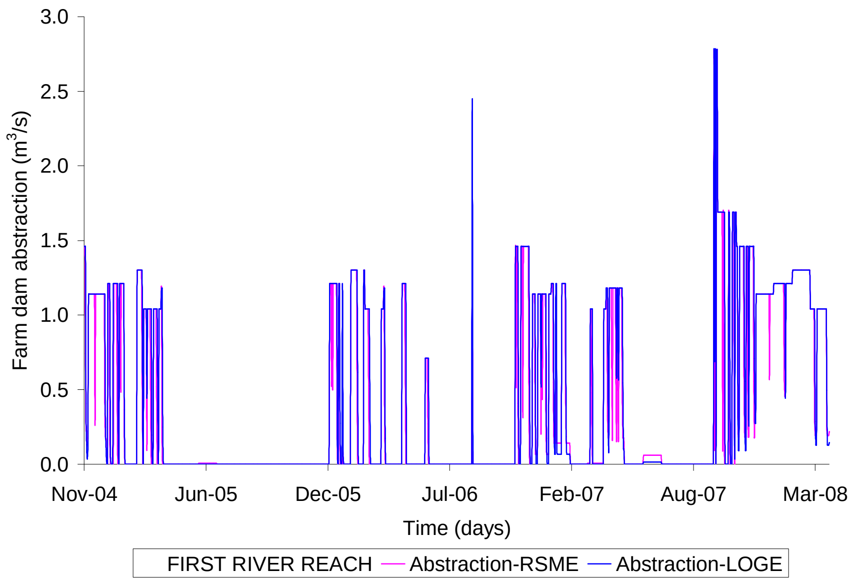


Figure 8.12: The abstraction series obtained by the hybrid model for the first river reach during the verification phase

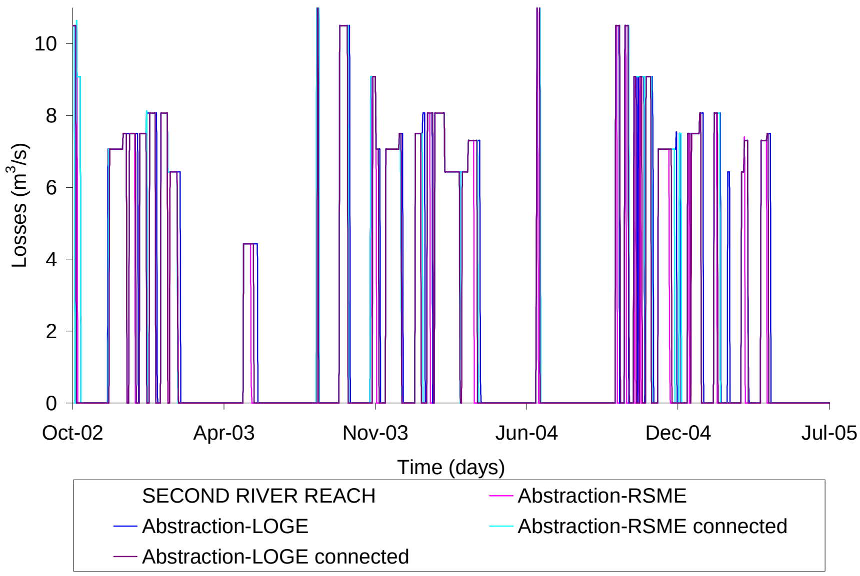


Figure 8.13: The abstraction series obtained by the hybrid model for the second river reach during the calibration phase

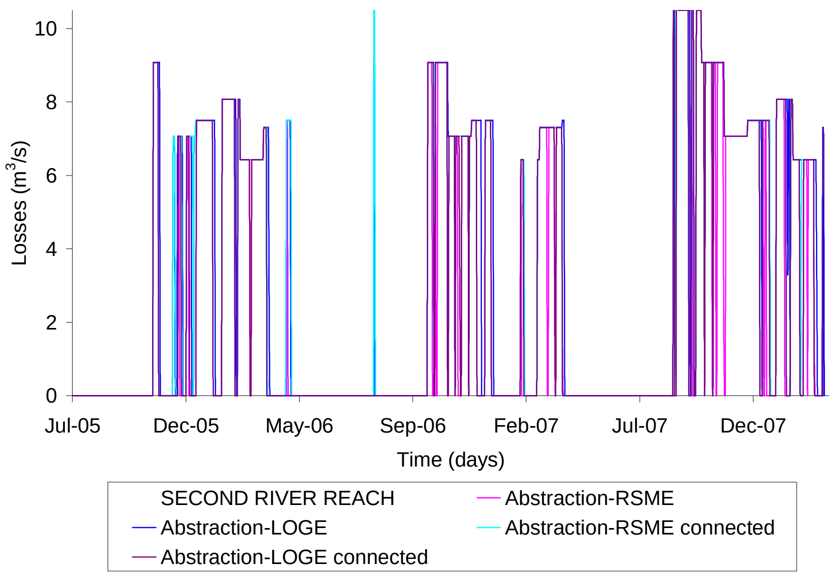


Figure 8.14: The abstraction series obtained by the hybrid model for the second river reach during the calibration phase

f. Simulated trajectories of the Farm dams and storage weirs

Figure 8.15 and Figure 8.16 show the trajectories of the storage weirs during the model calibration and verification phases respectively and Figure 8.17 and Figure 8.18 show the modelled status of the farm dams during the model calibration and verification phases respectively. During the calibration and verification phases of the storage weirs, the trajectory obtained when LOGE was used as an objective function were less than the trajectory obtained when the RMSE was used as an objective function. This is different from what was obtained for the farm dam capacities. The trajectory obtained when LOGE was used as an objective function has higher values than those obtained when RMSE was used as an objective function. The influence of the objective function on the simulation results may be responsible for these variations.

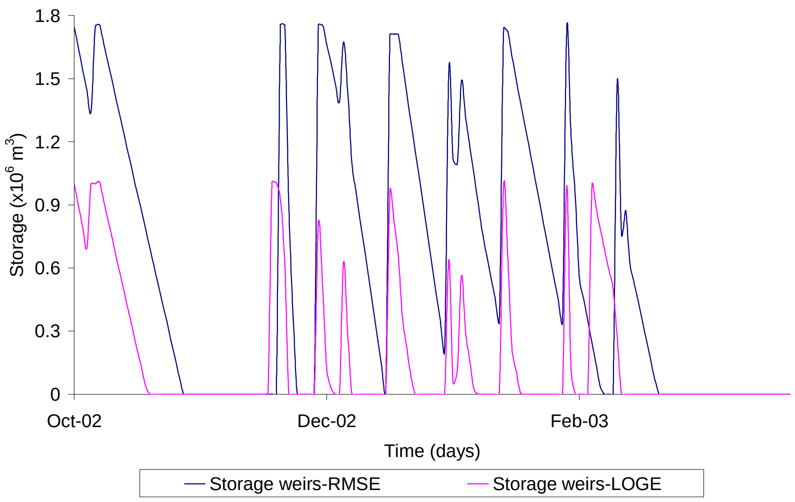


Figure 8.15: Typical storage trajectories of the storage weir in the second river reach during calibration phase

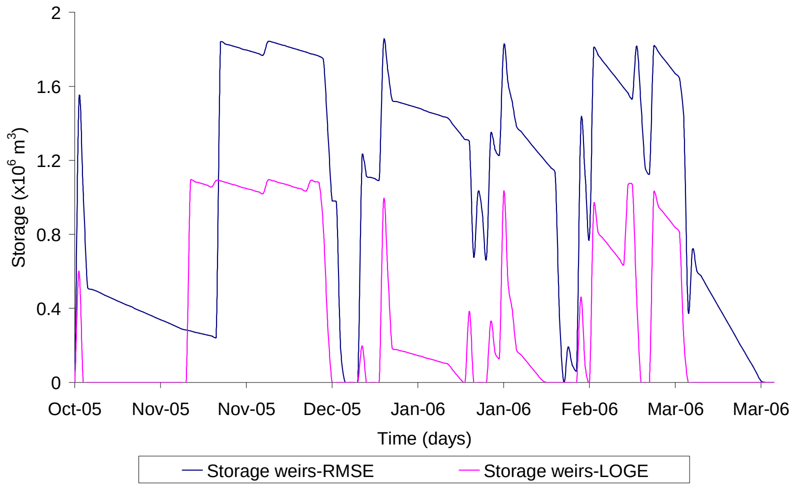


Figure 8.16: Typical storage trajectories of the storage weirs in the second river reach during verification phase

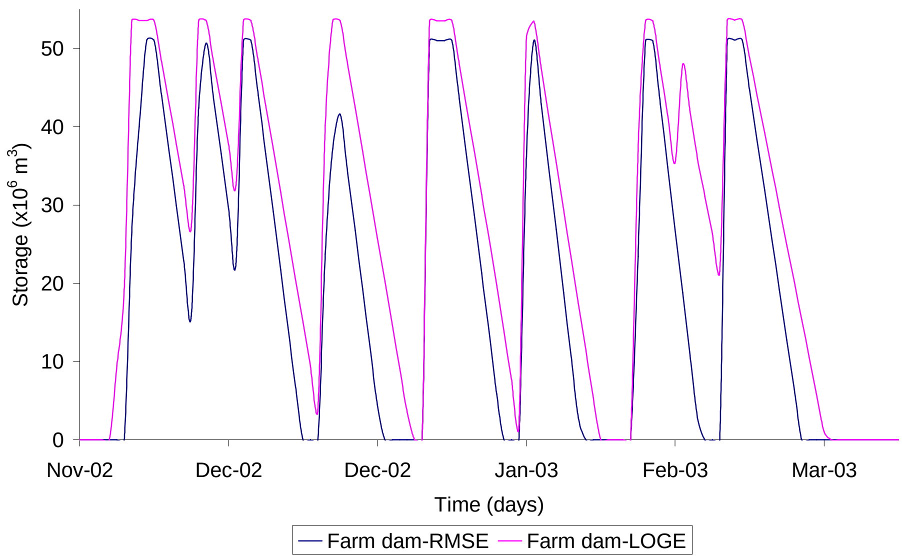


Figure 8.17: Typical storage trajectories of the farm dams in the second river reach during calibration phase

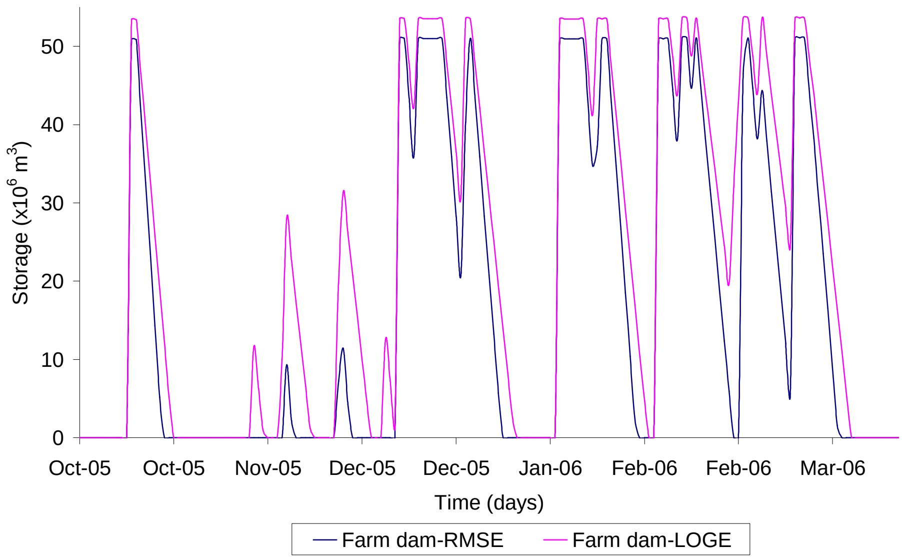


Figure 8.18: Typical storage trajectories of the farm dams in the second river reach during verification phase

g. Summary

The discussion in this section was with respect to the reproduction of catchment processes and various parameters of the model. The values obtained for the parameters were within the acceptable range. The various components of the models including the interception storage, soil storage, farm dams and storage weirs realistically represent catchments characteristics. The capacities of the storages within the catchment were observed to vary while it was also observed that the soil storage contributed more to evaporation than to the river flows. Visual and statistical evaluation of the performance of the models in simulating daily streamflows now follows.

