CHAPTER 7

ENTROPIC MEASURES FOR COMPARING FLOW SIMULATION MODELS

7.1 INTRODUCTION

Hydrological data (e.g. rainfall, river flow, etc) are used in water resources planning and management for planning reservoir and operation. However, it happens sometime that the appropriate site where a reservoir should be built has no available data due for example to inaccessibility to erect a flow gauging station. Bedford site where a dam has to be built in future does not make an exception. Physical models, semi-distributed models, statistical models, conceptual model, embracing probabilistic, fitting curve, black box, etc are often used to estimate flows. Thus in this chapter, two simulation models namely RAFLER and WRSM2000 (incorporating physical characteristics of the catchment area) are applied to rainfall data of the nearby stations to estimate the flow data series at Bedford where absolutely no hydrological data exist. The two models are assessed using entropy concept. Empirical comparison of the predictive accuracy, in terms of reduction of the uncertainty of flows (e.g. entropy at Bedford before and after applying any model) is then made (Ilunga and Stephenson, 2003b).

7.2 ENTROPY APPROACH AS A HYDROLOGICAL MODEL PERFORMANCE CRITERION

Amorocho and Espildora (1973) and Singh and Fiorentino (1992) suggested that the mutual information (between the observed values and the simulated ones) could be used as entropy criterion in the selection of hydrological models; e.g. rainfall-runoff prediction. Later, the directional information transfer index (DIT) appeared as a generalization of the mutual information and was used for streamflow network design (Yang and Burn, 1994). The DIT notion was only defined for streamflow gauging station pairs. Chapman proposed a more general criterion of model performance for data set, as being the ratio of the transinformation to the marginal entropy. Recently, it is argued that since the mutual information is used for model performance assessment, its generalization i.e. DIT can be extended to model performance evaluation (Ilunga and Stephenson, 2002a, 2003b).

The above considerations are valid when the estimated values are compared to the observed ones. In that respect, statistical criteria such as root mean square error, etc can be also used to crosscheck the results (Ilunga and Stephenson, 2002b, 2003a). However, it becomes difficult to be based on these considerations when missing values are encountered in the data series. Thus Panu (1992) introduced the notion of reduction of uncertainty of the hydrological variable before and after infilling the data series. Goodier and Panu (1994) used the same approach. The reduction of uncertainty at a given site as defined by Panu (1992) can be given as follows (refer to equation 2.90, Chapter 2):

$$\mathrm{Re}\, d(\%) = 100((H_{cc} - H_{comp}) / H_{cc}))$$

where H_{cc} and H_{comp} are entropy values before and after infilling the data series respectively. It should be noted that this concept was applied to cases of consecutive missing data values, e.g. hydrological data exist before and after the missing values.

In this chapter, the same expression, i.e. 2.90 is used for a case where no available flow data exist at all at the site (e.g. Bedford). It is more natural to say that in a case where no data is available, the uncertainty is higher than in a case where data exist. In other words, the amount of chaos or lack of information (ignorance) about a system is higher in the former case than in the latter one.

It is assumed that the uncertainty should be maximum (e.g. if all hydrological events would have occurred equally likely) in the sense of Amorocho and Espildora (1973).

The maximum uncertainty (i.e. maximum entropy) was given by expression 2.4 (refer to Chapter 2) as follows:

$$H_{\max}\left(X\right) = \log n$$

Thus, at a site where no hydrological data is known, the entropy can be set to that maximum as defined the expression above. This maximum entropy does not depend on the magnitude of the hydrological events (streamflow) at the site (e.g. Bedford), but it is only a function of the sample size (length of record n). Although the magnitude of the hydrological event is unknown at the site, its sample size will be assumed to be the one used concurrently for the simulation models as given briefly in the following section.

