Ecology and Development Series No. 34, 2005

Barnabas Akurigo Amisigo

Modelling riverflow in the Volta Basin of West Africa: A data-driven framework

Cuvillier Verlag Göttingen

To my parents for their foresight, and to my wife Hawa and daughter Mmalebna

ABSTRACT

The 400,000 km^2 Volta Basin is an international basin covering almost 28% of the West Coast. Basin of Africa. It extends from longitude 5^o 30 W to 2^o 00E and from latitude 5 30^o N to 14^o 30 N and is shared by the six West African countries Benin, Togo, Ghana, La Cote d'Ivoire, Burkina Faso and Mali. The water resources of this basin are under severe stress due to both human and natural causes. High population growth rate coupled with widespread and indiscriminate water mobilisation and use in the basin on the one hand, and high spatial and temporal variability of rainfall and high potential evapotranspiration on the other, are putting enormous pressure on the basin's water resources. As a result, there are serious water resources management problems such as flooding, water shortage, water pollution and loss of aquatic biodiversity. It is, therefore, recognised that integrated transboundary water resources management is necessary and urgent to ensure environmental integrity and sustainable water use in the basin. Streamflow modelling and prediction are essential components of any water resources management framework developed for the water-allocation and -use managers in the basin.

In this thesis, a riverflow modelling framework developed for monthly riverflow prediction in the Volta Basin is presented. By analysing available catchment rainfall, runoff and potential evapotranspiration series in the basin using methods such as correlation plots, autoregressive (AR) and autoregressive with exogenous input (ARX) modelling, it is shown that the monthly catchment rainfall-runoff process is better characterised by non-linear models.

First, a spatio-temporal linear dynamic model employing the Kalman smoother and the Expectation-Maximisation (EM) algorithm was developed and applied to filling in short gaps in daily riverflow series in the basin. This model was found to be a very good and powerful tool for filling in such data gaps.

Then, two non-linear modelling frameworks - a non-linear autoregressive and moving average with exogenous input (NARMAX) polynomial and a data-based mechanistic (DBM) modelling framework - were developed and applied to the monthly rainfall-runoff series in the basin for river catchment runoff prediction. The NARMAX model was able to capture much of the nonlinearity in the runoff generation process and provided good predictions of riverflow. However, it is a purely black-box formulation providing no physical interpretation of the runoff process in the basin. The DBM framework was very successful in representing the runoff mechanism in the basin, adequately predicting monthly river runoffs. Unlike the NARMAX models, the DBM framework is a grey-box that provided physically interpretable results at the catchment scale. Results from this modelling framework show that monthly runoff in the basin can be interpreted to occur in two pathways: a fast flow pathway and a slow, mainly delayed flow, pathway. Catchment effective rainfall in the basin was found to have a power law relationship with catchment runoff. In addition, the Identification of unit Hydrographs And Component flows from Rainfall, Evaporation and Streamflow (IHACRES) type effective rainfall-catchment wetness non-linear relationship in which the basin drying time constant is exponentially related to basin potential evapotranspiration, was found to be suitable for characterising the runoff processes in the basin.

Therefore, it is recommended that data-driven approaches be considered as the most appropriate for riverflow modelling in the Volta Basin. This is due, in part, to the fact that the approaches provide very good results that are, to some extent, physically interpretable and also because the quality, quantity and diversity of hydrological data used for riverflow modelling in the basin are too poor to enable effective use of the more elaborate distributed hydrological models.

TABLE OF CONTENTS

ABREVIATIONS AND ACRONYMS

AIC	Akaike information criterion
AR	Autoregressive
ARMA	Autoregressive and moving average
ARX	Autoregressive with exogenous input
CGIAR	Consultative Group on International Agricultural Research
DBM	Data based mechanistic model
DSS	Decision support system
EM	Expectation-Maximisation
ERR	Error reduction ratio
ETP	Potential evapotranspiration
FIS	Fixed interval smoother
GCI	Green Cross International
GEF	Global Environment Facility
GEM	General Expectation-Maximization
IWMI	International Water Management Institute
LTF	Linear transfer function
LTI	Linear time invariant
LTV	Linear time varying
LTV-SDPs	Linear time varying state dependent parameters
MSE	Mean square error
NARMAX	Non-linear autoregressive and moving average with exogenous input
NARX	Non-linear autoregressive with exogenous input
NSE	Nash-Sutcliffe efficiency
NVR	Noise variance ratio
SDP	State-dependent parameter
SISO	Single input-single output
SSG	Steady state gain
TC	Time constant
TV-SDPs	Time varying state-dependent parameters
UNEP	United Nations Environment Programme
VBRP	Volta Basin Research Project
YIC	Young information criterion
ZEF	Centre for Development Research

1 INTRODUCTION

Fresh water is increasingly becoming a scarce resource in many regions of the world due to both natural and man-made causes. Natural phenomena such as droughts cause water shortages in many areas, while floods cause pollution and degrade water sources. Often, however, it is the poor management of the resource by man that has resulted in the depletion and pollution of water bodies and made the resource less available. Over-extraction of water for domestic, agricultural, industrial and other purposes and the pollution of both surface water and groundwater from both point (industrial) and non-point (agricultural, mining) sources are major threats to water resources. For example, diversion of water from feeding rivers for irrigation, hydropower production and other purposes, and excessive pollution from agricultural chemicals and industrial and municipal wastes have contributed immensely to turning otherwise very productive water bodies such as the Aral Sea, Dead Sea and Lake Chad into environmental disasters (Micklin and Williams, 1996; Glanz, 1996; Coe and Grove, 1998; Devitt, 2001; FoEME, 1996; FoEME, 1998).

Severe water deficits can have disastrous consequences for the population of any region. The droughts in 1972, 1973, 1977 and 1982 – 1984 in the Sahel for example, caused the death of several hundred thousand people and forced millions to migrate to other less severely affected areas. In all, about 250 million people from 22 countries were affected by these droughts (ZEF, 2000).

In the Volta Basin of West Africa, there are competing demands for water use both within and among the riparian countries of the basin. This competition is mainly between industrial demands, particularly for power generation, and for agricultural water supplies, especially for irrigation. This is manifested in the numerous dams and reservoirs constructed throughout the basin for various purposes including industrial, agricultural and domestic water supplies. Thus, in Ghana, there is the world largest artificial lake, the Volta Lake, created by the dam on the Main Volta River at Akosombo for hydropower and covering 4% of the land area of the country. The over 1,000 MW of electrical power produced at this dam and the Kpong dam 100 km downstream, provide much of the electrical energy needs of the country. There are also smaller dams, particularly in the northern parts of the country, for irrigation and

domestic water supplies. The more than 600 small dams and lakes in Burkina Faso and many other similar dams in the other riparian countries provide various levels of electrical power, irrigation and domestic water supplies. The pressure on the water resources of the basin is bound to increase significantly in the future, as the high population growth rate would lead to an over 80% increase in population over the current level of about 18.6 million by the year 2025 (Water for Food, 2003). Despite the intensive and extensive use of the water resources, there is little consultation or co-operation between the countries involved in the use of these resources. In addition, rainfall in the region is erratic and unevenly distributed. Low rainfall in 1982/83 and 1997/1998, for example, saw water levels of reservoirs dropping to minimum operating levels and causing severe cuts in hydropower production and supply in Ghana in particular. The low rainfall also caused widespread crop failure, and consequently hunger, and a large part of the population in the basin suffered severe distress (Water for Food, 2003). There is, therefore, a great potential for conflicts within and between the involved countries with respect to the use of the basin's water resources, particularly in times of crisis.

Obviously, therefore, proper management of water resources is required in order to preserve and use them sustainably. Of particular concern are arid and semi-arid areas, where natural replenishment of water resources through precipitation is often inadequate or poorly distributed in space and time. Sustainable management of water resources is also urgent in areas of high population growth rates and expanding use of the resources as in the Volta Basin.

A scientifically sound decision support system (DSS) for the sustainable use of the water resources of the basin would be an important tool for the water resources managers in the basin. A key input to this DSS is the assessment of the resources in terms of quantity and distribution in space and time. This would provide information on how much water is available, where it is and when it is available.

An important indicator of the water yield of a given river catchment is streamflow; it provides information on both surface and subsurface flow processes and indicates to a large extent the level of interaction between these flow components. Analysis of the streamflows of the various river catchments in the Volta Basin could provide important insights into the level of river-aquifer interactions in the basin. In

addition, it would be possible to extract information on catchment-scale aquifer characteristics such as hydraulic conductivity and aquifer storage coefficients from long and high quality streamflow series. Unfortunately, existing streamflow series at gauging stations in the Volta Basin are short and full of gaps. In their present form it would be very difficult to extract the necessary information to enable proper assessment of the catchment response to rainfall inputs.

Therefore, a major contribution to the information needs for the water resources development and management of the Volta Basin would be the development of a modelling framework for riverflow prediction in the basin. A good modelling framework for streamflow prediction would provide strategies for:

(i) Filling in both short (a few days to a month) and long (more than a month to several continuous years) gaps in existing streamflow series
(ii) Extending flow series several years beyond their current lengths
(iii) A general assessment of the quality of existing flow series

It is the main objective of this study to develop such a modelling framework. The aim is to add to and improve upon important aspects of the water resources information in the Volta Basin. The modelling framework would also provide an important tool for the quantification in both space and time of the main variable resource in the basin, i.e., riverflow. It would provide an objective and scientific means of augmenting and extending streamflow records, which are necessary requirements for the proper assessment of the water resources of this very important basin. In undertaking this assignment, the main unit of analysis is the river catchment. The catchment will be considered a nonlinear dynamic system and the tools for the analysis of such systems, developed in systems engineering, used in a given time frame and temporal scale to model the runoff the catchment generates.

1.1 Structure of the thesis

The chapters of this thesis are stand alone chapters that are largely independent of each other. Chapter 1 begins with the main introduction to the thesis in which the philosophy

behind the study, i.e., the river catchment as a system, is presented along with the background to, research question and objectives of the study.

The study area, the 400,000 km^2 Volta Basin of West Africa, is described in Chapter 2. The high spatio-temporal rainfall and runoff variability in the basin, the poor groundwater potential, the almost indiscriminate exploitation of the water resources of the basin and the perceived water resources management problems facing the basin are highlighted. The need for transboundary co-operation between the riparian countries for the sustainable use of the basin's water resources is also highlighted.

In Chapter 3, the characteristics of the data available for the modelling activities are explored. In particular, the linear or non-linear nature of the data is examined and the modelling strategies applicable given the characteristics of the data ascertained.

The spatio-temporal state-space model is formulated, developed and applied to daily stream flow data at selected gauging stations in the basin in Chapter 4. The aim is to demonstrate the strengths of the developed model in patching small gaps of up to one month in daily stream flow data under natural conditions in the basin.

Non-linear rainfall-runoff modelling is examined more closely in Chapter 5, with the formulation, development and application of a NARMAX (Non-linear Autoregressive and Moving Average with exogenous input) polynomial model to the monthly rainfall-runoff series in the basin. Here it is demonstrated that the rainfall-runoff relationship is better described by non-linear models than by linear ones.

Chapter 6 presents the main work of this thesis – the formulation, development and application of a rainfall-runoff model, which is proposed as the most suitable for riverflow prediction in the data-poor Volta Basin. This model, a hybrid metric-conceptual, data-driven, grey-box model, is shown to provide very good predictions of monthly riverflows at gauging stations in all three principal sub-basins of the basin. In addition, it yields results that provide some physical interpretation as to the types of flow paths that runoff in the basin follows.

Finally, Chapter 7 summarizes the findings, conclusions and recommendations of the study.

1.2 Systems approach in hydrology and the river catchment as a system

Prediction of the output of a system is a major objective of any applied science such as hydrology. Streamflow is an output of the river catchment system, the prediction of which is a primary concern in hydrology. A system has both a structure and a function. Many definitions of a system emphasize either its structure or its function. The International Council on Systems Engineering (INCOSE, 2004), an international professional society for systems engineers based in the USA, defines a system as "a *construct or collection of different elements*[1] that together produce results not obtainable by the elements alone." This definition emphasizes the structure of the system, i.e. a collection of different elements. On the other hand, the following definition by Dooge (2003) emphasizes the system function. In this definition, "any structure, device, scheme, or procedure, real or abstract, that *interrelates in a given time reference, an input, cause, or stimulus of matter, energy, or information, and an output, effect, or response of information, energy, or matter*", the function of the system as interrelating an input and output is emphasized. A system can, therefore, be visualized as an input-output element as depicted in the block diagram in Figure 1.1, with input being external to the system, i.e., it influences but is not affected by the processes occurring in the system. Both input and output can be vector-valued.

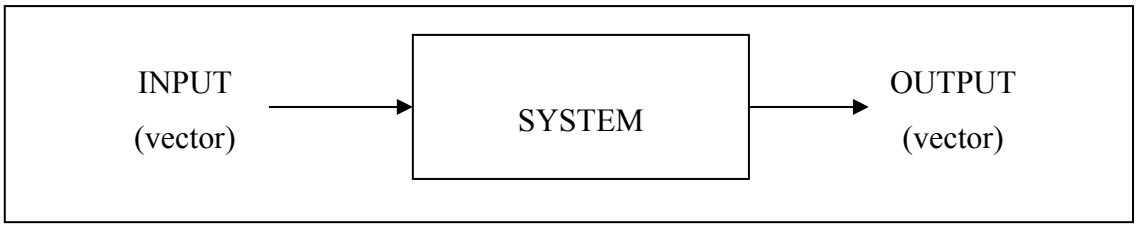


Figure 1.1 A system as a basic input-output element

A complex system may then consist of two or more subsystems each being a distinct input-output element. The subsystems in turn may be composed of components, the lowest elements in the system, each one being also a distinct input-output element. The open river catchment system shown in Figure 1.2 can be decomposed into four subsystems: surface, soil, groundwater and stream network, each with a distinct input-

[1] The emphasis here and in the following defintion are the author's.

output linkage. The soil system, for example, can further be divided into layers, each being an individual input-output element (Dooge, 2003).

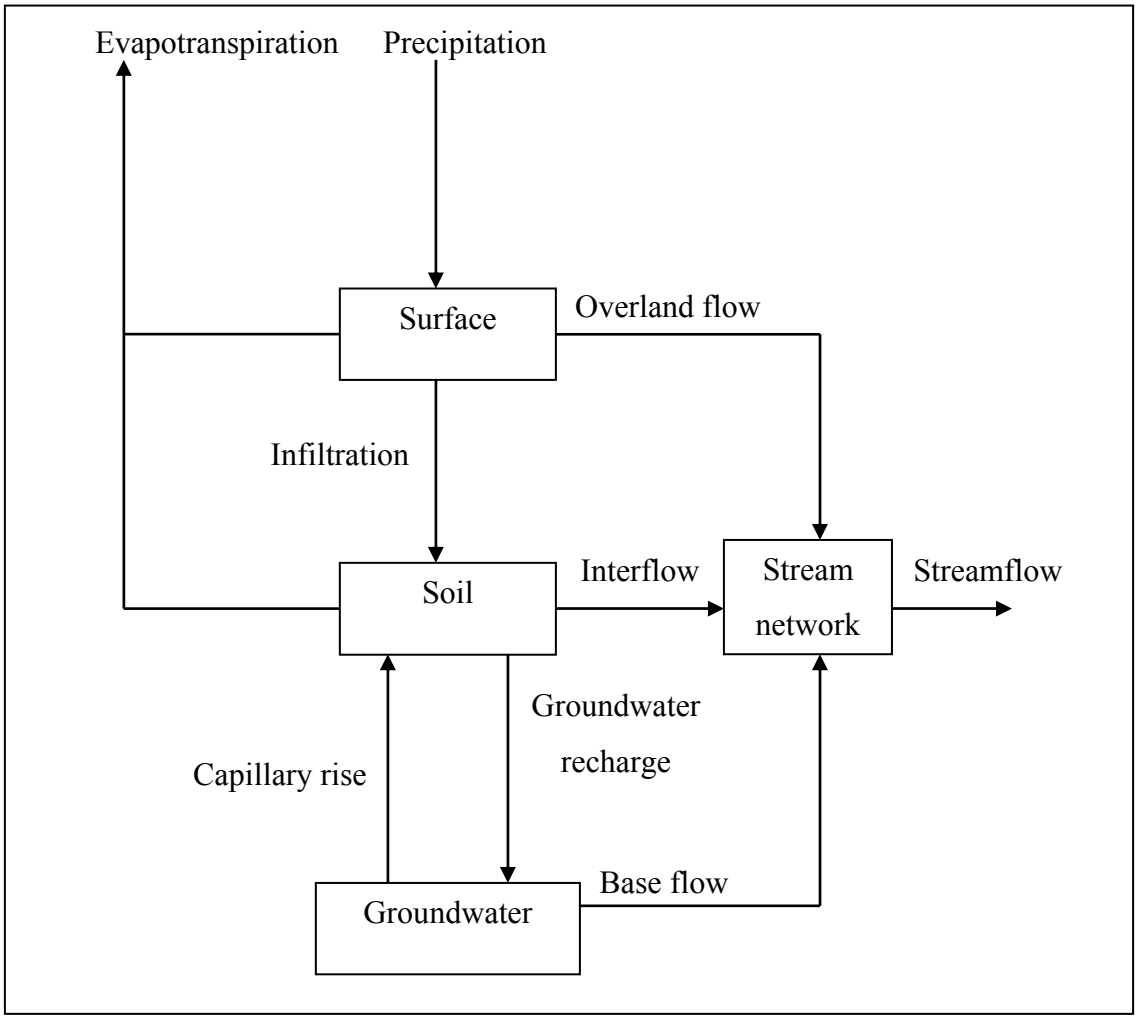


Figure 1.2 River catchment system (modified from Dooge, 2003)

A system is usually characterized by a number of variables that change with time. The state of the system at any instant is the set of values of all the variables that completely characterizes the system at that instant (Ljung, 1994; Wicox, 2002; Dooge, 2003). These variables are known as the state variables of the system. When there are direct, instantaneous relationships between the state variables, so that the state of the system at any instant does not depend on previous states, the system is termed static. In static systems, the variations in the output are dependent on only the instantaneous value of the input. If, however, the state of the system depends also on previous states, i.e., its state can change without current external input, so that the current output value

depends also on earlier input values, then the system is dynamic (Ljung, 1994). Dynamic systems, therefore, have memory or persistence which can be infinite if the current state depends on the entire past states of the system or finite if the current state depends on a fixed period of its past states. The fixed period of past states influencing the current state is also called the memory of the system (Dooge, 2003). Initial values of the state variables define the initial state of a dynamic system and enable the prediction of all future states of the system given all future input.

If the processes occurring in a system take place continuously, then that system is termed continuous, otherwise the system is discrete. When the relationship between the input and output does not depend on when the input is applied, the system is time invariant or stationary, and is time variant or unstationary otherwise. A system is also linear or nonlinear depending on whether the superposition and scaling properties apply or not. Systems are simple when they do not decompose into two or more subsystems or components, otherwise they are complex. Stable systems have bounded outputs when the inputs are bounded. A causal system is one that is not anticipative, i.e., it cannot have an output earlier than the corresponding input (Dooge, 2003).

The river catchment system is natural (inputs such as rainfall and temperature are uncontrollable), complex, nonlinear, time variant, causal, continuous (though the input-output observations are often discrete), extremely stable (rainfall often results in very attenuated runoff) and generally dynamic. However, it can be static depending on the size of the catchment, its drainage density, the climate of the region in which it is located and the input-output time scale. In arid and semi-arid regions, for example, while daily and monthly runoff from large catchments may depend also on previous daily or monthly rainfall, and the system in this case has memory, annual rainfall for previous years may not have any influence on current year annual runoff and the system at this time scale would have zero memory.

1.2.1 System identification

Prediction of the output of a system from input-output observations proceeds in two main steps. In the first step, the systems identification step, input-output observations are used to identify the input-output transformation mechanism of the system. In the prediction step, the identified mechanism is applied to new input observations to predict

the output. The system identification step is the more important and critical step. As in the definition of a system, the critical process of system identification can be undertaken with emphasis on the details of the system's structure or just its function.

Distributed models for system identification and output prediction consider in detail the nature of the system and the physical laws governing its behavior. They seek to provide output predictions for every component of the system and are generally deterministic. Such models require a very good understanding of the nature of the system - the internal workings of and connection and interaction between its subsystems and components of the subsystems, together with knowledge of the physical laws governing the processes occurring in the system - to formulate (Dooge, 2003). These physics-based models (Wheater *et al.*, 1993) are parametric (Heunecke and Welsch, 2004), as the parameters of the models are system or process parameters and are therefore physically interpretable. Distributed approaches therefore have the potential to provide the most useful and comprehensive information about the system, its nature and its functioning. However, the models have the serious drawback of being plagued with identifiability problems as a result of the very high number of parameters they usually require to be estimated from limited input-output observations (Young, 2001a). Distributed models in hydrology are based largely on the blueprint of Freeze and Harlan (1969), an example being the Systeme Hydrologique Europeen (SHE) model (Abbott *et al.*, 1986a).

The second approach to system identification is the so called black-box approach in which the details of the nature of the system and the physical laws governing the processes taking place in it are ignored completely and only its overall behavior is considered. This is the classical data-driven systems approach to system identification. It is also known as the metric approach (Wheater *et al*., 1993). As shown in Figure 1.3, the physical laws and the nature of the system are lumped together in this approach in what is called system operation (Dooge, 2003) and though, when the physical laws and/or the nature of the system change the system operation also changes, it is only the horizontal relationship shown in the figure that is considered important. These models are non-parametric in the sense that their parameters are not process parameters (Heunecke and Welsch, 2004) and so are not physically interpretable. The parameters result from the use of weighting functions in the overall input-output

transformation of the system. While the lack of physical interpretation of the parameters resulting from this systems approach combined with the inability of the approach to provide useful insights into the nature of the system being investigated may be a serious drawback, the approach often provides adequate and useful results and generally involves very few parameters to be estimated from input-output data. It thus avoids many of the identifiability problems encountered by the distributed approaches. The two polynomial models, nonlinear autoregressive with exogenous input (NARX) and nonlinear autoregressive and moving average with exogenous input (NARMAX), are models in this category that have been widely and successfully used in systems and control engineering (Chen and Billings, 1989) and to a limited extent in hydrology (Tabrizi *et al.*, 1998) to model the input-output nonlinearity in engineering and environmental systems.

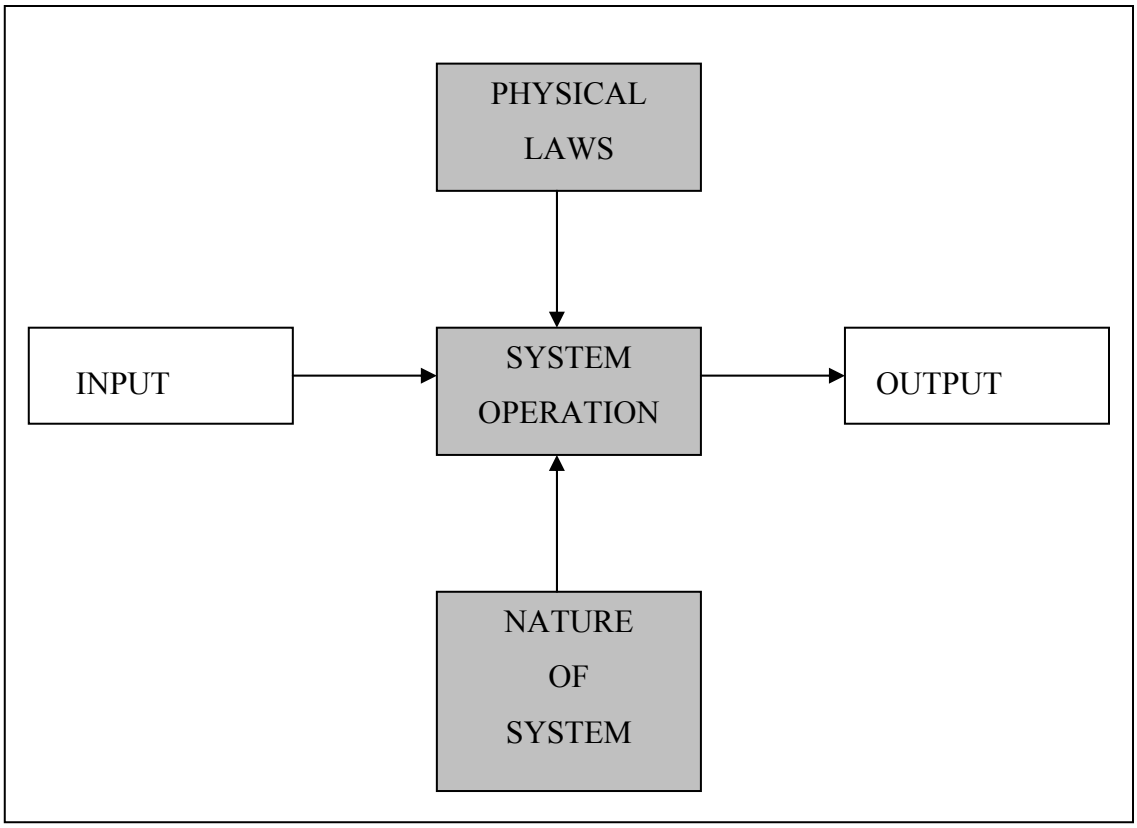


Figure 1.3 Concept of system operation (Dooge, 2003)

A third approach to system identification seeks a compromise between the fully distributed and the classical systems approaches. These are the conceptual approaches (Wheater *et al.*, 1993). They specify a priori the structure of the system

based on simplifying assumptions and the models are formulated based on the processes considered important in the system. They are therefore grey-box models with many of the parameters having physical interpretation. They generally have fewer parameters than the distributed approaches but still suffer heavily from identifiability problems. In hydrology, models in this category either use distribution functions to describe the spatial distribution of runoff such as in the TOPMODEL (Beven and Kirkby, 1979) or, as for the Stanford Watershed model (Crawford and Linsley, 1966), the runoff generation components and processes considered important in the catchment system are represented by internal storages and fluxes.

More recently, new systems approaches, the hybrid metric-conceptual approaches have been proposed for the system identification problem (Jakeman *et al*., 1990; Young, 1992). These exploit the parametric parsimony of the classical systems approaches combined with the ability of the conceptual approaches to provide physically interpretable results. These new „let the data determine the model" approaches can, therefore, be considered grey-box approaches. The Identification of unit Hydrographs And Component flows from Rainfall, Evaporation and Streamflow data (IHACRES, Jakeman *et al*., 1990) is an example of models in this category in which the structure of the catchment system (lumped) is determined a priori and the input-output data are only used to estimate the model parameters. The Data-Based Mechanistic (DBM) models (Young, 2001a), the other model type in the hybrid metric-conceptual approaches, makes no assumptions about the structure of the system but allows the data to determine both the system structure (lumped) and the parameter estimates. The hybrid metric-conceptual models usually have much fewer parameters than the conceptual models, and though the problem of identifiability is not completely eliminated in them, it is minimized. They, therefore, can be the most appropriate when the input-output observations available for system identification are limited in quantity, quality and diversity, which is often the case in river catchment rainfall-runoff modelling, particularly in the Volta Basin. Hybrid metric-conceptual models are transfer-function models based on digital filters and are particularly useful when the dynamic system is considered to have both deterministic and stochastic components; this is a concept that is very much applicable to river catchment systems that usually have noisy input-output observations.

1.2.2 Digital filters and the differential and difference equations description of dynamic systems

A real digital filter is a real-valued function, $F_t\{x\}$, which maps the real discrete input signal x (entire signal) to another real discrete output signal y_t at each sampling instant t (Smith III, 2004). Examples of real digital filters are:

$$F_t\{x\} = y_t = b_0 x_t + b_1 x_{t-1} \tag{1.1a}$$

$$F_t\{x\} = y_t = b_0 x_t + b_1 x_{t-1} + b_2 x_{t+1} \tag{1.1b}$$

$$F_t\{x\} = y_t = a_1 y_{t-1} + b_0 x_t + b_1 x_{t-1} \tag{1.1c}$$

$$F_t\{x\} = y_t = a_1 y_{t-1} + b_0 x_t + b_1 x_{t-1} + b_2 x_{t+1} \tag{1.1d}$$

$$F_t\{x\} = y_t = a_{1t} y_{t-1} + b_{0t} x_t + b_{1t} x_{t-1} \tag{1.1e}$$

$$F_t\{x\} = y_t = a_1 y_{t-1} + b_0 x_t^2 + b_1 x_{t-1}^3 \tag{1.1f}$$

where $\boldsymbol{y}$, $\boldsymbol{x}$ and all coefficients are real.

Filters 1.1a – 1.1e are all *linear filters, i.e.,* they possess both the *scaling* and *superposition* properties of linear systems. For all constant values of g and any signals ***x1*** and ***x2***, the *scaling* property of linear systems states that scaling the input of a linear system (multiplying it by a constant gain factor, g) scales the output by the same factor, i.e., $\boldsymbol{F}_t\{g\boldsymbol{x1}\}=g\boldsymbol{F}_t\{\boldsymbol{x1}\}$; the *superposition* property states that the response of a linear system to a sum of signals is the sum of the responses to each individual input signal, i.e., the input signals superimpose and do not interact ($\boldsymbol{F}_t\{\boldsymbol{x1}+\boldsymbol{x2}\} = \boldsymbol{F}_t\{\boldsymbol{x1}\} + \boldsymbol{F}_t\{\boldsymbol{x2}\}$) (Smith III, 2004). Filter 1.1f is a *nonlinear* filter (it doesn't possess the scaling and superposition properties). Filter 1.1e is a *linear time varying* filter, since the coefficients a_{1t}, b_{0t} and b_{1t} are time dependent; the rest are *time-invariant* filters, with constant coefficients. The filters in 1.1a, 1.1c, 1.1e and 1.1f are *causal filters*, as the output signal at each sampling instant does not depend on future samples of the input signal. Digital filters that involve past output samples such as 1.1c – 1.1f are *recursive* or *Infinite Impulse Response* (IIR) filters, and the past output terms are *feedback* terms. Filters without feedback are *non-recursive* or *Finite Impulse Response* (FIR) filters, i.e., the

filters have past, present and/or future input or *feed forward* terms only (Smith III, 2004).

In the mathematical modelling of a dynamic system, the relationships between the variables (input, output and internal or state variables) of the system are often described by differential equations for continuous representations and difference equations for discrete representations. These equations can be described in two ways: The inputs (external variables) are directly related to the outputs in one equation, or the inputs and outputs are related indirectly through a number of internal or state variables by means of a system of first-order differential or difference equations. The former description is termed external, while the latter is internal, also called state space (Ljung and Glad, 1994).

1.2.3 External model description

For the external model, the single differential equation relating the vector-valued input (***u***(t)) and output (***y***(t)) can be expressed as (Ljung and Glad, 1994):

$$g(y^{(n)}(t),y^{(n-1)}(t),...,y(t),u^{(m)}(t),u^{(m-1)}(t),...,u(t)) = 0 \tag{1.2}$$

where $y^{(k)}(t) = \frac{d^k}{dt^k} y(t)$

and ***g***(.,.,...,.) is an arbitrary, vector-valued, nonlinear function.

For linear ***g***, equation 1.1 can be written as:

$$\begin{aligned} a_n \frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \ldots + a_1 \frac{dy}{dt} + a_0 y = \\ b_m \frac{d^m u}{dt^m} + b_{m-1} \frac{d^{m-1} u}{dt^{m-1}} + \ldots + b_1 \frac{du}{dt} + b_0 u \end{aligned} \tag{1.3}$$

or in difference equation form:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + ... + a_n y_{t-n} + b_0 u_t + b_1 u_{t-1} + ... + b_m u_{t-m} \tag{1.4}$$

or, in the case of stochastic modelling,

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_n y_{t-n} + b_0 u_t + b_1 u_{t-1} + \ldots + b_m u_{t-m} + \varepsilon_t \qquad (1.5)$$

where u_t, y_t = values of u and y at time t,

a_0, a_1,.., a_n, b_0, b_1, ..., b_m are time invariant if the system is stationary and time variant otherwise, and ε_t is a random error that is the source of stochasticity in the system. The *differential* and *difference equations* 1.3, 1.4 and 1.5 are linear digital filter formulas for computing an output sample of a signal at time t based on past and present input samples of the input signal and past output samples.

The processes occurring in environmental systems, such as hydrological systems, are continuous. However, inputs, outputs and other observed variables of these systems are normally available as discrete time series. In modelling these systems, therefore, difference equations such as in equation 1.4 for deterministic and 1.5 for stochastic modelling are often used.

1.2.4 State- space representation of the general models of dynamic systems

When the outputs of the system are modeled as indirectly related to the inputs, internal state variables are introduced. Suppose $\boldsymbol{y_t} = (y_{1t}, y_{2t}, ..., y_{pt})'$ is the p x 1 output vector, $\boldsymbol{u_t} = (u_{1t}, u_{2t}, .., u_{mt})'$ the m x 1 input vector, and $\mathbf{x_t} = (x_{1t}, x_{2t}, .., x_{nt})'$ the n x 1 state vector of the system, which unambiguously define the state of the system at time instant t, and that the prime denotes transpose of the vectors. Then the following equations (ignoring stochastic components) constitute the internal model of the system (Ljung and Glad, 1994):

$$\begin{aligned}
\dot{x}_{1t} &= f_1(x_{1t}, \ldots, x_{nt}, u_{1t}, \ldots, u_{mt}) \\
\dot{x}_{2t} &= f_2(x_{1t}, \ldots, x_{nt}, u_{1t}, \ldots, u_{mt}) \\
&\cdot \\
&\cdot \\
&\cdot \\
\dot{x}_{nt} &= f_n(x_{1t}, \ldots, x_{nt}, u_{1t}, \ldots, u_{mt})
\end{aligned} \qquad (1.6a)$$

and

$$\begin{aligned} y_{1t} &= h_1\left(x_{1t},\ldots,x_{nt},u_{1t},\ldots,u_{mt}\right) \\ y_{2t} &= h_2\left(x_{1t},\ldots,x_{nt},u_{1t},\ldots,u_{mt}\right) \\ &\cdot \\ &\cdot \\ &\cdot \\ y_{pt} &= h_n\left(x_{1t},\ldots,x_{nt},u_{1t},\ldots,u_{mt}\right) \end{aligned} \tag{1.6b}$$

where $f_i(.,...,.)$ and $h_i(.,...,.)$ are generally nonlinear functions of the n+m components of variables $\boldsymbol{x}$ and $\boldsymbol{u}$.

In more compact form, equations 1.6 can be rewritten as:

$$\dot{x}_t = f\left(x_t,u_t\right) \tag{1.7a}$$

$$y_t = h\left(x_t,u_t\right) \tag{1.7b}$$

where $f(x_t,u_t) = (f_1(x_t,u_t),f_2(x_t,u_t),...,f_n(x_t,u_t))$ and $h(x_t,u_t) = (h_1(x_t,u_t),h_2(x_t,u_t),...,h_p(x_t,u_t))$ are n x 1 and p x 1 functions, respectively, that are generally nonlinear.

The discrete time equations corresponding to 1.7 are:

$$x_{t+1} = f\left(x_t,u_t\right) \tag{1.8a}$$

$$y_t = h\left(x_t,u_t\right) \tag{1.8b}$$

If $f(x_t,u_t)$ is continuously differentiable and u_t is a piecewise continuous function, then for all $t > t_0$, the initial time, and for a given initial state, $x_{t0} = x_0$, there always exists a unique solution to 1.7 (and 1.8) (Ljung and Glad, 1994). Models 1.7 and 1.8 are state-space models of order n, the dimension of the state vector x_t.

For the discrete time linear stochastic case, the state-space model is given as:

$$x_{t+1} = Fx_t + Bu_t + \omega_t \tag{1.9a}$$

$$y_t = Hx_t + Du_t + \upsilon_t \tag{1.9b}$$

where F, H, B and C are, respectively, n x n, p x n, n x m and p x m system matrices; and ω_t and υ_t are vector-valued random noise terms. When the input, output and state vectors are also spatially distributed, the state-space model would be a spatio-temporal model and is useful in environmental systems modelling. A form of spatio-temporal state-space model is employed in Chapter 4 in filling short gaps in daily riverflow series at a gauging station using flow series from neighboring gauging stations.

The external and internal model representation of a dynamic system are illustrated in figure 1.4

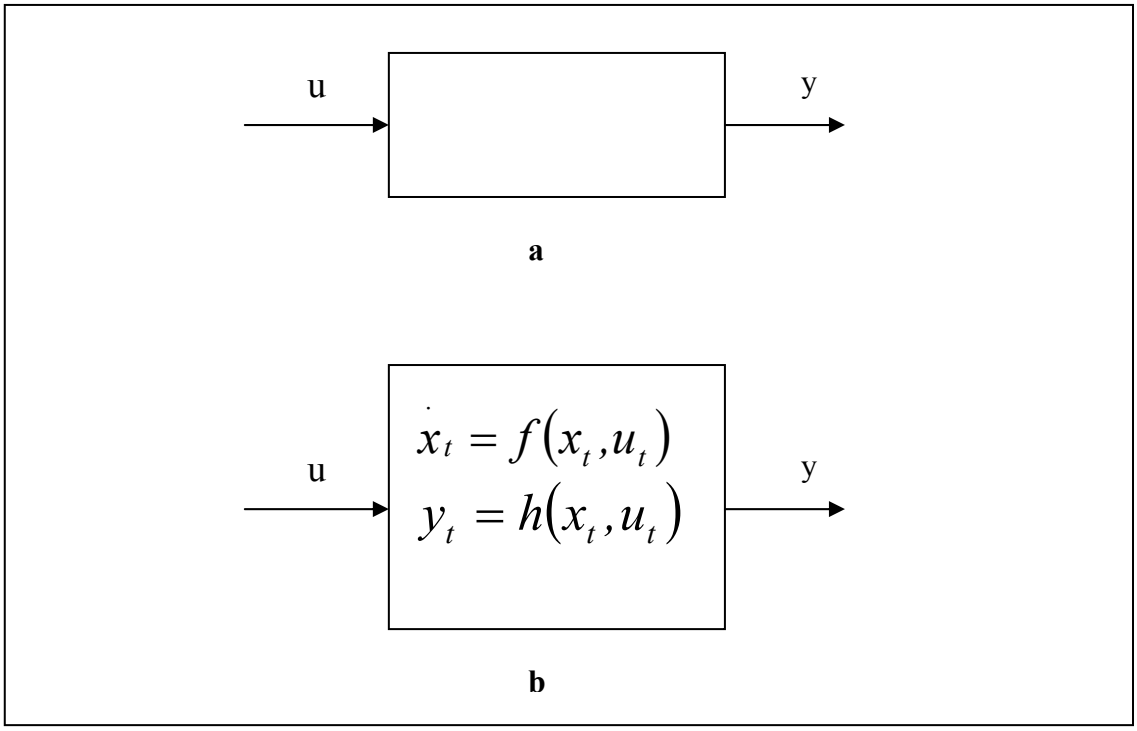


Figure 1.4 (a) External model, and (b) internal model (Ljung and Glad, 1994)

1.2.5 Transfer function (TF) models of dynamic stochastic systems

If z^{-1} is the backward shift operator such that $z^{-i}u_t = u_{t-i}$, then Equation 1.4 can be re-written as:

$$y_t = -a_1\left(z^{-1}y_t\right) - a_2\left(z^{-2}y_t\right) - \ldots - a_n\left(z^{-n}y_t\right) + b_0u_t + b_1\left(z^{-1}u_{t-1}\right) + \ldots + b_m\left(z^{-m}u_{t-m}\right) \quad (1.10a)$$

$$\Rightarrow y_t + \left(a_1z^{-1} + a_2z^{-2} + \ldots + a_nz^{-n}\right)y_t = \left(b_0 + b_1z^{-1} + \ldots + b_mz^{-m}\right)u_t \quad (1.10b)$$

$$\Rightarrow A\left(z^{-1}\right)y_t = B\left(z^{-1}\right)u_t \quad (1.10c)$$

$$\Rightarrow y_t = \frac{B\left(z^{-1}\right)}{A\left(z^{-1}\right)}u_t \quad (1.10d)$$

where

$$\mathrm{A}\left(\mathrm{z}^{-1}\right) = \left(1 + \mathrm{a}_1\mathrm{z}^{-1} + \mathrm{a}_2\mathrm{z}^{-2} + \ldots + \mathrm{a}_\mathrm{n}\mathrm{z}^{-\mathrm{n}}\right)$$

$$\mathrm{B}\left(\mathrm{z}^{-1}\right) = \left(\mathrm{b}_0 + \mathrm{b}_1\mathrm{z}^{-1} + \ldots + \mathrm{b}_\mathrm{m}\mathrm{z}^{-\mathrm{m}}\right)$$

Equation 1.10d is the Linear Transfer Function (LTF) representation of the input-output digital filter. The ratio $\frac{\mathrm{B}\left(\mathrm{z}^{-1}\right)}{\mathrm{A}\left(\mathrm{z}^{-1}\right)} = \mathrm{h}\left(\mathrm{z}^{-1}\right)$ is the transfer function. Stochasticity can be accounted for by adding a noise term to Equation 1.10d, so that the general stochastic linear transfer model for the dynamic system is then given as:

$$y = \frac{B\left(z^{-1}\right)}{A\left(z^{-1}\right)}u_t + \varepsilon_t \quad (1.11)$$

where ε_t is a noise term, which is the source of stochasticity in the output.

The feed-back and feed-forward coefficients in A and B in the LTF can either be time variant or invariant. As noted earlier, the Linear Time Invariant (LTI) transfer function characterizes stationary systems while the Linear Time Varying (LTV) function characterizes non-stationary systems. If the parameters of the LTV transfer function are allowed to be state dependent so that they vary as rapidly as the variation in the states, then the transfer function can characterize nonlinear systems (Young, 2001a).

This is the form of the transfer function used in the hybrid metric-conceptual modelling of the river catchment rainfall-runoff transformation process in Chapter 6.

1.3 Background to the study

The six countries within the Volta Basin have weak economies with low Gross National Incomes (GNI) and high debt to GNI ratios (Table 1.1). They are, therefore, classified as low-income countries by the World Bank (World Bank, 2004). Population growth rate in the basin is high (Table 1.1) with the total population estimated to reach 34 million by 2025 from the current level of about 18.6 million (GEF, 2003a). The high population growth rate coupled with the general widespread poverty in the region will result in enormous pressure on the natural resources, including water resources. Domestic and industrial water demand is projected to increase about 300% by 2025 due to the rapid population increase and the expected industrial expansion (GEF, 2003a). The problem is exacerbated by both high rainfall variability and uncertainty resulting in emphasis being placed more and more on irrigated rather than rain-fed agriculture in all the countries in the basin. For example, irrigation water demand is projected to increase by nearly 540 and 710% for Ghana and Burkina Faso, respectively, by 2025 (GEF, 2003a). Already, numerous dams and reservoirs of various sizes have been created in the basin to mobilize water for various purposes.

Table 1.1 National population and growth rates, incomes and indebtedness of the six riparian countries of the Volta Basin (Data Source: World Bank, 2004; UN, 2005). (GNI=Gross National Income)

Country	**Population (2002) million**	**Annual population growth rate (1980-2005)[a] %**	**GNI (2002) billion $**	**GNI/capita (2002) $**	**Total external debt (2002) million $**	**Debt as % of GNI (2002)[b]**
Benin	7	2.9	2.5	380	1,843	74
Burkina Faso	12	2.4	2.6	220	1,580	61
La Cote d'Ivoire	17	3.2	10.3	610	11,816	115
Ghana	20	2.8	5.4	270	7,338	136
Mali	11	2.5	2.8	240	2,803	100
Togo	5	2.9	1.3	270	1,581	122

[a] *UN (2005)*
[b] *Own computations using data in 4th and 6th columns*

The riparian countries do recognize the likely overexploitation of the natural resources in the region. Consequently, each country has established one national agency or another to regulate the use of the resources within its boundaries and to ensure national environmental integrity and sustainable water development and use.

Most of the water resources development projects in each country have been undertaken with little or no consultation with the other riparian countries. Currently, there are no formal institutional arrangements for managing the water resources of the basin nor are there any legal provisions for cooperation among the riparian countries for integrated multipurpose development and management of the shared water resources. However, this is beginning to change, as the countries recognize that the best way to sustainable water use and environmental integrity maintenance in the basin is through transboundary co-operation and consultation.

Thus in 1998, Ghana proposed an initiative on integrated ecosystems management of the basin. This resulted in the formation of the Volta River Basin Project (VRBP) involving all six riparian countries and financed by a grant from the Global Environment Facility (GEF) of UNEP from 1999 to 2002, following an inter-ministerial workshop held in Accra, Ghana, in 1999. At this workshop, the Accra Volta River Basin Declaration was adopted. In the declaration, the six riparian countries agreed to join forces and actively collaborate to achieve the following objectives (GEF, 2003a):

- The formulation of a strategic action plan for the Volta River Basin
- The formulation of a framework agreement of co-operation between the Basin States for the integrated management of the Volta River Basin
- The formulation of an agreed programme with a holistic vision for the integrated management of the Basin

Since this workshop, a flurry of activities involving concerned actors both from within the region and outside it have taken place to facilitate and promote collaboration between the riparian states for integrated land and water management of the basin. One of such initiatives resulted in the project “Addressing Transboundary Concerns in the Volta River Basin and its Downstream Coastal Area”, a project

formulated by representatives and endorsed by the governments of the six riparian countries with funding from the GEF and scheduled for execution from October 2003 to October 2007 (GEF, 2003b). According to GEF (2003b), "the global environmental objective of this project is to enhance the capacity of the countries to plan and manage the Volta catchment areas within their territories and aquatic resources and ecosystems on a sustainable basis." The following three project areas will be supported:

1. Build capacity and create a regional institutional framework for the effective management of the Volta Basin;
2. Develop regional policy, legal and regulatory frameworks for addressing transboundary concerns in the Volta Basin and its downstream coastal areas;
3. Initiate national and regional measures to combat transboundary environmental degradation in the Volta Basin.

Another initiative in the region is that by Green Cross International (GCI) through its Water for Peace project. The main aim of the project is "the prevention of conflicts and the promotion of dialogue and cooperation on the water and land resources of the Volta Basin. The project focuses on ensuring the fuller involvement of civil society in the development of transboundary basin management agreements, institutions and strategies."

In May 2004, in Accra, Ghana, the International Water Management Institute (IWMI) launched an 18-million dollar 15-year Water and Food Programme in the Volta Basin (Ghanaweb, 2004). To be financed under the Consultative Group on International Agricultural Research (CGIAR) Challenge Programme on Water and Food, the programme has 11 projects relevant to the basin. The first phase of five years is "expected to create research-based knowledge and methods for growing more food with less water, and develop a transparent framework for setting targets and monitoring progress." (Ghanaweb, 2004).

These and other such programs initiated in and for the benefit of the basin seek to position the people, the institutions and governments of the riparian countries to manage the natural resources of the Volta Basin in a sustainable and environmentally friendly manner for the benefit of the people and biodiversity in the basin. To achieve these laudable goals, support systems, i.e., scientifically based decision support systems, will be required. The Center for Development Research (ZEF) of Bonn University,

Germany, is undertaking the development of one such system, in the Sustainable Water Use under Changing Land Use, Rainfall reliability and Water Demands in the Volta Basin (Glowa-Volta) Project. Under this project, started in 2000, ZEF is coordinating and participating in research in the Volta Basin aimed at "*the development of a scientifically sound decision-support system for the assessment, sustainable use and development of water resources in the Volta Basin based on the analysis of the physical and socio-economic determinants of hydrologic cycles*" (ZEF, 2000). This decision support system (DSS) is to be made up of a set of dynamic models, which capture all first-order linkages between relevant processes in the atmosphere, soil and water and which readily interchange information with the correct scale and format. The project is interdisciplinary and is being undertaken in collaboration with other German and international institutions including the CSIR-Water Research Institute of Ghana.

The research is being conducted in three research clusters, each involving researchers of different disciplines. Each cluster consists of several subprograms geared towards addressing specific research questions handled in the cluster. One of these clusters is the Water Use cluster, which deals with five research questions being addressed in the subprojects runoff and hydraulic routing, integrated economic-hydrological optimization, health and water, communal and household water supply, and institutional analysis. This cluster will provide the necessary information for the optimal allocation of the available water to the various social, economic, agricultural and industrial sectors in the basin. The research question of the runoff and hydraulic routing subproject is:

What is the water availability over time throughout the Volta River network?

The research activities undertaken in this study seek to provide some of the important information required to answer this research question.

1.4 Research question and objectives

The research question in this study is:

What modelling framework is suitable for riverflow prediction in the Volta Basin?

To provide answers to the above research question, the following objectives were set:

(i) Formulate, develop and apply a suitable model for filling in short gaps in daily riverflows at gauging stations in the basin

(ii) Formulate, develop and apply a suitable rainfall-runoff model for predicting natural monthly riverflows in the basin.

2 THE STUDY AREA

2.1 Introduction

The Volta Basin is an international basin spanning six countries in West Africa. The FAO (FAO, 1997) groups it with the rest of the river basins draining to the sea from Senegal to Nigeria in one region called West Coast (Figure 2.1), a region that covers 4.7% of the African continent and spreads over 13 countries (FAO, 1997). The Volta basin covers almost 28% of the West Coast, extends from longitude 5^{o} 30 W to 2^{o} 00 E and from latitude 5 30^{o} N to 14^{o} 30 N and is shared by Benin, Togo, Ghana, La Cote d'Ivoire, Burkina Faso and Mali. However, most of the basin, about 85% of the total area of nearly 400,000 km^2, lies in Burkina Faso and Ghana. Table 2.1 shows the areas of each country covered by the basin.

According to the World Bank (2004) classification of economies, all the countries in the Volta Basin are Low Income Counties (LICs) with 2002 GNI per capita of less than $735. The vast majority of the population lives on primarily rain-fed agriculture, but irrigation is becoming increasingly important in the basin.

Table 2.1 Volta Basin areas by country (FAO, 1997)

Country	**Total area (km^2)**	**Area within the basin (km^2)**	**As % of total area of basin (%)**	**As % of total area of country (%)**
Mali	1,240,190	9,496	2.4	0.8
Burkina Faso	274,000	183,000	46.4	66.8
Benin	112,620	16,000	4.1	14.2
Togo	56,785	26,700	6.8	47.0
Côte d'Ivoire	322,462	7,000	1.8	2.2
Ghana	238,540	152,000	38.6	63.7
For Volta Basin		394,196	100.0	

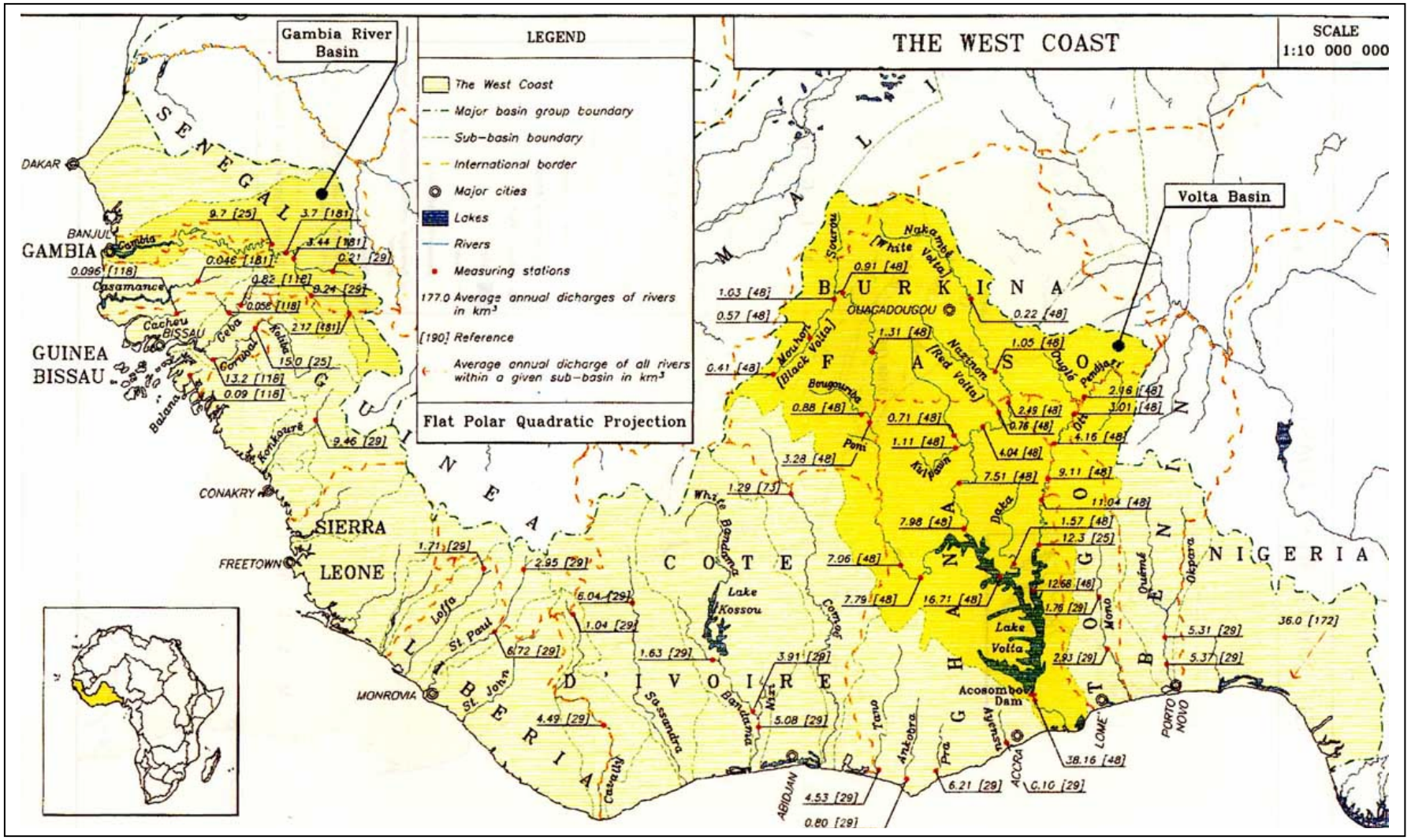


Figure. 2.1 Map of the West Coast showing the Volta Basin (FAO, 1997)

2.2 Climate

The climate of the Volta Basin, as for the rest of West Africa, is controlled by the south-north-south movement of the Inter Tropical Convergence Zone, ITCZ. The ITCZ is the belt into which the harmattan, the hot, dry and dusty tropical continental air mass from the Sahara to the north and the monsoon, the warm, moist tropical maritime air mass from the Atlantic to the south of the region, converge. The oscillation of the ITCZ produces two marked seasons in the region – the wet and dry seasons. In its complex oscillation across the region, the belt passes certain areas, particularly the lower latitudes, twice, giving the affected areas two rainy seasons. The basin can thus be divided into 3 main climatic zones: humid southern zone with two distinct rainy seasons, tropical transition zone with two rainy seasons close to each other, and tropical northern zone, covering most of the basin. This zone has one rainy season lasting from April to October, with rainfall peaking in September, and one dry season from November to March.

2.2.1 Rainfall

Rainfall is highly variable both spatially and temporally. It increases from north to south with mean annual values ranging from less than 500 mm in the extreme north to more than 1600 mm in the forested regions of the basin to the south in Ghana (Figure 2.2; MWH, 1998(2)). Over 70% of the annual total rainfall occurs in the three months of July, August and September with little or no rainfall in the months November – March in most of the basin.

2.2.2 Potential evapotranspiration

Potential evapotranspiration in the basin varies both spatially and temporally with an annual mean varying from 2500 mm in the north of the basin to 1800 mm in the coastal zone. Mean monthly potential evaportranspiration exceeds mean monthly rainfall for most of the year for the entire basin.

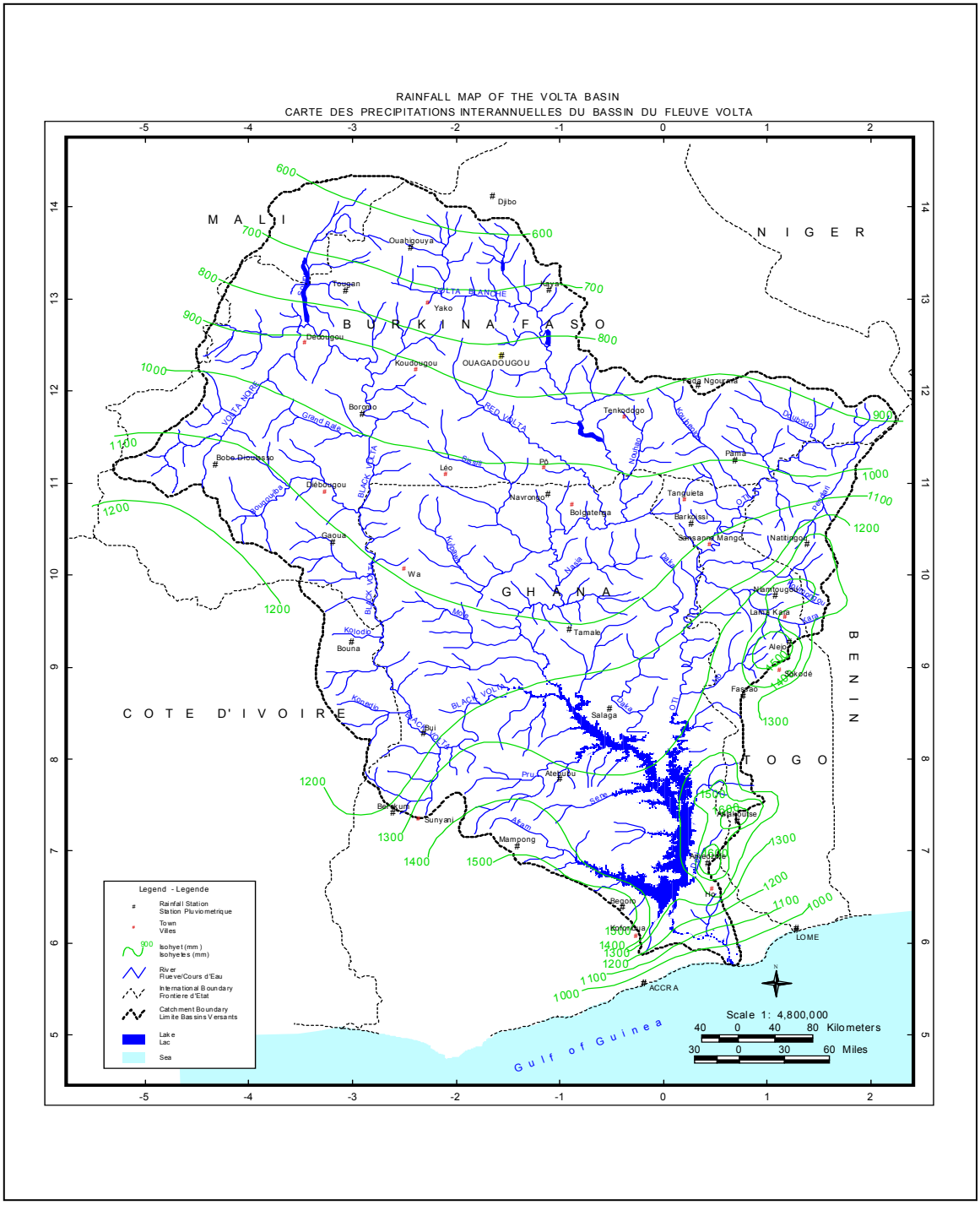


Figure. 2.2 Annual rainfall in the Volta Basin (VBRP, 2002)

2.3 Land cover and use

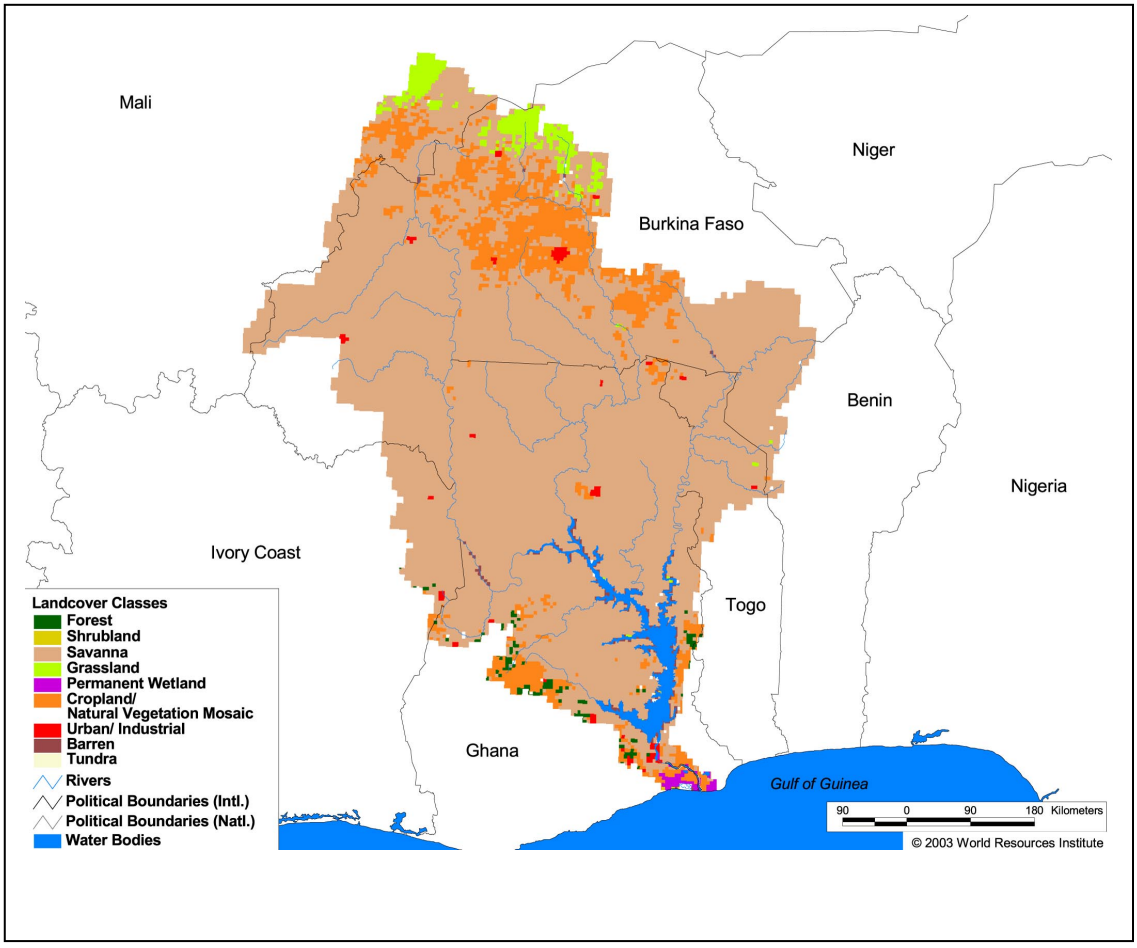


Figure 2.3 Land cover and use types in the Volta Basin (WRI, 2003)

As shown in Figure 2.3, the predominant land cover type in the Volta Basin is savanna. Table 2.2 below shows the relative areas occupied by each land cover use type in the basin:

Table 2.2 Relative coverage of each Land Cover and Use type (WRI, 2003)

Land cover and use type	**Percent of basin area**
Forest cover	0.7
Grassland, savanna and shrubland	85.6
Wetlands	10.4
Irrigated cropland	0.1
Dryland area	91.7
Urban and industrial area	0.5
Loss of original forest cover	96.6

2.4 Hydrology

2.4.1 Drainage

The basin is drained by numerous streams, most of which dry up in the dry season. These streams can be grouped into four main river systems formed from the four main rivers draining the basin. These rivers are the Black Volta (Mouhoun), White Volta (Nakanbe), Main Volta and Oti (Pendjari) Rivers (Figure 2.4). The Main Volta River is formed by the joining of the Black and White Volta Rivers and is joined further downstream by the Oti.

The Oti River starts out as the Pendjari River, with its source in northwest Benin. It first flows northeast, meanders west and then southwest along the boarder between Benin and Burkina Faso, continues its southwest flow through northern Togo forming, for a short while, the boarder between Togo and Burkina Faso and then flows south, entering Ghana as the Oti and forming part of the northern boarder between Togo and Ghana. It continues its southward flow until it joins the Main Volta River. Main tributaries of this river include the Doubodo and Koulpeolge, which originate from southeastern Burkina Faso; the Koumangou and Kara Rivers originating from northwestern Benin and flowing through northern Togo to join the Oti in Ghana; and the Mio River from northern Togo.

The White Volta River and its main tributary the Red Volta (Nazinon), have their source in north and central Burkina Faso, respectively. The White Volta flows south, then east and then southeast, flowing almost parallel to the Red Volta. Both then flow south when they enter Ghana, with the White Volta turning west to be joined by the Red Volta. The White Volta River then continues westwards through northern Ghana and then turns south where it is joined by tributaries Kulpawn, Nasia and Mole.

The Black Volta River originates from southwestern Burkina Faso. It flows northeast from its source, turns southeast and then south to Ghana. On entering Ghana, it continues its southward flow, forming part of the boarders between Burkina Faso and Ghana and then La Cote d'Ivoire and Ghana. It flows southeast after leaving the La Cote d'Ivoire-Ghana boarder, turns north, then east and then southeast joining the White Volta River to form the Main Volta. Important tributaries of the Black Volta River include the Grand Bale and the Bourgouriba, both of which originate in western

Burkina Faso, and the Tain River, which has it source in the western part of Ghana and joins the main river downstream of Bui.