8.3.3 Streamflow simulation performance of hybrid model

The simulation results indicate that the hybrid model performed better in the lower river reaches than the upper river reaches. Figures 8.19 through Figure 8.24 show the calibration and verification results of the models for all the three river reaches, while Figures 8.25 through Figure 8.28 show the calibration and verification of the connected river reaches. The results are based on the two objective functions, RMSE and LOGE. Figure 8.29 through 8.38 show all the values on logarithmic scale for the individual reaches and connected reaches during the calibration and verification phase.

a. Hybrid model performance based on statistics

Table 8.1 shows the performance of the hybrid model based on four performance measures. The performance values based on *CCoef* indicate an improvement towards the downstream for all the objective functions (RMSE and LOGE). There is an insignificant variation in the values obtained by the two objective functions (RMSE and LOGE). The *CCoef* values obtained range between 0.41 and 0.86 with the lowest value being for the first river reach and the highest being for the third river reach. The insignificant human activities in the third river reach and the model's ability to model the hydrological processes in the third river reach explain the good performance in this reach. The unsatisfactory performance of the model observed with respect to the first river reach is attributed to the model's inability to adequately model the human induced processes such as the operation of the storage weirs and intermittent water abstraction. Seibert and McDonnel (2000) argued that it is worthy accepting lower model performance values if a more real model of catchment behaviour is developed or used. It is also important to note though that the performance may be further improved if the monitoring network is improved including water abstraction activities and the operation of the individually and non-individually owned water storage facilities.

In general, the model overestimates the simulated values. The majority of the PBIAS values obtained are negative with values of up to 81.9 % (Table 8.1) for all the river reaches and based on the two objective functions. The obtained NSE values have a similar trend to that of the *CCoef* values. The NSE values obtained improved downstream. The average NSE values obtained when RMSE was used as an objective function are 0.58 and 0.52 for the calibration and verification respectively, while NSE values of 0.54 and 0.45 were obtained when LOGE was used as an objective function for the calibration and verification respectively.

The RSR values indicate a similar trend to that observed in the NSE and CCoef values. The average RSR values obtained when the RSME was used as an objective function are 0.63 and 0.52 for the calibration and verification respectively, while average values of 0.65 and 0.56 were observed when LOGE was used as an objective function for the calibration and verification respectively. Therefore, the RSR values obtained followed a similar trend to that observed in the NSE and CCoef values obtained.

The influence of the alluvial aquifer has been satisfactorily modelled. With respect to the performance, none of the objective functions significantly outperformed the other.

Table 8.1: Performance of the hybrid model

River Reach	Phase	RMSE				LOGE			
		CCoef	NSE	PBIAS	RSR	CCoef	NSE	PBIAS	RSR
FIRST	Calibration	0.62	0.36	-13.6	0.80	0.62	0.20	-35.6	0.89
	Verification	0.42	0.15	1.76	0.84	0.41	0.09	-8.58	0.87
SECOND	Calibration	0.75	0.58	-46. 7	0.66	0.75	0.51	-81.9	0.70
	Verification	0.69	0.29	-26.4	0.66	0.72	0.13	-44.9	0.73
THIRD	Calibration	0.91	0.84	-11.3	0.41	0.91	0.85	-9.84	0.40
	Verification	0.96	0.92	-0.19	0.20	0.94	0.90	-0.53	0.23
Connected									
SECOND	Calibration	0.74	0.45	-70	0.71	0.71	0.40	-75.3	0.72
	Verification	0.73	0.49	-31.5	0.56	0.75	0.38	-37.9	0.62
THIRD	Calibration	0.81	0.68	-32.8	0.58	0.85	0.74	-23.8	0.52
	Verification	0.85	0.74	-12.8	0.36	0.86	0.76	-9.26	0.35

b. Hybrid model performance based on graphical plots

The model was observed to overestimate the high flows for all the river reaches (Figures 8.19 through Figure 8.38). With respect to low flow simulation, it was observed that the model generally overestimates flows based on the two objective functions for the first and the second reach that could be attributed to an unsatisfactory model performance due to the impact of human activities. However, simulations obtained for the third reach where the need for water to meet the ecological requirements is significant, indicates a good performance.

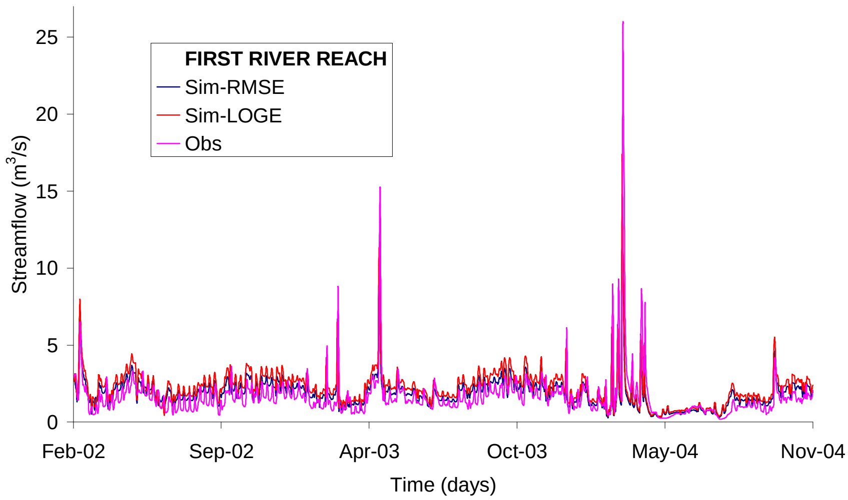


Figure 8.19: Observed and calibration of the hybrid model of the first river

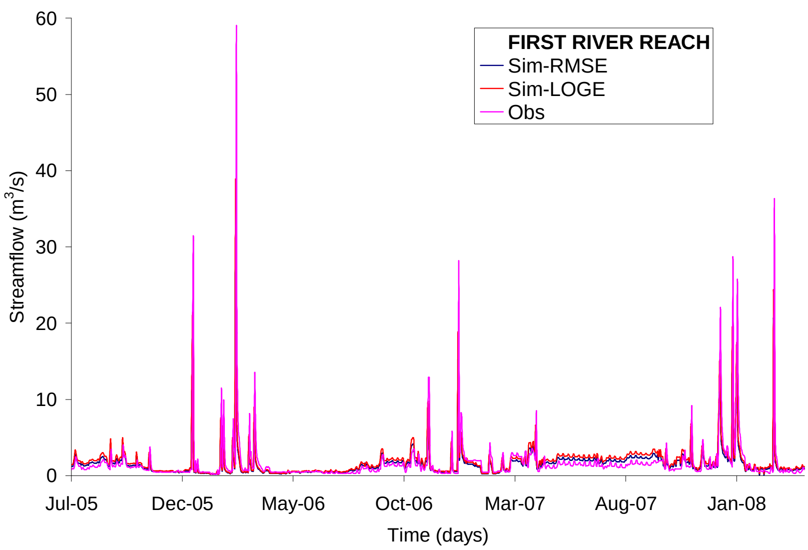


Figure 8.20: Observed and verification of the hybrid model of the first river reach

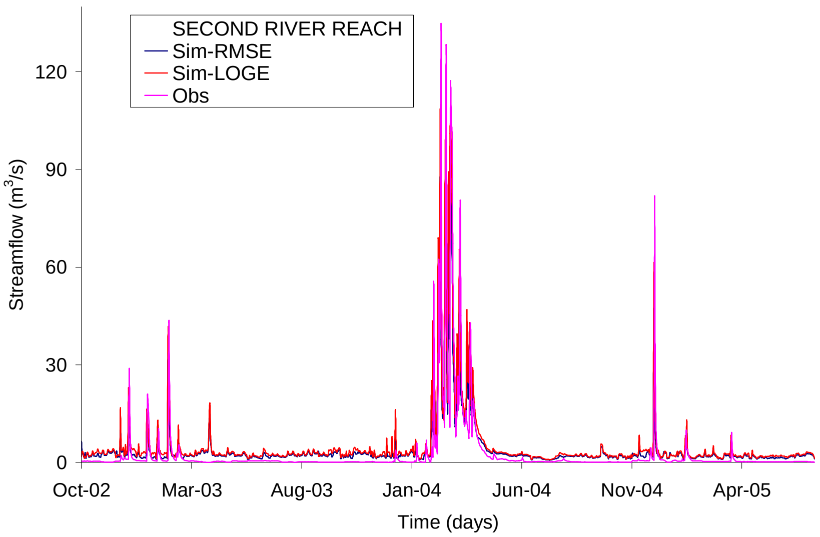


Figure 8.21: Observed and calibration of the hybrid model of the second river reach

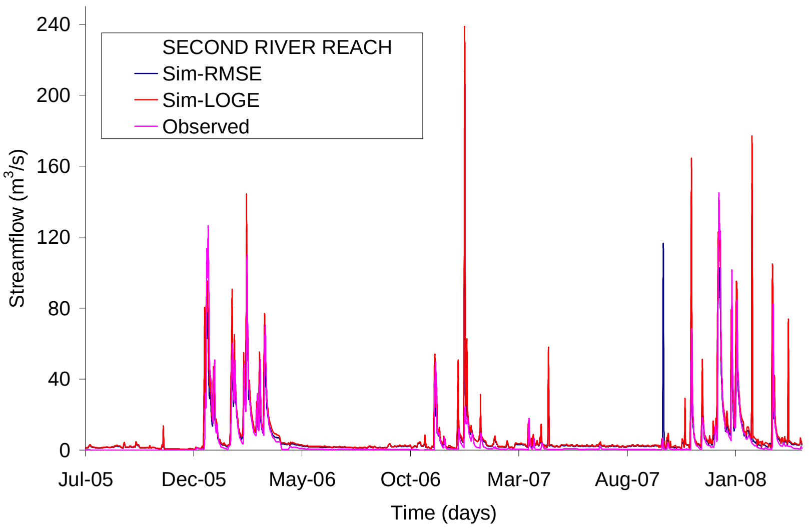


Figure 8.22: Observed and verification of the hybrid model of the second river reach

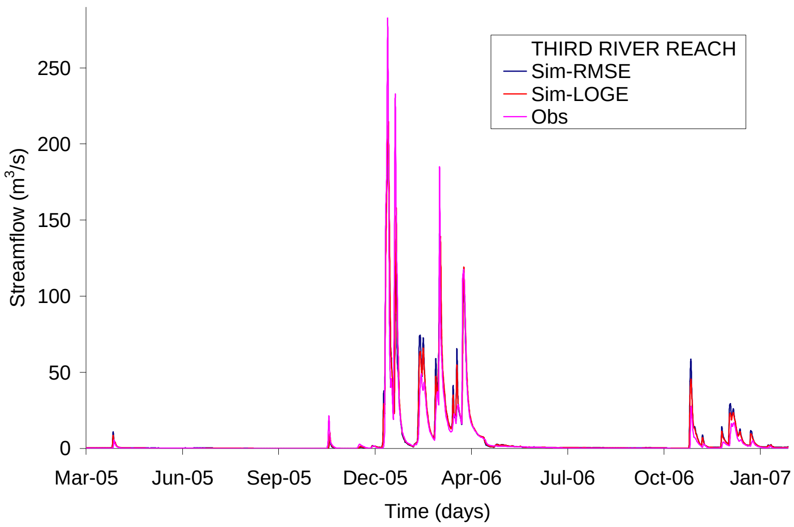


Figure 8.23: Observed and calibration of the hybrid model of the third river reach

