UNCERTAINTIES IN MODELLING HYDROLOGICAL RESPONSES IN GAUGED AND UNGAUGED SUB-BASINS

MADAKA H TUMBO

UNCERTAINTIES IN MODELLING HYDROLOGICAL RESPONSES IN GAUGED AND UNGAUGED SUB-BASINS

This thesis is submitted in fulfillment of the requirements for the degree of Doctor of Philosophy at Rhodes University

By

Madaka Harold Tumbo

May 2014

Supervised by

Professor D.A Hughes

And

Dr. Joel Nobert

ABSTRACT

The world is undergoing rapid changes and the future is uncertain. The changes are related to modification of the landscape due to human activities, such as large and small scale irrigation, afforestation and changes to the climate system. Understanding and predicting hydrologic change is one of the challenges facing hydrologists today. Part of this understanding can be developed from observed data, however, there often too few observations and those that are available are frequently affected by uncertainties. Hydrological models have become essential tools for understanding historical variations of catchment hydrology and for predicting future possible trends. However, most developing countries are faced with poor spatial distributions of rainfall and evaporation stations that provide the data used to force models, as well as stream flow gauging stations to provide the data for establishing models and for evaluating their success.

Hydrological models are faced with a number of challenges which include poor input data (data quality and poorly quantified human activities on observed stream flow data), uncertainties associated with model complexity and structure, the methods used to quantify model parameters, together with the difficulties of understanding hydrological processes at the catchment or sub-basin. Within hydrological modelling, there is currently a trend of dealing with equifinality through the evaluation of parameter identifiability and the quantification of uncertainty bands associated with the predictions of the model. Hydrological models should not only focus on reproducing the past behaviour of a basin, but also on evaluating the representativeness of the surface and sub-surface model components and their ability to simulate reality for the correct reasons. Part of this modelling process therefore involves quantifying and including all the possible sources of uncertainty.

Uncertainty analysis has become the standard approach to most hydrological modelling studies, but has yet to be effectively used in practical water resources assessment. This study applied a hydrological modelling approach for understanding the hydrology of a large Tanzanian drainage basin, the Great Ruaha River that has many areas that are ungauged and where the available data (climate, stream flow and existing water use) are subject to varying degrees of uncertainty. The

Great Ruaha River (GRR) is an upstream tributary of the Rufiji River Basin within Tanzania and covers an area of 86 000 km^2. The basin is drained by four main tributaries; the Upper Great Ruaha, the Kisigo, the Little Ruaha and the Lukosi. The majority of the runoff is generated from the Chunya escarpment, the Kipengere ranges and the Poroto Mountains. The runoff generated feeds the alluvial and seasonally flooded Usangu plains (including the Ihefu perennial swamp). The majority of the irrigation water use in the basin is located where headwater sub-basins drain towards the Usangu plains.

The overall objective was to establish uncertain but behavioural hydrological models that could be useful for future water resources assessments that are likely to include issues of land use change, changes in patterns of abstraction and water use, as well the possibility of change in future climates. The two models selected for use in this study are the GW_PITMAN and SWAT2009 models, both of which have been used successfully within the regions by previous studies and both of which can be applied within uncertainty frameworks. The GW_PITMAN uncertainty approach adopted in this study provides for the generation of ensemble predictions in both gauged and ungauged basins. Regionalised indices of hydrological response are extracted from observed stream flow data coupled with some information about physical basin properties and used to constrain ensembles of model predictions. The objective is to reduce the predictive uncertainty by constraining model output to expected sub-basin behaviour. Six constraints have been considered for this study and they are mean monthly discharge (MMQ in $m^3 * 10^6$), mean monthly ground water recharge (in mm), three flow duration curve percentage points (Q_{10}, Q_{50}, and Q_{90} expressed as fractions of MMQ) and the percentage of zero flows. The constraints are also quantified as uncertainty ranges (minimum and maximum) and were estimated for 48 sub-basins within the Great Ruaha River basin using six regional grouping that were assumed to have broadly similar response characteristics.

The approach that has been adopted involves two main steps. The first step only simulates the incremental contribution of each sub-basin (i.e. no downstream routing), only treats some of the 15 natural hydrology parameters as uncertain (using uniform or normal distributions) and is

designed to generate at least 2 000 behavioural parameter combinations based on ranges established for 6 constraints. The constraint values were quantified from observed flow data and available regional information on natural hydrological behaviour and it is expected that the constraint boundaries should reflect the uncertainty in this available information. Following this understanding, the model is run for each sub-basin up to 100 000 times with simple independent Monte Carlo sampling of the parameter distributions. An ensemble is considered behavioural if the simulated characteristics fall within all 6 constraint bounds and these parameter sets are saved for use in step 2. The second step in the process uses all the model components (including those used to quantify water use) and the full model is run with all sub-basins linked to generate cumulative stream flow volumes. The natural parameter values were sampled from the 2 000 saved behavioural sets (i.e. the full set of saved values), while the other parameters were sampled independently from their defined distributions. Water use parameters were included in all sub-basins with irrigation schemes, afforestation parameters were included in sub-basin 1ka32a, transmission losses (from the channel to ground water) in the downstream arid areas and wetland parameters in 1ka71a and 1ka27, representing the Usangu Plains.

The outputs from step 2 were compared with observed stream flow data and further constraints used to limit the final ensemble set. The results from this analysis are satisfactory in that the simulated flow duration curve ranges bracket the observed curves at two gauging stations that are important sites for water resources development decision making and are downstream of many ungauged sub-basins. The 2-step uncertainty analysis approach using regional constraints on hydrological response has proved to be useful in organising a substantial amount of uncertain information into a coherent model of the whole basin and is appropriate for the GRR basin and other data scarce basins given the available input data and the large spatial and temporal variability in both climate and geology.

The uncertainty analysis in SWAT was implemented using the Sequential Uncertainty Fitting (SUFI-2) approach. SUFI-2 parameter uncertainty accounts for all sources of uncertainties including the input data, model structure and parameters. All uncertainties are quantified by a measure referred

to as the P-factor, which is the percentage of measured data bracketed by the 95% prediction uncertainty (95PPU) and R-factor which is the measure of the width of the uncertainty band. Behavioural and non-behavioural parameters are separated based on the objective function. Parameter uncertainties and prediction uncertainties were calculated and statistics gave a good measure of the strength of calibration results. Calibration results yielded satisfactory results for both behavioural and non-behavioural parameters where the coefficient of efficiency (CE) was greater than 0.5. This is because the behavioural parameter sets are within the non-behavioural parameter sets. The SWAT results confirm quite large uncertainty of the simulated discharge due to the large equifinality in parameters and reliability of input data (precipitation and daily evaporation data). Thus further analysis is required to quantify the unexplained uncertainties.

The approach employed in this study represents a major step towards the identification of uncertain, but behavioural, parameters based on understanding rather than on calibration to local observations. The results are encouraging in that the simulated FDC ranges bracket the observed curves at two gauging stations downstream of many ungauged sub-basins that are important sites for water resources development decision making. Thus, the uncertainty approach that has been adopted in this study is appropriate for the GRR basin, given the available input data and the large spatial and temporal variability in both climate and geology. The overall conclusion is that the GW_PITMAN model has been satisfactorily established for the main part of the GRR and that it can be used in the future for further studies and assessments of stream flow scenarios associated with various development options or climate change.

TABLE OF CONTENTS

ACRONYMS

GCM.....Global Circulation Model

IPCC.....Intergovernmental Panel on Climate Change

IRA.....Institute of Resource Assessment

GRR.....Great Ruaha River

RCM.....Regional Circulation Model

SWAT.....Soil and Water Assessment Tool

UGRR.....Upper Great Ruaha River Catchment

URT.....United Republic of Tanzania

NAPA..... National Adaptation Program of Action

SAGCOT.....Southern Agricultural Corridor of Tanzania

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My greatest thanks to my supervisor Prof Denis Hughes for having guided and supported me through the journey towards becoming a scientist. Thanks also for contributing significantly to this PhD thesis with your experience, advice and constructive comments. My acknowledgements also goes to Dr Joel Nobert my second supervisor. Dr Simon Stisen, thanks for your kind and patient help with some of the analysis of rainfall input data and useful advice throughout the course of this study.

I am grateful for funding from DANIDA through CLIVET project that enabled me to pursue this study. I also want to acknowledge the people and organisations that have provided the advice and data that have been used in the thesis 1) Rufiji Basin Water Office 2.) Tanzania Meteorological Agency 3) Soil and Water Assessment Group, Sokoine University of Agriculture for sharing SMUWC reports and water use data 4) Geological Survey of Greenland and Denmark for the time I spent at GEUS during my study visits 5) Dr Raphael Tshimanga during the initial set up of the GW_PITMAN model 6) Dr Patrick Valimba for connecting me with Prof Hughes and useful advice. Thanks to all the friends, colleagues and everyone at the Institute for Water Research, Rhodes University for their nice and warm company, for the encouragement and for the advice and professional support.

Special thanks to my parents, family and friends, you are too many for listing down all your names, but always very dear to me. To my husband, Charles I honestly can never express my appreciation for your patience, or thank you enough for your support. I know I've made life difficult at times, and there are no better words to express this other than Thank you.

1 INTRODUCTION

1.1 Background information

The earth is undergoing rapid changes, and the future is uncertain. Efforts have been made to try to understand natural, economic and social systems and their relation to hydrological processes. This understanding is developed from observed data that are affected by observational uncertainties (Westerberg et al., 2011). Hydrological responses of a catchment are sensitive to changes in society (population increase), land cover/use and climate. Some hydrological and societal changes are a result of natural forces, and some are human induced and range from slow to abrupt (Montanari et al., 2013). The relationships between these changes are complex because of the feedback mechanisms acting on society and environmental systems (Warburton et al., 2012). This change may also lead to an increase in the amount of randomness in the catchment and a decrease in the predictability of the system (Tsonis, 2004). Without adequate observational data to monitor such changes, scientists rely increasingly on estimation methods and models of environmental systems.

In the Great Ruaha River basin, as well as many other basins in Tanzania, hydrological monitoring networks are generally inadequate, and there is a poor spatial distribution of gauging stations within the basin. Available gauging stations might be useful for both research and operational purposes, but a large amount of data is missing for each gauging station. Tanzania used to have good hydrological data during the colonial period (1920's -1960's) after which there has been a reduction in both data availability and quality. Part of this problem has been due to the IMF/World bank structural adjustment programmes in the 1990's coupled with poor economic development such that the government could not sustain the management of hydrological monitoring networks (SMUWC, 2001).

Hydrological models are fundamental tools for understanding historical variations of catchment hydrology and projecting trends into the future. The applications of hydrological models have a variety of objectives depending on the problem that needs to be investigated (Pechlivanidis et al., 2011; Beven, 2001). Although hydrological models have been used to solve

water resources problems around the world, they are faced with a number of challenges which include poor input data (quality, short time series, poor spatial distribution of gauging stations and lack of monitoring stations), uncertainties associated with model parameters, model complexity and structure and the influence of poorly quantified human activities on the hydrology of the basin. All these challenges affect the ability of the model to reproduce observations or represent reality. Both input data, as well as model simulated results, are associated with uncertainties and it is important to consider all sources of errors and their inherent uncertainties (Westerberg et al., 2011). Data unavailability is not the only source of uncertainties in hydrological modelling. The runoff generation process dynamics, underlying observed stream flow responses, are not yet well understood in most catchments, and this brings challenges to quantitative water resources assessments (Hrachowitz et al., 2013a). In hydrological modelling, a lack of understanding of basin hydrological processes means that we might be able to get the right answers, but not always for the right reasons (Kirchner, 2006). In complex environmental systems that are also impacted by social systems, getting the right answer is not always sufficient. If the real dynamics of the system are not properly represented, management decisions might be adversely affected.

The ongoing debate on the impact of climate and land use change on catchment hydrology has not reached a consensus about which change is the dominant driver of hydrological change. The improved understanding of different drivers is important for the understanding of hydrological changes, the direction of change and how it is going to impact water resources availability. Hydrological change involves a complex interaction between land use and climate with population increase influencing both drivers at temporal and spatial scales. The water sector is one of the major cross-cutting sectors as it affects all social and economic sectors and the availability and accessibility of water determines the achievement of the Sustainable Development (SD) and the Millennium Development Goals (MDG). A better understanding of the significant drivers of change in the hydrological cycle, in terms of environmental and social conditions, including the impacts of population increases, is crucial to improving water resources management. In the light of this challenge, the International Association of Hydrological Sciences(IAHS) has declared Panta Rhei (“Everything Flows”) as the new scientific decade focusing mainly on understanding of the major drivers of hydrological change and how

these drivers interact with the hydrological system and society (Montanari et al,2013). Efforts have been made to understand hydrological change but focused on individual drivers of change separately. For example, most climate change research inadequately considered land use change and population increase issues in the assessments and few studies considered both impacts of climate and land use change. Warburton et al. (2012) investigated the combined effects of environmental change on catchment hydrological response. Results from this study indicated that the impacts of environmental change varied across both spatial and temporal scales with the nature of land use and projected climate change having significant impacts on the hydrological response. Further it was found that catchment uniqueness and complexities have unique thresholds where changes start to impact the hydrological response of the basin (Wurburton et al., 2012).

To evaluate the variability in the hydrological response due to environmental change the most common method is to run a hydrological model driven by various climate scenarios from global (GCM) or regional circulation (RCM) models, together with trends in land use change as input forcing data (Gosling et al.,2011; Kim et al.,2013). The simulations of hydrological indices, such as river runoff, can then be used to assess the potential impact of land use or climate change and to inform policy- and decision-making. Although hydrological models are powerful tools for assessing impacts of environmental change, it is important to recognise the inherent uncertainties referred to above. Similarly, climate change projections are also subject to a number of uncertainties (Beven, 2011). A review of the current state of modelling by the World Climate Research Program (WCRP) indicated that projections from the current generation of climate models are insufficient for providing accurate and reliable predictions of regional climate change, including statistics of extreme events and high impact climate, which are required for local and regional adaptation strategies (UNEP, 2009). GCM/RCM insufficiencies are due to limitations in data availability, quality and model validity under changing environmental conditions and lack of understanding of atmospheric systems and their interactions with the oceans. Uncertainties in climate models come from multiple sources such as incomplete model formulations, future scenarios, input data and inability of the models to capture spatial and temporal variability in the climate system. Climate model evaluation is a fundamental step in estimating the uncertainty in future climate projections (Kim et al., 2013).

Since decision making for water resources, monitoring and management is sensitive to the information generated by GCM/RCM's it is essential to understand their strengths, weakness and uncertainties.

A good understanding of the capability of the model to simulate historical runoff and the inherent uncertainties is a pre-requisite before using hydrological models for impact studies. The estimation of the uncertainty of the simulated observations from hydrological models requires a formulation of the uncertainty estimation procedure and establishing the proper tools for implementing the established procedure. This involves the choice of the models, input data and observed stream flows which influence the uncertainty of the simulated output (Beven 2005). Establishing a general prediction error is difficult in hydrological modelling (Beven 2005), and one of the proposed solutions involves specification of behavioural parameter sets associated with the inherent error structure (Beven 2005). Uncertainty analysis has become the standard approach to hydrological modelling, but has yet to be efficiently used in practical water resources assessment. This study of the Great Ruaha River basin in Tanzania is based on the use of regional estimates of mean runoff, groundwater recharge and points on flow duration curves (FDCs) to constrain ensemble outputs from the Pitman monthly rainfall-runoff model (Hughes, 2013) using simple Monte Carlo parameter sampling and the implementation of the Sequential Uncertainty Fitting Version 2 (SUFI-2) approach for the uncertainty analysis in the Soil and Water Assessment Tool (SWAT) model (Abbaspour, 2008) The aim is to establish the models and a robust approach to their application for practical water resources assessments and impact studies.

1.2 Problem Statement

The Great Ruaha River (GRR) basin is critical to the economy and ecosystems of Tanzania. Water is the key resource in the Great Ruaha River basin, in that without it other resources, natural and human, cannot be sustained (SMUWC, 2001). A variety of stakeholder groups depend on water for human, wildlife and livestock sustenance, irrigation, and hydropower generation. The wetlands in the basin owe their existence and nature to the balance between the inflow and outflow of water and support ecosystems downstream of the wetlands (WWF, 2010). The ecology of the Ruaha National Park, which is the largest national park in Tanzania,

is supported by the flow regime of the GRR (SMUWC, 2001; WWF, 2010). Moreover, the distribution of flora and fauna in and around the wetlands is also mostly controlled by water availability. Any significant changes in the magnitude or timing of runoff induced by changes in climatic variables and/or land cover and water use would have significant implications for the economic prosperity of the Great Ruaha River basin in particular, and Tanzania in general (WWF,2010). Decreased flows in the GRR have been recorded since the early 1990s, resulting in complete drying of parts of the river in 1993 (SMUWC, 2001). This has had severe impacts on both terrestrial and aquatic ecosystems. Within the basin, there has been concerns about rising conflicts over water availability for irrigation, livestock, and hydroelectric power generation (SMUWC, 2001). Knowledge of the major changes in water resources is a key element to understanding the impacts on a variety of other sectors depending on the water resources of the GRR basin. It is likely that any future developments will occur against the background of hydrological change. Therefore a good understanding of the current hydrological regime before climate change impacts can be simulated and water management policies developed is necessary. This understanding will only be achieved through modelling approaches that can provide reliable predictions in an ungauged basin.

Based on the scientific contribution of the Predictions in Ungauged basins (PUB) decade, and the application of these in practice, it is possible to apply uncertainty analysis methods in water resources assessments. There has been some reluctance of practitioners to adopt new methods, the reason being (among many others) a lack of understanding of uncertainty methods and its importance. Other reasons are related to reliance on traditional model calibration and validation methods that are not applicable in ungauged basins as well as how scientists have communicated their uncertainty methods and results to practitioners. This thesis presents an uncertainty analysis approach to modelling the GRR basin and its practical application to enhance informed water resources management decisions.

1.3 Significance of the Study

Tanzania, under the policy of Kilimo Kwanza (Agriculture First), launched an initiative known as the Southern Agricultural Corridor of Tanzania (SAGCOT) at the World Economic Forum (WEF) Africa Summit in 2010. This is an inclusive, multi-stakeholder partnership, which aims at rapidly

developing the region's agricultural potential through large and small scale irrigation schemes. The Southern highlands zone which includes the Rufiji River Basin and its sub-basins (Great Ruaha, Kilombero, Luwegu and Lower Rufiji) has been identified as a potential area for the implementation of the SAGCOT. The key challenge is that the implementation of the SAGCOT initiative comes at the expense of water resources and other ecosystem services. In this regard, appropriate water resources assessment and practical adaptation to climate change within the SAGCOT initiatives will be important in ensuring that the implementation of the Kilimo Kwanza policy succeeds. Therefore, a good understanding of the factors influencing hydrological change in the basin including the potential impacts of climate and development change is of great importance. This can only be achieved through the availability of appropriate tools for water resources assessment and the understanding of all sources of uncertainties in a given assessment. This study will build on past research conducted in the area that were mostly concentrated on the Usangu plains. These studies include:

- RIPARWIN (2006): Raising Irrigation Productivity and Releasing Water for Inter-sectoral Needs.
- SMWUC (2001): Sustainable Management of The Usangu Wetlands Project
- Kashaigili (2006): Dynamics of the Usangu plains wetlands through the use of remote sensing and GIS as management decision tools.
- Kashaigili et al. (2008): Impacts of land-use and land-cover changes on flow regimes of the Usangu wetland.
- Lankford et al. (2009): Water Competition, Variability and River Basin Governance: Critical Analysis of the Great Ruaha River, Tanzania.
- WWF (2010): Environmental Flow Assessment of the Usangu Plains. DANIDA/WORLD BANK (1995): Water Resources Management in the Great Ruaha Basin.

These studies were influenced by the perceived needs of donor agencies and/or the Tanzania Government, where most of them concentrated on the causes of the drying up of the GRR, land cover changes, conflicts in water resources as well as wetlands management and environmental flow assessments. Further, the studies focused on the upstream part of the GRR basin (Usangu plains and Ihefu wetlands), and this has led to a bias in the availability of observed data (rainfall and discharge) compared to other parts of the basin. Thus, there is an

apparent knowledge gap on issues related to hydrological change due to climate and development and their inherent uncertainties. This study applies a hydrological modelling approach for understanding changes in the flow of the GRR basin as a result of development impacts and the propagation of uncertainties in hydrological modelling.

1.4 Aim and objectives of the research

Environmental changes are generally expected to lead to an intensification of the hydrologic cycle, which will impact the hydrologic regime and water resources availability (Ferguson and Maxwell, 2012; Montanari et al, 2013). This may lead to an increased need for integrated approaches to problem solving. Hydrological models have the ability to include all components of catchment runoff generation (natural and anthropogenic) and therefore they can be used to assess a range of scenarios of land use change, climate change and future water resources availability and management. ***The overall purpose of this thesis was to apply a hydrological modelling method for understanding the hydrology of the GRR basin taking into account uncertainties arising from various sources including data inputs, model structure and model parameters.*** The assessment involved an ensemble approach in understanding how different approaches to hydrological modelling can result in robust outputs and whether it is possible to select the best modelling approach for impact assessment studies. The key questions governing this research have been highlighted by different authors in the literature as fundamental issues in hydrological science and hydrological modelling approaches such as Beven (1989); Refsgaard (1997); Beven (2001); Yadav et al.,(2007); Montanary and Toth (2007) and Westerberg et al. (2011). The study was designed to answer the following three key questions.

1. *To what extent do physical basin properties affect runoff generation and how can this understanding be incorporated in model simulations?*

Hydrologic similarity between catchments is the basis for classification, for transferability of model parameters or model constraints, for regionalisation and also for understanding the potential impacts of environmental change. Catchments exhibit a degree of uniqueness and complexity (Beven, 2000), and this creates challenges in hydrological regionalisation. However, there is a level of organisation related to natural self-organization or co-evolution of climate,

soils, vegetation and topography (Sivapalan, 2005). This question recognizes the potential value of different types of data for development of first order grouping of hydrologically similar catchments. This part of the study is based on cluster analysis for improved understanding of how catchment groups are related based on identified characteristics and to use this understanding to develop regional constraints on model outputs and guide the development of appropriate model parameter sets in un-gauged basins. The constraints for all the sub-basins of the GRR were established and quantified on the basis of the available observed stream flow data, as well as some assumptions about the response characteristics of the un-gauged sub-basins.

2. *How can uncertainties in the input parameter and observed data be estimated and how can this information be used to inform the model calibration process?*

Uncertainty needs to be distributed appropriately and consistently across the sub-basins for both upstream and downstream areas. It is difficult to achieve this for an un-gauged basin or partially gauged basin characterized by high spatial variability in geology and climate and influenced by development impacts. The most important aspect of any uncertainty approach to hydrological modelling is not the method of generating uncertainty ensembles, but the method used to evaluate those ensembles and to decide which can be considered behavioural (Beven, 2012) in the face of uncertain observed data, or no observed data, for comparison purposes. An uncertainty approach has been developed to represent realistic parameter uncertainty for both gauged and ungauged- basins.

3. *What are the most appropriate model calibration or uncertainty parameter sampling approaches to use in data scarce regions?*

This research question is approached through the application of different modelling approaches that allow the integration of all available information within an uncertainty framework and is directed at establishing suitable models for water resources assessment which includes natural hydrology, development impacts and climate change. Two models that have been previously applied in the region (SWAT and GW_PITMAN) were selected for use and were applied with different approaches that included manual, automatic and uncertainty ensemble methods. The value of the different modelling approaches is assessed using selected

objective functions to compare simulated and observed stream flows at the available gauging sites within the GRR basin.

1.5 Thesis structure

Chapter 2 reviews the background concepts and research concerned with key issues in hydrological modelling, data and uncertainty and climate change. The chapter also provides an overview of the GW_PITMAN and SWAT hydrological models that are used in the research. Chapter 3 provides a detailed study area description and data synthesis of the GRR basin. It also provides a conceptual understanding of the basin and reviews the data that are subsequently used in the development of hydrological models. Hydrological modelling methods are presented in Chapter 4. This chapter includes details of the initial model calibration, performance criteria, and uncertainty analysis Similarity analysis, and development of regional constraints are detailed in Chapter 5. The GW_PITMAN model set up, calibration and uncertainty analysis methods and results are given in Chapter 6. Chapter 7 provides the details of the SWAT model set up and calibration and uncertainty approach. A parameter sensitivity analysis is also included in order to define those parameters that are then taken forward and subject to automatic calibration in the Sequential Uncertainty Fitting framework (SUFI). In chapter 6 and 7 modelling results and uncertainty are reviewed with reference to the identified research questions. Chapter 8 provides summaries of the major conclusions, a critical discussion of the research results and recommendation for further work. Some preliminary results of using the established behavioural model in climate change impact analysis are also presented in this chapter as an indicator of one of the directions for future research.

2 HYDROLOGICAL MODELLING, DATA AND UNCERTAINTY IN THE FACE OF ENVIRONMENTAL CHANGE

2.1 Introduction

A literature review on the building blocks, and some of the main challenges regarding data and uncertainties associated with hydrological modelling is presented in this chapter. Section 2.1 provides an overview of modelling hydrological processes, tools, methods, parameters and scale issues, as well as model calibration and validation. This section also reviews the frameworks for uncertainty estimation, performance measures and sensitivity analyses. Section 2.2 explores the issues related to data and uncertainties in hydrological modelling. An overview of approaches to the calibration of hydrological models is given in section 2.3 followed by model performance criteria and uncertainty estimation and analysis (Sections 2.4 and 2.5). Section 2.6 summarises hydrological model use within the southern African Region.

2.2 Modelling hydrological processes at the catchment scale

Hydrological model development involves a mathematical description of water partitioning and runoff generation processes (Gupta et al., 2012; McMillan et al, 2013). Understanding runoff generation processes is essential for successful hydrological model development as well as using models for impact assessments caused by changes in land use or climate (Sivapalan, 2003; Mul, 2009). According to Beven (2000), each catchment is unique, and there have been suggestions that hydrological model structures should be tailored to individual catchments (Kirchner, 2006; Gupta et al., 2008; McMillan et al., 2013.). This is generally not practical in large-scale studies involving many sub-basins and therefore the main issue becomes developing, or selecting, a model that is sufficiently flexible, or detailed, to represent the range of processes expected to occur within the specific study area (Hughes, 2013b). However, determining what level of detail is sufficient or optimal remains a contentious issue within the science of hydrological modelling (Hughes, 2010) and is strongly associated with problems of scale, equifinality and parameter uncertainty. In hydrological sciences, estimating parameter values related to water fluxes and storages at scales of interest is still a challenge, and this is partly associated with a lack of understanding of catchment heterogeneities (Beven, 2001;

Gentine et al., 2011). Catchment changes (land use and abstractions) affect the response of surface hydrology through processes that cannot easily be established at appropriate scales. Conceptualization of catchment hydrological processes is fundamental in any hydrological study as advocated by the Prediction in Ungauged Basins (PUB) science programme, which urged a re-think about the different ways in which the form and function of catchment systems are conceptualized (Sivapalan et al., 2003). A number of contributions have been made to the conceptual understanding of hydrological models, but issues of scale differences between field-based process research and model scale are yet to be resolved (Beven, 2001, Hughes, 2004; Kapangaziwiri, 2010; Gentine et al., 2011). According to Schertzer et al., (2010), nonlinearities are general features of natural physical systems, while the laws governing physical models are founded on linear approximations (Gentine et al., 2011). In these models, the laws governing hydrologic processes have been developed at scales ranging from ~1-100 m and may not accurately represent the processes that occur on a larger scale (Gentine et al., 2011; McDonnell et al., 2007). For example, observations that might have been made at the hillslope scale cannot be generalized over the whole catchment or even over the neighboring catchment (Weiler and McDonnell, 2007).

Catchments are heterogeneous at all scales because of the variability in input (precipitation), geology (material, bedrock topography, and fractures), soil (matrix, macropores, soil moisture, thermal status, and layering), vegetation or land cover, and topography (Tetzlaff et al., 2010). Given the importance of heterogeneity at all scales, process conceptualization requires the development of diagnostic classification tools that consider factors such as topography, topology and typology to develop indices of similarity for prediction in un-gauged or sparsely gauged basins (Buttle, 2006; Carillo et al.,2011; Tshimanga, 2012a). It is important to note that the description of the heterogeneity of processes and consequent complexity will not always improve the ability to extrapolate and predict, particularly when we recognize that forms of heterogeneity and complexity may themselves be subject to change over time (Troch et al., 2009). Similarly, methods that are used for catchment similarity analysis may also be subjective, and this may lead to biased results. Information on the internal behaviour of catchments is necessary to identify similarities and differences between catchments in terms of their response to natural and anthropogenic disturbance and change (Buttle, 2006;

Tshimanga, 2012a). For this reason, a universal approach to catchment characterization and process conceptualization remains important, but difficult, because catchments are unique (Sivapalan, 2005; Wagener et al., 2007). McMillan et al. (2013) developed a diagnostic method to assess spatial variability in hydrological processes targeting various catchment processes such as flow volume, storage–discharge relationships, threshold responses to rainfall and soil moisture and physical basin characteristics. The results showed a complex picture where different diagnostics showed different patterns of hydrological processes as well as large variations in processes even at the small scales involved ~10 km (McMillan et al., 2013). Based on this study, catchments that share similar characteristics showed similar patterns of response. Headwater sub-catchments were similar to each other (higher elevation, steep, forested), and were differentiated from downstream sub-catchments (lower elevation, shallower slopes, pasture). However, this pattern did not appear in the patterns of diagnostic indices, demonstrating the difficulties associated with defining realistic *a priori* estimates of spatial variability in catchment processes (McMillan et al., 2013).

2.3 Hydrological models

A model (whether mathematical, numerical or scale) is a simplification of reality. The simplification is required because of the complexity of the natural systems which makes it difficult to include all the processes and interactions occurring in nature (Davie, 2008). Models play an important role in hydrological modelling because of the natural variability of water resources in time and space and the fact that resources available to sustain long-term monitoring programmes are limited (Hughes, 2004). In designing a hydrological model, developers are normally trying to build as good a representation of hydrological reality as they can, given their understanding of the key hydrological processes and the ability to represent these as a series of equations.

The PUB decade achieved improvements in hydrological modelling science as well as in the tools and approaches for model applications in ungauged basins (Hrachowitz et al., 2013a; Blöschl et al., 2013). Despite these achievements, challenges still remain because models need to be useful in addressing water resources management problems under changing conditions (Montanari et al., 2013; Hughes, 2010; Hughes, 2013a). Any hydrological model must be able

to take into account both the influence of land use and the spatial variability of catchment hydrology. In theory, this can be achieved by using fully distributed models which describe each hydrological response unit through physically consistent formulation and parameters related to physical basin properties, for example, the MIKE SHE model (Refsgaard, 1995; 1997). However, it is practically impossible to apply these models at scales above very small catchments, so most models end up being merely conceptual because of the difficulties in obtaining good quality data. Even in small catchments, heterogeneity in the soil, topography and vegetation can lead to model bias which may affect model predictability (Tshimanga, 2012a; MacMillan et al., 2013). On the other hand, lumped hydrological models consider the entire hydrological system (catchment, sub-catchment) as a single unit and therefore parameters do not vary spatially and basin response can only be evaluated at the outlet without accounting for individual sub-basin response. According to Beven (2000), lumped models cannot be used for the analysis of event scale processes unless the focus is on discharge prediction only. Further, because of the inherent spatial variability of landscapes, lumped models are inadequate when calibrating ungauged basins because parameters used in lumped models represent spatially averaged characteristics which cannot be compared with field measurements (Beven, 2001; Sivapalan, 2003; Tshimanga, 2012; Wang et al., 2012). A compromise approach is to make use of essentially lumped conceptual model formulations, but apply them in a semi-distributed model, where individual sub-basins have their own climate inputs and parameter sets (Hughes, 2013a). Some lumped and semi-distributed models also make use of probability distribution function approaches to allow for internal heterogeneity (or sub-grid) effects within each sub-basin (Pitman, 1973; Moore, 1985).

2.4 Regionalisation in hydrological modelling

A lot has been gained from the implementation of the PUB principles. Some of the experiences are related to parameter estimation, scale and regionalization issues, model structures and various aspects of uncertainty quantification and assessment. Transfer of hydrologic characteristics of basins from gauged to ungauged catchments is one of the most fundamental challenges in hydrological modeling (Sivapalan, 2003). Regionalization approaches have been used in most hydrological studies (Schaake et al. 1997; Merz and Bloschl 2004; Tshimanga 2012; Westerberg et al., 2013) and this has stimulated the use of other sources of hydrologic

data such as spatial datasets and meteorological data including satellite and radar datasets. Similarity analysis approaches based on hydrological homogeneity provides a promising step towards the development of a robust regionalization approach.

Geomorphological and climatic characteristics of the basin play an important role in runoff generation and hydrological response of the basin (Tshimanga, 2012a). The relationship that exists between physical basin characteristics and hydrological indices is important in hydrological modelling because information generated from this relationship is important in the understanding and estimation of runoff generation in areas where data are limited. Even in areas where historical observed data exist, the quality of these data could be questionable. Spatial datasets (land cover, elevation, slope, and soils) are used to generate landscape characteristics which are then correlated with climate and hydrological signatures for the purpose of understanding runoff generation process in the basin and guide the regionalization process. Flow duration curves are powerful tools and have been used for understanding catchment characteristics (Westerberg et al., 2013). Various methods have been developed to construct and regionalize flow duration curves based on sub-basin characteristics (Tshimanga, 2012a). Following this understanding physical basin characteristics, climate information and hydrological datasets can be used to evaluate the relationship between various variables and use the correlation results to identify homogenous sub-basins. In turn, the information and understanding generated from this analysis and results, helps in guiding the process of assigning initial parameter sets and constraints on model outputs (Hughes et al.,2013c) for uncertainty analysis.

The process of runoff generation is influenced by watershed properties, such as topography, slope, soil and vegetation type, and by climatic driving forces, such as precipitation and temperature. For distributed models, a large amount of spatial data is needed to develop a proper understanding of the land surface for hydrological modelling. Land surface topography is a determinant variable for water flow, and it is, therefore, important to analyze the topographical properties of the basin before modelling the hydrological processes (Carrillo et al., 2011). Catchment characterization helps in synthesizing our understanding of how different catchment characteristics interact to give hydrological response. Variability in catchment

hydrological response can be quantified by means of specific indices of catchment behavior such as runoff coefficient, flow duration curve and characteristic drainage time scales (Carrillo et al., 2011).

In addition to quantification of specific indices, a catchment can be classified into units with the same hydrological response characteristics (hydrologic signatures). This gives a profile of hydrological behavior within the region (Carrillo et al., 2011). Topography alone is not sufficient to understand all the details of the rainfall-runoff transformation (McDonnel and Woods, 2004; Sawicz et al., 2011). Other spatially variable factors such as soil type, vegetation, and land use are also important. One way to simplify distributed hydrological modelling and still account for spatial variability is to divide the basin into homogeneous hydrological response units (HRUs). These are areas that have common properties which are important from the hydrological perspective; for instance similar slope, elevation, aspect, soil type, vegetation cover, and land-use.

2.5 Calibration and validation of hydrological models

Generally, models attempt to represent the complex spatially distributed interaction of water, energy and vegetation by means of mathematical equations (Wagener and Gupta, 2005) and parameters representing the specific response characteristics of a basin. In the past, parameter values were quantified through calibration against available observed data. Model calibration has been an integral part of hydrological modelling as many parameters in the model consist of different values for different slope, land cover and soil combinations for each hydrological response unit. Parameters obtained through calibration depend on the model structure, quality of data and physical basin characteristics of a particular area. However, the calibration process may be subject to the problem of equifinality, where there might be different parameter combinations within a model that can simulate the river flow of a catchment in the same way but may come from different regions of the total possible parameter space (Beven 2000; Beven, 2001; Beven and Freer, 2001). Therefore, it is possible that some of the parameter combinations are simulating the stream flow patterns correctly but for the wrong reasons with respect to hydrological processes. Hence, calibration approaches should focus on

trying to find behavioural parameters that represent what we know about the catchment response (Wagener et al., 2003; McDonnell et al., 2007).

Estimating errors in observed data (climate and stream flow) stem from both the instruments and the methods used for observation (Westerberg et al., 2013). This, in turn, causes uncertainty in parameterization since it is possible for models to be adequately calibrated, but with incorrect climate inputs or against incorrect river flow (Beven, 1989; Kirchner, 2006). Errors and uncertainties in the input data can lead to error in the representation of hydrological processes and boundary conditions further contributing to the problem of equifinality and uncertainty in the validity of the results (Westerberg et al., 2013). Any errors and uncertainty in the validity of results obtained through calibration in gauged sub-basins will inevitably be propagated to ungauged sub-basins regardless of the methods used to transfer parameters (Kapangaziwiri, 2010; Kapangaziwiri et al., 2012).

Fitting simulations against observed data is not enough to define catchment response and resolve parameter equifinality. The key issue is that a long-time series of good quality data are required, but these are scarce in many parts of the world, particularly Africa. Data scarcity and a decline of hydrological networks cause regional hydrologic predictions to be highly uncertain (Hughes et al., 2008; Koutsouris et al., 2010), further limiting our understanding of catchment processes in ungauged basins because it is difficult to understand and manage the unknown. Data collection efforts and methods need improvement and with improved measurements and quantification of catchment characteristics we should be more certain about our model predictions. Without an improvement in data collection methods, regionalization and parameter transferability is unlikely to work. In sparsely gauged basins, there are discharge data which can provide important information about catchment behaviour but it is not always easy to establish the quality and length of the time series that is necessary to obtain adequate model parameterization (Hrachowitz et al., 2013a). A number of research findings have suggested that different observed data lengths are necessary for parameters to be determined adequately, and a range of 2-5 years has been proposed (Vrugt et al.,2006; Merz et al.,2009), but this period is too short for highly variable systems. Regardless of the record length, the

quality of the collected data should be thoroughly evaluated, and their uncertainties quantified (Westerberg et al., 2011).

The progress in hydrological process understanding through improved modelling tools has highlighted the weaknesses of automatic, single-objective calibration approaches (Efstratiadis and Koutsoyiannis, 2010). This became evident when Gupta et al., (1998) discussed the advantages of multi-objective model calibration and its application as a way to extract more information from observed time series by using several measures of performance. This has led to the evolution of the model diagnostic area (Gupta et al.,2008) where evaluation and identification of parameters are supposed to satisfy multiple hydrologic responses such as low flows, high flows, the total water balance and others (Kollat et al.,2012). As a result, a realistic model should be able to satisfy both multi-objective and multi-criteria calibration and consistently reproduce signatures of the hydrological response so as to ensure a resemblance of the catchment behaviour and the statistical properties of the modelled variables (Kollat et al., 2012; Harchowitz et al., 2013b). Further, the use of hydrological signatures helps to obtain practical results and can reduce the impact of input data errors during the calibration process (Gupta et al., 2008; Wagener and Montanari, 2011; Casper et al., 2012; Euser et al., 2013). According to Casper et al. (2012), signature indices may be used to assess changes in water balance; therefore, they can be also be used as indirect measures of efficiency for model output. These indices can be derived from flow duration curves (FDCs) or distributions of runoff event coefficients (Yadav et al., 2007; Harchowitz et al., 2013b). Moreover, indices may also be developed using various data sources and information such as ground water levels; tracer information and satellite data (Euser et al., 2013). However, the use of signatures has both advantages and disadvantages; the advantage is that signatures are interpreted in terms of the underlying processes rather than the aggregate performance measure as they are constructed to reflect a specific aspect of the system behaviour (Euser et al., 2013). The disadvantage is that a signature represents certain aspects of the catchment response at the expense of others and it is necessary to consider multiple signatures to characterize system behaviour (Euser et al., 2013).

Validation of a hydrological model takes place after calibration to test if the model performs well on a portion of data, or in different areas, which were not used in the calibration process. Model verification aims to validate the model's robustness and ability to describe the hydrological response of the region and further detect any biases in the calibrated parameters (Gupta et al., 2005). However, model performance is usually better during the calibration than the verification period, a phenomenon called model divergence (Sorooshian and Gupta, 1995). When the degree of divergence is considered unacceptable, the modeller has to examine the model structure and the calibration procedure for inappropriate assumptions and revise the model or the parameter set accordingly.

2.6 Model performance evaluation

In the literature, a number of performance measures exist, however, the most common objective functions are based on the standard least squares and maximum likelihood methods (Pechlivanidis et al.,2011). Despite its shortcomings (Schaefli and Gupta, 2007; Gupta et al.,2009), the most popular objective function in hydrological research is the Nash-Sutcliffe Efficiency (NSE) criterion which measures the proportion of the variance of the data explained by the model (Nash and Sutcliffe, 1970). There has been an ongoing debate on how appropriate this criterion is for measuring the goodness of fit, as well as what is an acceptable NSE value (Gupta et al., 2009). There have also been proposals to develop modified versions of the Nash-Sutcliffe criterion to overcome some of the identified problems (Gupta et al., 2009; Pechlivanidis et al., 2011). Other studies have suggested that efficiency should only be used as an index of overall performance as there may be many different parameter sets that achieve equally good performance (Buytaert and Beven, 2011). Beven (2009; 2010) suggests an approach of testing using limits of acceptability and these need to be set before running the model. However, this will only be possible subject to the availability of good quality observed data to define such limits in any simulated hydrological variable.

Parameter identification can either be based on a single objective function or a multi-criteria/multi-objective approach. The use of a single objective function has limited value in that it is based on erroneous assumptions that all available information can be summarized using a single aggregate measure of model performance thus leading to loss of information

(Wagener et al., 2001). The existence of equifinality represents a further problem associated with testing model performance using a single performance measure (Beven, 2006). According to Westerberg et al. (2011) lumped global performance measures are typically influenced by the performance at certain flow magnitudes, such as high or low flows. This issue has been addressed in multi-criteria/multi-objective calibration approaches where different aspects of the fit between simulated and observed discharge are evaluated and also adding performance measures using observations such as soil moisture or actual evapotranspiration (Westerberg et al., 2011; Harchowitz et al., 2013a). Multi-objective approaches have been used in a variety of studies and results suggest that the approach can produce more robust parameterizations (Gupta et al., 1998; Koutsoyianis, 2010; Efstratiadis and Koutsoyiannis, 2010; Harchowitz et al., 2013a). In addition to the multi-objective approaches newer methods and tools (such as FITEVAL system) have been developed which highlights the objective reasoning for accepting Nash-Sutcliffe values above 0.65 instead of 0.5 (Ritter and Muñoz-Carpena,2013). In this approach, the probability distributions for (*NSE* and root mean square error) are derived with bootstrapping followed by bias corrected and accelerated calculation of confidence intervals (Ritter and Muñoz-Carpena (2013). The criteria for testing indicators exceeding threshold values have been proposed where it shows how model performance is not linearly related with *NSE* (Ritter and Muñoz-Carpena (2013) .The proposed tools and approach provide modelers with improved, less subjective and practical model evaluation guidance and tool (Ritter and Muñoz-Carpena (2013).

2.7 Uncertainties in hydrological modelling

Uncertainty is a major concern in hydrological science and water resources assessment planning and management. It is difficult to plan under conditions of uncertainty and yet all observational data, and model simulations are subject to uncertainties (Pappenberger and Beven, 2006). Sources of hydrological model uncertainty include model structure, model parameters, input data and lack of understanding of the catchment processes (Kapangaziwiri, 2010). Availability of good quality hydro-climatological data is still problematic in many parts of the world, and input data uncertainty will continue to be one of the major sources of uncertainties in hydrological modelling (Kapangaziwiri 2010; Hughes et al.,2010; Pechlivanidis et al.,2011; Westerberg et al.,2013; Hughes et al.,2013c). In the past decade, the recognition

of uncertainty in hydrological modelling and the need for quantification of these uncertainties has been advocated (Yadav et al., 2007; Wagener and Montanari, 2011; Kapangaziwiri et al., 2012; Westerberg et al., 2013). Following this recognition a number of approaches of uncertainty-based parameter estimation and quantification approaches (both simple and complex) have been developed and applied in different parts of the world (Beven and Binley, 1992; Wagener et al., 2003; Vrugt et al., 2003; Kavetski et al., 2006; Abbaspour et al., 2007; Kapangaziwiri, 2010). Most recent efforts to deal with uncertainties in ungauged basins concentrate on parameter and/or model structure uncertainty. Uncertainties in observed data have often been overlooked, while this should be the first step in any hydrological modelling research. Uncertainty in observed data affects the information generated by models (climate data inputs) and the evaluation of the model (stream flow data) and it is, therefore, essential to account for these uncertainties (Hrachowitz et al., 2013a; Westerberg et al., 2013).

The need for a realistic assessment of predictive uncertainty has been discussed by many authors (Beven, 2002; Sivapalan et al,2003; Vrugt et al.,2006; Hughes, 2006; Beven, 2006; Hughes et al., 2008; Kapangaziwiri et al., 2012; Hrachowitz et al., 2013; Hughes,2013a). In any attempt to understand uncertainty in hydrological modelling, it is important to note that uncertainty cannot be eliminated but can only be reduced through the improvements in methods of measurements and formulation of the model. Improved consideration of parameter uncertainty is a step forward, though this may not be sufficient because the whole process of identifying behavioural parameters is uncertain (Gupta et al., 1998; Buytaert and Beven, 2011). Uncertainty estimation depends on the quality and availability of data, the model in use, human and technical resources available as well as how the concept of hydrological uncertainty is propagated into decision making (Hughes et al., 2008). In the past two decades, uncertainty estimation efforts have been focused on investigating the sources of uncertainty and the development of approaches for quantifying uncertainty with more focus on methods to quantify parameter uncertainty and less effort on input and model structure uncertainty (Beven and Binley, 1992; Gupta et al., 1998; Vrugt et al., 2003a; Kapangaziwiri, 2010).

2.7.1 Sources of uncertainty

There are various ways in which sources of uncertainty in hydrological modelling can be classified (Beven 2000) and for the purposes of this study four main groups are used:

I) *Natural uncertainties*; These are associated with temporal and spatial variability of natural processes which lead to the introduction of the large amount of randomness in the physical processes that produce the responses of the hydrological system (Singh, 1997). Uncertainties that come from an inherent randomness in the behaviour of the system belong to the category of aleatory uncertainty (Bsalone, 2007; 2008). Aleatory uncertainty is a property of any natural system and cannot be reduced or eliminated by improvement in data (Bsalone, 2008).

II) *Input data uncertainties*; Data scarcity and a decline of hydro-meteorological networks cause regional hydrologic predictions to be highly uncertain. This may also lead to the introduction of errors when interpolation methods are applied across space and time based only on data from a few available observation stations or periods (Jung et al., 2012). Input data that are used to force hydrological models (rainfall and evaporation) and for calibration (discharge) are associated with errors due to measurement and estimation errors.

IIa) *Precipitation input data*: The spatial and temporal variability in rainfall contribute to uncertainty in precipitation data (Pechlivanidis et al., 2011). This variability is due to the complexity of the mechanisms that produce precipitation. Generally, precipitation uncertainty is regarded as the dominant source of uncertainty in rainfall-runoff modelling (Gupta et al., 2005). Remotely sensed rainfall data offer an alternative source of spatially integrated estimates but they are also associated with errors from methods used to analyze raw sensor data and produce rainfall estimates (Sawunyama and Hughes, 2008). Sawunyama and Hughes (2007) quantified input uncertainties associated with the use of rainfall data in the Pitman model for South African basins and results showed that the use of different combinations of rain gauges can have substantial impacts on spatially averaged

rainfall and therefore on simulated stream flow. The impact of spatial precipitation data uncertainty has been studied, and results have demonstrated that denser networks improve model performance (Younger, 2009). The importance of uncertainty in data may depend on whether the model parameters are determined from calibration or from physical measurements and principles. For example, Oudin et al. (2006) showed that systematic errors and uncertainties in rainfall data were transferred to the model as bias in the parameter values.

II.b) *Evaporation input data*: Evaporation demand data also affect uncertainty in model predictions. Potential evaporation is calculated directly through variables such as temperature, wind speed, relative humidity and radiation. Uncertainties in evaporation data arise from the data used in the calculations as well as the methods used for calculation (Sawunyama, 2009). In most applications, uncertainties in precipitation are considered to be more serious than uncertainties in evaporation (Gupta et al., 2005).

II.c) *Discharge data*: Discharge data are normally derived from the rating curve, where a series of streamflow measurement is collected using a current meter. These points are plotted versus the stage and a smooth curve is drawn through the points. There could be significant scatter around this curve. Thus, when using a rating curve, it is important to keep in mind that, the discharge read from the curve is the most likely value, but it can could be different form the measured value. Moreover, since rating curves are developed using few stage measurements, and measurements of high flows are rare, there can be significant errors in rating curve at high levels. Even though discharge values are not direct measurements, but rather estimates of the real and unknown discharge values, they are rarely presented with a statement of their uncertainty in practical applications (Herschy, 2002). Literature reports several studies focusing on the analysis of the different error sources and uncertainty affecting discharge measurements and rating-curves construction (Di Baldassarre and Montanari, 2009; Pappenberger et al., 2006; Di Baldassarre and Claps, 2011; Kiang et al., 2011; Westerberg et al., 2011; Moore et

al., 2012; Westerberg et al., 2013;Juston et al.,2014). These are related to uncertainties in stage-discharge measurements and the whole process of rating curve development (Moore et al., 2012). For example, uncertainty affects the velocity-area method, the mathematical interpolation of stage-discharge relationship, as well as the extrapolation of the curve beyond observed data (Domeneghetti et al., 2012). Moreover, the development of stage-discharge relationships is based on a number of assumptions, some of which inevitably introduce simplifications and errors (Domeneghetti et al., 2012). Again, errors may be associated with water level measurement and width of the river cross-section where the geometry of gauging cross-section is assumed to be stable in time, even though significant changes may occur during high flood events due to erosion, sediment transport, deposition and human impacts. The following possible errors should always be considered during evaluations of rating curves:

- Errors related to measurements combining instrumental, environmental and human errors.
- Time integration errors because of flow non-stationarity during measurements.
- Systemic errors due to flow non-stationarity during measurements.
- Extrapolation to stages that have not been included in the observations.
- Changes in the channel shape and bed-elevation during high flows leading to un-measured changes in the stage-discharge relationship.

The quantification of uncertainty in discharge measurements has been achieved through statistical techniques and on-site observations. The statistical method applies a replication of measurements approach known as Interpolated Variance Estimation (IVE) (Pelletier, 1988; Kiang et al., 2009; Cohn et al., 2013). The IVE estimated uncertainty based on the observed data collected during the stream flow measurement reflects site conditions and captures all sources of random uncertainty in the velocity depth measurements (Cohn et al., 2013).

III) *Model structure uncertainty*; Model structure uncertainty is associated with inappropriate model components relative to the natural system being simulated (Hughes et al., 2008). In hydrological modelling, consideration of hydrological processes and their mathematical representations lead to the selection of a model structure. However, this structure is controlled by our understanding of the hydrological system, which is determined by the available data. Unobserved processes are typically ignored leading to the introduction of uncertainties in modeling results. Other model structural uncertainties are related to the spatial and temporal scales at which the model is designed to operate and the extent to which these represent real spatial or temporal heterogeneity (Gentine et al., 2011; Hughes et al., 2008; Hughes, 2013). The research community has recently emphasized the need for explicit identification of model structure uncertainty and diagnosis of differences in hydrological behaviour between model structures in order to improve model performance (Hrachowitz et al., 2013; Hughes et al., 2013). In general, uncertainties related to the model structure are identified through the model behaviour for runoff properties such as peak discharge, time to peak, runoff volume (Butts et al., 2004) or through time-series diagnostics (Wagener et al., 2003). Using the Pitman model (Hughes et al., 2013a) demonstrated the potential value of reducing spatial scale when there are clear heterogeneities in catchment physical properties and hydrological process responses.

IV) Model parameter uncertainties; Parameter uncertainty arises from the inability of the model structure to locate unique optimum parameter sets given the information available (Wagener and Gupta, 2005). Bias in observed data will inevitably lead to biased parameter estimates while over-parameterization will result in the problem of equifinality (Beven and Freer, 2001). Parameter uncertainty may also result from scaling effects, where the value is not representative of the grid cell or sub-basin scale. The general practice in estimating parameter uncertainty is to encompass the effective model structure uncertainty in parameter uncertainty (Pechlivanidis et al., 2011). However, due to the propagation of uncertainty it is impossible to reduce uncertainty in parameter quantities even if a unique solution is available (Kapangaziwiri, 2010). One way to reduce uncertainty is to have enough information that will assist in parameter

identification e.g. the approach of *a priori* parameter estimation (Kapangaziwiri, 2010; Tshimanga, 2012) which helps in reducing uncertainties associated with the observed input and output data (Kapangaziwiri, 2008) but may be influenced by uncertainties related to measurements of physical basin characteristics (Kapangaziwiri, 2010). The problem of parameter uncertainty is more challenging in poorly gauged or completely ungauged sub-basins where there is no information to guide parameter estimation (Wagener and Wheater, 2006), and overcoming this was one of the major objectives of PUB decade (Hrachowitz et al., 2013a)

2.7.2 Uncertainty Analysis (UA) and Estimation Methods

As discussed in the previous section, uncertainty in model predictions arises from several sources and therefore a realistic assessment of the different sources of errors is important for science-based decision making as well as to direct the research towards model structural improvements and uncertainty reduction (Refsgaard et al., 2006). In hydrology, uncertainty assessment (UA) should be performed with understanding that interactions exist between uncertainty sources. Hughes et al., (2008) has highlighted that, during the model calibration process it is possible to have interactions between quantifying appropriate parameter sets and the quality and representativeness of input data. So, the search for behavioural parameter sets will always be influenced by errors in the input data as well as data used to evaluate the parameter sets (Hughes et al., 2008).

The selection of uncertainty analysis method should consider practical applications regarding computation time, the complexity of implementation within the hydrological model, the modeler's skills and the data availability (Sellami et al., 2013). Therefore, there will be variations in UA techniques used for different regions and countries (Hughes et al., 2008; Sellami et al., 2013). Most uncertainty methods are based on the concept of Bayesian model averaging (BMA) (Ajami et al., 2007) and multi-model ensemble methods (Butts 2004). The BMA method is aimed at evaluating both model structure and parameter uncertainties. The multi-model ensemble approach aims at evaluating model structure uncertainty by sampling from the output distribution. A variety of uncertainty analysis and estimation methods are presented in the literature.

Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992) is based on a Bayesian Monte Carlo simulation technique and was developed to address the problem of equifinality (Beven and Binley, 1992). The idea behind this framework is that if there is no unique optimal model, then parameter sets should be considered on a scale of likelihood because, while one parameter set may yield the best results, it is likely that there will be many other parameter sets which may give similar results. However, when further observational data are considered, the rank of the best parameter sets is likely to change. It should be noted that the GLUE concept is wider than just that of equifinality in parameter values, the basic idea is that all the different uncertainty sources that affect the model (model structure, parameter values, input data, evaluation data, commensurability errors between the scale of the model and the data etc.) results in equifinality in model representations of the studied system (Krueger 2010). In implementing GLUE, a Monte Carlo simulation sampling is used to derive a large number of parameter sets. The relative performance of each set is assessed by comparing model estimates with observed data and objective functions while retaining only those parameter sets that yield acceptable results (behavioural). The main strengths of GLUE (Beven and Freer, 2001; Beven 2006) are that it is straightforward and easy to implement; it is not model specific and allows for any model structure to be evaluated, and it allows for behavioural models to be scattered throughout the parameter space. The major drawback of this method is that since it is based on random Monte Carlo sampling, it requires a large number of model runs, making it computationally intensive. In addition, the selection of the likelihood function and the threshold value within the implementation of the GLUE framework is subjective, and this may add some additional uncertainty in the final model prediction uncertainty (Sellami et al., 2013).

The Dynamic Identifiability Analysis (DYNIA) (Wagener et al., 2003) approach was developed in order to improve the objectivity, applicability, and robustness of the approaches to hydrograph disaggregation for improved model identifiability. The method is based on locating periods of high identifiability for individual parameters and to detect failures of model structures in an objective manner. The approach uses a Monte Carlo sampling approach to assess the character of the model population in a feasible parameter space. The sampling strategy is based on an

uniform prior distribution, as in the case of the GLUE methodology, and a measure of performance is used to condition the parameter population resulting from uniform random sampling. The main strengths of DYNIA (Wagener et al., 2003; 2004) include flexibility in choice of model performance criteria and it is robust based on its ability to analyze parameter variation in time and to identify and separate periods of noise from information. The approach is also not model specific and allows for any model structure to be evaluated. Some of the weaknesses of the framework are subjectivity in the determination of feasible parameter ranges, it is computationally intensive as a large number of models are required to produce adequate representation of the shape of response surface and it does not account for parameter interactions due to the simple sampling strategy.

The Sequential Uncertainty Fitting (SUFI 2) (Abbaspour 2004; 2007) is a procedure for combined parameter estimation and uncertainty analysis. The approach has been implemented in the SWAT model (Schuol and Abbaspour 2006; Yang et al., 2008; Andersson et al., 2009 ;Sellami et al.,2013). Based on an inverse optimization approach the method uses the Latin hypercube sampling procedure along with a global search algorithm to examine the behaviour of objective functions. Parameterization is applied to a parameter set rather than to individual parameters (Abbaspour et al., 2004). The initial parameter ranges are updated at each iteration, and a narrower parameter uncertainty is obtained (Abbaspour et al., 2004). The main strengths of SUFI 2 includes, efficiency in providing behavioural simulations that satisfactorily match the observation data with less computational time, and it can be implemented with different performance measures (Schuol and Abbaspour 2006; Yang et al.,2008; Andersson et al. 2009; Sellami et al.,2013). The main limitations of this method are that the technique requires interactive checking of the final posterior parameter range by the modeller and can be subjective and also that the method has been tailored for use with only the SWAT model.

The shuffled complex evolution Metropolis (SCEM-UA) algorithm (Vrugt et al., 2003a) is an adaptive sampler that belongs to the family of Markov Chain Monte Carlo (MCMC) methods. MCMC techniques are stochastic simulation algorithms that successively visit solutions in the parameter space with a frequency proportional to their weight in the posterior distributions.

MCMC techniques have been successfully applied to hydrological models and have demonstrated to be superior to other methods in model calibration (Kuczera and Parent, 1998). The implementation of MCMC for Bayesian inference involves creation of a random walk (called Markov process) and then run the process long enough so that the resulting sample closely approximates the original population from which the sample was taken (Blasone et al.,2008). The samples are then used directly for parameter inference and prediction. This method is efficient in generating explicit estimates of parameter uncertainty and prediction of uncertainty bounds in model outputs. It is also comprehensive in exploring a feasible parameter range and produces estimates of parameter sensitivity over that range. However, the method is computationally demanding, and it ignores input and model structure uncertainty.

Sensitivity analysis evaluates the impact of changes in the model parameters or inputs on the model output of interest. Sensitivity analysis can be local or global, and the latter attempts to explore the full parameter space within predefined feasible ranges (Tang et al., 2007). Global sensitivity analysis methods include regional sensitivity analysis (Hornberger and Spear, 1981), variance based methods (Saltelli 2000), regression based approaches (Helton and Davis, 2002), and Bayesian sensitivity analysis (Freer et al., 1996; Vrugt and Robinson, 2007). Most sensitivity analyses consider only univariate or bivariate (first-order interactions) effects on model response (Pechlivanidis et al., 2011), which can be a problem especially when many factors are being considered. It is possible to analyse high order interactions, however, additional computational effort and analytical complexity is required (Saltelli, 2000). A powerful extension of sensitivity analysis is to evaluate the dynamic sensitivity of model parameters, for example, by evaluating the sensitivity based on a moving window passed through the observed and simulated time-series. Such a procedure was developed by Wagener et al., (2004) as part of DYNIA. This can be used to identify periods within an observed time-series for which specific parameters are sensitive, and hence the time-varying information content of data, and also can reveal conflicts within a model structure, for example, where parameters tend to have different optimal values at different points in time (Pechlivanidis et al., 2011).

Although different uncertainty analysis frameworks exist, they do not address the critical aspects of uncertainty analysis in an explicit and cohesive way (Liu and Gupta, 2007; Vrugt et al., 2003a; Wagener et al., 2003). A few frameworks have been introduced to deal with input, output and model structure uncertainty. Kavetski et al. (2006) introduced the BAyesian Total Error Analysis (BATEA) framework, which explicitly treats all sources of uncertainties and integrates these models into the posterior inference of model parameters and predictions. An Integrated Bayesian Uncertainty Estimator (IBUNE) framework was introduced by Ajami et al., (2007) combining Bayesian model combination techniques and a probabilistic parameter estimator algorithm to assess the integrated uncertainty propagation within a system.

2.7.3 Uncertainty analysis framework for gauged and ungauged sub-basins in the southern African Region

As established in the previous sections, the PUB initiative aimed at increasing the knowledge and understanding associated with reducing predictive uncertainty in data-sparse regions (Sivapalan et al.2003). Section 2.5 reviewed different approaches for incorporating uncertainty estimation in hydrological modelling that have been proposed. The applicability of the methods presented in section 2.5 is limited in data scarce areas. Based on this review it is clear that it is possible for different approaches/frameworks to be applied in different regions due to data availability, spatial variability of the catchment being studied and human and financial resources. Currently uncertainty approaches that are appropriate to ungauged or data-poor basins are ensemble predictions (McIntyre et al.,2005) and the use of indices of expected basin behaviour to constrain the output of *a priori* ensembles (Yadav et al., 2007). A study by Winsemius et al., (2009) addressed uncertainty in modelling a scarcely gauged basin in the southern African region and the setting of constraints based on soft and hard information where hydrological signatures are derived from historical discharge to be used in constraining model parameters. Kapangaziwiri et al. (2010) proposed an uncertainty framework (Figure 2.1) for hydrological predictions in both gauged and ungauged basins in Southern Africa based on the following;

- The use of local information on physical catchment characteristics to derive *a priori* values and the regionalization of model parameters from gauged catchments.

- ❖ Regionalization of stream flow signatures that were used to constrain acceptable model behaviour.

The general framework is model independent even though the development and testing have focused on the GW_PITMAN model widely used within the southern African region (Hughes, 2013a). The framework consists of a component for directly estimating local parameter priors using physical basin characteristics and their variability. The second part of the framework uses regionalized catchment functional features to constrain ensembles of flow predictions generated by sampling from the parameter priors (Kapangaziwiri et al., 2012). A regional sensitivity analysis component provides a feedback loop from the constraint analysis to the *a priori* parameter estimation method (Kapangaziwiri et al., 2012). The framework has been applied successfully in several sub-basins in Southern Africa (Kapangaziwiri 2010; Kapangaziwiri et al., 2012; Tshimanga, 2013; Hughes et al., 2013b) and the Congo basin (Tshimanga, 2012). So far, the framework has been applied with the GW_PITMAN model, while it is possible to apply the framework with any model structure (Kapangaziwiri et al., 2012).

2.8 Hydrological model use within the southern African region

The Pitman model is one of the first models to be developed and applied within the southern Africa region (Pitman, 1973, 1978). As well as its wide use in South Africa, the model has also been used in Zimbabwe, Namibia, Zambia, Tanzania, and the Democratic Republic of Congo (Mazvimavi, 2003; Mwelwa, 2004; Ndiritu 2009; Hughes et al.,2006; Sawunyama 2009; Kapangaziwiri, 2010; Thsimanga 2012;Tirivarombo, 2013). Hughes and Sami (1994) developed a Variable Time Interval (VT1), semi-distributed, partly physics based model with a facility to estimate some of the parameter values from physiographic data, but data limitations and a difficult calibration process limited the success in the use of this model in the region. Another widely used model in South Africa is the ACRU model, a daily agro-hydrological model named after the Agricultural Catchment Research Unit, which is frequently used in the assessment of hydrological responses to land use modifications and climate change (Schulze, 2000). The ACRU model has been designed around a multi-layer soil moisture accounting scheme and has a large number of parameters that require quantification. It is designed to be used in ungauged

basins on the basis that its parameters are evaluated through default relationships with measurable catchment properties.

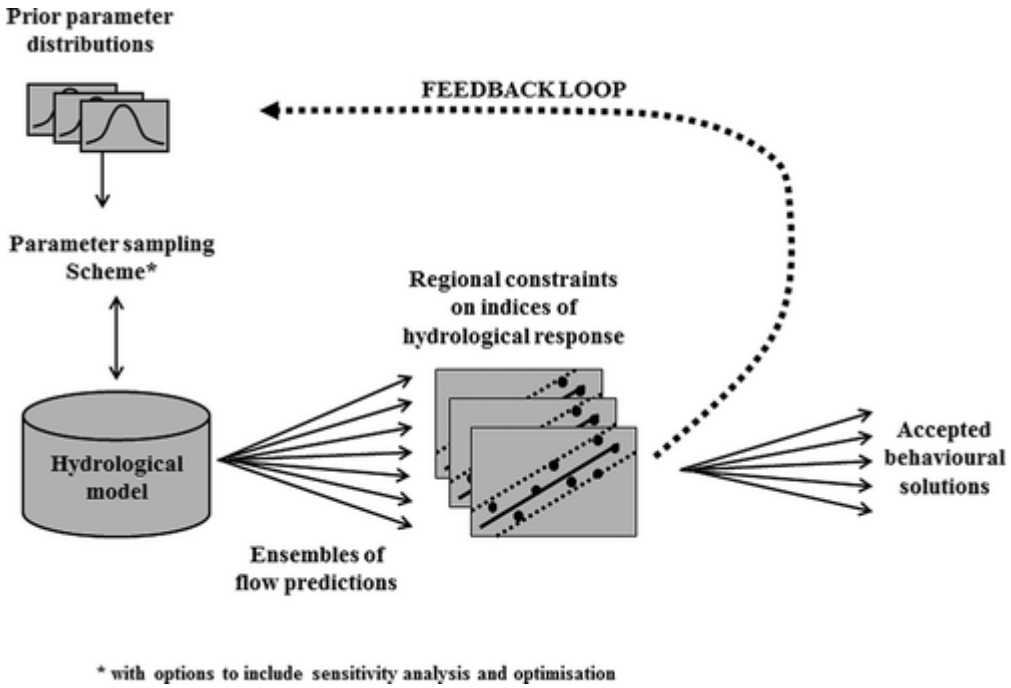


Figure 2.1 The model independent uncertainty application and evaluation framework for the southern African region (Kapangaziwiri et al., 2012)

The Soil and Water Assessment Tool (SWAT) is another model that has been used within the region. SWAT was developed by the US Department of Agriculture (USDA) to predict the impact of land management practices on water, sediment and agricultural chemical yields over extended periods of time in large complex watersheds of varying soil, land use and management conditions. The physically based and semi- distributed SWAT model operates on a daily time step. The SWAT model has been applied in Tanzania especially in the Pangani basin (Ndomba et al., 2008; Ndomba and Birhanu, 2008), Simiyu catchment (Rwetabula et al., 2007; Mulungu and Munishi 2007; Lubini and Adamowski, 2013). In Southern Africa, SWAT has also been used in the Okavango watershed (Milzow et al., 2011) and Zambezi River Basin (Lietchi et al., 2014).