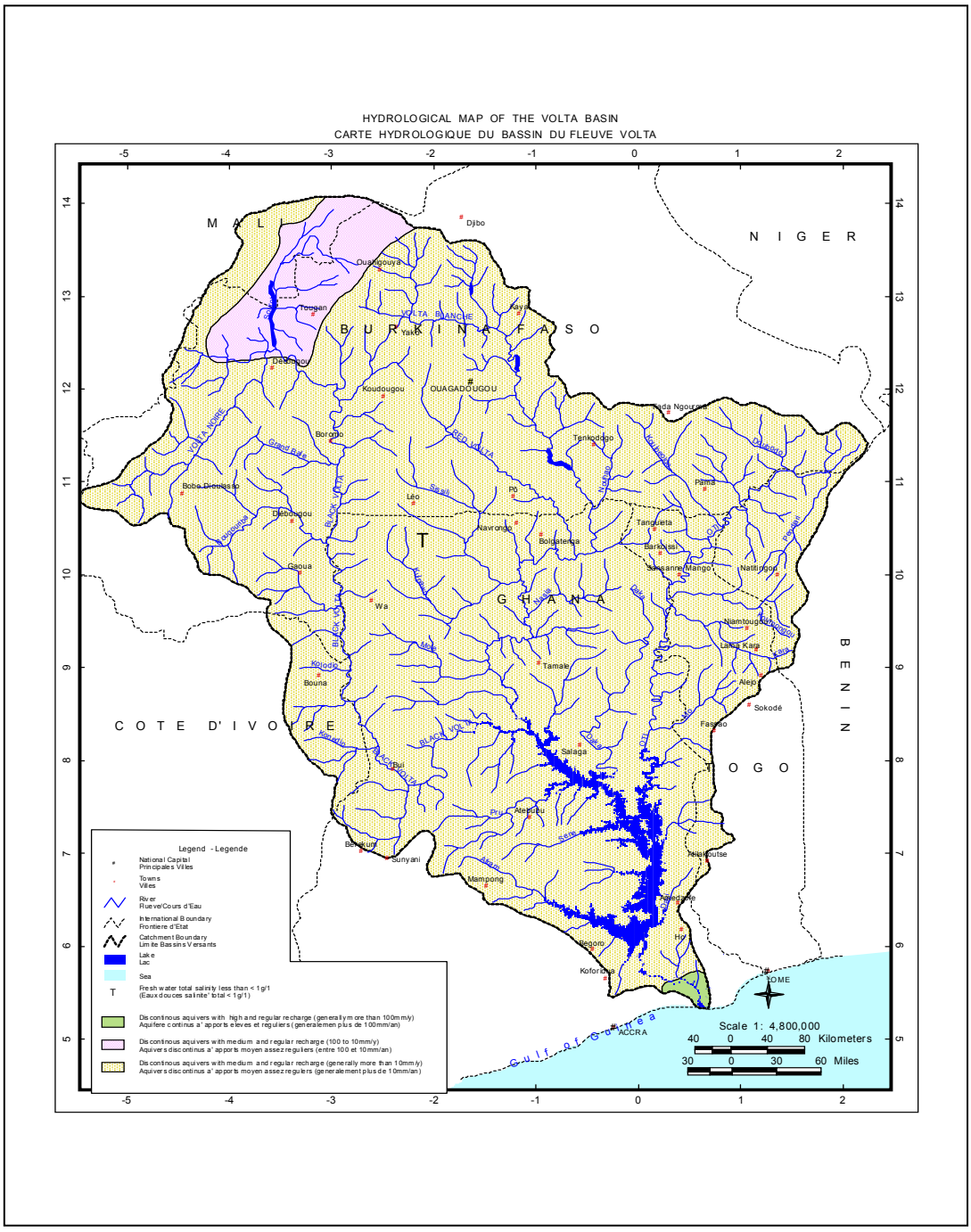


Figure. 2.4 The Drainage system of the Volta Basin (VBRP, 2002)

The Main Volta River is formed from the joining of the Black and White Volta Rivers and is joined further downstream by the Oti. Its main tributaries include the

Daka, Pru, Sene and Afram, all of which, together with the main river, flow entirely within Ghana.

The most prominent landmark of hydrological importance in the basin is the Volta Lake, formed from the hydropower dam built on the Main Volta River in southeastern Ghana. With a surface area of 8500 km^2 and a capacity of 148 km^3 (FAO, 1997), it is one of the largest man-made lakes in the world.

2.4.2 Stream flow distribution

Mean annual streamflow at various sections of the main rivers and some of their tributaries are shown in Figure. 2.1. For the Black Volta River, the mean annual runoff close to its source in western Burkina Faso is just above 0.4 km^3. The flow rises more than 3 fold just before entering Ghana and on entering Ghana the annual flow has increased to 8 times its value at the source of the river. By the time the flow leaves the Ghana-La Cote d'Ivoire boarder, it is close to its maximum of just under 7.8 $km^3\ yr^{-1}$ near the confluence with the White Volta. The mean annual flow of the White Volta starts at a little above 0.2 km^3 downstream of its source in northern Burkina Faso, increases to about 2.2 km^3 on entering Ghana and then to just over 4.0 km^3 downstream of the confluence with the Red Volta. The river joins the Black Volta at a slightly higher annual flow of just under 8 km^3. The Pendjari River attains an annual flow of nearly 2.2 km^3 before turning into the Oti, when its annual flow then reaches about 3.0 km^3 along the short Togo-Burkina Faso boarder. The flow enters Ghana with nearly 4.2 $km^3\ yr^{-1}$ and by the time it leaves the Togo-Ghana boarder it has increased to a little over 11.0 $km^3\ yr^{-1}$. It joins the Main Volta River at nearly 12.7 $km^3\ yr^{-1}$, more than one and half times the annual flow of the Black or White Volta Rivers at their confluence. Below the Akosombo dam, the controlled annual discharge of the Volta is about 38.2 km^3.

2.5 Hydrogeology

2.5.1 Geology

The basin overlies two main geological systems: a main Precambrian platform and a sedimentary layer (Figure 2.5). The igneous and metamorphic rocks of the Precambrian platform form part of the West African shield and consist primarily of granites, schists and basic rocks (van der Sommen and Geirnaert, 1988; MWH, 1998(2)). The

predominant rocks in these formations are the Dahomian, Togo series, Birimian and the folded series of the Tarkwaian (MWH, 1998(2)). Panafrican tectonic movements have subjected this crystalline basement to faulting and fracturing (van der Sommen and Geirnaert, 1988). The second geological system is less important in the basin as a whole but underlies a substantial area in Ghana. It is represented by the Voltaian system (underlying about 45% of all of Ghana, Dapaah-Siakwan and Gyau-Boakye (2000)) and recent formations found in the lateritic deposits in the northern parts of the basin. There are also recent alluvial deposits forming narrow bands along the main rivers of the basin.

2.5.2 Groundwater occurrence and flow

Two main aquifer systems occur in the basin – those developed in the weathered mantel of high porosity but low permeability (reservoir type) and the ones developed in the fractured bedrock of low porosity but high permeability (conductive type). Due to the absence of primary porosity in the crystalline basement complex and voltaian system, groundwater occurrence in much of the basin is associated with the development of secondary porosity such as from joints, fractures, shears and fissures (MWH, 1998(1), Dapaah-Siakwan and Boakye, 2000). It is, therefore, widely believed that the aquifer systems in the basin are highly discontinuous with individual compartments in which isolated groundwater circulation occurs. However, van der Sommen and Geirnaert (1988), citing hydrogeological studies in Ghana, Burkina Faso and Niger, hypothesize a regional groundwater recharge-discharge system in the basement complex that increases in importance from north to south in tandem with increasing rainfall. Since mean monthly potential evaportranspiration exceeds mean monthly rainfall for most of the year for the entire basin, groundwater loss in the dry season would likely be more to evapotranspiration than to baseflow to rivers. Baseflow is thought, therefore, to be insignificant in the basin.

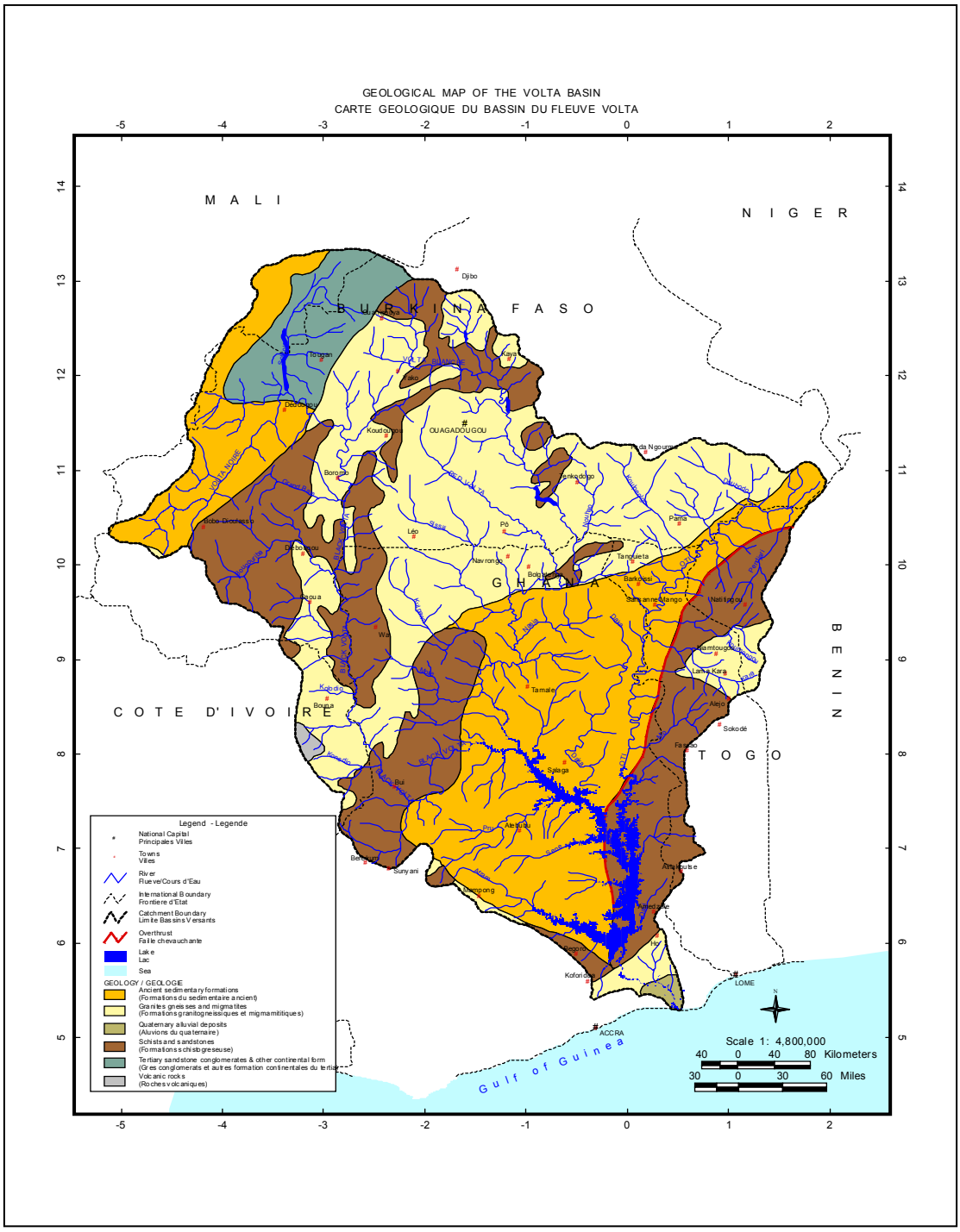


Figure 2.5 Geological Map of the Volta Basin (VBRP, 2002)

2.5.3 Borehole yields

Borehole yields in the basin are rather small – the mean yield being 2.1 – 5.7 m^3h^{-1} (MWH, 1998(1)). Van der Sommen and Geirnaert (1988) quote 2.5 m^3h^{-1} as the mean borehole yield in Burkina Faso with about 10% of the boreholes there yielding above 5 m^3h^{-1}. These mean yield values were obtained from analyses of borehole yields from

drilling programs that had concentrated largely on hand-operated wells developed mainly for rural water supply. In these programs, drilling was usually stopped at depths where the yields were found adequate for the rural communities served. These boreholes are, therefore, generally partially penetrating. Thus, while the low mean yields reported in the literature suggest that most of the aquifers in the basin are low yielding, it should be recognized that the yields reported in the literature were obtained from partially penetrating wells that may not have exploited the full potential of the aquifers involved. Nonetheless, given the general unfavorable conditions for groundwater flow and storage in the basin due to the rather poor geology, aquifer yields are not expected to be much higher than the available estimates.

2.6 Water use in the basin

The main water uses in the basin are hydropower, irrigation and domestic water supply. There are numerous infrastructural developments in the riparian countries of the basin for the mobilization particularly of surface water for various purposes. Important infrastructure includes (GCI, 2003):

- The Sourou works, Burkina Faso ($300x10^6$ m^3)
- The Ziga dam, Burkina Faso ($200x10^6$ m^3)
- The Kompienga dam, Burkina Faso ($2\ 050x10^6$ m^3)
- The Bagré dam, Burkina Faso ($1{,}700x10^6$ m^3)
- The Akosombo dam, Ghana ($150{,}000x10^6$ m^3)

In addition, thousands of boreholes have been drilled in the basin mainly for domestic water supply. Hydropower, particularly from the huge Volta Lake in Ghana, is by far the biggest water user in the basin. The hydropower production from the Akosombo dam and the much smaller Kpong dam downstream exceeds 1,000 megawatts and is the main energy source for Ghana (MWH, 1998(2)).

Surface water and groundwater are exploited independently for various purposes throughout the basin. No attention is paid to the effect of the consumptive use of one on the other. This attitude has developed because the two water systems are regarded as separate resources with little or no interaction between them. Groundwater recharge is considered to be largely from excess rainfall, and groundwater flow to rivers

in the basin is widely regarded as insignificant (van der Sommen and Geirnaert, 1988; MWH, 1998(1)).

2.7 Water resources management problems in the basin

The water resources management problems in the basin arise from two main sources. On the one hand, there is undue pressure on the resources from expanding populations and industrial activities coupled with the high rainfall variability and uncertainty in the basin. The result is largely water shortage or rather unavailability of water for large sections of the population, crop failure and the inability of hydraulic structures to function as designed due to reduced water levels. There is also pressure on other natural resources, resulting in land degradation, pollution of water bodies and environmental degradation in general. On the other hand, the numerous dams and reservoirs for water mobilisation in the basin result in reduced downstream flows and modification of downstream streamflow in general. Streamflow-dependent livelihoods of downstream populations and aquatic life are thus threatened. Floods that occur frequently in parts of the basin cause physical havoc and facilitate the spread of pollution to water bodies and other areas.

Stakeholders in the basin have identified twelve key water resources management problems in the basin urgently requiring attention. These are presented in Table 2.3. The rankings of the problems as perceived by each of the six riparian countries are shown in Table 2.4. This table shows that, for Ghana, the downstream country of the basin, all except 2 of the problems are high priority with none being low or no priority.

Table 2.4 also shows that flooding is of serious concern only in Burkina Faso and Ghana. Diminishing water resources, loss of biodiversity, water-borne diseases, inadequate/lack of information dissemination mechanisms and inadequate institutional and legal framework for basin management are the 5 top problems that are considered serious in all 6 countries.

Table 2.3 Key environmental problems and their causes as identified by technical representatives of the six riparian countries of the Volta Basin in 1999 (GEFa, 2003)

Problem	Cause
Upstream	
Diminishing water resources	Increased demands and increased pollution (which reduces availability or increases the cost of the available polluted waters); altered hydrology from changes in land use which affect runoff/infiltration.
Hydrological changes	Changes in land use that may affect runoff and infiltration patterns as well as sedimentation of canals/rivers, which may reduce hydraulic efficiency.
Soil erosion	Deforestation; bush fires; overgrazing; nomadism and human migration; uncontrolled human settlements along riverbanks and eroded soils resulting in decreased water quality.
Downstream	
Pollution	Dumping of human, domestic and industrial waste into water courses; leaching of agro-chemicals into rivers; salt water intrusion; oil spillage; waste from mining activities; use of agrochemicals in fishing.
Coastal erosion	Inadequate flow of sediments to the coast due to physical development within the basin.
Coastal pollution	Transport of pollutants to the coastal zone.
Flooding	Uncontrolled spilling from reservoirs; inadequate/lack of early warning systems; intense precipitation at short intervals; loss of wetlands.
Basin wide	
Loss of biodiversity	Deforestation; pollution; overexploitation of natural resources by humans and their livestock (overgrazing); changes in flow regimes downstream of dams; inundation of reservoir areas; dams as barriers.
Aquatic weeds	Introduction, deliberate in the case of florists, accidental in the case of fishermen and others, of exotic aquatic plants; the problem is exacerbated by increased nutrient availability, which promotes explosive growth, from both organic and inorganic sources.
Water-borne diseases	Creation of dams or impoundments; changes in flow regimes; contamination of water bodies with human waste and pollution; infestation of water bodies with aquatic weeds, habitat for hosts and vectors.
Inadequate / lack of information dissemination mechanisms	Inadequate resources (financial/human) for information gathering; absence of a regional mechanism for gathering information from member countries.
Inadequate institutional and legal framework for basin management	Lack of an enabling political environment

Table 2.4 Ranking of key environmental problems of the Volta River Basin by country and national priority (GEFa, 2003)

Environmental Problem	**BN**	**BF**	**CI**	**GH**	**ML**	**TL**
Diminishing water resources	3	3	2	3	3	3
Hydrological changes	2	2	2	2	2	2
Soil erosion	2	3	2	2	3	2
Pollution	2	2	3	3	2	2
Coastal erosion	3	1	3	3	1	3
Coastal pollution	3	1	2	3	1	3
Flooding	1	3	1	3	1	1
Loss of biodiversity	3	3	3	3	3	3
Aquatic weeds	2	3	3	3	2	1
Water-borne diseases	3	3	3	3	3	3
Inadequate/lack of information	3	3	3	3	3	3
Poor institutions/legal framework	3	3	3	3	3	3

1 = Low or no priority; 2 = Medium priority; 3 = High priority
BN = Benin, BF = Burkina Faso, CI = La Cote d'Ivoire, GH = Ghana, ML = Mali
TL = Togo

In order to effectively tackle these problems and to avoid future conflicts within and between the countries in the basin, there is the need for the countries to co-operate and collaborate in the integrated development of the water resources of the basin.

3 EXPLORATORY DATA ANALYSIS

3.1 Introduction

In this chapter, the data used for the rainfall-runoff modelling activities to be presented in chapters 5 and 6 are explored. The aim is to ascertain the characteristics of the data and determine, qualitatively, the level of persistence in the river runoff and the degree of nonlinearity in the rainfall-rainfall relationship. This would facilitate the choice of modelling framework suitable for the data. Two model-selection criteria – the complete K-fold cross validation and the Akaike information criteria – are also examined in this chapter in order to select the one most convenient for the non-linear model selection in later chapters.

The data used were selected from catchment monthly rainfall and corresponding streamflow and potential evapotranspiration (PET) series compiled and quality controlled by Taylor (2003) for river gauging stations in the Black and White Volta and Oti sub-basins. Selected stations for this study (Table 3.1 and Figure 3.1) were those with fairly natural flows and without gaps within their data series. This has resulted in very short flow series with almost none of the recent droughts captured. As Table 3.1 shows, most of the stations are in the Black Volta sub-basin. This is because the runoff data for the stations in this sub-basin have relatively fewer gaps and are of relatively better quality than those for the other two sub-basins. Also, since flows in the main White Volta River are highly controlled due to several hydro and irrigation dams upstream, particularly in Burkina Faso, no gauging stations on this river were selected. In general, runoff data are more complete and of better quality for gauging stations on the principal rivers of the Black and White Volta and Oti than for those on their tributaries.

At the annual scale (with runoff aggregated from the monthly series), there appears to be a distinct wet and dry period in the flow series as shown in Figure 3.2 using 4 of the stations with the longest flow series. The wet period can be taken as that up to 1971 and the dry period from 1972. Separate annual means for the wet and dry periods for these stations are presented in Table 3.2. A first check on the quality of the runoff data was a comparison of the computed mean annual runoffs with available values in the literature for the stations. Table 3.3 summarises the comparison between

mean annual runoff derived from the series used in this study and the FAO (1997) mean annual runoff available for most of the stations.

Table 3.1 Selected characteristics of gauging stations used in the study (Data Source: Taylor, 2003)

Station	**River**	**Co-ordinates (decimal degrees)**		**Drainage area (km^2)**	**Period of runoff Series**
		Longitude	**Latitude**		
Black Volta sub-Basin					
1. Banzo	Black Volta	4.80 W	11.32 N	3,024	1956-1978
2.Nwokuy	Black Volta	3.50 W	12.52 N	12,094	1956-1987
3.Manimenso	Black Volta	3.40 W	12.75 N	21,124	1956-1983
4.Tenado	Black Volta	2.80 W	12.17 N	24,086	1977-1985
5. Boromo	Black Volta	2.90 W	11.90 N	48,078	1955-1988
6. Debougou	Bougouriba	3.10 W	10.93 N	15,140	1963-1981
7.Lawra	Black Volta	2.90 W	10.60 N	96,000	1951-1973
8. Dapola	Black Volta	2.90 W	10.57 N	96,437	1951-1990
9. Bui	Black Volta	2.10 W	8.20 N	111,853	1954-1971
10. Bamboi	Black Volta	1.90 W	8.15 N	134,200	1951-1975
White Volta sub-Basin					
11. Wiasi	Sissili	1.30 W	10.33 N	12,105	1962-1973
12. Yagaba	Kulpawn	1.2 W	10.10 N	9,100	1958-1972
13. Nasia	Nasia	0.75 W	10.10 N	6,070	1969-1975
14. Nabogo	Nabogo	0.80 W	9.70 N	3,040	1963-1974
Oti sub-Basin					
15. Porga	Oti	0.90 E	11.05 N	27,197	1952-1984
16. Mango	Oti	0.40E	10.30 N	36,287	1953-1973
17. Koumangou	Koumangou	0.40 E	10.20N	6,070	1959-1973
18. Sabari	Oti	0.20 E	9.28 N	72,775	1960-1973

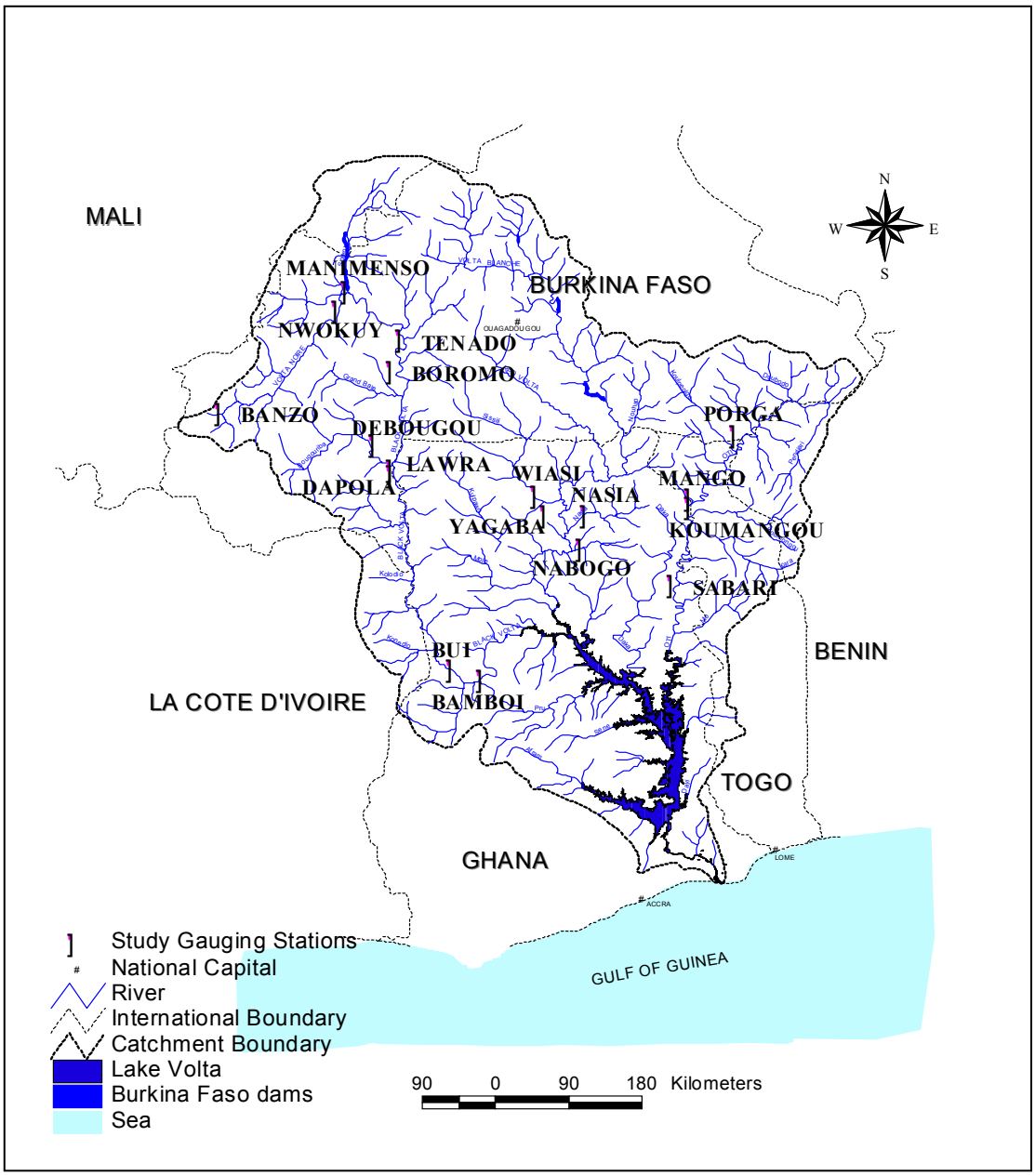


Figure 3.1 Map of the Volta Basin showing the gauging stations used in the study

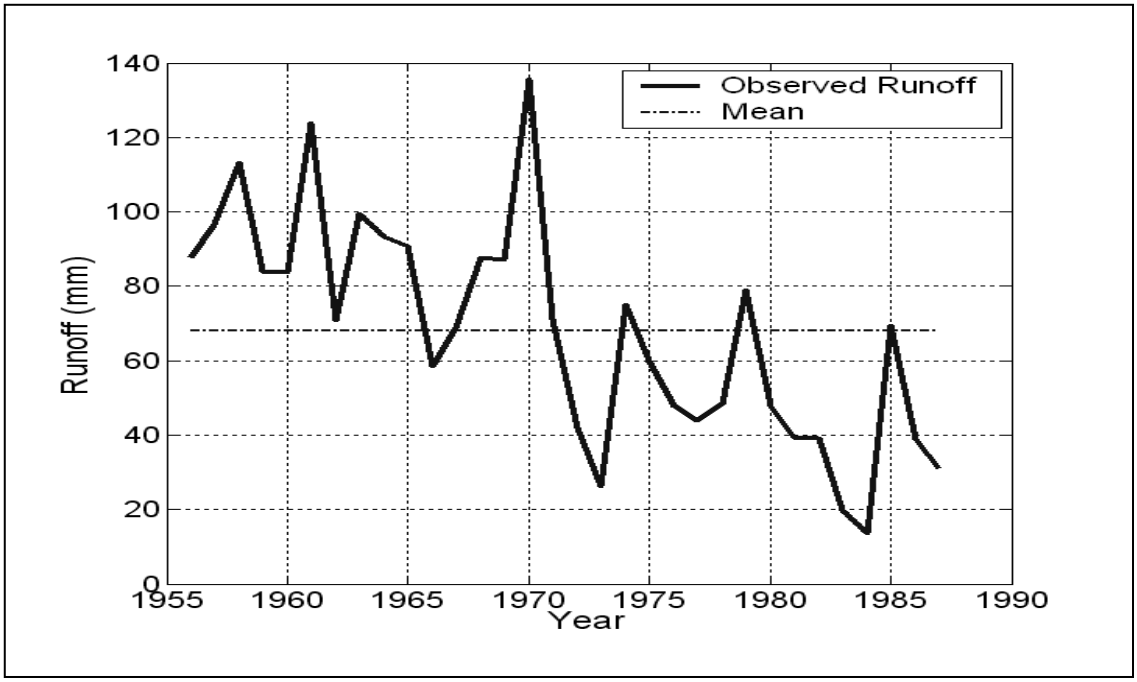


Figure 3.2a Annual hydrograph for Nwokuy on the Black Volta River

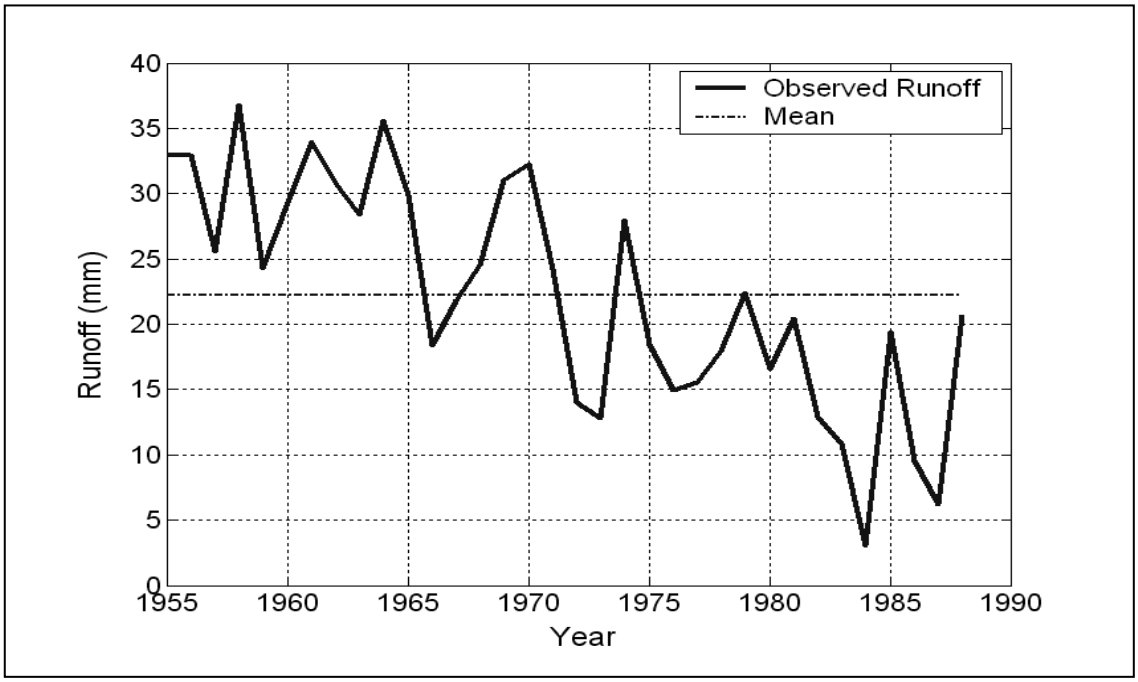


Figure 3.2b Annual hydrograph for Boromo on the Black Volta River

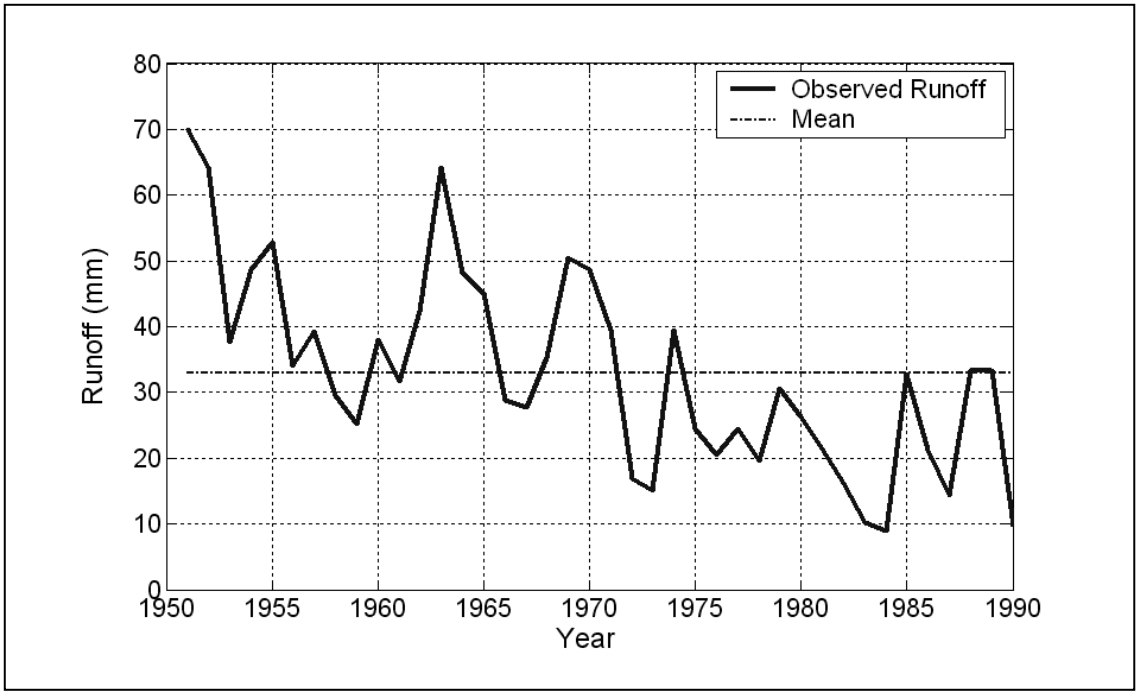


Figure 3.2c Annual hydrograph for Dapola on the Black Volta River

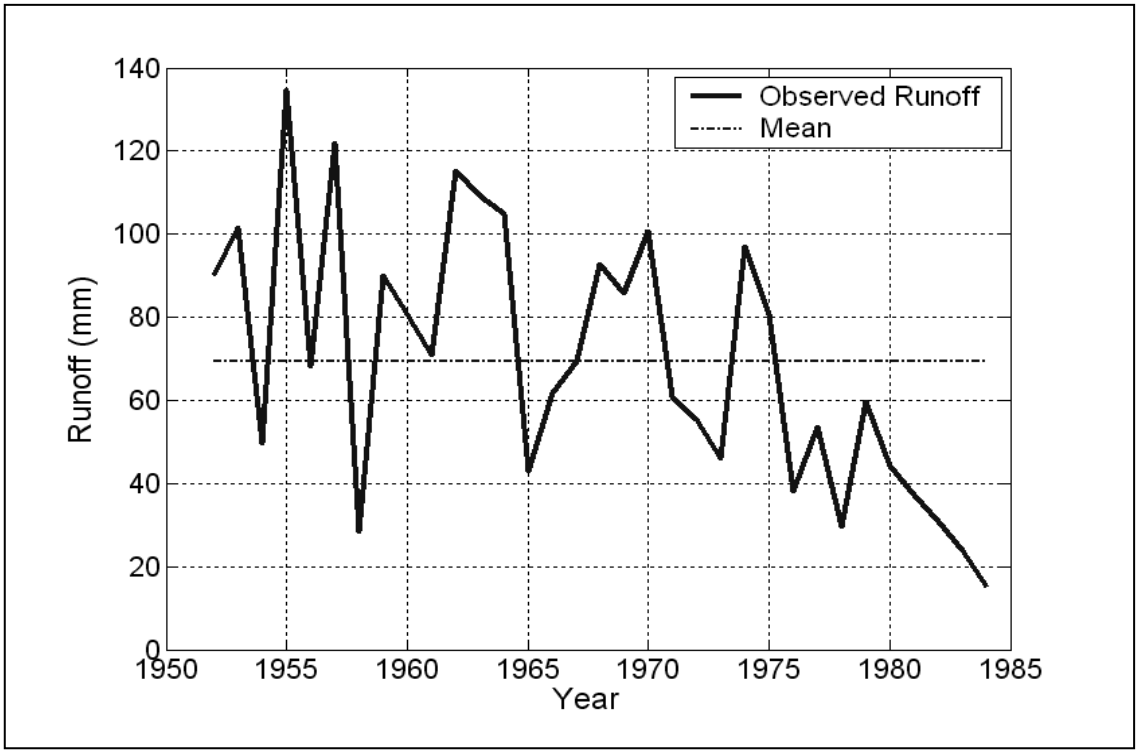


Figure 3.2d Annual hydrograph for Porga on the Oti River

Table 3.2 Mean annual flows for wet and dry periods for 4 of the stations with the longest flow series.

Period	Mean annual flow (km^3)			
	Nwokuy (Black Volta)	**Boromo (Black Volta)**	**Dapola (Black Volta)**	**Porga (Oti)**
Up to 1971 (wet)	1.09	1.39	4.13	2.28
From 1972 (dry)	0.54	0.74	2.14	1.28
Entire period	0.82	1.07	3.18	1.89

It can be observed that the FAO (1997) mean annual flows for these stations fall within the values for the wet and dry periods (Table 3.2). The discrepancies in the two sets of mean flow values presented in Table 3.3 may, therefore, be due largely to differences in the flow periods used in the two computations, though the flow periods of the data used in the computations of the means in the FAO (1997) case are not known. Therefore, the river flow data used in this study appear to represent the actual flows in the various catchments for the periods for which the have been compiled and, except for the rather short durations, should be good enough for the analysis undertaken in this study.

Each set of rainfall, runoff and ETP series at a gauging station was partitioned into estimation and validation series. These were used for model selection and validation purposes for the Autoregressive and Moving Average (ARMA) and Autoregressive with exogenous (ARX) models used in later sections of this chapter and for the models presented in chapters 5 and 6. This partitioning was done in such a way as to obtain estimation series long enough to be able to account for the variations in the validation series as much as possible.

Table 3.3 Comparison of mean annual flows at the gauging stations from FAO (1997) and computed from the monthly flows used in this study (NA = Not available)

Station	**Mean annual flow (km^3 yr^{-1})**	
	FAO (1997)	**Used in this study**
Black Volta Basin		
1. Banzo	0.41	0.37
2.Nwokuy	1.03	0.82
3.Manimenso	0.91	0.77
4.Tenado	NA	0.58
5. Boromo	1.31	1.07
6. Debougou	0.88	0.79
7.Lawra	NA	3.63
8. Dapola	3.28	3.18
9. Bui	7.08	7.29
10. Bamboi	7.79	8.20
White Volta sub-Basin		
11. Wiasi	0.71	0.77
12. Yagaba	1.11	1.15
13. Nasia	NA	0.80
14. Nabogo	NA	0.51
Oti sub-Basin		
15. Porga	2.16	1.89
16. Mango	4.16	4.32
17. Koumangou	NA	1.94
18. Sabari	11.04	11.28

3.2 Rainfall-runoff characteristics of the monthly riverflow data

Figures 3.3a and 3.3b show plots of total monthly catchment rainfall and corresponding total monthly catchment runoff and potential evapotranspiration (ETP) for Bamboi on the Black Volta River and for Sabari on the Oti River. The plots for the remaining stations are shown in figures 9.1i – 9.1xvi in the appendix. They indicate more variation in monthly runoff than in monthly rainfall and potential evapotranspiration for all stations (the series are compiled for the hydrological year March – February, as adopted by the Hydrological Services Department of Ghana). This observation is supported by the coefficient of variation (cv) values for each month listed in Table 3.4 for Bamboi and Sabari and in tables 9.1i – 9.1xvi in the appendix for the rest of the stations. These tables summarize important monthly rainfall, runoff and ETP characteristics for the stations. Plots of the coefficient of variation for each month for the three series as in figures 3.4a – 3.4d for representative stations in the 3 sub-basins show clearly the differences in the variation between the series. For the rainy season, when rainfall is expected and widespread, the monthly cv for rainfall is less than the corresponding cv for runoff. In the dry season, when rainfall is not expected and isolated, streamflow variation is less than rainfall variation. The variation in the potential evapotranspiration, in comparison, is much less significant. The difference in the variations of the rainfall and runoff (and also between runoff and potential evapotranspiration) suggests nonlinearity in their relationship (Adreini ***et al***., 2000).

The temporal distribution of catchment rainfall in the basin on the monthly scale is illustrated in figures 3.5a – 3.5c. These figures are plots of monthly rainfall as percent of total catchment rainfall for the entire rainfall series. In the case of the Black Volta Basin, the stations have been split into upstream (all in Burkina Faso) and downstream (2 in Burkina Faso) stations. It is observed from the plots that over 70% of the total rainfall in a year occurs in the 4 months June – September. This has important implications for water resources development and management in the basin, since this temporal variability in rainfall implies both flooding in some months and water shortage in others in much of the basin. Little wonder then these two have been identified as major water resources management problems by technical teams of the riparian countries (GEFa, 2003).

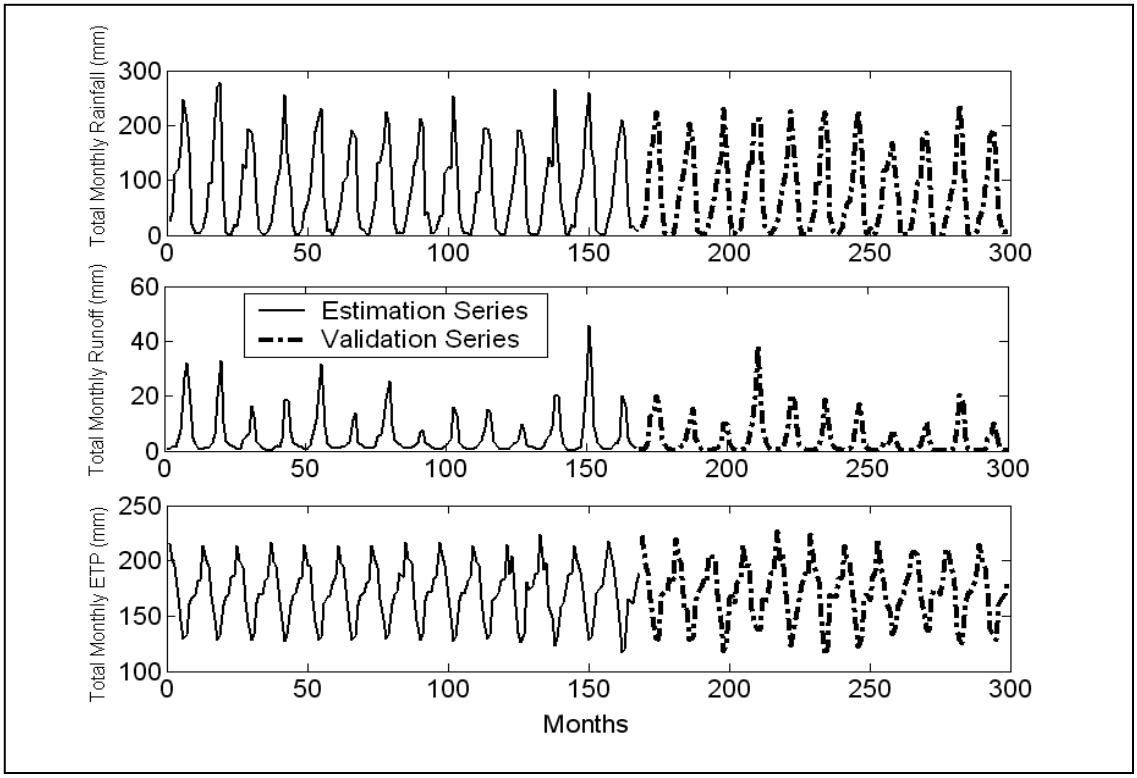


Figure 3.3a Observed monthly rainfall, runoff and ETP for Bamboi, Black Volta River

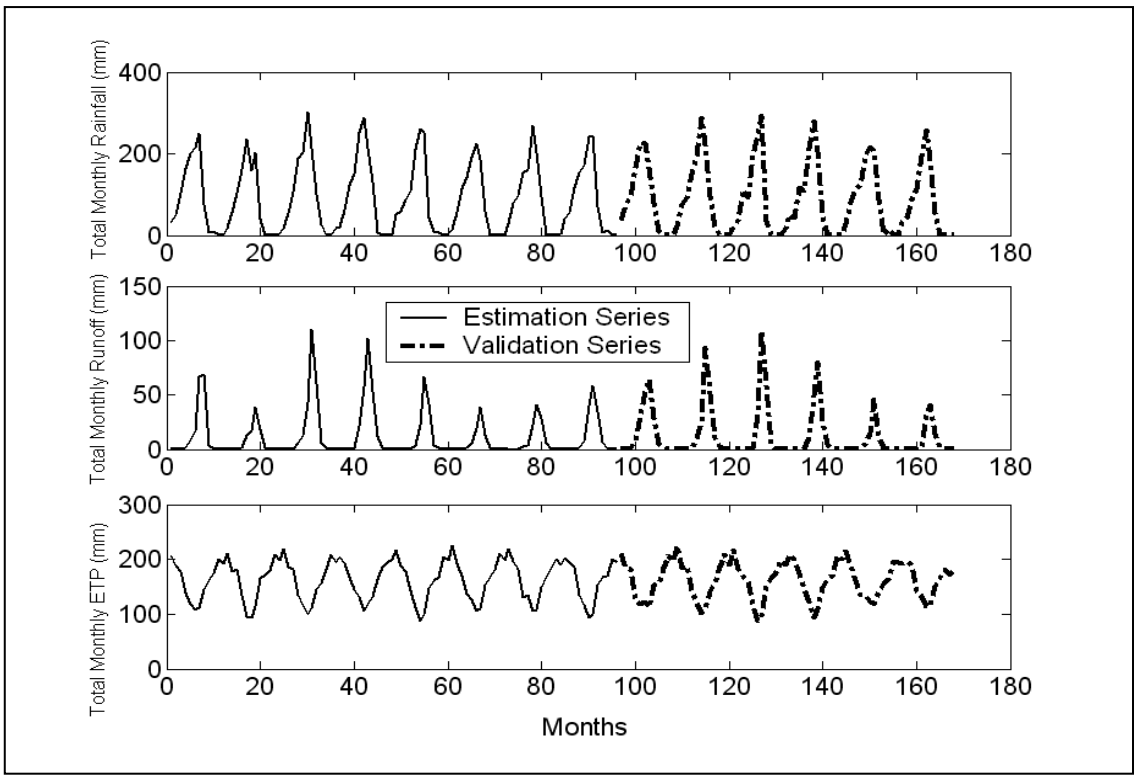


Figure 3.3b Observed monthly rainfall, runoff and ETP for Sabari, Oti River

Table 3.4 Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at (a) Bamboi and (b) Sabari. (All values in mm except CV = coefficient of variation, which is dimensionless)

(a) Bamboi

Statisticn	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	13.2	31.3	61.1	93.9	121.8	168.1	123.4	28.6	1.4	0.4	0.2	1.0
	Runoff	0.0	0.1	0.2	0.3	1.2	3.0	6.9	4.5	1.0	0.2	0.1	0.0
	ETP	205.5	187.9	181.3	155.0	134.3	114.9	119.7	147.7	161.6	161.0	169.8	179.9
Maximum	Rainfall	50.8	80.4	137.6	157.9	215.1	269.6	277.6	152.0	42.1	17.1	7.6	20.5
	Runoff	1.3	1.2	1.8	5.0	12.5	25.5	45.6	32.8	22.8	4.9	3.0	2.2
	ETP	227.1	208.1	206.7	181.7	157.8	138.3	138.7	180.7	173.4	178.8	189.3	193.0
Mean	Rainfall	25.4	53.9	95.5	121.4	173.0	220.5	176.9	63.9	11.9	4.6	1.7	7.3
	Runoff	0.6	0.6	1.0	2.1	4.2	8.1	18.1	16.8	5.6	2.2	1.3	0.8
	ETP	215.9	199.6	196.0	166.2	146.4	127.2	131.1	162.7	168.1	170.8	182.5	185.7
Std Dev.	Rainfall	9.4	12.8	18.7	17.0	24.8	27.7	33.5	30.3	9.9	4.5	2.2	5.6
	Runoff	0.4	0.3	0.4	1.3	2.7	5.2	8.8	8.6	4.5	1.1	0.8	0.5
	ETP	5.0	4.8	5.5	6.2	5.2	5.6	4.7	5.8	3.2	4.4	4.4	3.3
CV	Rainfall	0.37	0.24	0.20	0.14	0.14	0.13	0.19	0.47	0.83	0.98	1.29	0.76
	Runoff	0.67	0.57	0.44	0.65	0.65	0.64	0.49	0.51	0.81	0.51	0.65	0.71
	ETP	0.02	0.02	0.03	0.04	0.04	0.04	0.04	0.04	0.02	0.03	0.02	0.02

(b) Sabari

Statisticn	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	14.6	37.6	87.3	93.3	161.3	158.4	164.6	35.1	0.7	0.5	0.2	0.4
	Runoff	0.0	0.0	0.1	0.3	1.1	14.8	38.7	11.2	1.1	0.3	0.1	0.1
	ETP	192.1	177.2	161.2	123.5	96.2	86.8	98.0	131.8	154.2	163.4	170.9	179.5
Maximum	Rainfall	52.0	80.3	123.0	190.1	253.0	303.4	292.6	130.8	26.3	15.8	2.1	19.0
	Runoff	0.3	0.6	1.3	7.8	32.0	57.2	110.1	68.8	12.3	2.2	0.8	0.5
	ETP	224.7	198.2	190.0	162.1	135.1	121.9	119.1	166.2	171.4	185.7	210.6	200.6
Mean	Rainfall	29.5	59.4	105.3	141.4	203.4	247.9	217.2	72.4	6.6	4.2	0.7	4.4
	Runoff	0.1	0.2	0.4	1.9	9.8	31.7	68.3	36.0	4.8	1.0	0.4	0.2
	ETP	212.1	188.8	177.8	139.4	121.1	103.1	109.7	148.4	162.6	177.6	201.0	193.9
Std Dev.	Rainfall	12.0	13.4	11.5	27.9	23.8	38.3	33.0	27.8	7.3	5.1	0.6	6.2
	Runoff	0.1	0.2	0.3	2.2	8.4	14.4	26.5	19.4	3.3	0.6	0.2	0.1
	ETP	8.9	6.1	8.4	8.9	10.4	11.2	6.5	7.9	5.4	7.4	10.0	5.4
CV	Rainfall	0.41	0.23	0.11	0.20	0.12	0.15	0.15	0.38	1.11	1.21	0.90	1.43
	Runoff	0.76	0.92	0.73	1.17	0.86	0.45	0.39	0.54	0.68	0.56	0.57	0.64
	ETP	0.04	0.03	0.05	0.06	0.09	0.11	0.06	0.05	0.03	0.04	0.05	0.03

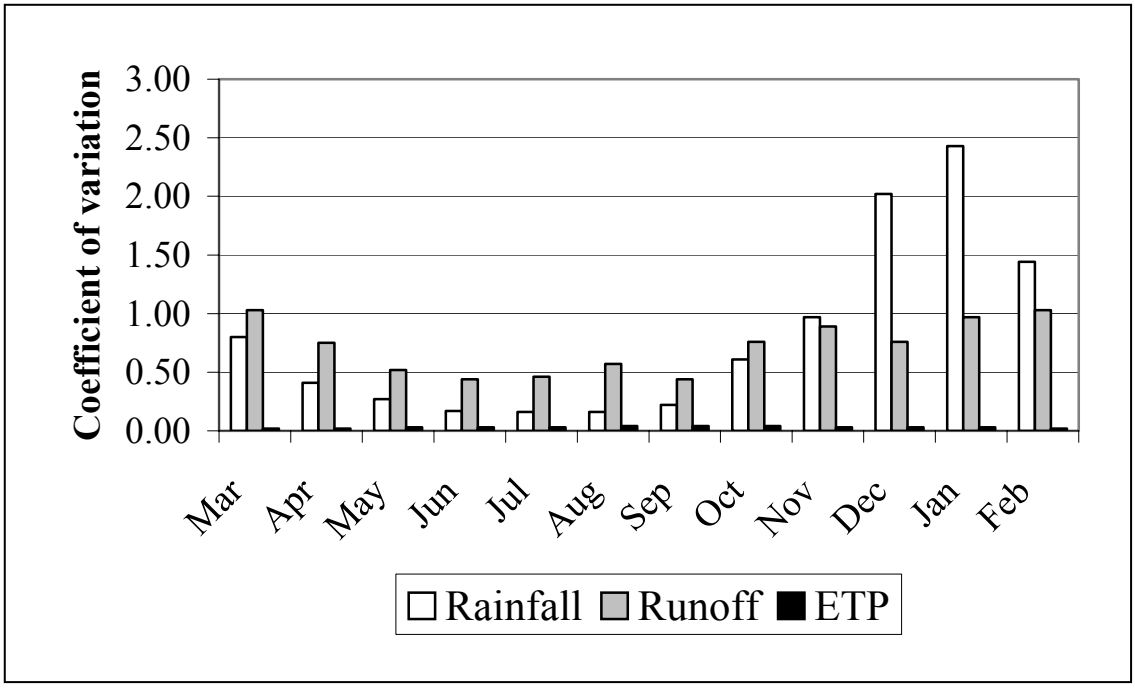


Figure 3.4a Coefficients of variation of monthly runoff, rainfall and potential evapotranspiration (ETP) at Dapola on the Black Volta River

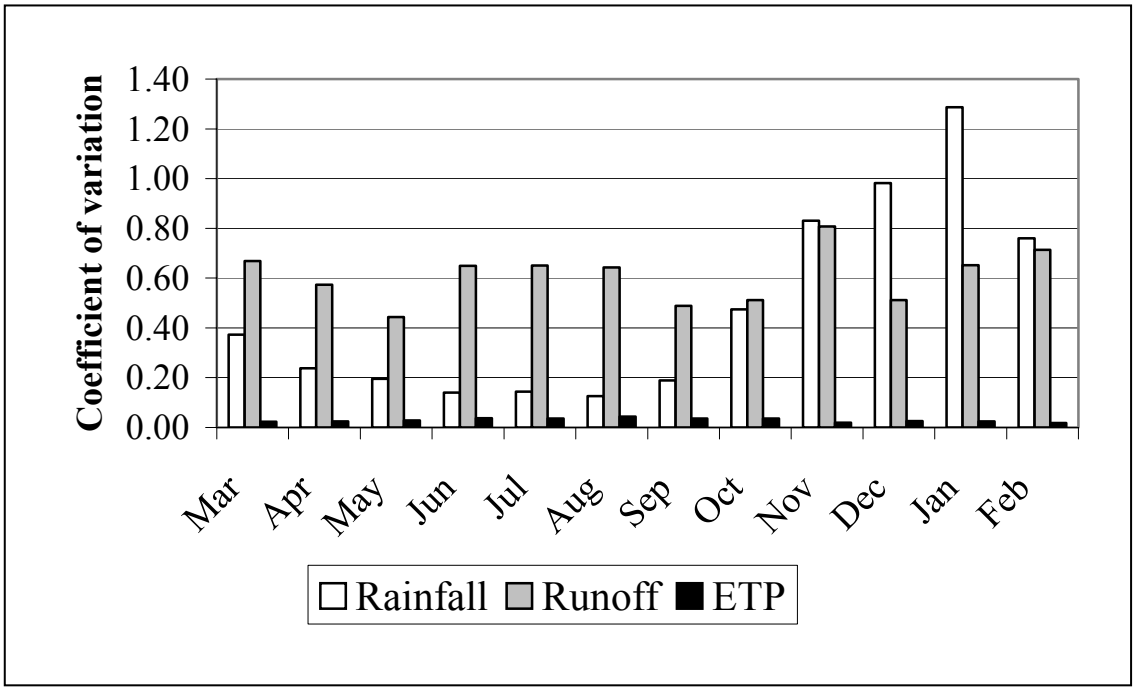


Figure 3.4b Coefficients of variation of monthly runoff, rainfall and potential evapotranspiration (ETP) at Bamboi on the Black Volta River

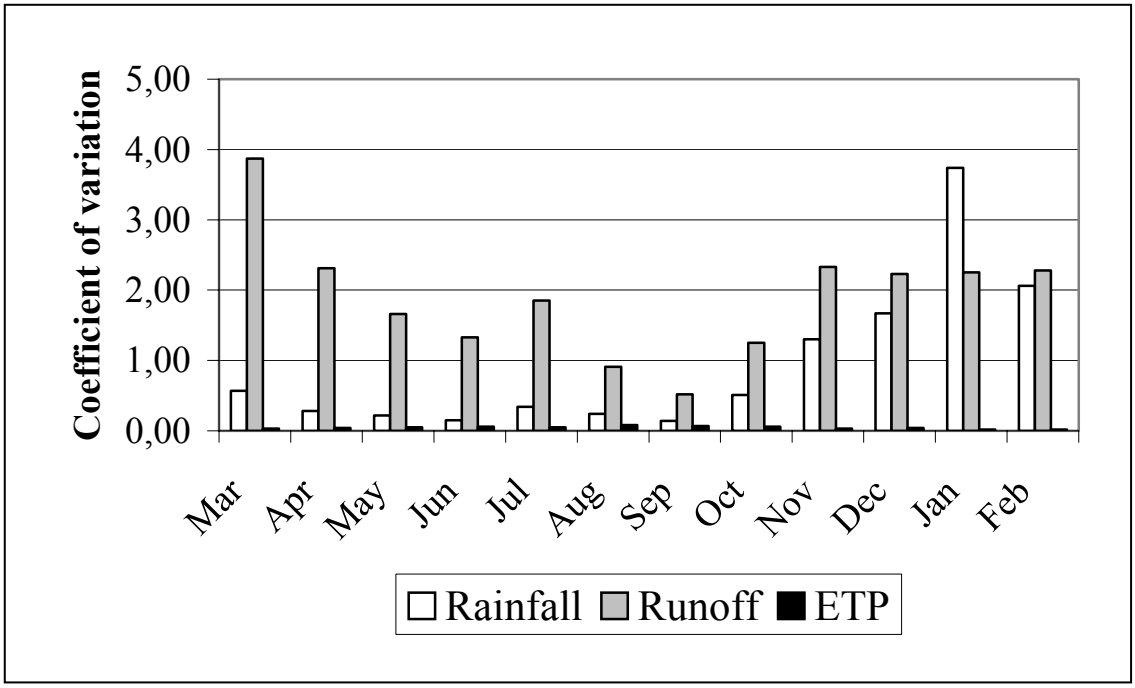


Figure 3.4c Coefficients of variation of monthly runoff, rainfall and potential evapotranspiration (ETP) at Yagaba on the Kulpawn River (White Volta Basin)

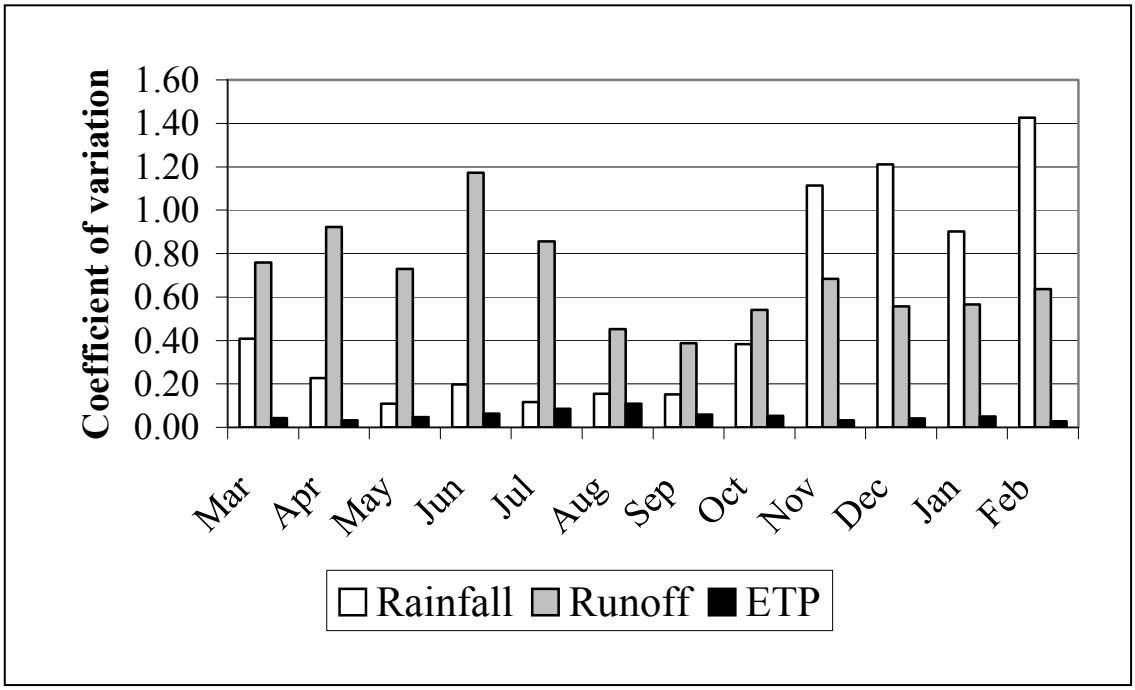
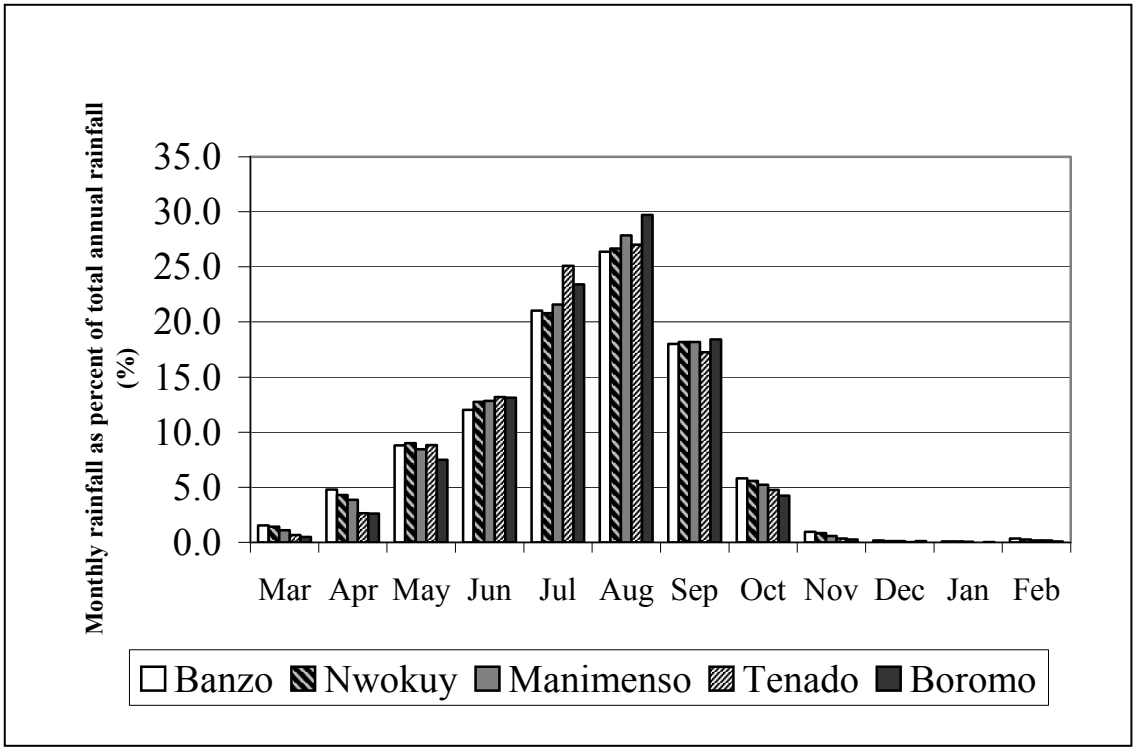
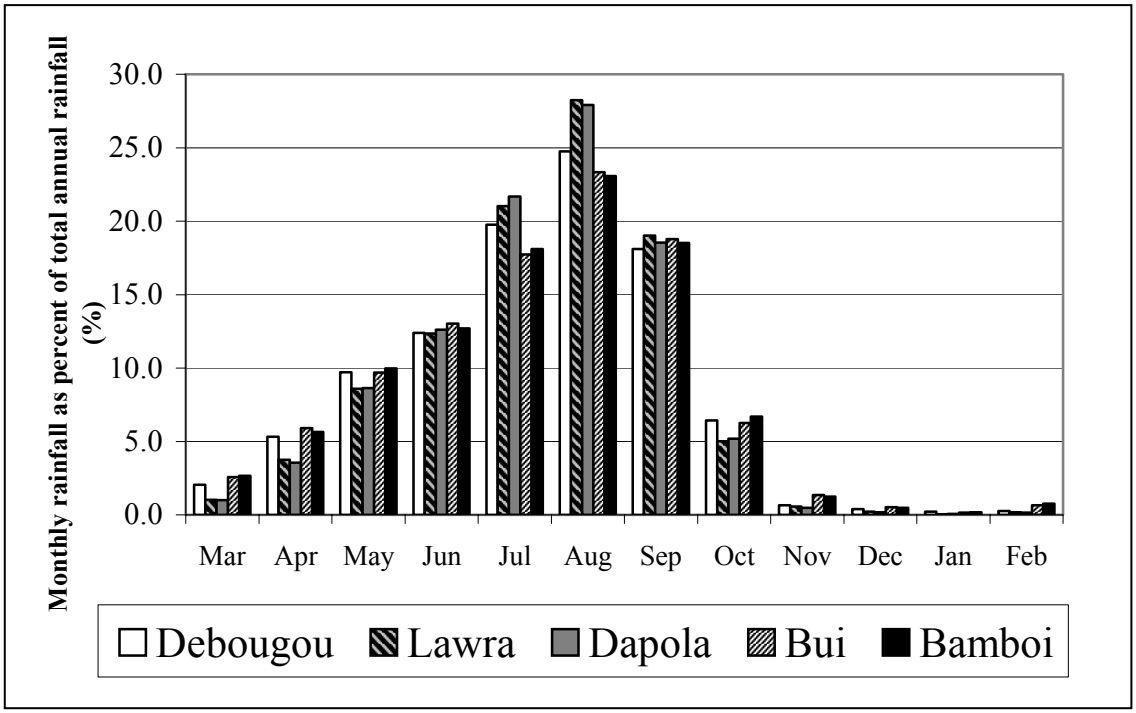


Figure 3.4d Coefficients of variation of monthly runoff, rainfall and potential evapotranspiration (ETP) at Sabari on the Oti River



(i)



(ii)

Figure 3.5a Percent of total catchment rainfall occurring in each month of the hydrological year – Black Volta Basin
(i) Upstream stations
(ii) Downstream stations

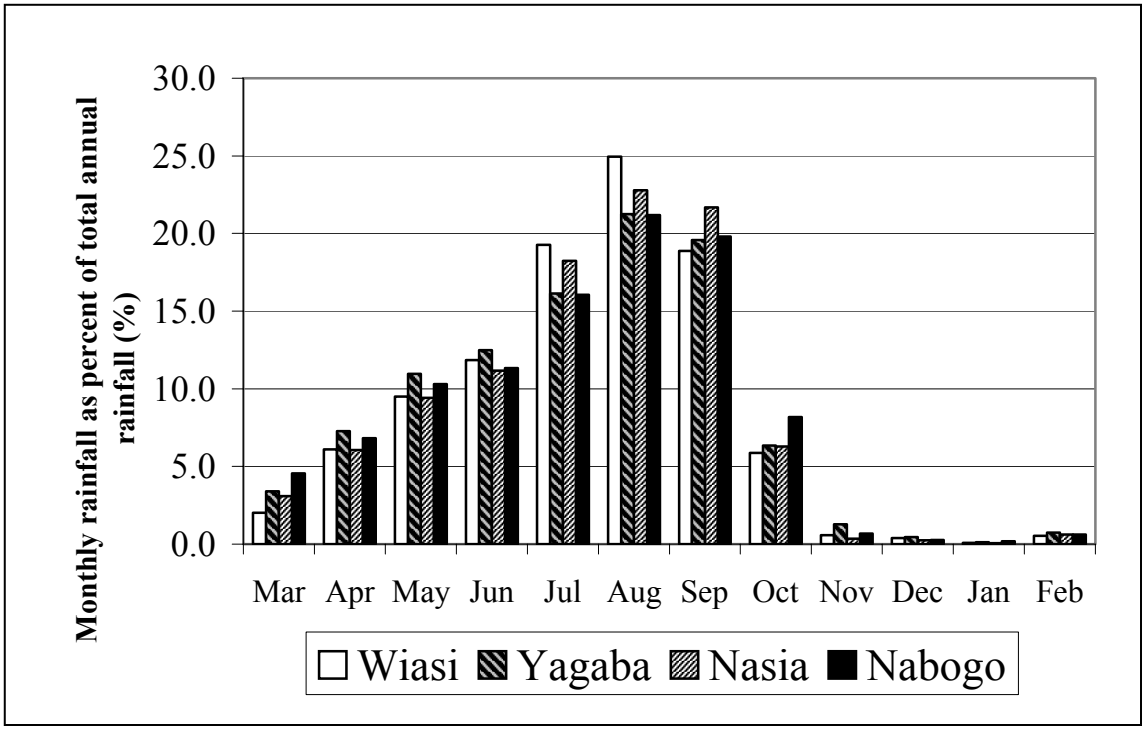


Figure 3.5b Percent of total catchment rainfall occurring in each month of the hydrological year - White Volta Basin

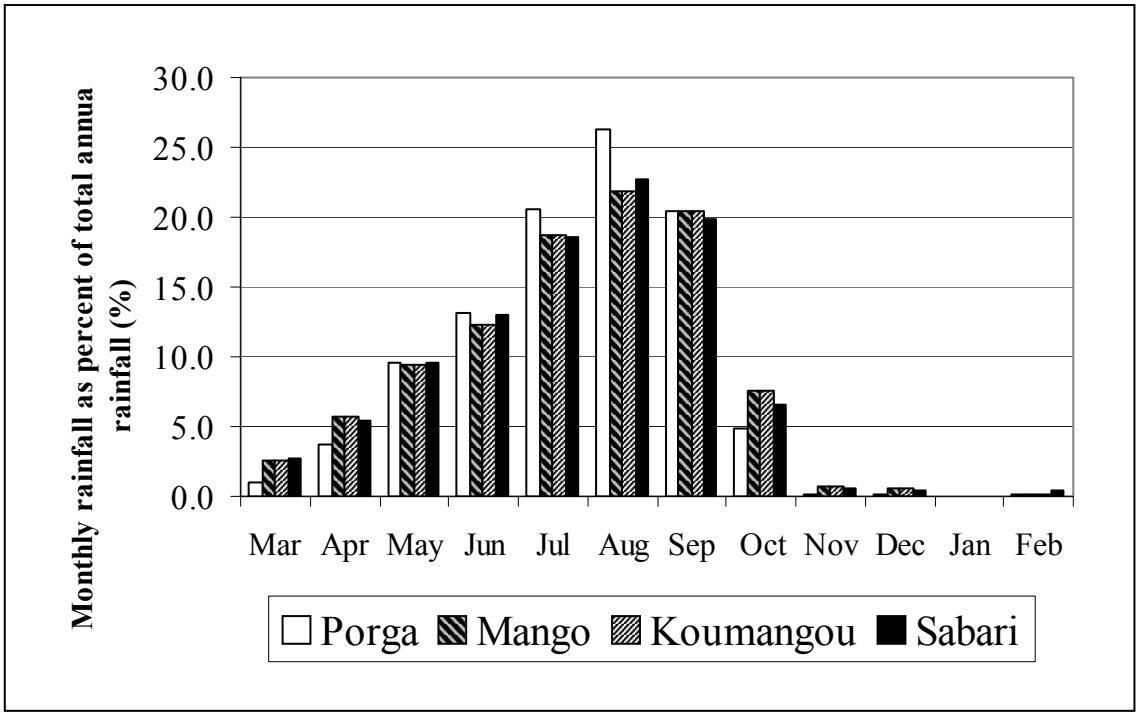


Figure 3.5c Percent of total catchment rainfall occurring in each month of the hydrological year - Oti Basin

The monthly means of rainfall and ETP for selected stations are plotted in figures 3.6a – 3.6d. The monthly mean rainfalls were aggregated from the corresponding daily values obtained from a network of rain gauges in the basin and interpolated to obtained mean values for each catchment using the Iddrisi GIS software (Taylor, 2004). It can be observed from these plots and tables 3.4 and 9.1i – 9.1xvi that except for the 3 (or 4 in some cases) wettest months of July, August and September (also June in some cases, tables 9.1i - 9.1xvi)), mean monthly ETP in the basin exceeds rainfall for all months. The rather high ETP for most of the year, particularly in the dry season, means much of the loss from groundwater would be to evapotranspiration rather than to baseflow. This implies that, for the Volta Basin, recession flow in streams would be a very poor indicator of groundwater recharge and storage. Another important implication of the high ETP for much of the year is that natural streamflow cannot be relied upon for water usage in the basin such as for agriculture and industry, and impoundments of some sort have to be resorted to. The numerous dams and reservoirs created for water storage throughout the basin is recognition of this implication.