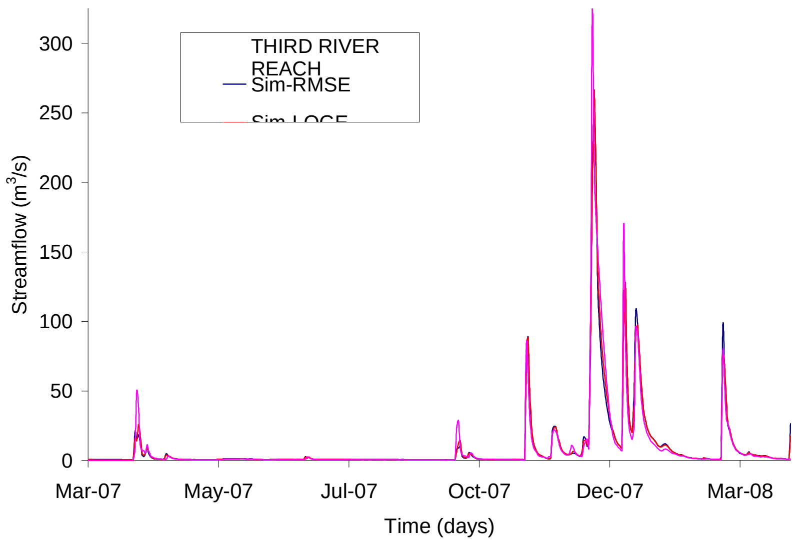


Figure 8.24: Observed and verification of the hybrid model of the third river reach

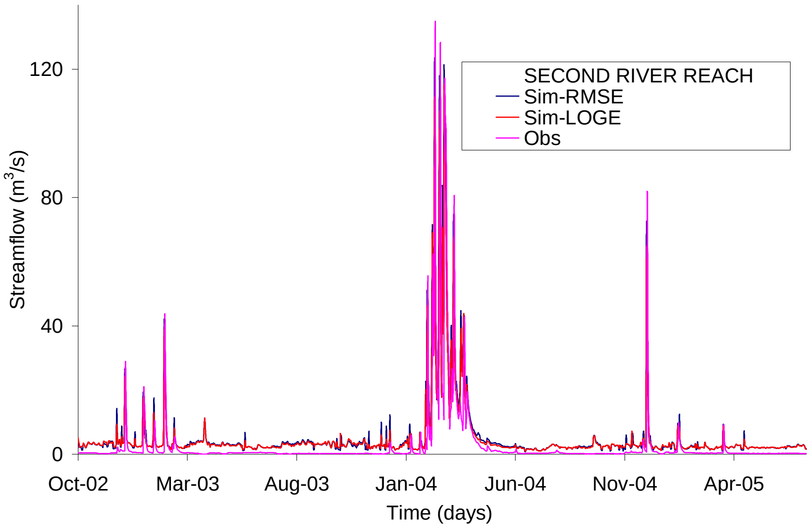


Figure 8.25: Observed and calibration of the hybrid model of the connected second river reach

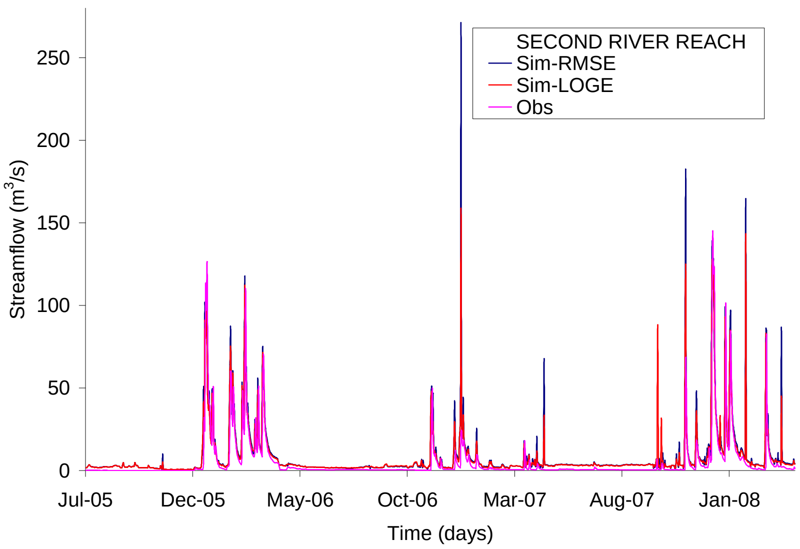


Figure 8.26: Observed and verification of the hybrid model of the connected second river reach

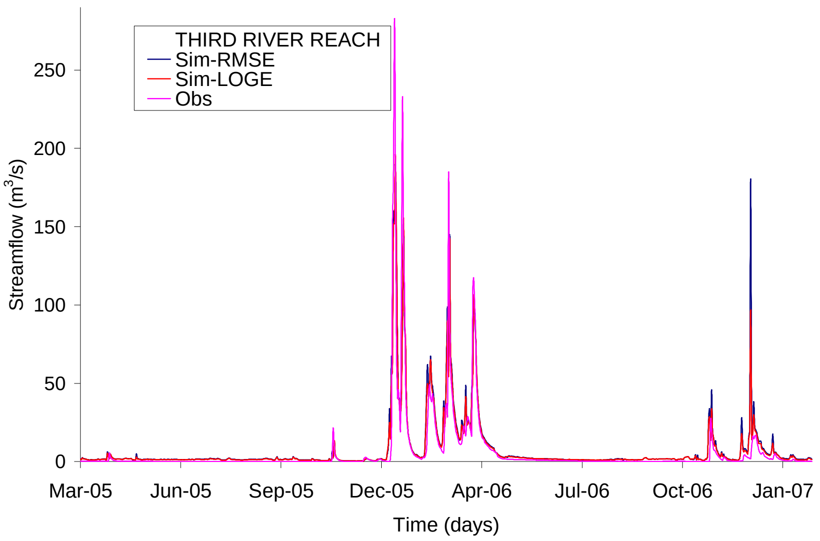


Figure 8.27: Observed and calibration of the hybrid model of the connected third river reach

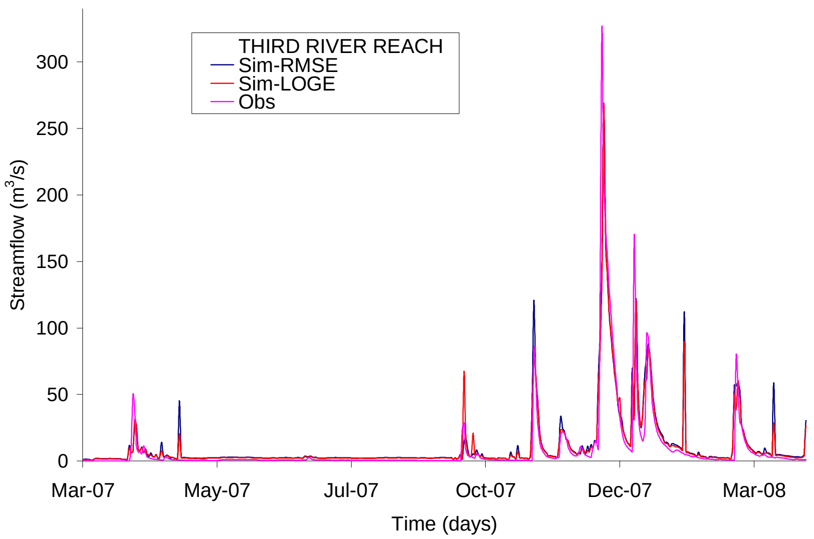


Figure 8.28: Observed and verification of the hybrid model of the connected third river reach

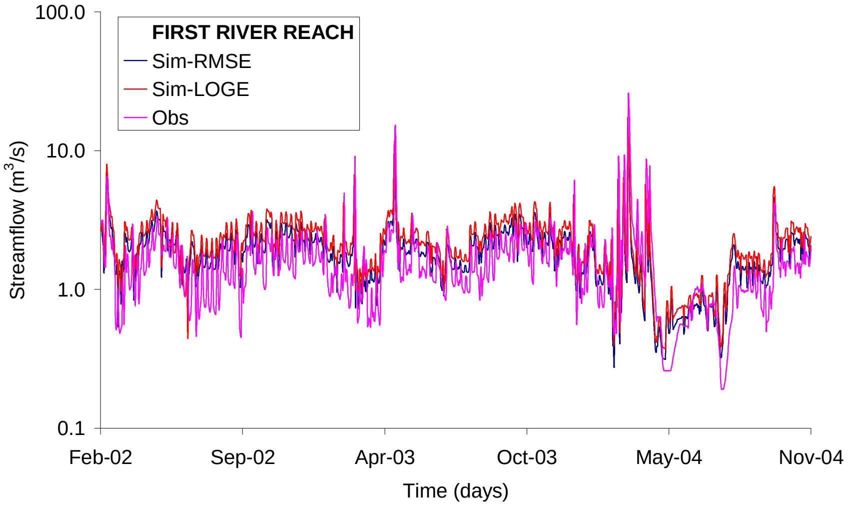


Figure 8.29: Observed and calibration of the hybrid model of the first river at logarithmic scale

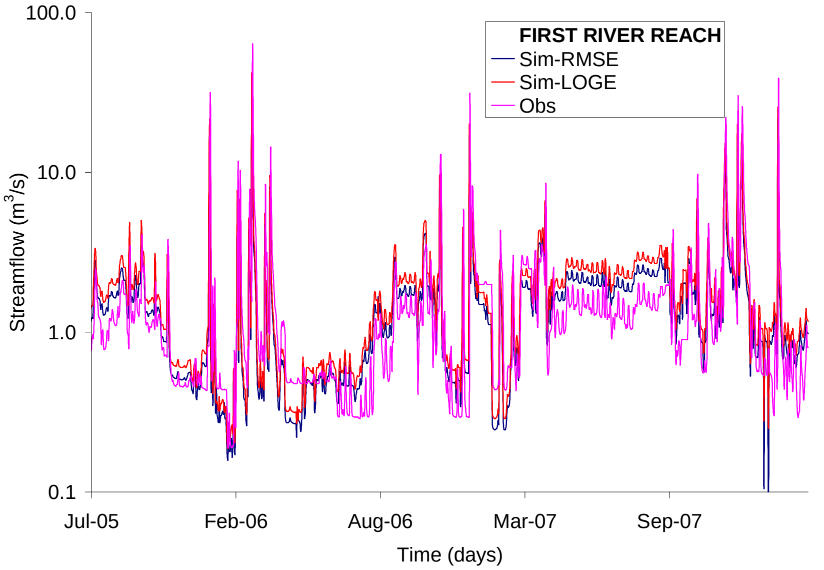


Figure 8.30: Observed and verification of the hybrid model of the first river reach at logarithmic scale

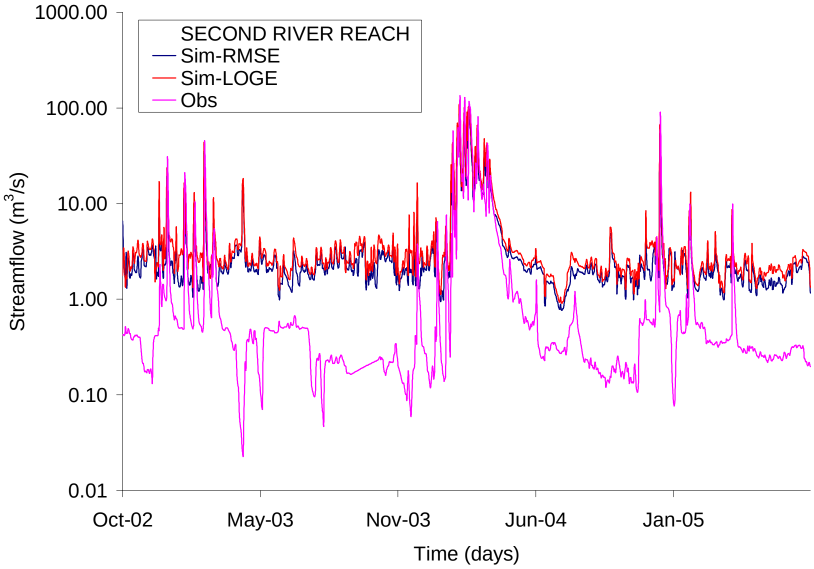


Figure 8.31: Observed and calibration of the hybrid model of the second river reach at logarithmic scale