Both SWAT and Pitman fall under the semi distributed model category and both models are capable of simulating the hydrological processes required for this study. The main advantage of semi-distributed models is that their structure is more physically-based than the structure of lumped models, and they are less demanding on input data than fully distributed models. Although the models can be considered applicable within the region, there are limitations associated with the data available, this includes the quantity and quality of data.

2.9 Uncertainty in assessing hydrological impacts of climate change

Uncertainty in the prediction of future hydrology comes from different sources, one of them related to future climates and include the projections of Global Climate Models (GCMs) and downscaling the outputs to scales appropriate for use with hydrological models. Figure 2.5 summarises the five main stages of performing a climate change hydrological impact assessment and the nature of uncertainties to be considered during the analysis. Although GCM's are capable of projecting changes in climatic variables, there are challenges in the predictions of precipitation and cloudiness because they occur at a smaller scale than the GCM's grid size. While high resolution meteorological inputs to fulfill hydrological needs can be generated through downscaling methods, downscaling of climatic variables can also introduce uncertainty in the hydrological model (Gosling et al., 2011).

Our ability to predict future hydrological effects related to change in climate is limited even if we have perfect hydrological models (Beven, 2001; Beven and Westerberg, 2011). Various studies have been undertaken on the use of downscaled outputs for assessing climate change impacts on hydrology (Graham et al.,2007; Horton et al., 2006; Gosling et al., 2011; Tshimanga and Hughes, 2012; Tirivarombo, 2013). Based on these studies and many more, GCM forcing has been identified as the major source of uncertainty, and this has a larger impact on the projected hydrological change than the selected emission scenario or RCM (Regional Climate Model) used for downscaling (Graham et al., 2007). Although GCMs have been reported to be the most important source of uncertainty in hydrologic climate change impact studies, it is important to quantify and estimate the uncertainties associated with the hydrological models themselves.

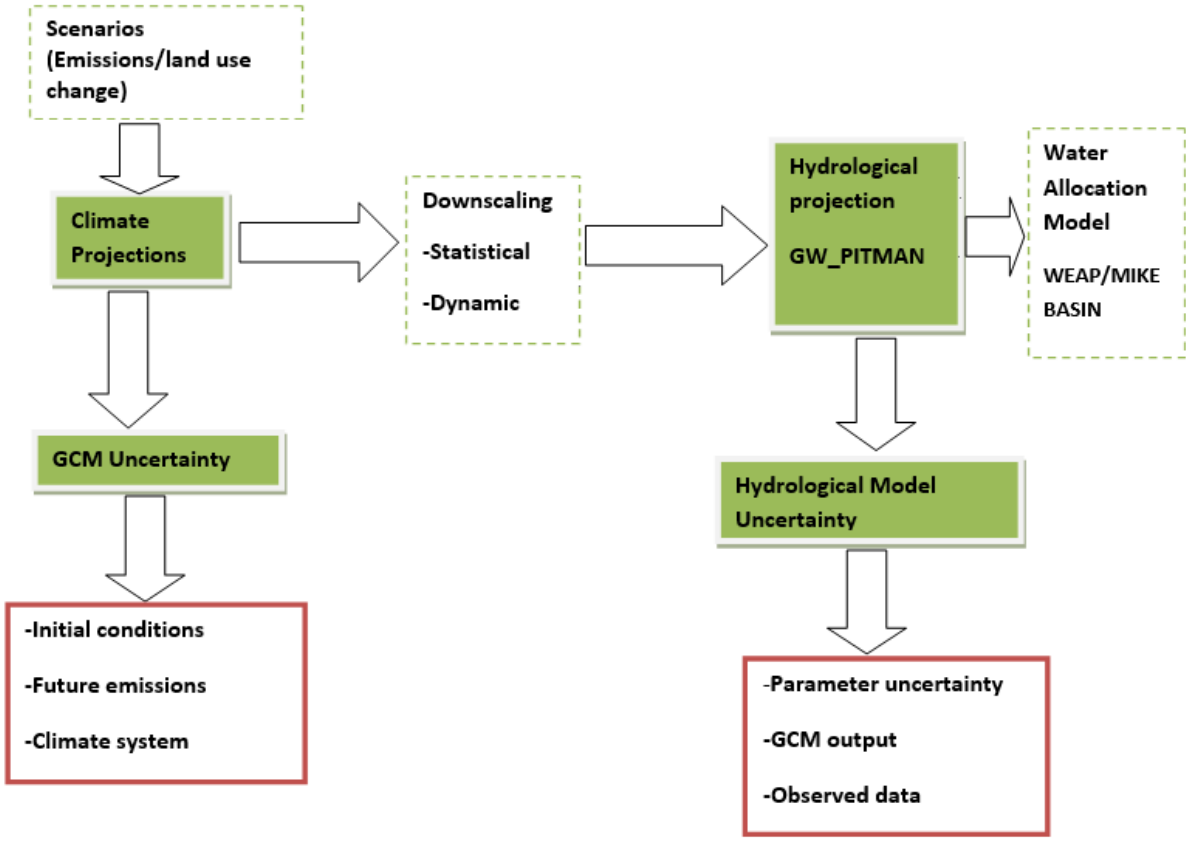


Figure 2.2 The five main steps in climate change hydrological impact assessment and the nature of uncertainties. *Uncertainty from the GCM and the downscaling is propagated to the hydrological model, and the uncertainty in the uncertainty model is propagated to the water allocation model.*

While there have been efforts to improve hydrologic model uncertainty for short to medium range forecasts in terms of hydrological model ensembles, for example, in the HEPEX experiment (Schaake et al.,2006), little quantitative knowledge is available about the role of hydrological model complexity for climate change impact assessments. For example Ludwig et al. (2009) investigated the varieties of different model response as a function of model complexity. Results from this analysis confirm that different degrees of complexity play a considerable role when evaluating respective model results because the physical basis of process descriptions is indispensable to maintain the predictive power of hydrological models (Ludwig et al., 2009). The importance of using an ensemble of hydrological models depends on the hydrological indicator considered and on the area of interest (Velazquez et al., 2013). An

ensemble of hydrological models can be used to evaluate a range of outputs associated with the differences in model process descriptions (Velazquez et al., 2013). However, the differences in physical basin properties can also influence the uncertainty arising from the hydrological model structure (Kay et al., 2007). Therefore, the most suitable level of model complexity and the implications for climate change related water resources assessments remains largely unresolved.

2.10 Concluding remarks

In hydrology, it is not possible to measure everything concerning catchment physical properties because of heterogeneity in basins and limitations of measurement techniques, resources and time. Because of these challenges, hydrological models play a very important role in advancing our understanding of basin hydrological processes. This chapter has introduced the principal concepts, issues and uncertainties associated with hydrological modelling. Based on this review, there is a large range of issues of uncertainty that are possible to assess within hydrological modelling, impact assessment related to development changes (such as water use abstractions) and climate change impact assessment studies.

3 THE GREAT RUAHA RIVER SUB-BASIN: Synthesis of Hydro-meteorological and Spatial Data

3.1 Introduction

The Great Ruaha River (GRR) is an upstream tributary of the Rufiji River Basin and covers an area of 86 000 km^2. The Rufiji Basin is the largest of the nine hydrological basins in Tanzania (Figure 3-1) with a drainage area of about 180 000 km^2, approximately 20% of Tanzania's area, and is populated by approximately six million people (RBWO, 2008; URT, 2012). The basin includes a number of varying biomes, from afromontane forests in the upper catchment areas in the south west of the country, wetlands, dryland and savannah biomes, to mangrove forests at the river delta where the Rufiji River flows into the Indian Ocean, approximately 200 km south of Dar es Salaam. The Rufiji River Basin comprises four sub-basins: 1) the Great Ruaha; 2) Kilombero; 3) Luwegu and; 4) Lower Rufiji. The drainage areas and their contributions to annual runoff are listed in Table 3.1.

Table 3.1 Sub-basins within the Rufiji River basin and their contributions to annual runoff

Sub-basin	Size		Contribution to annual runoff (%)
	Area (km^2)	Proportion of Total (%)	
Great Ruaha	83 979	47	15
Kilombero	39 990	23	62
Luwegu	26 300	15	18
Lower Rufiji	27 160	15	5
Total	177 429		

The Chunya, Poroto and Kipengere mountain ranges form the headwaters of the GRR (Figure 3.2). Several rivers rise from the mountains and drain toward the Usangu Plains (Figure 3.2). The basin is drained by the Great Ruaha River and its five main tributaries namely: 1) Great Ruaha; 2) Little Ruaha; 3) Kisigo; 4) Ndembera and; 5) Lukosi. Most of the basin lies

within the Iringa and Mbeya regions, while smaller parts of the northern portion of the basin lies within the Dodoma and Singida regions of central Tanzania.

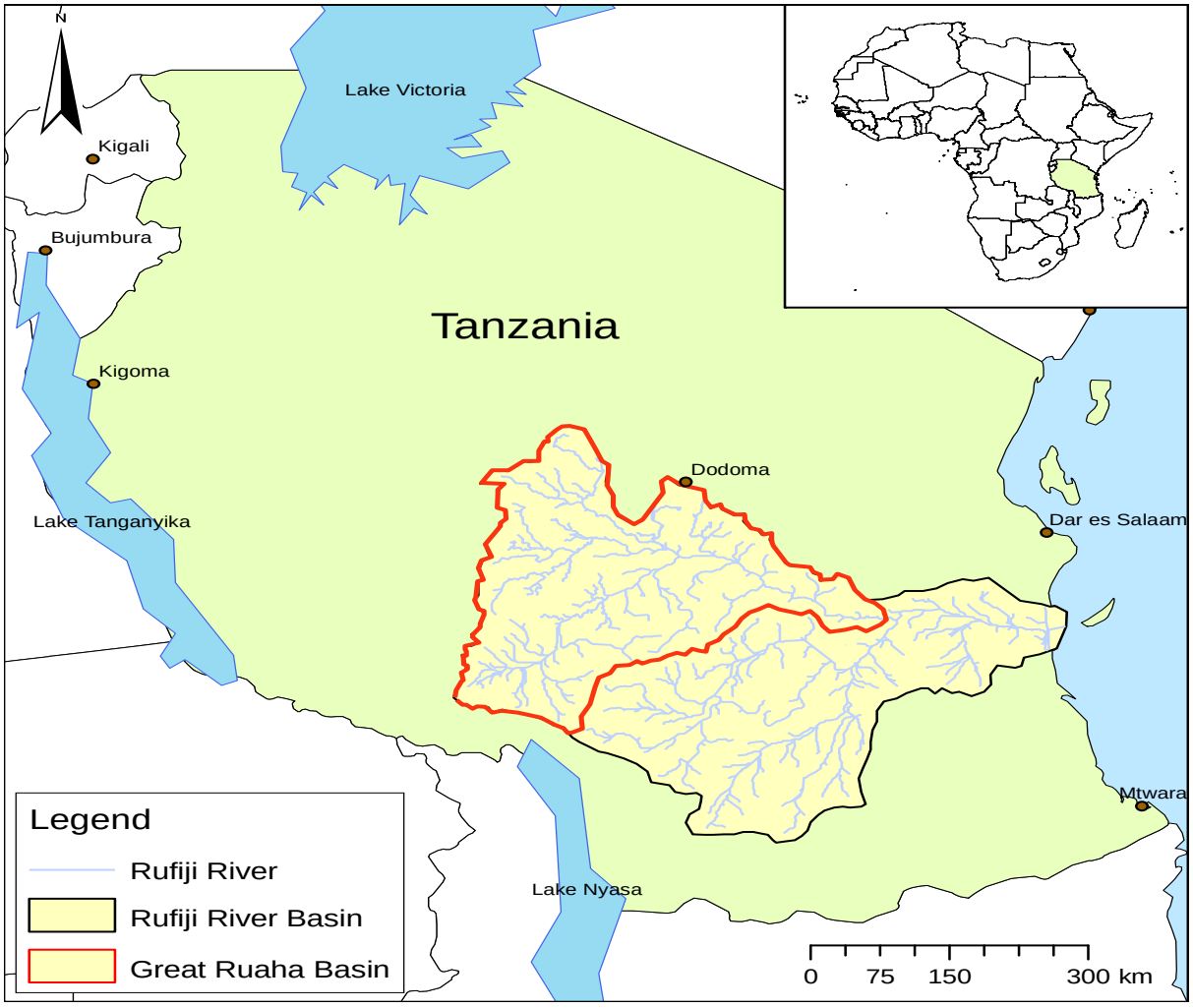


Figure 3.1 The Rufiji River Basin in Tanzania

The upstream part of the GRR comprises several tributaries which join at the wetlands of the Usangu Plains. The major rivers within the upper GRR are perennial, although flows can become minimal at the end of the dry season. The GRR discharges from the northern end of the Usangu Wetlands/Plains at NG'iriama, an outlet of the permanent Ihefu Swamp (Figure 3.3b). The catchment area at this point is approximately 21 500 km^2, and is commonly termed the upper Great Ruaha River catchment (UGRRC). From Usangu, the river flows through the Ruaha National Park (RNP) which starts 30 km downstream of NG'iriama. Downstream of the RNP, the Great Ruaha joins the Little Ruaha and Kisigo rivers to supply

water to the Mtera Reservoir and hydropower plant. The GRR provides about 56% of the inflow to Mtera Reservoir. After Mtera Reservoir, the river flows to the Kidatu hydro-electric plant (HEP). Downstream of Kidatu, the Kilombero and Luwegu rivers join the GRR to form the Rufiji River. The Rufiji eventually drains into the Indian Ocean through the Rufiji Delta.

3.2 The Usangu plains

The Usangu Plains have an area of approximately 5 800 km^2 and previously formed an inland lake (Lake Buhoro) that has been filled by erosional sediments transported from the highlands. The plains consist of a series of alluvial fans enclosing a floodplain wetland (area approximately 2 000 km^2) which, in turn, contains a small perennial swamp locally known as Ihefu (area approximately 80 km^2). The floodplain wetland shows a distinct 'figure-of-eight' form, with the western and eastern wetlands joined by a narrow corridor at Nyaluhanga (Figure 3.3). The majority of the rivers rise in the highlands and flow into the Usangu Plains through the western wetland. The main inflow to the eastern wetland is through the constriction at Nyaluhanga. The only other significant river flowing directly into the eastern wetland is the Ndembera River. Flow from the highlands rapidly fills the channel in the western wetland and spills over to inundate the seasonal floodplain. Within the eastern wetland the water again spreads out over the plain and then flows into the permanent swamp (Ihefu). As inflows increase during the wet season, the swamp fills and spills over the eastern wetland and over the rock bar at Ng'iriama, thus providing the steam flows for the GRR downstream. As inflows decline during the dry season, the combined effect of downstream flow and evapotranspiration from the swamp surface cause swamp water levels to drop to an extent that overflow over the rock bar ceases. Losses from evapotranspiration result in further declines in the swamp water level. Flow downstream returns once the swamp refills to above the level of the rock bar.

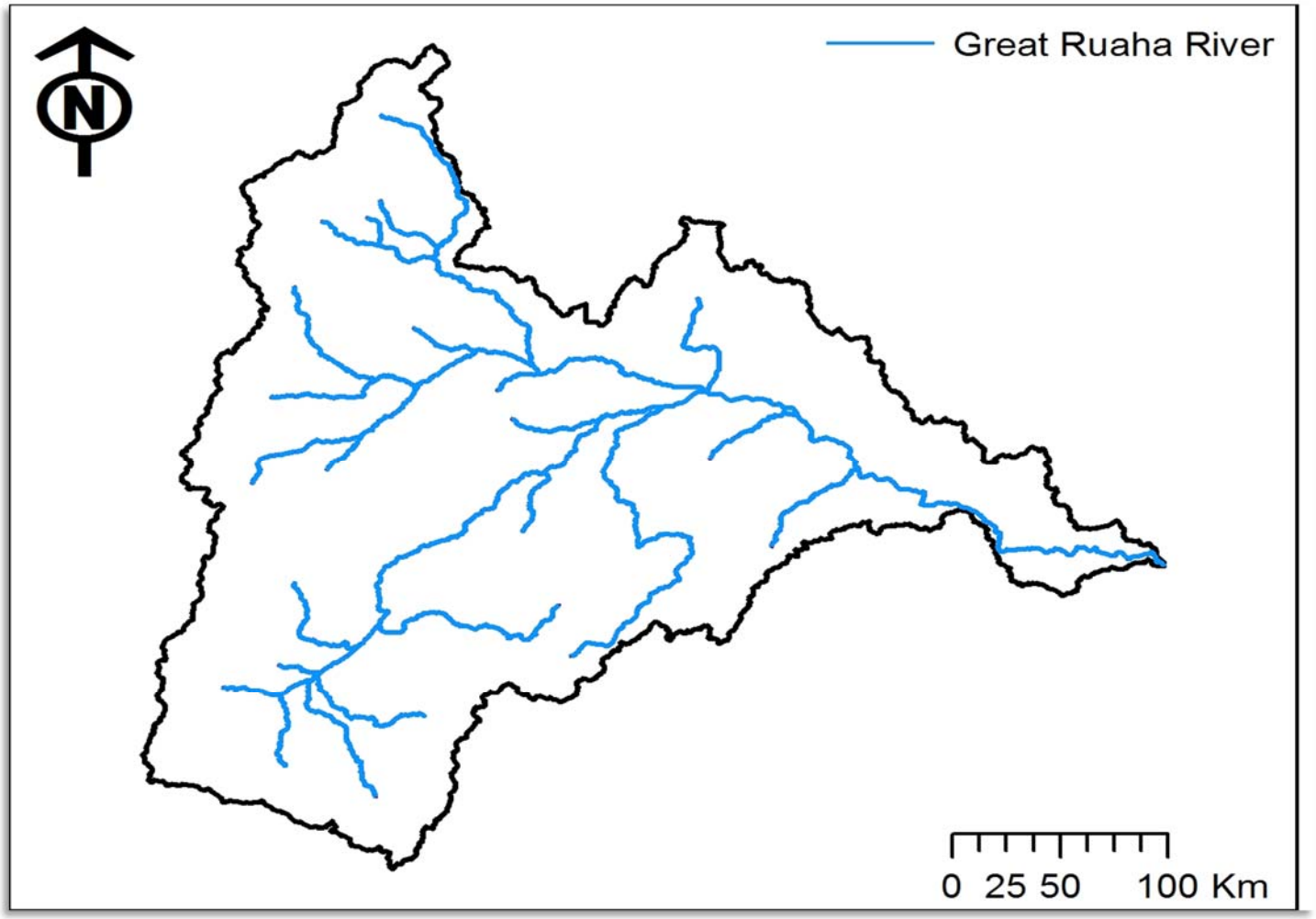


Figure 3.2 Great Ruaha River Sub-basin in Tanzania

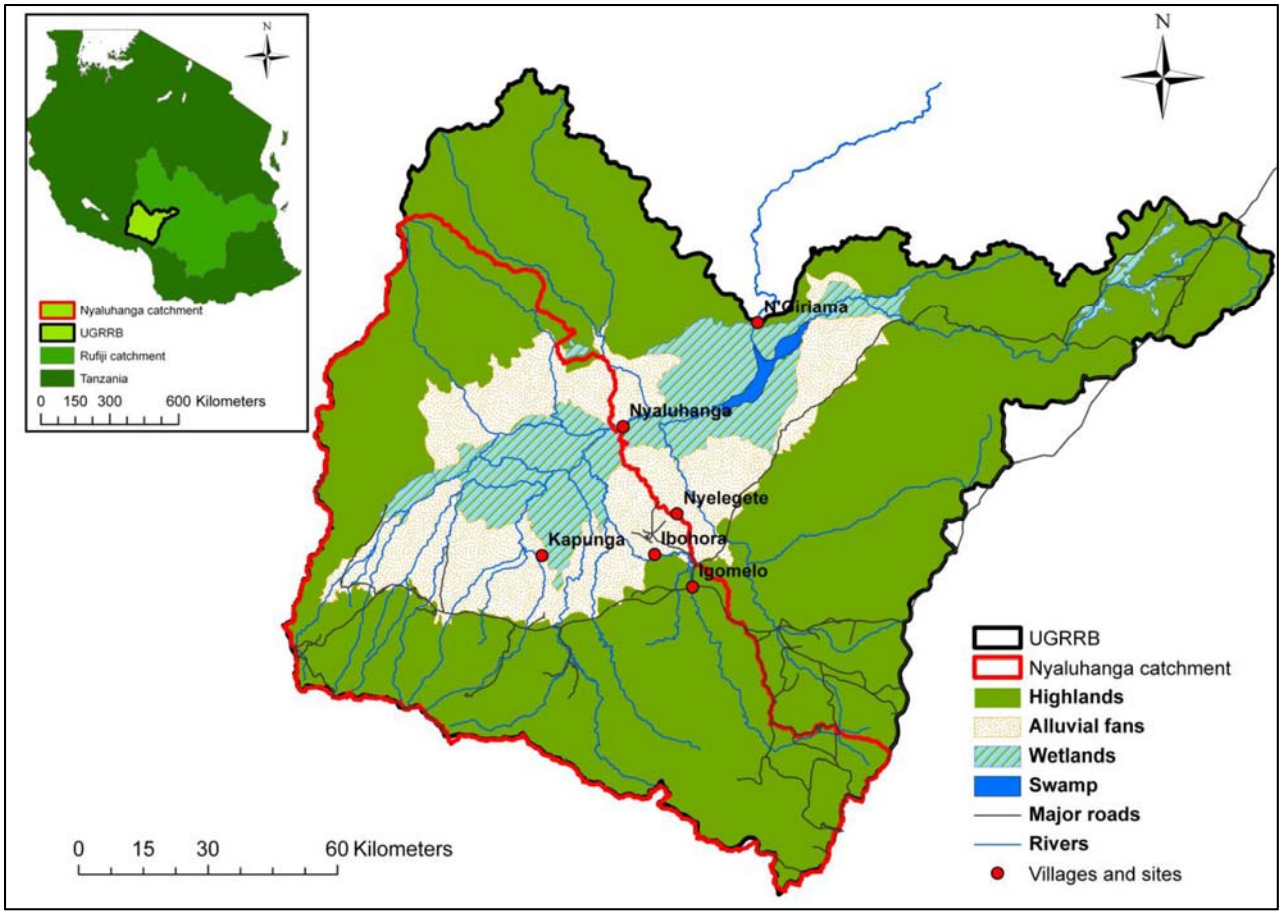


Figure 3.3 The Upper Great Ruaha River sub-basin (UGRRB) and the Usangu wetland (SMUWC, 2001)

3.3 Geomorphology

A landscape form is primarily a result of climatic, geological and geomorphological processes. The topographic characteristics are strongly related to erosion cycles which have prevailed over the continent since the late Jurassic. A cycle of erosion includes several stages of landscape evolution. Most parts of the study area are characterized by extensive elevated plateaus interrupted by isolated highlands and broad linear valleys of tectonic origin (CCKK, 1982). The hydrogeology is closely linked to the geomorphological landforms and therefore it is important to understand the geomorphology in detail. The description of the geomorphological zones (Figure 3.5) is summarised in Table 3.2 and a brief description of each zone is given in subsequent sections.

The African Zone is an extensive part of the basin with relatively flat topography. The main rivers originating from this zone are perennial, with dry-season discharge consisting of base flow derived from springs and seeps. Borehole profiles show that the lithography of this zone can generally be divided into three layers: 1) a sedimentary layer consisting of sand, silt or clay; 2) a weathered granite layer and; 3) bedrock. The infiltration capacity of the upper layer is good, and in many cases contains a secondary water table in the upper clay filled part of the saprolite. This water body is rather stagnant due to the low permeability of the clay but it provides leakage to the lower main aquifer in the weathered granite (SMUWC, 2001).

The Post-African Zone is characterised by a relatively young and unstable land surface where soil erosion is ongoing, flushing clay minerals out of the top soil layer and leaving a sandy surface behind. Profiles of this zone show a layer composition similar to that of the African Zone (SMUWC, 2001b).

The Gondwana Zone: This zone occupies the higher elevated areas of the study area except the Rungwe volcanic zone. The Gondwana erosion cycles are found on the Kipengere ranges and Pororo montains, the Gofio Plateau and the Mufindi highlands. The most common rocks are granite and gneiss, but limestone, shales and schists do occur. The south-western side of the catchment is dominated with granite and gneiss while some catchments on the eastern headwaters consists of crystalline limestone rocks. The Gondwana is characterised by a pronounced relief and thin or non-existent saprolite on the valley slopes, whereas saprolite can attain considerable thickness on the plateau on the south eastern part of the basin. Springs, seeps, bank storage and artesian discharge contribute to the base flow regime of the rivers within this zone (SMUWC, 2001).

The Rungwe Volcanic Zone is characterised by a considerably different geology compared to the other upland zones as it is dominated by a mixture of basalts, ash and pumice. The topography can be described as irregular, with deeply incised valleys formed by the easily eroded volcanic ash. Previous studies have identified a considerable amount of groundwater storage. Consequently, during the dry season, the majority of the rivers are entirely supported

by groundwater fed base flow (SMUWC, 2001b. This zone occupies the south west part of the basin and the headwaters of Little Ruaha and Lukosi Rivers.

Table 3.2 Geomorphological zones of the Great Ruaha River Basin

Zone	Description
Gondwana (2 960 mamsl*)	Erosion surfaces between 1 800–3 000 mamsl, representing the ancient remnants of the Gondwana Pediplain, heavily dissected by an erosion cycle following the break-up of Gondwanaland. Present in the south of the project area. Examples include the Kipengere Range (Gondwana surface), and the dissected landscape of Makete District (post-Gondwana).
Post Gondwana (1 400–2 900 mamsl*)	
African	Erosion surface between 1 200–1 800 mamsl forming a smooth landscape seen over large parts of Africa, and often containing abundant dambos. The African surface can be severely disrupted by the Rift Valley faulting. The plateau around Njombe and Sao Hill is largely an African surface.
Post African	Erosion surface between 1 100–1 200 mamsl. Erosion surface comprising a young pediplain ~100 m lower than the African surface, moderately to heavily dissected, forming more irregular and unstable terrain. This surface is partly covered by more recent deposits.
Scarp	Transition between one erosion surface and the next, but note that most of the scarps in the project area are related to Rift Valley faulting.
Lake deposits	Flat areas with elevation between 1 000–1 100 mamsl. Aggradational (depositional) surface of the lake bed deposits and alluvial fans.
Rungwe volcanics	Aggradational deposits from three main eruption centres around Rungwe Mountains.

* mamsl – metres above mean sea level.

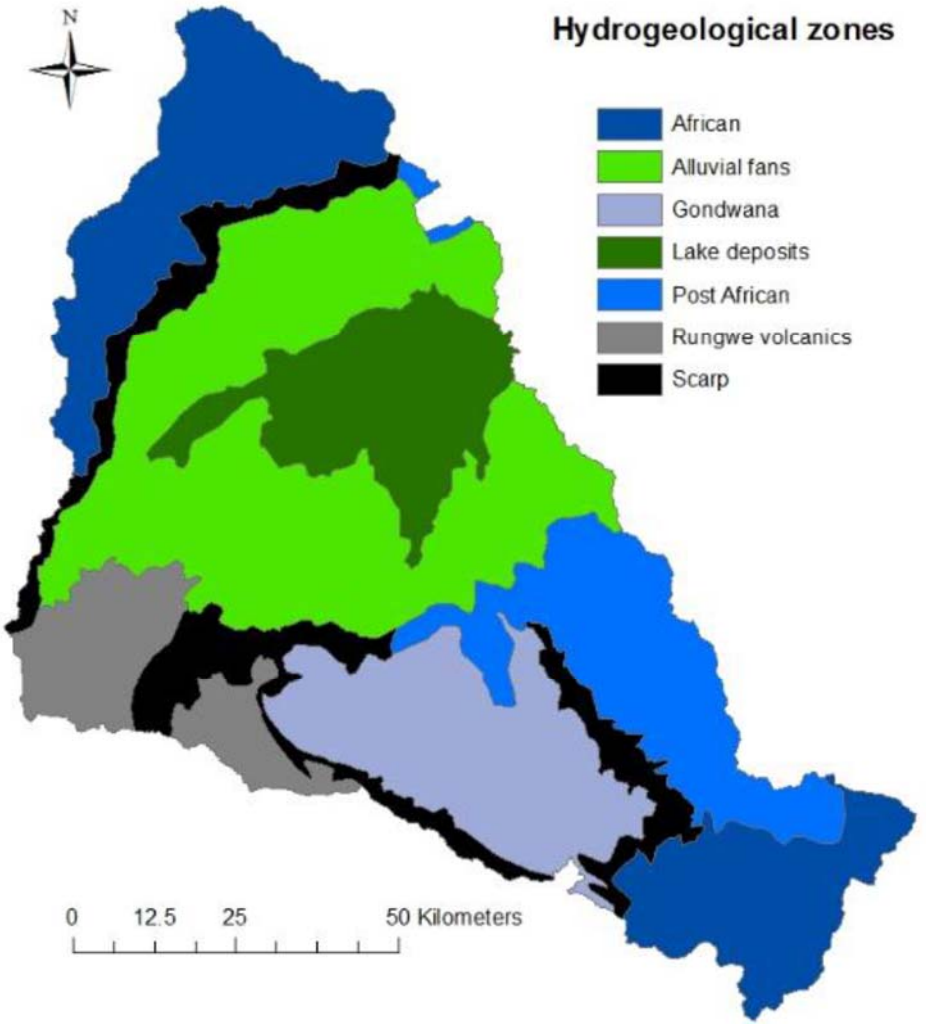


Figure 3.4 Spatial distribution of geomorphological zones within the Upper GRR sub-basin (Modified from SWUWC, 2001)

The Scarp Zone is located mainly along the southern and western sides of the Usangu Plains and refers to the actual scarp faces. The zone is characterised by steep topography, resulting in rapid runoff and drainage of groundwater (SMUWC, 2001).

The Alluvial Fans Zone borders the Scarp and the Post-African zones. The Alluvial Fans Zone is an important zone due to its groundwater potential, as the geological deposits are relatively permeable. The deposits in the plains consist of a mixture of clay, silt, sand and gravel layers, which make the hydrogeology of the plains highly heterogeneous. In general, the lower aquifers can be considered to be confined. The Alluvial Fans Zone is assessed to consist of 200 m of unconsolidated and heterogeneous deposits. The infiltration capacity on the sands is high, and groundwater levels respond rapidly to precipitation (SMUWC, 2001) and groundwater inputs from the surrounding catchments draining the highlands.

The Lake Deposits Zone is located inside the Alluvial Fans Zone and is in general similar in characteristics to the Alluvial Fans Zone. This zone consists of heterogeneous deposits, characterised by apparent buried river channels. The deposits are slightly less permeable than those of the Alluvial Fans Zone due to the spatial distribution of the alluvial fans and deposits. Figure 3.5 presents a hydrogeological conceptual profile of the study area. The figure is based on a conceptual model defined in SMUWC (2001).

3.4 Topography

The catchment is dominated by steep topography at the south-western borders consisting of the Poroto, Kipengere and Chunya mountains with altitudes up to 2 962 mamsl. Most highland parts in the GRR sub-basin are characterised by steep slopes. The mountains surround the nearly flat alluvial fans located at approximately 1 050 to 1 070 mamsl which have a longitudinal slope of 0 to 0.5 % towards the wetlands in the center of the catchment where the Ihefu swamp is located at approximately 1 020 mamsl. The topography and slopes of the area are depicted in Figure 3.6 and Figure 3.7. The two figures are based on a Digital Elevation Model (DEM) derived from Shuttle Radar Topographic Mission (SRTM) 90m digital elevation data (http://srtm.csi.cgiar.org/).

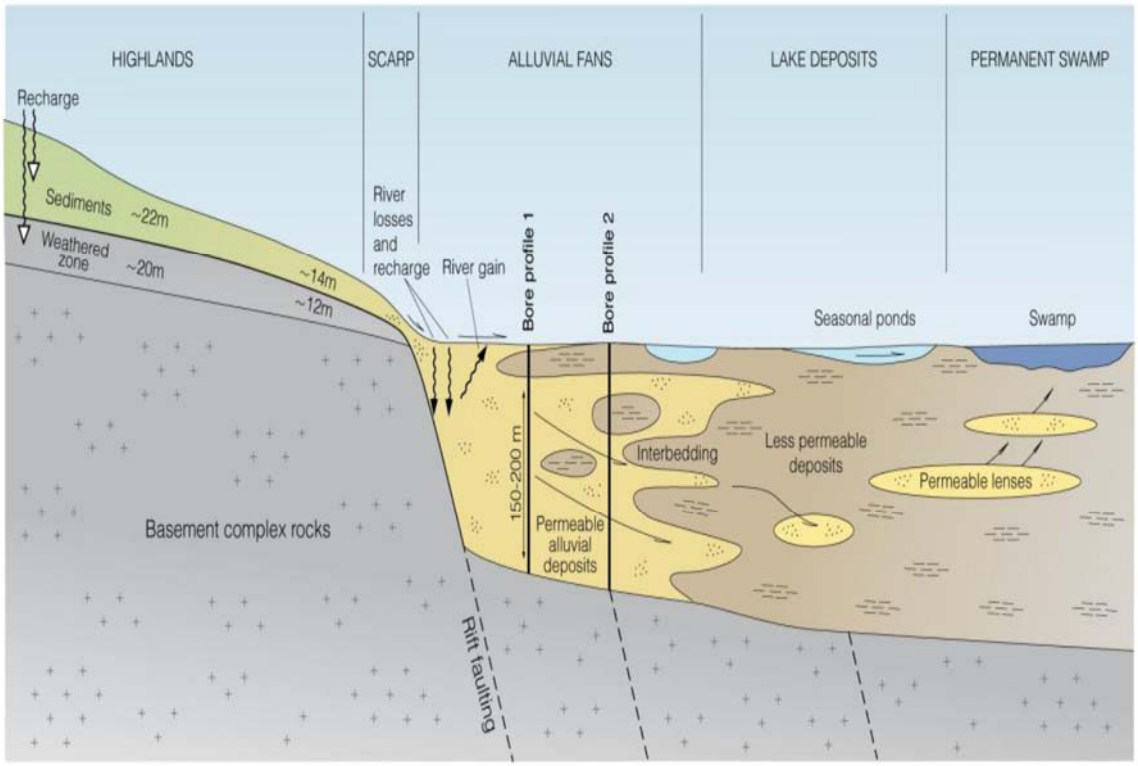


Figure 3.5 A hydrological conceptual model of the upstream part of the Great Ruaha sub-basin (SMUWC, 2001)

3.5 Geology

The major part of the basin is underlain by the basement complex of Precambrian rocks dominated by granite (Figure 3.9). In the south western part of the basin the parent material is volcanic ash deposit originating from the Rungwe-Mbozi volcanic complex. The eastern part of the basin consists of crystalline limestone rocks. Limestone rocks are characterized by caves which have high groundwater recharge potential as surface water drains downward through solution joints. The north western part of the basin is dominated by granite and rocks of the Dodoman series. The Dodoman series consists of sedimentary and highly, metamorphosed and migmatised igneous rocks. These types of rocks are also confined to central Tanzania. The rocks are very dense and not easily weathered and recharge occurs through fracture zones. The Usangu plain is dominated by limestone (Figure 3.9).

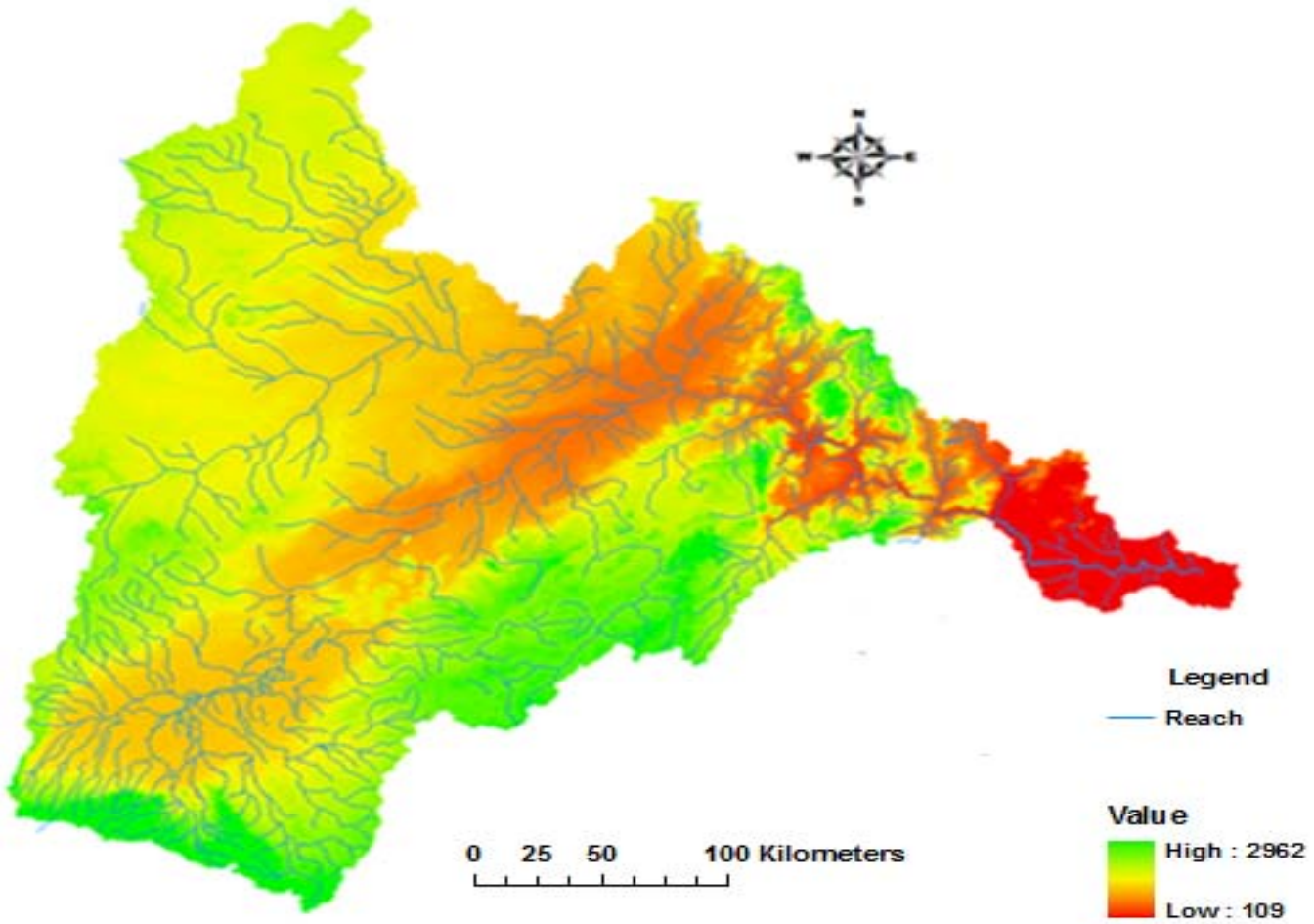


Figure 3.6 Topography of GRR sub-basin based on DEM

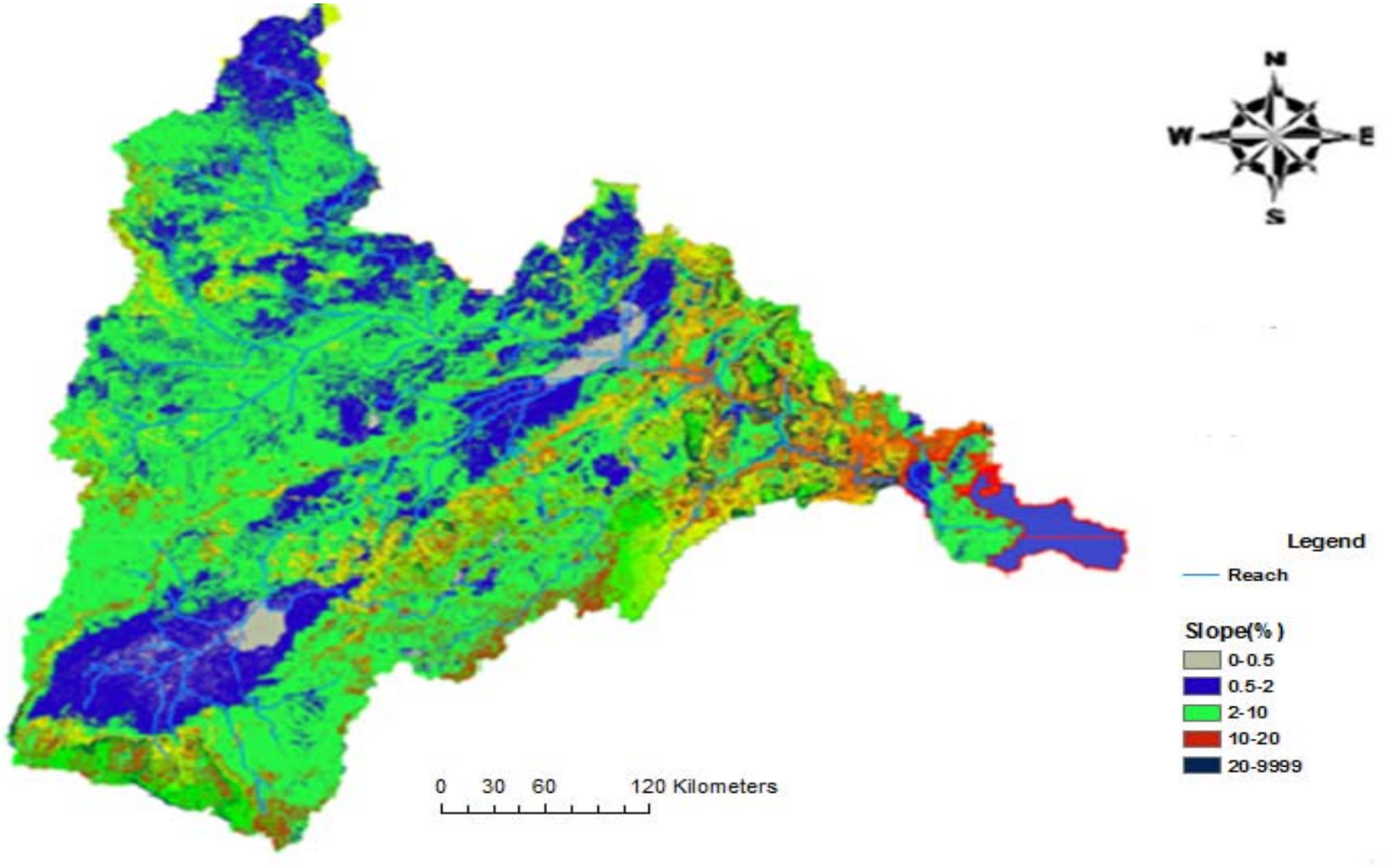


Figure 3.7 Spatial distribution of slope classes (%) in the Great Ruaha Sub-basin

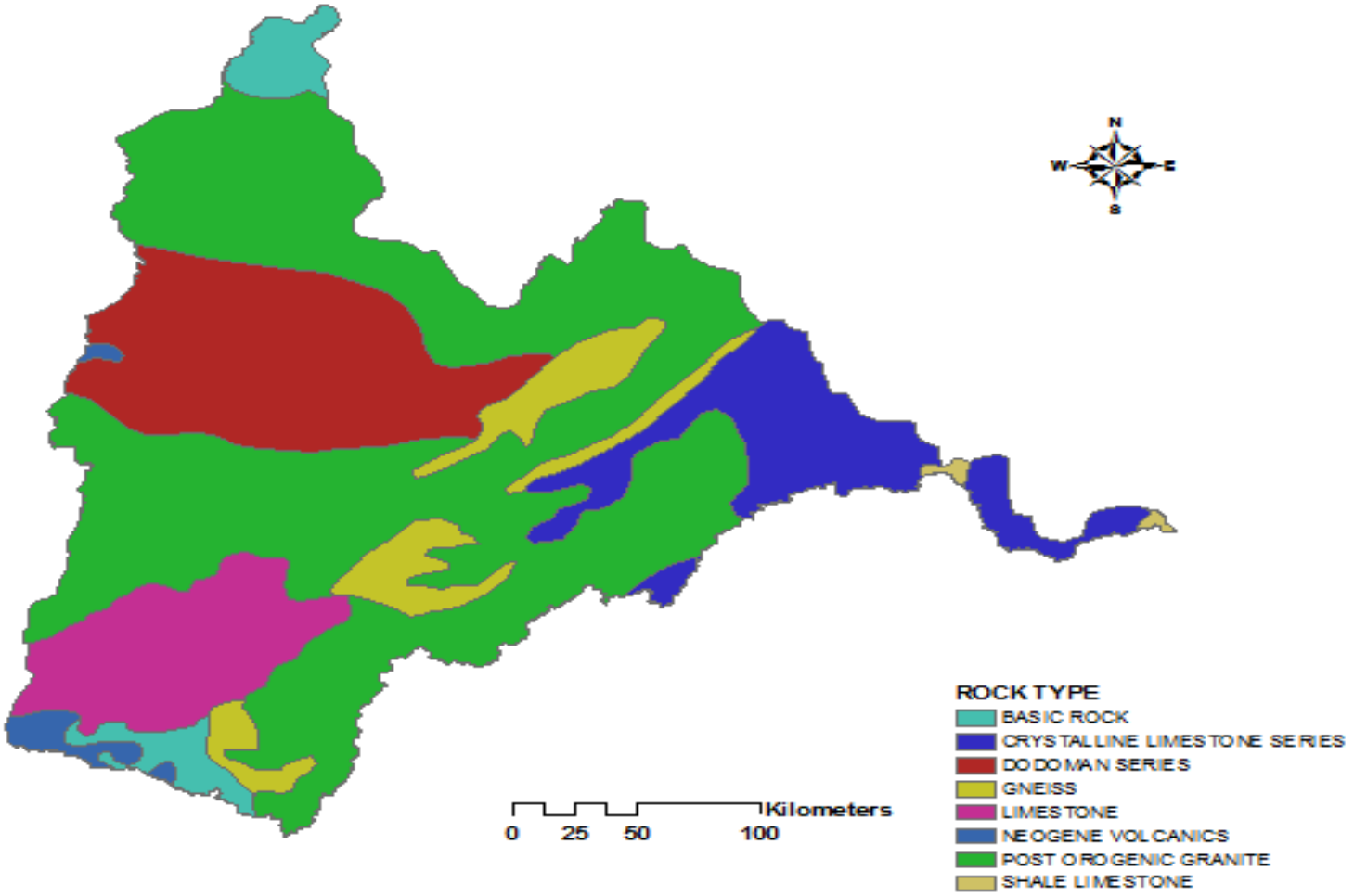


Figure 3.8 Map showing the broad geology of the Great Ruaha River Sub-basin

3.6 Soils

Data from the FAO's (Food and Agriculture Organization) Harmonized World Soil Database were used to describe the soil characteristics of the GRR. The top soil layer within the study area consists of 16 different soil classes (Figure 3.10). Most upland areas are dominated by Umbric Andosols, Umbric Nitisols, Vitric Andosols and Haplic Acrisols. The soil texture classes were derived using ArcGIS 9.3 (ESRI, Inc). Most soils are well drained sands, clays, loams and mixtures of these. The soils of the basin are dominated by loam soils (57.4%), followed by sandy-loam (26.9%). High elevation areas consist of well drained loamy soils. The Rungwe volcanics deposition also forms soils characterized by good drainage in high elevated areas. Lake deposit areas (Usangu Plains) are characterised by imperfect to poor drainage. The wetland area is characterized by clay soils (2.5%), whereas the alluvial fan area is covered by sandy-clay-loams. However, a variety of textural classes are present, depending on the variation in sedimentation conditions prevailing when the deposition took place (CCKK, 1982). Alluvial clay and clay loam soils occupy the greatest part of the existing paddy producing area (the alluvial fans area), and these are areas of high fertility (DANIDA/World Bank, 1995). Sandy-clay-loams cover 12.7% of the sub-basin and dominates the semi-arid part of the GRR, mostly the Kisigo Catchment.

3.7 Land cover

The land cover map (LCM) using 1 km raster data were obtained from the Global Land Cover Facility (Hansen et al., 2000). The land use classes for the GRR Sub-basin are mapped in Figure 3.11. The area and percentage basin cover for the classes are also given in Table 3.3. The predominant cover class is dry land cropland and pasture occupying 36.7% of the catchment (Table 3.3). Dry land cropland and pasture (CRDY) is especially prevalent in the north-western part of the basin, while the cropland-woodland mosaic (CRWO - 20.4%) dominates the UGRR. Savannah (SAVA - 21.5%) dominates the middle and north-western parts of the basin which is predominantly semi-arid.

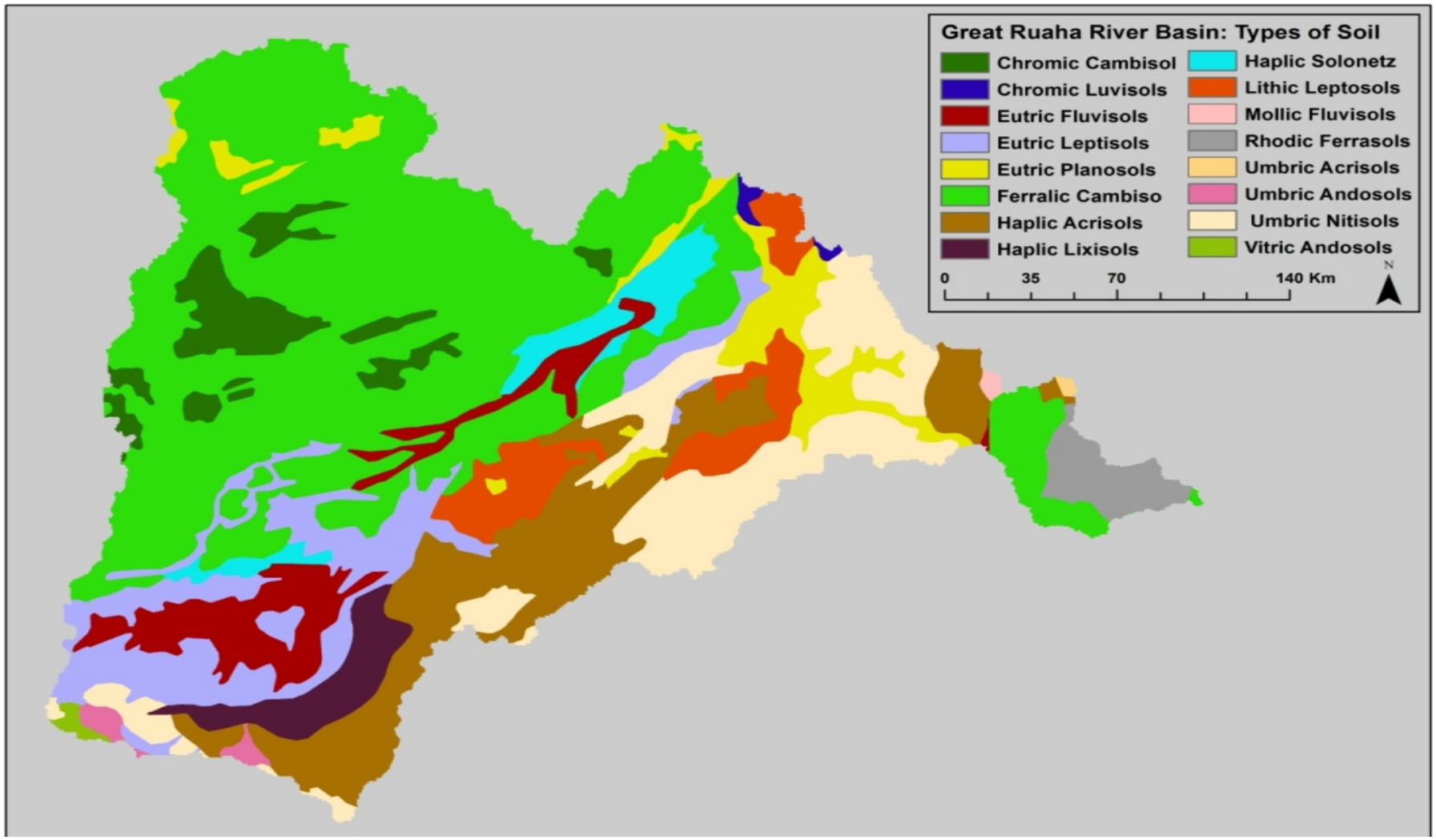


Figure 3.9 Top soil distribution within the GRR based on the FAO's Harmonised World Soil Database

6.1% of the area is designated as grassland (GRAS) which occupies the low lying areas of the GRR and the Usangu Plains. Another 6.1% is designated deciduous broadleaf forest (FODB) (Table 3.3). Mixed forest (FOMI) occupies 4.2% of the GRR.

Table 3.3 Land cover classes within the Great Ruaha Sub-basin

Land Cover Class	Key	Area(Km^2)	% total basin Area
Dryland cropland and pasture	CRDY	30 798.8	36.8
Cropland/woodland mosaic	CRWO	17 130.9	20.4
Grassland	GRAS	5 109.0	6.1
Shrubland	SHRB	3 069.1	3.7
Savanah	SAVA	18 019.4	21.5
Deciduous broadleaf forest	FODB	5 111.2	6.1
Evergreen broadleaf forest	FOEB	722.4	0.9
Mixed forest	FOMI	3 508.0	4.2
Water bodies	WATB	304.0	0.3
Barren or sparsely vegetated	BSVG	6.7	<0.1

The land use within the GRR is highly heterogeneous. The spatial distribution of the main cultivation areas is to a large extent determined by soil fertility and water availability. The agricultural areas in the Mbeya and Iringa regions coincide with areas of high rainfall. Hence, in Iringa, the cultivated area occurs along the African Plateau while in the Mbeya region, cultivation is concentrated along the south-western part of the GRR and the Usangu Plains. Originally, much of the area had been dominated by Miombo Forest, but due to increased population and migration which has led to increased demand for food supply, deforestation of natural vegetation for agricultural purposes has occurred. The GRR basin important for both irrigated and rain fed agriculture, livestock keeping forestry, fishing, and tourism activities in the Ruaha National Park (RNP). There are several plantations in the basin. Some of the major ones are the Sao Hill Forestry Plantations (about 40,000 ha) and Forest Escarpment in Iringa region. Other plantations are for tea and coffee in Mufindi and Njombe districts and paddy in the Usangu Plains.

3.8 Climate

The climate of the study area is influenced by its proximity to the Equator, the Indian Ocean and topography. Located between latitude 7° and 12°S, the climate is tropical, with high temperatures, low wind speeds, high humidity and no real winter season in low lying areas. The warm Indian Ocean effects large seasonal changes to the general circulation of the atmosphere of Tanzania's southern highlands, thus creating considerable seasonality in rainfall, cloudiness and surface wind conditions. Seasonal rainfall in Tanzania is driven by the movement of the Inter-Tropical Convergence Zone (ITCZ). The ITCZ is a narrow belt of very low pressure and heavy precipitation that forms near the Earth's Equator. The position of the ITCZ changes throughout the year, moving southwards through Tanzania during October–December, reaching the southern part of the country during January–February, and returning northwards during March–May. From December–February, the area is situated between a relatively high pressure over northern Africa and the Arabian Peninsula, and a large low pressure at about 10–15°S. Air masses moving from high to low pressure areas during this period give rise to the rather dry north-east monsoon (Kaskazi), which despite its relative dryness, does produce considerable rainfall in the regions. From about March, the ITCZ moves northward towards the Equator, placing the regions under the influence of the convergence between air masses from the Southern and Northern hemispheres. This situation dominates the climate through May and results in the heaviest rains of the year.

One of the reasons for this is the encounter between the north-east monsoon and air masses from the south-east at the Inter-Tropical Convergence Zone. The effect of which frequently extends from the north into southern Tanzania (Figure 3.13). This causes the southern, western and central parts of Tanzania to experience one wet season which starts in October and extends through to April or May. From June to September, the synoptic situation shows relatively little variation. During this period, the study area is under the influence of the south-east monsoon (Kusi) which carries air from a large high pressure area over South Africa and adjacent parts of the Indian Ocean to a very strong low pressure over Saudi Arabia. Originating largely from the South African winter, this monsoon is rather dry and cold, and the region experiences a pronounced dry season during this period.

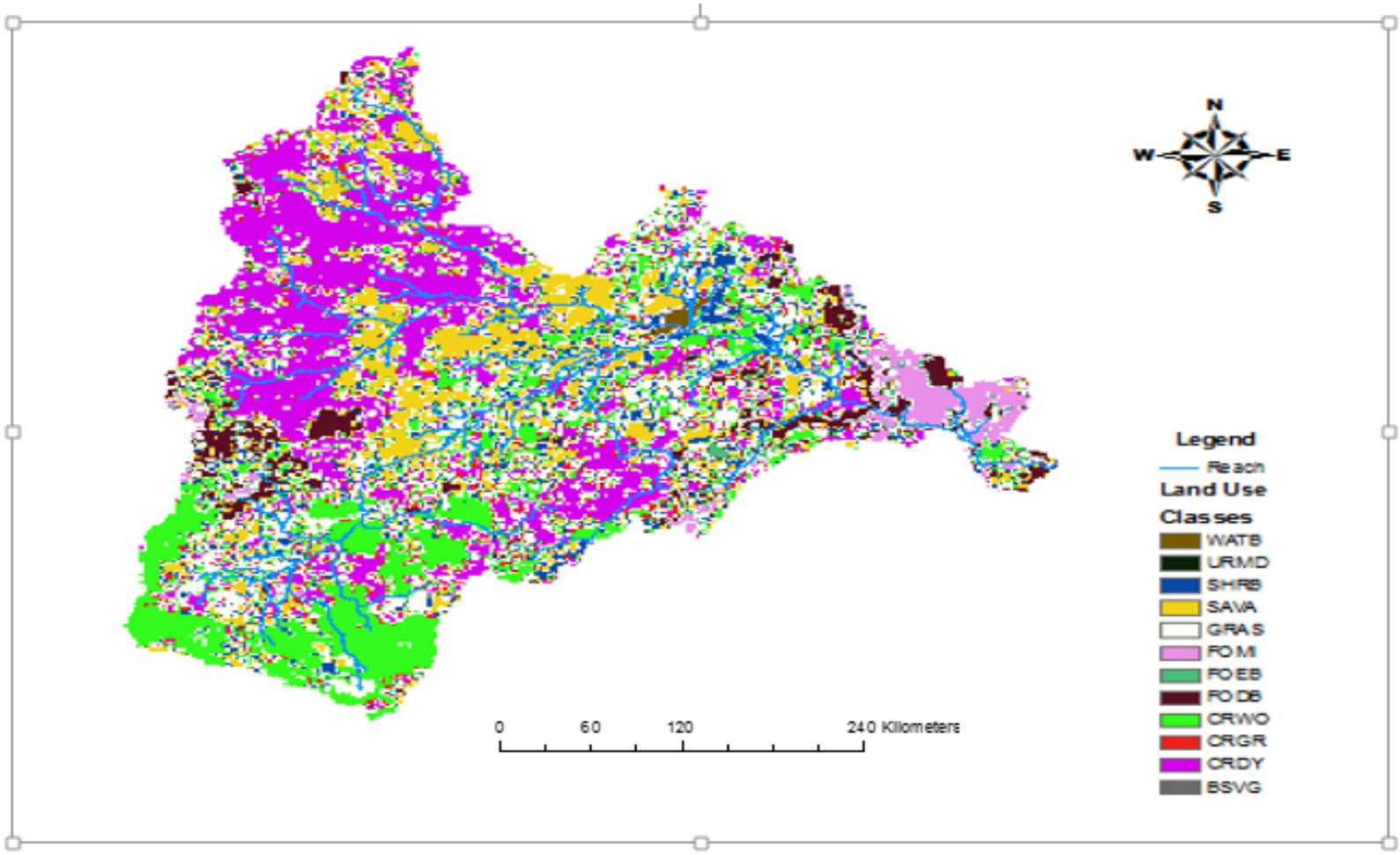


Figure 3.10 Spatial pattern of land cover in the Great Ruaha Sub-basin

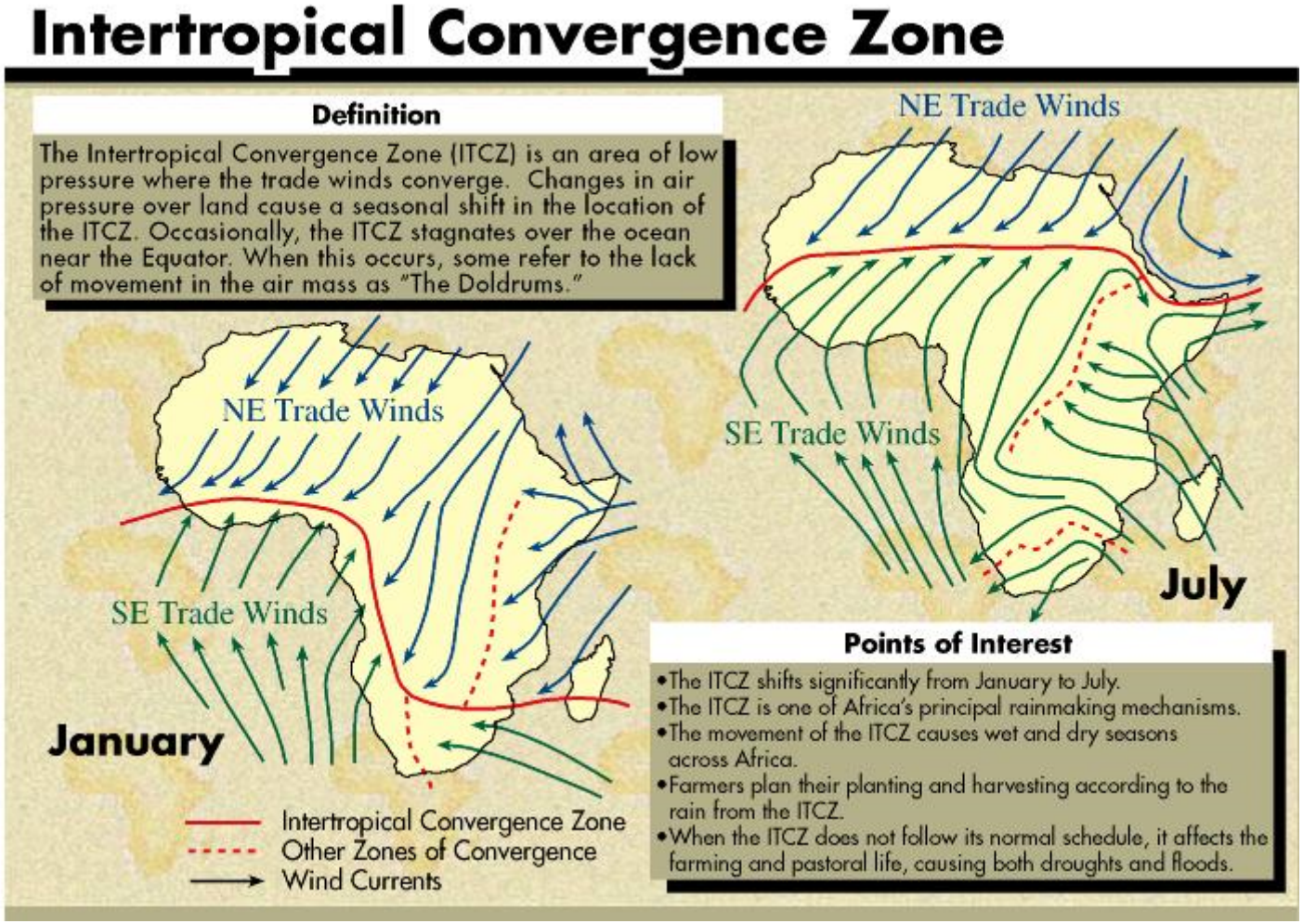


Figure 3.11 ITCZ Seasonal shifts over Africa (http://people.eku.edu/davisb/Africa/ITCZ.jpg)

The main convergence returns to Tanzania during October, reaching the project area during November and signaling the onset of the rainy season. The movement of the ITCZ is sensitive to the variations in the Indian Ocean sea surface temperatures and varies from year to year, hence the onset, duration and intensity of rainfall varies inter-annually.

3.9 Precipitation

Precipitation data were acquired from the Tanzania Meteorological Agency (TMA) and the Rufiji Basin Water Office (RBWO). These data have been used to establish the hydrological characterisation of the catchment and rainfall data interpolation. The identification of rainfall stations within and around the study area involved delineation of the catchment boundary using a digital elevation model (DEM) and the selection of the stations that fall within and around the catchment. The List of all stations is provided in Appendix I and their spatial distribution is shown in Figure 3.12.