Figures 3.7a – 3.7c are plots of monthly mean runoffs as percent of total annual runoff at the gauging stations and show the temporal distribution of runoff in the basin. The distribution in monthly runoff for the upstream stations in the Black Volta Basin are quite different from that of the rest of the stations including the downstream stations of the Black Volta. While for the other stations it is clear that most of the runoff occurs in the three months August to October, for the Black Volta upstream stations, June or July for some and November for others have comparatively substantial runoffs. This may be an abnormality and needs further investigation, since the rainfall pattern as shown in figures 3.5a – 3.5d is the same for all stations. However, for the ‚normal' stations, between 70 and 75%, 90 and 95% and 83 and 88% of the total annual runoff occurs in the period August-October for the Black Volta, White Volta and Oti basins, respectively. Thus, as in the case of the distribution of rainfall, this very skewed temporal distribution of runoff means that run-of-the-river water-use systems would require hydraulic structures to ensure all-year-round water availability.

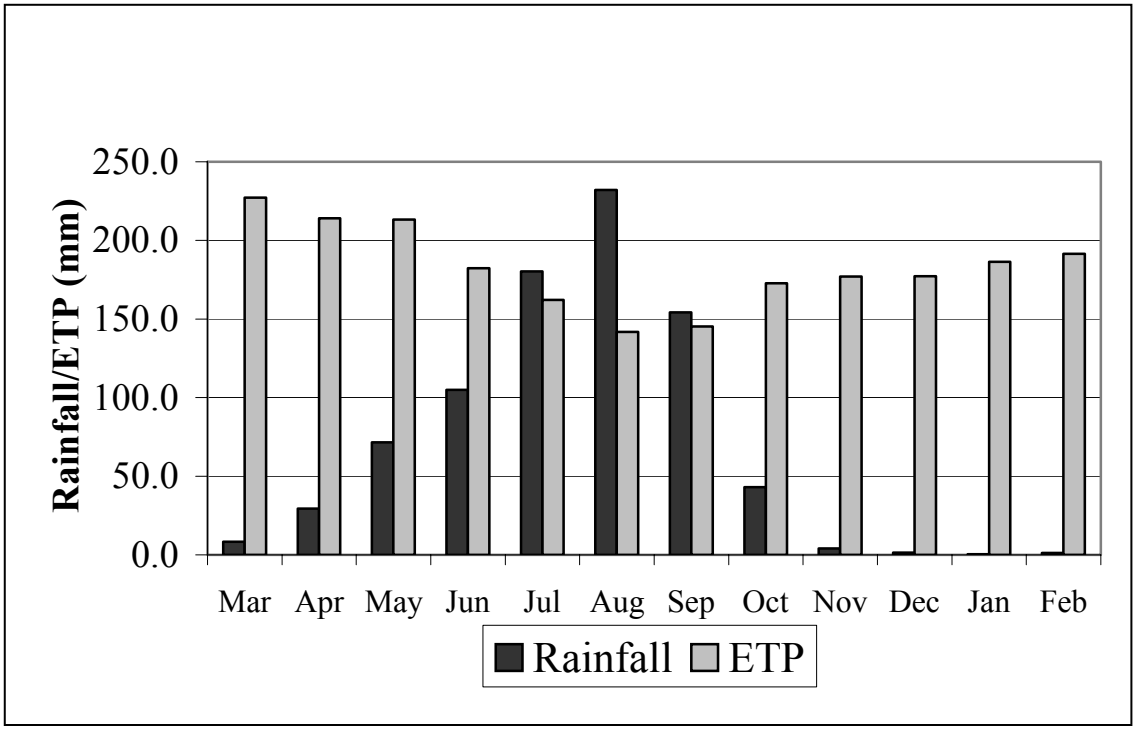


Figure 3.6a Monthly mean rainfall and potential evapotranspiration (ETP) for Dapola on the Black Volta River

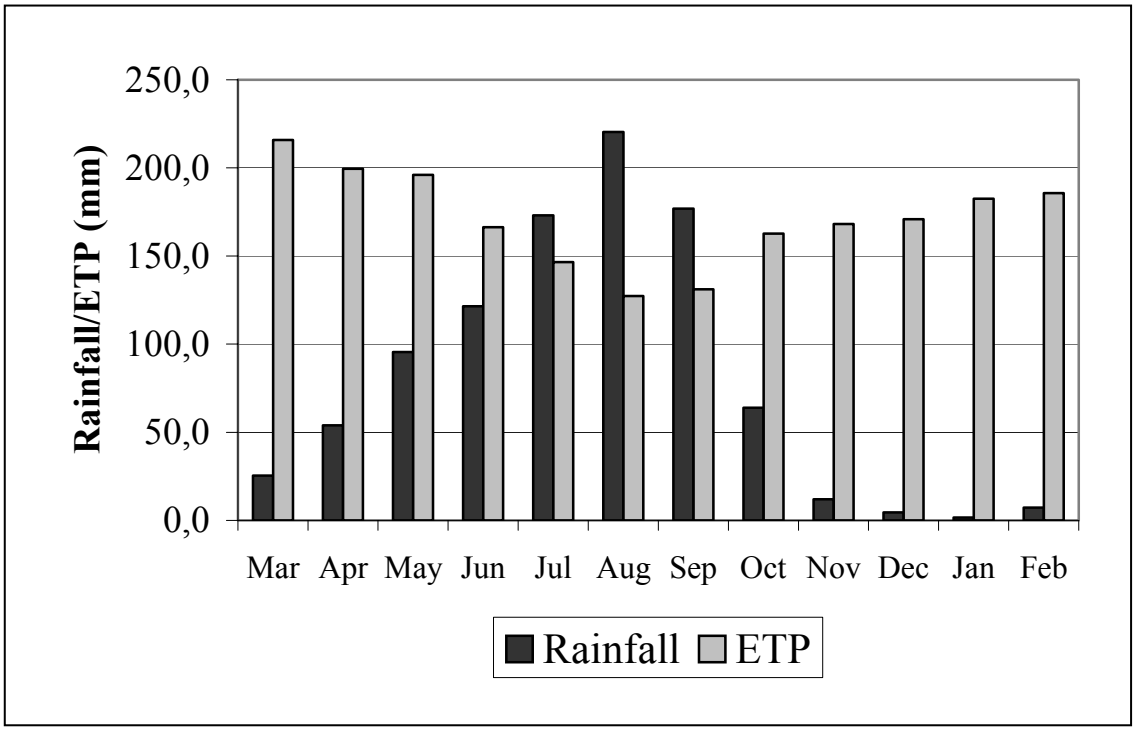


Figure 3.6b Monthly mean rainfall and potential evapotranspiration (ETP) for Bamboi on the Black Volta River

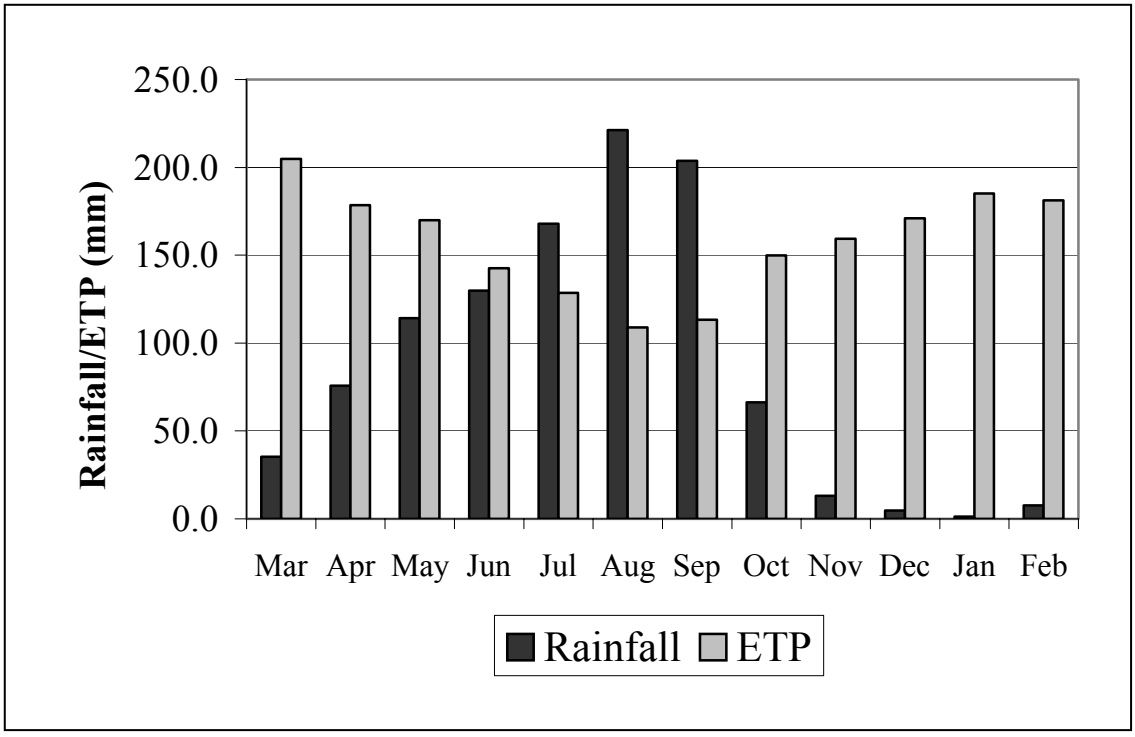


Figure 3.6c Monthly mean rainfall and potential evapotranspiration (ETP) for Yagaba on the Kulpawn River (White Volta Basin)

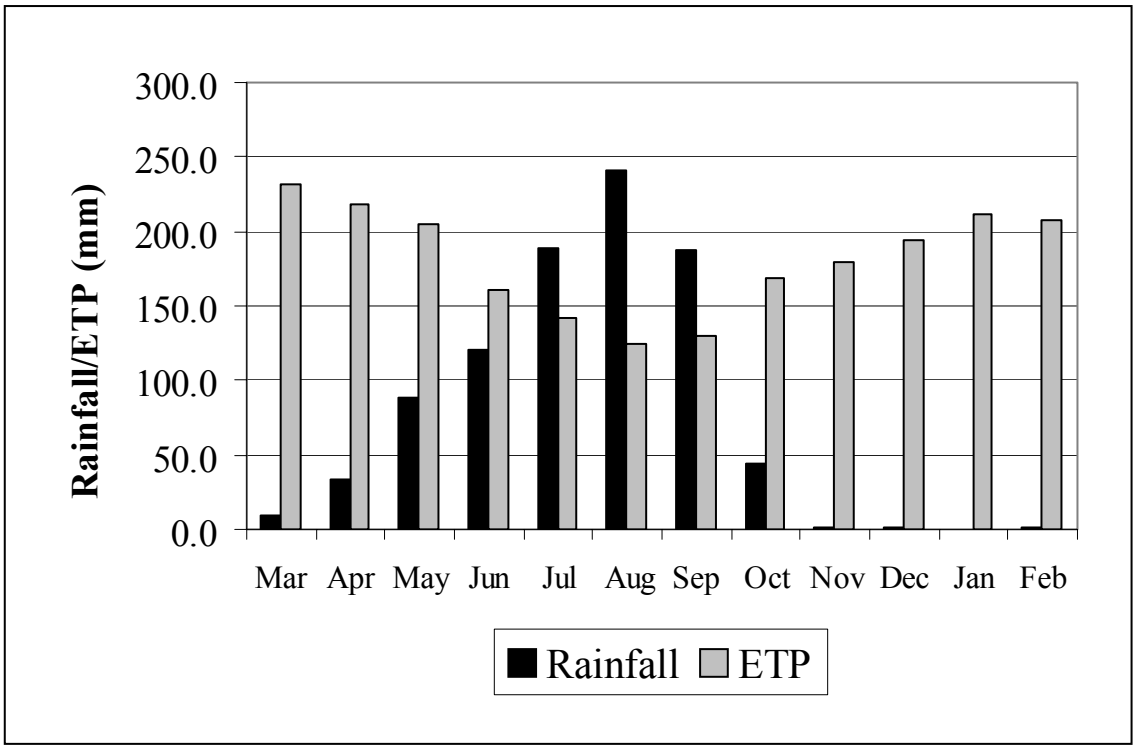
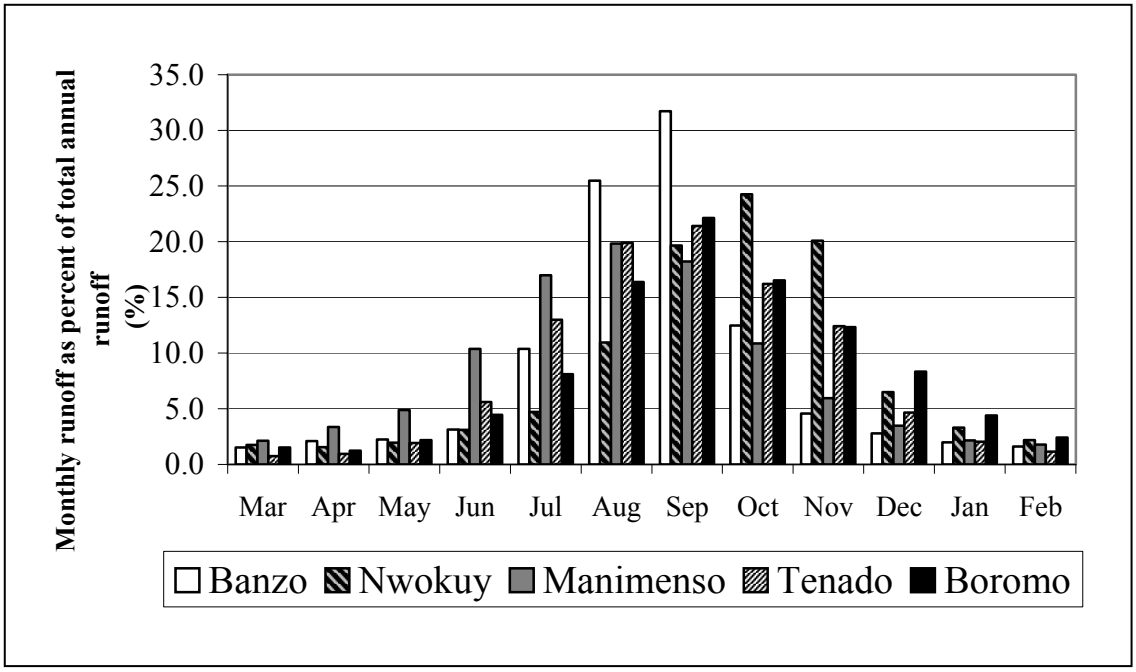
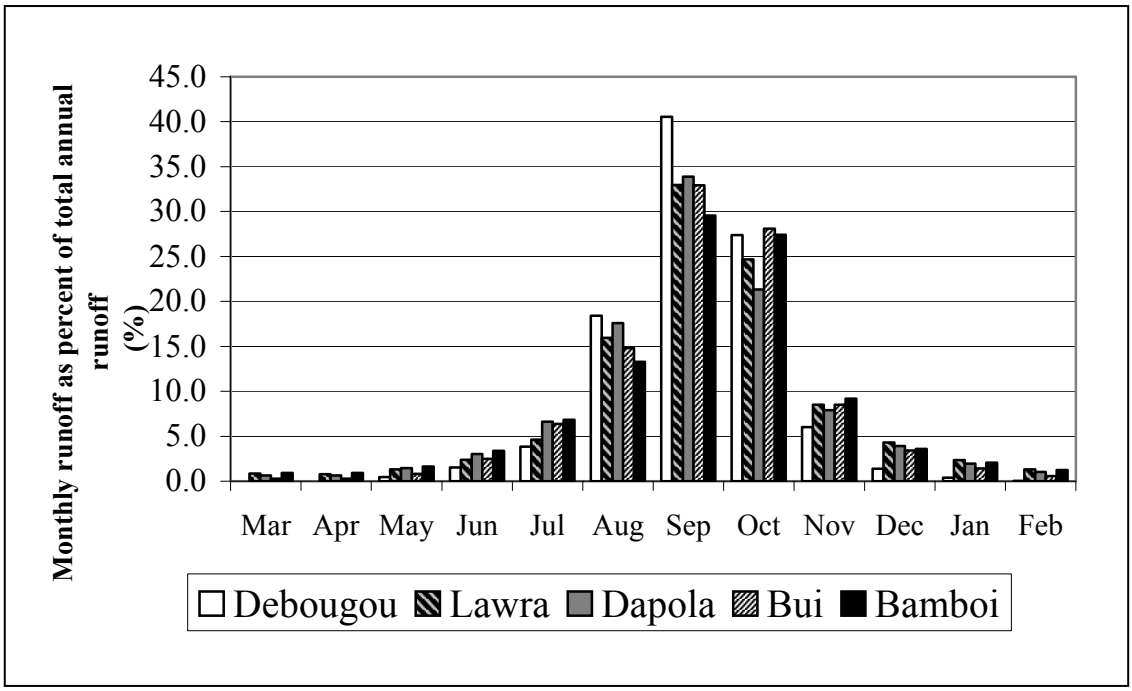


Figure 3.6d Monthly mean rainfall and potential evapotranspiration (ETP) for Sabari on the Oti River



(i)



(ii)

Figure 3.7a Percent of total catchment runoff occurring in each month of the hydrological year – Black Volta Basin
(i) Upstream stations
(ii) Downstream stations

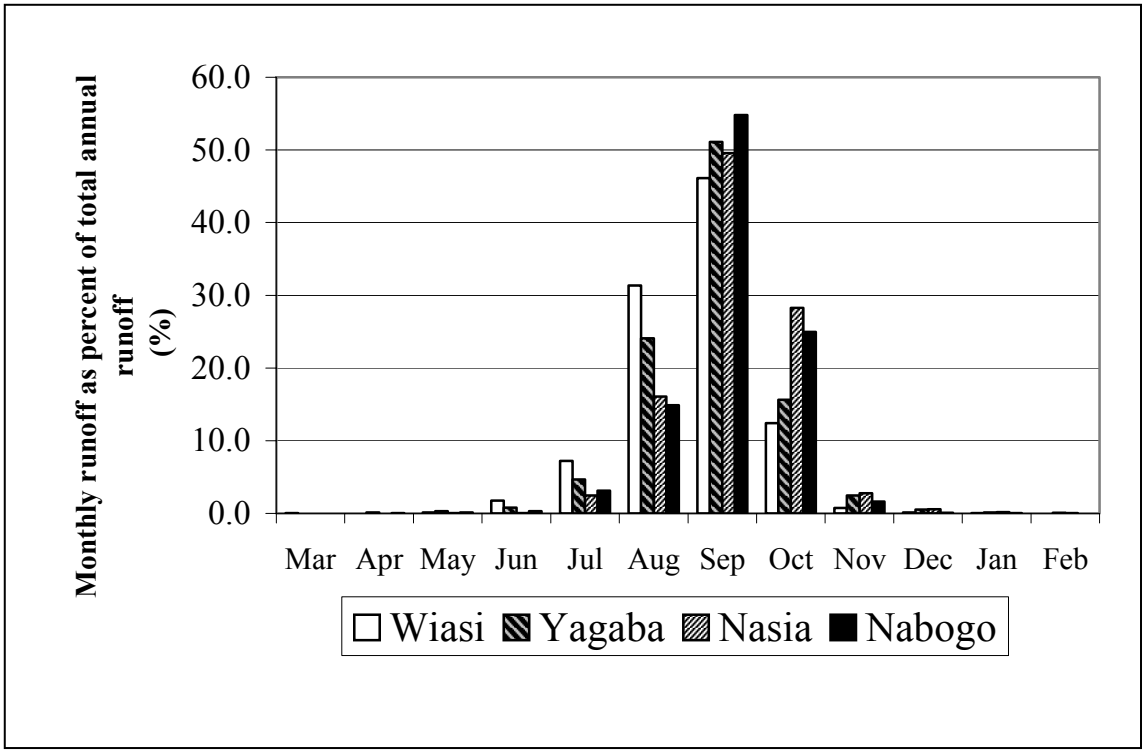


Figure 3.7b Percent of total catchment runoff occurring in each month of the hydrological year - White Volta Basin

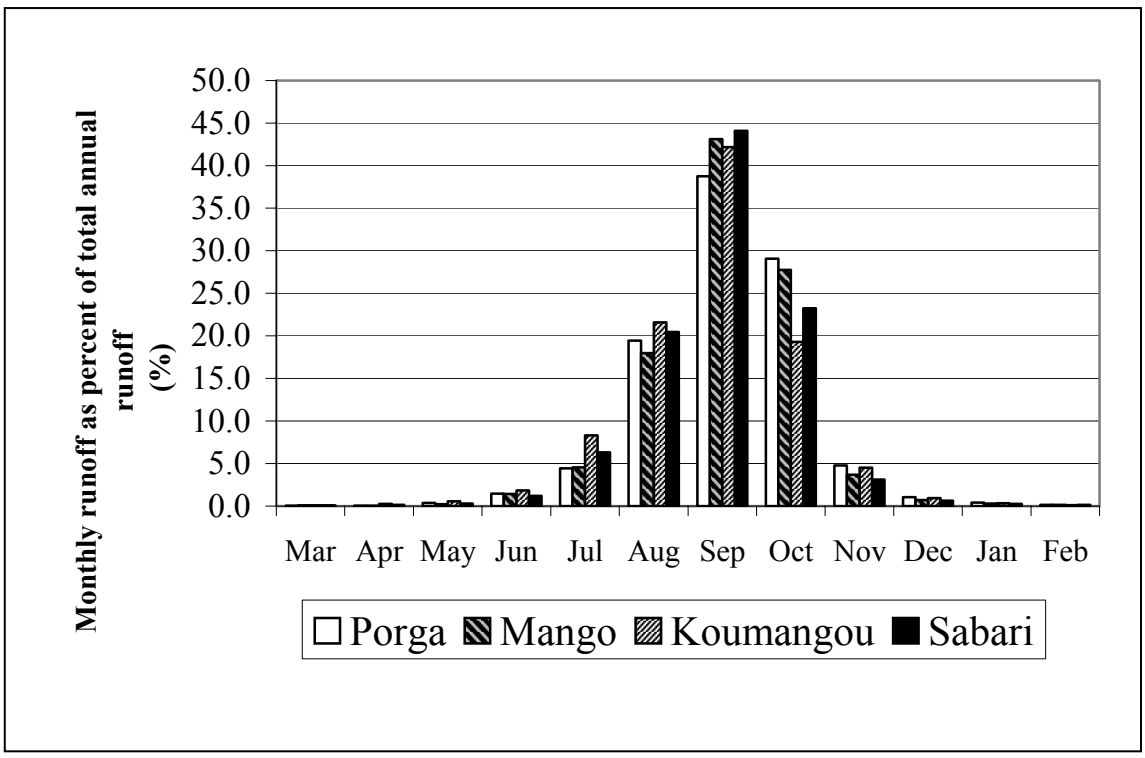


Figure 3.7c Percent of total catchment runoff occurring in each month of the hydrological year - Oti Basin

3.2.1 Persistence in river runoff

River runoffs often have memory or show persistence, i.e., the current runoff depends on past values. The fixed period of past runoffs influencing the current values is called the memory of the runoff system. Persistence of the runoff series used in this study was examined at both the monthly and annual scales by the use of the correlogram (plot of the autocorrelation function (ACF) against the lag), for the various runoff series. The ACF is the correlation coefficient r_k at various lags k = 1, 2, ...,nk of the flow series, nk being the number of feasible lags for the given length of the series.

The autocorrelation coefficient at lag k for a series $\boldsymbol{x} = (x_1, x_2, ..., x_N)$ of length N is given as (Anderson, 1976; Salas *et al*., 1980):

$$r_k = \frac{\sum_{i=1}^{N-k}\left(x_i - \bar{x}\right)\left(x_{i+k} - \bar{x}\right)}{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} \tag{3.1}$$

where x_i is the i^{th} element and $\bar{x}$ the mean of the series, respectively.

If the elements of x are independent and identically distributed random variables, then $Var(r_k)$, the variance of r_k, is given as:

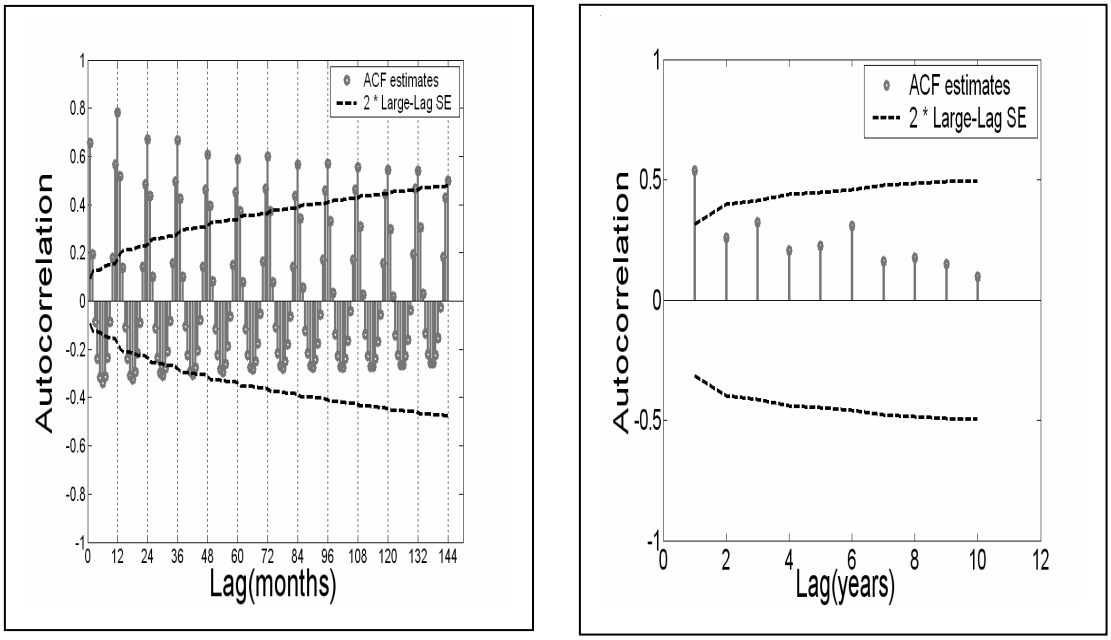
$$Var(r_k) = 1/N \tag{3.2}$$

Since successive values of r_k for a time series can be highly correlated, i.e., the value of a given r_k may be correlated with the K values at lower lags, $Var(r_k)$ is modified to account for this correlation as follows (Anderson, 1976):

$$Var(r_k) = \frac{1}{N}\left(1 + 2\sum_{i=1}^{K} r_i^2\right) \tag{3.3}$$

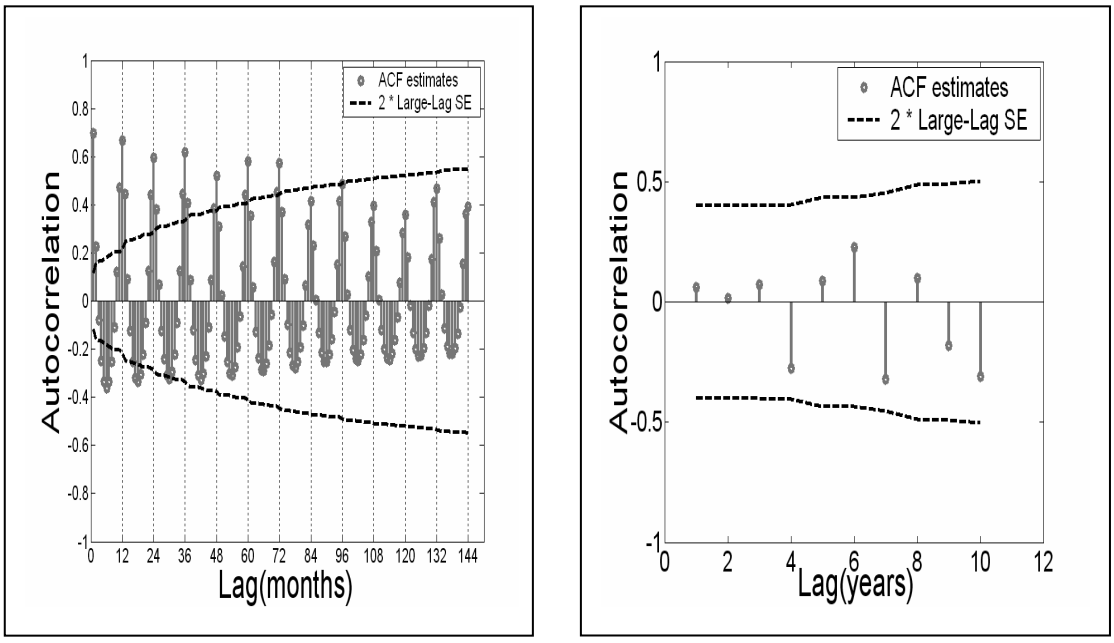
where K<k.

The standard errors derived from 3.3 are called the Large-lag standard errors of r_k (Anderson, 1976) and are used to establish the 95% confidence intervals for the correlation plots presented in figures 3.8a – 3.8d for both the monthly and annual runoff series.



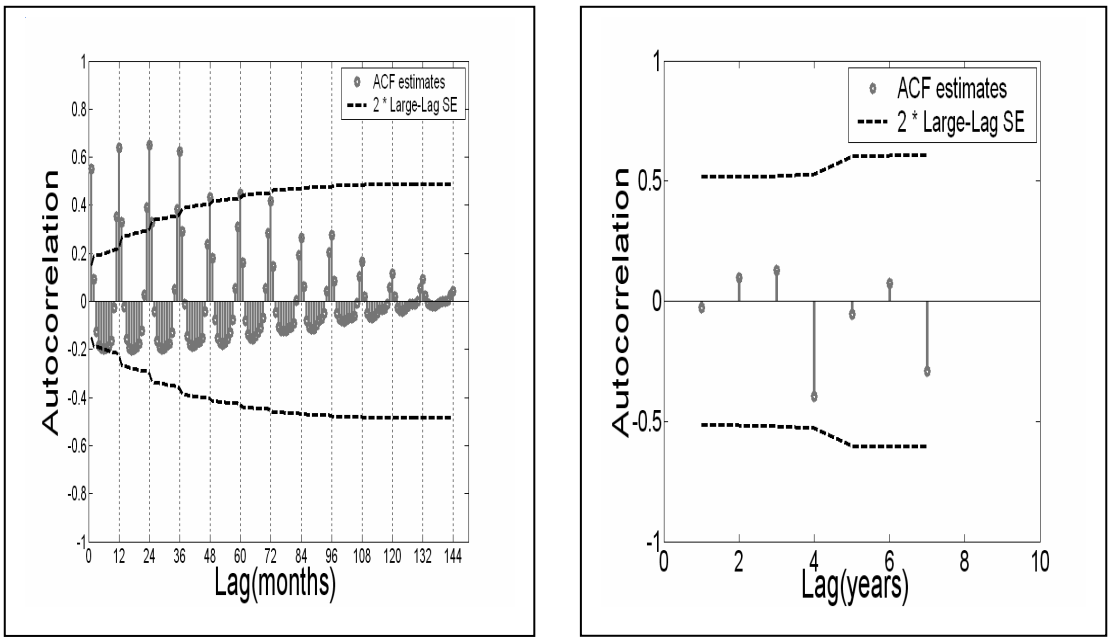
(i) (ii)

Figure 3.8a Autocorrelation function (ACF) for Dapola on the Black Volta River (SE = Standard error)
(i) Monthly series
(ii) Annual series



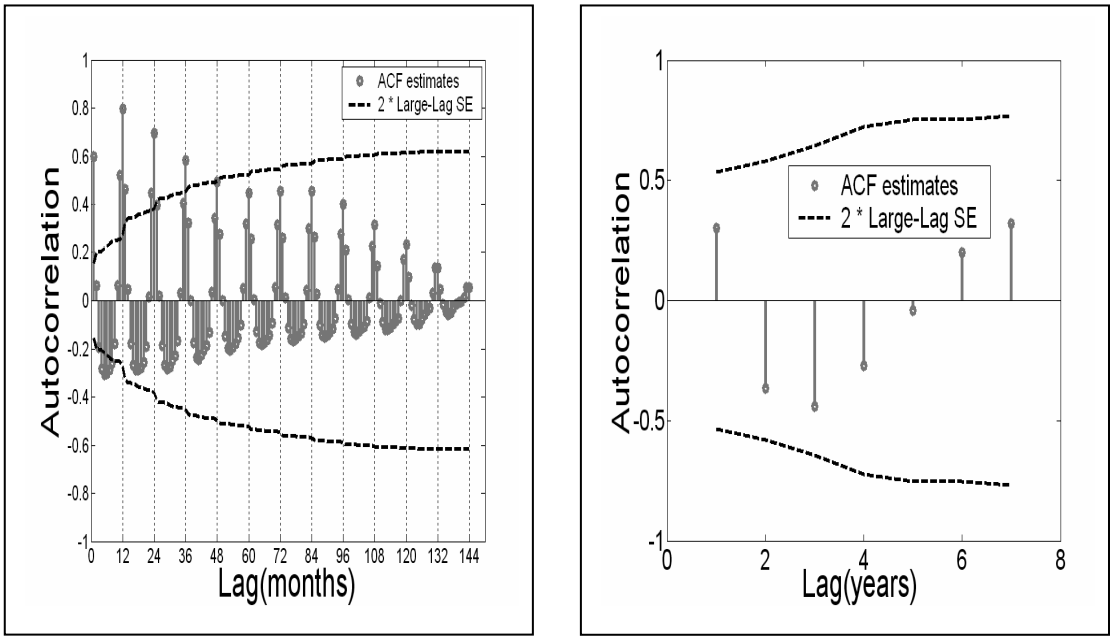
(i) (ii)

Figure 3.8b Autocorrelation unction (ACF) for Bamboi on the Black Volta River
(i) Monthly series
(ii) Annual series



(i) (ii)

Figure 3.8c Autocorrelation function (ACF) for Yagaba on the Kulpawn River (White Volta Basin)
(i) Monthly series
(ii) Annual series



(i) (ii)

Figure 3.8d Autocorrelation function (ACF) for Sabari on the Oti River
(i) Monthly series
(ii) Annual series

The correlation plots of figures 3.8a – 3.8d are stem plots with values of the correlation coefficient at a given lag indicated by the circle at the tip of the vertical line at that lag. The dotted lines are the ±2large-lag confidence bands at the 95% significance level. Though this level means that on average a given r_k value has a 5% chance of being outside the confidence bands, r_k values well above the bands should indicate significant autocorrelation.

The plots show both seasonality and persistence in the monthly runoff series. However, except for the rather small r_1 value shown as significant for Dapola, neither seasonality nor persistence exists for the annual series, an indication that for the Volta Basin, the annual series are independent and without memory. This means that any delayed or groundwater flow to the rivers ceases at the end of the dry season, so that there is little or no over-year baseflow. At the annual scale, therefore, direct runoff is the only important contributor to riverflow. On the other hand, persistence in monthly flow means that delayed direct runoff and/or baseflow may be relatively important at this scale.

3.2.2 Autoregressive and moving average modelling of monthly runoff

Since there is persistence in the monthly runoff series, two black box autoregressive models were investigated to determine their suitability for representing the flow series in the basin. Various orders of the polynomial autoregressive (AR) and autoregressive moving average (ARMA) models (Box *et al*, 1994) were fitted to monthly runoff series, and the best models (from evaluation criteria given below) selected for each runoff series. The ARX and ARMAX functions of the Systems Identification Toolbox of Matlab (Ljung, 2003) were used to estimate the models. The general ARMA (na,nc) model used is as given below (Ljung, 1999; Ljung, 2003):

$$A(q)y_t = C(q)e_t \tag{3.4}$$

where:

$\boldsymbol{y}_t$ = deseasonalized monthly river runoff at time t months

$\boldsymbol{e}_t$ = random error at time t months

$$A(q) = (1 - a_1 q - a_2 q^2 - \ldots\ldots - a_{na} q^{na})$$
$$C(q) = (1 - c_1 q - c_2 q^2 - \ldots\ldots - c_{nc} q^{nc})$$

q = backward operator, i.e., $qy_t = y_{t-1}$

$a_1, a_2, \ldots, a_{na}$ are the parameters of the AR component of the model to be estimated

$c_1, c_2, \ldots, c_{nc}$ are the parameters of the MA component of the model to be estimated

na = order of the AR model

nc = order of the MA model

Each runoff series was deseasonalized by subtracting the monthly means from each sample (Salas *et al.*, 1980). The AR ARMA modelling was undertaken using the deseasonalized series. It should be noted that deseasonalizing the original series means that the number of parameters obtained for any model fitted to the series should be augmented by 12, the number of seasonal means, to obtain the total number of parameters involved (Salas *et al.*, 1980). Two model selection criteria were used, i.e., the K-fold cross validation (Kohavi, 1995; Lendasse *et al*., 2003; Comp.ai.neural-nets FAQ, 2004) and the Akaike Information Criterion, AIC (Akaike, 1974). The reason for using both model selection criteria was to evaluate them and determine which is more convenient to use in the nonlinear modelling activities described in later chapters.

K-fold cross validation

For the K-fold cross validation, the following procedure (Lendasse *et al*., 2003) was followed for each model order (na, nc):

1. Divide the samples of the runoff series $\boldsymbol{y_t}$ into *K* series of roughly equal size. The samples of *the* k^{th} series form the validation series Yval. The other sets form a new estimation series Yest.

2. Fit the ARMA model to Yest and compute the mean error ME_k (na,nc*)* of Yval as:

$$ME_k(na,nc) = \frac{\sum_{t=1}^{nv} \left(y_t^{val} - y_t^{pval} \right)^2}{nv} \qquad (3.5)$$

where y_t^{val} is an observed sample, y_t^{pval} the corresponding predicted sample of the validation series Yval, and nv the number of samples in the validation series.

3. Repeat Steps 1 and 2 for *k* varying from 1 to *K*. The average error, *E*, is then computed according to:

$$E = \frac{\sum_{k=1}^{K} ME_k}{K} \tag{3.6}$$

4. Repeat steps 1 to 3 for all model orders (na, nc) being considered and select the best model as the one with minimum E.

In this study, K was fixed at 10 for series of 240 months and above and 5 for the rest. The division of the series into Yest and Yval was done such that each series had an integer number of whole hydrological years.

Single validation series method

In this method, each runoff series being fitted was divided into two sub-series only. The first sub-series was the estimation series and was used to estimate the parameters of the candidate model. The second sub-series was the validation series and was used to select the best model from those fitted to the estimation series using the Akaike Information Criterion, AIC (Akaike, 1974). The following form of the criterion was used:

$$AIC = log\ (V) + 2p/n \tag{3.7}$$

where V is the sum of squares of the residuals of the predicted from the observed validation series, p the number of parameters estimated (na+nc) and n the length of the series used.

The selected models for all the runoff series used were also compared by means of the Nash-Sutcliffe Efficiency coefficient, NSE (Nash and Sutcliffe, 1970), to

ascertain how much predictions from each model are better than the means of the respective series. The NSE (%) used is defined as:

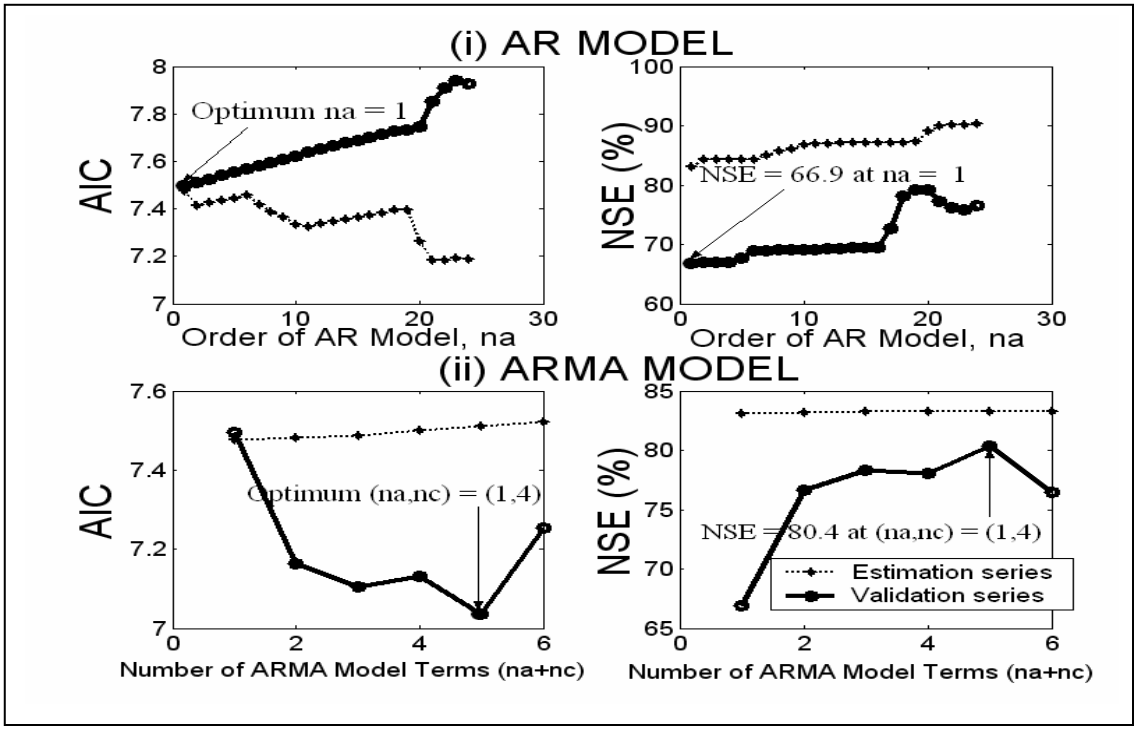
$$NSE = 100(1-\frac{\sigma_e}{\sigma_y}) \tag{3.8}$$

where σ_e is the variance of the residuals and σ_y that of the observed runoff series.

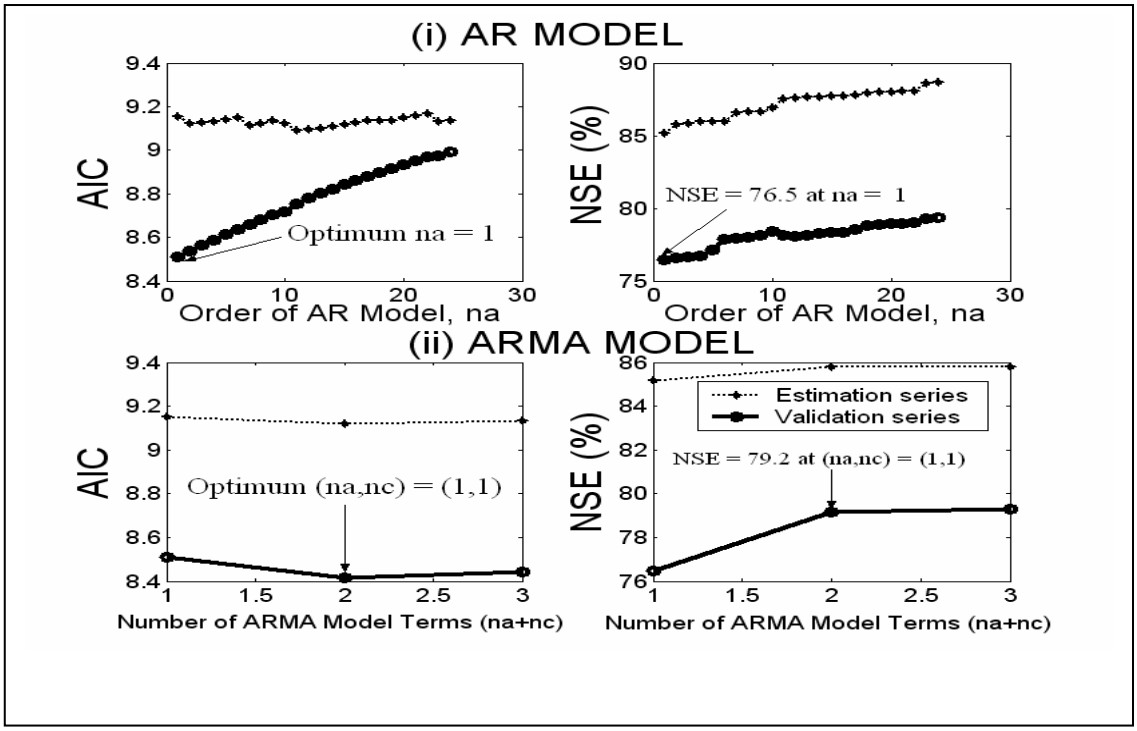
The estimation series in each case for this method was not chosen arbitrarily. It was chosen long enough to contain as much as possible the variation in the validation data (Young, 2001b). This was done subjectively by examining the series plot and ensuring that at least 5 years of the series was left for validation of very high orders. As a result, only stations with runoff series of 120 months or more were examined.

For both the K-fold cross validation and the single validation series methods, an AR model (na) was selected first and then an appropriate MA model (nc) fitted. This allowed better comparisons between the models selected by the two methods.

The values of the model selection criteria vs. model orders are plotted for selected stations in Figure 3.9 for the AIC fits and in Figure 3.10 for the K-fold cross validation. The optimum na or (na, nc) values shown in the figures are the respective values obtained at the minimum validation AIC, or minimum validation mean square errors in the case of the cross validation method. Results of the AR and ARMA model selection for the flow series investigated are summarized in Table 3.5. The entries for each station are the models selected and the corresponding NSE values computed for the estimation, validation and full series for the AIC case and for the full series only for the K-fold cross validation case. The NSE values at the minimum validation AIC of the AR model are presented in the first row and those at the minimum validation AIC of the ARMA model in the second. The NSE coefficients for the models selected by the K-fold cross validation AR and ARMA fits are shown in rows 3 and 4, respectively.

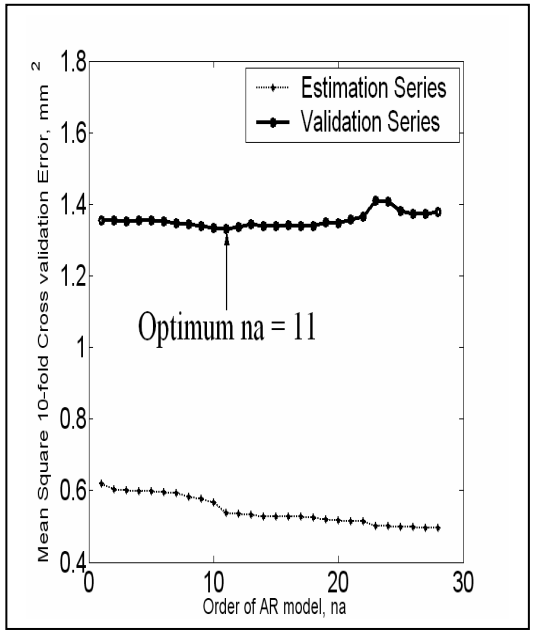


(a)

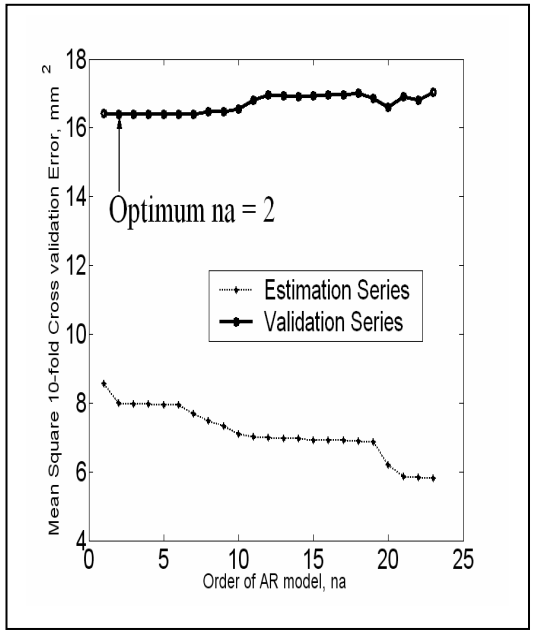


(b)

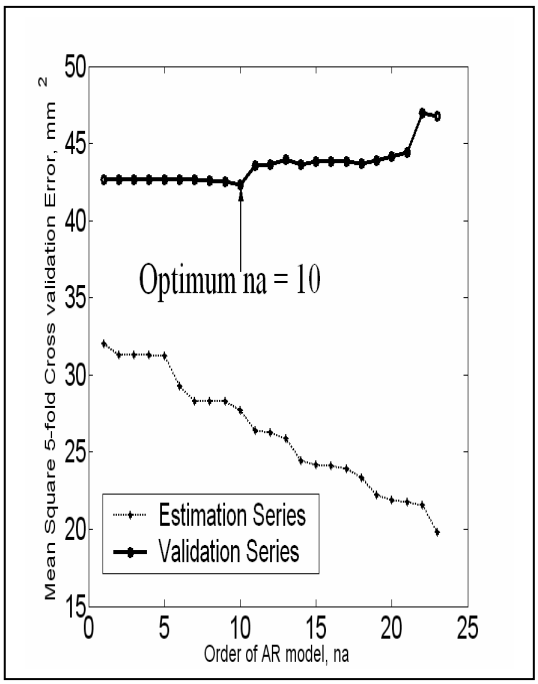
Figure 3.9 Akaike information criterion and Nash-Sutcliffe efficiency vs. model order
(a) Bamboi on the Black Volta River
(b) Mango on the Oti River



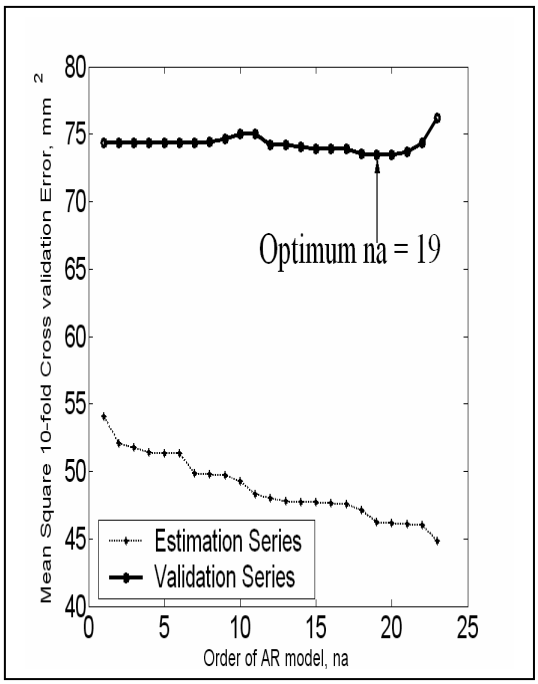
(a)



(b)



(c)



(d)

Figure 3.10 Cross validation mean square errors vs. AR model order na
(a) Boromo on the Black Volta River
(b) Bamboi on the Black Volta River
(c) Wiasi on the Sisilli River (White Volta Basin)
(d) Mango on the Oti River

Table 3.5 Nash-Sutcliffe efficiency values for the AIC and the K-fold cross validation Fits for the gauging stations studied. (BVB – Black Volta Basin; WVB – White Volta Basin; OB – Oti Basin. Entries in rows 1 and 2 for each station correspond, respectively, to the AR and ARMA models selected by the AIC; those in rows 3 and 4 are the corresponding values for the models (AR and then ARMA) selected by the K-fold cross validation fit)

Gauging station	**AR(ARMA) order na, (na,nc)**	**NASH-SUTCLIFFE EFFICIENCY (%)**		
		Estimation series (AIC)	**Validation series (AIC)**	**Full series at optimum na, or (na, nc)**
1. Banzo (BVB)	1	83.95	58.69	77.52
	(1,6)	84.00	66.71	77.68
	9	-	-	77.76
	(9,0)	-	-	77.76
2. Nwokuy (BVB)	1	89.41	37.08	87.32
	(1,2)	90.22	67.14	88.05
	2	-	-	88.38
	(2,1)	-	-	88.02
3. Manimenso (BVB)	1	95.32	48.54	92.71
	(1,5)	96.34	84.69	94.16
	1	-	-	92.71
	(1,4)	-	-	94.06
4. Boromo (BVB)	1	86.39	38.56	82.66
	(1,4)	86.98	65.72	83.18
	11	-	-	84.88
	(11,0)	-	-	84.88
5. Debougou (BVB)	1	78.94	42.83	75.11
	(1,3)	80.39	61.66	77.42
	2	-	-	76.69
	(2,1)	-	-	75.89
6. Lawra (BVB)	1	80.42	72.29	81.02
	(1,1)	81.13	79.78	81.85
	2	-	-	81.83
	(2,1)	-	-	81.88
7. Dapola (BVB)	1	80.68	32.74	77.84
	(1,1)	81.55	52.71	78.71
	2	-	-	78.63
	(2,1)	-	-	78.71
8. Bui (BVB)	1	84.17	83.70	86.15
	(1,2)	84.93	89.74	86.30
	11	-	-	87.43
	(11,0)	-	-	87.43
9. Bamboi (BVB)	1	83.06	66.89	83.56
	(1,4)	83.26	80.35	83.68
	2	-	-	84.66
	(2,1)	-	-	83.64

Table 3.5 Continued

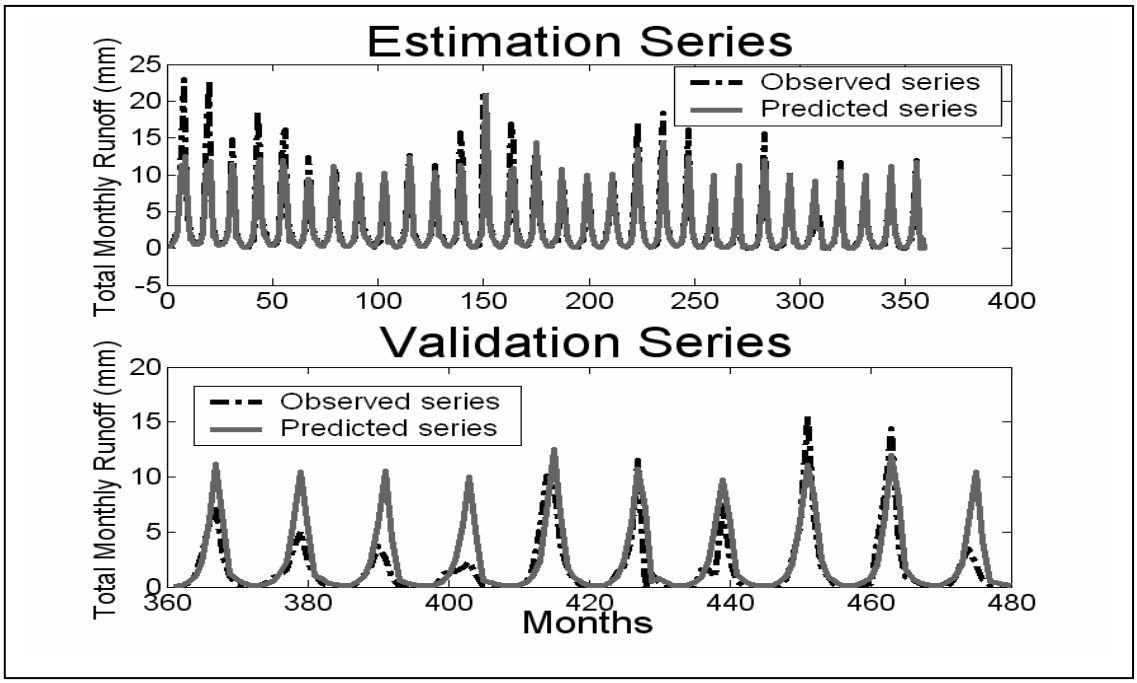
Gauging station	AR(ARMA) order na, (na,nc)	NASH-SUTCLIFFE EFFICIENCY (%)		
		Estimation series (AIC)	Validation series (AIC)	Full series at optimum na, or (na, nc)
10. Wiasi (WVB)	1	75.83	64.35	74.28
	(1,1)	76.02	71.08	74.51
	10	-	-	77.55
	(10,0)	-	-	77.55
11. Yagaba (WVB)	1	76.20	67.71	76.68
	(1,2)	74.98	76.52	75.02
	11	-	-	79.41
	(11,0)	-	-	79.41
12. Nabogo (WVB)	1	79.37	73.37	77.85
	(1,1)	78.92	74.57	77.27
	23	-	-	86.28
	(23,0)	-	-	86.28
13. Porga (OB)	1	85.27	43.35	80.79
	(1,1)	85.49	55.33	81.13
	9	-	-	83.05
	(9,0)	-	-	83.05
14. Mango (OB)	1	85.17	76.27	83.62
	(1,1)	85.80	78.97	84.23
	8	-	-	84.89
	(8,0)	-	-	84.89
15. Koumangou (OB)	1	83.50	82.52	83.47
	(1,1)	83.47	83.98	83.39
	9	-	-	83.77
	(9,0)	-	-	83.77
16. Sabari (OB)	1	86.12	81.28	85.85
	(1,1)	86.69	84.88	86.61
	10	-	-	88.76
	(10,0)	-	-	88.76

The model selection plots in Figure 3.9 show that the MA part of the ARMA models (obtained after fitting the AR models) improves the prediction of the validation series more than the same number of parameters for the AR model, as indicated by the validation NSE values. Since the MA terms model the random errors influencing the runoffs, stochasticity is an important consideration in modelling these series.