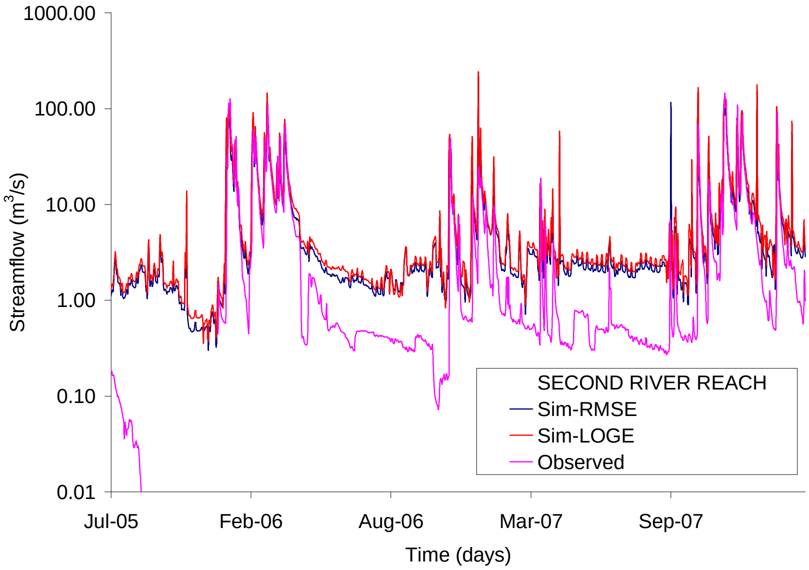


Figure 8.32: Observed and verification of the hybrid model of the second river reach at logarithmic scale

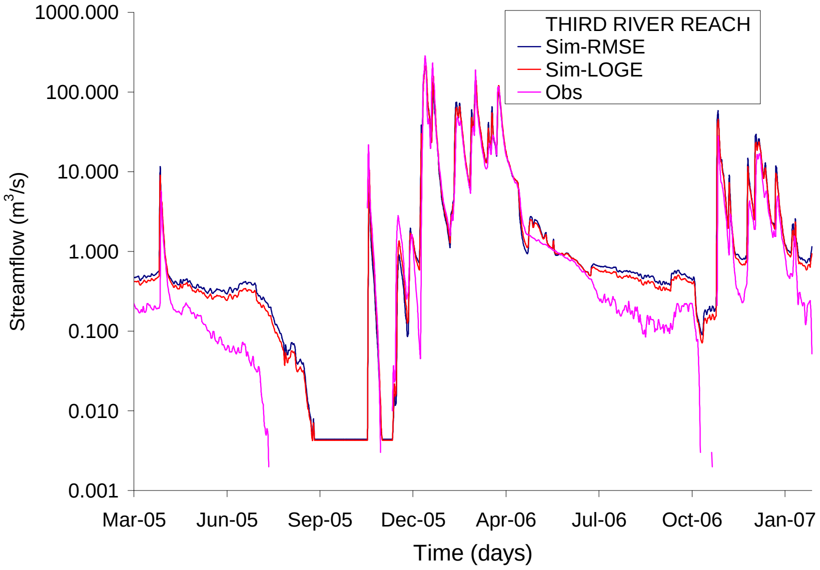


Figure 8.33: Observed and calibration of the hybrid model of the third river reach at logarithmic scale

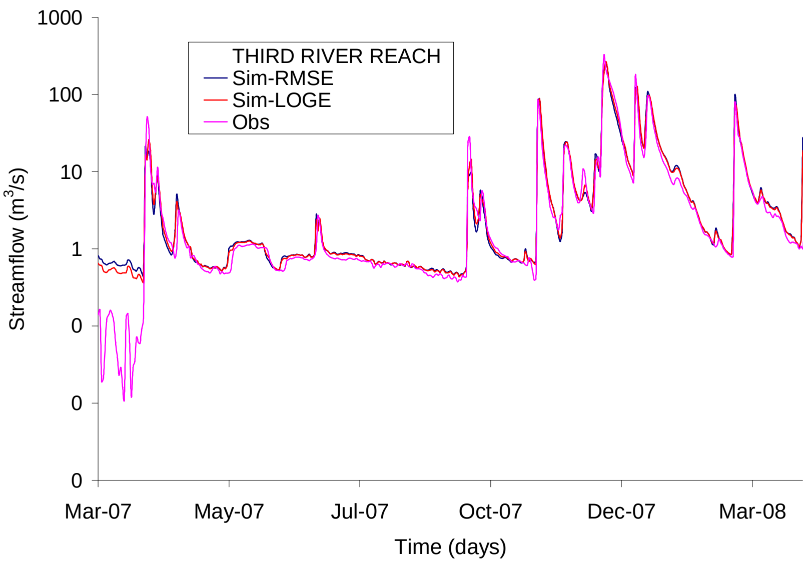


Figure 8.34: Observed and verification of the hybrid model of the third river reach at logarithmic scale

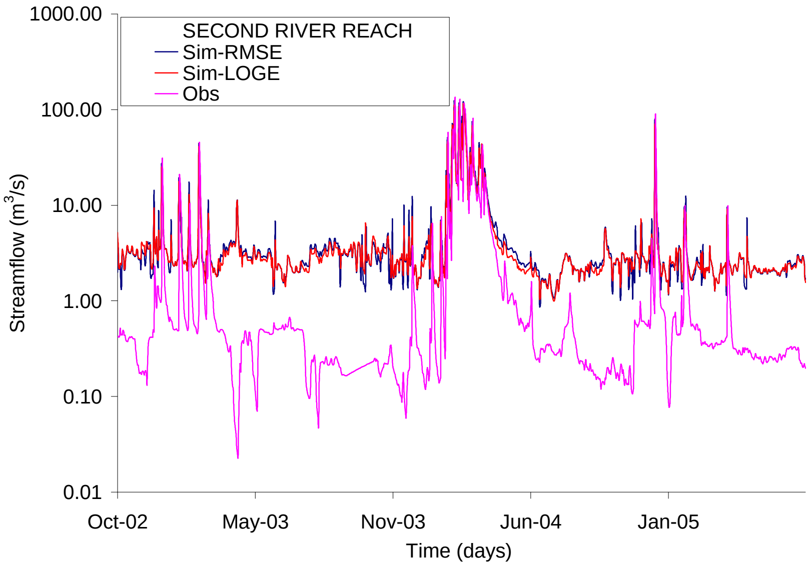


Figure 8.35: Observed and calibration of the conceptual model of the connected second river reach at logarithmic scale

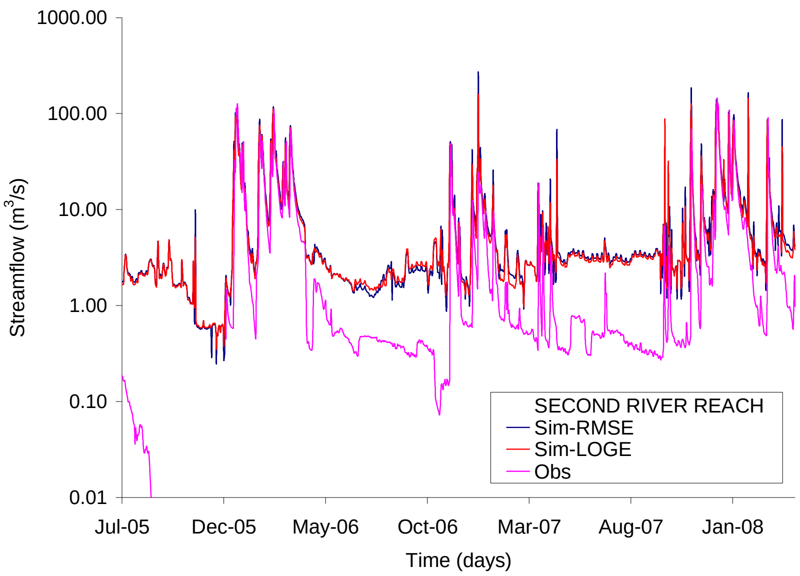


Figure 8.36: Observed and verification of the conceptual model of the connected second river reach at logarithmic scale

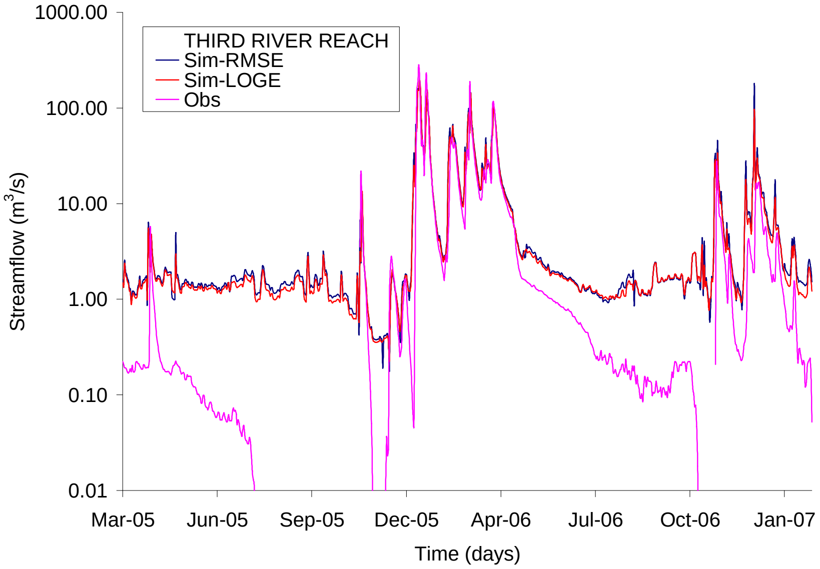


Figure 8.37: Observed and calibration of the conceptual model of the connected third river reach at logarithmic scale

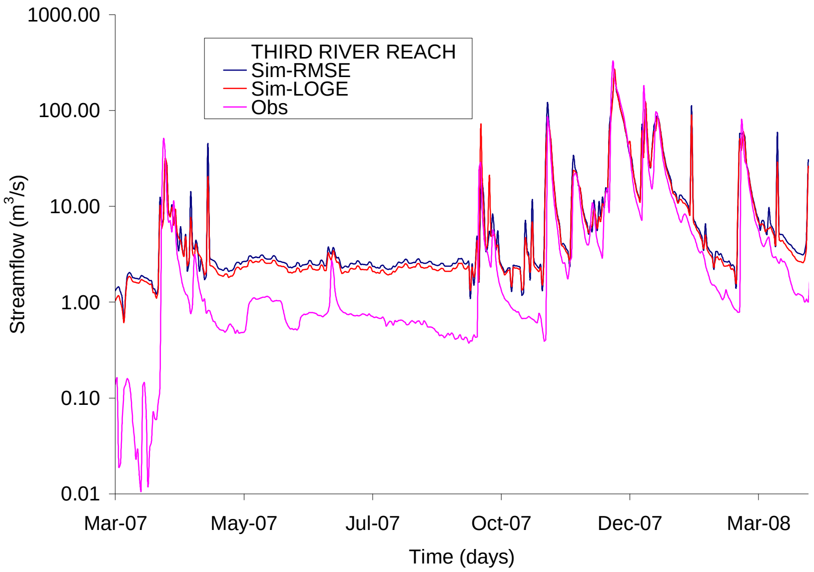


Figure 8.38: Observed and verification of the conceptual model of the connected third river reach at logarithmic scale

In catchment modelling, conceptual models attempt to simulate the system by representing it in a simplified form (Ndiritu and Daniell, 1999). However, while the model simulations may have low performance values associated with it, it is the representation of the catchment processes that may be more important (Seibert and McDonnell, 2002, Silberstein, 2006), as such, a stand-alone fuzzy inference model (Cheaper 6), a conceptual model (Cheaper 7), and a hybrid model (Cheaper 8) have been applied to model the flow in the Letaba River. The advantages and disadvantages of the respective models needs to be known and this, therefore, merits a comparison of the three models which is done in the next Chapter.