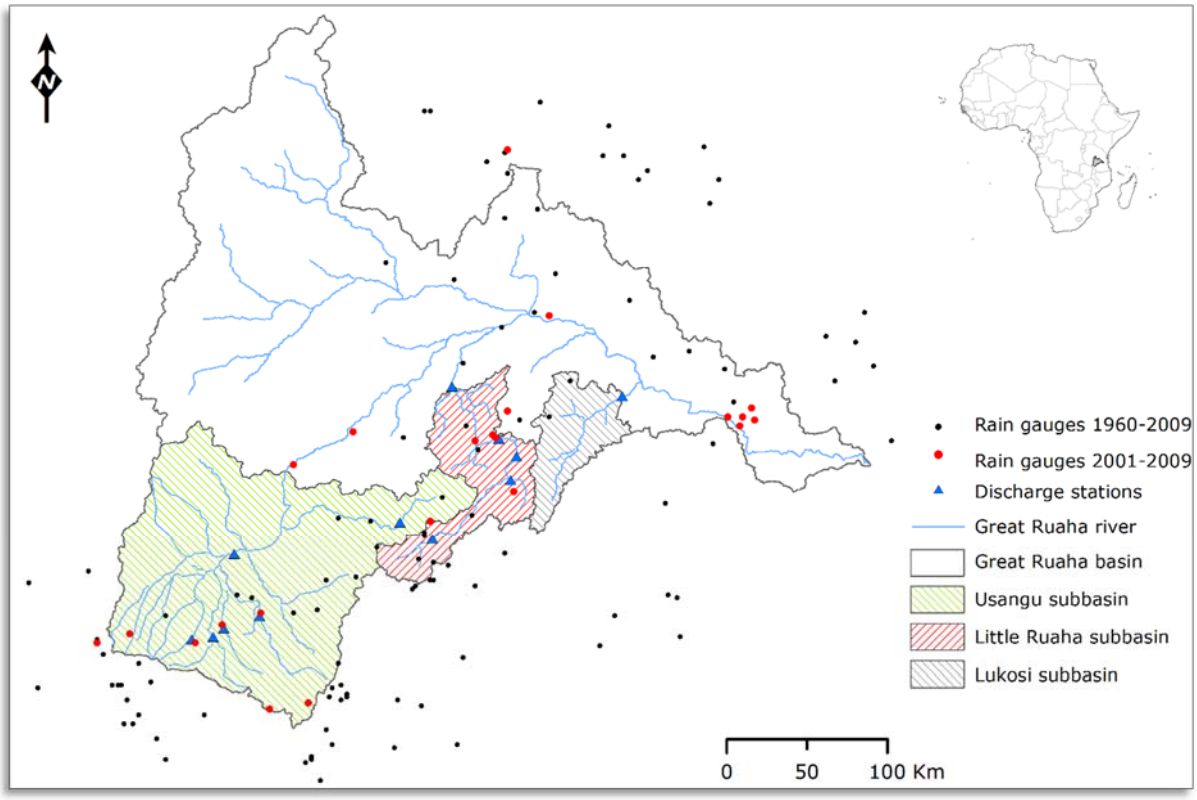


Figure 3.12 Spatial distribution of rain gauges within and around GRR basin.

The record length of the identified 175 rain gauge stations are very uneven. The region had many operational stations during the 1960s–1980s, followed by a rapid decline in the 1980s–1990s. In addition, the spatial distribution of stations within the basin is not uniform. Most stations are located in the upper part of the basin, and less in the north-western part. Few stations are located within the plains. Although lower numbers of rain gauges are operational within the plains, the spatial and temporal variation of rainfall within the plains is much smaller compared to that of the upper basin (SMUWC, 2001). A complete absence of rain gauges over a large area of the basin in the west and the north-west has serious implications for hydrological research and water management. This deficiency in the availability of data is likely to give rise to serious discrepancies in the estimation of runoff originating from this part of the basin, and presents a challenge to modelling the hydrology of the GRR sub-basin. Data scarcity and a decline of hydro-meteorological networks may lead to uncertain predictions (Hughes et al., 2008, Westerberg et al., 2011). This may also lead to the introduction of errors when interpolation methods are applied across space and time based on data from only a few available observational stations or periods (Westerberg et al., 2011; Jung et al., 2012).

3.9.1 Seasonal Variations

Seasonal variations indicate the dominance of a uni-modal rainfall regime, relating to a single rainy season between November–May, with a general rainfall peak occurring between January–April depending on region (Figure 3.13). For the high rainfall areas, the peak occurs around March–April, with the rainfall peak within low rainfall areas occurring around January–February (Figure 3.13). Mean monthly rainfall varies from less than 100 mm per month in the low rainfall areas to more than 300 mm in high rainfall areas (Figure 3.13). The seasonal variations further indicate the relatively dry period from June to September, with monthly rainfall amounts predominantly below 50 mm and August being the driest month. No rainfall occurs on the plains during the dry season from May to November, however, rainfall does occur on the southern side of the Kipengere Range because of orographic lifting of air moving north from Lake Nyasa. Sometimes this rain crosses the catchment, providing small amounts of rain during the dry season in the southern highlands. This accounts for the dry season cropping at higher elevations in the southern highlands.

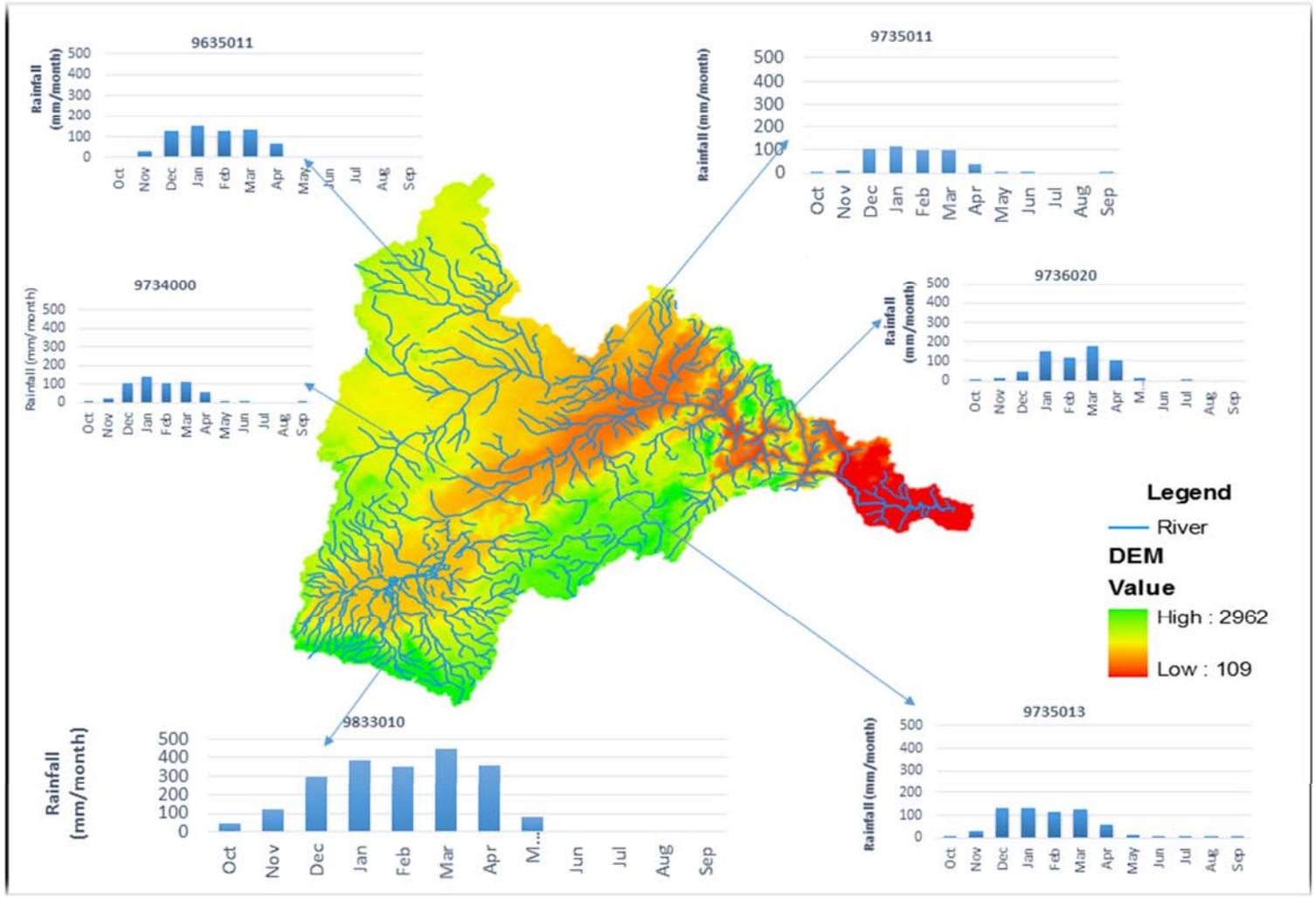


Figure 3.13 Spatial variability of long-term mean monthly precipitation in the GRR

3.9.2 Inter-annual variations

Rainfall in the area is subject not only to high spatial variability due to the characteristic convectional pattern, but also to considerable temporal variation from year to year. The spatial variation of mean annual precipitation (MAP) is shown in Figure 3.14 and 3.15. The highland areas, Kipengere Range and Poroto Mountains located in the southern and eastern part of the basin receive high rainfall of ~1 400–1 800 mm per year, which could be even higher at the summits of these mountains. Annual rainfall decreases towards the north-western part of the basin, where annual rainfall is only ~500 mm per year (Figure 3.14).

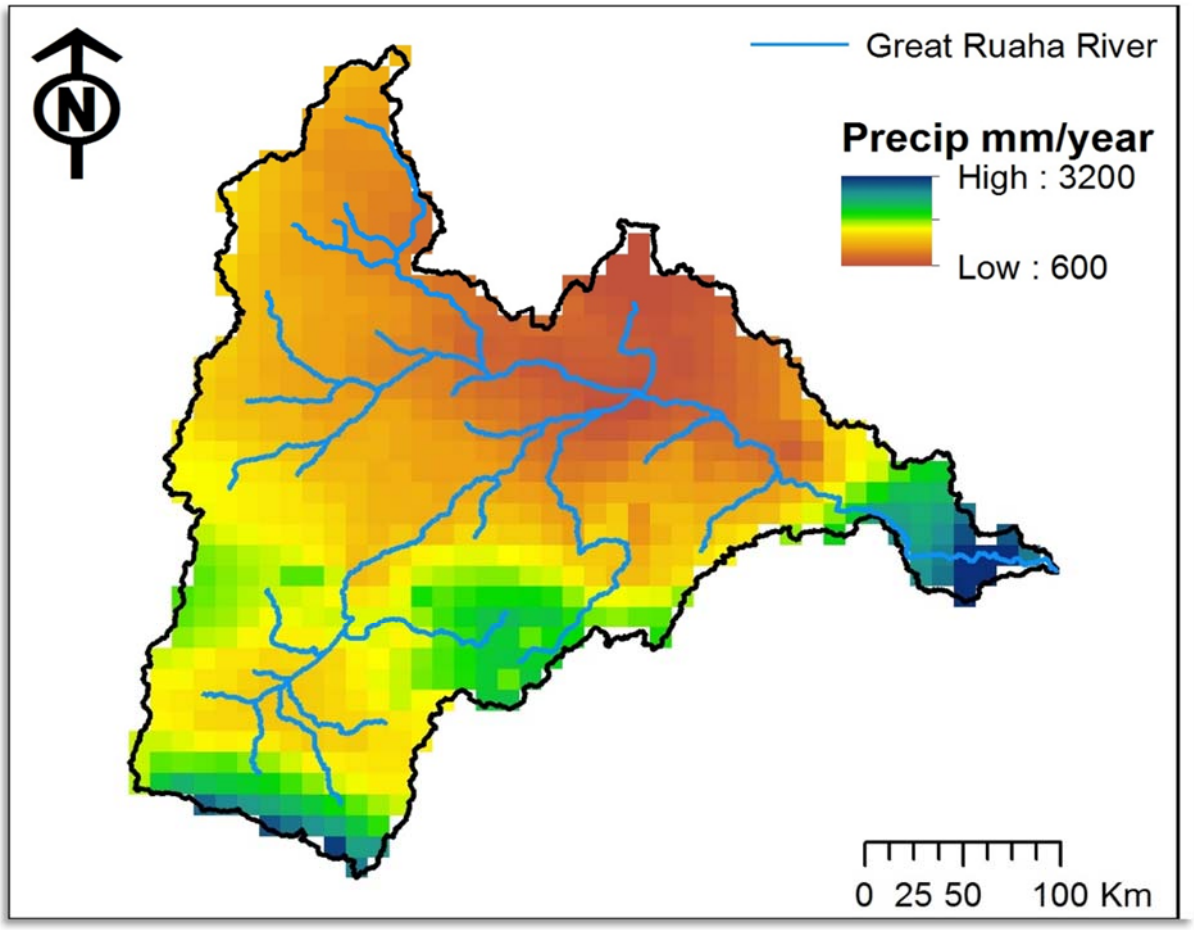


Figure 3.14 Spatial distribution of annual precipitation (mm/year) within the GRR Sub-basin

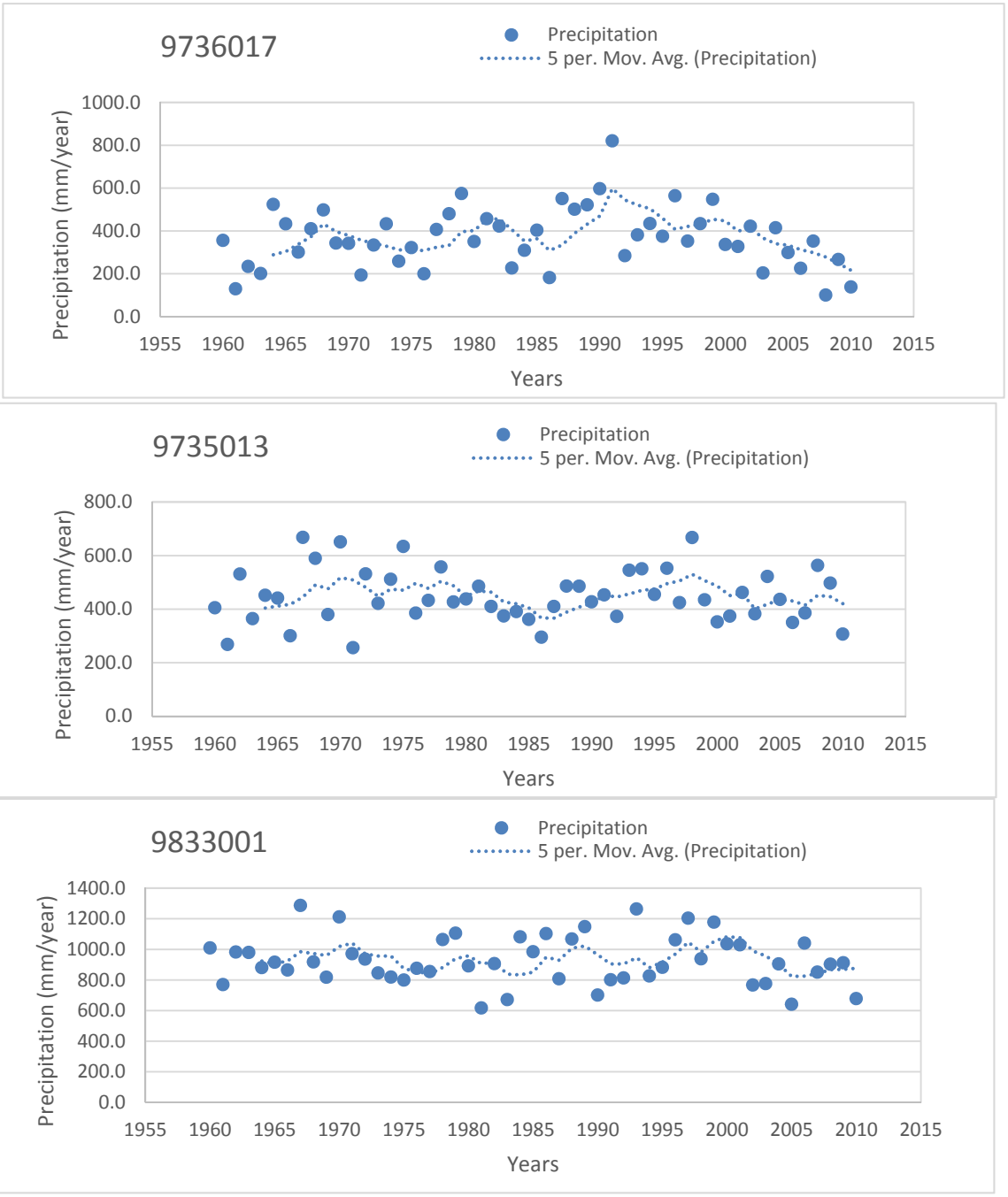


Figure 3.15 Spatial distribution of annual precipitation (mm/year) at individual gauges within the GRR Sub-basin

3.9.3 Interpolation of rainfall data

Accurate rainfall data are of prime importance for many hydrological applications. To obtain spatially distributed rainfall data, point measurements are interpolated. However, in areas with low density measurement networks, the use of different interpolation methods may result in highly variable results, and a large degree of uncertainty in the actual spatial distribution of

rainfall. An analysis of the few complete records from observed rainfall data showed that if five years of data are available from a given station, a reasonable annual average can be calculated with a maximum deviation from the true annual average of ~16–18% (mean deviation 5–8%) (Stisen,2013 unpublished). In this study, 130 rain gauge stations were selected from the 175 available (Figure 3.16). The selection was based on data consistency, length of record and proximity to the Great Ruaha Basin. The interpolation was done using the regression based Inverse Distance Weighting (IDW) method utilising monthly means of the Climate Prediction Center/Famine Early Warning System (CPC-FEWS v2) precipitation as a covariate. CPC FEWS RS is based on Global Telecommunication System (GTS) rain gauge reports and produced specifically for the African continent by use of the satellite METEOSAT TIR (Stisen et al, 2008; Stisen et al.,2009). The annual precipitation from the 130 rain gauge stations has been correlated with both elevation and annual precipitation from remote sensing (CPC-FEWS) (Figure 3.17). The interpolation of rainfall data was done as part of the CLIVET project by (Stisen,2013 unpublished) and the output is utilised for the hydrological modelling in this study.

3.10 Evaporation and Temperature

Pan evaporation data were available for two stations only (the Iringa and Mbeya stations). The aerial coverage of climatological and pan evaporation stations is inadequate for detailed evapotranspiration mapping within the GRR Sub-basin (Figure 3.18). In addition, the quality of data from these stations is poor, primarily because of station maintenance and operational problems. Therefore, any analysis of the evapotranspiration within the GRR Sub-basin will be subject to considerable uncertainty. Based on previous studies (CCKK, 1982), evaporation rates from free water surfaces are approximately 20% higher than the potential evapotranspiration rates. Moreover, unlike with rainfall which varies from year to year, evaporation rates in the basin showed little inter-annual variability, however, the spatial variability is considerable, ranging from 1 800 mm in the semi-arid part of the basin to less than 800 mm per year in the wet highlands.

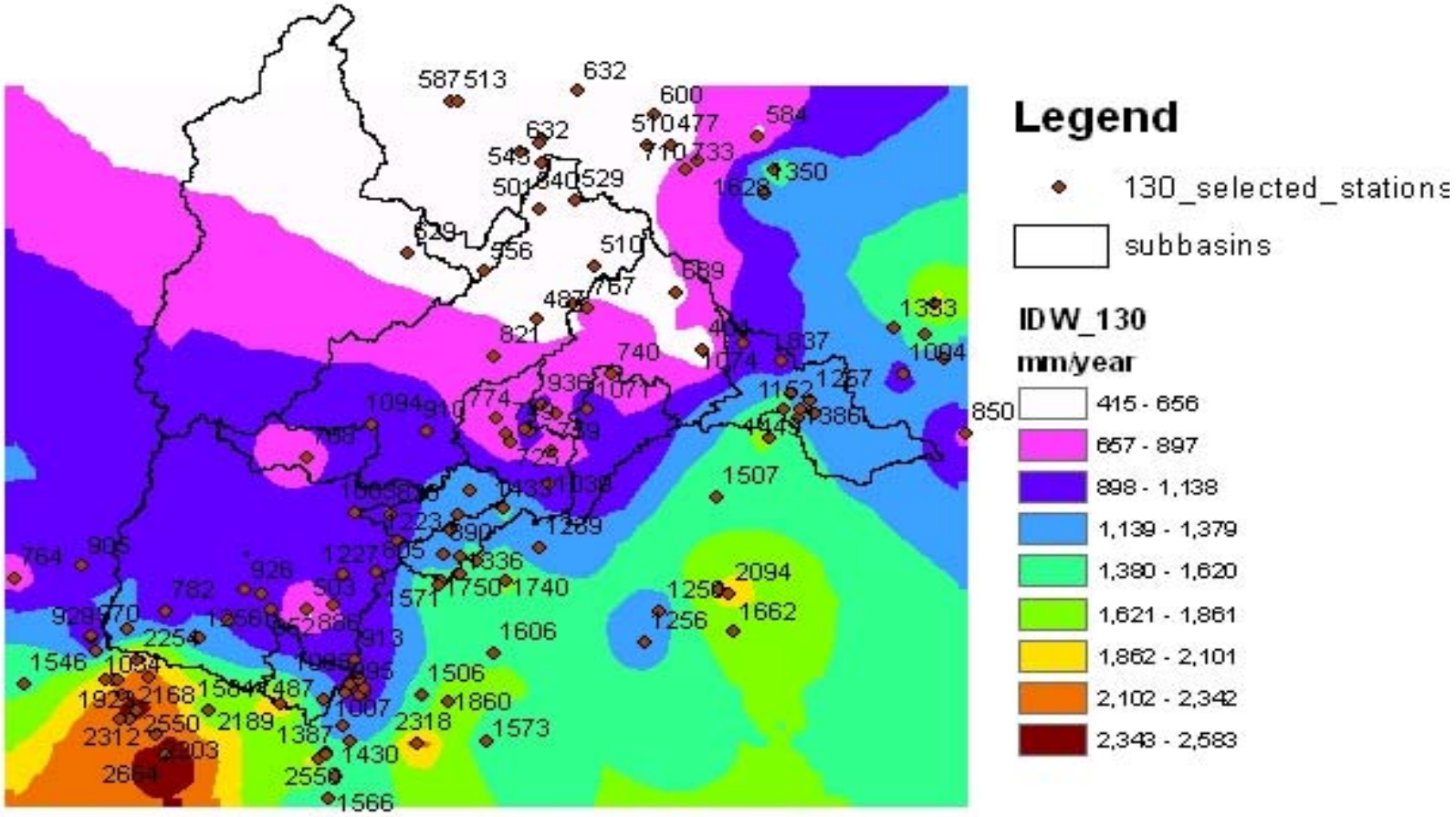
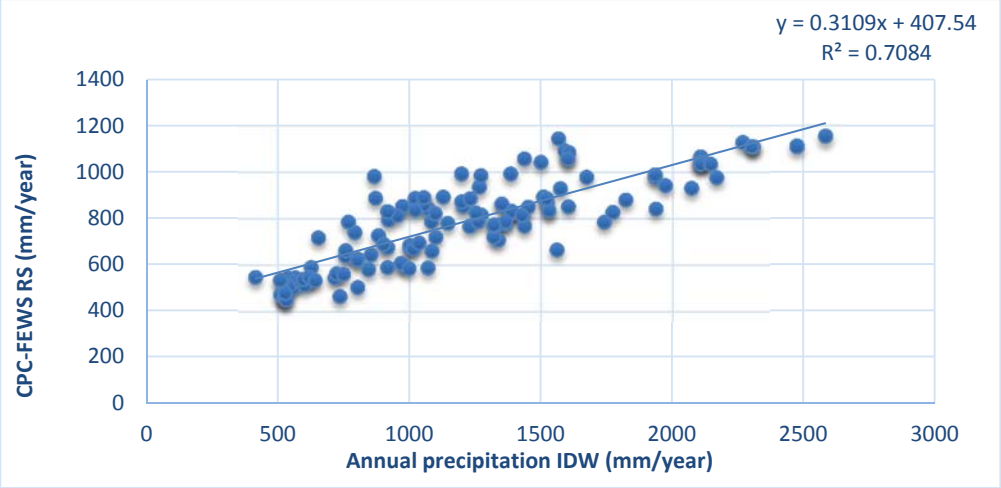


Figure 3.16 The 130 selected rain gauge stations within and outside the GRR basin



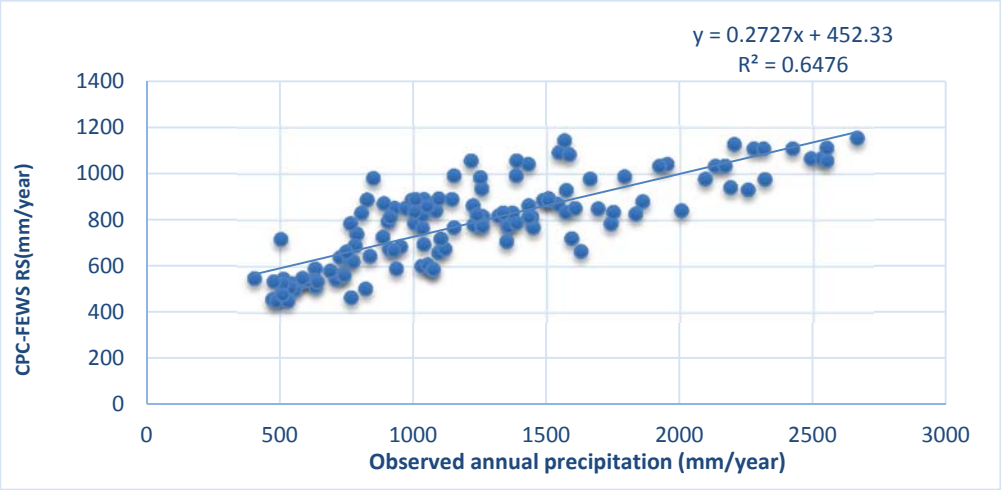


Figure 3.17 Correlation between CPC_FEWS annual precipitation IDW and observed annual precipitation

There is a tendency of decreasing evaporation with increasing altitude and rainfall (Figure 3.19). Due to poor quality and spatial distribution of the few pan evaporation data, evaporation data for use in this study were downloaded from the International Water Management Institute (IWMI) online climate service portal (http://wcatlas.iwmi.org). These data are based on the Penman-Montieth evapotranspiration estimates which compute the evaporation from open water surfaces using climate records of sunshine, temperature, humidity and wind speed (Allen et al.,1998). Monthly evaporation data have been extracted based on latitude and longitude coordinates for the specific location of the basin.

Temperature

Temperature data were collected from the Tanzania Meteorological Agency. Annual temperature within the basin ranges from 17 °C–23 °C where the high elevation areas are characterised by low temperatures (Figure 3.20). Minimum and maximum average monthly winter and summer temperatures vary from 5 °C to 13 °C and 22 °C to 27 °C, respectively in the higher altitudes and from 15 °C to 24 °C and 28 °C to 30 °C, respectively in the lower parts of the basin. Most of the lower parts of the study area comprising the Usangu Plains is semi-arid to sub-humid, whereas the highest part of the basin is humid with a sub-humid belt in between (Figure 3.20).

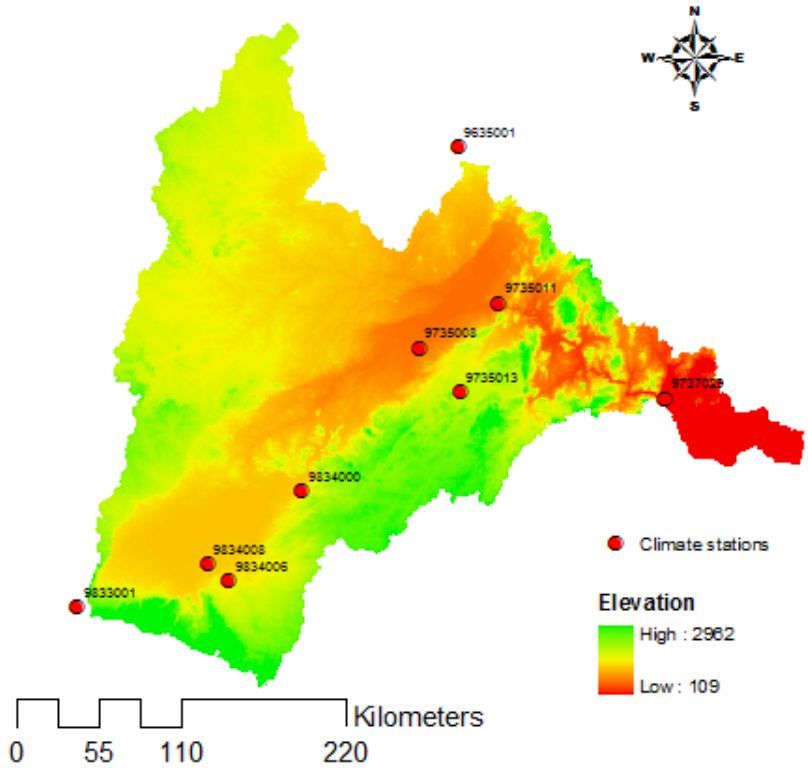


Figure 3.18 Spatial distribution of climatic stations in the GRR basin

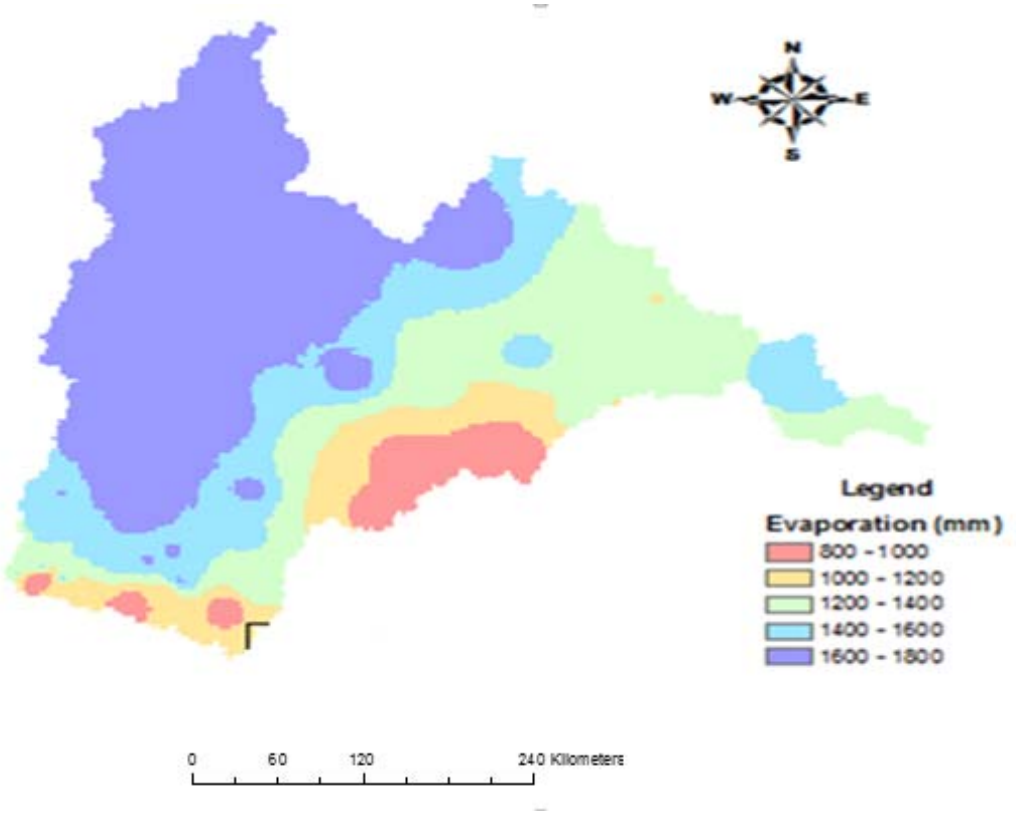


Figure 3.19 Spatial distribution of annual evaporation (mm/year)

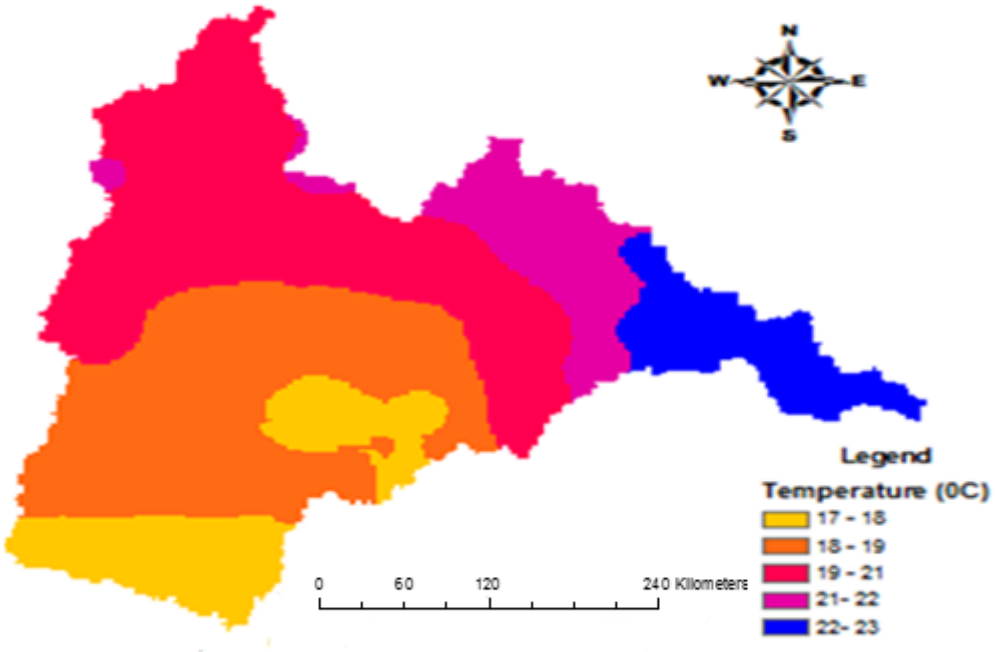


Figure 3.20 Spatial distribution of annual temperature (mm/year) within the GRR basin

3.11 The GRR network and River Flows

Daily flow data have been acquired from the Rufiji Basin Water Office (RBWO). Historically, Great Ruaha stream flow has been monitored by 33 river gauging stations. Most of the 33 stations were reported to be currently operational, but a large number of them are not. For this study, data from 26 gauging stations were available including both currently operational and non-operational stations (Table 3.4). The river network consists of stations established in the mid-1950s. The longest historical records span 30 to 35 years, but missing data can be as high as 50% (Table 3.4). The locations of the river gauging stations in the Great Ruaha Sub-basin are shown in Figure 3.23. Table 3.4 below shows discharge data available for each station, record length and percentage of missing values. Unfortunately, many of these stations are missing records over the last 30 years, and these are the most critical years in terms of the alleged recent flow changes. The distribution of stations is uneven, with most stations located along the base of the southern highlands, primarily along the main road. The SMUWC project installed one additional flow-gauging station at Nyaluhanga (1ka71a) located on the Great Ruaha River, approximately at the inlet of the eastern wetland (Ihefu).

Seasonal variations in river runoff

The runoff patterns in the sub-basin correspond closely to the uni-modal rainfall patterns prevailing in the whole basin. The streams start rising during November-December and experience a maximum flow during March–April and have their recession period from May to October/November. However, the seasonal variations of the runoff (flow regime) vary quite considerably for the three main rivers flowing to Mtera Dam in the basin as can be in Figure 3.22 below.

The Kisigo (1ka42), a seasonal river, is dry for four to five months a year as the low rainfall and high potential evaporation result in very little ground water recharge and consequently no dry season base flow. Due to the semi-arid nature of the catchment, significant rainfall occurs only during a small portion of the year therefore, runoff is very much dependent on rainfall. The effects of soil permeability, vegetation and topography are apparent and are affecting the water transmission properties of the soil. Topography is more important in determining the

rate of surface runoff. Based on the United States Soil Conservation Service (USSCS) method applied in the Singida Region Water Resources Surveys, it was indicated that the antecedent precipitation index (API) directly affects the curve number (Cn) value applicable to the catchment. For dry conditions (API 14 equal to zero) rainfall had to exceed a threshold of 13 mm before runoff occurred, and in wet conditions (high API 14) the majority of rainfall was converted to runoff (Singida Region Water Resources Surveys, 1984).

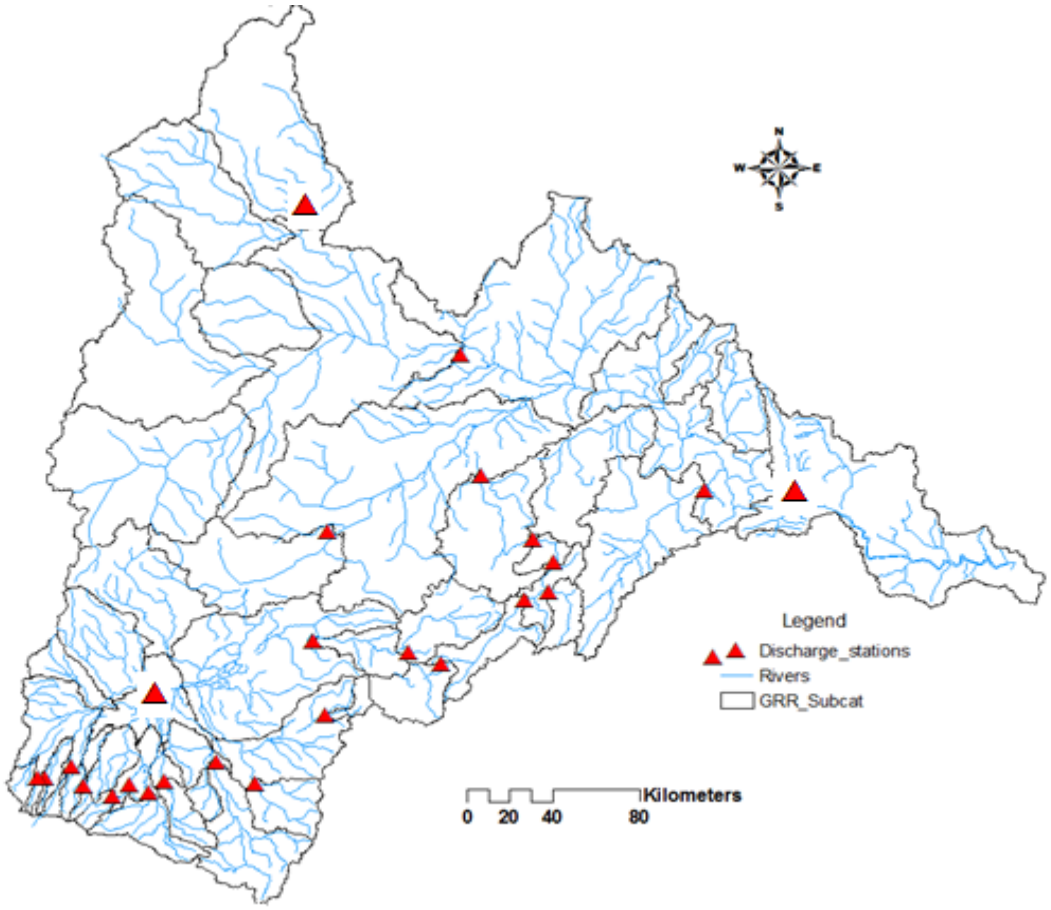


Figure 3.21 Spatial distribution of stream flow gauging station in the GRR basin

Table 3.4 Stream flow gauging stations in the Great Ruaha Sub-basin

Station ID	River	Start	End	Lat.	Lon.	% Missing data	Status
1KA2A	Little Ruaha	1954	2009	-7.8	35.7	30	Operational
1KA7A	Chimala	1962	1992	-8.9	34.0	52	Closed
1KA8A	Great Ruaha	1954	2009	-8.9	34.1	34	Operational
1KA9	Kimani	1954	2009	-8.9	34.2	17	Operational
1ka10	Mlomboji	1956	1983	-8.78	34.35	25	Closed
1KA11	Mbarali	1955	2009	-8.8	34.4	17	Operational
1ka12	Halali	1956	1983	-8.85	34.57	20	Closed
1KA15	Ndembera	1956	2010	-8.3	35.2	12	Operation
1KA16A	Lunwa	1964	1994	-8.9	33.8	58	Closed
1KA21	Little Ruaha	1957	2009	-7.9	35.8	10	Operational
1KA22	Mtitu	1957	2009	-8.0	35.8	3	Operational
1KA27	Great Ruaha	1965	1979	-8.00	34.58	36	Closed
1KA31	Little Ruaha	1957	2010	-7.5	35.5	21	Operational
1ka32A	Little Ruaha	1957	2009	-8.3	35.3	11	Operational
1KA33	Ndembera	1957	2009	-8.2	34.8	52	Closed
1KA37	Lukosi	1957	2010	-7.6	36.4	6	Operational
1KA39	Little Ruaha	1957	1980	-8.1	35.7	58	Closed
1KA41	Kisigo	1957	1994				Closed
1KA42	Kisigo	1957	1994	-7.0	35.4	66	Closed
1KA50A	Mswiswi	1958	1976	-8.8	33.8	68	Closed
1KA51	Umrobo	1960	1994	-8.8	33.7	32	Closed
1KA56	Ruaha	1961	1989	-8.6	34.9	12	Closed
1KA59	Great Ruaha	1963	2010	-7.8	34.9	15	Operational
1KA61	Great Ruaha	1965	1976	-7.6	36.8	77	Closed
1kA71A	Great Ruaha	2001	2008	-8.40	34.23	52	Closed
1ka5A	Great Ruaha	01957	1980	-7.07	35.98	58	Closed

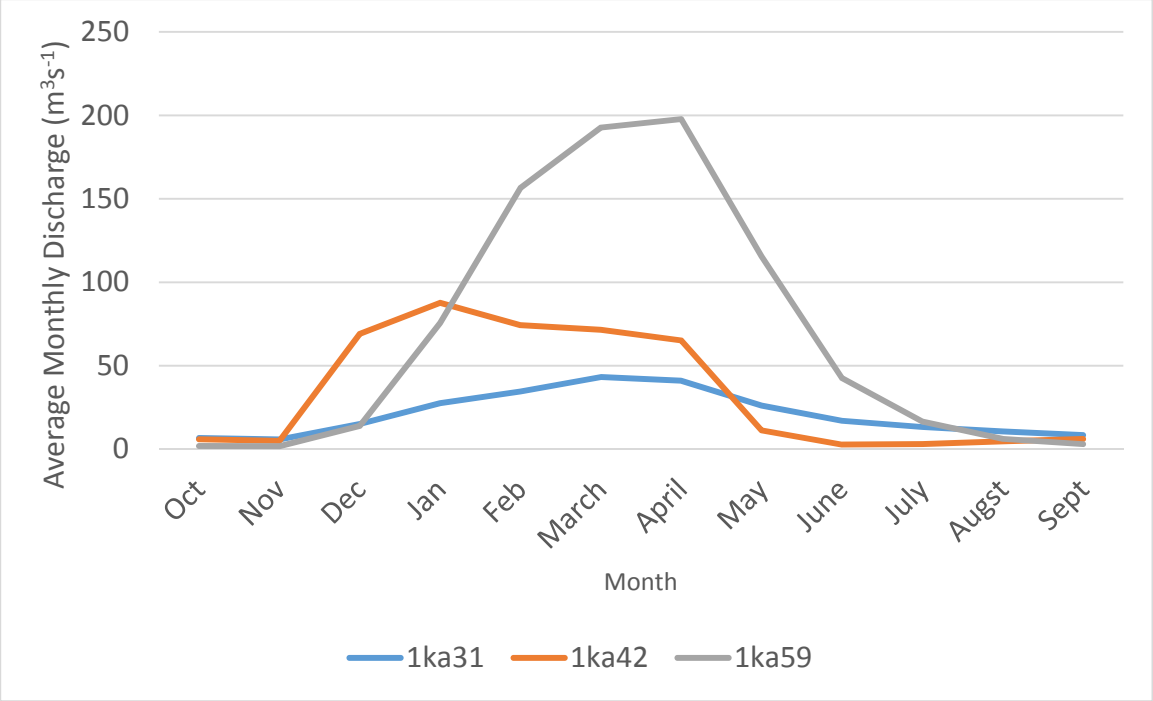


Figure 3.22 Average monthly discharge for the Little Ruaha River (1ka31), GR River (1ka59) and Kisigo (1ka42)

The headwaters for the Little Ruaha River (1ka31) originates from a permanent swamp covering an area of approximately 30–50 km^2. The seasonal variation of the runoff is less apparent for the Little Ruaha River, due to a considerable infiltration and ground water recharge during the wet season which are favoured by relatively high and often less intensive rainfall (DANIDA/Worldbank, 1995). The maximum and minimum recorded flows of the river are 775.0 and 2.8 $m^3 s^{-1}$ during the months of March and October, respectively. Estimates of groundwater recharge are discussed in the Water Master Plan for Iringa, Ruvuma and Mbeya regions (CCKK, 1982). Based on the CCKK report, the base flow component constituted about 80% of the total annual stream flow, which is consistent with the fact that the catchment is characterised by swamps in the headwaters, but also has highly permeable soils. This implies that there is high recharge. CCKK (1982) and water balance calculations for the catchment estimated quantities as given in Table 3.5

Table 3.5 Water balance calculated values for the Little Ruaha River

	Year	Rainfall(mm)	Recharge (mm)
Dry year	1977	85	455
Medium year	1960	955	173
Wet year	1962	1229	365

(Source: CCKK 1982)

The magnitude of the flow regime for the Great Ruaha River lies in between the flow regimes for Kisigo and Little Ruaha Rivers. The Great Ruaha River originates from an area with a similar amount of rainfall as the Little Ruaha Catchment, but the steep topography, often with exposed rocks, results in more surface runoff and less infiltration and ground water recharge during the rainy season.

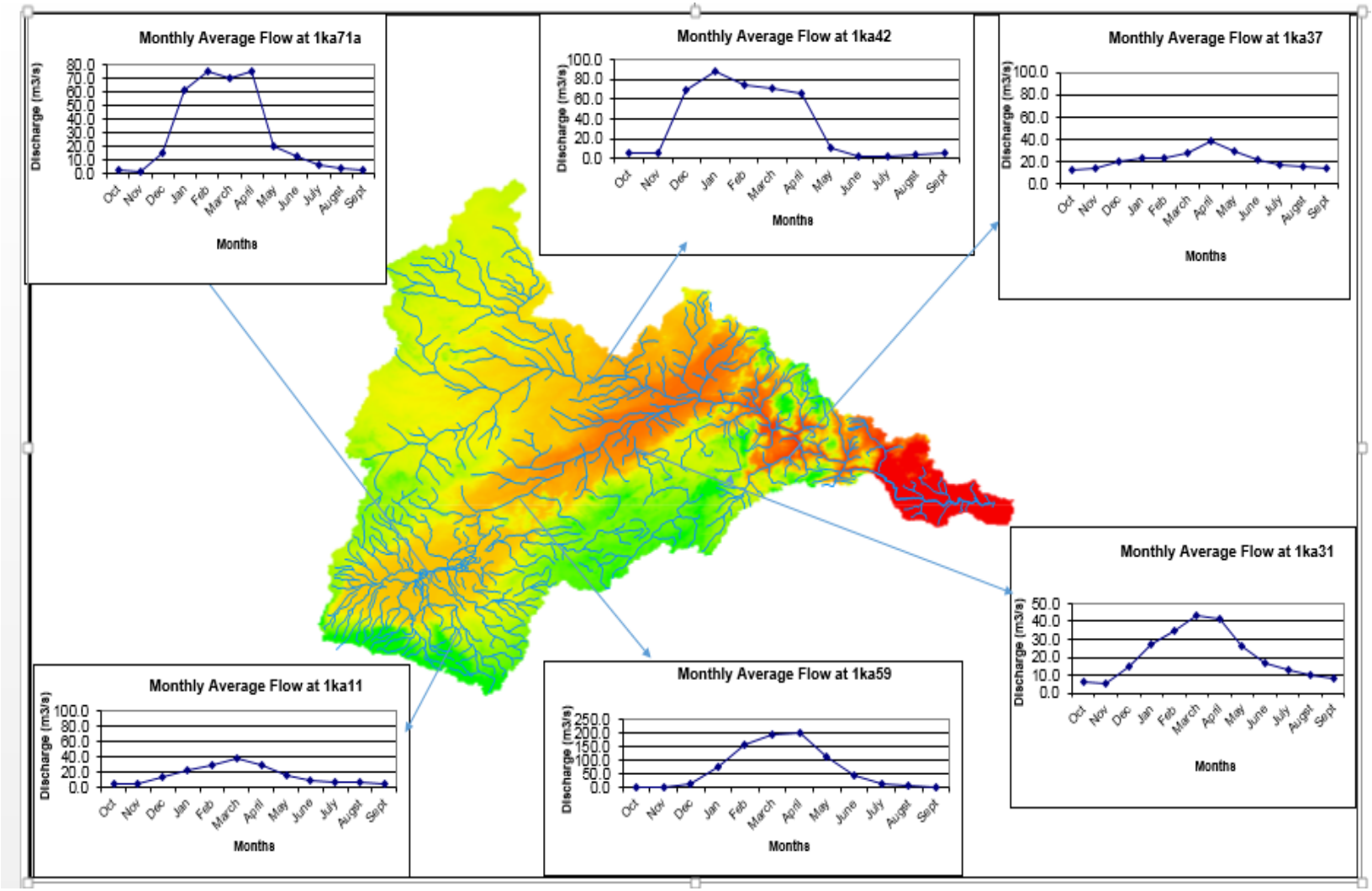


Figure 3.23 Seasonal distribution of average monthly discharge at individual gauges within GRR basin

Table 3.6 Characteristics for the catchment areas upstream of the main river gauging stations within the Great Ruaha basin

Station ID	River	Basin	Catchment km^2	Mean Annual discharge($m^3 s^{-1}$)	Mean Annual Rainfall(mm)	Mean Annual Runoff
1ka51	Umrobo	Great R	55	0.7	1 200	384
1ka16a	Liosi	Great R	77	1.5	1 200	611
1ka7a	Chumala	Great R	167	3.7	1 300	635
1ka8a	Great R	Great R	795	14.5	1 000	574
1ka9	Kimani	Great R	448	7.5	950	528
1ka10	Mlomboji	Great R	235	2.2	950	528
1ka11	Mbarali	Great R	1 600	17.2	1 000	340
1ka12	Halali	Great R	47	3.0	900	201
1ka23a	Huhuni	Great R	803	1.3	900	50
1ka33a	Ndembera	Great R	2 190	11.8	600	170
1ka27	Great R	Great R	19 941	50.0	800	79
1ka59	Great R	Great R	24 620	64.2	800	134
1ka32a	Little R	Little R	759	4.4	1 000	184
1ka39	Little R	Little R	1 740	12.5	1 000	226
1ka22	Mtitu	Little R	445	4.0	1 000	286
1ka31	Little R	Little R	5 193	19.2	850	116
1ka42	Kisigo	Kisigo	24 500	30.8	500	40
1ka5	Great R	Great R	67 950	114.0	500	53

3.12 Water use

The information on water use are based on previous projects (DANIDA/World Bank 1995; SMUWC, 2001) as well as estimates done using Google Earth and population census data (for domestic water use). Water use abstraction data was also obtained from the Rufiji Basin Water Office based on licensed irrigators. The 2012 population census results were used to estimate the domestic water uses at the sub-basin scale. The total domestic water uses for the whole of GRR Sub-basin was estimated to be 16.220 * 10^6 m^3 per year.

Irrigation Water use

The upstream part of the GRR Sub-basin has a high potential for paddy rice irrigation. Paddy rice is mainly grown in the Usangu plains during the wet season, hence the abstraction of water for the purpose of paddy irrigation is to supplement the rainfall. In wet years it is estimated that the area under irrigation can be up to 42 000 hectares, whereas in dry years the area under irrigation is estimated at 17 000 hectares. During the dry season (June to October) water is abstracted from rivers to irrigate other crops such as maize, beans, vegetables and some late planted rice fields. The estimated irrigation areas of each sub-basin are based on SMUWC project field surveys and aerial photographs interpretation (SMUWC, 2001). A calculation was carried out to determine an approximate value volume of return flows, (Approximate irrigation water use per crop * % of the water likely to contribute to return flows *approximate irrigated area (km^2)). The rice irrigation water use estimates was based on an annual demand of 3 007 mm with a high return flow fraction (0.33). The other irrigation areas used an annual demand of 431.5 mm with lower return flow fractions depending on the topography (0 in flat areas, 0.2 in steeper sub-basins). The GRR sub-basin were divided into six (6) regional groups based on similarities in their hydrological response characteristics (Chapter 5 , Table 5.5) as well as their physical properties (topography, geology, vegetation, etc.) and Ungauged sub-basins were denoted as UG (Chapter 4).

Table 3.7 Irrigated area and irrigation return flow fraction estimates

Group 1			Group 2			Group 3		
	Irrigation area (km^2)	Domestic use ($m^3 * 10^3 y^{-1}$)		Irrigation area (km^2)	Domestic use ($m^3 * 10^3 y^{-1}$)		Irrigation area (km^2)	Domestic use ($m^3 * 10^3 y^{-1}$)
1ka7a	21	341	1ka11	84	472	1ka2a	0	66
1ka8a	54	815	1ka12	0	900	1ka15	61	737
1ka9	37	409	1ka56	0	49	1ka21	0	93
1ka16a	0	65	ug1	126	557	1ka22	0	386
1ka50a	0	26	ug2	146	157	1ka31	0	219
1ka51	0	8	ug4	(Rice)47	151	1ka32a	0	352
ug15	0	54	ug6	(Rice)30	55	1ka33	76	413
ug16	0	74	ug20	(Rice)10	152	1ka37	0	590
ug17	0	48	ug21	0	40	1ka39	0	147
			ug22	0				

3.13 Concluding remarks

This chapter has provided a description of the Great Ruaha Sub-basin and its associated hydro-meteorological, topographic, soil and geological information. This chapter serves to present a conceptual understanding of the different components within the hydrological system, as synthesised on a geomorphologic scale where differences in key processes across the GRR Sub-basin have been identified. The description of the differences based on geomorphologic landforms facilitates the understanding of the different spatial variations internally within the basin, and later on, this description guides the similarity analysis, regionalisation and constraints development. One of the main differences across the catchment is the variable geology of the drainage sub-basins within the GRR sub-basin. This chapter also described the differences in area, topography, soils, precipitation, land cover and river flow within the GRR Sub-basin.

4 HYDROLOGICAL MODELLING METHODS AND UNCERTAINTY ANALYSIS

4.1 Introduction

The degree of confidence in predictions from hydrological models depends on how well the models represent natural hydrological processes taking place in the basin. Calibrating a hydrological model is a very challenging task due to the uncertainty related to the input data used to drive the model, the model structure and the parameters used to set up the model in any particular basin. It is also influenced by the shortage of historical observations against which calibrations of the model are performed, and the representativeness of the flow records to the natural hydrology of the sub-basin being modelled. Uncertainty is an inevitable problem in hydrological modelling and this must be taken into account before model prediction results are used for impact assessment studies and in the decision making process. This chapter gives a description of semi-distributed models GW_PITMAN and SWAT, their set up and the uncertainty frameworks used in the two models.

4.2 GW_PITMAN

The original Pitman model (Pitman, 1973) has undergone several modifications made to account for the continued challenges of water resources management in Africa. In this study, a modified version of the PITMAN model (GW_PITMAN model, Hughes, 2004; Hughes et al., 2006) is used to represent the processes at the sub-basin scale and includes simple components to represent groundwater processes; GW_PITMAN therefore is the groundwater (GW) version of the Pitman model. The model has demonstrated applicability in the southern African region and remains the most widely used hydrological model for research and practical water resources management. The GW_PITMAN model is a conceptual, semi-distributed monthly rainfall-runoff model consisting of storages (interception, soil moisture and groundwater) linked by functions designed to represent the main hydrological processes such as infiltration, saturation excess flow, direct overland flow, interflow and groundwater flow at the sub-basin scale (Hughes et al.,2006; Kapangaziwiri, 2008; Hughes, 2013a).The most detailed description of the algorithms and parameters of the GW_PITMAN model (except the wetland sub-model) is given in Kapangaziwiri (2008), while a more accessible description can

be found in Hughes et al. (2006). The main structure of the GW_PITMAN model is presented in Figure 4.1, and Table 4.1 lists the main components and parameters of the model. The description provided below represents a brief summary of the model parameters. It is necessary to briefly describe the parameters of the model given that the uncertainty related to their quantification is an integral part of this study.

Rainfall Distribution Factor (RDF): The RDF parameter accounts for distribution of the rainfall input in the model over the four iteration periods within each month of the time series. The distribution of total monthly rainfall is controlled by a symmetric S-curve function that is dependent on the total rainfall and the rainfall distribution factor. The lower values of RDF result in a more even distribution of rainfall (Mwelwa, 2004; Hughes et al., 2006). This parameter can be highly variable depending on the type of rainfall events that occur and the number of rain days. Arid zones that receive infrequent rainfall events are expected to have large values of RDF because this will ensure that rainfall is concentrated in two of the iteration periods. Temperate zones with more frequent rainfall events are expected to have lower values of RDF.

Interception and Evapotranspiration Parameters (PI1, PI2, AFOR, FF, R): The interception parameters account for the proportion of precipitation captured by the vegetation cover. The interception capacity within a sub-basin has a seasonal variation and is affected by the type of vegetation and the density of the vegetation cover. Interception storage parameters (P1 and P2) in the model account for two different vegetation types. Parameter AFOR is used to represent the proportion of the basin under vegetation type 2, while parameter FF is used as an evaporation scaling factor for vegetation type 2 relative to type 1. Based on the literature (Kapangaziwiri,2008) and limited empirical research, non-indigenous vegetation (plantation forests) can lose up to 40% more water through evapotranspiration than indigenous vegetation. Therefore, the value of FF is typically between 1.0 and 1.4. Greater values of FF in the model would imply higher evaporative losses through the secondary vegetation. R defines the relationship between the ratio of actual evapotranspiration to potential evaporation and the level of the soil moisture store (S). The value of R varies between 0 and 1, with lower values implying more effective evapotranspiration at low soil moisture levels.

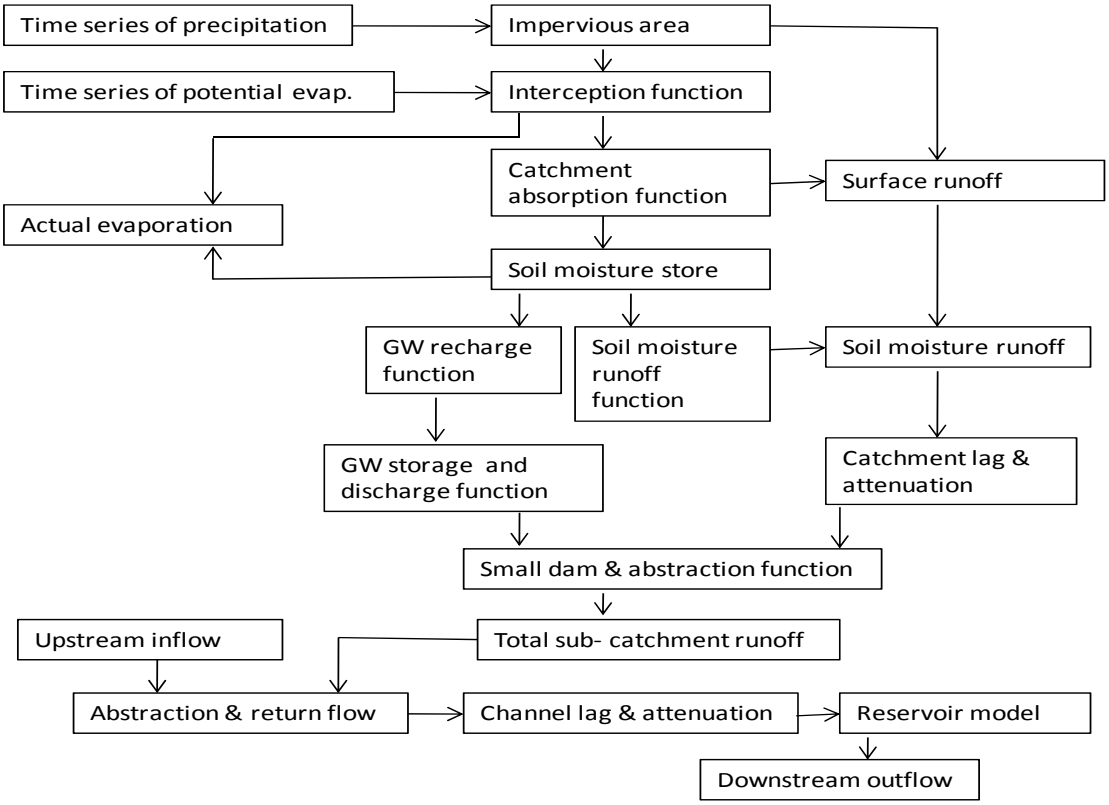


Figure 4.1 The main structure of the GW_PITMAN model

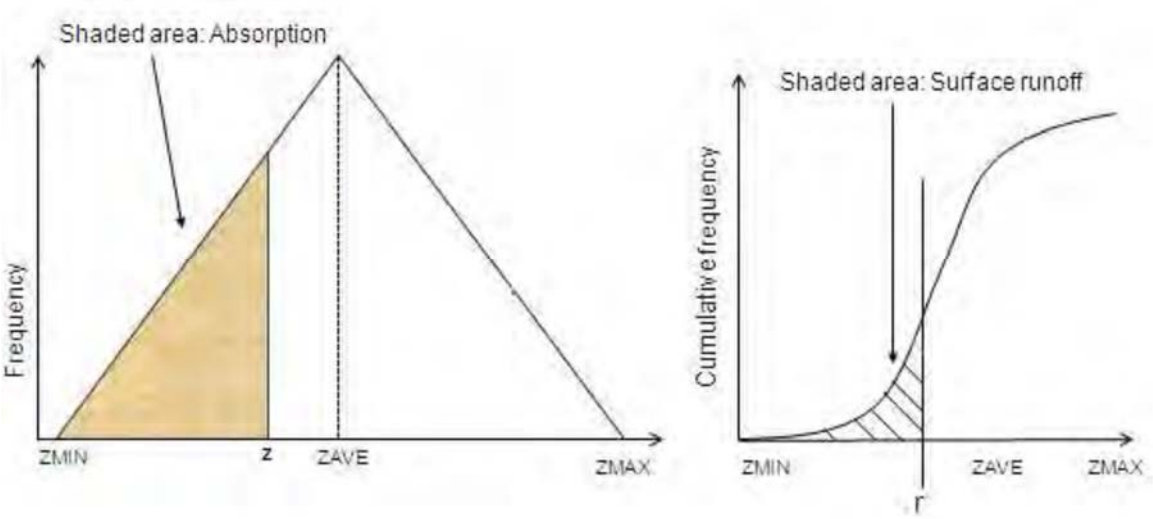


Figure 4.2 Frequency distribution of the catchment absorption rate Z in mm month^{-1} (left side) and cumulative frequency curve of the surface runoff generation (right side). r is the rate of rainfall input (Pitman, 1973)

Infiltration Parameters (AI, ZMIN, ZAVE, and ZMAX): The infiltration parameters account for surface runoff generation processes. The parameter AI represents the proportion of the basin that is impermeable, leading to infiltration excess surface runoff. The ZMIN, ZAVE and ZMAX parameters describe the absorption capacity of the sub-basin in response to different rates of rainfall input. The model makes use of an asymmetrical triangular distribution of basin absorption rates varying from a minimum value of ZMIN to a maximum value of ZMAX, and a mean value ZAVE (Figure 4.2). The values of the infiltration parameters are dependent upon a range of soil properties such as structure, porosity, texture, macro-pore density and surface sealing.

Soil Moisture Storage Parameters (ST): The soil moisture content of the GW_PITMAN model is controlled by the parameter ST, representing the maximum storage capacity of the soil in mm. The moisture storage of a soil increases due to infiltration. The water stored in the soil is absorbed by plant roots or evaporated from the topsoil into the atmosphere, or contributes runoff beyond soil saturation and contributes to ground water recharge. Pitman (1973) introduced this parameter as one that determined the ability of basins to regulate the runoff from a given precipitation input.

Soil Moisture Runoff Parameters (FT, POW): FT presents the maximum runoff generated from the soil (interflow) when the moisture level (S) is at its maximum value (ST). Runoff from the soil moisture store is assumed to be regulated through a non-linear relationship between discharge and soil moisture. This relationship is represented by a simple power (defined by parameter POW) function.

Groundwater Recharge Parameters (SL, GW, and GPOW): The conceptualization of the groundwater recharge subroutine is the same as unsaturated zone runoff using a similar non-linear power function but with parameters GW (maximum recharge in mm $month^{-1}$) and GPOW replacing FT and POW. SL represents the lower limit of soil moisture below which no groundwater recharge occurs.