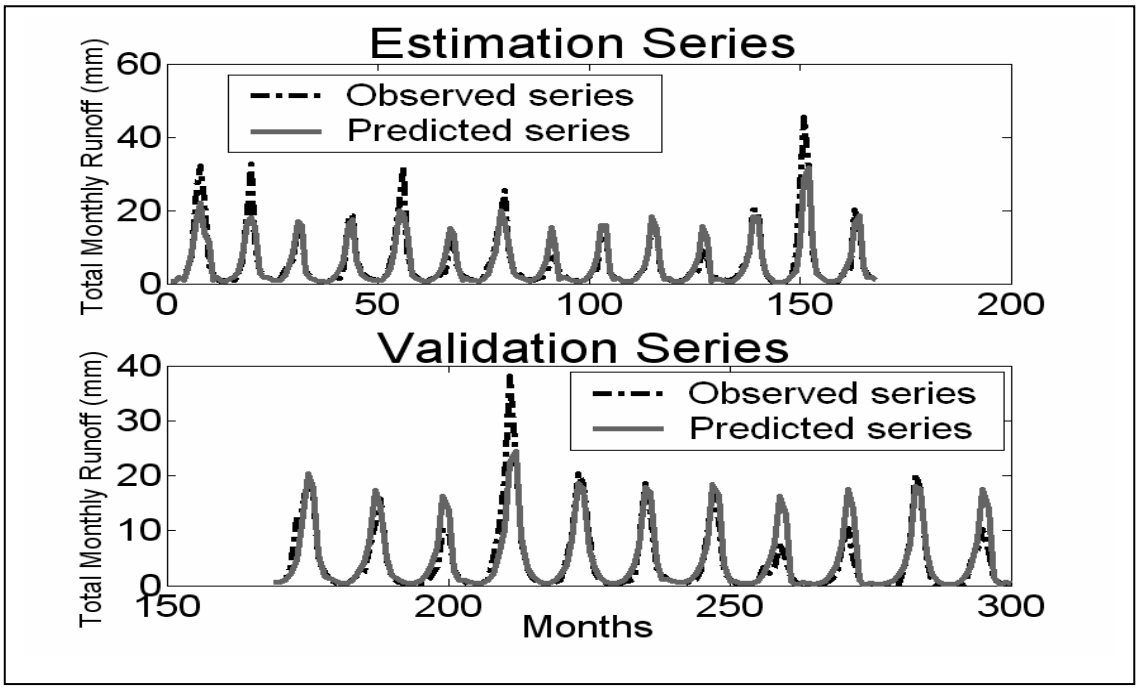
It is clear from the figures and the table, that the models selected by the two procedures are different for both types of models (except for Manimenso – station 3 - where an AR(1) model is selected by both). For the AR models, for example, the AIC applied to the one validation series consistently selects an AR(1) as the best model. The K-fold cross validation, on the other hand, selects quite a range of models in this category, and in all cases but one models of higher order than by the AIC are selected. The AIC is thus a more stringent model selection criterion. Nevertheless, considering the NSEs, the criterion has been able to pick very low order models whose performance is comparable to that of the higher order models selected by the K-fold cross validation.

The K-fold cross validation method averages the mean out-of-sample prediction errors over several validation series arbitrarily selected with a given value of k (5 or 10 in this study) and should, therefore, select models that are more reliable than those selected by the one validation series method where the models could be different with different validation series. However, for the series used here, the number of parameters it selects might not be parsimonious, given the fact that the 12 monthly means are an addition to each number of parameters selected. The flatness of the validation series mean error curves for the first few parameters as indicated in Figure 3.9 suggests that the significant number of parameters may well be lower and close to that selected by the AIC criterion in each case. That being the case, the AIC criterion may be more appropriate as a model selection criterion for these runoff series, as there is no ambiguity in the number of significant parameters it selects.

The observed and predicted estimation and validation series for the models selected based on the AIC are plotted in Figure 3.11 for some of the stations. These plots and the NSE values shown in Table 3.5, particularly those for the validation series, indicate that in general, the autoregressive models are not very good at representing the flows in the basin (except in one case, i.e., Manimenso, station 3 in Table 3.5). These models cannot be reliably used for prediction of future flows in the basin, since the results show that they lack the necessary information on such flows. Rainfall input would be necessary to adequately model these flows.

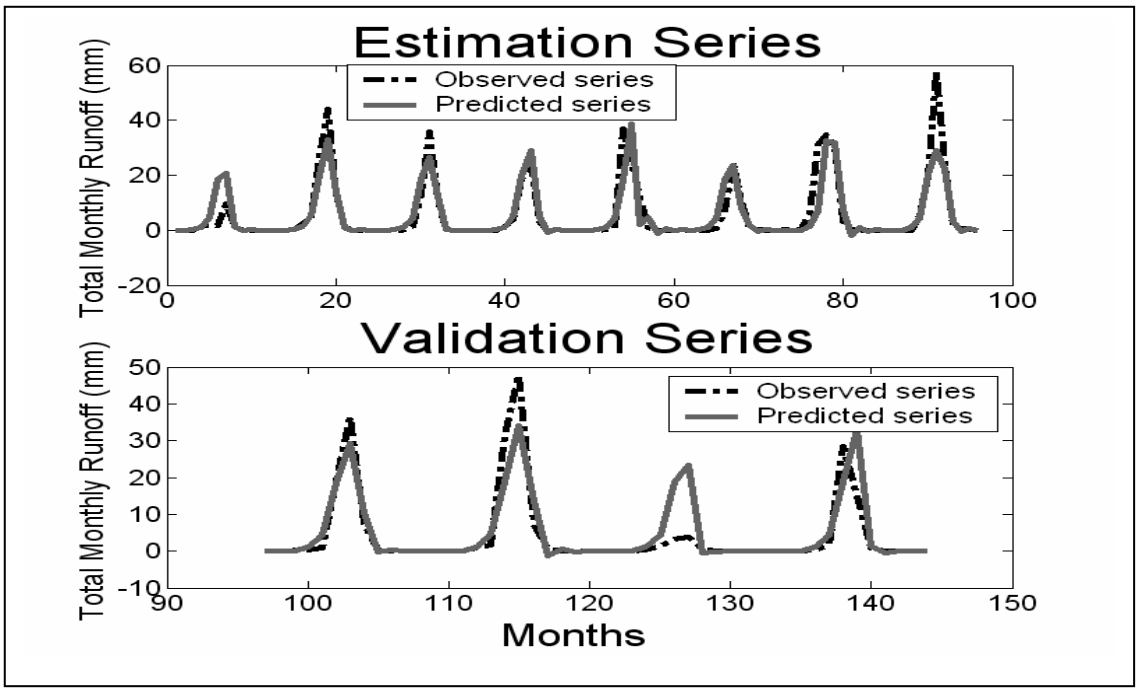


(a)

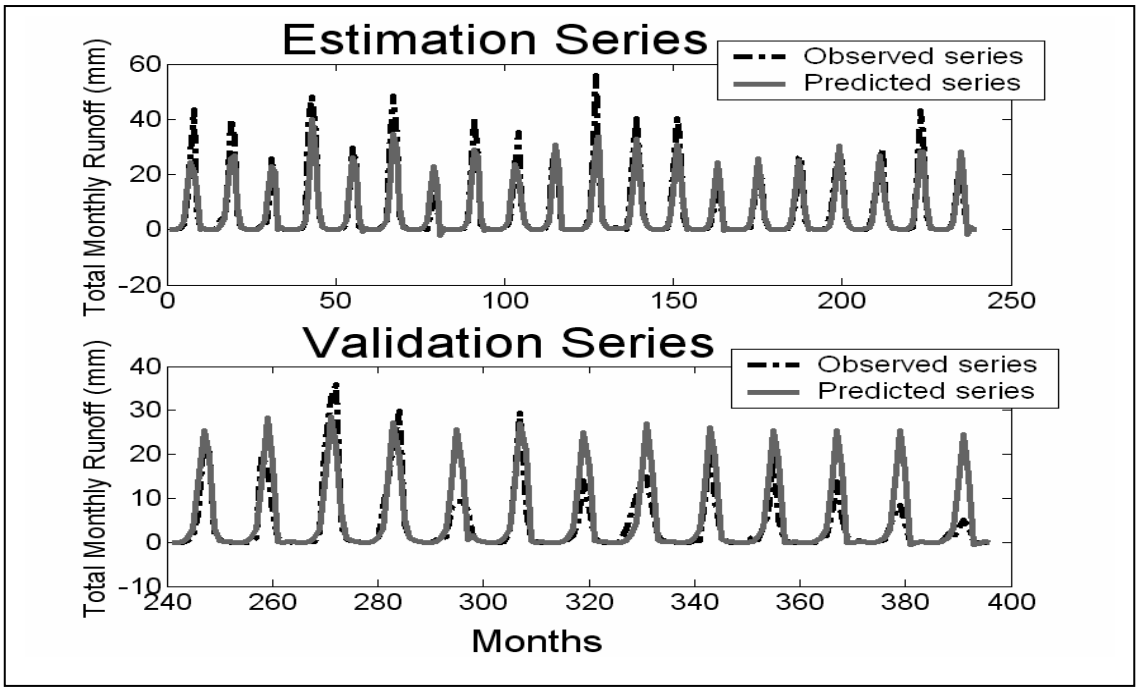


(b)

Figure 3.11 Observed and predicted total monthly runoff for AIC validated ARMA model
(a) Dapola on the Black Volta River ((na,nc) = (1,1))
(b) Bamboi on the Black Volta River ((na,nc) = (1,4))



(c)



(d)

Figure 3.11 Continued
(c) Wiasi on the Sisilli River (Black Volta Basin) ((na,nc) = (1,1))
(d) Porga on the Oti River ((na,nc) = (1,1))

3.2.3 Non-linearity in monthly rainfall-runoff relationship

The nonlinearity in the rainfall-runoff relationship is next investigated by determining the suitability of a linear time invariant input-output model representation of the relationship. The model considered is the black box ARX (Autoregressive with eXogenous input) model given as (Box *et al*, 1994; Ljung, 2003):

$$A(q)y_t = B(q)x_{t-nk} + e_t \tag{3.9}$$

where:

$\boldsymbol{y}_t$ = monthly river runoff at time t months

$\boldsymbol{x}_t$ = monthly catchment rainfall at time t months

$\boldsymbol{e}_t$ = random error at time t months

$$A(q) = (1 - a_1q - a_2q^2 - \ldots\ldots - a_{na}q^{na})$$
$$B(q) = (b_1q + b_2q^2 + \ldots\ldots + B_{nb}q^{nb})$$

q = backward operator, i.e., $qy_t = y_{t-1}$

$a_1, a_2, .., a_{na}$ are the parameters of the AR component of the model to be estimated

$b_1, b_2, .., b_{nc}$ are the parameters of the input component of the model to be estimated

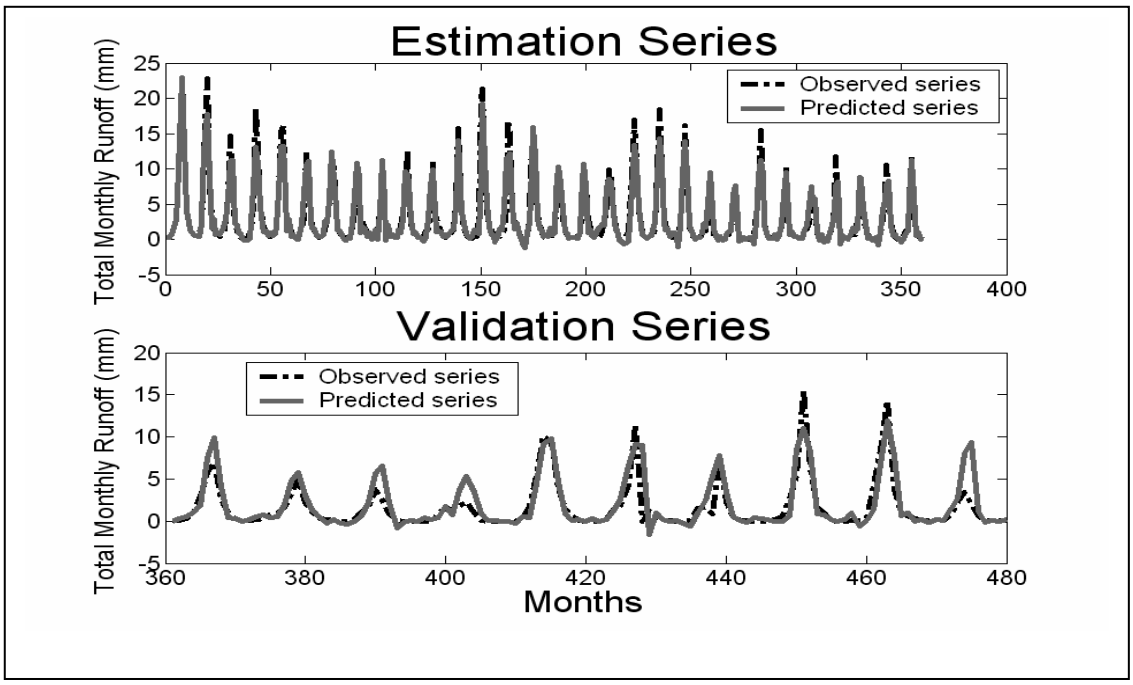
na = order of the AR model

nb = order of the input model

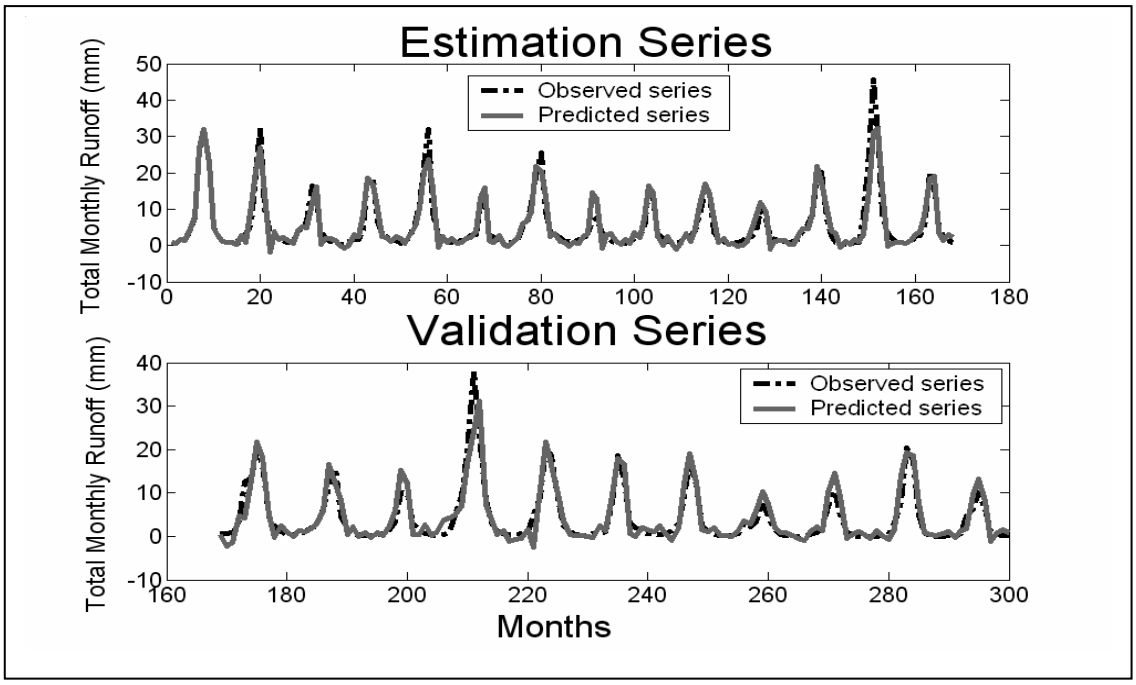
nk = input delay in months

Each model is thus represented by the values [na nb nk]. The Systems Identification toolbox of Matlab (Ljung, 2003) was used to identify and estimate the models.

Results of the ARX modelling activities are summarized in Figure 3.12 for some of the stations and in Table 3.6 for all the stations. It can be seen from the plots and the validation NSE values in Table 3.6 that although most of the fitted models have large numbers of parameters (15 or more) they still do not provide very good simulations.

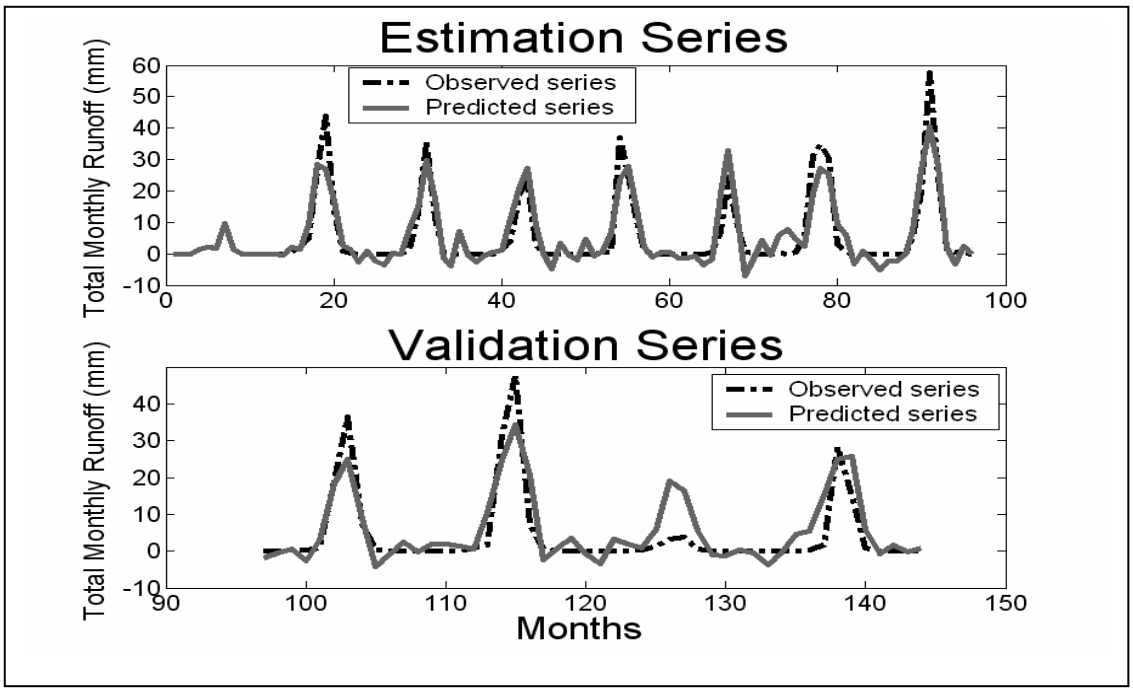


(a)

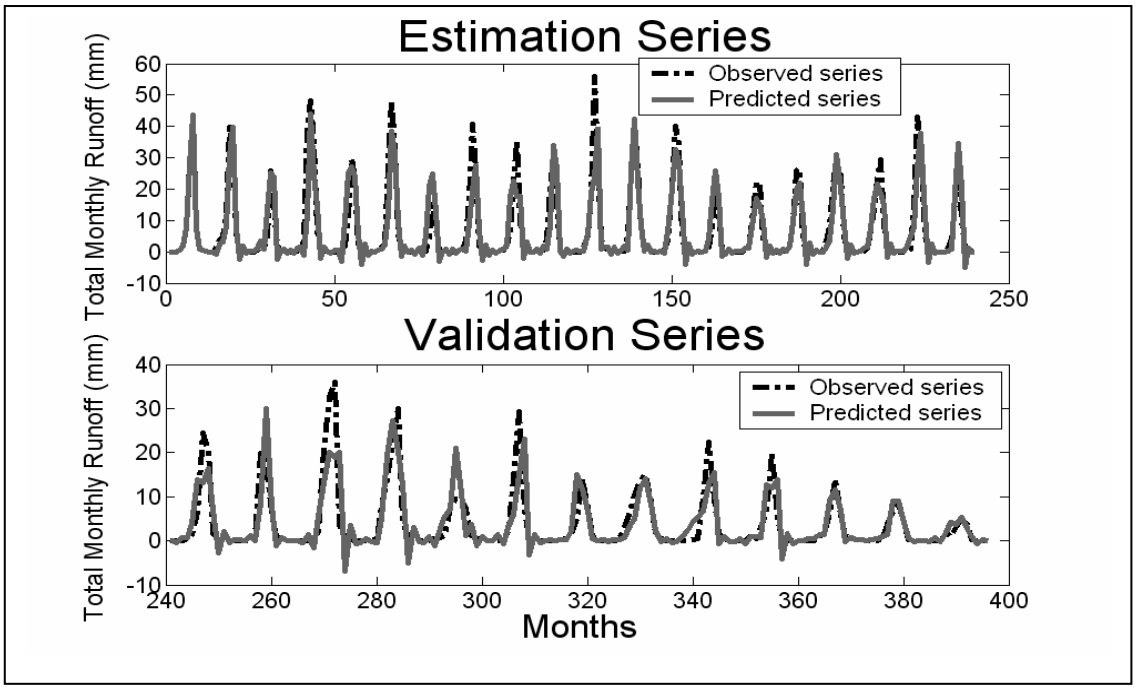


(b)

Figure 3.12 Observed and predicted total monthly runoff - ARX model
(a) Dapola on the Black Volta River [na nb nk] = [14 15 0]
(b) Bamboi on the Black Volta River [na nb nk] = [14 15 0]



(c)



(d)

Figure 3.12 Continued
(c) Wiasi on the Sisilli River (White Volta Basin)[na nb nk] = [1 14 0]
(d) Porga on the Oti River [na nb nk] = [14 5 5]

Table 3.6 Estimation and validation NSE values for the ARX Models

Station	River	Model [na nb nk]	NSE (%) Estimation Series	NSE (%) Validation Series
Black Volta Basin				
1. Banzo	Black Volta	[8 1 0]	74.08	71.53
2. Nwokuy	Black Volta	[13 1 0]	89.90	81.07
3. Manimenso	Black Volta	[14 3 0]	96.20	87.02
4. Tenado	Black Volta	[1 1 0]	83.00	70.59
5. Boromo	Black Volta	[1 1 0]	82.08	77.50
6. Debougou	Bougouriba	[13 1 0]	79.55	61.62
7. Lawra	Black Volta	[13 14 0]	89.57	78.44
8. Dapola	Black Volta	[14 15 0]	86.41	66.01
9. Bui	Black Volta	[1 15 0]	85.43	85.54
10. Bamboi	Black Volta	[14 15 0]	87.29	85.00
White Volta Basin				
11. Wiasi	Sissili	[1 14 0]	82.31	71.19
12. Yagaba	Kulpawn	[13 4 6]]	64.31	56.71
13. Nasia	Nasia	[5 10 3]	86.92	62.30
14. Nabogo	Nabogo	[12 8 8]	75.41	58.56
Oti Basin				
15. Porga	Oti	[14 5 5]	78.87	61.49
16. Mango	Oti	[14 14 0]	89.14	83.38
17. Koumangou	Koumangou	[15 3 0]	84.28	79.02
18. Sabari	Oti	[15 13 0]	90.41	88.02

The large number of parameters needed means that the linear models are unsuitable for representing the rainfall-runoff process, and nonlinear models should be considered. Appropriate nonlinear models should provide better predictions of the validation series with fewer than 15 parameters.

3.3 Conclusions

Rainfall and runoff variability in the Volta Basin, both temporally and spatially, is high. Most rainfall and runoff in the basin occurs within 3 to 4 months of the year. Potential evapotranspiration is also very high, particularly in the dry season. These in combination mean that riverflow in the dry season is low and unreliable and that groundwater contribution to streamflow is not substantial. There is no persistence in annual runoff in the basin, so runoff at this temporal scale can be considered as largely independent. The inadequacy of the linear models fitted to the monthly rainfall-runoff series have shown that modelling the rainfall-runoff process in this basin is a non-linear estimation problem.

Nonlinear rainfall-runoff modelling is described in the next chapters starting with a modelling framework suitable for data infilling to daily runoff series in the basin.

4 SPATIO-TEMPORAL MODEL FOR FILLING GAPS IN DAILY STREAM FLOW SERIES[2]

4.1 Introduction

A major requirement for the assessment, development and sustainable use of the water resources of any river basin is the availability of good quality runoff series of sufficiently long duration. In the Volta Basin, both daily and monthly river discharge series exist for a good number of gauging stations. However, many of these records are of poor quality and contain gaps of from several days to several years.

In an assessment of monthly flow series of river discharge from the main river gauging stations in the basin, Taylor (2003) observed that in general 20% of monthly discharge data over a 20-year period are missing from the available series in the basin, with some gauges having as many as 50% gaps in their series. By regressing rainfall with the various series, the above study determined that only half of the gauging stations examined had reliable flow series, though these also had gaps of varying lengths.

Filling gaps in existing river flow series in the Volta Basin is, therefore, a necessary and essential exercise if river runoff is to be satisfactorily modeled. This is because a full series of river runoff data is required to calibrate and test any models designed for predicting river flows in the basin.

Several methods are available for data infilling in general and for hydrological data infilling in particular. Data imputation methods (Dempster *et al*., 1977, Schafer 1997, Little and Rubin 2003) are generally difficult to apply to hydrological data such as monthly river runoff series because of autocorrelations at high lags and seasonal effects. A few reviews are available on methods that have been used successfully for hydrological data infilling (Kottegoda and Elgy1979, Gyau-Boakye and Schultz, 1994). Gyau-Boakye and Schultz (1994) have provided a framework for filling in gaps of various lengths in daily runoff series in West Africa including the Volta Basin. Among methods recommended for such data infilling are autoregression with or without rainfall, simple and multiple regressions with neighboring gauges, interpolation, recession methods and linear storage model formulations, and the method used depends,

[2] This chapter has been published in modified form in Hydrology and Earth System Sciences, 9, 209–224, 2005

among others, on the length of the data gaps to be filled and the season in which these gaps occur. Papadakis *et al* (1993) have also demonstrated the strength of satellite imagery with non-linear modelling in stream flow generation. Taylor *et al* (2004) used the Thornthwaite-Mather (TM) method to model river runoff in the Volta Basin and concluded that the method could be used to adequately fill in gaps in the runoff series in the basin if properly formulated.

Autoregressions without rainfall, simple and multiple linear regressions with neighboring gauges, interpolation and recession methods of data infilling, where they work, have the advantages of simplicity and not requiring any other input data such as rainfall, evaportranspiration, soil moisture status etc, that other models need. However, they still need separate complete and extensive data sets for their calibration and verification, a requirement that may be a dream in much of the Volta Basin. Recession and autoregressive methods are unsuitable for periods of flow with rainfall, as they ignore the effect of the driving rainfall input. Multiple regressions account for the rainfall input and catchment moisture status by their use of the runoff from neighboring gauges and so are more suitable for data infilling when there is rainfall. However, by being fitted to "global" data (runoff series for several years together), they may not provide very good estimates for short "local" gaps. Also, since in applying them to estimate missing values at a gauging station, the available observed runoff at the station for the period considered is completely ignored, optimum use of available runoff information is not being made. This means that the quality of the estimates would be sub-optimal.

The gap-filling method through the use of spatio-temporal dynamic models presented in this chapter makes optimum use of all available spatial and temporal information. Spatio-temporal dynamic models have been applied successfully to environmental systems (Shumway and Stoffer, 1982; Hasket, 1989; Rouhani and Myers, 1990; Goodall and Mardia, 1994; Guttorp and Sampson, 1994; Mardia et al., 1998). When cast in state-space form, they can be used with the Kalman smoother and the Expectation-Maximization (EM) algorithm to estimate missing values in environmental data including runoff data. Formulated this way, they have the advantage of not needing separate calibration data, as the use of the EM algorithm ensures that model parameters, missing observations and state vectors can be adequately estimated

with the same data set and concurrently. Another advantage is that available "local " measured streamflow data within the modelling period are also used in the estimation process, thereby ensuring that important information contained in these data is fully and optimally used.

In this chapter, the spatio-temporal state space dynamic model with time invariant parameters is presented. The presentation is in three parts. The first part is the model formulation using the Kalman smoother and the EM algorithm. This is followed by an outline of the necessary modifications that have to be made in the model framework to take care of missing values and to allow for the inclusion of constraints and prior knowledge of model structure and parameters. The final part is the application of the model to filling short gaps (a few days to a month) in daily riverflow data in the Volta Basin of West Africa using series lengths of up to one year. The aim of this modeling exercise was to ascertain the applicability and effectiveness of such a model to runoff patching in the basin. A typical daily river runoff time series with gaps is given in Figure 4.1.

4.2 The discrete spatio-temporal dynamic modelling framework

The discrete spatio-temporal dynamic model is formulated to predict an n x 1 state vector $\boldsymbol{x}_t=(x_{\alpha 1}(t), x_{\alpha 2}(t), \ldots, x_{\alpha n}(t))'$ of an unobserved spatio-temporal state process at a fixed network of n locations. In addition, there is the m x 1 vector $\boldsymbol{y}_t=(y_{\beta 1}(t), y_{\beta 2}(t), \ldots, y_{\beta m}(t))'$ of observed or measured values at m locations at time t, where the two sets of spatial locations need not be the same (Xu and Wikle, 2004). Here, and throughout, the prime marks the transpose vector or matrix. Bold small letters indicate vectors and bold capitals are matrices. The state-space representation for the prediction of the unobserved process without external input and for the linear dynamic case with time invariant parameters consists of the following process/state and measurement equations:

$$x_t = Fx_{t-1} + \omega_t, \qquad x_0 \sim N(\mu_0, \Sigma_0), \qquad \omega \sim N(0, Q) \tag{4.1a}$$

$$y_t = Hx_t + \upsilon_t, \qquad \upsilon \sim N(0, R) \tag{4.1b}$$

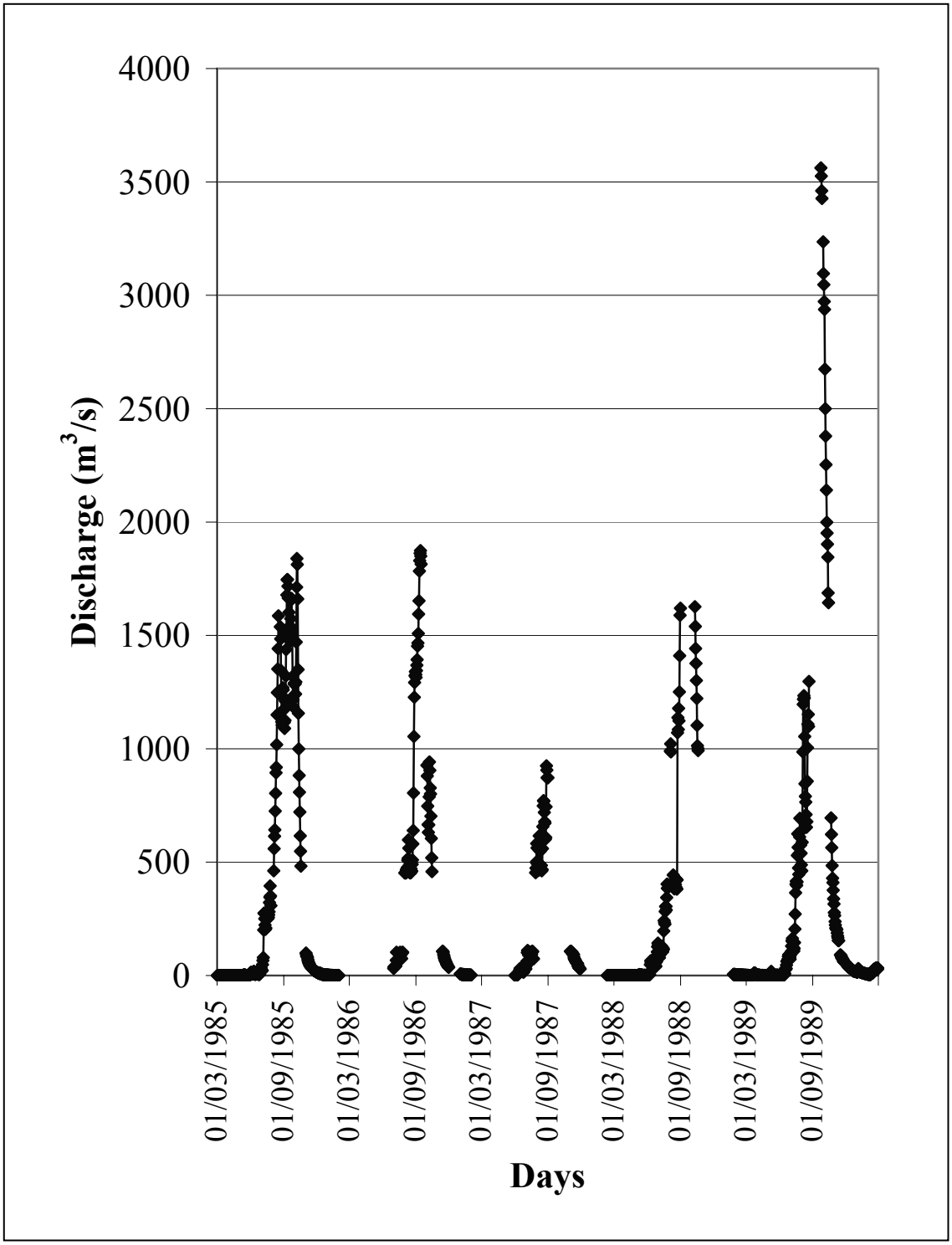


Figure 4.1 Daily hydrograph at Sabari on the Oti River (1985-1989). Typical pattern of gaps considered in the spatio-temporal modelling

F is the n x n transition or state propagation matrix that describes the dynamics of the system - a first-order Markov process. H is the m x n measurement matrix that relates the estimated state vectors to the vector of actual observations. The additive n x 1 state estimation errors, $\boldsymbol{\omega_t}$, and the m x 1 measurement errors, $\boldsymbol{\upsilon_t}$,, are uncorrelated Gaussian white noises with zero mean and covariance Q (n x n) and R (m x m). For convenience, the initial n x 1 state vector, x_0, is considered normally distributed with mean μ_0 and covariance Σ_0. Equation 4.1 is the finite dimensional linear dynamical system from which the vector time series of observations $Y = (y_0, y_1, y_2, \ldots, y_N)$ is assumed to be generated.

As noted in Ribeiro (2004), when x_0 is a Gaussian vector, ω_t and υ_t are Gaussian white noises and when the state and observation dynamics are linear, the conditional probability density function $p(x_t|Y)$ is normally distributed, i.e., $p\left(x_t \mid Y\right) \sim N\left(x_t^N, P_t^N\right)$. The conditional mean of this Gaussian probability density function is equivalent to the estimate $\boldsymbol{x_t^N}$ of the state x_t given the N observations at each of the m sites. The covariance matrix P_t^N quantifies the uncertainty of the state estimate given the same N observations.

In general, some or all of the system parameters $\Theta = \{F, H, Q, R, \mu_0, \Sigma_0\}$ are not known and will have to be estimated from the observations. This is a system identification problem and, in the Gaussian framework under consideration, the parameter estimation can be undertaken by the method of maximum likelihood. The maximum likelihood estimate of Θ given $X = (x_0, x_1, x_2, \ldots, x_N)$ and $Y = (y_0, y_1, y_2, \ldots, y_N)$ is obtained by maximizing the joint log-likelihood of X, Y, and Θ with respect to Θ. This log-likelihood function is given as (Shumway and Stoffer, 1982):

$$logL_Y\left(\Theta\right) = logL\left(X,Y,\Theta\right) = -\frac{1}{2}\left\{\begin{array}{l} log\left|\Sigma_0\right| + \left(x_0 - \mu_0\right)' \Sigma_0^{-1}\left(x_0 - \mu_0\right) \\ +Nlog\left|Q\right| + \sum_{t=1}^{N}\left(x_t - Fx_{t-1}\right)' Q^{-1}\left(x_t - Fx_{t-1}\right) \\ +Nlog\left|R\right| + \sum_{t=1}^{N}\left(y_t - Hx_t\right)' R^{-1}\left(y_t - Hx_t\right) \end{array}\right\} \quad (4.2)$$

When there are no constraints placed on the structure of the system matrices F, H, Q, and R, the estimates of the components of Θ are (Digalakis *et al.*, 1993; Xu and Wikle, 2004):

$$\hat{F} = A_4 A_3^{-1} \tag{4.3a}$$

$$\hat{H} = A_6 A_1^{-1} \tag{4.3b}$$

$$\hat{Q} = A_2 - A_4 A_3^{-1} A_4' \tag{4.3c}$$

$$\hat{R} = A_5 - A_6 A_1^{-1} A_6' \tag{4.3d}$$

$$\mu_0 = x_0 \tag{4.3e}$$

$$\Sigma_0 = P_0 \tag{4.3f}$$

where:

$$A_1 = \frac{1}{N+1} \sum_{t=0}^{N} x_t x_t' \tag{4.4a}$$

$$A_2 = \frac{1}{N} \sum_{t=1}^{N} x_t x_t' \tag{4.4b}$$

$$A_3 = \frac{1}{N} \sum_{t=1}^{N} x_{t-1} x_{t-1}' \tag{4.4c}$$

$$A_4 = \frac{1}{N} \sum_{t=1}^{N} x_t x_{t-1}' \tag{4.4d}$$

$$A_5 = \frac{1}{N+1} \sum_{t=0}^{N} y_t y_t' \tag{4.4e}$$

$$A_6 = \frac{1}{N+1} \sum_{t=0}^{N} y_t x_t' \tag{4.4f}$$

are known as the sufficient statistics.

In the state-space formulation of interest, the set of state vectors X (N x n) are not observed directly and not available a priori for the computation of the sufficient statistics and hence the parameter estimates. In addition, some of the observations in the set Y (N x m) may be missing. In these circumstances, the Expectation-Maximisation or

EM algorithm (Dempster et al., 1977) has been found to be a powerful tool for the maximum likelihood estimation of the system parameters (Shumway and Stoffer, 1982; Digalakis et. al., 1993; Ghahramani and Hinton, 1996; Bilmes, 1998; Xu and Wikle, 2004).

The EM algorithm is designed for parameter estimation of incomplete or missing data problems by the method of maximum likelihood (Dempster et. al., 1997). By treating the state vector as missing observations, the problem is now the same as a problem with incomplete data, which justifies the use of the EM algorithm. Thus, both the recursions for the computations of the state vector, such as by the Kalman filter used here, and the estimation of the model parameters can be undertaken concurrently and offline computations of the parameters is not necessary. The EM procedure involves computing the model parameter set Θ and then the state vector over and over again in a series of iterations until a set of convergence conditions is met. The computations for each iteration are carried out in two main steps, the E-step and the M-step.

Consider the $(r+1)^{th}$ iteration when $\Theta^{(r)}$, the parameter set at the r^{th} iteration, is known and it is required to find $\Theta^{(r+1)}$, the parameter set at the $(r+1)^{th}$ iteration. In the E-step of the EM algorithm, the expected value of the complete-data log-likelihood $\log p(X, Y| \Theta)$ with respect to the unobserved and missing data X, given the observations Y and the current parameter estimates $\Theta^{(r)}$, is evaluated. This expectation is defined as:

$$\begin{aligned} G(\Theta^{(r+1)}) = G(\Theta^{(r+1)},\Theta^{(r)}) &= E[\log p(X,Y|\ \Theta^{(r+1)})|\ Y,\ \Theta^{(r)}] \\ &= E[L(X,\ Y,\ \Theta^{(r+1)}\ |Y,\ \Theta^{(r)}] \end{aligned} \tag{4.5}$$

The expectation is evaluated with the known parameter set $\Theta^{(r)}$ (the new parameter set $\Theta^{(r+1)}$ is obtained by optimising G in the M step). The set Y, excluding any missing observations, constitutes the incomplete data set, while the full data set (X, Y) contains both observed and missing data.

In the M-step, $\Theta^{(r+1)}$, the new estimate of Θ, is computed by maximising the conditional expectation evaluated in the E-step, i.e.,

$$\Theta^{(r+1)} = \underset{\Theta}{argmax}\, G\left(\Theta^{(r+1)}\right) \tag{4.6}$$

The log-likelihood is guaranteed to increase with each iteration and the algorithm is guaranteed to converge to at least a local minimum (Dempster, 1977; Bilmes, 1998).

For the regular exponential distributions (e.g., normal, binomial, poisson, gamma distributions), the E-step of the EM algorithm consists of the computation of the conditional expectations of the complete data sufficient statistics as given in Equation 4.4 (Dempster, 1977). In the M-step, these conditional expectations of the complete-data sufficient statistics are then used instead of the (unknown) actual complete-data sufficient statistics. Thus, the following quantities are computed in the E-step and used to evaluate the sufficient statistics in Equation 4.4, (Digalakis, 1993):

$$E\left\{x_t \mid Y,\Theta^{(r)}\right\} = x_t^N \tag{4.7a}$$

$$E\left\{x_t x_t' \mid Y,\Theta^{(r)}\right\} = P_t^N + x_t^N \left(x_t^N\right)' \tag{4.7b}$$

$$\begin{aligned} E\left\{x_t x_{t-1}' \mid Y,\Theta^{(r)}\right\} &= E\left\{\left(x_t - x_t^N\right)\left(x_{t-1} - x_{t-1}^N\right)' \mid Y\right\} + x_t^N \left(x_{t-1}^N\right)' \\ &= P_{t,t-1}^N + x_t^N \left(x_{t-1}^N\right)' \end{aligned} \tag{4.7c}$$

$$E\left\{y_t x_t' \mid Y,\Theta^{(r)}\right\} = y_t E\left\{x_t' \mid Y,\Theta^{(r)}\right\} = y_t \left(x_t^N\right)' \tag{4.7d}$$

$$E\left\{y_t y_t' \mid Y,\Theta^{(r)}\right\} = y_t y_t', \text{ (no missing values in Y)} \tag{4.7e}$$

The required statistics in equations 4.7a-4.7d above at iteration r+1 can be computed for all t = 1, 2, … , N from the fixed interval Kalman smoother (also known as Rauch-Tung-Striebel or RTS smoother (S. Haykin, 2001)) using the parameter estimates obtained at iteration r. The smoother is given as (S. Haykin, 2001; Xu and Wikle, 2004):

Filter equation – Forward pass (t = 1, 2, …, N)

a. Prediction:

$$x_t^{t-1} = F x_{t-1}^{t-1} \tag{4.8a}$$

$$P_t^{t-1} = F P_{t-1}^{t-1} F' + Q \tag{4.8b}$$

with $$x_0^0 = \mu_0 \text{ and } P_0^0 = \Sigma_0 \tag{4.8c}$$

b.Update (filter):

$$e_t = y_t - Hx_t^{t-1} \tag{4.9a}$$

$$\Sigma_{et} = HP_t^{t-1}H' + R \tag{4.9b}$$

$$K_t = P_t^{t-1}H'\Sigma_{et}^{-1} \tag{4.9c}$$

$$x_t^t = x_t^{t-1} + K_t e_t \tag{4.9d}$$

$$P_t^t = P_t^{t-1} - K_t H P_t^{t-1} \tag{4.9e}$$

Smoothing – Backward pass (t =N, N-1, ..,1)

$$x_N^N = x_t^t, P_N^N = P_t^t, \text{ t = N only} \tag{4.10a}$$

$$J_{t-1} = P_{t-1}^{t-1}F'\left(P_t^{t-1}\right)^{-1} \tag{4.10b}$$

$$x_{t-1}^N = x_{t-1}^{t-1} + J_{t-1}\left(x_t^N - x_t^{t-1}\right) \tag{4.10c}$$

$$P_{t-1}^N = P_{t-1}^{t-1} + J_{t-1}\left(P_t^N - P_t^{t-1}\right)J_{t-1}' \tag{4.10d}$$

Smoothed lag-one covariance (t = N, N-1, …, 2)*:*

$$P_{N,N-1}^N = \left(I - K_N H\right)FP_{N-1}^{N-1}, \text{ t=N only} \tag{4.10e}$$

$$P_{t-1,t-1}^N = P_{t-1}^{t-1}J_{t-2} + J_{t-1}\left(P_{t,t-1}^N - FP_{t-1}^{t-1}\right)J_{t-2}' \tag{4.10f}$$

In the above equations, x_t^{t-1}, x_t^t, x_t^N and P_t^{t-1}, P_t^t, P_t^N are the predicted, filtered (updated) and smoothed values, respectively, of the state vector $\boldsymbol{x_t}$ and its covariance P_t. The values of interest are the smoothed values x_t^N and P_t^N, which are inserted on the right-hand side of equations 7a-e to calculate the expected values needed for the M-step. The log-likelihood function can also be conveniently computed as a by-product of the Kalman filter as follows:

$$logL_Y\left(\Theta\right) = -\frac{1}{2}\left\{\sum_{t=1}^{N} log\left|\Sigma_{et}\right| + \sum_{t=1}^{N} e_t'\Sigma_{et}^{-1}e_t\right\} \tag{4.11}$$

where $\boldsymbol{e_t}$ and $\boldsymbol{\Sigma_{et}}$, the innovation vector and innovation covariance matrix, respectively, are computed for t =1, 2, …, N as in equations 4. 9a and 4.9b.

The EM algorithm for the maximum likelihood estimation of the linear time invariant dynamic system represented in (4.1) can now be summarised as follows:

E-step at iteration r+1

(i) Use the parameter set $\Theta^{(r)} = \{F^{(r)}, H^{(r)}, Q^{(r)}, R^{(r)}, \mu_0^{(r)}, \Sigma_0^{(r)}\}$ obtained from the previous iteration, r, and the Kalman smoother presented in equations 4. 8 – 4.10 to compute the statistics in Equation 4.7. Also compute the log-likelihood, $L_Y(\Theta^{(r+1)})$ using Equation 4.11

(ii) Use the statistics computed in (i) to compute the conditional expectations of the sufficient statistics in Equation 4.4

M-step at iteration r+1

Re-estimate (update) the parameter set as $\Theta^{(r+1)} = \{F^{(r+1)}, H^{(r+1)}, Q^{(r+1)}, R^{(r+1)}, \mu_0^{(r+1)}, \Sigma_0^{(r+1)}\}$ using the relationships in (4.3), with the conditional expectations of the sufficient statistics calculated in (ii) under the E-step in equations 4.3a-d, x_0 in Equation 4.3e, and P_0 in Equation 4.3f.

Convergence

Test for the convergence of either the parameters or the log-likelihood, i.e., perform one of the following tests:

$$\left\| \Theta^{(r+1)} - \Theta^{(r)} \right\| < \varepsilon_\theta \tag{4.12a}$$

$$\left\| L_Y\left(\Theta^{(r+1)}\right) - L_Y\left(\Theta^{(r)}\right) \right\| < \varepsilon_L \tag{4.12b}$$

where ε_θ and ε_L are sufficiently small positive numbers.

If the test succeeds, the iterations are stopped and $\boldsymbol{\Theta}^{(r+1)}$ is retained as the final set of estimates of the system parameters, otherwise the iterations continue. Xu and Wikle (2004) prefer the use of the parameter values as the test criterion (Equation 4. 12a) to the use of the log-likelihood (Equation 4.12b) as the log-likelihood can be

unstable for spatio-temporal problems due to the high spatial correlations in the innovations as a result of processes at adjacent spatial locations often being very similar.

4.3 Missing observations

If there are missing observations, then y_t and H in the update equations of the forward pass of the Kalman filter would have to be modified as follows:

(i) Replace all missing values in y_t in Equation 4.9a with zeroes (observed zero values of y_t may be represented as very small numbers such as 0.0001).

(ii) Replace entries in the corresponding rows in H in Equation 4.9a with zeroes.

In addition, the conditional expectations given in Equation 4.7 would have to be modified as follows (Digalakis et al., 1993):

$$E\left\{y_t \mid Y,\Theta^{(r)}\right\} = \begin{Bmatrix} y_t & , \quad if\ observed \\ H^{(m)} E\left\{x_t \mid Y,\Theta^{(r)}\right\}, & if\ missing \end{Bmatrix} \tag{4.13a}$$

$$E\left\{y_t y_t^{'} \mid Y,\Theta^{(r)}\right\} = \begin{Bmatrix} y_t y_t^{'}, & if\ observed \\ R^{(m)} + H^{(m)} E\left\{x_t x_t^{'} \mid Y,\Theta^{(r)}\right\}\left(H^{(m)}\right)^{'}, & if\ missing \end{Bmatrix} \tag{4.13b}$$

$$E\left\{y_t x_t^{'} \mid Y,\Theta^{(r)}\right\} = \begin{Bmatrix} y_t E\left\{x_t \mid Y,\Theta^{(r)}\right\}, & if\ observed \\ H^{(m)} E\left\{x_t x_t^{'} \mid Y,\Theta^{(r)}\right\}, & if\ missing \end{Bmatrix} \tag{4.13c}$$

where $H^{(m)}$ and $R^{(m)}$ are the H and R matrices, respectively, corresponding to the missing values.

4.4 System matrices parameterisation

Equation 4.3 is used in the M-step to update the parameter values, when these parameters are not constrained or parameterised in any way. The parameters thus obtained are the maximum likelihood estimates. To avoid identifiability problems, some or all of the system matrices may be constrained or parameterised directly. In such cases, the parameters are no longer maximum likelihood estimates. However, the

parameterisation can be undertaken in such a way as to still result in an increase in the log likelihood at each iteration and lead to a convergence of the parameters. The algorithm is then called the General EM (GEM) (Xu and Wikle, 2004). Most often, Q is constrained to a diagonal matrix while R is modelled as $R = \sigma^2 I_m$, where I_m is an m x m identity matrix. The process matrix F can also be parameterised if its form and structure are known, while H may be specified a priori as a design matrix and would no longer be updated in the M-step.

In the GEM, the $(r+1)^{th}$ update formula for the general unconstrained Q, whether F is parameterised or not, is given as (Xu and Wikle, 2004):

$$Q^{(r+1)} = \frac{1}{N} A, \quad where\ A = \left(A_2 - A_4 F' - F A_4 + F A_3 F' \right) \tag{4.14a}$$

For the case when F is not parameterised and is estimated as in relation 4.3a, Equation 4.14a reduces to Equation 4.3c. For a diagonal Q matrix, its M-step update is:

$$diag\left(Q^{(r+1)}\right) = \frac{1}{N} diag\left(A\right) \tag{4.14b}$$

where diag (A) is the diagonal vector of A.

When R is parameterised as $R = \sigma^2 I_m$, σ^2 is updated in the M-step as follows:

$$\sigma^{2(r+1)} = \frac{1}{Nm} \sum_{t=1}^{N} tr\left\{ \left(y_t - H x_t^N \right) \left(y_t - H x_t^N \right)' + H P_t^N H' \right\} \tag{4.14c}$$

The relations in Equation 4.14 ensure that the log likelihood increases monotonically even though the final parameter estimates would not be maximum likelihood estimates.

4.5 Application of the modelling framework

The methodology developed here was designed to enable the estimation of short gaps of a few days to one month in the annual daily runoff series (temporal) at a given gauging station using the available observed runoff series of the same period measured at the station and at one or more other stations (spatial) in the same main sub-basin. Thus the

missing runoff data at a station in the Black Volta basin would be estimated using its available runoff data and those from other stations in the same main sub-basin for the year. The lengths of missing data considered in the study are comparable to the typical real gaps shown in the hydrograph in Figure 4.1.

Thus Equation 4.1 was used to model the runoff process both spatially, with discrete locations at the gauging stations, and temporally, with annual daily time series of runoff at the stations. The state vector, x_t, represents the unobserved actual catchment runoff and y_t is the vector with measured runoff at the m gauging stations at sampling time t. Both x_t and y_t are taken as m x 1 vectors of actual catchment and measured runoff series of length $N \leq 366$, i.e., x_t and y_t occur at the same locations. The process matrix F is unconstrained, $H = I_m$, Q is constrained to be diagonal and $R=\sigma^2 I_m$. At least one of the stations would have missing riverflow observations, the estimation of which is the main task at hand. Only gaps of a few days to a maximum of one month were considered. As convergence criterion, $\left\|\Theta^{(r+1)} - \Theta^{(r)}\right\| < 0.001$ was adopted. It should be mentioned that the results are relatively sensitive to the initial values used.

The model was applied to daily time series of riverflows of about one year measured at the stations Lawra (Black Volta River), Bui (Black Volta River), Bamboi (Black Volta River), Saboba and Sabari (Oti River) located as shown in Figure 3.1. By blacking out some of the observed data at a target gauging station, a maximum of 30 consecutive days of missing data in selected periods of the year were artificially created in the flow series of the target station. Three predictions of the missing data were made and compared by running the model with (i) the remaining samples of the target station as observed series (ii) the target station's series and the series from the rest of the stations in the same main sub-basin and (iii) the target station's series and that from only one other station in the same main sub-basin. Reliable daily flows for at least two stations for a full year could be obtained only for the main sub-basins of the Black Volta and Oti. The EM algorithm as described earlier not only provides estimates of the parameters but also those of the missing observations and state variables. It was used here to obtain estimates of the artificially created missing riverflow observations of the target stations. The performance of the model was evaluated in each case by the following Nash-Sutcliffe Efficiency (Nash and Sutcliffe, 1970) criterion:

$$NSE = 100\left(1 - \frac{\sigma_e^2}{\sigma_y^2}\right) \tag{4.15}$$

where σ_e^2 is the variance of the residuals and σ_y^2 the variance of the measured runoff at the target station.

4.6 Results and discussion

Figures 4.2a, 4.3a, and 4.4a are the observed runoff series used in the modelling exercise. They show varying degrees of correlation between the individual series in a plot. Figure 4.2a shows, for example, that there is a high correlation between the Bui and Bamboi flows but little correlation between the Lawra series and the others. The correlation between the two series in Figure 4.4a is not very good but not too poor either. As in Figure 4.2a, Figure 4.3a shows very good correlation between the two series plotted. A high correlation or lack of it in any set of runoff series could result in good or poor predictions of missing values in one series with the others as predictors.

Predicted missing values for the target station Bui in the Black Volta Basin are presented in figures 4.2b, 4.2c and 4.2d for the three cases (i), (ii) and (iii) above for the year 1964. The standard errors of the estimates for each case are also shown in the respective plots (at the bottom of the figures). As expected, these errors are small where observations are available and larger where they are missing. The “hills” and “valleys” in the error plots indicate clearly the ranges of the missing and observed flows.

Figures 4.2b and 4.2c and the relevant NSE values in Table 4.1 show that the use of the series from both Lawra and Bamboi, together with the remaining observations from Bui, provides better predictions of the missing values of Bui and with less uncertainty than using the observed series of Bui alone. The spatial interpolation obtained from the model is therefore adequate in this case. Figure 4.2d and Table 4.1 show that the use of the Bamboi series without that of Lawra did not reduce the quality of the predictions significantly, due to the fact that the correlation between the flows at Bui and at Lawra is low, as can be seen from the hydrographs in Figure 4.2a. Figures 4.3b and 4.3c show plots of predicted and observed flows at Bui for the year 1994 for cases (i) and (iii), again showing the good correlation between the Bui and Bamboi flows.

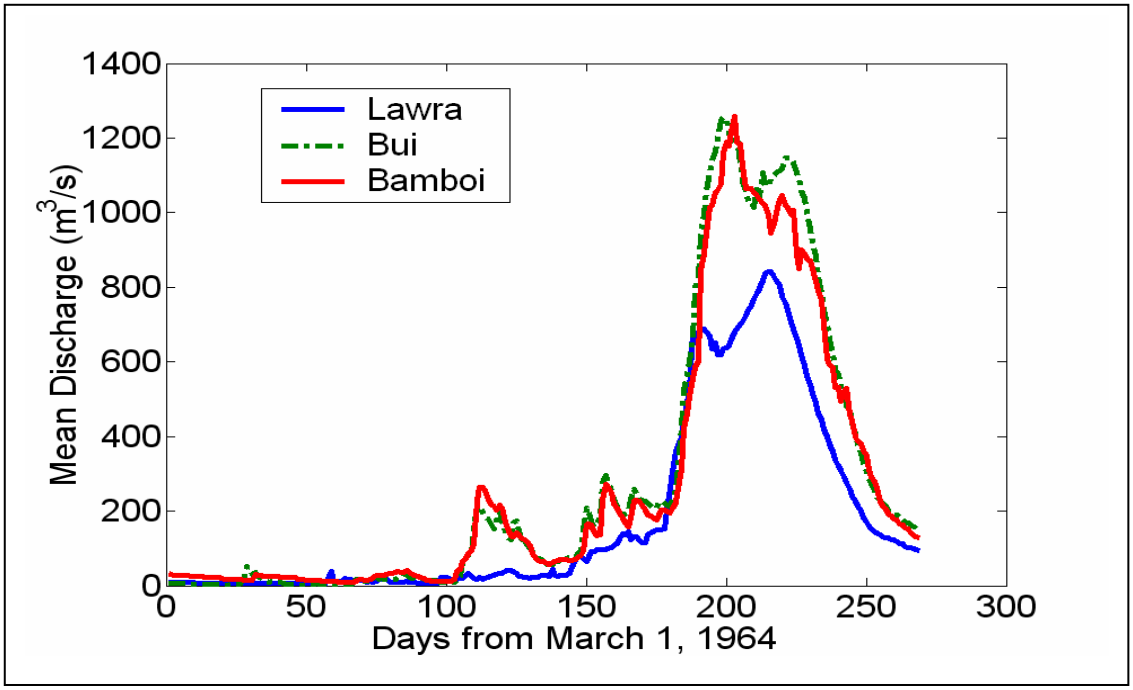


Figure 4.2a Daily hydrographs for stations Lawra, Bui and Bamboi on the Black Volta River (1964)

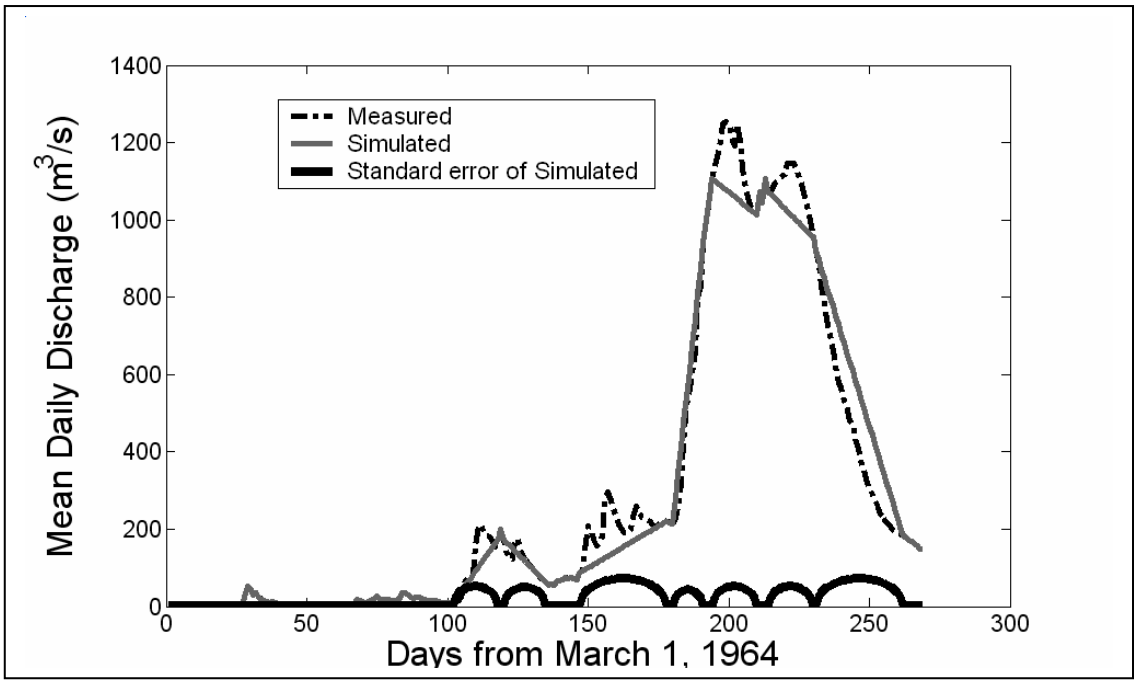


Figure 4.2b Measured and simulated daily hydrographs for station Bui on the Black Volta River; state space model on Bui discharge only (1964)

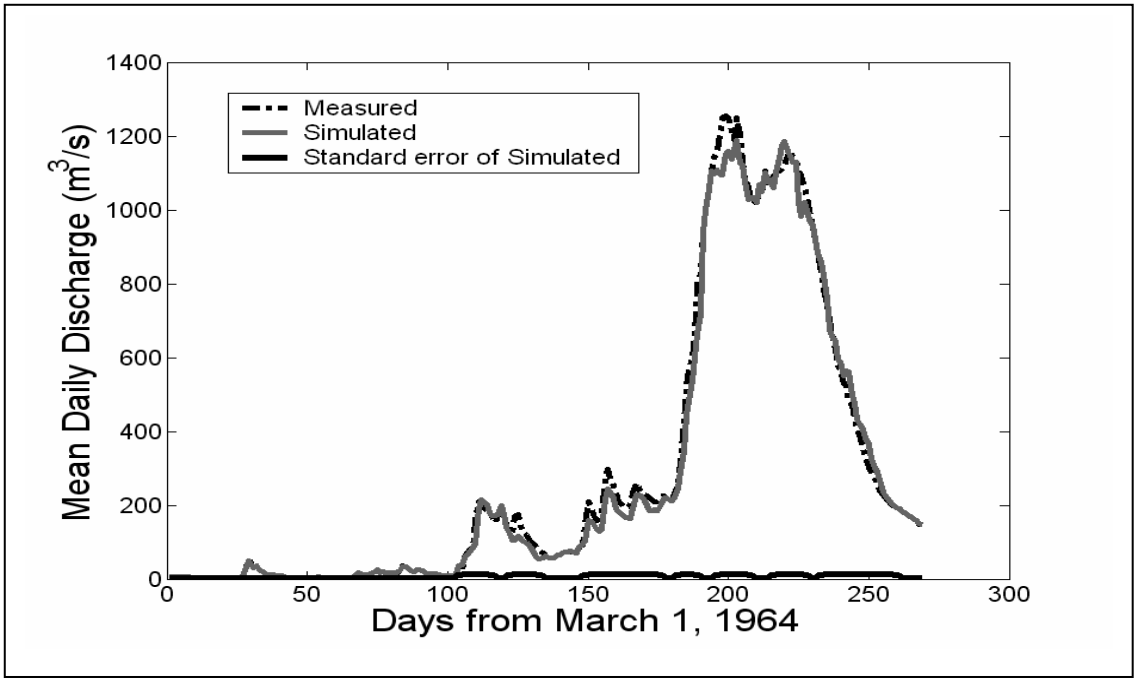


Figure 4.2c Measured and simulated daily hydrographs for station Bui on the Black Volta River; state space model on Bui, Lawra and Bamboi discharges (1964)

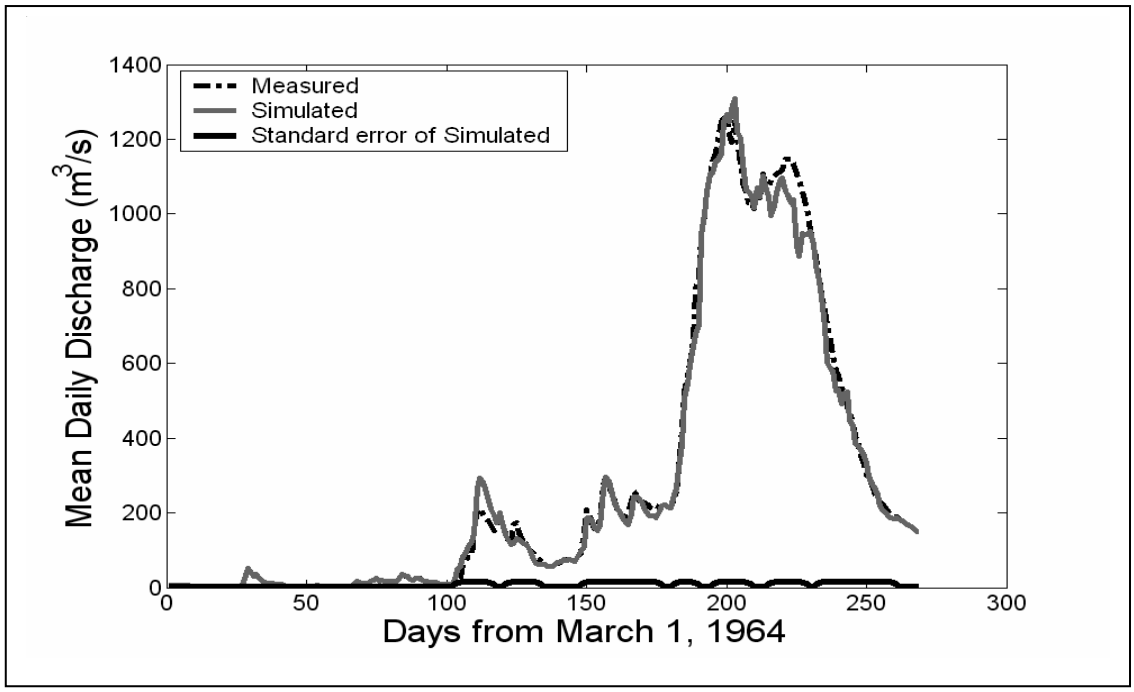


Figure 4.2d Measured and simulated daily hydrographs for Bui on the Black Volta River; state space model on Bui and Bamboi discharges (1964)