9 COMPARISON OF FUZZY INFERENCE, CONCEPTUAL AND CONCEPTUAL-FUZZY INFERENCE MODELS

9.1 THE NEED FOR COMPARISON

The fuzzy inference, conceptual and hybrid conceptual-fuzzy inference models have been applied to model flows in the Letaba river system. These models have demonstrated different capabilities with respect to modelling flow in the Letaba river system. In this chapter, a comparative analysis of the models based on the adequacy of representation of the system's characteristics, setting up, computation requirements and streamflow simulation performance is undertaken. These issues are some of the considerations in deciding the appropriateness of the various modelling approaches.

9.2 MODEL REPRESENTATION OF THE LETABA RIVER SYSTEM

The conceptual and the hybrid model were designed as an attempt to represent the system more closely than the fuzzy inference model. The complexity of the conceptual model was limited by the available data. The various catchment processes such as interception storage, soil storage, storage weirs farm dams, streamflow losses are represented in the model. The values obtained for the capacities of the various components are comparable to what has been reported in other studies. Although the conceptual and the hybrid model required more information, such as the catchment area, they provided results that improved the knowledge on the catchment characteristics. This is the advantage they possess over the fuzzy inference model. For those tasks where the knowledge of the catchment processes is not a requirement, however, the use of fuzzy inference model would be the preferred choice.

9.3 MODEL SETTING UP AND COMPUTATION REQUIREMENTS

All the models developed in this study were calibrated with several parameters that needed to be calibrated. The number of parameters varied, with the conceptual model having least number of parameters followed by the fuzzy inference model. The hybrid model had the largest number of parameters. The larger the number of calibrated parameters, the longer model calibration process takes. For instance, to calibrate a hybrid model the time required ranged between 45 minutes to 120 minutes, a stand-alone fuzzy inference model required 25 minutes to 60 minutes while the conceptual model was calibrated in less than 30 minutes. In addition, the conceptual and hybrid model did not require the data transformations that the fuzzy model needed. This highlights one of the disadvantages that fuzzy inference exhibited; it was found to become unstable when the variation of the input was significant as in the case of evaporation and when simulating low flows.

9.4 MODEL PERFORMANCE BASED ON STATISTICS

The performance rating by Moriasi et al., 2007 (Table 6.1) is used here to compare the models. The general performance of all the models improved towards the downstream suggesting the intractable human induced processes taking place in the upper river reaches (Table 9.1) greatly impact the models' performance. Considering the third river reach all models obtained very good performances as reflected in the Nash-Sutcliff efficiency values obtained. This, however, was contrasted by the conceptual model performance during the calibration phase. The Nash-Sutcliff efficiency values obtained for the second river reach range from unsatisfactory to satisfactory. The unsatisfactory values were obtained by the conceptual model during the calibration and verification phases and the hybrid model during the verification phase. When considering the first river reach, all the models obtained unsatisfactory Nash-Sutcliff efficiency values with exception of the first river reach during the calibration phase.

When assessing the performance of the models based on the percentage bias, the values obtained by all the model range from unsatisfactory to satisfactory. The majority of the PBIAS values obtained by the conceptual model are very good with the exception of the first river reach during the calibration phase. The fuzzy inference model obtained very good PBIAS values for all the river reaches except the second river reach during verification. The hybrid model obtained unsatisfactory values of the percentage bias for the second river reach, however, the model obtained good and very good values of the percentage bias for the third river reach. Based on the percentage bias values, the performance does not have a trend in the improvement toward the downstream for all the three models.

Table 9.1: Performance of the fuzzy inference, conceptual and hybrid model

River Reach	Phase	Fuzzy model				Conceptual model				Hybrid model			
		CCoef	NSE	PBIAS	RSR	CCoef	NSE	PBIAS	RSR	CCoef	NSE	PBIAS	RSR
FIRST	Calibration	0.72	0.51	-8.9	0.7	0.23	-2.22	57.62	1.794	0.62	0.36	-13.6	0.8
	Verification	0.47	0.08	-3.47	0.83	0.09	-2.87	12.54	1.789	0.42	0.15	1.76	0.84
SECOND	Calibration	0.8	0.63	-9.43	0.6	0.52	0.27	4	0.839	0.75	0.58	-46. 7	0.66
	Verification	0.79	0.56	-14	0.52	0.42	-0.14	-3	0.838	0.69	0.29	-26.4	0.66
THIRD	Calibration	0.92	0.85	3.96	0.38	0.86	0.69	2.16	0.57	0.91	0.84	-11.3	0.41
	Verification	0.95	0.9	6.95	0.21	0.94	0.78	6.63	0.334	0.96	0.92	-0.19	0.2

In general and when considering the performance measures, the standalone fuzzy inference is the best model followed by the conceptual-fuzzy inference model. The standalone conceptual model obtained unsatisfactory performance values.

9.5 MODEL PERFORMANCE BASED ON PLOTS

Figure 9.1 through Figure 9.6 show simulated flow from the three models using the RMSE as the objective function. The plots place more emphasis on the low flow simulation because of the ecological concerns in the Kruger National Park. There is a significant variation in the three models' low flow capabilities suggesting that the emphasis each model places on low flow simulation differs. Considering the simulation for the first river reach, the fuzzy inference model obtained simulations that were closer to the observed. The conceptual and the hybrid model underestimated the low flows during the calibration phase, however, during the verification phase, the hybrid model over estimated the low flows while the conceptual continued to underestimate them.

The simulated flows in the second river suggests that the hybrid model over estimated the low flows during the calibration and verification phases. The fuzzy inference model and the conceptual model displayed some characteristics suggesting instability with the simulated low flows varying highly within short periods. This can be attributed to the nature of the data resulting from the human induced processes in this river reach, however the hybrid model overcame these, may be due to initial catchment processes modelled by the conceptual component that result into more refined inputs and less input that are used in the fuzzy inference component of the hybrid model.

The performances of the models in the third river reach suggests that the hybrid model performed well in simulating the low flows although it was observed to slightly over estimate the low flows during the calibration period. The fuzzy inference model under estimated the low flows during the calibration and verification phases and the conceptual model overestimated the low flows.

Based on the nature of the simulated flows, the fuzzy inference model performed better than the other models in simulating the flows in the first river reach suggesting it capability in dealing with impacted flows. The hybrid model simulated better the low flows in the third river reach suggesting for its capabilities in modelling the low flows. If the task at hand requires the simulation of human induced processes, then the fuzzy inference based model or the hybrid model should be the preferred choice. But if some understanding of the impact of human activities is needed, then the hybrid or conceptual model may be used to inform this. However, the only drawback of the hybrid model may be the computation intensity. In order to maintain the minimum flows required in the Kruger National Park, the hybrid model would be the preferred choice.

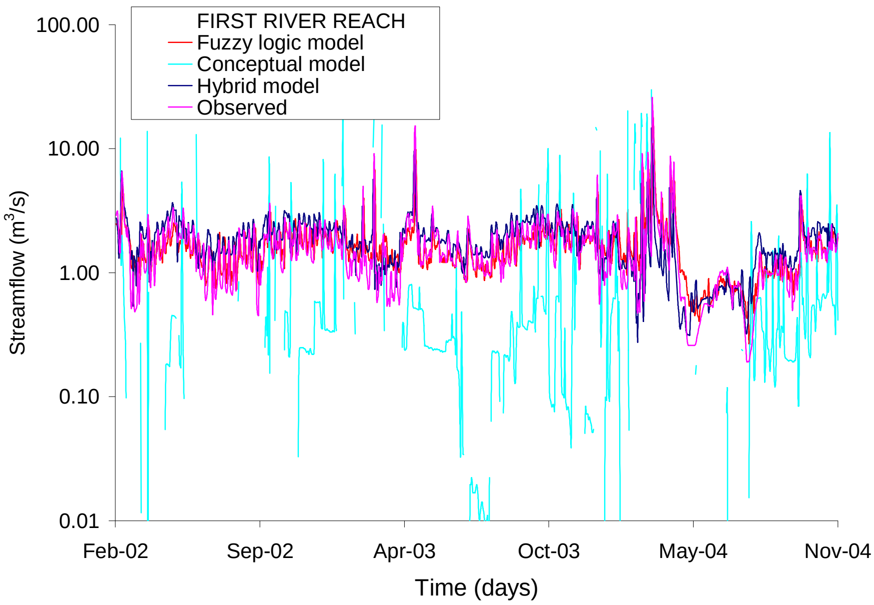


Figure 9.1: Simulated and observed flow for the first river reach during calibration phase

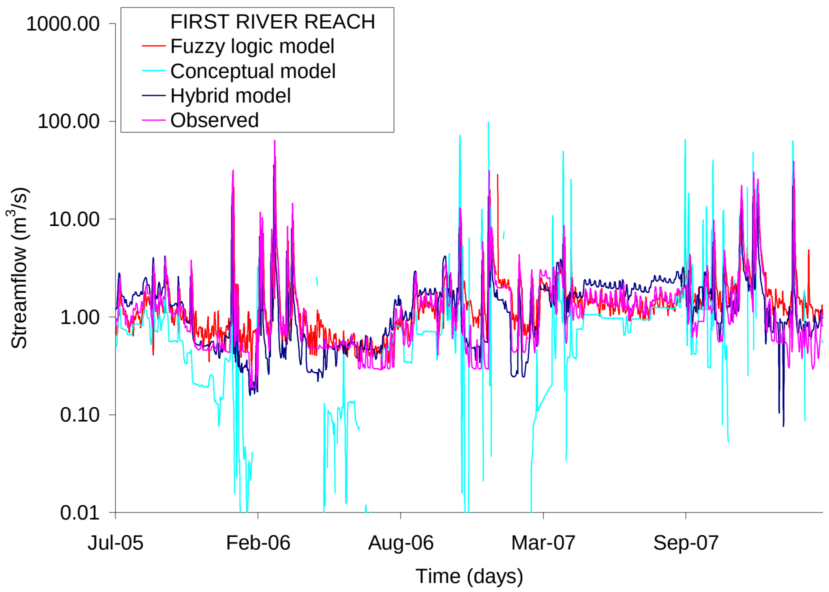


Figure 9.2: Simulated and observed flows for the first river reach during the verification phase

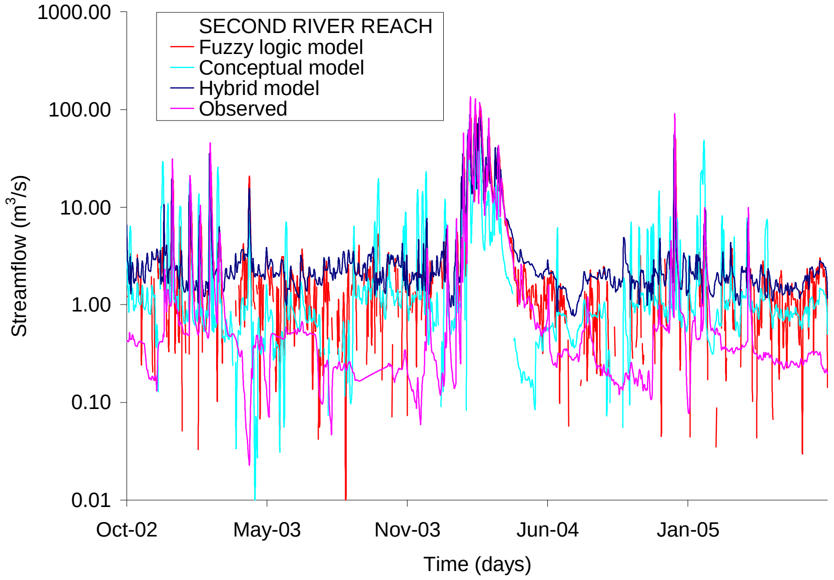


Figure 9.3: Simulated and observed flows for the second river reach during the calibration phase

