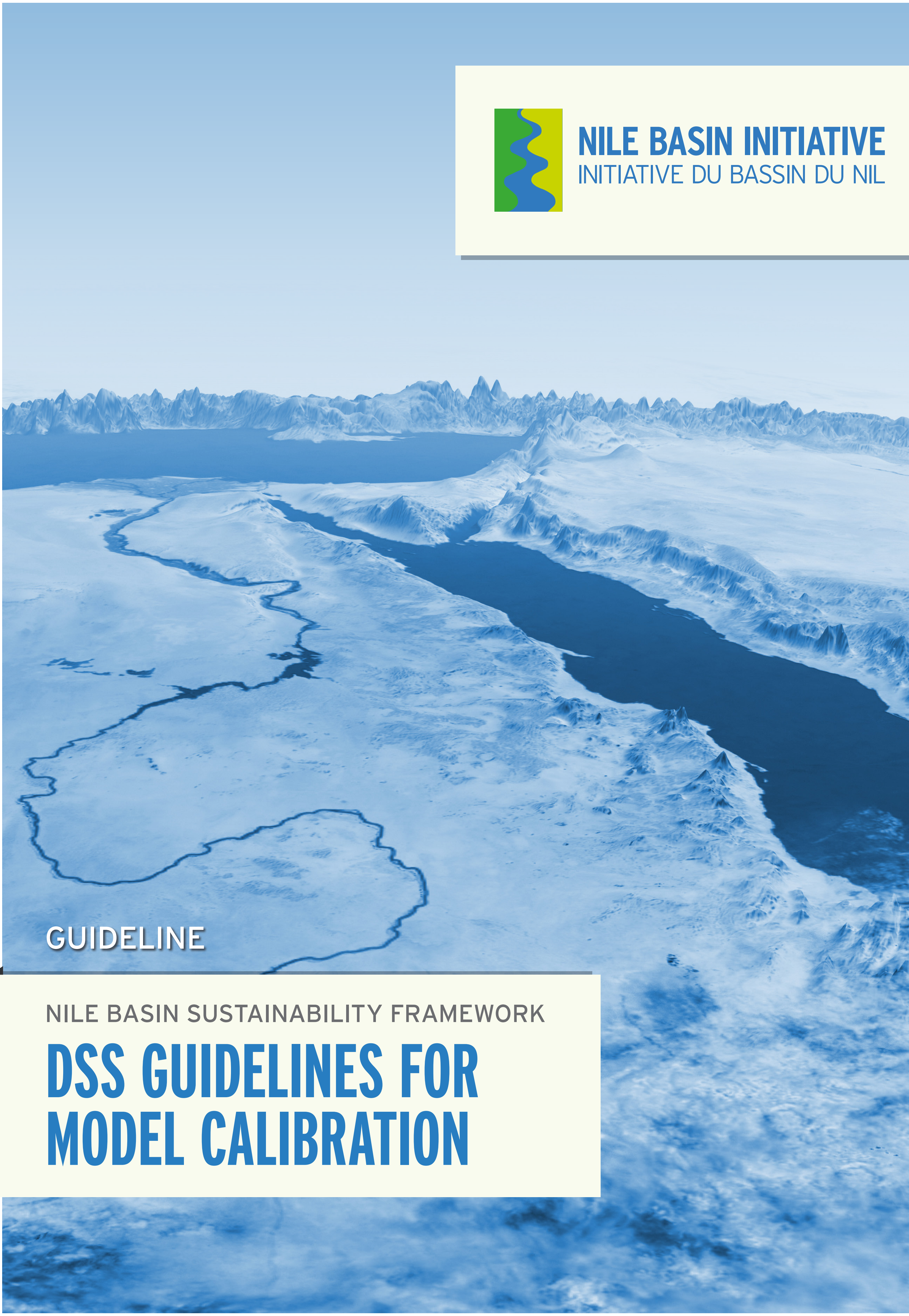


NILE BASIN INITIATIVE
INITIATIVE DU BASSIN DU NIL

GUIDELINE

NILE BASIN SUSTAINABILITY FRAMEWORK

DSS GUIDELINES FOR MODEL CALIBRATION

Photo: Shutterstock

Document Control Sheet

Title	DSS Guidelines for Model Calibration
Document type	☐ Policy ☐ Strategy ☒ Guidelines ☐ Legal and Foundational Document
Prepared by	☐ Nile-SEC ☐ ENTRO ☐ NELSAP-CU ☒ Other: WRPM Project
Status	☒ New Policy/Strategy/Guideline/Legal and Foundational Document ☐ Revision of existing Policy/Strategy/Guideline/Legal and Foundational Document
Revision Date	
Effective Date	2010

Consideration by Nile-COM/~~EN-COM/NEL-COM~~ (cross out whichever body is not applicable)	
Date of submission for consideration	
Action by Council of Ministers	
Comments satisfactorily addressed	☐ Yes ☐ No ☐ Not Applicable

Consideration by Nile-TAC~~/ENSAPT/NEL-TAC~~ (cross out whichever body is not applicable)	
Date of submission for consideration	
Action by the Technical Advisory Committee:	
Comments satisfactorily addressed	☐ Yes ☐ No ☐ Not Applicable

Responsible Officer: Dr. Abdulkarim H. Seid

Data Compilation and Pilot Application of the Nile Basin Decision Support System (NB-DSS): Work Package 2: Stage 2

MODEL CALIBRATION AND VALIDATION GUIDELINE

Leading. Vibrant. Global.
www.aurecongroup.com

MODEL CALIBRATION AND VALIDATION GUIDELINE

Final

Contact person:
V Jonker
Aurecon Centre
1 Century City Drive
Waterford Precinct
Century City, Cape Town, RSA
+27 21 526 9400
Verno.Jonker@aurecongroup.com

Submitted to:
Nile Basin Initiative
Water Resource Planning and Management Project
Dessie Road
Addis Ababa
Ethiopia

In association with:

SOLARIS Engineering & Consulting

BEUSTER, CLARKE & ASSOCIATES
GIS Application Development, Software Development, Database Development

Tony Barbour Consulting

Nepid Consultants

Conningarth Economists

Climate Systems Analysis Group (CSAG), University of Cape Town

PROJECT NAME	**:**	Data Compilation and Pilot Application of the Nile Basin Decision Support System (NB-DSS): Work Package 2: Stage 2
REPORT TITLE	**:**	Model Calibration and Validation Guideline
AUTHORS	**:**	Verno Jonker, Anton Sparks, Louise Dobinson
REPORT STATUS	**:**	Final
AURECON REPORT NO.	**:**	7331/107486
DATE	**:**	Dec 2012

Submitted by:

....

V JONKER (Date)

Study Leader

....

M KILLICK (Date)

Project Director

Approved for the NBI WRPMP

....

HA GHANY (Date)

Regional Manager

....

AH SEID (Date)

DSS Lead Specialist

Model Calibration and Validation Guideline

Project	Data Compilation and Pilot Application of the Nile Basin Decision Support System (NB-DSS): Work Package 2: Stage 2
Main Authors	Verno Jonker, Anton Sparks, Louise Dobinson
Status	Final
Date	Dec 2012

Reason for Circulation
Final

Circulation List
NBI DSS Core Team NBI Country representatives

Nature of comments required	
N.A.	
Date of Receipt of Comments	-

Dr V. Jonker: Study Leader
Aurecon Centre, 1 Century City Drive,
Waterford Precinct, Century City, Cape Town
Republic of South Africa
Tel: +27 21 526-9400
Fax: +27 21 526-9500
Verno.Jonker@aurecongroup.com

Table of Contents

Appendices

List of Figures

List of Tables

1. INTRODUCTION

1.1 BACKGROUND

For any model to be credible and gain acceptance among its users and stakeholder groups, it is imperative to have a well-defined and defensible model calibration and validation process as well as statistics or metrics that illustrates the model's ability to simulate changes in hydrologic conditions and water management options accurately. Hydrologic models are simplified mathematical representations of complex physical processes. As such, they cannot fully represent the numerous physical processes operating at various temporal and spatial scales across the domain of interest. However, when properly configured, models can simulate the known interrelationships between hydrologic variables and processes and can be useful in examining the response of hydrologic systems to perturbations in these variables.

1.2 PURPOSE

The purpose of these Guidelines is to provide general guidance on the calibration and validation of models which would typically be used for scenario analysis in the Nile Basin Decision Support System (NB-DSS). These include rainfall-runoff models, system or water balance models and hydrodynamic models.

1.3 TARGET AUDIENCE

These Guidelines are primarily aimed at modellers who will be tasked with configuring models to be used in the NB-DSS for the evaluation of alternative water resource development interventions and/or management options. However, it also provides water resource managers and other high level decision makers with an improved understanding of the challenges and limitations associated with the water resource models which drive the multi criteria decision analysis process in the NB-DSS.

2. KEY CONCEPTS

This Section introduces some of the key concepts associated with model calibration and validation.

2.1 MODEL SELECTION

One of the most important elements in the modelling process is the selection of the most relevant and appropriate model, which should be based on a clear understanding of the purpose of the modelling and the required modelling outputs. Dingman (2002) lists various considerations when choosing a model. These include the type, accuracy and precision of required model outputs, the spatial and temporal requirements for model configuration, and the availability of relevant and accurate input and calibration data.

2.2 MODEL CALIBRATION AND VALIDATION

Model calibration and validation are important steps in any water resource model application. In essence, calibration is an iterative procedure of parameter evaluation and refinement, as a result of comparing simulated and observed values of interest e.g. flows. Model validation aims to ensure that the calibrated model properly assesses all the variables and conditions which can affect model results, and demonstrates the model's ability to predict field observations for periods that are separate and independent from the calibration period.

2.3 MODEL PERFORMANCE

Model performance, i.e. the ability of the calibrated and validated model to accurately represent the physical system and/or processes being modelled, is measured by the degree in which the model complies with predefined model acceptance criteria. This is evaluated through a combination of quantitative and qualitative criteria, involving both graphical comparisons and statistical tests. It is important that the same model acceptance tests are employed during both the calibration and validation phases. The results of the calibration and validation process provide a measure of how accurately the model can be assumed to predict system responses to perturbed inputs. Basic measures of uncertainty associated with the validation process and an understanding of the model's limitations are keys to establishing model credibility and buy-in from stakeholders. Specific target values for quantitative measures are difficult to define precisely, as these depend on data availability and quality.

2.4 QUALITY ASSURANCE

Within the context of model calibration and validation, quality assurance refers to the identification of performance criteria that are appropriate for the intended model use (qualitative vs. quantitative), the specification of calibration methodologies (manual vs. automatic, etc.), the selection of relevant and quality controlled input data (time periods, temporal resolution, etc.), an audit trail of metadata describing all data and model development processes and the maintenance of detailed calibration logs.

The criteria for quality assurance acceptance include:

- Final quality assured data sets used in model calibration and validation; properly referenced in metadata files.

- Limited infilled data in model calibration and validation data sets; properly referenced in metadata files.
- Approach to model calibration and validation follows Guidelines; properly referenced in metadata files.
- Final calibrated and validated model properly referenced in metadata files.
- Results of model calibration and validation properly documented in study reports.
- Existing water resources facilities and operating rules correctly represented in models (spot checks)
- Proposed (scenario) water resources facilities and operating rules correctly represented in model

2.5 UNCERTAINTY

In any modelling process, it is important to understand, and if possible identify and quantify, the types of uncertainty likely to exist within the model. Sources of uncertainty in models could include:

- Errors in observed data
- Errors in model inputs (may be observed, or may be derived from other sources)
- Errors in model parameters
- Errors in model structure / physical process algorithms

In practice, it is very difficult to distinguish between some of the above sources of uncertainty. Model calibration will only impact model parameter errors, and only to the extent that the data against which the model is being calibrated is accurate. Errors in observed data include errors introduced due to infilling, which is a particular problem in data-scarce regions. Uncertainty analysis in its simplest form makes assumptions about the nature of the sources of uncertainty which are typically systematic errors, random errors, model structure errors or model parameter errors. A limiting factor to model performance is uncertainty in the input data set and it is therefore important to check the hydrological consistency of the data set. There are several types of uncertainty in models which relate the complexity of the following variables:

- Material properties of the hydrological response system including rainfall, evaporation, soils and others that can be easily measured at local scale, but may be difficult to quantify at a large scale
- Coupled processes which are dynamic and operating concurrently in time and space
- Model structure uncertainty, which can be an incomplete representation of processes

3. MODEL SELECTION

An important consideration before any model is configured, calibrated and validated is to ensure that the most appropriate model is selected. This decision should be guided by a clear understanding of:

- the purpose of the modelling
- the required modelling outputs
- spatial and temporal requirements for model configuration
- the availability of relevant and accurate input data which are required to 'build' the model
- the availability and quality of observed data to be employed in model calibration and validation

Models which would typically be used for scenario analysis in the NB-DSS include rainfall-runoff models, system / water balance models and hydrodynamic models.

3.1 RAINFALL-RUNOFF MODELS

Hydrological models are simplified, conceptual representations of a part of the hydrological cycle representing physical catchment processes. They are primarily used for hydrological prediction and for understanding hydrological processes. A rainfall-runoff model is used to generate synthetic flow sequences when a sufficiently long observed streamflow record is not available, and when there is suitable rainfall station data to undertake such an analysis. In many countries, there are generally more rainfall records than streamflow records, while rainfall records also tend to have longer periods of observed data.

Rainfall-runoff models are calibrated in catchments that are situated upstream of streamflow gauges with a sufficiently long historical record. Calibration parameters are adjusted until the goodness-of-fit of the simulated and observed flows complies with predefined model acceptance criteria.

The NAM model is an example of a rainfall-runoff model and was used to model streamflow in the Blue Nile and Tekeze catchments in this consultancy.

3.2 WATER BALANCE MODELS

A water balance model is used to represent an integrated system, usually at a larger scale of analysis than rainfall-runoff catchment models. A water balance can be useful to help manage water supply and predict where there may be water shortages. It uses flow inputs and represents physical features of the system including bulk infrastructure and associated operating rules. The model supports the water resource planning process by providing a tool to balance the available water resources in a system with the water requirements and losses to which the system is subjected.

Similar to rainfall-runoff models, water balance or system models are also calibrated against observed flows at key locations within the basin being modelled. This typically entails a comparison of flow characteristics at key flow gauges as well as an assessment of actual vs simulated water balances at major dam sites or key river nodes. In addition, it is useful if qualitative data e.g. historical areal extents of wetlands and lakes or historical dam storage trajectories are readily available as this would greatly assist with model calibration and validation.

MIKE Basin is an example of a system model and was used to model the water balance in the Nile Basin in this Study. The Mike Basin model can be used to simulate water allocation, conjunctive water use, reservoir operation or water quality.

3.3 HYDRODYNAMIC MODELS

Most hydrodynamic (HD) models use an implicit, finite difference scheme for the computation of unsteady flows in rivers and estuaries. The module can describe sub-critical as well as super critical flow conditions through a numerical scheme which adapts according to the local flow conditions (in time and space). Advanced computational modules are included for description of flow over hydraulic structures, including possibilities to describe structure operation.

HD models are calibrated against observed stage and discharge records at flow gauges and generally involve the adjustment of calibration parameters to improve the goodness of fit between simulated and observed flow and stage hydrographs in terms of the shape and timing of the hydrographs as well as peak values. In the case of extreme events, the areal extent of inundated floodplain areas as well as peak flood levels which have been marked on buildings or bridges, can also be very useful for model calibration.

MIKE 11 is an example of an HD model and was used to model flood propagation and inundation along the lower Blue Nile River in this Study.

4. APPROACH TO MODEL CALIBRATION AND VALIDATION

4.1 ALTERNATIVE APPROACHES

There are a multitude of approaches to model calibration and validation, and the selection of a particular methodology should be based on several factors, including the intended use of the model, the variable(s) of interest in the physical domain that are being simulated by the model, and the availability of data.

Some of the commonly used approaches to calibration and validation include split-sample calibration-validation, blind testing validation, optimization techniques for parameter identification, qualitative/subjective analysis and parameter uncertainty/perturbation analysis (Klemes, 1986; Donigian and Rao, 1990).

Split-sample techniques are perhaps the most commonly employed approach for river basin-scale model domains such as the Nile. The split-sample calibration / validation approach involves using a portion of the available data records as the basis for model calibration, and then performing a model simulation using a separate period of the available data, and evaluating the performance of the model against observed values from that period (model validation).

4.2 MANUAL VS AUTO CALIBRATION

Calibration procedures typically fall into one of two categories: manual (trial and error) or automated (parameter optimization, recursive algorithms, etc.).

4.2.1 Manual calibration

Manual calibration techniques are widely used, but they do require a degree of subjectivity and for complex systems can be time-consuming. The basic approach for manual calibration and validation of models is as follows (adopted from Abbott and Refsgaard, 1996; U.S. EPA, 2002):

1. Identify objective of calibration: variables of interest and acceptance criteria, including measures of fit.
2. Identify historical data suitable for calibration and validation periods. Criteria for selection of the periods include quality and completeness of the data, and may also relate to representativeness of the data (e.g. to include or exclude certain hydrological conditions, depending on the model purpose).
3. Identify calibration parameters, and the allowable range of values for model parameters. Identify appropriate observed field data or other references to support the analysis.
4. Simulate the calibration period and compare model outputs against observed data using the pre-defined acceptance criteria.
5. Adjust parameters and repeat simulation until acceptable values of the acceptance criteria are achieved.
6. Keep detailed calibration log of parameter values and calibration analysis results.
7. When the model calibration is acceptable, run a simulation using unchanged parameter values for the validation period, and evaluate the results using the same acceptance criteria.

4.2.2 Auto calibration

Automated approaches can quickly generate optimal solutions to parameter estimation, but users need to be careful to avoid physically unrealistic model parameterisations and the possibility that the optimal calibration result may not adequately recreate key hydro-meteorological events that are relevant to the overall model objective. Generally, auto-calibration should be used as an initial step towards model calibration, after which the parameters could be refined and "improved" manually.

Automated model calibration refers to generic tools that have been incorporated into many model applications (including the MIKE models), and which allow automatic calibration of parameters based on predefined criteria. Before performing an automatic calibration, it is necessary to specify initial values as well as upper and lower bounds for parameters to be adjusted. Furthermore, it is also possible to specify parameters as constant values or as functions of other parameters, i.e. a distinction is made between dependent and independent parameters.

5. MODEL ACCEPTANCE CRITERIA

Model acceptance criteria refer to the criteria that are used when evaluating the goodness-of-fit of model simulations against observed data. Generally, these can be categorised into visual or graphical assessments and quantitative assessments.

5.1 VISUAL (GRAPHICAL) CRITERIA

When comparing simulated model output against observed data at key locations for the acceptance of model calibration, the following visual assessments are typically used.