Thus, in this case, expression (2.90) can be re-written as:

$$\operatorname{Re} d(\%) = 100((H_{\max} - H_{comp}) / H_{\max})) \qquad (7.1)$$

7.3 SHORT NOTE ON RAFLER AND WRSM2000 MODELS

RAFLER is an acronym for Rainfall Flow Erosion. A model (RAFLER) is a deterministic model based on the physics of runoff, soil infiltration and soil transport and which converts rainfall data to runoff over a length of time, e.g. years. The model uses monthly rainfall figures to reproduce monthly stream flow series and soil erosion. Some simplification is made to enable the model to be run with a minimum of data. And the rainfall period each month is estimated from the number of rain days to enable true flow rates to be calculated. This model requires a number of modules including catchment, channels and reservoirs. The general theoretical background of the model can be traced in Stephenson (2002).

WRSM2000 is an acronym for the Water Resources Simulation Model for windows 2000. The WRSM2000 is a monthly model to produce runoff from a catchment. This model solves a problem allowing for a record period of up to 150 years. The WRSM2000 is new version of the WRSM90 with an old DOS network used. WRSM90 is an enhancement of the HYP09 program, which was also an enhancement of the first computer program, MORSIM. For more details, the reader is referred to Pitman et al. (2000). This model has basically fours different modules, viz. the runoff submodel, the channel reach submodel, the reservoir submodel and the irrigation submodel. The general theoretical background of the model can be traced in Pitman (1973).

7.4 STUDY AREA AND DATA AVAILABILITY

Bedford is situated in the Free State, in South Africa. The catchment area is about 10 km2. Neither rainfall data nor stream flow data is available at this particular site. It was possible to simulate flows at Bedford using rainfall data from the nearby sites; viz at Van Reen (MAP = 1002 mm/month); at Moorside (MAP = 839 mm/month) and at Baldergow (MAP = 887mm/month). The monthly rainfall data (1920-1989) were obtained from the Weather Bureau, South Africa.

7.5 MODEL PERFORMANCE EVALUATION FOR SIMULATED FLOWS AT BEDFORD SITE

The application of the two simulation models, i. e. RAFLER and WRSM2000 to simulate the total annual flows (from 1920-1989, e.g. 70 data points) at Bedford site gave the following results. Figure 1 below gives the estimated annual hydrographs at Bedford from the two models. Table 1 gives the values of different statistical parameters. Based on results in Figure 1 and Table 1, it is difficult to decide whether these two models can be used for flow simulation at Bedford (or which model can perform better than the other one). However, the respective values of different statistical parameters (obtained from the two models respectively) do not differ very much from one to the other.

Figures 2 and 3 show the probability (frequency) distribution estimated from the 2 models respectively. The number of class intervals was defined using the following expression (Yevjevich, 1972)

$$m = 1 + 1.33 * \log(number\ of\ data\ po\mathrm{int}\,s) \qquad (7.2)$$

As no flow record is yet known at that gauging station, it was assumed that the uncertainty of the hydrological variable (i.e. total annual flows) was very high. The value of this uncertainty (entropy) was set to the possible maximum value in the sense defined so far, e.g. the uncertainty would be the natural log of the sample size of the hydrological events (flows). So the possible maximum entropy as given by Amorocho and Epilsdora (1973) does not depend on the magnitude of the hydrological events; it is only a function of the sample size. The sample (or length of records, which are unknown) was assumed to be the one used concurrently by the 2 simulation models (e.g. sample size is 70).

Table 7.1 Simulated statistical parameters at Bedford

Model description	Mean	Standard deviation	Coefficient of variance
WRSM2000	3.21	1.27	0.39
RAFLER	2.35	0.61	0.26

Table 7.2 Model performance evaluation at Bedford

Model Description	Maginal entropy (napiers) at Bedford	Reduction of uncertainty (%) at Bedford
Before applying any model	4.25	0
WRSM2000	1.53	63.92
RAFLER	1.09	74.31

Table 7.2 shows the results of entropy calculations before and after applying the two models. It is therefore concluded that the values of reduction in uncertainty of the total annual flows at site Bedford were 63.92 % and 74.31 % by applying WRSM2000 and RAFLER models respectively. These values are the equivalent of information inferred about Bedford using the two models respectively. With a threshold value of 50% for the reduction of uncertainty at Bedford, both models are thought to perform well. Recall that the statistical parameters are not quite different from one another for the two data series simulated by the two models. Thus, they could be used for flow prediction at that station with regard to the annual total flows. However, RAFLER model performed better than WRSM2000 model for this specific flow regime. Nonetheless these two models need to be tested on other flow regimes for that specific site.