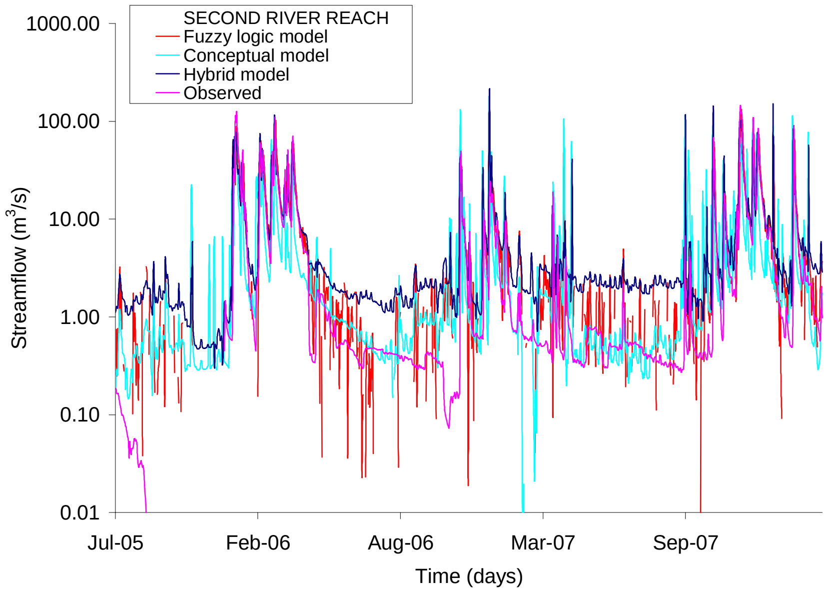


Figure 9.4: Simulated and observed flows for the second river reach during the verification phase

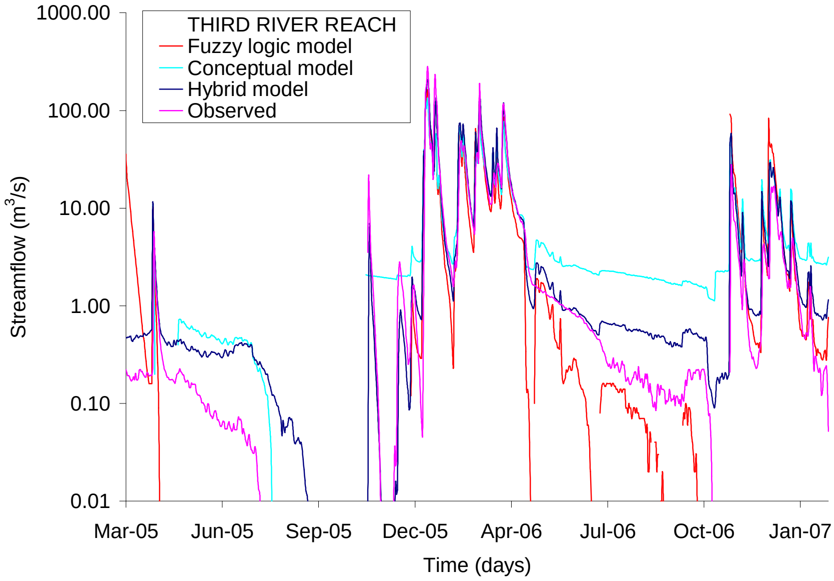


Figure 9.5: Simulated and observed flows for the third river reach during the calibration phase

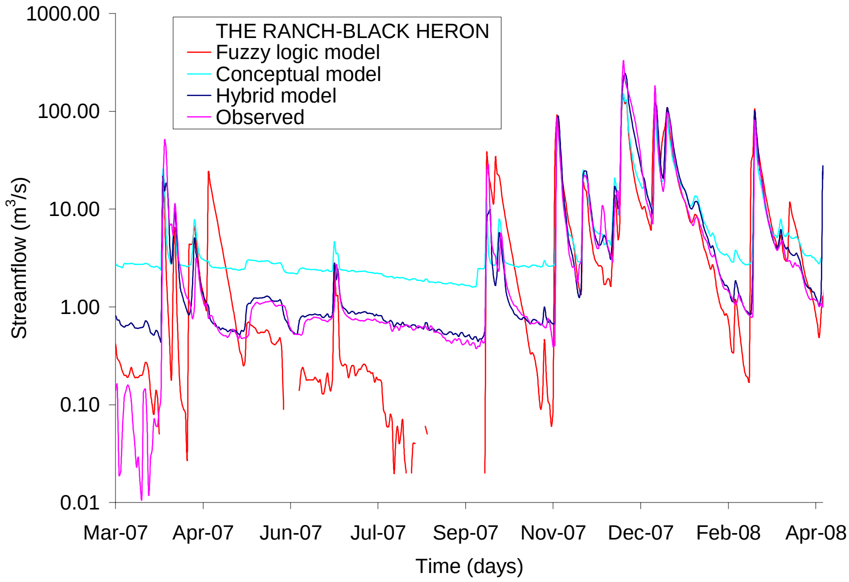


Figure 9.6: Simulated and observed flows for the third river reach during the verification phase

The conclusions, and recommendations with respect to the models applied in the Letaba River system then follow.

10 CONCLUSIONS AND RECOMMENDATIONS

10.1 CONCLUSIONS

In South Africa, water demand is increasing, the infrastructure in these river systems is highly developed and the catchments are characterised by scarce data. Adequate modelling of such systems for planning and operational purposes using many of the existing models is not easy owing largely to the mismatch between the data requirements of these models and the data available. The objective of this study was to examine the suitability of the fuzzy inference and hybrid conceptual-fuzzy inference models in modelling complex data-scarce river systems. The Letaba River system being an example of this has been used as a case study.

For the purpose of meeting the objective, a review of some of the commonly used catchments models was done, fuzzy inference and fuzzy based hybrid approaches were also reviewed, an assessment of the Letaba River system was done and finally the development, application and evaluation of fuzzy inference, conceptual and hybrid models was done.

The selected models were the Pitman model, ACRU model, MIKE models (MIKE SHE and MIKE 11), soil moisture assessment tool (SWAT) and Soil moisture accounting and routing model. The models were found unsuited for the task at hand for a variety of reasons; the modelling time step (Pitman model operate at monthly), intensive data requirements and expertise required for their application (MIKE models, ACRU and SWAT) and the inadequate representation of the catchment processes (SMAR). With these limitations in mind, a discussion of the capabilities of the fuzzy inference based approaches including their ability to deal with inadequate and vague information was done. Considering the finding that fuzzy inference has had limited prior application with respect to modelling flow in complex river systems, its application in complex river system with scarce data, as was done in this study, was another opportunity to probably extend the range of its applicability.

In addition to the soft data, the hard data essential for modelling the flows in the Letaba River system that was available included rainfall, streamflow, evaporation and abstractions. Assessment of the performance of the fuzzy inference model based on statistical performance measures and simulated time series was found not to improve the understanding of the poorly understood processes and the conceptual model was developed and applied so as to offer some basis for assessing the fuzzy inference model performance in a comparative manner. The conceptual model reproduced the catchment processes and the water resources system (anthropogenic activities) reasonable well thereby providing estimates of the fluxes of the main natural processes and anthropogenic activities. The fuzzy inference model however obtained better overall statistical performance than the conceptual model. The hybrid fuzzy-conceptual- model was perceived as superior and was found to perform comparably with the fuzzy model while representing the catchment processes as realistically as the stand-alone conceptual model. This indicated that the hybrid may be more suitable for applications where simulation accuracy and process representation are both important. The various quantities that represent the characteristics the Letaba River system were obtained and found to be inline with what was reported by other studies; suggesting that the models satisfactorily represented the system. All the model performance measures indicated improvements in model performance towards the downstream and were attributed to the reduction in anthropogenic activities along the river in the downstream direction. The modelling of flows in the Letaba River system has been done by innovatively using the available data. The modelling in this study helped to improve knowledge of the poorly understood and unmonitored processes while simultaneously taking into account the uncertainties associated with the use of inadequate data.

10.2 RECOMMENDATIONS

Since the hybrid fuzzy inference-conceptual model was able to provide superior flow simulations, it is imperative that further applications applying the hybrid modelling approach be undertaken in future studies aimed at making optimal use of the catchment data. The actual structure of an appropriate hybrid model would obviously depend on the specific modelling problem and the data at hand. For the system studied here, the structure of the hybrid model was obtained fairly subjectively and can be further improved. For example, only the streamflow output from the conceptual component 'filters' through the fuzzy component and the other outputs (e.g. farm dam and storage weirs trajectories, losses, evapotranspiration series) are intermediate though the impact of the fuzzy inference component on these is obtained indirect via calibration. The structure of the hybrid model can however be improved to obtain a better alignment of the outputs. In addition, modelling that excludes lagged observed outflows can be tested as this would increase the versatility of the hybrid modelling.

For any modelling task, the quality and quantity of the data plays an important role in improving the quality of the simulations obtained. In this regard, deliberate initiatives need to be undertaken to improve and increase the scale of monitoring in systems such as the Letaba River system. For the Letaba, this would require the operation of the individually owned and non-individually owned storage facilities and the water abstractions from both the surface and groundwater sources to be closely monitored and measured in addition to improving streamflow measurement including the contributions from tributaries.

REFERENCES

ABBOTT, M. B., BATHURST, J. C., CUNGE, J. A., O'CONNELL, P. E. & RASMUSSEN, J. (1986) An introduction to the European Hydrological System -- Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system. *Journal of Hydrology,* 87**,** 45-59.

ABULOHOM, M. S. (1997) Calibration of a mathematical model for generating monthly River Flows from Meteorological Data for a Selected Catchment. *M.Sc thesis, CEWRE, UET, Lahore, Pakistan.*

AFSHAR, A. & FATHI, H. (2009) Fuzzy multi-objective optimization of finance-based scheduling for construction projects with uncertainties in cost. *Engineering Optimization,* 41**,** 1063-1080.

ALTUNKAYNAK, A. & ÅŽEN, Z. (2007) Fuzzy logic model of lake water level fluctuations in Lake Van, Turkey. *Theoretical and Applied Climatology,* 90**,** 227-233.

ARNOLD, J. G. & FOHRER, N. (2005) SWAT2000: Current capabilities and research opportunities in applied watershed modelling, Hydrological Processes, 19, 563–572.

ARNOLD, J., WILLIAMS, J. & MAIDMENT, D. (1995) Continuous-time water and sediment-routing model for large basins. *Journal of Hydraulic Engineering,* 121**,** 171-183.

AZADI, H., SHAHVALI, M., VAN DEN BERG, J. & FAGHIH, N. (2007) Sustainable rangeland management using a multi-fuzzy model: How to deal with heterogeneous experts' knowledge. *Journal of Environmental Management,* 83**,** 236-249.

BABOVIC, V., DRÉCOURT, J.-P., KEIJZER, M. & FRISS HANSEN, P. (2002) A data mining approach to modelling of water supply assets. *Urban Water,* 4**,** 401-414.

BAILEY, A. (2008) Water resources simulation model for Windows, Theory Document, PWMA 04/000/00/6107, (DWAF, WRC, SSI).