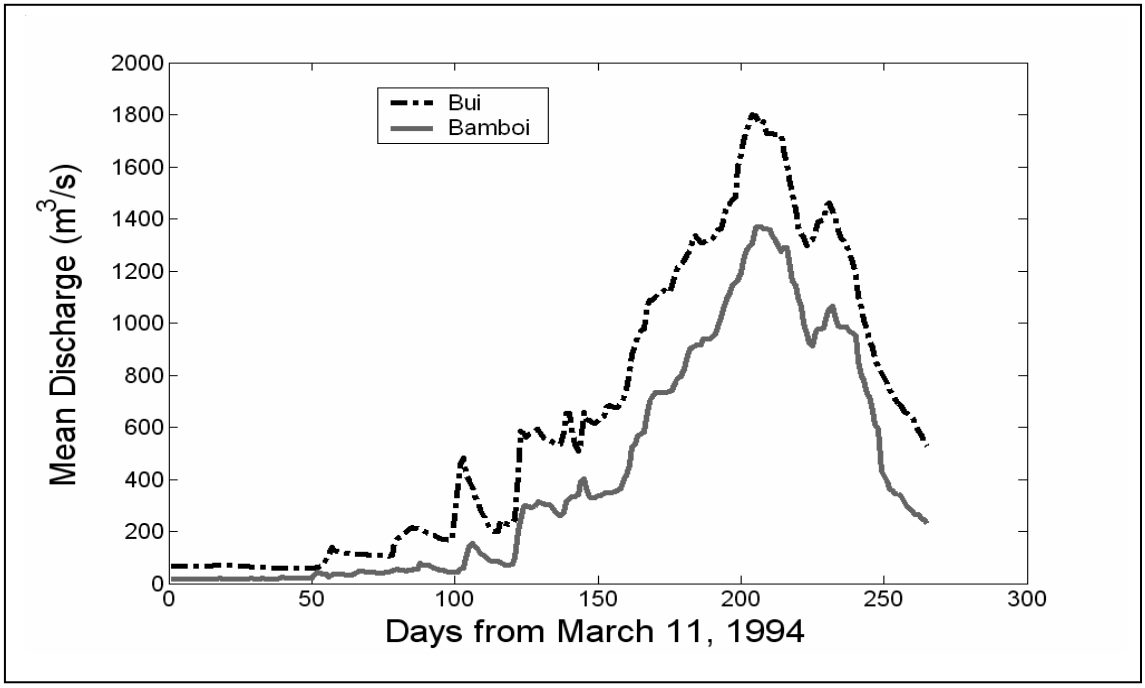


Figure 4.3a Daily hydrographs for Bui and Bamboi on the Black Volta River (1994)

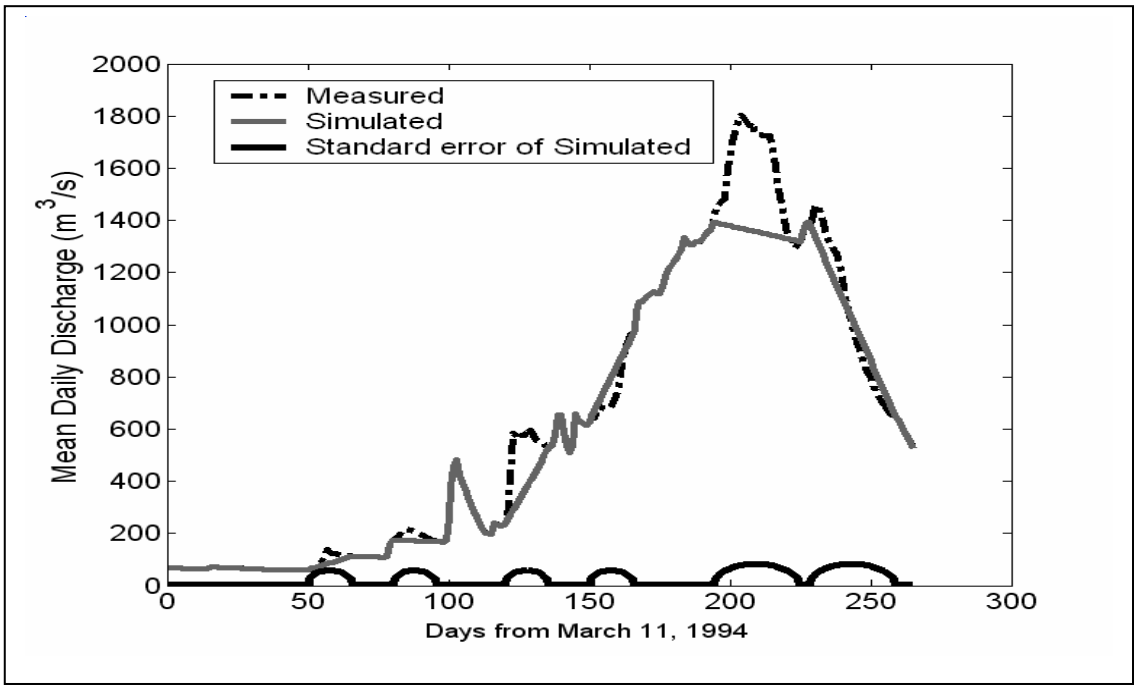


Figure 4.3b Measured and simulated daily hydrographs for Bui on the Black Volta River; state space model on Bui discharge only (1994)

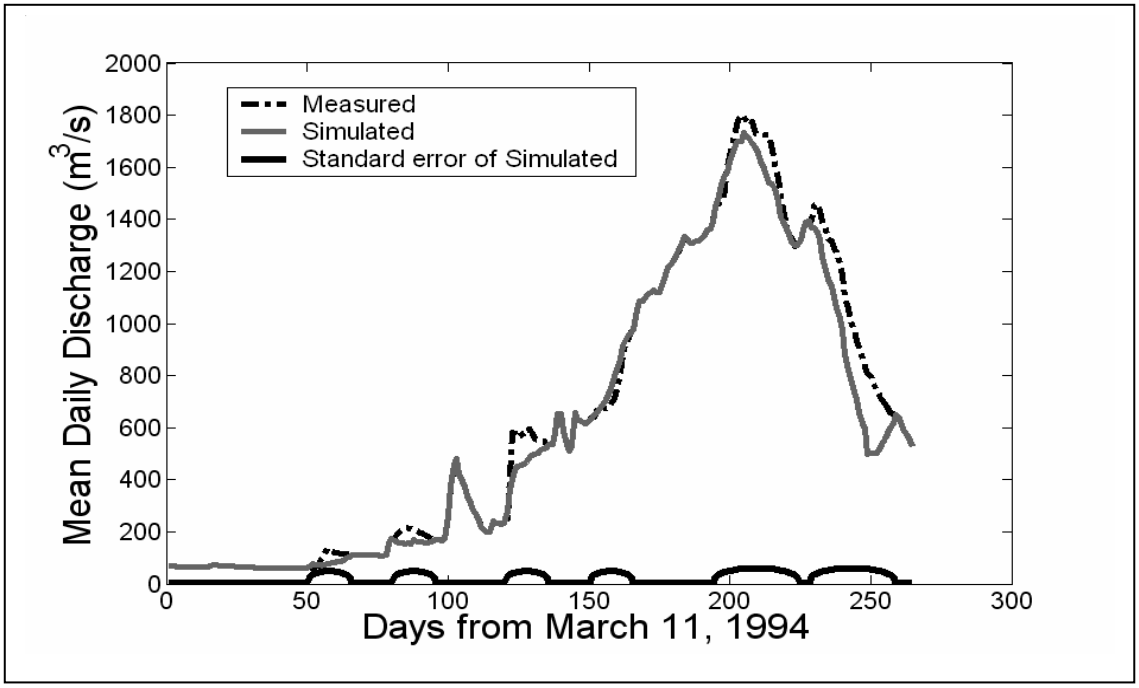


Figure 4.3c Measured and simulated daily hydrographs for Bui on the Black Volta River; state space model on Bui and Bamboi discharges (1994)

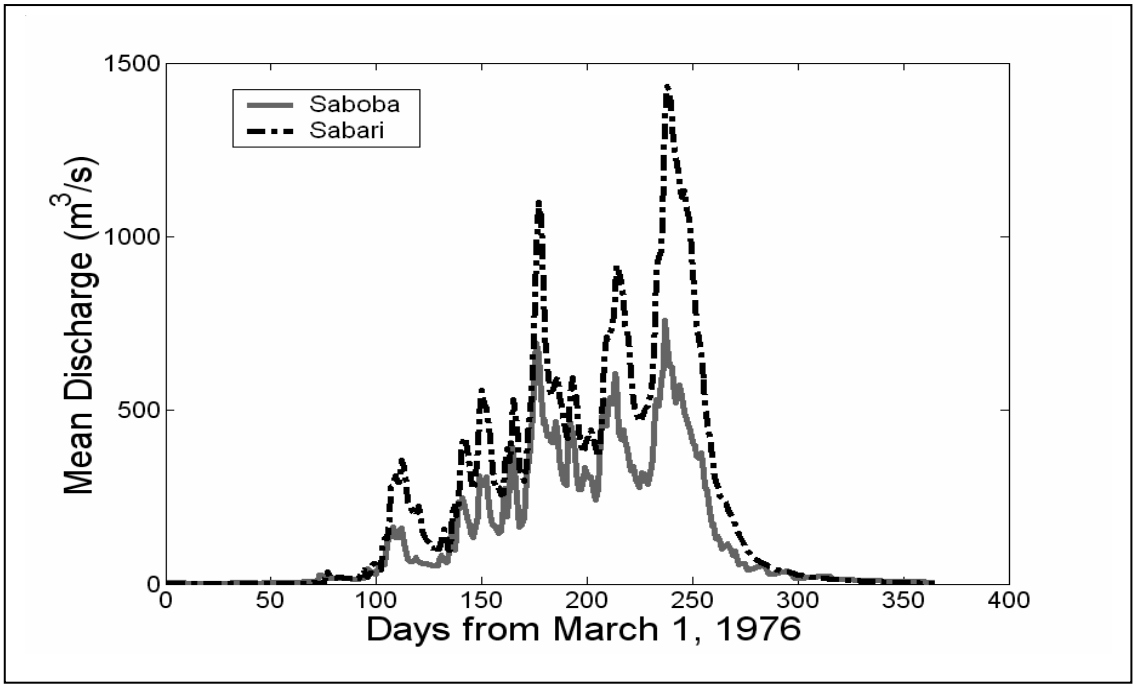


Figure 4.4a Daily hydrographs for Saboba and Sabari on the Oti River (1976)

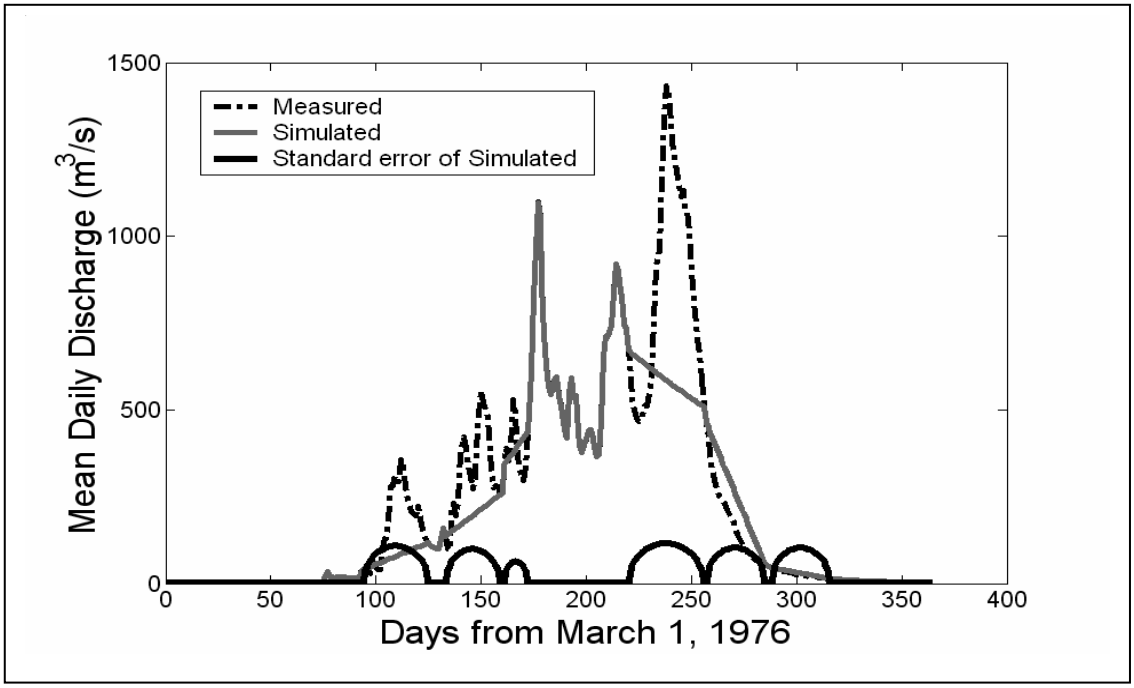


Figure 4.4b Measured and simulated daily hydrographs for Sabari on the Oti River; state space model on Sabari discharge only (1976)

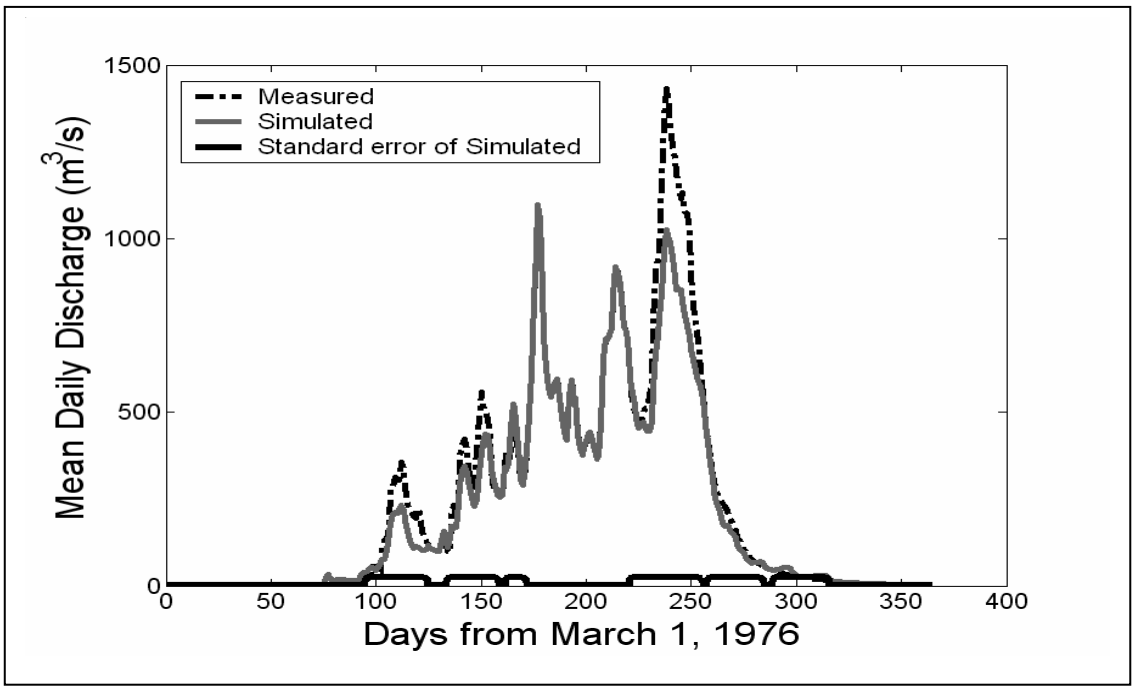


Figure4.4c Measured and simulated daily hydrographs for Sabari on the Oti River; state space model on Sabari and Saboba discharges (1976)

Table 4.1 NSEs for the various tests

TEST	SERIES	**NSE (%)**
1964	Flows in the Black Volta	
	Target station Bui only	95.0
	Bui, Lawra and Bamboi	99.0
	Bui and Bamboi only	98.6
1994	Flows in the Black Volta	
	Target station Bui only	93.5
	Bui and Bamboi	97.6
1976	Flows in the Oti	
	Target station Sabari only	61.6
	Sabari and Saboba	91.4

Predicted and observed flows at the station Sabari on the Oti River for 1976 are shown in the plots in figures 4.4b and 4.c for cases (i) and (iii) only, with the respective standard errors of the predictions plots. These and Table 4.1 show that though the spatial correlation between the flows at Saboba and those at Sabari may not be good enough to provide excellent predictions at Sabari using Saboba data, the predictions with the Saboba flows are much better than those from the use of the Sabari observations alone.

4.7 Conclusions

A spatio-temporal state space linear dynamic model was developed to fill short gaps (up to one month) in daily runoff series using other, spatially correlated, daily series of up to a year long. Parameter estimation was by the EM algorithm. Application of the model in the Volta Basin of West Africa shows that it is capable of providing good estimates of short gaps in river flows. The strength of the model is its ability to provide estimates of both the parameter and the missing values concurrently and without the need for separate calibration or training series and also provide error estimates of predictions. In addition, the errors of the predictions are smaller when using other spatially correlated series with the series with missing values than when this series is used alone to fill its gaps. It is, therefore, very suitable for filling in short gaps in riverflow series for basins such as the Volta, where missing flows in runoff series at many gauging stations abound, and for evaluating the accuracy of the model predictions.

The critical assumption in the spatio-temporal model is that the catchment runoffs for all stations considered in a model run are generated by the same process. The model thus works well when there is good spatial correlation between the runoff series involved. The results obtained in this study show that very clearly. It is, therefore, important that stations be grouped according to the degree of spatial correlation in their runoff series in order to obtain good predictions of missing values using the formulation presented here.

The presented method to fill data gaps is a relatively simple application of the powerful combination of the EM algorithm and Kalman smoother. While this combination allows the estimation of both states and parameters concurrently, using spatial and temporal data, use of non-linear updates and Gaussian error structures of the underlying processes would add versatility to the framework.

5 MODELLING STREAMFLOWS USING NARMAX POLYNOMIAL MODELS

5.1 Introduction

In Chapter 3, it was shown that the rainfall-runoff relationship in the Volta Basin, as in almost all basins, is very non-linear. It is, therefore, important that in modelling this relationship, this nonlinearity is adequately accounted for if streamflow predictions are to be reliable.

A non-linear input-output model suitable for representing a wide variety of non-linear systems including environmental systems is the NARMAX (Non-linear Autoregressive Moving Average with eXogenous Input) polynomial model (Billings and Leontaritis, 1982; Leontaritis and Billings, 1985; Chen and Billings, 1989; Tabrizi *et al*., 1998). In this chapter, the suitability of this model for river runoff prediction in the Volta Basin is investigated. The model is applied to monthly rainfall-runoff series at selected gauging stations in the basin and the results used to ascertain its ability to adequately account for the nonlinearities in the runoff generation process.

5.2 The NARMAX polynomial model

The NARMAX model is a metric or black-box model that been has been found to be very useful in modelling the nonlinear dynamics of both natural and devised systems. Its formulation includes a very elegant method of selecting parameters that are most significant in representing the nonlinearities in the system, thereby ensuring parameter parsimony in the estimated model. It is usually expressed as a non-linear polynomial function expansion of lagged input, output and noise terms, and, for the Single-Input-Single-Output (SISO) model, is given as:

$$y(t) = f^d\begin{bmatrix} y(t-1), y(t-2), \ldots, y(t-n_y), u(t-n_k), \ldots, u(t-n_k-n_u) \\ e(t-1), \ldots, e(t-n_e) \end{bmatrix} + e(t) \quad (5.1)$$

where

f^d	=	polynomial of degree d (d>1)
y(t), u(t), e(t)	=	output, input and white noise signals, respectively, at time t
n_y, n_u, n_e	=	maximum output, input and noise lags, respectively
$n_{k\,(>0)}$	=	input signal time delay (measured in sampling intervals)

In non-linear systems identification in general, n_k is usually taken as at least 1. However, since in the analysis here u(t) is monthly rainfall, n_k=0, i.e., y(t) is a function of both current and previous inputs, would be considered. Thus for $n_y = n_u = n_e = 1$, d = 2 and $n_k = 0$, for example, the polynomial expansion in Equation 5.1 for y(t) is:

$$y(t) = \begin{pmatrix} \theta_1 + \theta_2 y(t-1) + \theta_3 u(t) + \theta_4 u(t-1) + \\ \theta_5 y(t-1)^2 + \theta_6 u(t)^2 + \theta_7 u(t) u(t-1) + \theta_8 u(t-1)^2 + \\ \theta_9 y(t-1) u(t) + \theta_{10} y(t-1) u(t-1) + \\ \theta_{11} y(t-1) e(t-1) + \theta_{12} u(t) e(t-1) + \theta_{13} u(t-1) e(t-1) + \\ \theta_{14} e(t-1) + \theta_{15} e(t-1)^2 \end{pmatrix} + e(t) \quad (5.2)$$

where θ_1, θ_2, .., θ_{15} are parameters of the model.

A non-linear polynomial model of the form of Equation 5.1 but without the noise terms is the NARX (Non-linear Autoregressive with eXogenous Input) model given as:

$$y(t) = f^d\left[y(t-1), y(t-2), \ldots, y(t-n_y), u(t-n_k), \ldots, u(t-n_k-n_u)\right] + e(t) \quad (5.3)$$

This model is also general and can describe any non-linear system well (Stenman, 2000). In addition, it is not recursive, as the regressors are independent of previous model outputs, whereas in the NARMAX representation the noise terms can only be derived from previous model outputs. Therefore, the NARX model is more convenient to work with. However, the absence of a noise model in the structure means that a large

number of regressor terms has to be included so that the model can adequately represent both the system and noise dynamics (Stenman, 2000). Due to this limitation and also to avoid bias in the estimated parameters (Chiras, 2000), the full NARMAX model is preferred.

5.2.1 Formulation of the model

For a given input-output series and any set of n_y, n_u, n_e , n_k and d, the polynomial represented in Equation 5.1 above can generally be expressed as:

$$y(t) = \sum_{m=1}^{np} P_m \theta_m + e(t) \tag{5.4a}$$

where np (p not a subscript) is the number of parameters of the model and so equals the number of terms in the polynomial expansion, P_m is the m^{th} regressor term with $P_1 = 1$, and θ_m is the regression parameter for term m. The regressor terms are formed, as in Equation 5.2, by various combinations of lagged values of the output and noise terms and both lagged and current (when nk = 0) values of the input term.

In matrix form Equation 5.4a becomes:

$$y = P\theta + \varepsilon \tag{5.4b}$$

Here P is n by np and y n by 1 regressor and output matrices, respectively, Θ is np by 1, and ε is n by 1 parameter and white noise vectors, respectively, with n being the number of input-output samples.

Equation 5.4 is linear in the parameters, and so its parameters can be estimated by the use of well established parameter estimation techniques developed for linear systems identification, such as orthogonal least squares methods, although recursion is required to estimate the noise terms.

The system identification problem in modelling the output y using the NARMAX polynomial representation then consists of:

(i) Determination of the input, output and noise lags and input time delay (i.e., model order)

(ii) Selection of the polynomial degree

(iii) Estimation of the parameters of the model represented as in Equation 5.4

Procedures in (i) and (ii) involve model structure selection, while procedure (iii) is the straightforward parameter estimation.

5.2.2 Error reduction ratio and selection of significant terms

A major difficulty in systems identification using NARMAX models is obtaining a model that is parsimonious in the number of parameters and represents the dynamics of the system adequately. This is because of the enormous number of parameters that is often involved in such models. For example, a model with ny = nu = ne = d = 3 and nk = 0 results in 286 parameters. To assist in the selection of the most significant terms to be considered in a NARMAX model, the Error Reduction Ratio (ERR) algorithm (Billings *et al*., 1989, Korenberg *et al*., 1988), which is derived from the orthogonal least square algorithm used for solving equations such as Equation 5.4, is often used. The ERR is used to order regressors according to the levels to which each regressor reduces the mean square model error (MSE). The regressor with the largest reduction in the MSE is ranked first and is the first to be considered in a forward regression solution to Equation 5.4. The strength of the procedure lies in the fact that it does not require the estimation of the full model in order to rank the regressors. The number of terms to be included in the final model is determined by the application of information criteria (Billings *et. al*., 1989, Chiras *et al*., 2000) such as *final prediction error*, FPE (Akaike, 1974a), *Akaike information criterion*, AIC (Akaike, 1974b) and *Bayesian information criterion*, BIC (Kashyap, 1977) on validation series. The method is elaborated in the methodology sub-section below.

5.3 Application of the model

Monthly rainfall (u)-runoff (y) series for selected river catchments in the Volta Basin were used to fit and evaluate SISO NARMAX polynomial models of various structures.

QR function available in Matlab (Mathworks, 2004). The parameters of the model at each forward regression stage were obtained with the Matlab REGRESS function and used to compute the predicted values of the validation series. The predicted validation series at each stage of the forward regression were used with evaluation criteria to determine when to stop regressing.

A QR orthogonal-triangular decomposition can be performed on the n by np matrix P in (5.4b) to obtain

$$P = QR \tag{5.5}$$

where Q is an n by np orthogonal matrix such that $Q^TQ = I$, the np by np identity matrix, and R an np by np upper triangular matrix.
Let:

$$g = Q^T y \tag{5.6}$$

be an np by 1 vector, so that for the np by 1 parameter vector , Θ,

$$R\Theta = g \tag{5.7}$$

Then

$$y = P\Theta + \varepsilon = \left(PR^{-1}\right)\left(R\Theta\right) + \varepsilon = Qg + \varepsilon \tag{5.8}$$

Thus, the sum of squares of the observed output samples is

$$y^T y = \sum_{i=1}^{np} g_i^2 \left(q_i^T q_i\right) + \varepsilon^T \varepsilon = \sum_{1}^{np} g_i^2 + \varepsilon^T \varepsilon \tag{5.9}$$

where g_i is element i of vector g, q_i column i of matrix Q, and the orthogonality of Q and the mean of $\varepsilon = 0$ hold.

The error reduction ratio ERR_i due to the i^{th} regressor term is defined as the proportion of the observed output variance explained by that term (Chen *et. al*., 1989), i.e.:

$$ERR_i = \frac{g_i^2}{y^T y} \tag{5.10}$$

Equation 5.10 was used to order the regressors in each model according to their ERR_i values and the forward regression procedure used to select significant regressors and estimate parameter values. The process (NARX) model was fitted first, equation 5.10 used again to order the noise terms from the residuals generated by the fitted process model and the forward regression procedure used again to select the significant noise terms. The general steps adopted in the modelling procedure for each of the river catchments considered were as follows:

1. Select a model structure (i.e., pick n_a, n_b, and n_c; $n_k = 0$ and $d = 3$ for all structures in this study).
2. Form the regressors for the process or NARX model and rank them using their ERR values.
3. Perform forward regression and select process model (with optimum number of parameters) with validation series using both the AIC and NSE criteria
4. Form noise terms using the residuals from the model selected in 2 above and rank these using their ERR values.
5. Perform forward regression adding terms from the noise terms to the process model terms. Select full model (with optimum number of parameters) with validation series using both the AIC and NSE criteria.
6. Repeat for other model structures and select the best model(s) for the catchment from the values of the AIC and NSE of the validation series.

The following forms of the AIC (Ljung, 1999) and NSE criteria were used:

$$AIC = Log(V) + \frac{2np_s}{n} \tag{5.11a}$$

$$NSE = 100\left(1 - \frac{var(\varepsilon)}{var(y)}\right) \tag{5.11b}$$

where $V = \varepsilon^T\varepsilon$, np_s = number of selected regressors, and n is the sample size of the monthly runoff series (estimation or validation) being considered.

5.4 Results and discussion

Tables 5.1 and 5.2 summarize the results of the model selection process for Bamboi on the Black Volta River and Mango on the Oti River for various model structures considered in the study. Similar model selection procedures were undertaken for the rest of the selected stations. Tables 5.1 and 5.2 indicate that of the 23 model structures considered for each station only 8 of them in each case included significant terms from the noise model from the selection process. This suggests that for the remaining models enough process terms were fitted to account for both the system dynamics and the noise input. Considering both the NSE and AIC criteria for the validation series (columns 6, 10, 12 and 16), Table 5.1 shows that only 3 of the model structures for Bamboi are important. These are models 1 ([n_a n_b n_c] =[1 1 1]), 5 ([n_a n_b n_c] =[1 3 3]) and 7 ([n_a n_b n_c] = [1 4 2]). However, considering parameter parsimony, model 1 is the best with its 13 parameters (column 13). Models 5 and 7 with 22 and 23 parameters, respectively, are likely to be overparameterized.

A similar analysis of Table 5.2 shows that 5 good models can be identified, i.e., models 3 ([n_a n_b n_c] =[1 3 1]), 9 ([n_a n_b n_c] =[1 5 1]), 10 ([n_a n_b n_c] =[1 5 2]), 11 ([n_a n_b n_c] =[1 5 3]) and 23 ([n_a n_b n_c] =[5 1 1]). Here, models 9 and 11 are superior considering their lower AIC values (column 16). However, model 9, with 3 parameters less than and with an F value for the estimation series (the fitted series) higher than that of 11 (indicating the fit to 9 is more significant than for 11), should be considered the better of the two.

Table 5.1 Model structure selection layout for Bamboi on the Black Volta River

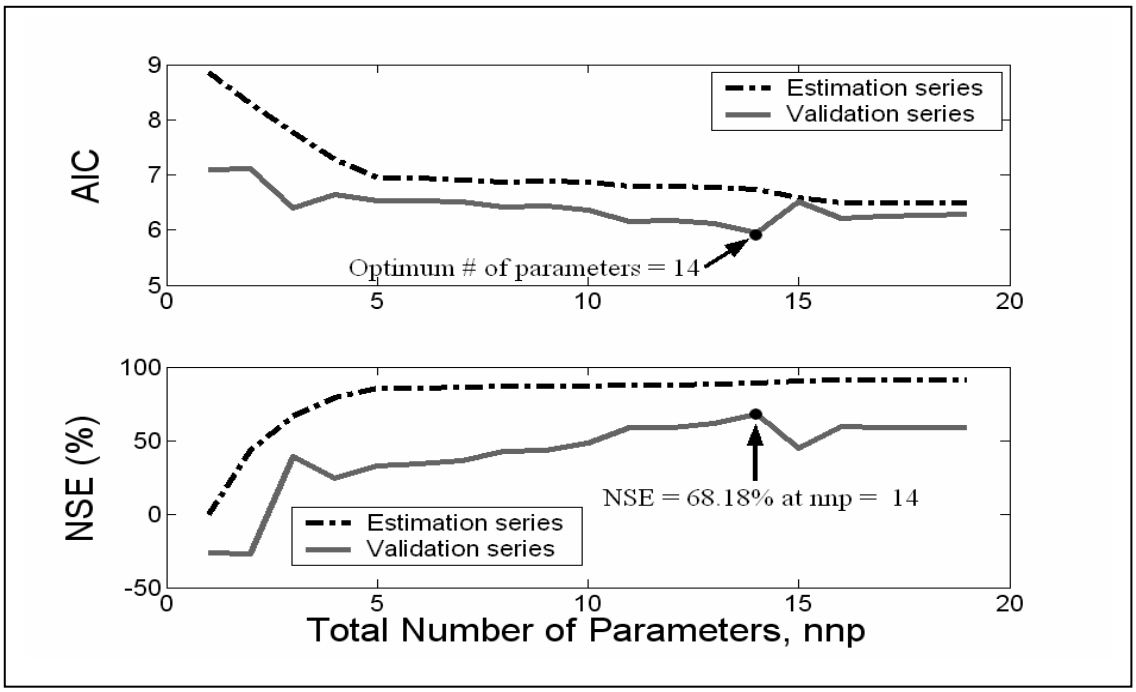
				PROCESS MODEL						FULL NARMAX MODEL					
Model No.	na	nb	nc	NSE (%)		Opt. np	R-square	F	AIC	NSE (%)		Opt. np	R-square	F	AIC
				Est. series	Val. series		Estimation	series	(Validation)	Est. series	Val. series		Estimation	series	(Validation)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
1	1	1	1	94.34	88.96	12	0.94	235.63	6.54	94.44	89.93	13	0.94	218.67	6.46
2	1	2	1	95.14	89.26	21	0.95	143.04	6.64	-	-	-	-	-	-
3	1	3	1	94.33	88.01	16	0.94	167.35	6.67	95.45	88.89	18	0.96	183.46	6.63
4	1	3	2	"	"	"	"	"	"	95.45	88.89	18	0.96	183.46	6.63
5	1	3	3							95.70	91.05	22	0.96	153.49	6.47
6	1	4	1	94.42	89.42	19	0.95	138.47	6.59	95.08	90.41	20	0.95	148.98	6.51
7	1	4	2							95.35	91.31	23	0.95	133.60	6.45
8	1	4	3	"	"	"	"	"	"	95.08	90.41	20	0.95	148.98	6.51
9	1	5	1	93.38	86.46	18	0.94	123.12	6.81	-	-	-	-	-	-
10	1	5	2	"	"	"	"	"	"	-	-	-	-	-	-
11	1	5	3	"	"	"	"	"	"	-	-	-	-	-	-
12	1	6	1	92.63	88.11	15	0.93	136.00	6.63	-	-	-	-	-	-
13	2	1	1	95.05	89.03	17	0.95	180.09	6.60	-	-	-	-	-	-
14	2	1	2	"	"	"	"	"	"	-	-	-			
15	2	2	1	94.88	87.89	17	0.95	173.70	6.70	94.96	88.81	18	0.95	165.14	6.64
16	2	2	2	"	"	"	"	"	"	-	-	-	-	-	-
17	2	3	1	93.62	87.27	17	0.94	137.34	6.75	-	-	-	-	-	-
18	2	4	1	93.73	88.57	19	0.94	122.50	6.67	-	-	-	-	-	-
19	3	1	1	94.90	88.33	17	0.95	174.19	6.66	-	-	-	-	-	-
20	3	2	1	94.70	87.67	18	0.95	156.26	6.73	-	-	-	-	-	-
21	3	3	1	93.66	86.70	17	0.94	138.28	6.79	-	-	-	-	-	-
22	4	1	1	94.70	88.37	17	0.95	167.02	6.65	-	-	-	-	-	-
23	5	1	1	93.60	86.31	15	0.94	158.20	6.78	-	-	-	-	-	-

Table 5.2 Model structure selection layout for Mango on the Oti River

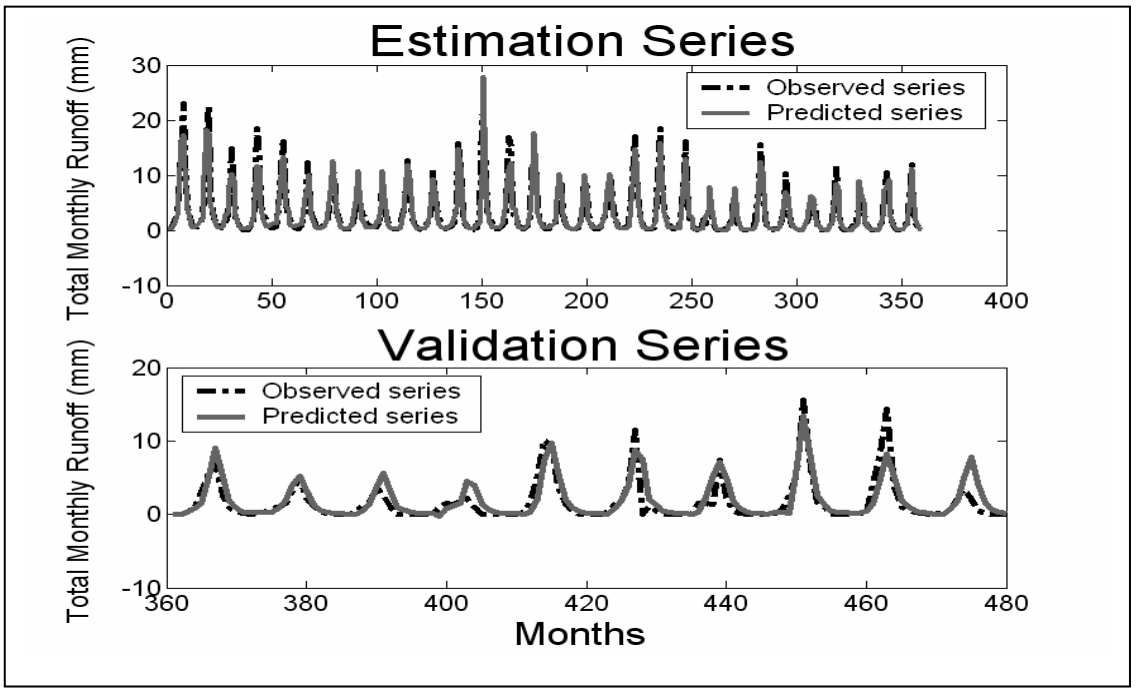
				NARX						NARMAX					
Model No.	na	nb	nc	NSE (%)		Opt. np	R-square	F	AIC	NSE (%)		Opt. np	R-square	F	AIC
				Est. series	Val. series		Estimation	series	(Validation)	Est. series	Val. series		Estimation	series	(Validation)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
1	1	1	1	94.12	87.18	9	0.94	341.63	8.11	-	-	-	-	-	-
2	1	2	1	94.10	86.68	12	0.94	242.74	8.22	94.12	87.66	13	0.94	221.98	8.18
3	1	3	1	93.48	89.34	14	0.94	182.04	8.05	93.50	90.13	15	0.94	168.74	8.00
4	1	3	2	"	"	"	"	"	"	-	-	-	-	-	-
5	1	3	3	"	"	"	"	"	"	-	-	-	-	-	-
6	1	4	1	93.26	88.69	12	0.93	210.20	8.04	-	-	-	-	-	-
7	1	4	2	"	"	"	"	"	"	-	-	-	-	-	-
8	1	4	3	"	"	"	"	"	"	-	-	-	-	-	-
9	1	5	1	93.60	88.86	12	0.94	222.43	8.02	93.70	91.22	13	0.94	206.06	7.81
10	1	5	2	"	"	"	"	"	"	94.34	90.02	14	0.94	211.70	7.97
11	1	5	3	"	"	"	"	"	"	94.46	92.28	16	0.95	185.42	7.77
12	1	6	1	93.66	89.62	12	0.94	224.08	7.94	-	-	-	-	-	-
13	2	1	1	91.26	85.30	10	0.91	196.56	8.27	91.51	86.34	11	0.92	181.62	8.22
14	2	1	2	"	"	"	"	"	"	91.51	86.34	11	1.92	181.62	8.22
15	2	2	1	93.41	84.96	13	0.93	196.52	8.37	-	-	-	-	-	-
16	2	2	2	"	"	"	"	"	"	-	-	-	-	-	-
17	2	3	1	93.51	87.67	15	0.94	168.81	8.22	-	-	-	-	-	-
18	2	4	1	93.66	89.18	17	0.94	149.56	8.14						
19	3	1	1	94.24	86.48	11	0.94	274.95	8.20	-	-	-	-	-	-
20	3	2	1	95.01	86.35	19	0.95	169.18	8.44	95.13	87.09	20	0.95	163.37	
21	3	3	1	90.98	84.29	9	0.91	214.70	8.29	-	-	-	-	-	-
22	4	1	1	94.97	87.90	14	0.95	239.86	8.17	-	-	-	-	-	-
23	5	1	1	95.35	90.48	17	0.95	207.68	8.01	-	-	-	-	-	-

Figures 5.1a, 5.2a, 5.3a and 5.4a show plots of AIC and NSE values against number of parameters included in the best NARMAX model for the representative stations Dapola and Bamboi on the Black Volta River, Yagaba on the Kulpawn River (White Volta Basin) and Mango on the Oti River. The best models for Dapola and Bamboi, for example, are $[n_a\ n_b\ n_c] = [3\ 1\ 0]$ and $[n_a\ n_b\ n_c] = [1\ 1\ 1]$, respectively. Such plots were used to select the best (optimum) model in each model class (of the same structure) such as listed in tables 5.1 and 5.2. The best in a given class was the one with minimum validation series AIC. The NSE values shown in the figures are the values for the validation series at the optimum number of parameters. The plots show that there is a continuous fall in the AIC and rise in NSE with increasing numbers of parameters for the estimation series. Since this series is the fitting series and is known to the model, an increase in the number of parameters always leads to a better fit to the series. Not so for the validation series. The model does not know the output component of this series and so cannot adjust to fit it. Thus, insignificant parameters soon show as an increase in the AIC. Therefore, the application of the AIC on validation data can be very effective in determining when to stop adding parameters to a model. However, the use of this criterion alone would not always result in a parsimonious model, as clearly shown in tables 5.1 and 5.2 where the number of parameters for some of the models is rather very high. Other goodness-of-fit criteria need also to be considered to arrive at an optimum model.

Visual evaluation of the selected model in each class was made through plots such as those in figures 5.1b, 5.2b, 5.3b and 5.4b for the best models for Dapola, Bamboi, Yagaba and Mango, respectively. These figures show plots of observed and predicted total monthly runoff for the NARMAX formulations. The plots and the validation NSEs shown in figures 5.1a – 5.4a indicate fairly good fits and predictions for the selected models. The number of fitted parameters, the estimation and validation NSEs for the best models for the stations modeled here and the corresponding values from Table 3.6 for the best linear ARX models in Chapter 3 are shown in Table 5.3. It can be seen from the validation NSEs in particular, that the nonlinear modelling has produced better predictions with much fewer parameters than the linear modelling. For Lawra, Dapola, Bamboi and Mango, in particular, better predictions have been obtained for the nonlinear model with less than half the number of parameters.

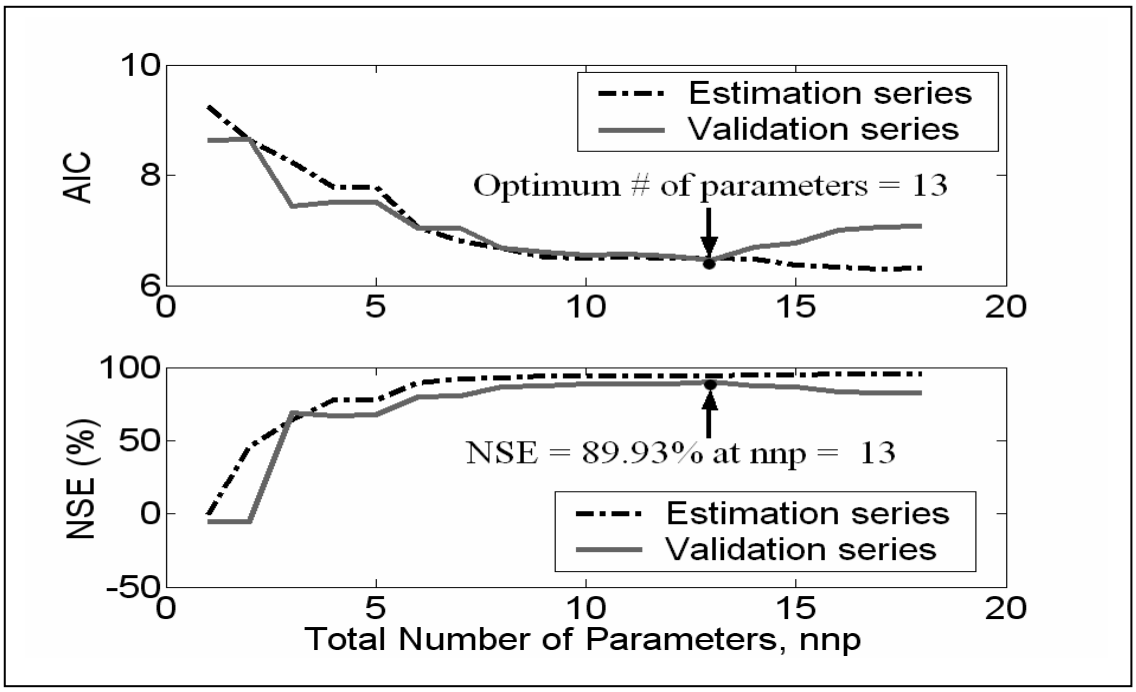


(a)

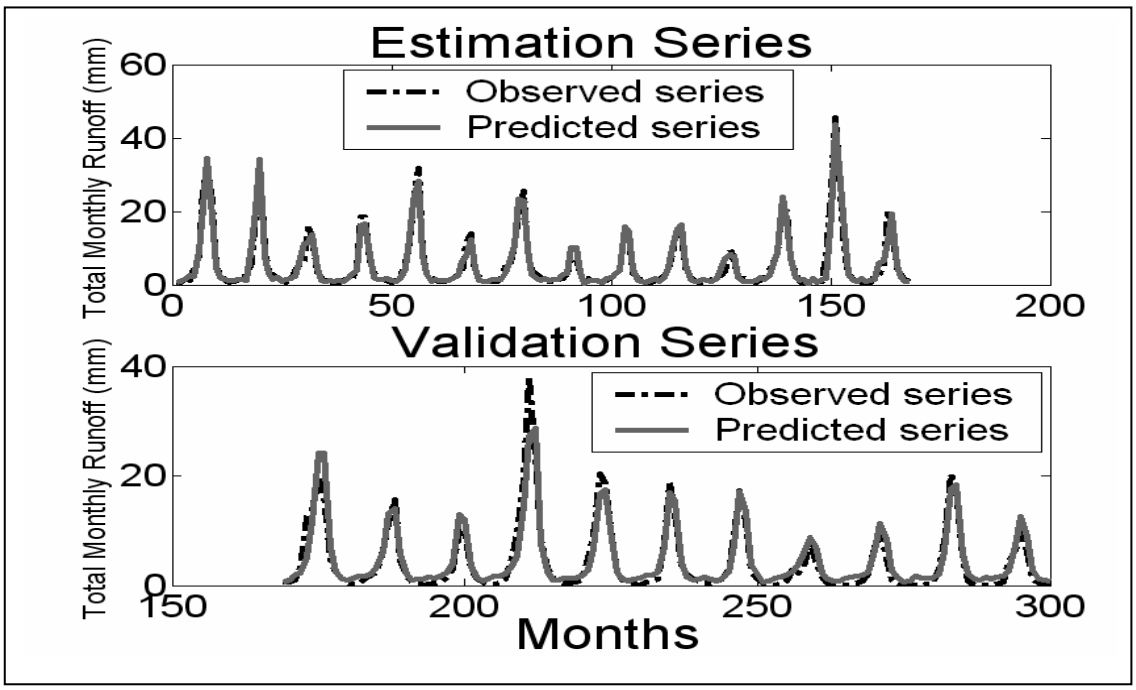


(b)

Figure 5.1 AIC, NSE, observed and predicted runoff plots for Dapola on the Black Volta River – [na nb nc] = [3 1 0]
(a) AIC and NSE vs. number of NARMAX parameters (nnp)
(b) Observed and predicted monthly runoff for both estimation and validation series

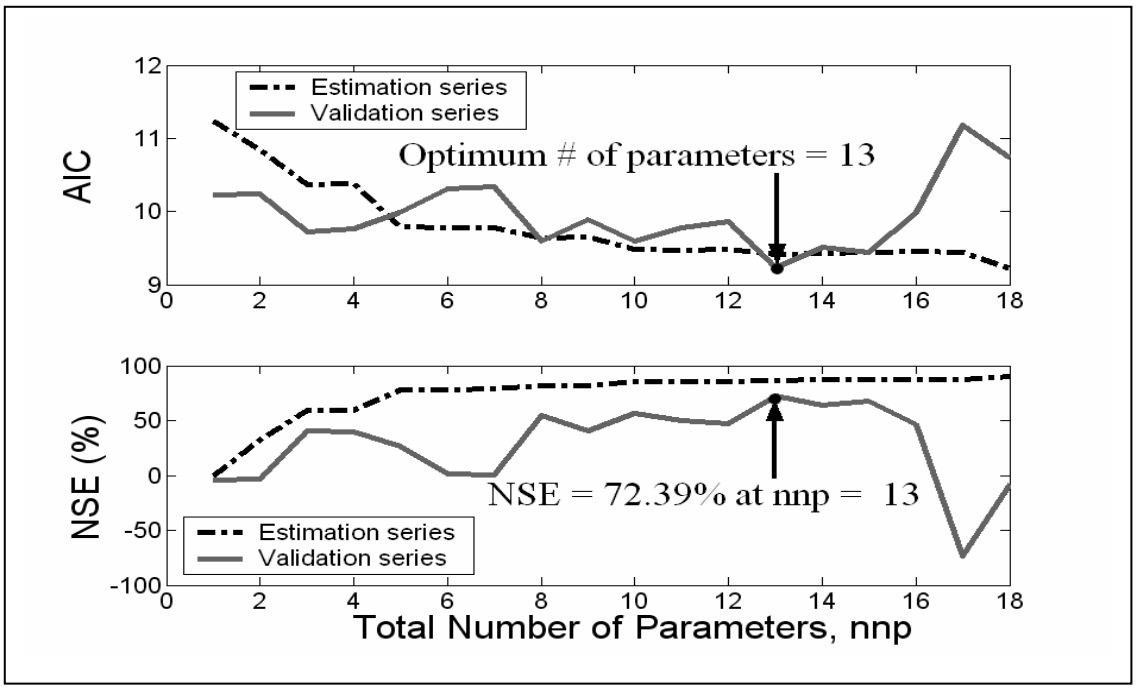


(a)

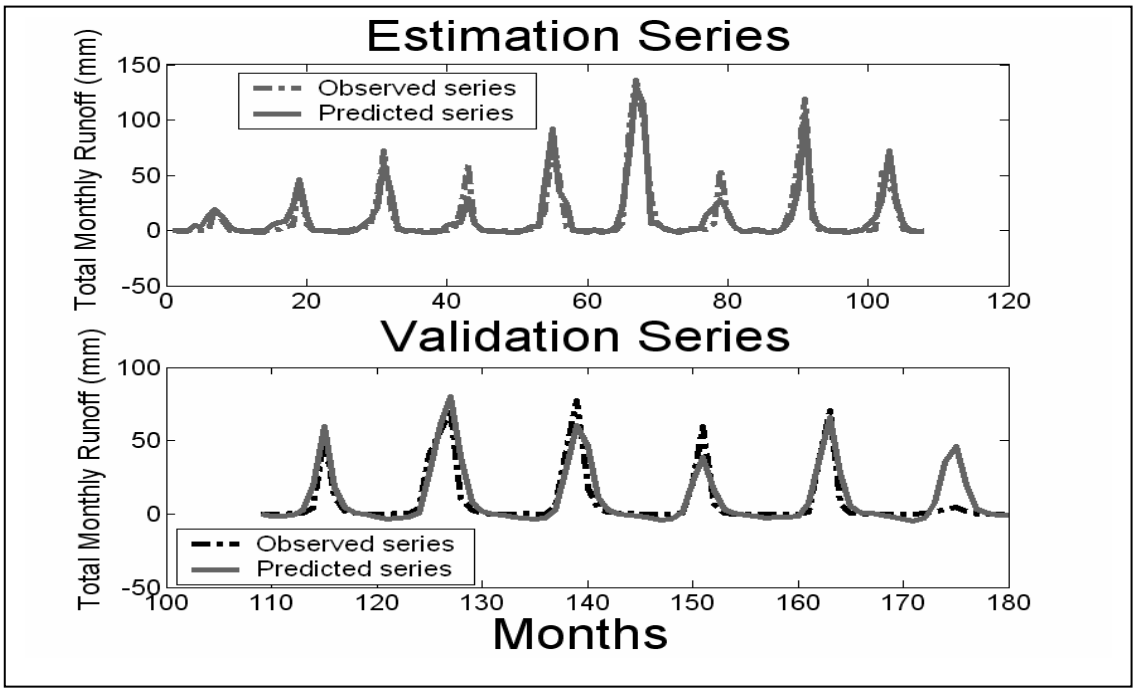


(b)

Figure 5.2 AIC, NSE, observed and predicted runoff plots for Bamboi on the Black Volta River – [na nb nc] = [1 1 1]
(a) AIC and NSE vs. number of NARMAX parameters (nnp)
(b) Observed and predicted monthly runoff for both estimation and validation series

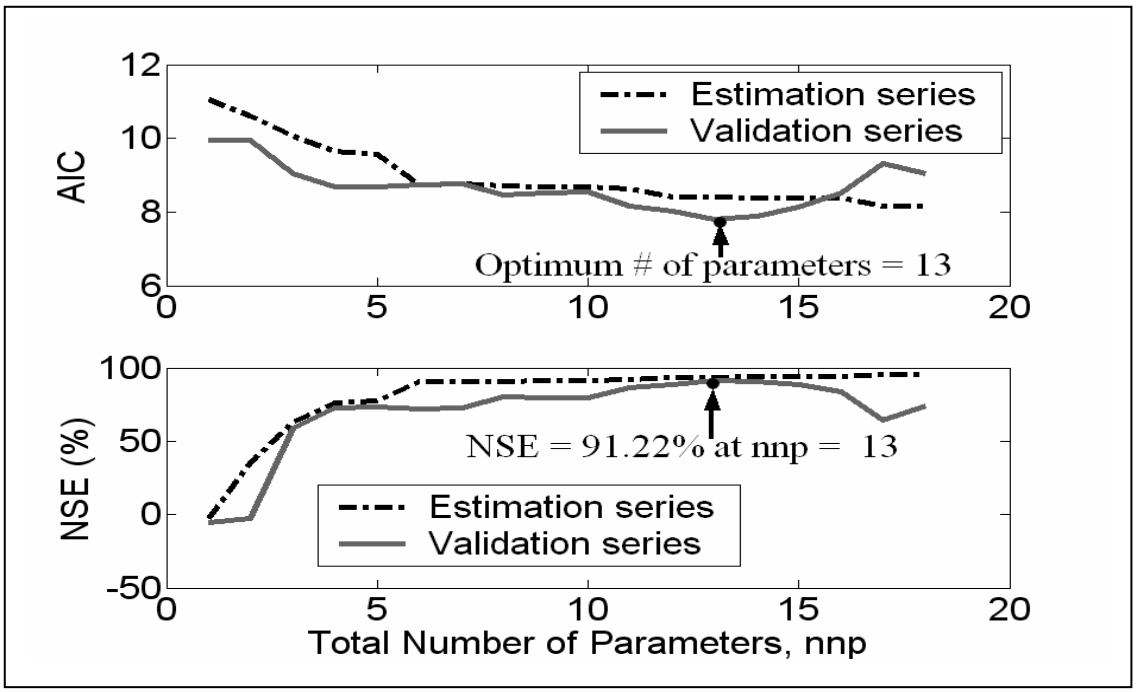


(a)

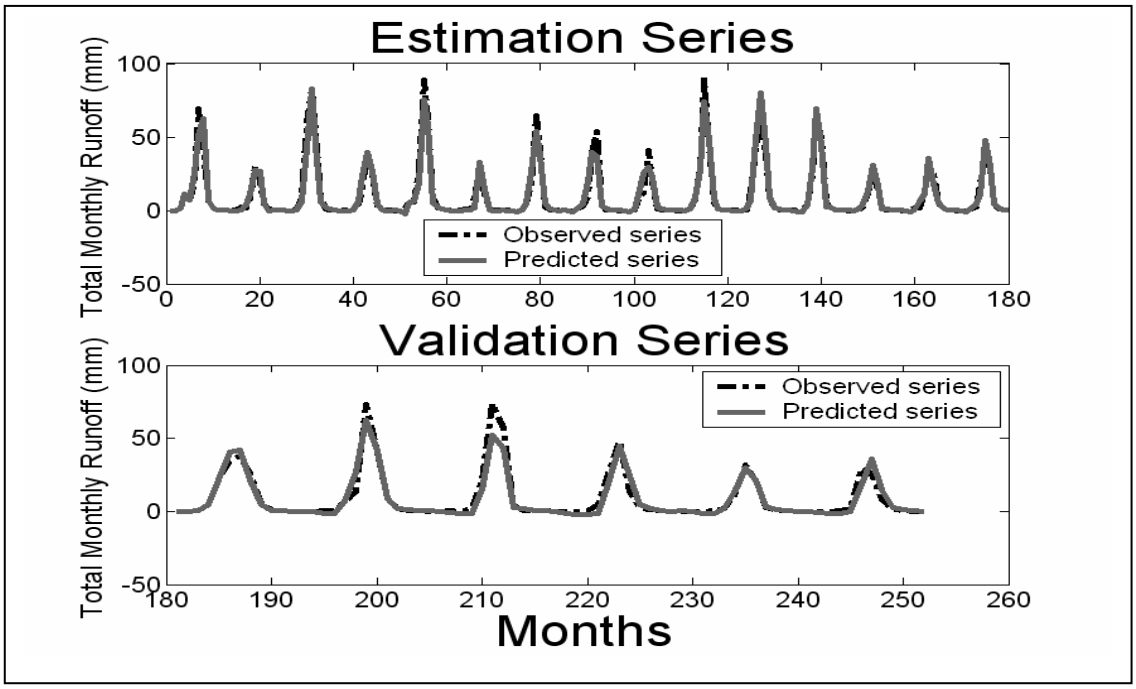


(b)

Figure 5.3 AIC, NSE, observed and predicted runoff plots for Yagaba on the Kulpawn River (White Volta Basin) – [na nb nc] = [1 1 1]
(a) AIC and NSE vs. number of NARMAX parameters (nnp)
(b) Observed and predicted monthly runoff for both estimation and validation series



(a)



(b)

Figure 5.4 AIC, NSE, observed and predicted runoff plots for Mango on the Oti River – [na nb nc] = [1 5 1]
(a) AIC and NSE vs. number of NARMAX parameters (nnp)
(b) Observed and predicted monthly runoff for both estimation and validation series

Table 5.3 Comparison of performance of the linear ARX and the NARMAX models (BV = Black Volta, KM = Koumangou)

Station	**River**	**Number of parameters in best model**		**NSE (%)**			
		ARX	**NARMAX**	**Estimation series**		**Validation series**	
				ARX	**NARMAX**	**ARX**	**NARMAX**
Black Volta Basin							
Lawra	BV	27	9	89.57	87.50	78.44	88.43
Dapola	BV	29	14	86.41	88.82	66.01	68.18
Bui	BV	16	10	85.43	95.22	85.54	91.27
Bamboi	BV	29	13	87.29	94.44	85.00	89.93
White Volta Basin							
Wiasi	Sisilli	15	11	82.31	87.89	71.19	84.88
Yagaba	Kulpawn	17	13	64.31	86.90	56.71	72.39
Nasia	Nasia	15	10	86.92	94.30	62.30	69.62
Nabogo	Nabogo	20	10	75.41	80.08	58.56	69.97
Oti Basin							
Porga	Oti	19	11	78.87	93.78	61.49	82.35
Mango	Oti	28	13	89.14	93.70	83.38	91.22
Koumangou	KM	18	13	84.28	88.37	79.02	87.10
Sabari	Oti	28	14	90.41	93.04	88.02	92.14

Clearly, nonlinear modelling of the rainfall-runoff process, as illustrated here, is very appropriate. The NARMAX models have been able to account for much of the nonlinearity in the rainfall-runoff transformation in this study, considering the very good predictions with relatively fewer parameters.

In Table 5.4, the selected regressors and the full regressor set for both the NARX and NARMAX formulations are presented for the best model for Bamboi. The rankings of the regressors as indicated in the table are obtained from the application of the ERR algorithm. The NSE values for the validation series as each regressor term is included in the model up to one term after the optimum number of terms are also indicated in the table. It is important to note that while the full regressor terms number 35, only 13 were found to be significant for the full NARMAX case for this model structure.

Table 5.4 Regressors for model [na nb nc) = [1 1 1] for Bamboi on the Black Volta (The NSE value refers to the value for the model with number of regressors up to the current regressor)

Optimum Regressor Terms				**Full Regressor terms**		
Regressor #	**Regressor Rank**	**Regressor (P)**	**Validation series NSE(%)***	**Regressor #**	**Regressor Rank**	**Regressor (P)**
NARX Model				**Process model terms**		
1	**1**	**1**	-5.35	1	**1**	**1**
2	**2**	y_{t-1}	-5.36	2	**2**	y_{t-1}
3	**3**	u_t	68.71	3	**3**	u_t
4	**4**	u_{t-1}	67.25	4	**4**	u_{t-1}
5	**5**	$(y_{t-1})^2$	67.56	5	**5**	$(y_{t-1})^2$
6	**6**	$(y_{t-1})* u_t$	79.85	6	**6**	$(y_{t-1})*u_t$
7	**7**	$(y_{t-1})*(u_{t-1})$	80.39	7	**7**	$(y_{t-1})*(u_{t-1})$
9	**8**	$u_t*(u_{t-1})$	86.55	9	**8**	$u_t*(u_{t-1})$
10	**9**	$(u_{t-1})^2$	87.72	10	**9**	$(u_{t-1})^2$
8	**10**	u_t^2	88.42	8	**10**	u_t^2
13	**11**	$(y_{t-1})^2*(u_{t-1})$	88.42	13	**11**	$(y_{t-1})^2*(u_{t-1})$
14	**12**	$(y_{t-1})* u_t^2$	**88.96**	14	**12**	$(y_{t-1})* u_t^2$
	13		88.82	15	**13**	$(y_{t-1})* u_t *(u_{t-1})$
NARMAX Model				16	**14**	$(y_{t-1})*(u_{t-1})$^2
1	**1**	**1**	-5.35	11	**15**	(y_{t-1})^3
2	**2**	y_{t-1}	-5.36	12	**16**	(y_{t-1})^2* u_t
3	**3**	u_t	68.71	17	**17**	ut^3
4	**4**	u_{t-1}	67.25	18	**18**	ut^2*(u_{t-1})
5	**5**	$(y_{t-1})^2$	67.56	19	**19**	$u_t*(u_{t-1})$^2
6	**6**	$(y_{t-1})* u_t$	79.85	20	**20**	(u_{t-1})^3
7	**7**	$(y_{t-1})*(u_{t-1})$	80.39			
9	**8**	$u_t*(u_{t-1})$	86.55	**Noise model terms**		
10	**9**	$(u_{t-1})^2$	87.72	29	**1**	$(e_{t-1})* u_t^2$
8	**10**	u_t^2	88.42	25	**2**	$(e_{t-1})^2$
13	**11**	$(y_{t-1})^2*(u_{t-1})$	88.42	33	**3**	$(e_{t-1})^2* u_t$
14	**12**	$(y_{t-1})* u_t^2$	88.96	34	**4**	$(e_{t-1})^2*(u_{t-1})$
29	**13**	$(e_{t-1})* u_t^2$	**89.93**	26	**5**	$(e_{t-1})*(y_{t-1})^2$
	14		87.51	30	**6**	$(e_{t-1})* u_t*(u_{t-1})$
				31	**7**	$(e_{t-1})*(u_{t-1})^2$
				32	**8**	$(e_{t-1})^2*(y_{t-1})$
				35	**9**	$(e_{t-1})^3$
				23	**10**	$(e_{t-1})*u_t$
				27	**11**	$(e_{t-1})*(y_{t-1})*u_t$
				21	**12**	e_{t-1}
				22	**13**	$(e_{t-1})*(y_{t-1})$
				24	**14**	$(e_{t-1})*(u_{t-1})$
				28	**15**	$(e_{t-1})*(y_{t-1})*(u_{t-1})$

5.5 Conclusions and recommendations

A SISO NARMAX polynomial model was successfully formulated and applied to monthly rainfall-runoff series for runoff prediction at selected river gauging stations in the Volta Basin. Several model structures for each station were evaluated using the AIC, NSE (for the validation series), R-square, and F (for the estimation series) goodness-of-fit criteria. The combination of these criteria enabled the identification of the "best" model from all the models tested for each river catchment. The AIC criterion was particularly useful in selecting the optimum models from the several available for each model structure. However, the other goodness-of-fit criteria were required to successfully identify parsimonious models in some of the cases.

Monthly rainfall predictions from the selected models were very good, and the polynomial models appeared to have captured a good part of the rainfall-runoff non-linearity, even though some peak and/or low flows were not adequately simulated in the cases investigated. The use of multiple goodness-of-fit criteria in which different criteria are employed for fitting simulations to different parts of the hydrograph, especially the peak and low flow sections, could result in improvements in the predicted hydrographs.

Nevertheless, the results show the appropriateness of nonlinear representation of the rainfall-flow process. They also indicate that the NARMAX modelling framework is suitable for monthly river runoff prediction in the Volta Basin. The drawback of the method, as applied here, is its inability to provide physically interpretable results. However, the several good models made available by the NARMAX modelling framework could be useful in the selection of model structures that also provide insight into the physical behavior of the catchment rainfall-runoff system. The results are a motivation, therefore, to seek a better nonlinear modelling framework that would also provide some plausible interpretation of the nature of the runoff process in the basin.