Table 4.1 A list of main components and parameters of GW_PITMAN model

Model Component	Model Parameter	Description	Unit
Surface Processes			
Precipitation	RDF	Rainfall distribution factor	[-]
Impervious area	AI	Impervious fraction of sub-basin	%
Potential ET	PEVAP	Annual sub-basin evaporation	mm
Interception	PI 1 and PI2	Interception storage for two vegetation types	mm
Evapotranspiration	AFOR	Proportion of the basin area covered by the second veg type	%
	FF	The ratio of forest/ grassland potential evapotranspiration	[-]
	R	Evaporation-moisture storage relationship parameter	[-]
Catchment absorption and surface runoff	ZMIN, ZAVE, ZMAX	Min, average and max catchment absorption rate	mm $month^{-1}$
Sub-surface processes			
Soil moisture store	ST	Maximum moisture storage capacity	mm
Soil moisture runoff	FT	Maximum interflow runoff at full capacity (ST)	mm $month^{-1}$
	POW	Power of moisture storage-interflow equation	[-]
Groundwater recharge	GW	Maximum groundwater recharge at full capacity (ST)	mm $month^{-1}$
	GPOW	Power of moisture storage-GW recharge equation	[-]
	SL	Soil moisture threshold below which no GW recharge occurs	Mm
Groundwater store and discharge	T	Groundwater transmissivity	m^2 d^{-1}
	S	Ground water storativity	[-]

	DDENS	Drainage Density	km km^{-2}
	Slope	Regional ground water gradient	%
	RWL	Rest water level	m
	RSF	Riparian strip factor for evapotranspiration losses from GW	%
Flow routing and water use			
Channel routing	CL	Channel routing coefficient	Months
	TL	Lag of surface and soil moisture runoff	Months
	TLGmax	Channel Losses	Months
Abstraction and return flow	IrrArea	Irrigated Area	Km^2
	IWR	Irrigation water return flow	Fraction
	NirrDmd	Non irrigation demand	m * 10^3 y^{-1}
	EffRF	Effective rainfall	Fraction

Groundwater storage and output parameters (DDENS, T, S, GWSlope, RWL, RSF):

The full details of the conceptualization of the groundwater components of the model are provided in Hughes (2004). DDENS refers to the drainage density of the basin and is expressed as a ratio of the channel length (using only channels expected to receive GW drainage) to the basin area given in km km^{-2}. DDENS is used to determine the basic geometry of the groundwater store (Hughes, 2004). T refers to the transmissivity (m^2 d^{-1}) of the aquifer and is a product of permeability and saturated aquifer thickness. S refers to the storativity, a measure of the capacity of the aquifer to store water. GWSlope represents the gradient used to determine downstream groundwater outflow volumes, while the Rest Water Level (RWL) represents the maximum depth below the channel that the aquifer is assumed to reach and groundwater movement is assumed to cease. RSF is the riparian strip factor parameter, which defines the volume of water lost through evapotranspiration close to the channel margin. Groundwater contributions to stream flow are determined from the current groundwater storage volume which is used to calculate local groundwater gradients (toward or away from the channel). The gradient in any month is used with T and the contributing channel length to determine the volume of discharge. If the gradient is negative, it is possible to simulate channel transmission losses using the TLGMAX parameter.

Runoff Routing Parameters (TL, CL): Runoff routing is presented by parameters TL and CL .The parameter TL refers to the runoff time lag in months that are applicable to the surface and soil moisture runoff components. CL is a parameter that performs the channel routing in large basins. A Muskingum function (Nash, 1959) is used in the model for the lag parameters, in which the weighting factor is set to zero to represent reservoir type storage attenuation (Hughes, 2004). TL is normally fixed at 0.25, but variation in the value is inevitable in large catchments.

Water Use Parameters (Airr, IWR, IrrAreaDmd, NIrrDmd, EffRf): The GW_PITMAN model has routines which account for direct abstractions for irrigation and non-irrigation purposes, and also other components to represent anthropogenic impacts on natural hydrology (distributed small farm dams and large reservoirs). Small farm dam routines in the model account for surface storage (and abstraction) of runoff generated within a sub-basin, while for large reservoir storages a main reservoir water balance component (Hughes, 1992) is included. The reservoir storage receives inflows from all upstream sub-basins and can be setup with a number of abstraction and release operating rules based on the level of storage in the reservoir. Both reservoir components (small farm dam and main reservoir) estimate evaporative losses on the basis of power relationships between simulated volumes (Vol - obtained from the model water balance) and surface area given by the equation below.

$$Area = A * Vol^{B} \quad \text{(Equation 1)}$$

While the model has been used quite frequently to simulate development effects, in complex situations (such as multiple water supply sources and complex operating rules) water resources systems models are more appropriate simulation tools.

The parameter Airr refers to the irrigated area in a sub basin given in km^2. The parameter accounts for the potential demand on river runoff by determining the size of area that can be irrigated. IWR refers to irrigation water return flow. EffRf is the effective rainfall fraction which is the proportion of a rain input that goes directly to satisfy irrigation demand. IrrAreaDmd refers to the total area irrigated from small dams and is expressed in km^2. NIrrDmd refers to the annual volume of non-irrigation demand given in $m^3 * 10^3 y^{-1}$. The non-irrigation parameter

is based on a specified annual demand value and is used together with a monthly distribution table.

4.3 GW_PITMAN Wetland sub-model and parameters

A wetland sub-model was developed at the IWR as a sub-component of the GW_PITMAN monthly time step model (Hughes et al., 2013d). The main aim of this development was to take into account wetland storage processes, thereby reducing model structural uncertainties and improving model simulations where the impacts of wetland or natural lakes are evident (Hughes et al., 2013d). The model was tested in some parts of the Congo; Okavango and Zambezi basins (Hughes et al., 2013d). The wetland model uses the following twelve parameters designed to account for the water balance of both lakes and wetlands.

i. Local area (km^2): The maximum wetland area that is permanently or periodically inundated and accounts for local runoff entering directly into the wetland. The use of aerial photos and site specific maps are necessary to determine the local wetland area.

ii. Residual wetland volume (RWV, $m^3 * 10^6$): The nominal storage capacity for the area permanently submerged and below which no downstream outflow occurs.

iii. Initial wetland volume (WV, $m^3 * 10^6$): The starting storage that depends on the season at the start of the model run.

iv. Area-volume relationship:

$$Area\ (km^2) = a * WV\ (Mm^3)^b \qquad \text{Equation 4.2}$$

Where a and b are two empirical parameters of the non-linear area-volume relationship with the constant and b the power of the equation.

v. Channel capacity for spillage (QCAP, $m^3 * 10^6$): The river channel monthly volume threshold below which there is no spillage to the wetland.

vi. Channel spill factor (Fraction): Proportion of the flow volume above the channel threshold that is assumed to spill to the wetland. Depending on the wetland type being studied, two situations need to be considered. The first case is a natural wetland where the river channel meanders through the wetland, thus spilling water into the wetland when the channel capacity for spillage is exceeded, and receiving water from the wetland during lower flows and when the wetland volume is greater than the residual

volume. The spill fraction will thus range from a value of 0 to 1 (Hughes et al., 2013d). The second example is for natural lakes or reservoirs where the river channel flows into the lake (in-channel wetland). So, the spill factor can be set to 1 and channel capacity for spillage to 0 to ensure that all flow enters the lake or wetland.

vii. Return Flow Factor (RFF) determines the amount of water that returns from the wetland to the river channel (Return Flow Volume, RFV) and that contributes to downstream outflow. A maximum fraction of 0.95 is used for the RFF (Hughes et al., 2013d).

$$RFV = RFF * (WV - RWV) \qquad \text{Equation 4.3}$$

The wetland storage-return flow relationship is given by:

$$RFV = AA * \left(\frac{WV}{RWV}\right)^{BB} * QCAP/Q \qquad \text{Equation 4.4}$$

Where AA is the return flow constant (RFC), BB is the power of the equation designed to account for a non-linear relationship, and Q is the flow in the river channel. Equation 4.3 accounts for the increase in return flow as wetland volume increases, while Equation 4.4 accounts for a reduction in return flow when the river volume is high. This last term of the equation is not used when QCAP is equal to zero, or when the lake is used. Return flow from the wetland to the channel occurs when WV exceeds the RWV. Losses in the wetland include mean annual evaporative losses (mm) from the freestanding water and the total annual abstractions (Mm^3) for water uses, both distributed as monthly percentages of the annual values.

4.4 PERFORMANCE MEAUSRES

Model performance assessment in the GW_PITMAN model is based on a set of several standard quantitative (dimensionless measures, error index and standard regression) and qualitative criteria that are used to reject or accept the model output simulations. These are: the percent bias of the mean monthly flows, percentage differences of standard deviations of monthly flows, coefficient of determination, Nash-Sutcliffe coefficient of efficiency, stream

flow hydrograph, flow duration curve and the monthly distribution graph. The quantitative assessments are taken for normal and natural logarithm transformed values. The transformation is necessary to more accurately account for medium to low flow components of the hydrographs.

Coefficient of determination (R^2): R^2 gives information about the goodness of fit of a model by describing the proportion of the variance in observed data explained by the model. R^2 ranges from 0 to 1. Although R^2 has been widely used for evaluating model performance, the statistics are sensitive to high extreme values and insensitive to additive and proportional differences between model predictions and measured data.

Coefficient of efficiency: The Nash–Sutcliffe model efficiency (Nash and Sutcliffe, 1970) coefficient is used to assess the predictive power of hydrological models. It is defined as:

$$E = 1 - \frac{\sum_{t=1}^{T} (Q_o^t - Q_m^t)^2}{\sum_{t=1}^{T} \left(Q_o^t - \overline{Q_o}\right)^2}$$

Equation 4.5

Where Q_o is observed discharge, and Q_m is modeled discharge. Q_o^t is observed discharge at time t. Nash–Sutcliffe efficiencies can range from $-\infty$ to 1. An efficiency of 1 (E = 1) corresponds to a perfect match of modeled discharge to the observed data. An efficiency of 0 (E = 0) indicates that the model predictions are as accurate as the mean of the observed data, whereas an efficiency less than zero (E < 0) occurs when the observed mean is a better predictor than the model or, in other words, when the residual variance (described by the numerator in the expression above), is larger than the data variance (described by the denominator). Essentially, the closer the model efficiency is to 1, the more accurate the model is. The coefficient of efficiency of inverse transformed data (CE (1/data)) can be used to further emphasize the fit to low flows.

Mean monthly percentage error of the simulated flows (Percent bias denoted as PBIAS): Percentage bias measures the average tendency of the simulated data to be larger or smaller than their observed counterpart. The optimal value of PBIAS is 0.0, with low-magnitude values

indicating accurate model simulation. Positive values indicate model underestimation bias, and negative values indicate model overestimation bias (Gupta et al., 1999). PBIAS is calculated with equation;

$$\text{PBIAS} = \frac{\sum_{i=1}^{n}(Y_i^{obs} - Y_i^{sim})}{\sum_{i=1}^{n}(Y_i^{obs}} * 100 \qquad \text{Equation 4.6}$$

where PBIAS is the deviation of data being evaluated, expressed as a percentage. According to Gupta et al. (1999) PBIAS has the ability to clearly indicate poor model performance.

Stream flow hydrograph: is a qualitative measure of model performance that involves visual inspection of the goodness of fit between the hydrographs of the simulated and the observed flows. It has the advantage of being able to evaluate the specific characteristics of stream flows such as timing, magnitude, early season flows, recession flows, peak flows, in order to judge the performance of the model outputs.

Flow duration curve (FDC): the FDC is the cumulative frequency distribution of the percentage of time a given flow magnitude in a river channel is equaled or exceeded. Visual comparison of observed and simulated FDCs provides a qualitative evaluation of the model performance based on the frequency distribution of high, medium and low flow. More quantitative assessments can be achieved by determining differences in frequency of exceedence at specific flows, or differences in flows at specific frequencies of exceedence.

4.5 SPATSIM

The version of the PITMAN model (GW_PITMAN) used here is applied within a modelling framework SPATSIM (Hughes and Forsyth, 2006). In SPATSIM, a database is managed through a spatial interface (shape files generated through GIS routines external to SPATSIM) that represents the basin features, which are linked to the data attributes. The software provides facilities which include routines for interpolation, patching, simulation, calibration and uncertainty analysis. There are a wide range of models linked to SPATSIM that cover various applications in hydrology and water resources estimation. Table 4.2 lists the SPATSIM

attributes (that form part of the common database) that are required by, and linked with, the GW_PITMAN model as part of a model set up.

Table 4.2 GW_PITMAN model setup requirements

Attribute Type	Attribute Requirement
Text	-Catchment ID -Downstream Area
Single real number	-Catchment area (km^2) -Catchment cumulative area (km^2)
Time series	-Average rainfall (mm) -Observed monthly flow (volume) -Downstream outflow (volume) -Uncertainty ensembles
One dimension array	-GW-model parameters -Mean monthly evaporation (monthly % of total annual) -Reservoir model parameters - Wetland model parameters
Two dimensional array	-Uncertain parameters -Monthly water distribution (fractions) -Reservoir monthly distribution (fraction) -Wetland seasonal distribution

4.5.1 GW_PITMAN uncertainty framework

The GW_PITMAN uncertainty framework and its implementation is presented in Kapangaziwiri et al.,2009; Kapangaziwiri 2010; Kapangaziwiri et al.,2012). The framework is applicable for both gauged and ungauged basins and has been applied successfully within South Africa. Part of the original framework involved the use of regional signatures of hydrological behaviour to constrain the uncertain outputs from a hydrological model so that only those that could be realistically considered to represent the behaviour of the catchment response would be included in further analyses (Kapangaziwiri, 2010). This study uses an improved version of the uncertainty framework (Figure 4.3) developed at the Institute for Water Research (IWR),

Rhodes University during the course of this study and its applicability is being tested in the GRR Sub-basin. This framework aimed at defining the uncertainty in the observed stream flow with respect to their ability to represent the natural conditions that the model simulates. The mean monthly runoff, mean monthly groundwater recharge, and flow duration curve values at the 10^{th}, 50^{th} and 90^{th} percentiles are extracted from observed discharge data and are used to guide the quantification of regional constraints.

The improved framework is implemented in the global options model. The GW_PITMAN Global Options model is a new component of SPATSIM, and consists of four different model types; *single run model with no uncertainty, incremental uncertainty run*, *cumulative uncertainty*, *cumulative uncertainty with stochastic rainfall inputs* and *structured uncertainty*. The single run, incremental and cumulative model runs were used for calibrations and uncertainty estimation and analysis. The description of each model version and its requirements is given in the next section. Figure 4.4 shows the main screen of GW_PITMAN Global Options Model which consists of options for model parameters, different available model types, sub-basins to be simulated, choice of output parameter sets, maximum parameter sample, the help facility and an option to explore stored parameter sets.

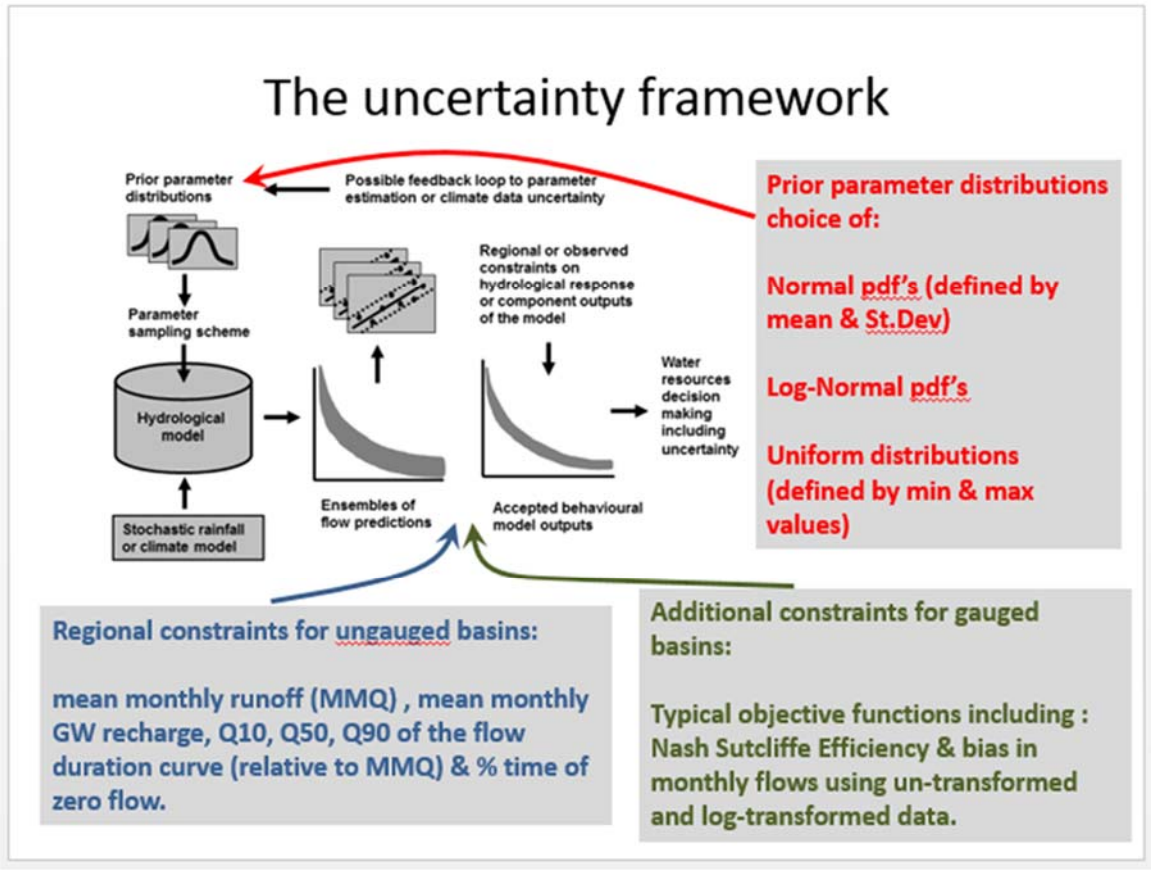


Figure 4.3 The GW_PITMAN uncertainty framework (Kapangaziwiri, 2010)

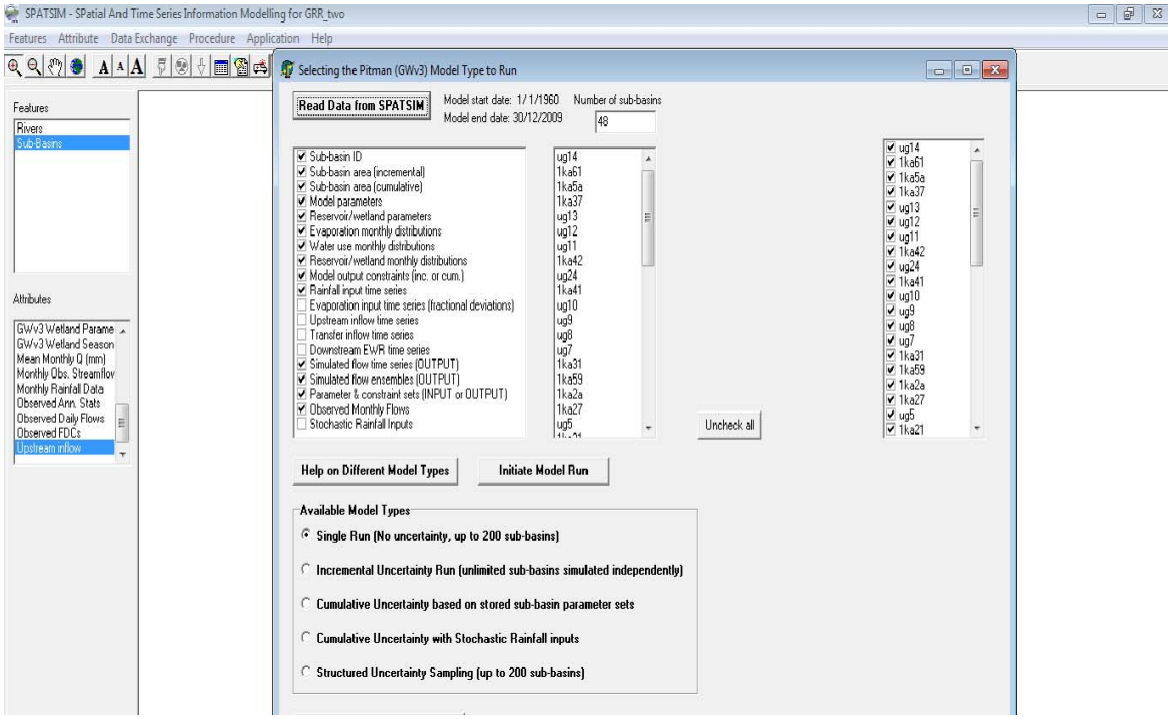


Figure 4.4 GW_PITMAN Global Options Model main screen

4.5.2 GW_PITMAN Global Options Model Types

This section summarises the different model types, while more details of the uncertainty sampling are provided in section 4.5.3.

4.5.2.1 Single run model

The single run uses one parameter set to generate a single simulation of the model. The model can simulate up to 200 sub-basins in the spatial distribution system. The simulated stream flow (cumulative) outputs are stored in a simulated flow time series attribute. The detailed outputs of individual model components are also stored in binary files and can be viewed and analysed to assess more than just the total stream flow outputs.

4.5.2.2 Incremental uncertainty run model

This model type simulates the incremental natural runoff only (i.e runoff without human abstractions) and it can be applied to one, all or selected sub-basins. The model is run independently on each sub-basin with Monte Carlo sampling of the parameter space. The outputs of each model ensemble are assessed against the output constraints (Step 1 of the

revised framework: Section 4.5.3) and if the constraints are met, the parameter set is saved to the database. The constraints consist of minimum and maximum values of mean monthly runoff volume (MMQ in MCM), mean monthly groundwater recharge depth (MMR in mm), Monthly flow duration curve percentiles (Q10,Q50,and Q90) extracted as fractions of MMQ (FDC Q_{10} / MMQ (fraction), FDC Q_{50} / MMQ (fraction), FDC Q_{90} / MMQ (fraction) and the percentage of zero flow (% time of zero flows) . The detailed description of the six constraints is given in chapter 5.

4.5.2.3 Cumulative uncertainty model

This version of the model utilizes the saved behavioral parameters from the incremental uncertainty run as well as the water use parameters (if quantified) and the downstream routing parameters. For this study water use abstractions and the wetland effect were applied in the model. The model generates 10 000 ensembles that are saved in the SPATSIM database. An additional output file contains the full list of parameter values and summary results data, as well as objective function values if observed data are available for that sub-basin.

4.5.3 Sampling procedure

All of the uncertainty versions of the GW_PITMAN model use a simple Monte Carlo sampling approach to generate ensembles of the model results. Within the uncertainty framework sampling can be done using either a structured or an unstructured approach. In the structured sampling approach sub-basins are grouped and the direction of the uncertainty for a specific parameter is always the same for sub-basins in the same group. Different groups have independent uncertainty samples. The objective of this approach is to preserve the degree of uncertainty in downstream sub-basins within model set-ups with many sub-basins. Without the grouping approach the downstream sub-basins tend to have very low uncertainty as the combinations of generally wet or generally dry simulations from upstream sub-basins cancel each other out.

Within large basins a further problem occurs if the range of behavioural parameter values is not known. The un-constrained outputs from an upstream sub-basin can be examined and the behavioural ensembles identified using regional or local constraints (Kapangaziwiri et al.,

2012). The same approach can be applied downstream, but the behavioural simulations at these sites can be made up of a mixture of behavioral and non-behavioural inputs from the upstream sub-basins. This could pose problems for spatially integrated water resources management.

The new sampling procedure based on two steps was subsequently developed in the IWR which is implemented within the Global Options Model. The first step is the incremental model, and is applied to the natural runoff parameters of the model and uses constraints based on natural runoff characteristics (such as mean monthly flow, mean monthly ground water recharge, and flow duration curve indices). The second step is implemented using the cumulative model where it is possible to include additional parameters such as water use, reservoirs, wetlands, etc. The descriptions of the two step parameter sampling procedure are given in the following sections.

4.5.3.1 Parameter sampling using the incremental model

The model uses constraints based on natural runoff characteristics (MMQ, MMR, Q_{10}, Q_{50}, Q_{90} and % of zero flows) to constrain model simulations. All sub-basins are simulated independently 100 000 times until 2 000 behavioral ensembles are attained. Behavioural parameter sets are defined as those that generate outputs that fall within the upper and lower bounds of the constraints (Figure 4.5A). Parameter sets giving results that will fall outside the constraints bounds will be rejected. Inconsistencies between the individual constraint bounds (Figure 4.5 B) or between the input parameter ranges and the constraint bounds (Figure 4.5 C) could lead to rejection of all possible 100 000 ensembles, Figure 4.6 show the results of a successful incremental model run with 2 000 behavioural parameter sets, and Figure 4.7 shows a partial failure with less than 2 000 saved parameter sets. These two figures represent screen shots of a utility developed to analyse the step 1 results and to refine the input parameter or constraint ranges if necessary.

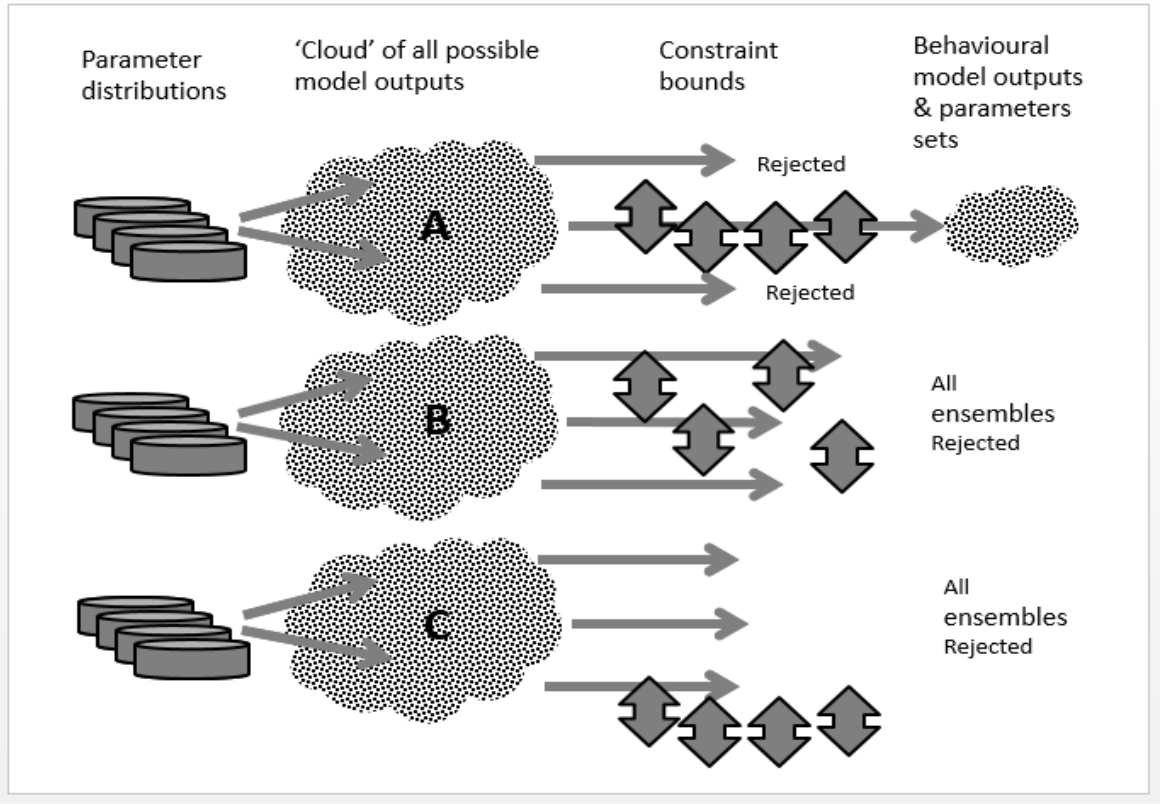


Figure 4.5 Parameter sampling procedure of the revised uncertainty framework. The thick arrows represent the constraint boundaries.

Figure 4.6 A successful sub-basin with 2 000 behavioural parameter sets generated with well distributed parameters and constraint values

Figure 4.7 An unsuccessful sub-basin showing parameter sets generated with a bias towards lower constraint values. Suggests adjustments needed to some parameter ranges

4.5.3.2 Parameter sampling using the cumulative model

This simulates the whole basin using structured random sampling from the behavioural parameter sets saved from step 1. This ensures that all of the sub-basin inputs to the whole basin hydrology are behavioural relative to the constraint bounds used. Additional parameters for such as wetlands, water use, and reservoirs are included in this step and their uncertainty distributions are independently sampled. The cumulative model outputs were used to examine whether the model had achieved practical and realistic uncertainty bounds throughout the basin. The model outputs can be examined using excel spreadsheets to sort the 10 000 parameter sets and summary statistics (including objective functions if observed data are available) in different ways. A facility is also available to extract two time series representing the ensemble minimum and maximum simulated flows for each month.

4.6 GW_PITMAN model setup for the Great Ruaha River Basin

The GW_PITMAN model was set up within SPATSIM for 48 sub-basins of the GRR basin. The GRR basin (86 000 km^2 in extent) (Figure 4.8) was delineated using the Shuttle Radar Topographic Mission (SRTM) Processed 90m Digital Elevation Data (http://srtm.csi.cgiar.org/). The International Centre for Tropical Agriculture (CIAT) has processed this data to provide seamless continuous topography surfaces. Areas with no data in the original SRTM data have been filled in using interpolation methods (SRTM, 2004). The basic characteristics of the delineated sub-basins are given in Table 4.3 and were named after the stream flow gauging stations at their outlets (e.g. 1ka7a) and for ungauged basins the UG symbol was used. The time series datasets consist of rainfall and stream flow. Rainfall data were obtained through spatial interpolation of the Climate Prediction Center (CPC), Famine Early Warning System (CPC_FEWS) remotely sensed rainfall data for the period 1960-2009, using the inverse distance weighted method (Chapter 3). The input data include the parameter values (main model and reservoir or wetland sub-models) and the seasonal distributions of some parameter values such as monthly water use and annual potential evaporation.

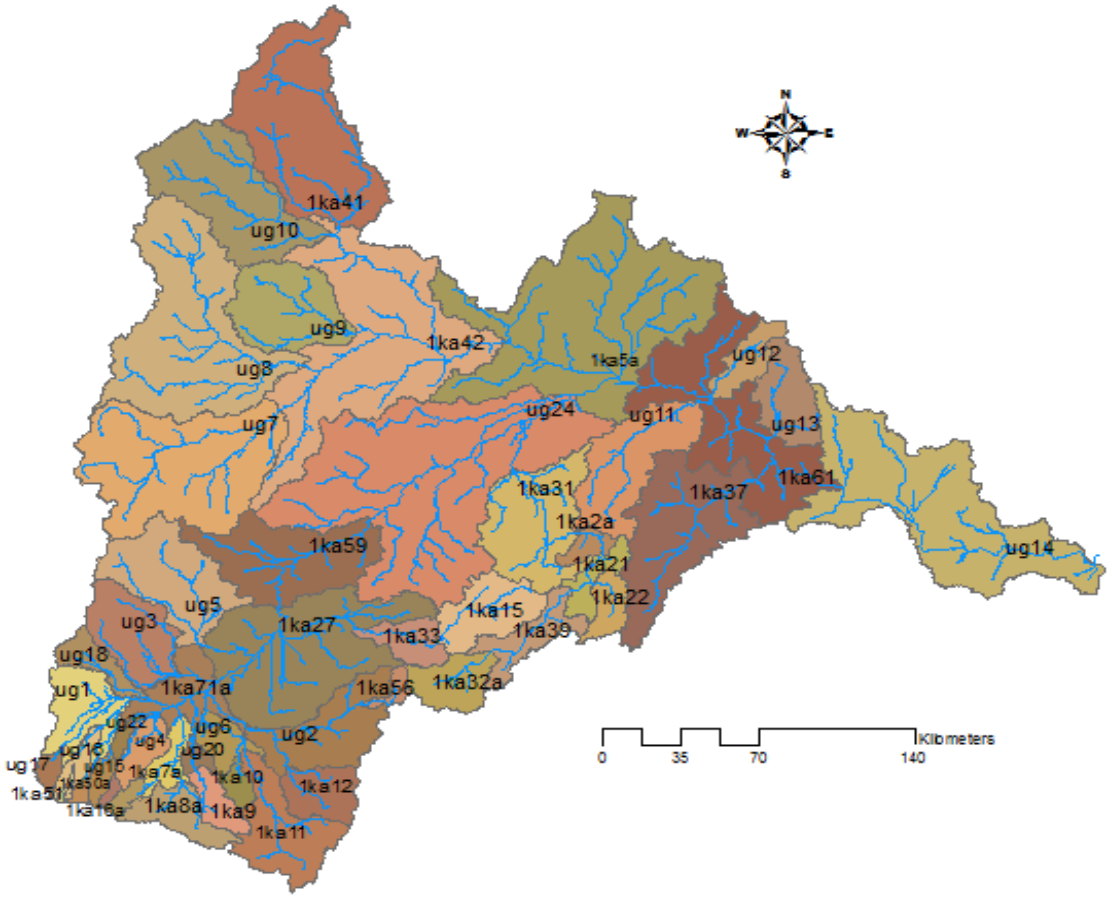


Figure 4.8 Delineated catchments of the GRR Sub-basin

Table 4.3 Description of the delineated catchments in the GRR Sub-basin

Catchment	River	Lat	Long	Catchment Area (km^2)	Cumulative _ Area (km^2)
1KA2A	Little Ruaha	-7.78	35.72	277	2896
1KA5A	Great Ruaha	-7.08	35.98	7315	69386
1KA7A	Chimala	-8.87	34.03	169	169
1KA8A	Great Ruaha	-8.9	34.12	783	783
1KA9	Kimani	-8.85	34.18	446	446
1KA10	Mlomboji	-8.78	34.35	233	233
1KA11	Mbarali	-8.77	34.4	1600	1620
1KA12	Halali	-8.85	34.57	807	807
1KA15	Ndembera	-8.28	35.2	1221	1221
1KA16A	Lunwa	-8.87	33.85	75	75
1KA21	Little Ruaha	-7.88	35.8	477	2619
1KA22	Mtitu	-8.02	35.78	461	461
1KA27	Great Ruaha	-8	34.58	4244	20746
1KA31	Little Ruaha	-7.5	35.5	2299	5195
1KA32A	Little Ruaha	-8.33	35.33	838	838
1KA33	Ndembera	-8.23	34.8	618	1839
1KA37	Lukosi	-7.57	36.43	3064	3064
1KA39	Little Ruaha	-8.05	35.68	843	1681
1KA41	Kisigo	-6.29	34.83	4368	4368
1KA42	Kisigo	-6.97	35.42	5999	25097
1KA50A	Mswiswi	-8.79	33.77	102	102
1KA51	Umrobo	-8.82	33.67	38	38
1KA56	Ruaha	-8.56	34.85	182	182
1KA59	Great Ruaha	-7.75	34.9	2539	23285
1KA61	Great Ruaha	-7.58	36.78	3691	79906
1ka71a	Great Ruaha	-8.4	34.2	1055	12514

Table 4.4 Description of the delineated catchments in the GRR Sub-basin

Catchment	River	Lat	Long	Catchment Area (km2)	Cumulative _ Area (km2)
UG13	Great Ruaha	-7.29	36.53	1107	1107
UG12		-7.09	36.31	612	612
UG11		-7.21	36.2	2046	2046
UG24	Great Ruaha	-7.18	35.7	8494	31779
UG10		-6.45	34.71	3038	3038
UG9		-6.9	34.82	1928	1928
UG8		-7.03	34.58	5155	5155
UG7	Njombe	-7.26	34.51	4609	4609
UG5		-8.27	34.18	2149	2149
UG3		-8.43	34.11	1463	1463
UG18		-8.51	33.96	501	501
UG1	Mkoji	-8.55	33.92	1176	1972
UG17	Mkoji	-8.77	33.64	181	181
UG16		-8.67	33.79	277	315
UG15		-8.77	33.84	198	198
UG22		-8.61	33.95	126	126
UG4		-8.62	34.04	500	575
UG21		-8.63	34.32	420	1372
UG20		-8.71	34.17	175	621
UG6		-8.6	34.22	480	2313
UG2	Kioga	-8.68	34.41	1709	2698

4.7 SOIL WATER ASSESSMENT TOOL (SWAT)

4.7.1 SWAT model description

The Soil and Water Assessment model (SWAT) is a physically-based, semi-distributed, continuous model that operates on a daily time step. The model is a result of 30 years of modelling experience of the Agricultural Research Service of the United State Department of Agriculture (Gassman et al., 2007). The model was developed to predict the impact of land management practices on water, sediment and agricultural chemical yields in large complex watersheds with varying soils, land use and management conditions in both gauged and ungauged basins (Arnold et al., 1998; Neitsch et al., 2005). Major model components relevant to this study include weather, hydrology, soil temperature and properties, plant growth and land management. The ArcGIS-SWAT (ArcSWAT) interface tool is designed to generate model inputs from ArcGIS data layers and execute SWAT2009. Only a brief summary of the model is provided in this section. SWAT uses a water balance equation (Equation 4.6) as the driving force behind everything that happens in the watershed (Neitsch et al., 2005).

$$SW_t = SW_0 + \sum_{i=1}^{t} \left(R_{day} - Q_{surf} - E_a - W_{perc} - Q_{gw}\right) \qquad \text{Equation 4.6}$$

Where SW_t and SW_o are, respectively, final and initial soil water content (mm^{-d}); t is the time (day); R_{day} is the precipitation (mm^{-d}); Q_{surf} is the runoff (mm^{-d}); E_a is the evapo-transpiration (mm^{-d}); W_{perc} is the percolation (mm^{-d}); Q_{gw} is the return flow (mm^{-d}).

SWAT partitions the basin into sub-basins which are further divided into hydrologic response units (HRUs) that possess unique land use, management and soil attributes. All model outputs can be evaluated at heterogeneous spatial scales ranging from HRUs to basins. The model requires several parameters which can be specified at the basin, sub-basin or HRU scale to simulate hydrologic and water quality processes. These include weather/climate, soils, ground water, channel, plant water use, plant growth, soil chemistry, and water quality parameters. The model also contains built-in climate, soils, and plant growth databases that can be used as data sources for climate, soil, and plant growth parameters. Land use, soil, topography and climate are the basic land use parameters.

Surface runoff is calculated using the modified Soil Conservation Service (SCS) curve number CN2 (USDA-SCS, 1972) approach when a daily time step is used or the Green and Ampt (1911) infiltration equation when an hourly or sub-daily time step is used. In this study, the SCS curve number method was used. Inter-flow is computed as a function of topographical and soil hydraulic features. Water percolating from the bottom of the soil profile can join the shallow or the deep aquifer. Seepage to the deep aquifer is considered as a loss from the model so only water from the shallow aquifer can produce slow flow in the river or re-enter the soil profile through capillary forces (Figure 4.10). The hydrological processes are individually simulated in each hydrological response unit (HRU) and aggregated at the sub-basin level. The volume of slow, interflow and quick flow generated by HRUs are aggregated per sub-basin and routed through the stream network to the outlet of catchment. In SWAT, water is routed through the channel network using either the variable storage routing or the Muskingum River routing method. More detailed descriptions of the different model components are listed in Arnold et al. (1998) and Neitsch et al. (2005).

Three evaporation estimations are available in SWAT: the Penman–Monteith method (Monteith, 1965), the Priestley–Taylor method (Priestley and Taylor, 1972) and the Hargreaves method (Hargreaves et al., 1985). Penman–Monteith is considered to be the best approach but it has high data demands most of which are unavailable. The Hargreaves method has an advantage over the Penman-Montheith because of its minimum data requirements and can be used even when there are missing data in a time series.

4.7.2 SWAT model parameter description

SWAT model parameters are presented in Table 4.4. These parameters are briefly summarized next, while a detailed description will be found in Nietsch et al. (2005)

- CN2: Initial Soil Conservation Service (SCS runoff curve number for moisture condition II. The SCS curve number is a function of the soil permeability, land use and soil water conditions.
- RCHRG_DP: Deep aquifer percolation fraction. The amount of water that will be diverted from the shallow aquifer to percolate into deep aquifer. The value for RCHRG_DP should be between 0.0 and 1.0.

- GWQMN: Threshold depth of water in the shallow aquifer required for return flow to occur (mm H_20). Ground water to the reach is allowed only if the depth of water in the shallow aquifer is equal or greater than GWQMN. The value for GWQMN varies between 0 and 5000 (mm H_20).
- SOL_K: Saturated hydraulic conductivity (mm h^{-1}). It relates soil water flow rate to hydraulic gradient and is a measure of the ease of water movement through the soil.
- GW_REVAP: This is also known as the ground water re-evaporation coefficient. This process is significant in catchments with the saturated zone not far below the surface or where deep rooted plants are growing. GW_ REVAP vary between 0-1, with low values indicating that the movement of water from the shallow aquifer to the root zone is restricted whereas high values indicate that the rate of transfer from the shallow aquifer to the root zone approaches the rate of potential evapotranspiration.
- SOL_AWC: Available water capacity of the soil layer (mm H_20/mm soil). The plant available water is calculated by subtracting the fraction of water present at permanent wilting point from that present at field capacity. AWC=FC-WP, where AWC is the plant available water content, FC is the water content at field capacity ad WP is he water content at permanent wilting point.
- SOL_Z: Depth for soil surface to bottom layer (mm). Since all the sub-catchments were conceptualized as single layer units SOLZ is equal to SOLZMX, the maximum rooting depth of the soil profile. If SOLZMX is not specified the model assumes the roots can develop throughout the entire depth of the soil profile.
- REVAP_MN: Threshold depth of water in the shallow aquifer for re-evaporation of percolation to the deep aquifer to occur, (mm H_20).
- ALPHA_BF: Baseflow alpha factor (days). The baseflow recession constant is a direct index of ground water flow response to changes in recharge. Values vary from o.1-0.3 for catchments with slow response to recharge, to 0.9-1.0 for catchments with rapid response.
- GW_DELAY: Ground water delay time (days). It represents the lag time between the time the water exists the soil profile and enters the shallow aquifer.

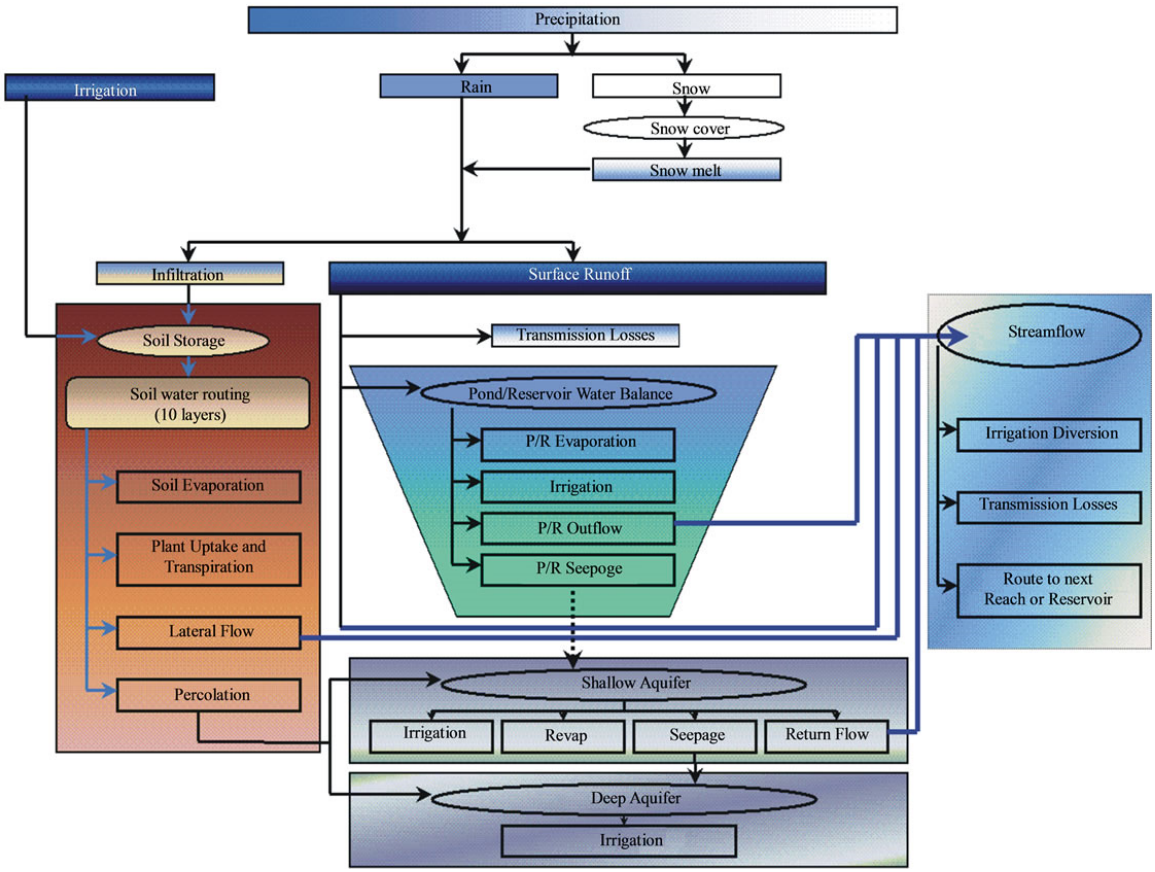


Figure 4.9 Representation of water movement in SWAT (SWAT2005 user manual)

4.7.3 SWAT model set up

Data required for setting up SWAT were collected from different sources as described in Chapter 3. Setting up SWAT for the GRR Sub-basin required topographic, soil, land use, climate and stream flow data for calibration and uncertainty analysis. The SWAT model setup involves two main steps; 1) Catchment delineation and 2) Hydrological response unit (HRU) analysis. The catchment delineation process includes five major steps; DEM set-up, stream definition, outlet and inlet definition, basin outlets selection and definition and calculation of sub-basin parameters. HRUs were defined based on land cover, soil, and slope information and forty (48) sub-basins were delineated for the GRR basin.

Table 4.5 SWAT model parameters and their description

Parameter	Description	Process
CN2	SCS runoff curve number for moisture condition	Runoff
Alpha_Bf	Base flow alpha factor for recession constant (days)	Groundwater
Ch_K2	Channel effective hydraulic conductivity (mm/hr)	Channel
Esco	Soil evaporation compensation factor	Evapotranspiration
Sol_K	Saturated hydraulic conductivity (mm/hr)	Soil
Surlag	Surface runoff lag time (days)	Runoff
Canmx	Maximum canopy storage (mm)	Runoff
Slope	Average slope steepness (m/m)	Geomorphology
Sol_Awc	Available soil water capacity (mm H_20/mm soil)	Soil
Ch_N2	Manning's ''n'' value for the main channel	Channel
Sol_Z	Soil depth (mm)	Soil
Rchrg_Dp	Deep aquifer percolation fraction	Groundwater
Gwqmn	Threshold water depth in the shallow aquifer for flow (mm)	Groundwater
Biomix	Biological mixing efficiency	Soil
Gw_Revap	Groundwater ''revap*'' coefficient	Groundwater
Slsubbsn	Average slope length (m/m)	Geomorphology
Epco	Plant evaporation compensation factor	Evaporation
Revapmn	Threshold water depth in the shallow aquifer for revap (mm)	Groundwater
Gw_Delay	Groundwater delay time (days)	Groundwater
Sol_Alb	Moist soil albedo	Soil
Blai	Leaf area index for crop	Crop

Realistic representation of the catchment processes is impacted by anthropogenic processes, and modelling the GRR basin does not only require the main hydrological processes to be well represented but also needs the main abstractions, reservoir, and wetland impacts to be considered. This means all this information needs to be collected, analyzed and have their spatial and temporal features specified before they can be included in the model. Discretization of the main basin into sub-basin and HRUs requires several input files in ASCII format. Each sub-basin requires six files to specify the sub-basin (.sub), weather (.wgn), water use (.wus), water quality (.swq), and main channel (.rte) and impoundments (.pnd) parameters.

Each HRU requires five files to describe soils (.sol), ground water (.gw), management (.mgt), topography and water cycle (.hru) and chemistry (.chem). Processing and populating all these files into the model is time consuming and subject to a number of errors. The size, scale, and number of sub-basins used in SWAT may impact the modelling process and model simulation results (Jha et al., 2004). On the other hand, calibration and uncertainty analysis involves a large number of runs in order to examine the performance of different parameter sets in representing hydrological processes within the basin. Consequently, the propagation of errors coming from the model parameterization into the hydrological model leads to large uncertainties. Based on these reasons only the Little Ruaha drainage system was used for the hydrological simulations using the SWAT model. The basin was chosen because it has a reasonable amount and quality of data. The location of the selected sub-basin is shown in Figure 4.10. In this thesis sub-basins were characterized by the dominant land use and soil type combination instead of considering several HRU for each sub-basin because there are not enough data to populate the large number of HRUs. This helped in reducing the number of parameters to be calibrated and reduced the time needed for processing large amounts of data.

4.7.4 Sensitivity analysis

Sensitivity analysis allows for the identification of model parameters that exert a strong influence on the model output, thus largely controlling the behaviour of the simulation process. In this study sensitivity analysis was carried out using the Latin Hypercube One-factor At a Time (LH-OAT) algorithm (Morris, 1991; van Griensven et al., 2006; van Griensven et al., 2008). The Sensitivity analysis minimizes the number of parameters to be used in the calibration step. The Latin Hypercube simulation is based on a Monte Carlo approach with a stratified sampling. The results of the sensitivity analysis are parameters arranged in ranks, where the parameter with a maximum effect obtains rank 1, and parameter with a minimum effect obtains the rank which corresponds to the number of all analyzed parameters. The parameter that has a global rank 1 is categorized as "very important", rank 2-7 as "important", rank 8-27 "slightly important" and rank 28 as "not important" (van Griensven et al., 2006).

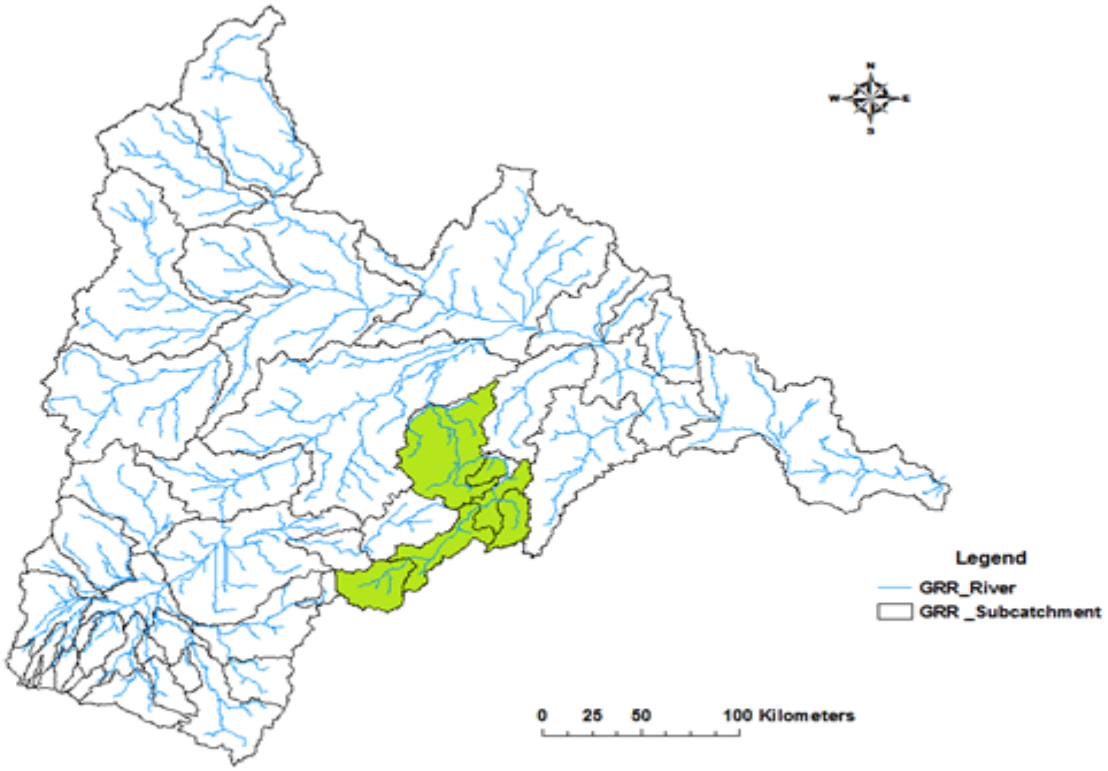


Figure 4.10 Little Ruaha drainage system (green)

4.7.5 Uncertainty analysis

Sequential Uncertainty Fitting version 2 (SUFI 2) framework is used for the implementation of uncertainty analysis in this study. The framework was selected because it takes fewer runs in comparison to other calibration procedures tailored for SWAT. According to Abbaspour et al. (2005; 2007; 2008) SUFI-2 parameter uncertainty accounts for all sources of uncertainties such as uncertainty in input data, model structure and parameters. All uncertainties are quantified by a measure referred to as the P-factor, which is the percentage of measured data bracketed by the 95% prediction uncertainty (95PPU) and R-factor which is the measure of the width of the uncertainty band.

The concept behind the uncertainty analysis of the SUFI-2 algorithm is illustrated graphically in Figure 4.11. The diagram illustrates that a single parameter value (black dot) leads to a single model response (Figure 4.11a), while the propagation of the uncertainty in a parameter (shown

by a line) leads to the 95PPU illustrated by the shaded region in Figure (4.11b). As parameter uncertainty increases, the output uncertainty also increases (Figure 4.11c). SUFI-2 normally begin with a large parameter uncertainty (within a physically meaningful range) in order to make sure that the observed data falls within the 95PPU, then decreases this uncertainty in steps while monitoring the P-factor and the R-factor. If the initial parameter ranges are equal to the maximum physically meaningful ranges and still cannot find a 95PPU that brackets any or most of the data (Figure 4.11d), then the the conceptual model needs to be re-examined (Abbaspour et at.,2007).In each step, initial parameter ranges are updated by calculating the sensitivity matrix, an equivalent of a Hessian matrix, followed by the calculation of the covariance matrix, 95% confidence intervals of the parameters, and correlation matrix. Parameters are then updated in such a way that the new ranges are smaller than the previous ranges, and are centered on the best simulation. More details on SUFI-2 and its algorithm can be found in Abbaspour et al. (2005; 2007). The uncertainty analysis in this study was implemented in the two stages;

Assigning initial parameter ranges: The complete physical range of each parameter was used to explore the surface response using Latin Hypercube sampling and to select the initial range for each parameter.

Derivation of a reduced parameter range and predictive uncertainty: The procedure identifies a range for each parameter in such a way that upon propagation:

- The 95% prediction uncertainty (95PPU) between the 2.5th and 97.5th percentiles contains (brackets) a predefined percentage of the measured data, and
- The average distance between the 2.5th and 97.5th prediction percentiles is less than the standard deviation of the measured data (Abbaspour 2005; 2007).

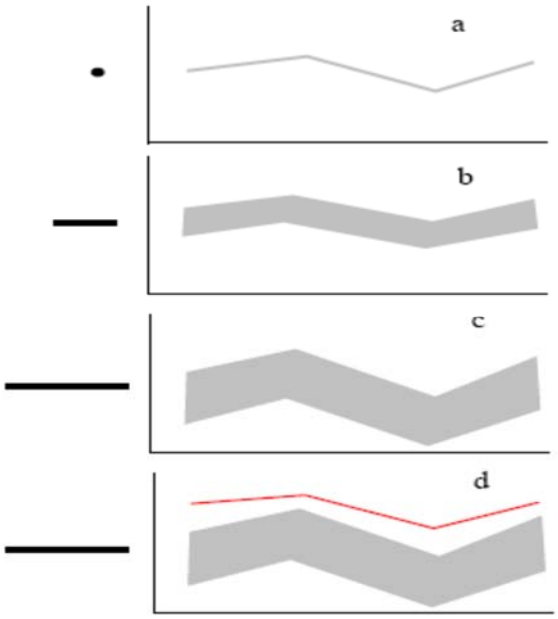


Figure 4.11 A conceptual illustration of the relationship between parameter uncertainty and prediction uncertainty (Abbaspour et al., 2007). *A single parameter value (black dot) leads to a single model response (Figure 4.11a), while the propagation of the uncertainty in a parameter leads to the 95PPU (Figure 4.11b). As parameter uncertainty increases, the output uncertainty also increases (Figure 4.11c). If the initial parameter ranges presents the physically meaningful ranges and still cannot find a 95PPU that brackets any or most of the data (Figure 4.11d).*

The model performance was assessed based on the two conditions being fulfilled and a good agreement between the simulated and the observed data for a calibration and validation period. In theory, the P factor values range between 0 and 100%, while the R-factor ranges between 0 and infinity (Abbaspour et al., 2007). A P-factor of 1 and R-factor of zero is a simulation that exactly corresponds to measured data. The degree of departure from these numbers is used to judge the strength of calibration. It is possible to achieve a good P-factor at the expense of a larger R-factor; therefore there should be a balance between the two factors (Abbaspour et al., 2007). Other performance measures used are the R^2 and Nash-Sutcliffe (CE) coefficient.

4.8 Concluding remarks

This chapter described the two different models (GW_PITMAN and SWAT) that were used to simulate flows in the GRR sub-basin models. The steps of the setups of each hydrological component were discussed to draw attention to the differences between the spatial representation of input data and parameterisation between the models. Similar spatial data sets and delineated catchments are used in both models. The calibration and uncertainty analyses of the two models are presented in Chapter 6 (GW_PITMAN) and Chapter 7 (SWAT).

5 SIMILARITY ANALYSIS AND DEVELOPMENT OF REGIONAL CONSTRAINTS

5.1 Introduction

Geomorphological and climatic characteristics play an important role in runoff generation and hydrological response of the basin. The relationship that exists between physical basin characteristics and hydrological indices is important in hydrological modelling because information generated from this relationship is important in understanding and estimating runoff generation in areas where data are limited. Even in areas where historical observed data exist, the quality of these data could be questionable. Hence, spatial datasets (land cover, elevation, slope, and soils) are used to generate landscape characteristics which are then correlated with climate and hydrological signature for the purpose of understanding runoff generation process in the basin and guide the regionalization process.

The development of an uncertainty framework that can be applied to both gauged and ungauged basins is a step towards development of the practical implementation of PUB principles (Hughes, 2013b). Uncertainty in model parameter estimation can be constrained using local observed data or regional hydrological indices of catchment response behavior (Kapangaziwiri *et al.,* 2012; Westerberg *et al.,* 2013). Developing suitable constraints is an important aspect in any practical uncertainty assessment in hydrology since constraints help in reducing uncertainty (within physically plausible boundaries), as well as ensuring that the uncertainty is realistically represented (Hughes, 2013b). Constraints can be derived from physical basin characteristics or hydro-meteorological data (Yadav *et al.,* 2007). The uncertainty approach adopted in this study provides for the generation of ensemble predictions in both gauged and ungauged basins and the use of regionalized indices extracted from observed stream flow data to constrain ensembles of model predictions. The objective is to reduce predictive uncertainty by constraining model output to expected catchment behaviour while maintaining consistent predictions (Yadav *et al.,* 2007; Kapangaziwiri, 2010).

5.2 Similarity analysis based on climate and landscape signatures

Classification of catchments was based on land use, soil types and climate conditions. The idea was to cluster catchments into hydrologically homogeneous regions. The physical characteristics were calculated for each basin using GIS information to obtain the average values of land cover, soil texture, and slope. The available spatial datasets are described in Chapter 3. Land cover and soils are presented in terms of the percentage area of each category within in a sub-basin. Although there are different ways in which average slope can be defined in a catchment, for this analysis the arithmetic mean of the slope in all cells of the sub-basin was used. Estimated values of the physical basin characteristics of the 48 sub-basins are presented in Appendix1 (A, B and C). Table 5.1 gives a summary of the estimated physical basin characteristics and climate. The climatic characteristics were calculated from rainfall and evaporation data. The aridity index was also calculated from the mean annual values of rainfall and evaporation.

Agglomerative hierarchical clustering (AHC) analysis

In the literature there are three common clustering algorithms: agglomerative hierarchical clustering, k-means clustering, and fuzzy partition clustering. In this study agglomerative cluster analysis was used to identify relatively homogenous groups of variables based on selected catchment functions. The Euclidean distance based on Ward's method (Ward, 1963) was used as this is the most widely used agglomerative clustering method in the regionalisation context (Tshimanga, 2012; Westerberg *et al.,* 2013). This method generated spherical clusters of approximately the same size with the aim of linking various objects into clusters using a distance or similarity measure. The method starts with *n* groups, each of which contains one case as there are as many clusters as cases. Clusters are then combined, sequentially reducing the number of clusters at each step until only one cluster is left. This method uses an analysis of variance approach to evaluate the distances between clusters and this makes it a very efficient method. Cluster membership is assessed by calculating the total sum of squared deviations from the mean of a cluster. The criterion for fusion is that it should produce the smallest possible increase in the error sum of squares. This cluster analysis was applied to 48 sub-basins using landscape and climate indices (Table 5.1) and 4 different classes of varying size were identified (Table 5.2).

Table 5.1 Estimated physical basin characteristics for the 48 catchments

Index	Description	Minimum	Maximum	Mean
Catch area	Catchment area	38	7315	1682
MAE	Mean Ann Evaporation	1380.1	1930.7	1669.3
MAP	Mean Ann Precipitation	640.2	1668.1	1153.6
AI	Aridity Index	0.9	2.8	1.5
CRDY	Dryland_cropland and pasture	0.0	87.9	30.0
CRWO	Cropland/woodland mosaic	0.2	97.5	28.8
GRASSLAND	Grassland	0.0	38.4	10.8
SHRUBLAND	Shrubland	0.0	13.1	3.6
SAVANNA	Savanna	0.0	40.6	15.9
FODB	Deciduous broadleaf forest	0.0	34.9	4.4
FOEB	Evergreen broadleaf forest	0.0	16.4	1.0
FOMI	Mixed forest	0.0	57.5	2.4
Elev	Elevation (Mean) m.a.s.l	692.2	2085.3	1369.1
Elev_Min	Elevation (Min) m.a.s.l	298.2	1714.0	1021.6
Elev_Max	Elevation (Max) m.a.s.l	1047.5	2950.0	1983.7
Slope	Slope	0.5	23.8	7.9
Loam	Loam	0.0	100.0	38.2
Sandy_Clay_Loam	Sandy_Clay_Loam	0.0	91.9	24.7
Sandy_Loam	Sandy_Loam	0.0	89.1	29.0
Clay_Loam	Clay_Loam	0.0	12.9	0.6
Clay	Clay	0.0	50.4	9.3

Similarity percentage analysis (SIMPER) was used to identify indices responsible for similarity within the four regions of homogeneous physical basin characteristics. SIMPER is a method of identifying similar and dissimilar characteristics between features and the approach has become widely used in aquatic ecology (Clarke, 1993). The results from the SIMPER analysis (Table 5.3) show the overall contribution of the main variables to the similarity within the four regions. The analysis shows that the variables MAP, MAE, Elev (Maximum and Average) account for the similarity within the identified four regions. The similarity within the regions is explained at more than 80%. It is important to note that, the cluster analysis approach is

influenced by subjectivity. There is no single correct solution to the unsupervised classification method; however the cluster analysis results give an indication of similarities in sub-basins. Both physical basin characteristics and climatic data provided information for the classification, however topography and climate-controlled indices dominated the classification, and this highlights the dominant role climate has in catchment runoff generation process. Moreover, the four regions formed exhibit a strong spatial pattern (Table 5.3).

Table 5.2 Homogenous regions of the physical basin properties

Class	1	2	3	4
Catchments	12	12	13	11
Within-class variance	43142.1	60742.4	89671.9	46320. 7
Minimum distance to centroid	79.7	95.70	135.7	62.3
Average distance to centroid	178.8	220.37	260.6	188.4
Maximum distance to centroid	330.0	424.4	609.3	335.2
List of sub-basins contained within each group				
	1ka10	1ka11	1ka41	1ka71a
	1ka12	1ka8a	1ka5a	UG4
	1ka56	1ka15	1ka61	UG21
	1ka37	1ka32a	1ka27	UG18
	1ka33	1ka51	1ka42	UG6
	1ka2a	1ka9	UG10	UG20
	1ka59	1ka16a	UG12	UG1
	1ka21	1ka22	UG13	UG3
	1ka31	1ka50a	UG14	UG22
	UG2	1ka7a	UG7	UG16
	UG11	1ka39	UG8	UG15
	UG5	UG17	UG9	
			UG24	

Table 5.3 Percentage contribution to the overall similarity within the groups (Cut-off for low contributions was set at 80%)

Indices	Region I	Region II	Region III	Region IV
ElevMax	27.14	33.40	33.43	26.91
MAE	20.69	19.33	24.13	19.06
Elev	19.10	16.78	18.95	17.72
MAP	18.09	16.18	12.28	17.10

The groups identified in Table 5.2 have a great deal of internal variability in their flow characteristics and are intuitively not very useful to establish sub-basin groupings for hydrological modelling. For example, one of the Usangu Plains wetland sub-basin (1ka71) has been grouped with a number of the ungauged foothills sub-basins. While the first two groups are made up of most of the wetter sub-basins no clear distinction is made between those with different flow characteristics. While the cluster analysis provides some guidelines, it was considered necessary to introduce additional 'soft', qualitative data on expected sub-basin runoff characteristics, as well as using more quantitative detail from the observed flow data to develop constraints on hydrological behaviour for regional groups.

5.3 Constraints development

This section discusses the methods that were used to quantify the constraints and to determine sub-basin groups so that constraint information quantified from the gauged sub-basins could be extrapolated to the ungauged sub-basins. The approach has used a combination of quantitative analysis together with some qualitative, and largely subjective assessments, to establish constraint uncertainty ranges and extrapolate to the ungauged basins. There is simply not enough 'hard' data to establish the regionalised constraints in a fully quantitative and objective manner.

5.3.1 Data quality and screening of inconsistency

Discharge data used in this study are presented in Chapter 3 (Table 3.4). Out of the total of 48 sub-basins, 22 are ungauged and the rest are gauged but with a lot of missing values in the time series data and some possible problems with the rating curves (Table 3.4). In addition, there are possible impacts of abstraction and land use change that have potentially non-stationary effects on the extent to which the observed time series can be considered to represent natural flow conditions. Of the 26 gauged sub-basins, only 16 represent headwater conditions, the others represent much more heterogeneous climate and runoff characteristics. All of these factors affect the value of the observed flow data for developing regional constraints on sub-basin hydrological response that can be used in uncertain hydrological models (Westerberg et al., 2013). Unfortunately, within the GRR basin, as in many other basins of Africa, there are few explicit sources of information with which to quantify the sources of uncertainty in the observed flow data.