BARDOSSY, A., BRONSTERT, A. & MERZ, B. (1995) 1-, 2- and 3-dimensional modeling of water movement in the unsaturated soil matrix using a fuzzy approach. *Advances in Water Resources,* 18**,** 237-251.

BARDOSSY, A., BRONSTERT, A. AND MERZ, B. (1995) 1-, 2- and 3-dimensional modeling of water movement in the unsaturated soil matrix using a fuzzy approach. Advances in Water Resources, 18(4): 237-251.

BÄRLUND, I., KIRKKALA, T., MALVE, O. & KÄMÄRI, J. (2007) Assessing SWAT model performance in the evaluation of management actions for the implementation of the Water Framework Directive in a Finnish catchment. *Environmental Modelling & Software,* 22**,** 719-724.

BENAMAN, J. & SHOEMAKER, C. (2004) Methodology for analyzing ranges of uncertain model parameters and their impact on total maximum daily load process. *Journal of environmental engineering,* 130**,** 648-656.

BERGSTRÖM, S. (1995) The HBV model, *in Computer Models of Watershed Hydrology, edited by V. J. Singh Water Resources Publications, Englewood, USA.*

BESSLER, F. T., SAVIC, D. A. & WALTERS, G. A. (2003) Water Reservoir Control with Data Mining. *Journal of Water Resources Planning and Management,* 129**,** 26-34.

BEVEN, K. & BINLEY, A. (1992) The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes,* 6**,** 279-298.

BEVEN, K. (2004) Rainfall-runoff modelling: the primer, *John Wiley & Sons Inc.*

BEVEN, K. J. & KIRKBY, M. J. (1979) A physically based, variable contributing area model of basin hydrology, *Hydrological Sciences Bulletin,* 24**,** 43 - 69.

BEVEN, K., (2006) A manifesto for the equifinality thesis. Journal of Hydrology, 320, 18-36.

BEZDEK, J. (1981) Pattern recognition with fuzzy objective function algorithm*s*, Kluwer Academic Publishers Norwell, MA, USA.

BEZDEK, J., HATHAWAY, R., SABIN, M. & TUCKER, W. (1987) Convergence theory for fuzzy c-means: Counter examples and repairs. *IEEE Transactions on Systems, Man, and Cybernetics,* 17**,** 873-877.

BOROTO, R. & GORGENS, A. (2003) Estimating transmission losses along the Limpopo River: an overview of alternative methods. *IAHS PUBLICATION***,** 138-143.

BRUNNER, G. (1998) HEC-RAS: River Analysis System. *Unpublished Users Manual, Version,* 2.

CHAUBEY, I., COTTER, A., COSTELLO, T. & SOERENS, T. (2005) Effect of DEM data resolution on SWAT output uncertainty. *Hydrological Processes,* 19**,** 621-628.

CHEN, J. & ADAMS, B. J. (2006) Integration of artificial neural networks with conceptual models in rainfall-runoff modelling. *Journal of Hydrology,* 318**,** 232-249.

CHETTY, K. & SMITHERS, J. (2005) Continuous simulation modelling for design flood estimation in South Africa: Preliminary investigations in the Thukela catchment. *Physics and Chemistry of the Earth,* 30**,** 634-638.

CHIEW, F. H. S., KAMALADASA, N. N., MALANO, H. M. & MCMAHON, T. A. (1995) Penman-Monteith, FAO-24 reference crop evapotranspiration and class-A pan data in Australia. *Agricultural Water Management,* 28**,** 9-21.

CHIU, S. L. (1994) Cluster estimation method with extension to fuzzy model identification. *IEEE International Conference on Fuzzy Systems, 1240-1245.*

CUI, L. J. & KUCZERA, G. (2003) Optimizing urban water supply headworks using probabilistic search methods. *JOURNAL OF WATER RESOURCES PLANNING AND MANAGEMENT,* 129**,** 380-387.

CUNDERLIK, J. M. (2003) Hydrological model selection for CFCAS Project: Assessment of Water Risk and Vulnerability to changing climate condition, *Project Report I, ISBN 978-0-7714-2623-0*

DEMETRIOU, C. & PUNTHAKEY, J. F. (1999) Evaluating sustainable groundwater management options using the MIKE SHE integrated hydrogeological modelling package. *Environmental Modelling and Software,* 14**,** 129-140.

DEMIRLI, K., CHENG, S. X. & MUTHUKUMARAN, P. (2003) Subtractive clustering based modeling of job sequencing with parametric search. *Fuzzy Sets and Systems,* 137**,** 235-270.

DEPARTMENT OF WATER AFFAIRS AND FORESTRY, (2004), Internal Strategic Perspective: Luvuvhu/Letaba WMA, Report No: P WMA 02/000/00/0304.

DEPARTMENT OF WATER AFFAIRS AND FORESTRY, (2006a), Letaba catchment reserve determination study, Report RDM/B800/01/CON/COMP/1304.

DEPARTMENT OF WATER AFFAIRS AND FORESTRY, (2006b), Letaba river system annual operating analysis 2005/06.

DUAN, Q.Y., SOROOSHIAN, S. & GUPTA, V. (1992) Effective and efficient global optimization for conceptual rainfall- runoff models. Water Resources Research, 28(4): 1015-1031.

DUNN, J. (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Cybernetics and Systems,* 3**,** 32-57.

DYE, P. J. & CROKE, B. F. W. (2003) Evaluation of streamflow predictions by the IHACRES rainfall-runoff model in two South African catchments. *Environmental Modelling & Software,* 18**,** 705-712.

ECKHARDT, K. & ARNOLD, J. G. (2001) Automatic calibration of a distributed catchment model. *Journal of Hydrology,* 251**,** 103-109.

ELFEKI, A.M.M. (2006) Reducing concentration uncertainty using the coupled Markov chain approach. Journal of Hydrology, 317(1-2): 1-16.

EVERSON, C. S. (2001) The water balance of a first order catchment in the montane grasslands of South Africa. *Journal of Hydrology,* 241**,** 110-123.

EWEN, J. & PARKIN, G. (1996) Validation of catchment models for predicting land-use and climate change impacts. 1. Method. *Journal of Hydrology,* 175**,** 583-594.

FENICIA, F., MCDONNELL, J. J. & SAVENIJE, H. H. G. (2008) Learning from model improvement: On the contribution of complementary data to process understanding. *Water Resources Research,* 44.

FENICIA, F., SAVENIJE, H. H. G., MATGEN, P. & PFISTER, L. (2006) Is the groundwater reservoir linear? Learning from data in hydrological modelling. *Hydrology and Earth System Sciences,* 10**,** 139-150.

FREER, J. E., MCMILLAN, H., MCDONNELL, J. J. & BEVEN, K. J. (2004) Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. *Journal of Hydrology,* 291**,** 254-277.

GAN, T. Y., DLAMINI, E. M. & BIFTU, G. F. (1997a) Effects of model complexity and structure, data quality, and objective functions on hydrologic modelling. *Journal of Hydrology,* 192**,** 81-103.

GAN, Y. T., DLAMINI, E. M. & BIFTU, G. F. (1997b) Effects of model complexity and structure, data quality, and objective functions on hydrologic modelling. *Journal of Hydrology,* 192**,** 81-103.

GERSHON, M., (1987) Heuristic approaches for mine planning and production scheduling. Geotechnical and Geological Engineering, 5, 1-13

GEZA, M. & MCCRAY, J. E. (2008) Effects of soil data resolution on SWAT model stream flow and water quality predictions. *Journal of Environmental Management,* 88**,** 393-406.

GRANGER, R. J. & GRAY, D. M. (1990) Examination of Morton's CRAE model for estimating daily evaporation from field-sized areas. *Journal of Hydrology,* 120**,** 309-325.

GYEDU-ABABIO, T. K., (2005, The water crisis in and around the Kruger National Park, South Africa: Which way? *Proceedings of the 12th SANCIAHS Symposium, Johannesburg, South Africa.*

HATHAWAY, R. J. & BEZDEK, J. C. (1986) Local convergence of the fuzzy c-Means algorithms. *Pattern Recognition,* 19**,** 477-480.

HAVNØ, K., MADSEN, M. & DØRGE, J. (1995) MIKE 11–A generalized river modelling package, *in Computer Models of Watershed Hydrology, edited by V. J. Singh Water Resources Publications, Englewood, USA.*

HUGHES, D. A. & SAMI, K. (1992) Transmission losses to alluvium and associated moisture dynamics in a semiarid ephemeral channel system in Southern Africa. *Hydrological Processes,* 6**,** 45-53.

HUGHES, D. A. & SAMI, K. (1993) The Bedford catchments: An introduction to their physical and hydrological characteristics. *Unpublished Report to the Water Research Commission by the Institute for Water Research, Rhodes University Grahamstown, South Africa.*

HUGHES, D. A. (2004) Incorporating groundwater recharge and discharge functions into an existing monthly rainfall-runoff model. *Hydrological Sciences Journal,* 49**,** 297-312.

HUGHES, D. A., ANDERSSON, L., WILK, J. & SAVENIJE, H. H. G. (2006) Regional calibration of the Pitman model for the Okavango River. *Journal of Hydrology,* 331**,** 30-42.

HUGHES, D.A. AND SAMI, K., (1994) A semi-distributed, variable time interval model of catchment hydrology--structure and parameter estimation procedures. Journal of Hydrology, 155, 265-291.

HUNDECHA, Y., BARDOSSY, A. & THEISEN, H. W. (2001) Development of a fuzzy logic-based rainfall-runoff model. *Hydrological Sciences Journal,* 46**,** 363-376.

JACQUIN, A. P. & SHAMSELDIN, A. Y. (2006) Development of rainfall-runoff models using Takagi-Sugeno fuzzy inference systems. *Journal of Hydrology,* 329**,** 154-173.

JACQUIN, A.P. AND SHAMSELDIN, A.Y. (2009) Review of the application of fuzzy inference systems in river flow forecasting. *Journal of Hydroinformatics, 11(3-4): 202-210.*

JAIN, S. K., STORM, B., BATHURST, J. C., REFSGAARD, J. C. & SINGH, R. D. (1992) Application of the SHE to catchments in India Part 2. Field experiments and simulation studies with the SHE on the Kolar subcatchment of the Narmada River. *Journal of Hydrology,* 140**,** 25-47.

JAKEMAN, A. J. & HORNBERGER, G. M. (1993) How much complexity is warranted in a rainfall-runoff model? *Water Resources Research,* 29**,** 2637-2649.

JEWITT, G. & SCHULZE, R. (1999) Verification of the ACRU model for forest hydrology applications. *Water S. A.,* 25**,** 483-490.

JEWITT, G., GARRATT, J., CALDER, I. & FULLER, L. (2004) Water resources planning and modelling tools for the assessment of land use change in the Luvuvhu Catchment, South Africa. *Physics and Chemistry of the Earth,* 29**,** 1233-1241.

JIANG, W. G., LI, J., LI, Z. W. & WU, Y. F. (2008) Fuzzy assessment of the population risk of flood disaster. *Journal of Hunan University Natural Sciences,* 35**,** 84-87.

KAGODA, P. & NDIRITU, J. (2008) ANALYSIS OF THE EFFECT OF PARAMETER UNCERTAINTY IN RAINFALL FREQUENCY ESTIMATION. *Proceedings of the Second IASTED Africa Conference*, *Gaborone, Botswana*.

KASABOV, N. (1996) *Foundations of neural networks, fuzzy systems, and knowledge engineering*, The MIT press.

KATAMBARA, Z. & NDIRITU, J. (2007) Developing a Surface Water-Groundwater Interaction Model for Letaba River System in South Africa. *In: Proceeding of the 8th WATERNET Conference, Lusaka, Zambia.*

KATAMBARA, Z. & NDIRITU, J. (2009) A fuzzy inference system for modelling streamflow: Case of Letaba River, South Africa. *Physics and Chemistry of the Earth, Parts A/B/C,* 34**,** 688-700.