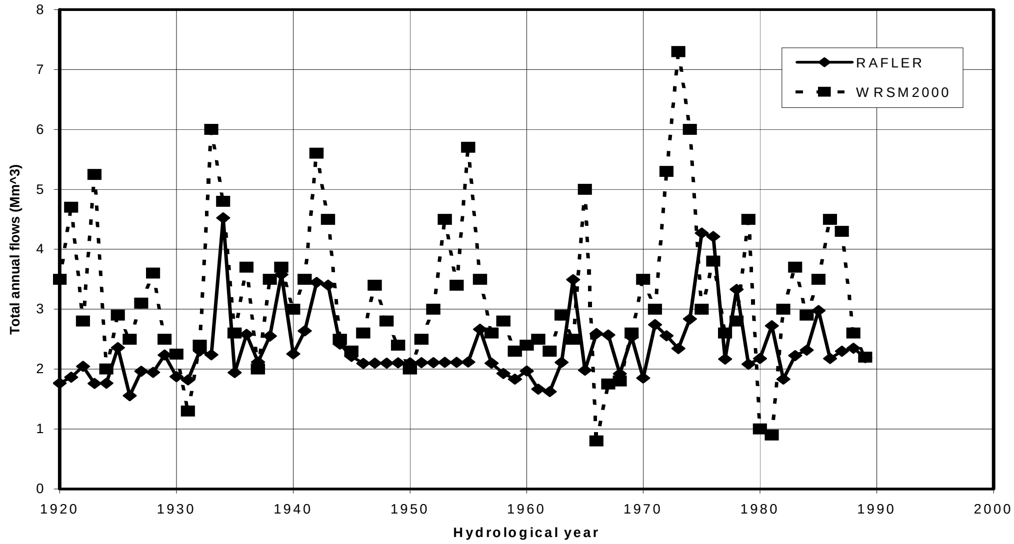

Figure 7.1 Simulated annual hydrographs at Bedford.