6 DATA-BASED MECHANISTIC MODELLING OF STREAMFLOWS

6.1 Introduction

The nonlinear rainfall-runoff NARMAX model developed in the previous chapter showed the appropriatenss of nonlinear representation of the rainfall-runoff process. Predictions of runoff were better with fewer parameters than for the linear ARX modeling case, though peak and low flows were not adequately simulated in some cases. However, it still had quite a number of parameters, indicating that the process nonlinearity was not accounted for very well. In addition, it was a purely black-box model, providing no physically interpretable results. Models that characterize the rainfall-flow nonlinearity very well and also provide some insights into the form of the flow process are necessary in hydrology. Such models promote better identification and understanding of the important hydrological issues and thereby assist in the design of better basin water resources management. This chapter focuses on such models and presents a modelling framework that is proposed as the most appropriate for riverflow modelling in the Volta Basin. The utility of properly formulated models cannot be overemphasized. The forecasting, backcasting and flow gap infilling abilities of well formulated streamflow models and the insights into the processes generating and controlling the flows that such models can provide make them very useful tools in many areas of catchment-scale water-resources management. Such areas include flood forecasting and control, drought management, assessment of water supply potentials and waste load carrying capacities of streams, engineering design of hydraulic structures such as on-the-river reservoirs and the assessment of the impacts of anthropogenic effects on both the spatial and temporal distribution of streamflow as well as the water quality of the streams (Tabrizi *et. al*., 1998; Beven, 2000).

6.2 Runoff Models in Hydrology

A wide variety of models for modelling streamflow exists in the literature. An extensive list of the most popular ones has been compiled by Singh and Woolhiser (2002). In general, these models can be grouped into four main categories (Wheater *et. al*., 1993; Beven 2000, Young 2001):

1. **Physics-based or fully distributed models:** These are the most complex and comprehensive rainfall-runoff models. They seek to model all aspects of the subsurface and surface flow processes in a catchment by representing the component processes with partial differential equations derived from consideration of the physics of the processes involved. They attempt to provide local outputs for every part of the model domain, are formulated based largely on the blueprint of Freeze and Harlan (1969) and have a very high number of parameters to be estimated from observed data. These models include the Systeme Hydrologique Europeen (SHE) model (Abbott *et. al.*, 1986a &b; Bathurst *et. al.,* 1995), the Australian TOPOG (Vertessy *et. al*., 1993) and THALES (Grayson *et. al.*, 1995) models, the Institute of Hydrology Distributed Model, IHDM, from the UK (Calver and Wood, 1995) and the Swiss Water balance Simulation Model ETH, WaSiM-ETH (Schulla, 2001).

2. **Conceptual models:** These models vary in complexity but are less ambitious than the fully distributed models. They specify the model structure a priori based on the component processes considered important in the runoff generation process. There are two main classes of models in this category. The first consists of simplified distributed models, also called distribution function models, that use distribution functions to represent the spatial variability of catchment runoff, thereby avoiding the use of detailed process representations of the fully distributed models (Beven, 2000). The TOPMODEL (Beven and Kirkby, 1979) and the Probability Distributed Model, PDM (Moore and Clark, 1981) are examples. The models in the second class are the Explicit Soil Moisture Accounting, ESMA, models (O'Connell, 1991). These models represent the important processes controlling catchment response to rainfall by a system of internal storages linked by mathematical description of the fluxes between them. Examples of such models are the Dawdy and O'Donnell (1965), the Stanford Watershed (Crawford and Linsley, 1966), the Sacramento Soil Moisture Accounting, SAC-SMA (Burnash *et. al*., 1973; Burnash 1995.) and the Xinanjiang (Zhao and Liu, 1995) models and the Australian Large Scale Catchment Model, LASCAM (Sivapalan et. al., 1996a, b, c). The

number of parameters to be calibrated for in most of these models is high but tend to be lower than in the fully distributed models.

3. **Metric models:** These are data-based, mainly black-box models, that rely largely on observed data to characterize the runoff response using statistical estimation or optimization methods. They are often based on the analysis of time series such as the Box-Cox type discrete-time transfer functions (Box *et. al*., 1994) and neural network models (Young, 2001). NARMAX (Non-Linear Autoregressive and Moving Average with eXogenous input) polynomial models are also in this category. These models are often parsimonious in their parameterization and many are able to explain the observed data well. However, it is usually difficult for such models to provide a physically meaningful interpretation of the flow process characterized, and this makes them unattractive as general modelling tools in hydrology.

4. **Hybrid metric-conceptual (HMC) models:** These are usually transfer function models and, as the name implies, combine the essential features of both the metric and conceptual models. In particular, they combine the ability of metric models to efficiently (parsimoniously) characterize the observational data in statistical terms and the advantages of conceptual models that have a prescribed physical interpretation (Young, 2001). They are therefore models of moderate parameterization. Two main modelling approaches can be identified for models in this category:
   a. **Deductive approach** in which the conceptual model structure is specified a priori. The Identification of unit Hydrographs And Component flows from Rainfall, Evaporation and Streamflow data, IHACRES, model (Jakeman *et al*., 1990; Jakeman and Hornberger, 1993; Jakeman *et al*., 1993a; Littlewood and Jakeman, 1994) is an example.
   b. **Data-based mechanistic (DBM)**: An inductive approach to modelling in which the model structure is not pre-specified by the modeler, i.e., the observed data are allowed, as much as possible, to determine the model structure from a more general class of models, and then the model is

interpreted in a physically meaningful manner (Young and Pedregal, 1997; Young, 1998; Young, 2001a). An essential feature in all DBM modelling is the physical interpretation of the identified models so that the selected working models are those that not only explain the observed data well but also can be interpreted in physically meaningful terms. Examples of the application of this approach to rainfall-runoff modelling are Young and Beven (1994), Young et al. (1997), Mwakalila et al. (2001a) and Young (2001).

The most serious drawback of distributed and many conceptual models is the problem of over-parameterization, i.e., they often have too many parameters to be estimated from a limited set of rainfall-flow, meteorological, soil and sub-surface hydraulic data (Loague and Freeze, 1985; Hornberger *et. al.*, 1985; Hooper *et. al*., 1988; Beven, 1989, Jakeman and Hornberger, 1993, Young and Beven, 1994; Young, 2001a). Though a good number of the parameters of these models may have physical meaning and can be estimated from soil, vegetation and other river catchment characteristics, deriving effective values from point measurements for the large scale applications they are intended for can be very problematic. Thus, there is often no unique set of parameters or combination of parameters that explain the observed data used for calibrating such models, unless prior restrictions are imposed on many of these parameters (Young, 1996). Several different parameterizations can usually be identified that explain the observed data equally well – a situation referred to as equifinality (Franks et. al., 1997; Beven, 2000; Beven, 2001; Beven and Freer, 2001).

The identifiability problem of these models may be attributed mainly to the inadequate information content of observed data, which makes it impossible to identify unique models (Jakeman and Hornberger, 1993; Kokkonen and Jakeman, 2001; Young 2001). The low information content of observational data may be due, on the one hand, to the inputs to the dynamic system not being *sufficiently exciting* (Young, 1984, Young 2001), and on the other, to the existence of dominant modes of dynamic systems in general (Young *et. al*., 1996; Hasselman *et. al.*, 1997; Hasselman, 1998; Young, 1998; Young 2001) resulting in the output of the systems being dominated by the cumulative effect of a few dynamic modes. Consequently, several researchers maintain that the information content of a typical set of rainfall-runoff data is enough to support the

estimation of only a small number of parameters (ten or less) and for the identification of non-linear models of dynamic order of up to only three (Kirby, 1976; Hornberger *et. al*., 1985, Jakeman and Hornberger, 1993; Wheater et. al., 1993; Young and Beven, 1994; Ye *et. al*., 1998; Young, *et. al*., 1997, 1998; Young, 2001). In the opinion of Young (2001), the search for a single, all encompassing model of any dynamic system is futile; therefore, the model-builder should look for *a model that suits the nature of the study objectives*.

HMC models appear then to be the models to use for rainfall-runoff modelling. They generally contain few parameters and while the identifiability problem is not entirely eliminated, it is reduced. Also, they usually have the necessary functionality to adequately represent the rainfall-runoff nonlinearity and lend themselves to plausible physical interpretation of the flow mechanisms characterized. The IHACRES model, for example, has only 6 or 7 parameters to be estimated from observational data (compared to up to 17 in the SAC-SMA model) but has been used successfully in a wide range of hydrological applications, including rainfall-runoff modelling, catchment characterization and assessment of the impact of climate change on the hydrological response of catchments and low flow analysis, in several regions and under various climatic conditions (Jakeman and Hornberger, 1993; Jakeman *et al*., 1993a,b; Sefton *et al.*, 1995; Post and Jakeman, 1996; Post *et. al*., 1996; Littlewood and Marsh, 1996; Littlewood *et. al*., 1996; Hansen *et. al*., 1997). In addition, the statistical procedures for assessing the validity and quantifying the uncertainties in both the parameter estimates and the outputs of HMC models have been well developed (Young, 1984;Young and Beven, 1994; Young, 2001).

Experience from extensive theoretical analysis and practical application of rainfall-runoff models in this category show that in hydrology, parallel processes, in which catchment runoff for example can be decomposed into parallel pathways of fast and slow flows, appear to be the rule (Young, 1992; Jakeman and Hornberger, 1993; Jakeman *et. al*., 1993; Young and Beven, 1994, Ye *et. al*., 1997,1998;Young *et. al.*, 1997; Young, 2001). The ability of these models to provide such important and useful information on the flows they characterize adds to their utility and, together with their parametric parsimony, enhances their credibility and makes them very attractive in catchment rainfall-runoff modelling.

In this chapter, monthly riverflows in selected catchments in the Volta Basin are modeled using a HMC rainfall-runoff modelling framework. This involves the use of the Fixed Interval Smoothing, FIS, algorithm (Young 1984, 88, 89) to estimate the parameters of a linear time-varying, state-dependent parameter (LTV-SDP) transfer function relating monthly catchment runoff to rainfall. The aim of the study is to obtain models that explain the observed data very well and in addition provide plausible decomposition of the runoffs in the basin into various parallel component (fast and slow or delayed) flows. The study also seeks to establish the relationship between catchment wetness index (computed from rainfall and potential evaportranspiration, ETP) and effective rainfall in the basin. This relationship would be useful in the estimation of the runoff series for ungaged catchments in the basin with measured rainfall and temperature or computed ETP.

In the rainfall-runoff modelling presented in this chapter, only **linear recursive causal** filters will be considered, with the use of **time varying, state-dependent coefficients** to characterize any **nonlinearities** in the rainfall-runoff transformation. The nonlinear rainfall-runoff modelling framework is developed in two steps. In the first step, a nonlinear transformation of catchment rainfall to effective rainfall using a linear time-varying (LTV) filter with state-dependent parameters is established. A linear time-invariant (LTI) relationship between effective rainfall and river runoff is then determined in the second step. Figure 6.1(a) illustrates the two-step rainfall-runoff modelling process.

Of particular interest is the plausibility of the modeled runoff being decomposed into two parallel components, i.e., slow or delayed and fast flows (Figure 6.1(b)). The slow flow would represent the sum of subsurface flow and baseflow or delayed flow in general, while the fast flow would be from surface flow and interflow (direct runoff).

6.2.1 Difference equation representation of the rainfall-runoff linear filter

The *difference equation* is one of the time domain linear filter formulas used in computing an output sample of a signal, such as runoff, at time t based on past and present input samples of the input signal (rainfall) and past output samples. The general

LTV difference equation (*explicit finite difference scheme*) for runoff at any sampling point t is given as follows:

$$y_t = -a_{1t}y_{t-1} - a_{2t}y_{t-2} - \ldots - a_{nt}y_{t-n} + b_{0t}r_{t-d} + b_{1t}r_{t-d-1} + \ldots + b_{1m}t_{t-d-m} \qquad (6.1a)$$

where

y_t = river runoff at sampling instant t,

r_t = rainfall at sampling instant t,

a_{1t}, a_{2t}, …, a_{nt} are feedback, and b_{0t}, b_{1t}, …, b_{mt} are feedforward, time varying, real coefficients.

d sampling intervals = pure time delay in the rainfall input.

Equation 6.1a is a causal recursive LTV real digital filter.

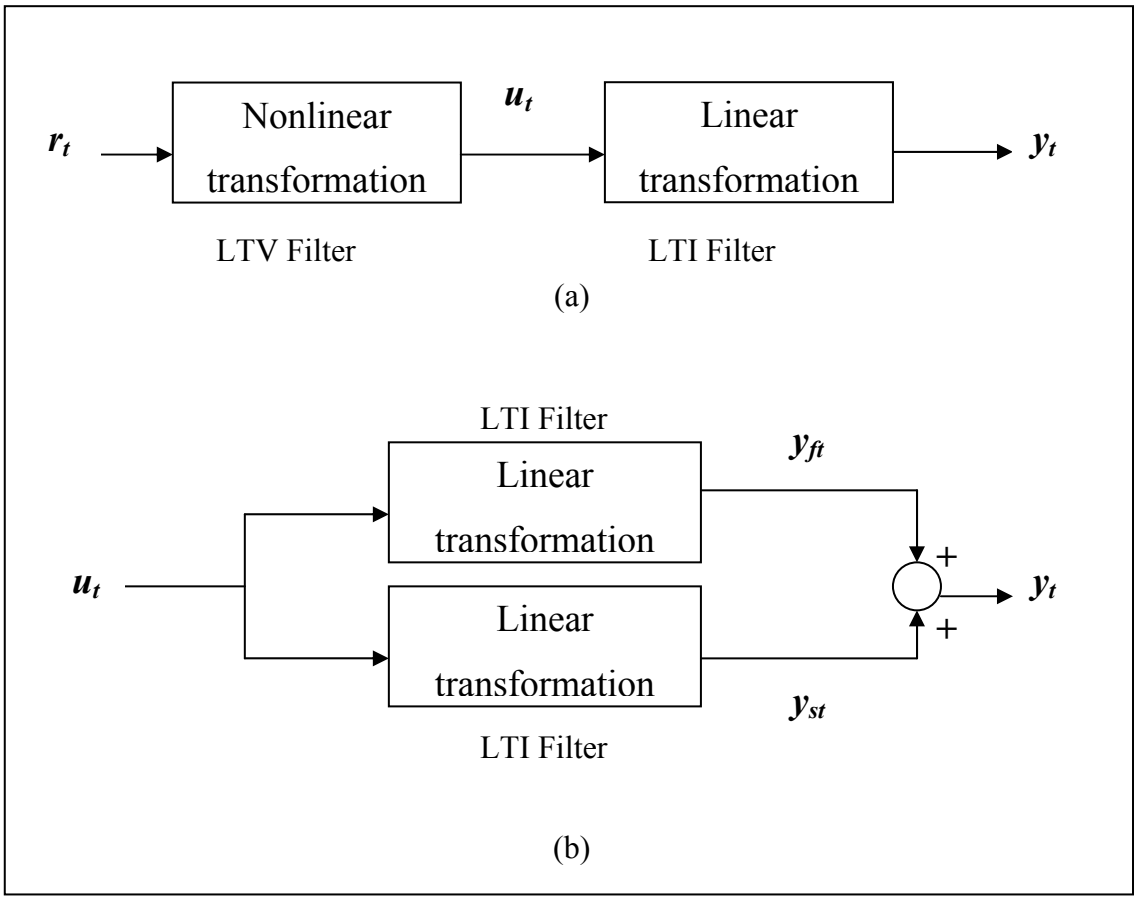


Figure 6.1 The rainfall-runoff transformation

(a) Rainfall (r_t) → Effective Rainfall (u_t) → Total Runoff (y_t) transformations

(b) Effective Rainfall transformations to two parallel flow components – Fast flow (y_{ft}) and Slow flow (y_{st}) components

6.2.2 Transfer function representation of the rainfall-runoff linear time-varying (LTV) filter

If z^{-1} is the backward shift operator such that $z^{-i}r_t = r_{t-i}$, then Equation 6.1a can be re-written as:

$$y_t = -a_{1t}\left(z^{-1}y_t\right) - a_{2t}\left(z^{-2}y_t\right) - \ldots - a_{nt}\left(z^{-n}y_t\right) + b_{0t}r_{t-d} + b_{1t}\left(z^{-1}r_{t-d}\right) + \ldots + b_{mt}\left(z^{-m}t_{t-d}\right) \quad (6.1b)$$

$$\Rightarrow y_t + \left(a_{1t}z^{-1} + a_{2t}z^{-2} + \ldots + a_{nt}z^{-n}\right)y_t = \left(b_{0t} + b_{1t}z^{-1} + \ldots + b_{mt}z^{-m}\right)r_{t-d} \quad (6.1c)$$

$$\Rightarrow A_t\left(z^{-1}\right)y_t = B_t\left(z^{-1}\right)r_{t-d} \quad (6.1d)$$

$$\Rightarrow y_t = \frac{B_t\left(z^{-1}\right)}{A_t\left(z^{-1}\right)}r_{t-d} \quad (6.1e)$$

where

$$\mathrm{A_t}\left(\mathrm{z}^{-1}\right) = \left(1 + \mathrm{a_{1t}z^{-1}} + \mathrm{a_{2t}z^{-2}} + \ldots + \mathrm{a_{nt}z^{-n}}\right)$$

$$\mathrm{B_t}\left(\mathrm{z}^{-1}\right) = \left(\mathrm{b_{0t}} + \mathrm{b_{1t}z^{-1}} + \ldots + \mathrm{b_{mt}z^{-m}}\right)$$

Equation 6.1e is the Linear Time Varying (LTV) transfer function (with time variable parameters, TVP) representation of the rainfall-runoff digital filter. The ratio $\frac{\mathrm{B_t}\left(\mathrm{z}^{-1}\right)}{\mathrm{A_t}\left(\mathrm{z}^{-1}\right)} = \mathrm{h_t}\left(\mathrm{z}^{-1}\right)$ is the time dependent transfer function. Stochasticity can be introduced into the deterministic or process model represented in Equation 6.1e by adding a noise term to it so that the general stochastic LTV transfer function model is then given as:

$$y_t = \frac{B_t\left(z^{-1}\right)}{A_t\left(z^{-1}\right)}r_{t-d} + \varepsilon_t \quad (6.2a)$$

or

$$y_t = \frac{b_{0t} + b_{1t}z^{-1} + \ldots + b_{mt}z^{-m}}{1 + a_{1t}z^{-1} + a_{2t}z^{-2} + \ldots + a_{nt}z^{-n}}r_{t-d} + \varepsilon_t \quad (6.2b)$$

where ε_t is a noise term that is the source of stochasticity in the runoff. The vector [n m d] is used as a short hand representation of the process part of the model in Equation 6.2 with n feedback, m feedforward coefficients and d samples time delay.

The model given in Equation 6.2 is used as the model describing the overall behavior of the rainfall-flow dynamics in the presentation given here. It is first used to determine effective rainfall in relation to the input or output series (the nonlinear part of the analysis). When the form of the effective rainfall is determined, the time invariant form of Equation 6.2 is then used to model the effective rainfall-runoff transformation (the linear part of the analysis). The time invariant form is given as:

$$y_t = \frac{B\left(z^{-}\right)}{A\left(z^{-1}\right)} u_{t-d} + \varepsilon_t \qquad 6.3a)$$

or

$$y_t = \frac{b_0 + b_1 z^{-1} + \ldots + b_m z^{-m}}{1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_n z^{-n}} u_{t-d} + \varepsilon_t \qquad (6.3b)$$

where u_t is effective rainfall at sample t and $\frac{\mathrm{B}\left(\mathrm{z}^{-1}\right)}{\mathrm{A}\left(\mathrm{z}^{-1}\right)} = \mathrm{h}\left(\mathrm{z}^{-1}\right)$ the Linear Time Invariant (LTI) transfer function. Note that, for example, equations 6.3a and 6.3b can be expressed equivalently as:

$$y_t = \frac{B\left(z^{-d}\right)}{A\left(z^{-1}\right)} u_t + \varepsilon_t \qquad (6.3c)$$

or

$$y_t = \frac{b_0 z^{-d} + b_1 z^{-1-d} + \ldots + b_m z^{-m-d}}{1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_n z^{-n}} u_t + \varepsilon_t \qquad (6.3d)$$

with B now a function of z^{-d} and the delay removed from u_t.

In general, when the poles (roots of the denominator polynomial) of Equation 6.3.a are real then the system can always be interpreted unambiguously as a series

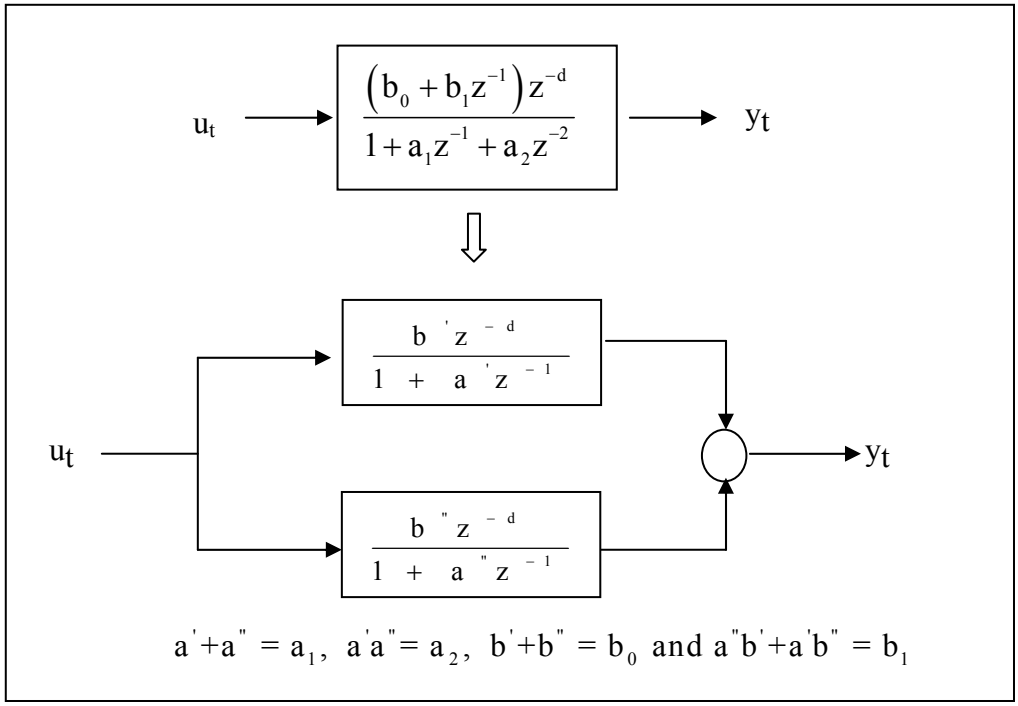
and/or parallel connection of first- and second-order sub-systems (Young, 1992). So, for example, the LTI transfer function models

$$y_t = \frac{\left(b_0 + b_1 z^{-1}\right) z^{-d}}{1 + a_1 z^{-1} + a_2 z^{-2}} u_t \tag{6.4a}$$

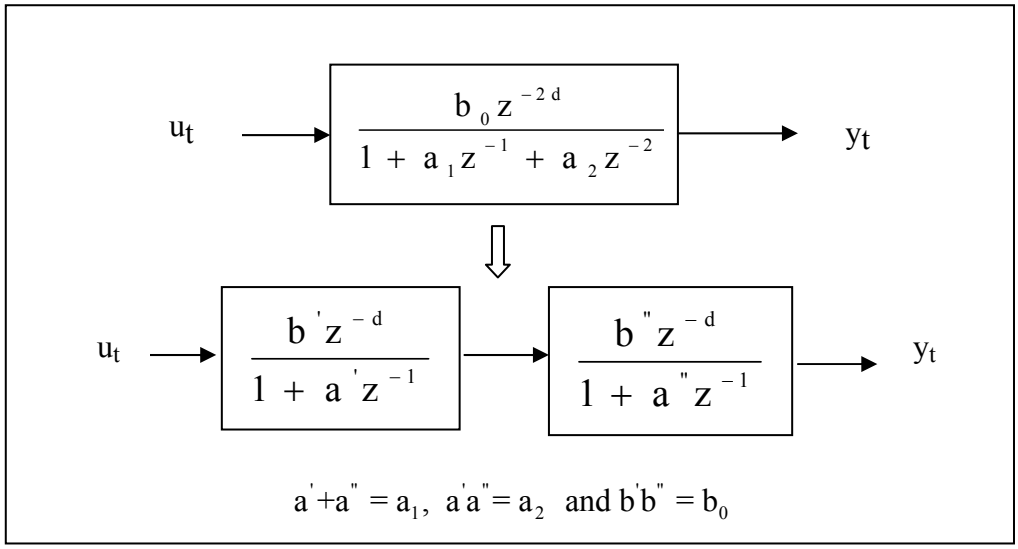
and

$$y_t = \frac{b_0 z^{-2d}}{1 + a_1 z^{-1} + a_2 z^{-2}} u_t \tag{6.4b}$$

can be decomposed into parallel and serial connections, respectively, of first-order sub-systems as shown in Figure 6.2.



(a) Parallel



(b) Serial

Figure 6.2. Decomposition of second order transfer functions

6.3 The HMC modelling framework

The HMC modelling framework used is based on linear time-varying, state-dependent parameter (LTV-SDP) transfer function modelling. The state dependency of the parameters allows them to vary in such a manner as to adequately account for the

nonlinearity in the rainfall-effective rainfall transformation. The procedure for this modelling framework is summarized below based on the outline for a Single-Input-Single-Output (SISO) system in Young and Beven (1994).

1. Start from a linear time invariant (LTI) transfer function relating the output variable y_t (e.g., catchment monthly runoff) to the primary input variable r_t (e.g., catchment monthly rainfall). Use this LTI model and the input-output series to determine [na nb d] of the simplest model that fits the data relatively well (na = number of feedback coefficients; nb = number of feedforward coefficients and d the time delay in the input). The model selection can be achieved by employing a combination of model order identification criteria such as the Akaike Information Criterion (AIC; Akaike, 1974), the Young Information Criterion (YIC; Young, 1989) and the coefficient of determination, R_T^2 (Nash-Sutcliffe Efficiency, NSE (Nash and Sutcliffe, 1970) in hydrology). This avoids over-parameterization of the LTV-SDP model to be fitted in the next step.
2. Proceed to a LTV-SDP transfer function relating y_t to r_t with the same order as identified in step 1 above using the relationship defined in Equation 6.2 and repeated in Equation 6.5 below:

$$y_t = \frac{B_t\left(z^{-1}\right)}{A_t\left(z^{-1}\right)} r_t + \varepsilon_t = \frac{b_{0t} + b_{1t}z^{-1} + \ldots + b_{mt}z^{-m}}{1 + a_{1t}z^{-1} + \ldots + a_{nt}z^{-n}} r_{t-d} + \varepsilon_t. \qquad (6.5)$$

$$t = 1, 2, \ldots, N$$

where

N is the number of samples in the output-input series,

$a_{1t}, \ldots, a_{nt}, b_{0t}, b_{1t}, \ldots, b_{mt}$ are n+m+1 are time-varying state-dependent parameters (TV-SDPs) that vary in such a manner as to account for the non-linearities and/or non-stationarities in the observed series.

z^{-1} is the backward shift operator, i.e., $z^{-i}r_t = r_{t-i}$

ε_t is a noise term that is statistically independent of the input variable r_t

d is the input delay, measured in sampling intervals.

3. Apply the Fixed Interval Smoothing (FIS) method of recursive estimation to the alternative vector form of Equation 6.5, given in Equation 6.6 below, to estimate the state dependent parameters of the LTV transfer function.

$$y_t = z_t^T a_t + e_t, \qquad e_t \sim N(0, \sigma^2) \tag{6.6}$$

where:

$z_t^T = (-y_{t-1} \quad -y_{t-2} \quad .... \quad -y_{t-n} \quad r_{t-d} \quad r_{t-d-1} \quad ... \quad r_{t-d-m})$, 1 by np vector,

np = n+m+1,

$a_t = (a_{1t} \quad a_{2t} \quad .... \quad a_{nt} \quad b_{0t} \quad b_{1t} \quad .... \quad b_{mt})^T$, np by 1 vector,

e_t is a stochastic noise vector with variance σ^2, is statistically independent of the input variable r_t and is related to ε_t in equation 6.3 by $\varepsilon_t = \frac{1}{A_t(z^{-1})} e_t$.

$(....)^T$ refers to the transpose of the vector.

When applied to equation 6.6, the FIS algorithm yields $\hat{a}_{t/N}$, an estimate of a_t at each sampling point t based on all N samples of the observed series. Also provided at each sampling point of the input-output series is an estimate of the covariance of the estimated parameters, $P^*_{t/N} = E\left\{\left(\tilde{a}_{t/N}\right)\left(\tilde{a}_{t/N}\right)^T\right\}$, also based on the N samples with $\tilde{a}_{t/N} = a_t - \hat{a}_{t/N}$, the estimation error (actual-estimated values).

4. Identify the elements of $\hat{a}_{t/N}$ that vary significantly over time (usually by visual means, from graphical plots of the elements of $\hat{a}_{t/N}$ vs. time). Then, while constraining all other parameters to constants, re-estimate $\hat{a}_{t/N}$ from the FIS algorithm and the input-output series.

5. Determine which of the variables in the vector **z** of Equation 6.6 is significantly related to the significantly varying state dependent parameters identified in step 4 above. This can be done through scatter plots and other correlation analyses. Then establish a functional relationship between these variables and the

significantly varying parameters by means of a weighted least squares estimation with weights given by the reciprocals of the estimated variances of the parameters, i.e., the reciprocals of the diagonal terms in $P^*_{t/N}$ for the significantly varying parameters.

6. Replace the input variable with the new variable, $u_{t,}$, obtained through the functions determined in step 5. It should then be possible to fit a final LTI model, with constant parameters, to the output variable $y_{t,}$, the new primary input variable u_t and any other identified secondary input variables. In rainfall-runoff modelling, u_t is effective rainfall and potential evapotranspiration or temperature is secondary input.

6.4 Fixed interval smoothing (FIS) method of parameter estimation of LTV-SDP models

Parameters used in the characterization of the input-output relationship of highly non-linear systems, such as the runoff generating process from rainfall, need to vary rapidly with time in order to adequately characterize the systems. They are, therefore, better estimated in a SDP setting, as the state dependency allows the necessary rapid temporal parametric changes commensurate with the variation in the system variables. The recursive fixed-interval smoothing (FIS) algorithm, combined with special data re-ordering and `back-fitting' procedures, has been widely used in non-linear systems identification, including catchment rainfall-runoff modelling, to obtain estimates of state dependent parameter variations (Young, 1984, 1988, 1989, 1993; Young and Beven, 1994; Young and Pedregal, 1998, Young *et **al***., 1999; Young *et al*., 2001). As a time series of observed data of fixed length or interval is also available, the estimation of the parameters can be undertaken in two directions, i.e., forward pass filtering using the versatile Kalman filter algorithm (Kalman, 1960) and backward pass smoothing (Bryson and Ho, 1969) using the optimal fixed interval smoothing method (Young, 1999; Young and Beven, 1994). Non-parametric methods, including graphical methods, are then employed to establish the relationship between the estimated SDPs and the associated state or observed variable(s). The FIS algorithm was used in this study to estimate the significantly varying parameters of the LTV-SDP model presented in Equation 6.6. The relationship of these parameters with the rainfall and runoff series

was then examined, and the series with a significant relationship identified. An appropriate function relating this series to the estimated time varying parameters was then established and used with the rainfall series to estimate the effective rainfall series. An outline of the SDP estimation procedure exploiting the powerful FIS algorithm is given below (details in the references cited above).

Consider the vector form of the Single-Input-Single–Output (SISO) LTV-SDP transfer function given in Equation 6.6 above. Each of the time variable parameters, i.e., each of the np elements of a_t , $a_{i,t}$, may be represented by a two-dimensional state vector $\boldsymbol{x}_{i,t} = \begin{bmatrix} l_{i,t} & s_{i,t} \end{bmatrix}^T$, where $l_{i,t}$ and $s_{i,t}$ are respectively the changing level (magnitude) and slope of the associated TV-SDP. The stochastic evolution of each $x_{i,t}$ can be described by the following Generalized Random Walk (GRW) process

$$x_{i,t} = F_i x_{i,t-1} + G_i \eta_{i,t} \qquad \text{i} = 1, 2, \ldots, \text{np} \tag{6.7}$$

where:

$$F_i = \begin{bmatrix} \alpha & \beta \\ 0 & \gamma \end{bmatrix}, \quad G_i = \begin{bmatrix} \delta & 0 \\ 0 & 1 \end{bmatrix}$$

α, β, γ and δ are *hyper parameters* that have to be estimated first from the observed data before the TVPs can be computed, and $\eta_{i,t}$ is a noise vector. Special cases of the GRW are the Integrated Random Walk, IRW ($\alpha = \beta = \gamma = 1$; $\delta = 0$) and the scalar Random Walk, RW (($\alpha = \beta = \delta = 0$; $\gamma = 1$).

By combining equations 6.6 and 6.7, the following state-space equations are obtained:

Observation equation: $$y_t = H_t x_t + e_t \tag{6.8ai}$$

State equation $$x_t = F x_{t-1} + G \eta_t \tag{6.8aii}$$

where

$$x_t = \begin{bmatrix} x_{1,t}^T & x_{2,t}^T & \ldots & x_{np,t}^T \end{bmatrix}^T \tag{6.8b}$$

F is a 2np × 2np block diagonal matrix with blocks defined by the F_i matrices in Equation 6.7; G is a 2np × 2np block diagonal matrix with blocks defined by the corresponding subsystem matrices G_i in Equation 6.7; and η_t is a 2np-dimensional vector containing the white noise input vectors $\eta_{i,t}$ in the equation. The white noise inputs provide the stochastic stimulus for parametric change in the model and are assumed to be independent of the observation noise e_t. H_t is a 1 × 2np vector of the following form:

$$H_t = [-y_{t-1} \;\; 0 \;\; -y_{t-2} \;\; 0 \ldots -y_{t-n} \;\; 0 \;\; u_{t-d} \;\; 0 \ldots u_{t-d-m} \;\; 0] \tag{6.8c}$$

Note that the number of parameters associated with y is n and that associated with u is m+1, so that the total number of parameters is np = n+m+1. Thus H_t in Equation 6.8c is a vector with 2np elements (number of zeroes is np).

The FIS algorithm, in relation to the series y_t of fixed interval N samples, i.e., t = 1, 2, …, N, is given as:

1. Forward pass recursive filtering equations (from beginning of sample set to end)

Prediction:

$$\hat{x}_{t|t-1} = F\hat{x}_{t-1} \tag{6.9ai}$$

$$P_{t|t-1} = FP_{t-1}F^T + GQ_rG^T \tag{6.9aii}$$

Correction:

$$\hat{x}_t = \hat{x}_{t|t-1} + P_{t|t-1}H_t^T\left[1 + H_tP_{t|t-1}H_t^T\right]^{-1}\left\{y_t - H_{t|t-1}\right\} \tag{6.9aiii}$$

$$P_t = P_{t|t-1} - P_{t|t-1}H_t^T\left[1 + H_tP_{t|t-1}H_t^T\right]^{-1}H_tP_{t|t-1} \tag{6.9aiv}$$

2. **Backward pass smoothing equations (from end of the sample set to beginning)**

$$\hat{x}_{t|N} = F^{-1}\left[x_{t+1|N} + GQ_rG^TL_t\right] \tag{6.9bi}$$

$$L_t = \left[I - P_{t+1} H_{t+1}^T H_t\right]^T \left[F^T L_{t+1} - H_{t+1}^T \left\{ y_{t+1} - H_{t+1} \hat{x}_{t+1} \right\} \right], \mathbf{L}_N = 0 \qquad (6.9bii)$$

$$P_{t|N} = P_t + P_t F^T P_{t+1|t}^{-1} \left[P_{t+1|N} - P_{t+1|t} \right] P_{t+1|t}^{-1} F P_t \qquad (6.9biii)$$

In the above algorithms, $Q_r = \frac{Q}{\sigma^2}$ and $P_t = \frac{P_t^*}{\sigma^2}$. Q_r is called the Noise Variance Ratio (NVR) matrix, Q, a 2np x 2np matrix (usually assumed diagonal), is the covariance of the white noise input η_t, and P_t^* is the error covariance matrix associated with the state estimates. If e_t is a discrete white noise, then P_t^* would be an accurate estimate of the covariance of the estimated TV-SDPs and thereby be an accurate representation of the estimated uncertainty in the parameters.

The NVR matrix Q_r is unknown prior to the FIS analysis, and so it and all unknown hyper-parameters in the state space model represented in equation 6.8 have to be estimated from the time series data y_t and r_t through appropriate optimization procedures before the FIS algorithms can be used.

6.5 Application of the modelling framework to rainfall-runoff series in the Volta Basin

For each river catchment studied, the LTV-SDP modelling framework as outlined above was applied first to the rainfall and runoff series of the data sets to determine effective rainfall as specified in steps 1-5 in the framework. The ETP series was added later as secondary input, after fitting to the effective rainfall series, to estimate the final LTI model. Model order identification was undertaken by examining the AIC, YIC and NSE order selection criteria together. The best models were those with low AIC, very negative YIC and high NSE. For the final LTI model identification, the model that, in addition to satisfying the above criteria, also provided a decomposition of the modeled flow into plausible fast and slow parallel component flows was chosen. Obvious feasible models, considering the last criterion, would include those with [na nb d] = [1 2 d], [2 2 d], [2 3 d] and [3 2 d]. If the poles (roots of the denominator polynomial) of the respective transfer functions are real, these models (with any identified d) provide decomposition of the flow into a bypass and 1 linear store, 2 linear stores in parallel, a

bypass and 2 linear stores in parallel and a linear store in parallel with 2 others in series, respectively. The time constants or residence times of the identified stores and the steady state gains were also computed.

If, for example, an identified [2 2 0] model given as:

$$y_t = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}} u_t + \varepsilon_t \tag{6.10a}$$

has real poles, then y_t can be unambiguously decomposed into the following parallel forms:

$$y_t = \frac{b'}{1 + a' z^{-1}} u_t + \frac{b''}{1 + a'' z^{-1}} u_t \tag{6.10b}$$

where:

$$a'+a''= a_1,\ a'a''= a_2,\ b'+b''= b_0 \text{ and } a''b'+a'b''= b_1$$

since from the above, $a''= a_2 / a'$ so that $a'+a_2 / a'= a_1$, then

$$(a')^2 - a_1 a'+a_2 = 0 \tag{6.10c}$$

The solution (real, as the poles of the TF are assumed real) to the quadratic Equation 6.10c provides the values of a' and a''.

The steady state gains (SSGs) in Equation 6.10b are $\mathrm{SSG1} = \frac{\mathrm{b'}}{1+\mathrm{a'}}$ and $\mathrm{SSG2} = \frac{\mathrm{b''}}{1+\mathrm{a''}}$, while the time constants (TCs) are given as TC1 = -Δt/log(-a') and TC2 = -Δt/log(-a'') where Δt is the time step. Obviously, both b' and b'' should be positive, while a' and a'' are both negative for the steady state gains and time constants to have physical meaning. The percentage of the total flow occurring in each pathway following such decompositions was obtained from the steady state gains.

Relevant functions in the CAPTAIN toolbox (Young *et. al.*, 2004) for Matlab (Mathworks, 2002) were used for all model identification and for both LTV-SDP and

constant parameter estimations. The following forms of the model selection criteria were used with the YIC form taken from Beven and Young, 1994.

$$NSE = 100\left(1 - \frac{\sigma_e^2}{\sigma_y^2}\right) \tag{6.11}$$

$$YIC = log_e \frac{\sigma_e^2}{\sigma_y^2} + log_e\left(NEVN\right) \tag{6.12a}$$

$$NEVN = \frac{1}{np}\sum \frac{\sigma_a^2 \cdot \hat{p}_{ii}}{\hat{a}_i^2}. \tag{6.12b}$$

$$AIC(np) = logV + \frac{2np}{N} \tag{6.13}$$

where:

σ_e^2 is the variance of the model residual; σ_y^2 the variance of the observed output series, y_t; np = m+n+1 is the number of estimated parameters in the $\hat{a}_{t/N}$ vector; $\hat{p}_{ii}$ is the i^{th} diagonal element of the $P^*_{t/N}$ covariance matrix obtained in the FIS analysis ($\sigma_e^2 \hat{p}_{ii}$ being an estimate of the variance of the estimated uncertainty on the i^{th} parameter estimate); a_i the i^{th} parameter estimate in the $\hat{a}_{t/N}$ vector and V is the sum of squares of the model residuals.

The NSE evaluates a model based on how well it explains the data, so that the closer this value is to 100% the better the model; the AIC selects the one that explains the data well and is parsimonious in the number of its parameters. The lower the value of the AIC the better the model. The YIC also selects models that explain the data well but penalizes them for large uncertainties in parameter estimates; the more negative the YIC, the better.

The set of rainfall, ETP and runoff series for a catchment was split into two different sets - estimation and validation series. The estimation series was used in the FIS analysis, the determination of effective rainfall and the estimation of the parameters

in the final LTI model. When effective rainfall was found to be a function of the observed runoff, the following form of the parametric non-linear rainfall filter used in various formulations of the IHACRES application (Jakeman ***et al***., 1990; Jakeman and Hornberger, 1993; Jakeman ***et al***., 1993) was fitted with the effective rainfall obtained in the estimation mode and used to model effective rainfall in the validation mode (as the runoff series is then assumed to be unknown).

$$u_t = S_t^B r_t \tag{6.14a}$$

$$S_t = 0.5(s_t + s_{t-1}) \tag{6.14b}$$

$$s_t = gr_t + \left(1 - \frac{1}{\tau_t}\right) s_{t-1} \tag{6.14c}$$

$$\tau_t = \tau_w e^{f(ETP_m - ETP_t)} \tag{6.14d}$$

Here B (> 0) is a free parameter to be estimated; s_t is a catchment wetness index; g is a scaling parameter that constraints s_t to have values between 0 and 1 and is also to be estimated from the data; τ_w is catchment drying time constant at a reference ETP, ETP_m; f is a parameter that controls the sensitivity of the catchment drying time constant, τ_t , to ETP_t, the ETP at sampling instant t. For all the catchments studied, the initial catchment index, s_0, was fixed at 0, τ_w at 1 and ETP_m at 240 mm, close to the maximum ETP for the catchments studied. Since equation 6.14 was used during model validation when u_t was estimated as $f(y_t).r_t$ in calibration mode, this estimate of u_t was used to fit the model in the equation and to estimate the parameters β, f and g.

6.6 Results and discussion

The results from each stage of the modelling process are presented and discussed for one catchment only, since similar discussions would also apply to the other catchments. However, where appropriate, selected results for all the catchments studied are presented and discussed.

The first 4 columns of Table 6.1 show the top five (based on YIC) models identified for the catchment in the model order identification for use in the LTV modelling. The simplest, which has a low YIC, relatively high NSE and both a and b

parameters (ie, neither n nor m is zero), is [1 1 1]. A similar result was obtained for the rest of the catchments with selected orders being [1 1 1] or [1 1 0] in some cases. Thus, only two time-variable parameters, one associated with runoff (the a parameter) and the other with rainfall (b parameter) were estimated in each case from the FIS algorithm. Figure 6.1a is a plot of the estimated values of these parameters, with the dotted lines representing the values ±2 times their estimated standard errors, for Bamboi. Clearly, the a parameter varies much less than the b parameter, so it was maintained constant for the next application of the FIS algorithm. The final estimates are shown in Figure 6.1b with the a parameter now constant and the b parameter being state dependent. The last two columns of Table 6.1 show the NSE for the LTV model fit with parameter a being state dependent (i.e., varying with time) in the first column and with a remaining constant in the second for the model [1 1 1] used in the estimation. These NSE values indicate that the FIS algorithm was successful in identifying the necessary variation in the b parameter to enable the model to represent the flow process accurately. Similar results were obtained for the rest of the catchments.

In all cases, the state-dependent b parameter was found to be correlated with runoff and not with rainfall. Figure 6.2 is a plot of the parameter and ±2 times its standard error against monthly runoff, for Bamboi. The data have been sorted in order of runoff. The weighted least squares power fit, with an NSE of 96.61%, is also shown on the graph. The weighted least squares linear fit (not shown on the graph) was obtained with a NSE of 95.93%. In each case, the weights used were the reciprocals of the estimated variances of the b parameter at each sampling instant. The slightly better fit power law relationship was found to be the case for the other catchments. The fitted parameters of the functional relationship $b_t = \alpha y_t^{\beta}$, between the state dependent b parameter and monthly runoff, for all the catchments studied are given in Table 6.2. The very high NSE values for all catchments indicate that a good estimate of the effective rainfall for these catchments can be obtained as $u_t = cr_t y_t^{\beta}$, the value of c selected such that total effective rainfall equals total runoff for the calibration period for each case. Thus, a surrogate (catchment runoff) replaces the catchment wetness index (such as given in Equation 6.14) that is used to define a rainfall filter for the estimation of effective rainfall in the IHACRES type models.

Table 6.1. Top 5 models from LTI model selection for LTV-SDP modelling for Bamboi on the Black Volta River (ordered by YIC for estimation series NSE of at least 65%; LTV-SDP NSEs are for selected model [1 1 1] used in the LTV-SDP analysis)

Model Order [m n d]	**YIC**	**AIC**	**NSE (%) (LTI)**	**LTV-SDP NSE (%)**	
				Varying a	**Constant a**
[1 1 1]	**-4.49**	**3.0754**	**65.29**	**98.35**	**99.17**
[1 2 0]	-3.25	2.9325	70.27	-	-
[2 3 0]	-2.98	2.9557	70.28	-	-
[3 2 0]	-1.72	2.9262	71.15	-	-
[2 1 1]	-0.49	3.0883	65.25	-	-

The non-linear rainfall filter defined in Equation 6.14 was then fitted to the u_t (of the calibration series) as defined above for each catchment in order to use the validation series in the selection of a final model for each basin. The estimated values of the B, f and g parameters used in Equation 6.14 are shown in Table 6.3 together with the NSE values for both the estimation and validation series. The NSE values show good fits between the estimated and predicted effective rainfall for all the catchments, and so the non-linear rainfall filters defined in Equation 6.14 with these parameters can be used to model the effective rainfall in the validation mode for the catchments.

The top 5 (based on YIC and for NSE > 90%) LTI models identified from the y_t and u_t calibration series for Bamboi are presented in Table 6.4. The validation NSE values confirm that these models represent the observed data very well and can all be used for the prediction of monthly runoff in the catchment. However, the second and third models ([2 1 0] and [3 1 0]) are respectively equivalent to 2 and 3 stores in series. They, therefore, do not decompose into parallel flow pathways like the other 4 models, and so are excluded from the feasible model set – the feasible set thus consisting of the remaining 4 models. The feasible models, their parameter estimates and the estimated standard errors of the parameters (in brackets) are presented in Table 6.5.

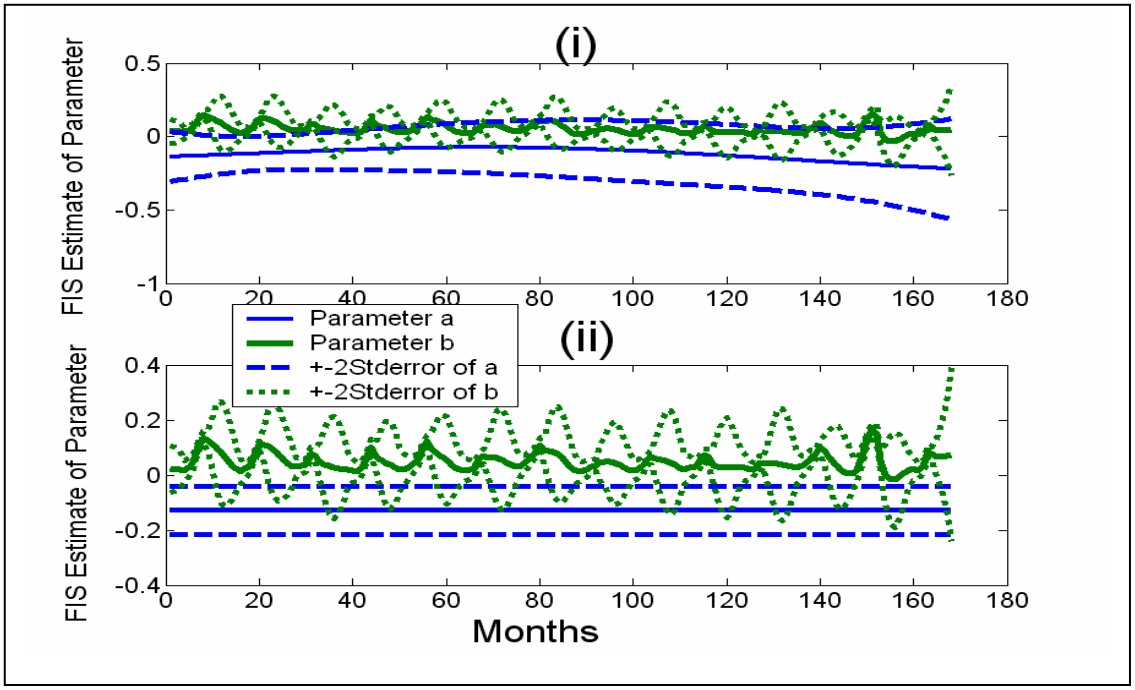


Figure 6.1 FIS estimated a and b parameters for Bamboi, Black Volta River
(i) Both a and b state dependent
(ii) Only b state dependent

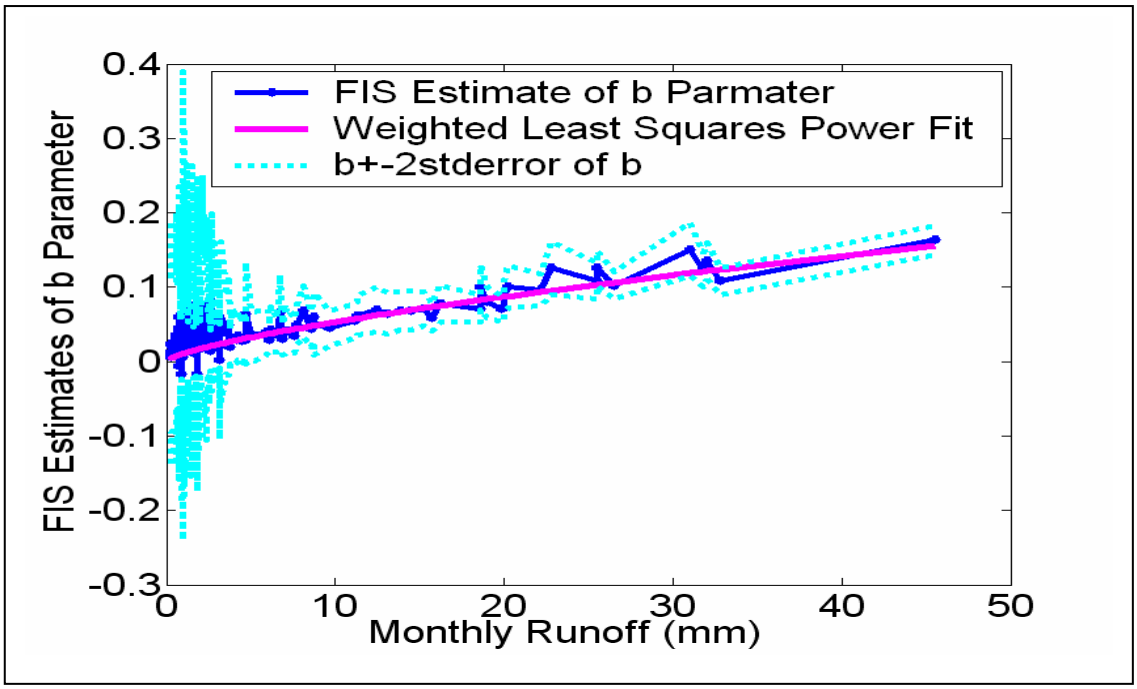


Figure 6.2 FIS estimates of the b parameter vs. monthly runoff for Bamboi, Black Volta River

Table 6.2 Parameter estimates from the weighted least squares power fit of monthly runoff to the FIS estimated b parameter, i.e., $b = \alpha y^{\beta}$ (WVB = White Volta Basin; OB = Oti Basin)

Gauging station, river	**Catchment area (km^2)**	α	β	**NSE(%)**
Lawra, Black Volta	96,000	0.005861	0.7989	94.73
Dapola , Black Volta	96,437	0.006015	0.7817	94.27
Bui, Black Volta	111,853	0.007344	0.7892	96.47
Bamboi, Black Volta	134,200	0.007947	0.7831	98.35
Yagaba, Kulpawn (WVB)	9,100	0.009160	0.8034	96.25
Nabogo, Nabogo (WVB)	3,040	0.008365	0.8178	95.85
Porga, Oti	27,197	0.011929	0.6654	97.56
Mango, Oti	36,287	0.011449	0.7126	98.21
Koumangou, Koumangou (OB)	6,070	0.010534	0.7919	94.73
Sabari, Oti	72,775	0.010916	0.7589	96.37

Table 6.3 Non-linear rainfall filter parameter estimates

Gauging station	**B**	**f**	**g**	**NSE (%)**	
				Estimation series	**Validation series**
Lawra	1.7817	0.027198	0.0004091	87.87	88.80
Dapola	1.7537	0.016005	0.0004895	89.28	84.63
Bui	2.7061	0.012476	0.0010461	83.62	92.37
Bamboi	2.0266	0.020744	0.0005847	85.90	90.09
Yagaba	3.3657	0.006191	0.0018730	89.25	88.59
Nabogo	3.6898	0.006506	0.0018538	91.91	87.49
Porga, Oti	2.0714	0.010491	0.0008728	95.48	91.86
Mango, Oti	2.6275	0.010022	0.0011999	95.14	92.76
Koumangou	2.4808	0.007356	0.0015325	92.86	92.45
Sabarii	2.8641	0.007680	0.0013608	96.89	93.60

Table 6.4 Identified linear time invariant (LTI) process models of the effective rainfall-runoff transformation for Bamboi on the Black Volta River (Est. = Estimation, Val. = Validation)

Model [n m d]	**YIC (Est. series)**	**AIC**		**NSE**	
		Est. series	**Val. series**	**Est. series**	**Val. series**
[0 2 0]	-8.5376	6.2675	6.4594	95.24	88.54
[3 1 0]	-5.9225	6.4886	6.4041	93.91	89.20
[2 1 0]	-5.8726	6.7170	6.4373	92.23	88.69
[2 2 0]	-4.6218	6.1723	6.4704	95.60	88.55
[2 3 0]	-4.5306	6.1031	6.4628	95.88	88.71
[1 2 0]	-3.4967	6.2686	6.4653	95.21	88.57

A parallel decomposition of the models is shown in Table 6.6. Clearly the last two models are unacceptable on physical grounds – the [2 2 0] model has a negative b" value (and hence a negative steady state gain) while the positive a" value of the [2 3 0] structure results in a complex valued time constant. The models that best describe the flow process in the catchment and have sound parallel decomposition are, therefore, the [0 2 0] and [1 2 0] models. They decompose as follows:

[0 2 0] $$y_t = b' u_t + b'' u_{t-1} \tag{6.15a}$$

[1 2 0] $$y_t = b' u_t + \frac{b''}{1 + a'' z^{-1}} u_{t-1} \tag{6.15b}$$

Table 6.5 Estimated parameters of the feasible LTI model set for Bamboi on the Black Volta River (estimated standard errors of parameters in brackets)

Model	**Estimated parameters**				
[n m d]	**a_1**	**a_2**	**b_0**	**b_1**	**b_2**
[0 2 0]	-	-	0.3273	0.60054	-
			(0.0184)	(0.0184)	
[1 2 0]	-0.0522	-	0.3381	0.5521	-
	(0.0323)		(0.0195)	(0.0356)	
[2 2 0]	0.1443	-0.1310	0.3301	0.6478	-
	(0.0568)	(0.0326)	(0.0186)	(0.0379)	
[2 3 0]	-0.4601	-0.1213	0.3300	0.4541	-0.3638
	(0.1531)	(0.0339)	(0.0181)	(0.0592)	(0.0838)

Table 6.6 Parallel decomposition of flow for the feasible LTI model set for Bamboi on the Black Volta River

Model	**Runoff Decomposition into Parallel Flows**				
[n m d]	**Instantaneous Flow**	**Fast Flow (Store 1)**		**Slow Flow (Store 2)**	
	b`	**a``**	**b``**	**a```**	**b```**
[0 2 0]	0.3273	-	0.6005	-	-
[1 2 0]	0.3381	-0.0522	0.5698	-	-
[2 2 0]	-	0.4412	-0.6803	-0.2969	1.0104
[2 3 0]	0.3301	0.1874	0.5238	-0.6475	0.0822

For model [0 2 0], Equation 6.15a shows that the total runoff is a sum of fractions of the current effective rainfall and that of the previous month. While the model is attractive because of the relatively low uncertainty in its parameter estimates, its drawback is the absence of the a parameter, i.e., no contribution of subsurface flow to the total catchment runoff is being captured. On the other hand, the [1 2 0] model decomposition of the flow into a fast bypass (within the month) and slow flow through a linear storage (Equation 6.15b) appears more plausible for the basin. The bypass would include flow resulting from rainfall falling directly in the channel, flows from saturated flood plains of the river and those from other saturated areas close enough to the outlet of the catchment as to be available within the month. However, the small value of the a parameter coupled with its rather high uncertainty means that very little baseflow is being accounted for and that this model may well be a [0 2 0] model.

For some of the catchments, only the [0 2 0] model was found to provide a plausible parallel decomposition of the flow and so was selected as the best model for those cases. Table 6.7 shows the estimated parameters and the NSEs for both estimation and validation series obtained from the best models fits for the selected catchments. The plots of the observed and predicted monthly runoff from the LTI models relating effective rainfall to runoff (without a noise term) for the catchments are shown in figures 6.3 – 6.12 for both estimation and validation series. Both table and plots show the very good fits of these process models to the observed data. However, in a few cases such as Dapola (Figure 6.3), Nabogo (Figure 6.8) and Mango (Figure 6.10), some of the peak flows have not been well predicted. The plots also show that for the case of the Black Volta River in particular, a part of the recession flow has not been fully captured in the selected models. These discrepancies might be a result of tradeoffs from choosing models that explain the data fairly well and also provide some physical interpretation of the results (grey box models) over those that fit the data extremely well but offer no plausible interpretation of the flow process (black box models). Also, the reliance of the modelling framework on effective rainfall means flows during periods of no rainfall (recession flows) are likely not to be completely represented (Mwakalila, ***et. al.***, 2001).

Table 6.7 Estimated parameters of the best LTI model for selected catchments (estimated standard errors of parameters in brackets; Est. = Estimation Series, Val. = Validation Series)

Gauging Station	**Estimated Parameters**			**NSE (%)**	
	a_1	**b_0**	**b_1**	**Est.**	**Val.**
Lawra	-0.2442(0.0340)	0.3962(0.0191)	0.3289(0.0354)	93.09	84.46
Dapola	-0.2791(0.0225)	0.4214(0.0127)	0.2819(0.0242)	93.45	65.36
Bui	-	0.4067(0.0203)	0.5478(0.0203)	96.26	90.94
Bamboi	-0.0522(0.0323)	0.3381(0.0195)	0.5521(0.0356)	95.22	88.57
Yagaba	-0.1614(0.0757)	0.7918(0.0180)	0.0773(0.0759)	97.02	78.04
Nabogo	-	0.7857(0.0243)	0.2487(0.0243)	95.49	83.73
Porga	-0.0259(0.0237)	0.4057(0.0143)	0.5786(0.0256)	95.89	82.26
Mango	-	0.5247(0.0149)	0.5192(0.0149)	96.61	89.81
Koumangou	-0.0838(0.0535)	0.6654(0.0185)	0.2667(0.0525)	96.78	86.58
Sabari	-	0.6265(0.0195)	0.4107(0.0195)	97.20	91.14

Table 6.8 lists the steady state gains for the best model for all the selected catchments. The table shows that, in general, the fast flow component is smaller than the slow one for the large catchments (in the Black Volta Basin) but is larger for the smaller catchments (White Volta and Oti Basins), Porga on the Oti River being an exception. As can be seen for the model [1 2 0] in the table, the time constants are very small, implying no significant baseflow contribution to streamflow.

No significant improvement in model fit was observed with the inclusion of the ETP_t series for the catchments. This may be because the use of the catchment runoff as a surrogate for the catchment wetness index in the estimation of effective rainfall was adequate to also account for the seasonal evaporative effects. Therefore, all model parameters were estimated with the runoff and the primary input, effective rainfall, series only.