In this analysis, the quality of observed monthly and daily discharge data was checked through graphical testing of the time series (plotting a time series) to identify inconsistencies and outliers. Judgements based on graphical displays are often indicative rather than prescriptive, therefore expert's judgement on the part of the data was still required. Presence of extreme events in a time series was confirmed by comparing a particular station with neighboring stations and correlation with rainfall data. Un-explainable extreme events were removed from the time series. Quality control was also done on the rating curves. The official rating curves and stage discharge measurements from the Rufiji Basin Water Office for 24 stations were used to identify possible problems with uncertainty in the ratings (high degrees of scatter in the stage-discharge relationships) and changes in the rating relationship, probably associated with unstable channel beds. A further problem that was identified at some of the stations was that some of the higher discharges in the recorded time series exceeded the limits of the rating curves, suggesting uncertain extrapolation. Table 5.4 provides a summary of identified problems related to the discharge-stage scatter plots, while Appendix C shows examples of poor rating curves. No attempts were made to explicitly quantify the rating curve uncertainty, but the likely quality of the rating information was identified in a qualitative manner (Table 5.4). The rating curve information was also used to identify possible problems in the recorded

time series data so that these could be excluded from the analysis of observed constraints and not used for comparisons between simulated and observed flows. The main focus was therefore on the identification of errors in the observed data so as to preserve the hydrological integrity of the observations that were used as far as possible. Some catchments were assumed to be influenced by upstream land use changes and abstractions (SMUWC, 2001) that were expected to decrease low flows and affect the observed quantification of the low flow constraint and possibly even the percentage of zero flows. These assumed influences on the observed data were taken into consideration when quantifying the local and regional constraint bounds.

The aridity index using the ratio of mean annual potential evaporation divided by mean annual rainfall (PE/P: Budyko, 1974) was used to assist in the regionalisation of the constraints using the observed discharge data. The aridity index plotted against the runoff ratio (Q/P: essentially a Budyko relation) was used to identify stations with inconsistent hydrological data, in a similar approach that was used by Westerberg *et al.* (2013). The points were plotted for all catchments with the headwater catchments identified by the quality of the rating curve data (Poor, Moderate or Good) (Figure 5.1). Four headwater catchments were identified to have somewhat anomalous runoff ratios and this result prompted further checks of the daily time series and rating curves. Catchment 1ka56 had extreme high flows at the beginning of the period and this had a substantial effect on the runoff ratio. The rating curve showed inconsistences between the two periods (pre- and post- 1996); therefore the period prior to 1996 was removed from the record. Catchment 1ka41 is a headwater catchment but is in a semi-arid region, so it was expected to have behaviour different from other headwater sub-basins. Catchment 1ka16a had one extreme event that was very high compared to the rest in the time series and neighbouring stations, and these data were removed from the record. Extreme high flows were also observed in sub-basin 1ka10 which affected the runoff ratio and the shape of the flow duration curve and consequently these records were also deleted. 1ka37 is geologically different from neighbouring catchments and is dominated by crystalline limestone rocks resulting in high base flows and a relatively high runoff ratio.

Table 5.4 Rating curve validation and quality check.

Station	Quality of rating curve	Description	Comment
1ka10	Poor	Large scatter at moderate to high flows. Daily flows show problems prior to 1960.	
1ka11	Poor	Unreliable for high flows, rating curve points show an increase in rating curve slope above 60 $m^3 s^{-1}$.	
1ka12	Poor	Recorded flows are well beyond the highest gauged flows	Unreliable, although the hydrographs appear to be reasonable. Extreme events were removed from the recorded flows.
1ka56	Poor	Early high flows exceed the rating curve	Extreme events were removed from the records. The data prior to 1966 were also identified as being unreliable and were removed.
1ka8a	Poor	Unreliable, lots of scatter in early gauging's	
1ka15	Moderate		
1ka32a	Moderate		
1ka37	Moderate	Very high base flow	
1ka51	Moderate	Lots of scatter in the observed gauging points.	
1ka9	Moderate	Lots of scatter.	
1ka16a	Good		Removed the period from 1989 onwards
1ka22	Good		Removed 2009 records
1ka41	Good		
1ka50a	Good		
1ka7a	Good		
1ka33	Poor	Potential problems with low flows	Removed 1980's onwards
1ka5a	Poor		
1ka2a	Moderate		
1ka59	Moderate		
1ka61	Moderate		
1ka21	Good		
1ka27	Good		
1ka39	Good		
1ka42	Good		

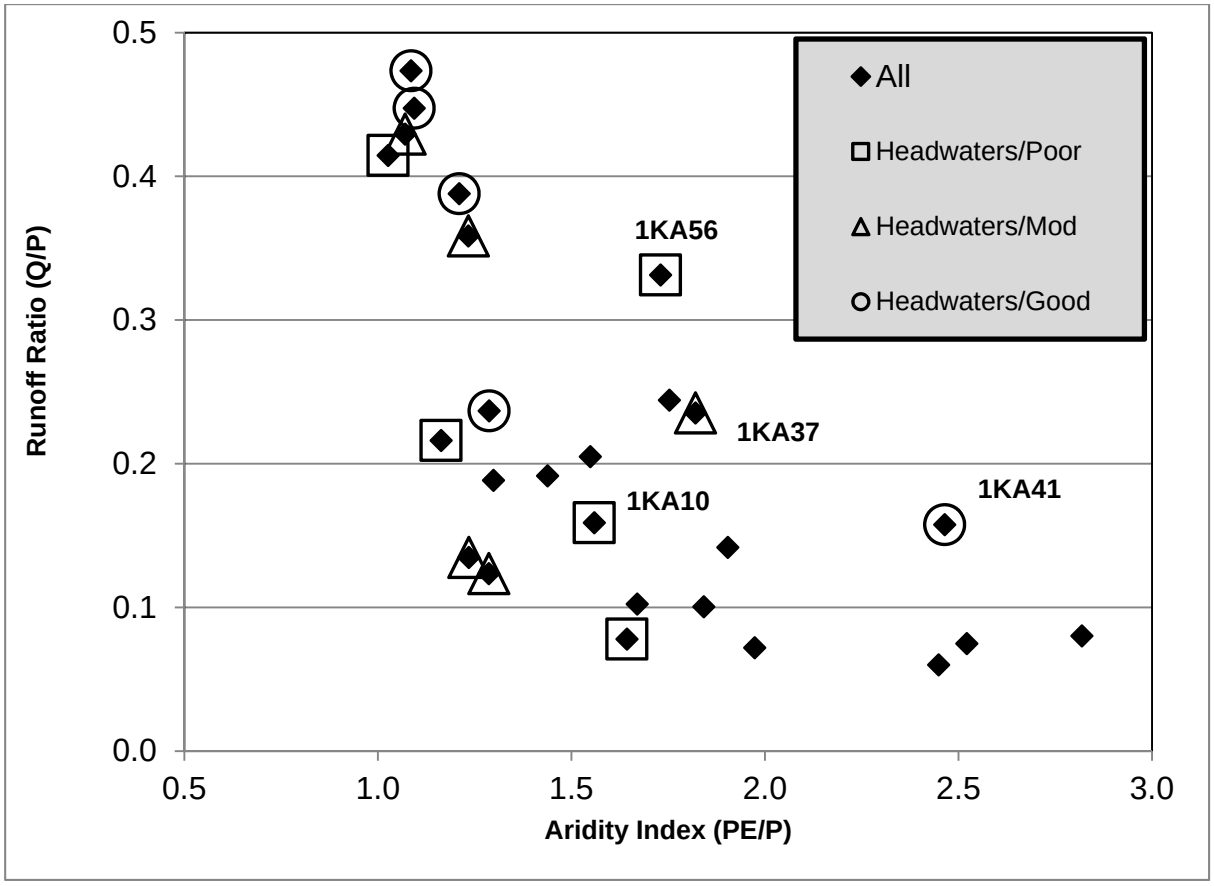


Figure 5.1 Budyko curve showing the relationship between the aridity index and the runoff ratio after quality check

5.3.2 Regional constraints

Six constraints have been considered for this study and they are mean monthly discharge (MMQ in $m^3 * 10^6$), mean monthly ground water recharge (mm), monthly flow duration curve percentiles (Q10, Q50,and Q90) extracted as fractions of MMQ and the percentage of zero flows (% zero flow). The constraints are quantified as ranges, defined by minimum and maximum values representing the uncertainty due to a lack of observed data in the ungauged sub-basins, or because of the assumed unreliability of the gauged data. All six constraints were extracted from the observed discharge data and are used to quantify uncertainty ranges for all the sub-basins of the GRR basin. It is important to emphasise that the constraints refer to the incremental runoff from the sub-basin area and not to the cumulative stream flow characteristics at the outlet of the sub-basin. To account for the sources of uncertainty percentage errors were subjectively assumed for the observed data and rating curves. A 15% error was applied to the MMQ, while 20% was applied to the Q10 and Q90 constraints as the

measurement of these high and low flows are associated with large errors. For the Q50 constraints a percentage error of 15% was applied (Di Baldassarre, G and Montanari, A (2009). There are no direct estimates of ground water recharge and therefore the quantification of this constraint was highly uncertain and largely based on a combination of the Q90 (low flow) signals in the observed data together with the topography and geology. The % zero flow constraint was similarly difficult to quantify.

The GRR sub-basins were divided into six (6) regional groups based on similarities in their hydrological response characteristics (Table 5.5) as well as their physical properties (topography, geology, vegetation, etc.). In certain cases the constraints varied within the groups, as well as between the groups so that localised differences could be accounted for, but sub-basins sharing similar characteristics were assigned comparable constraint values. The complete list of quantified constraint ranges is given in Appendix D but the groups and their main characteristics are summarised in Table 5.5.

Table 5.5 Regional groups and their hydrological response characteristics

Group	Catchments	Characteristics
1	1ka50a, 1ka51,1ka16a,1ka7a,1ka8a, 1ka9,UG17,UG15,UG16	South west headwaters, all with assumed high rainfall and runoff , relatively high recharge and base flow
2	1ka10,1ka11,1ka12,1ka56,UG2,UG6, UG20,UG22,UG21,UG4,UG1	Foothills of the south west, as well as some of the larger gauged catchments extending from the headwater areas. Lower rainfall and runoff ratio than Group 1.
3	1ka32a,1ka15,1ka39,1ka22,1ka21, 1ka2a,1ka33, 1ka31,1ka37	A mixed group representing the central headwater catchments and some in less elevated areas. Some of these are dominated by limestone geology with high recharge and base flow contributions
4	1ka41,1ka42,UG3,UG5,UG7,UG8, UG9,UG10,UG18	Arid western catchments with much lower rainfall, high evapotranspiration and low runoff ratios with some zero flow periods.
5	1ka27,1ka71,1ka59,UG24	Central catchments with low topography, low rainfall and assumed very low incremental contributions to flow
6	1ka5a,1ka61,UG11,UG11,UG12,UG13,UG14	Downstream catchments with similar incremental flow contributions as Group 5

5.3.2.1 Mean Monthly Flow (MMQ)

Group One consists of the majority of the headwater sub-basins covering the south west part of the basin (Table 5.5). These sub-basins fall within the Rungwe volcanic and Gondwana geomorphological zone (Chapter 3). The mean annual precipitation is around 1 200 mm^{-1} y^{-1} and can be higher in the elevated areas, reflecting the high runoff ratios and mean monthly flows. The upper and lower bounds of the constraints are 0.26 to 0.56 × 10^6 m^3 (Figure 5.2). The width of the block covering group one is relatively narrow, indicating that most stations used in the regionalization share similar aridity characteristics (Figure 5.2).

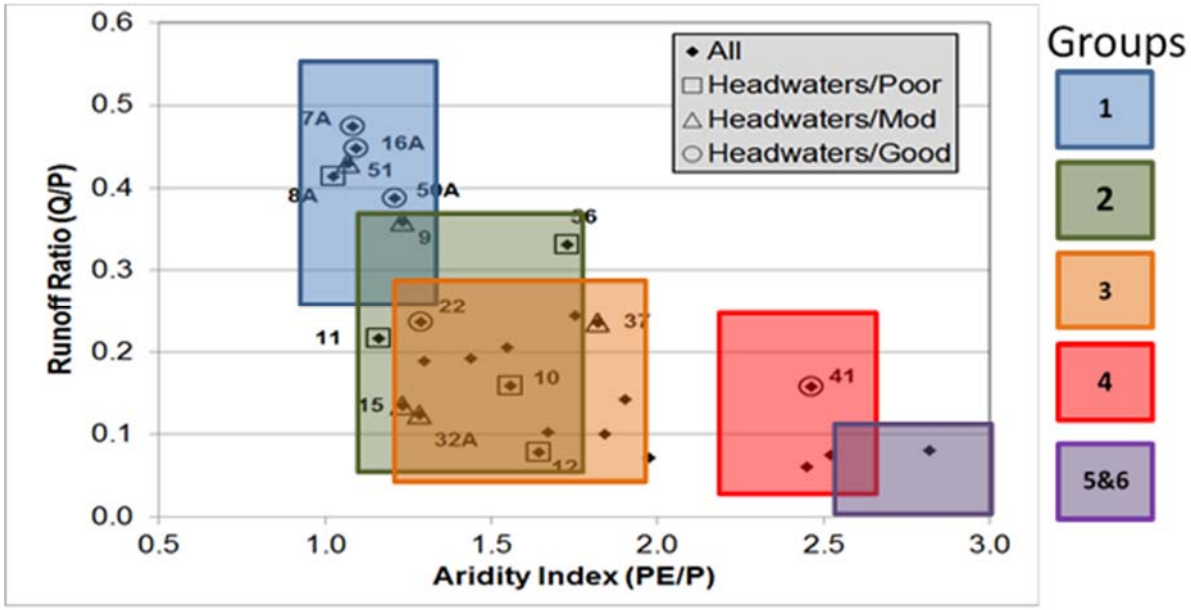


Figure 5.2 Observed stream flow constraints and regional constraint ranges for runoff ratio (used to determine the MMQ constraint), the number labels represent the headwater gauges but excluding the '1ka' prefix

Group Two (1ka11, 1ka10, 1ka12 and 1ka56) includes some upstream sub-basins, most of which fall within the African and Post African geomorphological zone. The group also includes the largely ungauged intermediate sub-catchments, which partly fall within the Alluvial fan zones. These zones start at the foot of the steep escarpments on the southern and western sides of the plains, and spread out toward the alluvial fans. The upper and lower bounds of the

constraints range are 0.06 to 0.37 × 10^6 m^3 (Figure 5.2). These values reflect reduced mean monthly flow due to less runoff from the foothills of the mountains. The width of the block covering group two is wide reflecting a substantial amount of variability in aridity characteristics within this group.

Group Three covers sub-basins in the Little Ruaha, Ndembera and Lukosi River drainage systems (1ka15, 1ka32a, 1ka39, 1ka22, 1ka21, 1ka2a, 1ka31 and 1ka37). The Little Ruaha and Ndembera fall within the African geomorphological zone while the Lukosi falls under the Gondwana zone. The upper and lower bounds of the constraints range are 0.04 to 0.28 × 10^6 m^3 (Figure 5.2). The width of the block covering group three is relatively wide because it consists of both upstream and downstream sub-basins with different climatic characteristics. Humid forest remnants cover the upper parts of the three Rivers. Accacia shrub land is more typical in the lower drier areas (1ka31 and 1ka33).

Group Four catchments fall within the arid part of the GRR basin and is drained by the Kisigo River and other tributaries. Due to the semi-arid nature of the catchment, significant rain only occurs during a small part of the year and stream flow is ephemeral. The upper and lower bounds of the MMQ constraint range are 0.03 and 0.24 × 10^6 m^3 (Figure 5.2). The upper value of the range has been set quite high, largely because of the unknown runoff characteristics of the western tributaries of the Kisigo River.

Groups Five and Six have been assigned the same constraints (Table 5.2) as group four. Group five is semi-arid and dominated by relatively low rainfall and high potential evaporation. The upper and lower bounds of the constraints have been set at 0.00 to 0.11 × 10^6 m^3 (Figure 5.2). The real lower value is unlikely to be zero but a realistic minimum is difficult to determine. The low volumes of incremental runoff are a reflection of the aridity and the low topography in these downstream sub-basins.

5.3.2.2 Flow duration curve indices (Q10, Q50 and Q90)

The shape of flow duration curves (FDCs) are related to the interactions of climate, catchment size and morphology, vegetation cover, and the properties of the subsurface domain, which

together control the various runoff components (Castellarin et al., 2012). The shape of FDCs is therefore governed by both precipitation and evapotranspiration variability and how water moves through the catchment (Castellarin et al., 2012). According to Yokoo and Sivapalan (2011) the FDC can be partitioned into three distinct parts, each of which is governed by different mechanisms or process controls:

(i) the upper part, which represent high flows, is governed by flood processes for which the dominant control is the interaction of extreme rainfall and fast runoff processes;

(ii) the middle part, relates to the mean runoff and its seasonality, for which the dominant control is the competition and seasonal interaction between available water, energy and storage, and

(iii) the lower part, is governed by base flow recession behaviour over dry periods for which the dominant control is the competition between deep drainage and riparian zone evaporation.

The objective of the observed flow data interpretation is to determine realistic uncertainty ranges for the natural flow FDC at the three percentage points. The FDC percentage points were calculated directly from the observed monthly discharge data and expressed as fractions of MMQ. Standardising the volumes at different percentage points by the MMQ is necessary for the constraint values to be used in the model, as the simulated MMQ will vary with each ensemble simulation. In the following sections, the observed Q10, Q50 and Q90 ratios are plotted against the aridity index (Figures 5.3 to 5.5), while it is recognised that other factors (such as topography, vegetation cover, soil characteristics and geology) will also play a major role in determining spatial variations. Only the data for the gauged headwater sub-basins are plotted as the downstream gauges will reflect multiple upstream influences. As noted previously, the upper and lower bounds for the three constraints were partly based on largely subjective assumptions of approximately 20% error in the high (Q10) and low (Q90) flow observations and 15% in the median (Q50) flow observations.

FDC Q10 (fraction of MMQ)

The upper and lower bounds for ***group one*** constraints are from 2.0 to 3.0 × 10^6 m^3 (Figure 5.3). The uncertainty bound in this group is relatively narrow despite the uncertainties in the high flow observations associated with some of the inadequacies in the rating curves. It is quite possible that many of these inadequacies manifest themselves at higher flows than the 10^{th} percentile. ***Group two*** constraints range from 1.5 to 3.5 × 10^6 m^3, reflecting a wide uncertainty bound (Figure 5.3) associated with a high level of within-group variability, However, a lower range of values was used for some of the gauged sub-basins within this group. The constraints for ***group three*** range from 1.3 to 3.0 × 10^6 m^3 and there is a large overlap between groups two and three, which is almost certainly due to the groups sharing similar climatic and topographical characteristics. ***Group four*** constraints have been set at 1.5 to 4.0 × 10^6 m^3. The semi-arid nature of the catchments together with little information on the catchment characteristics and only a single headwater gauge reflects the large amount of uncertainty in this group. A substantial amount of uncertainty is also expected in ***groups five and six*** for which there are no headwater gauges. The wide ranges of uncertainty for groups 2 to 6 are a reflection of a substantial amount of uncertainty in the high flow response characteristics of the majority of the basin and how these will be simulated by the model. The ranges were set to be quite wide to avoid a possible problem with no behavioural simulations being found in step 1 of the uncertainty simulation method.

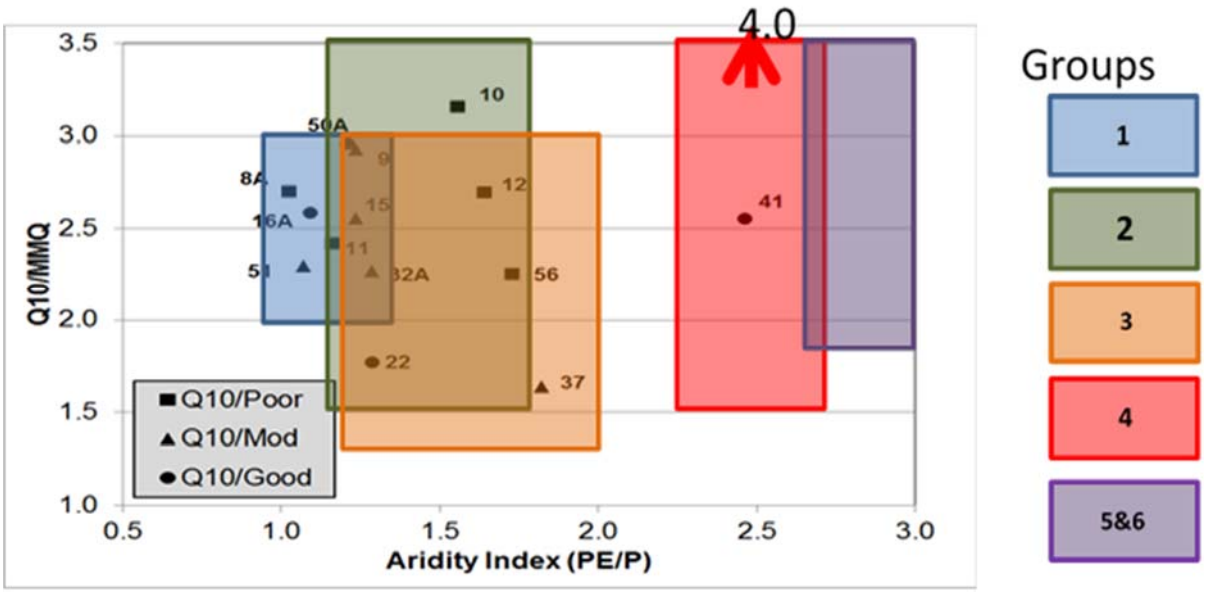


Figure 5.3 Observed stream flow constraints and regional constraint ranges for Q10 (relative to MMQ)

FDC Q50 (Fraction of MMQ)

The upper and lower bounds of the constraints range from 0.25 to 0.75 × 10^6 m^3 for ***group one***, and 0.1 to 0.7 × 10^6 m^3 for ***group two***. Most of the rivers in these groups are expected to have relatively high interflow and base flow runoff volumes that will impact on median flows as well as low flows (see next section). Group two includes foothill sub-basins where the interflow components are expected to be lower and this is reflected in the somewhat lower values for this constraint. ***Group three*** includes both headwater and downstream sub-basins that are characterised by moderate to high base flows associated with limestone geology. However, there is not enough information about the variation of the base flow contributions between the upstream and downstream sub-basins, hence the relatively large uncertainty. The constraint ranges from 0.3 to 1.1 × 10^6 m^3. ***Group four*** constraints range from 0.0 to 0.6 × 10^6 m^3, whereas ***groups five and six*** range from 0.0 to 0.7 × 10^6 m^3. Both group five and six have large uncertainties due to the lack of suitable gauged sub-areas to define the shape of the FDC for incremental flows from the semi-arid downstream sub-areas.

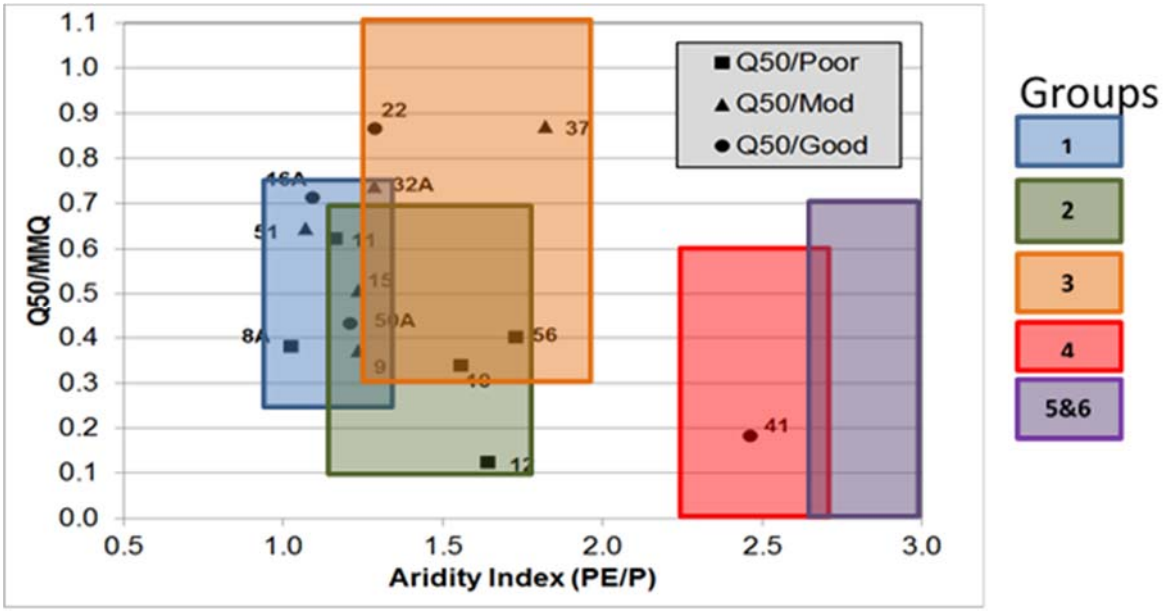


Figure 5.4 Observed stream flow constraints and regional constraint ranges for Q50 (relative to MMQ)

FDC Q90 (Fraction of MMQ)

Low flow (Q90) estimations are subject to the availability of data and understanding of the development and water abstraction impacts within the sub-basins. Some of the low flow uncertainties in the observed stream flow are related to the impacts of upstream land use change and abstractions, therefore prior understanding of the activities in the catchment was critical in developing these constraints. Sub-basins 1ka56, 1ka10 and 1ka12 are characterised by poor rating curves and overall poor quality of the observations, therefore large uncertainties were expected for these observation. Sub-basin 1ka32a is influenced by afforestation (pine and eucalyptus plantations), while 1ka16a is highly impacted by development impacts (abstractions) for both irrigation and domestic water uses.

Groups one (headwater) and ***two*** (foothills) have almost similar base flows, with constraint ranges between 0.02 to 0.25 and 0.0 to 0.25 × 10^6 m^3 respectively. The lower values are almost certainly unrealistic (notably for group 2) from the perspective of the natural flow regimes and were increased for most individual sub-basins. The very low values that were obtained from the observed data analysis are therefore assumed to be related to upstream abstractions. The

group three constraints range between 0.01 and 0.6 × 10^6 m^3, which reflects substantial uncertainty that is related to the diverse, but largely unknown, base flow conditions amongst the headwater and downstream sub-basins of this group*. Groups four to six* constraints range between 0.0 and 0.2 × 10^6 m^3. As with the other FDC constraints for these groups, setting realistic values is very difficult. It is likely that the majority of these sub-basins will experience zero flows at the 10^{th} percentile and therefore the upper value of this constraint is not very important.

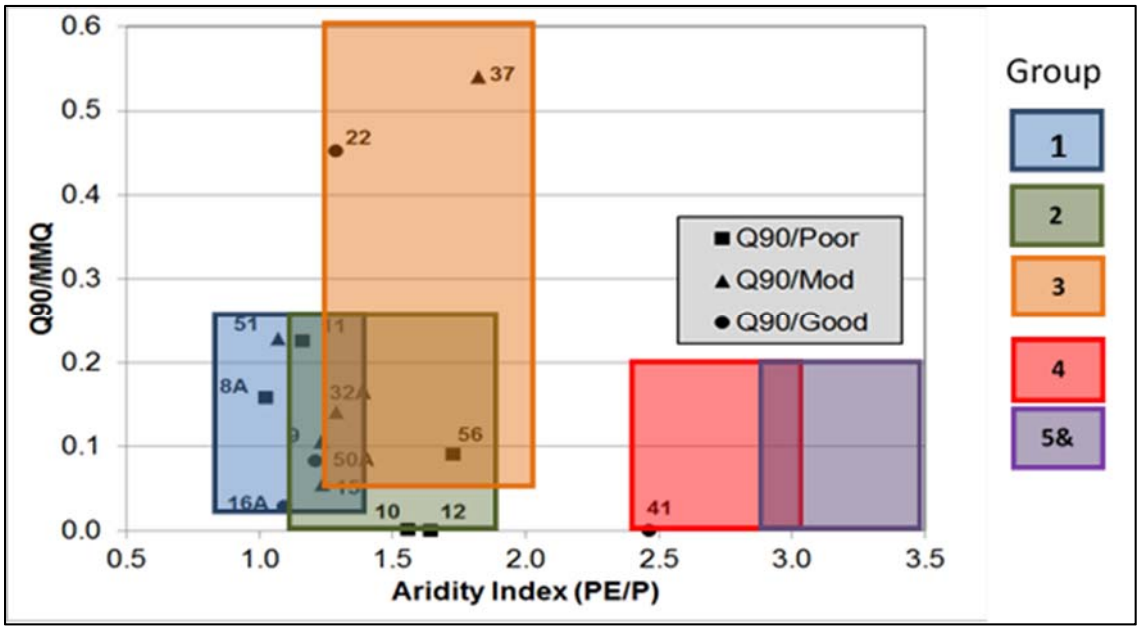


Figure 5.5 Observed stream flow constraints and regional constraints ranges for Q90 (relative to MMQ),

5.3.2.3 Mean monthly groundwater recharge (MMR)

Due to paucity of groundwater data, recharge in the GRR Sub-basin is hard to estimate. Most studies in the region have failed to do a detailed analysis of groundwater and have only made general assessments of the hydrogeology. Even when a few estimates are available, most of them are point observations that are normally associated with high uncertainty when extrapolating them to the basin scale. Given little confidence in the recharge estimates generated by CCKK (1982) on a few sites in the basin, some trial calibration of the model were used to establish realistic recharge values that matched observed low flow response.

Group one consists of catchments within the Gondwana and Rungwe volcanic geomorphological zones. In these zones recharge occurs in two ways; i) on the plateau where the saprolite has a considerable thickness and ii) directly through fractured rocks in areas with steep slopes. Group one constraints have been set to between 4 to 15 mm month^{-1}. For *group two* the MMR constraint was set to between 2 and 15 mm month^{-1}. Ground water recharge occurs on the summit areas and it reaches the aquifer situated in the saprolite, where groundwater flows by gravity into the valleys and then is discharged into the rivers (CCKK, 1982). *Group three* is characterized by a smooth African erosion surface, where infiltration is expected to be high (although some of this infiltration forms significant interflow). The constraints range between 1 to 15 mm month^{-1} (Table 5.6), with the exception of 1ka37 which has very high recharge associated with limestone geology. Most of the rivers in *group four* are assumed to be seasonal and are dry for four to five months a year. In these areas low rainfall and high potential evaporation result in very little groundwater recharge and consequently no dry season base flow (1ka41 and 1ka42). Constraints for group five and six range between 0 to 8 mm month^{-1}. The upper MMR constraint values for the last three groups are highly uncertain and possibly unrealistically high. An evaluation of the model results after stage 1 will help to confirm this possibility.

Table 5.6 Range of mean monthly recharge (mm month^{-1})

	Min	Max
Group 1	4	15
Group 2	2	15
Group 3	1	15
Group 4	0	0
Group 5&6	0	8

5.3.2.4 % of zero flow

Table 5.7 shows the percentage of zero flow constraints assigned to the different groups and illustrates the high degree of uncertainty in setting this constraint, especially for groups 2, 4, 5 and 6. The maximum value for ***group 2*** largely reflects the assumption that some of these foothill sub-basins may experience zero flows in the dry seasons of some years, however, the value of 20% is highly subjective. ***Groups 4 to 6*** are expected to experience some zero flows, but setting realistic minimum and maximum values was almost impossible given the lack of observed data on the incremental flow contributions from these areas. The wide range has therefore been set to ensure that this constraint is not incompatible with the ranges of the other constraints and the model parameter ranges used in step 1 of the uncertainty modelling (see Figure 4.6).

Table 5.7 Range of the percentage of zero flows

	Min	Max
Group 1	0	0
Group 2	0	20
Group 3	0	0
Group 4	0	50
Group 5&6	0	60

5.4 Discussions and conclusions

In the hydrological modelling approach adopted for this study (Chapter 4) the constraints are used to guide the process of checking simulated ensembles for behavioural results to reduce the uncertainty in the possible parameter sets. Constraints were developed based on local observed data where mean monthly recharge , mean groundwater recharge, percentage of zero flows, the 10^{th} ,50^{th} and 90^{th} percentiles of the monthly FDC were used. The quality and quantity of the available data was an important consideration in the quantification of the constraints. The constraints were regionalized based on assumed sub-basin similarities so that values could be assigned to the ungauged sub-basins as well as those that are gauged. Somewhat subjective levels of uncertainty were added to the constraint values calculated from the observed data and additional levels of uncertainty included in extrapolating the constraints

to ungauged basins within the six sub-basin groups. The assumed flow regime characteristics of the groups, and the assignment of the sub-basins to the groups, were partly based the observed flow data and partly on the documented variations in flow regimes associate with the different geomorphological zones (CCKK, 1982).

As explained in Chapter 4, the ranges of the six constraints, as well as the input model parameter ranges have to be compatible with each other to ensure that 2 000 behavioural ensembles are found during stage 1 of the uncertainty modelling method. Some of the highly uncertain constraint bounds have therefore been set to what are probably unrealistic values simply to ensure that they do not clash with other constraints. An example is the relatively high values for the maximum Q90 constraint for groups 4 to 6, as well as the minimum values for the % zero flow constraint in the groups. It is expected that these constraint ranges will be able to be reduced after the first complete run of the model and a detailed examination of the output constraint and parameter ranges contained within the 2 000 behavioural ensembles. This process can be considered the equivalent of 'calibrating' the uncertainty ranges and parameter ranges against expected sub-basin behaviour. It is much more difficult to determine *a priori* fully appropriate constraint ranges in the groups that have few or no representative observed stream flow records. Where these records do exist, as in group 1, the *a priori* constraints can be set with quite high confidence and with a relatively narrow range.

6 GW_PITMAN: RESULTS AND DISCUSSION

6.1 Introduction

Within current approaches to hydrological modelling, uncertainty analyses have become standard practice. Hence, the hydrological modelling process involves uncertainty analysis in both the inputs (input data and parameters) and the model results. This chapter present the results and discussions of the application of the improved GW_PITMAN uncertainty analysis framework. The results build upon previous applications of the model and the uncertainty framework (Hughes *et al.,* 2008; Kapangaziwiri, 2010; Kapangaziwiri *et al.,* 2012; Tshimanga, 2012). The whole methodology presented in this thesis is centred on uncertainty and reducing uncertainty. The approach uses a combination of uncertain parameters and uncertain output constraints that are quantified from uncertain observed data (for gauged sub-basins) and uncertain understanding of ungauged sub-basin runoff response. The basic concept is that the degree of uncertainty will vary spatially depending on the amount and quality of the available information that can be used to constrain the model outputs.

The input parameter uncertainty bounds were partly based on some trial manual calibrations of the model, but were generally set to be quite large with the expectation that these would be reduced using the output constraints (see Figure 4.5 – the concept figure in chapter 4). The objective was therefore to determine 2 000 behavioural parameter sets that generate outputs that are reasonably distributed within the set of constraint bounds. This was not expected to be a straightforward task because of the multi-dimensional nature of the problem, where there are many interactions between many different factors in the model. These include issues of parameter equifinality, the interactions between the six constraints that are used, as well as the complex interactions between the model (through its parameters), the input climate data and the simulated output values of the six constraint indices. It should also be emphasised that the whole process of setting up and quantifying the constraints is associated with uncertainties due to the poor quality of some of the available data and the influence of largely unquantified water abstractions on the extent to which the observed data can be considered to represent natural flow regime characteristics. It was expected that the first application of the 2 step

uncertainty analysis (refer to chapter 4) would require some degree of 'calibration' to ensure 2 000 successful outputs from step 1 that would involve adjustment of either some of the input parameter uncertainty ranges or some of the constraint ranges for the ungauged sub-basins. It was also anticipated that this first application would not necessarily represent the final uncertainty setup, but that a great deal would be learned from the integration of all the observed information and simulated model results across the whole GRR basin. The additional knowledge gained from this exercise would then be expected to contribute to a future re-assessment that would result in fully realistic uncertainty bounds at all locations within the basin that could be used for future water resources planning and management.

6.2 Overview of the methodology

The initial approach included the generation of an uncertain parameter set subjectively quantified using physical basin characteristic information. While all parameters are capable of being applied within the uncertainty framework, only the main runoff generation, moisture store accounting and ground water accounting parameters were assessed for uncertainty because these parameters represent the dominant natural hydrologic processes. The initial approach applied the structured uncertainty version of the model using the Monte Carlo structured sampling approach. Each sub-basin is treated independently with a different parameter set. The output file produced by the model consists of a text file of all the ensemble parameter sets (10 000), the summary statistics data (mean monthly runoff, mean monthly recharge, the slope of the flow duration curve, the 90^{th} , 50^{th} , and the 10^{th} percentiles of the flow duration curve) and five objective functions if observed data are included. These are the Nash-Sutcliffe coefficient of efficiency for the normal (CE n), natural logarithm transformed (CE ln) and inverse values (CE Inv) and the mean monthly runoff bias for both normal (% Diff n) and the natural logarithm transformed values (% Diff ln)). A second output file from each model run includes three time series based on the ranking (ascending order) of the 10 000 simulated flows for each month. The lower time series represents the 5^{th} percentile of the ranked values while the central and upper time series represent the 50^{th} and 95^{th} percentiles respectively. The lower and upper time series therefore represent the bands covering 90% of all simulated flows. The ensemble outputs were analyzed against the objective functions using excel and the ensembles producing the best fits to the observed data were identified. The uncertainty within

upstream sub-basins was realistically represented using this method. However, there was a problem associated with the downstream sub-basins and the uncertainties were not preserved (Figure 6.1). This was due to a cumulative uncertainty smoothing effect in downstream sub-basins. The other problem from a practical perspective of regional water resources assessment and management, is that there is no guarantee that behavioural downstream simulations are made up of behavioural simulations in all the upstream sub-basins.

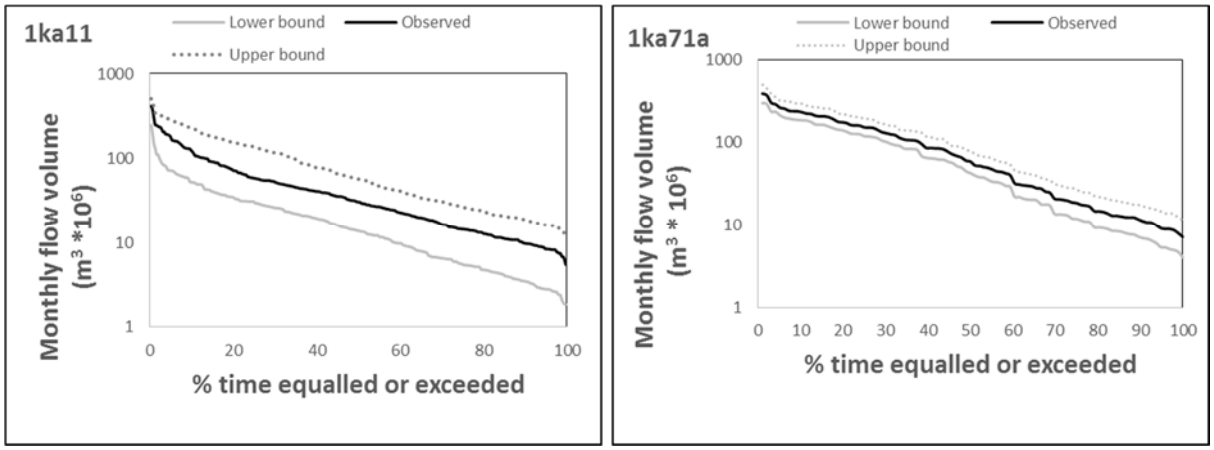


Figure 6.1 Simulated uncertainty bounds (5th and 95th percentiles of the total ensemble set) and observed flow for the sub-basin 1ka11 (upstream) and 1ka71a (downstream)

The revised sampling approach of the GW_PITMAN uncertainty framework was subsequently developed at the Institute for Water Research, Rhodes University. The revised sampling approach was implemented using the SPATSIM Global Options Model (Chapter 4). The methods followed for establishing appropriate parameter bounds is summarized in Figure 6.2 and outlined below:

1. Assessment of observed data inconsistences;
2. Catchment characterization and similarity analysis;
3. Assign regional constraints (upper and lower bounds) based on observed stream flow discharge data and the results of step 2 above;
4. Assign ranges (minimum and maximum) to the parameters affecting natural hydrology using information generated in steps 1 and 2;
5. Parameter sampling using the incremental model (with parameter ranges and regional constraints assigned in steps 3 and 4 respectively) to generate 2 000 behavioural natural

parameter sets for each sub-basin. If 2 000 parameter sets are not found then the ranges assigned in step 4 are adjusted and the model re-run for that sub-basin.

6. Additional water use and wetland parameters are incorporated into the model:
   - Data on the water use were estimated using previous reports and information from the Rufiji Basin Water Office (RBWO). Non-irrigated demand (domestic water use) was estimated using the 2012 Tanzanian population census results. The irrigated area was estimated using information from SWUWC project reports and google earth aerial images as well as information from previous studies in the area.
   - The wetland sub-model parameters were estimated using information from google earth and area-volume relationship curves.
7. The cumulative model is run 10 000 times with Monte Carlo sampling from the 2000 behavioural saved natural parameter sets for each sub-basin (step 5), plus independent sampling from the other parameters assigned under step 6. The outputs can be further examined using the statistical objective functions or time series and FDC plots compared with the available observed data.

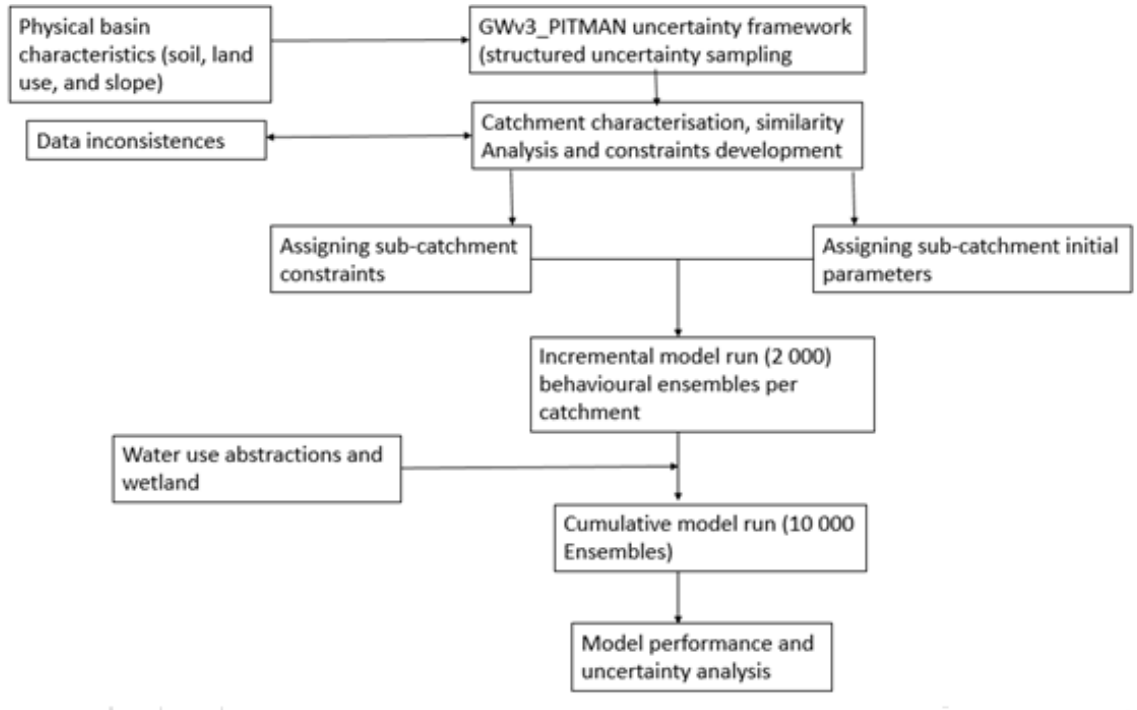


Figure 6.2 Hydrological modelling process for the GRR basin

It is important to note that the input constraints are regional values identified for the 6 groups of sub-basins (Chapter 5), although some within-group variations in the ranges were allowed for when it was assumed that better quality observed data were available. The input ranges for the natural parameters (step 4 above) were also regionalised for the same 6 groups. Variations in climate inputs will mean that the results for the incremental run of the model could be quite different for different sub-areas within each group. This effectively means that neither the full range of constraints, not the full range of input parameter values may appear within the constraint and parameter values of the 2 000 behavioural outputs. An analysis of these outputs is typically used to further guide the local (i.e. single sub-basin) quantification of the natural parameters (i.e. a form of calibration) or to prompt a re-evaluation of the local constraint ranges. This type of assessment is also necessary if less than 2 000 behavioural outputs are generated (from the maximum of 50 000 model runs used for the incremental analysis in this study). There were a number of sub-basins where it was necessary to repeat the incremental model run (step 5), but the details of this part of the process are not included here. The following sections, however, present a detailed analysis of the outputs from the final version (in this study) of the incremental uncertainty model run. As noted later in the conclusions, this does not necessarily represent the final version of the total uncertainty model that would be used for practical water resources analysis and further refinement of the constraint and parameter ranges is certainly possible.

6.3 Establishing the ranges of the natural hydrology parameters

The procedure followed for the development of constraints was detailed in Chapter 5. The GW_PITMAN model was established with single (no uncertainty) parameter inputs and the parameters were manually calibrated by comparing simulated outflows for key gauged sub-basins with the observed flow data. The original version of the uncertainty version of the model (without using any constraints on the outputs) was also run a number of times to help with the determination of plausible parameter ranges and combinations. These steps helped in establishing the initial parameter value ranges which were assigned to the sub-basin groups identified in Chapter 5. The same uncertain parameter sets were initially applied to all sub-basins within a group, before isolating specific sub-sub-basins based on their geomorphological differences. The uncertain parameter ranges for all the sub-basins are presented in Appendix

F. On the basis of the initial parameter explorations, as well as previous experience of applications of the model in other parts of southern Africa (Hughes, 2013a), it was decided that the ranges of some parameters could be fixed for all sub-basins, while the uncertainty ranges for other parameters would necessarily be dependent upon the characteristics of the different sub-basin groups. Thus, the interception (PI), evapotranspiration (R) and routing parameter ranges were kept the same, while the riparian evaporative loss parameter ranges were set to reflect valley bottom topography characteristics; low values for steep catchments with assumed narrow riparian margins, higher values for lower topography sub-basins. The primary focus for establishing variations between groups was therefore the main runoff, recharge and ground water drainage parameters and the intention was to ensure that the ranges established for these parameters would be compatible with the ranges already established for the output constraints. The focus was therefore on ZMIN and ZMAX, controlling surface runoff; ST, FT and POW controlling interflow; GW, GPOW controlling ground water recharge and, to a lesser extent, S and T that affect ground water contributions to stream flow.

Group 1 consists of the headwater sub-basins of the UGRR. The steep slopes on the Rungwe and Gondwana geomorphological surfaces suggest high FT values and the range was set between 10 and 40 mm month^{-1}. ZMIN values were set between 10-60 mm month^{-1}. Moderate ST (200-500 mm month^{-1}) and ZMAX (200-600 mm month^{-1}) values reflect the well-drained volcanic ash soils made up of predominantly sandy clays and loams. The values for the GW parameter were relatively high (10-40 mm month^{-1}) as recharge is thought to be significant to support large volumes of base flow in the sub-basin. Storativity and transmissivity were also assigned high values (fixed) as previous studies have identified significant ground water storage (CCKK 1982). *Group 2* consists of intermediate sub-basins with most falling in the foothill zones draining to the Usangu plains and the Post African geomorphological zone. Interflow was expected to be lower than in Group 1 due to decreased slopes and the FT values range from 0-25 mm month^{-1}. Infiltration capacity on the alluvial plains is high and this is reflected in moderate to high ZMIN and ZMAX values. ST values were set to between 500-1100 mm month^{-1}. Due to irregular topography, groundwater recharge varies across the landscape, ground water recharge parameters (GW) were set to between 5-25 mm month^{-1}.

Group 3 falls within the African zone characterized by good infiltration into the top soils and moderate to high absorptive capacity of the soil as reflected in ST values ranging from 600-1200 mm m^{-1}. ZMAX parameter values also reflect the high absorptive capacity of the soil. For the lower parts of the basin ZMIN values range between 30-100 mm $month^{-1}$ as the flat topography reduces overland flow and good soil structure explains the increased values in ZMIN compared to the lower values in the upper zones. Most sub-basins in *Group 4* fall in the semi-arid part of the GRR sub-basin drained by the Kisigo River and other small tributaries. The Kisigo River is seasonal and is dry for 4 to 5 months a year. FT parameter values were fixed at 0 while ground water recharge ranges from 5-10 mm $month^{-1}$. The minimum absorption rate ZMIN ranges from 20-60 and the maximum absorption rate ZMAX ranges from 300-500 mm $month^{-1}$. *Groups 5 and 6* have high ZMIN parameter values, ranging between 150-200 mm $month^{-1}$ together with a high ST parameter value range of 1200-1600 mm $month^{-1}$ reflecting the large infiltration capacity of the sands on the flood plains. For the same reason the ZMAX parameter value range is high (800-1200 mm $month^{-1}$). In the plains the deposits are slightly less permeable compared to the Alluvial fans zone due to the spatial distribution of the deposits and this explains the lower GW values (0.5-2 mm $month^{-1}$) that were used. Due to the flat topography in this zone FT values were fixed at 0.

6.4 Quantifying the water use, wetland and other parameters

The water abstractions have been assumed to consist of relatively small domestic abstractions for village communities located along the south western borders of the basin (mostly groups 1 and 2), together with large scale rice irrigation schemes at the base of the foothill zones from the tributaries that flow into the Usangu Plains (mostly group 2). Table 2 lists the main water use parameters and identifies the main rice growing areas. These data were obtained from previous reports and population census data (SMUWC, 2001; URT 2012) but are recognised to be highly uncertain. The rice irrigation water use was based on an annual demand of 1 070 mm y^{-1} with a high return flow fraction (0.33). The other irrigation areas used an annual demand of 475 mm y^{-1} with lower return flow fractions depending on the topography (0 in flat areas, 0.2 in steeper sub-basins). While the water use data can be input as uncertain ranges in step 2 of the modelling process, fixed values were used in the study to avoid confusion between

different sources of uncertainty. Water use uncertainty can be added at a later stage in the water resources assessment of the GRR.

According to Mistch and Gosselink (2000) wetlands have the capabilities to store runoff from direct precipitation as well as retain out-of-bank flows by providing additional storage areas during annual flood seasons. This enables the wetlands to release water back to the river channels during dry seasons, thus maintaining base flow of rivers and regulating the river runoff throughout the year (Wang *et al.,* 2009). Understanding hydrologic processes of wetlands is of importance due to their influence on the hydrological regimes of rivers. Processes occurring in wetlands are often overlooked and not fully incorporated in the conceptual development of many hydrological models of surface runoff (Hughes *et al.,* 2013d). The wetland parameters (Table 6.1) were included after an initial run of the model and an evaluation of the results against observed stream flow data at 1ka59. The physical characteristics of the Usangu wetlands were used to estimate some of the parameters of the wetland sub-model, while others were manually calibrated as there is no physical information that can be used to directly estimate them. The western flood plain is located within sub-basin1ka71a, while the eastern wetland model is within 1ka27. The estimated local catchment areas for the western and eastern flood plains are 800 km^2 and 600 km^2, respectively and were based on elevation-area-volume curves from previous studies (SWUWC, 2001; WWF, 2010). The documented understanding (WWF, 2010) of the dynamics of these wetland areas is not sufficient to confidently state that the parameter values used are physically representative, but they are believed to be within reasonable bounds of possibility. For example, the zero value for channel capacity for spillage (QCap) and high value for channel spill factor (QSF) reflect the fact that the main channel of the GRR disappears within the downstream part of the wetland within 1ka27, while a major channel feature is evident throughout the upper wetland area within 1ka71a.

In addition to the water use and wetland parameters, afforestation parameters were included in sub-basin 1ka32a, transmission losses (from the channel to ground water) in the downstream arid areas (e.g. 1ka42, 1ka59, ug24 and 1ka5a).

Table 6.1 Parameter estimates of the wetland model application at two gauging sites: 1ka71 (seasonal wetland) and 1ka27 (permanent wetland)

Array Parameter	1ka71a	1ka27
Local catchment area (km^2)	800	600
Residual Wetland storage ($m^3 * 10^6$)	40	80
Initial Storage (MCM)	20	40
A in Area(m^2) = A * Volume(m^3)^B	1000	2000
B in Area(m^2) = A * Volume(m^3)^B	0.8	0.8
Channel capacity for spillage ($m^3 * 10^6$)	10	0
Channel Spill Factor (Fraction)	0.1	0.8
AA in (Ret.Flow = AA*(Vol/RWS)^{BB})	0.2	0.1
BB in (Ret.Flow = AA*(Vol/RWS)^{BB})	0.6	0.8
Annual Evaporation (mm)	1870	1715

6.5 Modelling results

Analysis of the parameters and constraints

The results presented in this section are based on the output before water use (step 5). The extent of the uncertainty associated with each constraint and parameter was defined by upper (maximum) and lower (minimum) bounds for each sub-basin. The simulated minimum and maximum values of the constraints for the 2 000 behavioural outputs were compared with the input ranges of parameters and constraints to establish if the defined constraint uncertainties are compatible with each other and if the parameter bounds are compatible with the constraints. In the process of constraining the hydrological simulations using regionalised constraints a number of possible outcomes are expected and these could be:

a.) The model output has occupied the full range of uncertainty, which means the range of output is within the constraint boundaries suggesting that the two step sampling procedure has produced sets which generate acceptable outputs that are behavioural. However, in cases where the constraints and parameter bounds are wide there would

still remain a high degree of uncertainty in the interpretation of behavioural response (e.g. group 2).

b.) The model output has failed to occupy the full range and the output constraints concentrate either toward the lower, middle or upper end of the input constraint range. The reason for this could be a mismatch between the constraint limits and the input parameter ranges or the input data, preventing the model from finding solutions across the full constraint range. In extreme cases, the parameter ranges were modified and the model re-run. In less extreme cases, the issue was noted and could be addressed during a revision of the whole uncertainty model for all sub-basins as a follow-up on from this study.

The top part of Figure 6.3 illustrates the ranges of the surface runoff parameters ZMIN and ZMAX, while the lower part shows the ranges of soil moisture storage ST and ground water recharge parameter GW. The grey blocks represent the full input ranges, while the black blocks are the ranges of the final 2 000 behavioural outputs from stage 5 of the modelling process referred to in section 6.2. For the surface runoff parameters the behavioural parameter sets are within the input range except for catchment 1ka59 (Figure 6.3A). 1ka59 ensembles occupy the lower end of the range and this could be associated with the large input range assigned to this catchment. The behavioural range of ZMAX occupies the full range of the input constraints (Figure 6.3B). For parameter ST most sub-basins occupied the full input range except for catchment 1ka50a where there is a bias toward lower ST values (Figure 6.3D). Figure 6.3C shows the results for parameter GW, where the behavioural results suggest that the lower values of the input parameter range are incompatible with the range used for the ground water recharge constraint in many of the sub-basins.

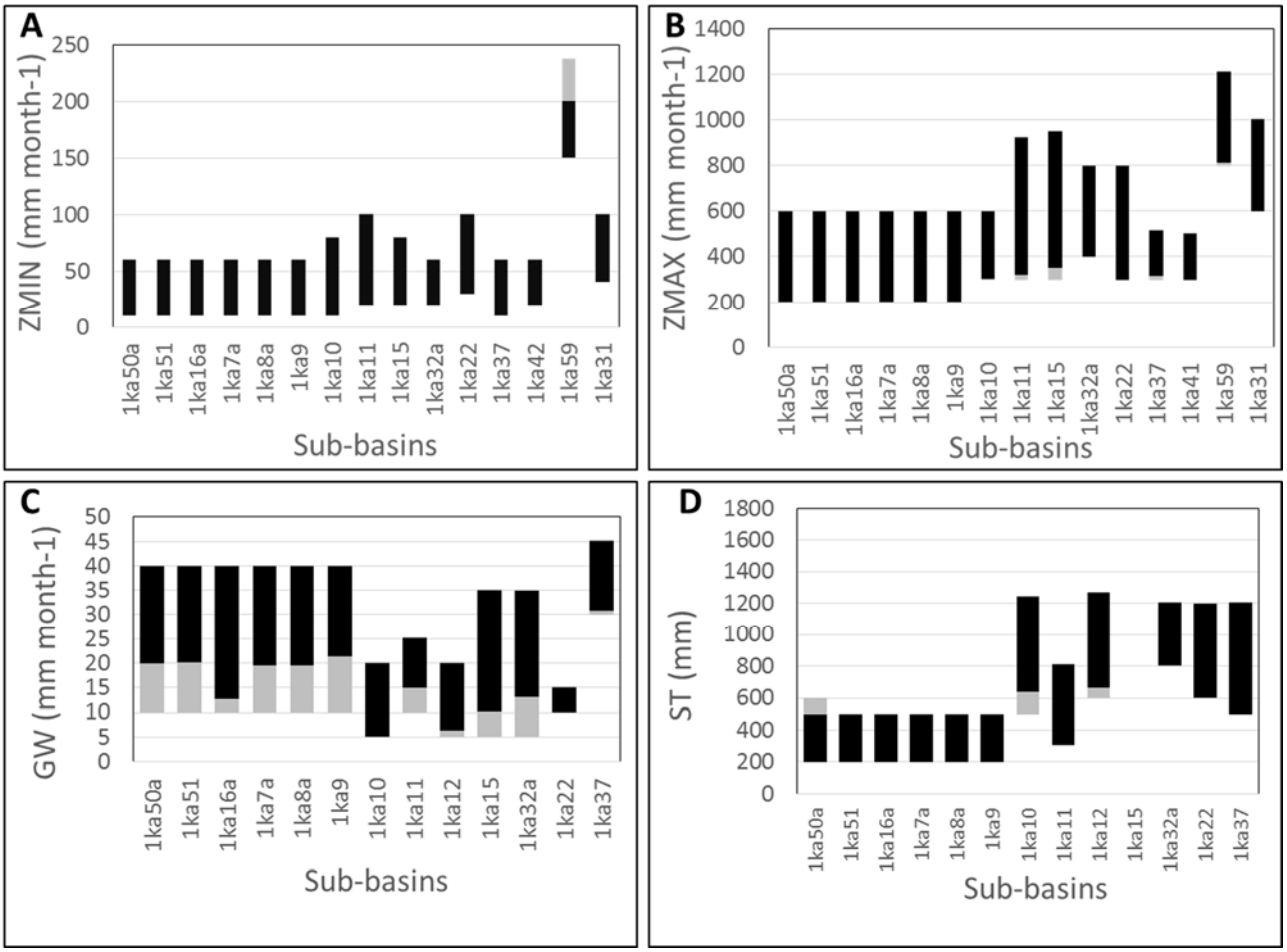


Figure 6.3 Uncertainty ranges for parameter ZMIN, ZMAX, ST and GW. The grey bars represent the input ranges and the black bars the behavioural ranges output form step 5 of the modelling process for natural flow conditions

The top part of Figure 6.4 illustrates the ranges of parameter FT and the input and output of the combined effect of the low flow parameters (FT, POW, GW and GPOW). The interactions of parameters FT (maximum interflow runoff per month) and GW (maximum ground water recharge per month) make it difficult for these parameters to be identifiable when considered separately. This is because both parameters can contribute to sustaining low flows, either directly (interflow) or indirectly through ground water discharge and it is hard to separate individual contributions of each of these parameters, representing one of the main sources of equifinality in the model. In addition, the two power parameters also affect the simulated quantities of low flows (higher powers, lower runoff). Figure 6.4B shows the combined effect

of FT, POW, GW and GPOW and it is clear that the range of behavioural combinations is less than the full possible range based on the inputs. Figure 6.4B illustrates that the constraints used to limit the total parameter space has resolved some of this equifinality in that the very high and very low combinations of the combined index of low flow parameters are not accepted as behavioural.

The lower part of Figure 6.4 illustrates the output behavioural ranges of the mean monthly recharge (MMR) and low flow (Q90/MMQ) constraints. There are a number of cases where the output ranges are substantially lower than the input ranges and this occurs particularly in group 2 where the regional uncertainty is quite high and in group 3 where some basins are characterised by high base flows. Figure 6.4C shows that the output range occupies the full input range. The output constraint ranges for the remaining sub-basins are toward the lower end of the constraint range. This could be due to the high values set for the upper constraint bounds and that a combination of other constraints or the input parameter bounds have prevented the higher values for MMR being found within the behavioural outputs. The range of Q90/MMR behavioural values are much closer to the input range (Figure 6.4D) and these results arguably reflect the greater confidence in estimates of the low flow constraints than the ground water recharge constraint. For some of the group 3 sub-basins high constraint values were set to account for large base flow volumes in 1ka22 and 1ka37. However, the relatively high values for the upper values of the MMR constraint in sub-basins 1ka10, 1ka12 and 1ka11, located at the foothills of the mountains, do not appear to be justified and did not appear in the final behavioural ensemble set. There are differences in the climate inputs to the model across sub-basins that were assumed to have broadly similar responses and this certainly affected the results. Thus the first 6 sub-basins (group 1) all have the same FT and GW parameter ranges, but have quite different responses in terms of MMR and Q90/MMQ.

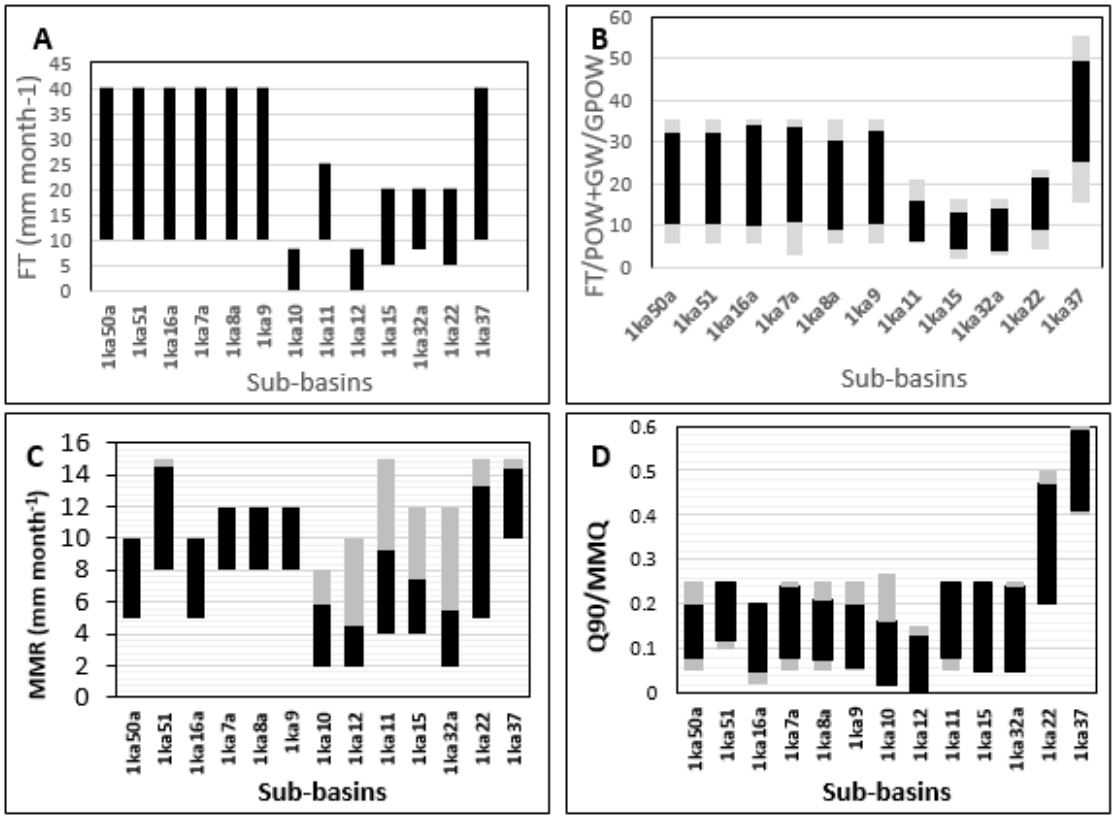


Figure 6.4 Range of FT parameter, MMR and Q90/MMQ constraints (A, C, D) as well as range of input and output values for an index of potential low flow generation using FT/POW+GW/GPOW. The grey bars represent the input ranges and the black bars presents the behavioural ranges output from step 1 of the modelling process for natural flow conditions.

The top part of Figure 6.5 illustrates the ranges of the mean monthly flow (MMQ) and high flow (Q10/MMQ) constraints while the lower part shows the ranges of the minimum flow (Q50/MMQ). The output ensembles for the MMQ have occupied the full range of the constraint uncertainty indicating that the ensembles are evenly distributed across all the constraints members, also there is consistency between constraints and parameters (see Figure 4.5A). Figure 6.5B and 6.5C illustrates the comparisons between the simulated ensembles and the input constraints for the Q10 and Q50 constraints and behavioural output are within the constraints bounds for some sub-basins, however, there are cases where output ensembles have not occupied the full range of input values. All regional groups have been fed with similar constraints and parameter sets, but the results shows difference response to Q10

and Q50. The last two sub-basins in Figure 6.5C have not occupied the full range, this could be explained by high maximum input values set for the two sub-basins as the two catchments are characterised by high base flows. Another explanations would be the constraints limits were too low to occupy the full input range or the model could not find behavioural sets at the upper end of the constraints.