- Good agreement between monthly and annual timeseries plots of simulated and observed catchment runoff
    - A good overall agreement of the shape of the hydrograph
    - A good agreement of the peak flows with respect to timing and volume
    - A good agreement for low flows.
- Seasonal flow distribution
- Standardised residuals
- Correlation between simulated and observed monthly or annual flow volumes
- Cumulative flow plots, Unit runoff plots and Storage-yield plots

5.1.1 Flow time series plots

Figure 5-1 shows a daily time series plot of flows from the NAM model at the Didessa flow gauge in the Blue Nile. Overall there is a good agreement of flows according to the criteria mentioned above. The daily or monthly time series plot is useful for detecting outliers (very large differences between observed and simulated flows) and, particularly in rivers with a strong base flow, for checking how well the dry-season recession is simulated. Specific aspects to consider include good overall agreement of the shape of the hydrographs, good agreement of the peak flows with respect to timing and volume and good agreement for low flows. A plot of the annual hydrograph is useful for assessing whether the simulated flows exhibit a similar pattern to the observed flows, to assess the range of simulated flows, flow variability and the sequence of wet and dry years. This plot is also useful for detecting outliers.

5.1.2 Plots of seasonal flow distribution

Plots of seasonal flow distribution provide the modeller with an indication of how well the model is simulating the seasonality of flows and can also show how well the wet season and dry season flows are being simulated. It will reveal consistent over- or underestimation of flows in any calendar month or sequence of calendar months. Typical problems include base (dry season) flows that are too low, simulated flows that are too low in the early wet season and too high in the late wet season and wet season flows that are too high or too low. Figure 5-2 below shows the seasonal flow distribution from the calibration of flows on the Didessa tributary in the Blue Nile using NAM. The high flows and low flows show good overall agreement, but the peak starts rising about a month after the observed flows. This indicates that the calibration parameters should be changed so as to decrease the lag in the catchment so that the flows in May, June and July are not so much lower than the observed.

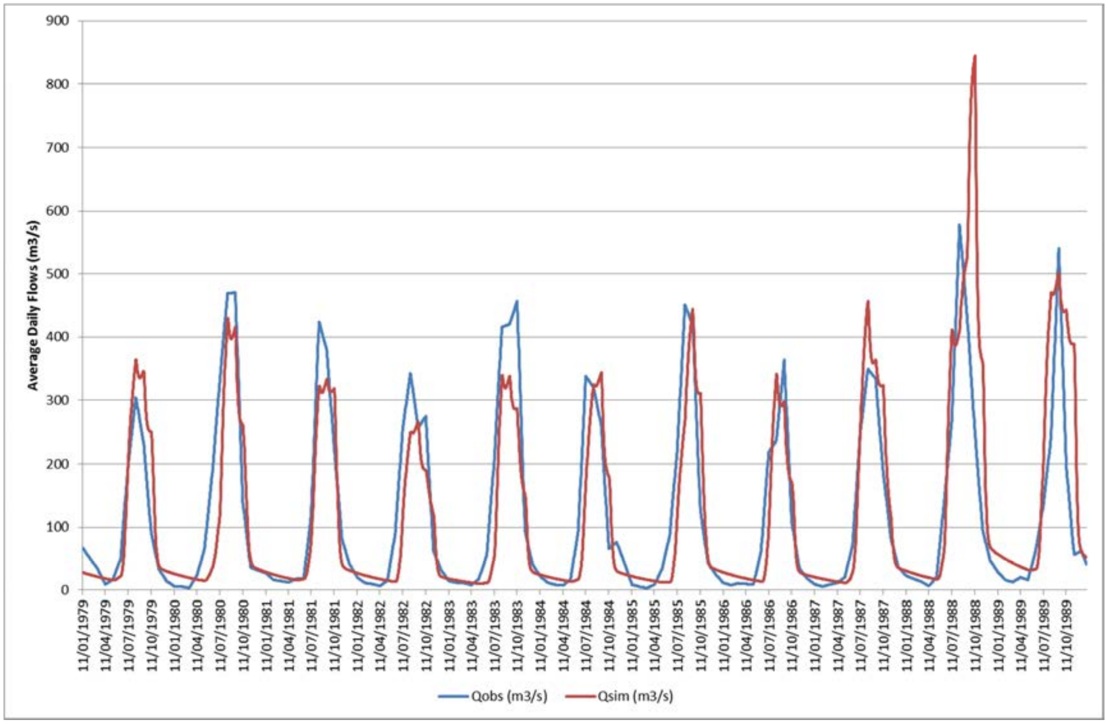


Figure 5-1 : Example of daily flow time series plot

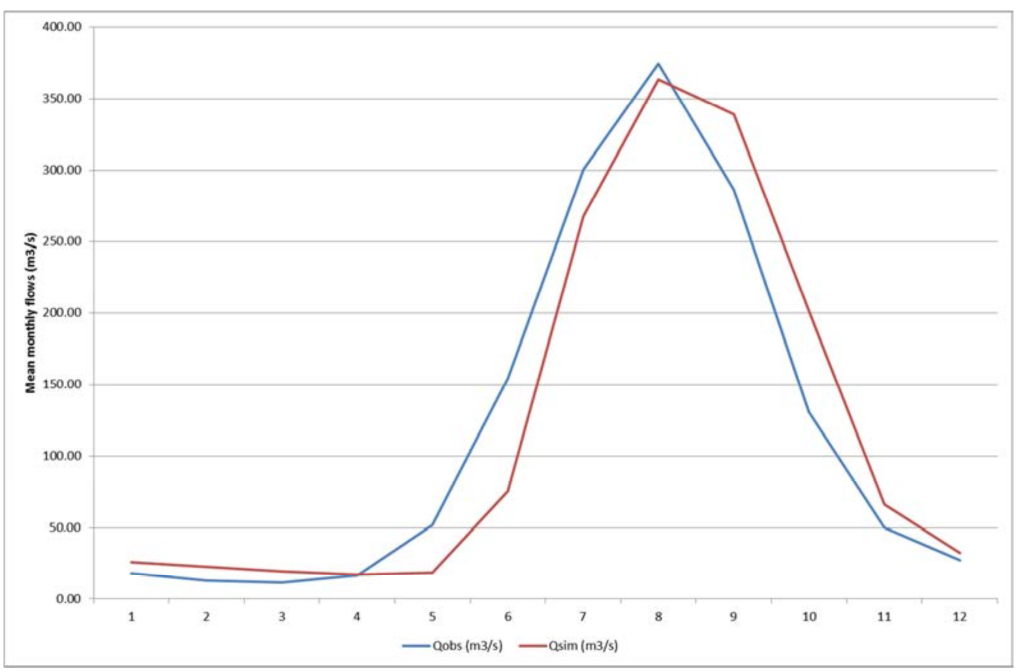


Figure 5-2 : Example of seasonal flow distribution

5.1.3 Standardised residuals (Flow)

The standardised flow residuals are calculated as:

(Simulated flow – Observed flow) / Observed flow

An even distribution around zero is desirable. In Figure 5-3 below, there is a tendency towards the positive which indicates that the simulated flows tend to be over-estimated.

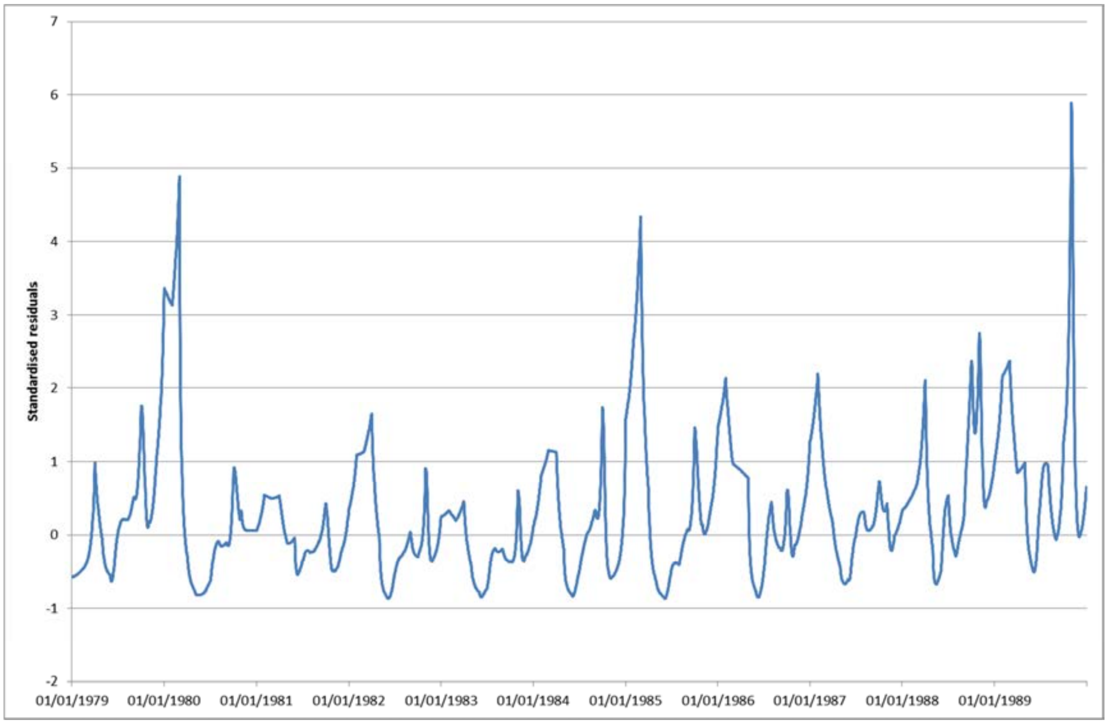


Figure 5-3 : Standardised flow residuals (monthly flows)

5.1.4 Scatter Plots

The scatter plot provides an indication of the correlation (or lack of correlation) between observed and simulated values and provides a measure of variability and of how well future outcomes are likely to be predicted by the model. Figure 5-4 below shows a scatter plot of simulated vs. observed monthly flows and also indicates the coefficient of determination (r^2 value). An r^2 value of 1 indicates that the regression line perfectly fits the data. A value of 0.8 as obtained in the example below is an acceptable r^2 value.

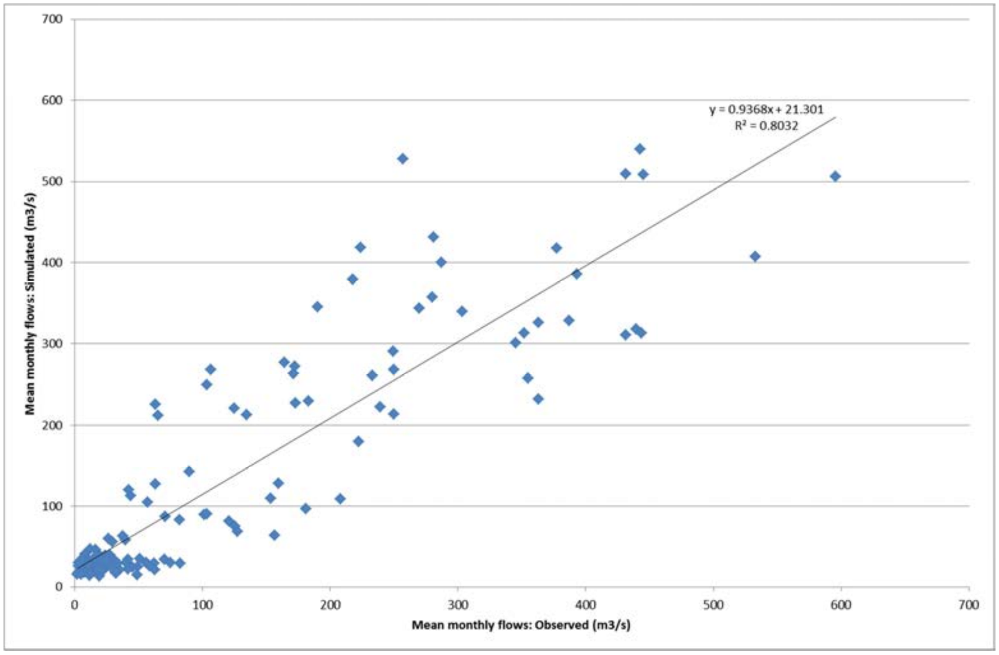


Figure 5-4 : Scatter plot of observed and simulated monthly flows

5.1.5 Cumulative flow plots

Cumulative flow plots are helpful to identify systematic deviations and discontinuities in the observed flow record or simulated flow sequence. Figure 5-5 shows the observed and simulated accumulated flows for the Didessa tributary of the Blue Nile RIver. For the initial part of the calibration period it shows a good agreement up to 1982, thereafter the simulated flows are lower than the observed.

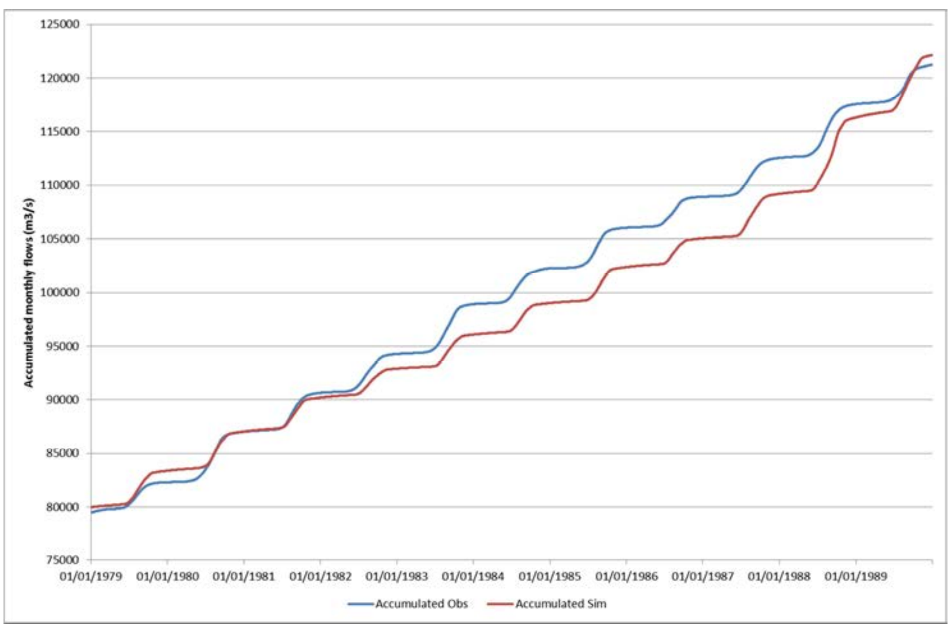


Figure 5-5 : Example of accumulated flows plot

5.1.6 Unit runoff plots

Another very useful plot to improve the understanding of the rainfall-runoff characteristics of a catchment and to identify uncertainties is a plot of Mean Annual Precipitation (MAP) to Unit Runoff (specific runoff). Unit runoff is determined by dividing the Mean Annual Runoff (MAR) of a catchment by the surface area and expressing the results in mm. When investigating unit runoff relationships in a catchment, it is important to take cognisance of the fact that runoff linked to any specific rainfall event is highly dependent on the antecedent soil moisture conditions in the catchment as well as on the spatial and temporal rainfall distribution of the particular rainfall event. For this reason, long term average values of rainfall and runoff e.g. MAP and MAR should be used when calculating unit runoff.

Figure 5-6 below shows a MAP- unit runoff plot prepared from data available in the Baro-Akobo Basin. Note the wide spread of unit runoff for an MAP of 1500mm, from 200 to about 700mm. If this spread is due to an error it will impact significantly on the viability of proposed schemes. These have not been investigated in depth, but some have been generated by scaling others and are not independent. In some instances there may be a good reason for a low reading. For example, large endoreic areas (marsh areas that do not drain from the catchment) may have low runoff, while a rocky mountainous area may have a much higher unit runoff. Catchments whose data allow good calibration will help to identify which points are more likely.

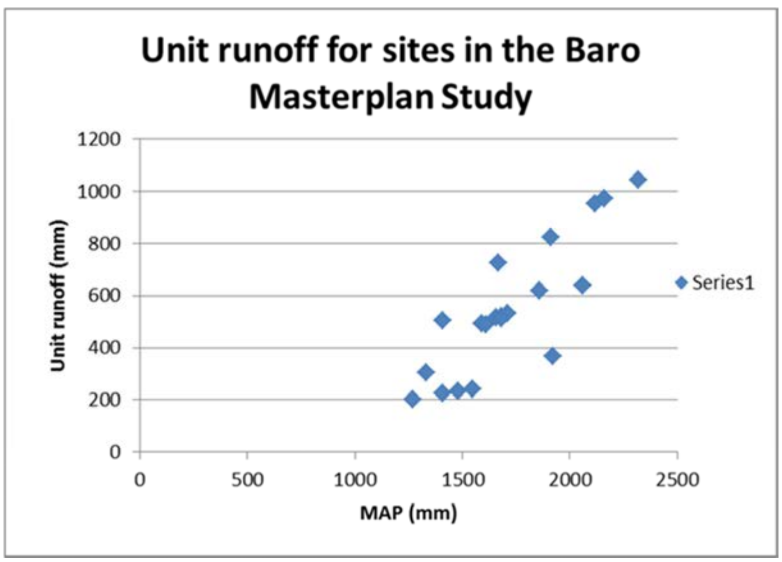


Figure 5-6 : MAP to unit runoff (mm)

5.1.7 Storage-yield plots

When performing rainfall-runoff modelling, an additional valuable check is to compare the storage-yield relationships of the observed and simulated flow sequences, which provide an indication of the similarity in terms of extended low flow periods. This plot is developed by evaluating the relationship between storage and gross firm yield for both the simulated and observed flow sequence. This is particularly important if the sequence is to be used for water yield planning purposes. An example of a storage yield plot is shown in Figure 5-7. Storage-draft analyses are usually done on a monthly time step.