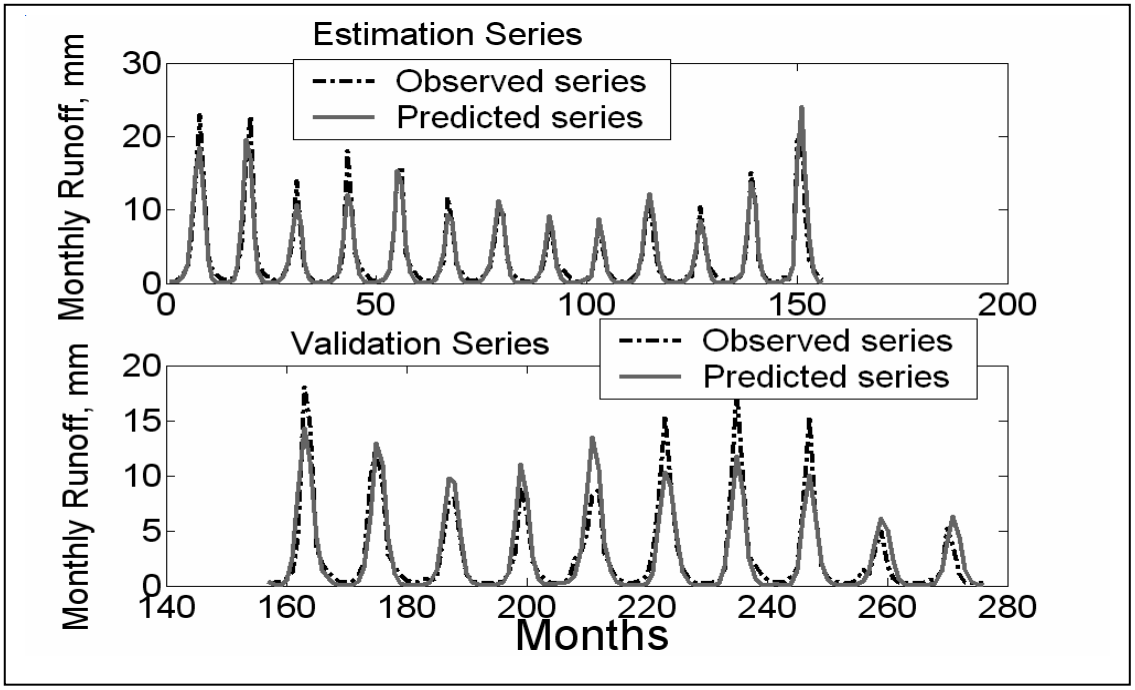


Figure 6.3 Observed and predicted monthly runoff for Lawra, Black Volta River LTI model [n m d] = [1 2 0]

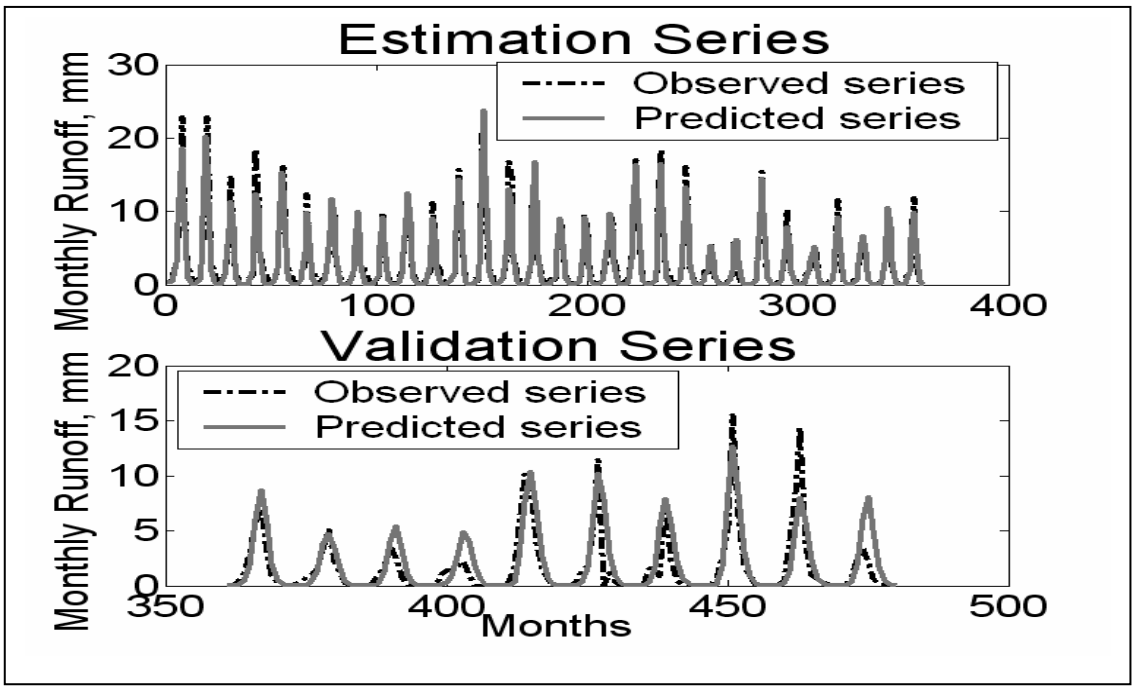


Figure 6.4 Observed and predicted monthly runoff for Dapola, Black Volta River LTI model [n m d] = [1 2 0]

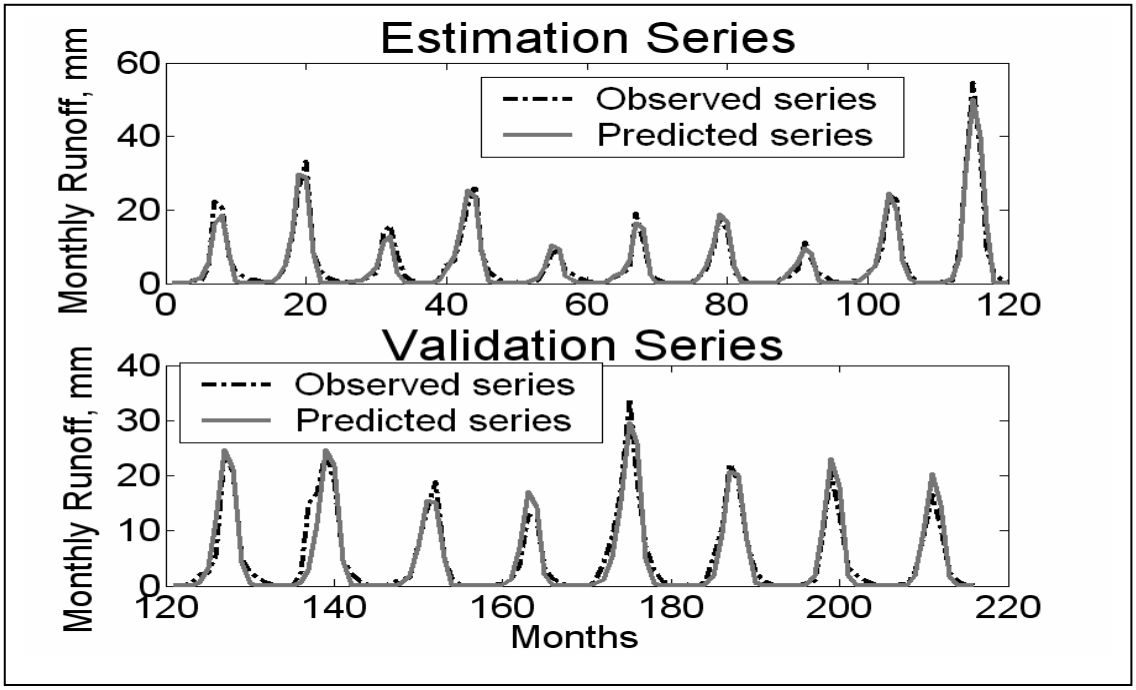


Figure 6.5 Observed and predicted monthly runoff for Bui, Black Volta River LTI model [n m d] = [0 2 0]

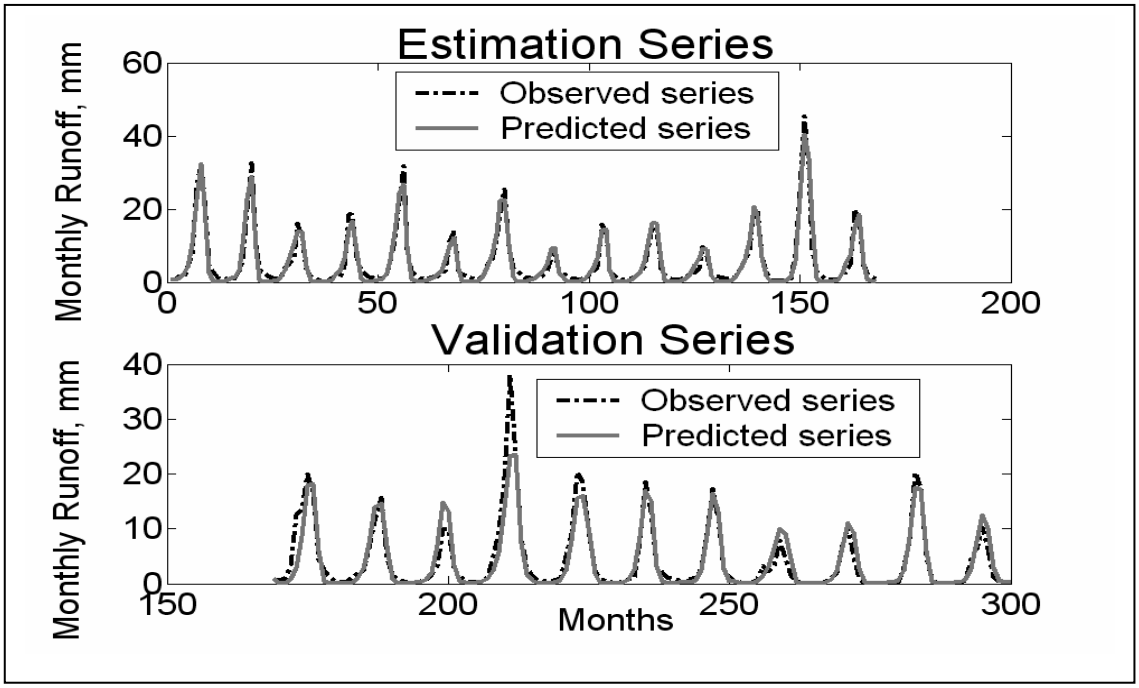


Figure 6.6 Observed and predicted monthly runoff for Bamboi, Black Volta River LTI model [n m d] = [1 2 0]

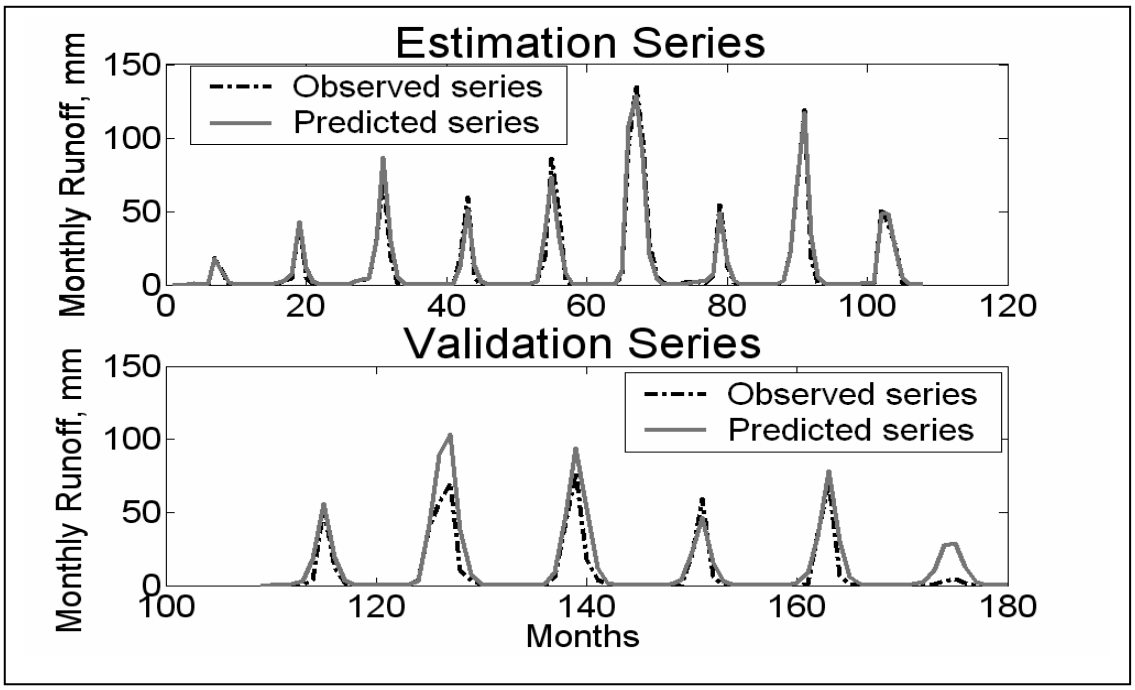


Figure 6.7 Observed and predicted monthly runoff for Yagaba, Kulpawn River (White Volta Basin) LTI model [n m d] = [1 2 0]

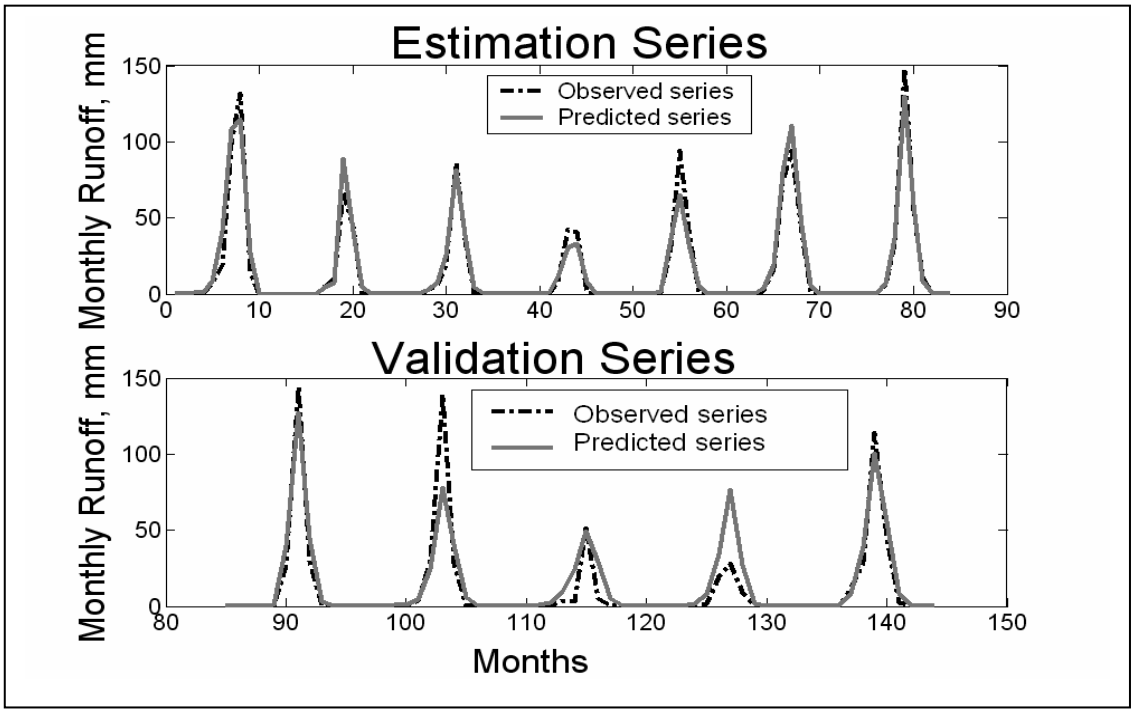


Figure 6.8 Observed and predicted monthly runoff for Nabogo, Nabogo River (White Volta Basin) LTI model [n m d] = [0 2 0]

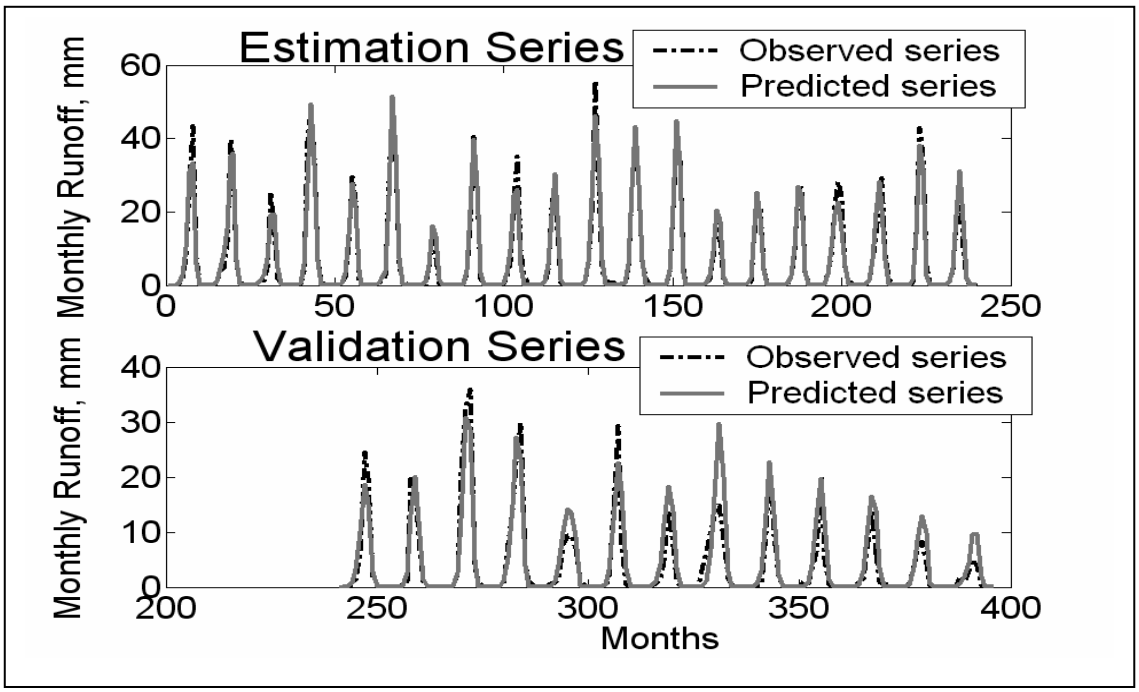


Figure 6.9 Observed and predicted monthly runoff for Porga, Oti River
LTI model [n m d] = [1 2 0]

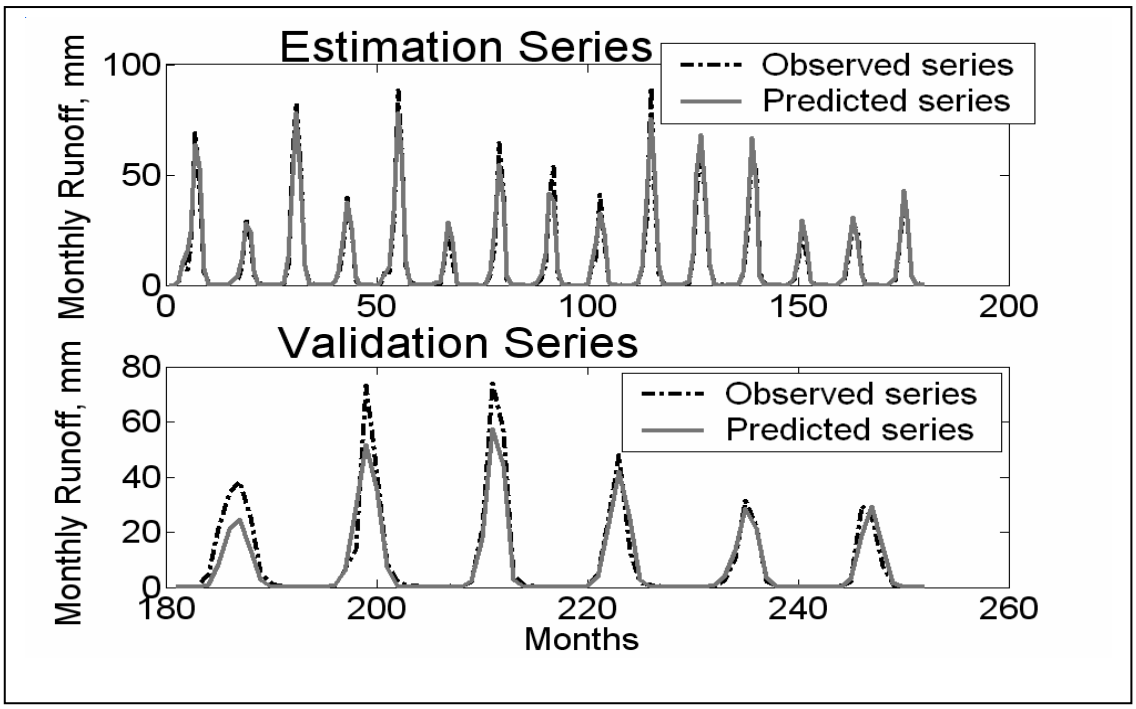


Figure 6.10 Observed and predicted monthly runoff for Mango, Oti River
LTI model [n m d] = [0 2 0]

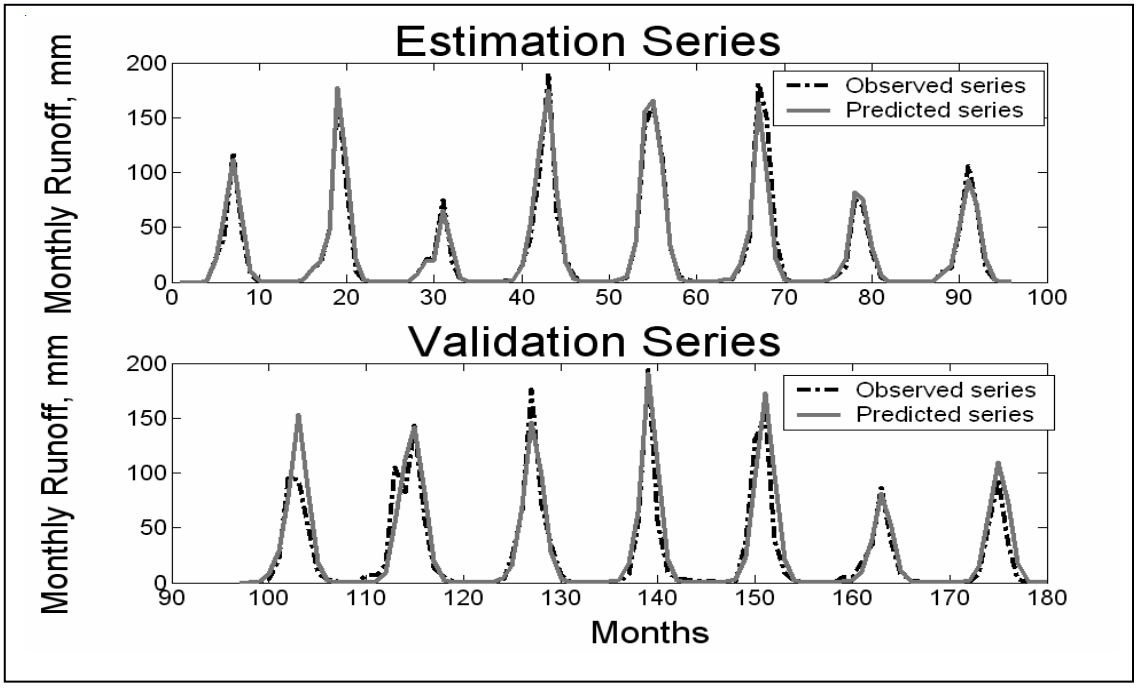


Figure 6.11 Observed and predicted monthly runoff for Koumangou, Koumangou River (Oti Basin) LTI model [n m d] = [1 2 0]

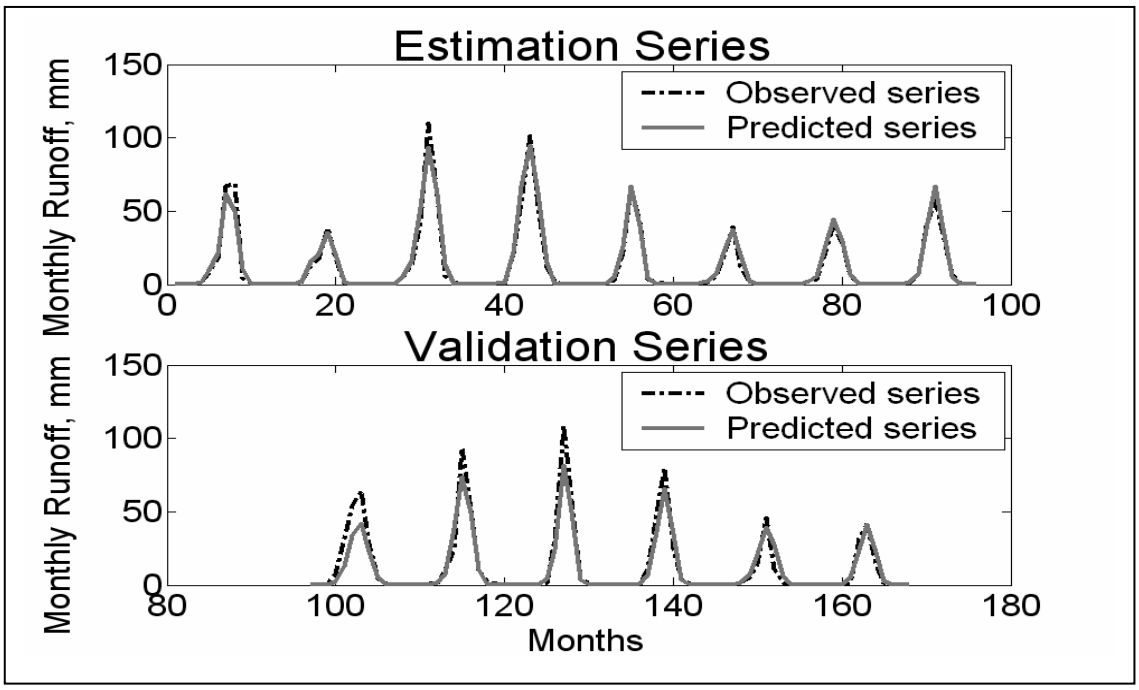


Figure 6.12 Observed and predicted monthly runoff for Sabari, Oti River LTI model [n m d] = [0 2 0]

Table 6.8 Steady state gains (SSG), percent of total monthly flow (% flow) and time constants (TC) of parallel flow components for the best model for each of the selected catchments.

Gauging Station	**Best Model**	**Flow Type**				
		Fast		**Slow**		
		SSG	**% Flow**	**SSG**	**% Flow**	**TC (days)**
Lawra	[1 2 0]	0.3962	41.30	0.5631	58.70	21.28
Dapola	[1 2 0]	0.4214	43.20	0.5541	56.80	23.51
Bui	[0 2 0]	0.4067	42.61	0.5478	57.39	-
Bamboi	[1 2 0]	0.3381	36.00	0.6531	64.00	10.16
Yagaba	[1 2 0]	0.7918	76.40	0.2446	23.60	16.45
Nabogo	[0 2 0]	0.7857	75.96	0.2457	24.04	-
Porga	[1 2 0]	0.4057	40.15	0.6048	59.85	8.21
Mango	[0 2 0]	0.5249	50.26	0.5192	49.74	-
Koumangou	[1 2 0]	0.6654	65.41	0.3519	34.59	12.10
Sabari	[0 2 0]	0.6265	60.40	0.4107	39.60	-

6.7 Conclusions and recommendations

A hybrid metric-conceptual modelling framework was successfully developed and applied to monthly runoff and rainfall series for selected river catchments in the Volta Basin of West Africa. Monthly runoff was found to be a suitable surrogate for the basin wetness index in defining a monthly rainfall filter for the estimation of effective rainfall. For simulation without observed runoff, the IHACRES-type definition of catchment wetness index for generating effective rainfall was found to be adequate for the catchments studied.

Results of the study show that monthly catchment runoff in the basin can be decomposed into parallel flow pathways, and the fitted models represented the river flows very well. However, the model form and the information content of the observed data appeared not to have fully captured the over-month recession flow for the case of the gauging stations on the Black Volta River where plots of the time series of the observed runoff seem to indicate that such recession flow exists. On the other hand, it is very clear from the results of this study and the observed runoff series for the

catchments investigated, that over-year (deep groundwater) flow is negligible in the basin (the flow drops to zero at the end of each year). This supports the results of the autocorrelation function analyses undertaken for the annual runoff series described in Chapter 3 in which no significant autocorrelation at any lag was found for the stations. This means that at the annual scale, current runoff is independent of any past values.

On the whole, this modelling framework is adequate for monthly river flow prediction in the basin given the limitations in the input-output data available in the basin in terms of quality, quantity and diversity. It is recommended that the framework be applied to daily rainfall, runoff and ETP series to establish it as the main framework for riverflow modelling at the two most important temporal scales in the basin.

7 SUMMARY AND RESEARCH FINDINGS

7.1 Introduction

The 400,000 km^2 Volta Basin is an international basin covering almost 28% of Africa's West Coast. It extends from longitude 5^o 30 W to 2^o 00E and from latitude 5 30^o N to 14^o 30 N and is shared by the six West African countries Benin, Togo, Ghana, La Cote d'Ivoire, Burkina Faso and Mali. There is great pressure on the land and water resources of the basin, due mainly to the poor economies of the riparian countries, rapid population growth rates, erratic and uncertain rainfall and increasing demands for water for agriculture and hydropower production. The geology of the basin does not favour massive groundwater storage and flow, so that this component of the resources has not contributed appreciably to water use in the basin except for rural-community water supplies. Water resources development in the basin for agriculture, industry and hydropower production has thus relied on surface water resources.

No institutional or legal framework currently exists for co-operation between the countries in the basin in the use of the water resources. Each country thus undertakes the mobilisation of its water requirements without consideration of the other users in the basin. The rather haphazard and indiscriminate harvesting and use of the resource have resulted in both upstream and downstream problems and are a potential source of conflicts both at local (within country) and international levels. Governments and water resources managers in the basin, at least at the national level, have recognised this potential conflict and the lack of any international mechanisms for consultation and co-operation between the riparian countries in the use of the water and other natural resources of the basin. They have identified the management problems needing priority attention, advocated an integrated land and water resources management and pledged transboundary co-operation to ensure environmental integrity and sustainable water use.

Concerned external actors have noted this willingness of the states in the basin to co-operate and are lending their support in the form of expertise and funds to ensure the maintenance of environmental integrity of the basin. The intervention of agencies and organisations such as GEF, UNEP, GCI and CGIAR seeks to develop the necessary human and institutional capacity within the riparian countries, develop appropriate legal framework and provide the necessary funding for a proper management of the basin's

water and other natural resources. External research institutions such as ZEF (Bonn University) are intervening in the scientific research area towards the development of both suitable scientifically-based decision support systems and human capacity in the basin to facilitate the water and other natural resources management tasks of the people and institutions of the riparian states. It is in furtherance of the objectives of the latter intervention that this thesis has been produced. It is contributing to the hydrological information base of the basin, and the results obtained in this study will hopefully form one of the several strands of the envisaged decision support system.

7.2 Exploratory data analysis

Exploratory analysis of the monthly catchment rainfall, runoff and potential evapotranspiration (ETP) as well as the annual series from the aggregation of the monthly series shows that:

- There is much more variation in monthly runoff than in the corresponding rainfall and ETP, suggesting a non-linear relationship between runoff and the two inputs.
- Except for 3 or 4 months of the year, mean monthly ETP far exceeds mean monthly rainfall for all months. This implies that groundwater loss in the dry season is mainly to evapotranspiration rather than to streamflow as baseflow.
- Between 70 and 95% of mean monthly runoff in much of the basin occurs in 4 months of the year. This very skewed temporal distribution of runoff and rainfall means, therefore, that run-on-the-river water-use systems require hydraulic structures to ensure all year round water availability.
- There is persistence in catchment monthly runoff. However, annual runoffs have no memory.
- The rainfall-runoff process is largely non-linear. While good linear models can be obtained for some of the catchments if large numbers of parameters are allowed, it is only non-linear modelling that can bring out the real character of the flow process.

7.3 Filling gaps in stream flow data

A large number of gaps exist in the available stream flow data in the basin, particularly in the daily flow series. The study has produced a novel method for filling in short gaps in daily streamflow series using a combination of spatio-temporal state-space modelling with Kalman smoothing and the Expectation-Maximization (EM) algorithm. The very good results obtained demonstrate both the ability of the method to adequately predict missing daily riverflow series and its power and utility as a tool in hydrological modelling in general.

7.4 Modelling streamflow using NARMAX polynomial models

NARMAX polynomial models have been shown to be suitable for non-linear modelling of input-output systems including environmental systems. In this study, the modelling framework produced very good results with much fewer parameters than in the case of the linear modelling. This indicates that much of the nonlinearity in the rainfall-runoff relationship had been accounted for in this modelling framework. It confirms the observation made in the exploratory data analysis phase of this study that the rainfall-runoff process is better modelled as a non-linear process. The main drawback of NARMAX models is their purely black-box nature.

7.5 Data-based mechanistic modelling of streamflow

The DBM framework takes advantage of the ability of the purely black-box models, such as NARMAX models, to adequately represent the rainfall-flow non-linear process parsimoniously. In addition, it seeks to select models from the feasible model set that also provide some insight into the flow mechanism. The very good results obtained from the application of this modelling framework to rainfall-runoff series establish it as the framework of choice for runoff modelling in the basin, particularly in the instances, such as here, where the available input-output data are limited in quantity, quality and diversity. Its main strength over the NARMAX formulations is its ability to produce results that depict parallel processes in the basin as has been found to be the case for hydrological processes in general.

Results from this modelling framework also show that the non-linear relationship between the catchment wetness index and effective rainfall, in which the

catchment drying time constant is an exponentially weighted function of potential evapotranspiration, is suitable for the estimation of effective rainfall from rainfall and potential evapotranspiration.

7.6 Recommendations for further research

The following issues are recommended for further study:

- The application of the spatio-temporal state-space modelling framework for missing flow series infilling should be extended beyond data infilling to areas such as adaptive flood forecasting in the Volta Basin. In such an application, forecasts of several hours ahead of water levels at a gauging station are made from real time measurements of water levels at several upstream gauging stations and previous water level measurements at the station of interest.
- The DBM framework should be applied to daily catchment rainfall, runoff and ETP (or temperature) in order to confirm that it is also suitable for runoff prediction at this temporal scale. Successful application of the framework at this scale would establish it as the model to use, when it is ascertained that it is not feasible to use distributed models in a given instance because of lack of requisite input-output data.
- Further research is needed to generalise the parameters B, f and g in the non-linear catchment wetness index-effective rainfall relationship of Equation 6.14 for a sub-basin such as Upper Black Volta or Middle White Volta, in order to be able to simulate runoff for ungaged river catchments. Results from the study have shown the ability of the model in 6.14 for runoff prediction in the catchments used in the study. Since this model does not use observed runoff as a surrogate for catchment wetness in the estimation of effective rainfall, it is very suitable for prediction in ungaged catchments.
- Results from the DBM modelling framework should be compared with those from distributed models such as WaSiM-ETH in order to determined areas in which these models complement each other. This would be appropriate for both daily and monthly runoff modelling in the basin and elsewhere and could be very useful in the calibration of such distributed models.

- Rainfall-runoff and hydrological modelling in general require very good and diverse observed input-output data, such as rainfall, temperature, soil moisture and runoff. Therefore, it is recommended that a few river catchments in the basin be selected and monitored for these input-output data, especially rainfall, runoff and temperature (for ETP computations). Continuous monitoring systems (e.g., weather stations, water level measuring divers) should be installed in the selected catchments to provide at least hourly measurements of the variables of interest. This would provide very good quality data with some diversity for testing any developed models for the basin thoroughly and adequately.

8 REFERENCES

Abbott M. B., Bathurst J. C., Cunge J. A., O'Connell P. E., Rasmussen J. (1986a). An introduction to the European Hydrologic System-Systeme Hydrologique Europeen, SHE, 1: History and philosophy of a physically based, distributed modelling system. J. Hydrol. 87: 45–59.

Abbott M. B., Bathurst J. C., Cunge J. A., O'Connell P. E., Rasmussen J. (1986b). An introduction to the European Hydrologic System - Systeme Hydrologique Europeen, SHE, 2: Structure of a physically based, distributed modelling system. J. Hydrol. 87: 61–77.

Akaike H. (1974). A new look at the statistical model identification. IEEE Trans. Autom. Control, AC-19: 716-723.

Anderson O. (1976). Time series analysis and forecasting: the Box-Jenkins approach. London, Butterworths. 182 pp.

Andreini M., van de Giesen N. C., van Edig A., Fosu M, Andah W. (2000). Volta Basin Water Valance; ZEF-Discussion Papers on Development Policy, Number 21. Bonn, Germany. Center for Development Research. 89 pp.

Bathurst J. C., Wicks J. M., O'Connell P. E. (1995). The SHE/SHESED basin scale water flow and sediment transport modelling system. In: V. P. Singh (ed.) Computer models of watershed hydrology, Water Resources Publications, Littleton, Colo. Chapter 16: 563–594.

Beven K. J. (2000). Rainfall-Runoff Modelling: A Primer. John Wiley and Sons Ltd., Baffins Lane, Chichester, England. 360 pp.

Beven K. J. (2001). Calibration, Validation and Equifinality in Hydrological Modelling: A Continuing Discussion. In: Anderson, M. G. and Bates, P. D. (eds), Model Validation - Perspectives in hydrological science, John Wiley and Sons, Chichester, Chapter 4.

Beven K. J. and Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 249: 11-29.

Beven K. J. and Kirkby M. (1979). A physically based, variable contributing area model of basin hydrology, Hydrol. Sci. Bull. 24(1): 43–69.

Billings S. A. and I. J. Leontaritis (1982). Parameter estimation techniques for nonlinear systems. 6th IFAC Symposium on System Identification and Parameter Estimation, Washington D.C., USA, pp. 505-510.

Billings S. A. and G. N. Jones (1992). Orthogonal least-squares parameter estimation algorithms for non-linear stochastic systems. International Journal of Systems Science 23(7): 1019-1032.

Bilmes J. A. (1998). A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Available at ftp://ftp.icsi.berkeley.edu/pub/techreports/1997/tr-97-021.pdf

Box G. E. P., Jenkins G.M., Reinsel G. C. (1994). Time Series Analysis: Forecasting and Control. 3rd Edition, Prentice Hall, Englewood Cliffs, New Jersey 07632

Bryson A. E. and Ho Y. C. (1969). Applied Optimal Control, Optimization, Estimation and Control, Waltham. Blaisdell Publishing Company.

Burnash R.J.C., Ferral R. L., McGuire R. A. (1973). A Generalized Streamflow Simulation System - Conceptual Modelling for Digital Computers. U.S. Department of Commerce, National Weather Service and State of California, Dept. of Water Resources.

Burnash R. J. C. (1995). The NWS River Forecast System - catchment modelling. In: Singh V. P. (ed), Computer Models of Watershed Hydrology, pp. 311-366.

Calver A. and Wood W. L. (1995). The Institute of Hydrology distributed model. In: V. P. Singh, (ed), Computer models of watershed hydrology, Water Resources Publications, Littleton, Colo. Chapter 17.

Chen S. and S. A. Billings (1989). Representations of non-linear systems: the NARMAX model, Int. J. of Control. 49(3): 1013-1032.

Chen S., Billings S. A., Luo W. (1989). Orthogonal least squares methods and their application to nonlinear system identification. Int. J. of Control. 50(5): 1873-1896.

Chiras N., Evans C. and Rees D. (2000). Nonlinear gas turbine computer modelling using NARMAX structures. Proceedings IEEE Instrumentation and Measurement Technology Conference IMTC2000, Baltimore. 3: 1278-1284.

Coe M. and A. T. Grove (1998). Lake Chad has shrunk into "a ghost". Available at http://www.afrol.com/Categories/Environment/env062_lake_chad.htm

Comp.ai.neural-nets FAQ (2004). What are cross-validation and bootstrapping? Available at http://www.faqs.org/faqs/ai-faq/neural-nets/part3/section-12.html

Crawford N. H. and Linsley R. K. (1966). Digital simulation in hydrology: Stanford Watershed Model IV. Tech. Rep. No. 39, Stanford Univ., Palo Alto, Calif.

Dapaah-Siakwan S., and Gyau-Boakye P. (2000). Hydrogeologic framework and borehole yields in Ghana. Hydrogeology Journal. 8: 405 – 416.

Dawdy D. R., and O'Donnell T. (1965). Mathematical models of catchment behavior. J. Hydraul. Div., Am. Soc. Civ. Eng. 91(HY4): 123–127.

Dempster A. P., Laird N. M., Rubin D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion). J. Roy. Stat. Soc. B. 39: 1-38.

Devitt, T. (2001). Africa's Lake Chad Disappearing, Thanks to Irrigation. UniSci Daily University Science News. Available at http://unisci.com/stories/20011/0301011.htm.

Digalakis. V., Rohlicek, J. R. and Ostendorf, M. (1993). ML Estimation of a Stochastic Linear System with the EM Algorithm and its Application to Speech Recognition. IEEE Transactions on Speech and Audio Processing 1(4): 431-442.

Dooge J. C. I. (2003). Linear Theory of Hydrologic Systems. EGU Reprint Series. European Geosciences Union, Katlenburg-Lindau, Germany. 327 pp.

FAO (1997). Irrigation potential in Africa: A basin approach. FAO Land and Water Bulletin 4. Available at http://www.fao.org/documents/show_cdr.asp?url_file=/docrep/W4347E/w4347e0u.htm.

FoEME (Friends of the Earth, Middle East). (1996). Dead Sea Challenges. Available at http://www.foeme.org/publications.htm.

FoEME (Friends of the Earth, Middle East). (1998). One Basin, One Strategy. Final Report of the Symposium on Promoting an Integrated Sustainable Regional Development Plan for the Dead Sea Bain. Middle East Environmental NGO Forum. Available at http://www.foeme.org/docs/dsfinrep.htm.

Franks S., Beven K. J., Quinn P. F., Wright, I. (1997). On the sensitivity of soil-vegetation-atmosphere transfer (SVAT) schemes: equifinality and the problem of robust calibration. Agricultural and Forest Meteorology 86: 63-75.

Freeze R. A. and Harlan R. L. (1969). Blueprint for a physically based, digitally simulated hydrologic response model. J. Hydrol. 9: 237–258.

Glanz M. H. (1998). Creeping Environmental problems in the Aral Sea Basin. In: Kobori, I. and M. H. Glanz (eds.) Central Eurasian water crisis: Caspian, Aral and Dead seas. United Nations University press. USA. Available at http://www.unu.edu/unupress/unupboks/uu18ce/uu18ce04.htm.

GCI (2003). Water for peace in the Volta River Basin. March 2003, Draft report Green Gross Internal. Available at http://www.greencrossinternational.net/ GreenCrossPrograms/WATERRES/wwf_03/gci_volta.pdf.

GEF (2003a). Addressing Transboundary Concerns in the Volta River Basin and its Downstream Coastal Area. Global Environment Facility, UN. Available at http://www.gefweb.org/Documents/Council_Documents/GEF_C21/IW_-_Regional_-_Volta_River_Basin_-_Project_Document.pdf.

GEF (2003b). GEF council work program submission - project executive summary – Volta River Basin. Global Environment Facility, UN. Available at http://www.gefweb.org/Documents/Council_Documents/GEF_C21/IW_-_Regional_-_Volta_River_Basin_-_Executive_Summary.pdf.

Ghahramani Z. and Hinton G. E. (1996). Parameter Estimation for Linear Dynamical Systems. Technical Report CRG-TR-96-2, Department of Computer Science, University of Toronto, Canada.

GhanaWeb (2004). 18-million dollar Programme on Volta Basin launched. General News of Wednesday, 19 May 2004. Available at http://www.waterforfood.org/newsroom/ volta/ghanaweb20040519.pdf.

Goodall C. and Mardia K. V. (1994). Challenges in Multivariate Saptial modelling. Proceedings of the XVIIth International Biometric Conference, Hamilton, Ontario, Canada, 8-12.

Grayson R. B., Bloschl G., Moore I. D. (1995). Distributed parameter hydrologic modelling using vector elevation data: THALES and TAPEC-C. In: V. P. Singh (ed.), Computer models of watershed hydrology, Water Resources Publications, Littleton, Colo. Chapter 19.

Guttorp P. and Sampson P. D. (1994). Methods for estimating heterogeneous spatial covariance functions with environmental applications. Environmental Statistics In: Patil G.P. and Rao C.R. (eds), Handbook of Statistics. 12, 661-689. North Holland, Amsterdam, New York.

Gyau-Boakye P. and Schultz G. A. (1994). Filling gaps in runoff time series in West Africa. Hydro.Sci.J. 39(6): 621-636.

Hansen D. P., Ye W., Jakeman A. J., Cooke R. and Sharma P. (1996). Analysis of the effect of rainfall and streamflow data quality and catchment dynamics on streamflow prediction using the rainfall-runoff model IHACRES. Environmental Software 11: 193-202.

Haslett J. (1989). Space time modelling in meteorology – a review. Bulletin of the International Statistical Institute 51: 229-246.

Haykin S. (2001). Kalman Filters. In: Haykin S: Kalman Filtering and Neural Networks. John Wiley & Sons, Inc, New York.

Heunecke G. and Welsch W. (2004). Terminology and Classification of Deformation Models in Engineering Surveys. J. Geospatial Engineering 2 (1): 35-44.

INCOSE (2004). A Consensus of the INCOSE Fellows. The International Council on Systems Engineering. Available at http://www.incose.org/ practice/fellowsconsensus.aspx.

Jakeman A. J., Littlewood I. G. and Whitehead P. G. (1990). Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. J. Hydrol. 117: 275-300.

Jakeman A. J. and G. M. Hornberger. (1993). How Much Complexity is warranted in a Rainfall-Runoff Model? Water Resour. Res. 29: 2637-2649.

Jakeman A. J., Littlewood I. G., Whitehead P. G. (1993). An assessment of the dynamic response characteristics of streamflow in the Balquhidder catchments. J. Hydrol. 145: 337-355.

Jakeman A. J., Chen T. H., Post D. A., Hornberger G. M., Littlewood I. G, Whitehead P. G. (1993). Assessing uncertainties in hydrological response to climate at large scale. In: Wilkinson, W. B. (ed.) Macroscale Modelling of the Hydrosphere. IAHS, Wallingford, UK, pp. 37-47.

Kashyap R.L. (1977). A bayesian comparison of different classes of dynamical models using empirical data. IEEE Trans. Automatic Control. 22(5): 715-727.

Kohavi R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Proc. of the 14th Int. Joint Conf. On A.I. 2. Canada.

Kokkonen T. S., and Jakeman A. J. (2001). A comparison of metric and conceptual approaches in rainfall-runoff modelling and its implications. Water Resour. Res. 38(9): 2345-2352.

Korenberg M.J., Billings S. A., Liu Y. P., Mollroy P. J. (1988). Orthogonal parameter estimation algorithm for nonlinear stochastic systems. Int. J. of Control. 48(1): 193-210.

Kottegoda N. T. and Elgy J. (1979). Infilling missing flow data. In: More-Seytoux H. J. et al. (eds), Modelling Hydrologic Processes. Proc. Fort Collins 3rd Int. Hydrol. Symp. On Theoretical and Applied Hydrology, Colorado State University, Fort Collins, Colorado, USA, 27-29 July, 1977, pp. 60-73.

Lendasse A., Wertz V., Verleysen M. (2003). Model Selection with Cross-Validation and Bootstraps – Application to Time Series Prediction with RBFN Models. In: O. Kaynak et. al. (eds.): ICANN/ICONIP 2003, LNCS 2714: 573–580.

Leontaritis I. J. and S. A. Billings (1985). Input-output parametric models for nonlinear systems. Part II: stochastic non-linear systems. Int. J. Control. 41(2): 329-344.

Little R. J. A. and Rubin D. B. (2003). Statistical Analysis with Missing Data. 2nd ed. Wiley, New York.

Littlewood I. G., and Jakeman A. J. (1994). A new method of rainfall-runoff modelling and its applications in catchment hydrology. In: P. Zannetti (ed), Environmental Modelling. II. Computational Mechanics Publications, Southampton, UK, pp. 143–171.

Littlewood I. G. and Marsh T. J. (1996). Re-assessment of the monthly naturalised flow record for the River Thames at Kingston since 1883, and the implications for the relative severity of historical droughts. Regulated Rivers: Research and Management 12: 13-26.

Ljung L. (1999). System Identification: Theory for the User. 2nd ed. Prentice-Hall, Upper Saddle River, New Jersey.

Ljung L. (2003). System Identification Toolbox for use with MATLAB, User's Guide. The MathWorks, 3 Apple Hill Drive, Natick, MA 01760-2098.

Mardia K., Goodall C., Redfern E., Alonso F. (1998). The kriged kalman filter (with discussion). Test. 7: 217-285.

Mashor M. Y. and Hamid H. A. (1998). Structure Detection of Hybrid-RBF Network Using Orthogonal Least Squares Algorithm. Available at http://www.journal. au. du ijcim/sep98/article1.html.

Micklin P. P. and Williams W. D. (1996). THE ARAL SEA BASIN. NATO ASI Series. 2- Environment Vol. 12. Springe-Verleg Berlin Heidelberg. Germany.

Moore R. J. and Clarke R. T., (1981). A distribution function approach to rainfall-runoff modelling. Water Resour. Res. 17:1367-1382.

Mwakalila S., Campling P., Feyen J., Wyseure G., Beven, K. (2001). Application of a data-based mechanistic modelling (DBM) approach for predicting runoff generation in semi-arid regions. Hydrolog. Process. 15(12): 2281-2295.

MWH (1998(1)). Ghana Water Resources Management study – International Waters "Building block" study. Main report (final). Ministry of Works and Housing, Accra, Ghana.

MWH (1998(2)). Ghana Water Resources Management study – Information "Building Block" study. Part III. Executive Eummary. Vol. 1: Summary of Information in the three basin systems (final report). Ministry of Works and Housing, Accra, Ghana.

Nash J. E. and Sutcliffe J. V. (1970). River Flow Forecasting through conceptual models. Part I – A discussion of principles. J. Hydrol. 10: 282-290.

O'Connell P. E. (1991). A historical perspective. In: Bowles D. S. and O'Connell P. E. (eds), Recent Advances in the Modelling of Hydrologic Systems, Kluwer. Dordrecht, pp. 3–30.

Papadakis I., Napiorkowski J.,Schultz G. A. (1993). Monthly runoff generation by non-linear model using multispectral and multitemporal satellite imagery. Adv. Space Res. 13(5): 181-186.

Post D. and Jakeman A. J. (1996). Relationships between catchment attributes and hydrological response characteristics in small Australian mountain ash catchments. Hydrolog. Process. 10: 877-892.

Post D. A., Jakeman A. J., Littlewood I. G., Whitehead P. G., Jayasuriya M. D. A. (1996). Modelling land cover induced variations in hydrologic response: Picaninny Creek, Victoria. Ecological Modelling 86: 177-182.

Riberiro M. I. (2004). Kalman and Extended Kalman Filters: Concept, Derivation and Properties. Available at http://omni.isr.ist.utl.pt/~mir/pub/kalman.pdf.

Rouhani S. and Myers D. E. (1990). Problems in space-time Kriging of geohydrological data. Math. Geology. 22: 611-623.

Salas J. D., Delleur J. W., Yevjecich V. M., Lane W. L. (1980). Applied modelling of hydrologic time series. Littleton, Colrado, Water Resources Publlications. 484 pp.

Schafer J. L. (1997). Analysis of Incomplete Multivariate Data. New York, CRC Press.

Schulla Jörg (2001). Water balance Simulation Model ETH (WaSiM-ETH). User Manual.

Sefton C. E. M., Whitehead P. G., Eatherall A., Littlewood I. G., Jakeman A. J. (1995). Dynamic response characteristics of the Plynlimon catchments and preliminary analysis of relationships to physical descriptors. Environmetrics. 6: 465-472.

Shumway R. H. and Stoffer D. S. (1982). An Approach to Time series Smoothing and Forecasting Using The EM Algortihm. J. Time Series Analysis 3(4): 253-264.

Singh V. J. and Woolhiser D. A. (2002). Mathematical Modelling of Watershed Hydrology. J. Hydrolog. Engineering 7(4): 270-292.

Sivapalan M., Ruprecht J. K., Viney N. R. (1996a). Water and salt balance modelling to predict the effects of land use changes in forested catchments: 1. Small catchment water balance model. Hydrolog. Process. 10: 393–411.

Sivapalan M., Viney N. R., Ruprecht J. K. (1996b). Water and salt balance modelling to predict the effects of land use changes in forested catchments: 2. Coupled model of water and salt balances. Hydrolog. Process. 10: 413–428.

Sivapalan M., Viney N. R., Jeevaraj C. G. (1996c). Water and salt balance modelling to predict the effects of land use changes in forested catchments: 3. The large catchment model. Hydrolog. Process. 10: 429–446.

Smith, J.O. (2004). Introduction to Digital Filters, May 2004 Draft. Available at http://ccrma.stanford.edu/~jos/filters/.

Stenman A. (2002). On Model Structure Selection for Nonparametric Prediction Methods. Available at http://www.control.isy.liu.se.

Tabrizi M. H. N., Said S. E., Badr A. W., Mashor Y., Billings S. A. (1998). Nonlinear Modelling and Prediction of a River Flow System. International Journal of AWRA. 34(8):1333-1339.

Taylor Joie (2003). Watershed management in the Volta River Basin, West Africa: Providing information through hydrological modelling. M.Sc. Thesis, Cornell University.

UN (2005). World Population Prospects: The 2004 Revision. Available at http://www.un.org/esa/population/unpop.htm.

van der Sommen J. J. and Geirnaert W. (1988). On the Continuity of Aquifer systems on the Crystalline Basement of Burkina Faso. In: Simmers I. (ed), Estimation of Natural Groundwater Recharge. D. Reidel Publishing Company, pp. 29 – 45.

VBRP (2002). Volta Basin Research Project Water Resources Management GIS Maps on the Volta Basin. Accra, Ghana.

Water For Food (2003). More demand, less water: Coping with West Africa's growing crisis. Volta River Basin: Basin Overview. Available at http://www.waterforfood.org /pdf/volta_br.pdf.

Wheater H. S., Jakeman A. J., Beven K. J. (1993). Progress and directions in rainfall-runoff modelling. In: Jakeman A. J., Beck M. B., McAleer M. J. (eds), Modelling Change in Environmental Systems, John Wiley & Sons. 101-132.

Wicox, K. E. (2002). Modelling Engineering Systems. Available at http://web.mit.edu/aa-math/www/modules/node1.html.

World Bank (2004). World Development Report 2004 : Making Services Work For Poor People. Selected World Development Indicators 2004. Available at http://econ.worldbank.org/wdr/wdr2004/text-30023/

WRI (2003). Watersheds of Africa Water Resources eAtlas Land Cover and Use Variables: A19 Volta. World Resources Institute, Available at http://www.earthtrends. wri.org/pdf_library/maps/watersheds/af26.pdf.

Xu, K. and Wikle, C. K. (2004). Estimation of Parameterized Spatio-Temporal Dynamic models. Available at http://www.stat.missouri.edu/~ wikle/xu_wikle_072104.pdf.

Ye W., Bates B. C, Viney N. R., Sivapalan M., Jakeman A. J. (1997). Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments. Water Resour. Res. 33:153-166.

Ye S., Jakeman A. J.,Young P. C. (1998). Identification of improved rainfall-runoff models for an ephemeral low yielding Australian catchment. Environmental modelling and Software 13: 59-74.

Young P. C. (1984). Recursive Estimation and Time-Series Analysis, Springer-Verlag, Berlin.

Young P. C. (1988). Recursive extrapolation, interpolation and smoothing of non-stationary time series. In: Chen C. F. (eds), Identification and System Parameter Estimation, Pergamon Press, Oxford, pp. 33-44.

Young P. C. (1989). Recursive estimation, forecasting and adaptive control. In: Leondes, C. T. (ed.), Control and Dynamic Systems, San Diego: Academic Press 30 (3): 119-165.

Young P. C. (1992). Parallel processes in hydrology and water quality: a unified time series approach. Journal of the Institute of Water and Environmental Management 6: 598-612.

Young P. C. (1993). Time variable and state dependent modelling of nonstationary and nonlinear time series. In: Subba Rao T. (eds), Developments in Time Series Analysis, Chapman and Hall, London, pp. 374-413.

Young P. C. (1998). Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. Environmental Modelling and Software 13: 105-122.

Young P. C. (2000). Stochastic, dynamic modelling and signal processing: time variable and state dependent parameter estimation. In: Fitzgerald W. J., Walden A., Smith R., Young P. C. (eds), Nonlinear and Nonstationary Signal Processing. Cambridge: Cambridge University Press, pp. 74-114.

Young P. C. (2001a). Data-based mechanistic modelling and validation of rainfall-flow processes. In: Anderson M. G. and Bates P. D. (eds), Model Validation - Perspectives in hydrological science, John Wiley & Sons Ltd, Chichester, pp. 117-161.

Young P. C. (2001b) Comment on 'A quasi-ARMAX approach to the modelling of nonlinear systems' by J. Hu et al. Int. J. of Control. 74: 1767-1771.

Young P. C. and Beven K. J. (1994). Data-based mechanistic modelling and the rainfall-flow nonlinearity. Environmetrics 5 :335-363.

Young P. C., Jakeman A. J. and Post, D. (1997). Recent Advances in Data-based Modelling and Analysis of Hydrological Systems. Water Science and Technology 36 (5):99-116.

Young P.C., McKenna P. and Bruun J. (2001). Identification of non-linear stochastic systems by state dependent parameter estimation. Int. J. Control. 74(18): 1837-1857.

Young P. C. and Pedregal D. J. (1998). Data-based mechanistic modelling. In: Heij C., Hanzon B. and Praagman, K. (eds), System Dynamics in Economic and Financial Models. J. Wiley, Chichester, pp. 169-213.

Young P. C. and Pedregal D. J. (1999). Recursive and en-block approaches to signal extraction. J. Applied Stat. 26:103-128.

Young P. C., Pedregal D. J., Tych, W. (1999). Dynamic harmonic regression. Journal of Forecasting 18:369-394.

Young P. C., Taylor C. J., Tych W., Pedregal D. J., McKenna P.G. (2004). The Captain Toolbox, Centre for Research on Environmental Systems and Statistics, Lancaster University. Available at www.es.lancs.ac.uk/cres/captain.

ZEF (2000). Sustainable Water Use under Changing Land Use, Rainfall reliability and Water Demands in the Volta Basin – Glowa Project. Centre for Development Research (ZEF), Bonn University, Germany.

Zhao R. J. and Liu X. R. (1995). The Xinjiang model. In: Singh V. P. (eds), Computer models of watershed hydrology. Water Resources Publications, Littleton, Colo. Chapter 7.

9 APPENDIX

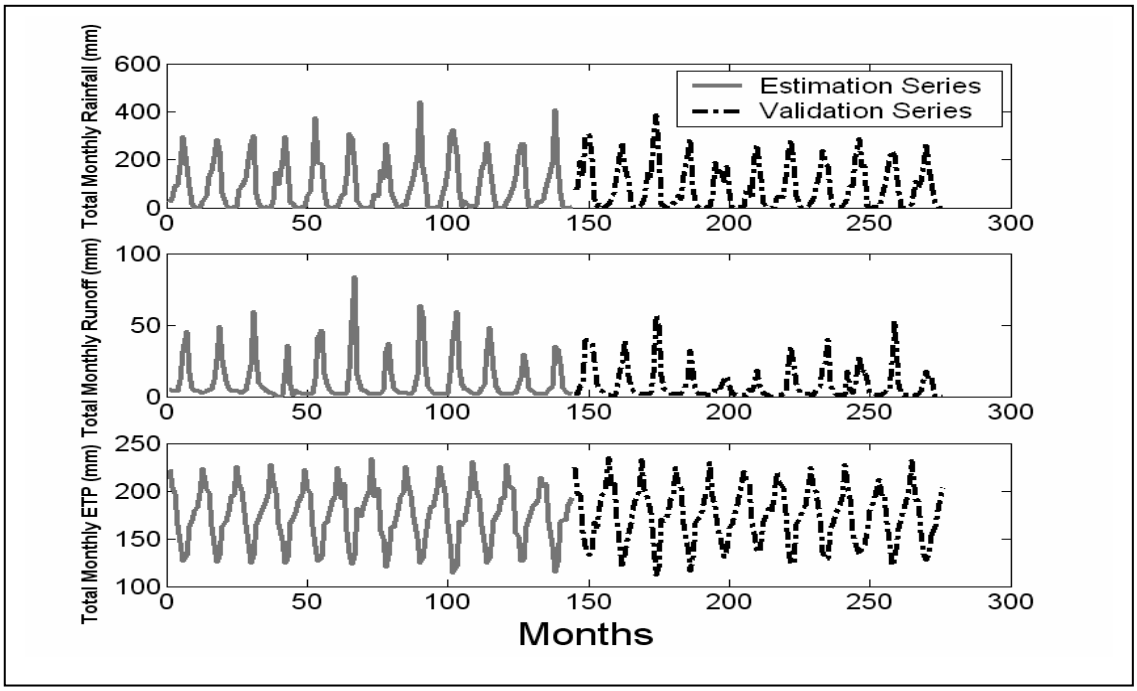


Figure 9.1i Observed monthly rainfall, runoff and ETP for Banzo, Black Volta River

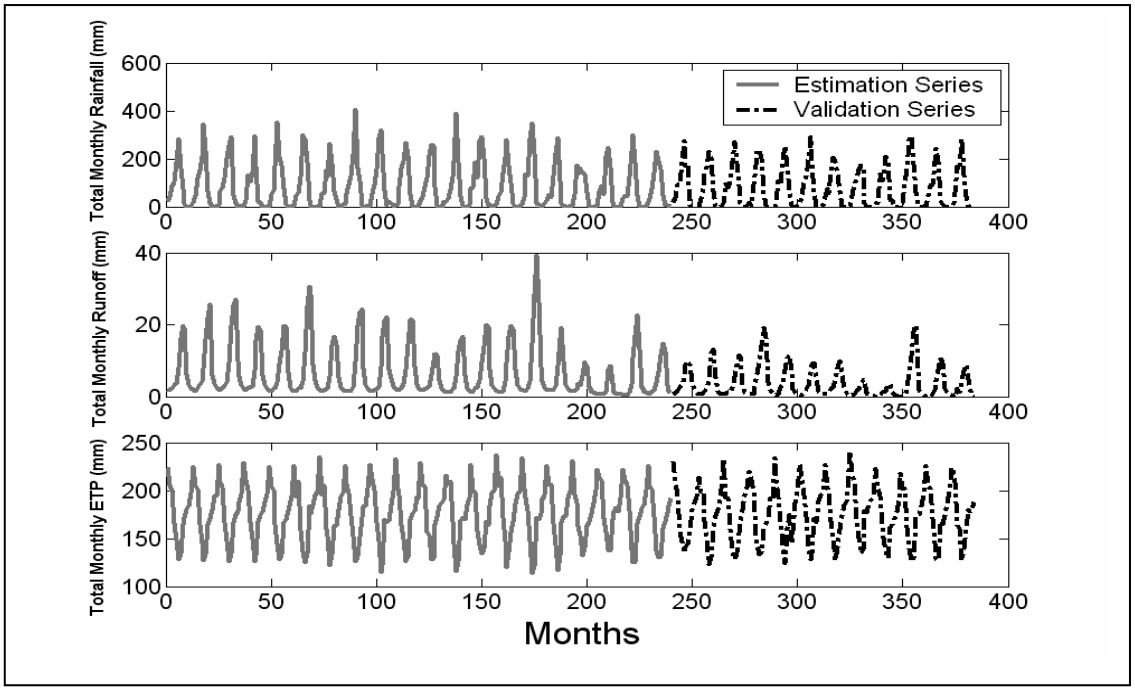


Figure 9.1ii Observed monthly rainfall, runoff and ETP for Nwokuy, Black Volta River

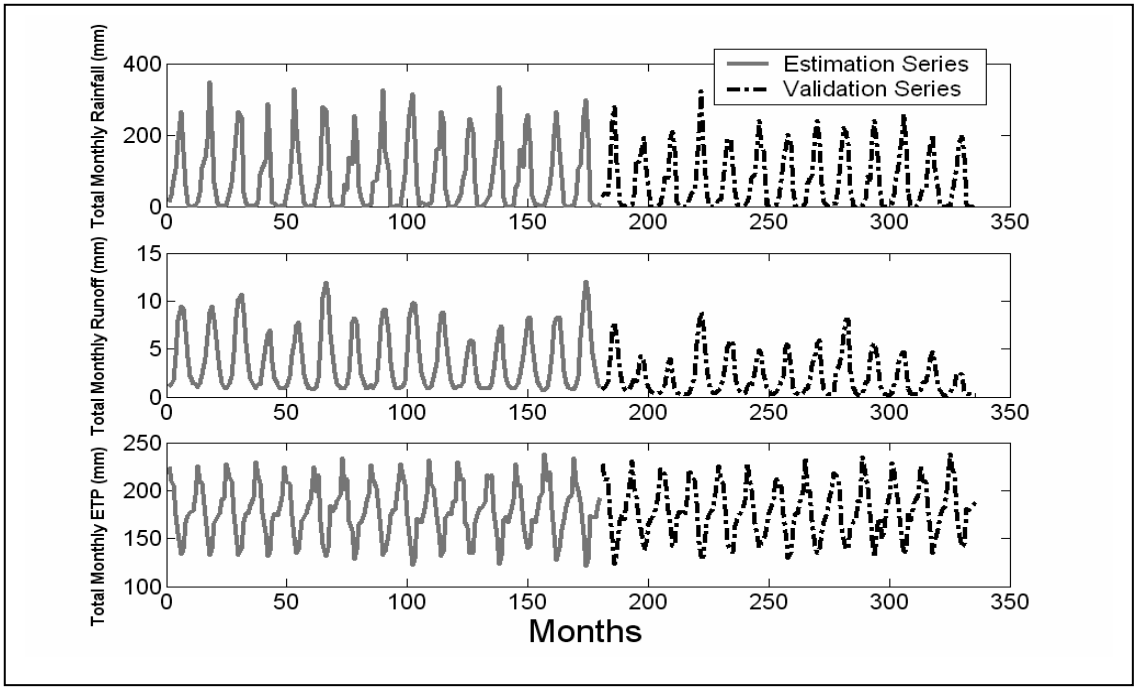


Figure 9.1iii Observed monthly rainfall, runoff and ETP for Manimenso, Black Volta River

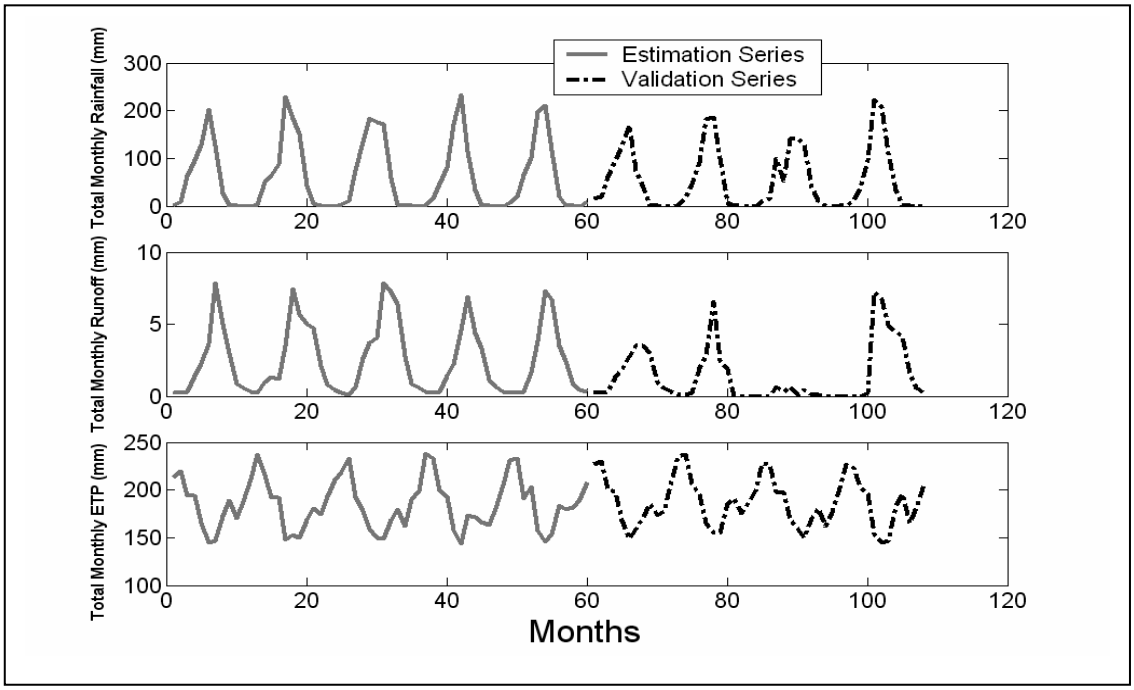


Figure 9.1iv Observed monthly rainfall, runoff and ETP for Tenado, Black Volta River

