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.

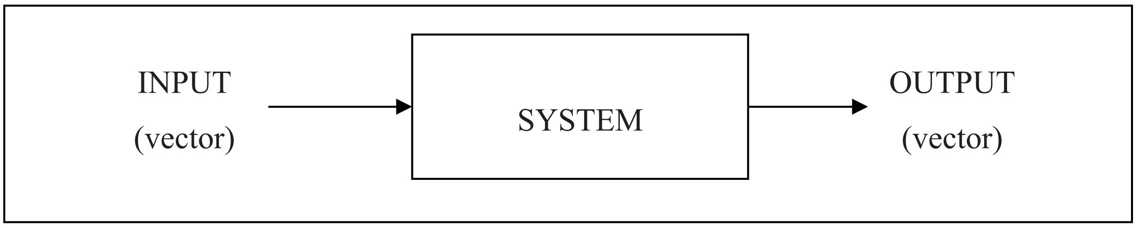

Figure 1.1 A system as a basic input-output element

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).

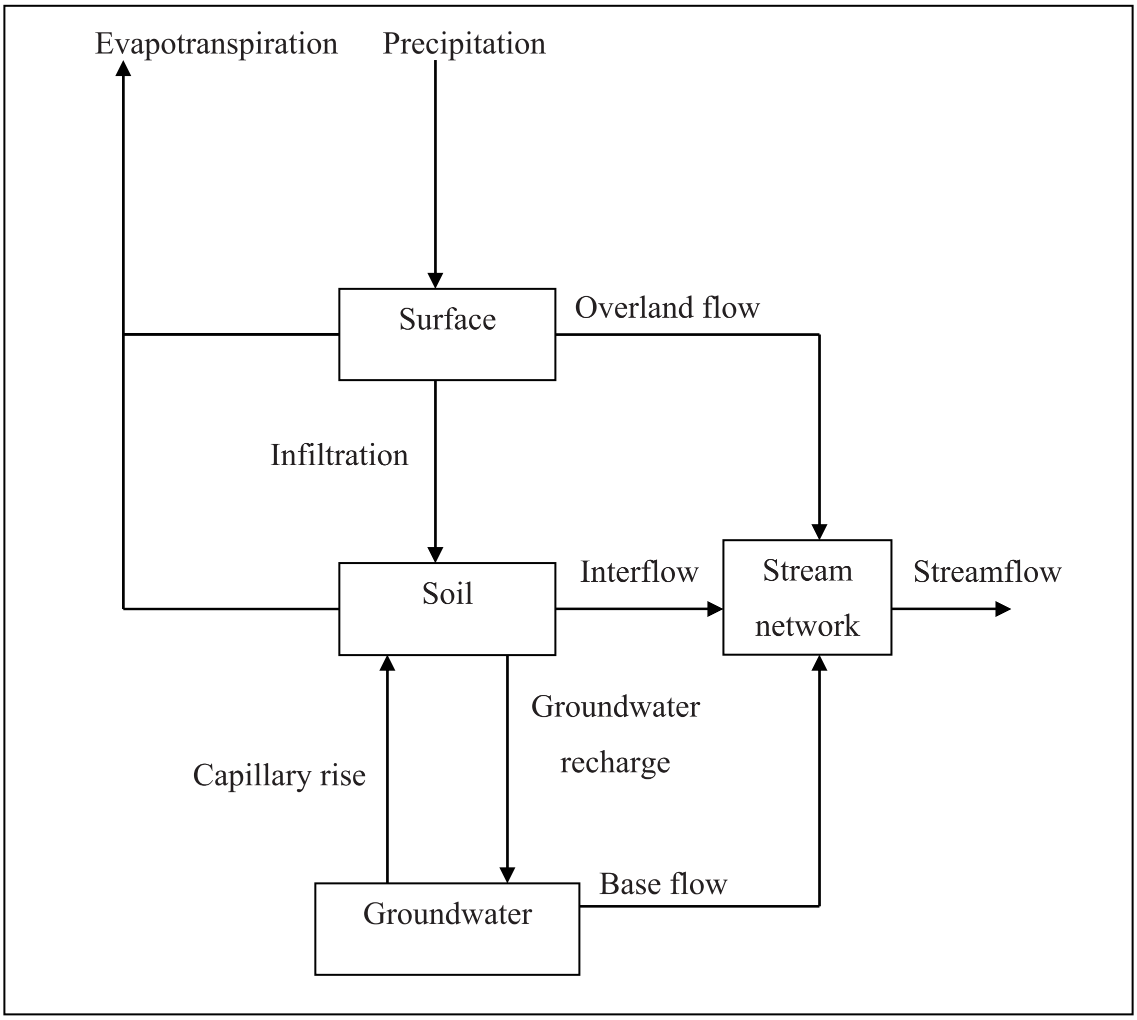

Figure 1.2 River catchment system (modified from Dooge, 2003)

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