KLAASSEN, W., BOSVELD, F. & DE WATER, E. (1998) Water storage and evaporation as constituents of rainfall interception. *Journal of Hydrology,* 212-213**,** 36-50.

KRISTENSEN, K. & JENSEN, S. (1975) A model for estimating actual evapotranspiration from potential evapotranspiration. *Nordic Hydrology,* 6.

LABADIE, J. W. (2004) Optimal operation of multireservoir systems: State-of-the-art review. *Journal of Water Resources Planning and Management,* 130**,** 93-111.

LAUZON, N. & LENCE, B. (2008) Hybrid fuzzy-mechanistic models for addressing parameter variability. *Environmental Modelling and Software,* 23**,** 535-548.

LIU, C., ZHANG, X. & ZHANG, Y. (2002) Determination of daily evaporation and evapotranspiration of winter wheat and maize by large-scale weighing lysimeter and micro-lysimeter. *Agricultural and Forest Meteorology,* 111**,** 109-120.

LLOYD, C. & ATKINSON, P. (2001) Assessing uncertainty in estimates with ordinary and indicator kriging. Computers & Geosciences, 27(8): 929-937.

LOHANI, A. K., GOEL, N. K. & BHATIA, K. K. S. (2006) Takagi-Sugeno fuzzy inference system for modelling stage-discharge relationship. *Journal of Hydrology,* 331**,** 146-160.

LOUCKS, D.P., VAN BEEK, E., STEDINGER, J.R., DIJKMAN, J.P.M. & VILLARS, M.T. (2005) Water resources systems planning and management: an introduction to methods, models and applications. Paris: UNESCO.

LYON, S. W., LEMBO, J. A. J., WALTER, M. T. & STEENHUIS, T. S. (2006) Defining probability of saturation with indicator kriging on hard and soft data. *Advances in Water Resources,* 29**,** 181-193.

MAMDANI, E. H. & ASSILIAN, S. (1975) An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies,* 7**,** 1-13.

MCKENZIE, R. S. & CRAIG, A. R. (2001) Evaluation of river losses from the Orange River using hydraulic modelling. *Journal of Hydrology,* 241**,** 62-69.

MCMICHAEL, C. & HOPE, A. (2007) Predicting streamflow response to fire-induced landcover change: Implications of parameter uncertainty in the MIKE SHE model. *Journal of Environmental Management,* 84**,** 245-256.

MORIASI, D. N., ARNOLD, J. G., VAN LIEW, M. W., BINGNER, R. L., HARMEL, R. D. & VEITH, T. L. (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE,* 50**,** 885-900.

NASH, J. E. & SUTCLIFFE, J. V. (1970) River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology,* 10**,** 282-290.

NDIRITU, J. (2009) A comparison of automatic and manual calibration using the Pitman model. *Physics and Chemistry of the Earth, Parts A/B/C,* 34**,** 729-740.

NDIRITU, J. G. & DANIELL, T. M. (1999) Assessing model calibration adequacy via global optimisation. *Water SA,* 25**,** 317-326.

NDIRITU, J. G. (2005) Maximising water supply system yield subject to multiple reliability constraints via simulation-optimisation. *Water SA,* 31**,** 423-434.

NDOMBA, P., MTALO, F. & KILLINGTVEIT, A. (2008) SWAT model application in a data scarce tropical complex catchment in Tanzania. *Physics and Chemistry of the Earth, Parts A/B/C,* 33**,** 626-632.

NEWMAN, B., VIVONI, E. & GROFFMAN, A. (2006) Surface water-groundwater interactions in semiarid drainages of the American southwest. *Hydrological Processes,* 20**,** 3371-3394.

NIELSEN, S. & HANSEN, E. (1973) Numerical simulation of the rainfall-runoff process on a daily basis. *Nordic Hydrology,* 4**,** 171–190.

O'CONNELL, P. E., NASH, J. E. & FARRELL, J. P. (1970) River flow forecasting through conceptual models part II - The Brosna catchment at Ferbane. *Journal of Hydrology,* 10**,** 317-329.

PILGRIM, D. H. & BLOOMFIELD, P. (1980) Problems in determining infiltration and soil store parameters of runoff models. *IAHS-AISH Publication* 129.

PITMAN, W. (1973) A mathematical model for generating monthly river flows from meteorological data in South Africa. *Report,* 2**,** 73.

QIU, G. Y., YANO, T. & MOMII, K. (1998) An improved methodology to measure evaporation from bare soil based on comparison of surface temperature with a dry soil surface. *Journal of Hydrology,* 210**,** 93-105.

QIU, Z. & PRATO, T. (1998) ECONOMIC EVALUATION OF RIPARIAN BUFFERS IN AN AGRICULTURAL WATERSHED1. *JAWRA Journal of the American Water Resources Association,* 34**,** 877-890.

QIU, Z. & PRATO, T. (2001) PHYSICAL DETERMINANTS OF ECONOMIC VALUE OF RIPARIAN BUFFERS IN AN AGRICULTURAL WATERSHED1. *JAWRA Journal of the American Water Resources Association,* 37**,** 295-303.

QIU, Z. & PRATO, T. (2007a) ECONOMIC EVALUATION OF RIPARIAN BUFFERS IN AN AGRICULTURAL WATERSHED1. *JAWRA Journal of the American Water Resources Association,* 34**,** 877-890.

QIU, Z. & PRATO, T. (2007b) Physical determinants of economic value of riparian buffers in an agricultural watershed. *JAWRA Journal of the American Water Resources Association,* 37**,** 295-303.

RAJA, P. A. & KUMAR, D. N. (1998) Ranking multi-criterion river basin planning alternatives using fuzzy numbers. *Fuzzy Sets and Systems,* 100**,** 89-99.

RAJURKAR, M. P., KOTHYARI, U. C. & CHAUBE, U. C. (2004) Modelling of the daily rainfall-runoff relationship with artificial neural network. *Journal of Hydrology,* 285**,** 96-113.

RAO, A. R. & SRINIVAS, V. V. (2006) Regionalization of watersheds by fuzzy cluster analysis. *Journal of Hydrology,* 318**,** 57-79.

REFSGAARD, J.& STORM, B. (1995) MIKE SHE *in Computer Models of Watershed Hydrology, edited by V. J. Singh Water Resources Publications, Englewood, USA.*

SALAS, J., KIM, H., EYKHOLT, R., BURLANDO, P. & GREEN, T. (2005) Aggregation and sampling in deterministic chaos: implications for chaos identification in hydrological processes. *Nonlinear Processes in Geophysics,* 12**,** 557-567.

SAMHOURI, M., ABU-GHOUSH, M., YASEEN, E. & HERALD, T. (2009) Fuzzy clustering-based modelling of surface interactions and emulsions of selected whey protein concentrate combined to [iota]-carrageenan and gum arabic solutions. *Journal of Food Engineering,* 91**,** 10-17.

SAVENIJE, H.H.G., (2001) Equifinality, a blessing in disguise? Hydrological Processes, 15, 2835-2838.

SCHULZ, K. & HUWE, B. (1997) Water flow modelling in the unsaturated zone with imprecise parameters using a fuzzy approach. *Journal of Hydrology,* 201**,** 211-229.

SCHULZE, R. (1989) ACRU: Background, concepts and theory. *Dept. of Agricultural Engineering., University. of Natal, Pietermaritzburg. ACRU Report,* 35.

SCHULZE, R. (1995) Hydrology and Agrohydrology: A text to accompany the ACRU 3.00 agrohydrological modelling system. *University of Natal, Pietermarizburg*.

SEE, L. & OPENSHAW, S. (1999) Applying soft computing approaches to river level forecasting. *Hydrological Science Journal,* 44**,** 763-778.

SEIBERT, J. & MCDONNELL, J. (2002) On the dialog between experimentalist and modeller in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research,* 38**,** 1241.

SHAMSELDIN, A. & O'CONNOR, K. (2001) A non-linear neural network technique for updating of river flow forecasts. *Hydrology and Earth System Sciences,* 5**,** 577-598.

SILBERSTEIN, R. P. (2006) Hydrological models are so good, do we still need data? *Environmental Modelling & Software,* 21**,** 1340-1352.

SOPHOCLEOUS, M. & PERKINS, S. P. (2000) Methodology and application of combined watershed and ground-water models in Kansas. *Journal of Hydrology,* 236**,** 185-201.

SOPHOCLEOUS, M. A., KOELLIKER, J. K., GOVINDARAJU, R. S., BIRDIE, T., RAMIREDDYGARI, S. R. & PERKINS, S. P. (1999) Integrated numerical modelling for basin-wide water management: The case of the Rattlesnake Creek basin in south-central Kansas. *Journal of Hydrology,* 214**,** 179-196.

SUMNER, D. M. & JACOBS, J. M. (2005) Utility of Penman-Monteith, Priestley-Taylor, reference evapotranspiration, and pan evaporation methods to estimate pasture evapotranspiration. *Journal of Hydrology,* 308**,** 81-104.

TAKAGI, T. & SUGENO, M. (1985) Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man and Cybernetics,* 15**,** 116-132.

TAN, B. & O'CONNOR, K. (1996) Application of an empirical infiltration equation in the SMAR conceptual model. *Journal of Hydrology,* 185**,** 275-295.

TAYFUR, G., OZDEMIR, S. & SINGH, V. P. (2003) Fuzzy logic algorithm for runoff-induced sediment transport from bare soil surfaces. *Advances in Water Resources,* 26**,** 1249-1256.

TAYFUR, G., OZDEMIR, S. & SINGH, V.P. (2003) Fuzzy logic algorithm for runoff-induced sediment transport from bare soil surfaces. Advances in Water Resources, 26(12): 1249-1256.

THOMPSON, J., SØRENSON, H., GAVIN, H. & REFSGAARD, A. (2004) Application of the coupled MIKE SHE/MIKE 11 modelling system to a lowland wet grassland in southeast England. *Journal of Hydrology,* 293**,** 151-179.

TSUKIMOTO, H. (2005) Logical Regression Analysis: From Mathematical Formulas to Linguistic Rules. *Foundations and Advances in Data Mining.* Springer Berlin / Heidelberg.

VAN HEERDEN, P., CROSBY, C., GROVÉ, B., BENADÉ, N., THERON, E., SCHULZE, R. & TEWOLDE, M. (2009) Integrating and updating of SAPWAT and PLANWAT to Create a powerful and user-friendly irrigation planning tool. *Water Research*

Commission, Pretoria, South Africa, WRC REPORT NO. TT 391/08 ISBN 978-1-77005-828-6.

VAN LITH, P. F., BETLEM, B. H. L. & ROFFEL, B. (2002) Fuzzy clustering, genetic algorithms and neuro-fuzzy methods compared for hybrid fuzzy-first principles modelling. *Systems Analysis Modelling Simulation,* 42**,** 597-630.

VENTER, J, (2008), Personal communication.

WHITTAKER, G. (2004) Use of a Beowulf cluster for estimation of risk using SWAT. *Agronomy Journal,* 96**,** 1495.

XIONG, L., SHAMSELDIN, A. Y. & O'CONNOR, K. M. (2001) A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system. *Journal of Hydrology,* 245**,** 196-217.

YAGER, R. & FILEV, D. (1994) Approximate clustering via the mountain method. *IEEE Transactions on Systems Man and Cybernetics,* 24**,** 1279-1284.

ZADEH, L. A. (1965) Fuzzy sets. *Information and control,* 8**,** 338-353.

ZHAO, L., YANG, Y. & ZENG, Y. (2009) Eliciting compact T-S fuzzy models using subtractive clustering and coevolutionary particle swarm optimization. *Neurocomputing,* 72**,** 2569-2575.

ZIMMERMANN, H. (2001) Fuzzy set theory--and its applications, *Kluwer Academic Publication.*