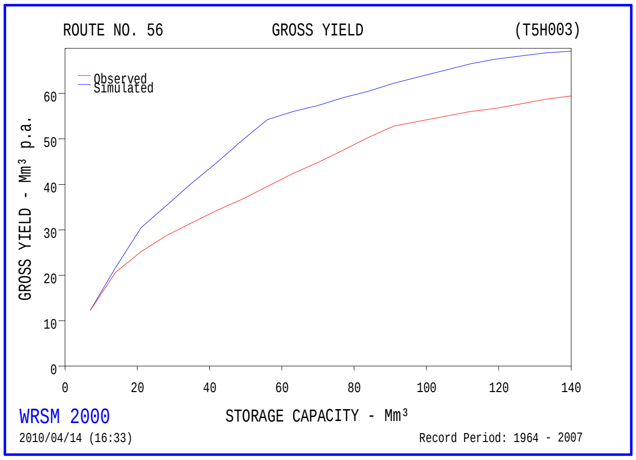


Figure 5-7 : Storage-Yield Plot

5.2 QUANTITATIVE CRITERIA

In addition to visual assessments, quantitative criteria can also be used to determine the acceptance of model calibration and validation. Quantitative measures may include simple mass (water) balance analysis, which represents the long-term difference between simulated and observed total flow over the simulation period, goodness-of-fit measures, a comparison of specified flow exceedence values, residual analysis, and other sensitivity analyses. These typically include statistically based indices such as the mean or median annual flow, standard deviation of flows, the seasonality index and the coefficient of determination (refer to Nash and Sutcliffe, 1970; Donigian and Rao, 1990; Abbott and Refsgaard, 1996; U.S. EPA, 2002).

5.2.1 Mean / Median Annual Runoff (Flow)

Due to its computational simplicity, Mean Annual Runoff (MAR) is a commonly used measure of central tendency. However, the median is sometimes to be preferred, especially for extremely skewed distributions.

5.2.2 Standard deviation

Standard deviation is a basic measure of variability.

5.2.3 Seasonality Index

The index of seasonal variability indicates by means of a simple coefficient the extent of month-by-month fluctuation. The seasonal index (SI) is obtained as the range of the monthly cumulative net deviations of flows as described by the equations below and illustrated in Figure 5-8.

$$SI = \text{maximum}\left(CumNetDev(j)\, j = 1,2,..12\right) \quad - \quad \text{minimum}\left(CumNetDev(j)\, j = 1,2,..12\right)$$

$$CumNetDev(j) = \sum_{k=1}^{J} NetDev(k) \qquad j = 1,2,....12$$

$$NetDev(k) = 100\left[\frac{\sum_{i=1}^{n} q_{i,k}}{\sum_{k=1}^{12}\sum_{i=1}^{n} q_{i,k}} - \frac{1}{12}\right] \qquad k = 1,2,....12$$

where $q_{i,k}$ is the flow in month k of year i.

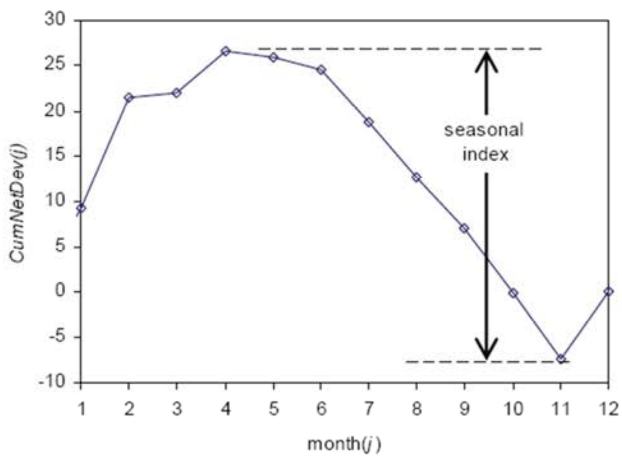


Figure 5-8 : Seasonality Index

5.2.4 Efficiency criteria

Efficiency criteria are mathematical measures of how well a model simulation fits the available observations. Various efficiency criteria exist. Two of the most common ones are discussed below.

Coefficient of determination (r^2)

The coefficient of determination (r^2) describes the degree of co-linearity between simulated and measured data and provides a measure of the proportion of the variance in measured data explained by the model. The coefficient of determination is defined as the squared value of the Pearson coefficient of correlation. It is calculated as:

$$r^2 = \left(\frac{\sum_{i=1}^{n} (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n} (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^{n} (P_i - \bar{P})^2}} \right)^2$$

with O observed and P predicted values.

r^2 values range from 0 to 1, with higher values indicating less error variance. Typically, values greater than 0.5 are considered acceptable (Santhi et al., 2001, Van Liew et al., 2003). Although r^2 is widely used for hydrological model evaluation, it is oversensitive to high extreme values (outliers) and insensitive to additive and proportional differences between model predictions and measured data (Legates and McCabe, 1999). As a result, attempts aimed at minimizing r^2 often lead to fitting the higher portions of the hydrograph (e.g. peak flows) at the expense of the lower portions (e.g. baseflow). Care should be exercised when using r^2 and it is preferable to use it in combination with other quantitative criteria.

Nash-Sutcliffe Coefficient of Efficiency

The Nash-Sutcliffe model efficiency coefficient (Nash and Sutcliffe, 1970) is defined as:

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

where Qo is observed discharge, and Qm is modeled discharge. Qot is observed discharge at time t.

Nash–Sutcliffe efficiencies can range from −∞ 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 largest disadvantage of the Nash-Sutcliffe efficiency is the fact that the differences between the observed and predicted values are calculated as squared values. As a result larger values in a time series are strongly overestimated whereas lower values are neglected (Legates and McCabe, 1999). For the quantification of runoff predictions this leads to an overestimation of the model performance during peak flows and an underestimation during low flow conditions. Similar to r^2, the Nash-Sutcliffe is not very sensitive to systematic model over- or underpredictions, especially during low flow periods.

5.2.5 Other quantitative criteria

Other quantitative acceptance criteria to further check model simulations include comparison of the modelled outputs with observed data at specific locations where relevant historical information is available. However, these checks are usually subject to high levels of uncertainty and very sensitive to dam operating rules for example. The following time series of model outputs could typically be compared to observed or recorded historical data:

- The variation of the areal extent of wetlands or lakes
- Fluctuations in dam levels (dam storage trajectories)

- Average generated hydropower at specific hydropower installations

5.3 TARGET VALUES

Specific target values for quantitative measures are difficult to define precisely, as these depend on data availability and quality. Typically, in terms of water balance, it should be aimed to achieve an error of less than 5% when comparing long-term mean annual flows at key nodes. When assessing the degree of correlation between monthly observed and simulated flows, an r^2 value larger than 0.8 indicates a good fit, while standardised residuals should preferably be confined to between -0.2 and +0.2. Indices such as standard deviation are generally very sensitive to outliers and this should be borne in mind when evaluating model performance.

6. PROCEDURAL GUIDANCE FOR MODEL CALIBRATION AND VALIDATION

This Section provides a step-by-step guide to model calibration and validation using the generic steps detailed in Section 4 as guidance. A distinction is made between rainfall-runoff models, system / water balance models and hydrodynamic models, which represent the three main types of models which would typically be used for scenario analysis in the Nile Basin Decision Support System (NB-DSS).

Appendix A provides an example of a NAM / MIKE Basin calibration and validation using one of the Nile Basin pilot areas as a case study.

6.1 RAINFALL-RUNOFF MODEL (NAM)

The NAM rainfall-runoff model is a deterministic, lumped model. It is based on physical processes (conceptual) and is continuous in that it accounts for moisture in four different and mutually inter-related storage zones. It requires limited data inputs namely rainfall and evaporation (and temperature for snow conditions). It simulates runoff (flow), groundwater levels, temporal variation in moisture content, infiltration, recharge, interflow, overland flow and base flow.

Step 1: Identify calibration objective(s)

The main purpose for the calibration of a rainfall-runoff model is the development of a physically- based model which is capable of simulating long-term flow sequences based on long-term observed rainfall. As such, the model can be used to extend observed flow records in gauged catchments, while calibration parameters can also be transferred to ungauged catchments in order to generate synthetic flow sequences. The following objectives are usually considered in model calibration:

- A good agreement between the average simulated and observed catchment runoff (i.e. a good water balance)
- A good overall agreement of the shape of the hydrographs
- A good agreement of the peak flows with respect to timing, rate and volume
- A good agreement for low flows

It is important to note that, in general, trade-offs are necessary between the different objectives. If the objectives are of equal importance, one should seek to balance all the objectives, whereas in the case of prioritised objectives, these objectives should be favoured.

Step 2: Identify calibration catchments

Rainfall-runoff models are calibrated in catchments that are situated upstream of streamflow gauges. The identification and selection of calibration catchments are therefore primarily dictated by the availability of good quality flow records at flow gauging stations within the study basin, which cover a sufficiently long period. These records should be continuous and should preferably not have more than 10% missing values, which should then be infilled. Furthermore, depending on the catchment size and rainfall variability within the calibration catchment, it should be ensured that there are a sufficient number of rainfall stations within the vicinity of each calibration catchment, providing sufficient spatial coverage, with historical data which overlap the observed flow record period. Another consideration includes the temporal resolution of the flow and rainfall data, e.g. monthly or daily data, depending on the modelling time step which will be used. Finally, especially in the case of larger river basins, it is important to ensure that the selected calibration catchments are spatially representative of the hydro-meteorological, topographical and physiographical variability within the basin.

Step 3: Select calibration / validation periods

Split-sample techniques are the most commonly employed approach for river basin-scale model domains such as the Nile. The split-sample calibration/validation approach involves using a portion of the available data records as the basis for model calibration, and then performing a model simulation using a separate period of the available data, and evaluating the performance of the model against observed values from that period (model validation). Satisfactory calibration and validation over a full range of flows usually require continuous observations of flow for a period of at least 10 to 15 years, which can then be split into two separate continuous periods for model calibration and validation respectively.

Step 4: Identify calibration parameters

The NAM model parameters include:

- Surface and root zone parameters
    - Maximum water content in surface storage U_{max}
    - Maximum water content in root zone storage L_{max}
    - Overland flow runoff coefficient CQOF
    - Time constant for interflow CKIF
    - Time constant for routing interflow and overland flow CK1 and CK2
    - Root zone threshold value for overland flow TOF
    - Root zone threshold value for interflow TIF
- Groundwater parameters
    - Baseflow time constant CKBF
    - Root zone threshold value for groundwater recharge TG
    - Ratio of groundwater catchment to topographical catchment area Carea
    - Snow module parameters

There are a maximum of 15 NAM parameters: five of these are usually fixed, leaving ten for calibration. Three parameters in particular are most important for the water balance: L_{max}, U_{max}, CQOF. The remaining parameters are used for minor adjustments and for routing. Table 6-1 presents a summary of the parameters in NAM, while Figure 6-1 shows a schematic of the NAM model structure.

Table 6-1 : Summary of key NAM parameters

NAM Parameter	ID	Information	Effect if increased	Effect if decreased	Typical range
Maximum water content in surface storage	U_{max}	Overland flow/infiltration Evapotranspiration Interflow	Less overland flow (especially in start of wet periods) Higher evapotranspiration Reduced infiltration Higher interflow		U_{max} ~ 0.1 * Lmax Umax ~ 10-20 mm
Maximum water content in lower zone / root zone storage	L_{max}	Overland flow/infiltration Evapotranspiration Baseflow	Higher evapotranspiration Reduced overland flow Higher infiltration Reduced baseflow		
Overland flow coefficient (Value range: 0-1)	C_{QOF}	CQOF is a very important parameter, determining the extent to which excess rainfall runs off as overland flow and the magnitude of infiltration.	Steep, impermeable soils, rocks : C_{QOF} ~ 1: Large overland flow Small infiltration	Flat, highly permeable soils : C_{QOF} ~ 0: Little overland flow High infiltration	0-1
Interflow drainage coefficient	C_{KIF} :	CKIF determines together with Umax the amount of interflow 1 / (CKIF) is the quantity of the surface water content U that is drained to interflow per time step	Linear amplification of interflow Reduced infiltration Reduced overland flow		C_{KIF} = 500-1000 hours
	CK1 CK2 CKBF	For routing overland flow and interflow along catchment slopes and through channels down to outlet of catchment.	Longer duration of flow; Lower peaks		Usually CK1 = CK2 Usually CKBF >> CK1/CK2

NAM Parameter	ID	Information	Effect if increased	Effect if decreased	Typical range
		CKBF for routing recharge through linear g.w. storage.			
Overland flow	TOF		Later start of overland flow in beginning of wet season; Higher infiltration		0-1
Threshold values for interflow	TIF		Later start of interflow in beginning of wet season; Higher infiltration and overland flow		0-1
Groundwater recharge	TG		Later start of groundwater recharge and flow in beginning of wet season; Quicker filling of root zone		0-1

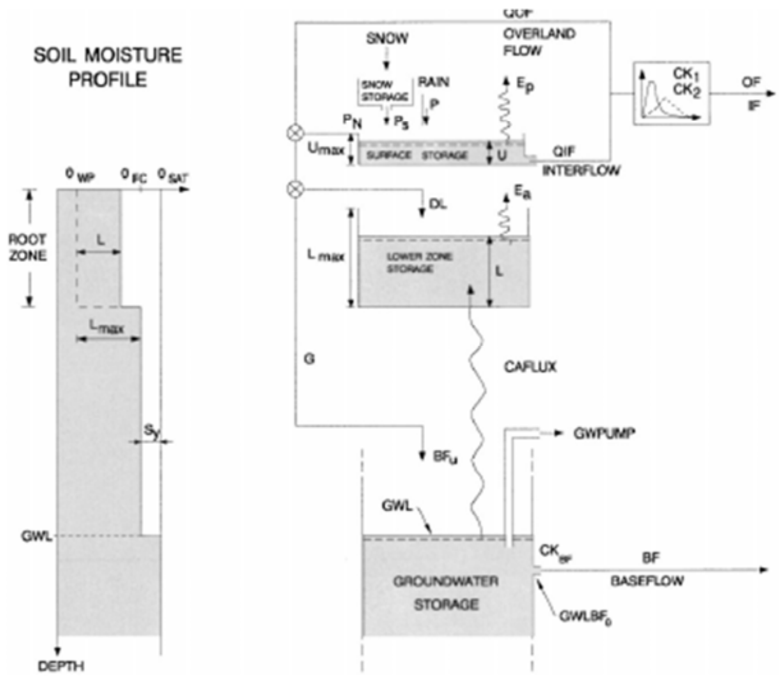


Figure 6-1 : Nam Model Structure (DHI, 2010)

Step 5: Specify model acceptance criteria / calibration targets

Both visual and quantitative criteria should be specified to evaluate model acceptance based on the calibration and validation results. Visual criteria refer to the use of graphs to gauge the goodness of fit of simulated vs observed flows. Typical plots include time series plots, seasonal flow distribution, standardised flow residuals, scatter plots to assess the degree of correlation, cumulative plots, storage-yield plots and unit runoff plots. In addition to visual assessments, quantitative criteria can also be used to determine the acceptance of model calibration and validation. These typically include statistically based indices such as the mean or median annual flow, standard deviation of flows, the seasonality index and the coefficient of determination.

Further details regarding model acceptance criteria are provided in Section 5 of this report.

Step 6: Conduct iterative simulations and evaluate model performance

Calibration parameters are adjusted until the goodness-of-fit of the simulated and observed flows complies with predefined model acceptance criteria.

A typical approach to carry out a successful NAM calibration involves:

- Sensitivity analysis to identify physical parameters characterizing the actual rainfall-runoff process.
- Fit of water balance of calibration period - adjust actual evapotranspiration by adjusting Lmax/Umax

- Fit of flood peaks – adjust overland flow:
    - Volume – CQOF
    - Timing – TOF
    - Shape – CK1 = CK2
- Fit of low flows - adjust baseflow:
    - Volume - adjust overland flow adjusted recharge i.e. again CQOF
    - Timing - TG
    - Shape - CKBF

Step 7: Complete calibration log

It is extremely important that a detailed log is maintained during model calibration. Not only does this facilitate the calibration process itself, but it is also extremely valuable for future re-calibrations and refinements of the model in light of updated data and information and or changes to the modelling routines. An example of a calibration log is included below.

Table 6-2 : Example of a Calibration Log

Calibration Run Id	Description	Effect of decreasing specific parameters	Parameter 1	Parameter 2	Parameter x	Model performance
1						
2						
3						

Step 8: Conduct model validation

Once satisfactory model calibration is achieved, it is imperative that model simulations should be performed using separate periods of the available data for evaluating the performance of the model against observed values from that period (model validation). The same set of model acceptance criteria should be used to assess model performance during validation and depending on the results, it might be necessary to revise the model calibration until both the calibration and validation results comply with the acceptance criteria.

6.2 SYSTEM / WATER BALANCE MODEL (MIKE BASIN)

The water balance and most of the operational management scenarios for the Nile Basin have been configured and simulated with the MIKE Basin software. MIKE Basin operates on the basis of a schematic river network, which is made up of building blocks or objects. These typically included river reaches, water user nodes, reservoir nodes, irrigation nodes, hydrological catchments, hydropower nodes and channels which link the various nodes. The characteristics of the individual building blocks and the interactions between them are governed by user-defined operating rules. The wide range of modelling routines embedded in the MIKE Basin model, allow the simulation of various multi-purpose water resource project scenarios and typically include reservoir and hydropower operation, irrigation demand modelling, hydrological routing, water allocation rules, and evaporation.

Step 1: Identify calibration objective(s)

The main purpose for the calibration of a system or network model is the development of a tool which is capable of simulating the long-term behaviour of a water resource system in terms of flows, operating rules, water losses, water use, water transfers, flow routing and hydropower generation. A key aspect linked to the configuration of a water balance model relates to the upstream boundary conditions or model inflows. These can be derived from a calibrated rainfall-runoff model, based on observed flows at gauging stations or scaled from observed flows in other parts of the basin using area-based or empirical rainfall-runoff relationships.

The following objectives are usually considered in model calibration:

- A good agreement between the average simulated and observed catchment runoff (i.e. a good water balance)
- A good overall agreement of the shape of the hydrographs
- A good agreement of the peak flows with respect to timing and volume
- A good agreement for low (base) flows
- A good agreement between dam storage trajectories
- A good agreement between the areal extent of wetland and/or lake inundation areas

It is important to note that, in general, trade-offs are necessary between the different objectives. If the objectives are of equal importance, one should seek to balance all the objectives, whereas in the case of priority to a certain objective, this objective should be favoured.

Step 2: Identify calibration locations

System models are typically calibrated at streamflow gauges with a sufficiently long historical record and at existing dams and/or lakes and wetlands where historical information on storage trajectories, areal extents etc. are available. The identification and selection of stream flow gauges within the study basin are primarily dictated by the availability of good quality flow data, which cover a sufficiently long period. These records should be continuous and should preferably not have more than 10% missing values, which should then be infilled. Another consideration includes the temporal resolution of the flow and rainfall data, e.g. monthly or daily data, depending on the modelling time step which will be used. Finally, it is important to ensure that at least one of the selected calibration locations are situated close to the outlet of the water resource system being modelled to ensure that the overall water balance can be accurately calibrated.