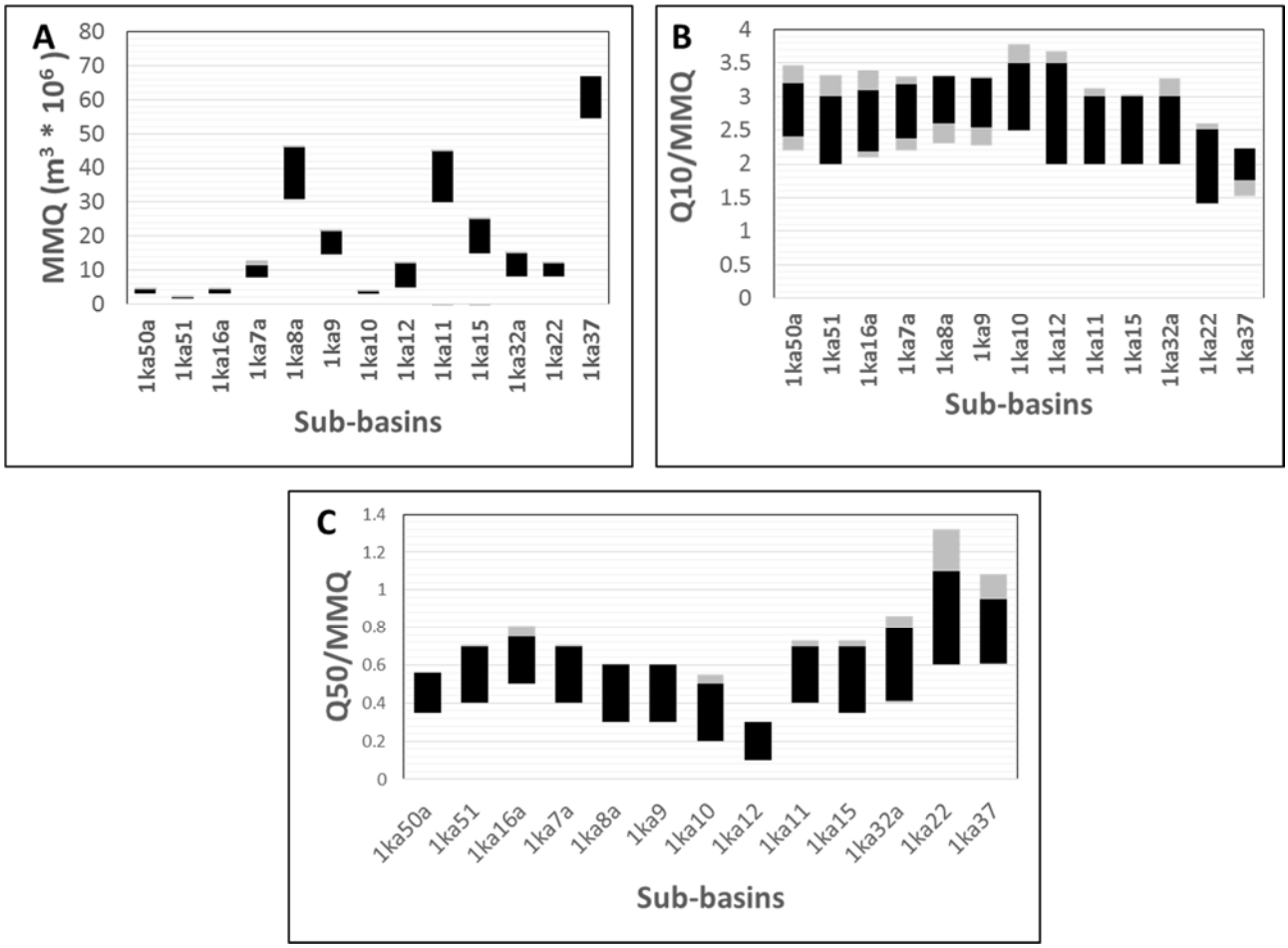


Figure 6.5 Range of MMQ, Q10/MMQ and Q50/MMQ constraint. The grey bars represent the input ranges and the black bars presents the behavioural ranges output form step 1 of the modelling process for natural flow conditions

At the completion of step 5 in the modelling process, 2 000 behavioural (relative to the constraint bounds used) parameter sets were generated for all sub-basins. However, the results suggest that there remain some inconsistencies both between constraints and between some of the parameter bounds and the constraints. This is particularly evident in the ground

water recharge constraints which were difficult to quantify because of the lack of direct information. Also, it is well known that there could be interactions between Q50 and Q90, where in some cases both interflow and ground water contribute to minimum or low flows through an interaction that is hard to resolve. Therefore, situations where the output has failed to occupy the full range indicates some difficulties associated with variations in input data, representation of basin processes (e.g presence of dolomites in group 3 sub-basins) or a mismatch between parameters and constraints. Despite some difficulties in getting the full range of behavioural output, the model has been able to generate 2 000 behavioural ensembles which fall within the constraints, but not all distributions within those constraints are evenly distributed. This will introduce some degree of bias in the results relative to the assumption of uniformly distributed constraints. Moreover, it should be noted that the behavioural model is the combination of all model input, parameters and the constraints, thus there is very high degree of equifinality that is hard to resolve.

There is clearly scope for further refinement of the input parameter and constraint bounds, but the objective of this part of the study to obtain the 2 000 behavioural parameter sets has been achieved. Further adjustment and alignment of the parameters and constraints is recommended for future work. This will involve the re-evaluation of the observed flow data to identify the more reliable records (or parts of some records) so that the constraint ranges for these sub-areas can be reduced (Westerberg et al., 2012; Sellami et al., 2013) or adjusted. Although getting realistic ground water recharge information will be extremely difficult, some large scale ground water recharge information (e.g www.whymap.org) could be sought to provide further guidance in setting up the ground water recharge constraints. The uncertainty in the input climate data was largely ignored in this part of study, but the results suggest that including these could be important to resolve some of the other uncertainties.

6.6 Quantification of the uncertainties in simulated results

The whole methodology presented in this thesis is centered on uncertainty and reducing uncertainties. The full range of uncertainty simulated ensembles (with water use and wetland parameters) is presented in this section. Table 6.2 shows the minimum and maximum values of performance measure (CE and CE (ln)) and Percentage bias (% bias) for selected gauged sub-

basins. Sub-basins with CE and CE (ln) values above 0.5 are highlighted in grey. Other sub-basins have poor results for one or both the CE and CE (ln) values. Apart from possible problems with the model itself (structure and parameters) this could be related to a number of factors including poor rating curve data (Chapter 5), inappropriate climate inputs, poor estimates of water use (including non-stationarity in the actual water use) or a lack of accounting for land use impacts in heavily populated areas. Uncertainties associated with the interpolation of the input rainfall data are very high in the western sub-basins where there are no rainfall stations. Figure 6.6 to 6.17 illustrate flow duration curves of the simulated uncertainty bounds and observed data for selected sub-basins.

Table 6.2 Performance measure statistics (with water use and wetland parameters)

Sub-basin	CE	CE(ln)	% bias	% bias (ln)
1ka51	-0.61 : -0.04	-0.10 : -0.27	-2.4 : 26.70	-17.9 : 25.0
1ka8a	0.58 : 0.70	-0.43 : 0.50	-21.8 : 14.80	-51.5 : 8.70
1ka9	0.49 : 0.60	-0.32 : 0.60	-27.2 : 10.40	-58.8 : 1.70
1ka11	-0.76 : 0.30	-7.76 : 0.50	36.5 : 3.10	-70.8 : -15.90
1ka56	0.39 : 0.54	0.04 : 0.80	-29.1 : 7.80	-44.4 : 33.40
1ka32a	-1.80 : 0.30	0.09 : 0.70	-25.7 : 39.10	-46.0 : 52.0
1ka22	-2.30 : 0.29	-1.31 : 0.50	-20.9 : 18.50	-41.8 : 4.50
1ka31	-0.47 : 0.30	0.38 : 0.70	-22.1 : 22.20	-32.3 : 22.80
1ka37	0.40 : 0.60	-0.32 : 0.60	-27.2 : 10.40	-58.8 : 1.70
1ka42	-2.50 : 0.30	-0.96 : 0.50	27.4 : 17.40	-13.2 : 3.00
1ka27a	0.58 : 0.73	0.45 : 0.74	4.4 : 42.70	-12.1 : 71.00
1ka71	0.29 : 0.69	-0.29 : 0.80	-24.1 : 35.40	-32.3 : 54.70
1ka59	0.40 : 0.50	0.20:0.60	-15.90:19.40	-10.17:83.20
1ka5a	-0.84 : 0.57	-0.60 : 0.84	25.50 : 95.90	-21.70:106.80

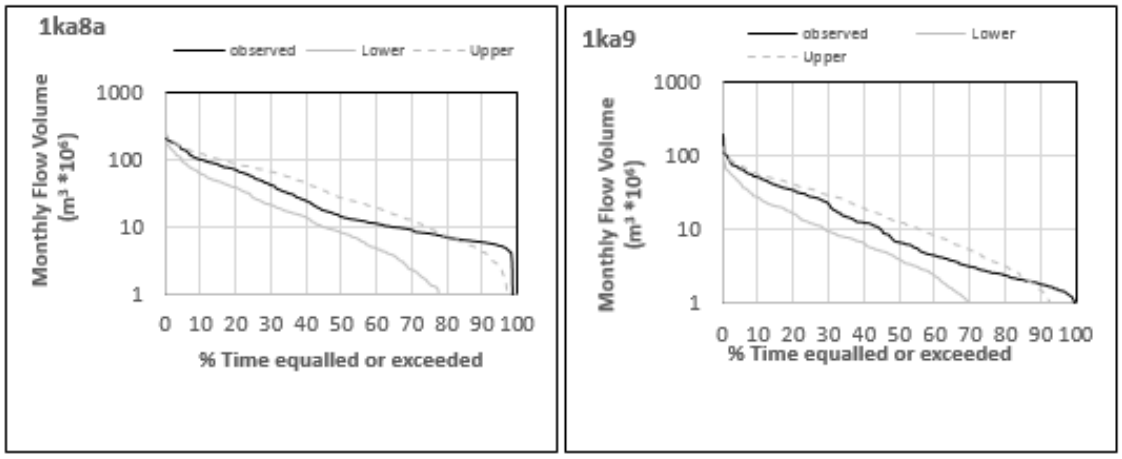


Figure 6.6 Flow duration curves of simulated bounds and observed data for sub-basin 1ka9 and 1ka8a, results are after step 6 and include water use

Figure 6.6 illustrates the results of the ensemble simulations for sub-basins 1ka8a and 1ka9. The results show that the uncertainty bounds generally bracket the observed flow. Figure 6.7 illustrates the results of the ensemble simulations for sub-basin 1ka11. The results shows that the upper and lower bounds brackets the observed flow expect for the low flows (90% exceedence). There is a bias in the simulations with most ensembles falling below the observed flow for moderate to low flows, while very high flows are apparently over-simulated. The latter could be related to uncertainty in the measurement of high flows. Rating curves for the two sub-basins (1ka8a and 1ka9) are characterised by lots of scatter in early gauging's hence unreliable.

Figure 6.8, illustrates that the upper and lower bounds have bracketed the frequency characteristics of the observed flow data for 1ka32a, 1ka22, 1ka37 and 1ka31. However, there is a bias towards under-simulation in 1ka32a and 1ka22. Low flow (90% exceedence) uncertainty is relatively large in sub-basin 1ka32a and this is partly explained by the uncertain effects of forest plantations. In sub-basin 1ka37 there is a similar high flow problem that was noted in 1ka11. However, most results for this group are very promising in that, uncertainty is evenly distributed for both upstream and downstream sub-basins (e.g. 1ka32a and 1ka 31). Sub-basins in this part of the basin have the most reliable observed time series with less than 10% of values missing except for catchment 1ka31 (21% missing values), suggesting that the

quality of the observed data plays an important role in both setting the constraints and assessing the model performance.

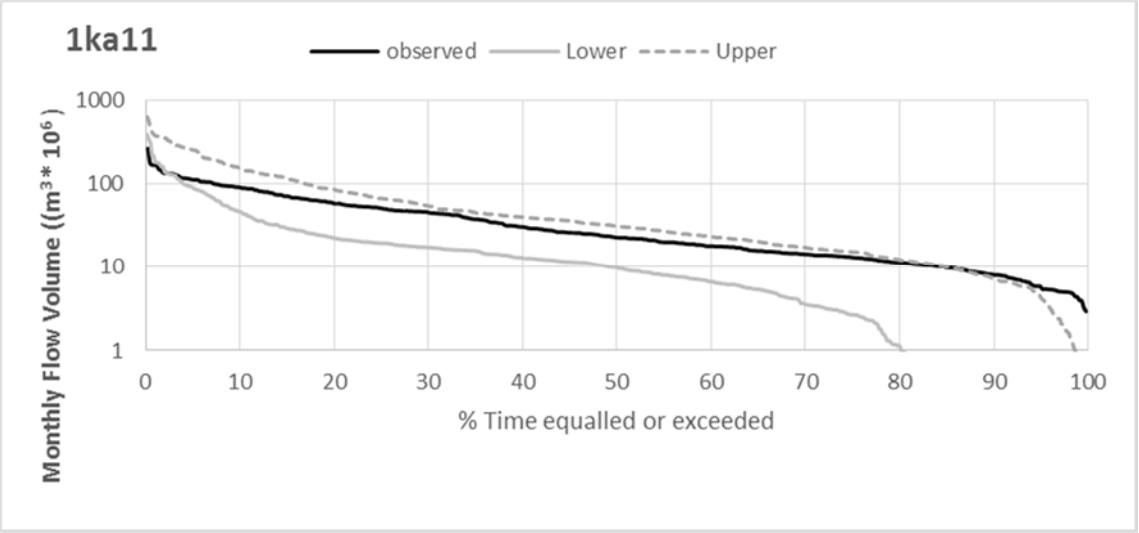


Figure 6.7 Flow duration curves of simulated bounds and observed data for sub-basin 1ka11, results are after step 6 and include water use .

Figure 6.9 illustrates the flow duration curve of simulated bounds and observed data for sub-basin 1ka42 in the more arid north-west of the basin. The upper and lower bounds brackets the high flows (10% exceedence) but there is large uncertainty and bias (relative to observed data) in moderate (50% exceedence) and low flows (90% exceedence). Despite several attempts to constrain the uncertainty in this sub-basin, the complexity of the processes and the fact that there are many upstream sub-basins that are not gauged has resulted in an unresolved amount of uncertainty for various reasons. The estimation of constraints from gauged data with poor time series is one possible reason. The data available to test various parts of this catchment are not generally available despite the fact that the catchment is gauged at the outlet (1ka42), the whole of the upstream part is ungauged. A further reason could be the influence of poor rainfall estimates in this part of the basin where there are few ground-based rainfall stations that can be used to correct any potential bias in the satellite data (Hughes and Sawunyama, 2009).

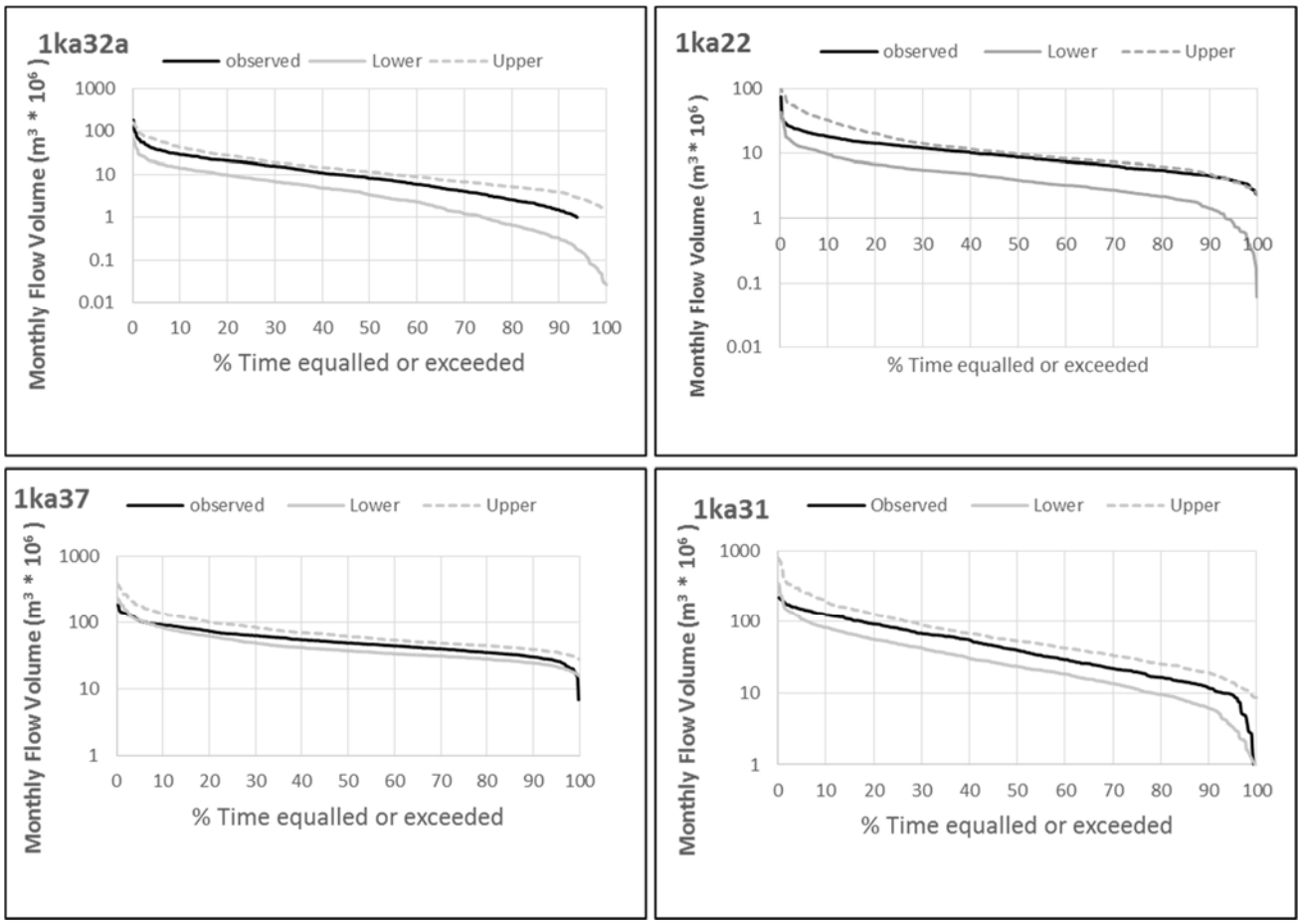


Figure 6.8 Flow duration curves of simulated bounds and observed data for catchment 1ka32a (upstream), 1ka31 (downstream), 1ka37 (upstream) and 1ka22 (upstream), results are after step 6 and include water use

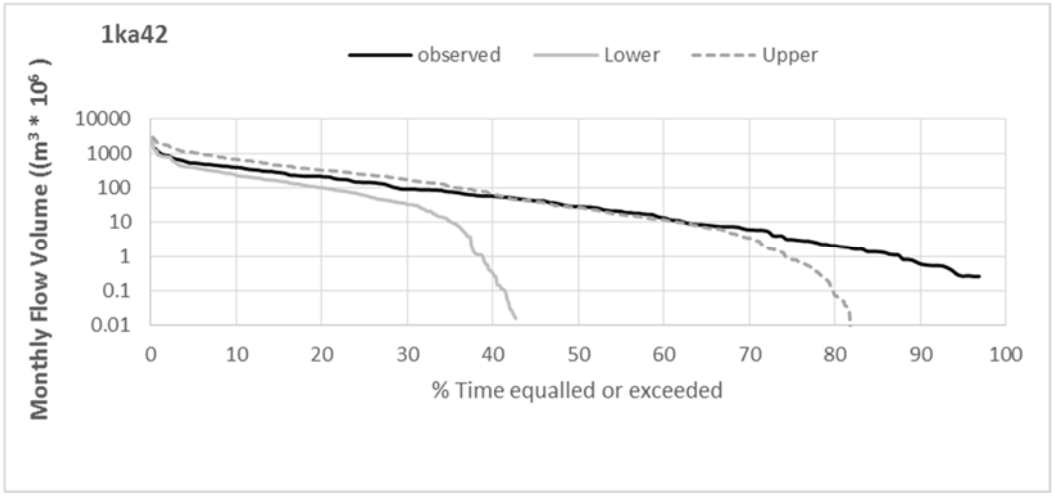


Figure 6.9 Flow duration curves of simulated bounds and observed data for sub-basin 1ka42, results are after step 6 and include water use

6.7 Simulation results and model performance

The results presented in section 6.6 illustrate the uncertainties associated with the simulated ensembles with water use. The results presented in this section are based on the best ensemble (based on all four objective functions listed in Table 6.2). The results are therefore based on a single ensemble with water use and wetland parameters included. The single ensemble was selected on the basis of the four main objective functions (CE (n), CE (ln), % Diff. (n) and % Diff. (ln). The final sets of best fit parameters are given in Appendix G. In general terms the model has been able to satisfactorily represent the basin hydrological behaviour. Before discussing these results several factors should be noted:

- All of the results are based on uncertain natural hydrology parameters that were derived using the step 3 and 4. There was no other form of 'calibration' used except for the parameters of the wetland function for 1ka71 and 1ka27a.
- The results are based on fixed but approximate values for the water abstractions. These are certainly not stationary throughout the observed data periods. In some cases the earlier (irrigation not established) and later (irrigation restricted to decrease downstream impacts) periods would reflect lower abstractions.
- The results are compared with very uncertain observed data that are affected by the historical patterns of abstractions, possible non-stationary rating curves and rating curves that do not include the full range of high flows in the observed time series records. This makes it very difficult to evaluate the results using conventional objective functions.

The results are shown both in the form of stream flow hydrographs for the simulated period and as FDCs representing the model simulations for the full range of the available observed data. The FDCs were used to show some simulation problems that could not be seen using the hydrographs. Table 6.3 presents the optimum ensemble performance statistics for all the sub-basins that have observed flow data.

Group one

The observed and simulated discharges are illustrated in Figure 6.10 and 6.11. The model is able to capture the basin behaviour based on the characteristics of the flow duration curve.

Calibration results show a good correlation between observed and simulated flows, although the statistics for two of the sub-basins (1ka50a and 1ka 51) are poor. The model under simulates high and medium flows in 1ka16a and 1ka50a sub-basins (Figure 6.10). For sub-basins (1ka7a, 1ka8a and 1ka9) calibration results show that there is a good correlation between observed and simulated flows (Table 6.3).

Table 6.3 Model performance statistics for all simulated sub-basins (simulations including water use)

Statistics of the GW_PITMAN model performance for the GRR sub-basin							
Gauging Site	R^2(n)	R^2(ln)	CE(n)	CE(ln)	%Diff(n)	%Diff(ln)	Group
1KA8a	0.74	0.54	0.74	0.40	-3.511	-3.34	1
1KA7	0.62	0.68	0.59	0.49	-10.71	-12.72	1
1KA16a	0.57	0.66	0.55	0.63	-3.17	76.84	1
1ka9	0.58	0.74	0.57	0.30	-1.94	-16.30	1
1ka50a	0.32	0.59	0.11	0.48	17.20	71.15	1
1ka51	0.32	0.33	-0.08	0.27	1.29	24.71	1
1ka10a	0.69	0.76	-0.01	0.73	-28.2	240.24	2
1ka11a	0.47	0.71	0.25	0.485	-3.89	-5.77	2
1ka12	0.37	0.44	0.33	0.31	-7.24	169.76	2
1ka15a	0.54	0.73	0.23	0.70	2.13	0.39	3
1ka32a	0.50	0.73	0.33	0.73	-6.53	0.76	3
1ka39	0.65	0.78	0.50	0.77	4.14	2.22	3
1ka21	0.60	0.74	0.49	0.73	2.51	0.99	3
1ka2a	0.50	0.63	0.42	0.62	-3.01	0.34	3
1ka31	0.47	0.68	0.40	0.68	-8.35	-0.26	3
1ka22a	0.47	0.63	0.28	0.57	-5.09	-2.71	3
1ka37a	0.42	0.57	0.16	0.49	2.23	0.21	3
1ka42	0.72	0.45	0.38	0.09	32.28	69.5	4
1ka5	0.78	0.85	0.53	0.81	32.4	5.03	4
1ka61	0.71	0.39	0.3	0.31	27	3.86	4
1kA71	0.68	0.65	0.68	0.52	0.26	1.51	5
1ka59	0.53	0.58	0.53	0.59	-7.31	1.13	5

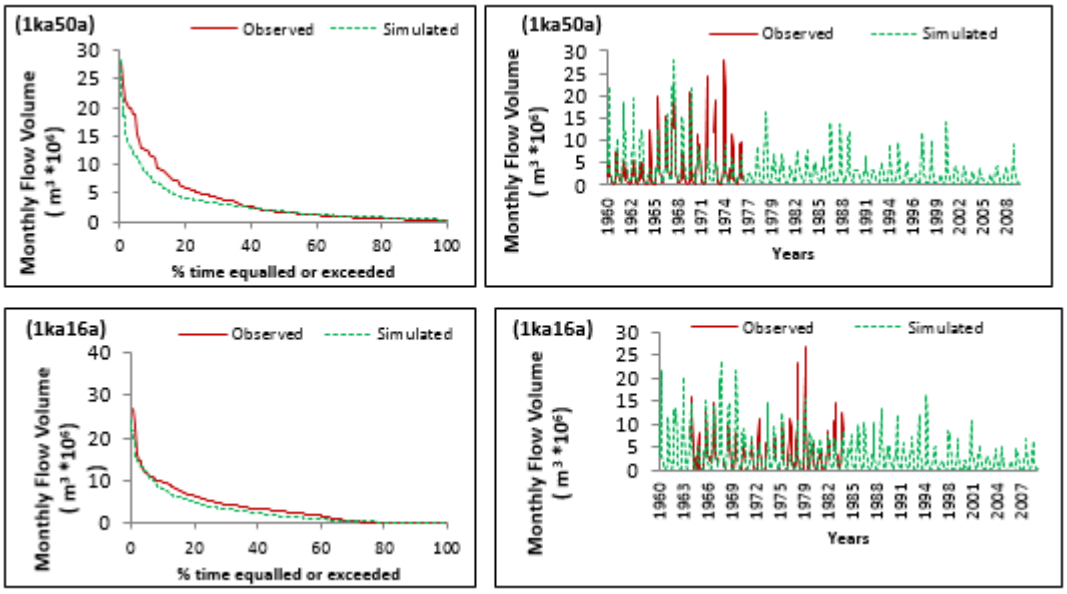


Figure 6.10 Observed and simulated monthly stream flow for gauging station 1ka50a and 1ka16a

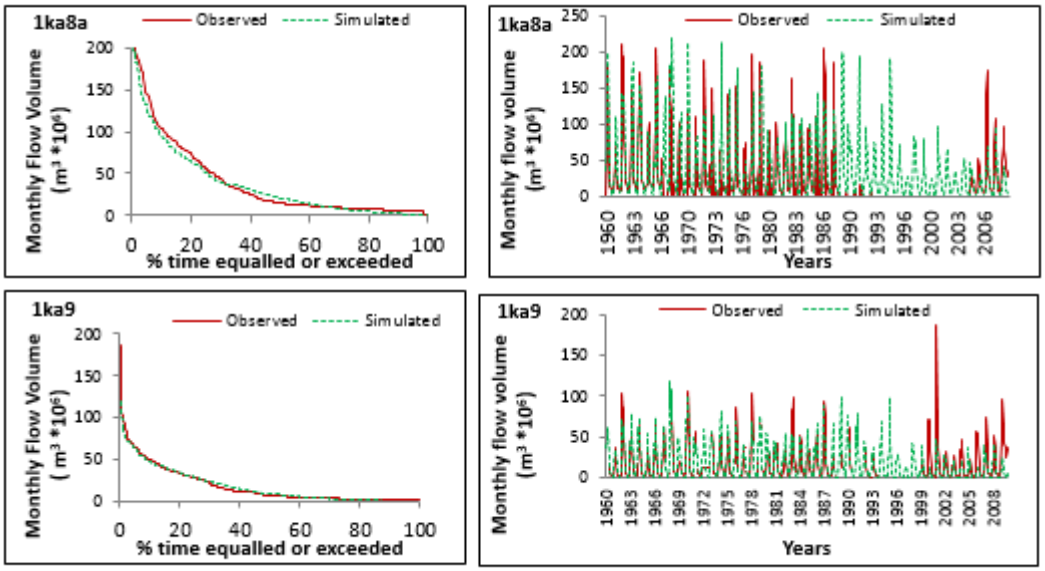


Figure 6.11 Observed and simulated monthly stream flow for stations 1ka8a and 1ka9

Group two

The hydrograph and flow duration curve for catchment 1ka11 are given in Figure 6.12. Based on the FDC the model captures all the important characteristics of the basin, however, high flows are apparently over simulated. While some parts are of the time series are well captured by the model (e.g. the period between, 1999-2006) the rest of the observed time series is not well represented because the model over simulates both high and low flows. The poor results could be explained by the large uncertainty surrounding both input and observed stream flow data.

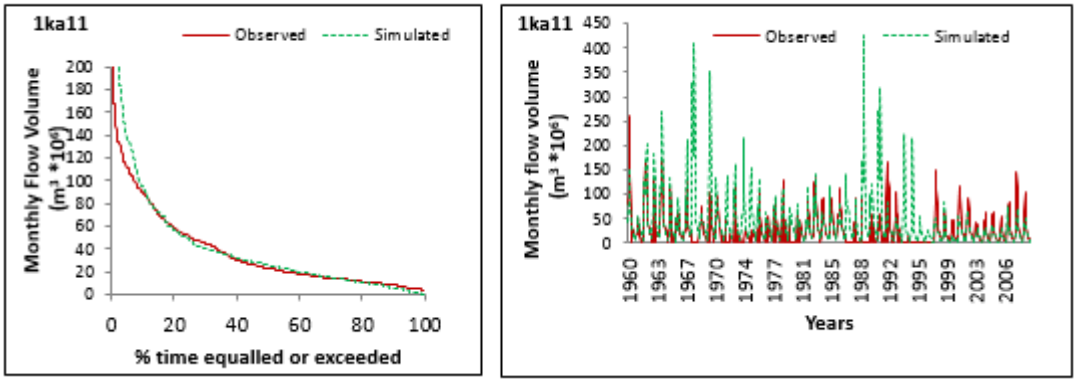


Figure 6.12 Observed and simulated monthly stream flow for gauging station 1ka11

Group three

In general, there is good agreement between the simulated and the observed monthly flow volumes (Figure 6.13). Both low and high flows are well simulated and fall within the (+/-5%) of the percentage bias for both natural and log transformed volumes (Table 6.3). The statistics for the headwater sub-basins (e.g 1ka32a) within the same group are poor while the statistics for downstream sub-basins are good (e.g 1ka31). The results are affected by extreme events that are not seen in the observed time series but the model has been able to simulate such events (e.g El-Nino years 1961, 1968 and 1997/1998). Moreover, the model under-simulates high flows in the post-1990 period, indicating that that there is a step change in the observed time series, hence the pre- and post-1990 periods are different and this could be related to rating curve problems.

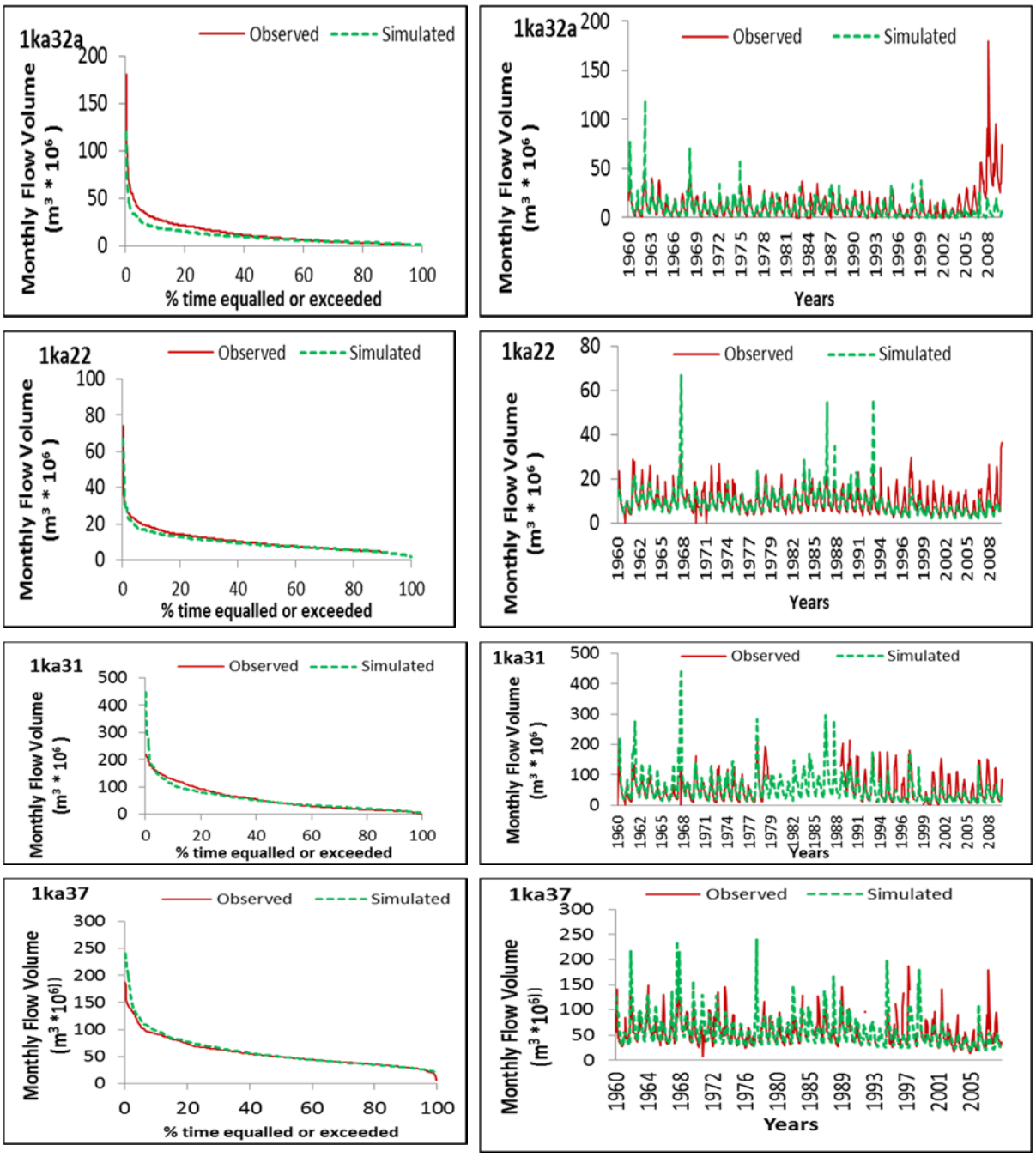


Figure 6.13 Observed and simulated monthly stream flow for gauging station 1ka32a, 1ka22, 1ka31 and 1ka37

Group four

Comparisons between the observed flow data and the simulated flow are poor, largely due to the difficulty of calibrating a semi-arid catchment with poor data. From the visual inspection of the hydrographs, the model is over simulating both high and low flows (high positive percentage bias values). It has been very difficult to get rid of the high flows through calibration and this problem could be related to uncertain rainfall data. The time series plots and flow duration curves for catchment 1ka42 are given in Figure 6.14

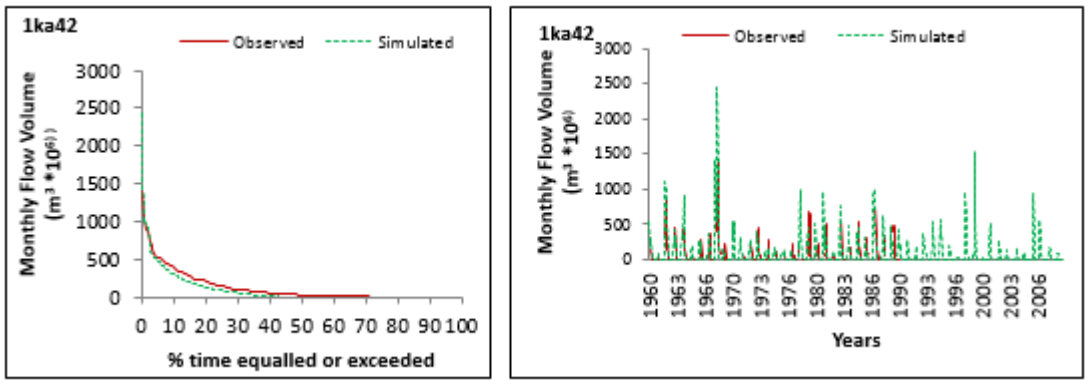


Figure 6.14 Observed and simulated monthly stream flow for catchment 1ka42

Group five and six

The Nyaluhanga gauging station (1ka71) is located in the wetland whereas Msembe ferry gauging station (1ka59) is downstream of 1ka71 below the Usangu wetlands. The observed and simulated discharges at Nyaluhanga for the calibration periods are illustrated in Figure 6.15. The statistics of the model performance for the 1ka71 are good indicating good correlation between observed and simulated flows This sub-basin is influenced by the flood plain and water use abstractions from upstream sub-basins, therefore the good or bad performance of the model depends on how well water use abstractions and wetland processes have been accounted for in addition to the quality of input data, and assigned parameters and constraints bounds.

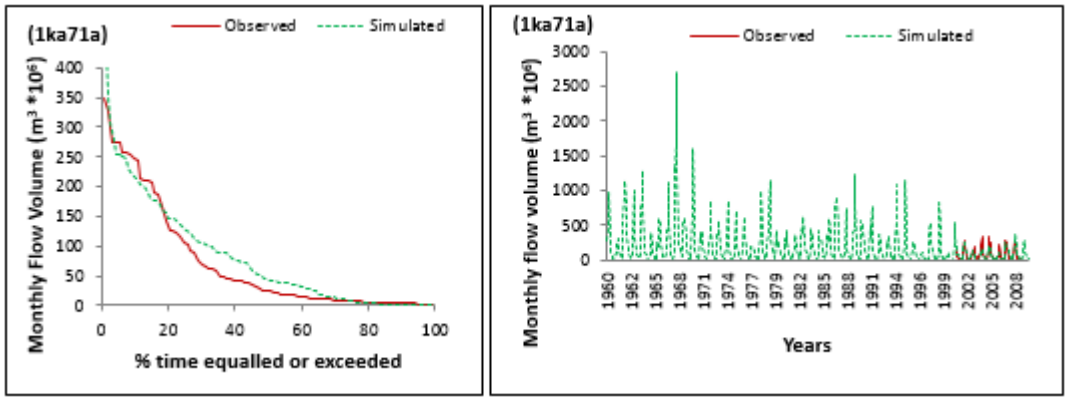


Figure 6.15 Observed and simulated monthly stream flow gauging station 1ka71

Gauging 1ka59 is regarded as the most downstream station of the upper Great Ruaha sub-basin which includes the Usangu wetlands. The catchment is influenced by both the wetlands and the irrigation schemes upstream of the Usangu plain and within the alluvial fans. In order to assess the irrigation impact on the downstream stream flow, the model was first run without water use abstractions (Figure 6.16), and water use abstractions were then added to the model. Figure 6.17 show the results of the model run with and without water use abstractions. Table 6.4 illustrates the comparisons of the model performance between the simulated and observed flow data for catchment 1ka59 for both the natural hydrology and with the inclusion of water use. The results shows that water use abstraction have significant impact on the downstream flow (Figure 6.17) and the percentage for both low and high flows have improved. The results of the model run with wetlands are shown in the subsequent section.

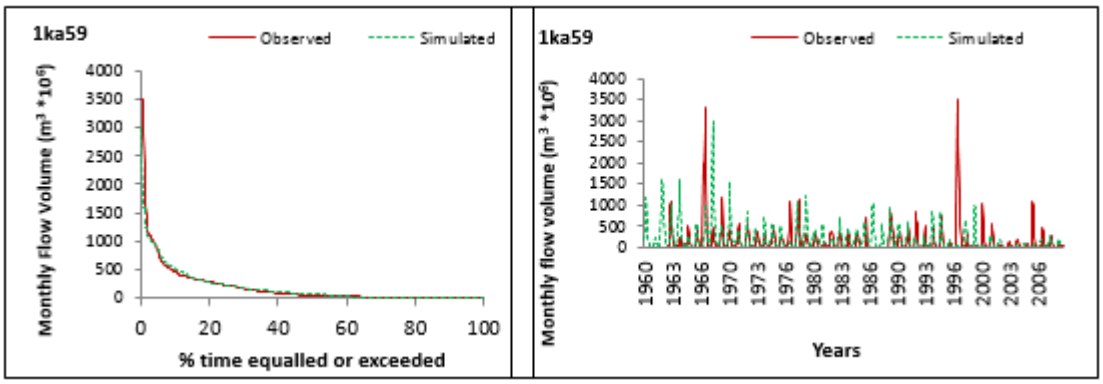


Figure 6.16 Observed and simulated monthly stream flow (without water use) for gauging station 1ka59

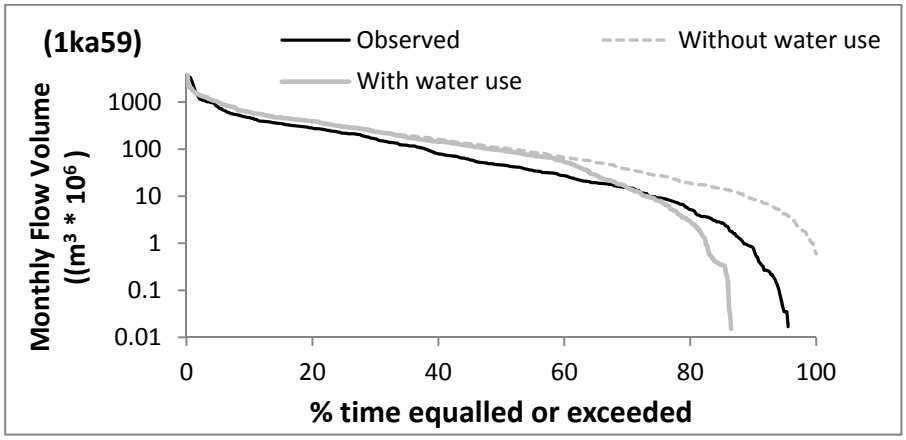


Figure 6.17 Comparisons of the observed flow data (natural hydrology) with the simulated data for the natural hydrology and the simulated flow with the incorporation of water use for sub-basin 1ka59

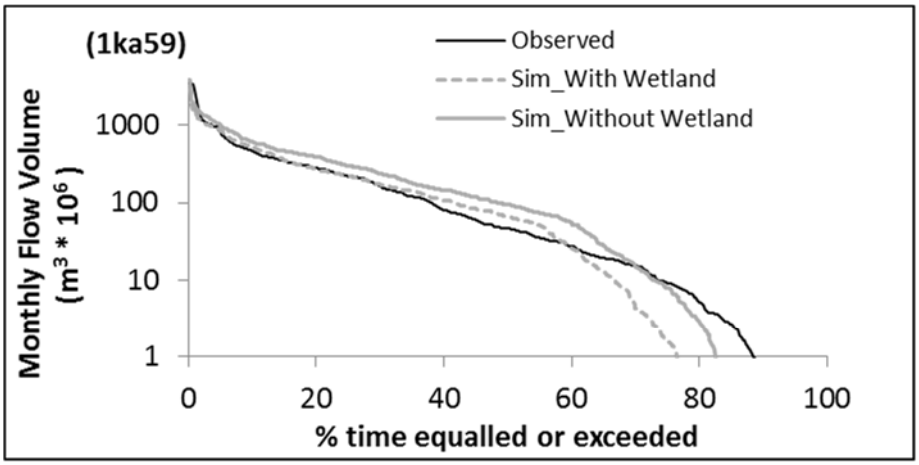
6.8 Accounting for wetland processes

This part of the study presents an attempt to simulate the surface hydrology of the Great Ruaha River sub-basin incorporating wetland processes into the existing GW_PITMAN rainfall runoff model (Chapter 3).

6.8.1 GW_PITMAN set up and calibration for the Usangu wetlands

The Usangu plains lie within the upper catchment of the Great Ruaha River. The inflow to the wetlands is supplied by 18 rivers, originating from the highlands, which are fed by high rainfall in the wet season. The main rivers meander across the plains frequently changing their courses and beds (CCKK, 1982). The western part of the wetland is seasonally flooded while the eastern part of the wetland is flooded throughout the year. The wetland acts as a flood plain in that, water flows in the river channel to the bankfull level, and any excess flow leads to overflow of the river bank. When water spreads over the flood plain, a substantial amount is lost through evaporation. The plains are characterized by erratic rainfall of about 400 to 700 mm y^{-1}, however the western end of the plains benefit from orographic effects of the Poroto and Chunya mountains and has rainfall up 800 mm y^{-1}.

The wetland parameters (Hughes et al., 2013d) were included in sub-basins 1ka71a and 1ka27, representing the Usangu Plains. The wetland parameters (Section 6.4) were included after an initial run of the model and an evaluation of the results against observed stream flow data at 1ka59. The wetland parameters were manually calibrated but within realistic parameter limits based on the documented characteristics of the two main wetland areas (Section 6.4). Figure 6.18 illustrates the results before and after the inclusion of the wetland functions. The results show that most of the components of the hydrological regime are well represented although the model over-simulates peak flow for the earlier period and under-simulates both peak and low flows for the late period. The model has also failed to capture some extreme events (e.g El-Nino events in 1967/68 and 1997/98). The reason for under-simulation during the period from 2000-2007 could be related to management interventions at the basin scale where there is improvement is water use efficiency in the basin and a reduction in water use for irrigation (Figure 6.19A and B). On the other hand, the overall under-simulation of low flows could also be due to over-estimation of irrigation abstractions for upstream sub-basins. Despite these observations the addition of the wetland function improved the overall results at the outlet of 1ka59.



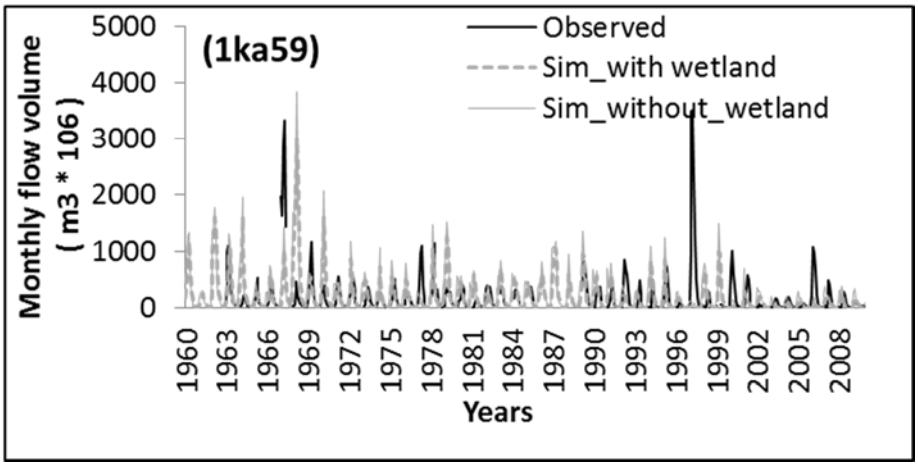


Figure 6.18 Observed and simulated flows (Before and after inclusion of wetland functions) for catchment 1ka59

6.8.2 Quantification of the uncertainties in wetland model simulations

There is large amount of uncertainty associated with understanding the dynamics of wetlands processes. The model was set up using available information as well as from estimations of spatial data sets and google earth. Channel losses were also added to sub-catchment 1ka59, however this transmission loss function is also subject to uncertainty. The estimated wetland parameters are also subject to uncertainty. Figure 6.19 shows the uncertainty estimation with incorporation of wetland parameters and water use abstractions.

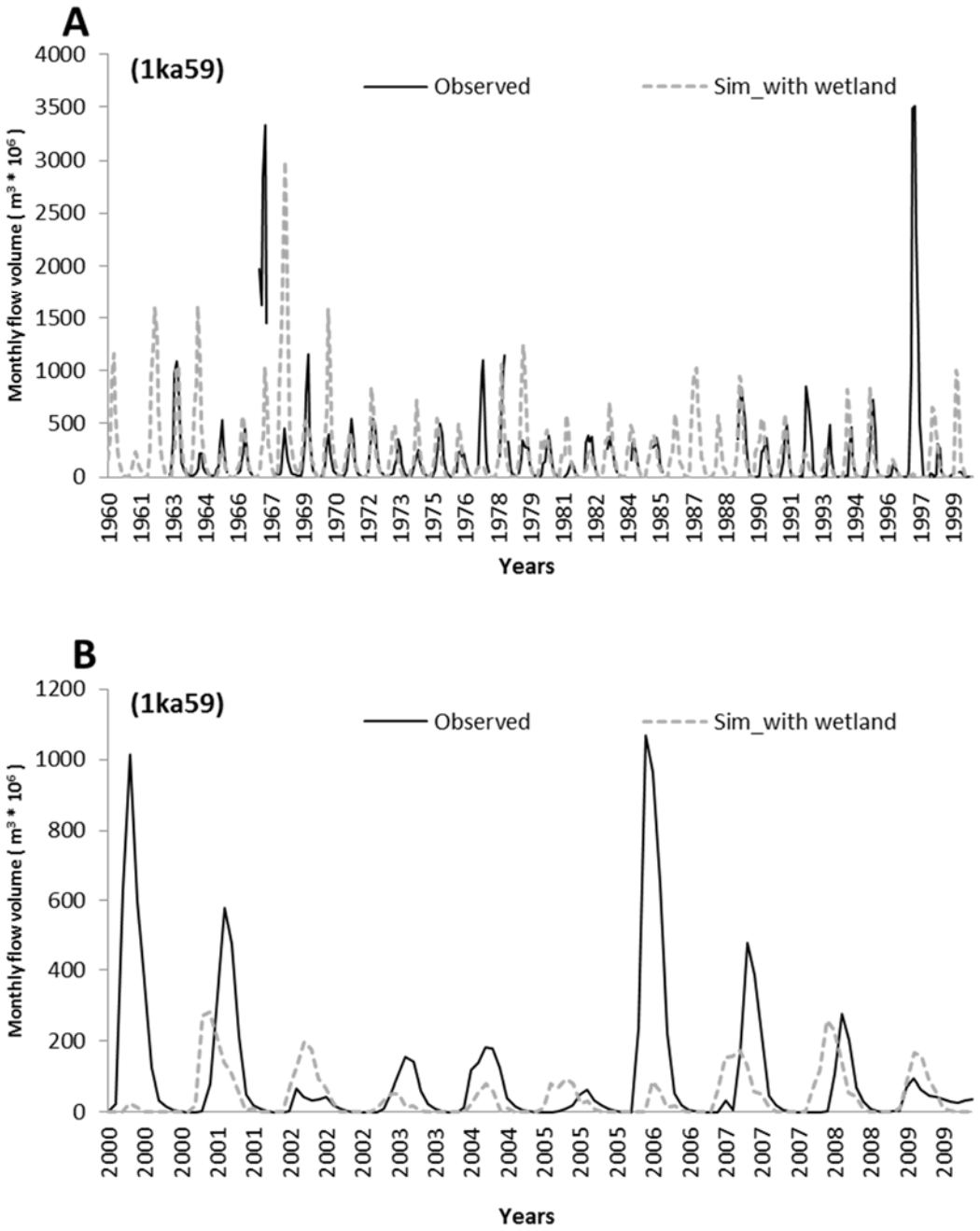


Figure 6.19 Observed and simulated flows prior 2000 (A) showing a good fit to low flow and high observed flows for the period post 2000 (B) for catchment 1ka59

Based on the relative spread of the simulated ensembles around the observed value, the uncertainty is well presented in high and moderate flows (The band of uncertainty encompasses the observed flow). There is more relative uncertainty in the simulation of low flows than for moderate to high flows and this is largely a consequence of the uncertainty

caused by the wetland effects and estimations of water use abstractions. Incorporating wetland parameters into the model simulation clearly lowered the simulated low flows although the uncertainty band is still wide.

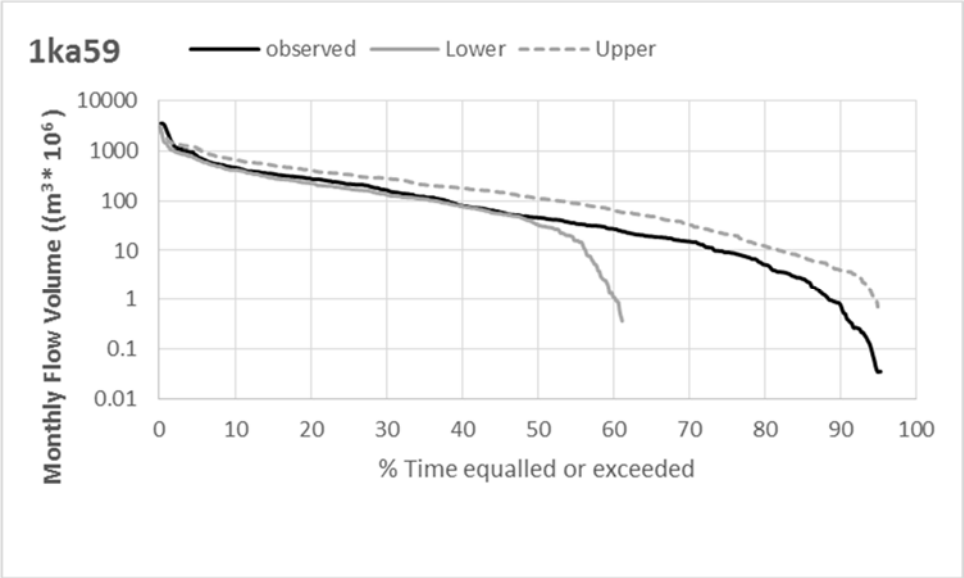


Figure 6.20 Uncertainty estimation for 1ka59 with the incorporation of wetland parameters (simulated with water use abstractions)

Figures 6.18 and 6.20 indicate that while the FDC results are generally adequate and the model is able to reproduce the patterns of inter-annual variability, the simulations of individual monthly flow volumes are not very good. This is quite a typical result in regions where there are likely to be large spatial variations in rainfall, but where there are not enough data to quantify the variability at the model time scale of 1 month. While the use of satellite rainfall data may be useful to fill some of the gaps, satellite rainfall data are not sensitive to orographic controls on rainfall and therefore cannot be expected to completely make up for deficiencies in ground-based observations in topographically diverse regions (Sawunyama and Hughes, 2008). The assessments of the validity of the ensemble simulations relative to the observed data at the outlet of 1ka5a (Mtera Dam: and the available data pre-date the dam construction in 1984), the other critical site from a water resources management perspective, yielded similar conclusions as those reached for 1ka59. There are ensembles that can be considered acceptable representations of the observed data, although there are many with some months of apparently excessive high flows. Most of these are generated within the Kisigo River sub-

basins upstream of 1ka42 and are difficult to ascribe to any obvious model parameter or rainfall errors.

6.9 Discussions and conclusions

The constraints were assigned to the 48 sub-basin within the GRR based on six (6) regional groups that were assumed to have similar response characteristics. This process was guided by available information such as physical basin characteristics, regional flow duration curves and geomorphological zones. The physical basin characteristic information was extracted from global data sets and therefore this broad scale information might not be representative of the characteristics of specific sub-basins. The same is true for the regional flow duration curves that were developed from observed stream flow data and there are a number of uncertainties associated with rating curve and stage observation errors. However, this was the best information available for the quantification of constraints. The process of assigning constraint ranges for specific groups was not straightforward due to the lack of relevant information to guide the whole process. The constraints in group 2 were more difficult to determine in that, most of the sub-basins are ungauged and the group contains a mixture of headwater and foothill response zones. Therefore, both the upper and lower ranges of this group were very uncertain. Group 3 is characterized by quite large differences in recharge (MMR) and low flow constraints (Q90/MMQ). Due to the paucity of data, recharge constraints were difficult to establish and assign to individual sub-basins and therefore the constraint ranges were large, however they were aligned with the Q90/MMQ constraint. Despite these challenges, the developed constraints proved to be adequate to generate the 2 000 behavioural parameters. However, there are a number of situations where the simulated ranges are lower than the input ranges (e.g. group 2) where the regional uncertainty is high. This could be related to differences in climate inputs to the model across sub-basins that were assumed to have similar responses.

The initial parameter ranges were guided by information from the manual calibration and physical basin characteristics information. The parameter ranges were assigned to the six (6) regional groups. It should be noted that the input parameter ranges might be the same for a group, but the output ranges may differ between sub-basins within a group. In some cases the

full parameter and constraint ranges were not represented in the 2 000 outputs for some sub-basins (e.g the uncertainty associated with GW parameter and recharge constraints). This could be associated with too high or too low constraint bounds in relation to the input parameter bounds for a specific sub-basin, relative to the group as a whole. However, this does not represent a failure of the modelling approach, but reflects the uncertainties in extrapolating from a limited number of observed sites (where the observations themselves are uncertain) to ungauged sub-basins. An assessment of the input and output values for a combined index of the parameters determining moderate to low flow responses (FT/POW+GW/GPOW) confirms the equifinality in the model and that the use of constraints can reduce the equifinality problem to a certain extent (Figure 6.4B) in that the very high and very low combinations of the index are not accepted as behavioural.

Objective functions based on a single (optimum) ensemble and ranges of objective functions values have been presented and it is clear that there are both good and poor results. Some of the poor results are associated with over-estimation of high flows (large percentage biases values) and one possible reason is that the use of the Q10/MMQ constraint does not adequately constrain very high flows during step 1 of the uncertainty approach (Chapter 4). This result could also be associated with uncertainties in the interpolated rainfall data or inadequacies in the rating curve data at very high flows.

The approach employed in this study represents a major step towards the identification of uncertain, but behavioural, parameters based on understanding rather than on calibration to local observations. The results from this assessment suggest that the uncertainty approach that has been adopted in this study is appropriate for the GRR basin given the available input data and the large spatial and temporal variability in both climate and geology. Starting from a situation with very little knowledge about the physical basin characteristics and basin processes, the quality of input data and initial parameter ranges, the constraints based on observed stream flow data have been developed to constrain the uncertainty in input data and parameters. Throughout the modelling process the dependence of the modelling results on the level of input data uncertainty has been emphasised. It is likely that (unknown) combinations in uncertainties in data errors and the quantification of regional constraints

significantly influence some of the poor results in some sub-basins and therefore the refinement of the regional constraints might help in improving the simulation results and narrowing the uncertainty bands in some sub-basins with large uncertainties. The conclusions from this chapter can be summarized as follows:

- The use of the constraints for guiding model setup and uncertainty analysis in both gauged and ungauged basins has advantages for use in regions where data for model calibration are scarce. However, there is a need to quantify the uncertainty associated with the observed stream flow data.
- The uncertainty analysis approach using regional constraints on hydrological response have proved to be useful in organising a substantial amount of uncertain information into a coherent model of the whole GRR basin. Through this approach a better understanding of the basin's hydrological response has been developed. Arguably, the approach has also helped to gain a better understanding of the sources of uncertainty and therefore where future data collection should be focussed.
- There is high degree uncertainty which could not be resolved in the western headwaters due to a lack of both rainfall and stream flow gauging stations.
- Results for the other parts of the basin and downstream sub-basins such as 1ka59 and 1ka5a are positive and provided a basin scale context of the spatial and temporal variations in simulated hydrological response.
- Availability of more input data (and the quality thereof) could enable a better determination of the constraint ranges and increase the quality of both input and output uncertainty estimations.
- Further adjustment and alignment of the parameters and constraints is recommended for future work.

7 SWAT APPLICATION

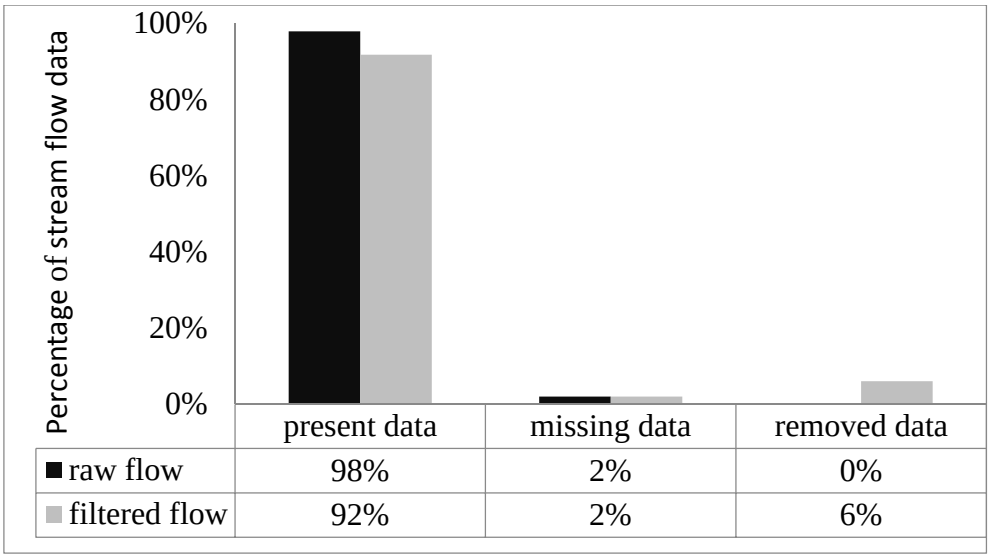
7.1 Introduction

The Soil and Water Assessment Tool (SWAT) is a hydrological simulation tool that is widely used by researchers and postgraduate students in Tanzania. This could be attributed to the free online spatial dataset (elevation, soil, land cover) necessary for setting up SWAT. However, the ease of setting up SWAT using the available information does not mean that the model will give behavioural results. It is clear that the calibration of hydrological models for water resources assessments is often difficult due to the large numbers of model parameters, and the difficulty increases with the model complexity. Similarly calibration and uncertainty analysis are a pre-requisite of any hydrological modelling study. Despite, the claimed wide use of SWAT in Tanzania, the whole issue of uncertainty has been ignored, where the uncertainty framework within SWAT is used for the optimisation of objective functions only (van Griensven et al., 2012). In this study the Sequential uncertainty fitting (SUFI-2) approach was combined with SWAT to quantify parameter uncertainty of the stream flow simulations for the Little Ruaha River (5 195 km^2). SWAT2009 was setup for the whole GRR basin but the analysis presented here is based on one major tributary only. The hydrological response units (HRU) were characterised using the dominant land use, soil and slope in order to keep the complexity of the analysis to a practical limit for the uncertainty propagation (Chapter 4). In this study, SWA2009 was used to explore the implementation of the uncertainty analysis framework for the meaningful application of the results.

7.2 Calibration approach

The gauging station 1ka31 (Little Ruaha at Mawande) was used for SWAT2009 model calibration for the period 1971-1979. Daily stream flow data from this station were checked for quality, and this involved the identification of errors from unexplained extreme values. Figure 7.1 illustrates the percentage of available data points, missing data points, and removed data points. Six percent (6%) of the data was deleted from the time series and two percent of the data was missing, therefore only 92% of the record was used for the calibration method. Both manual and automatic calibration approaches were used for this study. The pre-

calibration parameter sensitivity analysis was performed to identify parameters that are expected to have a strong influence on the model simulation results. The sensitivity analysis in this study was done using (i) automatic global sensitivity analysis in SUFI-2 ii) manual analysis of the sensitive parameters based on the output of the global sensitivity analysis.



	present data	missing data	removed data
raw flow	98%	2%	0%
filtered flow	92%	2%	6%

Figure 7.1 Summary of the screened daily stream flow data

The global sensitivity analysis in SUFI-2 is not able to analyse all the parameters in SWAT, it analyses the sensitivity of the pre-defined 27 parameters (Table 7.1). In this approach, parameter sensitivity is determined using the multiple regression equation, which regresses the Latin hypercube generated parameters against objective function values. The t-Stat and p-Value are statistical measures used to evaluate sensitivity in SWAT-CUP. A *t*-stat is used to identify the relative significance of each parameter by providing a measure of sensitivity (larger absolute values are more sensitive). p-Values determined the significance of the sensitivity where a value close to zero has more significance. The sensitivity analysis was followed by both manual and automatic calibration. The manual calibration was performed based on the understanding of the sub-basin characteristics. The results of the global sensitivity analysis gave an indication of the sensitive parameters and helped to guide the initial parameter ranges. The calibration procedure involved the following steps:

1. Sensitivity analysis
2. Manual calibration

3. SUFI-2 set up (automatic calibration)
4. Assigning initial parameter ranges
5. Latin Hypercube sampling is used to sample the parameter distributions
6. Model simulations are performed and objective functions are calculated for each of the n (n = 2 000 for this study) simulations

7.3 Parameter sensitivity analysis

Table 7.1 illustrates the parameter sensitivity ranking and values for t-statistics and p-statistics. The most sensitive parameters (highlighted in grey) for flow are sequentially ALPHA_BF, CN2, SURLAG, REVAPMN, CH_K2, GWQMN, SLSUBBSN, BLAI, and CANMX. Most of the sensitive parameters are the ones governing the surface and subsurface hydrological processes and stream flow routing. In literature, most of these parameters have been identified as sensitive for the SWAT model (van Griensven et al., 2006; Gassman et al., 2007). Parameter sensitivity in SWAT is influenced by climate, land use, topography and soil types. This implies that sensitivity analysis needs to be performed separately for each sub-basin since the results cannot be generalised (van Griensven et al., 2006).