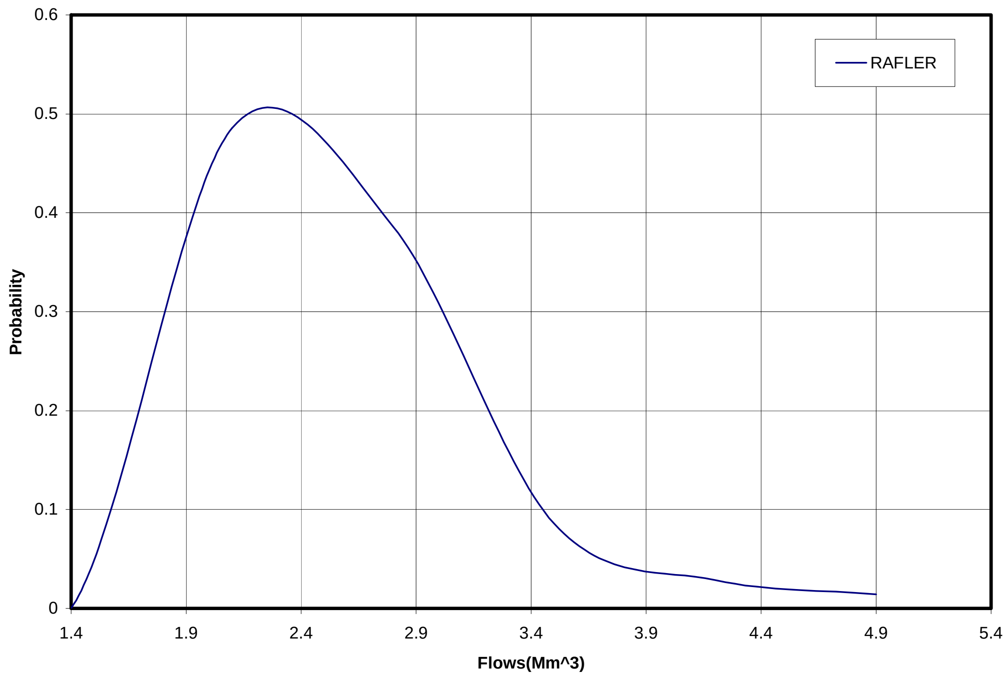

Figure 7.2. Simulated annual probability distribution at Bedford (using RAFLER).

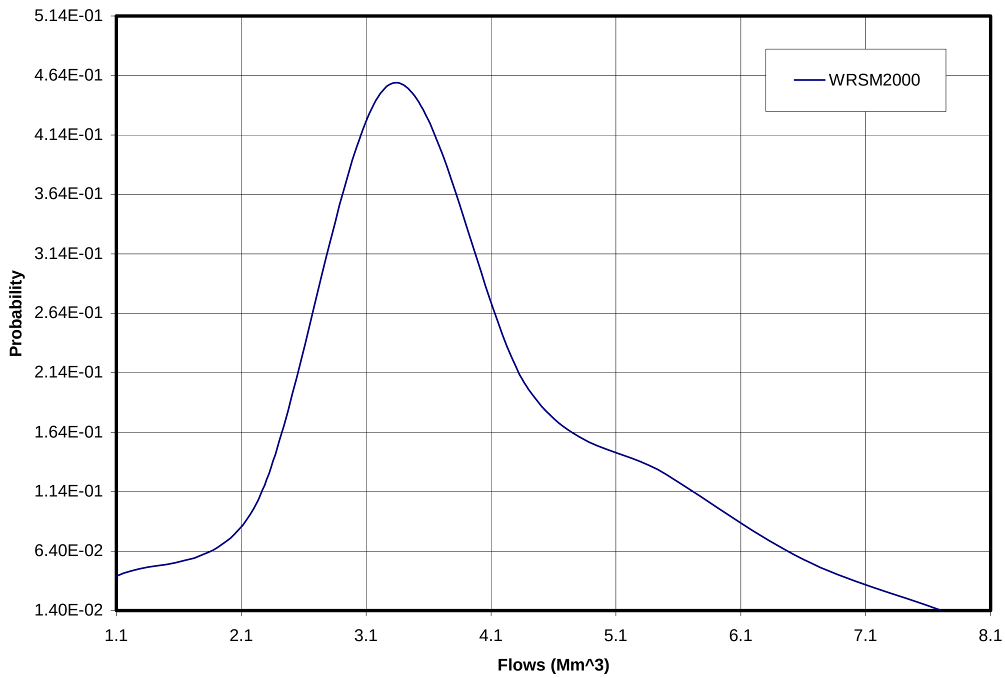

Figure 7.3 Simulated annual probability distribution at BedFord (using WRSM2000)

It should noticed that the concept as outlined in this chapter was also applied to Braamhoek (in the Free State, in South Africa) with only RAFLER model (Ilunga and Stephenson, 2004). The results were satisfactory.

7.5 SUMMARY

The focus of this chapter was to evaluate the performance of the two models, viz. RAFLER and WRSM2000 at Bedford, using entropy approach. The criterion used here is very similar to Panu (1992) who used it for infilling data flow series; however this criterion is extended to a case (e.g. Bedford) where absolutely no flow records exist at all. The computations from the entropy criterion showed that both models could be used for simulating the annual total flows at Bedford when a threshold value of 50% is considered for the reduction of uncertainty before and after applying the simulation models. Nonetheless, RAFLER model could perform better (than WRSM2000) when considering the annual total flows.