Step 3: Select calibration / validation periods

Split-sample techniques are the most commonly employed approach for river basin-scale model domains such as the Nile. The split-sample calibration/validation approach involves using a portion of the available data records as the basis for model calibration, and then performing a model simulation using a separate period of the available data, and evaluating the performance of the model against observed values from that period (model validation). Satisfactory calibration and validation over a full range of flows usually require continuous observations of flow for a period of at least 10 to 15 years, which can then be split into two separate continuous periods for model calibration and validation respectively.

Step 4: Identify calibration parameters

Calibration parameters for system models essentially involve refining parameters and inputs which are uncertain, while parameters which are based on accurate, quality controlled data should preferably not be adjusted. Potential calibration parameters include:

- model inflows at upstream boundaries (e.g. tributary inflows)
- dam operating rules (release rules, flood or sediment control rules, characteristics levels)
- losses (evaporation / infiltration losses)
- hydrological routing coefficients
- mean annual precipitation and evaporation

Step 5: Specify model acceptance criteria / calibration targets

Visual, quantitative and qualitative criteria should be specified to evaluate model acceptance based on the calibration and validation results. Visual criteria refer to the use of graphs to gauge the goodness of fit of simulated vs observed flows. Typical plots include time series plots, seasonal flow distribution, standardised flow residuals, scatter plots to assess the degree of correlation, cumulative plots and unit runoff plots. In addition to visual assessments, quantitative criteria can also be used to determine the acceptance of model calibration and validation. These typically include statistically based indices such as the mean or median annual flow, standard deviation of flows, the seasonality index and the coefficient of determination. Qualitative or semi-quantitative acceptance criteria to further check the model simulations would include comparison of the modelled outputs with observed data where this information is available e.g. variation of areal extent of lakes and wetlands, dam storage trajectories, generated hydropower etc.

Further details regarding model acceptance criteria are provided in Section 5 of this report.

Step 6: Conduct iterative simulations and evaluate model performance

Iterative simulations are carried out until the model performance complies with the model acceptance criteria as defined in Step 5.

Step 7: Complete the calibration log

It is extremely important that a detailed log is maintained during model calibration. Not only does this facilitate the calibration process itself, but it is also extremely valuable for future re-calibrations and refinements of the model in light of updated data and information and or changes to the modelling routines. An example of a calibration log is included in Table 6-2.

Step 8: Conduct model validation

Once satisfactory model calibration is achieved, it is imperative that model simulations should be performed using separate periods of the available data for evaluating the performance of the model against observed values from that period (model validation). The same set of model acceptance criteria should be used to assess model performance during validation and depending on the results, it might be necessary to revise the model calibration until both the calibration and validation results comply with the acceptance criteria.

6.3 HYDRODYNAMIC MODEL (MIKE11)

MIKE 11 is used for simulation of unsteady flow. It solves the one-dimensional St. Venant equations using an implicit finite difference scheme. Dynamic, diffusive and kinematic wave approximation is available as well as kinematic routing. The module includes a broad range of the most common hydraulic structures including weirs, culverts, bridges, pumps and tabulated structures. Hydrodynamic (HD) models are calibrated against observed stage and discharge records at flow gauges and generally involve the adjustment of calibration parameters to improve the goodness of fit between simulated and observed flow and stage hydrographs in terms of the shape and timing of the hydrographs as well as peak values. In the case of extreme events, the areal extent of inundated floodplain areas as well as peak flood levels which have been marked on buildings or bridges can also be very useful for model calibration.

Step 1: Identify calibration objective(s)

The main purpose for the calibration of a HD model is to develop a tool for simulating the propogation of flows along river reaches, while simultaneously predicting water levels at key locations within the river channel and floodplain based on stage-discharge relationships.

The following objectives are usually considered in model calibration:

- A good agreement between simulated and observed flow and stage hydrographs
- A good overall agreement of the shape of the hydrographs
- A good agreement of the peak flows with respect to timing, rate and volume
- A good agreement of stage (water levels)
- A good agreement of floodplain inundation in terms of timing, extent and duration

It is important to note that, in general, trade-offs are necessary between the different objectives. If the objectives are of equal importance, one should seek to balance all the objectives, whereas in the case of prioritised objectives, these objectives should be favoured.

Step 2: Identify calibration locations

HD models are typically calibrated at flow gauges with a sufficiently long historical stage and/or flow record and, in the case of high flows/extreme events, at locations along the floodplain where anecdotal information on historical flood levels or flood extents (e.g. from satellite imagery or photographs) exists. The identification and selection of gauges within the study basin are primarily dictated by the availability of good quality data, which cover a sufficiently long period. These records should be continuous and should preferably not have more than 10% missing values. Another consideration includes the temporal resolution of the flow data, e.g. monthly or daily data, depending on the modelling time step which will be used.

Step 3: Select calibration / validation periods

Split-sample techniques may be employed when calibrating HD models and entail the comparison of observed stage and flow data at flow gauging stations with HD model simulated values over different periods of observations. The split-sample calibration/validation approach involves using a portion of the available data records as the basis for model calibration, and then performing a model simulation using a separate period of the available data and, evaluating the performance of the model against observed values from that period (model validation). Satisfactory calibration and validation over a full range of flows usually require continuous observations of flow for a period of at least 10 to 15 years, which can then be split into two separate continuous periods for model calibration and validation respectively.

Step 4: Identify calibration parameters

Calibration parameters for HD models typically involve refining model parameters such as channel and floodplain roughness, routing parameters (e.g. kinematic routing parameters) and hydrodynamic parameters that are linked to the computational engine of MIKE 11. In the case of complex river and floodplain channels, it might also be necessary to refine the model parameters which govern the exchange of flow between the main river channel and the floodplain.

Step 5: Specify model acceptance criteria / calibration targets

Visual and qualitative criteria should be specified to evaluate model acceptance based on the calibration and validation results. Visual criteria refer to the use of graphs to gauge the goodness of fit of simulated vs observed flows and water levels (see Figure 6-2). These usually focus on time series plots (both flow and stage) of observed vs simulated results for particular flood events as well as for longer periods. In the case of extreme events, it is also useful to compare flood peaks and maximum water levels at key locations (e.g. in towns or at river crossings) where this data have been measured or estimated. Qualitative or semi-quantitative acceptance criteria to further check the model simulations would include comparison of the modelled outputs with observed data where this information is available e.g. variation of areal extent of inundation in floodplains, duration of floodplain inundation, initiation of river channel breaching etc. Further details regarding model acceptance criteria are provided in Section 5 of this report.

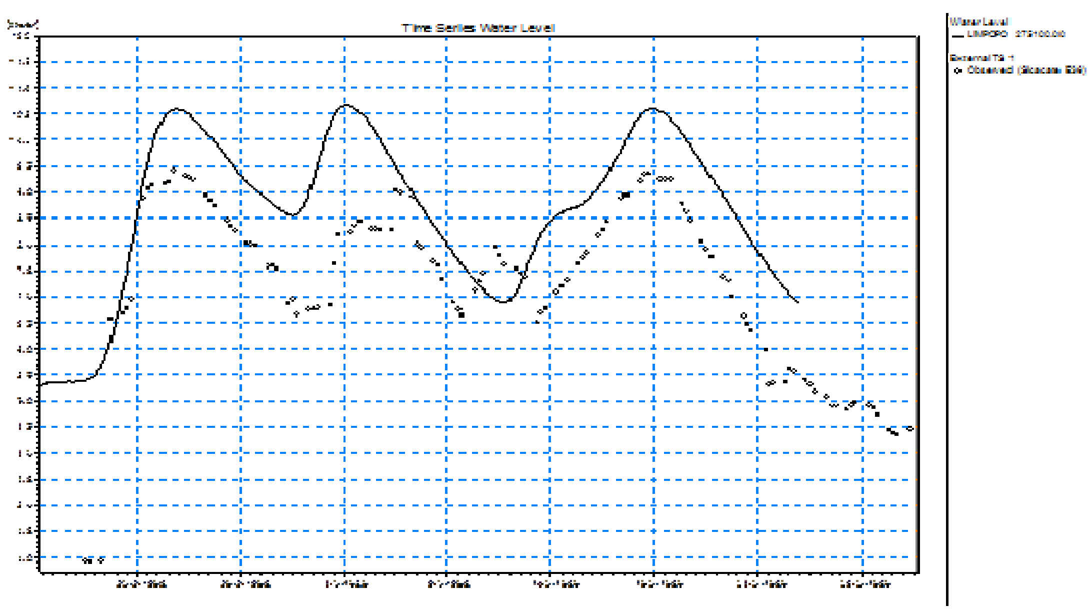


Figure 6-2 : Example of Simulated vs Observed Water Levels

Step 6: Conduct iterative simulations and evaluate model performance

Iterative simulations are carried out until the model performance complies with the model acceptance criteria as defined in Step 5.

Step 7: Complete the calibration log

It is extremely important that a detailed log is maintained during model calibration. Not only does this facilitate the calibration process itself, but it is also extremely valuable for future re-calibrations and refinements of the model in light of updated data and information and or changes to the modelling routines. An example of a calibration log is included in Table 6-2.

Step 8: Conduct model validation

Once satisfactory model calibration is achieved, it is imperative that model simulations should be performed using separate periods of the available data for evaluating the performance of the model against observed values from that period (model validation). The same set of model acceptance criteria should be used to assess model performance during validation and depending on the results, it might be necessary to revise the model calibration until both the calibration and validation results comply with the acceptance criteria.

7. REFERENCES

Dingman, S.L. 2002. Physical Hydrology Second Edition.

Abbott, M.B. and Refsgaard, J.C. (eds.) (1996) Distributed Hydrological Modeling, Dordrecht, The Netherlands: Kluwer Academic Press.

Parkin, G. and Carron, J (2011). Technical Note on Guidelines for Model Calibration and Validation, Pilot Application of the Nile Basin Decision Support System: Stage 1, NBI Water Resources Planning and Management Project

Carron, J., Parkin, G., O'Donnell, G.M. and O'Connell, P.E. (2011). Development of Pilot Case Models, Nile Basin Decision Support System (DSS), Data Processing and Quality Assurance, Pilot Application of the Nile Basin Decision Support System: Stage 1, NBI Water Resources Planning and Management Project, 125pp

Coccia G. and Todini E. (2010). Recent developments in predictive uncertainty assessment based on the model conditional processor approach. Hydrol. Earth Syst. Sci. Discuss., 7, 9219–9270

Corps of Engineers (1994). Engineering and Design – Hydrologic Engineering Studies Design. United States Army Corps of Engineers publication #EP1110-2-9, Washington, D.C. 40 p.

Donigian, A.S. Jr. and P.S.C. Rao. 1990. Selection, Application, and Validation of Environmental Models. Proceedings of International Symposium on Water Quality Modeling of Agricultural Nonpoint Sources. Part 2. June 19-23, 1988. Logan, UT. USDA-ARS Report No. ARS-81. D. G. Decoursey (ed). pp 577- 604

Ewen, J., O'Donnell, G.M., Burton, A., and O'Connell, P.E. (2006). Errors and uncertainty in physically-based rainfall-runoff modeling of catchment change effects. Journal of Hydrology, 330: 641-650.

Klemeš, V., (1986). Operational testing of hydrological simulation models. Hydrological Sciences Journal 31: 13-24

Legates, D. R. and McCabe Jr., G. J.: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35, 1, 233–241, 1999.

Nash, I.E. and Sutcliffe, I.V. (1970). River flow forecasting through conceptual models, Part 1. J. Hydrology, 140, 1-23.

O'Connell, P.E., Carron, J., Parkin, G., and O'Donnell, G.M. (2011) Inception Report, Nile Basin Decision Support System (DSS), Data Processing and Quality Assurance, Pilot Application of the Nile Basin Decision Support System: Stage 1, NBI Water Resources Planning and Management Project, 204pp

Parkin, G. and O'Donnell, G.M. (2011). Development of Nile Baseline Model, Nile Basin Decision Support System (DSS), Data Processing and Quality Assurance, Pilot Application of the Nile Basin Decision Support System: Stage 1, NBI Water Resources Planning and Management Project, 92pp

U.S. EPA. (2002) Guidance for Quality Assurance Project Plans for Modeling (EPA QA/G-5M). U.S. Environmental Protection Agency, Washington, D.C., 121 p.

Moriasi D. N., Arnold J. G., Van Liew M. W., Bingner R. L., Harmel R. D. and Veith T. L. 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. American Society of Agricultural and Biological Engineers Vol. 50(3): 885−900.

APPENDIX A

The Blue Nile Case Study

A1. INTRODUCTION

The approach to modelling the streamflow in the Blue Nile catchment was to select appropriate streamflow gauging stations based on pre-determined quality criteria and then to calibrate the NAM rainfall-runoff model at these gauges such that a representative streamflow time series was simulated when compared to the observed streamflow. Suitably defined objective functions were used to quantify the goodness-of-fit between the simulated and observed records. Thereafter, streamflow sequences for input to the Blue Nile Mike Basin model were prepared.

The calibration and validation of the Blue Nile MIKE Basin model in the ENJMP Pilot Case entailed three phases:

1. Calibration and validation of the NAM rainfall-runoff model at key flow gauging stations on tributaries and on the main Blue Nile River.
2. The calibrated NAM parameters were then used to generate long-term flow sequences based on observed rainfall for both the gauged and ungauged catchments. (Calibration parameters were transferred to the ungauged catchments.)
3. Using the generated flow sequences as inflow sequences in the Blue Nile MIKE Basin model, the Blue Nile MIKE Basin model was validated based on observed flows in the Blue Nile at Khartoum Soba, in the Main Nile at Dongola and observed water levels at Lake Tana and Roseires dams.

During model calibration and validation, existing infrastructure was included in the model simulations when these infrastructure components were present during the calibration / validation periods. In such cases, it was ensured that the operating rules which governed these infrastructure components were modelled accurately.

A2. APPROACH TO NAM MODEL CALIBRATION AND VALIDATION

Model selection

The NAM rainfall-runoff model was selected for calibrating flows in the Blue Nile. NAM is a deterministic, lumped model, it is based on physical processes (conceptual) and continuous in that it accounts for moisture in four different and mutually inter-related storage zones. It requires limited data inputs namely rainfall and evaporation (and temperature for snow conditions). The main model outputs include runoff, groundwater levels, temporal variation in moisture content, and time series of other physical processes e.g. infiltration, recharge, inter- , overland- and baseflows.

Calibration catchments

The sub-catchments in the Blue Nile as listed in Table A-1 below were selected for NAM model calibration and validation. Calibration sites were selected based on availability of observed streamflow records in the catchment as well as the availability of observed rainfall data. Consideration was also given to the observed streamflow record length and the number of missing values in the dataset.

Table A-1 : Flow gauges for NAM calibration and validation in the Blue Nile

Flow Gauge	Observed MAR (Mm^3/a)	Catchment area (km^2)	MAP (mm)	Runoff coefficient
Gumara at Bahir Dar	852	1394	1274	48%
Abbay at Kessie (incremental)	10899	49789	1092	20%
Didessa at Arjo	4027	9981	1186	34%
Dabus at Asossa	4894	10139	1735	28%
Abbay at Border/El Diem (incremental)	33915	109786	1307	24%

Calibration and Validation periods

The selection of calibration periods was based on the longest period of observed record that contained the fewest patched or infilled values from the Ethiopian Masterplan. Wherever possible a minimum flow record length of 10 years was selected to calibrate on. Table A-2 presents the calibration and validation periods that were selected.

Table A-2 : Calibration and validation periods at flow gauges in the Blue Nile

Flow gauge	Record period	% Patched / Infilled values	Calibration period	Validation period
Gumara at Bahir Dar	Jan 1960 – Dec 1992	5%	1970 – 1980	1981 - 1990
Abbay at Kessie (incremental)	Jan 1960 – Dec 1992	5%	1960 - 1968	1972 – 1982
Didessa at Arjo	Jan 1960 – Dec 1992	27%	1979 - 1989	1962-1972
Dabus at Asossa	Jan 1963 – Dec 1979	3%	1963 - 1974	1975-1980
Abbay at Border/El Diem	Jan 1961 – Dec 1979	8%	1969 -1979	1961-1968

Model input data

The subcatchment boundaries for the main tributaries and key flow gauging locations were delineated using GIS tools. The Mean Annual Rainfall was determined for each of the delineated subcatchments from the long term RSE v2 satellite rainfall estimates. A catchment rainfall file was generated from several patched rainfall stations in and around the sub-catchment which were considered to be representative of rainfall across the catchment and to obtain a rainfall record that was sufficiently long for calibration and validation purposes. The methodology for generation of catchment rainfall files is described in detail in the Data Report (NBI, 2012). Observed flow data were obtained from the Ethiopian Master Plan and from the Nile Encyclopaedia. Evaporation data was obtained from the FAO Climate database for representative stations in each calibration subcatchment. Where necessary during model calibration, existing infrastructure with its associated operating rules, were included in the simulations when this infrastructure affected the observed flow record used for model calibration.

Acceptance of fit criteria

Initially the observed and simulated streamflow sequences were compared graphically to identify outliers, periods of good and poor fit, correspondence of high flows, recession, and low flows. Once a reasonable calibration was obtained for a particular catchment the MAR's of the simulated streamflows for then calibration period is compared to the observed flows for the same period. The r^2 parameter was determined for record purposes. Normally an r^2 of about 0.8 is considered a good fit, and values of about 0.5 (Santhi et al 2001) might be deemed acceptable. The following criteria were considered when comparing simulated vs. observed model output at key locations for the acceptance of model calibration:

<u>Visual assessment</u>

- A good agreement between the monthly and annual timeseries plots of simulated and observed catchment runoff
    - A good overall agreement of the shape of the hydrograph
    - A good agreement of the peak flows with respect to timing and volume
    - A good agreement for low flows.
- Seasonal flow distribution; Standardized residuals; Correlation between simulated and observed monthly flow volumes; Cumulative flow; Unit runoff

<u>Quantitative assessment</u>

- Mean Annual Runoff (MAR); Standard deviation; Seasonality index; Total water balance; Dam storage trajectories

Specific target values for quantitative measures are difficult to define precisely, as these depend on data availability and quality. Typically, in terms of water balance, it should be aimed to achieve an error of less than 5% when comparing long-term mean annual flows at key nodes. When assessing the degree of correlation between monthly observed and simulated flows, an r^2 value larger than 0.8 indicates a good fit, while standardised residuals should preferably be confined to between -0.2 and 0.2. Indices such as standard deviation are generally very sensitive to outliers and this should be borne in mind when evaluating model performance.

A3. NAM CALIBRATION

As an example of a NAM model calibration, the observed flows for the Didessa tributary at Arjo were used as a calibration gauge.

Data preparation

The observed flows for the Didessa tributary at Arjo were sourced from the Ethiopian Master Plan in Stage 1. Observed monthly flows were available for the period January 1960 to December 1992. There were 27% missing values which were patched and extended during Stage 1 using PatchS.