Table 7.1 Parameter sensitivity ranking and category of the most sensitive parameters

Parameter	Description	t-stat	p-stat	Rank	Process
ALPHA_BF	Base flow alpha factor for recession constant (days)	-34.23	0.00	1	Ground water
CN2	SCS runoff curve number for moisture condition II	-12.90	0.00	2	Runoff
SURLAG	Surface runoff lag time (days)	-1.54	0.12	3	Runoff
REVAPMN	Threshold water depth in the shallow aquifer for revap (mm)	-1.51	0.13	4	Groundwater
SOL_K(2)	Saturated hydraulic conductivity soil layer 2 (mm h^{-1})	-1.39	0.16	5	Soil
GWQMN	Threshold water depth in the shallow aquifer for flow (mm)	-1.28	0.19	6	Groundwater
SLSUBBSN	Average slope length (mm^{-1})	1.17	0.19	7	Topography
BLAI	Leaf area index for crop	1.05	0.29	8	Crop
CANMX	Maximum canopy storage (mm)	0.60	0.54	9	Runoff
CH_N2	Manning's ''n'' value for the main channel	0.58	0.55	10	Channel
HRU_SLP	Average slope steepness of the HRU	-0.56	0.57	11	Topography
GW_REVAP	Groundwater ''revap'' coefficient	-0.46	0.63	12	Groundwater
BIOMIX	Biological mixing efficiency	-0.39	0.69	13	Soil
EPCO	Plant evaporation compensation factor	0.24	0.80	14	Evaporation
SOL_AWC	Available soil water capacity (mm H_20/mm soil)	0.21	0.82	15	Soil

RCHRG_DP	Deep aquifer percolation fraction	-0.21	0.83	16	Groundwater
ESCO	Soil evaporation compensation factor	-0.10	0.91	17	Evapotranspiration
GW_DELAY	Movement of water from shallow aquifer to the root zone	0.09	0.92	18	Groundwater
CH_K2	Channel effective hydraulic conductivity (mm h^{-1})	0.07	0.94	19	Channel
SOL_K(1)	Saturated hydraulic conductivity of soil layer 1 (mm h^{-1})	-0.04	0.09	20	Soil

7.4 Assigning initial parameter ranges

The results of the sensitivity analysis gave an indication of the sensitive parameters and helped in guiding the setup of the initial parameter ranges. It was important to consider the physical meaning of each parameter and its effects on the sub-basin behaviour. Therefore, the initial parameter sets were guided by the understanding of the physical basin characteristics and the default upper and lower limits established in SWAT (Nietsch et al., 2005). In SWAT default parameters can be modified for the whole sub-basin (lumped), or in a distributed way for individual sub-basins or hydrological response units. Table 7.2 shows the initial parameter ranges of the sensitive twenty parameters, where the most sensitive parameters are highlighted in grey.

Table 7.2 Defined upper and lower limits of initial parameter ranges, the extension of the files in which they are located, and the option used for carrying out changes

Parameter	Lower limit	Upper limit	Change option
v__ALPHA_BF.gw	0.00	1.00	replacement
r__CN2.mgt	30	80	relative
v__SURLAG.bsn	0.00	24.00	replacement
v__REVAPMN.gw	0.11	0.80	replacement
r__SOL_K(2).sol	0.39	4.28	relative
a__GWQMN.gw	1 983	2 889	absolute
r__SLSUBBSN.hru	0.13	0.33	relative
v__BLAI{120}.CROP.DAT	3.63	6.95	replacement
v__CANMX.hru	2.87	8.51	replacement
v__CH_N2.rte	0	0.3	replacement
r__HRU_SLP.hru	0	10	relative
a__GW_REVAP.gw	0.02	0.12	absolute
r__BIOMIX.mgt	0.11	0.69	relative
v__EPCO.hru	0	0.4	replacement
r__SOL_AWC(2).sol	0	0.9	relative
v__RCHRG_DP.gw	0	1	replacement
v__ESCO.hru	0	1	replacement
a__GW_DELAY.gw	0	129	absolute
v__CH_K2.rte	24.27	94.18	replacement
r__SOL_K(1).sol	0.66	5.55	relative

7.5 Parameter distributions

The identifiability of parameters was examined visually using scatter plots of model parameter values versus CE. Figure 7.1 show scatter plots with the values of each parameter defined versus their corresponding Nash-Sutcliffe efficiency (CE), where the parameter values were obtained from Latin Hypercube sampling of the initial range defined using 2 000 simulations. Scatter plots of the parameter values versus objective function were used to examine the

identifiability of individual parameters. Based on the scatter plots the identifiable parameters are expected to show a distinct maximum and lack of a distinct maximum indicates the difficulty in getting the optimal values that gives good model performance therefore the parameter becomes poorly identifiable. It is evident that none of the parameters are identifiable. However, it should be noted that non-identifiability of a parameter does not indicate that the model was not sensitive to these parameters. The sensitivity analysis results identifies the most sensitive parameters to be considered for calibration but does not consider the interactions between parameters, therefore having the most sensitive parameters does not mean that the parameter will be identifiable. Estimation of non-identifiable parameters is difficult because there may be many combinations of these parameters that would result in similar model performance (equifinallity). A number of factors might have led to the non-identifiability of parameters in this study. The interactions between parameters may have contributed to the equifinallity which might be associated with the simplified representation of the sub-basin (dominant HRU). Interactions between soil parameters (soil depth and available water capacity) and ground water parameters (Groundwater delay) is expected in SWAT. It is hard to explain these interactions since SWAT considers two soil layers (root zone and unsaturated zone) and ground water (conceptual shallow and deep aquifer stores) and there is not enough information regarding sub-surface water processes to will enable a better explanation of the parameter interactions.

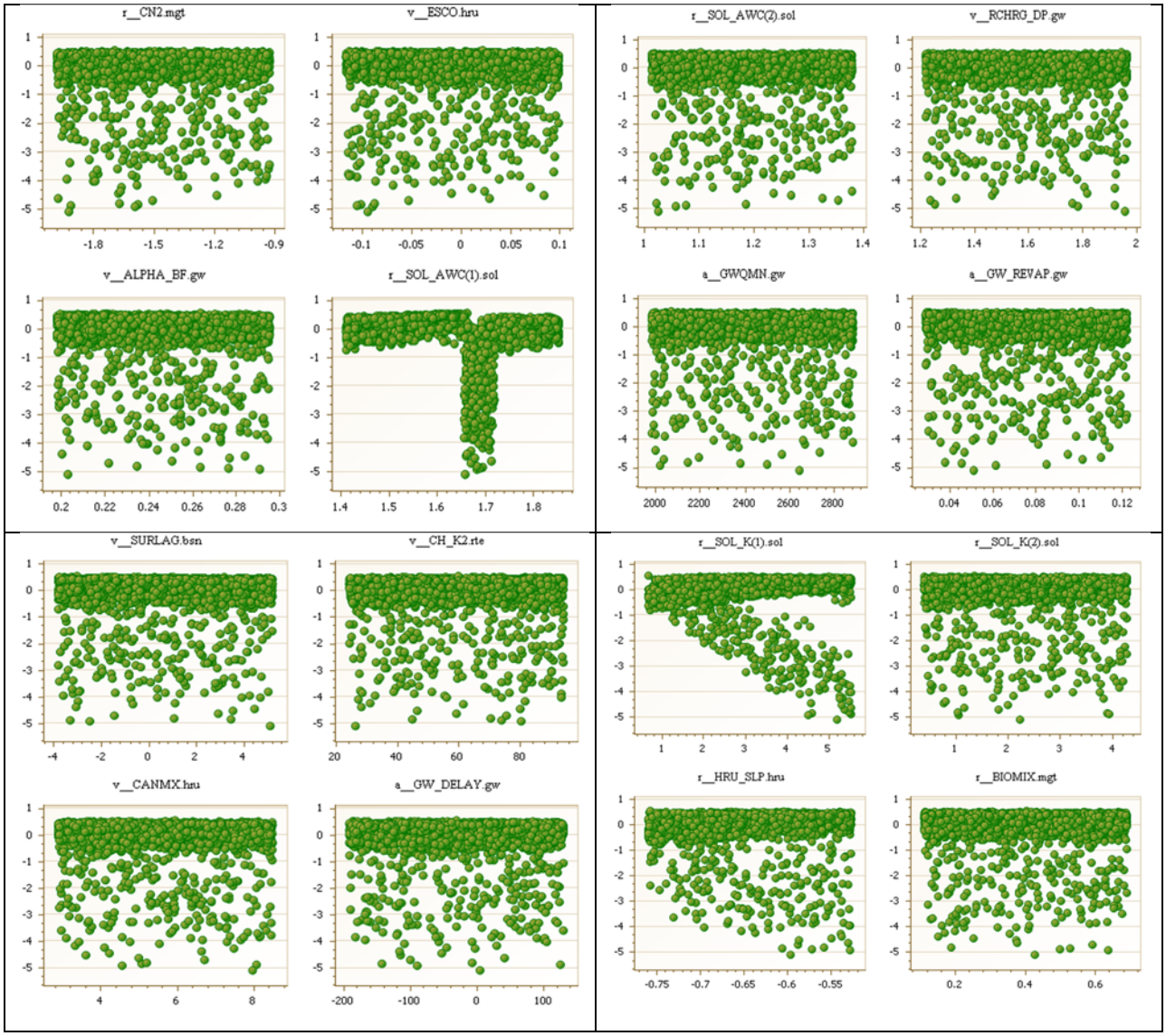


Figure 7.2 Scatter plots of the calibrated parameers of Little Ruaha River sub-basin (1ka31) versus Nash-Sutcliffe efficiency, obtained from Latin Hypercube sampling of the initial large range using 2 000 simulations

7.6 Final calibrated parameter ranges

Latin hypercube sampling was used to sample parameters within the initial ranges using 2 000 ensembles and a uniform distribution. The CE was used to get optimum parameter values and to separate behavioural from non-behavioural parameter sets, where a cut off limit of CE=0.45 was used. Table 7.3 shows the parameter range and optimal value for the best simulation. ALFA_BF is the most sensitive parameter followed by CN. ALPHA_BF parameter is a direct index of ground water flow response to changes in recharge. The ALPHA_BF value between 0.1 and 0.3 reflects an area with slow response to changes in flow, a value of 0.9 to 1 reflects an area with a rapid response to changes in flow. For the Little Ruaha sub-basin a value of 0.25 was obtained. The CN is the parameter that determines the amount of runoff to be generated from a sub-basin so it was expected to be sensitive for the Little Ruaha sub-basin with an optimal value of 48. SURLAG was the third most sensitive parameter, and is the fraction of runoff that reaches a sub-basin outlet on any given day. SURLAG was sensitive for this sub-basin because of the low time of concentration, and an optimum value of 3.5 days was obtained. REVAPMN presents the threshold depth of water in the shallow aquifer for return flow to the root zone to occur. This parameter is most important in areas where the water table is high or areas with deep rooted crops. An optimum value of 0.57 was obtained. SOL_K(2) is the saturated soil hydraulic conductivity (mm h^{-1}). In this study, a SOL_K value of 1.66 mm h^{-1} was used. This parameter relates water flow rate to the hydraulic gradient and is a measure of the rate of water movement though the soil.

The GWQMN is the threshold water level in shallow aquifer for return flow to occur (mm). The ground water flow to the main channel is allowed only when the depth of water in the shallow aquifer is equal to or greater than the threshold depth of water in the shallow aquifer required for the return flow to occur. An optimum value of 2071.38 (mm) was obtained. The obtained value for the mean slope steepness of the basin (SLSUBBSN) is 0.32, indicating that the sub-basin is influenced by low to moderate slopes and has implications for the runoff generation process. The optimum value for the maximum potential LAI is 4.82. The value corresponds to the MODIS data which indicates LAI for the Little Ruaha catchement ranges from low to moderate values (Figure 7.3). CANMIX represents the maximum canopy area, and an optimum

value of 5.95 mm was obtained. This value corresponds to the leaf area index indicated in Figure 7.3. The Manning roughness coefficient ''n'' for channel flow (CH_N(2)) is the parameter that influences channel roughness, an optimum value of 0.06 was obtained.

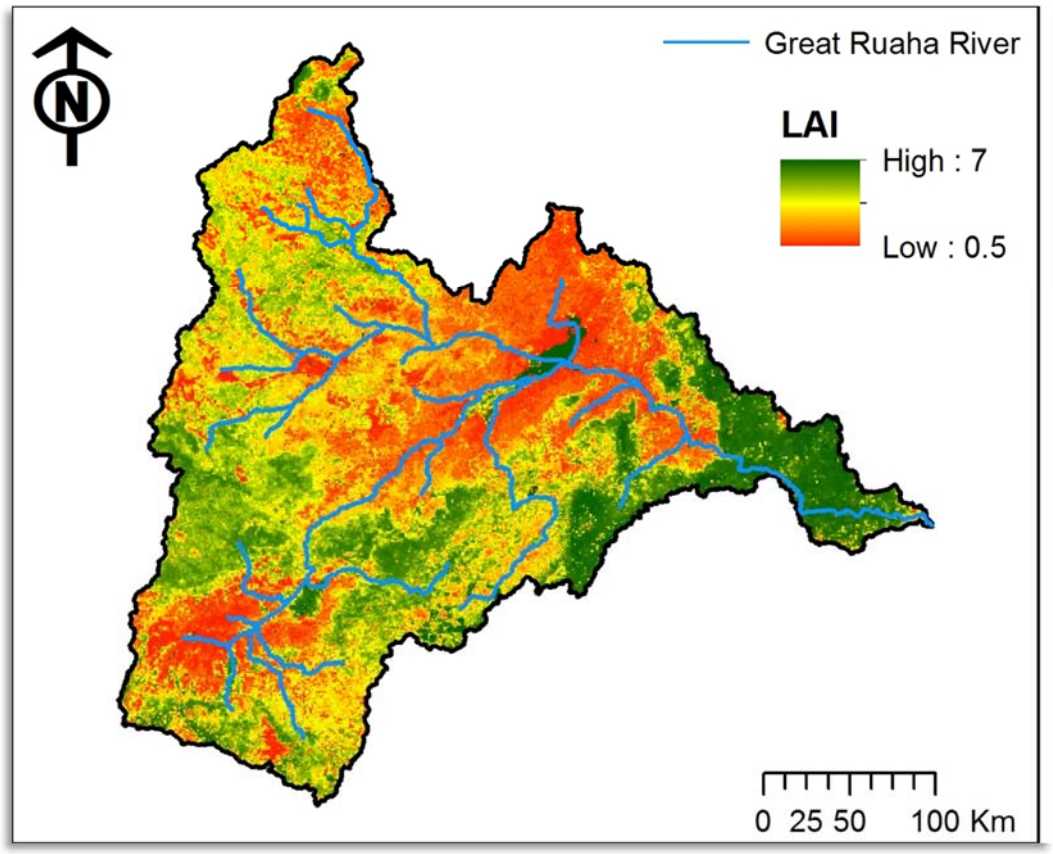


Figure 7.3 Spatial variations in leaf area index within the GRR basin

Table 7.3 Final parameter ranges calibrated using SUFI-2

Parameter name	Lower limit	Upper limit	Optimal SUFI-2
v__ALPHA_BF.gw	0.00	1.00	0.25
r__CN2.mgt	30	80	48
v__SURLAG.bsn	0.00	24.00	3.5
v__REVAPMN.gw	0.11	0.80	0.57
r__SOL_K(2).sol	0.39	4.28	1.36
a__GWQMN.gw	1983	2887.18	2071.38
r__SLSUBBSN.hru	0.13	0.33	0.32
v__BLAI{120}.CROP.DAT	3.63	6.95	4.82
v__CANMX.hru	2.87	8.51	5.95
v__CH_N2.rte	0	0.3	0.06
r__HRU_SLP.hru	0	10	0.75
a__GW_REVAP.gw	0.02	0.12	0.10
r__BIOMIX.mgt	0.11	0.69	0.40
v__EPCO.hru	0	0.4	0.004
r__SOL_AWC(2).sol	0	0.9	1.10
v__RCHRG_DP.gw	0	1	1.94
v__ESCO.hru	0	1	0.02
a__GW_DELAY.gw	0	129	-31.05
v__CH_K2.rte	24.27	94.18	59.94
r__SOL_K(1).sol	0.66	5.55	0.66

* (prefix v-indicates that the parameter value is replaced by a given value, prefix r-indicates that the parameter value is multiplied by (1+ a given value), and prefix a-indicates that the parameter value is added to a given value).

7.7 Model simulations results and uncertainty analysis

SWAT was calibrated against observed data for gauging station 1ka31 for the period 1970-1971. Calibration results yielded satisfactory results given the data scarcity. CE and R^2 values of 0.54 and 0.62 were achieved for the calibrated period. The P-factor (% of measured data bracketed by 95% prediction uncertainty) was 0.58 and 0.21 for the full range and behavioural simulations, respectively. The R factors for the full range and behavioural parameters were 1.91 and 0.36, respectively. These results confirm quite large uncertainty of the simulated discharge due to the large equifinality in parameters and reliability of input data (precipitation and daily evaporation data). Table 7.4 shows a summary of model performance for the calibrations and a comparison between all parameter sets (full range) and behavioural parameter sets. In presenting results the following performance measures were used;

- The relative distance between the observed data and the 95PPU (R-factor)
- The percentage of observations covered by the 95PPU (P-factor)
- Nash-Sutcliffe Efficiency (CE)
- Coefficient of Correlation (R^2)

Table 7.4 Summary of performance statistics for the best simulation

Station	Simulations	P-factor	R-factor	CE	R^2
1ka31	Full range	0.58	1.91	54%	62%
	Behavioural	0.21	0.36	54%	62%

The uncertainty analysis procedures were described in Chapter 4. Uncertainty analysis was implemented using the SUFI-2 algorithm. Figures 7.4 and 7.5 show the results of the daily flow uncertainty analysis carried out in the sub-basin for the full range and behavioural parameter sets respectively. The shaded area represents the 95% predictive uncertainty (95PPU), whereas the blue lines correspond to the observed discharges and the red lines correspond to the simulated flow at the sub-basin outlet. For the full range simulations (Figure 7.4) it was found that the observations fall within the lower and upper 95% prediction uncertainty in high and moderate flow but with large uncertainty. Figure 7.5 show that the 95% prediction uncertainty of behavioural simulations (CE ≥ 45%) does not bracket the observed flow, only

15% of the data were bracketed, indicating that some processes are not well represented in the model. The prediction limits obtained with SUFI-2 are highly dependent of the threshold selected to separate behavioural from non-behavioural parameter sets. It is also important to note that in SUFI-2 parameter uncertainty is presented as a uniform distribution in the final parameter range, while parameter interactions are ignored and contribute to the large equifinality observed in these results.

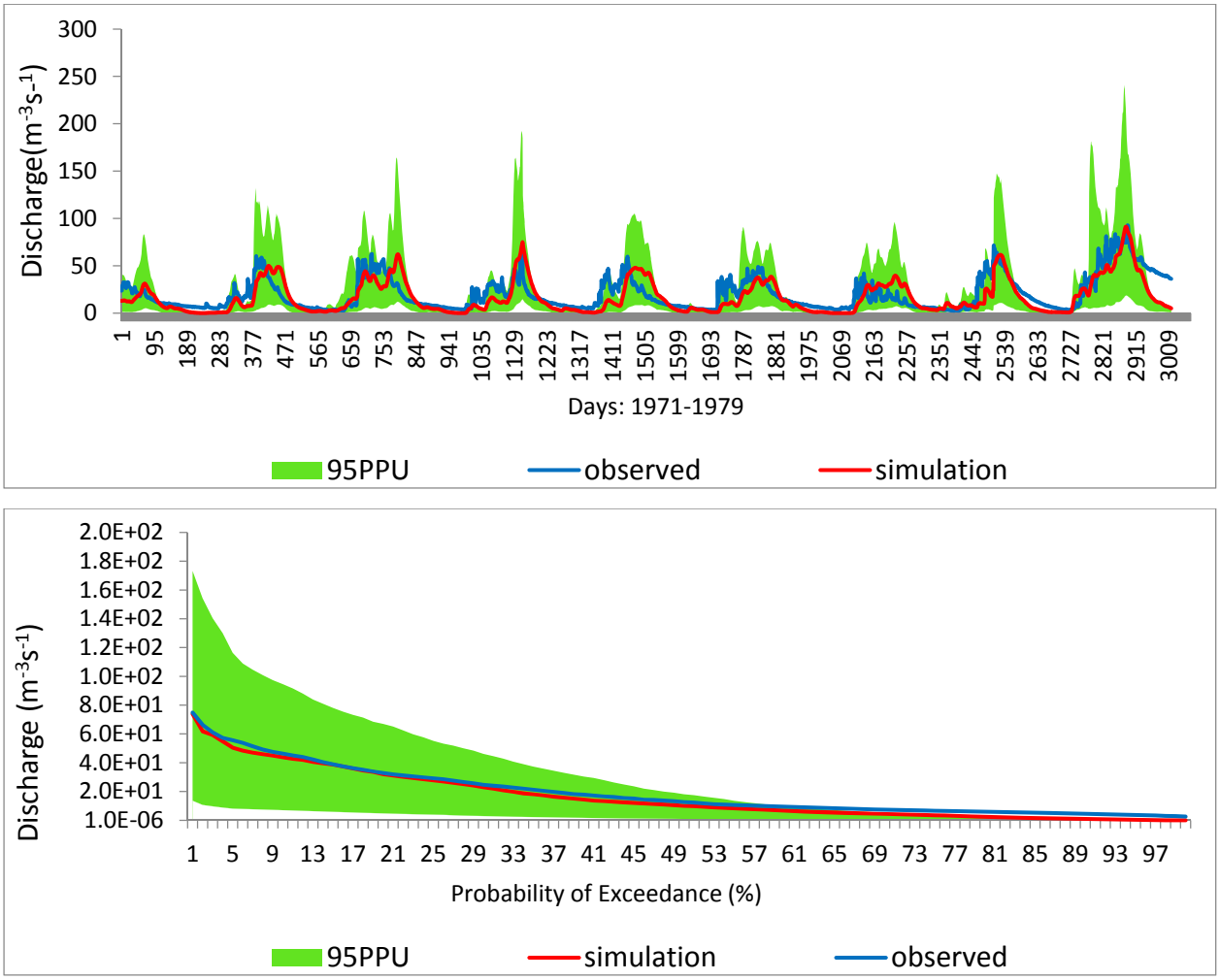


Figure 7.4 Calibration at 1ka31-Mawande (95PPU for full range simulations)

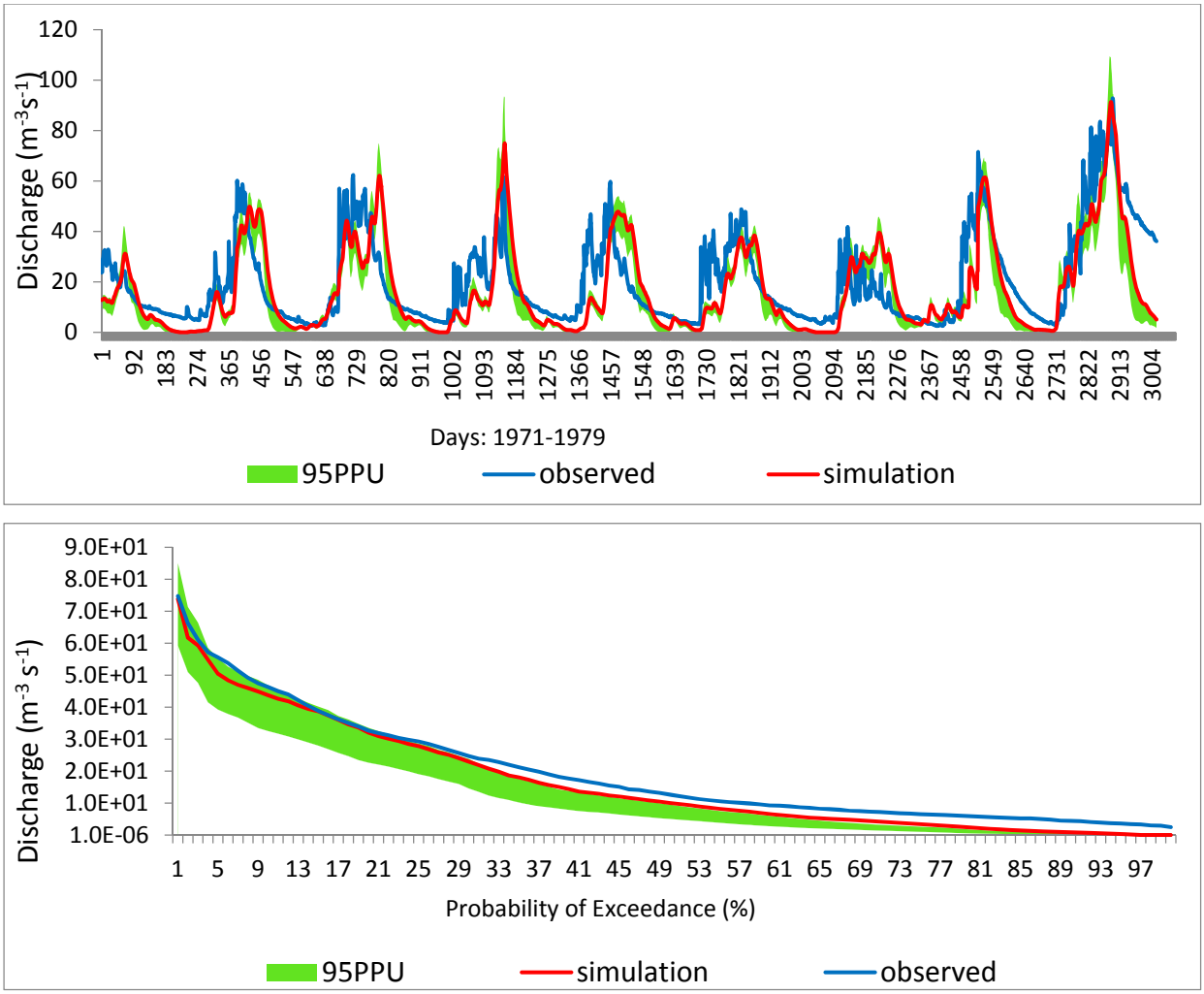


Figure 7.5 Calibration at 1ka31- Mawande (95PPU for behavioural simulations)

7.8 Discussions and conclusion

The SWAT2009 was applied to the Little Ruaha sub-basin. The model was set up using a coarse spatial dataset, interpolated rainfall data and a single dominant HRU. Sensitivity analysis results showed that ALPHA_BF, CN2, SURLAG, REVAPMN, CH_K2, GWQMN, SLSUBBSN, BLAI, and CANMX are the most sensitive parameters in the basin. The Little Ruaha drainage system falls within the African land surface where the infiltration of the top soil is good and interflow is an important part of the total river discharge. The soils in the upper part are deeply weathered and have a good soil structure. This explains the sensitivity of the surface and subsurface parameters. The drainage is dominated by steep topography and this explains the sensitivity of the mean slope length of the basin. Sensitivity analyses enabled the most sensitive model parameters to be identified for further calibration, but this does not mean that sensitive

parameters will also be identifiable. Out of the 27 parameters, 20 were identified as sensitive, but the interactions between these parameters were not considered during the sensitivity analysis.

Final calibration parameters for the Little Ruaha Drainage System are presented in Table 7.4, with a CE of 0.54 and R^2 of 0.62 for the best simulation regardless of the parameter set. This is due to the fact that the behavioural parameter sets are within the non-behavioural parameter sets. The results show reasonable performance in the hydrologic simulations but with large uncertainties (Figure 7.4). The model performance statistics achieved in this study are not different from the ones achieved in other studies in Tanzania (Ndomba et al., 2008), but one point that should be noted is that, after calibration, parameters should have physical meaning. With the large equifinality in the parameter sets (Figure 7.2), it was not possible to get identifiable parameter sets and it is hard to say that behavioural parameters sets are representatives of the basin's behaviour. van Griensven et al. (2012) reviewed the use of the SWAT model in the Nile Basin countries, including Tanzania, and found that the model produced satisfactory or good results, but almost all the case studies reviewed gave results based on the wrong process representation. These results were problematic because when different studies in the same or similar sub-basins are compared, they give different results. In peer reviewed papers (Mulungu and Munishi, 2007; Ndomba et al., 2008) some documented parameter values were not realistic but this information was not reported in those papers (van Griensven et al., 2012). This observation highlights the challenges associated with implementing SWAT for water resources use in Tanzania and other developing countries.

Despite the fact that the model gave satisfactory results based on the performance measures, a critical analysis of Figures 7.4 and 7.5 suggests a different picture. Figure 7.4 shows that there is good agreement between observed and simulated flow, but associated with very large uncertainty in high to moderate flows and the uncertainty band does not bracket the low flows. Running the model with the behavioural parameter sets shows a reduction in P-factor and R-factor values (Table 7.4). Figure 7.5 shows that while the uncertainty band has been reduced, the model is under-simulating both high and low flows, and does not bracket the moderate to low flows. This could be associated with input data uncertainties or some processes are not

well represented in the model. ALPHA_BF was the most sensitive parameter identified through the sensitivity analysis, and apart from a lack of observed ground water information, difficulties of SWAT in simulating ground water flow (Sellami et al., 2013) might have contributed to the negative aspects of these results.

This study assessed model uncertainty using a combined uncertainty approach that assumes all sources of uncertainty have been considered within the model. In such an approach it is hard to separate the sources of uncertainty and therefore a follow-up analysis of uncertainty should be undertaken by determining how erroneous input data influence model results. Although not assessed within the research questions of this study, the results highlight potential uncertainties in the input rainfall and evaporation data. The use of these data were justified and used in the simulations but could potentially have influenced the overall model performance and uncertainties that cannot be explained.

The uncertainty analysis was carried out using 20 sensitive parameters, which is a large number considering the interactions between them. Therefore, some less sensitive parameters should be fixed and allow only the most sensitive parameters to vary. This will reduce the effect of parameter interactions and hence the none-uniqueness problem. Although this model has been shown to generate reasonable results it is worthwhile to consider the challenges associated with setting up a distributed model. In this research, large scale spatial datasets have been used and a homogenous model was assumed because the spatial data resolution was insufficient to represent large numbers of hydrological response units. However, even when the resolution was sufficient, attribute values for most of the parameters are lacking. Because of difficulties associated with parameter representation across spatial scales, it is better to use a homogenous set up because biases and uncertainty can be added by the modeller when trying to parameterise values within the hydrological response unit at a size larger than its coverage. The overall conclusions from this assessment include;

- The SUFI-2 approach has capabilities of identifying behavioural parameter, however the results are influenced by large equifinality.

- The scatter plots of the parameter values against objective functions obtained after simulation provided an initial qualitative overview of the uncertainties involved in the representation of basin's behaviour.
- The 95% of the predictive uncertainty (95 PPU) for stream flow computed using SUFI-2 using the Latin Hypercube sampling with 2 000 runs, did not bracket all simulations, indicating that some processes are not represented in the model, hence additional information is needed to improve the results.
- It is also important to emphasize that the prediction limits obtained with SUFI-2 are highly dependent on the threshold selected to separate behavioural from non-behavioural parameter sets, and that the subjective choice of threshold value and objective function can lead to additional uncertainty in the simulation results.

8 CONCLUSIONS AND RECOMMENDATIONS

8.1 Summary of the research

Developing an understanding of the hydrological processes that occur in a system is critical for the effective assessment of water resources. However, the lack of observational data represents a serious challenge to understanding that is difficult to resolve, especially when there are so many factors that contribute to hydrological variation and change. Scientists and practitioners within the southern African region are attempting to develop the most effective methods for water resources assessment that will contribute to effective water resources management. The transfer of parameters from donor catchments (gauged) to other areas (ungauged) has been one of the long term challenges in hydrological modelling. The use of uncertainty and parameter value probability distributions are not new to hydrological science, it has taken a long time for these approaches to be used in hydrological analyses and in routine practice. For the Southern African region common uncertainty approach to all simulations has been developed. However, the remaining challenge is the realistic approach for constraining the ensembles based on information is available. For gauged catchment's, this would lead to the final ensemble set with a very narrow band of uncertainty, while in other areas the uncertainty band would be much greater depending of type and quality of information used as constraints.

Much of the discussion in this thesis focused on uncertainty (input data, parameter uncertainty, etc.) and the use of local and regional constraints for identifying behavioural (i.e. realistic) model simulations. Beven (2002) argues that a realistic level of uncertainty in model predictions has not been widely achieved, and hydrological models are expected to produce predictions for water resource management based on insufficient knowledge of the basin behaviour and poor data. Uncertainties in hydrological modelling limit both the skill of predictive models and the application of the models for water resources assessment and management. The International Association of Hydrological Science's (IAHS) Prediction in Ungauged Basins (PUB) has highlighted the unreliability of traditional approaches in hydrological modelling studies that are still practiced to date. While the traditional methods

have been useful, a number of weaknesses have been identified which are difficult to resolve. Given the importance of this problem, various approaches have been developed. These approaches range from establishing behavioural parameter sets to uncertainty analysis techniques. Resolving equifinality has been one of the most important issues in hydrological modelling and using multiple performance measures has proved to be a useful approach. In ungauged basins, model performance measures have to be based on regional information that represent our best knowledge of the likely response characteristics. This knowledge will always be uncertain and, therefore, the performance measures (or constraints on the hydrological behaviour) should also be uncertain and can be represented as bounds (maximum and minimum) or probability distributions. These then represent the limits of acceptability (Beven, 2012) and can be used to differentiate between behavioural and non-behavioural outputs when using uncertainty ensemble hydrological modelling methods. Using several constraints can make the problem more complex (potential for inconsistencies between different constraints) but has the advantage that different constraints can address different components of the model (surface runoff, groundwater recharge, low flows, etc.). Establishing appropriate constraint values will always, however, be subject to the availability and accuracy of some representative observed stream flow data within the region, as well as the accuracy of the available climate driving data. The extent to which the observed stream flow data can be considered to represent natural conditions is also of great importance.

This study has employed the concepts of the limits of acceptability approach for setting up the GW_PITMAN rainfall-runoff model (Hughes, 2013) for the Great Ruaha River basin and the assessment of uncertainties associated with simulations of naturally hydrological responses (including the effects of a large wetland) and the impacts of existing water abstractions. The aim was to explore uncertainties in modelling hydrological responses in the Great Ruaha River basin and to establish a behavioural model that can be used for water resources management and future decision making. This approach has addressed a range of key issues in hydrological modelling; these include the use of constraints based on natural hydrology, uncertainties associated with input data, parameter equifinality and the importance of realistic uncertainty representation through the use of constraints.

The SWAT model has been used in hydrological studies in Tanzania, but the uncertainty version of this model has not been fully explored. Previous studies (Schuol and Abbaspour, 2006) have indicated that the Sequential Uncertainty Fitting version-2 (SUFI-2) is an efficient uncertainty analysis procedure with model runs generating between 500 and 2 000 ensembles. However, the success and efficiency is very dependent on the type of computer used and the reliability of the electricity power supply as it can take 2 to 3 days to complete the 2 000 model runs. This is not a trivial issue in some countries of Africa, including Tanzania. The SWAT model has been used in this study as a comparison with the GW_PITMAN model and as a comparison between uncertainty approaches. In the SUFI-2 approach, the 95% model uncertainty bands (95PPU) are produced from the behavioural simulations together with the p-factor which expresses the percentage of observed data bracketed in the 95PPU as well as the R-factor, which is the average width of the 95PPU. The approach is useful in that it accounts for all sources of uncertainties.

To summarise, this study can be divided into four main parts; chapters 1 to 3 that provide the background to the research, chapters 4 to 5 the hydrological modelling methods and the development of constraints, chapter 6 presents the hydrological modelling and uncertainty analyses with GW_PITMAN, while chapter 7 provides the calibration and uncertainty analyses results for the SWAT model.

8.2 Assessment of research aims and objectives

The performance of the different uncertainty approaches indicated that it was possible to establish a behavioural GW_PITMAN model for the GRR basin with realistic uncertainty bounds that represent the full range of uncertainty sources. These include the input rainfall and evaporation demand data, the observed stream flow data, information about existing water uses, the information about physical basin properties and its interpretation, as well as the structure (including the spatial and temporal scales used) of the model itself. The main justification for this conclusion is based on the results obtained at the downstream gauges 1ka59 and 1ka5a. The extent to which the original research questions have been addressed is discussed in the following sections.

8.2.1 Catchment characterization and similarity analysis

A behavioural model should be based upon the understanding of the real catchment processes, such as flow routes and runoff generation mechanisms. Without this understanding, the behavioural model and subsequent quantitative assessments of the system will be highly uncertain. Moreover, without careful examination of the available information, model simulation results are subject to a high risk of misinterpretation. In gauged areas, this might result in getting the right answers for the wrong reasons (Krichner, 2006), which could lead to the wrong answers in ungauged areas where the model cannot be fully validated. In an attempt to improve the understanding of the dominant processes in the GRR basin, existing data and information were examined, and processes thought to be important for hydrological modelling were identified (Chapter 3 and Chapter 5). This understanding led to the development of six regional groups (Chapter 5) differentiated by their geomorphological and geological characteristics and which guided the whole hydrological modelling process including establishing the model output constraints, which was a critical step in the approach used. The level of understanding achieved resulted from an iterative process involving a number of hypotheses that were developed from the available information and tested using the model (Beven, 2012). The main conclusions include:

- The information derived from spatial datasets such as soil characteristics; land cover and slopes can be used for hydrologic similarity analysis because it reflects the climatic controls on the system behaviour. However, the spatial distribution of geological and geomorphological characteristics is large such that it is not possible to use a single approach to regionalisation. Hence, a combined approach should be used in regionalisation as this will help in understanding and estimating runoff generation in areas where data are limited.
- The six regional groups developed in this study were influenced by climatic gradients that had a strong influence on the formation of the groups. The identified regional groups indicate the main regions where there are similar patterns such as rainfall, temperature, evapotranspiration, topography, vegetation and soil types. Similarity percentage analysis was carried out, in order to identify the contribution of the each

variable to the average similarity within the regions of homogenous physical basin characteristics.

The complexity of sub-surface flow would mean that any hydrological model should involve simplifications and generalisations. The challenge that remain to hydrologists is that which simplifications and generalisations are the right ones (Kirchner, 2009). While our understanding of the GRR basin processes is still uncertain, the physical basin information has been valuable in providing basic insights identifying important basin characteristics. Many of the knowledge gaps in understanding are due to the difficulty in explaining the complex interactions between many interacting factors. For example, processes such as groundwater recharge are almost impossible to identify and measure, and hence, subject to very large uncertainty in hydrological modelling.

8.2.2 The use of constraints in hydrological modelling

The part of the thesis explored the use of constraints as proxies for observed historical data which were used to identify behavioural parameter sets and model results within an uncertain modelling approach in both gauged and ungauged basins. The initial constraint bounds were based on the available observed data, an analysis of their accuracy and how well they represent natural flow characteristics. These were developed for each observed sub-basin and extrapolated to the ungauged sub-basins based on the sub-basin groupings and the assumptions about their hydrological response characteristics derived from the catchment similarity analysis (section 8.2.1 above). The constraints were used to reduce the initial uncertain parameter space which was the important aspect of this study. The use of constraints has been useful in quantifying realistic parameter uncertainty across all sub-basins. Inevitably, the extent to which the parameter sets and model outputs can be considered behavioural depends upon the success of the process used to quantify the constraint bounds. Some of the constraints are affected by poorly observed data. For example, there are cases where very high peak flows were beyond the range of the rating curve while the accuracy of the low flow constraints is likely to be affected by assumptions about the volumes of abstraction. The accuracy of the groundwater recharge constraint values is highly uncertain due to the lack of observed groundwater data.

The success of the constraint-based uncertainty reduction process is also dependent on the input parameter uncertainty ranges. These were partly guided by some initial manual calibration model runs and partly by evaluations of the frequency distributions of the constraint values in the output ensembles of the uncertainty version of the model. Thus, it is possible to 'calibrate' the input parameter uncertainty ranges to ensure that the final behavioural output ensembles contain results that cover the full range of input constraints. However, achieving this is also dependent upon the six different constraints being fully compatible with each other. If they are not then, it is unlikely that the full range of constraint values will be represented in the final behavioural ensemble set. In most GRR sub-basins, the results were favourable, and the conclusion is that the constraint bounds used were compatible with each other and with the input parameter uncertainty bounds. There were, however, a few sub-basins where the results suggest a bias towards one or more of the upper or lower constraint bounds. This study, is a first attempt to test the applicability of constraints within the uncertainty framework. The failure of the simulation uncertainty bounds to occupy the full range of the uncertainty bound does not necessarily lead to outright rejection of the approach, but draws attention to its limitations, and where more adjustments should be done for future analyses. Most of the implications of this study will be applicable to distributed models, regardless of whether they are semi-distributed or fully distributed. However, even a fully distributed model cannot be applied using local parameters effectively, because techniques for estimating effective parameter values at grid scales does not exist and will certainly lead to errors in local predictions. One of the recommendations is that some further improvement of either the constraint bounds or the input parameter bounds would be beneficial. The overall conclusion is that, the approach has worked very well and has proved to be very useful in organising a substantial amount of uncertain information into a coherent model of the whole basin.

8.2.3 Realistic uncertainty to inform the model calibration process

All models are imperfect representations of reality, the knowledge of the imperfections is one of the basic aspects to be considered for practical application of hydrological models. Hydrological models can contribute to the understanding of hydrological processes within the

system; however, uncertainties related to the model input information should be realistically presented. The uncertainty version of the GW_PITMAN model has been being used to explore different parameter sets and their effects on a range of outputs such as surface runoff, storage, ground water recharge, etc. The inclusion of constraints has allowed for the parameter space to be explored in a more efficient way than earlier uncertainty versions and the single run version of the model.

Using the GW_PITMAN model, a two-step sampling approach has been used to establish uncertain but behavioural parameter sets for both natural and impacted conditions. The majority of the parameters have been quantified within relatively narrow bands of uncertainty, and it can be concluded that the approach has been appropriate in terms of the type of data that are available. Although a great deal of uncertainty has been accounted for, there are still some parts of the model that will remain highly uncertain in the absence of additional observed climate, stream flow and water use data. One of these is the transition of hydrological response characteristics from the relatively well gauged headwater sub-basins in the extreme southwest of the basin to the lower elevation foothill sub-basins (ungauged) and into the Usangu Plains. Another is the transition from the same headwater areas northwards through the ungauged basins forming the western boundary of the basin to the semi-arid gauged sub-basin 1ka41. These are not only ungauged with respect to stream flows, but they also contain no rainfall stations. Amongst the nine gauged sub-basins in the south-western headwaters there are substantial differences in the observed response characteristics. Some of these can be attributed to known geological and topographical differences, but by no means all of them. In addition to this, many parts of the GRR have large populations of rural communities, and their impacts on the hydrological regimes remain un-investigated and largely unknown. The possible impacts could include direct abstractions for domestic and garden irrigation purposes (already included in the model), as well as changes to vegetation cover and infiltration characteristics. Including land use and water abstractions, impacts in the model is likely to remain extremely uncertain.

The SWAT model and associated uncertainty analysis was applied only to the Little Ruaha sub-basin. In attempting to determine the most behavioural model, the uncertainty framework

together with sensitivity analysis provided guidance in conceptualising the model and understanding the processes. The SUFI-2 approach has managed to generate behavioural simulations, however, the parameters used in generating such simulations are influenced by large equifinality such that there are several combinations that may result in the same performance. The results show large model prediction uncertainty that could be partly due to the quality of the observed data use, but also the large equifinality in the model parameters. The P-factor and R factor are the only two statistical criteria used to assess the final acceptable model prediction uncertainty within SUFI-2, and there is a danger that the two measures can compensate each other such that a higher p-factor can be achieved at the expense of the R-factor. Thus, additional iterations may lead to a narrow uncertainty band, and this makes SUFI-2 lose its objectivity in deriving realistic uncertainty bounds.

The results from the two models have highlighted the existence of uncertainty in our available knowledge of the water resources availability in the GRR basin and this uncertainty should be communicated to water managers. In Tanzania, as in many other parts of the world, water resource decisions are still made without considering the uncertainties inherent in quantifying the system behaviour. Although some of the analyses show quite large uncertainties that may be difficult to use in decision making environment, it is still advantageous to communicate this information to water managers and policy makers. The uncertainties may encourage investment in further data collection and analysis to assist in reducing the uncertainties in the future.

8.2.4 Model calibrations and parameter sampling approaches in data scarce regions

This study represents a contribution towards the quantification of uncertainty in hydrological predictions in data scarce regions. Both models applied in this study provided reasonably acceptable simulation results despite the differences in spatial scale and time steps. Trade-offs exist between models in terms of data availability and realistic representation of parameter values across spatial scales. The ability of the uncertainty framework to incorporate constraints in guiding the parameter sampling and the recognition that observed data include uncertainties have been positive steps in addressing some of the uncertainty challenges in data scarce regions. The components of the GW_PITMAN uncertainty framework are considered

robust enough to enable its use with any model structure and capable of being used in both gauged and ungauged basins. While it would be imperative to be able to conclude this thesis by making some more concrete recommendations on the best calibration and or sampling approach, this was not the major aim of this thesis. The results have been quantified using uncertain data, and this means that some information has been quantified with unknown uncertainties, and it was not possible to disaggregate all the different sources of uncertainty that contribute to the unresolved uncertainty. Thus, the most important thing to conclude is that given the limited knowledge we have about all sources of uncertainty, the models are capable of generating estimates that reflect our understanding of the catchment processes and can be used for water resources management and climate change impacts assessments.

8.3 Application of the model for establishing hydrological impacts of climate change

The primary objective of this thesis was to establish a hydrological model for future predictions of hydrological responses to development and climate change impacts and assessments of future water availability. During the course of this study, it became quite clear that to include detailed climate change impact assessments would not be realistically possible, largely due to the amount of time required fully to understand the uncertainties in different future climate predictions. However, this section has been added to the thesis to illustrate possible approaches to future climate change studies using the established model. The success of climate impact assessment studies is partly dependent upon the uncertainties in the reliability of the future climate estimates derived from one or more GCMs or RCMs but also depend upon the ability of the hydrological model to represent the catchment responses to climate variations and change. Thus, both hydrological and climate model uncertainties propagate through the entire assessment process and are not always easy to isolate. No critical assessment of these preliminary results has been included as part of the study. Hence, further research will therefore be carried out as a follow-up study for future water resources assessment within the GRR basin.

The established hydrological model (GW_PITMAN) with behavioural parameter sets has been utilized to simulate the effects of climate change projections on stream flow as given by three (3) GCM/RCM combinations using the RCP 4.5 emission scenario for the near future period

2031-2060. The RCM projections of change in monthly precipitation and potential evapotranspiration were estimated using the delta change approach where future climate conditions are compared to a reference period 1976-2005. Preliminary results based on three GCM_RCM combinations (NCC_RCA4, MPI-ESM_CRCM5 and MIROC5_RCA4) clearly indicate a shift in mean discharge (Figure 8.1). There are notable changes in both high and low flows. In all three GCM-RCM combinations, low flows are lower than the historical flow conditions, while high flows are higher than historical flow conditions. The high flows correspond to the projected increase in precipitation where initial results indicate a clear tendency across the RCM to increased precipitation in the wet season and decreased precipitation in the dry season with an overall annual increase in precipitation. Potential evapotranspiration tends to increase due to a consistent 2-3 degree temperature increase across all RCM's. The decrease in low flows is largely a result of the increased evapotranspiration and, therefore, decreased soil moisture storage. While this effect could lead to increased irrigation demands, these were not included as part of the initial climate change assessment.

A comparison of the uncertainty arising from the hydrological model and GCM-RCMs combination is illustrated in Figure 8.2. The comparison is based on the upper and lower bounds of the simulated historical and future stream flows. Figure 8.2 illustrates the upper and lower bounds for the three GCM-RCM combinations where there is overlap between the historical and future predicted uncertainty.

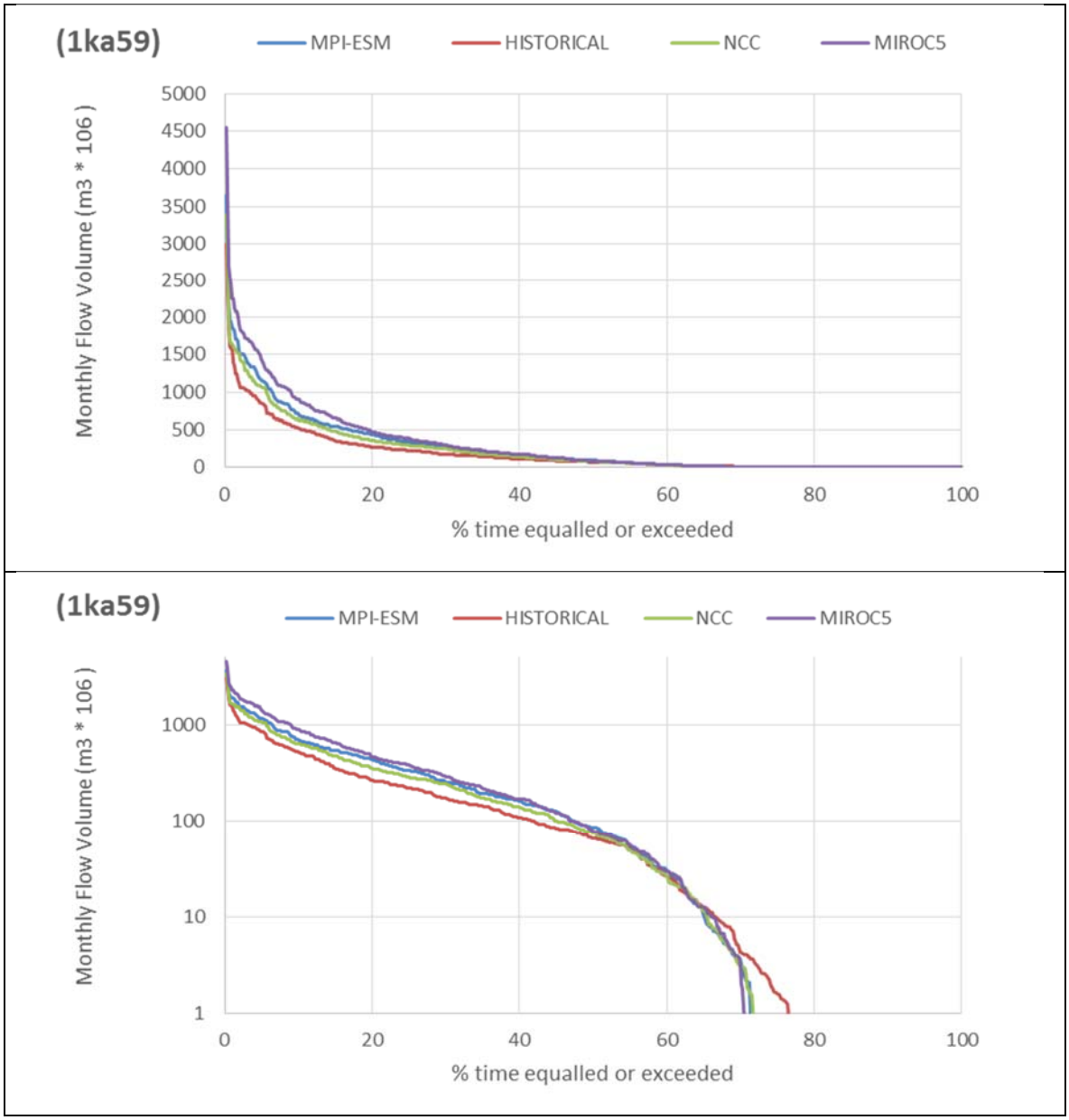


Figure 8.1 Flow duration curve plotted with normal scale (top) and logarithmic scale (below) indicating the historical and near future stream flows for sub-basin 1ka59. The red continuous line is the historical stream flow for the period (1976-2005)

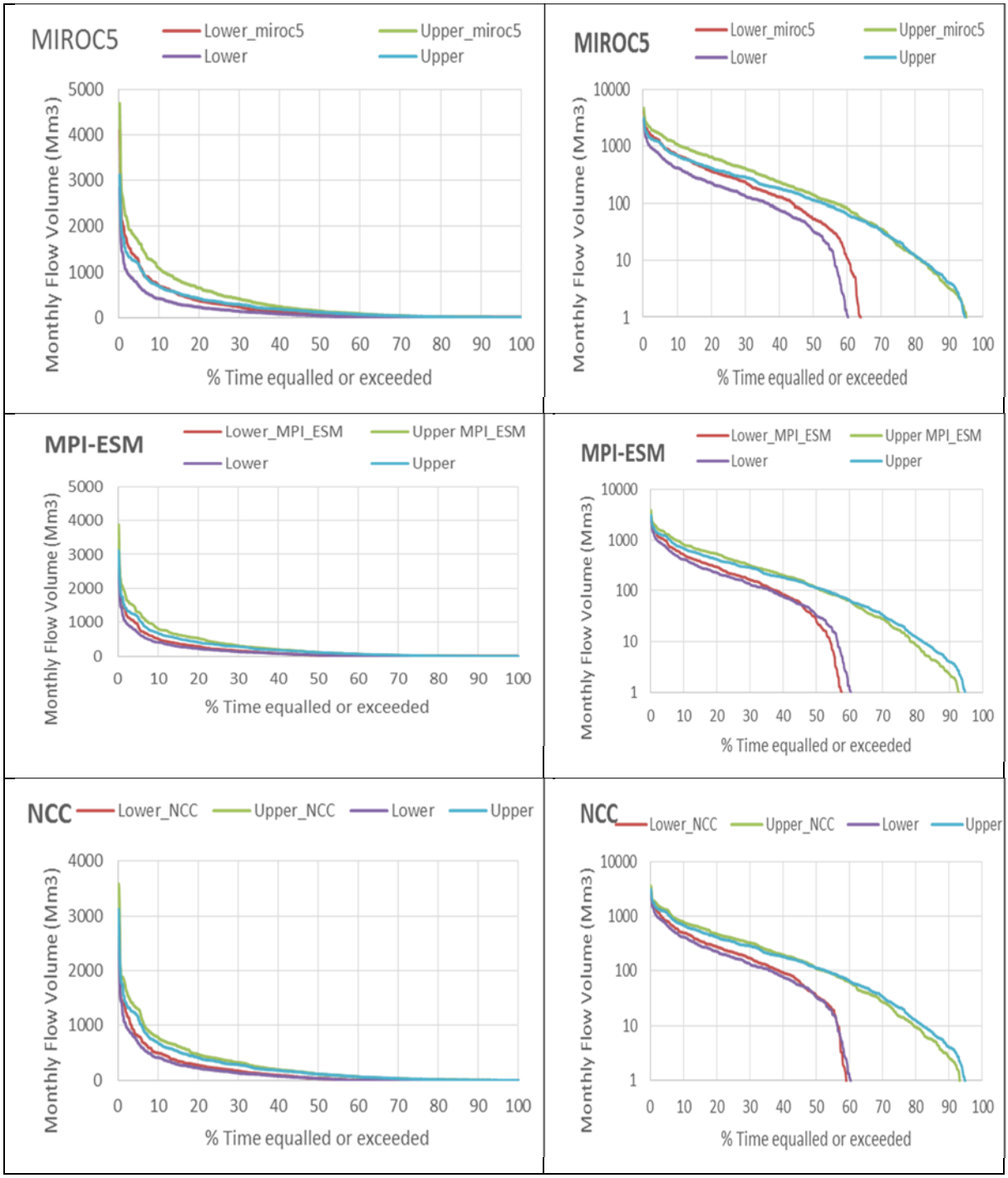


Figure 8.2 Comparison of the hydrological model uncertainty and GCM uncertainty

8.4 Recommendations for further research

8.4.1 Recommendations for the GW_PITMAN model

That, there are some parts of the basin where further improvements can be made is undeniable. However, the results for many parts of the basin, and particularly the downstream sites (1ka59 and 1ka5a) that are critical for water resources assessments and decision making, are very encouraging while some parts of the basin still need further improvements and these include:

- A re-evaluation of the observed flow data to identify the more reliable records (or parts of some records) so that the constraint ranges for these sub-areas can be reduced.
- An evaluation of the high flow simulations in some parts of the basin (notably the arid northwest sub-basins) to identify if these are related to excessively high rainfall inputs or some deficiencies in the existing parameter sets (e.g. too low maximum storage parameter values).
- An assessment of the sensitivity of the model results to changes in the water use parameters and whether or not any information (SMUWC, 2001; 2009) about the historical patterns of irrigation management practices are reflected in changes in the observed low flow regimes. Exploring licensed abstraction data within the basin that exist, but were not available for this study may also improve the water use abstraction calculations, and the associated model parameters.
- An overall re-evaluation of the spatial variations in low flows (linked to the previous point) that includes assumptions about surface – groundwater interactions.
- The inclusion of simulations of Mtera Dam and the Kidatu hydropower scheme to assist in identifying the key components of the upstream flow regimes that are likely to affect power production. This information would be valuable to provide focus for further model improvements and uncertainty reduction.
- Further exploration of the evapotranspiration representation in the model. It would be interesting to drive the model with alternative input evapotranspiration data (e.g., MODIS) and to assess its influence on the sensitivity of the calibrated models.

8.4.2 Recommendations for the SWAT model

Data availability is a general problem and the lack of data has been mentioned by many authors and in this thesis. The problem becomes more complicated when setting-up the distributed model with large scale datasets (e.g., soil and land use) but also limited climate input data (at the daily time scale required by SWAT). This thesis has presented an initial calibration and uncertainty analysis of the SWAT model, based on the results the following are the recommendations for future research:

- Evaluate the representativeness of the distributed processes and parameters especially when the model is set up for impact studies.
- Refining the model with aquifer boundary information in the saturated zone could also be a further area of research.
- Improving the spatial variability computations by performing the calibration at internal points (nodes).
- The crop and vegetation database currently used in SWAT has been developed for USA conditions and is being tailored for African conditions. These forms part one of many uncertainties of using SWAT in African conditions especially when the user is trying to localise the database. Therefore, the crop and vegetation database for Africa should be developed and shared.
- Due to the large equifinality in SWAT model parameters, the calibration process should concentrate on the most sensitive parameters dominating the hydrological processes of the basin, and try to resolve the uncertainties associated with the interaction between parameters e.g sub-surface storage and groundwater storage
- Explore other uncertainty frameworks tailored for SWAT and compare the results with SUFI-2.

REFERENCES

Abbaspour, K.C. (2005). Calibration of hydrologic models: when is a model calibrated? In MODSIM 2005 International Congress on Modelling and Simulation, Zerger A, Argent RM (eds). Modelling and Simulation Society of Australia and New Zealand, 2449–12455.

Abbaspour, K.C. (2007). SWAT-CUP, SWAT Calibration and Uncertainty Programs. A user manual, Swiss Federal Institute for Aquatic Science and Technology, Zurich, Switzerland, 84p.

Abbaspour, K.C. (2008). SWAT-CUP2: SWAT Calibration and Uncertainty Programs - A User Manual. Department of Systems Analysis, Integrated Assessment and Modelling (SIAM), Eawag, Swiss FederalInstitute of Aquatic Science and Technology, Duebendorf, Switzerland.

Abbaspour, K.C., Johnson, C.A., and van Genuchten, M.T. (2004). Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose Zone Journal*, **3** (4), 1340–1352.

Abbaspour, K.C., Yang J., Maximov I., and Siber R. (2007). Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of Hydrology*. **333** (2–4), 413–430

Ajami N.K., Duan Q. and Sorooshian S., (2007). An integrated hydrologic Bayessian multimodel combination framework. Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resources Research*, **43**(W01403).

Allen, R.G., Pereira, L.S., Raes, D., and Smith, M.(1998). Crop evapotranspiration – Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper, No. 56, FAO, Rome

Andersson, J.C.M., et al., (2009). Water availability, demand and reliability of in situ water harvesting in smallholder rain-fed agriculture in the Thukela River Basin, South Africa. *Hydrology and Earth System Sciences*, **13** (12), 2329–2347.

Arnold, J. G., Srinisvan, R., Muttiah, R. S. and Williams, J. R. (1998). Large area hydrologic modeling and assessment. Part I: Model development. *Journal of American Water Resources Association*. **34** (1), 73–89.

Beven, K.J. (1989). Changing ideas in hydrology – the case of physically-based models. *Journal of Hydrology*. **105**, 157-172.

Beven, K.J. (2000). On the future of distributed modelling in hydrology. *Hydrological Processes*, **14**, 16-17.

Beven, K.J. (2002). Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrological Processes*. **16** (2), 189-206.

Beven, K.J. (2005). On the concept of model structural error. *Water Science and Technology*, **52**(6), 167-175.

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

Beven, K.J. (2009). Environmental Modelling: An Uncertain Future? Routledge, Abingdon, UK.

Beven, K.J. (2010). Preferential flows and travel time distributions: defining adequate hypothesis tests for hydrological process models. *Hydrological Processes*.**24**, 1537–1547.

Beven, K.J. (2011). I believe in climate change but how precautionary do we need to be in planning for the future? *Hydrological Processes*. 25: 1517–1520.

Beven, K.J. (2012) Causal models as multiple working hypotheses about environmental processes. *Comptes Rendus Geoscience*, **344**, 77-88.

Beven, K. J. and Binley, A. M. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, **6**, 279-298

Beven, K.J., and Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems, *Journal of Hydrology*, **249,** 11-29.

Beven, K.J., and Westerberg, I. (2011). On red herrings and real herrings: disinformation and information in hydrological inference. *Hydrological Processes*. **25**: 1676–1680.

Blöschl, G., and Montanari, A. (2010): Climate change impacts – throwing ¨the dice? *Hydrological Process*. **24**, 374–381.

Blöschl, G., and Sivapalan., M. (1995). Scale issues in hydrology: a review. *Hydrological Processes*. **9**: 251–290.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A. and Savenije, H. (Eds.), (2013). Runoff Prediction in Ungauged Basins. Synthesis across processes, places and scales. *Cambridge University Press*, ISBN 978-1-107-02818-0

Bsalone, R.S. (2007). Parameter Estimation and Uncertainty Assessment in Hydrological Modelling. PhD Thesis, Institute of Environment and Resources, Technical University of Denmark.

Bsalone,R.S., Madsen,H. and Rosbjerg,D. (2008). Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. *Journal of Hydrology*, **353** (1–2), 18–32.

Budyko, M.I., (1974). Climate and life. New York: Academic Press.

Buttle J., (2006). Mapping first-order controls on stream flow from drainage basins: the T3 template. *Hydrological Processes*, **20**, 3415– 3422.

Butts, M.B., et al. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for stream flow simulation. *Journal of Hydrology*, **298** (1–4), 242–266.

Buytaert, W., and Beven K.J (2009). Regionalization as a learning process, *Water Resources Research*. **45**, (W11419)

Buytaert, W., and Beven, K.J (2011). Models as multiple working hypotheses: Hydrological simulation of tropical alpine wetlands. *Hydrological Processes*, **25**, 1784 - 1799

Carrillo G., Troch P.A., Sivapalan M., Wagener T., Harman C, and Sawicz K. (2011). Catchment classification: hydrological analysis of catchment behavior through process-based modeling along a climate gradient. *Hydrology and Earth System Science*. **15**, 3411–3430.

Casper, M.C., Grigoryan,G., Gronz, O., Gutjahr, O., Heinemann, G., Ley, R., and Rock, A. (2012). Analysis of projected hydrological behavior of catchments based on signature indices. *Hydrology and Earth System Science*, **16**, 409-421

Castellarin. A., Galeati.G., Brandimarte,L., Montanari.A., Brath, B. (2004). Regional flow-duration curves: reliability for ungauged basins. *Advances in Water Resources*. **27**, 953–965

CCKK (1982). Water master plans for Iringa, Ruvuma and Mbeya Regions, Volume 8, Hydrogeology. Carl Bro, Cowiconsult, Kampsax, Krüger, 1982.

Clarke, K.R., Warwick, R.M. (1994). Change in marine communities: an approach to statistical analysis and interpretation. Plymouth marine laboratory, UK.

Cohn,T., Kiang,J., and Mason,R.(2013). Estimating discharge measurement uncertainty using the interpolated variance estimator. *Journal of Hydraulic Engineering*, **139** (5), 502-510.

Danida/World Bank. (1995). Water resources management in the Great Ruaha basin: A study of demand driven management of land and water resources with local level participation. Rufiji Basin Water Office, Iringa, Tanzania.

Davie,T. (2008).Fundamentals of Hydrology. Taylor and Francis e-Library. ISBN10: 0–203–93366–4 (ebk)

DeFries, R. and Eshelman, K.N (2004). Land-use change and hydrologic processes: a major focus for the future. *Hydrological Process*, **18**, 2183–2186.

Di Baldassarre, G. and Claps, P. (2011). A hydraulic study on the applicability of flood rating curves, *Hydrology Research*, **42**, 10–19.

Di Baldassarre, G and Montanari, A. (2009). Uncertainty in river discharge observations: a quantitative analysis. *Hydrology and Earth System Sciences*. **13**, 913–921.

Domeneghetti, A., Castellarin, A., and Brath, A. (2012). Assessing rating-curve uncertainty and its effects on hydraulic model calibration. *Hydrology and Earth System Science*, **16**, 1191-1202.

Efstratiadis, A. and Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrological Sciences Journal*. **55**(1), 58–78.

Euser, T., Winsemius, H.C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H.H.G (2013). A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Science*. **17**, 1893–1912.

FAO, (2007). World Reference Base for Soil Resources 2006. A framework for international classification, correlation and communication. IUSS/ISRIC/FAO, Wageningen/Rome. Digital version

Ferguson, I.M., and Maxwell, R.M. (2012). Human impacts on terrestrial hydrology: climate change versus pumping and irrigation. *Environmental Resources Letters*. doi:10.1088/1748-9326/7/4/044022

Forestry and Beekeeping Division (2005). Hydrological analysis of the Eastern Arc Mountain forests. Compiled by Mtalo, F., D. Mulungu, F. Mwanuzi, S. Mkhandi, T. Kimaro and

Valimba P. for Conservation and Management of the Eastern Arc Mountain Forests, Forestry and Beekeeping Division, Dar es Salaam, Tanzania.

Freer, J.E., Beven, K.J. and Ambroise, B. (1996). Bayesian estimation on uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resources Research*, 32, 2161-2173.

Gassman, P.W., M.R. Reyes, C.H. Green and J.G. Arnold., (2007). The Soil and Water Assessment Tool: historical development, applications, and future research directions. *Transactions of the ASABE*. 50 (4), 1211-1250.

Gates, W.L. (1985). The use of general circulation models in the analysis of the ecosystem impacts of climatic change. *Climate Change*. 7, 267-284.

Gentine, P., Troy T., Lintner B.R., Findell K.L. (2011). Scaling in Surface Hydrology: Progress and Challenges. *Journal of Contemporary Water Research and Education*, 147(1), 28-40

Gosling, S. N., Taylor, R. G., Arnell, N. W., and Todd, M.C. (2011). A comparative analysis of projected impacts of climate change on river runoff from global and catchment-scale hydrological models. *Hydrology and Earth System Science*, *15*, 279–294.