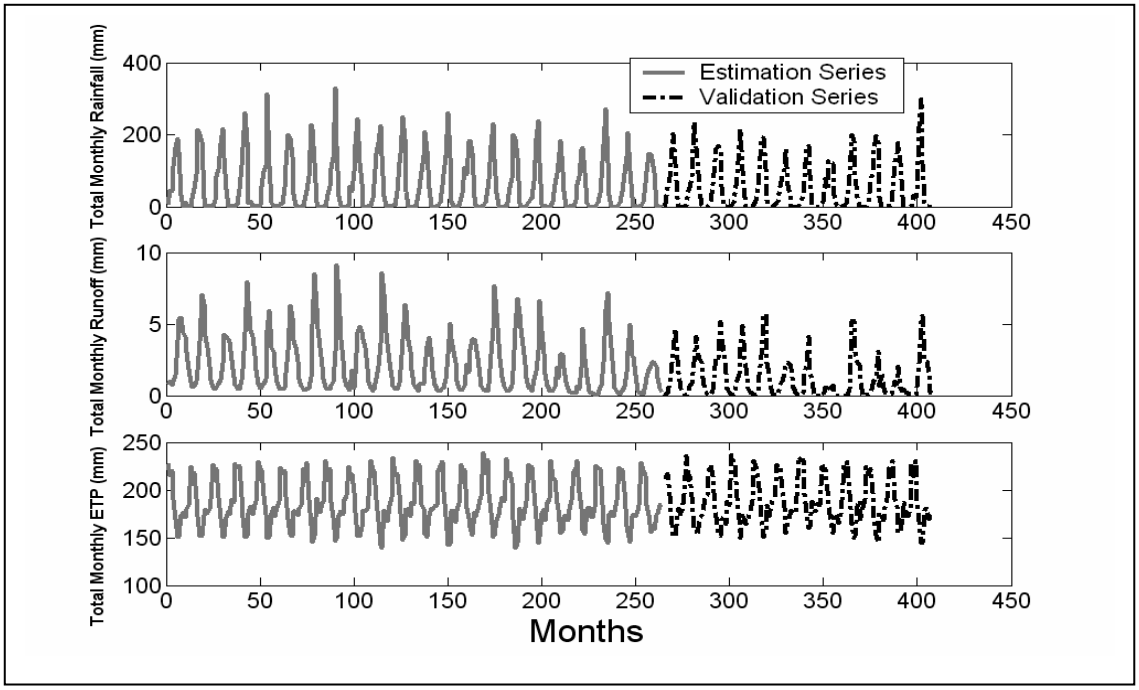


Figure 9.1v Observed monthly rainfall, runoff and ETP for Boromo, Black Volta River

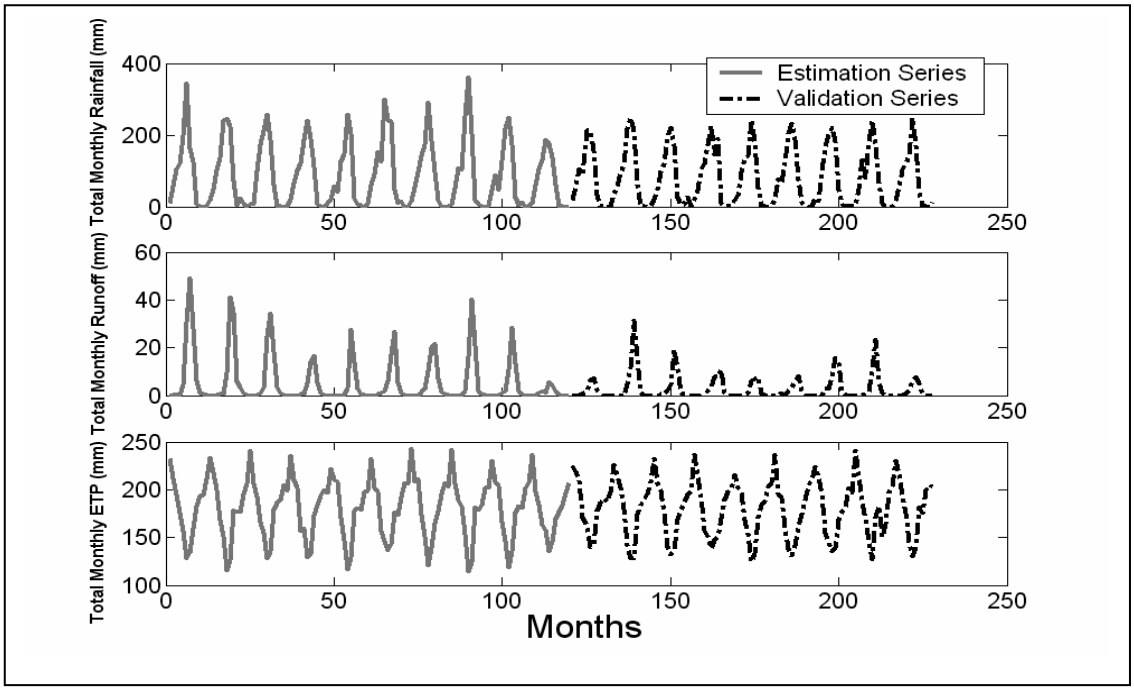


Figure 9.1vi Observed monthly rainfall, runoff and ETP for Debougou, Bourgouriba River (Black Volta Basin)

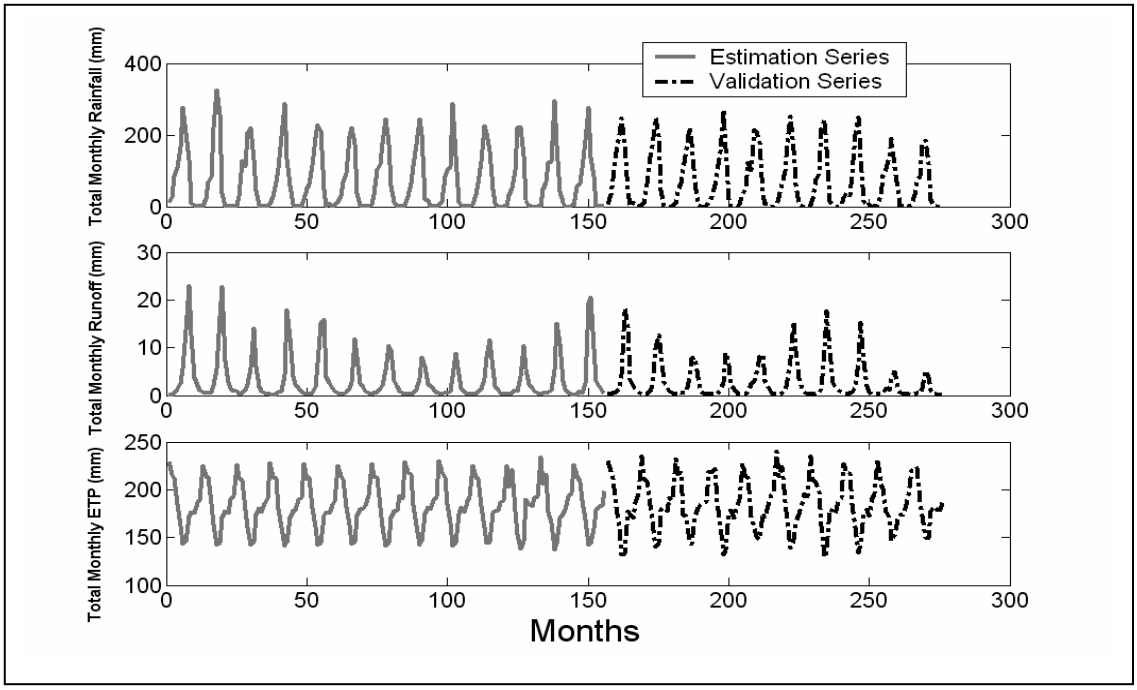


Figure 9.1vii Observed monthly rainfall, runoff and ETP for Lawra, Black Volta River

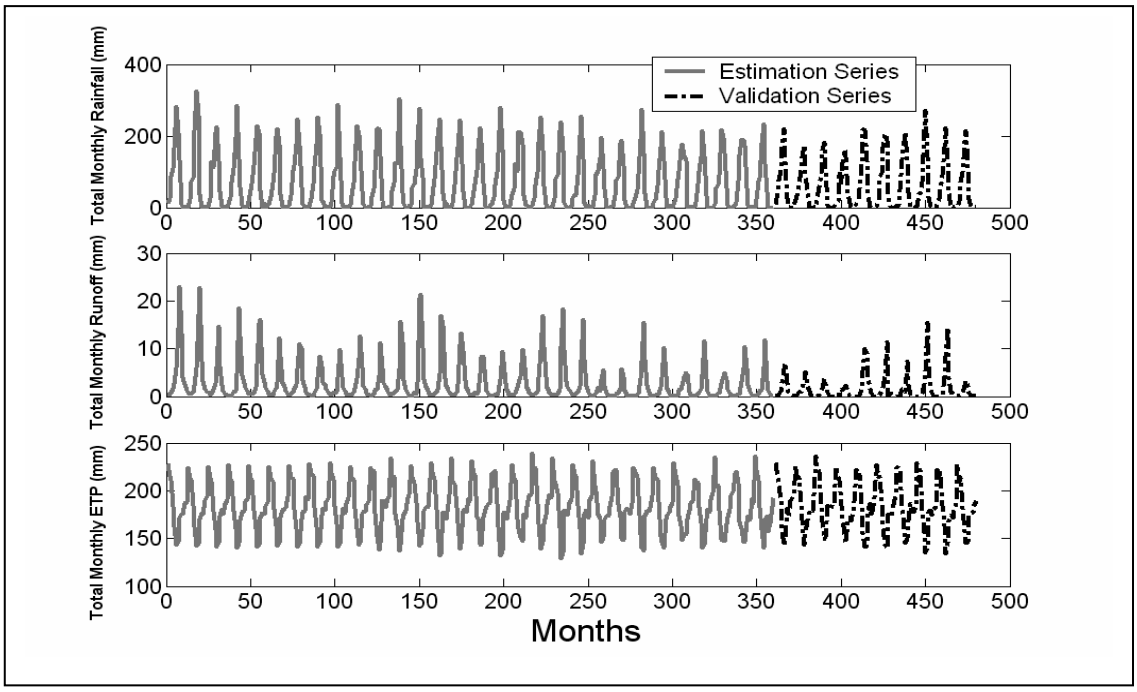


Figure 9.1viii Observed monthly rainfall, runoff and ETP for Dapola, Black Volta River

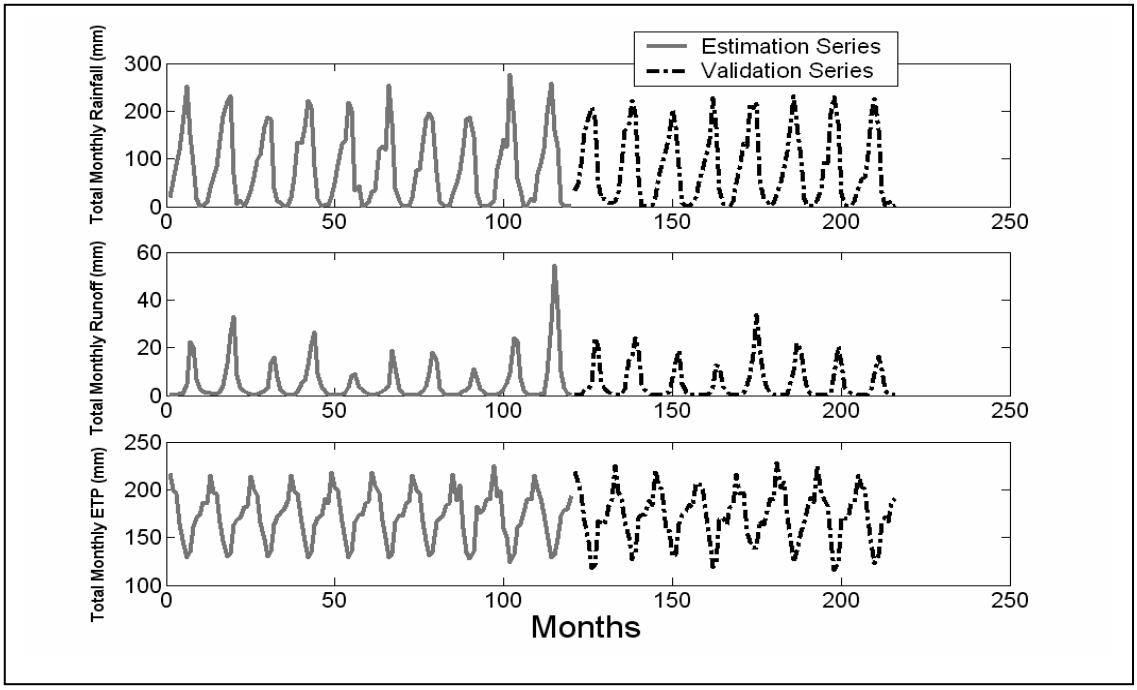


Figure 9.1ix Observed monthly rainfall, runoff and ETP for Bui, Black Volta River

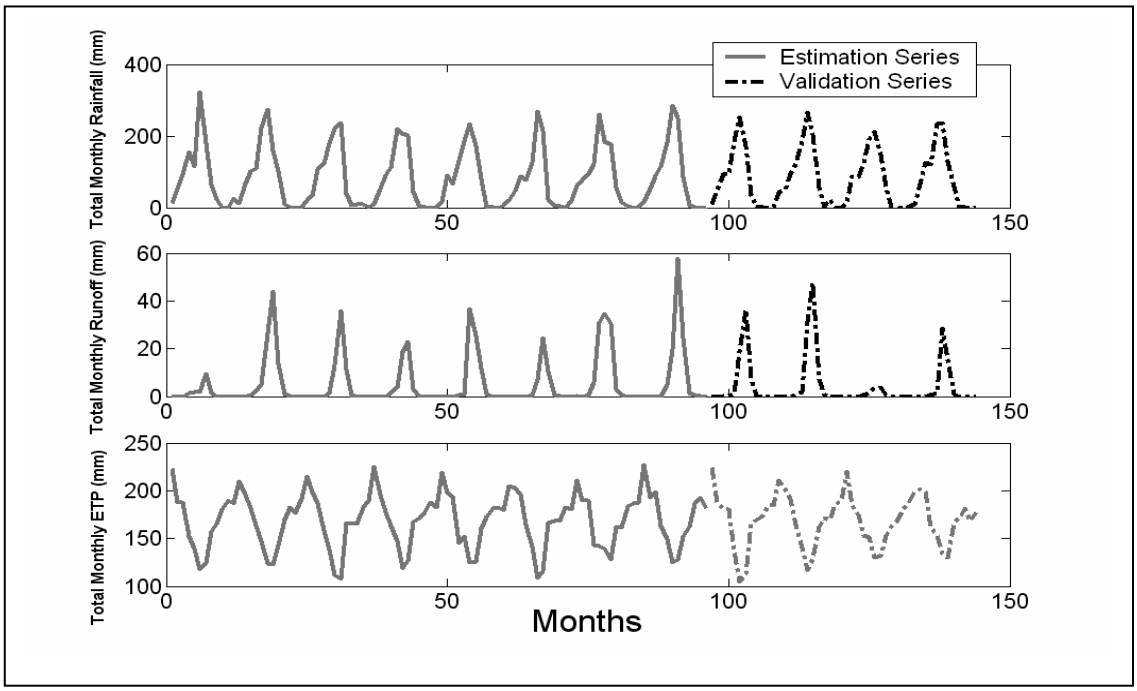


Figure 9.1x Observed monthly rainfall, runoff and ETP for Wiasi, Sisilli River (White Volta Basin)

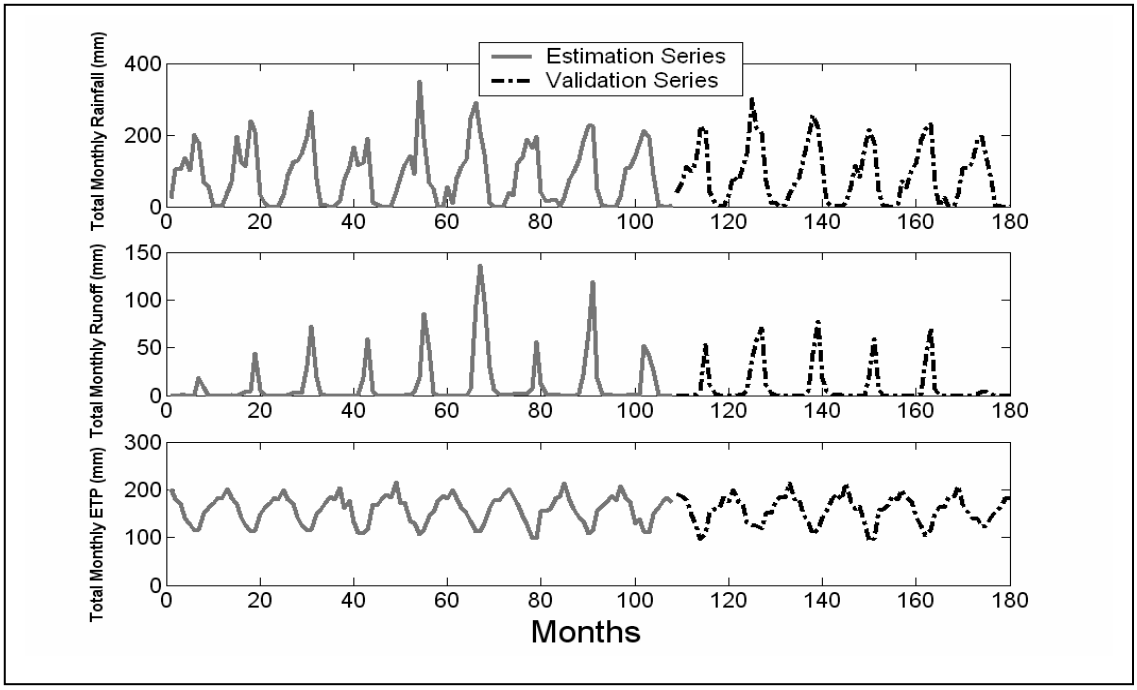


Figure 9.1xi Observed monthly rainfall, runoff and ETP for Yagaba, Kulpawn River (White Volta Basin)

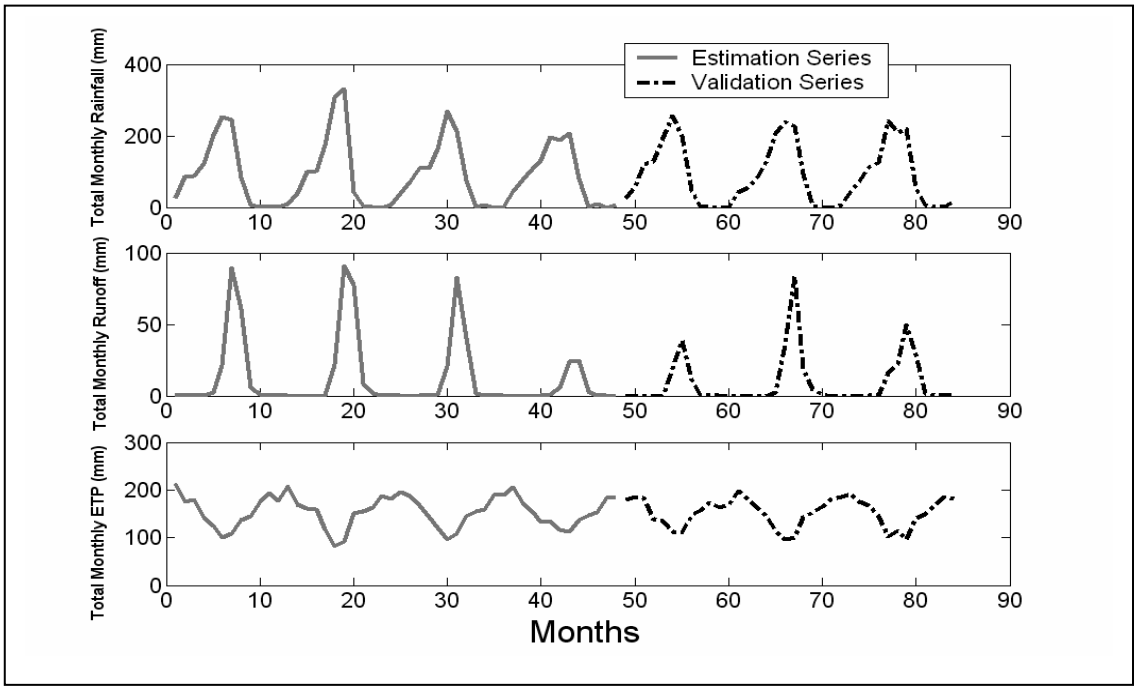


Figure 9.1xii Observed monthly rainfall, runoff and ETP for Nasia, Nasia River (White Volta Basin)

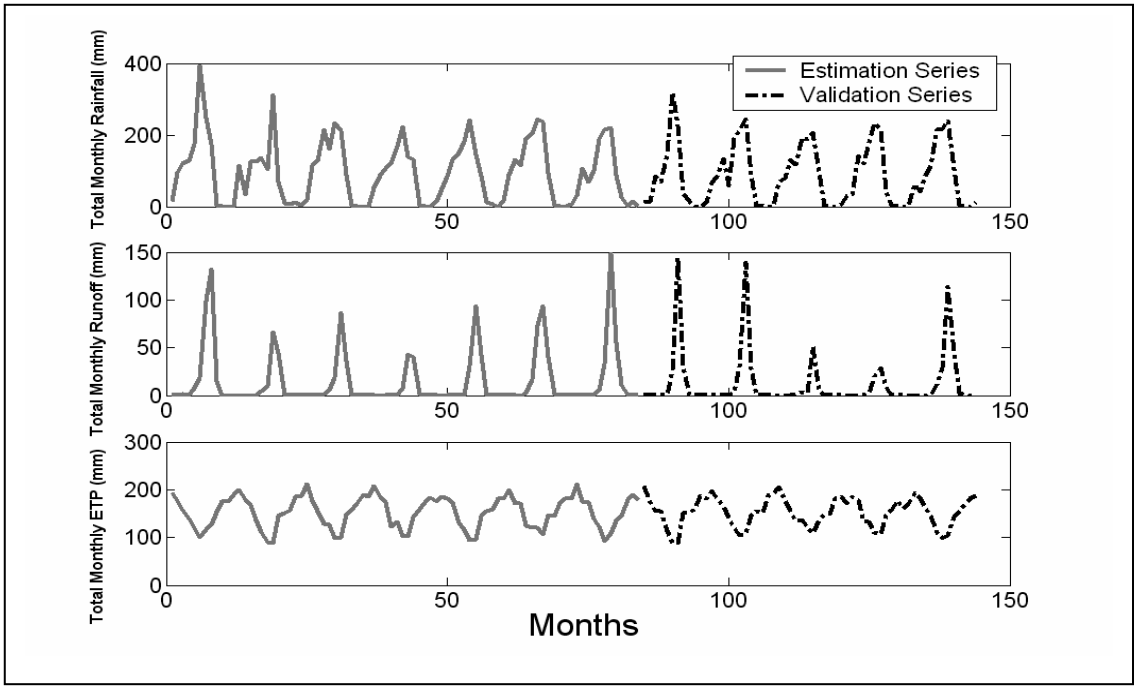


Figure 9.1xiii Observed monthly rainfall, runoff and ETP for Nabogo, Nabogo River (White Volta Basin)

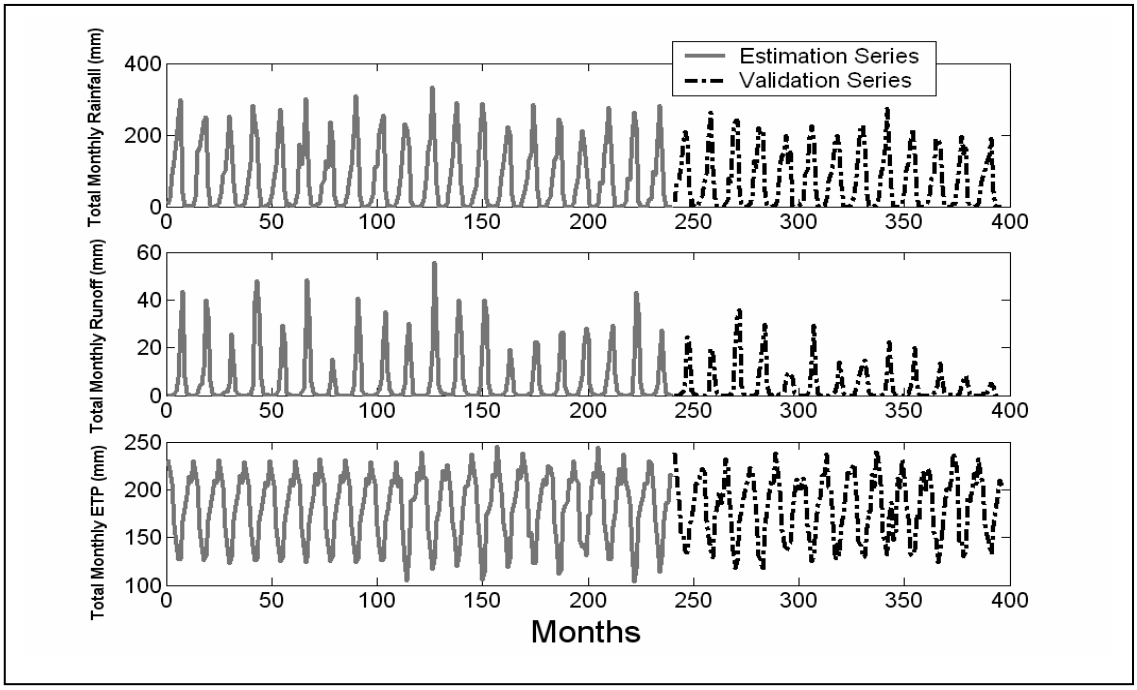


Figure 9.1xiv Observed monthly rainfall, runoff and ETP for Porga, Oti River

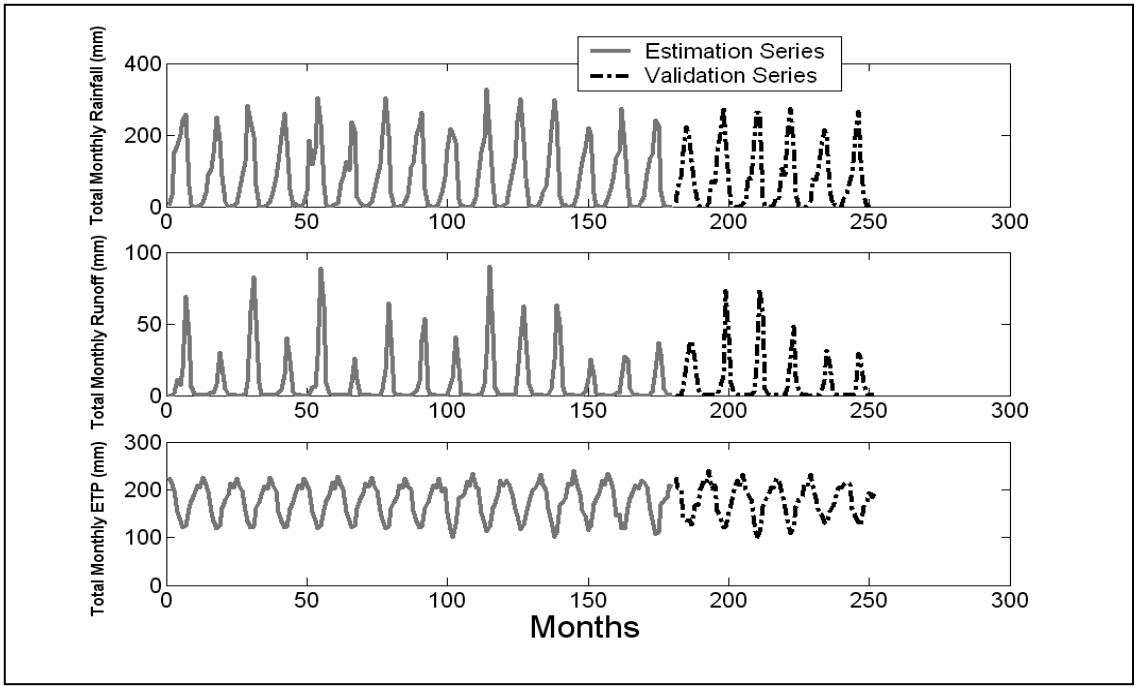


Figure 9.1xv Observed monthly rainfall, runoff and ETP for Mango, Oti River

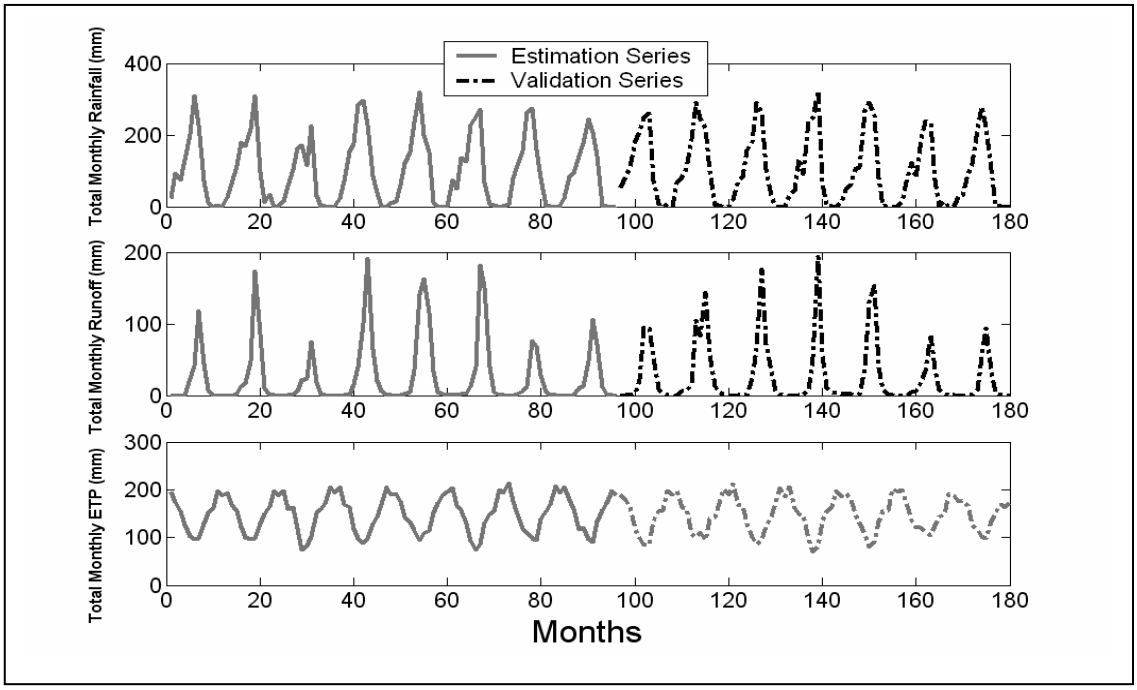


Figure 9.1xvi Observed monthly rainfall, runoff and ETP for Koumangou, Koumangou River (Oti Basin)

Table 9.1i Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Banzo. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	0.0	12.0	34.0	34.0	124.0	110.0	92.0	14.0	1.0	0.0	0.0	1.0
	Runoff	0.9	0.9	0.0	0.0	0.9	12.2	7.8	2.6	0.0	0.0	0.0	0.0
	ETP	212.2	190.9	184.3	155.2	133.9	111.6	119.5	143.2	158.0	168.7	177.7	190.8
Maximum	Rainfall	85.0	111.0	191.0	171.0	373.0	441.0	302.0	175.0	34.0	25.0	15.0	25.0
	Runoff	4.4	17.4	7.0	7.8	39.2	63.6	83.7	26.1	11.3	7.0	5.2	3.5
	ETP	235.4	213.3	211.4	177.5	155.2	137.6	139.8	182.1	174.9	185.3	196.6	203.7
Mean	Rainfall	16.5	51.0	93.8	128.3	224.6	281.7	192.4	62.1	10.3	2.0	1.0	3.8
	Runoff	1.9	2.6	2.7	3.8	12.7	31.1	38.8	15.2	5.6	3.4	2.4	2.0
	ETP	225.4	202.6	199.2	164.8	144.8	124.3	130.3	163.1	168.2	178.9	188.5	197.2
Std Dev.	Rainfall	17.7	19.6	41.9	29.1	56.1	69.2	42.0	36.7	7.7	5.8	3.3	6.7
	Runoff	0.9	3.4	1.6	1.8	8.0	13.5	18.5	6.7	3.2	1.7	1.3	1.0
	ETP	5.7	6.0	6.7	6.4	5.6	6.4	5.2	7.6	4.3	4.3	4.8	3.7
CV	Rainfall	1.08	0.38	0.45	0.23	0.25	0.25	0.22	0.59	0.74	2.88	3.12	1.78
	Runoff	0.48	1.31	0.58	0.48	0.63	0.43	0.48	0.44	0.57	0.49	0.55	0.52
	ETP	0.03	0.03	0.03	0.04	0.04	0.05	0.04	0.05	0.03	0.02	0.03	0.02

Table 9.1ii Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Nwokuy. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	0.5	7.0	38.3	52.5	122.5	145.8	109.0	14.5	1.0	0.0	0.0	0.5
	Runoff	0.0	0.0	0.0	0.7	1.1	1.5	3.3	3.5	0.9	0.7	0.2	0.0
	ETP	213.7	192.7	186.8	153.7	135.7	113.5	121.0	144.4	144.2	157.8	178.3	188.1
Maximum	Rainfall	75.0	94.3	185.5	174.5	354.0	407.5	291.3	175.8	30.8	21.3	13.0	18.0
	Runoff	2.2	1.7	2.4	4.8	7.2	14.8	28.5	39.4	28.5	10.0	5.4	3.1
	ETP	241.5	217.3	213.5	179.3	160.4	138.3	161.0	184.0	179.6	189.7	197.6	203.6
Mean	Rainfall	14.3	43.5	90.9	128.6	209.9	269.2	183.5	56.2	8.5	1.3	0.9	2.5
	Runoff	1.2	1.1	1.3	2.1	3.2	7.4	13.4	16.5	13.6	4.4	2.2	1.5
	ETP	227.0	205.4	201.1	166.1	146.6	126.8	133.6	165.5	169.7	178.5	188.7	196.5
Std Dev.	Rainfall	15.0	19.7	33.4	27.1	52.0	57.3	36.3	32.6	7.7	4.2	2.6	4.2
	Runoff	0.7	0.6	0.7	0.8	1.3	2.8	5.5	7.9	8.3	2.8	1.4	0.9
	ETP	6.1	6.4	6.3	6.2	5.8	6.0	7.1	7.8	6.5	6.2	5.1	4.1
CV	Rainfall	1.05	0.45	0.37	0.21	0.25	0.21	0.20	0.58	0.90	3.32	2.95	1.66
	Runoff	0.56	0.58	0.54	0.40	0.41	0.37	0.41	0.48	0.61	0.63	0.63	0.60
	ETP	0.03	0.03	0.03	0.04	0.04	0.05	0.05	0.05	0.04	0.03	0.03	0.02

Table 9.1iii Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Manimenso. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	1.1	6.9	33.3	74.6	119.3	187.6	104.1	8.3	0.9	0.0	0.0	0.1
	Runoff	0.2	0.9	1.1	2.1	2.5	2.2	0.5	0.4	0.3	0.2	0.1	0.1
	ETP	214.6	198.8	193.9	161.5	141.4	120.7	127.7	147.3	149.4	160.9	174.1	187.7
Maximum	Rainfall	50.1	69.3	160.1	178.7	328.0	347.4	251.3	158.3	22.9	13.3	7.9	8.6
	Runoff	1.4	2.6	3.4	6.1	10.4	12.1	10.9	8.9	5.6	3.5	1.9	1.4
	ETP	238.4	223.0	219.1	188.3	163.1	142.3	164.4	184.4	179.0	186.4	193.4	200.0
Mean	Rainfall	10.2	35.6	77.9	118.5	199.2	257.3	167.9	48.2	5.2	1.0	0.7	1.5
	Runoff	0.8	1.2	1.8	3.8	6.2	7.2	6.6	3.9	2.2	1.3	0.8	0.7
	ETP	227.8	210.7	207.9	174.9	153.2	132.6	139.0	167.1	171.5	175.9	185.1	193.8
Std Dev.	Rainfall	10.9	15.3	25.8	25.1	44.4	45.6	35.5	31.2	5.3	2.8	1.7	2.8
	Runoff	0.3	0.4	0.5	1.0	1.9	2.5	3.0	2.6	1.6	0.9	0.5	0.4
	ETP	5.9	6.3	6.1	6.6	5.4	5.7	6.9	7.6	5.9	5.5	4.7	3.8
CV	Rainfall	1.07	0.43	0.33	0.21	0.22	0.18	0.21	0.65	1.02	2.88	2.50	1.88
	Runoff	0.37	0.31	0.27	0.27	0.31	0.35	0.46	0.65	0.75	0.73	0.64	0.59
	ETP	0.03	0.03	0.03	0.04	0.04	0.04	0.05	0.05	0.03	0.03	0.03	0.02

Table 9.1iv Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Tenado. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	0.8	10.1	37.3	52.5	126.1	142.1	74.9	6.6	0.3	0.3	0.0	0.0
	Runoff	0.0	0.0	0.0	0.2	0.7	0.1	0.4	0.1	0.0	0.0	0.0	0.0
	ETP	213.2	215.8	190.8	177.8	147.9	144.1	146.4	167.9	166.2	162.4	176.1	198.7
Maximum	Rainfall	14.5	49.9	100.3	134.1	230.1	232.5	171.1	59.3	11.4	0.5	0.3	10.0
	Runoff	0.3	1.0	1.3	2.6	7.2	7.4	7.9	7.2	6.3	2.7	0.9	0.5
	ETP	238.0	237.5	206.7	202.9	169.3	158.6	173.3	185.6	195.9	182.1	193.3	211.3
Mean	Rainfall	4.7	18.7	62.0	92.6	176.4	189.9	121.3	33.4	2.4	0.3	0.0	1.1
	Runoff	0.2	0.2	0.5	1.3	3.1	4.8	5.2	3.9	3.0	1.1	0.5	0.3
	ETP	228.3	228.0	197.4	194.1	160.5	149.5	153.7	175.0	183.2	169.9	184.8	205.7
Std Dev.	Rainfall	4.9	12.2	18.8	21.2	38.3	26.6	28.2	15.0	3.7	0.1	0.1	3.3
	Runoff	0.1	0.3	0.4	0.7	1.8	2.5	2.6	2.0	2.0	0.9	0.3	0.2
	ETP	8.4	7.2	5.5	6.9	7.0	5.0	8.4	7.0	8.6	6.9	6.1	4.8
CV	Rainfall	1.03	0.65	0.30	0.23	0.22	0.14	0.23	0.45	1.55	0.30	3.00	2.96
	Runoff	0.67	1.29	0.82	0.55	0.59	0.51	0.50	0.52	0.68	0.80	0.63	0.66
	ETP	0.04	0.03	0.03	0.04	0.04	0.03	0.05	0.04	0.05	0.04	0.03	0.02

Table 9.1v Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Boromo. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	0.5	2.9	23.7	44.5	102.5	123.1	60.7	4.6	0.1	0.0	0.0	0.0
	Runoff	0.0	0.0	0.1	0.4	0.8	0.3	0.6	0.1	0.0	0.0	0.0	0.0
	ETP	212.6	208.6	210.8	177.7	158.8	139.0	139.6	159.0	163.6	164.5	171.0	176.9
Maximum	Rainfall	25.1	55.9	108.7	148.9	234.8	329.9	188.4	96.3	15.5	13.6	1.2	5.2
	Runoff	0.9	1.0	1.8	2.2	5.2	6.3	9.1	7.2	4.7	4.2	3.0	1.9
	ETP	239.3	230.6	231.9	208.0	182.9	165.3	173.0	190.7	188.6	185.3	193.3	195.3
Mean	Rainfall	3.6	18.6	53.2	93.3	166.5	211.1	130.9	30.1	1.8	0.9	0.2	0.7
	Runoff	0.3	0.3	0.5	1.0	1.8	3.6	4.9	3.7	2.7	1.9	1.0	0.5
	ETP	227.4	219.0	222.6	193.5	172.1	151.4	151.9	174.9	177.9	173.1	182.6	187.9
Std Dev.	Rainfall	4.8	12.8	16.7	20.9	32.6	47.1	29.8	18.6	3.0	2.5	0.3	1.4
	Runoff	0.3	0.2	0.3	0.5	0.9	1.4	2.2	2.0	1.4	1.3	0.9	0.5
	ETP	5.7	5.4	5.8	6.5	6.1	5.7	6.5	6.5	4.9	4.9	5.5	4.4
CV	Rainfall	1.35	0.69	0.31	0.22	0.20	0.22	0.23	0.62	1.64	2.89	1.99	1.99
	Runoff	0.81	0.79	0.62	0.46	0.47	0.37	0.45	0.54	0.50	0.71	0.92	0.95
	ETP	0.02	0.02	0.03	0.03	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02

Table 9.1vi Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Debougou. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	2.6	8.4	40.6	80.2	141.2	177.4	137.2	17.6	2.8	0.2	0.0	0.2
	Runoff	0.0	0.0	0.0	0.0	0.3	3.1	3.8	1.9	0.5	0.0	0.0	0.0
	ETP	216.2	195.8	183.7	154.6	138.3	113.9	123.4	150.6	150.3	164.1	187.2	196.4
Maximum	Rainfall	59.4	90.4	152.8	173.8	300.2	360.6	237.2	189.4	17.2	24.2	26.8	13.2
	Runoff	0.0	0.2	0.9	2.6	4.7	30.6	49.3	34.8	11.0	3.0	0.9	0.2
	ETP	243.6	219.2	212.0	183.3	161.7	140.0	170.8	182.5	186.4	200.0	207.3	211.0
Mean	Rainfall	20.8	53.7	98.2	125.4	199.9	250.6	183.1	64.9	6.7	4.0	2.1	2.6
	Runoff	0.0	0.0	0.2	0.8	2.0	9.6	21.2	14.3	3.1	0.7	0.2	0.0
	ETP	232.6	209.1	197.3	167.8	149.8	127.8	138.0	172.4	178.0	186.7	199.7	203.3
Std Dev.	Rainfall	13.8	20.4	25.7	21.9	38.8	42.5	25.2	42.1	4.1	6.8	6.4	4.0
	Runoff	0.0	0.0	0.3	0.7	1.1	7.3	13.6	9.1	2.9	0.7	0.2	0.1
	ETP	7.4	6.4	7.7	8.3	6.6	7.8	10.3	8.6	8.1	8.1	5.1	4.7
CV	Rainfall	0.7	0.4	0.3	0.2	0.2	0.2	0.1	0.7	0.6	1.7	3.0	1.6
	Runoff	0.0	4.4	1.1	0.8	0.6	0.8	0.6	0.6	0.9	1.0	1.2	2.4
	ETP	0.03	0.03	0.04	0.05	0.04	0.06	0.07	0.05	0.05	0.04	0.03	0.02

Table 9.1vii Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Lawra. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	1.3	10.1	32.2	81.7	111.0	185.2	108.5	12.8	0.7	0.0	0.0	0.1
	Runoff	0.2	0.2	0.2	0.4	0.7	2.8	3.7	1.1	0.3	0.2	0.2	0.2
	ETP	219.1	202.9	200.6	171.1	153.0	129.2	133.4	159.3	171.8	170.1	175.7	187.0
Maximum	Rainfall	36.1	56.2	128.4	153.6	226.1	324.5	259.5	130.5	17.5	12.0	3.9	7.1
	Runoff	0.9	0.5	0.9	2.4	3.6	19.3	20.6	23.0	12.7	3.7	2.1	1.1
	ETP	240.0	221.6	224.1	198.1	172.5	151.5	151.7	191.0	184.1	186.1	196.2	198.6
Mean	Rainfall	8.9	32.7	74.8	107.5	183.0	245.8	165.6	43.6	5.0	2.0	0.4	1.5
	Runoff	0.3	0.3	0.5	0.9	1.8	6.0	12.5	9.3	3.2	1.6	0.9	0.5
	ETP	228.7	213.6	213.7	182.4	163.6	140.7	144.2	173.7	178.3	177.9	188.0	192.3
Std Dev.	Rainfall	8.0	11.5	21.0	16.6	30.8	35.4	34.4	27.1	4.3	3.5	0.8	1.8
	Runoff	0.2	0.1	0.2	0.5	0.8	3.5	4.4	5.5	2.5	0.8	0.5	0.3
	ETP	4.8	4.6	5.5	5.7	4.4	5.3	4.0	5.8	3.1	4.2	4.6	3.3
CV	Rainfall	0.89	0.35	0.28	0.15	0.17	0.14	0.21	0.62	0.87	1.78	2.02	1.17
	Runoff	0.52	0.29	0.36	0.52	0.46	0.58	0.36	0.59	0.77	0.48	0.56	0.53
	ETP	0.02	0.02	0.03	0.03	0.03	0.04	0.03	0.03	0.02	0.02	0.02	0.02

Table 9.1viii Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Dapola. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	1.5	7.6	34.1	68.3	113.2	163.3	91.6	8.1	0.6	0.1	0.0	0.1
	Runoff	0.0	0.0	0.0	0.4	0.7	0.8	2.0	0.0	0.0	0.0	0.0	0.0
	ETP	212.1	202.1	200.1	167.8	149.4	128.9	132.8	154.5	157.5	166.6	175.1	182.1
Maximum	Rainfall	35.3	56.6	128.1	155.0	227.3	324.6	259.0	133.7	17.4	12.0	6.2	7.5
	Runoff	0.9	0.6	1.0	1.9	5.4	20.6	21.4	23.0	13.3	4.0	2.5	1.4
	ETP	239.2	224.2	223.7	197.6	173.3	154.6	170.6	190.8	186.3	189.7	195.7	198.4
Mean	Rainfall	8.4	29.5	71.7	104.9	180.2	232.0	154.1	43.1	4.0	1.4	0.5	1.3
	Runoff	0.2	0.2	0.5	1.0	2.2	5.8	11.2	7.0	2.6	1.3	0.6	0.3
	ETP	227.3	213.9	213.1	182.2	162.2	141.8	145.2	172.8	177.0	177.1	186.4	191.4
Std Dev.	Rainfall	6.8	12.0	19.1	18.3	29.1	37.9	33.5	26.4	3.9	2.8	1.2	1.8
	Runoff	0.2	0.2	0.3	0.4	1.0	3.3	5.0	5.4	2.3	1.0	0.6	0.4
	ETP	5.4	5.2	5.5	6.0	5.5	5.5	6.2	6.3	5.1	4.9	5.0	4.0
CV	Rainfall	0.80	0.41	0.27	0.17	0.16	0.16	0.22	0.61	0.97	2.02	2.43	1.44
	Runoff	1.03	0.75	0.52	0.44	0.46	0.57	0.44	0.76	0.89	0.76	0.97	1.03
	ETP	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04	0.03	0.03	0.03	0.02

Table 9.1ix Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Bui. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	11.0	29.6	60.8	89.3	119.4	186.8	147.9	27.4	1.5	0.4	0.1	0.9
	Runoff	0.0	0.0	0.2	0.2	0.7	2.9	8.4	6.9	2.4	0.9	0.2	0.1
	ETP	206.8	188.2	188.8	155.8	135.4	115.7	120.6	149.3	163.6	163.0	179.5	181.5
Maximum	Rainfall	49.7	79.9	134.2	158.2	215.6	276.5	231.6	125.2	44.3	17.3	7.3	18.2
	Runoff	0.8	0.7	1.2	5.0	15.0	30.4	54.6	36.1	9.9	3.5	2.0	1.0
	ETP	228.4	207.6	207.8	183.1	157.4	139.4	139.2	183.1	175.8	181.0	191.7	194.2
Mean	Rainfall	24.8	56.6	93.1	125.0	170.2	223.9	180.2	59.9	13.0	5.0	1.5	6.3
	Runoff	0.2	0.2	0.5	1.6	4.2	9.7	21.5	18.3	5.5	2.2	0.9	0.4
	ETP	218.0	199.9	197.5	166.2	147.6	126.8	132.0	165.2	170.2	173.3	185.8	187.2
Std Dev.	Rainfall	11.0	14.5	19.4	16.8	26.4	24.9	22.4	28.1	10.8	5.6	2.1	5.2
	Runoff	0.2	0.2	0.3	1.3	3.7	6.8	10.3	7.8	2.5	0.8	0.5	0.3
	ETP	5.3	4.7	5.1	6.5	4.5	5.6	4.9	6.8	3.3	5.1	3.8	3.3
CV	Rainfall	0.44	0.26	0.21	0.13	0.16	0.11	0.12	0.47	0.83	1.12	1.37	0.82
	Runoff	1.04	0.90	0.63	0.81	0.90	0.71	0.48	0.43	0.45	0.34	0.57	0.78
	ETP	0.02	0.02	0.03	0.04	0.03	0.04	0.04	0.04	0.02	0.03	0.02	0.02

Table 9.1x Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Wiasi. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	8.8	33.3	65.3	76.3	115.0	184.0	128.3	23.3	0.8	0.8	0.0	1.0
	Runoff	0.0	0.0	0.0	0.0	0.7	1.7	3.7	0.7	0.0	0.0	0.0	0.0
	ETP	198.6	184.4	174.0	143.5	137.2	104.3	107.5	142.7	161.0	165.4	168.3	177.1
Maximum	Rainfall	59.5	90.0	124.5	154.3	259.8	322.0	251.0	100.3	21.8	21.8	10.3	25.0
	Runoff	0.2	0.2	0.4	5.9	31.1	36.8	57.7	24.4	1.5	0.7	0.2	0.2
	ETP	227.0	202.5	199.1	180.6	153.6	139.4	131.8	167.5	172.8	186.8	192.7	191.3
Mean	Rainfall	19.9	60.3	93.8	117.0	190.3	246.5	186.4	58.0	5.7	3.9	0.9	5.3
	Runoff	0.0	0.0	0.1	1.1	4.6	20.0	29.4	7.9	0.5	0.1	0.0	0.0
	ETP	215.8	194.8	187.9	159.4	144.8	121.1	123.4	160.3	167.6	176.6	183.4	185.0
Std Dev.	Rainfall	15.3	17.0	13.9	19.1	41.5	38.6	36.2	21.7	6.7	6.3	3.0	8.1
	Runoff	0.1	0.1	0.2	1.7	8.5	12.0	15.9	6.9	0.5	0.2	0.1	0.1
	ETP	8.9	6.3	8.3	10.1	6.0	10.4	7.5	7.5	4.1	7.2	6.3	4.6
CV	Rainfall	0.77	0.28	0.15	0.16	0.22	0.16	0.19	0.37	1.18	1.59	3.46	1.52
	Runoff	3.46	3.46	1.90	1.48	1.85	0.60	0.54	0.87	1.01	2.16	3.46	3.46
	ETP	0.04	0.03	0.04	0.06	0.04	0.09	0.06	0.05	0.02	0.04	0.03	0.02

Table 9.1xi Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Yagaba. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	8.7	31.0	79.0	82.0	91.0	122.7	142.0	13.0	1.7	0.0	0.0	0.7
	Runoff	0.0	0.0	0.0	0.0	0.0	0.3	4.6	1.7	0.0	0.0	0.0	0.0
	ETP	190.9	161.1	154.0	129.1	110.6	92.3	97.7	131.0	151.9	158.5	177.1	173.5
Maximum	Rainfall	84.3	107.7	197.0	167.3	302.3	351.0	267.3	148.0	58.3	25.3	19.3	55.7
	Runoff	0.3	1.7	2.0	4.1	40.3	94.7	136.4	98.7	29.0	5.8	1.4	0.9
	ETP	215.7	188.3	181.0	165.0	139.2	125.0	124.4	169.6	168.0	180.3	191.4	189.1
Mean	Rainfall	35.4	75.7	114.1	129.9	167.9	221.2	203.8	66.2	13.2	4.8	1.3	7.7
	Runoff	0.0	0.2	0.4	1.0	5.9	30.4	64.6	19.7	3.1	0.7	0.2	0.1
	ETP	204.9	178.5	169.9	142.6	128.6	108.8	113.3	149.9	159.3	170.9	185.1	181.1
Std Dev.	Rainfall	20.1	21.2	25.2	19.8	56.7	52.1	28.9	33.9	17.2	8.0	5.0	15.8
	Runoff	0.1	0.4	0.7	1.4	10.9	27.7	33.7	24.7	7.2	1.5	0.4	0.3
	ETP	7.2	7.1	7.7	9.0	6.9	9.0	7.5	8.6	4.7	6.8	3.5	4.2
CV	Rainfall	0.57	0.28	0.22	0.15	0.34	0.24	0.14	0.51	1.30	1.67	3.74	2.06
	Runoff	3.87	2.31	1.66	1.33	1.85	0.91	0.52	1.25	2.33	2.23	2.25	2.28
	ETP	0.03	0.04	0.05	0.06	0.05	0.08	0.07	0.06	0.03	0.04	0.02	0.02

Table 9.1xii Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Nasia. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	10.0	39.0	79.5	100.5	164.0	187.5	197.5	43.0	1.0	1.0	0.0	2.5
	Runoff	0.0	0.0	0.0	0.0	0.0	5.7	23.7	11.3	0.8	0.2	0.0	0.0
	ETP	178.4	170.0	155.0	133.1	101.4	82.9	92.1	136.2	145.7	153.2	163.4	168.3
Maximum	Rainfall	45.5	86.0	120.0	130.5	242.0	307.0	331.0	93.0	8.0	8.5	3.0	17.5
	Runoff	0.0	0.1	0.3	0.2	15.8	35.8	91.0	77.6	8.4	1.4	0.6	0.2
	ETP	213.0	186.6	182.9	158.6	134.9	117.0	112.5	150.8	156.6	175.9	193.4	189.9
Mean	Rainfall	33.3	65.4	101.7	120.4	196.8	245.9	234.0	67.9	3.9	2.7	0.7	6.6
	Runoff	0.0	0.0	0.1	0.1	3.2	21.2	65.5	37.3	3.7	0.7	0.2	0.1
	ETP	198.8	178.7	168.4	143.9	120.8	102.8	104.0	142.7	151.8	165.7	183.4	181.3
Std Dev.	Rainfall	13.4	16.1	14.4	10.9	25.2	38.9	45.3	19.7	2.5	3.1	1.3	5.3
	Runoff	0.0	0.0	0.1	0.1	5.6	8.8	27.6	24.0	2.8	0.4	0.2	0.1
	ETP	11.6	6.6	9.6	7.9	11.8	12.4	7.9	5.5	4.4	8.3	9.7	6.8
CV	Rainfall	0.40	0.25	0.14	0.09	0.13	0.16	0.19	0.29	0.64	1.12	1.75	0.81
	Runoff	1.33	2.65	1.63	1.13	1.74	0.41	0.42	0.64	0.75	0.58	1.02	1.44
	ETP	0.06	0.04	0.06	0.05	0.10	0.12	0.08	0.04	0.03	0.05	0.05	0.04

Table 9.1xiii Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Nabogo. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	11.0	15.0	70.0	58.0	137.0	103.0	138.0	37.0	2.0	2.0	0.0	2.0
	Runoff	0.0	0.0	0.0	0.0	0.2	3.0	28.5	5.0	0.2	0.0	0.0	0.0
	ETP	172.2	171.4	155.1	121.9	110.6	87.4	87.7	125.8	145.1	148.2	161.3	169.7
Maximum	Rainfall	114.0	130.0	142.0	215.0	215.0	393.0	311.0	173.0	26.0	7.0	14.0	18.0
	Runoff	0.1	0.3	1.1	2.5	16.8	73.9	148.6	132.0	14.5	0.3	0.1	0.1
	ETP	212.8	185.6	177.0	155.2	136.7	121.4	115.6	151.1	159.3	179.1	189.6	190.8
Mean	Rainfall	50.8	75.9	114.7	126.1	178.4	235.5	220.3	90.9	7.5	3.0	2.2	6.8
	Runoff	0.0	0.0	0.3	0.5	5.3	25.1	92.5	42.1	2.7	0.1	0.0	0.0
	ETP	198.9	179.1	165.2	135.6	121.9	102.0	102.6	142.5	152.9	166.1	182.2	182.5
Std Dev.	Rainfall	31.5	36.1	23.3	44.3	22.5	69.6	45.8	36.4	7.1	1.8	5.1	5.9
	Runoff	0.0	0.1	0.3	0.7	4.9	18.2	39.9	31.8	4.8	0.1	0.0	0.0
	ETP	11.9	4.5	8.2	9.2	8.4	11.1	8.4	7.0	4.6	10.8	7.8	6.2
CV	Rainfall	0.62	0.48	0.20	0.35	0.13	0.30	0.21	0.40	0.94	0.59	2.34	0.86
	Runoff	1.63	1.96	1.28	1.33	0.93	0.72	0.43	0.76	1.76	0.51	0.70	0.94
	ETP	0.06	0.03	0.05	0.07	0.07	0.11	0.08	0.05	0.03	0.07	0.04	0.03

Table 9.1xiv Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Porga. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	2.9	8.0	52.1	70.0	109.1	142.3	106.4	12.4	0.1	0.0	0.0	0.0
	Runoff	0.0	0.0	0.0	0.1	0.1	3.0	5.0	3.8	0.2	0.0	0.0	0.0
	ETP	200.5	202.7	185.3	145.8	125.1	103.9	112.2	144.7	147.4	155.1	186.0	196.7
Maximum	Rainfall	22.1	71.1	173.9	165.7	281.1	334.4	297.6	152.0	10.3	10.1	1.9	6.2
	Runoff	0.7	0.2	3.1	6.6	10.5	39.3	55.9	43.6	9.6	2.6	0.9	0.8
	ETP	244.9	232.1	219.5	185.3	155.7	144.0	171.9	191.4	189.1	210.0	222.5	217.1
Mean	Rainfall	9.5	33.8	88.4	120.3	188.4	240.6	188.0	44.3	1.8	1.2	0.3	1.0
	Runoff	0.0	0.0	0.3	1.0	3.1	13.5	26.9	20.2	3.3	0.7	0.3	0.1
	ETP	231.3	217.7	204.6	161.1	142.4	125.0	129.4	168.3	179.4	193.9	212.1	207.2
Std Dev.	Rainfall	5.2	16.6	24.5	23.6	33.2	44.8	41.1	24.3	2.5	2.2	0.5	1.3
	Runoff	0.1	0.1	0.6	1.3	2.4	8.1	12.3	12.0	2.7	0.6	0.3	0.1
	ETP	8.7	6.9	7.5	7.8	7.2	10.1	9.9	9.2	7.4	9.6	7.2	4.7
CV	Rainfall	0.55	0.49	0.28	0.20	0.18	0.19	0.22	0.55	1.41	1.91	1.87	1.34
	Runoff	4.37	3.21	2.33	1.27	0.79	0.60	0.46	0.60	0.82	0.81	0.91	1.52
	ETP	0.04	0.03	0.04	0.05	0.05	0.08	0.08	0.05	0.04	0.05	0.03	0.02

Table 9.1xv Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Mango. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	3.3	15.9	55.9	77.3	100.2	189.8	125.5	19.3	0.2	0.0	0.0	0.0
	Runoff	0.0	0.0	0.0	0.2	0.6	7.8	25.1	7.8	0.8	0.1	0.1	0.1
	ETP	209.0	195.8	181.3	139.1	117.4	97.9	108.2	146.1	165.4	179.0	184.0	193.9
Maximum	Rainfall	28.3	76.5	185.4	171.8	281.1	328.2	264.3	85.1	16.1	9.6	2.1	5.9
	Runoff	0.3	0.1	1.7	11.3	21.3	59.1	90.1	62.5	8.7	1.8	0.7	0.4
	ETP	240.2	220.5	211.5	179.0	149.9	137.8	129.6	182.8	183.1	199.7	220.7	212.5
Mean	Rainfall	12.4	41.2	91.4	125.5	190.1	258.0	199.9	47.3	3.7	1.6	0.4	1.5
	Runoff	0.1	0.0	0.3	1.7	5.4	21.4	51.3	33.0	4.4	0.9	0.4	0.2
	ETP	227.3	209.4	198.6	154.9	136.7	117.1	122.1	162.7	175.2	190.8	211.0	205.0
Std Dev.	Rainfall	6.3	15.4	28.5	24.2	38.0	36.9	37.8	18.0	4.3	2.6	0.6	1.9
	Runoff	0.1	0.0	0.4	2.6	4.8	12.6	22.2	16.6	2.6	0.4	0.2	0.1
	ETP	7.0	5.8	6.9	7.6	7.3	10.1	5.4	7.0	4.5	5.8	8.0	4.2
CV	Rainfall	0.51	0.37	0.31	0.19	0.20	0.14	0.19	0.38	1.18	1.63	1.55	1.27
	Runoff	1.12	1.30	1.44	1.53	0.88	0.59	0.43	0.50	0.59	0.49	0.44	0.58
	ETP	0.03	0.03	0.03	0.05	0.05	0.09	0.04	0.04	0.03	0.03	0.04	0.02

Table 9.1xvi Selected monthly statistics of the monthly rainfall, runoff and potential evapotranspiration (ETP) series at Koumangou. (All values in mm except CV = coefficient of variation, which is dimensionless)

Statistic	Series	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
Minimum	Rainfall	7.5	30.5	75.0	80.5	168.0	114.5	158.5	32.0	1.0	0.0	0.0	0.0
	Runoff	0.0	0.0	0.0	0.4	7.4	23.0	68.6	18.7	2.2	0.4	0.0	0.0
	ETP	175.9	159.8	137.8	108.0	74.2	69.1	80.0	115.5	142.2	149.1	162.7	173.2
Maximum	Rainfall	74.0	92.0	151.5	188.0	291.0	320.5	323.5	152.5	38.5	35.0	3.0	13.5
	Runoff	1.7	6.5	6.5	16.5	105.1	141.5	194.1	151.1	39.9	6.5	3.0	2.2
	ETP	213.8	179.2	168.1	147.2	120.9	111.6	108.6	150.8	158.4	172.6	208.7	196.5
Mean	Rainfall	30.9	67.4	110.7	144.0	221.1	258.1	240.1	88.4	8.7	6.0	0.2	2.1
	Runoff	0.2	0.9	1.9	5.9	26.5	68.9	134.6	61.6	14.4	3.0	1.1	0.4
	ETP	198.8	171.2	156.4	124.2	103.5	91.1	95.2	131.9	150.4	165.2	197.4	189.0
Std Dev.	Rainfall	19.8	17.0	20.6	33.6	41.6	49.3	41.9	30.0	11.3	9.1	0.8	3.9
	Runoff	0.5	1.7	1.7	5.4	23.9	35.2	44.8	34.8	12.0	2.1	0.8	0.6
	ETP	9.8	6.3	9.0	9.1	12.2	12.3	7.7	7.8	5.4	7.5	10.8	5.7
CV	Rainfall	0.64	0.25	0.19	0.23	0.19	0.19	0.17	0.34	1.29	1.53	3.87	1.87
	Runoff	2.67	1.91	0.93	0.92	0.90	0.51	0.33	0.57	0.84	0.68	0.76	1.43
	ETP	0.05	0.04	0.06	0.07	0.12	0.14	0.08	0.06	0.04	0.05	0.05	0.03

ACKNOWLEDGEMENTS

I would like to acknowledge the several personalities and bodies that have contributed immensely to the successful completion of this study.

First of all, I should thank the people and government of Germany for providing the necessary funds through the Glowa-Volta Project to enable me undertake this study. I am grateful too, to the people and government of my home country, Ghana, for granting me study leave with pay through the Council for Scientific and Industrial Research (CSIR), and thus enabling me take care of my family back home while undertaking this study abroad.

Next, I would like to express my sincerest gratitude to Professor Dr. Ir. Nick Van de Giesen, who, as my tutor and supervisor, saw to the details of all the scientific work undertaken in this study and offered me all the encouragement and necessary pushes to bring out the best in me. Nick, I can't thank you enough!

I thank the directors of the Centre for Development Research (ZEF) of the University of Bonn and, in particular, Professor Dr. P. Vlek, Director of ZEFc and leader of the Glowa-Volta Project, for approving the necessary funds for my study for me, ensuring that these got to me on time and arranging the necessary courses for me to enable me register at the university of Bonn as a graduate student.

Professor Dr. Ing. H. Eggers assisted me in various ways in my academic work and in the preparation of my thesis. I gratefully acknowledge his contribution to the success of my work.

I express my gratitude to the members of my examination panel, Prof.Dr.-Ing.Dr.h.c. J. J. Bogardi, Prof.dr.ir. A.W. Heemink, Prof.dr.ir. A.E. Mynett, Prof.dr.ir. H.H.G. Savenije and Prof.dr.ir. G.S. Stelling, for accepting the nomination by the Technical University of Delft, the Netherlands, to the panel and critically but constructively reviewing my thesis.

I am also grateful to Dr. Manske, the staff of the Doctoral programme and Glowa-Volta Project at ZEF and the ZEFc secretariat for all the assistance given me in the course of my study, particularly in ensuring that my stay at ZEF, in Bonn and Germany in general, was fruitful and pleasant. Thanks to Mr. Volker Merx and the staff of the ZEF Library for ensuring that I got almost all of the literature I requested, particularly those through the inter-library loan system.

I acknowledge the contributions of Dr. Marc Andreini, former Glowa-Volta Project Coordinator in Ghana and Dr. Mathias Fosu of CSIR-SARI and Glowa-Volta Project in Tamale to this study. They provided the necessary logistics and made sure I got the necessary funds on time for my field research activities in Ghana and I cannot thank them enough.

I would like to thank the Director, Dr. Charles Biney, Dr. W. E. I. Andah and the rest of the staff of my home institute, CSIR-Water Research Institute, for all the support I received during this study. A lot of thanks to Mr. Wellens-Mensah, Director, Mr. Clottey, the Northern Regional Hydrologists and the technical staff of the Hydrological Services Department of Ghana, for assisting me in my field activities and providing me the daily stream flow data I asked for.

I should also say thank you to my uncles Bobena Yaro, Charles Laoni and Ben Nobiya, to Dr. and Mrs. Albert Tenga and to my close friend Thomas Akabzaa for their support and prayers, especially those rendered to my family while I was abroad.

Finally, special thanks to my wife, Hawa, and daughter, Mmalebna for their sacrifices, support and prayers and for giving me all the reasons to undertake this study.