Catchment rainfall for the Didessa catchment was generated using the rainfall stations in Table A-3.

Table A-3 : Rainfall stations used for catchment rainfall in Didessa catchment

Station name	Station number	Source	MAP	Record length
Bedele	R_003315	Ethiopian Masterplan	1609	1951-1992
Chora Kumbabe	R_003316	Ethiopian Masterplan	1767	1951-1992
Anger Gutin	R_003329	Ethiopian Masterplan	1593	1971-1992
Getema	R_003352	Ethiopian Masterplan	1414	1954-1988
Nekemte	633400000	GHCN	2076	1963-1998
Lekemti	ET96NKMT	FAO	2039	1970-1998

Visual data pre-checks

Visual pre-checks on flow data are useful as preliminary checks on data quality and to identify any obvious data anomalies. A plot of observed flow compared to catchment rainfall as well as a unit runoff plot is shown graphically below. The unit runoff plot displays the relationship between mean annual precipitation and specific runoff for the study catchment and is useful to obtain an indication of catchment response variability.

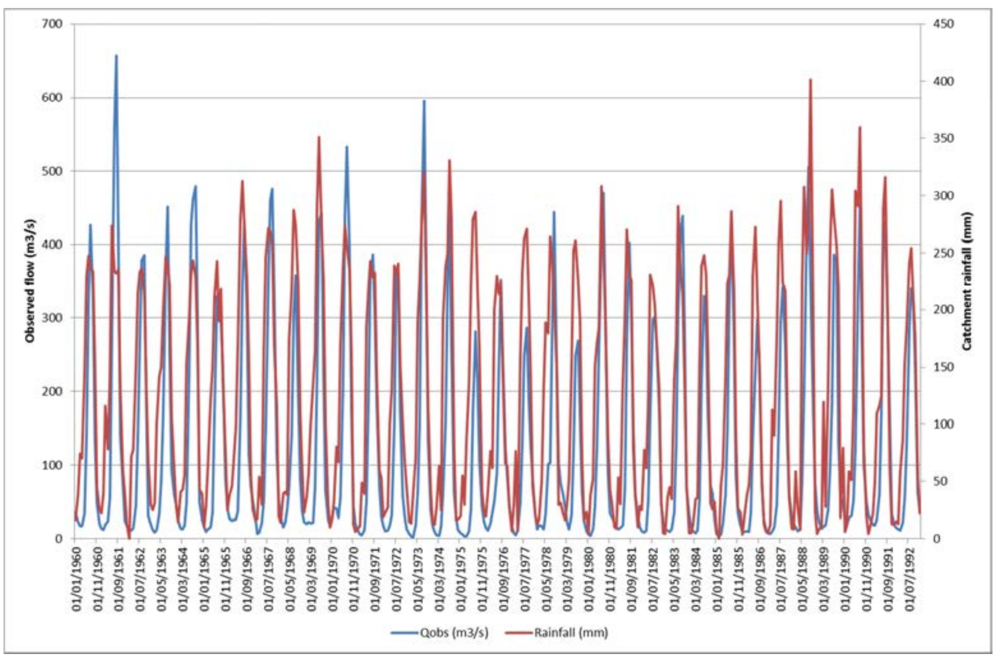


Figure A-1 : Comparison of observed streamflow and catchment rainfall

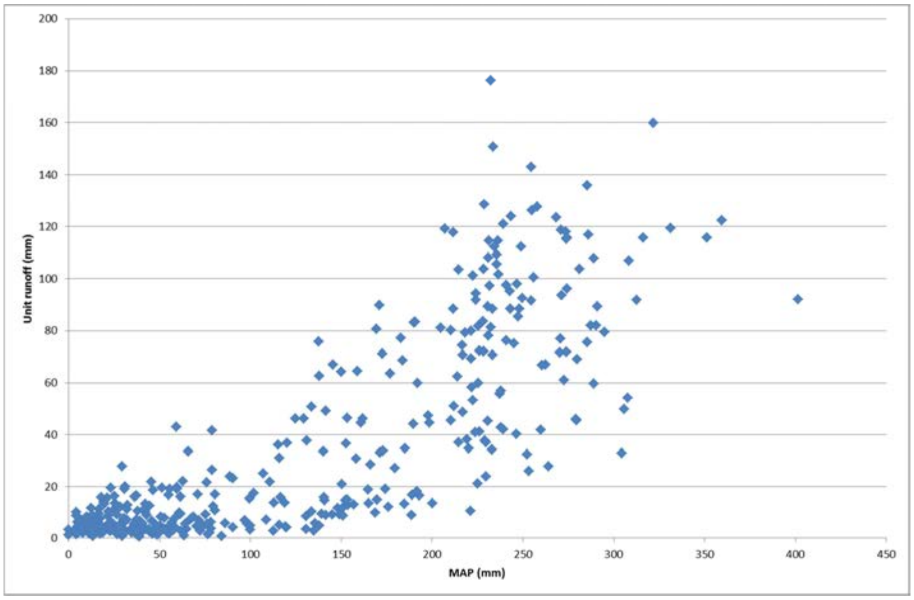


Figure A-2 : Unit runoff versus MAP for Didessa at Arjo

Calibration technique

The following process is typically followed to carry out a successful NAM calibration:

- Sensitivity analysis to identify physical parameters characterizing the actual rainfall-runoff process

A. Fit of water balance of calibration period - adjust actual evapotranspiration by adjusting Lmax/Umax

- Fit of flood peaks – adjust overland flow:
  - Volume – CQOF
  - Timing – TOF
  - Shape – CK1 = CK2

B. Fit of low flows - adjust baseflow:
  - Volume - adjust overland flow adjusted recharge i.e. again CQOF
  - Timing - TG
  - Shape - CKBF

- Verify on different data

Initial Run – Default parameters

Initially, the NAM model was run for the full observed period (1960-1992) with the default NAM parameters. The simulated results are shown below (Figure A-3). The r^2 is 0.685, observed runoff 407mm and simulated runoff 499mm.

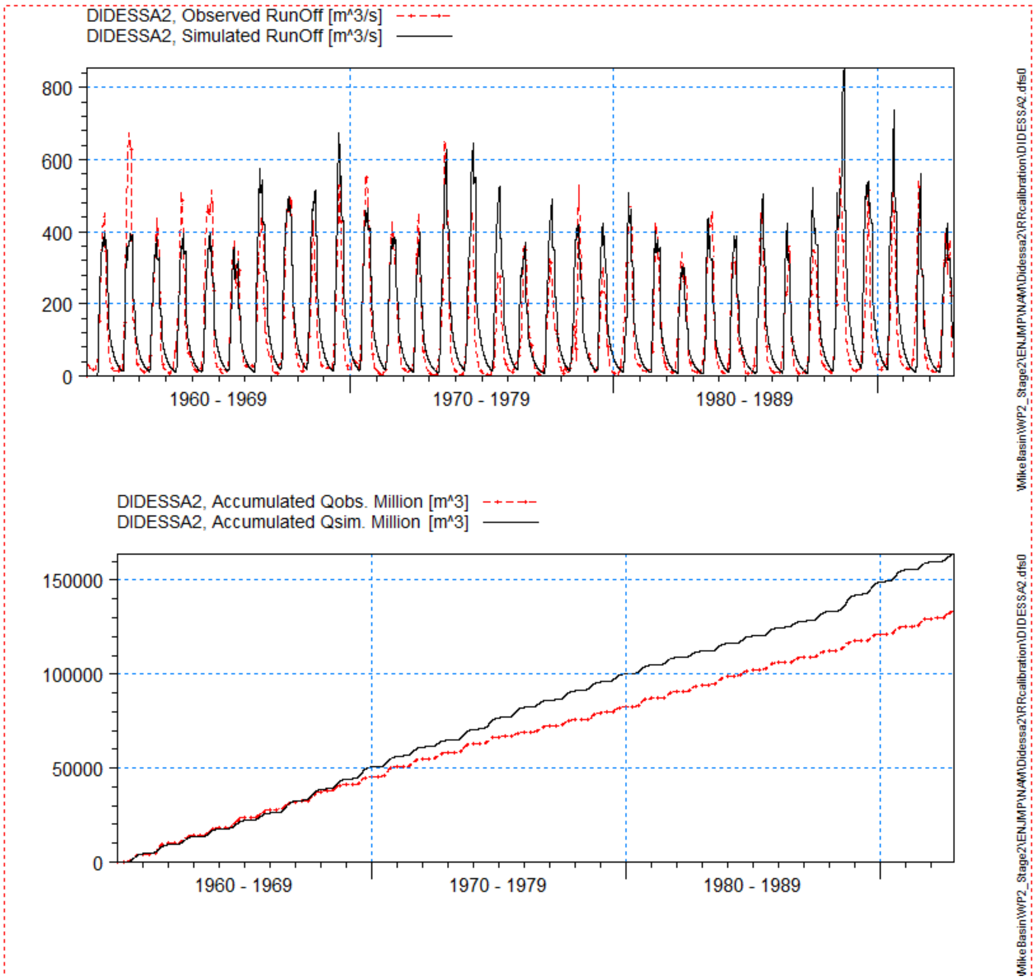


Figure A-3 :NAM simulation – default parameters, full period

Run 1 – Initial run

The calibration period from 1979 to 1989 was selected and NAM was re-run with default parameters (Run 1). The resulting simulated and observed flows are shown summarised below. The r^2 is 0.603, observed runoff 380 mm and simulated runoff 479 mm. The flows are over-simulated and therefore it is necessary to adjust L_{max} and U_{max} to fit the water balance.

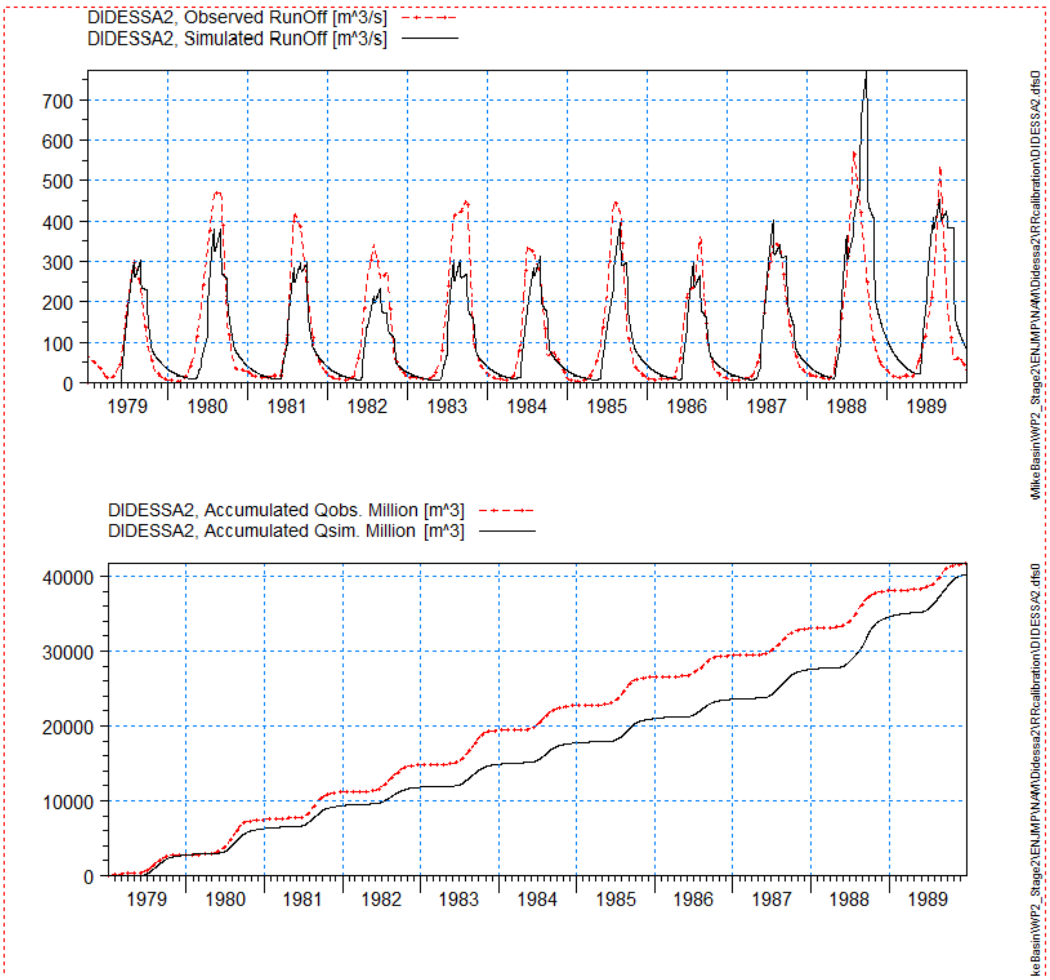


Figure A-4 : NAM simulation 1 – default parameters, calibration period 1979-1989

Run 2 – Fit water balance

Increase L_{max} and $U_{max,}$ to reduce overland flow and increase infiltration - keep U_{max} as 10% of L_{max}. The resulting simulated flows are shown below. The r^2 is 0.682, observed runoff 380 mm and simulated runoff 367 mm. The water balance shows a much closer fit with simulated flows slightly less than the observed flows. The next step is to try to fit the flood peaks by adjusting the CQOF, TOF and CK1,2 for volume, timing and shape respectively.

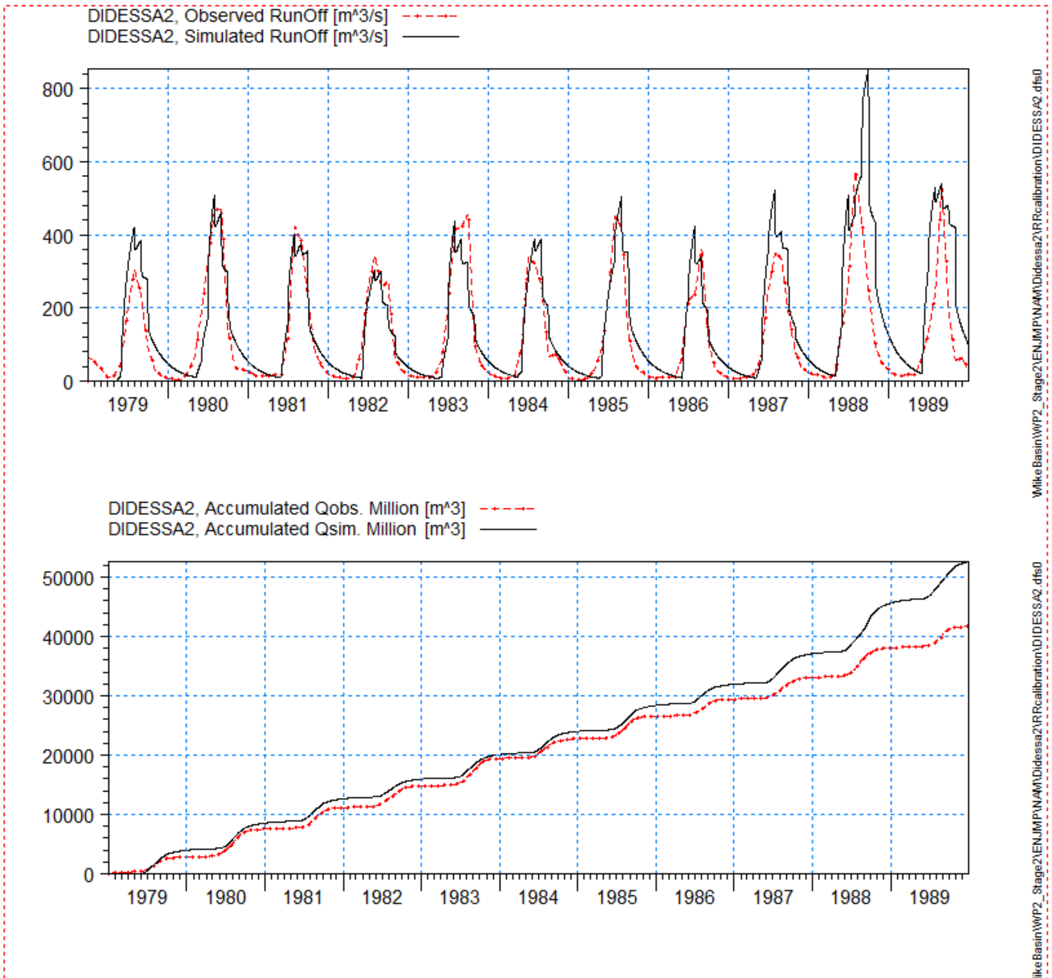


Figure A-5 : NAM simulation 2 – calibration period 1979-1989

Run 3 – Fit flood peaks

In order to fit the flood peaks better, adjustments were made to CQOF, TOF and CK1,2 parameters. The resulting simulated flows are shown below. The r^2 is 0.578, observed runoff 380 mm and simulated runoff 374 mm. The hydrograph plot shows an improved fit on the low flows (or volume of flood) and slightly better agreement with flood peaks except for the over simulation of the peak in late 1988. The r^2 value has now decreased however and the cumulative plot shows undersimulation over the period of calibration.