Graham, L. P., Hagemann, S., Jaun, S., and Beniston, M. (2007). On interpreting hydrological change from regional climate models. *Climate Change*, 81, 97–122.

Gupta, H.V., Beven, K.J., Wagener, T. (2005). Model Calibration and Uncertainty Estimation. In Anderson, M.G. (ed.) *Encyclopaedia of Hydrological Sciences*. John Wiley & Sons Ltd.

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M. (2012). Towards a comprehensive assessment of model structural adequacy. *Water Resources Research*. Doi: 10.1029/2011WR011044, 2012.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *Journal of Hydrology*. 377, 80–91.

Gupta, H. V., Sorooshian, S. and Yapo, P. O. (1998). Toward Improved Calibration of Hydrologic Models: Multiple and Non-commensurable Measures of Information. *Water Resources Research*, 34 (4), 751-763.

Gupta, H.V., Wagener, T., Liu, Y. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrological Processes*, 22 (18), 3802-3813.

Hansen, M., DeFries R., Townshend J.R.G., and Sohlberg, R. (2000). Global land cover classification at 1km resolution using a decision tree classifier. *International Journal of Remote Sensing*, **21**, 1331-1365.

Hargreaves, G. and Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied Engineering Agriculture*, **1**, 96–99.

Helton J., and Davis F. (2002). Illustration of sampling-based methods for uncertainty and sensitivity analysis. *Risk Analysis*, **22**, 622-691.

Herschy, W.(2002) The uncertainty in a current meter measurement, Flow Measurement and Instrumentation., **13**, 281–284.

Hornberger, G.M. and Spear, R.C. (1981). An approach to the preliminary-analysis of environmental systems. *Journal of Environmental Management*, **12**(1), 7–18.

Horton, P., Schaefli, B., Mezghani, A., Hingray, B., and Musy, A. (2006). Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty. *Hydrological Processes*, **20**, 2091–2109.

Hrachowitz, M., et al., (2013a). A decade of Predictions in Ungauged Basins (PUB) - a review. *Hydrological Sciences Journal*, doi:10.1080/02626667.2013.803183

Hrachowitz, M., et al., (2013b). What can flux tracking teach us about water age distribution patterns and their temporal dynamics? *Hydrology and Earth System Sciences*, **17**, 533–564.

Hughes, D.A. (1992). A monthly time step, multiple reservoir water balance simulation model, *Water South Africa*, **18**, 279-286.

Hughes, D.A. (2004). Incorporating ground water recharge and discharge functions into an existing monthly rainfall-runoff model. *Hydrological Sciences Journal*, **49**(2), 297-311.

Hughes, D.A. (2006). Comparison of satellite rainfall data with observations from gauging station networks. *Journal of Hydrology*, **327**(3-4), 399-410.

Hughes, D.A. (2010). Hydrological models: mathematics or science? *Hydrological Processes*, **24**, 2199–2201

Hughes, D.A. (2013a). A review of 40 years of hydrological science and practice in southern Africa using the Pitman rainfall-runoff model. *Journal of Hydrology,* **501**, 111-124.

Hughes, D.A. (2013b). PUB in practice at the national scale: the case of South Africa. In: Putting Prediction in Ungauged Basins into Practice. eds. J.W. Pomeroy,P.H. Whitfield, and C. Spence. *Canadian Water Resources Association*, 175-184.

Hughes, D.A., Andersson, L., Wilk, J. and Savenije, H.H.G. (2006). Regional calibration of the Pitman model for the Okavango River. *Journal of Hydrology*, **331**, 30-42.

Hughes, D.A., Forsyth, D.A. (2006). A generic database and spatial interface for the application of hydrological and water resource models. *Comput. Geosci.,* **32**, 1389-1402.

Hughes, D.A., Kapangaziwiri, E., and Sawunyama, T. (2010). Hydrological model uncertainty assessment in southern Africa. *Journal of Hydrology*, **387**, 221–232.

Hughes, D.A., Kapangaziwiri, E., Tanner, J. (2013a). Spatial scale effects on model parameter estimation and predictive uncertainty in ungauged basins. *Hydrology Research*, **44**, 441–453.

Hughes, D.A, Mantel, S and Mohobane T. (2013b). An assessment of the skill of downscaled GCM outputs in simulating historical patterns of rainfall variability. *Hydrology Research*, doi:10 2166/nh 2013 027

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**(1-2), 265-291.

Hughes, D.A., Sawunyama, T. and Kapangaziwiri, E. (2008). Incorporating estimation uncertainty into water resource development planning in ungauged basins in southern Africa. In: Sustainable Hydrology for the 21st Century, Proc. 10th BHS National Hydrology Symposium, Exeter, Sept. 2008.

Hughes, D.A., Spence C., and Woods, R. (2013c). Synthesis of major findings at PUB 2011 and recommendations for future directions In: Putting Prediction in Ungauged Basins into Practice. eds. J.W. Pomeroy, P.H. Whitfield, and C. Spence. *Canadian Water Resources Association*, 305-314.

Hughes, D.A., Tshimanga, R., Tirivarombo, S. and Tanner, J. (2013d). Simulating wetland impacts on stream flow in southern Africa using a monthly hydrological model. *Hydrological Processes,* doi 10.1002/hyp.9725.

IPCC, (1994). Technical Guidelines for Assessing Climate Change Impacts and Adaptations. Prepared by Working Group II [Carter, T.R., M.L. Parry, H. Harasawa, and S. Nishioka]

and WMO/UNEP. CGER-IO15-'94. University College London, UK and Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan, 59 pp.

Jha, M., Pan Z., Takle, E.S., and Gu, R. (2004). Impacts of climate change on stream flow in the Upper Mississippi River basin: A regional climate model perspective. *Journal of Geophysics Research*, DOI: 10.1029/2003JD003686.

Jung, G., Wagner, S., and Kunstmann, H. (2012). Joint climate-hydrology modeling: an impact study forthe data-sparse environment of the Volta Basin in West Africa. *Hydrology Research*, **43** (3), 231-248.

Juston,J., Jansson, P.E and Gustafsson, D. (2014). Rating curve uncertainty and change detection in discharge time series: case study with 44 year historic data from the Nyangores River, Kenya. Hydrological Processes. **28**, 2509-2523.

Kapangaziwiri, E., (2008). Revised parameter estimation methods for the Pitman monthly model, MSc thesis, Rhodes University, Grahamstown, South Africa. http://eprints.ru.ac.za/1310/

Kapangaziwiri, E., (2010). Regional application of the pitman monthly rainfall-runoff model in southern Africa incorporating uncertainty. Unpublished PhD Thesis, Rhodes University, Grahamstown, South Africa (available online at http://eprints.ru.ac.za/1777/).

Kapangaziwiri, E., Hughes, D.A., and Wagener, T. (2012). Constraining uncertainty in hydrological predictions for ungauged basins in southern Africa. *Hydrological Sciences Journal*, **57**(5), 1000-1019.

Kashaigili, J.J. (2008). Impacts of land-use and land-cover changes on flow regimes of the Usangu wetland and the Great Ruaha River, Tanzania. *Physics and Chemistry of the Earth*, **33**(8–13): 640-647.

Kashaigili, J.J., Mbilinyi, B.P., McCartney, M., Mwanuzi, F.L. (2006). Dynamics of Usangu plains wetlands: Use of remote sensing and GIS as management decision tools. *Physics and Chemistry of the Earth*, 31(15–16), 967-975

Kavetski, D., Franks, S.W., and Kuczera, G. (2003). Confronting input uncertainty in environmental modelling. In: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R. (eds.), Calibration of watershed models, *Water Science and Application,* **6**, American Geophysical Union, Washington DC, 48-68

Kavetski D., Kuczera G. and Franks S.W., (2006). Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. *Water Resources Research*, **42**(W03408).

Kay, A.L., Jones D.A., Crooks S. M., Kjeldsen T. R., and Fung C. . (2007). An investigation of site-similarity approaches to generalisation of a rainfall-runoff model, *Hydrology and Earth System Science*, **11**(1), 500–515

Kiang, J.E., Cohn, T.A., and Mason, R.R. (2009). Quantifying Uncertainty in Discharge Measurements: A New Approach, Proceedings of World Environmental and Water Resources Congress 2009, doi: 10.1061/41036(342)599

Kiang, J.E., Cohn T.A., Mason R., Fulford J., and Hamilton. S. (2011). A comparison of methods for estimating uncertainty in discharge measurements. In Proceedings of Canadian Water Resources Association 64th Annual Conference, St. John's, Newfoundland, June 27–30, 2011. St. John's, NFLD: Canadian Water Resources Association.

Kim J, D. E. et al., (2013). Evaluation of the CORDEX-Africa multi-RCM hind-cast: systematic model errors, *Climate Dynamics*. DOI 10.1007/s00382-013-1751-7.

Kirchner, J.W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resources Research*, **42**, (W03S04).

Kirchner, J.W. (2009), Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resources Research*, **45**, (W02429).

Kollat, J.B., Reed P. M., and Wagener T. (2012). When are multi-objective calibration trade-offs in hydrologic models meaningful?, *Water Resources Research*, **48**, (W03520).

Koutsouris, A.J., Jarsjö, J., Destouni, G., & Lyon, S.W., (2010). Hydro-climatic trends and water resource management implications based on multi-scale data in the Lake Victoria region, Kenya. *Environmental Resource Letters,* **5** (034005) (http://iopscience.iop.org/1748-9326/5/3/034005)

Koutsoyiannis, D. (2010). A random walk on water. *Hydrology and Earth System Sciences*,**14**, 585–60

Krueger, T., J. Freer, J. N. Quinton, C. J. A. Macleod, G. S. Bilotta, R. E. Brazier, P. Butler, and P. M. Haygarth. (2010). Ensemble evaluation of hydrological model hypotheses, Water Resour. Res., 46, W07516

Kuczera, G. and Parent, E. (1998). Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm. *Journal of Hydrology*, 211, 69-85.

Lankford, B.A., Tumbo, S. and Rajabu, K. (2009) 'Water competition, variability and river basin governance: A critical analysis of the Great Ruaha River, Tanzania' in Molle, F. and Wester, P. (eds) River Basin Development in Perspective, CABI. pp 171-195.

Liechti, T.C., José Pedro Matos, David Ferràs Segura, Jean-Louis Boillat and Anton J. Schleiss. (2014). Hydrological modelling of the Zambezi River Basin taking into account floodplain behaviour by a modified reservoir approach. *International Journal of River Basin Management*, DOI: 10.1080/15715124.2014.880707

Liu, Y., and Gupta H.V. (2007). Uncertainty in hydrologic modelling: Toward an integrated data assimilation framework, *Water Resources Research*, 43(W07401).

Lubini, A and Adamowski J. (2013). Assessing the Potential Impacts of Four Climate Change Scenarios on the Discharge of the Simiyu River, Tanzania Using the SWAT Model. *International Journal of Water Sciences*, 2, 1:2013

Ludwig, R., et al., (2009). The role of hydrological model complexity and uncertainty in climate change impact assessment, *Advances in Geosciences*, 21, 63-71.

Mazvimavi, D. (2003). Estimation of Flow Characteristics of Ungauged Catchments: Case Study in Zimbabwe. Unpublished PhD thesis, University of Wageningen, Netherlands.

McDonnell J.J, Sivapalan M, Vach´e K, Dunn S, Grant G, Haggerty R, Hinz C, Hooper R, Kirchner J, Roderick ML, Selker J, Weiler M. (2007). Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research* 43 (WR005467).

McDonnell, J.J., and Woods, R. A. (2004). On the need for catchment classification, *Journal of hydrology*, 299, 5 (2–3).

McIntyre, N., Lee H., Wheater H., Young A., and Wagener T. (2005). Ensemble predictions of runoff in ungauged catchments, *Water Resources. Research*, 41, (W12434).

McMillan, H., Gueguen, M., Grimon, E., Woods, R., Clark, M. and Rupp, D.E. (2013). Spatial variability of hydrological processes and model structure diagnostics in a 50 km^2 catchment. *Hydrological Processes*. DOI: 10.1002/hyp.9988

Mearns, L. O., W. Easterling, C. Hays, and D. Marx., (2001). Comparison of agricultural impacts of climate change calculated from high and low resolution climate model scenarios: Part I. The uncertainty due to spatial scale. *Climatic Change*, **51**, 131-172

Merz, R., and Bloschl, G. (2004). Regionalisation of catchment model parameters. *Journal of Hydrology*, **287**, 95-123

Merz, R., Parajka, J. Blöschl, G. (2009) Scale effects in conceptual hydrological modelling. *Water Resources Research*, **45**, 91944-7973

Milzow, C., Krogh P.E., Bauer Gottwein P., (2011). Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment, *Hydrology and Earth System Sciences*, **15**, 17291743.

Mitsch,W.J., and Gosselink J.G. (2007) .Wetlands (4th ed) John Wiley, Inc, New York

Montanari, A., Toth, E. (2007). Calibration of hydrological models in the spectral domain: An opportunity for ungauged basins? *Water Resources Research*, **43**, (W05434)

Montanari, A., Young, G., H. Savenije, D. A. Hughes, T. Wagener, L Ren, D. Koutsoyiannis, C. Cudennec, S. Grimaldi, G. Bloeschl, M. Sivapalan, K. Beven, H. Gupta, B. Arheimer, Y. Huang, A. Schumann, D. Post, V. Srinivasan, E. Boegh, P. Hubert, C. Harman, S. Thompson, M. Rogger, M. Hipsey, E. Toth, A. Viglione, G. Di Baldassarre, B. Schaefli, H. McMillan, S.J. Schymanski, G. Characklis, B. Yu, Z. Pang and V. Belyaev (2013). "Panta Rhei – Everything Flows": Change in hydrology and society – The IAHS Scientific Decade 2013-2022. *Hydrological Sciences Journal*, **58**(6), 1256-1275

Monteith, J.L. (1965). Evaporation and Environment. 19th Symposia of the Society for Experimental Biology, University Press, Cambridge, **19**:205-234.

Moore, R.D., Hamilton A.S, Whitfield P.H, (2012). North American Stream Hydrographers [NASH] Special Issue. *Canadian Water Resources Journal,* **37**, 1, (1-2).

Moore, R.J., (1985). The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal*, **30** (2), 273–297.

Morris, M.D.(1991). "Factorial Sampling Plans for Preliminary Computational Experiments". *Technometrics*, **33**, 161–174.

Mul, M. (2009). Understanding Hydrological Processes in an Ungauged Catchment in sub-Saharan Africa. PhD Thesis. Delft University of Technology

Mulungu, D. M. M., and Munishi, S. E. (2007). Simiyu river catchment parameterization using SWAT model. *Journal of Physics and Chemistry of the Earth*, **32**, 1032–1039.

Mwelwa, E.M. (2004). The application of a monthly time step Pitman rainfall-runoff model to the Kafue river basin of Zambia. MSc thesis, Rhodes University, Grahamstown, South Africa. http://eprints.ru.ac.za/173/.

Nash, J. E., and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. A discussion of principles. *Journal of Hydrology*, **10**(3), 282-290

Ndiritu, J., (2009). A comparison of automatic and manual calibration using the pitman model. *Physics and Chemistry of the Earth*, **34**,729-740

Ndomba, P. M., and Birhanu, B. Z. (2008). Problems and Prospects of SWAT Model Applications in NILOTIC Catchments: A Review, *Nile Basin Water Engr. Sci. Magazine*, **1**, 41–52.

Ndomba, P., Mtalo, F., and Killingtveit, A. (2008). SWAT model application in a data scarce tropical complex catchment in Tanzania, *Physics and Chemistry of the Earth*, **33**, 1–19.

Neitsch, S.L., Arnold, J., Kiniry G., Williams J. R., King, K. W. (2005). Soil and Water Assessment Tools: Theoretical documentation version 2005, Grassland, Soil and Water Reasearch Laboratory, ARS, 808 East Blackland Road- temple, Texas 76502.0092009 Report No: 061.

Nikulin G., Jones C., Giorgi F., Asrar G., Büchner M., Cerezo-Mota R., Bøssing Christensen O., Déqué, J M., Fernandez, A. Hänsler, van Meijgaard E., Samuelsson P., Bamba Sylla M., and Sushama L., (2012). Precipitation Climatology in an Ensemble of CORDEX-Africa Regional Climate Simulations. *Journal of Climate*, **25**, 6057–6078.

Oudin L., Perrin C., Mathevet T., Andréassian V. and Michel C. (2006). Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. Journal of Hydrology, 320(01-févr) :62-83

Pappenberger, F., Beven, K. (2006). Ignorance is bliss: or seven reasons not to use uncertainty analysis. *Water Resources Research*, http://dx.doi.org/10.1029/2005WR004820.

Pappenberger, F., Matgen, P., Beven, K. J., Henry, J. B., Pfister, L., and de Fraipont, P. (2006). Influence of uncertain boundary conditions and model structure on flood inundation predictions, Adv. Water Resour., 29, 1430–1449, 2006

Pechlivanidis I.G., Jackson, B.M., McIntyre N.R., Wheater.H.S. (2011). Catchment scale hydrological modelling: a review of model types, calibration approaches and

uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST Journal* .13 (3) 193-214

Pelletier, P.M.,(1988). Uncertainties in the single determination of river discharge: a literature review, *Canadian Journal of Civil Engineering*, 15, 834-850.

Pitman, W.V., (1973). A mathematical model for generating Monthly River flows from meteorological data in South Africa. Report No. 2/73, Hydrological Research Unit, University of the Witwatersand, Johannesburg, South Africa.

Pitman, W.V., (1978). Trends in stream flow due to upstream land-use changes. *Journal of Hydrology*. 39, 227–237.

Priestley, C. H. B. and Taylor, R. J. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, 100, 81–92.

Raising Irrigation Productivity and Releasing Water for Inter-Sectoral Needs (RIPARWIN) (2006). University of East Anglia, Norwich, UK/Sokoine University of Agriculture, Morogoro, Tanzania/International Water Management Institute (IWMI), Pretoria, South Africa, 37 pp

Refsgaard, J. C., (1997). Parameterization, calibration and validation of distributed Hydrological models. *Journal of Hydrology*, 198, 69-97.

Refsgaard, J. C., and Storm, B., (1995). MIKE SHE, in Computer Models of Watershed Hydrology, edited by V. P. Singh, pp. 809-846, Water Resources Publications, Highlands Ranch, Colorado.

Refsgaard, J. C., van der Sluijs, J. P., Brown, J. and van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources*, 29, 1586-1597.

Ritter, A. and Muñoz-Carpena., R., (2013). Predictive ability of hydrological models: objective assessment of goodness-of-fit with statistical significance. Journal of Hydrology 480(1):33-45.

Rufiji Basin Water Office (RBWO) Irrigation. (2008). Hydrogeological and Geophysical survey report for Usangu, at district Prison area (Isisi village), Montfort Secondary School (Mwakaganga village), Mambi and Majenje villages in Mbarali district - Mbeya region, Tanzania.

Rwetabula J., Desmedt F., Rebhun M. (2007). Prediction of runoff and discharge in the Simiyu River using the wetspa model. *Hydrology and Earth System Sciences Discussions*, **4**(2), 881908.

Saltelli, A., Chan, K., and Scott, M., (eds) (2000). Sensitivity analysis. John Wiley & Sons Ltd., Chichester, UK.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., Carrillo,G., (2011). Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Science Discussions*, **8**, 4495–4534.

Sawunyama, T., (2009). Evaluating Uncertainty in Water Resources Estimation in Southern Africa: A Case Study of South Africa. PhD thesis, Rhodes University, Grahamstown, South Africa. http://eprints.ru.ac.za/.

Sawunyama, T., and Hughes, D.A., (2007) Assessment of rainfall-runoff model input uncertainties on simulated runoff in southern Africa. Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resource Management (Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007). *IAHSPubl.,* **313**, 98-106.

Sawunyama, T., and Hughes D.A., (2008) Application of satellite-derived rainfall estimates to extend water resource simulation modelling in South Africa. *Water SA*, **34**(1), 1-9.

Schaake, J.V., Duan, Q., Koren, V.I., Cong, S. (1997). Regional parameter estimation of land surface parameterizations for GCIP large-scale area southwest, Paper presented at 13th Conference on Hydrology. American Meteorology Society, Long Beach, CA.

Schaake, J., Franz, K., Bradley, A., and Buizza, R. (2006). The Hydrologic Ensemble Prediction Experiment (HEPEX), *Hydrological and Earth System Science Discussion*, 3321–3332.

Schaefli B., and Gupta H., (2007).Do Nash values have value?, *Hydrological Processes*, **21**,2075-2080.

Schertzer, D., L Tchiguirinskaia, S. Lovejoy, and P. Hubert. (2010). No monsters, no miracles: In nonlinear sciences hydrology is not an outlier! *Hydrological Sciences Journal*, **55**(6): 965-979.

Schmidt, E.J., and Schulze, R.E., (1987). Flood volume and peak discharge from small catchments in southern Africa, based on SCS technique. Water Research Commission, Pretoria, and Technology Transfer report TT/3/87. pp 164 (ISBN 78 1)

Schulze, R.E., (2000). Modelling hydrological responses to land use and climate change: A southern Africa perspective, *Ambio*. **29**, 12-22.

Schuol, J., and Abbaspour K.C. (2006). Calibration and uncertainty issues of a hydrological model (SWAT) applied to West Africa. *Advanced Geosciences*, **9**,137 – 143

Sellami, H., La Jeunesse, I., Benabdallah, S., and Vanclooster, M., (2013). Parameter and rating curve uncertainty propagation analysis of the SWAT model for two small Mediterranean watersheds. *Hydrological Sciences Journal*, **58** (8), 1635–1657.

Singh, V.P., (1997). Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph, *Hydrological Processes*, **11**(12), 1649-1669.

Sivapalan, M., (2003). Process complexity at hill slope scale, process simplicity at the watershed scale: is there a connection? *Hydrological Processes*, **17**, 1037-1041.

Sivapalan, M., (2005). Pattern, process and function: elements of a unified theory of hydrology at the catchment scale. In Encyclopedia of Hydrological Sciences, Anderson MG (ed). John Wiley & Sons, Inc.: Chichester, UK; 193–220.

Sivapalan, M., Takeuchi K.,. Franks S.W, Gupta V.K., Karambiri H., Lakshmi V., Liang X.,. McDonnell J.J, Mendiondo E.M., O'Connell P.E., Oki T., Pomeroy J.W., Schertzer D., Uhlenbrook S., Zehe, E. (2003). IAHS Decade on predictions in ungauged basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, **48** (6), 857-880.

Sivapalan, M., Yaeger M. A., Harman C. J., Xu X., and Troch P. A., (2011). Functional model of water balance variability at the catchment scale: Evidence of hydrologic similarity and space-time symmetry, *Water Resources Research*, **47**, (W02522)

SMUWC (2001). Sustainable Management of the Usangu Wetland Final Report. Supporting Report 14, Environmental functions study. Directorate of Water Resources Dar es Salaam, Tanzania. 70p.

Sorooshian, S., and Gupta V.K., (1995).Model Calibration, Computer models of watershed hydrology, edited by Singh, V.P., Water Resources Publications, USA.

SRTM (2004). DEM data from International Centre for Tropical Agriculture (CIAT), available from the CGIAR-CSI SRTM 90m Database: (http://srtm.csi.cgiar.org)

Stisen, S., Jensen K. H., Sandholt I., and Grimes, D., (2008). A remote sensing driven distributed Hydrological model of the Senegal River basin, *Journal of Hydrology*, **354**(1-4), 131-148

Stisen, S., Sandholt, I., (2009). Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modeling. *Hydrological processes* **24**, p879-891.

Tang, Y., Reed P., van Werkhoven K., and Wagener T., (2007). Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis, *Water Resources Research*, **43**, (W06415)

Tetzlaff D. Carey S.K, Laudon H, and and McGuire K., (2010). Catchment processes and heterogeneity at multiple scales—benchmarking observations, conceptualization and prediction. *Hydrological Process*. **24**, 2203–2208

Troch, P.A., Carrillo GA, Heidb¨uchel I, Rajagopal S, Switanek M, Volkmann THM, Yaeger M., (2009). Dealing with landscape heterogeneity in watershed hydrology: a review of recent progress toward new hydrological theory. *Geography Compass* **3**, 375–392.

Tshimanga, R.M., (2012).Hydrological uncertainty analysis and scenario-based stream flow modelling for the Congo River Basin. PhD thesis, Rhodes University repository (http://eprints.ru.ac.za/2937/), South Africa.

Tshimanga, R.M., and Hughes, D.A, (2012). Climate change and impacts on the hydrology of the Congo Basin: the case of the northern sub-basins of the Oubangui and Sangha rivers. *Physics and Chemistry of the Earth*, 50–52, 72–83

Tsonis, A.A., (2004). Is global warming injecting randomness into the climate system? *American Geophysical Union*. **85** (38) 361–364

UNEP (2009). Climate Change Science Compendium. http://www.unep.org/pdf/

URT (2012). Population and Housing Census 2012 Report. National Bureau of Statistics, Dar es Salaam, Tanzania.

Valimba, P., Camberlin P., Yves R., Servat E., Hughes D.A (2006). Influences of ENSO and SST variations on the interannual variability of rainfall amounts in southern Africa. *AHS-AISH Publication*, **308**, 362-368.

Van Griensven, A., Meixner T.,Srinivasan R, and Grunwals S,. (2008). Fit-for-purpose analysis of uncertainty using split-sampling evaluations. *Hydrological Sciences Journal*, **53**(5), 1090-1103

Van Griensven, A., Ndomba, P., Yalew, S., and Kilonzo, F. (2012). Critical review of SWAT applications in the upper Nile basin countries, *Hydrology and Earth System Science*. 16, 3371-3381

Velazquez, ´J., Schmid J., Ricard S., Muerth M. J., Gauvin St-Denis B., Minville M., Chaumont D., Caya D., Ludwig R., and Turcotte R., (2013). An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources. *Hydrology and Earth System Science*, 17, 565–578.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W. and Sorooshian, S. (2003b). Effective and efficient algorithm for multi-objective optimization of hydrologic models. *Water Resources Research*, 39(8), 1214

Vrugt, J. A., Gupta, H. V., Bouten, W. and Sorooshian, S. (2003a). A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*. 39(8), 1201

Vrugt, J.A., Gupta, H.V., Dekker, S.C., Sorooshian, S., Wagener, T., Bouten, W., (2006). Confronting parameter uncertainty in hydrologic modeling: Application of the SCEM-UA algorithm to the Sacramento Soil Moisture Accounting model. *Journal of Hydrology*, 325, 288-307.

Vrugt, J.A, and Robinson B. (2007). Improved evolutionary optimization from genetically adaptive multi-method search. *Proceedings of the National Academy of Science U. S. A*. 104(3), 708 – 711.

Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V. and Sorooshian, S., (2001). A framework for development and application of hydrological models. *Hydrology and Earth Systems Science*, 5(1), 13-26.

Wagener, T., Franks, S., Bøgh, E., Gupta, H.V., Bastidas, L., Nobre, C., Oliveira Galvão, C.,(2005). Regional hydrologic impacts of climate change – Impact assessment and decision making. *IAHS Redbook Publ.,* 295, 356pp. ISBN 1-901502-08-2.

Wagener, T., Gupta, H.V., (2005). Model identification for hydrological forecasting under uncertainty. Stochastic. *Environmental Research and Risk Assessment*. doi: 10.1007/s00477-005-0006-5.

Wagener, T., Kollat, J., (2007). Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox. *Environmental Model. Software*, **22**, 1021-1033.

Wagener, T., McIntyre N., Lees M., Wheater H. and Gupta H., (2003). Towards reduced uncertainty in conceptual rainfall-runoff modeling: dynamic identifiability analysis. *Hydrological Processes*. **17**(2), 455-476

Wagener, T., and Montanari A. (2011). Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, *Water Resources Research*, **47**, (W06301)

Wagener, T, Sivapalan M, Troch P, Woods R. (2007). Catchment classification and hydrologic similarity. *Geography Compass*, **1**: 10.

Wagener, T., Wheater H.S. and Gupta H.V., (2004). Rainfall-Runoff Modelling in Gauged and Ungauged Catchments. Imperial College Press, London, UK, 1-306 pp.

Wagener, T., Wheater, H.S., (2006). Parameter estimation and regionalisation for continuous rainfall-runoff models including uncertainty. *Journal of Hydrology*, **320**(1-2), 132-154.

Wang, Q. J., Shrestha, D. L., Robertson, D. E., and Pokhrel,P. (2012). A log-sinh transformation for data normalization and variance stabilization. *Water Resources Research*, **48**, W05514

Wang, M., Qin D., Lu, C., Li Y., (2009). Modeling Anthropogenic Impacts and Hydrological Processes on a Wetland in China. *Water Resources Management*. **24** (11) 2743-2757

Warburton, M.L., Schulze, R.E., Jewitt, G.P.W., (2012). Hydrological impacts of land use change in three diverse South African catchments. *Journal of Hydrology*. doi:10.1016/j.jhydrol.2011.10.028

Weiler, M., and McDonnell. J.J.(2007). Conceptualizing lateral preferential flow and flow networks and simulating the effects on gauged and ungauged hillslopes. *Water Resources Research*.**43** (W03403).

Westerberg, I.K., Gong L., Beven K.J., Seibert J., Semedo A. C., Xu, Y and Halldin, S. (2013). Regional water-balance modelling using flow-duration curves with observational uncertainties (2013). *Hydrology and Earth System Science Discussion*, **10**, 15681–15729.

Westerberg, I.K, Guerrero J.-L., Younger P. M., Beven K. J. Seibert J., Halldin S., Freer J. E, and. Xu., C.Y (2011). Calibration of hydrological models using flow-duration curves. *Hydrology and Earth System Science*, **15**, 2205–2227

Wheater, H.S., Jakeman, A.J. and Beven, K.J., (1993). Chapter 5 –Progress and directions in rainfall-runoff modelling, In: Modelling change in environmental systems (Eds. A.J. Jakeman, M.B. Beck and M.J. McAleer), Wiley, Chichester, UK, 101-132.

Winsemius, H. C., Schaefli B, Montanari A, and Savenije H. H. G. (2009). On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, Water Resources. Research., **45**, W12422.

WWF Tanzania Country Office (WWF-TCO), (2010). Assessing Environmental Flows for the Great Ruaha River, and Usangu Wetland, Tanzania. Technical Report. WWF Tanzania Country Office.

Yadav, M., Wagener, T., Gupta, H.V., (2007). Regionalisation of constraints on expected watershed response behaviour. *Advances in Water Resources*, **30**, 1756-1774.

Yang, J., Reichert P., Abbaspour K.C., Xia J., Yang H., (2008). Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China, *Journal of Hydrology*, **358** (1–2), 1-23

Yokoo, Y., and Sivapalan, M. (2011). Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis. *Hydrology and Earth System Science*, **15**, 2805–2819

Younger, P. M., Freer, J. E. Beven, K. J., (2009). Detecting the effects of spatial variability of rainfall on hydrological modelling within an uncertainty analysis framework. *Hydrological Processes*, **23**, 1099-1085

APPENDICES

APPENDIX A: Estimated elevation and slope values

Sub-basinID	Catch Area(km^2)	ElevMin	ElevMax	Elevation	Slope
1ka10	233	1031	2522	1408.2	8.2
1ka11	1600	1353	2457.7	1758.6	10.1
1ka12	807	1122	1899	1521	4.8
1ka56	182	1122	1885	1455	5.2
1ka8a	783	1025	2950	1882.3	16.7
1ka15	1221	1043	2080	1722	6.6
1ka32a	838	1714	2130	1874.5	6.8
1ka37	3064	661.3	2390.3	1393.8	17.9
1ka51	38	1036	2852	1607	11.7
1ka9	446	1031	2720	1535.6	10
1ka16a	75	1036	2852	1607	11.7
1ka22	461	1714	2140	1927.8	12
1ka41	4368	1115.7	1431.3	1279.4	2.3
1ka50a	102	1036	2852	1607	11.7
1ka7a	169	1025	2950	1882.3	16.7
1ka33	618	1020	1870	1216.6	5.8
1ka5a	7315	811.3	1417.7	948.2	4.2
1ka2a	277	772	2301	1654.5	12
1ka59	2539	900	1691.3	1130.7	5.7
1ka61	3691	532.6	2173.6	962.2	18
1ka21	477	772	2301	1654.5	12
1ka27	4244	1010.1	1438.9	1097	2.7
1ka39	843	1714	2130	1874.5	6.8
1ka42	5999	993	1457.1	1095.3	2.8
1ka31	2299	771.7	2090.3	1420.5	9.6
1ka71a	1055	1023.8	1187	1087.3	1.1
UG2	1709	1122	1885	1455	5.2
UG10	3038	1176.7	1586.7	1317.4	3.6
UG12	612	566	2184	1195.2	19.9
UG13	1107	534	2288	1358.3	23.8
UG4	500	1026	2528	1168.1	4.6
UG21	420	1026	2528	1168.1	4.6

Appendix A continued.....

UG18	501	1025	1597	1237.5	3
UG6	480	1022.5	1047.5	1030.9	0.5
UG20	175	1022.5	1047.5	1030.9	0.5
UG14	5540	298.2	1687.4	692.2	15.7
UG11	2046	933.3	2220	1368.8	11.4
UG1	1176	1041	2102	1316.4	7.4
UG3	1463	1026	1723.7	1281.8	3.5
UG5	2149	1052	1822.5	1439.3	5
UG7	4609	1191	1818.3	1345.8	5.3
UG8	5155	1176.4	1539.2	1285.8	4.2
UG9	1928	1091.5	1669	1212.4	4.7
UG24	8494	740.7	1536.4	987.2	5.1
UG22	126	1020	1134	1051.8	1.3
UG17	181	1516	2798	2085.3	14.9
UG16	277	1020	1232	1047.9	0.8
UG15	198	1026	1074	1039.5	0.6

APPENDIX B.: Estimated soil texture classes

Catch ID	Loam	Sandy_Clay_Loam	Sandy_Loam	Clay_Loam	Clay
1ka10	0	32.8	67.2	0	0
1ka11	50.2	10.9	38.8	0	0
1ka12	39.3	0.3	60.4	0	0
1ka56	9.5	1.8	88.7	0	0
1ka8a	21	28.9	43.7	2.6	3.8
1ka15	71.7	1	26.2	0	1.2
1ka32a	92.7	0	7.3	0	0
1ka37	20.9	0	81.6	0	0
1ka51	0	52.3	45.2	1.7	0.7
1ka9	0.6	55.8	43.6	0	0
1ka16a	0	52.3	45.2	1.7	0.7
1ka22	100	0	0	0	0
1ka41	73.7	26.1	0	0.3	0
1ka50a	0	52.3	45.2	1.7	0.7
1ka7a	21	28.9	43.7	2.6	3.8
1ka33	0	39.9	42	0	18.1
1ka5a	91.9	91.9	0	0.3	0
1ka2a	28.9	0	71.1	0	0
1ka59	1.1	14.5	84.2	0	0.2
1ka61	73	9.7	17.3	0	0
1ka21	28.9	0	71.1	0	0
1ka27	0	54.3	19.4	0	26.3
1ka39	92.7	0	7.3	0	0
1ka42	98.3	0	1.7	0	0
1ka31	11.6	0	89.1	0	0

Appendix B continued……

1ka71a	13.1	36.5	0	0	50.4
UG2	9.5	1.8	88.7	0	0
UG10	100	0	0	0	0
UG12	100	0	0	0	0
UG13	100	0	0	0	0
UG4	0.1	43.5	8.1	0	48.4
UG21	0.1	43.5	8.1	0	48.4
UG18	32.8	41.1	0	0	26.1
UG6	0	52.6	0	0	47.4
UG20	0	52.6	0	0	47.4
UG14	71.7	0	26.9	1.4	0
UG11	20	0.3	66.9	12.9	0
UG1	21.6	71.5	5.4	1.5	0
UG3	49.8	42.4	0	0	7.8
UG5	30	18.2	51.7	0	0
UG7	84.3	0	15.7	0	0
UG8	100	0	0	0	0
UG9	100	0	0	0	0
UG24	6.3	48.9	37	0	0
UG22	0	53.8	0	0	46.2
UG17	65.2	0	34.8	0	0
UG16	0	80	0	0	20.1
UG15	0.1	43.5	8.1	0	48.4

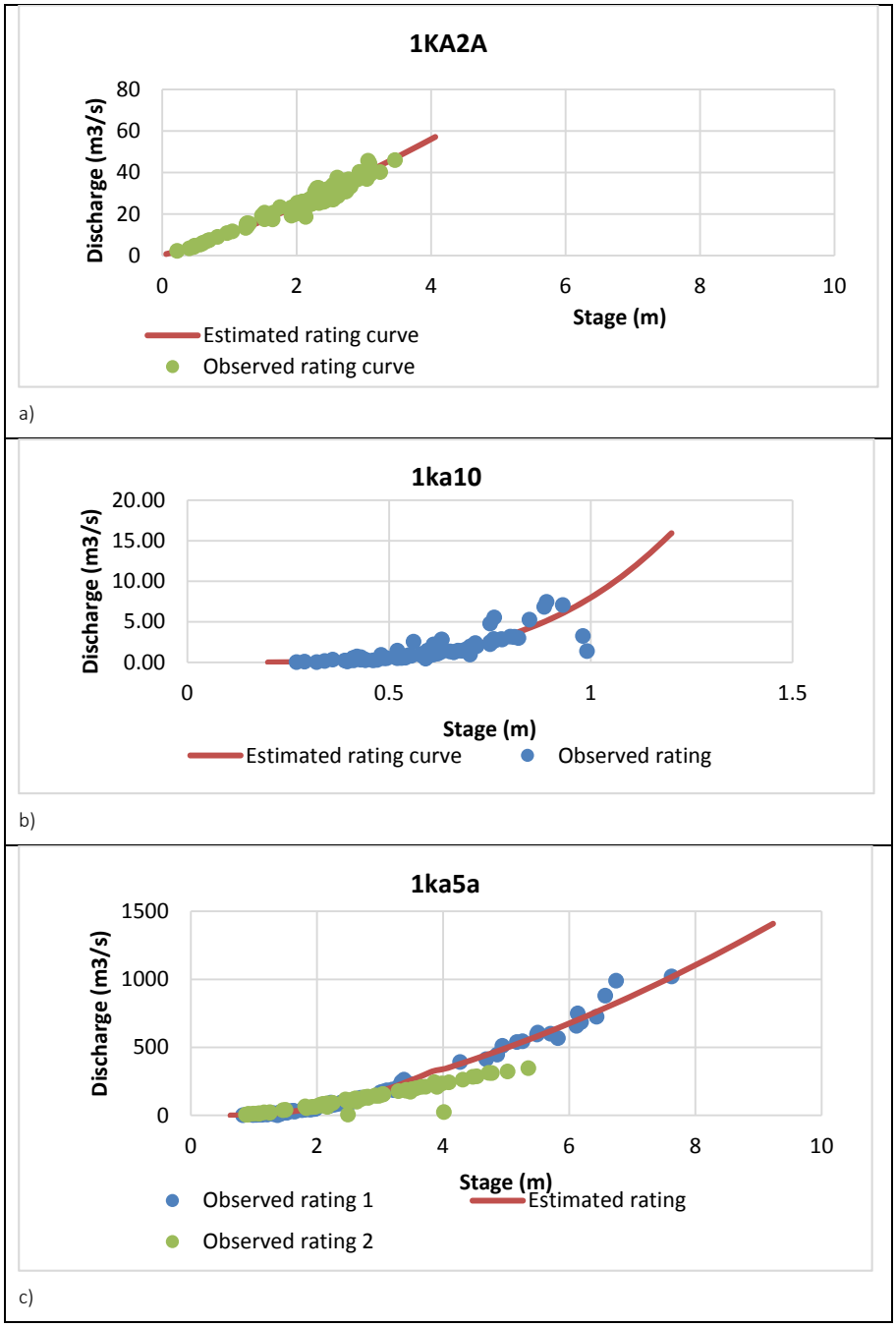
APPENDIX C.: Estimated land cover classes

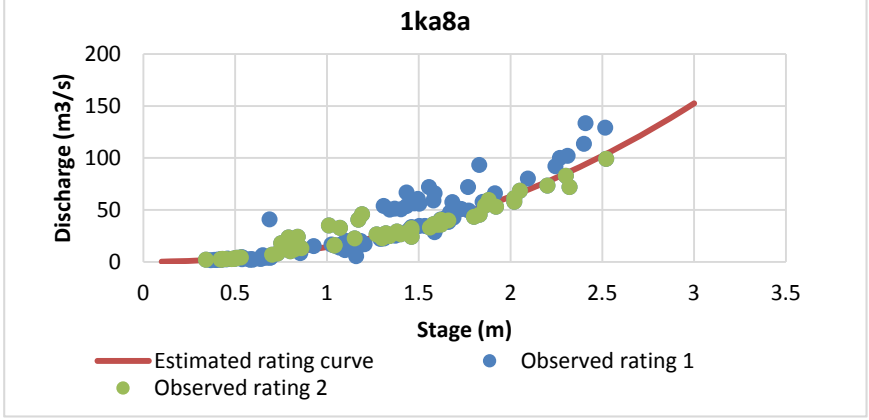
ID	Area(km^2)	CRDY	CRWO	GRASS	SHRUB	SAVAN	FODB	FOEB	FOMI	WATB
1ka10	233	8.8	72.5	3.6	3.8	11.3	0	0	0	0
1ka11	1600	23.5	69.6	9.6	2.1	7.2	1.1	0.7	1.2	0
1ka12	807	2	97.5	0.1	0.3	0.2	0	0	0	0
1ka56	182	16.9	18	7.9	0	7.1	0	0	0	0
1ka8a	783	13.5	58.7	2.6	4.7	18.9	1.5	0.1	0	0
1ka15	1221	67.1	14.4	0	3.6	12.7	1.7	0.3	0.1	0
1ka32a	838	39.3	27.4	0.1	13.1	13.2	1	3.2	2.4	0
1ka37	3064	25.7	5.5	12.2	6.9	16.9	16.5	8.2	8.2	0
1ka51	38	15.3	66.6	9.5	0.7	7.8	0	0	0	0
1ka9	446	16.6	55	4.9	3.7	19.5	0.3	0	0	0
1ka16a	75	15.3	66.6	9.5	0.7	7.8	0	0	0	0
1ka22	461	28.5	9.7	14.8	12.5	10.8	7.4	16.4	0	0
1ka41	4368	62.2	0.2	1.3	0	35.9	0.5	0	0	0
1ka50a	102	15.3	66.6	9.5	0.7	7.8	0	0	0	0
1ka7a	169	13.5	58.7	2.6	4.7	18.9	1.5	0.1	0	0
1ka33	618	30.6	43.5	6.5	1.6	15.6	2.3	0	0	0
1ka5a	7315	22.3	12.4	15.5	6.2	40.6	1.1	0	0	2
1ka2a	277	64.4	3.1	4.7	8	16.6	1.6	0.9	0.8	0
1ka59	2539	37.5	16.8	4.8	0.1	32.6	8	0	0.2	0
1ka61	3691	23.1	17.6	11.7	10.7	26.6	4.3	4.5	0.9	0.5
1ka21	477	64.4	3.1	4.7	8	16.6	1.6	0.9	0.8	0
1ka27	4244	30.2	38.9	14.1	1.8	13.9	1	0	0	0
1ka39	843	19.7	13.7	0.1	6.6	6.6	0.5	1.6	1.2	0
1ka42	5999	64.5	1.6	0.3	0	31.8	1.7	0	0	0
1ka31	2299	0	13.2	11.9	7.7	22.8	2.2	0.3	0.3	0
1ka71a	1055	30.4	12.3	24.9	10	16.1	5.3	0	1	0
UG2	1709	33.9	36.1	15.9	0	14.2	0	0	0	0
UG10	3038	75	2.2	0	0	21.4	1.4	0	0	0
UG12	612	10.3	22	14.6	2.1	24.7	21.6	0.4	4.3	0
UG13	1107	22.8	12.9	0.7	1.8	8.6	29.7	5.6	17.9	0
UG4	500	17.2	22.1	38.4	1.1	21.2	0	0	0	0
UG21	420	17.2	22.1	38.4	1.1	21.2	0	0	0	0
UG18	501	17.4	50.3	19	0.2	13.1	0	0	0	0
UG6	480	47.2	1.6	38.4	1.1	11.8	0	0	0	0
UG20	175	47.2	1.6	38.4	1.1	11.8	0	0	0	0
UG14	5540	5.4	7.5	0	4.3	3.1	19.7	2.5	57.5	0
UG11	2046	26.6	15.5	18.3	7.9	24.9	5.2	1	0.5	0

Appendix C continued…….

UG1	1176	14	61.8	10.6	0.1	13.3	0.2	0	0	0
UG3	1463	14.9	37.3	11.5	2.1	21.3	11.2	0	1.7	0
UG5	2149	10.4	23.8	4	8.9	5.7	34.9	0.5	11.8	0
UG7	4609	56.5	10.8	1.4	1	7.7	18.9	0	3.7	0
UG8	5155	87.9	0.7	1.5	0	8.1	1.8	0	0	0
UG9	1928	79.5	0.4	0.4	0	18.5	1.3	0	0	0
UG24	8494	22.9	24.1	4.7	6.1	39.8	2.2	0.1	0.1	0
UG22	126	12.4	52	0.7	12.2	0	0	0	0	0
UG17	181	7.8	62.6	23.2	2	3.5	0.7	0.1	0.2	0
UG16	277	46.8	27.8	12.9	0.5	11.7	0.3	0	0	0
UG15	198	17.2	22.1	38.4	1.1	21.2	0	0	0	0

APPENDIX D: Examples of poor rated rating curves




1ka8a
Discharge (m3/s)
200
150
100
50
0
0
0.5
1
1.5
2
2.5
3
3.5
Stage (m)
Estimated rating curve
Observed rating 1
Observed rating 2

d)

APPENDIX E : Minimum and Maximum constraints values

Catch ID	Status	Location	Bounds	MMQ	MMGW Rech	FDC10%	FDC50%	FDC90%	% Zero Flows
1ka7a	gauged	Headwater	Min	7.7	8	2	0.4	0.05	0
			Max	11.5	15	3	0.7	0.25	0
1ka12	gauged	Headwater	Min	4.3	2	2	0.1	0	0
			Max	12	10	3.5	0.3	0.15	20
1ka9	gauged	Headwater	Min	14.5	8	2	0.3	0.05	0
			Max	21.5	12	3	0.6	0.25	0
1ka8a	gauged	Headwater	Min	31	8	2	0.3	0.05	0
			Max	46	12	3	0.6	0.25	0
1ka11	gauged	Headwater	Min	30	4	2	0.4	0.05	0
			Max	45	15	3	0.7	0.25	0
1ka51	gauged	Headwater	Min	1.5	8	2	0.4	0.1	0
			Max	2	15	3	0.7	0.25	0
1ka50a	gauged	Headwater	Min	3	5	2	0.35	0.05	0
			Max	4.5	10	3	0.6	0.2	0
1ka16a	gauged	Headwater	Min	3	5	2	0.5	0.02	0
			Max	4.5	10	3	0.75	0.2	0
1ka10	gauged	Headwater	Min	2.7	2	2.5	0.2	0	0
			Max	3.9	8	3.5	0.5	0.15	10
ug17	ungauged	Headwater	Min	4.2	4	2	0.25	0.05	0
			Max	6.5	12	3	0.6	0.25	0
ug18	ungauged	Headwater	Min	3	2	1.8	0.35	0.05	0
			Max	5	8	3	0.7	0.25	20
1ka56	gauged	Headwater	Min	3.6	4	2	0.35	0.05	0
			Max	5.5	12	3	0.6	0.2	0
1ka32a	gauged	Headwater	Min	8	2	2	0.4	0.05	0
			Max	15	12	3	0.8	0.25	0
ug3	ungauged	Headwater	Min	7.3	2	1.8	0.35	0.05	0
			Max	14.6	8	3	0.7	0.25	15
1ka15	gauged	Headwater	Min	15	4	2	0.35	0.05	0
			Max	25	12	3	0.7	0.25	0

APPENDIX E continued…….

Catch ID	Status	Location	Bounds	MMQ	MMGW Rech	FDC10%	FDC50%	FDC90%	% Zero Flows
1ka22	gauged	Headwater	Min	8	5	1.4	0.6	0.2	0
			Max	12	15	2.5	1.1	0.5	0
ug5	ungauged	Headwater	Min	6	1	2.2	0.1	0	0
			Max	17	5	3.5	0.2	0.005	10
1ka37	gauged	Headwater	Min	54.5	10	1.3	0.6	0.4	0
			Max	66.8	15	2	1	0.6	0
ug11	ungauged	Headwater	Min	6.5	0.5	2	0.1	0	10
			Max	20	2	3.5	0.2	0.005	30
ug7	ungauged	Headwater	Min	15	0.5	1.8	0	0	10
			Max	40	2	3.5	0.2	0.005	50
ug8	ungauged	Headwater	Min	20	0.5	2.2	0.1	0	10
			Max	50	2	3.5	0.2	0.005	30
ug13	ungauged	Headwater	Min	3.5	0.5	1.8	0.1	0	10
			Max	10	2	3.5	0.2	0.005	30
1ka41	Gauged	Headwater	Min	12	0	1.5	0.05	0	10
			Max	40	2	4	0.2	0.005	50
ug10	ungauged	Headwater	Min	10	0.5	2.2	0.1	0	10
			Max	30	2	3.5	0.2	0.005	30
ug9	ungauged	Headwater	Min	6	0.5	1.8	0	0	10
ug12	ungauged	Headwater	Min	1.5	0.5	1.8	0.1	0	10
			Max	5	2	3.5	0.2	0.005	40
ug15	ungauged	Intermediate	Min	5.6	8	2	0.35	0.05	0
			Max	8.4	15	3	0.7	0.25	0
ug21	ungauged	Intermediate	Min	4	4	2	0.25	0.05	0
			Max	7	10	3	0.6	0.25	15
ug16	ungauged	Intermediate	Min	6.6	8	2	0.25	0.05	0
			Max	10	15	3	0.6	0.25	0
ug22	ungauged	Intermediate	Min	1.2	2	2	0.25	0	0
			Max	2	10	3	0.6	0.25	15

Appendix E continued……..

Catch ID	Status	Location	Min_Max	MMQ	MMGW Rech	FDC10%	FDC50%	FDC90%	% Zero Flows
ug1	ungauged	Intermediate	Min	12	3	2	0.2	0.05	0
			Max	19	10	3	0.5	0.25	15
ug20	ungauged	Intermediate	Min	1.7	3	2	0.25	0	0
			Max	2.8	10	3	0.6	0.15	15
ug6	ungauged	Intermediate	Min	4.8	3	2	0.25	0.005	0
			Max	7.6	10	3	0.6	0.15	15
ug2	ugauged	Intermediate	Min	17	4	2	0.25	0.005	0
			Max	27	12	3	0.6	0.15	15
ug4	ungauged	Intermediate	Min	5	4	2	0.25	0.05	0
			Max	8	10	3	0.6	0.25	15
1ka39	gauged	Intermediate	Min	16	2	1.5	0.5	0.1	0
			Max	24	10	2.5	1	0.5	0
1ka21	gauged	Intermediate	Min	5.5	2	1.5	0.5	0.1	0
			Max	8.6	10	2.5	1	0.5	0
1ka27	gauged	Downstream	Min	0	0	0	0	0	0
			Max	20	8	3.5	1.2	0.5	60
1ka71a	gauged	Downstream	Min	0	0	1	0	0	0
			Max	2	8	3	1	0.5	60
1ka33	gauged	Downstream	Min	10	4	1.5	0.35	0.05	0
			Max	15	12	3	0.7	0.25	0
1ka2a	gauged	Downstream	Min	2	2	1.6	0.3	0.05	0
			Max	4	10	3	0.8	0.4	0
1ka31	gauged	Downstream	Min	7	1	1.6	0.3	0.01	0
			Max	11	5	2.6	0.7	0.2	0
ug14	ungauged	Downstream	Min	100	5	1.8	0.35	0.05	0
			Max	150	15	3	0.6	0.2	0
1ka59	gauged	Downstream	Min	0	0	1	0	0	0
			Max	2	8	3	1	0.5	60
1ka42	gauged	Downstream	Min	15	0	1.8	0	0	10
			Max	50	2	3.5	0.2	0.005	50
1ka61	gauged	Downstream	Min	10	0.5	1.8	0.1	0	10
			Max	20	2	3.5	0.2	0.005	50
1ka5a	gauged	Downstream	Min	15	0.5	1.8	0	0	10
			Max	40	2	3.5	0.2	0.005	50

APPENDIX F: Initial parameter ranges

Group one and two

Catch ID	1ka16a,1ka50a											
	1ka51,1ka7a,1ka8a,											
	1ka16a and 1ka9	1ka11	1ka10	1ka12	Ug1	Ug22	Ug21	Ug4	Ug20	Ug6	Ug2	1ka56
Paramete	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max
PI1	1.5-3	1.5-3	1.5-3	1.5-3	1.5-3	1.5-3	1.5-3	1.5-3	1.5-3	1.5-3	1.5-3	1.5-3
ZMIN	10.0-60	20-100	10.0-80	20-80	30-80	30-80	30-80	30-80	30-80	30-80	30-80	Oct-80
ZMAX	200-600	300-900	300-600	300-800	300-600	300-600	400-600	400-600	400-600	400-600	400-600	200-600
ST	250-500	500-1100	300-600	300-800	200-500	200-500	300-700	300-700	300-700	300-700	300-700	200-600
POW	1.8-3	2-3.5	2.5-3.5	2.5-4	2.5-4.0	2.5-4.0	2.5-4	2.5-4	2.5-4	2.5-4	2.5-4	42.0-4.0
FT	10.0-40	10.0-25	5.0-20	0-8	2.0-8.0	2.0-10	2.0-15	2.0-15	2.0-15	2.0-15	2.0-15	10.0-40.0
GW	10.0-40	10.0-25	5.0-20	5.0-20	5.0-12	5.0-15	10.0-25	10.0-25	10.0-25	10.0-25	10.0-25	10.0-40.0
R	0-0.7	0-0.5	0.0.6	0.0.5	0.0-0.5	0.0-0.5	0.0-0.5	0.0-0.5	0.0-0.5	0.0-0.5	0.0-0.5	0.0-0.7
GPOW	2.0-4.0	3.0-4.0	3.0-4.0	3-4.5	3.0-5	3.0-5	3.0-5.0	3.0-5	3.0-5	3.0-5	3.0-5	3.0-5.0
D.DENS	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.3
T	20	20	10	10	20	20	20	20	20	20	20	20
S	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
GW SLOP	0.01	0.01	0.01	0.01	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.001
RSF	0.2-1	0.4-1	0.2-1	0.4-2	0.2-1.0	0.2-1.0	0.2-1.0	0.2-1	0.2-1	0.2-1	0.2-1	0.4-2.0

Group three

Catch ID	1ka56	1ka32a	1ka15	1ka22	1ka37	1ka39	1ka21	1ka2a	1ka31	1ka33
Parameters	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max
PI1	1.5-3	1.5-3.0	1.5-3.0	1.5-3.0	1.5-2.0	1.5-3.0	1.5-3.0	1.5-3.0	1.5-3.0	1.5-3.0
ZMIN	10.0-80	20-60	20-80	30-100	10.0-60	30-100	30-100	30-60	40-100	40-80
ZMAX	200-600	400-800	300-900	300-800	300-500	300-800	300-800	400-800	600-1000	400-800
ST	200-600	800-1200	600-1200	600-1200	500-1200	500-1200	500-1200	400-1000	600-1200	600-1200
POW	42.0-4.0	2.0-4.0	2.0-4.0	2.0-4.0	1.8-3.0	2.0-4.0	2.0-4.0	2.0-4.0	2.0-4.0	2.0-4.0
FT	10.0-40.0	8.0-20	5.0-20	5.0-20	10.0-40	5.0-30	5.0-20	5.0-20	2.0-10	5.0-20
GW	10.0-40.0	5.0-20	5.0-20	10.0-40	30-60	5.0-40	5.0-40	5.0-30	2.0-10	5.0-20
R	0.0-0.7	0.0-0.5	0.0-0.4	0.0-0.6	0.4-0.8	0.0-0.6	0.0-0.6	0.0-0.6	0.0-0.4	0.0-6.0
GPOW	3.0-5.0	3.0-5.0	3.0-5.0	3.0-4.0	1.8-2.5	3.0-5.0	3.0-5.0	3.0-5.0	3.0-5.0	3.0-5.0
D.D	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
T	20	20	20	20	15	20	20	20	20	20
S	0.004	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
GWslope	0.001	0.001	0.001	0.001	0.005	0.001	0.001	0.001	0.001	0.001
RSF	0.4-2.0	0.4-2.0	0.6-3.0	0.4-2.0	0.2-0.4	0.4-2.0	0.4-2.0	0.6-1.5	0.4-2.0	0.6-1.0

Group four

	1ka41	Ug10	Ug8	Ug7	Ug9	1ka42	Ug3	Ug5	ug18
	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max
PI1	1.5-2	1.5-2	1.5-2	1.5-2	1.5-2	1.5-2	1.5-2	1.5-2	1.5-2
ZMIN	20-60	20-60	20-60	20-60	20-60	20-60	20-60	20-60	20-60
ZMAX	300-500	300-500	300-500	300-500	300-500	300-500	300-500	300-500	300-500
ST	20-500	20-500	20-500	20-500	20-500	20-500	20-500	20-500	20-500
POW	0	0	0	0	0	0	0	0	0
FT	0	0	0	0	0	0	0	0	0
GW	5.0-10	5.0-10	5.0-10	5.0-10	5.0-10	5.0-10	5.0-10	5.0-10	5.0-10
R	0.0-0.7	0.0-0.7	0.0-0.7	0.0-0.7	0.0-0.7	0.0-0.7	0.0-0.7	0.0-0.7	0.0-0.7
GPOW	3	3	3	3	3	3	3	3	3
D.D	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
T	20	20	20	20	20	20	20	20	20
S	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
GWslope	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
RSF	0.6-2	0.6-2	0.6-2	0.6-2	0.6-2	0.6-2	0.6-2	0.6-2	0.6-2

Group five

	1ka71a	1ka27	1ka59	UG24
	Min-Max	Min-Max	Min-Max	Min-Max
PI1	1.5-3	1.5-3	1.5-3	1.5-3
ZMIN	150-200	150-200	150-200	40-120
ZMIN	150-200	150-200	150-200	40-120
ZMAX	800-1200	800-1200	800-1200	500-1200
ST	1200-1600	1200-1600	1200-1600	500-1200
POW	2.5	2.5	2.5	2.5
FT	0	0	0	0
GW	0.5-2	0.5-2	0.5-2	5.0-20
R	0.0-0.2	0.0-0.2	0.0-0.2	0-0.5
Channel Loss TLGMax	0	0	0	0
GPOW	2.0-3	2.0-3	2.0-3	2.0-3
D.D	0.4	0.4	0.4	0.4
T	20	20	20	20
S	0.004	0.004	0.004	0.004
GW slope	0.002	0.002	0.002	0.002
RSF	2.0-6	1.0-4	1.0-4	1.0-4

Group six

	1ka5a	Ug11	Ug12	Ug13	1ka61	Ug14
	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max
PI1	1.5-2.5	1.5-2.5	1.5-2.5	1.5-2.5	1.5-2.5	1.5-3
ZMIN	10.0-60	10.0-60	10.0-60	10.0-60	10.0-60	40-100
ZMAX	300-600	300-600	300-600	300-600	300-600	300-800
ST	300-500	300-500	300-500	300-500	300-500	400-800
POW	0	0	0	0	0	2.5-4
FT	0	0	0	0	0	5.0-15
GW	2.0-15	2.0-15	2.0-15	2.0-15	2.0-15	10.0-30
R	0.0-0.5	0.0-0.5	0.0-0.5	0.0-0.5	0.0-0.5	0.0-0.7
GPOW	3.0-5	3.0-5	3.0-5	3.0-5	3.0-5	3.0-4.0
D.D	0.4	0.3	0.4	0.4	0.4	0.4
T	20	20	20	20	20	20
S	0.004	0.002	0.004	0.004	0.004	0.004
GWslope	0.01	0.005	0.01	0.01	0.01	0.005
RSF	0.4-4	0.4-4	0.4-4	0.4-4	0.4-4	0.4

APPENDIX G: Final parameter values

Group One

Parameters	1ka7a	1ka9	1ka8a	1ka51	1ka50a	1ka16a	ug15	ug16	ug17
ZMIN	35	35	35	35	35	35	35	35	35
ZMAX	400	400	400	400	400	400	400	400	400
ST	350	350	350	350	300	350	350	350	350
POW	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4
FT	25	25	25	25	25	25	25	25	25
GW	25	25	25	30	30	25	25	25	25
TLGMax(mm)	0	0	0	0	0	0	0	0	0
R	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
GPOW	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5
DD	0.3	0.4	0.4	0.2	0.3	0.4	0.3	0.4	0.2
T	20	20	20	20	20	20	20	20	20
S	0.004	0.004	0.004	0.008	0.008	0.008	0.004	0.008	0.008
Gwslope	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
RSF	0.2	0.4	0.6	0.6	0.6	0.6	0.6	0.4	2
Water use parameters									
Irrig.Area (km2) AIRR	21	36.6	54.3	0	0	0	0	0	0
Irrig. return flow fraction IWR	0.15	0.15	0.15	0	0	0	0	0	0
Non-irrg.Direct Demand Ml/year	341	409	815	7.8	26.1	64.5	53.6	73.6	48.1

Group two

Paramenters	1ka10	1ka12	1ka11	ug22	ug1	ug20	ug6	ug2	ug4	ug21	1ka56
ZMIN	45	50	60	55	30	55	55	55	55	55	45
ZMAX	450	550	600	450	450	450	450	450	450	450	400
ST	450	550	800	400	400	400	400	400	400	400	400
POW	3	3.25	2.75	3.25	3.25	3.25	3.25	3.25	3.25	3.25	3
FT	4	4	17.5	8.5	12.5	8.5	11	8.5	11	11	25
GW	12.5	10	17.5	20	20	20	20	20	20	20	25
R	0.3	0.25	0.25	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
GPOW	3.5	3.75	3.5	4	4	4	4	4	4	4	4
DD	0.4	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.3
T	20	20	20	20	20	20	20	20	20	20	20
S	0.004	0.008	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
Gwslope	0.01	0.01	0.01	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.01
RSF	1	0.8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.8
Water use parameters											
Irrig.Area (km2) AIRR	0	0	84	0	126	10	30	146	47	0	0
Irrig. return flow fraction IWR		0	0.15	0	0.2	0.33	0.33	0.05	0.33	0	0
Non-irrg.Direct Demand Ml/year	0	900	472	0	557	152	55	157	151	40	49

Group three

Parameters	1ka15	1ka32a	1ka39	1ka22	1ka21	1ka2a	1ka31	1ka37	1ka33
ZMIN	50	40	65	65	65	45	70	35	60
ZMAX	600	600	550	550	550	600	800	400	600
ST	1200	1000	850	900	850	700	900	850	1200
POW	3	3	3	3	3	3	3	2.4	3
FT	12.5	14	17.5	12.5	12.5	12.5	6	25	12.5
GW	12.5	12.5	22.5	25	22.5	17.5	11	45	12.5
R	0.2	0.25	0.3	0.3	0.3	0.3	0.2	0.6	0.3
GPOW	4	4	4	3.5	4	4	4	2.15	4
DD	0.3	0.3	0.3	0.3	0.3	0.4	0.3	0.3	0.3
T	20	20	20	20	20	20	20	15	20
S	0.008	0.008	0.002	0.008	0.002	0.002	0.002	0.008	0.002
Gwslope	0.005	0.01	0.01	0.01	0.005	0.008	0.005	0.005	0.008
RSF	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.2	0.8
Water use parameters									
Irrig.Area (km2) AIRR	61	0	0	0	0	0	0	0	76.2
Irrig. return flow fraction IWR	0.2	0	0	0	0	0	0	0	0.2
Non-irrg.Direct Demand Ml/year	737	352	147	386	93	66	219	590	413

Group four

	1ka41	ug10	ug9	ug8	ug7	ug5	ug3	ug18	1ka42
ZMIN	25	25	25	25	25	70	70	70	25
ZMAX	400	400	450	450	450	650	650	650	400
ST	350	350	450	450	450	450	450	450	350
POW	2.8	2.8	2.8	2.8	2.8	3.25	3.25	3.25	2.8
FT	0	0	0	0	0	3	8.5	8.5	0
GW	5	5	5	5	5	8.5	20	20	5
R	0.3	0.3	0.3	0.3	0.3	0.35	0.35	0.35	0.3
GPOW	3.75	3	3	3	3	4	4	4	3
DD	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.3	0.4
T	20	20	20	20	20	20	20	20	20
S	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.002	0.004
Gwslope	0.005	0.005	0.005	0.005	0.005	0.012	0.012	0.001	0.005
RSF	1	1	1	1	1	0.4	0.4	0.6	1

Group five and six

Parameters	1ka27	1ka71a	1ka59	ug24	1ka5a	ug11	ug12	ug13	1ka61	ug14
ZMIN	200	200	200	80	35	35	35	35	35	70
ZMAX	1000	1000	1000	850	450	450	400	400	400	550
ST	1500	1500	1500	850	400	350	350	350	350	600
POW	0	0	2.8	2.8	2.8	3.1	0	2.8	2.8	0
FT	0	0	0	0	0	0	0	0	0	0
GW	1.25	1.25	2	50	2	2	8	2	2	2
R	0	0	0	0.3	0.3	0.2	0.3	0.3	0.3	0.3
GPOW	2.5	2.5	2.5	2.5	4	4	4	4	4	3.5
DD	0.4	0.4	0.4	0.4	0.4	0.3	0.4	0.4	0.4	0.4
T	20	20	20	20	20	20	20	20	20	20
S	0.004	0.004	0.004	0.004	0.004	0.002	0.004	0.004	0.004	0.004
Gwslope	0.002	0.002	0.002	0.002	0.01	0.005	0.01	0.01	0.01	0.005
RSF	0.4	5	5	0.4	2	0.8	0.6	2	2	0.4