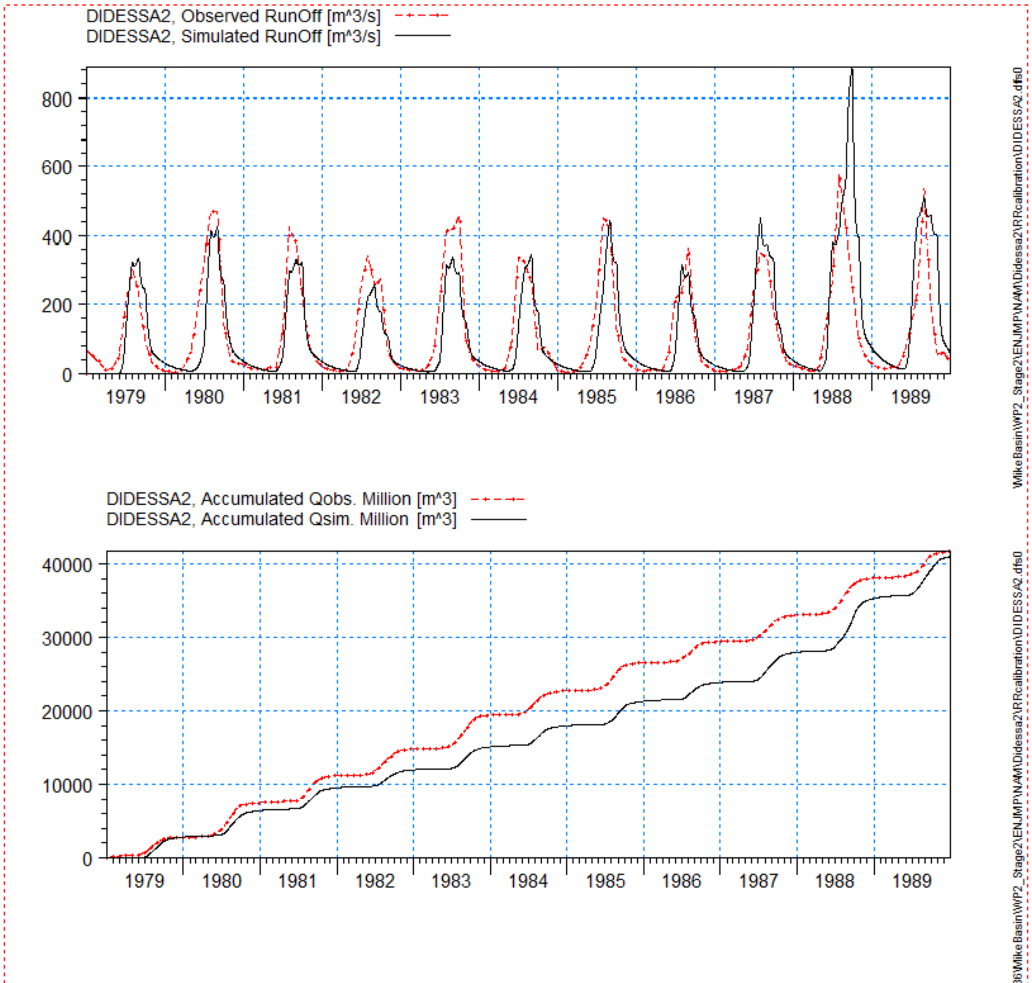


Figure A-6 : NAM simulation 3 – calibration period 1979-1989

Run 4 – Fit low flows

In order to fit the low flows better, the baseflows need to be adjusted by changing the CQOF, TG and CKBF parameters. Increasing TG and CKBF will help to delay the start of groundwater recharge and flow in beginning of wet season and allow for quicker filling of the root zone. It will also allow for longer duration of flow and lower peaks. The resulting simulated flows are shown below. The r^2 is 0.628, observed runoff 380 mm and simulated runoff 355 mm.

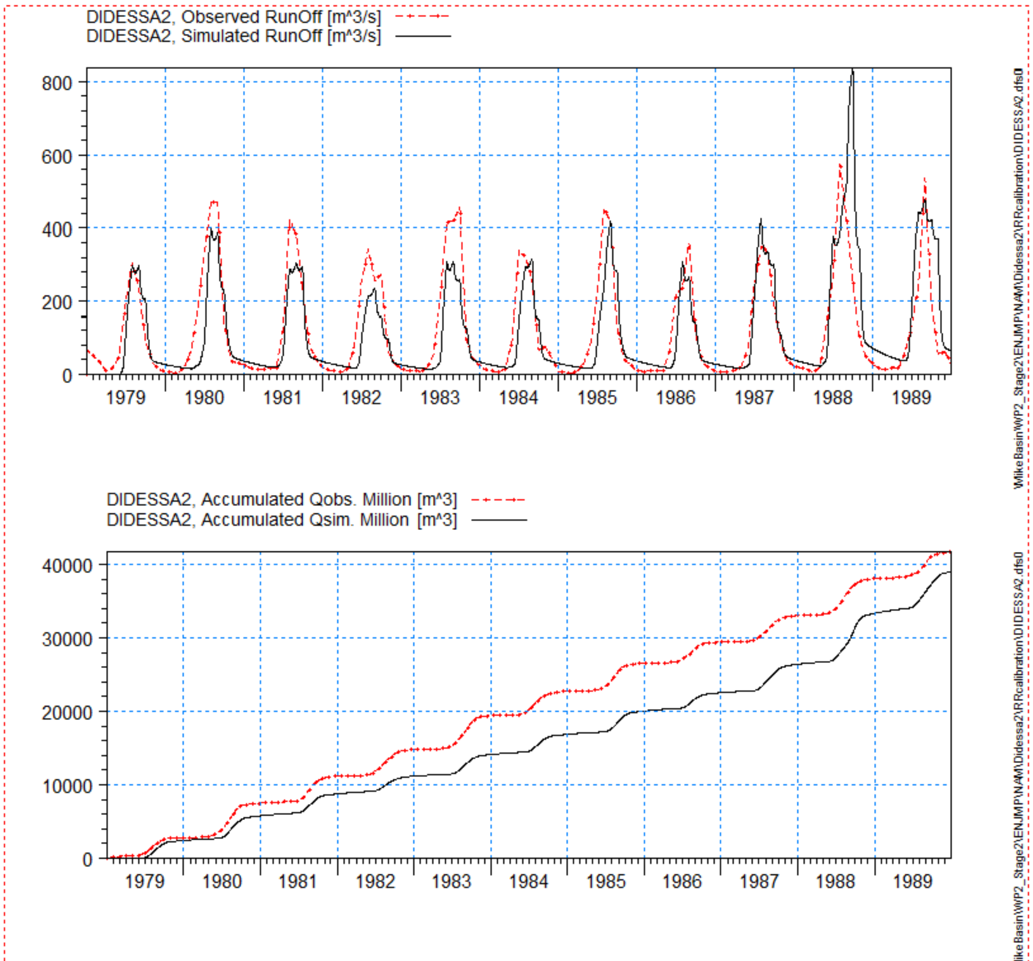


Figure A-7 : NAM simulation 4 – calibration period 1979-1989

Run 5 – Final adjustment

For the final adjustment, CKIF and TIF were adjusted in order to improve the calibration statistics. The resulting simulated flows are shown below. The r^2 is 0.647, observed runoff 380 mm and simulated runoff 377 mm.

Figures A-8 to A-10 display the goodness of fit for the final calibration and Table A-4 summarises the final calibration statistics.

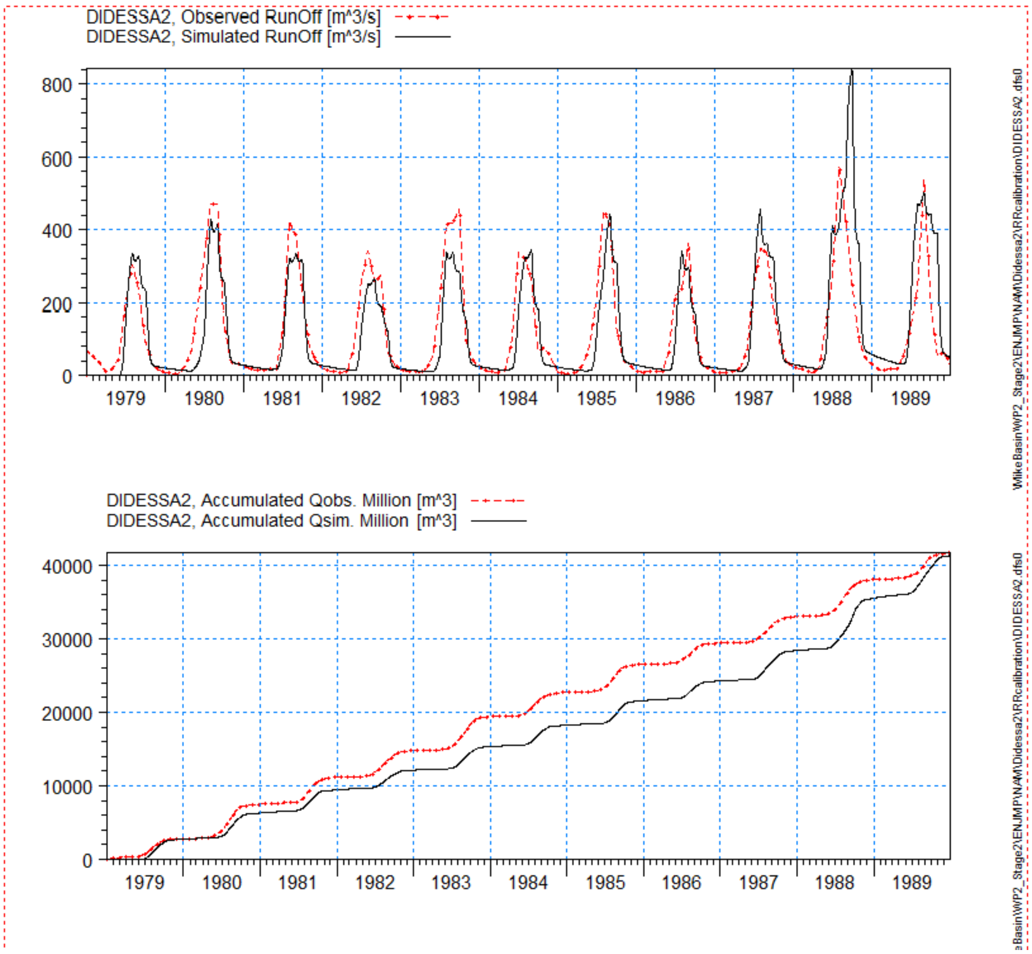


Figure A-8 : NAM simulation 5 – calibration period 1979-1989

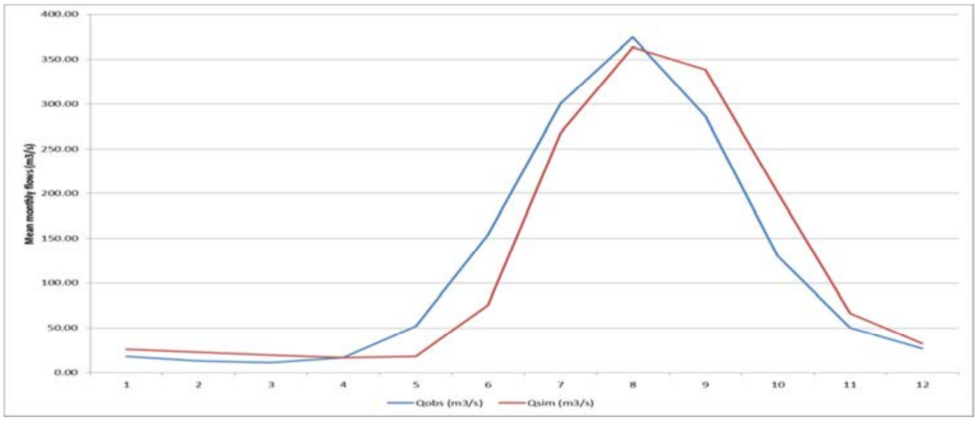


Figure A-9 : Observed and simulated seasonal flows for Didessa at Arjo

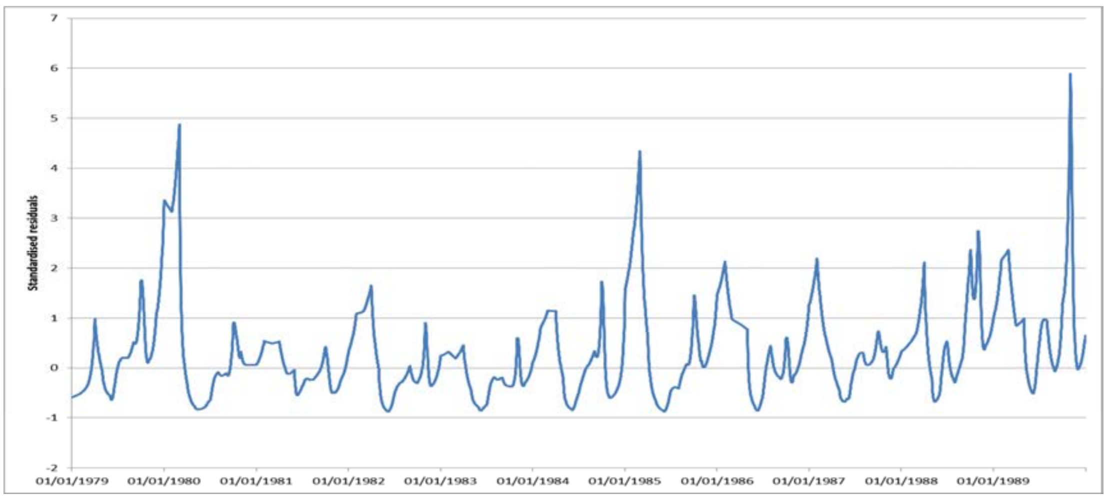


Figure A-10 : Standardised flow residuals for calibration period at Didessa

Table A-4 : Calibration statistics for Didessa at Arjo

1979-1989	Observed	Simulated	Difference
Mean Annual Flow (million m^3/a)	3796	3833	+1%
Std.Dev. (m^3/s)	137	149	+8%
r^2 = 0.65			

The NAM parameters used for each of the runs presented above are summarised below:

Table A-5 : NAM calibration parameters for Didessa

Run	U_{max}	L_{max}	CQOF	CKIF	CK1,2	TOF	TIF	TG	CKBF
1 (Default)	10	100	0.5	1000	10	0	0	0	2000
2 (Fit water balance)	**30**	**300**	0.5	1000	10	0	0	0	2000
3 (Fit flood peaks)	30	300	**0.85**	1000	**100**	**0.25**	0	0	2000
4 (Fit low flows)	30	300	**0.85**	1000	100	0.25	0	**0.2**	**5000**
5 (Final adjustment)	30	300	0.85	**350**	100	0.25	**0.05**	0.2	5000

A4. NAM VALIDATION

Flows were simulated for the period 1962-1972 to check the final calibration parameters for a validation period from 1962 to 1972. The resulting simulated flows are shown below. The R^2 is 0.783, observed runoff 437 mm and simulated runoff 412 mm. The comparison of flows during validation period shows a good overall agreement and the statistics are within acceptable limits, although there is a slight lag in the simulated flows – similar to the calibration period. There is also a tendency to overestimate dry season flows and to underestimate high flows during the wet season.

Figures A-11 to A-13 display the goodness of fit for the validation and Table A-6 summarises the final validation statistics.

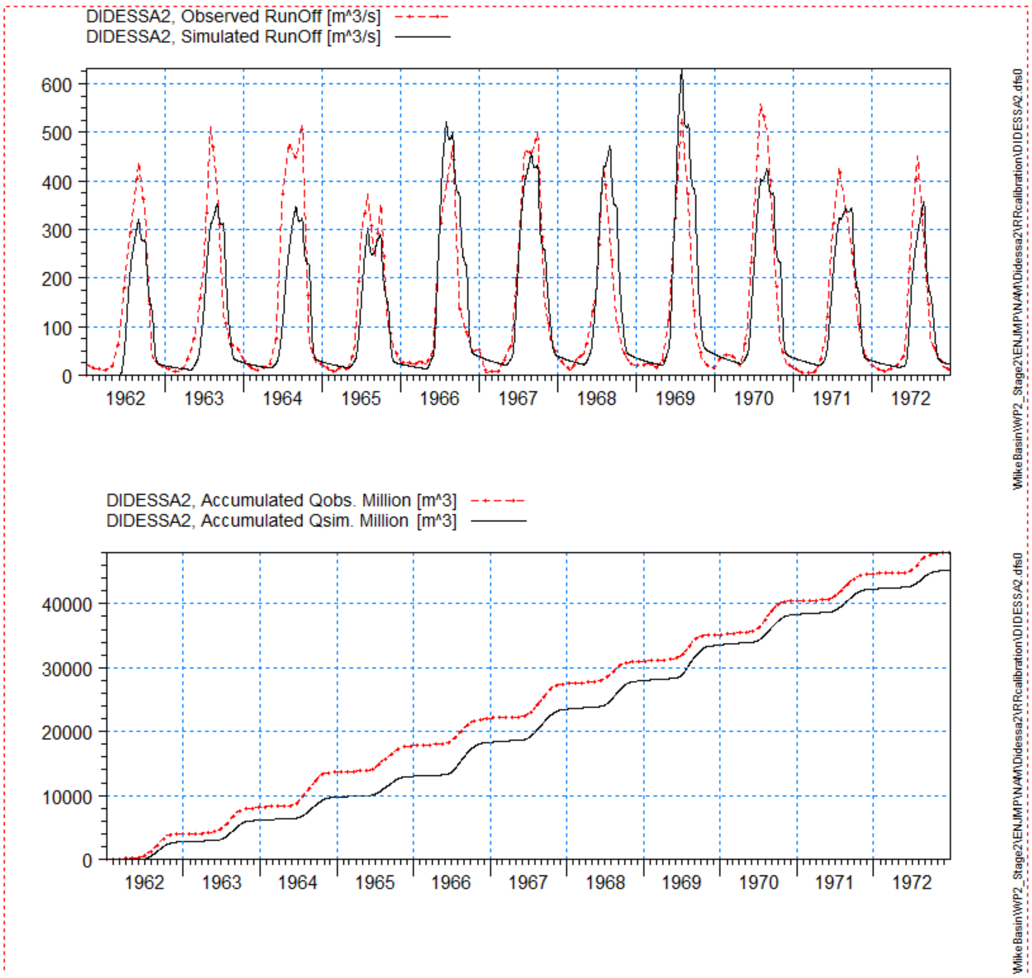


Figure A-11 : NAM simulation – validation period 1962-1972

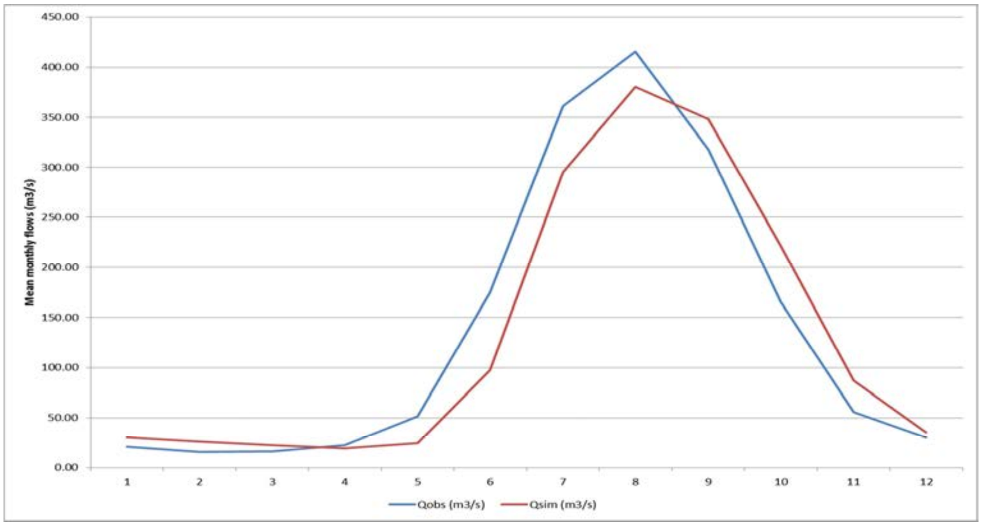


Figure A-12 : Observed and simulated mean monthly flows at Didessa at Arjo for the validation period (1962-1972)

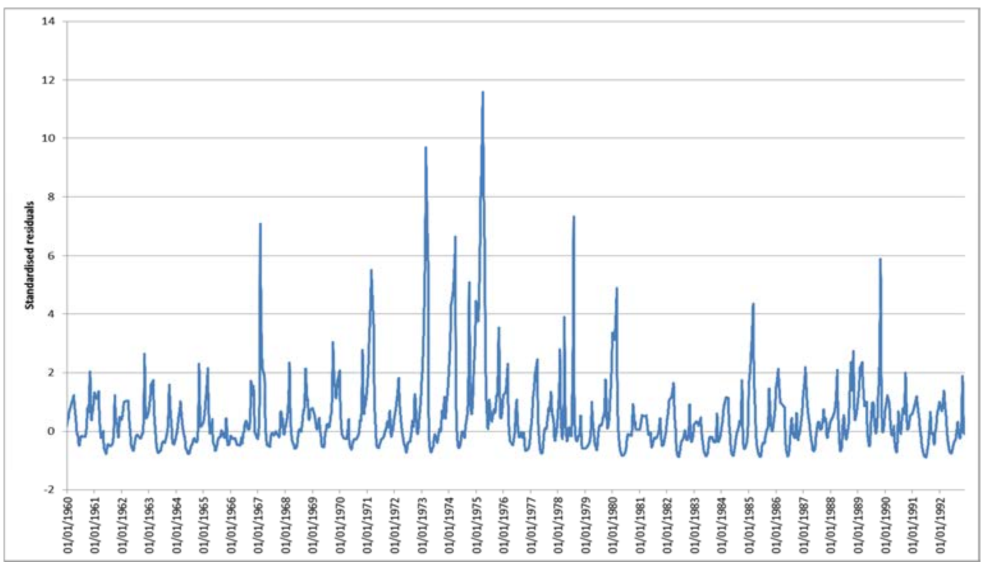


Figure A-13 : Standardised flow residuals, validation period at Didessa at Arjo (1962-1972)

Table A-6 : Validation statistics: Didessa at Arjo

1962-1972	Observed	Simulated	Difference
MAR (million m^3/a)	4365	4201	-4%
Std.Dev. (m^3/s)	155	145	-6%
r^2 = 0.79			79%

A5. BLUE NILE MIKE BASIN VALIDATION

Having calibrated and validated the NAM model on individual sub-catchments of the Nile Basin, the next step entailed the validation of the Blue Nile and Main Nile rivers down to Aswan Dam based on water balances (mean annual flows) and flow patterns in the Blue Nile at Khartoum Soba, in the Main Nile at Dongola and against observed water levels at Lake Tana and Roseires dams. For the Blue Nile system, all flows were simulated by means of the NAM model. This entailed using the calibrated parameters for each of the calibration subcatchments. Runoff in ungauged tributaries was generated by the NAM model based on transferred calibration parameters.

Blue Nile at Khartoum Soba

The simulated flows for the full model simulation period (1951-1990) in the Blue Nile River at Khartoum Soba, upstream of the confluence with the Main Nile, were compared to the observed flows. A comparison of the flow statistics is included in Table A-7. The validation results are presented graphically in Figure A-14 to A-17 below. Overall, there is a good agreement between the simulated flows and the observed flows at Khartoum Soba. The mean annual flows are 5% less than the observed with an r^2 value of 68%. The peak flows are slightly over simulated and the standardised residuals originally indicated a high error in February 1983, which, after closer inspection, revealed an error in the observed flow value, which was repatched.

Table A-7 : Validation statistics for the Blue Nile at Khartoum Soba

1951-1990	Observed	Simulated	Difference
Mean Annual Flow (million m^3/a)	42322	40123	-5%
Std.Dev. monthly flows (m^3/s)	1843	2226	+21%
R^2 = 68%			

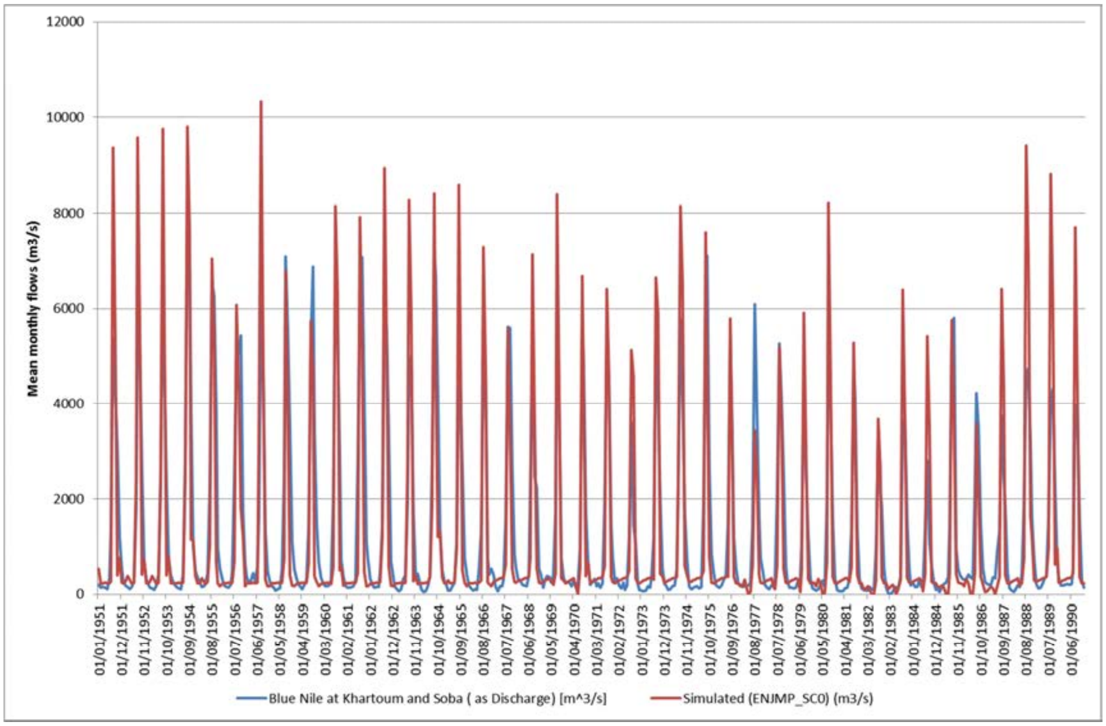


Figure A-14 : Blue Nile observed and simulated monthly flows at Khartoum Soba

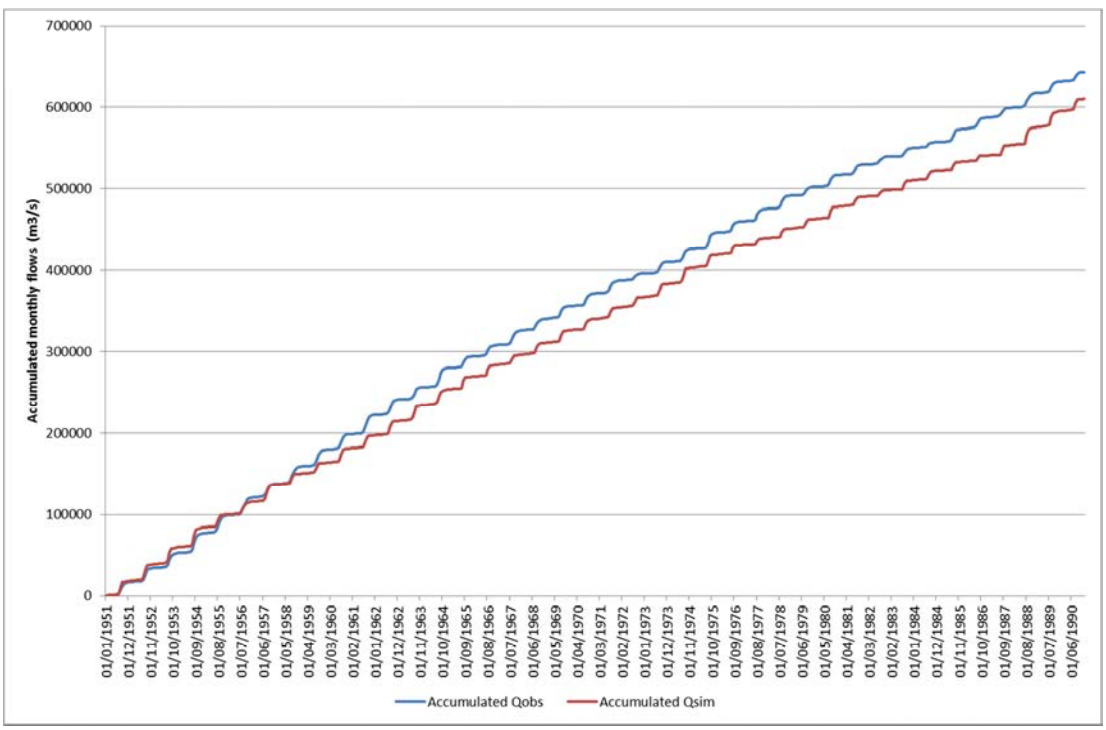


Figure A-15 : Observed and simulated accumulated monthly flows at Khartoum Soba

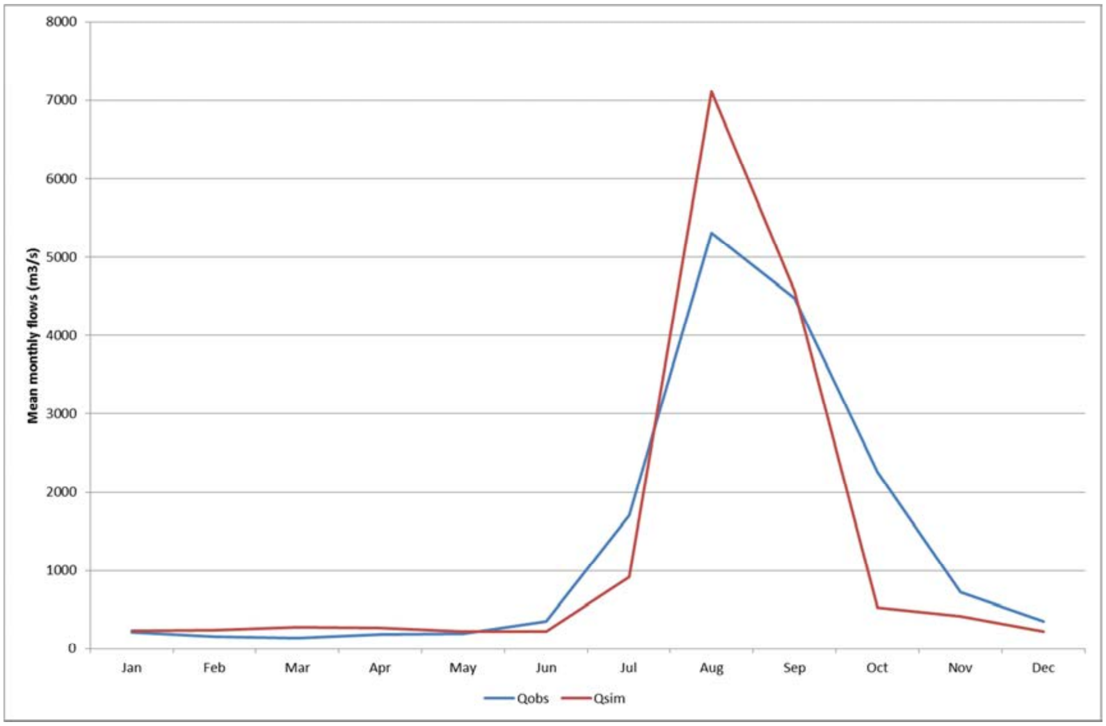


Figure A-16 : Observed and simulated mean monthly flows at Khartoum Soba

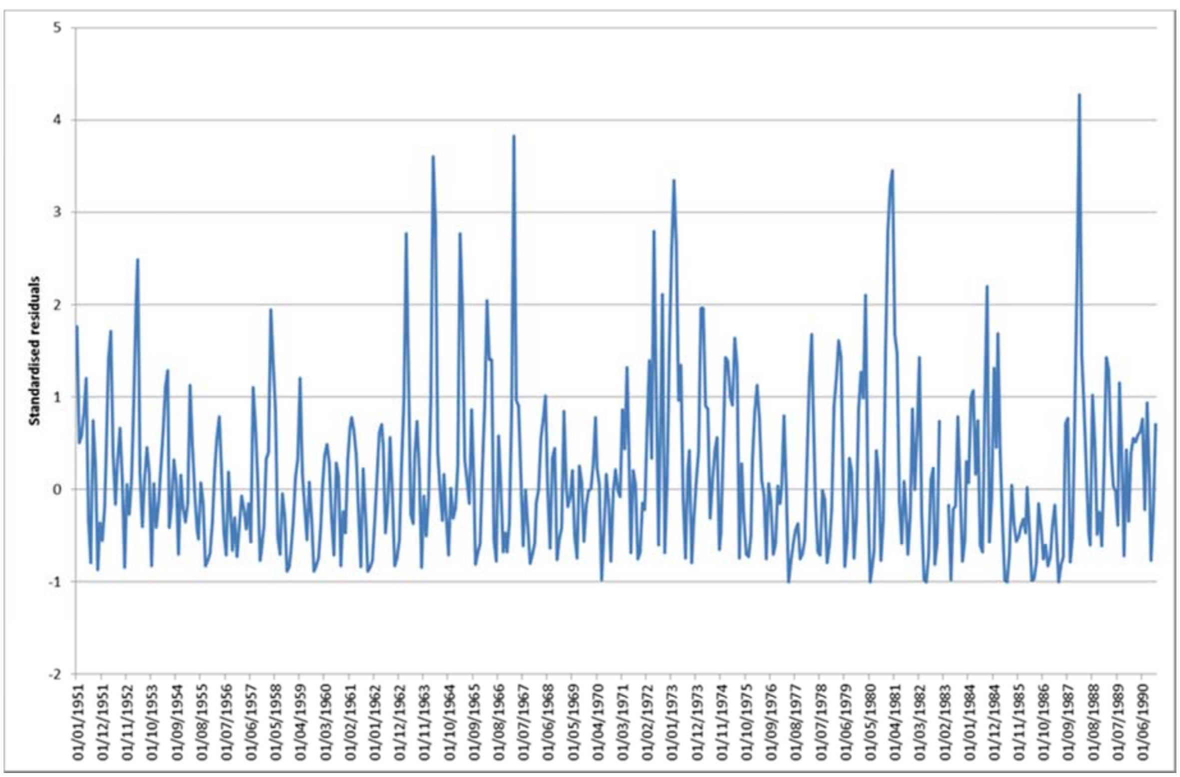


Figure A-17 : Standardised flow residuals, validation at Khartoum Soba

Main Nile at Dongola

Observed flows at Dongola were available for the period 1962 – 1997 from the Nile Encyclopaedia. The simulated flows for the period 1962-1990 in the Main Nile at Dongola were compared to the observed flows. A comparison of the flow statistics is included in Table A-8. The validation results are presented graphically in Figure A-18 to A-21 below. The accumulated monthly totals (Figure A-19) show a consistent under estimation of total flow volume after about 1975, however the statistics show that overall, there is a good agreement between the simulated flows and the observed flows at Dongola with only 4% difference in MAR and with an overall R^2 of 74%, which is reasonable and was accepted. The comparison of mean monthly flows (Figure A-20) shows an over estimation of the peak flows, however the observed peak is flattened which may be an indication that peak flows are under estimated in the observed record. It is recommended that this be investigated in future studies using this flow record.

Table A-8 : Validation statistics for the Main Nile at Dongola

1962-1990	**Observed**	**Simulated**	**Difference**
Mean Annual Flow (million m^3/a)	71311	68505	-4%
Std.Dev. monthly flows (m^3/s)	2097	2300	+10%
R^2 = 74%			

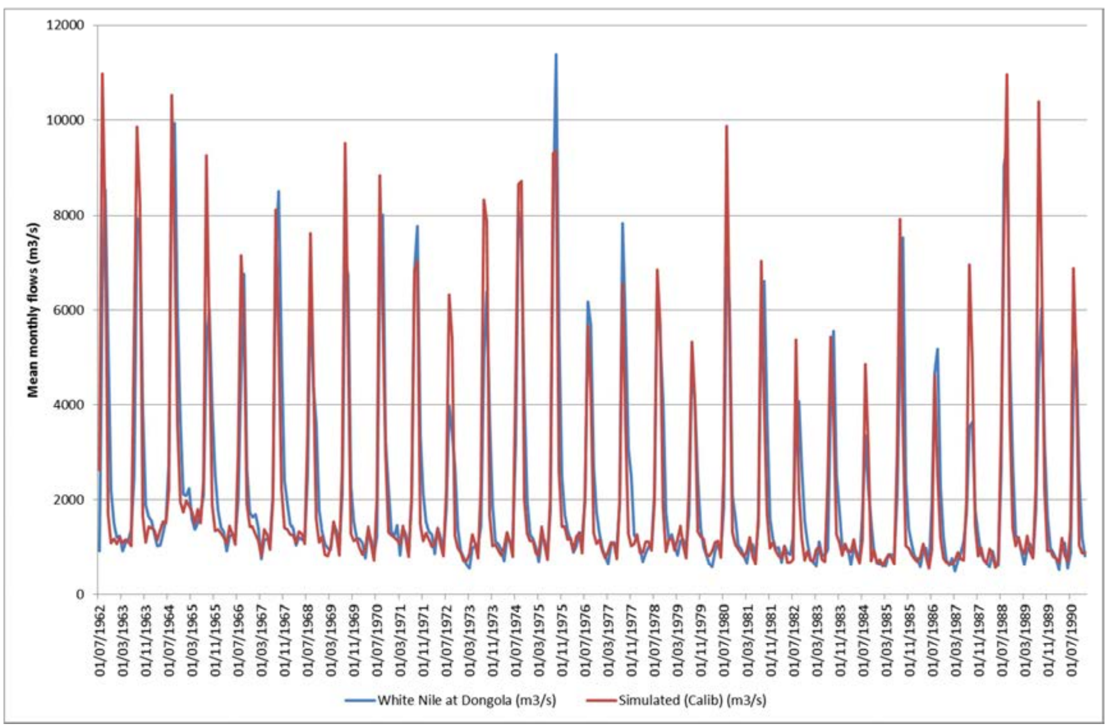


Figure A-18 : Main Nile observed and simulated monthly flows at Dongola

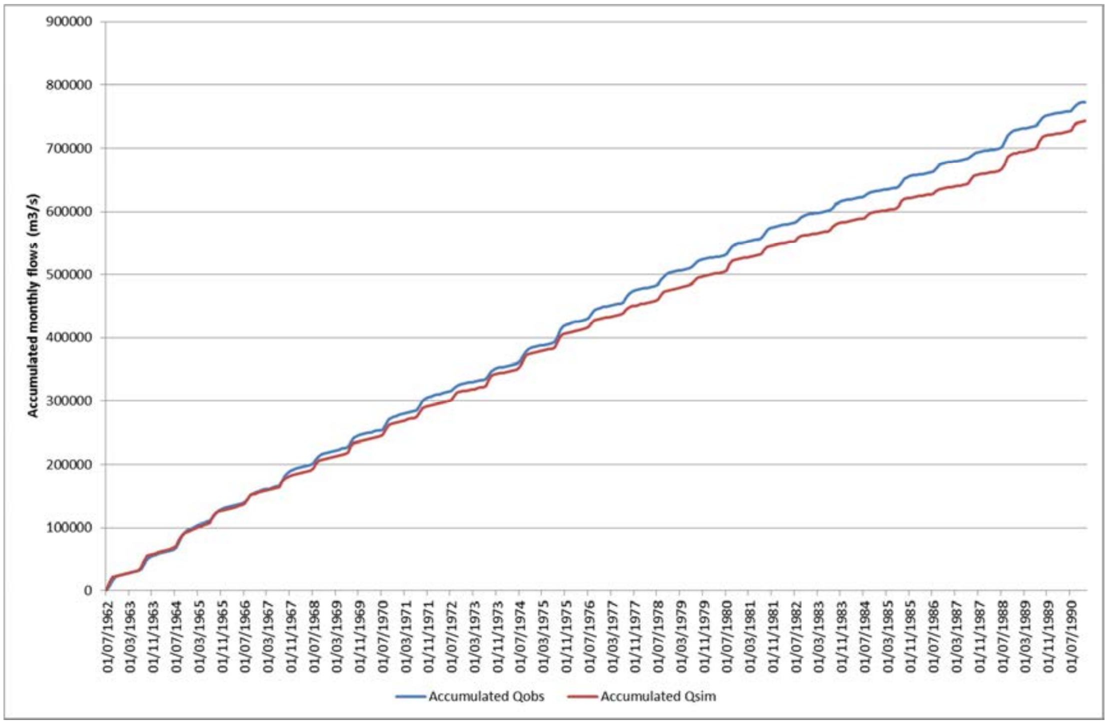


Figure A-19 : Observed and simulated accumulated monthly flows at Dongola

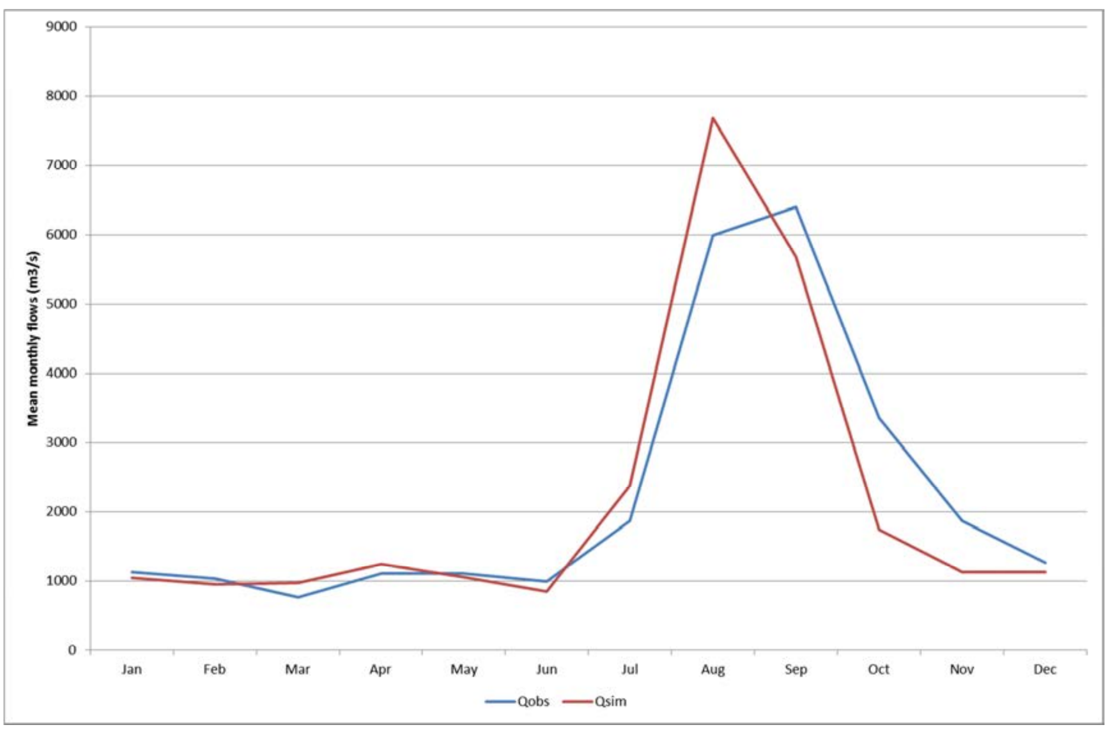


Figure A-20 : Observed and simulated mean monthly flows at Dongola

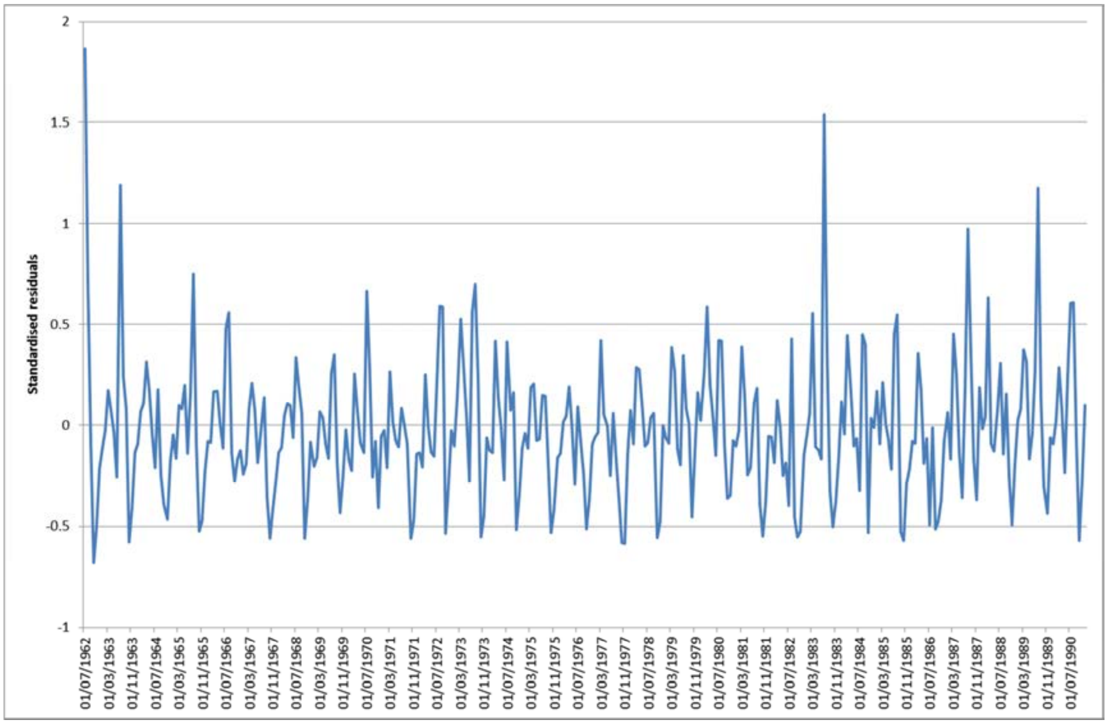


Figure A-21 : Standardised flow residuals, validation at Dongola

Lake Tana Water Levels

The calibrated NAM parameters, based on the calibration in the Gumara River at Bahir Dar, were transferred to the other tributary catchments of Lake Tana and inflows into Lake Tana were simulated for the full model simulation period (1951-1990). The simulated water levels in Lake Tana were then compared to the observed water levels (see Figure A-22) and were found to have a good agreement.

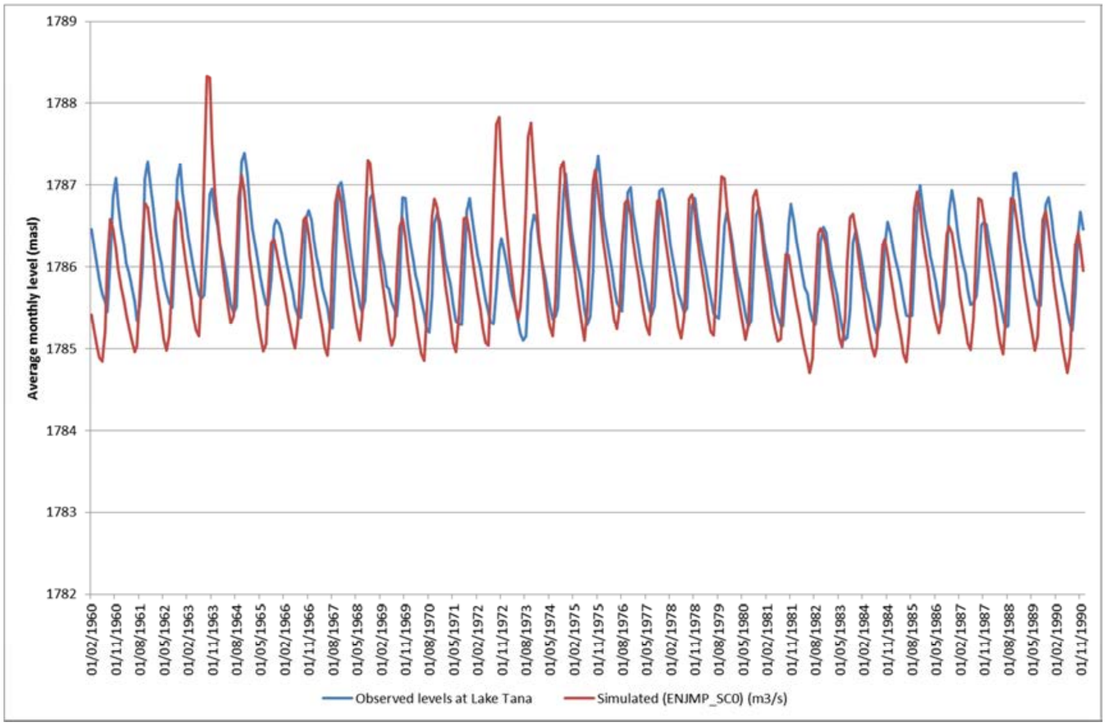
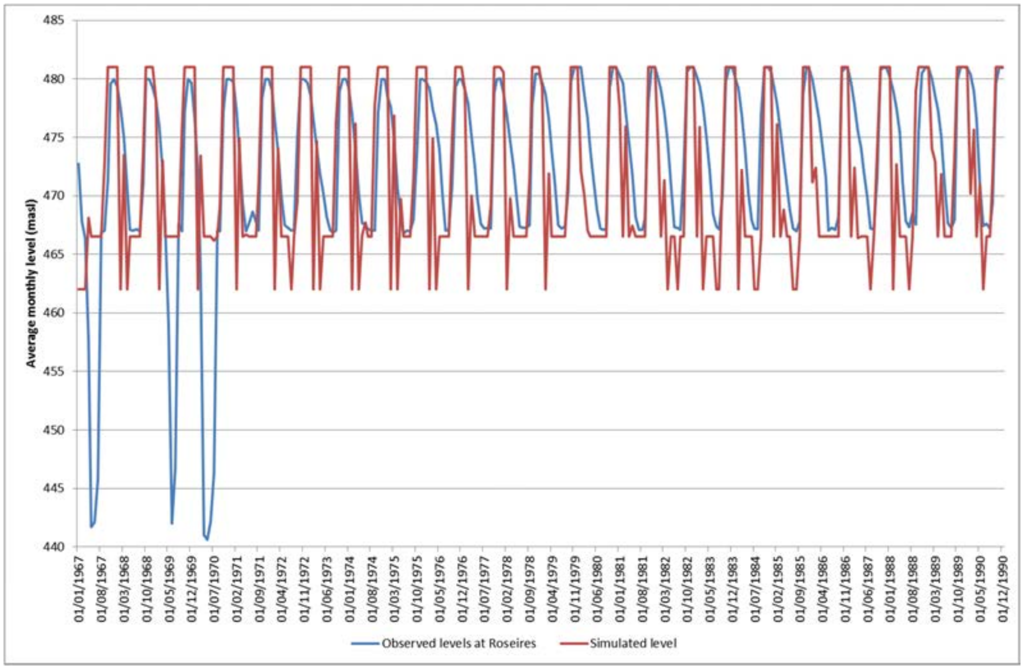


Figure A-22 : Observed and simulated water levels in Lake Tana, 1951-1990

Roseires Water Levels

Finally, the simulated water level in Roseires Dam, after its construction in 1966, were compared to the observed water level, based on simulated flows in the Blue Nile catchment using the calibrated NAM parameters. A comparison of the observed and simulated water levels in Figure A-23 shows an overall good agreement. The difference between the minimum simulated water levels and observed minimum water levels relate to the variable implementation of minimum drawdown levels in Roseires Dam during the simulation period.



Fiure A-23 : Observed and simulated water levels at Roseires

From the above, it was concluded that the calibrated NAM model provides an acceptable simulation of runoff within the Blue Nile Basin and this model was subsequently taken forward into the scenario analysis component of this study, which included a climate change scenario.

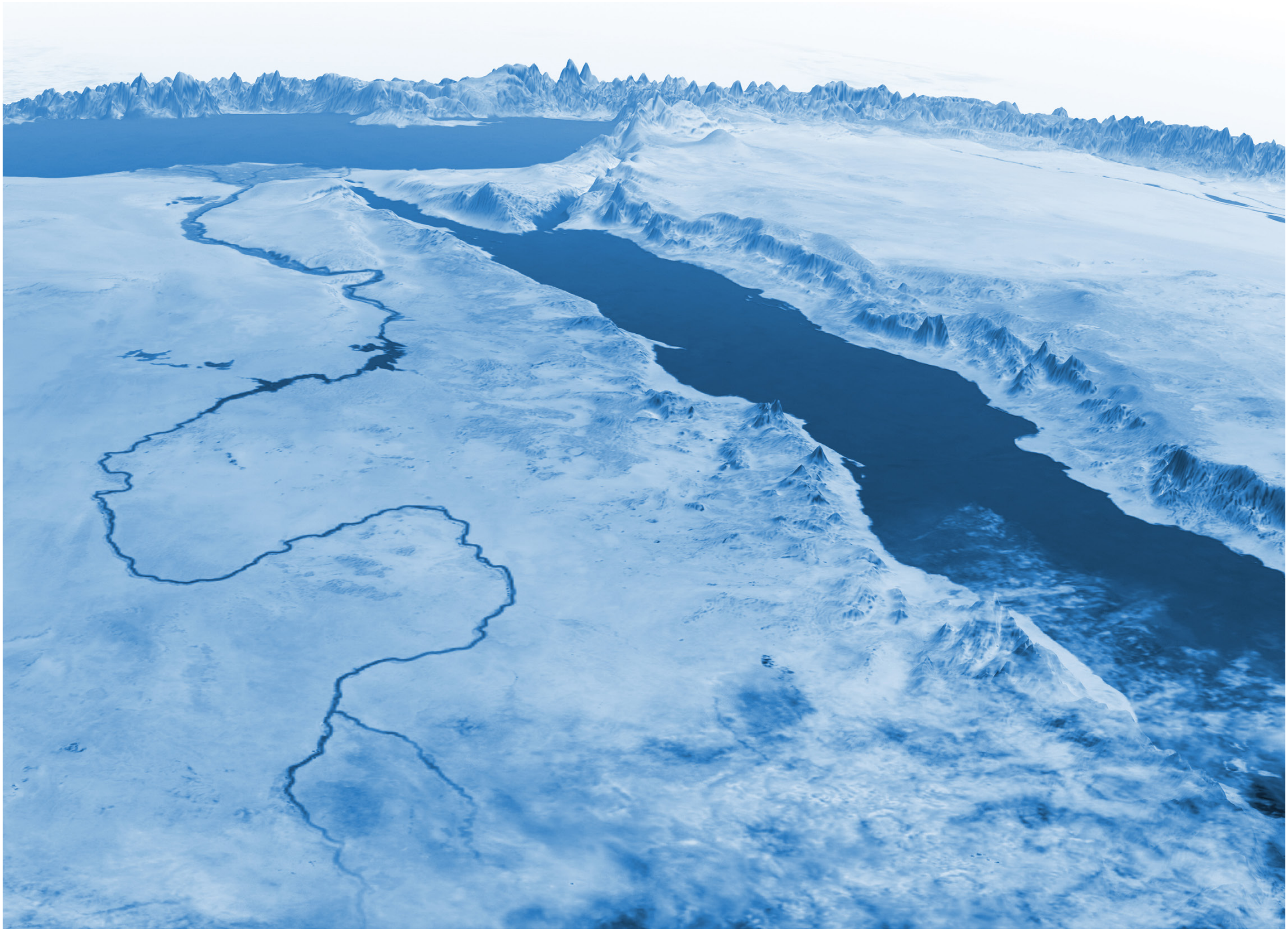


Photo: Shutterstock

ONE RIVER ONE PEOPLE ONE VISION

NILE BASIN INITIATIVE
INITIATIVE DU BASSIN DU NIL

Nile Basin Initiative Secretariat
P.O. Box 192
Entebbe - Uganda
Tel: +256 417 705 000
+256 417 705 117
Email: nbisec@nilebasin.org
Website: http://www.nilebasin.org
Facebook: /Nile Basin Initiative
Twitter: @nbiweb

Eastern Nile Technical Regional Office
Dessie Road
P.O. Box 27173-1000
Addis Ababa - Ethiopia
Tel: +251 116 461 130/32
Fax: +251 116 459 407
Email: entro@nilebasin.org
Website: http://ensap.nilebasin.org

Nile Equatorial Lakes Subsidiary Action Programme Coordination Unit
Kigali City Tower
KCT, KN 2 St, Kigali
P.O. Box 6759, Kigali Rwanda
Tel: +250 788 307 334
Fax: +250 252 580 100
Email: nelsapcu@nilebasin.org
Website: http://nelsap.nilebasin.org

NBI MEMBER STATES

Burundi | DR Congo | Egypt | Ethiopia | Kenya | Rwanda | South Sudan | The Sudan | Tanzania | Uganda

/Nile Basin Initiative @nbiweb

#NileCooperation; #NileBasin; #OneNile